

Cornelius Fritz

Statistical Approaches to Dynamic Networks in Society

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 25.05.2022



Cornelius Fritz

Statistical Approaches to Dynamic Networks in Society

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 25.05.2022

Erster Berichterstatter: Prof. Dr. Göran Kauermann (LMU Munich)
Zweiter Berichterstatter: Prof. Dr. Paul W. Thurner (LMU Munich)
Dritter Berichterstatter: Prof. Dr. David R. Hunter (Pennsylvania State University)

Tag der Disputation: 27.07.2022

Acknowledgments

First and foremost, I thank Göran Kauermann for being a great supervisor and giving me both the freedom and guidance I needed during the work on this dissertation. Working together with you has not only taught me a lot but was also truly fun. Besides, I want to thank you for consistently supporting me and having my back during all my endeavors from when I started working for you during my bachelor years as a student assistant. This trust made me put additional care into my work to show that your confidence was, at least to some point, legitimate. I also want to express my gratitude to Paul Thurner for introducing me to political network analysis, which led to many interesting joint papers and fostered my motivation for interdisciplinary research.

Next, I want to thank David Hunter for agreeing to be the external reviewer of my dissertation. The first talk I ever got to hear about statistical network data analysis was given by you, and much of this thesis builds directly on your work. Further, I want to thank Christian Heumann and Helmut Küchenhoff for agreeing to steer the examination committee.

I believe that, at its best, research is a collaborative effort; thus, I now want to thank my collaborators. I thank Giacomo De Nicola for being the objectively speaking best office mate and always having an open ear for some research idea or general remark that had to get out of my head. For all the tremendous substantive input and evening beers, I am grateful to Marius Mehrl. Our joint projects made me believe in synergy effects and the need for interdisciplinary research. Additionally, I want to thank my two former seminar supervisors, Michael Lebacher and David Rügamer. Both of you showed me during my Master's studies how interesting scientific work can be and paved the road to this dissertation. In particular, I want to thank Michael for our never-ending discussions on the foundations of statistical inference and the future of data analysis. Furthermore, I want to thank the COVID-19 data analysis group (CODAG) for the great discussions and joint work in the last two years. The pandemic is the elephant in the room of this dissertation, but this vital work showed me the impact our work as statisticians can have during such challenging times.

Through great tea talks, lunch gatherings, summer retreats, football games, and a generally optimistic workplace atmosphere, the department of statistics contributed tremendously to this work. From our chair, I will always be grateful for all the impromptu 45-minutes chat breaks with Benjamin Sischka and Martje Rave, the understanding of Marc Schneble for my lack of understanding of poker, the great times in the Staatskanzlei Saarland with Michael Windmann, the entertaining working hours for Data Science with Constanze Schmaling, and the multiple C++ crash courses of Nurzhan Sapargali. Next, my thanks go to Alexander Bauer and Max Weigert from the Stablab for advice on the latest vegan restaurants in Munich and a convincing tour through Augsburg, respectively.

To not forget the world outside of the department of statistics, I want to thank the Gesauft Gruppe, the Schöner Leben and Gaudi WG, and all my other friends. Finally, this thesis would not have been possible without the unlimited support of my family. All the walks, talks, dinners, and visits to my parents and brother were vital to this work. I couldn't have done this without your help.

Summary

Networks encode relational structures between entities that do not generally abide by the conditional independence assumption employed in most statistical models. In particular, network analysis has gained traction in the past decades in the Social Sciences, where many applications originate. This cumulative dissertation is dedicated to studying dynamic networks with a focus on applications within the Social Sciences. The thesis is divided into four parts, with the first part providing the background on the contributing manuscripts and putting them into a general context. Each subsequent part is composed of two articles.

More specifically, the second part provides an overview of methods to model dynamic networks as well as a substantively meaningful example of using dynamic networks as covariates to study infections during the first wave of the COVID-19 pandemic. The first article introduces two modeling frameworks, one based on Markov chains to study networks observed at discrete time points and another one building on counting processes if the network is continuously monitored in a fine-grained temporal resolution. We showcase the available methods, software implementations, and their applicability and interpretation for each framework by two data examples. The second article uses dynamic spatial and weighted networks as covariates to investigate how regional mobility and social connectivity affect COVID-19 infections in Germany.

The third part focuses on studying networks observed at discrete points in time as the outcome of a Markov chain. In the first article encompassed in this part, building on the theorization and description of signed networks, starting with the structural balance theory of [Heider \(1946\)](#), and the Temporal Exponential Random Graph model of [Hanneke et al. \(2010\)](#) the Signed Exponential Random Graph Model (SERGM) for the study of dynamic signed networks is introduced. With the theoretical foundation of structural balance theory in mind, novel simultaneous statistics are proposed that provide better performance than operationalizing them by lagged covariates, as commonly done by other authors. The second article examines the co-inventorship of patents viewed as a bipartite network. Here one mode comprises the inventors, and the second mode is composed of all patents on which the inventors collaborate. In particular, we propose a bipartite variant of the TERGM with varying actor compositions, differentiating between inventors that already submitted patents and those that did not while accounting for pairwise statistics of inventors.

Finally, in the fourth part of the dissertation, we analyze event data observed in continuous and discrete time. Within the first article, a tie-oriented model for longitudinal event network data is proposed to explore the international trade of combat aircraft. Motivated by the observation that automated or human-coded events often suffer from non-negligible false-discovery rates in event identification, the second article offers the Relational Event Model for Spurious Events (REMSE) as a flexible solution for modeling data while controlling for spurious events. Moreover, it is possible to use the REMSE to assess the robustness of any Relational Event Model specification for the studied event data.

Zusammenfassung

Netze kodieren Beziehungen zwischen Akteuren, welche im Allgemeinen nicht bedingt unabhängig voneinander sind. Diese Annahme liegt den meisten statistischen Modellen zugrunde. In dem Großteil der praktischen Anwendungen ist es natürlich, Netzwerke als sich im Laufe der Zeit entwickelnde Systeme zu betrachten. Vor allem in den Sozialwissenschaften, wo viele Anwendungen ihren Ursprung haben, hat die Netzwerkanalyse in den letzten Jahrzehnten an Bedeutung gewonnen. Diese kumulative Dissertation beschäftigt sich mit der Untersuchung von dynamischen Netzwerken mit einem Schwerpunkt auf Anwendungen in den Sozialwissenschaften. Die Arbeit ist in vier Teile gegliedert, wobei der erste Teil den Hintergrund zu den Beiträgen liefert und diese in einen allgemeinen Kontext einordnet. Jeder weitere Teil besteht jeweils aus zwei Artikeln.

Im zweiten Teil wird nach einem Überblick über Methoden zur Modellierung dynamischer Netzwerke ein aussagekräftiges Beispiel für die Verwendung dynamischer Netzwerke als Kovariaten zur Untersuchung von Infektionen während der ersten Welle der COVID-19-Pandemie vorgestellt. Der erste Artikel stellt zwei Modellierungsverfahren vor, eines auf der Grundlage von Markov-Ketten zur Untersuchung von Netzwerken, die zu diskreten Zeitpunkten beobachtet werden, und ein anderes, das auf Zählprozessen basiert, wenn das Netzwerk kontinuierlich in einer feinkörnigen zeitlichen Auflösung beobachtet wird. Anhand von zwei Datenbeispielen werden die gebotenen Methoden, Softwareimplementierungen sowie deren Anwendbarkeit und Interpretation für die beiden Verfahren vorgestellt. Der zweite Artikel nützt dynamische und räumliche Netzwerke, sowie solche mit gewichteten Kanten als Kovariaten, um zu ermitteln, wie regionale Mobilität und soziale Konnektivität die COVID-19-Infektionen in Deutschland beeinflussen.

Der dritte Teil konzentriert sich auf die Untersuchung von Netzwerken, die zu diskreten Zeitpunkten als Ausgang einer Markov-Kette untersucht werden. Der erste Artikel dieses Teils baut auf der Theoretisierung und Beschreibung von Netzwerken mit positiven und negativen Kanten auf. Basierend auf der strukturellen Balancetheorie von [Heider \(1946\)](#) und dem Temporal Exponential Random Graph Model von [Hanneke et al. \(2010\)](#) wird das Signed Exponential Random Graph Modell (SERGM) zur Untersuchung dynamischer Netzwerke mit positiven und negativen Kanten vorgestellt. Auf der theoretischen Grundlage der strukturellen Gleichgewichtstheorie werden simultane Statistiken vorgeschlagen, die eine bessere Performanz bieten als die Operationalisierung von anderen Autoren. Der zweite Artikel untersucht die Ko-Erfinderschaft von Patenten als bipartites Netzwerk. Dabei besteht das erste Set von Akteuren aus den Erfindern, das zweite aus den Patenten, an denen die Erfinder zusammenarbeiten können. Dabei wird eine bipartite Variante des TERGM mit einer sich ändernden Akteurszusammensetzung eingeführt, welche zwischen Erfindern, die bereits Patente angemeldet haben, und solchen, die dies nicht getan haben, unterschieden wird. Gleichzeitig werden paarweise Statistiken der Erfinder berücksichtigt.

Im vierten Teil der Dissertation werden schließlich Eventdaten analysiert, die in kontinuierlicher und diskreter Zeit erfasst wurden. Der erste Artikel schlägt ein verbindungsorientiertes Modell für longitudinale Eventdaten vor, um den internationalen Handel mit Kampfflugzeugen zu beschreiben. Motiviert durch die Beobachtung, dass automatisierte oder menschlich kodierte Events oft zu falsch-positiven Events führen, bietet der zweite Artikel das Relational Event Model for Spurious Events (REMSE) als flexible Lösung für die Modellierung von Daten unter Kontrolle von diesen fehlerhaften Events. Darüber hinaus ist es möglich, das REMSE zu verwenden, um die Robustheit einer beliebigen Spezifikation des REMs zur Modellierung von bestimmten Eventdaten zu überprüfen.

Contents

I. Introduction and Background	1
1. Overview	1
2. Formalization of Network Data	5
2.1. From Networks to Graphs	5
2.2. From Interactions to Counting Processes	9
3. Network Models	11
3.1. Exponential Random Graph Model	11
3.2. Relational Event Model	18
4. Network Attribute Models	25
5. Discussion and Future Work	29
Bibliography	33
II. Statistical Modeling of Networks	43
6. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time	45
7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany	85
III. Temporal Exponential Random Graph Models	127
8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations	129
9. Modelling the large and dynamically growing bipartite network of German patents and inventors	169
IV. Relational Event Models	199
10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades	201
11. All that Glitters is not Gold: Modeling Relational Events with Spurious Events	239
Contributing Publications	265
Eidesstattliche Versicherung (Affidavit)	267

Part I.

Introduction and Background

1. Overview

“As statisticians, we should know what we are talking about and should care that what we say is true, in the sense of agreeing with phenomena in the real world. If we statisticians are to become statistical scientists we must become thoroughly familiar with the processes of science.”

J. A. Nelder

This quote by Nelder (1999, p. 258) highlights the vital connection between statistical models and real-world phenomena. To ensure that statistical models mirror real-world phenomena, we have to “*play in everyone’s backyard*” as John Wilder Tukey put it so famously. Kass (2011) even argued for substituting the classical view of statistical inference based on drawing inference from a sample on some unreachable population for a “*pragmatist*” view trying to best align the “*real*” and “*theoretical world*”, which can be seen as synonyms for the worlds where statistical models and real-world phenomena live in. Along these lines, one might describe the statistician’s job as a process starting with the collection of data and the theories posed within the respective substantive sciences. Consecutively, we formalize this data and theory with the help of mathematical objects. In this “*theoretical world*” (Kass, 2011), we define a statistical model which is characterized by random variables and parameters that we can, in turn, estimate from the observed data. In the last step, we trace the estimated parameters back to a substantially meaningful interpretation. As described in Box (1976), this process is iterative and model criticism is used to ascertain whether the model is a good fit to the observed data or if further loops, i.e., changes to the model itself or its specification, are required. This dissertation consists of six manuscripts with the aim to show how “*statistics*” or “*statistical science*” (as Nelder, 1999 would have put it) can assist in analyzing data collected in the Social Sciences. The type of data we are predominantly concerned with is network data, as “*networks are ubiquitous in science and have become a focal point for discussion in everyday life*” (Goldenberg et al., 2010, p.129). Most networks are also emergent of a dynamic process and seldom are inherently static since “[*e*]very social fact is situated, surrounded by other contextual facts and brought into being by a process relating it to past contexts” (Abbott, 1997, p.1152). Therefore, our main focus lies on the analysis of temporally evolving networks.

The Perspective of Social Network Data Analysis Network data analysis in the Social Sciences has gained traction in the past decades (Borgatti et al., 2009). The field can be traced back to Moreno (1934), who investigated the remarkably high number of runaways of students from a girl school. During this project, he raised concerns about the independence of each separate runaway and introduced the “*sociogram*”, which later evolved into graph drawing and the adjacency matrix, to uncover that the embedding of the girls in a social space leads to clustered outbreaks. Suppose one were to apply standard statistical models, e.g., logistic regression, to the binary indicator of whether a specific girl ran away or not. Moreno’s theory violates the conditional independence assumption inherent to the classic regression model, rendering its mere application

questionable. This observation embodies a core concept of the social network perspective in that attributes of individuals of a study are not viewed as autonomous but “*arising out of structural or relational processes*” (Wasserman and Faust, 1994, p.7). As a result, the central differentiating characteristic of network theories is that the relations between the studied subjects are a vital part of them. More abstractly, the question Moreno (1934) brought up can be stated as follows: How is individual behavior affected by the behavior of other actors in the same network and the network itself? A term coined in the Social Sciences for this type of mechanism are “*influence*” processes, whereas “*selection*” concerns processes in which behavior affects the relations between actors in the network. We are, therefore, also interested in studying how a particular network can be the product of endogenous and exogenous factors. These two types of canonical problems tackled in social network analysis provide a clear division visible in the whole thesis.

Statistical Network Analysis From the statistical perspective, the data example of Moreno (1934) leading to the area of sociometry and later social network analysis requires methods that adequately account for interdependencies between the studied relationships. While dependencies are often treated as technical issues when applying statistical models in, e.g., longitudinal settings (Zeger and Qaqish, 1988), understanding instead of merely controlling for these dependencies is core to network theories as well as the statistical network analysis methods used to test hypotheses originating from these theories (Robins and Pattison, 2005). In a series of articles (Goldenberg et al., 2010; Kim et al., 2018; Butts, 2008) and books (Kolaczyk, 2009, 2017; Kolaczyk et al., 2009; Cranmer et al., 2021; Carrington et al., 2005), thorough treatments of these methods as well as general introductions to the field are provided. Under the risk of oversimplification, one can differentiate these modeling endeavors along one dimension measuring how explicit the dependence between relations is stated in the model. On the one end of this spectrum is the Exponential Random Graph Model (ERGM, Robins et al., 1999), where the dependence assumptions are precisely stated and interwoven with the model specification. In the other limit case we get the Graphon model (for an introduction see Sischka and Kauermann, 2022), leading to a nonparametric representation of very complex structural patterns. Latent Space Models (Hoff et al., 2002), Infinite Relational Models (Kemp et al., 2006), Infinite Feature Models (Miller et al., 2009), p_2 Models (van Duijn et al., 2004), and Stochastic Actor-oriented Models (Snijders, 1996) can be placed somewhere between those two poles. While the books by Lusher et al. (2012) and Carrington et al. (2005) cover the ERGM, we refer to Matias and Robin (2014) and Kim et al. (2018) for an introduction to models based on node-specific latent variables.

Outline of the Introduction Within the first part of this dissertation, we provide an overview of this statistical toolbox of models for dynamic networks forming the foundation for the contributed manuscripts. In Chapter 2 we focus on two different mathematical formalizations of networks and show how they allow the representation of a taxonomy of networks and, in general, systems of interrelated entities. For dynamic networks, the settings differ in that we either observe the temporal networks only at specific time points or continuously in time. In the next step, Chapter 3 introduces approaches to model networks in the two respective settings. For the former situation, Section 3.1 focuses on Exponential Random Graph Models (ERGM, Robins et al. 1999), as this model class provides a unified and flexible framework for testing and modeling complex dependencies with theoretical underpinnings of the available exponential family theory (Schweinberger et al., 2020; Barndorff-Nielsen, 1978). Conversely, Section 3.2 covers the framework for events introduced by Butts (2008) for the latter setting. Besides modeling the relations in networks

themselves, the second canonical task in social network analysis is to detect how actor attributes are affected by networks, which we cover in Chapter 4. To conclude, attention is drawn to some open issues that can serve as the basis of future research. In summary, the goal of the introduction is to provide the means to carry out the statistician's task outlined above applied to network data.

2. Formalization of Network Data

Following the seminal work of [Moreno \(1934\)](#) a novel scientific area called “*sociometry*” arose and sociologists ([Granovetter, 1973](#)) and anthropologists ([Bott, 1957](#)) started to use matrix algebra to formalize these network effects, eventually accumulating in what is now known as “*social network analysis*” ([Wasserman and Faust, 1994](#)). For this chapter, we build on this work and introduce a framework to represent network data as mathematical objects, based on which the statistical models introduced in subsequent chapters are defined. In [Section 2.1](#) we present a formalization for network data in which the relationships are temporally extensive, and entire networks are observed at specific points in time. [Section 2.2](#) instead suggests a framework to represent interactions by counting processes.

Some notational remarks; we use upper case letters for random variables, X , lower case letters for their observation, x , calligraphic letters for sets, intervals or tuples, \mathcal{X} , and write the letters in bold if they are matrices or vectors, \mathbf{x} .

2.1. From Networks to Graphs

Graphs and Networks In general, the term “*networks*” describes “*an abstract concept referring to a system of interrelated entities*” ([Crane, 2018](#)), while “*graphs*” are defined as particular mathematical objects used to represent certain characteristics of a “*network*”. After obtaining some network measurements, one can formalize the network as a graph. Formally, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of two sets: $\mathcal{V} = \{1, \dots, N\}$ is the set of actors for N actors (also called vertices or nodes in the literature), while $\mathcal{E} = \{(i_1, j_1), \dots, (i_M, j_M)\} \subseteq \mathcal{P} = \{(i, j); i, j \in \mathcal{V}\}$ with $|\mathcal{E}| = M$ is the set of M observed edges or ties. We call each element in \mathcal{P} a “*pair*” or “*relation*”. When working with multiple graphs simultaneously $\mathcal{V}(\mathcal{G})$ denotes the actor labels and $\mathcal{E}(\mathcal{G})$ the edges of graph \mathcal{G} . Examples of what those two sets can represent are provided over the course of this section. If $(i, j) \in \mathcal{E} \Rightarrow (j, i) \in \mathcal{E} \forall i \neq j$ with $i, j \in \mathcal{V}$ holds, the corresponding graph is “*undirected*”, while it is otherwise “*directed*”. Put differently, for “*directed*” graphs, the order of actors carries additional information, while this is not the case for “*undirected*” graphs. If not stated otherwise, self-loops cannot be observed, i.e., $\forall i \in \mathcal{V}, (i, i) \notin \mathcal{E}$. Mathematically, we can express a graph by its “*adjacency matrix*” $\mathbf{y} = (y_{ij})_{i,j=1,\dots,N}$ where the entries are given by

$$y_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{E} \\ 0, & \text{else} \end{cases} \quad (2.1)$$

Let \mathcal{Y} denote the set of all possible adjacency matrices between M actors that could be potentially observed. Henceforth, we refer to graph \mathcal{G} by \mathbf{y} . Returning to the example of [Moreno \(1934\)](#) on runaways in the Hudson School for Girls, the set $\mathcal{V} = \{1, \dots, 44\}$ encompasses $N = 44$ girls in the school, and the $M = 71$ edges in \mathcal{E} are directed and represent feelings of attraction between these girls. For unconstrained directed networks without loops of 44 actors, the size of \mathcal{Y} is $2^{44(44-1)}$,

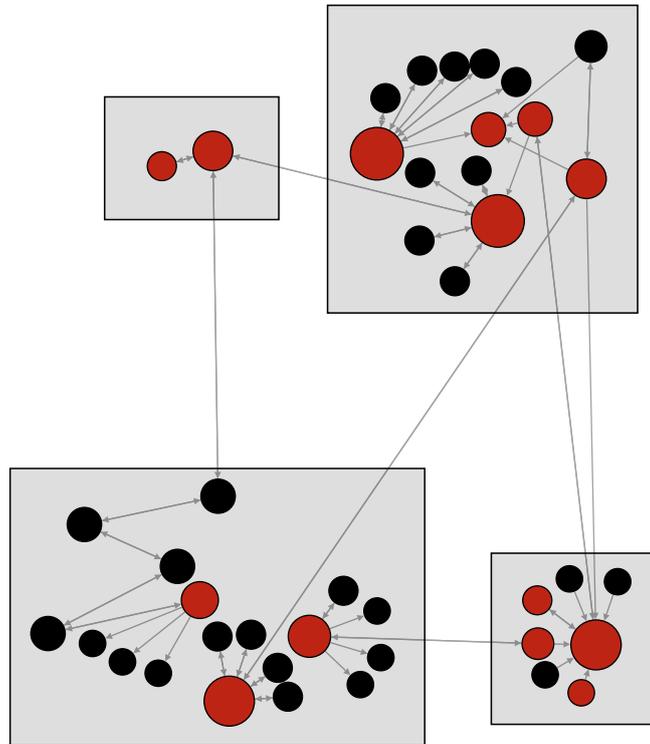


Figure 2.1.: Network indicating the feelings of attraction between pupils at Hudson School for Girls. The color of nodes is red if the respective girl was among the runaways and black otherwise. Additionally, the rectangles represent the cottages in which the girls lived. The size of the nodes is given by the sum of in- and out-degree. This figure is inspired by Figure 1 in [Borgatti et al. \(2009\)](#).

which is already astronomically large. Graph drawing helps to visualize the network on a two-dimensional plane by, e.g., applying algorithms proposed by [Fruchterman and Reingold \(1991\)](#) or [Brandes and Pich \(2007\)](#). Figure 2.1 depicts the network emerging from feelings of attraction between pupils at Hudson School for Girls, where pupils are represented as nodes and attraction between them as directed edges.

Descriptive Network Statistics This basic notion of graphs already allows to measure abstract concepts such as centrality and cohesion. To begin, one may quantify centrality on the level of actors in a directed graph as their in- and out-degree. The degree statistics are in the case of

2.1 From Networks to Graphs

directed networks given by the number of in- and out-going ties, respectively, and can be stated in terms of \mathbf{y} for actor i and $i = 1, \dots, N$:

$$\begin{aligned} s_{i,\text{Out}}(\mathbf{y}) &= \sum_{j \neq i} y_{ij} \\ s_{i,\text{In}}(\mathbf{y}) &= \sum_{j \neq i} y_{ji}. \end{aligned} \tag{2.2}$$

The related total degree of this actor i is given by $s_i^{\text{total}}(\mathbf{y}) = s_{i,\text{Out}}(\mathbf{y}) + s_{i,\text{In}}(\mathbf{y})$, which is captured by the size of the nodes in Figure 2.1. Freeman (1978) proposed a general technique to aggregate this type of micro level centrality index to the macro level and hence characterize the whole adjacency matrix \mathbf{y} . Let $S_{\text{Out}}(\mathbf{y}) = \max_{i=1, \dots, N} s_{i,\text{Out}}(\mathbf{y})$ be the maximum out-degree, then the out-degree centrality of \mathbf{y} is given by:

$$s_{\text{Out}}(\mathbf{y}) = \frac{\sum_{i=1}^N |S_{\text{Out}}(\mathbf{y}) - s_{i,\text{Out}}(\mathbf{y})|}{\max_{\tilde{\mathbf{y}} \in \mathcal{Y}} \sum_{i=1}^N |S_{\text{Out}}(\tilde{\mathbf{y}}) - s_{i,\text{Out}}(\tilde{\mathbf{y}})|}. \tag{2.3}$$

For this type of degree measure for directed graphs, the network $\tilde{\mathbf{y}}$ in the denominator is the star graph in which only one actor has outgoing edges to all other actors; hence the denominator can be reduced to $(N - 1)^2$. By substituting $s_{i,\text{Out}}(\mathbf{y})$ with any other actor-level measure, such as $s_{i,\text{In}}(\mathbf{y})$ or $s_{i,\text{Total}}(\mathbf{y})$, we obtain a rich class of “centrality” indices. Next to centrality, proposals to quantify the formation of cohesive subgroups are of vital interest in social network analysis. Cohesive subgroups are subsets of actors in \mathcal{V} among whom many edges are observed (Wasserman and Faust, 1994). The extent to which the entire graph is connected may be measured by counting the number of edges in \mathbf{y} ,

$$s_{\text{Edges}}(\mathbf{y}) = \sum_{i,j} y_{ij} = M, \tag{2.4}$$

which, when normalized with the number of pairs, that is $N(N - 1)$ for directed graphs, yields the density of \mathbf{y} . Based on the work of Luce and Perry (1949), we can further count the number of transitive triangles leading to the global clustering coefficient given by

$$s_{\text{Trans}}(\mathbf{y}) \propto \sum_{i,j,k} y_{ij} y_{ih} y_{hj}. \tag{2.5}$$

Normalizing (2.5) with the number of observed two-paths $\sum_{i,j,k} y_{ij} y_{hj}$, we get the clustering coefficient, which can also be viewed as the percentage of closed two-paths among all two-paths. For the network depicted in Figure 2.1, the density, i.e. proportion of realized edges, is 0.04 and the clustering coefficient 0.03 indicating a relatively sparse graph without a strong tendency to close triangles.

Data on the Edge, Pair and Node Level In most applications involving network data, auxiliary information is available one wants to incorporate into this framework. To represent data on the edge level, we define the set $\{w_e\}_{e \in \mathcal{E}}$. Graphs in which this information is present are often called “weighted graphs”; thus w_e is the weight of edge e . One instance for a weighted graphs results from the network comprising federal districts of Germany as nodes in which weighted edges indicate the proportion of Facebook friendships between users located in the respective districts, which we

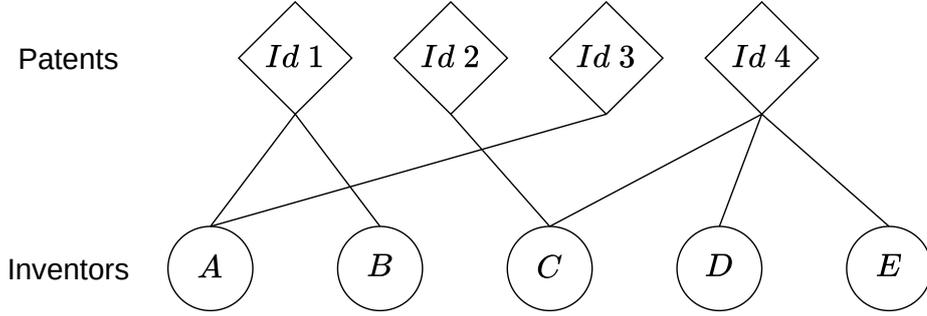


Figure 2.2.: Illustrative example of a bipartite network structure, where one mode consists of patents and another mode of inventors.

use in Chapter 7. Further, “*signed graphs*” are particular graphs studied in Chapter 8, where the edges can either be positive or negative; thus $w_e \in \{+, -\}$ holds. Sometimes, we are given further data $\{x_{ij}\}_{i,j=1,\dots,N}$ or $\{x_i\}_{i=1,\dots,N}$ for each pair or actor in \mathcal{V} , respectively. Often the latter type of data is termed nodal attribute or behavior. In the running example, these include the binary indicator of whether a girl ran away and the cottages where each girl lived. The color of the actors in Figure 2.1 conveys whether or not a girl ran away, and the rectangles provide information on the cottages in which they resided.

Bipartite Networks It is often not possible to observe the complete set of edges in the network. One common example where the restriction of observable ties has a particular structure are bipartite networks¹, where there are two types of actors, and edges are only possible between different types of actors. Hence \mathcal{V} can be split into \mathcal{V}_1 and \mathcal{V}_2 , and the space of possible ties is restricted to \mathcal{E} where $\mathcal{E} \subseteq \{(i, k); i \in \mathcal{V}_1, k \in \mathcal{V}_2\}$ holds. In Chapter 9, we investigate the driving forces behind collaboration and innovation through bipartite networks of inventors and patents. Thus, inventors are included in the set of actors in the first mode \mathcal{V}_1 , while the patents make up the set of actors \mathcal{V}_2 from the second mode. Figure 2.2 provides a schematic representation of such graphs, where inventors are labeled from *A* to *D* and patents from *Id 1* to 4. For instance, inventors *A* and *D* jointly author the patent *Id 1*.

Temporal Networks If one were to observe the year of submission for each patent, one could include this information as an additional actor mode leading to a tripartite network. The resulting network can be understood as a particular case of a temporal network in which the actors of the second mode (patents in the example above) are only active during one snapshot of the network. In the more general setting, temporal networks are observed at discrete points in time $1, \dots, T$ and represent them as separate adjacency matrices for each step, i.e., $\mathbf{y}_1, \dots, \mathbf{y}_T$. Technically, the actor sets $\mathcal{V}_1, \dots, \mathcal{V}_T$ can vary over time as well. Instances of temporal networks appearing in this thesis include the international trade of Major Conventional Weapons in 2016 and 2017 studied in Chapter 6 and the signed network consisting of Defense Cooperation Agreements (Kinne, 2020) and Militarized Interstate Disputes (Palmer et al., 2015) between 2000 and 2010 from Chapter 8.

¹Alternative names used in the literature for bipartite networks are two-mode or affiliation networks.

2.2 From Interactions to Counting Processes

These examples also clarify that actors in a network are not necessarily individuals but may be conglomerate entities, i.e., countries.

2.2. From Interactions to Counting Processes

Another characteristic of relations observed over time is their duration, i.e., how long they are active. Within the longitudinal setting of Section 2.1, relationships are implicitly assumed to be lasting at least for the time between each observation (Butts and Marcum, 2017). This simultaneity enables the states of pairs to be codependent and follows from a certain temporal extensiveness forming the basis of the presented “*graph-as-networks*” paradigm (Crane, 2018). However, if the duration of the studied relations becomes shorter, ties become actions, whereby the classic network structure becomes elusive and unstable. Ties, as conceptualized in Section (2.1), cannot be classified into being present and absent anymore but emerge from event recurrences. These recurrences are deduced from observed event sequences, where time is observed on a continuous scale (Vu et al., 2017). Data that behaves in this manner is increasingly available due to the automated collection of network data on social platforms and sensory data (Lazer et al., 2009). To represent such network data, an alternative formalization based on counting processes and relational events can be used. Following Fritz et al. (2020) and Fritz et al. (2021), we use the temporal indicator t if the temporal scale is discrete and \tilde{t} if it is continuous.

Relational Events The main building block representing such data is the “*relational event*”, which we denote by the tuple $a = (i, j, \tilde{t})$. Butts (2008) coined this term to encode pairwise actions, which we observe in a stream of M events $\mathcal{M} = \{a_1, \dots, a_M\}$. For $m = 1, \dots, M$, the m th entry a_m of \mathcal{A} provides information about an event from actor $i_m \in \mathcal{S} \subseteq \mathcal{V}$ to actor $j_m \in \mathcal{R} \subseteq \mathcal{V}$ at time point $\tilde{t}_m \in \mathcal{T}$. For the settings covered in this dissertation, it suffices to only regard events without loops, i.e., $i_m \neq j_m$ holds for $m = 1, \dots, M$. The observational period \mathcal{T} is assumed to span from 0 to the exogenously determined time point τ with $\tilde{t}_1 = 0$ and $\tilde{t}_m = \tau$, although relaxations are straightforward where needed. In most cases, the event times are not time clustered; thus, all event times are unique, and the events in \mathcal{A} are ordered chronologically, i.e., $\tilde{t}_m < \tilde{t}_{m+1}$ for $m = 1, \dots, M$. One example for event data are 21,635 emails sent between 1998 and 2002 among 156 coworkers of the Enron corporation investigated in Perry and Wolfe (2013). Here, an event represents a sent mail from coworker i to an associate j at time point \tilde{t} . Similarly to Section 2.1, an event is “*directed*” if the order of sender and receiver provides additional information and “*undirected*” otherwise. In the latter case, events are restricted such that $i < j$ holds. For instance, for Chapter 10 the deliveries of combat aircraft directed from one particular country i to another country j in year \tilde{t} are examined. In contrast, we study undirected combat events between rebel groups and state-based actors that occurred in the Syrian civil war in Chapter 11.

Characteristics of Events Weighted events with discrete or continuous weights can be represented by appending the value w to the event a . To give an example for discrete weights, Salathé et al. (2013) study the spread of positive and negative sentiments towards a novel vaccine expressed by tweets², whereas Lerner et al. (2013) and Brandes et al. (2009) model political events with continuous weights in the interval $[-10, 8.3]$ provided by the Kansas Event Data System (Schrodt

²One can also regard these observations as “*signed*” events.

et al., 1994). As already stated, events are, for the most part, assumed to be instantaneous; hence their duration is negligible. However, for enduring relationships, one can describe with a categorical weight w the start and end of a relationship by two different event types as discussed in (Stadtfeld et al., 2017). In the context of this dissertation, the analysis of pairwise events, i.e., events encompassing exactly one sender and receiver, is of particular interest. However, more complex scenarios with set-valued senders and/or receivers are covered in Perry and Wolfe (2013), Lerner et al. (2021) and Lerner and Lomi (2022). One example for such data is the original setting of the Enron mail events mentioned above, where the number of receivers of an email can be greater than one.

Exogenous Information Further information exogenous to the events is often available on the level of the pair or the node, which is denoted by the corresponding sets $\{x_{ij}\}_{i,j=1,\dots,N}$ and $\{x_i\}_{i=1,\dots,N}$. To give an example, Chapter 11 studies the co-location events in university housing while incorporating self-reported friendships between students as pairwise and the floor on which they live as nodal information.

Counting Processes Similar to the adjacency matrix from (2.1), all pairwise events can be represented in a matrix-valued counting process, whose state at $\tilde{t} \in \mathcal{T}$ is:

$$\mathbf{N}(\tilde{t}) = (N_{ij}(\tilde{t})|i, j = 1, \dots, N), \quad (2.6)$$

where $N_{ij}(\tilde{t})$ counts how many events have been observed between actors i and j in $[0, \tilde{t})$. Contrasting (2.1), the entries of (2.6) are positive integers. After observing event $a = (i, j, \tilde{t})$, the corresponding entry in $\mathbf{N}(\tilde{t})$ would increase by one, i.e. $N_{ij}(\tilde{t}) + 1 = N_{ij}(\tilde{t} + h)$ for $h \downarrow 0$. Thus, we assume that, in general, the realized counting process in \mathcal{T} is right-continuous with left-hand limits (Aalen et al., 2008). To accommodate for settings with a changing set of actors, we define the risk set encompassing all pairs at time \tilde{t} between which an event can happen by $\mathcal{U}(\tilde{t})$. The risk sets can change at all time points where events are observed, i.e., $\tilde{t}_1, \dots, \tilde{t}_m$, if they are exogenous to the stochastic events. These compositional changes occur in several settings, e.g., when certain actors drop out early, or in the application of Chapter 10, when countries can start but also stop existing. For completeness, let $\mathcal{U} = \bigcup_{m=1}^M \mathcal{U}(\tilde{t}_m)$ be the set of all possible relations. One way to guarantee the restriction to events without loops is to exclude the respective events from \mathcal{U} . Further we establish $\mathcal{H}(\tilde{t}) = \{\mathbf{N}(u); u < \tilde{t}\}$ as the entire past of the counting process from time point 0 up to but not including \tilde{t} . For brevity in notation, covariate processes of exogenous covariates resulting from the pairwise and nodal characteristics of events introduced in the previous paragraph are implicitly incorporated in $\mathcal{H}(\tilde{t})$.

3. Network Models

Returning to the statistician’s job described in Chapter 1, the step following the formalization of available data is modeling it within the constructed “*theoretical world*”. In Chapter 3, we introduce network models as probabilistic models for the data generating process of network data, while models for actor attributes are covered in Chapter 4. To start, Section 3.1 discusses ERGMs as a general family of models for network data that can be viewed from the graphs-as-networks perspective from Section 2.1. Next, Section 3.2 proposes REMs as a flexible solution for modeling event data detailed in Section 2.2.

3.1. Exponential Random Graph Model

Our first aim is to formulate a probabilistic model for the network \mathbf{Y} parameterized by θ abiding particular dependency between the studied relations. This model can be used to empirically test network theories that posit a particular dependency structure in the network while controlling for alternative structural mechanisms and exogenous information. Further, we would like to assess multiple competing theories simultaneously and use model selection criteria as a vehicle to decide between different specifications based on empirical evidence in the observed network. Moreover, by assuming that \mathbf{y} is the outcome of a stochastic process, we recognize that there is a certain structure in the network constraining the randomness of pairwise relations. Since Section 2.1 provides means to represent a whole taxonomy of networks, the proposed model should not be limited to static and binary networks. To achieve all these goals, we introduce the ERGM as a regression for network data (Lusher et al., 2012). Since the historical roots of ERGMs lie in the p_1 model proposed in Holland and Leinhardt (1981), some literature uses the alternative name p^* for the ERGM to signal that it generalizes the p_1 model (see Wasserman and Pattison, 1996, Robins et al., 1999, Pattison and Wasserman, 1999, or Wasserman et al., 2007). Next, we introduce dependence graphs to flexibly formulate dependencies between relations. Based on a predefined form of dependence, one can subsequently derive a distribution over adjacency matrices of binary and static networks. Subsequently, we discuss issues of interpretation of the parameters, degeneracy, and estimation. Finally, we generalize the model class to bipartite and temporal network data, because Chapters 6, 9, and 8 make use of and extend these models.

Dependence Graph Our aim is to specify a probabilistic model for the random variable $\mathbf{Y} = (Y_{ij})_{i,j=1,\dots,n}$, where each contained random pair might be conditionally dependent on other pairs. To formalize this dependency, we introduce the dependence graph³ \mathcal{D} , where $\mathcal{V}(\mathcal{D})$ consists of all pairs, i.e., $\{Y_{ij}; i \neq j \text{ and } i, j = 1, \dots, n\}$, and the edges in $\mathcal{E}(\mathcal{D})$ translate to a conditional dependence structure. If there is an edge between the nodes Y_{ij} and Y_{hk} , the corresponding pairs depend on one another conditional on the rest of the network. Conversely, this means

³Note that the closely related literature on graphical models calls this graph the “*independence graph*”.

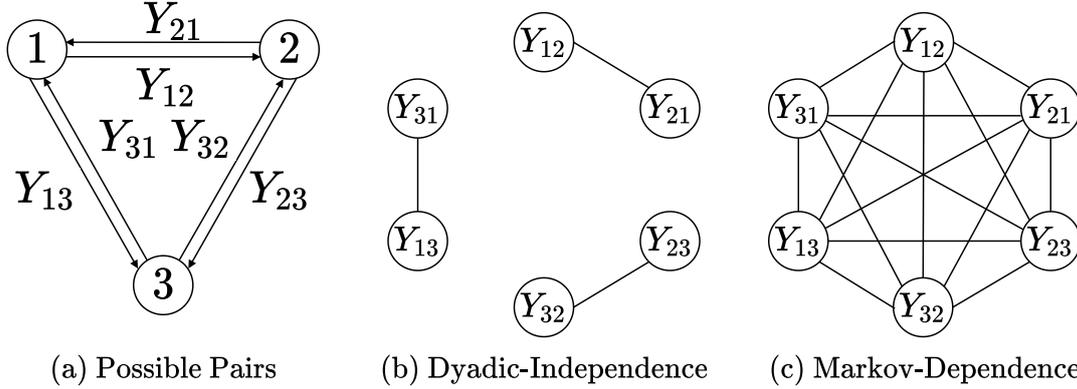


Figure 3.1.: Visualization of pairs between three actors (a) and the dependence graphs representing the dyad-independent (b) and Markov-dependence (c).

$(Y_{ij}, Y_{hk}) \notin \mathcal{E}(\mathcal{D})$ if and only if $Y_{ij} \perp Y_{hk}$ holds conditional on all other pairs in \mathbf{Y} (Lauritzen, 1996). We define a subgraph of \mathcal{D} to be a graph where $\mathcal{V}(\mathcal{D}) \subseteq \mathcal{V}(\mathcal{C})$ and $\mathcal{E}(\mathcal{D}) \subseteq \mathcal{E}(\mathcal{C})$ holds and a “clique” \mathcal{C} to be a subgraph of \mathcal{D} in which all possible edges are observed. In other words, cliques are subsets of nodes in \mathcal{D} that are assumed to be mutually dependent. For completeness, let \mathcal{A} be the set of all cliques in \mathcal{D} , which can be identified for any graph. The importance of this object will become clearer in the ensuing paragraph.

In general, dependence graphs are assumed to be undirected, nonrandom, and known, although the assumption of undirectedness is relaxed when exogenous covariates are included (see Robins et al., 2001). What dependence assumptions are reasonable is seldom self-evident and posits one of the main challenges when modeling networks, since too simple assumptions often imply an unrealistic data generating process while too complex assumptions can lead to unstable models and also be unrealistic for larger networks. An example of the specification of an ERGM for a larger network is provided in Chapter 9. We here provide increasingly complex examples of common dependence assumptions to guide this process and refer to Robins et al. (2007) for additional information on the topic. The *iid* regime, which most statistical models abide, forms the simplest dependence assumption called the Bernoulli assumption, where no edges are in $\mathcal{E}(\mathcal{D}) = \emptyset$. Under the dyad-independent assumption, each pair also depends on its reciprocal value (Holland and Leinhardt, 1981). Beyond dyadic-dependence, Frank and Strauss (1986) proposed Markov dependence graphs, where pairs are assumed to be independent if they do not share an actor, i.e., Y_{ij} is conditionally independent of Y_{hk} if $\{i, j\}$ and $\{h, k\}$ are disjoint sets. Figure 3.1 illustrates these two types of dependence graphs together with all possible pairs between three actors. In recent years, more complex dependency structures, such as partial conditional dependence (Pattison and Robins, 2002) or conditional dependence (Snijders et al., 2006; Hunter and Handcock, 2006), were proposed in the literature.

Model Formulation Given the dependence graph \mathcal{D} , the Hammersley-Clifford theorem (Hammersley and Clifford, 1971; Besag, 1974) establishes the joint distribution of all pairs in \mathbf{Y} in the following form:

$$\mathbb{P}_{\theta}(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\sum_{c \in \mathcal{A}} s_c(\mathbf{y})\theta_c\}}{\sum_{\tilde{\mathbf{y}} \in \mathcal{Y}} \exp\{\sum_{c \in \mathcal{A}} s_c(\tilde{\mathbf{y}})\theta_c\}}, \quad (3.1)$$

3.1 Exponential Random Graph Model

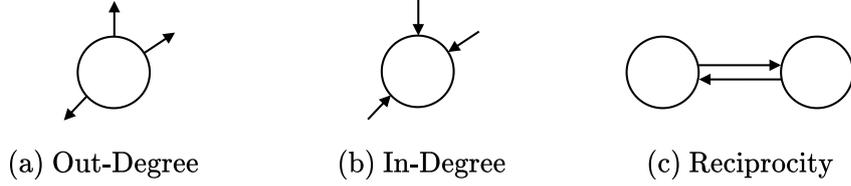


Figure 3.2.: Visualization of all sufficient statistics incorporated in the p_1 model. Circles represent actors and lines edges, arrow heads indicate the direction of the ties.

where the “sufficient statistics” are

$$s_{\mathcal{C}}(\mathbf{y}) = \prod_{(i,j) \in \mathcal{E}(\mathcal{C})} y_{ij}$$

and $\theta_{\mathcal{C}}$ for all \mathcal{C} in \mathcal{A} are the “natural parameters” determining the distribution (Frank and Strauss, 1986). On one hand, the Hammersley-Clifford theorem establishes the missing link between the dependence graph \mathcal{D} and the joint probability distribution of all pairs in \mathbf{Y} and, on the other hand, guarantees that one only has to worry about the set of cliques in \mathcal{D} as they are the sufficient statistics of model (3.1). Additional information on the derivation of ERGMs based on dependence graphs is provided in Wasserman and Robins (2005) and Robins and Pattison (2005), while Casella and Berger (2001) supplies a formal background on the concept of statistical sufficiency.

To make the abstract form of (3.1) concrete, consider the example of modeling a network with 3 actors under the the dyadic-independence assumption illustrated in Figure 3.1 (b). For this example, the set of cliques consists of each single pair and the three sets of reciprocal pairs, thus $\mathcal{A} = \{\{Y_{12}\}, \{Y_{21}\}, \{Y_{23}\}, \{Y_{32}\}, \{Y_{13}\}, \{Y_{31}\}, \{Y_{12}, Y_{21}\}, \{Y_{23}, Y_{32}\}, \{Y_{31}, Y_{13}\}\}$ and applying (3.1) gives

$$\mathbb{P}_{\theta}(\mathbf{Y} = \mathbf{y}) \propto \exp\{\theta_{12}y_{12} + \dots + \theta_{31}y_{31} + \theta_{12-21}y_{12}y_{21} + \theta_{23-32}y_{23}y_{32} + \theta_{31-13}y_{31}y_{13}\}. \quad (3.2)$$

Since model (3.2) includes 9 parameters to be estimated, it is not identifiable given a single realized network \mathbf{y} which only consists of 6 pairs. Note that heuristically the dependence between pairs translates to less information we can draw from a single realization of one pair from a larger network, thus determining the effective degrees of freedom of a network is still subject of current research (Krivitsky and Kolaczyk, 2015). To still obtain a parsimonious model, we assume that a priori all actors in the networks are indistinguishable from one another, which can be represented by a “homogeneity condition” yielding

$$\theta_{ij} = \theta_{i,\text{Out}} + \theta_{j,\text{In}} + \theta_{\text{Edge}} \quad (3.3)$$

for $i \neq j$ and

$$\theta_{12-21} = \theta_{23-32} = \theta_{31-13} = \theta_{\text{Reci}}.$$

We refer to [Robins and Pattison \(2005\)](#) and [Frank and Strauss \(1986\)](#) for a more technical and general definition of the “*homogeneity condition*” based on isomorphic graphs. Under these constraints, (3.2) simplifies to:

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) &\propto \exp \left\{ \theta_{\text{Edge}} \sum_{i \neq j} y_{ij} + \sum_{i=1}^N \theta_{i,\text{Out}} \left(\sum_{j=1}^N y_{ij} \right) + \sum_{i=1}^N \theta_{i,\text{In}} \left(\sum_{j=1}^N y_{ji} \right) + \theta_{\text{Reci}} \sum_{i \neq j, i < j} y_{ij} y_{ji} \right\} \\ &= \exp \left\{ \theta_{\text{Edge}} s_{\text{Edges}}(\mathbf{y}) + \sum_{i=1}^N \theta_{i,\text{Out}} s_{i,\text{Out}}(\mathbf{y}) + \sum_{i=1}^N \theta_{i,\text{In}} s_{i,\text{In}}(\mathbf{y}) + \theta_{\text{Reci}} s_{\text{Reci}}(\mathbf{y}) \right\}. \end{aligned} \quad (3.4)$$

Model (3.4) was first derived by [Holland and Leinhardt \(1981\)](#) as the p_1 model and its sufficient statistics are illustrated in Figure 3.2. [Frank and Strauss \(1986\)](#) carry out a similar exercise for the Markov dependence for undirected graphs (shown in Figure 3.1 (c) for directed networks) to derive that the sufficient statistics are counts of k -star configurations for $k = 1, \dots, N - 1$, which is a subgraph where one actor is connected to $k - 1$ other actors, and triangles. The latter statistic is stated in (2.5). In general, note that also the sufficient functions in (3.4) were already defined as global descriptive statistics in (2.2) and (2.4).

Ensuing from this observation, we can define the ERGM the other way around and hence determine the sufficient statistics implying a specific dependence and homogeneity assumption. Generally, this view on ERGMs is more widespread, and [Robins et al. \(1999\)](#) introduced the model in the following convenient representation:

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y})\}}{\kappa(\boldsymbol{\theta})} \quad (3.5)$$

where

- $\boldsymbol{\theta} \in \mathbb{R}^p$ is a p -dimensional vector of natural parameters;
- $\mathbf{s}: \mathcal{Y} \rightarrow \mathbb{R}^p$ is a function calculating the vector of p natural sufficient statistics for any network in \mathcal{Y} . As detailed in the previous paragraph, it determines the form of dependence between the pairs in the network. Note that in addition to endogenous statistics capturing specific structural patterns, we can incorporate exogenous in the sufficient statistics as detailed underneath;
- $\kappa(\boldsymbol{\theta}) = \sum_{\tilde{\mathbf{y}} \in \mathcal{Y}} \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\tilde{\mathbf{y}})\}$ is a normalizing constant, which is sometimes also called the partition function, to ensure that (3.5) sums up to one over all $\mathbf{y} \in \mathcal{Y}$.

Note that (3.5) is a canonical exponential family model with known properties that guarantee meaningful inference and guide estimation procedures (see [Barndorff-Nielsen, 1978](#) for a general introduction to exponential families and Chapters 9 and 8 for details on this particular topic). Building on the expression that “*all roads lead to Rome*”, we point out that one can also end up with the specific form of (3.5) from the perspective of mechanical physics. In particular, one can comprehend (3.5) as the solution to an optimization problem in the space of all possible probability distributions, where we want to maximize the Gibbs entropy subject to the constraint that the expectation of the sufficient statistics needs to be equal to their observed value ([Park and Newman, 2004](#)).

3.1 Exponential Random Graph Model

We can also incorporate exogenous statistics in the sufficient statistics $\mathbf{s}(\mathbf{y})$ on the level of the pairs or nodes by including statistics in the following form

$$s_{\text{Pair}}(\mathbf{y}) = \sum_{i \neq j} y_{ij} x_{ij}. \quad (3.6)$$

For categorical and continuous nodal attributes one can define x_{ij} in (3.6) for $i \neq j$ through

$$s_{\text{Nodematch}}(\mathbf{y}) = \sum_{i \neq j} y_{ij} \mathbb{I}(x_i = x_j)$$

$$s_{\text{Nodeabs}}(\mathbf{y}) = \sum_{i \neq j} y_{ij} |x_i - x_j|,$$

to, e.g., test whether there is the tendency for homophily or heterophily in the network (McPherson et al., 2001). The theoretical underpinning with corresponding dependence graphs to incorporate exogenous variables is provided by Robins et al. (2001) based on so-called chain graphs, that at the same time include undirected edges between pairs and directed edges from exogenous variables and the pairs.

Interpretation Interpreting the coefficients $\boldsymbol{\theta}$ is possible both at the global network and single tie level. We here illustrate the interpretation of θ_q for $q \in \{1, \dots, p\}$ corresponding to the sufficient statistic $s_q(\mathbf{y})$. For the global perspective, $\theta_q > 0$ implies that networks with increasing values of $s_q(\mathbf{y})$ are also increasingly more likely according to (3.5) if all other statistics remain constant. The contrary holds under $\theta_q < 0$. To obtain an interpretation on the tie level reminiscent of logistic regression, we need additional notation. By \mathbf{y}_{ij}^x we denote the network \mathbf{y} where the (i, j) th entry is fixed at value x for $x \in \{0, 1\}$ and let $\mathbf{y}_{-(ij)}$ be \mathbf{y} excluding the relation y_{ij} . Next we define change statistics as a function of $\mathbf{y}_{-(ij)}$ returning the change in the sufficient statistics caused by switching the entry y_{ij} from 0 to 1:

$$\Delta_{ij}(\mathbf{y}_{-(ij)}) = \mathbf{s}(\mathbf{y}_{ij}^1) - \mathbf{s}(\mathbf{y}_{ij}^0). \quad (3.7)$$

One can then specify the probability to observe a tie between actors i and j conditional on $\mathbf{y}_{-(ij)}$:

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij} = 1 | Y_{-(ij)} = \mathbf{y}_{-(ij)}) = \frac{\exp\{\boldsymbol{\theta}^\top \Delta_{ij}(\mathbf{y}_{-(ij)})\}}{1 + \exp\{\boldsymbol{\theta}^\top \Delta_{ij}(\mathbf{y}_{-(ij)})\}}. \quad (3.8)$$

Thus we get the following interpretation on the level of the tie: if switching the value of y_{ij} from 0 to 1 raises only the q th entry of $\Delta_{ij}(\mathbf{y}_{-(ij)})$ by one, the conditional log-odds of Y_{ij} are changed by the additive factor θ_q (Goodreau et al., 2009).

Estimation and Inference Given a set of sufficient statistics and realized network \mathbf{y} , we want to find an estimator $\hat{\boldsymbol{\theta}}$ maximizing (3.5), called the maximum-likelihood estimator. Although the likelihood has a simple mathematical form, maximizing it is notoriously difficult. The primary obstacle in this endeavor is the evaluation of $\kappa(\boldsymbol{\theta})$, which in the case of directed networks necessitates calculating the sum of $|\mathcal{Y}| = 2^{n(n-1)}$ terms which is only feasible for small networks (Vega Yon et al., 2021). Not provided the current computational possibilities, Strauss and Ikeda (1990) base the estimation of $\boldsymbol{\theta}$ on the pseudolikelihood, defined through (3.8) and equivalent to common logistic regression. Since the properties of this approach are, however, unknown in the general case of (3.5) (van Duijn et al., 2009), different types of Markov Chain Monte Carlo techniques were proposed to help in this setting:

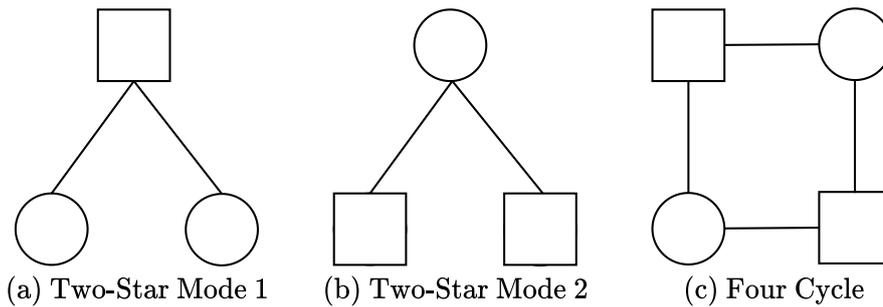


Figure 3.3.: Visualization of sufficient statistics for bipartite networks. Circles represent actors of mode 1, while rectangles indicate actors of mode 2.

1. *Stochastic Approximation* via a Robbins-Monroe algorithm was proposed by [Snijders \(2002\)](#), which can be comprehended as a stochastic version of the known *Newton Raphson Algorithm*.
2. [Handcock \(2003\)](#) first adapted the methods of [Geyer \(1992\)](#) to the ERGM, where we approximate the logarithmic likelihood ratio of θ and a fixed θ_0 via a Monte Carli quadrature.
3. Motivated by the importance of picking an adequate value for θ_0 to approximate the likelihood ratio, [Hummel et al. \(2012\)](#) suggest a *Partial Stepping* to successively find θ_0 .

All these approaches heavily rely on sampling networks from (3.5) under fixed parameter θ , which is possible by employing a Metropolis-Hastings sampler that generates a Markov chain of graphs that has (3.5) as its stationary distribution. See [Hunter et al., 2013](#) and [Byshkin et al., 2016](#) for more detail on this topic. We refer to the references provided above and Chapters 9 and 8 for additional information on the optimization procedures. In particular, Chapter 8 extends the proposed methods of [Handcock \(2003\)](#) and [Hummel et al. \(2012\)](#) to dynamic signed networks.

Degeneracy Issues After or while fitting model (3.5) to realized data, an often occurring phenomenon first described in [Snijders \(2002\)](#) and [Handcock \(2003\)](#) but formalized by [Schweinberger \(2011\)](#) is obtaining a degenerate distribution over networks. If this happens, the probability distribution (3.5) puts most probability mass under a particular set of parameters either on the empty or full graph, which is unreasonable for most applications. One way of identifying this behavior is through the goodness-of-fit procedure proposed by [Hunter et al. \(2008\)](#), where observed network statistics are compared to statistics of networks simulated under the estimated model. Degeneracy appears to be particularly prevalent for Markov graphs; thus [Snijders et al. \(2006\)](#) and [Hunter and Handcock \(2006\)](#) propose novel weighted statistics that, in many cases, have better empirical behavior. In Chapter 6, we explain how these statistics are constructed and can be interpreted, while we adapt them in Chapter 8 to signed networks representing the predictions of structural balance theory ([Heider, 1946](#)).

Bipartite Networks The entire derivation of the ERGM given above holds similarly for bipartite networks. However, note that bipartite networks are constrained in that ties are only possible between actors of different modes of actors. Resulting from this, all pairs (i, j) with $i, j \in \mathcal{V}_1$ or $i, j \in \mathcal{V}_2$ are nonstochastic and hence excluded from the actor set of the dependence graph. This characteristic leads to slightly changed forms of dependence ([Wang et al., 2013](#)) and sufficient

3.1 Exponential Random Graph Model

statistics (Faust and Skvoretz, 1999; Wang et al., 2009), exemplary statistics are depicted in Figure 3.3. Moreover, we introduce a novel suite of network terms in Chapter 9 for pairwise covariates between actors of \mathcal{V}_1 or \mathcal{V}_2 .

Temporal Networks Extending the framework to encompass networks observed over time, i.e., $\mathbf{y}_1, \dots, \mathbf{y}_T$, can be achieved in two ways.

First, we can comprehend the evolution of these observed networks as the outcome of a Markov chain, by specifying a joint model for

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_1, \dots, \mathbf{Y}_T | \mathbf{Y}_0 = \mathbf{y}_0) = \prod_{t=1}^T \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \dots, \mathbf{Y}_0 = \mathbf{y}_0), \quad (3.9)$$

conditional on the initial network \mathbf{y}_0 . Under a temporal Markov dependence structure of first order⁴, the separate terms of the product of (3.9) simplify to

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \dots, \mathbf{Y}_1 = \mathbf{y}_1) = \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}). \quad (3.10)$$

In the last step, we plug (3.5) into (3.10), although now the sufficient statistics can additionally depend on the lagged network \mathbf{y}_{t-1} next to \mathbf{y}_t :

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})\}}{\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1})}. \quad (3.11)$$

Sufficient statistics encompassed in $\mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ can capture within-network or simultaneous dependencies through statistics that only depend on \mathbf{y}_t and between-network dependencies by also incorporating \mathbf{y}_{t-1} . While statistics for within-network dependencies were already illustrated in Figure 3.2, Figure 3.4 provides three examples of statistics that relate to between-network dependencies. This general model class is called the Temporal Exponential Random Graph Model (TERGM, Hanneke et al., 2010) building on Robins and Pattison (2001). Afterwards the model was substantially improved via the Separable Temporal Exponential Random Graph Model (STERGM, Krivitsky and Handcock, 2014) by specifying a separable version of the model differentiating between the formation and dissolution of edges. See Chapter 6 for a survey covering the model and Chapter 10 for its adaption to event data.

Second, we can perceive the networks as evolving over time, guided by a Markov process. Snijders and Koskinen (2013) suggest using the conditional distribution (3.10) in conjunction with random opportunities for a change in the state of pairs. The resulting process has an ERGM as its stationary distribution and can be considered a tie-oriented variant of the actor-oriented model (SAOM, Snijders, 1996). In Koskinen and Lomi (2013) and Koskinen et al. (2015), this approach is further broadened and applied to networks representing foreign direct investments between countries.

⁴Higher order Markov dependence assumptions are possible in the same manner.

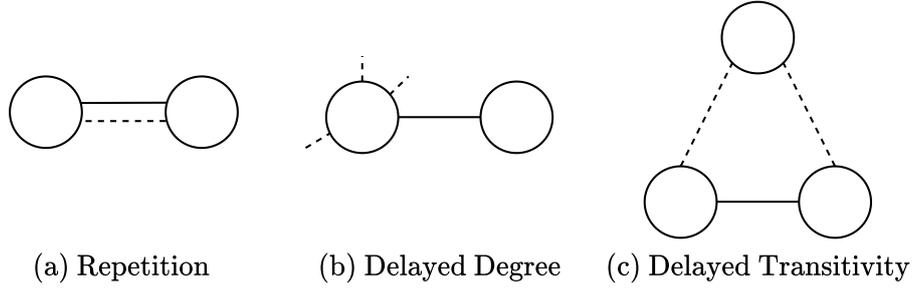


Figure 3.4.: Visualization of sufficient statistics for temporal networks. Contemporaneous edges are drawn as solid lines, while edges observed in the past are dashed.

3.2. Relational Event Model

Being surrounded by sensor technologies every day in all shapes and forms, such as video cameras, logs of phone calls, Twitter notifications, Facebook friendships, and sociometric badges (Lazer et al., 2009; Eagle and Pentland, 2006), we can measure and map social movements at a much more fine-grained resolution than at particular time points $1, \dots, T$. While one could still aggregate the provided event traces over some time intervals and study them with the means detailed in Section 3.1, this approach necessarily comes with a loss of information (Stadtfeld, 2018). Further, it is not trivial to assess the extent to which the aggregation of a specific number of events constitutes a durable tie such as a friendship (Kitts, 2014). Take, for instance, the communication network in response to the World Trade Center attack in 2001. For this data, much information lies in the sequence of the calls, i.e., participation shifts and conversational norms discussed in Gibson (2003), not necessarily in the aggregated networks⁵. One way to exploit the complete available temporal information is the Relational Event Model (REM, Butts, 2008). In the following paragraphs, we first formulate the model from the perspective of counting processes. Second, we detail how events from the past can affect the intensity of this counting process in the present through sufficient statistics. Then, the interpretation, estimation, and inference are detailed. Finally, extensions of the REM are sketched.

Model Formulation Taking the matrix-valued counting process $\mathbf{N}(\tilde{t})$ defined in (2.6) as a starting point, we observe that each cell is nondecreasing in time and is thus a submartingale. For an introduction to stochastic processes we refer to Daley and Vere-Jones (2008) and Aalen et al. (2008, Chap.2). For technical reasons, we assume that $N_{ij}(0) = 0$ holds and that $N_{ij}(\tilde{t})$ is adapted to a history $\{\mathcal{H}(\tilde{t})\}$ for all $(i, j) \in \mathcal{U}$. We can then apply the Doob-Meyer decomposition:

$$\mathbf{N}(\tilde{t}) = \mathbf{\Lambda}(\tilde{t}) + \mathbf{M}(\tilde{t}), \quad (3.12)$$

where $\mathbf{\Lambda}(\tilde{t}) = \int_0^{\tilde{t}} \boldsymbol{\lambda}(u) du$ is defined as the matrix-valued intensity process and $\mathbf{M}(\tilde{t})$ is a matrix-valued local martingale, which is a process with expectation $\mathbf{0}$ and uncorrelated increments (Aalen et al., 2008). One can think of the martingale process as random noise and of the intensity process as the deterministic part of the observed counting process which we can learn. Put differently,

⁵For analyses of this data, see Butts, 2008 and Renshaw et al., 2022.

3.2 Relational Event Model

(3.12) serves as the analog to the decomposition of observed values in linear models into the predicted value and an unexplained residual. Formally, the (i, j) th cell of $\boldsymbol{\lambda}(\tilde{t})$ is defined as

$$\lambda_{ij}(\tilde{t}) = \lim_{h \downarrow 0} \frac{\mathbb{P}(N_{ij}(\tilde{t} + h) = N_{ij}(\tilde{t}) + 1)}{h},$$

which is the instantaneous probability of an event between actors i and j at time point \tilde{t} . In other words, it suffices to specify a model for the intensity process to fully characterize the counting process $\mathbf{N}(\tilde{t})$. Note that this setting is equivalent to time-to-event analysis on which the textbooks by [Kalbfleisch and Prentice \(2002\)](#) and [Lawless \(2003\)](#) give an overview.

We assume that $\mathbf{N}(\tilde{t})$ is a matrix-valued inhomogenous Poisson process, whereby the increments of each cell are Poisson distributed random variables between times h and t with $t > h$:

$$N_{ij}(t) - N_{ij}(h) \sim \text{Pois} \left(\int_h^t \lambda_{ij}(u) \, du \right). \quad (3.13)$$

We here opt for a local characterization of Poisson processes encompassing intensities that are stochastic and explicitly depend on previous events, covering doubly stochastic Cox processes such as self-exciting Hawkes processes ([Hawkes, 1971](#)).

Next, we parametrize the intensity of (2.6) conditional on the history of events and additional exogenous covariate processes. On the tie-level, the conditional intensity to observe an event between actors i and j at \tilde{t} is specified through:

$$\lambda_{ij}(\tilde{t} | \mathcal{H}(\tilde{t}), \boldsymbol{\theta}) = \begin{cases} \lambda_0(\tilde{t}) \exp\{\boldsymbol{\theta}^\top \mathbf{s}_{ij}(\mathcal{H}(\tilde{t}))\}, & \text{if } (i, j) \in \mathcal{U}(\tilde{t}) \\ 0, & \text{else} \end{cases} \quad (3.14)$$

where

- $\lambda_0(\tilde{t}) \in \mathbb{R}^+$ defines the baseline intensity, which one can either specify parametrically ([Butts, 2008](#)), semiparametrically ([Fritz et al., 2022a](#)), or nonparametrically ([Vu et al., 2011a,b](#));
- $\boldsymbol{\theta} \in \mathbb{R}^p$ is a p -dimensional vector of parameters to be estimated;
- $\mathbf{s}_{ij}(\mathcal{H}(\tilde{t})) \in \mathbb{R}^p$ are sufficient statistics calculated for the relation between actors i and j that can be defined similar to the sufficient statistics in (3.5) and will be discussed in more depth in the following paragraph.

We only constrain ourselves to intensities that are piecewise constant between the event times, $\tilde{t}_1, \dots, \tilde{t}_M$, although exogenous known shocks as introduced in [Butts and Marcum \(2017\)](#) are easily incorporated. Resulting from this restriction, we can comprehend the implied counting process, in most cases, as an homogeneous Poisson process we restart after each observed event with accordingly updated intensities ([Butts, 2008](#)).

Sufficient Statistics Contrasting the definition of sufficient statistics in (3.5), the statistics $\mathbf{s}_{ij}(\mathcal{H}(\tilde{t}))$ from (3.14) are defined on the level of each pair. This characteristic highlights the differing levels of the two models introduced in Section 3.1 and 3.2: while the ERGM models networks from a global perspective, the REM operates on the local tie-level. We refer to Chapter 6 for a more thorough discussion of the differences between the corresponding two models. That being

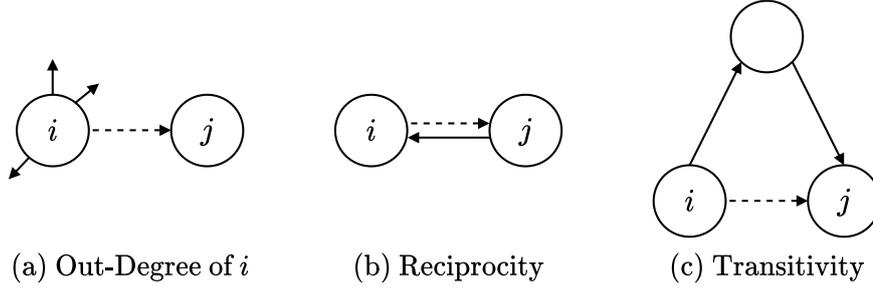


Figure 3.5.: Visualization of sufficient statistics on the tie level for REMs. Solid lines represent events observed in the past, while the dotted lines indicate the modeled event. Arrowheads indicate the direction of the event.

said, most global statistics of the ERGM do have an interpretation from the local perspective as well (see (3.10) and Pattison and Robins, 2002).

Pairwise equivalents to the statistics visualized in Figure 3.2 (a) and (b) can be included in $\mathbf{s}_{ij}(\mathcal{H}(\tilde{t}))$ as follows:

$$s_{ij, \text{Out-Sender}}(\mathcal{H}(\tilde{t})) = \sum_{k=1}^N \mathbb{I}(N_{ik}(\tilde{t}^-) > 0) \quad (3.15)$$

$$s_{ij, \text{Recei}}(\mathcal{H}(\tilde{t})) = \mathbb{I}(N_{ji}(\tilde{t}^-) > 0), \quad (3.16)$$

where \tilde{t}^- is the time point immediately before \tilde{t} and the corresponding illustrations are supplied in Figure 3.5 (a) and (b). Similar to (3.15), one can likewise define a statistic for the out-degree of the receiver or the in-degree statistic. Beyond the dyadic level, we may incorporate the number of transitive partners from Figure 3.5 (c) via the statistic

$$s_{ij, \text{Transitivity}}(\mathcal{H}(\tilde{t})) = \sum_{k=1}^N \mathbb{I}(N_{ik}(\tilde{t}^-) > 0) \mathbb{I}(N_{kj}(\tilde{t}^-) > 0)$$

to cover tendencies of transitive clustering of events. Other types of triangular statistics are described in Chapter 10 and legitimized in the application to the international combat aircraft trade. If exogenous covariates on each pair are available, one may integrate such information in (3.14) directly through

$$s_{ij, \text{Pair}}(\mathcal{H}(\tilde{t})) = x_{ij}$$

for continuous data, while dummy-coded covariates can be generated for categorical data. We further comment that categorical or continuous actor covariates can be transformed to pairwise information by setting x_{ij} equal to $\mathbb{I}(x_i = x_j)$ for checking whether actors i and j match on a categorical attribute or $|x_i - x_j|$ in the continuous case. While these two examples reflect processes of homophily and heterophily, other types of transformations can be used as well. In Chapter 10, we suggest statistics to mirror the separable parametrization of the STERGM mentioned in Section 3.1. These statistics allow the differentiation between the onset and repetition of events.

In analogy to the parallel between the dependence graph and set of sufficient statistics detailed in Section 3.1, the dependency structure between the individual pairwise counting processes encompassed in $\mathbf{N}(\tilde{t})$ is also determined by the statistics $\mathbf{s}_{ij}(\mathcal{H}(\tilde{t}))$. We assume that, after conditioning

3.2 Relational Event Model

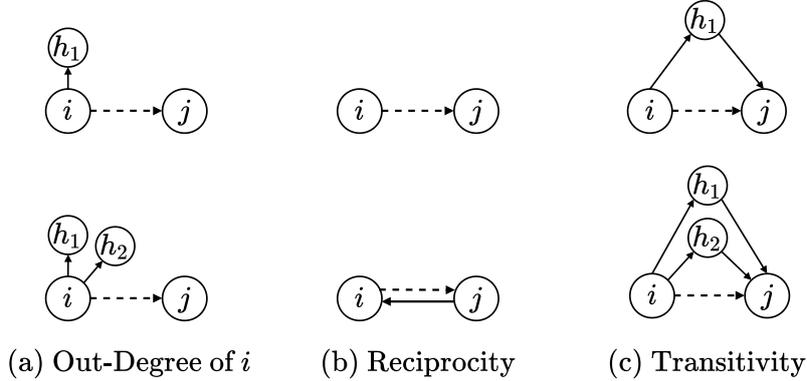


Figure 3.6.: Graphical illustration of endogenous statistics. Solid lines represent past interactions, while dotted lines are possible but unrealized events. The relative risk of the events in the second row compared to the events in the first row is $\exp\{\theta_q\}$ if all other covariates are fixed, where θ_q is the coefficient of the respective statistic of each row.

on the history of events, all events are independent of one another. This does, however, not imply the marginal independence of events, as an event observed at a particular point in time can affect future events. Put differently, “*past history creates the context for present action*” (Butts, 2008, p. 160). For a mathematically rigorous definition of possible dependence assumptions we refer to Kreiss (2021).

Interpretation One can interpret the coefficient θ_q with $q \in \{1, \dots, p\}$ as the multiplicative change in the intensity (3.14) if the corresponding statistic $s_{ij,q}(\mathcal{H}(\tilde{t}))$ is raised by one unit and all other statistics are left unchanged. In the jargon of proportional hazards commonly employed in time-to-event analysis, the covariates are assumed to have a proportional effect on the intensity function. To give an example, a positive effect of the sender’s out-degree, shown in Figure 3.5 (a), may be viewed as the tendency to interact with actors that were already active a lot in the past. A graphical illustration of this interpretation for all statistics of Figure 3.5 is provided in Figure 3.6. For instance, the intensities of the events shown in the first row multiplied by $\exp\{\theta_q\}$, here defined as the corresponding coefficient for each term, gives the intensity of the respective event in the second row. Relaxations of this proportionality of the effect are possible by letting θ vary over time as described later in this section and Chapters 10 and 11.

Estimation and Inference Based on a sufficient set of statistics, we construct the likelihood of θ with observed stream of events \mathcal{A} from the increments between all time points that according to (3.13) are realizations of Poisson-distributed random variables. Under piecewise constant intensity functions, (3.13) simplifies to:

$$Y_{ij,m} = N_{ij}(\tilde{t}_m) - N_{ij}(\tilde{t}_{m-1}) \sim \text{Pois}((\tilde{t}_m - \tilde{t}_{m-1})\lambda_{ij}(\tilde{t}_m|\mathcal{H}(\tilde{t}_m), \theta)) \quad (3.17)$$

for all t_m with $m = 1, \dots, M$ and $(i, j) \in \mathcal{U}(t_m)$. Note that in the case of a continuously time-varying baseline intensity as employed in Fritz et al. (2022a), (3.17) still serves as a simple rectangular

approximation of the integral in (3.13). The likelihood that we maximize with respect to θ can then be formulated as

$$\mathcal{L}(\theta; \mathcal{M}) = \prod_{m=1}^M \prod_{(i,j) \in \mathcal{U}(\tilde{t}_m)} ((\tilde{t}_m - \tilde{t}_{m-1}) \lambda_{ij}(\tilde{t}_m | \mathcal{H}(\tilde{t}_m), \theta))^{y_{ij,m}} \exp\{-(\tilde{t}_m - \tilde{t}_{m-1}) \lambda_{ij}(\tilde{t}_m | \mathcal{H}(\tilde{t}_m), \theta)\}. \quad (3.18)$$

Note that (3.18) falls into the more general family of generalized linear models (GLMs) proposed by Nelder and Wedderburn (1972), where the target variable $Y_{ij,m}$ follows a Poisson distribution with the offset term $\log(\tilde{t}_m - \tilde{t}_{m-1})$. Due to this observation, we do not have to write ad hoc software implementations but can rely on stable implementations for GLMs, such as the software package `stats` (R Core Team, 2022) to find $\hat{\theta}$. Consistency and asymptotic normality of $\hat{\theta}$ can be derived as a special case of known results for GLMs (Fahrmeir and Kaufmann, 1985). These asymptotics allow the construction of tests and quantification of the uncertainty associated with the obtained estimates.

A second remark concerns the case where $\lambda_0(\tilde{t})$ from (3.14) is left unspecified. In fact, this setting is equivalent to assuming $\lambda_0(\tilde{t}) = \lambda_{0,m}$ holds for $\tilde{t} \in (t_{m-1}, t_m]$ and $m = 1, \dots, M$, i.e., a separate constant intercept is estimated for each observed event. Consequentially, we use the reverse argument of Whitehead (1980) to prove that (3.18) is equivalent to the partial likelihood developed by Cox (1972, 1975):

$$\mathcal{L}(\theta; \mathcal{A}) = \prod_{m=1}^M \frac{\lambda_{i_m j_m}(t_m | \mathcal{H}(t_m), \theta)}{\sum_{(i,j) \in \mathcal{U}(t_m)} \lambda_{ij}(t_m | \mathcal{H}(t_m), \theta)}. \quad (3.19)$$

Maximizing (3.19) in respect to θ is still possible via a Newton-Raphson algorithm (for the technical details see Perry and Wolfe, 2013). One may alternatively derive this likelihood by assuming that no information on the exact time points of events but solely their order is provided in the data as detailed in Butts (2008). In this context, inference about $\hat{\theta}$ can be based on the theory derived in Kreiß et al. (2019) if $|\mathcal{V}|$ increases and on the results of Perry and Wolfe (2013) if the time span of observed events goes to infinity.

Incorporating nonlinear and random effects The connection to GLMs detailed in the prior paragraph also enables us to utilize the literature extending GLMs to Generalized Additive Models (GAM, Hastie and Tibshirani, 1987), varying-coefficient models (VCM, Hastie and Tibshirani, 1993), and random effects.

GAMs allow the inclusion of nonlinear effects in (3.14) in the following manner:

$$\lambda_{ij}(\tilde{t} | \mathcal{H}(\tilde{t}), \theta) = \begin{cases} \lambda_0(\tilde{t}) \exp\{\sum_{q=1}^p f_q(s_{ij,q}(\mathcal{H}(\tilde{t})))\}, & \text{if } (i, j) \in \mathcal{U}(\tilde{t}) \\ 0, & \text{else} \end{cases}$$

where $f_q(s_{ij,q}(\mathcal{H}(\tilde{t})))$ is a smooth function parametrized through basis functions. Assuming that we use K_q basis functions for $s_{ij,q}(\mathcal{H}(\tilde{t}))$, we write:

$$f_q(s_{ij,q}(\mathcal{H}(\tilde{t}))) = \sum_{k=1}^{K_q} \theta_{q,k} B_k(s_{ij,q}(\mathcal{H}(\tilde{t}))) = \theta_q^\top \mathbf{B}(s_{ij,q}(\mathcal{H}(\tilde{t}))), \quad (3.20)$$

3.2 Relational Event Model

where the k th basis function is $B_k(\cdot)$ weighted by $\theta_{q,k}$. To ensure that the function is adequately smooth, we incorporate a multiplicative penalization term in (3.18). As for the type of basis function and penalization, there are numerous options to choose from. In Chapter 7 we rely on thin-plate splines (Wood, 2003) penalizing the curvature of the squared function, while P-splines as proposed by Eilers and Marx (1996) are employed in Chapter 11. By including interaction variables between smooth and linear effects we obtain time-varying effects via VCMs by setting:

$$\boldsymbol{\theta}_q(\tilde{t}) = \boldsymbol{\theta}_q^\top \mathbf{B}(\tilde{t}).$$

In this context, the function $\boldsymbol{\theta}_q(\tilde{t})$ can be specified in the same manner as the smooth functions in (3.20). In Chapter 10, we study an event network over a period of more than 50 years, thus time-varying effects are urgently needed. Eventually, random effects can be used to control for latent heterogeneity on the actor level, e.g., by including sender and receiver-specific effects in (3.14):

$$\lambda_{ij}(\tilde{t}|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}) = \begin{cases} \lambda_0(\tilde{t}) \exp\{\boldsymbol{\theta}^\top \mathbf{s}_{ij}(\mathcal{H}(\tilde{t})) + u_i^S + u_j^R\}, & \text{if } (i, j) \in \mathcal{U}(\tilde{t}) \\ 0, & \text{else} \end{cases}$$

with $u_i^S \sim N(0, \tau_S^2)$ and $u_i^R \sim N(0, \tau_R^2)$ for $i = 1, \dots, N$ in the case with directed events. Following the arguments presented in Kimeldorf and Wahba (1970) as well as in Chapters 7 and 11, estimation routines proposed for Generalized Additive Models suffice for estimating models with these extensions. We refer to Wood (2017) for a thorough treatment of these estimation techniques as implemented in the versatile R package `mgcv`.

Extensions In some cases, the observed events suffer from measurement errors. One example of this phenomenon is the non-negligible false-discovery rate in event identification when automated or human-coded events are studied. In Chapter 11, we offer a method to control for the resulting spurious events by defining two separate counting processes, one counting the real events and one spurious event, each characterized by a different intensity function. Another particular type of measurement error results from clustered observations, where the events are not observable in continuous time, but we can only identify a time interval in which each event falls. We treat such situations in Chapter 10 under the assumption that the exact order of events observed in the same temporal interval carries no information.

Stadtfeld (2012) proposed an actor-oriented variant of the REM, which we introduced in a tie-oriented version in (3.14). For this approach we assume that the intensity of directed events can be split into a sender-specific intensity and a probability of selecting the receiver conditional on the sender:

$$\lambda_{ij}(\tilde{t}|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}) = \begin{cases} \lambda_i(\tilde{t}|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}_1) p_i(j|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}_2), & \text{if } (i, j) \in \mathcal{U}(\tilde{t}) \\ 0, & \text{else} \end{cases}$$

where

$$\lambda_i(\tilde{t}|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}_1) \exp\{\boldsymbol{\theta}_1^\top \mathbf{s}_i(\mathcal{H}(\tilde{t}))\} \tag{3.21}$$

is the intensity for actor i to be the sender of any event and $p_i(j|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}_2)$ describes the conditional probability of choosing actor j as the receiver if actor i is the sender. The actor-specific intensity

can be related to a counting process of how often actors in the network are senders of events and is hence sometimes called an egocentric REM (see [Vu et al., 2011b](#) for an example of how this is sub-model can be of interest by itself). For the selection model, a discrete choice model as introduced by [McFadden \(1973\)](#) is employed:

$$p_i(j|\mathcal{H}(\tilde{t}), \boldsymbol{\theta}_2) = \frac{\exp\{\boldsymbol{\theta}_2^\top \mathbf{s}_{ij}(\mathcal{H}(\tilde{t}))\}}{\sum_{k=1, k \neq i}^N \exp\{\boldsymbol{\theta}_2^\top \mathbf{s}_{ik}(\mathcal{H}(\tilde{t}))\}}. \quad (3.22)$$

While all sufficient statistics in (3.21) and (3.22) can be employed in (3.14), the other way around does not hold. For $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ to be identifiable, the corresponding statistics have to vary with the sender in (3.21) and with the receiver in (3.22). An example for the former type of statistic is the out-degree of the sender and for the latter the reciprocity statistic both illustrated in [Figure 3.5](#) (a) and (b).

[Perry and Wolfe \(2013\)](#) amend the model beyond pairwise interactions to cover multicast events, where the receiver of an event can be set-valued, i.e., events like $a = (i, \{j_1, \dots, j_K\}, \tilde{t})$ are possible. One setting in which such data naturally arise is when studying email traffic. Going in a similar direction, [Lerner et al. \(2021\)](#) and [Lerner and Lomi \(2022\)](#) study so-called hyper events, which are events not between two but multiple actors. The application covered in [Lerner et al. \(2021\)](#) consists of approximately 2000 meeting events of Margaret Thatcher with her cabinet ministers.

4. Network Attribute Models

For many investigations, the principal goal lies in quantifying how an attribute is affected by the actor’s embedding in the network instead of how the random relations are generated. If dynamics are observed, we can distinguish between two settings for this type of study. First, we might assume that the attributes and the network evolve simultaneously. Second, it is reasonable to formulate a model where the network is an obvious antecedent to a change in the attribute in particular applications with no simultaneous but temporal dependency.

A canonical case from sociology for the former setting is how smoking or drug consumption is caused by one’s friend’s behavior (Snijders et al., 2018). The central question is whether the behavior is explained by influence (“*friends of smokers start to smoke*”) or selection (“*smokers befriend predominantly other smokers*”). While we can tackle hypotheses related to selection processes with adequately defined exogenous covariates affecting the generation of the network in the form of (3.6) (Robins et al., 2001), we require novel methods to investigate the influence of social relations on attributes. Regarding the second data setting, one might consider how mobility patterns relate to the number of people infected with COVID-19. For this example, assume that a weighted network is provided in which the actors are spatial units, and the edges indicate the mobility flows between the corresponding spatial units. In this example, the infection counts in a particular temporal interval and spatial unit are the actor attribute. Based on the finding that each infection becomes symptomatic only after an incubation period, which at the beginning of the pandemic was estimated to be around five days (Li et al., 2020), one can argue that there should only be a temporally lagged dependency between the mobility network and the infection counts. Since this problem was at the time of writing this dissertation solely tackled for data observed in discrete time representable with the tools of Section 2.1, our treatment of the topic focuses on the corresponding data setting.

Simultaneous Dependence Robins et al. (2001) suggest Autologistic Actor Attribute Models as a way to study social influences processes, where the roles of the network and attributes are swapped in that the outcome of the attributes is stochastic and the network is fixed. In line with the notation of Section 2.1, we change perspective from modeling the network \mathbf{Y} to modeling the attribute $\mathbf{X} = (X_1, \dots, X_N)$, being the random variable corresponding to a binary actor attribute \mathbf{x} , e.g., smoker vs. non-smoker. We obtain a probability distribution for \mathbf{X} conditional on \mathbf{y} similar to (3.5):

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{x}, \mathbf{y})\}}{\kappa(\boldsymbol{\theta}, \mathbf{y})}, \quad (4.1)$$

where $\kappa(\boldsymbol{\theta}, \mathbf{y}) = \sum_{\tilde{\mathbf{x}} \in \mathcal{X}} \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\tilde{\mathbf{x}}, \mathbf{y})\}$ is the normalizing constant with \mathcal{X} being the set of all observable behaviors in the network. Relying on the theory of chain graphs, (4.1) follows again from the Hammersley-Clifford theorem and the sufficient statistics can again be derived through

particular forms of dependence and homogeneity assumptions (Robins et al., 2001; Daraganova and Robins, 2013).

If the actor attributes are observed over time, we denote their evolution by $\mathbf{x}_1, \dots, \mathbf{x}_T$. In the same manner that (3.5) was expanded to cover temporal networks in (3.11), it is straightforward to model time-varying behavior with (4.1) through a first-order Markov assumption:

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X}_t = \mathbf{x}_t | \mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{y})\}}{\kappa(\boldsymbol{\theta}, \mathbf{x}_{t-1}, \mathbf{y})} \quad (4.2)$$

providing a Temporal Autologistic Actor Attribute Model akin to the TERGM proposed by Hanneke et al. (2010) and Robins and Pattison (2001). To model count-valued attributes, we extend (4.2) following Krivitsky (2012):

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X}_t = \mathbf{x}_t | \mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{Y} = \mathbf{y}) = \frac{h(\mathbf{x}_t) \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{y})\}}{\kappa(\boldsymbol{\theta}, \mathbf{x}_{t-1}, \mathbf{y})}, \quad (4.3)$$

where we additionally have to specify a reference function $h: \mathcal{X} \rightarrow [0, \infty)$ that characterizes the distribution of \mathbf{X}_t under $\boldsymbol{\theta} = \mathbf{0}$. We refer to Krivitsky (2012) for details on different ways to choose the reference function, but for this section it suffices to work with the Poisson reference measure $h(\mathbf{x}_t) = \prod_{i=1}^N \frac{1}{x_i!}$.

On a side note, extensions of the ERGM to multivariate (Krivitsky et al., 2020), continuous (Desmarais and Cranmer, 2012), and ordered (Krivitsky and Butts, 2017) networks are directly transferable to modeling multivariate, continuous, and ordered attributes. Moreover, the estimation routines sketched in Section 3.1 are applicable for this model class if the employed sampled is adapted to (4.1), (4.2), or (4.3) depending on the employed model. Contrasting this approach, Snijders et al. (2018) introduced a joint model for the co-evolution of networks and behavior based on the SAOM, which was later enhanced to cover continuous behavior (Niezink and Snijders, 2017). These two approaches, however, not only differ in being actor- vs. tie-oriented models but also in the fact that in (4.1) and (3.5), either processes of influence or selection are studied conditionally on the other, whereas Snijders et al. (2018) posits a joint model for the network and attribute.

Temporal Dependence Under the assumption that the dependence between attributes at time point t is captured by past observations of the attributes themselves, the network, and other exogenous factors, the sufficient statistics can be stated as $\mathbf{s}(\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{y}) = \sum_{i=1}^N x_{i,t} \mathbf{s}_i(\mathbf{x}_{t-1}, \mathbf{y})$. Incorporating these types of statistics in (4.3) under the Poisson reference measure yields:

$$X_{i,t} | \mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{Y} = \mathbf{y} \sim \text{Pois}(\lambda = \exp\{\mathbf{s}_i(\mathbf{x}_{t-1}, \mathbf{y})\}), \quad (4.4)$$

where $\mathbf{s}_i(\mathbf{x}_{t-1}, \mathbf{y})$ is an arbitrary function of past attributes, network \mathbf{y} , and exogenous information, which we again omit for notational brevity. Similar models for binary networks were proposed in Hanneke et al. (2010) and Almquist and Butts (2014), while they were applied in Almquist and Butts (2013) and Lebacher et al. (2021).

Returning to the example stated above concerning people infected with COVID-19, the attributes are often multivariate. In Germany, the federal government agency responsible for health reporting and disease control provides, for instance, infection data stratified according to the age and gender of the population in a particular district. Let therefore the infection count in week t in age/gender group g and district i be given by $X_{i,g,t}$. Moreover, we might not only have information on one but

multiple networks, where some are time-constant and others vary over time. In our example, all networks have 401 nodes representing all federal districts in Germany. In particular, we are given a weighted network detailing the number of Facebook friends between these districts, denoted by \mathbf{y}^1 , as well as networks of the co-location probabilities, i.e., the probability of meeting one another. While the first network \mathbf{y}^1 is measured before $t = 1$, therefore, constant over time, the co-location networks are subject to change each week and denoted by $\mathbf{y}_1^2, \dots, \mathbf{y}_T^2$. Extending (4.4) to such a data setting yields:

$$X_{i,g,t} | X_{i,g,t-1} = x_{i,g,t-1}, \mathbf{Y}^1 = \mathbf{y}^1, \mathbf{Y}_{t-1}^2 = \mathbf{y}_{t-1}^2 \sim \text{Pois}(\lambda = \exp\{\mathbf{s}_{i,g,t}(x_{i,g,t-1}, \mathbf{y}^1, \mathbf{y}_{t-1}^2)\}). \quad (4.5)$$

Note that the temporal delay of one week is legitimized by the finding that the incubation period is around that time frame (Li et al., 2020). The specification of $\mathbf{s}_{i,g,t}(\mathbf{y}^1, \mathbf{y}_{t-1}^2)$ can include any actor-level statistics such as the statistics in (2.2). One can also view the positions of the nodes found through a graph drawing algorithm as a two-dimensional measure of how actors are embedded in a particular network. We follow this path in Chapter 7 to find the position of each district in \mathbf{y}^1 through multidimensional scaling as proposed by Brandes and Pich (2007). Subsequently, we can incorporate the found coordinates in (4.5) via isotropic multivariate splines (Wood, 2003). We use the Gini index to process the time-varying networks of co-location probabilities to the actor level. This measure allows us to differentiate between mobility patterns of districts being rather dispersed or restricted. In Chapter 7, we detail how to carry out those two tasks. Finally, note that by using different reference measures in (4.3) than the Poisson reference measure, one may generalize the Poisson distribution in (4.5) to any count distribution, such as the negative binomial distribution for a fixed dispersion parameter.

5. Discussion and Future Work

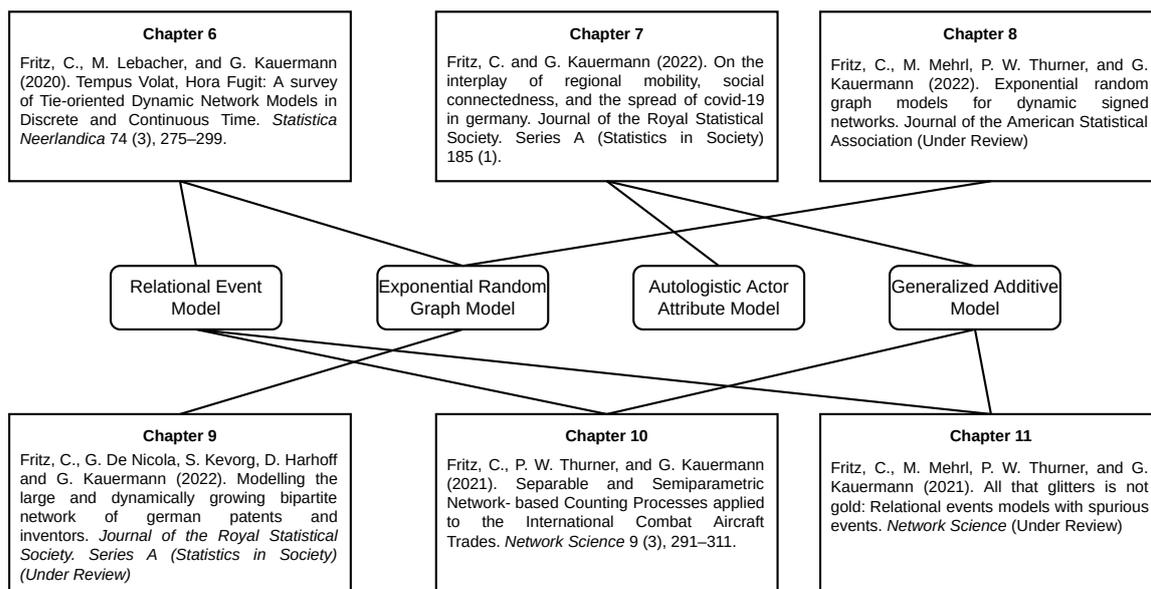


Figure 5.1.: Bipartite graph illustrating which contributed manuscripts relate to or build on which core methods introduced in this introduction. Rectangles indicate the core methods if they have rounded corners and the manuscripts if they have angular corners. Ties translate to manuscripts relating to a particular core method.

The field of social network data analysis has grown continuously since its origin in [Moreno \(1934\)](#). Given the resulting scientific breadth and depth, this overview is restricted to statistical network analysis with the ERGM and REM. Nevertheless, this sub-field already offers a powerful modeling framework taking the dependencies between the observed relations and attributes seriously. This is a meaningful step towards specifying realistic models in more complex data scenarios. How the ensuing manuscripts contributing to this dissertation are interwoven with those foundational techniques can be best illustrated using techniques introduced in [Section 2.1](#). [Figure 5.1](#) displays a bipartite graph where \mathcal{V}_1 contains the contributed manuscript, shown by rectangles with angular corners, and \mathcal{V}_2 consists of the core methods introduced in the last chapter, depicted by rectangles with round corners. Each chapter encompasses one of these manuscripts, which are provided in the published version for the most part. Now, the contributions and aims of each manuscript are detailed separately for each project:

Chapter 6, [Fritz, Lebacher, and Kauermann \(2020\)](#): We offer a survey of dynamic network models for networks observed at specific times and networks available at a more fine-grained temporal resolution. By doing that, we showcase the theoretical properties, fitting procedures, applicability, and interpretation of the canonical models for each setting.

Chapter 7, Fritz and Kauermann (2022): Motivated by the surge of the COVID-19 pandemic starting in 2020, we investigate the effect of human mobility and social connectivity on the spread of the virus on a local scale. With data provided by Facebook, we can address this task based on the weekly rate of new infections in Germany during the first COVID-19 wave. First, we find evidence that reduced social activities lower the incidence of infections. Second, our results suggest that social distances affect infections next to geographical distances.

Chapter 8, Fritz, Mehrl, Thurner, and Kauermann (2022b): Stemming from the importance of Structural Balance Theory (Heider, 1946), signed networks are among the most frequently studied matters in social network analysis. Therefore, we extend the ERGM to such networks in static and dynamic settings. We then apply the novel method to recent data on militarized interstate disputes and defense cooperation agreements and provide a software implementation with the R package `ergm.sign`.

Chapter 9, Fritz, De Nicola, Kevorg, Harhoff, and Kauermann (2022): The goal of this chapter is to study the dynamically growing bipartite network of German patents and inventors by means of a Temporal Exponential Random Graph Model (TERGM, Hanneke et al., 2010). Standard procedures are infeasible when considering the emergent network of patents only within the main area of electrical engineering since the network encompasses 78.412 actors. To circumvent this, we decompose the network into several smaller subnetworks by exploiting inventors' "mortality" and "natality". The reasoning for this is that inventors are, in most cases, only active during a relatively short period of the studied period from 2000 to 2015. To explore the driving forces behind innovation, we further introduce a suite of network terms to capture and test for network characteristics, such as team persistence and collaboration interlocking.

Chapter 10, Fritz, Thurner, and Kauermann (2021): The duration of ties in networks varies greatly. For instance, friendships are generally assumed to be active over a long period, while emails are better comprehended as instantaneous events. While durable links are commonly only available at specific instances, most events are observed continuously. For some events, namely the delivery of combat aircraft, we can, however, only determine the temporal resolution up to a specific year leading to time-clustered observations. To model such data while accounting for time-varying and random coefficients, we generalize the Relational Event Model (REM, Butts, 2008). We further rely on separable models as introduced by Krivitsky and Handcock (2014) and model the onset and repetition of yearly observed events with two separate intensity functions. The proposed model is applied to the international combat aircraft trade network spanning from 1950 to 2017.

Chapter 11, Fritz, Mehrl, Thurner, and Kauermann (2022a): To ease the collection of event data, researchers increasingly rely on automated or human-coded event data, which may include spurious events, i.e., false-positive events. We propose the Relational Event Model for Spurious Events (REMSE) as an extension to the REM to control for such spurious events. The practical relevancy of the proposed model is exhibited in applications to event data from combat events in the Syrian civil war and spatial proximity of students.

Beyond these topics, there are still numerous challenges in network science, guiding the way for future research. For instance, one might wonder how dependence graphs from Section 3.1 behave as networks grow. As noted by Strauss (1986), the number of neighbors of a node in the dependence graph under Markov dependence is $2(N - 2)$. Thus, one implicitly assumes that the dependency of relations increases as the network expands. On the other hand, this behavior goes against our heuristic understanding of sparsity in large graphs that the density decreases as the network increases. Also, from the perspective of analytical sociology, it is hardly reasonable

that actors have the capacities to maintain increasing relations or are even aware of the complete growing network. Recalling the discussion at the beginning of Section 3.2, the availability of large networks through novel technologies further underlines how important the consideration of models for large network data is. It thus does not suffice to propose scalable estimation routines for the network models introduced in this overview if the estimated models cannot mimic the behavior of observed networks due to unrealistic dependence assumptions. A promising concept to help in this regard is local dependence as described in [Schweinberger and Handcock \(2015\)](#). Under this type of dependence, the network is decomposed into several sub-networks based on an either observed or unobserved grouping of the actors. For this approach, complex dependence structures akin to the ones from Section 3.1 are assumed solely within the sub-networks, while ties between sub-networks abide by simpler forms of dependence. Within this framework, smaller dense networks nested within larger, more sparse networks are represented more naturally. Although much theoretical work considers attractive properties under local dependence, such as consistency under a growing number of groups ([Schweinberger and Stewart, 2020](#)) and the recoverability of unknown groups ([Schweinberger, 2020](#)), the presentation of the model is restricted to static binary networks; hence extensions to temporal or bipartite networks, where adapted motions of local dependency are necessary, represent an interesting future research direction. Moreover, extending the principles of local dependence to other model classes, e.g., the REM, seems to be a fruitful premise.

The topic of large networks immediately brings the sampling of networks to mind. Generally, the classical statistical setting, in which models are estimated only with a sample from the population of interest and inference is drawn from the sample onto the population, should be critically reassessed when working with network data. Contrasting the standard setting in applied statistical work, different types of infinite population scenarios are possible for network data as described in [Schweinberger et al. \(2020\)](#). One setting is where we only observe a sub-network nested within a larger network, and the researcher wants to learn about this unobserved network from the sub-network. Although [Shalizi and Rinaldo \(2013\)](#) show that only specific dependence assumptions lead to consistency under sampling for ERGMs, one can account for a particular type of sampling design within the estimation procedure of the model. A general framework to account for sampling is provided by [Handcock and Gile \(2012\)](#) if the size of the large network is known and not too big, while more niche techniques were proposed for snowball ([Pattison et al., 2013](#)) and egocentric sampling ([Krivitsky and Morris, 2017](#)). Especially within the local dependence framework, both approaches have much potential, and designing sampling strategies under local dependence can greatly ease the estimation of models. One might argue that those issues on inference go well beyond technicalities but should also be brought to the center of attention in the substantive sciences, where network data and theories naturally arise. Besides, concerns about sampling mainly affect networks representable under the “*graph-as-networks*” paradigm, although they are equivalently relevant for event data covered in Section 2.2. In Chapter 11, we take the first steps by controlling for spurious events, but more general treatments of sampling designs for this type of model are clearly needed.

The surge of event data further suggests that methods to study the effects of events on continuous traces of attributes are needed to catch up with recent advances in digital technologies. Paralleling the development of the REM, one practical way forward could be adapting known techniques for the analysis of marked point processes ([Sun et al., 2009](#)) or egocentric REMs ([Vu et al., 2011b](#)) to attributes. In general, these methods should be tailored to the type of continuous attribute measured. For instance, continuously measured heart rates necessitate models that are different than those used for categorical states, such as sleeping, walking, or running. Data sources similar

to [Aharony et al. \(2011\)](#) further include attributes measured at the discrete and continuous scale; hence methods combining event and network data are also needed. To sum up, there still remains much to be done for the statistical analysis of network data. This thesis presents first steps in several directions.

Bibliography

- Aalen, O. O., O. Borgan, and H. K. Gjessing (2008). *Survival and event history analysis: A process point of view*. New York: Springer.
- Abbott, A. (1997). Of time and space: The contemporary relevance of the Chicago school. *Social Forces* 75(4), 1149.
- Aharony, N., W. Pan, C. Ip, I. Khayal, and A. Pentland (2011). SocialfMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing* 7(6), 643–659.
- Almquist, Z. W. and C. T. Butts (2013). Dynamic network logistic regression: A logistic choice analysis of inter- and intra-group blog citation dynamics in the 2004 us presidential election. *Political Analysis* 21(4), 430–448.
- Almquist, Z. W. and C. T. Butts (2014). Logistic network regression for scalable analysis of networks with joint edge/ vertex dynamics. *Sociological Methodology* 44(1), 273–321.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 36(2), 192–236.
- Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca (2009). Network analysis in the social sciences. *Science* 323(5916), 892–895.
- Bott, E. (1957). *Family and social network*. Hove: Psychology Press.
- Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association* 71(356), 791–799.
- Brandes, U., J. Lerner, and T. A. Snijders (2009). Networks evolving step by step: Statistical analysis of dyadic event data. In *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, pp. 200–205.
- Brandes, U. and C. Pich (2007). Eigensolver methods for progressive multidimensional scaling of large data. In *Graph Drawing 14th International Symposium*, pp. 42–53.
- Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology* 11(1), 13–41.
- Butts, C. T. and C. S. Marcum (2017). A relational event approach to modeling behavioral dynamics. In A. Pilny and M. S. Poole (Eds.), *Group Processes*, pp. 51–92. Cham: Springer.
- Byshkin, M., A. Stivala, A. Mira, R. Krause, G. Robins, and A. Lomi (2016). Auxiliary Parameter MCMC for Exponential Random Graph Models. *Journal of Statistical Physics* 165(4), 740–754.

-
- Carrington, P. J., J. Scott, and S. S. Wasserman (2005). *Models and methods in social network analysis*. Structural analysis in the social sciences. Cambridge: Cambridge University Press.
- Casella, G. and R. Berger (2001). *Statistical inference*. Belmont: Brooks/Cole.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 34(2), 187–202.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* 62(2), 269–276.
- Crane, H. (2018). *Probabilistic Foundations of Statistical Network Analysis*. Boca Raton: CRC Press.
- Cranmer, S. J., B. A. Desmarais, and J. W. Morgan (2021). *Inferential network analysis*. Cambridge: Cambridge University Press.
- Daley, D. J. and D. Vere-Jones (2008). *An introduction to the theory of point processes: Volume I: Elementary Theory and methods*. New York: Springer.
- Daraganova, G. and G. L. Robins (2013). Autologistic actor attribute models. In *Exponential random graph models for social networks*, Chapter 9, pp. 102–113. Cambridge: Cambridge University Press.
- Desmarais, B. A. and S. J. Cranmer (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PLoS ONE* 7(1), 1–12.
- Eagle, N. and A. Pentland (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* 10(4), 255–268.
- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.
- Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13(1), 342–368.
- Faust, K. and J. Skvoretz (1999). Logit models for affiliation networks. *Sociological Methodology* 31, 253–280.
- Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association* 81(395), 832–842.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks* 1(3), 215–239.
- Fritz, C., G. De Nicola, S. Kevorg, D. Harhoff, and G. Kauermann (2022). Modelling the large and dynamically growing bipartite network of german patents and inventors. *arXiv preprint 2201.09744*.
- Fritz, C. and G. Kauermann (2022). On the interplay of regional mobility, social connectedness, and the spread of covid-19 in germany. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 185(1).

Bibliography

- Fritz, C., M. Lebacher, and G. Kauermann (2020). Tempus Volat, Hora Fugit: A survey of Tie-oriented Dynamic Network Models in Discrete and Continuous Time. *Statistica Neerlandica* 74(3), 275–299.
- Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2022a). All that glitters is not gold: Relational events models with spurious events. *arXiv preprint 2109.10348*.
- Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2022b). Exponential random graph models for dynamic signed networks. *arXiv preprint*.
- Fritz, C., P. W. Thurner, and G. Kauermann (2021). Separable and Semiparametric Network-based Counting Processes applied to the International Combat Aircraft Trades. *Network Science* 9(3), 291–311.
- Fruchterman, T. M. J. and E. M. Reingold (1991). Graph drawing by force-directed placement. *Software: Practice and Experience* 21(11), 1129–1164.
- Geyer, C. J. (1992). On the convergence of monte carlo maximum likelihood calculations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 56(1), 261–274.
- Gibson, D. R. (2003). Participation shifts: Order and differentiation in group conversation. *Social Forces* 81(4), 1335–1380.
- Goldenberg, A., A. X. Zheng, S. E. Fienberg, E. M. Airoldi, and Others (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning* 2(2), 129–233.
- Goodreau, S. M., J. A. Kitts, and M. Morris (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography* 46(1), 103–125.
- Granovetter, M. (1973). The strength of weak ties. *The American Journal of Sociology* 78(6), 1360–1380.
- Hammersley, J. M. and P. Clifford (1971). Markov fields on finite graphs and lattices. *Unpublished Manuscript*, 1–26.
- Handcock, M. (2003). Assessing degeneracy in statistical models of social networks. Technical report, University of Washington.
- Handcock, M. S. and K. J. Gile (2012). Modeling social networks from sampled data. *Annals of Applied Statistics* 6(1), 5–25.
- Hanneke, S., W. Fu, and E. P. Xing (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* 4, 585–605.
- Hastie, T. and R. Tibshirani (1987). Generalized additive models: some applications. *Journal of the American Statistical Association* 82(398), 371–386.
- Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 55(4), 757–796.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1), 83–90.

- Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology* 21(1), 107–112.
- Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97(460), 1090–1098.
- Holland, P. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* 76(373), 33–50.
- Hummel, R. M., D. R. Hunter, and M. S. Handcock (2012). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics* 21(4), 920–939.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481), 248–258.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2013). Ergm.userterms: A template package for extending statnet. *Journal of Statistical Software* 52(2), 1–25.
- Hunter, D. R. and M. S. Handcock (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3), 565–583.
- Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data*. Hoboken: J. Wiley.
- Kass, R. E. (2011). Statistical Inference: The Big Picture. *Statistical Science* 26(1), 1–9.
- Kemp, C., J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence* 1, pp. 381–388.
- Kim, B., K. H. Lee, L. Xue, and X. Niu (2018). A review of dynamic network models with latent variables. *Statistical Surveys* 12, 105–135.
- Kimeldorf, G. S. and G. Wahba (1970). A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics* 41(2), 495–502.
- Kinne, B. J. (2020). The defense cooperation agreement dataset (DCAD). *Journal of Conflict Resolution* 64(4), 729–755.
- Kitts, J. A. (2014). Beyond networks in structural theories of exchange: Promises from computational social science. In *Advances in Group Processes*, Volume 31, Chapter 9, pp. 263–298. Bingley: Emerald Group Publishing Limited.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data. Methods and models*. New York: Springer.
- Kolaczyk, E. D. (2017). *Topics at the frontier of statistics and network analysis: (Re)visiting the foundations*. Cambridge: Cambridge University Press.
- Kolaczyk, E. D., G. Csárdi, and K. Nordhausen (2009). *Statistical analysis of network data with R*. New York: Springer.
- Koskinen, J., A. Caimo, and A. Lomi (2015). Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to foreign direct investments. *Network Science* 3(1), 58–77.

Bibliography

- Koskinen, J. and A. Lomi (2013). The local structure of globalization: The network dynamics of foreign direct investments in the international electricity industry. *Journal of Statistical Physics* 151(3-4), 523–548.
- Kreiss, A. (2021). Correlation bounds, mixing and m-dependence under random time-varying network distances with an application to Cox-processes. *Bernoulli* 27(3), 1666–1694.
- Kreiß, A., E. Mammen, and W. Polonik (2019). Nonparametric inference for continuous-time event counting and link-based dynamic network models. *Electronic Journal of Statistics* 13(2), 2764–2829.
- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics* 6, 1100–1128.
- Krivitsky, P. N. and C. T. Butts (2017). Exponential-family random graph models for rank-order relational data. *Sociological Methodology* 47(1), 68–112.
- Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 76(1), 29–46.
- Krivitsky, P. N., L. M. Koehly, and C. S. Marcum (2020). Exponential-family random graph models for multi-layer networks. *Psychometrika* 85(3), 630–659.
- Krivitsky, P. N. and E. D. Kolaczyk (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science* 30(2), 184–198.
- Krivitsky, P. N. and M. Morris (2017). Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Annals of Applied Statistics* 11(1), 427–455.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford: Oxford University Press.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Hoboken, NJ: Wiley.
- Lazer, D., A. Pentland, L. Adamic, S. Aral, A. L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne (2009). Computational social science. *Science* 323(5915), 721–723.
- Lebacher, M., P. W. Thurner, and G. Kauermann (2021). A dynamic separable network model with actor heterogeneity: An application to global weapons transfers*. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 184(1), 201–226.
- Lerner, J., M. Bussmann, T. A. B. Snijders, and U. Brandes (2013). Modeling Frequency and Type of Interaction in Event Networks. *Corvinus Journal of Sociology and Social Policy* 4(1), 3–32.
- Lerner, J. and A. Lomi (2022). A dynamic model for the mutual constitution of individuals and events. *Journal of Complex Networks* 10(2), 1–30.
- Lerner, J., A. Lomi, J. Mowbray, N. Rollings, and M. Tranmer (2021). Dynamic network analysis of contact diaries. *Social Networks* 66, 224–236.

- Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine* 382(13), 1199–1207.
- Luce, R. D. and A. D. Perry (1949). A method of matrix analysis of group structure. *Psychometrika* 14(2), 95–116.
- Lusher, D., J. Koskinen, and G. Robins (2012). *Exponential random graph models for social networks*. Cambridge: Cambridge University Press.
- Matias, C. and S. Robin (2014). Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys* 47, 55–74.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1), 415–444.
- Miller, K. T., T. L. Griffiths, and M. I. Jordan (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems 22*, pp. 1276–1284.
- Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human interrelations*. Washington: Nervous and Mental Disease Publishing Co.
- Nelder, J. A. (1999). From statistics to statistical science. *Journal of the Royal Statistical Society. Series D (The Statistician)* 48, 257–269.
- Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 135(3), 370.
- Niezink, N. M. and T. A. Snijders (2017). Co-evolution of social networks and continuous actor attributes. *Annals of Applied Statistics* 11(4), 1948–1973.
- Palmer, G., V. D’Orazio, M. Kenwick, and M. Lane (2015). The MID4 data set, 2002-2010. *Conflict Management and Peace Science* 32(2), 222–242.
- Park, J. and M. E. J. Newman (2004). The statistical mechanics of networks. *Physical Review E* 70(6), 1–13.
- Pattison, P. and G. Robins (2002). Neighborhood-based models for social networks. *Sociological Methodology* 32(1), 301–337.
- Pattison, P. E., G. L. Robins, T. A. B. Snijders, and P. Wang (2013). Conditional estimation of exponential random graph models from snowball sampling designs. *Journal of Mathematical Psychology* 57, 284–296.

Bibliography

- Pattison, P. E. and S. S. Wasserman (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52(2), 169–193.
- Perry, P. O. and P. J. Wolfe (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 75(5), 821–849.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Renshaw, S. L., S. M. Livas, M. G. Petrescu-Prahova, and C. T. Butts (2022). Modeling Complex Interactions in a Disrupted Environment: Relational Events in the WTC Response. *arXiv preprint 2204.07890*.
- Robins, G., P. Elliott, and P. Pattison (2001). Network models for social selection processes. *Social Networks* 23(1), 1–30.
- Robins, G. and P. Pattison (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology* 25(1), 5–41.
- Robins, G., P. Pattison, and P. Elliott (2001). Network models for social influence processes. *Psychometrika* 66(2), 161–189.
- Robins, G., P. Pattison, Y. Kalish, and D. Lusher (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks* 29(2), 173–191.
- Robins, G., P. Pattison, and S. Wasserman (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika* 64(3), 371–394.
- Robins, G. L. and P. E. Pattison (2005). Interdependencies and social processes: Dependence graphs and generalized dependence structures. In P. J. Carrington, J. Scott, and S. S. Wasserman (Eds.), *Models and methods in social network analysis*, Chapter 10, pp. 192–214. Cambridge: Cambridge University Press.
- Salathé, M., D. Q. Vu, S. Khandelwal, and D. R. Hunter (2013). The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science* 2(1), 1–12.
- Schrodt, P. A., S. G. Davis, and J. L. Weddle (1994). Political science: KEDS- A program for the machine coding of event data. *Social Science Computer Review* 12(4), 561–587.
- Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association* 106(496), 1361–1370.
- Schweinberger, M. (2020). Consistent structure estimation of exponential-family random graph models with block structure. *Bernoulli* 26(2), 1205–1233.
- Schweinberger, M. and M. S. Handcock (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 77(3), 647–676.
- Schweinberger, M., P. N. Krivitsky, C. T. Butts, and J. R. Stewart (2020). Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios. *Statistical Science* 35(4), 627–662.

- Schweinberger, M. and J. Stewart (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *The Annals of Statistics* 48(1), 374–396.
- Shalizi, C. R. and A. Rinaldo (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics* 41(2), 508–535.
- Sischka, B. and G. Kauermann (2022). EM-based smooth graphon estimation using MCMC and spline-based approaches. *Social Networks* 68, 279–295.
- Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology* 21, 149–172.
- Snijders, T. A. B. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure* 3(2), 1–40.
- Snijders, T. A. B. and J. Koskinen (2013). Longitudinal Models. In *Exponential random graph models for social networks*, Structural Analysis in the Social Sciences, Chapter 11, pp. 130–140. Cambridge: Cambridge University Press.
- Snijders, T. A. B., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology* 36(1), 99–153.
- Snijders, T. A. B., C. Steglich, and M. Schweinberger (2018). Modeling the co-evolution of networks and behavior. In *Longitudinal Models in the Behavioral and Related Sciences*, pp. 41–71.
- Stadtfeld, C. (2012). *Events in social networks : A stochastic actor-oriented framework for dynamic event processes in social networks*. Ph. D. thesis, KIT.
- Stadtfeld, C. (2018). The micro–macro link in social networks. In R. A. Scott, M. Buchmann, and S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences*, Chapter 1, pp. 1–15. Hoboken: John Wiley & Sons, Ltd.
- Stadtfeld, C., J. Hollway, and P. Block (2017). Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology* 47(1), 1–40.
- Strauss, D. (1986). On a general class of models for interaction. *SIAM Review* 28(4), 513–527.
- Strauss, D. and M. Ikeda (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* 85(409), 204–212.
- Sun, Y., P. B. Gilbert, and I. W. Mckeague (2009). Proportional hazards models with continuous marks. *Annals of Statistics* 37(1), 394–426.
- van Duijn, M. A., K. J. Gile, and M. S. Handcock (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks* 31(1), 52–62.
- van Duijn, M. A., T. A. B. Snijders, and B. J. H. Zijlstra (2004). p2: A random effects model with covariates for directed graphs. *Statistica Neerlandica* 58(2), 234–254.
- Vega Yon, G. G., A. Slaughter, and K. de la Haye (2021). Exponential random graph models for little networks. *Social Networks* 64, 225–238.

Bibliography

- Vu, D. Q., A. U. Asuncion, D. R. Hunter, and P. Smyth (2011a). Continuous-time regression models for longitudinal networks. In *Advances in Neural Information Processing Systems 24*, pp. 2492–2500.
- Vu, D. Q., A. U. Asuncion, D. R. Hunter, and P. Smyth (2011b). Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 857–864.
- Vu, D. Q., A. Lomi, D. Mascia, and F. Pallotti (2017). Relational event models for longitudinal network data with an application to interhospital patient transfers. *Statistics in Medicine* 36(14), 2265–2287.
- Wang, P., P. Pattison, and G. Robins (2013). Exponential random graph model specifications for bipartite networks-A dependence hierarchy. *Social Networks* 35(2), 211–222.
- Wang, P., K. Sharpe, G. L. Robins, and P. E. Pattison (2009). Exponential random graph (p^*) models for affiliation networks. *Social Networks* 31(1), 12–25.
- Wasserman, S. and K. Faust (1994). *Social network analysis : methods and applications*. Cambridge: Cambridge University Press.
- Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika* 61(3), 401–425.
- Wasserman, S., G. Robins, and D. Steinley (2007). Statistical models for networks: A brief review of some recent research. In *ICML 2006 workshop on statistical network analysis*, pp. 45–56.
- Wasserman, S. S. and G. L. Robins (2005). An Introduction to Random Graphs, Dependence Graphs, and p^* . In P. J. Carrington, J. Scott, and S. S. Wasserman (Eds.), *Models and methods in social network analysis*, Chapter 8, pp. 148–161. Cambridge: Cambridge University Press.
- Whitehead, J. (1980). Fitting cox’s regression model to survival data using glim. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29(3), 268.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65(1), 95–114.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Boca Raton: CRC Press.
- Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* 44(4), 1019–1031.

Part II.

Statistical Modeling of Networks

6. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time

Contributing article

Fritz, C., Lebacher, M. and Kauermann, G. (2020) Tempus Volat, Hora Fugit: A survey of Tie-oriented Dynamic Network Models in Discrete and Continuous Time. *Statistica Neerlandica*, 74 (3), 275–299. doi:10.1111/stan.12198.

Replication code

The complete replication code for this article is available as supporting information under <https://onlinelibrary.wiley.com/doi/full/10.1111/stan.12198>.

Copyright information

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited.

Author contributions

The general idea of contrasting temporal network models in continuous- and discrete-time can be attributed to Göran Kauermann. Cornelius Fritz, as the leading author, wrote the section on continuous-time network models together with the implementation of these models in R. He also created most visualizations in the paper. The introduction, the definitions section, as well as the application section, were written in joint work by Cornelius Fritz and Michael Lebacher. Michael Lebacher wrote Section 3 on dynamic exponential random graph models and implemented all models as well as goodness-of-fit measures in this section in R. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

Received: 24 May 2019 | Revised: 29 October 2019 | Accepted: 4 November 2019

DOI: 10.1111/stan.12198

SPECIAL ISSUE ARTICLE

WILEY

Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time

Cornelius Fritz  | Michael Lebacher | Göran Kauermann

Department of Statistics,
Ludwig-Maximilians-Universität
München, Munich, Germany

Correspondence

Cornelius Fritz, Department of Statistics,
Ludwig-Maximilians-Universität
München, Ludwigstr. 33, 80539 Munich,
Germany.

Email:
cornelius.fritz@stat.uni-muenchen.de

Funding information

European Cooperation in Science and
Technology, Grant/Award Number:
COST Action CA15109 (COSTNET);
German Research Foundation (DFG),
Grant/Award Number: KA 1188/10-1;
Munich Center for Machine Learning
(MCML)

Given the growing number of available tools for modeling dynamic networks, the choice of a suitable model becomes central. The goal of this survey is to provide an overview of tie-oriented dynamic network models. The survey is focused on introducing binary network models with their corresponding assumptions, advantages, and shortfalls. The models are divided according to generating processes, operating in discrete and continuous time. First, we introduce the temporal exponential random graph model (TERGM) and the separable TERGM (STERGM), both being time-discrete models. These models are then contrasted with continuous process models, focusing on the relational event model (REM). We additionally show how the REM can handle time-clustered observations, that is, continuous-time data observed at discrete time points. Besides the discussion of theoretical properties and fitting procedures, we specifically focus on the application of the models on two networks that represent international arms transfers and email exchange, respectively. The data allow to demonstrate the applicability and interpretation of the network models.

KEYWORDS

continuous-time, discrete-time, ERGM, event modeling, random graphs, REM, STERGM, TERGM

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Statistica Neerlandica* published by John Wiley & Sons, Ltd. on behalf of VVS

1 | INTRODUCTION

The conceptualization of systems within a network framework has become popular within the last decades; see Kolaczyk (2009) for a broad overview. This is mostly because network models provide useful tools for describing complex dependence structures and are applicable to a wide variety of research fields. In the network approach, the mathematical structure of a graph is utilized to model network data. A graph is defined as a set of nodes and relational information (ties) between them. Within this concept, nodes can represent individuals, countries, or general entities, whereas ties are connections between those nodes. Dependent on the context, these connections can represent friendships in a school (Raabe, Boda, & Stadtfeld, 2019), transfers of goods between countries (Ward, Ahlquist, & Rozenas, 2013), sexual relations between people (Bearman, Moody, & Stovel, 2004), or hyperlinks between websites (Leskovec, Lang, Dasgupta, & Mahoney, 2009) to name just a few. Given a suitable data structure for the system of interest, the conceptualization as a network enables analyzing dependencies between ties. A central statistical model that allows this is the exponential random graph model (ERGM, Robins & Pattison, 2001). This model permits the inclusion of monadic, dyadic, and hyperdyadic features within a regression-like framework.

Although the model allows for an insightful investigation of *within-network* dependencies, most real-world systems are typically more complex. This is especially true if a temporal dimension is added, which is relevant, as most systems commonly described as networks evolve dynamically over time. It can even be argued that most static networks are *de facto* not static but snapshots of a dynamic process. A friendship network, for example, typically evolves over time and influences like reciprocity often follow a natural chronological order.

Of course, this is not the first paper concerned with reviewing temporal network models. Goldenberg, Zheng, Fienberg, and Airoldi (2010) wrote a general survey covering a wide range of models. The authors laid the foundation for further articles and postulated a soft division of statistical network models into latent space (Hoff, Raftery, & Handcock, 2002) and p_1 models (Holland & Leinhardt, 1981), all originating in the Erdős-Rényi-Gilbert random graph models (Erdős & Rényi, 1959; Gilbert, 1959). Kim, Lee, Xue, and Niu (2018) give a contemporary update on the field of dynamic models building on latent variables. Snijders (2005) discusses continuous-time models and reframes the independence and reciprocity model as a stochastic actor-oriented model (SAOM; Snijders, 1996). Block, Koskinen, Hollway, Steglich, and Stadtfeld (2018) provide an in-depth comparison of the temporal ERGM (TERGM, Hanneke, Fu, & Xing, 2010) and the SAOM with special focus on the treatment of time. Furthermore, the ERGM and SAOM for networks that are observed at single time points are contrasted by Block, Stadtfeld, and Snijders (2019), deriving theoretical guidelines for model selection based on the differing mechanics implied by each model.

In the context of this compendium of articles, the scope is to give an update on the dynamic variant of the second strand of models relating to p_1 models. We therefore extend the summarizing diagram of Goldenberg et al. (2010), as depicted in Figure 1. Generally, we divide temporal models into two sections, by differentiating between discrete and continuous-time network models. This review paper will focus on tie-oriented models. Tie-oriented models are concerned with formulating a stochastic model for the existence of a tie in contrast to the actor-oriented approach by Snijders (2002), which specifies the model from the actor's point of view (Block et al., 2018). The dynamical actor-oriented model (DyNAM, Stadtfeld & Block, 2017) adopts this actor-oriented paradigm to event data. This type of model was formulated with a focus on social networks

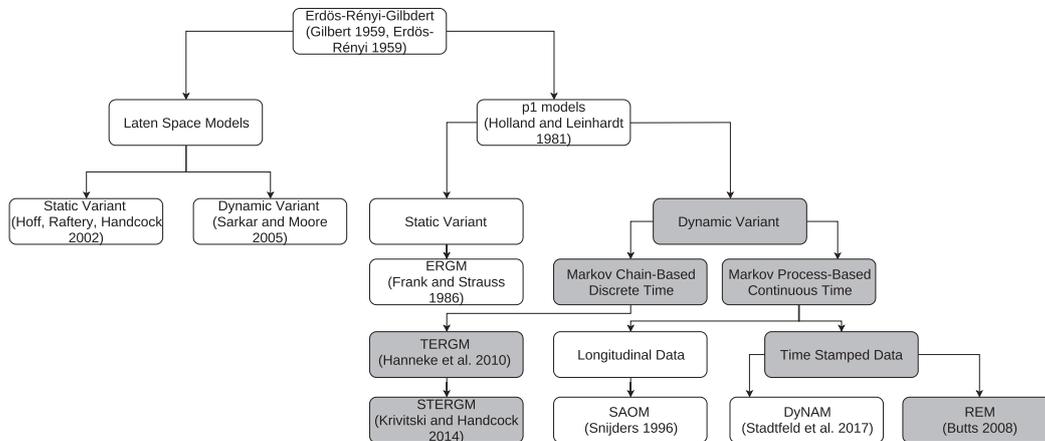


FIGURE 1 Tree diagram summarizing the dependencies between models originating in the Erdős-Rényi-Gilbert graph model; the models situated in a box with a grey background are discussed in this article. This graph is an update of Figure 6.1 in Goldenberg, Zheng, Fienberg, and Airoldi (2010). ERGM = exponential random graph model; TERGM = temporal exponential random graph model; STERGM = separable temporal exponential random graph model; SAOM = stochastic actor-oriented model; DyNAM = dynamical actor-oriented model; REM = relational event model

(Snijders, 1996). Tie-oriented models, on the other hand, can be viewed as more general because they are also applicable to nonsocial networks.

Statistical models for time discrete data rely on an autoregressive structure and condition the state of the network at time point t on previous states. This includes the TERGM and the separable TERGM (STERGM; Krivitsky & Handcock, 2014). There exists a variety of recent applications of the TERGM. White, Forester, and Craft (2018) use a TERGM for modeling epidemic disease outcomes and Blank, Dincecco, and Zhukov (2017) investigate interstate conflicts. In He, Dong, Wu, Jiang, and Zheng (2019), Chinese patent trade networks are inspected, and Benton and You (2017) use a TERGM for analyzing shareholder activism. Applications of STERGMs are given, for example, by Stansfield et al. (2019) that model sexual relationships and by Broekel and Bednarz (2018) that study the network of research and development cooperation between German firms.

In case of time-continuous data, the model regards the network as a continuously evolving system. Although this evolution is not necessarily observed in continuous time, the process is taken to be latent and explicitly models the evolution from the state of the network at time point $t - 1$ to t (Block et al., 2018). In this paper, we discuss the relational event model (REM, Butts, 2008) for the analysis of event data. Applications of the REM are manifold and range from explaining the dynamics of health behavior sentiments via Twitter (Salathé, Vu, Khandelwal, & Hunter, 2013), interhospital patient transfers (Vu, Lomi, Mascia, & Pallotti, 2017), online learning platforms (Vu, Pattison, & Robins, 2015), and animal behavior (Tranmer, Marcum, Morton, Croft, & de Kort, 2015) to structures of project teams (Quintane, Pattison, Robins, & Mol, 2013). Eventually, the REM is adapted to time-discrete observations of networks. That is, we observe the time-continuous developments of the network at discrete observation times only. Henceforth, we use the term time-clustered for this special data structure.

In reviewing dynamic network models, we assume a temporal first-order Markov dependency. To be more specific, this implies that the network at time point t only depends on the previous observation of the network. This characteristic is widely used in the analysis of longitudinal

networks (Hanneke et al., 2010; Krivitsky & Handcock, 2014) and the resulting conditional independence among states of the network facilitates the estimation with an arbitrary number of time points. In that respect, it suffices to only include two observational moments for illustrative purposes because the interpretation and estimation with a longer series of networks is unchanged. This allows for a clear-cut comparison of the methods at hand.

This paper is structured as follows. In Section 2, we give basic definitions that are used throughout this paper and present the two data examples that will be analyzed as illustrative examples. After that, Section 3 introduces a time-discrete network model and Section 4, a time-continuous network model. They are applied in Section 5 on two data sets and Section 6 concludes. Additional results relating to the applications can be found in the Supplementary Material.

2 | DEFINITIONS AND DATA DESCRIPTION

2.1 | Definitions

This article regards directed binary networks, with ties representing directed relations between two nodes at a time point. The respective information can be represented in an adjacency matrix $Y_t = (Y_{ij,t})_{i,j=1,\dots,n} \in \mathcal{Y}$, where $\mathcal{Y} = \{Y : Y \in \{0, 1\}^{n \times n}\}$ represents the set of all possible networks with n nodes. The entry (i, j) of Y_t is “1” if a tie is outgoing from node i to j in year t and “0,” otherwise. Furthermore, the discrete time points of the observations of Y_t are denoted as $t = 1, \dots, T$. We restrict our analysis to two time points in both exemplary networks, which suffices for comparison. Hence, we set $T = 2$. In many networks, including our running examples, self-loops are meaningless. We therefore fix $Y_{ii,t} \equiv 0 \forall i \in \{1, \dots, n\}$ throughout the article. Furthermore, all subscripted temporal indices (Y_t) are assumed to take discrete values and all indices in brackets ($Y(t)$) continuous values. The temporal indicator t denotes the observation times of the network, and to notationally differ this from time-continuous model, we write \tilde{t} for continuous time.

To sufficiently compare different models, we use two application cases. The first one represents the international trade of major weapons, which is given by discrete snapshots of networks that are yearly aggregated over time-continuous trade instances, that is, the time-stamped information is not observed. However, the second application, a network of email traffic, comes in time-stamped format, which can be aggregated to discrete-time observations.

2.2 | Data set 1: International arms trade

The data on international arms trading for the years 2016 and 2017 are provided by the Stockholm International Peace Research Institute (SIPRI, 2019). To be more specific, information on the exchange of major conventional weapons together with the volume of each transfer is included. In order to have a binary network representation, we discretize the data and set edges $Y_{ij,t}$ to “1” if country i sent arms to country j in t .

The left side of Table 1 gives some descriptive measures (Csardi & Nepusz, 2006) and Figure 2 visualizes the arms trade network using the software G_ep_hi (Bastian, Heymann, & Jacomy, 2009). The density of a network is the proportion of realized edges out of all possible edges and is similar in both years, indicating the sparsity of the modeled network. Clustering can be expressed by the transitivity measure, providing the percentage of triangles out of all connected triplets. Reciprocity in a graph is the ratio of reciprocated ties and is similar in both years. As expressed by the high percentage of repeated ties, most countries seem to continue trading with the same partners.

TABLE 1 Descriptive statistics for the international arms trade network (left) and the European research institutions email correspondence (right)

		Arms trade network		Email network	
Time	t	2016	2017	Period 1	Period 2
Number of events		–	–	4,957	2,537
Number of nodes	n	180	180	88	88
Number of possible ties	$n(n - 1)$	32,220	32,220	7,656	7,656
Density		0.021	0.020	0.123	0.087
Transitivity		0.195	0.202	0.407	0.345
Reciprocity		0.081	0.083	0.7	0.687
Repetition		–	0.641	–	0.574

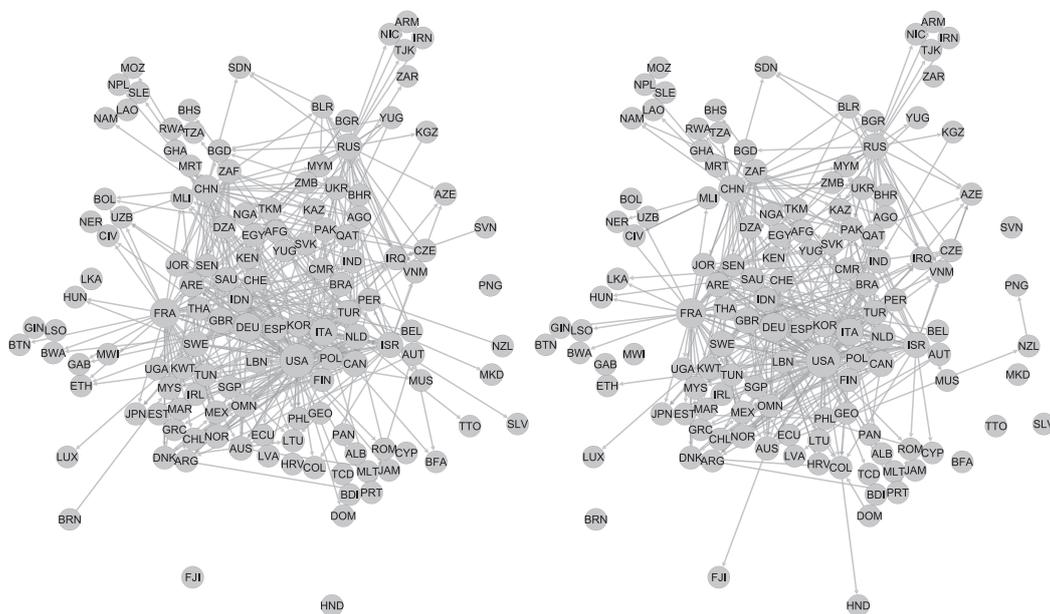


FIGURE 2 The international arms trade as a binary network in 2016 (left) and 2017 (right). Nodes that are isolated in both years are not depicted for clarity and the node size relates to the sum of the out- and in-degrees. The labels of the nodes are the ISO3 codes of the respective countries

Additionally, different kinds of exogenous covariates may be controlled for in statistical network models. In the given example, we use the logarithmic gross domestic product (GDP; World Bank, 2019) as monadic covariates concerning the sender and receiver of weapons. We also include the absolute difference of the so-called polity IV index (Center for Systemic Peace, 2017), ranging from zero (no ideological distance) to 20 (highest ideological distance), as a dyadic exemplary covariate. These covariates are assumed to be nonstochastic and we denote them by x_t . See the Supplementary Material for a list of all included countries and their ISO3 code.

2.3 | Data set 2: European research institution email correspondence

The second network under study represents anonymized email exchange data between institution members of a department in a European research institution (see Email EU Core, 2019; Paranjape,

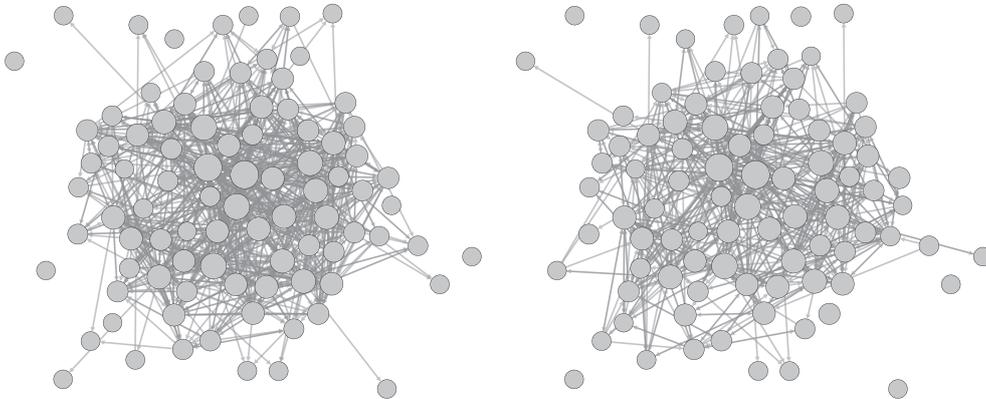


FIGURE 3 The European research institution email correspondence aggregated to a binary network divided into Period 1 (days 1–365, left) and Period 2 (days 1–365, right). The node size relates to the sum of the out- and in-degrees

Benson, & Leskovec, 2017). In this data set, we observe events $\omega = (i, j, \tilde{t})$ that represent emails sent from department member i to department member j at a specific time point \tilde{t} .

The data contain $n = 89$ persons and are recorded over 802 days. For this paper, we select the first two years and split them again into two years, labeled Period 1 and Period 2. Within the first period, 8,068 events are recorded, and in the second period, 4,031 events. We only regard one-to-one email correspondences; therefore, we exclude all group mails from the analysis. In the right column of Table 1, the descriptive measures for the two aggregated networks are given and in Figure 3, they are visualized. All descriptive statistics are higher in the email exchange network as compared to the arms trade network. In comparison to the arms trade network, the aggregated network is more dense with more than 10% of all possible ties being realized. In both years, the transitivity measure is relatively higher in both time periods. The high share of reciprocated ties is intuitive given that the network represents email exchange between institution members that may collaborate. No covariates are available for this network. See the Appendix for the visualization of the degree distributions of both applications.

3 | DYNAMIC EXPONENTIAL RANDOM GRAPH MODELS

3.1 | Temporal ERGM

The ERGM is among the most popular models for the analysis of static network data. Holland and Leinhardt (1981) introduced the model class, which was subsequently extended with respect to fitting algorithms and network statistics (see Lusher, Koskinen, & Robins, 2012; Robins, Pattison, Kalish, & Lusher, 2007). Spurred by the popularity of ERGMs, dynamic extensions of this model class emerged, pioneered by Robins and Pattison (2001) who developed time-discrete models for temporally evolving social networks. Before we start with a description of the model, we want to highlight that the TERGM and the STERGM are most appropriate for equidistant time points. That is, we observe the networks Y_t at discrete and equidistant time points $t = 1, \dots, T$. Only in this setting, the parameters allow for a meaningful interpretation. See Block et al. (2018) for a deeper discussion.

Hanneke et al. (2010) is the main reference for the TERGM, a model class that utilizes the Markov structure and, thereby, assumes that the transition of a network from time point $t - 1$ to time point t can be explained by exogenous covariates as well as structural components of the present and preceding networks. Using a first-order Markov dependence structure and conditioning on the first network, the resulting dependence structure of the model can be factorized into

$$\mathbb{P}_\theta(Y_T, \dots, Y_2 | Y_1, x_1, \dots, x_T) = \mathbb{P}_\theta(Y_T | Y_{T-1}, x_T) \dots \mathbb{P}_\theta(Y_3 | Y_2, x_3) \mathbb{P}_\theta(Y_2 | Y_1, x_2). \quad (1)$$

In the formulation above, the joint distribution is decomposed into yearly transitions from Y_{t-1} to Y_t . Furthermore, it is assumed that the same parameter vector θ governs all transitions. Often, this is an unrealistic assumption for networks observed at many time points because the generative process may change other time. Therefore, it can be useful to allow for different parameter vectors for each transition probability (i.e., $\theta_T, \theta_{T-1}, \dots$). In such a setting, the parameters for each transition can either be estimated sequentially (e.g., Thurner, Schmid, Cranmer, & Kauermann, 2019) or by using smooth time-varying effects (e.g., Lebacher, Thurner, & Kauermann, 2019).

Given the dependence structure (1), the TERGM assumes that the transition from Y_{t-1} to Y_t is generated according to an exponential random graph distribution with the parameter θ :

$$\mathbb{P}_\theta(Y_t = y_t | Y_{t-1} = y_{t-1}, x_t) = \frac{\exp\{\theta^T s(y_t, y_{t-1}, x_t)\}}{\kappa(\theta, y_{t-1}, x_t)}. \quad (2)$$

Generally, $s(y_t, y_{t-1}, x_t)$ specifies a p -dimensional function of sufficient network statistics, which may depend on the present and previous network, as well as on covariates. These network statistics can include static components, designed for cross-sectional dependence structures (see Morris, Handcock, & Hunter, 2008, for more examples). However, the statistics $s(y_t, y_{t-1}, x_t)$ explicitly allow temporal interactions, for example, delayed reciprocity

$$s_{\text{delrecip}}(y_t, y_{t-1}) \propto \sum_{i \neq j} y_{j,t} y_{i,t-1}. \quad (3)$$

This statistic governs the tendency whether a tie (i, j) in $t - 1$ will be reciprocated in t . Another important temporal statistic is stability

$$s_{\text{stability}}(y_t, y_{t-1}) \propto \sum_{i \neq j} (y_{i,j,t} y_{i,j,t-1} + (1 - y_{i,j,t})(1 - y_{i,j,t-1})). \quad (4)$$

In this case, the first product in the sum measures whether existing ties in $t - 1$ persist in t and the second term is one if nonexistent ties in $t - 1$ remain nonexistent in t . The proportionality sign is used because in many cases the network statistics are scaled into a specific interval (e.g., $[0, n]$ or $[0, 1]$). Such a standardization is especially sensible for networks where the actor set changes with time. Additionally, exogenous covariates can be included, for example, time-varying covariates $x_{ij,t}$

$$s_{\text{dyadic}}(y_t, x_t) = \sum_{i \neq j} y_{ij,t} x_{ij,t}. \quad (5)$$

There exists an abundance of possibilities for defining interactions between ties in $t - 1$ and t . From this discussion and Equation (2), it also becomes evident that, in a situation where the interest lies in the transition between two periods, a TERGM can be modeled simply as an ERGM, including lagged network statistics. This can be done for example by incorporating $y_{ij,t-1}$ as explanatory variable in (5), which is mathematically equivalent to the stability statistic (4). In the application we call this statistic repetition (Block et al., 2018).

Concerning the estimation of the model, maximum likelihood estimation appears to be a natural candidate due to the simple exponential family form (2). However, the normalization constant in the denominator of model (2) often poses an inhibiting obstacle when estimating (T)ERGMs. This can be seen by inspecting the normalization constant $\kappa(\theta, y_{t-1}, x_t) = \sum_{\tilde{y} \in \mathcal{Y}} \exp\{\theta^T s(\tilde{y}_t, y_{t-1}, x_t)\}$, which requires summation over all possible networks $\tilde{y} \in \mathcal{Y}$. This task is virtually infeasible, except for very small networks. Therefore, Markov Chain Monte Carlo (MCMC) methods have been proposed in order to approximate the logarithmic likelihood function (see Geyer & Thompson, 1992, for Monte Carlo maximum likelihood, and Hummel, Hunter, & Handcock, 2012, for its adaption to ERGMs). The article by Caimo and Friel (2011) provides an alternative algorithm that uses MCMC-based inference in a Bayesian model framework. Another approach is to employ maximum pseudolikelihood estimation (MPLE, Strauss & Ikeda, 1990) that can be viewed as a local alternative to the likelihood (van Duijn, Gile, & Handcock, 2009) but is often regarded as unreliable and poorly understood in the literature (Handcock, 2003; Hunter, Goodreau, & Handcock, 2008). However, the MPLE is claimed to be consistent and asymptotically efficient (Desmarais & Cranmer, 2012) and the biased standard errors can be corrected via bootstrap (Leifeld, Cranmer, & Desmarais, 2018). A notable special case arises if the network statistics are restricted such that they decompose to

$$s(y_t, y_{t-1}, x_t) = \sum_{i \neq j} y_{ij,t} \tilde{s}_{ij}(y_{t-1}, x_t), \quad (6)$$

with \tilde{s}_{ij} being a function that is evaluated only at the lagged network y_{t-1} and covariates x_t for tie (i, j) . With this restriction, we impose that the ties in t are independent, conditional on the network structures in $t - 1$. This greatly simplifies the estimation procedure and allows to fit the model as a logistic regression model (see, e.g., Almquist & Butts, 2014) without the issues related to the MPLE.

A problem, that is very often encountered when fitting (T)ERGMs with endogenous network statistics is called degeneracy (Schweinberger, 2011) and occurs if most of the probability mass is attributed to network realizations that provide either full or empty networks. One way to circumvent these problems is the inclusion of modified statistics, called geometrically weighted statistics (Snijders, Pattison, Robins, & Handcock, 2006). Using the definitions of Hunter (2007), the geometrically weighted out-degree distribution (GWOD) controls for the out-degree distribution with one statistic, via

$$s_{\text{GWOD}}(y_t) = \exp\{\alpha_O\} \sum_{k=1}^{n-1} (1 - \exp\{-\alpha_O\})^k O_k(y_t), \quad (7)$$

with $O_k(y_t)$ being the number of nodes with out-degree k in t and α_O being the weighting parameter. Correspondingly, the in-degree distribution is captured by the geometrically weighted in-degree distribution (GWID) statistic by exchanging $O_k(y_t)$ with $I_k(y_t)$, which counts the number of nodes with in-degree k , and α_O with α_I . While on the one hand, the weighting often effectively counteracts the problem of degeneracy, the statistics become more complicated to interpret. Negative values of the associated parameter typically indicate a centralized network structure.

Regarding statistics capturing clustering, the most common geometrically weighted triangular structure is called geometrically weighted edge-wise shared partners (GWESP) and builds on the number of two-paths that indirectly connect two nodes i and j given the presence of an edge (i, j) :

$$s_{\text{GWESP}}(y_t) = \exp\{\alpha_S\} \sum_{k=1}^{n-2} (1 - \exp\{-\alpha_S\})^k S_k(y_t), \quad (8)$$

where α_S is a weighting parameter. The number of edges with k shared partners ($S_k(y_t)$) is uniquely defined in undirected networks. If the edges are directed it must be decided which combination should form a triangle; see Lusher et al. (2012) for a discussion. As a default, the number of directed two-paths is chosen (Goodreau, Kitts, & Morris, 2009). Generally, a positive coefficient for GWESP indicates that triadic closure increases the probability of edge occurrence, and globally, a positive value for the associated parameter means more triadic closure as compared to a regime with a negative value (Morris et al., 2008).

3.2 | Separable TERGM

A useful improvement of the TERGM (2) is the STERGM proposed by Krivitsky and Handcock (2014). This model can be motivated by the fact that the stability term leads to an ambiguous interpretation of its corresponding parameter. Given that we include (4) in a TERGM and obtain a positive coefficient after fitting the model, it is not clear whether the network can be regarded as “stable” because existing ties are not dissolved (i.e., $y_{ij,t} = y_{ij,t-1} = 1$) or because no new ties are formed (i.e., $y_{ij,t} = y_{ij,t-1} = 0$). To disentangle this, the authors propose a model that allows for the separation of formation and dissolution.

Krivitsky and Handcock (2014) define the formation network as $Y^+ = Y_t \cup Y_{t-1}$, being the network that consists of the initial network Y_{t-1} together with all ties that are newly added in t . The dissolution network is given by $Y^- = Y_t \cap Y_{t-1}$ and contains exclusively ties that are present in t and $t-1$. Given the network in $t-1$ together with the formation and the dissolution network, we can then uniquely reconstruct the network in t because $Y_t = Y^+ \setminus (Y_{t-1} \setminus Y^-) = Y^- \cup (Y^+ \setminus Y_{t-1})$. Define $\theta = (\theta^+, \theta^-)$ as the joint parameter vector that contains the parameters of the formation and the dissolution model. Building on that, Krivitsky and Handcock (2014) define their model to be separable in the sense that the parameter space of θ is the product of the parameter spaces of θ^+ and θ^- together with conditional independence of formation and dissolution given the network in $t-1$:

$$\mathbb{P}_\theta(Y_t = y_t | Y_{t-1} = y_{t-1}, x_t) = \underbrace{\mathbb{P}_{\theta^+}(Y^+ = y^+ | Y_{t-1} = y_{t-1}, x_t)}_{\text{Formation Model}} \underbrace{\mathbb{P}_{\theta^-}(Y^- = y^- | Y_{t-1} = y_{t-1}, x_t)}_{\text{Dissolution Model}}. \quad (9)$$

The structure of the model is visualized in Figure 4. On the left-hand side, the state of the network Y_{t-1} is given, consisting of two ties (i, h) and (h, j). In the top network all ties that could possibly be formed are shown dashed and the actual formation in this example (i, j) is shown solid. On the bottom, the two ties that could possibly be dissolved are shown, and in this example, (h, j) persists while (i, j) is dissolved. On the right-hand side of Figure 4, the resulting network at time point t is displayed. Given this structure and the separability assumption (9), it is assumed that the formation model is given by

$$\mathbb{P}_{\theta^+}(Y^+ = y^+ | Y_{t-1} = y_{t-1}, x_t) = \frac{\exp\{(\theta^+)^T s(y^+, y_{t-1}, x_t)\}}{\kappa(\theta^+, y_{t-1}, x_t)}, \quad (10)$$

with $\kappa(\theta^+, y_{t-1}, x_t)$ being the normalization constant. Accordingly, the dissolution model can be defined. It becomes apparent how the STERGM is a subclass of the TERGM by inserting the

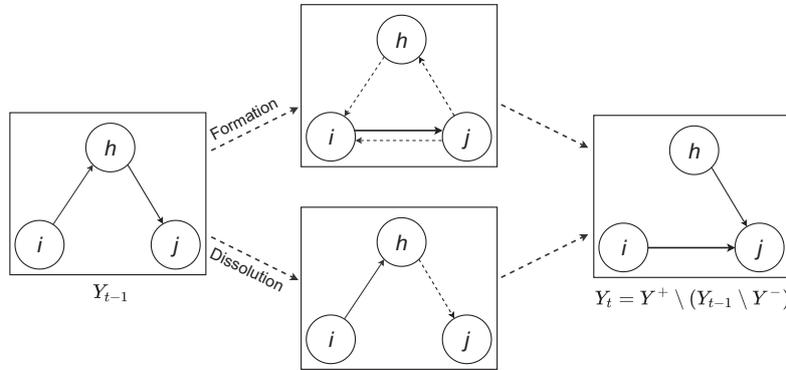


FIGURE 4 Conceptual representation, illustrating formation, and dissolution in the separable temporal exponential random graph model

separable models in (9) because

$$\begin{aligned} \mathbb{P}_\theta(Y_t = y_t | Y_{t-1} = y_{t-1}, x_t) &= \frac{\exp\{(\theta^+)^T s(y^+, y_{t-1}, x_t)\} \exp\{(\theta^-)^T s(y^-, y_{t-1}, x_t)\}}{\kappa(\theta^+, y_{t-1}, x_t) \kappa(\theta^-, y_{t-1}, x_t)} \\ &= \frac{\exp\{\theta^T s(y_t, y_{t-1}, x_t)\}}{\kappa(\theta, y_{t-1}, x_t)} \end{aligned}$$

with $\theta = (\theta^+, \theta^-)^T$, $s(y_t, y_{t-1}, x_t) = (s(y^+, y_{t-1}, x_t), s(y^-, y_{t-1}, x_t))^T$, and the normalization constant set accordingly.

For practical reasons, it is important to understand that the term dissolution model is somewhat misleading because a positive coefficient in the dissolution model implies that nodes (or dyads) with high values for this statistic are less likely to dissolve. This is also the standard implementation in software packages but can simply be changed by switching the signs of the parameters in the dissolution model.

The network statistics are used similarly as in a cross-sectional ERGM. In Krivitsky and Handcock (2014), they are called *implicitly dynamic* because they are evaluated either at the formation network y^+ or the dissolution network y^- , which are both formed from y_{t-1} and y_t . For example, the number of edges is separately computed now for the formation and the dissolution network, giving either the number of edges that newly formed or the number of edges that persisted. For example, reciprocity in the formation network is defined as

$$s_{\text{recip}}(y^+, y_{t-1}) = s_{\text{recip}}(y^+) \propto \sum_{i \neq j} y_{ji}^+ y_{ij}^+, \quad (11)$$

and in case of the dissolution model, y^+ is simply exchanged with y^- . Similarly, edge covariates or the geometrically weighted statistics shown in Equations (5), (7), and (8) are now functions of y^+ or y^- , not y_t .

3.3 | Model assessment

In analogy to binary regressions models, the (S)TERGM can be evaluated in terms of their receiver-operator-characteristic (ROC) curve or precision-recall (PR), where the latter puts more emphasis on finding true positives (e.g., Grau, Grosse, & Keilwagen, 2015). A comparison between

different models is possible using, for example, the Akaike information criterion (AIC, Claeskens & Hjort, 2008). Here, we want to highlight that the AIC fundamentally builds on the log likelihood, which in most realistic applications is only available as an approximation; see Hunter et al. (2008) for further discussion.

However, in statistical network analysis, it is often argued that suitable network models should not exclusively provide good predictions for individual edges but also be able to represent topologies of the observed network. The dominant approach to assess the goodness-of-fit of (S)TERGMs is based on sampling networks from their distribution under the estimated parameters and then comparing network characteristics of these sampled networks with the same ones from the observed network (Hunter et al., 2008). For this approach, it is recommendable to utilize network characteristics that are not used for specifying the model. For instance, models that include the GWOD statistic (7) may not be compared to its simulated values but against the out-degree distribution.

Hanneke et al. (2010) point out that for networks with more than one transition from $t - 1$ to t available, it is possible to employ a “cross-validation-type” assessment of the fit. The parameters can be fit repetitively to all observed transitions except one hold-out transition. It is then checked, how well the network statistics from the hold-out transition period are represented by the ones sampled from the coefficients obtained from all other transitions.

4 | RELATIONAL EVENT MODEL

4.1 | Time-continuous event processes

The second type of dynamic network models results by comprehending network changes as a continuously evolving process (see Girardin & Limnios, 2018 as a basic reference for stochastic processes). The idea was originally introduced by Holland and Leinhardt (1977). In their work, changes in the network are not occurring at discrete time points but as a continuously evolving process, where only one tie can be toggled at a time. This framework was extended by Butts (2008) to model behavior, which is understood as a directed event at a specific time that potentially depends on the past. Correspondingly, the observations in this section are behaviors that are given as triplets $\omega = (i, j, \tilde{t})$ and encode sender i , receiver j , and exact time point \tilde{t} . This fine-grained temporal information is often called time-stamped or time-continuous; we adopt the latter name. Furthermore, we only regard dyadic events in this article, that is, a behavior only includes one sender and receiver.

The concept of behavior, hereinafter called event, generalizes the classical concept of binary relationships based on graph theory as promoted by Wasserman and Faust (1994). This event framework does not intrinsically assume that ties are enduring over a specific time frame (Butts, 2009; Butts & Marcum, 2017). For example in an email exchange network, sending one email at a specific time point is merely a brief event, which does not convey the same information as a durable relationship. Therefore, the time-stamped information cannot adequately be represented in a binary adjacency matrix without having to aggregate the relational data at the cost of information loss (Stadtfeld, 2012). Nevertheless, a friendship between actors i and j at a given time point can still be viewed as an event that has a one-to-one analogy to a tie in the classical framework.

The overall aim of relational event models (REMs, Butts, 2008) is to understand the dynamic structure of events conditional on the history of events (Lerner, Bussmann, Snijders, & Brandes, 2013). This dynamic structure, in turn, controls how past interactions shape the propensity of

future events. To make this model feasible, we leverage results from the field of time-to-event analysis or survival analysis, respectively (see, e.g., Kalbfleisch & Prentice, 2002, for an overview).

The central concept of the REM can be motivated by the introduction of a multivariate time-continuous counting process

$$N(\tilde{t}) = (N_{ij}(\tilde{t})|i, j \in \{1, \dots, n\}), \quad (12)$$

where $N_{ij}(\tilde{t})$ counts how often actors i and j interacted in $[0, \tilde{t}]$. Note that we indicate continuous time \tilde{t} with a tilde to distinguish from the discrete time setting with $t = 1, 2, \dots, T$ assumed in the previous section. Process (12) is characterized by an intensity function $\lambda_{ij}(\tilde{t})$ for $i \neq j$, which is defined as

$$\lambda_{ij}(\tilde{t}) = \lim_{dt \downarrow 0} \frac{\mathbb{P}(N_{ij}(\tilde{t} + dt) = N_{ij}(\tilde{t}) + 1)}{dt}.$$

This is the instantaneous probability of observing a jump of size “1” in $N_{ij}(\tilde{t})$, which indicates observing the event (i, j, \tilde{t}) . Because we assume that there are no self-loops $\lambda_{ii}(\tilde{t}) \equiv 0 \forall i = 1, \dots, n$ holds.

4.2 | Time-continuous observations

Butts (2008) introduced the REM to analyze the intensity $\lambda_{ij}(\tilde{t})$ of process (12) when time-continuous data on the events are available. He assumed that the intensity is constant over time but depends on time-varying relational information of past events and exogenous covariates. Vu, Hunter, Smyth, and Asuncion (2011) extended the model by postulating a semiparametric intensity similar to Cox (1972):

$$\lambda_{ij}(\tilde{t}|N(\tilde{t}), x(\tilde{t}), \theta) = \lambda_0(\tilde{t}) \exp\{\theta^T s_{ij}(N(\tilde{t}), x(\tilde{t}))\}, \quad (13)$$

where $\lambda_0(\tilde{t})$ is an arbitrary baseline intensity, $\theta \in \mathbb{R}^p$ is the parameter vector, and $s_{ij}(N(\tilde{t}), x(\tilde{t}))$ is a statistic that depends on the (possibly time-continuous) covariate process $x(\tilde{t})$ and the counting process just prior to \tilde{t} .

Generally, similar statistics as already introduced in Section 3 can be included in $s_{ij}(N(\tilde{t}), x(\tilde{t}))$. Solely, the differing level of the model needs to be accounted for because model (13) takes a local time-continuous point of view to understand the relational nature of the observed events. This necessitates defining the statistics $s_{ij}(N(\tilde{t}), x(\tilde{t}))$ from the position of specific ties, in contrast to the globally defined statistics $s(y_t, y_{t-1}, x_t)$ in (2). To give an example, the tie-level version of reciprocity for the event (i, j) is defined as

$$s_{ij, \text{reciprocity}}(N(\tilde{t}), x(\tilde{t})) = \mathbb{1}(N_{ji}(\tilde{t}) > 0),$$

where $\mathbb{1}(\cdot)$ is the indicator function. It only regards, whether already having observed the event (j, i) prior to \tilde{t} has an effect on $\lambda_{ij}(\tilde{t}|N(\tilde{t}), x(\tilde{t}), \theta)$, in comparison to the network level version (3) of delayed reciprocity that counted all reciprocated ties between the networks y_t and y_{t-1} .

Degree statistics can be specified as either sender- or receiver-specific. If we, for example, want to control for the out-degree of the sender, the corresponding tie-oriented statistic is

$$s_{ij, \text{SOD}}(N(\tilde{t}), x(\tilde{t})) = \sum_{h=1}^n \mathbb{1}(N_{ih}(\tilde{t}) > 0).$$

The in-degree of the receiver can be formulated accordingly.

Clustering in event sequences may be captured by different types of nested two-path configurations. For instance, the tie-oriented version of directed two-paths, henceforth called transitivity, is given by

$$s_{ij,TRA}(N(\tilde{t}), x(\tilde{t})) = \sum_{h=1}^n \mathbb{1}(N_{ih}(\tilde{t}) > 0) \mathbb{1}(N_{hj}(\tilde{t}) > 0).$$

The inclusion of monadic and dyadic exogenous covariates becomes straightforward by setting $s_{ij,dyadic}(N(\tilde{t}), x(\tilde{t}))$ equal to the covariate values of interest. Because the effect of a past event at time δ , say, on a present event at time \tilde{t} may vary according to the elapsed time $\tilde{t} - \delta$, Stadtfeld and Block (2017) introduced windowed effects, which only regard events that occurred in a prespecified time window, for example, a year. We will come back to this point in the next section.

If time-continuous observations are available, each dimension of the observed counting process is conditionally independent given the past. This, in turn, enables the construction of a likelihood, which can subsequently be maximized. Assuming that Ω is the set of all observed events and \mathcal{T} is the interval of observation, the likelihood can be written as:

$$\mathcal{L}(\theta) = \prod_{(i,j,\tilde{t}) \in \Omega} \lambda_{ij}(\tilde{t}|N(\tilde{t}), x(\tilde{t}), \theta) \exp \left\{ - \int_{\mathcal{T}} \sum_{k,h=1}^n \lambda_{kh}(u|N(u), x(u), \theta) du \right\}. \quad (14)$$

It is straightforward to maximize the likelihood in the case of a parametric baseline intensity $\lambda_0(t)$, for example, Butts (2008) assumes $\lambda_0(t) = \gamma_0$. Alternatively, Butts (2008) analyzed events with ordinal temporal information. In this setting, the likelihood is equal to the partial likelihood introduced by Cox (1972) for estimating parameters of semiparametric intensities as in (13). Letting U_t denote the set of all possible events that could have occurred at time point t but did not, the partial likelihood for continuous event data is defined as

$$\mathcal{PL}_{\text{cont}}(\theta) = \prod_{(i,j,\tilde{t}) \in \Omega} \frac{\lambda_{ij}(\tilde{t}|N(\tilde{t}), x(\tilde{t}), \theta)}{\sum_{(k,h) \in U_{\tilde{t}}} \lambda_{kh}(\tilde{t}|N(\tilde{t}), x(\tilde{t}), \theta)}. \quad (15)$$

Consecutively, $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ can be estimated with a Nelson Aalen estimator (see Kalbfleisch & Prentice, 2002, for further details on the estimation).

When dealing with large amounts of event data, the main obstacle is evaluating the sum over the intensities of all possible ties in (15) (Butts, 2008). One exact option is to trade a longer running time for a slimmer memory footprint by means of a coaching data structure. Vu et al. (2011) exploit this by saving prior values of the sum and subsequently changing it event-wise by elements of $U_{\tilde{t}}$ whose covariates changed. Alternatively, Vu et al. (2015) proposes approximate routines that utilize case-control sampling and stratification for the Cox model (Langholz & Borgan, 1995). More precisely, the sum is only calculated over a sampled subset of possible events in addition to stratification. Lerner and Lomi (2019) go one step further and sample events out of Ω for the calculation of $\mathcal{PL}(\theta)$ in (15).

Numerous extensions of this model that build on already well-established methods in social networks and time-to event analysis have been proposed. Perry and Wolfe (2013) used a stratified Cox model in (13). Stadtfeld, Hollway, and Block (2017) adopted the SAOM to events. DuBois and Smyth (2010) and DuBois, Butts, and Smyth (2013) extended the stochastic block model for time-stamped relational events. Furthermore, DuBois, Butts, McFarland, and Smyth (2013) adopted a Bayesian hierarchical model to event data when information is only available in smaller groups.

4.3 | Time-clustered observations

Generally, the approach discussed above requires time-continuous network data, meaning that we observe the precise time points of all events. To give an instance, in the first data example, this means that we need the exact time point \tilde{t} of an arms trade between countries i and j . Often, such exact time-stamped data are not available and, in fact, trading between states can hardly be stamped with a single time point \tilde{t} . Indeed, we often only observe the time-continuous network process at discrete time points $t = 1, \dots, T$. In such setting, we may assume a Markov structure in that we do not look at the entire history of the process $N(\tilde{t})$ but just condition the intensity (13) on the history of events from the previous observation $t - 1$ to \tilde{t} . Technically, this means that $N(t)$ is adapted to $\tilde{Y}(\tilde{t}) := N(\tilde{t}) - N(t - 1)$ and $x(\tilde{t})$ for $\tilde{t} \in [t - 1, t]$. We then reframe (13) as

$$\lambda_{ij}(\tilde{t} | \tilde{Y}(\tilde{t}), x(\tilde{t}), \theta) = \lambda_0(\tilde{t}) \exp\{\theta^T s_{ij}(\tilde{Y}(\tilde{t}), x(\tilde{t}))\}. \quad (16)$$

In other words, we assume that the intensity of events between $t - 1$ and t does not depend on states of the multivariate counting process (12) prior to $t - 1$. For this reason, all endogenous statistics introduced in Section 4.2 are now evaluated on $\tilde{Y}(\tilde{t})$ instead of $N(\tilde{t})$. This is a reasonable assumption, if one is primarily interested in short-term dependencies between the individual counting processes. It enables a meaningful comparison to the models from Section 3 that assume an analog discrete Markov property. However, we want to emphasize that this dependence structure is not vital to inferential results.

If we observe the continuous process at discrete time points, it is inevitable that we observe time-clustered observations, meaning that two or more events happen at the same time point. Under the term tied observations, this phenomenon is well known in time-to-event analysis and treated with several approximations. One option is the so-called Breslow approximation (see Breslow, 1974; Peto, 1972). Let therefore

$$O_t = \{(i, j) \mid N_{ij}(t) - N_{ij}(t - 1) > 0\},$$

where element (i, j) is replicated $N_{ij}(t) - N_{ij}(t - 1)$ times in O_t , that is, if an event between i and j occurred multiple times in the interval from $t - 1$ to t , then (i, j) appears respective times in O_t . Given that we have not observed the exact time point of an event, we also get no information on the baseline intensity $\lambda_0(\tilde{t})$ in (13) for $\tilde{t} \in [t - 1, t]$ so that the model simplifies to a discrete choice model structure (see, e.g., Train, 2009), which resembles the partial likelihood (15) and is defined as

$$\mathcal{P}\mathcal{L}_{\text{clust}}(\theta) = \prod_{t=1}^T \frac{\prod_{(i,j) \in O_t} \exp\{\theta^T s_{ij}(\tilde{Y}(t), x(t))\}}{\left(\sum_{(k,h) \in U_t} \exp\{\theta^T s_{kh}(\tilde{Y}(t), x(t))\}\right)^{n_t}}, \quad (17)$$

where $n_t = |O_t|$. Alternatively, one can replace the denominator in (17) by considering all possible orders of the unobserved events in O_t giving the average likelihood as introduced by Kalbfleisch and Prentice (2002). Because this can be a combinatorial and, hence, numerical challenge, random sampling of time point orders among the time-clustered observations can be used with subsequent averaging, which we call Kalbfleisch–Prentice approximation (see Kalbfleisch & Prentice, 2002). Further techniques to deal with unknown time ordering are augmenting the clustered events into possible paths of ordered events and adapting the maximum likelihood estimation proposed for the SAOM by Snijders, Koskinen, and Schweinberger (2010) or using random

sampling of the ordering. This can be legitimized in cases where we may assume independence among events happening in one year because the events take a long time to materialize (Snijders, 2017).

4.4 | Model assessment

In comparison to the assessment for models operating in discrete time, widely accepted methods dealing with relational event data are scarce. The proposals either stem from time-to-event analysis or regard link prediction, which is the task of predicting the most likely next event given the history of past events (Liben-Nowell & Kleinberg, 2007). One example of the former option is the usage of *Schönfeld residuals* by Vu, Asuncion, Hunter, and Smyth (2011) to check the assumption of proportional intensities, which is central to semiparametric models as the one proposed by Cox (1972). For the latter approach, we need to define a predictive measure that quantifies how well the next event is predicted. Vu et al. (2011) proposed the recall measure that estimates the percentage of test events, which are in the list of K most likely next events according to a given model. Evaluating this percentage for different values of K permits a visualization of the predictive capabilities of the model. The strength of the predicted intensity allows the ordering of events according to the probability of being observed next. If we model the propensity of time-clustered events that represent binary adjacency matrices, one can alternatively adopt the analysis of the ROC and PR curve introduced in Section 3.3.

5 | APPLICATION

When it comes to software, there exist essentially three main R packages that are designed for fitting TERGMs and STERGMs. Most important is the extensive `statnet` library (Goodreau, Handcock, Hunter, Butts, & Morris, 2008) that allows for simulation-based fitting of ERGMs. The library contains the package `tergm` with implemented methods for fitting STERGMs using MCMC approximations of the likelihood. However, currently the package `tergm` (version 3.5.2) does not allow for fitting STERGMs with time-varying dyadic covariates for more than two time periods jointly. The package `btergm` (Leifeld et al., 2018) is designed for fitting TERGMs using either maximum pseudolikelihood or MCMC maximum likelihood estimation routines. In order to obtain Bayesian Inference in ERGMs, the package `bergm` by Caimo and Friel (2014) can be used. Besides implementations in R, the stand-alone program `PNet` (Wang, Robins, & Pattison, 2006) allows for simulating, fitting, and evaluating (T)ERGMs. In order to ensure comparable estimates, we estimate the TERGM, as well as the STERGM, with the `statnet` library, using MCMC-based likelihood estimation techniques. We use the package `ergm` and include delayed reciprocity and the repetition of previous ties as dyadic covariates. The STERGM is fitted using the `tergm` package.

Marcum and Butts (2015) implemented the R package `relevent` (version 1.0-4) to estimate the REM for time-stamped data. It was followed by the package `goldfish` (version 1.2) by Stadtfeld and Hollway (2018) for modeling event data with precise and ordinal temporal information with an actor- and tie-oriented variant of the REM. Furthermore, it is highly customizable in terms of endogenous and exogenous user terms and will be used in the following applications.

We want to remark that the STERGM coefficients are implicitly dynamic, whereas in the TERGM, all network statistics except the lagged network and delayed reciprocity terms are evaluated on the network in t . All covariates of the REM are continuously updated and the intensity

at time point $\tilde{t} \in [t - 1, t]$ only depends on events observed in $[t - 1, \tilde{t})$. Like the building period proposed by Vu et al. (2011), the events in $t - 1$ are only used for building up the covariates and not directly modeled. Moreover, the coefficients of the REM affect the intensity of a specific event on the tie level. For the (S)TERGM, the estimates can be interpreted as the effect of a global statistic $s(y_t, y_{t-1}, x_t)$ on the probability of observing the network y_t . They also allow for a tie-level interpretation based on the so-called change statistics (Cranmer & Desmarais, 2011). Due to no compositional changes, we did not scale any statistics.

5.1 | Data set 1: International arms trade

The results obtained for the arms trading data section are displayed in Table 2. For a detailed interpretation of effects focusing on political, social, and economic aspects, we refer to the relevant literature (e.g., Thurner et al., 2019). Here, we want to comment on a few aspects only. While we do not have time stamps for the arms trades, the longitudinal networks can still be viewed as time-clustered observations enabling the techniques from Section 4.3.

TABLE 2 Arms trade network: Comparison of parameters obtained from the TERGM (first column), STERGM (formation in the second column, dissolution in the third column), and REM (fourth column)

	TERGM	STERGM		REM	
		Formation	Dissolution		
Repetition	3.671*** (0.132)	–	–	2.661*** (0.143)	
Edges	–15.632*** (1.809)	–17.186*** (2.168)	–16.987*** (3.587)	–	
Reciprocity	–0.258 (0.306)	–0.620 (0.436)	–0.058 (0.619)	–0.109 (0.181)	
In-degree (GWID)	–1.823*** (0.278)	–2.106*** (0.379)	–0.412 (0.442)	0.060** (0.015)	In-degree receiver
Out-degree (GWOD)	–3.220*** (0.304)	–4.126*** (0.462)	–0.326 (0.533)	0.010** (0.004)	Out-degree sender
GWESP	0.050 (0.066)	0.076 (0.071)	0.150 (0.126)	0.010 (0.029)	Transitivity
Polity score (absolute difference)	–0.024* (0.010)	–0.028* (0.014)	–0.016 (0.017)	–0.016 (0.009)	
log(GDP) sender	0.313*** (0.048)	0.394*** (0.054)	0.323*** (0.088)	0.395*** (0.039)	
log(GDP) receiver	0.165*** (0.043)	0.135* (0.054)	0.327*** (0.087)	0.192*** (0.032)	
Log likelihood	–949.833	–675.327	–258.425		
AIC	1,917.666	1,366.654	532.849		
\sum AIC	1,917.666	1,899.503			

Note. Standard errors in brackets and stars according to p values smaller than 0.001 (***), 0.05 (**), and 0.1 (*). The decay parameter of the geometrically weighted statistics is set to $\log(2)$ and the Kalbfleisch–Prentice approximation was used with 100 random orderings of the events to find the estimates of the REM. TERGM = temporal exponential random graph model; STERGM = separable temporal exponential random graph model; REM = relational event model; GWID = geometrically weighted in-degree distribution; GWOD = geometrically weighted out-degree distribution; GWESP = geometrically weighted edge-wise shared partners; GDP = gross domestic product; AIC = Akaike information criterion.

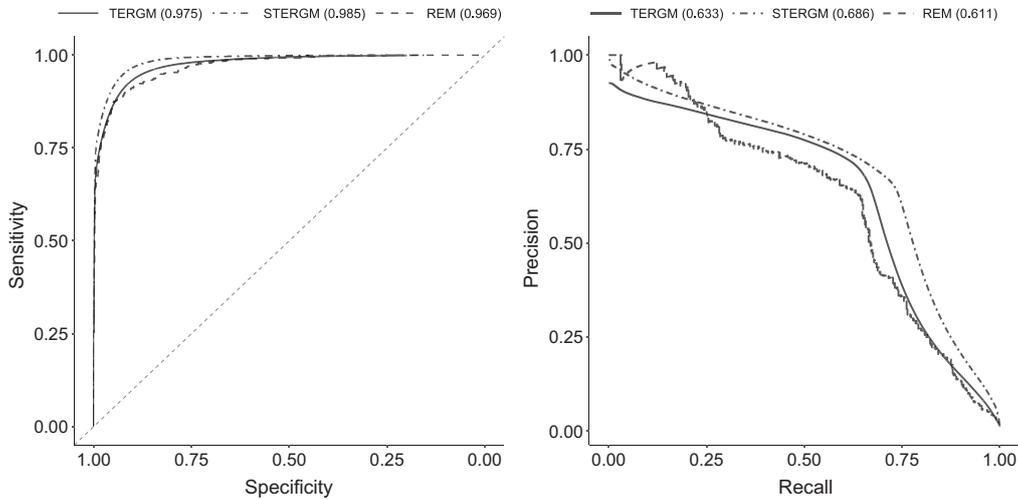


FIGURE 5 Arms trade network: ROC (left) and PR (right) curves from the TERGM (dotted line), STERGM (dot-dashed line), and REM (solid line). The AIC values of the respective curves are indicated in brackets. ROC = receiver-operator-characteristic; PR = precision-recall; TERGM = temporal exponential random graph model; STERGM = separable temporal exponential random graph model; REM = relational event model; AIC = Akaike information criterion

Both the TERGM (column 1) and the REM (column 4) identify the repetition of previous ties as a driving force in the dynamic structure of the network. Degree-related covariates, which are GWID and GWOD in the (S)TERGM and the in- and out-degrees in the REM, capture centrality in the network. The coefficients of the GWID and GWOD are negative and have low p values in the TERGM. This stands in contrast to the STERGM, where these effects are only pronounced in the formation model (column 2); however, they are insignificant effects in the dissolution model (column 3). Hence, these effects suggest a centralized pattern in the formation network, which is also captured by the TERGM. In the REM, an analogous pattern can be detected because a higher in-degree of the receiver increases the respective intensity; thus, spurs trade relations. Similar interpretations hold for the out-degree of the sender. Overall, countries that have a high out-degree are more likely to send weapons and countries with a high in-degree to receive them, which again results in a centralized network structure as indicated by the estimates in the TERGM and STERGM. This example illustrates how seemingly contrasting coefficients of the (S)TERGM and REM can still imply a similar interpretation in terms of the implied global network characteristics.

Lastly, consistent effects among the models were also found for the exogenous covariates. Consider, for instance, the coefficient of the logarithmic GDP of the importing country. The TERGM assigns a significantly higher probability to observe in-going ties to countries with a high GDP just like the REM. However, disentangling the model towards formation and dissolution, we see strongly significant coefficients in the dissolution model, whereas the effect for the formation model is weakly significant.

Based on the independence assumption in (9), we can sum up the two AIC values and see that the AIC value of the STERGM is smaller than of the TERGM. Furthermore, the ROC and PR curves of all three models are shown in Figure 5. Again, both measures indicate that the STERGM provides a slightly improved fit when compared to the TERGM and REM. The results

of the simulation-based model assessment of the (S)TERGM can be found in the Supplementary Material.

5.2 | Data set 2: European research institution email correspondence

As already indicated by the descriptive statistics in Table 1, the email network seems to be driven by three major structural influences: repetition, reciprocity, and transitive clustering. The estimates from Table 3 demonstrate that all models were able to identify these forces.

According to the REM (column 4), the event network of email traffic in the research institution is not centralized and primarily based on collaboration between coworkers. We can draw those conclusions from insignificant estimates of degree-related statistics and highly significant estimates regarding reciprocity and repetition. In the TERGM (column 1), we find a positive and significant effect of GWID, whereas no effect can be found in the STERGM (columns 2 and 3). The estimates of repetition and reciprocity in the REM and TERGM are very pronounced. For instance, the estimates of the REM imply that a reciprocated event is 19.6 times more likely than an event with the same covariates only not being reciprocated. Interestingly, the STERGM detects a lower effect of GWESP in the formation and dissolution than the TERGM. The effect of the delayed reciprocity in the TERGM is less relevant than reciprocity in the formation and dissolution model. This strongly differing effect size results from the mathematical formulation of the statistics given in Equations (3) and (11).

Contrasting the AIC values of the TERGM and STERGM shows that the dynamic structure of the email network is again better explained by the STERGM. In the Supplementary Material, we

TABLE 3 Email exchange network: Comparison of parameters obtained from the TERGM (first column), STERGM (formation in the second column, dissolution in the third column), and REM (fourth column)

	TERGM	STERGM		REM	
		Formation	Dissolution		
Repetition	1.367*** (0.107)	–	–	2.27*** (0.084)	
Edges	–5.755*** (0.237)	–4.853*** (0.247)	–2.237*** (0.224)	–	
Reciprocity	0.398*** (0.112)	2.498*** (0.157)	2.586*** (0.226)	1.655*** (0.075)	
In-degree (GWID)	1.060** (0.333)	1.349* (0.648)	0.709 (0.415)	–0.004 (0.003)	In-degree receiver
Out-degree (GWOD)	0.031 (0.312)	–0.411 (0.431)	–0.369 (0.397)	–0.0001 (0.003)	Out-degree sender
GWESP	1.560*** (0.110)	0.655*** (0.111)	0.429*** (0.086)	0.070*** (0.008)	Transitivity
Log likelihood	–1,723.732	–1,000.506	–505.431		
AIC	3,459.464	2,011.012	1,020.862		
\sum AIC	3,459.464	3,031.874			

Note. Standard errors in brackets and stars according to p values smaller than 0.001 (***), 0.05 (**), and 0.1 (*). Decay parameter of the geometrically weighted statistics is set to $\log(2)$. TERGM = temporal exponential random graph model; STERGM = separable temporal exponential random graph model; REM = relational event model; GWID = geometrically weighted in-degree distribution; GWOD = geometrically weighted out-degree distribution; GWESP = geometrically weighted edge-wise shared partners; GDP = gross domestic product; AIC = Akaike information criterion.

give the results of model assessment for the (S)TERGM and REM, as well as an application of the TERGM and STERGM to multiple time points.

6 | CONCLUSION

6.1 | Further models

Snijders (1996) formulated a two-stage process model operating in a continuous-time framework. The dynamics are considered to evolve according to unobserved microsteps. At first, a sender out of all eligible actors gets the opportunity to change the state of all his outgoing ties. Consecutively, the actor needs to evaluate the probability of changing the present configuration with each possible receiver, which entails each actor's knowledge of the complete graph whenever he has the possibility to toggle one of his ties. Lastly, the decision is randomly drawn relative to the probabilities of all possible actions. In general, the SAOM is a well-established model for the analysis of social networks, which was successfully applied to a wide array of network data, for example, in sociology (Agneessens & Wittek, 2012; de Nooy, 2002), political science (Bichler & Franquez, 2014; Kinne, 2016), economics (Castro, Casanueva, & Galán, 2014), and psychology (Jason, Light, Stevens, & Beers, 2014). Estimation of this model variant is predominantly carried out with the R package *RSiena* (Ripley, Boitmanis, & Snijders, 2013).

Another notable model that can be regarded as a bridge between the ERGM and continuous-time models is the longitudinal ERGM (LERGM; Koskinen, Caimo, & Lomi, 2015; Snijders & Koskinen, 2013). In contrast to the TERGM, the LERGM assumes that the network evolves in microsteps as a continuous-time Markov process with an ERGM being its limiting distribution. Similar to the SAOM, the model builds on randomly assigning the opportunity to change, followed by a function that governs the probability of a tie change. This model is still tie-oriented, meaning that dyadic ties instead of actors are chosen and then have the option to change the current network.

6.2 | Summary

In this article, we emphasize tie-oriented dynamic network models. Comparisons between these models can be drawn on the level at which each implied generating mechanism works and how time is perceived. The overall aim in the TERGM is to find an adequate distribution of the adjacency matrix Y_t conditioning on information of previous realizations of the network. In the separable extension, the aim remains unchanged, only splitting Y_t into two subnetworks that include all possible ties that were and were not present in Y_{t-1} separately. While the (S)TERGM proceeds in discrete time, the REM tackles modeling the intensity on the tie level in continuous time conditional on past events. Therefore, the (S)TERGM takes a global and REM a local point of view. Even though this results in substantially different interpretations of the estimates, they can still be related to one another by focusing on how the global effects of the (S)TERGM can be emergent from tie-level effects treated in the REM. Furthermore, our discussion exposed how the model assessment of the REM focuses on analogies with time-to-event analysis, that is, looking at the adequate behavior of the model on the tie-level, and the (S)TERGM on simulation-based procedures, which regards the capabilities of the model to capture global characteristics.

Finally, the typically required data structure differs between the model classes. While the time-discrete generating mechanism of the (S)TERGM naturally processes network data that were

observed at discrete time points, the REM is based on time-continuous data. Of the two analyzed data sets, the international arms trade network represents the former data structure, whereas the email traffic data are an example of the latter. By extending the REM to time-clustered observations and aggregating events to binary adjacency matrices, a meaningful comparison between the STERGM, TERGM, and REM is enabled.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their careful reading and constructive comments. The project was supported by the European Cooperation in Science and Technology (COST Action CA15109 [COSTNET]). We also gratefully acknowledge funding provided by the German Research Foundation (DFG) for project KA 1188/10-1: *International Trade of Arms: A Network Approach*. Furthermore, we like to thank the Munich Center for Machine Learning (MCML) for the funding.

ORCID

Cornelius Fritz  <https://orcid.org/0000-0002-7781-223X>

REFERENCES

- Agneessens, F., & Wittek, R. (2012). Where do intra-organizational advice relations come from? The role of informal status and social capital in social exchange. *Social Networks*, 34(3), 333–345.
- Almquist, Z. W., & Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, 44(1), 273–321.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. <http://www.aiai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Bearman, P., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1), 44–91.
- Benton, R. A., & You, J. (2017). Endogenous dynamics in contentious fields: Evidence from the shareholder activism network, 2006–2013. *Socius: Sociological Research for a Dynamic World*, 3.
- Bichler, G., & Franquez, J. (2014). *Conflict cessation and the emergence of weapons supermarkets* (pp. 189–215). Cham, Switzerland: Springer International Publishing.
- Blank, M., Dincecco, M., & Zhukov, Y. M. (2017). Political regime type and warfare: Evidence from 600 years of european history. Available at SSRN: <https://dx.doi.org/10.2139/ssrn.2830066>.
- Block, P., Koskinen, J., Hollway, J., Steglich, C., & Stadtfeld, C. (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52, 180–191.
- Block, P., Stadtfeld, C., & Snijders, T. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research*, 48(1), 202–239.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1), 89–99.
- Broekel, T., & Bednarz, M. (2018). Disentangling link formation and dissolution in spatial networks: An application of a two-mode STERGM to a project-based R&D network in the German biotechnology industry. *Networks and Spatial Economics*, 18(3), 677–704.
- Butts, C. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1), 155–200.
- Butts, C. (2009). Revisiting the foundations of network analysis. *Science*, 325, 414–416.
- Butts, C., & Marcum, C. (2017). A relational event approach to modeling behavioral dynamics. In *Group processes data-driven computational approaches*. Cham, Switzerland: Springer.
- Caimo, A., & Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1), 41–55.

- Caimo, A., & Friel, N. (2014). Bergm: Bayesian exponential random graphs in R. *Journal of Statistical Software*, 61(2), 1–25.
- Castro, I., Casanueva, C., & Galán, J. L. (2014). Dynamic evolution of alliance portfolios. *European Management Journal*, 32(3), 423–433.
- Center for Systemic Peace (2017). Polity IV Annual Time-Series, 1800–2015. Version 3.1. <http://www.systemicpeace.org>
- Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge, UK: Cambridge University Press.
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 34(2), 187–220.
- Cranmer, S., & Desmarais, B. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1), 66–86.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
- Desmarais, B. A., & Cranmer, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1865–1876.
- DuBois, C., Butts, C., McFarland, D., & Smyth, P. (2013). Hierarchical models for relational event sequences. *Journal of Mathematical Psychology*, 57(6), 297–309.
- DuBois, C., Butts, C., & Smyth, P. (2013). Stochastic blockmodeling of relational event dynamics. In C. M. Carvalho & P. Ravikumar (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics* (pp. 238–246).
- DuBois, C., & Smyth, P. (2010). Modeling relational events via latent classes. *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 803–812.
- van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31(1), 52–62.
- Email EU Core. (2019). Email EU core temporal. Department 3. <https://snap.stanford.edu/data/email-Eu-core-temporal-Dept3.txt.gz>
- Erdős, P., & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6, 290.
- Frank, O., & Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association*, 81(395), 832–842.
- Geyer, C. J., & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 54(3), 657–699.
- Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4), 1141–1144.
- Girardin, V., & Limnios, N. (2018). *Applied probability* (2nd ed.). Heidelberg, Germany: Springer.
- Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–233.
- Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., & Morris, M. (2008). A statnet tutorial. *Journal of Statistical Software*, 24(9), 1.
- Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1), 103–125.
- Grau, J., Grosse, I., & Keilwagen, J. (2015). PRROC: Computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics*, 31(15), 2595–2597.
- Handcock, M. S. (2003). *Assessing degeneracy in statistical models of social networks* (Working Paper No. 39).
- Hanneke, S., Fu, W., Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4, 585–605.
- He, X., Dong, Y.-B., Wu, Y.-Y., Jiang, G.-R., & Zheng, Y. (2019). Factors affecting evolution of the interprovincial technology patent trade networks in China based on exponential random graph models. *Physica A: Statistical Mechanics and Its Applications*, 514, 443–457.
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.
- Holland, P., & Leinhardt, S. (1977). A dynamic model for social networks. *The Journal of Mathematical Sociology*, 5(1), 5–20.

- Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33–50.
- Hummel, R. M., Hunter, D. R., & Handcock, M. S. (2012). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4), 920–939.
- Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, 29(2), 216–230.
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248–258.
- Jason, L. A., Light, J. M., Stevens, E. B., & Beers, K. (2014). Dynamic social networks in recovery homes. *American Journal of Community Psychology*, 53(3-4), 324–334.
- Kalbfleisch, J., & Prentice, R. (2002). *The statistical analysis of failure time data*. Hoboken, NJ: Wiley-Blackwell.
- Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12, 105–135.
- Kinne, B. J. (2016). Agreeing to arm: Bilateral weapons agreements and the global arms trade. *Journal of Peace Research*, 53(3), 359–377.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer.
- Koskinen, J., Caimo, A., & Lomi, A. (2015). Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to Foreign Direct Investments. *Network Science*, 3(1), 58–77.
- Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 29–46.
- Langholz, B., & Borgan, Ø. (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika*, 82(1), 69–79.
- Lebacher, M., Thurner, P. W., & Kauermann, G. (2019). A dynamic separable network model with actor heterogeneity: An application to global weapons transfers. arXiv e-prints arXiv:1803.02707v4.
- Leifeld, P., Cranmer, S. J., & Desmarais, B. A. (2018). Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. *Journal of Statistical Software*, 83(6).
- Lerner, J., Bussmann, M., Snijders, T., & Brandes, U. (2013). Modeling frequency and type of interaction in event networks. *Corvinus Journal of Sociology and Social Policy*, 4(1), 3–32.
- Lerner, J., & Lomi, A. (2019). Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events. arXiv preprint arXiv:1905.00630.
- Leskovec, J., Lang, K. J., Dasgupta, A., & Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1), 29–123.
- Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology*, 58(7), 1019–1031.
- Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge, UK: Cambridge University Press.
- Marcum, C., & Butts, C. (2015). Constructing and modifying sequence statistics for relevant Using informR in R. *Journal of Statistical Software*, 64(5), 1–36.
- Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software*, 24(4), 1548.
- de Nooy, W. (2002). The dynamics of artistic prestige. *Poetics*, 30(3), 147–167.
- Paranjape, A., Benson, A. R., & Leskovec, J. (2017). Motifs in temporal networks. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 601–610.
- Perry, P., & Wolfe, P. (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5), 821–849.
- Peto, R. (1972). Contribution to the discussion of the paper by Dr. Cox. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 34(2), 187–220.
- Quintane, E., Pattison, P., Robins, G., & Mol, J. (2013). Short- and long-term stability in organizational networks: Temporal structures of project teams. *Social Networks*, 35(4), 528–540.
- Raabe, I. J., Boda, Z., & Stadtfeld, C. (2019). The social pipeline: How friend influence and peer exposure widen the STEM gender gap. *Sociology of Education*, 92(2), 105–123.
- Ripley, R., Boitmanis, K., & Snijders, T. (2013). RSiena: Siena–simulation investigation for empirical network analysis. R package version 1.2-12. <https://CRAN.R-project.org/package=RSiena>

- Robins, G., & Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1), 5–41.
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29(2), 173–191.
- Salathé, M., Vu, D., Khandelwal, S., & Hunter, D. (2013). The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science*, 2(1), 4.
- Sarkar, P., & Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7(2), 31–40.
- Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496), 1361–1370.
- SIPRI (2019). SIPRI arms transfers database. <https://www.sipri.org/databases/armstransfers>
- Snijders, T. (1996). Stochastic actor-oriented models for network change. *The Journal of Mathematical Sociology*, 21(1–2), 149–172.
- Snijders, T. (2002). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1), 361–395.
- Snijders, T. (2005). Models for longitudinal network data. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). Cambridge, UK: Cambridge University Press.
- Snijders, T. (2017). Comment: Modeling of coordination, rate functions, and missing ordering information. *Sociological Methodology*, 47(1), 41–47. Available at <https://doi.org/10.1177/0081175017723346>
- Snijders, T., & Koskinen, J. (2013). *Longitudinal models*. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks* (pp. 130–140). Cambridge, UK: Cambridge University Press.
- Snijders, T., Koskinen, J., & Schweinberger, M. (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2), 567.
- Snijders, T., Pattison, P., Robins, G., & Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99–153.
- Stadtfeld, C. (2012). Events in social networks: a stochastic actor-oriented framework for dynamic event processes in social networks. Karlsruhe, Germany: KIT Scientific Publishing.
- Stadtfeld, C., & Block, P. (2017). Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science*, 4(14), 318–352.
- Stadtfeld, C., & Hollway, J. (2018). goldfish: Goldfish – Statistical network models for dynamic network data. Retrieved from R package version 1.2. www.social-networks.ethz.ch/research/goldfish.html
- Stadtfeld, C., Hollway, J., & Block, P. (2017). Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology*, 47(1), 1–40.
- Stansfield, S. E., Mittler, J. E., Gottlieb, G. S., Murphy, J. T., Hamilton, D. T., Detels, R., ... Goodreau, S. M. (2019). Sexual role and hiv-1 set point viral load among men who have sex with men. *Epidemics*, 26, 68–76.
- Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409), 204–212.
- Turner, P. W., Schmid, C. S., Cranmer, S. J., & Kauermann, G. (2019). Network interdependencies and the evolution of the international arms trade. *Journal of Conflict Resolution*, 63(7), 1736–1764.
- Train, K. (2009). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press.
- Tranmer, M., Marcum, C. S., Morton, B., Croft, D., & de Kort, S. (2015). Using the relational event model (REM) to investigate the temporal dynamics of animal social networks. *Animal Behaviour*, 101, 99–105.
- Vu, D., Asuncion, A., Hunter, D., & Smyth, P. (2011). Dynamic egocentric models for citation networks. *Proceedings of the 28th International Conference on Machine Learning*, pp. 857–864.
- Vu, D., Hunter, D., Smyth, P., & Asuncion, A. (2011). Continuous-time regression models for longitudinal networks. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 24* (pp. 2492–2500). Red Hook, NY: Curran Associates.
- Vu, D., Lomi, A., Mascia, D., & Pallotti, F. (2017). Relational event models for longitudinal network data with an application to interhospital patient transfers. *Statistics in Medicine*, 36(14), 2265–2287.
- Vu, D., Pattison, P., & Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, 43, 121–135.
- Wang, P., Robins, G., & Pattison, P. (2006). PNet: A program for the simulation and estimation of exponential random graph models. Available from <http://sna.unimelb.edu.au/PNet>
- Ward, M., Ahlquist, J. S., & Rozenas, A. (2013). Gravity's rainbow: A dynamic latent space model for the world trade network. *Network Science*, 1(1), 95–118.

Wasserman, S., & Faust, K. (1994). *Structural Analysis in the Social Sciences. Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

White, L. A., Forester, J. D., & Craft, M. E. (2018). Covariation between the physiological and behavioral components of pathogen transmission: Host heterogeneity determines epidemic outcomes. *Oikos*, *127*(4), 538–552.

World Bank (2019). World Bank Open Data, Real GDP.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Fritz C, Lebacher M, Kauermann G. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time. *Statistica Neerlandica*. 2020;74:275–299. <https://doi.org/10.1111/stan.12198>

APPENDIX

ADDITIONAL DESCRIPTIVES

Figures A1 and A2 depict the distributions of in- and out-degrees in the two networks. Building on the in- and out-degrees of all nodes, these distributions represent the relative frequency of all possible in- and out-degrees in the observed networks, which is calculated with the *igraph* package in R (Csardi & Nepusz, 2006).

In the arms trade network, a strongly asymmetric relation is revealed, indicating that about 70% of the countries do not export any weapons, whereas a small percentage of countries account

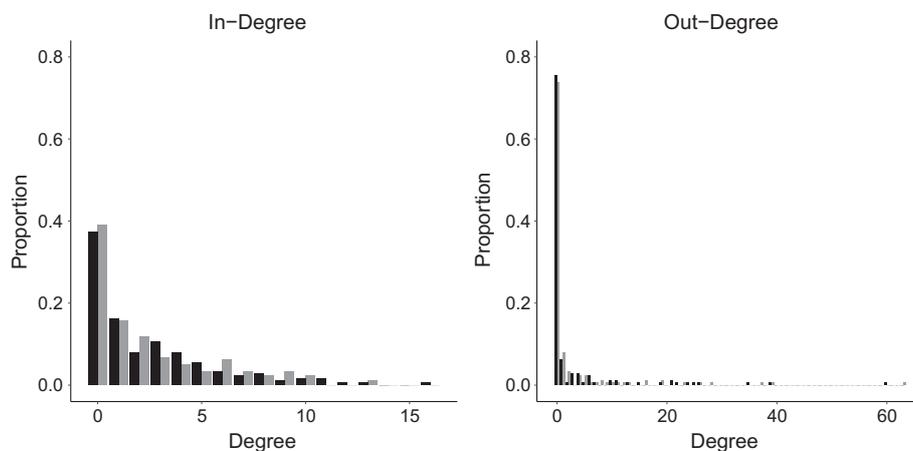


FIGURE A1 Arms trade network: Bar plots indicating the distribution of the in- and out-degrees. Black bars indicate the values of year 2016 and the grey bars, year 2017

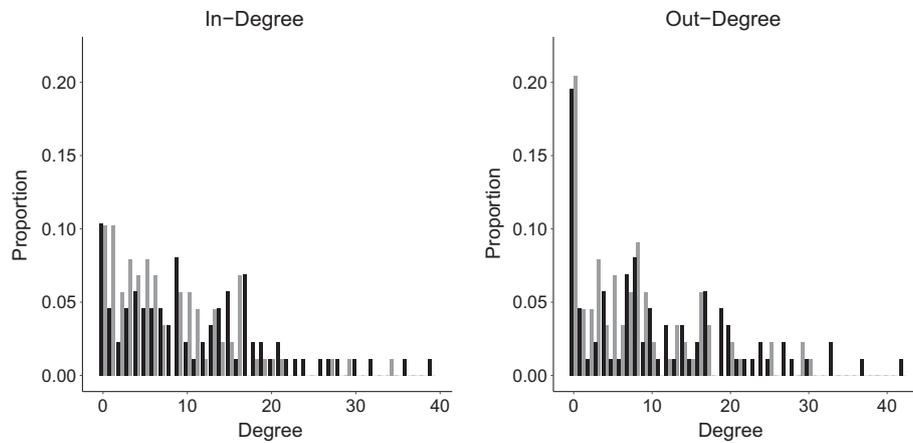


FIGURE A2 Email exchange network: Bar plots indicating the distribution of the in- and out-degrees. Black bars indicate the values of Period 1 and the grey bars, Period 2

for the major share of trade relations. The distribution of the in-degree is not that extreme, but still, we have roughly one third of all countries not importing at all.

The email exchange network shows a different structure. Here, many medium-sized in-degrees can be found, and only roughly 10% of all nodes have received no emails. For the out-degree, this number doubles (roughly 20% have not sent emails). Furthermore, the distribution of the out-degree is more skewed than the one for the in-degree.

SUPPLEMENTARY MATERIAL

Tempus Volat, Hora Fugit - A Survey of Dynamic Tie-Oriented Network Models in Discrete and Continuous Time

Cornelius Fritz¹ | Michael Lebacher¹ |
Göran Kauermann¹

Contents

1	Countries in the International Arms Trade Network	2
2	Simulation-based Model Assessment in (S)TERGMs	3
2.1	Data Set 1: International Arms Trade	4
2.2	Data Set 2: European Research Institution Email Correspondence	7
3	ROC-based Model Assessment	10
3.1	Data Set 1: International Arms Trade	10
3.2	Data Set 2: European Research Institution Email Correspondence	11
4	Application with multiple time points	12

6. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time

2

Fritz et al.

1 | COUNTRIES IN THE INTERNATIONAL ARMS TRADE NETWORK

Country Name	ISO3	Country Name	ISO3	Country Name	ISO3
Afghanistan	AFG	Germany	DEU	Niger	NER
Albania	ALB	Ghana	GHA	Nigeria	NGA
Algeria	DZA	Greece	GRC	Norway	NOR
Andorra	AND	Grenada	GRD	Oman	OMN
Angola	AGO	Guatemala	GTM	Pakistan	PAK
Antigua and Barbuda	ATG	Guinea	GIN	Palau	PLW
Argentina	ARG	Guinea-Bissau	GNB	Panama	PAN
Armenia	ARM	Guyana	GUY	Papua New Guinea	PNG
Australia	AUS	Haiti	HTI	Paraguay	PRY
Austria	AUT	Honduras	HND	Peru	PER
Azerbaijan	AZE	Hungary	HUN	Philippines	PHL
Bahamas	BHS	Iceland	ISL	Poland	POL
Bahrain	BHR	India	IND	Portugal	PRT
Bangladesh	BGD	Indonesia	IDN	Qatar	QAT
Barbados	BRB	Iran	IRN	Romania	ROM
Belarus	BLR	Iraq	IRQ	Russia	RUS
Belgium	BEL	Ireland	IRL	Rwanda	RWA
Belize	BLZ	Israel	ISR	Saint Kitts and Nevis	KNA
Benin	BEN	Italy	ITA	Saint Lucia	LCA
Bhutan	BTN	Jamaica	JAM	Saint Vincent and the Grenadines	VCT
Bolivia	BOL	Japan	JPN	Samoa	WSM
Botswana	BWA	Jordan	JOR	San Marino	SMR
Brazil	BRA	Kazakhstan	KAZ	Sao Tome and Principe	STP
Brunei Darussalam	BRN	Kenya	KEN	Saudi Arabia	SAU
Bulgaria	BGR	South Korea	KOR	Senegal	SEN
Burkina Faso	BFA	Kosovo	KOS	Serbia	YUG
Burundi	BDI	Kuwait	KWT	Seychelles	SYC
Cambodia	KHM	Kyrgyzstan	KGZ	Sierra Leone	SLE
Cameroon	CMR	Laos	LAO	Singapore	SGP
Canada	CAN	Latvia	LVA	Slovakia	SVK
Cape Verde	CPV	Lebanon	LBN	Slovenia	SVN
Central African Republic	CAF	Lesotho	LSO	Solomon Islands	SLB
Chad	TCD	Liberia	LBR	South Africa	ZAF
Chile	CHL	Libya	LYB	Spain	ESP
China	CHN	Lithuania	LTU	Sri Lanka	LKA
Colombia	COL	Luxembourg	LUX	Sudan	SDN
Comoros	COM	Macedonia (FYROM)	MKD	Suriname	SUR
DR Congo	ZAR	Madagascar	MDG	Swaziland	SWZ
Congo	COG	Malawi	MWI	Sweden	SWE
Costa Rica	CRI	Malaysia	MYS	Switzerland	CHE
Cote d'Ivoire	CIV	Maldives	MDV	Tajikistan	TJK
Croatia	HRV	Mali	MLI	Tanzania	TZA
Cuba	CUB	Malta	MLT	Thailand	THA
Cyprus	CYP	Marshall Islands	MHL	Timor-Leste	TMP
Czech Republic	CZE	Mauritania	MRT	Togo	TGO
Denmark	DNK	Mauritius	MUS	Trinidad and Tobago	TTO
Dominica	DMA	Mexico	MEX	Tunisia	TUN
Dominican Republic	DOM	Micronesia	FSM	Turkey	TUR
Ecuador	ECU	Moldova	MDA	Turkmenistan	TKM
Egypt	EGY	Mongolia	MNG	Uganda	UGA
El Salvador	SLV	Montenegro	YUG	Ukraine	UKR
Equatorial Guinea	GNQ	Morocco	MAR	United Arab Emirates	ARE
Estonia	EST	Mozambique	MOZ	United Kingdom	GBR
Ethiopia	ETH	Myanmar	MYM	United States	USA
Fiji	FJI	Namibia	NAM	Uruguay	URY
Finland	FIN	Nauru	NRU	Uzbekistan	UZB
France	FRA	Nepal	NPL	Vanuatu	VUT
Gabon	GAB	Netherlands	NLD	Viet Nam	VNM
Gambia	GMB	New Zealand	NZL	Zambia	ZMB
Georgia	GEO	Nicaragua	NIC	Zimbabwe	ZWE

TABLE 1 Countries included in the analysis of the international trade network with the ISO3 codes, that are used in the graphical representations of the network.

2 | SIMULATION-BASED MODEL ASSESSMENT IN (S)TERGMS

In Figures 1 and 4 we show simulation-based goodness-of-fit (GOF) diagnostics for the the TERGM model and in Figures 2, 3, 5 and 6 for the STERGM in the formation and dissolution model, respectively. The figures are created by the R package `ergm` (version 3.10.4) and follow the approach of Hunter et al. [2008]. In all three models, the fitted model is used in order to simulate 100 new networks. Based on these, different network characteristics are computed and visualized in boxplots.

The standard characteristics used are the complete distributions of the in-degree, out-degree, edge-wise shared partners and minimum geodesic distance (i.e. number of node pairs with shortest path of length k between them). The solid black line indicates the measurements of these characteristic in the observed network. These statistics show whether measures like GWID, GWOD and GWESP are sufficient to reproduce global network patterns. Because many shares are rather small, we visualize the simulated and observed measures on a log-odds scale.

On the bottom of the figures it is shown how well the actual network statistics are reproduced. Note, that both models compare different things as the TERGM is evaluated at y_i while the STERGM regards y^+ and y^- . Overall, all plots indicate a satisfying fit of the respective models.

2.1 | Data Set 1: International Arms Trade

Goodness-of-fit diagnostics

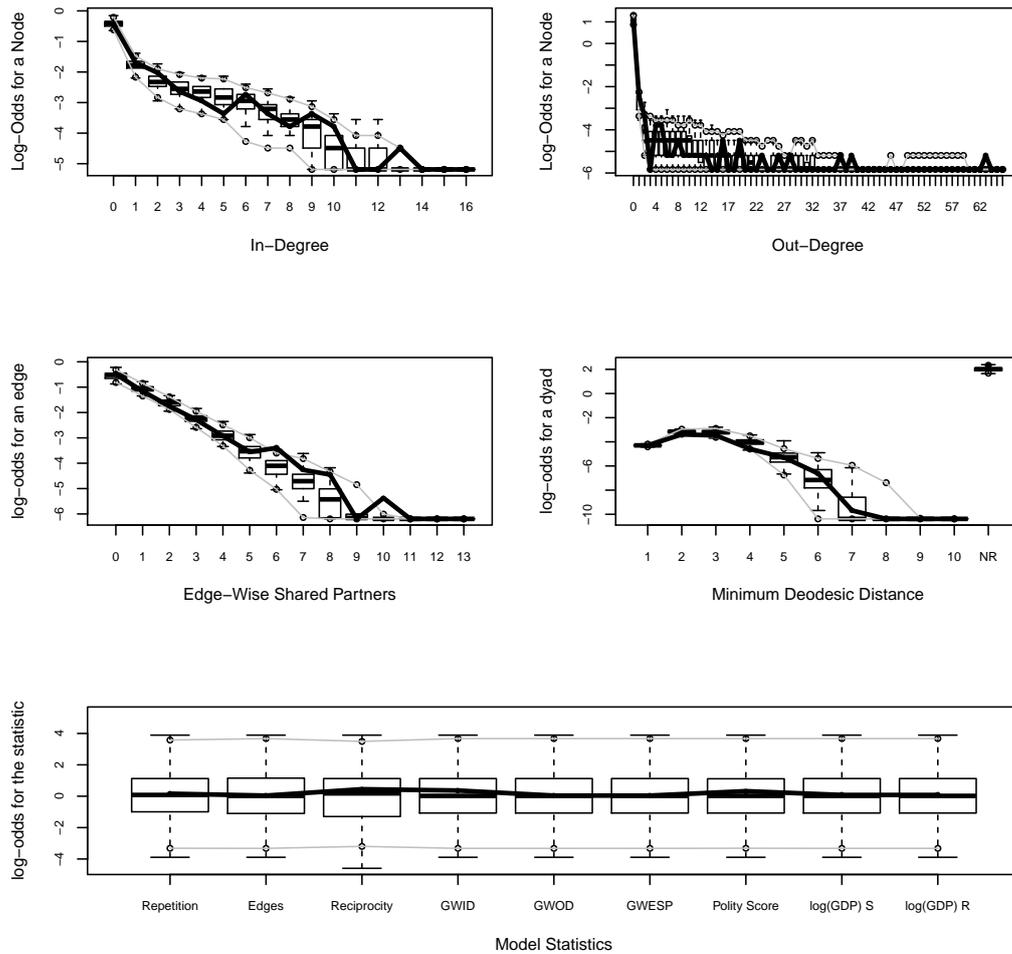


FIGURE 1 Arms trade network: Simulation-based goodness-of-fit diagnostics in the TERGM. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (middle right). All included rescaled network statistics on the bottom panel.

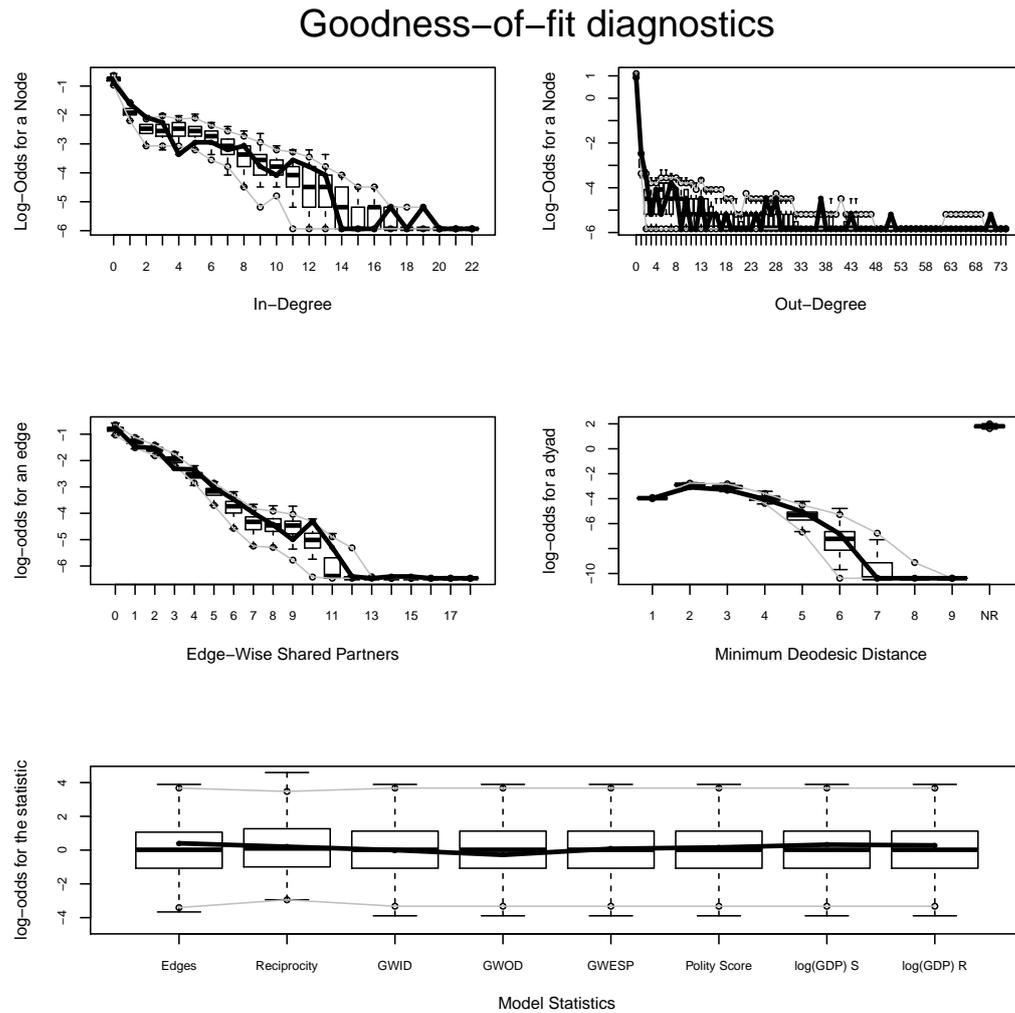


FIGURE 2 Arms trade network: Simulation-based goodness-of-fit diagnostics in the STERGM for the formation model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (middle right). All included rescaled network statistics on the bottom panel.

Goodness-of-fit diagnostics

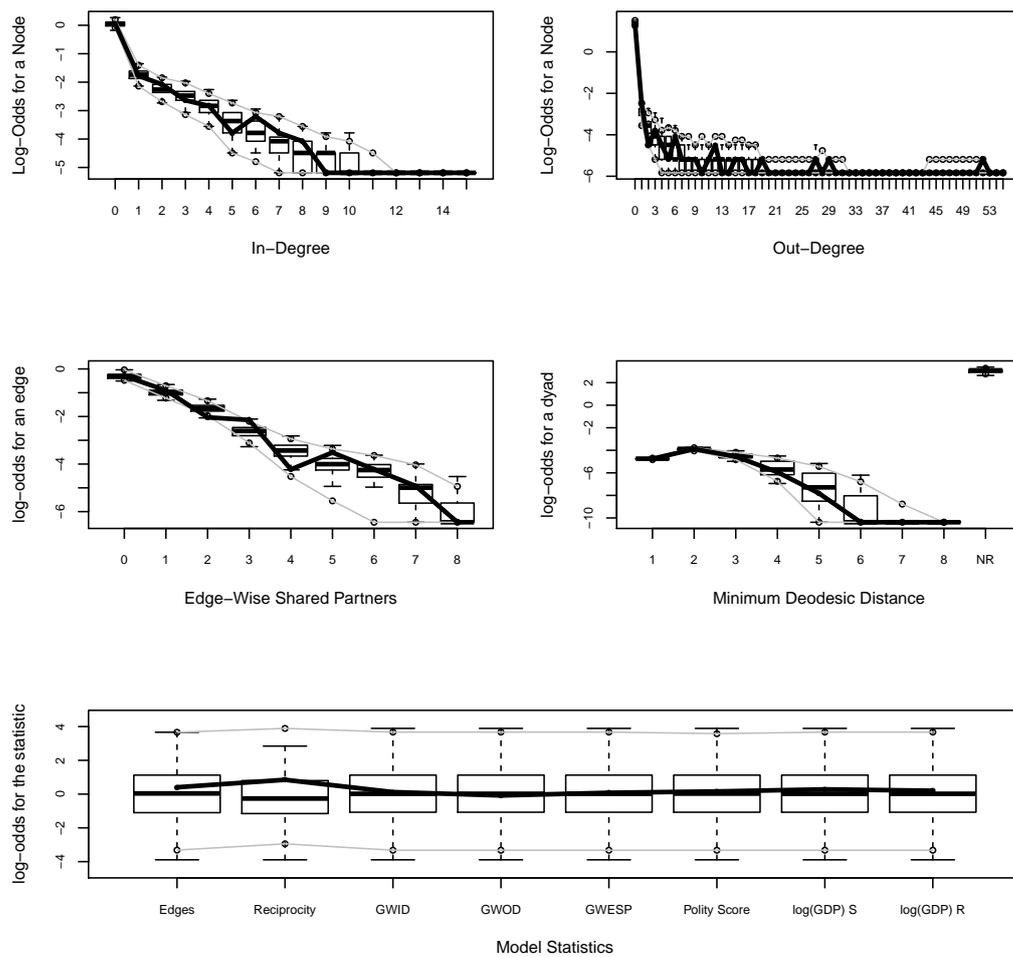


FIGURE 3 Arms trade network: Simulation-based Goodnes-of-fit diagnostics in the STERGM for the dissolution model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (between them, middle right). All included rescaled network statistics on the bottom panel.

2.2 | Data Set 2: European Research Institution Email Correspondence

Goodness-of-fit diagnostics

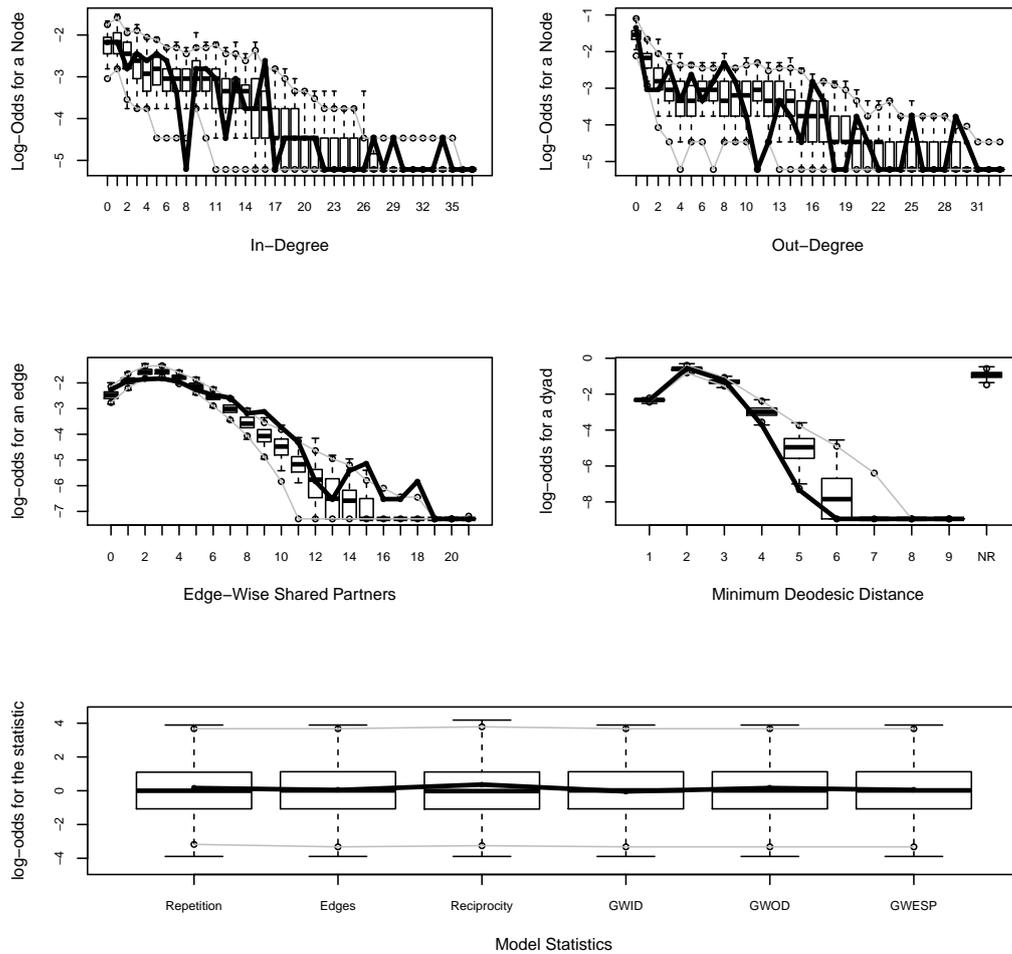


FIGURE 4 Email exchange network: Simulation-based Goodnes-of-fit diagnostics in the TERGM. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (middle right). All included rescaled network statistics on the bottom panel.

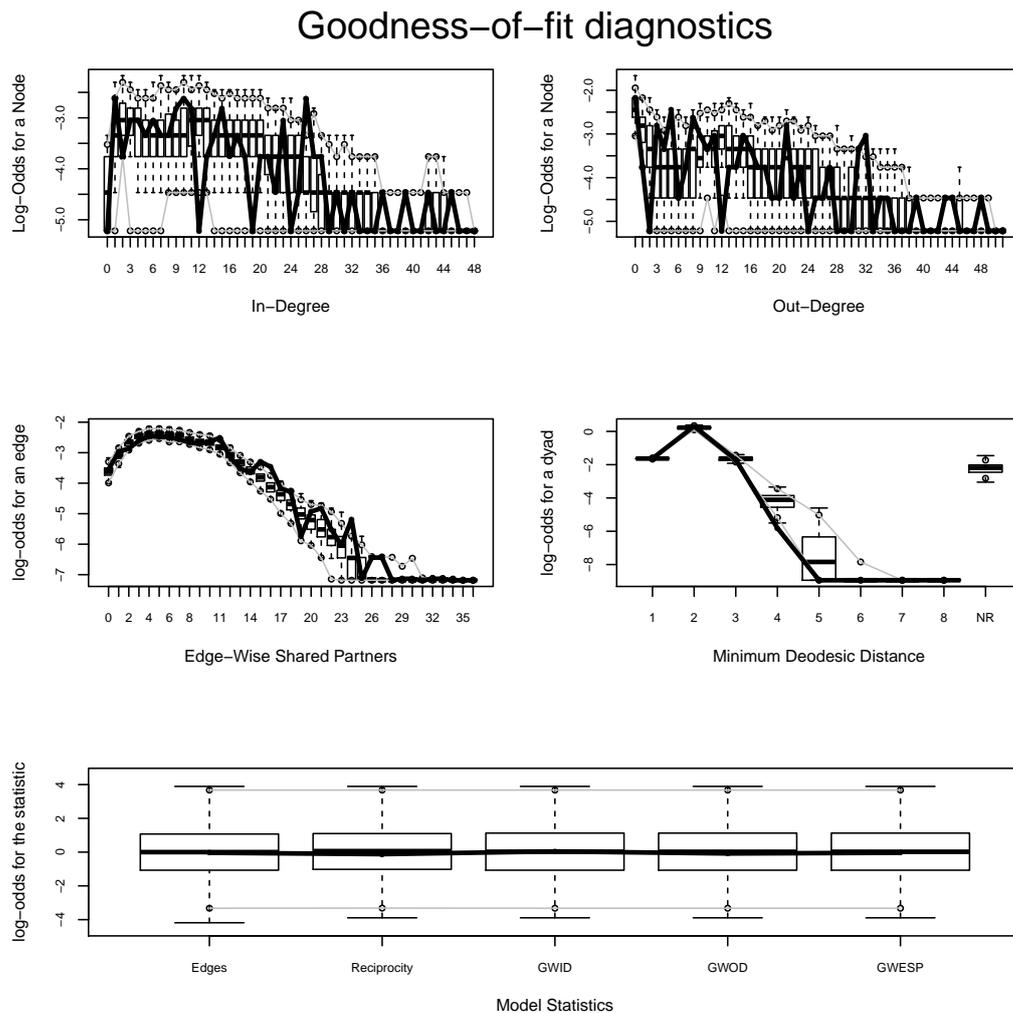


FIGURE 5 Email exchange network: Simulation-based Goodnes-of-fit diagnostics in the STERGM for the formation model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (middle right). All included rescaled network statistics on the bottom panel.

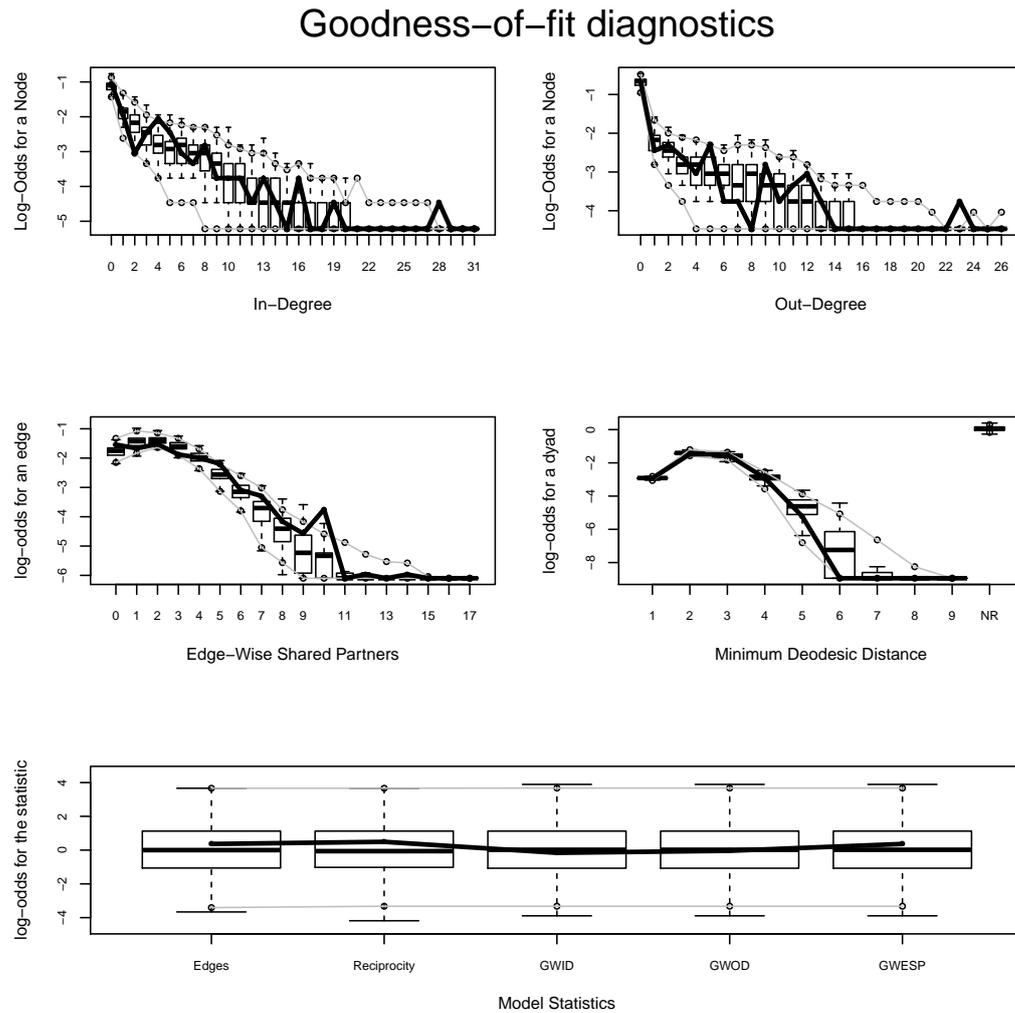


FIGURE 6 Email exchange network: Simulation-based Goodness-of-fit diagnostics in the STERGM for the dissolution model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node for different in-degrees (top left), out-degrees (top right), edge-wise shared partners (middle, left) and minimum geodesic distance (between them, middle right). All included rescaled network statistics on the bottom panel.

3 | ROC-BASED MODEL ASSESSMENT

3.1 | Data Set 1: International Arms Trade

As already stated in Section 3.3 of the main paper techniques for assessing the fit of a probabilistic classification can be used when working with binary network data. In the case of observations at discrete time points this allows an informal comparison of the models proposed in Section 3 and 4.3.

In the case of the TERGM and STERGM the application of the ROC- and PR-curve follows from the conditional probability of observing a specific tie (see equation (2.5) of Hunter and Handcock, 2006). For the REM we predict the intensities of all possible events given the information of $t - 1$ and use this value as a score in the calculation of the ROC curve. While the latter approach is non-standard and can only be applied to REMs that regard durable ties, it enables a direct comparison between the models as shown in Figure 7. The results of the ROC curve indicate a generally good fit of all models. In the STERGM more parameters are estimated, which seems to lead to a slightly bigger area under curve (AUC) values as compared to the REM and TERGM. Similar to the conclusions from the ROC curve, the PR curve favors the TERGM and STERGM over the REM.

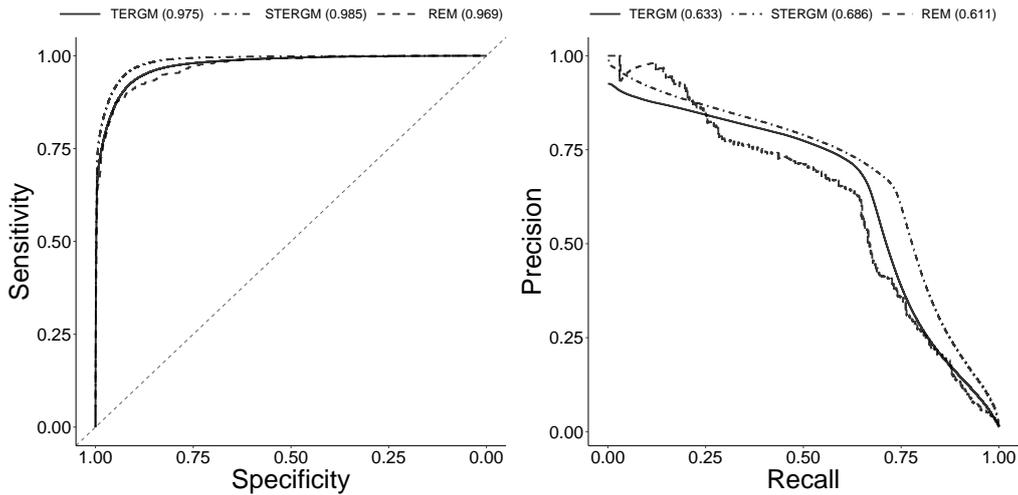


FIGURE 7 Arms trade network: ROC (left) and PR (right) curves from the TERGM (dotted line), STERGM (dot-dashed line), and REM (solid line). The AUC values of the respective curves are indicated in brackets.

3.2 | Data Set 2: European Research Institution Email Correspondence

The second data set regards reoccurring events in the REM, which are aggregated for the analysis of the TERGM and STERGM. Therefore, the ROC and PR curve are only available for the TERGM and STERGM. The results are depicted in Figure 8. For this data set, the ROC curves favor the TERGM. Yet, when emphasis is put on finding the true positives, the PR curve detects a better model fit of the STERGM.

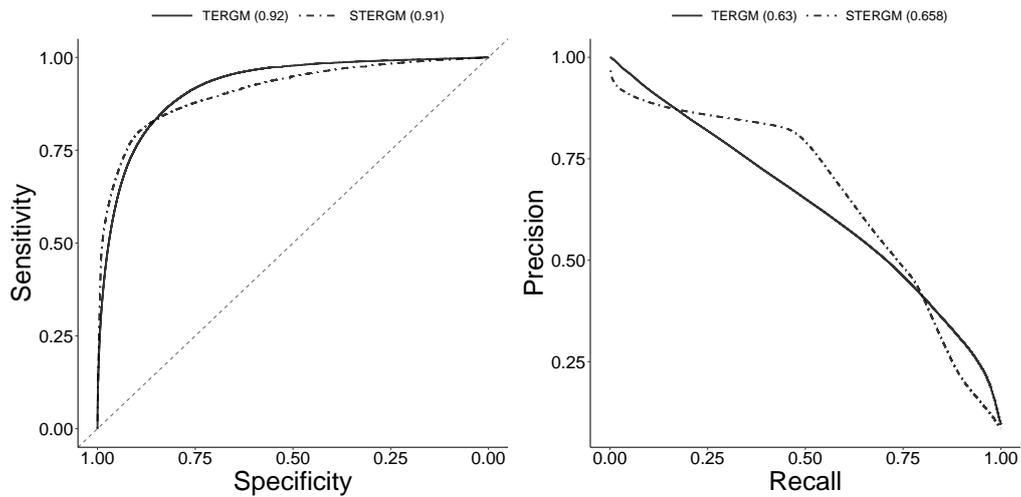


FIGURE 8 Email exchange network: ROC (left) and PR (right) curves from the TERGM (dotted line), STERGM (dot-dashed line), and REM (solid line). The AUC values of the respective curves are indicated in brackets.

As explained in Section 4.4 one option to assess the goodness-of-fit of REMs is the recall measure as proposed by Vu et al. [2011a,b]. We apply the measure in three different situations that may be of interest when measuring the predictive performance of relational event models: predict the next tie, next sender, and next receiver. The worst case scenario in terms of predictions of a model would be random guessing of the next sender, receiver, or event, the resulting recall rates are indicated by the dotted lines. The results in Figure 9 exhibit a good predictive performance of the REM, i.e. in about 75% of the events the right sender and receiver is among the 25 most likely senders and receivers.

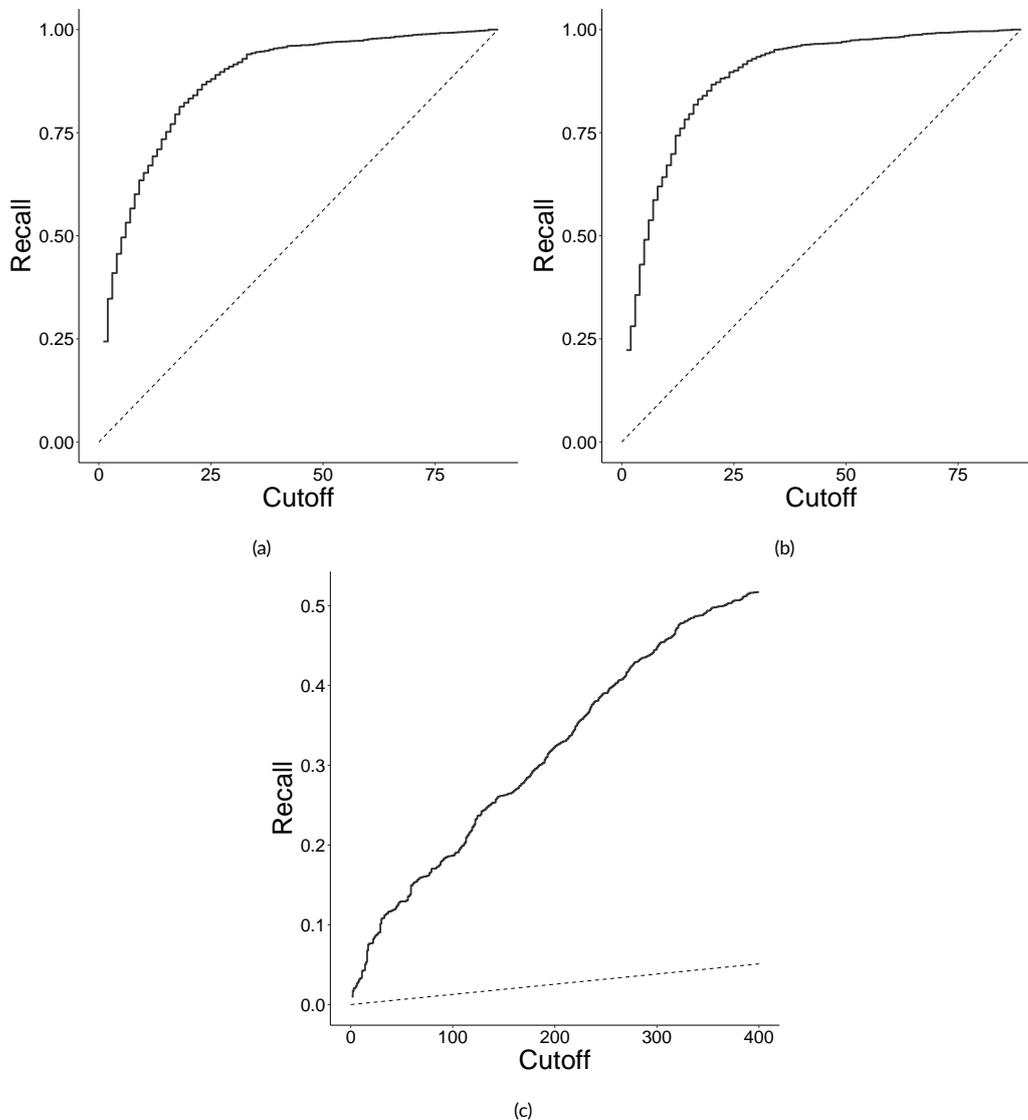


FIGURE 9 Email exchange network: Recall curves of REM regarding the next sender (a), receiver (b), and event (c). The dotted line indicates the measure under random guessing of the next sender, receiver, or event.

4 | APPLICATION WITH MULTIPLE TIME POINTS

In the main article, we fitted a TERGM as well as a STERGM to two time points, called *period 1* and *period 2*. However, it is possible to fit these models to multiple transitions. In order to do so, we took the first two years of the email exchange network and aggregated the deciles into 10 binary networks. Using again a first order Markov assumption and conditioning on the first network, this allows to fit a TERGM as well as a STERGM to the remaining nine networks.

	TERGM	STERGM		REM	
		Formation	Dissolution		
Repetition	1.986*** (0.058)	–	–	2.354*** (0.048)	–
Edges	–3.435*** (0.059)	–4.485*** (0.073)	–0.769*** (0.107)	–	–
Reciprocity	1.090*** (0.069)	2.755*** (0.079)	1.761*** (0.113)	1.699*** (0.043)	
In-Degree (GWID)	–1.239*** (0.099)	–0.870*** (0.156)	–0.307* (0.152)	0.007*** (0.002)	In-Degree Receiver
Out-Degree (GWOD)	–1.574*** (0.097)	–1.757*** (0.155)	–0.056 (0.159)	0.009*** (0.001)	Out-Degree Sender
GWESP	0.497*** (0.028)	0.508*** (0.029)	0.124** (0.043)	0.199*** (0.011)	Transitivity

TABLE 2 Email exchange network: Comparison of parameters obtained from the TERGM (first column) and STERGM (Formation in the second column, Dissolution in the third column). Standard errors in brackets and stars according to p -values smaller than 0.001 (***), 0.05 (**) and 0.1 (*). Decay parameter of the geometrically weighted statistics is set to $\log(2)$.

For comparison we additionally fit a REM to the data set. The estimation is done using the function `mtergm` from the package `btergm` (version 3.6.1) (Leifeld et al., 2018) that implements MCMC-based maximum likelihood. The STERGM is fitted again using the package `tergm` (version 3.5.2).

The corresponding results can be found in Table 2 in column 1 to 3. Note that the parameter estimates still refer to the transition from $t - 1$ to t and can be interpreted in the same way as in the main article. In that regard note, that it is now assumed that the coefficients stay constant with time. Possible approaches to relax this assumption were given in the main article. The estimates of the REM (column 4) are consistent with the (S)TERGM but slightly differ to the main article, since now we condition only on the first out of 10 *periods* and hence model more events.

references

- Hunter, D. R., Goodreau, S. M. and Handcock, M. S. (2008) Goodness of fit of social network models. *Journal of the American Statistical Association*, **103**, 248–258.
- Hunter, D. R. and Handcock, M. S. (2006) Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, **15**, 565–583.
- Leifeld, P., Cranmer, S. J. and Desmarais, B. A. (2018) Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. *Journal of Statistical Software*, **83**.
- Vu, D., Asuncion, A., Hunter, D. and Smyth, P. (2011a) Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning*, 857–864.
- Vu, D., Hunter, D., Smyth, P. and Asuncion, A. (2011b) Continuous-Time Regression Models for Longitudinal Networks. In *Advances in Neural Information Processing Systems 24* (eds. J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira and K. Q. Weinberger), 2492–2500.

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

Contributing article

Fritz, C. and Kauermann, G. (2022). On the Interplay of Regional Mobility, Social Connectedness, and the Spread of COVID-19 in Germany (2022). *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 185 (1), 400-424 doi:10.1111/rssa.12753..

Replication code

Facebook collected the mobility and connectivity data and provided as access to an anonymized version of the data. We cannot share the raw data due to a data agreement. Still, we are allowed to provide all data aggregated onto the level of federal districts. To guarantee the replicability of our results, we make the complete code to obtain the results from this article available as supporting information under <https://rss.onlinelibrary.wiley.com/doi/10.1111/rssa.12753>.

Copyright information

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited.

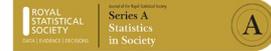
Author contributions

The general idea of contacting Facebook to ask for data on human mobility and network data can be attributed to Göran Kauermann. Discussions between Göran Kauermann and Cornelius Fritz resulted in the proposed methodological approach that combines methods from multivariate statistics and generalized additive models. Cornelius Fritz, as the leading author, prepared an initial draft of the complete paper and implemented the model. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

Received: 6 August 2020 | Accepted: 31 August 2021

DOI: 10.1111/rssa.12753

ORIGINAL ARTICLE



On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

Cornelius Fritz | Göran Kauermann

Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence

Cornelius Fritz, Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539, Munich, Germany.
Email: cornelius.fritz@stat.uni-muenchen.de

Funding information

European Cooperation in Science and Technology [COST Action CA15109 (COSTNET)]; Munich Center for Machine Learning (MCML); German Federal Ministry of Education and Research (BMBF), Grant/Award Number: 01IS18036A

Abstract

Since the primary mode of respiratory virus transmission is person-to-person interaction, we are required to reconsider physical interaction patterns to mitigate the number of people infected with COVID-19. While research has shown that non-pharmaceutical interventions (NPI) had an evident impact on national mobility patterns, we investigate the relative regional mobility behaviour to assess the effect of human movement on the spread of COVID-19. In particular, we explore the impact of human mobility and social connectivity derived from Facebook activities on the weekly rate of new infections in Germany between 3 March and 22 June 2020. Our results confirm that reduced social activity lowers the infection rate, accounting for regional and temporal patterns. The extent of social distancing, quantified by the percentage of people staying put within a federal administrative district, has an overall negative effect on the incidence of infections. Additionally, our results show spatial infection patterns based on geographical as well as social distances.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. Journal of the Royal Statistical Society: Series A (Statistics in Society) published by John Wiley & Sons Ltd on behalf of Royal Statistical Society

KEYWORDS

COVID-19, infectious disease modelling, semiparametric regression, social connectedness, social networks, spatial network data

1 | INTRODUCTION

The COVID-19 virus outbreak originating in mainland China leapt over to Europe and quickly evolved to a global pandemic in March 2020. Only through strict non-pharmaceutical interventions (NPI) could most national health systems rapidly react to this new threat. In numerous scientific efforts, physical distancing measures were discovered to be the most effective interventions (Prem et al., 2020) and found to be necessary maybe until 2022 (Kissler et al., 2020). The measures' effectiveness emanates from researchers confirming that the main form of virus transmission is person-to-person interaction (Chan et al., 2020). The virus can be spread by inhaling microscopic aerosol particles that contain COVID-19 and remain viable in the air with a half-life of about 1 h (Asadi et al., 2020) or direct contact through the exchange of virus-containing droplets with infected individuals (Guan et al., 2020). Since also a high proportion of cases is asymptomatic (Lavezzo et al., 2020) and gets infected by cases in the presymptomatic stage (Li et al., 2020b), human mobility can explain the spread of COVID-19 to a considerable extent (Kraemer et al., 2020).

Stemming from the consequential need to account for contact patterns when investigating the spread of COVID-19, Oliver et al. (2020) list multiple possibilities of how one may utilise mobile phone data to do so. To enable this type of research, Facebook extended the *Data for Good* program to a broader audience of researchers and provided so-called *Disease Prevention Maps* for multiple countries (Maas et al., 2019). This database includes measurements on quantities like co-location, user counts and movement ranges on a regional level derived from information of more than 26 million Facebook users. Additionally, a measure for the social connectedness between geographical regions is supplied (Bailey et al., 2018). In various studies, this data source was employed to demonstrate how the impact of lockdown measures in Italy was more severe for municipalities with higher fiscal capacities (Bonaccorsi et al., 2020), quantify social and geographical spillover effects from relaxations of shelter-in-place orders (Holtz et al., 2020) and predict the number of infections on a granular spatiotemporal resolution using contact tracing data (Lorch et al., 2020).

This article uses the same data source to analyse how regional differences in mobility patterns and friendship proximity affect the spread of COVID-19 in Germany. While NPIs, for example, the nationwide shutdown in Germany that started 22 March, had an evident impact on national human mobility and ceased the exponential spread of the virus (Flaxman et al., 2020), the effect of the relative movement between regional districts was not yet fully assessed. So far, studies concerning human movements during the current pandemic are focused mainly on how the lockdown affected national human mobility (Galeazzi et al., 2020) or specific regions regarding their economic status (Bonaccorsi et al., 2020). To fill this gap, we derive covariates from the mobility data to quantify the overall dispersion of meeting patterns and compliance with social distancing. Through weekly standardisation of the covariates, we control for the dynamics therein, which are, in turn, driven by NPIs. As a result, our research enables a quantitative assessment of different mobility strategies relative to the

national average. Also, we infer positions of the federal administrative regions in a social space from the information on the relative friendship links among them using multidimensional scaling (Cox & Cox, 2000). Subsequently, we relate the processed data to Germany's weekly rate of local COVID-19 infections between 3 March and 22 June 2020. This time frame permits the analysis of the dynamic spread starting with the WHO declaring COVID-19 a pandemic (WHO, 2020).

We employ a spatiotemporal regression model for the ratio of local COVID-19 infections that takes autoregressive structures, age- and gender-specific effects, contagion by nearby districts in the geographical and social space, as well as latent heterogeneities between the districts into account. Our method is closely related to the surveillance model introduced by Held et al. (2005). They extend generalised linear models to analyse surveillance data from epidemic outbreaks. This approach was expanded to handle multivariate surveillance data (Paul et al., 2008), control for seasonality and spatial heterogeneity (Held & Paul, 2012) and include neighbourhood information from social contact data (Meyer & Held, 2017). In contrast to this type of model, our model's objective is to investigate the connection between mobility patterns, social connectivity and the spread of COVID-19 in an interpretable manner. While forecasting infections is undoubtedly a central objective in epidemic surveillance, this is not the main focus of our work (see also Held et al., 2017).

The rest of the article will be structured as follows: We discuss the data sources, its measures on social interaction as well as mobility in Section 2. In Section 3, we detail our proposed modelling approach. We propose an imputation model for missing onset dates and use a semiparametric spatiotemporal model to analyse the ratio of local COVID-19 infections with a specific disease onset date. The results of the analysis are presented in Section 4. Section 5 concludes the article.

2 | DATA DESCRIPTION

2.1 | Data on infections

Our application's outcome of interest is the ratio of COVID-19 infections in a federal administrative district (NUTS-3 level), which we define as the quotient of the number of COVID-19 infections over the corresponding population size. In Germany, there are $n = 401$ federal administrative districts (a complete list is given by the German Federal Statistical Office). At a higher hierarchical level, each federal district also belongs to a federal state (NUTS-1 level). In most figures, for example Figure 2, we colour-code the district-specific time series according to this allocation. If we refer to a specific district in the text, we generally specify the corresponding federal state in brackets.

Infection count: The Robert-Koch-Institute provides timely data on the daily number of COVID-19 infections in Germany for each federal district. We limit the present analysis to individuals between 15 and 59 years old due to the age structure in the Facebook population. Besides, the given surveillance counts are stratified by age group (15–35 and 36–59) and gender. For each entry, dates of symptom onset and reporting are given, although the onset date is partially missing. Our principal analysis is based on the disease onset date since it ensures more valid information on the infection incidence (Günther et al., 2020). Imputation of the missing values is required (we present our method in Section 3.1). By $y_{i,g,t}$ we denote the observed (and partially imputed) counts of new onsets within district i , age/gender-group g and week t . For completeness, we define with $x_{i,g}$ the corresponding indicator for the age/gender group.

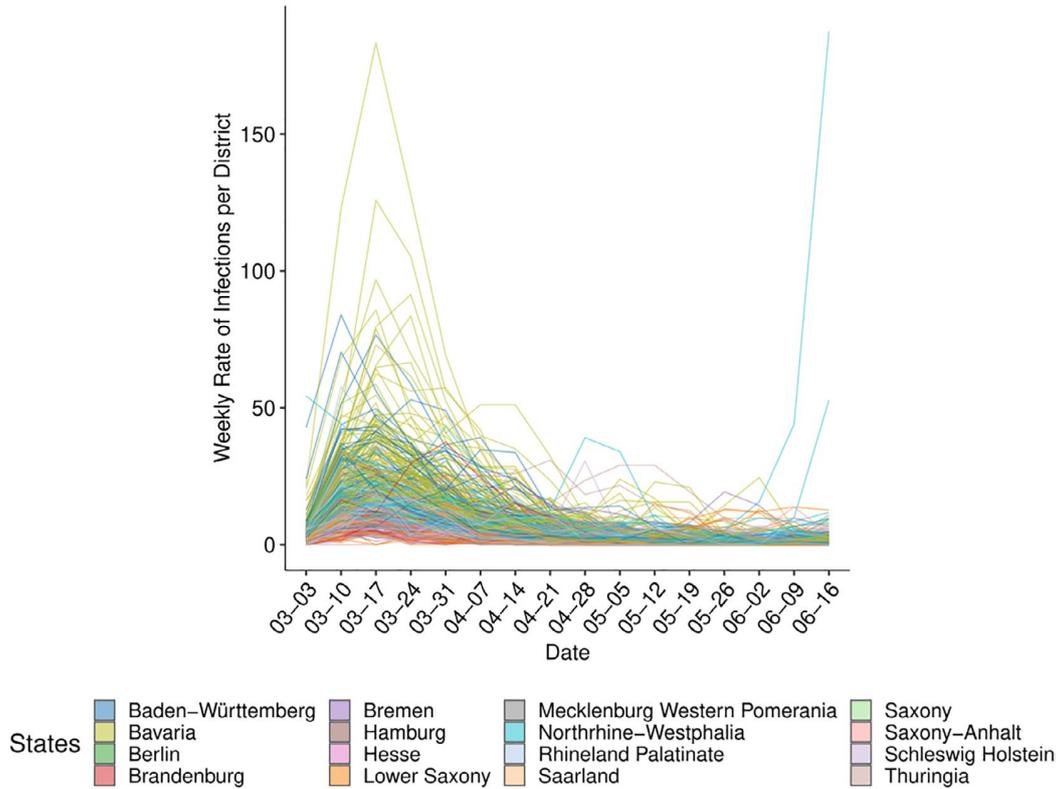


FIGURE 1 Observed Rate of Weekly Infections for each federal district. The colour of the lines indicate the federal state in which each district is located and the dates (mm:dd) are the first day of the corresponding week

Population: We obtained district-, age- and gender-specific population data from the German Federal Statistical Office. To guarantee a consistent definition of age-groups, we categorised the data according to the two primary age groups according to which the infection data are reported, namely people between 15–35 and 36–59 years old. The corresponding time-constant covariate is denoted for age/gender-group g in district i by $x_{i,g,pop}$.

The observed rates per 10.000 inhabitants $\bar{y}_{i,g,t} = \frac{10,000y_{i,g,t}}{x_{i,g}}$ are visualised in Figure 1 colour-coded according to the different states. For each week, we plot the rate of disease onsets that we partially impute in case of missingness, as described in detail in the next section. Once the first peak of infections could be overcome, the cases in the aftermath are increasingly attributed to local outbreaks. Two districts, namely Guetersloh and Warendorf (North Rhine-Westphalia), experience a local outbreak in a meat factory during the last weeks of the observational period (Kottasová, 2020). This local outbreak encompasses 48% of all infections with disease onset in the week starting on 16 June.

2.2 | Data on social activity during COVID-19

All data related to social activities during the COVID-19 pandemic are generated from approximately 10 million Facebook users in Germany, who enabled geolocation features in the Facebook

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

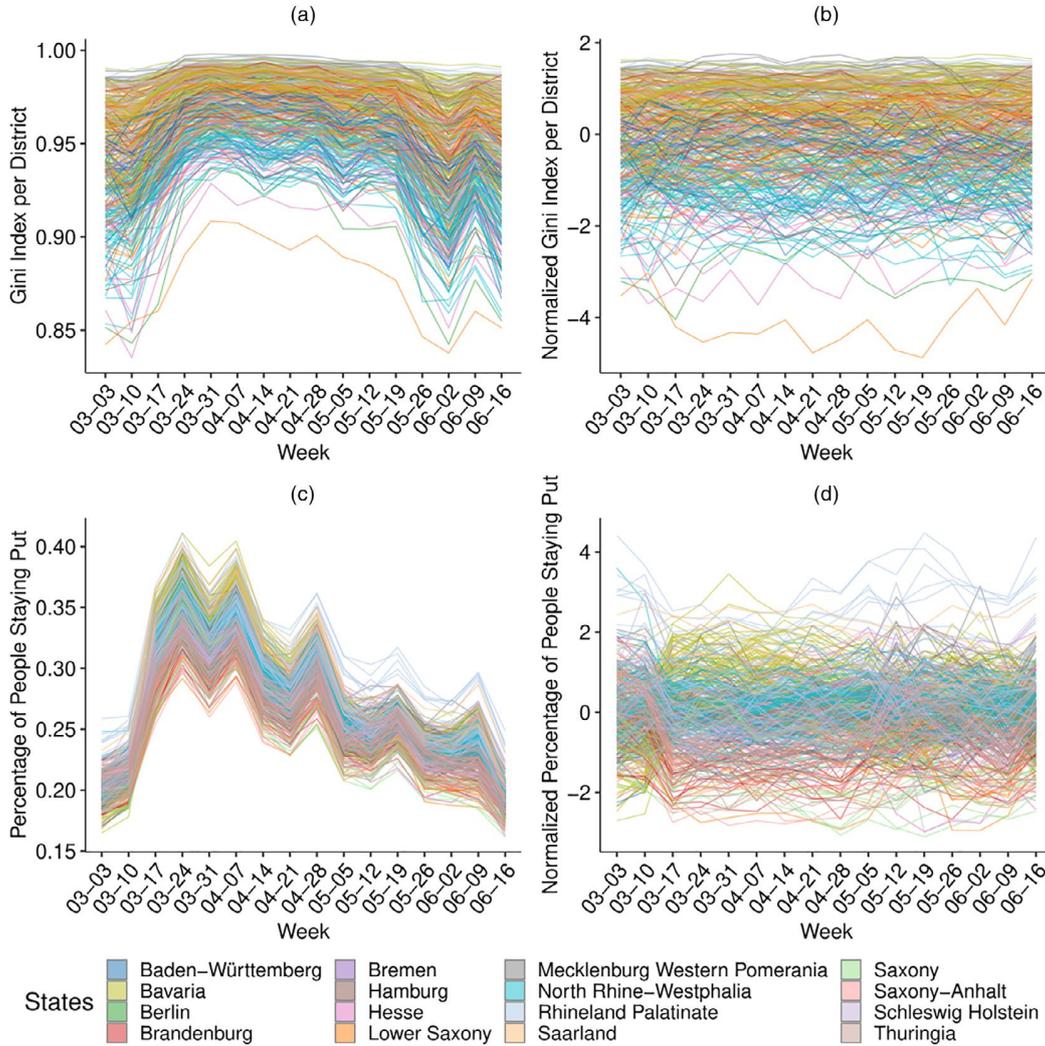


FIGURE 2 (a) Gini indices for each district over time. (b) Standardised Gini indices for each district over time. (c) Percentages of people staying put for each district over time. (d) Standardised percentages of the people staying put. The colour of the lines indicate the state in which each district is located and the dates (mm:dd) are the first day of the corresponding week

app on their mobile devices. To abide by the privacy policies, the observations are anonymised through aggregation onto tile bins polygons, censoring if we observed not enough users in the spatial region, as well as randomisation using additional noise and spatial smoothing (Maas et al., 2019). We aggregate the polygons to the same spatial units for our application on which we have the infection data. We propose the following measures describing social interaction and mobility. All measurements are taken weekly, where we use simple averaging for quantities available at a more granular temporal resolution.

Co-location: Co-location in week t is measured by the probability $p_{ij,t}$ of a random person from district i to be located in the same $0.6 \text{ km} \times 0.6 \text{ km}$ square as another random person from district j (Iyer et al., 2020). These probabilities are then used to construct a district-wise quantity for the concentration of meeting patterns using the Gini index, which is given by:

$$x_{i,t,gini} = \frac{\sum_{m,l \neq i} |p_{im,t} - p_{il,t}|}{2(n-1) \sum_{j \neq i} p_{ij,t}}.$$

If we were to observe the maximal value of 1 in $x_{i,t,gini}$ all people within federal district i would only meet people (i.e. Facebook users) from only one further district. This behaviour is exemplary of extremely restricted mobility. Conversely, a lower value heuristically indicates dispersed meeting patterns. Due to this intuitive interpretation, we opt for the Gini index as a measure of concentration. The Gini indices' temporal paths for the 401 districts in Germany are depicted in Figure 2a. Overall, the meeting patterns become more concentrated on a few other districts as the crisis evolves. This behaviour contrasts rather dispersed practices before the pandemic. An upward trend is visible until the nationwide lockdown on 22nd of March, 2020.¹ Thereupon, meeting patterns continue to be overall condensed, although the indices slowly decline. To enable a meaningful comparison between the respective estimates in the regression setting of Section 3, we standardise the Gini indices per week. The standardised covariate $\tilde{x}_{i,t,gini}$ is shown in Figure 2b and given by

$$\tilde{x}_{i,t,gini} = \frac{x_{i,t,gini} - \hat{\mu}_{t,gini}}{\hat{\sigma}_{t,gini}}, \quad (1)$$

where $\hat{\mu}_{t,gini} = \frac{1}{n} \sum_{j=1}^n x_{j,t,gini}$ and $\hat{\sigma}_{t,gini} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{j,t,gini} - \hat{\mu}_{t,gini})^2}$.

Percentage staying put: Besides the relative attribution of co-location probabilities to other districts, we investigate a measure that expresses how people (Facebook users) comply with social distancing. We quantify this concept by the covariate $x_{i,t,sp}$, which is defined as the average percentage of people in district i staying put during week t . Respective data were collected using geolocation traces of mobile devices and users are defined to be staying put, if they are only observed in one $0.6 \text{ km} \times 0.6 \text{ km}$ square throughout a day (Facebook, 2020). In Figure 2c, clear break-points are visible, giving evidence of the temporary lockdown that started between 17 and 24 March. During the following weeks, the observed values gradually level off around pre-lockdown values. We also observe some peaks in the weeks starting on 7 and 28 April, which could be traced back to the different mobility behaviour during national holidays, namely *Good Friday* on 10 April and *Labour Day* on 1 May 2020.

Similarly to the treatment of the Gini index, we standardise the percentages in the regression setting. While the visual impression from Figure 2c insinuates that the dynamics of people staying put are similar between districts, the standardised paths, given in Figure 2d, reveal local differences between them. For instance, the early look-down in Bavaria resulted in a substantial relative increase of the respective districts between the 10th and 17th of March, see the yellow-green lines.

Friendship distance: Spatial distance is found to be strongly associated with the spread between regions (Kang et al., 2020). Beyond the geographical proximity, Cho et al. (2011) argued that friendship ties explain specifically long-distance mobility, which is fundamental for understanding the early spread of the pandemic (Chinazzi et al., 2020). To accommodate this possible line of infection, we include a measure for the strength of friendship ties between the districts of Germany. More precisely, we employ the social connectedness index proposed by Bailey et al. (2018), which is based

¹In Bavaria, the lockdown started already on 16 March 2020.

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

on an anonymised snapshot of all active Facebook users and their friendship networks from April 2020. For the administrative district i and j , the time-invariant measure $x_{ij,soc}$ is given by:

$$x_{ij,soc} = \frac{\#\{\text{Friendship Ties between users in district } i \text{ and } j\}}{\#\{\text{Users in district } i\} \#\{\text{Users in district } j\}} \quad \forall i, j \in \{1, \dots, n\}. \quad (2)$$

In a note, Kuchler et al. (2021) uncover high correlations between the social connectedness indices and the spread of COVID-19. This index is further processed to provide a spatial allocation based on social instead of Euclidean distances. To do so, we first transform social connectedness to social distance d_{soc} by taking the reciprocal of connectedness, that is, $d_{ij,soc} = \frac{1}{x_{ij,soc}}$. Consecutively, we process these distances to coordinates using *multidimensional scaling* (Cox & Cox, 2000). In our application, this procedure's result is a two-dimensional representation of each district's location in the network defined through Equation (2) that is only identifiable up to the scale and rotation. Using *Procrustes analysis*, we map the rotation of the inferred coordinates in the friendship space to be most similar to the geographical coordinates (Cox & Cox, 2008). Technical details on both procedures are given in Annex A. The outcome of the algorithm for each district i is denoted by $x_{i,soc}$ and gives the geo-coordinates in the friendship space as shown in Figure 3. Robust connectivity within federal states and neighbouring districts are visible in the friendship coordinates. We also observe that the capital, Berlin, is situated in the very centre, reflecting its unique and highly connected position. One can also detect a persisting corridor between districts located in former East- and West Germany. Next

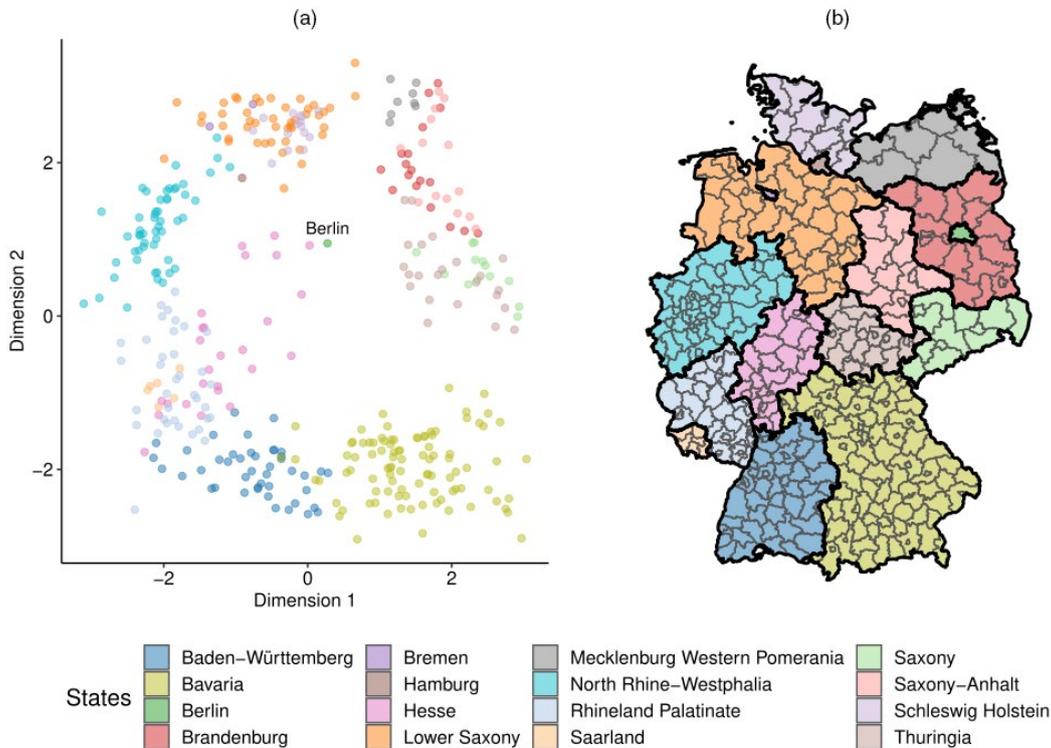


FIGURE 3 (a) Coordinates representing the friendship distances. The colour of the points indicates the state in which each district is located. (b) Map representing the colour legend. The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts

to the social coordinates, we incorporated each district's geographical coordinates $x_{i,coord}$ that is, the longitude and latitude of each districts centroid, in our application.

3 | MODELLING

We start by proposing a model to impute missing dates of the disease onset. Subsequently, these partially imputed infection data are modelled with a negative binomial regression.

3.1 | Imputation model

We can see in Figure 4 that approximately 30% of the onset dates are missing. To still make use of all available information, we propose to impute missing disease onset dates under the assumption of missingness at random. This allows for unbiased findings which are not guaranteed when using a complete case analysis (Little & Rubin, 2002). In particular, we leverage the fact that the chronologically later reporting date is available for all cases. Thereby, the problem of imputing the date of disease onset for a single infection is reduced to imputing the time between onset of disease and its reporting through a positive test, which we call test delay. Following Günther et al. (2020), we use the subset of all data without any missing disease onset dates to fit a distributional regression model for this test delay. In the next step, we predict all distributional parameters under this model for all cases with a missing disease onset date and sample the missing onset date.

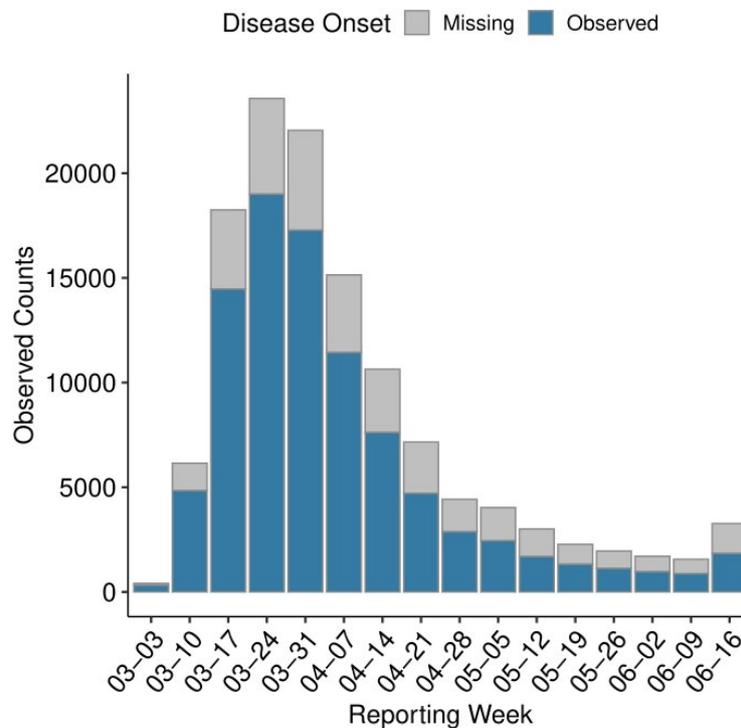


FIGURE 4 Count of missing and observed disease onset dates per reporting week

To fit the imputation model, we first disaggregate the given surveillance counts to the patient level. For each complete case l , the data include the age/gender group indicator ($x_{l,g}$), an indicator whether the reporting date was during a weekend ($x_{l,weekend}$), and the state ($x_{l,state}$) and district ($x_{l,district}$) where it was observed. Regarding the temporal information of each infection, we are given the date of disease onset ($t_{l,o}$) and its reporting ($t_{l,r}$). For complete-case data, the test delay is then given by $d_l = t_{l,r} - t_{l,o}$. As regressors in the imputation model, we include dummy covariates $x_l = (x_{l,g}, x_{l,weekend}, x_{l,state}, x_{l,district})$ and the metric covariate $t_{l,r}$ itself to account for changing testing strategies, for example, during the early spread the test capacities were limited and patients needed to wait longer for a test to be conducted. We assume that d_l is a realisation of random variable D_l , which follows a negative binomial model:

$$D_l | x_l, t_{l,r} \sim NB \left(\mu_l = \exp \left\{ \theta_\mu^\top x_l + f_\mu(t_{l,r}) \right\}, \sigma_l = \exp \left\{ \theta_\sigma^\top x_l + f_\sigma(t_{l,r}) \right\} \right), \quad (3)$$

where $\mathbb{E}(D_l | x_l, t_{l,r}) = \mu_l$ and $\text{Var}(D_l | x_l, t_{l,r}) = \mu_l + \sigma_l \mu_l^2$ holds. A discrete-valued distribution appears most suitable since the patient-level data are available daily, making the test delay inherently discrete. As indicated in Equation (3), we model the location and scale parameters of the distribution by separate linear predictors. Note that the linear predictors are defined by $\eta_\mu = \theta_\mu^\top x_l + f_\mu(t_{l,r})$ and $\eta_\sigma = \theta_\sigma^\top x_l + f_\sigma(t_{l,r})$ for the corresponding distributional parameters and that the linearity only refers to linearity in the coefficients not in the covariates. Therefore, the model lies within the family of generalised additive models for location, scale and shape (Rigby & Stasinopoulos, 2005). While all components of x_l have a log-linear effect, we parameterise the trend effect of the reporting date $t_{l,r}$ by nonlinear penalised splines (see Eilers & Marx, 1996 for details). The district-specific effects are assumed to be Gaussian. After having obtained the estimates, we calculate $\hat{\mu}_{\bar{l}} = \exp \{ \hat{\theta}_\mu^\top x_{\bar{l}} + \hat{f}_\mu(t_{\bar{l},r}) \}$ and $\hat{\sigma}_{\bar{l}} = \exp \{ \hat{\theta}_\sigma^\top x_{\bar{l}} + \hat{f}_\sigma(t_{\bar{l},r}) \}$ for all observations \bar{l} with missing disease onset. We can now simulate $d_{\bar{l}}$ from Equation (3) to acquire a full data set by setting $t_{\bar{l},o} = t_{\bar{l},r} - d_{\bar{l}}$. Through aggregation from the daily patient-level data to the infection counts per district i and age/gender group g with disease onset in week t , denoted by $y_{i,g,t}$, we build a single partially imputed data set. This procedure is repeated K times to represent the uncertainty associated with the missing information of all disease onsets.

3.2 | Infection model

To model the rate of infections with partially imputed data, we apply a negative binomial ‘*observation-driven*’ model for count data including the population as an offset term (Cox, 1981). By doing so, we assume

$$Y_{i,g,t} | x_{i,g,t-1}, a_i, b_i \sim NB(\mu_{i,g,t}, \sigma), \quad \forall i \in \{1, \dots, 401\}, g \in \mathcal{G}, \text{ and } t = 2, \dots, T, \quad (4)$$

where $x_{i,g,u} = (u, x_{i,g}, x_{i,g,pop}, \tilde{x}_{i,u,gini}, \tilde{x}_{i,u,sp}, x_{i,coord}, x_{i,soc}, \tilde{y}_{i,g,t-1})$ are the covariates at arbitrary week u specified in Section 2 and \mathcal{G} denotes the set of age/gender groups used from the data. Furthermore, let T be the final week of data we use in the analysis. We assume in Equation (4) that the random variable $Y_{i,g,t}$ follows a negative binomial distribution conditional on $x_{i,g,t-1}$, a_i and b_i to compensate overdispersion in the observed counts.

Aligned with models for the spread of infectious diseases (Held et al., 2005), we decompose $\mu_{i,g,t}$ into an endemic and epidemic component:

$$\mu_{i,g,t} = \exp \left\{ v_{i,g,t}^{END} + v_{i,g,t}^{EPI} \right\}, \quad (5)$$

where each part is parameterised as follows:

$$v_{i,g,t}^{EPI} = \theta_{AR(1)} \log(\tilde{y}_{i,g,t-1} + c) \quad (6)$$

$$\begin{aligned} v_{i,g,t}^{END} = & \theta_t + \theta_{gen} \mathbb{1}(x_{i,gen} = \text{"Male"}) + \theta_{age} \mathbb{1}(x_{i,age} = \text{"36-59"}) \\ & + \theta_{age:gen} \mathbb{1}(x_{i,gen} = \text{"Male"}) \cdot \mathbb{1}(x_{i,age} = \text{"36-59"}) + \theta_{t,gini} x_{i,t-1,gini} \\ & + \theta_{t,sp} x_{i,t-1,sp} + f_{coord}(x_{i,coord}) + f_{soc}(x_{i,soc}) + a_i + b_i \mathbb{1}(t = T) + \log(x_{i,g,pop}). \end{aligned} \quad (7)$$

We include a first-order autoregressive term of this rate, since path dependencies and self-exciting behaviour are common with infectious diseases and should therefore be accounted for Held et al. (2005). In addition, we transform the respective term by $h(x) = \log(x + c)$ to bypass problems with absorbing states of the implied counting process when $\tilde{y}_{i,g,t-1} = 0$. The value $c \in (0, 1]$ is estimated from the data. More general types of these autoregressive models are proposed by Zeger and Qaqish (1988).

As is evident from Equation (5), we constitute that both the epidemic and endemic components have a multiplicative effect on the observed infection rates. As an alternative, Held et al. (2005) replace the log link by an identity link, although Fokianos et al. (2020) argue for the logarithmic link implied in Equation (5) if additional covariates are available. They further derive theoretical properties, such as ergodicity, in the case of Poisson-distributed target variables under the condition $\theta_{AR(1)} < 1$.

Time-varying effects: For the endemic part (7), the temporal trend is reflected by piecewise constant fixed effects separately for each week, θ_t . By means of group-specific covariates we control for gender- and age-related effects and their interaction, θ_{gen} , θ_{age} and $\theta_{age:gen}$ (Walter & McGregor, 2020). The principal covariates, Gini Index and Percentage Staying Put, are modelled by piecewise constants in each week for maximal flexibility. To account for the stylised fact, that the incubation period, that is, the time between being infected and symptom onset, for COVID-19 is around 5 days (Li et al., 2020a), we lag the information on Gini Index and Percentage Staying Put by one week as indicated in Equation (7).

Isotropic smooth effects: The bivariate functions $f_{coord}(x_{i,coord})$ and $f_{soc}(x_{i,soc})$ display the effects of geographical coordinates and social coordinates on the incidence rate. To properly incorporate $x_{i,coord}$ and $x_{i,soc}$ in our regression framework, we propose the usage of isotropic splines. These kind of flexible functions were proposed by Duchon (1977) to model multiple covariates by a multivariate term as an alternative to anisotropic tensor products. Isotropic smoothers have the property of giving the identical predictions of the response under arbitrary rotation and reflection of the respective covariates (Wood, 2017). This characteristic is commonly reasonable when working with geographical coordinates $x_{i,coord}$ and in accordance with the uniqueness of the multidimensional scaling results, thus also for $x_{i,soc}$. With respect to the form of the smooth terms, we follow Wood (2003) and use a low-rank approximation of the thin-plate splines introduced in Duchon (1977). To obtain a smooth fit, we impose a penalty that is controlled by τ_{soc} and τ_{coord} for the respective isotropic splines.

Random effects: Because super spreader events such as carnival sessions (Streeck et al., 2020) or local outbreaks in major slaughterhouses (Dyal et al., 2020) lead to unobserved heterogeneities, our model comprises two district-specific Gaussian random effects. The random effect a_i governs long-term heterogeneities, while short-term dependencies, that is, sudden locally confined outbreaks as visible in the last week of Figure 1, are captured by b_i . We assume $a = (a_1, \dots, a_n)^T \sim N(0, I_n \tau_a^2)$ and $b = (b_1, \dots, b_n)^T \sim N(0, I_n \tau_b^2)$. Relying on the duality

between semiparametric regression and random effects (Ruppert et al., 2003), we can equivalently write the random effects as semiparametric terms. Hence we may replace a_i and b_i by $f_a(i) = (a_1, \dots, a_n)^\top X_a$ and $f_b(i) = (b_1, \dots, b_n)^\top X_b$, respectively, and introduce a ridge penalty for each coefficient vector. In this context, the design matrices X_a and X_b each consist of n dummy variables indicating to which district a specific observation refers. As a result of this reformulation, we can estimate the additional parameters τ_a and τ_b as tuning parameters in semiparametric regression (see Annex B for further information).

Modelling rates via count regression: Effectively, we model the rate of infections by including the term $\log(x_{i,g,pop})$ as an offset in Equation (7) since the infections rates $\tilde{Y}_{i,g,t}$ relate to the counts through $Y_{i,g,t} = \tilde{Y}_{i,g,t} x_{i,g,pop}$ via (note the slight abuse of notation as we here do not regard the infection rate among 10.000 inhabitants but the percentage of people infected with a disease onset in a specific week). As a byproduct, we implicitly assume that the entire population is susceptible, which is reasonable when considering the low prevalence of COVID-19 in Germany during the first wave. However, the model is still applicable in the later stages of the pandemic by replacing this offset with the number of susceptible inhabitants in each region.

3.3 | Estimation

At first, we propose an estimation procedure for the imputation model from Section 3.1. Given a partially imputed data set, we specify how to get estimates for the infection model from Section 3.2. Finally, the multiple imputation scheme combining both approaches is presented. Generally, we carry out all computations conditional on the observations in $t = 1$, that is, the week between the 3rd and 9th of March.

Imputation model: We get estimates for the imputation model through maximising the likelihood function resulting from Equation (3). As mentioned in Section 3.2, we can rewrite all random effects as smooth terms and penalise this likelihood to obtain smooth functions. By repeatedly updating the estimators through a backfitting algorithm, we optimise this objective (see Rigby & Stasinopoulos, 2005 for details). This procedure is readily implemented in the software package `gamlss` (Stasinopoulos et al., 2020).

Infection model: The infection model is characterised by the parameters c and θ , relating to the log-transformation of the autoregressive component and all other parameters. Given a partially imputed data set, we first consider θ to be a nuisance parameter and find c via a profile likelihood approach. Here the profile likelihood is given by

$$\mathcal{L}_{Profile}(c) = \max_{\theta} \mathcal{L}(c, \theta) = \mathcal{L}(c, \hat{\theta}(c)),$$

where $\mathcal{L}(c, \theta)$ is the joint likelihood resulting from Equation (4) and $\hat{\theta}(c)$ is the maximum likelihood estimator of θ for a fixed value of c . For any c , we can find $\hat{\theta}(c)$ by carrying out the estimation as explained in Annex B, hence it is straightforward to evaluate $\mathcal{L}_{Profile}(c)$. Building on this result, we use standard optimisation software, that is, the `optimise` routine within the software environment `R` (R Core Team, 2020), to obtain $\hat{c} = \arg \max_c \mathcal{L}_{Profile}(c)$. In the consecutive step, we fix c at \hat{c} to get $\hat{\theta}$ again by following Annex B.

Multiple imputation: Since information on the onset of symptoms is missing for approximately 30% of the cases, we proposed an imputation model in Section 3.1 to generate K partially imputed data sets. To correct the uncertainty quantification of the infection model for this multiple imputation procedure, we use the Rubin's rule. At first, we sample K imputed

data sets according to Section 3.1. Let $\hat{\vartheta}_{(k)} = (\hat{\theta}_{(k)}, \hat{c})$ be the resulting estimator from the two-stage maximum profile likelihood procedure explained in the previous paragraph given the partially imputed data set from the k th imputation step. By $\hat{V}_{(k)}$ we denote the corresponding variance estimate that results from Bayesian large sample properties (Wood, 2013). We then average the coefficients over all K iterations to obtain $\hat{\vartheta}_{MI} = \frac{1}{K} \sum_{k=1}^K \hat{\vartheta}_{(k)}$ and estimate its variance through:

$$\widehat{\text{Var}}(\hat{\vartheta}_{MI}) = \bar{V} + (1 + K^{-1})\bar{B},$$

where its components are given by

$$\bar{V} = \frac{1}{K} \sum_{k=1}^K \hat{V}_{(k)}$$

$$\bar{B} = \frac{1}{K-1} \sum_{k=1}^K (\hat{\vartheta}_{(k)} - \hat{\vartheta}_{MI})(\hat{\vartheta}_{(k)} - \hat{\vartheta}_{MI})^\top.$$

In our application, setting $K = 20$ proved to be sufficient since the estimates of different imputed data sets varied only marginally.

4 | RESULTS

We only report the findings of the infection model detailed in Section 3.2. A detailed analysis of the imputation model as well as a robustness check for the infection model can be found in the Supplementary Material.

4.1 | Temporal effect

To start, the estimate of θ_t is shown in Figure 5. The progression of the weekly estimates confirms generally decreasing infection rates over time. Due to the standardisation employed for the principal covariates in the analysis, the temporal trend can be interpreted as the log-transformed expected infection rate of female individuals aged between 15 and 35 in a district where the standardised Gini Index and Percentage Staying Put are zero. Since observing a zero in the standardised covariates translates to the mean observed values where we observed most information, the standard errors are also extremely narrow.

4.2 | Sociodemographic and epidemic effects

The linear time-constant estimates are given in Table 1 and exhibit in general a negative effect on male patients compared to female patients, 3% in the younger and 9.6% in the older age cohort.²

²One can derive these percentages by computing the expected multiplicative change that results from alternating the prediction from one to another demographic group. For instance, $\exp\{0.03\} \approx 0.97$, which is equivalent to a 3% decrease, is the multiplicative change *ceteris paribus* between females and males both aged between 15 and 35.

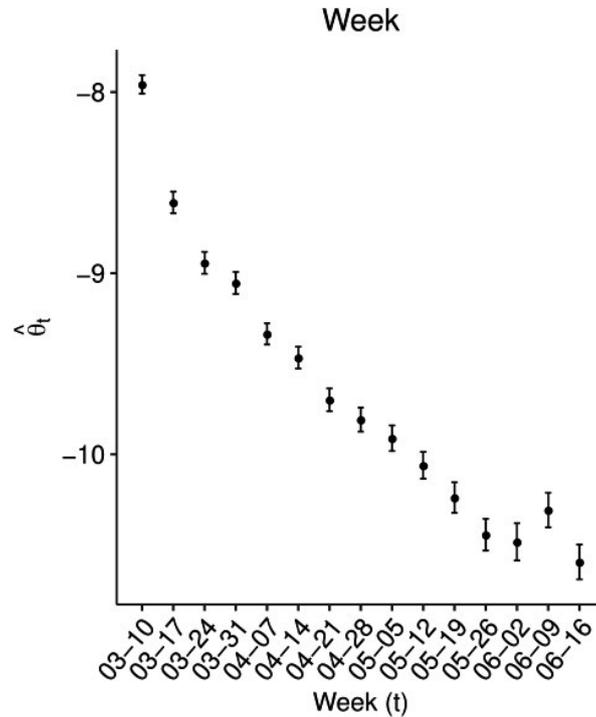


FIGURE 5 Estimate of temporal effect θ_t . The 95% confidence interval accompanies the estimates, and the shown dates (mm:dd) on the x-axis are the first days of the corresponding weeks

TABLE 1 Estimates of linear time-constant effects

Covariable	Estimate (standard error)	$\exp\{\text{Estimate}\}$ (standard error)
Male	-0.03 (0.015)	0.97 (0.014)
A35-A59	-0.031 (0.014)	0.969 (0.013)
Male: A35-A59	-0.071 (0.02)	0.931 (0.017)
$\log(\tilde{y}_{i,g,t-1} + c)$	0.623 (0.009)	1.865 (0.031)

Notes The reference group are female individuals aged between 15 and 35. By use of the delta rule, we approximated the standard errors of the transformed coefficients in the third row. The value c is estimated at 0.499 with a standard error of 0.027.

According to its partial effect, we also predict that the older age group has a lower infection rate than the younger group encompassing individuals aged between 15 and 35, for men 9.7% and women 3.1%. The autocorrelation coefficient $\theta_{AR(1)}$ expresses that one more infection among 10.000 inhabitants in a district during the past week almost doubles the predicted infections for the present week. This dominant finding confirms strong path dependencies in the data. In this context, we need to remark that the coefficients are partial effects that condition on all other

covariates used in the model. Therefore, a positive coefficient of a dummy variable does not necessarily translate to the same finding in the raw numbers.

4.3 | Mobility effects

The time-varying estimates regarding the relative mobility pattern are displayed in Figure 6. Overall, the estimated effects of the measures proposed in Section 2.2 on the rate of local COVID-19 infections are negative. In regards to relative importance, both variables rank similarly during the lockdown period that persists until early May. Subsequently, the Gini Index in a region gains weight, while the effect of People Staying Put becomes more volatile. The temporal changes of the respective estimates illustrate nonlinearities, which would not have been sufficiently captured by linear effects.

Gini index of co-location: Given all other covariates, Figure 6a suggests that inhabitants with meeting patterns that are centred around a few other districts entail reduced infection rates for a specific district. This tendency is only suspended in the week starting on March 17th during the early lockdown in Bavaria. The corresponding estimate is positive and significant. Right after the national lockdown on 22 March 2020, is ordered, the effect is not significantly different from zero for one week (03–24). The estimated effects remain low but negative until the German government introduces compulsory masks in public areas on 22 April (Mitze et al., 2020). Thereupon, the effect has a clear downwards tendency. Once policymakers slowly lift the lockdown measures, the estimate declines further until its maximum in the penultimate week of our observational period. This development may be viewed as evidence that a more focused attribution of co-location probabilities in a district becomes more crucial over time.

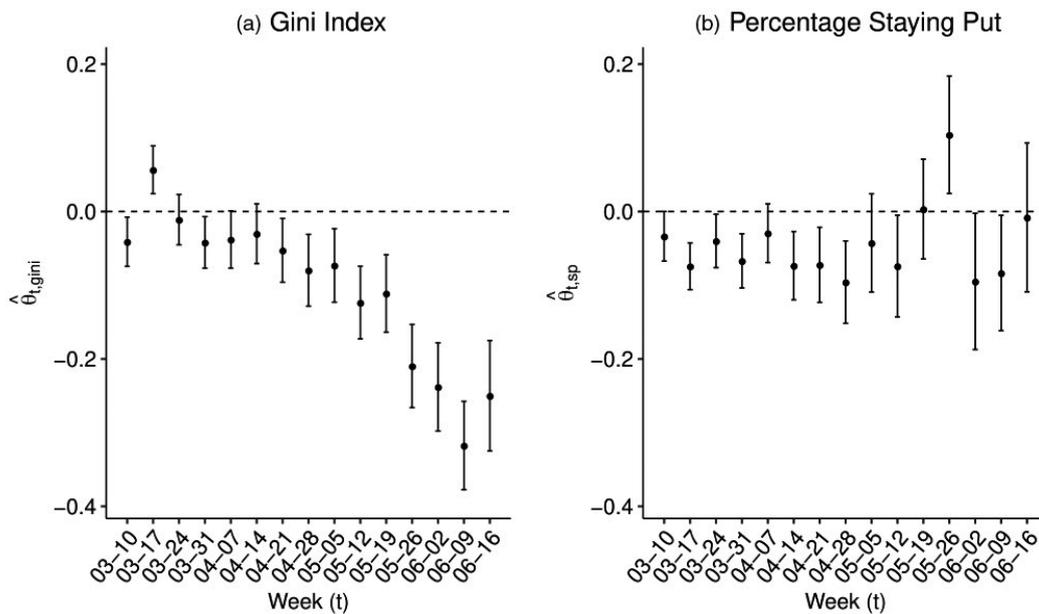


FIGURE 6 (a) Time-varying effects of the Gini index $\hat{\theta}_{t,gini}$. (b) Time-varying effects of the Percentage of People Staying Put $\hat{\theta}_{t,sp}$. The 95% confidence interval accompanies the estimates, and the shown dates (mm:dd) on the x-axis are the first days of the corresponding weeks

Percentage staying put: Suppose the percentage of inhabitants in a district staying put is large relative to the national tendency. In that case, we expect the incidence of infections throughout the lockdown period to be lower. We deduce this result from the largely negative estimates in Figure 6b for the weeks between 10 March and 12 May. Once the orders are relaxed, on the other hand, the standard errors of the respective covariate become relatively large, and the effect vanishes in the final week of the study. A possible explanation for this phenomenon is that when daily infections decline, most diseases are related to local outbreaks (as already mentioned in Section 2). These breakouts, in turn, cannot be associated with the percentage of people staying put. One exception to this finding is the estimate in the week starting on 26 May, where we encounter a significant positive effect.

4.4 | Spatial and social connectedness effects

In our model specification, we incorporate the friendship coordinates and geographical coordinates as two spatial effects. In combination with the two unstructured latent variables, we can disentangle separate influences on the local infection rates of spatial and friendship proximity as well as short- and long-term district-specific deviations from it.

Spatial effects: Let us start with the smooth spatial effect in Figure 7. Overall, the geographical effects within federal states, indicated by the black borders in Figure 7, are mostly heterogeneous. To give some examples, an almost uniformly augmented risk of infections is estimated in Baden-Württemberg and Thuringia. At the same time, we remark a negative spatial effect in Germany's northern districts, that is, Schleswig Holstein and Mecklenburg Western Pomerania. On the other hand, the fit for districts in North Rhine-Westphalia varies between positive, negative and no effect.

We visualise the result of the friendship coordinates in two manners. One may plot the smooth bivariate function in the friendship space, Figure 8a, or map the smooth fit on the geographical space, Figure 8b. The re-mapping allows for sharp edges in the geographical coordinates. Broadly, the fit differentiates between districts allocated in former East Germany (corresponding in Figure 8a to MDS coordinates located in the first quadrant) and former West Germany. We observe that the predicted infections are *ceteris paribus* lower if a district is situated in former East Germany. Districts allocated in the second and fourth quadrant of Figure 8a (mainly including districts from the states Bavaria, North Rhine-Westphalia and parts of Lower Saxony) are negatively affected by social proximity. Figure 8b demonstrates how the partial effects sometimes change abruptly between large cities and neighbouring districts. For instance, Berlin's central position is unrelated to the infection rates compared to the negative effect evaluated in Brandenburg. We observe a similar phenomenon for Hamburg when contrasting its partial effect with surrounding districts in Schleswig Holstein and Lower Saxony.

Unobserved heterogeneity effect: In Figure 9, the posterior modes of both random effects evince strong heterogeneities between districts and underpin local differences in the spread of COVID-19. Noticeable estimates of the long-term random effects, Figure 9a, reflect early outbreaks in the districts Greiz (Thuringia) and Coesfeld (North Rhine-Westphalia). Some estimates may also be related to heterogeneous testing practices between the districts.

We can trace back most high estimates of the short-term random effect to locally confined outbreaks, for instance, Guethersloh and Warendorf (North Rhine-Westphalia). As already stated in Section 2 the proportion of infections attributed to these local events rises once the general level of new cases declines. This result is supported by the different scales of the two types of random

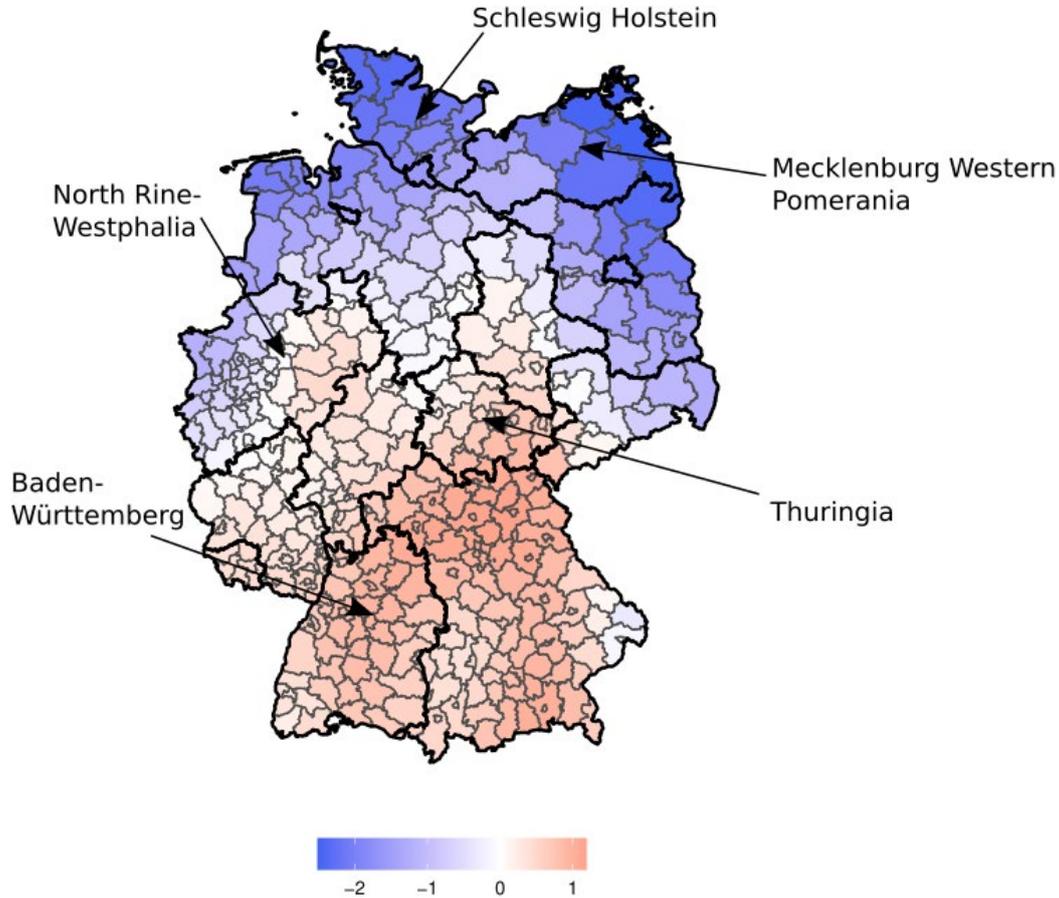


FIGURE 7 Estimated smooth spatial effect f_{coord} . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts. Through arrows, we highlight selected states mentioned in the text

effects and apparent in the estimates $\hat{\tau}_a = 0.2 < \hat{\tau}_b = 0.585$. Therefore, the posterior modes of the short-term effects exhibit higher variances and are larger in absolute terms than the long-term effects.

4.5 | Model assessment

We compare various alternative model specifications to check the robustness of our conclusions. In particular, we estimate separate models, adding dummy covariates for each state and leaving out one of the spatial terms, the Gini index, the Percentage of People Staying Put, all Facebook-related covariates and random effects. For this endeavour, we utilise the corrected Akaike information criterion (cAIC) introduced by Wood et al. (2016) since the effective degrees of freedom need to be adjusted for the additionally estimated variance components if random effects are included (we average the respective values over the results of all imputed data sets). The results in Table 2 support the appropriateness of our final model since the corresponding cAIC value is the lowest. Besides, the change in the cAIC value to the model (4), denoted by $\Delta cAIC$, permits

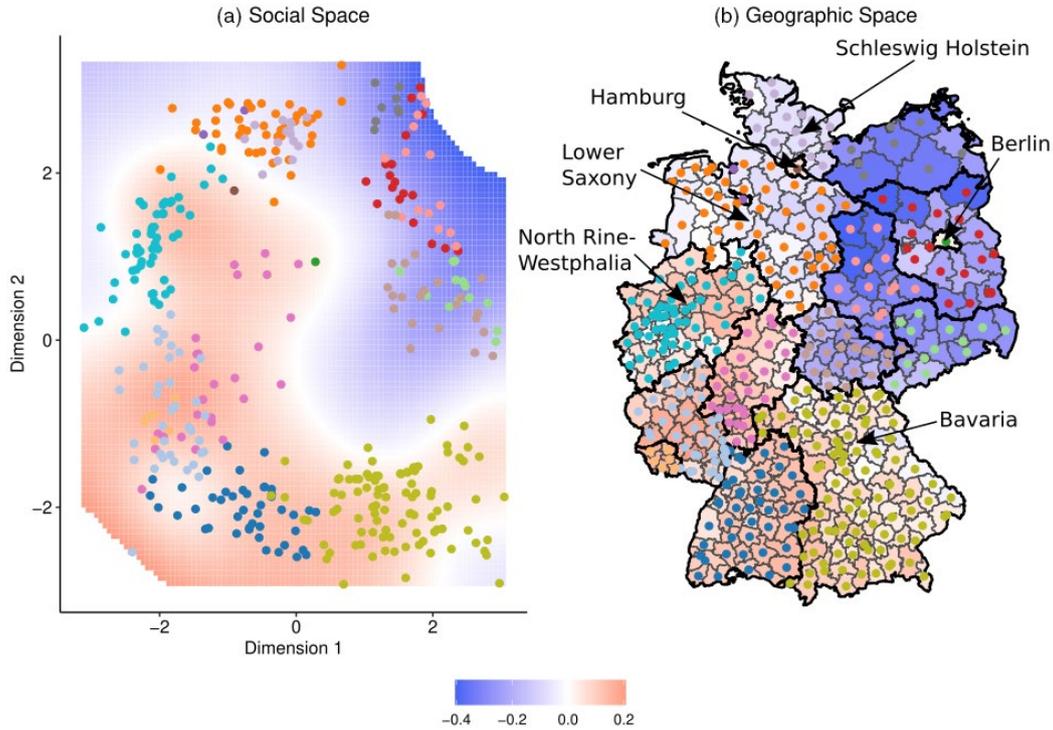


FIGURE 8 (a) Coordinates of the districts in the friendship space with the smooth partial effect of f_{soc} in the background. We only show the predictions in the range of observed values. (b) Coordinates of the districts in the geometric space with the smooth partial effect of f_{soc} again shown in the background for each district. The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts. Through arrows, we highlight selected states mentioned in the text

an evaluation of the variable importance of each eliminated covariate. We can conclude from Table 2 that the exclusion of the Gini Index induces the highest loss in cAIC value. Concerning the different types of distances, the friendship distance is more important than the geographical distance.

For further validation, we plot one draw of the randomised quantile residuals in Figure 10a. Dunn and Smyth (1996) proposed this type of residual based on the result that evaluating the cumulative distribution function at all observed values of $y_{i,g,t}$ under the estimated parameters should yield uniformly distributed random variables. Transforming these uniform values by the quantile function of the standard normal gives the quantile residuals. To obtain continuous residuals, the values are randomised since the negative binomial distribution in Equation (4) has discrete support. On average, the empirical quantiles are close to the theoretical expectations and do not indicate problems regarding the statistical fit. At the right tail of the distribution, 38 (out of 24.060) observations exhibit higher deviations from the normal quantiles, which we coloured in red. The underlying counts are mainly credited to local outbreaks that could not be completely captured by the random effects, namely Coesfeld (Thuringia), Cuxhaven (Lower Saxony), Aichach-Friedberg (Bavaria), Guetersloh and Warendorf (North Rhine-Westphalia). Additionally, we assess the predictions of the final model through plotting the predicted infections against the observed infections, Figure 10b, and a rootogram proposed by Kleiber and

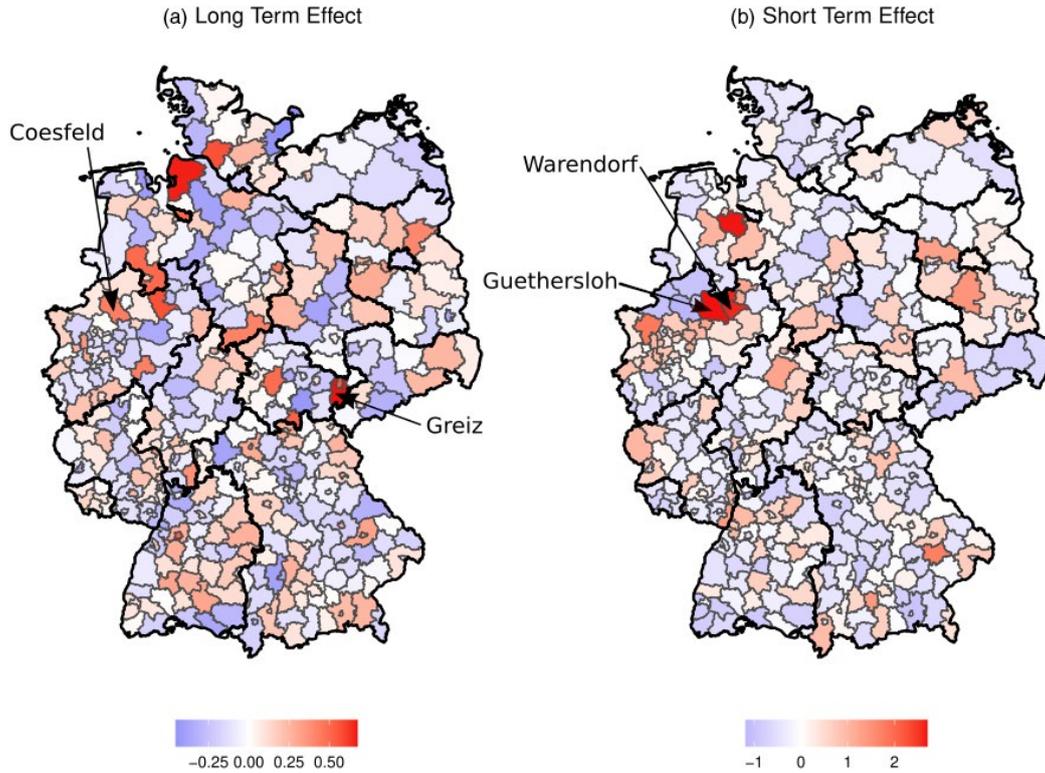


FIGURE 9 (a) Maximum posterior modes of the long-term random effects a_i . (b) Maximum posterior modes of the short-term random effects b_i . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts. Through arrows, we highlight selected districts mentioned in the text

TABLE 2 Alternative model specifications with resulting corrected Akaike information criterion (cAIC) value and change in corrected AIC value when compared to our model from Section 3

Model description	cAIC (Model)	Δ cAIC (Model)
Our model	86694	–
With state effect	86694.79	0.790
Without geographical distance	86699.42	5.422
Without friendship distance	86701.26	7.262
Without Age:Gender interaction	86707.34	13.336
Without percentage staying put	86732.46	38.461
Without Gini index	86974.32	280.319
Without Facebook covariates	87033.87	339.867
Without long-term effect	87452.62	758.620
Without short-term effect	87900.03	1206.034
Without long- and short-term effect	88624.38	1930.382

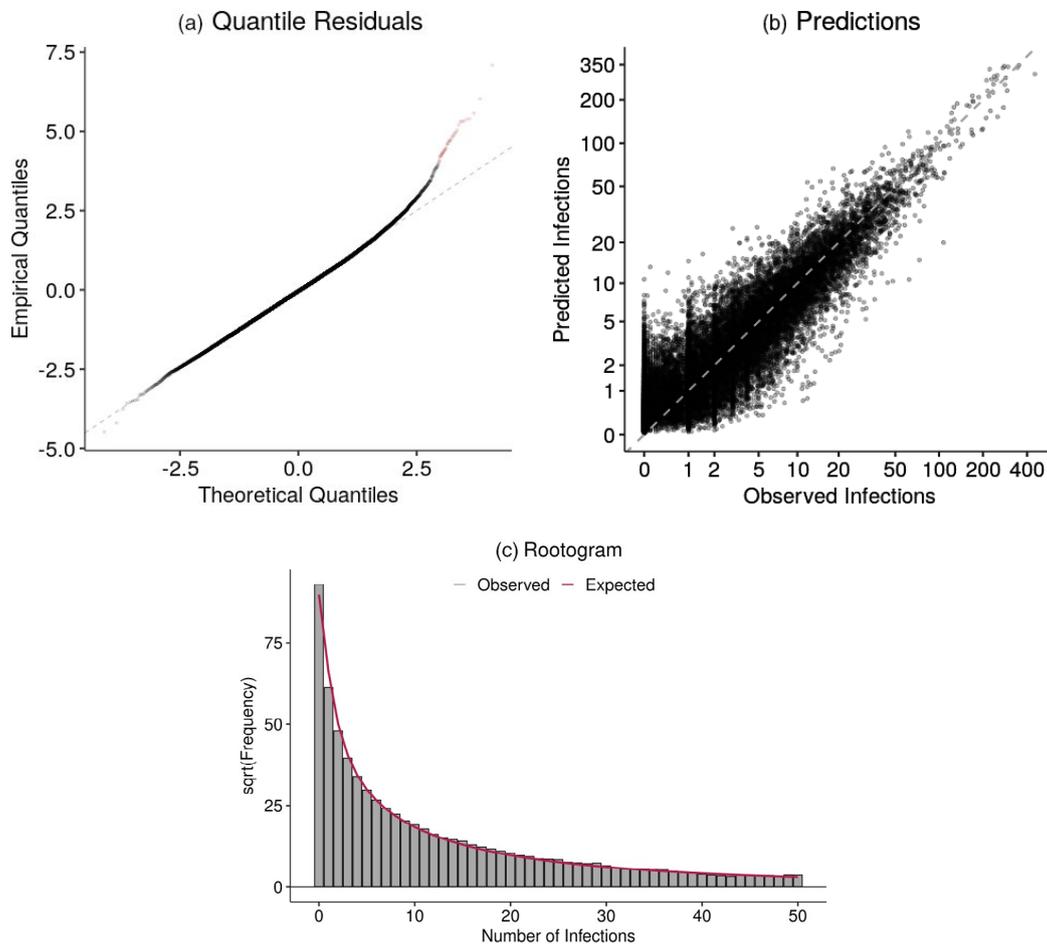


FIGURE 10 (a) QQ Plot of randomised quantile residuals, observations with a distance larger than 1 to the theoretically expected values are drawn in red. (b) Scatter plot of the observed and predicted infection count, for the x and y -axis, we used a $\log(\cdot + 1)$ scale. The dotted grey line is the best-case scenario of the prediction and has intercept 0 and slope 1. (c) Rootogram comparing the observed and expected counts. The grey barplot specifies the observed counts, while the red line gives the expected values under Equation (4)

Zeileis (2016), Figure 10c. Both visualisations confirm a strong fit of the presented model and proof that the model can sufficiently capture the observed counts of infected individuals. Due to the multiple imputation scheme specified in Sections 3.1 and 3.3, we carry the model assessment out for each imputation separately and report the averaged results.

5 | CONCLUSION

In this writing, our contributions are twofold. First, we used state-of-the-art regression models to quantify the importance of human mobility for understanding the spread of COVID-19 on a local level accounting for their temporal dynamic, latent effects and other covariates. Concerning the relative importance, the Gini index of meeting probability attribution proved to be a primary

driver of the infection rates. Second, we used methods from multivariate statistics to derive friendship coordinates for the federal districts in Germany. Consecutively, we coupled the result with a regression model via isotropic splines and, thereby, revealed a perpetual clustering of communities in former East- and West Germany that remains existent for COVID-19 infections because the social geographical system proves to be an essential regressor in our application. Moreover, our findings enable an evaluation of the district-wise policies undertaken between March and June 2020. The results corroborate the usefulness of interventions limiting trans-district movements and concentrating meeting patterns. Especially during the last weeks of this study, local lockdowns could mitigate further national outbreaks.

Still, we need to address some limitations of our work, which require additional investigation. The data sources for the infection data include all individuals in Germany that tested positive on COVID-19. During the peak phase in March, these tests were mainly carried out with patients who showed symptoms or had contact with an infected individual. Due to an unknown dark figure of infected persons missing in the public records (Lavezzo et al., 2020), the observed data are a proxy for the current epidemiological situation. To control for this possible bias, further research on the prevalence of COVID-19 in Germany and the representability of the official statistics of the real infection occurrence akin to the REACT Study in England (Riley et al., 2021) would be necessary.

Even with these caveats, the combination of infection, mobility and connectivity data can serve for a fruitful application of other methods as well. Contrasting our approach, one may tackle the regression task in Section 3 by incorporating the spatial dependencies directly in the correlation structure, as is done in the literature on spatial econometric models (LeSage & Pace, 2009). We could also employ novel clustering algorithms that naturally exploit different proximity dimensions, such as the geographical and social space, to identify similar districts while taking into account spatial dependencies (D'Urso & Vitale, 2020; D'Urso et al., 2019). Furthermore, the research questions posed in this article would greatly benefit from an examination through the lens of analytical sociology (Hedström & Bearman, 2011). Nevertheless, this type of analysis usually necessitates individual-level data, which are not readily available. Therefore, we can only verify some of the theoretical results of Block et al. (2020) on the macro scale, which does not necessarily translate to the micro scale (Stadtfeld, 2018). Therefore, additional empirical work on the implications of individual behaviour on the spread of COVID-19 is still needed. Nevertheless, our work can give valuable pointers in that regard contingent on the assumption that the corresponding district average adequately represents the mobility patterns of an individual.

6 | DATA AND CODE AVAILABILITY

Facebook collected the anonymised mobility and connectivity data. We cannot share the raw data due to a data agreement. Still, we are allowed to provide all data aggregated onto the level of federal districts. To guarantee the replicability of our results, we make the complete code to obtain the results from this article available online. We also supply a visualisation of the entire pipeline of our analysis in the Supplementary Material for transparency.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their careful reading and constructive comments. Further, we thank Alex Pompe from Facebook for his useful comments and explanations of the data from Facebook. The project was supported by the European Cooperation in Science and

Technology [COST Action CA15109 (COSTNET)]. This work has been funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibility for its content.

ORCID

Cornelius Fritz  <http://orcid.org/0000-0002-7781-223X>

REFERENCES

- Asadi, S., Bouvier, N., Wexler, A.S. & Ristenpart, W.D. (2020) The coronavirus pandemic and aerosols: does COVID-19 transmit via expiratory particles? *Aerosol Science and Technology*, 54(6), 635–638.
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. (2018) Social connectedness: measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–280.
- Block, P., Hoffman, M., Raabe, I.J., Beam Dowd, J., Rahal, C., Kashyap, R. et al. (2020) Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behavior*, 4, 588–596.
- Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F. et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117(27), 15530–15535.
- Borg, I., Groenen, P.J.F. & Mair, P. (2013) *Applied multidimensional scaling*. New York: Springer.
- Cailliez, F. (1983) The analytical solution of the additive constant problem. *Psychometrika*, 48(2), 305–308.
- Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J. et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223), 514–523.
- Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S. et al. (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400.
- Cho, E., Myers, S.A., & Leskovec, J. (2011) Friendship and mobility: user movement in location-based social networks. In: *KDD '11: proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*.
- Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8(2), 93–115.
- Cox, T. & Cox, M. (2000) *Multidimensional scaling*. Boca Raton: CRC Press.
- Cox, M. & Cox, T. (2008) Multidimensional scaling. In: Chen, C., Härdle, W.K., & Unwin, A. (Eds.) *Handbook of data visualization*. Berlin: Springer, pp. 315–347.
- D'Urso, P. & Vitale, V. (2020) A robust hierarchical clustering for georeferenced data. *Spatial Statistics*, 35, 1–33.
- D'Urso, P., De Giovanni, L., Disegna, M. & Massari, R. (2019) Fuzzy clustering with spatial-temporal information. *Spatial Statistics*, 30, 71–102.
- Duchon, J. (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Schempp, W. & Zeller, K. (Eds.) *Constructive theory of functions of several variables*. Berlin: Springer, pp. 85–100.
- Dunn, P.K. & Smyth, G.K. (1996) Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236–244.
- Dyal, J.W., Grant, M.P., Broadwater, K., Bjork, A., Waltenburg, M.A., Gibbins, J.D. et al. (2020) COVID-19 among workers in meat and poultry processing facilities—19 states, April 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(18), 557–561.
- Eilers, P.H. & Marx, B.D. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–102.
- Facebook (2020) GeoInsights Help. Available from <https://www.facebook.com/help/geoinsights>.
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H. et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584, 257–261.
- Fokianos, K., Støve, B., Tjøstheim, D. & Doukhan, P. (2020) Multivariate count autoregression. *Bernoulli*, 26(1), 471–499.

- Galeazzi, A., Cinelli, M., Bonaccorsi, G., Pierri, F., Schmidt, A.L., Scala, A. et al. (2020). Human mobility in response to COVID-19 in France, Italy and UK. *arXiv preprint*.
- Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X. et al. (2020). Clinical characteristics of Coronavirus disease 2019 in China. *New England Journal of Medicine*, 382(18), 1708–1720.
- Günther, F., Bender, A., Katz, K., Küchenhoff, H. & Höhle, M. (2020) Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal* (OnlineFirst).
- Hedström, P. & Bearman, P. (2011) What is analytical sociology all about? An introductory essay. In: Bearman, P. & Hedström, P. (Eds.) *The Oxford handbook of analytical sociology*. Oxford: Oxford University Press.
- Held, L. & Paul, M. (2012) Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54(6), 824–843.
- Held, L., Höhle, M. & Hofmann, M. (2005) A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5(3), 187–199.
- Held, L., Meyer, S. & Bracher, J. (2017) Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. *Statistics in Medicine*, 36(22), 3443–3460.
- Holtz, D., Zhao, M., Benzell, S.G., Cao, C.Y., Rahimian, M.A., Yang, J. et al. (2020) Interdependence and the cost of uncoordinated responses to COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, 117(33), 19837–19843.
- Iyer, S., Karrer, B., Citron, D., Kooti, F., Maas, P., Wang, Z. et al. (2020) Large-scale measurement of aggregate human colocation patterns for epidemiological modeling. *medRxiv preprint*.
- Kang, D., Choi, H., Kim, J.H. & Choi, J. (2020) Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, 94, 96–102.
- Kimeldorf, G.S. & Wahba, G. (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2), 495–502.
- Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H. & Lipsitch, M. (2020) Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493), 860–868.
- Kleiber, C. & Zeileis, A. (2016) Visualizing count data regressions using rootograms. *The American Statistician*, 70(3), 296–303.
- Kottasová, I. (2020) Germany reports 650 new coronavirus cases in a meat processing plant - CNN. Available from <https://edition.cnn.com/2020/06/18/europe/germany-meat-processing-plant-coronavirus-cases-intl>.
- Kraemer, M.U., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M. et al. (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497.
- Kuchler, T., Russel, D. & Stroebel, J. (2021) The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics* (OnlineFirst).
- Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., Barzon, L., Del Vecchio, C. et al. (2020) Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*, 584, 425–429.
- LeSage, J. & Pace, R.K.P. (2009) *Introduction to spatial econometrics*. Boca Raton: CRC Press.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y. et al. (2020a) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. et al. (2020b) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 368(6490), 489–493.
- Little, R.J.A. & Rubin, D.B. (2002) *Statistical analysis with missing data*. New York: Wiley.
- Lorch, L., Trouleau, W., Tsirtsis, S., Szanto, A., Schölkopf, B. & Gomez-Rodriguez, M. (2020) A spatiotemporal epidemic model to quantify the effects of contact tracing, testing and containment. *arXiv preprint*.
- Maas, P., Shankar, I., Gros, A., Wonhee, P., McGorman, L., Nayak, C. et al. (2019) Facebook disaster maps: aggregate insights for crisis response & recovery. In: *Proceedings of the 16th ISCRAM conference*, pp. 1–12.
- Mardia, K.V. (1978) Some properties of classical multi-dimensional scaling. *Communications in Statistics-Theory and Methods*, 7(13), 1233–1241.
- Meyer, S. & Held, L. (2017) Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2), 338–351.
- Mitze, T., Kosfeld, R., Rode, J. & Walde, K. (2020) Face masks considerably reduce COVID-19 cases in Germany. *Proceedings of the National Academy of Sciences of the United States of America*, 117(51), 32293–32301.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D. et al. (2020) vegan: community ecology package. R package version 2.5-7.

- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Delataille, S., De Nadai, M. et al. (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6(23), 1–6.
- Paul, M., Held, L. & Toschke, A.M. (2008) Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, 27(29), 6250–6267.
- Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N. et al. (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5), e261–e270.
- R Core Team. (2020) R: a language and environment for statistical computing.
- Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554.
- Riley, S., Ainslie, K.E.C., Eales, O., Walters, C.E., Wang, H., Atchison, C. et al. (2021) Resurgence of SARS-CoV-2: detection by community viral surveillance. *Science*, 372(6545), 990–995.
- Ruppert, D., Wand, M. & Carroll, R.J. (2003) *Semiparametric regression*. Cambridge: Cambridge University Press.
- Stadtfeld, C. (2018) The micro–macro link in social networks. *Emerging Trends in the Social and Behavioral Sciences* (OnlineFirst).
- Stasinopoulos, M., Rigby, B., Voudouris, V. & Kiose, D. (2020) gamlss.add: extra additive terms for generalized additive models for location scale and shape. R package version 5.1-6.
- Streeck, H., Schulte, B., Kuemmerer, B., Richter, E., Hoeller, T., Fuhrmann, C. et al. (2020) Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *Nature Communications*, 11(5829), 1–12.
- Walter, L.A. & McGregor, A.J. (2020) Sex-and gender-specific observations and implications for COVID-19. *Western Journal of Emergency Medicine*, 21(3), 507–509.
- WHO (2020) Coronavirus disease (COVID-2019) situation reports. Available from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- Wood, S.N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 95–114.
- Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1), 3–36.
- Wood, S.N. (2013) On p -values for smooth components of an extended generalized additive model. *Biometrika*, 100(1), 221–228.
- Wood, S.N. (2017) *Generalized additive models: an introduction with R*. Boca Raton: CRC Press.
- Wood, S.N., Pya, N. & Säfken, B. (2016) Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.
- Young, G. & Householder, A.S. (1938) Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3(1), 19–22.
- Zeger, S.L. & Qaqish, B. (1988) Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, 44(4), 1019–1031.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Fritz, C., Kauermann, G. (2021) On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 00, 1–25. <https://doi.org/10.1111/rssa.12753>

APPENDIX A

A. MULTIDIMENSIONAL SCALING AND PROCRUSTES ANALYSIS

In order to determine the information given in the pairwise social connectedness indices $x_{soc} = (x_{ij,soc})_{i,j=1,\dots,n}$ for explaining the spread of COVID-19 in Germany, we use techniques from multivariate statistics (Cox & Cox, 2000). Thereby, we can derive a low-dimensional representation of the network on the actor level and guarantee interpretable as well as transparent results. More specifically, we apply *metric multidimensional scaling* (MDS) to represent dissimilarity matrices in a lower-dimensional geometric space that preserves the dissimilarities through Euclidean distances (Borg et al., 2013). To illustrate the application of this algorithm, one can think of MDS as a technique to reverse-engineer geographical coordinates that are unique up to scale and rotation from distances between cities (Young & Householder, 1938).

At first, we transform the similarities expressed by the counts of friendship ties between the districts x_{soc} to dissimilarities. In our application, the measure of dissimilarity is given by $d_{soc} = (\frac{1}{x_{ij,soc}})_{i \neq j=1,\dots,n}$ and $d_{ii,soc} = 0$. While this dissimilarity matrix is symmetric and nonnegative, there is no general guarantee that the entries of d_{soc} are Euclidean. Therefore, we add the constant c to the off-diagonal elements to ensure that the distances between the found coordinates are Euclidean (Cailliez, 1983; Mardia, 1978). In order to estimate these p -dimensional coordinates $x_{i,soc} = (x_{i,1}, \dots, x_{i,p}) \forall i = 1, \dots, n$ from the dissimilarity matrix d_{soc} , the objective is to minimise the squared error between the pairwise entries of d_{soc} and the Euclidean distances calculated with the respective coordinates:

$$x_{soc} = \left(x_{1,soc}^\top, \dots, x_{n,soc}^\top \right) = \underset{\tilde{x} \in \mathbb{R}^{p \times n}}{\operatorname{argmin}} \left(\sum_{i \neq j} (d_{ij,soc} + c - \|\tilde{x}_i - \tilde{x}_j\|^2) \right)^{1/2}, \quad (\text{A1})$$

in our case we set $p = 2$. See Cox and Cox (2000) and Borg et al. (2013) for methods to find x such that Equation (A1) holds, which are implemented in the R-package `stats` (R Core Team, 2020).

Since arbitrary transformations, rotations and reflections of any coordinates that optimise (A1), represented by $x_{soc} = (x_{1,soc}, \dots, x_{n,soc})$, are equally valid, we further process the solution to guarantee uniqueness and an intuitive understanding of the result. To achieve this goal, we use *Procrustes Analysis* (Cox & Cox, 2008) and find an optimal solution x_{soc} to Equation (A1) that is also most similar to the geographical coordinates $x_{coord} = (x_{1,coord}, \dots, x_{n,coord})$ given in Figure 8. As a measure of similarity between the matrices x_{soc} and x_{coord} , commonly $R^2 = \sum_{i=1}^n (x_{i,soc} - x_{i,coord})^\top (x_{i,soc} - x_{i,coord})$ is used. Furthermore, we can parameterise the desired class of functions that transform an according to Equation (A1) optimal solution $x_{soc,i}$ to $\tilde{x}_{soc,i}$ by:

$$\tilde{x}_{soc,i} = \rho \mathcal{A}^\top x_{soc,i} + b, \quad (\text{A2})$$

where ρ is scalar determining the dilation, \mathcal{A} an orthogonal matrix defining the rotation and reflection, and b a two-dimensional vector for a possible translation. From an optimisation point of view, we now have to find ρ , \mathcal{A} , and b such that the resulting R^2 is minimised, which we can do in closed form (see Cox & Cox, 2000). This type of transformation is implemented in the R-package `vegan` (Oksanen et al., 2020) and does not change the estimates or inference because we apply isotropic smooth terms.

B. ESTIMATION OF θ GIVEN c AND COMPLETE DATA

From Equation (4), we construct a likelihood for each district and age/gender group tuple. Combining these separate contributions under independence leads to a joint logarithmic likelihood given by:

$$\ell(\theta, c) \propto \sum_{i=1}^n \sum_{g \in \mathcal{G}} \sum_{t=1}^T \log \left(\frac{\Gamma(\phi + y_{i,g,t})}{y_{i,g,t}! \Gamma(y_{i,g,t})} \right) + \phi \log \left(\frac{\phi}{\phi + \mu_{i,g,t}} \right) + y_{i,g,t} \log \left(\frac{\mu_{i,g,t}}{\phi + \mu_{i,g,t}} \right). \quad (\text{B1})$$

note that ϕ is the dispersion parameter of the negative binomial distribution and that the likelihood of the imputation model from Equation (3) in Section 3.1 has the same form with $\phi^{-1} = \sigma_l$. Suppose we plug $\mu_{i,g,t}$ as defined in Equation (5) into (B1) and fix the value of c . In that case, we observe that the result is a function of θ and resembles the likelihood of a generalised additive model with negative binomial distributed target variables and denote the likelihood by $\ell(\theta|c)$ (Ruppert et al., 2003). To obtain a smooth fit of θ , we extend this function by an additive penalisation component:

$$\ell_p(\theta|c) = \ell(\theta|c) - \tau^\top S, \quad (\text{B2})$$

where $\tau = (\tau_a, \tau_b, \tau_{coord}, \tau_{soc})^\top$ are smoothing parameters weighting the term-specific penalties $S = (S_a, S_b, S_{coord}, S_{soc})^\top$. The choice of these penalties differs between the random effects and bivariate spacial effects. For the random effects, we follow Ruppert et al. (2003) and define S_a and S_b through ridge penalties, hence, for instance, $S_a = \sum_{i=1}^2 a_i^2$. In the case of the isotropic semiparametric terms, we chose the penalty terms in accordance with Duchon (1977). Here, S_{coord} penalises the roughness of the bivariate function $f_{coord}(x_{i,coord}) = f_{coord}(x_{i,coord,1}, x_{i,coord,2})$, where $x_{i,coord,p}$ denotes the p th dimension of $x_{i,coord} \forall p \in \{1, 2\}$, in our application the longitude and latitude of district i . Given this notation, we can state the functional form of the penalty term:

$$S_{coord} = \int \frac{\partial^2}{\partial^2 x_{coord,1}} f_{coord}(x_{coord})^2 + 2 \frac{\partial^2}{\partial x_{coord,1} \partial x_{coord,2}} f_{coord}(x_{coord})^2 + \frac{\partial^2}{\partial^2 x_{coord,2}} f_{coord}(x_{coord})^2 dx_{coord,1} dx_{coord,2}.$$

Besides we ensure identifiability of all smooth effects by incorporating a sum-to-zero constraint per term, which translates to $\sum_{i=1}^n f_{coord}(x_{i,coord}) = 0$ for $f_{coord}(\cdot)$ (Wood, 2017).

To maximise (B2) in terms of θ and τ , we follow the nested optimisation approach of Wood (2011). Hence, we find $\hat{\tau}$ in an outer iteration and $\hat{\theta}$ consecutively in an inner iteration. Generally, the validity of this procedure rests on the finding that $\hat{\theta}$ is the posterior mode of $\theta|y$ under the assumption that θ follows a zero-mean normal prior with improper variance (Kimeldorf & Wahba, 1970). Viewing θ as random coefficients enables us to estimate all smoothing parameters τ via restricted maximum likelihood estimation. More specifically, we set up $f(y, \theta|c)$ given $\ell(\theta|c)$ and $f(\theta)$. Through integrating θ out of $f(y, \theta|c)$ by deploying a Laplace approximation we obtain an approximate REML criterion, which is a function of τ and ϕ , the dispersion parameter from Equation (B1). Maximising the derived function in terms of these parameters gives $\hat{\tau}$ and $\hat{\phi}$ (see Wood, 2011 for additional details). Given the tuning parameters, we consecutively find $\hat{\theta}$ through standard penalised iterative re-weighted least squares estimates (PIRLS, Wood, 2017) in the inner iteration. We repeat this iterative scheme until convergence to obtain $\hat{\theta}$ and $\hat{\tau}$ given a fixed value of c . A scalable implementation of this routine that we used is available in the software package `mgcv` (Wood, 2017).

On the Interplay of Regional Mobility, Social Connectedness, and the Spread of COVID-19 in Germany

Cornelius Fritz | Göran Kauermann
Department of Statistics, LMU Munich

CONTENTS

1	Analysis of Representativeness of Facebook Data	2
2	Pipeline of the Analysis	3
3	Imputation of the Test Delay	4
3.1	Stochasticity of Missings	4
3.2	Missing at Random Analysis	6
3.3	Results	6
3.4	Sensitivity Analysis of Disease Onset Imputation	9
4	Alternative Infection Model	9

1 | ANALYSIS OF REPRESENTATIVENESS OF FACEBOOK DATA

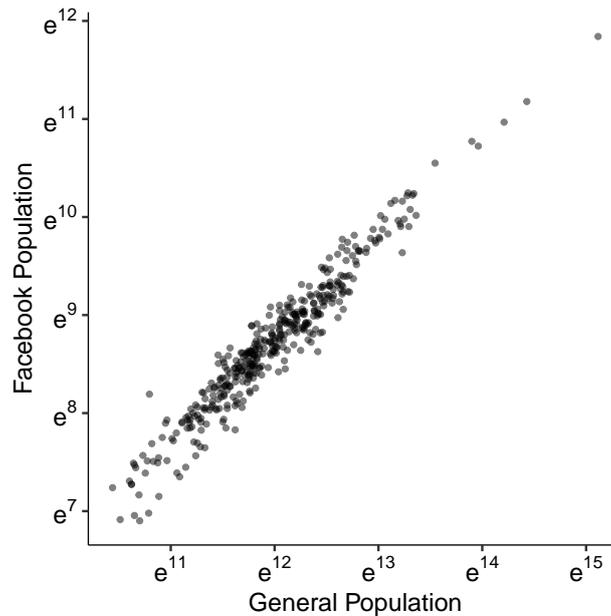


FIGURE 1 Scatter-plot of the Facebook population that opted in the geolocation features and the general population for each district.

To guarantee robust estimates, we investigate to what extent the mobility data provided by Facebook in the context of the *Data for Good* program (Maas et al., 2019; Iyer et al., 2020) are representative of the general German population and in line with other mobility data. We carry out the respective assessment in three parts:

1. Compare the spatial distribution of Facebook users that opted in geolocation features with the general population for each district.
2. Contrast the age structure of Facebook users with the demographic pyramid in Germany.
3. Check if mobility data from other providers, i.e., Apple and Google, measure similar information as the Facebook data.

Spatial Distribution: Besides mobility data, the datasets made available by Facebook also include the daily counts of Facebook users who enabled geolocation features in each federal district. For this assessment, we only look at the average of users per district over the study period from the 3rd of March to the 16th of June 2020 and calculate the same quantity for the general population provided by the German Federal Statistical Office. Figure 1 is a scatter-plot of 401 points representing the general and Facebook population measured for each federal district. We can conclude from this plot that the spatial distribution of Facebook users is positively correlated with official statistics, i.e., the estimated Pearson correlation is 0.98.

Age Representation: In addition to people's spatial distribution, we want to compare the age structure of general Facebook users and the general German population. The Facebook users' age structure that turned on the geolocation features is unknown due to differential privacy. As a possible proxy for this information, we compile information about the age structure of the entire

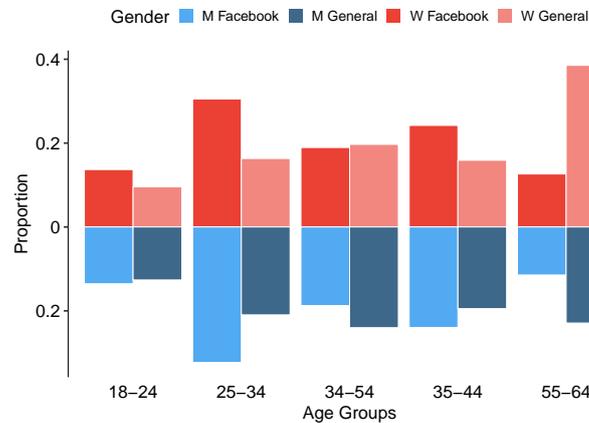


FIGURE 2 Comparison of the age- and gender-structure within Facebook and in the general population of Germany.

Facebook population in Germany from Facebook Audience Insights. Unfortunately, Facebook's age-groups are not aligned with the age groups used by the RKI. As can be seen in Figure 2, the Facebook population is younger than the general population. While we can conclude that the age groups 18-24, 35-54 are adequately represented in the Facebook sample, there is a surplus of people in the 25-to-34-year-old cohort. Simultaneously, the age group of the oldest individuals is in relative terms less populated than the general population. Still, this finding does not necessarily indicate a bias in the mobility data that we use in the principal analysis. Due to the standardisation of the mobility-related covariates given in Formula (1) of the main paper, the covariates are robust to a nation-wide under- and over-representation of older and younger individuals. In that case, the measurement bias would only be a constant, and hence the standardised covariates would be the same. We cannot quantify with the available data whether the percentage compositions of those age structures vary within each district. However, the promising results of the spatial representativeness make it reasonable to assume that this is not the case.

Mobility Measurement from other Sources: We now contrast the mobility data provided by two other major technology companies, namely Apple¹ and Google², with the data made available by Facebook. We use the Google Community Mobility Reports, including relative mobility trends in retail or grocery stores, transit stations, and places of work. Similarly, Mobility Trends Reports from Apple use information on the relative requests for directions to walk, drive or use transit transportation in Apple Maps to measure the mobility trends. Since solely the Percentage of People Staying Put relates to absolute movements in the Facebook data, we investigate the Pearson correlation matrix of all given mobility indices in Figure 3. The high correlations in the first row and column are conclusive in that the Percentage of People Staying Put captures the information of all other variables reasonably well since the absolute correlation coefficients range from 0.88 to 0.97.

2 | PIPELINE OF THE ANALYSIS

To make the performed analysis of the main article as transparent as possible, Figure 4 depicts the complete pipeline, including all stages needed for this work. Information on the raw data and needed pre-processing steps are given in Section 2 and Annex A. The imputation and infection models, together with Ruben's formula to pool estimates from different imputations, are specified

¹<https://covid19.apple.com/mobility>

²<https://www.google.com/covid19/mobility/>

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

4

FRITZ AND KAUEMANN

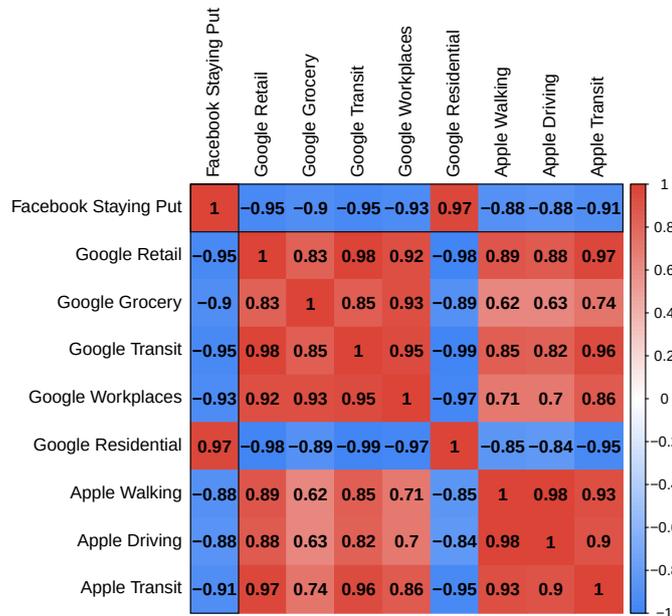


FIGURE 3 Correlation matrix of the weekly national mobility indicators from Facebook, Google, and Apple.

in Section 3 (all referred content relates to the main article).

3 | IMPUTATION OF THE TEST DELAY

The imputation procedure given in the main article relies on all observations to be missing at random (Little and Rubin, 2002). To check this assumption, we first argue for stochasticity of the mechanism driving the binary indicator whether the target variable, i.e., the date of disease onset, is missing. In a second step, we perform a missing at random analysis to discover which covariates affect this mechanism. We consecutively employ the detected covariates in the imputation procedure introduced in Section 2.3 of the main article and report the full estimates. Finally, we conduct a sensitivity analysis to check whether our findings change if we only regard cases with observed disease onset.

3.1 | Stochasticity of Missings

Each registered COVID-19 case can be characterised by three dates representing distinct stages of the disease, namely the time of infection, disease onset, and registration at local health authorities. This progression is illustrated in Figure 5. While we do not observe the disease onset date for some cases, we can still argue that all samples progress through the same three stages, although the second date is in some cases latent. Hence under the assumption that the incubation period from the infection date to the (observed or latent) disease onset date is stochastic, it is also appropriate to assume that the same holds for the test delay,

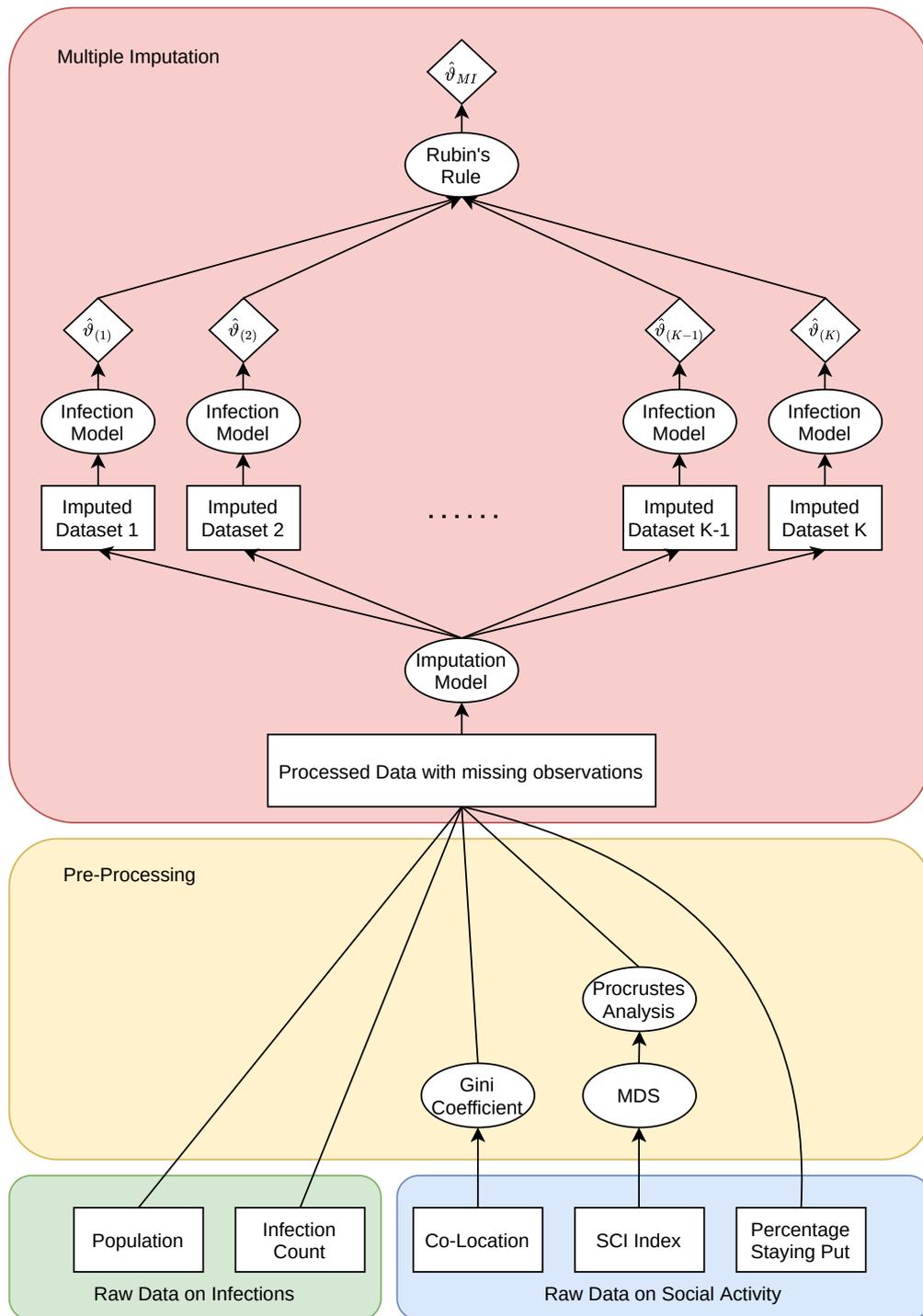


FIGURE 4 Visualisation of the complete pipeline from pre-processing raw data to obtaining the final estimates. Rectangles relate to data, ovals to procedures, and tilted squares to estimators.

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

6

FRITZ AND KAUEMANN

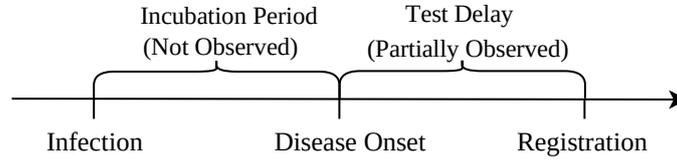


FIGURE 5 Illustrative temporal path of a COVID-19 infection, collected surveillance data and resulting delays.

	Time Trend	Age	Gender	Age:Gender	District	State	cAIC	Rank
Model 1	✓	✓	✓	✓	✓	✓	121608.4	1
Model 2	✓	✓	✓	✓	✓		121610.7	2
Model 3	✓	✓	✓	✓		✓	142637.8	5
Model 4	✓	✓	✓		✓	✓	126562	7
Model 5	✓	✓		✓	✓	✓	121632.1	6
Model 6	✓		✓	✓	✓	✓	121741.9	4
Model 7		✓	✓	✓	✓	✓	121731.1	3

TABLE 1 Different specifications of the missing at random process. The sign ✓ signals the inclusion of a specific covariate in the respective model. In the last two rows the corresponding corrected AIC (cAIC) values and their rankings are given.

defined as the time-span between disease onset and registration date. Given this argument, it is legitimate to impute the test delay through a stochastic model. We can then use this information to project the disease onset date back from the registration date for all cases given the observed data. Besides, we carry out a sensitivity analysis in Section 3.4 to check for structural differences in the dynamics of infection rates of cases with a disease onset and all cases. The estimates excluding all observations with missing disease onset date are in accordance with our findings of the main article.

3.2 | Missing at Random Analysis

Having established that the missing values are random, we need to identify the covariates driving the missing values' mechanism. To do that, we generate a binary indicator of whether the test delay was observed for each case i . In the consecutive step, we take this binary indicator to be the target variable of a logistic regression with seven different sets of covariates given in Table 1, which we compare utilising the corrected AIC values (Wood et al., 2016). Further, we include all categorical covariates as dummy-coded regressors. All effects are fixed beside the Gaussian random district-specific effects. We parametrise the temporal trend by a penalised spline (Eilers and Marx, 1996). In the final row of Table 1, the corresponding cAIC values indicate that Model 1 is the best specification of the missing at random process; hence we employ the corresponding set of covariates in our imputation model.

3.3 | Results

Originating from the missing at random analysis of Section 3.2, our imputation model incorporates a random district-specific effect, a temporal trend, and dummy covariates for the age, gender, and state cohort as well as an interaction between age and

Covariable	Parameter	
	μ_l	σ_l
	(Standard Error)	(Standard Error)
Intercept	1.852	-1.011
	(< 0.01)	0.022
Male	0.005	-0.069
	(< 0.01)	(0.024)
A35-A59	0.017	0.101
	(< 0.01)	(0.02)
A35-A59:Male	0.038	-0.027
	(< 0.01)	(0.030)
Weekend	-0.034	-0.049
	(< 0.01)	(0.017)

TABLE 2 Imputation Model: Sociodemographic linear estimates. The reference group are female individuals aged between 15 and 35 living in Baden-Württemberg.

gender groups. In our proposed imputation procedure, we parametrise not only the mean of the test delay (μ) but also its scale factor (σ). Therefore, we provide the full estimates of the corresponding model defined in Formula (2) of the main article separately for each coefficient. Besides, we decompose all reported effects into sociodemographic, state, smooth and random effects.

Sociodemographic Effects: The results of demographic terms are given in Table 2 and should be interpreted regarding the mean and dispersion parameter of the period between disease onset date and reporting date, which we define as test delay.

State Effects: In addition to the demographic covariates, we included a fixed effect for each state. For the results shown in Figure 6, we use Baden-Württemberg as the reference category. Hence, all estimates should be interpreted relative to Baden-Württemberg. The state effects on μ in Figure 6 (a) indicate that the average test delay in eastern states, e.g., Saxony-Anhalt and Saxony, is lower than in the reference category. The test delay volatility also varies significantly between the states, as shown in Figure 6 (b), e.g., it is the highest in Mecklenburg Western Pomerania and Hamburg.

Smooth Effects: We observe a negative temporal trend for reporting dates during the beginning of March. But once we see more reported cases in April, the average test delay lengthens. During this period, the σ parameter is higher, leading to a higher variance of the respective observations. After May, the average delay decreases but increases again during the last week of this study.

Random Terms: The district-specific intercept effects for μ are given in Figure 8 (a), while Figure 8 (b) shows them for σ .

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

8

FRITZ AND KAUEMANN

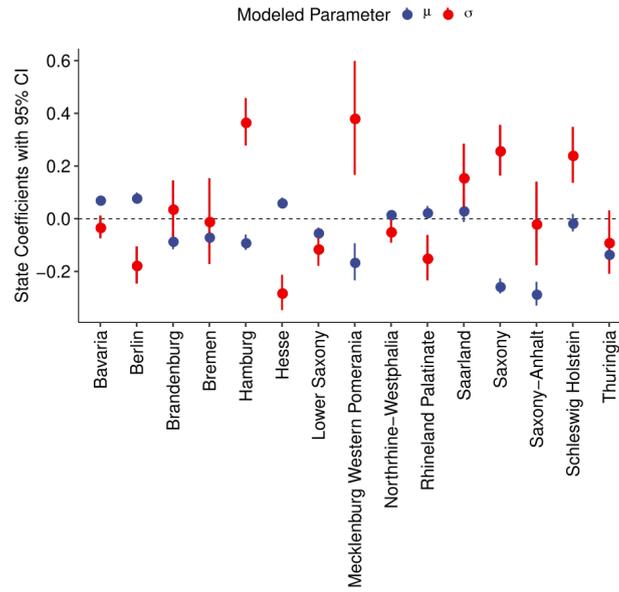


FIGURE 6 Imputation model: Linear state effects on μ and σ . The reference category is Baden-Württemberg, hence the respective coefficient is fixed at zero.

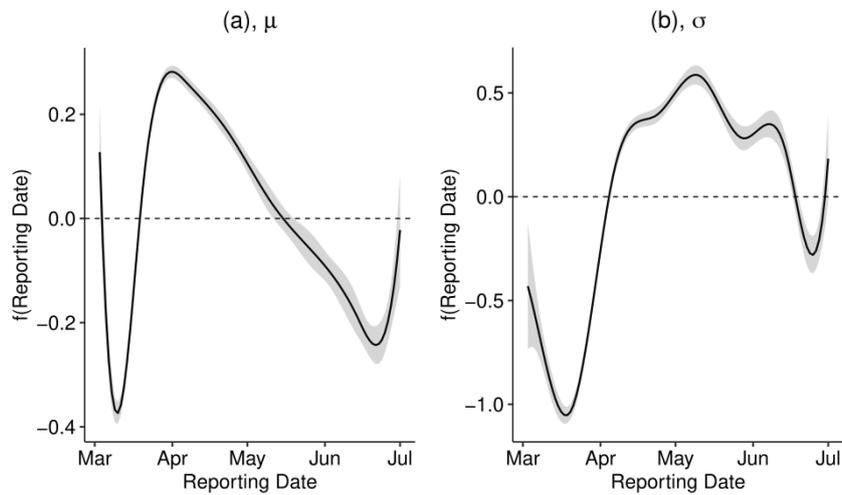


FIGURE 7 Imputation model: Smooth temporal trends of the μ (a) and σ (b) parameter.

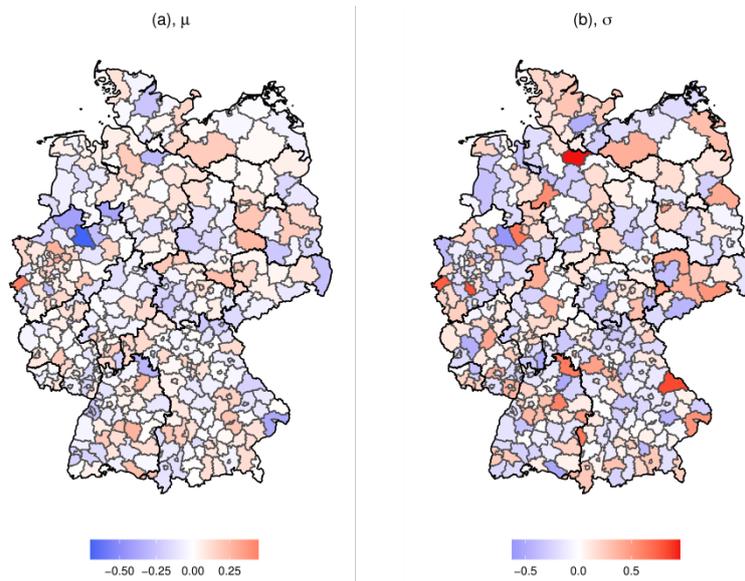


FIGURE 8 Imputation model: Random district-specific effects regarding μ (a) and σ (b) parameter.

3.4 | Sensitivity Analysis of Disease Onset Imputation

We carry out a sensitivity analysis of the imputation procedure introduced in Section 3.1 of the main manuscript to check whether our findings differ if we only regard cases with observed disease onset. In Figures 10 to 13, the model results are given if we only include the cases where the disease onset date is recorded in the surveillance data. All findings of the principal analysis are robust to the imputation method used; hence we see no structural differences that arise due to including all cases rather than only the cases with an observed disease onset.

4 | ALTERNATIVE INFECTION MODEL

To exhibit the robustness of our findings, we compare the results under an alternative model specification based on the quasi-likelihood (Wedderburn, 1974) combined with multiplicative random effects (comparable to Firth and Harris, 1991). In contrast to the main analysis, not the complete conditional distribution but solely its first two moments are specified. To enable a clear comparison, we carry out the exact same procedure detailed in the main article but substitute the negative binomial likelihood conditional on c for a conditional random variable $Y_{i,g,t} \mid x_{i,g,t-1}, y_{i,g,t-1}, a_i, b_i$ with $\mathbb{E}(Y_{i,g,t} \mid x_{i,g,t-1}, y_{i,g,t-1}, a_i, b_i) = \exp\{v_{i,g,t}^{END} + v_{i,g,t}^{EPI}\}$ and $\text{Var}(Y_{i,g,t} \mid x_{i,g,t-1}, y_{i,g,t-1}, a_i, b_i) = \exp\{v_{i,g,t}^{END} + v_{i,g,t}^{EPI}\}\phi$, where ϕ is an additional dispersion parameter to be estimated. Consecutively, we again correct for the multiple imputation scheme and the estimates relating to the social activity during COVID-19 are given together with the original findings in Figure 14 and 15. It is clearly visible that the findings hardly differ from the ones presented in the main article and that all conclusions drawn in the main article based on the negative binomial approach are consistent with the ones of the quasi likelihood model. The results of all other covariates are in line with the results of the principal analysis (see Figure 16 to 18).

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

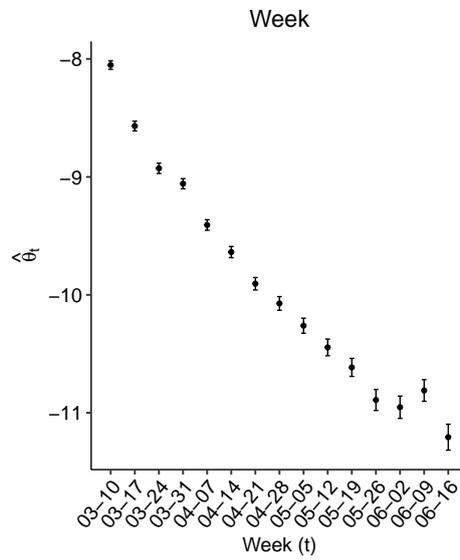


FIGURE 9 Sensitivity model: Estimate of temporal effect θ_t . The 95% confidence interval accompanies the estimates, and the shown dates (mm:dd) on the x-axis are the first days of the corresponding weeks.

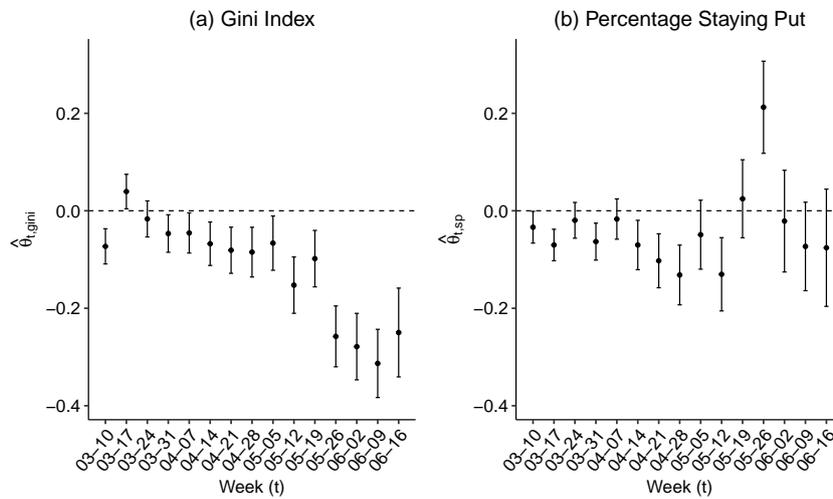


FIGURE 10 Sensitivity model: (a) Time-varying effects of the Gini index $\hat{\theta}_{t,gini}$. (b) Time-varying effects of the Percentage of People Staying Put $\hat{\theta}_{t,sp}$. The 95% confidence interval accompanies the estimates, and the shown dates (mm:dd) on the x-axis are the first days of the corresponding weeks.

REFERENCES

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.

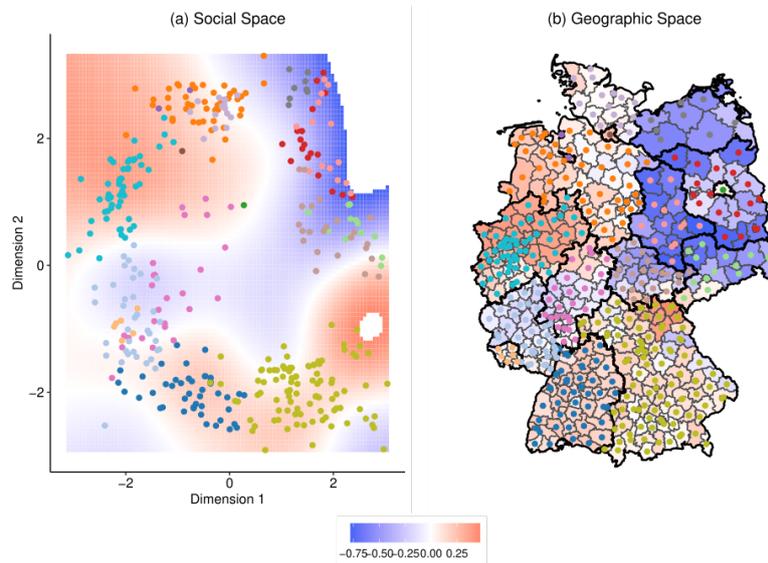


FIGURE 11 Sensitivity model: (a) Coordinates of the districts in the friendship space with the smooth partial effect of f_{soc} in the background. We only show the predictions in the range of observed values. (b) Coordinates of the districts in the geometric space with the smooth partial effect of f_{soc} again shown in the background for each district. The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

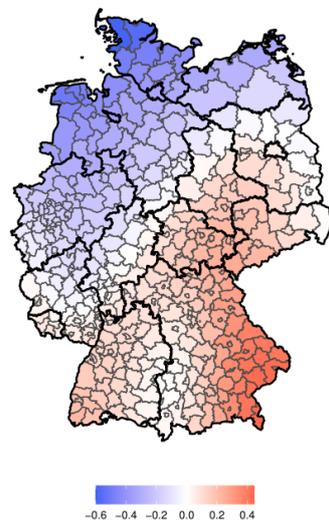


FIGURE 12 Sensitivity model: Estimated smooth spatial effect f_{coord} . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

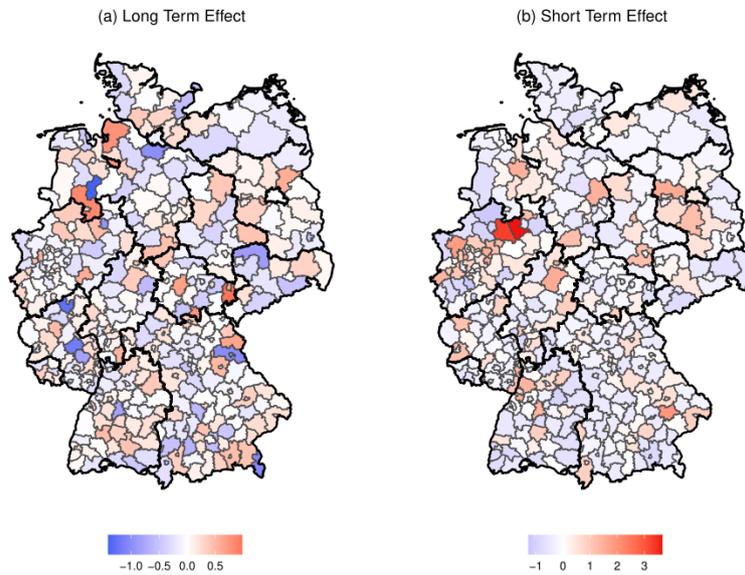


FIGURE 13 Sensitivity model: (a) Maximum posterior modes of the long-term random effects a_j . (b) Maximum posterior modes of the short-term random effects b_j . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

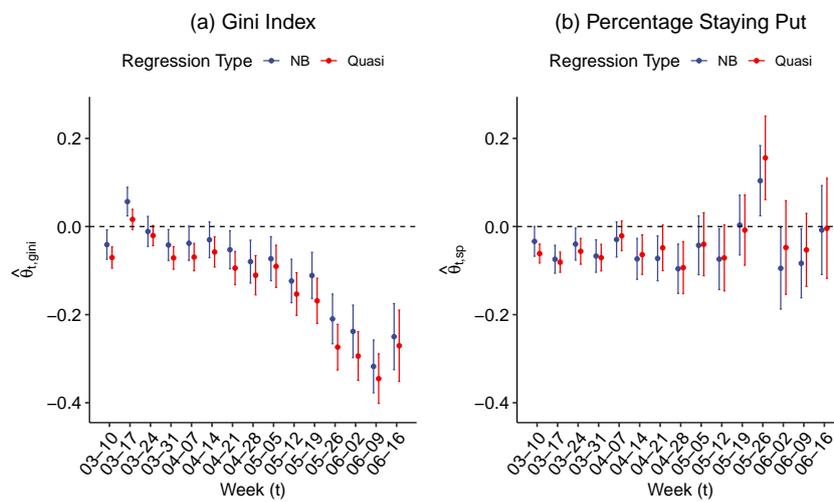


FIGURE 14 (a) Time-varying effects of the Gini index $\hat{\theta}_{t,gini}$ in the Quasi-Likelihood (Quasi) and the Negative Binomial model (NB). (b) Time-varying effects of the Percentage of People Staying Put $\hat{\theta}_{t,sp}$ in the Quasi-Likelihood and the Negative Binomial model. All estimates are accompanied by 95% confidence interval and the colour indicates whether the coefficients relate to the original Negative binomial fit presented in the main analysis or the quasi likelihood approach.

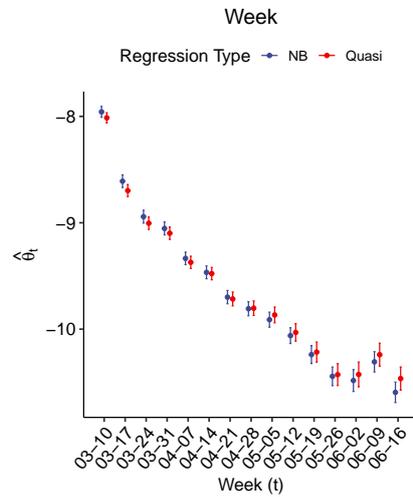


FIGURE 15 Time-varying baseline intensity θ_t in the Quasi-Likelihood (Quasi) and the Negative Binomial model (NB). The estimates are accompanied by the 95% confidence interval and the colour indicates whether the coefficients relate to the original Negative binomial fit presented in the main analysis or the quasi likelihood approach.

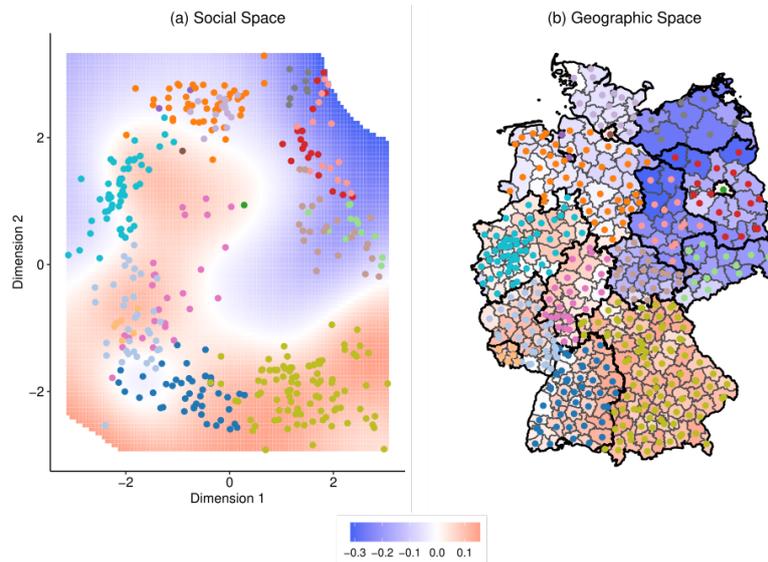


FIGURE 16 Quasi-likelihood model: (a) Coordinates of the districts in the friendship space with the smooth partial effect of f_{soc} in the background. We only show the predictions in the range of observed values. (b) Coordinates of the districts in the geometric space with the smooth partial effect of f_{soc} again shown in the background for each district. The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

7. On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

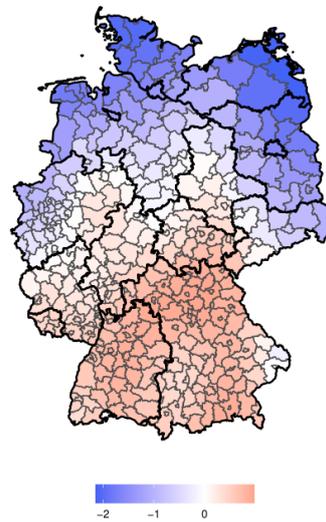


FIGURE 17 Quasi-likelihood model: Estimated smooth spatial effect f_{coord} . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

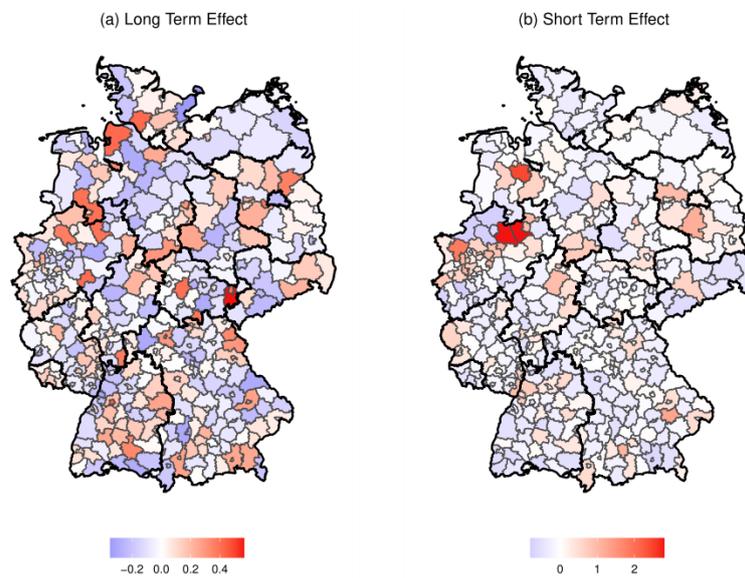


FIGURE 18 Quasi-likelihood model: (a) Maximum posterior modes of the long-term random effects a_j . (b) Maximum posterior modes of the short-term random effects b_j . The thick black lines represent borders between federal states, while the thinner grey borders separate federal districts.

- Firth, D. and I. R. Harris (1991). Quasi-likelihood for Multiplicative Random Effects. *Biometrika* 78(3), 545.
- Iyer, S., B. Karrer, D. Citron, F. Kooti, P. Maas, Z. Wang, E. Giraudy, P. A. Dow, and A. Pompe (2020). Large-scale measurement of aggregate human colocation patterns for epidemiological modeling. *medRxiv preprint*.
- Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. New York: Wiley.
- Maas, P., I. Shankar, A. Gros, P. Wonhee, L. McGorman, C. Nayak, and A. P. Dow (2019). Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In *Proceedings of the 16th ISCRAM Conference*, pp. 1–12.
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61(3), 439.
- Wood, S. N., N. Pya, and B. Säfken (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* 111(516), 1548–1563.

Part III.

**Temporal Exponential Random Graph
Models**

8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

Contributing article

Fritz, C., Mehrl, M., Thurner, W. P. and Kauermann, G. (2022) Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations. *Under review in the Journal of the American Statistical Association (Applications and Case Studies)*.

Replication code

The complete replication code for this article is available under https://github.com/corneliusfritz/sergm_replication A R package `ergm.sign` implementing the introduced methods is available under <https://github.com/corneliusfritz/ergm.sign>.

Author contributions

Based on the goal of Paul W. Thurner to model signed networks, Cornelius Fritz and Marius Mehrl had the idea to extend Exponential Random Graph Models for such networks. Göran Kauermann proposed the extension to dynamic signed networks. Cornelius Fritz implemented the accompanying R package `ergm.sign` and wrote the draft of all methodological parts of the manuscript (Sections 2, 3, 4). Marius Mehrl and Cornelius Fritz wrote the draft of the rest of the manuscript jointly, although Marius Mehrl wrote the majority of Sections 2.2 and 4. Cornelius Fritz, Marius Mehrl, Göran Kauermann, and Paul W. Thurner were all heavily involved with proofreading the manuscript.

Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

Cornelius Fritz*

Department of Statistics, LMU Munich

Marius Mehrl

Paul W. Thurner

Geschwister Scholl Institute of Political Science

Göran Kauermann

Department of Statistics, LMU Munich

May 24, 2022

Abstract

Substantive research in the Social Sciences regularly investigates signed networks, where edges between actors are either positive or negative. For instance, schoolchildren can be friends or rivals, just as countries can cooperate or fight each other. This research often builds on structural balance theory, one of the earliest and most prominent network theories, making signed networks one of the most frequently studied matters in social network analysis. While the theorization and description of signed networks have thus made significant progress, the inferential study of tie formation within them remains limited in the absence of appropriate statistical models. In this paper we fill this gap by proposing the Signed Exponential Random Graph Model (SERGM), extending the well-known Exponential Random Graph Model (ERGM) to networks where ties are not binary but negative or positive if a tie exists. Since most networks are dynamically evolving systems, we specify the model for both cross-sectional and dynamic networks. Based on structural hypotheses derived from structural balance theory, we formulate interpretable signed network statistics, capturing dynamics such as “the enemy of my enemy is my friend”. In our empirical application, we use the SERGM to analyze cooperation and conflict between countries within the international state system.

Keywords: Exponential Random Graph Models, Signed Networks, Structural Balance Theory, International Relations, Inferential Network Analysis

*The authors gratefully acknowledge support from the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A and the German Research Foundation (DFG) for the project TH 697/11-1: Arms Races in the Interwar Period 1919-1939. Global Structures of Weapons Transfers and Destabilization.

1 Introduction

In February 2022, Russia invaded Ukraine. This invasion shifted the relations that numerous European countries had with the belligerents. The EU member states, including previously Russia-aligned countries such as Hungary, sanctioned Russia and provide support to Ukraine. Belarus, a close ally of Russia, followed its partner into the conflict and was accordingly also sanctioned by the EU member states. And Turkey, a political and economic partner of both Ukraine and Russia, struggled to remain neutral in the conflict and thus sought to mediate between the belligerents. A meaningful geopolitical adjustment thus followed the Russian attack which demonstrates the importance of positive *and* negative ties in the international network of states, showing how pairwise cooperation and conflict between countries are interdependent.

Political scientists have studied this interplay of positive and negative ties between states since the early 1960s (Harary, 1961). In this context, international relations are conceived as signed networks, where the nodes are states and the edges are either positive, corresponding to bilateral cooperation, negative, expressing bilateral conflict, or non-existent. Most of this research builds on structural balance theory, which postulates that triads are balanced if they include an odd number of positive relations and unbalanced if that number is either even (“strong” structural balance; Heider, 1946; Cartwright and Harary, 1956) or exactly two (“weak” structural balance; Davis, 1967). Accordingly, International Relations scholars have studied whether specific triangular constellations correspond with these propositions (Harary, 1961; Healy and Stein, 1973; McDonald and Rosecrance, 1985; Doreian and Mrvar, 2015) and what implications structural balance has for community formation and system polarization (Hart, 1974; Lee et al., 1994). More recently, studies seek to test whether structural balance affects interstate conflict and cooperation in an inferential framework (Maoz et al., 2007; Lerner, 2016; Kinne and Maoz, 2022).

However, the study of signed networks is not restricted to International Relations. There are also applications to friendship and bullying between children (Huitsing et al., 2012, 2014), alliances and conflicts between tribal (Hage and Harary, 1984) or criminal groups (Nakamura et al., 2020), statements of support and opposition between politicians (Arinik et al., 2020; De Nooy and Kleinnijenhuis, 2013), and even to interactions within ecological networks (Saiz et al., 2017). In the setting of online social media and multiplayer games signed networks are also frequently studied (Leskovec et al., 2010; Bramson et al., 2021). Signed networks are thus a substantively important subject of study across and beyond the Social Sciences.

When working with signed networks, most techniques known from binary networks are not directly appropriate. A significant amount of work thus focuses on adapting blockmodels (Doreian and Mrvar, 2009; Jiang, 2015) as well as network statistics, such as centrality (Everett and Borgatti, 2014) and status (Bonacich and Lloyd, 2004), to signed networks. From an inferential perspective, the study of signed networks so far has mainly relied on logistic regression (Maoz et al., 2007; Lerner, 2016) or perceiving the observations as multivariate networks with multiple layers (Huitsing et al., 2012, 2014; Stadtfeld et al., 2020), where one level relates to the positive and another to the negative edges. While the former approach disregards endogenous dependence, the latter only allows for dependence between the separate observed layers of the network. Moreover, the multilayer approach does not adequately capture that most interactions in signed networks are either positive, negative,

or non-existent. In other words, countries having negative and positive relations at the same time is unrealistic.

In the context of binary networks, Frank and Strauss (1986) proposed Exponential Random Graph Models (ERGMs) as a generative model for a network encompassing n actors represented by the adjacency matrix $\mathbf{y} = (\mathbf{y}_{ij})_{i,j=1,\dots,n}$, where $y_{ij} = 1$ translates to an edge between actors i and j and $y_{ij} = 0$ indicates that there is no edge. Henceforth, we use lowercase letters for variables when referring to the realized value of a random variable, i.e., the observed network \mathbf{y} , and capitalize the name to indicate that they are stochastic random variables, for instance, \mathbf{Y} . Within this framework, Wasserman and Pattison (1996) formulate a probability distribution over all possible $\mathbf{y} \in \mathcal{Y}$ by a canonical exponential family model:

$$\mathbb{P}_{\theta}(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}^{\top} \mathbf{s}(\mathbf{y})\}}{\kappa(\boldsymbol{\theta})} \quad \forall \mathbf{y} \in \mathcal{Y}, \quad (1)$$

where \mathcal{Y} is the set of all observable binary adjacency matrices among n fixed actors, $s : \mathcal{Y} \rightarrow \mathbb{R}^q$ is a function of sufficient statistics weighted by the coefficients $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, and $\kappa(\boldsymbol{\theta}) := \sum_{\tilde{\mathbf{y}} \in \mathcal{Y}} \exp\{\boldsymbol{\theta}^{\top} \mathbf{s}(\tilde{\mathbf{y}})\}$ is a normalizing constant. Possible choices for the sufficient statistics $\mathbf{s}(\mathbf{y})$ of directed networks include the number of edges and triangles in the network (see Lusher et al., 2012 for a detailed overview of the model and other possible statistics). Depending on the specific sufficient statistics, ERGMs relax the often unrealistic conditional independence assumption inherent to most standard regression tools in dyadic contexts and allow general dependencies between the observed relations. Note that in many applications, auxiliary information \mathbf{x} exogenous to the network is available, which can also be used in the sufficient statistics. For brevity of the notation, we, however, omit the dependence of \mathbf{s} on \mathbf{x} . Due to this ability to flexibly specify dependence among relations, account for exogenous information, the desirable properties of exponential families, and versatile implementation in the `ergm` R package (Handcock et al., 2008; Hunter et al., 2008), the ERGM is a core inferential approach in the statistical analysis of networks.

In this article, we extend (1) to cover signed networks under general dependency assumptions and coin the term Signed Exponential Random Graph Model (SERGM) for the resulting model. The SERGM provides an inferential framework to test the predictions of, e.g., structural balance theory (Heider, 1946; Cartwright and Harary, 1956) without assuming that all observed relations are independent of one another. This characteristic is of vital importance given that balance theory explicitly posits that the sign of one relation depends on the state of other relations in the network. As the introductory examples suggest, interdependence-driven sign changes occur rapidly between states, necessitating the use of endogenous network statistics to adequately capture them. Along these lines, Lerner (2016, p. 75) notes that “tests of structural balance theory” should not rely on “models that assume independence of dyadic observations” and thereby flags the importance of developing an ERGM for signed networks. We answer this call by introducing, applying, and, via the R package `ergm.sign`, providing statistical software in R (R Core Team, 2021) to implement the SERGM for static and dynamic networks, which is currently available at:

<https://github.com/corneliusfritz/ergm.sign>

We proceed as follows: In the consecutive section, we formally introduce the SERGM and a novel suite of sufficient statistics to capture network topologies specific to signed

networks. In Section 3, we detail how to estimate the parameters of the SERGM and quantify the uncertainty of the estimates. Next, we apply the introduced model class to the interstate network of cooperation and conflict in Section 4. Finally, we conclude with a discussion of possible future extensions.

2 The Signed Exponential Random Graph Model

2.1 Model Formulation

First, we establish some notation to characterize signed networks. Assume that the signed adjacency matrix $\mathbf{y} = (y_{ij})_{i,j=1,\dots,n}$ was observed between n actors. Contrasting the binary networks considered in (1), the entries of this signed adjacency matrix y_{ij} are either “+”, “-”, or “0”, indicating a positive, negative, or no edge between actors i and j . To ease notation, we limit ourselves to undirected networks without any self-loops, i.e., $\forall i, j = 1, \dots, n$ $y_{ij} = y_{ji}$ and $y_{ii} = “0”$ holds. Nevertheless, the proposed model naturally extends to directed settings. We denote the space encompassing all observable signed networks between n actors by \mathcal{Y}^\pm and specify a distribution over this space analogous to (1) in the following log-linear form:

$$\mathbb{P}_\theta(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y})\}}{\kappa(\boldsymbol{\theta})} \forall \mathbf{y} \in \mathcal{Y}^\pm. \quad (2)$$

The function of sufficient statistics in (2) takes a signed network as its argument and determines the type of dependence between dyads in the network. A theoretically motivated suite of statistics one can incorporate as sufficient statistics follows in Section 2.2 but mirroring the term counting edges in binary networks, we can use the count of positive ties in signed network \mathbf{y} via

$$EDGE^+(\mathbf{y}) = \sum_{i < j} \mathbb{I}(y_{ij} = “+”),$$

where $\mathbb{I}(\cdot)$ is the indicator function. Along the same lines, one can define a statistic for the number of negative edges $EDGE^-(\mathbf{y})$ and use both statistics as intercepts in the model.

We can extend (2) to dynamic networks, which we denote by $\mathbf{Y}_1, \dots, \mathbf{Y}_T$ for observations at $t = 1, \dots, T$, by assuming a first-order Markov dependence structure to obtain

$$\mathbb{P}_\theta(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})\}}{\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1})} \forall \mathbf{y}_t \in \mathcal{Y}^\pm. \quad (3)$$

The sufficient statistics encompassed in $\mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ capture within-network or endogenous dependencies through statistics that only depend on \mathbf{y}_t and between-network dependencies when incorporating \mathbf{y}_{t-1} . One instance for network statistics for between-network dependency is the stability statistic for positive edges

$$STABILITY^+(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i < j} \mathbb{I}(y_{ij,t} = “+”) \mathbb{I}(y_{ij,t-1} = “+”),$$

which can equivalently be defined for negative ties. Thus, we assume that the observed network is the outcome of a Markov chain with state space \mathcal{Y}^\pm and transition probability (3). Of course, we may also include exogenous terms in (3), i.e., any pairwise- or actor-specific information external to \mathbf{y}_t .

For the interpretation of the estimates, techniques from binary ERGMs can be adapted. To derive a local tie-level interpretation, let θ_q with $q \in \{1, \dots, p\}$ denote the q th entry of $\boldsymbol{\theta}$ corresponding to the q th sufficient statistic, $s_q(\mathbf{y}_t, \mathbf{y}_{t-1})$. We further define $\mathbf{y}_t = (y_{ij,t})_{i,j=1,\dots,n}$ for $t = 1, \dots, T$ and by $\mathbf{y}_{ij,t}^+$ denote the network \mathbf{y}_t with the entry $y_{ij,t}$ fixed at “+”, $\mathbf{y}_{ij,t}^-$ and $\mathbf{y}_{ij,t}^0$ are established accordingly. Let $\mathbf{y}_{(-ij),t}$ refer to the network \mathbf{y}_t excluding the entry $y_{ij,t}$. Due to the added complexity of signed networks, the distribution of $Y_{ij,t}$ conditional on $\mathbf{Y}_{(-ij),t}$ is a multinomial distribution where the event probability of entry “+” is:

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij,t} = \text{“+”} | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t-1}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}_{ij,t}^+, \mathbf{y}_{t-1})\}}{\sum_{k \in \{+, -, 0\}} \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}_{ij,t}^k, \mathbf{y}_{t-1})\}}. \quad (4)$$

In the same manner, we can state the conditional probability of “-” and “0”. In accordance with change statistics from binary ERGMs, we subsequently define positive and negative change statistics through

$$\begin{aligned} \Delta_{ij,t}^{0 \rightarrow +}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}) &= \mathbf{s}(\mathbf{y}_{ij,t}^+, \mathbf{y}_{t-1}) - \mathbf{s}(\mathbf{y}_{ij,t}^0, \mathbf{y}_{t-1}) \\ \Delta_{ij,t}^{0 \rightarrow -}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}) &= \mathbf{s}(\mathbf{y}_{ij,t}^-, \mathbf{y}_{t-1}) - \mathbf{s}(\mathbf{y}_{ij,t}^0, \mathbf{y}_{t-1}). \end{aligned} \quad (5)$$

While the positive change statistic $\Delta_{ij,t}^{0 \rightarrow +}(\mathbf{y}_t, \mathbf{y}_{t-1})$ is the change in the sufficient statistics resulting from flipping the edge value of $y_{ij,t}$ from “0” to “+”, the negative change statistic $\Delta_{ij,t}^{0 \rightarrow -}(\mathbf{y}_t, \mathbf{y}_{t-1})$ relates to the change from “0” to “-”. By combining (4) and (5), we can obtain the relative log odds of $Y_{ij,t}$ to be “+” and “-” rather than “0”:

$$\begin{aligned} \log \left(\frac{\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij,t} = \text{“+”} | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})}{\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij,t} = \text{“0”} | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})} \right) &= \boldsymbol{\theta}^\top \Delta_{ij,t}^{0 \rightarrow +}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}) \\ \log \left(\frac{\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij,t} = \text{“-”} | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})}{\mathbb{P}_{\boldsymbol{\theta}}(Y_{ij,t} = \text{“0”} | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})} \right) &= \boldsymbol{\theta}^\top \Delta_{ij,t}^{0 \rightarrow -}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}). \end{aligned} \quad (6)$$

This allows us to relate $\boldsymbol{\theta}$ to the conditional distribution of $Y_{ij,t}$ given the rest of the network and derive two possible interpretations of the coefficients reminiscent of multinomial and logistic regression: the conditional log-odds of $Y_{ij,t}$ to be “+” rather than “0” are changed by the additive factor θ_p , if the value of $y_{ij,t}$ changing from “0” to “+” raises the p th entry of $\Delta_{ij,t}^{0 \rightarrow +}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1})$ by one unit, while the other statistics remain unchanged. A similar interpretation holds for the negative change statistic.

Second, one can employ a global interpretation to understand the parameters on a network level. Then, $\theta_q > 0$ indicates that higher values of $s_q(\mathbf{y}_t, \mathbf{y}_{t-1})$ are expected under (2) than under a multinomial graph model, which we define as a simplistic network model where the value of each dyad is “+”, “-” and “0” with equal probability. In the opposing regime with $\theta_q < 0$, we expect lower values than under this multinomial graph model.

2.2 From Structural Balance Theory to Sufficient Statistics

As discussed in the introduction, structural balance theory is a natural approach to signed networks. But so far, inferential work on it remains limited and uses, as we show below, suboptimal measures of its structural expectations. We thus shortly introduce the core logic of structural balance theory, discuss previous measures of it, and then derive sufficient statistics from it for inclusion in the SERGM. These statistics enable us to test the structural expectations formulated by structural balance theory in a principled manner within the framework introduced in Section 2.1.

Theory The main implication of structural balance theory relates to the existence of triads between actors. Triads are the relations between three actors (Wasserman and Faust, 1994) and generally called balanced if they consist solely of positive ties (“the friend of my friend is my friend”) or one positive and two negative ties (“the enemy of my enemy is my friend”). According to structural balance theory, this type of triad should be observed more often than expected by chance in empirical signed networks. In contrast, triads that include a single negative tie are structurally imbalanced as the node participating in both positive relations has to cope with the friction of its two “friends” being opposed to each other. This actor should thus try to turn the negative tie into a positive tie to achieve a balanced constellation where all three actors share positive connections. But if this proves impossible, the actor will eventually have to choose a side, making one of its previously positive ties negative and resulting in structural balance. In triads where relations between all three actors are negative, the actors at least have incentives to make similar changes; these triads are thus also considered structurally imbalanced (Heider, 1946; Cartwright and Harary, 1956). In particular, two actors could reap benefits by developing a positive relationship, pooling their resources, and ganging up on the third node. However, later work views these triads without any positive ties as weakly balanced (Heider, 1958; Davis, 1967), as Davis (1967) notes that enemies of enemies being enemies indicates structural imbalance only if there are two subsets of nodes in the network. Triadic constellations with one negative relation are thus structurally imbalanced, should be empirically rare, and, where they exist, tend to turn into balanced states. Where only one negative tie exists, there is strong pressure to either eliminate it or create an additional one. And where there are three negative ties, actors at least have a clear incentive to turn one of them into a positive relation opportunistically, though their (im-)balance depends on the nature of the wider system (see also Easley and Kleinberg, 2010, ch. 5).

Testing Structural Balance via Lagged Statistics In interstate relations, this theory implies that two countries that are on friendly terms with the same other state should not wage war against each other. If three states all engage in conflict with each other, two of them may also find it beneficial to bury their hatchet, focus on their common enemy, and pool their resources against it. Along these lines, existing research asks whether two countries’ probability to cooperate or to fight is affected by them sharing common friends or foes (Maoz et al., 2007; Lerner, 2016). In particular, these authors investigate whether having shared allies or enemies at time $t - 1$ affects the presence of positive and negative ties at t . The resulting “friend of my friend is my friend” statistic we can incorporate in

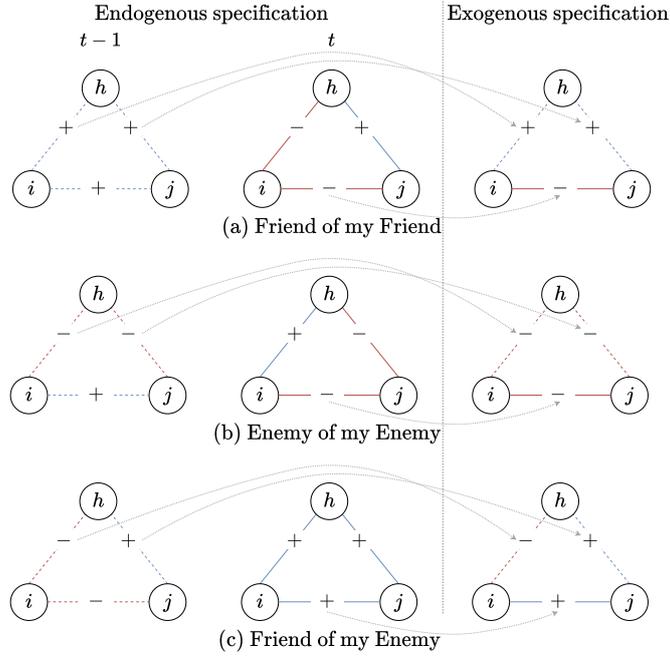


Figure 1: Combining past and present ties can misrepresent structural (im-)balance: Triads observed at $t - 1$ and t are balanced (left side), combined triads are imbalanced (right side). Dashed lines indicate tie at $t - 1$, solid ones at t . Dotted arrows show which ties from $t - 1$ and t contribute to the exogenous specification.

the sufficient statistics of (3) is:

$$CF^+(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i < j} \mathbb{I}(y_{ij,t} = "+") \left(\sum_{h \neq i, h \neq j} \mathbb{I}(y_{ih,t-1} = "+") \mathbb{I}(y_{jh,t-1} = "+") \right). \quad (7)$$

Similar delayed statistics can be defined for all other implications of the theory by treating the existence of common friends and foes as exogenous covariates. However, this approach comes with both theoretical and methodological problems. It is unclear whether actors wait a period (a calendar year in the case of Maoz et al., 2007 and Lerner, 2016) to adjust their relations towards structural balance and why they should do so as other applications of structural balance theory view these changes as instantaneous (see e.g. Kinne and Maoz, 2022). If the countries do not wait for a period, this approach can misrepresent the dynamics of signed networks as contradicting structural balance theory when they do not.

To illustrate this point, the right side of Figure 1 visualizes three structurally imbalanced constellations which Maoz et al. (2007) and Lerner (2016) uncover in the network of cooperation and conflict between states: (a) The friend of a friend being an enemy, (b) the enemy of an enemy being an enemy, and (c) the friend of an enemy being a friend. The left

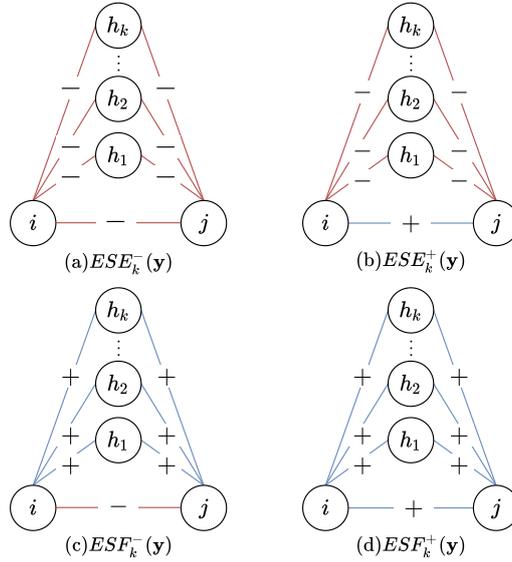


Figure 2: Sufficient statistics for signed networks.

side of Figure 1 presents the triads at $t - 1$ and t these constellations are potentially made up of as ties are not observed simultaneously. The links of i and j to h were observed at $t - 1$ but those between i and j at t . The structurally imbalanced triads on the right side of Figure 1 thus consist of observations of the same triad made at two different points in time. Crucially, the left side of Figure 1 shows that both of these observations can themselves be structurally balanced. Exogenous measures of common friends and enemies can thus only capture the predictions of structural balance theory if (i) actors i and j wait a period until they change their tie sign due to their links to h and (ii) their links to h remain unchanged. Both of these conditions require strong assumptions regarding how actors behave within a network. In particular, structural balance theory implies that the edges between i , j , and h are interdependent. But its exogenous operationalization assumes two of these edges as fixed while waiting to observe the third. An example shows that this is not just a theoretical issue, but mischaracterizes empirically observed relations between states: The US and Iran had common foes in 1978 but, in 1979, had become outright enemies themselves. The exogenous operationalization of structural balance regards this situation as unbalanced although it is an example of the scenario of Figure 1b.

Testing Structural Balance via Endogenous Statistics Therefore, endogenous network terms are necessary to capture the endogenous network dynamics postulated by structural balance theory. We next define endogenous statistics that mirror each constellation described by structural balance theory to test its predictions empirically. Building on the k -Edgewise-Shared Partner statistic developed to measure transitive closure in binary ERGMs (Hunter, 2007), we can define k -Edgewise-Shared Friends, $ESF_k(\mathbf{y})$, and k -Edgewise-Shared Enemies, $ESE_k(\mathbf{y}_t)$, for signed networks. $ESF_k(\mathbf{y}_t)$ counts the edges

with k shared friends and $ESE_k(\mathbf{y}_t)$ those with k shared enemies. We further differentiate between the state of the edge at the center of each triangular configuration and, e.g., write $ESF_k^+(\mathbf{y}_t)$ and $ESF_k^-(\mathbf{y}_t)$ as the version of the statistic where the value of y_{ij} is “+” and “-”, respectively. Figure 2 illustrates the resulting four statistics.

For $k = 2$ these statistics reduce to specific types of common triangle configurations (Holland and Leinhardt, 1972). However, as shown in Snijders et al. (2006), these types of statistics frequently lead to degenerate distributions where most of the probability mass is put on the empty or full graph (Handcock, 2003; Schweinberger, 2011). Moreover, the implied avalanche effect is particularly pronounced if the corresponding parameters are positive, as structural balance theory suggests. For binary ERGMs, it is thus standard to employ a statistic of the weighted sum of statistics in which the weights are proportional to the geometric sequence (Snijders et al., 2006; Hunter and Handcock, 2006). We follow this practice and define the geometrically weighted statistic for negative edgewise-shared enemies, as portrayed in Figure 2a, with a fixed decay parameter α as

$$GWSESE^+(\mathbf{y}_t, \alpha) = \exp\{\alpha\} \sum_{k=1}^{n-2} (1 - \exp\{-\alpha\})^k ESE_k^+(\mathbf{y}_t). \quad (8)$$

We establish the geometrically weighted variants of $ESE_k^-(\mathbf{y}_t)$, $ESF_k^+(\mathbf{y}_t)$, and $ESF_k^-(\mathbf{y}_t)$ accordingly. Each of these statistics reflects a specific type of triadic closure in signed networks as visualized in Figure 2. To interpret the coefficient θ_{GWSESE^+} one can consider the logarithmic relative change in the probability according to (3) when increasing the number of common enemies of a befriended edge by one and keeping all other statistics constant. If the befriended actors already had k prior common enemies before this change, this relative change is given by

$$\theta_{GWSESE^+} (1 - \exp\{-\alpha\})^k.$$

Thus, if $\theta_{GWSESE^+} > 0$, each additional common enemy raises the probability to observe the signed network, although the increments become smaller for higher values of k . Hunter (2007) shows that these geometrical weighted statistics are equivalent to the alternating k -triangle statistics proposed by Snijders et al. (2006).

These triadic structures fully capture the logic of structural balance as they allow us to study the prevalence of triads where positive ties account for zero (Figure 2a), one (Figure 2b), two (Figure 2c), and all three (Figure 2d) of the edges. According to this logic, we would expect the statistics $GWSESE^+(\mathbf{y}_t)$ and $GWSESE^-(\mathbf{y}_t)$ to be higher in empirical networks than expected by chance, but not $GWSESE^-(\mathbf{y}_t)$ and, particularly, $GWSESE^+(\mathbf{y}_t)$. If, on the other hand, the coefficients corresponding to $GWSESE^-(\mathbf{y}_t)$ or $GWSESE^+(\mathbf{y}_t)$ turn out to be positive in a network, this would offer empirical support for modifications of structural balance theory that also see the constellation in Figure 2a as balanced (Heider, 1958; Davis, 1967) or combine it with insights about, e.g., opportunism or reputation (Maoz et al., 2007). Mirroring the development of edge-wise shared enemy and friend statistics, it is also possible to compute dyad-wise statistics that do not require i and j to share a tie.

Other Sufficient Statistics Besides these substantively informed statistics developed from structural balance theory, there are - as in the binary case - numerous other statistics

one may incorporate into the model. Some of these are even necessary to isolate the effects of structural balance. In binary networks, closed triads where each node is connected to the others are more likely to form if the involved actors are highly active due to processes such as popularity. In the context of ERGMs, this phenomenon is captured by degree statistics counting the number of actors in the network with a specific number of edges. For signed networks, similar but more complicated processes may be at work and, to capture them, we define $DEG_k^+(\mathbf{y}_t)$ and $DEG_k^-(\mathbf{y}_t)$ as statistics that, respectively, count the number of actors in the signed network \mathbf{y}_t with degree $k \in \{1, \dots, n-1\}$ for “+”- and “-”-signed links, respectively. Since the degree statistics are also prone to the degeneracy issues detailed above, we define geometrically-weighted equivalents for the positive and negative degrees. One can also incorporate exogenous statistics for the propensity to observe either a positive tie, similar to (7), via the following statistic:

$$EXO^+(\mathbf{y}_t) = \sum_{i < j} \mathbb{I}(y_{ij,t} = "+")x_{ij,t},$$

where $x_{ij,t}$ can be any pairwise scalar information. Similar statistics can be defined for negative, $EXO^-(\mathbf{y}_t)$, and any, $EXO^\pm(\mathbf{y}_t)$, tie. To test whether there is a tendency for homo- or heterophily based on actor attribute $x = (x_1, \dots, x_n)$ in the network, one may transform the nodal information to the pairwise level by setting $x_{ij,t} = |x_{i,t} - x_{j,t}|$ or $x_{ij,t} = \mathbb{I}(x_{i,t} = x_{j,t})$ for continuous and categorical attributes, respectively.

3 Estimation and Inference

To estimate θ for a fully specified set of sufficient statistics, we maximize the likelihood of (3) conditional on the initial network \mathbf{y}_0 :

$$\mathcal{L}(\theta; \mathbf{y}_1, \dots, \mathbf{y}_T) = \prod_{t=1}^T \frac{\exp\{\theta^\top \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})\}}{\kappa(\theta, \mathbf{y}_{t-1})} = \frac{\exp\{\theta^\top (\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1}))\}}{\prod_{t=1}^T \kappa(\theta, \mathbf{y}_{t-1})}. \quad (9)$$

We can observe that this joint probability of the observed networks is still an exponential family, where the sufficient statistic is the sum of the individual statistics, the normalizing constant is composed of the product of the normalizing constants at each time point, and the canonical parameter is unchanged. Evaluating the normalizing constant in (9), on the other hand, necessitates the calculation of $T \cdot \binom{n(n-1)}{2}$ summands, making the direct evaluation of the likelihood prohibitive even for small networks. Fortunately, these difficulties are known from the analysis of binary networks and have been tackled in numerous articles (see, e.g., Strauss and Ikeda, 1990; Hummel et al., 2012; Snijders, 2002; Hunter and Handcock, 2006), which guide our estimation approach for the SERGM.

To circumvent the direct evaluation of (9), we can write the logarithmic likelihood ratio

of θ and a fixed θ_0 without a normalizing constant but an expected value

$$r(\theta, \theta_0; \mathbf{y}) = (\theta - \theta_0)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1}) \right) - \log \left(\mathbb{E}_{\theta_0} \left(\exp \left\{ (\theta - \theta_0)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{Y}_t, \mathbf{y}_{t-1}) \right) \right\} \right) \right). \quad (10)$$

We approximate the expectation in (10) by sampling networks over time, denoted by $\mathbf{Y}^{(m)} = (\mathbf{Y}_1^{(m)}, \dots, \mathbf{Y}_T^{(m)})$ for the m th sample, whose dynamics are governed by (3) under θ_0 . Due to the Markov assumption, it suffices to specify only how to sample $\mathbf{Y}_t^{(m)}$ conditional on \mathbf{y}_{t-1} for $t = 1, \dots, T$ via Gibbs sampling. In particular, we generate a Markov chain with state space \mathcal{Y}^\pm that, after a sufficient burn-in period, converges to samples from \mathbf{Y}_t conditional on \mathbf{y}_{t-1} . Since we toggle one dyad in each iteration, the conditional probability distribution we sample from is the multinomial distributions stated in (4). In a setting where we sample $Y_{ij,t}$ conditional on $\mathbf{y}_{(-ij),t}$ and \mathbf{y}_{t-1} with its present value given by $\tilde{y}_{ij,t}$, we can restate this conditional probability for “+” in terms of change statistics:

$$\mathbb{P}_\theta(Y_{ij,t} = “+” | \mathbf{Y}_{(-ij),t} = \mathbf{y}_{(-ij),t}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) = \frac{\exp \left\{ \theta^\top \Delta_{ij}^{\tilde{y}_{ij,t} \rightarrow +}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}) \right\}}{\sum_{k \in \{+, -, 0\}} \exp \left\{ \theta^\top \Delta_{ij}^{\tilde{y}_{ij,t} \rightarrow k}(\mathbf{y}_{(-ij),t}, \mathbf{y}_{t-1}) \right\}}.$$

This reformulation speeds up computation, since for most statistics the calculation of global statistics is computationally more demanding than the calculation of the change statistics defined in (5). Given M sampled networks, we get

$$r(\theta, \theta_0; \mathbf{y}) \approx (\theta - \theta_0)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1}) \right) - \log \left(\frac{1}{M} \sum_{m=1}^M \exp \left\{ (\theta - \theta_0)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(m)}, \mathbf{y}_{t-1}) \right) \right\} \right), \quad (11)$$

as an approximation of (10). However, according to standard theory of exponential families, the parameter θ maximizing (11) only exists if the sum of all observed sufficient statistics $\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ under θ_0 is inside the convex hull spanned by the sum of the sampled sufficient statistics (see Theorem 9.13 in Barndorff-Nielsen, 1978). Since this condition does not hold for arbitrary values of θ_0 , we modify the partial stepping algorithm under a log-normal assumption on the sufficient statistics introduced by Hummel et al. (2012) to dynamic signed networks for finding an adequate value for θ_0 (details can be found in the Supplementary Material). We seed our algorithm with θ_0 maximizing the pseudo-likelihood given by (4). To obtain estimates in the cross-sectional setting of (2), we can use the same procedure by setting $T = 1$.

To quantify the sampling error of the estimates, we rely on the theory of exponential families stating that the Fisher information $\mathcal{I}(\theta)$ equals the variance of $\sum_{t=1}^T \mathbf{s}(\mathbf{Y}_t, \mathbf{y}_{t-1})$ under the maximum likelihood estimate $\hat{\theta}$. We can estimate the Fisher information by again sampling networks $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(M)}$ and calculating the empirical variance of $\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(m)}, \mathbf{y}_{t-1})$ for $m = 1, \dots, M$. Due to the employed MCMC approximation, we follow standard practice of the `ergm` and `coda` packages (Handcock et al., 2008; Plummer et al.,

2006) and estimate the MCMC standard error by the spectral density at frequency zero of the Markov chains of the statistics. For the final variance estimate, we sum up both types of errors. By extending the bridge sampler introduced in Hunter and Handcock (2006) to the SERGM for dynamic networks, we can also evaluate the AIC value of the model to carry out a model selection (see Supplementary Material).

4 Testing Structural Balance in International Cooperation and Conflict

4.1 Motivation

We now employ the SERGM to investigate relations of cooperation and conflict in the interstate network over the years 2000-2010. This application speaks directly to Maoz et al. (2007), Lerner (2016), and the many other studies on structural balance in international relations cited above. We focus on this period since it is the most current period for which we have comprehensive and reliable data and because 9/11 provided a structural break in international relations. We do not let θ vary over time here, but it would be reasonable to assume that 9/11 altered the dynamics of the interstate network (see Thurner et al., 2019). Hence θ likely changed from before to after 9/11 and we analyze only the 2000s. One example of this phenomenon is how states cooperate on their defense and security policies after 9/11. While alliances remain important, there is nowadays relatively little change in the alliance network from one year to another as “only a dozen new alliances have emerged since 9/11” (Kinne, 2020, p.730). Instead, a new type of formal commitment between states, defence cooperation agreements (DCA), have become widely used throughout the 1990s and 2000s (see Kinne, 2018, 2020). To ensure that we capture interstate cooperation in a meaningful manner for the period we are interested in, we depart from previous studies of structural balance in international relations and use DCAs instead of alliances to operationalize interstate cooperation. We do so for several reasons.

First, as noted, the contemporary alliance network is basically static, experiencing little to no shifts over time. This is a challenge for estimation but, substantively, also severely limits the extent to which alliance relations could be affected by conflict between states. In contrast, DCAs are both initiated and terminated regularly (Kinne, 2018). Second, contemporary alliances are often multilateral and strongly institutionalized, meaning that if e.g. a new member joined NATO, it would result in the creation of several new alliance ties at once, but also that terminating these alliances, which have own secretariats, headquarters, and command structures, is challenging and thus empirically rare. Alliances hence do not clearly correspond to dyadic ties and have a life of their own which restricts tie deletion. In contrast, DCAs are bilateral and not as institutionalized, making them correspond much better to positive dyadic ties which can be formed but also removed (Kinne, 2018). Third, as opposed to alliances, DCAs are also signed by countries which have a policy of neutrality, thus reducing the risk that some ties are structural zeros, i.e. ineligible to be formed (Kinne, 2020). And fourth, most alliances only become active during armed conflict, stipulating wartime cooperation between their members (Leeds et al., 2002), but their goal is to deter enemies from instigating conflict in the first place. In other words, states’ formal commitment to cooperate, as demonstrated in an alliance, becomes realized

only in a fraction of cases which are those where the alliance’s main goal, deterrence, has failed. In contrast, DCAs specify states’ commitment to and framework for peacetime, day-to-day defence cooperation regarding activities such as joint defence policies, military exercises, the co-development of military technology, and bilateral arms transfers (Kinne, 2018, 2020). DCAs therefore present a better dynamic measure of regular, bilateral defence cooperation between states for the 2000s than alliances do.

4.2 Model Specification

To measure cooperative, positively-signed interstate relations, we thus use the DCA data collected by Kinne (2020) and consider a tie as existent and positive if a pair of states shares at least one active DCA in year t . For conflictious, negatively-signed relations, we follow Maoz et al. (2007) and Lerner (2016) by using the Militarized Interstate Dispute (MID) Data provided by Palmer et al. (2021). MIDs are defined as “united historical cases of conflict in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state” (Jones et al., 1996, p.163). We consider a tie to be existent and negative in year t if a pair of states has at least one MID between them. We plot the resulting interstate network, consisting of positive DCA- and negative MID-ties, in the Supplementary Material.

To specify a SERGM for modeling this evolving network, we first follow Maoz et al. (2007) and Lerner (2016) by including several exogenous covariates, namely i ’s and j ’s political difference, military capability ratio, the difference in wealth, and geographical distance. These variables’ sources are discussed in the Supplementary Material. Stemming from (3), we condition on the first year for the estimation and hence effectively model the network between 2001 and 2010.

Regarding endogenous statistics, the SERGM includes, most importantly, the four triadic terms developed above to capture the network’s tendency towards or against structural balance. Theoretically, we would expect the coefficients concerning $GWESSE^+(\mathbf{y}_t)$ and $GWESF^+(\mathbf{y}_t)$ but not $GWESF^-(\mathbf{y}_t)$ to have positive and statistically significant coefficients. For $GWESSE^-(\mathbf{y}_t)$, the expectation depends on whether we believe the state system to consist of two or of more groups (Davis, 1967). The latter appears more likely for the 2000s and we may thus expect to observe a positive coefficient. Furthermore, we include the positive and negative degree statistics, to capture highly active nodes’ propensity to (not) form more ties, and statistics that count the number of positive and negative edges as well as how many isolate nodes exist in each part of the network. Finally, stability terms are included to capture positive and negative ties remaining from the previous period. We term this specification Model 1 and present the results on the left side of Table 1.

We further compare Model 1 to a model specification where we replace the endogenous terms of structural balance, as depicted in Figure 2, with the exogenous versions used by Maoz et al. (2007) and Lerner (2016), stated in (7), where i ’s and j ’s ties with h are observed not contemporaneously but in $t - 1$. We denote the corresponding statistics by $CF^+(\mathbf{y}_t, \mathbf{y}_{t-1})$ and $CF^-(\mathbf{y}_t, \mathbf{y}_{t-1})$ to quantify the effect of common friends on positive and negative ties, while the number of common enemies are $CE^+(\mathbf{y}_t, \mathbf{y}_{t-1})$ and $CE^-(\mathbf{y}_t, \mathbf{y}_{t-1})$. Each of these exogenous measures corresponds to one of our triadic endogenous statistics, e.g. $CF^+(\mathbf{y}_t, \mathbf{y}_{t-1})$ to $GWESF^+(\mathbf{y}_t)$ and $CE^-(\mathbf{y}_t, \mathbf{y}_{t-1})$ to $GWESSE^-(\mathbf{y}_t)$. Otherwise, the

Table 1: Estimated coefficients and confidence intervals of the two model specifications detailed above. Dashes indicate the exclusion of covariates in a model specification. ΔAIC indicates the difference between the AIC values of Model 1 and 2.

	Model 1		Model 2	
	Coef.	CI	Coef.	CI
Edges +	-1.161	[-1.59,-0.732]	-0.689	[-1.203,-0.175]
Edges -	-1.754	[-2.142,-1.366]	-1.469	[-1.912,-1.026]
Isolates +	0.667	[-0.203,1.537]	0.462	[-0.422,1.346]
Isolates -	-1.188	[-2.319,-0.057]	-0.474	[-1.617,0.669]
Stability +	7.447	[7.331,7.563]	7.502	[7.379,7.625]
Stability -	5.531	[5.262,5.8]	5.594	[5.306,5.882]
Abs. Polity Diff. +	-0.022	[-0.032,-0.012]	-0.017	[-0.027,-0.007]
Abs. Polity Diff. -	0.004	[-0.016,0.024]	0.012	[-0.01,0.034]
CINC Ratio +	0.186	[0.117,0.255]	0.202	[0.129,0.275]
CINC Ratio -	-0.168	[-0.293,-0.043]	-0.14	[-0.279,-0.001]
Abs. GDP Diff. +	-0.521	[-0.57,-0.472]	-0.495	[-0.554,-0.436]
Abs. GDP Diff. -	-1.04	[-2.651,0.571]	-1.311	[-3.069,0.447]
Abs. Distance \pm	3.324	[0.515,6.133]	2.796	[-0.428,6.02]
<i>GWESE</i> ⁺ (Fig. 2a)	0.618	[0.308,0.928]	-	
<i>GWESE</i> ⁻ (Fig. 2b)	0.515	[0.199,0.831]	-	
<i>GWESF</i> ⁺ (Fig. 2c)	0.489	[0.415,0.563]	-	
<i>GWESF</i> ⁻ (Fig. 2d)	0.319	[0.178,0.46]	-	
<i>GWD</i> ⁺	-2.214	[-2.577,-1.851]	-2.625	[-3.015,-2.235]
<i>GWD</i> ⁻	-0.321	[-1.617,0.975]	-0.998	[-2.276,0.28]
<i>CF</i> ⁺	-		0.069	[0.051,0.087]
<i>CF</i> ⁻	-		0.077	[0.04,0.114]
<i>CE</i> ⁺	-		0.374	[-0.042,0.79]
<i>CE</i> ⁻	-		0.304	[-0.239,0.847]
ΔAIC	0		599.894	

two models are identical as Model 2 includes the other endogenous statistics specified in Model 1. We can thus adjudicate whether operationalizing structural balance dynamics in an endogenous manner, implying that they occur instantaneously, is preferable over the exogenous specification where these dynamics occur with a one-period time delay.

4.3 Results

Below, we interpret the results of the endogenous network terms and their exogenous equivalents. We discuss the coefficient estimates of the exogenous covariates in the Supplementary Material. As expected, both the $GWESF^+(\mathbf{y}_t)$ and the $GWESF^-(\mathbf{y}_t)$ terms exhibit positive and statistically significant coefficients, with neither confidence interval encompassing zero. These results align with structural balance theory in that both “the friend of my friend” and “the enemy of my enemy” are my friend. But we also find that the $GWESF^-(\mathbf{y}_t)$ and $GWESF^+(\mathbf{y}_t)$ coefficients are positive and statistically significant, albeit with smaller effects and confidence intervals closer to zero than in the case of the first two statistics. In the studied interstate network, there is thus also a tendency towards enemies of enemies being enemies. This echoes the point that triangles with three negative ties are imbalanced only in systems with two subsets (Davis, 1967), a condition that may have been present in the highly bipolar first half of the Cold War, but not more than a decade after its termination. This result is thus consistent with the verdict that, against early formulations of structural balance theory (Heider, 1946; Cartwright and Harary, 1956), “if two negative relations are given, balance can be obtained either when the third relationship is positive or when it is negative” (Heider, 1958, p.206). Observing that the effect of $GWESF^-(\mathbf{y}_t)$ is positive and statistically significant underlines the importance of overall network structure for the predictions of structural balance theory.

We also find that friends of friends have an increased probability of being enemies. In the international relations of the 2000s, what seems to hold is that both enemies of enemies and friends of friends are more likely to interact than if they did not share relations with a common third state. Friends of friends being more likely to fight than to have no relation at all suggests that shared relations may also indicate the “reachability” of one state to another within a system where some dyads, e.g., that between Lesotho and Belize, have a very low structural probability of ever being active (see, e.g. Quackenbush, 2006). Triadic closure, regardless of the sign, thus exists also in the network of cooperation and conflict between states. However, we observe that the tendency towards such closure is stronger for structurally balanced relations than for structurally imbalanced ones.

A comparison of the two model specifications shown in Table 1 allows us to ascertain whether specifying the triadic relationships endogenously affects substantive results and model performance. Here, it is visible that the AIC of the model with the endogenous statistics is lower than that with their exogenous versions. Specifying interdependent dynamics in the interstate network via endogenous covariates hence increases model performance compared to trying to capture them by including lagged, exogenous variables.

More strikingly, Table 1 shows that the substantive results of the corresponding endogenous and exogenous measures of structural balance dynamics differ significantly. Contrasting the results under the endogenous and exogenous model specification, the latter offers much more limited support for these notions. While the coefficient of $CF^+(\mathbf{y}_t, \mathbf{y}_{t-1})$ is positive and statistically significant, its effect size is still very close to zero. The “friends of friends are friends”-effect is thus found to be substantively negligible in Model 2. In contrast, the coefficient of $CE^+(\mathbf{y}_t, \mathbf{y}_{t-1})$ is positive and substantively larger, while its 95%-confidence interval includes zero, indicating that the model cannot statistically distinguish it from zero as its estimation is very imprecise. The statistics $CF^-(\mathbf{y}_t, \mathbf{y}_{t-1})$ and $CE^-(\mathbf{y}_t, \mathbf{y}_{t-1})$ mirror their corresponding endogenous terms from Model 1 in that both ex-

hibit positive coefficients but, again, the first is substantively much smaller and the second one very imprecisely estimated. On the whole, this comparison of an endogenous and an exogenous specification of the triadic configurations motivated by structural balance theory thus shows that Model 1 is preferable over Model 2. The model including endogenous terms thus not only provides better performance than that with their exogenous counterparts but these terms are also estimated to be more influential and more precisely.

4.4 Model Assessment

To assess the fit of the estimated SERGM, we employ a graphical tool inspired by Hunter et al. (2008) to evaluate whether it can adequately represent topologies of the observed network not explicitly incorporated as sufficient statistics in (3). Therefore, we sample networks from (3), compute the statistics, summarize them, and then compare this summary to the statistics evaluated on the observed network. Heuristically, a model generating simulations that better reflect the observed values also has a better goodness-of-fit. To cover signed networks, we investigate the observed and simulated distributions of positive and negative degrees, and edgewise-shared enemies and friends in the interstate network.

We report the goodness-of-fit plots for Model 1 from Table 1 in Figure 3 for the year 2005. In each subplot, a series of box plots display the distribution of a given value of the statistic under consideration over the networks simulated from the model via the Gibbs Sampler detailed in Section 3. The red line indicates where the statistic is measured in the observed network and should thus, ideally, lie close to the median value of the simulated networks, i.e., the center of the box plots. In Figure 3, this is the case for all four statistics, indicating that Model 1 under the estimated parameters generalizes well to network topologies not explicitly incorporated in the sufficient statistics.

Together, the results presented here indicate that the SERGM is able to uncover structural balance dynamics in the interstate network and is preferable over approaches that seek to model signed interstate networks under conditional independence, but also that further substantial research on structural balance in international relations is needed. The Supplementary Material employs the SERGM to analyze a cross-sectional network, representing enmity and friendship among New Guinean Highland Tribes (Hage and Harary, 1984), and shows its applicability when there is no observable temporal dependence structure.

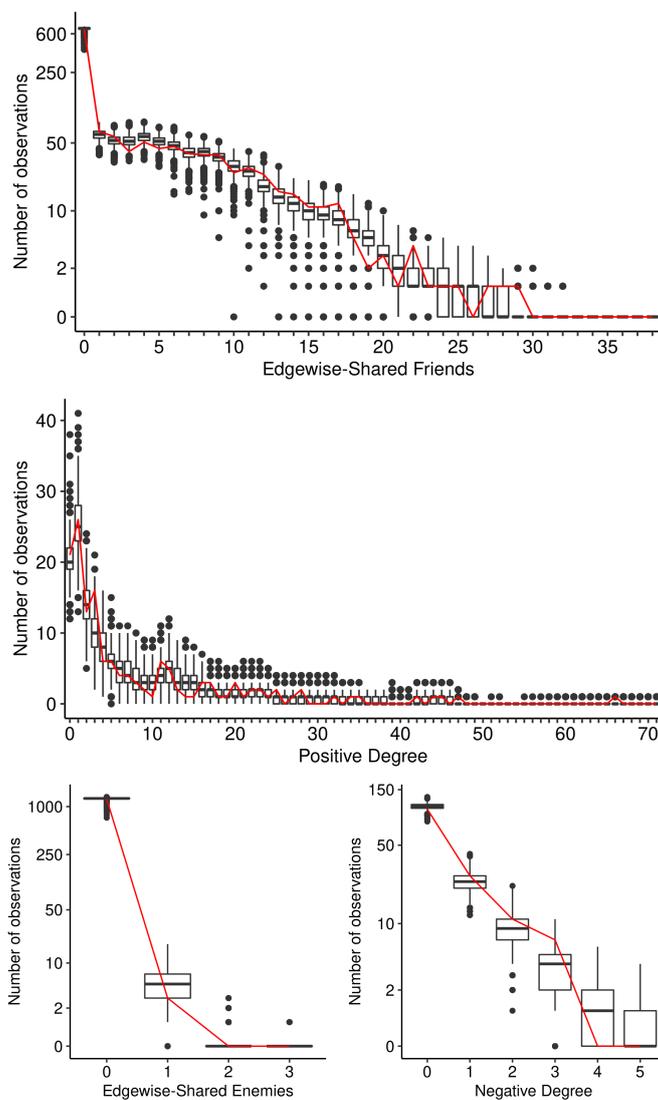


Figure 3: Goodness-of-fit assessment in year 2005.

5 Discussion

We extended the core regression model for network data to dynamic and cross-sectional signed networks. Given the theoretical foundation of structural balance, we introduce novel endogenous statistics that offer better performance than operationalizing them by lagged covariates, as commonly done in previous research. Finally, we apply the method to recent

data on militarized interstate disputes and defense cooperation agreements and provide a software implementation with the R package `ergm.sign`.

From a substantive point of view, this research offers new insights on the empirical testing of structural balance theory and challenges earlier inferential studies on the topic. How one captures structural balance matters. We show that an approach relying on past observations of some ties within a triad to measure structural balance as an exogenous variable can mischaracterize triadic (im-)balance. We thus develop endogenous balance measures that can be used in the SERGM framework and show empirically that these endogenous measures result in different substantive results as well as increased model performance as compared to the exogenous ones. Most importantly, the exogenous measures do not affect tie formation consistent with structural balance theory, whereas when employing the endogenous ones, we find evidence in line with it. States are thus more likely to cooperate if they share common partners or are hostile to the same enemies. This indicates that there is structural balance in interstate cooperation and conflict, at least when studying the 2000s. Future work in International Relations should seek to build on this fundamental result to test whether it also holds for earlier periods, for instance the bipolar Cold War years, and how structural balance interacts with exogenous factors such as military capabilities. Beyond International Relations, the SERGM will also serve to advance research across all Social Sciences, allowing researchers to investigate tie formation in networks of friendship and enmity between school children, gangs, or social media accounts.

At the same time, we find that, generally, states appear more likely to interact, positively or negatively, when they share friends or enemies. Substantively, this result suggests that, additional to structural balance, something else is at play and may indicate that some state dyads are structurally very unlikely to ever be active, due to the countries' distance, lack of economic development, and/or power projection capabilities, mirroring research on politically "relevant" or "active" dyads (see Quackenbush, 2006). But this implied variation in "reachability" between states also points to the fact that structural balance theory was developed on complete networks, where every possible ties is realized with either a negative or a positive sign, while empirical networks are usually incomplete (see Easley and Kleinberg, 2010, ch.5). It thus lends some support to Lerner's (2016) argument that tests of structural balance theory should not examine states' marginal probability to cooperate or fight, but instead their probability of cooperating or fighting *conditional upon them interacting*. However, following Lerner's (2016, Sec. 4.2.1) argument on the use of ERGMs in conjunction with this conditional viewpoint, it becomes evident that (3) is consistent with it. Defining $\mathbf{Y}^{|\pm|}$ with $Y_{ij}^{|\pm|} = 1$ if $Y_{ij} \neq "0"$ as the random adjacency matrix describing any type of interaction and \mathcal{Y} , be it positive or negative, one can derive the following conditional probability distribution

$$\mathbb{P}_{\theta}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_t^{|\pm|} = \mathbf{y}_t^{|\pm|}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})\}}{\tilde{\kappa}(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \mathbf{y}_t^{|\pm|})} \quad \forall \mathbf{y}_t \in \mathcal{Y}^{\pm}, \quad (12)$$

where $\tilde{\kappa}(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \mathbf{y}_t^{|\pm|}) = \sum_{\tilde{\mathbf{y}} \in \mathcal{Y}^{\pm}} \mathbb{I}(\tilde{\mathbf{y}}^{|\pm|} = \mathbf{y}_t^{|\pm|}) \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\tilde{\mathbf{y}}, \mathbf{y}_{t-1})\}$. The conditional distribution (12) is thus a SERGM with support limited to networks where $\mathbf{y}_t^{|\pm|}$ is equal to the observed network and the coefficients of (12) are unchanged. Therefore, (3) implies (12).

Alternatively, some dyads' lack of "reachability" may also indicate that dependency

structures are not fully global, even in international relations where all actors know of each other. Major powers should generally be able to reach all other states in the system, thus also making their actions globally relevant, but smaller countries' reach and relevance will be more locally limited. Since the more general framework of ERGMs in (1) relies on homogeneity assumptions implying that each endogenous mechanism has the same effect in the entire network, model (3) might assume dependence between relations where, in reality, there is none. One possible endeavor for future research would be adapting local dependence (Schweinberger and Handcock, 2015) to signed and dynamic networks. This approach assumes complex dependency solely within either observed or unobserved groups of the actors, solving the obstacle of "reachability" between some countries in the network. At the same time, other extensions of ERGMs, be it actor-specific random effects or curved ERGMs where α in (8) is estimated from the data, are also feasible under (2) and (3).

References

- Arinik, N., R. Figueiredo, and V. Labatut (2020). Multiple partitioning of multiplex signed networks: Application to european parliament votes. *Social Networks* 60, 83–102.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Bonacich, P. and P. Lloyd (2004). Calculating status with negative relations. *Social Networks* 26(4), 331–338.
- Bramson, A., K. Hoefman, K. Schoors, and J. Ryckebusch (2021). Diplomatic relations in a virtual world. *Political Analysis*, 1–22.
- Cartwright, D. and F. Harary (1956). Structural balance: A generalization of heider’s theory. *Psychological Review* 63(5), 277–293.
- Davis, J. A. (1967). Clustering and structural balance in graphs. *Human Relations* 20(2), 181–187.
- De Nooy, W. and J. Kleinnijenhuis (2013). Polarization in the media during an election campaign: A dynamic network model predicting support and attack among political actors. *Political Communication* 30(1), 117–138.
- Doreian, P. and A. Mrvar (2009). Partitioning signed social networks. *Social Networks* 31(1), 1–11.
- Doreian, P. and A. Mrvar (2015). Structural balance and signed international relations. *Journal of Social Structure* 16(1), 1–49.
- Easley, D. and J. Kleinberg (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- Everett, M. G. and S. P. Borgatti (2014). Networks containing negative ties. *Social Networks* 38(1), 111–120.
- Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association* 81(395), 832–842.
- Hage, P. and F. Harary (1984). *Structural Models in Anthropology*. Cambridge: Cambridge University Press.
- Handcock, M. (2003). Assessing degeneracy in statistical models of social networks. Technical report, University of Washington.
- Handcock, M. S., D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software* 24(1), 1–11.
- Harary, F. (1961). A structural analysis of the situation in the middle east in 1956. *Journal of Conflict Resolution* 5(2), 167–178.

- Hart, J. (1974). Symmetry and polarization in the european international system, 1870 - 1879: A methodological study. *Journal of Peace Research* 11(3), 229–244.
- Healy, B. and A. Stein (1973). The balance of power in international history: Theory and reality. *Journal of Conflict Resolution* 17(1), 33–61.
- Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology* 21(1), 107–112.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Holland, P. W. and S. Leinhardt (1972). Some evidence on the transitivity of positive interpersonal sentiment. *American Journal of Sociology* 77(6), 1205–1209.
- Huitsing, G., T. A. B. Snijders, M. A. J. Van Duijn, and R. Veenstra (2014). Victims, bullies, and their defenders: A longitudinal study of the coevolution of positive and negative networks. *Development and Psychopathology* 26(3), 645–659.
- Huitsing, G., M. A. van Duijn, T. A. Snijders, P. Wang, M. Sainio, C. Salmivalli, and R. Veenstra (2012). Univariate and multivariate models of positive and negative networks: Liking, disliking, and bully–victim relationships. *Social Networks* 34(4), 645–657.
- Hummel, R. M., D. R. Hunter, and M. S. Handcock (2012). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics* 21(4), 920–939.
- Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks* 29(2), 216–230.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481), 248–258.
- Hunter, D. R. and M. S. Handcock (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3), 565–583.
- Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software* 24(3), 1–29.
- Jiang, J. Q. (2015). Stochastic block model and exploratory analysis in signed networks. *Physical Review E* 91(6), 062805.
- Jones, D. M., S. A. Bremer, and J. D. Singer (1996). Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science* 15(2), 163–213.
- Kinne, B. J. (2018). Defense cooperation agreements and the emergence of a global security network. *International Organization* 72(4), 799–837.
- Kinne, B. J. (2020). The defense cooperation agreement dataset (dcad). *Journal of Conflict Resolution* 64(4), 729–755.

-
- Kinne, B. J. and Z. Maoz (2022). Local politics, global consequences: How structural imbalance in domestic political networks affects international relations. *Journal of Politics (OnlineFirst)*.
- Lee, S. C., R. G. Muncaster, and D. A. Zinnes (1994). The friend of my enemy is my enemy: Modeling triadic international relationships. *Synthese* 100(3), 333–358.
- Leeds, B., J. Ritter, S. Mitchell, and A. Long (2002). Alliance treaty obligations and provisions, 1815-1944. *International Interactions* 28(3), 237–260.
- Lerner, J. (2016). Structural balance in signed networks: Separating the probability to interact from the tendency to fight. *Social Networks* 45, 66–77.
- Leskovec, J., D. Huttenlocher, and J. Kleinberg (2010). Signed networks in social media. In *Proceedings of the 28th international conference on Human factors in computing systems*, pp. 1361–1370.
- Lusher, D., J. Koskinen, and G. Robins (2012). *Exponential random graph models for social networks*. Cambridge: Cambridge University Press.
- Maoz, Z., L. G. Terris, R. D. Kuperman, and I. Talmud (2007). What is the enemy of my enemy? causes and consequences of imbalanced international relations, 1816–2001. *Journal of Politics* 69(1), 100–115.
- McDonald, H. B. and R. Rosecrance (1985). Alliance and structural balance in the international system: A reinterpretation. *Journal of Conflict Resolution* 29(1), 57–82.
- Nakamura, K., G. Tita, and D. Krackhardt (2020). Violence in the “balance”: A structural analysis of how rivals, allies, and third-parties shape inter-gang violence. *Global Crime* 21(1), 3–27.
- Palmer, G., R. W. McManus, V. D’Orazio, M. R. Kenwick, M. Karstens, C. Bloch, N. Dietrich, K. Kahn, K. Ritter, and M. J. Soules (2021). The mid5 dataset, 2011–2014: Procedures, coding rules, and description. *Conflict Management and Peace Science* 39(4), 470–482.
- Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News* 6(1), 7–11.
- Quackenbush, S. L. (2006). Identifying opportunity for conflict: Politically active dyads. *Conflict Management and Peace Science* 23(1), 37–51.
- R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Saiz, H., J. Gómez-Gardeñes, P. Nuche, A. Girón, Y. Pueyo, and C. L. Alados (2017). Evidence of structural balance in spatial ecological networks. *Ecography* 40(6), 733–741.
- Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association* 106(496), 1361–1370.

- Schweinberger, M. and M. S. Handcock (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77(3), 647–676.
- Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*.
- Snijders, T. A. B., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology* 36(1), 99–153.
- Stadtfeld, C., K. Takács, and A. Vörös (2020). The emergence and stability of groups in social networks. *Social Networks* 60, 129–145.
- Strauss, D. and M. Ikeda (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* 85(409), 204–212.
- Thurner, P. W., C. S. Schmid, S. J. Cranmer, and G. Kauermann (2019). Network interdependencies and the evolution of the international arms trade. *Journal of Conflict Resolution* 63(7), 1736–1764.
- Wasserman, S. and K. Faust (1994). *Social network analysis : Methods and applications*. Cambridge: Cambridge University Press.
- Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika* 61(3), 401–425.

Supplementary Material – Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

Cornelius Fritz*

Department of Statistics, LMU Munich

Marius Mehrl

Paul W. Thurner

Geschwister Scholl Institute of Political Science

Göran Kauermann

Department of Statistics, LMU Munich

May 25, 2022

1 Technical Details

1.1 Partial Stepping Algorithm

Following standard theory of exponential families, θ maximizing the approximate likelihood detailed in (11) of the main article only exists if the observed sufficient statistics $\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ are inside the convex hull spanned by the sampled sufficient statistics $(\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(m)}, \mathbf{y}_{t-1}), \dots, \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(M)}, \mathbf{y}_{t-1}))$ (Barndorff-Nielsen, 1978). Since this condition does not hold for arbitrary values of θ_0 , we adapt the partial stepping algorithm introduced by Hummel et al. (2012) to find an adequate θ_0 .

In the k th step of this iterative procedure, we substitute $\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ in (11) of the main article by

$$\xi^{(k)} = \gamma^{(k)} \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1}) + (1 - \gamma^{(k)}) \widehat{\mathbf{m}}^{(k)}, \quad (1)$$

where $\gamma^{(k)} \in (0, 1]$ and $\widehat{\mathbf{m}}^{(k)} = \frac{1}{M} \sum_{m=1}^M \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(m)}, \mathbf{y}_{t-1})$ is the estimated mean of the sufficient statistics of networks sampled under $\theta^{(k)}$. We select the largest possible value

*Corresponding Autor: cornelius.fritz@stat.uni-muenchen.de

of $\gamma^{(k)}$ in (1) such that even the point marginally closer to $\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$, defined by $1.05\gamma^{(k)} \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1}) + (1 - 1.05\gamma^{(k)}) \mathbf{m}^{(k)}$, is inside the convex hull spanned by the sampled statistics. One can test whether a point $\in \mathbb{R}^p$ lies in this convex hull via a linear programming problem (details can be found in Hummel et al., 2012 and Krivitsky et al., 2022).

To update $\boldsymbol{\theta}^{(k)}$ to $\boldsymbol{\theta}^{(k+1)}$ for a given $\gamma^{(k)}$, we thus optimize

$$\left(\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\right)^\top \boldsymbol{\xi}^{(k)} - \log \left(\frac{1}{M} \sum_{m=1}^M \exp \left\{ \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(m)}, \mathbf{y}_{t-1}) \right) \right\} \right), \quad (2)$$

with a Newton-Raphson algorithm and $\mathbf{y}_t^{(m)} \forall m = 1, \dots, M$ and $t = 1, \dots, T$ sampled from model (3) of the main article. To ease this step, we assume that $\sum_{t=1}^T \mathbf{s}(\mathbf{Y}_t, \mathbf{y}_{t-1})$ follows a p -variate Gauss distribution with mean $\mathbf{m}^{(k)}$ and covariance matrix $\boldsymbol{\Sigma}^{(k)}$, which is the covariance matrix of the sufficient statistics under $\boldsymbol{\theta}^{(k)}$. Both terms can be estimated with samples $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(M)}$. Thereby we can state the optimal value of (2) in closed form:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \left(\widehat{\boldsymbol{\Sigma}}^{(k)}\right)^{-1} \left(\boldsymbol{\xi}^{(k)} - \widehat{\mathbf{m}}^{(k)}\right).$$

The algorithm terminates when we estimate $\gamma^{(k)} = 1$ two iterations in a row, we then continue the procedure with $\boldsymbol{\xi}^{(k)} = \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t, \mathbf{y}_{t-1})$ until the estimates stabilize.

1.2 Evaluation of the AIC

To decide between alternative specifications of the sufficient statistics, a common method is to select the model with the lowest AIC value. The AIC is defined as

$$\text{AIC}(M) = 2p - 2\ell(\hat{\boldsymbol{\theta}}; \mathbf{y}), \quad (3)$$

where M is a SERGM for temporal networks with a particular specification of the sufficient statistics and estimated parameters $\hat{\boldsymbol{\theta}}$ and $\ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) = \log \left(\prod_{t=1}^T \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) \right)$ is the log likelihood. To evaluate (3), we have to calculate the value of the intractable logarithmic likelihood at $\hat{\boldsymbol{\theta}}$, which we can restate by

$$\ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) = r(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_{\text{Ind}}; \mathbf{y}) + \ell(\hat{\boldsymbol{\theta}}_{\text{Ind}}; \mathbf{y}), \quad (4)$$

where $\hat{\boldsymbol{\theta}}_{\text{Ind}} \in \mathbb{R}^p$ is the estimate of the sub-model including only the subset from the sufficient statistics that abide the conditional dependence assumption (the coefficients of all other (endogenous) statistics are fixed at 0). Due to this characteristic, $\ell(\hat{\boldsymbol{\theta}}_{\text{Ind}}; \mathbf{y})$ is equivalent to the log likelihood in a multinomial regression and can be computed in closed form. To evaluate $r(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_{\text{Ind}}; \mathbf{y})$, we follow Hunter and Handcock (2006) and apply path sampling (Gelman and Meng, 1998) to approximate

$$\log \left(\mathbb{E}_{\boldsymbol{\theta}_0} \left(\exp \left\{ \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^\top \left(\sum_{t=1}^T \mathbf{s}(\mathbf{Y}_t, \mathbf{y}_{t-1}) \right) \right\} \right) \right) = \log \left(\frac{\boldsymbol{\kappa}(\boldsymbol{\theta})}{\boldsymbol{\kappa}(\boldsymbol{\theta}_0)} \right)$$

from (11) in the main article in a more precise manner. If we specify a smooth mapping $\boldsymbol{\theta} : [0, 1] \rightarrow \mathbb{R}^p$ with $\boldsymbol{\theta}(0) = \hat{\boldsymbol{\theta}}_{\text{Ind}}$ and $\boldsymbol{\theta}(1) = \hat{\boldsymbol{\theta}}$ and let $0 = u_0 < u_1 < \dots < u_J = 1$ for $J \in \mathbb{N}$ be a fixed grid of so-called bridges and its finite support, the following approximation holds:

$$\begin{aligned} \log \left(\frac{\kappa(\boldsymbol{\theta})}{\kappa(\boldsymbol{\theta}_0)} \right) &= \sum_{j=1}^J \frac{1}{u_j - u_{j-1}} \mathbb{E}_{\boldsymbol{\theta}(u_j)} \left(\left(\frac{d}{du} \boldsymbol{\theta}(u) \Big|_{u=u_j} \right)^\top \sum_{t=1}^T \mathbf{s}(\mathbf{Y}_t, \mathbf{y}_{t-1}) \right) \\ &\approx \frac{1}{M} \sum_{j=1}^J \sum_{m=1}^M \frac{1}{u_j - u_{j-1}} \left(\left(\frac{d}{du} \boldsymbol{\theta}(u) \Big|_{u=u_j} \right)^\top \sum_{t=1}^T \mathbf{s}(\mathbf{y}_t^{(j,m)}, \mathbf{y}_{t-1}) \right), \end{aligned} \quad (5)$$

where $\mathbf{y}_t^{(j,1)}, \dots, \mathbf{y}_t^{(j,M)}$ are networks sampled conditional on \mathbf{y}_{t-1} under $\boldsymbol{\theta}(u_j) \forall j = 1, \dots, J$. For our implementation, we set $\boldsymbol{\theta}(u) = \hat{\boldsymbol{\theta}}_{\text{Ind}} + u(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{\text{Ind}})$, corresponding to a linear path from $\hat{\boldsymbol{\theta}}_{\text{Ind}}$ to $\hat{\boldsymbol{\theta}}$ and $\frac{d}{du} \boldsymbol{\theta}(u) = \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{\text{Ind}}$. Plugging (5) into the first row of (11) of the main article permits the computation of (3). For a more technical derivation of (5), we refer to Hunter and Handcock, 2006 or Gelman and Meng, 1998.

2 Details on the Application to International Cooperation and Conflict

2.1 Data Visualization and Covariate Details

Here, we visualize the network and offer additional details regarding the data sources for the application of the SERGM to interstate relations presented in Section 4. As discussed in this section, we source data from Kinne (2020) and Palmer et al. (2021) to construct a network, spanning the years 2000-2010, where positive ties represent Defense Cooperation Agreements (DCAs) and negative ties Militarized Interstate Disputes (MIDs) between states. A snapshot of the resulting network, as observed in 2005, is presented in Figure 1.

For this application, we also use additional data to construct our exogenous covariates. The information underlying these variables, as well as the MID data, are sourced from the `peacsciencr` package (Miller, 2021), but the original data sources are as follows: We measure countries' absolute political difference (Abs. Polity Diff.) using their polity scores (Marshall et al., 2018), their relative military power by taking the ratio of their Composite Indicators of National Capabilities* (CINC Ratio; Singer et al., 1972), their difference in wealth via their absolute GDP difference (Abs. GDP Diff; Anders et al., 2020) and obtain their geographical distance from Schvitz et al. (2022), log-transforming it before inclusion (Abs. Distance). For each covariate, we separately estimate effects on the propensity of a positive and negative edge. The only exception to this rule is the effect of the absolute distance, which is assumed to be equal for both types of edges.

We now shortly discuss the estimation results for these covariates, as reported for Model 1 in Table 1 of the main article. These estimates are *ceteris paribus*, i.e. when accounting for network dependencies via the endogenous terms. Regarding cooperation, countries are found to be more likely to formally work together via defense cooperation agreements if they

*We use the higher CINC value in the ratio's numerator.

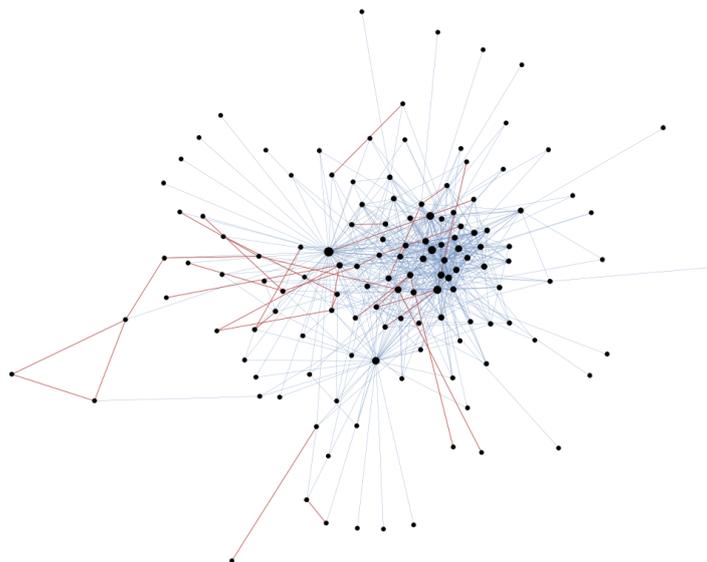


Figure 1: Network of MIDs (red) and DCAs (blue) in 2005. The size of each node relates to the degree (positive plus negative) of the respective country.

are politically more similar, more comparable in their wealth, but also differ more in their material military capabilities. In particular the first result is in line with previous research showing that similar regimes are more likely to ally (Lai and Reiter, 2000; Warren, 2016) while the second indicates that for DCAs, which regulate activities such as the joint research and development of military technology, countries' economic match also plays a role. That countries are more likely to cooperate as their CINC ratio increases indicates, instead, that DCAs also follow a hierarchical structure where powerful states enter agreements with less powerful ones (Lake, 2009). In contrast, we see that states are more likely to fight if their CINC_s are more similar, and hence military capabilities, are more similar whereas their differences in terms of regime type and wealth are not found to play a role. Finally, countries' absolute distance exhibits a positive coefficient, indicating that, surprisingly, they are more likely to interact the farther they are away from each other.

2.2 MCMC Diagnostics

In figures 2–6, we present some diagnostic plots of the MCMC chain used in the final iteration of Model 1 in the application of Section 4 of the main article. We average the

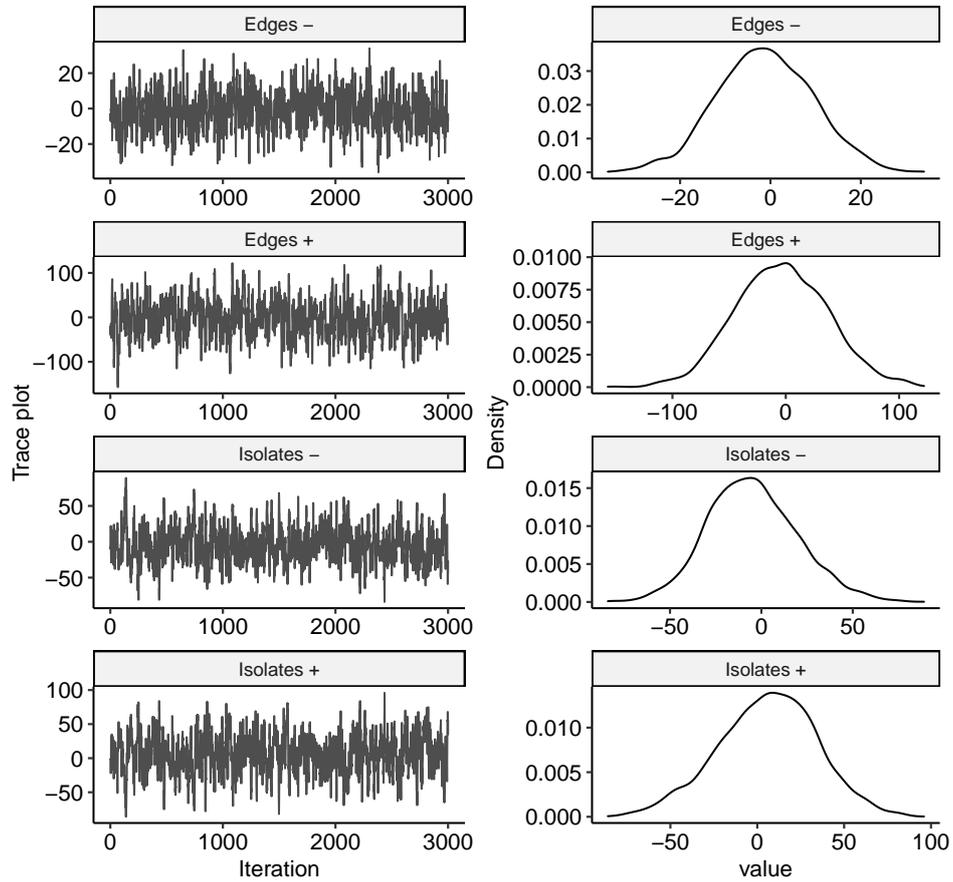


Figure 2: MCMC diagnostics of Model 1.

Markov chain of each sufficient statistic around its observed value for better readability. Overall, one can observe that the model's estimates converged, are not degenerate, and are equal to the maximum likelihood estimates since the Markov chain oscillates around 0.

8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

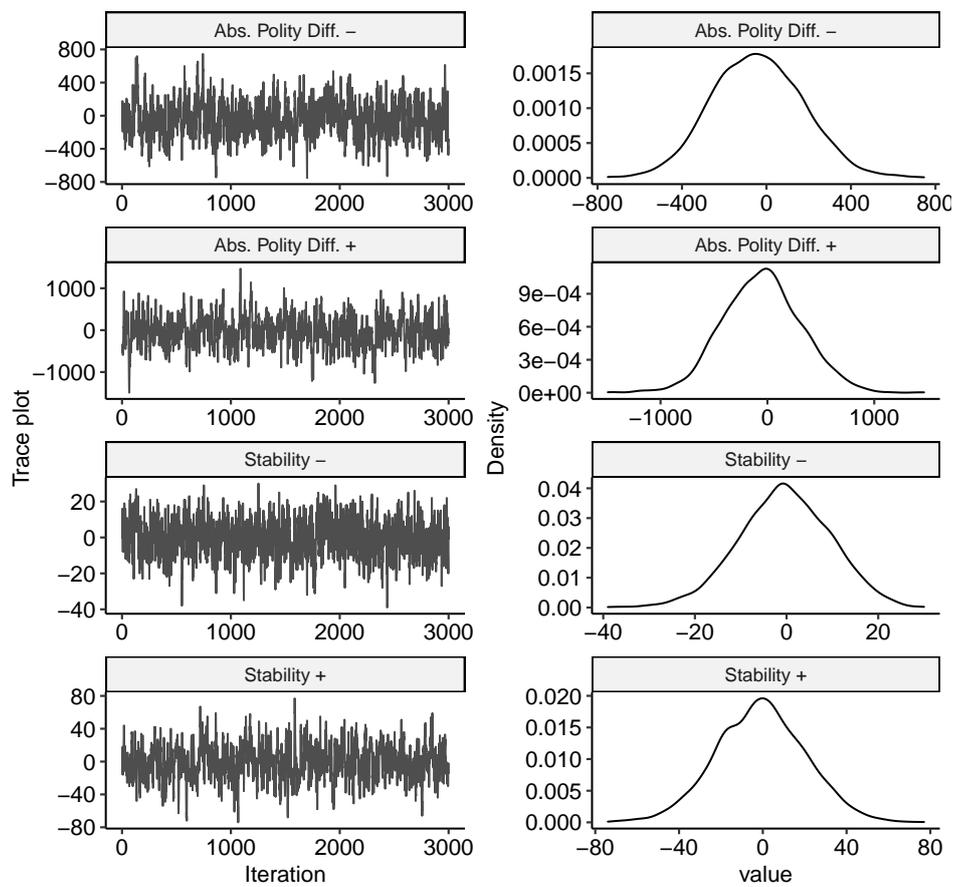


Figure 3: MCMC diagnostics of Model 1.

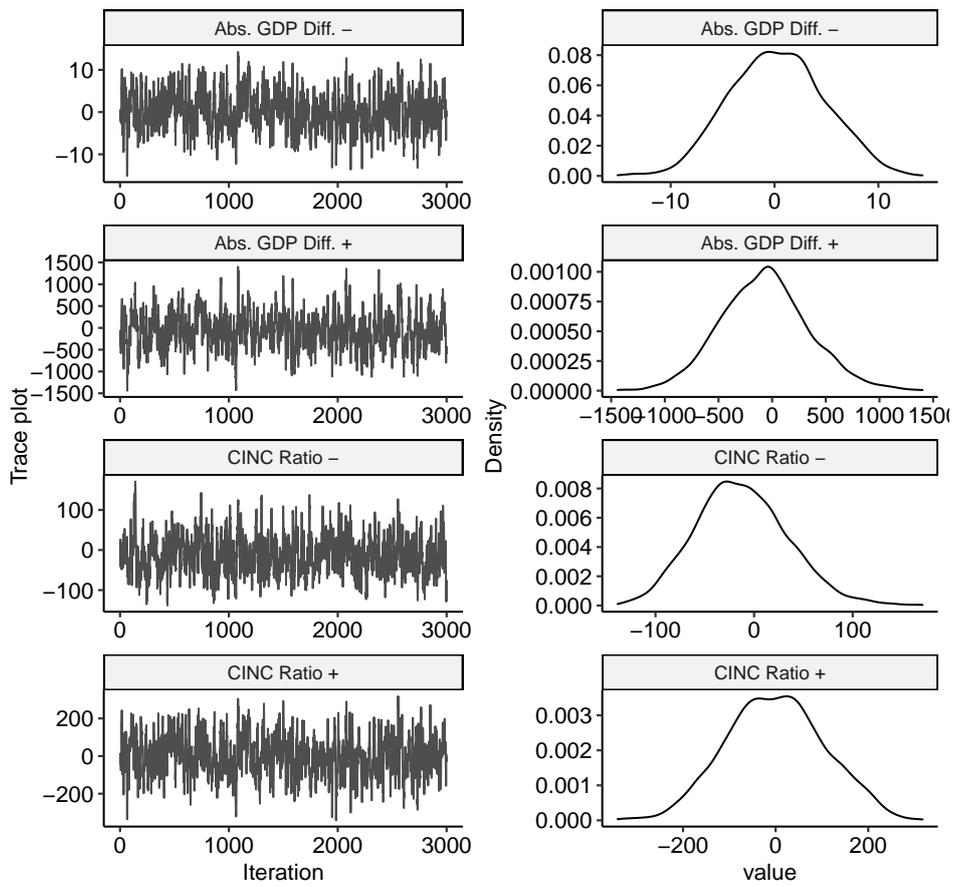


Figure 4: MCMC diagnostics of Model 1.

8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

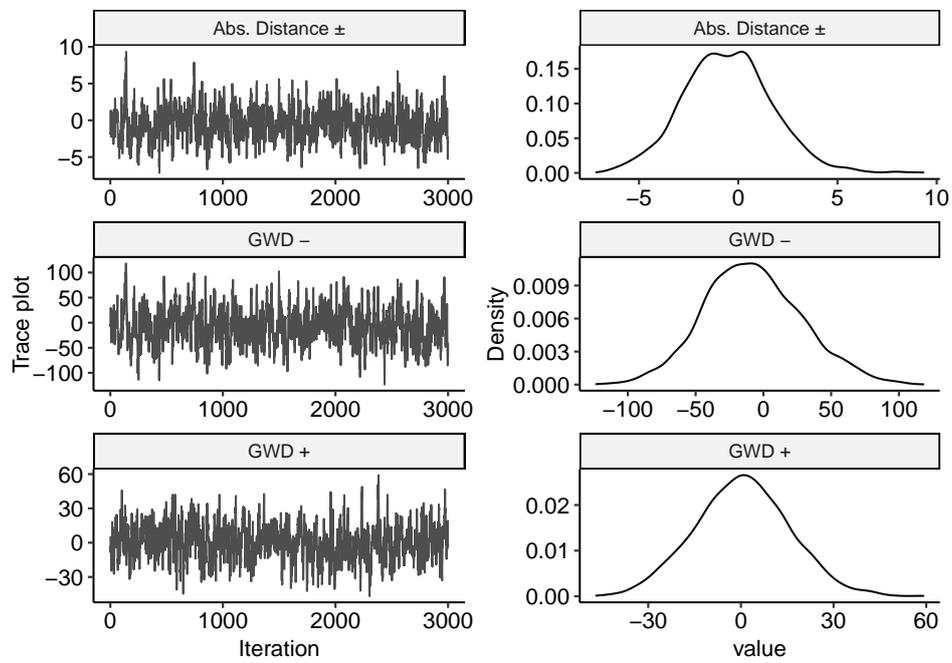


Figure 5: MCMC diagnostics of Model 1.

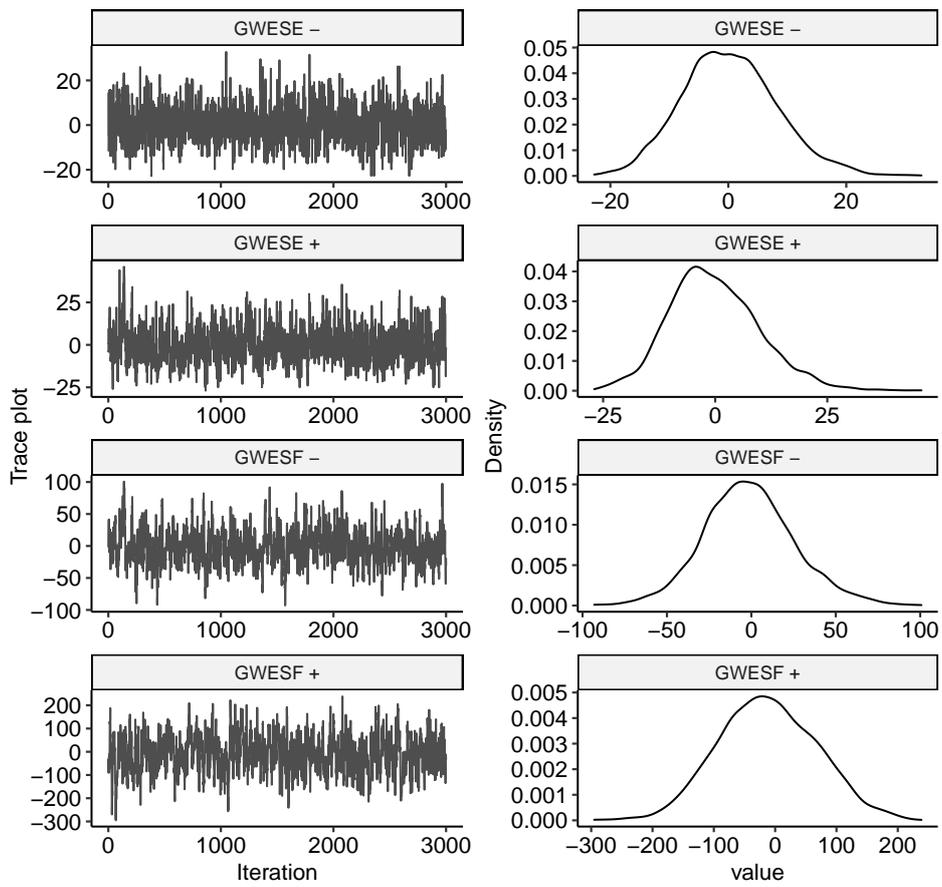


Figure 6: MCMC diagnostics of Model 1.

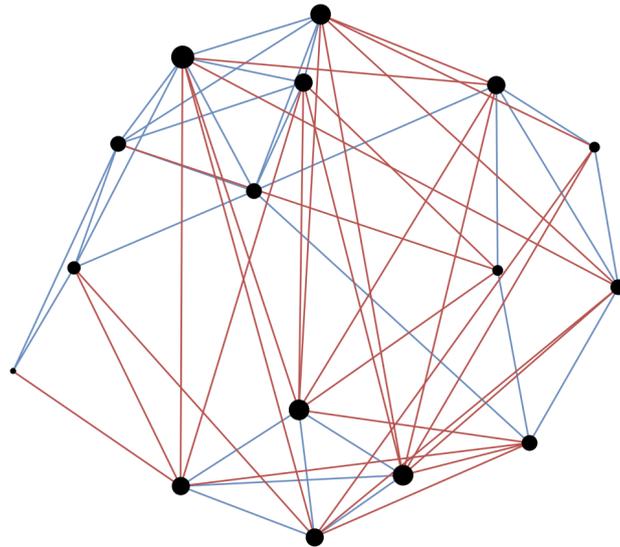


Figure 7: Network of enmity (red) and friendship (blue) among New Guinean Highland tribes. The size of each node relates to the degree (positive plus negative) of the respective people.

3 Application to a static network: Enmity and Friendship among New Guinean Highland Tribes

3.1 Data Visualization and Sources

Next to dynamic networks one can also apply the SERGM to static networks. We demonstrate this with network data on interactions between the New Guinean Highland Tribes originally collected by Read (1954) and presented in Hage and Harary (1984). We source these data from the R package `signet` (Schoch, 2020). The network covers relations of enmity and friendship among sixteen subtribes of the Gahuku-Gama, based on the anthropological work of Read (1954). Hage and Harary (1984) introduce it as an example of a network which is not perfectly balanced due to the existence of triads with zero or two positive ties but note that 82% of triads are balanced nonetheless. The full network is plotted in Figure 7. We now apply the SERGM to this static network to test whether structural balance effects can be recovered from it. The SERGM we specify includes edge terms, $GWESF$, $GWESF$ as well as a degree statistic. To show the flexibility with which these statistics can be specified, we include the edge and $GWESF$ terms separately for positive and negative ties, but the $GWESF$ and GWD statistics only for positive ties. For

Table 1: Results of the models.

	Dependence		Independence	
	Coef.	CI	Coef.	CI
Edges +	-8.744	[-12.182,-5.306]	-0.76	[-1.13,-0.39]
Edges -	-1.647	[-2.766,-0.528]	-0.76	[-1.13,-0.39]
<i>GWES</i> ⁺	0.45	[0.015,0.885]	-	
<i>GWES</i> ⁺	0.068	[-0.228,0.364]	-	
<i>GWES</i> ⁻	0.932	[0.616,1.248]	-	
<i>GWD</i> ⁺	5.492	[2.252,8.732]	-	
AIC	139.593		170.355	

the sake of comparison, we also estimate a model that drops all endogenous network terms and hence includes only the two edge terms. Results of both models are presented in Table 1.

3.2 Results and Model Assessment

In Table 1, it is apparent that the fully specified Dependence model has a lower AIC than the Independence model which does not account for endogenous network terms, indicating that it is preferable in terms of performance. Table 1 also offers some evidence that structural balance drives tie formation among the Gahuku-Gama: *GWES*⁺ has a positive effect on positive ties whose 95%-Confidence Intervals clearly exclude zero while its effect on negative ties is very close to zero. This implies that here, friends of friends are indeed more often friends but not less often enemies than one would expect by chance. *GWES*⁺ also exhibits a positive and statistically significant effect, suggesting that subtribes with a common enemy are more likely to share an alliance than in a random network of the same size. Finally, the effect of *GWD*⁺ is also positive and statistically significant, meaning that a subtribe’s probability of gaining a further positive tie increases with the number of such ties it already has.

Figure 8 offers a visual assessment of the goodness-of-fit of the Dependence model. Here, we can see that while the observed network lines up quite well with the simulated ones in terms of Edgewise-Shared Friends, model fit is more problematic for Edgewise-Shared Enemies where the observed network is regularly outside the interquartile range of the simulated networks. Similarly, the model does not do a good job of capturing the observed network’s negative degree distribution. Based on these plots, one may consider re-running the Dependence model while specifying *GWES* for both positive and negative ties and including *GWD*⁻. Nonetheless this application demonstrates the possibility to use the SERGM for the analysis of static networks.

8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

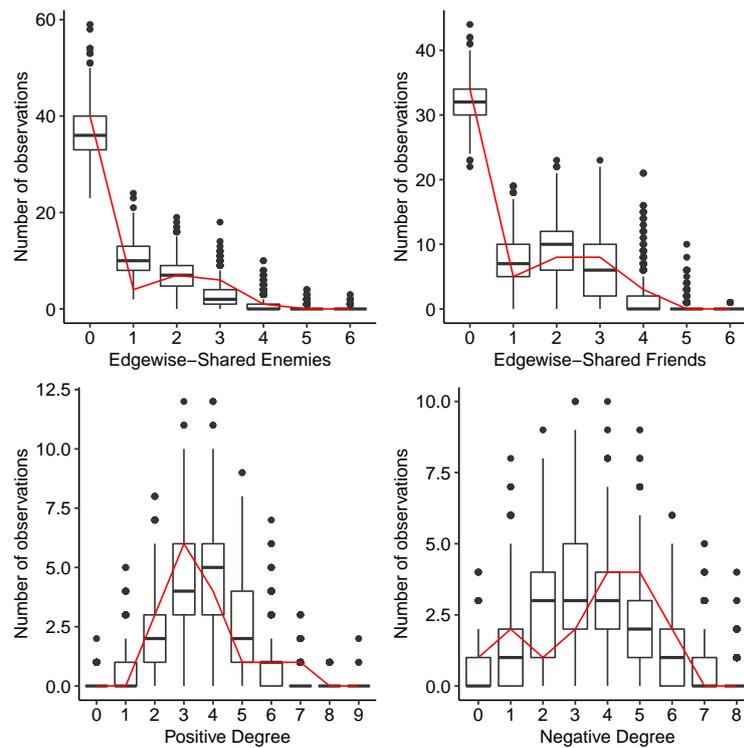


Figure 8: Model Assessment of the New Guinean Highland Tribes.

3.3 MCMC Diagnostics

Finally, we also present the MCMC diagnostics for this additional application. Below are thus shown the MCMC trace plots for all its covariates (Figures 9–10). Overall, these MCMC diagnostics indicate a good convergence of the model.

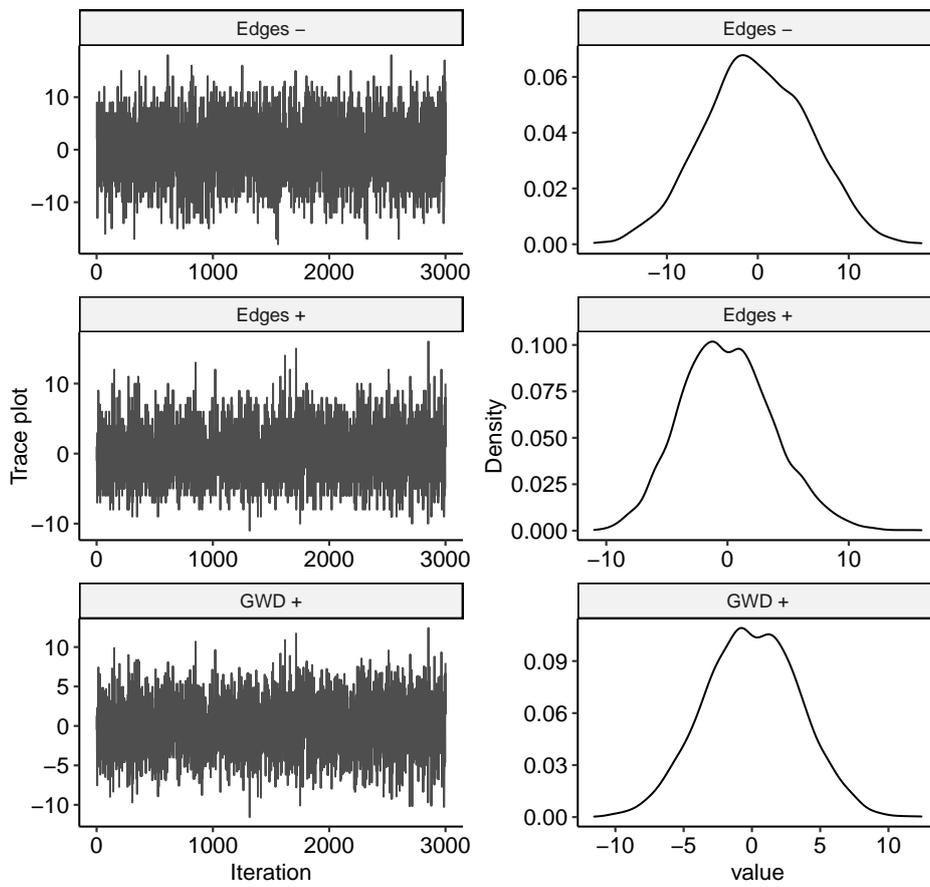


Figure 9: MCMC diagnostics of Dependence Model.

8. Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations

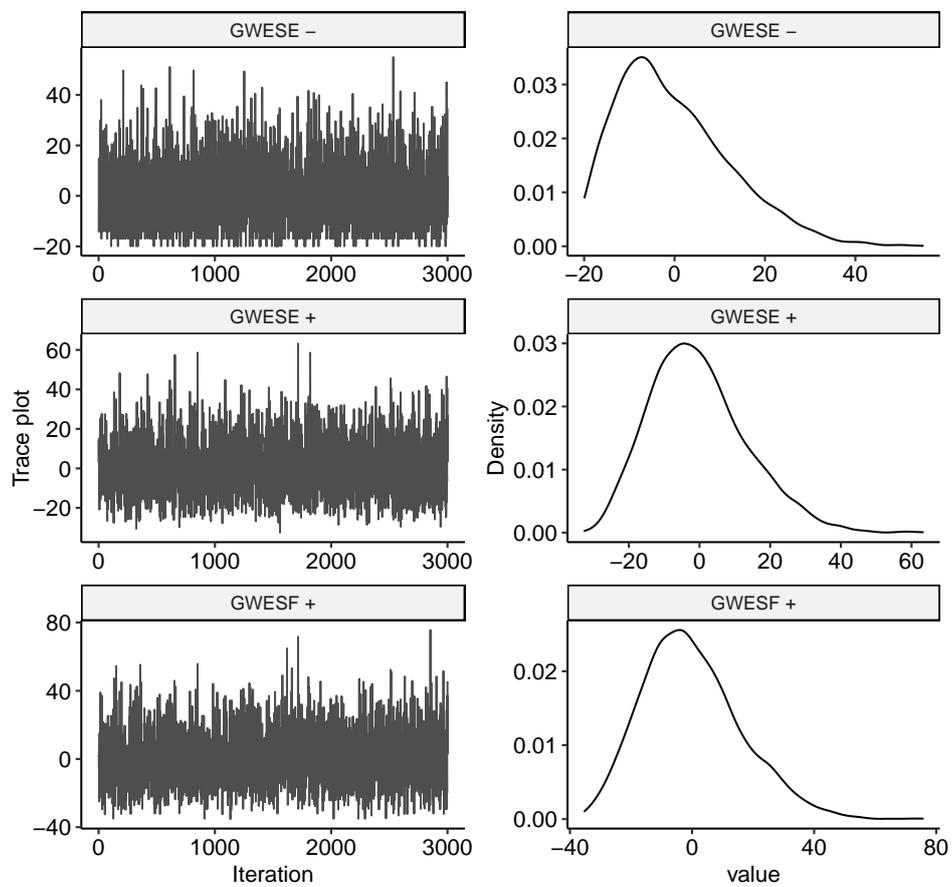


Figure 10: MCMC diagnostics of Dependence Model.

	Cooperation and Conflict	Tribes
Fixed α	$\log(2)$	1.5
Grid size for γ	3,000	2,000
Number of Bridges	16	16
Estimation	Burn-In	10,000
	MCMC Interval	1,000
	M	2,000
	Start Empty	True
Variance	Burn-In	100,000
	MCMC Interval	10,000
	M	3,000
	Start Empty	True
Bridge	Burn-In	10,000
	MCMC Interval	1,000
	M	3,000
	Start Empty	True

Table 2: Setting of the parameters of the fitting of the MCMC estimation procedure. One can define separate configurations for the Gibbs sampler used for the estimation of the parameters, the quantification of their variance, and the evaluation of the AIC.

4 Computational Settings

We provide the set tuning parameters of the MCMC algorithm and the models for both applications in Table 2. We performed sensitivity checks to guarantee that the reported findings do not depend on the fixed parameters. In general, the values should depend on the density (the lower the density, the higher the burn-In and MCMC interval), size (the larger the size, the higher the burn-In and MCMC interval), and the strength of exogenous covariates (the stronger the influence of exogenous factors, the higher the burn-In and MCMC interval) of the network.

Supplementary Material References

- Anders, T., C. J. Fariss, and J. N. Markowitz (2020). Bread before guns or butter: introducing surplus domestic product (sdp). *International Studies Quarterly* 64(2), 392–405.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Gelman, A. and X. L. Meng (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* 13(2), 163–185.
- Hage, P. and F. Harary (1984). *Structural Models in Anthropology*. Cambridge: Cambridge University Press.

- Hummel, R. M., D. R. Hunter, and M. S. Handcock (2012). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics* 21(4), 920–939.
- Hunter, D. R. and M. S. Handcock (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3), 565–583.
- Kinne, B. J. (2020). The defense cooperation agreement dataset (dcad). *Journal of Conflict Resolution* 64(4), 729–755.
- Krivitsky, P. N., A. R. Kuvelkar, and D. R. Hunter (2022). Likelihood-based Inference for Exponential-Family Random Graph Models via Linear Programming. *ArXiv*. 2202.03572.
- Lai, B. and D. Reiter (2000). Democracy, political similarity, and international alliances, 1816-1992. *Journal of Conflict Resolution* 44(2), 203–227.
- Lake, D. A. (2009). *Hierarchy in International Relations*. Ithaca: Cornell University Press.
- Marshall, M. G., T. R. Gurr, and K. Jaggers (2018). Polity iv project: Political regime characteristics and transitions, 1800–2017. *Center for systemic peace*.
- Miller, S. V. (2021). {peacesciencer}: An r package for quantitative peace science research. *Conflict Management and Peace Science (OnlineFirst)*.
- Palmer, G., R. W. McManus, V. D’Orazio, M. R. Kenwick, M. Karstens, C. Bloch, N. Dietrich, K. Kahn, K. Ritter, and M. J. Soules (2021). The mid5 dataset, 2011–2014: Procedures, coding rules, and description. *Conflict Management and Peace Science* 39(4), 470–482.
- Read, K. E. (1954). Cultures of the central highlands, new guinea. *Southwestern Journal of Anthropology* 10(1), 1–43.
- Schoch, D. (2020). *signnet: An R package to analyze signed networks*. <https://github.com/schochastics/signnet>.
- Schvitz, G., L. Girardin, S. Rügger, N. B. Weidmann, L.-E. Cederman, and K. S. Gleditsch (2022). Mapping the international system, 1886-2019: The cshapes 2.0 dataset. *Journal of Conflict Resolution* 66(1), 144–161.
- Singer, J. D., S. Bremer, and J. Stuckey (1972). Capability distribution, uncertainty, and major power war, 1820-1965. In B. Russett (Ed.), *Peace, war, and numbers*, pp. 19–48. Beverly Hills: Sage.
- Warren, T. C. (2016). Modeling the coevolution of international and domestic institutions. *Journal of Peace Research* 53(3), 424–441.

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

Contributing article

Fritz, C., De Nicola, G., Kevork, S., Harhoff, D. and Kauermann, G. (2022) Modelling the large and dynamically growing bipartite network of German patents and inventors. *Under review in the Journal of the Royal Statistical Society. Series A (Statistics in Society)*.

Replication code

The complete replication code for this article is available under <https://github.com/corneliusfritz/Replication-Modelling-large-and-dynamically-growing-bipartite-networks>. A R package `ergm.patent` implementing the novel suite of covariates is attached to the repository.

Author contributions

The idea to use patent data came from Cornelius Fritz, Giacomo De Nicola, Sevag Kevork, and Göran Kauermann. Göran Kauermann then came up with the general idea to model the patent data as a bipartite network that grows over time. Cornelius Fritz then proposed a novel suite of covariates and adopted the offset adjustment introduced in prior work for bipartite network data. Cornelius Fritz and Giacomo De Nicola then specified the model, and Cornelius Fritz implemented it. The introduction was written jointly by Cornelius Fritz, Giacomo De Nicola, and Göran Kauermann. Section 2, on the other hand, was composed by Giacomo De Nicola, Cornelius Fritz, and Dietmar Harhoff. Cornelius Fritz further wrote Section 3 and composed Section 4 jointly with Giacomo De Nicola. Sevag Kevork contributed to Sections 1 and 3.3. Cornelius Fritz, Giacomo De Nicola, Göran Kauermann, and Dietmar Harhoff were all proofreading the manuscript.

ORIGINAL ARTICLE

Modelling the large and dynamically growing bipartite network of German patents and inventors

Cornelius Fritz¹ | Giacomo De Nicola¹ | Sevag Kevork¹
| Dietmar Harhoff² | Göran Kauermann¹

¹Department of Statistics,
Ludwig-Maximilians-Universität München

²Max Planck Institute for Innovation and
Competition

Correspondence

Cornelius Fritz, Department of Statistics,
Ludwig-Maximilians-Universität München,
Ludwigstr. 33, 80539 Munich, Germany
Email: cornelius.fritz@stat.uni-muenchen.de

Funding information

Munich Center for Machine Learning (MCML)

We analyse the bipartite dynamic network of inventors and patents registered within the main area of electrical engineering in Germany to explore the driving forces behind innovation. The data at hand leads to a bipartite network, where an edge between an inventor and a patent is present if the inventor is a co-owner of the respective patent. Since more than a hundred thousand patents were filed by similarly as many inventors during the observational period, this amounts to a massive bipartite network, too large to be analysed as a whole. Therefore, we decompose the bipartite network by utilising an essential characteristic of the network: most inventors tend to stay active only for a relatively short period, while new ones become active at each point in time. Consequently, the adjacency matrix carries several structural zeros. To accommodate for these, we propose a bipartite variant of the Temporal Exponential Random Graph Model (TERGM) in which we let the actor set vary over time, differentiate between inventors that already submitted patents and those that did not, and account for pairwise statistics of inventors. Our results corroborate the hypotheses that inventor characteristics and knowledge flows play a crucial role in the dynamics of invention.

Keywords – Bipartite networks, Patent collaboration, Temporal exponential random graph models, Inventors, Co-inventorship networks, Knowledge flows

1 | INTRODUCTION

In the social sciences, bipartite networks are often used to represent and study affiliation of the actors to some groups (such as directors on boards, Friel et al., 2016, or football players in teams, Onody and de Castro, 2004) and participation of people to events (such as researchers citing papers, Small, 1973, or actors in movies, Ahmed et al., 2007). Research on bipartite structures was initially focused on unimodal projections of the networks (Breiger, 1974), where we consider two nodes of one type to be tied if they share at least one alter of the other kind. This practice forces the researcher to give priority to one type of node over another and thus comes with a loss of possibly relevant information (Koskinen and Edling, 2012). Direct bipartite network analysis has first been considered in Borgatti and Everett (1997), where an introduction and traditional network analysis techniques are systematically discussed. Latapy et al. (2008) further adjusted known concepts from unipartite networks, such as clustering and redundancy, to the bipartite case, with a focus on large networks.

For this paper, we also consider high-dimensional bipartite networks where actors are related to one another through instantaneous events, which by definition only occur once. In particular, we focus on the network formed by inventors residing in Germany and patents submitted between 1995 and 2015, where a tie between an inventor and a patent is present if the focal individual is listed among the patent’s inventors. The resulting data structure is visualised in Figure 1, where we can assign each patent (or event, in the jargon of bipartite network analysis) to a time point and a set of co-inventors. For instance, inventors A and B filed the joint patent with ID 1. We may represent the bipartite network structure as an adjacency matrix with entries Y_{ij} , where

$$Y_{ij} = \begin{cases} 1 & \text{if actor } i \text{ is on patent ID } j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and $i \in \mathcal{I}$ and $j \in \mathcal{K}$, where we denote the complete set of inventors and patents by \mathcal{I} and \mathcal{K} , respectively. In our example this bipartite network is of massive dimensions, with $|\mathcal{I}| = 78.412$ inventors on a total of $|\mathcal{K}| = 126.388$ filed patents.

The data allow us to gain insight into the dynamics and drivers of innovation, collaboration and knowledge flows in the private sector. Moreover, inventorship status on a patent is legally more binding than authorship of academic papers, suggesting a greater degree of validity of the results of network analysis in this context. The data, however, present some obstacles to their study. First, the complete network is too massive, making analysis with most traditional network techniques prohibitive. Second, the data carry structural zero entries since not all inventors are active during the entire time period between 1995 and 2015. This phenomenon is partially due to the retirement of inventors, who hence have natural “actor mortality”. Concurrently, inventors are often active for a short period before changing careers, thereby ending their patent output and further reinforcing the aforementioned actor mortality. Vice versa, new inventors continuously enter the market by producing their first patent, resulting in what we can call “actor natality” in the network. These aspects imply that the bipartite network matrix (1) contains structural zeros for inventors which are not active at particular time points. To incorporate this feature into a statistical network model, we consider the network dynamically and discretise the time dimension by looking at yearly data, such that time takes values $t = 1, 2, \dots, T$, as sketched in Figure 1. In this context, T denote the number of observed time points. We then allow the actor set to change at each time point. For the adjacency matrix of Figure 1, this leads to the matrix structure in Figure 2, where e.g. inventor A retires after time point $t = 1$ and hence does not take part in the patent market at $t = 2$. We, therefore, define activity sets \mathcal{I}_t to include all actors that are active at time point t . We also define the event set \mathcal{K}_t , which contains all patents submitted in a particular time window. We assume that both sets are known for each time point $t = 1, \dots, T$. Therefore, we decompose the observed massive bipartite network matrix into smaller dimensional bipartite submatrices denoted by

$$\mathbf{Y}_t = (Y_{ij} : i \in \mathcal{I}_t, j \in \mathcal{K}_t), \quad (2)$$

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

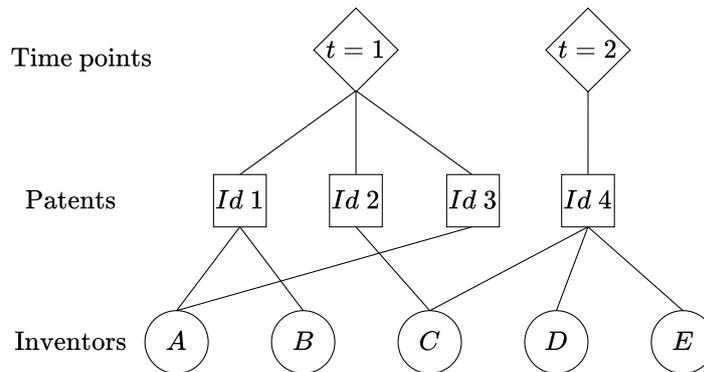


FIGURE 1 Illustrative example of the network structure.

which are visualised for $t = 1$ and 2 by the grey-shaded areas in Figure 2. Instead of modelling the entire bipartite network, we break down our analysis to modelling \mathbf{Y}_t given the previous bipartite networks $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$. Incorporating the varying actor set as such in the analysis allows to structurally account for the observed actor mortality and natality while also making the difficulty of the problem more manageable, thus solving both issues simultaneously.

This change in perspective induces a structure that deviates from conventionally analysed networks. To accommodate for it in a probabilistic modelling framework, we extend the Temporal Exponential Random Graph Model (TERGM, Hanneke et al., 2010) towards dynamic bipartite networks with varying actor sets. For TERGMs, we assume that a discrete Markov chain can describe the generating process of the networks observed over time. The transition probabilities of jumping from one network to another one are determined by an Exponential Random Graph Model (ERGM, Wasserman and Pattison, 1996). ERGMs, on the other hand, were adapted to bipartite data by Faust and Skvoretz (1999), while adjustments to incorporate the model specifications of Snijders et al. (2006) were proposed in Wang et al. (2013). These types of network models were already successfully applied to static (Agnessens and Roose, 2008) as well as dynamic networks (Broekel and Bednarz, 2018).

In addition to the dynamically varying actor set, the network at hand presents another particular feature for which we need to account in the modelling. Collaborations and knowledge flows generally build up over time, rather than being confined to single time points. To adequately represent these mechanisms, we need to include covariate information from the past and on the pairwise level of one alter set in the model, which has not yet been implemented in the bipartite ERGM framework. We, therefore, define and include novel sufficient network statistics in our model to account for this particular kind of dynamic interdependence.

Overall, the contributions of this paper are the following. First, we demonstrate how massive bipartite networks can be broken down in a way that allows their analysis. Secondly, we extend temporal network models towards bipartite network data with varying actor sets. And lastly, with these requisites, we can analyse patent data with respect to innovation dynamics and collaboration in a more refined way than has been feasible to date.

The remainder of the paper is organised as follows: Section 2 gives a literature overview of the research in patent data. In this section, we also describe the data in detail. Section 3 motivates the model and introduces its novelties in more detail from a theoretical perspective. We present the results of our empirical analysis in Section 4, while Section 5 wraps up the paper with some concluding remarks.

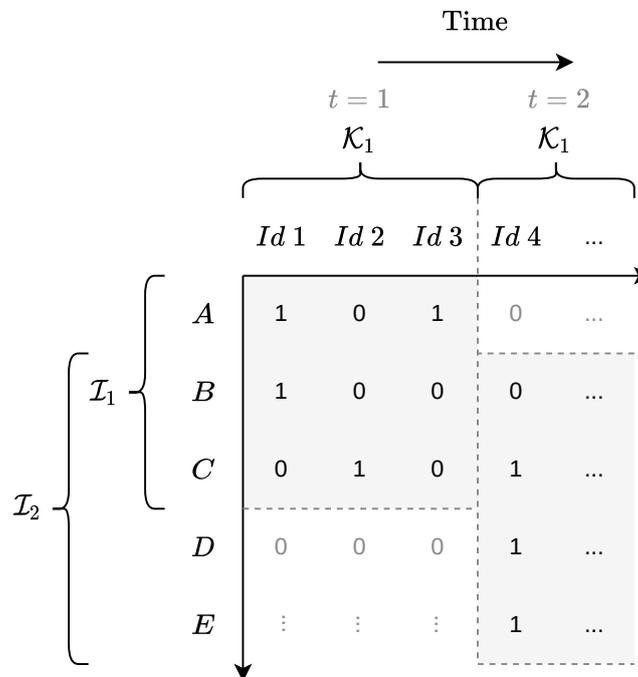


FIGURE 2 Illustrative example of the tripartite network matrix. The two sub-matrices shaded in grey are \mathbf{Y}_1 in the top left and \mathbf{Y}_2 on the bottom right.

2 | PATENT DATA

2.1 | Research in patent data

The analysis of patents and their impact and evolution over time is an important area of current economic research. Hall and Harhoff (2012) provide a general overview of the field and its recent developments. The existence of patents induces a trade-off for society, namely between short-term monopoly rights to the use of inventions as an incentive to invent and early publication of inventions (rather than their secrecy for personal gain by the inventors), which may invite others to build on the patented technology. The study of patents in much of classic economic literature revolves around the economic consequences of their existence from a regulatory standpoint, or, in other words, whether the aforementioned trade-off is worth it for society. Patent data are often also used in innovation research to explore how new technologies develop and spread, which innovation areas are the most active, how innovation areas and sub-areas are connected with one another, and how productive firms or nations are with regards to their patenting output.

Two of the main categories of methods for analysing patent data in this latter regard are keyword-based morphological approaches and network-based approaches. Under keyword-based patent analysis, we understand the process of gaining insights on core technology information of patents through text mining of the document content of each patent. Tseng et al. (2007) review this approach and describe different text mining techniques that conform to the analytical process used by patent analysts, see, e.g., Yoon and Park (2007) or Lee et al. (2009).

In contrast to text mining approaches, network-based patent analysis starts with the idea of constructing a network map of the technology space to understand how technology as a whole behaves and evolves. In this context, patents (or technologies) are often considered as nodes of the network, where an edge links them if they are close to one another according to some proximity measure. Such measure is often based on the citations that occur between patents (Alstott et al., 2017). The study of citation networks has generally been an important area of research at least since the work of Garfield (1955) (see also de Solla Price, 1965 and Egghe and Rousseau, 1990), and the techniques developed for general citation networks can naturally be applied to map patent citation networks (see, e.g., von Wartburg et al., 2005; Li et al., 2007; Verspagen, 2012).

As an alternative to using citation-based approaches for measuring proximity, one can draw a network map of the technology space by focusing on the co-inventorship of patents. Using inventorship data instead of citations to construct the network entails a different perspective on the technological space: The focus is shifted from the content of the patents towards the people coming up with inventions. Thereby, we can gain insights into the network of inventors' underlying collaborative structure and investigate how behaviour differs between and among areas. Co-authorship networks have been extensively studied within the area of research publications (see, e.g., Melin and Persson, 1996 and Newman, 2004). For patent data, it is possible to construct the co-inventorship network in two main ways. One can directly analyse the bipartite network formed by the patents and their inventors, see, e.g., Balconi et al. (2004). Alternatively, one projects the bipartite structure on one of the two modes, which in the context of patent data is usually that of inventors. This entails a network composed only of inventors, in which two nodes are connected if they have at least one patent in common (Ejeremo and Karlsson, 2006; Bauer et al., 2021). Much of the literature in this area utilises such projection since the focus is generally on knowledge flows between inventors, and models for unimodal networks are developed to a greater extent. As explained in the introduction, however, projecting everything on one mode inevitably loses information on the mode that is excluded.

2.2 | Data description

We consider patent applications submitted to the European Patent Office or the German Patent and Trademark Office (Deutsches Patent- und Markenamt) between 1995 and 2015. More specifically, we look at patents filed within the main area of electrical engineering, and for which at least one of the inventors listed on the patent has a residential address in Germany. For assigning each patent to a single time point we use the priority date, i.e., the first-time filing date of a patent (which precedes the publication and the grant date). We focus on electrical engineering as it is one of the largest main areas and as it has seen particularly high growth rates since 2010. Moreover, collaborations between inventors are particularly frequent in this field. For our analyses, we focus on the data starting 2000 and condition on the information from the first five years considered (i.e. from 1995 to 1999) to derive covariates from them. The dataset can be represented in a massive bipartite network, for which the observed adjacency matrix (1) is visualised in Figure 3.

As described in Section 1, we instead consider this a dynamic bipartite network, discretising the time steps yearly such that time takes values $t = 1, 2, \dots, T$. In our notation, $t = 1$ translates to the year 2000. We also allow the actor set to change at each time point so that we end up with T bipartite networks in which the nodes are given by the active inventors at each time point. Resulting from this, we include new inventors that are active for the first time and remove inactive ones from the network at each time point t . The latter point is motivated by the empirical data, which suggests that if previously active inventors don't produce any patents for a long time, it is likely that they will not be active anymore. This phenomenon can stem from a changed career path (moved up to a management position where writing patents is not among the work tasks) or retirement. To this point, we show in Figure 4 the Kaplan-Meier estimate of the time passing between two consecutive patents by the same inventor. As indicated by the dashed grey lines, about 85% of patents by a specific inventor that already had at least one patent are submitted within two years from the previous one. Given this, we define an inventor as active at time t if they had at least one patent in the two years prior to t . Note that by doing so we do not disregard the remaining 15% of the data, but simply label these inventors as

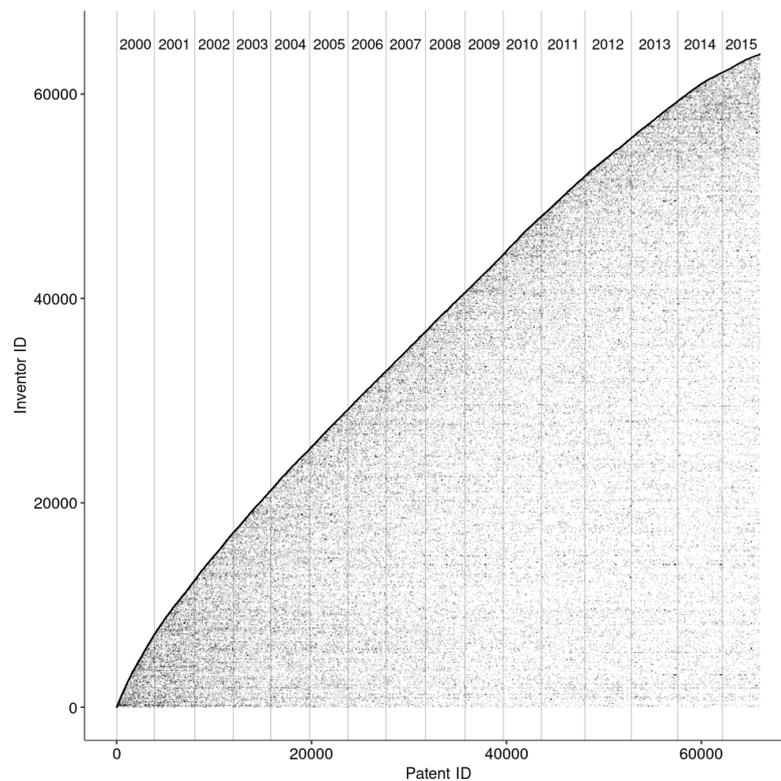


FIGURE 3 Adjacency matrix of the patent-inventor network between 2000 and 2015 with black points in the (i, k) th row indicating that inventor i is a co-owner of patent k . Note that the points are heavily over-represented pixel-wise, hence the network is more sparse than it appears in the plot.

inactive for a specific period, at least until they appear on another patent.

As we are interested in investigating the drivers of patented innovation and inventor collaboration, we exclude patents developed by a single inventor from the modelled patent set. Moreover, we exclude inventors with no address in Germany from the actor set, as they make up less than 1% of the population. In addition to the residence address of each inventor and the date of each patent, we also incorporate information on the gender of each inventor in our model.

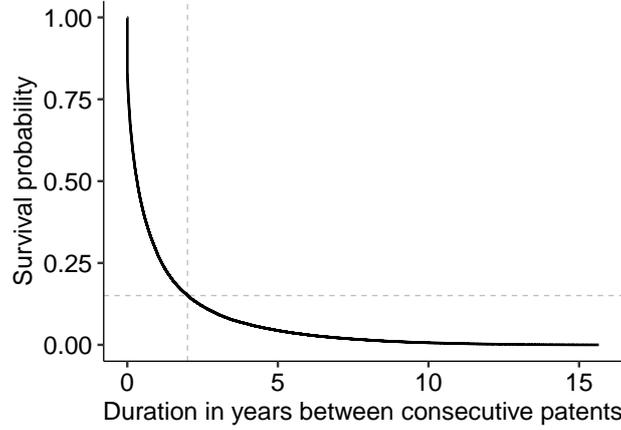


FIGURE 4 Kaplan-Meier estimate of the duration between consecutive patents submitted between 2000 and 2015.

3 | MODELLING PATENT DATA AS BIPARTITE NETWORKS

3.1 | Temporal Exponential Random Graph Models for bipartite networks

Having laid out the available data, we now formulate a generative network model for the bipartite networks at hand. This framework should allow us to differentiate between random and structural characteristics of the network to support or disregard our substantive expectations, such as, for example, whether or not two inventors that teamed up in the past are likely to produce another patent together in the future. To do so we first need to introduce some additional notation. As a general rule, we write \mathbf{Y}_t for the network when viewed as a random variable and $\mathbf{y}_t = (y_{t,ik} : i \in \mathcal{I}_t, k \in \mathcal{K}_t)$ if we relate to the observed counterpart. In this context, $y_{t,ik} = 1$ translates to inventor i being a co-owner of patent k , while $y_{t,ik} = 0$ indicates the contrary. As a result, the observed networks are binary and undirected, i.e. $\mathbf{y}_t \in \{0, 1\}^{|\mathcal{I}_t| \times |\mathcal{K}_t|}$. We denote the space of all networks that could potentially be observed at time point t by \mathcal{Y}_t . For our application, as explained in the previous section, the latter is restricted to only allow for patents which have at least two inventors.

We specify the joint probability for the set of networks through

$$\mathbb{P}_\theta(\mathbf{Y}_1, \dots, \mathbf{Y}_T) = \prod_{t=1}^T \mathbb{P}_\theta(\mathbf{Y}_t | \mathcal{H}_t), \quad (3)$$

where \mathcal{H}_t defines the history, composed of previous bipartite networks $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ and covariates $\mathbf{x}_1, \dots, \mathbf{x}_{t-1}$. The covariates can encompass dyadic and nodal information, but to make the notation less cumbersome, we suppress the explicit inclusion of the covariates in the formulae. We simplify (3) by assuming a fixed time lag, i.e.

$$\mathbb{P}_\theta(\mathbf{Y}_t | \mathcal{H}_t) = \mathbb{P}_\theta(\mathbf{Y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \dots, \mathbf{Y}_{t-s} = \mathbf{y}_{t-s}), \quad (4)$$

for $s \in \mathbb{N}$. The Markov property then allows us to postulate an ERGM for the transition probability (4) in the following form:

$$\mathbb{P}_\theta(\mathbf{Y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \dots, \mathbf{Y}_{t-s} = \mathbf{y}_{t-s}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})\}}{\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s})}, \quad (5)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ is a q -dimensional vector of parameters, $\mathbf{s}() : \mathcal{Y}_t \times \dots \times \mathcal{Y}_{t-s} \rightarrow \mathbb{R}^q$ is the vector of sufficient statistics and $\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s}) := \sum_{\mathbf{y} \in \mathcal{Y}_t} \exp\{\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s})\}$ is a normalising factor. We obtain a canonical exponential family model with known characteristics (Barndorff-Nielsen, 1978), which come in handy when quantifying the uncertainty of the estimates of $\boldsymbol{\theta}$. Note that for the application to patent data, the coefficients governing the transition from one time point to another are not necessarily constant over time due to external shocks, such as for example the dot-com bubble and the 2008 financial crisis, which may affect the activity of inventors. For this reason, we let $\boldsymbol{\theta}$ in (5) flexibly depend on time, and estimate it separately for each time point t . We omit the subscript t from the formulae for notational simplicity though.

Interpreting the coefficients $\boldsymbol{\theta}$ can be done both at the global network level as well on the single tie level. For the former, $\theta_p > 0$ implies that the expected value of the p th statistic of $\mathbf{s}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$ for networks generated from (5) is higher than under a Bernoulli graph, while $\theta_p < 0$ implies that it is lower. In this context, a Bernoulli graph is a simplistic network model where every edge is present with equal probability of 0.5. For the latter, we define so called change statistics, which are the change in the sufficient statistics caused by switching the entry $y_{t,ik}$ from 0 to 1. Formally,

$$\Delta_{ik}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = \mathbf{s}(\mathbf{y}_{t,ik}^+, \dots, \mathbf{y}_{t-s}) - \mathbf{s}(\mathbf{y}_{t,ik}^-, \dots, \mathbf{y}_{t-s}), \quad (6)$$

where $\mathbf{y}_{t,ik}^+$ is the network \mathbf{y}_t with entry $y_{t,ik}$ fixed at 1, while the entry is set to 0 in $\mathbf{y}_{t,ik}^-$. For each possible inventor-patent connection, we can then state the corresponding probability conditionally on the remaining bipartite network denoted by $\mathbf{y}_{t,ik}^C$, i.e. the complete network \mathbf{y}_t excluding the single entry $y_{t,ik}$. This leads to

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{t,ik} = 1 | \mathbf{y}_{t,ik}^C = \mathbf{y}_{t,ik}^C) = \frac{\exp\{\boldsymbol{\theta}^\top \Delta_{t,ik}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})\}}{1 + \exp\{\boldsymbol{\theta}^\top \Delta_{t,ik}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})\}}. \quad (7)$$

Through this expression we can relate $\boldsymbol{\theta}$, the canonical parameter of (5), to the conditional probability of inventor i to be co-owner of patent k . We can thereby derive an interpretation of the coefficients reminiscent of the common logistic regression: if adding the tie $y_{t,ik}$ to the network raises the p th entry of $\Delta_{t,ik}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$ by one unit, the conditional log-odds of $Y_{t,ik}$ are ceteris paribus altered by the additive factor θ_p (Goodreau et al., 2009).

3.2 | Sufficient statistics for bipartite patent data

The main ingredient of model (5) is the set of sufficient statistics, which translates to a particular dependence structure assumed for the edges in the observed bipartite network (Wang et al., 2013). A statistic that is typically included is the number of edges at time point t , i.e. $s_{\text{edges}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = |\mathbf{y}_t|$, which can be comprehended as the equivalent of an intercept term in standard regression models (Goodreau et al., 2009). As we are in a dynamic setting in which additional information on past networks is available, we can define statistics that depend on the past networks, such as the number of patents in the previous s years for each actor active at time point t :

$$s_{\text{pastpatent}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \sum_{u=t-s}^{t-1} \sum_{l \in \mathcal{K}_u} y_{u,il} \quad (8)$$

As the patent network presents some particular dependence structures, novel types of statistics are needed, which we describe in the following.

3.2.1 | Pairwise statistics of inventors

One drawback of representing the patent data as a bipartite adjacency matrix instead of the one-mode-projected version is that incorporating information on the pairwise inventor-to-inventor level is not straightforward. We therefore introduce assortative two-star statistics extending the work of Bomirya (2014, Chapter 2) on homophily, which is defined as the mechanism driving ties between similar individuals (McPherson et al., 2001), for bipartite networks. We take the patent-based two-star statistic as starting point, which for \mathbf{y}_t is defined by

$$s_{\text{twostar.patent}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \right). \quad (9)$$

The tendency to interact with one another is often based on the similarity of a factor variable $u_t = (u_{t,i}; i \in \mathcal{I}_t)$. We therefore define the indicator matrix $\mathbf{x}_t \in \{0, 1\}^{|\mathcal{I}_t| \times |\mathcal{I}_t|}$ with entries $x_{t,ij} = \mathbb{1}(u_{t,i} = u_{t,j})$. In line with Bomirya (2014, Chapter 2), this allows to augment the two-star statistic (9) in the form

$$s_{\text{homophily.x}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} x_{t,ij} \right). \quad (10)$$

Next, we generalise (10) by not restricting ourselves to any particular definition of \mathbf{x}_t but letting the matrix be an arbitrary function of the networks from the past s years and other exogenous information. To further correct for different sizes of patents, i.e. the number of inventors co-owning the patent, we normalise the statistic by the degree of each patent, whereby the novel statistic is defined through:

$$s_{\text{assort.x}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{ik} \left(100 \times \frac{\sum_{j \neq i} y_{t,jk} x_{t,ij}}{\sum_{j \neq i} y_{t,jk}} \right) \quad (11)$$

The corresponding change statistic for an edge between inventor i and patent k is then

$$\Delta_{ik, \text{assort.x}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = 100 \times \frac{\sum_{j \neq i} y_{t,jk} x_{t,ij}}{\sum_{j \neq i} y_{t,jk}}, \quad (12)$$

which can simply be interpreted as the percentage of inventors on patent k that match with inventor i in matrix \mathbf{x} . We multiply the statistic by 100, which does not affect the model itself but eases interpretation (as a unit increase is now equivalent to a single percentage change). To give an example of a statistic of this type, we can combine (12) with matrix \mathbf{x}_t^P , for which entry $x_{t,ij}^P$ is 1 if inventor i and j already had a joint patent in the last s years and 0 otherwise. The resulting statistic measures how previous collaboration among inventors affects the propensity of future collaboration. More examples for such statistics are provided in Section 4.

3.2.2 | Node set statistics

As a result of the actor natality and mortality described in the introduction, we can split the set of inventors \mathcal{I}_t at each time step $t \in \mathcal{T}$ into new inventors with their first patent in t , $\mathcal{I}_t^+ = \{i \in \mathcal{I}_t; \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik} = 0\}$, and inventors that were already active prior to t , $\mathcal{I}_t^- = \{i \in \mathcal{I}_t; \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik} > 0\}$. We here use the term ‘‘new inventors’’ for actors in \mathcal{I}_t^+ and ‘‘experienced inventors’’ for those in \mathcal{I}_t^- . We can then define $\mathbf{y}_t^+ = (y_{t,ik})_{i \in \mathcal{I}_t^+, k \in \mathcal{K}_t}$ and $\mathbf{y}_t^- = (y_{t,ik})_{i \in \mathcal{I}_t^-, k \in \mathcal{K}_t}$ to be the sub-networks of \mathbf{y}_t made up of new and experienced inventors, respectively.

As it turns out, statistics on past behaviour such as (8) are not meaningful for inventors from \mathcal{I}_t^+ , since no historical data is available for those inventors at time t . To account for this, we decompose the statistics $\mathbf{s}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$ into three types of terms, namely $\mathbf{s}^+(\mathbf{y}_t^+)$, $\mathbf{s}^-(\mathbf{y}_t^-, \dots, \mathbf{y}_{t-s}^-)$, and $\mathbf{s}^\pm(\mathbf{y}_t)$, which are defined as statistics that only relate to either \mathbf{y}_t^+ , \mathbf{y}_t^- and past networks or the full set of inventors \mathbf{y}_t , respectively. Defining the corresponding coefficients $(\theta^+, \theta^-, \theta^\pm)$ and change statistics $(\Delta_{t,ik}^+, \Delta_{t,ik}^-, \Delta_{t,ik}^\pm)$ accordingly yields

$$\mathbb{P}_{\theta} \left(Y_{t,ik} = 1 \mid \mathbf{Y}_{t,ik}^C = \mathbf{y}_{t,ik}^C \right) = \begin{cases} \pi_{t,ik}^+(\mathbf{y}_t), & \text{if } i \in \mathcal{I}_t^+ \text{ (new inventor)} \\ \pi_{t,ik}^-(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}), & \text{if } i \in \mathcal{I}_t^- \text{ (experienced inventor)} \end{cases} \quad (13)$$

where $\pi_{t,ik}^+(\mathbf{y}_t)$ and $\pi_{t,ik}^-(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$ are given by

$$\begin{aligned} \pi_{t,ik}^+(\mathbf{y}_t) &= \frac{\exp\{(\theta^+)^\top \Delta_{t,ik}^+(\mathbf{y}_t^+) + (\theta^\pm)^\top \Delta_{t,ik}^\pm(\mathbf{y}_t)\}}{1 + \exp\{(\theta^+)^\top \Delta_{t,ik}^+(\mathbf{y}_t^+) + (\theta^\pm)^\top \Delta_{t,ik}^\pm(\mathbf{y}_t)\}} \\ \pi_{t,ik}^-(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) &= \frac{\exp\{(\theta^-)^\top \Delta_{t,ik}^-(\mathbf{y}_t^-, \dots, \mathbf{y}_{t-s}^-) + (\theta^\pm)^\top \Delta_{t,ik}^\pm(\mathbf{y}_t)\}}{1 + \exp\{(\theta^-)^\top \Delta_{t,ik}^-(\mathbf{y}_t^-, \dots, \mathbf{y}_{t-s}^-) + (\theta^\pm)^\top \Delta_{t,ik}^\pm(\mathbf{y}_t)\}}. \end{aligned}$$

As an example, for the common edge statistic $s_{\text{edges}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$, the aforementioned decomposition means we can define $s_{\text{New}}(\mathbf{y}_t^+) = |\mathbf{y}_t^+|$ and $s_{\text{Experienced}}(\mathbf{y}_t^-, \dots, \mathbf{y}_{t-s}^-) = |\mathbf{y}_t^-|$, to allow for new and experienced inventors to generally have a different propensity to be part of a patent. Note that the splitting of the node set as in (13) does not assume any (in)dependence structure between \mathbf{Y}_t^+ and \mathbf{Y}_t^- , but rather serves as an aid to specify additional terms and interpret the coefficients at a finer level, as just exemplified for the edge statistic.

3.2.3 | Adjustment for varying network size:

As argued in Krivitsky et al. (2011), the task of comparing estimated coefficients of two models with identical specifications but different network sizes is non-trivial. This behaviour is due to the fact that including the edge count statistic from the previous paragraph in a TERGM assumes density invariance as the network grows. This characteristic seldom holds for real-world networks as it implies a linearly growing mean degree of all involved actors. In the case of our longitudinal patent network, the number and composition of inventors and patents change from year to year, thus correcting for this is of practical importance to be able to compare coefficient estimates at different time points. To solve the issue, we follow the suggestion of Krivitsky et al. (2011) and incorporate the offset term $\frac{1}{|\mathcal{I}_t^+| |\mathcal{K}_t|}$ to achieve asymptotically constant mean-degree scaling as the composition of inventors and patents change over time.

3.3 | Estimation and inference

We now seek to estimate the parameter θ , by maximising the logarithmic likelihood constructed from (5) for the transition between time points $t-1$ and t . To do so, we follow the Markov Chain Monte Carlo Maximum Likelihood Estimation procedure introduced by Geyer and Thompson (1992) and adapted to ERGMs by Hunter and Handcock (2006). In our application, we repeat this for each available time step $t = 1, \dots, T$.

First, note that subtracting any constant from the logarithmic likelihood constructed from (5) does not change its maximum. We can therefore subtract the logarithmic likelihood evaluated at an arbitrary value of the parameter θ , i.e. θ_0 , which yields the

equivalent objective function

$$l(\theta) - l(\theta_0) = (\theta - \theta_0)^\top \mathbf{s}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) - \log \left(\mathbb{E}_{\theta_0} \left(\exp \{ (\theta - \theta_0)^\top \mathbf{s}(\mathbf{Y}_t, \dots, \mathbf{Y}_{t-s}) \} \right) \right), \quad (14)$$

where $\mathbb{E}_{\theta}(f(\mathbf{X}))$ is the expected value of random variable \mathbf{X} characterised by parameter θ and transformed through the arbitrary function $f(\cdot)$. As described in Hunter and Handcock (2006), one can evaluate this objective function by approximating the expected value by generating random networks $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(M)}$ from (5) under θ_0 . In particular, we approximate the expected value in (14) through a Monte Carlo quadrature:

$$\mathbb{E}_{\theta_0} \left(\exp \{ (\theta - \theta_0)^\top \mathbf{s}(\mathbf{Y}_t, \dots, \mathbf{Y}_{t-s}) \} \right) \approx \frac{1}{M} \sum_{m=1}^M \exp \left\{ (\theta - \theta_0)^\top \mathbf{s}(\mathbf{y}^{(m)}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s}) \right\} \quad (15)$$

For sufficiently large M the convergence of this expectation is guaranteed, and we can plug (16) into (14) and apply Newton-Raphson type of methods to maximise it with respect to θ . Sampling from a probability distribution with intractable normalisation constant, such as (5), is achieved by a Metropolis-Hastings algorithm. In particular, we first sample an edge, defined as the tuple (i, k) , at random, and consecutively toggle the corresponding entry of \mathbf{Y}_t from 0 to 1 with probability equal to (7) (for more details see Hunter et al., 2013). Due to the large size of the patent networks, we start with the observed network, propose 15.000 of such changes and then stop the Markov chain. This procedure is hence equivalent to contrastive divergence as introduced by Hinton (2002) and adapted to ERGMs by Krivitsky (2017).

Inference on the estimates is drawn based on the Fisher matrix $\mathbf{l}(\theta)$, which equals the variance of the sufficient statistics for exponential family distributions (Wassennan, 2004). Thus, we can approximate the Fisher matrix through

$$\widehat{\mathbf{l}}(\theta) = \text{Var}_{\theta}(\mathbf{s}(\mathbf{Y}_t, \dots, \mathbf{Y}_{t-s})) \approx \frac{1}{M} \sum_{m=1}^M \left(\mathbf{s}(\mathbf{y}^{(m)}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s}) - \bar{\mathbf{s}}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}) \right) \left(\mathbf{s}(\mathbf{y}^{(m)}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s}) - \bar{\mathbf{s}}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}) \right)^\top$$

where $\bar{\mathbf{s}}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}) = \frac{1}{M} \sum_{m=1}^M \mathbf{s}(\mathbf{y}^{(m)}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s})$ is the vector of average of the sufficient statistics from the simulated networks $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}$, which are, in turn, drawn from the fitted model with parameter θ set to its maximum likelihood estimate.

4 | APPLICATION

We can now present the results of the application of our model to the patent data introduced in Section 2. For each statistic included in the model, we first explain its meaning and subsequently interpret the corresponding estimated coefficient. Further details on the specification of each sufficient statistic can be found in Appendix A. MCMC diagnostics, and goodness-of-fit assessments as proposed by Hunter et al. (2008) are provided in the Supplementary Material.

4.1 | Network effects

Propensity to invent: To account for the changing activity levels over time, we incorporate a statistic that counts how many edges are in the network. Following Section 3.2, we split this term into separate statistics for experienced and new inventors. Heuristically, one can interpret the corresponding coefficients as the general propensity to form ties, i.e. participate in a patent,

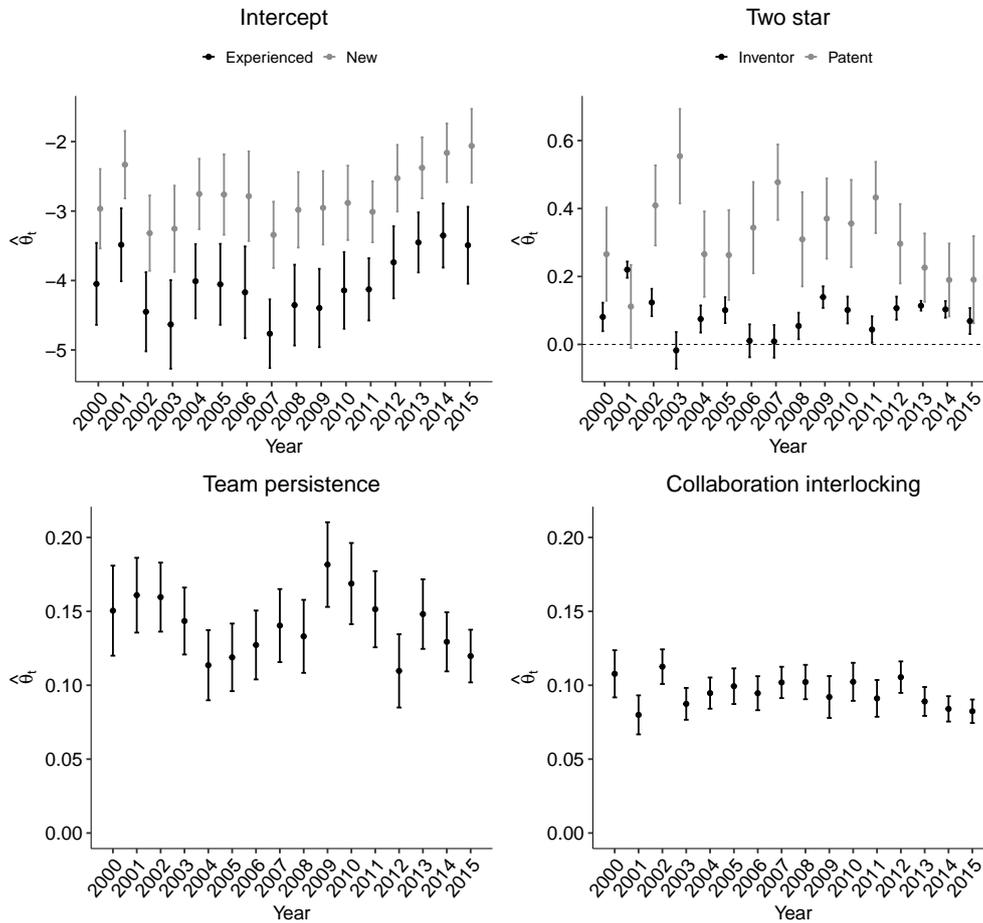


FIGURE 5 Estimated time-varying coefficients regarding the propensity to invent, two-star statistics, team persistence and collaboration interlocking.

for the two inventor sets, respectively. The plot of the corresponding estimates is shown in the upper left panel of Figure 5. It exhibits a different level of activity for new and experienced inventors. We expect this by design, as new inventors enter the network precisely because they are active at time t , while experienced ones might only have been active in the past. Overall, we observe a steady increase in activity in the network over time from 2008 onward for both sets of inventors.

Two-star statistics: Two-star statistics relate to the concept of centrality (Wasserman and Faust, 1994). For bipartite networks, they can be defined with respect to each of the two modes (inventors and patents, respectively). For inventors the statistic is given in Appendix A and expresses whether inventor i is more or less likely to invent an additional patent in year t , given that he/she is (co-)owner of at least another patent in that year. For patents, the statistic relates to the number of inventors per patent and is given in (9). The top right panel of Figure 6 depicts the two corresponding estimates. For inventors, the estimates take for most time points small positive values without much temporal variation. This indicates a slight tendency towards centralisation for inventors, i.e. inventors aiming to submit multiple patents per year. For patents the corresponding

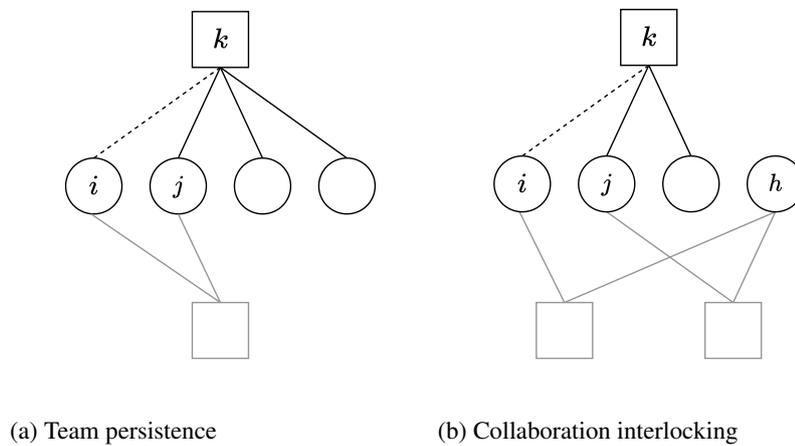


FIGURE 6 Illustration of the change statistics related to assortative network statistics for team persistence (a) and collaboration interlocking (b). Circles represent inventors, and squares are patents. The dashed line indicates a possible edge at time point t , while black lines represent edges given at time point t . Grey lines, on the other hand, display past connections, and grey squares stand for past patents.

two-star estimates are larger, i.e. patents tend to be owned by multiple inventors. The two star effect slowly decreases since 2011, meaning that the number of owners per patent is getting smaller. The variance for the estimated two-star patent effect is generally larger than the estimate of the corresponding two-star inventor effect, which stems from the fact that there are fewer patents than inventors in a single year.

Team persistence: Most patented inventions are the result of team work (Giuri et al., 2007), which leads to the build-up of valuable team-specific capital (Jaravel et al., 2018). We therefore expect past collaboration to positively affect the propensity for two inventors to collaborate again. We include the team persistence statistic based on the pairwise statistics of inventors proposed in 3.2 in the model to account for this effect. The statistic, which could also be defined as “repetition statistic”, is visually represented in Figure 6 (a), and rests on the definition of matrix \mathbf{x}_t^p , whose (i, j) th entry is 1 if inventor i and j have already co-invented a patent in the previous five years and 0 otherwise. The bottom left panel of Figure 5 depicts the corresponding coefficient estimate, which is significantly positive over time. This finding corroborates our anticipations that, controlling for the other factors, two inventors are more likely to jointly produce a patent if they already worked on an invention together in the past. Hence teams of inventors play an important role in patent creation.

Collaboration interlocking: In addition to investigating the persistence of collaborations, it is also of interest to understand how having had a common partner in the past influences the tendency to develop a patent together in the present. We account for this by including the collaboration interlocking statistic in our model. By common partners we are referring to actors such as inventor h for inventors i and j in Figure 6 (b). We define the statistic again by pairwise statistics of inventors through the matrix \mathbf{x}_t^{qI} , encoding in the (i, j) th entry the binary information if inventors i and j have at least one common partner or not. The related coefficient estimates are shown in the bottom right panel of Figure 5, where we notice that the estimate attains significantly positive values throughout the observational period. This result suggests that if two inventors i and j both had a patent with the same inventor h in the past, they are generally more likely to co-invent in the future. The finding holds controlling for all other features in our model (including the previously described team persistence statistic), and can be viewed as akin to triadic closure in unimodal networks, i.e. “a collaborator of my collaborator is more likely to become my collaborator”. The result thus supports

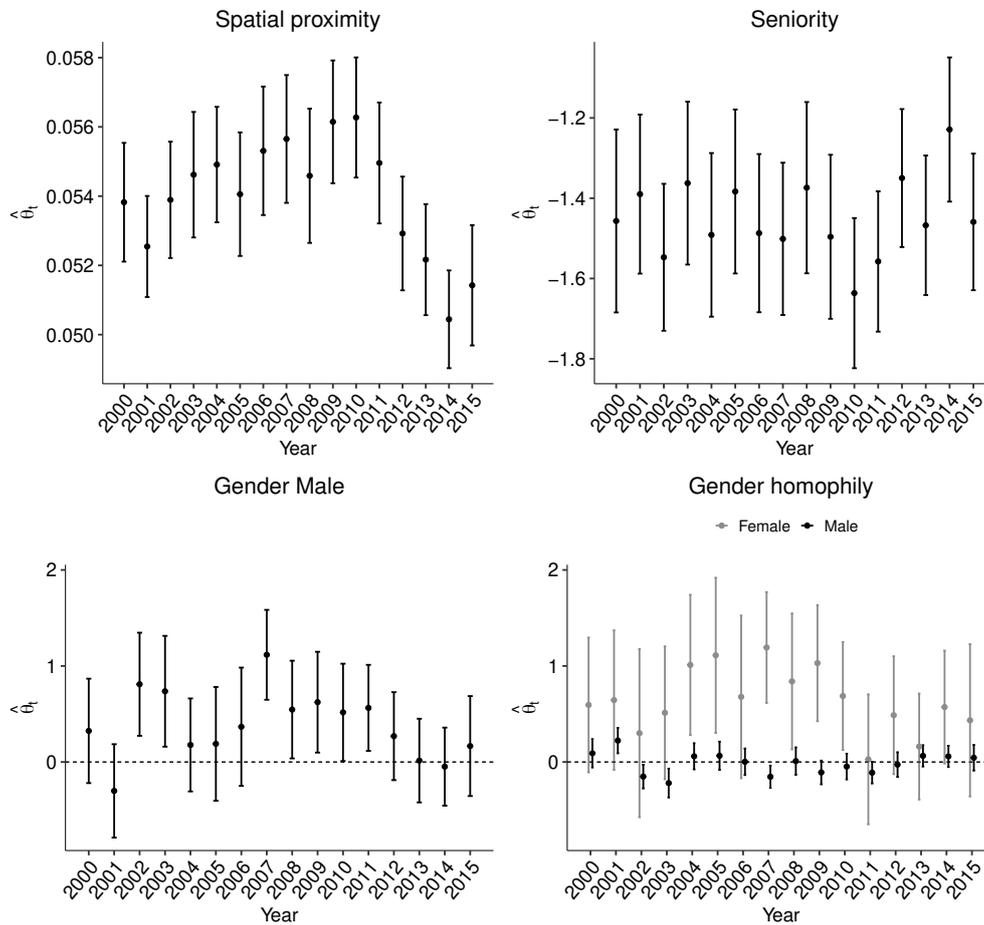


FIGURE 7 Estimated time-varying coefficients regarding the spatial proximity, seniority, and gender of inventors.

the idea that the creation of inventor teams is often promoted via common colleagues and that informal knowledge flows are key to the invention process (see Giuri and Mariani, 2013 and references cited therein).

4.2 | Effects of inventor-specific covariates

Spatial proximity: Many patents are created in a workplace environment (Giuri et al., 2007). For this reason, we would expect inventors that live close to each other to be more likely to invent together. Moreover, there is empirical evidence that collaboration is more likely between inventors that live close to one another even if they do not share the same employer (e.g., Crescenzi et al., 2016). For these reasons, we include a spatial proximity statistic in our model, where we define spatial proximity as living within a range of 50 km. We encode this proximity information in a binary matrix \mathbf{x}^{SP} and incorporate it in the model as a pairwise statistic of inventors. The top left panel of Figure 7 depicts the estimated coefficients for the statistic. The positive values attained over time confirm that inventors living near each other have a higher chance to collaborate. We can also see that the effect goes

down over time from 2010 onward; this makes sense in an increasingly interconnected society, where more and more connections are formed through the web in addition to physical ones.

Seniority: The top right panel in Figure 7 depicts the effect of the number of previously owned patents by each inventor in the past five years. The corresponding statistic can be viewed as a measure of inventor seniority, where inventors with more patents in the past are considered to be senior. Note that this statistic is only computed for the set of inventors which were previously active in the network (experienced inventors), as for new inventors it would trivially be a structural zero. The negative coefficient estimate here suggests that, conditional on all other statistics included in the model, senior inventors have a lower propensity to create new patents. Prior research has shown that career dynamics of inventors are complex as economic opportunities, productivity and personal preferences interact (see, e.g., Allen and Katz, 1992 and Bell et al., 2019). But our results would be consistent with earlier results that with greater seniority, inventors take over managerial responsibilities within the same firm, or that high visibility of their invention output also leads them to move to new employers and tasks, thus lowering (or halting) their invention output.

Gender and gender homophily: Another variable of interest in the realm of innovation research is gender. Many researchers have expressed concerns about the extremely low representation of women among inventors (typically far less than 10 %) and possible wage discrimination, see, e.g., Hoisl and Mariani (2017) and Jensen et al. (2018). These studies made gender an important topic in innovation economics. We incorporate gender in our model in two ways, i.e. as a main effect and as homophily effect introduced in (10). The two plots at the bottom of Figure 7 show the effects of gender on the propensity to create patents (left) and on homophily, i.e. the tendency of inventing together with people of the same gender (right). Note that both effects need to be interpreted keeping in mind that the vast majority of the actors in the network are male (96%). From the plot on the bottom left, we can see how, while male inventors seem to be slightly more active, all in all male and female inventors did not show significant differences in their propensity to invent. Note that this holds given the inclusion of those inventors in the network, i.e. given that they were already inventors. The gender homophily plot shows different results; here we see that, while male inventors seem to have the same likelihood to form patents with both genders, female inventors tend to have more collaborations with other females than with males. While the effect is quite sizeable in absolute value, the uncertainty here is considerable given the small number of female actors in the network. Still, we can see this as weak evidence for a gender homophily effect for female inventors. These results are consistent with earlier findings by Whittington (2018) who studies the role of gender in life science inventor teams.

5 | DISCUSSION

This paper analyses a massive bipartite network consisting of all inventors and collaborative patents submitted between 1995 and 2015 in electrical engineering. To account for the sheer size of the complete network and the structural zeros in the related bipartite adjacency matrix, we suggested a temporal decomposition of the data into multiple smaller networks. Guided by substantive questions posed by innovation research, we then proposed a set of novel bipartite network statistics focused on gender issues, team persistence, collaboration interlocking and spatial proximity.

Time-varying actor sets due to actor mortality and natality are often observed in networks beyond the realm of patent data. For instance, scientific collaboration behaves similarly, as many PhD students do not pursue an academic career and hence have a short lifespan in the scientific collaboration network. At the same time, new PhD students continuously enter the scientific world. Therefore, we argue that the proposed temporal decomposition and the novel class of network terms can be employed in other settings.

In addition to the methodological contributions, our study offers several empirical findings. In particular, we show how spatial proximity, team work and interlocking of collaborations each have a positive impact on the output of inventors. We also

demonstrate how inventors' characteristics, such as gender and seniority, play a significant role in the process.

All in all, our application to inventor teams presents an alternative to classical forms of analysis of patenting and inventorship networks. While prior studies are almost exclusively focused on analysing the underlying mechanisms one at a time, we model them simultaneously in the framework of bipartite networks. We argue that this can provide an effective alternative to classical forms of regression-based analysis of this important phenomenon.

FUNDING

The work has been partially supported by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibility for its content.

REFERENCES

- Agneessens, F. and H. Roose (2008). Local structural properties and attribute characteristics in 2-mode networks: p^* models to map choices of theater events. *Journal of Mathematical Sociology* 32(3), 204–237.
- Ahmed, A., V. Batagelj, X. Fu, S.-H. Hong, D. Merrick, and A. Mrvar (2007). Visualisation and analysis of the internet movie database. In *6th International Asia-Pacific Symposium on Visualization*, pp. 17–24. IEEE.
- Allen, T. J. and R. Katz (1992). Age, education and the technical ladder. *IEEE Transactions on Engineering Management* 39(3), 237–245.
- Alstott, J., G. Triulzi, B. Yan, and J. Luo (2017). Mapping technology space by normalizing patent networks. *Scientometrics* 110(1), 443–479.
- Balconi, M., S. Breschi, and F. Lissoni (2004). Networks of inventors and the role of academia: An exploration of Italian patent data. *Research Policy* 33(1), 127–145.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Bauer, V., D. Harhoff, and G. Kauermann (2021). A smooth dynamic network model for patent collaboration data. *ASIA Advances in Statistical Analysis*, 1–20.
- Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019). Do tax cuts produce more einsteins? the impacts of financial incentives versus exposure to innovation on the supply of inventors. *Journal of the European Economic Association* 17(3), 651–677.
- Bomiriya, R. P. (2014). *Topics in exponential random graph modeling*. Phd thesis, Pennsylvania State University.
- Borgatti, S. P. and M. G. Everett (1997). Network analysis of 2-mode data. *Social Networks* 19(3), 243–269.
- Breiger, R. (1974). Duality of persons and groups. *Social Forces* 52(2), 181–190.
- Broekel, T. and M. Bednarz (2018). Disentangling link formation and dissolution in spatial networks: An Application of a Two-Mode STERGM to a Project-Based R&D Network in the German Biotechnology Industry. *Networks and Spatial Economics* 18(3), 677–704.
- Crescenzi, R., M. Nathan, and A. Rodríguez-Pose (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy* 45(1), 177–194.
- de Solla Price, D. J. (1965). The science of science. *Bulletin of the Atomic Scientists* 21(8), 2–8.
- Egghe, L. and R. Rousseau (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.
- Ejermo, O. and C. Karlsson (2006). Interregional inventor networks as studied by patent coinventorships. *Research Policy* 35(3), 412–430.

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

- Faust, K. and J. Skvoretz (1999). Logit models for affiliation networks. *Sociological Methodology* 31(1), 253–280.
- Friel, N., R. Rastelli, J. Wyse, and A. E. Raftery (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences* 113(24), 6629–6634.
- Garfield, E. (1955). Citation indexes for science. *Science* 122(3159), 108–111.
- Geyer, C. J. and E. A. Thompson (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 54(3), 657–699.
- Giuri, P. and M. Mariani (2013). When distance disappears: Inventors, education, and the locus of knowledge spillovers. *Review of Economics and Statistics* 95(2), 449–463.
- Giuri, P., M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes, A. Geuna, R. Gonzales, D. Harhoff, K. Hoisl, C. Le Bas, A. Luzzi, L. Magazzini, L. Nesta, Önder Nomaler, N. Palomeras, P. Patel, M. Romanelli, and B. Verspagen (2007). Inventors and invention processes in europe: Results from the patval-eu survey. *Research Policy* 36(8), 1107–1127.
- Goodreau, S., J. A. Kitts, and M. Morris (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography* 46(1), 103–125.
- Hall, B. H. and D. Harhoff (2012). Recent research on the economics of patents. *Annual Review of Economics* 4(1), 541–565.
- Hanneke, S., W. Fu, and E. P. Xing (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* 4, 585–605.
- Hinton, G. E. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation* 14(8), 1771–1800.
- Hoisl, K. and M. Mariani (2017). It’s a man’s job: Income and the gender gap in industrial research. *Management Science* 63(3), 766–790.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481), 248–258.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2013). Ergm.userterms: A template package for extending statnet. *Journal of Statistical Software* 52(2), 1–25.
- Hunter, D. R. and M. S. Handcock (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3), 565–583.
- Jaravel, X., N. Petkova, and A. Bell (2018). Team-specific capital and innovation. *American Economic Review* 108(4-5), 1034–73.
- Jensen, K., B. Kovács, and O. Sorenson (2018). Gender differences in obtaining and maintaining patent rights. *Nature Biotechnology* 36(4), 307–309.
- Koskinen, J. and C. Edling (2012). Modelling the evolution of a bipartite network—peer referral in interlocking directorates. *Social Networks* 34(3), 309–322.
- Krivitsky, P. N. (2017). Using contrastive divergence to seed monte carlo MLE for exponential-family random graph models. *Computational Statistics and Data Analysis* 107, 149–161.
- Krivitsky, P. N., M. S. Handcock, and M. Morris (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology* 8(4), 319–339.
- Latapy, M., C. Magnien, and N. Del Vecchio (2008). Basic notions for the analysis of large two-mode networks. *Social Networks* 30(1), 31–48.
- Lee, S., B. Yoon, and Y. Park (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* 29(6-7), 481–497.

- Li, X., H. Chen, Z. Huang, and M. C. Roco (2007). Patent citation network in nanotechnology (1976–2004). *Journal of Nanoparticle Research* 9(3), 337–352.
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1), 415–444.
- Melin, G. and O. Persson (1996). Studying research collaboration using co-authorships. *Scientometrics* 36(3), 363–377.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* 101(suppl 1), 5200–5205.
- Onody, R. N. and P. A. de Castro (2004). Complex network study of brazilian soccer players. *Physical Review E* 70(3), 037103.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24(4), 265–269.
- Snijders, T. A., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology* 36(1), 99–153.
- Tseng, Y. H., C. J. Lin, and Y. I. Lin (2007). Text mining techniques for patent analysis. *Information Processing & Management* 43(5), 1216–1247.
- Verspagen, B. (2012). Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. *Advances in Complex Systems* 10(1), 93–115.
- von Wartburg, I., T. Teichert, and K. Rost (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy* 34(10), 1591–1607.
- Wang, P., P. Pattison, and G. Robins (2013). Exponential random graph model specifications for bipartite networks - A dependence hierarchy. *Social Networks* 35(2), 211–222.
- Wang, P., G. Robins, P. Pattison, and E. Lazega (2013). Exponential random graph models for multilevel networks. *Social Networks* 35(1), 96–115.
- Wassennan, L. (2004). *All of Statistics*. New York, NY: Springer.
- Wasserman, S. and K. Faust (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika* 61(3), 401–425.
- Whittington, K. B. (2018). “A tie is a tie? Gender and network positioning in life science inventor collaboration”. *Research Policy* 47(2), 511–526.
- Yoon, B. and Y. Park (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management* 54(3), 588–599.

A SUFFICIENT STATISTICS

In the following, we detail the mathematical definitions of all sufficient statistics incorporated in our model.

Propensity to invent: As already stated in Section 3.1, the standard term to incorporate in any ERGM specification is an edge statistic that counts how many edges are realised in the network. In accordance with Section 3.2, we split this term into the statistics $s_{\text{New}}(\mathbf{Y}_t^+) = |\mathbf{Y}_t^+| = \sum_{i \in \mathcal{I}_t^+} \sum_{k \in \mathcal{K}_t} y_{t,ik}$ and $s_{\text{Experienced}}(\mathbf{Y}_t^-, \dots, \mathbf{Y}_{t-s}^-) = |\mathbf{Y}_t^-| = \sum_{i \in \mathcal{I}_t^-} \sum_{k \in \mathcal{K}_t} y_{t,ik}$. Figures 8(a) and 8(b) visualise the corresponding two network configurations.

Two-star statistics: Two-star statistics can be stated regarding either set of actors in the case of bipartite networks. The definition of the two-star statistic for the patents is shown in Figure 8(c) and given by

$$s_{\text{twostar.patent}}(\mathbf{Y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \right),$$

while the version for the inventors is visualised in Figure 8(d) and defined as:

$$s_{\text{twostar.inventor}}(\mathbf{Y}_t) = \frac{1}{2} \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \left(\sum_{l \neq k} y_{t,il} \right).$$

Pairwise statistics of inventors: We include three versions of pairwise statistics of inventors introduced in Section 3.2. The statistics are given by

$$s_{\text{assort.x}}(\mathbf{Y}_t, \dots, \mathbf{Y}_{t-s}) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(100 \times \frac{\sum_{j \neq i} y_{t,jk} x_{t,ij}}{\sum_{j \neq i} y_{t,jk}} \right).$$

Note that, in general, the matrix \mathbf{x} can be an arbitrary function of the past networks and nodal or dyadic exogenous information. Its definition differs between the three statistics of pairwise statistics of inventors:

1. Team persistence: For $i, j \in \mathcal{I}_t$ and $i \neq j$ the entries of \mathbf{x}_t^P are given by

$$x_{t,ij}^P = \begin{cases} 1, & \text{if } \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik} y_{u,jk} > 0 \\ 0, & \text{else} \end{cases}$$

and a graphical illustration of the statistic is provided in Figure 8(g). One can comprehend this statistic as a particular type of the four-cycle statistic (Wang et al., 2013) where one half already occurred in the past, and the other half might occur in the present.

2. Collaboration interlocking: For $i, j \in \mathcal{I}_t$ and $i \neq j$ the entries of \mathbf{x}_t^{CI} are defined by

$$x_{t,ij}^{CI} = \begin{cases} 1, & \text{if } \sum_{u=t-s}^{t-1} \sum_{h \in \mathcal{I}_t} \sum_{k, l \in \mathcal{K}_u} y_{u,ik} y_{u,hk} y_{u,jl} y_{u,hl} > 0 \\ 0, & \text{else} \end{cases}$$

and a graphical illustration of the statistic is provided in Figure 8(h). Coming back to the representation as cycle-statistics, this term is a six-cycle statistic in which four of the six edges happened in the time frame from $t-5$ to $t-1$ and two in year t .

3. Spatial proximity: For $i, j \in \mathcal{I}_t$ and $i \neq j$ the entries of \mathbf{x}_t^{SP} are defined as

$$x_{t,ij}^{SP} = \begin{cases} 1, & \text{if } \text{dist}(x_{\text{coord},i}, x_{\text{coord},j}) > 50\text{km} \\ 0, & \text{else} \end{cases}$$

where $x_{\text{coord},i}$ and $x_{\text{coord},j}$ define the longitude and latitude of inventors i and j , respectively, and the function $\text{dist}(x_{\text{coord},i}, x_{\text{coord},j})$ computes the distance in kilometers between them via the haversine formula.

Seniority: The respective binary indicator is based on the `pastpatent` statistic given in (8), but in this case we define it on the inventor level:

$$s_{\text{seniority},i}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s}) = \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik}$$

We binarise this inventor-specific covariate by first computing the median of $s_{\text{seniority},i}(\mathbf{y}_t, \dots, \mathbf{y}_{t-s})$ over all inventors and then using this value to split the inventors into two groups (i.e. seniors and juniors). The resulting categorical covariate relates to the number of patents in the past, and is represented in Figure 8(e).

Gender and gender homophily: The main effect of gender is depicted in Figure 8(f) and defined by:

$$s_{\text{gender}}(\mathbf{y}_t) = \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \mathbb{1}(x_{\text{gender},i} = \text{"male"}),$$

where $x_{\text{gender},i} \in \{\text{"male"}, \text{"female"}\}$ indicates the gender of inventor i . The homophily effect, on the other hand, is for males defined by:

$$s_{\text{homophily,male}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \mathbb{1}(x_{\text{gender},i} = \text{"male"}) \mathbb{1}(x_{\text{gender},j} = \text{"male"}) \right). \quad (16)$$

and for females the formula reads :

$$s_{\text{homophily,female}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \mathbb{1}(x_{\text{gender},i} = \text{"female"}) \mathbb{1}(x_{\text{gender},j} = \text{"female"}) \right). \quad (17)$$

Figure 8(i) visualises the homophily statistic for females. The equivalent statistic for males can be defined in the same manner.

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

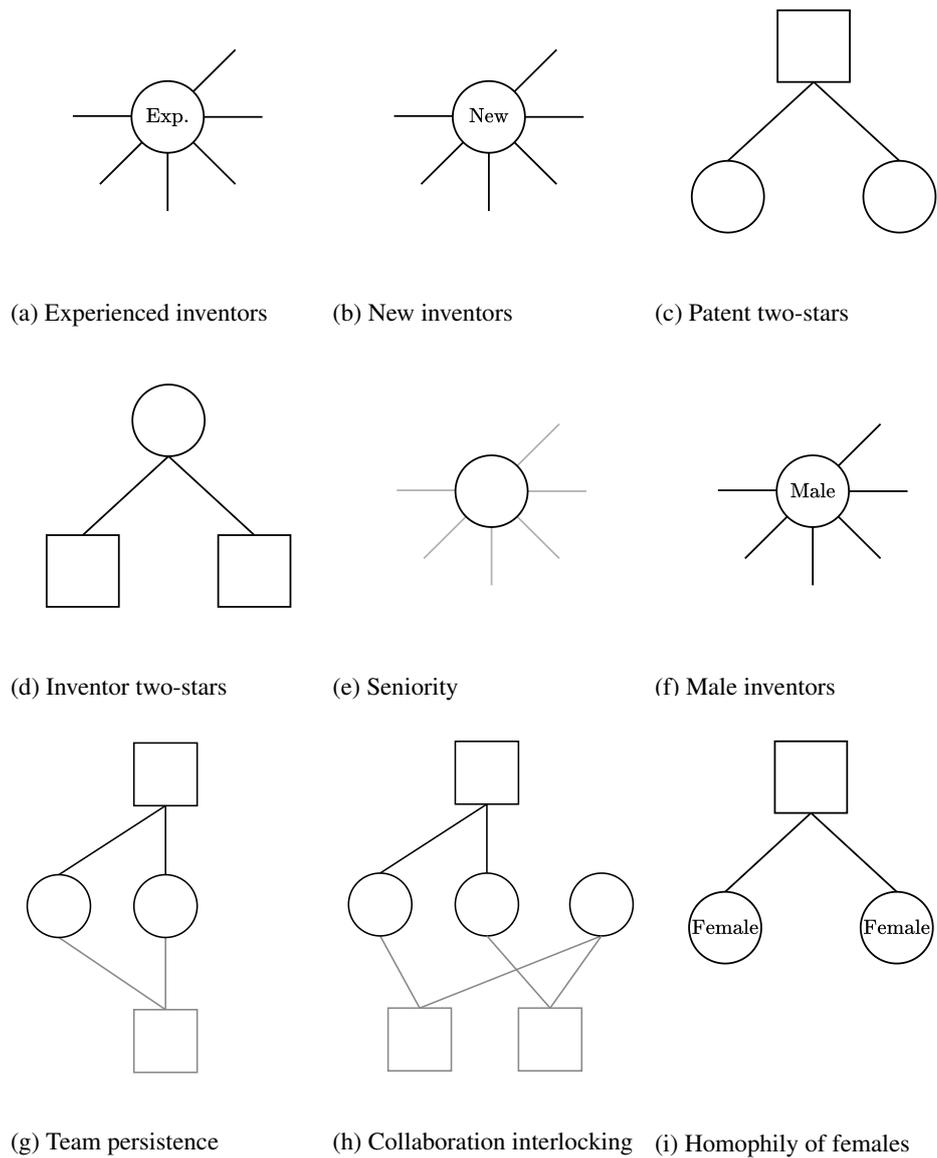


FIGURE 8 Network configurations for the general edge and two-star terms. Circles are inventors and squares patents and black lines are observed edges in the network at time point t , while grey lines are edges in the past.

Modelling the large and dynamically growing bipartite network of German patents and inventors

Cornelius Fritz¹ | Giacomo De Nicola¹ | Sevag Kevork¹
| Dietmar Harhoff² | Göran Kauermann¹

1 | DESCRIPTIVE STATISTICS OF THE PATENT NETWORKS

The top left panel of Figure 1 depicts the absolute number of active inventors and patents in each year. In general, the number of patents per year increase over time, although we observe a drop in recent years. Moreover, there are fewer active inventors than patents in the networks, and the number of inventors is relatively constant over time. On the right side of Figure 1, we plot the density over time. The density declines over time, although we see a boost in the last three years.

2 | IMPLEMENTATION

To carry out the analysis of the presented manuscript, we heavily rely on the `statnet` suite of packages. In particular, we implement wrapper functions to use the `ergm` package for the estimation (Hunter et al., 2008; Krivitsky et al., 2021) and the package template `ergm.userterms` (Hunter et al., 2013) to employ the novel pairwise statistics of inventors. To enable the use of those statistics in other applications, we make the software package `patent.ergm` available for the software package R (R Core Team, 2021). Moreover, we can provide the complete replication code, including the goodness-of-fit analysis and MCMC diagnostics.

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

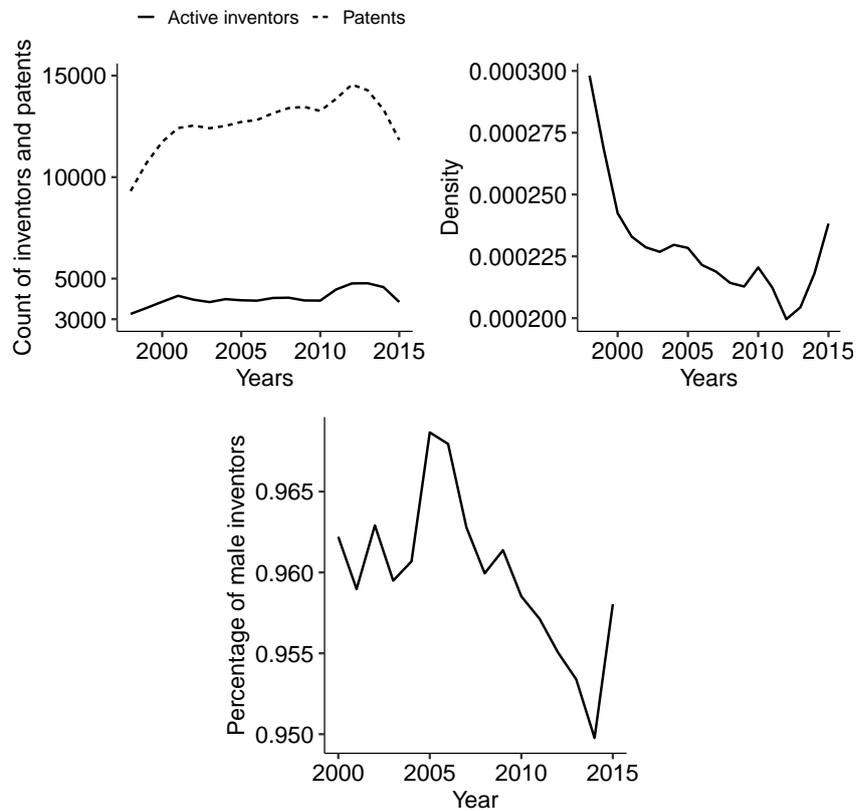


FIGURE 1 (a) Number of inventors and patents over time. (b) Density of the patent network. (c) Percentage of male inventors over time.

3 | GOODNESS-OF-FIT ANALYSIS

To assess whether the estimated model adequately represents the generative mechanisms of the dynamic network of the patent data at hand, we rely on the goodness-of-fit methods proposed by Hunter et al. (2008) and simulate 200 networks for each time point $i \in \mathcal{T}$. We then compute the degree distributions for inventors and patents and compare them to the respective observed statistics. For the analysis carried out in this paper, we have 16 time points for which we separately estimated the model. Resulting from this, we also have to carry out the goodness-of-fit assessment for each time point separately. To save space in the supplementary materials, we only show the results for the first and last year in the analysis, namely 2000 and 2015, in Figure 4.

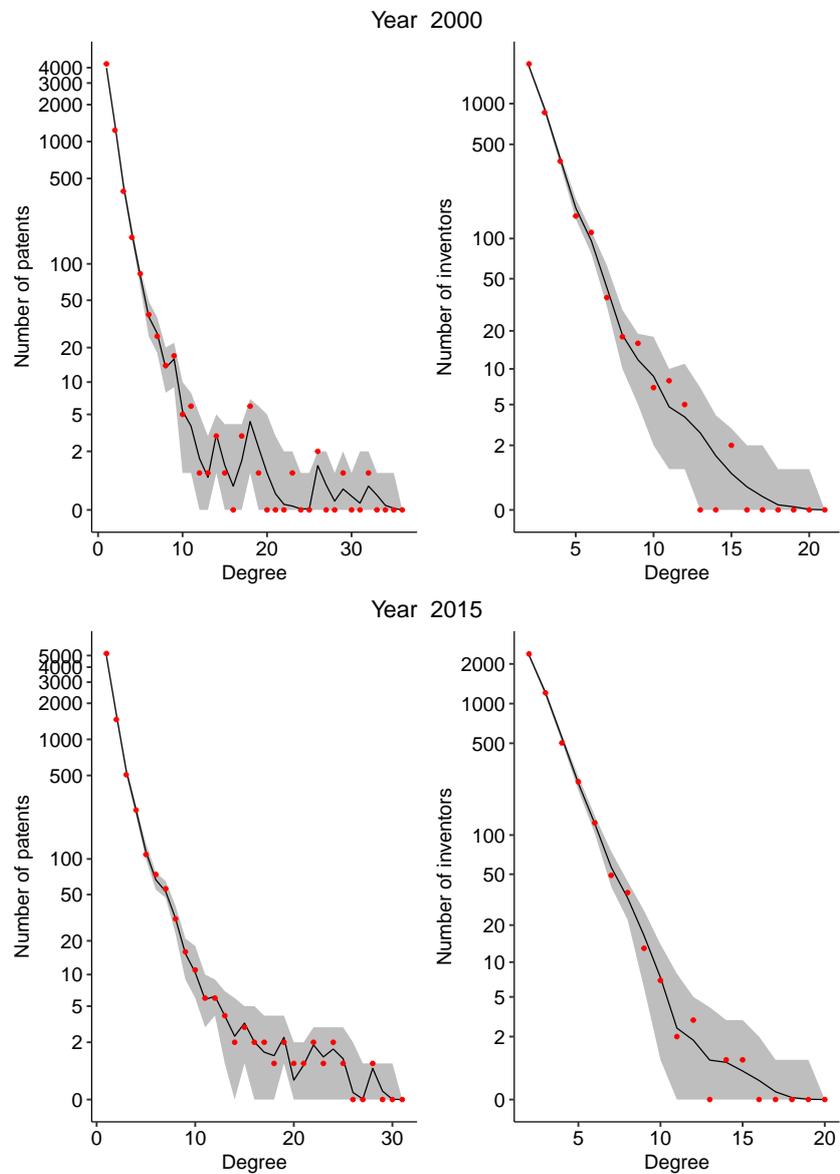


FIGURE 2 Goodness-of-fit assessment via the degree distributions of inventors and patents on a logarithmic scale in year 2000 and 2015. The red dots show the observed networks, the shades are gives the range for the simulated networks

4 | MCMC DIAGNOSTICS

We also provide the standard MCMC diagnostics to guarantee converged estimates of θ . Similarly to the goodness-of-fit assessment, we limit the respective trace and density plots to the first and last year of the analysis.

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

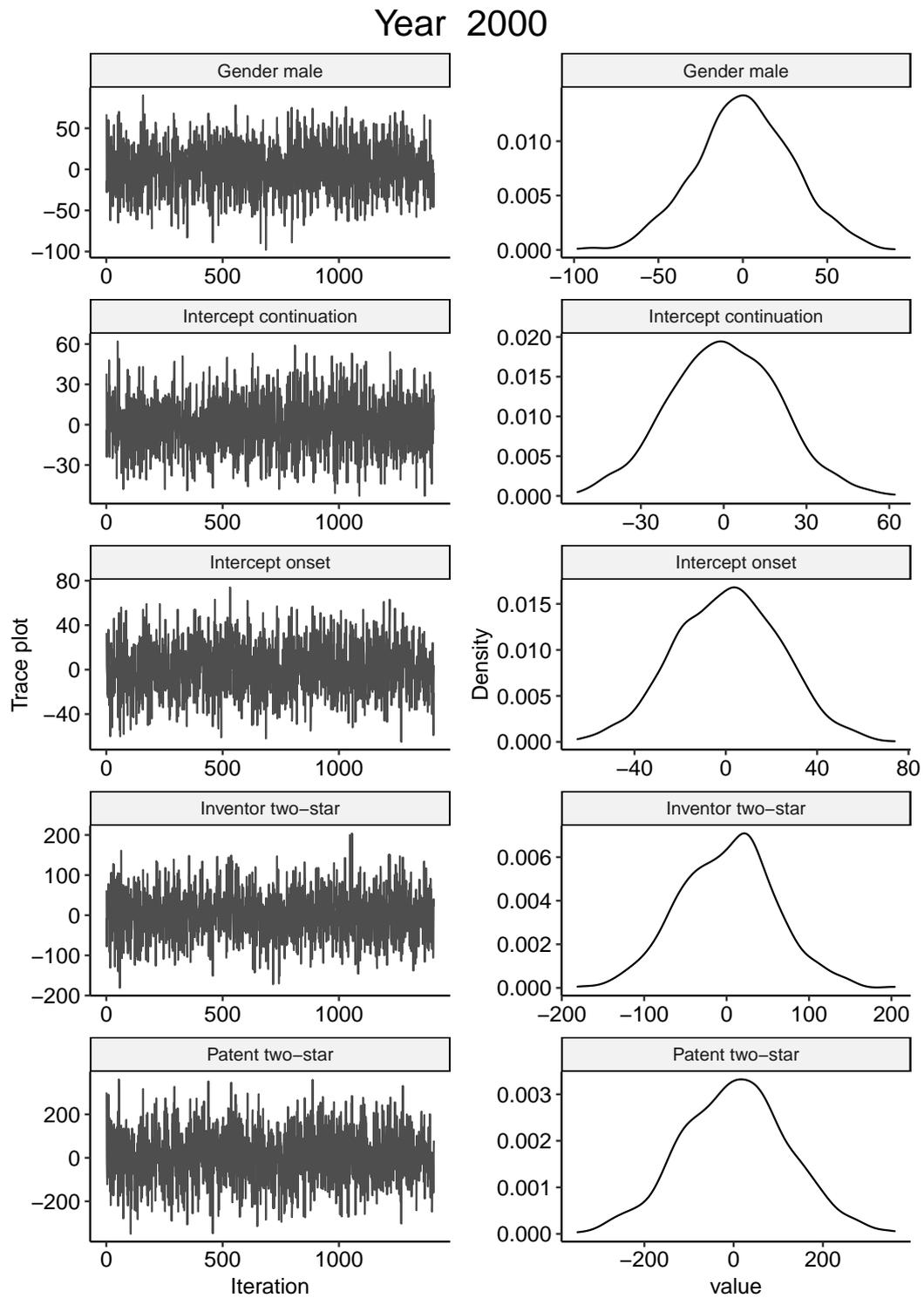


FIGURE 3 MCMC diagnostics of year 2000

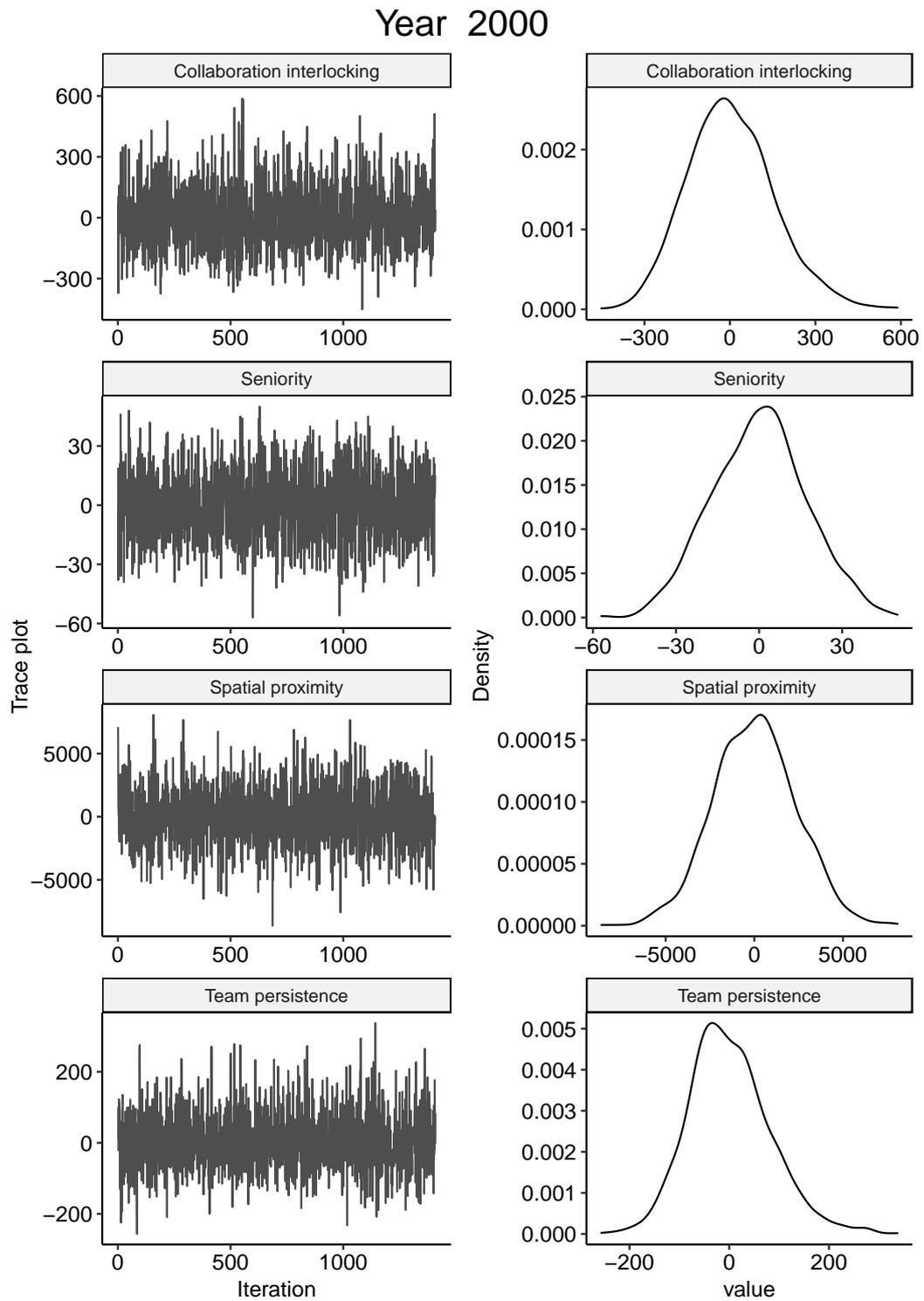


FIGURE 3 MCMC diagnostics of year 2000

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

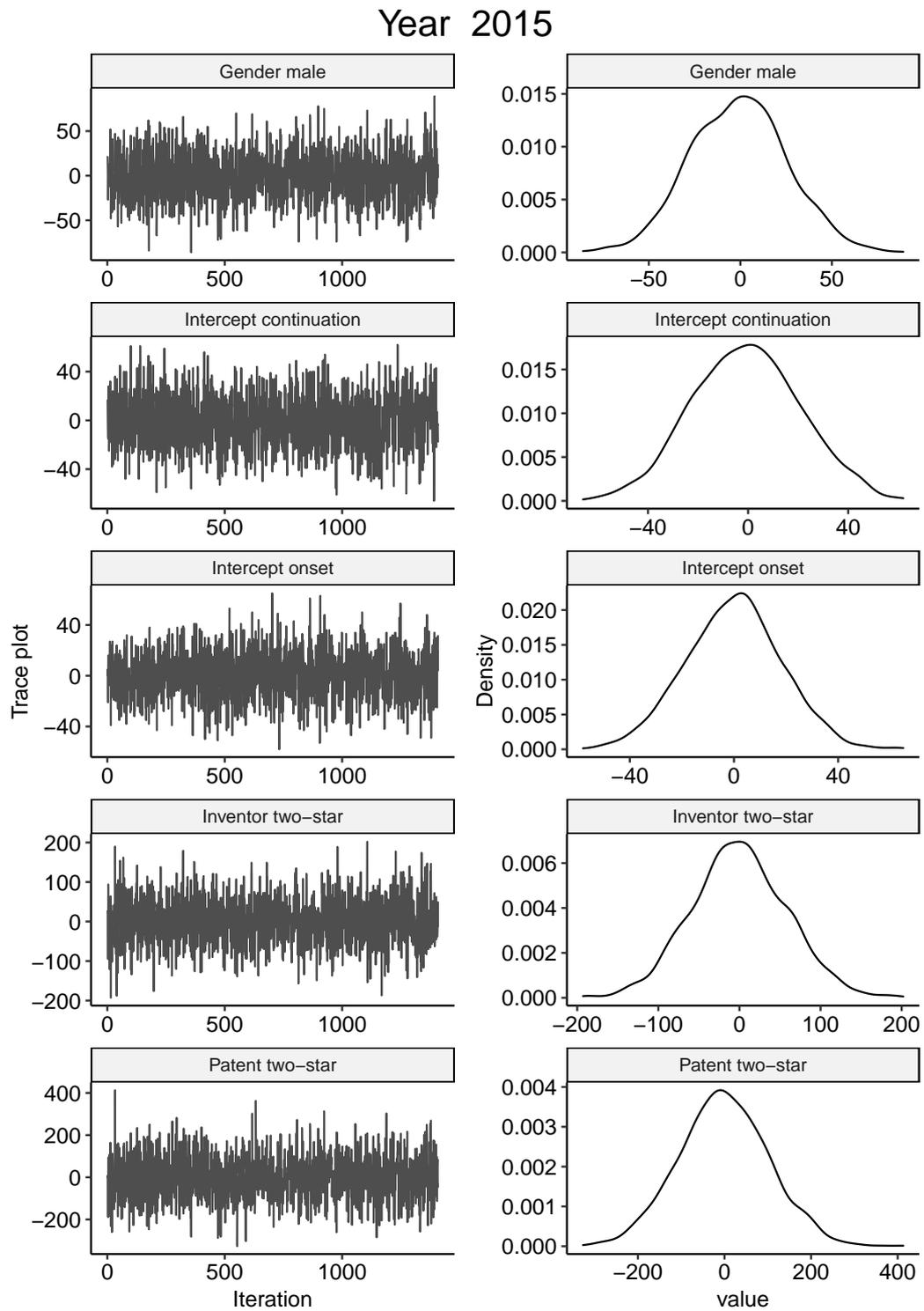


FIGURE 4 MCMC diagnostics of year 2015

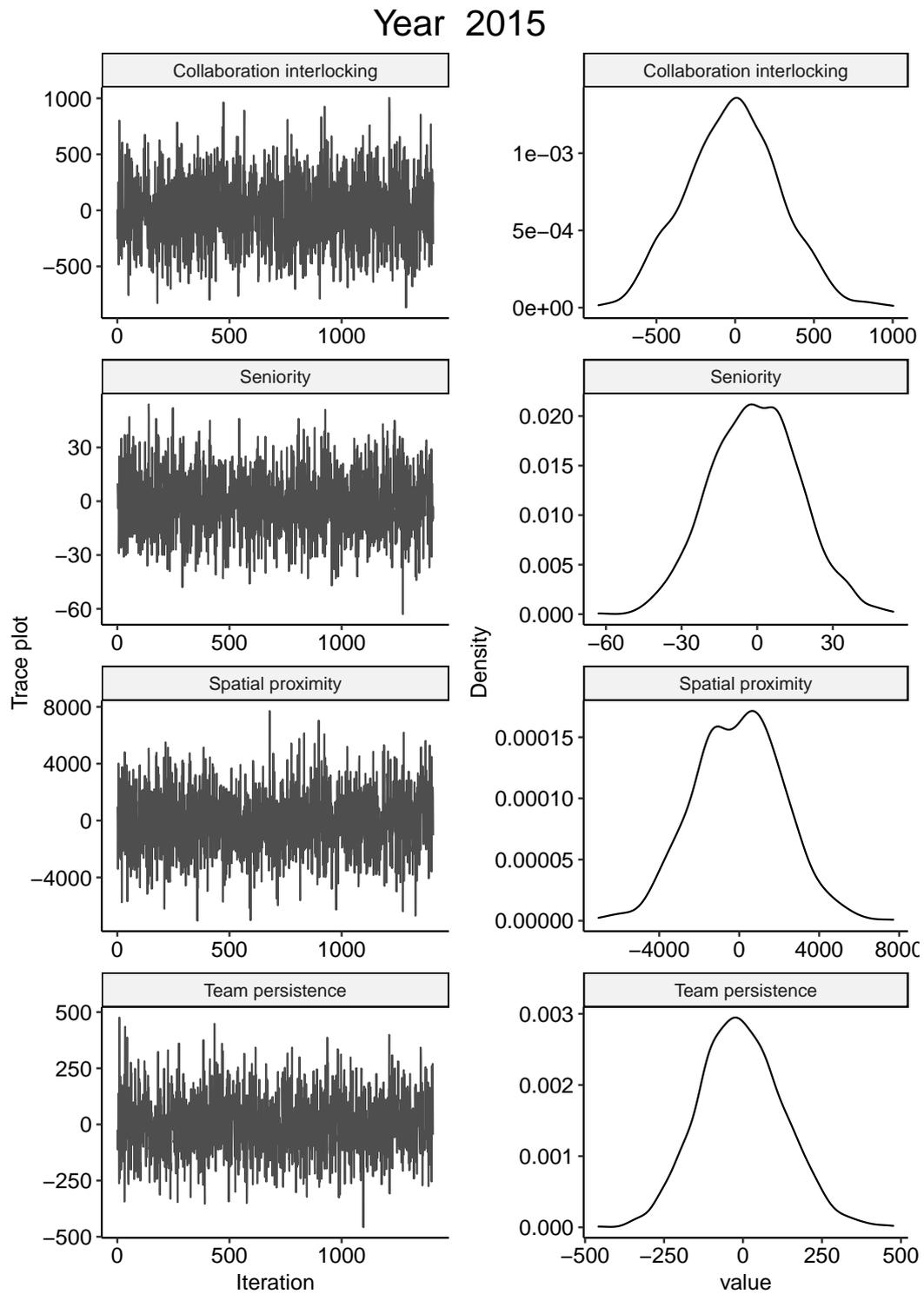


FIGURE 4 MCMC diagnostics of year 2015

9. Modelling the large and dynamically growing bipartite network of German patents and inventors

REFERENCES

- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481), 248–258.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2013). Ergm.userterms: A template package for extending statnet. *Journal of Statistical Software* 52(2), 1–25.
- Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software* 24(3), 1–29.
- Krivitsky, P. N., D. R. Hunter, M. Morris, and C. Klumb (2021). ergm 4.0: New features and improvements.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Part IV.

Relational Event Models

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

Contributing article

Fritz, C., Thurner, W. P. and Kauermann, G. (2021) Separable and Semiparametric Network-based Counting Processes applied to the International Combat Aircraft Trades. *Network Science*, 9 (3), 291–311. doi:10.1017/nws.2021.9.

Replication code

The complete code and data used for this article is openly available as supporting information under <https://www.cambridge.org/core/journals/network-science/article/separable-and-semiparametric-networkbased-counting-processes-applied-to-the-international-combat-aircraft-trades/OD57EC7B7E1775B0BEF72BDE101E507F>.

Copyright information

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited.

Author contributions

The concept of the publication was jointly developed by Cornelius Fritz and Göran Kauermann. Cornelius Fritz implemented the method and wrote the initial draft of the publication. Paul W. Thurner contributed to Section 3.1 and the putting the interpretations of the results in Section 3.3 in the substantive context. Göran Kauermann and Paul W. Thurner were closely involved in proofreading all parts and heavily revised the overall linguistic style.

ORIGINAL ARTICLE

Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

Cornelius Fritz^{1,*} , Paul W. Thurner² and Göran Kauermann¹

¹Department of Statistics, LMU Munich, Munich, Germany (e-mail: goeran.kauermann@stat.uni-muenchen.de) and

²Geschwister Scholl Institute of Political Science, LMU Munich, Munich, Germany (e-mail: paul.thurner@gsi.uni-muenchen.de)

*Corresponding author. Email: cornelius.fritz@stat.uni-muenchen.de

Action Editor: Stanley Wasserman

Abstract

We propose a novel tie-oriented model for longitudinal event network data. The generating mechanism is assumed to be a multivariate Poisson process that governs the *onset* and *repetition* of yearly observed events with two separate intensity functions. We apply the model to a network obtained from the yearly dyadic number of international deliveries of combat aircraft trades between 1950 and 2017. Based on the trade gravity approach, we identify economic and political factors impeding or promoting the number of transfers. Extensive dynamics as well as country heterogeneities require the specification of semiparametric time-varying effects as well as random effects. Our findings reveal strong heterogeneous as well as time-varying effects of endogenous and exogenous covariates on the *onset* and *repetition* of aircraft trade events.

Keywords: arms trade network; combat aircraft; longitudinal network analysis; relational event model

1. Introduction

Network data capture information on relations between actors. The various types of links between actors in the network encompass stable ties associated with some duration. For example, in political science, military alliance agreements are active for a certain number of years (Cranmer et al., 2012; Leeds, 2019). A different type of link consists of instantaneous bilateral events—like hostile actions measured in real-time (Boschee et al., 2018). Note that instantaneous events can be viewed as the limit case of stable ties if the duration of these ties goes to zero (Butts & Marcum, 2017). While instantaneous events can happen anytime, they are not always observable in a high resolution of time. Under these circumstances, we can count the instantaneous events occurring in a given time interval, which implies a network-based counting process. We define the respective class of processes as a multivariate counting process that simultaneously guides all dyadic interactions within an event network and dedicate this article to its analysis. Comprehensive monographs and survey articles on statistical network analysis are available in Kolaczyk (2009), Kolaczyk (2017), Goldenberg et al. (2010), Lusher et al. (2012). Recent overviews of dynamic network modeling can be found in Fritz et al. (2020), Kim et al. (2018).

In real-life applications, most networks exhibit dynamics, that is, structural changes over time are driven by endogenous and exogenous determinants, being covariates that capture the present or past network dependencies and additional information external to the evolution of the network,

respectively. One way to conceive the generating process of networks is to represent it as a discrete Markov chain, where the realized path consists of the observed networks, and the state space is the set of all observable networks. The transition probabilities defining the chain are given by a distribution over all possible networks (Robins & Pattison, 2001). For stable ties, this view results in the temporal exponential random graph model (TERGM, Hanneke et al., 2010).

Alternatively, we can perceive the networks as evolving over time, guided by a continuous Markov process (Holland & Leinhardt, 1977). In this case, network dynamics are often modeled by the stochastic actor-oriented model (SAOM, Snijders, 1996) or in the case of instantaneous events with a precise time stamp by the relational event model (REM) as proposed by Butts (2008). Although modern sensory technology eases the collection of such fine-grained data (Lazer et al., 2009), exact continuous information is usually not obtainable for every observed event. In our case, for example, data on the transactions of combat aircraft trades are collected yearly, but the exact time point of each event (e.g., day of delivery) is impossible to verify (SIPRI, 2019). Therefore, instead of observing instantaneous events, we only protocol the counts of events during given intervals. Consequently, the resulting event data can be comprehended as valued networks, weighted by the count of events that happened within the given intervals. Though the body of literature on dynamic network models is steadily growing, the consideration of valued dynamic networks is less developed and mainly limited to cross-sectional analyses (see Desmarais & Cranmer, 2012; Krivitsky, 2012; Robins et al., 1999; Krivitsky et al., 2009).

In this article, we introduce a tie-oriented model for the analysis of network-based event data. Tie-oriented models assume a bilateral intensity governing the occurrence of events within a dyad, as opposed to actor-oriented models suggested by Stadtfeld (2012) and extended in Hoffman et al. (2020), Stadtfeld et al. (2017). This approach partitions the intensity into an egocentric sender-specific intensity and a probability selecting the receiver conditional on the sender along the lines of the discrete choice model of McFadden (1973). To represent the dynamic evolution of the network-based process, we start with a framework that operates in continuous time at the tie level. Because the ranking of events in our application is not unique due to the lack of exact time stamps, standard REMs (Butts, 2008; Vu et al., 2011) cannot be readily applied. Therefore, we develop our model under the assumption that the exact ordering of aircraft deliveries within the window of a year is unknown and uninformative. Given that a perennial interplay between policymakers of the involved countries as well as a lengthy order process precludes each trade, this assumption seems reasonable (Snijders, 2017).

Our approach extends existing models in multiple ways. Firstly, we generalize the separable decomposition of network dynamics differentiating between the formation and dissolution of ties introduced by Krivitsky & Handcock (2014), Holland & Leinhardt (1977). In particular, we extend the separable decomposition to event and count data instead of the continuous specification given in Krivitsky & Handcock (2014) and Holland & Leinhardt (1977), where solely binary and durable ties are regarded. Thereby, we enhance recently introduced *windowed* effects by Stadtfeld et al. (2017). Furthermore, we propose a semiparametric specification and use penalized B-splines to obtain flexible time-varying coefficients (Eilers & Marx, 1996). In a similar approach, Bauer et al. (2021) employ non-linear effects to investigate the collaboration between inventors through joint EU patents. Kreiß et al. (2019) propose a nonparametric model with time-varying coefficients that necessitates time-continuous observations, although focusing on the estimator's properties as the number of actors goes to infinity. To capture latent actor-specific heterogeneity, we include random effects for each actor in the network differentiating between the sender and receiver of events. As an application case, we focus on the strategically most crucial international deliveries of weapons, namely combat aircraft from 1950 to 2017 (Forsberg, 1994; SIPRI, 2020a). Combat aircraft comprises all "unmanned aircraft with a minimum loaded weight of 20 KG" (SIPRI, 2020b). They are very costly, and the number of units transferred constitutes highly valuable information for military strategists (Forsberg, 1997). Therefore, we propose to focus on yearly unit sales as a substantial quantity.

The remainder of this article is structured as follows: the next section formally introduces the tie-oriented model based on a network-based counting process together with extensions to separable, time-varying, and random effects and an estimation procedure. Consecutively, we introduce the application case and apply our novel method. The paper concludes with Section 4.

2. Network-based counting process

2.1 A framework for discrete and continuous time event data

We start by proposing the model for time-continuous event data, which are observed at discrete time points. We use the temporal indicator $\tilde{t} \in \mathcal{T} = [0, T)$ and mathematically define the network-valued process as a Poisson process on a valued network given by:

$$N(\tilde{t}) = (N_{ij}(\tilde{t}) \mid i, j \in \{1, \dots, n\}) \quad (1)$$

where $n \in \mathbb{N}$ is the total number of actors in the network. Process (1) counts the relational events between all actors in the network during the interval $[0, \tilde{t})$. It is characterized by the network-valued intensity rate $\tilde{\lambda}(\tilde{t}) = (\tilde{\lambda}_{ij}(\tilde{t}) \mid i, j \in \{1, \dots, n\})$. The (i, j) th entry of this intensity is defined as the probability that we observe an instantaneous jump of size 1 in $N_{ij}(\tilde{t})$. Heuristically, this is the probability of the occurrence of a directed event from actor i to j at time point \tilde{t} . By definition, we set $\tilde{\lambda}_{ii}(\tilde{t}) = 0 \forall i \in \{1, \dots, n\}$ and $\tilde{t} \in \mathcal{T}$.

Assuming the process is observed at discrete time points $t \in \{0, \dots, T\}$ leads to the time-discrete observations \mathbf{Y}_t , which are defined as cumulated events through:

$$\mathbf{Y}_t = \mathbf{N}(t) - \mathbf{N}(t-1) \forall t \in \{1, \dots, T\}$$

with $\mathbf{N}(0)$ set to 0. Based on the properties of a Poisson process, these increments follow a matrix-valued Poisson distribution:

$$\mathbf{Y}_t \sim \text{Pois} \left(\int_{t-1}^t \tilde{\lambda}(\tilde{u}) d\tilde{u} \right) \forall t \in \{1, \dots, T\} \quad (2)$$

Given that the exact orderings of events within each observation window are not known and assumed to be uninformative, the integrated intensity on the time interval $(t-1, t]$ simplifies to a constant, so that $\int_{t-1}^t \tilde{\lambda}(\tilde{u}) d\tilde{u} = \lambda(t)$ holds. Accordingly, we define the observed values of \mathbf{Y}_t as \mathbf{y}_t . As a result of Equation (2), the waiting times between subsequent events follow an exponential distribution. Therefore, our model is equivalent to the REM as introduced in Butts (2008) in the special case where $\|\mathbf{y}_t\|_1 = 1 \forall t \in \{1, \dots, T\}$ holds under piece-wise constant intensities.

Generally, we are interested in modeling $\lambda(t)$ conditional on the past network topology and exogenous covariates, which are denoted by \mathbf{x}_t . Covariates can be node-specific (regarding either a feature of the sender or receiver), dyadic (regarding a relation between the sender and receiver), or global (regarding the complete network). Building on a first-order Markov property, we allow the intensity to depend on the past network behavior and exogenous covariates through:

$$Y_{ij,t} \sim \text{Pois}(\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1})) \forall t \in \{1, \dots, T\}; i, j \in \{1, \dots, n\}, i \neq j \quad (3)$$

This is equivalent to the assumption of dyadic independence of events to occur in each time interval given information on the past and exogenous covariates. Similar assumptions were made by Lebacher et al. (2021) in the context of separable TERGMs (Krivitsky & Handcock, 2014). Almquist & Butts (2014) justify this method for network panel data where little simultaneous dependence between possible ties is present. For our application to the international combat aircraft trades, this can be legitimized by the long time span of aircraft trades between the order and delivery of units.¹

Accordingly, we specify the intensity in time-varying semiparametric form through:

$$\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}) = \lambda_0(t) \exp\{\boldsymbol{\theta}(t)^\top s_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})\} \quad \forall t \in \{1, \dots, T\} \quad (4)$$

where $\lambda_0(t)$ is the baseline intensity, $s_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ is a multidimensional vector consisting of network statistics and theoretically derived exogenous covariates in $t - 1$. We discuss different specifications of the statistics in Section 3 where we describe the application case in more detail. The coefficient vector $\boldsymbol{\theta}(t)$ is possibly time-varying and needs to be estimated from the data.

In many application cases, compositional changes of the actor set occur. To compensate for this phenomenon in the model, we include indicator functions similar to risk indicators in time-to-event analysis (Kalbfleisch & Prentice, 2002). To be specific, we multiply the intensity by an indicator function, determining whether actors i and j are both present in the network at time t :

$$\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}) = \mathbb{I}(i, j \in \mathcal{R}_t) \lambda_0(t) \exp\{\boldsymbol{\theta}(t)^\top s_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})\} \quad \forall t \in \{1, \dots, T\} \quad (5)$$

with \mathcal{R}_t denoting the set of actors partaking in the network at time point t . By including the indicator functions $\mathbb{I}(i, j \in \mathcal{R}_t)$, we decompose our observed network into a stochastic and deterministic component. The latter component consists of structural zeros at time point t in the modeled network between all actors where at least one side is not present. With these actor set changes, the possible range of the network statistics changes as well, leading to values which are not scaled coherently for a comparison across years. To solve this issue, we divide all network statistics by their maximal value to allow for a cohesive interpretation.

2.2 Extensions

2.2.1 Separability assumption

Interaction patterns are commonly substantially different for already linked and still unlinked actors. To adequately capture this characteristic, Holland & Leinhardt (1977) proposed a process-based model for binary ties taking the values “0” or “1” by two separate intensity functions. One intensity toggles entries from “0” to “1” (formation of ties) and another one from “1” to “0” (dissolution of ties). Thereby, separate and potentially differential effects of statistics depending on previous interaction behavior are enabled. This model, henceforth called *separable* model, was later adopted to the SAOM by incorporating a *so-called* gratification function (Snijders & van Duijn, 1997; Snijders, 2003) and to the TERGM by extending it to the separable TERGM (Krivitsky & Handcock, 2014). However, one should keep these *separable* models apart from the *separability condition* introduced in Almquist & Butts (2014). In the following, we combine the framework of REMs with the separability approach as coined by Krivitsky & Handcock (2014).

More specifically, we postulate two different conditions for the network-based process under which the effect of all covariates changes. One condition governs events between unlinked actors and is characterized by the *onset* intensity. The second condition only regards events among actors that already interacted with each other and is driven by the *repetition* intensity. In accordance with the Markov assumption specified in Equation (4), we define the *onset* intensity at time t to control all events which did not occur in \mathbf{y}_{t-1} . Accordingly, the *repetition* intensity drives the events that did occur at least once in \mathbf{y}_{t-1} . This can be incorporated by splitting the intensity into two conditional intensities:

$$\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}) = \begin{cases} \lambda_{ij}^+(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}), & \text{if } y_{ij,t-1} = 0 \\ \lambda_{ij}^-(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}), & \text{if } y_{ij,t-1} > 0 \end{cases} \quad (6)$$

where $\lambda_{ij}^+(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ and $\lambda_{ij}^-(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ are defined along the lines of Equation (4) and specified by the corresponding time-varying parametric effects $\boldsymbol{\theta}^+(t)$ and $\boldsymbol{\theta}^-(t)$ jointly represented by $\boldsymbol{\theta}(t) = (\boldsymbol{\theta}^+(t), \boldsymbol{\theta}^-(t))$. The possibly overlapping vectors of statistics are denoted accordingly as $s_{ij}^+(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ and $s_{ij}^-(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$, respectively. Setting $s_{ij,0}^+(\mathbf{y}_{t-1}, \mathbf{x}_{t-1}) = 1$ enables the inclusion

of a time-varying intercept $\lambda_0^+(t) = \exp\{\theta_0^+(t)\}$ in the *onset* model, this holds similarly for the *repetition* model. Consecutively, the complete separable model is given by replacing Equation (4) with

$$\begin{aligned} \lambda_{ij}(t, \mathbf{Y}_{t-1}, \mathbf{x}_{t-1}) &= \exp \left\{ \mathbb{I}(y_{ij,t} = 0) [\boldsymbol{\theta}^+(t)^\top s_{ij}^+(y_{t-1}, \mathbf{x}_{t-1})] \right. \\ &\quad \left. + \mathbb{I}(y_{ij,t} > 0) [\boldsymbol{\theta}^-(t)^\top s_{ij}^-(y_{t-1}, \mathbf{x}_{t-1})] \right\} \\ &= \exp \left\{ \boldsymbol{\theta}(t)^\top s_{ij}(y_{t-1}, \mathbf{x}_{t-1}) \right\} \end{aligned} \quad (7)$$

where $\boldsymbol{\theta}(t) = (\boldsymbol{\theta}^+(t), \boldsymbol{\theta}^-(t))$ and

$$s_{ij}(y_{t-1}, \mathbf{x}_{t-1}) = (\mathbb{I}(y_{ij,t} = 0) \cdot s_{ij}^+(y_{t-1}, \mathbf{x}_{t-1}), \mathbb{I}(y_{ij,t} > 0) \cdot s_{ij}^-(y_{t-1}, \mathbf{x}_{t-1}))$$

2.2.2 Spline-based time-varying effects

Let the k th component of statistic $s_{ij}(y_{t-1}, \mathbf{x}_{t-1})$ be defined as $s_{ij,k}(y_{t-1}, \mathbf{x}_{t-1})$ with the matching coefficient $\theta_k(t)$. We expand each component $\theta_k(t)$ in a semiparametric way by replacing it with a B-spline basis function (see de Boor, 2001). More specifically, we place equidistant knots on a grid in \mathcal{T} , where the number of knots can be chosen relatively high (Kauermann & Opsomer, 2011). In principle, we could choose individual grids for each component of $\boldsymbol{\theta}(t)$, but for the sake of a simple notation, we select the same one for all covariates. We now rewrite each coefficient as:

$$\theta_k(t) = B(t)\alpha_k \quad \forall k \in \{0, \dots, K\} \quad (8)$$

where $B(t) \in \mathbb{R}^q$ is the B-spline basis evaluated at t and $\alpha_k \in \mathbb{R}^q$ denotes the corresponding coefficient vector. In our context, q constitutes the dimension of the B-spline basis and hence gives the number of separate B-spline bases used for each covariate. To obtain a smooth fit, we penalize the difference of adjacent basis coefficients α_k as proposed by Eilers & Marx (1996). This leads to the overall penalized log-likelihood function:

$$\ell_p(\alpha_0, \dots, \alpha_K, \gamma_0, \dots, \gamma_K) \propto \sum_{t=1}^T \sum_{i,j \in \mathcal{R}_t} (y_{ij,t} \log(\lambda_{ij,t}) - \lambda_{ij,t}) - \frac{1}{2} \sum_{k=0}^K \gamma_k \alpha_k^\top D_k \alpha_k \quad (9)$$

with $\lambda_{ij,t} = \lambda_{ij}(t, y_{t-1}, \mathbf{x}_{t-1})$. The penalty results from the quadratic form with penalty matrix D_k constructed from pairwise differences of the spline coefficients and γ_k as the penalty (and hence tuning) parameter. This vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$ controls the smoothness of the fit and is chosen data based following a mixed model approach as described in detail in Ruppert et al. (2003), see also Wood (2017). The incorporation of a penalization in Equation (9) results in a biased estimator and a *so-called* bias-variance tradeoff, which is thoroughly discussed for penalized spline smoothing in Ruppert et al. (2003). Kauermann & Opsomer (2011) extend the theoretical results toward a data-driven finite-sample version, and Kauermann et al. (2009) show that the estimates from Equation (9) are consistent.

2.2.3 Accounting for nodal heterogeneity

The specification of the model introduced so far implicitly implies that the nodal heterogeneity is fully captured by the structural statistics $s_{ij}(y_{t-1}, \mathbf{x}_{t-1})$. As already thoroughly discussed by Thiemichen et al. (2016) or Box-Steffensmeier et al. (2018), this can be considered a questionable assumption. It seems, therefore, advisable to include sender- and receiver-specific random effects to account for unobserved heterogeneity. Let therefore u_i^S denote a latent sender-specific effect of actor i and u_j^R the receiver-specific effect of actor j . This leads to the heterogeneous intensity

$$\lambda_{ij}(t, y_{t-1}, \mathbf{x}_{t-1}, u^S, u^R) = \lambda_{ij}(t, y_{t-1}, \mathbf{x}_{t-1}) \exp\{u_i^S + u_j^R\} \quad \forall t \in \{1, \dots, T\} \quad (10)$$

We assume $u^S = (u_1^S, \dots, u_n^S)^\top \sim N(0, I_n \tau_S^2)$ and $u^R = (u_1^R, \dots, u_n^R)^\top \sim N(0, I_n \tau_R^2)$ with I_n as the $n \times n$ identity matrix. The expression $\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ may be specified through Equation (4) or (7). Conditional on the random effects u^S and u^R , the distributional assumption (3) still holds

$$Y_{ij}(t) \mid u^S, u^R \sim \text{Pois}(\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}, u^S, u^R)) \quad (11)$$

$$\forall t \in \{1, \dots, T\}; i, j \in \{1, \dots, n\}, i \neq j$$

where $\lambda_{ij}(t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}, u^S, u^R)$ is specified in Equation (10).

2.3 Estimation

The vector-valued function $\theta(t) = (\theta^+(t), \theta^-(t))$ is estimated by finding the argument maximizing the penalized likelihood resulting from Equation (11) and viewing the penalty on coefficient vector α as an improper prior distribution. This leads to a generalized additive mixed model, which is extensively discussed in Wood (2017), Ruppert et al. (2003), Ruppert et al. (2009). To leverage the advanced optimization techniques proposed for this model class, we initially calculate all covariates $s_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$ for each actor-tuple and at each point in time. By doing that, we transform the data into a generalized version of the *so-called* counting-process representation, which is known from time-to-event analysis (Tutz & Schmid, 2016; Friedman, 1982; Whitehead, 1980). For each snapshot of the event network at time point t , this procedure generates a design matrix of $|\mathcal{R}_t|$ conditionally independent observations with a target variable $y_{ij,t}$ expressing the number of events that occurred between a specific tuple of actors and covariates given by $s_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$.

For the estimation, we use the versatile R package `mgcv` (Wood, 2017, version 1.8-31). Thereby, we follow Wood et al. (2017) who enhance the pseudo-quasi-likelihood method by Breslow & Clayton (1993) for the analysis of larger data sets. The main extensions are threefold:

- (1) The tuning parameters γ are not estimated until convergence in each iteration of the estimation procedure but updated by only one Newton step.
- (2) Efficient methods for computing the matrix cross-products in each iteration are run in parallel (Li & Wood, 2020).
- (3) The covariates are discretized along a marginal grid. Hence, the design matrices for the smooth covariates take significantly less memory.

Wood et al. (2017) describe the method in detail as it is implemented in the function `bam` of the already mentioned R package. Well-calibrated frequentist confidence bands for the estimated function $\theta(t)$ are guaranteed by Bayesian large sample properties (Wood, 2006).

3. Application

3.1 Data

So far, quantitative work on the international arms trade utilizing statistical network analysis has mostly been restricted to binarized networks. Here, the occurrence of a trade relationship between two countries in a specific year was modeled conditional on endogenous and exogenous statistics by the gravity model of trade by employing TERGMs and extensions of it (Lebacher et al., 2021; Thurner et al., 2019). Contrary, Lebacher et al. (2020) fit a network disturbance model on the yearly aggregated trend indicator values (SIPRI, 2020b) of the international arms trades, maintaining the valued character of deliveries. All these contributions rely on data provided by the Stockholm International Peace Research Institute (SIPRI, 2020a), and they consider each type of major conventional weapons indiscriminately.

In the following, we concentrate on the counts of combat aircraft deliveries, as reported in the SIPRI data, where each combat aircraft delivery is perceived as an event. We focus on the

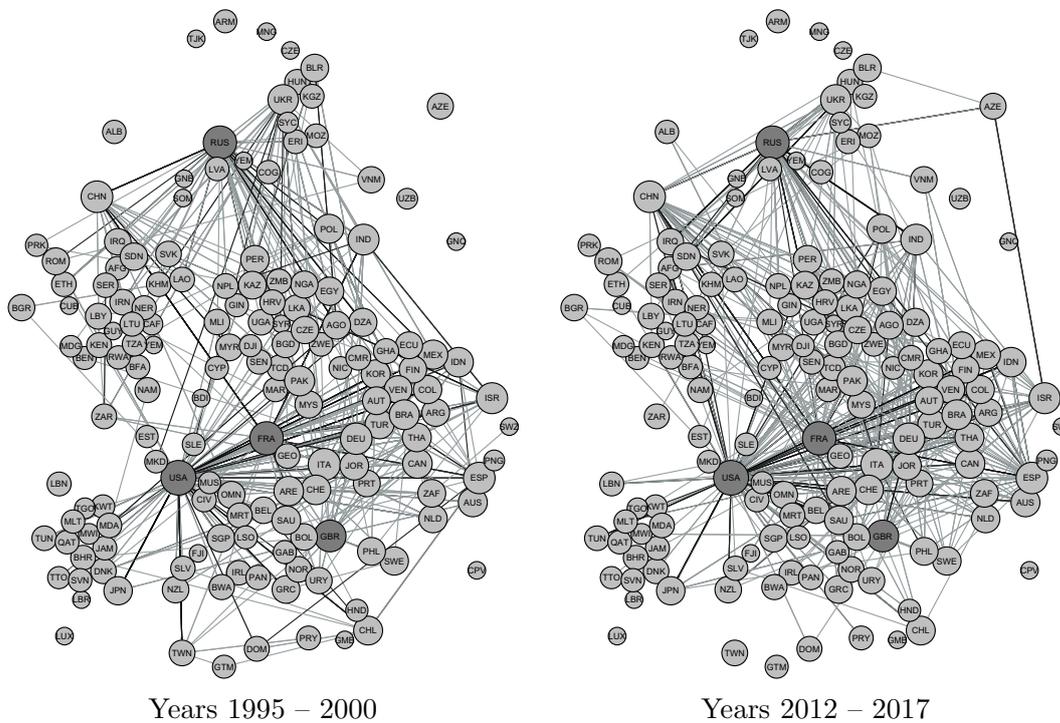


Figure 1. The international network of combat aircraft trades in two periods. Node size is proportional to the sum of involved deals and the grey-scale of each tie indicates the aggregated amount of deals in the specific time frame. The labels of the nodes are the ISO3 codes of the respective countries. The four major sender countries are drawn in a darker shade.

transfers of aircraft because these weapon systems usually incorporate the highest technological sophistication. Therefore, they are being restricted to close allies. Furthermore, they are of crucial strategic importance for international deterrence and counterinsurgency in intrastate conflict (Hoeffler & Mérand, 2016; Mehrl & Thurner, 2020). Lastly, their sizes and cost make the available data highly reliable (Forsberg, 1994, 1997). Previous research on combat aircraft trade was limited to the quantitative analysis of a small subset of countries or fighter programs (Hoeffler & Mérand, 2016; Vucetic, 2011; Vucetic & Nossal, 2012). Contrasting these endeavors, we take a global point of view on the combat aircraft trade. Here, a closer look at the data reveals how countries commonly partition major deals with their stable trade partners into multiple deliveries occurring over the span of several years. For instance, the United States and Japan signed a deal in 1984 comprising 32 quantities of aircraft, which were realized between 1988 and 2016. The additional information provided by this segmentation of trade deals into isolated deliveries would be lost when only regarding binarized networks.²

Two examples of the network representing aggregated events over 6 years are depicted in Figure 1. Generally, the networks exhibit a structure with hubs around the United States (USA), Russia (RUS), France (FRA), and United Kingdom (UK). Coincidentally, this set of countries also demonstrate the highest average hub-scores over time (Kleinberg, 1999). Analog to the distribution of the in- and out-degrees in binary networks, we can examine the distribution of the concatenated in- and outgoing event counts for all years. We call the respective statistics in- and out-count, although they are equivalent to the generalized degree proposed by Opsahl et al. (2010). The empirical distribution of those statistics enables a better understanding of the topology of the observed networks. Figure 2(a) suggests a strong centralization in the outward event count distribution. Some countries are the sender of up to 130 deliveries in one year. Still, on average, 82%

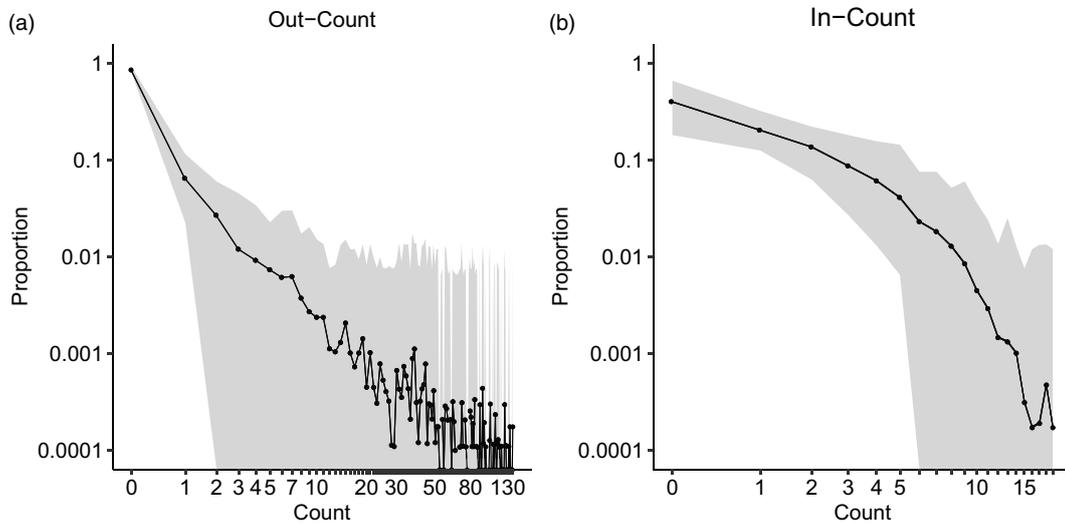


Figure 2. Distributions of the Out- and In-Counts for all included countries concatenated over all years. The shaded area represents the minimum and maximum of the observed values. Both graphs are represented on a logarithmic scale.

of the countries do not export. The inward count distribution is not as skewed and centralized, as shown in Figure 2(b). Nonetheless, the mode of the distribution is still at zero.

3.2 Model specification

We now employ the outlined model to the international combat aircraft trade network spanning from 1950 to 2017. The event networks are observed yearly. In this context, $y_{ij,t}$ denotes the number of observed combat aircraft units delivered in year t between country i and j and its distribution follows from Equation (3). Given this information, we estimate the time-continuous intensities of all country-dyads, which are per assumption governed by the *repetition* intensity if the respective countries traded in the previous year and by the *onset* intensity otherwise as defined in Equation (7).³ All network actors are countries, and an event represents the delivery of combat aircraft between them. To appropriately capture interdependencies of the observed event counts, we incorporate a wide range of endogenous statistics, whose mathematical representation is given in Table 1 and visualized in Figure 3. Generally, we define all non-binary structural statistics to be bounded between 0 and 100 to guarantee a consistent interpretation independent of the varying network size and prevent the implied autoregressive counting process from unrealistic behavior (Gjessing et al., 2010).

As already investigated in multiple applications (Snijders, 2003; Newman et al., 2002), the degree structure plays a crucial role in the observed event network. In the case of directed events, the in- and out-degree of a country determine its relative location in the network (Wasserman & Faust, 1994). In our application, the degrees reflect the number of different countries with whom a specific country had at least one transaction in a particular year as an importer (in-degree) and exporter (out-degree). To reveal the impact of these measures on the intensity of observing an event, we include four degree-related statistics concerning the sender and receiver in our specification, as illustrated in Figure 3(a)–(d). For instance, one can interpret a positive effect of the sender's out-degree as the tendency to trade with countries that are already sending a lot in the previous year.

Besides degree-based statistics, Holland & Leinhardt (1971), Davis (1970) highlight the role of triangular structures in networks. When adapted to event relations, it refers to the change in intensity of an event between countries i and j , if they are indirectly connected by an additional

Table 1. Mathematical formulations of the structural covariates as calculated for $S_{ij}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})$. The number of countries that are present in the network at time point t is denoted by n_t . The identifying letters concern the respective graphical illustrations in Figure 3

Name	Mathematical representation
(a) In-degree sender	$\frac{100}{n_t-1} \sum_{h=1}^n \mathbb{I}(y_{hi,t-1} > 0)$
(b) In-degree receiver	$\frac{100}{n_t-1} \sum_{h=1}^n \mathbb{I}(y_{hj,t-1} > 0)$
(c) Out-degree sender	$\frac{100}{n_t-1} \sum_{h=1}^n \mathbb{I}(y_{ih,t-1} > 0)$
(d) Out-degree receiver	$\frac{100}{n_t-1} \sum_{h=1}^n \mathbb{I}(y_{jh,t-1} > 0)$
(e) Transitivity	$\frac{100}{n_t-2} \sum_{h=1}^n \mathbb{I}(y_{ih,t-1} > 0) \mathbb{I}(y_{hj,t-1} > 0)$
(f) Shared supplier	$\frac{100}{n_t-2} \sum_{h=1}^n \mathbb{I}(y_{hi,t-1} > 0) \mathbb{I}(y_{hj,t-1} > 0)$
(g) Reciprocity	$\mathbb{I}(y_{ji,t-1} > 0)$

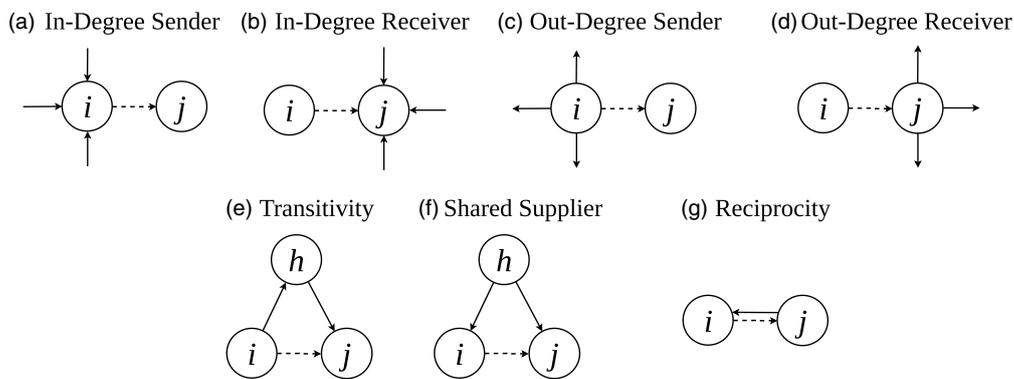


Figure 3. Graphs consisting of three arbitrary actors i, j , and h that illustrate the included triangular and dyadic covariates in the first row. Dashed arrows represent the event that is modeled and solid arrows in $t - 1$.

two-path, i.e., third country. Since the aircraft deliveries between countries are directed, there are multiple ways to define two-paths. We incorporate two triadic structures: transitivity, Figure 3(e), and shared supplier, Figure 3(f). While transitivity in an event network suggests that already having observed a delivery from country i to k and k to j affects the intensity of an event from i to j , the shared supplier mechanism reflects the tendency toward trading with countries that import combat aircraft from a common exporter. These triangular structures were the only variants found to be relevant for the trade of combat aircraft. Likewise, we control for reciprocity, which is the tendency of countries to respond to previous events directed at them, Figure 3(g).

Political economy models of arms trade (Levine et al., 1994; Thurner et al., 2019) as well as the gravity model of arms trade guide the selection of appropriate exogenous covariates. Thurner et al. (2019), Akerman & Seim (2014) included the dyadic distance in kilometers between the capitals of country i and j as well as the logarithmic gross domestic product (GDP in US \$) of the sender and receiver countries as covariates in the model. Pamp et al. (2018), Lebacher et al. (2021) emphasize the impact of military expenditures as a proxy for the Newtonian power of attraction, which we include in logarithmic form as a sender- and receiver-specific covariate. The respective yearly data were collected by SIPRI (2019) in US \$ and combined by Nordhaus et al. (2012) with data from Singer et al. (1972). We use this combined data set but employ linear interpolation if at least 60% of the time series for a specific country is observed. Moreover, we incorporate two dyadic variables controlling whether country i and j signed an alliance treaty or are similar to each other in terms of their regimes in power, following Martínez-Zarzoso & Johannsen (2019), Thurner et al. (2019). The alliance treaty obligations and provisions project identified military alliance

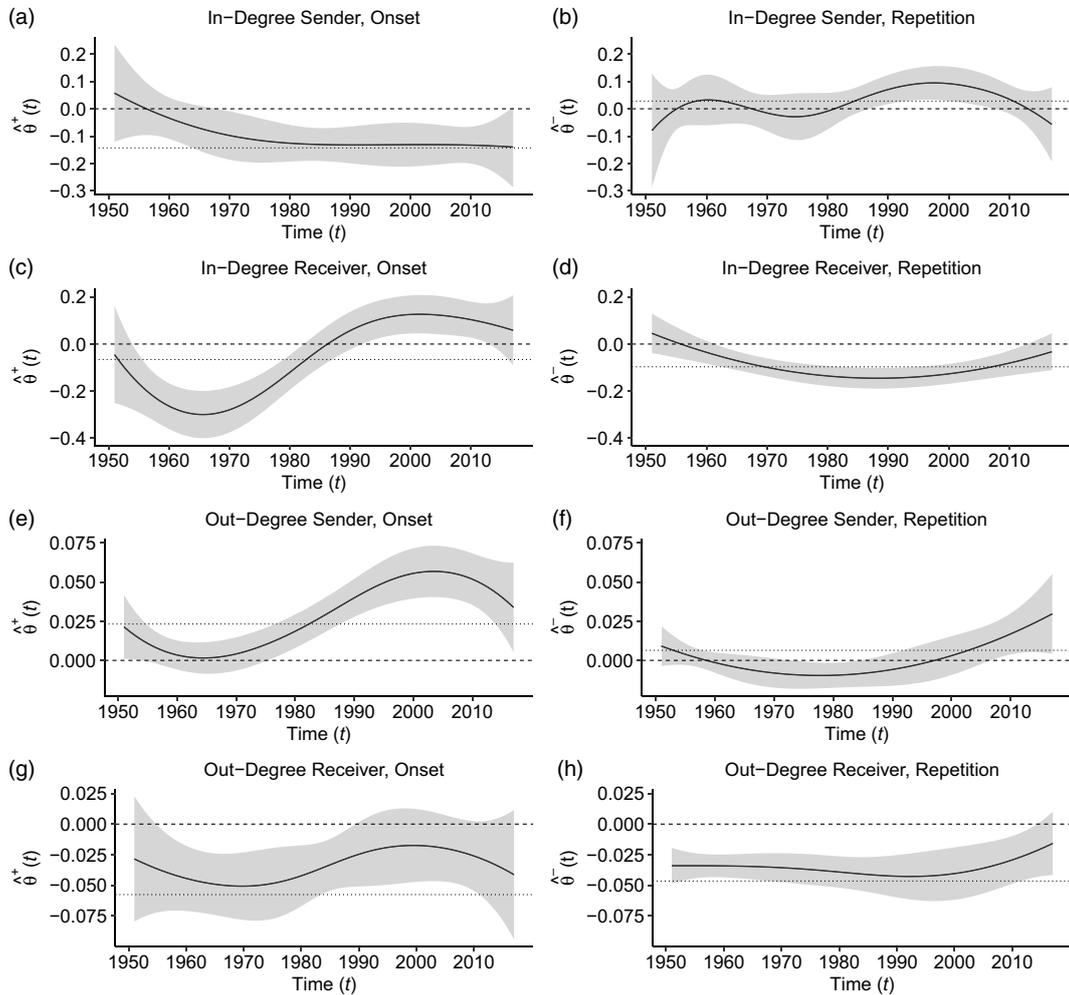


Figure 4. Results of endogenous statistics relating to centrality. The shaded area indicates the 95% confidence bands of the estimates and the dotted horizontal lines represent the time-constant parameters.

agreements (Leeds, 2019) and we operationalize regime dissimilarity by the absolute difference in the Polity IV scores of countries i and j (Marshall, 2017). This measure indicates all countries' year-wise regime characteristics and takes values from -10 (strongly autocratic) to 10 (strongly democratic). Thus, the absolute differences lie between 0 (strong similarity) and 20 (strong dissimilarity) for each country-dyad and year. The sources and used period of all incorporated exogenous covariates are described in more detail in the Supplementary Material.

3.3 Results

3.3.1 Fixed effects

In Figures 4–7, the full results of the time-varying estimates are given accompanied by alternative time-constant coefficients as dotted horizontal lines. The latter are obtained by setting $\theta(t) \equiv \theta$. All exponentially transformed estimates at a specific point in time can be interpreted (*ceteris paribus*) as the multiplicative change of the intensity (6) corresponding to the effect of covariates in relative risk models (Kalbfleisch & Prentice, 2002). Therefore, an effect estimated at zero does not change the relative risk of an event to happen, but positive or negative coefficients lead to a

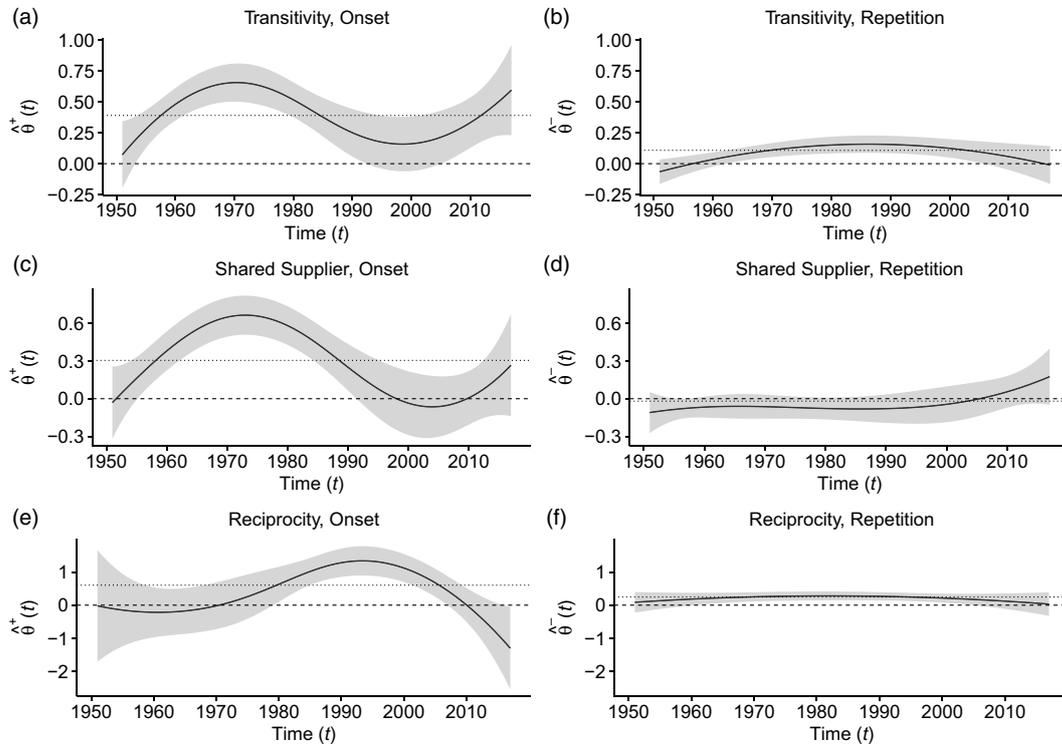


Figure 5. Results of endogenous statistics relating to past dyadic interaction and clustering. The shaded area indicates the 95% confidence bands of the estimates and the dotted horizontal lines represent the time-constant parameters.

higher or lower relative risk of the event to occur, respectively. Additionally, an event's occurrence is equivalent to the increment of one in the counts of aircraft units since one event represents a combat aircraft delivery in our application case.

From simple inspection, it can be concluded that in all cases, time-varying coefficients are carrying completely different information as compared to time-constant coefficients. This is evidence of the necessity to account for the multiple systemic changes within the international aircraft market during the considered time interval. From a statistical point of view, the time-varying effects can also be underpinned by a lower cAIC value when compared to time-constant effects (see Section 3.4 for additional details on the cAIC).

Moreover, we observe different shapes of the curves of the time-varying coefficients when comparing *onset* and *repetition* conditions leading to the conclusion that the import of all covariates on these two separate conditions is different.

Time-varying effects relating to the degree structure are shown in Figure 4. Figure 4(a) indicates a steady negative influence of the sender's in-degree in the *onset* condition from around 1965 onward. It can be concluded that the count of dyadic events is lower if the sender's in-degree is high, which may be justified by the observation that only a small subset of countries is adequately equipped to produce and export aircraft. This technological possibility, in turn, increases self-sufficient behavior, thus alleviating the need for additional imports. Contrary, in the *repetition* condition, the in-degree of the receiver exhibits a positive effect for the post-Cold War period from 1990 to 2010, Figure 4(b). Otherwise, the effect is insignificant. Concerning the receiver, a negative effect of the in-degree can be observed from 1950 to 1980 in the *onset* model, Figure 4(c). When proceeding to deliver aircraft, the receiver's in-degree effect is similar to the sender's in-degree effect, Figure 4(d). For the sender's out-degree, the effect in the *onset* model is negative until around 1980 and thereupon positive. In the latter case, the effect mirrors a higher tendency

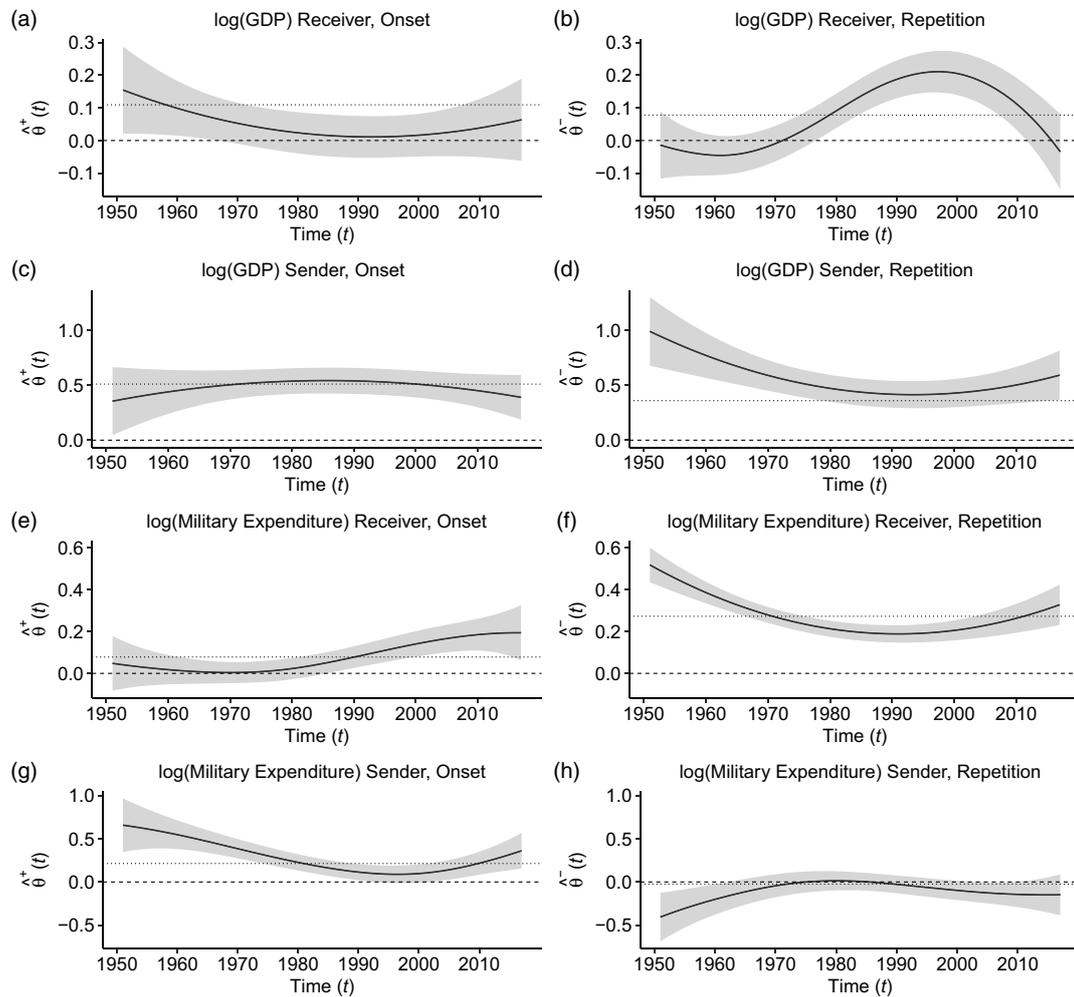


Figure 6. Results of exogenous statistics relating to economic factors. The shaded area indicates the 95% confidence bands of the estimates and the dotted horizontal lines represent the time-constant parameters.

of delivering combat aircraft if the sender is already a prolific exporter country. During the entire observational period, we observe that receivers are not senders themselves, thus exhibiting low out-degrees, Figure 4(g) and (h). This behavior does not depend on the condition of the dyadic intensity.

The specified triadic structures play a substantial role during the Cold War. Afterwards, the impact disappears but is again strengthened after 2000 under the *onset* condition, Figure 5(a) and (c). In particular, an increasing number of indirect transitive connections between country i and j results in a greater count of aircraft deliveries between 1950 and 1990. Similarly, receiving combat aircraft from the same third country increases the unit sales between the receivers during the Cold War period, Figure 5(c). A possible consequence of this process is the strengthening of a block structure. For a consecutive delivery, the triadic effects are less pronounced, and in the case of shared suppliers, Figure 5(d), constantly insignificant. The count of reciprocal events, on the other hand, raises trade from 1990 to 2005, Figure 5(e). This result may be a consequence of an international market opening after the Soviet Union's fall, leading to multiple emergent countries. If the relationship is maintained, reciprocal events are encouraged throughout the period of observation, although to a smaller degree, Figure 5(f).

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

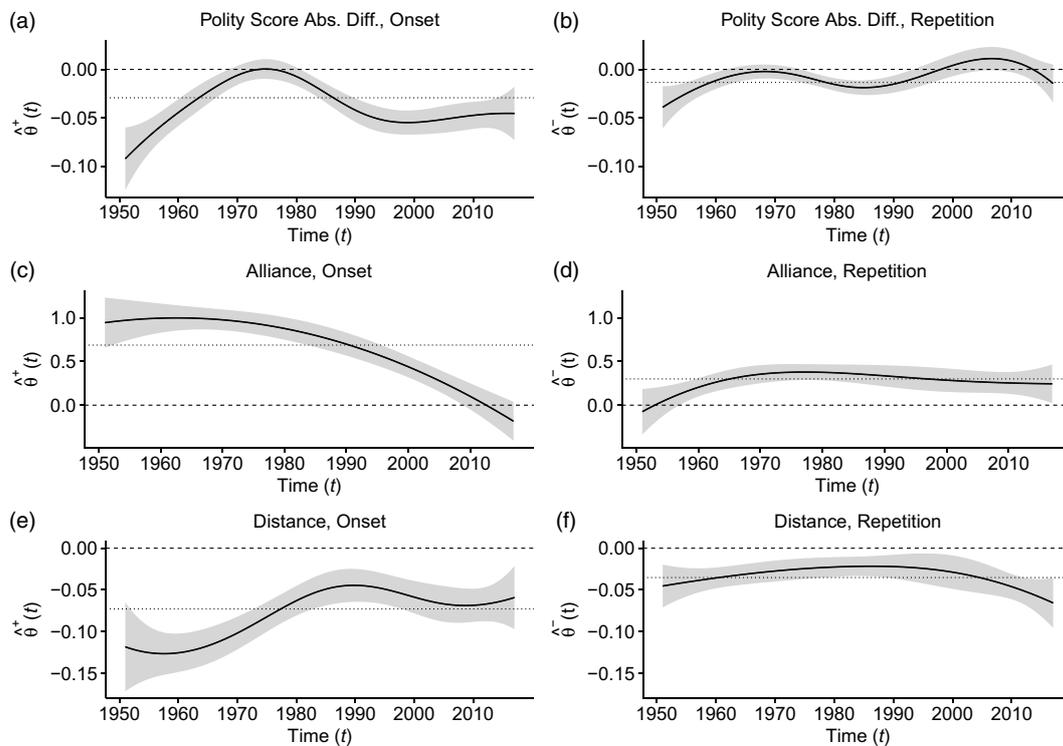


Figure 7. Results of exogenous statistics relating to political, security, and geographical factors. The shaded area indicates the 95% confidence bands of the estimates and the dotted horizontal lines represent the time-constant parameters.

While the logarithmic GDP of the receiver has a relatively weak positive influence when starting a trade relation, Figure 6(a), its repetition is only affected after the end of the Cold War, Figure 6(b). On the sender-side, the estimates of both models are constantly positive, Figure 6(c) and (d). In contrast to the effect in the onset model, the sender’s logarithmic GDP has a higher effect from 1950 to 1980 in the *repetition* condition. Moreover, the military expenditure of the receiver is one of the main drivers in this model, Figure 6(f). Here, higher military spending of possible sender countries augments the count of receiving combat aircraft deliveries, specifically during the 1950s. Conversely, the exogenous covariate only slowly gains attention in the *onset* condition after the Cold War, Figure 6(e). While the effect of the military expenses of the sender stays overall positive when delivering aircraft for the first time, it inhibits it to be repeated in the next year, Figure 6(g) and (h).

The findings in Figure 7(a) and (b) indicate that similar regimes are overall more likely to start trading combat aircraft. Only at the height of the Cold War from 1970 to 1980, the effect is estimated at approximately 0, Figure 7(a). The strength of the effect is less salient in the *repetition* condition than in the *onset* condition of our model, Figure 7(b). Furthermore, the time-varying coefficients discover a steadily decreasing influence of beginning to transact with allies, Figure 7(c). This finding suggests evidence of the overall deteriorating importance of international alliances in combat aircraft transactions if they did not trade in the previous year. We do not observe a similar downward trend when repeating an event, Figure 7(d). Lastly, a larger distance between the respective capitals generally hinders events from occurring, Figure 7(e) and (f). Therefore, countries tend to trade with spatially more close than distant partners. This may be caused by the relatively lower transportation cost and is in line with the expectations of the gravity model of trade (Martínez-Zarzoso & Johannsen, 2019; Thurner et al., 2019, see corrigendum).

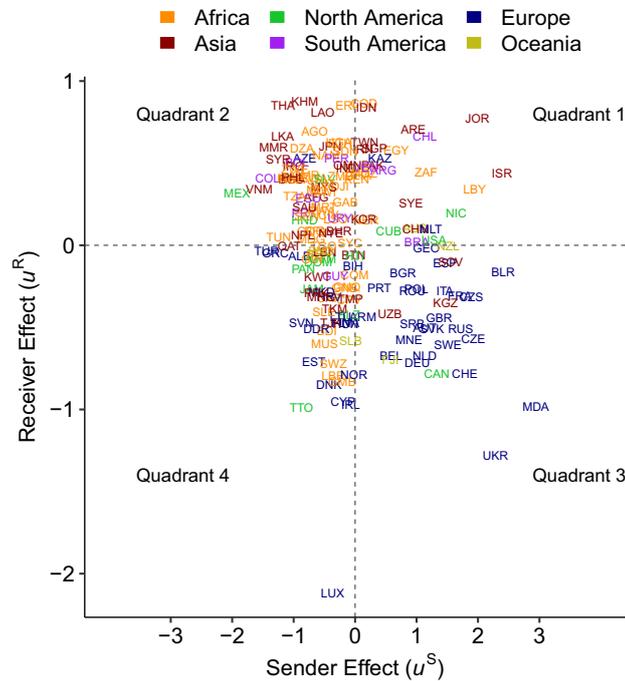


Figure 8. Country-specific random sender and receiver effects. The drawn label represents the respective ISO3 code of the represented country.

3.3.2 Random effects

The random effects permit an extended analysis of the unexplained heterogeneity in the model. More precisely, the random effects express country-specific deviations from an overall behavioral trend captured by the time-varying effects. Additionally, they correct the countries' repeated measurements as simultaneous senders and receivers of events in each year. The model introduced in Section 2 comprises two country-specific random effects for all countries as a sender and receiver of combat aircraft deliveries. The results are given in Figure 8 and visualized on a world map in Figure 9.

In the first quadrant of Figure 8, countries with a positive random sender and receiver effect are shown. This composition of random effects suggests that the respective countries are senders and receivers of more combat aircraft events than marginally expected. Countries in the Middle East, for example, Israel (ISR), Libya (LBY), and Jordanian (JOR), are allocated to this group.

Negative sender but positive receiver effects are identified for countries in South-East Asia (Thailand (THA), Cambodia (KHM), Laos (LAO), Myanmar (MYR), and Sri Lanka (LKA)). Compared to the average behavior, these countries are somewhat reluctant as senders and confident as receivers of combat aircraft deliveries. The latent sender effect of Mexico (MEX) is the most negative coefficient estimated. This suggests Mexico's reliance on the import of combat aircraft, although its high economic status would imply additional participation in the event network as a sender.

The third quadrant contains all countries, which were less active than expected as a sender and receiver of events. This strand of countries is either economically strong, yet exhibiting a passive trading behavior, for example, Luxembourg (LUX), or relatively poor and missing preconditions to send or receive weapons, for example, Trinidad and Tobago (TTO).

Lastly, a negative random coefficient regarding receiving arms is mostly associated with European countries. The corresponding sender effect is positive. Hence, these countries are

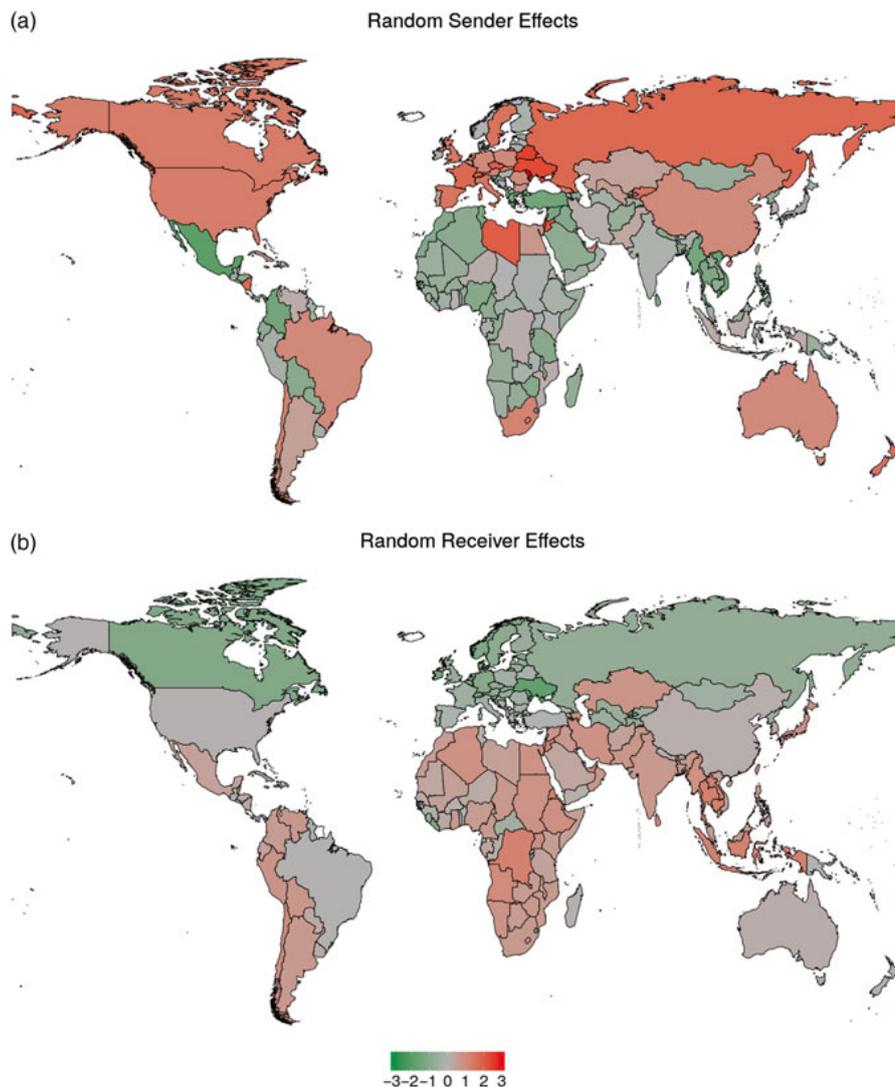


Figure 9. Random country-specific sender (a) and receiver (b) effects. The layout represents the borders as of 2020.

situated in the fourth quadrant of Figure 8. The East European countries Moldova (MDA), Ukraine (UKR), and Belarus (BLR) have the highest positive sender effect paired with relatively low receiver effects.

In terms of continent-wide tendencies, we locate Africa in the first three quadrants. South America is principally assigned to the first and second quadrant. Asia, Oceania, and North America are more dispersed and exhibit less homogeneous country behavior.

3.4 Model comparison and assessment

We compare the estimated model to alternative specifications, which are chosen to reflect all subsequent extensions of Section 2.2 and are indicated in Table 2. Model 1 includes all effects linearly without the separable extension. This is we assume that $\theta(t) \equiv \theta$ and omit the separation of the statistics $s_{ij}(y_{t-1}, x_{t-1})$ into $s_{ij}^+(y_{t-1}, x_{t-1})$ and $s_{ij}^-(y_{t-1}, x_{t-1})$. This separability is added in

Table 2. Specifications of the compared models and resulting corrected AIC (cAIC) values

	Separability	Time-varying effects	Random effects	cAIC
Model 1				84,622.47
Model 2	✓			65,614.86
Model 3	✓	✓		63,174.49
Model 4	✓	✓	✓	59,717.77

Model 2 according to Section 2.2.1. Model 3 includes time-varying coefficients as introduced in Section 2.2.2. Lastly, Model 4 is the model whose findings were presented in Section 3.3. Hence, also random effects are taken into account, which are explained in Section 2.2.3.

One way to compare these models is by means of information criteria, that is, the Akaike information criterion (AIC, Akaike, 1974). As already discussed in the context of linear mixed models (Greven & Kneib, 2010) and generalized mixed models (Saefken et al., 2014), the usage of the conditional or marginal AIC does not appropriately incorporate the uncertainty of estimating the covariance parameters of the random effects (in our application τ_S^2 and τ_R^2). Therefore, we utilize a corrected conditional AIC proposed by Wood et al. (2016). The resulting cAIC values are given in Table 2 and indicate a superior model fit when all extensions introduced in Section 2.2 are included.

We assess the selected Model 4 with a graphical tool proposed by Hunter et al. (2008) for general network models. The procedure's basic idea is to evaluate whether networks randomly generated according to the estimated network model at hand conserve pre-specified characteristics of the observed network reasonably well. In our particular case, we simulate yearly increments of our network counting process from Equation (11) and consider the result as a count-valued network. However, most network statistics commonly used for this assessment are solely defined for binary networks. Therefore, we propose a suite of novel statistics for our application case. To detect whether our model adequately replicates possible over- or underdispersion in the count data, we rely on the statistics from rootograms, that is, the empirical and simulated frequencies of the counts in the networks. For general regression tasks involving count data, rootograms were proposed by Kleiber & Zeileis (2016) and date back to Tukey (1977). Usually, one compares the square-root-transformed observed and expected frequencies of the target variable. However, in our application, we substitute the square-root transformation with a log transformation due to the high percentage of zeros and use the simulated rather than expected frequencies to fit into the framework of Hunter et al. (2008). Secondly, we investigate to what extent the performance of our model is stable over the time frame we analyze. To do so, we compute the clustering coefficient for weighted networks as proposed by Opsahl & Panzarasa (2009)⁴ for the yearly networks y_t . Besides, we examine the average in-count per year, which is directly related to the average count per year. In the Supplementary Material, we show how the distribution of the observed counts of in- and outgoing events given in Figure 2 is reproduced in the simulated networks and provide the mathematical formulations of all statistics.

Figure 10 shows the variability of all specified statistics computed for all 1,000 simulated networks through boxplots and displays the average value by a blue triangle. Red lines indicate the observed measurements. We can infer from Figure 10(a) that the estimated model captures even high event counts between countries averaged over the entire period. At the same time, our proposed model is capable of representing the yearly clustering as well as the average in-count, see Figure 10(b) and (c). Therefore, we gather that the performance of the proposed model is consistently good throughout the observational period.

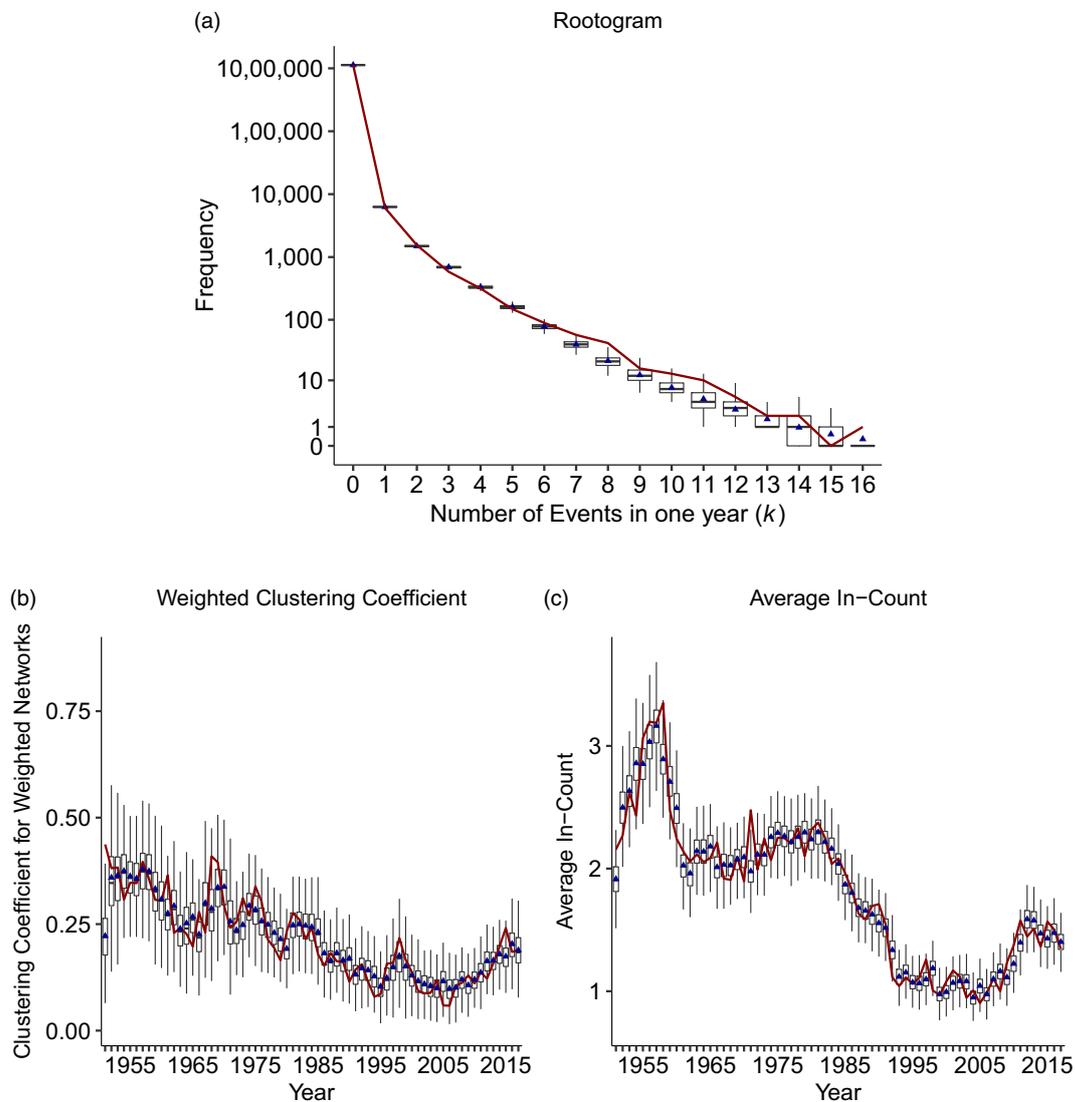


Figure 10. Comparison of the observed and simulated frequencies of the dyadic event counts (a), weighted clustering coefficients over time (b), and average in-count over time (c). The red lines indicate the observed values of each statistic, whereas the boxplots are the result of drawing 1,000 networks according to Equation (11) and the blue triangles the average values.

4. Conclusion

We introduced a novel model for the analysis of relational event data. Originating in a counting process operating in continuous time that we only observe at specific time points, we derived a tie-level intensity, whose parameters can be estimated according to the maximum likelihood principle. Extensions to separable models, which govern the *onset* and *repetition* of events by two functions, and the incorporation of time-varying and random coefficients are given. Eventually, we applied the procedure to the international combat aircraft network from 1950 to 2017. By doing that, we use the additional information provided by the counts of yearly aircraft deliveries to estimate a time-continuous intensity, contrary to existing work on binarized networks. Moreover, the separability detects fundamentally different processes governing the *onset* and *repetition* of event relationships, while the time-varying effects uncover a systemic change during the Cold War period. Furthermore, we identified triangular network statistics and the sender's economic

nodal covariates as the principal drivers of the *onset* condition of the proposed intensity. Here, a decaying effect of bilateral military alliances became apparent. For the *repetition* condition, this effect remained consistently positive, and the receiver's high military expenditure was shown to be the driving force. Finally, the random effects enable a visual comparison of the unexplained heterogeneity between the modeled countries (Figure 9) and correct the estimates for repeated measurements as well as possible overdispersion.

Supplementary materials. For supplementary material for this article, please visit <http://dx.doi.org/10.1017/nws.2021.9>

Acknowledgments. We thank the anonymous reviewers for their careful reading and constructive comments.

Financial support. The project was supported by the European Cooperation in Science and Technology [COST Action CA15109 (COSTNET)]. We also gratefully acknowledge funding provided by the German Research Foundation (DFG) for the project KA 1188/10-1 and TH 697/9-1: *International Trade of Arms: A Network Approach*. Furthermore, the work has been partially supported by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibilities for its content.

Conflict of interest. None.

Notes

1 We further provide a descriptive analysis in the Supplementary Material to demonstrate high positive auto-correlation of the endogenous covariates between consecutive years; therefore, they are a reliable proxy of simultaneous dependence.

2 In the Supplementary Material, we deliver the results regarding alternative models for the data. Overall, there is no relevant difference to the findings presented subsequently.

3 As a robustness check, we compare different time frames to define which events are driven by the *onset* and *repetition* intensity, for example, having delivered combat aircraft the last one or two years in the Supplementary Material.

4 We opt for the variant of the statistic that aggregates triplets of event counts within a year via the arithmetic mean.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Akerman, A., & Seim, A. L. (2014). The global arms trade network 1950–2007. *Journal of Comparative Economics*, 42(3), 535–551.
- Almquist, Z. W., & Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, 44(1), 273–321.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.
- Bauer, V., Harhoff, D., & Kauermann, G. (2021). A smooth dynamic network model for patent collaboration data. In *AStA advances in statistical analysis*.
- Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., & Starz, J. (2018). ICEWS automated daily event data. Retrieved from <https://doi.org/10.7910/DVN/QI2T9A> (visited 2019-09-23).
- Box-Steffensmeier, J. M., Christenson, D. P., & Morgan, J. W. (2018). Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. *Political Analysis*, 26(1), 3–19.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9–25.
- Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1), 155–200.
- Butts, C. T., & Marcum, C. S. (2017). A relational event approach to modeling behavioral dynamics. In A. Pilny, & M. S. Poole (Eds.), *Group processes* (pp. 51–92). Cham: Springer.
- Cranmer, S. J., Desmarais, B. A., & Menninga, E. J. (2012). Complex dependencies in the alliance network. *Conflict Management and Peace Science*, 29(3), 279–313.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
- Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review*, 35(5), 843–851.
- de Boor, C. (2001). *A practical guide to splines*. New York: Springer.
- Desmarais, B. A., & Cranmer, S. J. (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PLoS ONE*, 7(1), 1–12.

- Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–102.
- Forsberg, R. (1994). *The arms production dilemma: Contraction and restraint in the world combat aircraft industry*. Cambridge: MIT Press.
- Forsberg, R. (1997). The contraction of the world military aircraft industry. In M. Kaldor, B. Vashee, & G. Scheder (Eds.), *The end of military fordism: Restructuring the global military sector*. London: Bloomsbury.
- Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *Annals of Statistics*, 10(1), 101–113.
- Fritz, C., Lebacher, M., & Kauermann, G. (2020). Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time. *Statistica Neerlandica*, 74(3), 275–299.
- Gjessing, H. K., Røysland, K., Pena, E. A., & Aalen, O. O. (2010). Recurrent events and the exploding Cox model. *Lifetime Data Analysis*, 16(4), 525–546.
- Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–233.
- Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, 97(4), 773–789.
- Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4, 585–605.
- Hoefler, C., & Mérand, F. (2016). Buying a fighter jet: European lessons for Canada. *Canadian Foreign Policy Journal*, 22(3), 262–275.
- Hoffman, M., Block, P., Elmer, T., & Stadtfeld, C. (2020). A model for the dynamics of face-to-face interactions in social groups. *Network Science*, 8(S1), 1–22.
- Holland, P., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, 2(2), 107–124.
- Holland, P., & Leinhardt, S. (1977). A dynamic model for social networks. *The Journal of Mathematical Sociology*, 5(1), 5–20.
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248–258.
- Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data*. Hoboken: Wiley.
- Kauermann, G., Krivobokova, T., & Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 487–503.
- Kauermann, G., & Opsomer, J. D. (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika*, 98(1), 225–230.
- Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12, 105–135.
- Kleiber, C., & Zeileis, A. (2016). Visualizing count data regressions using rootograms. *The American Statistician*, 70(3), 296–303.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data. Methods and models*. New York: Springer.
- Kolaczyk, E. D. (2017). *Topics at the frontier of statistics and network analysis: (Re)visiting the foundations*. Cambridge: Cambridge University Press.
- Kreiß, A., Mammen, E., & Polonik, W. (2019). Nonparametric inference for continuous-time event counting and link-based dynamic network models. *Electronic Journal of Statistics*, 13(2), 2764–2829.
- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6, 1100–1128.
- Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 29–46.
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3), 204–213.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723.
- Lebacher, M., Thurner, P. W., & Kauermann, G. (2020). Exploring dependence structures in the international arms trade network: A network autocorrelation approach. *Statistical Modelling*, 20(2), 195–218.
- Lebacher, M., Thurner, P. W., & Kauermann, G. (2021). A dynamic separable network model with actor heterogeneity: An application to global weapons transfers. *Journal of Royal Statistical Society: Series A (Statistics in Society)*.
- Leeds, B. A. (2019). Alliance treaty obligations and provisions (ATOP 4.01). Retrieved from <http://www.atopdata.org/> (visited 2019-09-30).
- Levine, P., Sen, S., & Smith, R. (1994). A model of the international arms market. *Defence and Peace Economics*, 5(1), 1–18.
- Li, Z., & Wood, S. N. (2020). Faster model matrix crossproducts for large generalized linear models with discretized covariates. *Statistics and Computing*, 30, 19–25.
- Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential random graph models for social networks*. Cambridge: Cambridge University Press.

- Marshall, M. G. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2016. Retrieved from <http://www.systemicpeace.org/inscrdata.html> (visited 2019-09-16).
- Martínez-Zarzoso, I., & Johannsen, F. (2019). The gravity of arms. *Defence and Peace Economics*, 30(1), 2–26.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.
- Mehrl, M., & Thurner, P. W. (2020). Military technology and human loss in intrastate conflict: The conditional impact of arms imports. *Journal of Conflict Resolution*, 64(6), 1172–1196.
- Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(S1), 2566–2572.
- Nordhaus, W., Oneal, J. R., & Russett, B. (2012). The effects of the international security environment on national military expenditures: A multicountry study. *International Organization*, 66(3), 491–513.
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245–251.
- Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, 31(2), 155–163.
- Pamp, O., Dendorfer, F., & Thurner, P. W. (2018). Arm your friends and save on defense? The impact of arms exports on military expenditures. *Public Choice*, 177, 165–187.
- Robins, G., & Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1), 5–41.
- Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika*, 64(3), 371–394.
- Ruppert, D., Wand, M., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge: Cambridge University Press.
- Ruppert, D., Wand, M., & Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*, 3(3), 1193–1256.
- Saeffken, B., Kneib, T., van Waveren, C.-S., & Greven, S. (2014). A unifying approach to the estimation of the conditional Akaike information in generalized linear mixed models. *Electronic Journal of Statistics*, 8(1), 201–225.
- Singer, J. D., Bremer, S., & Stuckey, J. (1972). Capability distribution, uncertainty, and major power war, 1820–1965. In B. Russett (Ed.), *Peace, war, and numbers* (pp. 19–48), vol. 19. Sage.
- SIPRI. (2019). Military expenditure database. Retrieved from <https://www.sipri.org/databases/milex> (visited 2020-09-03).
- SIPRI. (2020a). Arms transfers database. Retrieved from <https://www.sipri.org/databases/armstransfers> (visited 2020-03-09).
- SIPRI. (2020b). Arms transfers database: Sources and methods. Retrieved from <https://www.sipri.org/databases/armstransfers/sources-and-methods> (visited 2020-03-09).
- Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, 21, 149–172.
- Snijders, T. A. B. (2003). Accounting for degree distribution in empirical analysis of network dynamics. In *Dynamic social network modeling and analysis: Workshop summary and papers* (pp. 146–161). The National Academies Press.
- Snijders, T. A. B. (2017). Comment: Modeling of coordination, rate functions, and missing ordering information. *Sociological Methodology*, 47(1), 41–47.
- Snijders, T. A. B., & van Duijn, M. (1997). Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena*. Cham: Springer.
- Stadtfeld, C. (2012). *Events in social networks: A stochastic actor-oriented framework for dynamic event processes in social networks*. Ph.D. thesis, KIT.
- Stadtfeld, C., Hollway, J., & Block, P. (2017). Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology*, 47(1), 1–40.
- Thiemichen, S., Friel, N., Caimo, A., & Kauermann, G. (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46, 11–28.
- Thurner, P. W., Schmid, C. S., Cranmer, S. J., & Kauermann, G. (2019). Network interdependencies and the evolution of the international arms trade. *Journal of Conflict Resolution*, 63(7), 1736–1764.
- Tukey, J. W. (1977). *Exploratory data analysis*. London: Pearson.
- Tutz, G., & Schmid, M. (2016). *Modeling discrete time-to-event data*. Cham: Springer.
- Vu, D., Hunter, D., Smyth, P., & Asuncion, A. (2011). Continuous-time regression models for longitudinal networks. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, vol. 24.
- Vucetic, S. (2011). Canada and the F-35: What's at stake? *Canadian Foreign Policy Journal*, 17(3), 196–203.
- Vucetic, S., & Nossal, K. R. (2012). The international politics of the F-35 Joint Strike Fighter. *International Journal*, 68(1), 3–12.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.
- Whitehead, J. (1980). Fitting cox's regression model to survival data using glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(3), 268.

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

Network Science 311

- Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian and New Zealand Journal of Statistics*, 48(4), 445–464.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Boca Raton: CRC press.
- Wood, S. N., Li, Z., Shaddick, G., & Augustin, N. H. (2017). Generalized additive models for gigadata: Modeling the U.K. black smoke network daily data. *Journal of the American Statistical Association*, 112(519), 1199–1210.
- Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.

Cite this article: Fritz C., Thurner P.W. and Kauermann G. (2021). Separable and semiparametric network-based counting processes applied to the international combat aircraft trades. *Network Science* 9, 291–311. <https://doi.org/10.1017/nws.2021.9>

Supplementary Material: Separable and Semiparametric Network-based Counting Processes applied to the International Combat Aircraft Trades

Cornelius Fritz*, Paul W. Thurner[†], Göran Kauermann*
Department of Statistics, LMU Munich*
Geschwister Scholl Institute of Political Science, LMU Munich[†]

August 16, 2021

Contents

A	Correlation of Endogenous Statistics between Subsequent Years	2
B	Data Sources	3
C	Further Descriptive Analysis	3
D	Robustness Checks	5
	D.1 Weighted Fit	5
	D.2 Alternative Time-Spans defining Separability	8
	D.3 Thresholds for TIV of Events	10
	D.4 Corrected AIC for Finite Sample Size	14
E	Further Results of the Model Assessment	14

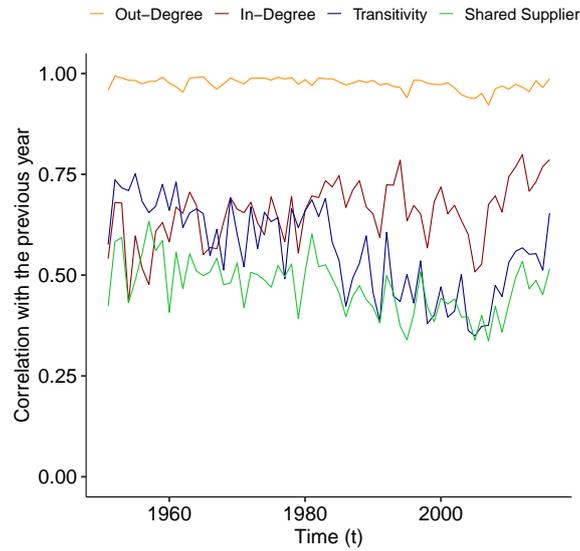


Figure 1: Yearly auto-correlations of endogenous covariates.

A Correlation of Endogenous Statistics between Subsequent Years

To further legitimize the usage of lagged endogenous covariates, we investigate the yearly auto-correlations of the corresponding statistics. Therefore, we construct time series on the monadic level for the in- and out-degrees of each country and at the dyadic level for triangular statistics, i.e., regarding a tuple of countries. In Figure 1, we then descriptively analyze the yearly correlation between all statistics, where measurements are available at both time points. The results again highlight the reliability of using the endogenous statistics of the past year as a proxy for the current year, as we observe exceptionally high correlations.

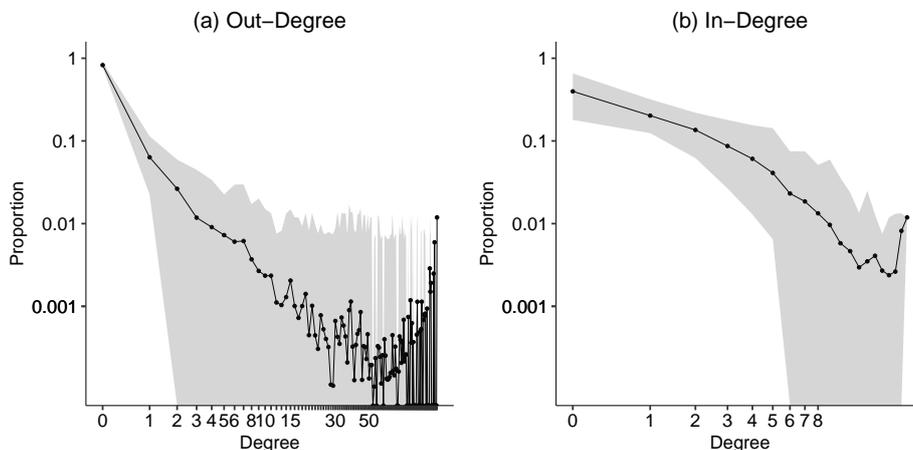


Figure 2: Average Degree Distributions of the Out- and In-Degree for all included countries. The shaded area represents the minimum and maximum observed value. All graphs are represented on a logarithmic scale.

B Data Sources

Table 1: Data sources of the exogenous covariates. Versions are indicated where available.

Covariate	From	To	Data Source
GDP, Base-Year 2005	1950	2011	Gleditsch (2002), v4.1
	2012	2017	World Bank (2017)
Military Expenditure, Base-Year 2017	1950	2000	Singer et al. (1972), v5.0
	2000	2017	SIPRI (2019)
Polity Score	1950	2017	Marshall (2017)
Alliance	1950	2017	Leeds (2019), v4.01
Distance of Capitals	1950	2017	Gleditsch (2013)

C Further Descriptive Analysis

The distribution of the in- and out-degrees can be used to analyze the topology of general networks (Barabási and Albert, 1999; Snijders, 2003; Newman et al., 2002). Similar to the findings in Figure 2 of the main article, 2 (a) underpins the strong centralization of the out-degree distribution. Again mirroring the results of the main article, the in-degree distribution is not as skewed, Figure 2 (b). There are few high degree countries, but the mode is still at zero.

Alternatively, we can focus the descriptive analysis on the top 10 sender and receiver in the network. The yearly counts of the respective countries are represented as boxplots in Figure 3 and 4. The exposed situation of USA is clearly visible, especially in Figure 3. This role was already thoroughly analyzed in Lorell (2003). India predominantly buys combat aircraft from Great Britain, which reflects the dyadic colonial history. Japan, on

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

the other hand, obtains 95% of the delivered aircraft from USA, being the second highest receiving country.

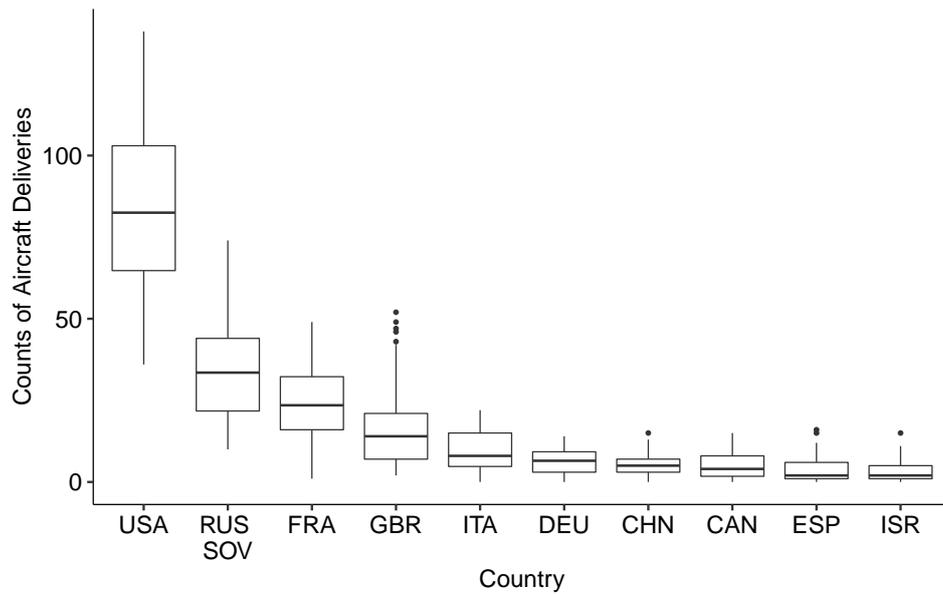


Figure 3: Boxplot of the observed counts over the years of the top 10 sender countries. The labels are the ISO3 codes of the respective countries.

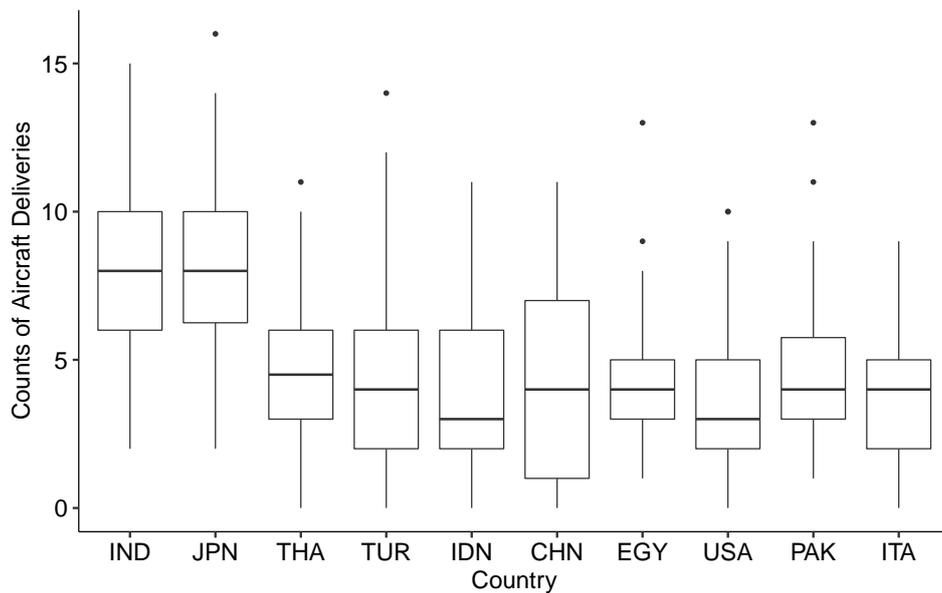


Figure 4: Boxplot of the observed counts over the years of the top 10 sender countries. The labels are the ISO3 codes of the respective countries.

D Robustness Checks

D.1 Weighted Fit

Each event can be comprehended as having a weight given by its TIV. As most possible events in our application were not realized, the respective TIVs are set to zero. Therefore, the weight of the tuple between country i and j at time point t is given by $w_{ij}(t) \propto \log(\text{TIV}_{ij}(t) + 1) + 1$, where $\text{TIV}_{ij}(t)$ denotes the aggregated TIVs of the same country tuple in the year t . The proportionality stems from the fact, that the weights are subsequently standardized so that their sum equals 1.

Figures 5, 6, and 7 contrast the estimates resulting from the original and weighted fit. The substantial conclusions drawn in Section 4 of the main article are paralleled by the weighted estimates.

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

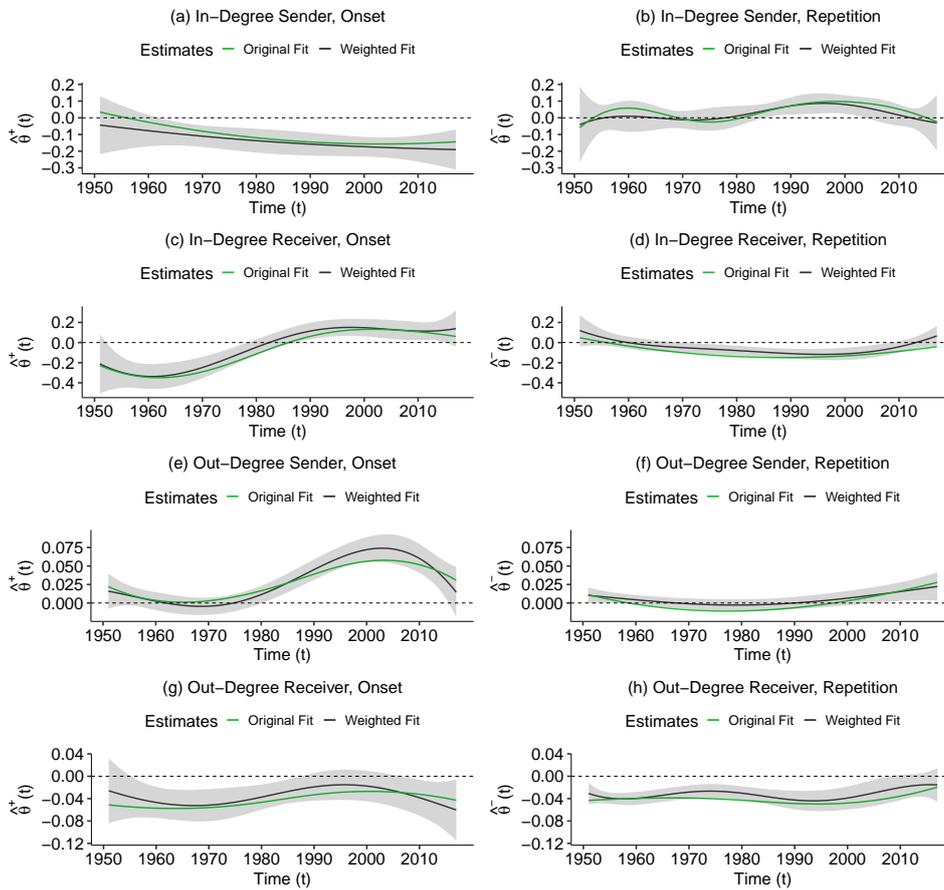


Figure 5: Robustness checks of the estimated parameters comparing the original fit to the model that weighted the observations according to the respective TIV. The green line represents the original fit, while the shaded area indicates the 95% quantile confidence bands of the weighted estimation.

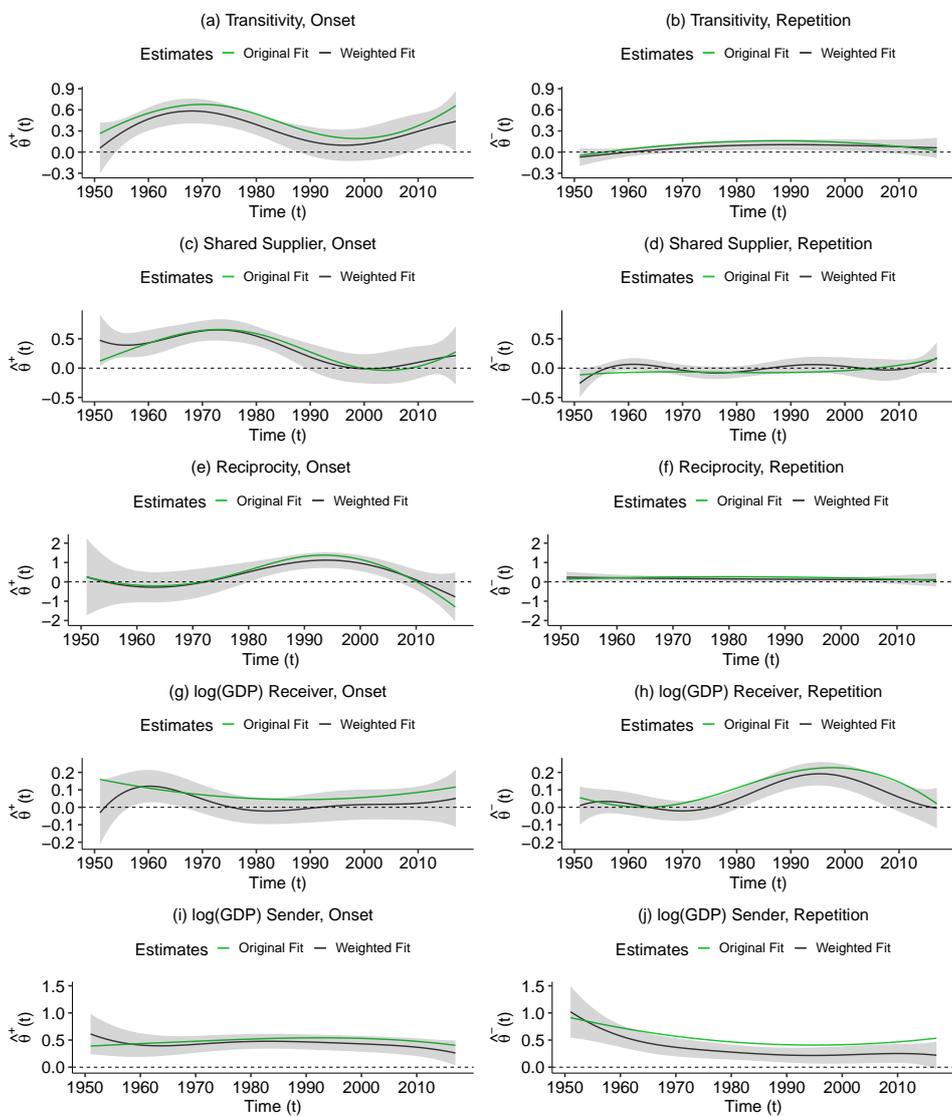


Figure 6: Robustness checks of the estimated parameters comparing the original fit to the model that weighted the observations according to the respective TIV. The green line represents the original fit, while the shaded area indicates the 95% quantile confidence bands of the weighted estimation.

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

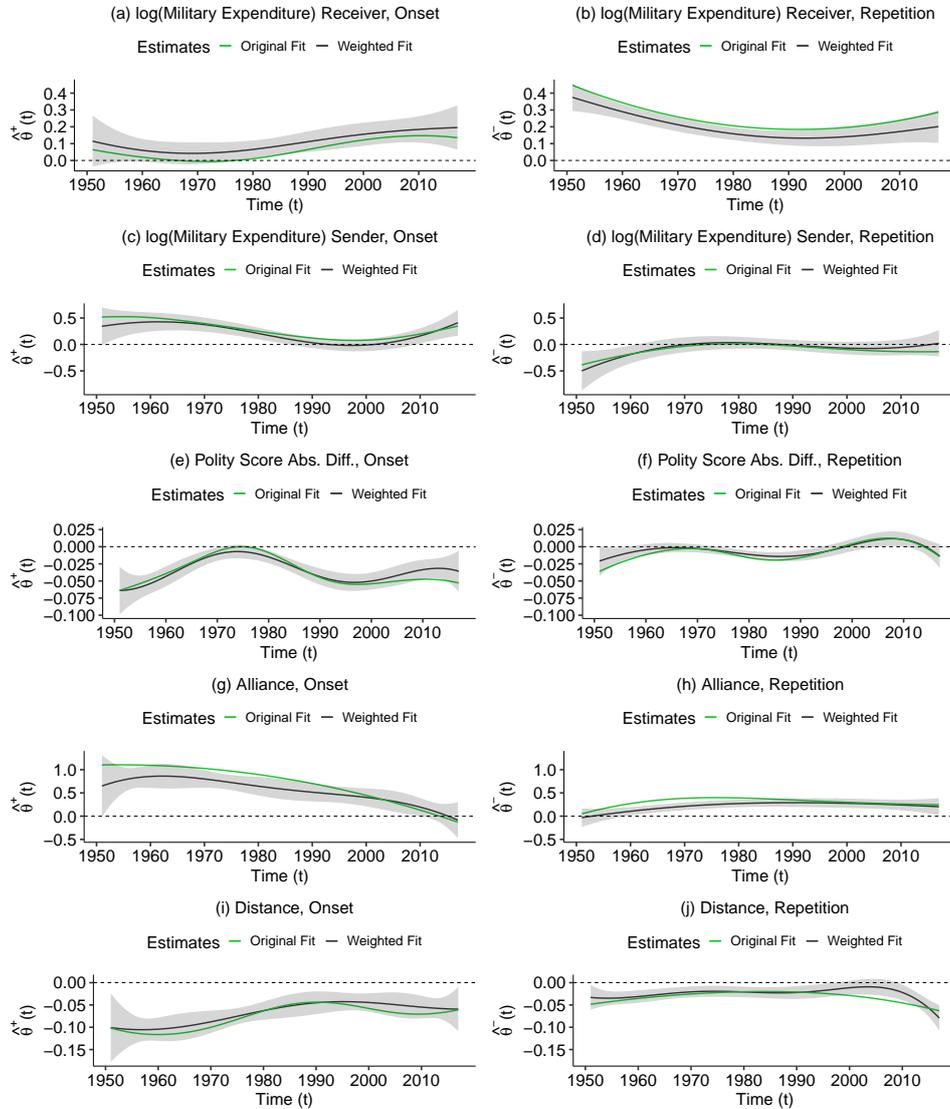


Figure 7: Robustness checks of the estimated parameters comparing the original fit to the model that weighted the observations according to the respective TIV. The green line represents the original fit, while the shaded area indicates the 95% quantile confidence bands of the weighted estimation.

D.2 Alternative Time-Spans defining Separability

The separability assumption can be adapted by changing the time frame, dictating which intensity governs which event. In the application case we fixed this interval to be one year. In order to legitimize this decision, we estimated the exact same model with a varying interval length defining from when an event tuple is, e.g., driven by the *onset* intensity.

For instance, a lag of 10 years would translate to being driving by the *onset* intensity if two countries did not trade with each other in the last 10 years. Figure 8 plots the AIC scores and values of the log likelihood evaluated at the final estimates of the respective models over the lag. Apparently, there are only slight differences between using a lag of one or two years, yet longer lags lead to a steadily deteriorating performance of the model.

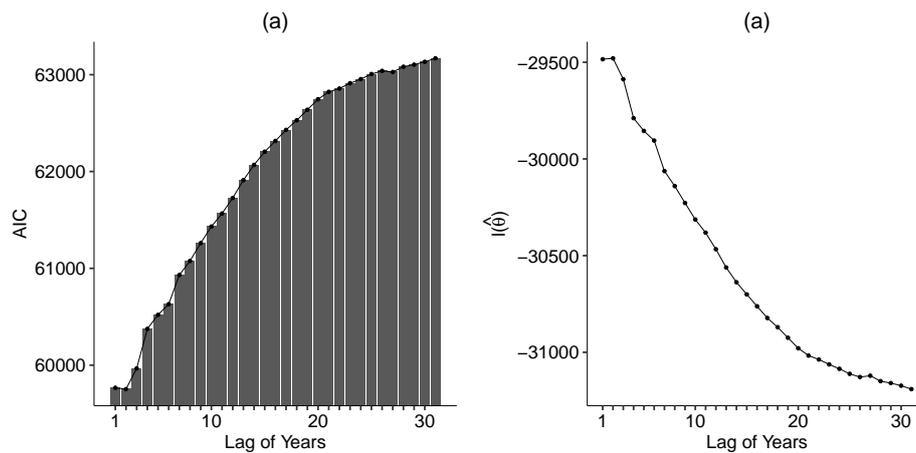


Figure 8: (a): Resulting AIC value by varying the length of the interval defining the separability. (b): The value of the log likelihood evaluated at the final estimates of the respective models.

D.3 Thresholds for TIV of Events

In the application of Section 3 all events were regarded unconditional of their extent. Alternatively, one may only include events above a certain threshold in terms of TIVs of the events. As a robustness check of the findings in the article, we, therefore, repeat the parameter estimation in three different scenarios, which are defined as follows:

1. Include events, if their TIV is above the 0.05 quantile of all TIVs ($> z_{0.05}$)
2. Include events, if their TIV is above the 0.1 quantile of all TIVs ($> z_{0.1}$)
3. Include events, if their TIV is above the 0.15 quantile of all TIVs ($> z_{0.15}$)
4. Include all events (Full Data)

The resulting estimates are shown in Figures 9 to 11 and proof the robustness of Figures 4 to 7. More specifically, equal interpretations and conclusions stated in Section 3.3.1 still hold. Only slight variations are visible in Figure 9 (g) concerning the out-degree of the receiver. Comparing the confidence bands of the original model with the estimates of the conditional models, we observe full coverage in most cases.

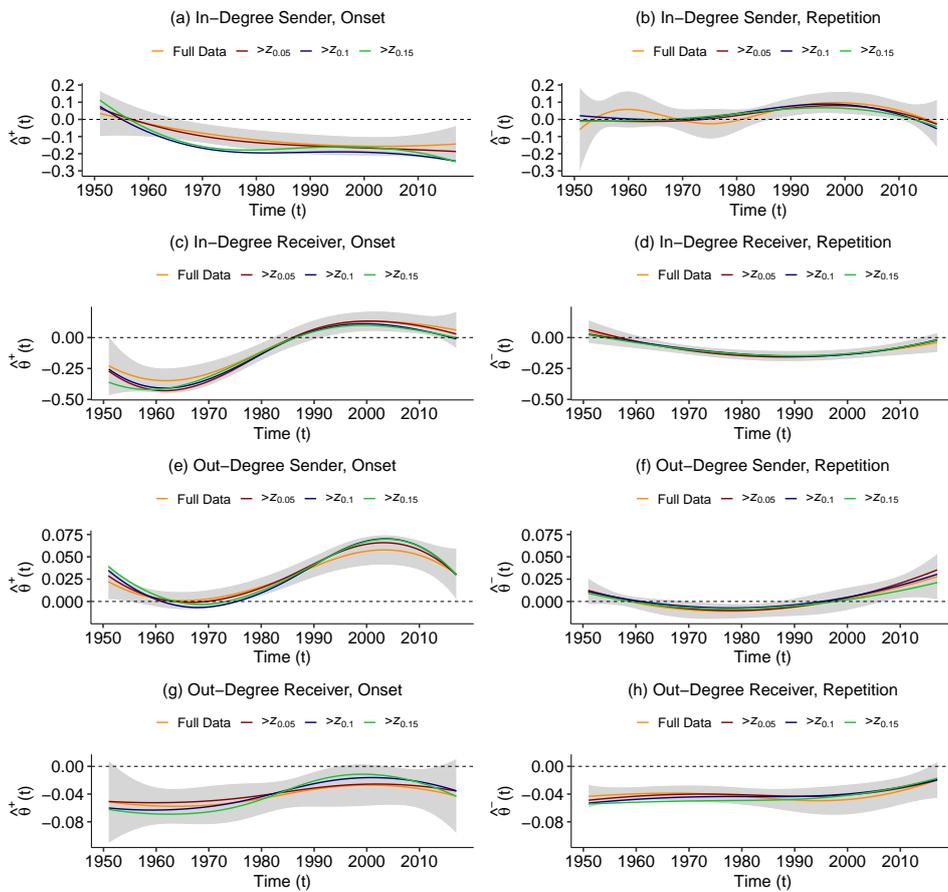


Figure 9: Robustness checks of the estimated parameters when only events with a specific TIV are regarded. The shaded area indicates the 95% confidence bands of the estimates from the unconditional model including all events.

10. Separable and semiparametric network-based counting processes applied to the international combat aircraft trades

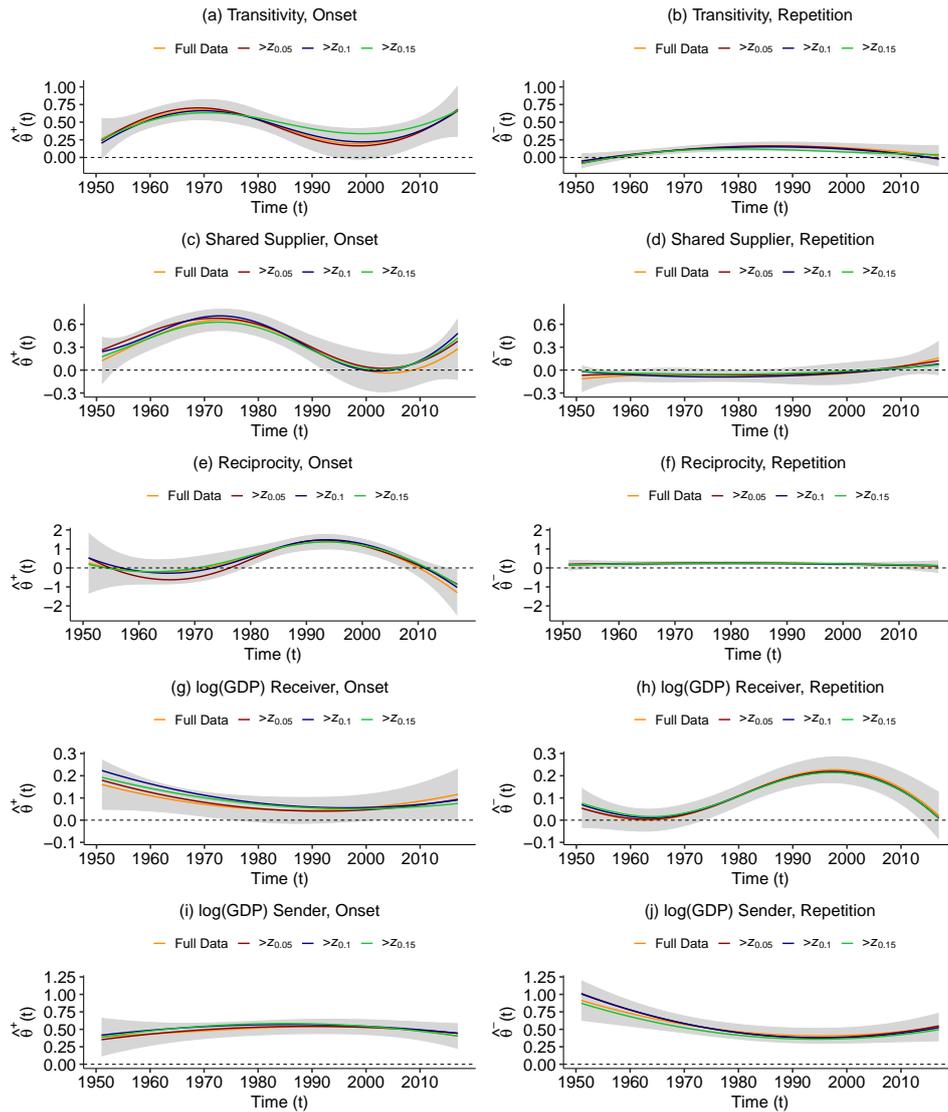


Figure 10: Robustness checks of the estimated parameters when only events with a specific TIV are regarded. The shaded area indicates the 95% confidence bands of the estimates from the unconditional model including all events.

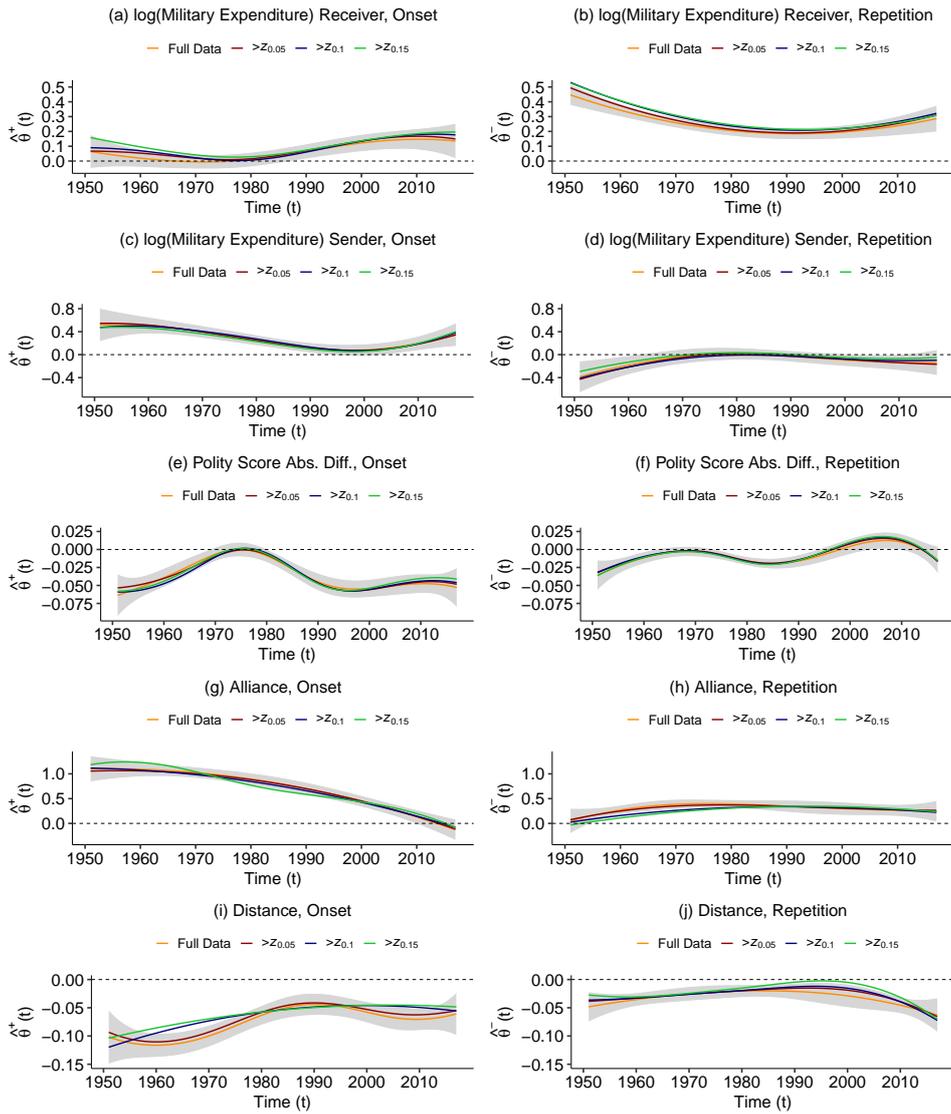


Figure 11: Robustness checks of the estimated parameters when only events with a specific TIV are regarded. The shaded area indicates the 95% confidence bands of the estimates from the unconditional model including all events.

Table 2: Specifications of the compared models and resulting corrected AIC_c value.

	Separability	Time-Varying Effects	Random Effects	AIC_c
Model 1	✗	✗	✗	84622.47
Model 2	✓	✗	✗	65614.86
Model 3	✓	✓	✗	63174.54
Model 4	✓	✓	✓	59718.04

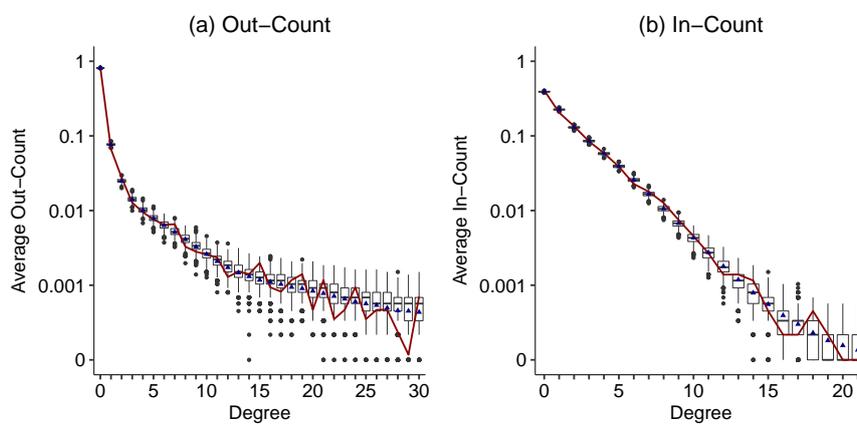


Figure 12: Comparison of the observed and simulated count distributions of the Out- (a) and In-Counts (b) for all included countries summed up over all years. The red lines indicate the observed values of each respective case, whereas the boxplots are the result of drawing 1000 networks and the blue triangles the average values.

D.4 Corrected AIC for Finite Sample Size

Besides correcting for the uncertainty resulting from estimating the variance and tuning parameters of the random and smooth components, we can define a version of the same AIC value that corrects for finite sample sizes as proposed by Hurvich and Tsai (1989). Table 2 reports this type of AIC value, although the results do not change compared to the values reported in the main article.

E Further Results of the Model Assessment

We begin by giving the mathematical formulations of the three network statistics for weighted networks analyzed in Section 3.4 of the main article. For the rootogram, we compute the frequencies h_k of combat aircraft deliveries $k \in \{1, \dots\}$ over all year. We calculated the weighted clustering coefficient proposed by Opsahl and Panzarasa (2009) for the increments of our network counting process in each year. For the increments \mathbf{y}_t in year t , we hence count the total value of the closed triplets and all triplets and define the generalized clustering coefficient by their ratio. We specify a triplet's value as the arithmetic mean of all observed weights, i.e., the number of yearly deliveries in our application case. The in-count of all countries in year t determines the yearly average

in-count. For country i the in-count in year t is defined by $\text{in-count}(i, t) = \sum_{j=1}^n y_{ji,t}$. Taking the arithmetic mean over all in-count(i, t) $\forall i \in \mathcal{A}_t$, where the set \mathcal{A}_t includes all countries present in the trade network in year t , gives the average in-count per year. The resultant statistic is proportional to the average events per year. We define the out-count in the same line. If we then concatenate all in- or out-counts over all years, the resulting empirical distribution represents the in- or out-counts irrespective of time. Figure 12 gives visual proof that our model can conserve both the in- and out-count distributions.

References

- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Gleditsch, K. S. (2002). Expanded trade and GDP data. *Journal of Conflict Resolution*, 46(5):712–724.
- Gleditsch, K. S. (2013). Distance between capital cities.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.
- Leeds, B. A. (2019). Alliance areaty obligations and provisions (ATOP 4.01). <http://www.atopdata.org/> (visited 2019-09-30).
- Lorell, M. A. (2003). *The U.S. combat aircraft industry, 1909-2000 : structure, competition, innovation*. RAND Corporation, Santa Monica, CA; Arlington, VA; Pittsburgh, PA.
- Marshall, M. G. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2016. <http://www.systemicpeace.org/inscrdata.html> (visited 2019-09-16).
- Newman, M. E. J., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(S 1):2566–2572.
- Opsahl, T. and Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, 31(2):155–163.
- Singer, J. D., Bremer, S., and Stuckey, J. (1972). Capability distribution, uncertainty, and major power war, 1820-1965. In Russett, B., editor, *Peace, War, and Numbers*, volume 19, pages 19–48. Sage.
- SIPRI (2019). Military expenditure database. <https://www.sipri.org/databases/milex> (visited 2020-09-03).
- Snijders, T. A. B. (2003). Accounting for degree distribution in empirical analysis of network dynamics. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 146–161. The National Academies Press.
- World Bank (2017). World Bank Open Data.

11. All that Glitters is not Gold: Modeling Relational Events with Spurious Events

Contributing article

Fritz, C., Mehrl, M., Thurner, W. P. and Kauermann, G. (2022) All that Glitters is not Gold: Modeling Relational Events with Spurious Events. *Under review in Network Science*.

Replication code

The complete replication code for this article is available under <https://github.com/corneliusfritz/Replication-Code-All-that-Glitters>

Author contributions

The methodological idea to propose a method for spurious events came from Cornelius Fritz, who implemented the method and wrote the initial draft of the publication. The introduction and application Sections were written jointly with Marius Mehrl, who especially contributed to the framing in Section 1 and the substantive context in Section 4. Paul W. Thurner contributed to Section 1 and Section 4. Every author was involved in proofreading all parts of the manuscript.

All that Glitters is not Gold: Relational Events Models with Spurious Events

Cornelius Fritz^{*1}, Marius Mehrl[†], Paul W. Thurner[†], Göran Kauermann^{*}

Department of Statistics, LMU Munich^{*}

Geschwister Scholl Institute of Political Science, LMU Munich[†]

May 24, 2022

Abstract

As relational event models are an increasingly popular model for studying relational structures, the reliability of large-scale event data collection becomes more and more important. Automated or human-coded events often suffer from non-negligible false-discovery rates in event identification. And most sensor data is primarily based on actors' spatial proximity for predefined time windows; hence, the observed events could relate either to a social relationship or random co-location. Both examples imply spurious events that may bias estimates and inference. We propose the Relational Event Model for Spurious Events (REMSE), an extension to existing approaches for interaction data. The model provides a flexible solution for modeling data while controlling for spurious events. Estimation of our model is carried out in an empirical Bayesian approach via data augmentation. Based on a simulation study, we investigate the properties of the estimation procedure. To demonstrate its usefulness in two distinct applications, we employ this model to combat events from the Syrian civil war and student co-location data. Results from the simulation and the applications identify the REMSE as a suitable approach to modeling relational event data in the presence of spurious events.

1 Introduction

In recent years, event data have become ubiquitous in the social sciences. For instance, interpersonal structures are examined using face-to-face interactions (Elmer and Stadtfeld, 2020). At the same time, political event data are employed to study and predict the occurrence and intensity of armed conflict (Fjelde and Hultman, 2014; Blair and Sambanis, 2020; Dorff et al., 2020). Butts (2008a) introduced the Relational Event Model (REM) to study such relational event data. In comparison to standard network data of durable relations observed at specific time points, relational events describe instantaneous actions or, put differently, interactions at a fine-grained temporal resolution (Borgatti et al., 2009).

However, in some contexts there arise problems regarding the reliability of event data. While data gathered from e.g. direct observations (Tranmer et al., 2015) or parliamentary records (Malang et al., 2019) should prove unproblematic in this regard, other data collection methods may be prone to *spurious events*, i.e. events that are recorded but did not actually occur as such. For instance, data collection on face-to-face interactions

¹Corresponding Autor: cornelius.fritz@stat.uni-muenchen.de

relies on different types of *sociometric badges* (Eagle and Pentland, 2006) for which a recent study reports a false-discovery rate of the event identification of around 20% when compared to video coded data (Elmer et al., 2019). Political event data on armed conflict, in contrast, are generally collected via automated or human coding of news and social media reporting (Kauffmann, 2020). Spurious events may arise in this context if reports of fighting are wrong, as may be the case for propaganda reasons or due to reporters' reliance on rumors, or when fighting took place between different belligerents than those named. Such issues are especially prevalent in machine-coded conflict data where both false-positive and false-discovery rates of over 60% have been reported (King and Lowe, 2003; Jäger, 2018). However, even human-coded data suffer from this problem (Dawkins, 2020; Weidmann, 2015).

This discussion suggests that specific types of event data can include unknown quantities of spurious events, which may influence the substantive results obtained from models such as the REM (Butts, 2008a) or the Dynamic Actor-Oriented Model (Stadtfeld et al., 2017; Stadtfeld, 2012). We thus propose a Relational Events Model with Spurious Events (REMSE) as a method that allows researchers to study relational events from potentially error-prone contexts or data collections methods. Moreover, this tool can assess whether spurious events are observed under a particular model specification and, more importantly, whether they influence the substantive results. The REMSE can thus serve as a straightforward robustness check in situations where the researcher, due to their substantive knowledge, suspects that there are spurious observations and wants to investigate whether they distort their empirical results.

We take a counting process point of view where some increments of the dyadic counting processes are *true* events, while others may be attributed to *spurious* events, i.e., exist due to measurement error. This decomposition results in two different intensities governing the two respective types of events. The *spurious* events are described by a *spurious-event* intensity that we specify independently of the *true-event* intensity of *true* events. We present the model under the assumption that the *spurious* events are purely random. Therefore, we can model the respective intensity solely as a constant term. However, more complex scenarios involving the specification of exogenous and endogenous covariates for the *spurious-event* intensity are also possible. In general, we are however primarily interested in studying what factors drive the intensity of *true* events. We model this intensity following Butts (2008a), but the methodology is extendable to other model types such as Stadtfeld et al. (2017); Vu et al. (2015); DuBois et al. (2013); Perry and Wolfe (2013) or Lerner et al. (2021).

This article is structured as follows: We begin in Section 2 by introducing our methodology. In particular, we lay out the general framework to study relational event data proposed by Butts (2008a) in Section 2.1 and introduce an extension to this framework, the REMSE, to correct for the presence of *spurious* events in the remainder of Section 2. Through a simulation study in Section 3, we investigate the performance of our proposed estimator when spurious events are correctly specified and when they are nonexistent. We then apply the proposed model in Section 4 to analyze fighting incidents in the Syrian civil war as well as social interaction data from a college campus. A discussion of possible implications and extensions for the analysis of events concludes the article in Section 5.

2 A Relational Events Model with Spurious Events

2.1 Modeling framework for relational events

We denote observed events in an event stream $\mathcal{E} = \{e_1, \dots, e_M\}$ of M elements. Each object $e \in \mathcal{E}$ consists of a tuple encoding the information of an event. In particular, we denote the two actors of an event by $a(e)$ and $b(e)$ and the time of the event with $t(e)$. For simplicity of notation, we omit the argument e for $a()$ and $b()$ when no ambiguity exists and write a_m for $a(e_m)$, b_m for $b(e_m)$, and t_m for $t(e_m) \forall m \in \{1, \dots, m\}$. Stemming from our application cases, we mainly focus on undirected events in this article; hence the events $e = (a, b, t)$ and $\tilde{e} = (b, a, t)$ are equivalent in our framework. Note however that the proposed method also generalizes to the directed case. We denote the set of actor-tuples between which events can possibly occur by \mathcal{R} , where, for simplicity, we assume that \mathcal{R} is time-constant.

Following Perry and Wolfe (2013) and Vu et al. (2011a), we assume that the events in \mathcal{E} are generated by an inhomogeneous matrix-valued counting process

$$\mathbf{N}(t) = (N_{ab}(t) | (a, b) \in \mathcal{R}), \quad (1)$$

which, in our case, is assumed to be a matrix-valued Poisson process (see Daley and Vere-Jones, 2008 for an introduction to stochastic processes). Without loss of generality, we assume that $\mathbf{N}(t)$ is observed during the temporal interval \mathcal{T} , starting at $t = 0$. The cells of (1) count how often all possible dyadic events have occurred between time 0 and t , hence $\mathbf{N}(t)$ can be conceived as a standard social network adjacency matrix with integer-valued cell entries (Butts, 2008b). For instance, $N_{ab}(t)$ indicates how often actors a and b have interacted in the time interval $[0, t]$. Therefore, observing event $e = (a, b, t)$ constitutes an increase in $N_{ab}(t)$ at time point t , i.e. $N_{ab}(t - h) + 1 = N_{ab}(t)$ for $h \rightarrow 0$. We denote with $\boldsymbol{\lambda}(t)$ the matrix-valued intensity of process $\mathbf{N}(t)$. Based on this intensity function we can characterize the instantaneous probability of a unit increase in a specific dimension of $\mathbf{N}(t)$ at time-point t (Daley and Vere-Jones, 2008). We parametrize $\boldsymbol{\lambda}(t)$ conditional on the history of the processes, $\mathcal{H}(t)$, which may also include additional exogenous covariates. Hence, $\mathcal{H}(t) = (\mathbf{N}(u), X(u) | u < t)$, where $X(t)$ is some covariate process to be specified later. Note that we opt for a rather general characterization of Poisson processes, including stochastic intensities that explicitly depend on previous events. We define the intensity function at the tie-level:

$$\lambda_{ab}(t | \mathcal{H}(t), \vartheta) = \begin{cases} \lambda_0(t, \boldsymbol{\alpha}) \exp\{\boldsymbol{\theta}^\top s_{ab}(\mathcal{H}(t))\}, & \text{if } (a, b) \in \mathcal{R} \\ 0, & \text{else} \end{cases} \quad (2)$$

where $\vartheta = (\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top = \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\theta})$ is defined with the help of a dyadic operator $\text{vec}(\cdot, \cdot)$ that stacks two vectors and $\lambda_0(t, \boldsymbol{\alpha})$ is the baseline intensity characterized by coefficients $\boldsymbol{\alpha}$, while the parameters $\boldsymbol{\theta}$ weight the statistics computed by $s_{ab}(\mathcal{H}(t))$, which is the function of sufficient statistics. Based on $s_{ab}(\mathcal{H}(t))$, we can formulate endogenous effects, which are calculated from $(N(u) | u < t)$, exogenous variables calculated from $(X(u) | u < t)$, or a combination of the two which results in complex dependencies between the observed events. Examples of endogenous effects for undirected events include degree-related statistics like the absolute difference of the degrees of actors a and b or hyperdyadic effects, e.g., investigating how triadic closure influences the observed events. In our first application case, exogenous factors include a dummy variable whether group a and b share an

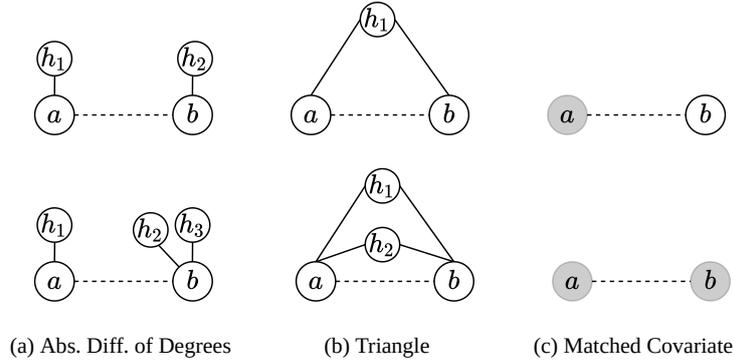


Figure 1: Graphical illustrations of endogenous and exogenous covariates. Solid lines represent past interactions, while dotted lines are possible but unrealized events. Node coloring indicates the node’s value on a categorical covariate. The relative risk of the events in the second row compared to the events in the first row is $\exp\{\theta_{end}\}$ if all other covariates are fixed, where θ_{end} is the coefficient of the respective statistic of each row.

ethno-religious identity. Alternatively, one may incorporate continuous covariates, e.g., computing the absolute geographic distance between group a and b . We give graphical representations of possible endogenous effects in Figure 1 and provide their mathematical formulations together with a general summary in Annex A. When comparing the structures in the first row with the ones in the second row in Figure 1, the respective sufficient statistic of the event indicated by the dotted line differs by one unit. Its intensity thus changes by the multiplicative factor $\exp\{\theta_{endo}\}$, where θ_{endo} is the respective parameter of the statistic if all other covariates are fixed. The interpretation of the coefficients is, therefore, closely related to the interpretation of relative risk models (Kalbfleisch and Prentice, 2002).

Previous studies propose multiple options to model the baseline intensity $\lambda_0(t)$. Vu et al. (2011a,b) follow a semiparametric approach akin to the proportional hazard model by Cox (1972), while Butts (2008a) assumes a constant baseline intensity. We follow Etezadi-Amoli and Ciampi (1987) by setting $\lambda_0(t, \alpha) = \exp\{f(t, \alpha)\}$, with $f(t, \alpha)$ being a smooth function in time parametrized by B-splines (de Boor, 2001):

$$f(t, \alpha) = \sum_{k=1}^K \alpha_k B_k(t) = \alpha^\top B(t), \quad (3)$$

where $B_k(t)$ denotes the k th B-spline basis function weighted by coefficient α_k . To ensure a smooth fit of $f(t, \alpha)$, we impose a penalty (or regularization) on α which is formulated through the priori structure

$$p(\alpha) \propto \exp\{-\gamma \alpha^\top \mathbf{S} \alpha\}, \quad (4)$$

where γ is a hyperparameter controlling the level of smoothing and \mathbf{S} is a penalty matrix that penalizes the differences of coefficients corresponding to adjacent basis functions as proposed by Eilers and Marx (1996). We ensure identifiability of the smooth baseline

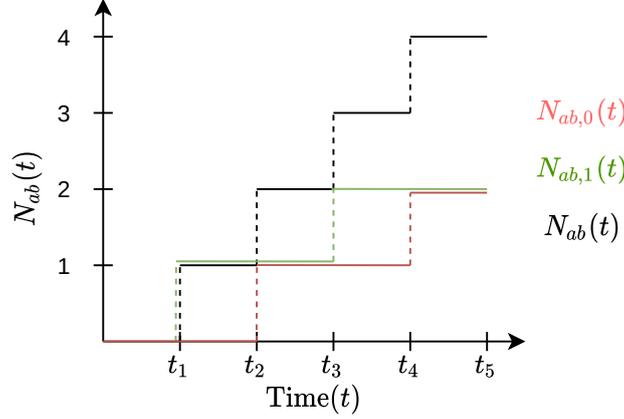


Figure 2: Graphical illustration of a possible path of the counting process of observed events ($N_{ab}(t)$) between actors a and b that encompasses spurious ($N_{ab,0}(t)$) and true events ($N_{ab,1}(t)$).

intensity by incorporating a sum-to-zero constraint and refer to Ruppert et al. (2003) and Wood (2017) for further details on penalized spline smoothing. Given this notation, we can simplify (2):

$$\lambda_{ab}(t|\mathcal{H}(t), \vartheta) = \begin{cases} \exp\{\vartheta^\top \mathcal{X}_{ab}(\mathcal{H}(t), t)\}, & \text{if } (a, b) \in \mathcal{R} \\ 0, & \text{else,} \end{cases} \quad (5)$$

with $\mathcal{X}_{ab}(\mathcal{H}(t), t) = \text{vec}(\mathbf{B}(t), s_{ab}(\mathcal{H}(t)))$.

2.2 Accounting for spurious relational events

Given the discussion in the introduction, we may conclude that some increments of $\mathbf{N}(t)$ are true events, while others stem from spurious events. Spurious events can occur because of coding errors during machine- or human-based data collection. To account for such erroneous data points, we introduce the Relational Events Model with Spurious Events (REMSE).

First, we decompose the observed Poisson process into two separate matrix-valued Poisson processes, i.e. $\mathbf{N}(t) = \mathbf{N}_0(t) + \mathbf{N}_1(t) \forall t \in \mathcal{T}$. On the dyadic level, $N_{ab,1}(t)$ denotes the number of true events between actors a and b until t , and $N_{ab,0}(t)$ the number of events that are spurious. Assuming that $N_{ab}(t)$ is a Poisson process, we can apply the so-called *thinning* property, stating that two separate processes that sum up to a Poisson process are also Poisson processes (Daley and Vere-Jones, 2008). A graphical illustration of the three introduced counting processes, $N_{ab,0}(t)$, $N_{ab,1}(t)$, and $N_{ab}(t)$, is given in Figure 2. In this illustrative example, we observe four events at times t_1 , t_2 , t_3 , and t_4 , although only the first and third constitute true events, while the second and fourth are spurious. Therefore, the counting process $N_{ab}(t)$ jumps at all times of an event, yet $N_{ab,1}(t)$ does so only at t_1 and t_3 . Conversely, $N_{ab,0}(t)$ increases at t_2 and t_4 .

The counting processes $\mathbf{N}_0(t)$ and $\mathbf{N}_1(t)$ are characterized by the dyadic intensities $\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0)$ and $\lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta_1)$, where we respectively denote the history of all spurious and true processes by $\mathcal{H}_0(t)$ and $\mathcal{H}_1(t)$. This can also be perceived as a competing risks setting, where events can either be caused by the *true-event* or *spurious-event* intensity (Gelfand et al., 2000). To make the estimation of θ_0 and θ_1 feasible and identifiable (Heckman and Honoré, 1989), we assume that both intensities are independent of one another, which means that their correlation is fully accounted for by the covariates. Building on the *superpositioning* property of Poisson processes, the specification of those two intensity functions also defines the intensity of the observed counting process $N_{ab}(t)$. In particular, $\lambda_{ab}(t|\mathcal{H}(t), \vartheta) = \lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0) + \lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta_1)$ holds (Daley and Vere-Jones, 2008).

The *true-event* intensity $\lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta_1)$ drives the counting process of true events $\mathbf{N}_1(t)$ and only depends on the history of true events. This assumption is reasonable since if erroneous events are mixed together with true events, the covariates computed for actors a and b at time t through $s_{ab}(\mathcal{H}(t))$ would be confounded and could not anymore be interpreted in any consistent manner. We specify $\lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta_1)$ in line with (2) at the dyadic level by:

$$\lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta) = \begin{cases} \exp\{\vartheta_1^\top \mathcal{X}_{ab,1}(\mathcal{H}_1(t), t)\}, & \text{if } (a, b) \in \mathcal{R} \\ 0, & \text{else.} \end{cases} \quad (6)$$

At the same time, the *spurious-event* intensity $\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0)$ determines the type of measurement error generating spurious events. One may consider the spurious-event process as an overall noise level with a constant intensity. This leads to the following setting:

$$\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0) = \begin{cases} \exp\{\alpha_0\}, & \text{if } (a, b) \in \mathcal{R} \\ 0, & \text{else.} \end{cases} \quad (7)$$

The error structure, that is, the intensity of the spurious-event process can be made more complex, but to ensure identifiability, $\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0)$ cannot depend on the same covariates as $\lambda_{ab,1}(t|\mathcal{H}_1(t), \vartheta)$. We return to the discussion of this point below and focus on model (7) for the moment.

2.3 Posterior inference via data augmentation

To draw inference on $\vartheta = \text{vec}(\vartheta_0, \vartheta_1)$, we employ an empirical Bayes approach. Specifically, we will sample from the posterior of ϑ given the *observed* data. Our approach is thereby comparable to the estimation of standard mixture (Diebolt and Robert, 1994) and latent competing risk models (Gelfand et al., 2000).

For our proposed method, the *observed* data is the event stream of all events \mathcal{E} regardless of being a real or a spurious event. To adequately estimate the model formulated in Section 2, we lack information on whether a given event is spurious or not. We denote this formally as a latent indicator variable $z(e)$ for event $e \in \mathcal{E}$:

$$z(e) = \begin{cases} 1, & \text{if event } e \text{ is a true event} \\ 0, & \text{if event } e \text{ is a spurious event} \end{cases}$$

We write $z = (z(e_1), \dots, z(e_M))$ to refer to the latent indicators of all events and use z_m to shorten $z(e_m)$. Given this notation, we can apply the data augmentation algorithm developed in Tanner and Wong (1987) to sample from the joint posterior distribution of (Z, ϑ) by iterating between the I Step (Imputation) and P Step (Posterior) defined as:

I Step: Draw $Z^{(d)}$ from the posterior $p(z|\vartheta^{(d-1)}, \mathcal{E})$;

P Step: Draw $\vartheta^{(d)}$ from the augmented $p(\vartheta|z^{(d)}, \mathcal{E})$.

This iterative scheme generates a sequence that (under mild conditions) converges to draws from the joint posterior of (ϑ, Z) and is a particular case of a Gibbs' sampler. Each iteration consists of an Imputation and a Posterior step, resembling the Expectation and Maximization step from the EM algorithm (Dempster et al., 1977). Note, however, that Tanner and Wong (1987) proposed this method with multiple imputations in each I Step and a mixture of all imputed *complete-data* posteriors in the P Step. We follow Little and Rubin (2002) and Diebolt and Robert (1994) by performing one draw of Z and ϑ in every iteration, which is a specific case of data augmentation. As Noghrehchi et al. (2021) argue, this approach is closely related to the stochastic EM algorithm (Celeux et al., 1996). The main difference between the two approaches is that in our P Step, the current parameters are sampled from the *complete-data* posterior in the data augmentation algorithm and not fixed at its mean as in Celeux et al. (1996). Consequently, the data augmentation algorithm is a *proper* multiple imputation procedure (MI, Rubin, 1987), while the stochastic EM algorithm is *improper* MI (see Noghrehchi et al., 2021). We choose the data augmentation algorithm over the stochastic EM algorithm because Rubin's combination rule to get approximate standard errors can only be applied to *proper* MI procedures (Noghrehchi et al., 2021).

In what follows, we give details and derivations on the I and P Steps and then exploit MI to combine a relatively small number of draws from the posterior to obtain point and interval estimates for ϑ .

Imputation-step: To acquire samples from $Z = (Z_1, \dots, Z_M)$ conditional on \mathcal{E} and ϑ , we first decompose the joint density by repeatedly applying the Bayes theorem:

$$\begin{aligned} p(z|\vartheta, \mathcal{E}) &= p(z_M, \dots, z_1|\vartheta, \mathcal{E}) \\ &= \prod_{m=1}^M p(z_m|z_1, \dots, z_{m-1}, \vartheta, \mathcal{E}). \end{aligned} \quad (8)$$

The distribution of z_m conditional on $z_1, \dots, z_{m-1}, \vartheta$ and \mathcal{E} is:

$$Z_m|z_1, \dots, z_{m-1}, \vartheta, \mathcal{E} \sim \text{Bin} \left(1, \frac{\lambda_{a_m b_m, 1}(t_m|\mathcal{H}_1(t_m), \vartheta_1)}{\lambda_{a_m b_m, 0}(t_m|\mathcal{H}_0(t_m), \vartheta_0) + \lambda_{a_m b_m, 1}(t_m|\mathcal{H}_1(t_m), \vartheta_1)} \right). \quad (9)$$

Note that the information of z_1, \dots, z_{m-1} and \mathcal{E} allows us to calculate $\mathcal{H}_1(t_m)$ as well as $\mathcal{H}_0(t_m)$. By iteratively applying (9) and plugging in $\vartheta^{(d)}$ for ϑ , we can draw samples in the I Step of $Z = (Z_1, \dots, Z_M)$ through a sequential design that sweeps once from Z_1 to Z_M . The mathematical derivation of (9) is provided in Annex B.

Posterior-step: As already stated, we assume that the *true-event* and *spurious-event* intensities are independent. Hence, the sampling from the *complete-data* posteriors of ϑ_0

and ϑ_1 can be carried out independently. In the ensuing section, we therefore only show how to sample from $\vartheta_1|z, \mathcal{E}$, but sampling from $\vartheta_0|z, \mathcal{E}$ is possible in the same manner. To derive this posterior, we begin by showing that the likelihood of \mathcal{E} and z with parameter ϑ_1 is the likelihood of the counting process $\mathbf{N}_1(t)$, which resembles a Poisson regression. Consecutively, we state all priors to derive the desired *complete-data* posterior.

Given a general z sampled in the previous I Step and \mathcal{E} , we reconstruct a unique complete path of $\mathbf{N}_1(t)$ by setting

$$N_{ab,1}(t) = \sum_{\substack{e \in \mathcal{E}; \\ z(e)=1, t(e) \leq t}} \mathbb{I}(a(e) = a, b(e) = b) \quad \forall (a, b) \in \mathcal{R}, t \in \mathcal{T}, \quad (10)$$

where $\mathbb{I}(\cdot)$ is an indicator function. The corresponding likelihood of $\mathbf{N}_1(t)$ results from the property that any element-wise increments of the counting process between any times s and t with $t > s$ and arbitrary actors a and b with $(a, b) \in \mathcal{R}$ are Poisson distributed:

$$N_{ab,1}(t) - N_{ab,1}(s) \sim \text{Pois} \left(\int_s^t \lambda_{ab,1}(u | \mathcal{H}_1(u), \vartheta_1) du \right). \quad (11)$$

The integral in (11) is approximated through simple rectangular approximation between the observed event times to keep the numerical effort feasible, so that the distributional assumption simplifies to:

$$Y_{ab,1}(t_m) = N_{ab,1}(t_m) - N_{ab,1}(t_{m-1}) \sim \text{Pois}((t_m - t_{m-1}) \lambda_{ab,1}(t_m | \mathcal{H}_1(t_m), \vartheta_1)) \quad (12) \\ \forall m \in \{1, \dots, M\} \text{ with } z_m = 1 \text{ and } (a, b) \in \mathcal{R}.$$

We specify the priors for α_1 and θ_1 separately and independent of one another. The prior for α_1 was already stated in (4). Through a restricted maximum likelihood approach, we estimate the corresponding hyperparameter γ_1 such that it maximizes the marginal likelihood of z and \mathcal{E} given γ_1 (for additional information on this estimation procedure and general empirical Bayes theory for penalized splines see Wood, 2011, 2020). Regarding the linear coefficients θ_1 , we assume flat priors, i.e. $p(\theta_1) \propto k$, indicating no prior knowledge.

In the last step, we apply Wood's (2006) result that for large samples, the posterior distribution of ϑ_1 under likelihoods resulting from distributions belonging to the exponential family, such as the Poisson distribution in (12), can be approximated through:

$$\vartheta_1|z, \mathcal{E} \sim N(\hat{\vartheta}_1, \mathbf{V}_1). \quad (13)$$

Here, $\hat{\vartheta}_1$ denotes the penalized maximum likelihood estimator resulting from (12) with the extended penalty matrix $\tilde{\mathbf{S}}_1$ defined by

$$\tilde{\mathbf{S}}_1 = \begin{bmatrix} \mathbf{S}_1 & \mathbf{O}_{p \times q} \\ \mathbf{O}_{p \times q} & \mathbf{O}_{q \times q} \end{bmatrix}$$

with $\mathbf{O}_{p \times q} \in \mathbb{R}^{p \times q}$ for $p, q \in \mathbb{N}$ being a matrix filled with zeroes and \mathbf{S}_1 defined in accordance with (4). For $\vartheta_1 = \text{vec}(\alpha_1, \theta_1)$, let p be the length of α_1 and q of θ_1 . The penalized likelihood is then given by:

$$\ell_p(\vartheta_1; z, \mathcal{E}) = \ell(\vartheta_1; z, \mathcal{E}) - \gamma_1 \vartheta_1^\top \tilde{\mathbf{S}}_1 \vartheta_1, \quad (14)$$

Result: $(\vartheta^{(1)}, z^{(1)}), \dots, (\vartheta^{(D)}, z^{(D)})$
Set: $\vartheta^{(0)}$ to be the posterior mean of the *true* and *spurious* events, which are sampled randomly from the observed events with equal probability
for $d \in \{1, \dots, D\}$ **do**
 Imputation Step: Sample $Z^{(d)} | \vartheta^{(d-1)}, \mathcal{E}$
 for $m \in \{1, \dots, M\}$ **do**
 • Sample $Z_m^{(d)} | z_1^{(d)}, \dots, z_{m-1}^{(d)}, \vartheta^{(d-1)}, \mathcal{E}$ according to (9)
 • If $z_m^{(d)} = 1$ update $s_{ab}(\mathcal{H}_1(t_m)) \forall (a, b) \in \mathcal{R}$
 end
 Posterior Step: Sample $\vartheta^{(d)} | z^{(d)}, \mathcal{E}$
 • Reconstruct $\mathbf{N}_0(t)$ and $\mathbf{N}_1(t) \forall t \in \mathcal{T}$ from $z^{(d)}$ and \mathcal{E} according to (10)
 • Obtain $\hat{\vartheta}_0$ and \mathbf{V}_0 by maximizing the penalized Poisson likelihood given in (12) (only for $\mathbf{N}_0(t)$ instead of $\mathbf{N}_1(t)$)
 • Sample $\vartheta_0^{(d)} | z^{(d)}, \mathcal{E} \sim N(\hat{\vartheta}_0, \mathbf{V}_0)$
 • Obtain $\hat{\vartheta}_1$ and \mathbf{V}_1 by maximizing the penalized Poisson likelihood given in (12)
 • Sample $\vartheta_1^{(d)} | z^{(d)}, \mathcal{E} \sim N(\hat{\vartheta}_1, \mathbf{V}_1)$
end

Algorithm 1: Pseudo-Code to obtain D samples from the data augmentation algorithm.

which is equivalent to a generalized additive model; hence we refer to Wood (2017) for a thorough treatment of the computational methods needed to find $\hat{\vartheta}_1$. The variance matrix in (13) has the following structure:

$$\mathbf{V}_1 = \left(\mathcal{X}_1^\top \mathbf{W}_1 \mathcal{X}_1 + \gamma_1 \tilde{\mathbf{S}}_1 \right)^{-1}.$$

Values for γ_1 and $\hat{\vartheta}_1$ can be extracted from the estimation procedure to maximize (14) with respect to ϑ_1 , while $\mathcal{X}_1 \in \mathbb{R}^{(M|\mathcal{R}|) \times (\rho+q)}$ is a matrix whose rows are given by $\mathcal{X}_{ab,1}(\mathcal{H}_1(t_m), t_{m-1})$ as defined in (6) for $m \in \{1, \dots, M\}$ and $(a, b) \in \mathcal{R}$. Similarly,

$\mathbf{W}_1 = \text{diag}(\lambda_{ab,1}(t | \mathcal{H}_1(t), \vartheta_1); t \in \{t_1, \dots, t_M\}, (a, b) \in \mathcal{R})$ is a diagonal matrix.

For the P Step, we now plug in $z^{(d)}$ for z in (13) to obtain $\hat{\vartheta}_1$ and \mathbf{V}_1 by carrying out the corresponding *complete-case* analysis. In the case where no spurious events exist, the complete estimation can be carried out in a single P Step. In Algorithm 1, we summarize how to generate a sequence of random variables according to the data augmentation algorithm.

Multiple imputation: One could use the data augmentation algorithm to get a large amount of samples from the joint posterior of (ϑ, Z) to calculate empirical percentiles for obtaining any types of interval estimates. However, in our case this endeavor would be very time-consuming and even infeasible. To circumvent this, Rubin (1976) proposed multiple imputation as a method to approximate the posterior mean and variance. Coincidentally, the method is especially successful when the *complete-data* posterior is multivariate normal as is the case in (13), thus only a small number of draws is needed to obtain good approximations (Little and Rubin, 2002). To be specific, we apply the law of iterative

expectation and variance:

$$\mathbb{E}(\boldsymbol{\vartheta}|\mathcal{E}) = \mathbb{E}(\mathbb{E}(\boldsymbol{\vartheta}|\mathcal{E}, z)|z) \quad (15)$$

$$\text{Var}(\boldsymbol{\vartheta}|\mathcal{E}) = \mathbb{E}(\text{Var}(\boldsymbol{\vartheta}|\mathcal{E}, z)|z) + \text{Var}(\mathbb{E}(\boldsymbol{\vartheta}|\mathcal{E}, z)|z). \quad (16)$$

Next, we approximate (15) and (16) using a Monte Carlo quadrature with K samples from the posterior obtained via the data augmentation scheme summarized in Algorithm 1 after a burn-in period of D iterations:

$$\mathbb{E}(\boldsymbol{\vartheta}|\mathcal{E}_{\text{obs}}) \approx \frac{1}{K} \sum_{k=D+1}^{D+K} \hat{\boldsymbol{\vartheta}}^{(k)} = \bar{\boldsymbol{\vartheta}} \quad (17)$$

$$\begin{aligned} \text{Var}(\boldsymbol{\vartheta}|\mathcal{E}_{\text{obs}}) &\approx \frac{1}{K} \sum_{k=D+1}^{D+K} \mathbf{V}^{(k)} + \frac{K+1}{K(K-1)} \sum_{k=D+1}^{D+K} (\hat{\boldsymbol{\vartheta}}^{(k)} - \bar{\boldsymbol{\vartheta}}) (\hat{\boldsymbol{\vartheta}}^{(k)} - \bar{\boldsymbol{\vartheta}})^\top \\ &= \bar{\mathbf{V}} + \bar{\mathbf{B}}, \end{aligned} \quad (18)$$

where $\hat{\boldsymbol{\vartheta}}^{(k)} = \text{vec}(\hat{\boldsymbol{\vartheta}}_0^{(k)}, \hat{\boldsymbol{\vartheta}}_1^{(k)})$ encompasses the *complete-data* posterior means from the k th sample and $\mathbf{V}^{(k)} = \text{diag}(\mathbf{V}_0^{(k)}, \mathbf{V}_1^{(k)})$ is composed of the corresponding variances defined in (13). We can thus construct point and interval estimates from relatively few draws of the posterior based on a multivariate normal reference distribution (Little and Rubin, 2002).

3 Simulation Study

We conduct a simulation study to explore the performance of the REMSE compared to a REM, which assumes no spurious events, in two different scenarios, including a regime where measurement error is correctly specified in the REMSE and one where spurious events are instead non-existent.

Simulation design: In $S = 1000$ runs, we simulate event data between $n = 40$ actors under known true and spurious intensity functions in each example. For exogenous covariates, we generate categorical and continuous actor-specific covariates, transformed to the dyad level by checking for equivalence in the categorical case and computing the absolute difference for the continuous information. Generally, we simulate both counting processes $\mathbf{N}_1(\mathbf{t})$ and $\mathbf{N}_0(\mathbf{t})$ separately and stop once $|\mathcal{E}_1| = 500$.

The data generating processes for *true* events is identical in each case and given by:

$$\begin{aligned} \lambda_{ab,1}(t|\mathcal{H}_1(t), \boldsymbol{\vartheta}_1) &= \exp\{-5 + 0.2 \cdot s_{ab, Degree\ Abs.}(\mathcal{H}_1(t)) \\ &\quad + 0.1 \cdot s_{ab, Triangle}(\mathcal{H}_1(t)) - 0.5 \cdot s_{ab, Repetition\ Count}(\mathcal{H}_1(t)) \\ &\quad + 2 \cdot s_{ab, Sum\ cont.}(\mathcal{H}_1(t)) - 2 \cdot s_{ab, Match\ cat.}(\mathcal{H}_1(t))\}, \end{aligned} \quad (\text{DG 1-2})$$

where we draw the continuous exogenous covariate (cont.) from a standard Gaussian distribution and the categorical exogenous covariates (cat.) from a categorical random variable with seven possible outcomes, all with the same probability. Mathematical definition of the endogenous and exogenous statistics are given in Annex A. In contrast, the *spurious-event* intensity differs across regimes to result in correctly specified (DG 1) and

11. All that Glitters is not Gold: Modeling Relational Events with Spurious Events

Table 1: Result of the simulation study for the REMSE and REM with the two data-generating processes (DG 1, DG 2). For each DG and covariate, we note the AVE (Average Estimate), RMSE (Root-Mean-Squared Error), and CP (Coverage Probability). We report the average Percentage of False Events (PFE) for each DG in the last row.

	Coefs.	REMSE			REM		
		AVE	RMSE	CP	AVE	RMSE	CP
DG 1 (PFE: 4.819 %)							
Intercept	-5.0	-4.936	0.337	0.944	-3.510	1.523	0.003
Degree abs	0.2	0.198	0.009	0.940	0.168	0.033	0.018
Triangle	0.1	0.101	0.019	0.949	0.094	0.019	0.932
Repetition	-0.5	-0.494	0.035	0.946	-0.385	0.120	0.039
Cov. cont.	2.0	1.982	0.101	0.951	1.557	0.453	0.003
Cov. cat.	-2.0	-1.986	0.246	0.952	-1.594	0.461	0.515
\widehat{PFE} (in %)			4.835				
DG 2 (PFE 0 %)							
Intercept	-5.0	-5.040	0.286	0.954	-5.027	0.281	0.955
Degree abs	0.2	0.201	0.008	0.958	0.201	0.008	0.956
Triangle	0.1	0.102	0.018	0.955	0.102	0.018	0.952
Repetition	-0.5	-0.505	0.030	0.969	-0.504	0.030	0.964
Cov. cont.	2.0	2.009	0.087	0.952	2.006	0.086	0.948
Cov. cat.	-2.0	-2.007	0.231	0.952	-2.004	0.230	0.952
\widehat{PFE} (in %)			0.0001				

nonexistent (DG 2) measurement errors:

$$\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0) = \exp\{-2.5\} \quad (\text{DG 1})$$

$$\lambda_{ab,0}(t|\mathcal{H}_0(t), \vartheta_0) = 0 \quad (\text{DG 2})$$

Given these intensities, we follow DuBois et al. (2013) to sample the events.

Although the method is estimated in a Bayesian framework, we can still assess the frequentist properties of the estimates of the REMSE and REM. In particular, the average point estimate (AVE), the root-mean-squared error (RMSE) and the coverage probabilities (CP) are presented in Table 1. The AVE of a specific coefficient is the average over the posterior modes in each run:

$$\text{AVE} = \frac{1}{S} \sum_{s=1}^S \bar{\vartheta}_s,$$

where $\bar{\vartheta}_t$ is the posterior mean (17) of the t th simulation run. To check for the average variance of the error in each run, we further report the RMSEs of estimating the coefficient

vector ϑ :

$$\text{RMSE} = \sqrt{\frac{1}{S} \sum_{s=1}^S (\bar{\vartheta}_s - \vartheta)^\top (\bar{\vartheta}_s - \vartheta)},$$

where ϑ is the ground truth coefficient vector defined above. Finally, we assess the adequacy of the uncertainty quantification by computing the percentage of runs in which the real parameter lies within the confidence intervals based on a multivariate normal posterior with mean and variance given in (17) and (18). According to standard statistical theory for interval estimates, this coverage probability should be around 95% (Casella and Berger, 2001).

Results: DG 1 shows how the estimators behave if the true and false intensities are correctly specified. The results in Table 1 suggest that the REMSE can recover the coefficients from the simulation. On the other hand, strongly biased estimates are obtained in the REM, where not only the average estimates are biased, but we also observe high RMSEs and violated coverage probabilities.

In the second simulation, we assess the performance of the spurious event model when it is misspecified. In particular, we investigate what happens when there are no spurious events in the data, i.e., all events are real, and the intensity of $N_{ab,2}(t)$ is zero in DG 2. Unsurprisingly, the REM allows for valid and unbiased inference under this regime. But our stochastic estimation algorithm proves to be robust as for most runs, the simulated events were at some point only consisting of true events. In other words, the REMSE can detect the spurious events correctly and is unbiased if none occur in the observed data.

For both DG 1 and DG 2, the PFE estimated by the REMSE closely matches the observed one whereas the REM, by constraining it to zero, severely underestimates the PFE in DG 1. In sum, the simulation study thus offers evidence that the REMSE increases our ability to model relational event data in the presence of measurement error while being equivalent to a standard REM when spurious events do not exist in the data.

4 Application

Next we apply the REMSE on two real-world data sets motivated by the types of event data discussed in the introduction, namely human-coded conflict events in the Syrian civil war and co-location event data generated from the Bluetooth devices of students in a university dorm. Information on the data sources, observational periods and numbers of actors and events is summarized in Table 2. Following the above presentation, we focus on modeling the *true-event* intensity of the REMSE and limit the *spurious-event* intensity to the constant term. Covariates are thus only specified for the *true-event* intensity. In our applications, the samples drawn according to Algorithm 1 converged to a stationary distribution within the first 30 iterations. To obtain the reported point and interval estimates via MI, we sampled 30 additional draws. Due to space restrictions, we keep our

²We include all actors that, within the two-year period, participated in at least five events. To verify their existence and obtain relevant covariates, we compared them first to data collected by Gade et al. (2019) and then to various sources including news media reporting. We omitted two actors on which we could not find any information as well as actor aggregations such as “rioters” or “syrian rebels”.

³To capture new events instead of repeated observations of the same event, we omit events where the most recent previous interaction between a and b occurred less than 20 minutes before.

Table 2: Descriptive information on the two analyzed data sets.

	Conflict Event Data in 4.1	Co-location Events in 4.2
Source	ACLED (Raleigh et al., 2010)	MIT Human Dynamics Lab (Madan et al., 2012)
Observational Period	2017:01:01 - 2019:01:01	2008:11:01 - 2008:11:04
Number of Actors	68 ²	58
Number of Events	4,362	2,489 ³

discussions of the substantive background and results of both applications comparatively short.

4.1 Conflict events in the Syrian civil war

In the first example, we model conflict events between different belligerents as driven by both exogenous covariates and endogenous network mechanisms. The exogenous covariates are selected based on the literature on inter-rebel conflict. We thus include dummy variables indicating whether two actors share a common ethno-religious identity or receive material support by the same external sponsor as these factors have previously been found to reduce the risk of conflict (Popovic, 2018; Gade et al., 2019). Additionally, we include binary indicators of two actors being both state forces or both rebel groups as conflict may be less likely in the former but more likely in the latter case (Dorff et al., 2020).

Furthermore, we model endogenous processes in the formation of the conflict event network and consider four statistics for this purpose. First, we account for repeated fighting between two actors by including both the count of their previous interactions as well as a binary indicator of repetition, which takes the value 1 if that count is at least 1. We use this additional endogenous covariate as a conflict onset arguably comprises much more information than subsequent fighting. Second, we include the absolute difference in a and b 's degree to capture whether actors with a high extent of previous activity are prone to engage each other or, instead, tend to fight less established groups to pre-empt their rise to power. Finally, we model hyper-dyadic dependencies by including a triangle statistic that captures the combat network's tendency towards triadic closure.

Given that fighting should be a relatively obvious event, one may wonder why conflict event data may include spurious observations. This is because all common data collection efforts on armed conflict cannot rely on direct observation but instead use news and social media reporting. Spurious events thus occur when these sources report fighting which did not actually take place as such. In armed conflict, this can happen for multiple reasons. For instance, pro-government media may falsely report that state security forces engaged with and defeated rebel combatants to boost morale and convince audiences that the government is winning. Social media channels aligned with a specific rebel faction may similarly claim victories by its own forces or, less obviously, battles where a rival faction fought and suffered defeat against another group. In war-time settings, journalists may also be unable or unwilling to enter conflict areas and thus base their reporting on local contacts, rumors, or hear-say. Finally, spurious observations may arise here when reported fighting occurred but was attributed to the wrong belligerent faction at some point in the data collection process. From a substantive perspective, it is thus advisable to check for the influence of spurious events when analysing these data.

Table 3: Combat events in the Syrian civil war: Estimated coefficients with confidence intervals noted in brackets in the first column, while the Z values are given in the second column. The results of the REMSE are given in the first two columns, while the coefficients of the REM are depicted in the last two columns. The last row reports the estimated average Percentage of False Events (PFE).

	REMSE		REM	
	Coef./CI	Z Val.	Coef./CI	Z Val.
Intercept	-10.047 [-10.24,-9.854]	-102.124	-9.944 [-10.112,-9.775]	-115.723
Degree Abs	0.03 [0.026,0.035]	12.677	0.03 [0.027,0.034]	17.295
Repetition Count	0.009 [0.009,0.01]	45.28	0.009 [0.009,0.01]	56.342
First Repetition	5.052 [4.87,5.235]	54.321	4.911 [4.763,5.059]	64.946
Triangle	0.074 [0.06,0.089]	10.229	0.073 [0.065,0.08]	18.989
Match Ethno-Religious Id.	-0.387 [-0.532,-0.242]	-5.225	-0.393 [-0.525,-0.262]	-5.852
Match Rebel	0.159 [0.064,0.255]	3.27	0.171 [0.094,0.247]	4.381
Match State Force	-0.087 [-0.323,0.149]	-0.721	-0.077 [-0.287,0.132]	-0.723
Common Sponsor 1	-0.218 [-0.388,-0.048]	-2.51	-0.227 [-0.378,-0.077]	-2.957
\widehat{PFE} (in %)	1.1		0	

Table 3 accordingly presents the results of an REM and the REMSE. Beginning with the exogenous covariates, belligerents are found to be less likely to fight each other when they share an ethno-religious identity or receive resources from the same external sponsor. In contrast, there is no support for the idea that state forces exhibit less fighting among each other than against rebels in this type of internationalized civil war, whereas different rebel groups are more likely to engage in combat against one another. Furthermore, we find evidence that endogenous processes affect conflict event incidence. The binary repetition indicator exhibits the strongest effect across all covariates, implying that two actors are more likely to fight each other if they have done so in the past. As indicated by the positive coefficient of the repetition count, the dyadic intensity further increases the more they have previously fought with one another. The absolute degree difference also exhibits a positive effect, meaning that fighting is more likely between groups with different levels of previous activity. And finally, the triangle statistic's positive coefficient suggests that even in a fighting network, triadic closure exists. This may suggest that belligerents engage in multilateral conflict, attacking the enemy of their enemy, in order to preserve the existing balance of capabilities or change it in their favor (Pischedda, 2018).

This discussion holds for both the results of REM and REMSE. Their point estimates are generally quite similar in this application, suggesting that spurious events do not substantively affect empirical results in this case. That being said, there are two noticeable differences between the two models. First, the coefficient estimates for the binary indicator of belligerents having fought before differs between the two models. In the REM, it implies a multiplicative change of $\exp\{4.911\} = 135.775$ while for the REMSE, it is estimated at $\exp\{5.059\} = 157.433$. While both models thus identify this effect to be positive and significant, it is found to be substantively stronger when spurious events are accounted for. Second, the two models differ in how precise they deem estimates to be. This difference is clearest in their respective Z-values, which are always farther away from zero for the REM than the REMSE. As a whole, these results nonetheless show that spurious events have an overall small influence on substantive results in this application. The samples from the latent indicators z also indicate that only approximately 1% of the observations, about 50 events, are on average classified as spurious events. These findings offer reassurance for the increasing use of event data to study armed conflict.

4.2 Co-location events in university housing

In our second application, we use a subset of the co-location data collected by Madan et al. (2012) to model when students within an American university dorm interact with each other. These interactions are deduced from continuous (every 6 minutes) scans of proximity via the Bluetooth signals of students' mobile phones. Madan et al. (2012) used questionnaires to collect a host of information from the participating students. This information allows us to account for both structural and more personal exogenous predictors of social interaction. We thus include binary indicators of whether two students are in the same year of college or live on the same floor of the dorm to account for the expected homophily of social interactions (McPherson et al., 2001). In addition, we incorporate whether two actors consider each other close friends⁴. Given that the data were collected around a highly salient political event, the 2008 US presidential election, we also incorporate a dummy variable to measure whether they share the same presidential preference

⁴We symmetrized the friendship network, i.e., if student a nominated student b as a close friend, we assume that the relationship is reciprocated.

Table 4: Co-location Events in University Housing: Estimated coefficients with confidence intervals noted in brackets in the first column, while the Z values are given in the second column. The results of the REMSE are given in the first two columns, while the coefficients of the REM are depicted in the last two columns. The last row reports the estimated average Percentage of False Events (PFE).

	REMSE		REM	
	Coef./CI	Z Val.	Coef./CI	Z Val.
Intercept	-10.077 [-10.235,-9.919]	-124.905	-10.012 [-10.153,-9.871]	-139.269
Degree Abs	0.025 [0.016,0.035]	5.361	0.025 [0.017,0.032]	6.369
Repetition Count	0.066 [0.061,0.07]	27.263	0.065 [0.061,0.069]	29.988
First Repetition	2.714 [2.587,2.84]	42.024	2.615 [2.501,2.73]	44.704
Triangle	0.049 [0.035,0.064]	6.597	0.049 [0.037,0.061]	8.109
Match Floor	0.117 [0.013,0.221]	2.197	0.123 [0.024,0.222]	2.439
Match Presidential Pref	0.195 [0.108,0.282]	4.374	0.188 [0.106,0.27]	4.499
Match Year	-0.003 [-0.109,0.104]	-0.051	-0.012 [-0.112,0.088]	-0.236
Dyad. Friendship	0.157 [0.059,0.254]	3.145	0.15 [0.057,0.243]	3.15
Sim. Interested In Politics	-0.018 [-0.064,0.029]	-0.74	-0.021 [-0.065,0.024]	-0.917
\widehat{PFE} (in %)	3.264		0	

and a variable measuring their similarity in terms of interest in politics (Butters and Hare, 2020). In addition, we include the same endogenous network statistics here as in section 4.1. These covariates allow us to capture the intuitions that individuals tend to socialize with people that they have interacted with before, are not equally popular as they are, and they share more common friends with (Rivera et al., 2010). Compared to the first application, sources of spurious events here are more evident as students may not actually interact with but be physically close to and even face each other, e.g., riding an elevator, queuing in a store, or studying in a common space.

We present the results in Table 4. Beginning with the exogenous covariates, we find that the observed interactions tend to be homophilous in that students have social encounters with people they live together with, consider their friends, and share a political opinion with. In contrast, neither a common year of college nor a similar level of political

interest are found to have a statistically significant effect on student interactions. At the same time, these results indicate that the social encounters are affected by endogenous processes. Having already had a previous true event is found to be the main driver of the corresponding intensity; hence having a very strong and positive effect. Individuals who have socialized before are thus more likely to socialize again, an effect that, as indicated by the repetition count, increases with the number of previous interactions. Turning to the other endogenous covariates, the result for absolute degree difference suggests that students a and b are more likely to engage with each other if they have more different levels of previous activity, suggesting that e.g. popular individuals attract attention from less popular ones. As is usual for most social networks (Newman and Park, 2003), the triangle statistic is positive, meaning that students “socialize” with the friends of their friends.

As in the first application, the REM and REMSE results presented in Table 4 are closely comparable but also show some differences. Again, the effect estimate for binary repetition, at $\exp\{2.715\} = 15.105$, is higher in the REMSE than in the REM ($\exp\{2.615\} = 13.667$) while Z-values and confidence intervals obtained in the REM are substantially smaller in the REM than in the REMSE. In the co-location data too, the results are thus not driven by the presence of spurious events but accounting for these observations does affect results to some, albeit rather negligible, extent. This is the case even though the average percentage of spurious events here is comparatively high at 3%. That leaving out the corresponding 81 events yielded similar estimates may indicate that spurious events were mainly observed at the periphery of the interaction network and hardly affected the behavior in the network’s core. More generally, these results may assuage concerns over sensor data reliability (see Elmer et al., 2019).

5 Discussion

In summary, this paper extends the relational event framework to handle spurious events. In doing so, it offers applied researchers analyzing instantaneous interaction data a useful tool to explicitly account for measurement errors induced by spurious events or to investigate the robustness of their results against this type of error. Our proposed method controls for one explicit measurement error, namely that induced by spurious events. The simulation study showed that our approach can detect such false events and even yield correct results if they are not present. Still, we want to accentuate that numerous other types of measurement error may be present when one analyses relational events, which we disregard in this article. For instance, true events may be missing. These false negatives, e.g., unreported conflict events between different belligerents, are difficult to tackle because of a lack of information.

We explicitly recommend the use of the REMSE as a method for checking robustness. When substantive knowledge suggests the presence of spurious events, the REMSE can be used to assess whether REM results hold when accounting for them. Spurious events may be common in datasets which come from sensors or are coded from journalistic sources, as discussed above, and more generally seem credibly present in data that are based on secondary sources instead of direct observation. Spurious events also occur, and may possibly be more influential, in data where relations are directed, the model we introduce accordingly also generalizes to directed event data. Especially for politically contentious data, where some events may be openly claimed to be false, the REMSE offers a possibility to adjudicate whether overall findings depend on such contested observations. But also

where the data content is non-political, it is recommendable to check how common and influential false observations are. We provide replication code implementing the REMSE for this purpose.

When specifying the REMSE, two aspects require caution so that identifiability is ensured. First, given we know which events are spurious, our model simplifies to a competing risk model; thus, the identifiability issues discussed in Heckman and Honoré (1989) or Tsiatis (1975) apply. For this reason, we presented our model under the assumption of independence between the *true-event* and *spurious-event* intensities. Second, the particular specification of the covariates might also affect the identifiability of the model. This may occur when one assumes complex dependencies of spurious events and exogenous covariates are unavailable, or the prior information about the coefficients is too weak. For the model specification employed in this article, this is not an issue due to the simple form of the *spurious-event* intensity as long as at least one exogenous or endogenous term has a nonzero effect on the *true-event* intensity. For more complex models, one may use multiple starting values of the data augmentation algorithm or formulate more informative priors for θ_1 and possibly θ_0 .

Our latent variable methodology can also be extended beyond the approach presented here. A straightforward refinement along the lines of Stadtfeld and Block (2017) would be to include *windowed* effects, i.e., endogenous statistics that are only using history ranging into the past for a specific duration, or exogenous covariates calculated from additional networks to the one modeled. The first modification could also be extended to separable models as proposed in Fritz et al. (2021). A relatively simplistic version of the latter type of covariate was incorporated in Section 4.2 to account for common friendships but more complex covariates are possible. This might be helpful, for instance, when we observe proximity and e-mail events between the same group of actors. Moreover, with minor adaptations, the proposed estimation methodology could handle some of the exogenous or endogenous covariates having nonlinear effects on the intensities.

Finally, the framing of the simultaneous counting processes may be modified and their number extended. To better understand the opportunities our model framework entails, it is instructive to perceive the proposed model as an extension to the latent competing risk model of Gelfand et al. (2000) with two competing risks. For time-to-event data, one could thus employ an egocentric version⁵ of our model for model-based clustering of general duration times, which could prove to be a valuable tool for medical applications. Or our proposed methodology could be conceived as a general tool to correct for additive measurement errors in count data and extend it to spatial data analysis to be used in settings described in (Raleigh et al., 2010).

A Definition of undirected network statistics

As REMs for undirected events are so far sparse in the literature, there are no standard statistics that are commonly used (one exception being Bauer et al., 2021). Thus we define all statistics based on prior substantive research (Rivera et al., 2010; Wasserman and Faust, 1994) and undirected statistics used for modeling static networks (Robins et al., 2007). Generally, nondirected statistics have to be invariant to swapping the positions of actor a and b . For the following mathematical definitions, we denote the set of all actors by \mathcal{A} .

⁵See Vu et al., 2011b for further information on egocentric models.

For degree-related statistics, we include the absolute difference of the degrees of actors a and b :

$$S_{ab, Degree\ Abs.}(\mathcal{H}(t)) = \left| \sum_{h \in \mathcal{A}} (\mathbb{I}(N_{ah}(t^-) > 0) + \mathbb{I}(N_{ha} > 0)(t^-)) - \sum_{h \in \mathcal{A}} (\mathbb{I}(N_{bh}(t^-) > 0) + \mathbb{I}(N_{hb}(t^-) > 0)) \right|,$$

where t^- is the point-in-time just before t . Alternatively, one might also employ other bivariate functions of the degrees as long as they are invariant to swapping a and b , such as the sum of degrees. When simultaneously using different forms of degree-related statistics, collinearities between the respective covariates might severely impede the interpretation.

To capture past dyadic behavior, one can include $N_{ah}(t^-)$ directly as a covariate. Since the first event often constitutes a more meaningful action than any further observed events between the actors a and b , we additionally include a binary covariate to indicate whether the respective actors ever interacted before, leading to the following endogenous statistics:

$$S_{ab, Repition\ Count}(\mathcal{H}(t)) = N_{ah}(t^-)$$

$$S_{ab, First\ Repition}(\mathcal{H}(t)) = \mathbb{I}(N_{ab}(t^-) > 0).$$

Hyperdyadic statistics in the undirected regime are defined as any type of triadic closure, where actor a is connected to an entity that is also connected to actor b :

$$S_{ab, Triangle} = \sum_{h \in \mathcal{A}} \mathbb{I}(N_{ah}(t^-) > 0) \mathbb{I}(N_{bh}(t^-) > 0) + \mathbb{I}(N_{ha}(t^-) > 0) \mathbb{I}(N_{bh}(t^-) > 0) + \mathbb{I}(N_{ah}(t^-) > 0) \mathbb{I}(N_{hb}(t^-) > 0) + \mathbb{I}(N_{ha}(t^-) > 0) \mathbb{I}(N_{hb}(t^-) > 0)$$

Finally, actor-specific exogenous statistics can also be used to model the intensities introduced in this article. We denote arbitrary continuous covariates by $x_{a,cont} \forall a \in \mathcal{A}$. On the one hand, we may include a measure for the similarity or dissimilarity for the covariate through:

$$S_{ab, Sim. cont} = |x_{a,cont} - x_{b,cont}|$$

$$S_{ab, Dissim. cont} = \frac{1}{|x_{a,cont} - x_{b,cont}|}.$$

For multivariate covariates, such as location, we only need to substitute the absolute value for any given metric, e.g., euclidean distance. In other cases, it might be expected that high levels of a continuous covariable result in higher or lower intensities of an event:

$$S_{ab, Sum cont} = x_{a,cont} + x_{b,cont}$$

Which type of statistic should be used depends on the application case and the hypotheses to be tested. Categorical covariates, that we denote by $x_{a,cat} \forall a \in \mathcal{A}$, can also be used to parametrize the intensity by checking for equivalence of two actor-specific observations of

the variable:

$$S_{ab,Match\ cat} = \mathbb{I}(x_{a,cat} = x_{b,cat})$$

Besides actor-specific covariates also exogenous networks or matrices, such as $x_{Network} \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$, can also be incorporated as dyadic covariates in our framework:

$$S_{ab,Dyad. Network} = \frac{x_{Network,ab} + x_{Network,ba}}{2},$$

where $x_{Network,ab}$ is the entry of the a th row and b th column of the matrix $x_{Network}$. Extensions to time-varying networks are straightforward when perceiving changes to them as exogenous to the modeled events (Stadtfeld and Block, 2017).

B Mathematical derivation of (9)

For $m \in \{1, \dots, M\}$, let $Y_{abm,1}(t_m) = N_{abm,1}(t_m) - N_{abm,1}(t_{m-1})$ be the increments of the latent counting process of true events between the time points t_m and t_{m-1} , where we additionally define $t_0 = 0$ without the loss of generality. We observe \mathcal{E} , hence we can reconstruct the respective increment $Y_{abm}(t_m) = N_{abm}(t_m) - N_{abm}(t_{m-1}) = Y_{abm,0}(t_m) + Y_{abm,1}(t_m)$, where $Y_{abm,0}(t_m)$ is the increment of the spurious-event counting process. The second equality holds since by design the sum of increments of the processes counting the true and false events is the increment of the observed counting process, i.e. $N_{ab}(t) = N_{ab,0}(t) + N_{ab,1}(t)$. To sample from $Z_m | z_1, \dots, z_{m-1}, \mathcal{E}$, note that $Z_m = Y_{abm,1}(t_m) | Y_{abm}(t_m)$ holds. Heuristically, this means that if we know that one of the two *thinned* counting processes jumps at time t_m , the probability of the jump being attributed to $N_{abm,1}(t)$ is the probability that the m th event is a true event. For the increments of the involved counting processes, we can then use the properties of the Poisson processes and the fact that the intensities are piecewise constant between event times to derive the following distributional assumptions $\forall m = 1, \dots, M$:

$$Y_{abm,0}(t_m) | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta \sim \text{Pois}(\delta_m \lambda_{abm,0}(t_m | \mathcal{H}_0(t_m), \vartheta_0)) \quad (19)$$

$$Y_{abm,1}(t_m) | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta \sim \text{Pois}(\delta_m \lambda_{abm,1}(t_m | \mathcal{H}_1(t_m), \vartheta_1)) \quad (20)$$

$$Y_{abm}(t_m) | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta \sim \text{Pois} \left(\delta_m \left(\lambda_{abm,0}(t_m | \mathcal{H}_0(t_m), \vartheta_0) + \lambda_{abm,1}(t_m | \mathcal{H}_1(t_m), \vartheta_1) \right) \right), \quad (21)$$

where we set $\delta_m = t_m - t_{m-1}$. We can now directly compute the probability of $Z_m = 1|z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta$:

$$\begin{aligned}
 p(z_m = 1|z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta) &= p(Y_{a_m b_m, 1}(t_m) = 1 | Y_{a_m b_m}(t_m) = 1, z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta) \\
 &= \frac{p(Y_{a_m b_m, 1}(t_m) = 1, Y_{a_m b_m}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)}{p(Y_{a_m b_m}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)} \\
 &= \frac{p(Y_{a_m b_m, 1}(t_m) = 1, Y_{a_m b_m, 2}(t_m) = 0 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)}{p(Y_{a_m b_m}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)} \\
 &= \frac{p(Y_{a_m b_m, 1}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)}{p(Y_{a_m b_m}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)} \\
 &\quad \times \frac{p(Y_{a_m b_m, 2}(t_m) = 0 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)}{p(Y_{a_m b_m}(t_m) = 1 | z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)} \\
 &= \frac{\lambda_{a_m b_m, 1}(t_m | \mathcal{H}_1(t_m, \vartheta_1))}{\lambda_{a_m b_m, 0}(t_m | \mathcal{H}_0(t_m), \vartheta_0) + \lambda_{a_m b_m, 1}(t_m | \mathcal{H}_1(t_m), \vartheta_1)}
 \end{aligned}$$

In the last row we plug in (19) and (20) for the probabilities in the numerators and (21) in the denominator to prove claim (9). The calculation for $p(z_m = 0|z_1, \dots, z_{m-1}, \mathcal{E}, \vartheta)$ is almost identical to the one shown here.

References

- Bauer, V., D. Harhoff, and G. Kauermann (2021). A smooth dynamic network model for patent collaboration data. *AStA Advances in Statistical Analysis*.
- Blair, R. A. and N. Sambanis (2020). Forecasting civil wars: Theory and structure in an age of “big data” and machine learning. *Journal of Conflict Resolution* 64(10), 1885–1915.
- Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca (2009). Network analysis in the social sciences. *Science* 323(5916), 892–895.
- Butters, R. and C. Hare (2020). Polarized networks? new evidence on american voters’ political discussion networks. *Political Behavior*.
- Butts, C. T. (2008a). A relational event framework for social action. *Sociological Methodology* 38(1), 155–200.
- Butts, C. T. (2008b). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology* 11(1), 13–41.
- Casella, G. and R. Berger (2001). *Statistical Inference*. Belmont: Brooks/Cole.
- Celeux, G., D. Chauveau, and J. Diebolt (1996). Stochastic versions of the EM algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation* 55(4), 287–314.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2), 187–202.
- Daley, D. J. and D. Vere-Jones (2008). *An introduction to the theory of point processes: Volume I: Elementary Theory and Methods*. New York: Springer.
- Dawkins, S. (2020). The problem of the missing dead. *Journal of Peace Research* (OnlineFirst).
- de Boor, C. (2001). *A practical guide to splines*. New York: Springer.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39(1), 1–22.
- Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)* 56(2), 363–375.
- Dorff, C., M. Gallop, and S. Minhas (2020). Networks of violence: Predicting conflict in nigeria. *The Journal of Politics* 82(2), 476–493.
- DuBois, C., C. T. Butts, D. McFarland, and P. Smyth (2013). Hierarchical models for relational event sequences. *Journal of Mathematical Psychology* 57(6), 297–309.
- Eagle, N. and A. Pentland (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* 10(4), 255–268.

- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.
- Elmer, T., K. Chaitanya, P. Purwar, and C. Stadtfeld (2019). The validity of RFID badges measuring face-to-face interactions. *Behavior Research Methods* 51(5), 2120–2138.
- Elmer, T. and C. Stadtfeld (2020). Depressive symptoms are associated with social isolation in face-to-face interaction networks. *Scientific Reports* 10(1).
- Etezadi-Amoli, J. and A. Ciampi (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the baseline hazard function. *Biometrics* 43(1), 181–192.
- Fjelde, H. and L. Hultman (2014). Weakening the enemy: A disaggregated study of violence against civilians in africa. *Journal of Conflict Resolution* 58(7), 1230–1257.
- Fritz, C., P. W. Thurner, and G. Kauermann (2021). Separable and Semiparametric Network-based Counting Processes applied to the International Combat Aircraft Trades. *Network Science* (OnlineFirst).
- Gade, E. K., M. M. Hafez, and M. Gabbay (2019). Fratricide in rebel movements: A network analysis of syrian militant infighting. *Journal of Peace Research* 56(3), 321–335.
- Gelfand, A. E., S. K. Ghosh, C. Christiansen, S. B. Soumerai, and T. McLaughlin, J. (2000). Proportional hazards models: A latent competing risk approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49(3), 385–397.
- Heckman, J. J. and B. E. Honoré (1989). The identifiability of the competing risks model. *Biometrika* 76(2), 325–330.
- Jäger, K. (2018). The limits of studying networks via event data: Evidence from the ICEWS dataset. *Journal of Global Security Studies* 3(4), 498–511.
- Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data*. Wiley Series in Probability and Statistics. Hoboken: Wiley.
- Kauffmann, M. (2020). Databases for defence studies. In D. Deschaux-Dutard (Ed.), *Research Methods in Defence Studies*, pp. 129–150. London: Routledge.
- King, G. and W. Lowe (2003). An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization* 57(3), 617–642.
- Lerner, J., A. Lomi, J. Mowbray, N. Rollings, and M. Tranmer (2021). Dynamic network analysis of contact diaries. *Social Networks* 66, 224–236.
- Little, R. J. A. and D. B. Rubin (2002). *Statistical analysis with missing data*. Hoboken: John Wiley & Sons, Inc.
- Madan, A., M. Cebrian, S. Moturu, K. Farrahi, and A. S. Pentland (2012). Sensing the health state of a community. *IEEE Pervasive Computing* 11(4), 36–45.
- Malang, T., L. Brandenberger, and P. Leifeld (2019). Networks and Social Influence in European Legislative Politics. *British Journal of Political Science* 49(4), 1475–1498.

-
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27, 415–444.
- Newman, M. E. and J. Park (2003). Why social networks are different from other types of networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 68(3), 8.
- Noghrehchi, F., J. Stoklosa, S. Penev, and D. I. Warton (2021). Selecting the model for multiple imputation of missing data: Just use an IC! *Statistics in Medicine* 40(10), 2467–2497.
- Perry, P. O. and P. J. Wolfe (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society. Series B (Methodological)* 75(5), 821–849.
- Pischedda, C. (2018). Wars within wars: Why windows of opportunity and vulnerability cause inter-rebel fighting in internal conflicts. *International Security* 43(1), 138–176.
- Popovic, M. (2018). Inter-rebel alliances in the shadow of foreign sponsors. *International Interactions* 44(4), 749–776.
- Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research* 47(5), 651–660.
- Rivera, M. T., S. B. Soderstrom, and B. Uzzi (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology* 36, 91–115.
- Robins, G., T. Snijders, P. Wang, M. Handcock, and P. Pattison (2007). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks* 29(2), 192–215.
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika* 63(3), 581.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Ruppert, D., M. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge: Cambridge University Press.
- Stadtfeld, C. (2012). *Events in social networks : A stochastic actor-oriented framework for dynamic event processes in social networks*. Ph. D. thesis, KIT.
- Stadtfeld, C. and P. Block (2017). Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science* 4(14), 318–352.
- Stadtfeld, C., J. Hollway, and P. Block (2017). Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology* 47(1), 1–40.
- Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82(398), 528–540.
- Tranmer, M., C. S. Marcum, F. B. Morton, D. P. Croft, and S. R. de Kort (2015). Using the relational event model (rem) to investigate the temporal dynamics of animal social networks. *Animal behaviour* 101, 99–105.

- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America* 72(1), 20–22.
- Vu, D., P. Pattison, and G. Robins (2015). Relational event models for social learning in MOOCs. *Social Networks* 43, 121–135.
- Vu, D. Q., A. U. Asuncion, D. R. Hunter, and P. Smyth (2011a). Continuous-time regression models for longitudinal networks. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pp. 2492–2500.
- Vu, D. Q., A. U. Asuncion, D. R. Hunter, and P. Smyth (2011b). Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 857–864.
- Wasserman, S. and K. Faust (1994). *Social network analysis : methods and applications*. Cambridge: Cambridge University Press.
- Weidmann, N. B. (2015). On the accuracy of media-based conflict event data. *Journal of Conflict Resolution* 59(6), 1129–1149.
- Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian and New Zealand Journal of Statistics* 48(4), 445–464.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1), 3–36.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Boca Raton: CRC press.
- Wood, S. N. (2020). Inference and computation with generalized additive models and their extensions. *Test* 29(2), 307–339.

Contributing Publications

- Fritz, C., Lebacher, M. and Kauermann, G. (2020) Tempus Volat, Hora Fugit: A survey of Tie-oriented Dynamic Network Models in Discrete and Continuous Time. *Statistica Neerlandica*, 74 (3), 275–299. doi:10.1111/stan.12198.
- Fritz, C. and Kauermann, G. (2022). On the Interplay of Regional Mobility, Social Connectedness, and the Spread of COVID-19 in Germany (2022). *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 185 (1), 400-424 doi:10.1111/rssa.12753..
- Fritz, C., Mehrl, M., Thurner, W. P. and Kauermann, G. (2022) Exponential Random Graph Models for Dynamic Signed Networks: An Application to International Relations. *Under review in the Journal of the American Statistical Association (Applications and Case Studies)*.
- Fritz, C., De Nicola, G., Kevork, S., Harhoff, D. and Kauermann, G. (2022) Modelling the large and dynamically growing bipartite network of German patents and inventors. *Under review in the Journal of the Royal Statistical Society. Series A (Statistics in Society)*.
- Fritz, C., Thurner, W. P. and Kauermann, G. (2021) Separable and Semiparametric Network-based Counting Processes applied to the International Combat Aircraft Trades. *Network Science*, 9 (3), 291–311. doi:10.1017/nws.2021.9.
- Fritz, C., Mehrl, M., Thurner, W. P. and Kauermann, G. (2022) All that Glitters is not Gold: Modeling Relational Events with Spurious Events. *Under review in Network Science*.

Eidesstattliche Versicherung (Affidavit)

(Siehe Promotionsordnung vom 12. Juli 2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 25.5.2021

Cornelius Fritz

