

New Approaches in Statistical Network Data Analysis

Dissertation

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

vorgelegt von

Michael Lebacher

Eingereicht am 05.09.2019

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

1. Berichterstatter: Prof. Dr. Göran Kauermann (LMU München)
2. Berichterstatter: Prof. Dr. Paul W. Thurner (LMU München)
3. Berichterstatter: Prof. Dr. Thomas A. B. Snijders (Rijksuniversiteit Groningen)

Tag der Einreichung: 05.09.2019

Tag der Disputation: 15.11.2019

New Approaches in Statistical Network Data Analysis

Dissertation

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

vorgelegt von

Michael Lebacher

Eingereicht am 05.09.2019

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

1. Berichterstatter: Prof. Dr. Göran Kauermann (LMU München)
2. Berichterstatter: Prof. Dr. Paul W. Thurner (LMU München)
3. Berichterstatter: Prof. Dr. Thomas A. B. Snijders (Rijksuniversiteit Groningen)

Tag der Einreichung: 05.09.2019

Tag der Disputation: 15.11.2019

Acknowledgment

Many persons, directly and indirectly, contributed to the realization of this thesis. First and foremost, I want to thank my supervisor Göran Kauermann for shaping my view on statistical matters, his guidance and advice. Furthermore, I want to express my gratitude to Paul Thurner who helped me with input and encouragement through many dark hours of major revisions.

I also want to thank Tom Snijders who kindly agreed to be part of my examination committee and to review this thesis. Further thanks go to Christian Heumann and Helmut Küchenhoff for their willingness to steer the examination committee.

I want to thank Aude Fleurant and the other employees of the Stockholm International Peace Research Institute for explaining the arms trade data and inviting me to Stockholm to visit their institute. Additionally, my gratitude goes to Nic Marsh from the Norwegian Initiative on Small Arms Transfers for helping me understanding the small arms and ammunition data set. Thanks for fruitful cooperation go to Samantha Cook and Nadja Klein.

Many people from the statistics department in Munich gave me assistance and contributed to a good atmosphere at the institute. Thank you, Michael Windmann, for being a sociable room-mate and for withstanding the strange sounds my computer made under computational overload. Thank you, Benjamin Sischka, for sharing cigarettes on “Smokey Friday” and thank you, Cornelius Fritz, for providing me again and again with a new enthusiasm for frequentist statistics. Thank you, Sevag Kevork, for the many discussions about statistical network data analysis and controversial discussions about politics. Thank you, Marc Schneble, for your administrative help in organizing exercises and partnering me as “civilian service” in Portugal. Thank you, Christoph Striegel, for sharing your dust-dry opinions about books, statistics, economics, and finance with me. Thank you, Verena Bauer, for valuable insights about the drawbacks of canteen food. And thank you, Iris Burger, for kind help whenever I was overwhelmed by organizational stuff.

After this non-exhaustive list, I certainly have to thank my parents Hermann and Hilde and especially my beloved girlfriend Diana.

Kurzfassung

Diese kumulative Dissertation beschäftigt sich mit der statistischen Analyse von Netzwerkdaten. Der generelle Ansatz, interdependente Systeme als Netzwerke zu konzeptualisieren um sie anschließend mit statistischer Methodik zu analysieren, hat in den vergangenen Jahren deutlich an Relevanz gewonnen. Insbesondere die Flexibilität der Methodik, zusammen mit der Möglichkeit komplexe Abhängigkeitsstrukturen zu modellieren, hat zu ihrer Popularität beigetragen.

Ein Netzwerk ist ein System, das sich aus Knoten und Kanten zusammensetzt. Dabei sind die Knoten generelle Einheiten, die durch die Kanten miteinander in Verbindung gebracht werden. Je nach Forschungsfrage interessieren entweder die Abhängigkeiten zwischen den Knoten oder die Verteilung der Kanten mit gegebenen Knoten. Diese Arbeit greift mit insgesamt sechs Artikeln den zweiten Ansatz auf. Unter Zuhilfenahme von statistischen Modellen werden die Kanten in verschiedenen binären und gewichteten Netzwerken analysiert, beziehungsweise rekonstruiert.

Um der Arbeit einen generellen Kontext zu geben, wird den angehängten Artikeln ein Mantelteil vorangestellt. In diesem wird auf zentrale Konzepte und Modelle der statistischen Netzwerkanalyse eingegangen. Dabei werden die Vorteile als auch die Nachteile der Modelle diskutiert und potenzielle Erweiterungen und Modifikationen beschrieben.

Die in dieser Dissertation enthaltenen Artikel lassen sich grob in zwei verschiedene Projekte einordnen. In einem Projekt steht die statistische Modellierung des internationalen Waffenhandels im Fokus. Zwei Artikel untersuchen den globalen Austausch von Großwaffen (Major Conventional Weapons), dabei wird sowohl die dynamische Struktur als auch das gehandelte Waffenvolumen analysiert. Ein weiterer Artikel widmet sich den latenten Strukturen im internationalen Kleinwaffenhandel (Small Arms and Ammunition). Weiterhin werden die Waffenhandelsdaten in einem Übersichtsartikel, der sich mit dynamischen Netzwerkmodellen beschäftigt, verwendet. Das zweite Projekt befasst sich, verteilt über zwei Artikel, mit der Rekonstruktion von finanziellen Netzwerken basierend auf den Randsummen von Netzwerkmatrizen.

Alle in dieser Dissertation angehängten Artikel befinden sich in der Form, in der sie als Vorabversion veröffentlicht wurden. Bei Veröffentlichungen in Fachjournalen wird die jeweilige Quelle angegeben. Zudem wird vor jedem Artikel der Beitrag des jeweiligen Autors angegeben. Sämtliche Analysen wurden mit der statistischen Software R durchgeführt. Der dazugehörige Code ist über Github verfügbar.

Summary

This cumulative dissertation is dedicated to the statistical analysis of network data. The general approach of combining network science with statistical methodology became very popular in recent years. An important reason for this development lies in the ability of statistical network data analysis to provide a means to model and quantify interdependencies of complex systems.

A network can be comprehended as a structure consisting of nodes and edges. The nodes represent general entities that are related via the edges. Depending on the research question at hand, it is either of interest to analyze the dependence structure among the nodes or the distribution of the edges given the nodes. This thesis consists of six contributed manuscripts that are concerned with the latter. Based on statistical models, edges in different dynamic and weighted networks are investigated or reconstructed.

To put the contributing articles in a general context, the thesis starts with an introductory chapter. In this introduction, central concepts and models from statistical network data analysis are explained. Besides giving an overview of the available methodology, the advantages and drawbacks of the models are given, supplemented with a discussion of potential extensions and modifications.

Content-wise it is possible to divide the articles into two projects. One project is focused on the statistical analysis of international arms trade networks. Two articles are devoted to the global exchange of major conventional weapons with a focus on the dynamic structure of the system and the volume traded. A third article explores latent patterns in the international trade system of small arms and ammunition. Additionally, the arms trade data is used in a survey paper that is concerned with dynamic network models. The second project regards the reconstruction of financial networks from their marginals and includes two articles.

All contributing articles are attached in the form as published as a preprint. For publications in scientific journals the respective sources are given. Additionally, the contributions of all authors are included. All computations were done with the statistical software `R` and the corresponding code is available from Github.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Static binary networks	3
1.2.1	Exponential random graph model	6
1.2.2	Further models for static binary networks	8
1.3	Dynamic binary networks	11
1.3.1	Dynamic exponential random graph models	11
1.3.2	Further models for dynamic binary networks	13
1.4	Weighted networks	15
1.4.1	Models for weighted networks	15
1.4.2	Models for network flow data	17
1.5	Networks with limited information	22
1.5.1	Network tomography	23
1.5.2	Maximum-entropy network reconstruction	24
1.5.3	Approaches for sparse networks	26
1.6	Lessons learned	27
	Literature	28
2	A dynamic separable model with actor heterogeneity: An application to global weapons transfers	35
3	Tempus volat hora fugit - A survey of tie-oriented dynamic network models in discrete and continuous time	93
4	Exploring dependence structures in the international arms trade network: A network autocorrelation approach	141
5	Censored regression for modelling international small arms trading and its “forensic” use for exploring unreported trades	143
6	Regression-based network reconstruction with nodal and dyadic covariates and random effects	183
7	In search of lost edges: A case study on reconstructing financial networks	211
	Affidavit	256

Chapter 1

Introduction

Science may be described as the art of systematic oversimplification.
Karl Popper (Popper, 1992, p. 44)

1.1 Overview

The methodology from statistical network data analysis provides a toolbox for investigating various kinds of interdependent systems. Taken as a conceptual framework, it allows to analyze phenomena that can be described in terms of a graph, consisting of a set of nodes and a set of edges. In this setting, the nodes are general entities, e.g. persons, countries or simply things, and the edges represent relations between them. This connectivity typically implies a non-trivial dependence structure among the edges. Additionally, most network systems are characterized not only by static interdependencies but often exhibit a temporal dependence structure. Quantifying these dependencies and drawing inference on it is the goal of the respective statistical analysis and the focus of the models and applications that constitute the contributing manuscripts of this thesis. To do so, different tools from statistical network data analysis are employed to model the dynamic evolution of binary networks, investigate dependence structures of valued edges and to reconstruct networks from limited information.

The first chapter of this thesis is intended to serve as a general introduction to the articles and embeds them in a wider context. Obviously, such an overview is subjectively influenced and, therefore, far from encompassing everything relevant in this broad area. The focus lies on models that are, to a certain degree, of practical importance to the contributing articles. For a general overview and survey articles about the state of the art in statistical network data analysis, refer to Goldenberg et al. (2010), Fienberg (2012), Kolaczyk (2009, 2017) and Squartini et al. (2018).

Section 1.2 introduces static binary networks and gives central definitions and explanations. Based on that, the exponential random graph model (ERGM) is motivated, followed by a brief outline of alternative binary network models. Building on the description of the ERGM, its dynamic extensions, the temporal ERGM (TERGM) and the separable TERGM (STERGM), are explained in Section 1.3. Both sections can be thought of as being related to Chapters 2 and 3 that are concerned with dynamic binary network models. Apart from that, comments and literature for other models designed for analyzing dynamic binary networks are given.

Section 1.4 is focused on weighted networks and models designed for analyzing network flow data. In the first part of Section 1.4, standard models for weighted networks are described. The second part relates to Chapters 4 and 5 and introduces the gravity model as well as two spatial econometric models: the spatial autocorrelation (SAR) model and the

spatial error model (SEM). Based on that, it is explained how spatial econometric models can be used for modeling network flow data.

The main topic of Chapters 6 and 7, reconstructing networks from incomplete information, is picked up in Section 1.5. This comes with a general introduction to the problem of network reconstruction and a discussion of models from the network tomography literature. Additionally, maximum-entropy probability distributions for network reconstruction and density-adjusted methods are motivated. Section 1.6 concludes the introductory chapter.

In Chapters 2-7 all manuscripts are included in their original form¹ as published as a preprint together with a description of the authors contributions. For the manuscripts that are published in scientific journals the respective sources are given.

¹Note that the notation of the original manuscripts partly differs from the notation in the introductory chapter.

1.2 Static binary networks

The difficulty of describing these systems lies partly in their topology: Many of them form rather complex networks whose vertices are the elements of the system and whose edges represent the interactions between them.

Barabási and Albert (1999, p. 509)

Network analysis has its theoretical fundamentals in graph theory and is, therefore, sometimes dated back to Euler’s “*Königsberger Brückenproblem*” (also known as “*puzzle of Königsberg’s bridges*”, Fortunato, 2010, p. 76) in 1736. Although modern statistical network data analysis has emancipated to a certain degree from mathematical graph theory, it is a good starting point to introduce networks by the notion of a *graph*.

Formulation as a graph

A graph is a mathematical structure consisting of *nodes* and *edges*. In short, a graph allows to describe systems that can be characterized in terms of relations between entities. In this thesis, exclusively directed graphs are treated and the terms graph and network are used interchangeably. Formally, a graph with directed edges can be represented as a tuple

$$G = (V, E),$$

where $V = \{v_1, \dots, v_n\}$ is a set of $|V| = n$ nodes and

$$E \subseteq \{(i, j) : i, j \in V, i \neq j\} \quad (1)$$

represents a set of directed edges with $|E| = N_E$. An edge (i, j) can be understood as a directed relation from v_i to v_j . General relations between two nodes, irrespective of the presence or absence of an edge, are referred to as *dyads*. As formalized in the edge set (1), no *self-loops* are considered in this thesis. Therefore, the analysis is restricted to directed networks without edges originating and ending at the same node. As a consequence, the edge set E may contain a maximum of $N = n(n - 1)$ edges. If N edges are present, the graph is said to be *fully connected*.

Formulation as an adjacency matrix

Graph G can be summarized in a (random) $n \times n$ *adjacency matrix* $\mathbf{Y} = (Y_{ij})_{i,j=1,\dots,n} \in \mathcal{Y}$, where \mathcal{Y} represents the set of all 2^N possible directed binary networks with n nodes. For equivalence to graph G set

$$Y_{ij} = \begin{cases} 1 & , \text{ if } (i, j) \in E \\ 0 & , \text{ else} \end{cases}.$$

Hence, the matrix \mathbf{Y} contains information about the presence of edges that originate from the nodes of the corresponding rows and direct to the nodes in the respective columns. In a setting without self-loops, the diagonal elements can be set to zero as a convention, i.e. $Y_{ii} = 0 \forall i \in V$. Considering a directed graph implies that \mathbf{Y} is not symmetric in general.

Typically, binary networks are represented graphically as a network graph with circles (the nodes) and arrows between them (the edges). In Figure 1, a simple directed binary network with $n = 3$ nodes and $N_E = 4$ directed edges is shown formally as a graph in subplot (a), as an adjacency matrix in subplot (b) and as a network graph in the last subplot (c).

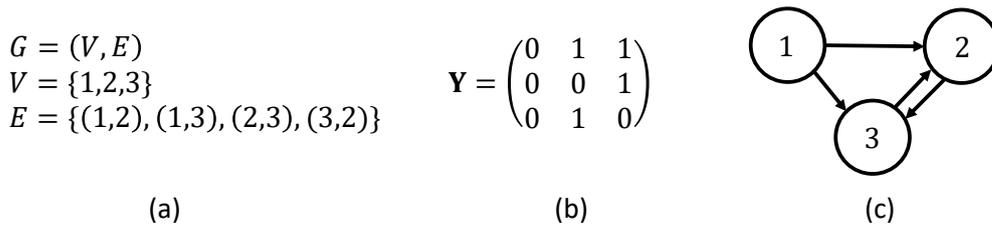


Figure 1: Three different representations of a binary, directed network. Formal representation as a graph (a), as an adjacency matrix (b) and a plot of the network graph (c).

Descriptive network statistics

Although graphical representations of networks can be very conclusive, network graphs often become very chaotic with increasing network size (sometimes called “hairball effect”, Röttgers and Faust, 2018) and demand complex visualization algorithms (e.g. Csardi and Nepusz, 2006). Therefore, *network topologies*, are often described in terms of *network statistics*, i.e. functions $g(\mathbf{Y})$ defined on the adjacency matrix, that quantify network properties of interest.

One of the most important basic measures is given by the number of edges within a network. The *edge count* is defined as

$$g_{edges}(\mathbf{Y}) = \sum_{i \neq j} Y_{ij}. \quad (2)$$

If the number of edges is scaled to the unit interval, by dividing the edge count by the number of possible edges, the statistic is called *density*

$$g_{density}(\mathbf{Y}) = \frac{1}{N} \sum_{i \neq j} Y_{ij}. \quad (3)$$

Using this measure, a network with a high density is called dense, while a network with a low one is called sparse.

As a general philosophy in network analysis, it is often argued that complex global network patterns can be decomposed into simple local ones (Morris et al., 2008). In Figure 2 some stylized local network structures are visualized. As illustrated in the subplots (a) and (b), edges (either outgoing, ingoing or both) might cluster at certain nodes. This can be measured with the so-called *degrees*, a quantity often used as a centrality measure for the nodes. In a network with directed edges, the *outdegree* of node i is defined as

$$g_{outdeg,i}(\mathbf{Y}) \propto \sum_{k \in V, k \neq i} Y_{ik} \quad (4)$$

and the *indegree* of node j is given by

$$g_{indeg,j}(\mathbf{Y}) \propto \sum_{k \in V, k \neq j} Y_{kj}. \quad (5)$$

Based on these network statistics, it is possible to construct the degree distribution, i.e. the share of nodes that comes with a certain degree. Although this is a very simple description, networks are very often characterized by the shape of their degree distributions (see Barabási and Albert, 1999).

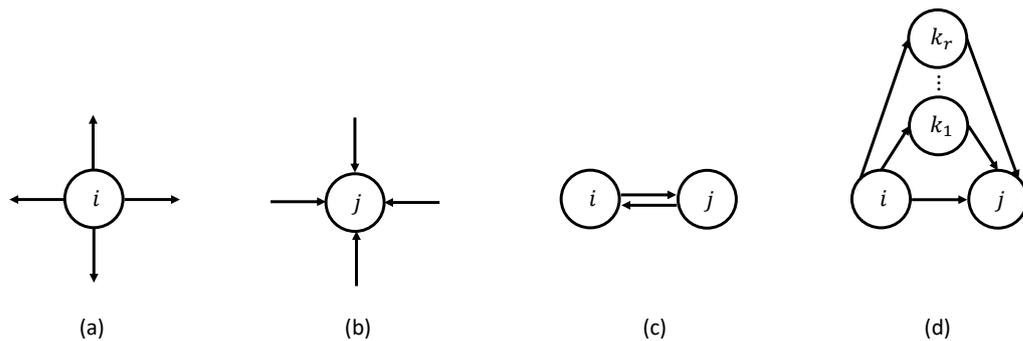


Figure 2: Scheme representing different local network structures. Outdegree of node i (a), indegree of node j (b), reciprocity between nodes i and j (c) and directed two-paths between nodes i and j (d).

Another important topological factor is *reciprocity* as shown in subplot (c) of Figure 2. In many real-world networks, the number of reciprocated edges is quite high (Garlaschelli and Loffredo, 2004). Reciprocal structures in a network can be measured by counting the number of mutual edges

$$g_{recip}(\mathbf{Y}) \propto \sum_{i \neq j} Y_{ij} Y_{ji}.$$

Often, more complicated network statistics that go beyond dyadic relations are considered. They are sometimes called *hyperdyadic* statistics and can be used to measure clustering in a network. Frequently used hyperdyadic measures are *triadic* statistics that count triangle-like structures in the network. An example is shown in the subplot (d) of Figure 2. The statistic

$$g_{dpath,ij}(\mathbf{Y}) \propto \sum_{k \in V, k \neq i, j} Y_{ij} Y_{ik} Y_{kj}. \quad (6)$$

counts the number of directed two-paths that indirectly relate the nodes i and j . A two-path refers to two edges, the first starts at node i and goes to node k and the second one goes from node k to node j . This can be interpreted as the number of directed triangles that are closed by the edge (i, j) .

There exists an abundance of further statistics that allow to describe many different topologies. See Csardi and Nepusz (2006) for an implementation of many descriptive network statistics in R.

Remarks on network statistics

Note that in the last four equations, the network statistics are shown with proportionality signs. Similar to the density (3), being a scaled version of the edge count (2), it can be useful to norm network statistics for ease of interpretation. For example, the outdegree can be normalized by $(n - 1)$, the highest possible outdegree, or the reciprocity statistic could be normed with N_E , giving the share of reciprocated edges among all edges. This can also be of importance to compare networks with differing node sets or dynamic networks where the node set varies with time.

The network statistics introduced above are defined on different levels, i.e. some are global and some are local. While the density gives a description of the whole network, the degree statistics refer to specific nodes and the directed path statistic (6) relates to one edge. Depending on their usage and the model employed, local statistics can be aggregated to global ones, for example by summing the statistic (6) over all edges.

The presented network statistics hint on a crucial feature of statistical network data analysis. If there is a high share of reciprocity in the network or if the edges cluster at certain nodes, it seems to be unrealistic that the edges are distributed randomly in the network. Hence, they are almost certainly not independent. Therefore, network statistics are not only useful for describing topologies but also important for building models that are intended to take into account network interdependencies appropriately. This motivates the construction of models that can reproduce topologies that are reflected by network statistics.

1.2.1 Exponential random graph model

In this section, one of the most popular binary network models, the exponential random graph model (ERGM) is introduced (see Frank and Strauss, 1986, Wasserman and Pattison, 1996, Robins et al., 2007 and Lusher et al., 2013 for a book-length treatment). The ERGM is very attractive for motivating model-based network data analysis because it provides a good example of how binary network dependencies can be modeled with network statistics.

In the following, random variables are denoted in upper case letters while realizations are denoted in lower case. For notational clarity, the dependence on potential covariates is suppressed.

Model motivation

A very simple starting point for analyzing binary networks is the Bernoulli random graph model also known as the Gilbert model (Gilbert, 1959). This model assigns equal probabilities to all edges, such that

$$\mathbb{P}(Y_{ij} = 1; p) = p \text{ for all } i \neq j.$$

The resulting probability distribution for the whole network

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}; p) = p^{\sum_{i \neq j} y_{ij}} (1 - p)^{N - \sum_{i \neq j} y_{ij}}, \quad (7)$$

is an exponential family model (e.g. Fahrmeir et al., 2007) with the edge count (2) as sufficient statistic. To see this, define $\theta = \text{logit}(p) := \log(p/(1 - p))$. Then it holds that

$$\begin{aligned} \mathbb{P}(\mathbf{Y} = \mathbf{y}; \theta) &= p^{\sum_{i \neq j} y_{ij}} (1 - p)^{N - \sum_{i \neq j} y_{ij}} = \frac{\left(\frac{p}{1-p}\right)^{\sum_{i \neq j} y_{ij}}}{(1 - p)^{-N}} \\ &= \frac{\exp\{\theta g_{edges}(\mathbf{y})\}}{(1 + \exp\{\theta\})^N} = \frac{\exp\{\theta g_{edges}(\mathbf{y})\}}{\kappa(\theta)}, \end{aligned}$$

with $\kappa(\theta)$ denoting the normalization constant.

However, in practice it is often very unlikely that the edge probabilities are all equal and hence, this model is much too simplistic for explaining the topologies of complex real-world networks. Technically, this means that the expected values for certain network statistics under model (7) do not match with the observed ones. This motivates an extension of the model by incorporating more network statistics

$$g(\mathbf{Y}) = (g_1(\mathbf{Y}), \dots, g_p(\mathbf{Y}))^T,$$

including for example reciprocity or triadic structures (see Morris et al., 2008 for a broad range of statistics). Extending the simple model (7) results in the ERGM as typically defined in the literature

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}; \theta) = \frac{\exp\{\theta^T g(\mathbf{y})\}}{\kappa(\theta)}. \quad (8)$$

Here, $\theta \in \Theta$ represents a p -dimensional column vector of coefficients. Because the ERGM is an exponential family model, it holds that

$$\mathbb{E}_{\hat{\theta}}[g_l(\mathbf{Y})] = g_l(\mathbf{y}), \text{ for } l = 1, \dots, p$$

with $\hat{\theta}$ being the maximum likelihood estimator for θ . Therefore, the model provides a probability distribution for the graph that, in expectation, replicates the topological structures of the observed network.

Local interpretation

Although model (8) can be said to be global, in the sense that it gives a probability distribution with, e.g. an expected number or reciprocated edges in the network, it is also possible to derive a local interpretation. Define \mathbf{y}_{-ij} to be the network \mathbf{y} , excluding the dyad (i, j) , and \mathbf{y}_+ as the network \mathbf{y} with element $y_{ij} = 1$ and \mathbf{y}_- correspondingly with y_{ij} set to zero. Then, the conditional logarithmic odds are given by

$$\begin{aligned} \text{logit}(\mathbb{P}(Y_{ij} = 1 | \mathbf{Y}_{-ij} = \mathbf{y}_{-ij}; \theta)) &= \log \left(\frac{\mathbb{P}(\mathbf{Y} = \mathbf{y}_+; \theta)}{\mathbb{P}(\mathbf{Y} = \mathbf{y}_-; \theta)} \right) \\ &= \theta^T \{g(\mathbf{y}_+) - g(\mathbf{y}_-)\} = \theta^T \delta(\mathbf{y}). \end{aligned} \quad (9)$$

The *change statistics* $\delta(\mathbf{y})$ measure how $g(\mathbf{y})$ changes if the edge (i, j) is switched from zero to one, leaving the rest of the network untouched. Equation (9) bears similarities to the logistic regression model but comes with the caveat that the change statistics vary with changes of the network and the logarithmic odds interpretation is valid only conditional on the remaining state of the network.

Estimation

The seemingly simple mathematical form of the ERGM (8) might lead to the tempting conclusion that maximum likelihood provides a straightforward solution to the problem of obtaining parameter estimates. The log-likelihood

$$\ell(\theta; \mathbf{y}) = \theta^T g(\mathbf{y}) - \log(\kappa(\theta)),$$

looks innocent but if endogenous network statistics are considered, the normalization constant

$$\kappa(\theta) = \sum_{\mathbf{y} \in \mathcal{Y}} \exp\{\theta^T g(\mathbf{y})\}$$

becomes intractable, except for very small networks (Yon and de la Haye, 2019), impeding direct maximization of the log-likelihood. Consider the comparison of Chandrasekhar and Jackson (2014). Given a small network with $n = 17$ nodes, calculating the normalization constant requires to sum 2^{272} terms, while the estimated number of atoms in the universe is roughly 2^{58} .

One very popular solution to the problem is given by approximations of the log-likelihood. This becomes possible by fixing a parameter vector θ_0 and noting that

$$\begin{aligned} \ell(\theta; \mathbf{y}) - \ell(\theta_0; \mathbf{y}) &= (\theta - \theta_0)^T g(\mathbf{y}) - \log \left(\frac{\kappa(\theta)}{\kappa(\theta_0)} \right) \\ &= (\theta - \theta_0)^T g(\mathbf{y}) - \log \{ \mathbb{E}_{\theta_0} [\exp\{(\theta - \theta_0)^T g(\mathbf{Y})\}] \} \\ &\approx (\theta - \theta_0)^T g(\mathbf{y}) - \log \left\{ \frac{1}{m} \sum_{i=1}^m \exp\{(\theta - \theta_0)^T g(\mathbf{y}_i^S)\} \right\}. \end{aligned}$$

By the law of large numbers, the log-likelihood $\ell(\theta; \mathbf{y})$ can be approximated if it is possible to sample network realizations $\mathbf{y}_1^S, \dots, \mathbf{y}_m^S$ from distribution (8) parameterized with $\theta = \theta_0$. This can be done based on *Monte Carlo Markov Chain* (MCMC) methods, see Geyer and Thompson (1992) for MCMC maximum likelihood and Hunter and Handcock (2006) as well as Hummel et al. (2012) for a more detailed description of the fitting procedure for ERGMs. Further approaches include Bayesian inference (Caimo and Friel, 2011) and maximum pseudolikelihood estimation (MPLE, Strauss and Ikeda, 1990) that can be seen as a local approximation of the log-likelihood, building on a logistic regression model using equation (9). See Chapter 3 for further notes and literature concerning estimation.

Degeneracy

An endemic problem to this model class is called *degeneracy* (see e.g. Snijders et al., 2006, Schweinberger, 2011 or Chatterjee et al., 2013) and describes the circumstance that many regions from the parameter space Θ may lead to distributions where most probability mass is concentrated either to an empty or a full network. This poses a substantial problem for simulation-based fitting procedures and might prevent convergence of MCMC algorithms. Certain statistics or combinations of statistics are especially prone to degeneracy. For example, the inclusion of triadic structures is often impossible without running into degeneracy issues. A potential way out of this problem is the usage of geometrically weighted statistics, see Snijders et al. (2006), Hunter and Handcock (2006) and Chapter 3. The inclusion of the statistic stabilizes the fitting procedure but comes at the cost of simplicity and interpretability.

Node-specific heterogeneity

In the model formulation above, it is assumed that the network statistics are sufficient to capture the heterogeneity of the nodes. Hence, given the included network statistics all nodes are assumed to be equal in a statistical sense. This might not be a very realistic assumption. One possibility to give up this assumption is to incorporate nodal heterogeneity via nodal random effects. This direction is taken by Duijn et al. (2004) in an approach called p_2 model. For further discussion on nodal heterogeneity and appropriate modeling with random effects, see Thiemichen et al. (2016) and Chapter 2.

1.2.2 Further models for static binary networks

Although the ERGM can be said to be the workhorse for modeling static binary networks, there exist further important model classes applicable to binary network data. Two very popular ones are latent space (latent factor) models, introduced by Hoff et al. (2002) (see also Hoff, 2005 and Handcock et al., 2007) and stochastic block models (SBM, Holland et al., 1983, Wang and Wong, 1987, Nowicki and Snijders, 2001). Although they are not applied in the contributing articles, some general notes on these models help to gain a broader understanding of the different approaches in the field.

Latent space model

Latent space models are very general and can be interpreted as an approach that tries to transfer the flexibility of generalized linear models (GLM, e.g. Fahrmeir et al., 2007) to models designed for network data. Given this generality, it is possible to analyze various kinds of networks, including binary networks as well as network with continuous edge values or edges with count values. In contrast to the motivation of the ERGM, the model class does not

explicitly quantify network topologies by utilizing network statistics but assumes that the network dependencies can be represented by positions $Z = (Z_1, \dots, Z_n)^T$ in a latent space. Conditional on the position in the latent space, and potentially covariates, the edges are assumed to be independent

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} | Z = z; \theta) = \prod_{i \neq j} \mathbb{P}(Y_{ij} = y_{ij} | Z_i = z_i, Z_j = z_j; \theta).$$

The probability model is an exponential family distribution, allowing to use a logit model for binary networks

$$\text{logit}(\mathbb{P}(Y_{ij} = 1 | Z_i = z_i, Z_j = z_j; \theta)) = \eta_{ij}(z_i, z_j)$$

and the central idea of latent space models is described best in terms of the predictor

$$\eta_{ij}(z_i, z_j) = \mu_{ij} - d(z_i, z_j). \quad (10)$$

Here, μ_{ij} gives a dyad-specific term that can be expressed as a linear combination of covariates. The function $d(\cdot, \cdot)$ evaluates the distance between the latent space positions z_i and z_j . This function can be any distance measure satisfying the triangle inequality, for example Euclidean with $d(z_i, z_j) = \|z_i - z_j\|_2$ or the absolute distance, $d(z_i, z_j) = \|z_i - z_j\|_1$. With increasing distance in the latent space, the predictor decreases, leading to a reduction of the probability associated with the occurrence of an edge. Hence, nodes that are close in the latent space are more probable to connect. Hoff et al. (2002) explain that the positions in the latent space capture many network dependencies, for example, reciprocity or transitivity. Estimation is possible via maximum likelihood or Bayesian methods and the authors also provide methodology for making inference on the latent positions Z .

Stochastic block model

Stochastic block models (SBM) have their roots in mixture models (e.g. McLachlan and Peel, 2004) and in their basic form, they can be understood as a mixture of simple random graph models. As a general foundation, it is assumed that the nodes $V = \{v_1, \dots, v_n\}$ can be assigned into Q different blocks (also called classes). Within each block, the nodes are homogeneous and only between the blocks they are heterogeneous. An intuitive way to think about SBMs is imagining a re-ordering of the adjacency matrix such that it contains dense blocks on the diagonal while the off-diagonal blocks are sparse.

Define $K_i = (K_{i1}, \dots, K_{iQ})^T$ to be a vector containing indicators, with K_{iq} equal to one if node v_i belongs to block $q \in Q$. The probability of node i belonging to block q is given by

$$\mathbb{P}(K_{iq} = 1) = \alpha_q, \text{ for } q = 1, \dots, Q$$

together with $\sum_{q=1}^Q \alpha_q = 1$. Conditional on the block membership K_i and K_j , the edges Y_{ij} are assumed to be independent Bernoulli random variables with

$$\mathbb{P}(Y_{ij} = 1 | K_{iq} = 1, K_{jr} = 1) = \pi_{qr}.$$

Consequently, the edge probabilities are equal within the blocks and differ for the connections between the blocks. In case, the block membership $K = (K_1, \dots, K_n)$ is known, the model is a non-stochastic block model that can be fitted with maximum likelihood. Obviously, this is very unrealistic and the block membership is unobserved in most real applications. The traditional estimation method in mixture models, the Expectation-Maximization (EM, Dempster et al., 1977) algorithm is not always applicable in its standard form but the SBM can be estimated via variational EM algorithms. Additionally, there exist other methods as

Bayesian estimation based on Gibbs sampling. Hunter et al. (2012) provides further overview on available methodology for estimation. The number of blocks can be chosen by information criteria. Airoldi et al. (2011) and Sweet (2015) provide extensions with covariates. See Goldenberg et al. (2010) for further descriptions and references as well as a comparison of latent space models and SBMs.

1.3 Dynamic binary networks

*The line it is drawn; The curse it is cast; The slow one now; Will later be fast;
As the present now; Will later be past; The order is rapidly fadin'; And the first
one now; Will later be last; For the times they are a-changin.*
Bob Dylan (“The Times They Are a-Changin’” in 1964)

In Kolaczyk (2017, p. 98) modeling dynamic networks is highlighted as one of the important emerging topics of statistical network data analysis: “*Most complex systems are dynamic in nature. So, realistically, the corresponding network graphs and processes thereon are dynamic as well and, ideally, should be analyzed as such*”. From a general statistical perspective, incorporating dynamics can be viewed as adding another dependence structure that either supplements or even replaces the static network dependencies (see e.g. Almquist and Butts, 2014).

1.3.1 Dynamic exponential random graph models

The static ERGM introduced in Section 1.2 can be extended very naturally to models suitable for dynamic networks. In the following, two popular approaches for modeling dynamic networks building on the ERGM class are introduced.

In this section it is assumed that the observed network data is recorded over time, i.e. the binary networks $\mathbf{Y}_t = (Y_{t,ij})_{i,j=1,\dots,n}$ under study are observed at discrete equidistant time points $t = 1, \dots, T$.

Temporal exponential random graph model

The temporal ERGM (TERGM) as proposed by Hanneke et al. (2010) combines an ERGM with a first-order Markov dependence structure to model the transition from $t - 1$ to t via

$$\mathbb{P}(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}; \theta) = \frac{\exp\{\theta^T g(\mathbf{y}_t, \mathbf{y}_{t-1})\}}{\sum_{\mathbf{y}^* \in \mathcal{Y}} \exp\{\theta^T g(\mathbf{y}^*, \mathbf{y}_{t-1})\}}.$$

The network statistics $g(\mathbf{y}_t, \mathbf{y}_{t-1})$ can be evaluated on the network in t but also topologies from the network in $t - 1$ can be incorporated. This gives the possibility to include interactions between the network in $t - 1$ and the network in t . As an example, consider the visualization of a dynamic network with $n = 4$ nodes at $T = 3$ time points in Figure 3.

The exemplary network appears to be relatively stable. Over time, most edges stay present and only two edges are added and deleted, respectively. This pattern can be interpreted as an autoregressive dependence structure that can be measured for example with the statistic

$$g_{\text{autoreg}}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) \propto \sum_{i \neq j} Y_{t,ij} Y_{t-1,ij},$$

meaning that the occurrence of an edge between i and j in $t - 1$ influences the probability of observing an edge between i and j in t .

Another dynamic interpretation of network patterns in Figure 3 could be dynamic mutuality because the mutual edges between nodes i and k are stable over time. This can be quantified with delayed reciprocity

$$g_{\text{drecip}}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) \propto \sum_{i \neq j} Y_{t,ij} Y_{t-1,ji}$$

and helps to investigate whether an edge (j, i) in $t - 1$ alters the edge probability for (i, j) in t .

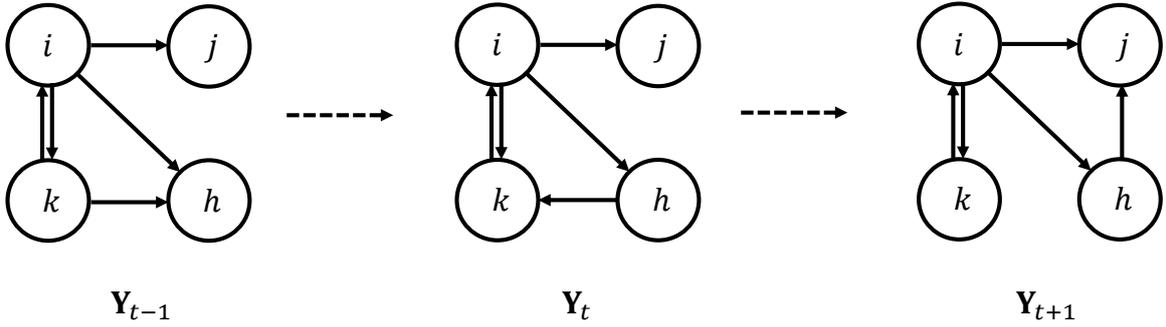


Figure 3: Scheme representing a dynamic network. Three snapshots of a network at different time points.

It also can be proposed that edges that originate at a node with a high outdegree have a low tendency to dissolve over time because the outdegree of node i stays equal to three for all time points in Figure 3. Such a pattern can be covered by including the lagged outdegree of the nodes in a statistic

$$g_{loutdeg}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) \propto \sum_{i \neq j} Y_{t,ij} \left(\sum_{k \in V, k \neq i} Y_{t-1,ik} \right). \quad (11)$$

Additionally, it could be investigated whether triadic structures contribute to the stability of edges. Again, there exist many possibilities to define dynamic network statistics, see Leifeld et al. (2018) for further examples.

Separable temporal exponential random graph model

The concept of the TERGM was refined further by Krivitsky and Handcock (2014). Taking again Figure 3 as an example, it might be plausible to assume that the persistence of edges, for example, the reciprocal relation between i and k is governed by mechanisms that differ from those that lead to the appearance of the edges (h, k) in t or (h, j) in $t + 1$. In their separable TERGM (STERGM) the authors formalize this idea and propose to separate the transition from \mathbf{Y}_{t-1} to \mathbf{Y}_t into two processes, the *formation*, that allows to investigate the occurrence of edges in t without a preceding one in $t - 1$ and the *dissolution* of edges in t , already present in the network in $t - 1$. See Chapters 2 and 3 for further explanations and a comparison between the TERGM and the STERGM.

Changes in the process

When modeling dynamic networks it is important to consider the possibility of changes in the data generating process. Both models, the TERGM and the STERGM allow for different layers of flexibility regarding their parameters. Most restrictively, it can be assumed that the parameter vector θ remains constant for all transitions in the whole time period. However, especially if a long series of networks is investigated, this might be an unrealistic assumption. Given the first-order Markov assumption, this restriction can be relaxed by allowing the parameter vectors to differ for each transition. The obtained parameters can then be plotted against time to show changes in the generative process. Another possibility is the usage of time-varying coefficients (Hastie and Tibshirani, 1993) as proposed in Chapter 2. In such

a framework, the coefficients are allowed to change with time but a penalization on the difference between sequential parameter estimates ensures that these changes are smooth.

Conditional dyadic independence

Similar as in the static ERGM, estimation of θ is possible by approximating the log-likelihood of the TERGM and the STERGM via MCMC sampling. Although these models are less affected by the degeneracy problem (Hanneke et al., 2010) the fitting procedure can be tedious and even infeasible for big and for medium-sized networks. Further, they rely on nodal homogeneity. In Chapter 2 a variant of the STERGM is proposed that circumvents the simulation-based fitting procedure and allows for the inclusion of random heterogeneity components and time-varying effects. This is possible if the network statistics are restricted to the following structure

$$g_l(\mathbf{y}_t, \mathbf{y}_{t-1}) = \sum_{i \neq j} y_{t,ij} \tilde{g}_{l,ij}(\mathbf{y}_{t-1}), \text{ for } l = 1, \dots, p \quad (12)$$

with $\tilde{g}_{l,ij}$ being a network statistic evaluated on \mathbf{y}_{t-1} for dyads (i, j) . An example for such a statistic is the lagged outdegree (11) but also delayed reciprocity and the autoregressive statistic. Given this restriction, the model collapses to a logit model. This can be illustrated using arguments from Strauss and Ikeda (1990) building on the conditional logarithmic odds from (9) with $\mathbf{y}_{t,-ij}$ being the network \mathbf{y}_t , excluding the dyad (i, j) . Matrix $\mathbf{y}_{t,+}$ gives the network \mathbf{y}_t with element $y_{t,ij} = 1$ and $\mathbf{y}_{t,-}$ with $y_{t,ij} = 0$:

$$\text{logit}(\mathbb{P}(Y_{t,ij} = 1 | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \mathbf{Y}_{t,-ij} = \mathbf{y}_{t,-ij}; \theta)) = \theta^T \{g(\mathbf{y}_{t,+}, \mathbf{y}_{t-1}) - g(\mathbf{y}_{t,-}, \mathbf{y}_{t-1})\}. \quad (13)$$

Inserting now the restricted statistics (12) in (13) results in

$$\begin{aligned} \text{logit}(\mathbb{P}(y_{t,ij} = 1 | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \mathbf{Y}_{t,-ij} = \mathbf{y}_{t,-ij}; \theta)) &= \theta^T \tilde{g}_{ij}(\mathbf{y}_{t-1}) \\ &= \text{logit}(\mathbb{P}(Y_{t,ij} = 1 | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}; \theta)), \end{aligned} \quad (14)$$

with $\tilde{g}_{ij}(\mathbf{y}_{t-1}) = (\tilde{g}_{1,ij}(\mathbf{y}_{t-1}), \dots, \tilde{g}_{p,ij}(\mathbf{y}_{t-1}))^T$. Clearly, the resulting change statistic is invariant to changes of \mathbf{y}_t and the model can be represented as a logit model, conditional on the network in $t - 1$. This approach has many advantages, including the possibility to fit big networks very quickly without the need to apply simulation-based fitting procedures. On top of that, all popular extensions of generalized linear models, for example random effects or smooth covariate effects can be employed in network analysis without any difficulties. Nevertheless, assuming (12) effectively implies that conditioning on the network embedding in $t - 1$ is sufficient for assuming dyadic independence in t which may be questionable (e.g. Lerner et al., 2013).

1.3.2 Further models for dynamic binary networks

Stochastic actor-oriented model

Especially for the analysis of social networks, stochastic actor-oriented models (SAOM, Snijders, 2017, Snijders et al., 2010) are very popular. Most importantly, they differ from the TERGM family because in the SAOM it is explicitly assumed that the observed networks \mathbf{y}_{t-1} to \mathbf{y}_t are in fact realizations of a *time-continuous process* that governs the dynamic evolution of the network over time. Within the continuous time intervals actors (nodes), are selected and get a chance to alter the state of one outgoing dyadic relation.

Although being different models with different assumptions, designed for different data situations, there is a growing (partially even controversial) literature that draws comparisons between the (T)ERGM and the SAOM, see Desmarais and Cranmer (2012a), Block et al. (2018), Block et al. (2019), Leifeld and Cranmer (2019).

Dynamic latent space model

The latent space models as introduced in Section 1.2 can be extended toward dynamic networks. In this model class, the difficulties of modeling dynamics come with the handling of the latent space over time. If separate positions in the latent space are estimated for each time point, the representation might become unstable and does not necessarily yield an interpretable pattern. Different solutions are proposed in the literature. Sarkar and Moore (2006) propose to extend the static latent space model by allowing for an autoregressive structure of the latent positions. Hoff (2015) gives a very pragmatic approach and proposes to assume constant latent space positions over time. In Ward et al. (2013), the latent space positions in $t - 1$ can affect the response in t and in Sewell and Chen (2015) much effort is put into providing a meaningful dynamic latent space representation with temporal trajectories.

Dynamic stochastic block model

In stochastic block models, group-level heterogeneity is assumed to be captured with the blocking but dynamics necessarily imply that the assignment of individual nodes to different blocks may change with time. In Yang et al. (2011) a discrete-time SBM is proposed where the nodes can change their blocks but the edge probabilities remain constant with time. This is extended by Xu and Hero (2013) with a state-space model where both, the edge probabilities and the block membership, can vary with time. In Matias and Miele (2017) SBMs combined with Markov chains for the time-varying block memberships are investigated. The authors also note that at present the dynamic block assignments can be very hard to compare and often suffer from identifiability issues.

1.4 Weighted networks

What nature hath joined together, multiple regression analysis cannot put asunder.
Richard E. Nisbett (Nisbett, 2015, p. 187)

While many models designed for network data are suited very well to investigate complex network interdependencies between binary edges, network data very often comes with valued edges. The edge values usually convey meaningful properties of the system under study and hence binarization of the network typically implies a loss of information. This motivates the need for models that can extract information from the edge values.

This section starts with some brief definitions. Further, it introduces models designed for networks with valued edges. The last part of this section is devoted to the analysis of network flow data, including the gravity model and models from spatial econometrics.

Definitions

Weighted networks can be formalized by supplementing the graph $G = (V, E)$, as defined in Section 1.2, with values $\mathcal{W}((i, j))$ for each edge $(i, j) \in E$. In the following, this is characterized by defining the weighted adjacency matrix $\mathbf{Y} = (\tilde{Y}_{ij})_{i,j=1,\dots,n}$. As a convention, diagonal elements are not regarded and edges not present in the edge set are set to zero:

$$\tilde{Y}_{ij} = \begin{cases} \mathcal{W}((i, j)) & , \text{ if } (i, j) \in E \\ 0 & , \text{ else} \end{cases} .$$

Similar as in the binary case, the density $g_{density}(\tilde{\mathbf{Y}})$ of a weighted network can be defined by dividing the number of edges within E by the number of all potential edges N . Also other concepts from binary networks can be translated directly to weighted networks but potentially come with a different interpretation. While it is possible to define measures like weighted degrees or reciprocity for weighted networks, the task becomes much more complicated for hyperdyadic statistics. Once the network is not binary anymore, many of the appealing and accessible properties of statistics that encode for example triadic structures are lost.

1.4.1 Models for weighted networks

It is fair to say that the development of models for weighted networks is still in its infancy and a canonical model is not found yet. In the following, three network models that are capable to analyze weighted adjacency matrices are presented. Two of them are offspring of the ERGM family: the generalized ERGM (GERGM, Desmarais and Cranmer, 2012b, Wilson et al., 2017) and the valued ERGM by Krivitsky (2012). The third model class are latent space models (Hoff et al., 2002) as already presented in Section 1.2.

Generalized exponential random graph model

One of the main challenges of generalizing the ERGM towards valued edges comes with the problem that convergence of the normalization constant is no longer ensured if the edge values are unbounded. However, continuous edge values within the unit interval ensure a convergent normalization constant. Assuming that $\tilde{\mathbf{X}}$ represents a restricted network with edges $\tilde{X}_{ij} \in [0, 1]$ allows to define the model

$$\mathbb{P}(\tilde{\mathbf{X}} = \tilde{\mathbf{x}}; \theta) = \frac{\exp\{\theta^T g(\tilde{\mathbf{x}})\}}{\int_{[0,1]^N} \exp\{\theta^T g(\tilde{\mathbf{x}}^*)\} d\tilde{\mathbf{x}}^*},$$

where $g(\cdot)$ gives a vector of network statistics. The central model ingredient of the GERGM is given by an one-to-one monotone transformation $\tilde{Y}_{ij} = R^{-1}(\tilde{X}_{ij}, \beta)$ that maps the edge values $\tilde{X}_{ij} \in [0, 1]$ to $\tilde{Y}_{ij} \in \mathbb{R}$. The transformation is parameterized with the vector β and the function

$$\frac{\partial R_{ij}(\tilde{\mathbf{y}}, \beta)}{\partial \tilde{Y}_{ij}} = r_{ij}(\tilde{Y}_{ij}, \beta)$$

gives the partial derivatives with respect to \tilde{Y}_{ij} . Using standard results on transformed random variables, it holds that

$$\mathbb{P}(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}; \theta, \beta) = \frac{\exp\{\theta^T g(R(\tilde{\mathbf{y}}, \beta))\}}{\int_{[0,1]^N} \exp\{\theta^T g(R(\tilde{\mathbf{y}}^*, \beta))\} d\tilde{\mathbf{y}}^*} \prod_{i \neq j} r_{ij}(\tilde{y}_{ij}, \beta).$$

The choice of $R(\tilde{\mathbf{y}}, \beta)$ is left to the researcher but Desmarais and Cranmer (2012b) suggest to use cumulative distribution functions for the transformation. Following their advice, r_{ij} is simply the corresponding density. If a network with real-valued edges is to be analyzed, choosing a cumulative normal distribution for $R(\tilde{\mathbf{y}}, \beta)$ might be an appropriate choice. This leads to an attractive interpretation in the case the data does come without endogenous network dependencies ($\theta = 0$). Then, the GERGM collapses into a standard linear regression model with independent observations \tilde{Y}_{ij} (potentially conditional on further covariates). For model inference and log-likelihood approximation via MCMC, see Wilson et al. (2017).

The GERGM is an attractive model for small and fully connected networks. On the other hand, these two properties also represent the major limitations of the approach. It is hard or even infeasible to work with zero-inflated data (i.e. networks with a density less than one) and the estimation procedure does not scale well to big networks. See the articles by Simpson et al. (2013), Simpson and Laurienti (2015), Ward et al. (2013), Boivin and D'Elia (2017) and Schoeneman et al. (2017) for comments on the shortfalls of the GERGM for networks that are big or come with a low density.

Valued exponential random graph model

Another direction, but also starting from the ERGM family, is taken by Krivitsky (2012). The author presents a model, called *valued ERGM*, that is most suitable for count-valued edges, i.e. $\tilde{Y}_{ij} \in \mathbb{N}_0$. The model is defined by the following probability mass function

$$\mathbb{P}(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}; \theta) = \frac{h(\tilde{\mathbf{y}}) \exp\{\theta^T g(\tilde{\mathbf{y}})\}}{\sum_{\tilde{\mathbf{y}}^* \in \tilde{\mathcal{Y}}} h(\tilde{\mathbf{y}}^*) \exp\{\theta^T g(\tilde{\mathbf{y}}^*)\}} = \frac{h(\tilde{\mathbf{y}}) \exp\{\theta^T g(\tilde{\mathbf{y}})\}}{\tilde{\kappa}(\theta)},$$

with $h(\tilde{\mathbf{y}})$ being the reference measure. Given that convergence of the normalization constant is not generally ensured for valued edges, the parameter space Θ must be restricted to finite normalization constants

$$\Theta \subseteq \{\theta^* \in \mathbb{R}^p : \tilde{\kappa}(\theta^*) < \infty\}.$$

In order to get a better understanding for the role of the reference measure, regard an example by Krivitsky (2012). He proposed the reference measure

$$h(\tilde{\mathbf{y}}) = \prod_{i \neq j} \frac{1}{\tilde{y}_{ij}!}.$$

In a simplistic network model, including only the sum of the edge values as network statistic, this results in

$$\mathbb{P}(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}; \theta) = \prod_{i \neq j} \frac{\mu(\theta)^{\tilde{y}_{ij}} \exp\{-\mu(\theta)\}}{\tilde{y}_{ij}!},$$

being the distribution of independently, identically distributed Poisson random variables with expectation $\mu(\theta) = \exp\{\theta\}$. Although the reference measure determines the general shape of the distribution, this does not necessarily imply that a model with more complicated network statistics also results in a Poisson regression model. Selecting network statistics can be tricky because they must be constructed such that a convergent normalization constant is ensured.

Krivitsky (2012) also discusses the possibility of using the valued ERGM for continuous edge values but so far such extensions are neither implemented in software, nor applied or extensively discussed in the literature.

Latent space model

Although already discussed in Section 1.2, it is appropriate to mention here again that the flexibility of the latent space models allows for investigating weighted networks. Since latent space models build on the generalized linear model infrastructure, different data structures can be modeled by simply choosing an adequate distribution from the exponential family. For example a Normal distribution for continuous edge values or a Poisson distribution for count data associated with the edges. Otherwise, the same basic equation (10) for the predictor is valid.

While the approach can be criticized for not relying on interpretable network statistics, in a setting with a weighted network this is actually an advantage. Because the network dependencies must not be specified explicitly, the problem of defining statistics for weighted networks can be circumvented. However, similar to in the GERGM, the model class is limited to fully connected networks and cannot handle zero-inflated data.

1.4.2 Models for network flow data

Depending on the system of interest, the values associated with the edges have plenty of interpretations. Very often these values can be interpreted as flows between the nodes. The flows might be physical, representing transfers of goods in trade networks or traffic in transportation networks. Further, the flows can be digital, for example in computer networks or financial networks. Because network flow data emerges in many disciplines and comes with a very specific interpretation, there is a long history of analyzing network flow data (see Kolaczyk, 2009). In Chapters 4, 5, 6 and 7 physical and digital network flow data is analysed which motivates the introduction of some related models that are important for these articles.

Gravity model

A model specifically designed for network flow data is the so-called *Gravity Model*. The general idea is rather old and early versions can be found for example in the social sciences (Stewart, 1941) or economics (Tinbergen, 1963). As a general motivation, the model borrows from the Newtonian law of universal gravitation and postulates that the edge (flow) values $\check{Y}_{ij} \in \mathbb{R}_+$ can be described as follows

$$\check{Y}_{ij} \propto GS_i^{\rho_S} R_j^{\rho_R} D_{ij}^{\rho_D}, \quad (15)$$

with S_i being a sender-related factor, R_j a receiver-related factor, D_{ij} gives a factor on the dyadic level and G is a constant. The parameters ρ_S , ρ_R and ρ_D give the strengths of the factors. Taking the original (undirected) model of Newton's law of universal gravitation, the sender- and receiver-related factors represent the masses of objects i and j with $\rho_S = \rho_R = 1$, the sender-receiver related factor gives the distance between them with $\rho_D = -2$ and G gives the gravitational constant.

While relation (15) is assumed to hold exactly in physics, in social science applications typically a stochastic component ϵ_{ij} is added and the factors of the model are exchanged with suitable covariates. Furthermore, most often a linearized representation

$$\log(\tilde{Y}_{ij}) = \tilde{Y}_{ij} = \log(G) + \rho_S \log(S_i) + \rho_R \log(R_j) + \rho_D \log(D_{ij}) + \epsilon_{ij}, \quad (16)$$

is considered for estimation.² For example in economics, the model is amended such that the economic “mass” of countries is measured with the gross domestic product and for D_{ij} again physical distance is taken (Head and Mayer, 2014, Disdier and Head, 2008).

In most applications of the gravity model to trade networks, endogenous network dependencies are not considered and it is assumed that exogenous covariates are sufficient to justify conditional independence of the edge values \tilde{Y}_{ij} . However, one research strand, originating in spatial econometrics (LeSage and Pace, 2009) has started to doubt this assumption. LeSage and Pace (2008) write in their widely accepted textbook on spatial econometrics on page 211: “[...] *assuming independence between flows is heroic* [...]” and propose to add origin-, destination- and origin-destination dependence to international flows of commercial goods.

Taking a network perspective, the analog to this model is a representation that allows for the dependence of edge values based on the network structure.

Spatial econometric models

Spatial econometric models are used for decades in statistical network data analysis but they are typically *not considered* for modeling edge values. Most often, the approach is employed for characterizing real-valued variables that embedded in binary networks, i.e. they are used to investigate node-specific characteristics. In this literature, the models are called *network autocorrelation models* and have found many applications in sociology, political sciences and other fields (see Doreian, 1980, Dow et al., 1982, Franzese Jr and Hays, 2007, Hays et al., 2010, Metz and Ingold, 2017 and Silk et al., 2017).

There are two canonical models in spatial econometrics (Kauermann et al., 2012) that allow formalizing spatial dependence structures. In the spatial autocorrelation (SAR) model the dependence is assumed to directly affect the response and in the spatial error model (SEM) the dependencies are related to the error terms. In the general discussion of these models, a real-valued (after suitable transformation) response vector \tilde{Y} is used without explicitly defining whether \tilde{Y} represents edge values or node-specific values.

Spatial error model

The SEM can be characterized by the following set of equations

$$\begin{aligned} \tilde{Y} &= X\beta + u \\ u &= W(\rho)u + \epsilon \\ \epsilon &\sim \mathcal{N}_N(0, \sigma^2 I_N), \end{aligned} \quad (17)$$

with X being a $N \times k$ design matrix of covariates, I_N is the $N \times N$ identity matrix, $\beta \in \mathbb{R}^k$ represents the vector of coefficients for the covariates and $W(\rho)$ is a $N \times N$ weight matrix that depends on one or more parameters ρ , encoding the strength of the spatial correlation. In this model, it is assumed that covariates determine the “correct” conditional expectation while the correlated errors lead to deviations from it (Leenders, 2002). An advantage of the model

²Another version is to model the logarithm of the expectation instead of the expected logarithm. See Kolaczyk (2009, Chap. 9.2) for a discussion.

is given by its straightforward interpretation. To see this, define $B(\rho) := (I_N - W(\rho))^{-1}$ and note that model (17) is equivalent to

$$\begin{aligned}\tilde{Y} &= X\beta + u \\ u &\sim \mathcal{N}_N(0, \sigma^2 B(\rho)B(\rho)^T).\end{aligned}$$

Hence, the whole dependence structure can be shifted into the variance-covariance matrix and

$$\mathbb{E}[\tilde{Y}|X] = X\beta$$

with β providing simply the effect of changes in the covariates on the conditional expectation as in a standard linear model.

Spatial autoregressive model

Turning to the SAR model, similar but re-arranged ingredients are used in order to define the model

$$\begin{aligned}\tilde{Y} &= X\beta + W(\rho)\tilde{Y} + \epsilon \\ \epsilon &\sim \mathcal{N}_N(0, \sigma^2 I_N).\end{aligned}$$

The main difference in comparison to the SEM comes with the assumption that the response variables are directly related, leading to so-called *spill-over effects*. Again this can be demonstrated with an alternative representation of the model:

$$\begin{aligned}\tilde{Y} &= B(\rho)X\beta + u \\ u &\sim \mathcal{N}_N(0, \sigma^2 B(\rho)B(\rho)^T).\end{aligned}$$

Now, the conditional expectation

$$\mathbb{E}[\tilde{Y}|X] = B(\rho)X\beta$$

is slightly more involved because $B(\rho)$ can be thought of as acting like a “filter” on the covariates. Consequently, if some covariates of a certain variable change, close neighbors (in the spatial or the network sense) get affected. To be precise, if an explanatory variable changes for one entity, the effect “spills over” to all related neighbors and again to the neighbors of the neighbors and so forth. This holds because $B(\rho)$ is an infinite sum

$$B(\rho) = (I_N - W(\rho))^{-1} = I_N + W(\rho) + W^2(\rho) + W^3(\rho) + \dots$$

and typically, the spatial weight matrix comes with the simple parametric form

$$W(\rho) = \rho W$$

with $\rho \in (-1, 1)$ and hence the spill over effects get weaker, the more one moves away from the direct neighbors.

Spatial econometric models for valued edges

Assume now that the response is explicitly defined to represent logarithmically transformed edge values $\tilde{Y} \in \mathbb{R}^{N_E}$. Based on that, it can be illustrated, how to formulate the SAR model or the SEM in such a way that it can be utilized for analyzing the edges instead of the nodes.

In the subplot (a) of Figure 4 a weighted network with three nodes and edges $\tilde{Y} = (\tilde{Y}_{12}, \tilde{Y}_{13}, \tilde{Y}_{23}, \tilde{Y}_{32})^T$ is illustrated. Following the “gravity hypothesis” introduced above, it

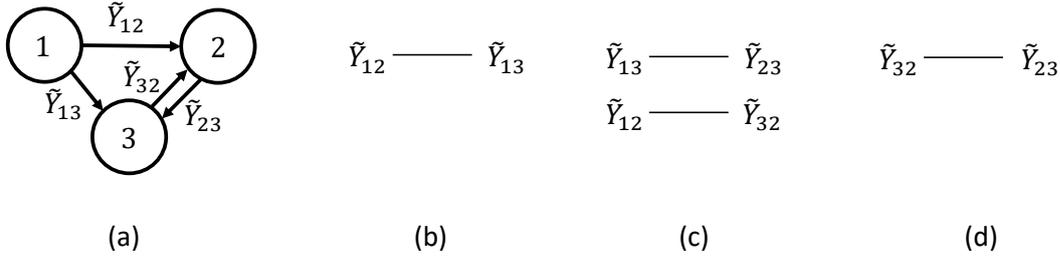


Figure 4: Illustrative scheme. Exemplary network (a) and dependence structure among edges: Sender-related dependence (b), receiver-related dependence (c) and dyadic dependence (d).

could be plausible to assume that edges originating from the same node, pointing to the same node, or those that are mutual are correlated. To model these structures, the spatial weight matrix can be extended to incorporate edge dependencies

$$\begin{aligned}
 W(\rho) &= \rho_S W_S + \rho_R W_R + \rho_D W_D \\
 &= \rho_S \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \rho_R \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} + \rho_D \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (18)
 \end{aligned}$$

A graphical representation is shown in the subplots (b), (c) and (d) of Figure 4. The sender-related dependence (b) assumes the two transfers \tilde{Y}_{12} and \tilde{Y}_{13} to be related because both originate at node 1. Similar, the receiver-related dependence (c) relates \tilde{Y}_{13} and \tilde{Y}_{23} as well as \tilde{Y}_{12} and \tilde{Y}_{32} because the share they same receiver nodes 3 and 2, respectively. The dyadic dependence (d) allows for a correlation of the mutual edges \tilde{Y}_{23} and \tilde{Y}_{32} .

Estimation

In contrast to many other network models, direct maximization of the log-likelihood is possible if spatial econometric models are used. Therefore, the model is, in principle, scalable to big networks. For fitting and inference in the SEM, see the supplementary material of Chapter 4. However, for more complex models time-consuming simulation-based methods might be necessary. This is the case in Chapter 5, where a censored SAR model is estimated using Monte Carlo Expectation Maximization (MCEM, Wei and Tanner, 1990).

Types of network dependencies

The SAR model and the SEM are restrictive in the sense that exclusively linear relations between endogenous variables are possible. Linear means here, that the model only allows for relating \tilde{Y} to $W(\rho)\tilde{Y}$, i.e. it is not possible to model, for example, an association between \tilde{Y}_{13} and $\tilde{Y}_{12}\tilde{Y}_{23}$. Depending on the application, the lack of ability to include non-linear relations may not necessarily be a serious shortcoming because hyperdyadic structures are typically very hard to interpret anyway. Nevertheless, this is a drawback in comparison to the GERGM, which can include triadic network statistics evaluated at the transformed network. Similarly, the linear endogenous relations that can be covered are potentially not as rich as the ones that can be captured in the latent space of latent space models.

Model specification

In comparison to spatially related data, network data has the drawback of lacking a clear physical measure for distance or closeness. Hence, the construction of the weight matrices can be regarded as somewhat arbitrary. Although the dependence structure represented by the weight matrices follows clear rules that derive from the original network, there are myriads of possible combinations that can be defined on a given network, potentially even including weight matrices constructed from interactions with exogenous covariates. These are issues that raise the need for carefully specifying the model or using model selection techniques. The two articles of Chapters 4 and 5 differ in the way network dependencies are included. While in Chapter 4, three rather abstract structures are defined and then compared using the Akaike Information Criterion (AIC, Claeskens and Hjort, 2008), Chapter 5 builds on defining three interpretable weight matrices, similar to the decomposition shown in equation (18).

Sparsity

Weighted networks are often sparse, leading to zero-inflated data. In such a case, modeling the whole weighted adjacency matrix is clearly in conflict with the assumption of multivariate normality. Unlike the GERGM or latent space models, the problem can be dealt with in a rather simple way with spatial econometric models. One approach is to use a conditional model, i.e. including only observations where an edge value is given. This approach also corresponds to the example illustrated in Figure 4 with equation (18) referring only to non-zero edges. Another approach is pursued in Chapter 5 where a censored regression model is applied, i.e. a latent, fully connected network with continuous edge values is assumed to lie behind the observed one and all edge values below a certain threshold are assumed to be censored.

1.5 Networks with limited information

There are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know.

Donald Rumsfeld in February, 2012 during the Pentagon news briefing

In the previous sections, it was simply assumed that networks, either binary or weighted, are fully observed over all time points and all network properties of interest can be investigated with suitable models. Nevertheless, dyadic data is expensive to collect, or even impossible to obtain, for various reasons. It is, therefore, not surprising that in many cases the observed network data is incomplete and only partial information is available, for example only node-specific information, only relatively high edge values or only the binary structure without information on the corresponding edge values.

In this section, a special case of studying networks with partial information is considered. It is assumed that interest is in the specific edge values but only the sum of out- and ingoing edge values of the nodes are observed. In the literature, this problem is often called *network reconstruction*, see Squartini et al. (2018) for a recent overview from a methodological perspective.

Formulation of the problem

To illustrate the problem it is helpful to regard an exemplary weighted network $\tilde{\mathbf{y}}$ with $n = 7$ nodes and edge values $\tilde{y}_{ij} \in \mathbb{R}_+$ shown in Table 1.1. Unlike the settings discussed in the previous sections, neither the binary structure nor the corresponding values are observed. This means, the data available to the researcher is the one given in Table 1.2 and the information is restricted to the row and column sums, shown on the rightmost column and the last row of the table. The task is then to make the best use of the available information in Table 1.2 to reconstruct the edge values as accurate as possible.

The row and column sums can be defined in analogy to the binary in- and outdegree from equations (4) and (5). The *weighted outdegree* is given by

$$g_{woutdeg,i}(\tilde{\mathbf{Y}}) = \sum_{k \in V, k \neq i} \tilde{Y}_{ik} \quad (19)$$

and the *weighted indegree* of node j can be defined as

$$g_{windeg,j}(\tilde{\mathbf{Y}}) = \sum_{k \in V, k \neq j} \tilde{Y}_{kj}. \quad (20)$$

Using these definitions and denoting $\tilde{\mathbf{y}} = \text{vec}(\tilde{\mathbf{y}})$ to be a vectorized network realization (excluding the diagonal elements), the stacked unknown matrix elements $\tilde{\mathbf{y}}$ and the known weighted degrees can be related linearly via

$$g_{wdeg_s}(\tilde{\mathbf{y}}) := \begin{pmatrix} g_{woutdeg,1}(\tilde{\mathbf{y}}) \\ \vdots \\ g_{woutdeg,n}(\tilde{\mathbf{y}}) \\ g_{windeg,1}(\tilde{\mathbf{y}}) \\ \vdots \\ g_{windeg,n}(\tilde{\mathbf{y}}) \end{pmatrix} = \tilde{\mathbf{y}}_+ = A\tilde{\mathbf{y}}, \quad (21)$$

with $\tilde{\mathbf{y}}_+$ representing the stacked $2n$ row- and column sums of the weighted adjacency matrix and A is a $2n \times N$ *routing matrix*. The routing matrix is assumed to be deterministic and

	1	2	3	4	5	6	7	Σ
1	-	3.78	1.05	0.45	0.99	1.23	2.67	10.17
2	2.46	-	2.33	17.45	9.12	2.34	1.56	35.26
3	1.23	1.11	-	2.52	3.44	1.34	0.90	10.54
4	1.93	2.43	5.45	-	0.47	0.54	2.34	13.16
5	1.20	8.89	2.00	3.45	-	2.22	1.23	18.99
6	9.89	10.67	22.8	6.78	10.23	-	9.12	69.49
7	17.67	3.43	9.97	2.12	7.65	10.80	-	51.64
Σ	34.38	30.31	43.60	32.77	31.90	18.47	17.82	

Table 1.1: Unobserved real network $\tilde{\mathbf{y}}$ with matrix entries \tilde{y}_{ij} . Row- and column sums in the rightmost column and last row, respectively.

	1	2	3	4	5	6	7	Σ
1	-							10.17
2		-						35.26
3			-					10.54
4				-				13.16
5					-			18.99
6						-		69.49
7							-	51.64
Σ	34.38	30.31	43.60	32.77	31.90	18.47	17.82	

Table 1.2: Observed information $\tilde{\mathbf{y}}_+$. Row- and column sums of the unobserved matrix entries in the rightmost column and last row, respectively.

simply extracts and sums up the elements of the vector $\tilde{\mathbf{y}}$ according to the weighted out- and indegrees.

1.5.1 Network tomography

One research strand that evolved in parallel to network reconstruction originates in computer science and is called *network tomography* (Vardi, 1996). Network tomography is concerned with the estimation of origin-destination traffic matrices. The edge values are assumed to represent *known* flows. The difficulties enter the problem because the traffic volumes are allowed to pass through the nodes, i.e. some nodes are only inter-stations. It is then the goal to predict the traffic volume between arbitrary origin and destination nodes that potentially passes intermediate nodes.

Although the problem appears to be different, the mathematical formulation is almost identical to the reconstruction problem formulated above. Therefore, some of the methodology developed for network tomography applies to the network reconstruction problem. There are, nevertheless, some subtle differences. Most importantly, in network tomography, many edges are known to be zero which can greatly reduce the dimensionality of the problem (i.e. a lower number of columns in the routing matrix A). Consequently, not all methods from network tomography can be translated to network reconstruction problems. Here, the popular approach of regularized least squares, that applies to both fields, is presented.

Regularized least-squares

Looking at equation (21), it seems natural to approach the problem in terms of a linear model

$$\tilde{y}_+ = A\mu + \epsilon,$$

where ϵ is an error term and $\mu = (\mu_{12}, \dots, \mu_{n(n-1)})^T$ gives the vector of non-negative expected values. However, $A^T A$ is singular and yields an infinite number of solutions. This fundamental identification problem can be tackled using a regularized least-squares model with the following loss function

$$L(\mu) = (\tilde{y}_+ - A\mu)^T (\tilde{y}_+ - A\mu) + \psi J(\mu), \quad (22)$$

together with the restriction $\mu_{ij} \geq 0$ for $i \neq j$. Typical regularization terms in statistics (e.g. Hastie et al., 2009) are for example

$$J(\mu) = \mu^T \mu$$

yielding a ridge regression model or

$$J(\mu) = \sum_{i \neq j} |\mu_{ij}|$$

resulting in the least absolute shrinkage and selection operator (LASSO) regression model. For example Chen et al. (2017) combine elements of the LASSO and ridge regression to reconstruct traffic in biking networks. In Meinshausen et al. (2013) a LASSO-type penalty is recommended for network tomography and Castro et al. (2004) introduces shrinkage estimators for network tomography in a wider context. A model with $J(\mu)$ based on information-theoretic reasoning is proposed by Zhang et al. (2003).

All of these approaches rely on finding good parameter values for the penalization strength ψ which is apparently a hard task if no training set is available and hence some authors simply recommend values based on simulation experiments or assume training data to be available. Another potential shortfall of these approaches is given by the shrinkage property of regularized models, i.e. the reconstructed edge values sum up to marginals that are smaller than the observed ones.

1.5.2 Maximum-entropy network reconstruction

Instead of viewing network reconstruction as an ill-posed least-squares problem, it can be considered as a task that requires to find a distribution for the edges that is consistent with the given limited information. From that point of view, the *maximum-entropy* (ME) probability distribution formalism (see e.g. Squartini et al., 2018, Koller et al., 2009) provides a very useful tool and allows to express the available information with the least-informative distribution. In short, the ME approach tries to find the probability distribution with the maximum entropy among all distributions that match pre-specified expectation-constraints.

Model derivation

In the exemplary case of Table 1.2, the edge values are defined on the space $\tilde{\mathcal{Y}} = \mathbb{R}_+^N$. Therefore, the continuous *Shannon entropy functional* must be considered

$$S[f] = - \int_{\tilde{\mathcal{Y}}} f(\tilde{y}) \log(f(\tilde{y})) d\tilde{y}, \quad (23)$$

where the argument f of the functional is a continuous density function. The requirement that a density integrates to one is formalized by the constraint

$$\int_{\tilde{y}} f(\tilde{y}) d\tilde{y} = 1. \quad (24)$$

For the given network reconstruction problem, equation (21) provides $2n$ restrictions that can be used to specify the expectation-constraints

$$\mathbb{E}[g_{wdegsl}(\tilde{\mathbf{Y}})] = \tilde{y}_{+,l}, \text{ for } l = 1, \dots, 2n \quad (25)$$

with $\tilde{y}_{+,l}$ being the l th element of \tilde{y}_+ .

Then, the maximum-entropy solution (see Chapter 6 for concrete derivation) can be found by maximizing (23) with respect to f and conditional on the constraints (24) and (25). The resulting maximum-entropy distribution is an exponential family distribution

$$\hat{f}(\tilde{y}; \lambda) = \frac{\exp \left\{ - \sum_{l=1}^{2n} \lambda_l \tilde{y}_{+,l} \right\}}{\tilde{\kappa}(\lambda)},$$

with parameters $\lambda = (\lambda_1, \dots, \lambda_{2n})^T$ and normalization constant $\tilde{\kappa}(\lambda)$ satisfying restriction (24).

Estimation

The maximum likelihood principle, implies that the moment equations

$$\mathbb{E}_{\hat{\lambda}}[g_{wdegsl}(\tilde{\mathbf{Y}})] = \tilde{y}_{+,l}, \text{ for } l = 1, \dots, 2n \quad (26)$$

must be satisfied at the maximum likelihood estimator $\hat{\lambda}$. This can be achieved by employing iterative proportional fitting (IPF, Deming and Stephan, 1940, Kolaczyk, 2009). Denote k to be the iterator and define $\mathbb{E}_{\lambda^k}[\tilde{Y}_{ij}] =: \mu_{ij}^k$ with starting values $\mu_{ij}^0 = 1$ for all $i \neq j$. Then, the IPF algorithm iterates by adjusting with respect to the weighted outdegree

$$\mu_{ij}^k = \mu_{ij}^{k-1} \frac{\tilde{y}_{+,i}}{\sum_{j \in V, j \neq i} \mu_{ij}^{k-1}}, \text{ for } i \neq j,$$

and with respect to the weighted indegree

$$\mu_{ij}^{k+1} = \mu_{ij}^k \frac{\tilde{y}_{+,n+j}}{\sum_{i \in V, i \neq j} \mu_{ij}^k}, \text{ for } j \neq i.$$

After convergence, the procedure ensures that the moment equations (26) hold, i.e. the parameters λ are set implicitly in such a way that the expected values sum up to the observed weighted in- and outdegree. In the special case, where diagonal elements are allowed (with self-loops), the solution is especially easy to find and the IPF algorithm converges in two steps. The first iteration leads to

$$\mu_{ij}^1 = \frac{\tilde{y}_{+,i}}{N},$$

and the second one gives

$$\mu_{ij}^2 = \frac{\tilde{y}_{+,i}}{N} \frac{\tilde{y}_{+,n+j}}{\sum_{i=1}^n \frac{\tilde{y}_{+,i}}{N}} = \frac{\tilde{y}_{+,i} \tilde{y}_{+,n+j}}{\sum_{i=1}^n \tilde{y}_{+,i}} = \hat{\mu}_{ij}.$$

This is a variant of the gravity model (15) with $S_i = y_{+,i}$, $R_j = y_{+,n+j}$, $G^{-1} = \sum_{i=1}^n \tilde{y}_{+,i}$ and $D_{ij} = 1$.

Extensions

Combining both ideas, i.e. searching for the ME solution and the gravity model from equation (15) is the basic motivation for the model derived in Chapter 6. The proposed model parametrizes the expectation of the matrix entries as a log-linear gravity model such that

$$\mu_{ij} = S_i^{\rho_S} R_j^{\rho_R} D_{ij}^{\rho_D}.$$

This formulation extends the standard IPF model from above because it adds an exogenous dyadic factor D_{ij} and parametric weights ρ_S , ρ_R and ρ_D . If the dyadic factor D_{ij} is reasonably well associated with the unknown matrix entries, the predictive quality can be increased relative to the standard IPF solution.

1.5.3 Approaches for sparse networks

In the explanations above, it was implicitly assumed that the network under study is fully connected and N edge values are to be predicted. As long as all elements of $g_{wdegs}(\tilde{\mathbf{y}})$ are greater than zero, any maximum-entropy method will allocate strictly positive values to all edges. This is a clear drawback because most weighted networks come with less than N valued edges and in some applications, it is of vital interest to obtain good estimates for the binary network structure. However, from a theoretical point of view, it is not possible to infer the number of edges from the weighted degrees and hence, further information or assumptions are needed to overcome this lack of identifiability (Gandy and Veraart, 2017, Proposition 3.1).

Minimum density

One approach is to assume that the network is very (very) sparse. This is proposed by Anand et al. (2015) in a model, called *minimum density* with the goal of reducing the number of edges as far as possible. The model also can be understood in terms of penalized least-squares (22) with penalty

$$J(\mu) = \sum_{i \neq j} I(\mu_{ij} > 0).$$

Hence, the algorithm seeks a solution with a minimal amount of edges with the smallest deviation from the weighted in- and outdegree.

Density-calibrated approaches

Other approaches build on the assumption that the real density, or at least a good guess for the density, is known. This allows to provide edge probabilities $p_{ij}(\tilde{y}_+; \theta)$ as a function of the marginals and potentially exogenous covariates. Given a target density, the parameter θ is chosen such that

$$\frac{1}{N} \sum_{i \neq j} p_{ij}(\tilde{y}_+; \hat{\theta}) = g_{density}(\tilde{\mathbf{y}}).$$

The respective models can be constructed as two-step procedures (Cimini et al., 2015, Gandy and Veraart, 2019). In the first step, the probabilities are calibrated to the target density and in the second step, a model for the edge values is applied conditional on a sampled binary network structure. Another approach that relies on joint estimation, based on adding a density constraint to the ME problem, is proposed by Bargigli (2014). See Chapter 7 for more details regarding sparse network reconstruction and density calibration.

1.6 Lessons learned

*There are no routine statistical questions;
there are only questionable statistical routines.*

David R. Cox (found in Havránek et al., 2013, p. 250)

Statistical network data analysis provides, without doubt, a powerful framework that allows uncovering relations and dependencies that would remain hidden otherwise. From a statistical perspective, the most important aspect of the network data analysis literature is probably the explicit focus on dependent observations. This strong emphasis can be seen as a vital step towards modeling complex real-world processes more appropriately.

Alas, statistical network data analysis as a methodological toolkit is also a good example for the fact that allowing for more complexity can open many trapdoors. Regarding models for binary networks, for example the ERGM and its dynamic extensions, it becomes obvious that the appealing flexibility is bought at a high price. This includes tedious and painfully slow computation as well as laborious interpretability. Therefore, Chapter 2 tries to escape these shortcomings with the assumption of dyadic independence conditional on the past state of the network and in the survey article of Chapter 3 much effort is spent on the explanation and interpretation of different network statistics. Still the ERGM and its temporal extensions are among the most established models.

Regarding weighted network analysis, most proposals in the literature appear to be somewhat unfinished and incomplete concerning scalability and interpretation. While there exist some ambitious and carefully constructed approaches, one is still tempted to conclude that the question of how to model real-world weighted networks appropriately is unresolved so far. Regarding the network applications of Chapters 4 and 5, common models for weighted networks haven't even proved to be infeasible, forcing the authors to circumvent the inadequacy of these models by a rather unusual choice of modifying spatial econometric models for investigating edges in weighted networks.

In the case of network reconstruction (Chapters 6 and 7), neither interpretability nor feasibility is that much of an issue. However, network reconstruction has a focus on prediction rather than interpretation. Furthermore, in settings where networks are to be reconstructed, most of the complicated network dependence structure is already lost and the reconstruction depends only on simple network statistics like the weighted degrees or the density.

As a consequence, some of the contributing articles can be regarded to be “workarounds” trying to bypass the problems that come with complex networks. A less cynic description of these approaches would be that the articles show how recycling and re-arrangement of established concepts and models can be fruitfully combined with the ideas of statistical network data analysis.

Having said this, two conclusions can be drawn. First, there is certainly still a lot of room for improvement in the field of statistical network data analysis. Second, it seems like much of the answers to this challenge can be found in already established and long-standing methods scattered over the whole literature of quantitative research and statistics.

Bibliography

- AIROLDI, E. M., D. S. CHOI, AND P. J. WOLFE (2011): “Confidence sets for network structure,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4, 461–469.
- ALMQUIST, Z. W. AND C. T. BUTTS (2014): “Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics,” *Sociological Methodology*, 44, 273–321.
- ANAND, K., B. CRAIG, AND G. VON PETER (2015): “Filling in the blanks: Network structure and interbank contagion,” *Quantitative Finance*, 15, 625–636.
- BARABÁSI, A.-L. AND R. ALBERT (1999): “Emergence of scaling in random networks,” *Science*, 286, 509–512.
- BARGIGLI, L. (2014): “Statistical ensembles for economic networks,” *Journal of Statistical Physics*, 155, 810–825.
- BLOCK, P., J. KOSKINEN, J. HOLLWAY, C. STEGLICH, AND C. STADTFELD (2018): “Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power,” *Social Networks*, 52, 180–191.
- BLOCK, P., C. STADTFELD, AND T. A. SNIJDERS (2019): “Forms of dependence: Comparing SAOMs and ERGMs from basic principles,” *Sociological Methods & Research*, 48, 202–239.
- BOIVIN, R. AND M. D’ELIA (2017): “A network of neighborhoods: Predicting crime trips in a large Canadian city,” *Journal of Research in Crime and Delinquency*, 54, 824–846.
- CAIMO, A. AND N. FRIEL (2011): “Bayesian inference for exponential random graph models,” *Social Networks*, 33, 41–55.
- CASTRO, R., M. COATES, G. LIANG, R. NOWAK, AND B. YU (2004): “Network tomography: Recent developments,” *Statistical Science*, 19, 499–517.
- CHANDRASEKHAR, A. G. AND M. O. JACKSON (2014): “Tractable and consistent random graph models,” Tech. rep., National Bureau of Economic Research.
- CHATTERJEE, S., P. DIACONIS, ET AL. (2013): “Estimating and understanding exponential random graph models,” *The Annals of Statistics*, 41, 2428–2461.
- CHEN, L., X. MA, G. PAN, J. JAKUBOWICZ, ET AL. (2017): “Understanding bike trip patterns leveraging bike sharing system open data,” *Frontiers of Computer Science*, 11, 38–48.
- CIMINI, G., T. SQUARTINI, D. GARLASCELLI, AND A. GABRIELLI (2015): “Systemic risk analysis on reconstructed economic and financial networks,” *Scientific Reports*, 5, 15758.
- CLAESKENS, G. AND N. L. HJORT (2008): *Model selection and model averaging*, Cambridge: Cambridge University Press.
- CSARDI, G. AND T. NEPUSZ (2006): “The igraph software package for complex network research,” *InterJournal, Complex Systems*, 1695, 1–9.

- DEMING, W. E. AND F. F. STEPHAN (1940): “On a least squares adjustment of a sampled frequency table when the expected marginal totals are known,” *The Annals of Mathematical Statistics*, 11, 427–444.
- 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 (Statistical Methodology)*, 39, 1–38.
- DESMARAIS, B. A. AND S. J. CRANMER (2012a): “Micro-level interpretation of exponential random graph models with application to estuary networks,” *Policy Studies Journal*, 40, 402–434.
- (2012b): “Statistical inference for valued-edge networks: The generalized exponential random graph model,” *PLOS ONE*, 7, 1–12.
- DISDIER, A.-C. AND K. HEAD (2008): “The puzzling persistence of the distance effect on bilateral trade,” *The Review of Economics and Statistics*, 90, 37–48.
- DOREIAN, P. (1980): “Linear models with spatially distributed data: spatial disturbances or spatial effects?” *Sociological Methods & Research*, 9, 29–60.
- DOW, M. M., M. L. BURTON, AND D. R. WHITE (1982): “Network autocorrelation: a simulation study of a foundational problem in regression and survey research,” *Social Networks*, 4, 169–200.
- DUIJN, M. A., T. A. SNIJDERS, AND B. J. ZIJLSTRA (2004): “p2: a random effects model with covariates for directed graphs,” *Statistica Neerlandica*, 58, 234–254.
- FAHRMEIR, L., T. KNEIB, S. LANG, AND B. MARX (2007): *Regression: Models, methods and applications*, Berlin, Heidelberg: Springer.
- FIENBERG, S. E. (2012): “A brief history of statistical models for network analysis and open challenges,” *Journal of Computational and Graphical Statistics*, 21, 825–839.
- FORTUNATO, S. (2010): “Community detection in graphs,” *Physics Reports*, 486, 75–174.
- FRANK, O. AND D. STRAUSS (1986): “Markov graphs,” *Journal of the American Statistical Association*, 81, 832–842.
- FRANZESE JR, R. J. AND J. C. HAYS (2007): “Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data,” *Political Analysis*, 15, 140–164.
- GANDY, A. AND L. A. VERAART (2017): “A Bayesian methodology for systemic risk assessment in financial networks,” *Management Science*, 63, 4428–4446.
- GANDY, A. AND L. A. M. VERAART (2019): “Adjustable network reconstruction with applications to CDS exposures,” *Journal of Multivariate Analysis*, 172, 193 – 209.
- GARLASCHELLI, D. AND M. I. LOFFREDO (2004): “Patterns of link reciprocity in directed networks,” *Physical Review Letters*, 93, 268701.
- 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, 657–699.

- GILBERT, E. N. (1959): “Random graphs,” *The Annals of Mathematical Statistics*, 30, 1141–1144.
- GOLDENBERG, A., A. X. ZHENG, S. E. FIENBERG, E. M. AIROLDI, ET AL. (2010): “A survey of statistical network models,” *Foundations and Trends in Machine Learning*, 2, 129–233.
- HANDCOCK, M. S., A. E. RAFTERY, AND J. M. TANTRUM (2007): “Model-based clustering for social networks,” *Journal of the Royal Statistical Society: Series A (Statistics and Society)*, 170, 301–354.
- HANNEKE, S., W. FU, E. P. XING, ET AL. (2010): “Discrete temporal models of social networks,” *Electronic Journal of Statistics*, 4, 585–605.
- HASTIE, T. AND R. TIBSHIRANI (1993): “Varying-coefficient models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 757–796.
- HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The elements of statistical learning*, New York: Springer.
- HAVRÁNEK, T., Z. SIDAK, AND M. NOVAK (2013): *COMPSTAT 1984: Proceedings in Computational Statistics*, New York: Springer.
- HAYS, J. C., A. KACHI, AND R. J. FRANZESE (2010): “A spatial model incorporating dynamic, endogenous network interdependence: A political science application,” *Statistical Methodology*, 7, 406–428.
- HEAD, K. AND T. MAYER (2014): “Gravity equations: Workhorse, toolkit, and cookbook,” in *Handbook of international economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Amsterdam: Elsevier Science Publishing, vol. 4, 131–195.
- HOFF, P. D. (2005): “Bilinear mixed-effects models for dyadic data,” *Journal of the American Statistical Association*, 100, 286–295.
- (2015): “Dyadic data analysis with amen,” *arXiv preprint arXiv:1506.08237*.
- 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, 1090–1098.
- HOLLAND, P. W., K. B. LASKEY, AND S. LEINHARDT (1983): “Stochastic blockmodels: First steps,” *Social Networks*, 5, 109–137.
- 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, 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, 565–583.
- HUNTER, D. R., P. N. KRIVITSKY, AND M. SCHWEINBERGER (2012): “Computational statistical methods for social network models,” *Journal of Computational and Graphical Statistics*, 21, 856–882.
- KAUERMANN, G., H. HAUPT, AND N. KAUFMANN (2012): “A hitchhiker’s view on spatial statistics and spatial econometrics for lattice data,” *Statistical Modelling*, 12, 419–440.

- KOLACZYK, E. D. (2009): *Statistical analysis of network data. Methods and Models*, New York: Springer.
- (2017): *Topics at the frontier of statistics and network analysis: (Re)Visiting the foundations*, Cambridge: Cambridge University Press.
- KOLLER, D., N. FRIEDMAN, AND F. BACH (2009): *Probabilistic graphical models: principles and techniques*, Cambridge: MIT Press.
- KRIVITSKY, P. N. (2012): “Exponential-family random graph models for valued networks,” *Electronic Journal of Statistics*, 6, 1100–1128.
- 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, 29–46.
- LEENDERS, R. T. A. (2002): “Modeling social influence through network autocorrelation: constructing the weight matrix,” *Social Networks*, 24, 21–47.
- LEIFELD, P. AND S. J. CRANMER (2019): “A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model,” *Network Science*, 7, 20–51.
- LEIFELD, P., S. J. CRANMER, AND B. A. DESMARAIS (2018): “Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals,” *Journal of Statistical Software*, 83, doi: 10.18637/jss.v083.i06.
- LERNER, J., N. INDLEKOFER, B. NICK, AND U. BRANDES (2013): “Conditional independence in dynamic networks,” *Journal of Mathematical Psychology*, 57, 275–283.
- LESAGE, J. AND R. K. PACE (2009): *Introduction to spatial econometrics*, Boca Raton: CRC press.
- LESAGE, J. P. AND R. K. PACE (2008): “Spatial econometric modeling of origin-destination flows,” *Journal of Regional Science*, 48, 941–967.
- LUSHER, D., J. KOSKINEN, AND G. ROBINS (2013): *Exponential random graph models for social networks: Theory, methods, and applications*, Cambridge: Cambridge University Press.
- MATIAS, C. AND V. MIELE (2017): “Statistical clustering of temporal networks through a dynamic stochastic block model,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1119–1141.
- MCLACHLAN, G. AND D. PEEL (2004): *Finite mixture models*, New York: John Wiley & Sons.
- MEINSHAUSEN, N. ET AL. (2013): “Sign-constrained least squares estimation for high-dimensional regression,” *Electronic Journal of Statistics*, 7, 1607–1631.
- METZ, F. AND K. INGOLD (2017): “Politics of the precautionary principle: assessing actors’ preferences in water protection policy,” *Policy Sciences*, 50, 721–743.
- MORRIS, M., M. S. HANDCOCK, AND D. R. HUNTER (2008): “Specification of exponential-family random graph models: terms and computational aspects,” *Journal of Statistical Software*, 24, 1548.

- NISBETT, R. E. (2015): *Mindware: Tools for smart thinking*, New York: Farrar, Straus and Giroux.
- NOWICKI, K. AND T. A. B. SNIJDERS (2001): “Estimation and prediction for stochastic blockstructures,” *Journal of the American Statistical Association*, 96, 1077–1087.
- POPPER, K. R. (1992): *The open universe: An argument for indeterminism*, London: Routledge.
- R DEVELOPMENT CORE TEAM (2008): *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
- ROBINS, G., P. PATTISON, Y. KALISH, AND D. LUSHER (2007): “An introduction to exponential random graph (p^*) models for social networks,” *Social Networks*, 29, 173–191.
- RÖTTJERS, L. AND K. FAUST (2018): “From hairballs to hypotheses—biological insights from microbial networks,” *FEMS Microbiology Reviews*, 42, 761–780.
- SARKAR, P. AND A. W. MOORE (2006): “Dynamic social network analysis using latent space models,” in *Advances in Neural Information Processing Systems*, 18, 1145–1152.
- SCHOENEMAN, J., B. ZHU, AND B. A. DESMARAIS (2017): “The Network of Foreign Direct Investment Flows: Theory and Empirical Analysis,” *SSRN*, [dx.doi.org/10.2139/ssrn.3018031](https://doi.org/10.2139/ssrn.3018031).
- SCHWEINBERGER, M. (2011): “Instability, sensitivity, and degeneracy of discrete exponential families,” *Journal of the American Statistical Association*, 106, 1361–1370.
- SEWELL, D. K. AND Y. CHEN (2015): “Latent space models for dynamic networks,” *Journal of the American Statistical Association*, 110, 1646–1657.
- SILK, M. J., D. P. CROFT, R. J. DELAHAY, D. J. HODGSON, N. WEBER, M. BOOTS, AND R. A. McDONALD (2017): “The application of statistical network models in disease research,” *Methods in Ecology and Evolution*, 8, 1026–1041.
- SIMPSON, S. L., F. D. BOWMAN, AND P. J. LAURIENTI (2013): “Analyzing complex functional brain networks: fusing statistics and network science to understand the brain,” *Statistics Surveys*, 7, 1–36.
- SIMPSON, S. L. AND P. J. LAURIENTI (2015): “A two-part mixed-effects modeling framework for analyzing whole-brain network data,” *NeuroImage*, 113, 310–319.
- SNIJDERS, T. A. (2017): “Stochastic actor-oriented models for network dynamics,” *Annual Review of Statistics and Its Application*, 4, 343–363.
- SNIJDERS, T. A., P. E. PATTISON, G. L. ROBINS, AND M. S. HANDCOCK (2006): “New specifications for exponential random graph models,” *Sociological Methodology*, 36, 99–153.
- SNIJDERS, T. A., G. G. VAN DE BUNT, AND C. E. STEGLICH (2010): “Introduction to stochastic actor-based models for network dynamics,” *Social Networks*, 32, 44–60.
- SQUARTINI, T., G. CALDARELLI, G. CIMINI, A. GABRIELLI, AND D. GARLASCHELLI (2018): “Reconstruction methods for networks: The case of economic and financial systems,” *Physics Reports*, 757, 1 – 47.
- STEWART, J. Q. (1941): “An inverse distance variation for certain social influences,” *Science*, 93, 89–90.

- STRAUSS, D. AND M. IKEDA (1990): “Pseudolikelihood estimation for social networks,” *Journal of the American Statistical Association*, 85, 204–212.
- SWEET, T. M. (2015): “Incorporating covariates into stochastic blockmodels,” *Journal of Educational and Behavioral Statistics*, 40, 635–664.
- THIEMICHEN, S., N. FRIEL, A. CAIMO, AND G. KAUEMANN (2016): “Bayesian exponential random graph models with nodal random effects,” *Social Networks*, 46, 11–28.
- TINBERGEN, J. (1963): “Shaping the world economy,” *Thunderbird International Business Review*, 5, 27–30.
- VARDI, Y. (1996): “Network tomography: Estimating source-destination traffic intensities from link data,” *Journal of the American Statistical Association*, 91, 365–377.
- WANG, Y. J. AND G. Y. WONG (1987): “Stochastic blockmodels for directed graphs,” *Journal of the American Statistical Association*, 82, 8–19.
- WARD, M. D., J. S. AHLQUIST, AND A. ROZENAS (2013): “Gravity’s rainbow: A dynamic latent space model for the world trade network,” *Network Science*, 1, 95–118.
- WASSERMAN, S. AND P. PATTISON (1996): “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* ,” *Psychometrika*, 61, 401–425.
- WEI, G. C. AND M. A. TANNER (1990): “A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms,” *Journal of the American Statistical Association*, 85, 699–704.
- WILSON, J. D., M. J. DENNY, S. BHAMIDI, S. J. CRANMER, AND B. A. DESMARAIS (2017): “Stochastic weighted graphs: Flexible model specification and simulation,” *Social Networks*, 49, 37 – 47.
- XU, K. S. AND A. O. HERO (2013): “Dynamic stochastic blockmodels: Statistical models for time-evolving networks,” in *International conference on social computing, behavioral-cultural modeling, and prediction*, Springer, 201–210.
- YANG, T., Y. CHI, S. ZHU, Y. GONG, AND R. JIN (2011): “Detecting communities and their evolutions in dynamic social networks—a Bayesian approach,” *Machine Learning*, 82, 157–189.
- YON, G. G. V. AND K. DE LA HAYE (2019): “Exponential random graph models for little networks,” *arXiv preprint arXiv:1904.10406*.
- ZHANG, Y., M. ROUGHAN, C. LUND, AND D. DONOHO (2003): “An information-theoretic approach to traffic matrix estimation,” in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, Association for Computing Machinery, 301–312.

Chapter 2

A dynamic separable model with actor heterogeneity: An application to global weapons transfers

Contributing Article:

Michael Lebacher, Paul W. Thurner and Göran Kauermann (2019): *A dynamic separable model with actor heterogeneity: An application to global weapons transfers*.

Under review in the Journal of the Royal Statistical Society: Series A (Statistics in Society).

arXiv preprint <https://arxiv.org/abs/1803.02707>

Code at https://github.com/lebacher/dynamic_separable_network_model

Further Versions:

Michael Lebacher, Paul W. Thurner and Göran Kauermann (2018): *International arms transfers: A dynamic separable network model with heterogeneity components*.

Proceedings of the 33th International Workshop on Statistical Modelling, 1:163–168.

Author Contributions:

The general idea of extending the STERGM towards time-varying effects and smooth random effects stems from Göran Kauermann. Paul Thurner has provided the political and economic theory and valuable domain knowledge needed to legitimize, specify and interpret the models and results. The contribution of Michael Lebacher is given by the implementation of the model in R, including data manipulation, fitting and simulation. Furthermore, Michael Lebacher wrote the major part of the manuscript related to statistical matters and developed the idea of using a functional principal component analysis to analyze the fitted smooth random effects with advice and input from Göran Kauermann. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

A Dynamic Separable Network Model with Actor Heterogeneity: An Application to Global Weapons Transfers

Michael Lebacher^{*}, Paul W. Thurner[†] and Göran Kauermann^{‡§}

Abstract

In this paper we propose to extend the separable temporal exponential random graph model (STERGM) to account for time-varying network- and actor-specific effects. Our application case is the network of international major conventional weapons transfers, based on data from the Stockholm International Peace Research Institute (SIPRI). The application is particularly suitable since it allows to distinguish the potentially differing driving forces for creating new trade relationships and for the endurance of existing ones. In accordance with political economy models we expect security- and network-related covariates to be most important for the formation of transfers, whereas repeated transfers should prevalently be determined by the receivers' market size and military spending. Our proposed modelling approach corroborates the hypothesis and quantifies the corresponding effects. Additionally, we subject the time-varying heterogeneity effects to a functional principal component analysis. This serves as exploratory tool and allows to identify countries that stand out by exceptional increases or decreases of their tendency to import and export weapons.

Keywords: Arms Transfers, Functional Principal Component Analysis, Generalized Additive Mixed Model, Security and Defence Network, Varying Coefficient Model

^{*}Department of Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Germany, michael.lebacher@stat.uni-muenchen.de

[†]Department of Political Science, Ludwig-Maximilians-Universität München, 80538 Munich, Germany, paul.thurner@gsi.uni-muenchen.de

[‡]Department of Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Germany, goeran.kauermann@stat.uni-muenchen.de

[§]The authors gratefully acknowledge funding provided by the german research foundation (DFG) for the project *International Trade of Arms: A Network Approach*.

1. Introduction

In this paper we present a data-driven extension of the separable temporal exponential random graph model (STERGM, Krivitsky and Handcock 2014) applied appropriately to a highly relevant case: The international weapons exchange. The STERGM allows to differentiate between the *formation*, i.e. new arms trades, and the *persistence* of existing edges, i.e. continued arms transfers. To introduce into the field, we first sketch and motivate network analysis for (arms) trade data. We then put the model in a broader context of statistical network models, supplemented by a description and discussion of international arms trade.

Trade networks

Statistical network analysis provides a good framework to conceptualize international trade systems. Schweitzer et al. (2009) highlight the enormous interdependencies of economic transactions and propose a network approach for capturing the systemic complexity. Gravity models, as standard approach in econometrics for modelling trade data (Head and Mayer 2014), are usually focussed on dyadic relations. Hence, the models exclude highly important hyper-dyadic dependencies, and especially indirect relations. Squartini et al. (2011a,b) showed that gravity models of international trade are, therefore, necessarily incomplete. In particular, they demonstrated that analysing the determinants of link creation is highly important as the binary network carries information that goes beyond the classical gravity model representation. Barigozzi et al. (2010) demonstrated that trade networks are commodity-specific, i.e. their topologies are quite different across commodities - leading us to conclude that there is also a need to consider arms transfers separately. This is theoretically challenging since arms transfers constitute a very special trade relationship. The transferred products and services can potentially lead to deadly quarrels between or within states, or they may contribute to stabilization and deterrence. The delivery is not always a purely economic exchange but may also serve the support of aligned countries or groups. In sum, the exchange of weapons is a politically sensible and security-related, but also an economically beneficial relationship. For this reason, we make use of flexible statistical models for network data that allow us to investigate the special incentives in the international arms trade network.

Statistical network models

Statistical models that are suitable for temporal networks have been developed just in the recent years, and different techniques have been proposed. Robins and Pattison (2001) were the first to extend the static exponential random graph model (ERGM, Holland and Leinhardt 1981; Lusher et al. 2012) to discrete-time Markov chain models, see also Snijders et al. (2010). Hanneke et al. (2010) or Leifeld et al. (2018) also consider network dynamics on a discrete time scale. They propose the temporal exponential random graph model (TERGM) which makes use of a Markov structure conditioning on previous network statistics as covariates in the model. A related approach is presented by Almquist and Butts (2014),

discussing assumptions that allow for circumventing the often computationally intractable fitting process of dynamic network models by applying logistic regression models. Koskinen et al. (2015) expand the model using Bayesian methods which allows the parameters in the dynamic network model to change with time. A general perspective on dynamic networks is provided by Holme (2015). It also includes models for continuous time, such as stochastic actor-oriented models (SAOM, see Snijders et al. 2010) or dynamic stochastic block models (SBM, see for instance Xu 2015).

A recent novel modelling strategy for networks observed at discrete time points has been proposed by Krivitsky and Handcock (2014). They do not model the state of the network itself but rather focus on network changes which either occur because of the formation of new edges or because of the (non-)persistence of existing ones. Assuming independence between the two processes, conditional on the previous network, leads to the so called *separable* TERGM. The separation is motivated by the fact that the two processes under study are highly likely to be driven by different mechanisms and factors. The authors argue that the inclusion of a stability term (being mathematically equivalent to the inclusion of the lagged edge values as explanatory variable) in a TERGM could lead to ambiguous conclusions because it is not clear whether a positive stability parameter means that non-existing ties remain non-existent (no formation) or whether existent ties remain existent (persistence).

For many real world dynamic networks the process change with time and therefore the assumption of stationarity seems to be inappropriate. This is especially the case for network data that span a long time period and potentially subject to structural breaks. Under such conditions it appears necessary to allow the model parameters to change with time. We take up this idea and extend the STERGM by allowing for time-varying coefficients. More specifically, we propose to rely on so called generalized additive models (GAM). This model class has been proposed by Hastie and Tibshirani (1987) and extended fundamentally by Wood (2017) to allow for smooth, semi-parametric modelling of time-varying parameters in a generalized regression framework (see also Ruppert et al. 2009).

Furthermore, the assumption of node homogeneity must be regarded as questionable. We therefore allow for heterogeneity in the model (see Thiemichen et al. 2016 for a discussion on node heterogeneity). Accordingly, we follow the p_2 -model developed by Duijn et al. (2004) and enrich the STERGM with functional time-varying random effects (Durbán et al. 2005) which leads to smooth node-specific effects. We propose to investigate the fitted functional heterogeneity effects with techniques from functional data analysis (FDA), see for instance Ramsay and Silverman (2005). This allows to identify countries (nodes) that have fundamentally changed their role in the arms-trading network over the observation period.

Global weapons transfers

At present, there are only a few empirical binary network analyses of the international arms trade. Akerman and Seim (2014) pioneered in analysing topological features of the binary arms trade network. Their descriptive network analysis is supplemented by an empirical investigation using a binarized gravity model without considering network dependencies. In

this article we build on the recently published paper by Thurner et al. (2018) that uses a TERGM. However, our approach extends the TERGM in many aspects. Most importantly, we treat dynamic dependencies in a fundamentally different way. In Thurner et al. (2018), the authors found that previous arms trading has a highly determining impact on the occurrence of subsequent transfers due to the enormous inertia. This finding implies that the information whether trade happened in the preceding time period(s) has a considerable impact on the probability to trade again, leading to the same ambiguities as mentioned in the stability term example by Krivitsky and Handcock (2014). In order to disentangle the driving network formation forces due to pure inertia, we propose to incorporate this distinction directly in the model. More precisely, the STERGM allows us to investigate whether the mechanisms that result in transfers being formed without immediate predecessor differ from those that lead to consecutive transfers. This is also of practical importance because governments carefully reflect the decision whether to authorize arms transfers based on economic and security considerations. Furthermore, they continuously reconsider this decision whether to maintain such trade relations or whether to dissolve because the importer potentially jeopardises strategic interests or violates once shared normative standards (see Garcia-Alonso and Levine 2007 for the general model and for Blanton 2005 as well as Erickson 2015 for normative considerations).

We expect several necessary conditions to hold for the formation of transfers: the receiving country must be considered at least marginally trustworthy and politically and economically reliable. Hence, passing a threshold of trustworthiness is required for formation, i.e building new trades. The special role of trustworthiness in arms transfers stems from the fact that security concerns play an important role when governments decide whether to license the delivery. We expect network statistics, as well as regime dissimilarity and formal alliances to play a prominent role in the formation stage to raise a relationship above the minimum threshold level of reservation. Follow-up trades and their repetition should then be rather dominated by economic considerations like the size of a receiver economy and by the size of the military expenditures (see Schulze et al. 2017).

While differentiation between formation and repetition, respectively, legitimates the use of the STERGM per se, our extensions of the model towards time-varying coefficients are important and in our view inevitable because the observational time covers more than 65 years. Hence, the introduction of smooth dynamic effects is needed to build a realistic model. Given the dynamic evolution of the network, the historical developments and the presence of at least one system-wide structural break with the collapse of the Soviet Union, we expect that the generative mechanisms change over time and differ with respect to the included variables if we compare the pre- and post cold war time period (see also Akerman and Seim 2014 and Thurner et al. 2018).

Finally, we argue that not all network activities and trades can be explained by observables and, thus, unobserved heterogeneity remains. We expect primarily actor-specific heterogeneity which is accentuated by systematic historical accounts (Harkavy 1975; Krause

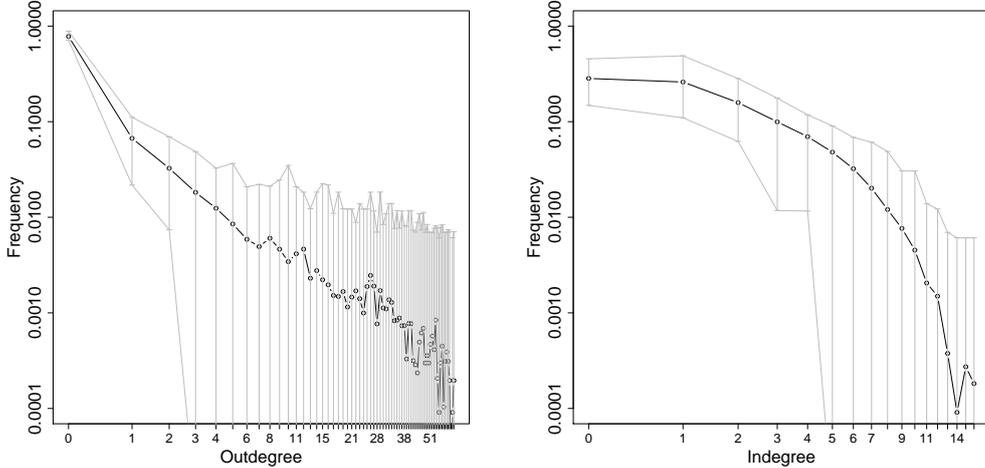


Figure 1: Degree distributions of the included countries for the outdegree (number of outgoing edges) on the left and indegree (number of ingoing edges) on the right. Averages over all years are represented by the solid line. The whiskers in grey show the minimum and maximum Values realized in all years. Both axes are in logarithmic scale.

1995). This highlights the self-reinforcing tendencies of technological advantages of highly developed countries which results in strong heterogeneity of the countries' abilities to export (and import). Therefore, the inclusion of actor-specific random effects seems necessary and we expect strong heterogeneity among the countries with respect to imports and exports.

We proceed as follows. Section 2 presents the data provided by the Stockholm International Peace Research Institute (SIPRI). Section 3 introduces the statistical models used to analyse the data. Section 4 provides the results and their interpretation. Section 5 concludes the paper.

2. Data description and preprocessing

Data on the international trade of major conventional weapons (MCW) are provided by the Stockholm International Peace Research Institute (see SIPRI 2017a). They include for example aircrafts, armoured vehicles and ships (see Table 1 in the Appendix A.1 for an overview of the types of arms). The countries included and their three-digit country codes are given in Table 2 of Appendix A.1. Note that we have excluded all non-state organizations like the Khmer Rouge or the Lebanon Palestinian Rebels from the dataset as well as countries with no reliable covariate information available.

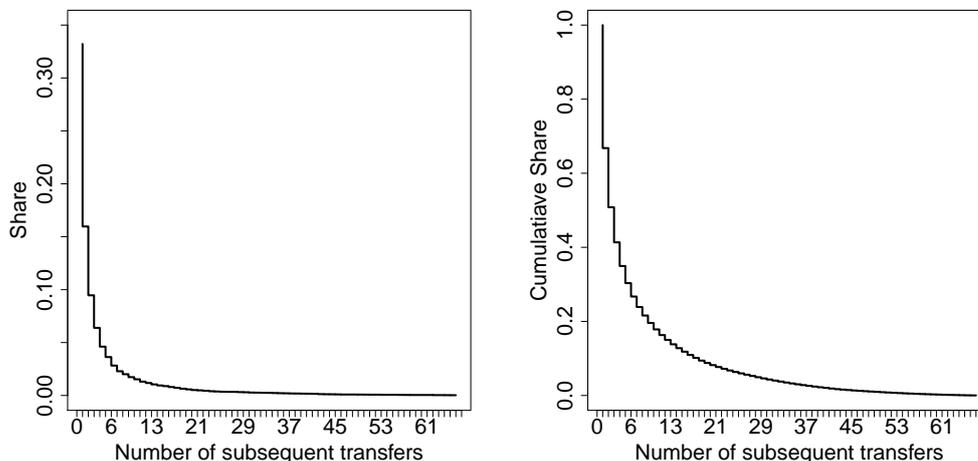


Figure 2: Share of subsequent arms transfers (left) and cumulative share of subsequent arms transfers (right). Number of subsequent transfers on the horizontal axis and share of observations on the vertical axis.

Figure 9 in the Appendix A.1 shows binary networks for the years 2015 and 2016 and Figure 10 in the Appendix A.1 provides a collection of summary statistics for the networks.

We focus on the binary occurrence of trade thereby disregarding the exact transfer volumes and follow Akerman and Seim (2014) and Thurner et al. (2018) in setting the edge value to one if there is a trade flow greater zero between two countries and zero else. Additionally, we re-estimated our model with different thresholds and found that the results are quite robust, for details see the Supplementary Material.

The analysis of the degree distributions is of vital interest in statistical network analysis (Barabási and Albert 1999) and gives important insights into the basic properties of the network under study. With more than 65 networks to analyse, we compute the period-average degree distribution and provide information on the minimal and maximal value of the realized degree distribution. This is represented in a log-log version in Figure 1 for both, outdegree and indegree. The plot shows the enormous heterogeneity in the networks. Most of the countries have no exports at all with a time-average share of 78% of countries exhibiting outdegree zero, while the outdegree distribution has a long tail, indicating that there are a few countries, having a very high outdegree. The highest observed outdegree in a year is 66 and is observed for the United States. Other countries with exceptional high outdegree for almost the whole time period are Russia (Soviet Union), France, Germany, United Kingdom, China, Italy and Canada. In the right plot, the indegree distribution can be seen. Here the pattern is different. The highest value observed in a year is 16 and corresponds to Saudi

Arabia. In contrast to the outdegree distribution, the countries with a high indegree are changing with time. In the beginning of the observational period the countries with the highest indegree were Germany, Indonesia, Italy, Turkey and Australia, but in more recent times these are the United Arab Emirates, Saudi Arabia, Singapore, Thailand and Oman.

In Figure 2 we provide a graphical representation of the change and stability patterns in the network. On the left hand side we present the share of observations (vertical axis) against the number of subsequent transfers (i.e. repeated transfers) on the horizontal axis. Out of roughly 19,000 recorded trading instances only 33% do not have at least one consecutive transfer in the follow-up year of a trade. Looking on the right hand side of Figure 2 we visualize the share of observations (vertical axis) that has at least as much subsequent transfers as indicated by the horizontal axis. It can be seen that roughly the same share of observations (35%) lasts at least five periods and almost 10% of all dyadic relations last more than 20 consecutive years without any interruption. Therefore, a differentiated approach to the explanation of formation and persistence could be fruitful in this application case.

3. Model

3.1. Dynamic formation and Persistence model

In this section we formalize our network model. Let Y^t be the network at time point t , which consists of a set of actors, labelled as A^t and a set of directed edges, represented through the index set $E^t = \{(i, j) : i, j \in A^t\}$. Note that this is a slight misuse of index notation since Y_{ij}^t does not necessarily refer to the (i, j) -th element if we consider Y^t as adjacency matrix. This is because the actor set A^t is allowed to change with time, so that i and j are not running indices from 1 to n_t , where n_t is the number of elements in A^t . Instead indices i and j represent the i -th and j -th country, respectively. We define $Y_{ij}^t = 1$ if country i exports weapons to country j and since self-loops are meaningless, elements Y_{ii}^t are not defined.

We aim to model the network in t based on the previous year network in $t - 1$. To do so we have to take into account that the actor sets A^{t-1} and A^t may differ. In particular we have to consider the case of newly formed countries. New countries of interest are those that are present in t but do not provide information about their network embedding in the previous period. For exports this is not a concern as it is almost never the case that a new country starts sending arms immediately after entering the network. Notable exceptions are Russia, the Czech Republic and Slovakia. However, these countries have clear defined predecessor states (the Soviet Union and Czechoslovakia) which can be used in order to gain information about the position of these countries in the precedent network. Regarding the imports, there is a share of countries that start receiving arms immediately with entering the network. Notwithstanding, those transactions represent a share of less than 0.3% of the observed trade flows. Therefore, we regard these cases as negligible and include in the model only countries

where information on the current and previous time period is available. We formalize this approach by defining $Y^{t,t-1}$ as the subgraph of Y^t with actor set $B^{t,t-1} = A^t \cap A^{t-1}$ containing $n_{t,t-1} := |B^{t,t-1}|$ elements. Accordingly, $Y^{t-1,t}$ represents the subgraph of Y^{t-1} with actor set $B^{t,t-1}$. Note that both subgraphs share the same set of actors and $Y^{t-1} = Y^{t-1,t}$ if A^{t-1} and A^t coincide.

From a modelling perspective, we follow Hanneke et al. (2010) and assume that the network in t can be modelled given preceding networks, using a first-order Markov structure to describe transition dynamics for those actors included in the set $B^{t,t-1}$. Furthermore, we want to identify the driving forces of a transfer in t if there was a preceding transfer in $t-1$ in the persistence model while the formation model considers the process of forming a trade relationship without a preceding transfer, i.e. biannual data. The notion of formation and persistence can be amended by using broader time windows. We demonstrate the robustness of our results with respect to broader time windows in the Supplementary Material.

Let $Y^+ = Y^{t,t-1} \cup Y^{t-1,t}$ represent the formation network, that consists of edges that are either present in t or in $t-1$. For the persistence network, we define $Y^- = Y^{t,t-1} \cap Y^{t-1,t}$, being the network that consists of edges that are present in t and in $t-1$. Based on the actor set $B^{t,t-1}$ and given the formation and persistence network as well as the network in $t-1$ the network in t is uniquely defined by

$$Y^{t,t-1} = Y^+ \setminus (Y^{t-1,t} \setminus Y^-) = Y^- \cup (Y^+ \setminus Y^{t-1,t}). \quad (1)$$

Note that both, Y^+ as well as Y^- depend on time t as well, which we omitted in the notation for ease of readability. We assume that for each discrete time step, the processes of formation and persistence are separable. That is, the process that drives the formation of edges does not interact with the process of the persistence of the edges conditional on the previous network. Formally this is given by the conditional independence of Y^+ and Y^- :

$$\begin{aligned} P(Y^{t,t-1} = y^{t,t-1} | Y^{t-1,t} = y^{t-1,t}; \theta) = \\ P(Y^+ = y^+ | Y^{t-1,t} = y^{t-1,t}; \theta^+) P(Y^- = y^- | Y^{t-1,t} = y^{t-1,t}; \theta^-), \end{aligned}$$

where the lower case letters denote the realizations of the random networks and $\theta = (\theta^+, \theta^-)$ gives the parameters of the model. We will also include non-network related covariates in our analysis, but we suppress this here in the notation for simplicity.

Note that it is not possible to use the lagged response as predictor, as by construction $Y_{ij}^{t-1,t} = 1 \Rightarrow Y_{ij}^+ = 1$ and $Y_{ij}^{t-1,t} = 0 \Rightarrow Y_{ij}^- = 0$. That is, an edge that existed in $t-1$ cannot be formed newly and an edge that was not existent in $t-1$ cannot be dissolved. It follows that the formation model exclusively focuses on the binary variables Y_{ij}^+ with $(i, j) \in E^+ = \{(i, j) : i, j \in B^{t,t-1}, Y_{ij}^{t-1,t} = 0\}$. This assures that in $t-1$ no edge between actors i and j was present and both actors are observable at both time points. Equivalently, the model for Y^- consists of observations Y_{ij}^- with $(i, j) \in E^- = \{(i, j) : i, j \in B^{t,t-1}, Y_{ij}^{t-1,t} = 1\}$, assuring

that only edges that could potentially persist enter the model. The time-dependence of E^+ and E^- is omitted for ease of readability.

If we use an ERGM for the transition, this would yield the following probability model for the formation

$$P(Y^+ = y^+ | Y^{t-1,t} = y^{t-1,t}; \theta^+) = \frac{\exp\{\theta^+ g(y^+, y^{t-1,t})\}}{\sum_{\tilde{y}^+ \in \mathcal{Y}^+(y^{t-1,t})} \exp\{\theta^+ g(\tilde{y}^+, y^{t-1,t})\}}.$$

The sum in the denominator is over all possible formation networks from the set of potential edges that can form given the network $y^{t-1,t}$. The inner product $\theta^+ g(y^+, y^{t-1,t})$, relates a vector of statistics $g(\cdot)$ to the parameter vector θ^+ . The analogous model is assumed for the persistence of edges and not explicitly given here for the interest of space.

We will subsequently work with a simplified model which is computationally much more tractable. We assume that the formation or persistence of an edge at time point t does solely depend on the past state but not on the current state of the network. This is achieved by restricting the statistics such, that they decompose to

$$g(y^+, y^{t-1,t}) = \sum_{(i,j) \in E^+} y_{ij}^+ \tilde{g}_{ij}(y^{t-1,t})$$

for some statistics $\tilde{g}(\cdot)$. This assumption is extensively discussed by Almquist and Butts (2014) and can be well justified by the notion that the lagged network accounts for the major share of the dependency among the edges in the current network. It also allows for intuitive interpretations as can be seen as follows. Let Y_{-ij}^+ represent the formation network Y^+ , excluding the entry Y_{ij}^+ . Then, for $(i, j) \in E^+$ the following logistic model holds

$$\begin{aligned} \log \left\{ \frac{P(Y_{ij}^+ = 1 | Y_{-ij}^+ = y_{-ij}^+, Y^{t-1,t} = y^{t-1,t}; \theta^+)}{P(Y_{ij}^+ = 0 | Y_{-ij}^+ = y_{-ij}^+, Y^{t-1,t} = y^{t-1,t}; \theta^+)} \right\} &= \log \left\{ \frac{P(Y_{ij}^+ = 1 | Y^{t-1,t} = y^{t-1,t}; \theta^+)}{P(Y_{ij}^+ = 0 | Y^{t-1,t} = y^{t-1,t}; \theta^+)} \right\} \\ &= \theta^+ \tilde{g}_{ij}(y^{t-1,t}). \end{aligned} \tag{2}$$

Note that model (2) describes network dynamics, but the model itself is static. Hence we model dynamics but do not allow for dynamics in the model itself. This is a very implausible restriction which we give up by allowing the model parameters to change with time t , that is we replace the parameter θ^+ by $\theta^+(t)$, representing a smooth function in time. In other words, we allow the parameters in the model to smoothly interact with time. This leads to a time-varying coefficient model in the style as proposed by Hastie and Tibshirani (1993). The focus of interest is therefore not only on the formation and persistence of edges (trade flows) but also on how these effects change in the 67 years long observation period.

3.2. Network statistics and explanatory variables

From a statistical point of view, network statistics are required in order to capture network dependencies. However, as social network literature has shown, network statistics usually are not just statistical controls but convey substantial meaning (see e.g. Snijders 2011). In the given context, they can be motivated by political, strategic and economic arguments that refer to real-world processes (see Thurner et al. 2018). Note, that we norm all network statistics (with the exception of *Reciprocity*) to be within a percentage range between 0 and 100, this is necessary in order to make the statistics independent from the varying network size and allows to compare them.

Outdegree: The outdegree of a node is a standard statistic in network models. Formally, the outdegree of actor i at time point $t - 1$ is defined as

$$outdeg_{t-1,i} = \frac{100}{n_{t,t-1} - 1} \sum_{k \in B^{t,t-1}} y_{ik}^{t-1,t}.$$

The arms trade network exhibits a highly oligopolistic structure with a few high-intensity traders, hence a positive coefficient for the outdegree of the sender ($sender.outdeg_{t-1,i}$) is plausible. However, we incorporate country-specific random effects in the model and it is therefore not clear whether the senders' outdegree as a global measure is still of relevance once controlled for the random country heterogeneity.

Only few advanced countries within NATO export and import at the same time. They have a highly differentiated portfolio, rendering specialization economically reasonable and strategically non-hazardous. In order to better represent this world-wide asymmetry we include the outdegree of the importer ($receiver.outdeg_{t-1,j}$). This should not be captured by the random effects and we expect a clear negative effect, indicating that strong exporters seldom match with strong importers.

Reciprocity: This statistic is intended to detect whether there is a general tendency of arms transfers to be mutual. The statistic measures whether the potential receiver was a sender in the dyadic relationship in the previous period:

$$recip_{t-1,ij} = y_{ji}^{t-1,t}.$$

Reciprocation is an essential mechanism in human relations in general, and in trade more specifically. Similar as noted above, in the context of arms transfers, especially highly developed countries exhibit this feature. Since this group of countries is rather small, and specialization-induced transfers between developed countries do not lead to continuous inflows we expect this mechanism to be rather visible at the formation stage, whereas it should not be a dominant feature for permanent repetition.

Transitivity: Hyperdyadic trade relationships are an effective mechanism for pooling risks in buyer-seller networks (Bramoullé et al. 2019) and for the emergence of generalized trust

which is especially important in exchanging security goods. As a measure for higher-order dependencies we include transitivity, defined as

$$trans_{t-1,ij} = \frac{100}{n_{t,t-1} - 2} \sum_{k \in B^{t,t-1}, k \neq i,j} y_{ik}^{t-1,t} y_{kj}^{t-1,t}.$$

This statistic essentially counts the directed two-paths from i to j in $t - 1$ and can be interpreted as a direct application of the *Friend of a Friend* logic from social networks to arms trade. Clearly, this kind of network embeddedness of weapons transfer deals is important for establishing for new ones but is also likely to be relevant for the continuation of already existing ones.

Shared Suppliers: We also include a statistic that we call shared-suppliers in this context. This statistic counts the shared number of actors that export to a given pair of countries:

$$sup_{t-1,ij} = \frac{100}{n_{t,t-1} - 2} \sum_{k \in B^{t,t-1}, k \neq i,j} y_{ki}^{t-1,t} y_{kj}^{t-1,t}.$$

This statistic allows to investigate whether two countries that share multiple suppliers have the tendency to engage in trade with each other. Such a pattern is likely to be induced by a general hierarchy in the network (see Krause 1995). While the first tier consists of strong exporters, the second tier is populated by countries with the ability to produce and export that are nevertheless mainly supplied by the big exporters. Countries with many shared partners are likely to engage in trade with each other but on the other hand they are typically dependent on imports from the first tier. Therefore, relationships among those countries are rather of a sporadic nature and unlikely to endure. Consequently, we expect a positive coefficient in the formation model and a negative one in the persistence model.

Naturally, the network of international arms trade is not exclusively driven by endogenous network processes but also influenced by variables from the realms of politics and economics. We lag all exogenous covariates by one year, first in order to be consistent with the idea that the determination of the network in t is based on the preceding time period and second, to account for the time lag between the ordering and the delivery of MCW.

Formal Alliance: We regard dyadic formal alliances (including defence agreements and non-aggression pacts) as an important security related criteria that plays a central role for the formation during the cold war period. Therefore, the binary variable $alliance_{ij}$ is included in the model, being one if countries i and j had a formal alliance in the previous period. Given the restriction that the data is available only until 2012 (Correlates of War Project 2017a) we extrapolate the data, thereby assuming that the formal alliances did not change between 2012 and 2015.

Regime Dissimilarity: Another important security related variable that potentially acts on the formation of arms transfers is given by the differences in political regimes between two

potential trading partners. Hence we include the so called polity IV score, ranging from the spectrum -10 (hereditary monarchy) to $+10$ (consolidated democracy). This data can be downloaded as annual cross-national time-series until 2015, see Center for systemic Peace (2017) for the data and Marshall (2017) as a basic reference. In our model we operationalise the distance between political regimes by using the absolute differences between the scores: $poldiff_{ij} = |polity_i - polity_j|$.

GDP: Following the standard gravity model, we include market sizes and distance in our model. The standard measure for market size is the gross domestic product (GDP, in millions). We include the GDP in logarithmic form for the sender (gdp_i) as well as the receiver (gdp_j). The GDP data are taken from Gleditsch (2013b) and merged from the year 2010 on with recent real GDP data from the World Bank real GDP dataset (World Bank 2017). Clearly, the market size and economic reliability of the exporter is a prerequisite for forming and maintaining arms exports.

Distance: For gravity models applied to trade in commercial goods, there exists mounting empirical evidence that distance is a relevant factor for determining trade relations (Disdier and Head 2008). We do not expect that trade costs and geographical distance impede arms trade because arms transfers establish world-wide alignments of exporters pursuing global strategic interests. Nevertheless, we include the logarithmic distance between capital cities in kilometres (Gleditsch 2013a) in order to fulfil the gravity model specification.

Military Expenditures: We propose to include military expenditures of the sending and receiving country. This measure can be used as representing the size of the defence industrial base of the exporter, and the spending power and the intensity of the threat perceptions of the importing country. Accordingly, military expenditure is added separately for the exporter and the importer in logarithmic form ($milex_i$, $milex_j$). With regard to the distinction between formation and persistence, our expectation is related to the hypothesis that countries with high military expenditures are attractive customers for repeated importing. We therefore expect a positive and high coefficient for the military expenditures of the importer in the persistence model. The data are available from Correlates of War Project (2017b) in the national material capabilities data set with Singer et al. (1972) as the basic reference on the data.

3.3. Modelling heterogeneity

The proposed network model assumes homogeneity, meaning that all differences between nodes in the network are fully described by the gravity model, enriched by security related criteria and network statistics as proposed above. However, the arms transfer network exhibits a rather small number of countries that are high-intensity exporters and a large number of countries that are restricted to imports. Furthermore, there are some countries that change their relative position in the trade network during the course of time. This mirrors a substantial amount of dynamic heterogeneity which need to be taken into account.

This dynamic heterogeneity is accommodated by the inclusion of latent country effects, capturing the unobserved heterogeneity. We follow the idea of Durbán et al. (2005) and model country specific random curves which are fitted with penalized splines. This can be written in a mixed model representation such that the smooth country-specific effects are constructed using a B-spline basis with (a-priori) normally distributed spline coefficients. We follow the modelling strategy of Durban and Aguilera-Morillo (2017) and assume that the model includes two time-dependent random coefficients $\phi_{i, \text{sender}}^+(t)$ and $\phi_{j, \text{receiver}}^+(t)$. The effects are assumed to be a realization of a stochastic process with continuous and integrable functions. For each sender and receiver in both models the country-specific curves are given by

$$\phi_i(t) = B(t)a_i \quad (3)$$

where $B(t) = (B_1(t), \dots, B_Q(t))$ is a B-spline basis covering the time range of observations and $a_i = (a_{i1}, \dots, a_{iQ})$ is the coefficient vector. We impose the prior distribution

$$a_i \sim N(0, \sigma_a^2 D_Q), \text{ i.i.d. for } i = 1, \dots, n$$

where D_Q is the inverse of a difference based penalty matrix which guarantees smoothness of the fitted curves $\phi_i(t)$ (see e.g. Eilers and Marx 1996, for details on smoothing with B-splines). Note that for time windows where a country did not exist, the corresponding B-spline does take value zero, so that no heterogeneity effect is present.

3.4. Complete model and estimation

Putting all the above elements together, the specification of the formation model of equation (2) is given by

$$\begin{aligned} \theta^+(t) \tilde{g}_{ij}(y^{t-1,t}, x_{ij}^{t-1,t}) = & \theta_0^+ \text{sender.outdeg}_{t-1,i} \theta_1^+(t) + \text{receiver.outdeg}_{t-1,j} \theta_2^+(t) \\ & + \text{recip}_{t-1,ij} \theta_3^+(t) + \text{trans}_{t-1,ij} \theta_4^+(t) + \text{sup}_{t-1,ij} \theta_5^+(t) \\ & + \text{distance}_{t-1,ij} \theta_6^+(t) + \text{alliance}_{t-1,ij} \theta_7^+(t) + \text{poldif}_{t-1,ij} \theta_8^+(t) \\ & + \text{gdp}_{t-1,i} \theta_9^+(t) + \text{gdp}_{t-1,j} \theta_{10}^+(t) + \text{miles}_{t-1,i} \theta_{11}^+(t) + \text{miles}_{t-1,j} \theta_{12}^+(t) \\ & + \phi_{i, \text{sender}}^+(t) + \phi_{j, \text{receiver}}^+(t). \end{aligned}$$

Analogously we get the persistence model. Estimation is carried out with spline smoothing. That is, we replace the coefficients by

$$\theta_k(t) = B(t)u_k,$$

where u_k is penalized through

$$u_k \sim N(0, \sigma^2 D).$$

Like above, the penalty matrix is appropriately chosen (see e.g. Wood 2017) and $B(t)$ is a B-spline basis. Hence, smooth functions and smooth random heterogeneity can be estimated in a coherent framework (see Durbán et al. 2005). The entire model can be integrated in the flexible generalized additive model (GAM) framework provided by Wood (2017) (see also Wood 2006) which is implemented in the `mgcv` package (version 1.8-28) by Wood (2011). The identification of the smooth components and the intercept term is ensured by a "sum-to-zero" constraint (Wood 2017). For further details see the Appendix A.2.

4. Results

4.1. Time-varying fixed effects

The results of the time-varying effects are grouped into network-related covariates (presented in Figure 3) and political and economic covariates (presented in Figure 4). The left columns give the coefficients for the formation model and the right columns for the persistence model, respectively. In the case of the network statistics, a schematic representation of the corresponding network effects is added on the right hand side. The values for the coefficients are presented as solid lines with shaded regions, indicating two standard error bounds. The zero-line is indicated as dashed line and the estimates for time-constant coefficients are given by the dotted horizontal line. Note that the coefficients at a given time point can be interpreted just as the coefficients in a simple logit model. Additionally, for the same coefficient (or coefficients with the same norming) in the formation and persistence model, the effect size can be compared directly.

Network-Effects (see Figure 3)

Outdegree: The senders' outdegree has a coefficient that is almost time-constant and close to zero for both models. This stands in contrast to the findings of Thurner et al. (2018), where a strong effect is present. Hence, once controlled for country-specific heterogeneity (especially the sender-specific country effect), no population-level outdegree effect for the exporter is present (we show in the Supplementary Material that the effect is indeed present when country-specific heterogeneity is excluded).

However, the inclusion of country-specific sender and receiver effects does not affect the effect of the receivers' outdegree and the coefficient is consistently negative, and slightly increasing over time in the formation model. For the persistence model, we find a less pronounced but significant negative effect. We interpret this as clear evidence that countries with a high outdegree are comparatively less frequently importing, and importers usually have relatively less frequent export relations. According to our experience this specification captures the trade asymmetries of the oligopolistic market better than just specifying the indegrees of the receiver.

Reciprocity: Controlling for the distinguished asymmetrical nature of the weapons transfers,

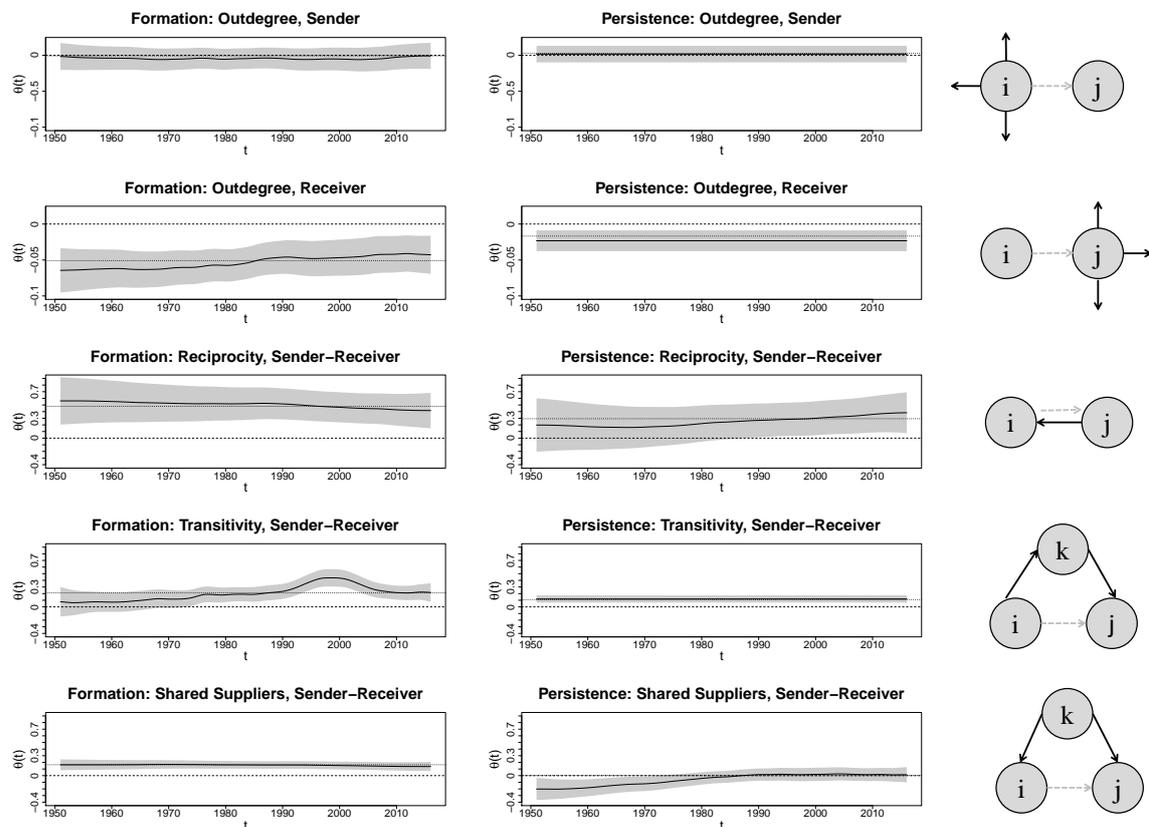


Figure 3: Time-varying coefficients of network statistics in solid black. Shaded areas give two standard error bounds. Time-constant effects in dashed grey and zero line in dotted black. Schematic representation of the network effects on the right hand side.

we identify a positive and significant impact of reciprocity in the formation model. Reciprocity in repeated transfers is only a relevant feature after the breakdown of the bipolar block structure. We conclude that the asymmetric structure is more present in persistent trade relations with importing countries that are typically dependent on big exporters.

Transitivity: Looking at three-node statistics it can be seen that the variable transitivity has a positive impact on the formation and persistence. In the formation model, the effect is insignificant in the first years. This may be influenced by the clear hegemony of the United States and the Soviet Union, respectively, immediately after World War II which did not require a shared control over the recipient country, because the donor was powerful enough to secure the terms of a deal. In the 1980s middle power countries became technologically

more advanced and especially in the West, they joined the US in delivering to other countries. The pronounced change between 1990 and 2010 can be explained by the break up of the two hostile blocs and the interruption of long-standing arm-trading partnerships leading to a fundamental reorganization until 2010 when the effect came back to the level of 1990. Although these arguments are also valid for the persistence model, we see that transitivity is less relevant for ongoing, repeated transfers (the time constant effect in the formation model has twice the size as the one in the persistence model). This impression is also strengthened by the fact that the coefficient is not subject to changes over time.

Shared Suppliers: The coefficients related to the shared suppliers corroborate our expectation that many shared suppliers lead to the formation of transfers (positive and significant coefficient for the whole time period in the formation model). This indeed mirrors the phenomenon described above: there is a hierarchy of producing countries in the world. Receiver countries i and j should become acquainted with these technologies and should have similar levels of production capacities. This allows them to exchange arms. Also, the act of receiving both from the same supplier means that this country places trust to both receivers - such that this facilitates trust giving one to another. On the other hand, in the persistence model, the effect is indeed significantly negative and virtually zero from the 1975 on, showing that repetitive trading is not promoted by many shared suppliers.

Covariate Effects (see Figure 4)

Formal Alliance: The impact of a bilateral formal alliances on the formation of a transfer is positive and significant for both, the formation and with a more modest effect for the persistence, corroborating our expectation that formal alliances are most relevant for the formation, i.e. by passing the required threshold of starting weapons transfers. The required threshold of trustfulness to start seems to decline over time for the initiation. Hence, while formal alliances play a central role for arms trading after the second world war, the formation of arms trades is less and less influenced by the existence of a formal alliance by the sending and receiving state. However, given there exists an alliance, the impact (despite being smaller) continues to be relevant for repeated transfers. This is an important insight as we show for the first time that formalized alliance actually breed a dense web of arms transfers.

Regime Dissimilarity: For the formation model, the coefficient on the absolute difference of the polity scores is all along negative, significant and shows some time variation. With the decay of the eastern bloc, the resistance to send new arms to dissimilar regimes increases until 2000. After that, the absolute effect of different polity scores declines again, coming back to the long-term constant effect. Interestingly, we find that regime dissimilarity is irrelevant in the persistence model, showing that given a relationship is started, repetition does no more require regimes to exhibit shared governance values.

GDP: As expected, the coefficients on the logarithmic GDP for sender and receiver are positive and constant for both models. However, the effect for the senders' GDP is much

stronger in the formation model, showing that indeed mostly economically strong countries are able to open new markets for arms exports. Together, the coefficients support the "gravity hypothesis", i.e. greater economic power and market sizes of the sender as well as the receiver increases the probability of forming and maintaining trade relations. However, given a transfer relation is started, this effect becomes smaller for repetition.

Distance: In accordance with previous insights (Thurner et al. 2018), the results on the logarithmic distance contradicts the standard gravity model and distance proves to be insignificant in both models.

Military Expenditures: For the military expenditures of the sender, we find very comparable and declining effects that become insignificant from 1990 on in both models. This indicates that with the end of the cold war the dominance of exporting countries with high military budgets has decreased. For the receivers' military expenditures in the formation model, the effect is positive and turns significant with time. This clearly illustrates that the military expenditures of the receiver are not as important in the Cold War period where super powers often granted military assistance. Only with end of the 1980s there begins a marketization of the weapons transfers with suppliers demanding money for delivery. Given there is a preceding exchange, we find a very strong effect for the military expenditures of the receiver for the full observational period, indicating that the availability of huge military expenditures is a key for understanding the continuous yearly inflow of weapons.

Overall, the results confirm our initial hypothesis. Judged by the size of the coefficients and their significance we find that the network statistics (reciprocity, transitivity, shared suppliers) and security related covariates (formal alliance, regime dissimilarity) prove to be highly influential in the formation model. On the other hand, we find weaker (or insignificant) network effects in the persistence model combined with a high dominance of the GDP and especially the military expenditures of the receiving country. This is not to say that we regard for example the positive effect of transitivity or alliances in the persistence model as irrelevant for repeated trading since the special nature of arms trading clearly demands trust for the formation and the persistence of transfers but the effects nevertheless show that the two processes are guided by different mechanisms that attach different priorities to security-related and economic variables.

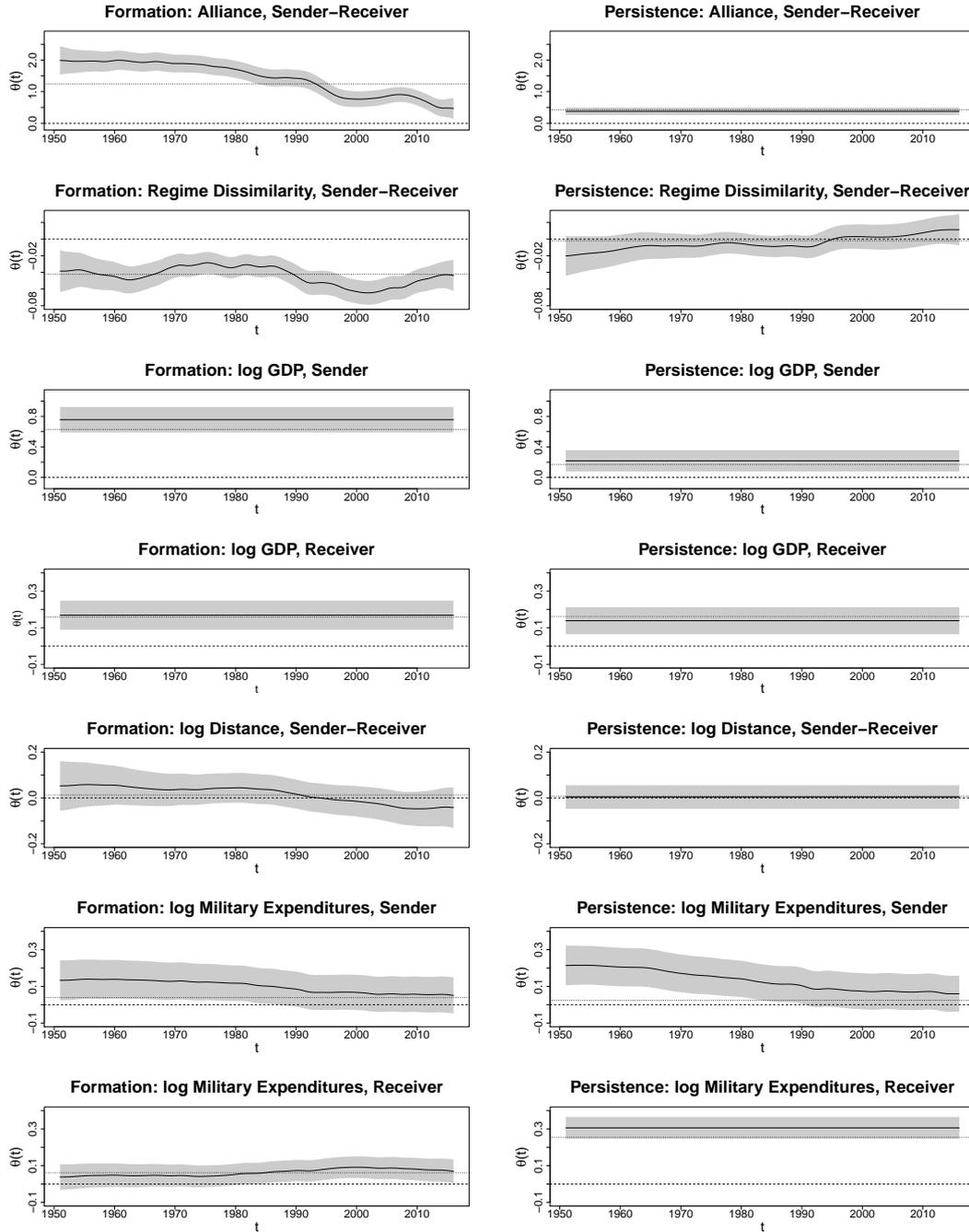


Figure 4: Time-varying coefficients of political and economic covariates in solid black. Shaded areas give two standard error bounds. Time-constant effects in dashed grey and zero line in dotted black.

4.2. Time-varying smooth random effects

4.2.1. Functional component analysis

We now pay attention to the actor-specific heterogeneity. In Figure 5, the country-specific effects for the sender, as well as the receiver countries are visualized for the formation model on the left and the persistence model on the right. Note that in these plot we have truncated the curves for the years where countries are not existent.

At a first sight, interpretation of these plots looks clumsy. We therefore retrieve information by employing a functional principal component analysis to the multivariate time series of random effects seen in Figure 5 (see also Ramsay and Silverman 2005 and the Appendix A.3). The results are shown in Figure 6 for the formation model and in Figure 7 for the persistence model. On the left hand side the scores of the first two principal components are plotted, where the latter are visualized on the right hand side. The share of variance explained by the respective component is provided in the brackets. The basic idea of the approach is to show the effect of the principal components as perturbations from the mean random effects curves. By adding (the "+" line) or subtracting (the "-" line) a multiple of the principal component curve we get the visualized perturbation from the mean.

The first principal component is close to be constant and represents the share of variance induced by different overall levels of the random effect curves. The dynamic of the random effects is captured by the second principal component, delivering a tendency for an upward movement if positive and downward if negative. Hence, looking on the horizontal axes, we see countries that build up their arm trade links over the years as exporters (importers) on the right hand side while countries that are reluctant to building up export (import) links are plotted on the left hand side. Looking on the vertical axes, we see countries that decrease their role as exporter (importer) over the time on the bottom, and vice versa countries that increase the number of export (import) links over time on the top. All these effects are conditional on the remaining covariate effects discussed before. Hence, these random effects capture the remaining heterogeneity not included in the remaining model.

4.2.2. Results of the functional component analysis

Because of the great amount of information condensed in Figures 6 and 7 we restrict our interpretation to a few global patterns and selected countries that take either very special positions in the arms trade network (high or low values for component 1) or exhibit variation over time (high or low values for component 2). Overall regarding the different levels of the random effects, it can already be seen in Figure 5, that the heterogeneity is much more pronounced in the formation model in comparison to the persistence model. Furthermore, in the formation model, the countries differ more strongly in their ability to export in comparison to their ability to import while this contrast is not present in the persistence model.

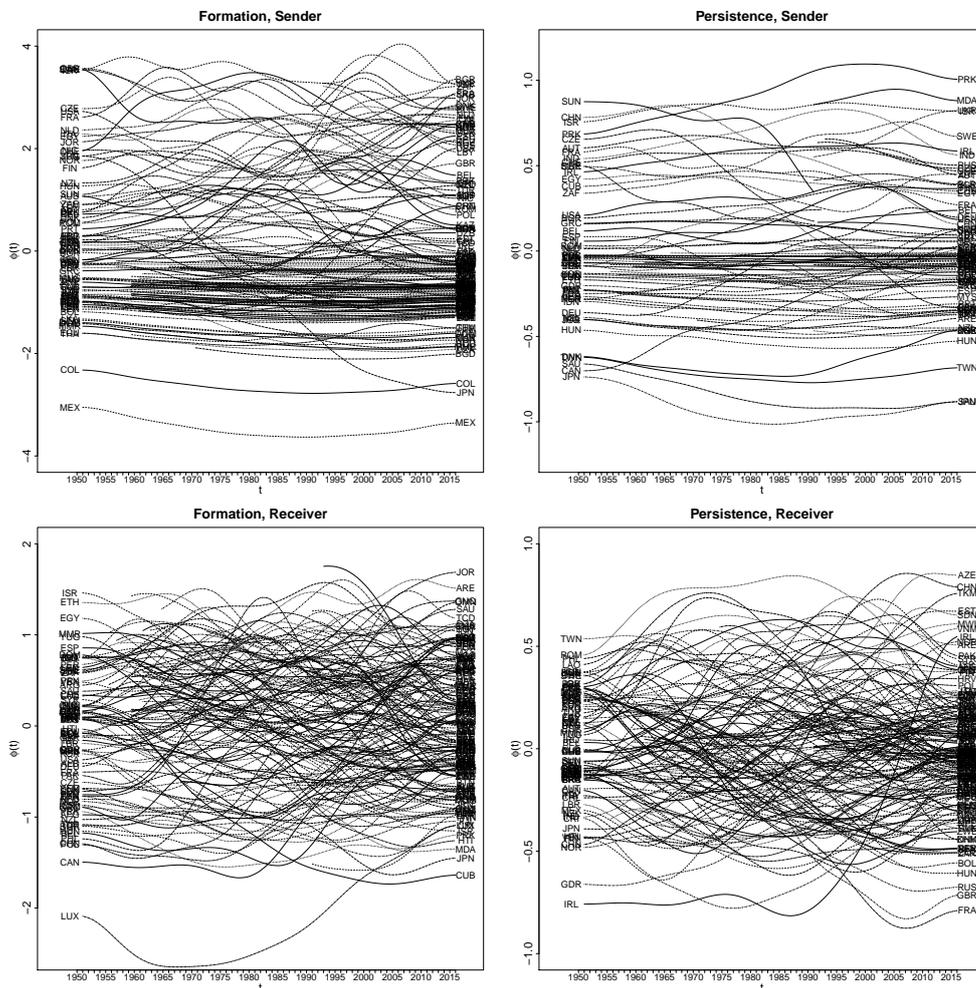


Figure 5: Fitted time-varying smooth random effects $\phi(t)$ plotted against time with country codes. The respective models are in the columns (formation on the left and persistence on the right) and the type of random effects in the rows (sender effect on the top and receiver effect on the bottom).

A global pattern regarding the dynamics of the sender effect becomes visible since the top left in Figure 6 looks like a lying mushroom. That is, countries that started on a low level (i.e. negative component 1) show, with the exception of Japan (JPN) and Turkey (TUR), not very much upward or downward variability (i.e. low level for component 2). In contrast, countries that have a random effect above zero move more strongly up or down with time.

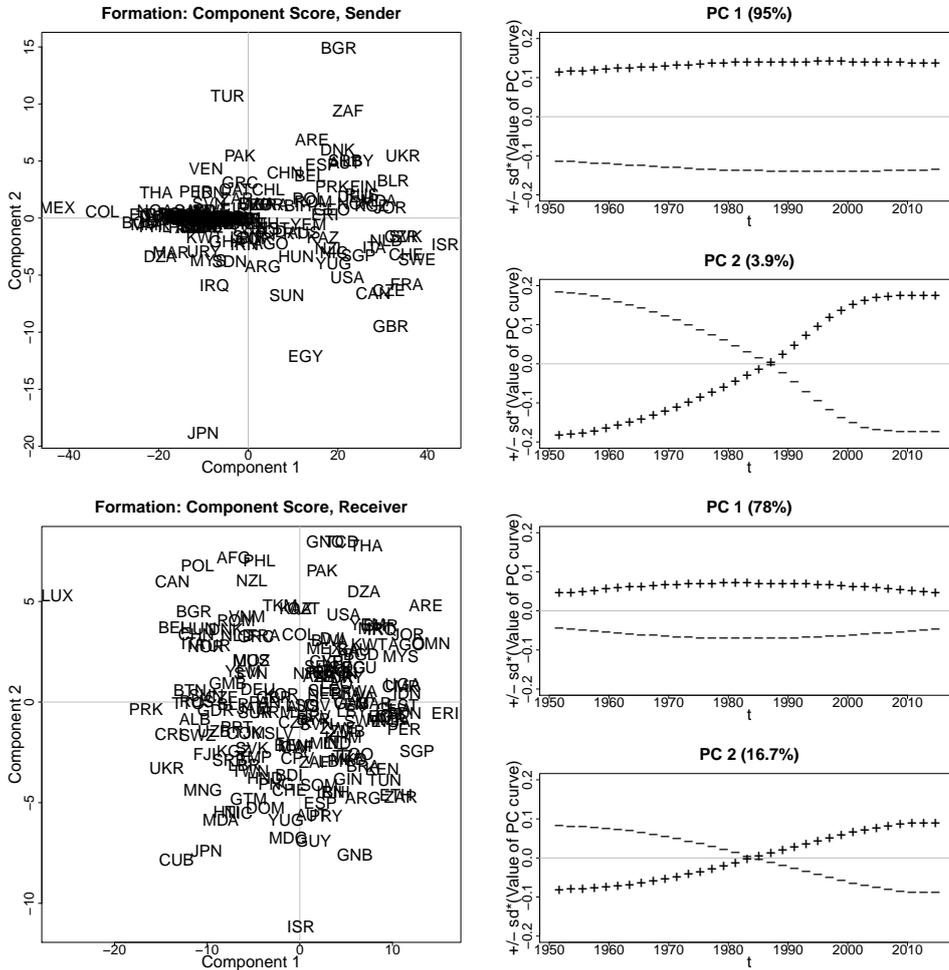


Figure 6: Functional principal component analysis of the smooth random effects in the formation model for the sender (top) and the receiver (bottom). Scores of the random effects for the first two principal components are given on the left. Mean principal component curve (zero line) and the effects of adding (+) and subtracting the principal component curve are given on the right.

This means that the export dynamics are mainly driven by countries with relatively high sender effects.

Figures 6 and 7 show very well that fundamental changes of the system are driven by the end of the cold war. This can be seen exemplary regarding the position of the Soviet Union (SUN) and Czechoslovakia (CZE) in the top left in Figures 6 and 7 (both with a high level

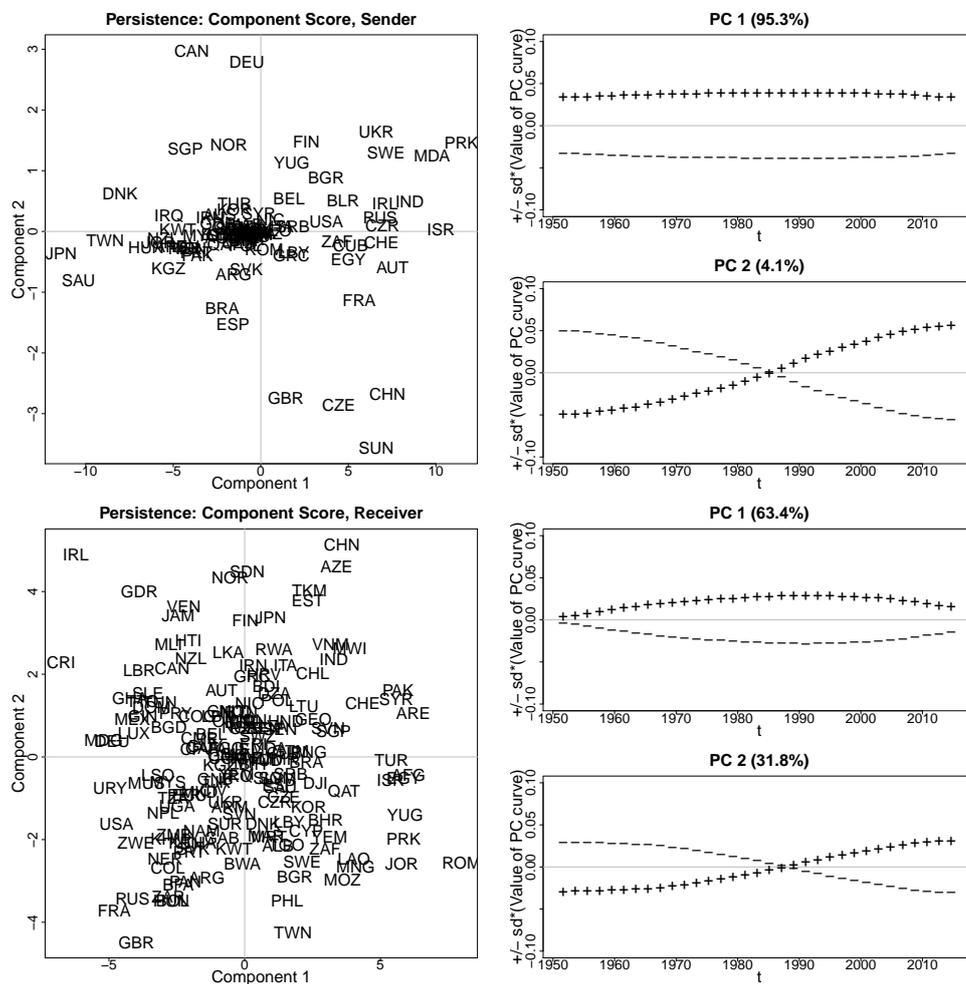


Figure 7: Functional principal component analysis of the smooth random effects in the persistence model for the sender (top) and the receiver (bottom). Scores of the random effects for the first two principal components are given on the left. Mean principal component curve (zero line) and the effects of adding (+) and subtracting the principal component curve are given on the right.

for component 1 and a low level for component 2). This mirrors that these countries left the system shortly after the collapse of the eastern bloc. However, this turning point affected not only exporters but also importers and consequently the representation of the receiver effects of the formation model at the bottom left of Figure 6 is populated with (former) socialist countries such as Cuba (CUB), Ukraine (UKR), North Korea (PRK), Yugoslavia (YUG) and

Moldova (MDA). Additionally, we find a prominent position for Romania (ROM), being a country that has a high level (high value for component 1) but decreased its' tendency to be a receiver in persistent trade relations (low value for component 2) in Figure 7. However, while some of the countries of the eastern block ceased to exist or strongly reduced their exports or imports we also find a contrary pattern. Countries like Ukraine (UKR) and Bulgaria (BGR) have managed to increase their sender effect in the formation as well as in the persistence model with time (high value for component 1 and component 2 in the top left of Figures 6 and 7). This indicates that some left overs of the collapsed Soviet Union defence industries sold off their stocks and rushed into the global market of military products.

Besides the massive shift initiated by the end of the cold war, we see that some dominant exporting countries, especially Great Britain (GBR), France (FRA) and Egypt (EGY), lost importance over time. This countries can be found in the fourth quadrant of the top left panels in Figures 6 and 7, meaning their high sender effects decreased strongly with time. This might seem surprising since France and Great Britain are still among the countries with the highest exported volumes. However, France and Great Britain have left their dominance over former colonies leading to a loss of control over many potential importers. The general pattern also carries over to their receiver effects. Looking at the scores of Great Britain (GBR) and France (FRA) at the bottom left of Figure 7 we see a strong decrease of their receiver effects in the persistence model.

Apart from global patterns, some countries exhibit exceptional scores that can be traced back to country-specific circumstances. We find that Japan (JPN) stands out among the countries with the lowest proclivity to import (see the low scores for components 1 and 2 at the bottom left of Figure 6). Even more pronounced is the astonishing low tendency to export, mirrored by Japan's sender effect in the persistence models (Figure 7, top left) and the strongly declining sender effect in the formation model (Figure 6, top left). This stands in contrast to the fact that Japan is among the wealthiest countries with a highly developed export industry and is clearly due to the highly restrictive arms export principles introduced in 1967, and tightened in 1976. This ban on exports was only lifted in 2014 (Hughes 2018; Ministry of Foreign Affairs of Japan 2014).

Another, very notable case is Israel (ISR), being somehow the opposite pole in comparison to Japan (JPN). The sender effects on the top left of Figures 6 and 7 show that Israel (ISR) has an outstanding tendency to establish and maintain arms exports. On the other hand, Israel (ISR) takes a very polar position in the bottom left of Figure 6 as a consequence of a strongly decreased (i.e. low level for component 2) receiver effect in the formation model. These results reflect the country's path of developing highly internationally competitive weapons systems and its' rise to be one of the most important exporters. This stands in contrast to countries like Mexico (MEX), being the country with the least tendency to form new trade exports (top left in Figure 6). It appears that this country is not able to be a relevant player in the market despite being among the worlds' largest economies. We consider these special paths as induced by cumulative advantages and learning over time in

the one case (Israel), whereas in the case of Mexico (MEX) we observe the stickiness and path inertia of a country having not been able to make its defence products sold externally.

There remain many other interesting cases. For example the rise of South Africa (ZAF) as an exporter in the formation model (top left in Figure 6), mirroring the history of the country, being initially dependent on imports and now among the major exporters of MCW. We also find that Ireland (IRL) strongly increased its tendency to be a persistent importer after its entry to the European Union (bottom left in Figure 7) while Germany (DEU) and Canada (CAN) strongly increased their roles as persistent exporters (top left in Figure 7).

4.3. Model evaluation

The evaluation of the out-of-sample predictive power is based on the following steps. We first fit the formation model as well as the persistence model, based on the information in $t - 1$, to the data in t and use the estimated coefficients for the prediction of new formation or persistence of existing ties in $t + 1$. As the predictions are probabilistic by their nature, we weight the recall (true positive rate) against the false positive rate for varying threshold levels, yielding the ROC curve and the AUC for each year of prediction. Because arms transfers can be regarded as rare events we also compute the PR curve and the corresponding AUC. The results are plotted in Figure 8 with the AUC values that correspond to the PR curves on the left and the one corresponding to the ROC curves on the right. The first row gives the evaluation of the formation model and the second row shows the persistence model. While the AUC values in the formation model are very high when evaluated at the ROC curves they are much lower with the PR curves. This is a consequence of being right quite frequently if a zero is predicted, while it is hard to forecast the actual transfers in the next period in case of the formation model. Interestingly, the opposite holds for the persistence model. In the combined version at the bottom of Figure 8 the AUC values derived from the PR curve show that the model does quite well.

Additionally, we evaluate how well global network structures like the mean outdegree, the share of reciprocity and observed transitivity can be mirrored by the predictions using a simulation-based approach (see Hunter et al. 2008). To do so, we fit the models for the transition between $t - 1$ and t and simulate from the formation model and the persistence model 1,000 times based on the information in t . Then, based on equation (1), the predicted network for $t + 1$ is constructed. From this, we evaluate global network characteristics and compare them to the actual characteristics from the true MCW trade network in $t + 1$. The corresponding Figure 11 is given in the Appendix A.4. The results are reassuring and the simulated networks mirror the real network properties in an acceptable way.

Clearly the proposed model is not the only suitable network model. Alternatively, it is possible to analyse the data with a STERGM without random effects and with various variants of the ERGM or the TERGM with and without random effect. We discuss this extensively in the Supplementary Material and show that the out-of-sample predictive power

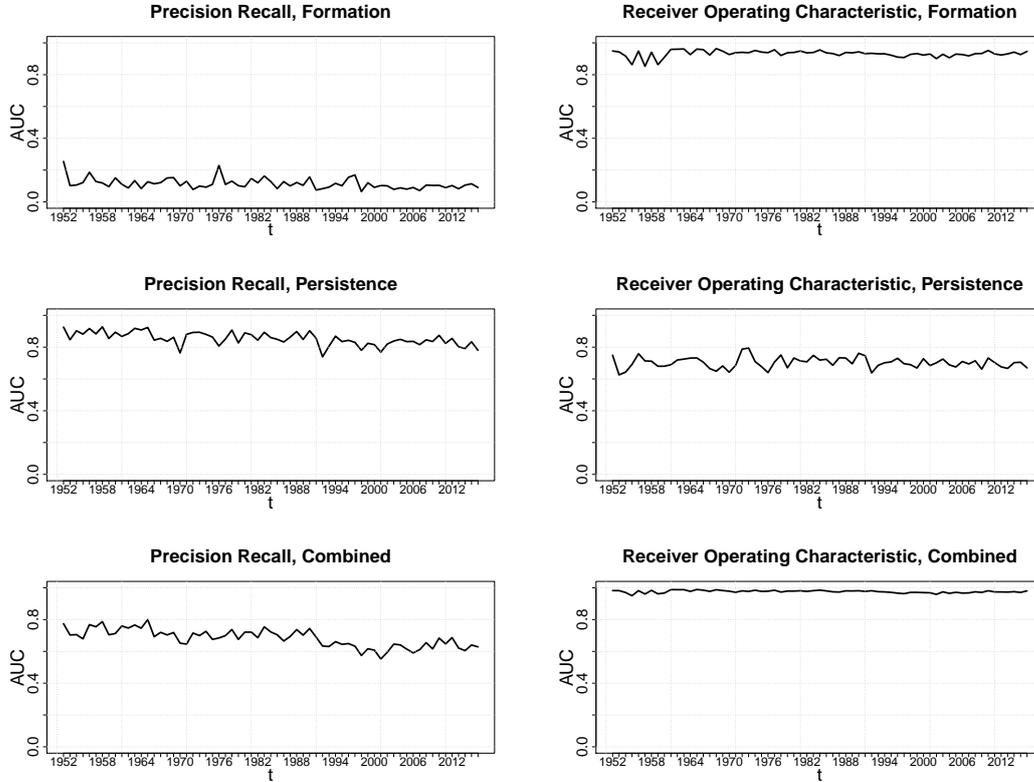


Figure 8: Time series of area under the curve (AUC) values for precision recall (PR) on the left and AUC values for the receiver operating characteristic (ROC) on the right. Formation model in the first row, persistence model in the second row and their combination in the last row.

of our model is superior to all other fitted candidate models.

5. Conclusion

In this paper we employ a dynamic separable network model as introduced by Krivitsky and Handcock (2014) and add techniques proposed by Hastie and Tibshirani (1993) and Durbán et al. (2005). This enables us to study the process of formation and persistence separately as well as the inclusion of time-varying coefficients and smooth time-varying random effects that are further analysed by methods from functional data analysis as described in Ramsay and Silverman (2005).

Applied to the discretized MCW networks from 1950 to 2016 we find that the mechanisms leading to formation and persistence differ fundamentally. Most importantly, the formation is driven by network effects and security related variables, while the persistence of transfers is dominated the military expenditures of the receiving country. A careful analysis of the random effects exhibits a high variation among the countries as well as along the time dimension. By using functional principal component analysis we decompose the functional time series of smooth random effects in order find countries that have increased or decreased their relative importance in the network. The evaluation of the fit confirms that the chosen model is able to give good out-of-sample predictions.

References

- Akerman, A. and A. L. Seim (2014). The global arms trade network 1950–2007. *Journal of Comparative Economics* 42(3), 535–551.
- 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.
- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science* 286(5439), 509–512.
- Barigozzi, M., G. Fagiolo, and D. Garlaschelli (2010). Multinetwork of international trade: A commodity-specific analysis. *Physical Review E* 81(4), 046104.
- Blanton, S. L. (2005, 11). Foreign Policy in Transition? Human Rights, Democracy, and U.S. Arms Exports. *International Studies Quarterly* 49(4), 647–667.
- Block, P., J. Koskinen, J. Hollway, C. Steglich, and C. Stadtfeld (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks* 52, 180 – 191.
- Bramoullé, Y., A. Galeotti, B. Rogers, and T. Chaney (2019). *Networks in International Trade*. Oxford: Oxford University Press.
- Center for systemic Peace (2017). Polity IV annual time-series, 1800-2015, version 3.1. Accessed: 2017-06-02.
- Correlates of War Project (2017a). International military alliances, 1648-2012, version 4.1. Accessed: 2017-05-03.
- Correlates of War Project (2017b). National material capabilities, 1816-2012, version 5.0. Accessed: 2017-02-06.
- Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5), 1–9.
- Disdier, A.-C. and K. Head (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics* 90(1), 37–48.
- Duijn, M. A., T. A. Snijders, and B. J. Zijlstra (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica* 58(2), 234–254.
- Durban, M. and M. C. Aguilera-Morillo (2017). On the estimation of functional random effects. *Statistical Modelling* 17(1-2), 50–58.

- Durbán, M., J. Harezlak, M. Wand, and R. Carroll (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 24(8), 1153–1167.
- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.
- Erickson, J. L. (2015). *Dangerous Trade: Arms Exports, Human Rights, and International Reputation*. New York: Columbia University Press.
- García-Alonso, M. D. and P. Levine (2007). Arms trade and arms races: A strategic analysis. In T. Sandler and K. Hartley (Eds.), *Handbook of Defense Economics: Defense in a globalized world*, Volume 2, pp. 941–971. Amsterdam: Elsevier Science Publishing.
- Gleditsch, K. S. (2013a). Distance between capital cities. Accessed: 2017-04-07.
- Gleditsch, K. S. (2013b). Expanded trade and GDP data. Accessed: 2017-04-07.
- Grau, J., I. Grosse, and J. Keilwagen (2015). PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* 31(15), 2595–2597.
- 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), 1548–7660.
- Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based clustering for social networks. *J. R. Statist. Soc. A* 170(2), 301–354.
- Hanneke, S., W. Fu, E. P. Xing, et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* 4, 585–605.
- Harkavy, R. E. (1975). *The Arms Trade and International Systems*. Cambridge: Cambridge University Press.
- Hastie, T. and R. Tibshirani (1987). Generalized additive models: some applications. *J. Am. Statist. Ass.* 82(398), 371–386.
- Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *J. R. Statist. Soc. B* 55(4), 757–796.
- Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of international economics*, Volume 4, pp. 131–195. Amsterdam: Elsevier Science Publishing.
- Hlavac, M. (2013). stargazer: Latex code and ascii text for well-formatted regression and summary statistics tables.

- Hoff, P., B. Fosdick, A. Volfovsky, and Y. He (2015). `amen`: Additive and multiplicative effects models for networks and relational data. R package version 1.3.
- Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *J. Am. Statist. Ass.* 97(460), 1090–1098.
- Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *J. Am. Statist. Ass.* 76(373), 33–50.
- Holme, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B* 88(9), 1–30.
- Hughes, C. (2018). Japan’s emerging arms transfer strategy: diversifying to re-centre on the us–japan alliance. *The Pacific Review* 31(4), 424–440.
- Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *J. Am. Statist. Ass.* 103(481), 248–258.
- 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.
- Krause, K. (1995). *Arms and the state: patterns of military production and trade*. Cambridge: Cambridge University Press.
- Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *J. R. Statist. Soc. B* 76(1), 29–46.
- Leifeld, P., S. J. Cranmer, and B. A. Desmarais (2018). Temporal exponential random graph models with `btergm`: estimation and bootstrap confidence intervals. *Journal of Statistical Software* 83(6).
- Lusher, D., J. Koskinen, and G. Robins (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.
- Marshall, M. G. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2016. Accessed: 2017-06-02.
- Ministry of Foreign Affairs of Japan (2014). Japan’s policies on the control of arms exports. Accessed: 2017-02-21.
- Moritz, S. (2016). `imputeTS`: Time series missing value imputation. R package version 2.6.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

- Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. New York: Springer Science & Business Media.
- Robins, G. and P. Pattison (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology* 25(1), 5–41.
- Ruppert, D., M. Wand, and R. J. Carroll (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics* 1(3), 1193–1256.
- Schulze, C., O. Pamp, and P. W. Thurner (2017, 10). Economic Incentives and the Effectiveness of Nonproliferation Norms: German Major Conventional Arms Transfers 1953–2013. *International Studies Quarterly* 61(3), 529–543.
- Schweitzer, F., G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White (2009). Economic networks: The new challenges. *Science* 325(5939), 422–425.
- Singer, J. D., S. Bremer, and J. Stuckey (1972). Capability distribution, uncertainty, and major power war, 1820–1965. *Peace, War, and Numbers* 19, 19–48.
- SIPRI (2017a). Arms transfers database. Accessed: 2017-06-02.
- SIPRI (2017b). Arms transfers database - methodology. Accessed: 2017-03-23.
- Snijders, T. A. (2011). Statistical models for social networks. *Annual Review of Sociology* 37(1), 131–153.
- Snijders, T. A., J. Koskinen, and M. Schweinberger (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics* 4(2), 567.
- Snijders, T. A., G. G. Van de Bunt, and C. E. Steglich (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32(1), 44–60.
- Squartini, T., G. Fagiolo, and D. Garlaschelli (2011a). Randomizing world trade. I. A binary network analysis. *Physical Review E* 84(4), 046117.
- Squartini, T., G. Fagiolo, and D. Garlaschelli (2011b). Randomizing world trade. II. A weighted network analysis. *Physical Review E* 84(4), 046118.
- Thiemichen, S., N. Friel, A. Caimo, and G. Kauermann (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks* 46, 11–28.
- Thurner, P. W., Schmid, C. Christian, Skyler, and G. Kauermann (2018). Network interdependencies and the evolution of the international arms trade. Online First: *Journal of Conflict Resolution*.

- Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* 62(4), 1025–1036.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Statist. Soc. B* 73(1), 3–36.
- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Boca Raton: CRC press.
- Wood, S. N., Y. Goude, and S. Shaw (2015). Generalized additive models for large data sets. *J. R. Statist. Soc. C* 64(1), 139–155.
- World Bank (2017). World bank open data, real GDP. Accessed: 2017-04-01.
- Xu, K. (2015). Stochastic block transition models for dynamic networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Volume 18, pp. 1079–1087.

A. Appendix

A.1. Descriptives

In Figure 9, the binary network is shown for the years 2015 and 2016. Table 1 gives the categories of arms that are included in the analysis. All types with explanations are taken from SIPRI (2017b). The 171 countries that are included in our analysis can be found in Table 2, together with the three-digit country codes that are used to abbreviate countries in the paper. In addition to that, the time periods, for which we coded the countries as existent are included. Note that the SIPRI data set contains more than 171 arm trading entities but we excluded non-states and countries with no (reliable) covariates available. In the covariate GDP some missings are present in the data. No time series of covariates for the selected countries is completely missing (those countries are excluded from the analysis) and the major share of them is complete but there are series with some missing values. This is sometimes the case in the year 1990 and/or 1991 where the former socialist countries splitted up or had some transition time. In other cases values at the beginning or at the end of the series are missing. We have decided on three general rules to fill the gaps: First, if a value for a certain country is missing in t but there are values available in $t - 1$ and $t + 1$, the mean of those is used. If the values are missing at the end of the observational period, the last value observed is taken. In case of missing values in the beginning, the first value observed is taken.

The series on military expenditures are imputed similarly, using linear interpolation by employing the R package `imputeTS` by Moritz (2016).

The number of countries included each year in the network is provided in the upper left panel of Figure 10. It can be seen that the network is growing almost each year until 1992, with two big leaps that show the effects of the decolonization, beginning in 1960 and the end of the Soviet Union after 1991. Typical descriptive statistics for the analysis of networks are *Density*, *Reciprocity* and *Transitivity*, all shown in Figure 10. The *Density* is defined as the number of edges divided by the number of possible edges. *Reciprocity* is defined as the share of trade flows being reciprocal. *Transitivity* is defined as the ratio of triangles and connected triples in the graph.

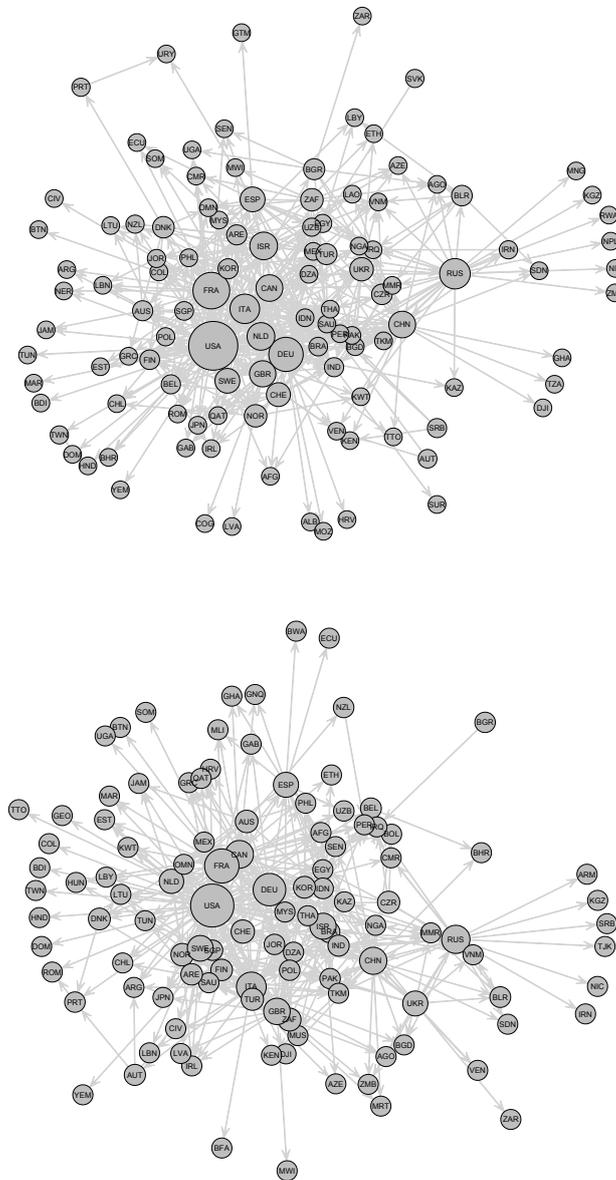


Figure 9: Network of international transfers of major conventional weapons (MCW) in 2015 (top) and 2016 (bottom). Countries are represented by vertices and directed edges represent arms exports. Vertex sizes are scaled proportional to the logarithmic outdegree (number of outgoing edges).

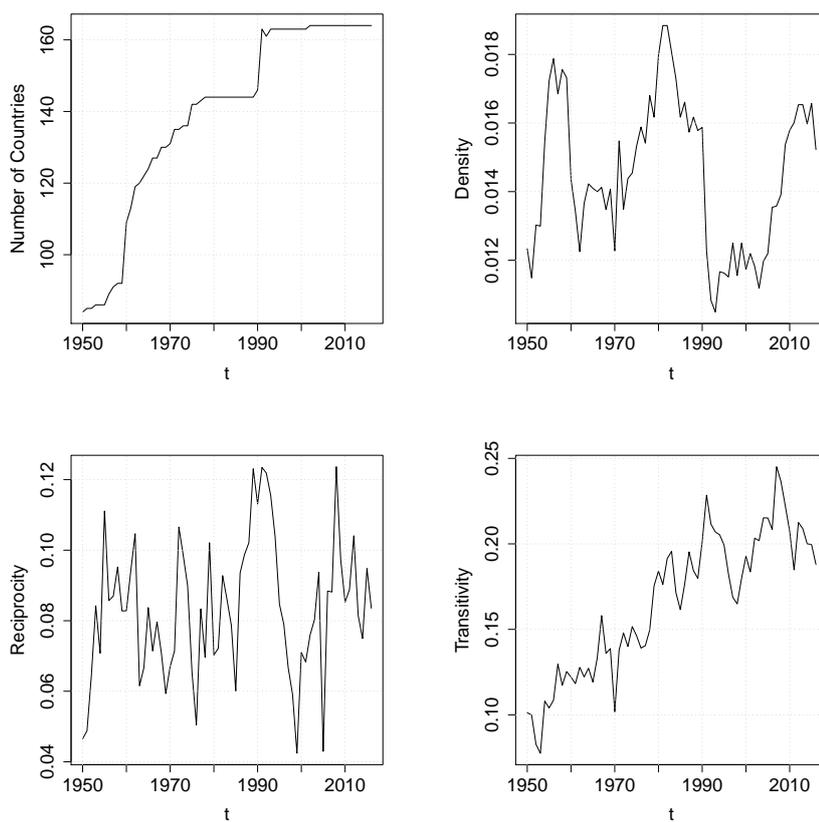


Figure 10: Time series from 1950 to 2016 of global network statistics for the international arms trade network for the included countries. Number of countries included on the top left panel. Number of realized transfers relative to the number of possible transfers on the top right Panel. Share of reciprocated transfers on the bottom left panel. Ratio of undirected triangles relative to connected triples on the bottom right.

Type	Explanation
Aircraft	All fixed-wing aircraft and helicopters, including unmanned aircraft with a minimum loaded weight of 20 kg. Exceptions are microlight aircraft, powered and unpowered gliders and target drones.
Air-defence systems	(a) All land-based surface-to-air missile systems, and (b) all anti-aircraft guns with a calibre of more than 40 mm or with multiple barrels with a combined caliber of at least 70 mm. This includes self-propelled systems on armoured or unarmoured chassis.
Anti-submarine warfare weapons	Rocket launchers, multiple rocket launchers and mortars for use against submarines, with a calibre equal to or above 100 mm.
Armoured vehicles	All vehicles with integral armour protection, including all types of tank, tank destroyer, armoured car, armoured personnel carrier, armoured support vehicle and infantry fighting vehicle. Vehicles with very light armour protection (such as trucks with an integral but lightly armoured cabin) are excluded.
Artillery	Naval, fixed, self-propelled and towed guns, howitzers, multiple rocket launchers and mortars, with a calibre equal to or above 100 mm.
Engines	(a) Engines for military aircraft, for example, combat-capable aircraft, larger military transport and support aircraft, including large helicopters; (b) Engines for combat ships - fast attack craft, corvettes, frigates, destroyers, cruisers, aircraft carriers and submarines; (c) Engines for most armoured vehicles - generally engines of more than 200 horsepower output.*
Missiles	(a) All powered, guided missiles and torpedoes, and (b) all unpowered but guided bombs and shells. This includes man-portable air defence systems and portable guided anti-tank missiles. Unguided rockets, free-fall aerial munitions, anti-submarine rockets and target drones are excluded.
Sensors	(a) All land-, aircraft- and ship-based active (radar) and passive (e.g. electro-optical) surveillance systems with a range of at least 25 kilometres, with the exception of navigation and weather radars, (b) all fire-control radars, with the exception of range-only radars, and (c) anti-submarine warfare and anti-ship sonar systems for ships and helicopters.*
Ships	(a) All ships with a standard tonnage of 100 tonnes or more, and (b) all ships armed with artillery of 100-mm calibre or more, torpedoes or guided missiles, and (c) all ships below 100 tonnes where the maximum speed (in kmh) multiplied with the full tonnage equals 3500 or more. Exceptions are most survey ships, tugs and some transport ships
Other	(a) All turrets for armoured vehicles fitted with a gun of at least 12.7 mm calibre or with guided anti-tank missiles, (b) all turrets for ships fitted with a gun of at least 57-mm calibre, and (c) all turrets for ships fitted with multiple guns with a combined calibre of at least 57 mm, and (d) air refueling systems as used on tanker aircraft.*

*In cases where the system is fitted on a platform (vehicle, aircraft or ship), the database only includes those systems that come from a different supplier from the supplier of the platform.

The Arms Transfers Database does not cover other military equipment such as small arms and light weapons (SALW) other than portable guided missiles such as man-portable air defence systems and guided anti-tank missiles. Trucks, artillery under 100-mm calibre, ammunition, support equipment and components (other than those mentioned above), repair and support services or technology transfers are also not included in the database.

Source: SIPRI (2017b)

Table 1: Types of weapon systems included in the SIPRI arms trade database.

Country	Code	Included	Country	Code	Included	Country	Code	Included
Afghanistan	AFG	1950 - 2016	German Dem. Rep.	GDR	1950 - 1991	Pakistan	PAK	1950 - 2016
Albania	ALB	1950 - 2016	Germany	DEU	1950 - 2016	Panama	PAN	1950 - 2016
Algeria	DZA	1962 - 2016	Ghana	GHA	1957 - 2016	Papua New Guin.	PNG	1975 - 2016
Angola	AGO	1975 - 2016	Greece	GRC	1950 - 2016	Paraguay	PRY	1950 - 2016
Argentina	ARG	1950 - 2016	Guatemala	GTM	1950 - 2016	Peru	PER	1950 - 2016
Armenia	ARM	1991 - 2016	Guinea	GIN	1958 - 2016	Philippines	PHL	1950 - 2016
Australia	AUS	1950 - 2016	Guinea-Bissau	GNB	1973 - 2016	Poland	POL	1950 - 2016
Austria	AUT	1950 - 2016	Guyana	GUY	1966 - 2016	Portugal	PRT	1950 - 2016
Azerbaijan	AZE	1991 - 2016	Haiti	HTI	1950 - 2016	Qatar	QAT	1971 - 2016
Bahrain	BHR	1971 - 2016	Honduras	HND	1950 - 2016	Romania	ROM	1950 - 2016
Bangladesh	BGD	1971 - 2016	Hungary	HUN	1950 - 2016	Russia	RUS	1992 - 2016
Belarus	BLR	1991 - 2016	India	IND	1950 - 2016	Rwanda	RWA	1962 - 2016
Belgium	BEL	1950 - 2016	Indonesia	IDN	1950 - 2016	Saudi Arabia	SAU	1950 - 2016
Benin	BEN	1961 - 2016	Iran	IRN	1950 - 2016	Senegal	SEN	1960 - 2016
Bhutan	BTN	1950 - 2016	Iraq	IRQ	1950 - 2016	Serbia	SRB	1992 - 2016
Bolivia	BOL	1950 - 2016	Ireland	IRL	1950 - 2016	Sierra Leone	SLE	1961 - 2016
Bosnia Herzegov.	BIH	1992 - 2016	Israel	ISR	1950 - 2016	Singapore	SGP	1965 - 2016
Botswana	BWA	1966 - 2016	Italy	ITA	1950 - 2016	Slovakia	SVK	1993 - 2016
Brazil	BRA	1950 - 2016	Jamaica	JAM	1962 - 2016	Slovenia	SVN	1991 - 2016
Bulgaria	BGR	1950 - 2016	Japan	JPN	1950 - 2016	Solomon Islands	SLB	1978 - 2016
Burkina Faso	BFA	1960 - 2016	Jordan	JOR	1950 - 2016	Somalia	SOM	1960 - 2016
Burundi	BDI	1962 - 2016	Kazakhstan	KAZ	1991 - 2016	South Africa	ZAF	1950 - 2016
Cambodia	KHM	1953 - 2016	Kenya	KEN	1963 - 2016	Soviet Union	SUN	1950 - 1991
Cameroon	CMR	1960 - 2016	North Korea	PRK	1950 - 2016	Spain	ESP	1950 - 2016
Canada	CAN	1950 - 2016	South Korea	KOR	1950 - 2016	Sri Lanka	LKA	1950 - 2016
Cape Verde	CPV	1975 - 2016	Kuwait	KWT	1961 - 2016	Sudan	SDN	1956 - 2016
Central Afr. Rep.	CAF	1960 - 2016	Kyrgyzstan	KGZ	1991 - 2016	Suriname	SUR	1975 - 2016
Chad	TCD	1960 - 2016	Laos	LAO	1950 - 2016	Swaziland	SWZ	1968 - 2016
Chile	CHL	1950 - 2016	Latvia	LVA	1991 - 2016	Sweden	SWE	1950 - 2016
China	CHN	1950 - 2016	Lebanon	LBN	1950 - 2016	Switzerland	CHE	1950 - 2016
Colombia	COL	1950 - 2016	Lesotho	LSO	1966 - 2016	Syria	SYR	1950 - 2016
Comoros	COM	1975 - 2016	Liberia	LBR	1950 - 2016	Taiwan	TWN	1950 - 2016
DR Congo	ZAR	1960 - 2016	Libya	LBY	1951 - 2016	Tajikistan	TJK	1991 - 2016
Congo	COG	1960 - 2016	Lithuania	LTU	1990 - 2016	Tanzania	TZA	1961 - 2016
Costa Rica	CRI	1950 - 2016	Luxembourg	LUX	1950 - 2016	Thailand	THA	1950 - 2016
Cote d'Ivoire	CIV	1960 - 2016	Macedonia	MKD	1991 - 2016	Timor-Leste	TMP	2002 - 2016
Croatia	HRV	1991 - 2016	Madagascar	MDG	1960 - 2016	Togo	TGO	1960 - 2016
Cuba	CUB	1950 - 2016	Malawi	MWI	1964 - 2016	Trinidad Tobago	TTO	1962 - 2016
Cyprus	CYP	1960 - 2016	Malaysia	MYS	1957 - 2016	Tunisia	TUN	1956 - 2016
Czech Republic	CZR	1993 - 2016	Mali	MLI	1960 - 2016	Turkey	TUR	1950 - 2016
Czechoslovakia	CZE	1950 - 1991	Mauritania	MRT	1960 - 2016	Turkmenistan	TKM	1991 - 2016
Denmark	DNK	1950 - 2016	Mauritius	MUS	1968 - 2016	Uganda	UGA	1962 - 2016
Djibouti	DJI	1977 - 2016	Mexico	MEX	1950 - 2016	Ukraine	UKR	1991 - 2016
Dominican Rep.	DOM	1950 - 2016	Moldova	MDA	1991 - 2016	Un. Arab Emirates	ARE	1971 - 2016
Ecuador	ECU	1950 - 2016	Mongolia	MNG	1950 - 2016	United Kingdom	GBR	1950 - 2016
Egypt	EGY	1950 - 2016	Morocco	MAR	1956 - 2016	United States	USA	1950 - 2016
El Salvador	SLV	1950 - 2016	Mozambique	MOZ	1975 - 2016	Uruguay	URY	1950 - 2016
Equatorial Guin.	GNQ	1968 - 2016	Myanmar	MMR	1950 - 2016	Uzbekistan	UZB	1991 - 2016
Eritrea	ERI	1993 - 2016	Namibia	NAM	1990 - 2016	Venezuela	VEN	1950 - 2016
Estonia	EST	1991 - 2016	Nepal	NPL	1950 - 2016	Vietnam	VNM	1976 - 2016
Ethiopia	ETH	1950 - 2016	Netherlands	NLD	1950 - 2016	South Vietnam	SVM	1950 - 1975
Fiji	FJI	1970 - 2016	New Zealand	NZL	1950 - 2016	Yemen	YEM	1991 - 2016
Finland	FIN	1950 - 2016	Nicaragua	NIC	1950 - 2016	North Yemen	NYE	1950 - 1991
France	FRA	1950 - 2016	Niger	NER	1960 - 2016	South Yemen	SYE	1950 - 1991
Gabon	GAB	1960 - 2016	Nigeria	NGA	1960 - 2016	Yugoslavia	YUG	1950 - 1992
Gambia	GMB	1965 - 2016	Norway	NOR	1950 - 2016	Zambia	ZMB	1964 - 2016
Georgia	GEO	1991 - 2016	Oman	OMN	1950 - 2016	Zimbabwe	ZWE	1950 - 2016

Table 2: Countries included in the analysis (columns 1, 4 and 7) with three-digit country codes (columns 2, 5 and 8) and time period of inclusion in the model (columns 3, 6 and 9).

A.2. Details on the estimation procedure

The recent implementation of Generalised Additive Models (GAM) in the R package `mgcv` allows for smooth varying coefficients as proposed by Hastie and Tibshirani (1993). These models can be represented in GAMs by multiplying the smooths by a covariate (in the given application the smooths of time are multiplied by the covariates. See Wood (2017) for more details.

The functions for the smooths are based on P-Splines as proposed by Eilers and Marx (1996), giving low rank smoothers using a B-spline basis using a simple difference penalty applied to the parameters. For the smooth time-varying coefficients on the fixed effects a maximum number of 65 knots is used, combined with a second-order P-spline basis (quadratic splines) and a first-order difference penalty on the coefficients.

The non-linear random smooths are estimated similar to those proposed by Durbán et al. (2005). As a basic idea, one views the individual smooths as splines with random coefficients, i.e. each country has a random effect, that is in fact a function of time that is approximated by regression splines. The parameters of the splines are assumed to be normally distributed with mean zero and the same variance for all curves, which translates into having the same smoothness parameter for all curves. This concept is implemented efficiently in the GAM structure of the `mgcv` package by using the nesting of the smooth within the respective actor. In order to avoid overfitting and keeping computation tractable, a first-order penalty with nine knots is employed. The smoothness selection is done for all smooths by the restricted maximum likelihood criterion (REML).

As the data set is rather big with more than 1.3 million observations in the formation model, the fitting procedure of the model is computationally expensive and was virtually impossible with standard implementations in R before the introduction of the `bam()` function in the `mgcv` package in 2016 that needs less memory and is much faster than other comparable packages. The estimation routine employs techniques as proposed in Wood et al. (2015). Those methods use discretization of covariate values and iterative updating schemes that require only subblocks of the model matrix to be computed at once which allows for the application of parallelization tools.

For all computations we also used the statistical programming language R (R Development Core Team 2008). Important packages used for visualization of networks and computation of network statistics are the `statnet` suite of network analysis packages (Handcock et al. 2008) as well as the package `igraph` (Csardi and Nepusz 2006). For the Tables the `stargazer` package from Hlavac (2013) was employed. For the model evaluation and visualization we used the `PRROC` package of Grau et al. (2015).

A.3. Details on the PCA of the time-varying smooth random effects

For the analysis of the smooth random effects we are following the discretization approach of Ramsay and Silverman (2005, Chap. 8). As noted in Section 3.3 we assume the random effects $\phi_i(t)$ ($\phi_i^+(t)$ and $\phi_i^-(t)$ in the formation and the persistence model, respectively) to be realizations of a stochastic process $\Phi = \{\phi(t), t \in \tau\}$, for $i = 1, \dots, N$ individual countries and $\tau = [1951, 2016]$.

In order to summarize the information provided by these functions we are searching for a weight function $\beta(t)$ that gives us the principal component scores $\phi_i = \int_{\tau} \beta(t) \phi_i(t) dt$. In order to do so, the weight function $\xi_1(t)$ among all possible functions $\beta(t)$ must be found that maximizes $N^{-1} \sum_{i=1}^N (\int_{\tau} \beta(t) \phi_i(t) dt)^2$ subject to the constraint $\int_{\tau} \xi_1^2(t) dt = 1$. From our model we get N individual estimated functions $\hat{\phi}_i(t)$ for all observations (countries) and can discretize the functions $\hat{\phi}_i(t)$ on a grid. We use $T = 100$ equidistant points $\{t_1, \dots, t_{100}\}$ on the interval τ of length $|\tau| = \mathcal{T}$. This gives a discretized $(N \times T)$ time series matrix $\hat{\Phi}$ with N country specific observations in the rows and the estimated functions, evaluated at the discrete time points, in the columns:

$$\hat{\Phi} = \begin{pmatrix} \hat{\phi}_1(t_1) & \cdots & \hat{\phi}_1(t_{100}) \\ \vdots & \ddots & \vdots \\ \hat{\phi}_N(t_1) & \cdots & \hat{\phi}_N(t_{100}) \end{pmatrix}$$

Therefore, in fact we are searching for a solution for the discrete approximation of

$$\int_{\tau} \beta(t) \hat{\phi}_i(t) dt \approx (\mathcal{T}/T) \sum_{j=1}^T \beta(t_j) \hat{\phi}_i(t_j) = \sum_{j=1}^T \tilde{\beta}(t_j) \hat{\phi}_i(t_j)$$

such that the solution $\tilde{\xi}_1$ that maximizes the mean square satisfies $\|\tilde{\xi}_1\|^2 = 1$. This is now a standard problem, with the solution $\tilde{\xi}_1$ being found by the eigenvector that corresponds to the largest eigenvalue of the covariance matrix of $\hat{\Phi}$.

A.4. Out-of-sample-predictions for simulated networks

As a standard principle in network analysis, a model should be able to reflect global network characteristics. We evaluate six of them for our out-of-sample forecasts. The first three characteristics are related to the number of actors that are actively engaged in the arms trade. The statistic *Size* is defined as the count of predicted edges in each year. This measure helps to evaluate the ability of the model to predict amount of realized arms trade in each year. As it is also of interest to measure how dense the predicted arms trade network

is, we include *Density*, relating the size of the network to the number of edges that could have potentially realized. We define the *Order* of the network as number of actors that are engaged in either exporting or importing arms. The results will provide an impression whether the model has the ability not only to classify the right amount of edges (as in *Size*), but also their nesting within the countries.

As we have emphasized the importance of local network statistics we evaluate whether the local network statistics are able to generate the corresponding global statistics. Therefore, we include the *Mean Indegree* (being the same as *Mean Outdegree*), as well as the share of *Reciprocity*. In order to evaluate the accuracy of our predictions with respect to triangular relationships we furthermore include the measure *Transitivity*, that divides the number of triangles by the number of connected triples in the graph. In this statistic, the direction of the edges is ignored. The analysis of this measure gives an impression how well the two chosen transitivity measures capture the overall clustering in the network.

The results are presented in Figure 11. In each of the six panels we see the respective network statistics plotted against time. The solid red line gives the network statistics, evaluated at the real MCW network. The boxplots show the network statistics, evaluated for each year for the 1.000 simulated networks.

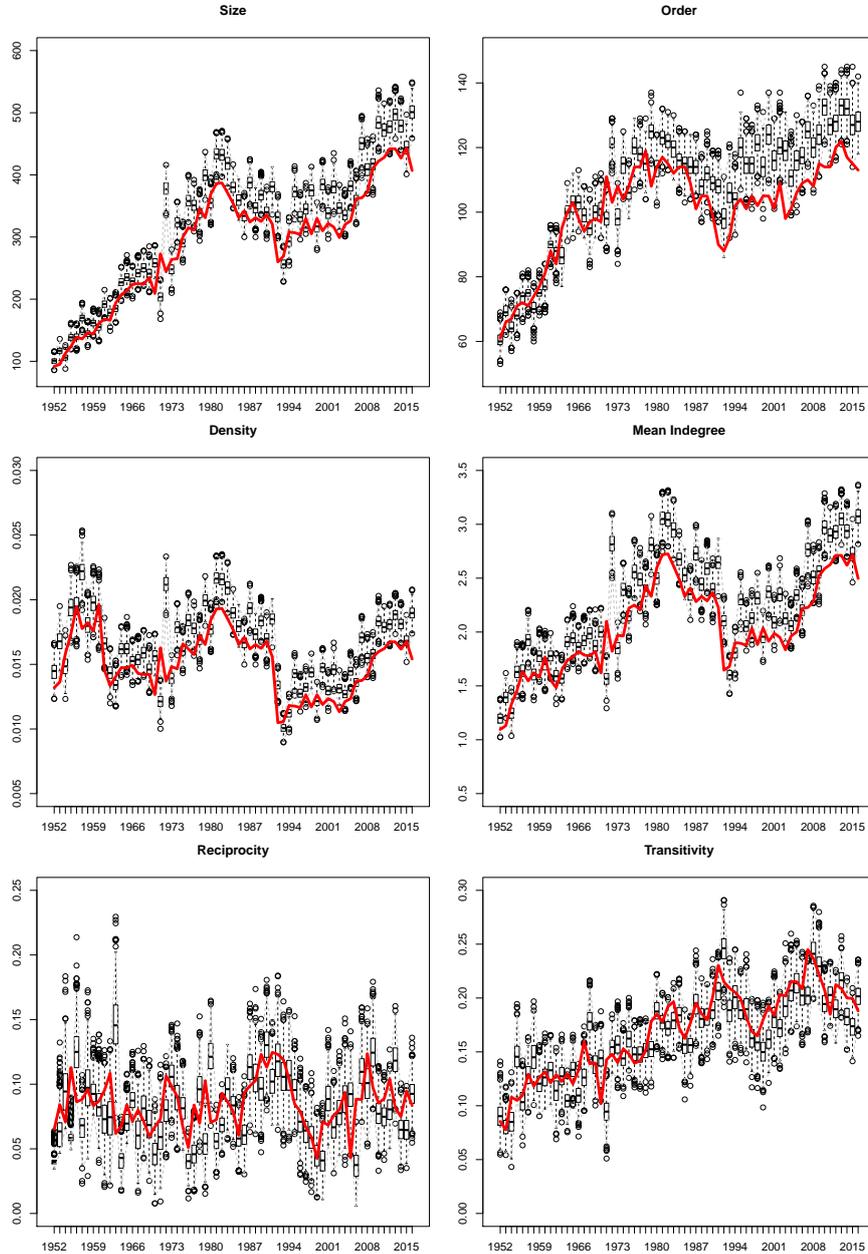


Figure 11: Comparison of realized and simulated network topologies. The boxplots give statistics from the simulated networks. The solid line gives the statistics for the real networks. Number of edges (Size), number of active countries (Order), number of realized transfers relative to possible transfers (Density), average indegree (Indegree), share of reciprocated transfers and ratio of undirected triangles relative to connected triples (Transitivity).

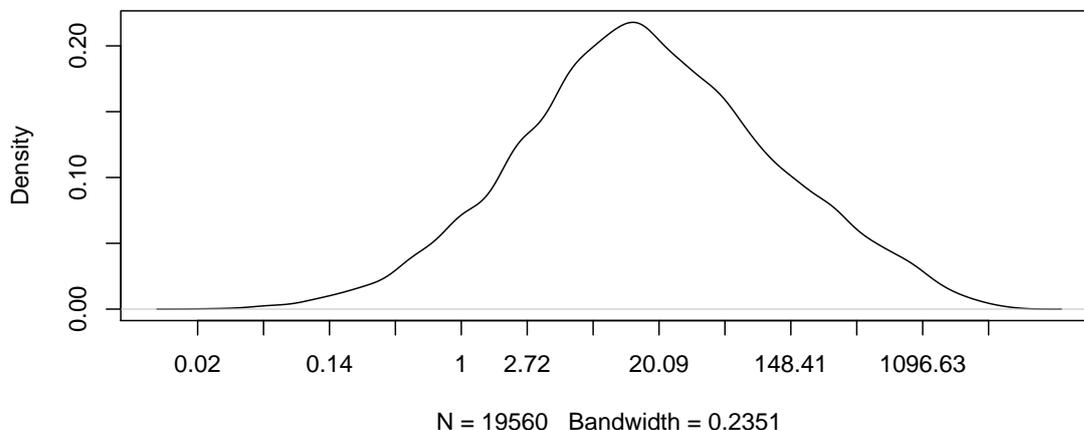


Figure 12: Kernel density estimate (KDE) of arms exports, measured in TIV and pooled over all years from 1950 to 2016. Logarithmic x-axis.

B. Supplementary Material

B.1. Different threshold values

B.1.1. Distribution of TIVs

In Figure 12 we present a kernel density estimate (KDE) of the pooled TIVs for the whole time period. The distribution of the TIVs is highly skewed and has a long tail. Therefore, we give a logarithmic representation. In order to give an impression of the left tail, Table 3 provides the lower quantiles of the distribution. From this it can be seen that roughly 20% of all observations are below a threshold of 3.

0%	4.75%	9.5%	14.25%	19%
0.020	0.700	1.332	2.200	3

Table 3: Lower quantiles of the TIV distribution pooled over all years from 1950 to 2016.

In order to demonstrate the effect of different binarization thresholds on the estimated coefficients we pursue the following strategy. As a baseline we use the "original effects" from the paper with a threshold of zero and plot them in solid black together with two standard error confidence bounds in dark grey. Additionally, we include the estimated coefficients with thresholds incrementing from zero to three in steps of 0.5 as dashed black lines. (i)

By comparison of the solid line with the dashed lines it can be seen how strong the point-estimates vary with different thresholds. (ii) If the dashed lines are within the confidence intervals in dark grey they can be said to be statistically indistinguishable from the original estimates. (iii) Furthermore, we show how the confidence bounds of the new estimates exceed the ones of the original estimation, displayed in light grey. These areas represent the highest upper limit and the lowest lower limit that exceeds the original confidence bounds. Hence, if only the dark grey confidence bound is visible, then the effects of all estimates are the same, if the light grey confidence bound is above and/or below the dark grey, this means that the bounds of the estimates with higher thresholds are wider.

B.1.2. Fixed effects with different thresholds

The degree-related statistics are shown in the top four panels of Figures 13. We find no significant changes of the results as the dashed lines stay in almost all cases within the original confidence bounds. An exception is the senders' outdegree in the formation model but even with the highest threshold this effect does not become significant. On the contrary, partly the results get even more clear. For example the outdegree effect for the receiver in the formation and dissolution model (second row) becomes even more negative in tendency with increasing thresholds.

For the reciprocity effect in the third row we find that the coefficients stay almost the same for all different thresholds of binarization. The same applies for the transitivity effect in the formation model (left panel in the fourth row). Here the upper confidence bound even indicates a potentially higher effect. For the transitivity in the persistence model (right panel in the fourth row) we find that the effect becomes insignificant in the beginning if we set roughly 20% of the lowest observations to zero. Otherwise the effect stays significant and very close to the point estimates of the original estimation.

For the Shared Suppliers Effect (the two panels at the bottom) the results are very similar to the transitivity effects, i.e. we find an potentially stronger effect in the formation model (left panel at the bottom) and an insignificant effect for the first years in the persistence model (right panel) with the highest binarization threshold.

Given that the network statistics are constructed from the network and therefore directly and potentially strongly affected by different thresholds, the robustness of the effects is reassuring. Only if we replace almost one fifth of the existing edges by zeros the effects of the hypderdyadic statistics in the persistence model start to become partly insignificant. However, the affected statistics are in line with our theoretical expectation that the network effects matter mostly for the formation.

From this result it might not come as a surprise that the effects of the non-network related covariates are even more robust because their construction is not affected by the thinning of the network. We can confirm Akerman and Seim (2014) with Figure 14 that shows virtually no noteworthy changes of the effects.

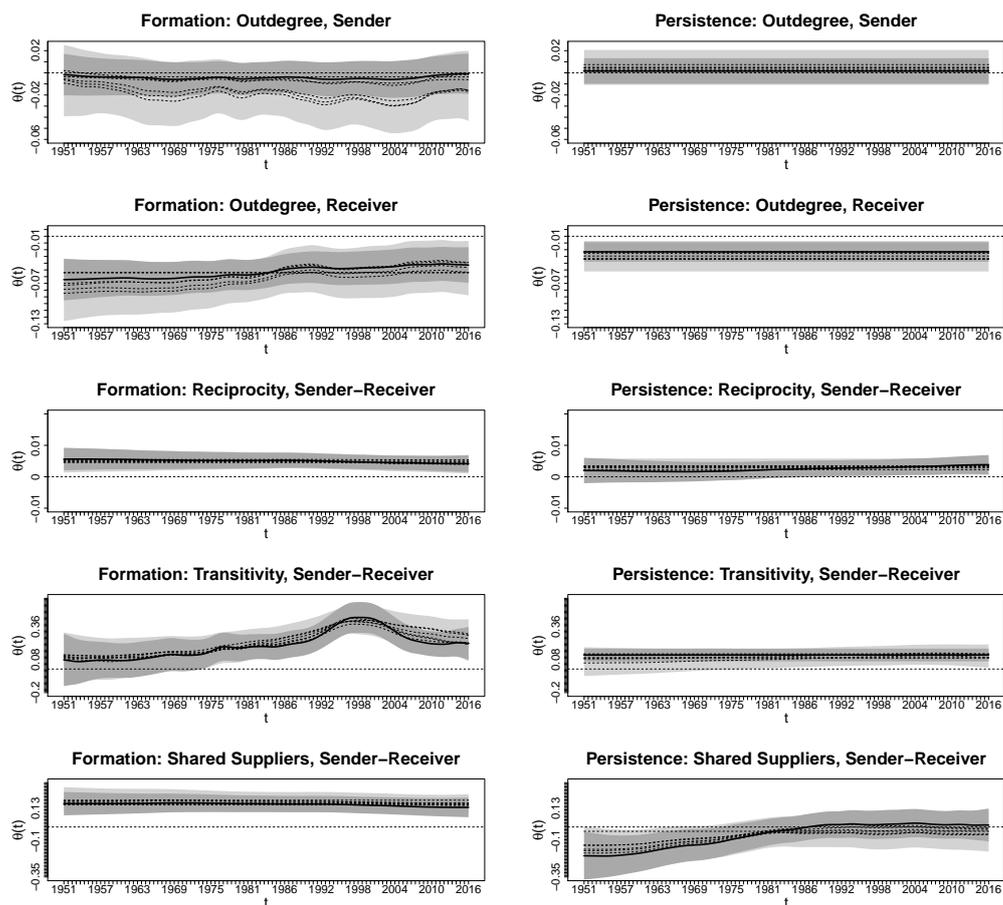


Figure 13: Comparison of fixed effects for network statistics with different binarization thresholds ranging from zero to three, incrementing by 0.5. The estimates with a threshold of zero are given in solid black with dark grey confidence bounds. All other estimates are indicated by dashed lines. Confidence bounds derived from estimates with higher thresholds are in light grey.

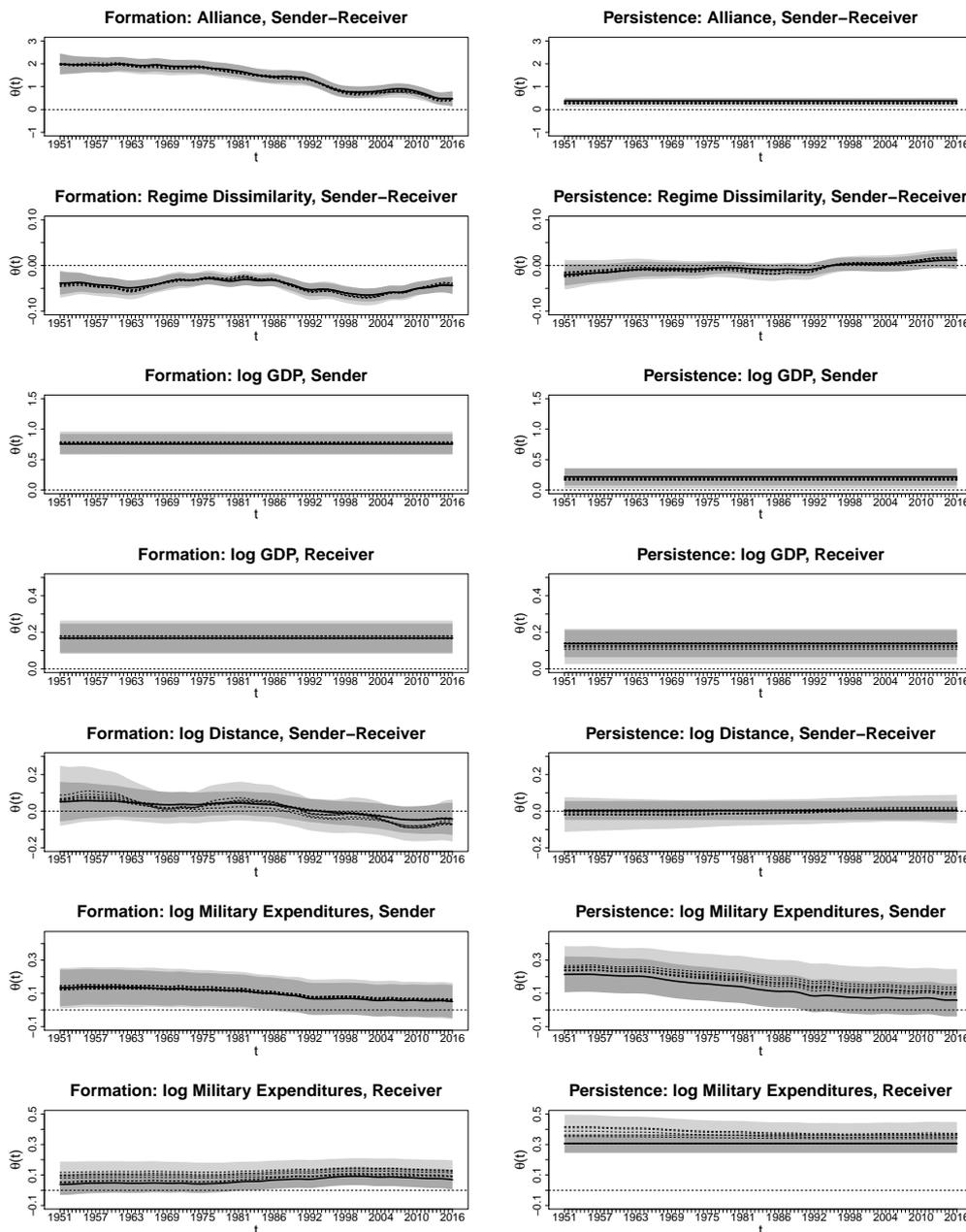


Figure 14: Comparison of fixed effects for economic and political Covariates with different binarization thresholds ranging from zero to three, incrementing by 0.5. The estimates with a threshold of zero are given in solid black with dark grey confidence bounds. All other estimates are indicated by dashed lines. Confidence bounds derived from estimates with higher thresholds in light grey.

B.2. Different time windows

In the paper we assume that the STERGM process applies to two consecutive years. As a robustness check, we define the periods t and $t + 1$ such that they contain multiple years. If we take the years 2013, 2014, 2015 and 2016 as an example for time windows of length two, we set $Y_{ij}^{t-1,t} = 1$ if country i exports arms to j in 2013 or 2014 and $Y_{ij}^{t,t-1} = 0$ if country i has not exported to j neither in 2015 nor in 2016. We also extend this concept such that we combine three years into one period.

For the non-network related covariates we are using the time averages for the respective time windows for continuous variables (e.g. if a period contains two years, the average of the GDP in these two years is taken) and we set binary variables to one if the respective feature was present in all year (e.g. the indicator for a formal alliance is one if the alliance was present in all two or three years).

The corresponding estimates can be seen in Figures 15 and 16. (i) The Figures are constructed such that the baseline is given by the "original effects" from the paper. These effects are plotted in solid black together with two standard error confidence bounds in dark grey. (ii) The coefficients with two or three years within one time period are given in dotted (two years) and dotted-dashed (three years). (iii) The area where the original confidence bounds are exceeded is given in light grey.

In Figure 15 it can be seen that the dotted and dashed-dotted line rarely is outside of the dark grey confidence bound in the panels of rows one to three. One really noteworthy exception applies to the transitivity effect in the formation model. Here the estimates with periods containing three years behave somewhat more time-constant than the original estimates. This is, however, a natural result because with the broader time windows major changes in the data as the collapse of the Soviet Union become more smooth. All in all, the panels clearly show that the coefficients are very robust and do not change fundamentally. This impression is confirmed by Figure 16 where again no notable exceptions can be found influence the interpretation in terms of variation with time or significance. This is again in line with our theoretical expectations since the coefficient for the military expenditures of the receiver is almost not affected by different time windows.

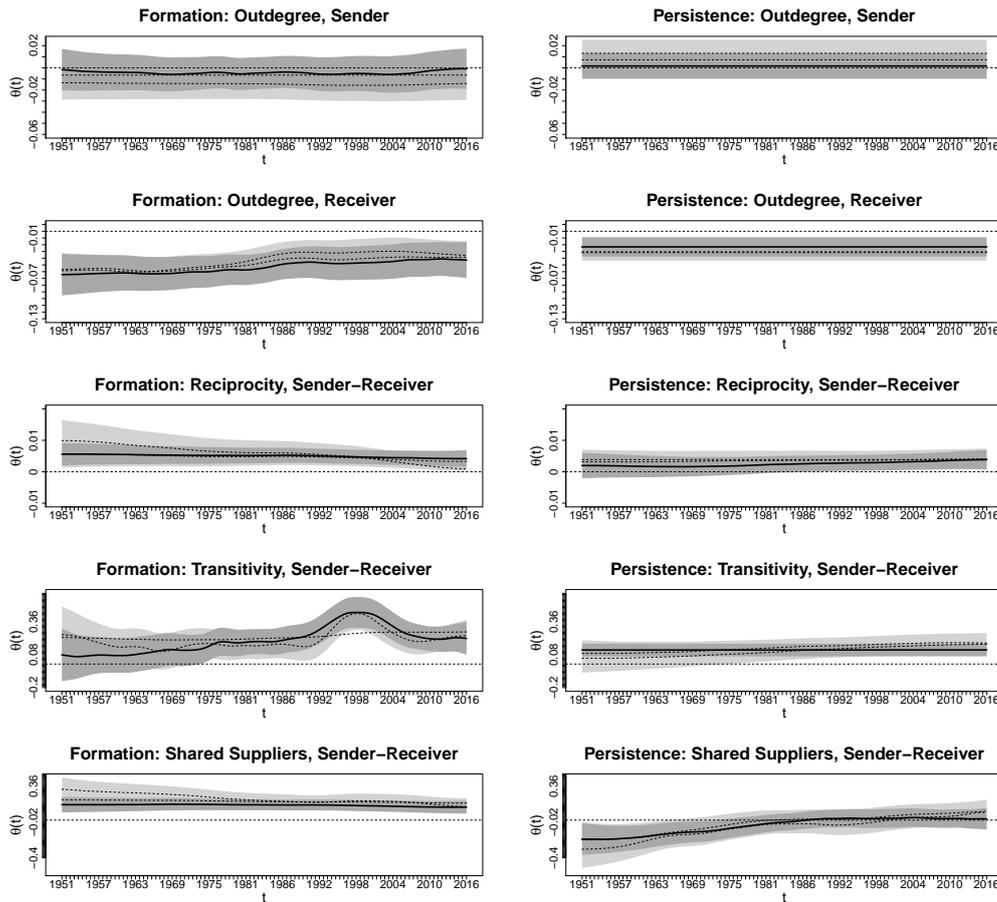


Figure 15: Comparison of fixed effects for network statistics with different time windows. The original estimates are given in solid black with dark grey confidence bounds. The estimates with time windows containing two years are given in dotted and those with three years in dotted-dashed. Confidence bounds derived from estimates with broader time windows in light grey.

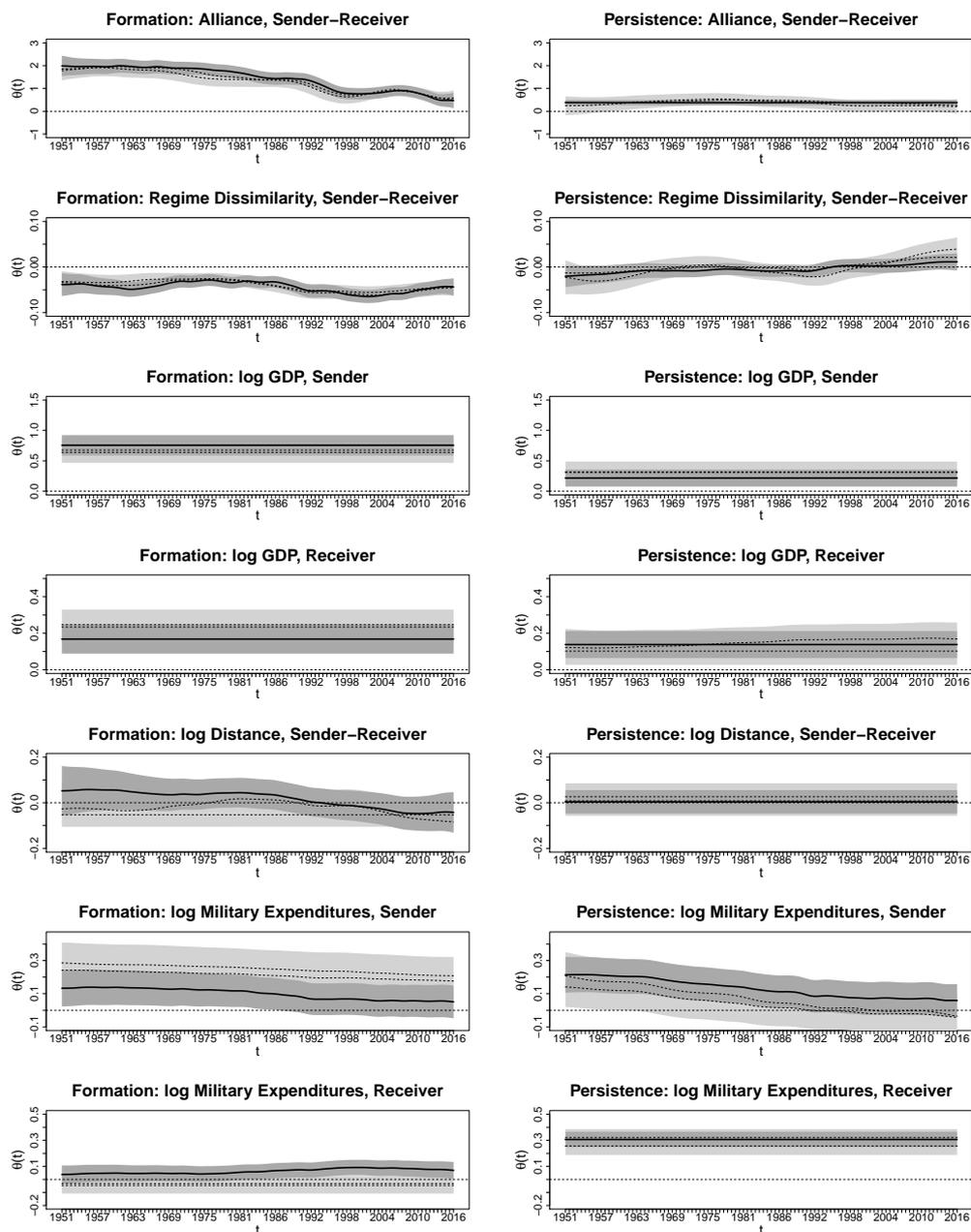


Figure 16: Comparison of fixed effects for economic and political covariates with different time windows. The original estimates are given in solid black with dark grey confidence bounds. The estimates with time windows containing two years are given in dotted and those with three years in dotted-dashed. Confidence bounds derived from estimates with broader time windows in light grey.

B.3. Model without random effects

In the main article we mentioned that the inclusion of the random effects leads to a vanishing global effect of the senders outdegree. I.e once controlled for the sender-specific random effect the coefficient on the outdegree statistic is insignificant. Here, we show that once we exclude random effects from the models, all results are very robust with the exception of the coefficient on the outdegree. This can be seen in Figures 17 and 18 with coefficients that are very comparable to the ones from the main paper. The main exception is given by the senders outdegree (top panels in Figure 17). Here the coefficients are now in both models positive and significant. Especially in the formation effect the coefficient is very high. This shows that there is indeed a global effect of the senders outdegree that, however, vanishes if one controls for country-specific heterogeneity.

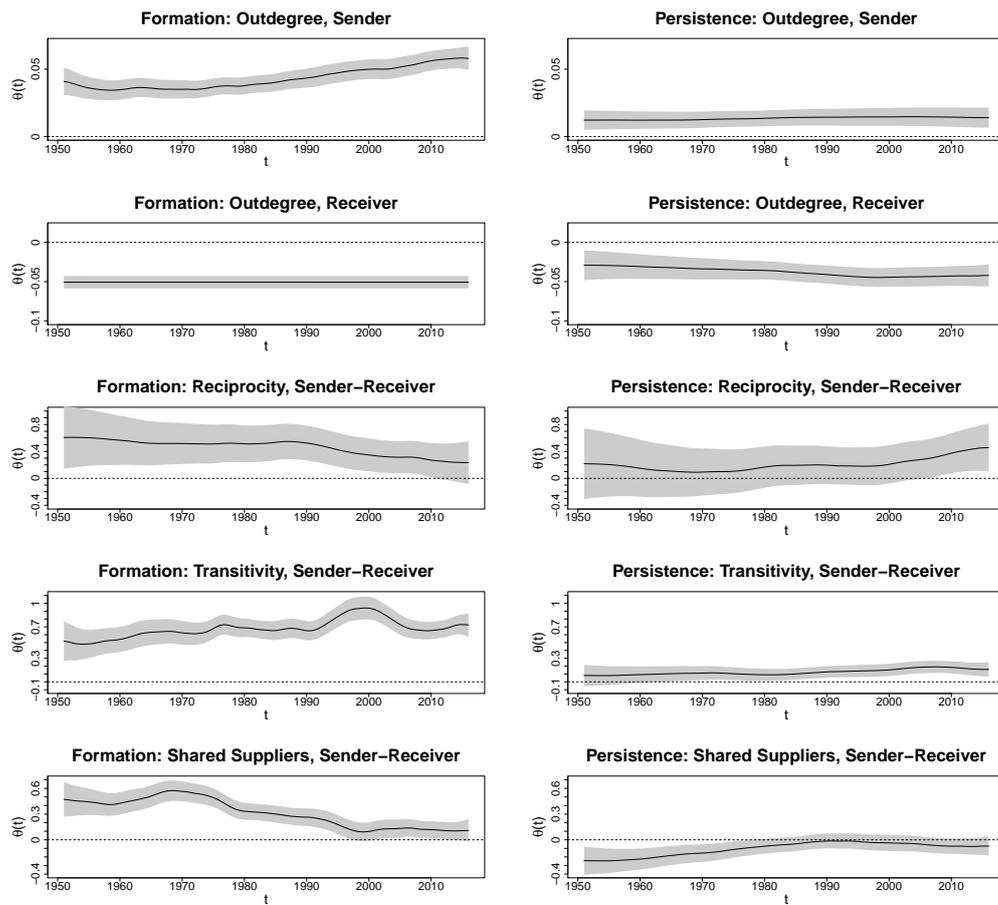


Figure 17: Time-varying coefficients of network statistics without random effects in solid black. Shaded areas give two standard error bounds.

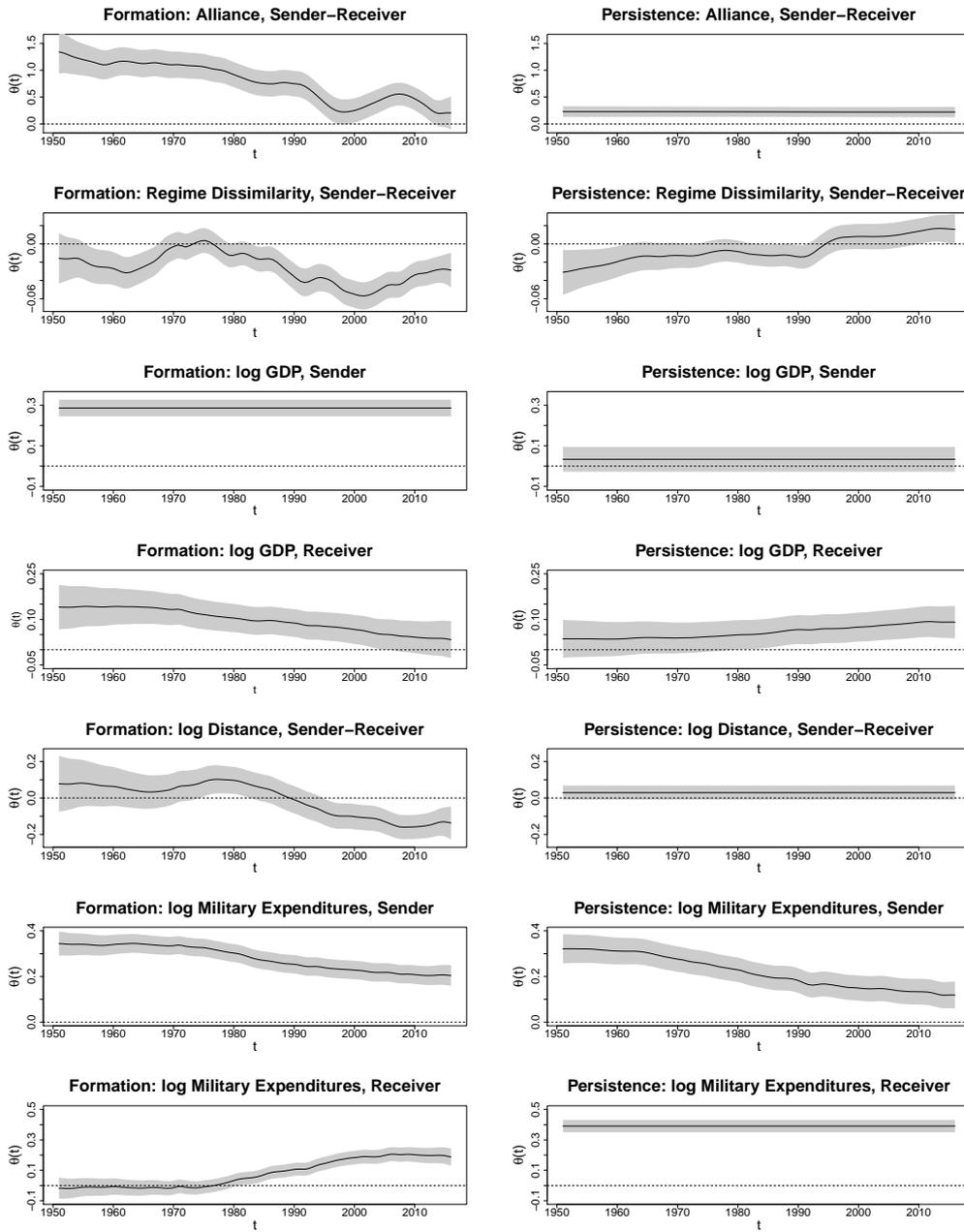


Figure 18: Time-varying coefficients of economic and political covariates without random effects in solid black. Shaded areas give two standard error bounds.

B.4. Comparison of different methods

B.4.1. Theoretical discussion

Besides the STERGM, there are many other models suitable for dynamic networks. In the following, we provide an overview of alternative approaches with a discussion of their suitability for the given dataset.

Latent Space Models: A potential alternative is given by the family of latent space models (e.g. Hoff et al. 2002; Handcock et al. 2007). However, it is hard to model arms trade appropriately within this model class. The need of estimating time-varying coefficients results in yearly separate estimations. The resulting yearly latent space representations are not very stable and do not work well with isolates, i.e. countries without transfers in a given year. If one goes for a panel approach with this model class (see the package `amen` by Hoff et al. 2015) one must accept that the coefficients stay the same for the whole time-period and, even more problematic, that the positions within one latent space are sufficient to capture all network dependencies for the time-period 1950-2016, which is clearly a heroic assumption. On top of that, the latent space approach does not allow for the evaluation of complex network statistics we are interested in.

Stochastic-Actor Oriented Models: Stochastic-Actor Oriented Model (SAOM, see e.g. Snijders et al. 2010) are built for modelling dynamic networks and have the virtue of allowing for the estimation of dyadic and hyper-dyadic network effects. However, the model is tailored for social networks and some of its' assumptions are very problematic for modelling the arms trade network. Most importantly, it assumes actor-homogeneity, an assumption that is clearly violated for the dataset. Additionally, the SAOM fundamentally builds on the idea that the networks observed represent snapshots of a continuous underlying process of edge formation and persistence, i.e. we would need to assume that between t and $t + 1$ multiple changes could have been realized in the network. This is not an acceptable assumption because if a transfer between i and j was recorded in t and $t + 1$, it is not meaningful to assume that there exists an in-between state where the transfer has the change to disappear and re-emerged multiple times. Otherwise, if no transfer is recorded in t and $t + 1$ we have no reason to believe that there was trade in between. In the basic description of the SAOM (Snijders et al. 2010, p. 54) the authors write: "*A foundational assumption of the models discussed in this paper is that the network ties are not brief events, but can be regarded as states with a tendency to endure over time. Many relations commonly studied in network analysis naturally satisfy this requirement of gradual change, such as friendship, trust, and cooperation.*" Apparently it is hard to argue that recording whether there was a transfer between two countries in a given year can be viewed as an enduring state.

Exponential Random Graph Models: Our model is in fact motivated by recent advances within the exponential random graph model (ERGM) family, i.e. the TERGM (Hanneke et al. 2010, Snijders et al. 2010, Leifeld et al. 2018) and the STERGM (Krivitsky and Handcock

2014). However, our model differs from the standard cross-sectional ERGM (but also from the conventional TERGM and STERGM) because it does not allow for simultaneous network dependencies. In general, a static cross-sectional ERGM seems to be an implausible choice for the modelling of a dynamic network with strong actor heterogeneity. A dynamic formation and the TERGM, however, can be constructed in a very similar manner as the STERGM. Those models are natural candidate model for comparison.

B.4.2. Candidate models

In the following we present several alternative candidate models by increasing level of complexity.

Autoregressive ERGM (Model 1): The most simplistic stochastic model is an autoregressive model that assumes time-dependence of all individual dyads such that

$$\log \left\{ \frac{P(Y_{ij}^{t,t-1} = 1 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)}{P(Y_{ij}^{t,t-1} = 0 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)} \right\} = \theta_0 + \theta_1 y_{ij}^{t-1,t}. \quad (4)$$

This temporal dependence structure can be interpreted as a cross sectional ERGM with the lagged response as a dyadic exogenous covariate or as a TERGM with a dyadic stability term (see e.g. Block et al. 2018). It is motivated by the idea that the probability of a transfer in t might change if there was a transfer in $t - 1$.

TERGM with covariates (Model 2): We can change model (4) by including all the network effects and covariates as specified in Section 3.2 of the main paper:

$$\log \left\{ \frac{P(Y_{ij}^{t,t-1} = 1 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)}{P(Y_{ij}^{t,t-1} = 0 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)} \right\} = \theta \tilde{g}_{ij}(y^{t-1,t}, x^{t-1,t}) \quad (5)$$

In this formulation, we include the autoregressive component only indirectly, the lagged network statistics give some information about the network embedding of a transfer but not whether there was a preceding transfer.

TERGM with covariates and random effects (Model 3): With the inclusion of smooth time-varying random effects for the sender and the receiver we have

$$\log \left\{ \frac{P(Y_{ij}^{t,t-1} = 1 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)}{P(Y_{ij}^{t,t-1} = 0 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)} \right\} = \theta \tilde{g}_{ij}(y^{t-1,t}, x^{t-1,t}) + \phi_{i, \text{sender}}(t) + \phi_{j, \text{receiver}}(t). \quad (6)$$

The main difference to the STERGM is now that we do not model the processes of formation and persistence separately but within one model and that we do not include the information on the lagged response here, that is implicitly included in the STERGM mechanics.

Number	Model Name	lagged edge	covariates	rand. eff.	PR %	ROC %
1	Autoreg. ERGM	yes	no	no	0	0
2	TERGM	no	yes	no	0	0
3	TERGM	no	yes	yes	0	1.54%
4	TERGM with dyadic stability	yes	yes	no	4.62%	6.15%
5	TERGM with dyadic stability	yes	yes	yes	18.46%	44.62%
6	STERGM	implicit	yes	no	24.62%	6.15%
7	STERGM	implicit	yes	yes	52.31%	41.54%

Table 4: Different dynamic network models from the ERGM family included in the comparison of prediction. Name of the models in the second column. Model specification in columns three, four and five. The share of years where the respective model performed best according to the AUC of the PR or the ROC curve are given in the two rightmost columns.

TERGM with covariates and dyadic stability (Model 4): The inclusion of the lagged response to the TERGM with covariates gives

$$\log \left\{ \frac{P(Y_{ij}^{t,t-1} = 1 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)}{P(Y_{ij}^{t,t-1} = 0 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)} \right\} = \theta \tilde{g}_{ij}(y^{t-1,t}, x^{t-1,t}) + \theta_1 y_{ij}^{t-1,t}. \quad (7)$$

TERGM with network effects, dyadic stability and random effects (Model 5): And as a last step we allow for sender- and receiver-specific random effects in the TERGM with network statistics and lagged response:

$$\log \left\{ \frac{P(Y_{ij}^{t,t-1} = 1 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)}{P(Y_{ij}^{t,t-1} = 0 | Y^{t-1,t} = y^{t-1,t}, X^{t-1,t} = x^{t-1,t}; \theta)} \right\} = \theta \tilde{g}_{ij}(y^{t-1,t}, x^{t-1,t}) + \theta_1 y_{ij}^{t-1,t} + \phi_{i, \text{sender}}(t) + \phi_{j, \text{receiver}}(t). \quad (8)$$

STERGM without random effects (Model 6): Together with our main model from the paper (Model 7), the STERGM with random effects we additionally include a STERGM without random effects. The formal representations are given in the main paper in equations (4) and (5).

B.4.3. Comparison of out-of-sample predictions

We evaluate the proposed models in the following way. In a first step, we fit coefficients based on the information of $t - 1$ to the response in t . In the second step we use the fitted models in order to predict the edges in $t + 1$. As a result we obtain probabilistic out-of-sample predictions. The evaluation is done with area under the curve (AUC) measures for the Receiver-Operating-Characteristic (ROC) curve and the Precision-Recall (PR) curve.

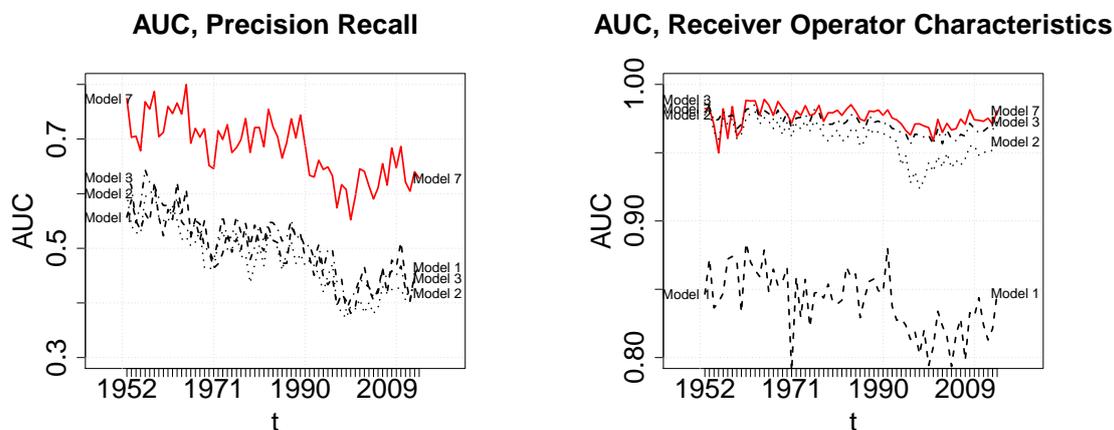


Figure 19: AUC values for out-of-sample predictions based on precision recall (left) and receiver-operator-characteristics (right). The STERGM with covariates and random effects (Model 7) in solid and red, the autoregressive ERGM (Model 1) and the TERGM with covariates (Model 2) and random effects (Model 3) in black and dashed.

All models are fitted using the package `mgcv` (Wood 2017; Version 1.8-24) and evaluated with the package `PRROC` (Grau et al. 2015; Version 1.3).

Table 4 gives an overview of the predictive performance of all models included. In the two rightmost columns we present the share of years where the respective model has the highest out-of-sample predictive power. It can be seen that evaluated by the PR, the STERGM with random effects clearly represents the superior model with the highest predictive performance. Judged by the ROC the performance of the TERGM and the STERGM (both with random effects and as a model class) are very similar. We give a detailed description of the results below.

In order to give a clear impression of how the out-of-sample fits are related we provide multiple plots where the baseline model, the STERGM with covariates and random effects, is always indicated in solid red.

In Figure 19 we compare the AUC values of the autoregressive ERGM (Model 1), the TERGM with covariates (Model 2) and with random effects (Model 3) with the baseline STERGM model (Model 7). The Precision Recall AUC values are shown on the left hand side of Figure 19 and provide a clear message since all selected candidate models have AUC values clearly below the STERGM used in the paper (Model 7). Looking at the AUC values from the ROC measure (right panel) shows that the two TERGM models (Models 2 and 3) come partly close to the baseline model in the first years of the observational period, while the simplistic autoregressive ERGM has considerable lower AUC values. However as

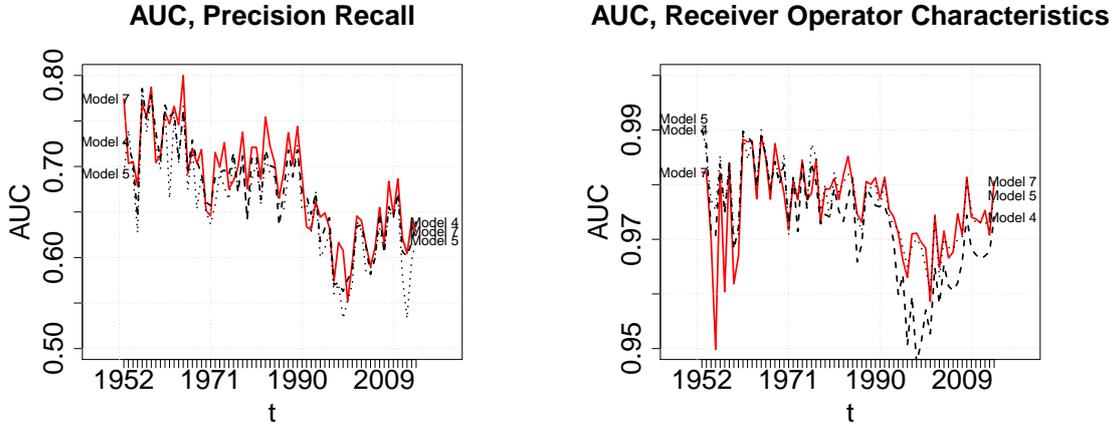


Figure 20: AUC Values for out-of-sample predictions based on precision recall (left) and receiver-operator-characteristics (right). The STERGM with covariates and random effects (Model 7) in solid and red, the TERGM with covariates, dyadic stability (Model 4) and random effects (Model 5) in black and dashed.

a general picture the STERGM with random effects is clearly the superior model.

Including the lagged response as explanatory variable makes the TERGM models (Models 4 and 5) pretty similar to our baseline model (Model 7). Naturally, this is reflected in Figure 20. However, the STERGM model is more flexible because it allows for different coefficients for the processes of formation and persistence. This fact leads to superior predictions of the STERGM model (Model 7) in the left panel of Figure 20. The red line provides the upper boundary in most time points, but there are some instances where the autoregressive TERGMs (Models 4 and 5) provide the better predictions when evaluated with the PR curve. If the AUC measure for the ROC curve is compared, we see again that the predictions of the TERGM partly outperform the STERGM. Nevertheless, this is mostly the case in the beginning of the observational period and it seems like the superiority of the STERGM increases with the more recent periods.

In Figure 21 the predictive performance of the STERGM with (Model 7) and without (Model 6) random effects is compared. Although the AUC values are very close to each other, the model that includes the random effects clearly provides better out-of-sample predictions. On the right hand side of Figure 21 the contrast becomes visible more clearly and in almost all years our baseline model provides the better predictions.

All in all we conclude the following. (i) The STERGM gives, among all other candidate models the best predictions when judged by Precision Recall, being the more important measure when predicting rare vents. (ii) Furthermore, the STERGM has a much richer

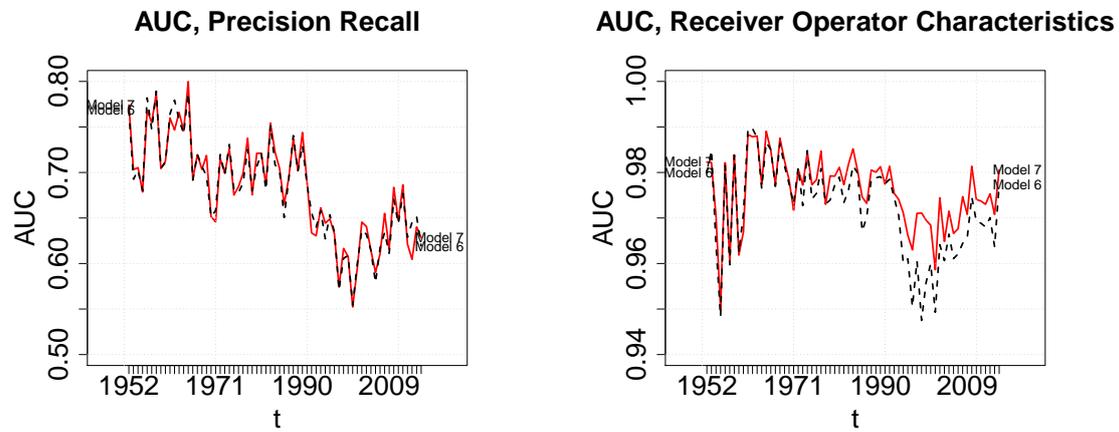


Figure 21: AUC values for out-of-sample predictions based on precision recall (left) and receiver-operator-characteristics (right). The STERGM with covariates and random effects (Model 7) in solid and red, the STERGM with covariates but without random effects (Model 6) in black and dashed.

interpretation than the TERGM and (iii) the random effects provide an substantial benefit to the inferential part of the model. We therefore conclude that the choice of the STERGM with random effects seems to be very appropriate regarding both, the predictive performance as well as the ability to gain new insights.

Chapter 3

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

Contributing Article:

Cornelius Fritz, Michael Lebacher and Göran Kauermann (2019): *Tempus volat hora fugit - A survey of tie-oriented dynamic network models in discrete and continuous time*. Under review in *Statistica Neerlandica*.¹

arXiv preprint <https://arxiv.org/abs/1905.10351>

Code at https://github.com/lebachelor/tempus_volat_hora_fugit

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.

¹The article was accepted after handing in the thesis and is now available in a revised version at <https://doi.org/10.1111/stan.12198>.

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

Cornelius Fritz*, Michael Lebacher and Göran Kauermann
Department of Statistics, Ludwig-Maximilians-Universität München

Abstract

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, i.e., 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. The data allow to demonstrate the applicability and interpretation of the network models.

Keywords — Continuous-Time, Discrete-Time, Event Modeling, ERGM, Random Graphs, REM, STERGM, TERGM

*cornelius.fritz@stat.uni-muenchen.de

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, while ties are connections between those nodes. Dependent on the context, these connections can represent friendships in a school (Raabe et al., 2019), transfers of goods between countries (Ward et al., 2013), sexual relations between people (Bearman et al., 2004) or hyperlinks between websites (Leskovec et al., 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 and 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, e.g., 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 et al. (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 et al., 2002) and p_1 models (Holland and Leinhardt, 1981), all originating in the Edös-Rényi-Gilbert random graph models (Erdös and Rényi, 1959; Gilbert, 1959). Kim et al. (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 et al. (2018) provide an in-depth comparison of the Temporal Exponential Random Graph Model (TERGM, Hanneke et al., 2010) and the SAOM with special focus on the treatment of time. Further, the ERGM and SAOM for networks which are observed at single time points are contrasted by Block et al. (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

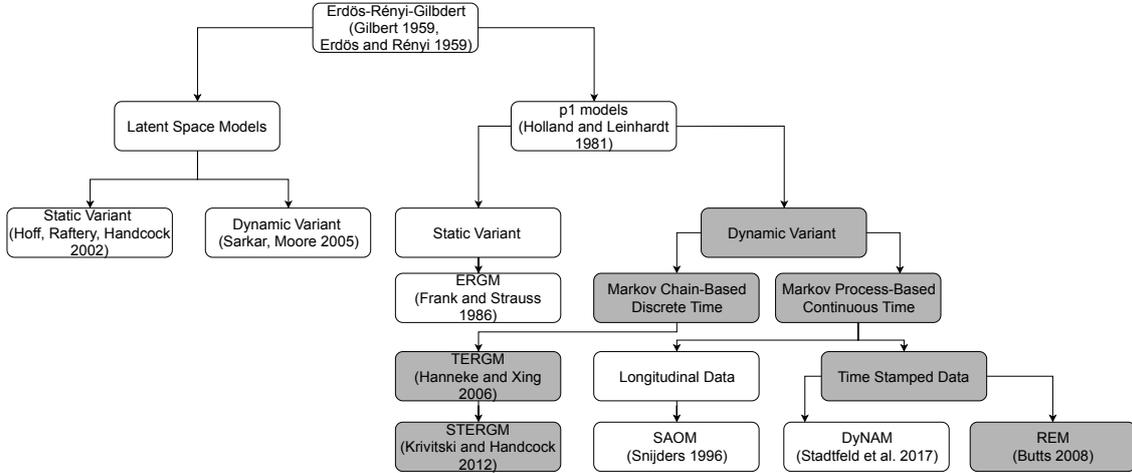


Figure 1: Tree diagram summarizing the dependencies between models originating in the Erdős-Rényi-Gilbert graph model, models situated in a box with a grey background are discussed in this article. This graph is an update of Figure 6.1 in Goldenberg et al. (2010).

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. In this endeavor, we emphasize reviewing tie-oriented models. Tie-oriented models are concerned with formulating a stochastic model for the existence of a tie contrasting the actor-oriented approach by Snijders (2002), which specifies the model from the actors point of view (Block et al., 2018). The Dynamical Actor-Oriented model (DyNAM, Stadtfeld and Block, 2017) adopts this actor-oriented paradigm to event data. This type of model was formulated with a focus on social networks (Snijders, 1996). Contrasting this, tie-oriented models can be viewed as more general, since they are also applicable to non-social 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 and Handcock, 2014). There exists a variety of recent applications of the TERGM. White et al. (2018) use a TERGM for modeling epidemic disease outcomes and Blank et al. (2017) investigate interstate conflicts. In He et al. (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 Broekel and Bednarz (2019) 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 et al., 2013), inter-hospital patient transfers (Vu et al., 2017), online learning platforms (Vu et al., 2015), and animal behavior (Tranmer et al., 2015) to structures of project teams (Quintane et al., 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 and 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, since the interpretation and estimation with a longer series of networks is unchanged. Lastly, the comparison of the methods at hand in a clear-cut manner is hence enabled.

The paper is structured as follows. In Section 2 we give basic definitions that are used throughout the paper and present the two data examples that will be analyzed as illustrative examples. After that, Section 3 introduces time-discrete and Section 4 time-continuous network models. 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. Further, 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

		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

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

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. Further, all sub-scripted temporal indices (Y_t) are assumed to take discrete 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, i.e., the time-stamped information is not observed. Whereas the second application, a network of email traffic, comes in time-stamped format, that 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 (MCW) 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 and Nepusz, 2006) and Figure 2 visualizes the arms trade network using the software **Gephi** (Bastian et al., 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

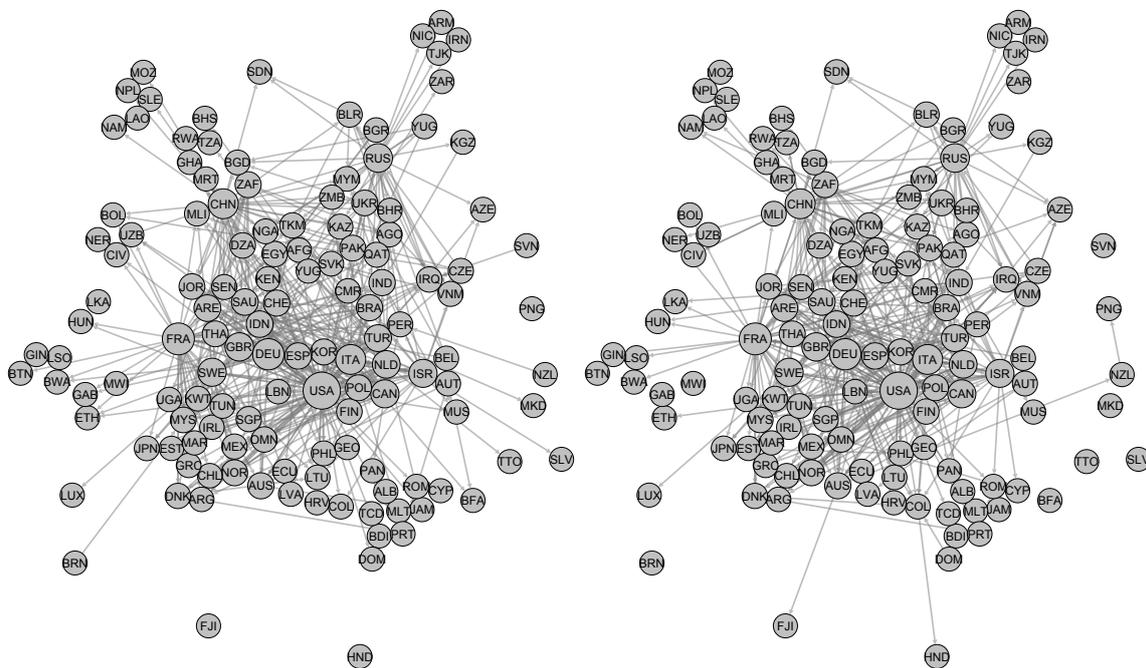


Figure 2: The International Arms Trade as a binary network in 2016 (left) and 2017 (right). Nodes that are isolated in both years are not depicted for clarity and the node size relates to the sum of the out- and in-degree. The labels of the nodes are the ISO3 codes of the respective countries.

continue trading with the same partners.

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, 2017) 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 non-stochastic 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 Paranjape et al., 2017, Email EU Core, 2019). 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 contains $n = 89$ persons and is 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, 8068 events are recorded and 4031 in the second period. 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 Annex A for the visualization of the degree distributions of both applications.

3 Dynamic Exponential Random Graph Models

3.1 Temporal Exponential Random Graph Model

The Exponential Random Graph Model (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 et al., 2012, Robins et al., 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 as well as 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

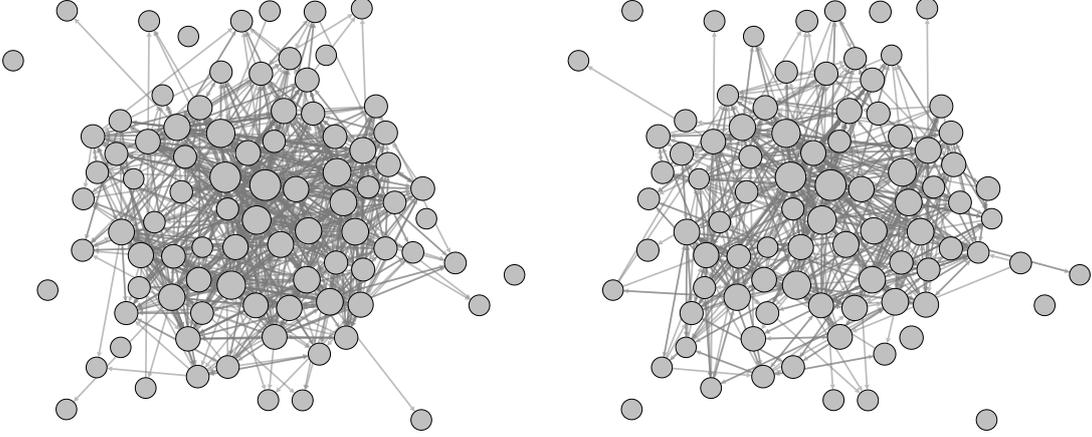


Figure 3: The European research institution email correspondence aggregated to a binary network divided into *Period 1* (day 1-365, left) and *Period 2* (day 1-365, right). The node size relates to the sum of the out- and in-degree.

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

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

In the formulation above, it is assumed that the joint distribution can be decomposed into yearly transitions from Y_{t-1} to Y_t . Further, 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 over 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 et al., 2018) or by using smooth time-varying effects (e.g. Lebacher et al., 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 et al., 2008 for more examples). However, the statistics $s(y_t, y_{t-1}, x_t)$ explicitly allow temporal interactions, e.g. delayed reciprocity

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

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

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

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 non-existent ties in $t - 1$ remain non-existent in t . The proportionality sign is used since 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, e.g., time-varying covariates $x_{ij,t}$

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

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 $\kappa(\theta, y_{t-1}, x_t) = \sum_{\tilde{y} \in \mathcal{Y}} \exp\{\theta^T s(\tilde{y}_t, y_{t-1}, x_t)\}$, that 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 and Thompson (1992) for Monte Carlo max-

imum likelihood and Hummel et al. (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 and Ikeda, 1990) that can be viewed as a local alternative to the likelihood (van Duijn et al., 2009) but is often regarded as unreliable and poorly understood in the literature (Hunter et al., 2008, Handcock et al., 2003). However, the MPLE is asymptotically consistent (Desmarais and Cranmer, 2012) and the often suspect standard errors can be corrected via bootstrap (Leifeld et al., 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 for example Almquist and 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 this problems is the inclusion of modified statistics, called *geometrically weighted statistics* (Snijders et al., 2006). Using the definitions of Hunter (2007), the *geometrically weighted out-degree distribution* (GWOD) controls for the out-degree distribution with one statistic, via

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

with $O_k(y_t)$ being the number of nodes with out-degree k in t and α_O as 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) :

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

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 et al., 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 Temporal Exponential Random Graph Model

An 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 , since $Y_t = Y^+ \setminus (Y_{t-1} \setminus Y^-) = Y^- \cup (Y^+ \setminus Y_{t-1})$. 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 $t-1$:

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

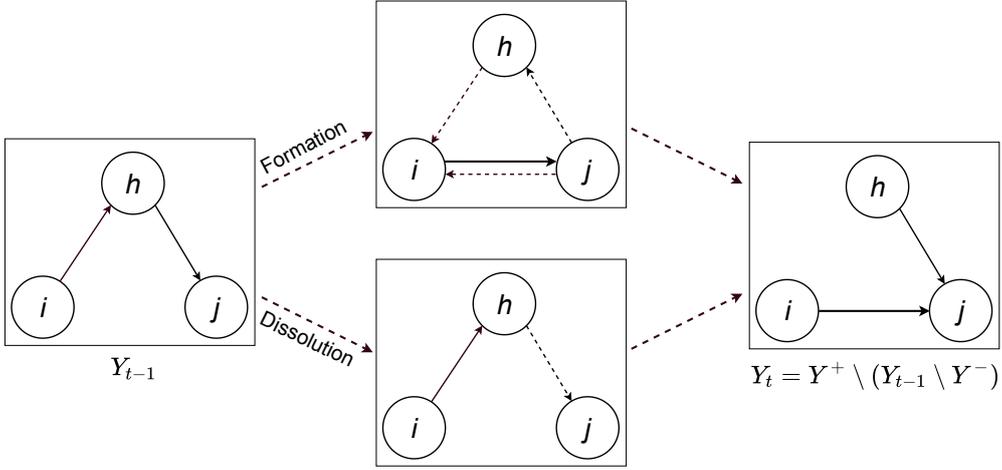


Figure 4: Conceptual representation, illustrating formation and dissolution in the STERGM.

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 in dashed and the actual formation in this example (i, j) is shown in 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. Inserting the separable models in (9) makes it apparent, that the STERGM is a subclass of the TERGM since

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

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 since 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

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

and in case of the dissolution model, y^+ is simply exchanged with y^- . Similarly, and 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 et al., 2015). A comparison between different models is possible using, for example, the Akaike Information Criterion (AIC, Claeskens and 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 asses 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 the simulated

values of it 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 and Limnios, 2018 as a basic reference for stochastic processes). The idea was originally introduced by Holland and Leinhardt (1977). According to their view, 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 which are given as triplets $\omega = (i, j, \tilde{t})$ and encode the 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, i.e., 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 and 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 actor i and j at a given time point can still be viewed as an event that has an one-to-one analogy to a tie in the classical framework.

The overall aim of Relational Event Models (REM, Butts, 2008) is to understand the dynamic structure of events conditional on the history of events (Lerner et al., 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 and Prentice, 2002 for an

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

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

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}) . Since 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 et al. (2011) extended the model by postulating a semi-parametric intensity similar to Cox (1972):

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

where $\lambda_0(\tilde{t})$ is an arbitrary baseline intensity, $\theta \in \mathbb{R}^p$ the parameter vector and $s_{ij}(N(\tilde{t}), x(\tilde{t}))$ 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, since 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

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

where $\mathbb{I}(\cdot)$ 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, e.g., want to control for the out-degree of the sender the corresponding tie-oriented statistic is:

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

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:

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

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. Since 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 pre-specified time window, e.g. 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 conditional on the past independent. 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} the interval of observation, the likelihood can be written as:

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

This likelihood is straightforward to maximize 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 semi-parametric 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:

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

Consecutively, $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ can be estimated with a Nelson Aalen estimator (see Kalbfleisch and 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 and 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 $\mathcal{P}\mathcal{L}(\theta)$ in (15).

Extensions of this model building on already well-established methods in social network and time-to-event analysis were numerously proposed. Perry and Wolfe (2013) used a stratified Cox model in (13). Stadtfeld et al. (2017) adopted the Stochastic Actor oriented Model (SAOM) to events. DuBois and Smyth (2010) and DuBois et al. (2013) extended the Stochastic Block Model (SBM) for time-stamped relational events. Further, DuBois et al. (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 country 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 Peto, 1972; Breslow, 1974). 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:

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

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). Since 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 and 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 et al. (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, since 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 and Kleinberg, 2007). One example of the former option is the usage of *Schönfeld residuals* by Vu et al. (2017) to check the assumption of proportional intensities, which is central to semi-parametric 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 et al., 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 pseudo-likelihood 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 et al., 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 tem-

poral 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, while 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. Due to no compositional changes, we did not scale any statistics. Moreover, we refer to the Supplementary Material for the model assessment.

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., 2018). Here we want to comment on a few aspects only. While we do not have timestamps for the arms trades, the longitudinal networks can still be viewed as time-clustered observations enabling the techniques from Section 4.3.

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 in- and out-degree 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), while they are insignificant effect 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, since 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 weapons, which again results in a centralized network structure as indicated by the estimates in the TERGM and STERGM.

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 while 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

	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	–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	1917.666	1366.654	532.849		
\sum AIC	1917.666	1899.503			

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). 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 $\log(2)$ and the Kalbfleisch-Prentice approximation was used with 100 random orderings of the events to find the estimates of the REM.

see that the AIC value of the STERGM is smaller than of the TERGM.

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.

	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	–1723.732	–1000.506	–505.431		
AIC	3459.464	2011.012	1020.862		
\sum AIC	3459.464	3031.874			

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). 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 $\log(2)$.

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, while 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 struc-

ture of the email network is again better explained by the STERGM. In the Supplementary Material we fit 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 micro-steps. 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 actors 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, that was successfully applied to a wide array of network data, e.g., in Sociology (Agneessens and Wittek, 2012; de Nooy, 2002), Political Science (Kinne, 2016; Bichler and Franquez, 2014), Economics (Castro et al., 2014), and Psychology (Jason et al., 2014). Estimation of this model variant is predominantly carried out with the R package `RSiena` (Ripley et al., 2013).

Another notable model that can be regarded as a bridge between the ERGM and continuous-time models is the Longitudinal ERGM (LERGM, Snijders and Koskinen, 2013; Koskinen et al., 2015). In contrast to the TERGM, the LERGM assumes that the network evolves in micro-steps 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 put emphasis on 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 smaller sub-networks 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 TERGM and STERGM take a global and REM a local point-of-view, which results in substantially different interpretations of the estimates.

Furthermore, we analyzed two data sets that represent two types of network data that are traditionally either modeled by the TERGM and REM. 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.

Acknowledgement

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 the 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 funding.

References

- Agneessens, F. and R. Wittek (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. and C. T. Butts (2014). Logistic Network Regression for Scalable Analysis of Networks with joint Edge/Vertex Dynamics. *Sociological methodology* 44(1), 273–321.
- Bastian, M., S. Heymann, and M. Jacomy (2009). Gephi: An open source software for exploring and manipulating networks.
- Bearman, P., J. Moody, and K. Stovel (2004). Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology* 110(1), 44–91.
- Benton, R. A. and J. You (2017). Endogenous dynamics in contentious fields: Evidence from the shareholder activism network, 2006–2013. *Socius* 3, 2378023117705231.
- Bichler, G. and J. Franquez (2014). *Conflict Cessation and the Emergence of Weapons Supermarkets*, pp. 189–215. Cham: Springer International Publishing.
- Blank, M., M. Dincecco, and Y. M. Zhukov (2017). Political regime type and warfare: evidence from 600 years of european history. *Available at SSRN 2830066*.

- Block, P., J. Koskinen, J. Hollway, C. Steglich, and C. Stadtfeld (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks* 52, 180 – 191.
- Block, P., C. Stadtfeld, and T. Snijders (2019). Forms of Dependence: Comparing SAOMs and ERGMs From Basic Principles. *Sociological Methods & Research* 48(1).
- Breslow, N. (1974). Covariance Analysis of Censored Survival Data. *Biometrics* 30(1), 89–99.
- Broekel, T. and M. Bednarz (2019, Jan). 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*.
- 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(5939), 414–416.
- Butts, C. and C. Marcum (2017). A Relational Event Approach to Modeling Behavioral Dynamics. *ArXiv e-prints*.
- Caimo, A. and N. Friel (2011). Bayesian inference for exponential random graph models. *Social Networks* 33(1), 41–55.
- Caimo, A. and N. Friel (2014). Bergm: Bayesian exponential random graphs in r. *Journal of Statistical Software* 61(2), 1–25.
- Castro, I., C. Casanueva, and J. L. Galán (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>. Accessed: 2017-06-02.
- Claeskens, G. and N. L. Hjort (2008). *Model selection and model averaging*. Cambridge: Cambridge University Press.
- Cox, D. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2), 187–220.
- Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5), 1–9.

- de Nooy, W. (2002). The dynamics of artistic prestige. *Poetics* 30(3), 147 – 167.
- Desmarais, B. A. and S. J. Cranmer (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications* 391(4), 1865–1876.
- DuBois, C., C. Butts, D. McFarland, and P. Smyth (2013). Hierarchical models for relational event sequences. *Journal of Mathematical Psychology* 57(6), 297 – 309. Social Networks.
- DuBois, C., C. Butts, and P. Smyth (2013). Stochastic blockmodeling of relational event dynamics. In C. M. Carvalho and P. Ravikumar (Eds.), *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, Volume 31 of *Proceedings of Machine Learning Research*, Scottsdale, Arizona, USA, pp. 238–246. PMLR.
- DuBois, C. and P. Smyth (2010). Modeling Relational Events via Latent Classes. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, New York, NY, USA, pp. 803–812. ACM.
- Email EU Core (2019). Email EU core temporal, Department 3. <https://snap.stanford.edu/data/email-Eu-core-temporal-Dept3.txt.gz>. Accessed: 2019-08-05.
- Erdős, P. and A. Rényi (1959). On Random Graphs i. *Publicationes Mathematicae Debrecen* 6, 290.
- Geyer, C. J. and E. A. Thompson (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Statist. Soc. B*, 657–699.
- Gilbert, E. N. (1959, 12). Random graphs. *Ann. Math. Statist.* 30(4), 1141–1144.
- Girardin, V. and N. Limnios (2018). *Applied Probability* (2 ed.). Heidelberg: Springer.
- Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoldi (2010). A Survey of Statistical Network Models. *Foundations and Trends® in Machine Learning* 2(2), 129–233.
- Goodreau, S. M., M. S. Handcock, D. R. Hunter, C. T. Butts, and M. Morris (2008). A statnet Tutorial. *Journal of Statistical Software* 24(9), 1.
- 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.
- Grau, J., I. Grosse, and J. Keilwagen (2015). PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* 31(15), 2595–2597.

- Handcock, M. S., G. Robins, T. Snijders, J. Moody, and J. Besag (2003). Assessing degeneracy in statistical models of social networks. Technical report, Citeseer.
- Hanneke, S., W. Fu, E. P. Xing, et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* 4, 585–605.
- He, X., Y. bo Dong, Y. ying Wu, G. rui Jiang, and Y. Zheng (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., 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 (1977). A dynamic model for social networks. *The Journal of Mathematical Sociology* 5(1), 5–20.
- Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *J. Am. Statist. Ass.* 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. (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. *J. Am. Statist. Ass.* 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.
- Jason, L. A., J. M. Light, E. B. Stevens, and K. Beers (2014). Dynamic Social Networks in Recovery Homes. *American Journal of Community Psychology* 53(3-4), 324–334.
- Kalbfleisch, J. and R. Prentice (2002). *The Statistical Analysis of Failure Time Data*. Wiley-Blackwell.
- Kim, B., K. H. Lee, L. Xue, and X. Niu (2018). A review of dynamic network models with latent variables. *Statist. Surv.* 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: Springer Science & Business Media.
- 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.
- Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *J. R. Statist. Soc. B* 76(1), 29–46.
- Langholz, B. and O. Borgan (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* 82(1), 69–79.
- Lebacher, M., P. W. Thurner, and G. Kauermann (2019, Mar). A Dynamic Separable Network Model with Actor Heterogeneity: An Application to Global Weapons Transfers. *arXiv e-prints*.
- Leifeld, P., S. J. Cranmer, and B. A. Desmarais (2018). Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. *Journal of Statistical Software* 83(6), doi: 10.18637/jss.v083.i06.
- Lerner, J., M. Bussmann, T. 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 (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., K. J. Lang, A. Dasgupta, and M. W. Mahoney (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. and J. Kleinberg (2007). The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* 58(7), 1019–1031.
- Lusher, D., J. Koskinen, and G. Robins (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.

- Marcum, C. and C. Butts (2015). Constructing and Modifying Sequence Statistics for relevant Using informR in R. *Journal of Statistical Software* 64(5), 1–36.
- Morris, M., M. S. Handcock, and D. R. Hunter (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software* 24(4), 1548.
- Paranjape, A., A. R. Benson, and J. Leskovec (2017). Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 601–610. ACM.
- Perry, P. and P. Wolfe (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 (Methodological)* 34(2), 187–220.
- Quintane, E., P. Pattison, G. Robins, and J. Mol (2013). Short- and long-term stability in organizational networks: Temporal structures of project teams. *Social Networks* 35(4), 528 – 540.
- Raabe, I. J., Z. Boda, and C. Stadtfeld (2019). The Social Pipeline: How Friend Influence and Peer Exposure Widen the STEM Gender Gap. *Sociology of Education* 92(2), 105–123.
- Ripley, R., K. Boitmanis, and T. Snijders (2013). RSiena: Siena - Simulation Investigation for Empirical Network Analysis. <https://CRAN.R-project.org/package=RSiena>. R package version 1.2-12.
- 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, Y. Kalish, and D. Lusher (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks* 29(2), 173–191.
- Salathé, M., D. Q. Vu, S. Khandelwal, and D. Hunter (2013). The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science* 2(1), 4.
- Sarkar, P. and A. W. Moore (2006). Dynamic Social Network Analysis using Latent Space Models. In Y. Weiss, B. Schölkopf, and J. C. Platt (Eds.), *Advances in Neural Information Processing Systems* 18, pp. 1145–1152. MIT Press.
- Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association* 106(496), 1361–1370.

- SIPRI (2019). Arms Transfers Database. <https://www.sipri.org/databases/armstransfers>. Accessed: 2019-03-01.
- 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*, pp. 215–247. Structural Analysis in the Social Sciences. Cambridge University Press.
- Snijders, T. (2017). Comment: Modeling of coordination, rate functions, and missing ordering information. *Sociological Methodology* 47(1), 41–47.
- Snijders, T. and J. Koskinen (2013). *Longitudinal Models*, pp. 130–140. Structural Analysis in the Social Sciences. Cambridge University Press.
- Snijders, T., J. Koskinen, and M. Schweinberger (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics* 4(2), 567.
- Snijders, T., 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. (2012). *Events in social networks : a stochastic actor-oriented framework for dynamic event processes in social networks*. Ph. D. thesis.
- 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. and J. Hollway (2018). *goldfish: Goldfish – Statistical network models for dynamic network data*. R package version 1.2.
- Stadtfeld, C., J. Hollway, and P. Block (2017). Dynamic Network Actor Models: Investigating Coordination Ties through Time. *Sociological Methodology* 47(1), 1–40.
- Stansfield, S. E., J. E. Mittler, G. S. Gottlieb, J. T. Murphy, D. T. Hamilton, R. Detels, S. M. Wolinsky, L. P. Jacobson, J. B. Margolick, C. R. Rinaldo, et al. (2019). Sexual role and hiv-1 set point viral load among men who have sex with men. *Epidemics* 26, 68–76.
- 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 (2018). Network Interdependencies and the Evolution of the International Arms Trade. *Journal of Conflict Resolution*.
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Tranmer, M., C. S. Marcum, B. Morton, D. Croft, and S. de Kort (2015). Using the relational event model (rem) to investigate the temporal dynamics of animal social networks. *Animal Behaviour* 101, 99 – 105.
- 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.
- Vu, D., L. Alessandro, M. Daniele, and P. Francesca (2017). Relational event models for longitudinal network data with an application to interhospital patient transfers. *Statistics in Medicine* 36(14), 2265–2287.
- Vu, D., A. Asuncion, D. Hunter, and P. Smyth (2011). Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pp. 857–864.
- Vu, D., D. Hunter, P. Smyth, and A. Asuncion (2011). Continuous-Time Regression Models for Longitudinal Networks. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 24*, pp. 2492–2500. Curran Associates, Inc.
- Vu, D., P. Pattison, and G. Robins (2015). Relational event models for social learning in MOOCs. *Social Networks* 43, 121–135.
- Wang, P., G. Robins, and P. Pattison (2006). Pnet: A program for the simulation and estimation of exponential random graph models. *University of Melbourne*.
- Ward, M., J. S. Ahlquist, and A. Rozenas (2013). Gravity’s Rainbow: A Dynamic Latent Space Model for the World Trade Network. *Network Science* 1.
- Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press.
- White, L. A., J. D. Forester, and M. E. Craft (2018). Covariation between the physiological and behavioral components of pathogen transmission: Host heterogeneity determines epidemic outcomes. *Oikos* 127(4), 538–552.

World Bank (2017). World Bank Open Data, Real GDP. <http://data.worldbank.org/>.
Accessed: 2017-04-01.

A Annex: Additional Descriptives

Figures 5 and 6 depict the distributions of in- and out-degrees in the two networks. Building on in- and out-degree 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 and 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, while a small percentage of countries accounts 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 receiver no emails. For the out-degree, this number doubles (roughly 20% have not sent emails). Further, the distribution of the out-degree is more skewed than the one for the in-degree.

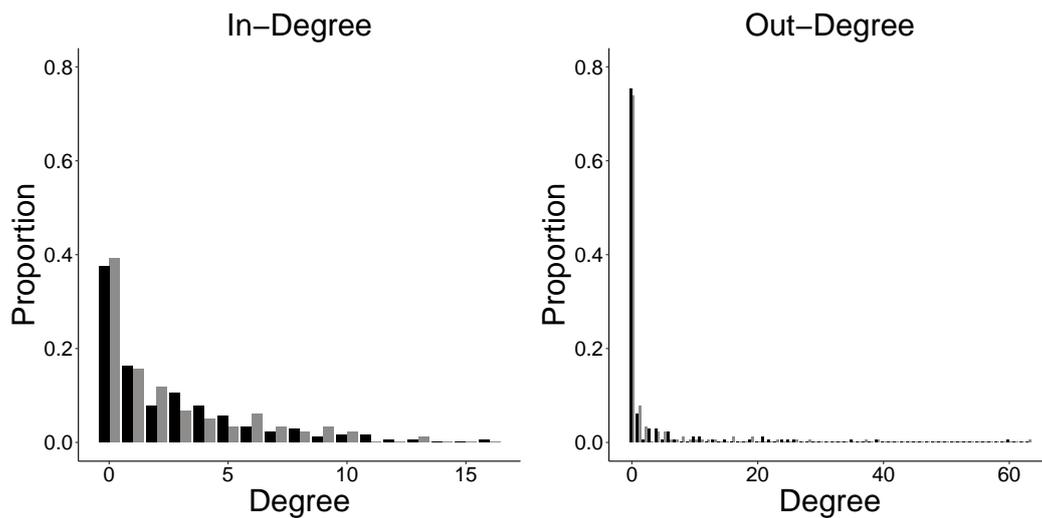


Figure 5: Arms trade network: Barplots indicating the distribution of the in- and out-degrees. Black bars indicate the values of year 2016 and grey bars of 2017.

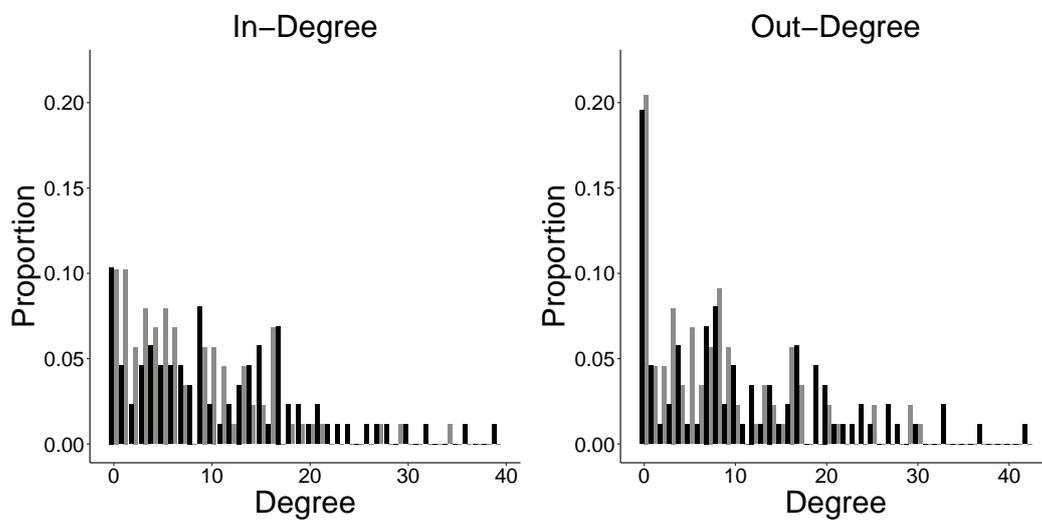


Figure 6: Email exchange network: Barplots indicating the distribution of the in- and out-degrees. Black bars indicate the values of period 1 and grey bars of period 2.

B Supplementary Material

B.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 4: Countries included in the analysis of the international trade network with the ISO3 codes, that are used in the graphical representations of the network.

B.2 Simulation-based Goodness-of-fit in (S)TERGMs

In Figures 7 and 10 we show simulation-based goodness-of-fit (GOF) diagnostics for the TERGM model and in Figures 8, 9, 11 and 12 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_t while the STERGM regards y^+ and y^- . Overall, all plots indicate a satisfying fit of the respective models.

B.2.1 Data Set 1: International Arms Trade

Goodness-of-fit diagnostics

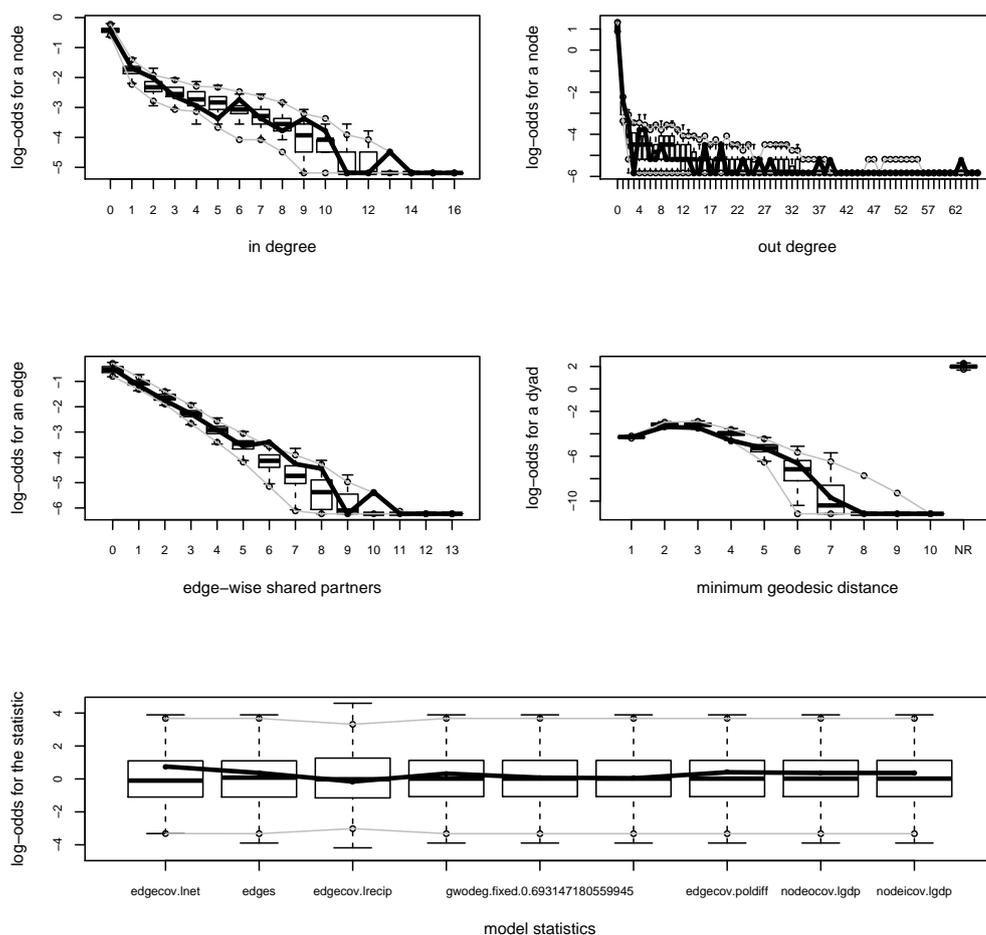


Figure 7: Arms trade network: Simulation-based goodness-of-fit diagnostics in the TERGM. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of a node 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.

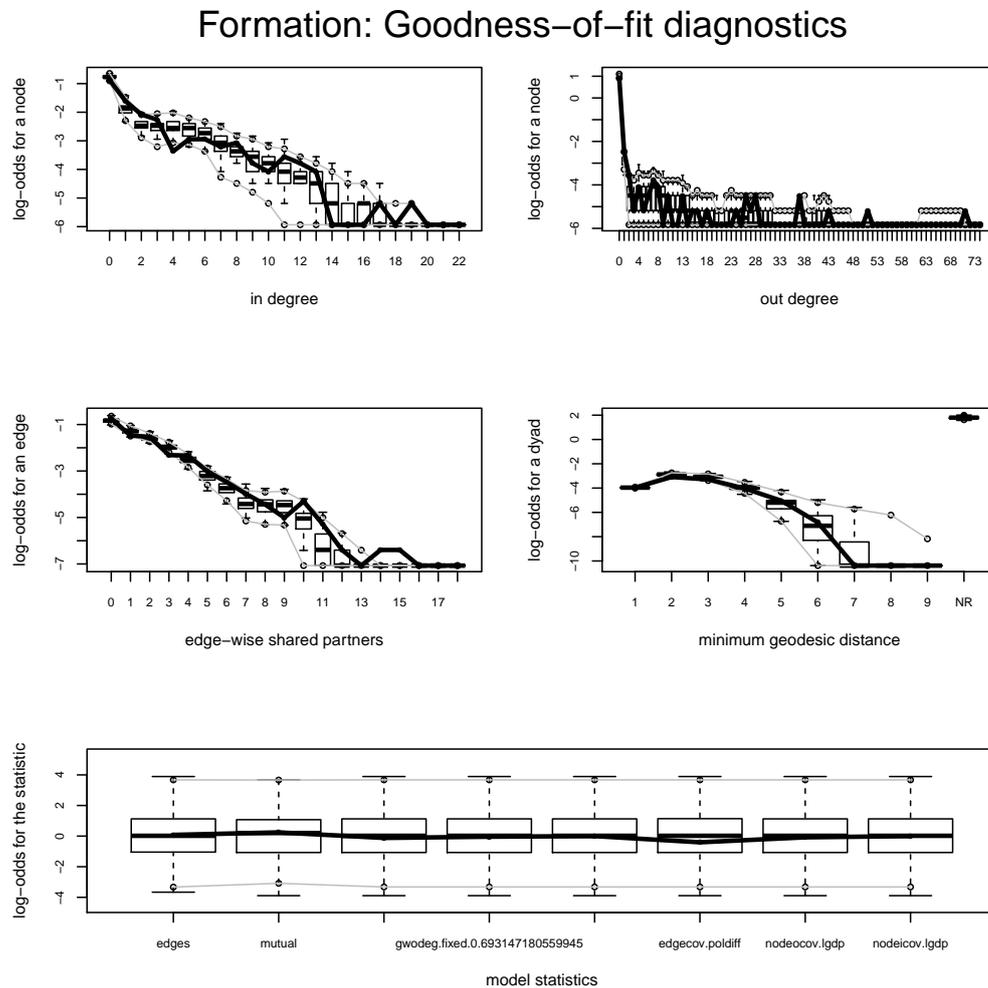


Figure 8: Arms trade network: Simulation-based goodness-of-fit diagnostics in the STERGM for the formation model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give 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.

Dissolution: Goodness-of-fit diagnostics

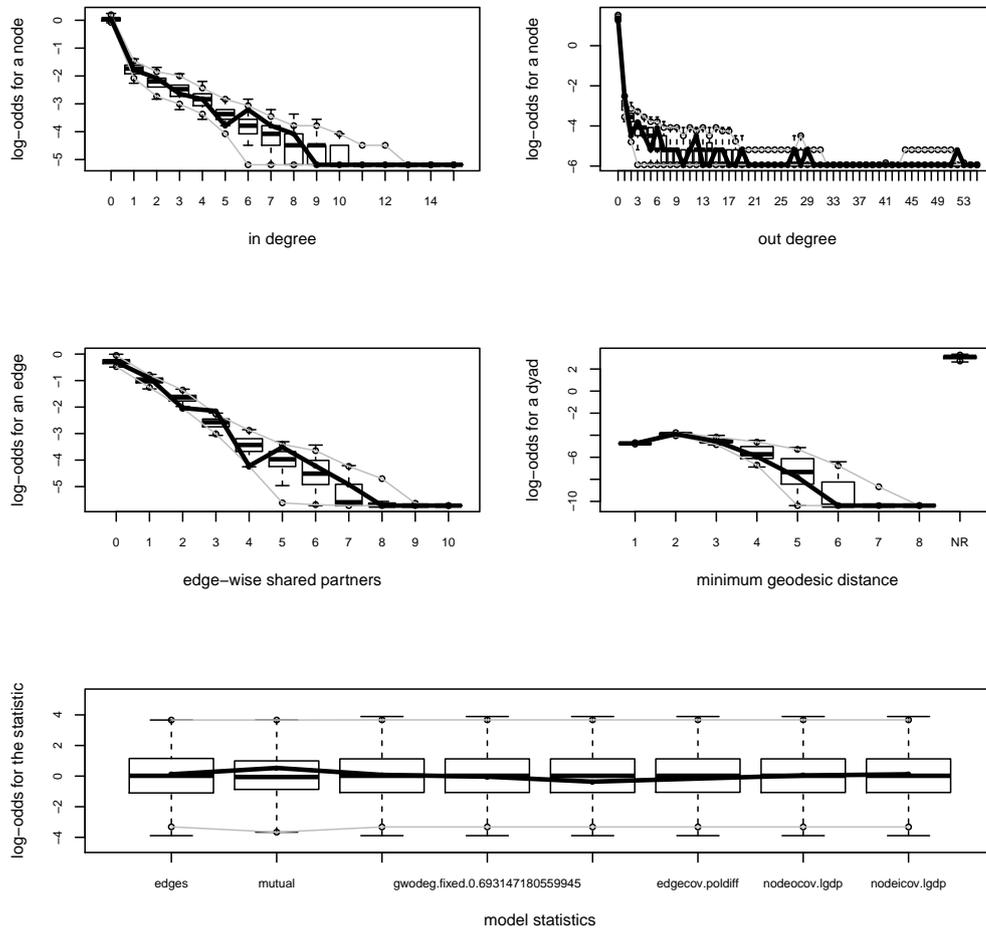


Figure 9: Arms trade network: Simulation-based Goodness-of-fit diagnostics in the STERGM for the dissolution model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels 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.

B.2.2 Data Set 2: European Research Institution Email Correspondence

Goodness-of-fit diagnostics

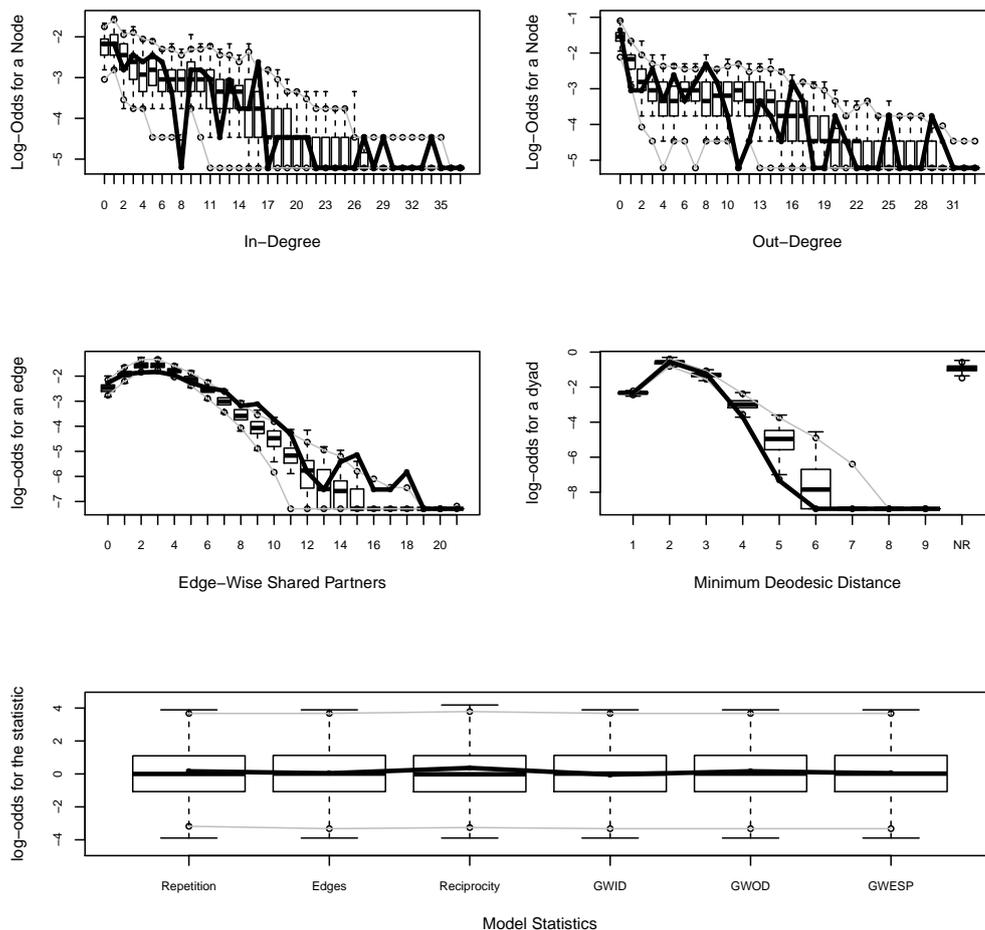


Figure 10: Email exchange network: Simulation-based Goodness-of-fit diagnostics in the TERGM. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels give the log-odds of 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.

Formation: Goodness-of-fit diagnostics

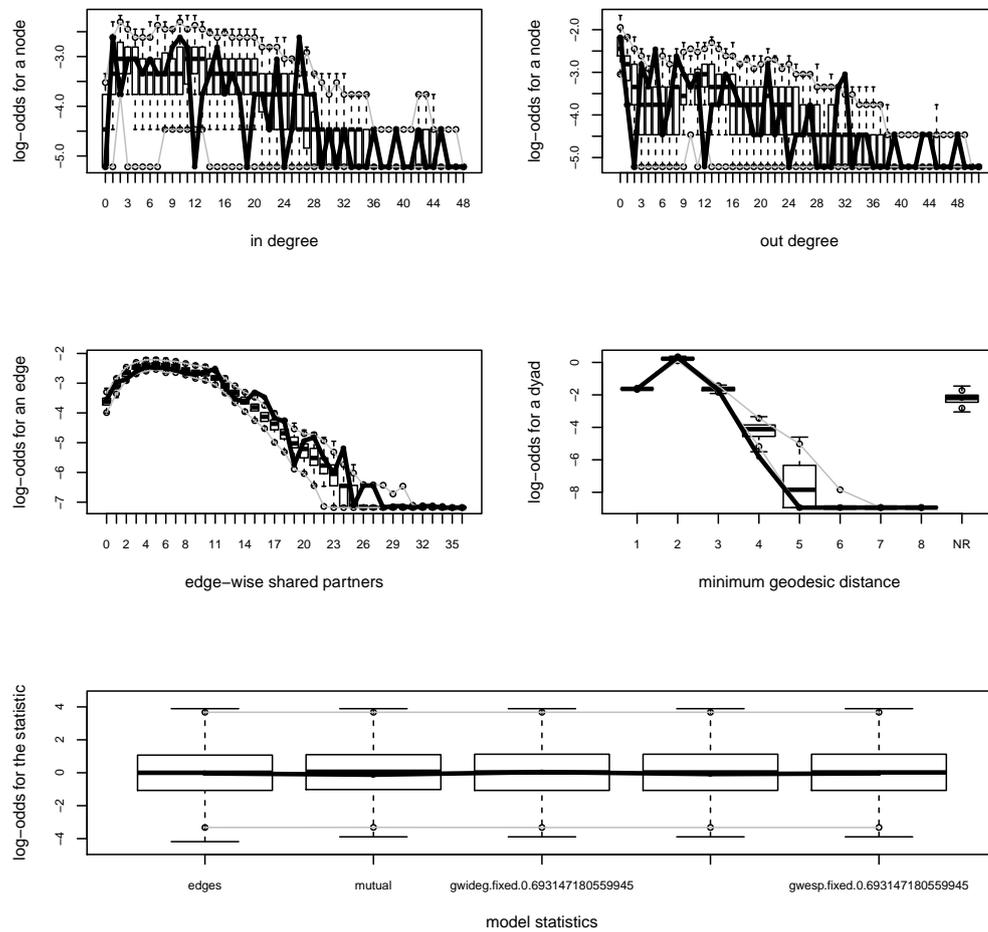


Figure 11: Email exchange network: Simulation-based Goodness-of-fit diagnostics in the STERGM for the formation model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels 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.

Dissolution: Goodness-of-fit diagnostics

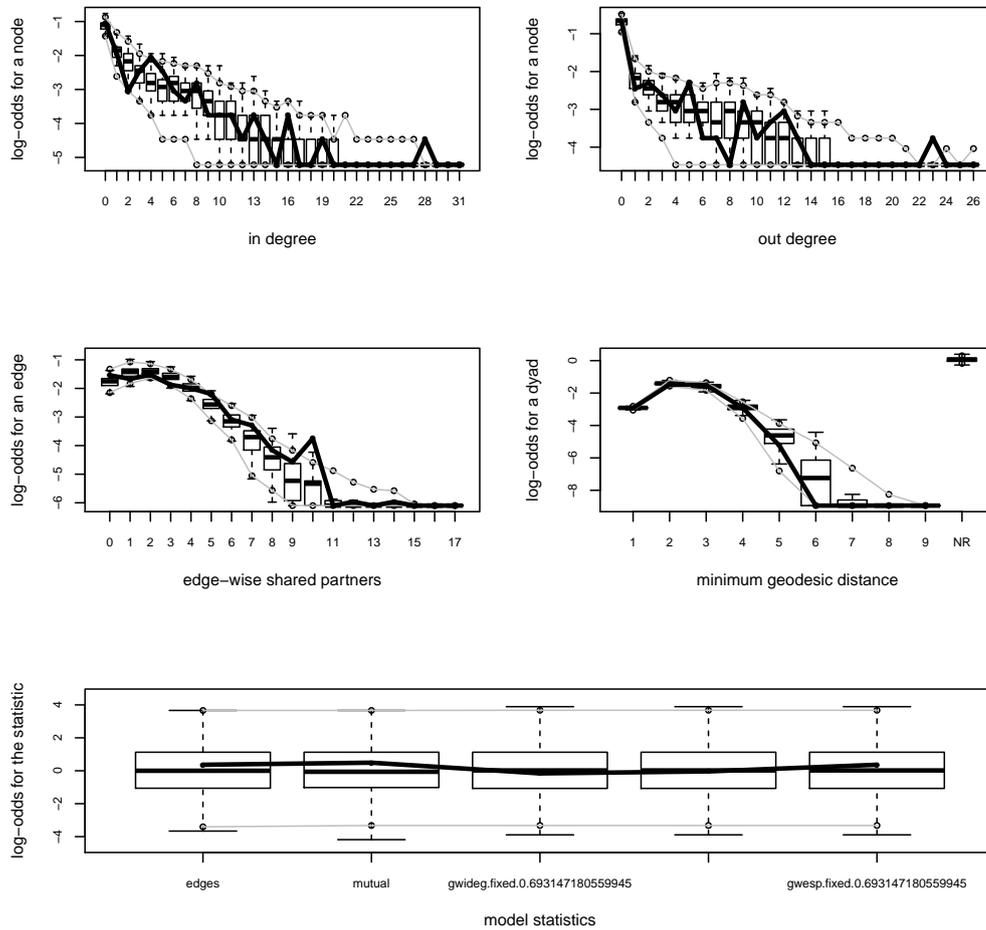


Figure 12: Email exchange network: Simulation-based Goodnes-of-fit diagnostics in the STERGM for the dissolution model. Boxplots give the evaluations of the respective network characteristics at the simulated networks and the solid line gives the actual values from the observed network. First four panels 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.

B.3 ROC-based Goodness-of-fit

B.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 13. 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.

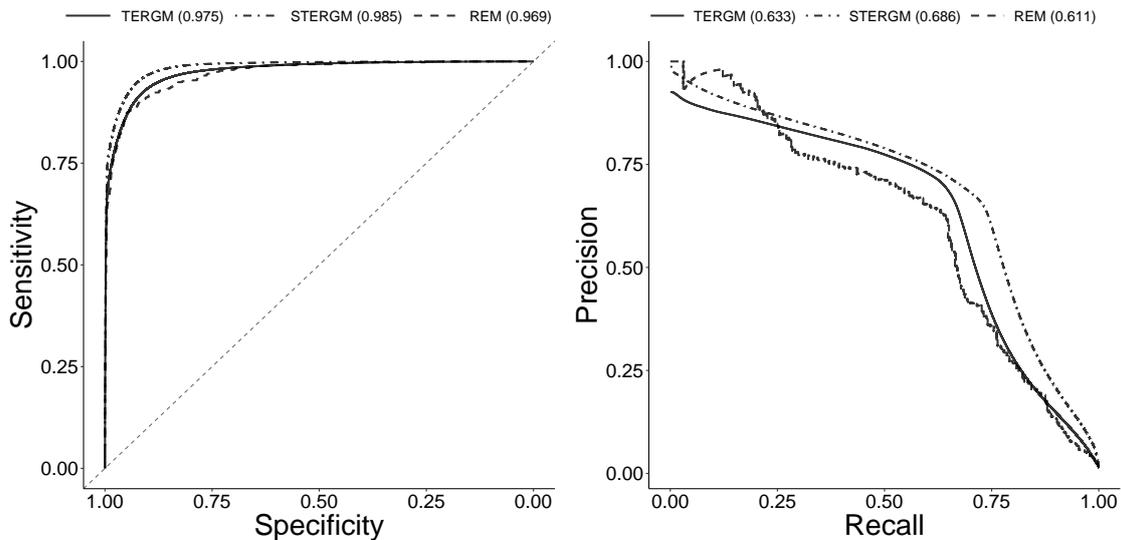


Figure 13: Arms trade network: ROC and PR curves from the TERGM (dotted line), STERGM (dotdashed line), and REM (solid line). The AUC values of the respective curves are indicated in brackets.

B.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 14. 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.

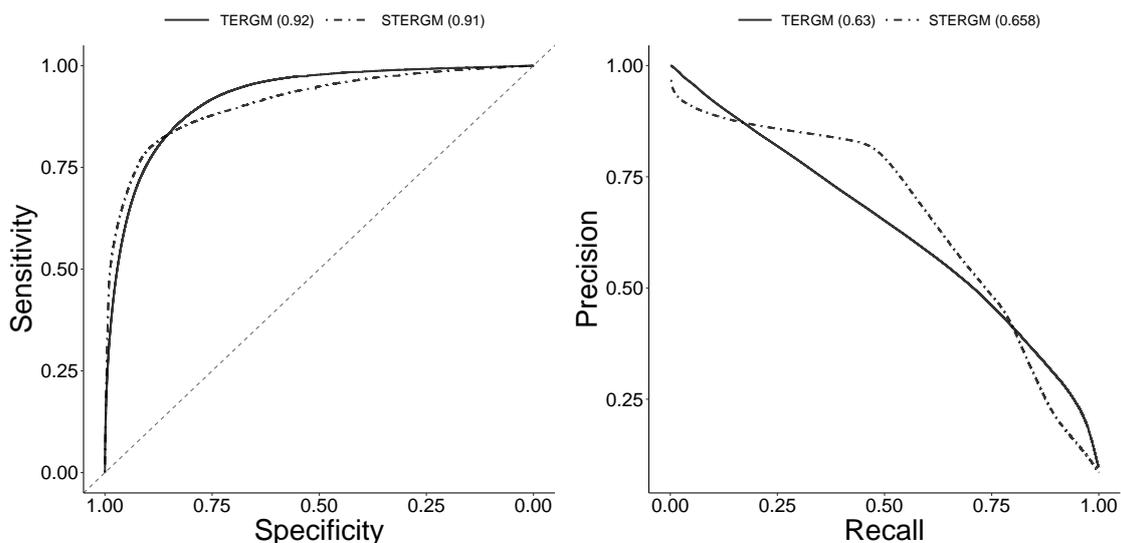


Figure 14: Email exchange network: ROC and PR curves from the TERGM (dotted line), STERGM (dotdashed line), and REM (solid line). The AUC values of the respective curves are indicated in brackets.

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. (2011). 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 15 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.

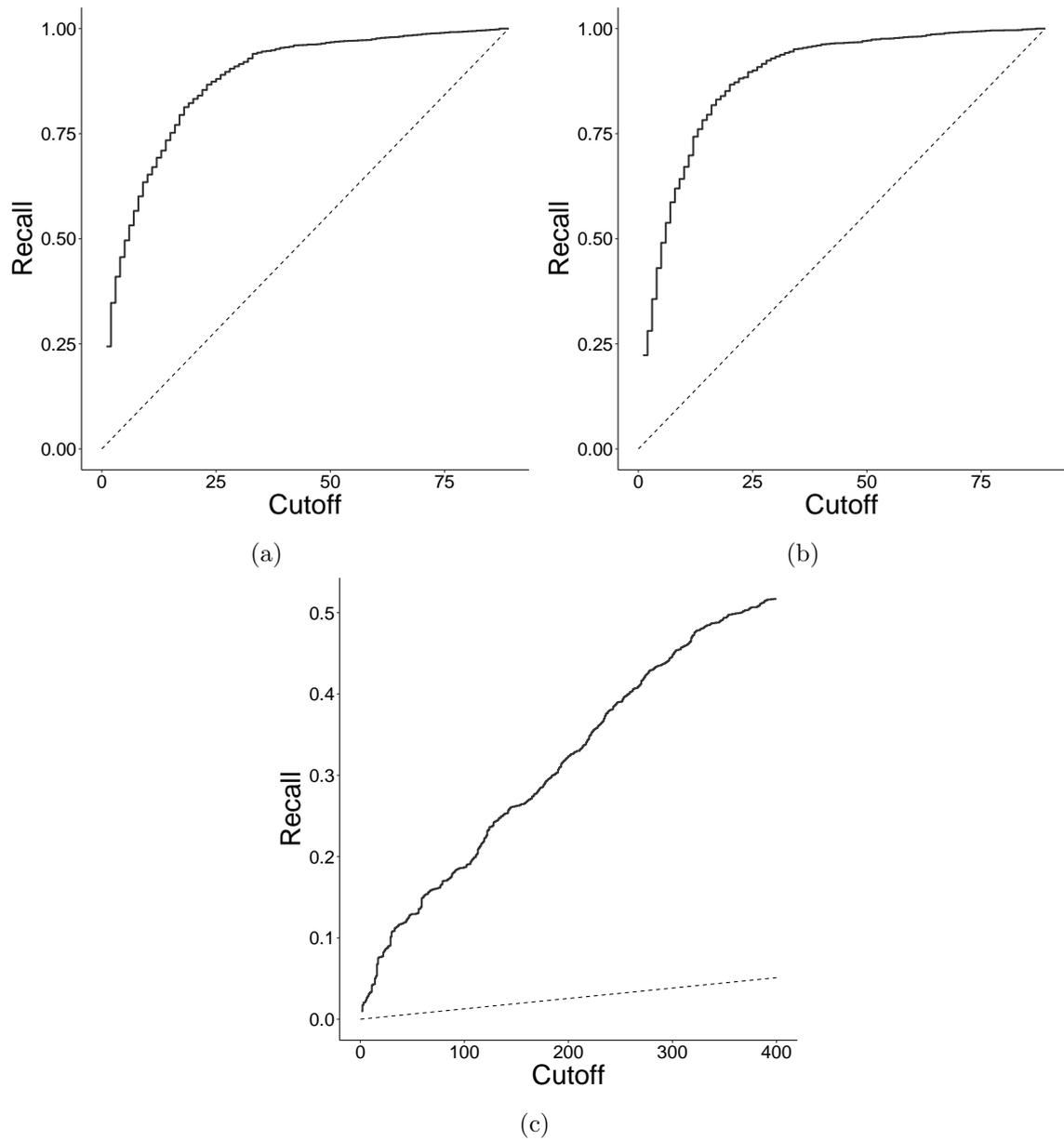


Figure 15: Email exchange network: Recall Curves of REM regarding the next sender (a), receiver (b), and event (c). The dotted line indicates the measure under random guessing of the next sender, receiver, or event.

B.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. 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 5 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.

	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 5: 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 $\log(2)$.

Chapter 4

Exploring dependence structures in the international arms trade network: A network autocorrelation approach

Contributing Article:

Michael Lebacher, Paul W. Thurner and Göran Kauermann (2019): *Exploring dependence structures in the international arms trade network: A network autocorrelation approach* Statistical Modelling.

Online first <https://doi.org/10.1177/1471082X18817673>

Code at <http://www.statmod.org/smij/archive.html>

Copyright:

Statistical Modelling Society, SAGE Publications Ltd, 2019.

Author Contributions:

The general idea of applying models from spatial econometrics in order to analyse edges in weighted networks stems from Göran Kauermann. He also had the idea of focusing on model comparison. Paul Thurner has provided the political and economic theory and valuable domain knowledge needed to legitimize, specify and interpret the models and results. The contribution of Michael Lebacher is given by the implementation of the model in R, including data manipulation, fitting and simulation. Furthermore, Michael Lebacher wrote the major part of the manuscript related to statistical matters. Also, the idea of comparing and visualizing different dependence structures as well as investigating differences between the normal residuals and the spatially correlated residuals can be attributed to Michael Lebacher. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

Chapter 5

Censored regression for modelling international small arms trading and its “forensic” use for exploring unreported trades

Contributing Article:

Michael Lebacher, Paul W. Thurner and Göran Kauermann (2019): *Censored regression for modelling international small arms trading and its “forensic” use for exploring unreported trades*.

Under review in the Journal of the Royal Statistical Society: Series C (Applied Statistics).

arXiv preprint <https://arxiv.org/abs/1902.09292>

Code at https://github.com/lebachelor/censored_network_regression

Author Contributions:

The general idea of applying models from spatial econometrics in a censored regression framework to model valued edges stems from Göran Kauermann. Paul Thurner has provided the political and economic theory and valuable domain knowledge needed to legitimize, specify and interpret the models and results. The concrete contribution of Michael Lebacher is given by the implementation of the model in R, including data manipulation, fitting and simulation. Furthermore, Michael Lebacher wrote the major part of the manuscript related to statistical matters. The initial idea of investigating the latent structures stems from Michael Lebacher and was extended together with Göran Kauermann toward the forensic approach. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

Censored Regression for Modelling International Small Arms Trading and its "Forensic" Use for Exploring Unreported Trades

Michael Lebacher*, Paul W. Thurner[†] and Göran Kauermann^{‡§}

Abstract

In this paper we use a censored regression model to investigate data on the international trade of small arms and ammunition (SAA) provided by the Norwegian Initiative on Small Arms Transfers (NISAT). Taking a network based view on the transfers, we not only rely on exogenous covariates but also estimate endogenous network effects. We apply a spatial autocorrelation (SAR) model with multiple weight matrices. The likelihood is maximized employing the Monte Carlo Expectation Maximization (MCEM) algorithm. Our approach reveals strong and stable endogenous network effects. Furthermore, we find evidence for a substantial path dependence as well as a close connection between exports of civilian and military small arms. The model is then used in a "forensic" manner to analyse latent network structures and thereby to identify countries with higher or lower tendency to export or import than reflected in the data. The approach is also validated using a simulation study.

Keywords: Gravity Model; Latent Variable; Maximum Likelihood; Monte Carlo EM Algorithm; Network Analysis; Spatial Autocorrelation; Zero Inflated Data

*Department of Statistics, Ludwig-Maximilians-Universität München, michael.lebacher@stat.uni-muenchen.de

[†]Department of Political Science, Ludwig-Maximilians-Universität München

[‡]Department of Statistics, Ludwig-Maximilians-Universität München

[§]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.

1. Introduction

The Small Arms Survey Update 2018 indicates transfers of small arms in 2015 amounting to 5,7 billion (Holtom and Pavesi, 2018, p. 19) with a major share and highest increases in ammunitions (Holtom and Pavesi, 2018, p. 22). Given the often fatal consequences - civilian or military - of the availability of these arms for intrastate conflict and shootings as well as for interstate war, the absence of empirical evidence for supplier-recipient networks is surprising. A major reason behind this research gap are the notorious data deficiencies due to non-reporting and illicit trafficking (see Holtom and Pavesi, 2018, p. 29-46). Based on the only large-scale data base for small arms (Marsh and McDougal, 2016) we aim to analyse for the first time the small arms trading network. We integrate gravity models in a statistical network design to apply a forensic statistical analysis.

Starting with the seminal work of Tinbergen (1962), the gravity equation was quickly established as a valuable tool of empirical trade research. The success of the model stems from its intuitive interpretation as well as its surprisingly strong empirical validity, see e.g. Head and Mayer (2014). It is therefore not surprising that the concept was applied to all kinds of trade relations, including the international exchange of arms. An early example for these applications is the work of Bergstrand (1992). Although he doubted the suitability of the model for arms trade because of the strong political considerations in this area, the approach was taken up more recently. Akerman and Seim (2014) and Thurner et al. (2018) used the gravity model in order to explain whether Major Conventional Weapons (MCW) are exchanged. Martinez-Zarzoso and Johannsen (2017) rely on the framework of Helpman et al. (2008) to investigate the influence of economic and political variables on the so-called extensive and intensive margin of MCW trade. The interplay between oil imports and arms exports is determined using a gravity model in Bove et al. (2018). While the papers above focus on the exchange of MCW, in our paper we investigate transfers of small arms and ammunition (SAA) provided by the Norwegian Initiative on Small Arms Transfers (NISAT). This data is arguably even better suited for a gravity model since small arms are potentially less dependent on political decision making and many more trade occurrences are recorded.

We propose a network perspective on international SAA trade and conceptualize countries as nodes and transfers between them as directed, valued edges. Although gravity models are a standard tool for the analysis of dyadic data (Kolaczyk, 2009), endogenous network effects are rarely incorporated in these models. We do so by connecting the idea of gravity models with the spatial autoregressive (SAR) model adjusted to network data. Especially in sociology, SAR models are regularly used in a network context since the early eighties (Dow et al., 1982; Doreian et al., 1984; Doreian, 1989). They are called *network*

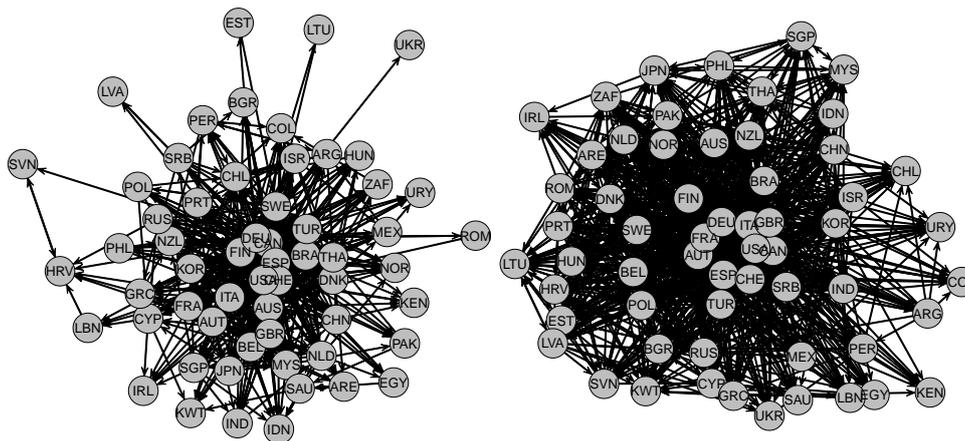
autocorrelation models in this strand of literature. More recently, network autocorrelation models became popular in political science applications, see for example Franzese and Hays (2007), Hays et al. (2010) and Metz and Ingold (2017). Here, it is assumed that actors with certain characteristics are embedded in a network and this embedding leads to contagion and/or spillover effects transmitted through the edges that relate the actors (Leenders, 2002). Hence, one presumes that the characteristics of actors are correlated because a specific social, political or economic mechanism is connecting them. Note that the design of these models is different as compared to the usual set-up of gravity models since the outcome is related to the nodes, and the edges only represent indicators for node dependence. In this paper, we are interested in the dependencies among the *transfers* (instead of the actors) and account for outdegree, indegree, reciprocity and exogenous covariates. A similar model in a non-network context is the spatial gravity model (LeSage and Pace, 2008), that accounts for spatial dependence of the exporter, the importer as well as for the spatial importer-exporter dependencies.

Contrary to the typical structure of trade data we observe a high degree of reported non-trade in SAA. In other words, the trade network has a large percentage of zero entries. To accommodate the zero inflation problem we employ a censored SAR model that can be fitted using the Monte Carlo Expectation Maximization (MCEM) algorithm (Dempster et al., 1977; Wei and Tanner, 1990). There are already several similar EM-based approaches that have been pursued. For instance Suesse and Zammit-Mangion (2017) use the EM algorithm in spatial econometric models, Schumacher et al. (2017) apply an EM-based application to a censored regression model with autoregressive errors, and Vaida and Liu (2009) utilize EM estimation in a censored linear mixed effects model. In Augugliaro et al. (2018) a similar estimation procedure is used in the context of fitting a graphical LASSO to genetic networks.

While the model application per se provides new insights into SAA trading, the ultimate objective in this paper is to make use of the model to explore the validity of reported zero trades. This reflects a "forensic" objective, i.e. we estimate, whether unreported trades are likely to have happened based on the fitted model. Despite this idea is in line with forensic statistics and forensic economics (Aitken and Taroni, 2004; Zitzewitz, 2012) our goal is apparently less ambitious. We do not aim to provide statistical evidence that some states are under-reporting but we do want to investigate *potential* under-reporting by utilizing the fitted network model.

This paper is organized as follows: after presenting the data in Section 2, we explain the model and show how to proceed with estimation and inference in Section 3. In Section 4 the results of the censored regression analysis are given and Section 5 provides the "forensic" analysis, accompanied by a simulation study. Section 6 concludes the paper.

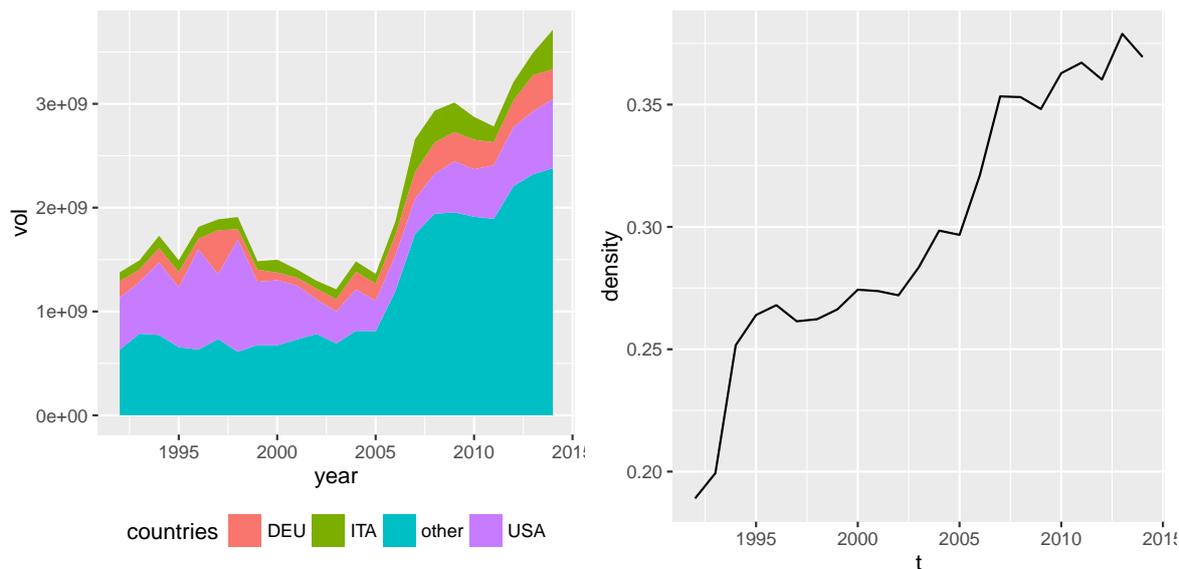
Figure 1: Binary SAA trade network for the 59 most relevant countries in 1992 (left) and 2014 (right). Countries are indicated by grey nodes and transfers by edges in black.



2. Data description

Since 2001, the Geneva-based Small Arms Survey specializes on documenting the international flows of the respective products. However, only the Norwegian Initiative on Small Arms Transfers (NISAT, see Marsh, 2017) provides truly relational data necessary for applying network analysis. The NISAT database contains relational information on the trade of small arms, light weapons and ammunition (see also Marsh and McDougal, 2016). This information is collected from different sources as described in Haug et al. (2002). Although NISAT represents the most reliable source of data regarding the exchange of small arms and light weapons, there is nevertheless an enormous amount of uncertainty inherent to arms trading data. This is especially true for light weapons where data quality and availability is partly very poor (Herron et al., 2011). Therefore, we restrict our analysis to small arms and the associated ammunition (SAA). See Table 2 in Annex 2 for the types of small arms and ammunition included in the dataset. Note, that the NISAT database also contains data on sporting guns, which we excluded from the dataset since we are particularly interested in the export of small arms with potential military value. Actually, we will rely on transferred sporting guns volumes later as a useful explanatory variable. In the remaining dataset, more than 86 000 SAA transfers are recorded for the years 1992-2014, providing the exporting country, the importing country as well as the transferred arms category. The value of the export is measured in constant 2012 USD. In order to make estimation feasi-

Figure 2: Aggregated exports (left) and density (right) in the SAA trade networks of the 59 most relevant countries. Countries with the highest export volume (Germany DEU, Italy ITA and United States USA) are highlighted.



ble, we restrict our analysis to a subnetwork and select those countries that account for the major share of the SAA trade activity. The resulting 59 countries (see Annex A.1, Table 3) account for 73% – 91% (depending on the year) of the total transfer volume and have participated in arms trade at least once in each year under study. Hence, we investigate the "core" of the international small arms trade network, balancing the trade-off between the number of countries included, the share of trade volume and the density of the sub-networks. In Figure 1, we show two binary networks for 1992 and 2014, with the countries represented as nodes and the arms transfers as directed edges among them.

In the left panel of Figure 2 we show the aggregated exports for the most important exporters United States (USA), Germany (DEU) and Italy (ITA) together with the exports of the 56 other countries (other). On the right hand side of Figure 2 we present the density, defined as the sum of existent edges divided by the number of potential edges. Although the network can be without doubt described as a dense one (as compared to the density of social networks), the density is smaller than 0.2 in the beginning and remains below 0.4 in the subsequent recent years.

3. Regression model

3.1. General model

Let $Y_t = (Y_{t,ij}) \in \mathbb{R}^{n \times n}$ represent a network of transfers at the discrete time points $t = 1, \dots, T$. At each time point Y_t consists of n nodes and $N = n(n - 1)$ directed, continuous real valued edges, with diagonal elements $Y_{t,ii}$ left undefined. We set $\tilde{Y}_t = \text{vec}(Y_t) \in \mathbb{R}^N$ as the row wise vectorization of Y_t , excluding the diagonal elements. In the following, we suppress the time index t for ease of the notation and assume (after some suitable transformation) that in each time point \tilde{Y} follows the autoregressive network model

$$\tilde{Y} = \sum_{k=1}^q \rho_k W_k \tilde{Y} + X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}_N(0, \sigma^2 I_N) \quad (1)$$

with β being a p -dimensional parameter vector for the design matrix X . The matrices W_k are row-normalized weight matrices representing linear endogenous network effects, with parameters ρ_k as their strength. Model (1) is usually known as spatial autoregressive (SAR) model and we refer to LeSage and Pace (2009) for a more detailed discussion and to Lacombe (2004) or LeSage and Pace (2008) for similar models with multiple weight matrices. Standard software implementations that allow for a likelihood based estimation of the model are mostly restricted to the special case with $q = 1$, for example in the R package `spdep` (Bivand et al., 2013; Bivand and Piras, 2015). The package `tnam` by Leifeld et al. (2017) allows for multiple weight matrices but is based on pseudo-likelihood estimation and therefore valid only if the weight matrices exclusively apply to exogenous covariates. Another possibility to estimate similar models is given by the package `ARCensReg` (Schumacher et al., 2017), initially designed to fit models with autoregressive errors. Because of the similar mathematical structure, the package could be used to fit models with spatially dependent errors known as Spatial Error Models (SEM). In the given case however, the network structure is assumed to influence the response directly which prevents us from using the package.

Model (1) can be rewritten as

$$\tilde{Y} = \underbrace{\left(I_N - \sum_{k=1}^q \rho_k W_k \right)}_{\equiv A(\boldsymbol{\rho})}^{-1} (X\beta + \epsilon) = \underbrace{(A(\boldsymbol{\rho}))^{-1}}_{\equiv B(\boldsymbol{\rho})} (X\beta + \epsilon) = B(\boldsymbol{\rho})(X\beta + \epsilon),$$

where the dependence on the q -dimensional parameter vector $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$ is made explicit for notational clarity. Similar as in Besag (1974) and given that all N edges in the

network are observed, their distribution is given by

$$P(\tilde{Y}|X, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} |A(\boldsymbol{\rho})| \exp \left\{ - \frac{(A(\boldsymbol{\rho})\tilde{Y} - X\beta)^T (A(\boldsymbol{\rho})\tilde{Y} - X\beta)}{2\sigma^2} \right\}. \quad (2)$$

The parameter space of this model is restricted such that $A(\boldsymbol{\rho})$ is non-singular, which is ensured if the eigenvalues of $A(\boldsymbol{\rho})$ are real valued and greater than zero.

3.2. Censored regression model

The above model is not directly applicable to our data since a large proportion of the SAA trade values is zero, expressing no (reported) SAA trade between countries i and j . We therefore adapt model (1) towards a utility model with censored observations. For each potential transfer from i to j there exists a utility of the transfer. This utility, however, only materializes in a transfer if it is higher than a certain threshold. Therefore, we assume that the probability model (2) applies to a network of partly observed latent variables, say $Z = (Z_{ij})$. The relation among Y_{ij} and Z_{ij} is given by $Y_{ij} = \max(c, Z_{ij})$ for $i, j = 1, \dots, n$ and c some threshold. Accordingly, we now set $\tilde{Y} = \text{vec}(Z)$ and label the observed utility $\tilde{Y}_o \in \mathbb{R}^{N_o}$ and the N_m unobserved ones as \tilde{Y}_m . A reordering according to the observational pattern of Y gives

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}_o \\ \tilde{Y}_m \end{pmatrix} \sim \mathcal{N}_N \left(\begin{pmatrix} \mu_o \\ \mu_m \end{pmatrix}, \begin{pmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{pmatrix} \right),$$

where $\tilde{Y}_m < c$ and $N = N_o + N_m$. Since the density of the network (see Figure 2) is roughly between 0.2 and 0.4 in all years, N_m is always substantially larger than N_o . The mean-covariance structure is given by

$$B(\boldsymbol{\rho})X\beta = \begin{pmatrix} \mu_o \\ \mu_m \end{pmatrix}, \quad B(\boldsymbol{\rho})(B(\boldsymbol{\rho}))^T \sigma^2 = \begin{pmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{pmatrix}.$$

In the following, we will denote all reordered matrices in the notation with double subscripts, i.e. A_{oo} refers to the submatrix of A where only interactions of observed variables \tilde{Y}_o enter.

3.3. Monte Carlo EM estimation

In order to estimate the unknown parameter vector $\theta = (\boldsymbol{\rho}, \beta, \sigma^2) \in \mathbb{R}^{q+p+1}$, we employ the EM algorithm (Dempster et al., 1977). The *complete log-likelihood* $\ell_{comp}(\theta)$ is simply

derived from (2). We are interested in maximizing the *observed log-likelihood* $\ell_{obs}(\theta) = \ell_{\tilde{Y}_o}(\theta) + \ell_{\tilde{Y}_m|\tilde{Y}_o}(\theta)$, where the first part is simply the multivariate normal density of the observed transfers. The second part equals

$$\ell_{\tilde{Y}_m|\tilde{Y}_o}(\theta) = \log \left(\int_{(-\infty, c]^{N_m}} \frac{1}{\sqrt{(2\pi)^{N_m} |\Sigma_{m|o}|}} \exp \left\{ -\frac{(U - \mu_{m|o})^T \Sigma_{m|o}^{-1} (U - \mu_{m|o})}{2} \right\} dU \right),$$

where $\mu_{m|o}$ and $\Sigma_{m|o}$ are the first and second conditional moments. Because N_m is greater than 2000 in each year, the observed log-likelihood is numerically hard to evaluate (and even more so to maximize) with state of the art software implementation. As a solution, we apply the EM algorithm and maximize $Q(\theta|\theta_0) := \mathbb{E}_{\theta_0}[\ell_{comp}(\theta)|\tilde{Y}_o, X, \mathcal{M}]$ iteratively. The observation space is given by

$$\mathcal{M} = \{\tilde{Y}_m : \tilde{Y}_{m,1} < c, \dots, \tilde{Y}_{m,N_m} < c\}. \quad (3)$$

3.4. E-Step

The E-Step essentially boils down to calculating the first two moments of a multivariate normally distributed variable \tilde{Y}^c

$$\tilde{Y}^c \sim \mathcal{N}_{N_m}(\mu_m + \Sigma_{mo}\Sigma_{oo}^{-1}(\tilde{Y}_o - \mu_o), \Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{om}) \quad (4)$$

with restriction \mathcal{M} from (3) applied to \tilde{Y}^c . Let those truncated moments be $\mu_{m|o}^c$ and $\Sigma_{m|o}^c$ and define

$$\tilde{Y}^* = \begin{pmatrix} \tilde{Y}_o \\ \mu_{m|o}^c \end{pmatrix} \quad (5)$$

as the vector that contains the observed values as well as the conditional expectation of the non-observed ones. Given the two moments, we can calculate the conditional expectation of the quadratic form (see Mathai and Provost, 1992):

$$\begin{aligned} S^*(\boldsymbol{\rho}) &= \mathbb{E}_{\theta_0}[\tilde{Y}^T (A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho}) \tilde{Y} | \tilde{Y}_o, X, \mathcal{M}] = \\ &= \text{tr} \left((A_{mm}(\boldsymbol{\rho}))^T A_{mm}(\boldsymbol{\rho}) \Sigma_{m|o}^c \right) + (\tilde{Y}^*)^T (A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho}) \tilde{Y}^*. \end{aligned} \quad (6)$$

Then, the function to maximize in the M-step is given by

$$Q(\theta|\theta_0) = -\frac{N}{2} \log(2\pi\sigma^2) + \log(|A(\boldsymbol{\rho})|) - \frac{(S^*(\boldsymbol{\rho}) - 2\beta^T X^T A(\boldsymbol{\rho}) \tilde{Y}^* + \beta^T X^T X \beta)}{2\sigma^2}. \quad (7)$$

In order to find the first and second moment of a truncated multivariate normally distributed variable, Vaida and Liu (2009) use the results of Tallis (1961) on the moment

generating function to provide closed form expressions of the E-Step. This, however, is not practicable in our setting as (a) software implementations of a multivariate normal distribution function are overstrained by the high dimension of our problem (the standard package in R, `mvtnorm` by Genz et al. (2016) is not able to process dimensions higher than 1 000) and (b) as noted by Schumacher et al. (2017), even if the distribution function could be evaluated, the closed form solution is computationally very expensive which leads to infeasible convergence times in applications with a high number of non observed values. The same is true for the direct calculation using the moment generating function implemented in R by Wilhelm et al. (2012).

A practicable alternative consists in using the Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990) where intractable expectations are replaced by sample based approximations. In our specific case we use the R package `TruncatedNormal` by Botev (2017) in order to draw from the truncated multivariate normal distribution. An alternative would be to enrich the E-Step with a stochastic approximation step (SAEM algorithm, see Schumacher et al., 2017 for a detailed description) which reduces the number of simulations needed and is very efficient if the M-step is faster than the E-Step. In our specific application, the computational bottleneck comes with the M-Step and simulations showed that the SAEM converges more slowly than the MCEM algorithm.

3.4.1. M-Step

It is numerically more efficient to reduce the log-likelihood to a profile log-likelihood by first maximizing with respect to β and σ^2 and then with respect to $\boldsymbol{\rho}$. Using the derivatives of (7) with respect to β and σ^2 and defining $\hat{\beta}(\boldsymbol{\rho})$ and $\hat{\sigma}^2(\boldsymbol{\rho})$ as the solutions of the score equations as functions of $\boldsymbol{\rho}$ it follows that

$$\begin{aligned}\hat{\beta}(\boldsymbol{\rho}) &= (X^T X)^{-1} X^T A(\boldsymbol{\rho}) \tilde{Y}^* \\ \hat{\sigma}^2(\boldsymbol{\rho}) &= \frac{S^*(\boldsymbol{\rho}) - \tilde{Y}^{*T} (A(\boldsymbol{\rho}))^T H A(\boldsymbol{\rho}) \tilde{Y}^*}{N},\end{aligned}\tag{8}$$

where $H = X(X^T X)^{-1} X^T$ is the hat matrix. With κ being a constant we can write the profiled function $\tilde{Q}(\cdot)$ as

$$\tilde{Q}(\boldsymbol{\rho}|\theta_0) = \kappa + \log(|A(\boldsymbol{\rho})|) - \frac{N}{2} \log \left(S^*(\boldsymbol{\rho}) - \tilde{Y}^{*T} (A(\boldsymbol{\rho}))^T H A(\boldsymbol{\rho}) \tilde{Y}^* \right).\tag{9}$$

The expressions $(A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho})$ and $(A(\boldsymbol{\rho}))^T H A(\boldsymbol{\rho})$ have derivatives

$$\begin{aligned}\frac{\partial(A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho})}{\partial \rho_k} &= -W_k - W_k^T + 2\rho_k W_k^T W_k + \sum_{l \neq k} \rho_l (W_k^T W_l + W_l^T W_k) =: R_k(\boldsymbol{\rho}) \\ \frac{\partial(A(\boldsymbol{\rho}))^T H A(\boldsymbol{\rho})}{\partial \rho_k} &= -H W_k - W_k^T H + 2\rho_k W_k^T H W_k + \sum_{l \neq k} \rho_l (W_k^T H W_l + W_l^T H W_k) =: H_k(\boldsymbol{\rho}).\end{aligned}$$

Now define

$$R_k^*(\boldsymbol{\rho}) = \text{tr}(R_{k,mm}(\boldsymbol{\rho}) \Sigma_{m|o}^c) + (\tilde{Y}^*)^T R_k(\boldsymbol{\rho}) \tilde{Y}^*$$

which gives

$$\frac{\partial \tilde{Q}(\boldsymbol{\rho} | \theta_0)}{\partial \rho_k} = -\text{tr}(B(\boldsymbol{\rho}) W_k) - \frac{N}{2} \frac{R_k^*(\boldsymbol{\rho}) - \tilde{Y}^{*T} H_k(\boldsymbol{\rho}) \tilde{Y}^*}{S^*(\boldsymbol{\rho}) - \tilde{Y}^{*T} (A(\boldsymbol{\rho}))^T H A(\boldsymbol{\rho}) \tilde{Y}^*}. \quad (10)$$

Iteration between the E- and the M-step provide the final estimate $\hat{\theta}$. The variance of $\hat{\theta}$ can be calculated using Louis (1982) formula with more details on the practical implementation provided in the Supplementary Material.

4. Application to the data

4.1. Covariates

Considering model (1) we need to specify the two major components of the model, namely (a) the covariates included in matrix X and (b) the network related correlation structure. *Node Specific Variables:* Following standard applications (Ward et al., 2013; Head and Mayer, 2014; Egger and Staub, 2016; Thurner et al., 2018), we control for the logarithmic real GDP in constant 2010 USD as a measure for the market size of the exporting and importing country. The data are provided by the World Bank (2017). For the two years 1993-1994 no reliable GDP data are available for Serbia, Croatia, Estonia, Latvia, Lithuania and Slovenia, we therefore assume that the GDP remained constant in the first three years for this countries. In order to control for the potential influence of intrastate conflicts we insert a binary variable that is one if there is an intrastate conflict in the receiving country in the respective year and zero otherwise. The corresponding data is available from the webpage of the Uppsala Conflict Data Program (UCDP, 2019).

Edge Specific Distance Measures: Because of the strong empirical evidence that geographic distance is a relevant factor in trade (Disdier and Head, 2008), we control for the

logarithmic distance between capital cities in kilometres (Gleditsch, 2013). In recent applications of the gravity model to arms trade (Akerman and Seim, 2014; Martinez-Zarzoso and Johannsen, 2017; Bove et al., 2018; Thurner et al., 2018) it is argued that political distance measures in terms of regime dissimilarity must also be inserted in the gravity equation. We use the absolute difference of the polity IV index (Marshall, 2017) between two countries, ranging from 20 (highest ideological distance) to zero (no ideological distance). Additionally, we include a dummy variable for formal alliances between the exporting and importing country, being one if the two countries have a formal alliance. The data is available from Correlates of War Project (2017) until 2012 and we assume that the alliances stay constant for the years 2013 and 2014.

Edge Specific Trade Measures: We control for lagged logarithmic SAA transfers by smoothing the past observed trade volume using a five-year moving average. In the years with less than five lagged periods available, the moving average is shortened accordingly. We call this path dependency, leading to inertias that arises because of diminishing transaction costs, trust relations, security aspects and potentially interoperability, and is a very important determinant in the MCW trade network (Thurner et al., 2018).

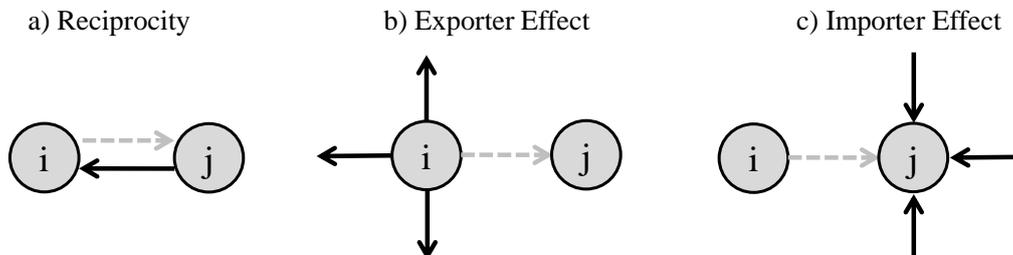
Additionally, we enrich the model with a five year moving average of logarithmic civilian weapon transfers. The intuition behind that is that exports of SAA for military usage and civilian usage might be correlated. This is plausible because countries that export massive amounts of civilian arms also have the capabilities to produce military arms. The data is also provided by NISAT (Marsh and McDougal, 2016). Furthermore it seems plausible that there is a connection between the volume of small arms traded and the volume of MCW. MCW transfers are recorded by the Stockholm International Peace Research Institute (SIPRI) and measured in so called trend indicator values (TIV). This measure represents the military value and the production costs of the transferred products. For detailed explanation of the data and the TIV see SIPRI (2017b,a) and Holtom et al. (2012). We use a dummy variable that is one if there was an MCW transfer from country i to j in the actual year or in the four preceding years, zero otherwise. Additionally, we use the logarithmic sum of the exported TIV volumes in the actual year and the four preceding ones.

4.2. Network structure

Next we specify the network specific effects represented by matrices W_k in model (1). We include three effects which are explained subsequently and visualized in Figure 3.

Reciprocity: The reciprocity effect measures whether the export volume from country i to country j increases in the export volume from j to i . In the given context it is a plausible assumption that countries tend to specialize in certain types of small arms and/or

Figure 3: Schematic representation of linear network effects. The focal edge is in dashed grey.



ammunition and therefore complement each other with their products. Mutual trade is likely to be encouraged by political partnerships and indicates strategic elements, induced by bilateral agreements. The measure is also investigated in the context of commercial trade (e.g. Garlaschelli and Loffredo, 2005; Barigozzi et al., 2010; Ward et al., 2013). In the arms trade literature, reciprocity is specified by Thurner et al. (2018), with the finding that this is rather unusual in the context of MCW.

Exporter and Importer Effect: The exporter and the importer effect have their analogies in binary networks and can be interpreted as the valued versions of the outdegree and the indegree. The coefficient of the exporter effect measures whether the transfers going out from a certain exporter i are correlated. A positive effect indicates the presence of "super-exporters". Contrary, the importer effect measures whether the imports of a certain importer j are related, with a positive effect indicating "super-importers". The degree structure is a crucial feature of the SAA network because a rather small number of countries accounts for the major share of the trade volume, while a small share of (potentially identical) importing countries accounts for a great amount of the import volume.

Before fitting the model we apply the natural logarithm to the data. This is necessary, because in its raw form the data is strongly skewed with a long tail. In the Supplementary Material the distribution of the log-transformed response is investigated

Hence, if the original trade matrices are given by $Y_t = (Y_{t,ij})$, the elements of \tilde{Y}_t are given by $\tilde{Y}_{t,ij} = \log(Y_{t,ij})$ if $Y_{t,ij} > 0$ and are not defined if $Y_{t,ij} = 0$. Furthermore, we define $d_t = \min(\{Y_{t,ij} > 0\})$ as the lowest strictly positive value in the network at year t and set $c_t = \log(d_t)$. That is, the threshold c_t is defined such that at a given time point t all transfers below the smallest observed log-transformed transfer in that sample are censored. Utility below the threshold c_t implies that no transfer was carried out or was not recorded. Furthermore, we allow for time-varying coefficients by estimating each time-period separately. This relaxes the unrealistic assumption of time-constant effects for more than 20 years and reduces the computational effort. Given these specifications, the

final model is now given by

$$\tilde{Y}_{t,ij} = X_{t,ij}^T \beta_t + \underbrace{\rho_{t,1} \tilde{Y}_{t,ji}}_{\text{Reciprocity}} + \underbrace{\rho_{t,2} \frac{1}{n-2} \sum_{u \neq j} \tilde{Y}_{t,iu}}_{\text{Exporter Effect}} + \underbrace{\rho_{t,3} \frac{1}{n-2} \sum_{u \neq i} \tilde{Y}_{t,u,j}}_{\text{Importer Effect}} + \epsilon_{t,ij},$$

$$\epsilon_{t,ij} \sim \mathcal{N}(0, \sigma^2) \text{ for } i, j = 1, \dots, n, i \neq j, t = 1993, \dots, 2014, n = 59 \text{ and } N = 3422.$$

4.3. Results: Coefficients

In Figure 4 the time series of coefficients are plotted against time for the years 1993-2014. The shaded areas around the coefficients give two standard error bounds and the colouring of the point estimates reflects the respective significance level, the zero-line is depicted by solid black.

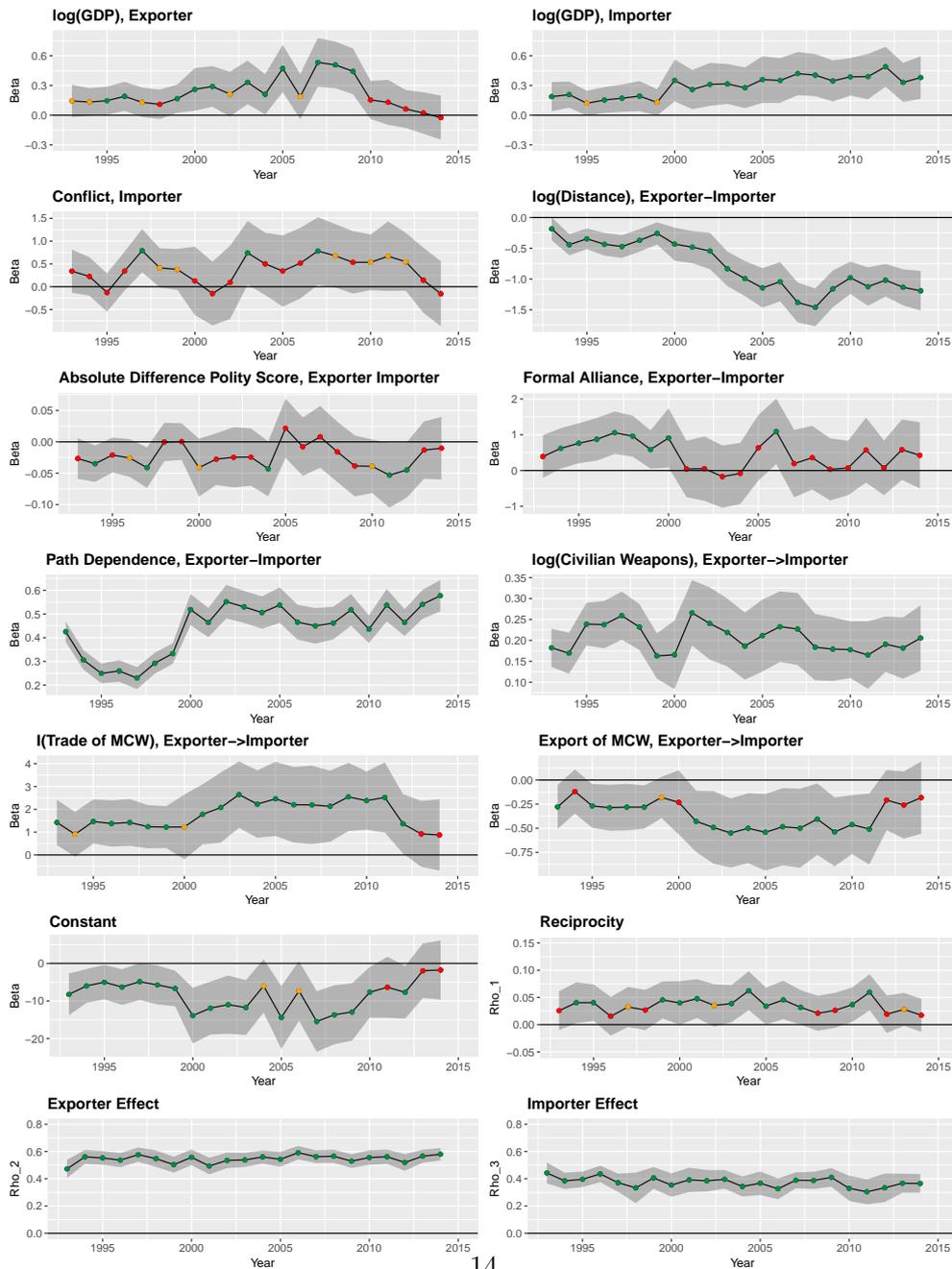
Covariates: The exogenous covariates in the first row and in the second row on the right represent the standard gravity variables logarithmic GDP of the exporter and the importer as well as the geographic distance between them (second row, right panel). Overall, the expected results of the gravity equation hold, except for the logarithmic GDP of the exporter that tends to be insignificant and is close to zero in the most recent years. This is an interesting result, because it highlights the fact, that market size is not a prerequisite for producing and exporting internationally competitive SAA. This finding is in stark contrast to the insights on MCW by Thurner et al. (2018).

The strong negative effect of the geographic distance is significant in all years. Again, this is different as compared to MCW transfers where geopolitical strategy disregards distance.

Regarding the political security measures we find that the presence of a conflict in the importing country (second row, left panel) has a mostly positive but seldom significant effect while the coefficient on the dissimilarity of political regimes (third row, left panel) is mostly negative but also often insignificant. The coefficient on the dummy variable for formal alliances (third row, right panel) is positive in the beginning but almost permanently insignificant from 2001 on.

The large and consistently significant coefficients of the lagged moving average (fourth row, left panel) illustrates an important feature of the network, namely path inertia. Intensive transfer relationships in the past, strongly increase the export volume in the present. Similarly we find a strong connection between exporting civilian and military arms (fourth row, right panel). Looking at the relation between SAA and MCW trade we find that having traded MCW (fifth row, left panel) in the actual year or the in the four preceding ones has a strong positive effect - at least until the last two years. However, at the same time

Figure 4: Time-series of annually estimated regression coefficients. Shaded areas give ± 2 standard errors. Colouring according to p-values, green: $p < 0.05$, yellow: $p < 0.1$ and red: $p > 0.1$.



the effect of the logarithmic sum of the TIV values (fifth row, right panel) has a negative and mostly significant effect. I.e. rather small transfers of MCW tend to coincide with SAA exports while dyads with high amounts of MCW exchange tend to transfer small arms to a relatively lower degree.

Network Structure: On the right panel in the sixth row of Figure 4 the coefficients for reciprocity are shown. The coefficients remain almost constant and positive with values between 0.02 and 0.06. As the coefficients often changes from significant to insignificant we infer that there is at least a tendency that mutuality increases the volume of arms exchanged.

The strongest endogenous effect is the exporter effect (bottom row, left panel) with coefficients that are consistently positive and significant. This indicates that the transfers stemming from the same exporter are indeed highly correlated and reflects the existence of "super-sellers" like the United States, Germany, Brazil or Italy. On the other hand, we also find a stable, positive and significant importer effect (bottom row, right panel). The fact that the two coefficients on the exporter and the importer effect are much higher than the reciprocity effect provides structural information about heterogeneity in the network. Being a strong exporter or sending to a strong importer increases the export volume more than simply having imported high amounts from the respective partner.

5. "Forensic" statistical analysis

5.1. Under- and over-reporting

Our model rests on the assumption that the SAA network is determined by a latent utility network Z_t . Based on the joint distribution (2) we can in fact estimate the probability of $Z_{t,ij}$ being greater than the threshold c_t , given the covariates, the endogenous effects and the rest of the network. In order to do so, let $Z_{t,-ij}$ represent the $(N - 1)$ -dimensional vector that contains the realized and the expected values of the latent variables, except the entry that corresponds to the transfer from i to j . Because we are interested whether some latent transfers could have realized according to the model, we form the expectations *without* the restriction that the latent transfers must be smaller than c_t . Based on this, we define the conditional probability of a specific latent transfer being greater than the threshold by

$$\pi_{t,ij} = \mathbb{P}(Z_{t,ij} > c_t | X_{t,ij}, Z_{t,-ij}; \hat{\theta}_t).$$

By construction (see the Supplementary Material for the derivation), $\pi_{t,ij}$ is high for transfers that are observed in the dataset ($Y_{t,ij} > 0$) and small for transfers that are not observed ($Y_{t,ij} = 0$). However, we may calculate a high value of $\pi_{t,ij}$, i.e. a high probability for a

realized transfer of arms, despite the data actually indicates $Y_{t,ij} = 0$. We propose to consider this as potential *under-reporting*. Such a zero-record can happen due to random fluctuation, factors beyond the model as for example historical relationships, or because de-facto existent transfers have not been reported. Vice versa, we may obtain a low value of $\pi_{t,ij}$ although $Y_{t,ij} > 0$. We label this as *over-reporting*. This label is not intended to suggest that potentially over-reported transfers in fact never happened, but highlights transfers where our model attaches a lower level of latent utility than manifested in the data. Naturally, our main "forensic" interest is in uncovering potential under-reporting.

Apparently, this requires the fixation of a threshold value for the probabilities. Based on Receiver-Operating Characteristic (ROC) curves, an optimal threshold value J_t can be found using Youden's J statistic (Youden, 1950). This value is optimal in the sense that it allows for a separation such that both, sensitivity and specificity are maximized. This defines the binary network

$$\Pi_t = (I(\pi_{t,ij} > J_t)).$$

This network is now set into relation with the observed binary SAA trade

$$\Gamma_t = (I(Y_{t,ij} > 0)).$$

Comparing Π_t and Γ_t , we can define the "forensic" network

$$\Omega_t = (\omega_{t,ij}) = \Pi_t - \Gamma_t$$

which in turn creates two new binary networks

$$\begin{aligned}\Omega_t^+ &= (I(\omega_{t,ij} = 1)) \\ \Omega_t^- &= (I(\omega_{t,ij} = -1)).\end{aligned}$$

For $\omega_{t,ij} = 1$, the model predicted a transfer that is not present in the dataset, and for $\omega_{t,ij} = -1$, the model did not predict an actual transfer. Following our convention from above we label Ω_t^+ as the *under-reporting network* and to Ω_t^- as the *over-reporting network* of unpredicted but realized transfers.

5.2. Simulation study of "forensic" power

Before we apply our model in a "forensic" matter to identify transfers with potential under-reporting we demonstrate the behavior of the model in a simulation study to explore its detection properties. We use two different settings in order to investigating how well the

Table 1: Schematic representation of the evaluation scheme used in the simulation study. True conditions in the columns and Estimated in the rows. **UR** denotes under-reporting and **UR** denotes censored observations that are not under-reported. Further abbreviations: True positive (TP), false positive (FP), false negative (FN) and true negative (TN).

TRUE					TRUE				
		UR	UR	Σ			UR	UR	Σ
Estimated	UR	0	FP	FP	Estimated	UR	TP	FP	TP+FP
	UR	0	TN	TN		UR	FN	TN	FN+TN
	Σ	0	$0.75N$	$0.75N$		Σ	$0.1N$	$0.65N$	$0.75N$

(a) Classifier evaluation DGP1.

(b) Classifier evaluation DGP2.

proposed approach identifies under-reporting. The first setting builds on the following Data Generating Process (DGP1)

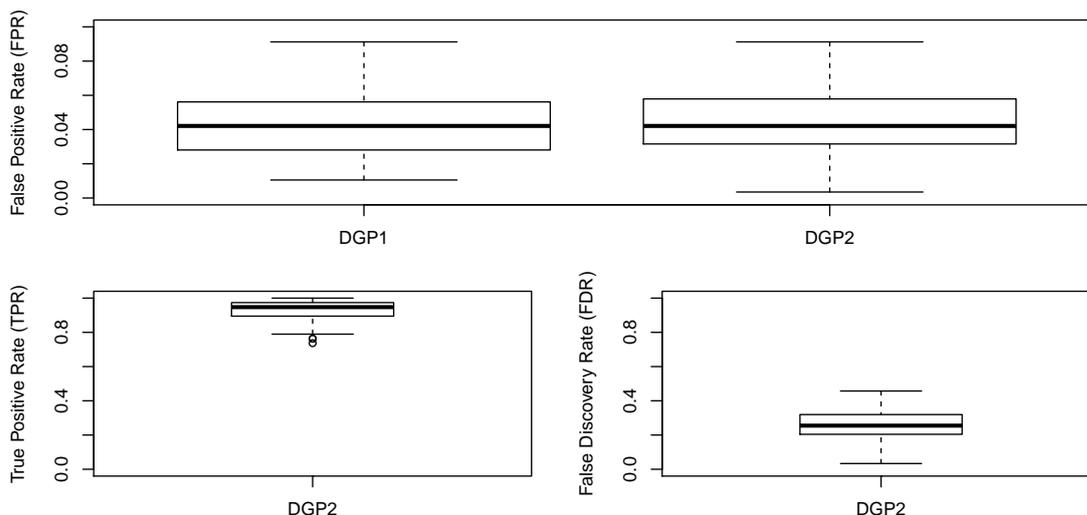
$$\begin{aligned}
 \boldsymbol{\rho} &= (0.1, 0.2, 0.3)^T, \beta = (1, 2, 3, 4, 5)^T, p = 5, n = 20, N = 380 \\
 X &\sim \mathcal{N}_p(\mathbf{1}, I_p) \\
 Z &\sim \mathcal{N}_N(B(\boldsymbol{\rho})X\beta, B(\boldsymbol{\rho})B(\boldsymbol{\rho})^T) \\
 \tilde{Z}_{ij} &= I(Z_{ij} > q_{0.75}(Z))Z_{ij}, \text{ for } i \neq j = 1, \dots, n.
 \end{aligned} \tag{11}$$

Here, $q_{0.75}(Z)$ denotes the 75% quantile and we are censoring the network towards an observed density of 0.25. Note, that DGP1 is not subject to under-reporting and all censored responses are in fact below the censoring threshold. The results of running DGP1 100 times and applying the estimation procedure are summarized in the Supplementary Material, indicating that the expected values approximate the latent variables very well and that we are able to find unbiased estimates despite the enormous amount of censoring.

In order to validate the forensic power of the model, we run a second experiment (DGP2), being a modified version of DGP1. To be precise, we are censoring again 75% of the observations but only 65% correspond to the lowest ones, while the remaining 10% are randomly selected among the observations that are in fact *higher* than the threshold $q_{0.65}(Y)$. This share of observations represents the under-reporting. Again we run DGP2 100 times.

In order to make the following evaluation transparent, we represent the evaluation scheme (e.g. Fawcett, 2006) for both DGPs in Table 1. On the left hand side, we regard the simulation without under-reporting (DGP1). In this setting we can investigate the false positive rate (FPR), being the sum of the false positives (FP) relative to the number of all observations that are in fact not under-reported. A low value for this measures

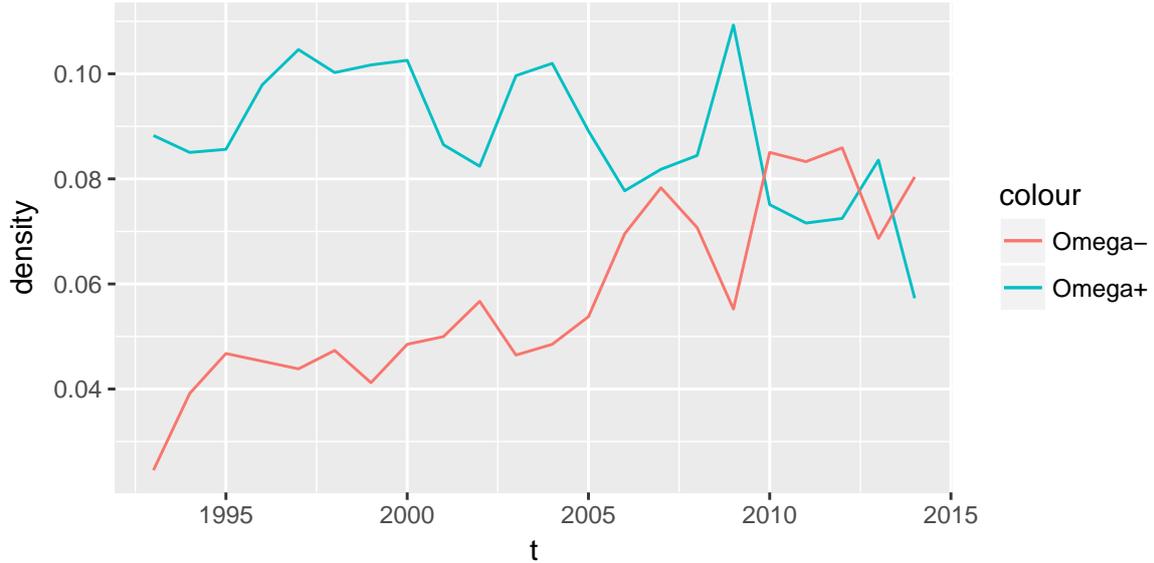
Figure 5: Results of DGP1 and DGP2. The top panel shows boxplots for the false positive rate (FPR) in DGP1 (left) and DGP2 (right). On the bottom boxplots for the true positive rate (TPR) and the false discovery rate (FDR) are provided for DGP2.



means in DGP1 that in a setting without under-reporting a low share is classified as under-reporting. In DGP2, the measure tells us whether including true under-reporting in the simulation leads to an increase of misclassified under-reporting. The corresponding results are visualized on the top panel of Figure 5. The FPR shows a higher variability in DGP2 (right panel) and is slightly higher as compared to DGP1 (left panel). However, the results provide evidence for an overall low FPR in both setting.

Furthermore, DGP2 allows to evaluate the share of under-reported observations that is identified. This is assessed based on the true positive rate (TPR) and shown on the bottom left panel of Figure 5. In fifty percent of all simulation runs we are able to identify at least 95% of the falsely censored observations and even in the simulation runs with the worst performance, the TPR are does not fall below 74%. Additionally, we investigate the False discovery rate (FDR) that relates the observations that are wrongly classified to be under-reporting to the sum of all observations that are classified for under-reporting. A low value for this measure provides evidence, whether the model is able to keep the number of potential over-reporting that are in fact not under-reporting low. The corresponding results are shown in the south-east panel of Figure 5. We find a median share of less than 26% to be classified incorrectly.

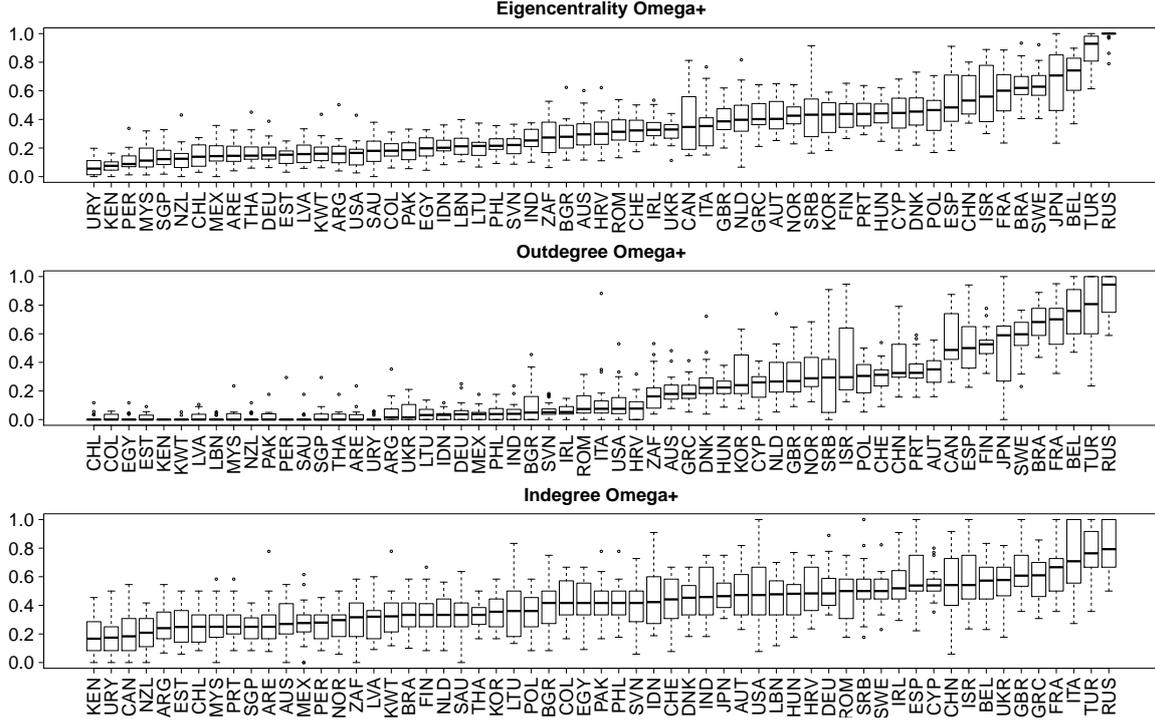
Figure 6: Densities of the under-reporting network Ω_t^+ and the over-reporting network Ω_t^- over time.



5.3. "Forensic" analysis of arms trade data

We now turn back to the data and provide the development of the densities for the latent networks in Figure 6. In the real data, over- and under-reporting is certainly not random but potentially clustered among countries. We therefore evaluate node (i.e. country) specific network topologies of Ω_t^+ and Ω_t^- for each year and summarize the information in box-plots for each country, ordered according to the median of the respective feature. This is shown in Figure 7 for potential under-reporting and in Figure 8 for over-reporting. In the first row, we represent the Eigenvector centrality scores. This measure is undirected and constructed such that the centrality of each country is proportional to the sum of the centralities of its trading partners. Hence, countries with high scores have many potentially under-reported (over-reported) import- and export-relations with many other countries that themselves have many under-reported (over-reported) import- and export-relations, see e.g. Csardi and Nepusz (2006). In the middle row, we present the outdegree, that is the number of potentially under-reported (over-reported) exports for a country. The bottom row in Figures 7 and 8 gives the indegree, that is the number of potentially under-reported (over-reported) imports. All measures are scaled to take values between 0 and 1. Countries at the right hand side in the plots of Figure 7 are potentially under-reporting and in Figure

Figure 7: Ordered box-plot representation of topological network features of the under-reporting networks Ω_t^+ for $t = 1993, \dots, 2014$: Eigencentrality (top), outdegree (middle) and indegree (bottom).



8, the right hand side of the plots mirrors high over-reporting.

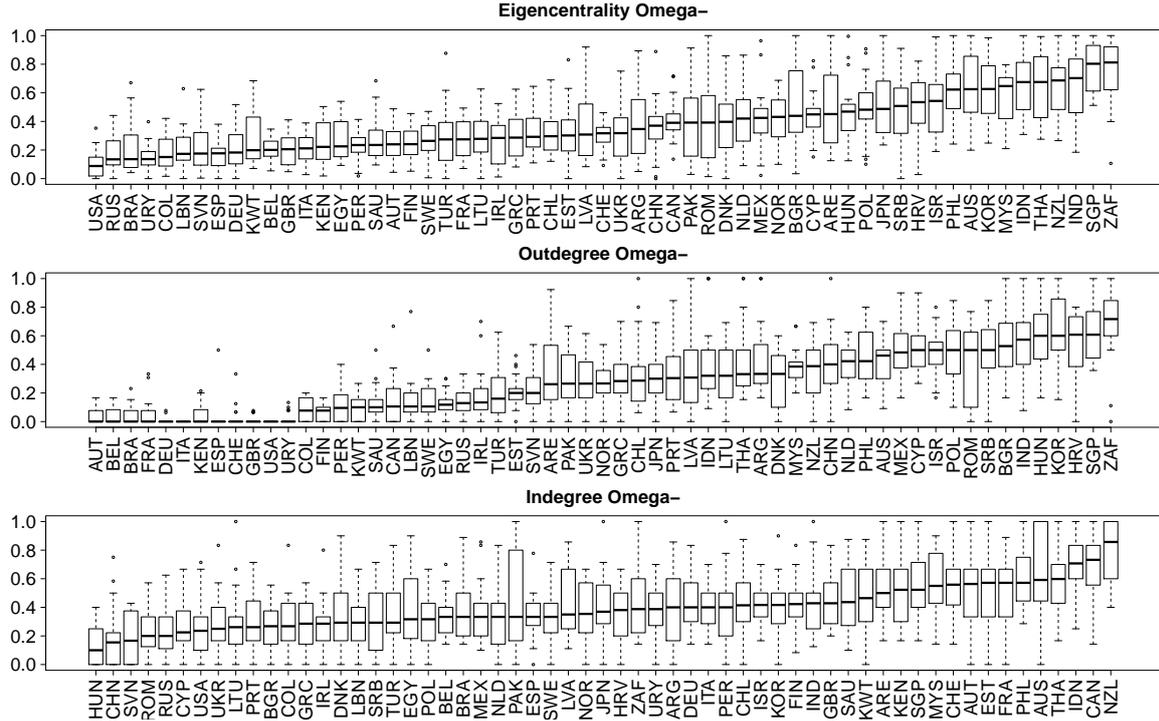
To detect persistent patterns in the networks on a dyadic level, we check whether potential under-reporting or over-reporting occurs frequently, i.e. counting instances of $\omega_{t,ij}^+ = 1$ and $\omega_{t,ij}^- = 1$ for $t \in \mathcal{T} = \{1993, \dots, 2014\}$. Denote the aggregated "forensic" networks as

$$\Omega_{\mathcal{T}}^+ = \sum_{t \in \mathcal{T}} \Omega_t^+,$$

$$\Omega_{\mathcal{T}}^- = \sum_{t \in \mathcal{T}} \Omega_t^-.$$

We look at the distribution of elements of $\Omega_{\mathcal{T}}^+$ and $\Omega_{\mathcal{T}}^-$, which is plotted in Figure 9. On the horizontal axis we show the possible values of the matrix entries, that is the number

Figure 8: Ordered box-plot representation of topological network features of the over-reporting networks Ω_t^- for $t = 1993, \dots, 2014$: Eigenvector centrality (top), outdegree (middle) and indegree (bottom).

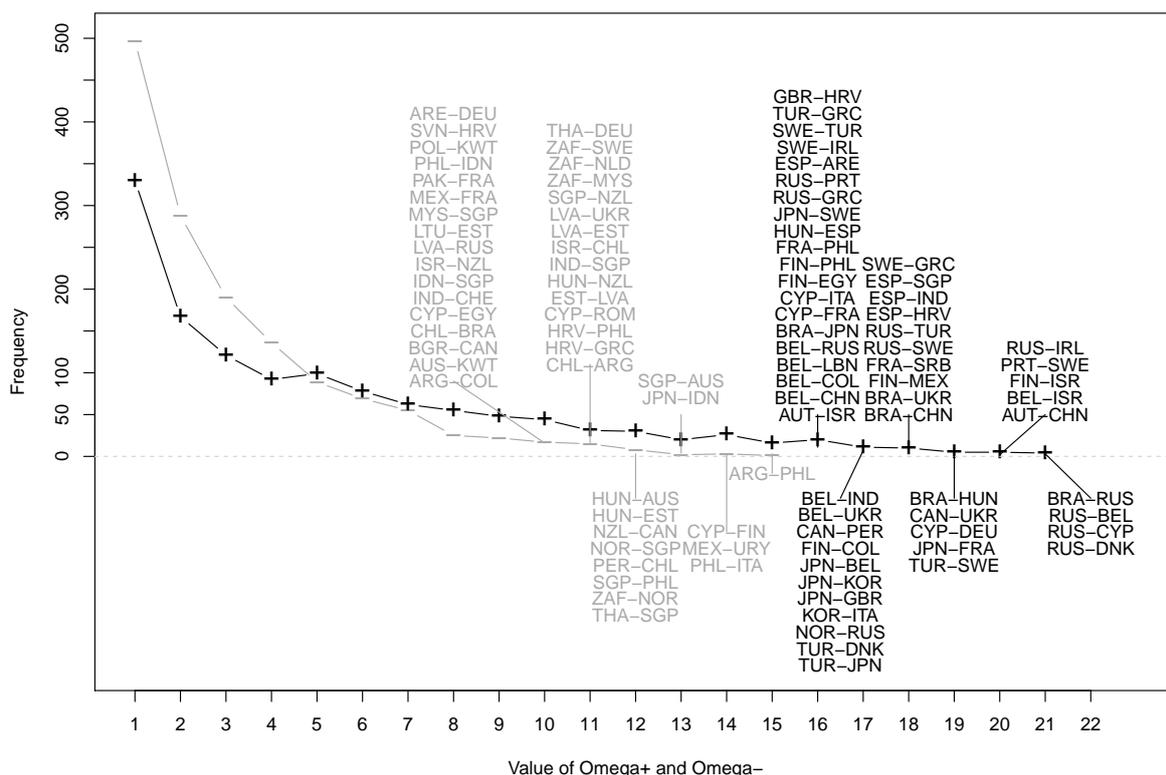


years where transfers in the forensic networks occur. This ranges from 1 (potential under-reporting or over-reporting in one year) to 22 (potential under-reporting or over-reporting in all years). The maximum entry of $\Omega_{\mathcal{T}}^+$ is thereby less than 22, namely 21, while the maximum value of $\Omega_{\mathcal{T}}^-$ is 15. On the vertical axis of Figure 9 we show the frequency of the entries of $\Omega_{\mathcal{T}}^+$ and $\Omega_{\mathcal{T}}^-$. Apparently for "forensic" purposes, large values of $\Omega_{\mathcal{T}}^+$ are of particular interest, since they report pairs of countries which are likely to under-reporting.

The line in solid black with the "+" symbols represents $\Omega_{\mathcal{T}}^+$ and the line in grey with the "-" symbols represents the under-reporting network. Additionally, we indicate for both networks the pairs of countries (i.e. sender and receiver) which are of particular interest for "forensic" purposes. This means for example for an element of $\Omega_{\mathcal{T}}^+$ that has value 21, that the respective transfer from i to j is one of the four transfers appeared that appeared 21 times in the under-reporting network.

Under-reporting networks Ω_t^- : Looking at the Eigenvector centrality scores of Figure 7

Figure 9: Frequency distribution of transfers in the aggregated under-reporting network ($\Omega_{\mathcal{T}}^+$, black "+"") and over-reporting ($\Omega_{\mathcal{T}}^-$, grey "-"") networks on the vertical axis. Number of years with under-reporting ($\omega_{\mathcal{T},ij}^+$) or over-reporting ($\omega_{\mathcal{T},ij}^-$) on the horizontal axis. Transfers with the most years predicted are indicated in the form "exporter-importer" in black for $\Omega_{\mathcal{T}}^+$ and in grey for $\Omega_{\mathcal{T}}^-$.



on the top provides conclusive results about countries that are central in the network series Ω_t^+ . Among the countries where arms transfers are potentially under-reported, we find many Western European countries such as Belgium (BEL), Sweden (SWE), France (FRA), Spain (ESP) and Denmark (DNK). However, the list of presumed under-reporting is headed by Russia (RUS) and Turkey (TUR) but also Brazil (BRA), Israel (ISR) and China (CHN) have high scores. These countries also play a dominant role in Figure 9. In particular, exports from Brazil (BRA) to Russia (RUS), Hungary (HUN), Ukraine (UKR), China (CHN) and Japan (JPN) are likely to be frequently under-reported. Similarly, exports from Russia (RUS) to Cyprus (CYP), Denmark (DNK), Ireland (IRL), Turkey (TUR),

Sweden (SWE), Portugal (PRT) and Greece (GRC) are listed. We also find imports of Israel (ISR) from Finland (FIN), Belgium (BEL) and Austria (AUT) as well as exports of Belgium (BEL) to India (IND), Ukraine (UKR), Russia (RUS), Lebanon (LEB), Colombia (COL) and China (CHN).

Over-reporting networks Ω_t^- : Among the twelve countries with the highest Eigenvector centrality in Figure 8 is Croatia (HRV) as the only European country. There are however many countries from Asia such as Singapore (SGP), India (IND), Thailand (THA), Indonesia (IDN), Malaysia (MYS), South Korea (KOR) and Philippines (PHL). Furthermore, South Africa (ZAF), New Zealand (NZL), Australia (AUS) and Israel (ISR) are among the countries where trade activity is often over-reported. For Asian countries as well as for Australia and New Zealand this might mirror the fact that those countries export many SAA to Europe and the United States despite the strongly negative distance effect of the model. Furthermore, this network is very likely to be driven by bilateral agreements and historical developments not covered by the covariates. See for example in Figure 9 the number of over-reporting related to the Baltic countries Estonia (EST) Lithuania (LTU) and Latvia (LVA).

It remains to be emphasized that the constructions of the forensic networks relies on our model with corresponding assumptions and admittedly high degrees of uncertainty. As a consequence, it does not allow for definite statements about actual hidden transfers. However, many of the dyads listed in Figure 9 indeed have either traded massive amounts of civilian arms (e.g. AUT-CHN, BRA-RUS, RUS-BEL, RUS-DNK) or had frequent MCW trade relations (e.g. RUS-CYP) but almost no documented small arms transfers for military usage. Additionally, many of the countries that take central positions in the forensic networks are known for not being very transparent with respect to their SAA exports and imports (see e.g. the small arms transparency barometer).

6. Conclusion

In this paper we have modelled the volumes of international transfers of small arms and ammunition for the years 1992-2014 based on data provided by NISAT. As an analytical tool we combined the gravity model of trade with a modified SAR model that allows to enrich the analysis by endogenous network dependencies, accounting for exporter-related, importer-related and reciprocal dependency among the transfers in the network. Using a censored normal regression model we are able to include information provided by zero-valued transfers. The infeasible likelihood of the censored model is maximized using a Monte Carlo EM algorithm. The fitted model shows strong and stable endogenous network effect, especially related to the sender effect and the receiver effect but also some evidence

for reciprocity. Additionally, we find a high coefficient on path dependency and a close connection to the exports of civilian small arms. Conditional on that, the classical gravity hypothesis is confirmed with respect to the GDP of the importer and physical distance but only exceptionally with respect to political distance measures and the GDP of the exporter. This contrasts with the MCW network where distance plays no role, where political similarity and GDP of the exporter have a strong impact (see Thurner et al., 2018). Actually, this difference is plausible, as the technological requirements for the production of small and ammunition are relatively low, and strategic considerations of world-wide acting countries make geographic distances a negligible factor for MCW trade.

Building on our latent utility framework we were able to explore latent utility networks. With the construction of under-reporting and over-reporting networks we perform for the first time a forensic approach in this area highlighting especially potentially under-reported exports of Russia and Turkey. We refrain, of course, from making too far-reaching assertions. Note that we do not claim to provide unambiguous claims for intentional false reporting. However, we demonstrate that some zero entries in the SAA trading network tend to be not plausible.

Acknowledgements

We would like to thank Nic Marsh from NISAT for his useful comments and explanations on the data.

References

- Aitken, C. G. and F. Taroni (2004). *Statistics and the evaluation of evidence for forensic scientists*. Chichester: John Wiley & Sons.
- Akerman, A. and A. L. Seim (2014). The global arms trade network 1950–2007. *Journal of Comparative Economics* 42(3), 535–551.
- Augugliaro, L., A. Abbruzzo, and V. Vinciotti (2018). L1-penalized censored gaussian graphical model. *Biostatistics*, <https://doi.org/10.1093/biostatistics/kxy043>. to appear.
- Barigozzi, M., G. Fagiolo, and D. Garlaschelli (2010). Multinetwork of international trade: A commodity-specific analysis. *Physical Review E* 81(4), 046104.
- Bergstrand, J. H. (1992). On modeling the impact of arms reductions on world trade. In C. Isard and W. Anderton (Eds.), *Economics of arms reduction and the peace process*, pp. 121–142. Amsterdam: Elsevier Science Publishing.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B* 36(2), 192–236.
- Bivand, R., J. Hauke, and T. Kossowski (2013). Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis* 45(2), 150–179.
- Bivand, R. and G. Piras (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63(18), 1–36.
- Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79(1), 125–148.
- Bove, V., C. Deiana, and R. Nisticò (2018). Global arms trade and oil dependence. *The Journal of Law, Economics, and Organization*, <http://dx.doi.org/10.1093/jleo/ewy007>.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics* 6(1), 76–90.

- Correlates of War Project (2017). Formal interstate alliance dataset, 1648-2012, version 4.1. <http://www.correlatesofwar.org/data-sets/formal-alliances>. Accessed: 2017-02-06.
- Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5), 1–9.
- 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* 39(1), 1–38.
- Disdier, A.-C. and K. Head (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics* 90(1), 37–48.
- Doreian, P. (1989). Models of network effects on social actors. In L. Freeman, D. White, and K. Romney (Eds.), *Research Methods in Social Network Analysis*, pp. 295–317. Washington DC: George Mason University Press.
- Doreian, P., K. Teuter, and C.-H. Wang (1984). Network autocorrelation models: some monte carlo results. *Sociological Methods & Research* 13(2), 155–200.
- Dow, M., M. Burton, and D. White (1982). Network autocorrelation: A simulation study of a foundational problem. *Social Networks* 4, 169–200.
- Egger, P. H. and K. E. Staub (2016). Glm estimation of trade gravity models with fixed effects. *Empirical Economics* 50(1), 137–175.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters* 27(8), 861–874.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal* 13(3), 317–322.
- Franzese, R. J. and J. C. Hays (2007). Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis* 15(2), 140–164.
- Garlaschelli, D. and M. I. Loffredo (2005). Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications* 355(1), 138–144.

- Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn (2016). `mvt-norm`: Multivariate normal and t distributions. <http://CRAN.R-project.org/package=mvtnorm>. R package version 1.0-5.
- Gleditsch, K. S. (2013). Distance between capital cities. <http://privatewww.essex.ac.uk/~ksg/data-5.html>. Accessed: 2017-04-07.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation* 24(109), 23–26.
- Haug, M., M. Langvandslie, L. Lumpe, and N. Marsh (2002, January). Shining a light on small arms exports: The record of state transparency. Occasional Paper 4, Norwegian Initiative of Small Arms Transfers.
- Hays, J. C., A. Kachi, and R. J. Franzese (2010). A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology* 7(3), 406–428.
- Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of international economics*, Volume 4, pp. 131–195. Amsterdam: Elsevier Science Publishing.
- Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics* 123(2), 441–487.
- Henningsen, A. (2013). `censReg`: Censored regression (tobit) models. <https://cran.r-project.org/web/packages/censReg/censReg.pdf>. R package version 0.5-26.
- Herron, P., N. Marsh, and M. Schroeder (2011). Larger but less known - authorized light weapons transfers. <http://www.smallarmssurvey.org/fileadmin/docs/S-Trade-Update/SAS-Trade-Update-2018.pdf>. Accessed: 2019-21-05.
- Holtom, P., M. Bromley, and V. Simmel (2012). *Measuring international arms transfers*. Stockholm International Peace Research Institute.
- Holtom, P. and I. Pavesi (2018). Small arms survey - trade update 2018. <http://www.smallarmssurvey.org/fileadmin/docs/S-Trade-Update/SAS-Trade-Update-2018.pdf>. Accessed: 2019-06-02.

- Kang, L., R. Carter, K. Darcy, J. Kauderer, and S.-Y. Liao (2013). A fast monte carlo em algorithm for estimation in latent class model analysis with an application to assess diagnostic accuracy for cervical neoplasia in women with agc. *Journal of applied Statistics* 40(12), 2699.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data. Methods and models*. New York: Springer Science & Business Media.
- Lacombe, D. J. (2004). Does econometric methodology matter? an analysis of public policy using spatial econometric techniques. *Geographical Analysis* 36(2), 105–118.
- Leenders, R. T. A. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks* 24(1), 21–47.
- Leifeld, P., S. J. Cranmer, and B. A. Desmarais (2017). tnam: Temporal network autocorrelation models. <https://cran.r-project.org/package=tnam>. R package version 1.6.5.
- LeSage, J. P. and R. K. Pace (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science* 48(5), 941–967.
- LeSage, J. P. and R. K. Pace (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B* 44(2), 226–233.
- Magnus, J. R. and H. Neudecker (1988). *Matrix differential calculus with applications in statistics and econometrics*. Chichester: John Wiley & Sons.
- Marsh, N. (2017). Norwegian initiative on small arms transfers, firearms and ammunition trade data 1992-2014. <http://www.nisat.prio.org>. Accessed: 27.03.2017.
- Marsh, N. J. and T. L. a. McDougal (2016). Illicit small arms prices: Introducing two new datasets. Technical report, Small Arms Data Observatory.
- Marshall, M. G. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2016. <http://www.systemicpeace.org/inscrdata.html>. Accessed: 2017-06-02.
- Martinez-Zarzoso, I. and F. Johannsen (2017). The gravity of arms. *Defence and Peace Economics*, <https://doi.org/10.1080/10242694.2017.1324722>.

- Mathai, A. M. and S. B. Provost (1992). *Quadratic forms in random variables: theory and applications*. London: Taylor & Francis.
- McLachlan, G. and T. Krishnan (2007). *The EM algorithm and extensions*. Hoboken: Wiley & Sons.
- Metz, F. and K. Ingold (2017). Politics of the precautionary principle: assessing actors' preferences in water protection policy. *Policy Sciences* 50(4), 721–743.
- Oakes, D. (1999). Direct calculation of the information matrix via the em. *Journal of the Royal Statistical Society. Series B* 61(2), 479–482.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robert, C. and G. Casella (2004). *Monte Carlo methods*. New York: Springer Science & Business Media.
- Schumacher, F. L., V. H. Lachos, and D. K. Dey (2017). Censored regression models with autoregressive errors: A likelihood-based perspective. *Canadian Journal of Statistics* 45(4), 375–392.
- Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation* 24(111), 647–656.
- SIPRI (2017a). Arms transfers database. Accessed: 2017-03-10.
- SIPRI (2017b). Arms transfers database - methodology. Accessed: 2017-03-10.
- Suesse, T. and A. Zammit-Mangion (2017). Computational aspects of the em algorithm for spatial econometric models with missing data. *Journal of Statistical Computation and Simulation* 87(9), 1767–1786.
- Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society. Series B* 23(1), 223–229.
- Thurner, P. W., Schmid, C. Christian, Skyler, and G. Kauermann (2018). The network of major conventional weapons transfers 1950-2013. <http://journals.sagepub.com/doi/10.1177/0022002718801965>. Online First: Journal of Conflict Resolution.
- Tinbergen, J. (1962). Shaping the world economy: An analysis of world trade flows. *New York Twentieth Century Fund* 5(1), 27–30.

- UCDP (2019). UCDP. <http://ucdp.uu.se/downloads/>. Accessed: 2018-12-01.
- Vaida, F. and L. Liu (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics* 18(4), 797–817.
- Ward, M. D., J. S. Ahlquist, and A. Rozenas (2013). Gravity’s rainbow: A dynamic latent space model for the world trade network. *Network Science* 1(1), 95–118.
- Wei, G. C. and M. A. Tanner (1990). A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association* 85(411), 699–704.
- Wilhelm, S. et al. (2012). Moments calculation for the doubly truncated multivariate normal density. *arXiv preprint arXiv:1206.5387*.
- World Bank (2017). World bank open data, real GDP. <http://data.worldbank.org/>. Accessed: 2017-04-01.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3(1), 32–35.
- Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature* 50(3), 731–69.

A. Annex

A.1. Descriptives

Table 2: Different arms types included in the NISAT dataset with three digit arms category code, weapon type, subcategory and number of transfers in the dataset.

Code	PRIO Weapons Type	Subcategories
200	Small Arms	
210		Pistols & Revolvers
230		Rifles/Shotguns (Military)
233		Assault Rifles
234		Carbines
235		Sniper Rifles
237		Semi-automatic Rifles (Military)
239		Shotguns (Military)
240		Machine Guns
243		Sub Machine Guns
245		Light Machine Guns
247		General Purpose Machine Guns
250		Military Weapons
260		Military Firearms
270		Machine Guns All Types
300	Light Weapons	
310		Heavy Machine Guns $\leq 12.7\text{mm}$
400	Ammunition	
415		Small Arms Ammunition
417		Small Calibre Ammunition $\leq 12.7\text{mm}$
418		Shotgun Cartridges

Source: nisat.prio.org.

Table 3: The 59 major exporting and importing countries of the small arms and ammunition dataset with ISO 3 country codes.

Country	ISO3 Code	Country	ISO3 Code	Country	ISO3 Code
Argentina	ARG	India	IND	Poland	POL
Australia	AUS	Indonesia	IDN	Portugal	PRT
Austria	AUT	Ireland	IRL	Romania	ROM
Belgium	BEL	Israel	ISR	Russia	RUS
Brazil	BRA	Italy	ITA	Saudi Arabia	SAU
Bulgaria	BGR	Japan	JPN	Serbia	SRB
Canada	CAN	Kenya	KEN	Singapore	SGP
Chile	CHL	South Korea	KOR	Slovenia	SVN
China	CHN	Kuwait	KWT	South Africa	ZAF
Colombia	COL	Latvia	LVA	Spain	ESP
Croatia	HRV	Lebanon	LBN	Sweden	SWE
Cyprus	CYP	Lithuania	LTU	Switzerland	CHE
Denmark	DNK	Malaysia	MYS	Thailand	THA
Egypt	EGY	Mexico	MEX	Turkey	TUR
Estonia	EST	Netherlands	NLD	Ukraine	UKR
Finland	FIN	New Zealand	NZL	Un. Arab Emirates	ARE
France	FRA	Norway	NOR	United Kingdom	GBR
Germany	DEU	Pakistan	PAK	United States	USA
Greece	GRC	Peru	PER	Uruguay	URY
Hungary	HUN	Philippines	PHL	-	-

B. Supplementary Material

B.1. Derivatives of the complete log-likelihood

The complete log-likelihood is given by

$$\ell_{comp}(\theta) = -\frac{N}{2} \log(2\pi\sigma^2) + \log(|A(\boldsymbol{\rho})|) - \frac{(A(\boldsymbol{\rho})\tilde{Y} - X\beta)^T(A(\boldsymbol{\rho})\tilde{Y} - X\beta)}{2\sigma^2},$$

with score vector

$$\begin{aligned} \frac{\partial \ell_{comp}(\theta)}{\partial \beta} &= \frac{1}{\sigma^2} X^T [A(\boldsymbol{\rho})\tilde{Y} - X\beta] \\ \frac{\partial \ell_{comp}(\theta)}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{\tilde{Y}^T(A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho})\tilde{Y} - 2\beta^T X^T A(\boldsymbol{\rho})\tilde{Y} + \beta^T X^T X \beta}{2\sigma^4} \\ \frac{\partial \ell_{comp}(\theta)}{\partial \rho_k} &= -\text{tr}(B(\boldsymbol{\rho})W_k) \\ &\quad - \frac{\tilde{Y}^T[-W_k - W_k^T + 2\rho_k W_k^T W_k + \sum_{l \neq k} \rho_l (W_k^T W_l + W_l^T W_k)]\tilde{Y} + 2\beta^T X^T W_k \tilde{Y}}{2\sigma^2}. \end{aligned} \quad (12)$$

And the corresponding Hessian results in

$$\begin{aligned} \frac{\partial^2 \ell_{comp}(\theta)}{\partial \beta \partial \beta^T} &= -\frac{1}{\sigma^2} X^T X \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4} X^T [A(\boldsymbol{\rho})\tilde{Y} - X\beta] \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \beta \partial \rho_k} &= -\frac{1}{\sigma^2} X^T W_k^T \tilde{Y} \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \sigma^2 \partial \sigma^2} &= \frac{N}{2\sigma^4} - \frac{\tilde{Y}^T(A(\boldsymbol{\rho}))^T A(\boldsymbol{\rho})\tilde{Y} - 2\beta^T X^T A(\boldsymbol{\rho})\tilde{Y} + \beta^T X^T X \beta}{\sigma^6} \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \rho_k \partial \sigma^2} &= \frac{\tilde{Y}^T[-W_k - W_k^T + 2\rho_k W_k^T W_k + \sum_{l \neq k} \rho_l (W_k^T W_l + W_l^T W_k)]\tilde{Y} + 2\beta^T X^T W_k \tilde{Y}}{2\sigma^4} \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \rho_k \partial \rho_k} &= -\text{tr}\left(B(\boldsymbol{\rho})W_k B(\boldsymbol{\rho})W_k\right) - \frac{\tilde{Y}^T W_k^T W_k \tilde{Y}}{\sigma^2} \\ \frac{\partial^2 \ell_{comp}(\theta)}{\partial \rho_k \partial \rho_l} &= -\text{tr}\left(B(\boldsymbol{\rho})W_l B(\boldsymbol{\rho})W_k\right) - \frac{\tilde{Y}^T (W_k^T W_l + W_l^T W_k) \tilde{Y}}{2\sigma^2}. \end{aligned} \quad (13)$$

Where we use Jacobi's formula (see Magnus and Neudecker, 1988) that allows to express the derivative of a matrix determinant in terms of the derivative of the matrix and its adjugate ($\text{adj}(\cdot)$). Resulting in

$$\frac{\partial \log(|A(\boldsymbol{\rho})|)}{\partial \rho_k} = -|A(\boldsymbol{\rho})|^{-1} \text{tr}[\text{adj}(A(\boldsymbol{\rho}))W_k] = -\text{tr}(B(\boldsymbol{\rho})W_k),$$

for the third equation in (12). The differentiation of the trace

$$\frac{\partial \text{tr}(B(\boldsymbol{\rho})W_k)}{\partial \rho_l} = \text{tr}\left(\frac{\partial B(\boldsymbol{\rho})}{\partial \rho_l}W_k\right) = -\text{tr}\left(B(\boldsymbol{\rho})\frac{\partial A(\boldsymbol{\rho})}{\partial \rho_l}B(\boldsymbol{\rho})W_k\right) = \text{tr}(B(\boldsymbol{\rho})W_lB(\boldsymbol{\rho})W_k)$$

is used for the sixth and seventh equation in (13).

B.2. Practical Implementation of the Algorithm

The gradient

$$\frac{\partial \tilde{Q}(\boldsymbol{\rho}|\theta_0)}{\partial \rho_k} = -\text{tr}(B(\boldsymbol{\rho})W_k) - \frac{N}{2} \frac{R_k^*(\boldsymbol{\rho}) - \tilde{Y}^{*\text{T}}H_k(\boldsymbol{\rho})\tilde{Y}^*}{S^*(\boldsymbol{\rho}) - \tilde{Y}^{*\text{T}}(A(\boldsymbol{\rho}))^{\text{T}}HA(\boldsymbol{\rho})\tilde{Y}^*}. \quad (14)$$

can be used to maximize

$$\tilde{Q}(\boldsymbol{\rho}|\theta_0) = \kappa + \log(|A(\boldsymbol{\rho})|) - \frac{N}{2} \log\left(S^*(\boldsymbol{\rho}) - \tilde{Y}^{*\text{T}}(A(\boldsymbol{\rho}))^{\text{T}}HA(\boldsymbol{\rho})\tilde{Y}^*\right). \quad (15)$$

by applying the BFGS optimization routine (see Broyden, 1970, Fletcher, 1970, Goldfarb, 1970 and Shanno, 1970). The implementation of the BFGS algorithm in R (R Core Team, 2016) is provided by the base function `optim`. More computational stability for the maximization of equation (15) is reached by defining $\lambda = (\lambda_1, \dots, \lambda_N)^{\text{T}}$ as the vector of eigenvalues of $A(\boldsymbol{\rho})$ and replacing $\log(|A(\boldsymbol{\rho})|)$ by $\sum_{r=1}^N \log(\lambda_r(\boldsymbol{\rho}))$ in equation (15), see Bivand and Piras (2015). The starting value for the algorithm can be found by using a maximum pseudolikelihood estimate (MPLE), using W_1Y, \dots, W_qY as exogenous covariates in a censored regression model, provided by the R package `censReg` (Henningsen, 2013). Since the observed log-likelihood cannot be evaluated, we define $\hat{\theta}$ as the solution of the maximization problem if $(\hat{\theta} - \theta_0)^{\text{T}}(\hat{\theta} - \theta_0) < 0.1$, otherwise we set $\theta_0 = \hat{\theta}$ and re-iterate until the stopping criteria is satisfied.

B.3. Approximation of the Fisher Information

Louis (1982) and Oakes (1999) provide formulas for the Fisher information of the observed likelihood. We follow the recommendation of McLachlan and Krishnan (2007), arguing that Louis's formula is best suited for the MCEM and provides a conservative measure of the standard errors. Therefore, we calculate the observed information based on

$$-\frac{\partial^2 \ell_{obs}(\theta)}{\partial \theta \partial \theta^T} = \mathbb{E}_\theta \left[-\frac{\partial^2 \ell_{comp}(\theta)}{\partial \theta \partial \theta^T} \Big| \tilde{Y}_o, X, \mathcal{M} \right] - \mathbb{E}_\theta \left[\frac{\partial \ell_{comp}(\theta)}{\partial \theta} \left(\frac{\partial \ell_{comp}(\theta)}{\partial \theta} \right)^T \Big| Y_o, X, \mathcal{M} \right] \\ + \mathbb{E}_\theta \left[\frac{\partial \ell_{comp}(\theta)}{\partial \theta} \Big| Y_o, X, \mathcal{M} \right] \left(\mathbb{E}_\theta \left[\frac{\partial \ell_{comp}(\theta)}{\partial \theta} \Big| Y_o, X, \mathcal{M} \right] \right)^T. \quad (16)$$

Note that the second term of (16) depends not only on the first and second but also on the third and fourth conditional moment of the truncated multivariate normal and cannot be evaluated analytically therefore.

In order to approximate the observed information we are using the results of Robert and Casella (2004, p. 187) and Kang et al. (2013, Section 3.7) that allow for an approximation of the observed information based on the score and Hessian of the complete likelihood. Hence, we can use the results from Section B.1 for the following procedure.

We draw $w = 1000$ times potential realizations $\tilde{Y}_{s,sim}$ from the truncated version of

$$\tilde{Y}^c \sim \mathcal{N}_{N_m}(\mu_m + \Sigma_{mo} \Sigma_{oo}^{-1} (\tilde{Y}_o - \mu_o), \Sigma_{mm} - \Sigma_{mo} \Sigma_{oo}^{-1} \Sigma_{om}) \quad (17)$$

using the package `TruncatedNormal` (Botev, 2017). Those are stored for each draw s in a vector $\tilde{Y}_{s,sim}^* = (\tilde{Y}_o, \tilde{Y}_{s,sim})$.

Then we calculate the score and Hessian from equations (12) and (13) w times, where we replace \tilde{Y} by $\tilde{Y}_{s,sim}^*$ in each equation and index them by s , allowing to calculate the empirical version of (16) by approximating the expectations by means.

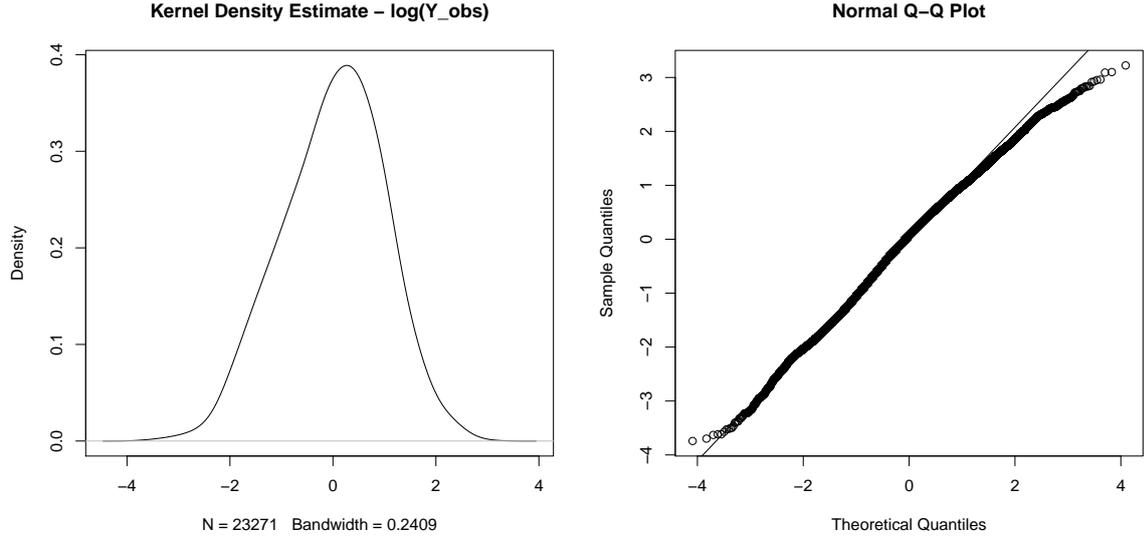
$$-\frac{\partial^2 \ell_{obs}(\theta)}{\partial \theta \partial \theta^T} \approx \frac{1}{w} \sum_{s=1}^w \left[-\frac{\partial^2 \ell_{s,comp}(\theta)}{\partial \theta \partial \theta^T} - \left(\frac{\partial \ell_{s,comp}(\theta)}{\partial \theta} - \frac{1}{w} \sum_{s=1}^w \frac{\partial \ell_{s,comp}(\theta)}{\partial \theta} \right) \left(\frac{\partial \ell_{s,comp}(\theta)}{\partial \theta} - \frac{1}{w} \sum_{s=1}^w \frac{\partial \ell_{s,comp}(\theta)}{\partial \theta} \right)^T \right].$$

This gives an estimator for the observed information. Standard errors are obtained by the square root of the diagonal elements of the inverted approximated matrix.

B.4. Data Transformation

In Figure 10 we show the distribution of the observed log-transformed response variable. The data is pooled over all years and standardized to have mean zero and variance one.

Figure 10: Kernel Density Estimate of the log-transformed standardized observed response variable pooled for all time periods (left). Q-Q plot for the log-transformed standardized response variable (right).



The panel on the left side shows a kernel density estimate and the panel on the right gives a Q-Q plot.

B.5. Conditional Probabilities

Based on the fitted coefficients $\hat{\theta}_t = (\hat{\beta}_t, \hat{\rho}_t, \hat{\sigma}_t^2)$ and our model assumptions, we can represent the joint distribution of the latent utility network Z_t via a multivariate normal

$$Z_t \sim \mathcal{N}_N(\hat{\mu}_t, \hat{\Sigma}_t),$$

where $\hat{\mu}_t = B(\hat{\rho}_t)X\hat{\beta}_t$ and $\hat{\Sigma}_t = B(\hat{\rho}_t)(B(\hat{\rho}_t))^T\hat{\sigma}_t^2$. Given that, define $Z_{t,-ij}$ as the $(N-1)$ -dimensional vector, containing all entries of Z_t except $Z_{t,ij}$. Additionally, for example in the case that ij is the first entry of Z_t , rearrange $\hat{\Sigma}_t$ such that

$$\hat{\Sigma}_t = \begin{pmatrix} \hat{\Sigma}_{t,ij,ij} & \hat{\Sigma}_{t,ij,-ij} \\ \hat{\Sigma}_{t,-ij,ij} & \hat{\Sigma}_{t,-ij,-ij} \end{pmatrix}.$$

Then, the conditional distribution of $Z_{t,ij}$ is given by a univariate normal distribution

$$\begin{aligned} Z_{t,ij}|X, Z_{t,-ij} &\sim \mathcal{N}(\hat{\mu}_{t,ij|-ij}, \hat{\Sigma}_{t,ij|-ij}), \text{ where} \\ \hat{\mu}_{t,ij|-ij} &= \hat{\mu}_{t,ij} + \hat{\Sigma}_{t,ij,-ij} \hat{\Sigma}_{t,-ij,-ij}^{-1} (Z_{t,-ij} - \hat{\mu}_{t,-ij}) \text{ and} \\ \hat{\Sigma}_{t,ij|-ij} &= \hat{\Sigma}_{t,ij,ij} - \hat{\Sigma}_{t,ij,-ij} \hat{\Sigma}_{t,-ij,-ij}^{-1} \hat{\Sigma}_{t,-ij,ij}. \end{aligned}$$

We are interested in a possible state of the network, where the latent utility is allowed to be greater c_t . Therefore, we insert the expectation for the non-observed utility in $Z_{t,-ij}$ and denote this by $\tilde{Z}_{t,-ij}$. Consequently, we can calculate the probability of $Z_{t,ij}$ being greater than c_t using

$$\begin{aligned} \pi_{t,ij} &= \mathbb{P}(Z_{t,ij} > c_t | X_{t,ij}, \tilde{Z}_{t,-ij}; \hat{\theta}) = 1 - \mathbb{P}(Z_{t,ij} \leq c_t | X_{t,ij}, \tilde{Z}_{t,-ij}; \hat{\theta}) \\ &= 1 - \int_{-\infty}^{c_t} \frac{1}{\sqrt{2\pi \hat{\Sigma}_{t,ij|-ij}^2}} \exp\left(-\frac{(U - \hat{\mu}_{t,ij|-ij})^2}{2\hat{\Sigma}_{t,ij|-ij}^2}\right) dU. \end{aligned}$$

The probability $\pi_{t,ij}$ can be interpreted as the probability that the latent utility of a transfer from country i to country j is higher than the threshold c_t conditional on the covariates X_t and the remaining network, where no transfer is restricted to be smaller c_t .

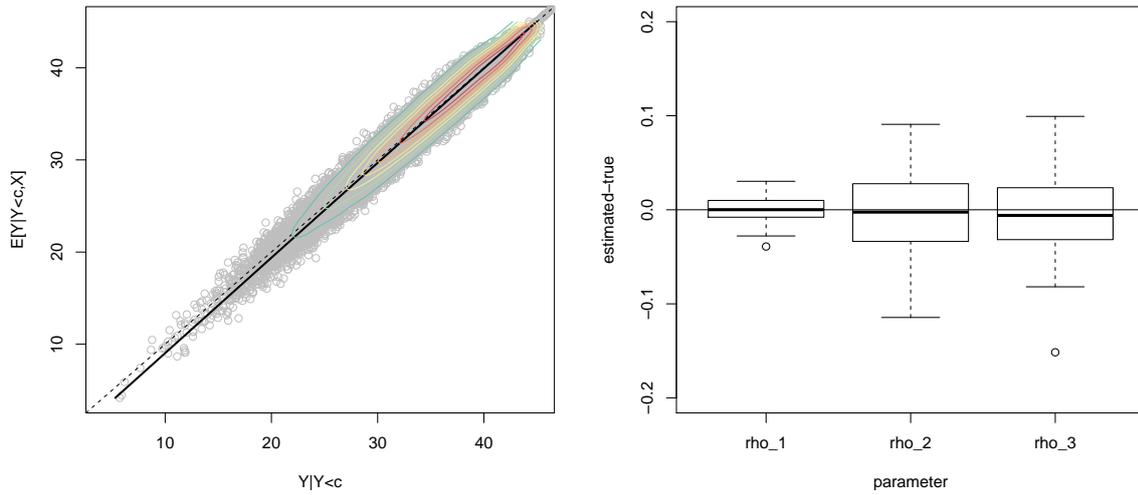
B.6. Simulation study - Endogenous effects and approximation of censored variables

In order to analyse the properties of our estimator, we use the following Data Generating Process (DGP1)

$$\begin{aligned} \boldsymbol{\rho} &= (0.1, 0.2, 0.3)^T, \boldsymbol{\beta} = (1, 2, 3, 4, 5)^T, p = 5, n = 20, N = 380 \\ X &\sim \mathcal{N}_p(\mathbf{1}, I_p) \\ Z &\sim \mathcal{N}_N(B(\boldsymbol{\rho})X\boldsymbol{\beta}, B(\boldsymbol{\rho})B(\boldsymbol{\rho})^T) \\ \tilde{Z}_{ij} &= I(Z_{ij} > q_{0.75}(Z))Z_{ij}, \text{ for } i \neq j = 1, \dots, n. \end{aligned} \tag{18}$$

Here, $q_{0.75}(Z)$ denotes the 75% quantile and we are censoring the network towards an observed density of 0.25. Note, that DGP1 is not subject to under-reporting and all censored responses are in fact below the censoring threshold. The results of running DGP1 100 times and applying the estimation procedure are summarized in Figure 11. On the left panel, we show the true but censored values against the expected values from the last E-Step, together with contour curves and a non-parametric fit for the mean in solid black.

Figure 11: Results of DGP1. Expected values against true censored values for all simulations (left). Angle bisector in dashed black, non-parametric mean in black and colored contours. Boxplots for the difference between estimated and true values for ρ (right).



It can be seen that the expected values approximate the latent variables very well. The right panel of Figure 11 shows boxplots for the difference between the true values of ρ and the estimated parameters. It indicates that as we are able to find unbiased estimates of the endogenous parameters despite the enormous amount of censoring.

Chapter 6

Regression-based network reconstruction with nodal and dyadic covariates and random effects

Contributing Article:

Michael Lebacher and Göran Kauermann (2019): *Regression-based network reconstruction with nodal and dyadic covariates and random effects*.

Under review in the Journal of the American Statistical Association (Theory and Methods).

arXiv preprint <https://arxiv.org/abs/1903.11886>

Code at https://github.com/lebachelor/regression_network_reconstruction

Further Versions:

Michael Lebacher and Göran Kauermann (2019): *Regression-based network reconstruction with nodal and dyadic covariates and random effects*

Proceedings of the 33th International Workshop on Statistical Modelling, 1:220–226.

Author Contributions:

The initial ideas for solving the network reconstruction problem (including approaches with random effects) were provided by Göran Kauermann together with the idea of using bootstrap prediction intervals. Furthermore, Göran Kauermann gave very valuable input for the simulation section. The contribution of Michael Lebacher is given by deriving the model with and without random effects, the implementation of the model in R, including data manipulation, fitting and simulation. The concrete formulation and the algorithm is also due to Michael Lebacher. Furthermore, Michael Lebacher wrote the major part of the manuscript. Both authors contributed to the manuscript writing and were involved in extensive proof-reading.

Regression-based Network Reconstruction with Nodal and Dyadic Covariates and Random Effects

Michael Lebacher and Göran Kauermann*

Department of Statistics, Ludwig-Maximilians Universität München

Abstract

Network (or matrix) reconstruction is a general problem which occurs if the margins of a matrix are given and the matrix entries need to be predicted. In this paper we show that the predictions obtained from the iterative proportional fitting procedure (IPFP) or equivalently maximum entropy (ME) can be obtained by restricted maximum likelihood estimation relying on augmented Lagrangian optimization. Based on the equivalence we extend the framework of network reconstruction towards regression by allowing for exogenous covariates and random heterogeneity effects. The proposed estimation approach is compared with different competing methods for network reconstruction and matrix estimation. Exemplary, we apply the approach to interbank lending data, provided by the Bank for International Settlement (BIS). This dataset provides full knowledge of the real network and is therefore suitable to evaluate the predictions of our approach. It is shown that the inclusion of exogenous information allows for superior predictions in terms of L_1 and L_2 errors. Additionally, the approach allows to obtain prediction intervals via bootstrap that can be used to quantify the uncertainty attached to the predictions.

Keywords: Bootstrap; Interbank lending; Inverse problem; Iterative proportional fitting; Network analysis; Maximum entropy; Matrix estimation

*We would like to thank Samantha Cook and Kimmo Soramäki from Financial Network Analytics (www.fna.fi) for providing and explaining the data and the problem to us. The project was supported by the European Cooperation in Science and Technology [COST Action CA15109]. We also acknowledge funding provided by the German Research Foundation (DFG) for the project KA 1188/10-1.

1. Introduction

The problem of how to obtain predictions for unknown entries of a matrix, given restrictions on the row and column sums is a problem that comes with many labels. Without a sharp distinction of names and fields, some non exhaustive examples for keywords that are related to very similar settings are *Network Tomography* and *Traffic Matrix Estimation* in Computer Sciences and Machine Learning (e.g. Vardi, 1996, Coates et al., 2002, Hazelton, 2010, Airoidi and Blocker, 2013, Zhou et al., 2016), *Input-Output Analysis* in Economics (e.g. Bacharach, 1965, Miller and Blair, 2009), *Network Reconstruction* in Finance and Physics (e.g. Sheldon and Maurer, 1998, Squartini and Garlaschelli, 2011, Mastrandrea et al., 2014, Gandy and Veraart, 2018), *Ecological Inference* in Political Sciences (e.g. King, 2013, Klima et al., 2016), *Matrix Balancing* in Operation Research (e.g. Schneider and Zenios, 1990) and many more.

An old but nevertheless popular solution to problems of this kind is the so called *iterative proportional fitting procedure* (IPFP), firstly introduced by Deming and Stephan (1940) as a mean to obtain consistency between sampled data and population-level information. In essence, this simple procedure iteratively adjusts the estimated entries until the row and column sums of the estimates match the desired ones. In the statistics literature, this procedure is frequently used as a tool to obtain maximum likelihood estimates for log-linear models in problems involving three-way and higher order tables (Fienberg et al., 1970, Bishop et al., 1975, Haberman, 1978, 1979). Somewhat parallel, the empirical economics literature, concerned with the estimation of Input-Output matrices, proposed a very similar approach (Bacharach, 1970), often called *RAS algorithm*. Here, the entries of the matrix must be consistent with the inputs and the outputs. The solution to the problem builds on the existence of a prior matrix that is iteratively transformed to a final matrix that is similar to the initial one but matches the input-output requirements. Although the intention is somewhat different, the algorithm is effectively identical to IPFP (Onuki, 2013). The popularity of the procedure can also be explained by the fact that it provides a solution for the so called *maximum entropy* (ME) problem (Malvestuto, 1989, Upper, 2011, Elsinger et al., 2013). In Computer Sciences, flows within router networks are often estimated using *Raking* and so called *Gravity Models* (see Zhang et al., 2003). Raking is in fact identical to IPFP and the latter can be interpreted as a special case of the former.

In this paper, we propose an estimation algorithm that builds on augmented Lagrangian optimization (Powell, 1969, Hestenes, 1969) and can provide the same predictions as IPFP but is flexible enough to be extended toward more general concepts. In particular we propose to include exogenous covariates and random effects to improve the predictions of the missing matrix entries. Furthermore, we compare our approach with competing models using real data. To do so, we look at an international financial network of claims and liabilities where we pretend that the inner part of the matrix is unknown. Since in the data at hand the full matrix is in fact available we can carry out a competitive comparison with alternative

routines. Note that commonly the inner part of the financial network remains unknown but finding good estimates for the matrix entries is essential for central banks and financial regulators. This is because it is a necessary prerequisite for evaluating systemic risk within the international banking system. See e.g. a very recent study by researchers from 15 different central banks (Anand et al., 2018) where the question of how to estimate financial network linkages was identified as being crucial for contagion models and stress tests. Our proposal has therefore a direct practical contribution.

The paper is structured as follows. In Section 2 we relate maximum entropy, maximum likelihood and IPFP. In Section 3 we introduce our model and discuss estimation and inference as well as potential extensions. After a short data description in Section 4 we apply the approach, compare different models and give a small simulation study that shows properties of the estimator. Section 5 concludes our paper.

2. Modelling approach

2.1. Notation

Our interest is in predicting non-negative, directed dyadic variables x_{ij}^t among $i, j = 1, \dots, n$ observational units at time points $t = 1, \dots, T$. The restriction to non-negative entries is henceforth referred to as *non-negativity constraint*. We do not allow for self-loops and leave elements x_{ii}^t undefined. Hence, the number of unknown variables at each time point t is given by $N = n(n - 1)$. Let $\mathbf{x}^t = (x_{12}^t, \dots, x_{1n}^t, x_{21}^t, \dots, x_{n(n-1)}^t)^T$ be an N -dimensional column vector and define $\mathcal{I} = \{(i, j) : i, j = 1, \dots, n; i \neq j\}$ as the corresponding ordered index set. We denote the i th row sum by $y_i^t = x_{i\bullet}^t = \sum_{j \neq i} x_{ij}^t$ and the j th column sum by $y_{n+j}^t = x_{\bullet j}^t = \sum_{i \neq j} x_{ij}^t$. Stacking the row and column sums, results in the $2n$ -dimensional column vector \mathbf{y}^t . Furthermore, we define the known binary $(2n \times N)$ routing matrix \mathbf{A} such that the linear relation

$$\mathbf{y}^t = \mathbf{A}\mathbf{x}^t, \text{ for } t = 1, \dots, T \quad (1)$$

holds. Henceforth, we will refer to relation (1) as *marginal restrictions*. Furthermore, we denote each row of \mathbf{A} by the row vector $\mathbf{A}_r = (a_{r1}, \dots, a_{rN})$. Hence, we can represent the marginal restrictions row wise by

$$\mathbf{A}_r \mathbf{x}^t = y_r^t, \text{ for } r = 1, \dots, 2n \text{ and } t = 1, \dots, T.$$

Note that in cases where some elements of \mathbf{y}^t are zero, the number of unknown variables to predict decreases and matrix \mathbf{A} must be rearranged accordingly. In the following we will ignore this issue and suppress the time-superscript for ease of notation. Random variables and vectors are indicated by upper case letters, realizations as lower case letters.

2.2. Maximum entropy, iterative proportional fitting and maximum likelihood

Besides the long known relation between maximum entropy (ME) and IPFP, there also exists an intimate relation between maximum entropy and maximum likelihood that is formalized for example by Golan and Judge (1996) and is known as the *Duality theorem*, see for example Brown (1986) and Dudík et al. (2007). Also in so-called configuration models (Squartini and Garlaschelli, 2011, Mastrandrea et al., 2014) the connection between maximum entropy and maximum likelihood is a central ingredient for network reconstruction.

In the following (i) we rely on the work of Golan and Judge (1996), Squartini and Garlaschelli (2011) and Muñoz-Cobo et al. (2017) in order to briefly derive the ME-distribution in the given setting. (ii) After that, we show that IPFP indeed maximizes the ME-distribution. (iii) Based on the first two results, we show that we can arrive at the same result as IPFP by constrained maximization of a likelihood where each matrix entry comes from an exponential distribution.

(i) *Maximum entropy distribution:* We formalize the problem by defining the Shannon entropy functional of the system as

$$H[f] = - \int_{\mathcal{X}} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x},$$

where we make it explicit in the notation that the functional $H[f]$ takes the function f as input. The support of f is given by $\mathcal{X} \in \mathbb{R}_+^N$, ensuring the non-negativity constraint. Furthermore, we require that the density function $f : \mathcal{X} \rightarrow \mathbb{R}_+$ integrates to unity

$$\int_{\mathcal{X}} f(\mathbf{x}) d\mathbf{x} = 1. \quad (2)$$

We denote the expectation of the random vector \mathbf{X} by $\boldsymbol{\mu}$ and formulate the marginal restrictions in terms of linear restrictions on $\boldsymbol{\mu}$ which we specify as

$$\int_{\mathcal{X}} \mathbf{A}_r \mathbf{x} f(\mathbf{x}) d\mathbf{x} = \mathbf{A}_r \boldsymbol{\mu} = y_r \text{ for } r = 1, \dots, 2n. \quad (3)$$

Combining the constraints (2) and (3) results into the Lagrangian functional

$$\mathcal{L}[f] = - \int_{\mathcal{X}} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} - \lambda_0 \left(\int_{\mathcal{X}} f(\mathbf{x}) d\mathbf{x} - 1 \right) - \sum_{r=1}^{2n} \lambda_r \left(\int_{\mathcal{X}} \mathbf{A}_r \mathbf{x} f(\mathbf{x}) d\mathbf{x} - y_r \right) \quad (4)$$

with Lagrange multipliers $\lambda_r > 0$ for $r = 0, \dots, 2n$. The solution can be found using the Euler-Lagrange equation (Dym and Shames, 2013), stating that a functional of the form

$\int_{\mathcal{X}} L(\mathbf{x}, f(\mathbf{x}), f'(\mathbf{x}))d\mathbf{x}$ is stationary (i.e. its first order derivative is zero) if

$$\frac{\partial L}{\partial f} = \frac{d}{d\mathbf{x}} \frac{\partial L}{\partial f'}. \quad (5)$$

If the Lagrangian functional does not depend on the derivative of $f(\cdot)$, we find the right hand side in equation (5) to be zero so that no derivative appears. For the Lagrangian functional (4) this provides

$$-\log(\hat{f}(\mathbf{x})) - 1 - \lambda_0 - \sum_{r=1}^{2n} \lambda_r \mathbf{A}_r \mathbf{x} = 0. \quad (6)$$

Rearranging the terms in (6) results in the maximum entropy distribution

$$\hat{f}(\mathbf{x}) = \exp \left\{ - \sum_{r=1}^{2n} \lambda_r \mathbf{A}_r \mathbf{x} - 1 - \lambda_0 \right\}, \text{ for } \mathbf{x} \in \mathcal{X}. \quad (7)$$

In order to ensure restriction (2) we set $\exp(1 + \lambda_0) = c(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{2n})$ is the parameter vector and

$$c(\boldsymbol{\lambda}) = \int_{\mathcal{X}} \exp \left\{ - \sum_{r=1}^{2n} \lambda_r \mathbf{A}_r \mathbf{x} \right\} d\mathbf{x},$$

where $\lambda_r > 0$ for $r = 1, \dots, 2n$ ensures integration to a finite value. Taken together, this leads to the exponential family distribution

$$\hat{f}(\mathbf{x}) = \frac{1}{c(\boldsymbol{\lambda})} \exp \left\{ - \sum_{r=1}^{2n} \lambda_r \mathbf{A}_r \mathbf{x} \right\}, \text{ for } \mathbf{x} \in \mathcal{X}. \quad (8)$$

Apparently, the sufficient statistics in (8) result through

$$\mathbf{A}_r \mathbf{x} = y_r, \text{ for } r = 1, \dots, 2n$$

and hence, we can characterize the N dimensional random variable \mathbf{X} in terms of $2n$ parameters $\boldsymbol{\lambda}$. Using (8), the second order condition results from

$$-\frac{1}{\hat{f}(\mathbf{x})} < 0, \forall \mathbf{x} \in \mathcal{X}$$

and ensures that \hat{f} is indeed a maximizer.

(ii) *IPFP and the maximum entropy distribution:* In order to solve for the parameters of the maximum entropy distribution we take the first derivative of the log-likelihood obtained

from (8), i.e.

$$\hat{\ell}(\boldsymbol{\lambda}) = -\log(c(\boldsymbol{\lambda})) - \sum_{r=1}^{2n} \lambda_r y_r. \quad (9)$$

Since (8) is an exponential family distribution we can use the relation

$$\frac{\partial \log(c(\boldsymbol{\lambda}))}{\partial \lambda_r} = -\mathbb{E}_{\boldsymbol{\lambda}}[\mathbf{A}_r \mathbf{x}], \text{ for } r = 1, \dots, 2n$$

and the maximum likelihood estimator $\hat{\boldsymbol{\lambda}}$ results from the score equations

$$1 = \frac{\mathbf{A}_r \mathbf{x}}{\mathbb{E}_{\hat{\boldsymbol{\lambda}}}[\mathbf{A}_r \mathbf{x}]} = \frac{y_r}{\mathbb{E}_{\hat{\boldsymbol{\lambda}}}[\mathbf{A}_r \mathbf{x}]}, \text{ for } r = 1, \dots, 2n. \quad (10)$$

If we now multiply the left and the right hand side of (10) by parameter λ_r we get

$$\lambda_r = \lambda_r \frac{y_r}{\mathbb{E}_{\hat{\boldsymbol{\lambda}}}[\mathbf{A}_r \mathbf{x}]}, \text{ for } r = 1, \dots, 2n$$

and can solve the problem using fixed-point iteration (Dahmen and Reusken, 2006). That is we fix the right hand side to λ_r^{k-1} and update the left side to λ_r^k through

$$\lambda_r^k = \lambda_r^{k-1} \frac{y_r}{\mathbb{E}_{\boldsymbol{\lambda}^{k-1}}[\mathbf{A}_r \mathbf{x}]}, \text{ for } r = 1, \dots, 2n. \quad (11)$$

But this is in fact iterative proportional fitting, a procedure that iteratively rescales the parameters until the estimates match the marginal constraints. Convergence is achieved when $\lambda_r^{k-1} = \lambda_r^k$, satisfying the score equations (10). More generally, the log-likelihood (9) is monotonically non-decreasing in each update step (11) and convergence of (11) is achieved only if the log-likelihood is maximized (Koller et al., 2009, Theorem 20.5).

(iii) *IPFP and constrained maximum likelihood*: If we re-sort the sufficient statistics and re-label the elements of $\boldsymbol{\lambda}$ we get

$$\begin{aligned} \sum_{r=1}^{2n} \lambda_r \mathbf{A}_r \mathbf{x} &= \lambda_1(x_{12} + x_{13} + \dots + x_{1n}) + \dots + \lambda_{2n}(x_{1n} + x_{2n} + \dots + x_{(n-1)n}) \\ &= \sum_{q=(q_1, q_2) \in \mathcal{I}} (\lambda_{q_1} + \lambda_{n+q_2}) x_q = \sum_{q \in \mathcal{I}} \frac{x_q}{\mu_q} \end{aligned}$$

with $\mu_q = (\lambda_{q_1} + \lambda_{n+q_2})^{-1}$ for $q \in \mathcal{I}$. This leads to

$$c(\boldsymbol{\lambda}) = \int_{\mathcal{X}} \exp \left\{ - \sum_{q \in \mathcal{I}} \frac{x_q}{\mu_q} \right\} d\mathbf{x} = \prod_{q \in \mathcal{I}} \mu_q, \quad (12)$$

and where with (10) $\mathbf{A}\boldsymbol{\mu} = \mathbf{y}$. Hence, we can represent the whole system as the product of densities from exponentially distributed random variables X_q for $q \in \mathcal{I}$. That is

$$\hat{f}(\mathbf{x}) = \exp \left\{ - \sum_{q \in \mathcal{I}} \frac{x_q}{\mu_q} - \sum_{q \in \mathcal{I}} \log(\mu_q) \right\} = \prod_{q \in \mathcal{I}} \frac{1}{\mu_q} \exp \left\{ - \frac{x_q}{\mu_q} \right\} \quad (13)$$

with observed margins $\mathbf{A}\mathbf{x} = \mathbf{A}\boldsymbol{\mu} = \mathbf{y}$ and $x_q \geq 0 \forall q \in \mathcal{I}$.

3. Maximum likelihood-based estimation strategy

3.1. Parametrization and estimation

From result (13) it follows that we can use a distributional framework in order to build a generalized regression model. We exemplify this with a model which includes a sender-effect denoted as $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ and a receiver-effect $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$. We stack the coefficients in a $2n$ parameter vector $\boldsymbol{\theta} = (\boldsymbol{\delta}^T, \boldsymbol{\gamma}^T)^T$ such that the following log-linear expectation results

$$\mathbb{E}_{\boldsymbol{\theta}}[X_{ij}] = \mu_{ij}(\boldsymbol{\theta}) = \exp(\delta_i + \gamma_j). \quad (14)$$

Based on this structural assumption, we can now maximize the likelihood derived from (13) with respect to $\boldsymbol{\theta}$ subject to the observed values $\mathbf{A}\mathbf{x} = \mathbf{y}$ and the *moment condition* $\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{y}$. In the given formulation, the moment condition is linear in $\boldsymbol{\mu}(\boldsymbol{\theta})$ but not in $\boldsymbol{\theta}$. Consequently, the numerical solution to the problem might be burdensome. We therefore propose to use an iterative procedure that is somewhat similar to the Expectation Conditional Maximization (ECM, Meng and Rubin, 1993) algorithm, since it involves iteratively forming the expectation of X_{ij} based on the previous parameter estimate (E-Step) and constrained maximization afterwards (M-Step). To be specific, define the $(N \times 2n)$ design matrix \mathbf{Z} , that contains indices for sender- and receiver-effects. Matrix \mathbf{Z} has rows z_q indexed by $q \in \mathcal{I}$ where for $q = (i, j)$ we have the i -th and the $(n + j)$ -th element of z_q equal to 1 and all other elements are equal to zero. Starting with an initial estimate $\boldsymbol{\theta}_0$ that satisfies $\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}_0) = \mathbf{y}$, we form the expectation of the log-likelihood

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0} \left[\sum_{q \in \mathcal{I}} \left(-z_q^T \boldsymbol{\theta} - \frac{X_q}{\mu_q(\boldsymbol{\theta})} \right) \right] = \sum_{q \in \mathcal{I}} \left(-z_q^T \boldsymbol{\theta} - \exp\{z_q^T(\boldsymbol{\theta}_0 - \boldsymbol{\theta})\} \right).$$

Then, the maximization problem in the M-step is given by

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^{2n}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}_0) \text{ subject to } \mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{y}. \quad (15)$$

A suitable optimizer for non-linear constraints is available by the augmented Lagrangian (Hestenes, 1969, Powell, 1969)

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{\xi}_k, \zeta, \boldsymbol{\theta}_k) = -Q(\boldsymbol{\theta}; \boldsymbol{\theta}_k) - \boldsymbol{\xi}_k^T (\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) - \mathbf{y}) + \frac{\zeta}{2} \|\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) - \mathbf{y}\|_2^2, \quad (16)$$

with $\boldsymbol{\xi}_k$ and ζ being auxiliary parameters. The augmented Lagrangian method decomposes the constrained problem (15) into iteratively solving unconstrained problems. In each iteration we start with an initial parameter $\boldsymbol{\xi}_k$ in order to find the preliminary solution $\boldsymbol{\theta}_{k+1}$. Then, we update $\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \zeta(\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}_{k+1}) - \mathbf{y})$ in order to increase the accuracy of the estimate. In case of slow convergence, also ζ can be increased. An implementation in R is given by the package `nloptr` by Johnson (2014).

3.2. Confidence and prediction intervals

Considering the data entries as exponentially distributed allows for a quantification of the uncertainty of the estimates. We pursue this by bootstrapping (Efron and Tibshirani, 1994) here. Given a converged estimator $\hat{\boldsymbol{\theta}}$ we can draw for each matrix entry X_{ij} from an exponential distribution with expectation $\mu_{ij}(\hat{\boldsymbol{\theta}})$ in order to obtain B bootstrap samples $\mathbf{X}^* = (\mathbf{X}_{(1)}^*, \dots, \mathbf{X}_{(B)}^*)$. For each bootstrap sample $\mathbf{X}_{(b)}^*$ we calculate the marginals $\mathbf{A}\mathbf{X}_{(b)}^* = \mathbf{Y}_{(b)}^*$ and re-run the constrained estimation procedure resulting in B vectors of estimated means $\hat{\boldsymbol{\mu}}^* = (\hat{\boldsymbol{\mu}}_{(1)}^*, \dots, \hat{\boldsymbol{\mu}}_{(B)}^*)$. Consequently, the moment condition $\mathbf{A}\hat{\boldsymbol{\mu}}_{(b)}^* = \mathbf{Y}_{(b)}^*$ holds for all mean estimates of the bootstrap and by model-construction, the expected marginal restrictions from the bootstrap sample match the observed ones:

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}}[\mathbf{A}\mathbf{X}_{(b)}^*] = \mathbf{A}\mathbb{E}_{\hat{\boldsymbol{\theta}}}[\mathbf{X}_{(b)}^*] = \mathbf{y}.$$

Based on the bootstrap estimates, we can easily derive confidence intervals for μ_{ij} using the variability of $\hat{\mu}_{(b),ij}^*$ for $b = 1, \dots, B$. Additionally, we define the prediction error as $e_{ij} = x_{ij} - \hat{\mu}_{ij}$ and construct prediction intervals for the unknown x_{ij} based on the quantiles of the empirical distribution of

$$\hat{\mu}_{ij} + e_{(b),ij}^* = \hat{\mu}_{ij} + x_{(b),ij}^* - \hat{\mu}_{(b),ij}^*, \text{ for } b = 1, \dots, B.$$

3.3. Extensions with exogenous information and random effects

The regression framework allows to extend the model by including exogenous information. We consider again model (13) and parametrize the expectation through

$$\mathbb{E}_{\boldsymbol{\theta}}[X_{ij}] = \mu_{ij}(\boldsymbol{\theta}) = \exp(\delta_i + \gamma_j + \tilde{\mathbf{z}}_{ij}^T \boldsymbol{\beta}) = \exp(\mathbf{z}_{ij}^T \boldsymbol{\theta}), \quad (17)$$

with δ_i and γ_j again being the subject-specific sender- and receiver-effects. Furthermore, $\tilde{\mathbf{z}}_{ij}$ represents a l -dimensional covariate vector and $\boldsymbol{\beta}$ is the corresponding parameter vector. We can use the augmented Lagrangian approach from above to estimate the $p = l + 2n$ dimensional parameter vector $\boldsymbol{\theta}$. It is important to note here that only dyadic covariates have the potential to increase the predictive performance of the approach. If we only include subject-specific (monadic) information the expectation can be multiplicatively decomposed and the model collapses back to the IPFP model (14). This is easily seen through

$$\mu_{ij}(\boldsymbol{\theta}) = \exp(\delta_i + \tilde{\mathbf{z}}_i^T \boldsymbol{\beta}_i + \gamma_j + \tilde{\mathbf{z}}_j^T \boldsymbol{\beta}_j) = \exp(\tilde{\delta}_i + \tilde{\gamma}_j).$$

Nevertheless, the inclusion of subject-specific information may be valuable if it is the goal to forecast future networks based on new covariate information. This holds in particular in dynamic networks. We give an example for predictions based on lagged covariates in the next section.

We can also easily add additional structure to model (17) and assume a distributional form for some or all coefficients. A simple extension arises if we assume random effects. This occurs by the inclusion of normally distributed sender- and receiver-effects:

$$(\boldsymbol{\delta}, \boldsymbol{\gamma})^T \sim \mathcal{N}_{2n}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\vartheta})), \quad (18)$$

where we take $\boldsymbol{\vartheta}$ as the vector of parameters that determines the covariance matrix of the random effects. The latter could be parametrized for example with $\boldsymbol{\vartheta} = (\sigma_{\delta}^2, \sigma_{\delta, \gamma}^2, \sigma_{\gamma}^2)^T$ such that

$$\begin{pmatrix} \delta_i \\ \gamma_j \end{pmatrix} \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{pmatrix} \sigma_{\delta}^2 & \sigma_{\delta, \gamma}^2 \\ \sigma_{\delta, \gamma}^2 & \sigma_{\gamma}^2 \end{pmatrix} \right), \text{ for } i, j = 1, \dots, n \text{ and } i \neq j, \quad (19)$$

where we assume separate variance components for the sender- and the receiver effects, respectively. In order to fit the model, we follow a Laplace approximation estimation strategy similar to Breslow and Clayton (1993). Details are given in the Appendix A.

AT	Austria	ES	Spain	JP	Japan
AU	Australia	FI	Finland	KR	South Korea
BE	Belgium	FR	France	NL	Netherlands
CA	Canada	GB	United Kingdom	SE	Sweden
CH	Switzerland	GR	Greece	TR	Turkey
CL	Chile	IE	Ireland	TW	Taiwan
DE	Germany	IT	Italy	US	United States of America

Table 1: Countries included in the analysis

4. Application

4.1. Data description

The dataset under study is provided by the Bank for International Settlements (BIS) and freely available from their homepage. In general, the *locational banking statistics* (LBS) provide information about international banking activity by aggregating the financial activities (in million USD) to the country level. Within each country, the LBS accounts for outstanding claims (conceptualized as a valued dyad x_{ij} that consists of all claims banks from country i to banks of country j) and liabilities of internationally active banks located in reporting countries (conceptualized as the reverse direction x_{ji}). We have selected the 21 most important countries (see Table 1) for the time period from January 2005 to December 2017 as a quarterly series for the subsequent analysis. In Figure 1 the density of the network (number of existing edges relative to the number of possible edges) is shown on the left, the share of the zero-valued marginals in the middle and the development of the aggregated exposures on the right. Especially in the first years some marginals of the financial networks are zero and the corresponding matrix entries are therefore not included in the estimation problem. Correspondingly, it can be seen that most countries do have some claims and liabilities to other countries but especially in the beginning, many dyads x_{ij} are zero valued.

Since it is plausible that financial interactions are related to the economic size of a country, we consider the annual *Gross Domestic Product* (GDP, in current USD Billions) as covariate. The data is provided for the years 2005-2017 by the International Monetary Fund on their homepage. Furthermore, there might be relationship between trade in commercial goods and financial transfers and we use data on *dyadic trade flows* (in current USD) between states as additional covariate. The data is available annually for the years 2005 to 2014 by the Correlates of War Project online. We do not have available information on trade for the years 2015, 2016 and 2017 and we therefore extrapolate the previous values using an autoregressive regression model. Apparently, by doing so we have covariate information which is subject to uncertainty. We ignore this issue subsequently. In order to have an

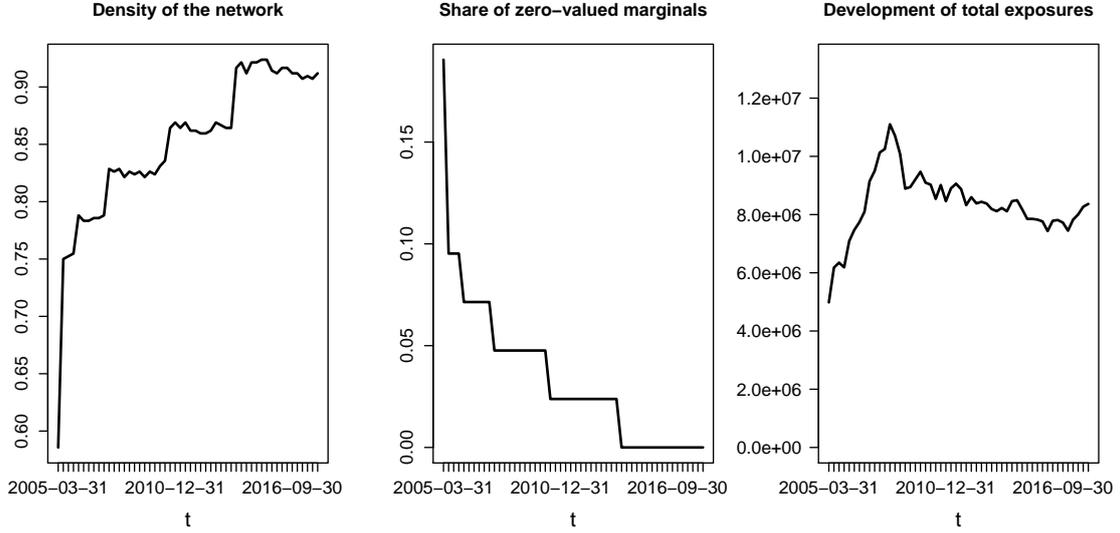


Figure 1: Density (left), share of zero-valued marginals (middle) and aggregated volume in million USD (right) of the network as quarterly time series.

Variable	Description	Type	Correlation
x_{ij}	Claim from country i to country j	dyad specific	1.0000
gdp_i	Gross Domestic Product of country i	node specific	0.4716
gdp_j	Gross Domestic Product of country j	node specific	0.1858
$trade_{ij}$	Bilateral trade flows of commercial goods	dyad specific	0.4349
$dist_{ij}$	Distance between the capital cities of countries i and j	dyad specific	-0.0953
x_{ij}^{t-1}	Lagged claim from country i to country j	dyad specific	0.9935

Table 2: Covariates used for the regression-based network-reconstruction

example for uninformative dyadic information, we use time-invariant data on the *dyadic distance* in kilometres between the capital cities of the countries under study (provided by Gleditsch, 2013). Finally, in some matrix reconstruction problems, the matrix entries of previous time points become known after some time. Typically, lagged values are strongly correlated with the actual ones. We therefore also consider the matrix entries, lagged by one quarter as covariates. See Table 2 for an overview of the variables, together with the overall correlation of the actual claims and the respective covariate. In the subsequent analysis we include all covariates in logarithmic form.

Method	Covariates	Rand. eff.	Model	overall L_1	overall L_2	average L_1	SE	average L_2	SE	
1	IPFP	-	(11,14)	4204.212	75.778	80.850	12.564	10.445	1.168	
2	Regression	-	(14,19)	4204.212	75.778	80.850	12.564	10.445	1.168	
3	Regression	$gdp_i, gdp_j, trade_{ij}$	(17)	3300.242	56.794	63.466	9.246	7.802	1.085	
4	Regression	$gdp_i, gdp_j, trade_{ij}$	$\sigma_\delta^2, \sigma_\gamma^2, \sigma_{\delta,\gamma}^2$	(17, 19)	3315.673	57.104	63.763	9.222	7.850	1.052
5	Regression	$gdp_i, gdp_j, dist_{ij}$	-	(17)	4884.728	91.575	93.937	15.864	12.565	1.857
6	Regression	$gdp_i, gdp_j, dist_{ij}$	$\sigma_\delta^2, \sigma_\gamma^2, \sigma_{\delta,\gamma}^2$	(17, 19)	4843.641	90.623	93.147	15.490	12.435	1.833
7	Regression	$gdp_i, gdp_j, x_{ij}^{t-1}$	-	(17)	2280.235	41.805	43.851	12.833	5.591	1.549
8	Regression	$gdp_i, gdp_j, x_{ij}^{t-1}$	$\sigma_\delta^2, \sigma_\gamma^2, \sigma_{\delta,\gamma}^2$	(17, 19)	2341.796	43.483	45.035	11.715	5.787	1.710

Table 3: Comparison of different regression models with the BIS Dataset. All values scaled by 100 000 and lowest values in bold.

4.2. Model performance

We evaluate the proposed models in terms of their L_1 and L_2 errors. The corresponding results are provided in Table 3. As a baseline specification, all models contain sender- and receiver effects. In the first row, we provide the maximum entropy model (14) that coincides with the IPFP solution (11). The second row shows model (14) together with the random effects structure (19). In the third row, we provide the errors for model (17) where we included the covariates logarithmic GDP (gdp_i, gdp_j) as well as the logarithmic trade data ($trade_{ij}$). In row four, we use the same model as in row three but additionally added the random effects structure from (19). In rows five and six, the same models as in rows three and four are used but with logarithmic distance ($dist_{ij}$) instead of trade as dyadic explanatory variable. In the last two rows we consider models with lagged claims (x_{ij}^{t-1}) with and without random effects. This comparison might be somewhat unfair because of the strong correlation and because it is not clear whether it can safely be assumed that such data is always available. Therefore, we have separated this specification from the other models.

In the first four columns the different specifications together with the related equations are provided. Columns five and six show the aggregated errors over all 52 quarters and the last four columns show the errors averaged over all years together with their corresponding standard errors. It can be seen that the first two models provide the same predictions and the inclusion of the random effects has no impact other than giving estimates for the variance of the sender- and receiver-effects as well as their correlation, shown in Figure 2. It becomes visible that the variation of the receiver-effect is much higher than the variation of the sender-effect which is almost constant. The correlation between the sender and the receiver effect is consistently positive and increases strongly within the first years.

Furthermore, Table 3 shows that in the four models that include exogenous information (rows three to six) the extension towards the random effects structure has an impact on the predictive quality. It decreases in the model that includes the variable $trade_{ij}$ and increases in the one that includes $dist_{ij}$. Nevertheless, the models with and without random effects are rather close to each other and in fact they are statistically indistinguishable with respect

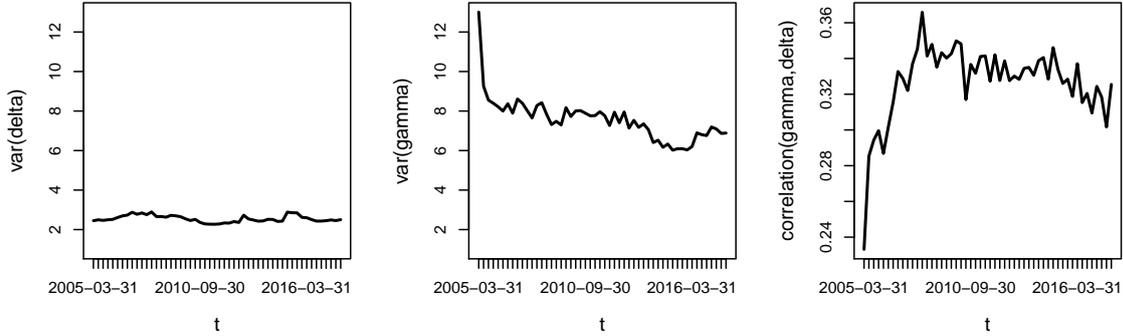


Figure 2: Quarterly time series of the estimated variances (models (14) and (19)) of the sender-effect ($\hat{\sigma}_\delta^2$) on the left, the receiver effect ($\hat{\sigma}_\gamma^2$) in the middle and the correlation between the sender- and the receiver effect ($\hat{\sigma}_{\delta,\gamma}^2/(\hat{\sigma}_\gamma\hat{\sigma}_\delta)$) on the right.

to their L_2 differences. While the model with the covariate $dist_{ij}$ performs even worse than the IPFP solution, the model that includes $trade_{ij}$ but includes no random effects (row three) gives superior predictions relative to all other models in the upper part of Table 3. However, the two models that use the information on the lagged values give by far the best predictions. We nevertheless continue with the best model from the upper part of Table 3 lagged data are not necessarily available. The corresponding fitted values are provided as time series in Figure 4 and in Figure 3 we provide the estimates for the coefficients of the model. In the first row, the estimated sender- ($\hat{\delta}$, left) and receiver-effects ($\hat{\gamma}$, right) are shown as a time series. In the second row of Figure 3 the estimates for the coefficients on the exogenous covariates ($\hat{\beta}$) can be seen. The estimated coefficients provide the intuitive result that the claims from country i to country j increase with gdp_i and gdp_j and the trade volume between them ($trade_{ij}$). It is reassuring that the ordering of the average height of the coefficients approximately matches with the order of the correlations reported in Table 2. Note however, that the size of the coefficients is to be interpreted with care because of the limited information available on the unknown claims. We also provide prediction intervals in Figure 5, based on the share of real values x_{ij} located in the interval $[q_{0.005}, q_{0.955}]$. Here, $q_{0.005}$ and $q_{0.955}$ are the 0.005 and 0.955-quantiles derived from the bootstrap distribution (bootstrap sample size $B = 100$). On the left, we illustrate the real values against the predicted ones together with grey 95% prediction intervals for the most recent network. Observations that do not fall within the prediction interval are indicated by red circles. Because of the quadratic mean-variance relation of the exponential distribution it is much easier to capture high values within the prediction intervals than low ones. A circumstance that materializes in the fact that exclusively small values are outside the prediction intervals.

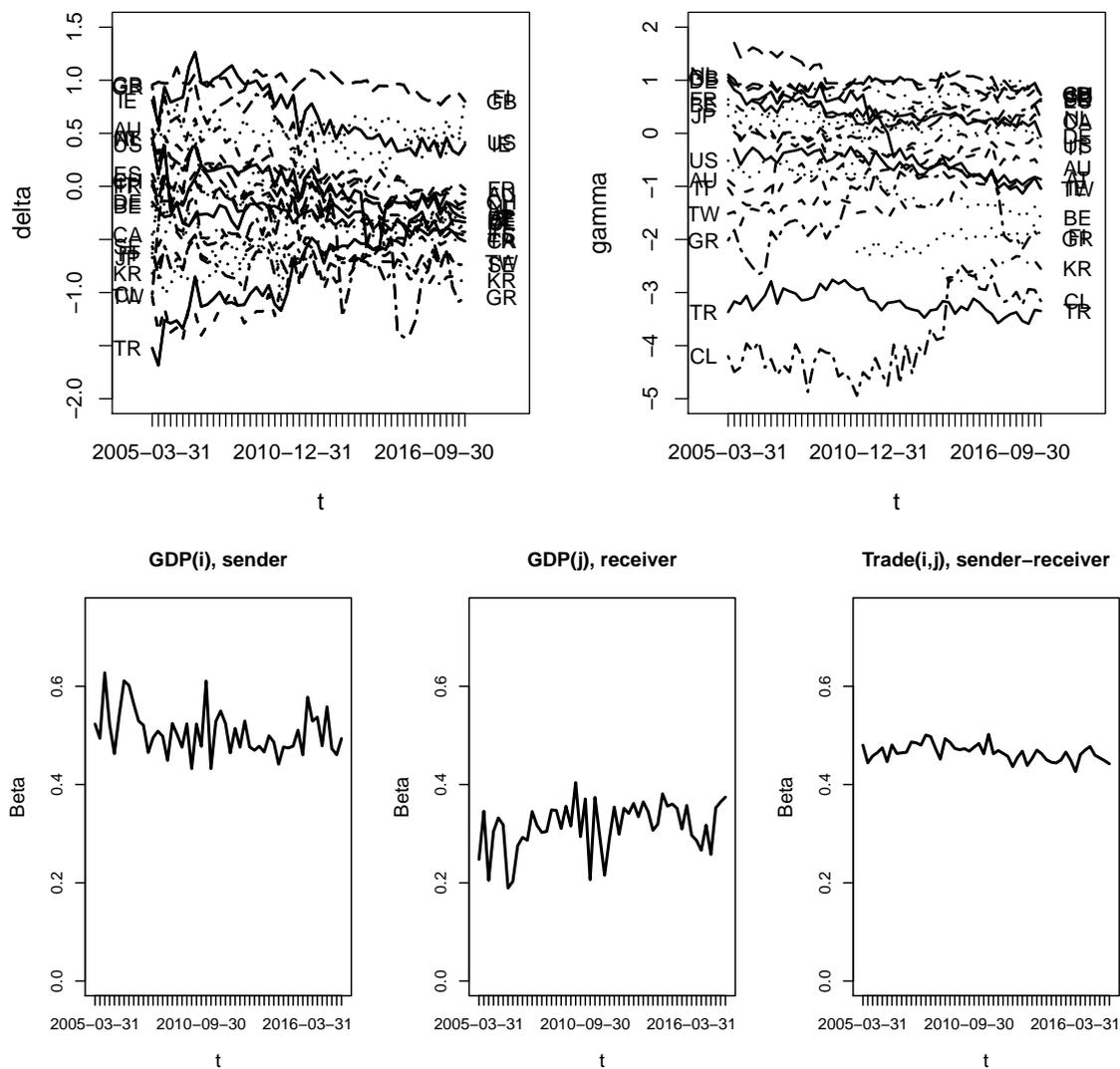


Figure 3: Estimated coefficients of model (17) with gdp_i , gdp_j and $trade_{ij}$ as covariates. Time series of sender- ($\hat{\delta}$, left) and receiver-effects ($\hat{\gamma}$, right) in the first row. Time series of estimated coefficients on exogenous covariates ($\hat{\beta}$) in the second row.

The share of real values within the prediction intervals against time is shown on the right hand side of Figure 5. We cover on average 96% of all true values with our prediction intervals over all time periods and regard the bootstrap approach therefore as satisfying.

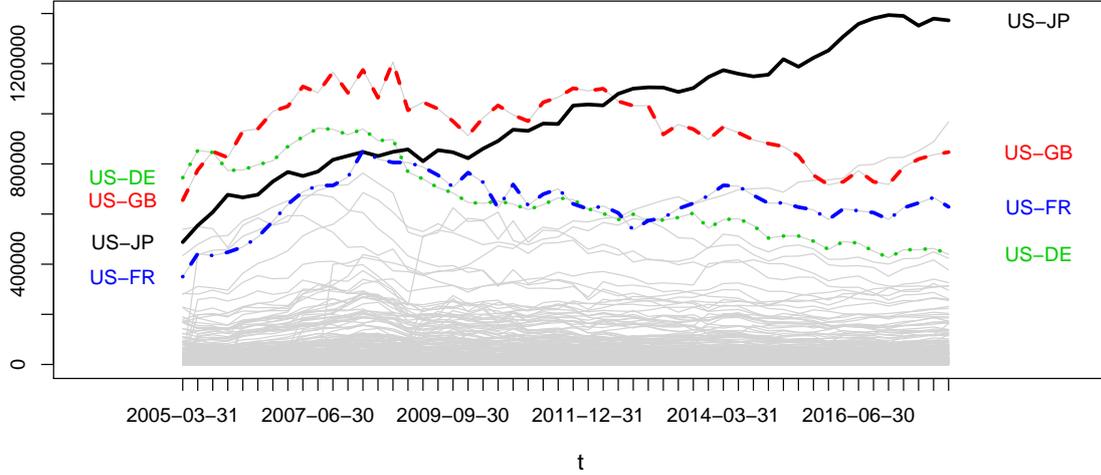


Figure 4: Fitted exposures, shown as quarterly series (52 time points) in million USD. All estimated with model (17) with gdp_i , gdp_j and $trade_{ij}$ as covariates. Four dyads with the overall highest values are highlighted.

4.3. Comparison to alternative routines

Gravity model: A standard solution to the problem is the gravity model (e.g. Sheldon and Maurer, 1998). In essence it represents a multiplicative independence model

$$\hat{\mu}_{ij} = \frac{x_{i\bullet}x_{\bullet j}}{x_{\bullet\bullet}}. \quad (20)$$

The model is simple, easy to implement and very intuitive. In situations where diagonal elements x_{ii} are not restricted to be zero it even coincides with the maximum entropy solution.

Tomogravity model: An extension of the gravity model is given by the tomogravity approach by Zhang et al. (2003). The model was initially designed to estimate point-to-point traffic volumes from dyadic loads and builds on minimizing the loss-function

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \left\{ (\mathbf{A}\boldsymbol{\mu} - \mathbf{y})^T (\mathbf{A}\boldsymbol{\mu} - \mathbf{y}) + \psi^2 \sum_{i \neq j} \frac{\mu_{ij}}{N} \log \left(\frac{\mu_{ij}}{x_{i\bullet}x_{\bullet j}} \right) \right\} \quad (21)$$

subject to the non-negativity constraint. Here, the gravity model (20) serves as a null model in the penalization term and the strength of penalization is given by ψ . The approach is implemented in the R package `tomogravity` (see Blocker et al., 2014). Zhang et al. (2003) show in a simulation study, that $\psi = 0.01$ is a reasonable choice if no training set is available. In our competitive comparison we optimize the tuning parameter in order to minimize the

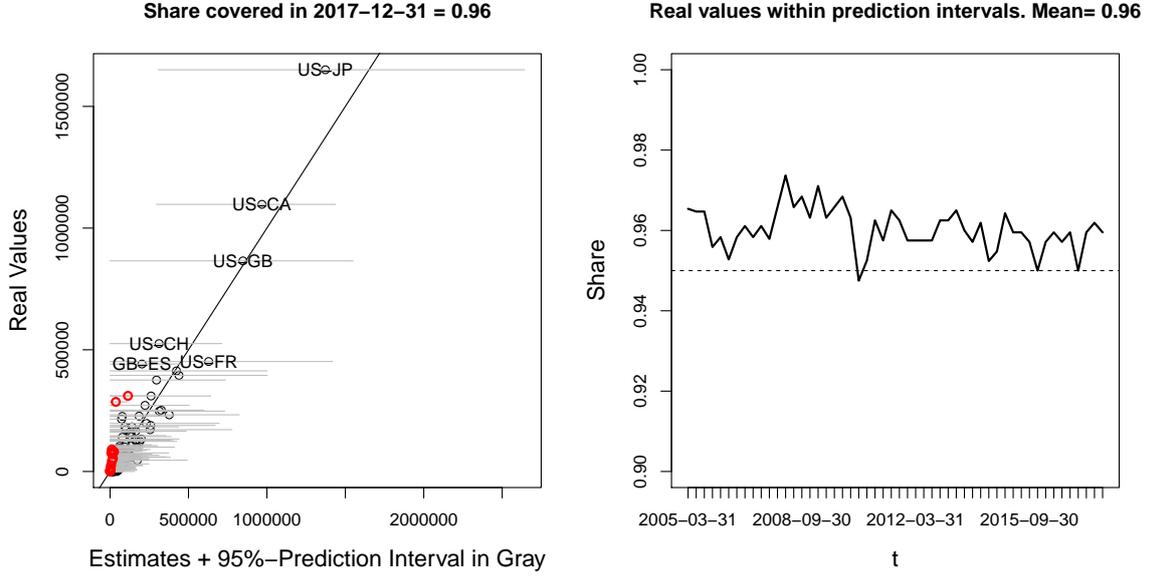


Figure 5: 95% Prediction intervals for the estimated means of model (17) with gdp_i , gdp_j and $trade_{ij}$ as covariates, for the most recent network on the left. Prediction intervals in grey. Real values on the horizontal axis and estimated means on the vertical axis. Bisecting line in solid black. Predictions (not) covered by prediction intervals in black (red) circles. Share of real values within prediction intervals as a time series on the right.

overall L_2 error with grid search and find $\psi = 0.011$ to be an optimal value.

Non-negative LASSO: The LASSO (Tibshirani, 1996) was already applied to predict flows in bike sharing networks by Chen et al. (2017) and works best with sparse networks. Using this approach, we minimize the loss function

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \left\{ (\mathbf{A}\boldsymbol{\mu} - \mathbf{y})^T (\mathbf{A}\boldsymbol{\mu} - \mathbf{y}) + \tau \sum_{i \neq j} |\mu_{ij}| \right\} \quad (22)$$

where we can drop the absolute value in the penalization term because of the non-negativity constraint (see Wu et al., 2014 for the non-negative LASSO). In order to use the approach in the competitive comparison, we optimize the penalty parameter τ on a grid for the minimum L_2 error and use $\tau = 45,483.6$. The models are estimated using the R package `glmnet` by Friedman et al. (2009).

Ecological regression: In Ecological Inference (see e.g. Klima et al., 2016, King, 2013), it is often assumed that the observations at hand are independent realizations of a linear model,

parametrized by time-constant transition-shares β_{ij} . Define the stacked column sums in t by \mathbf{y}_c^t and the stacked row sums in t by \mathbf{y}_r^t . Then, the model can be represented as

$$\mathbb{E}[\mathbf{Y}_r^t | \mathbf{Y}_c^t = \mathbf{y}_c^t] = \mathbf{B}_c \mathbf{y}_c^t, \text{ for } t = 1, \dots, T \quad (23)$$

where the $(n \times N)$ matrix \mathbf{B}_c contains the parameters β_{ij} . In the give form, the problem is not identified for one time period t and it must be assumed that multiple time-points can be interpreted as independent realizations. Additionally, the model is not symmetric, implying that the solution to equation (23) does not coincide with the solution to

$$\mathbb{E}[\mathbf{Y}_c^t | \mathbf{Y}_r^t = \mathbf{y}_r^t] = \mathbf{B}_r \mathbf{y}_r^t, \text{ for } t = 1, \dots, T. \quad (24)$$

Since, the estimated transition shares are not guaranteed to be non-negative and sum up to one they must be post-processed to fulfil this conditions. Both models are fitted via least-squares in R.

Hierarchical Bayesian models: Gandy and Veraart (2017) propose to use simulation-based methods. In their hierarchical models the first step consists of estimating the link probabilities and given that there is a link, the weight is sampled from an exponential distribution:

$$\begin{aligned} P(X_{ij} > 0) &= p \\ X_{ij} | X_{ij} > 0 &\sim \text{Exp}(\mu). \end{aligned} \quad (25)$$

In order to estimate the link probabilities p , knowledge of the density or a desired target density is needed. In their basic model it is proposed to use an Erdős-Rényi model with p consistent with the target density. In an extension of the model, inspired by Graphon models, Gandy and Veraart (2018) propose a so called empirical fitness model. Here the link probability is determined by the logistic function

$$P(X_{ij} > 0) = p_{ij} = \frac{1}{1 + \exp(-\alpha - z_i - z_j)}, \quad (26)$$

with α being some constant that is estimated for consistency with the target density. For the fitness variables z_i , the authors propose to use an empirical Bayes approach, incorporating the information of the row and column sums as $z_i = \log(x_{\bullet i} + x_{i \bullet})$. An implementation of both models is given by the R package `systemicrisk`. In order to make the approach as competitive as possible we use for each quarter the real (but in principle unknown) density of the networks. Because the results of the method differ between each individual estimate, we average the estimates and evaluate the combined dataset.

Method	Model	overall L_1	overall L_2	average L_1	SE	average L_2	SE
1	Regression ($gdp_i, gdp_j, trade_{ij}$)	(17) 3 300.242	56.794	63.466	9.246	7.802	1.085
2	Gravity model	(20) 4 300.927	79.342	82.710	13.003	10.935	1.232
3	Tomogravity Model	(21) 4 241.299	75.760	81.563	12.774	10.442	1.168
4	Non-negative LASSO	(22) 7 233.821	127.638	139.112	20.399	17.572	2.150
5	Ecological Regression, columns	(23) 9 785.014	163.422	188.173	24.575	22.573	2.032
6	Ecological Regression, rows	(24) 10 776.570	184.636	207.242	27.662	25.438	2.946
7	Hierarchical, Erdős-Rényi	(25) 5 328.834	101.639	102.478	17.439	14.004	1.610
8	Hierarchical, Fitness	(26) 5 316.036	102.072	102.231	17.912	14.039	1.827

Table 4: Comparison of different methods with the BIS Dataset. All values scaled by 100 000 and lowest values in bold.

4.4. Competitive comparison

Again we compare the different algorithms in terms of their L_1 and L_2 errors in Table 4. In the first row of Table 4 we show the restricted maximum likelihood model with the best predictions from Table 3 and in the following rows, the models introduced in Section 4.3 above are shown. The results can be separated roughly into three blocks. The models that fundamentally build on some kind of Least Squares criterion without referring in some way to the gravity model or the maximum entropy solution (ecological regression, and non-negative LASSO in rows four, five and six) have the highest values in terms of their L_1 and L_2 errors. Somewhat better are the Hierarchical Bayesian Models (rows seven and eight) that can be considered as the second block. However, although they provide better predictions than the models in the first block, we used the real density of the network in order to calibrate them which gives them an unrealistic advantage. The third group is given by the gravity and tomogravity model (rows two and three). Those are statistically indistinguishable and provide considerably better results than the models from the former blocks. Nevertheless, the regression model that uses exogenous information on $trade_{ij}$ (first row) yields the best predictions in this comparison.

4.5. Performance of the estimator

We hope to see improvements in the predictions if we include informative exogenous variables in the model. Informative means in this context, that variation in \tilde{z}_{ij} is able to explain variation in the unknown X_{ij} . Apparently, including information with a low association to X_{ij} simply adds noise into the estimation procedure. In this case we expect inferior predictions as compared to the IPFP solution. We illustrate the properties of the estimator

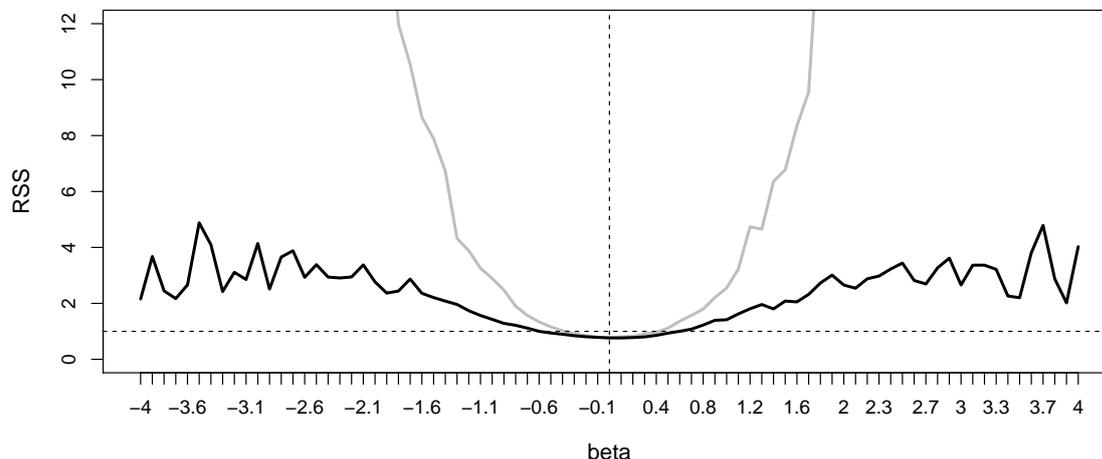


Figure 6: Median (solid black) and mean (solid grey) of the relative squared error $RSS_s(\beta)$ for different levels of β .

using a simulation study with the following data generating process

$$\begin{aligned} \delta_i &\sim N(0, 1), \gamma_j \sim N(0, 1), \tilde{\mathbf{z}}_{ij} \sim N(0, 1), \text{ for } i, j = 1, \dots, 10 \text{ and } i \neq j \\ \mu_{ij}(\beta) &= \exp(\delta_i + \gamma_j + \tilde{\mathbf{z}}_{ij}\beta) \\ X_{ij} &\sim \text{Exp}(\mu_{ij}(\beta)). \end{aligned} \quad (27)$$

Since the association between $\tilde{\mathbf{z}}_{ij}$ and the unknown X_{ij} is crucial, we vary the parameter β from -4 to 4 and denote with $\mu_{ij}(\beta)$ the mean based on β . For each parameter β we re-run the data generating process (27) $S = 1000$ times and calculate for the s -th simulation the IPFP solution $\check{\mu}_{s,ij}(\beta)$ and the restricted maximum likelihood solution $\hat{\mu}_{s,ij}(\beta)$. Based on that, we calculate in each simulation the ratio of the squared errors

$$RSS_s(\beta) = \frac{\sum_{i \neq j} (X_{s,ij} - \check{\mu}_{s,ij}(\beta))^2}{\sum_{i \neq j} (X_{s,ij} - \hat{\mu}_{s,ij}(\beta))^2}, \text{ for } s = 1, \dots, 1000.$$

This ratio is smaller than one if the IPFP estimates yield a lower mean squared error than the restricted maximum likelihood estimates and higher than one if the exogenous information improves the predictive quality in the terms of the mean squared error.

In Figure 6, we show the median (solid black) and the mean (solid grey) of $RSS_s(\beta)$ for different values of β as well as a horizontal line indicating the value one (dashed black) and

a vertical line for $\beta = 0$ (dashed black). It can be seen, that the mean and the median of $RSS_s(\beta)$ are below one for values of β that are roughly between -0.5 and 0.5 but increase strongly with higher absolute values of β . Apparently, the distribution of $RSS_s(\beta)$ is skewed with a long tail since the mean is much higher than the median. With very high or low values of β , the median of the relative mean squared error becomes more volatile and partly decreases.

Furthermore, we investigate how well the estimated expectations match the true ones. By construction, it holds that $\mathbf{A}\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}] = \mathbf{A}\mathbb{E}[\hat{\boldsymbol{\mu}}]$ and consequently, the regression-based approach as well as IPFP assume that the sum of realized values equals the sum of the true expectations. Nevertheless, the moment condition does not imply that $\mathbb{E}[\hat{\mu}_{s,ij}] = \mu_{ij}$. In order to investigate potential bias, we draw again from

$$\delta_i \sim N(0, 1), \gamma_j \sim N(0, 1), \tilde{\mathbf{z}}_{ij} \sim N(0, 1), \text{ for } i, j = 1, \dots, 10 \text{ and } i \neq j,$$

and fix $\mu_{ij}(\beta) = \exp(\delta_i + \gamma_j + \tilde{\mathbf{z}}_{ij}\beta)$ to be the true expectation and draw and re-estimate again $S = 1\,000$ times from $X_{ij} \sim \text{Exp}(\mu_{ij}(\beta))$. Apparently, estimating the true expectations is a hard task as only sums of random variables with different expectations are available. Consequently, the variation of the mean estimates is rather high and we report boxplots of the normalized difference between the true value and mean estimate

$$\Delta_{s,ij}(\beta) = \frac{\hat{\mu}(\beta)_{s,ij} - \mu_{ij}(\beta)}{S^{-1} \sum_{s=1}^S (\hat{\mu}(\beta)_{s,ij} - S^{-1} \sum_{s=1}^S \hat{\mu}(\beta)_{s,ij})^2},$$

and accordingly for $\check{\mu}(\beta)_{s,ij}$. In Figure 7 we illustrate three different cases with $\beta = 0$ (top), $\beta = 1$ (middle) and $\beta = -1$ (bottom). On the left hand side, boxplots for $\Delta(\beta)_{s,ij}$ are shown for the regression-based model and on the right hand side for IPFP. The solid black line represents zero and the dashed black lines give ± 1.96 . The results for the case $\beta = 0$ on the top, match with the previous analysis illustrated in Figure 6 and show that IPFP identifies the true expectations somewhat better than the regression-based approach when the exogenous information is non-informative. In such a case, including $\tilde{\mathbf{z}}_{ij}$ adds noise in the estimation procedure, resulting in a greater variance around the true expectations. However, this changes strongly if $\tilde{\mathbf{z}}_{ij}$ is informative. Especially for $\beta = -1$ on the bottom of Figure 7, some estimates obtained from IPFP are seriously biased because this procedure does not have the ability to account for the dyad-specific heterogeneity. The regression-based method, however does a reasonable job in recovering the unknown true expectations.

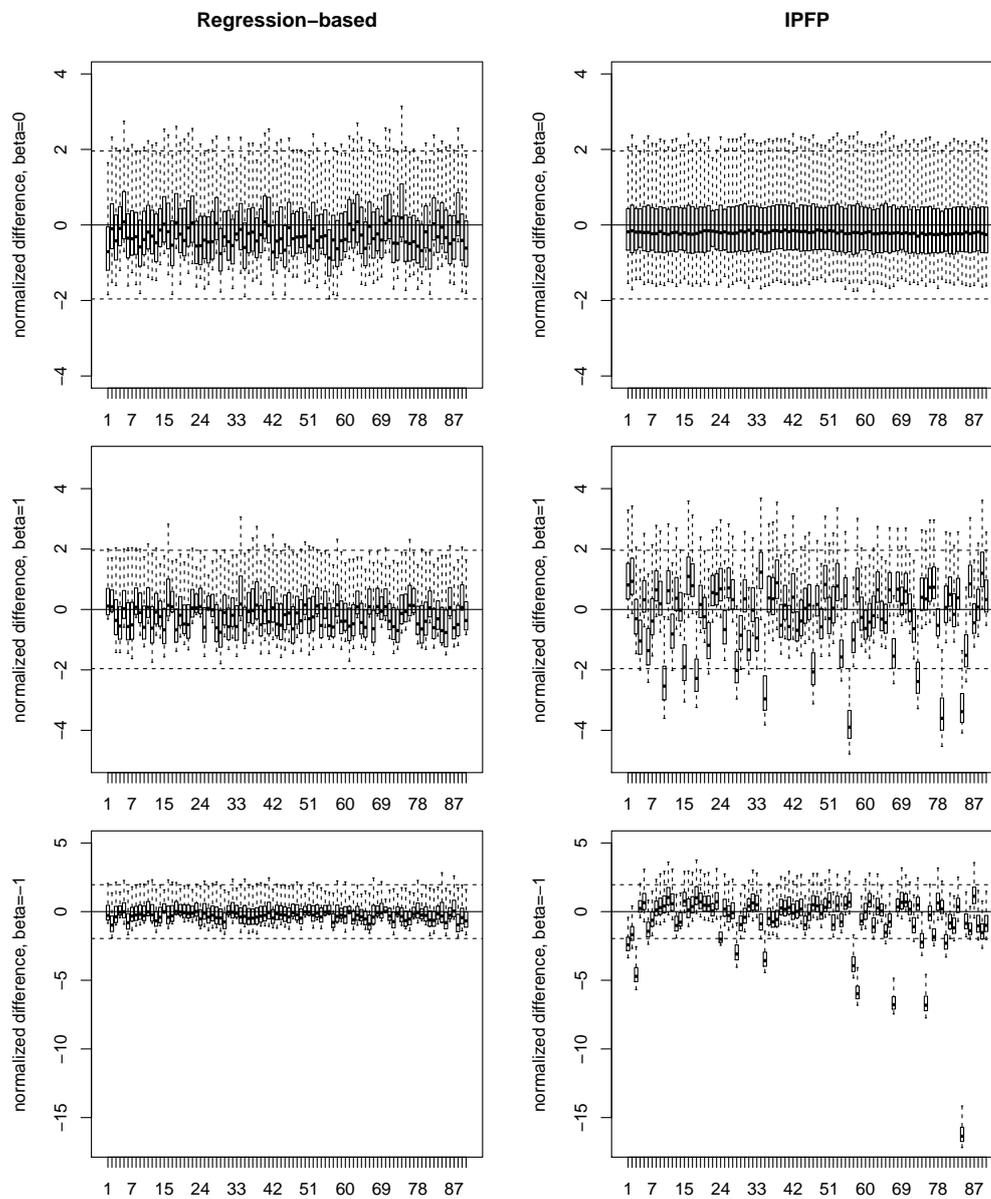


Figure 7: Boxplots for the normalized differences between the true expectation and the mean of the estimated ones $\Delta_{s,ij}(\beta)$ for $\beta = 0$ (top), $\beta = 1$ (middle) and $\beta = -1$ (bottom). Regression-based model on the left, IPFP on the right. Outliers are excluded from the boxplot representation for better clarity.

5. Discussion

In this paper we propose a method that allows for network reconstruction within a regression framework. This approach makes it easy to add and interpret exogenous information. It also allows to construct bootstrap prediction intervals to quantify the uncertainty of the estimates. Furthermore, the framework is flexible enough to deal with problems that involve partial information. For example if some elements of the network are known or if we have information on the binary network, then we can model the expected values of the matrix entries conditional this information, simply by changing the routing matrix and the E-Step.

However, we also want to list some shortfalls of the method. An obvious drawback of the method is its derivation from the maximum entropy principle that tries to allocate the matrix entries as even as possible and is therefore not very suitable for sparse networks as long as the sparseness cannot be inferred from the marginals. Furthermore, the estimated coefficients must be interpreted with care, as they are estimated based on a data situation with much less information than in usual regression settings. As a last but most important point, if the association between the exogenous explanatory variable(s) and the unknown matrix entries is low, the method is likely to deliver predictions that are worse than simple IPFP. It is therefore highly recommendable to use expert knowledge when selecting the exogenous dyadic covariates for regression-based network reconstruction.

- Dudík, M., S. J. Phillips, and R. E. Schapire (2007). Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *Journal of Machine Learning Research* 8(6), 1217–1260.
- Dym, C. L. and I. H. Shames (2013). Introduction to the calculus of variations. In *Solid Mechanics*, Chapter 2, pp. 71–116. New York: Springer.
- Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*. Boca Raton: CRC press.
- Elsinger, H., A. Lehar, and M. Summer (2013). Network models and systemic risk assessment. In J.-P. Fouque and J. A. Langsam (Eds.), *Handbook on Systemic Risk*, Chapter IV, pp. 287–305. Cambridge: Cambridge University Press.
- Fienberg, S. E. et al. (1970). An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics* 41(3), 907–917.
- Friedman, J., T. Hastie, and R. Tibshirani (2009). glmnet: Lasso and elastic-net regularized generalized linear models. R package version 2.0-1.6.
- Gandy, A. and L. A. Veraart (2017). A bayesian methodology for systemic risk assessment in financial networks. *Management Science* 63(12), 4428–4446.
- Gandy, A. and L. A. M. Veraart (2018). Adjustable network reconstruction with applications to CDS exposures. *Journal of Multivariate Analysis*, in press.
- Gleditsch, K. S. (2013). Distance between capital cities. <http://privatewww.essex.ac.uk/~ksg/data-5.html>. Accessed: 2017-04-07.
- Golan, A. and G. Judge (1996). Recovering information in the case of underdetermined problems and incomplete economic data. *Journal of statistical planning and inference* 49(1), 127–136.
- Haberman, S. J. (1978). *Analysis of qualitative data 1: Introductory topics*. New York: Academic Press.
- Haberman, S. J. (1979). *Analysis of qualitative data 2: New Developments*. New York: Academic Press.
- Hazelton, M. L. (2010). Statistical inference for transit system origin-destination matrices. *Technometrics* 52(2), 221–230.
- Hestenes, M. R. (1969). Multiplier and gradient methods. *Journal of optimization theory and applications* 4(5), 303–320.

- Johnson, S. G. (2014). The NLOpt nonlinear-optimization package. R package version 1.2.1.
- King, G. (2013). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton: Princeton University Press.
- Klima, A., P. W. Thurner, C. Molnar, T. Schlesinger, and H. Küchenhoff (2016). Estimation of voter transitions based on ecological inference: An empirical assessment of different approaches. *AStA Advances in Statistical Analysis* 100(2), 133–159.
- Koller, D., N. Friedman, and F. Bach (2009). *Probabilistic graphical models: principles and techniques*. Cambridge: MIT Press.
- Malvestuto, F. (1989). Computing the maximum-entropy extension of given discrete probability distributions. *Computational Statistics & Data Analysis* 8(3), 299–311.
- Mastrandrea, R., T. Squartini, G. Fagiolo, and D. Garlaschelli (2014). Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics* 16(4), 043022.
- Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 80(2), 267–278.
- Miller, R. E. and P. D. Blair (2009). *Input-output analysis: foundations and extensions*. Cambridge: Cambridge University Press.
- Muñoz-Cobo, J.-L., R. Mendizábal, A. Miquel, C. Berna, and A. Escrivá (2017). Use of the principles of maximum entropy and maximum relative entropy for the determination of uncertain parameter distributions in engineering applications. *Entropy* 19(9), 486.
- Onuki, Y. (2013). Extension of the iterative proportional fitting procedure and its evaluation using agent-based models. In M. Tadahiko, T. Takao, and T. Shingo (Eds.), *Agent-Based Approaches in Economic and Social Complex Systems VII*, pp. 215–226. Berlin, Heidelberg: Springer.
- Powell, M. J. (1969). A method for nonlinear constraints in minimization problems. In R. Fletcher (Ed.), *Optimization*, pp. 283–298. New York: Academic Press.
- Schneider, M. H. and S. A. Zenios (1990). A comparative study of algorithms for matrix balancing. *Operations Research* 38(3), 439–455.
- Sheldon, G. and M. Maurer (1998). Interbank lending and systemic risk: An empirical analysis for Switzerland. *Swiss Journal of Economics and Statistics* 134(4), 685–704.

- Squartini, T. and D. Garlaschelli (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* 13(8), 083001.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58(1), 267–288.
- Tutz, G. and A. Groll (2010). Generalized linear mixed models based on boosting. In *Statistical Modelling and Regression Structures*, pp. 197–215. Berlin, Heidelberg: Springer.
- Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability* 7(3), 111–125.
- Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association* 91(433), 365–377.
- Wu, L., Y. Yang, and H. Liu (2014). Nonnegative-lasso and application in index tracking. *Computational Statistics & Data Analysis* 70, 116–126.
- Zhang, Y., M. Roughan, C. Lund, and D. Donoho (2003). An information-theoretic approach to traffic matrix estimation. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 301–312. Association for Computing Machinery.
- Zhou, H., L. Tan, Q. Zeng, and C. Wu (2016). Traffic matrix estimation: A neural network approach with extended input and expectation maximization iteration. *Journal of Network and Computer Applications* 60, 220 – 232.

A. Estimation with random effects

In order to fit a model of the form

$$\begin{aligned} \mu_{ij}(\boldsymbol{\theta}) &= \exp(\delta_i + \gamma_j + \tilde{\mathbf{z}}_{ij}^T \boldsymbol{\beta}) = \exp(\mathbf{z}_{ij}^T \boldsymbol{\theta}), \\ \begin{pmatrix} \delta_i \\ \gamma_j \end{pmatrix} &\sim \mathcal{N}_2 \left(\mathbf{0}, \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta,\gamma}^2 \\ \sigma_{\delta,\gamma}^2 & \sigma_\gamma^2 \end{pmatrix} \right), \text{ for } i, j = 1, \dots, n \text{ and } i \neq j, \end{aligned}$$

we follow a Laplace approximation estimation strategy similar to Breslow and Clayton (1993) and fix $\boldsymbol{\vartheta}$ to some value $\boldsymbol{\vartheta}_0$, see also Tutz and Groll (2010). The moment condition implies a restriction for $\boldsymbol{\theta}$ but not for $\boldsymbol{\vartheta}$ and given some starting value $\boldsymbol{\theta}_0$, we can maximize the *penalized log-likelihood* (constant terms omitted)

$$\ell_{pen}(\boldsymbol{\theta}; \boldsymbol{\vartheta}_0, \boldsymbol{\theta}_0) = \sum_{q \in \mathcal{I}} \left(-\mathbf{z}_q^T \boldsymbol{\theta} - \exp\{\mathbf{z}_q^T (\boldsymbol{\theta}_0 - \boldsymbol{\theta})\} \right) - \frac{1}{2} (\boldsymbol{\delta}^T, \boldsymbol{\gamma}^T) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\vartheta}_0) (\boldsymbol{\delta}^T, \boldsymbol{\gamma}^T)^T$$

subject to the moment condition $\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{y}$. Therefore, the new optimization problem is given by

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{\xi}, \zeta, \boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) = -\ell_{pen}(\boldsymbol{\theta}; \boldsymbol{\vartheta}_0, \boldsymbol{\theta}_0) - \boldsymbol{\xi}^T (\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) - \mathbf{y}) + \frac{\zeta}{2} \|\mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta}) - \mathbf{y}\|_2^2. \quad (28)$$

Define $\tilde{\mathbf{Z}}$ as the $(N \times l)$ design matrix for the fixed effects and \mathbf{U} as the $(N \times 2n)$ random effects design matrix, this allows to write the generic mean as $\log(\boldsymbol{\mu}) = \mathbf{Z}\boldsymbol{\theta} = \tilde{\mathbf{Z}}\boldsymbol{\beta} + \mathbf{U}(\boldsymbol{\delta}^T, \boldsymbol{\gamma}^T)^T$. Given that we have some estimate of $\boldsymbol{\theta}$, call it $\boldsymbol{\theta}_1$ we can estimate the variance parameters $\boldsymbol{\vartheta}$ with an approximation of the *marginal restricted log-likelihood*:

$$\ell_R(\boldsymbol{\vartheta}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_0) = -\frac{1}{2} \log(|\mathbf{V}(\boldsymbol{\vartheta})|) - \frac{1}{2} \log(|\tilde{\mathbf{Z}}^T \mathbf{V}(\boldsymbol{\vartheta})^{-1} \tilde{\mathbf{Z}}|) - \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\mathbf{Z}}\boldsymbol{\beta})^T \mathbf{V}(\boldsymbol{\vartheta})^{-1} (\tilde{\mathbf{y}} - \tilde{\mathbf{Z}}\boldsymbol{\beta}) \quad (29)$$

where $\mathbf{V}(\boldsymbol{\vartheta}) = (\mathbf{D} \text{diag}\{\mathbf{V}(\mathbf{X})\}^{-1} \mathbf{D})^{-1} + \mathbf{U}\boldsymbol{\Sigma}(\boldsymbol{\vartheta})\mathbf{U}^T$, with $\mathbf{D} = \text{diag}\{\boldsymbol{\mu}(\boldsymbol{\theta}_1)\}$ and $\text{diag}\{\mathbf{V}(\mathbf{X})\}^{-1} = \text{diag}\{\boldsymbol{\mu}(\boldsymbol{\theta}_1)^{-2}\}$ and consequently $\mathbf{V}(\boldsymbol{\vartheta}) = \mathbf{I}_N + \mathbf{U}\boldsymbol{\Sigma}(\boldsymbol{\vartheta})\mathbf{U}^T$. The pseudo-observations $\tilde{\mathbf{y}}$ are given by $\log(\boldsymbol{\mu}(\boldsymbol{\theta}_1)) + \mathbf{D}^{-1}(\boldsymbol{\mu}(\boldsymbol{\theta}_0) - \boldsymbol{\mu}(\boldsymbol{\theta}_1))$. Estimators can be obtained by iteratively optimizing firstly (28) and secondly (29) in each iteration until convergence.

Chapter 7

In search of lost edges: A case study on reconstructing financial networks

Contributing Article:

Michael Lebacher, Samantha Cook, Nadja Klein and Göran Kauermann (2019): *In search of lost edges: A case study on reconstructing financial networks*.

Submitted to the Journal of Network Theory in Finance.

arXiv preprint <https://arxiv.org/abs/1909.01274>

Code at https://github.com/lebachem/lost_edges

Disclaimer:

Data relating to SWIFT messaging flows is published with permission of S.W.I.F.T. SCRL. SWIFT© 2019. All rights reserved. Because financial institutions have multiple means to exchange information about their financial transactions, SWIFT statistics on financial flows do not represent complete market or industry statistics. SWIFT disclaims all liability for any decisions based, in full or in part, on SWIFT statistics, and for their consequences.

Author Contributions:

The idea of reviewing and comparing different methods that are suitable for reconstructing networks from their marginals came from Samantha Cook and Göran Kauermann. Samantha Cook contributed extensive preprocessing of the raw data and added valuable domain knowledge. Further, the introduction is written by Samantha Cook and Michael Lebacher. The selection of the methods, their implementation in R together with the competitive comparison was done by Michael Lebacher. Also the main body of the article was written by Michael Lebacher. Nadja Klein gave input regarding the evaluation and the presentation of the reviewed methodology. All authors contributed to the manuscript writing and were involved in extensive proof-reading.

In Search of Lost Edges: A Case Study on Reconstructing Financial Networks*

Michael Lebacher[†], Samantha Cook[‡], Nadja Klein[§] and Göran Kauermann[¶]

Abstract

To capture the systemic complexity of international financial systems, network data is an important prerequisite. However, dyadic data is often not available, raising the need for methods that allow for reconstructing networks based on limited information. In this paper, we are reviewing different methods that are designed for the estimation of matrices from their marginals and potentially exogenous information. This includes a general discussion of the available methodology that provides edge probabilities as well as models that are focussed on the reconstruction of edge values. Besides summarizing the advantages, shortfalls and computational issues of the approaches, we put them into a competitive comparison using the SWIFT (Society for Worldwide Interbank Financial Telecommunication) MT 103 payment messages network (MT 103: Single Customer Credit Transfer). This network is not only economically meaningful but also fully observed which allows for an extensive competitive horse race of methods. The comparison concerning the binary reconstruction is divided into an evaluation of the edge probabilities and the quality of the reconstructed degree structures. Furthermore, the accuracy of the predicted edge values is investigated. To test the methods on different topologies, the application is split into two parts. The first part considers the full MT 103 network, being an illustration for the reconstruction of large, sparse financial networks. The second part is concerned with reconstructing a subset of the full network, representing a dense medium-sized network. Regarding substantial outcomes, it can be found that no method is superior in every respect and that the preferred model choice highly depends on the goal of the analysis, the presumed network structure and the availability of exogenous information.

Keywords: Density Calibration, Financial Networks, Inverse Problems, Maximum-Entropy, MT 103 Messages, Network Reconstruction, Network Tomography, SWIFT

***Disclaimer** Data relating to SWIFT messaging flows is published with permission of S.W.I.F.T. SCRL. SWIFT[®] 2019. All rights reserved. Because financial institutions have multiple means to exchange information about their financial transactions, SWIFT statistics on financial flows do not represent complete market or industry statistics. SWIFT disclaims all liability for any decisions based, in full or in part, on SWIFT statistics, and for their consequences.

[†]Department of Statistics, Ludwig-Maximilians Universität München, michael.lebacher@stat.uni-muenchen.de

[‡]Financial Network Analytics, Ltd. (FNA), sam@fna.fi

[§]School of Business and Economics, Humboldt-Universität zu Berlin, nadja.klein@hu-berlin.de

[¶]Department of Statistics, Ludwig-Maximilians Universität München, goeran.kauermann@stat.uni-muenchen.de

1. Introduction

In recent years, interest in applying network-based methodology to financial data has strongly increased (see e.g. Soramäki et al., 2007, Schweitzer et al., 2009, Imakubo et al., 2010, Baek et al., 2014, Battiston et al., 2016). A huge amount of this research effort is directed to the study and assessment of systemic risk (see e.g. Gai and Kapadia, 2010, Kauê Dal’Maso Peron et al., 2012, Billio et al., 2012, Chinazzi et al., 2013, Thurner and Poledna, 2013, Soramäki and Cook, 2013, Bardoscia et al., 2017 and Caccioli et al., 2018). This focus stems from the fact that in the aftermath of the financial crisis it became clear that the banking system forms a complex network with inherent interdependencies and feedback loops. As a consequence, the centrality and connectedness of a financial institution can be just as important as size for its potential to wreak havoc on the system overall (Markose et al., 2012, Liu et al., 2015). Battiston et al. (2012) even suggest to add the term “too-central-to-fail” to the discussion of “too-big-to-fail” institutions. Given that, investigation of the topologies of financial networks is very important for regulators, central banks and other institutions concerned with the stability of the financial system. Although considerable effort is put into modeling system risk, those methods generally require information from the full network that is most often not observed. This raises the need for a methodology that allows providing an accurate reconstruction of the networks derived from the limited information available.

The canonical examples of network reconstruction in finance are exposure networks created by interbank loans. In these networks, the total assets and liabilities of a given bank are mostly known, but the actual loans made to other banks, i.e. the binary edge structure (existence or non-existence of loans) and their corresponding edge weights (loan volume), are unobserved. Knowledge of the edges and their values is nevertheless crucial to measure the systemic risk in the exposure network. If one bank fails to meet its’ obligations that could lead its’ creditor(s) unable to make their obligations which leads to further contagion, potentially affecting all banks or a large portion of the network. An example of how to process such information, if available, is *DebtRank* (Battiston et al., 2012), being a popular metric for assessing systemic importance in exposure networks based on the values of loans between bank pairs.

Although the reconstruction problem is introduced here as a task that belongs to the realms of Finance or Economics, it emerges in many different disciplines. In order to get an overview, from the perspective of Economics, see for example Sheldon and Maurer (1998), Upper (2011) and Elsinger et al. (2013). The article by Squartini et al. (2018) provides a very broad overview from a methodological perspective, based on maximum-entropy methods and Statistical Physics (see also Cimini et al., 2015 and Mastrandrea et al., 2014). In Computer Sciences and Statistics, a similar problem is often called traffic matrix estimation or network tomography and this research branch developed its’ own methodological toolkit (e.g. Castro et al., 2004, Zhang et al., 2003b, Airoidi and Blocker, 2013, Zhou et al., 2016 and Nie et al., 2017). In the given paper, not all models proposed in different research fields can be included but we have selected the ones that are feasible and potentially useful for the given data situation.

A good reference point for this paper is certainly the extensive study by Anand et al. (2018). In their paper, they employ seven different reconstruction methods to 25 different networks. Although we are not that ambitious regarding the variety of use cases, our approach can be seen as a related paper that focusses on other aspects. First of all, we do not restrict our methodology to methods that rely only on aggregated row- and column sums but also include density-calibrated methods and models that are capable of incorporating exogenous covariates. Further, it is tried to propose regularized least-squares models inspired by the network tomography literature and new methodology not considered by Anand et al. (2018). Additionally, we provide a more de-

tailed technical exposition of the models in a manner that is comprehensible for practitioners. Regarding the evaluation techniques, we separate the evaluation of the binary and valued reconstruction more clearly and employ measures that are more standard in Statistics and Machine Learning.

To compare the different models, we use data provided by the Society for Worldwide Interbank Financial Telecommunication (SWIFT, www.swift.com). SWIFT acts as an infrastructure for financial institutions and enables them to send and receive information about financial transactions encoded in the form of secure standardized messages. One of the most important types of messages is the *MT 103 single customer credit transfer*, representing payments sent between clients of financial institutions. The MT 103 data under study consist of monthly bilateral message counts aggregated at the country level between January 2003 and February 2018. Note that the concept of a country here is not limited to independent, passport-granting states, but also includes territories (e.g. Turks and Caicos Islands), dependencies (e.g. Guernsey and Jersey) and autonomous constituent states (e.g. Greenland).

The SWIFT network is especially suitable for testing network reconstruction methods because it is an economically meaningful data set (see Cook and Soramaki, 2014 for an extensive investigation). Further, the data provides a long time series available for testing with full link data available. Hence, in this dataset, it is known exactly how well different methods work allowing to compare different models.

The article is structured as follows. In Section 2 we formalize the problem and give general notation for the paper. This is followed by a description of the SWIFT data in Section 3. In Section 4 we introduce the models under study and their evaluation is provided in Section 5. Section 6 discusses the results and concludes the paper.¹

2. Notation

The SWIFT MT 103 messages can be represented as a series of matrices $\mathbf{X}^t = (x_{ij}^t)$ containing dyadic count data. The elements of \mathbf{X}^t can be interpreted as directed edge values $x_{ij}^t \in \mathbb{N}_0$ among $i, j = 1, \dots, n$ countries at time points $t = 1, \dots, T$. We exclude self-loops from our study and, therefore, elements x_{ii}^t are left undefined for $i = 1, \dots, n$. Accordingly, within-country payments are not regarded. We also assume that the number of nodes n is invariant with respect to time so that at each time point t the number of variables is given by $N = n(n - 1)$.

2.1. Binary Network Structure

Although the binary networks structure is readily available if the valued structure is given, both aspects of the network need to be modelled separately. To account for this aspect, we also introduce notation for the binary network structure. Let $\mathbf{Z}^t = (z_{ij}^t)$ denote the binary networks, defined via

$$z_{ij}^t = I(x_{ij}^t > 0), \text{ for } i \neq j,$$

¹We provide the code online at Github https://github.com/lebachem/lost_edges. Because the used data set is confidential, the code is not accompanied with the actual dataset but with a “fake dataset” that does *not* represent the original data but only the same dimension and a similar density.

with elements z_{ij}^t being indicators whether the corresponding entry of the matrix is zero or greater than zero. Let the *density* (also called the connectivity) of the network be

$$\mathcal{D}^t = \frac{1}{N} \sum_{i \neq j} z_{ij}^t, \text{ for } t = 1, \dots, T$$

providing the number of non-zero edges in the network relative to the number of possible edges at time point t . Additionally, we define the number of outgoing edges to be the *outdegree* and the number of ingoing edges is measured with the *indegree*. Formally, the outdegree and the indegree for node i at time point t are given by

$$\begin{aligned} z_{i\bullet}^t &= \sum_{k \neq i} z_{ik}^t, \text{ for } i = 1, \dots, n \\ z_{\bullet i}^t &= \sum_{k \neq i} z_{ki}^t, \text{ for } i = 1, \dots, n. \end{aligned} \tag{1}$$

2.2. Valued Network

Similarly, we are interested in the row and column sums of the valued network, i.e. the *valued in-* and *outdegree*. Other names in the network literature describing the same concepts are the in-strength and out-strength or the weighted in- and outdegree. Let the i th valued outdegree and valued indegree be

$$\begin{aligned} x_{i\bullet}^t &= \sum_{k \neq i} x_{ik}^t, \text{ for } i = 1, \dots, n \\ x_{\bullet i}^t &= \sum_{k \neq i} x_{ki}^t, \text{ for } i = 1, \dots, n. \end{aligned} \tag{2}$$

For a more compact formulation, we stack the row and column sums, resulting in a $2n$ -dimensional column vector of marginals

$$\mathbf{y}^t = (x_{1\bullet}^t, \dots, x_{n\bullet}^t, x_{\bullet 1}^t, \dots, x_{\bullet n}^t)^T, \text{ for } t = 1, \dots, T.$$

Furthermore, let

$$\mathbf{x}^t = (x_{12}^t, \dots, x_{1n}^t, x_{21}^t, \dots, x_{n(n-1)}^t)^T, \text{ for } t = 1, \dots, T$$

be an N -dimensional column vector containing the values of the edges (without diagonal elements) and define the known binary $(2n \times N)$ routing matrix \mathbf{A}^t such that the linear relation

$$\mathbf{y}^t = \mathbf{A}^t \mathbf{x}^t \tag{3}$$

holds for $t = 1, \dots, T$. Note that relation (3) is just a compact way of writing equations (2) in matrix notation. Henceforth, we will refer to relation (3) as *marginal restrictions*. The restriction that all matrix entries are non-negative is referred to as *non-negativity constraint*. If we refer to methods that yield stochastic solutions we adopt the nomenclature from Physics and label a collection of sampled networks as *network ensemble* (e.g. Bargigli, 2014). In the model description we will suppress the time-superscript in most representations for ease of notation.

As a general convention, vectors and matrices are given in bold and (with the exception of the deterministic routing matrix \mathbf{A}), random variables are given by upper case and realisations

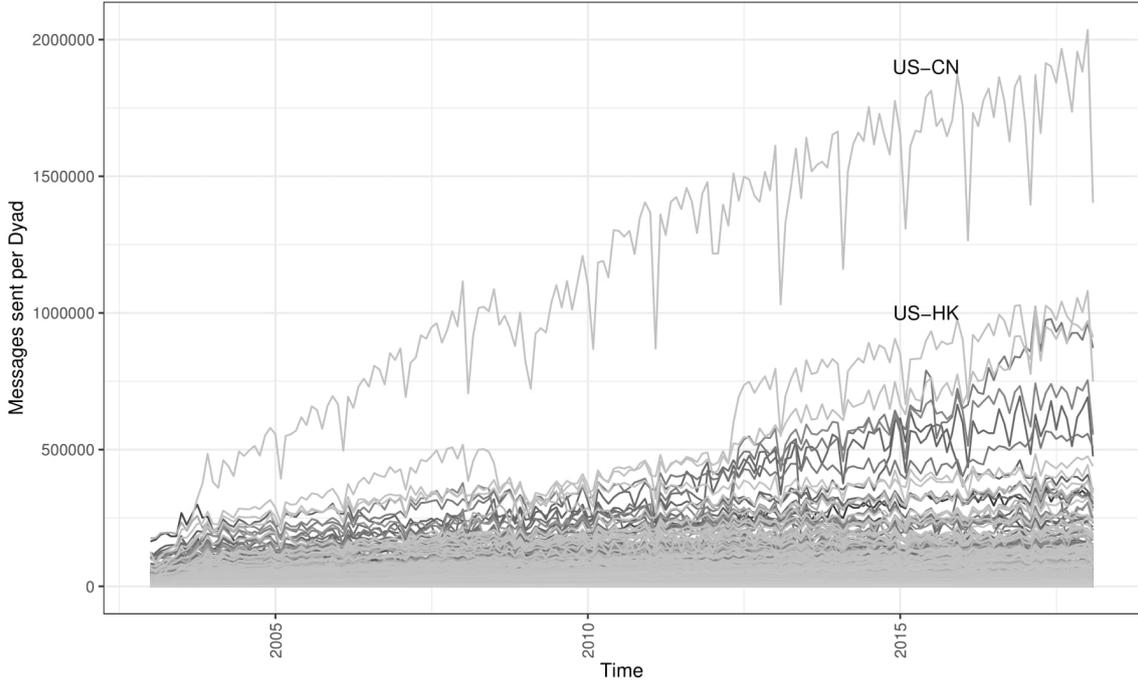


Figure 1: Time series of valued edges x_{ij}^t in the full MT 103 network on a monthly basis. Messages per edge on the vertical axis, time measured in months on the horizontal axis. Source: SWIFT BI Watch.

by lower case letters.

3. Data description

The data under study is provided by the Society for Worldwide Interbank Financial Telecommunication (SWIFT, www.swift.com) and provides standardized messages called MT 103, representing payment transfers. The data is aggregated to the country level and allow to construct a network \mathbf{x}^t where the countries are the nodes and the directed, valued edges x_{ij}^t between them represent the number of messages sent from country i to country j at time point t . The available database covers $T = 182$ time points on a monthly basis, ranging from January 2003 to February 2018.

We restrict our analysis to the $n = 203$ countries that are existent during the whole observational period. This includes one entity that is not a country or a territory but represents international market infrastructure, referring to monetary organizations that operate in many countries (see the country list in Table 4 of Appendix A). The network including all 203 countries is set to be the baseline for all models that can deal with “big” networks and is labeled *full network*.

In Figure 1 we plot all individual edges of the full network against time. Notably, there is a great deal of country-related heterogeneity in the data. The time series with the highest time-averaged amount of messages sent corresponds to the edge United States - China (US-CN) and is on average almost ten times higher than the second-highest valued edge (United States - Hong Kong, US-HK). Furthermore, already the yearly 80% quantile of the number of messages

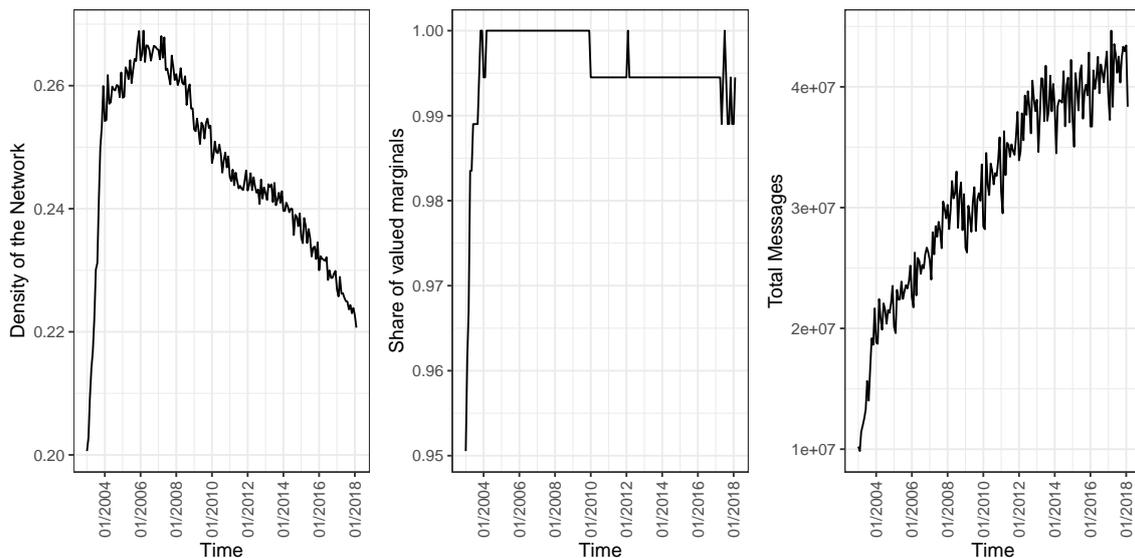


Figure 2: Summary statistics for the full MT 103 network as monthly time series. Density of the network (left), share of non-zero marginals \mathbf{y}^t (middle) and cumulative edge values (right). Source: SWIFT BI Watch.

within each month ranges only between one and five messages per edge. This implies that the major share of all messages is sent and received by a small subset of countries characterized by a high-intensity exchange.

To analyze both, the full network and its’ “dense core” we additionally investigate a reduced dataset, containing the 59 most important countries. This network is labeled to be the *reduced network*. Depending on the month, the reduced network accounts for about 85% up to 91% of all messages sent in the whole system. In the following, we show descriptive measures for the full network but in the Annex B the same descriptives are provided for the reduced network.

The structure of the full binary networks \mathbf{z}^t is summarized in Figure 2. On the left-hand side, it can be seen that the density of the network decreases steadily from 2006 on whereas, the development of the total MT 103 messages (right-hand side) follows a clear upward trend. This pattern implies that increasingly more messages are sent per edge. Similarly, but in a more modest form, this can also be concluded for the reduced network (see Figure 10 in the Annex B). The SWIFT data exclusively contains countries that send or receive MT 103 messages. Therefore, each country can only have either a zero out- or indegree. However, if many countries would be restricted to only receive or send messages, the dimensionality of the problem could be greatly reduced. This can be investigated by calculating the share of non-zero marginals \mathbf{y}^t for each year. The resulting plot is given in the middle plot of Figure 2 and we find that the low density is not mirrored by a low share of valued marginals. This means that almost no information about the density can be inferred from the marginals since the vast majority of them are greater zero. In the reduced network, the density is much higher (about 0.85 averaged over all months) without any zero marginal.

In Figure 3 the degree structure of the full network is visualized. The first two panels show the cumulative degree distribution (indegree on the left and outdegree on the right) aggregated for all months. The realizations between the monthly minimum and maximum values are indicated in grey. It can be seen that both, the indegree as well as the outdegree grow close

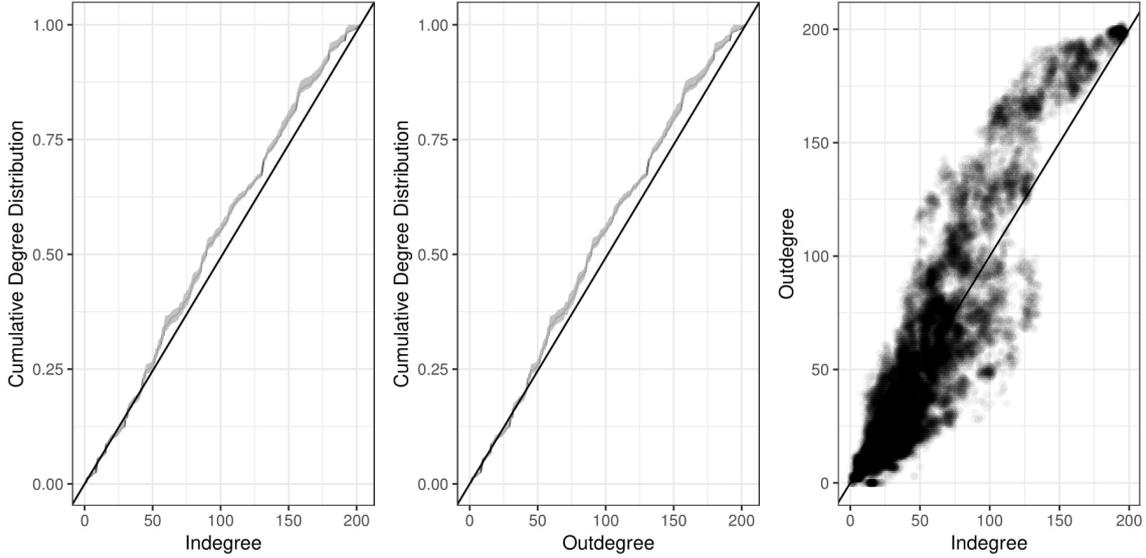


Figure 3: Binary network topology of the full MT 103 network aggregated for all time points. Cumulative indegree (left) and outdegree distributions (middle) with maximum and minimum values indicated in grey. Outdegree against indegree (right) for all months in dotted with colour intensity by frequency. 45 degree line in solid black.

Source: SWIFT BI Watch.

to linear and are, therefore, almost uniformly distributed. This is a rather uncommon finding and does not match with common structures like scale-free networks (Barabási and Albert, 1999; Albert and Barabási, 2002) or random graphs (Erdős-Rényi graphs, Erdős and Rényi, 1959). This is consistent with the findings of Cook and Soramaki (2014) who noted that the data cannot appropriately be described with standard power-law distributions. Again similar and even more pronounced results can be found for the reduced network, shown in Figure 11 of Annex B.

The right panel of Figure 3 plots shows for all nodes and all months the indegree versus the outdegree. This can be thought of a check how “symmetric” the network is and it appears that there is a strong positive (non-linear) relationship between the in- and outdegrees.

In some models, exogenous data can be incorporated. Based on the empirical investigation by Cook and Soramaki (2014), we assume it to be plausible that financial activity in a given country is related to its’ economic size and consider the annual *Gross Domestic Product* (GDP, in current USD Billions) as a valid covariate. The data is provided by the International Monetary Fund (IMF) and we denote the GDP of country i by gdp_i .

4. Models for Network Reconstruction

4.1. Overview

Since almost none of the zeros in the network can be inferred from the marginals, most of the models that provide edge probabilities rely on two crucial assumptions: (i) The true density is known and (ii) the row- and column sums of the valued edges carry information about the binary structure. Both points are highly related and evolve around the basic problem that

Method	Abbreviation	Section	Full netw.	Calibrated
Maximum-Entropy	<i>IPFP</i>	4.2	X	
Maximum-Entropy, GDP	<i>IPFP-GDP</i>	4.2		
Maximum-Entropy, lag. values	<i>IPFP-LAG</i>	4.2		
Gravity Model	<i>GRAVITY</i>	4.3	X	
Dens. cor. Gravity Model	<i>DC-GRAVITY</i>	4.3	X	X
Dens. cor. Gravity, GDP	<i>DC-GRAVITY-GDP</i>	4.3		X
Dens. cor. Gravity, lag. values	<i>DC-GRAVITY-LAG</i>	4.3		X
Tomogravity Model	<i>TOMOGRAVITY</i>	4.4		
LASSO	<i>LASSO</i>	4.5	X	X
Hierarchical Erdős-Rényi Model	<i>H-ER</i>	4.6	X	X
Hierarchical Fitness Model	<i>H-FIT</i>	4.6	X	X
Minimum Density	<i>MINDENS</i>	4.7	X	

Table 1: Summary of reconstruction methods used in this article together with abbreviations, their ability to fit the full network (Full netw.) and whether calibration to the true density is needed (Calibrated).

knowledge of the marginals is not sufficient to provide information about the edge probabilities (Gandy and Veraart, 2017, Proposition 3.1). This fundamental identification problem can be overcome only by adding additional constraints (i.e. knowledge of the density). For the sake of this article, we assume the true density to be known. In practice, this implies that models that are found to perform very well in our comparison might not do so with an incorrectly specified density. Given knowledge about the true density, the second assumption is less problematic and, depending on the presumed structure of the network under study, it can be plausible that the marginals provide information that helps to determine the edge probabilities.

Another issue that complicates network reconstruction is the high dimensionality of the full network. With $n = 203$ nodes, the number of dyads amounts to $N = 41\,006$. This brings many methods to their computational limits. In the reduced network, the problem greatly simplifies as N shrinks to 3422 which allows to apply almost all methods considered in this paper. An exception is the density-corrected directed weighted configuration model (DWCM) by Bargigli (2014) that is not considered in this paper because the algorithm failed to converge even in the small network. In the model description, we will mention which methods are computationally tractable in the full network and in case they are not, they are only applied to the reduced network.

All methods used are summarized in Table 1, including their names, abbreviations and references to the corresponding sections with a detailed description. Additionally, it is shown whether the methods are applied to the full network and whether knowledge of the true density is needed to calibrate the models.

4.2. Iterative Proportional Fitting

A very simplistic, but nevertheless powerful method to reconstruct *dense* networks is given by the iterative proportional fitting procedure (IPFP, Deming and Stephan, 1940, Fienberg et al., 1970). The algorithm has gained much attention in the matrix reconstruction literature under the name maximum-entropy method because it allows for estimating the parameters of the maximum-entropy probability distribution (the methodological backbone of many reconstruc-

tion tasks, Squartini et al., 2018).

In the Statistics literature, the procedure is originally intended to provide maximum likelihood estimates for parameters of log-linear models in contingency tables (Bishop et al., 1975, Haberman, 1978, 1979). In the given case, this interpretation is convenient because it allows for a specific interpretation of the outcomes as the maximum likelihood estimates for the expectation of a Poisson-distributed random variable

$$X_{ij} \sim Poi(\mu_{ij}), \text{ for } i \neq j, \quad (4)$$

with log-linear expectation $\mathbb{E}[X_{ij}] = \mu_{ij} = \exp\{\delta_i + \gamma_j\}$. The two parameters δ_i and γ_j correspond to row- and column-effects. Furthermore, the model provides a model-based possibility to calculate the probability of observing a value X_{ij} greater than zero:

$$p_{ij} := \mathbb{P}(X_{ij} > 0) = 1 - \mathbb{P}(X_{ij} = 0) = 1 - \exp\{-\mu_{ij}\}, \text{ for } i \neq j.$$

However, with high values for X_{ij} , the probabilities approach zero exponentially fast, meaning that for high marginals most probabilities will be almost or numerically even equal to one.

The model is a dense reconstruction method that provides edge values for all rows and columns with valued marginals. Besides this drawback, the model has the merit of being computationally efficient (see the R package `ipfp` by Blocker et al., 2014) and due to the construction of the algorithm, it is guaranteed that the row- and column sums of the predicted entries match the observed marginals exactly.

In Lebacher and Kauermann (2019) the IPFP model from above is extended to incorporate informative dyadic exogenous information, which is labeled here as C_{ij} . In particular, the covariates C_{ij} can be included in the log-linear expectation

$$\mathbb{E}[X_{ij}|C_{ij} = c_{ij}] = \mu_{ij} = \exp\{\delta_i + \gamma_j + c_{ij}\beta\}, \text{ for } i \neq j.$$

If the association between C_{ij} and the unknown X_{ij} is high, the prediction accuracy increases relative to the standard IPFP solution. However, this approach comes at a price since model fitting is based on constrained non-linear optimization making computation significantly more demanding than standard IPFP. Furthermore, only dyadic covariates have the potential to increase the predictive power but in practice often only monadic information is available. We take a pragmatic approach and use a transformation of the GDP values of countries i and j that is not linearly separable and can be interpreted as dyadic overall GDP, defined through:

$$c_{ij} = \log(gdp_i + gdp_j), \text{ for } i \neq j. \quad (5)$$

This yields the expectation

$$\mathbb{E}[X_{ij}|C_{ij} = c_{ij}] = \mu_{ij} = \exp\{\delta_i + \gamma_j\} + (gdp_i + gdp_j)^\beta, \text{ for } i \neq j.$$

To test the power of the model in situations where a covariate with a strong association is available, we also include a logarithmic transformation of the lagged edge values:

$$c_{ij}^t = \log(1 + x_{ij}^{t-1}), \text{ for } i \neq j. \quad (6)$$

Estimation is pursued as described in Lebacher and Kauermann (2019), with a constrained Poisson likelihood. In the following, the IPFP-type models using GDP and the lagged variables are denoted *IPFP-GDP* and *IPFP-LAG*, respectively.

4.3. Gravity Models

Gravity models are at the heart of many methods related to the analysis of network flow data (Kolaczyk, 2009). Besides their successful application to economic trade data (Disdier and Head, 2008, Head and Mayer, 2014) they are also among the preferred models for network tomography in Computer Sciences (Vardi, 1996). Network tomography relates to a problem that often appears when analyzing computer networks. Here, the individual edge loads are assumed to be *known* but the flow is allowed to intersect the nodes in the network. The task is then to provide accurate predictions for flows between arbitrary nodes. Very often the gravity model is found to be among the best algorithms to solve this problem (Zhang et al., 2003a). Although the formulation of the problem seems to be very different compared to the network reconstruction task, it leads to the same mathematical structure.

From a methodological point of view, the gravity model is simply a special case of the IPFP model discussed above and in fact, the gravity model is the immediate maximum-entropy solution in each network reconstruction problem where self-loops are allowed (Squartini et al., 2018, Sheldon and Maurer, 1998). Mathematically, the model builds on a simple multiplicative structure

$$\hat{\mu}_{ij} = \frac{x_{i\bullet}x_{\bullet j}}{x_{\bullet\bullet}}, \text{ for } i \neq j, \quad (7)$$

with $x_{\bullet\bullet}$ representing the sum over all valued in- or outdegrees. Though simple in structure and fast to compute, the model has two main drawbacks. First, the model yields biased results if the diagonal elements are restricted to be zero because then the row and column sums of the predictions do not match the marginal restrictions exactly. However, in big networks, the bias is often negligible. Second, as in all maximum-entropy models, the approach relies on inferring sparseness from the marginals and predicts exclusively non-zero matrix entries if all marginals are greater than zero.

Because economic and financial networks most often exhibit a density smaller than one, Cimini et al. (2015) proposed a model that is designed for reconstructing the binary structure of networks with limited information available. Basically, they extend the gravity model from above towards a two-step procedure. In the first step they propose to model the probabilities of observing an edge with a parameter α such that they match with the pre-defined targeted density

$$\mathcal{D} = \frac{1}{N} \sum_{i \neq j} P(X_{ij} > 0; \hat{\alpha}) = \frac{1}{N} \sum_{i \neq j} \frac{\hat{\alpha} \chi_i \psi_j}{1 + \hat{\alpha} \chi_i \psi_j}, \quad (8)$$

where the parameters χ_i and ψ_j are node-specific *fitness variables*. Following the idea that the marginals carry information about the binary network structure, they are typically set equal to the marginals (i.e. $\chi_i = x_{i\bullet}$ and $\psi_j = x_{\bullet j}$) or some transformation of them. Another interpretation is that the economic strength determines the fitness of a country or previous bilateral exchanges influence the fitness of dyadic relations. We include the transformed GDP values and set $\chi_i \psi_j = c_{ij}$ as defined in equation (5). For the logarithmic lagged exchange we set $\chi_i \psi_j = \log(1.1 + x_{ij}^{t-1})$ as fitness variables. Note that adding 1.1 instead of 1 prevents the probabilities from being zero irrespective of α in cases with $x_{ij}^{t-1} = 0$.

The parameter $\hat{\alpha}$ can be found by any precise root-search program. In applications with larger dimensionality, the values for $\hat{\alpha}$ might become numerically very small and we use a genetic algorithm (implemented in the R package `GA` by Scrucca, 2013) to overcome this problem. Given an estimate for α that satisfies (8), Cimini et al. (2015) propose to sample binary networks network ensembles with variables \hat{z}_{ij} and use a density-corrected version of model (7) for the

edge values

$$\hat{\mu}_{ij} = \frac{\hat{\alpha}^{-1} + x_{i\bullet}x_{\bullet j}}{x_{\bullet\bullet}} I(\hat{z}_{ij} > 0), \text{ for } i \neq j. \quad (9)$$

In the following we refer to the density-corrected gravity model by *DC-GRAVITY* and the models with GDP and lagged variables are abbreviated by *DC-GRAVITY-GDP* and *DC-GRAVITY-LAG*.

4.4. Tomogravity model

An important model candidate from the network tomography literature is proposed by Zhang et al. (2003b). In their article, the problem of learning origin-destination flows from link load data in IP networks motivates the estimation of a traffic matrix. The authors regard the problem as an ill-posed regression problem that must be regularized with the Kullback-Leibler divergence from an independence model. The predicted values can be found by minimizing the loss-function

$$L(\boldsymbol{\mu}) = (\mathbf{A}\boldsymbol{\mu} - \mathbf{y})^T(\mathbf{A}\boldsymbol{\mu} - \mathbf{y}) + \psi^2 \sum_{i \neq j} \frac{\mu_{ij}}{N} \log \left(\frac{\mu_{ij}}{x_{i\bullet}x_{\bullet j}} \right) \quad (10)$$

with respect to $\boldsymbol{\mu} = (\mu_{12}, \dots, \mu_{n(n-1)})^T$ and subject to the non-negativity constraint. The first term is simply the sum of squared deviations from the marginals. In the penalization term, the gravity model serves as a null model together with a regularization parameter ψ . Note that the model is a dense reconstruction technique and does neither provide probabilities nor do the predictions match with the observed marginals.

Although this appears to be an appealing combination between the successful gravity model and information-theoretic reasoning, the procedure is so far seldom applied to the reconstruction of networks. The approach is implemented in the R package `tomogravity` (see Blocker et al., 2014). The implementation is computationally expensive and we, therefore, apply this model only for the reduced data set. Zhang et al. (2003b) show in a simulation study, that the performance of the algorithm is not very sensitive to varying values of ψ and as a rule of thumb they recommend to use $\psi = 0.01$ if no training data are available and we follow their rule in the application section.

4.5. LASSO Model

Regarding the network reconstruction problem again as an ill-posed regression problem, it might not even be necessary to make use of a new penalization term. Instead, the least absolute shrinkage and selection operator (*LASSO*) approach proposed by Tibshirani (1996) can be employed, which uses a L_1 penalty to enforce sparsity in the model. Although approaches with some kind of regularization are common in network tomography (Castro et al., 2004) the LASSO is applied rather rarely for network reconstruction. An exception is given by Chen et al. (2017) who propose a LASSO-type model to predict flows in a bike-sharing network from station traffic (number of ingoing and outgoing bikes at each station).

Technically, the quadratic deviation from the marginals is combined with a regularization term that penalizes the sum of the predicted matrix entries, yielding the following loss function

$$L(\boldsymbol{\mu}) = (\mathbf{A}\boldsymbol{\mu} - \mathbf{y})^T(\mathbf{A}\boldsymbol{\mu} - \mathbf{y}) + \tau \sum_{i \neq j} |\mu_{ij}|. \quad (11)$$

By the non-negativity constraint, the absolute value in the penalization term can be dropped.

The R package `glmnet` by Friedman et al. (2009) allows for efficient and scalable estimation.

In principle, the model might appear to be attractive because the regularization shrinks some predictions exactly to zero. However, it is not clear how to derive the penalization parameter τ because cross-validation aiming at the marginals does not lead to satisfactory results. Chen et al. (2017) propose to use a training data set - information that might not always be available. To use the approach nevertheless in the competitive comparison without a training set available, we optimize the penalty parameter τ on a grid such that the number of non-zero coefficients is consistent with the real density.

Note further, that the predicted marginals are, by construction, always be smaller than the observed ones because of the shrinkage property of the LASSO. On the other hand, the model has much potential for exploratory analysis by investigating the path plots of the coefficients, i.e. the values of the coefficients against increasing values of τ .

4.6. Hierarchical Fitness Models

A central finding of the study by Anand et al. (2018) states that no method works equally well for different reconstruction tasks. Based on this insight, Gandy and Veraart (2017) proposed that a construction method should be adjustable to topological characteristics and especially to the density of a network. To do so, they present a hierarchical model designed for the reconstruction of financial networks. In the hierarchy of the model, the first step consists of estimating the edge probabilities consistent with the target density \mathcal{D} . As a baseline model, the authors propose an Erdős-Rényi model with

$$p_{ij} = p, \text{ for } i \neq j,$$

treating each edge to be equally likely. Given the obtained set of probabilities, edge weights are sampled from an exponential distribution with common expectation

$$\mu = \mathbb{E}[X_{ij}|Z_{ij} = 1], \text{ for } i \neq j. \quad (12)$$

The sampling algorithm is constructed such that the sampled networks provide stochastic network ensembles but each realization is consistent with the marginal restrictions.

Additionally, they proposed a model that is inspired by fitness-based approaches similar as in equation (9). In this model, the edge probability is determined by the logistic function

$$p_{ij}(\alpha) = \frac{1}{1 + \exp\{-\alpha - \log(x_{\bullet i} + x_{i \bullet}) - \log(x_{\bullet j} + x_{j \bullet})\}}, \text{ for } i \neq j, \quad (13)$$

with α being some constant that is estimated for consistency with the target density. In this model, the marginals serve as log-transformed fitness variables. In principle, any kind of variables could be used for the fitness model but only the marginals are yet implemented in the R package `systemicrisk`. The software implementation is very efficient and not overstrained by the dimensionality of the full network. Nevertheless, the algorithm is in trouble with the high values of the marginals. In the given application the marginals are scaled down in the estimation procedure and the predictions are then rescaled again.

By construction, the model puts much more emphasis on the binary network structure than on the prediction of the edge values. This is because the marginals are used directly only in the first step to estimate the edge probabilities. In the second step, all edge values are assumed to share the same expectation (12) and the marginal constraints enter only indirectly as a

restriction.

In the comparison, the hierarchical Erdős-Rényi model is abbreviated by *H-ER* and the hierarchical fitness model is called *H-FIT*.

4.7. Minimum Density

Anand et al. (2015) noted that the problem of binary network reconstruction can be viewed as finding a solution between two extreme points in the space of possible networks. Either, a maximally dense solution is searched for (maximum-entropy approaches), or it is the goal to find a solution with a minimal number of non-zero edges that are still consistent with the marginal constraints. Given that financial networks are typically sparse and disassortative, maximum-entropy solutions almost certainly provide an incorrect binary network structure.

In principle, if the density of the network is driven to the lowest level possible, the allocation of the edge weights might even become a simple task because of the small number of possibilities that are left. In its original form, the loss function of the minimum density model is simply given by the number of non-zero edges

$$L(\boldsymbol{\mu}) = \sum_{i \neq j} I(\mu_{ij} > 0),$$

subject to the marginal constraints and the non-negativity constraint. The loss function is not differentiable and direct minimization is computationally expensive. To circumvent this obstacle, Anand et al. (2015) relax the problem by giving up the assumption that the marginal constraint must hold exactly and shift the focus on the quadratic deviations from the marginals. Then, the authors propose an algorithm that implements two Markov processes, one adds new edges and weights and the second one deletes edges. Initialized with an arbitrary network the algorithm iterates as long as the loss function does not decrease any more together with a sufficient fit for the marginals.

The proposed algorithm is stochastic with non-unique solutions and generates ensembles of low-density networks. Typically, the realization with the lowest density is taken to be the optimal estimate (called *MINDENS* henceforth). By definition, the method does not rely on knowledge of the real density \mathcal{D} . Therefore, it is appropriate to regard the model as a lower-bound (in the space of feasible networks that satisfy the marginal constraints) instead of viewing it as an accurate reconstruction. This also has implications for the edge values, because a minimal number of edges in the system leads to maximal concentration of the edge values on a few nodes.

5. Evaluation

5.1. Binary Network Reconstruction

We evaluate the quality of the binary network reconstruction with different measures. For models that provide edge probabilities, we use the area under the curve (AUC) of the receiver-operating characteristic (ROC) curve and the precision-recall (PR) curve (see Grau et al., 2015). We regard both measures as complementary for model evaluation. While the ROC curve is, so to speak, ignorant about how good we predict either $Z_{ij}^t = 1$ or $Z_{ij}^t = 0$, the PR curve describes how well the models do in predicting $Z_{ij}^t = 1$. This is relevant because in low-density networks it is simpler to predict a zero than a one. Further, we look at the Bier score decomposition

proposed by Murphy (1973) (see also Siegert, 2017). For each time period t and model, we obtain K^t different probabilities $\hat{p}_{ij}^t \in \{\hat{p}_1^t, \dots, \hat{p}_{K^t}^t\}$ with n_k^t equal probabilities that correspond to edges z_k^t for $k = 1, \dots, K^t$. The Bier score decomposition is given by

$$BR_t = \frac{1}{N} \sum_{i \neq j} (\hat{p}_{ij}^t - z_{ij}^t)^2 = \underbrace{\sum_{k=1}^{K^t} \frac{n_k^t}{N} \left(\frac{z_k^t}{n_k^t} - \hat{p}_k^t \right)^2}_{REL_t} + \underbrace{\sum_{k=1}^{K^t} \frac{n_k^t}{N} \left(\frac{z_k^t}{n_k^t} - \mathcal{D}^t \right)^2}_{RES_t} - \underbrace{\mathcal{D}^t(1 - \mathcal{D}^t)}_{UNC_t}.$$

The reliability REL_t measures the distance between the estimated probabilities and the average real frequencies, with 0 being the best value that can be achieved. This means that a *low reliability* actually is the preferred outcome. The term labeled UNC_t is called uncertainty and gives the variability of the edges in the sample. Resolution (RES_t) gives the difference between the different share of empirical probabilities for each of the K^t categories and their overall average. Hence, it is a measure for the ability to discriminate between zero and one. A higher value indicates a better resolution. If the ability to discriminate is at its maximum, all probabilities are either one and zero, in this situation it holds that $RES_t = UNC_t$. We report these measures aggregated for all years and show the aggregated difference $UNC_t - RES_t$ which can be interpreted as the reduction of the uncertainty due to resolution.

We follow Squartini et al. (2018) and provide graphical representations of the reconstructed networks in the Appendices C.1 and D.1. There, the reconstructed adjacency matrices (based on binarization with a threshold according to the true density) for the most recent full and reduced network are shown.

We are not only interested in the prediction of individual edge occurrences but also in the quality of the reconstructed network topology. Given the strong heterogeneity in the network, the degree distribution can be regarded as a very important measure for the binary structure. We evaluate the fit of the outdegree distribution using the square root of the mean squared error of the real and the reconstructed outdegree distribution

$$RMSE_{od}^t = \sqrt{\frac{1}{n} \sum_{j=1}^n \left\{ \sum_{i=1}^n I(z_{i\bullet}^t = j) - \sum_{i=1}^n I(\hat{z}_{i\bullet}^t = j) \right\}^2}, \text{ for } t = 1, \dots, T$$

and correspondingly for the indegree.

To make the models comparable, we calibrate all estimates to the same target density. For models that are not scaled to the real density, we use a pragmatic approach and take the highest \mathcal{DN} (the number of edges in the real network) estimates to be one and all other estimates to be zero while in the probability-based models, we use the highest \mathcal{DN} probabilities to predict a one.

A visual impression of the quality of the degree reconstruction is given in Appendix C.2, plotting the predicted outdegree (indegree) against the real outdegree (indegree) for the most recent network observation of the full network and in D.2 for the reduced network.

5.1.1. Full Network

In the full network, four different models can be compared using AUC values and the decomposed Brier score. These four models include the iterative proportional fitting model from Section 4.2 (IPFP), the density-corrected gravity model by Cimini et al. (2015) from Section 4.3 (DC-GRAVITY) and the two hierarchical models from Section 4.6, with edge-probabilities coming

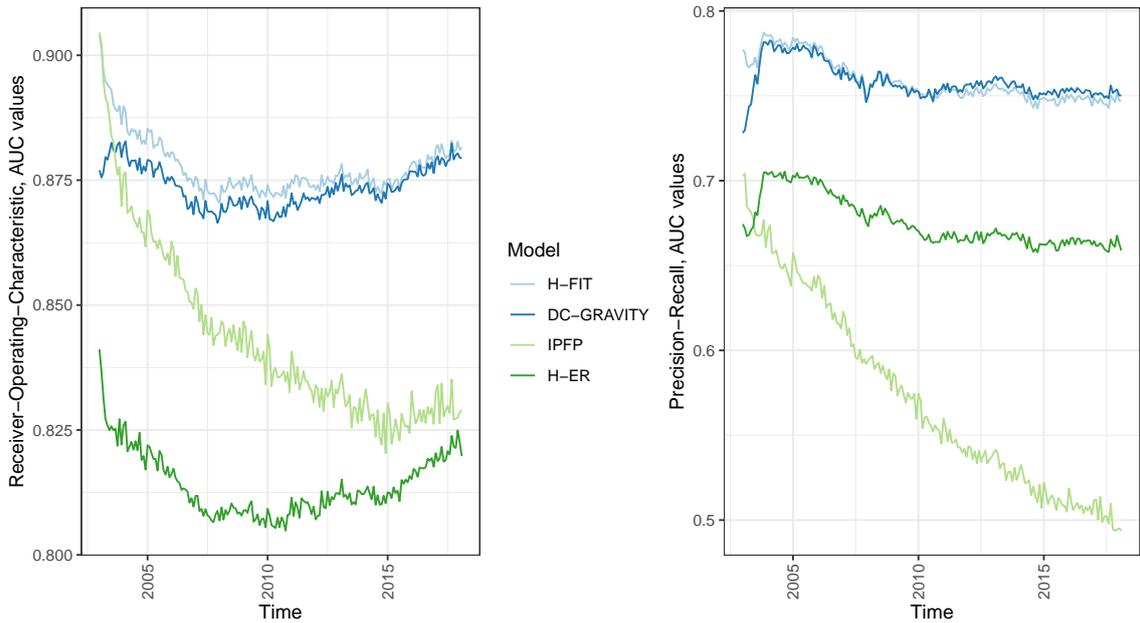


Figure 4: Evaluation of probabilities in the full network. Time series of the area under the curve (AUC) values for receiver-operating-characteristics (ROC, left panel) curve and precision-recall (PR, right panel) curve for the IPFP model, the degree corrected Gravity model (DC-GRAVITY), the hierarchical Erdős-Rényi model (H-ER) and the hierarchical fitness model. Source: SWIFT BI Watch.

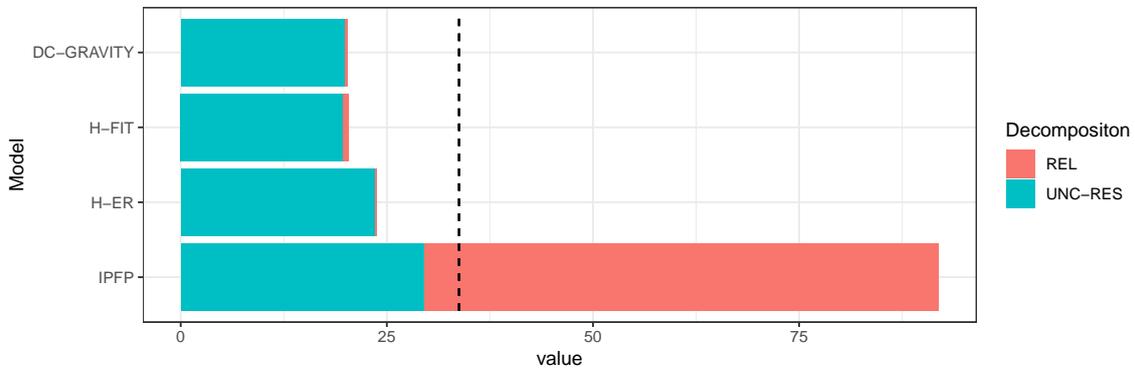


Figure 5: Decomposition of the Brier score for the full network into reliability (REL) and remaining uncertainty after subtracting resolution (UNC-RES) for the IPFP model, the Gravity model (GRAVITY), the degree corrected gravity model model (DC-GRAVITY), the hierarchical fitness models (H-ER, H-FIT). Uncertainty (UNC) as a dashed line. Source: SWIFT BI Watch.

either from the Erdős-Rényi (H-ER) or the fitness model (H-FIT). In Figure 4, we plot the AUC values for the ROC (left panel) and PR (right panel) curves against time. In Figure 5 we show the decomposition of the Brier score, with uncertainty (UNC) as a dashed vertical line.

For the reconstruction of the degrees, additionally the Gravity model (GRAVITY) from

Section 4.3, the LASSO from Section 4.5 and the minimum density (MINDENS) method of Section 4.7 enter the comparison. This is visualized in Figure 6 with the root mean squared errors for the outdegree in the left panel and for the indegree on the right. In both figures, the abbreviations of models are ordered to approximately match the time-averaged height of the respective measures.

Edge Probabilities

The hierarchical Erdős-Rényi (H-ER) model performs worst in the left panel and second-worst in the right panel of Figure 4. Seemingly, the assumption of equal probabilities for all dyads is strongly violated in this network. This relates to the discussion in Section 3 where we showed that the patterns of the degree distribution do not match with the Erdős-Rényi model.

However, also the IPFP model that allows for differing edge probabilities in its' Poisson interpretation does not perform satisfactorily and we see a declining trend of the prediction accuracy with time. As a consequence, the AUC values of the ROC curve decrease strongly and when evaluated with the PR curves, the model even provides the worst outcomes. This is a result of the growing values of the marginals, implying that the IPFP probabilities become very close to one or even numerically equal to one, leading to a loss of variation among the probabilities.

The two winners of this comparison, the density-corrected gravity model (DC-GRAVITY) and the hierarchical fitness model (H-FIT), give very similar accuracy measures in both panels of Figure 4. The AUC values for the ROC curves provided by the H-FIT model are slightly better than the ones of the DC-GRAVITY model and the other way round when evaluated with the PR curves. The strong similarity of the models' predictive power is, in fact, intuitive and results from the comparable choice of functions for determining the edge probabilities.

These results can be supported by the decomposition of the aggregated Brier score shown in Figure 5. The DC-GRAVITY model and the H-FIT model both provide a very low reliability measure and a comparatively high resolution. Interestingly, they are closely followed by the H-ER model that does not appear to be much worse with respect to the Brier score. Different from that, we find that the IPFP model has a low resolution and a high reliability measure, indicating that provided probabilities deviate strongly from the real ones and the ability to separate the predictions into "0" and "1" is rather low. Again this is because the IPFP model is not calibrated and many predictions are numerically just equal to one.

Degree Structure

Turning to the reconstruction of the degree structure, the different scaling of the two panels in Figure 6 shows that it is simpler to reconstruct the indegrees as compared to the outdegrees. The minimum density solution (MINDENS) marks an extreme case, resulting in the worst reconstruction of the out- and indegree structure. However, MINDENS has the comparative disadvantage of not being calibrated to the density and predicts far fewer edges than present in the real networks. Therefore, fewer edges can be allocated to certain nodes. With the exception of the United States, the model predicts no out- or indegrees above 65 at all (see also Figure 24 in the Annex C.2).

The LASSO provides the most unstable behavior and exhibits a high variance. Although the model is calibrated to the real density, the edge reconstruction is second-worst and delivers unsatisfactory reconstructions for the out- and the indegree. In Figure 25 of Annex C.1 it can be seen that the reconstructed degrees look almost random and Figure 18 indicates that the model is not able to make efficient use of the provided information on the row and column sums.

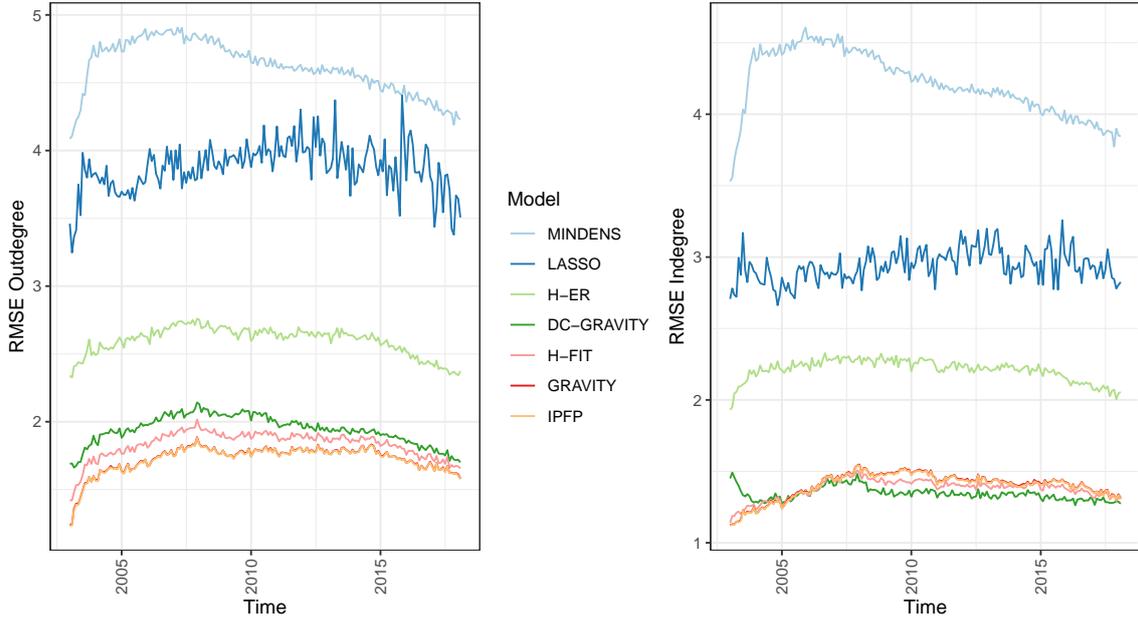


Figure 6: Time series of root mean squared error (RMSE) for the reconstruction of the outdegree (right) and the indegree (left) of the full network for the IPFP model, the Gravity model (GRAVITY), degree corrected gravity model model (DC-GRAVITY), the hierarchical fitness models (H-ER, H-FIT), the LASSO model and the minimum density model (MINDENS).

Source: SWIFT BI Watch.

The performance of the hierarchical Erdős-Rényi model (H-ER) shows that ignorance about the marginals for predicting the binary structure also can lead to unsatisfactory outcomes. The H-ER model “over-estimates” the out- and indegree for countries with medium-sized degrees and “under-estimates” the out- and indegree for countries with high degrees. This is clearly a result of the assumption that all edges are equally likely (as long as consistent with the marginals), leading to a random block structure in the network.

The Gravity model (GRAVITY) and the IPFP model make the best use of the information provided by the marginals to reconstruct the outdegree but not for the indegree. Their predictive quality concerning the degrees is almost identical and it is hard to distinguish both models in the left panel (IPFP overlays GRAVITY in both plots).

The hierarchical fitness model (H-FIT) together with the degree corrected gravity model (DC-GRAVITY) perform slightly worse than the GRAVITY and IPFP models concerning the outdegree but can be said to be the winner in the competition for the indegree reconstruction.

5.1.2. Reduced Network

In the reduced network, a greater variety of models can be investigated. Essentially, we can add four additional models in our comparative study, by extending the degree corrected gravity model (DC-GRAVITY) from Section 4.3 with the usage of GDP (DC-GRAVITY-GDP) and the lagged values (DC-GRAVITY-LAG) for determining the edge probabilities as well as the extended IPFP approach from Section 4.2 using the GDP values (IPFP-GDP) and the lagged values (IPFP-LAG) as covariates. In the degree reconstruction part, additionally the TOMO-

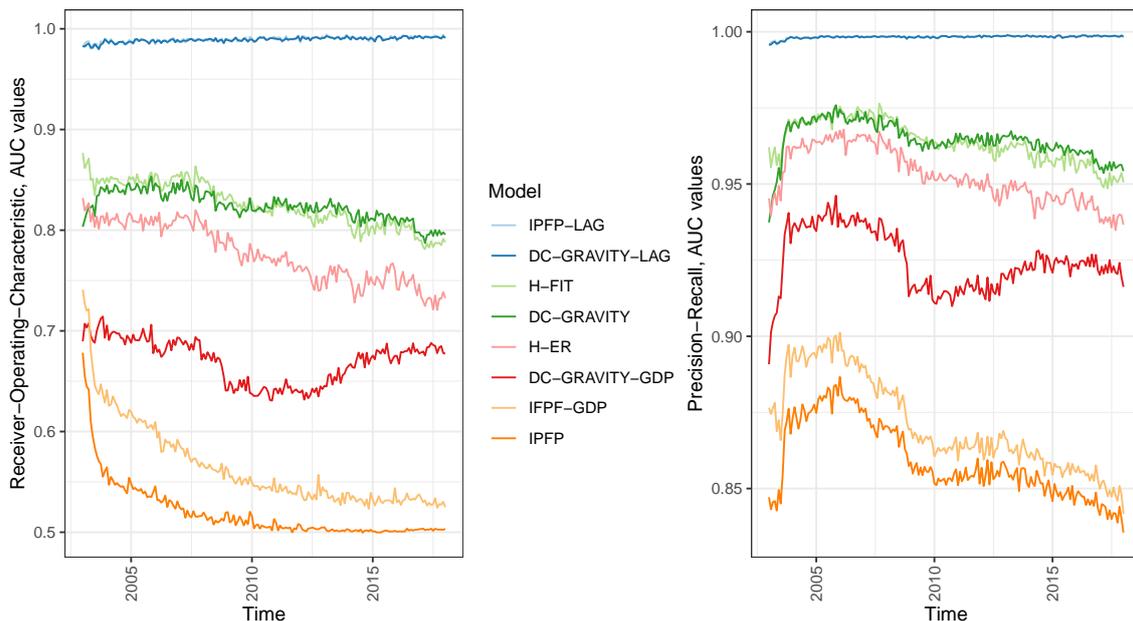


Figure 7: Evaluation of probabilities in the reduced network. Time series of area under the curve (AUC) values for receiver-operating-characteristics (ROC, left panel) and precision-recall (PR, right panel) for the IPFP model, the degree corrected gravity model (DC-GRAVITY) with covariates (DC-GRAVITY-GDP, DC-GRAVITY-LAG), the hierarchical Erdős-Rényi model (H-ER), the hierarchical fitness model (H-FIT) and the IPFP-based models with covariates (IPFP-GDP, IPFP-LAG).

Source: SWIFT BI Watch.

GRAVITY model (Section 4.4) is considered. The MINDENS model, however, is not considered for the reduced network because it is very dense.

Edge Probabilities

The IPFP probabilities are among the worst in both panels of Figure 7. Similar to the full network, the AUC values for the ROC curve are strongly decreasing with time. An almost parallel pattern can be found for the IPFP-based reconstruction with GDP values (IPFP-GDP). Although the exogenous information helps to improve the performance relative to IPFP, the outcome is still very bad in comparison to the other models.

While the information on the GDP nevertheless improves the fit in the IPFP-based models, this is not the case for the density-corrected gravity model (DC-GRAVITY). It turns out that the version that includes GDP values (DC-GRAVITY-GDP) performs even worse than without (DC-GRAVITY) with both measures. Again, we find that the DC-GRAVITY and the H-FIT model behave very similar.

The two models with lagged variables as covariates, the DC-GRAVITY-LAG model and the IPFP model combined with the lagged values (IPFP-LAG), have the unfair advantage of incorporating much more information than all others and reach outstanding AUC values in both panels of Figure 7 (both lines overlay in the plots). In Figures 30 and 35 it can be seen that the reconstructed network based on the lagged covariates is almost identical to the original one, showing that having observed an edge in $t - 1$ is almost deterministic for predicting an edge in

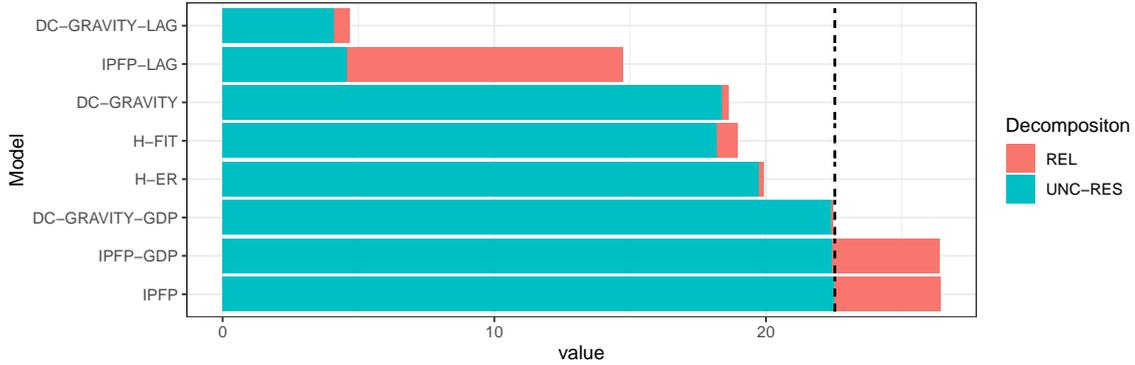


Figure 8: Decomposition of the Brier score for the reduced network into reliability (REL) and remaining uncertainty after subtracting resolution (UNC-RES) for the IPFP model, the degree corrected gravity model (DC-GRAVITY) with covariates (DC-GRAVITY-GDP, DC-GRAVITY-LAG), the hierarchical Erdős-Rényi model (H-ER), the hierarchical fitness model (H-FIT) and the IPFP-based models with covariates (IPFP-GDP, IPFP-LAG). Uncertainty (UNC) as a dashed line.

Source: SWIFT BI Watch.

t.

The decomposition of the Brier score in Figure 8 mirror the results discussed above. However, it is striking that the IPFP-LAG model has a much higher reliability measure in comparison to the DC-GRAVITY-LAG model which results from not being calibrated to the real density. Further note that the three models IPFP, IPFP-GDP and DC-GRAVITY-GDP have a resolution score of almost zero indicating that the knowledge GDP does not contribute much information about the binary edge structure.

Degree Structure

Given that the reduced network is very dense, the degree structures might be more easily reconstructed as compared to the sparse full network. However, the predicted edges still need to be allocated correctly to the corresponding nodes which are not a trivial task. This becomes obvious when regarding the visualization of the degree reconstruction in Supplementary Material. There it can be seen that the reconstruction of the binary degrees is partly very bad. Quantified with the root mean squared errors as shown in Figure 9, the models can be compared directly.

In both panels of Figure 9 it can be seen very clearly that the models that incorporate the lagged matrix entries (IPFP-LAG, DC-GRAVITY-LAG) lead to degree reconstructions that are superior in every respect. Except for some spikes, that might reflect a kind of seasonality pattern, the root mean squared errors are close to zero.

If no information from exogenous covariates is available, the DC-GRAVITY model is found to perform very well for the indegree. Especially regarding the outdegree, almost all methods (amongst others H-FIT, GRAVITY, IPFP) give good and very comparable results.

Again the LASSO proves to be a bad choice for reconstructing the degree structure, exhibiting a high variance over time as well as large deviations from the actual degrees.

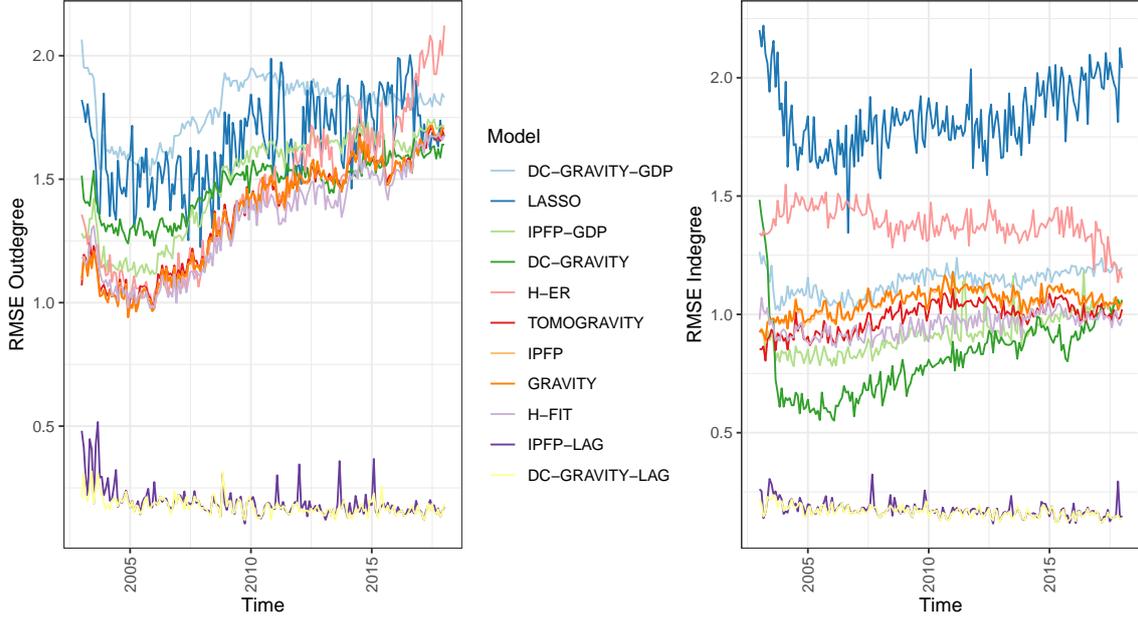


Figure 9: Time series of root mean squared error (RMSE) for the reconstruction of the outdegree (right) and the indegree (left) of the reduced network for the IPFP model, the Gravity model (GRAVITY), the degree corrected gravity model (DC-GRAVITY) with covariates (DC-GRAVITY-GDP, DC-GRAVITY-LAG), the hierarchical Erdős-Rényi model (H-ER), the hierarchical fitness model (H-FIT), the IPFP-based models with covariates (IPFP-GDP, IPFP-LAG), the LASSO model and the TOMOGRAVITY model.

Source: SWIFT BI Watch.

5.2. Valued Network Prediction

The mechanisms that determine the edge probabilities might differ fundamentally from the ones that lead to certain edge values. Additionally, some models are restricted to the prediction of edge values and the prediction of binary networks constructed with threshold values is not the usage they are originally built for. Therefore, we now pay attention to the predictive quality of the valued reconstruction in terms of the L_1 errors

$$L_1^t = \sum_{i \neq j} |x_{t,ij} - \hat{\mu}_{t,ij}|, \text{ for } t = 1, \dots, T$$

and the L_2 errors

$$L_2^t = \sqrt{\sum_{i \neq j} (x_{t,ij} - \hat{\mu}_{t,ij})^2}, \text{ for } t = 1, \dots, T.$$

These measures are regarded in terms of overall errors aggregated over all time points as well as their monthly averages and the corresponding standard errors.

Method	overall L_1	overall L_2	average L_1	SE	average L_2	SE
IPFP	48,443.070	229.987	266.171	68.568	16.225	5.246
GRAVITY	47,971.710	211.125	263.581	68.321	14.886	4.843
DC-GRAVITY	51,595.870	239.899	283.494	70.855	16.917	5.495
LASSO	452,296.700	824.632	2,485.147	778.114	58.177	18.807
H-FIT	67,171.460	479.524	369.074	100.739	33.669	11.425
H-ER	67,282.440	481.793	369.684	101.241	33.810	11.534
MINDENS	90,610.240	459.455	497.858	114.582	33.021	8.361

Table 2: Evaluation of the reconstructed valued full MT 103 networks, Method in the first column. Aggregated L_1 and L_2 errors in columns two and three as well as average errors and their standard errors over time in the last four columns. Minimal values in bold.

Source: SWIFT BI Watch.

Method	overall L_1	overall L_2	average L_1	SE	average L_2	SE
IPFP	39,087.480	216.339	214.766	54.409	15.270	4.912
GRAVITY	38,646.430	203.317	212.343	54.268	14.321	4.709
DC-GRAVITY	41,470.410	232.719	227.859	55.489	16.396	5.378
TOMOGRAVITY	39,691.210	215.081	218.084	55.635	15.182	4.878
LASSO	290,061.700	1,042.029	1,593.746	486.963	73.577	23.569
H-FIT	55,979.590	471.112	307.580	83.433	33.426	10.136
H-ER	56,071.170	471.710	308.083	83.580	33.460	10.178
DC-GRAVITY-GDP	39,759.120	232.127	218.457	53.322	16.337	5.415
DC-GRAVITY-LAG	40,110.850	232.127	220.389	53.532	16.351	5.373
IPFP-GDP	38,231.110	214.741	210.061	54.639	15.018	5.290
IPFP-LAG	4,137.484	34.178	22.859	11.523	2.128	1.391

Table 3: Evaluation of the reconstructed valued reduced MT 103 networks, Method in the first column. Aggregated L_1 and L_2 errors in columns two and three as well as average errors and their standard errors over time in the last four columns. The last four rows give models with exogenous information included. Minimal values in bold.

Source: SWIFT BI Watch.

5.2.1. Full Network

In Table 2, it can be seen that the two dense reconstruction models IPFP and GRAVITY give the best reconstruction evaluated with the L_1 and L_2 errors with the GRAVITY model being slightly ahead. The third-best prediction quality is delivered by the DC-GRAVITY model. It can be inferred that the risk of guessing the wrong edges to be zero or one (and placing a high weight or no weight to the false edges) strongly counterweights the seeming disadvantage of the dense reconstruction methods. This effect is pronounced in the MINDENS model and even more so in the LASSO model that comes with extremely high errors. However, also the hierarchical fitness model (H-FIT), one of the best models for binary network reconstruction is found to provide edge value predictions that are by far worse compared to the GRAVITY solution.

5.2.2. Reduced Network

In the reduced network, we conclude that IPFP, the GRAVITY model and the TOMOGRAVITY model results in very similar aggregated L_1 and L_2 errors. These models are closely followed by the DC-GRAVITY-GDP model. The hierarchical models (H-FIT, G-ER) perform comparable and by far better than the LASSO.

Models that include exogenous information are separated and given in the last four rows of Table 3. Among these models, the second-best result is given by the IPFP-GDP model, showing that the GDP values provide useful information that improves the quality of the edge value reconstruction, for example, relative to IPFP or the GRAVITY model. The IPFP model that incorporates lagged edge values (IPFP-LAG) as covariates performs outstandingly well. But again it might be unrealistic to assume the availability of lagged data points. Interestingly, both density-corrected gravity models with exogenous covariates (DC-GRAVITY-GDP, DC-GRAVITY-LAG) are only slightly better than the DC-GRAVITY model. This can be explained by the fact, that the exogenous information is only used to determine the edge probabilities.

6. Discussion

In this paper, we have compared different models for network construction using the SWIFT MT 103 networks. The models are compared along different dimensions, including the accuracy of edge prediction, degree reconstruction, and edge value estimation. Overall, four conclusions that can be drawn from this competitive comparison.

(i) The task of reconstructing edge values differs fundamentally from the task of estimating edge probabilities. Technically, this is very intuitive because the marginals give exclusively information about the edge values and all approaches that output edge probabilities are necessarily dependent on further restrictions (the real density).

Even if the true density is assumed to be known, no model emerged that can be said to be great in achieving outstanding predictions of the edge probabilities and their values. This conclusion is also in line with the findings of the extensive comparison by Anand et al. (2018). We, therefore, recommend that the model choice should be governed by the specific use case and depending on the importance attached to either reconstruction. If the binary structure is of interest and the model is presumed to be sparse, the hierarchical fitness model (H-FIT) and the density-corrected gravity model (DC-GRAVITY) are good choices. While Anand et al. (2018) highlight the ability of the minimum density method (MINDENS) to detect absent edges we must supplement this by noting that the method nevertheless performs not that good if interest lies in detecting present edges.

Regarding the quality of the edge value prediction, either in sparse or dense networks the maximum entropy models (IPFP and GRAVITY) work very well. The same was found by Anand et al. (2018), pointing on the good quality of maximum entropy solutions. However, in contrast to their findings, we highlight here more clearly the potential shortfalls for sparse reconstruction methods for the prediction of edge values.

(ii) Other than Anand et al. (2018), we do not find that the preferred models change when either a dense or a sparse network is to be reconstructed. However, this statement must be taken with care since in our analysis the dense network is, in fact, a subset of the sparse one.

(iii) Including exogenous information can help to improve both, the binary and the valued network reconstruction and partly leads to dramatic increases in the predicted performance. However, this increase in predictive accuracy is not guaranteed. If variables with a low association to the unknown edge values are chosen, the quality of the reconstruction might even

decline (see also Lebacher and Kauermann, 2019). Especially regarding the binary network reconstruction, the inclusion of GDP led to mixed results.

(iv) As an “off the shelf” model in situations without exogenous information available, the density-corrected gravity model (DC-GRAVITY) can be recommended because it is found to work well on the big sparse network as well as on the small dense network with respect to the edge probabilities and the edge values. A similar conclusion can be found in Anand et al. (2018, p. 116), stating that among the probabilistic methods the model is the “*clear winner across all measures of interest*”. Similarly, Gandy and Veraart (2019) report that this model is performing very well in binary and valued reconstruction. Further, the model can be extended towards the inclusion of exogenous information in a simple way.

For further research, it seems to be necessary to compare the performance of edge probabilities when using calibration densities that differ from the real one. Another important research question relates to the ability of reconstruction models to provide uncertainty quantification. Many approaches introduced above results in network ensembles or come with an associated stochastic structure that can be used to construct prediction intervals.

Acknowledgement

We would like to thank Peter Ware and Nancy Murphy for providing the SWIFT data.

Declaration of Interest

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: *International Trade of Arms: A Network Approach*.

References

- AIROLDI, E. M. AND A. W. BLOCKER (2013): “Estimating latent processes on a network from indirect measurements,” *Journal of the American Statistical Association*, 108, 149–164.
- ALBERT, R. AND A.-L. BARABÁSI (2002): “Statistical mechanics of complex networks,” *Rev. Mod. Phys.*, 74, 47–97.
- ANAND, K., B. CRAIG, AND G. VON PETER (2015): “Filling in the blanks: Network structure and interbank contagion,” *Quantitative Finance*, 15, 625–636.
- ANAND, K., I. VAN LELYVELD, ÁDÁM BANAI, S. FRIEDRICH, R. GARRATT, G. HALAJ, J. FIQUE, I. HANSEN, S. M. JARAMILLO, H. LEE, J. L. MOLINA-BORBOA, S. NOBILI, S. RAJAN, D. SALAKHOVA, T. C. SILVA, L. SILVESTRI, AND S. R. S. DE SOUZA (2018): “The missing links: A global study on uncovering financial network structures from partial data,” *Journal of Financial Stability*, 35, 107 – 119.
- BAEK, S., K. SORAMAKI, AND J. YOON (2014): “Network indicators for monitoring intraday liquidity in bok-wire+,” *Bank of Korea Working Paper*, 1.
- BARABÁSI, A.-L. AND R. ALBERT (1999): “Emergence of Scaling in Random Networks,” *Science*, 286, 509–512.

- BARDOSCIA, M., S. BATTISTON, F. CACCIOLI, AND G. CALDARELLI (2017): “Pathways towards instability in financial networks,” *Nature Communications*, 8, 14416.
- BARGIGLI, L. (2014): “Statistical ensembles for economic networks,” *Journal of Statistical Physics*, 155, 810–825.
- BATTISTON, S., J. D. FARMER, A. FLACHE, D. GARLASCHELLI, A. G. HALDANE, H. HEESTERBEEK, C. HOMMES, C. JAEGER, R. MAY, AND M. SCHEFFER (2016): “Complexity theory and financial regulation,” *Science*, 351, 818–819.
- BATTISTON, S., M. PULIGA, R. KAUSHIK, P. TASCA, AND G. CALDARELLI (2012): “DebtRank: Too central to fail? Financial networks, the Fed and systemic risk,” *Scientific reports*, 2, 541.
- BILLIO, M., M. GETMANSKY, A. W. LO, AND L. PELIZZON (2012): “Econometric measures of connectedness and systemic risk in the finance and insurance sectors,” *Journal of Financial Economics*, 104, 535–559.
- BISHOP, Y. M., P. W. HOLLAND, AND S. E. FIENBERG (1975): *Discrete multivariate analysis: theory and practice*, Cambridge: MIT Press.
- BLOCKER, A. W., P. KOULLICK, AND E. AIROLDI (2014): “networkTomography: Tools for network tomography,” R package version 0.3.
- CACCIOLI, F., P. BARUCCA, AND T. KOBAYASHI (2018): “Network models of financial systemic risk: A review,” *Journal of Computational Social Science*, 1, 81–114.
- CASTRO, R., M. COATES, G. LIANG, R. NOWAK, AND B. YU (2004): “Network tomography: Recent developments,” *Statistical Science*, 19, 499–517.
- CHEN, L., X. MA, G. PAN, J. JAKUBOWICZ, ET AL. (2017): “Understanding bike trip patterns leveraging bike sharing system open data,” *Frontiers of computer science*, 11, 38–48.
- CHINAZZI, M., G. FAGIOLO, J. A. REYES, AND S. SCHIAVO (2013): “Post-mortem examination of the international financial network,” *Journal of Economic Dynamics and Control*, 37, 1692–1713.
- CIMINI, G., T. SQUARTINI, D. GARLASCHELLI, AND A. GABRIELLI (2015): “Systemic risk analysis on reconstructed economic and financial networks,” *Scientific reports*, 5, 15758.
- COOK, S. AND K. SORAMAKI (2014): “The global network of payment flows,” *SWIFT Institute Working Paper*.
- DEMING, W. E. AND F. F. STEPHAN (1940): “On a least squares adjustment of a sampled frequency table when the expected marginal totals are known,” *The Annals of Mathematical Statistics*, 11, 427–444.
- DISDIER, A.-C. AND K. HEAD (2008): “The puzzling persistence of the distance effect on bilateral trade,” *The Review of Economics and Statistics*, 90, 37–48.
- ELSINGER, H., A. LEHAR, AND M. SUMMER (2013): “Network models and systemic risk assessment,” in *Handbook on Systemic Risk*, ed. by J.-P. Fouque and J. A. Langsam, Cambridge: Cambridge University Press, chap. IV, 287–305.

- ERDÖS, P. AND A. RÉNYI (1959): “On Random Graphs I,” *Publicationes Mathematicae Debrecen*, 6, 290.
- FIENBERG, S. E. ET AL. (1970): “An iterative procedure for estimation in contingency tables,” *The Annals of Mathematical Statistics*, 41, 907–917.
- FRIEDMAN, J., T. HASTIE, AND R. TIBSHIRANI (2009): “glmnet: LASSO and elastic-net regularized generalized linear models,” R package version 2.0-1.6.
- GAI, P. AND S. KAPADIA (2010): “Contagion in financial networks,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466, 2401–2423.
- GANDY, A. AND L. A. VERAART (2017): “A Bayesian methodology for systemic risk assessment in financial networks,” *Management Science*, 63, 4428–4446.
- GANDY, A. AND L. A. M. VERAART (2019): “Adjustable network reconstruction with applications to CDS exposures,” *Journal of Multivariate Analysis*, 172, 193 – 209.
- GRAU, J., I. GROSSE, AND J. KEILWAGEN (2015): “PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R,” *Bioinformatics*, 31, 2595–2597.
- HABERMAN, S. J. (1978): *Analysis of qualitative data 1: Introductory topics*, New York: Academic Press.
- (1979): *Analysis of qualitative data 2: New Developments*, New York: Academic Press.
- HEAD, K. AND T. MAYER (2014): “Gravity equations: Workhorse, toolkit, and cookbook,” in *Handbook of international economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Amsterdam: Elsevier Science Publishing, vol. 4, 131–195.
- IMAKUBO, K., Y. SOEJIMA, ET AL. (2010): “The transaction network in Japan’s interbank money markets,” *Monetary and Economic Studies*, 28, 107–150.
- KAUÊ DAL’MASO PERON, T., L. DA FONTOURA COSTA, AND F. A. RODRIGUES (2012): “The structure and resilience of financial market networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22, 013117.
- KOLACZYK, E. D. (2009): *Statistical analysis of network data. Methods and Models*, New York: Springer.
- LEBACHER, M. AND G. KAUEMANN (2019): “Regression-based network reconstruction with codal and dyadic covariates and random effects,” *arXiv preprint arXiv:1903.11886*.
- LIU, Z., S. QUIET, AND B. ROTH (2015): “Banking sector interconnectedness: what is it, how can we measure it and why does it matter?” *Bank of England Quarterly Bulletin*, Q2.
- MARKOSE, S., S. GIANANTE, AND A. R. SHAGHAGHI (2012): “Too interconnected to fail - financial network of US CDS market: Topological fragility and systemic risk,” *Journal of Economic Behavior & Organization*, 83, 627–646.
- MASTRANDREA, R., T. SQUARTINI, G. FAGIOLO, AND D. GARLASCHELLI (2014): “Enhanced reconstruction of weighted networks from strengths and degrees,” *New Journal of Physics*, 16, 043022.

- MURPHY, A. H. (1973): “A new vector partition of the probability score,” *Journal of applied Meteorology*, 12, 595–600.
- NIE, L., D. JIANG, AND Z. LV (2017): “Modeling network traffic for traffic matrix estimation and anomaly detection based on Bayesian network in cloud computing networks,” *Annals of Telecommunications*, 72, 297–305.
- SCHWEITZER, F., G. FAGIOLO, D. SORNETTE, F. VEGA-REDONDO, A. VESPIGNANI, AND D. R. WHITE (2009): “Economic networks: The new challenges,” *Science*, 325, 422–425.
- SCRUCCA, L. (2013): “GA: A package for genetic algorithms in R,” *Journal of Statistical Software*, 53, 1–37.
- SHELDON, G. AND M. MAURER (1998): “Interbank lending and systemic risk: An empirical analysis for Switzerland,” *Swiss Journal of Economics and Statistics*, 134, 685–704.
- SIEGERT, S. (2017): “Simplifying and generalising Murphy’s Brier score decomposition,” *Quarterly Journal of the Royal Meteorological Society*, 143, 1178–1183.
- SORAMÄKI, K., M. L. BECH, J. ARNOLD, R. J. GLASS, AND W. E. BEYELER (2007): “The topology of interbank payment flows,” *Physica A: Statistical Mechanics and its Applications*, 379, 317–333.
- SORAMÄKI, K. AND S. COOK (2013): “SinkRank: An algorithm for identifying systemically important banks in payment systems,” *Economics: The Open-Access, Open-Assessment E-Journal*, 7, 1–27.
- SQUARTINI, T., G. CALDARELLI, G. CIMINI, A. GABRIELLI, AND D. GARLASCHELLI (2018): “Reconstruction methods for networks: The case of economic and financial systems,” *Physics Reports*, 757, 1 – 47.
- THURNER, S. AND S. POLEDNA (2013): “DebtRank-transparency: Controlling systemic risk in financial networks,” *Scientific reports*, 3, 1888.
- TIBSHIRANI, R. (1996): “Regression shrinkage and selection via the LASSO,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58, 267–288.
- UPPER, C. (2011): “Simulation methods to assess the danger of contagion in interbank markets,” *Journal of Financial Stability*, 7, 111–125.
- VARDI, Y. (1996): “Network tomography: Estimating source-destination traffic intensities from link data,” *Journal of the American Statistical Association*, 91, 365–377.
- ZHANG, Y., M. ROUGHAN, N. DUFFIELD, AND A. GREENBERG (2003a): “Fast accurate computation of large-scale IP traffic matrices from link loads,” in *ACM SIGMETRICS Performance Evaluation Review*, ACM, vol. 31, 206–217.
- ZHANG, Y., M. ROUGHAN, C. LUND, AND D. DONOHO (2003b): “An information-theoretic approach to traffic matrix estimation,” in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, Association for Computing Machinery, 301–312.
- ZHOU, H., L. TAN, Q. ZENG, AND C. WU (2016): “Traffic matrix estimation: A neural network approach with extended input and expectation maximization iteration,” *Journal of Network and Computer Applications*, 60, 220 – 232.

A. Countries included

ISO	name	reduced	ISO	name	reduced	ISO	name	reduced
AD	Andorra	0	GQ	Equatorial Guinea	0	PE	Peru	0
AE	United Arab Emirates	1	GR	Greece	1	PF	French Polynesia	0
AG	Antigua & Barbuda	0	GT	Guatemala	0	PG	Papua New Guinea	0
AI	Anguilla	0	GY	Guyana	0	PH	Philippines	1
AL	Albania	0	HK	Hong Kong SAR China	1	PK	Pakistan	0
AM	Armenia	0	HN	Honduras	0	PL	Poland	1
AO	Angola	0	HR	Croatia	1	PR	Puerto Rico	0
AR	Argentina	0	HT	Haiti	0	PS	Palestinian Territories	0
AT	Austria	1	HU	Hungary	1	PT	Portugal	1
AU	Australia	1	ID	Indonesia	1	PY	Paraguay	0
AW	Aruba	0	IE	Ireland	1	QA	Qatar	0
AZ	Azerbaijan	0	IL	Israel	1	RE	Réunion	0
BA	Bosnia & Herzegovina	0	IM	Isle of Man	0	RO	Romania	1
BB	Barbados	0	IMI	Internat. Market Infrastruct.	0	RU	Russia	1
BD	Bangladesh	0	IN	India	1	RW	Rwanda	0
BE	Belgium	1	IR	Iran	0	SA	Saudi Arabia	1
BF	Burkina Faso	0	IS	Iceland	0	SB	Solomon Islands	0
BG	Bulgaria	1	IT	Italy	1	SC	Seychelles	0
BH	Bahrain	0	JE	Jersey	0	SD	Sudan	0
BI	Burundi	0	JM	Jamaica	0	SE	Sweden	1
BJ	Benin	0	JO	Jordan	0	SG	Singapore	1
BM	Bermuda	0	JP	Japan	1	SI	Slovenia	1
BN	Brunei	0	KE	Kenya	0	SK	Slovakia	1
BO	Bolivia	0	KG	Kyrgyzstan	0	SL	Sierra Leone	0
BR	Brazil	1	KH	Cambodia	0	SM	San Marino	0
BS	Bahamas	0	KN	St. Kitts & Nevis	0	SN	Senegal	0
BW	Botswana	0	KR	South Korea	1	SR	Suriname	0
BY	Belarus	1	KW	Kuwait	1	SV	El Salvador	0
BZ	Belize	0	KY	Cayman Islands	0	SY	Syria	0
CA	Canada	1	KZ	Kazakhstan	1	TC	Turks & Caicos Islands	0
CF	Central African Republic	0	LA	Laos	0	TG	Togo	0
CH	Switzerland	1	LB	Lebanon	0	TH	Thailand	1
CI	Côte d'Ivoire	0	LC	St. Lucia	0	TJ	Tajikistan	0
CL	Chile	0	LI	Liechtenstein	0	TL	Timor-Leste	0
CM	Cameroon	0	LK	Sri Lanka	0	TM	Turkmenistan	0
CN	China	1	LS	Lesotho	0	TN	Tunisia	0
CO	Colombia	0	LT	Lithuania	1	TO	Tonga	0
CR	Costa Rica	0	LU	Luxembourg	1	TR	Turkey	1
CU	Cuba	0	LV	Latvia	1	TT	Trinidad & Tobago	0
CV	Cape Verde	0	LY	Libya	0	TW	Taiwan	1
CY	Cyprus	1	MA	Morocco	0	TZ	Tanzania	0
CZ	Czechia	1	MC	Monaco	0	UA	Ukraine	1
DE	Germany	1	MD	Moldova	0	UG	Uganda	0
DJ	Djibouti	0	MG	Madagascar	0	US	United States	1
DK	Denmark	1	MK	Macedonia	0	UY	Uruguay	0
DM	Dominica	0	ML	Mali	0	UZ	Uzbekistan	0
DO	Dominican Republic	0	MN	Mongolia	0	VC	St. Vincent & Grenadines	0
DZ	Algeria	0	MO	Macau SAR China	0	VE	Venezuela	0
EC	Ecuador	0	MR	Mauritania	0	VG	British Virgin Islands	0
EE	Estonia	1	MS	Montserrat	0	VI	U.S. Virgin Islands	0
EG	Egypt	0	MT	Malta	0	VN	Vietnam	1
ES	Spain	1	MU	Mauritius	0	VU	Vanuatu	0
ET	Ethiopia	0	MV	Maldives	0	WS	Samoa	0
FI	Finland	1	MW	Malawi	0	YE	Yemen	0
FJ	Fiji	0	MX	Mexico	1	YT	Mayotte	0
FO	Faroe Islands	0	MY	Malaysia	1	ZA	South Africa	1
FR	France	1	MZ	Mozambique	0	ZM	Zambia	0
GA	Gabon	0	NA	Namibia	0	ZW	Zimbabwe	0
GB	United Kingdom	1	NC	New Caledonia	0	GF	French Guiana	0
GD	Grenada	0	NE	Niger	0	KI	Kiribati	0
GE	Georgia	0	NG	Nigeria	1	CD	Congo - Kinshasa	0
GG	Guernsey	0	NI	Nicaragua	0	CG	Congo - Brazzaville	0
GH	Ghana	0	NL	Netherlands	1	MQ	Martinique	0
GI	Gibraltar	0	NO	Norway	1	SZ	Swaziland	0
GL	Greenland	0	NP	Nepal	0	CK	Cook Islands	0
GM	Gambia	0	NZ	New Zealand	1	VA	Holy See	0
GN	Guinea	0	OM	Oman	0	TD	Chad	0
GP	Guadeloupe	0	PA	Panama	1			

Table 4: Countries included in the analysis with ISO 2 country code (ISO), name of the country (name) and occurrence in the small MT 103 network set (reduced=1).

Source: SWIFT BI Watch.

B. Descriptives for the reduced data set

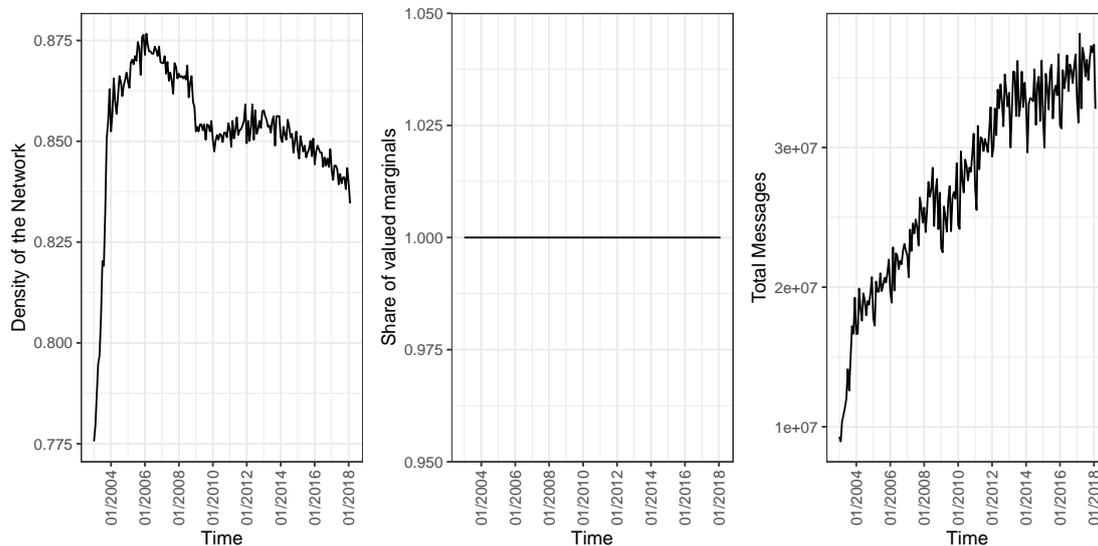


Figure 10: Summary statistics for the reduced MT 103 network as monthly time series. Density of the network (left), share of non-zero marginals (middle) and cumulative edge values (right). Source: SWIFT BI Watch.

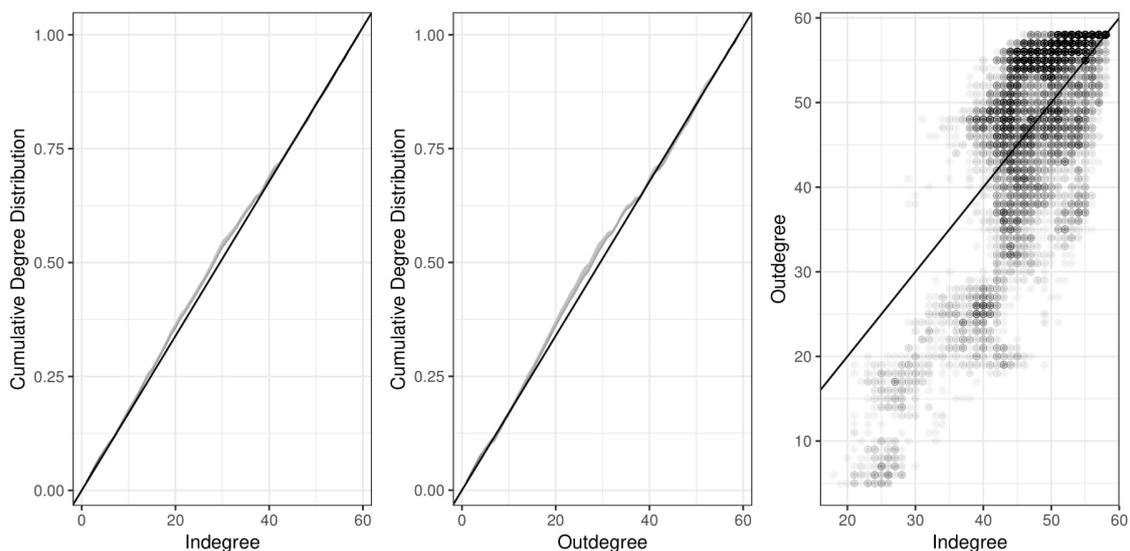


Figure 11: Binary network topology of the reduced MT 103 network aggregated for all time points. Cumulative indegree (left) and outdegree distributions (middle) with maximum and minimum values indicated in grey. Outdegree against indegree (right) for all months in dotted with colour intensity by frequency. 45 degree line in solid black. Source: SWIFT BI Watch.

C. Binary Reconstruction: Full Network

C.1. Predicted Adjacency Matrices: Full Network

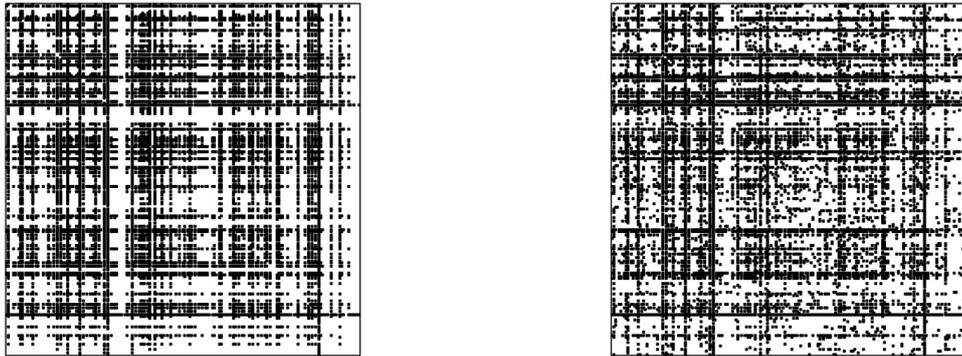


Figure 12: Adjacency matrices, representing the full MT 103 network in February 2018. IFPF reconstruction (left), real network (right).

Source: SWIFT BI Watch.

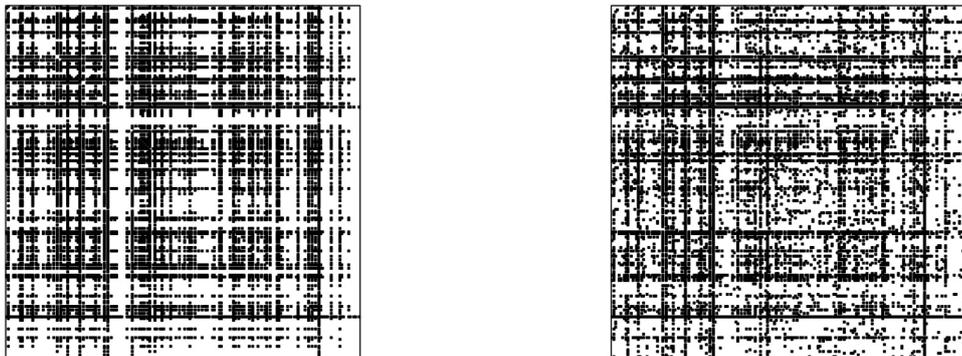


Figure 13: Adjacency matrices, representing the full MT 103 network in February 2018. GRAVITY reconstruction (left), real network (right).

Source: SWIFT BI Watch.

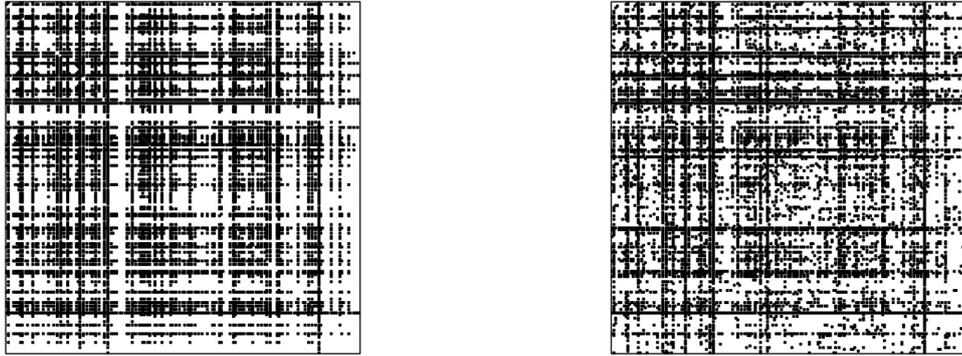


Figure 14: Adjacency matrices, representing the full MT 103 network in February 2018. DC-GRAVITY reconstruction (left), real network (right).
Source: SWIFT BI Watch.

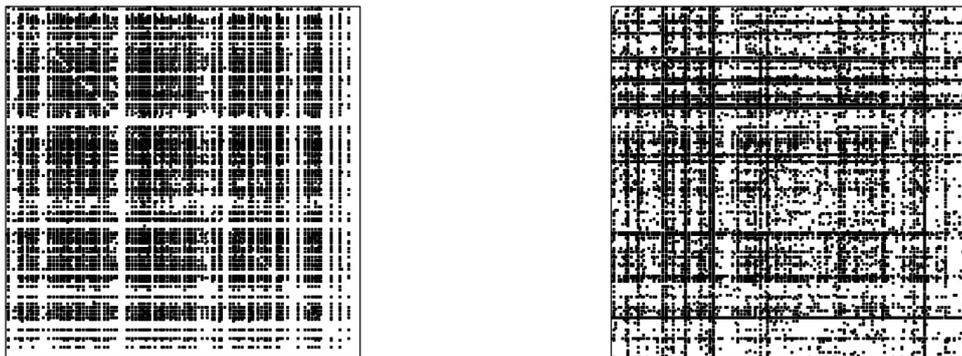


Figure 15: Adjacency matrices, representing the full MT 103 network in February 2018. H-ER reconstruction (left), real network (right).
Source: SWIFT BI Watch.

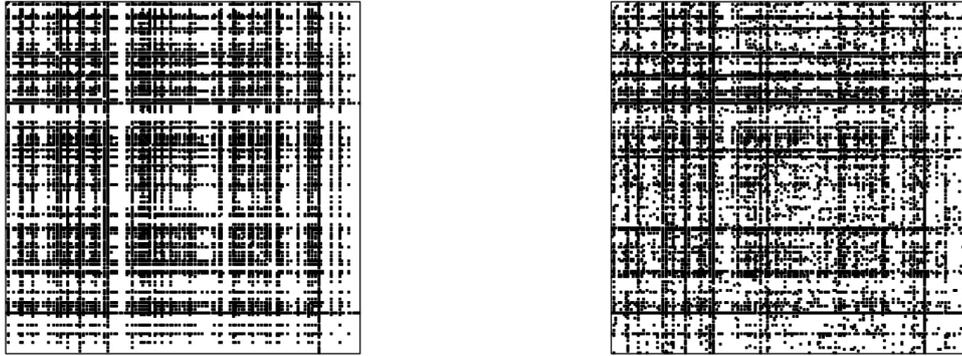


Figure 16: Adjacency matrices, representing the full MT 103 network in February 2018. H-FIT reconstruction (left), real network (right).

Source: SWIFT BI Watch.

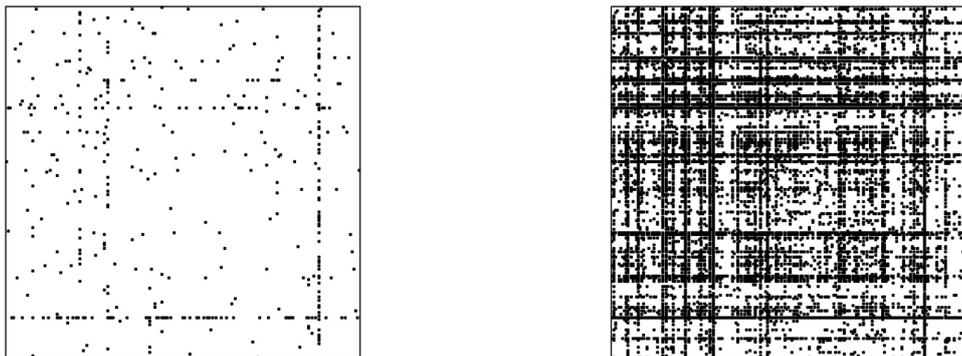


Figure 17: Adjacency matrices, representing the full MT 103 network in February 2018. MIN-DENS reconstruction (left), real network (right).

Source: SWIFT BI Watch.

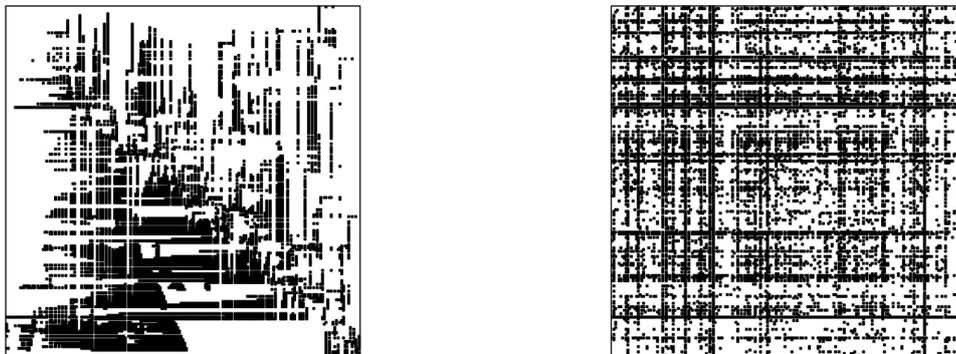


Figure 18: Adjacency matrices, representing the full MT 103 network in February 2018. LASSO reconstruction (left), real network (right).
Source: SWIFT BI Watch.

C.2. Degree Reconstruction: Full Network

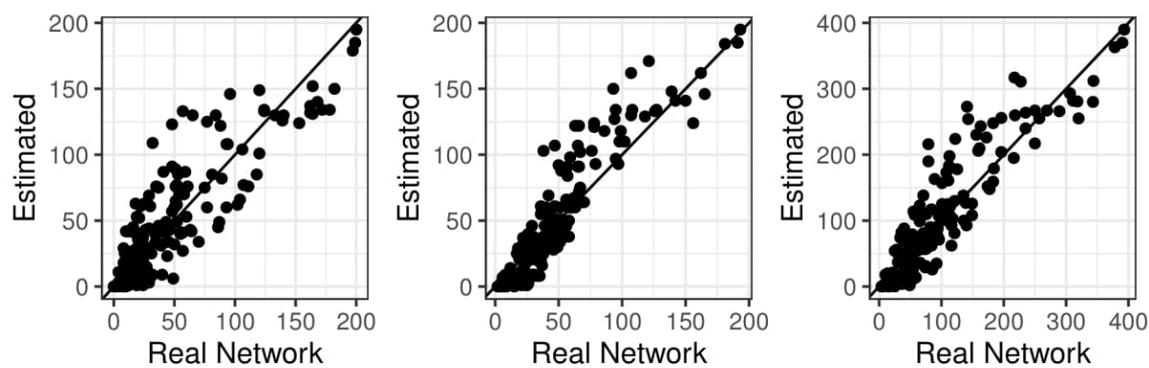


Figure 19: Degree Reconstruction in the full MT 103 network in February 2018. IPFP reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right).
Source: SWIFT BI Watch.

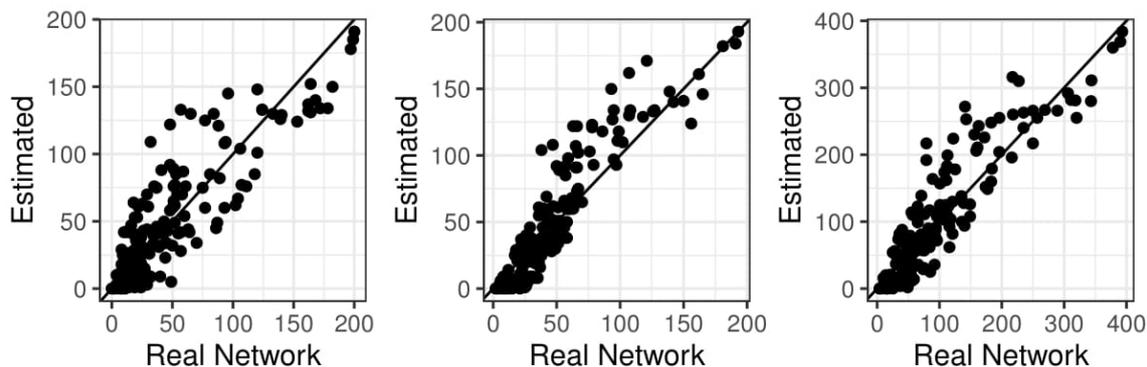


Figure 20: Degree Reconstruction in the full MT 103 network in February 2018. GRAVITY reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

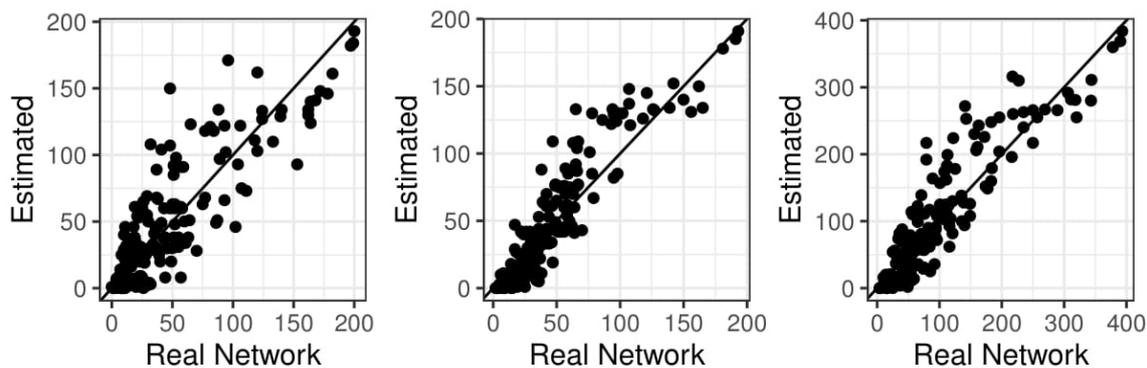


Figure 21: Degree Reconstruction in the full MT 103 network in February 2018. DC-GRAVITY reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

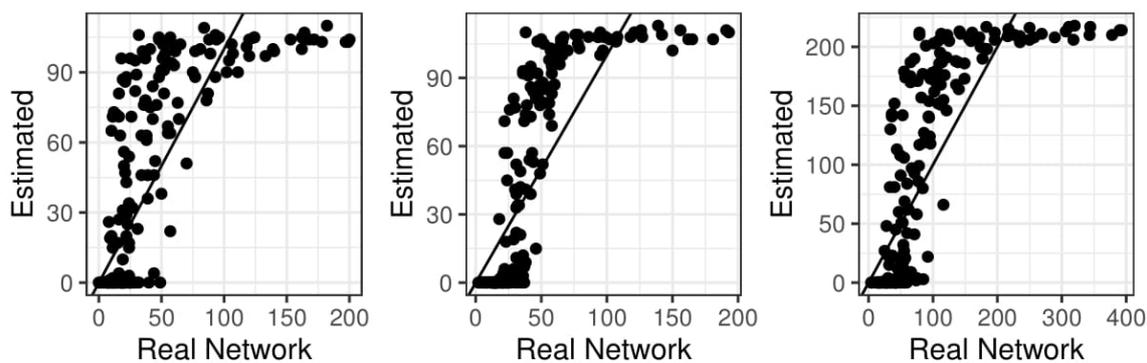


Figure 22: Degree Reconstruction in the full MT 103 network in February 2018. H-ER reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

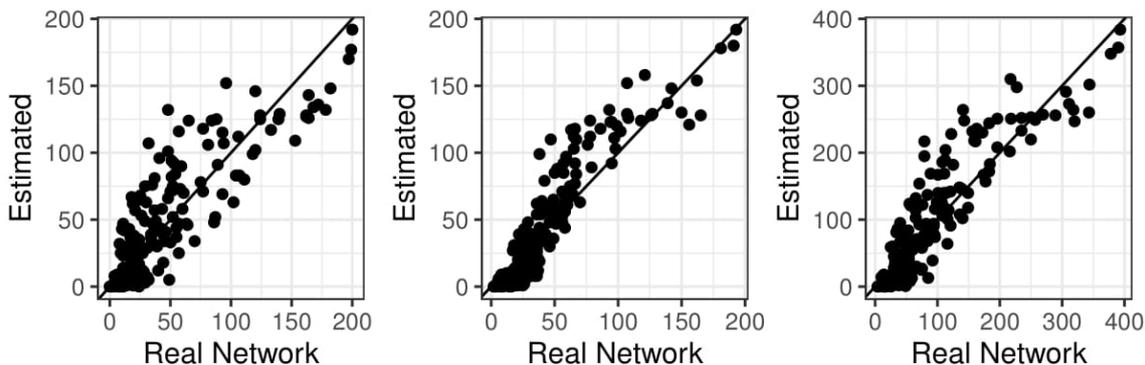


Figure 23: Degree Reconstruction in the full MT 103 network in February 2018. H-FIT reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

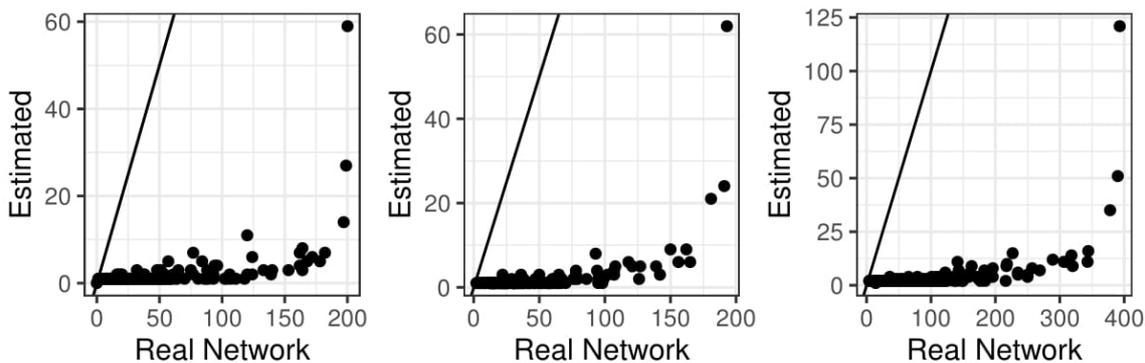


Figure 24: Degree Reconstruction in the full MT 103 network in February 2018. MINDENS reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

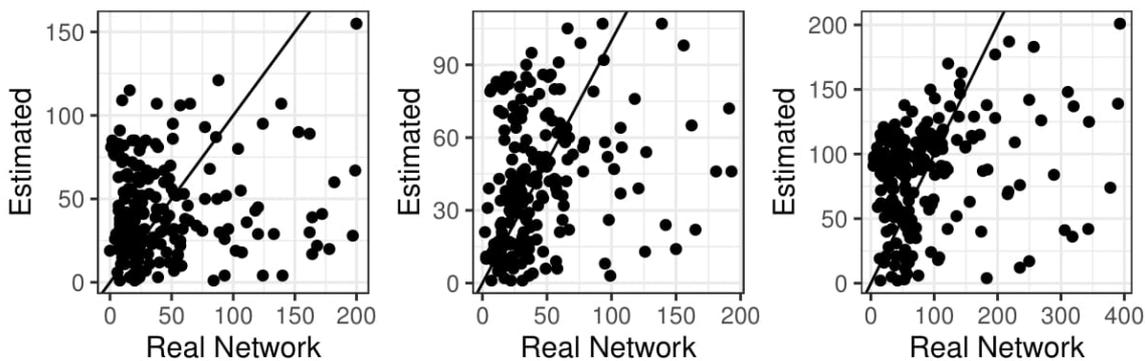


Figure 25: Degree Reconstruction in the full MT 103 network in February 2018. LASSO reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

D. Binary Reconstruction: Reduced Network

D.1. Predicted Adjacency Matrices: Reduced Network



Figure 26: Adjacency matrices, representing the reduced MT 103 network in February 2018. IPFP reconstruction (left), real network (right).

Source: SWIFT BI Watch.



Figure 27: Adjacency matrices, representing the reduced MT 103 network in February 2018. GRAVITY reconstruction (left), real network (right).

Source: SWIFT BI Watch.

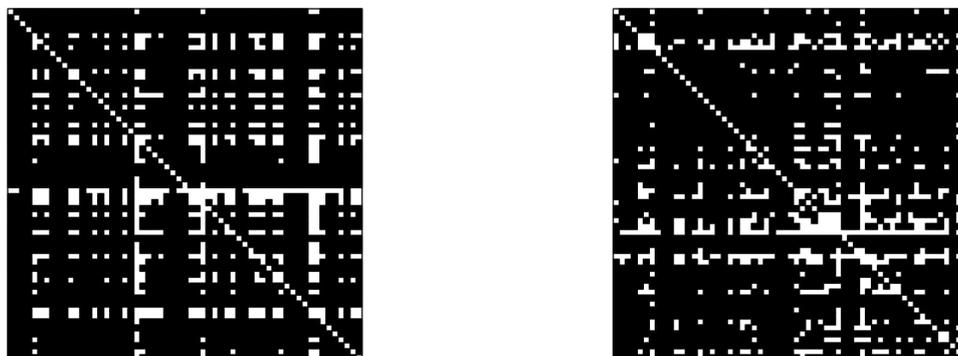


Figure 28: Adjacency matrices, representing the reduced MT 103 network in February 2018. DC-GRAVITY reconstruction (left), real network (right).

Source: SWIFT BI Watch.



Figure 29: Adjacency matrices, representing the reduced MT 103 network in February 2018. DC-GRAVITY-GDP reconstruction (left), real network (right).

Source: SWIFT BI Watch.



Figure 30: Adjacency matrices, representing the reduced MT 103 network in February 2018. DC-GRAVITY-LAG reconstruction (left), real network (right).
Source: SWIFT BI Watch.

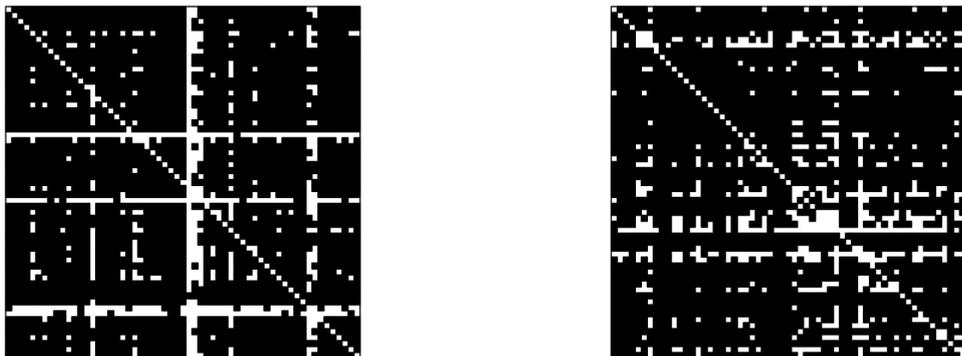


Figure 31: Adjacency matrices, representing the reduced MT 103 network in February 2018. H-ER reconstruction (left), real network (right).
Source: SWIFT BI Watch.



Figure 32: Adjacency matrices, representing the reduced MT 103 network in February 2018. H-FIT reconstruction (left), real network (right).
Source: SWIFT BI Watch.



Figure 33: Adjacency matrices, representing the reduced MT 103 network in February 2018. LASSO reconstruction (left), real network (right).
Source: SWIFT BI Watch.



Figure 34: Adjacency matrices, representing the reduced MT 103 network in February 2018. IPFP-GDP reconstruction (left), real network (right).

Source: SWIFT BI Watch.



Figure 35: Adjacency matrices, representing the reduced MT 103 network in February 2018. IPFP-LAG reconstruction (left), real network (right).

Source: SWIFT BI Watch.

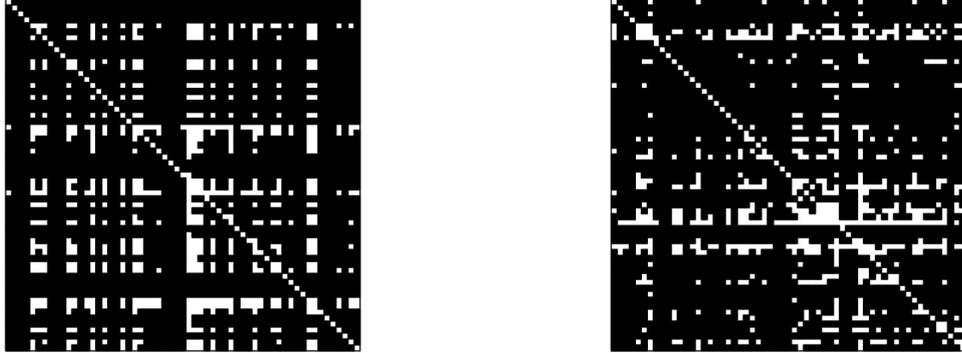


Figure 36: Adjacency matrices, representing the reduced MT 103 network in February 2018. TOMOGRAPHY reconstruction (left), real network (right). Source: SWIFT BI Watch.

D.2. Degree Reconstruction: Reduced Network

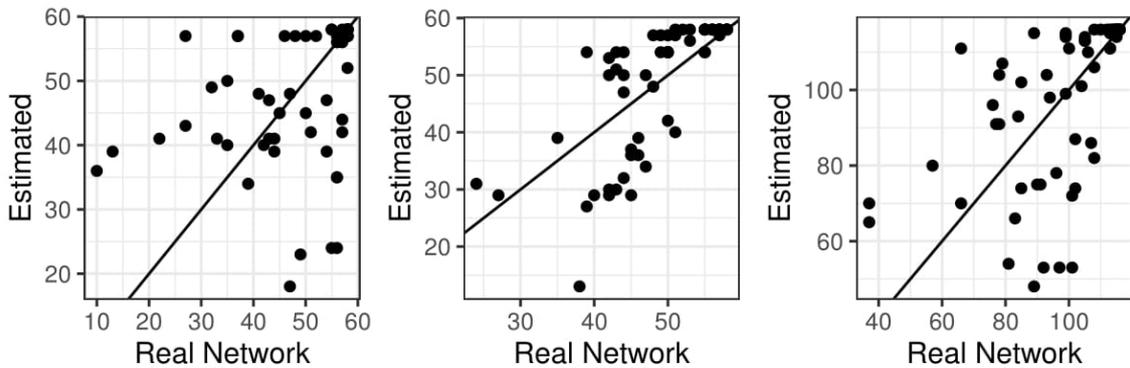


Figure 37: Degree Reconstruction in the reduced MT 103 network in February 2018. IPFP reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

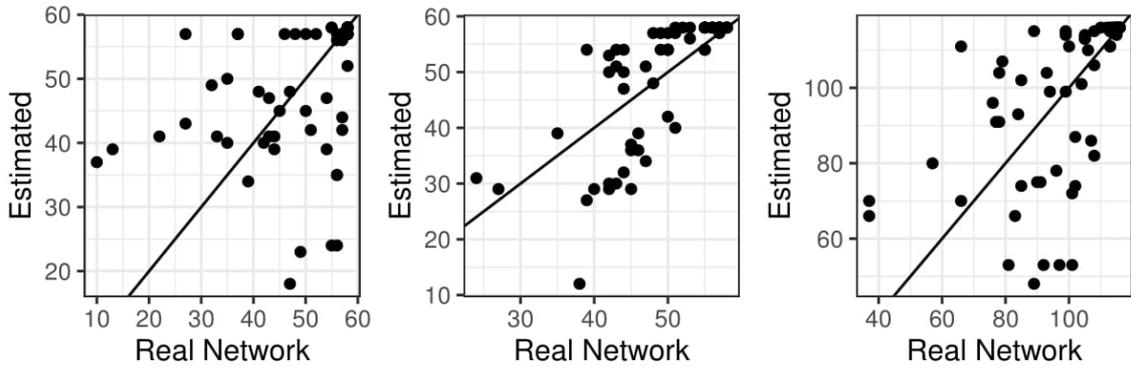


Figure 38: Degree Reconstruction in the reduced MT 103 network in February 2018. GRAVITY reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

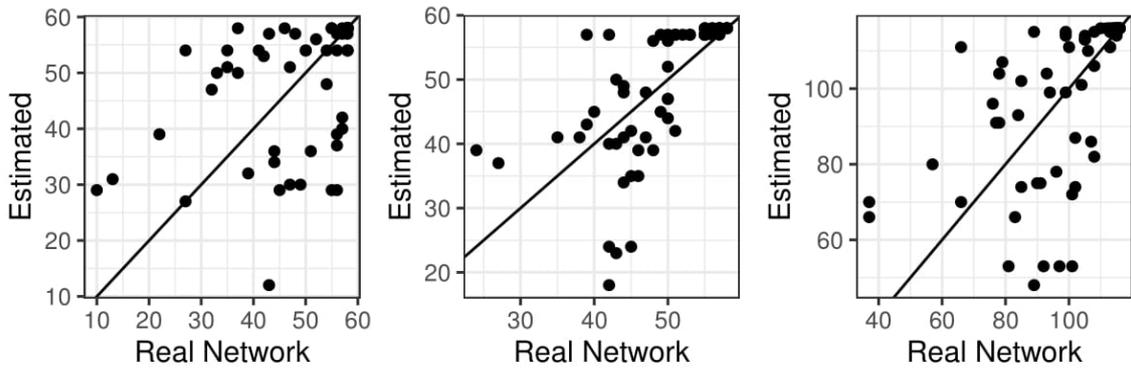


Figure 39: Degree Reconstruction in the reduced MT 103 network in February 2018. DC-GRAVITY reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

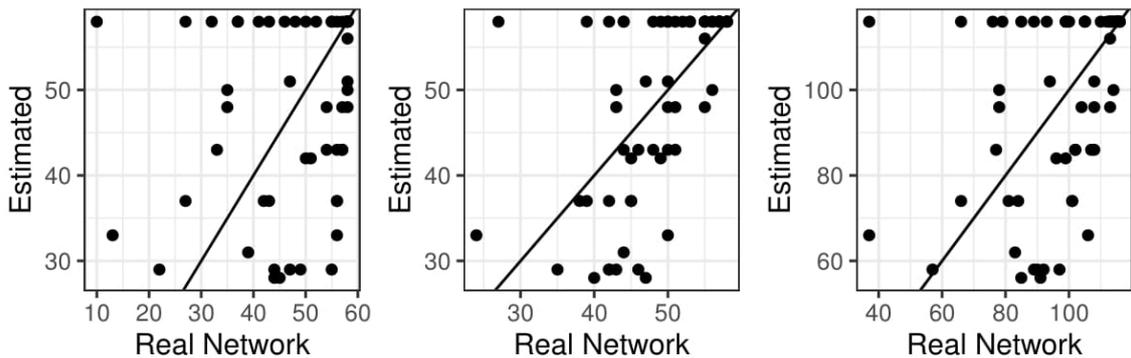


Figure 40: Degree Reconstruction in the reduced MT 103 network in February 2018. DC-GRAVITY-GDP reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

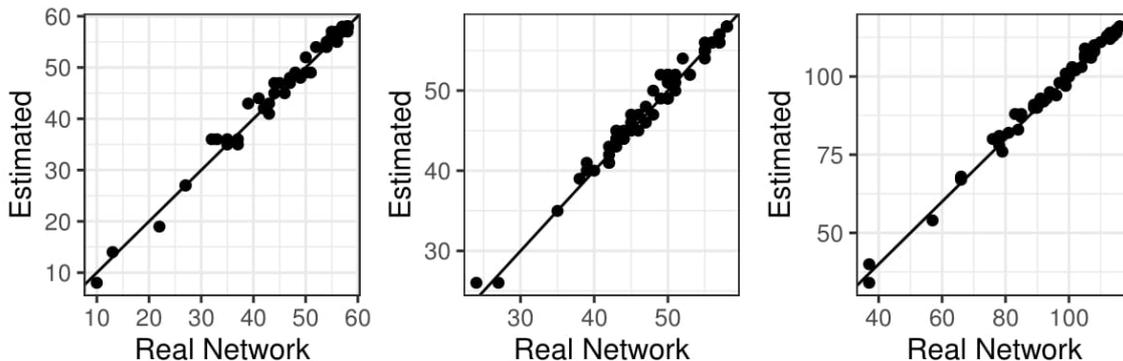


Figure 41: Degree Reconstruction in the reduced MT 103 network in February 2018. DC-GRAVITY-LAG reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right).
Source: SWIFT BI Watch.

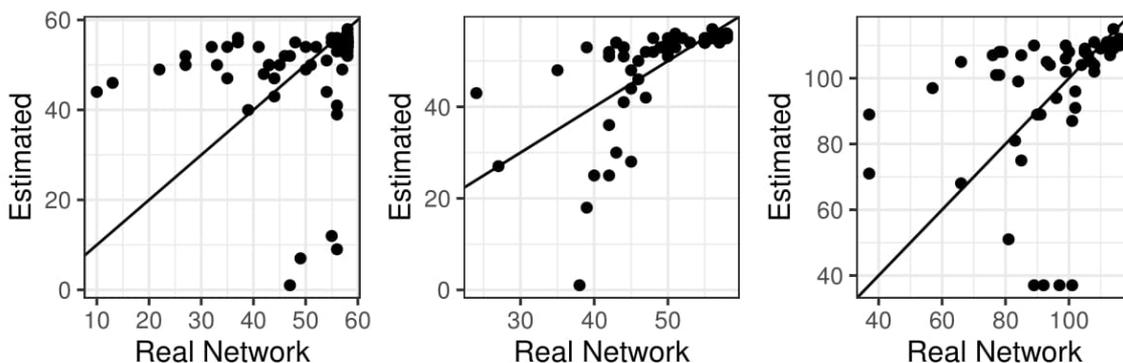


Figure 42: Degree Reconstruction in the reduced MT 103 network in February 2018. H-ER reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right).
Source: SWIFT BI Watch.

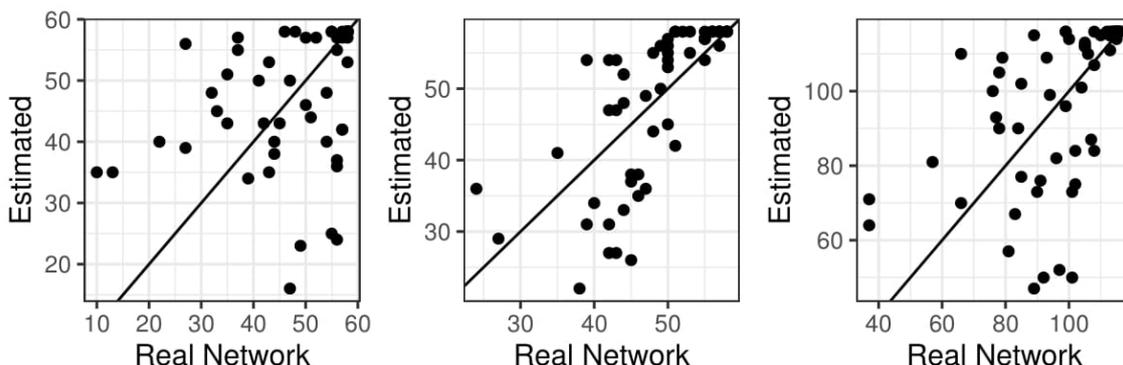


Figure 43: Degree Reconstruction in the reduced MT 103 network in February 2018. H-FIT reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right).
Source: SWIFT BI Watch.

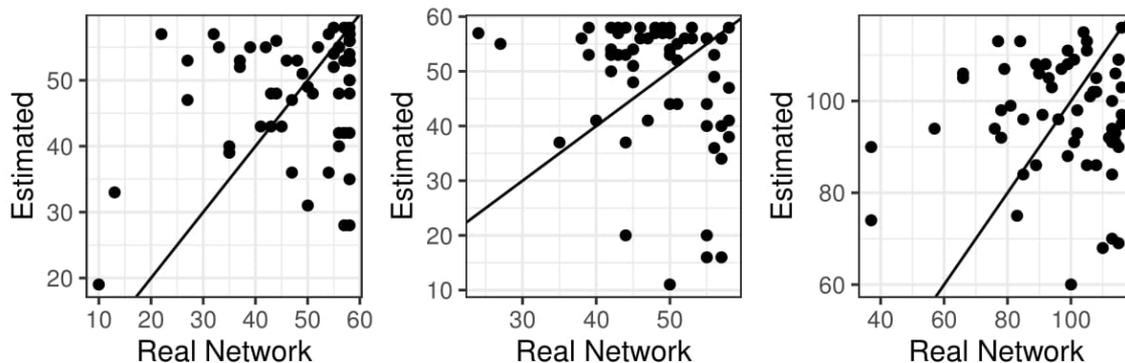


Figure 44: Degree Reconstruction in the reduced MT 103 network in February 2018. LASSO reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

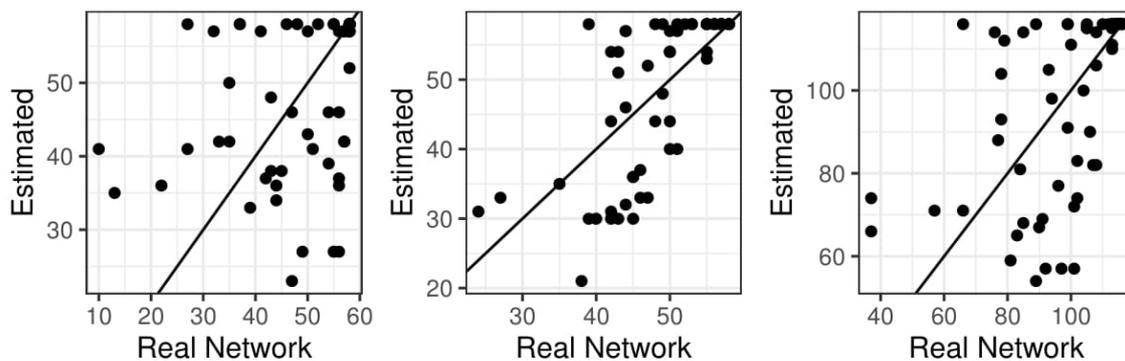


Figure 45: Degree Reconstruction in the reduced MT 103 network in February 2018. IPFP-GDP reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

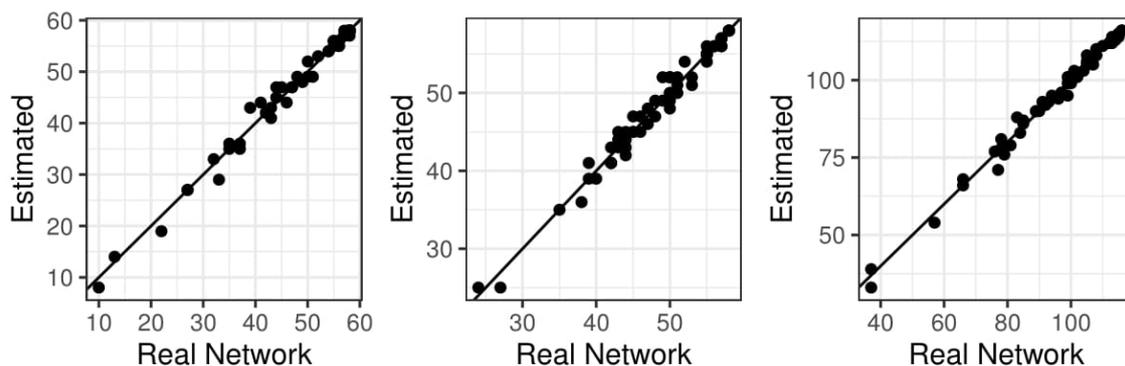


Figure 46: Degree Reconstruction in the reduced MT 103 network in February 2018. IPFP-LAG reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right). Source: SWIFT BI Watch.

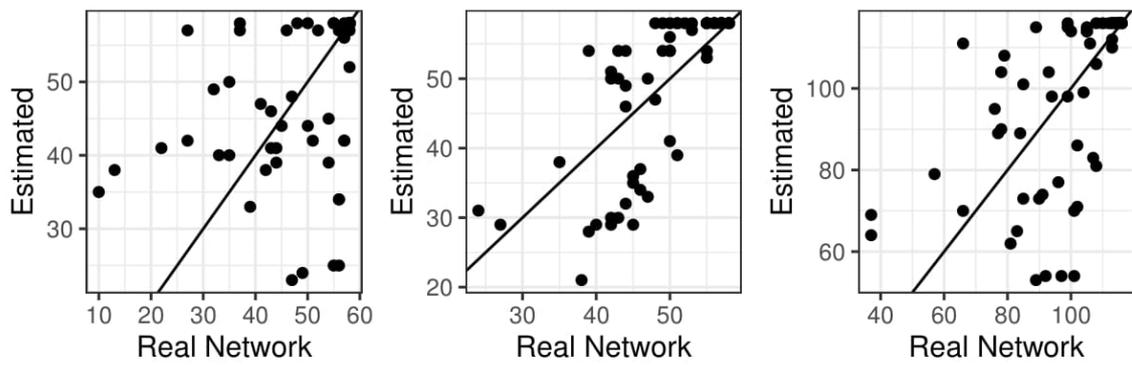


Figure 47: Degree Reconstruction in the reduced MT 103 network in February 2018. TOMO-GRAVTIY reconstruction of the outdegree (left), outdegree (middle) and in- and outdegree (right).

Source: SWIFT BI Watch.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 05.09.2019

Michael Lebacher