

Benjamin Sischka

# **Graphon Models for Network Data: Estimation, Extensions and Applications**

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

Eingereicht am 30. März 2023





Benjamin Sischka

# **Graphon Models for Network Data: Estimation, Extensions and Applications**

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

Eingereicht am 30. März 2023

Erster Berichterstatter: Prof. Dr. Göran Kauermann (LMU München)  
Zweite Berichterstatterin: Prof. Dr. Gesine Reinert (University of Oxford)  
Dritter Berichterstatter: Prof. Dr. Thomas Nagler (LMU München)

Tag der Disputation: 17. Juli 2023

## Acknowledgments

Clearly, this dissertation would not have been possible without the help and support of many people, to whom I would like to express my sincere gratitude. First and foremost, I want to thank my supervisor Göran Kauermann, for his unwavering support and guidance throughout my dissertation, and especially for his patience even when I was once again on one of my many detours. Without his expert advice, insightful feedback, and constant encouragement, I might still be searching for any crucial detail to find. Furthermore, I would like to thank Gesine Reinert, not only for agreeing to be part of my examination committee but also for her gentle and fruitful feedback after some talks I gave and for the very nice conversations we had at the always enjoyable CostNet events. I also want to express my gratitude to Thomas Nagler, who was willing to take on the responsibility of reviewing my thesis without hesitation. Beyond that, I am grateful for the many great football games we played together and other social events we both participated in.

Another important factor on this journey was the fellowship we had among the colleagues at the chair, which made for a very good working atmosphere. Being one of the first companions, I thank Michael Lebacher for many light-hearted moments and for the interesting exchanges we had, not only on many joint conference trips. Moreover, I thank Giacomo De Nicola for not only being an excellent co-author but also for inviting the whole crew to his hometown Pisa. I am also grateful for the many subject-related discussions with Cornelius Fritz and the always delicious pizza evenings at his home. My further appreciation goes to Marc Schneble for being an awesome office mate and for having many joint beers after work or on weekends. Likewise, I was pleased when Martje Rave joined our office, who not only brought more balance but was also a wonderful person to talk to about emotional situations. In addition, I would like to thank Verena Bauer, Christoph Striegel, and Shuai Shao for giving me a warm welcome into the chair, as well as Michael Windmann for many stimulating discussions about what could be improved at the department. Moreover, I appreciate having had a great co-organizer for the chair-wide hinking day in Constanze Schmaling, a colleague to talk about graphons in Nurzhan Sapargali, and a student who has become even more involved as a colleague in Cornelia Gruber. I further thank Iris Burger, who hopefully is now enjoying her retirement, and Martina Brunner for their great administrative work and always being of help if needed.

This list could surely be extended and is far from being exhaustive. What definitely needs to be added, though, is my deep gratefulness for the help of my family, who always backed me up, and the immense support from my beloved girlfriend Fiona. Although the last few months have not been easy, or even just because of that, I am very much looking forward to the time that lies immediately ahead of us.



## Summary

Network data are nowadays prevalent in various fields such as social and political sciences, economics, biology, neurosciences, and others. This is due to the fact that the structure within many systems can be described as connectivity pattern between entities. Together with modern surveying and measuring technologies, a systematic collection of data in such a structural format has recently become common practice. As a result, the statistical modeling of complex networks has gained traction over the last decades. Analyzing network data from the statistical perspective, however, encounters specific difficulties stemming from high interdependencies between the actors' connections. Hence, such data structures require the development of specifically tailored models.

This dissertation is concerned with the so-called *graphon model*, a specific framework for network data, and it addresses aspects of estimation, extensions, and applications. More precisely, this thesis is divided into three parts: (i) the development of an estimation procedure for smooth graphons, (ii) the additional incorporation of block structures by employing the graphon model in a mixture model configuration, and (iii) the elaboration of a joint graphon estimation strategy for modeling multiple networks simultaneously, which allows developing a procedure for nonparametric testing on structural equivalence of networks.

The smooth graphon model for network data is a very flexible approach that enables capturing diverse structural aspects, making it a valuable tool for learning the underlying network structure. Moreover, the graphon model belongs to the class of *node-specific latent variable models*, where “node” refers to the interacting objects. As for other latent variable models, the unobserved quantities (in this case, the node positions) and the concrete model structure (the graphon) need to be estimated based on each other. In the context of (i), we design a simultaneous estimation by applying principles of the EM algorithm. In the E-step, MCMC techniques are employed to approximate the marginal conditional mean of the node positions. To achieve a smooth functional estimate for the graphon, we make use of linear B-spline regression, which represents the M-step.

Regarding general concepts of statistical network analysis, a common assumption is the emergence of blockwise patterns as a result of the group-formation phenomenon. Following the intention of (ii), we explicitly incorporate such a structure into the graphon model by combining its smooth version with stochastic blockmodels. Applying such a modeling strategy allows to divide the network into groups of actors which still possess smooth differences in their connectivity behavior. Thus, this approach exploits the advantages of both the smooth graphon model and the stochastic blockmodel. An adapted EM-type algorithm can be employed to fit this generalized model to given network data.

The graphon model as a modeling framework has certain advantages compared to other approaches. On the one hand, the graphon can be viewed more broadly as general nonparametric density function on networks. On the other hand, the size and structure of a network are decoupled in the model formulation, which is a deficit in many other models. In line with (iii), we exploit these properties to develop a joint graphon estimation routine for modeling multiple networks simultaneously. Corresponding estimation results enable to compare the structures of networks directly. In this line, we construct a chi-squared test based on differences in the local connectivity behavior, providing information on whether the networks are drawn from the same distribution.

All the analytic concepts outlined above are elaborated in-depth, and their usability and applicability are demonstrated by considering the performance on both simulated data and real-world networks. Corresponding implementations are provided by open source Python packages which are publicly available on <https://github.com/BenjaminSischka>.



## Zusammenfassung

Netzwerkdaten sind heute in verschiedenen Bereichen wie Sozial- und Politikwissenschaften, Wirtschaft, Biologie, Neurowissenschaften und anderen weit verbreitet. Dies ist darauf zurückzuführen, dass sich die Struktur vieler Systeme als Verbindungen zwischen Einheiten beschreiben lässt. Zusammen mit modernen Umfrage- und Messtechniken ist die systematische Erfassung von Daten in einem solchen strukturellen Format in letzter Zeit zur gängigen Praxis geworden. Infolgedessen hat die statistische Modellierung komplexer Netzwerke in den letzten Jahrzehnten an Bedeutung gewonnen. Die Analyse von Netzwerkdaten unter statistischen Gesichtspunkten stößt jedoch auf besondere Schwierigkeiten, die sich aus den starken Abhängigkeiten zwischen den Verbindungen der Akteure ergeben. Solche Datenstrukturen erfordern daher die Entwicklung spezifisch zugeschnittener Modelle.

Diese Dissertation befasst sich mit dem sogenannten *Graphonmodell*, einem spezifischen Rahmen für Netzwerkdaten, und behandelt Aspekte der Schätzung, Erweiterungen und Anwendungen. Genauer gesagt gliedert sich diese Arbeit in drei Teile: (i) die Entwicklung eines Schätzverfahrens für glatte Graphone, (ii) die zusätzliche Einbeziehung von Blockstrukturen durch die Verwendung des Graphonmodells in einer Mischmodellkonfiguration und (iii) die Ausarbeitung einer gemeinsamen Graphonschätzungsstrategie zur gleichzeitigen Modellierung mehrerer Netzwerke, die die Entwicklung eines Verfahrens zur nichtparametrischen Prüfung auf strukturelle Äquivalenz von Netzwerken ermöglicht.

Das glatte Graphonmodell für Netzwerkdaten ist ein sehr flexibler Ansatz, der es ermöglicht, verschiedene strukturelle Aspekte zu erfassen, was es zu einem wertvollen Instrument für das Lernen der zugrunde liegenden Netzwerkstruktur macht. Außerdem gehört das Graphonmodell zur Klasse der *knotenspezifischen latenten Variablenmodelle*, wobei "Knoten" der konzeptionelle Begriff für die interagierenden Objekte ist. Wie bei anderen latenten Variablenmodellen müssen die unbeobachteten Größen (hier die Knotenpositionen) und die konkrete Modellstruktur (das Graphon) auf gegenseitiger Grundlage geschätzt werden. Im Zusammenhang mit (i) entwerfen wir eine simultane Schätzung, indem wir die Prinzipien des EM-Algorithmus anwenden. Im E-Schritt werden MCMC-Verfahren eingesetzt, um den marginalen bedingten Mittelwert der Knotenpositionen zu approximieren. Um eine glatte funktionale Schätzung für das Graphon zu erhalten, verwenden wir die lineare B-Spline-Regression, die den M-Schritt darstellt.

Im Hinblick auf allgemeine Konzepte der statistischen Netzwerkanalyse ist eine gängige Annahme das Auftreten von blockweisen Mustern als Ergebnis des Gruppenbildungsphänomens. Im Sinne von (ii) beziehen wir eine solche Struktur explizit in das Graphonmodell ein, indem wir seine glatte Version mit stochastischen Blockmodellen kombinieren. Die Anwendung einer solchen Modellierungsstrategie ermöglicht es, das Netzwerk in Gruppen von Akteuren zu unterteilen, die dennoch glatte Unterschiede in ihrem Konnektivitätsverhalten aufweisen. Somit nutzt dieser Ansatz die Vorteile sowohl des glatten Graphonmodells als auch des stochastischen Blockmodells. Ein angepasster EM-Algorithmus kann verwendet werden, um dieses verallgemeinerte Modell an gegebene Netzwerkdaten anzupassen.

Das Graphonmodell als Modellierungsrahmen hat im Vergleich zu anderen Ansätzen bestimmte Vorteile. Einerseits kann das Graphon im weiteren Sinne als allgemeine nichtparametrische Dichtefunktion für Netzwerke betrachtet werden. Zum anderen sind Größe und Struktur eines Netzwerks in der Modellformulierung entkoppelt, was ein Defizit vieler anderer Modelle darstellt. Im Einklang mit (iii) nutzen wir diese Eigenschaften, um eine gemeinsame Graphon-SchätZRoutine für die gleichzeitige Modellierung mehrerer Netzwerke zu entwickeln. Entsprechende Schätzergebnisse ermöglichen es, die Strukturen von Netzwerken direkt zu vergleichen. In diesem Zusammenhang konstruieren wir einen Chi-Quadrat-Test, der auf Unterschieden im lokalen Konnektivitätsverhalten basiert und Auskunft darüber gibt, ob die Netzwerke aus der gleichen Verteilung stammen.

Alle oben skizzierten Analysekonzepte werden eingehend erläutert, und ihre Nutzbarkeit und Anwendbarkeit wird anhand der Tauglichkeit sowohl in Bezug auf simulierte Daten als auch auf reale Netzwerke demonstriert. Entsprechende Implementierungen werden durch Open-Source Python-Pakete bereitgestellt, die auf <https://github.com/BenjaminSischka> öffentlich zugänglich sind.



# Contents

<b>I. Introduction</b>	<b>1</b>
1. Overview	1
<b>2. Statistical Network Analysis</b>	<b>3</b>
2.1. Formalization of Network Data . . . . .	3
2.2. The Nature of Networks . . . . .	4
2.3. Statistical Modeling Approaches . . . . .	5
2.3.1. Stochastic Network Processes . . . . .	5
2.3.2. Network Topology . . . . .	6
2.3.3. Node-Specific Latent Variable Models . . . . .	7
2.4. Network Comparison . . . . .	9
<b>3. Graphon Model</b>	<b>12</b>
3.1. Formulation and Interpretation . . . . .	12
3.2. Identifiability and Complexity . . . . .	15
3.3. Estimation and Inference . . . . .	18
3.4. Applications and Links to Other Models . . . . .	20
<b>4. Current Research, Contributions, and Outlook</b>	<b>26</b>
<b>Bibliography</b>	<b>29</b>
<b>Appendix</b>	<b>37</b>
<b>II. Smooth Graphon Estimation / Chapter 5</b>	<b>42</b>
<b>III. Incorporating Block Structures</b>	<b>62</b>
6. The Stochastic Blockmodel and Its Graphon Representation	63
7. A Mixture of Stochastic Blockmodels and Smooth Graphon Models	64
<b>IV. Graphon-Based Network Comparison / Chapter 8</b>	<b>124</b>
Contributing Publications	176
Eidesstattliche Versicherung (Affidavit)	178



**Part I.**

**Introduction**

# 1. Overview

Network data arise these days in various fields as an intuitive measurement for describing different kinds of systems. In general, network-structured data represent some sort of connectivity pattern between some type of entities, entailing a versatility that makes them applicable to abstracting diverse real-world situations. A network as such consists of a set of *nodes*, representing the actors or objects that are somehow related, and a set of *edges*, which reflect the formed relationships. The growing amount of network data, evoked by an increasing interest in analyzing corresponding systems, consequently requires the development of statistical tools for modeling and analyzing such kind of data. As a particular challenge in this context, network-structured data usually possess a complex dependency structure. Hence, customized methodological approaches are necessary for analyzing this specific data format. This dissertation contributes to the field of statistical network analysis by extending previous methods and developing new ones. The subject matters of these contributions can be sketched as follows.

**Nonparametric Modeling of Network Data.** One very flexible framework in the network context is the so-called *graphon model*. As a theoretical construct, it arises from the theories of graph limits and exchangeable random graphs. Moreover, the structural specification in this model is constituted by a bivariate function called *graphon*, which can be interpreted more broadly as density or intensity function on networks. Since this function can flexibly vary in shape and, in principle, does not rely on any parameterization, the graphon model is considered a nonparametric representation of network structures. As such, the graphon model is known to be able to represent complex structural aspects and even to fully cover other models, such as the *stochastic blockmodel* and the *latent distance model*. Yet, in contrast to the latter approaches, the graphon model cannot be reduced to such explicit structural properties as clusterability or distance-based representability (which specifically also implicates assortativity). Moreover, the graphon model can be linked to methods that rely on network statistics. In this line, corresponding representations have been formulated for simple *exponential random graph models*. Note that all these capabilities become possible because of the graphon model's high flexibility, which, on the other hand, entails a high complexity when it comes to estimation. In Part II of this thesis, we demonstrate how concepts of the EM algorithm, in combination with MCMC techniques and spline-based approaches, can be used to develop a smooth graphon estimation routine.

**Stochastic Equivalence and Its Relaxation.** One phenomenon that has been experienced regularly during the extensive analysis of networks over the last decades is the formation of blocks. The constitution of these blocks is characterized by a grouping of nodes that exhibit a similar connectivity behavior. A model that is specifically dedicated to uncovering such structural behavior is the stochastic blockmodel. In order to fit this model to network data, one usually has to estimate both the group memberships and the blockwise connectivity structure. The crucial assumption for the stochastic blockmodel to be legitimate is the stochastic equivalence. This implies that nodes from the same block have the exact same connection probabilities to all other nodes, which appears to be a rather rigorous assumption. A relaxation of that might be the setup where the

nodes decompose into blocks that permit slight differences in the stochastic connectivity behavior. Following the terming of “stochastic equivalence”, this might be referred to as “stochastic similarity.” In fact, real-world networks could most likely be represented much more accurately by such a modified framework since strict equivalence seems to be an unrealistic assumption. In Part III, we rely on this intuition and incorporate smooth differences into the stochastic blockmodel by formulating a piecewise smooth graphon model. The lines of discontinuity might then also be interpreted as structural breaks.

**Statistical Testing on Structural Equivalence.** Although the literature on statistical network analysis has extensively increased in recent years, it mainly focused on describing and modeling global structural patterns or specific structural aspects. One facet that has not been dealt with in such depth is the statistical testing on structural equivalence of networks. With regard to general network comparison, several strategies have been developed to contrast the overall structural behavior. Yet, these methods rely almost exclusively on descriptive statistics and thus do not provide any distributional assumptions on the resulting distance metric. Apparently, such approaches prohibit drawing inference in the probabilistic sense. A reason for the shortcoming of inference-based methods lies in the fact that testing on whether two networks are drawn from the same distribution is a particularly difficult task per se. A given network is, in its entirety, usually considered a single observation of a specific system, which, in the comparison context, means to contrast unknown underlying distributions based on only one realization each. As an additional issue in this regard, the networks might be of different sizes, causing a natural aspect of dissimilarity. A corresponding distributional specification thus requires the ability to reasonably define a probability measure on networks in a size-detached manner. We formulate such a model-based distribution in Part IV by making use of the graphon model. In particular, we apply a joint graphon estimation to model multiple networks simultaneously. This provides a general network alignment, which, in turn, can be used to contrast networks of different sizes on the edge level.

**Outline of this Thesis.** The first part of this dissertation briefly introduces the general research topic dealt with in the three subsequent main parts. This introductory part is organized as follows. In Chapter 2, we outline the characteristics of network data and what distinguishes it from other kinds of data structures. In addition, this includes a brief overview of the most common models for analyzing network data. To complement the chapter, we discuss potential strategies for comparing the structures between networks. A detailed introduction to the graphon model as the central modeling approach in this dissertation follows in Chapter 3. Starting with its general formulation and interpretation, we further discuss difficulties and potential strategies with respect to its estimation, as well as possible applications in the context of the concrete analysis of network structures. Chapter 4 concludes the introduction with a short review of the current research in the graphon literature, a small summary of the contributions of this thesis, and a brief outlook. This first part of the dissertation is followed by the actual contributions, which can be roughly outlined as follows. In Part II, we develop a smooth graphon estimation routine based on MCMC techniques and spline-based approaches. An extension of the class of *smooth* graphon models is formulated in Part III, where we propose a unification with the stochastic blockmodel. Finally, in Part IV, we formulate a joint graphon estimation strategy which we utilize to develop a nonparametric test on networks. In particular, this allows to draw statistical inference with regard to structural equivalence, which still appears to be an open challenge in the network analysis literature.

## 2. Statistical Network Analysis

To make the concepts and formulations used in this thesis more comprehensible, we start by introducing and formalizing the type of data we are dealing with. In that line, we elucidate the particular character of network-structured data and emphasize the difficulties arising when pursuing to analyze them. Based on that, we give a brief outline of statistical modeling approaches specifically developed for such data structures. To complement the notion of network data and corresponding modeling concepts, we discuss possible strategies for network comparison.

### 2.1. Formalization of Network Data

The usage of network-structured data allows to capture the essential features of diverse systems. In general, a network construct consists of entities and relations or interactions between them. For example, this might represent friendships among members of a social group (Eagle et al., 2009), the trading between nations (Bhattacharya et al., 2008), interactions of proteins (Schwikowski et al., 2000), or the functional coactivation within the human brain (Bassett et al., 2018, Crossley et al., 2013).

**Networks and Graphs.** As a universal concept, the formulation of networks allows abstracting the structure within a system of interrelated objects. In a more mathematical way, a network is usually conceptualized as a *graph*  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  and  $\mathcal{E} \subseteq \mathcal{V}^2(\times \mathcal{W})$  represent the set of nodes (or vertices, inspiring the notation of  $\mathcal{V}$ ) and edges, respectively. The additional usage of  $\mathcal{W} \subseteq \mathbb{R}$  facilitates edges to possess weights, allowing to indicate differing strengths of connectedness. Hence, in the general graph representation, the set of edges  $\mathcal{E}$  comprises pairs of nodes equipped with corresponding weights, i.e.  $(v_i, v_j, w_{v_i v_j})$  with  $i, j \in \{1, \dots, N\}$ . The *size* (or *order*) of a graph—or of the corresponding network—is usually defined by  $N = |\mathcal{V}|$ , where  $|\cdot|$  is the cardinality of a set. (Note that, contrarily, some works define  $|\mathcal{E}|$  as the size of a network.) While the term “network” refers mostly to a collection of interrelated elements as a notional object or, similarly, to its graphical representation consisting of dots or circles and connecting lines, a “graph” formalizes this vague concept. As such, a graph enables to properly define (mathematical) operations. In this light, the field of graph theory provides a whole bunch of definitions, propositions, and useful techniques for accomplishing meaningful transformations, extracting desired information, and solving specific problems. In the present thesis, we specifically focus on *simple undirected graphs*, i.e. graphs that only indicate whether two nodes are connected or not. Transferred to the network context, this refers to instances where links do not possess a distinct strength or a certain direction. As for the graph specification, we consequently neither observe any associated weights (which is tantamount to setting  $w_{v_i v_j} = 1$  for all present edges) nor distinguish between the (unweighted) edges  $(v_i, v_j)$  and  $(v_j, v_i)$ . In addition, throughout the entire dissertation, we assume the graphs to be free of self-loops, meaning  $(v_i, v_i) \notin \mathcal{E}$  for all  $v_i \in \mathcal{V}$ . As a consequence of

the undirectedness, we can further substitute the two tuples  $(v_i, v_j)$  and  $(v_j, v_i)$  by the (unordered) set  $\{v_i, v_j\}$ .

**Adjacency Matrix.** Instead of relying on  $\mathcal{E}$ , often also referred to as *edge list* (where the ordering aspect of a typical list object is neglected), the connectivity within a simple undirected graph can alternatively be described by the so-called *adjacency matrix*  $\mathbf{Y} = [Y_{v_i v_j}]_{i,j=1,\dots,N} \in \{0, 1\}^{N \times N}$ . The entries of this adjacency matrix are given by the individual *edge variables*  $Y_{v_i v_j} \in \{0, 1\}$  for  $i, j = 1, \dots, N$ , representing either the presence ( $\{v_i, v_j\} \in \mathcal{E}$ , encoded as 1) or the absence ( $\{v_i, v_j\} \notin \mathcal{E}$ , encoded as 0) of the connection between nodes  $v_i$  and  $v_j$ . Taking into account that edges are undirected and self-loops are omitted, it generally holds that  $Y_{v_i v_j} \equiv Y_{v_j v_i}$  and  $Y_{v_i v_i} \equiv 0$  for all  $i, j = 1, \dots, N$ . The realizations of the edge variables and the adjacency matrix are denoted by  $y_{v_i v_j}$  and  $\mathbf{y} = [y_{v_i v_j}]_{i,j=1,\dots,N}$ , respectively. Note that to simplify notation, the nodes are commonly labeled from 1 to  $N$ . This shortens the expression of the adjacency matrix to  $\mathbf{Y} = [Y_{ij}]_{i,j=1,\dots,N}$ , and of other quantities analogously. The structure exhibited in any simple undirected graph can be fully represented by the corresponding adjacency matrix. Hence, a graph of this type can equivalently be specified through  $\mathcal{G} = (\mathcal{V}, \mathbf{y})$ . In contrast to the edge list format, a matrix representation allows for performing graph calculations by applying calculations on matrices. One commonly used method from the ensemble of matrix calculations is, for example, the spectral decomposition, often applied for graph clustering (Rohe et al., 2011) or estimating the edge probability matrix under specific model assumptions (Chatterjee, 2015). In this dissertation, we also rely on the adjacency matrix as the object that quantifies the present structure (or, at least, on the edge variables contained therein).

## 2.2. The Nature of Networks

**Dependency Structure.** The aspect that differentiates networks as complex structural objects from ordinarily structured data considered in classical statistical frameworks—i.e. tabular data which are assumed to consist of independent and identically distributed (i.i.d.) observations—is the inherent dependency structure. This dependency is very pronounced in the context of networks since the relationships between individuals usually strongly affect each other. Putting it more contrastively, in an ordinary data set as a collection of individual measurements, the interrelation among observations is “(nearly) structureless,” whereas “in the case of network data, the [dependency] structure is the data” (Crane, 2018, p. 3). In other words, classical statistics treats dependence as a confounding factor that is tackled, if at all, by simple structural assumptions, while in the analysis of networks, the inherent dependency structure is what one aims—more or less explicitly—to model. In the various disciplines applying statistical network analysis, different recurring phenomena with respect to such dependencies have been identified. In most cases, these phenomena can be brought into line and explained with behavioral patterns well-known in the respective domains. Common behaviors in networks are denominated, for instance, in the social sciences, which is the field with perhaps the longest tradition in network analysis. Popular observed patterns are, for example, “the friend of my friend is my friend” (also known as *triadic closure effect*) or more wide-ranging cohesive structures, i.e. group-forming effects. However, the dependencies among the relations might be much more complex. In any case, classical approaches for statistical modeling cannot be easily adopted. Thus, the analysis of networks requires the development of novel methods taking these particular circumstances into account.

## 2.3 Statistical Modeling Approaches

---

**Special Structural Features.** Besides the complex dependency structure, there are further intrinsic aspects that contradict classical statistical principles and thus need to be paid specific attention to when modeling real-world networks. A common issue in such data structures is that of *sparsity*, indicating a marginal proportion of existing versus possible edges, i.e.  $|\mathcal{E}| \ll \binom{N}{2}$ . This often appears in the form of a power-law degree distribution (Albert and Barabási, 2002) or other compositions that entail a still well-connected graph in terms of global connectedness. Since the latter property is contrary to corresponding expectations under classical paradigms, the sparsity condition is especially challenging. Such a well-connectedness of graphs can be similarly conceptualized by the *small-world property* (Watts and Strogatz, 1998), which additionally implies the triadic closure effect. Another connectivity pattern associated with heterogeneous behavior—which is closer to what has already previously been pursued to capture by statistical models—is the formation of groups of nodes (Holland et al., 1983). If these groups appear to be more strongly connected within than between themselves, they are usually referred to as *communities* (Girvan and Newman, 2002).

Altogether, modeling network data requires the consideration of diverse aspects, many of which are contrary to the way of thinking that has prevailed in the statistical modeling literature in the past.

## 2.3. Statistical Modeling Approaches

In the last decades, many different methods for modeling complex random graphs have been proposed and extensively developed. Survey articles that demonstrate the state of the art in statistical network analysis have been published by Goldenberg et al. (2009), Snijders (2011), Hunter et al. (2012), Fienberg (2012), and Salter-Townshend et al. (2012). Monographs in this field outlining the extensive methodological development since its beginnings are given by Kolaczyk (2009), Lusher et al. (2013), Kolaczyk and Csardi (2014), Kolaczyk (2017), Newman (2018), and Crane (2018). Altogether, there exist various strategies for describing and modeling network data, which rely on a compendium of diverse concepts. We emphasize that some of these methods are also naturally able to model directed, weighted, or bipartite networks or to incorporate covariates on the node or edge level. For other models, there might exist corresponding extensions. Nonetheless, since the focus of this dissertation is on simple undirected graphs, the following overview is restricted to models for this data setting. The description for “broader” models is narrowed down accordingly.

As a major distinction besides different capacities, the existing modeling strategies can also be distinguished according to their underlying paradigm, which can be of either *static* or *dynamic* nature. Under the static paradigm, the purpose is to model the topology of a single network instance. That is, to uncover the structure of a fully evolved connectivity pattern. In contrast, the dynamic paradigm implies considering a network not only as a fixed and “final” construct but from the perspective of an evolutionary process consisting of successively occurring events.

### 2.3.1. Stochastic Network Processes

**Evolutionary Structure.** Investigating the structural formation within a network on the level of the underlying stochastic process apparently involves some kind of temporal component. This

consequently entails data in a longitudinal format. Under this paradigm, the goal is to model the underlying stochastic process that forms the network. Possible events serving as the intermediate evolutionary steps are the formation (or disappearance) of edges, but potentially also the emergence (or disappearance) of nodes. The most popular modeling strategy in this direction is the *stochastic actor-oriented model* (Snijders et al., 2010), which was specifically developed for social networks. The basic idea here is a scenario where actors are aware of the occurrences and (at least to some extent) in control of their relationships. The intention of this approach is to model the upcoming events conditional on the respective current state of the network. A similar framework can be developed for an edge-oriented perspective (Snijders and Koskinen, 2013).

**Network Dynamics.** In the last years, methods that have been primarily developed for covering the topology of networks were extended towards the dynamic perspective. Works in this direction are, among others, Hanneke et al. (2010) and Krivitsky and Handcock (2014) (for exponential random graph models), Matias and Miele (2017) (for stochastic blockmodels), and Pensky (2019) (for graphon models, see below for a description of the respective original model). Yet, these approaches do not strive for actually modeling the underlying stochastic process. Instead, they aim for incorporating information on different states of the network to improve the precision of the model or to extend the framework towards a time-varying specification.

Considering the network’s stochastic process instead of its topology seems to provide richer information simply because it does not only exhaust knowledge about the presence or absence of edges but also about the time of their emergence or disappearance. More specifically, edges might occur only temporarily and disappear after a while, which would not be recorded by the final network topology at all. Nonetheless, networks are not always the outcome of an evolutionary process or, at least, do not simply disclose information about the immanent intermediate steps. Moreover, one might often be particularly interested in the topological structure of a network rather than how it evolved over time.

### 2.3.2. Network Topology

In this thesis, we follow the static paradigm, which means pursuing the intention to model the topology of a fully evolved network. Approaches of this category can be distinguished by the strategies used for formalizing the dependency structure.

**Subgraph Frequencies.** One of the two tendencies in this direction can be confined to capturing a network’s structure by specific network statistics. To be precise, these statistics are commonly defined as the frequencies of prespecified subgraphs, providing information on the prevalence of local connectivity patterns. The workhorse model in this context is the *exponential random graph model*, often also referred to as *p\* model* (Frank and Strauss, 1986, Robins et al., 2007). The underlying intuition of this framework is to formulate a probability distribution on networks on the basis of subgraph counts. In the classical design, this can be specified through

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}; \boldsymbol{\eta}) = \frac{\exp(\boldsymbol{\eta}^\top \mathbf{s}(\mathbf{y}))}{v(\boldsymbol{\eta})},$$

where  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)^\top \in \mathbb{R}^m$  is a parameter vector,  $\mathbf{s}(\cdot)$  is a vector operation returning the counts of  $m$  (prespecified) subgraphs, and  $v(\boldsymbol{\eta})$  is a normalizing constant, ensuring that  $\sum_{\mathbf{y} \in \mathcal{Y}_N} \mathbb{P}(\mathbf{Y} =$

## 2.3 Statistical Modeling Approaches

---

$\mathbf{y}; \boldsymbol{\eta}) = 1$  with  $\mathcal{Y}_N$  being the set of all simple undirected graphs of size  $N$ . The concept of exponential random graph models arises from the generalization of Markov graphs (Wasserman and Pattison, 1996). Concurrently, the exponential random graph model describes an extension of the  $p_1$  model, which, as a model for directed graphs, specifically aims at capturing the effects of reciprocity as well as in- and outgoing attractiveness (Holland and Leinhardt, 1981). The concrete specification of this model, meaning the predefinition of what kind of subgraphs are relevant, needs to be done based on subject-matter expertise (Robins et al., 2007, Lusher et al., 2013).

**Latent Variable Models.** The second main model class focused on capturing the completed network topology comprises the node-specific latent variable models (see Matias and Robin, 2014 for an overview). Methods of this class follow the universal assumption that the complex dependency structure within a network can be ascribed to nodal quantities with respect to some underlying (more or less complex) structural construct. A generic specification of this model class can be given through

$$Y_{ij} \mid \boldsymbol{\xi}_i, \boldsymbol{\xi}_j \stackrel{\text{ind.}}{\sim} \text{Bernoulli}(h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)) \quad \text{for } i < j, \quad (1)$$

where  $0 \leq h(\cdot, \cdot) \leq 1$  represents a corresponding overall connectivity pattern. Adopting the assumptions of simple undirected graphs, it generally holds that  $Y_{ji} \equiv Y_{ij}$  and  $Y_{ii} \equiv 0$ . Model formulation (1) clearly underlines that the connection probability for node pair  $(i, j)$  depends exclusively on the corresponding nodal quantities  $\boldsymbol{\xi}_i$  and  $\boldsymbol{\xi}_j$ . Depending on the concrete model choice, these quantities are considered either as random variables themselves (due to either a corresponding model specification or a Bayesian perspective) or simply as unknown but fixed parameters. Moreover, we stress that  $\boldsymbol{\xi}_i$  is used as a scalar in many frameworks, although it can generally be of multivariate form.

This general class of latent variable models accounts for a large part of the literature on statistical network analysis. In fact, it includes some of the most extensively discussed and frequently applied methods for modeling network data, the most popular of which we will briefly introduce in the next section. In this regard, it seems worth noting that these approaches possess different abilities and qualifications to cover the diverse structural aspects that might be present in network data. This is despite the fact that they are all inspired by the same notion, namely that the structure within networks can be boiled down to nodal quantities referring to some more or less complex structural construct.

### 2.3.3. Node-Specific Latent Variable Models

**Latent Node Positions.** In line with general latent variable framework (1), Hoff et al. (2002) developed an approach according to which the network’s nodes are situated in an (Euclidean) latent space in which distances provide full information about connection probabilities. In its simple form, the *latent distance model* can be formalized as

$$\mathbb{P}(Y_{ij} = 1 \mid \mathbf{X}_i = \mathbf{x}_i, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta}) = \frac{\exp(\beta_0 - \beta_1 \|\mathbf{x}_i - \mathbf{x}_j\|)}{1 + \exp(\beta_0 - \beta_1 \|\mathbf{x}_i - \mathbf{x}_j\|)}, \quad (2)$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top \in \mathbb{R}^2$  is a coefficient vector and  $\mathbf{X}_i$  is the latent position of node  $i$ —potentially located in the metric space  $(\mathbb{R}^J, \|\cdot\|)$  with  $J \in \mathbb{N}$ . Note that the coefficient  $\beta_1$  is only required (and identifiable) if the latent space is bounded. In the normalized version,  $\mathbf{X}_i$  is assumed to be lying

within  $[0, 1]^J$ . The domain of the node positions is often referred to as latent *social* space, implying that coordinates might be translated into certain social attributes or vice versa. Beyond its classical interpretation, this approach can be extended towards a model-based clustering technique, where the latent positions are assumed to follow a mixture of multivariate normal distributions (Handcock et al., 2007).

**Node Clustering.** A more tailored approach for clustering the nodes of a network is given by the *stochastic blockmodel*. This model was originally formulated by Holland et al. (1983), whereas Snijders and Nowicki (1997, 2001) first tackled the issue of a *posteriori blockmodeling*. In this approach, the latent variables represent group memberships, denoted by  $Z_i \in \{1, \dots, K\}$  for  $i = 1, \dots, N$ , where  $K$  represents the number of groups. These memberships are then brought together with edge probabilities specified between and within groups, denoted by  $p_{kl} \in [0, 1]$  for  $k, l = 1, \dots, K$ . To be precise, the network data is assumed to be generated through

$$Y_{ij} \mid Z_i, Z_j \sim \text{Bernoulli}(p_{Z_i Z_j}).$$

The individual probabilities  $p_{kl}$  with  $k, l = 1, \dots, K$  are often gathered in the blockwise edge probability matrix  $\mathbf{P} = [p_{kl}]_{k,l=1,\dots,K} \in [0, 1]^{K \times K}$ , where, in the undirected setting, it holds that  $p_{lk} = p_{kl}$ . Commonly, the group memberships are, in turn, assumed to independently follow a categorical distribution in the form of

$$\mathbb{P}(Z_i = k; \boldsymbol{\alpha}) = \alpha_k \quad \text{for } k = 1, \dots, K$$

with  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in [0, 1]^K$  and  $\sum_k \alpha_k = 1$ . The entries of the vector  $\boldsymbol{\alpha}$  can consequently be interpreted as the (expected) group proportions. In this overall setup, the specification of  $K$  is apparently an essential aspect of the model itself. Information on that is only rarely given a priori in real-world scenarios, which is why it usually needs to be inferred from the data as well. To do so, there exist various methods in the blockmodeling literature, which rely on, for example, log-likelihood ratio statistics (Wang and Bickel, 2017), Bayesian inference under MCMC-based approximations (Newman and Reinert, 2016, Riolo et al., 2017), or cross-validation techniques (Chen and Lei, 2018).

Apparently, the general intuition of the stochastic blockmodel is the assumption that the overall heterogeneity between the nodes can be explained by global node set  $\mathcal{V}$  dividing into groups of individuals with homogeneous behavior. In formulae, this means that there exists a partition  $\mathcal{C}_1, \dots, \mathcal{C}_K$  with  $\bigcup_k \mathcal{C}_k = \mathcal{V}$  such that, for all  $k = 1, \dots, K$ ,

$$\mathbb{P}(Y_{ij} = 1) = \mathbb{P}(Y_{i'j} = 1) \quad \text{for all } i, i' \in \mathcal{C}_k \text{ and } j \in \mathcal{V}.$$

In turn, this implies that the set of all edge variables,  $\{Y_{ij} : i, j \in \mathcal{V}, i < j\}$ , as the complete connectivity pattern, can be divided into subsets of edge variables,  $\{Y_{ij} : i \in \mathcal{C}_k, j \in \mathcal{C}_l\}$  with  $k, l = 1, \dots, K$ , within which a homogeneous density applies. Such a strict homogeneity, however, might not always be appropriate and thus might be relaxed towards a more continually adapted edge density.

**Local Edge Density.** To specify the edge density on a more local level, one could make use of a more detailed specification. More precisely, instead of relying on the somehow artificial assumption of strict blockwise homogeneity—defined through the specification of  $(\boldsymbol{\alpha}, \mathbf{P})$ —, the edge density

## 2.4 Network Comparison

---

could be determined pointwise. To do so, the groups of homogeneous nodes could be chosen to be infinitesimally small (in relation to the overall number of nodes). This conception is closely related to Szemerédi’s regularity lemma (1978). According to that, the set of nodes of every large enough graph can be partitioned into a bounded number of groups such that the connectivity behavior between different groups is almost homogeneous, i.e. can be well approximated by a constant edge density. In the end, this breaks down to specifying an individual edge probability for each pair of nodes. This can be achieved by defining a local edge density in the form of a bivariate function. Note that “local” here refers to the position within the latent domain, which can alternatively be illustrated by ordering the adjacency matrix according to this latent scale (cf. Figure 1 below). In this sense, such a framework can be interpreted as mapping the entries of the two-dimensional adjacency matrix grid-wise into the domain of the bivariate function to get corresponding edge probabilities. Following this intuition, one ends up with the *graphon model* (Lovász and Szegedy, 2006, Borgs et al., 2008, Borgs et al., 2012), which covers exactly the formulated specification of a local edge density. Apart from general definitions, this model is fully specified through the eponymous bivariate function, i.e. the *graphon*. We will present a precise formulation of the graphon model in the next chapter. For the moment, however, we want to remain with its non-formalized conceptualization. As a further aspect, the graphon model arises from the theories of *graph limits* and *exchangeable random graphs* (Diaconis and Janson, 2008, see also Section 4). Still, it can likewise be interpreted as the local edge density. In this regard, it also covers the stochastic blockmodel, which is exemplarily demonstrated through a corresponding graphon specification by Latouche and Robin (2016, Fig. 1), see also Part III, Section 3.3 of Chapter 6. Conversely, many works have argued that the graphon model can be approximated by an extensive stochastic blockmodel and presented corresponding results on theoretical convergence rates (Wolfe and Olhede, 2013, Gao et al., 2015, Klopp et al., 2017). Hence, the graphon model can be considered a nonparametric extension of the stochastic blockmodel (Borgs et al., 2018a, p. 2) or as a blockmodel on an arbitrarily fine scale (Bickel and Chen, 2009, p. 21069). From a more practical view, other authors exploited the blockwise approximation to construct a concrete estimation technique (Airoldi et al., 2013, Chan and Airoldi, 2014, Yang et al., 2014).

Overall, the graphon model enables to specify a local edge density on networks, i.e. to assign an individual probability to each edge variable. Hence, the graphon model appears as a suitable choice for capturing the structure within a network in a nonparametric fashion. Lastly, this makes it specifically appropriate for comparing networks on a distributional basis, as it is elaborated in detail in Part IV. The following section presents common techniques for contrasting network data.

## 2.4. Network Comparison

As in classical statistics, comparing different samples, i.e. observed data sets in the form of  $\mathbf{y}^{(g)} = [y_{ij}^{(g)}]_{i,j=1,\dots,N^{(g)}}$  with  $g \in \{1, \dots, G\}$ , is an important yet not fully resolved problem in the network context. To give an example, one might be interested in whether and how the functional connectivity in the human brain differs for people suffering from autism spectrum disorder compared to a typical-development group (Song et al., 2019, Subbaraju et al., 2017, Pascual-Belda et al., 2018). Yet, for the complex structure inherent in network data, this task becomes much more difficult than for classical data formats. This problem aggravates even more when one seeks

a comparison on the basis of statistical inference, which is due to the difficulty of defining a universal network distribution that is invariant to specific individual network features such as size or even edge density.

**Node Correspondence-Based Approaches.** The most straightforward setting in the network comparison context is the situation where several connectivity patterns have been observed on the same set of nodes. That is, for all pairs of networks,  $(g, g') \in \{1, \dots, G\}^2$ , there exists a permutation  $\kappa : \{1, \dots, N\} \rightarrow \{1, \dots, N\}$  such that  $v_{\kappa(i)}^{(g')} = v_i^{(g)}$  for all  $i = 1, \dots, N$ , where  $N = N^{(1)} = \dots = N^{(G)}$ . This allows to directly calculate distances between corresponding adjacency matrices or other network-related matrices that carry substantial information (Koutra et al., 2013, Koutra et al., 2016, Liu et al., 2018). However, in most cases, there is no a priori knowledge about this *node correspondence*. That means a direct comparison of connection statuses between networks, i.e. checking  $y_{ij}^{(g)}$  against  $y_{\kappa(i)\kappa(j)}^{(g')}$ , requires first to infer the node mapping  $\kappa(\cdot)$  (Milenković et al., 2010, Kuchaiev and Pržulj, 2011, Memišević and Pržulj, 2012). This is usually done by maximizing

$$\begin{cases} \sum_{j>i} \left[ y_{ij}^{(g)} y_{\kappa(i)\kappa(j)}^{(g')} + (1 - y_{ij}^{(g)})(1 - y_{\kappa(i)\kappa(j)}^{(g')}) \right], & \text{if the connectivity behavior} \\ & \text{underlies the dense regime} \\ \sum_{j>i} y_{ij}^{(g)} y_{\kappa(i)\kappa(j)}^{(g')}, & \text{otherwise} \end{cases} \quad (3)$$

with respect to  $\kappa(\cdot)$ . Such a mapping can also be accomplished for networks of unequal sizes, where, without loss of generality,  $|\mathcal{V}^{(g)}| \leq |\mathcal{V}^{(g')}|$ , although this obviously prohibits a bijective node mapping. As a consequence, some information within the data gets lost, namely the one about the  $|\mathcal{V}^{(g')}| - |\mathcal{V}^{(g)}|$  nodes that drop out of network  $g'$  as being not assigned. Moreover, the computational task of finding the optimal node mapping according to (3) is, in principle, NP-complete and thus can only be solved approximately. Lastly, the direct comparison of connection statuses might not always be the most accurate strategy for uncovering structural differences between separate networks.

**Structural Network Comparison.** As an alternative, networks could be compared by relying on a more general representation of the present structure, i.e. without exploiting an explicit node correspondence. To do so, one needs first to quantitatively capture the underlying structures in a node label-independent manner, where, in the next step, one can appropriately compare these structural representations between networks. The specific approach chosen for representing the underlying structure apparently affects the capability of uncovering the differences of interest. As for classical frameworks, the structure can generally be captured by relying either on descriptive or model-based methods. Regarding descriptive approaches, Newman (2018, pp. 364 ff.) and Anderson et al. (1999, pp. 242 f.) formulate a comparison based on standard network features. One step further, Wilson and Zhu (2008) and Gera et al. (2018) rely on the graph spectrum to uncover differences in network structures. Another large part of the literature on feature-based strategies deals with graphlets, where the intuition is representing structure through frequencies of prespecified subgraphs, see e.g. Pržulj et al. (2004), Pržulj (2007), Yaveroğlu et al. (2014), Ali et al. (2014), and Faisal et al. (2017). However, as a general drawback of descriptive comparison methods, they apparently do not allow drawing statistical inference. As a potential way

## 2.4 Network Comparison

---

out, [Anderson et al. \(1999\)](#) and [Butts \(2008\)](#) suggest applying (a multivariate generalization of) conditional uniform graph distributions. On the other hand, model-based approaches for network comparison are rather rare. This is due to the fact that most network models impose specifications that are related to individual network attributes. For example, exponential random graph models involve coefficients that have different meanings for different total numbers of nodes. Similar issues arise for other frameworks. A stochastic blockmodel-based approach has been developed by [Onnela et al. \(2012\)](#), who compare summary statistics of the networks' disintegration processes. But still, this approach does not provide the possibility of drawing statistical inference. For further reading into this direction, survey articles for network comparison are given by [Soundarajan et al. \(2014\)](#), [Yaveroğlu et al. \(2015\)](#), [Emmert-Streib et al. \(2016\)](#), and [Tantardini et al. \(2019\)](#). Beyond the limitation to networks, [Marron and Alonso \(2014\)](#) discuss more generally how complex data objects—such as adjacency matrices—might be compared.

**Graphon-Based Comparison of Network Structures.** The graphon model as a probabilistic framework for network data can be viewed as outstanding with regard to the comparison task. In this context, we stress that the graphon can serve as both the central underlying object of a model-induced probability measure and a descriptive entity that captures the network structure on a fine resolution. From both perspectives, the graphon can be interpreted as (local) density or intensity function on networks. (These aspects are described in detail in the next section.) Given that, the graphon can be considered as an overall characterizing network feature that, like nothing else, uniquely covers the global structure. Furthermore, in the graphon model, the general probabilistic specification and individual network properties, like the number of nodes, are naturally decoupled. Taking all this together makes the graphon model an optimal choice for network comparison. Thus, as a first intuition, one might think of separately fitting the graphon model to different networks and subsequently comparing the resulting estimates. However, the graphon model is known to suffer from severe identifiability issues, which arise from the fact that any permutation of a graphon yields the same generating process as the original one ([Diaconis and Janson, 2008](#), Sec. 7). Consequently, the comparison of separate graphon model fits appears to be a complex matter. To circumvent this issue, the graphon approach can instead be employed to model multiple networks simultaneously. The results of the joint graphon estimation can then be utilized to formulate an appropriate network comparison strategy. To be precise, the aligned node positions—under certain smoothness assumptions—can be interpreted as a relaxed node correspondence that does not stipulate a one-to-one mapping but rather identifies small fuzzy groups of nodes with similar structural roles. With respect to the underlying graphon model, which induces a proper network distribution, this approach allows to formalize a statistical testing procedure on equivalence of the underlying network structures and to detect relevant microscopic differences. A concrete elaboration of this approach is given in detail in Part IV. In the next chapter, we will introduce the graphon model in detail, and, in doing so, it will also become more apparent why the graphon model is a specifically useful tool for comparison purposes.

## 3. Graphon Model

In this chapter, we will introduce and discuss the graphon model in detail. This includes its formal definition, distinctive properties, specific difficulties in the estimation, and a potential scope of application. In the following formulations, we will illuminate the graphon from two perspectives, namely as a naturally emerging limiting construct and as the central object of a nonparametric data-generating process. In this light, the graphon model provides an outstanding framework for statistical network analysis.

To further provide a theoretical background on the graphon's origination, in the Appendix, we give an overview of the concepts relevant to the development of the graphon model. These are graph limits and exchangeable random graphs, described in detail from page 37 on.

### 3.1. Formulation and Interpretation

**Graphon Model as Data-Generating Process.** In classical graphon theory, which has been mainly introduced and developed by Lovász and Szegedy (2006), Borgs et al. (2006), Diaconis and Janson (2008), Bollobás et al. (2007), and Borgs et al. (2008), a *graphon* is a bivariate and symmetric  $[0, 1]$ -valued function defined on a probability space. In particular, the graphon model can be formulated as a data-generating process through the following two successive steps. First, we draw the node-specific latent quantities  $U_1, \dots, U_N$  through

$$U_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1] \quad \text{for all } i = 1, \dots, N. \quad (4a)$$

Secondly, and conditional on the realizations of the latent quantities, i.e. given  $U_1 = u_1, \dots, U_N = u_N$ , we sample the edge variables independently in the form of

$$Y_{ij} \mid U_i = u_i, U_j = u_j \sim \text{Bernoulli}(w(u_i, u_j)), \quad (4b)$$

where the function  $w : [0, 1]^2 \rightarrow [0, 1]$  with  $w(u, v) = w(v, u)$  for all  $u, v \in [0, 1]$  represents the graphon. This formulation specifically implies that the network entries only depend on the latent nodal variables, which, with respect to their latent domain  $[0, 1]$ , are often also referred to as latent node positions. Given that, it is easy to see that the graphon model belongs to the class of node-specific latent variable models as defined through generic model specification (1).

**Generating Exchangeable Graphs.** The graphon model can further be used to formalize the generalized process of exchangeable arrays (cf. formulation (15) of the Appendix) in an itemized manner. This can be done by defining

$$Y_{ij} \equiv H(U_0, U_i, U_j, U_{ij}) \equiv \begin{cases} 1, & \text{if } U_{ij} \leq w_{U_0}(U_i, U_j) \\ 0, & \text{otherwise,} \end{cases}$$

### 3.1 Formulation and Interpretation

where  $\{w_{u_0}(\cdot, \cdot) : u_0 \in [0, 1]\}$  is a corresponding uncountable set of graphons from which an element is chosen conditional on  $U_0$  (Lloyd et al., 2012, Sec. 2.2). From the perspective of ergodic systems, one can alternatively formulate the distribution of a simple exchangeable random graph of infinite size through

$$\mathbb{P}(\mathbf{Y}_\infty \in \mathbb{A}) = \int \mathbb{P}(\mathbf{Y}_\infty \in \mathbb{A}; w_{U_0}(\cdot, \cdot)) \mu(dU_0) = \int_0^1 \mathbb{P}(\mathbf{Y}_\infty \in \mathbb{A}; w_{u_0}(\cdot, \cdot)) du_0$$

for any (symmetric)  $\mathbb{A} \subseteq \{0, 1\}^{\infty \times \infty}$ , where  $\mu(\cdot)$  is the probability measure induced by the standard uniform distribution, see Diaconis and Janson (2008, Thm. 1.2) and Orbanz and Roy (2015, Sec. 3–B). Apparently, when using this process to model a single network, one neither has information about the entire graphon set (instead, information is only given about  $w_{u_0}(\cdot, \cdot)$  with  $U_0 = u_0$ ) nor is one interested in inferring all the graphons. This is especially the case since we are usually “only” interested in the structure of the network at hand rather than in the structures of networks as they might result under another  $U_0$ . Still, this makes it impossible to infer the complete exchangeable array process based on a single observation. Bickel and Chen (2009) thus call the variable  $U_0$  unidentifiable for single samples (where even if the graph were infinite, it would not provide further information in that regard). The plain graphon model (4) could instead be interpreted as a constrained version of exchangeable array process (15) of the Appendix, where one assumes  $H(\cdot, \cdot, \cdot, \cdot)$  to be constant in its first argument, i.e. for any  $u_0 \in [0, 1]$  it holds that

$$H(u_0, \cdot, \cdot, \cdot) \equiv H(1, \cdot, \cdot, \cdot) \quad \text{which is due to } w_{u_0}(\cdot, \cdot) \equiv w_1(\cdot, \cdot).$$

Crane (2018) calls exchangeable arrays of this subclass to be *dissociated* since  $U_0$  becomes irrelevant. Nonetheless, random graphs that are generated by graphon process (4) generally possess the property of exchangeability (see formulation (14) of the Appendix).

**Graphon as Graph Limit Object.** This limitation on the general representation of exchangeable graphs does, however, not affect the graphon model’s ability to serve as graph limit object. As described in the Appendix, for a convergent graph sequence, the homomorphism density of any subgraph  $\mathcal{G}'$  converges towards a fixed value  $c_{\mathcal{G}'}$ . Moreover, according to the induced sampling distribution (cf. formulation (12) of the Appendix), the probability of a graph sample reflecting the connectedness of a prespecified subgraph also converges towards  $c_{\mathcal{G}'}$ . Such a connectivity behavior can be shown to apply to the graphon model as a data-generating process, under which it can also be directly quantified. More precisely, with respect to a prespecified subgraph  $\mathcal{G}' = (\mathcal{V}', \mathbf{y}')$ , let  $\mathbf{Y} = [Y_{ij}]_{i,j=1,\dots,N'}$  be the connectivity pattern of a graph  $\mathcal{G}_{N'}^{w(\cdot, \cdot)} = (\mathcal{V}_{N'}, \mathbf{Y}_{N'})$  of size  $N' = |\mathcal{V}'|$  that is generated according to process (4). Then we can formulate that

$$\mathbb{P}(Y_{ij} = 1, i < j : y'_{ij} = 1; w(\cdot, \cdot)) = \mathbb{P}(\mathcal{G}' \subseteq \mathcal{G}_{N'}^{w(\cdot, \cdot)}) = \underbrace{\int_{[0,1]^{N'}} \prod_{\substack{i < j: \\ y'_{ij}=1}} w(u_i, u_j) \prod_{i=1}^{N'} du_i}_{=: t(\mathcal{G}', w(\cdot, \cdot))}$$

where  $\mathbf{y}' = [y'_{ij}]_{i,j=1,\dots,N'}$  represents the concrete connectivity pattern of  $\mathcal{G}'$ . That means the probability that a graphon-generated graph exhibits the connectedness of prespecified pattern  $\mathcal{G}'$  can be calculated by solving the integral on the right-hand side. Based on that, one furthermore can formulate a two-stage sampling scheme by assuming  $\mathcal{G}_N^{w(\cdot, \cdot)} = (\mathcal{V}_N, \mathbf{Y}_N)$  to be generated

through graphon process (4) and  $\mathcal{G}_N^{w(\cdot, \cdot)}|_{N'}$  being a corresponding induced subgraph sample under uniform node drawing with replacement. (This can be interpreted as the reverse perspective of getting from graph limit formulation (11) to limit sampling distribution (12).) Under this setting, we get

$$\begin{aligned}
 t(\mathcal{G}', \mathcal{G}_N^{w(\cdot, \cdot)}) &= \mathbb{P}(\mathcal{G}' \subseteq \mathcal{G}_N^{w(\cdot, \cdot)}|_{N'}) \\
 &= \sum_{\mathbf{y} \in \mathcal{Y}_N} \mathbb{P}(\mathcal{G}' \subseteq \mathcal{G}_N^{w(\cdot, \cdot)}|_{N'} | \mathbf{Y}_N = \mathbf{y}) \cdot \mathbb{P}(\mathbf{Y}_N = \mathbf{y}; w(\cdot, \cdot)) \\
 &= \sum_{\mathbf{y} \in \mathcal{Y}_N} t(\mathcal{G}', \mathcal{G}_N^{w(\cdot, \cdot)}(\mathcal{V}_N, \mathbf{Y}_N = \mathbf{y})) \cdot \mathbb{P}(\mathbf{Y}_N = \mathbf{y}; w(\cdot, \cdot)) \xrightarrow{p} t(\mathcal{G}', w(\cdot, \cdot)). \tag{5}
 \end{aligned}$$

Hence, for a graphon-generated graph, we can clearly attribute the a priori probability of a random node collection to form a prespecified connectivity pattern. Note that the only reason why the expression in (5) “only” converges towards  $t(\mathcal{G}', w(\cdot, \cdot))$ —instead of being equal—is given through the possibility of drawing from  $\mathcal{G}_N^{w(\cdot, \cdot)}$  the same nodes multiple times. This effect, however, vanishes for a growing graph  $\mathcal{G}_N^{w(\cdot, \cdot)}$ . To complete the reasoning and get from result (5) to the concept of graph limits, it is easy to argue that  $\mathcal{G}_N^{w(\cdot, \cdot)}$  can be considered as the current state of a corresponding graph sequence. To demonstrate this, we start with  $\mathcal{G}_1^{w(\cdot, \cdot)} = (\mathcal{V}_1 = \{1\}, \mathbf{y}_1 = (0))$  and  $\mathbf{u}_1 = (u_1)$ , where  $u_1$  is the realization of  $U_1 \sim \text{Uniform}[0, 1]$ . For obtaining graph  $\mathcal{G}_{N+1}^{w(\cdot, \cdot)} = (\mathcal{V}_{N+1}, \mathbf{y}_{N+1})$ , we then perform the following steps:

1. Draw  $U_{N+1} \sim \text{Uniform}[0, 1]$ .
2. Conditional on  $U_{N+1} = u_{N+1}$ , draw  $Y_{i,N+1} | U_i = u_i, U_{N+1} = u_{N+1} \sim \text{Bernoulli}(w(u_i, u_{N+1}))$  for all  $i = 1, \dots, N$ .
3. Set  $\mathbf{y}_{N+1} = \begin{pmatrix} \mathbf{y}_N & \mathbf{y}_*^\top \\ \mathbf{y}_* & 0 \end{pmatrix}$  with  $\mathbf{y}_* = (y_{1,N+1}, \dots, y_{N,N+1})$ .
4. Finally, define  $\mathcal{G}_{N+1}^{w(\cdot, \cdot)} = (\mathcal{V}_N \cup \{N+1\}, \mathbf{y}_{N+1})$ .

Together with result (5), we can conclude that a graph sequence  $(\mathcal{G}_N^{w(\cdot, \cdot)} = (\mathcal{V}_N, \mathbf{Y}_N))_{N=1,2,\dots}$  that is generated using graphon process (4) is stochastically convergent (see definition (11) of the Appendix). To be precise, the homomorphism density of any subgraph  $\mathcal{G}'$  converges with respect to  $\mathcal{G}_N^{w(\cdot, \cdot)}$  towards  $t(\mathcal{G}', w(\cdot, \cdot))$ . To highlight the capability of the graphon model in this regard, we refer to the major finding of Lovász and Szegedy (2006), which is comprehensibly expressed by Lovász (2012) in Theorem 11.21: “For any convergent sequence  $[(\mathcal{G}_N)_{N=1,2,\dots}]$  of simple graphs, there exists a graphon  $[w(\cdot, \cdot)]$  such that  $[t(\mathcal{G}', \mathcal{G}_N) \rightarrow t(\mathcal{G}', w(\cdot, \cdot))]$  for every simple graph  $[\mathcal{G}']$ .” In other words, no matter how complex the structure of a convergent graph sequence is, it can be described in the form of a graphon. With regard to the property of a graph sequence being convergent, we lastly emphasize that this can be illustrated by considering the sequence of adjacency matrices. Looking at Figure 1 reveals a clearly evolving structure in the three progressive adjacency matrices on the left. Scaling the adjacency matrices towards  $[0, 1]^2$  and considering them as graphons with codomain  $\{0, 1\}$  allows for a proper definition of convergence with respect to the so-called *cut distance* (Lovász, 2012, Sec. 8). Note, however, that for this kind of reconciliation, information about the ordering of the nodes must be given. For real-world data, this can be inferred from the estimation.

### 3.2 Identifiability and Complexity

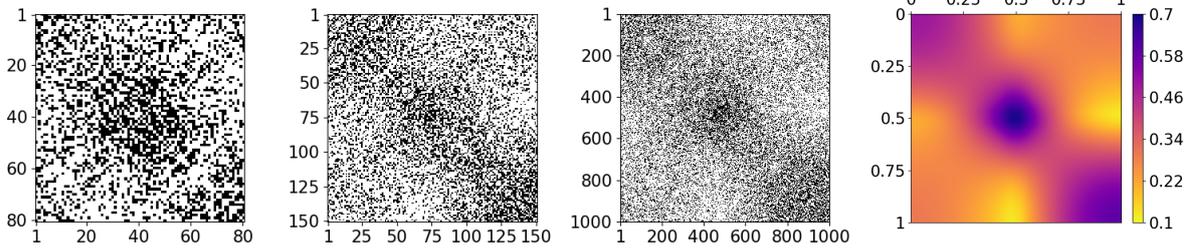


Figure 1.: Graphon as the limiting object of a convergent graph sequence. The first three plots show the (ordered) adjacency matrices of the growing graph at states with  $N = 80, 150,$  and  $1000,$  respectively. The depicted “pixel pictures” are coded with white and black for 0 and 1, respectively. The rightmost plot illustrates the corresponding graphon.

### 3.2. Identifiability and Complexity

When applying the graphon framework to describe and model network data, it is important to discuss the general identifiability and complexity of this type of model. This is specifically relevant when it comes to developing strategies for inferring the graphon model from data, as further discussed below.

**Non-Uniqueness of the Graphon Model.** An important aspect of the graphon model concerning its estimation and interpretation is the identifiability issue. With regard to the fact that the graphon model can be applied as a flexible nonparametric framework for network data (cf. Section 2.3.3, paragraph about local edge density), it is important to note that its representation is not unique. To be precise, the specification of data generating process (4) is ambiguous with respect to  $w(\cdot, \cdot)$ . This is due to the fact that for any measure-preserving bijection  $\varphi : [0, 1] \rightarrow [0, 1]$ , the jointly permuted graphon  $w'(u, v) := w(\varphi(u), \varphi(v))$  describes the exact same network model as  $w(\cdot, \cdot)$  itself. Here, “jointly permuted” refers to the two arguments of  $w(\cdot, \cdot)$  and “measure-preserving” means that  $\mu(\varphi^{-1}(\mathbb{A})) = \mu(\mathbb{A})$  for any measurable sets  $\mathbb{A} \subseteq [0, 1]$ , where  $\varphi^{-1}(\mathbb{A}) := \{a \in [0, 1] : \varphi(a) \in \mathbb{A}\}$  and  $\mu(\cdot)$  is again the probability measure induced by the standard uniform distribution. Diaconis and Janson (2008) state that the non-uniqueness in process (4) is even more complicated. More generally, they show that two graphons  $w(\cdot, \cdot)$  and  $w'(\cdot, \cdot)$  represent the same network model if and only if there exist two measure-preserving mappings  $\varphi, \varphi' : [0, 1] \rightarrow [0, 1]$ —which are not necessarily bijections—such that

$$w(\varphi(u), \varphi(v)) = w'(\varphi'(u), \varphi'(v)) \quad \text{for almost all } (u, v)^\top \in [0, 1]^2. \quad (6)$$

As a simple example to illustrate this, they consider the two graphons  $w(u, v) = uv$  and  $w'(u, v) = (2u \bmod 1)(2v \bmod 1)$  in combination with the two mappings  $\varphi(u) = 2u \bmod 1$  and  $\varphi'(u) = u$ . Under these specifications, relation (6) holds, but  $\varphi(\cdot)$  is not bijective and thus there exists no transformation  $\tilde{\varphi}'(\cdot)$  to achieve  $w'(\tilde{\varphi}'(\cdot), \tilde{\varphi}'(\cdot)) \equiv w(\cdot, \cdot)$ .

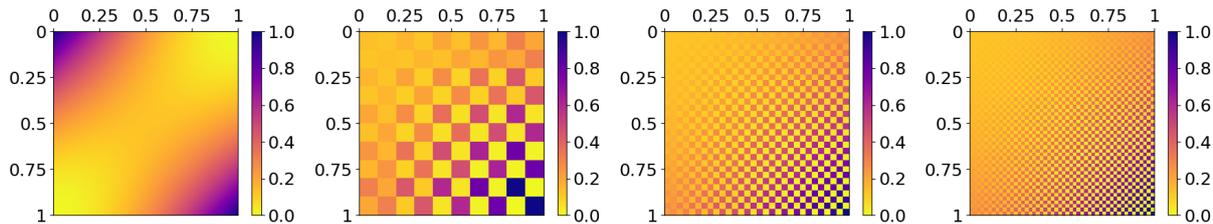


Figure 2.: Histogram approximation of graphon (8) with ordering according to canonical representation (7). Left: graphon representation according to definition. Three right plots: histogram approximation in canonical representation based on a resolution of 10, 30, and 50 blocks, respectively.

**Guaranteeing Uniqueness.** To circumvent this identifiability issue and to conceive uniqueness, some papers postulate that

$$g(u) = \int w(u, v) dv \quad \text{is strictly increasing,} \quad (7)$$

see e.g. Bickel and Chen (2009), Chan and Airoldi (2014), Yang et al. (2014), or Latouche and Robin (2016). Graphon formulations that fulfill this assumption are called to be of *canonical* form. (Since some works suggested that  $g(\cdot)$  being monotone non-decreasing would be sufficient for guaranteeing uniqueness, it seems important to stress that this is actually not true, see e.g. Borgs et al., 2010.) Note that the distribution of  $g(U_i)$  with  $U_i$  following distributional assumption (4a) can be interpreted as the (asymptotic) distribution of the degree proportion. So together, condition (7) appears to be well interpretable and, more importantly, very convenient for the estimation. Specifically, Yang et al. (2014) and Chan and Airoldi (2014) exploit the implication of  $u_i < u_j$  if  $g(u_i) < g(u_j)$  to justify their degree-based node ordering strategy. More precisely, they assume the empirical degree to be a reasonable estimate of the expected degree and thus order the rows and columns of the adjacency matrix according to the degree realizations. In this regard, condition (7) might yield only an imperfect identification, especially when the marginal function possesses a low slope. The failure to identify plausible node positions under such a setting is demonstrated in Figure 2 of Part II. In addition, see Nowicki and Snijders (2001, Sec. 4) for an argumentation against an analogous identification strategy in the stochastic blockmodel framework.

**Requirements for Graphon Estimation.** With regard to condition (7), we stress that there is an even more important aspect to consider, namely the strong restriction on the generality of the graphon model. To give an example, condition (7) excludes the model specification

$$w(u, v) = (uv)^2 + ((1-u)(1-v))^2 \quad (8)$$

since there exists no measure-preserving function  $\varphi(\cdot)$  such that  $w(\varphi(\cdot), \varphi(\cdot))$  is well-defined and fulfills condition (7). To be precise, it holds for all  $u \in [0, 1]$  that  $g(u) = g(1-u)$  while the slices  $w(u, \cdot)$  and  $w(1-u, \cdot)$  are completely different, at least for values of  $u$  that are not close to 0.5. As a potential workaround, one could choose a histogram approximation and order the blocks accordingly. This is illustrated in Figure 2. (Note that, strictly speaking, the ordering here only fulfills a *non-decreasing* profile of  $g(\cdot)$ .) This strategy, however, does not lead to a well-defined limit, and an intermediate stage with many blocks results in the loss of the smoothness of  $w(\cdot, \cdot)$ .

### 3.2 Identifiability and Complexity

---

In this regard, the smoothness property is an essential requirement to make graphon estimation feasible (Wolfe and Olhede, 2013, Gao et al., 2015). The canonical condition, on the other hand, only serves to reduce complexity and to simplify the estimation. Hence, it seems worthwhile to look out for an estimation procedure that does not require restrictive condition (7). This is exactly what we aim for in Part II of this thesis, where we allow graphons to be descended from a more general model class without the canonical constraint.

Apparently, when refusing such a strong restriction on the graphon’s marginal function, inferring the node positions becomes much more complicated. A natural approach to overcome this issue seems to be estimating  $\mathbf{U}$  and  $w(\cdot, \cdot)$  simultaneously. This, in turn, requires an iterative estimation procedure. In this regard, we want to point out potential issues that might occur when solving the estimation problem in a self-referential fashion.

**Pitfalls under Iterative Estimation.** Applying an iterative procedure for tackling graphon estimation means estimating  $\mathbf{U}$  and  $w(\cdot, \cdot)$  separately but conditional on the current state of the respective other quantity. Such an approach is generally suitable to estimate models that depend on unobserved latent variables (McLachlan and Krishnan, 2007). However, it also requires paying particular attention to possible identifiability issues as they are highly predominant in the graphon model. More precisely, as a consequence of identifiability issue (6), there is not just one ground-truth graphon  $w(\cdot, \cdot)$  that could be identified as the prescribed estimation target. Instead, one aims to estimate any graphon from its *equivalence class*. Such an equivalence class can be defined with respect to formulation (6) and the existence of corresponding measure-preserving mappings  $\varphi, \varphi' : [0, 1] \rightarrow [0, 1]$ , or, equivalently, through a set of all  $w' : [0, 1]^2 \rightarrow [0, 1]$  that fulfill

$$\mathbb{P}(\mathbf{Y}_N = \mathbf{y}; w(\cdot, \cdot)) = \mathbb{P}(\mathbf{Y}_N = \mathbf{y}; w'(\cdot, \cdot)) \quad \text{for all } \mathbf{y} \in \mathcal{Y}_N \text{ and } N \geq 2.$$

The question which then arises is whether there exists a representative in the equivalence class to which an iterative procedure might converge. In this dissertation, we apply an EM-type algorithm as a technique often used in the context of latent variables (Dempster et al., 1977). Specifically, this implies the usage of the marginal conditional expectations. Based on that, we can directly formulate a requirement on the class of graphon models for the estimation via an EM-type algorithm. That is, the ground-truth graphon needs to represent an equivalence class that includes at least one representative  $w(\cdot, \cdot)$  which is identifiable with respect to  $\mathbb{E}(U_i | \mathbf{y}; w(\cdot, \cdot))$ . This graphon representation then serves as a stationary state of the algorithm since performing the E-step conditional on  $w(\cdot, \cdot)$  would yield persistent estimates for the node positions. (Note that under a fully Bayesian approach, e.g. when employing an overall MCMC technique, this kind of identifiability issue is often referred to as *label switching problem*, see Stephens, 2000. Due to the additional stochasticity in the Bayesian context, this becomes even more complicated.) To give an intuition of what this is about, we consider the counterexamples from Figure 3. Regarding the left graphon, it exhibits a global symmetry, which is manifested in a way such that, for any observed network  $\mathbf{y}$ , it holds for any  $\mathbf{u} = (u_1, \dots, u_N) \in [0, 1]^N$  that  $f(\mathbf{u} | \mathbf{y}) = f(\mathbf{1} - \mathbf{u} | \mathbf{y})$  with  $\mathbf{1} - \mathbf{u} = (1 - u_1, \dots, 1 - u_N)$ . This especially implies that the marginal conditional distribution  $f(u_i | \mathbf{y})$  for any node  $i$  is symmetric around 0.5. Hence, this specification leads to  $\mathbb{E}(U_i | \mathbf{y}; w(\cdot, \cdot)) = 0.5$  for all  $i = 1, \dots, N$ . Consequently, the latent positions are *not* marginally identifiable. The model on the right is an analogous example of the partially symmetric case. Here, the latent positions that correspond to the segments delimited by white and black lines, respectively, are, as a whole, arbitrarily exchangeable. Nevertheless, with respect to such (partially) symmetric models, we emphasize that this kind of behavior can be viewed as exceptional.

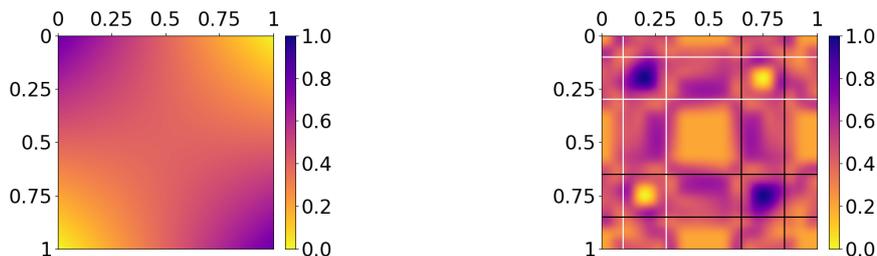


Figure 3.: Graphons that are not properly estimable with an EM-type algorithm. Left: globally symmetric graphon specification where for any  $u \in [0, 1]$  it holds that  $w(u, v) = w(1 - u, 1 - v)$  for all  $v \in [0, 1]$ . Right: partially symmetric model where segments delimited by white and black lines, respectively, are reversely symmetric, i.e.  $w(u, v) = w(\varphi(u), 1 - v)$  for all  $v \in [0, 1]$  with  $\varphi(u) = 0.85 - (u - 0.1)$  for any  $u \in [0.1, 0.3]$ .

Therefore, it should only rarely represent the ground truth for real-world networks. In fact, this *symmetry-free* condition can be interpreted as a modification of the *twin-free* condition formulated in Borgs et al. (2010). In that regard, two points  $u_i, u_j \in [0, 1]$  are called twins if  $w(u_i, \cdot) \equiv w(u_j, \cdot)$ , and  $w(\cdot, \cdot)$  is called twin-free if there exist no twins. Lastly, we emphasize that the assumption of  $w(\cdot, \cdot)$  being not symmetric in the above sense is a much less restrictive condition than canonical condition (7).

Having discussed the specific difficulties in graphon estimation, we next give a brief overview of the estimation approaches proposed so far.

### 3.3. Estimation and Inference

As for other models with latent variables, the estimation becomes particularly difficult because neither the latent node positions nor the desired graphon specification is known. In addition, unlike in ordinary models with latent variables, the node positions are mutually dependent conditional on the connectivity pattern even when the graphon is fixed, i.e.

$$[(U_{i_1}, \dots, U_{i_n}) \perp\!\!\!\perp (U_{j_1}, \dots, U_{j_{n'}})] \mid \mathbf{Y} = \mathbf{y}; w(\cdot, \cdot) \quad (9)$$

for any disjoint, non-empty subsets  $\{i_1, \dots, i_n\}, \{j_1, \dots, j_{n'}\} \subseteq \{1, \dots, N\}$  with  $n, n' \in \{1, \dots, N - 1\}$ . As a consequence of that, the node positions can be determined in a plausible way only with respect to each other. The severe identifiability issue described above exacerbates the estimation's complexity. Nonetheless, the graphon model has gained a lot of attention during the last decade, mainly stimulated by Diaconis and Janson (2008), Lovász (2012), and Bickel and Chen (2009), and various estimation procedures have been proposed in the statistical literature.

**Theoretical Approaches.** Wolfe and Olhede (2013) develop a nonparametric graphon approximation by applying converging stochastic blockmodels, see also Rohe et al. (2011) and Choi et al. (2012) for the general idea of stochastic blockmodels with a growing number of clusters. Choi and Wolfe (2014) and Choi (2017) extend this strategy towards the situation of bipartite graphs with separately exchangeable arrays. Olhede and Wolfe (2014) adopt the stochastic blockmodel approximation to formulate a histogram estimator with a global bandwidth considered as a tuning parameter. Gao et al. (2015) and Klopp et al. (2017) exploit these representation forms to discuss

### 3.3 Estimation and Inference

---

optimal convergence rates for nonparametric graphon estimation in the dense and sparse regime, respectively. To do so, they rely on constrained least squares estimators, see also Gao et al. (2016, Sec. 4.3) for (partially observed and potentially sparse) bipartite graphs or Gao et al. (2020, Sec. 5.7) for a Bayesian approach. Following these approaches, however, generally requires information about either the oracle or the optimally assigned node ordering. Except for very small networks, such an optimization is a computationally infeasible task. Nonetheless, the stochastic blockmodel approximation is generally one of the most propagated strategies in the graphon estimation literature, which can be concretely implemented by appropriate node ordering techniques.

**Feasible Blockmodeling Approaches.** An efficient implementation for fitting the histogram approximation is proposed by Amini and Levina (2018, Sec. 7). To be precise, they employ semidefinite programming approaches that naturally entail convenient regularization effects. As an alternative, Chan and Airolidi (2014) and Yang et al. (2014) focus on graphon estimation under restrictive canonical condition (7). This makes the graphon identifiable with respect to its marginal function  $g(\cdot)$  and thus allows to solve the ordering task by sorting the nodes with respect to their observed or estimated degrees. Applying the histogram estimator and potentially smoothing the resulting histogram completes their approaches. Assuming multiple observations of the connectivity pattern among a given node set, Airolidi et al. (2013) calculate distances between node pairs based on which they infer corresponding blocks. By subsequently applying the stochastic blockmodel approximation, they are able to show consistency of the corresponding graphon estimate. In this direction, Zhang et al. (2017) develop an estimation of the distances between node positions under the scenario of a single graph observation. By inferring individual local neighborhoods, they construct a consistent smoothing estimator.

**Probability Matrix Estimation.** Many papers reduce the graphon estimation task to the problem of estimating the probability matrix that generates the observed adjacency matrix in the sense of independent Bernoulli trials. To be precise, they pursue to infer information about the individual edge probabilities, i.e.

$$\mathbb{P}(Y_{ij} = 1 \mid U_i = u_i, U_j = u_j) = w(u_i, u_j),$$

which is often referred to as estimating an “inhomogeneous” Erdős-Rényi model (Klopp et al., 2017, Sec. 1). Taking this perspective, the estimation is no longer reliant on a concrete specification of the model formulation ( $\mathbf{U} = (U_1, \dots, U_N)$ ,  $w(\cdot, \cdot)$ ). Instead, for estimating the edge probabilities  $w(u_i, u_j)$ , it is sufficient to smooth over “neighboring” edges, i.e. edges with position  $(u_{i'}, u_{j'})$  such that  $\|(u_i, u_j)^\top - (u_{i'}, u_{j'})^\top\| \approx 0$ , where  $\|\cdot\|$  is the Euclidean distance. The rationale for doing so is apparently the assumption of

$$w(u_i, u_j) \approx w(u_{i'}, u_{j'}) \quad \text{if } \|(u_i, u_j)^\top - (u_{i'}, u_{j'})^\top\| \approx 0.$$

In turn, the distances of node positions  $U_i$  and  $U_j$  can, in one way or another, be inferred from the nodal connectivity behaviors  $\mathbf{y}_{i\bullet} = (y_{i1}, \dots, y_{iN})$  with  $i = 1, \dots, N$ . Note that in this context, the distances do not rely on specific node positions. As a consequence, such approaches allow for circumventing the identifiability issue described above. Among others, some of the previously mentioned works follow this narrowed perspective, see Gao et al. (2015), Klopp et al. (2017), Gao et al. (2016), Gao et al. (2020), and Zhang et al. (2017). In a different direction, Chatterjee (2015, Sec. 2.6) relies on spectral analysis and uses singular value decomposition with a universal thresholding rule for graphon estimation. Based on that, he provides error rates which implicate

consistency, see also Xu (2018). Note that the intuition of such a spectral method is to reduce the data complexity by curtailing the rank of the matrix, hence often referred to as low-rank matrix completion technique (Li et al., 2020). This can also be interpreted as smoothing out unusual appearances.

**General Shortcomings of Previous Approaches.** As outlined above, graphon estimation has been extensively elaborated so far from a theoretical point of view. Works following this perspective mostly apply the histogram approach or a more general stochastic blockmodel approximation. This specifically means to formulate the graphon as a piecewise constant, discontinuous function. In many real-world settings, however, it is more reasonable to assume a smooth transition in the connectivity behavior (Airoldi et al., 2013, Olhede and Wolfe, 2014). In addition, discontinuous estimates are often more difficult to interpret since nodes with nearby positions can be assigned to different bins, providing no direct information about the present similarity in connectivity behavior. On the other hand, most concretely implemented estimation procedures make unfavorable but crucial assumptions. This includes repeated observations for the same node set (which is only rarely the case in real-world applications) and the existence of a canonical representation, describing a strong restriction on the generality of graphon models. Moreover, estimating merely the edge probability matrix is apparently different from estimating the graphon itself and provides less global structural information. For example, it does not yield a complete picture of the relations between nodes since it usually does not specify concrete node positions on a measurable scale. Furthermore, the probability matrix estimate does not enable to simulate new networks larger than the observed one. Altogether, the previous compendium of graphon estimation routines lacks techniques for a more fundamental graphon estimation. This is addressed in Part II and Part III, where we aim to estimate smooth graphons and mixtures of smooth graphons, respectively.

### 3.4. Applications and Links to Other Models

With reference to the last section, we have seen that graphon estimation is a complex matter due to the high complexity inherent in the model. On the other hand, its great flexibility makes it a very useful tool for modeling complex networks. In fact, while graphons were initially studied as limiting objects of large graphs, they have meanwhile been demonstrated to serve as a rich nonparametric modeling framework for finite networks as well. The structural scope covered by this model class is quite large and involves diverse structural aspects. Overall, the graphon model is very powerful when it comes to analyzing real-world networks.

**Practical Applications.** The graphon model can be employed for addressing different application problems, some of which are the following. First, the model provides a well-interpretable visualization of the structure within a network, which allows to draw conclusions about the network’s composition. Second, and with regard to an interpretation on the node level, the model provides information for investigating the “position” of an actor within the network. Other node-specific characteristics, like centrality measures, can also be more accurately specified. While the empirical counterparts are naturally subject to randomness, this can be offset by relying on the (inferred) graphon model (Avella-Medina et al., 2020). Third, the graphon model can be utilized as a simple

### 3.4 Applications and Links to Other Models

---

simulation tool that allows to sample from a generative process that mirrors the data. Such a simulation approach facilitates networks of arbitrary sizes since the specification of  $w(\cdot, \cdot)$  is unrelated to  $N$ . Fourth, as for other node-specific latent variable models, estimation procedures for graphons can be constructed and implemented such that missing values can be easily handled by just removing them from the calculations. In addition, predictions for individual edges can also be easily deduced. Taken together, the graphon model can be employed as a sophisticated edge-prediction method for incomplete network data (Zhang et al., 2017). Lastly, since the graphon can be interpreted as density or intensity function on networks, this model seems particularly suitable for network comparison. This is also supported by a model specification under which the network’s size and structure are decoupled.

Note that the aspects listed above are only a selection of useful applications that by no means claims to be exhaustive. In any case, these aspects underpin the exploratory character of the graphon framework and support its applicability for analyzing network data. The first two issues mentioned above are dealt with in Part II and Part III, where the former part additionally considers aspect four. Part IV focuses on the last aspect, i.e. network comparison.

Furthermore, as a very flexible tool that is able to represent different types of network structures, graphon models are intrinsically related to other models. In particular, a more or less direct connection can be established to other node-specific latent variable models.

**Formulation of the Erdős-Rényi-Gilbert Model.** The works of Erdős and Rényi (1960) and Gilbert (1959) are one of the earliest in the field of formalizing and modeling random graphs and, together with other papers of these authors, they perhaps can be seen as laying the foundation of statistical network analysis. Considering an undirected graph of size  $N$ , Erdős and Rényi (1960) formulated a generative process where the set of edges results from drawing uniformly at random  $|\mathcal{E}|$  connections from all possible ones. Specifically, this means that, under a fixed number of present connections, all graphs with  $|\mathcal{E}|$  edges are equally likely. On the other hand, Gilbert (1959) introduced a slightly different model where each connection emerges with a fixed overall probability. Consequently, the Gilbert model involves an additional layer of randomness. Moreover, it can be exactly represented by the graphon model with  $w(\cdot, \cdot)$  being specified as globally constant at the predefined edge probability. The general assumption of this model specification is apparently an overall homogeneous behavior among all nodes. Although many theoretical properties could be derived for this model class, it turned out quickly that it is most often not suitable for modeling real-world networks. As a consequence of this shortcoming, it was the concrete intention of later models to incorporate some heterogeneity in one way or another. In this direction, also the graphon model can be interpreted as an “inhomogeneous” Erdős-Rényi-Gilbert model (Klopp et al., 2017).

**Stochastic Blockmodel Representation.** It is well known that the stochastic blockmodel can also be interpreted as a mixture of Erdős-Rényi-Gilbert models and, in fact, some works have made use of this modeling perspective, see e.g. Daudin et al. (2008). As such, the stochastic blockmodel is also covered by the graphon model and can be accordingly reformulated (see Latouche and

Robin, 2016 or, for the reverse link, Olhede and Wolfe, 2014 and Airoldi et al., 2013). To do so,  $w(\cdot, \cdot)$  is formulated as a piecewise constant step function with a rectangular pattern, i.e.

$$w(u, v) = \sum_{k=1}^K \sum_{l=1}^K \mathbb{1}_{\{\zeta_{k-1} \leq u < \zeta_k\}} \mathbb{1}_{\{\zeta_{l-1} \leq v < \zeta_l\}} p_{kl}, \quad (10)$$

where  $\zeta_k = \sum_{l=1}^k \alpha_l$  for  $k = 1, \dots, K$  and  $(\mathbf{P} = [p_{kl}]_{k,l=1,\dots,K}, \boldsymbol{\alpha} = [\alpha_k]_{k=1,\dots,K})$  is the specification of the stochastic blockmodel with  $K$  groups as defined in Section 2.3.3. In this regard, the graphon model can be seen as a generalization of the stochastic blockmodel. As a theoretical use case of this unifying perspective, Bickel and Chen (2009) derive conditions under which different types of modularities yield consistent assignments in the blockmodel framework. Regarding the practical modeling aspect, blockmodels have a long-standing tradition in applied network analysis, especially in the context of social sciences, see e.g. the collective volume edited by Scott and Carrington (2011). This applicability might, to some extent, transfer to the more general graphon model. Its estimation, however, usually does not provide information about a suitable clustering of nodes. Instead, the graphon model can be interpreted as an arrangement of nodes that—at least partially—follows a smooth transition in the connectivity behavior (which is more appropriate than a strict block structure in many situations). Nonetheless, grouping the network’s nodes is often a proclaimed goal in concrete applications and specifically convenient to interpret. Despite the graphon model’s general capability to cover block structures, so far, its estimation has not been intended to explicitly do so. We address this deficit in Part III by adapting the specification of the graphon model accordingly. First, in Chapter 6, we estimate stochastic blockmodels by relying on the corresponding graphon representation. Based on that, and with regard to model specification (10), we then relax the restriction of  $w(\cdot, \cdot)$  possessing only strictly constant plateaus, see Chapter 7. The resulting model specification consequently provides a framework that is commonly more flexible and more realistic than the stochastic blockmodel. At the same time, its estimation results still allow to infer groups of actors.

**Extension to the Degree-Corrected Stochastic Blockmodel.** As for the classical stochastic blockmodel, it can be illustrated that the graphon model also covers its degree-corrected version (Karrer and Newman, 2011). The difference between the classical and the degree-corrected blockmodel refers to the individuality of nodes’ attractiveness. While in the standard variant, the stochastic equivalence within groups also involves the expected degree, this is implemented in a node-wise individual fashion in the degree-corrected approach. The latter model thus implies incorporating a degree heterogeneity into the fundamentally equivalent behavior within blocks. To represent such a connectivity through the graphon model, we can again rely on the model specification from Chapter 7 of Part III. According to those formulations, there are potentially smooth shifts within groups that induce a varying behavior for nodes from the same block. These shifts can also be exploited to represent differences in attractiveness. To be precise, for  $k = 1, \dots, K$  one can formulate

$$w(u, \cdot) \equiv a_k(u) \cdot w(\zeta_{k-1}, \cdot) \quad \text{for all } u \in (\zeta_{k-1}, \zeta_k),$$

where  $a_k : [0, 1] \rightarrow \mathbb{R}_+$  is a continuous monotone non-decreasing function fulfilling  $a_k(\zeta_{k-1}) = 1$ . This setting entails that two nodes from the same group will reveal the same basic connectivity behavior (stochastically) but differ in the expected degree. Note that  $a_k(\cdot)$  is restricted through  $a_k(\zeta_k) \leq \min_v \{1/w(\zeta_{k-1}, v)\}$ , which is a consequence of the Bernoulli-type stochastic blockmodel

### 3.4 Applications and Links to Other Models

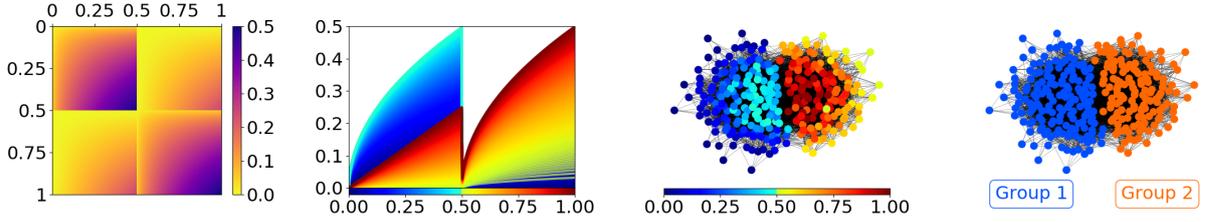


Figure 4.: Degree-corrected stochastic blockmodel with two communities of same size in graphon representation. The two left plots illustrate the graphon specification  $w(\cdot, \cdot)$  and its corresponding slices  $w(u, \cdot)$  with coloring according to  $u \in [0, 1]$ , respectively. The two plots on the right represent twice the same simulated network of size  $N = 300$ , where node colors indicate once  $u_i \in [0, 1]$  (left) and once the community membership (right). Note that, according to the representation, the node colors of the left network directly correspond to the colors of the slices, meaning that, for example, dark reddish nodes follow the behavior described by the dark red profile.

that does not occur in the Poisson variant. As a workaround, slices might not only be strict multiples of each other but differ slightly in profiles. Such a graphon specification is illustrated in Figure 4, where slices of the same group exhibit similar but shifted profiles that imply fundamentally similar connectivity behaviors with varying attractiveness.

**Approximation of the Latent Distance Model.** Furthermore, the graphon model is able to represent the latent distance model of any dimension  $J \in \mathbb{N}$  (Bickel et al., 2011, p. 2282), or at least to approximate it arbitrarily well. A concrete specification with respect to the one-dimensional distance model is given by Chan and Airolidi (2014, Sec. 5.1), see also Matias and Robin (2014, Sec. 2.2). To formulate a general connection, we again rely on the graphon specification from Chapter 7 of Part III. The intuition of this strategy is then decomposing the latent metric space  $\mathcal{S}^J$ —which is assumed to be bounded—into  $\mathcal{S}^{J-1} \times \mathcal{S}$  and subsequently partitioning  $\mathcal{S}^{J-1}$  into fine segments. If the volumes of the segments are small enough, it is reasonable to neglect distances within these segments. In this line, referring to the node positions contained therein, these segments specify the corresponding blocks of the blockwise graphon specification. The behavioral differences resulting through the distances in the direction of the  $J$ -th dimension are captured by the smooth differences within the blocks formed in the graphon model. Increasing the number of segments and letting their volumes converge towards zero improves the accuracy of the deduced graphon representation with respect to the initial latent distance model. By employing this concrete relation, the graphon model allows to approximate the latent distance model arbitrarily well. An illustration of this link is depicted in Figure 5. As latent distance model used in this example, we employ the normalized Euclidean variant of two dimensions. That means the latent space is given by  $[0, 1]^2$  and equipped with the Euclidean distance measure  $\|\cdot\|$ . Referring to model specification (2), the left plot visualizes the edge probabilities under the setting with  $\beta_0 = 0$  and  $\beta_1 = \sqrt{2}^{-1}$  and from the perspective of a node with position  $\mathbf{x}_i = (0.1, 0.45)^\top$ . In this graphic, the x-axis and y-axis represent the scales of the values  $x_{j1}$  and  $x_{j2}$ , respectively, for the reference position  $\mathbf{x}_j = (x_{j1}, x_{j2})^\top$ . The second plot illustrates the corresponding discretization under a resolution of  $K = 10$ , which seems to preserve the original structure quite accurately. For this discretized version, a corresponding graphon specification can be directly formulated, which is shown in the third plot. Comparing the edge probabilities induced by the two-dimensional latent distance model with those from the corresponding graphon approximation (rightmost plot of Figure 5)

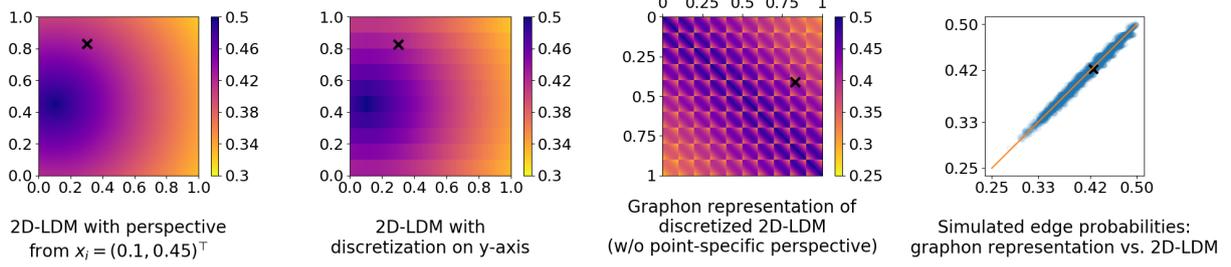


Figure 5.: Approximating the normalized two-dimensional latent distance model through the graphon model. The two left plots illustrate the latent distance model with  $\beta_0 = 0$  and  $\beta_1 = \sqrt{2}^{-1}$  (see formulation (2)) and its “discretization”, each with the perspective from  $\mathbf{x}_i = (0.1, 0.45)^\top$ . Here, the x-axis and y-axis refer to  $x_{j1}$  and  $x_{j2}$ , respectively, yielding  $\mathbf{x}_j = (x_{j1}, x_{j2})^\top$ . The third plot shows the graphon representation that serves as a corresponding approximation. On the right, simulated edge probabilities from the graphon representation are plotted against the corresponding equivalents from the latent distance model. For illustrative purposes, the black cross visualized in all plots the position of the edge probability for nodes  $i$  and  $j$  with  $\mathbf{x}_i = (0.1, 0.45)^\top$  and  $\mathbf{x}_j = (0.3, 0.83)^\top$ . Specifically, for the graphon approximation, the position is given as  $(\{\lfloor 0.45/\frac{1}{K} \rfloor + 0.1\}/K, \{\lfloor 0.83/\frac{1}{K} \rfloor + 0.3\}/K)^\top = (0.41, 0.83)^\top$  with  $K = 10$ .

reveals that a resolution of  $K = 10$  already leads to a close approximation. Note, however, that such a graphon specification does not converge towards a continuous function, which is why there is no well-defined model for  $K = \infty$ . Yet, such an approximation appears appropriate and similar results can also be achieved for latent distance models of higher dimensions. Lastly, according to this relation, one can, to some extent, transfer the interpretation of the latent distance model. That is, the latent space of the  $U_i$ , i.e. the unit interval, can be considered as reflecting an underlying latent social space, where nearby positions imply similar social attributes, at least within blocks.

**Connection to Exponential Random Graph Models.** As emphasized in the first paragraph of Section 3.1, the graphon model belongs to the latent space approaches (see Section 2.3.2). As such, it conceptually differs from approaches that explicitly focus on modeling and explaining local structural patterns like the exponential random graph model. Nonetheless, the graphon model also captures endogenous structural processes, although this is often difficult to attribute or to explicitly incorporate. More precisely, the distribution of the frequency of any motif (meaning simple finite subgraph) is uniquely characterized by the graphon model. This is described in detail by Lovász and Szegedy (2006). An intuition on that can be given by performing the calculations from (5) under an induced subgraph sampling *without* replacement for the drawing of nodes (cf. injective homomorphism density (13) of the Appendix). For this purpose, let  $\mathcal{G}_N^{w(\cdot, \cdot)} = (\mathcal{V}_N, \mathbf{Y}_N)$  be a random graph generated through graphon process (4) and  $\mathcal{G}_N^{w(\cdot, \cdot)}|_{N'}^{w/o-r}$  a corresponding induced subgraph sample of size  $N'$  under uniform node drawing *without* replacement. Then it holds that

$$\mathbb{P}(\mathcal{G}' \subseteq \mathcal{G}_N^{w(\cdot, \cdot)}|_{N'}^{w/o-r}) = t(\mathcal{G}', w(\cdot, \cdot)) = \int_{[0,1]^{N'}} \prod_{\substack{i < j: \\ y'_{ij}=1}} w(u_i, u_j) \prod_{i=1}^{N'} du_i.$$

Hence, the probability that  $N'$  nodes—which have been drawn uniformly at random and without replacement from a graphon-generated graph—exhibit the connectedness prescribed by  $\mathcal{G}'$  can be directly calculated by solving the integral on the right-hand side. (Considering exemplarily

### 3.4 Applications and Links to Other Models

---

triangular structure  $\mathcal{G}'$  from Figure 6, then the probability of drawing a triangle can be specified as  $\iiint w(u_1, u_2) w(u_1, u_3) w(u_2, u_3) du_1 du_2 du_3$ .) Given additionally that the graphon model describes a stationary random graph-generating process, the expectation for the number of motif occurrences can be directly calculated, see [Picard et al. \(2008\)](#). To be precise, and with regard to the classical notation in the exponential random graph model (see also Section 2.3.2), one can formulate that

$$\mathbb{E}(s_{\mathcal{G}'}(\mathbf{Y}_N); w(\cdot, \cdot)) = \binom{N}{N'} \rho(\mathcal{G}') t(\mathcal{G}', w(\cdot, \cdot)),$$

where  $s_{\mathcal{G}'}(\mathbf{Y}_N)$  specifies the counts of subgraph  $\mathcal{G}'$  and  $\rho(\mathcal{G}')$  is the number of non-redundant permutations of  $\mathcal{G}'$ . For example,  $\rho(\mathcal{G}')$  is three in the two-star case and one for the triangle, where possible concrete forms of realization can be specified as

- (1)  $\{\{v'_1, v'_2\}, \{v'_1, v'_3\}\}$       (2)  $\{\{v'_1, v'_2\}, \{v'_2, v'_3\}\}$       (3)  $\{\{v'_1, v'_3\}, \{v'_2, v'_3\}\}$       and
- (1)  $\{\{v'_1, v'_2, v'_3\}\}$ ,

respectively. Note that, however, deriving the full distribution of  $s_{\mathcal{G}'}(\mathbf{Y}_N)$  is rather complex. This is because different subgraphs of  $\mathcal{G}_N^{w(\cdot, \cdot)}$ , that are potential locations where the connectedness of  $\mathcal{G}'$  might occur, can overlap, which thus induces a dependency on the individual occurrence probability. Addressing this issue, [Latouche and Robin \(2016\)](#) formulate a closed-form expression of the (variational posterior) distribution of motif frequencies under particular circumstances. [Coulson et al. \(2016\)](#) apply the Stein–Chen method to achieve an appropriate Poisson approximation. Conversely, [Bickel et al. \(2011\)](#) establish a sequence of subgraph patterns that is sufficient for the graphon. Moreover, they study the asymptotic behavior when applying corresponding plug-in estimators based on the method of moments approach. Further direct connections between the graphon model and the exponential random graph model are elaborated by [Chatterjee and Diaconis \(2013\)](#), [Yin et al. \(2016\)](#), and [Krioukov \(2016\)](#). More precisely, they aim to construct graphon specifications that are closely related to exponential random graph models with basic network statistics like the number of edges, two-stars, or triangles. Overall, defining a concrete distribution for subgraph counts as in the exponential random graph model is not straightforward for the graphon model, but it is possible in principle.

Taking together all the relations outlined above, a link to all key methods of statistical network analysis can be established.

## 4. Current Research, Contributions, and Outlook

In this last introductory chapter, we briefly want to recap the role of network data, outline the current state of research in the graphon model literature, and highlight the contributions of this thesis. Lastly, we will discuss research questions that have been left open and describe further paths that could be taken.

**Emergence of Network Science.** The field of statistical network analysis has started to grow tremendously in the last two decades, especially in recent years. The popularity of modeling such data structures is driven by the fact that many systems from various disciplines can be represented as networks. Thus, the competence to study and analyze these objects offers the potential to gain deeper insights into the phenomena taking place therein. Even though the analysis of real-world networks in the form of graphs can be traced back to the first half of the 20th century (e.g. see [Moreno, 1934](#) and [Simmel and Wolff, 1950](#)), the more recent interest in this area has perhaps been stimulated in large part by [Watts and Strogatz \(1998\)](#). Already during the early stages, it became apparent quite quickly that classical statistical approaches could not be transferred directly to the network context. As a consequence, a new methodological branch arose, intending to cover the particular dependency structure inherent in such type of data. In fact, as the network science community grew and researchers worldwide started to intensively investigate network data, a vast methodological compendium has evolved. This includes a considerable ensemble of different modeling frameworks, which all have their own motivations and capabilities for covering different structural aspects. More general hurdles in the modeling context, with which research has been concerned over the last years, are graphs with directed or weighted connections, large networks, sparsity, dynamic structures, and the presence of node- or edge-wise covariates. In particular, these issues were also discussed for the graphon model, which, on the one hand, is a quite new method compared to other network approaches and, on the other hand, has so far been considered primarily as a theoretical construct rather than a practical modeling strategy.

**Current State of Graphon Research.** In its classical version, the graphon model has been introduced for simple undirected graphs without additional exogenous effects ([Lovász and Szegedy, 2006](#), [Borgs et al., 2006](#), [Bollobás et al., 2007](#)). Moreover, the first proposed approaches for fitting the model to network data described estimation routines that are based on an optimal node positioning ([Wolfe and Olhede, 2013](#), [Gao et al., 2015](#)). Such an optimization task, however, is known to be NP-hard. Thus, to make the model more applicable, several works in the graphon literature aimed for extending the model and developing feasible estimation algorithms. In this line, [Bickel and Chen \(2009\)](#) introduced a global scaling parameter depending on the network size to also allow for modeling sparse networks. A more flexible approach in this regard is given by the graphon processes of [Caron and Fox \(2017\)](#), [Borgs et al. \(2018a\)](#), and [Borgs et al. \(2019a\)](#), which extend the graphon's domain from  $[0, 1]^2$  to  $\mathbb{R}_+^2$  and assume a point process for the emergence of nodes with respect to the time scale  $\mathbb{R}_+$ . Further works that address representing graph

---

limits under the sparse regime are given by Bollobás and Riordan (2013), Borgs et al. (2015), Borgs et al. (2018b), and Borgs et al. (2019b). Moreover, an approach to extend the graphon model towards dynamic networks is presented by Pensky (2019). To formulate such a dynamic framework, the graphon can be augmented by an additional dimension for time parameter  $t$ , i.e.  $w_t(u, v) = w(u, v, t)$ . For the directed case, Diaconis and Janson (2008, Sec. 9) proposed what they call *digraphon*, a specification that basically consists of multiple graphons determining point-wise the probabilities of a multinomial-distributed dyadic event  $(Y_{ij}, Y_{ji})$ . By formulating a joint probability for the dyadic pattern, this strategy aims to take account of reciprocity directly, see also Cai et al. (2016). Lovász and Vesztegombi (2013) generalized this approach by allowing for even more categories than the four possible dyadic outcomes. A notion for limits of weighted graphs has been elaborated in Lovász (2012, Sec. 17), Diao et al. (2015), and Borgs et al. (2018a). Note that this requires first to agree on an appropriate notion of graph homomorphisms under the weighted setting, which is not unambiguous (Freedman et al., 2006). Furthermore, modeling strategies for incorporating covariates into the graphon model have been elaborated. Su et al. (2020), Mao et al. (2021), and Chandna et al. (2022) followed the intuition that the latent space of the node positions, i.e. the interval  $[0, 1]$ , reflects a compressed scale for latent (social) attributes. In this line, they formulated a model framework where similar node features imply nearby node positions and vice versa. In contrast, Latouche et al. (2018) applied a logistic regression model where, under the usage of a link function, a *residual* graphon complements the effect of covariates on the edge probability. Taking all this together, the graphon model can apparently be extended into various directions and thus presumably will still unfold a great potential to model real-world networks. To this end, the development of graphon estimation approaches has been the subject of large parts of the literature, see Section 3.3. Yet, most of these works are limited in different ways, such as taking an exclusively theoretical perspective (Olhede and Wolfe, 2014, Gao et al., 2015), considering the reduced problem of matrix estimation (Chatterjee, 2015, Zhang et al., 2017), or applying a strongly restricted model class (Chan and Airolidi, 2014, Yang et al., 2014).

**Main Contributions.** Until now, the graphon model has hardly been used by practitioners at all. For one thing, this is because estimation results are difficult to interpret. However, more importantly, estimation procedures proposed so far are not suitable for applications because they either do not provide feasible algorithms or imply too restrictive assumptions (cf. Section 3.3). In Part II, we aim to address these shortcomings by presenting a practicable estimation routine that yields reasonable results and showcasing the graphon model’s potential applicability. With regard to simulations and real-world examples, it can be clearly demonstrated that this method allows to appropriately capture the underlying network structure. One step further, the graphon model is well known to be generally able to cover diverse structural aspects. This leads to intrinsic links to the most common frameworks developed for modeling network structures, as emphasized in Section 3.4. We can show that some of these modeling approaches can even be precisely covered by the graphon model when applying appropriate representation formats. Following this direction, in Chapter 6 of Part III, we adapt the methodological approach developed in Part II to enable capturing block structures via the graphon model. These formulations are extended in subsequent Chapter 7, where we additionally incorporate smooth behavioral changes within blocks. From a different perspective, this can also be interpreted as including structural breaks into the smooth graphon model. Again, by analyzing synthetic and real-world networks, we can demonstrate that such a model specification has high capabilities to capture complex structures. As an additional use case of the model formulation in Chapter 7 of Part III, we can show that this allows to

establish graphon-based representations of further related network models. An attribute that most of these other modeling approaches lack, however, is the potential to compare networks, see Section 2.4. The graphon model, on the other hand, provides a predestined framework in this regard. In Part IV, we take advantage of this circumstance and develop a size-independent testing procedure on the equivalence of network structures. In this regard, the nonparametric comparison of networks on the basis of statistical inference has remained an unsolved challenge up to now.

**Open Issues and Future Work.** Even though many works extend the graphon model in a more applicable direction, most of these approaches cannot be readily applied in practice. This is primarily because corresponding model formulations still lack proper estimation strategies. In this regard, an iterative approach as the one proposed in Part II—which allows for estimating the graphon and the latent positions simultaneously—might also provide an appropriate approach for “upgraded” graphon models. In fact, first attempts in this direction lead to satisfying results. This includes scenarios of networks with weighted edges, dynamic structure, and covariates, where corresponding frameworks can be taken from the theoretical approaches developed in the papers above. Yet, formulating and implementing proper extensions of this estimation procedure is a complicated task in itself, at which one will face diverse difficulties. Moreover, when applying corresponding algorithms to large networks, the bottleneck is the iteration step of determining the highly dependent node positions (cf. expression (9)). Potential strategies to overcome this issue might be, for example, a variational EM approach or a limited evaluation by relying on a representative node subsample. Moreover, to achieve consistency, which is generally not provided for the EM-type algorithm, one might switch to the MCEM strategy (McLachlan and Krishnan, 2007, Sec. 6.3). That is, employing intermediate states of the Gibbs-sampled node positions individually instead of taking summary values such as marginal means. The theoretical guarantees generally provided by the MCEM algorithm might potentially also be transferable to the graphon estimation setting. Having said that, in our studies, we found that the MCEM approach results in less pronounced and blurrier outcomes, which is why we decided to utilize the EM-type approach. (Note that MCEM techniques would also allow to overcome the issue of “symmetric” graphons as a pitfall under iterative estimation, see last paragraph of Section 3.2. This is because the intermediate states of the sampling sequence of node positions are individually reliable even though their summarization is not.) Furthermore, with regard to the model formulation of Chapter 7 of Part III, it seems worthwhile to elaborate the relations to other models more thoroughly. This should also include studies about how accurately one can recover structures of theoretical graphon representations imitating other modeling frameworks. This is particularly relevant considering that it often remains unclear which modeling strategy is optimal for capturing the present network structure. To tackle this issue, Li et al. (2020) developed a cross-validation procedure for model selection in the network context, whereas Ghasemian et al. (2020) and Li and Le (2021) discussed the mixing of several model fits based on different weighting strategies. However, such computationally intensive methods could be circumvented if a unified, overall superior model could be found. Altogether, modeling and capturing the structure in complex networks remains a challenge that will continue to concern scientific communities of various fields for some time to come. This is especially so since such data structures have just demonstrated their ability to reveal fascinating phenomena in complex systems.

## Bibliography

- Airoldi, E. M., T. B. Costa, and S. H. Chan (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems 26*, pages 692–700.
- Albert, R. and A.-L. Barabási (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
- Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4), 581–598.
- Ali, W., T. Rito, G. Reinert, F. Sun, and C. M. Deane (2014). Alignment-free protein interaction network comparison. *Bioinformatics*, 30(17), i430–i437.
- Amini, A. A. and E. Levina (2018). On semidefinite relaxations for the block model. *Annals of Statistics*, 46(1), 149–179.
- Anderson, B. S., C. Butts, and K. Carley (1999). The interaction of size and density with graph-level indices. *Social networks*, 21(3), 239–267.
- Avella-Medina, M., F. Parise, M. T. Schaub, and S. Segarra (2020). Centrality Measures for Graphons: Accounting for Uncertainty in Networks. *IEEE Transactions on Network Science and Engineering*, 7(1), 520–537.
- Bassett, D. S., P. Zurn, and J. I. Gold (2018). On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, 19(9), 566–578.
- Bhattacharya, K., G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna (2008). The international trade network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(02), P02002.
- Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21068–21073.
- Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *Annals of Statistics*.
- Bollobás, B., S. Janson, and O. Riordan (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1), 3–122.
- Bollobás, B. and O. Riordan (2013). Metrics for sparse graphs. In *Surveys in Combinatorics 2009*, pages 211–288. Cambridge University Press.

- Borgs, C., J. Chayes, and L. Lovász (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geometric and Functional Analysis*, 19(6), 1597–1619.
- Borgs, C., J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi (2006). Counting graph homomorphisms. In *Topics in discrete mathematics*, pages 315–371. Springer.
- Borgs, C., J. T. Chayes, H. Cohn, and N. Holden (2018a). Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research*, 18, 1–71.
- Borgs, C., J. T. Chayes, H. Cohn, and V. Veitch (2019a). Sampling perspectives on sparse exchangeable graphs. *Annals of Probability*, 47(5), 2754–2800.
- Borgs, C., J. T. Chayes, H. Cohn, and Y. Zhao (2018b). An  $L^p$  theory of sparse graph convergence II: LD convergence, quotients and right convergence. *The Annals of Probability*, 46(1).
- Borgs, C., J. T. Chayes, H. Cohn, and Y. Zhao (2019b). An  $L^p$  theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5), 3019–3062.
- Borgs, C., J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6), 1801–1851.
- Borgs, C., J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi (2012). Convergent sequences of dense graphs II: Multiway cuts and statistical physics. *Annals of Mathematics*, pages 151–219.
- Borgs, C., J. T. Chayes, and A. Smith (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems*, volume 2015-January, pages 1369–1377.
- Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology*, 11(1), 13–41.
- Cai, D., N. Ackerman, and C. Freer (2016). Priors on exchangeable directed graphs. *Electronic Journal of Statistics*, 10(2), 3490–3515.
- Caron, F. and E. B. Fox (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79(5), 1295–1366.
- Chan, S. H. and E. M. Airoidi (2014). A consistent histogram estimator for exchangeable graph models. In *31st International Conference on Machine Learning, ICML 2014*.
- Chandna, S., S. C. Olhede, and P. J. Wolfe (2022). Local linear graphon estimation using covariates. *Biometrika*, 109(3), 721–734.
- Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1), 177–214.
- Chatterjee, S. and P. Diaconis (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*, 41(5), 2428–2461.
- Chen, K. and J. Lei (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521), 241–251.

## Bibliography

---

- Choi, D. (2017). Co-clustering of nonsmooth graphons. *Annals of Statistics*, 45(4), 1488–1515.
- Choi, D. and P. J. Wolfe (2014). Co-clustering separately exchangeable network data. *Annals of Statistics*, 42(1), 29–63.
- Choi, D. S., P. J. Wolfe, and E. M. Airoldi (2012). Stochastic blockmodels with a growing number of classes. *Biometrika*.
- Coulson, M., R. E. Gaunt, and G. Reinert (2016). Poisson approximation of subgraph counts in stochastic block models and a graphon model. *ESAIM - Probability and Statistics*, 20, 131–142.
- Crane, H. (2018). *Probabilistic Foundations of Statistical Network Analysis*. CRC Press, Boca Raton.
- Crossley, N. A., A. Mechelli, P. E. Vértes, T. T. Winton-Brown, A. X. Patel, C. E. Ginestet, P. McGuire, and E. T. Bullmore (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), 11583–11588.
- Daudin, J. J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and Computing*.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
- Diaconis, P. and S. Janson (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni*, 28, 33–61.
- Diao, P., D. Guilloit, A. Khare, and B. Rajaratnam (2015). Differential calculus on graphon space. *Journal of Combinatorial Theory. Series A*, 133, 183–227.
- Eagle, N., A. Pentland, and D. Lazer (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15274–15278.
- Emmert-Streib, F., M. Dehmer, and Y. Shi (2016). Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346-347, 180–197.
- Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17–61.
- Faisal, F. E., K. Newaz, J. L. Chaney, J. Li, S. J. Emrich, P. L. Clark, and T. Milenković (2017). Grafene: Graphlet-based alignment-free network approach integrates 3d structural and sequence (residue order) data to improve protein structural comparison. *Scientific Reports*, 7(1), 1–15.
- Fienberg, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics*, 21(4), 825–839.
- Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832–842.

- Freedman, M., L. Lovász, and A. Schrijver (2006). Reflection positivity, rank connectivity, and homomorphism of graphs. *Journal of the American Mathematical Society*, 20(1), 37–51.
- Gao, C., Y. Lu, Z. Ma, and H. H. Zhou (2016). Optimal estimation and completion of matrices with biclustering structures. *Journal of Machine Learning Research*, 17.
- Gao, C., Y. Lu, and H. H. Zhou (2015). Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6), 2624–2652.
- Gao, C., A. W. van der Vaart, and H. H. Zhou (2020). A general framework for bayes structured linear models. *Annals of Statistics*, 48(5), 2848–2878.
- Gera, R., L. Alonso, B. Crawford, J. House, J. A. Mendez-Bermudez, T. Knuth, and R. Miller (2018). Identifying network structure similarity using spectral graph theory. *Applied Network Science*, 3(1), 1–15.
- Ghasemian, A., H. Hosseinmardi, A. Galstyan, E. M. Airoidi, and A. Clauset (2020). Stacking models for nearly optimal link prediction in complex networks. *Proceedings of the National Academy of Sciences of the United States of America*, 117(38), 23393–23400.
- Gilbert, E. N. (1959). Random Graphs. *The Annals of Mathematical Statistics*, 30(4), 1141–1144.
- Girvan, M. and M. E. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826.
- Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoidi (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(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 in Society)*, 170(2), 301–354.
- Hanneke, S., W. Fu, and E. P. Xing (2010). Discrete temporal models of social networks. *Electronic journal of statistics*, 4, 585–605.
- 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. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137.
- Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33–50.
- Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*.
- Hunter, D. R., P. N. Krivitsky, and M. Schweinberger (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4), 856–882.
- Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 83(1).

## Bibliography

---

- Klopp, O., A. B. Tsybakov, and N. Verzelen (2017). Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1), 316–354.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data*. Springer, New York.
- Kolaczyk, E. D. (2017). *Topics at the frontier of statistics and network analysis*. Cambridge University Press, Cambridge.
- Kolaczyk, E. D. and G. Csardi (2014). *Statistical analysis of network data with R*. Springer, New York, 2nd edition.
- Koutra, D., N. Shah, J. T. Vogelstein, B. Gallagher, and C. Faloutsos (2016). DELTACON: Principled massive-graph similarity function with attribution. *ACM Transactions on Knowledge Discovery from Data*, 10(3).
- Koutra, D., J. T. Vogelstein, and C. Faloutsos (2013). DELTACON: A principled massive-graph similarity function. In *Proceedings of the 2013 SIAM International Conference on Data Mining, SDM 2013*, pages 162–170.
- Krioukov, D. (2016). Clustering Implies Geometry in Networks. *Physical Review Letters*, 116(20).
- Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 29–46.
- Kuchaiev, O. and N. Pržulj (2011). Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 27(10), 1390–1396.
- Latouche, P. and S. Robin (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing*, 26(6), 1173–1185.
- Latouche, P., S. Robin, and S. Ouadah (2018). Goodness of Fit of Logistic Regression Models for Random Graphs. *Journal of Computational and Graphical Statistics*, 27(1), 98–109.
- Li, T. and C. M. Le (2021). Network estimation by mixing: Adaptivity and more. *arXiv preprint arXiv:2106.02803*.
- Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika*, 107(2), 257–276.
- Liu, Q., Z. Dong, and E. Wang (2018). Cut Based Method for Comparing Complex Networks. *Scientific Reports*, 8(1).
- Lloyd, J. R., P. Orbanz, Z. Ghahramani, and D. M. Roy (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, volume 2, pages 998–1006.
- Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60. American Mathematical Society, Providence.
- Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory. Series B*, 96(6), 933–957.
- Lovász, L. and K. Vesztegombi (2013). Non-deterministic graph property testing. *Combinatorics Probability and Computing*, 22(5), 749–762.

- Lusher, D., J. Koskinen, and G. Robins (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press, Cambridge.
- Mao, X., D. Chakrabarti, and P. Sarkar (2021). Consistent Nonparametric Methods for Network Assisted Covariate Estimation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 7435–7446.
- Marron, J. S. and A. M. Alonso (2014). Overview of object oriented data analysis. *Biometrical Journal*, 56(5), 732–753.
- 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(4), 1119–1141.
- Matias, C. and S. Robin (2014). Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys*, 47, 55–74.
- McLachlan, G. J. and T. Krishnan (2007). *The EM Algorithm and Extensions: Second Edition*. John Wiley & Sons, New Jersey, 2nd edition.
- Memišević, V. and N. Pržulj (2012). C-GRAAL: Common-neighbors-based global GRAph ALignment of biological networks. *Integrative Biology (United Kingdom)*, 4(7), 734–743.
- Milenković, T., W. L. Ng, W. Hayes, and N. Pržulj (2010). Optimal network alignment with graphlet degree vectors. *Cancer Informatics*, 9, 121–137.
- Moreno, J. L. (1934). *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. Nervous and Mental Disease Publishing Company, Washington, D.C.
- Newman, M. (2018). *Networks*. Oxford University Press, Oxford.
- Newman, M. E. and G. Reinert (2016). Estimating the number of communities in a network. *Physical review letters*, 117(7), 078301.
- Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, 96(455), 1077–1087.
- Olhede, S. C. and P. J. Wolfe (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(41), 14722–14727.
- Onnela, J.-P., D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones (2012). Taxonomies of networks from community structure. *Physical Review E*, 86(3), 036104.
- Orbanz, P. and D. M. Roy (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 437–461.
- Pascual-Belda, A., A. Díaz-Parra, and D. Moratal (2018). Evaluating functional connectivity alterations in autism spectrum disorder using network-based statistics. *Diagnostics*, 8(3), 51.
- Pensky, M. (2019). Dynamic network models and graphon estimation. *The Annals of Statistics*, 47(4), 2378–2403.

## Bibliography

---

- Picard, F., J. J. Daudin, M. Koskas, S. Schbath, and S. Robin (2008). Assessing the exceptionality of network motifs. *Journal of Computational Biology*, 15(1), 1–20.
- Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2), e177–e183.
- Pržulj, N., D. G. Corneil, and I. Jurisica (2004). Modeling interactome: Scale-free or geometric? *Bioinformatics*, 20(18), 3508–3515.
- Riolo, M. A., G. T. Cantwell, G. Reinert, and M. E. Newman (2017). Efficient method for estimating the number of communities in a network. *Physical Review E*.
- 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.
- Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4), 1878–1915.
- Salter-Townshend, M., A. White, I. Gollini, and T. B. Murphy (2012). Review of statistical network analysis: Models, algorithms, and software. *Statistical Analysis and Data Mining*, 5(4), 243–264.
- Schwikowski, B., P. Uetz, and S. Fields (2000). A network of protein–protein interactions in yeast. *Nature biotechnology*, 18(12), 1257–1261.
- Scott, J. and P. Carrington (2011). *The SAGE Handbook of Social Network Analysis*. SAGE Publications, Thousand Oaks.
- Simmel, G. and K. H. Wolff (1950). *The Sociology of Georg Simmel*, volume 92892. The Free Press, New York.
- Snijders, T. and J. Koskinen (2013). Longitudinal models. In *Exponential random graph models for social networks*, pages 130–140. Cambridge University Press.
- Snijders, T. A. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37, 131–153.
- Snijders, T. A. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1), 75–100.
- 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.
- Song, Y., T. M. Epalle, and H. Lu (2019). Characterizing and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis. *Frontiers in Human Neuroscience*, 13, 203.
- Soundarajan, S., T. Eliassi-Rad, and B. Gallagher (2014). A guide to selecting a network similarity method. In *SIAM International Conference on Data Mining 2014, SDM 2014*, volume 2, pages 1037–1045.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 795–809.

- 
- Su, Y., R. K. Wong, and T. C. Lee (2020). Network Estimation via Graphon with Node Features. *IEEE Transactions on Network Science and Engineering*, 7(3), 2078–2089.
- Subbaraju, V., M. B. Suresh, S. Sundaram, and S. Narasimhan (2017). Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. *Medical Image Analysis*, 35, 375–389.
- Szemerédi, E. (1978). Regular partitions of graphs. In *Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260, pages 399–401, Paris.
- Tantardini, M., F. Ieva, L. Tajoli, and C. Piccardi (2019). Comparing methods for comparing networks. *Scientific Reports*, 9(1), 1–19.
- Wang, Y. R. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2), 500–528.
- Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3), 401–425.
- Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440–442.
- Wilson, R. C. and P. Zhu (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9), 2833–2841.
- Wolfe, P. J. and S. C. Olhede (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.
- Xu, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *35th International Conference on Machine Learning, ICML 2018*, volume 12, pages 8646–8661.
- Yang, J. J., Q. Han, and E. M. Airoldi (2014). Nonparametric estimation and testing of exchangeable graph models. In *Artificial Intelligence and Statistics*, pages 1060–1067. PMLR.
- Yaveroglu, Ö. N., N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic, and N. Pržulj (2014). Revealing the hidden language of complex networks. *Scientific Reports*, 4(1), 1–9.
- Yaveroglu, Ö. N., T. Milenković, and N. Pržulj (2015). Proper evaluation of alignment-free network comparison methods. *Bioinformatics*, 31(16), 2697–2704.
- Yin, M., A. Rinaldo, and S. Fadnavis (2016). Asymptotic quantization of exponential random graphs. *Annals of Applied Probability*, 26(6), 3251–3285.
- Zhang, Y., E. Levina, and J. Zhu (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4), 771–783.

# Appendix

## Graph Limits and Exchangeable Random Graphs

As stated in the main part of the Introduction, the graphon model is a theoretical construct that derives likewise from the theory of graph limits and the theory of exchangeable random graphs. These two general concepts follow theoretical notions about complex graphs, which as such, often appear to be reasonable and beneficial when modeling real-world networks in the form of their graph representations.

**Graph Homomorphism.** To start with, the concept of graph limits is, in turn, based on the concept of *graph homomorphisms*, meaning adjacency-preserving maps (Lovász, 2012). To be precise, a graph homomorphism from graph  $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$  to graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is defined through a node mapping  $\Gamma : \mathcal{V}' \rightarrow \mathcal{V}$  such that  $\{v'_i, v'_j\} \in \mathcal{E}'$  implies  $\{\Gamma(v'_i), \Gamma(v'_j)\} \in \mathcal{E}$ . In this context,  $\mathcal{G}'$  is said to be homomorphic to  $\mathcal{G}$ , if there exists any homomorphism from  $\mathcal{G}'$  to  $\mathcal{G}$ . In addition, the *number of homomorphisms* of  $\mathcal{G}'$  into  $\mathcal{G}$  is usually denoted by  $\text{hom}(\mathcal{G}', \mathcal{G})$ . Moreover, the *homomorphism density* as normalized homomorphism number is defined as

$$t(\mathcal{G}', \mathcal{G}) = \frac{\text{hom}(\mathcal{G}', \mathcal{G})}{|\mathcal{V}|^{|\mathcal{V}'|}} \in [0, 1].$$

To give an example, consider the two graphs  $\mathcal{G}'$  and  $\mathcal{G}$  in Figure 6. For calculating  $\text{hom}(\mathcal{G}', \mathcal{G})$ , we are interested in the number of node triplets  $(v_1, v_2, v_3)_{\mathcal{G}}$  in  $\mathcal{G}$  that mirror the connectedness in the node triple  $(1, 2, 3)_{\mathcal{G}'}$  of  $\mathcal{G}'$ . In general, with “connectedness” we mean exclusively the present, not the absent edges. Here, we find that

$$\begin{aligned} \text{hom}(\mathcal{G}', \mathcal{G}) = & |\{(1, 2, 4)_{\mathcal{G}}, (1, 4, 2)_{\mathcal{G}}, (2, 1, 4)_{\mathcal{G}}, (2, 4, 1)_{\mathcal{G}}, (4, 1, 2)_{\mathcal{G}}, (4, 2, 1)_{\mathcal{G}}, \\ & (1, 3, 4)_{\mathcal{G}}, (1, 4, 3)_{\mathcal{G}}, (3, 1, 4)_{\mathcal{G}}, (3, 4, 1)_{\mathcal{G}}, (4, 1, 3)_{\mathcal{G}}, (4, 3, 1)_{\mathcal{G}}\}| = 12. \end{aligned}$$

Based on that, we can directly calculate  $t(\mathcal{G}', \mathcal{G}) = 12/6^3 \approx 5.56\%$ .

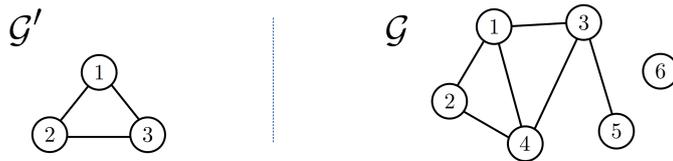


Figure 6.: Exemplary graphs to illustrate the concepts of graph homomorphism, number of homomorphisms, and homomorphism density.

**Graph Limits.** As one step further, we consider the scenario of graphs arising from (potentially stochastic) processes and, based on that, we define corresponding graph sequences as  $(\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N))_{N=1,2,\dots}$  with  $|\mathcal{V}_N| = N$ , for which holds that  $\mathcal{V}_N \subset \mathcal{V}_{N+1}$  and  $\mathcal{E}_N \subseteq \mathcal{E}_{N+1}$ . The theory of graph limits now postulates that the graph sequence  $(\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N))_{N=1,2,\dots}$  is *convergent* if the homomorphism density of any fixed and finite subgraph  $\mathcal{G}'$  with respect to  $\mathcal{G}_N$  converges (Lovász and Szegedy, 2006, Borgs et al., 2006, Borgs et al., 2008). That means, for any finite subgraph  $\mathcal{G}'$  there exists a value  $c_{\mathcal{G}'} \in [0, 1]$  such that

$$t(\mathcal{G}', \mathcal{G}_N) \longrightarrow c_{\mathcal{G}'}. \quad (11)$$

As far as it does not hold that  $c_{\mathcal{G}'} = 0$  for every  $\mathcal{G}'$  with  $|\mathcal{E}'| \geq 1$ , this apparently supposes an underlying process that describes the dense regime, i.e. under which  $|\mathcal{E}_N| = \mathcal{O}\left(\binom{N}{2}\right)$ . If the generative process is assumed to be stochastic, convergence (11) needs to be understood as convergence in probability. Note that, in fact, the graph sequence does not necessarily comprise a record for all  $N \in \mathbb{N}$ . Specifically, some processes might imply that multiple nodes emerge simultaneously. In any case, the associated graph size is assumed to converge to infinity.

**Induced Distribution on Induced Subgraph Sampling.** In general, the homomorphism density can also be interpreted as the probability that  $N' = |\mathcal{V}'|$  nodes from  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , which are chosen uniformly at random—with replacement and under consideration of the order—, exhibit the connectedness prespecified by  $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ . Thus, homomorphism densities can be used to specify corresponding sampling distributions. As a sampling strategy, we here assume the induced subgraph sampling under uniform drawing from  $\mathcal{V}$  with replacement. That means, the graph sample  $\mathcal{G}|_{N'} = (\mathcal{V}|_{N'}, \mathcal{E}|_{N'})$  consists of the sampled nodes  $v_1^*, \dots, v_{N'}^*$  and all the edges to be found in  $\mathcal{G}$  among these nodes. Nodes that are drawn repeatedly are just labeled separately, where the edges connected to the origin nodes will be included for all repeated nodes. To be precise, let  $\Gamma^* : \mathcal{V}|_{N'} \rightarrow \mathcal{V}$  be the node assignment resulting from the random draw. Then  $\{v_i^*, v_j^*\} \in \mathcal{E}|_{N'}$  if and only if  $\{\Gamma^*(v_i^*), \Gamma^*(v_j^*)\} \in \mathcal{E}$ . Under this setting, the sampling distribution induced by the homomorphism densities can be defined through

$$\mathbb{P}(\{v_i^*, v_j^*\} \in \mathcal{E}|_{N'}, i < j : \{v'_i, v'_j\} \in \mathcal{E}') = t(\mathcal{G}', \mathcal{G}).$$

Combined with the graph limit concept, that means that, for convergent graph sequence  $(\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N))_{N=1,2,\dots}$ , the probability of  $\mathcal{G}_N|_{N'}$  mirroring the connectedness within  $\mathcal{G}'$  converges towards  $c_{\mathcal{G}'}$  (where we still assume that the sample size  $N'$  reflects the size of  $\mathcal{G}'$ ). In formulae, that can be expressed as

$$\mathbb{P}(\mathcal{G}' \subseteq \mathcal{G}_N|_{N'}) = t(\mathcal{G}', \mathcal{G}_N) \longrightarrow c_{\mathcal{G}'}, \quad (12)$$

where  $\mathcal{G}' \subseteq \mathcal{G}_N|_{N'}$  means that the connectedness of  $\mathcal{G}'$  is reflected in  $\mathcal{G}_N|_{N'}$ , i.e.

$$\{v'_i, v'_j\} \in \mathcal{E}' \Rightarrow \{v_i^*, v_j^*\} \in \mathcal{E}|_{N'} \quad \text{for all } i, j = 1, \dots, N' \text{ with } i < j.$$

**Applicability of Graph Limit Theory.** When describing or analyzing networks represented as graphs, the utility of the graph limit concept is two-fold. On the one hand, this theoretical perspective allows to conceptualize graph structures in a more customized way and thus to establish new approaches for solving specific problems. In fact, “many results [e.g. about regularity] can be stated and proved for graphons in a more natural and cleaner way.” (Lovász,

2012, p. 18.) On the other hand, and with regard to a more practical evaluation, the frequencies of local patterns can be used to specify sensible model-based processes. This is especially suitable because these frequencies already provide much information about the inherent structure, even if the consideration is restricted to a few prescribed subgraph patterns. For instance, again consider the triangular pattern on the left-hand side of Figure 6, generally defined as  $\mathcal{G}' = (\{1, 2, 3\}, \{\{1, 2\}, \{1, 3\}, \{2, 3\}\})$ . Setting its frequency in relation to the frequency of two-stars (defined as  $\mathcal{G}'' = (\{1, 2, 3\}, \{\{1, 2\}, \{1, 3\}\})$ ) can then be used to quantify transitivity, which in turn serves as an adequate measure to assess the triadic closure effect. Focusing on the frequency of subgraphs is also the strategy that motivates the exponential random graph model (Robins et al., 2007).

**Link to Subgraph Frequencies.** Note that the homomorphism density is not equivalent to the (relative) subgraph frequency. This is simply because graph homomorphisms are not necessarily injective; plus, for the homomorphism number, all isomorphic pattern repetitions are counted separately. The latter source of deviation can be easily overcome by dividing  $\text{hom}(\mathcal{G}', \mathcal{G})$  by the number of isomorphic transformations. This might be useful to avoid redundancy in the counting process and to get results that are easier to interpret. For example, with regard to the triangle example from Figure 6, considering  $\{1, 2, 4\}_{\mathcal{G}}$  as a single occurrence of pattern  $\mathcal{G}'$  rather than treating all six isomorphic representations as individual ones might seem more intuitive. On the other hand, the non-injective nature can be solved by introducing a corresponding counterpart. Let therefore  $\Gamma_{\text{inj}}(\cdot)$  be a node-mapping as before but injective, and  $\text{inj}(\mathcal{G}', \mathcal{G})$  be the corresponding number of *injective* homomorphisms from  $\mathcal{G}'$  to  $\mathcal{G}$ . Then we can define the injective homomorphism density as

$$t_{\text{inj}}(\mathcal{G}', \mathcal{G}) = \frac{\text{inj}(\mathcal{G}', \mathcal{G})}{|\mathcal{V}|! / (|\mathcal{V}| - |\mathcal{V}'|)!}, \tag{13}$$

see Borgs et al. (2008). Analogously, this can be interpreted as the probability that  $N'$  nodes that are randomly chosen from  $\mathcal{G}$  *without* replacement possess the connectedness of  $\mathcal{G}'$ . We stress, however, that the difference between  $t(\mathcal{G}', \mathcal{G})$  and  $t_{\text{inj}}(\mathcal{G}', \mathcal{G})$  vanishes for a growing graph since the number of possibilities for choosing the same nodes multiple times gets negligible in comparison with all possible sample combinations. Taking additionally the number of isomorphic transformations into account allows for formulating a direct connection to subgraph frequencies. Again, with regard to the graphs from Figure 6, the absolute and relative frequency of triangles in  $\mathcal{G}$  can be calculated as  $12/6 = 2$  and  $(12/6)/(120/6) = 10\%$ , respectively.

Overall, we emphasize that graph limit theory provides powerful tools and opens up new opportunities for analyzing networks. Moreover, the limiting object resulting from this theory—which can be defined as the graphon (see main part)—can be used as an approximation of a large dense graph. At that, the limiting object is often easier to handle for performing analytic calculations.

**Exchangeable Random Graphs.** Apparently, when dealing with graph limits, the connectivity-related objective is the *relative* structural pattern rather than the absolute one. More precisely, for specifying the homomorphism densities, one is *not* interested in how concrete nodes contribute to specific subgraph frequencies. A differentiation of nodes might be carried out *a posteriori* according to their structural role in the network but in no way *a priori*. Hence, the concrete labels of the nodes are of no meaning and can be arbitrarily exchanged. This concept is thus called *exchangeability*. In that sense, applying a (node-)exchangeable model means that the perspective

on the network is always the same, no matter which nodal viewpoint is taken. With regard to an underlying generative process  $\mathcal{M}$  (also interpretable as a concrete model specification), exchangeability can be formulated as the probabilistic invariance with respect to arbitrary node relabeling, i.e.

$$\mathbb{P}(\mathbf{Y} = [y_{ij}]_{i,j=1,\dots,N}; \mathcal{M}) = \mathbb{P}(\mathbf{Y} = [y_{\psi(i)\psi(j)}]_{i,j=1,\dots,N}; \mathcal{M}) \quad (14)$$

for any permutation  $\psi : \{1, \dots, N\} \rightarrow \{1, \dots, N\}$ . This concept of exchangeability can be further transferred to infinite binary arrays  $\mathbf{Y} = [Y_{ij}]_{i,j=1,2,\dots}$ , which, in the network context, would represent the adjacency matrix of an infinite graph. The infinite array  $\mathbf{Y} = [Y_{ij}]_{i,j=1,2,\dots}$  is then called (*jointly*) *exchangeable* if for any  $N \in \mathbb{N}$

$$\mathbb{P}(Y_{ij} = y_{ij}, i, j \in \{1, \dots, N\}) = \mathbb{P}(Y_{ij} = y_{\psi(i)\psi(j)}, i, j \in \{1, \dots, N\})$$

for all  $[y_{ij}]_{i,j=1,\dots,N} \in \{0, 1\}^{N \times N}$  and all permutations  $\psi(\cdot)$  on  $\{1, \dots, N\}$ . Hoover (1979) and Aldous (1981) found out that the generative process of any exchangeable random  $\{0, 1\}$ -valued array  $\mathbf{Y} = [Y_{ij}]_{i,j=1,2,\dots}$  can be described as follows. Let  $U_0$ ,  $U_i$ , and  $U_{ij}$  with  $i, j = 1, 2, \dots$  be independently Uniform $[0, 1]$ -distributed random variables, then there exists a corresponding function  $H : [0, 1]^4 \rightarrow \{0, 1\}$  such that one can formulate

$$Y_{ij} \equiv H(U_0, U_i, U_j, U_{ij}). \quad (15)$$

For symmetric arrays, which are the objects emerging in the undirected graph setting, one additionally assumes symmetry in the second and third arguments, i.e.  $H(\cdot, u_i, u_j, \cdot) \equiv H(\cdot, u_j, u_i, \cdot)$  for all  $u_i, u_j \in [0, 1]$ . Moreover, to prevent self-loops, one specifies  $H(\cdot, u_i, u_i, \cdot) \equiv 0$ . In the context of graphs,  $U_i$  and  $U_{ij}$  correspond to node-specific and edge-specific variables, respectively, whereas  $U_0$  represents a global effect. In honor of the founders of this theory, proposition (15) is nowadays often referred to as the Aldous-Hoover theorem.

Given the notions of graph limits and exchangeable random graphs, this brings us directly to the graphon model. In this sense, it is demonstrated in the main part of the Introduction that the graphon model derives from the generic process (15) and, at the same time, addresses graph limit properties (11) and (12).



**Part II.**

# **Smooth Graphon Estimation**

## 5. Smooth Graphon Estimation

### Contributing Article.

Sischka, B. and Kauermann, G. (2022). EM-Based Smooth Graphon Estimation Using MCMC and Spline-Based Approaches. *Social Networks*, 68, 279–295. doi:10.1016/j.socnet.2021.08.007.

### Software Implementation.

The method developed and formulated in the paper is implemented in a free and open source Python package that is publicly available on GitHub:

<https://github.com/BenjaminSischka/GraphonPy.git>

Moreover, all data sets used for demonstrating the applicability of our approach are freely accessible. For information on concrete sources, see the specifications in the paper.

### Copyright Information.

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Author Contributions.

As the original motivation of this paper, Göran Kauermann came up with the initial idea of improving the sorting-and-smoothing algorithm of Chan and Airola (2014) by adjusting the node order in accordance with the resulting graphon estimate. To develop such a resorting strategy, he proposed to apply MCMC techniques. Benjamin Sischka extended this idea towards an iterative EM-based estimation routine, where he suggested estimating the node positions and the graphon with respect to each other until convergence is achieved. In addition, Benjamin Sischka proposed to replace the former histogram estimator with a smooth estimator (first implemented methods were kernel smoothing and kriging). Göran Kauermann eventually elaborated the B-spline regression approach. Putting the graphon model into the context of social network analysis has primarily been worked out by Benjamin Sischka. Also the implementation and application of the method were carried out by Benjamin Sischka, which involves the simulation studies (Section 4) and real-world network analyses (Section 5). Both authors were extensively involved in structuring and writing the article.



# EM-based smooth graphon estimation using MCMC and spline-based approaches

Benjamin Sischka\*, Göran Kauermann

Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany

## ARTICLE INFO

### Keywords:

Graphon model  
EM algorithm  
MCMC  
Gibbs sampling  
B-spline surface  
Social network  
Political network  
Connectome

## ABSTRACT

This paper proposes the estimation of a smooth graphon model for network data analysis using principles of the EM algorithm. The approach considers both variability with respect to ordering the nodes of a network and smooth estimation of the graphon by nonparametric regression. To do so, (linear) B-splines are used, which allow for smooth estimation of the graphon, conditional on the node ordering. This provides the M-step. The true ordering of the nodes arising from the graphon model remains unobserved and MCMC techniques are employed to obtain position samples conditional on the network. This yields the E-step. Combining both steps gives an EM-based approach for smooth graphon estimation. Unlike common other methods, this procedure does *not* require the restriction of a monotonic marginal function. The proposed graphon estimate allows to explore node-ordering strategies and therefore to compare the common degree-based node ranking with the ordering conditional on the network. Variability and uncertainty are taken into account relying on the MCMC sequences. Examples and simulation studies support the applicability of the approach.

## 1. Introduction

The analysis of network data has achieved increasing interest in the last years. Goldenberg et al. (2010), Hunter et al. (2012), Fienberg (2012), Snijders (2011), and Salter-Townshend et al. (2012), respectively, published survey articles demonstrating the state of the art in the field. We also refer to Kolaczyk (2009), Kolaczyk and Csárdi (2014), and Lusher et al. (2013) for monographs in the field of statistical network data analysis, see also Kolaczyk (2017). The statistical workhorse models for explaining structural patterns in static network data are *exponential random graph models* (ERGM), *stochastic blockmodels* (SBM) and *latent distance models*. More explicitly, ERGMs make use of an exponential family distribution to model the network's adjacency matrix based on frequencies of local structural patterns. This model class was proposed by Frank and Strauss (1986) and is extensively discussed in Snijders et al. (2006). In comparison, SBMs as well as latent distance models make use of nodal latent quantities to determine specifically pairwise connection probabilities between nodes, which is done with reference to a parameterized model specification. These model classes were proposed in their fundamental forms by Holland et al. (1983) and Hoff et al. (2002), respectively. Regarding the SBM, we also refer to Snijders and Nowicki (1997) and Nowicki and Snijders (2001) for the

introduction of a *posteriori* blockmodeling.

A different modeling strategy, which is also based on nodal latent quantities, results from comprehending the network adjacency matrix  $Y = [Y_{ij}]_{i,j=1,\dots,N} \in \{0,1\}^{N \times N}$  to be generated by a *graphon model*. This model as data generating process comes into play by assuming that we draw  $N$  random variables

$$U_1, \dots, U_N \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1] \quad (1)$$

and, given  $U_i$  and  $U_j$ , simulate the network entries  $Y_{ij}$  conditionally independently through

$$Y_{ij} | U_i = u_i, U_j = u_j \sim \text{Binomial}(1, w(u_i, u_j)). \quad (2)$$

In this context, the function  $w(\cdot, \cdot)$  is called a *graphon* (=graph function). In case of undirected networks, we additionally require symmetry so that  $Y_{ij} = Y_{ji}$  for  $i < j$ , and hence in principle we assume  $w(u, v) = w(v, u)$ . This is what we focus on in this work. (Note that although the extension to directed networks seems convincing, one would need to take care of consequences concerning the theory of graph limits and exchangeable random graphs from which the graphon model concept is originated, see Lovász and Szegedy, 2006 or Diaconis and Janson, 2008.) We also stick with the commonly used convention of excluding self-loops, which

\* Corresponding author.

E-mail addresses: [benjamin.sischka@stat.uni-muenchen.de](mailto:benjamin.sischka@stat.uni-muenchen.de) (B. Sischka), [goeran.kauermann@stat.uni-muenchen.de](mailto:goeran.kauermann@stat.uni-muenchen.de) (G. Kauermann).

<https://doi.org/10.1016/j.socnet.2021.08.007>

Received 20 August 2020;

Available online 26 August 2021

0378-8733/© 2021 Elsevier B.V. All rights reserved.

means  $Y_{ii} = 0$  for  $i = 1, \dots, N$ . In general, this model specification implies a dense network. To be precise, the number of edges increases proportionally to the squared number of nodes. Note that there also exist extensions for sparse networks. For example, [Bickel and Chen \(2009\)](#) make use of a global scale parameter which depends on the network size. However, such straightforward modifications often have other drawbacks like the prevention of finite but densely connected clusters or hubs. This is why here, we stick with the initial model formulation.

The graphon model is of practical use for different application problems. First, the model yields a convenient visualization of the network and its embedded structure, which allows characterizing the network’s organization and composition. In this regard, the graphon estimate can be used to compare networks of different sizes. Secondly, at the actor level, the model enables to investigate the “position” of a node in the network. Besides, other measures of node relevance such as centrality can also be derived faithfully, even in the presence of uncertainty under which the empirical counterparts might fail ([Avella-Medina et al., 2018](#)). Thirdly, the graphon model provides a simple simulation tool for sampling from a general model that mirrors the data. In this light, with regard to incomplete network data, missing links can be easily predicted ([Zhang et al., 2017](#)). All these aspects underpin the exploratory character of the graphon approach and support its applicability for network data analysis.

In order to properly apply the graphon model and correctly interpret the results in the later sections, the interpretation of the model is subsequently discussed in more detail. In addition, we illustrate particular difficulties and common pitfalls that need to be considered when developing a suitable estimation routine for the graphon model.

### 1.1. Interpretability and comparison with other models

The graphon model is a very flexible tool for representing different types of network structures. Hence, it is inherently related to other models, especially to those that are also based on latent variables. In this regard, the graphon model can be seen as a generalization of the SBM if smoothness of the graphon is neglected. The associated graphon is discontinuous because the SBM was developed to cover blockwise structural equivalence, which intrinsically describes a discrete, i.e., non-smooth concept. To be precise, any blockmodel can be represented as a graphon model by formulating  $w(\cdot, \cdot)$  as a piecewise constant step function with a rectangular pattern, as illustrated by [Latouche and Robin \(2016\)](#), for example. The height and size of the rectangles are determined by the edge probabilities and community proportions in the given SBM, respectively. As a theoretical use case of this unifying perspective, [Bickel and Chen \(2009\)](#) derive conditions under which modularities yield consistent assignments in the blockmodel framework. Regarding the practical data modeling aspect, blockmodels have a long-standing tradition in applied network analysis, especially in the context of social sciences, see e.g. the collective volume edited by [Scott and Carrington \(2011\)](#). Hence, seeing the graphon model in this light provides an intuition for its applicability and emphasizes its field of application. This perspective can be broadened by relaxing the graphon representation as a strict step function. More precisely, continuous shifts within the blocks can be incorporated, yielding an extension of the SBM framework. The *smooth* graphon model, on the other hand, technically excludes strict block structures but can be interpreted as a partitioning of the network with smooth transitions in the connectivity behavior, which is often more plausible (especially when many actors would fall between clusters). Furthermore, the smooth graphon model involves simple latent distance models (e.g., see [Matias and Robin, 2014](#), Section 2.2 or [Chan and Airoldi, 2014](#), Section 5.1). From this perspective, the latent space  $[0, 1]$  of the  $U_i$  can be considered as a covering of, for example, social space spanned by (unobserved) social attributes, as motivated by [Hoff et al. \(2002\)](#). In general, supposing that  $w(\cdot, \cdot)$  is smooth implies that two nodes  $i$  and  $j$  with  $u_i \approx u_j$  are associated with similar slices, i.e.,  $w(u_i, v) \approx w(u_j, v)$  for all  $v \in [0, 1]$ . Therefore, a fundamental implication

of the smooth graphon approach is a structural similarity (in probabilities) of nodes which are close with respect to the latent quantity. However, the changes in connectivity are smooth (as opposed to SBMs), and edge probabilities are not monotonic in relation to the distance of latent quantities (as opposed to latent distance models). In this light, node positions cannot simply be interpreted as community memberships and do not directly reflect the relationships between actors. Rather, they specify a more sophisticated (local) position within the network.

In addition to the points made above, the more conceptual issue of endogenous network structure in graphon models is discussed in Section A.1 of the Appendix. This also includes a comparison with ERGMs. Thus, a connection can be established to all key methods of statistical network analysis.

### 1.2. Identifiability

As a very flexible approach for modeling network data, the graphon model suffers from non-identifiability, meaning that  $w(\cdot, \cdot)$  is not unique with regard to the data generating process (2). This is because, for any measure-preserving bijection  $\varphi: [0, 1] \rightarrow [0, 1]$ , the permuted graphon  $w'(u, v) = w(\varphi(u), \varphi(v))$  yields the same network model as  $w(\cdot, \cdot)$  itself. More generally, as has been stated by [Diaconis and Janson \(2008\)](#), two graphons  $w(\cdot, \cdot)$  and  $w'(\cdot, \cdot)$  represent the same generating model if and only if there exist two measure-preserving mappings – not necessarily bijections –  $\varphi$  and  $\varphi': [0, 1] \rightarrow [0, 1]$  such that  $w(\varphi(u), \varphi(v)) = w'(\varphi'(u), \varphi'(v))$  for almost all  $(u, v) \in [0, 1]^2$ . Some papers therefore add a further attribute to achieve uniqueness, see e.g. [Bickel and Chen \(2009\)](#), [Chan and Airoldi \(2014\)](#), or [Yang et al. \(2014\)](#). The common setting to do so is to postulate that

$$g(u) = \int w(u, v)dv \tag{3}$$

is strictly increasing in  $u$ , which leads to the so-called *canonical representation* of the graphon  $w(\cdot, \cdot)$ . Note that the distribution of  $g(U_i)$  with  $U_i$  following (1) can be interpreted as the (asymptotic) distribution of the degree proportion. However, the additional condition (3) implies a strong restriction of the model’s generality. This appears especially in combination with the common assumption of smoothness, meaning that  $w(\cdot, \cdot)$  satisfies (at least piecewise) some Lipschitz or Hölder condition, e.g. see [Olhede and Wolfe \(2014\)](#), [Chan and Airoldi \(2014\)](#), or [Gao et al. \(2015a\)](#). For instance, considering the smooth graphon  $w(u, v) = (uv)^2 + ((1-u)(1-v))^2$ , it holds for all  $u \in [0, 1] \setminus \{0.5\}$  that  $g(u) = g(1-u)$  while the slices  $w(u, \cdot)$  and  $w(1-u, \cdot)$  are completely different. Therefore, a rearrangement with respect to fulfilling (3) does not yield a well-defined graphon in this case, which implies that the data generating model does not possess a canonical representation. Regarding the general intention of these graphon attributes, the smoothness of  $w(\cdot, \cdot)$  is required to make graphon estimation feasible, while the canonical condition yields a unique representation and therefore reduces the complexity. However, the simplified arrangement of  $w(\cdot, \cdot)$  under (3) is only required for certain estimation procedures. Specifically, approaches that rely on a degree-based node ordering exploit the implication of  $g(u_i) < g(u_j) \Rightarrow u_i < u_j$ , meaning that the higher the expected degree (given by  $N \cdot g(u_i)$ ), the higher the latent quantity. Considering that this also applies asymptotically to the empirical degrees provides a justification for the degree-based ranking strategy. As an advantage in terms of generalization, this is not required for our approach, as will be illustrated in the following sections. We therefore consider the more general class of graphon models without the canonical constraint. This is a substantial advance compared to all methods requiring the canonical form because a smooth transition in the connectivity behavior only in the direction of an increasing degree does not seem to be a plausible assumption for real-world networks in general. Still, to enable equal conditions, we additionally discuss estimation under restriction (3).

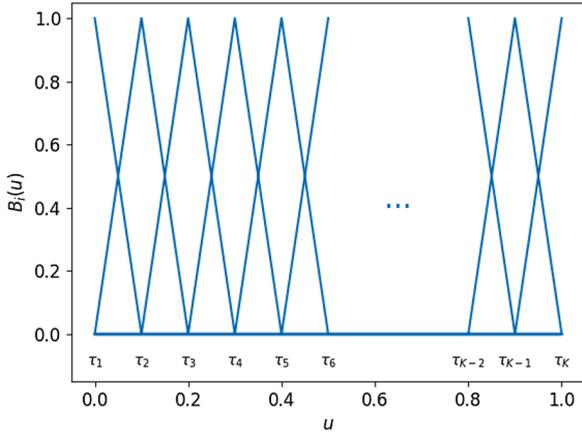


Fig. 1. Normalized univariate linear B-spline basis, used as input for the tensor product operation to reformulate the graphon  $w(\cdot, \cdot)$  as a B-spline function. The (equidistant) inner knots are denoted by  $\tau_j$  with  $j = 1, \dots, K$ .

### 1.3. Developments in graphon estimation

The estimation of graphon models has recently found attention in the statistical literature. Wolfe and Olhede (2013) and Yang et al. (2014) discuss nonparametric graphon estimation including tests on the validity of prespecified graphon shapes, see also Bickel and Chen (2009), Airolidi et al. (2013), Chan and Airolidi (2014), and Olhede and Wolfe (2014). In all these works, the formulated estimation procedure is based on approximating the graphon by an SBM, often also called histogram estimator. This is generally one of the most propagated strategies in the graphon estimation literature, see also Choi and Wolfe (2014) and Choi (2017). Gao et al. (2015a) discuss optimal graphon estimation for this kind of approximation. SBMs are, however, discontinuous models by definition, meaning that the associated graphon is discontinuous, i.e. not smooth. In this paper, in contrast, we focus on smooth graphon estimation. For a general discussion on graphon models, we refer to Borgs et al. (2008), Lovász (2012), Diaconis and Janson (2008), Bickel and Chen (2009), and Orbanz and Roy (2014). Further, You (2020) lately launched an  $\mathbb{R}$  package for graphon estimation.

Expanding the scope of the objective, graphon estimation can generally be reduced to probability matrix estimation, where the goal is to gain information about the specific edge probabilities  $P(Y_{ij} = 1 | U_i = u_i, U_j = u_j) = w(u_i, u_j)$ . From this perspective, the  $U_i$  are not well-defined components. (Instead, only their mutual proximity is employed.) Thus, it allows circumventing the identifiability issue mentioned above, which results from a model formulation that is not unique with respect to those  $U_i$ . Following this intuition, Chatterjee (2015) provides convergence results for general matrix estimation using singular value decomposition, see also Xu (2017). Gao et al. (2015a) and Klopp et al. (2017) make use of constrained least squares estimators for the edge probabilities to establish optimal rates of convergence in theory. Gao et al. (2016) extend this towards partially observed matrices/networks with biclustering structures, see also Gao et al. (2015b) for a Bayesian approach. Zhang et al. (2017) develop a routine to estimate the edge probability matrix by applying neighborhood smoothing. Su et al. (2020) build upon this approach to incorporate nodal covariates, where they assume that nearby nodes with respect to  $U_i$  have similar node features and vice versa. However, all those works about probability matrix estimation describe a procedure to estimate specific graphon values at unspecific positions. This does apparently not lead to a smooth graphon estimate on the domain  $[0, 1]^2$ , which is the focus of this work.

In this paper, we propose to use *penalized linear B-splines* for graphon estimation since this easily allows to accommodate side constraints for the resulting estimate, e.g., symmetry in the form of  $w(u, v) = w(v, u)$  or condition (3) if required. This, in contrast, is difficult to accommodate in

histogram or kernel-based estimation. For this B-spline regression approach, we borrow ideas suggested in Kauermann et al. (2013) for copula estimation. In this context, penalized estimation with B-splines has a long-standing tradition in data smoothing, starting with Eilers and Marx (1996) and Ruppert et al. (2003, 2009), see also Wood (2017b). Here, we extend this idea to graphon estimation. However, for smoothing the network data, we need information about the latent  $U_i$ , which in turn can only be estimated in relation to  $w(\cdot, \cdot)$ . Since the B-spline regression coefficients and the latent quantities  $U_i$  need to be estimated simultaneously, this is a typical task for an *EM-type algorithm*. Having a closer look at the E-step, *MCMC techniques* can be applied to approximate the complex conditional distribution of the  $U_i$  numerically. Together, this yields an *MCEM algorithm*. We recommend applying this procedure to networks with roughly a hundred to a thousand nodes, which should ensure both satisfying results and an acceptable bound for computational complexity.

The rest of the paper is organized as follows. Section 2 displays the main ideas of pursuing an EM-based algorithm for smooth graphon estimation. Section 3 describes the procedure in detail. Sections 4 and 5 showcase results for both simulations and real-world data examples, respectively. A discussion concludes the paper.

## 2. Graphon representation and EM motivation

We assume that the graphon  $w : [0, 1]^2 \rightarrow [0, 1]$  is a smooth function, meaning that it satisfies some Lipschitz condition in the sense that there exists a constant  $M \geq 0$  such that

$$|w(u, v) - w(u', v')| \leq M \|(u, v)^\top - (u', v')^\top\| \quad \text{for all } u, u', v, v' \in [0, 1] \quad (4)$$

with  $\|\cdot\|$  being the Euclidean norm. We say that  $w(\cdot, \cdot)$  has a canonical representation if  $g(u) = \int w(u, v) dv$  is strictly increasing. We further assume that  $w(\cdot, \cdot)$  is symmetric and generates a network of size  $N$  through the following process. For  $N$  independently and uniformly distributed variables

$$U_i \stackrel{i.i.d.}{\sim} \text{Uniform}[0, 1], \quad i = 1, \dots, N,$$

we obtain the conditionally independent and symmetric network through

$$P(Y_{ij} = 1 | U_i = u_i, U_j = u_j) = w(u_i, u_j) \quad (5)$$

for  $1 \leq i < j \leq N$ , where  $Y_{ji} = Y_{ij}$  and  $Y_{ii} = 0$ . Regarding the modeling perspective, the variables  $U_i$  usually remain unobservable and as data we only obtain the observed network adjacency matrix  $\mathbf{y} = [y_{ij}]_{i,j=1,\dots,N}$ . Hence, the estimation of  $w(\cdot, \cdot)$  involves the exploration of these latent variables. We therefore propose to tackle the estimation of  $w(\cdot, \cdot)$  by using an EM algorithm. Specifically, we calculate (or rather approximate by simulation) the expected value  $E(\mathbf{U} | \mathbf{y})$ , representing the E-step. This, in turn, allows the estimation of  $w(\cdot, \cdot)$  by using smoothing techniques, representing the M-step. For the E-step, we look at the conditional distribution of  $\mathbf{U} = (U_1, \dots, U_N)$  given  $\mathbf{Y}$ . Since  $U_1, \dots, U_N$  are independently  $\text{Uniform}[0, 1]$ -distributed, for  $\mathbf{u} = (u_1, \dots, u_N)$ , we obtain

$$f(\mathbf{u} | \mathbf{y}) \propto \prod_{\substack{i,j \\ j>i}} w(u_i, u_j)^{y_{ij}} (1 - w(u_i, u_j))^{1-y_{ij}}.$$

If we look at the univariate distribution of a single variable  $U_k$  given the entire network  $\mathbf{Y}$ , this results through

$$f_k(u_k | \mathbf{y}) \propto \int \dots \int \prod_{\substack{i,j \\ j>i}} w(u_i, u_j)^{y_{ij}} (1 - w(u_i, u_j))^{1-y_{ij}} du_1 \dots du_{k-1} du_{k+1} \dots du_N. \quad (6)$$

Apparently, both the joint and the marginal conditional distributions are too complex to be calculated analytically, in particular if  $N$  is large. We will therefore explore (6) by pursuing an MCMC approach, from which estimates of the latent values can easily be derived. Subsequently, in the M-step, we smooth the data relying on these current estimates for  $U_1, \dots, U_N$ . To do so, we make use of penalized B-spline regression, meaning that in (5), we express the graphon  $w(\cdot, \cdot)$  as (approximate) B-spline function  $\mathbf{B}(\cdot, \cdot)\boldsymbol{\theta}$  and then estimate the vector of unknown spline coefficients  $\boldsymbol{\theta}$  by maximizing the penalized likelihood. Here,  $\mathbf{B}(\cdot, \cdot) = \mathbf{B}(\cdot) \otimes \mathbf{B}(\cdot)$  represents a fixed bivariate spline basis constructed as tensor product of univariate linear spline bases, see Fig. 1. In theory, we require that there exists a parameter specification  $\boldsymbol{\theta}$  such that the true  $w(\cdot, \cdot)$  is approximated sufficiently well by  $\mathbf{B}(\cdot, \cdot)\boldsymbol{\theta}$ . This, however, is guaranteed due to the smoothness assumption from (4) since any smooth function can be approximated through the B-spline representation with arbitrary accuracy (if the dimension of the basis is chosen high enough). Both the E- and the M-step will be introduced in detail below. Before doing so, however, we propose a simple first E-step which, in other approaches, is the presented estimate but here should serve as initialization of the EM algorithm.

In recent literature, the need of the latent  $U_i$  for estimating  $w(\cdot, \cdot)$  is usually circumvented by smoothing the ordered observed adjacency matrix  $\mathbf{y}$ . Yet, in doing so, the preceding rearrangement of the rows and columns of the adjacency matrix should somehow reflect the ordering of the  $U_i$ . To be precise, this proceeding pursues to imitate the permutation  $\psi: \{1, \dots, N\} \rightarrow \{1, \dots, N\}$  under which  $U_{\psi(i)} \leq U_{\psi(j)}$  for  $i < j$ , implying that  $U_{\psi(i)} = U_{(i)}$  with  $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N)}$  defining the ordered variables  $U_i$ . However, since the  $U_i, i = 1, \dots, N$ , are not observable, we also have no direct information about  $\psi(\cdot)$ , which therefore needs to be estimated differently. A common approach to do so is to make use of the degree, i.e., to exploit the permutation  $\hat{\psi}: \{1, \dots, N\} \rightarrow \{1, \dots, N\}$  under which

$$\text{degree}(\hat{\psi}(i)) \leq \text{degree}(\hat{\psi}(j)) \tag{7}$$

for  $i < j$ . This ordering strategy is applied directly for graphon estimation by Yang et al. (2014) and Chan and Airolidi (2014). The latter authors even derive asymptotic pointwise convergence rates for  $\hat{\psi}(\cdot)$  towards  $\psi(\cdot)$  in the sense that  $|\psi(k) - \hat{\psi}(k)|/N \xrightarrow{P} 0$  for all  $k = 1, \dots, N$  under  $N \rightarrow \infty$ . However, that being said, applying  $\hat{\psi}(\cdot)$  as an estimate for  $\psi(\cdot)$  requires the assumption from (3) of a strictly increasing  $g(\cdot)$ , and even under the canonical condition, it might lead to unsatisfying results. We therefore consider it as a starting estimate in the initial E-step which will be improved in further iterations. Thus, we are also able to compare the degree ranking proposed by others and our final EM-based ordering. Note that B-spline regression, in contrast to matrix smoothing techniques, requires specific positions of the latent quantities. Hence, in order to make the degree-based ordering applicable, we derive an initial estimate for  $U_k$  through

$$\hat{u}_k^{(0)} = \frac{\text{rank}(\text{degree}(k))}{N + 1}, \tag{8}$$

where  $\text{rank}(\text{degree}(k))$  is the rank from smallest to largest of the  $k$ th element of the tuple  $(\text{degree}(i): i = 1, \dots, N)$ . At that, tied ranks are broken up and assigned uniquely to consecutive integer values in random order. This is equivalent to defining  $\hat{u}_{\psi(k)}^{(0)} = k/(N + 1)$ , where  $i/(N + 1), i = 1, \dots, N$ , represent the expected values of  $N$  ordered independently Uniform[0, 1]-distributed variables. These estimates can be considered as an eligible initialization which allows to properly proceed with the EM algorithm. Besides, for an initial M-step, we can replace  $w(\cdot, \cdot)$  in (5) by its empirical version  $\hat{w}^{(0)}(\cdot, \cdot)$ , which we define by

$$\hat{w}^{(0)}(u, v) = y_{\hat{\psi}([uN])\hat{\psi}([vN])},$$

where  $[uN]$  represents the smallest integer value greater or equal to  $uN$ .

Note that  $\hat{w}^{(0)}(\cdot, \cdot)$  just mimics the degree-ordered adjacency matrix scaled towards the unit square. Taken together, these calculations provide the initial estimates in the EM algorithm introduced in the next section.

### 3. EM algorithm for smooth graphons

#### 3.1. MCMC approach for the E-step

We pursue the E-step by exploiting the conditional distribution of  $U$ . This is done by constructing an appropriate MCMC Gibbs sampling scheme based on the full-conditional distribution of  $U_k$ . Note that by conditioning on  $\mathbf{Y}$  and all  $U_i$  except for  $U_k$ , one gets

$$f_k(u_k | u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_N, \mathbf{y}) \propto \prod_{j \neq k} w(u_k, u_j)^{y_{kj}} (1 - w(u_k, u_j))^{1-y_{kj}}. \tag{9}$$

Additionally following the standard setting in the E-step and pretending that the graphon  $w(\cdot, \cdot)$  is known, this allows us to easily draw from (9) using Gibbs sampling. To do so, we assume  $\mathbf{u}^{<t>} = (u_1^{<t>}, \dots, u_N^{<t>})$  to be the current state of the Markov chain. To update the  $k$ th component, we then set  $u_i^{<t+1>} := u_i^{<t>}$  for  $i \neq k$ , while  $u_k^{<t+1>}$  is obtained by drawing from (9). For this purpose, we make use of a normal proposal using a logit link. To be specific, let  $z_k^{<t>} = \log(u_k^{<t>}/(1 - u_k^{<t>})) = \text{logit}(u_k^{<t>})$ . We then propose to draw  $z_k^* = z_k^{<t>} + \text{Normal}(0, \sigma^2)$  under an appropriate choice for the variance  $\sigma^2$  and set  $u_k^* = \text{logit}^{-1}(z_k^*) = \exp(z_k^*)/(1 + \exp(z_k^*))$ . Hence, the proposal density for  $U_k$  is proportional to

$$q(u_k^* | u_k^{<t>}) = \frac{\partial u_k^*}{\partial z_k^*} \phi(z_k^* | z_k^{<t>}) \propto \frac{1}{u_k^*(1 - u_k^*)} \exp\left(-\frac{1}{2} \frac{(\text{logit}(u_k^*) - \text{logit}(u_k^{<t>}))^2}{\sigma^2}\right),$$

where  $\phi(\cdot)$  is the standard normal density. Consequently, the ratio of proposals equals

$$\frac{q_k(u_k^{<t>} | u_k^*)}{q_k(u_k^* | u_k^{<t>})} = \frac{u_k^*(1 - u_k^*)}{u_k^{<t>}(1 - u_k^{<t>})}.$$

The proposed value  $u_k^*$  is accepted (which means setting  $u_k^{<t+1>} := u_k^*$ ) with probability

$$\min\left\{1, \prod_{j \neq k} \left[ \left( \frac{w(u_k^*, u_j^{<t>})}{w(u_k^{<t>}, u_j^{<t>})} \right)^{y_{kj}} \left( \frac{1 - w(u_k^*, u_j^{<t>})}{1 - w(u_k^{<t>}, u_j^{<t>})} \right)^{1-y_{kj}} \right] \frac{u_k^*(1 - u_k^*)}{u_k^{<t>}(1 - u_k^{<t>})} \right\}$$

If we do not accept  $u_k^*$ , we set  $u_k^{<t+1>} := u_k^{<t>}$ . This update strategy will be performed for all components  $k = 1, \dots, N$  throughout the sequence  $t + 1, \dots, t + N$ . The consecutive continuation of successively updating all components completes the MCMC sampling. Based on the resulting Markov chain, we can then approximate the marginal conditional mean  $E(U_k | \mathbf{y})$  by taking the sample mean of the simulated values, observing an appropriate burn-in phase of the Gibbs sampler. To be specific, we use the MCMC sequence to estimate the conditional mean in the  $m$ th iteration of the EM algorithm through

$$\bar{u}_k^{(m)} = \frac{1}{n} \sum_{s=1+b}^{n+b} u_k^{<sN-r>}, \tag{10}$$

where  $b \in \mathbb{N}$  represents a burn-in parameter,  $r \in \mathbb{N}$  describes a thinning factor, and  $n$  is the number of MCMC states which are taken into account. We then use the ranks of the conditional means to reorder the network matrix accordingly. This corresponds to setting the value of  $U_k$  according to (8) to

$$\hat{u}_k^{(m)} = \frac{\text{rank}(\hat{u}_k^{(m)})}{N + 1} \tag{11}$$

We denote the final estimate resulting from (11) after convergence of the EM algorithm by  $\hat{u}_k^{\text{EM}}$ . Returning to the fact that  $w(\cdot, \cdot)$  is unknown, in the next section, we describe its estimation conditional on the results from the E-step, which provides the M-step.

### 3.2. Spline-based graphon estimation for the M-step

#### 3.2.1. Linear B-spline regression

For smooth estimation of the graphon  $w(\cdot, \cdot)$ , we first formulate a spline-based approximation through

$$w_\theta^{\text{spline}}(u, v) = \mathbf{B}(u, v)\boldsymbol{\theta} = [\mathbf{B}(u) \otimes \mathbf{B}(v)]\boldsymbol{\theta}, \tag{12}$$

where  $\otimes$  is the Kronecker product and  $\mathbf{B}(u) \in \mathbb{R}^{1 \times K}$  is a linear B-spline basis on  $[0, 1]$ , normalized to have a maximum value of one, cf. Fig. 1.

The parameter vector  $\boldsymbol{\theta} \in \mathbb{R}^{K^2}$  is indexed through

$$\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1K}, \theta_{21}, \dots, \theta_{K1}, \dots, \theta_{KK})^\top.$$

Using (12), we obtain the likelihood

$$l(\boldsymbol{\theta}) = \sum_{\substack{ij \\ j \neq i}} [y_{ij} \log(\mathbf{B}_{ij}\boldsymbol{\theta}) + (1 - y_{ij}) \log(1 - \mathbf{B}_{ij}\boldsymbol{\theta})],$$

where  $\mathbf{B}_{ij} = \mathbf{B}(u_i) \otimes \mathbf{B}(u_j)$ . Taking the derivative leads to the score function

$$s(\boldsymbol{\theta}) = \sum_{\substack{ij \\ j \neq i}} \mathbf{B}_{ij}^\top \left( \frac{y_{ij}}{w_\theta^{\text{spline}}(u_i, u_j)} - \frac{1 - y_{ij}}{1 - w_\theta^{\text{spline}}(u_i, u_j)} \right)$$

Moreover, taking the expected second order derivative leads to the Fisher matrix

$$\mathbf{F}(\boldsymbol{\theta}) = \sum_{\substack{ij \\ j \neq i}} \mathbf{B}_{ij}^\top \mathbf{B}_{ij} [w_\theta^{\text{spline}}(u_i, u_j) \cdot (1 - w_\theta^{\text{spline}}(u_i, u_j))]^{-1}.$$

Our intention is to maximize  $l(\boldsymbol{\theta})$ , which could be done by Fisher scoring. The resulting maximizer does, however, not lead to a proper estimate by default, meaning to fulfill symmetry and boundedness. Furthermore, in case we aim to estimate a graphon with canonical representation, we need to incorporate the constraint from (3). We therefore impose additional (linear) side constraints on  $\boldsymbol{\theta}$ . Considering the canonical condition, we get the marginal function from (12) through

$$g_\theta^{\text{spline}}(u) = \left[ \mathbf{B}(u) \otimes \int_0^1 \mathbf{B}(v) dv \right] \boldsymbol{\theta}. \tag{13}$$

For normalized B-splines, we can easily calculate the integral, and for a standardized basis with equidistant knots, we obtain

$$\begin{aligned} \int_0^1 \mathbf{B}(v) dv &= \left( \int_0^1 B_1(v) dv, \int_0^1 B_2(v) dv, \dots, \int_0^1 B_K(v) dv \right) \\ &= \underbrace{\frac{1}{K-1} \left( \frac{1}{2}, 1, \dots, 1, \frac{1}{2} \right)}_{=: \mathbf{A}}. \end{aligned}$$

This allows rewriting (13) as  $g_\theta^{\text{spline}}(u) = [\mathbf{B}(u) \otimes \mathbf{A}]\boldsymbol{\theta}$ . Hence, the marginal function  $g_\theta^{\text{spline}}(\cdot)$  is also expressed as a (univariate) linear B-spline function and a monotonicity constraint is easily accommodated by postulating monotonicity at the knots  $\tau_1, \dots, \tau_K$ . That is to say, we need

$$g_\theta^{\text{spline}}(\tau_l) - g_\theta^{\text{spline}}(\tau_{l-1}) > 0 \Leftrightarrow [(\mathbf{B}(\tau_l) - \mathbf{B}(\tau_{l-1})) \otimes \mathbf{A}]\boldsymbol{\theta} > 0 \tag{14}$$

for  $l = 2, \dots, K$ , which is a linear constraint on the coefficient vector. Imposing symmetry on the graphon can also easily be accommodated as linear constraints in the form of  $\theta_{pq} = \theta_{qp}$  for  $p \neq q$ . Finally, we need  $w_\theta^{\text{spline}}(\cdot, \cdot)$  to be bounded to  $[0, 1]$ , which is again a linear constraint in the form of  $0 \leq \theta_{pq} \leq 1$ . All in all, we can formulate the side constraints as  $\mathbf{C}\boldsymbol{\theta} \geq (\mathbf{0}, -\mathbf{1})^\top$  and  $\mathbf{D}\boldsymbol{\theta} = \mathbf{0}$  for matrices  $\mathbf{C}$  and  $\mathbf{D}$  chosen accordingly, where the constraint from (14) to impose a canonical representation can be added as desired (with  $>0$  then being replaced by  $\geq 0$ ). Incorporating the above linear constraints into the task of maximizing  $l(\boldsymbol{\theta})$ , we obtain a quadratic programming problem, which can be solved using standard software (see e.g. Andersen et al., 2004 or Turlach and Weingessel, 2013).

#### 3.2.2. Penalized estimation

Following the motivation and idea underlying penalized spline estimation (see Eilers and Marx, 1996 or Ruppert et al., 2009), we additionally impose a penalty on the coefficients to achieve smoothness. This is necessary since we intend to choose a large  $K$ , and unpenalized estimation will lead to wiggled estimates. To do so, we penalize the difference between “neighboring” elements of  $\boldsymbol{\theta}$ . Let therefore

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(K-1) \times K}$$

be the first order difference matrix. We then penalize  $[\mathbf{L} \otimes \mathbf{I}]\boldsymbol{\theta}$  and  $[\mathbf{I} \otimes \mathbf{L}]\boldsymbol{\theta}$ , where  $\mathbf{I}$  is the identity matrix of size  $K$ . This leads to the penalized likelihood

$$l_P(\boldsymbol{\theta}, \lambda) = l(\boldsymbol{\theta}) - \frac{1}{2} \lambda \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta},$$

where  $\mathbf{P} = (\mathbf{L} \otimes \mathbf{I})^\top (\mathbf{L} \otimes \mathbf{I}) + (\mathbf{I} \otimes \mathbf{L})^\top (\mathbf{I} \otimes \mathbf{L})$  and  $\lambda$  serves as smoothing parameter. The corresponding penalized score function is given through

$$s_P(\boldsymbol{\theta}, \lambda) = s(\boldsymbol{\theta}) - \lambda \mathbf{P} \boldsymbol{\theta}$$

and the penalized Fisher matrix in the form of

$$\mathbf{F}_P(\boldsymbol{\theta}, \lambda) = \mathbf{F}(\boldsymbol{\theta}) + \lambda \mathbf{P}.$$

Following this methodology, the estimate apparently depends on the penalty parameter  $\lambda$ , which is expressed in the notation. Setting  $\lambda \rightarrow 0$  yields an unpenalized fit, while setting  $\lambda \rightarrow \infty$  leads to a constant graphon, i.e. an Erdős-Rényi model. Therefore, a data-driven approach is necessary in order to determine the smoothing parameter  $\lambda$ . Here, we follow Kauermann et al. (2013) and make use of the Akaike Information Criterion (AIC) (Hurvich and Tsai, 1989, see also Burnham and Anderson, 2010). To do so, we define the corrected AIC through

$$\text{AIC}_c(\lambda) = -2l(\hat{\boldsymbol{\theta}}_P) + 2 \text{df}(\lambda) + \frac{2 \text{df}(\lambda)(\text{df}(\lambda) + 1)}{(N(N-1) - \text{df}(\lambda) - 1)},$$

where  $\hat{\boldsymbol{\theta}}_P$  is the penalized parameter estimate and  $\text{df}(\lambda)$  represents the degrees of freedom of the model. We define the latter in the common way as the trace of the product of the inverse penalized Fisher matrix and the unpenalized Fisher matrix, see Wood (2017a, page 211 ff.), i.e.,

$$\text{df}(\lambda) = \text{tr} \left\{ \mathbf{F}_P^{-1}(\hat{\boldsymbol{\theta}}_P, \lambda) \mathbf{F}(\hat{\boldsymbol{\theta}}_P) \right\}$$

with  $\text{tr}(\mathbf{M})$  being the trace of a matrix  $\mathbf{M}$ . We subsequently denote by  $\hat{w}^{(1)}(\cdot, \cdot)$  and  $\hat{w}^{\text{EM}}(\cdot, \cdot)$  the penalized B-spline estimates of  $w(\cdot, \cdot)$  in the first and the final EM iteration (i.e. after convergence), meaning that  $\hat{w}^{(1)}(\cdot, \cdot)$  and  $\hat{w}^{\text{EM}}(\cdot, \cdot)$  are based on  $\hat{u}^{(0)} = (\hat{u}_1^{(0)}, \dots, \hat{u}_N^{(0)})$  and  $\hat{u}^{\text{EM}} = (\hat{u}_1^{\text{EM}}, \dots, \hat{u}_N^{\text{EM}})$ .

**Table 1**

Exemplary graphons considered for simulations. For the second graphon specification,  $\Phi(\cdot)$  and  $F_{\mathcal{N}(0,0.25)}(\cdot)$  denote the cumulative distribution functions of the normal distribution with parameterization  $(\mu, \sigma^2) = (0, 1)$  and  $(0, 0.25)$ , respectively.

ID	Graphon
1	$w_1(u, v) = 0.8(1 - u)(1 - v) + 0.85(u \cdot v)$
2	$w_2(u, v) = 0.5 \cdot \{F_{\mathcal{N}(0,0.25)}(\Phi^{-1}(u)) \cdot F_{\mathcal{N}(0,0.25)}(\Phi^{-1}(v)) + [1 - F_{\mathcal{N}(0,0.25)}(\Phi^{-1}(u))] \cdot [1 - F_{\mathcal{N}(0,0.25)}(\Phi^{-1}(v))]\}$

...,  $\hat{u}_N^{\text{EM}}$ ), respectively. Hence, these two estimates allow to evaluate the gain achieved by our EM approach in comparison with the unprocessed degree ordering proposed by others. As is well known, the EM algorithm can be trapped at local maxima of the likelihood (or, in this case, at local minima of the corrected AIC). Thus, depending on the specific data situation, it might be recommendable to repeat the algorithm several times to achieve an optimal fit.

**3.3. Information on ranking**

As discussed earlier, the  $U_i$  are uniformly distributed and it is helpful to order them such that  $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N)}$ . Considering the degree-based ordering  $\hat{\psi}(\cdot)$  from (7) as appropriate representative that provides a permutation-invariant labeling (under appropriate handling of ties), this allows to define the (degree-related) ranking density  $f_{\hat{\psi}^{(k)}}(u_{\hat{\psi}^{(k)}} | \mathbf{y})$ . For the sake of simplicity, we collapse  $\hat{\psi}^{(k)}$  to  $(k)$  so that henceforth  $f_{(k)}(u_{(k)} | \mathbf{y})$  and  $U_{(k)}$  refer to the node with the  $k$ th lowest degree. (For

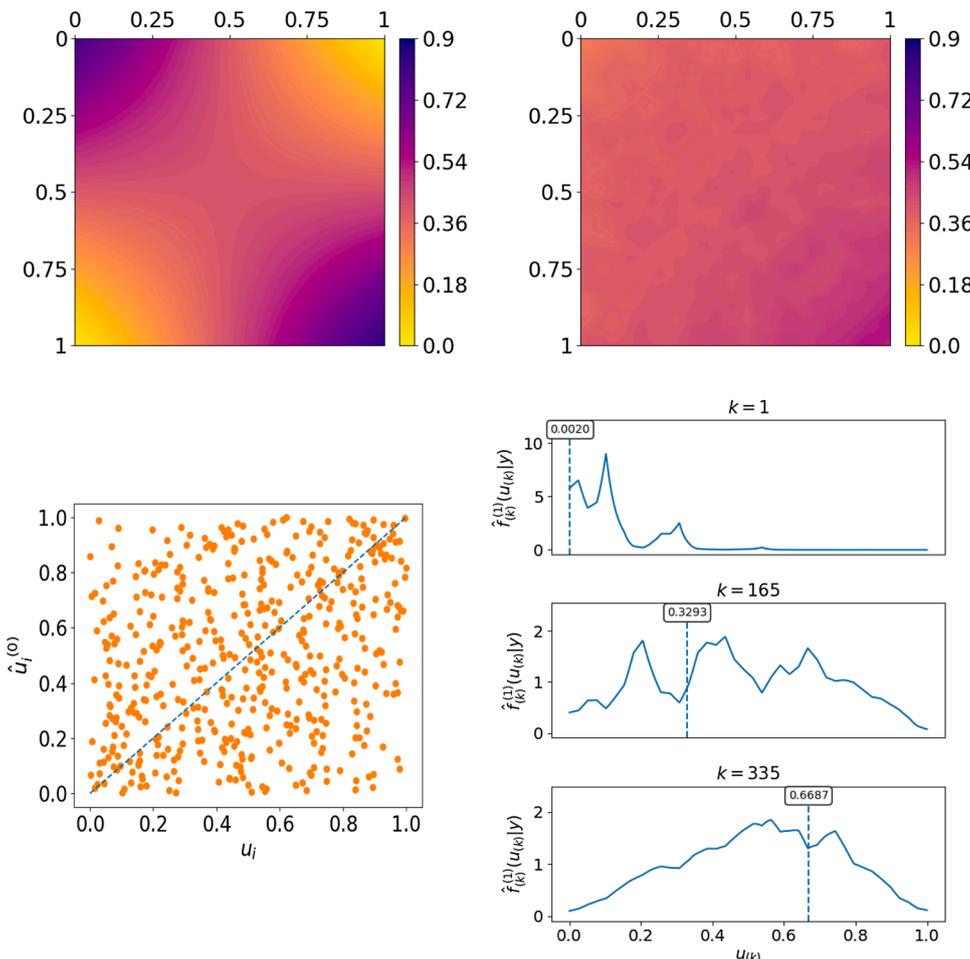
other quantities, the notation applies accordingly.) This ranking density, however, is again difficult or even impossible to calculate analytically. For a numerical approximation, the full-conditional density of  $U_{(k)}$  can be given as

$$f_{(k)}(u_{(k)} | \mathbf{u}_{(-k)}, \mathbf{y}) \propto \prod_{i \neq k} w(u_{(k)}, u_{(i)})^{y_{(k)(i)}} (1 - w(u_{(k)}, u_{(i)}))^{1 - y_{(k)(i)}}, \tag{15}$$

where the indices of  $\mathbf{u}_{(-k)} = (u_{(1)}, \dots, u_{(k-1)}, u_{(k+1)}, \dots, u_{(N)})$  and  $y_{(k)(i)}$  refer to the node labeling *after* the permutation according to the degree-based ordering. In that regard, the MCMC sequence (after burn-in phase and appropriate thinning) provides information about the conditional distribution of  $\mathbf{U}$  given the network  $\mathbf{Y}$ . Hence, for the marginal conditional distribution of  $U_{(k)}$ , we can follow a Monte Carlo integration approach, see Gelfand and Smith (1990, Sections 2.2 and 2.3), and calculate

$$f_{(k)}(u_{(k)} | \mathbf{y}) \approx \frac{1}{n} \sum_{s=1}^{n+b} f_{(k)}(u_{(k)} | \mathbf{u}_{(-k)}^{<s>}, \mathbf{y}), \tag{16}$$

where  $\mathbf{u}_{(-k)}^{<s>} = (u_{(1)}^{<s>}, \dots, u_{(k-1)}^{<s>}, u_{(k+1)}^{<s>}, \dots, u_{(N)}^{<s>})$  is the  $t$ th state of the Gibbs sampling sequence without the  $k$ th component *after* degree-related permutation and  $b$ ,  $r$ , and  $n$  are interpreted as in (10). For this purpose, the unspecified normalizing constant in (15) can be approximated by a Riemann sum since we assume  $w(\cdot, \cdot)$  to fulfill certain continuity properties. Moreover, again considering  $w(\cdot, \cdot)$  as unknown, we employ  $\hat{w}^{(1)}(\cdot, \cdot)$  or  $\hat{w}^{\text{EM}}(\cdot, \cdot)$  as estimate in (15) and denote by  $\hat{f}_{(k)}^{(1)}(\cdot | \mathbf{y})$  or  $\hat{f}_{(k)}^{\text{EM}}(\cdot | \mathbf{y})$  the resulting outcome of (16), respectively. This allows to directly assess the appropriateness of both the degree-based and the EM-based node



**Fig. 2.** Graphon estimation based on linear B-splines (including the canonical restriction from (14) and  $\hat{u}^{(0)}$  for Graphon 1 (top left) from Table 1. The “one-step” graphon estimate  $\hat{w}^{(1)}(\cdot, \cdot)$  given a simulated network of size  $N = 500$  is depicted at the top right. The plot at the bottom left illustrates the estimated  $\hat{u}_i^{(0)}$  versus the true simulated  $u_i$ . The three lower right plots show the approximated conditional distribution of  $U_{(k)}$  (based on the MCMC sequence and with respect to the given graphon estimate) for some selected indices. The dashed vertical lines (see also numbers in the box annotations) represent the estimates  $\hat{u}_{(k)}^{(0)}$ .

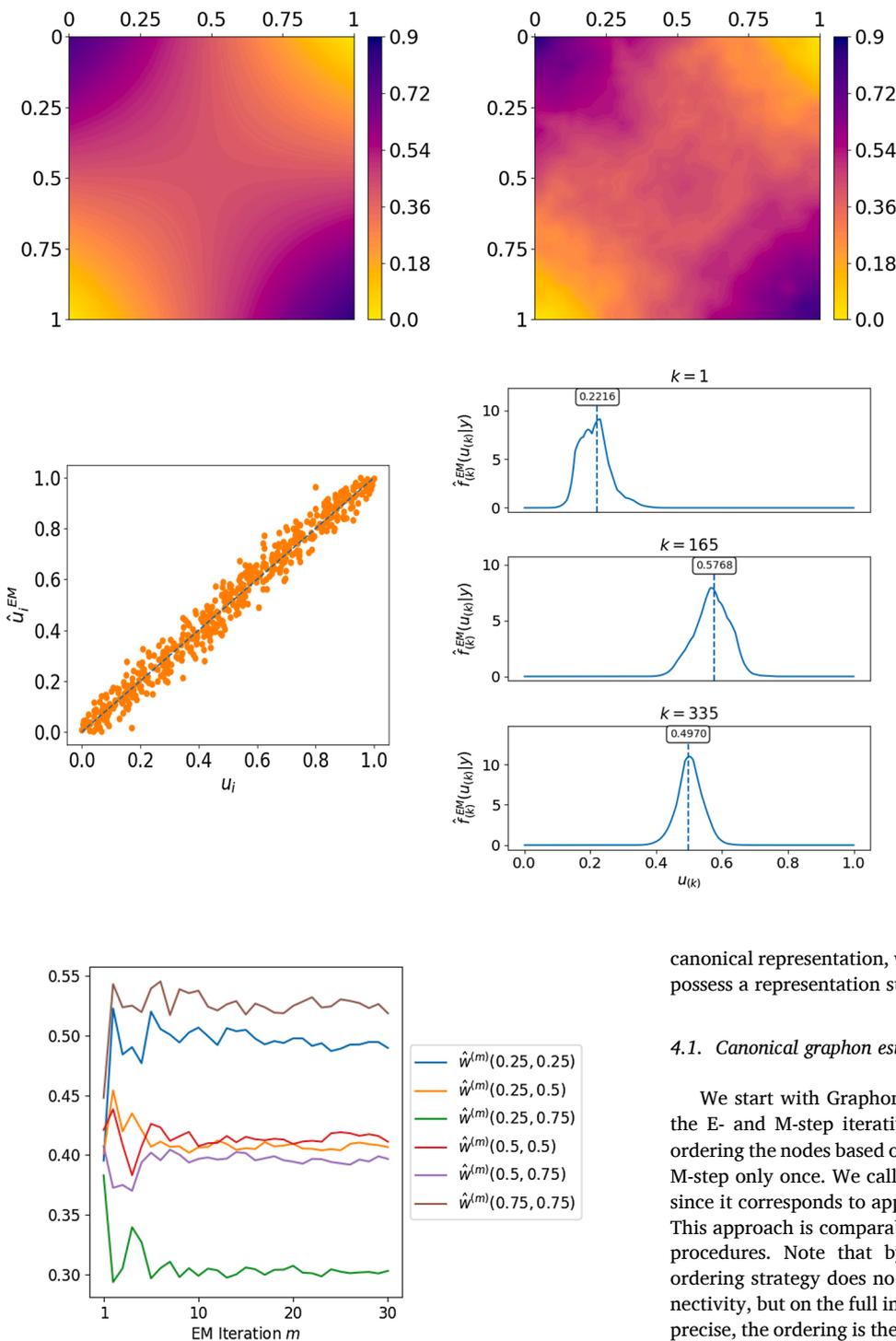


Fig. 4. Trajectory of the estimate  $\hat{w}^{(m)}(u, v)$  for Graphon 1 from Table 1 at selected positions  $(u, v) \in [0, 1]^2$  for the proceeding EM iterations  $m = 1, \dots, 30$ .

ordering, relating them to the corresponding graphon estimates. This completes the procedure of the EM-type algorithm and any related evaluation concepts.

#### 4. Simulation studies

For evaluating our graphon estimation approach, we first consider networks generated from a known ground truth. More precisely, for each of the two graphons from Table 1, we simulate networks with dimension  $N = 500$  using the data generating process (5). The first graphon has a

Fig. 3. Graphon estimation for Graphon 1 (top left) from Table 1, based on the EM algorithm using  $\hat{u}^{(0)}$  from (8) as initialization and excluding the canonical restriction from (14). The final graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  given a simulated network of size  $N = 500$  is depicted at the top right. The plot at the bottom left illustrates the comparison between the estimated  $\hat{u}_i^{EM}$  and the true simulated  $u_i$ . The three plots at the bottom right show for some selected indices the approximated conditional distribution of  $U_{(k)}$  with respect to the graphon estimate in the top right panel. The dashed vertical lines (see also numbers in the box annotations) represent the estimates  $\hat{u}_{(k)}^{EM}$ .

canonical representation, while the second is more general and does not possess a representation such that  $g(\cdot)$  from (3) is strictly increasing.

#### 4.1. Canonical graphon estimation

We start with Graphon 1 and demonstrate the benefits of applying the E- and M-step iteratively. This will be done in comparison with ordering the nodes based on the degree, i.e., using  $\hat{\psi}(\cdot)$ , and applying the M-step only once. We call the latter approach the “one-step” estimator since it corresponds to applying one iterative step of the EM algorithm. This approach is comparable to other degree-based graphon estimation procedures. Note that by proceeding with the EM algorithm, the ordering strategy does no longer rely merely on the marginalized connectivity, but on the full information about connectivity behavior. To be precise, the ordering is then based on  $\mathbf{y}_{i\bullet} = (y_{i1}, \dots, y_{iN})$  instead of degree  $(i) = \sum_j y_{ij}$ . Thus, applying further iterations should generally improve the model fit. To evaluate this with regard to Graphon 1, we first consider the “one-step” estimate  $\hat{w}^{(1)}(\cdot, \cdot)$ , which is shown in the top right panel in Fig. 2. Apparently, this model fit is not convincing when compared to the true graphon at the top left, which can be traced back to poor estimates of the  $U_i$ . To illustrate this, we compare  $\hat{u}_i^{(0)}$  as defined in (8) with the true simulated values  $u_i$ , see the bottom left plot in Fig. 2. Obviously, no concordance is visible, which reveals the inadequacy of the degree-based node ordering. In addition, we estimate the ranking density of  $U_{(k)}$  as proposed in (16), which, for three selected indices, is plotted in the bottom right panels, including the corresponding initial estimates  $\hat{u}_{(k)}^{(0)}$  (vertical dashed lines). This shows that the  $\hat{u}_{(k)}^{(0)}$  are not well represented by the respective conditional distributions.

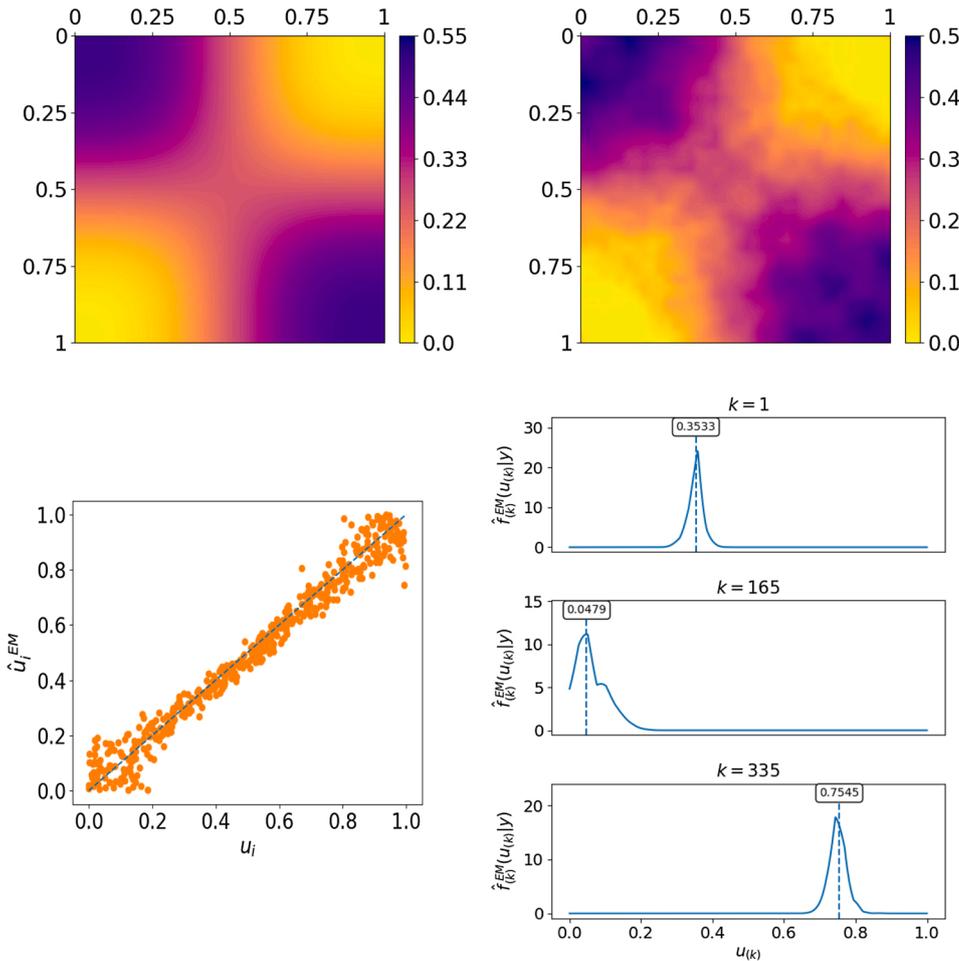


Fig. 5. Graphon estimation for the non-canonical Graphon 2 (top left) from Table 1, based on the EM algorithm using  $\hat{u}^{(0)}$  from (8) as initialization and excluding the canonical restriction from (14). The final graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  given a simulated network of size  $N = 500$  is depicted at the top right. The plot at the bottom left illustrates the estimated  $\hat{u}_i^{EM}$  versus the true simulated  $u_i$ . The three lower right plots show for some selected indices the approximated conditional distribution of  $U_{(k)}$  with respect to the graphon estimate in the top right panel. The dashed vertical lines (see also numbers in the box annotations) represent the estimates  $\hat{u}_{(k)}^{EM}$ .

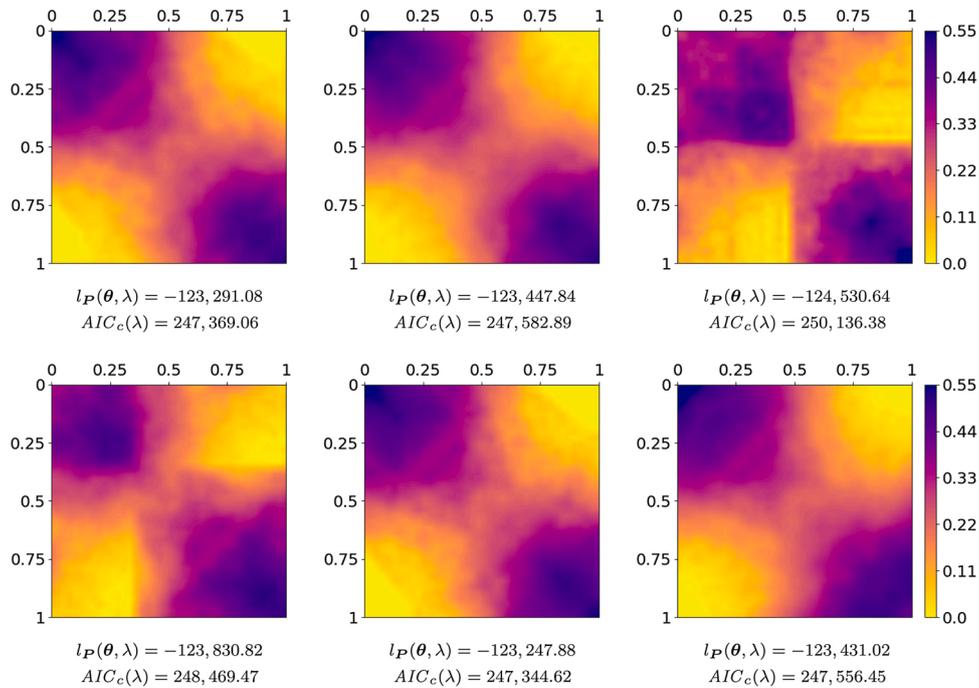


Fig. 6. Final graphon estimates  $\hat{w}^{EM}(\cdot, \cdot)$  for the non-canonical Graphon 2 from Table 1, based on the EM algorithm using different uninformed random initializations. For all six repetitions, the same simulated network of size  $N = 500$  was used. Additionally, the corresponding penalized likelihood and the corrected AIC are given beneath each estimation result.

**Table 2**  
Details about real-world networks used as application examples.

	Number of nodes	Average degree	Overall density
Facebook friendships	333	15.13	0.046
Military alliances	141	24.16	0.173
Human brain functional coactivations	638	58.39	0.092

Accordingly, the conformity between  $\hat{w}^{(1)}(\cdot, \cdot)$  and  $\hat{u}^{(0)}$  is rather poor. Summarizing the estimation results, we can conclude that in this case the degree ordering is not suitable to achieve an appropriate estimate of the true node ordering with respect to the data generating model. This happens even though the graphon has a canonical representation. Consequently, the “one-step” estimator also results in a very poor fit to the underlying true graphon.

To correct the node ordering and to improve the graphon estimate, we subsequently iterate between the E- and the M-step. The final EM estimate (i.e., after convergence) for Graphon 1 is shown in the top right plot in Fig. 3, where, in contrast to  $\hat{w}^{(1)}(\cdot, \cdot)$ , the true structure is clearly captured. Regarding the final EM-based estimates  $\hat{u}_i^{EM}$ , the comparison with the true values  $u_i$  now reveals a reasonable ordering (see Fig. 3,

bottom left). Moreover, here the conditional distributions of  $U_{(k)}$  in relation to the estimates  $\hat{u}_{(k)}^{EM}$  indicate a plausible positioning for the same selected indices as above (bottom right plots). As an overall conclusion, the proposed EM algorithm provides convincing results even if the initial node ordering based on the degree is not adequate.

Finally, to explore the convergence rate, we evaluate graphon values at selected positions throughout the algorithm, which is illustrated in Fig. 4. In doing so, we see that after about eight to twelve iterations a reasonable convergence occurs. Note that although Graphon 1 has a canonical representation, the canonical restriction from (14) has not been incorporated in the proceeding EM iterations. This is because the unrestricted estimate turned out to be similarly good.

4.2. Non-canonical graphon estimation

For a further evaluation, we now look at Graphon 2, which does not provide a canonical representation and thus is intractable for simple degree-based estimation procedures. We demonstrate that our proposed method, however, is also able to handle such unrestricted graphons. The marginal function  $g(\cdot)$  as defined in (3) is constant here at 0.25, meaning that the degree is completely uninformative with regard to the node ordering. Nonetheless, we stick with the degree-based initialization as

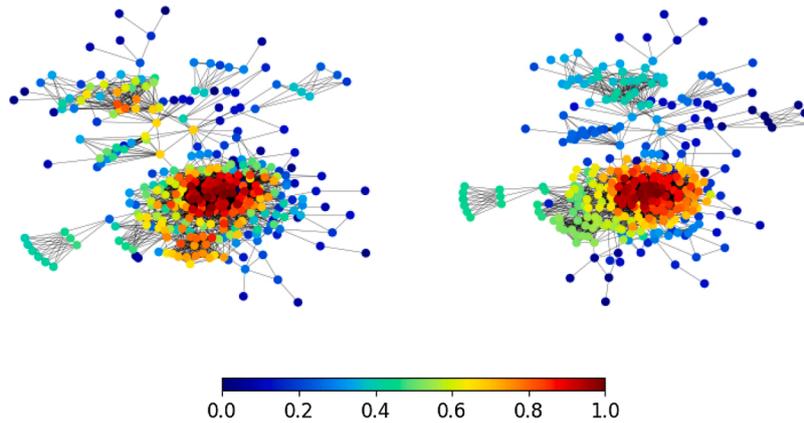


Fig. 7. Facebook ego network with node coloring referring to  $\hat{u}_i^{(0)}$  (left) and  $\hat{u}_i^{EM}$  (right).

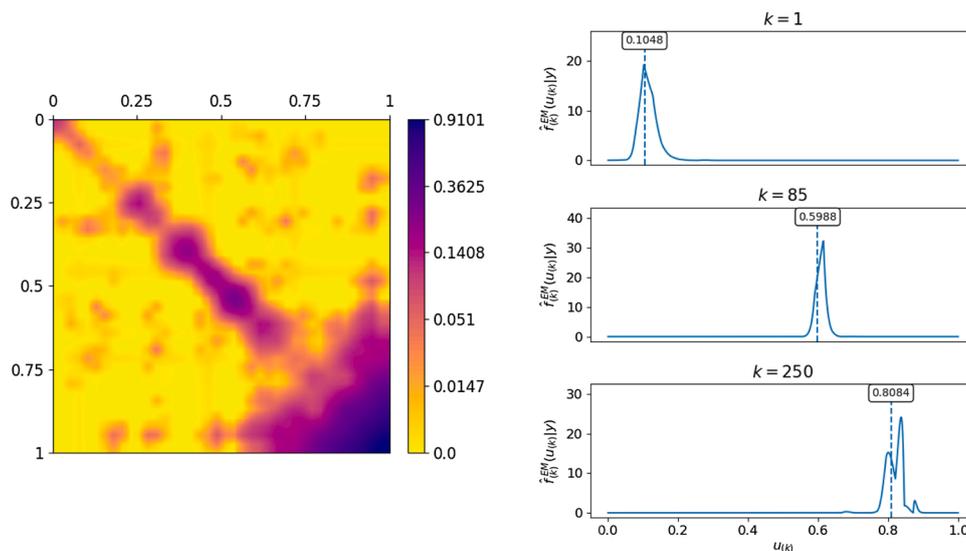


Fig. 8. Graphon estimation for the Facebook ego network. The graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  (in log scale) is depicted on the left. The three plots on the right show the approximated conditional distribution of  $U_{(k)}$  (with respect to the given graphon estimate) for some selected indices. The dashed vertical lines (see also numbers in the box annotations) represent the estimates  $\hat{u}_{(k)}^{EM}$ .

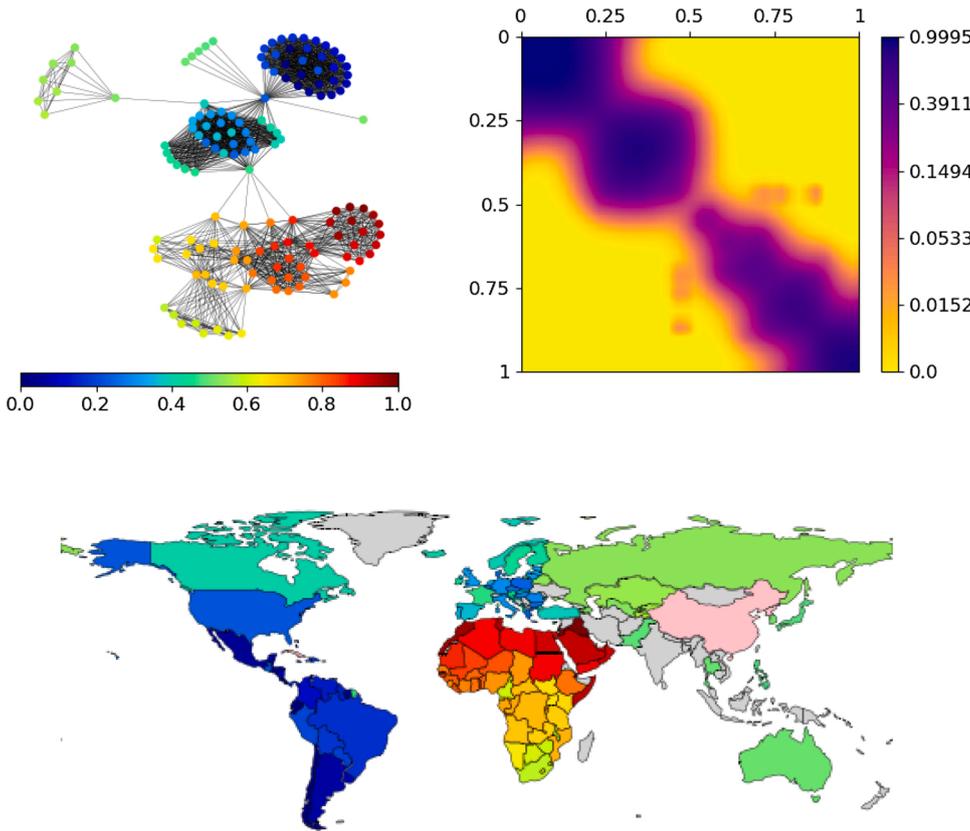


Fig. 9. Graphon estimation for the military alliance network (top left, with coloring referring to  $\hat{u}_i^{EM} \in [0, 1]$ ). The graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  (in log scale) is depicted at the top right. The lower plot shows the world map, using the same color scheme for  $\hat{u}_i^{EM}$ . As an isolated group, China, Cuba, and North Korea (colored in pink) have not been included in the estimation routine. Countries which have no strong agreement with any other country and hence do not appear in the data set are colored in gray.

**Table 3**  
List of non-allied states with highest edge probabilities and allied states with lowest edge probabilities, first five relationships each. The estimated edge probabilities are derived from the fitted graphon model.

Status	States	Edge probability
Non-allied	Ethiopia – Ivory Coast	0.5673
	Mauritania – Sudan	0.5651
	Finland – Turkey	0.5578
	Mauritania – Yemen	0.5533
	Ethiopia – Sierra Leone	0.5495
Allied	France – Senegal	0.0199
	Central African Republic – France	0.0214
	France – Gabon	0.0218
	Comoros – France	0.0234
	Sudan – Uganda	0.0679

proposed above to show that the EM algorithm is capable of capturing the underlying structure even under improper initialization. We exclude the canonical side constraint from (14) since we now explicitly want to enable a flexible marginal function. The final graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  for a simulated network of dimension  $N = 500$  is depicted in the top right plot in Fig. 5, where the structure of the true graphon (top left) is fully captured. This is accompanied by a very good concordance between the estimated  $\hat{u}_i^{EM}$  and the true  $u_i$ . Consequently, the estimates  $\hat{u}_i^{EM}$  are well covered by the corresponding ranking densities after (16), which is illustrated for three selected indices in the stacked plots at the bottom right.

To showcase the performance of our method in more depth, we consider another simulated network ( $N = 500$ ) under Graphon 2 and apply the estimation procedure several times with different random initializations. More precisely, instead of using the degree-based ordering from (8), we set  $\hat{u}^{(0)}$  as a random permutation of  $(i/(N + 1)) : i = 1, \dots, N$ . This proceeding then also exemplarily characterizes the

appearance of the EM algorithm being trapped at local minima of the AIC. Fig. 6 illustrates the final estimates for six such randomly initialized repetitions. Although we start each run with a completely uninformative random initialization, in four out of six final estimates, the structure of the original graphon can instantly and clearly be recognized. Yet, the estimates in the top right and bottom left panels seem to display differing structures. In fact, these estimates exhibit the appearance of segment swaps and reversals. For instance, in the graphon estimate in the bottom left panel, the upper and the lower parts of the domain  $[0, 1]$  are swapped and, in addition, the originally lower, here upper part is reversed. However, as mentioned in Section 1.2, applying permutations to the graphon has no effect on the network generating model itself. In this light, the structure of the true graphon is always well captured, albeit in some cases in a different form of representation. Thus, our algorithm yields very good estimates even under uninformative initial node ordering and is therefore not (crucially) dependent on the initialization. Nevertheless, the results can be distinguished. Looking at the respective AIC values beneath the graphon estimates in Fig. 6, we see that the four smoother estimates result in smaller values, while the occurrence of a “jump” – caused by an only piecewise correct merging of nodes – leads to increased AIC values. This illustrates that a random node ordering as initial E-step can be applied in combination with the AIC in order to obtain an optimal smooth graphon estimate.

Having demonstrated the applicability of our approach, the above results can furthermore be employed to exemplify possible use cases. Regarding the fact that the estimates from Figs. 5 and 6 are based on different simulated networks, they can be considered as independent and therefore allow to conduct a structural comparison on the basis of graphon estimates. This is elaborated in Section A.2 of the Appendix.

### 5. Real-world data examples

We complete the paper by analyzing real-world networks. To do so, we consider network data from three different domains, namely from

sociology, political science, and neuroscience. Details about the networks are given in Table 2, which reveal differences in size and average degree. However, all considered networks have an overall density of almost or more than five percent and therefore can be seen as rather dense graphs. For the EM-based graphon estimation, we use degree ordering as initialization where the structure seems appropriate. Otherwise, we apply random initialization and select the best outcome over several repetitions.

5.1. Facebook ego network

A very common application and one of the roots of social network analysis are friendship networks. Here, we consider a Facebook ego network which has been collected by McAuley and Leskovec (2012) and is available on the Stanford Large Network Dataset Collection (Leskovec and Krevl, 2014). This ego network, consisting of 333 actors, is depicted in Fig. 7 with two different node orderings (represented by coloring). The left panel illustrates the initial degree ordering, while on the right, coloring refers to the final EM ordering. The latter seems much more appropriate with respect to the network structure. Hence, the initial degree ordering can be improved distinctly through our iterative proceeding. Moreover, the inherent structure of the network can be recognized in the corresponding final graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$ , which is shown on the left in Fig. 8. For example, the connectivity pattern among the bundle of nodes in the center of the lower network part with roughly  $\hat{u}_i^{EM} \in [0.65, 1]$  (see right network) is captured in the graphon estimate through the intense region at the bottom right. Similarly, other connected node bunches can also be allocated. In addition to this recognizable structure, the estimates  $\hat{u}_{(k)}^{EM}$  for some selected indices are adequately represented by the corresponding conditional distributions of  $U_{(k)}$ , which underlines the appropriateness of the graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$ . Exploiting the possibility of representing network structure in the form of graphons, we can, in addition, compare this ego network to another one. This is outlined in Section

A.2 of the Appendix.

5.2. Military alliance network

As second real-world network example, we consider strong military alliances among the world’s nations. For that purpose, we use data from the Alliance Treaty Obligations and Provisions project (Leeds et al., 2002), which provide information about all kinds of military alliance agreements over an extensive period. To extract a network with substantial connections, here we define the presence of a strong military alliance when two states have entered into an offensive or defensive pact, meaning when they have signed a treaty which forces the one country to intervene by active military support if the other country comes into a conflict with offensive or defensive actions, respectively. Furthermore, we truncate the data to agreements that were in force in 2016 as the most recent available year. The best final estimation results of the EM algorithm over several repetitions with random initial node ordering are illustrated in Fig. 9. The graphon estimate in the top right panel exhibits a very pronounced assortative structure, meaning that links predominantly occur between pairs of nodes whose latent quantities are close. In the smooth graphon model, such an elevated region around the diagonal generally implies that similar connectivity behavior and connectedness are accompanied by each other, comparable to an assortative community structure in the SBM. To verify the assortativity and compare the discovered structure, the results of an SBM fitted to this network are provided in Section A.3 of the Appendix. The SBM result confirms the assortative structure but lacks information on within- and between-group positions, which, in contrast, is provided by the graphon model. The ordering illustrated in the top left network in Fig. 9 reflects, for example, the position of the United States (light blue in-between node in the upper network part) in between the South American states (bluish node bundle) and the European states (cyanish node bundle), whereas, in the blockmodeling approach, it would be assigned either to one of those two groups or as a group on its own (as in the fit in Section A.3). The final ordering of the other nodes also appears reasonable with

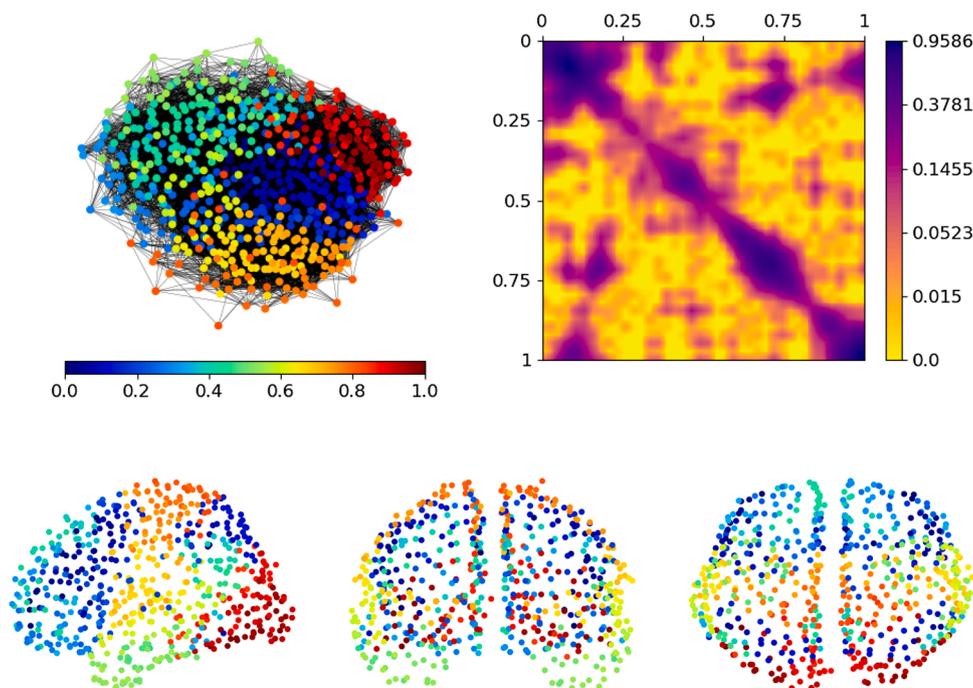


Fig. 10. Graphon estimation for the functional coactivation network of the human brain (top left, with coloring referring to  $\hat{u}_i^{EM} \in [0, 1]$ ). The graphon estimate  $\hat{w}^{EM}(\cdot, \cdot)$  (in log scale) is depicted at the top right. The lower three plots show the local positions of the human brain regions in anatomical space in side view (left), front view (middle), and top view (right), using the same color scheme for  $\hat{u}_i^{EM}$ .

respect to the network structure. Transferring this ordering by coloring the world map (see the bottom part of Fig. 9) reveals a strong conformity between geographic closeness and closeness with respect to the latent quantity. Together with the assortative structure, this means that countries which are geographically close form similar military alliances and, moreover, are more likely to be allied with each other. Furthermore, taking the perspective of link prediction, we can easily evaluate which relationships are extraordinary. To do so, we select the non-allied states which have the highest edge probabilities and the allied ones which have the lowest, see Table 3. This indeed reveals exceptional cases such as the relation between Finland and Turkey (both integrated into the Western states but not allied) or the outstanding alliances between France and some African states. (Note that links which are to be estimated can easily be left out in the graphon estimation procedure if desired, i.e., set to “N/A”).

### 5.3. Human brain functional coactivation network

To conclude the real-world examples, we consider a functional coactivation network of the human brain which has been constructed through a meta-analysis by Crossley et al. (2013) and which is available in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010a, see also Rubinov and Sporns, 2010b for detailed description). Their weighted network matrix represents the “estimated [...] similarity (Jaccard index) of the activation patterns across experimental tasks between each pair of 638 brain regions” (Crossley et al., 2013). In addition, to preserve sparsity, “[t]he coactivation matrix was probabilistically thresholded such that [...] no edge was drawn [...] if the Jaccard index between a pair of regions was not significantly greater than expected under the null hypothesis” ( $p$ -value of 0.01 and corrected for false discovery rate). Consequently, we apply a threshold of slightly above zero to obtain a binary adjacency matrix, meaning that we include a link between each pair of brain regions whose activation coincide significantly. The resulting network with a density of approximately 9.2 percent is depicted in the top left panel in Fig. 10. For the evaluation, we again consider the best outcome of the EM algorithm over several repetitions with random initialization. The resulting node ordering is illustrated in the top left network with coloring referring to  $\hat{u}_i^{\text{EM}}$ . The corresponding graphon estimate in the top right plot again reveals an assortative structure but, in addition, exhibits a conspicuous pattern of functional coactivations of some segments which are separated with respect to the latent dimension. Regarding the spatial positions of the brain regions, the lower three plots show a strong relation between closeness in anatomical space and closeness of the  $\hat{u}_i^{\text{EM}}$ . Nevertheless, there also seem to be areas which have a similar color pattern (indicating a similar coactivation pattern) but are spatially separated. In fact, the interaction and coactivation of distant brain areas is a familiar phenomenon in neuroscience and can also be seen in Crossley et al. (2013), who pursue a blockmodeling strategy. Yet, we argue that through the graphon model more insights can again be gained in terms of within- and between-group positions. To directly compare the insights acquired, the SBM with four communities in Crossley et al. (2013) is reconstructed in Section A.3 of the Appendix. This clearly shows that a strict clustering into groups does not fully suit the smooth transition in coactivation, which, however, can be discovered by the graphon model. Overall, we can demonstrate that the EM approach for estimating smooth graphons provides additional insights into network structures.

## 6. Discussion and conclusion

This paper proposes a novel estimation routine for smooth graphon estimation which explicitly takes the variability of ordering the nodes into account. The proposed semiparametric approach based on (linear) B-splines allows to incorporate relevant properties into the estimation, such as symmetry or the common canonical constraint if desired.

Exploring the conditional distribution of the latent positions by applying Gibbs sampling illuminates the uncertainty about the degree ordering and its distribution. Both steps combined give an EM-type algorithm which enables flexible graphon estimation even in large networks. The proposed procedure outperforms available feasible routines in three aspects, where we draw comparisons with the sorting-and-smoothing algorithm (Chan and Airoldi, 2014), the stochastic blockmodel approximation (Airoldi et al., 2013), the universal singular value thresholding (Chatterjee, 2015), and the neighborhood smoothing (Zhang et al., 2017). First, the B-spline estimate can guarantee a smooth outcome, which goes also beyond estimating “merely” the edge probability matrix. Secondly, the constraint from (3) of a strictly increasing marginal function, which is a major limitation of the generality of graphon models, is *not* required. (This also applies in some aspects to the universal singular value thresholding and the neighborhood smoothing approach, but only with regard to the estimation. For the (visual) representation, ordering the resulting edge probability matrix appropriately remains an open issue if (3) does not apply.) Thirdly, based on the calculations of the conditional distribution of  $\mathbf{U}$  given  $\mathbf{Y}$  embedded in the EM algorithm, one can gain information about the applied node ordering and assess its underlying uncertainty.

The proposed approach can also be used in other related models like the stochastic blockmodel (SBM), where one assumes that nodes cluster and form simple Erdős-Rényi models within and between the clusters. Apparently, SBMs do not have a smooth underlying graphon structure, so that the procedure presented in this paper would need to be adjusted and further developed. This is accomplished in a separate paper (De Nicola et al., 2020, Section 4.3). Although MCEM-based approaches already exist in the SBM estimation literature (see e.g. Daudin et al., 2008), an extension of our method could provide an innovative perspective by additionally enabling the estimation of a mixture of SBM and smooth graphon model, i.e., a model of distinct communities but with smoothly differing profiles within these communities. This extension, however, lies beyond the scope of this paper.

Moreover, the introduced procedure makes an extension towards directed networks seem to be convincing. Yet, one needs to bear in mind the consequences concerning the theory of graph limits and exchangeable random graphs. Besides, also incorporating exogenous covariates is conceivable by formulating  $\text{logit}(P(Y_{ij} = 1|U_i = u_i, U_j = u_j), \mathbf{X} = \mathbf{x}) = w(u_i, u_j) + \mathbf{x}^\top \boldsymbol{\beta}$ , which should be estimable by extending the EM approach. However, both extensions require a thorough elaboration.

Overall, graphon estimation provides an interesting tool for network visualization, as demonstrated in the examples. This allows for exploring network structure and classifying node relevance at the individual level. From the perspective of link prediction, missing edges can be estimated and extraordinary connections can be detected. Besides, the resulting graphon estimate captures network heterogeneity independently of the network size, and therefore it can be used to compare the structure of more than one network. In this respect, graphon estimation is more than a modeling exercise but also serves as a tool for exploratory network data analysis and could potentially contribute to statistical inference on random graphs.

We have implemented the EM-based graphon estimation routine described in the paper in a free and open source Python package, which is publicly available on <https://github.com/BenjaminSischka/GraphonPy.git> (Sischka, 2021).

### Declarations of interest

None.

### Acknowledgments

We thank the referees who greatly improved the paper with constructive and inspiring comments and suggestions, especially with

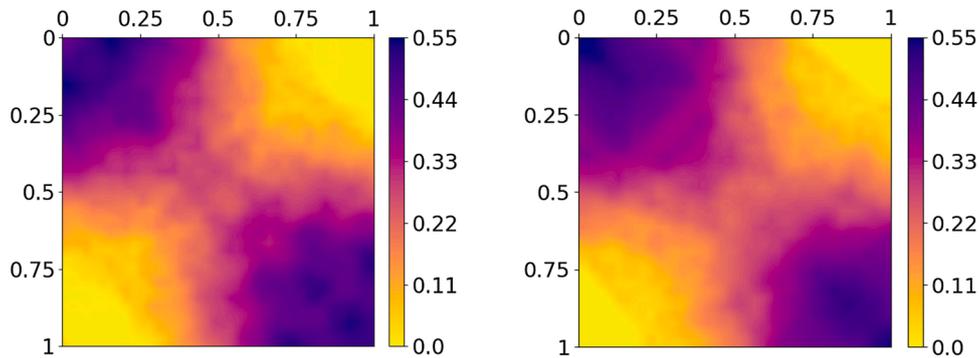
regard to a suitable framing towards the field of *social* network analysis. The project was partially supported by the European Cooperation in Science and Technology [COST Action CA15109 (COSTNET)]. This

research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

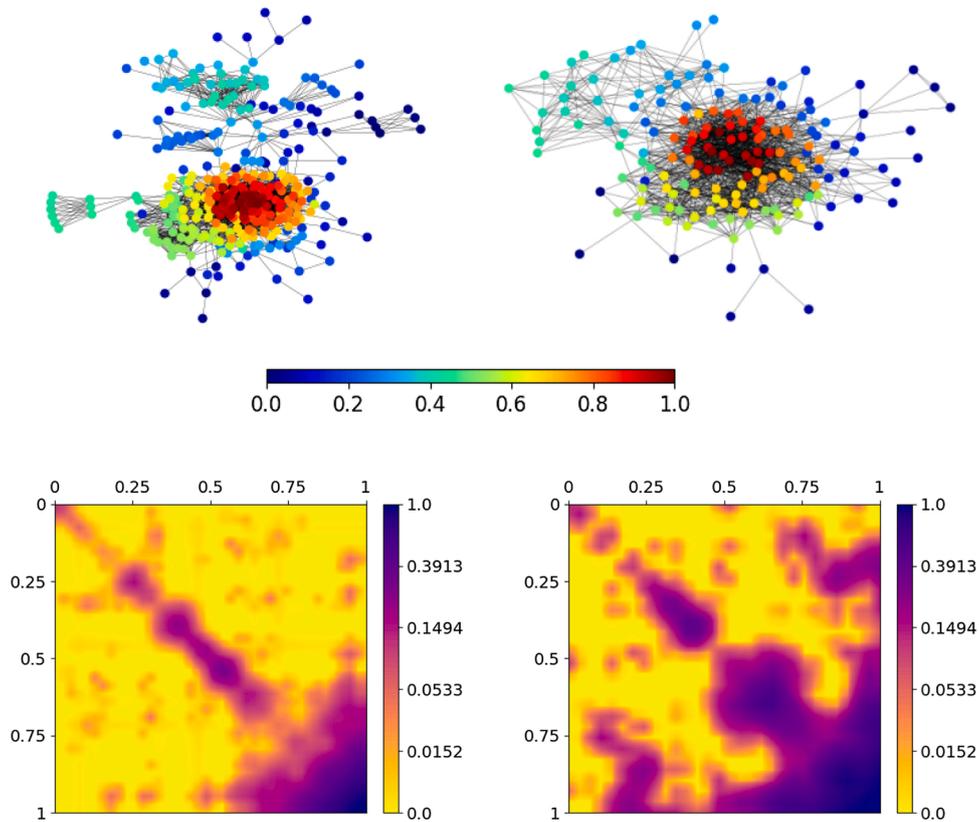
## Appendix

### A.1 Endogenous network structure in graphon models

The graphon model belongs to the latent space approaches (see [Matias and Robin, 2014](#) for an overview). Hence, it conceptually differs from approaches which explicitly focus on modeling and explaining local structural patterns like the ERGM. However, the graphon model also involves *endogenous structural processes*, although this is often difficult to attribute. More precisely, the distribution of the frequency of any motif (meaning simple finite subgraph) is uniquely characterized by the graphon model, as described by [Lovász and Szegedy \(2006\)](#). [Latouche and Robin \(2016\)](#) derive a closed-form expression of the (variational posterior) distribution of motif frequencies under particular circumstances. Even though this



**Fig. 11.** Two graphon estimates  $\hat{w}^{EM}(\cdot, \cdot)$  for the non-canonical Graphon 2 from [Table 1](#), based on different simulated networks of size  $N = 500$ . Apparently, the inherent structure in these two models is very similar, which implies comparable behavioral patterns in the two underlying networks.



**Fig. 12.** Graphon estimation for Facebook ego networks. The upper row shows the two ego networks with respective sizes of 333 and 168 actors, where the coloring refers to the latent quantities. The corresponding graphon estimates are shown in the lower row. Although the inherent structures exhibit similarities, connections between separated actors are more pronounced in the right model.

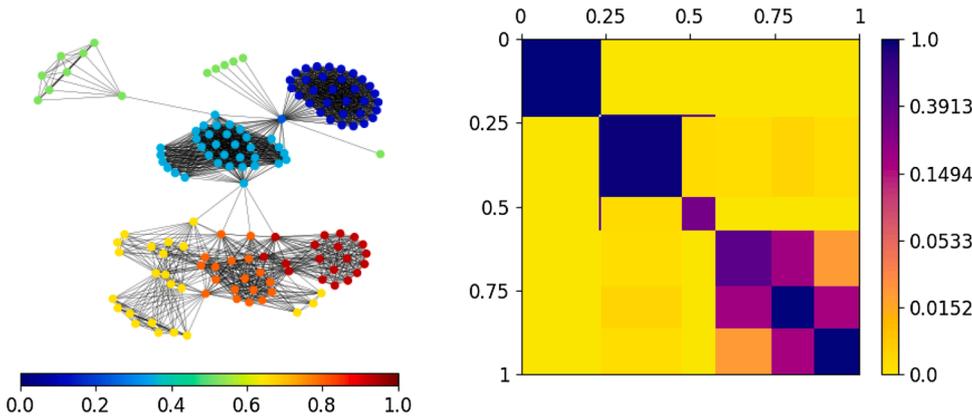
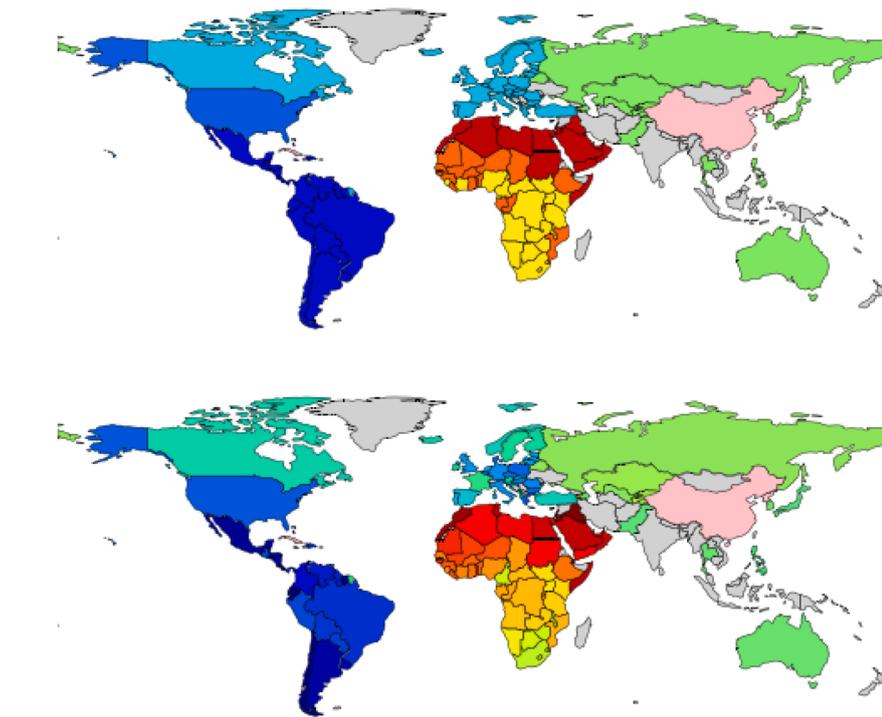


Fig. 13. Stochastic blockmodeling of the military alliance network. The network with communities indicated by different colors is illustrated at the top left. The estimate edge probability matrix – represented as graphon in log scale – is shown at the top right. The two lower plots depict the world map with coloring referring to the results of the SBM (upper map) and the graphon model (lower map, cf. Fig. 9). In both models, the isolated group, consisting of China, Cuba, and North Korea (colored in pink), was not included in the estimation routine. Countries which have no strong agreement with any other country and hence do not appear in the data set are colored in gray.



relation does not describe a universal local behavioral pattern as in the ERGM, it provides information about a global pattern. For instance, global transitivity can be expressed by

$$\int \int \int w(u_i, u_j) w(u_i, u_k) w(u_j, u_k) du_i du_j du_k,$$

which is the probability that three arbitrary nodes  $i, j$ , and  $k$  form a triangle. Conversely, [Bickel et al. \(2011\)](#) establish a sequence of subgraph patterns, called wheels, which fully determine the graphon structure and from which they derive a method of moments estimator. We also refer to [Chatterjee and Diaconis \(2013\)](#) as well as [Yin et al. \(2016\)](#) who elaborate a direct connection between the graphon model and the ERGM, at least for simple statistics, such as the number of two-stars or triangles, and in the case of dense graphs. [He and Zheng \(2015\)](#) make use of this connection and propose to use asymptotic properties of graphons to derive estimates in high dimensional ERGMs. Moreover, a connection in the situation of sparse graphs is given by [Krioukov \(2016\)](#), who also focuses on the edge-triangle model. Altogether, endogenous network structure exists in graphon models as well (at least on a global scale), but it is not straightforward to either derive or include a concrete specification of this kind of structure since the model follows a different conception.

A.2 Examples of network structure comparison

Graphon models represent an outstanding tool for comparing the structure of networks. To demonstrate this, we resume two of the examples presented above, namely the non-canonical graphon model from Section 4.2 and the Facebook ego network from Section 5.1.

First, regarding the artificial non-canonical graphon model from Section 4.2, graphon estimates have been carried out for two different simulated networks of size  $N = 500$ , confer Figs. 5 and 6, respectively. To compare the structure of these two networks, we contrast the corresponding estimates in Fig. 11. To be precise, here we illustrate once again the one-run estimate from the top right panel in Fig. 5 and the AIC-optimal estimate in the middle of the lower row from Fig. 6. Apparently, these two graphons exhibit a very similar structural pattern. In both model fits, there are regions at both ends which imply an aggregation of nodes that are densely connected among themselves but sparsely connected to the respective other ones. The transition between those regions is modeled smoothly in both cases. In summary, comparing the estimation results reveals that the two underlying networks exhibit a very similar structure. Moreover, it provides a strong indication of what is in fact the ground truth, namely that the two networks originate from the same distribution model.

As a second example, we consider the Facebook ego network presented in Section 5.1, which we now want to compare with another ego network with 168 actors that has also been collected by McAuley and Leskovec (2012). The two ego networks with coloring referring to node positioning and the corresponding graphon estimates are illustrated in Fig. 12. Comparing the two graphons reveals similar structural patterns, where a global assortative structure is prominent in both model fits. This includes in particular a bundle of densely connected nodes represented by the intense region at the bottom right. However, regarding the right network, the corresponding graphon estimate also exhibits intense regions offside the diagonal, which, in this form, is not present in the left estimate. Therefore, the two networks reveal relevant structural differences and do presumably not originate from the same distribution model.

A.3 Advantages of the graphon model over the SBM

Making use of the graphon model often provides more detailed insights into the network structure than applying an SBM. To demonstrate this, we again consider the military alliance network from Section 5.2 and the human brain network from Section 5.3, where we now additionally fit an SBM to both networks. The results for the military alliance network are illustrated in Fig. 13, where the number of seven communities can be deduced by applying the AIC. The estimated edge probability matrix, which is represented as a graphon on the top right, confirms the assortative structure found by the graphon estimate in Fig. 9. The assignment of the nodes is depicted by the coloring in the top left network and exhibits a reasonable clustering.

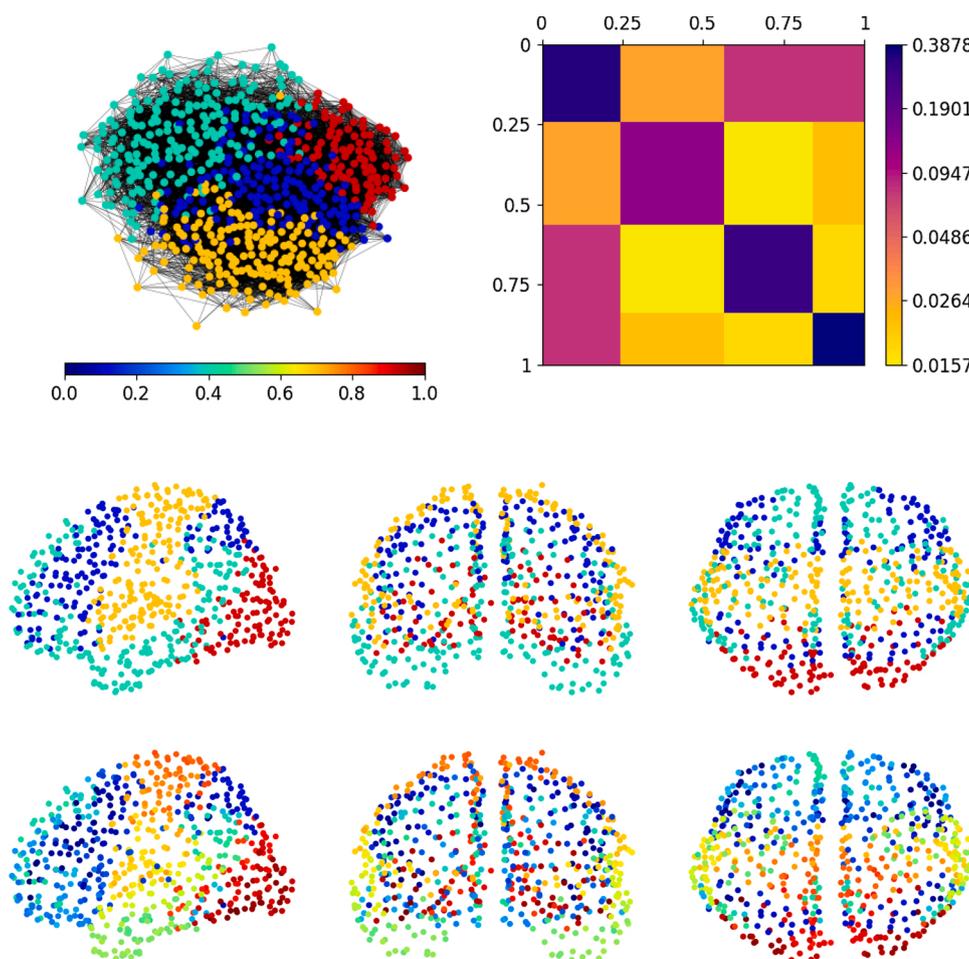


Fig. 14. Stochastic blockmodeling of the functional coactivation network of the human brain. The top left plot shows the network with clusters indicated by colors. The estimated edge probability matrix is depicted as a graphon (in log scale) at the top right. The lower two rows show the regions' assignment according to the SBM (middle row) and the graphon model (lower row, cf. Fig. 10) with respect to anatomical space (in side, front, and top view, respectively).

Comparing these results transferred to the world map with the results from the graphon model (see the two lower plots, respectively) reveals a very similar assignment. However, in this context, the graphon model provides not only a (smooth) division into groups but also information about within- and between-group positions. For instance, the relations between the African countries are captured on a finer and fluent scale, yielding a more detailed picture of the constellation. To be precise, this higher resolution allows to compare countries individually and not only based on their community memberships, where countries from the same community (different communities) might still be quite different (similar). On top of that, in the SBM, the United States simply forms a group on its own, whereas the graphon model reveals its position between the European states and the South American states and therefore its role as a connective actor between these two groups. (Note that the chosen group ordering in the SBM, which might suggest similar conclusions, here is based on the ordering in the graphon model but is generally arbitrary.) Similar circumstances can be pointed out for the human brain network, with results of the fitted SBM illustrated in Fig. 14. For the blockmodel estimation, here we follow Crossley et al. (2013) and perform the procedure of Newman (2006) with four groups. Again, the division of the nodes shows a reasonable clustering, where the blockmodel structure now exhibits high connectivity not only within but also between groups (see the two plots in the upper row, respectively). This pattern of an intense connectivity between separated node bundles is also exhibited in the graphon estimate from Fig. 10. Comparing the node assignments in the SBM and the graphon model with regard to anatomical space (see plots in the two lower rows, respectively), we see that the nodes are merged similarly. However, the graphon model additionally provides positions within and between groups. For example, considering the blockmodel's orange central cluster in the left representation (side view), the results from the graphon model reveal that the brain regions in the upper and lower area (with dark and light orange shades, respectively) are more similar to each other than to regions from the respective other area. Moreover, according to the positioning in the graphon model, the transition from the orange to the cyan community occurs at the lower central border between those communities. In conclusion, considering the graphon model as a smooth node clustering, we gain additional insights into both the within-group positions and the transitions between groups.

## References

- Airoldi, E.M., Costa, T.B., Chan, S.H., 2013. Stochastic blockmodel approximation of a graphon: theory and consistent estimation. *Adv. Neural Inf. Process. Syst.* 692–700.
- Andersen, M.S., Dahl, J., Vandenberghe, L., 2004. CvxOpt: Open Source Software for Convex Optimization (Python) <https://cvxopt.org>, version 1.2.0 [online; accessed 04-20-2018].
- Avella-Medina, M., Parise, F., Schaub, M.T., Segarra, S., 2018. Centrality measures for graphons: accounting for uncertainty in networks. *IEEE Trans. Netw. Sci. Eng.* 7 (1), 520–537.
- Bickel, P.J., Chen, A., 2009. A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci.* 106 (50).
- Bickel, P.J., Chen, A., Levina, E., 2011. The method of moments and degree distributions for network models. *Ann. Stat.* 39 (5), 2280–2301.
- Borgs, C., Chayes, L., Lócaś, L., Sós, V.T., Vesztegombi, K., 2008. Convergent sequences of dense graph I: subgraph frequencies, metric properties and testing. *Adv. Math.* 219, 1801–1851.
- Burnham, K.P., Anderson, D.R., 2010. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media, New York.
- Chan, S.H., Airoldi, E.M., 2014. A consistent histogram estimator for exchangeable graph models. *Int. Conf. Mach. Learn.* 208–216.
- Chatterjee, S., 2015. Matrix estimation by universal singular value thresholding. *Ann. Stat.* 43 (1), 177–214.
- Chatterjee, S., Diaconis, P., 2013. Estimating and understanding exponential random graph models. *Ann. Stat.* 41 (5), 2428–2461.
- Choi, D., 2017. Co-clustering of nonsmooth graphons. *Ann. Stat.* 45 (4), 1488–1515.
- Choi, D., Wolfe, P.J., 2014. Co-clustering separately exchangeable network data. *Ann. Stat.* 42 (1), 29–63.
- Crossley, N.A., Mechelli, A., Vértes, P.E., Winton-Brown, T.T., Patel, A.X., Ginestet, C.E., McGuire, P., Bullmore, E.T., 2013. Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci.* 110 (28), 11583–11588.
- Daudin, J., Picard, F., Robin, S., 2008. A mixture model for random graph. *Stat. Comput.* 18, 1–36.
- De Nicola, G., Sischka, B., Kauermann, G., 2020. Mixture Models and Networks – Overview of Stochastic Blockmodelling [arXiv preprint arXiv:2005.09396](https://arxiv.org/abs/2005.09396).
- Diaconis, P., Janson, S., 2008. Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni* 28, 33–61.
- Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 11 (2), 89–102.
- Fienberg, S.E., 2012. A brief history of statistical models for network analysis and open challenges. *J. Comput. Graph. Stat.* 21 (4), 825–839.
- Frank, O., Strauss, D., 1986. Markov graphs. *J. Am. Stat. Assoc.* 81 (395), 832–842.
- Gao, C., Lu, Y., Ma, Z., Zhou, H.H., 2016. Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* 17 (1), 5602–5630.
- Gao, C., Lu, Y., Zhou, H.H., 2015a. Rate-optimal graphon estimation. *Ann. Stat.* 43 (6), 2624–2652.
- Gao, C., van der Vaart, A.W., Zhou, H.H., 2015b. A General Framework for Bayesian Structured Linear Models [arXiv preprint arXiv:1506.02174](https://arxiv.org/abs/1506.02174).
- Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85 (410), 398–409.
- Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M., 2010. A survey of statistical network models. *Found. Trends Mach. Learn.* 2, 129–233.
- He, R., Zheng, T., 2015. GLMLE: graph-limit enabled fast computation for fitting exponential random graph models to large social networks. *Soc. Netw. Anal. Min.* 5 (1), 8.
- Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. *J. Am. Stat. Assoc.* 97 (460), 1090–1098.
- Holland, P.W., Laskey, K.B., Leinhardt, S., 1983. Stochastic blockmodels: first steps. *Soc. Netw.* 5 (2), 109–137.
- Hunter, D.R., Krivitsky, P.N., Schweinberger, M., 2012. Computational statistical methods for social network analysis. *J. Comput. Graph. Stat.* 21 (4), 856–882.
- Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. *Biometrika* 76 (2), 297–307.
- Kauermann, G., Schellhase, C., Ruppert, D., 2013. Flexible copula density estimation with penalized hierarchical B-splines. *Scand. J. Stat.* 40 (4), 685–705.
- Klopp, O., Tsybakov, A.B., Verzelen, N., 2017. Oracle inequalities for network models and sparse graphon estimation. *Ann. Stat.* 45 (1), 316–354.
- Kolaczyk, E.D., 2009. *Statistical Analysis of Network Data*. Springer, New York.
- Kolaczyk, E.D., 2017. *Topics at the Frontier of Statistics and Network Analysis*. Cambridge University Press.
- Kolaczyk, E.D., Csárdi, G., 2014. *Statistical Analysis of Network Data with R*. Springer, New York.
- Krioukov, D., 2016. Clustering implies geometry in networks. *Phys. Rev. Lett.* 116 (20), 208302.
- Latouche, P., Robin, S., 2016. Variational Bayes model averaging for graphon functions and motif frequencies inference in w-graph models. *Stat. Comput.* 26 (6), 1173–1185.
- Leeds, B.A., Ritter, J.M., Mc Laughlin Mitchell, S., Long, A.G., 2002. Alliance treaty obligations and provisions, 1815–1944. *Int. Interact.* 28 (3), 237–260.
- Leskovec, J., Krevl, A., 2014. SNAP Datasets: Stanford Large Network Dataset Collection <http://snap.stanford.edu/data> [online; accessed 06-14-2018].
- Lovász, L., 2012. *Large Networks and Graph Limits*. American Mathematical Soc.
- Lovász, L., Szegedy, B., 2006. Limits of dense graph sequences. *J. Combin. Theory Ser. B* 96 (6), 933–957.
- Lusher, D., Koskinen, J., Robins, G., 2013. *Exponential Random Graph Models for social Networks*. Cambridge University Press.
- Matias, C., Robin, S., 2014. Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proc. Surv.* 47, 55–74.
- McAuley, J., Leskovec, J., 2012. Learning to discover social circles in ego networks. *Adv. Neural Inf. Process. Syst.* 25, 539–547.
- Newman, M.E., 2006. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* 103 (23), 8577–8582.
- Nowicki, K., Snijders, T.A.B., 2001. Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Assoc.* 96 (455), 1077–1087.
- Olhede, S.C., Wolfe, P.J., 2014. Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci.* 111 (41), 14722–14727.
- Orbanz, P., Roy, D.M., 2014. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (2), 437–461.
- Rubinov, M., Sporns, O., 2010a. Brain Connectivity Toolbox (MATLAB). <https://sites.google.com/site/bctnet> [online; accessed 10-15-2019].
- Rubinov, M., Sporns, O., 2010b. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52 (3), 1059–1069.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression*. Cambridge University Press.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2009. Semiparametric regression during 2003–2007. *Electron. J. Stat.* 3, 1193.
- Salter-Townshend, M., White, A., Gollini, I., Murphy, T.B., 2012. Review of statistical network analysis: models, algorithms and software. *Stat. Anal. Data Min.* 5 (4), 243–264.
- Scott, J., Carrington, P.J., 2011. *The SAGE Handbook of Social Network Analysis*. SAGE Publications.
- Sischka, B., 2021. EM-Based Smooth Graphon Estimation (Python) <https://github.com/BenjaminSischka/GraphonPy.git> [online; accessed 01-21-2021].

- Snijders, T.A., 2011. Statistical models for social networks. *Annu. Rev. Sociol.* 37, 131–153.
- Snijders, T.A., Nowicki, K., 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classif.* 14 (1), 75–100.
- Snijders, T.A.B., Pattison, P.E., Robins, G.L., Handcock, M.S., 2006. New specifications for exponential random graph models. *Sociol. Methodol.* (1), 99–153.
- Su, Y., Wong, R.K., Lee, T.C., 2020. Network estimation via graphon with node features. *IEEE Trans. Netw. Sci. Eng.*
- Turlach, B.A., Weingessel, A., 2013. quadprog: Functions to Solve Quadratic Programming Problems (R) <https://CRAN.R-project.org/package=quadprog>, version 1. 5-5 [online; accessed 04-26-2018].
- Wolfe, P.J., Olhede, S.C., 2013. Nonparametric Graphon Estimation arXiv preprint arXiv:1309.5936.
- Wood, S.N., 2017a. Generalized Additive Models: An Introduction with R. CRC press.
- Wood, S.N., 2017b. P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Stat. Comput.* 27 (4), 985–989.
- Xu, J., 2017. Rates of Convergence of Spectral Methods for Graphon Estimation arXiv preprint arXiv:1709.03183.
- Yang, J.J., Han, Q., Airolidi, E.M., 2014. Nonparametric estimation and testing of exchangeable graph models. *Artif. Intell. Stat.* 1060–1067.
- Yin, M., Rinaldo, A., Fadnavis, S., et al., 2016. Asymptotic quantization of exponential random graphs. *Ann. Appl. Probab.* 26 (6), 3251–3285.
- You, K., 2020. graphon: A Collection of Graphon Estimation Methods (R) <https://CRAN.R-project.org/package=graphon>, version 0.3.4 [online; accessed 06-13-2020].
- Zhang, Y., Levina, E., Zhu, J., 2017. Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* 104 (4), 771–783.



**Part III.**

## **Incorporating Block Structures**

## 6. The Stochastic Blockmodel and Its Graphon Representation

### Contributing Article.

De Nicola, G., Sischka, B., and Kauermann, G. (2022). Mixture Models and Networks: The Stochastic Blockmodel. *Statistical Modelling*, 22(1-2), 67–94. doi:10.1177/1471082X211033169.

### Data Sets.

All data sets used to showcase the importance of modeling block structures in networks are freely accessible. For information on concrete sources, see the specifications in the paper.

### Copyright Information.

Copyright © 2022 by Statistical Modelling Society, SAGE Publications Ltd.

### Author Contributions.

The conceptual idea of framing the stochastic blockmodel as a mixture model approach for network data was developed by Göran Kauermann. The literature review for this survey article was mainly done by Giacomo De Nicola, who also took care of categorizing the papers and putting them into a meaningful relation. Moreover, the writing was done for the most part by Giacomo De Nicola, where the methodological (re-)formulation of the reviewed estimation techniques (Section 4) was strongly supported by Göran Kauermann. Benjamin Sischka was mainly responsible for formulating the graphon representation (Section 3.3) as well as for developing and implementing the corresponding estimation routine (Section 4.3). Besides that, Benjamin Sischka stood by for advice with regard to the selection of articles and their reasonable integration into the compilation of works on stochastic blockmodels. The application study in Section 5 was carried out by Giacomo De Nicola, where Benjamin Sischka provided the analysis results for the military alliance network (which has been explored by exploiting the graphon representation from Section 3.3).

## 7. A Mixture of Stochastic Blockmodels and Smooth Graphon Models

### Contributing Article.

Sischka, B. and Kauermann, G. (2023). Stochastic Block Smooth Graphon Model. *arXiv preprint arXiv:2203.13304*. Under review in *Journal of Computational and Graphical Statistics*.

### Software Implementation.

The method developed and formulated in the paper is implemented in a free and open source Python package that is publicly available on GitHub:

<https://github.com/BenjaminSischka/SBSGMest.git>

Moreover, all data sets used for demonstrating the applicability of our approach are freely accessible. For information on concrete sources, see the specifications in the paper.

### Author Contributions.

The initial idea of incorporating smooth variations into the graphon representation of stochastic blockmodels was proposed by Göran Kauermann. A more detailed conception of this notion was elaborated by Benjamin Sischka, where he also dealt with the elucidation from different perspectives and the meaningful integration into the network modeling context. The modeling framework of B-spline mixtures was formulated by Göran Kauermann (Section 2.4). The corresponding adaptation of the EM-type estimation routine was mainly elaborated by Benjamin Sischka (see Section 3), including the formulations necessary for adjusting the node positions. Furthermore, Benjamin Sischka developed the criterion for choosing the number of communities (Section 3.3). The application studies (Section 4), which consider both synthetic and real-world networks, were almost exclusively carried out and described by Benjamin Sischka. Lastly, the links to other models (see Supplementary Material) were entirely formulated by Benjamin Sischka.

# Stochastic Block Smooth Graphon Model

Benjamin Sischka and Göran Kauermann

March 5, 2023

## Abstract

In this paper, we propose combining the stochastic blockmodel and the smooth graphon model, two of the most prominent modeling approaches in statistical network analysis. In doing so, we bring both perspectives together and utilize their modeling capacities in a joint framework. Stochastic blockmodels are generally used for partitioning the individual actors of a network into blocks within which the connectivity behavior is assumed to be stochastically equivalent. Smooth graphon models instead follow the intuition that the nodes can be arranged on a one-dimensional scale such that closeness implies a similar behavior in connectivity. Both frameworks belong to the class of node-specific latent variable models, entailing a natural relationship. While these two modeling concepts have developed more or less independently, this paper proposes their generalization towards *stochastic block smooth graphon models*. Such a combined approach enables to exploit the advantages of both worlds. Employing concepts of the EM-type algorithm allows us to develop a corresponding estimation routine, where MCMC techniques are used to accomplish the E-step. Simulations and real-world applications support the practicability of our new method and demonstrate its advantages.

*Keywords:* Stochastic blockmodel; Graphon model; Latent space model; EM algorithm; Gibbs sampling; B-spline surface; Social network; Political network; Connectome

---

Benjamin Sischka is Research Assistant, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany (E-mail: [benjamin.sischka@stat.uni-muenchen.de](mailto:benjamin.sischka@stat.uni-muenchen.de)). Göran Kauermann is Professor, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany (E-mail: [goeran.kauermann@stat.uni-muenchen.de](mailto:goeran.kauermann@stat.uni-muenchen.de)).

# 1 Introduction

The statistical modeling of complex networks has gained increasing interest over the last two decades, and much development has taken place in this area. Network-structured data arise in many application fields and corresponding modeling frameworks are used in sociology, biology, neuroscience, computer science, and others. To demonstrate the state of the art in statistical network data analysis, survey articles have been published by Goldenberg et al. (2009), Snijders (2011), Hunter et al. (2012), Fienberg (2012), and Salter-Townshend et al. (2012). Moreover, monographs in this field are given by Kolaczyk (2009), Lusher et al. (2013), Kolaczyk and Csardi (2014), and Kolaczyk (2017).

In order to capture the underlying structure within a given network, various modeling strategies based on different concepts have been developed. One ubiquitous model class in this context is given by the Node-Specific Latent Variable Models, see Matias and Robin (2014) for an overview. The general notion in this rather broad model class is the assumption that, for a network comprising nodes  $1, \dots, N$ , the edge variables  $Y_{ij}$ ,  $i, j = 1, \dots, N$ , can be characterized as independent when conditioning on the node-specific latent quantities  $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N$ . To be precise, the generic model design can be formulated via independent Bernoulli random variables with corresponding success probabilities, i.e.

$$Y_{ij} \mid \boldsymbol{\xi}_i, \boldsymbol{\xi}_j \stackrel{\text{ind.}}{\sim} \text{Bernoulli}(h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)), \quad (1)$$

where  $0 \leq h(\cdot, \cdot) \leq 1$  characterizes an overall connectivity pattern. This especially means that the connection probability for node pair  $(i, j)$  depends exclusively on the associated quantities  $\boldsymbol{\xi}_i$  and  $\boldsymbol{\xi}_j$ . Depending on the concrete model, these quantities are either random variables themselves or simply unknown but fixed parameters. Moreover,  $\boldsymbol{\xi}_i$  can be multivariate, even though it is used as a scalar in many frameworks. Lastly, data-generating process (1) is generally defined for any  $i, j = 1, \dots, N$ . However, in the case of undirected networks without self-loops, it is only performed for  $i < j$ , with the additional setting of  $Y_{ji} \equiv Y_{ij}$  and  $Y_{ii} \equiv 0$ . This scenario is what we focus on in this work.

The general framework sketched above includes several well-known models in the field

of statistical network analysis. The most popular ones of this type are the Stochastic Blockmodel (Holland et al., 1983, Snijders and Nowicki, 1997 and 2001) as well as its variants (Airoldi et al., 2008, Karrer and Newman, 2011), the Latent Distance Model (Hoff and coauthors, 2002, 2007, 2009, 2021, Ma et al., 2020), and the Graphon Model (Lovász and coauthors, 2006, 2007, 2010, Diaconis and Janson, 2007). Following the same notion, namely a connectivity structure that only depends on node-specific latent variables, entails a natural relationship among these modeling approaches. Yet, it is well known that they possess very different capacities for representing the diverse structural aspects discovered in real-world networks. However, it is often unknown beforehand what the requirements for the model are in terms of structural expressiveness, implying that the question of which modeling strategy is best able to capture the present network structure cannot be clearly answered. To detect the best method from an ensemble of models and corresponding estimation algorithms, Li et al. (2020) recently developed a cross-validation procedure for model selection in the network context, see also Gao and Ma (2020). One step further, Ghasemian et al. (2020) and Li and Le (2021) discuss mixing several model fits based on different weighting strategies.

Although all node-specific latent variable models are more or less closely related by construction, little attention has been paid to the proper representation or integration of one model by another or to combining multiple models in a joint framework. As an advantage, such a model unification offers a new perspective that potentially provides a more flexible modeling strategy. Steps in this direction have been taken by, for example, Fosdick et al. (2019, Sec. 3), who developed a Latent Space Stochastic Blockmodel, where the within-community structure is modeled as a latent distance model. In a similar direction, Schweinberger and Handcock (2015) combined the stochastic blockmodel with the Exponential Random Graph Model (ERGM). The resultant Hierarchical Exponential Random Graph Model (HERGM) intends to fit a global ERGM only to the denser within-community structure. ERGMs are, however, beyond formulation (1) and instead seek to model the frequency of prespecified structural patterns.

In this paper, we pick up the idea of model (1) but aim to formalize and estimate the un-

derlying connectivity structure  $h(\cdot, \cdot)$  in a way that allows us to unite previous approaches. To do so, we combine stochastic blockmodels with (smooth) graphon models, leading to an extension that is able to capture the expressiveness of both models simultaneously. In order to fit this model to network data, we utilize previous results on smooth graphon estimation (Sischka and Kauermann, 2022) and EM-based stochastic blockmodel estimation (Daudin et al., 2008, De Nicola et al., 2022). The resulting method is flexible and feasible for even large networks.

The rest of the paper is structured as follows. In Section 2, we first briefly describe stochastic blockmodels and smooth graphon models. Building on that, we formalize their combination, leading to our novel modeling approach. For its estimation, we develop an EM-type algorithm, which is elaborated in Section 3. This includes the formulation of a criterion for choosing the number of groups. The capability of our method is demonstrated in Section 4, which is done with respect to simulations and real-world networks. The discussion in Section 5 completes the paper.

## 2 Conceptualizing the Stochastic Block Smooth Graphon Model

### 2.1 Stochastic Blockmodel

In the literature on statistical network analysis, the stochastic blockmodel (SBM) is an extensively developed tool for modeling clustering structures in networks, see Newman (2006), Choi et al. (2012), Peixoto (2012), Bickel et al. (2013), and others. In its classical version, one assumes that each node,  $i = 1, \dots, N$ , can be uniquely assigned to one of  $K \in \mathbb{N}$  groups—often also referred to as communities—, such that the probability of two nodes being connected only depends on their group memberships. More precisely, the data-generating process can be formulated as drawing at first the node assignments  $Z_i$ ,

$i = 1, \dots, N$ , independently from a categorical distribution given through

$$\mathbb{P}(Z_i = k; \boldsymbol{\alpha}) = \alpha_k \quad \text{with } k = 1, \dots, K, \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in [0, 1]^K \text{ and } \sum_k \alpha_k = 1. \quad (2)$$

Based on that, the edge variables are simulated under conditional independence through

$$Y_{ij} \mid Z_i, Z_j \sim \text{Bernoulli}(p_{Z_i Z_j}) \quad (3)$$

for  $i < j$ , where  $Y_{ji} \equiv Y_{ij}$  and  $Y_{ii} \equiv 0$  by definition. In these formulations,  $\boldsymbol{\alpha}$  represents the vector of (expected) group proportions and  $p_{Z_i Z_j}$  is the corresponding entry of the block-related edge probability matrix  $\mathbf{P} = [p_{kl}]_{k,l=1,\dots,K} \in [0, 1]^{K \times K}$ . Referring to formulation (1), this construction is apparently equivalent to setting  $\boldsymbol{\xi}_i = Z_i$  and  $h(Z_i, Z_j) = p_{Z_i Z_j}$ . Moreover, this modeling framework can also be viewed as a mixture of Erdős-Rényi-Gilbert models (Daudin et al., 2008). This is because the edge variables between all pairs of nodes from two particular communities or within one community are assumed to be independent and to have the same probability.

Although the model formulation is simple, the estimation is not straightforward. This is because both the latent community memberships and the model parameters need to be estimated. The literature of *a posteriori* blockmodeling starts with the work of Snijders and Nowicki (1997, 2001) and since then has been elaborated extensively (Handcock et al., 2007, Decelle et al., 2011, Rohe et al., 2011, Choi et al., 2012, Peixoto, 2017, and others). As an additional hurdle in this framework, also the number of communities has usually to be inferred from the data. Works taking this issue into account or specifically focusing thereon are, among others, Kemp et al. (2006), Wang and Bickel (2017), Chen and Lei (2018), Newman and Reinert (2016), Riolo et al. (2017), and Geng et al. (2019).

## 2.2 Smooth Graphon Model

Another modeling approach that makes use of latent quantities to capture complex network structures is the graphon model. In contrast to the SBM, the latent variables in the graphon model are scaled continuously on  $[0, 1]$ , but again the nodes' connectivity is

assumed to depend only on those quantities. The data-generating process induced by the graphon model can more precisely be formulated as follows. First, the latent quantities are independently drawn from a uniform distribution, i.e.

$$U_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1) \quad (4)$$

for  $i = 1, \dots, N$ . Secondly, the network entries are sampled conditionally independently in the form of

$$Y_{ij} \mid U_i, U_j \sim \text{Bernoulli}(w(U_i, U_j)) \quad (5)$$

for  $i < j$ , where again  $Y_{ji} \equiv Y_{ij}$  and  $Y_{ii} \equiv 0$ . The bivariate function  $w : [0, 1]^2 \rightarrow [0, 1]$  is the so-called graphon. Choosing  $\boldsymbol{\xi}_i = U_i$  and  $h(U_i, U_j) = w(U_i, U_j)$  yields the representation in the form of (1). In comparison with the SBM, the graphon model does usually *not* decompose a network into groups of equally behaving actors. Instead, it facilitates more flexible structures, which is due to the unlimited possibilities of specifying  $w(\cdot, \cdot)$ . In other words,  $w(\cdot, \cdot)$  has a higher complexity than  $(\boldsymbol{\alpha}, \mathbf{P})$ , especially for small  $K$ , which is the more usual setting. On the other hand, it is possible to construct the graphon model such that it covers any stochastic blockmodel, see e.g. Latouche and Robin (2016, Fig. 1) or De Nicola et al. (2022, Sec. 3.3). However, when it comes to estimation, the high complexity of  $w(\cdot, \cdot)$  is problematic and thus must be brought under control by applying additional constraints. A common approach to do so is to assume smoothness, meaning that  $w(\cdot, \cdot)$  fulfills some Hölder or Lipschitz condition (Olhede and Wolfe, 2014, Gao et al., 2015, Klopp et al., 2017). This framework is what we call Smooth Graphon Model (SGM). Relying on this smoothness assumption, many works apply histogram estimators, often also interpreted as SBM-type approximation, see e.g. Wolfe and Olhede (2013), Airoldi et al. (2013), Chan and Airoldi (2014), or Yang et al. (2014). In order to get a continuous function out of such a piecewise constant representation, Li et al. (2022) specify graphons by constructing the value at each position  $(u, v)$  through a mixture over such blockwise probabilities, with weights as continuous functions of  $(u, v)$ . As another approach to continuous graphon estimation,

Lloyd et al. (2012) formulate a Bayesian modeling framework with Gaussian process priors, see also Orbanz and Roy (2015) for a more general formulation. Sischka and Kauermann (2022) guarantee a smooth and stable estimation of  $w(\cdot, \cdot)$  by making use of (linear) B-spline regression. In contrast, some other works waive strict smoothness assumptions on  $w(\cdot, \cdot)$ , see Chatterjee (2015) and Zhang et al. (2017). However, these methods do *not* aim at estimating  $w(\cdot, \cdot)$  but solely the edge probabilities  $\mathbb{P}(Y_{ij} = 1 \mid U_i, U_j; w(\cdot, \cdot))$ .

## 2.3 Stochastic Block Smooth Graphon Model

Both the SBM and the SGM are based on underlying assumptions which appear to be restrictive conditions—namely strict homogeneity within the communities and overall smoothness, respectively. Therefore, we pursue to design a new model class which does not suffer from such limitations. To do so, we combine the two modeling approaches towards what we call a Stochastic Block Smooth Graphon Model (SBSGM). To be specific, we assume the node assignments  $Z_i$  to be drawn from (2) and draw independently  $U_i$  from (4) for  $i = 1, \dots, N$ . Then (3) and (5) are replaced by

$$Y_{ij} \mid Z_i, Z_j, U_i, U_j \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\tilde{w}_{Z_i Z_j}(U_i, U_j)), \quad (6)$$

where, for each pair of blocks and also within blocks, connectivity is now formulated by an individual smooth graphon  $\tilde{w}_{kl}(\cdot, \cdot)$ ,  $k, l = 1, \dots, K$ . Apparently, if  $K = 1$  we obtain an SGM, while all  $\tilde{w}_{kl}(\cdot, \cdot)$  being constant yields an SBM.

This model can be reformulated in a compact form by conflating the node assignments (2) and the latent quantities (4) in the following way. We draw  $U_i$  from (4) and, given  $U_i$ ,  $i = 1, \dots, N$ , we formulate for  $i < j$

$$Y_{ij} \mid U_i, U_j \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(w_{\zeta}(U_i, U_j)), \quad (7)$$

where  $w_{\zeta}(\cdot, \cdot)$  is a partitioned graphon which is smooth within the blocks spanned by  $\zeta = (\zeta_0 = 0, \zeta_1, \dots, \zeta_K = 1)$ , meaning within  $(\zeta_{k-1}, \zeta_k) \times (\zeta_{l-1}, \zeta_l)$  for  $k, l = 1, \dots, K$ . To

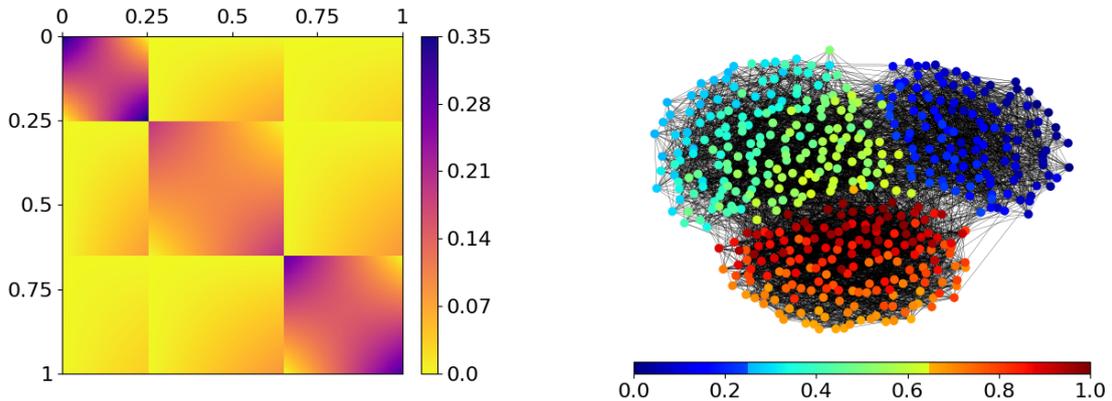


Figure 1: Exemplary stochastic block smooth graphon model with three communities.  $w_\zeta(\cdot, \cdot)$  is represented as heat map on the left. A simulated network of size 500 based on this SBSGM is given on the right, with node coloring referring to the simulated  $U_i$ 's. This network exhibits a clear community structure (global division) but also smooth transitions within the communities (local structure).

be precise, transforming the formulation from (6) to (7) implies that  $\zeta_k = \sum_{l=1}^k \alpha_l$  and

$$w_\zeta(u, v) = \tilde{w}_{k_u k_v} \left( \frac{u - \zeta_{k_u - 1}}{\zeta_{k_u} - \zeta_{k_u - 1}}, \frac{v - \zeta_{k_v - 1}}{\zeta_{k_v} - \zeta_{k_v - 1}} \right)$$

with  $k_u \in \{1, \dots, K\}$  being given through  $\zeta_{k_u - 1} \leq u < \zeta_{k_u}$ , i.e.  $k_u = \sum_k \mathbb{1}_{\{u \geq \zeta_k\}}$ . We also here remain with the common convention of symmetry ( $Y_{ji} \equiv Y_{ij}$ ) and the absence of self-loops ( $Y_{ii} \equiv 0$ ). An exemplary SBSGM together with a simulated network is illustrated in Figure 1. As a special property in terms of expressiveness, this model allows for smooth local structures under a global division into groups.

Note that the assumption of such a piecewise smooth structure in the context of graphon models has also been proposed before, see e.g. Airoldi et al. (2013) or Zhang et al. (2017). Nonetheless, there is a major conceptual distinction in the modeling perspective pursued here. While in previous works, lines of discontinuity were *merely allowed*, we now *explicitly incorporate* them as structural breaks. We stress that this novel modeling approach—which also rules the estimation—has a substantial impact on uncovering the network's underlying structure. The fact that previous graphon estimation approaches are not able to fully capture block structures is showcased, for example, by Li and Le (2021), who

observe an improvement in accuracy when mixing graphon fits with SBM estimates. By the applications in Section 4 and the comparative analysis in Section 3.2 of the Supplementary Material, we demonstrate that our method is able to capture both the block structure and the smooth differences.

## 2.4 Piecewise Smoothness and Semiparametric Model Formulation

In general, we define the SBSGM to be specified by a piecewise Lipschitz graphon with lines of discontinuity. In this context, a graphon  $w(\cdot, \cdot)$  satisfies piecewise the Lipschitz condition if there exist boundaries  $0 = \zeta_0 < \zeta_1 < \dots < \zeta_K = 1$  and a constant  $M \geq 0$  such that for all  $u, u' \in (\zeta_{k-1}, \zeta_k)$ ,  $v, v' \in (\zeta_{l-1}, \zeta_l)$

$$|w(u, v) - w(u', v')| \leq M \|(u, v)^\top - (u', v')^\top\| \quad (8)$$

for any  $k, l = 1, \dots, K$ , where  $\|\cdot\|$  is the Euclidean norm. We indicate this attribute in the notation by making use of the subscript  $\zeta$ , meaning that  $w_\zeta(\cdot, \cdot)$  is piecewise Lipschitz continuous with respect to  $\zeta$ . In the case of  $M = 0$ , this implies the representation of an SBM.

To achieve a semiparametric framework from this theoretical model formulation, we follow the spline-based approach of Sischka and Kauermann (2022), extending it to the piecewise smooth format. This can be realized by constructing a mixture of B-splines. To be precise, we formulate blockwise B-spline functions on disjoint bases in the form of

$$w_{\zeta, \gamma}^{spline}(u, v) = \sum_{k, l} \mathbb{1}_{\{\zeta_{k-1} \leq u < \zeta_k\}} \mathbb{1}_{\{\zeta_{l-1} \leq v < \zeta_l\}} [\mathbf{B}_k(u) \otimes \mathbf{B}_l(v)] \gamma_{kl}, \quad (9)$$

where  $\otimes$  is the Kronecker product and  $\mathbf{B}_k(\cdot) = (B_{k1}(\cdot), \dots, B_{kL_k}(\cdot)) \in \mathbb{R}^{1 \times L_k}$  is a linear B-spline basis on  $[\zeta_{k-1}, \zeta_k]$ , normalized to have maximum value 1 (see Figure 2 for a color-coded exemplification). The inner knots of the  $k$ -th one-dimensional B-spline component with length  $L_k$  are denoted by  $\boldsymbol{\tau}_k = (\tau_{k1}, \dots, \tau_{kL_k})$ , where  $\tau_{k1} = \zeta_{k-1}$  and  $\tau_{kL_k} = \zeta_k$ .

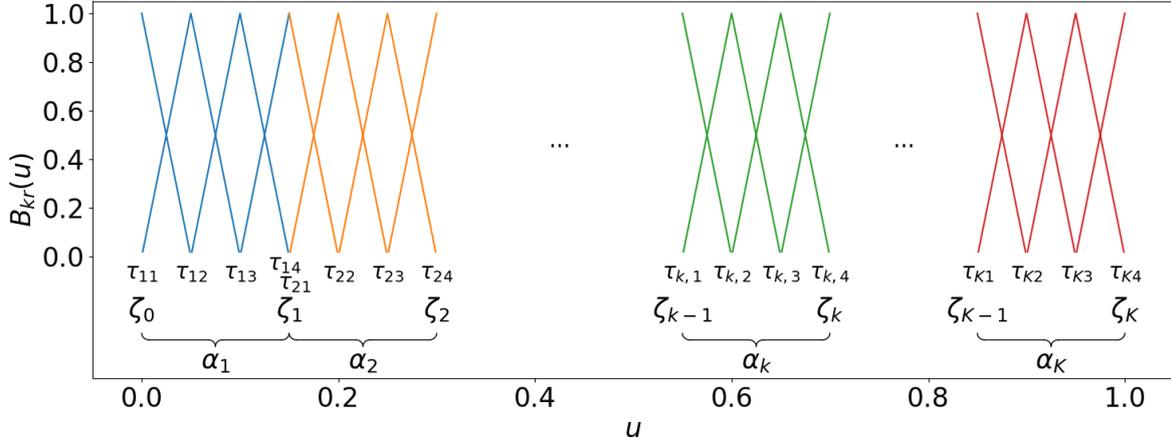


Figure 2: Disjoint univariate linear B-spline bases, colored in blue, orange, green, and red, respectively. Applying the tensor product yields the basis to construct blockwise independent B-spline functions for approaching SBSGMs. Note that this illustration shows the special case of equal community proportions.

Moreover,  $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_K)$  denotes the overall vector of B-spline knots, and the complete parameter vector is given in the form of

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}^\top, \dots, \boldsymbol{\gamma}_{1K}^\top, \boldsymbol{\gamma}_{21}^\top, \dots, \boldsymbol{\gamma}_{KK}^\top)^\top$$

with  $\boldsymbol{\gamma}_{kl} = (\gamma_{kl,11}, \dots, \gamma_{kl,1L_l}, \gamma_{kl,21}, \dots, \gamma_{kl,L_k L_l})^\top$ .

This piecewise B-spline representation serves as suitable approximation of  $w_\zeta(\cdot, \cdot)$ , where the approximation error

$$\sqrt{\iint |w_\zeta(u, v) - w_{\zeta, \boldsymbol{\gamma}}^{spline}(u, v)|^2 du dv}$$

can be arbitrarily reduced by increasing  $L_1, \dots, L_K$  accordingly. In general, we choose a sufficiently large total basis length  $L = |\boldsymbol{\tau}|$ , which is then divided among the segments in proportion to their extents. The  $L_k$  inner knots of the  $k$ -th component are subsequently located equidistantly within  $[\zeta_{k-1}, \zeta_k]$ , which completes the specification of B-spline formulation (9).

The above representation allows to apply penalized B-spline regression readily. The

capability of such a framework as well as the general role of penalized semiparametric modeling concepts are discussed, for example, by Eilers and Marx (1996), Ruppert et al. (2003), Wood (2017), and Kauermann and Opsomer (2011).

## 2.5 Identifiability Issue

As discussed above, SBSGMs describe specific forms of the graphon model, which is why it likewise suffers from non-identifiability. To be precise, Diaconis and Janson (2007) show that two graphons  $w(\cdot, \cdot)$  and  $w'(\cdot, \cdot)$  describe the same network-generating process if and only if there exist two measure-preserving functions  $\varphi, \varphi' : [0, 1] \rightarrow [0, 1]$  such that

$$w(\varphi(u), \varphi(v)) = w'(\varphi'(u), \varphi'(v)) \quad (10)$$

for almost all  $(u, v)^\top \in [0, 1]^2$ . To circumvent this identifiability issue and to guarantee uniqueness, some papers have postulated that

$$g(u) = \int w(u, v) \, dv \quad (11)$$

is strictly increasing, see e.g. Bickel and Chen (2009) or Chan and Airolidi (2014). This, however, is a strong restriction on the generality of the graphon model. To give an example, it excludes the model with  $w(u, v) = (uv)^2 + ((1 - u)(1 - v))^2$  since there exists no measurable-preserving function  $\varphi : [0, 1] \rightarrow [0, 1]$  such that  $w(\varphi(\cdot), \varphi(\cdot))$  is well-defined and fulfills condition (11). We therefore avoid employing such a restrictive uniqueness assumption. Instead, we emphasize that identifiability issues such as label switching are an inherent problem in all mixture models (see e.g. Stephens, 2000), which can often be handled through appropriate estimation routines. A further discussion on this issue—including conditions that allow us to derive a proper estimate—is given in the Supplementary Material.

### 3 EM-type Algorithm

For fitting the SBSGM to a given network, the latent positions  $U_1, \dots, U_N$  and the parameters  $\zeta$  and  $\gamma$  need to be estimated simultaneously. This is a typical task for an EM algorithm, which aims at deriving information about unknown quantities in an iterative way. Regarding the inherent community structure, we assume the number of groups,  $K \in \mathbb{N}$ , as given for now. A discussion on that issue is provided in Section 3.3.

#### 3.1 MCMC E-Step

The conditional distribution of  $\mathbf{U}$  given  $\mathbf{y}$  is rather complex, and hence calculating the expectation cannot be solved analytically. Therefore, we apply MCMC techniques for carrying out the E-step. In that regard, the full-conditional distribution of  $U_i$  can be formulated as

$$f(u_i \mid u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N, \mathbf{y}) \propto \prod_{j \neq i} w_\zeta(u_i, u_j)^{y_{ij}} (1 - w_\zeta(u_i, u_j))^{1-y_{ij}}. \quad (12)$$

Based on that, we can construct a Gibbs sampler, which allows consecutive drawings for  $U_1, \dots, U_N$ . For its concrete implementation, we replace  $w_\zeta(\cdot, \cdot)$  by its current estimate. Finally, we derive reliable means for the node positions by appropriately summarizing the MCMC sequence. Technical details are provided in the Appendix.

We are however faced with an additional identifiability issue, which we want to motivate as follows. Assume first an SBSGM as in (7), but allow the distribution of the latent quantities  $U_i$  to be not necessarily uniform but arbitrarily continuous instead. In this scenario, the SBSGM specification is broadened to  $(F(\cdot), w_\zeta(\cdot, \cdot))$  with  $F(\cdot)$  as the distribution of the  $U_i$ 's. An equivalent network-generating process can then not only be constructed through permutations as in (10) but also by taking any strictly increasing continuous transformation  $\varphi' : [0, 1] \rightarrow [0, 1]$  and specifying  $F'(\cdot) \equiv F(\varphi'^{-1}(\cdot))$  and

$$w'_{\zeta'}(\cdot, \cdot) \equiv w_\zeta(\varphi'^{-1}(\cdot), \varphi'^{-1}(\cdot)) \quad (13)$$

with  $\zeta'_k = \varphi'(\zeta_k)$  for  $k = 1, \dots, K - 1$ . (Note that in comparison with formulation (10), here  $\varphi'(\cdot)$  implies a modification of the probability measure on the domain  $[0, 1]$ , and therefore it is *no* measure-preserving transformation—except for  $\varphi'(u) = u$ .) The two models  $(F(\cdot), w_\zeta(\cdot, \cdot))$  and  $(F'(\cdot), w'_{\zeta'}(\cdot, \cdot))$  are then not distinguishable in terms of the probability mass function induced on networks, causing an additional identifiability issue which, in fact, cannot be resolved by the EM algorithm. As a matter of conception, the EM algorithm aims at optimally fitting the data and not at accurately recovering the underlying model framework. Consequently, this needs to be handled post hoc, where we aim for model specification  $(F'(\cdot), w'_{\zeta'}(\cdot, \cdot))$  with  $F'(\cdot)$  representing the uniform distribution. We tackle this issue by making use of two separate transformations  $\varphi'_1, \varphi'_2 : [0, 1] \rightarrow [0, 1]$ , adjusting between and within groups, respectively. This is sketched in Figure 3, where we demonstrate both the theoretical and the empirical implementation. Adjustment 1 ensures that the extent of a component complies with its associated probability mass or relative frequency, respectively. This apparently involves adapting the component boundaries, which, compared to the blockmodeling framework, is analogous to adjusting community proportions. Adjustment 2 effects that the distribution within groups is uniform. Further details on the concrete implementation in the algorithm are given in the Appendix. We denote the final result of the E-step in the  $m$ -th iteration, i.e. the outcome achieved through Gibbs sampling and applying Adjustment 1 and Adjustment 2, by  $\hat{\mathbf{U}}''^{(m)} = (\hat{U}_1''^{(m)}, \dots, \hat{U}_N''^{(m)})$ .

## 3.2 M-Step

### 3.2.1 Linear B-Spline Regression

In consequence of representing the SBSGM as a mixture of (linear) B-splines as in (9), we are now able to view the estimation as semiparametric regression problem, which can be solved via maximum likelihood approach. Given the spline formulation, the full log-

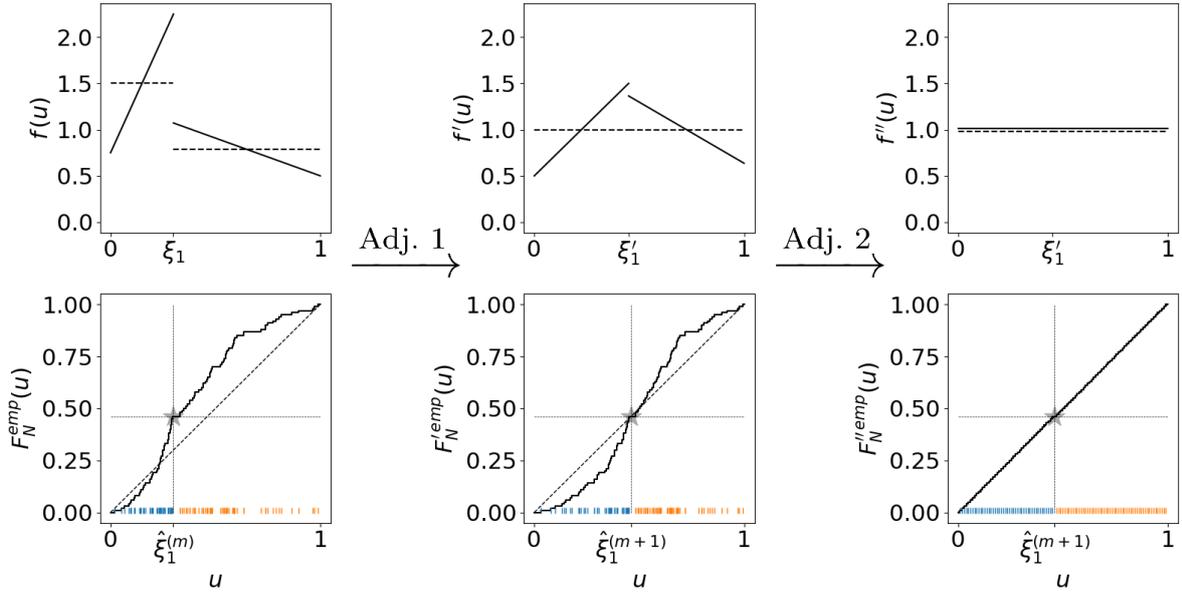


Figure 3: Adjustment of the distribution of the  $U_i$ 's and the community boundaries. Top: Three distributions for the latent quantities  $U_i$  which are equivalent in terms of representing the same data-generating process (under applying transformation (13) to  $w_\zeta(\cdot, \cdot)$  accordingly). The solid line represents the density  $f(u)$ , while the dashed line illustrates the frequency density over the communities, i.e.  $\mathbb{P}(U_i \in [\zeta_{k-1}, \zeta_k]) / (\zeta_k - \zeta_{k-1})$ . Bottom: Implementation of the adjustment in the algorithm with regard to the empirical cumulative distribution function (including the realizations of  $U_1, \dots, U_N$  as vertical bars at the bottom). The gray star illustrates the proportion of the two communities against the community boundary.

likelihood results in

$$\begin{aligned} \ell(\boldsymbol{\gamma}) = & \sum_{\substack{i,j \\ j \neq i}} \sum_{k,l} \mathbb{1}_{\{\hat{\zeta}_{k-1}^{(m+1)} \leq \hat{U}_i''^{(m)} < \hat{\zeta}_k^{(m+1)}\}} \mathbb{1}_{\{\hat{\zeta}_{l-1}^{(m+1)} \leq \hat{U}_j''^{(m)} < \hat{\zeta}_l^{(m+1)}\}} \\ & \cdot \left[ y_{ij} \log \left( \mathbf{B}_{kl,ij}^{(m+1)} \boldsymbol{\gamma}_{kl} \right) + (1 - y_{ij}) \log \left( 1 - \mathbf{B}_{kl,ij}^{(m+1)} \boldsymbol{\gamma}_{kl} \right) \right], \end{aligned}$$

where  $\mathbf{B}_{kl,ij}^{(m+1)} = \mathbf{B}_k^{(m+1)}(\hat{U}_i''^{(m)}) \otimes \mathbf{B}_l^{(m+1)}(\hat{U}_j''^{(m)})$  and  $\mathbf{B}_k^{(m+1)}(\cdot)$  is the B-spline basis on  $[\hat{\zeta}_{k-1}^{(m+1)}, \hat{\zeta}_k^{(m+1)}]$ . Taking the derivative leads to the score function

$$\begin{aligned} \mathbf{s}(\boldsymbol{\gamma}) = & \sum_{\substack{i,j \\ j \neq i}} \sum_{k,l} \mathbb{1}_{\{\hat{\zeta}_{k-1}^{(m+1)} \leq \hat{U}_i''^{(m)} < \hat{\zeta}_k^{(m+1)}\}} \mathbb{1}_{\{\hat{\zeta}_{l-1}^{(m+1)} \leq \hat{U}_j''^{(m)} < \hat{\zeta}_l^{(m+1)}\}} \\ & \cdot \mathbf{B}_{kl,ij}^{(m+1)\top} \left( \frac{y_{ij}}{w_{\hat{\zeta}^{(m+1)}, \boldsymbol{\gamma}}(\hat{U}_i''^{(m)}, \hat{U}_j''^{(m)})} - \frac{1 - y_{ij}}{1 - w_{\hat{\zeta}^{(m+1)}, \boldsymbol{\gamma}}(\hat{U}_i''^{(m)}, \hat{U}_j''^{(m)})} \right). \end{aligned}$$

Moreover, taking the expected second-order derivative gives us the Fisher matrix

$$\begin{aligned} \mathbf{F}(\boldsymbol{\gamma}) = & \sum_{\substack{i,j \\ j \neq i}} \sum_{k,l} \mathbb{1}_{\{\hat{\zeta}_{k-1}^{(m+1)} \leq \hat{U}_i''^{(m)} < \hat{\zeta}_k^{(m+1)}\}} \mathbb{1}_{\{\hat{\zeta}_{l-1}^{(m+1)} \leq \hat{U}_j''^{(m)} < \hat{\zeta}_l^{(m+1)}\}} \\ & \cdot \mathbf{B}_{kl,ij}^{(m+1)\top} \mathbf{B}_{kl,ij}^{(m+1)} \left[ w_{\hat{\zeta}^{(m+1)}, \boldsymbol{\gamma}}(\hat{U}_i''^{(m)}, \hat{U}_j''^{(m)}) \cdot \left( 1 - w_{\hat{\zeta}^{(m+1)}, \boldsymbol{\gamma}}(\hat{U}_i''^{(m)}, \hat{U}_j''^{(m)}) \right) \right]^{-1}. \end{aligned}$$

Based on that,  $\ell(\boldsymbol{\gamma})$  could now be maximized using Fisher scoring. However, we additionally need to ensure that the resulting estimate of  $w_{\hat{\zeta}(\cdot, \cdot)}$  in the  $(m+1)$ -th EM iteration fulfills symmetry and boundedness, which is why we impose additional (linear) side constraints on  $\boldsymbol{\gamma}$ . To guarantee symmetry, we accommodate  $\gamma_{kl,pq} = \gamma_{lk,qp}$  for all  $k, l \in \{1, \dots, K\}$  and  $p \neq q$ . Moreover, the condition of  $w_{\hat{\zeta}, \boldsymbol{\gamma}}^{spline}(\cdot, \cdot)$  being bounded to  $[0, 1]$  can be formulated as  $0 \leq \gamma_{kl,pq} \leq 1$ . These two side constraints are of linear form and can be incorporated through  $\mathbf{G}\boldsymbol{\gamma} \geq (\mathbf{0}^\top, -\mathbf{1}^\top)^\top$  and  $\mathbf{A}\boldsymbol{\gamma} = \mathbf{0}$  with matrices  $\mathbf{G}$  and  $\mathbf{A}$  chosen accordingly.  $\mathbf{0} = (0, \dots, 0)^\top$  and  $\mathbf{1} = (1, \dots, 1)^\top$  are vectors of corresponding sizes. Consequently, maximizing  $\ell(\boldsymbol{\gamma})$  with respect to the postulated side constraints can be considered as an (iterated) quadratic programming problem, which can be solved using standard software

(see e.g. Andersen et al., 2016 or Turlach and Weingessel, 2013).

### 3.2.2 Penalized Estimation

Following the ideas underlying the penalized spline estimation (see Eilers and Marx, 1996 or Ruppert et al., 2009), we additionally impose a penalty on the coefficients to achieve smoothness. This is necessary since we intend to choose the total basis length  $L$  of the mixture of B-splines to be large and unpenalized estimation will lead to wiggled estimates. For inducing smoothness within the components, we penalize the difference between “neighboring” elements of  $\boldsymbol{\gamma}_{kl}$ . Let therefore

$$\mathbf{D}_k = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(L_k-1) \times L_k}$$

be the first-order difference matrix. We then penalize  $[\mathbf{D}_k \otimes \mathbf{I}_l] \boldsymbol{\gamma}_{kl}$  and  $[\mathbf{I}_k \otimes \mathbf{D}_l] \boldsymbol{\gamma}_{kl}$ , where  $\mathbf{I}_k$  is the identity matrix of size  $L_k$ . This leads to the penalized log-likelihood

$$\ell^p(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\gamma}) - \frac{1}{2} \boldsymbol{\gamma}^\top \mathbf{Q}_\lambda \boldsymbol{\gamma},$$

where  $\mathbf{Q}_\lambda$  is the diagonal matrix  $\text{diag}\{\lambda_{11} \mathbf{Q}_{11}, \dots, \lambda_{1K} \mathbf{Q}_{1K}, \lambda_{21} \mathbf{Q}_{21}, \dots, \lambda_{KK} \mathbf{Q}_{KK}\}$  with  $\mathbf{Q}_{kl} = (\mathbf{D}_k \otimes \mathbf{I}_l)^\top (\mathbf{D}_k \otimes \mathbf{I}_l) + (\mathbf{I}_k \otimes \mathbf{D}_l)^\top (\mathbf{I}_k \otimes \mathbf{D}_l)$  and  $\boldsymbol{\lambda} = (\lambda_{11}, \dots, \lambda_{1K}, \lambda_{21}, \dots, \lambda_{KK})$  serving as vector of smoothing parameters for the respective blocks. In this configuration, the resulting estimate apparently depends on the penalty parameter vector  $\boldsymbol{\lambda}$ . Setting  $\lambda_{kl} \rightarrow 0$  for  $k, l = 1, \dots, K$  yields an unpenalized fit, while setting  $\lambda_{kl} \rightarrow \infty$  leads to a piecewise constant SBSGM, i.e. an SBM. Therefore, the smoothing parameter vector  $\boldsymbol{\lambda}$  needs to be chosen in a data-driven way. For example, this can be realized by relying on the Akaike Information Criterion (AIC) (see Hurvich and Tsai, 1989 or Burnham and

Anderson, 2002). In the present context, this can be formulated as

$$\text{AIC}(\boldsymbol{\lambda}) = -2 \ell(\hat{\boldsymbol{\gamma}}^p) + 2 \text{df}(\boldsymbol{\lambda}), \quad (14)$$

where  $\hat{\boldsymbol{\gamma}}^p$  is the penalized parameter estimate and  $\text{df}(\boldsymbol{\lambda})$  represents the cumulated degrees of freedom within the blocks. We define the latter in the usual way as the trace of the product of the inverse penalized Fisher matrix  $[\mathbf{F}^p]^{-1}(\hat{\boldsymbol{\gamma}}^p, \boldsymbol{\lambda})$  and the unpenalized Fisher matrix, see Wood (2017, page 211 and the following pages). To be precise, we define

$$\text{df}(\boldsymbol{\lambda}) = \text{tr} \{ [\mathbf{F}^p]^{-1}(\hat{\boldsymbol{\gamma}}^p, \boldsymbol{\lambda}) \mathbf{F}(\hat{\boldsymbol{\gamma}}^p) \}$$

with  $\text{tr}\{\cdot\}$  as the trace of a matrix. Making use of  $\text{df}_{kl}(\lambda_{kl}) = \text{tr}\{[\mathbf{F}_{kl}^p]^{-1}(\hat{\boldsymbol{\gamma}}_{kl}^p, \lambda_{kl}) \mathbf{F}_{kl}(\hat{\boldsymbol{\gamma}}_{kl}^p)\}$  with  $\mathbf{F}_{kl}(\hat{\boldsymbol{\gamma}}_{kl}^p)$  being the submatrix of  $\mathbf{F}(\hat{\boldsymbol{\gamma}}^p)$  which refers to the subvector  $\hat{\boldsymbol{\gamma}}_{kl}^p$  (and for the penalized fisher matrix equivalently), this calculation can be reduced to  $\text{df}(\boldsymbol{\lambda}) = \sum_{k,l} \text{df}_{kl}(\lambda_{kl})$  since  $[\mathbf{F}^p]^{-1}(\hat{\boldsymbol{\gamma}}^p, \boldsymbol{\lambda})$  and  $\mathbf{F}(\hat{\boldsymbol{\gamma}}^p)$  are both block diagonal matrices. Applying this simplification, we can rephrase (14) to

$$\text{AIC}(\boldsymbol{\lambda}) = \sum_{k,l} \{ -2 \ell_{kl}(\hat{\boldsymbol{\gamma}}_{kl}^p) + 2 \text{df}_{kl}(\lambda_{kl}) \}, \quad (15)$$

where  $\ell_{kl}(\cdot)$  is the partial likelihood of all potential connections falling into the  $(k, l)$ -th component. This representation allows us to optimize for  $\lambda_{kl}$  separately. Following this procedure finally leads us to parameter estimate  $\hat{\boldsymbol{\gamma}}^{(m+1)}$  in the  $(m + 1)$ -th iteration of the EM algorithm.

Note that the presented estimation routine might appear to be time-consuming and computationally intensive when compared to more efficient algorithms for related models. However, we emphasize that such approaches cannot be simply adopted due to the higher complexity of the SBSGM. Moreover, this iterative procedure promises a high accuracy, as demonstrated by the applications in Section 4 and the comparative analysis in Section 3.2 of the Supplementary Material. A detailed discussion on that is given in Section 1.2 of the Supplementary Material.

### 3.3 Choice of the Number of Communities

As for the SBM, the number of communities,  $K$ , is an essential parameter that, in real-world networks, is usually unknown. Preferably, this should also be inferred from the data. For this purpose, we take up two different intuitions, which can finally be brought together to construct an appropriate model selection criterion.

On the one hand, one might think of adopting methods for determining the number of communities from the SBM context. A common strategy under this framework relies on the Integrated Classification Likelihood (ICL) criterion (Daudin et al., 2008, Côme and Latouche, 2015, Mariadassou et al., 2010). However, for applying this to the SBSGM, one needs to observe the more complex structure. This is especially so because the structural complexity within communities and the number of groups can compensate for each other to a certain extent.

As an alternative approach, one could exploit the already formulated AIC from (14) by extending it towards a model selection strategy with respect to  $K$ . This can be accomplished by adding a corresponding term, for which we propose to apply the concept of the Bayesian Information Criterion (BIC). Combined with the previous formulation, this yields a mixing of AIC and BIC. To be precise, we select the smoothing parameter using the AIC as described above, but for the number of blocks, we impose a stronger penalty by replacing the factor of two with the logarithmized sample size. We consider this to be in line with Burnham and Anderson (2004), who conclude that the AIC is more reliable when the ground truth can be described through many tapering effects (smooth within-community structure), whereas the BIC is preferable when there exist only a few large effects (number of groups).

To formulate the BIC part, we first need to carefully think about how the model complexity grows with an increasing number of groups and what the corresponding sample size is. The degrees of freedom originating from the number of groups comprise two aspects, the  $K - 1$  boundary parameters  $\zeta_1, \dots, \zeta_{K-1}$  and the  $K^2$  basis connectivity parameters between and within communities (comparable to  $\mathbf{P}$  in the SBM context). The corresponding sample sizes are  $N$  (= number of nodes) for the boundary parameters and  $N(N - 1)$  (= number

of edges) for the connectivity parameters. In regard to extending criterion (14), we additionally have to take into account that the second component thereof already includes the degrees of freedom that are induced by the basis connectivity parameters. This can be easily seen by setting  $\lambda_{kl} \rightarrow \infty$ , leading to  $\text{df}(\boldsymbol{\lambda}) = K^2$ . To correct for this,  $K^2$  needs to be subtracted from  $\text{df}(\boldsymbol{\lambda})$ , which, however, has no effect on the optimization with respect to  $\boldsymbol{\lambda}$ . Putting all together, we achieve the complete model selection criterion

$$-2\ell(\hat{\boldsymbol{\zeta}}_K, \hat{\boldsymbol{\gamma}}_K) + 2\{\text{df}(\hat{\boldsymbol{\lambda}}_K) - K^2\} + \log\{N(N-1)\}K^2 + \log\{N\}(K-1), \quad (16)$$

where  $\hat{\boldsymbol{\zeta}}_K$ ,  $\hat{\boldsymbol{\gamma}}_K$ , and  $\hat{\boldsymbol{\lambda}}_K$  are the final estimates according to the above EM procedure under a fixed  $K$ . We emphasize that criterion (16) is equivalent to the ICL up to the different parameterization of the log-likelihood and the term  $2\{\text{df}(\hat{\boldsymbol{\lambda}}_K) - K^2\}$  for penalizing the smooth structure within communities. When setting  $\lambda_{kl} \rightarrow \infty$  for  $k, l = 1, \dots, K$ , the criterion even reduces exactly to the ICL for SBMs.

## 4 Application

We investigate the performance of our approach with respect to both simulated and real-world networks. For an “uninformative” implementation, we initialize the algorithm by setting  $(U_1, \dots, U_N)$  to a random permutation of  $(i/(N+1) : i = 1, \dots, N)$ . At the same time, we place the community boundaries equidistantly within  $[0, 1]$ , i.e. we set  $\hat{\zeta}_k^{(0)} = k/K$  for  $k = 0, \dots, K$ . Since different initializations might lead to different final results, we repeat the estimation procedure with varying random permutations for  $\hat{\boldsymbol{U}}^{(0)}$  and, in the end, choose the best outcome. As a rough guideline, we recommend five to ten random repetitions. Moreover, if the number of groups is unknown beforehand, we fit the SBSGM with varying  $K$  and employ criterion (16) to select the optimal estimate. The variability of the final result under different initializations of  $\boldsymbol{U}$  and settings of  $K$  as well as the overall prediction performance is illustrated for all of the following applications in Figure 4 and Figure 5 of the Supplementary Material, respectively. These evaluations include the comparison with state-of-the-art methods for uncovering the structure in complex networks.

$K$	1	2	3	4	5	6
Assortative network (see Section 4.1.1)	8.960	8.978	<b>8.953</b>	8.956	8.963	8.981
Core-periphery network (Section 4.1.2)	<b>9.799</b>	9.804	9.825	9.833	9.843	9.846
Network with differing preferences (Section 4.1.3)	<b>10.287</b>	10.330	10.394	10.373	10.402	10.422
Political blogs (Section 4.2.1)	15.928	<b>15.663</b>	15.785	15.991	16.214	16.173
Human brain functional coactivations (Section 4.2.3)	13.256	13.197	<b>13.133</b>	13.142	13.206	13.287
$K$	5	6	7	8	9	10
Military alliances (Section 4.2.2)	2.578*	2.379*	<b>2.268*</b>	2.412*	2.807*	2.973*

Table 1: Resulting values of criterion (16) for all network examples considered in Section 4. Specification refers to the factor of  $10^4$  (\*or  $10^3$ ). The lowest value per network is highlighted in bold.

## 4.1 Synthetic Networks

In the scenario of simulations, we order the final estimate according to the ground-truth model, involving the within- and between-group arrangement. That is, we apply  $\varphi : [0, 1] \rightarrow [0, 1]$  from (10) to swap communities and reverse within-group arrangements from back to front if visibly adequate.

### 4.1.1 Assortative Structure with Smooth Within-Group Differences

To showcase the general applicability of our method, we first consider again the SBSGM from Figure 1. Starting with determining the number of groups, the first row of Table 1 shows the corresponding values for criterion (16). This suggests choosing the correct number of three communities. The corresponding estimation results, i.e. under  $K = 3$ , are illustrated in Figure 4. It can be clearly seen that the estimate of  $w_{\zeta}(\cdot, \cdot)$  (top right plot)

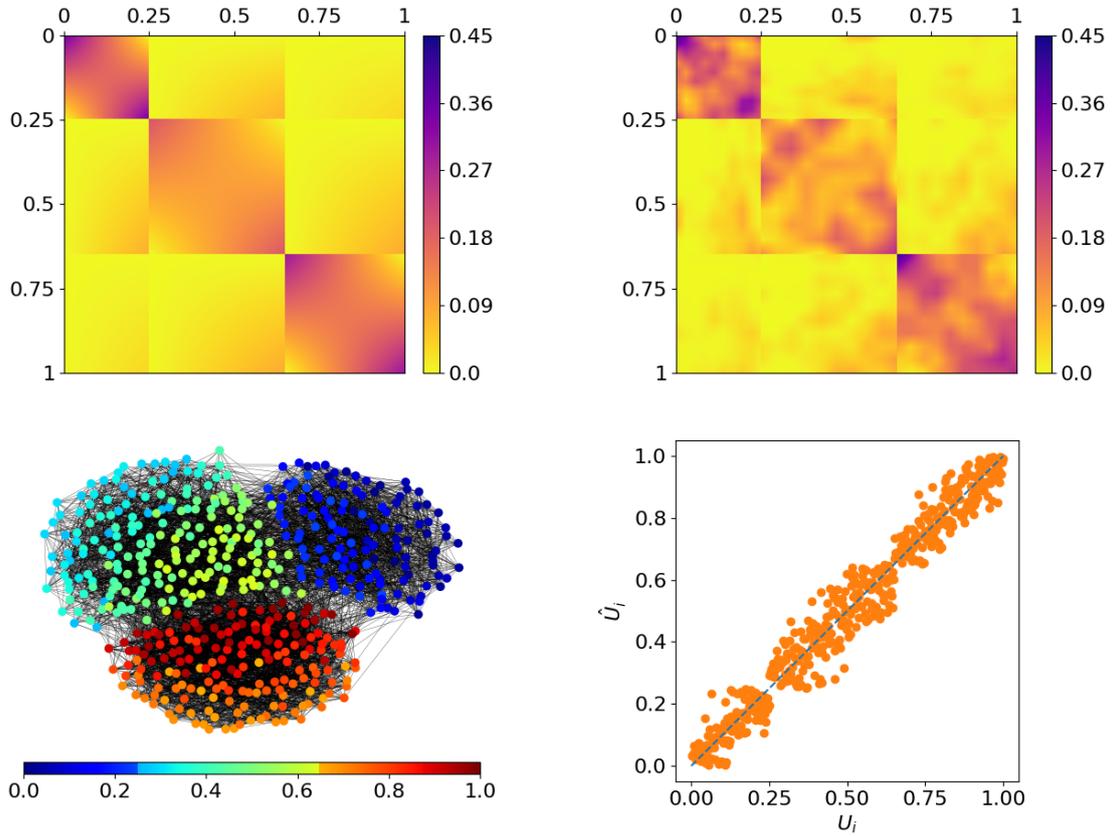


Figure 4: Estimation results for the synthetic SBSGM from Figure 1 (shown again at top left, rescaled according to the estimate's range from 0 to 0.45). The top right plot shows the final estimate of  $w_\zeta(\cdot, \cdot)$ , i.e. after convergence of the algorithm. The estimation is based on the simulated network of size  $N = 500$  at the bottom left, where nodes are colored according to  $\hat{U}_i \in [0, 1]$ . A comparison between the simulated  $U_i$ 's and their estimates is illustrated at the bottom right.

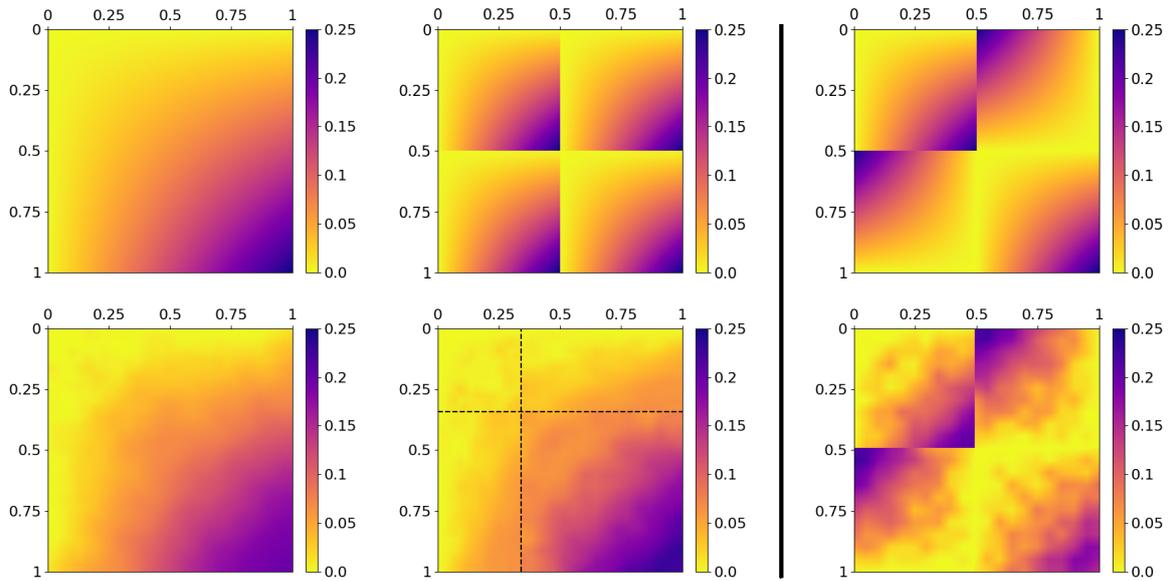


Figure 5: Estimation results for two synthetic SBSGMs (upper row). The left model is illustrated by two different representations, including different numbers of groups. The corresponding estimates with number of groups adopted from the model representations above are illustrated in the lower row. The estimation is based on simulated networks of size  $N = 500$ .

precisely captures the structure of the true model (top left). In line with this, comparing the estimated node positions with the true simulated ones (bottom right) reveals that the latent quantities are appropriately recovered. More precisely, all three truly underlying groups are clearly separated, and, in addition, also the within-community positions are well replicated. Altogether, the underlying structure can be accurately uncovered.

#### 4.1.2 Core-Periphery Structure

As a second simulation example, we consider the model at the top left of Figure 5, which is illustrated as two different formations. Apparently, the “true” number of groups here is  $K = 1$ , meaning that the left model is preferred. Hence, the estimation procedure should only follow that representation, regardless of the applied  $K$ . We demonstrate that our algorithm adheres to that by fitting the model with both settings,  $K = 1$  and  $K = 2$ , which also helps to understand the algorithm’s intuition. The results for simulated networks of size  $N = 500$  are illustrated in the lower row of the left-hand side of Figure 5.

Both estimates follow the “single-community” representation, demonstrating the method’s implicit strategy of merging similar nodes. In addition, a comparison of the estimates with regard to minimizing criterion (16) (see second row of Table 1) reveals that the model fit with  $K = 1$  seems preferable over the one with  $K = 2$ .

### 4.1.3 Mixture of Assortative and Disassortative Structures under Equal Overall Attractiveness

We now amend the previous situation in the following spirit, where we refer to the representation with two communities. Instead of nodes being generally either weakly or strongly connected, they should now be either strongly connected within their own group and weakly connected into the respective other group or vice versa. This leads us to the SBSGM specification represented in the top right plot of Figure 5. More precisely, all nodes in this model have the same expected degree, where nodes being weakly connected within their own community compensated for their lack of attractiveness by reaching out to members of the respective other community. This kind of structure clearly cannot be collapsed to a single-community SBSGM. More importantly, it also cannot be captured by the SBM even under degree correction. Our algorithm is however still able to fully capture the underlying structure, as shown by the estimate at the bottom right. Note that applying criterion (16) in this case actually suggests choosing  $K = 1$  (third row of Table 1). This wrong specification might be caused by the fact that the structural break at 0.5 is only half. Besides, the decision is quite close compared to the setting of  $K = 2$ .

## 4.2 Real-World Networks

For evaluating our method with regard to real-world examples, we consider three networks from different domains, comprising social and political sciences as well as neurosciences. Besides their different domains, the networks vary in their inherent structure, including the overall density. An overview of the networks’ most relevant summarizing attributes is given in Table 2.

	Number of nodes	Average degree	Overall density
Political blogs	1222	27.31	0.022
Military alliances	141	24.16	0.173
Human brain functional coactivations	638	58.39	0.092

Table 2: Details about real-world networks used as application examples.

#### 4.2.1 Political Blogs

The political blog network has been assembled by Adamic and Glance (2005) and consists of 1222 nodes (after extracting the largest connected component). These nodes represent political blogs of which 586 are liberal and 636 are conservative, according to manual labeling (Adamic and Glance, 2005). An edge between two blogs illustrates a web link pointing from one blog to another within a single-day snapshot in 2005. For our purpose, these links are interpreted in an undirected fashion. The resultant network is illustrated in the top plot of Figure 6, with nodes exhibiting political labels. (Note that this is the same network used by Karrer and Newman (2011) for demonstrating the enhancement achieved through their degree-corrected variant of the SBM.) Criterion (16) here suggests an SBSGM with  $K = 2$  (see fourth row of Table 1), which is in accordance with the number of political orientations. Moreover, the predicted group assignments (bottom right plot) reveal a broad concordance with the manually assigned labels. This is additionally underpinned by a similar size ratio of 565 (mostly liberals) to 657 (mostly conservatives). Information beyond community memberships can be gained through the within-community positions visualized in the middle right plot. Together with the estimate of  $w_{\zeta}(\cdot, \cdot)$  (middle left plot), they reveal, for example, a local structure that is dominated by a division into core and periphery nodes. The emergence of such a core-periphery structure is a well-known phenomenon in the linking of the World Wide Web, of which the considered blogs are an extraction. In a more specific sense, this core-periphery structure mirrors the blogs’ “sociability,” i.e. their involvement in the political discourse. The narrow intense regions in  $\hat{w}_{\zeta}(\cdot, \cdot)$  and the steep slopes in its marginal function (cf. formulation (11); bottom left

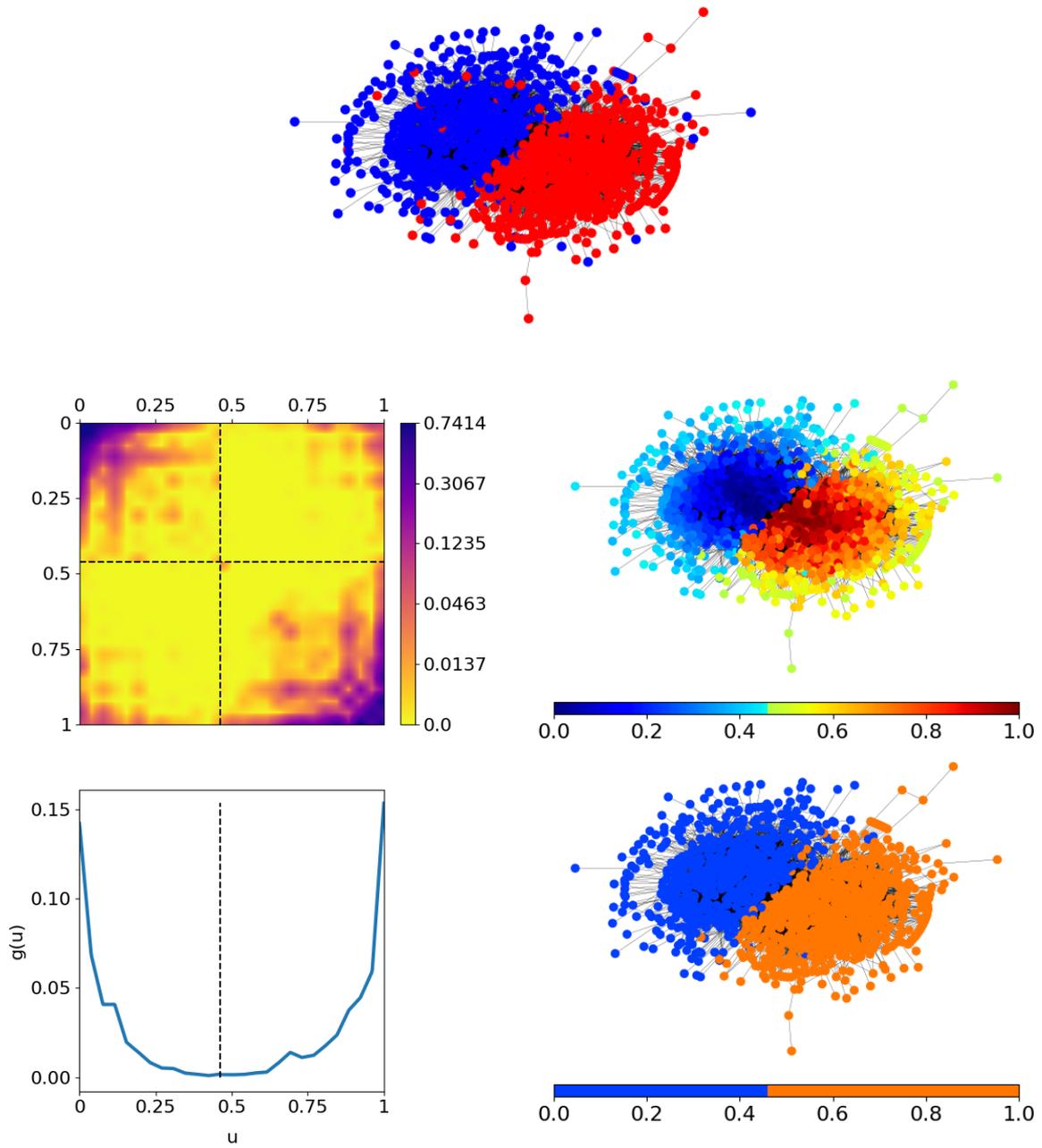


Figure 6: SBSGM estimation for the political blog network (top plot, with blue for ‘liberal’ and red for ‘conservative’). The estimate of  $w_{\zeta}(\cdot, \cdot)$  (in log scale) is depicted at the middle left. The plot at the bottom left illustrates the corresponding marginal function  $\hat{g}_{\zeta}(u) = \int \hat{w}_{\zeta}(u, v) dv$ . In the right column of the two lower rows, the network is shown with coloring referring once to the node positions  $\hat{U}_i$  (top) and once to the derived community memberships (bottom).

plot) further indicate the presence of hubs, meaning small subgroups of nodes that are much more densely connected than others. Lastly, the model fit reveals a domination of assortative structures since the overall intensity within the two communities is much higher than between them. Together with the inferred grouping by political orientations, this indicates that blogs tend to link to blogs from the same political side rather than to blogs of the respective other side.

#### 4.2.2 Military Alliances

As a second real-world example, we consider the military alliances among the world's nations. These data have been gathered and are provided by the Alliance Treaty Obligations and Provisions project (Leeds et al., 2002). More specifically, from the available data, we extract only the strong military alliances which were lately in force. That means an edge between two countries is inserted if they have a current military agreement that constitutes an offensive or defensive pact which would force the one country to militarily intervene when the other one has come into an offensive or defensive military conflict. This network, which, according to criterion (16), decomposes into seven communities (see last row of Table 1), is illustrated in the top right plot of Figure 7. Also in this data example, the estimate of  $w_{\zeta}(\cdot, \cdot)$  (top left) reveals a very dominant assortative structure, even though few groups also strongly connect to other groups. Further conclusions about political constellations can be drawn by transferring the inferred node positions and the deduced community memberships to the world map, as shown in the two lower plots. The resulting communities exhibit a composition of almost exclusively neighboring states, implying that those arrange similar strong military alliances. Relating this to the discovered assortative structure allows to additionally conclude that countries which are geographically close are likely to form military alliances. Consulting further the within-community positions provides insight into the communities' local formation. For example, the arrangement of the yellow community goes from Southern Africa via Central/East Africa to West Africa. In general, the structure within the network seems to be well captured by node clustering with smooth within-group differences.

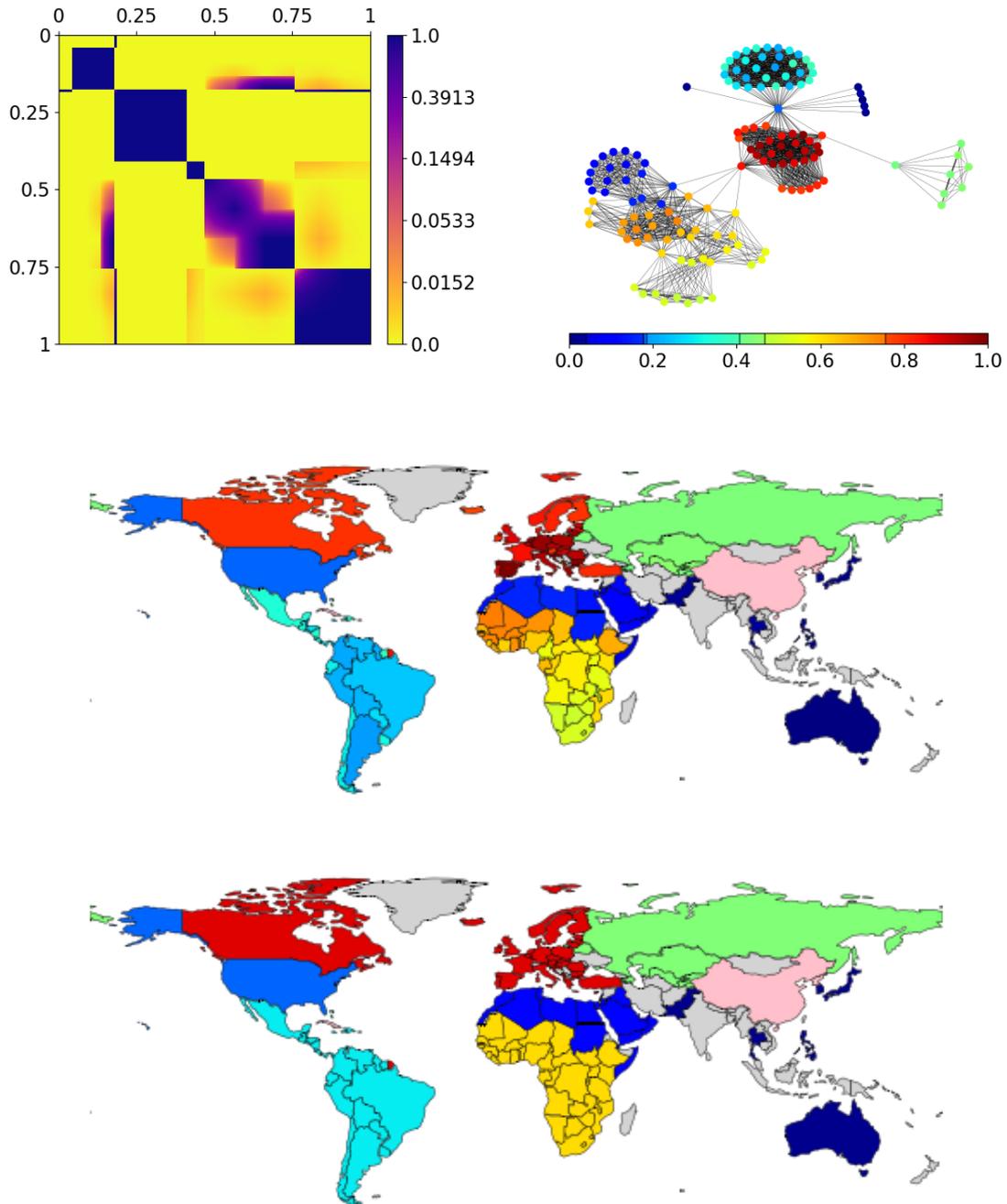


Figure 7: SBSGM estimation for the military alliance network. The estimate of  $w_\zeta(\cdot, \cdot)$  (in log scale) and the network with node coloring referring to  $\hat{U}_i \in [0, 1]$  are depicted at the top left and top right, respectively. The two lower plots show the world map with colors indicating the precise positions in the SBSGM (middle) and the resulting community memberships (bottom). China, Cuba, and North Korea (colored in pink) form an isolated group and therefore have been excluded from the estimation procedure. Countries which do not appear in the data set and thus are assumed not to have any strong military alliance are colored in gray.

### 4.2.3 Human Brain Functional Coactivations

We conclude the real-world data examples by considering the human brain functional coactivation network. This network has been assembled by Crossley et al. (2013) via meta-analysis and is accessible through the Brain Connectivity Toolbox (Rubinov and Sporns, 2010). More precisely, the provided weighted network matrix represents the “estimated [...] similarity (Jaccard index) of the activation patterns across experimental tasks between each pair of 638 brain regions” (Crossley et al., 2013), where this similarity is additionally “probabilistically thresholded.” Based on that, we construct an unweighted graph by including a link between all pairs of brain regions which have a significant similarity, meaning a positive score in the original data set. For the arising network, determining the number of groups using criterion (16) yields three communities (see last-but-one row of Table 1). The decision, however, seems rather tight, and Crossley et al. (2013) choose a regular SBM with four communities for fitting these data. Hence, we also here choose  $K = 4$ , which allows a direct comparison of the results. The resulting SBSGM fit is illustrated in Figure 8. In this application,  $\hat{w}_\zeta(\cdot, \cdot)$  (top left) again reveals an assortative structure. However, it is comparatively less pronounced, implicating that similar behavior and connectedness are less strongly associated. Regarding the inferred clustering in anatomical representation (bottom row), the results look very similar to those derived by Crossley et al. (2013, Fig. 1, A). They label their inferred communities anatomically as occipital, central, frontoparietal, and default-mode module. As the biggest difference to that, our method proposes to merge the occipital and the frontoparietal part (yielding the red group) and instead to split the central module into an upper and a lower section (light green and cyan node set, respectively). Yet, the within-group positions of the merged red community reveal analogically a certain disintegration into brain regions of the occipital and the frontoparietal module. Hence, the additional information about structural patterns within communities here allows to uncover a clustering of nodes at a higher resolution. (For the sake of completeness, the SBSGM fit with  $K = 3$  is illustrated in the Supplementary Material. These results look quite similar, but they reveal a more extensive exploitation of capturing group structures through smooth differences.)

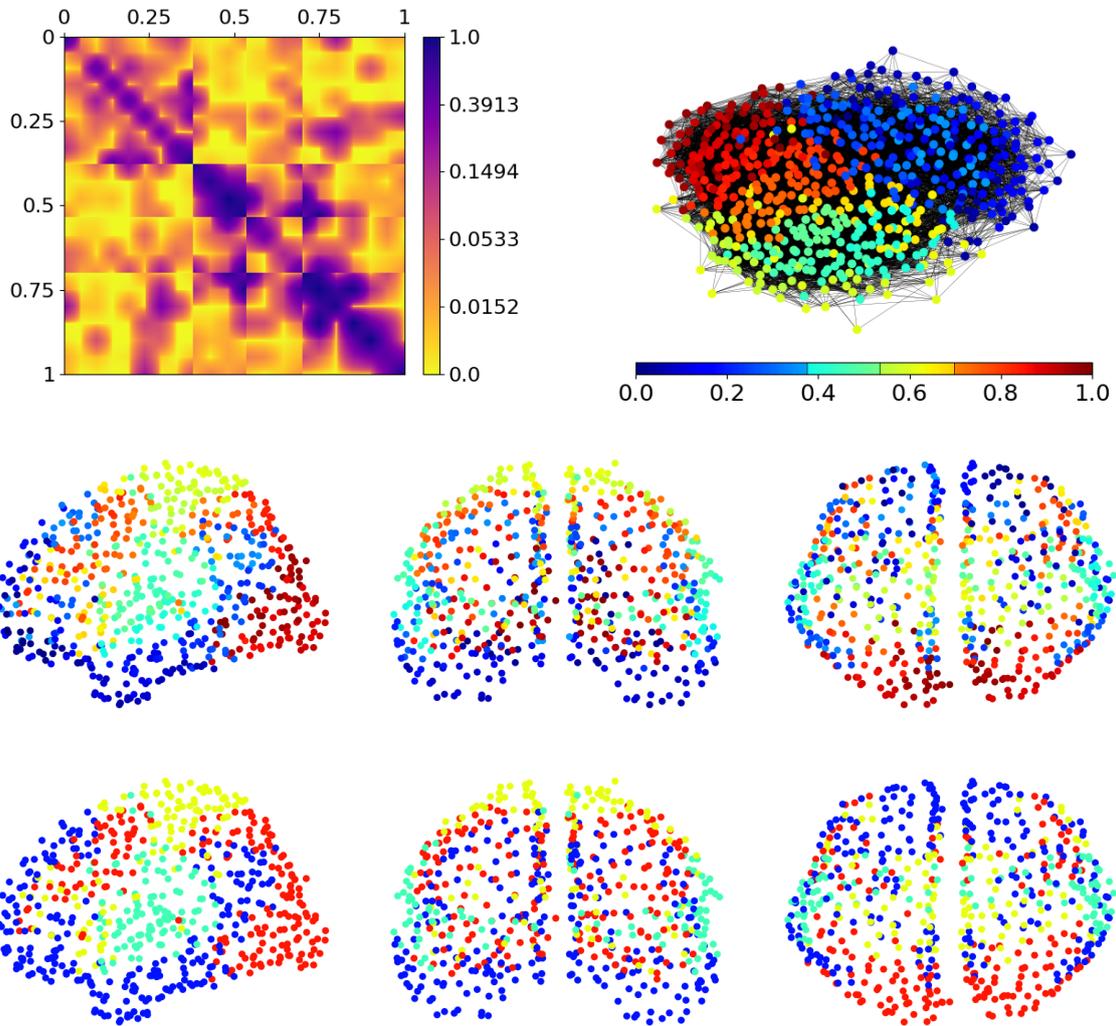


Figure 8: SBSGM estimation for the human brain functional coactivation network (top right, coloring referring to  $\hat{U}_i \in [0, 1]$ ). The estimate of  $w_\zeta(\cdot, \cdot)$  (in log scale) is depicted at the top left. The lower six plots show the local positions of the human brain regions in anatomical space with coloring referring to  $\hat{U}_i \in [0, 1]$  (middle row) and derived community memberships (bottom row). The different angles show side view (left column), front view (middle column), and top view (right column).

Altogether, the applications demonstrate that our novel approach is a very flexible modeling tool. As such, the SBSGM allows to precisely capture the structure within complex networks, which consequently yields a high prediction accuracy (see Supplementary Material). Moreover, the strict node clustering paired with within-group positions provides favorable interpretability. These features make the SBSGM a helpful tool that enables to gain further insight and to draw more profound scientific conclusions.

## 5 Discussion and Conclusion

Despite their close relationship, the stochastic blockmodel and the (smooth) graphon model have mainly been developed separately until now. To address this shortcoming, the paper aimed at combining both model formulations to create a novel modeling approach that generalizes the two detached concepts in a unified framework. The resulting stochastic block smooth graphon model consequently unites the individual capabilities of the two initial approaches. These are the clustering of networks (SBM) and the representation of local structures by formulating smoothly changing connectivity behaviors (SGM). By utilizing previous results on SBM and SGM estimation, we developed an EM-type algorithm for reliably estimating the (semiparametric) model specification.

Although the SBSGM arises from combining the SBM and the SGM, connections to other statistical network models can be established. In this line, an approximation of the latent distance model (Hoff et al., 2002) can be formulated for any dimension by appropriately partitioning and mapping the latent space to the unit interval as the domain of the SBSGM. Moreover, our method allows to cover the structure of the degree-corrected SBM (Karrer and Newman, 2011) in a natural way. This can be accomplished by restricting the slices  $w_{\zeta}(u, \cdot)$  to be proportional within communities, i.e. by setting  $w_{\zeta}(u, \cdot) \equiv c \cdot w_{\zeta}(u_0, \cdot)$  for  $u \in [\zeta_{k_{u_0}-1}, \zeta_{k_{u_0}})$  with  $c \in \mathbb{R}_+$ . Such a specification consequently implies an equivalent connectivity behavior with different attractiveness. The applicability to such a situation has been demonstrated by the political blog example of Section 4.2.1. Finally, we argue that—from a conceptual perspective—the SBSGM is also related to the hierarchical exponential random graph model developed by Schweinberger and Handcock (2015). This

is because in both models, the set of nodes is divided into “neighborhoods” (Schweinberger and Handcock, 2015) within which the structure is then modeled by a further approach, namely an ERGM or multiple SGMs, respectively. In this regard, the HERGM could be enhanced by modeling the between-neighborhood structure through an SBSGM instead of an SBM. Lastly, a connection between the graphon model and the ERGM has been elaborated by Chatterjee and Diaconis (2013), Yin et al. (2016), and Krioukov (2016). Further details on the links to other models are provided in the Supplementary Material.

Besides its theoretical capabilities and connections, we demonstrated the practical applicability of the SBSGM. This has been done with respect to both simulated and real-world networks. The estimation results of Section 4 clearly reveal that our novel modeling approach of clustering nodes into groups with smooth structural differences is able to capture various types of complex structural patterns. Overall, the SBSGM is a very flexible tool for modeling complex networks, which helps to uncover the network’s structure in more detail and thus enables to get a better understanding of the underlying processes.

## Acknowledgment

The project was partially supported by the European Cooperation in Science and Technology [COST Action CA15109 (COSTNET)].

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest: none.

## SUPPLEMENTARY MATERIAL

The Supplementary Material comprises elaborations about intuition and justification of the EM-type algorithm, computational issues, links to other models, and an evaluation of the algorithm’s concrete performance in comparison with other methods. Moreover, we have implemented the EM-based estimation routine described in the paper in a free and open source `Python` package, which is publicly available on github (<https://github.com/BenjaminSischka/SBSGMest.git>, Sischka, 2022).

## References

- Adamic, L. A. and N. Glance (2005). The political blogosphere and the 2004 U.S. Election: Divided they blog. In *3rd International Workshop on Link Discovery, LinkKDD 2005 - in conjunction with 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 36–43.
- Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* 9, 1981–2014.
- Airoldi, E. M., T. B. Costa, and S. H. Chan (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems 26*, pp. 692–700.
- Andersen, M., J. Dahl, and L. Vandenberghe (2016). CvxOpt: Open source software for convex optimization (Python). Available online: <https://cvxopt.org> [accessed 11-27-2020]. Version 1.2.7.
- Bickel, P., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Annals of Statistics*.
- Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences of the United States of America* 106(50), 21068–21073.
- Borgs, C., J. Chayes, and L. Lovász (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geometric and Functional Analysis* 19(6), 1597–1619.
- Borgs, C., J. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi (2007). Counting graph homomorphisms. In *Topics in Discrete Mathematics*, pp. 315–371.
- Burnham, K. and D. Anderson (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. (2nd ed.). New York: Springer.

- Burnham, K. P. and D. R. Anderson (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research* 33(2), 261–304.
- Chan, S. H. and E. M. Airoldi (2014). A consistent histogram estimator for exchangeable graph models. In *31st International Conference on Machine Learning, ICML 2014*.
- Chatterjee, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *Annals of Statistics* 43(1), 177–214.
- Chatterjee, S. and P. Diaconis (2013). Estimating and understanding exponential random graph models. *Annals of Statistics* 41(5), 2428–2461.
- Chen, K. and J. Lei (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association* 113(521), 241–251.
- Choi, D. S., P. J. Wolfe, and E. M. Airoldi (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* 99(2), 273–284.
- Côme, E. and P. Latouche (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling* 15(6), 564–589.
- Crossley, N. A., A. Mechelli, P. E. Vértes, T. T. Winton-Brown, A. X. Patel, C. E. Ginestet, P. McGuire, and E. T. Bullmore (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences of the United States of America* 110(28), 11583–11588.
- Daudin, J. J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and Computing* 18(2), 173–183.
- De Nicola, G., B. Sischka, and G. Kauermann (2022). Mixture models and networks: The stochastic blockmodel. *Statistical Modelling*. [Available online.] doi:10.1177/1471082X211033169.

- Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 84(6).
- Diaconis, P. and S. Janson (2007). Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*.
- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.
- Fienberg, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics* 21(4), 825–839.
- Fosdick, B. K., T. H. McCormick, T. B. Murphy, T. L. J. Ng, and T. Westling (2019). Multiresolution network models. *Journal of Computational and Graphical Statistics* 28(1), 185–196.
- Gao, C., Y. Lu, and H. H. Zhou (2015). Rate-optimal graphon estimation. *Annals of Statistics* 43(6), 2624–2652.
- Gao, C. and Z. Ma (2020). Discussion of 'Network cross-validation by edge sampling'. *Biometrika* 107(2), 281–284.
- Geng, J., A. Bhattacharya, and D. Pati (2019). Probabilistic Community Detection With Unknown Number of Communities. *Journal of the American Statistical Association* 114(526), 893–905.
- Ghasemian, A., H. Hosseinmardi, A. Galstyan, E. M. Airoldi, and A. Clauset (2020). Stacking models for nearly optimal link prediction in complex networks. *Proceedings of the National Academy of Sciences of the United States of America* 117(38), 23393–23400.
- Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoldi (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning* 2(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 in Society* 170(2), 301–354.
- Hoff, P. D. (2007). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*.
- Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory* 15(4), 261–272.
- Hoff, P. D. (2021). Additive and multiplicative effects network models. *Statistical Science* 36(1), 34–50.
- 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. W., K. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks* 5, 109–137.
- Hunter, D. R., P. N. Krivitsky, and M. Schweinberger (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics* 21(4), 856–882.
- Hurvich, C. M. and C. L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* 76(2), 297–307.
- Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 83(1).
- Kauermann, G. and J. D. Opsomer (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika* 98(1), 225–230.
- Kemp, C., J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence*, Volume 1, pp. 381–388.

- Klopp, O., A. B. Tsybakov, and N. Verzelen (2017). Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics* 45(1), 316–354.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data*. New York: Springer.
- Kolaczyk, E. D. (2017). *Topics at the frontier of statistics and network analysis*. Cambridge: Cambridge University Press.
- Kolaczyk, E. D. and G. Csardi (2014). *Statistical analysis of network data with R* (2nd ed.). New York: Springer.
- Krioukov, D. (2016). Clustering implies geometry in networks. *Physical Review Letters* 116(20).
- Latouche, P. and S. Robin (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing* 26(6), 1173–1185.
- Leeds, B. A., J. M. Ritter, S. M. L. Mitchell, and A. G. Long (2002). Alliance treaty obligations and provisions, 1815-1944. *International Interactions* 28(3), 237–260.
- Li, T. and C. M. Le (2021). Network estimation by mixing: Adaptivity and more. *arXiv preprint arXiv:2106.02803*.
- Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika* 107(2), 257–276.
- Li, Y., X. Fan, L. Chen, B. Li, and S. A. Sisson (2022). Smoothing graphons for modelling exchangeable relational data. *Machine Learning* 111(1), 319–344.
- Lloyd, J. R., P. Orbanz, Z. Ghahramani, and D. M. Roy (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, Volume 2, pp. 998–1006.
- Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory. Series B* 96(6), 933–957.

- Lusher, D., J. Koskinen, and G. Robins (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.
- Ma, Z., Z. Ma, and H. Yuan (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research* 21, 1–67.
- Mariadassou, M., S. Robin, and C. Vacher (2010). Uncovering latent structure in valued graphs: A variational approach. *Annals of Applied Statistics* 4(2), 715–742.
- Matias, C. and S. Robin (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. *ESAIM: Proceedings and Surveys* 47, 55–74.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* 103(23), 8577–8582.
- Newman, M. E. and G. Reinert (2016). Estimating the number of communities in a network. *Physical Review Letters* 117(7).
- Nowicki, K. and T. A. Snijders (2001). Estimation and prediction for stochastic block-structures. *Journal of the American Statistical Association* 96(455), 1077–1087.
- Olhede, S. C. and P. J. Wolfe (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences* 111(41), 14722–14727.
- Orbanz, P. and D. M. Roy (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(2), 437–461.
- Peixoto, T. P. (2012). Entropy of stochastic blockmodel ensembles. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 85(5).
- Peixoto, T. P. (2017). Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 95(1).

- Riolo, M. A., G. T. Cantwell, G. Reinert, and M. E. Newman (2017). Efficient method for estimating the number of communities in a network. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 96(3).
- Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics* 39(4), 1878–1915.
- Rubinov, M. and O. Sporns (2010). Brain connectivity toolbox (MATLAB). Available online: <https://sites.google.com/site/bctnet> [accessed 02-15-2021].
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2009). Semiparametric regression during 2003-2007. *Electronic Journal of Statistics* 3, 1193–1256.
- Salter-Townshend, M., A. White, I. Gollini, and T. B. Murphy (2012). Review of statistical network analysis: Models, algorithms, and software. *Statistical Analysis and Data Mining* 5(4), 243–264.
- Schweinberger, M. and M. S. Handcock (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 77(3), 647–676.
- Sischka, B. (2022). EM-based estimation of stochastic block smooth graphon models (Python). Available online: <https://github.com/BenjaminSischka/SBSGMest.git> [accessed 03-24-2022].
- Sischka, B. and G. Kauermann (2022). EM-based smooth graphon estimation using MCMC and spline-based approaches. *Social Networks* 68, 279–295.
- Snijders, T. A. (2011). Statistical models for social networks. *Annual Review of Sociology* 37, 131–153.
- Snijders, T. A. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* 14(1), 75–100.

- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 62(4), 795–809.
- Turlach, B. A. and A. Weingessel (2013). quadprog: Functions to solve quadratic programming problems (R). Available online: <https://CRAN.R-project.org/package=quadprog> [accessed 11-27-2020]. Version 1.5-5.
- Wang, Y. X. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. *Annals of Statistics* 45(2), 500–528.
- Wolfe, P. J. and S. C. Olhede (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton: CRC Press.
- Yang, J. J., Q. Han, and E. M. Airolidi (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Journal of Machine Learning Research, Conference and Workshop Proceedings 33, PMLR*, pp. 1060–1067.
- Yin, M., A. Rinaldo, and S. Fadnavis (2016). Asymptotic quantization of exponential random graphs. *Annals of Applied Probability* 26(6), 3251–3285.
- Zhang, Y., E. Levina, and J. Zhu (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* 104(4), 771–783.

## Appendix

### Gibbs Sampling of Node Positions and Subsequent Adjustments

In the EM-type algorithm presented in the paper, we apply the Gibbs sampler in the E-step to achieve appropriate node positions conditional on  $\mathbf{Y} = \mathbf{y}$  and given  $w_{\zeta}(\cdot, \cdot)$ . That

means, we aim to stepwise update the  $i$ -th component of the current state of the Markov chain,  $\mathbf{U}^{<t>} = (U_1^{<t>}, \dots, U_N^{<t>})$ . This is done by setting  $U_j^{<t+1>} := U_j^{<t>}$  for  $j \neq i$  and for  $U_i^{<t+1>}$  drawing from the full-conditional distribution as formulated in (12) of the paper. To do so, we make use of a mixture proposal which differentiates between remaining within and switching the community. This is appropriate due to different structural relations with respect to  $U_i^{<t>}$ , where nearby positions within the same community imply similar connectivity patterns. To this end, we split the proposing procedure into two separate steps. First, we randomly choose the proposal type, i.e. either remaining within or switching the community. This is done by drawing from a Bernoulli distribution with  $\nu \in [0, 1]$  as the probability of remaining within the group. Secondly, conditional on the proposal type, we either draw from  $[\zeta_{k_i-1}, \zeta_{k_i})$  or from  $[0, 1] \setminus [\zeta_{k_i-1}, \zeta_{k_i})$ , where  $k_i \in \{1, \dots, K\}$  is the community including  $U_i^{<t>}$ , i.e.  $U_i^{<t>} \in [\zeta_{k_i-1}, \zeta_{k_i})$ . For a proposal within the current community, we employ a normal distribution under a compressed logit link. To be precise, we first define  $V_i^{<t>} = \log\{(U_i^{<t>} - \zeta_{k_i-1})/(\zeta_{k_i} - U_i^{<t>})\}$ , then we draw  $V_i^*$  from  $\text{Normal}(V_i^{<t>}, \sigma^2)$  with an appropriate value for the variance  $\sigma^2$ , and finally we calculate  $U_i^* = \{\exp(V_i^*)/(1 + \exp(V_i^*))\} \cdot (\zeta_{k_i} - \zeta_{k_i-1}) + \zeta_{k_i-1}$ . Consequently, for  $U_i^* \in [\zeta_{k_i-1}, \zeta_{k_i})$ , the proposal density follows

$$p(U_i^* | U_i^{<t>}) \propto \nu \cdot \frac{1}{(U_i^* - \zeta_{k_i-1})(\zeta_{k_i} - U_i^*)} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \left[ \log \left\{ \frac{(U_i^* - \zeta_{k_i-1})}{(\zeta_{k_i} - U_i^*)} \right\} - \log \left\{ \frac{(U_i^{<t>} - \zeta_{k_i-1})}{(\zeta_{k_i} - U_i^{<t>})} \right\} \right]^2 \right\},$$

yielding a ratio of proposals in the form of

$$\frac{p(U_i^{<t>} | U_i^*)}{p(U_i^* | U_i^{<t>})} = \frac{(U_i^* - \zeta_{k_i-1})(\zeta_{k_i} - U_i^*)}{(U_i^{<t>} - \zeta_{k_i-1})(\zeta_{k_i} - U_i^{<t>})}.$$

Regarding the proposal under switching the community, no information about the relation to  $U_i^{<t>}$  is given beforehand. Hence, in this case, we apply a uniform proposal restricted to the segments of all other groups. To be precise, for  $U_i^*$  we draw from a uniform distribution with the support  $[0, \zeta_{k_i-1}) \cup [\zeta_{k_i}, 1]$ . This means that the proposal density is given as  $p(U_i^* | U_i^{<t>}) = (1 - \nu)/(1 - (\zeta_{k_i} - \zeta_{k_i-1})) \cdot \mathbb{1}_{\{U_i^* \in [0, \zeta_{k_i-1}) \cup [\zeta_{k_i}, 1]\}}$ , yielding for  $U_i^* \in [\zeta_{k_i^*-1}, \zeta_{k_i^*})$

with  $k_i^* \neq k_i$  a proposal ratio of

$$\frac{p(U_i^{<t>} | U_i^*)}{p(U_i^* | U_i^{<t>})} = \frac{1 - (\zeta_{k_i} - \zeta_{k_i-1})}{1 - (\zeta_{k_i^*} - \zeta_{k_i^*-1})}.$$

Having defined the proceeding for proposing a new position for node  $i$ , including the calculations of the corresponding density ratios, we are now able to specify the acceptance probability. Hence, we accept the proposed value and therefore set  $U_i^{<t+1>} := U_i^*$  with a probability of

$$\min \left\{ 1, \prod_{j \neq i} \left[ \left( \frac{w_\zeta(U_i^*, U_j^{<t>})}{w_\zeta(U_i^{<t>}, U_j^{<t>})} \right)^{y_{ij}} \left( \frac{1 - w_\zeta(U_i^*, U_j^{<t>})}{1 - w_\zeta(U_i^{<t>}, U_j^{<t>})} \right)^{1-y_{ij}} \right] \frac{p(U_i^{<t>} | U_i^*)}{p(U_i^* | U_i^{<t>})} \right\}.$$

If we do not accept  $U_i^*$ , we set  $U_i^{<t+1>} := U_i^{<t>}$ . The consecutive drawing and updating of the components  $U_1, \dots, U_N$  then provides a proper Gibbs sampling sequence. After cutting the burn-in phase and appropriate thinning, calculating the sample mean of the simulated values consequently yields an approximation of the marginal conditional mean  $\mathbb{E}(U_i | \mathbf{y})$ . To be precise, for appropriately estimating  $U_i$  in the  $m$ -th iteration of the EM algorithm, we define

$$\hat{U}_i^{(m)} = \frac{1}{n} \sum_{s=1+b}^{n+b} U_i^{<s \cdot N \cdot r>}, \quad (17)$$

where  $b \in \mathbb{N}$  represents a burn-in parameter,  $r \in \mathbb{N}$  describes a thinning factor, and  $n$  is the number of MCMC states which are taken into account.

However, as discussed in Section 3.1 of the paper, these estimates need to be further adjusted in a two-fold manner, which also involves adjusting the community boundaries. Starting with Adjustment 1, we relocate the boundaries  $\zeta_k$  such that the group allocations correspond to the proportions of the realized groups, meaning we set

$$\hat{\zeta}_k^{(m+1)} = \frac{\sum_i \mathbb{1}_{\{\hat{U}_i^{(m)} < \hat{\zeta}_k^{(m)}\}}}{N}.$$

Note that this calculation represents an estimate of the transformation  $\varphi'_1(\hat{\zeta}_k^{(m)})$  with  $\varphi'_1(\cdot)$

as described in Section 3.1 of the paper. In fact, it is advisable to make small adjustments in early iterations since, in the beginning, the result of the E-step is rather rough. We therefore make use of step-size adjustments in the form of

$$\hat{\zeta}_k^{(m+1)} = \delta^{(m+1)} \frac{\sum_i \mathbb{1}_{\{\hat{U}_i^{(m)} < \hat{\zeta}_k^{(m)}\}}}{N} + (1 - \delta^{(m+1)}) \frac{k}{K}.$$

In this specification, the weighting  $\delta^{(m+1)} \in [0, 1]$  with  $\delta^{(m+1)} \geq \delta^{(m)}$  induces a step-size adaptation from a priori equidistant boundaries to boundaries implied by observed frequencies. Such step-size adaptation is recommendable to prevent the community size from shrinking too substantially before the structure of the community has properly evolved. In general,  $\delta^{(m+1)}$  is chosen to be one in the last iteration. This concludes Adjustment 1 with respect to the community boundaries.

We proceed with applying Adjustment 1 and Adjustment 2 to the posterior means derived from expression (17). To do so, we order all  $\hat{U}_i^{(m)}$  in the original blocks by ranks and rescale them to the new blocks defined through  $[\hat{\zeta}_{k-1}^{(m+1)}, \hat{\zeta}_k^{(m+1)})$ . That means, we first assign communities through  $\mathcal{C}_k^{(m)} = \{i \in \{1, \dots, N\} : \hat{\zeta}_{k-1}^{(m)} \leq \hat{U}_i^{(m)} < \hat{\zeta}_k^{(m)}\}$  with sizes  $N_k^{(m)} = |\mathcal{C}_k^{(m)}|$ . To enforce equidistant adjusted positions within the new community boundaries, we then calculate for all  $j \in \mathcal{C}_k^{(m)}$

$$\hat{U}_j^{(m+1)} = \frac{\text{rank}_k(\hat{U}_j^{(m)})}{N_k^{(m)} + 1} (\hat{\zeta}_k^{(m+1)} - \hat{\zeta}_{k-1}^{(m+1)}) + \hat{\zeta}_{k-1}^{(m+1)} \quad (18)$$

with  $\text{rank}_k(\hat{U}_j^{(m)})$  being the rank from smallest to largest of the element  $\hat{U}_j^{(m)}$  within all positions in community  $k$ , i.e. within the tuple  $(\hat{U}_i^{(m)} : i \in \mathcal{C}_k^{(m)})$ . These calculations, which represent an estimate of  $\varphi'_2 \circ \varphi'_1(\hat{U}_j^{(m)})$  with  $\varphi'_1(\cdot)$  and  $\varphi'_2(\cdot)$  as described in Section 3.1 of the paper, are applied to all communities  $k = 1, \dots, K$ . This concludes applying Adjustment 1 and Adjustment 2 to the latent quantities.

# Supplementary Material for ‘Stochastic Block Smooth Graphon Model’

Benjamin Sischka and Göran Kauermann

July 19, 2023

## 1 Notes on the EM-type Algorithm

### 1.1 Intuition and Justification

As a consequence of the identifiability issue (see Section 2.5 of the paper), there is not just one ground-truth SBSGM  $w_\zeta(\cdot, \cdot)$  that could be identified as the prescribed estimation target. Instead, assuming that the network  $\mathbf{Y} \in \{0, 1\}^{N \times N}$  is in fact generated from SBSGM  $w_\zeta(\cdot, \cdot)$ , we aim at reconstructing any SBSGM  $w'_{\zeta'}(\cdot, \cdot)$  from its equivalence class. In this context, such an equivalence class could be specified by referring to Diaconis and Janson (2007) and defining it as the set of all SBSGMs  $w'_{\zeta'}(\cdot, \cdot)$  for which two measure-preserving functions  $\varphi, \varphi' : [0, 1] \rightarrow [0, 1]$  exist such that  $w_\zeta(\varphi(u), \varphi(v)) = w'_{\zeta'}(\varphi'(u), \varphi'(v))$  for almost every  $(u, v)^\top \in [0, 1]^2$ . In the spirit of the cut distance from Lovász (2012, eq. 8.16), this means for  $w_\zeta(\cdot, \cdot)$  and  $w'_{\zeta'}(\cdot, \cdot)$

$$\inf_{\varphi, \varphi' \in \mathcal{M}} \iint |w_\zeta(\varphi(u), \varphi(v)) - w'_{\zeta'}(\varphi'(u), \varphi'(v))|^2 du dv = 0,$$

where  $\mathcal{M}$  is the set of all measure-preserving permutations on  $[0, 1]$ . However, as an alternative specification which omits the rearrangements  $\varphi(\cdot)$  and  $\varphi'(\cdot)$ , we define an equivalence class  $\mathcal{W}_{w_\zeta(\cdot, \cdot)}$  of  $w_\zeta(\cdot, \cdot)$  as the set of all symmetric measurable functions  $w'_{\zeta'} : [0, 1]^2 \rightarrow [0, 1]$  for which

$$\lim_{N \rightarrow \infty} \sum_{\mathbf{y} \in \{0, 1\}^{N \times N}} \mathbb{P}_{w_\zeta(\cdot, \cdot)}(\mathbf{Y} = \mathbf{y}) \log \frac{\mathbb{P}_{w_\zeta(\cdot, \cdot)}(\mathbf{Y} = \mathbf{y})}{\mathbb{P}_{w'_{\zeta'}(\cdot, \cdot)}(\mathbf{Y} = \mathbf{y})} = 0.$$

That means the Kullback-Leibler divergence between the probability functions induced by  $w_\zeta(\cdot, \cdot)$  and  $w'_{\zeta'}(\cdot, \cdot)$  on a network with diverging size should be zero. As is well-known,

---

Benjamin Sischka is Research Assistant, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany (E-mail: [benjamin.sischka@stat.uni-muenchen.de](mailto:benjamin.sischka@stat.uni-muenchen.de)). Göran Kauermann is Professor, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany (E-mail: [goeran.kauermann@stat.uni-muenchen.de](mailto:goeran.kauermann@stat.uni-muenchen.de)).

maximizing the likelihood is asymptotically equivalent to minimizing the Kullback-Leibler divergence, where the former one is what we pursue by our estimation procedure. More precisely, we exploit the concepts of the EM-type algorithm to derive the (penalized) maximum likelihood estimate for  $(\zeta, \gamma)$ . In this context, note that despite the non-identifiability,

- (a) an objective log-likelihood with respect to  $(\zeta, \gamma)$  and given  $\mathbf{y}$  can still be formulated properly in the form of  $\max_{\mathcal{U}} \ell(\zeta, \gamma; \mathbf{y}, \mathcal{U})$ , and
- (b) under mild conditions, both the E- and the M-step still lead to an increase of the log-likelihood, i.e.

$$\ell(\hat{\zeta}^{(m)}, \hat{\gamma}^{(m)}; \mathbf{y}, \hat{\mathcal{U}}^{(m-1)}) \underset{\text{E-step}}{\leq} \ell(\hat{\zeta}^{(m)}, \hat{\gamma}^{(m)}; \mathbf{y}, \hat{\mathcal{U}}^{(m)}) \underset{\text{M-step}}{\leq} \ell(\hat{\zeta}^{(m+1)}, \hat{\gamma}^{(m+1)}; \mathbf{y}, \hat{\mathcal{U}}^{(m)}) \quad (1)$$

if a (local) maximum has not already been reached.

So in combination with the implication on the Kullback-Leibler divergence, this maximization of the log-likelihood ensures that the algorithm approaches an SBSGM from the equivalence class  $\mathcal{W}_{w_{\zeta}(\cdot, \cdot)}$  for  $N \rightarrow \infty$ .

Following the above argumentation, the EM-type algorithm yields proper results if formulation (1) applies. At that, the second inequality is apparently always true since maximizing the log-likelihood is exactly what the M-step aims at. Therefore, the required conditions can be brought down to ensuring that the E-step yields proper predictions of the latent positions. According to the E-step's construction (see equation (17) of the paper's Appendix), this is based on a marginalization of the results from the Gibbs sampler conditional on  $\mathbf{y}$  and given an estimate of  $w_{\zeta}(\cdot, \cdot)$ . Thus, we require that at least for one  $w'_{\zeta'}(\cdot, \cdot) \in \mathcal{W}_{w_{\zeta}(\cdot, \cdot)}$ , the latent positions are identifiable with respect to marginalizing the conditional distribution. This is only violated if the SBSGM possesses a (partially) symmetric structure. Exemplary cases of such a scenario are given in Figure 1. Regarding the left model, for any observed network  $\mathbf{y}$ , it holds for any  $\mathbf{u} = (u_1, \dots, u_N) \in [0, 1]^{1 \times N}$  that  $f(\mathbf{u} | \mathbf{y}) = f(\mathbf{u}' | \mathbf{y})$  with  $\mathbf{u}' = (1 - u_1, \dots, 1 - u_N)$ . This especially implies that the marginalized conditional distribution for any node is symmetric around 0.5 and thus leads to  $\mathbb{E}(U_i | \mathbf{y}) = 0.5$  for all  $i = 1, \dots, N$ . Consequently, for this SBSGM, the latent positions are *not* marginally identifiable. The model on the right is an analogous example for the smooth case, where the latent positions which correspond to the segments delimited by white and black lines, respectively, are as a whole arbitrarily exchangeable (in reverse fashion, meaning  $0.1 \mapsto 0.85$ ,  $0.3 \mapsto 0.65$ , and mapping in-between values accordingly).

However, these kind of (partially) symmetric models can be viewed as very exceptional cases and thus should only rarely represent the ground truth in real-world networks. Nevertheless, we require the assumption that the true underlying model is not (partially) symmetric. If this assumption holds, a reasonable positioning by performing the E-step is guaranteed when it is based on the true model  $w'_{\zeta'}(\cdot, \cdot)$ . From there, it is easy to see that  $w'_{\zeta'}(\cdot, \cdot)$  is a stationary state of the described algorithm. Thus, its approximation in the form of  $w'^{spline}_{\zeta', \gamma'}(\cdot, \cdot)$  is (asymptotically) one of the global maxima—among the approximations

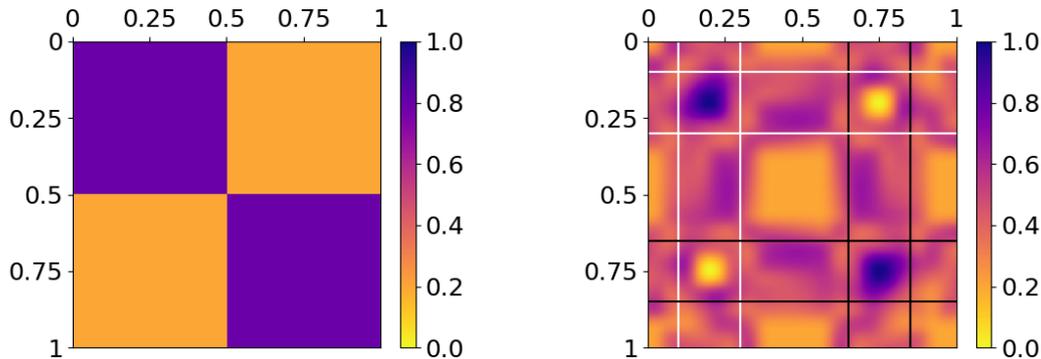


Figure 1: SBSGMs with non-identifiable latent positions under marginalization of the conditional distribution. Left: overall symmetric model in the SBM context. Right: partially symmetric model in the SGM setting.

of other elements from  $\mathcal{W}_{w_\zeta(\cdot, \cdot)}$ —of the log-likelihood with appropriate latent positions, i.e.  $\max_{\mathcal{U}} \ell(\zeta, \gamma; \mathbf{y}, \mathcal{U})$ .

## 1.2 Computational Intensiveness

Besides the fact that our proposed estimation routine can be demonstrated to provide accurate SBSGM fits, the algorithm appears to be computationally intensive. In regard to efficiency, it is well known that iterative procedures are inherently more expensive in computational terms than non-iterative techniques providing immediate estimation results. In the context of graphon estimation, the latter includes approaches such as histogram estimators based on marginal sorting strategies (Chan and Airolidi, 2014, Yang et al., 2014), spectral methods (Chatterjee, 2015, Xu, 2018), and neighboring smoothing (Zhang et al., 2017). Yet, we stress that these procedures are associated with a reduction in goodness of fit, accepted in favor of beneficial structural concepts. To be precise, degree-based ordering strategies require restrictive condition (11), spectral methods assume an (efficient) linear separability, and determining neighbors based on observed connectivity behaviors involves the total randomness inherent in the stochastic difference  $\sum_j (Y_{ij} - Y_{i'j})^2$ . (To underline the latter, Avella-Medina et al., 2020 have demonstrated that centrality can be measured more robustly by relying on derived graphon estimates than by simply extracting it from the observed network.) In particular, these drawbacks are detached from and exist despite potential consistency properties. Iterative estimation methods, on the other hand, are known to be able to capture present structures in more detail since they aim at aligning the estimates of multiple unknown quantities. This strategy consequently entails a finer-tuned model fit.

To make iterative estimation approaches more efficient, different strategies have been proposed in recent years. In the context of modeling networks based on latent variables, the most computationally expensive part is determining the latent nodal quantities. Aim-

ing for a more efficient estimation of mixed-membership SBMs, Mørup et al. (2011) and Li et al. (2016) adapted MCMC techniques by additionally incorporating a group-related split-merge scheme and a stochastic gradient approach, respectively. The resulting methods have lower computational complexity and therefore become more feasible for larger networks. A further approach for speeding up iterative estimation in the context of community detection is given by Narasimhamurthy et al. (2008), where they propose a two-stage clustering scheme. In the first stage, they achieve a rough partition by applying a computationally tractable, non-iterative technique. Based on that, they utilize more demanding approaches to get a more precise splitting in the second stage. Note, however, that the SBSGM involves a much more complex structure both within and between groups, which is why more efficient grouping strategies cannot be readily applied. As a different alternative, Karrer and Newman (2011) make use of greedy search techniques for estimating degree-corrected SBMs, see also Airolidi et al. (2013) for graphon estimation based on multiple graph observations. Such a strategy can be interpreted as a maximization-maximization approach with respect to the log-likelihood. In this scenario, the latent quantities are individually optimized based on the full-conditional likelihood. Yet, as stated by Fortunato (2010, p. 9), “despite the improvements and refinements of the last years, the accuracy of greedy optimization is not that good, as compared with other techniques.” A common concept to make specifically the EM algorithm more feasible is to rely on variational approaches; see e.g. Daudin et al. (2008) for its application in the SBM framework. However, this requires the existence of a suitable approximate distribution in the E-step. In the SBSGM context, this would need to appropriately cover the complex within- and between-group distribution.

Overall, adopting more computationally efficient methods for the SBSGM framework seems challenging, especially when the model’s accuracy is a key aspect. Nonetheless, as a straightforward improvement in terms of speeding up the estimation, one could employ an informative initialization that results in faster convergence. Strategies into this direction are discussed in Section 3.3.

## 2 Link to Other Models

The SBSGM possesses relationships to other network models which are beyond the regular SBM and the SGM. Here, we frame the connection to some selected models.

### 2.1 Degree-Corrected SBM

The SBSGM is not only able to cover the regular SBM but can further reflect its degree-corrected version that has been introduced by Karrer and Newman (2011). To do so, we define the following pattern within the SBSGM segments. Denoting by  $[\zeta_{k-1}, \zeta_k)$  with  $k = 1, \dots, K$  the group-specific intervals, then we specify for the SBSGM slices at  $u \in (\zeta_{k-1}, \zeta_k)$  (under particular consideration of symmetry in the form of  $w_\zeta(u, v) = w_\zeta(v, u)$ )

$$w_\zeta(u, \cdot) \equiv a_k(u) \cdot w_\zeta(\zeta_{k-1}, \cdot)$$

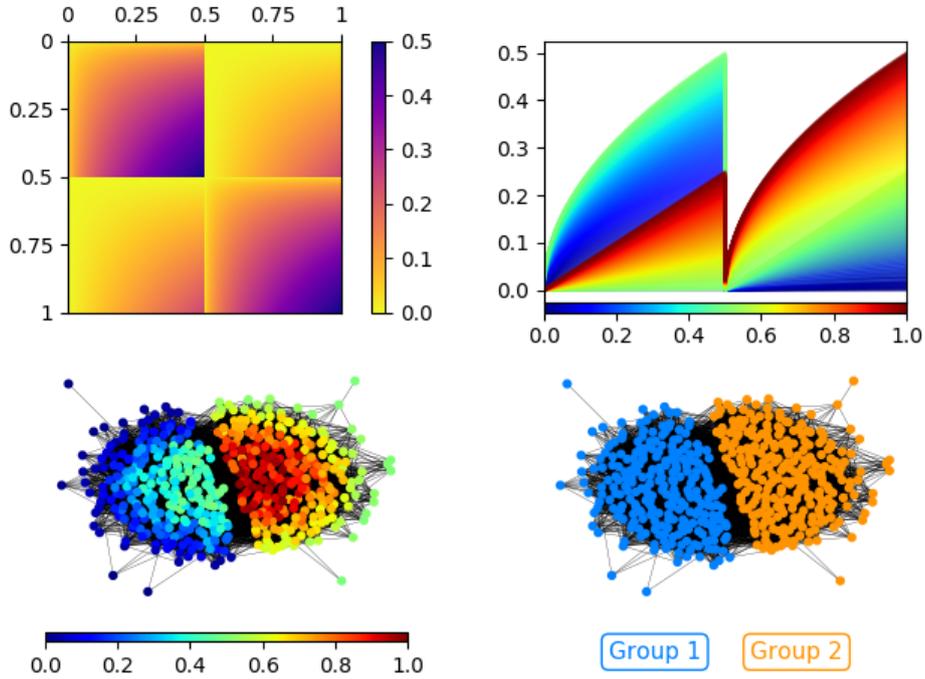


Figure 2: SBSGM representing a degree-corrected SBM with two communities. The SBSGM  $w_\zeta(\cdot, \cdot)$  is shown in the top left, while the corresponding slices  $w_\zeta(u, \cdot)$  are illustrated in the top right, exhibiting a more or less proportional profile within the groups. The bottom row shows twice the same simulated network of size  $N = 500$  but with different coloring; on the left, it refers to the  $U_i$ , and on the right, it illustrates the resulting community memberships (specified by  $U_i < 0.5$  and  $U_i \geq 0.5$ ).

with  $a_k : [0, 1] \rightarrow \mathbb{R}_+$  being a continuous monotone non-decreasing function which fulfills that  $a_k(\zeta_{k-1}) = 1$ . Apparently, in this setting, two nodes from the same community will reveal the same basic connectivity behavior (stochastically) but will differ in the degree proportion. For example, to model additionally the presence of hubs, one could choose an exponential profile for  $a_k(\cdot)$ , resulting in a node majority with lower degrees and a few nodes (hubs) having considerably higher degrees. Note that  $a_k(\cdot)$  is restricted through  $a_k(\zeta_k) \leq \min_v \{1/w_\zeta(\zeta_{k-1}, v)\}$ , meaning that the growth factor is limited by the inverse of the highest edge probability at the lower bound. This restriction of the degree heterogeneity within a community does not exist in the Poisson-type SBM but is a consequence of the Bernoulli-type SBM. However, one can circumvent this type-specific issue by allowing the slices to not only be strict multiples of each other but also to differ slightly in the profile, meaning especially to have a higher increase at low values than at values that are already high. An example of such a model is given in Figure 2. Considering the SBSGM slices  $w_\zeta(u, \cdot)$  at the top right, they have similar but shifted profiles. Consequently, in the two representations of a simulated network at the bottom row, the nodes from the same community exhibit fundamentally similar connectivities but differ in attractiveness.

## 2.2 Latent Distance Model

A further model class which has relation to the SBSGM is the latent distance model (LDM) introduced by Hoff et al. (2002). In its simple form it can be formulated as

$$\mathbb{P}(Y_{ij} = 1 | \mathbf{X}_i = \mathbf{x}_i, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta}) = \frac{\exp(\beta_0 - \beta_1 \|\mathbf{x}_i - \mathbf{x}_j\|)}{1 + \exp(\beta_0 - \beta_1 \|\mathbf{x}_i - \mathbf{x}_j\|)},$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top \in \mathbb{R}^2$  is a coefficient vector and  $\mathbf{X}_i$  is the latent position of node  $i$ —potentially located in the metric space  $(\mathbb{R}^J, \|\cdot\|)$  with  $J \in \mathbb{N}$ . The coefficient  $\beta_1$  is only required (and identifiable) if the latent space is bounded. We here consider the standardized version, meaning that we assume  $\mathbf{X}_i$  to be lying within  $[0, 1]^J$ . An approximate representation in the form of an SBSGM as formulated in (8) of the paper can then be achieved by decomposing the latent space into  $[0, 1]^{J-1} \times [0, 1]$  and subsequently partitioning  $[0, 1]^{J-1}$  into an increasing number of segments with volumes converging to zero. In doing so, the number of segments determines the number of “communities”, i.e.  $K$ , and the changes resulting through the distances in direction of the  $J$ -th dimension are captured within the intervals  $(\zeta_{k-1}, \zeta_k)$ . Consequently, starting with the one-dimensional version with prior distribution  $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$ , this can be exactly represented by the SGM, choosing

$$w_{\boldsymbol{\beta}}(u, v) = \frac{\exp(\beta_0 - \beta_1 |u - v|)}{1 + \exp(\beta_0 - \beta_1 |u - v|)}. \quad (2)$$

Similar connections are formulated by Matias and Robin (2014, Sec. 2.2), as well as Chan and Airoldi (2014, Sec. 5.1). In contrast, the two-dimensional LDM with uniform prior distribution on  $[0, 1]^2$  can be approximated arbitrarily well by the SBSGM through

$$w_{K, \boldsymbol{\beta}}(u, v) = \sum_{k=1}^K \sum_{l=1}^K \mathbb{1}_{\{u \in [\frac{k-1}{K}, \frac{k}{K}]\}} \mathbb{1}_{\{v \in [\frac{l-1}{K}, \frac{l}{K}]\}} w_{kl; K, \boldsymbol{\beta}}(u, v), \quad (3)$$

where  $K^2 \in \mathbb{N}$  is the number of segments into which the latent space  $[0, 1]^2$  is partitioned and

$$w_{kl; K, \boldsymbol{\beta}}(u, v) = \frac{\exp\left(\beta_0 - \beta_1 \sqrt{\left(\frac{|k-l|}{K}\right)^2 + |(Ku - k) - (Kv - l)|^2}\right)}{1 + \exp\left(\beta_0 - \beta_1 \sqrt{\left(\frac{|k-l|}{K}\right)^2 + |(Ku - k) - (Kv - l)|^2}\right)}$$

for  $k, l = 1, \dots, K$ . For  $K \rightarrow \infty$ , this SBSGM specification converges towards the exact representation of the two-dimensional distance model. Specifically, the odds ratio of an

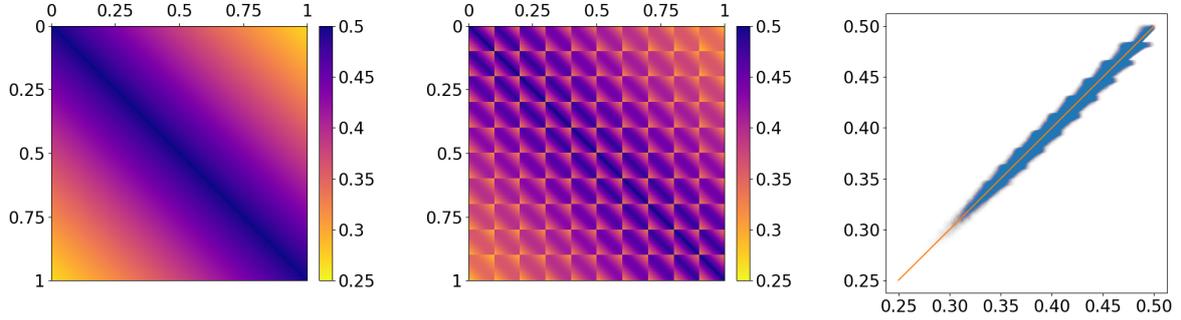


Figure 3: SBSGM representation of the LDM with uniform prior distribution on the unit space. Left: exact representation from (2) of the one-dimensional version with  $(\beta_0, \beta_1) = (0, 1)$ ; middle: SBSGM-type approximation from (3) of the two-dimensional version with  $(\beta_0, \beta_1) = (0, \sqrt{2}^{-1})$  and partition parameter  $K = 10$ . The coefficient  $\beta_1$  has been adapted such that the edge probability for the widest possible distance is equivalent for 1D and 2D. Right: edge probabilities for node pairs with simulated positions for the 2D distance model (horizontal axis) transformed into the SBSGM (vertical axis).

edge with respect to the LDM and its SGBSM-type approximation can be bounded by

$$\frac{\frac{\mathbb{P}_{\text{LDM}}(Y_{ij} = 1 | \mathbf{X}_i = \mathbf{x}_i, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta})}{1 - \mathbb{P}_{\text{LDM}}(Y_{ij} = 1 | \mathbf{X}_i = \mathbf{x}_i, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta})}}{\frac{\mathbb{P}_{\text{SBSGM}}(Y_{ij} = 1 | U_i = \psi(\mathbf{x}_i), U_j = \psi(\mathbf{x}_j); \boldsymbol{\beta})}{1 - \mathbb{P}_{\text{SBSGM}}(Y_{ij} = 1 | U_i = \psi(\mathbf{x}_i), U_j = \psi(\mathbf{x}_j); \boldsymbol{\beta})}} \in \left[ \exp\left(\pm \frac{\beta_1}{K}\right) \right] \xrightarrow{K \rightarrow \infty} [1 \pm 0],$$

where the mapping  $\psi : [0, 1]^2 \rightarrow [0, 1]$  for  $\mathbf{x}_i = (x_{i1}, x_{i2}) \in [0, 1]^2$  is defined through

$$\psi(\mathbf{x}_i) = \frac{x_{i1}}{K} + \sum_{k=1}^K \mathbb{1}_{\{x_{i2} \in [\frac{k-1}{K}, \frac{k}{K}]\}}. \quad (4)$$

However, the SBSGM from (3) does not converge towards a continuous function, which is why for  $K = \infty$  this does not yield a well-defined model. The SBSGM representation of the one- and two-dimensional LDM with uniform prior distribution in the unit space are illustrated in Figure 3. In addition, with regard to the SBSGM-type approximation of the two-dimensional LDM, the edge probabilities from the exact model and the approximation are plotted against each other for a hundred thousand simulated node pairs (right panel). This illustration demonstrates that the approximation is already close for a partition parameter of  $K = 10$ . In a similar way, also LDMs of higher dimension can be approximated.

Note that under the more general circumstances of  $\mathbf{X}_i$  being not uniformly distributed within  $[0, 1]^J$ , it follows that also the regular mapping  $\psi(\mathbf{X}_i)$  (like in (4) for the two-dimensional case) is not uniformly distributed. However, there exists a monotonically

increasing function  $\varphi : [0, 1] \rightarrow [0, 1]$  such that  $\varphi \circ \psi(\mathbf{X}_i) \sim \text{Uniform}(0, 1)$ . Referring to the regular SBSGM-type approximation  $w_{K,\beta}(\cdot, \cdot)$  (like in (3) for the two-dimensional case), the SBSGM then instead could be specified as  $w'_{K,\beta}(u, v) = w_{K,\beta}(\varphi^{-1}(u), \varphi^{-1}(v))$ , where  $\varphi^{-1}(u) := \max\{v : \varphi(v) = u\}$ .

## 2.3 Hierarchical Exponential Random Graph Model

A conceptual connection to the SBSGM is also given for the hierarchical exponential random graph model (HERGM) developed by Schweinberger and Handcock (2015). Starting with the standard exponential random graph model (ERGM), a link to the graphon model has been described by Chatterjee and Diaconis (2013) as well as Yin et al. (2016), at least for simple statistics like the number of two-stars and triangles and in the case of dense graphs. They make use of this relation to achieve asymptotic properties about the ERGM, which primarily concerns the normalizing constant and extremal behaviors. Krioukov (2016) has independently established a connection for sparse networks through expanding the domain of the graphon onto  $\mathbb{R}^2$ , where the focus also lies on the rather simple edge-triangle ERGM. The groundwork for the connection between these two models with differing conceptions was in particular laid by Lovász and Szegedy (2006) and Borgs et al. (2007). While the ERGM aims to capture structural patterns directly, the graphon model as limiting graph object has the intention to cover network structure through making use of additional latent quantities (see Matias and Robin, 2014 for an overview of latent variable models). Yet, recent works pursued to derive motif frequencies for graphon models (see e.g. Latouche and Robin, 2016) and therefore give further rise to the connection of these two models. For example, the overall probability of three random nodes forming a triangle

$$\iiint w(u_i, u_j) w(u_i, u_k) w(u_j, u_k) du_i du_j du_k$$

can be interpreted as expected global transitivity. Conversely, Bickel et al. (2011) established a sequence of patterns—called wheels—for which they can show that the graphon model is identifiable. So altogether, considering that there is a link between the ERGM and the graphon model, we draw the conclusion that there is also a connection between the HERGM and the SBSGM. Both approaches aim to divide the former global model into multiple local models to describe local dependence (see Schweinberger and Handcock, 2015). Therefore, both model extensions follow the same intuition.

## 3 Investigation on the Method’s Behavior and Performance

Beyond theoretical considerations about the capacity of our model and the effectiveness of the proposed estimation procedure, we want to evaluate the method’s concrete applicability. By visual checks, this has already been carried out in Section 4 of the paper. Sticking with these applications, we now want to get more detailed insights into the method’s behavior and how well it performs, also with respect to competing state-of-the-art approaches. To

do so, we investigate two quantitative aspects. On the one hand, we analyze its sensitivity with respect to initialization, providing information about the algorithm’s stability. On the other hand, an assessment of its capacity to uncover underlying structures is given. As competing methods, we consider the degree-corrected stochastic blockmodel (DCSBM, Karrer and Newman, 2011), the universal singular value thresholding strategy (USVT, Chatterjee, 2015), and the neighborhood smoothing approach (NBSM, Zhang et al., 2017). For these modeling techniques, we employ the software packages `graspologic` (DCSBM, Jaewon et al., 2019) and `randnet` (USVT and NBSM, Li et al., 2021).

### 3.1 Robustness with Respect to Initialization

The impact of the concrete starting values for  $\mathbf{U}$  can be investigated by considering the variability of the log-likelihood resulting from the final estimate. To this end, we repeatedly fit an SBSGM with different starting values, yielding an empirical distribution of  $\mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \hat{\mathbf{U}}; \hat{\boldsymbol{\zeta}}_K, \hat{\boldsymbol{\gamma}}_K)$ . The results are illustrated in Figure 4, where, for each number of groups, we repeated the estimation routine ten times.

Regarding the simulation examples illustrated in the left column, the variability of the log-likelihood seems relatively low compared to the variation among the competing methods. Moreover, for each of the networks, the center of the likelihood distribution is close to the log-likelihood under the true underlying model. Lastly, the outcomes distinguished by the number of groups reveal a clear separation in their tendencies.

The results for the real-world networks are depicted in the right column of Figure 4. The variability of the derived log-likelihood values here seems to be a bit higher but still within a very acceptable range. The sub-distributions reflecting the different group numbers are again clearly separated, demonstrating the explicit impact of  $K$ . Regarding the location of the likelihood distribution, it always lies between the log-likelihood results of the competing methods. This suggests that the SBSGM acts as a composite model that unites the different structural aspects covered by the other approaches. Referring to the illustrated variability under fixed group number, the precision of the model fit seems to depend only weakly on the model’s initialization. Note that in the military alliance network, the deviation between differing numbers of groups is relatively high, indicating that the group structure plays a predominant role.

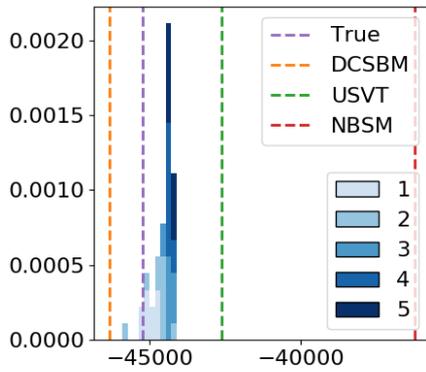
### 3.2 Performance in Prediction Accuracy

In a next step, we aim for investigating and comparing the capacity of covering underlying network structures. In the scenario of simulations, this can be done by relying on the differences between true and inferred edge probabilities. To be precise, we consider the root-mean-square error

$$\sqrt{\frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ j>i}} [h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) - \widehat{h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)}]^2}.$$

The results for the graphon models from Section 4.1 and the corresponding simulated network data are illustrated in the top plot of Figure 5. For the graphon with smoothly

Fig. 4



Political blogs

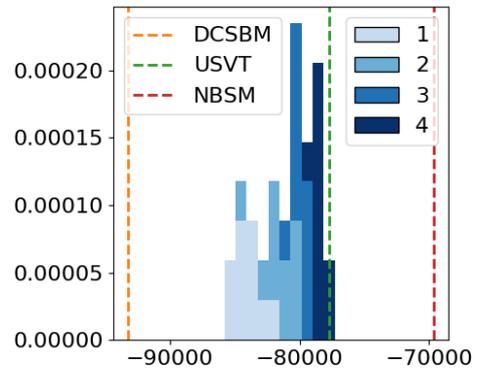
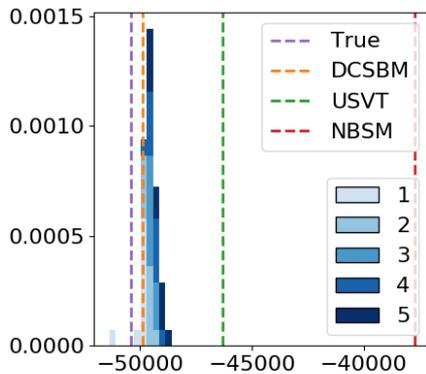


Fig. 5/a



Military alliances

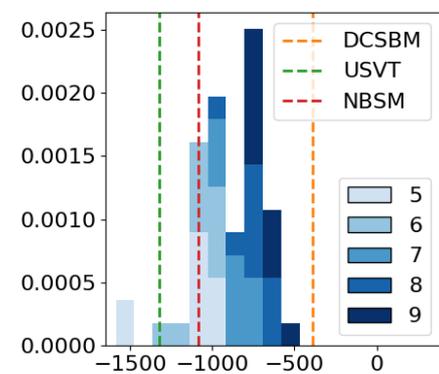
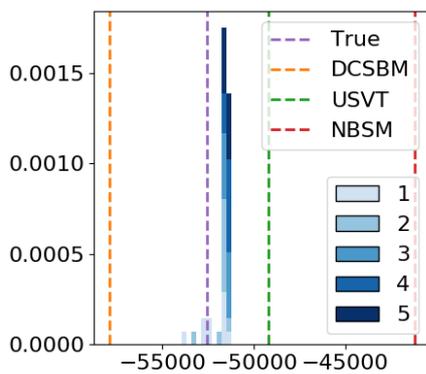


Fig. 5/b



Human brain functional coactivations

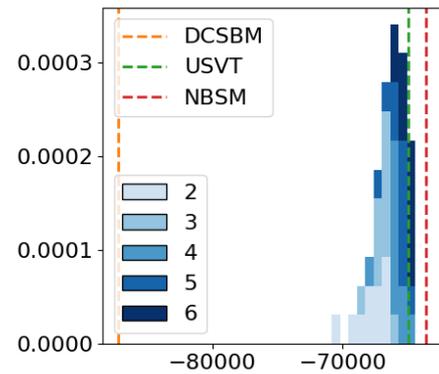


Figure 4: Distribution of the log-likelihood as it results under repeatedly fitting the SBSGM with different random initializations; illustrations refer to the networks from Section 4. The stacked histogram illustrates the overall outcome as well as its subdivision according to different numbers of groups. For comparison reasons, the log-likelihood is additionally depicted for the three competing methods DCSBM, USVT, and NBSM, as well as for the true underlying model (in the simulation scenarios).

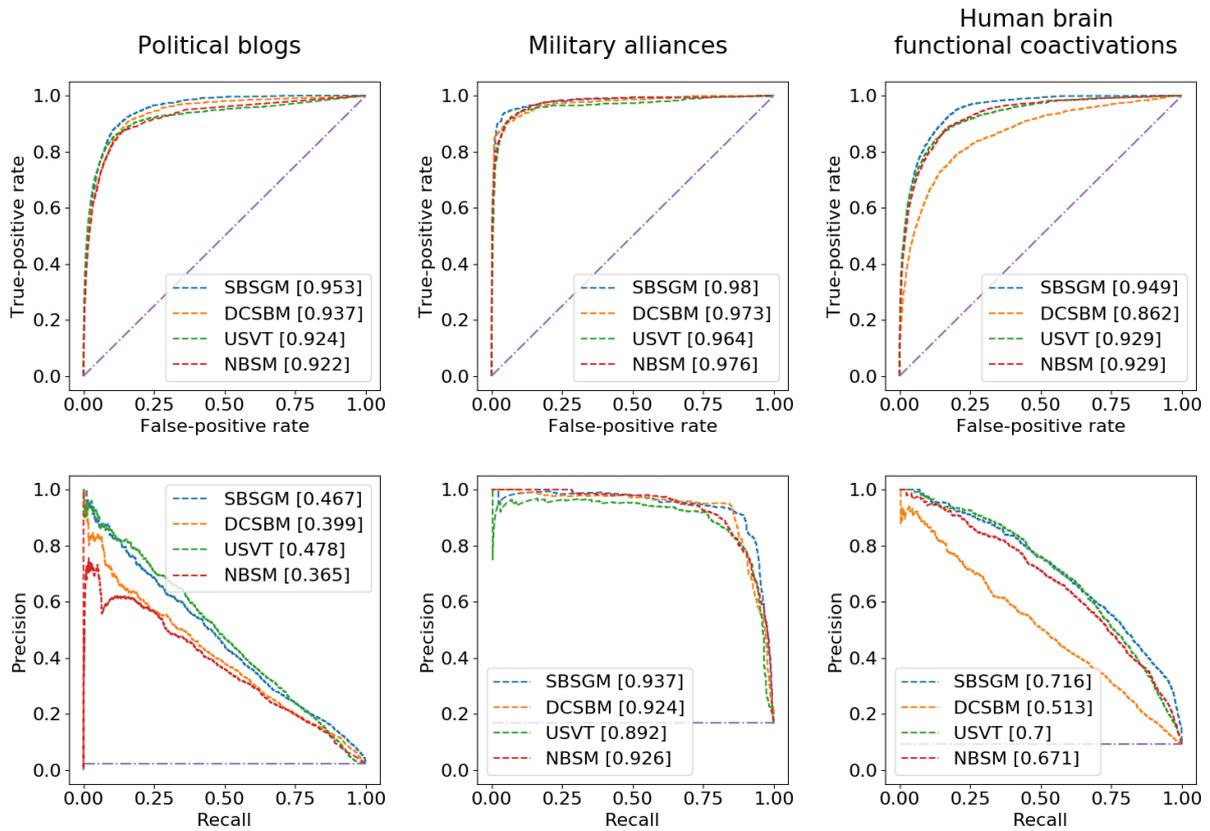
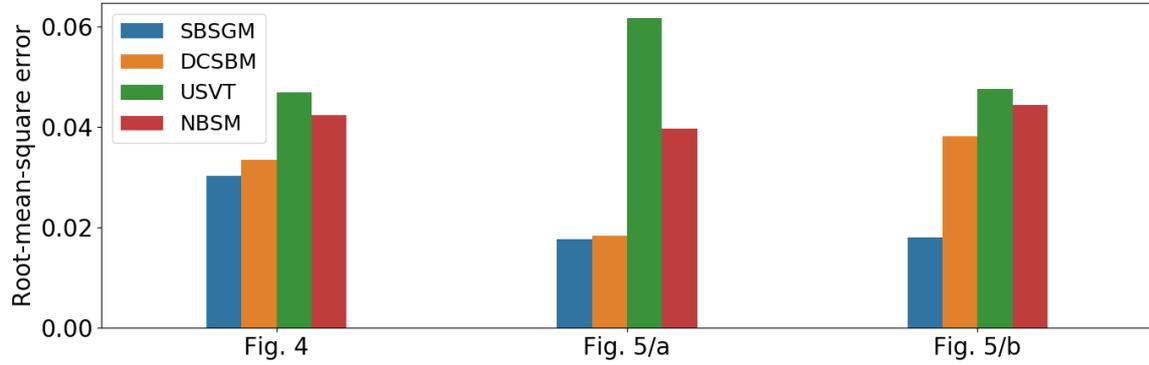


Figure 5: Top plot: root-mean-square error between true underlying and inferred edge probabilities based on indicated modeling frameworks. Results refer to the simulation scenarios from Section 4.1 of the paper with simulated networks of size  $N = 500$ . Bottom part: receiver operating characteristic curves (upper row) and precision-recall curves (lower row) for real-world networks from Section 4.2 (main paper). Predictions refer to network entries that have been randomly modified to imitate unstructured noise. Proportions of (potentially) modified entries are 10% for political blog and human brain network and 30% for military alliance network. Values given in the legend (in square brackets after model specifications) declare the area under the curve.

varying assortative structures from Section 4.1.1, illustrated in Figure 4 (main paper), our method is slightly better than the competing approaches, where, altogether, the differences between the methods’ performances are rather small. For the core-periphery structured graphon from Section 4.1.2 (main paper; left-hand side of Figure 5), the DCSBM and the SBSGM perform distinctly better than the methods USVT and NBSM. The high accuracy of the DCSBM can be explained by the fact that the data-generating model can be exactly represented as single-community SBM with degree heterogeneity. In contrast, the graphon from Section 4.1.3, which describes a mixture of assortative and disassortative structures—depicted at the right-hand side of Figure 5 of the paper—, can noticeably be best modeled by the SBSGM. This is as expected since the structural assumptions of all the competing methods are only poorly met.

To evaluate the performance on the real-world examples, we investigate the prediction accuracy. Apparently, to not give advantages to approaches that tend to overfit, the evaluation needs to be done on a “left-out” sample. Since the competing methods can however not (directly) deal with missing information about the presence of edges, this must be handled differently. For a comparative analysis, Zhang et al. (2017) thus consider the task of predicting missing edges. To do so, they randomly choose a set of network entries beforehand and set the values to zero by default. For this set of edge variables, the predicted edge probabilities and the actually observed values can then be compared. Yet, removing present edges only and not adding additional edges describes an unbalanced modification of the network. Li et al. (2020) therefore propose a different cross-validation strategy, where they substitute the randomly drawn network entries with the estimates obtained from a low-rank-based matrix completion algorithm. This, however, tends to give advantage to the USVT strategy, which can be interpreted as and is highly related to other low-rank-based matrix completion techniques. Hence, the USVT model fit will only be marginally affected by such network modifications. As a consequence, we propose a different approach which aims at a natural way of removing structure. In this respect, we consider the Erdős-Rényi-Gilbert model as the “structureless” representative among network-generating processes. We therefore propose to substitute the observations of the randomly chosen network entries by a sample from the Erdős-Rényi-Gilbert model with the same density as present in the observed network.

Based on this modification strategy, we investigate the prediction accuracy with respect to the (potentially) modified network entries. A proper quantification can then be constructed by relying on the Receiver Operating Characteristic (ROC) or the Precision Recall (PR) relation, depending on how balanced present versus absent edges are. The Area-Under-the-Curve (AUC) metric additionally provides a scalar reference value. The corresponding results for the real-world networks from Section 4.2 of the paper are illustrated in the bottom part of Figure 5. For all three examples, the ROC curve of the SBSGM exhibits a profile that is preferable over the other ones. That is also underpinned by the corresponding AUC, which is the highest for the SBSGM. Comparing the three competing methods among each other, the political blog and the military alliance network reveal better results under the DCSBM, indicating that they are inherently characterized by a predominant block structure. This is not the case for the human brain functional coactivation network. Similar results can be observed for the PR curve, according to which the

SBSGM yields the best predictions except for the political blog network. For the latter, we see that the USVT strategy provides slightly better results when the focus is on correctly recovering present edges. Altogether, the results obtained here coincide with the findings of Li and Le (2021), according to which a mixing of inferred block structures and graphon estimates leads to an improvement in prediction accuracy.

### 3.3 Running Time

Finally, note that with regard to running times, it takes the considered competitors only a few seconds to complete for networks with a few hundred nodes, whereas the current implementation of our SBSGM algorithm takes several minutes to a few hours (including finding the optimal  $K$ ). However, we stress that there are crucial differences in the objective. While other methods only seek to estimate the edge probabilities, our algorithm aims for recovering the underlying data-generating process, which additionally provides information on, for example, grouping structures. In particular, reconstructing the generating process often enables to easily extend the model in terms of covariates, representing weighted or dynamic networks, or incorporating other competencies. In contrast, this is often very difficult for heuristic approaches, where the key intention lies in fast prediction under straight structural assumptions.

Nonetheless, as a strategy to cut computational costs in the proposed approach, applying an “informative” initialization appears promisingly in order to achieve much faster convergence. Reasonable starting values for  $\mathbf{U}$  could exemplarily be derived through employing multidimensional scaling to the nodes’ connectivity, i.e.  $\mathbf{y}_{i\bullet} = (y_{i1}, \dots, y_{iN})$ , utilizing the reduction to one dimension. This follows the intuition of the SBSGM, according to which (per block) nearby nodes behave similarly. In this framework,  $\zeta$  can be initialized by determining the largest gaps within  $\hat{\mathbf{U}}^{(0)}$  or the highest differences between connectivity after ordering the nodes accordingly. Alternatively, the results of the DCSBM can be used as initialization, which seems particularly reasonable after the close connection formulated in Section 2.1. For this initialization strategy, the grouping of nodes can be directly adopted, whereas the ordering within communities can be derived from the nodes’ individual attractiveness.

These two informative initializations are additionally implemented in the provided Python package. In several simulations, we have found these informative starting configurations to yield very good results while cutting computational costs by more than half. Yet, if the focus is on finding the best possible estimate rather than getting results fast, as in the presented applications, we recommend the repetition strategy under different random initializations.

## 4 Human Brain Functional Coactivation Network – Model Fit with $K = 3$

As pointed out in Section 4.2.3 of the paper, we applied an SBSGM with four communities for fitting the human brain functional coactivation network even though the algorithm

inferred three, cf. Table 1 of the paper. This has been done for comparison purposes. For the sake of completeness, we here additionally illustrate the SBSGM estimate for the setting of  $K = 3$ , see Figure 6. The differences to the model fit from Figure 8 of the paper can be broken down quite clearly into two aspects, where we again refer to the categorization made in Crossley et al. (2013). On the one hand, the frontoparietal part is now detached from the occipital scope and instead merged with the lower section of the central module. On the other hand, here the upper section of the central module is integrated into the default-mode part. All other arrangement facets are more or less identical.

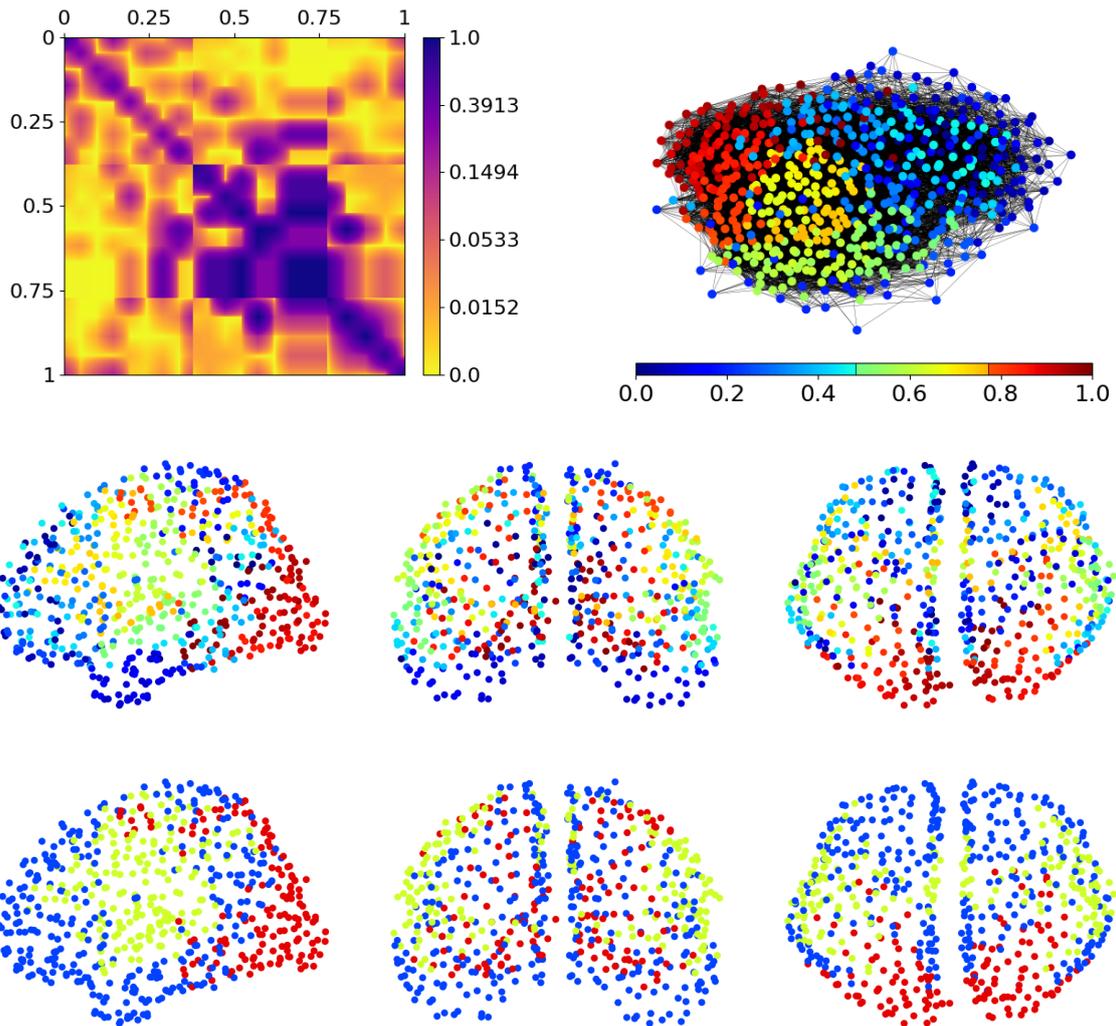


Figure 6: SBSGM estimation results for the human brain functional coactivation network under the setting of  $K = 3$ . Top left: estimate of  $w_\zeta(\cdot, \cdot)$  (in log scale). Top right: network with node coloring referring to  $\hat{U}_i \in [0, 1]$ . Bottom part: local positions of the human brain regions in anatomical space with coloring illustrating precise node positions (middle row) and derived community memberships (bottom row). Different angles illustrate side view (left column), front view (middle column), and top view (right column).

## References

- Airoldi, E. M., T. B. Costa, and S. H. Chan (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems 26*, pp. 692–700.
- Avella-Medina, M., F. Parise, M. T. Schaub, and S. Segarra (2020). Centrality Measures for Graphons: Accounting for Uncertainty in Networks. *IEEE Transactions on Network Science and Engineering*.
- Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *Annals of Statistics* 39(5), 2280–2301.
- Borgs, C., J. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi (2007). Counting graph homomorphisms. In *Topics in Discrete Mathematics*, pp. 315–371.
- Chan, S. H. and E. M. Airoldi (2014). A consistent histogram estimator for exchangeable graph models. In *31st International Conference on Machine Learning, ICML 2014*.
- Chatterjee, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *Annals of Statistics* 43(1), 177–214.
- Chatterjee, S. and P. Diaconis (2013). Estimating and understanding exponential random graph models. *Annals of Statistics* 41(5), 2428–2461.
- Crossley, N. A., A. Mechelli, P. E. Vértes, T. T. Winton-Brown, A. X. Patel, C. E. Ginestet, P. McGuire, and E. T. Bullmore (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences of the United States of America* 110(28), 11583–11588.
- Daudin, J. J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and Computing* 18(2), 173–183.
- Diaconis, P. and S. Janson (2007). Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*.
- Fortunato, S. (2010). Community detection in graphs.
- 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.
- Jaewon, C., D. Pedigo Benjamin, W. Bridgeford Eric, K. Varjavand Bijan, S. Helm Hayden, and T. Vogelstein Joshua (2019). GraSPy: Graph statistics in Python. *Journal of Machine Learning Research* 20(158), 1–7.
- Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 83(1).

- Krioukov, D. (2016). Clustering implies geometry in networks. *Physical Review Letters* 116(20).
- Latouche, P. and S. Robin (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing* 26(6), 1173–1185.
- Li, T. and C. M. Le (2021). Network estimation by mixing: Adaptivity and more. *arXiv preprint arXiv:2106.02803*.
- Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika* 107(2), 257–276.
- Li, T., E. Levina, J. Zhu, and C. M. Le (2021). randnet: Random network model estimation, selection and parameter tuning (R). Available online: <https://mran.microsoft.com/snapshot/2021-08-04/web/packages/randnet> [accessed 11-22-2022]. Version 0.4.
- Li, W., S. Ahn, and M. Welling (2016). Scalable MCMC for mixed membership stochastic blockmodels. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016*, pp. 723–731.
- Lovász, L. (2012). Large networks and graph limits. 2012, 487.
- Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory. Series B* 96(6), 933–957.
- Matias, C. and S. Robin (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. *ESAIM: Proceedings and Surveys* 47, 55–74.
- Mørup, M., M. N. Schmidt, and L. K. Hansen (2011). Infinite multiple membership relational modeling for complex networks. In *IEEE International Workshop on Machine Learning for Signal Processing*.
- Narasimhamurthy, A., D. Greene, N. Hurley, and P. Cunningham (2008). Community finding in large social networks through problem decomposition. In *Proc. 19th Irish Conference on Artificial Intelligence and Cognitive Science, AICS*.
- Schweinberger, M. and M. S. Handcock (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 77(3), 647–676.
- Xu, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *35th International Conference on Machine Learning, ICML 2018*, Volume 12, pp. 8646–8661.
- Yang, J. J., Q. Han, and E. M. Airoldi (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Journal of Machine Learning Research, Conference and Workshop Proceedings* 33, PMLR, pp. 1060–1067.

Yin, M., A. Rinaldo, and S. Fadnavis (2016). Asymptotic quantization of exponential random graphs. *Annals of Applied Probability* 26(6), 3251–3285.

Zhang, Y., E. Levina, and J. Zhu (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* 104(4), 771–783.

**Part IV.**

# **Graphon-Based Network Comparison**

## 8. Graphon-Based Network Comparison

### Contributing Article.

Sischka, B. and Kauermann, G. (2023). Nonparametric Two-Sample Test for Networks Using Joint Graphon Estimation. *arXiv preprint arXiv:2303.16014*. Planned submission (after adequate shortening): *Journal of the American Statistical Association (Applications and Case Studies)*.

### Data Sets.

All data sets used to demonstrate the applicability of our testing procedure are freely accessible. For information on concrete sources, see the specifications in the paper.

### Author Contributions.

The conceptual idea of utilizing the graphon model to compare the structures between networks came from Göran Kauermann (see Section 3). The joint graphon estimation routine (Section 4) was mainly formulated by Benjamin Sischka, who also entirely elaborated the subsequent testing procedure (Section 5). Moreover, Benjamin Sischka reviewed and put into context the previously developed concepts for network comparison (Section 2) and conducted the simulation and application studies (Section 6). The latter task specifically included the implementation of the whole strategy. An initial draft of the complete manuscript was prepared by Benjamin Sischka. Besides the initial idea and his general support, Göran Kauermann was involved in extensive proofreading. Lastly, the idea of extending the method towards detecting differences between networks at the microscopic level can be attributed to Benjamin Sischka, which also applies to the corresponding elaborations (see Supplementary Material).

# Nonparametric Two-Sample Test for Networks Using Joint Graphon Estimation

Benjamin Sischka\* and Göran Kauermann

Department of Statistics, Ludwig-Maximilians-Universität München

March 28, 2023

## Abstract

This paper focuses on the comparison of networks on the basis of statistical inference. For that purpose, we rely on smooth graphon models as a nonparametric modeling strategy that is able to capture complex structural patterns. The graphon itself can be viewed more broadly as density or intensity function on networks, making the model a natural choice for comparison purposes. Extending graphon estimation towards modeling multiple networks simultaneously consequently provides substantial information about the (dis-)similarity between networks. Fitting such a joint model—which can be accomplished by applying an EM-type algorithm—provides a joint graphon estimate plus a corresponding prediction of the node positions for each network. In particular, it entails a generalized network alignment, where nearby nodes play similar structural roles in their respective domains. Given that, we construct a chi-squared test on equivalence of network structures. Simulation studies and real-world examples support the applicability of our network comparison strategy.

*Keywords:* Network Comparison, EM Algorithm, Gibbs Sampler, B-Spline Regression, Chi-Squared Test

---

\*This work was partially funded by the BERD@NFDI Consortium (<https://www.berd-nfdi.de>).

# 1 Introduction

The field of statistical modeling and analysis of complex networks has gained strongly increasing interest over the last two decades. This is driven by the fact that different types of systems can be reasonably formalized as relationships between individuals or interactions between objects. Analyzing such structures consequently allows to uncover and describe the phenomena that affect these systems. Network-structured data arise in various fields, for example, social and political sciences, economics, biology, neurosciences, and many others. In this regard, a connectivity pattern between entities might describe friendships among members of a social group (Eagle et al., 2009), the trading between nations (Bhattacharya et al., 2008), interactions of proteins (Schwikowski et al., 2000), or the functional coactivation within the human brain (Bassett et al., 2018, Crossley et al., 2013).

In many situations, uncovering the underlying connectivity structure is not the only concern but also the comparison of akin networks and the exploration of potential differences. For example, this might be of interest in the context of brain coactivation. Recently, a lot of work has been going on investigating how the functional connectivity in the brain differs when people are affected by cognitive disorders like Alzheimer’s disease or autism spectrum disorder (Song et al., 2019, Subbaraju et al., 2017, Pascual-Belda et al., 2018). Two such functional coactivation networks—resulting from respectively averaging over the measurements of two different subject groups—are illustrated in Figure 1. The posed research question in this context apparently also involves the inquiry of whether a significant difference in the brain processes is observable at all, which additionally might depend on the environmental conditions like resting state, external stimuli, etc. More generally, this can be phrased as a hypothesis test on structural equivalence of two networks.

To this end, we pursue constructing a model-based approach for network comparison

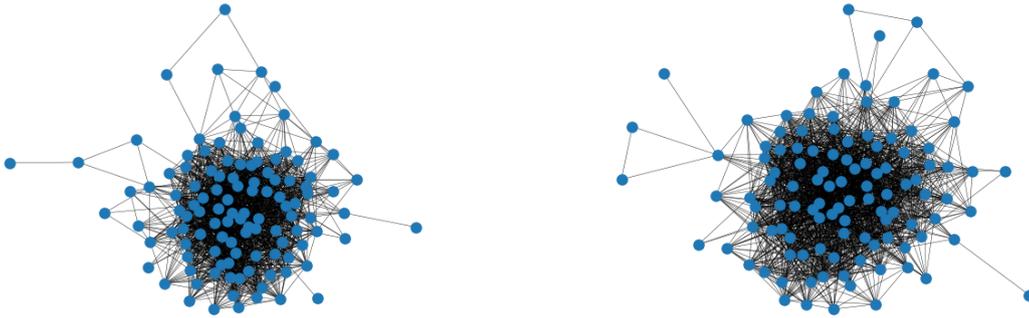


Figure 1: Functional coactivation networks of the human brain. The illustrated connectivity patterns result from averaging over multiple measurements for subjects with autism spectrum disorder (left) and typical development (right). Do these networks reveal a significant structural difference?

that allows for formal statistical testing. More precisely, we aim to test whether two networks can be considered independent samples drawn from the same probability distribution. This is apparently in itself a technically difficult question since the two networks can have different sizes, and hence the two distributions need to be somehow different. Therefore, it is crucial that the applied distributional framework constitutes a rather universal probability measure. In fact, this is not a trivial property, and many network models entail conceptual issues that impede a direct comparison. For example, in the Exponential Random Graph Model (Robins et al., 2007), a concrete model parameterization has different implications for different network sizes, making corresponding coefficient estimates hardly comparable. Hence, it is necessary to rely on distributional models where the specification of size and edge probabilities can be explicitly disentangled.

For such a comparative analysis of networks, we will demonstrate that Graphon Models (Lovász and Szegedy, 2006, Diaconis and Janson, 2008) are a very useful tool. First, the graphon model is very flexible and able to capture complex network structures. Secondly, the graphon itself can be interpreted as nonparametric density or intensity function on

networks. Both together make the graphon an overall characterizing network feature that, like nothing else, uniquely covers the structure in a comprehensive way. Hence, the graphon framework appears as a natural choice for comparison purposes. Lastly, the model’s design fulfills the above requirement of decoupling the network’s structure and size, allowing for modeling multiple networks simultaneously (Navarro and Segarra, 2022).

The rest of the paper is organized as follows. In Section 2, we start with reviewing methods from the network comparison literature. A formalization of the test problem we want to tackle in this work is then concretely specified and discussed in Section 3. The involved smooth graphon estimation in its capacity as the joint modeling approach is formulated in Section 4. Based on this joint graphon model, in Section 5, we develop a network comparison strategy for testing on equivalence of the underlying structures. The general applicability of the complete approach is demonstrated in Section 6, where we consider its performance on simulated and real-world data. This involves the method’s ability to uncover the underlying structure by joint graphon estimation, as well as the qualification of the subsequent testing procedure. The discussion and conclusion in Section 7 completes the paper.

## 2 Concepts for Network Comparison

When reviewing the literature on network comparison, it is worth noting that all proposed strategies are naturally based on a specific concept of capturing network structures. In general, the various approaches available for comparing networks can be broadly distinguished according to whether they rely on a descriptive or a model-based structural framework. Survey articles in this field are given by Soundarajan et al. (2014), Yaveroğlu et al. (2015), Emmert-Streib et al. (2016), and Tantardini et al. (2019). A more general perspective on

how complex data objects—such as adjacency matrices—might be compared is pointed out by Marron and Alonso (2014). Consulting this compendium clearly reveals a lack of model-based approaches in the context of network comparison. This is specifically deficient since drawing statistical inference is only possible under some kind of distributional assumption. The different strategies for network comparison proposed in the literature—irrespective of the capacity for drawing inference—are briefly reviewed hereafter.

Starting with approaches that are based on extracted network statistics, the most intuitive strategy is probably to simply compare global characteristics such as the clustering coefficient or the average path length (Newman, 2018, pp. 364 ff., Butts, 2008, p. 31). However, this captures the overall network structure only very poorly since completely differently structured networks can apparently still possess the same global statistics. As a more advanced approach, Wilson and Zhu (2008) consider the differences in the graph spectra, see also Gera et al. (2018). Yet, for the spectrum, it is often unclear which information in terms of local structural properties is extracted from the network. In addition, spectral methods can be strongly affected by small structural changes under specific circumstances. Taken together, such approaches often ascribe too little importance to the attributes of interest, leading to an over- or underrating of the structural dissimilarity at hand.

Another branch of the literature on descriptive network comparison relies on the concept of graphlets, i.e. prespecified subgraph patterns that are assumed to be sufficient for describing the present structure. Papers that, in one way or another, consider differences in the frequencies of graphlets are, among others, Pržulj et al. (2004), Pržulj (2007), Ali et al. (2014), and Faisal et al. (2017). Since the counting procedure is rather complex for larger graphlets, it is sort of a consensus to include only those that consist of no more than five nodes. However, this seems somehow arbitrary and incomplete in terms of capturing

all structural aspects. Moreover, Yaveroglu et al. (2014) found high correlations among the graphlet-related statistics, including complete redundancies. In contrast, model-based approaches specify the complete distribution of frequencies over all kinds of subgraphs. A connection between subgraph frequencies and the concrete specification of the graphon model is exemplarily elaborated in Borgs et al. (2008), Bickel et al. (2011), and Latouche and Robin (2016).

Overall, descriptive network statistics entail two general shortcomings. First, it is very difficult to assess in which parts of the networks the key differences are accommodated. To be precise, the nodes or edges (present or absent) that contribute most to a quantified structural discrepancy can only hardly be detected. Second, and more importantly, descriptive methods lack, by nature, the ability to draw inference in the probabilistic sense. Specifically, they provide no information on whether the found deviation between networks is plausible to be ascribed to randomness or whether there is a significant structural dissimilarity. Butts (2008) aims to overcome this deficit by applying a simplistic conditional uniform graph distribution.

On the other hand, probabilistic models for network data allow to induce distributions on network patterns that extend to desired distributional assumptions on structural differences. As a consequence, these modeling approaches might potentially serve as structural construct used for comparison purposes. Yet, they possess individual conceptual shortcomings that often impede a direct comparison. While, for example, the Latent Distance Model (Hoff et al., 2002) does not provide any model-related key component to be compared, coefficient estimates from the exponential random graph model are not directly comparable across separated networks. The graphon model and the Stochastic Blockmodel (Holland et al., 1983, Snijders and Nowicki, 1997 and 2001) suffer from identifiability issues (see e.g.

Diaconis and Janson, 2008, Thm. 7.1) that make a comparison of corresponding individual estimates complicated. The latter model’s adaptivity is additionally highly dependent on the choice of the number of blocks. Onnela et al. (2012) tackle this issue by observing the networks’ complete disintegration processes, which they subsequently summarize by the profiles of well-known network statistics. Integrating over the profiles’ differences and applying principle component analysis for summarization yields the final distance measure, which has been demonstrated to provide reasonable results in terms of leading to a good classification. However, to the best of our knowledge, there exists no (model-based) *nonparametric* test on the equivalence of network structures.

In this paper, we aim to address this shortcoming, which we tackle by striking new paths. As a general concept for this approach, we follow the intuition of fitting a joint model to multiple networks simultaneously. For that purpose, we resort to the smooth graphon model as an appropriate and very powerful framework. Such a joint modeling strategy consequently circumvents the need for post-hoc alignment of individual model fits and yields an outcome that provides substantial information for comparison purposes. More precisely, it allows for directly relating the networks at hand on the microscopic scale, which, in the literature, is often referred to as “network alignment” (Kuchaiev et al., 2010). However, as an essential distinction to classical network alignment strategies, this method does not seek to find a node-wise one-to-one mapping. Instead, it implies a mapping of local components, meaning small fuzzy groups of nodes with similar structural roles in their respective domains. Based on this network alignment, a structural comparison at the microscopic level becomes possible. Aggregating local differences finally enables to construct a test on structural equivalence of networks.

### 3 Notation and Formulation of the Test Problem

We consider the setting where two undirected networks of possibly different sizes  $N^{(1)}$  and  $N^{(2)}$  have been observed. Let  $\mathbf{y}^{(g)} = [y_{ij}^{(g)}]_{i,j=1,\dots,N^{(g)}}$  for  $g = 1, 2$  denote the two respective adjacency matrices, where, for  $i, j = 1, \dots, N^{(g)}$ ,  $y_{ij}^{(g)} = 1$  if in network  $g$  an edge between nodes  $i$  and  $j$  exists and  $y_{ij}^{(g)} = 0$  otherwise. That specifically means that  $\mathbf{y}^{(g)} \in \{0, 1\}^{N^{(g)} \times N^{(g)}}$ . We assume the networks to be undirected so that  $y_{ij}^{(g)} = y_{ji}^{(g)}$ . Additionally, the diagonal elements are set to zero, i.e.  $y_{ii}^{(g)} = 0$ , reflecting the absence of self-loops. In general, we consider  $\mathbf{y}^{(g)}$  to be a realization of a random network  $\mathbf{Y}^{(g)}$  of size  $N^{(g)}$  which is subject to probability mass  $\mathbb{P}(\mathbf{Y}^{(g)} = \mathbf{y}^{(g)}; N^{(g)})$ . The question we aim to tackle is whether  $\mathbf{y}^{(1)}$  and  $\mathbf{y}^{(2)}$  are drawn from the same distribution. To suitably specify such a distribution, we rely on the smooth graphon model. The data-generating process is thereby as follows. Assume that we independently draw uniformly distributed random variables

$$U_i^{(g)} \sim \text{Uniform}[0, 1] \quad \text{for } i = 1, \dots, N^{(g)} \text{ and } g = 1, 2. \quad (1)$$

Conditional on  $\mathbf{U}^{(g)} = (U_1^{(g)}, \dots, U_{N^{(g)}}^{(g)})$ , we then draw the edges i.i.d. through

$$Y_{ij}^{(g)} \mid (\mathbf{U}^{(g)} = \mathbf{u}^{(g)}) \sim \text{Binomial}(1, w^{(g)}(u_i^{(g)}, u_j^{(g)})) \quad \text{for } j > i \quad (2)$$

with  $\mathbf{u}^{(g)} = (u_1^{(g)}, \dots, u_{N^{(g)}}^{(g)}) \in [0, 1]^{N^{(g)}}$  and under the setting of  $Y_{ij}^{(g)} \equiv Y_{ji}^{(g)}$  for  $j < i$  and  $Y_{ii}^{(g)} \equiv 0$ . In this modeling framework, the function  $w^{(g)} : [0, 1]^2 \rightarrow [0, 1]$ , which specifies the structural behavior of the emerging network, is called graphon (see Lovász and Szegedy, 2006 and Diaconis and Janson, 2008). Here, in particular, we assume  $w^{(g)}(\cdot, \cdot)$  to be smooth according to some Hölder or Lipschitz condition (cf. Wolfe and Olhede, 2013 or Chan and Airoldi, 2014). Relying on this data-generating process, the graphon-based probability

model can be defined through

$$\mathbf{Y}^{(g)} \sim \mathbb{P}(\mathbf{Y}^{(g)} = \cdot; w^{(g)}(\cdot, \cdot), N^{(g)}). \quad (3)$$

In this distribution model, the network's size and structure are apparently dissociated, which therefore allows for a size-independent comparison of underlying structures. Hence, our goal is to develop a statistical test on the hypothesis

$$H_0 : w^{(1)}(\cdot, \cdot) \equiv w^{(2)}(\cdot, \cdot). \quad (4)$$

In this context, we emphasize that data-generating process (2) is not unique because it is invariant to permutations of  $w^{(g)}(\cdot, \cdot)$ , as discussed in detail by Diaconis and Janson (2008, Sec. 7). Thus, the formulation of  $H_0$  needs to be understood from the perspective of corresponding equivalence classes, implying that  $w^{(1)}(\cdot, \cdot)$  and  $w^{(2)}(\cdot, \cdot)$  are rather viewed from a theoretical perspective. Nonetheless, for the concrete implementation of the test procedure, we employ a concrete representation of the two graphons. Specifically, under the assumption of  $H_0$  being true, we call the coinciding manifestation the *joint graphon*. This can be formalized as

$$w^{\text{joint}}(u, v) := w^{(1)}(u, v) = w^{(2)}(u, v) \quad \text{for all } (u, v)^\top \in [0, 1]^2.$$

Since  $w^{(1)}(\cdot, \cdot)$  and  $w^{(2)}(\cdot, \cdot)$  are assumed to be smooth—at least for one possible arrangement, and, in particular, the one we consider here—, this also holds for  $w^{\text{joint}}(\cdot, \cdot)$ . Given such a concrete representation of the joint graphon, the node position vectors  $\mathbf{u}^{(1)}$  and  $\mathbf{u}^{(2)}$  referring to  $w^{\text{joint}}(\cdot, \cdot)$  then provide a specific type of network alignment. This is what we

utilize for a direct comparison of  $\mathbf{y}^{(1)}$  and  $\mathbf{y}^{(2)}$ . However, one typically observes neither  $\mathbf{u}^{(1)}$  and  $\mathbf{u}^{(2)}$  nor  $w^{\text{joint}}(\cdot, \cdot)$ . Thus, in order to achieve this alignment, we first need to formulate an appropriate estimation procedure for the joint graphon model.

## 4 EM-Based Joint Graphon Estimation

In this section, we present an iterative estimation procedure for the joint smooth graphon model under the assumption that null hypothesis (4) is true. To do so, we follow the EM-based estimation routine of Sischka and Kauermann (2022a), extending it to the situation of two networks.

### 4.1 MCMC E-Step

Starting with the E-step of our iterative algorithm, we assume the joint graphon  $w^{\text{joint}}(\cdot, \cdot)$  to be known for the moment. Based on that, the latent positions of the networks can be separately determined using MCMC techniques. To be precise, we apply Gibbs sampling by formulating the full conditional distribution of  $U_i^{(g)}$  through

$$f(u_i^{(g)} \mid u_1^{(g)}, \dots, u_{i-1}^{(g)}, u_{i+1}^{(g)}, \dots, u_{N^{(g)}}^{(g)}, \mathbf{y}^{(g)}) \propto \prod_{j \neq i} w^{\text{joint}}(u_i^{(g)}, u_j^{(g)})^{y_{ij}^{(g)}} [1 - w^{\text{joint}}(u_i^{(g)}, u_j^{(g)})]^{1-y_{ij}^{(g)}} \quad (5)$$

for all  $i = 1, \dots, N^{(g)}$  and  $g = 1, 2$ . Details on the concrete implementation of the Gibbs sampler are given in Section A of the Appendix. The resulting MCMC sequence (after cutting the burn-in period and appropriate thinning) then reflects the joint conditional distribution  $f(\mathbf{u}^{(g)} \mid \mathbf{y}^{(g)})$ . Thus, the marginal conditional means of the latent positions, i.e.  $\mathbb{E}(U_i^{(g)} \mid \mathbf{Y}^{(g)} = \mathbf{y}^{(g)})$  for  $i = 1, \dots, N^{(g)}$ , can be approximated by taking the mean

over the MCMC samples, which we denote by  $\bar{\mathbf{u}}^{(g)} = (\bar{u}_1^{(g)}, \dots, \bar{u}_{N^{(g)}}^{(g)})$ . This posterior mean vector, however, requires further adjustments due to additional identifiability issues which cannot be coped with the standard EM-type algorithm. To illustrate this, let model assumption (1) be more relaxed in the sense that the  $U_i^{(g)}$ 's might follow any continuous distribution  $F^{(g)}(\cdot)$ . Under this configuration, the model  $(F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot))$  is equivalent to any other model  $(F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot))$  constructed through

$$F^{(g)' }(u') := F^{(g)}(\varphi(u')) \quad \text{and} \quad w^{(g)' }(u', v') := w^{(g)}(\varphi(u'), \varphi(v'))$$

with  $\varphi : [0, 1] \rightarrow [0, 1]$  being a strictly increasing continuous function (that is, in contrast to Diaconis and Janson, 2008, Sec. 7, not measure-preserving). Specifically, that means

$$\mathbb{P}(\mathbf{Y}^{(g)} = \cdot; F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot), N^{(g)}) \equiv \mathbb{P}(\mathbf{Y}^{(g)} = \cdot; F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot), N^{(g)})$$

for all  $N^{(g)} \geq 2$ . As a matter of conception, this issue cannot be solved by the EM algorithm since it aims at specifying a model that adapts optimally to the given data instead of perfectly recovering the underlying model structure. Consequently, the EM approach is not able to distinguish between the two conceptually equivalent model specifications  $(F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot))$  and  $(F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot))$ . Nonetheless, this identifiability issue can simply be tackled by adjusting  $\bar{\mathbf{u}}^{(g)}$  before estimating the graphon in the M-step. To do so, we just impose that the inferred node positions follow an ideal sample drawn from the standard uniform distribution. That is, we set

$$\hat{u}_i^{(g)} = \frac{\text{rank}(\bar{u}_i^{(g)})}{N^{(g)} + 1},$$

where  $\text{rank}(\bar{u}_i^{(g)})$  is the rank from smallest to largest of element  $\bar{u}_i^{(g)}$  within  $\bar{\mathbf{u}}^{(g)}$ . In this context, note that the values  $i/(N^{(g)} + 1)$  with  $i = 1, \dots, N^{(g)}$  represent the expectations of  $N^{(g)}$  ordered random variables that are independently drawn from the standard uniform distribution. As a result, with  $\hat{\mathbf{u}}^{(g)} = (\hat{u}_1^{(g)}, \dots, \hat{u}_{N^{(g)}}^{(g)})$  we obtain a plausible realization of the node positions of network  $g$ . Apparently, this relies on the current joint graphon estimate  $\hat{w}^{\text{joint}}(\cdot, \cdot)$ , which is applied as substitute in conditional distribution (5). In the next step, we formulate the procedure for updating  $\hat{w}^{\text{joint}}(\cdot, \cdot)$  given  $\hat{\mathbf{u}}^{(1)}$  and  $\hat{\mathbf{u}}^{(2)}$ .

## 4.2 Spline-Based M-Step

For a semiparametric estimation of the joint smooth graphon, we choose a linear B-spline regression approach. To this end, we assume the joint graphon to be approximated through

$$w_{\boldsymbol{\theta}}^{\text{joint}}(u, v) = \mathbf{B}(u, v) \boldsymbol{\theta} = [\mathbf{B}(u) \otimes \mathbf{B}(v)] \boldsymbol{\theta},$$

where  $\otimes$  is the Kronecker product,  $\mathbf{B}(u) \in \mathbb{R}^{1 \times L}$  is a linear B-spline basis on  $[0, 1]$ , normalized to have a maximum value of one, and  $\boldsymbol{\theta} \in \mathbb{R}^{L^2}$  is the parameter vector to be estimated. The inner B-spline knots are specified as lying equidistantly on a regular 2D grid within  $[0, 1]^2$ , where  $\boldsymbol{\theta}$  is indexed accordingly through  $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1L}, \theta_{21}, \dots, \theta_{LL})^\top$ . Based on this representation and given the node positions  $\hat{\mathbf{u}}^{(1)}$  and  $\hat{\mathbf{u}}^{(2)}$ , we formulate the marginal log-likelihood over both networks as

$$\ell(\boldsymbol{\theta}) = \sum_g \sum_{\substack{i,j \\ j \neq i}} \left[ y_{ij}^{(g)} \log \left( \mathbf{B}_{ij}^{(g)} \boldsymbol{\theta} \right) + \left( 1 - y_{ij}^{(g)} \right) \log \left( 1 - \mathbf{B}_{ij}^{(g)} \boldsymbol{\theta} \right) \right], \quad (6)$$

where  $\mathbf{B}_{ij}^{(g)} = \mathbf{B}(\hat{u}_i^{(g)}) \otimes \mathbf{B}(\hat{u}_j^{(g)})$ . Furthermore, through standard calculations, we are able to derive the score function  $\mathbf{s}(\boldsymbol{\theta})$  and the Fisher information  $\mathbf{F}(\boldsymbol{\theta})$ , as demonstrated in

Section B of the Appendix. Fisher scoring can then be used to maximize  $\ell(\boldsymbol{\theta})$ . In addition, we include side constraints to ensure that  $w_{\boldsymbol{\theta}}^{\text{joint}}(\cdot, \cdot)$  is bounded to  $[0, 1]$  and symmetric. In the linear B-spline setting, this means restricting the parameters by the conditions

$$\theta_{kl} \geq 0, \quad \theta_{kl} \leq 1, \quad \text{and} \quad \theta_{kl} - \theta_{lk} = 0$$

for all  $l > k$ . Apparently, all three conditions are of linear form and thus can be written in matrix format. Taken together, the Fisher scoring becomes a quadratic programming problem that can be solved using standard software (see Andersen et al., 2016 or Turlach and Weingessel, 2013).

Moreover, we intend to add penalization on the B-spline estimate. As outlined in Eilers and Marx (1996) and Ruppert et al. (2003, 2009), penalized spline estimation under the setting of a rather large spline basis yields a preferable outcome since it guarantees a functional fit that covers the data adequately but is still smooth. Thus, this approach enables to precisely capture the underlying structure while avoiding overfitting. To realize this, we add a first-order penalty, meaning that “neighboring” elements of  $\boldsymbol{\theta}$  get penalized. For the log-likelihood, the score function, and the Fisher information, this leads to the penalized versions in the form of

$$\begin{aligned} \ell_p(\boldsymbol{\theta}, \lambda) &= \ell(\boldsymbol{\theta}) - \frac{1}{2} \lambda \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta}, \quad \mathbf{s}_p(\boldsymbol{\theta}, \lambda) = \mathbf{s}(\boldsymbol{\theta}) - \lambda \mathbf{P} \boldsymbol{\theta}, \\ \text{and} \quad \mathbf{F}_p(\boldsymbol{\theta}, \lambda) &= \mathbf{F}(\boldsymbol{\theta}) + \lambda \mathbf{P}, \end{aligned} \tag{7}$$

respectively, where  $\mathbf{P}$  is a penalization matrix of appropriate shape (see Section B of the Appendix). For an adequate choice of the penalty parameter  $\lambda$  in the two-dimensional spline regression, we follow Kauermann et al. (2013) and apply the corrected Akaike Information Criterion ( $\text{AIC}_c$ , see Hurvich and Tsai, 1989 and Burnham and Anderson, 2002). This is

defined as

$$\text{AIC}_c(\lambda) = -2 \ell(\hat{\boldsymbol{\theta}}_p) + 2 \text{df}(\lambda) + \frac{2 \text{df}(\lambda)[\text{df}(\lambda) + 1]}{N(N - 1) - \text{df}(\lambda) - 1},$$

where  $\hat{\boldsymbol{\theta}}_p$  is the corresponding penalized parameter estimate and  $\text{df}(\lambda)$  specifies the degrees of freedom of the penalized B-spline function. More precisely, according to Wood (2017, pp. 211 ff.), the latter is defined through

$$\text{df}(\lambda) = \text{tr} \left\{ \mathbf{F}_p^{-1}(\hat{\boldsymbol{\theta}}_p, \lambda) \mathbf{F}(\hat{\boldsymbol{\theta}}_p) \right\}$$

with  $\text{tr}\{\cdot\}$  being the trace of a matrix. A numerical optimization of the corrected AIC with respect to  $\lambda$  concludes the estimation of  $\boldsymbol{\theta}$ , resulting in the eventual estimate  $\hat{w}^{\text{joint}}(\cdot, \cdot)$  of the current M-step.

Finally, the EM-type estimation procedure described above—meaning the consecutive repetition of the E- and M-step until convergence is achieved—allows us to adequately estimate both the joint graphon  $w^{\text{joint}}(\cdot, \cdot)$  and the corresponding node positions  $\mathbf{u}^{(1)}$  and  $\mathbf{u}^{(2)}$  of the two networks. Based on these results, we are now able to formulate an appropriate test procedure.

## 5 Two-Sample Test on Network Structures

Returning to the test problem raised in Section 3, we now develop a statistical test procedure on hypothesis (4), i.e. whether  $\mathbf{y}^{(1)}$  and  $\mathbf{y}^{(2)}$  are drawn from the same distribution. To do so, we utilize the network alignment resulting from the (inferred) joint smooth graphon model. More precisely, we exploit the fact that two edge variables  $Y_{i_1 j_1}^{(1)}$  and  $Y_{i_2 j_2}^{(2)}$  that have nearby positions—i.e. for which the distance between  $(u_{i_1}^{(1)}, u_{j_1}^{(1)})^\top$  and  $(u_{i_2}^{(2)}, u_{j_2}^{(2)})^\top$  is small—possess similar probabilities to form a connection. In a more formalized way, this

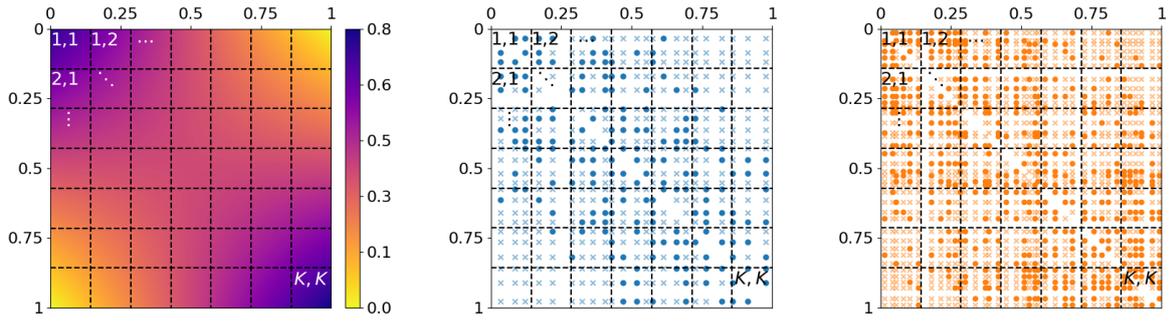


Figure 2: Dividing the unit square as domain of the graphon model into small segments for comparing network structure on a microscopic level. Left: division of  $w^{\text{joint}}(\cdot, \cdot)$  into approximately piecewise-constant rectangles. Middle and right: edge positions  $(u_i^{(g)}, u_j^{(g)})^\top$  of two simulated networks with respect to  $w^{\text{joint}}(\cdot, \cdot)$ ; weakly colored crosses and intensively colored circles represent absent and present edges, respectively. The two networks can be compared by pairwise contrasting the edge proportions within the labeled rectangles.

means that, from  $\|(u_{i_1}^{(1)}, u_{j_1}^{(1)})^\top - (u_{i_2}^{(2)}, u_{j_2}^{(2)})^\top\| \approx 0$ , it follows that

$$\begin{aligned} \mathbb{P}(Y_{i_1 j_1}^{(1)} = 1 \mid U_{i_1}^{(1)} = u_{i_1}^{(1)}, U_{j_1}^{(1)} = u_{j_1}^{(1)}) \\ \approx \mathbb{P}(Y_{i_2 j_2}^{(2)} = 1 \mid U_{i_2}^{(2)} = u_{i_2}^{(2)}, U_{j_2}^{(2)} = u_{j_2}^{(2)}), \end{aligned}$$

where  $\|\cdot\|$  is the Euclidean distance. Following this intuition, we divide the unit square into small segments and compare between networks the ratio of present versus absent edges occurring in these segments (see Figure 2 for an exemplary division). For that purpose, we choose a suitable  $K \in \mathbb{N}$ , specify a corresponding boundary sequence  $a_0 = 0 < a_1 < \dots < a_K = 1$ , and define the following two quantities for  $l, k = 1, \dots, K$  with  $l \geq k$ :

$$\begin{aligned} d_{kl}^{(g)} &= \sum_{\substack{i,j \\ j>i}} y_{ij}^{(g)} \mathbb{1}_{\{u_i^{(g)} \in [a_{k-1}, a_k]\}} \mathbb{1}_{\{u_j^{(g)} \in [a_{l-1}, a_l]\}} \\ m_{kl}^{(g)} &= \sum_{\substack{i,j \\ j>i}} \mathbb{1}_{\{u_i^{(g)} \in [a_{k-1}, a_k]\}} \mathbb{1}_{\{u_j^{(g)} \in [a_{l-1}, a_l]\}}. \end{aligned} \tag{8}$$

This means  $d_{kl}^{(g)}$  and  $m_{kl}^{(g)}$  represent the number of present ( $y_{ij}^{(g)} = 1$ ) and a priori potential ( $y_{ij}^{(g)} \in \{0, 1\}$ ) edges of network  $g$ , respectively, within the constructed rectangle  $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$ . The corresponding cross-network counts can be calculated by  $d_{kl} = d_{kl}^{(1)} + d_{kl}^{(2)}$  and  $m_{kl} = m_{kl}^{(1)} + m_{kl}^{(2)}$ . Since  $w^{\text{joint}}(\cdot, \cdot)$  is smooth, we further assume that the induced probability on edge variables within  $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$  is approximately constant. That allows for putting the observed ratios between present and absent edges in direct relation. In this light, we formulate the following contingency table to keep track of homogeneity between the networks within rectangle  $(k, l)$ :

$d_{kl}^{(1)}$	$d_{kl}^{(2)}$	$d_{kl}$
$m_{kl}^{(1)} - d_{kl}^{(1)}$	$m_{kl}^{(2)} - d_{kl}^{(2)}$	$m_{kl} - d_{kl}$
$m_{kl}^{(1)}$	$m_{kl}^{(2)}$	$m_{kl}$

Apparently, if  $H_0$  is assumed to be true, we would expect the proportions of present edges,  $d_{kl}^{(1)}/m_{kl}^{(1)}$  and  $d_{kl}^{(2)}/m_{kl}^{(2)}$ , to be similar. This can be assessed by contrasting the observed numbers of edges with their expectations conditional on the given margin totals, which is in line with the construction of Fisher's exact test on  $2 \times 2$  contingency tables. In this regard, the theoretical random counterpart of  $d_{kl}^{(1)}$  can be defined as

$$D_{kl}^{(1)} = \sum_{\substack{i,j \\ j>i}} Y_{ij}^{(1)} \mathbb{1}_{\{u_i^{(1)} \in [a_{k-1}, a_k)\}} \mathbb{1}_{\{u_j^{(1)} \in [a_{l-1}, a_l)\}},$$

for which under  $H_0$  it approximately holds that

$$D_{kl}^{(1)} \mid d_{kl} \sim \text{Hyp} \left( m_{kl}, d_{kl}, m_{kl}^{(1)} \right) \quad \text{with} \quad E_{kl}^{(1)} := \mathbb{E}(D_{kl}^{(1)} \mid d_{kl}) = m_{kl}^{(1)} \frac{d_{kl}}{m_{kl}} \tag{9}$$

$$\text{and} \quad V_{kl}^{(1)} := \mathbb{V}(D_{kl}^{(1)} \mid d_{kl}) = m_{kl}^{(1)} \frac{d_{kl}}{m_{kl}} \frac{m_{kl} - d_{kl}}{m_{kl}} \frac{m_{kl} - m_{kl}^{(1)}}{m_{kl} - 1}.$$

Based on these specifications, we define our test statistic as

$$T = \sum_{\substack{k,l \\ l \geq k}} \frac{\left(D_{kl}^{(1)} - E_{kl}^{(1)}\right)^2}{V_{kl}^{(1)}} \quad \text{with realization} \quad t = \sum_{\substack{k,l \\ l \geq k}} \frac{\left(d_{kl}^{(1)} - E_{kl}^{(1)}\right)^2}{V_{kl}^{(1)}}. \quad (10)$$

Note that we only include the quantities of the first network due to the symmetry of the hypergeometric distribution. Moreover, summands for which  $V_{kl}^{(1)} = 0$ —resulting from  $m_{kl}^{(1)}$ ,  $m_{kl} - m_{kl}^{(1)}$ ,  $d_{kl}$ , or  $m_{kl} - d_{kl}$  being zero—carry no relevant information and thus can simply be omitted from the calculation. In contrast, if  $m_{kl}^{(1)}$  is large,  $m_{kl}$  and  $d_{kl}$  are large compared to  $m_{kl}^{(1)}$ , and  $d_{kl}/m_{kl}$  is not close to zero or one, then  $D_{kl}^{(1)}$  is known to be approximately normally distributed. Given that, we can conclude that

$$T \stackrel{a}{\sim} \chi_{K(K+1)/2}^2 \quad (11)$$

since, in this scenario, the test statistic is essentially the sum of squared (conditionally) independent random variables that approximately follow a standard normal distribution. If the latter condition does not apply, and assumption (11) is not reasonable to hold, we still can simulate a sample of the theoretical distribution by drawing  $D_{kl}^{(1)} \mid d_{kl}$  according to (9) and calculating  $T$  as in (10). In both cases, we can easily derive a critical value  $c_{1-\alpha}$  to be compared with the realization  $t$  of the test statistic. To do so, we pick the corresponding  $(1 - \alpha)$ -quantile of either the theoretical distribution  $\chi_{K(K+1)/2}^2$  or the simulated sample. Finally, we reject null hypothesis (4) at the significance level of  $\alpha$  if  $t > c_{1-\alpha}$ . The choice of an appropriate  $K$  applied for these calculations is discussed in Section C of the Appendix. Note that altogether the presented test procedure follows a conception similar to the one underlying the log-rank test for time-to-event data.

Apparently, when conducting the test procedure on real-world networks, we obtain the

joint graphon and the corresponding alignment of the networks by applying the estimation procedure described in Section 4. In the end, this enables us to appropriately approximate test statistic (10). In this context, it is important to consider the general behavior of the joint graphon estimation under the alternative, that is, if hypothesis (4) does not hold. We stress that the intuition of the estimation procedure is to align the two networks as well as possible with respect to some suitable joint graphon model. Consequently, the expectation of  $T$  will be higher the more the true graphons  $w^{(1)}(\cdot, \cdot)$  and  $w^{(2)}(\cdot, \cdot)$  differ after “optimal” alignment. This clearly implies that the power of our test is higher for instances that deviate more strongly from the null hypothesis.

## 6 Applications

In this section, we showcase the applicability of the joint graphon estimation routine and the subsequent testing procedure. To give a comprehensive insight, this comprises both simulated and real-world networks. For an optimal estimation result and to best approximate test statistic (10), we repeat the estimation and testing procedure several times. In a modeling-oriented context, we would then typically pick the outcome with the lowest corrected AIC. However, since here the focus is on the statistical testing aspect, we choose the estimation result which leads to the highest  $p$ -value, assuming that this provides an optimal lower bound for the outcome under the (potentially existing) oracle network alignment.

## 6.1 Simulation Studies

### 6.1.1 Exemplary Application to Synthetic Data

To demonstrate the general capability of the joint graphon estimation and the performance of the subsequent testing procedure, we consider the graphon in the top left plot of Figure 3. Its formation is inspired by and can be interpreted as a stochastic blockmodel with smooth transitions between communities. Based on this ground-truth model specification, we simulate two networks with  $N^{(1)} = 200$  and  $N^{(2)} = 300$  by making use of data-generating process (2). To recover the underlying structure, we then apply the presented EM-type algorithm, where, for initialization, we make use of an uninformative random node positioning. After several iterations, we achieve the reasonable joint graphon estimate at the top right, which fully captures the structure of the ground-truth graphon. Relying on the accompanying estimated node positions, we subsequently conduct the testing procedure on whether the underlying distributions are equivalent. To this end, we start with calculating the rectangle-wise differences according to the construction of test statistic (10). The results are depicted as a heat map at the bottom left plot of Figure 3. This reveals that the difference in the local edge density is rather low to moderate in most rectangles, whereas it is distinctly higher in a few others. Aggregating these differences yields a test statistic of 203.2 as depicted by the black solid vertical line at the bottom right. Contrasting this result with the simulated 95% quantile of the distribution of  $T$  under  $H_0$  as the critical value (red dashed vertical line) yields no rejection. Hence, the underlying distributions of the two networks do not significantly differ with respect to a significance level of 5%. As a final remark with regard to the bottom right plot, the simulated distribution of  $T$  (black solid step function) and its theoretical approximation (blue dashed curve)—both relying on the assumption of  $H_0$  being true—are very close to one another. Consequently,

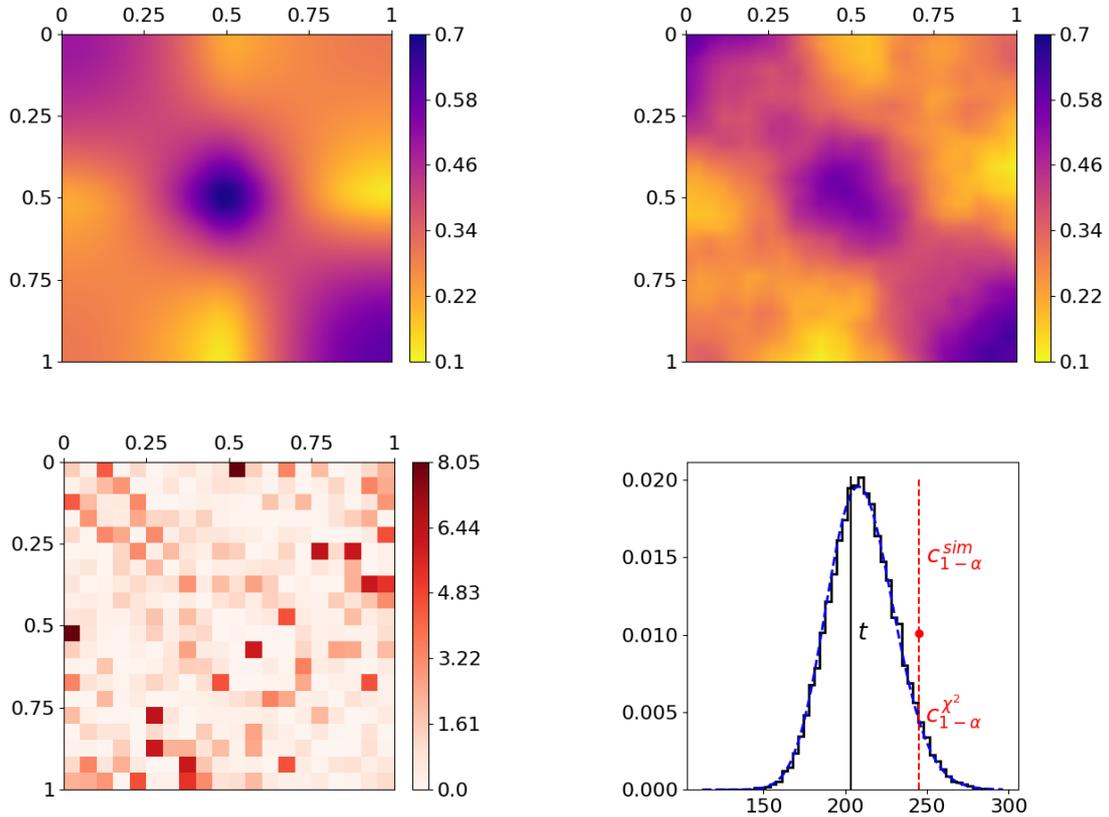


Figure 3: Joint graphon estimation for simulated networks with subsequent testing on equivalence of the underlying distribution models. The top row shows the true and the jointly estimated graphon on the left and right, respectively. The realizations of the terms of test statistic (10), representing the dissimilarities of the two networks per rectangle, are visualized at the bottom left, where  $m_{kl}^{(g)} \geq 100$  for  $k \neq l$  and  $\geq 45$  otherwise. The final result of the test statistic (black solid vertical line) as well as its distribution under  $H_0$  are illustrated at the bottom right, where the black solid step function and the blue dashed curve depict the simulated and the asymptotic chi-squared distribution, respectively. The red dashed vertical lines visualize the critical values at a significance level of 5%, derived from the simulated (upper line) and the asymptotic distribution (lower line).

they also provide very similar critical values, namely 243.6 and 244.8, respectively. This demonstrates that asymptotic distribution (11) represents a good approximation.

### 6.1.2 Performance Analysis under $H_0$

To evaluate the performance of the testing procedure in this example more profoundly, we repeat the above proceeding 400 times, with newly simulated networks in each trial (remaining with  $N^{(1)} = 200$  and  $N^{(2)} = 300$ ). Note that we run the estimation procedure always ten times (with different random node positions as varying initialization) and finally pick the highest  $p$ -value as the actual result for the given network pair. These repetitions already provide a broad insight into the method's performance under the given setting. An even more extensive evaluation becomes possible when, in contrast to the proceeding above, the testing procedure is performed on the basis of the oracle node positions. This allows us to dramatically reduce the computational burden since it releases us from the preceding (computationally expensive) model estimation. As a consequence, we are able to increase the number of conducted tests to 10,000. From these two repetition studies (using either  $\hat{\mathbf{u}}^{(g)}$  or  $\mathbf{u}^{(g)}$ ), we obtain rejection rates of 6.5% and 6.15% under the estimated and oracle node positioning, respectively. That means the test is slightly overconfident relative to the nominal significance level of 5%. The top row of Figure 4 shows additionally the empirical distributions of the observed  $p$ -values, illustrated as densities (left) and cumulative distribution functions (right). In accordance with the mildly inflated rejection rates, this exhibits a slight tendency to underestimate the  $p$ -value, i.e. interpreting the discrepancy as too high in distributional terms.

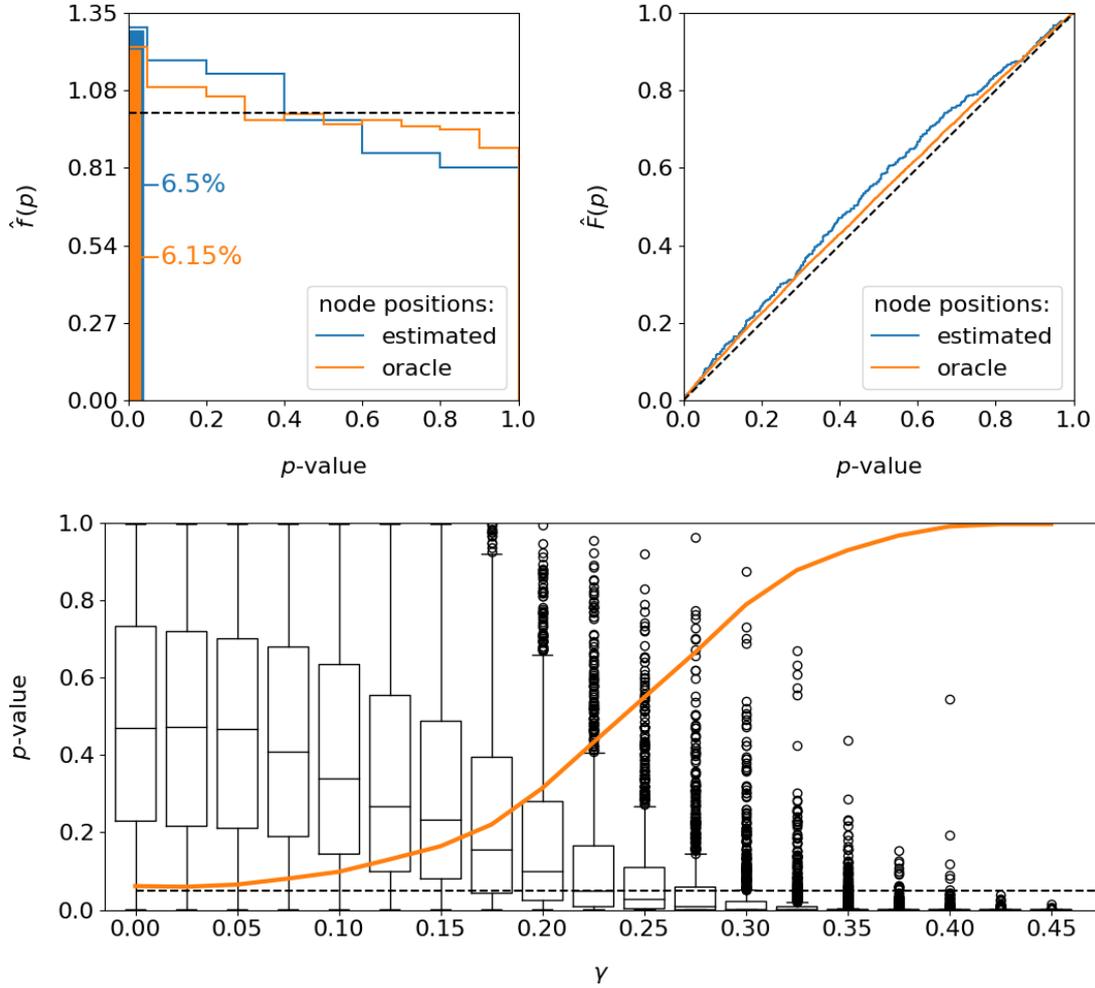


Figure 4: Performance of the testing procedure with regard to the resulting  $p$ -value; results are simulation-based. Top: empirical distribution of the  $p$ -value under  $H_0$ , illustrated as density and cumulative distribution function on the left and right, respectively. The black dashed lines illustrate the desired distributional behavior of an optimal test. Number of repetitions for estimated / oracle node positions: 400 / 10,000. Bottom: distribution of the  $p$ -value under  $H_1$  and the usage of oracle node positions (in box plot format); based on 1,000 repetitions each. The x-axis illustrates different settings according to formulation (12) (higher value of  $\gamma$  implies stronger deviation from  $H_0$ ). The black dashed horizontal line represents the 5% significance level, and the orange curve illustrates the corresponding power.

### 6.1.3 Performance Analysis under $H_1$

Conclusively, we are interested in evaluating the test performance under a false null hypothesis, which apparently requires formulating a suitable alternative. To this end, we “shrink” the heterogeneity within the graphon such that the present structure becomes less pronounced. The resulting graphon specification consequently tends more towards an Erdős–Rényi model, with the global density remaining unchanged. To be precise, based on  $w^{(1)}(\cdot, \cdot)$ , we formulate

$$w^{(2)}(u, v) := (1 - \gamma) w^{(1)}(u, v) + \gamma \bar{w}^{(1)} \quad (12)$$

with  $\gamma \in [0, 1]$  and  $\bar{w}^{(1)} = \iint w^{(1)}(u, v) du dv$ . Apparently, increasing the mixing parameter  $\gamma$  leads to a stronger deviation from  $H_0$ . At the same time, this setting guarantees an optimal alignment of  $w^{(1)}(\cdot, \cdot)$  and  $w^{(2)}(\cdot, \cdot)$ , meaning that there exists no rearrangement of  $w^{(2)}(\cdot, \cdot)$  that is closer to  $w^{(1)}(\cdot, \cdot)$  than specification (12). For this experiment, we again choose  $N^{(1)} = 200$  and  $N^{(2)} = 300$ . Moreover, here we rely exclusively on the oracle node positions. This provides a lower bound of the power since the rejection rate can be expected to be higher when using estimated node positions instead (cf. previous analysis under  $H_0$ ). The results for this setup are presented in the bottom plot of Figure 4, where the distribution of the resulting  $p$ -value is illustrated for different settings of  $\gamma$ . The orange curve additionally visualizes the resulting power, i.e. the proportion of cases with  $p < 0.05$ . These results clearly show that the probability of detecting the false null hypothesis monotonically increases as the parameter  $\gamma$  gets larger, which underpins the appropriateness of our test procedure.

Overall, the obtained simulation results demonstrate that the elaborated estimation and

testing procedure yields reasonable results for assessing structural differences between networks. Building upon these findings, we next want to investigate the method’s performance on real-world data.

## 6.2 Real-World Examples

### 6.2.1 Facebook Ego Networks

As a first real-world example, we consider two Facebook ego networks which have been assembled by Leskovec and McAuley (2012) and are publicly available on the Stanford Large Network Dataset Collection (Leskovec and Krevl, 2014). The two ego networks consist of 333 and 168 individuals, respectively, where the ego nodes are not included. An illustration of these networks is given in the top row of Figure 5, with nodes being colored according to resulting positions. In both networks, these final estimated node positions appear to be in line with the given network structure in terms of reflecting the nodes’ embedding within the network. Moreover, they seem to be aligned across networks. For example, in both networks, the reddish nodes represent the rather centric individuals, whereas the nodes from the dark blue spectrum constitute a moderately interconnected branch that is more detached from the rest of the network. However, the segment-wise differences depicted at the bottom left in Figure 5 exhibit some severe deviations. This can be clearly traced back to the blockwise division of the adjacency matrices as it results from partitioning the domain of edge positions (middle row). The aggregated differences ultimately result in a test statistic that is far from the sector of plausible values under the null hypothesis, as illustrated at the bottom right. Consequently, we can conclude that the structural behavior in the two networks differs significantly.

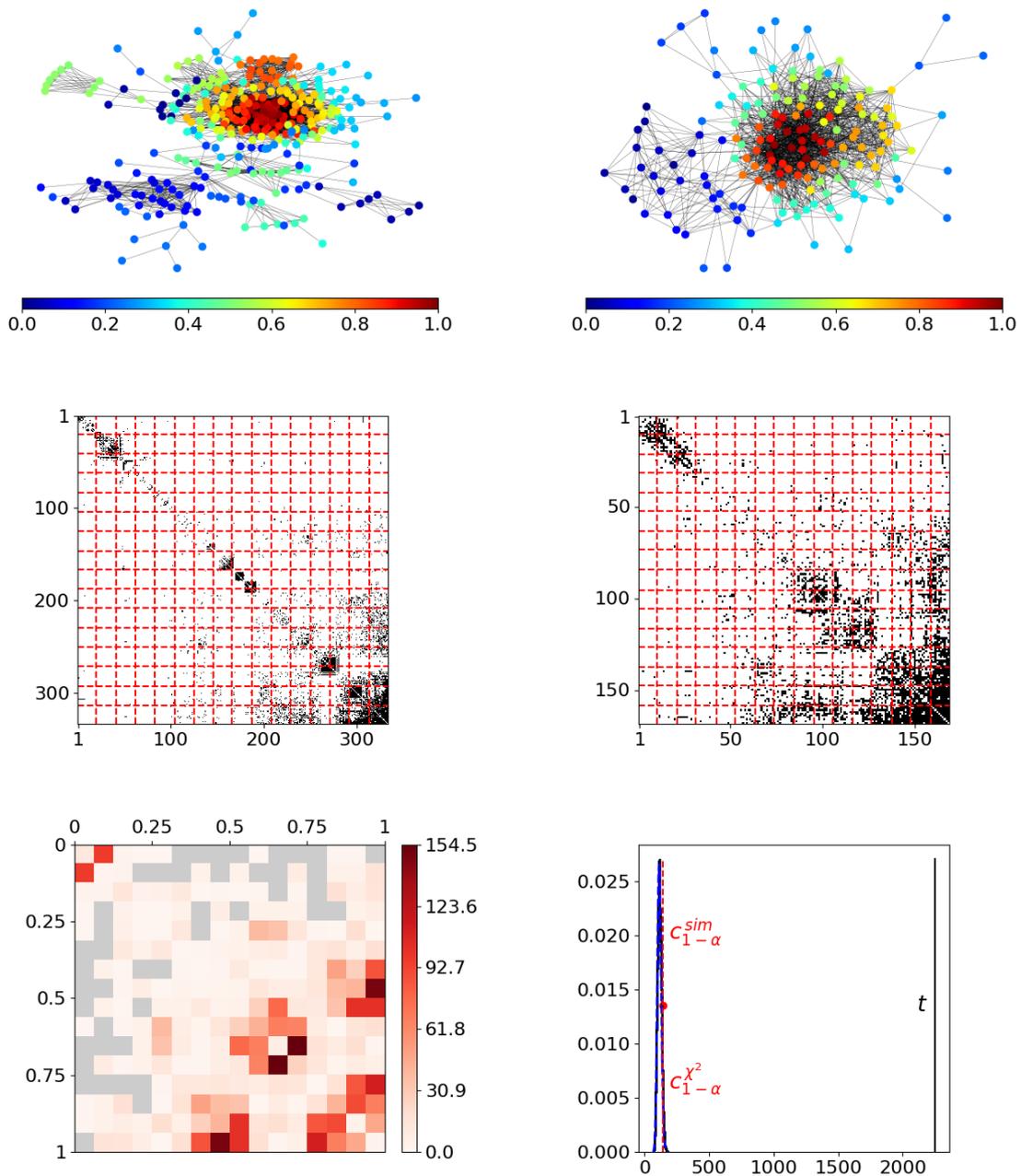


Figure 5: Comparison of two Facebook ego networks. Top: illustration of networks with coloring referring to estimated node positions. Middle: ordered adjacency matrices divided into blockwise segments. Bottom left: segment-wise differences between the two networks with  $m_{kl}^{(g)} \geq 100$  for  $k \neq l$  and  $\geq 45$  otherwise; gray rectangles do not contain any observed edges ( $d_{kl} = 0$ ) and thus provide no information. Bottom right: realization of test statistic (black solid vertical line) plus corresponding distribution under  $H_0$  (black solid step function and blue dashed curve represent simulated and asymptotic chi-squared distribution, respectively); critical values derived from the two types of distributions are represented by the upper and the lower red dashed vertical line.

### 6.2.2 Human Brain Functional Coactivation Networks

In the second real-world application, we are concerned with differences in the human brain coactivation structure. To be precise, we compare two types of individuals, one with autism spectrum disorder (ASD) and the other with typical development (TD). In particular, we are interested in whether the functional connectivity within the brain significantly differs between these two groups (cf. the introductory example from Figure 1). For this analysis, we use resting-state functional magnetic resonance imaging data from the Autism Brain Imaging Data Exchange project (ABIDE I, 2013). More specifically, we employ preprocessed data provided by the Preprocessed Connectomes Project platform (PCP, 2015). Based on these person-specific datasets, we first calculate correlations between brain regions with respect to concurrent activation over time. Aggregating the results of participants from the same clinical group and employing an appropriate threshold finally yields the network-structured data which we aim to compare. To be precise, by performing the described preprocessing, we achieve for both groups, ASD and TD, a global connectivity pattern between 116 prespecified relevant brain regions. Note that these regions are the same for both groups, which is why this could also be viewed as a comparison task under known node correspondence. However, we emphasize that in neurosciences, the transfer of competencies between brain regions is a well-known phenomenon, wherefore the general functional connectivity structure might be of greater relevance than the functional connection between specific regions. Further details on the acquisition and adequate transformation of the data are provided in Section D of the Appendix.

For analyzing the differences in the brain coactivation structure between the two diagnostic groups, we again start with appropriately aligning the two networks. This is apparently done by employing the joint graphon estimation routine. The resulting node

positions in relation to the embedding of nodes within the networks are illustrated in the top row of Figure 6. Again this reveals a plausible allocation of the nodes in the joint graphon model. The structural evolution can further be evaluated by consulting the correspondingly ordered adjacency matrices (see middle row), where the red dashed lines represent the blockwise division resulting from the assignment of the edge positions to the rectangles in  $[0, 1]$ . At first view, the formed structure looks quite similar in both matrices. Yet, on closer inspection, some blocks can be found where the density seems considerably different. This is also observed in the rectangle-wise differences depicted in the bottom left plot. In the end, aggregating these differences leads to rejecting the null hypothesis, as represented at the bottom right. To be precise, this test decision is based on a resulting  $p$ -value of 0.013 (with reference to the simulated distribution under  $H_0$ ).

Given that this outcome does not support an utterly unambiguous decision of the conducted test procedure, one might additionally be interested in the nature of the inferred differences. To address this, in Section 1.2 of the Supplementary Material, we localize different behavior between the networks on the microscopic scale. Note that this can also be derived more or less directly from the joint graphon model. Moreover, for comparison reasons, we repeated the above analysis for two randomly selected disjoint subgroups of the TD group. According to the results illustrated in Figure 3 of the Supplementary Material, in this scenario, we do not observe a significant overall deviation. This further underlines the findings about the dissimilarity between the ASD and the TD group.

## 7 Discussion and Conclusion

In the network comparison literature, the task of drawing statistical inference appears to be an open challenge up to now. We addressed this shortcoming in this paper by

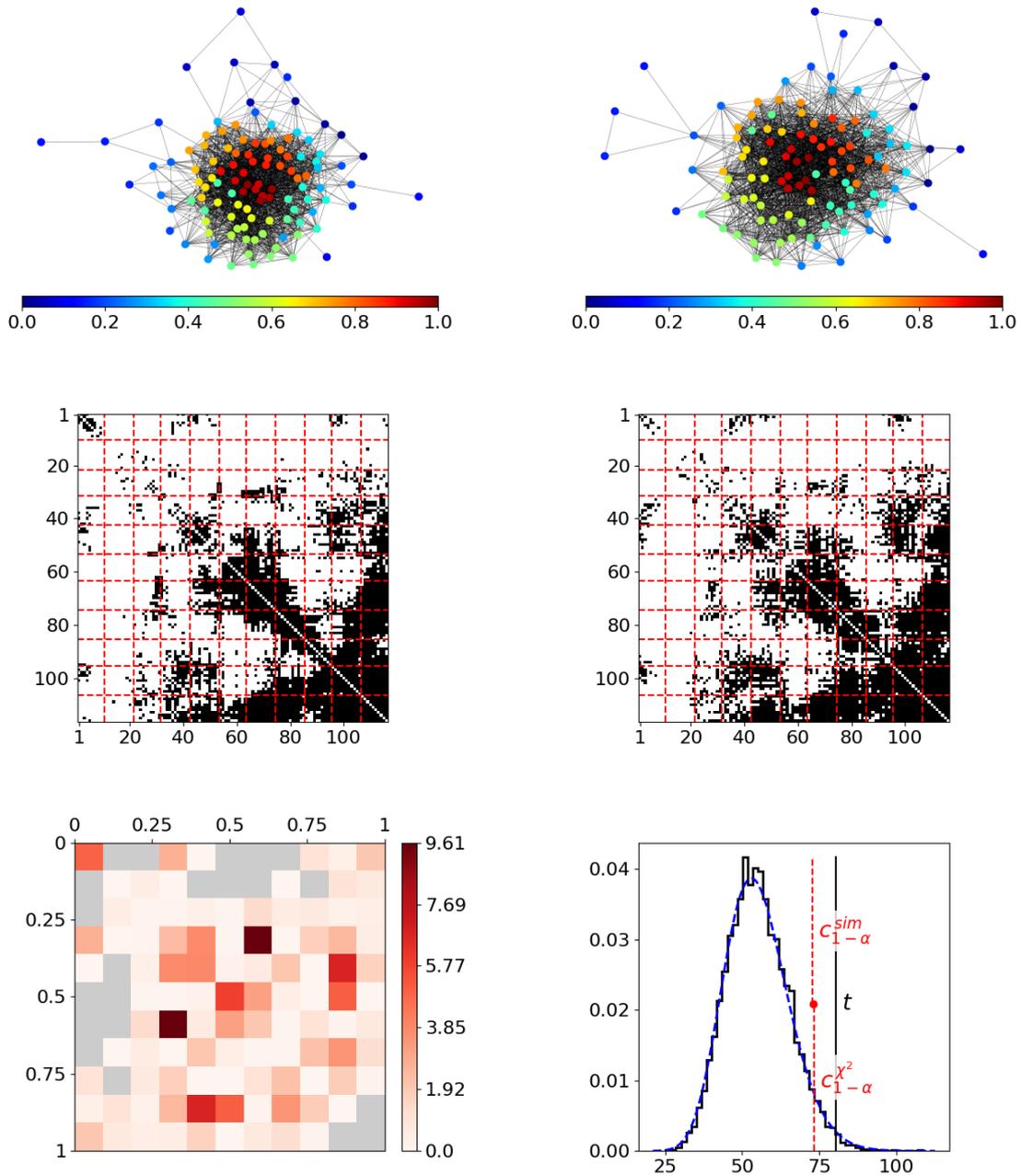


Figure 6: Comparison of functional coactivation in the human brain between groups of subjects with autism spectrum disorder and with typical development. The top row shows the networks of the ASD and the TD group on the left- and right-hand side, respectively. All illustration aspects are equivalent to the representation in Figure 5. The number of nodes per rectangle is again given by  $m_{kl}^{(g)} \geq 100$  for  $k \neq l$  and  $\geq 45$  otherwise, where N/A's in the blockwise differences result from  $d_{kl}$  or  $m_{kl} - d_{kl}$  being zero.

developing a nonparametric test on the equivalence of network structures. To do so, we utilized the smooth graphon model as a powerful tool for both describing and modeling the structure in complex networks. More precisely, extending previous estimation approaches towards a joint modeling framework allowed us to formulate a more generalized network alignment. Given that, local structure comparison can be carried out to uncover differences on the microscopic scale. Adequately aggregating these local differences finally enables to construct an appropriate nonparametric testing procedure on network data. Applying this comparison strategy to simulated and real-world networks clearly demonstrated its general applicability.

As outlined before, a crucial point for the proposed approach to work is the graphon model's property of decoupling structure and size. In the same line, one could think of further decoupling the global density by following the approach of Bickel and Chen (2009), i.e. by introducing network-related quantities  $\rho^{(g)}$  that serve as individual density coefficients. Specifically, this means modifying formulation (2) by employing  $\rho^{(g)}w^{(g)}(\cdot, \cdot)$ , where  $\hat{\rho}^{(g)} = [N^{(g)}(N^{(g)} - 1)]^{-1} \sum_{i,j} y_{ij}^{(g)}$  could serve as an estimate that is independent of the rest of the structure. Such a framework consequentially might lead to a more balanced comparison strategy.

Beyond the applications presented in the previous section, which all refer to the situation with two networks, the method could easily be extended to cases with multiple or even a single network. For example, in the one-sample setting, to test whether a given network follows a hypothetical distribution  $\mathbb{P}(\mathbf{Y} = \cdot; w(\cdot, \cdot), N)$ , we could first align the network with the theoretical graphon. That is, applying the E-step based on  $w(\cdot, \cdot)$ . Given this alignment, we could then turn distributional assumption (9) into a binomial distribution with the rectangle-specific mean over  $w(\cdot, \cdot)$  as success probability, and, based on that,

calculate the test statistic as in (10).

Besides the testing aspect, our approach could further be used to uncover relevant differences between networks on the microscopic scale. To be precise, determining the nodes or edges (present or absent) that contribute most to a quantified structural discrepancy between networks is possibly interesting in many situations. This is further elaborated in Section 1.1 of the Supplementary Material.

As a last application case, the joint graphon estimation could further be used to predict edges *between* separated networks by considering the cross-sample probabilities  $w^{\text{joint}}(u_{i_1}^{(1)}, u_{i_2}^{(2)})$ . This might be of interest when two (or more) networks are assumed to be samples of a larger global network. To the best of our knowledge, this has not been pursued by any other approach so far and hence constitutes a novel perspective. As a particular hurdle in this framework, the sampling strategy that is supposed for the drawing of subnetworks needs to be taken into account in the estimation routine. As far as this is not the Simple Induced Subgraph Sampling, where one selects a simple random sample of nodes within which all edges are observed, further adaptations are required. Hence, this lies beyond the scope of the paper.

## SUPPLEMENTARY MATERIAL

**Supplementary Manuscript:** (i) Description for deriving differences between networks at the microscopic level. As exemplary application, the two brain networks from Section 6.2.2 are considered.

(ii) Replication of the test on functional coactivation networks for two subgroups of the typical-development group (confer Section 6.2.2).

**Python-package for testing on structural equivalence:** Python-package containing the

code to perform the comparison methods described in the paper. The package also contains the preprocessed data of the human brain functional coactivation networks (see Section 6.2.2). (GNU zipped tar file)

## References

- ABIDE I (2013). Autism Brain Imaging Data Exchange I. Available online: [http://fcon\\_1000.projects.nitrc.org/indi/abide/](http://fcon_1000.projects.nitrc.org/indi/abide/) [accessed 06-21-2022].
- Ali, W., T. Rito, G. Reinert, F. Sun, and C. M. Deane (2014). Alignment-free protein interaction network comparison. *Bioinformatics* 30(17), i430–i437.
- Andersen, M., J. Dahl, and L. Vandenberghe (2016). CvxOpt: Open source software for convex optimization (Python). Available online: <https://cvxopt.org> [accessed 11-03-2022]. Version 1.2.7.
- Bassett, D. S., P. Zurn, and J. I. Gold (2018). On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience* 19(9), 566–578.
- Bhattacharya, K., G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna (2008). The international trade network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment* 2008(02), P02002.
- Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences* 106(50), 21068–21073.
- Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *The Annals of Statistics* 39(5), 2280–2301.

- Borgs, C., J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics* 219(6), 1801–1851.
- Burnham, K. and D. Anderson (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. (2nd ed.). New York: Springer.
- Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology* 11(1), 13–41.
- Chan, S. H. and E. M. Airoldi (2014). A consistent histogram estimator for exchangeable graph models. In *31st International Conference on Machine Learning, ICML 2014*, pp. 208–216.
- Craddock, C., Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, et al. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics* 7, 27.
- Crossley, N. A., A. Mechelli, P. E. Vértes, T. T. Winton-Brown, A. X. Patel, C. E. Ginestet, P. McGuire, and E. T. Bullmore (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences* 110(28), 11583–11588.
- Di Martino, A., C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry* 19(6), 659–667.

- Diaconis, P. and S. Janson (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni* 28, 33–61.
- Eagle, N., A. Pentland, and D. Lazer (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences* 106(36), 15274–15278.
- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–102.
- Emmert-Streib, F., M. Dehmer, and Y. Shi (2016). Fifty years of graph matching, network alignment and network comparison. *Information Sciences* 346-347, 180–197.
- Faisal, F. E., K. Newaz, J. L. Chaney, J. Li, S. J. Emrich, P. L. Clark, and T. Milenković (2017). Grafene: Graphlet-based alignment-free network approach integrates 3d structural and sequence (residue order) data to improve protein structural comparison. *Scientific Reports* 7(1), 1–15.
- Gera, R., L. Alonso, B. Crawford, J. House, J. A. Mendez-Bermudez, T. Knuth, and R. Miller (2018). Identifying network structure similarity using spectral graph theory. *Applied Network Science* 3(1), 1–15.
- 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. W., K. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks* 5(2), 109–137.
- Hurvich, C. M. and C. L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* 76(2), 297–307.

- Kauermann, G., C. Schellhase, and D. Ruppert (2013). Flexible copula density estimation with penalized hierarchical B-splines. *Scandinavian Journal of Statistics* 40(4), 685–705.
- Kuchaiev, O., T. Milenković, V. Memišević, W. Hayes, and N. Pržulj (2010). Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface* 7(50), 1341–1354.
- Latouche, P. and S. Robin (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing* 26(6), 1173–1185.
- Leskovec, J. and A. Krevl (2014). SNAP Datasets: Stanford large network dataset collection. Available online: <http://snap.stanford.edu/data> [accessed 05-27-2022].
- Leskovec, J. and J. McAuley (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems (NIPS)*, Volume 25, pp. 539–547.
- Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory. Series B* 96(6), 933–957.
- Marron, J. S. and A. M. Alonso (2014). Overview of object oriented data analysis. *Biometrical Journal* 56(5), 732–753.
- Navarro, M. and S. Segarra (2022). Joint network topology inference via a shared graphon model. *arXiv preprint arXiv:2209.08223*.
- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press.
- Nowicki, K. and T. A. Snijders (2001). Estimation and prediction for stochastic block-structures. *Journal of the American Statistical Association* 96(455), 1077–1087.

- Olhede, S. C. and P. J. Wolfe (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences* 111(41), 14722–14727.
- Onnela, J.-P., D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones (2012). Taxonomies of networks from community structure. *Physical Review E* 86(3), 036104.
- Pascual-Belda, A., A. Díaz-Parra, and D. Moratal (2018). Evaluating functional connectivity alterations in autism spectrum disorder using network-based statistics. *Diagnostics* 8(3), 51.
- PCP (2015). Preprocessed Connectomes Project: ABIDE. Available online: <http://preprocessed-connectomes-project.org/abide/> [accessed 06-21-2022].
- Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics* 23(2), e177–e183.
- Pržulj, N., D. G. Corneil, and I. Jurisica (2004). Modeling interactome: Scale-free or geometric? *Bioinformatics* 20(18), 3508–3515.
- 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.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2009). Semiparametric regression during 2003—2007. *Electronic Journal of Statistics* 3, 1193–1256.
- Schwikowski, B., P. Uetz, and S. Fields (2000). A network of protein–protein interactions in yeast. *Nature biotechnology* 18(12), 1257–1261.

- Sischka, B. and G. Kauermann (2022a). EM-based smooth graphon estimation using MCMC and spline-based approaches. *Social Networks* 68, 279–295.
- Sischka, B. and G. Kauermann (2022b). Stochastic block smooth graphon model. *arXiv preprint arXiv:2203.13304*.
- Snijders, T. A. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* 14(1), 75–100.
- Song, Y., T. M. Epalle, and H. Lu (2019). Characterizing and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis. *Frontiers in Human Neuroscience* 13, 203.
- Soundarajan, S., T. Eliassi-Rad, and B. Gallagher (2014). A guide to selecting a network similarity method. In *SIAM International Conference on Data Mining 2014, SDM 2014*, Volume 2, pp. 1037–1045.
- Subbaraju, V., M. B. Suresh, S. Sundaram, and S. Narasimhan (2017). Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. *Medical Image Analysis* 35, 375–389.
- Tantardini, M., F. Ieva, L. Tajoli, and C. Piccardi (2019). Comparing methods for comparing networks. *Scientific Reports* 9(1), 1–19.
- Turlach, B. A. and A. Weingessel (2013). quadprog: Functions to solve quadratic programming problems (R). Available online: <https://CRAN.R-project.org/package=quadprog> [accessed 11-03-2022]. Version 1.5-5.

- Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1), 273–289.
- Wilson, R. C. and P. Zhu (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition* 41(9), 2833–2841.
- Wolfe, P. J. and S. C. Olhede (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC Press.
- Xu, T., Z. Yang, L. Jiang, X.-X. Xing, and X.-N. Zuo (2015). A connectome computation system for discovery science of brain. *Science Bulletin* 60(1), 86–95.
- Yaveroglu, Ö. N., N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic, and N. Pržulj (2014). Revealing the hidden language of complex networks. *Scientific Reports* 4(1), 1–9.
- Yaveroglu, Ö. N., T. Milenković, and N. Pržulj (2015). Proper evaluation of alignment-free network comparison methods. *Bioinformatics* 31(16), 2697–2704.

# Appendix

## A Implementation of the Gibbs Sampler

In an iterative joint graphon estimation procedure, the joint posterior distribution of the node positions given  $w^{\text{joint}}(\cdot, \cdot)$  can be simulated by constructing a Gibbs sampler. We stress that the node positions are independent between networks and thus the Gibbs sampling procedure can be conducted for each network separately. The MCMC framework is then build upon full conditional distribution (5) and can be formulated as follows. For the successive updating procedure, we consider  $\mathbf{u}^{(g), \langle t \rangle} = (u_1^{(g), \langle t \rangle}, \dots, u_{N^{(g)}}^{(g), \langle t \rangle})$  to be the current state of the Markov chain. In the  $(t + 1)$ -th step, component  $i$  is then updated according to (5), where all other components remain unchanged, i.e.  $u_j^{(g), \langle t+1 \rangle} := u_j^{(g), \langle t \rangle}$  for  $j \neq i$ . To do so, we propose a new position  $u_i^{(g), *}$  by drawing from a normal distribution under the application of a logit link. To be precise, we first calculate

$$v_i^{(g), \langle t \rangle} = \text{logit}(u_i^{(g), \langle t \rangle}) = \log \left( \frac{u_i^{(g), \langle t \rangle}}{1 - u_i^{(g), \langle t \rangle}} \right),$$

then we add a normal term in the form of  $v_i^{(g), *} = v_i^{(g), \langle t \rangle} + \text{Normal}(0, \sigma_v^2)$ , and finally we accomplish the retransformation through

$$u_i^{(g), *} = \text{logit}^{-1}(v_i^{(g), *}) = \frac{\exp(v_i^{(g), *})}{1 + \exp(v_i^{(g), *})}.$$

In this setting, the variance  $\sigma_v^2$  should be chosen such that a balance between a wide-ranging exploration and a high acceptance rate is achieved. Given these formulations, the proposal

density can be written as

$$\begin{aligned}
q(u_i^{(g),*} | u_i^{(g),\langle t \rangle}) &= \frac{\partial u_i^{(g),*}}{\partial z_i^{(g),*}} \phi(z_i^{(g),*} | z_i^{(g),\langle t \rangle}) \\
&\propto \frac{1}{u_i^{(g),*} (1 - u_i^{(g),*})} \\
&\quad \cdot \exp\left(-\frac{1}{2} \frac{(\text{logit}(u_i^{(g),*}) - \text{logit}(u_i^{(g),\langle t \rangle}))^2}{\sigma^2}\right),
\end{aligned}$$

which leads to a proposal ratio of

$$\frac{q(u_i^{(g),\langle t \rangle} | u_i^{(g),*})}{q(u_i^{(g),*} | u_i^{(g),\langle t \rangle})} = \frac{u_i^{(g),*} (1 - u_i^{(g),*})}{u_i^{(g),\langle t \rangle} (1 - u_i^{(g),\langle t \rangle})}.$$

In combination with the likelihood ration, the acceptance probability of the proposal, i.e.

the probability for setting  $u_i^{(g),\langle t+1 \rangle} := u_i^{(g),*}$ , can be calculated through

$$\min \left\{ 1, \prod_{j \neq i} \left[ \left( \frac{w(u_i^{(g),*}, u_j^{(g),\langle t \rangle})}{w(u_i^{(g),\langle t \rangle}, u_j^{(g),\langle t \rangle})} \right)^{y_{ij}} \cdot \left( \frac{1 - w(u_i^{(g),*}, u_j^{(g),\langle t \rangle})}{1 - w(u_i^{(g),\langle t \rangle}, u_j^{(g),\langle t \rangle})} \right)^{1 - y_{ij}} \right] \frac{u_i^{(g),*} (1 - u_i^{(g),*})}{u_i^{(g),\langle t \rangle} (1 - u_i^{(g),\langle t \rangle})} \right\}.$$

In case the decision yields a rejection of the proposal, we set  $u_i^{(g),\langle t+1 \rangle} := u_i^{(g),\langle t \rangle}$ . Applying this updating strategy, which comprises the proposal of a new position plus the decision about its acceptance, to all  $i = 1, \dots, N^{(g)}$  completes one global update. Finally, we achieve a proper Gibbs sampling routine through consecutively repeating this global updating scheme. After cutting the burn-in period and applying an appropriate thinning, this approach yields a sample of the desired joint posterior distribution of the node positions.

## B Derivative and Penalization of the B-Spline Function

As has been show in Section 4.2, the log-likelihood of a B-spline function can be straightforwardly extended towards the situation with multiple datasets. Given the formulation from (6), the score function can be calculated as

$$\begin{aligned} \mathbf{s}(\boldsymbol{\theta}) &= \left[ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^\top \\ &= \sum_g \sum_{\substack{i,j \\ j \neq i}} [\mathbf{B}_{ij}^{(g)}]^\top \left( \frac{y_{ij}^{(g)}}{w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)})} - \frac{1 - y_{ij}^{(g)}}{1 - w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)})} \right). \end{aligned}$$

This, in turn, leads to the Fisher information in the form of

$$\begin{aligned} \mathbf{F}(\boldsymbol{\theta}) &= -\mathbb{E} \left( \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &= \sum_g \sum_{\substack{i,j \\ j \neq i}} [\mathbf{B}_{ij}^{(g)}]^\top \mathbf{B}_{ij}^{(g)} \left[ w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)}) \cdot \left( 1 - w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)}) \right) \right]^{-1}. \end{aligned}$$

These results can then be used to implement the Fisher scoring procedure, where in (7) we additionally add a penalization term to guarantee a smooth estimation result. For penalizing “neighborhood” elements of the parameter vector  $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1L}, \theta_{21}, \dots, \theta_{LL})^\top$ , the penalization matrix can be formulated through

$$\mathbf{P} = (\mathbf{J}_L \otimes \mathbf{I}_L)^\top (\mathbf{J}_L \otimes \mathbf{I}_L) + (\mathbf{I}_L \otimes \mathbf{J}_L)^\top (\mathbf{I}_L \otimes \mathbf{J}_L),$$

where  $\mathbf{I}_L$  is the identity matrix of size  $L$  and

$$\mathbf{J}_L = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(L-1) \times L}.$$

## C Choosing the Number and Extent of Rectangles

In order to appropriately test null hypothesis (4), in Section 5, we have developed an approach that relies on the partition of the graphon's domain. According to formulation (8), that involves the number of rectangles,  $K$ , as well as their concrete specification in the form of  $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$ . In this regard, we emphasize that two aspects need to be observed. On the one hand, the joint graphon should be approximately constant within rectangles, requiring  $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$  to be not too expansive. On the other hand, the amount of edge variables per network falling into these blocks should be high, which needs rather broad rectangles. Thus, a trade-off between these two opposed requirements should be reached. In general, we choose  $K$  to grow more slowly than both network dimensions, e.g. scaling as  $\min_g \sqrt{N^{(g)}}$ . Note that choosing  $K = 1$  would imply to test whether the two networks possess the same global density under the assumption of a joint Erdős–Rényi model. Having determined a suitable value for  $K$ , we then simply specify the boundaries of the rectangles through  $a_k = k/K$  for  $k = 0, \dots, K$ . In combination with the subsequent adjustment of the latent positions as described in Section 4.1, which leads to equidistance of the estimates  $\hat{u}_i^{(g)}$ , a general lower bound for the amount of contained nodes per interval,

$N_k^{(g)}$ , can be derived. To be precise, we can formulate

$$N_k^{(g)} = \left| \left\{ i \in \{1, \dots, N^{(g)}\} : \frac{i}{N^{(g)} + 1} \in [a_{k-1}, a_k) \right\} \right| \geq \left\lfloor \frac{1}{K} (N^{(g)} + 1) \right\rfloor,$$

where  $\lfloor x \rfloor$  returns the largest integer smaller than or equal to  $x$ . Given that,  $K$  could also be chosen such that, per network, a prescribed minimum amount of edge variables per rectangle ( $N_k^{(g)} N_l^{(g)}$  for  $l > k$  and  $N_k^{(g)}(N_k^{(g)} - 1)/2$  for  $l = k$ ) is guaranteed.

As a final remark, we emphasize that with regard to the rectangle-based test statistic, it seems natural to alternatively apply a histogram estimator (Chan and Airoldi, 2014 or Olhede and Wolfe, 2014). However, the smooth graphon estimation adapted from Sischka and Kauermann (2022b) considers a global node ordering which refers not only to separated intervals but to the entire domain of  $[0, 1]$ . This consequently facilitates the iterative estimation procedure and thus leads to a more plausible and faster converging node positioning.

## D Acquiring and Processing of Brain Functional Activation Data

The data we use for analyzing differences in the functional brain activation are originally provided by the Autism Brain Imaging Data Exchange project (ABIDE I, 2013, Di Martino et al., 2014). However, we make use of preprocessed data that are directly accessible through the Preprocessed Connectomes Project platform (PCP, 2015, Craddock et al., 2013). To be precise, we here apply the Connectome Computation System pipeline (Xu et al., 2015) and the reduction to the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002). For each participant, this yields a dataset that consists of activity measurements over time for 116 prespecified brain regions (a.k.a. regions of interest). Based on these temporal activity measurements, we calculate Pearson’s correlation coefficient between all pairs of

brain regions which, per participant, leads to the corresponding functional connectivity matrix (Song et al., 2019, Subbaraju et al., 2017). For this analysis, we rely on the data from New York University, comprising 73 ASD patients and 98 TD subjects. For aggregating these connectivity patterns per clinical group, we apply Fisher's transformation to the pairwise correlation coefficients, calculate their mean for all pairs of brain regions, and finally retransform these means (Pascual-Belda et al., 2018). This yields for both diagnostic groups a  $116 \times 116$  weighted connectivity matrix which we binarize by employing a threshold of 0.4. Based on that, the two final networks we obtain both possess a global density of about 30%. With regard to the choice of the threshold, Song et al. (2019) have found that this has only minor effects when comparing the networks.

# Supplementary Material for 'Nonparametric Two-Sample Test for Networks Using Joint Graphon Estimation'

Benjamin Sischka and Göran Kauermann

March 28, 2023

## 1 Detecting Differences on the Microscopic Scale

### 1.1 Methodological Construction

In addition to the testing aspect, the network alignment based on the joint smooth graphon model also allows for determining structural differences on the edge level. In that regard, we are especially interested in differences that occur when inferring the structure separately. To address this, we first fit two separate graphon models to the two networks individually. To be precise, for this operation, we employ the node positions obtained from estimating the *joint* graphon model and then perform the M-step as described in Section 4.2 of the paper but reduced to the use of only  $\mathbf{y}^{(1)}$  or  $\mathbf{y}^{(2)}$ . This yields the separate estimates  $\hat{w}^{(1)}(\cdot, \cdot)$  and  $\hat{w}^{(2)}(\cdot, \cdot)$ , respectively. Based on that, for  $g_1, g_2 = 1, 2$  with  $g_1 \neq g_2$ , we calculate

$$\hat{w}_{(g_1)(g_2)}^{\text{diff}}(u, v) = \frac{\hat{w}^{(g_1)}(u, v) - \hat{w}^{(g_2)}(u, v)}{\sqrt{\{\hat{w}^{(1)}(u, v) [1 - \hat{w}^{(1)}(u, v)] + \hat{w}^{(2)}(u, v) [1 - \hat{w}^{(2)}(u, v)]\} / 2}}$$

for  $(u, v)^\top \in [0, 1]^2$ . With regard to data-generating process (2), this can be interpreted as the difference between expectations in relation to the averaged standard deviation. Hence,  $|\hat{w}_{(g_1)(g_2)}^{\text{diff}}(\cdot, \cdot)|$  provides an appropriate measure to quantify the local differences between

the found graphon structures. Moreover, it can be considered as a smoothed version of the impact on test statistic (10). In turn, the contribution of the present or absent edge between node pair  $(i, j)$  of network  $g_1$  to the difference in the inferred structure can be quantified by evaluating

$$\begin{cases} \hat{w}_{(g_1)(g_2)}^{\text{diff},+}(\hat{u}_i^{(g_1)}, \hat{u}_j^{(g_1)}), & \text{if } y_{ij}^{(g_1)} = 1 \\ \hat{w}_{(g_2)(g_1)}^{\text{diff},+}(\hat{u}_i^{(g_1)}, \hat{u}_j^{(g_1)}), & \text{otherwise} \end{cases}$$

with  $\hat{w}_{(g_1)(g_2)}^{\text{diff},+}(u, v) = \max\{0, \hat{w}_{(g_1)(g_2)}^{\text{diff}}(u, v)\}$ . In particular, this means that the contribution is zero if the considered present or absent edge is contrary to the direction of the detected difference. Note that also here, the estimated node positions  $\hat{u}_i^{(g_1)}$  are the ones stemming from the network alignment, meaning the positions resulting from estimating the joint graphon model. We stress that this approach is different from determining the deviation of present or absent edges from their “transferred” expectation, i.e. from simply calculating  $|y_{ij}^{(g_1)} - \hat{w}^{(g_2)}(\hat{u}_i^{(g_1)}, \hat{u}_j^{(g_1)})|$ . Instead, here we aim to highlight those connections that, in one way or another, *collectively* have enough impact to actually affect the inferred structure. To provide an illustrative intuition, this approach for detecting deviating behavior on the microscopic scale is additionally illustrated in Figure 1. This representation allows to graphically demonstrate the single steps and thus to make the procedure much clearer.

## 1.2 Application to Human Brain Coactivation Example

The analysis of the functional coactivation in the human brain has yielded a rather narrow test decision with regard to the two clinical groups, see Section 6.2.2 of the paper. Hence, we now additionally analyze the differences at the microscopic level. To quantify these differences, we make use of the approach described above. The corresponding results for this

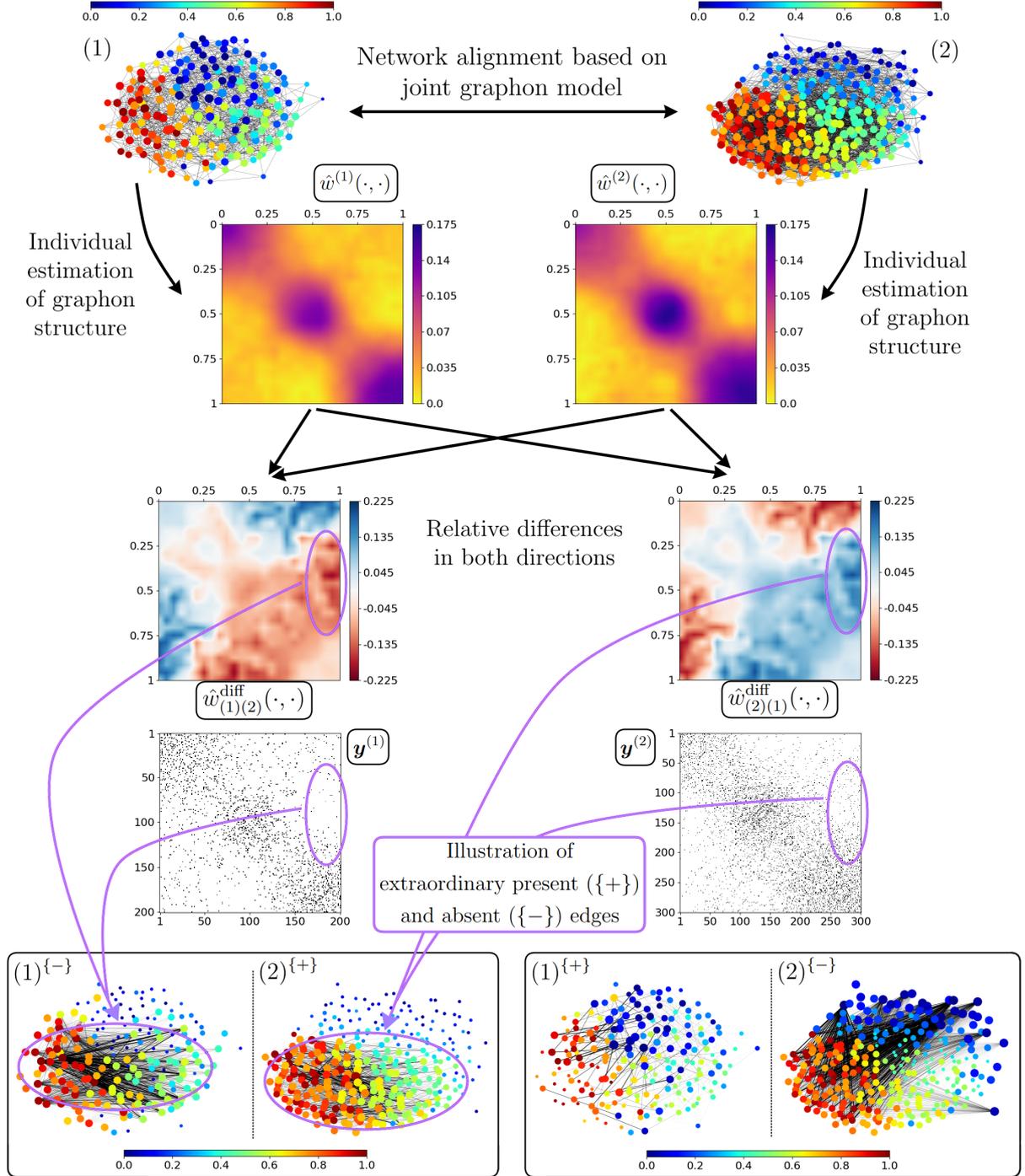


Figure 1: Detecting differences in networks at the microscopic level by employing the joint graphon-based alignment technique. The final network representations at the bottom row illustrate extraordinary *absent* edges in network 1 and extraordinary *present* edges in network 2 on the left and the converse on the right. The steps to get there are as follows: (i) Align networks based on joint smooth graphon estimation. (ii) Estimate individual graphons for disjoint networks separately. (iii) Calculate relative differences between graphon estimates [ $\rightarrow \hat{w}_{(g_1)(g_2)}^{\text{diff}}(\cdot, \cdot)$ ]. (iv) Evaluating the function of relative differences at the edge positions provides information about contributions to the inferred structural deviation. This assessment is restricted to present or absent edges that are opposed to the formation of equivalent structures.

method are illustrated in Figure 2, where the ASD and the TD group are represented on the left and right, respectively. In all these illustrations, the node coloring refers to the positions inferred through the joint graphon estimation procedure. For comparison reasons, the first row depicts again the functional connectivity networks just as obtained after preprocessing (cf. Figure 6 of the paper). The second row shows the present edges in both networks which form collectives that are exceptional with respect to the structure uncovered from the respective other network. Here, the intensity of the depicted connections represents the magnitude of distinctiveness on an inverse log scale. The most extraordinary absent edges—also with respect to the structure of the respective other network and with magnitude represented by the inversely log-scaled intensity—are visualized at the bottom row. Based on these illustrations, we can conclude that, for example, the interconnection between the dark bluish nodes is much denser in the ASD group than in the TD group. Similar results are revealed for the interconnection between nodes from the green to the yellow color spectrum. In contrast, the connections between the dark orange and the cyan node bundle seem to be much more for the TD group. With regard to the test procedure carried out in the paper, these microscopic differences can be seen as a rough division of the calculated overall discrepancy represented in the form of the test statistic  $t$ .

Taking these analytical results together with the finding from the paper, we can (i) infer that the functional connectivity significantly differs between the ASD and the TD group and (ii) provide information on what these differences are composed of.

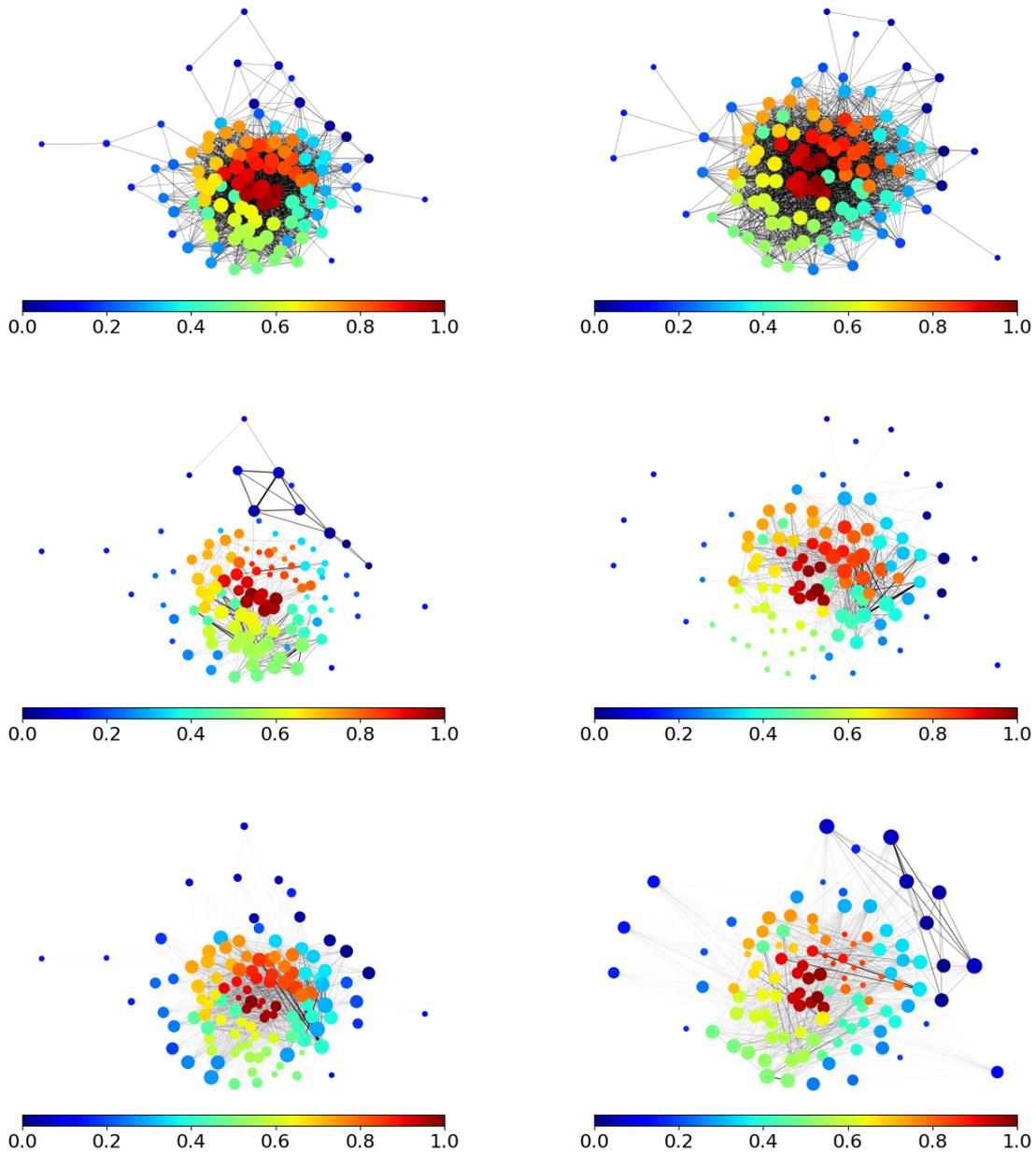


Figure 2: Localization of differences in the functional coactivation within the human brain. The results for the two clinical groups ASD and TD are represented in the left and the right column, respectively. Top: resulting connectivity between the 116 considered brain regions after preprocessing; degree of nodes is illustrated by log-scaled node size. The lower four plots show observed present edges (middle) and absent edges (bottom) that form extraordinary structural patterns with respect to the structure found in the respective other network. The node sizes visualize the nodes' impact (log scale) as aggregation over the impact of attached edges.

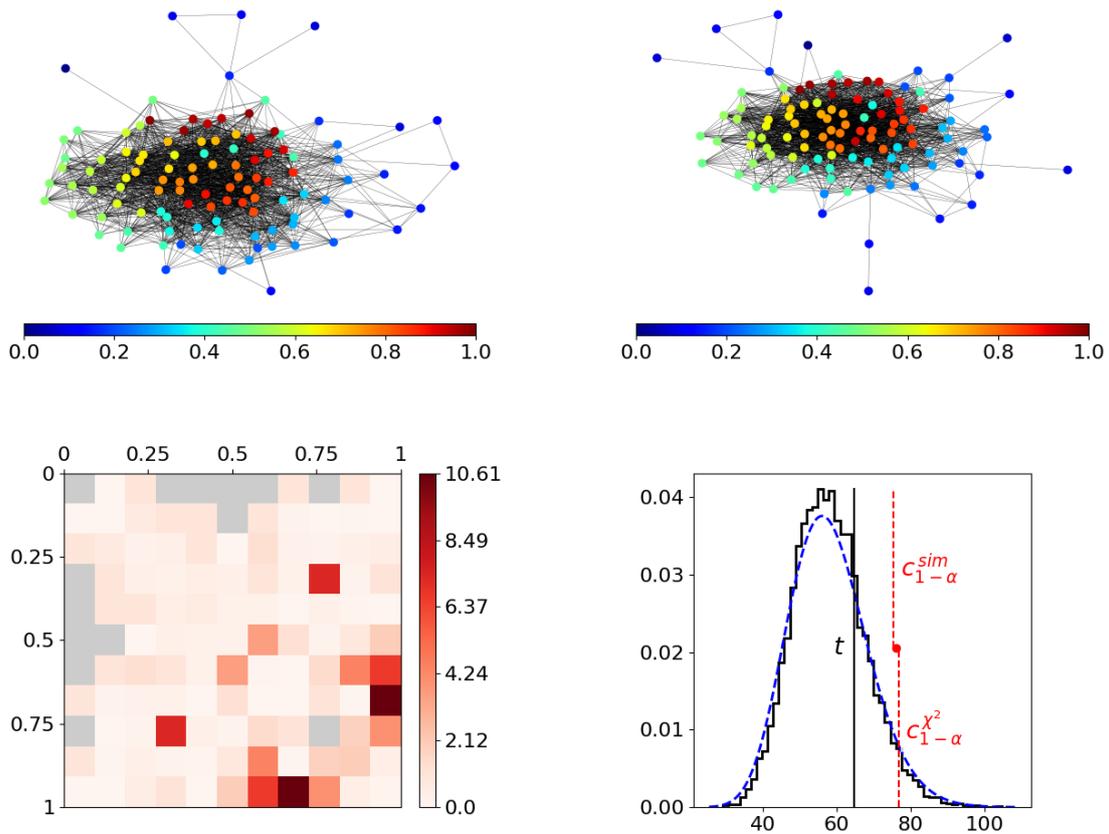


Figure 3: Comparison of the human brain functional coactivation between two subgroups of the TD group. Top: networks of subgroups with coloring referring to the inferred node positions. Bottom left: differences in rectangles according to the construction of test statistic (10) in the paper. Bottom right: result of the test statistic, including the simulated and the theoretical distribution plus their corresponding critical values used for comparison.

## 2 Comparison of Brain Coactivation Between Typical-Development Subgroups

To further demonstrate that the found significant differences in the brain coactivation between the ASD and the TD group are meaningful (see Section 6.2.2 of the paper), we here repeat the analysis for two randomly selected subgroups of the TD group. The results are illustrated in Figure 3. Considering the formation of the two networks in the top row,

inclusive of the nodes' positional embedding found by the algorithm, this seems very similar for the two subgroups. This is also reflected by the differences in rectangles (bottom left) and the final realization of the test statistic (bottom right). Comparing the latter with the corresponding critical value shows that no significant difference on the global scale can be found.

## Contributing Publications

Sischka, B. and Kauermann, G. (2022). EM-Based Smooth Graphon Estimation Using MCMC and Spline-Based Approaches. *Social Networks*, 68, 279–295. doi:10.1016/j.socnet.2021.08.007.

De Nicola, G., Sischka, B., and Kauermann, G. (2022). Mixture Models and Networks: The Stochastic Blockmodel. *Statistical Modelling*, 22(1-2), 67–94. doi:10.1177/1471082X211033169.

Sischka, B. and Kauermann, G. (2023). Stochastic Block Smooth Graphon Model. *arXiv preprint arXiv:2203.13304*. Under review in *Journal of Computational and Graphical Statistics*.

Sischka, B. and Kauermann, G. (2023). Nonparametric Two-Sample Test for Networks Using Joint Graphon Estimation. *arXiv preprint arXiv:2303.16014*. Planned submission (after adequate shortening): *Journal of the American Statistical Association (Applications and Case Studies)*.



# Eidesstattliche Versicherung (Affidavit)

(Siehe Promotionsordnung vom 12. Juli 2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 30. März 2023

---

Benjamin Sischka

