Modeling Unobserved Heterogeneity in Social Network Data Analysis

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

vorgelegt von
Sevag Kevork

Eingereicht am 03.02.2022

Modeling Unobserved Heterogeneity in Social Network Data Analysis

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

vorgelegt von
Sevag Kevork

Eingereicht am 03.02.2022

1. Berichterstatter: Prof. Dr. Göran Kauermann
2. Berichterstatter: Prof. Dr. Christoph Stadtfeld
3. Berichterstatter: Prof. Dr. Helmut Küchenhoff

Tag der Einreichung: 03.02.2022
Tag der Disputation: 27.04.2022

für Nelly und Raffi

Danksagung

Mein besonderer Dank gilt allen, die auf die eine oder andere Weise zur Vollendung dieser Dissertation beigetragen haben:

In erster Linie meinem Betreuer Göran Kauermann, der mich während der letzten Jahre mit Rat und Tat unterstützt und betreut hat.

Helmut Küchenhoff für seine Tätigkeit als Prüfer und dafür, dass er mich am StaBLab angestellt und in das akademische Leben eingeführt hat.

Christoph Stadtfeld für die Bereitschaft, die Aufgabe des externen Prüfers zu übernehmen.

Christian Heumann für die Tätigkeit des Vorsitzenders der Promotionskomission und Volker Schmid dafür, als Ersatzgutachter zur Verfügung zu stehen.

Constanze Schmaling fürs Korrekturlesen, das Teilen unseres Büros und natürlich für die Hilfe und Unterstützung in jeder Hinsicht während der gesamten Zeit.

Michael Windmann fürs gelegentliche Vorbeischauen und für die interessanten Diskussionen über Statistik, Lehre und Politik.

Sowie den vielen Mitarbeiter/innen am Institut, vor allem Brigitte Maxa, Elke Höffner und Pia Förg.

Und schließlich meiner Familie und meinen Eltern für die unendliche Unterstützung.

Zusammenfassung

Die Analyse von Netzwerkdaten hat sich in den letzten Jahren zu einem anspruchsvollen und aufstrebenden Bereich der Statistik entwickelt. In diesem Zusammenhang ist das sogenannte Exponential Random Graph Model (ERGM) ein vielversprechender Ansatz für die Modellierung von Netzwerkdaten. Allerdings erweist sich die Parameterschätzung als schwierig, nicht nur aufgrund von Berechnungs- und Stabilitätsproblemen, insbesondere bei großen Netzwerken, sondern auch wegen des unbeobachteten Vorhandenseins der Knotenheterogenität im Netzwerk.

Diese Dissertation beginnt mit einer allgemeinen Einführung in die Graphentheorie, der eine ausführlichen Diskussion über Exponential Random Graph Modelle und die konventionellen Parameterschätzungsansätze folgt. Zudem werden die Vorteile dieser Modellklasse präsentiert und das Problem der Modelldegeneration diskutiert.

Der erste Beitrag der Dissertation schlägt einen neuen iterativen Schätzansatz für Exponential Random Graph Modelle mit knotenspezifischen Zufallseffekten in unipartiten Netzwerken vor, der die unbeobachtete Knotenheterogenität berücksichtigt und sowohl Maximum-Likelihood- als auch Pseudolikelihood-Schätzverfahren zur Schätzung der Netzwerkstatistiken beziehungsweise der knotenspezifischen Zufallseffekte kombiniert, um eine stabile Parameterschätzung zu gewährleisten. Außerdem wird eine Modellauswahlstrategie entwickelt, um das Vorhandensein von Knotenheterogenität im Netzwerk zu bewerten.

Im zweiten Beitrag wird der Ansatz der iterativen Schätzung auf bipartite Netzwerke ausgedehnt, wobei die Schätz- und Bewertungstechniken erläutert werden. Darüber hinaus wird eine gründliche Untersuchung und Interpretation von zufälligen knotenspezifischen Zufallseffekten in bipartiten Netzwerken für das vorgeschlagene Modell diskutiert.

Simulationsstudien und Datenbeispiele illustrieren beide Beiträge. Alle entwickelten Methoden sind mit der Open-Source-Statistiksoftware `R` implementiert.

Summary

The analysis of network data has become a challenging and growing field in statistics in recent years. In this context, the so-called Exponential Random Graph Model (ERGM) is a promising approach for modeling network data. However, the parameter estimation proves to be demanding, not only because of computational and stability problems, especially in large networks but also because of the unobserved presence of nodal heterogeneity in the network.

This thesis begins with a general introduction to graph theory, followed by a detailed discussion of Exponential Random Graph Models and the conventional parameter estimation approaches. In addition, the advantages of this class of models are presented, and the problem of model degeneracy is discussed.

The first contribution of the thesis proposes a new iterative estimation approach for Exponential Random Graph Models incorporating node-specific random effects that account for unobserved nodal heterogeneity in unipartite networks and combines both maximum likelihood and pseudolikelihood estimation methods for estimating the structural effects and the nodal random effects, respectively, to ensure stable parameter estimation. Furthermore, a model selection strategy is developed to assess the presence of nodal heterogeneity in the network.

In the second contribution, the iterative estimation approach is extended to bipartite networks, explaining the estimation and the evaluation techniques. Furthermore, a thorough investigation and interpretation of nodal random effects in bipartite networks for the proposed model is discussed.

Simulation studies and data examples are provided to illustrate both contributions. All developed methods are implemented using the open-source statistical software `R`.

Contents

1 Introduction

The analysis of network data is of great interest in many scientific fields. Typical examples can be found in social sciences, where friendship networks play an essential role, or in political sciences, where, for instance, international arms trade networks are of importance. In financial mathematics, one tries to achieve an optimal portfolio composition with the help of network analysis, among other things. In computer science, medicine, and biology, statistical methods of network data analysis are also used to investigate networks.
The unique feature of networks is their dependency structure since their observations, also called actors, are related to each other, which is vital in the analysis. Accordingly, networks are not limited to variables that affect individual actors but primarily map their relationships. Thus, the actors in a network cannot be regarded as independent of each other. Ordinary statistical analyses, however, assume that observations are independent of each other. Consequently, alternative methods of statistical modeling are needed for the analysis of networks. Kolaczyk (2009) gives a considerate and thorough introduction to the field of statistical network analysis. The survey articles of Goldenberg et al. (2010), Fienberg (2012), Hunter et al. (2012), and Salter-Townshend et al. (2012), discuss recent statistical approaches, challenges, and advancements in this field.

The Exponential Random Graph Model (ERGM) is one such alternative method suitable for modeling networks. ERGMs are a class of stochastic models that use network local structures and, if available, actor/edge specific covariates to model the formation of an edge for a network with a fixed number of nodes. Lusher et al. (2013) give a general introduction to ERGMs and their properties.

This dissertation is structured as follows: The introductory chapter begins with a general analysis of unipartite and bipartite networks. Section 1.2 introduces some pre-ERGM model specifications, followed by the definition and properties of ERGMs in Section 1.3. Network simulation and a brief introduction to available algorithms are discussed in Section 1.4. The estimation framework in ERGMs is presented in Section 1.5 with a detailed discussion of the estimation methods, followed by a summary demonstrating the advantages and disadvantages of the introduced methods. The degeneracy issue in network data analysis is addressed in Section 1.6; the ways of dealing with this problem are reflected in Section 1.7, where alternative network specifications are presented. A general discussion on ERGMs, challenges, limitations, and drawbacks is given in Section 1.8. Chapter 2 introduces the first

contribution of this thesis, which proposes a new iterative estimation approach for ERGMs, incorporating node-specific random effects that account for unobserved nodal heterogeneity in unipartite networks and combines both maximum likelihood and pseudolikelihood estimation methods to ensure stable parameter estimation. In addition, a model selection strategy is developed to assess the presence of nodal heterogeneity in the network (Kevork and Kauermann, 2021). The iterative estimation idea is extended to bipartite networks in Chapter 3, which is the second contribution of this dissertation, demonstrating the estimation and the evaluation techniques. Furthermore, a thorough investigation and interpretation of nodal random effects in bipartite networks for the proposed model are discussed (Kevork and Kauermann, 2022).

1.1 Analyzing Networks

1.1.1 Analyzing Unipartite Networks

A unipartite or one-mode network is a group or a system of interconnected actors (vertices, nodes), whereby an actor can be many things, such as a person or a city. Therefore, unipartite network data is relational data as it represents the relationships (edges, ties) between the individual actors. Regarding networks in general, one has to distinguish between directed and undirected networks. Undirected networks focus on any relation between two actors, whereas directed networks consider the direction of the relation between two actors. In this dissertation, we focus on undirected unipartite networks since the data examples in the contributed manuscript are of this type.

A graph $G = (V, E)$ is the mathematical description of a unipartite network. It consists of a node-set V and an edge set E. A node embodies an actor of a network. Usually, it is assumed that the number of nodes $N_v = |V|$ is less than infinity. Edge $\{i, j\}$ is an element of the set E and connects two actors. It describes the relationship between the nodes i and j. If an edge exists between the nodes i and j, it denotes a relationship between them (see Kolaczyk 2009).

The structure of a graph is determined by its adjacency matrix. The adjacency matrix Y of a graph is a matrix of $N_v \times N_v$ whose elements Y_{ij} indicate the number of edges between nodes i and j. For a simple graph, these elements are either 0 or 1. If there is an edge between node i and node j, $Y_{ij} = 1$, otherwise $Y_{ij} = 0$ with $i, j \in \{1, \ldots, N_v\}$ and $i \neq j$. For simplicity we assume undirected edges, that is $Y_{ij} = Y_{ji}$. An example of such an adjacency matrix with its associated graph is shown in Figure 1.1. Moreover, node-specific covariates (e.g., gender) or dyad-wise covariates (e.g., indicators for the same gender) can be available. We should bear in mind that the link variables Y_{ij}, which are of interest for $i, j \in \{1, \ldots, n\}$, are treated as random, while the nodes $i = 1, \ldots, n$ are fixed.

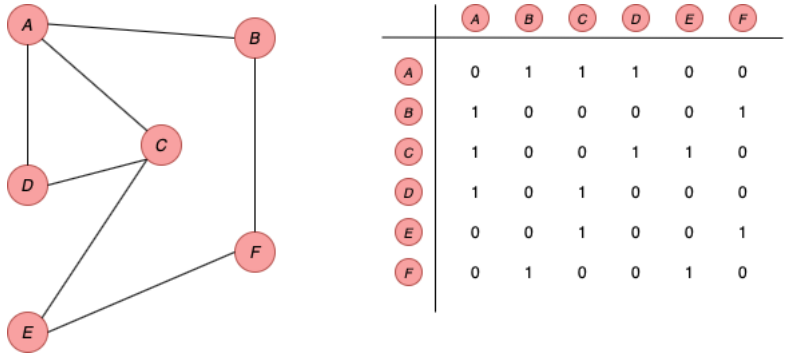


Figure 1.1 *Schematic representation of a unipartite network.*

1.1.2 Analyzing Bipartite Networks

A bipartite network can be represented as a triple $Y = (R, C, E)$ where R and C, also often called first- and second-mode nodes, are two disjoint sets of nodes, and $E \subseteq R \times C$ is the set of edges of the network. This representation can be demonstrated as an $n \times m$ rectangular matrix Y (see Figure 1.2), where n and m are the number of nodes in R (mode-1 nodes) and C (mode-2 nodes) respectively, and with elements $Y_{ij} = 1$ if there is an edge between i and j and $Y_{ij} = 0$ otherwise. Bipartite networks are also known as

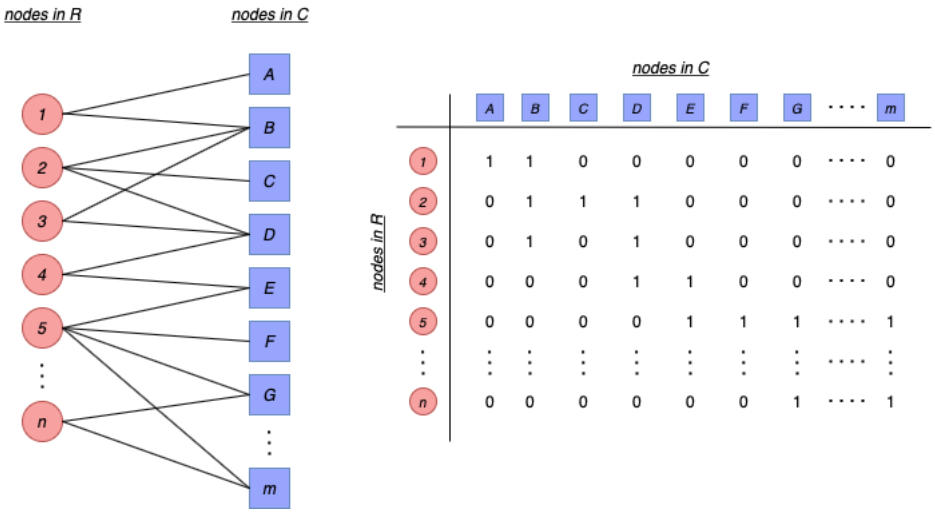


Figure 1.2 *Schematic representation of a bipartite network.*

affiliation or two-mode networks. Some widely used examples of this type of network are citation networks, collaboration networks, actor-film networks, or patent networks. We often encounter two unipartite network representations that are derived from a bipartite network. For instance, in a patent network, one set of nodes represents the patents filed to the patent office, and the other set represents the inventors, with ties representing inventors filing a patent to the patent office. Therefore, an inventor to inventor network can be derived

such that if two inventors share a filed patent, there is an edge between them; a patent to patent network can be constructed in a similar way to form an inventor sharing network. Figure 1.3 illustrates the projection of a toy bipartite patent network example (center) to

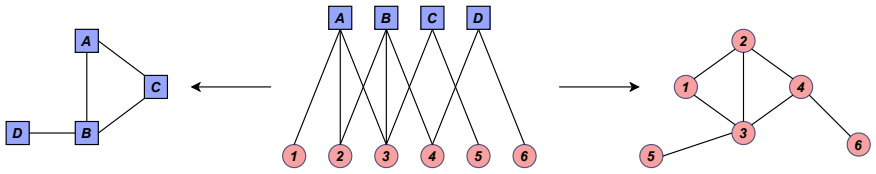


Figure 1.3 *Projection of a toy bipartite patent network example to two unipartite networks.*

a patent to patent network projection (left) and an inventor to inventor network projection (right). Using projection techniques, one may analyze a bipartite network by looking at two unipartite networks. However, by projecting a bipartite network and ignoring the edge values, we lose information about the number of nodes in the other set acting as edges between connected pairs in the unipartite networks. In this dissertation, we seek to analyze unprojected bipartite networks to avoid information loss and focus on the global structures of bipartite networks.

Borgatti and Everett (1997) provide a general overview of the basic ideas of bipartite network analysis and discuss ways of applying and interpreting traditional network analytic techniques to two-mode network data. More recently, Latapy et al. (2008) propose extensions of basic tools for analyzing large one-mode networks (the classical case) to the two-mode case. A general survey of different approaches for bipartite network data is provided in Shi et al. (2017).

1.2 Model Specifications

Social scientists have studied social networks since the 1930s. The sociogram, a mathematical graph in which a group of individuals is represented as nodes and the relationship between pairs of individuals is represented by an edge, was first developed by Moreno and Jennings (1938). The Bernoulli Random Graph Model introduced by Gilbert (1959), also referred to as the simplest ERGM, assumes that each actor (node) in the network has the same probability of forming an edge, where the only network statistic is the count of edges. Mathematically speaking, the probability of an edge between nodes i and j has the following form:

$$\mathbb{P}\Big(Y_{ij} = 1\Big) = p, \quad \forall\, i, j \in \{1, \ldots, n\}, \text{ and } j > i, \text{with } p \in (0, 1). \tag{1.1}$$

However, the independence assumption of this model is too unrealistic or nonsensical for all but the simple of cases. Therefore, this model may be most useful as a "null" model, although it is arguably too simple even for this.
Holland and Leinhardt (1981) addressed the independence assumption for directed networks (digraphs) with their p_1 model, focusing on two empirical observations from sociometric studies, namely reciprocation (mutual relationships) and attractivity/productivity (in- and out-degrees of the nodes). They then constructed a family of distributions with parameters that allow controlling the probability of observing different values of mutual relationships and in-/out-degrees. After some algebraic manipulation, the model can be written as:

$$\operatorname{logit}\Big[\mathbb{P}\Big(Y_{ij}=1\Big)\Big] = \log\left\{\frac{\mathbb{P}\Big(Y_{ij}=1\Big)}{1-\mathbb{P}\Big(Y_{ij}=1\Big)}\right\} = \alpha_i + \alpha_j + \boldsymbol{z}_{ij}^t\boldsymbol{\beta}, \tag{1.2}$$

where α_i and α_j are fixed nodal effects, and $\boldsymbol{z}_{ij}$ is a set of covariates related to i and j.
The early steps of ERGMs are introduced by Frank and Strauss (1986) with their implementation of the Markov dependence assumption (Markov model) in which two dyads are independent, conditional on the rest of the graph if they do not have a node in common. The Markov graph model in an undirected network consists of three configurations, edges, stars, and triangles, which has the following form:

$$\mathbb{P}\Big(Y=y\Big) = \frac{1}{\kappa\Big(\theta,\sigma_k,\tau\Big)}\exp\Big\{\theta E + \sigma_k S_k + \tau T\Big\}, \tag{1.3}$$

where θ, σ_k, and τ are the edge, k-stars and triangle parameters respectively, $E = \sum_{i<j} y_{ij}$ denotes the total number of edges, S_k represents the number of $k-$stars (S_2 would therefore be a 2-star with $S_2 = \sum_{i<j<k} y_{ij}y_{ik}$ representing the total number of 2-stars), $T = \sum_{i<j<k} y_{ij}y_{jk}y_{ik}$ denotes the total number of triangles in a network, and $\kappa(\theta,\sigma_k,\tau)$ is the normalizing constant to ensure that (1.3) is a legitimate probability distribution.
Although the Markov model is a promising approach, the probability for all possible edges across the graph is assumed to be homogeneous since node-specific covariates, such as gender or age, are not incorporated in the model, which limits the utility of the model.
About ten years later, Wasserman and Pattison (1996) extend the Markov model and introduce the so-called $p^\star$ model, which assumes a more general conditional dependence among edges. To be more specific, two edges are conditionally independent if, given all other network edges, the conditional probability that both edges exist does not equal the product of their marginal conditional probabilities (Wasserman and Pattison, 1996). The generalization of this assumption allows ERGMs with most dependence assumptions to be considered $p^\star$ models.

1.3 Definition of Exponential Random Graph Model

The Exponential Random Graph Model describes the probability distribution for all graphs or networks with a fixed number of nodes. Let Y be a matrix-valued random variable representing the adjacency matrix of a network. The probability distribution of the random variable Y can be expressed in the following form:

$$\mathbb{P}\Big(Y = y|\boldsymbol{\theta}\Big) = \frac{\exp\{\boldsymbol{\theta}^T \boldsymbol{s}(y)\}}{\kappa(\boldsymbol{\theta})}, \tag{1.4}$$

where $\boldsymbol{\theta} = \Big(\theta_0, \ldots, \theta_p\Big)^T$ is the vector of model parameters and $\boldsymbol{s}(y) = \Big(s_0(y), \ldots, s_p(y)\Big)^T$ is the vector of network statistics, such as edges, 2-stars, and triangles which can be extended to $\boldsymbol{s}(y, X)$ if covariates X such as gender or age are included in the model, see, e.g., Snijders et al. (2006) for more details. Note that there is no general answer to which network statistics should be included in (1.4) to capture the structural properties, as Figure 1.4 illustrates some of the possible network configurations in an undirected network.

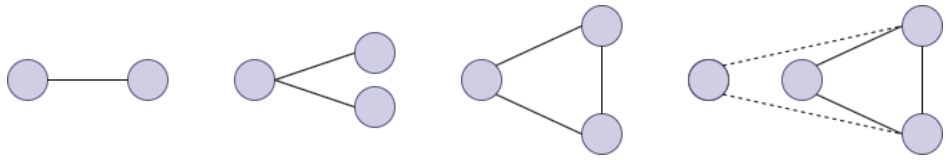


Figure 1.4 *Examples of network statistics: edge, 2-star, triangle, and gwesp.*

The denominator $\kappa(\boldsymbol{\theta})$ in (1.4) represents the normalizing factor to ensure that (1.4) is a legitimate probability mass function

$$\kappa\Big(\boldsymbol{\theta}\Big) = \sum_{\boldsymbol{y} \in \mathcal{Y}} \exp\Big\{\boldsymbol{\theta}^T \boldsymbol{s}(y)\Big\},$$

where $\mathcal{Y}$ is the set of all possible $n \times n$ adjacency matrices. An exact calculation of $\kappa\Big(\boldsymbol{\theta}\Big)$ is problematic except for tiny networks, since the normalizing constant sums over all possible expressions of the random variable Y. For instance, for an undirected network with n nodes, $2^{\frac{n(n-1)}{2}}$ potential networks are possible. With only 15 nodes, there are already $1.57 \cdot 10^{57}$ possible expressions of the random variable Y. Therefore, it is usually infeasible to determine the normalizing constant numerically. Thus, handling the normalization constant becomes the crucial dilemma of the ERGM (Hummel et al., 2012a) because a calculation of the normalizing constant is inevitable for a classical maximum likelihood estimation. Since (1.4) belongs to exponential family type distributions, we use the properties of those types of distributions to overcome the computational burden of $\kappa(\boldsymbol{\theta})$. Eventually, for the

log normalizing constant, we get

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log\Big(\kappa\left(\boldsymbol{\theta}\right)\Big) = \mathbb{E}\Big(\boldsymbol{s}(Y)|\boldsymbol{\theta}\Big) \tag{1.5}$$

which can be approximated using simulation-based approaches (see Snijders 2002). More details on network simulation are given in Section 1.4.

Frequently, one also encounters (1.4) in its conditional form, clarifying also the interpretation of the model parameters, which has the following form:

$$\text{logit}\Big[\mathbb{P}(Y_{ij} = 1|Y_{-ij} = y_{-ij}, \boldsymbol{\theta})\Big] = \boldsymbol{\theta}^T \Delta_{ij}\boldsymbol{s}(y), \tag{1.6}$$

where Y_{ij} is a binary random variable, indicating whether an edge between the nodes i and j exists or not. Y_{-ij} denotes the state of all remaining edges from Y, i.e., all edges except the one between nodes i and j, and $\Delta_{ij}\boldsymbol{s}(y) = \boldsymbol{s}(y_{ij} = 1, y_{-ij}) - \boldsymbol{s}(y_{ij} = 0, y_{-ij})$ is the vector of so-called change statistics. Hence, $\boldsymbol{\theta}$ in (1.6) can be interpreted as the increase in the full conditional log-odds of a connection between actors i and j induced by the formation of the edge, conditional on all other edges remaining unchanged.
To provide a better understanding of the change statistics $\Delta_{ij}\boldsymbol{s}(y)$ in (1.6), we look exemplarily at the number of edges and triangles, where a triangle corresponds to three nodes that are directly connected. Consider the change from a network with no edge between nodes i and j to a network with an edge between them, given that the rest of the network is identical for both. Therefore, the change statistics for the number of edges always takes the value 1 since adding an edge between i and j increases the number of edges by precisely 1. On the other hand, the triangle change corresponds to the number of two-paths between i and j; for instance, in a friendship network, this is the number of mutual friends of i and j given the rest of the network; the two red nodes in Figure 1.5 represent nodes i and j. To determine the change statistics for the number of triangles, we need to determine how many new triangles are created by adding the red dashed edge to the network; in this case, this would be exactly three. Note that some network statistics

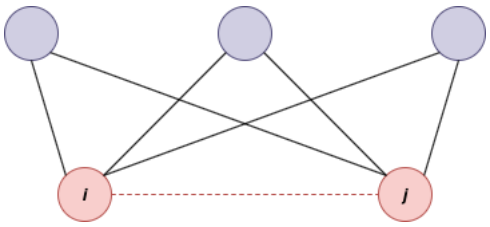


Figure 1.5 *Visualization of change statistics for the number of triangles.*

are much harder to interpret. An overview with some examples is discussed in Section 1.7. More detailed discussions on properties of Exponential Random Graph Models are presented in Robins et al. (2007a), Robins et al. (2007b), and Lusher et al. (2013).

1.4 Simulating Random Networks

Network simulation is an essential ingredient in the inference of ERGMs. Although the normalizing constant in (1.4) is usually intractable, network simulation in ERGMs can be handled easily, since evaluating the conditional probability of an edge, given the rest of the network, the normalizing constant cancels out. The Markov Chain Monte Carlo algorithm (Metropolis et al. 1953; Hastings 1970) allows generating networks efficiently. The goal is to generate a sequence of M networks $y^{(0)}, y^{(1)}, \ldots, y^{(m)}, \ldots, y^{(M)}$ from the target distribution, that are updated sequentially by small changes; see also Hunter et al. (2008a) and Lusher et al. (2013). Convergence is ensured on the Markov chain with a large M, resulting in a draw from the target distribution. Some common approaches for network simulation are presented below.

1.4.1 Gibbs Sampling

The Gibbs sampling approach begins by selecting a pair of nodes (i, j) uniformly at random from all possible pairs in the network. Given (1.6), we can define the conditional distribution of Y_{ij} given the rest of the network Y_{-ij} as follows

$$\mathbb{P}(Y_{ij} = 1 | Y_{-ij} = y_{-ij}, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\}}{1 + \exp\{\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\}} .$$

We then set Y_{ij} either to 1 or 0 according to the conditional probabilities $\mathbb{P}(Y_{ij} = 1 | Y_{-ij} = y_{-ij}, \boldsymbol{\theta})$ and $\mathbb{P}(Y_{ij} = 0 | Y_{-ij} = y_{-ij}, \boldsymbol{\theta})$, respectively, as described in Hunter et al. (2008b). We repeat this process until the desired number of networks is obtained.

1.4.2 Metropolis Algorithm

In the Metropolis algorithm approach, as the Gibbs sampling approach, we select a pair of nodes (i, j) uniformly at random from all possible pairs in the network. Then we calculate the acceptance ratio π as

$$\pi = \frac{\mathbb{P}(Y_{ij}^{\text{changes}} | Y_{-ij} = y_{-ij})}{\mathbb{P}(Y_{ij}^{\text{stays the same}} | Y_{-ij} = y_{-ij})} = \exp\{\pm \boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\}.$$

If the current state of $Y_{ij} = 1$, then the acceptance ratio $\pi = \frac{\mathbb{P}(Y_{ij}=0 | Y_{-ij}=y_{-ij})}{\mathbb{P}(Y_{ij}=1 | Y_{-ij}=y_{-ij})} = \exp\{-\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\}$, whereas if the current state $Y_{ij} = 0$, then the acceptance ratio $\pi = \frac{\mathbb{P}(Y_{ij}=1 | Y_{-ij}=y_{-ij})}{\mathbb{P}(Y_{ij}=0 | Y_{-ij}=y_{-ij})} = \exp\{\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\}$. We accept the change of Y_{ij} with probability $\min\{1, \pi\}$. Finally, this process is repeated until the desired number of networks is obtained

(see Hunter et al. 2008b).

In both approaches described above, a dyad, which is simply pair of nodes connected or not, is uniformly selected at random to be toggled. However, in sparse networks, edges exist only between a small proportion of all possible dyads. Selecting a pair of nodes uniformly at random therefore suggests switching an edge rather than removing an existing one. The low density in sparse networks will cause the Markov chains to be stuck in the same state for many iterations.

1.4.3 TNT (Tie/No Tie) Sampler

The TNT (tie/no tie) sampler, introduced by Morris et al. (2008), is one way to overcome this problem by selecting connected dyads more frequently than would be the case with uniform random selection. Thus, the TNT sampler moves more efficiently across the space of possible networks than the other two methods mentioned above.

The TNT sampler is a Metropolis-Hastings algorithm implemented in the `ergm` package, which begins by choosing a connected dyad with a probability $\omega = 0.5$ to toggle. We calculate the Hastings factor ρ based on the current state of the selected dyad $(i,\ j)$. If $y_{ij} = 1$ is chosen,

$$\rho = \begin{cases} \frac{1}{\omega \times \text{ndyads} + (1-\omega)} & \text{if nedges} = 1 \\ 1 + \left(\frac{\omega}{1-\omega}\right)\left(\frac{\text{ndyads}}{\text{nedges}+1}\right) & \text{if nedges} > 1 \end{cases}$$

and if $y_{ij} = 0$ is chosen,

$$\rho = \begin{cases} \omega \times \text{ndyads} + (1-\omega) & \text{if nedges} = 0 \\ 1 + \left(\frac{\omega}{1-\omega}\right)\left(\frac{\text{ndyads}}{\text{nedges}+1}\right) & \text{if nedges} > 0 \end{cases}$$

with ndyads denoting the number of all possible dyads in the network and nedges the total number of edges in the network in the current state. We then calculate the acceptance probability π as

$$\pi = \begin{cases} \exp\{-\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\} \times \rho & \text{if } \ y_{ij} = 1 \\ \exp\{\boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)\} \times \rho & \text{if } \ y_{ij} = 0 \end{cases}$$

and we accept the change of Y_{ij} with probability $\min\{1, \pi\}$. Subsequently, we repeat this process until the desired number of networks is obtained.

After a sufficiently large number of iterations, the result is a random network, which can be seen as a draw from the target distribution of an ERGM. All previous draws are called

burn-in and are disregarded. This burn-in is necessary so that the Markov chain no longer depends on its arbitrarily chosen initial state because only with a sufficiently large burn-in is a drawn network independent of the starting point. Each subsequently drawn network can also be considered as a draw from the target distribution, which is why multiple random networks can be drawn from the same chain. However, two networks drawn directly after each other are highly correlated since, at most, one edge in the network is changed in each iteration. Therefore, one omits t iterations between each of the drawn networks. The value of t is called "thinning" and should be as large as possible to minimize the autocorrelation between the drawn networks; see, e.g., Koskinen and Snijders (2013) and Lusher et al. (2013).

1.5 Estimation of Exponential Random Graph Models

For an ERGM as given in (1.4), the maximum likelihood estimator (MLE) for the vector of model parameters $\boldsymbol{\theta}$ is defined as $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\, l(\boldsymbol{\theta})$, where $l(\boldsymbol{\theta})$ is the log-likelihood

$$l(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{s}(y) - \log(\kappa(\boldsymbol{\theta})). \tag{1.7}$$

However, such direct maximization of the log-likelihood is essentially impossible due to the intractable normalization constant $\kappa(\boldsymbol{\theta})$ in the log-likelihood.

As a result, several different approaches for model parameter estimation in the context of ERGMs are introduced. The major challenge for all alternative methods mentioned below is to find a solution for dealing with the normalization constant. We refer to Hunter et al. (2012) and Koskinen and Snijders (2013) for more detailed discussion.

1.5.1 Maximum Pseudolikelihood Estimation

Probably the most straightforward and least computationally expensive approach to estimating the ERGM parameters is the maximum pseudolikelihood estimation (MPLE) presented in Strauss and Ikeda (1990). This approach is based on the so-called pseudolikelihood, which is defined as follows:

$$\prod_{\substack{i,j \in \{1,2,\ldots,N_v\} \\ j>i}} \mathbb{P}(Y_{ij} = y_{ij} | Y_{-ij} = y_{-ij}, \boldsymbol{\theta}).$$

Accordingly, for each potential edge of the network the conditional probability is calculated given the rest of the network. The product of these individual conditional probabilities yields the pseudolikelihood. The maximum pseudolikelihood estimator maximizes the

following log pseudolikelihood

$$l_{\text{pseudo}}(\boldsymbol{\theta}) = \sum_{\substack{i,j\in\{1,2,\ldots,N_v\}\\ j>i}} \log[\mathbb{P}(Y_{ij} = y_{ij}|Y_{-ij} = y_{-ij}, \boldsymbol{\theta})]\,.$$

Hence, the maximum pseudolikelihood estimator is defined as:

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\, l_{\text{pseudo}}(\boldsymbol{\theta}). \tag{1.8}$$

The maximization of the log pseudolikelihood can be done by simple logistic regression, as shown by Strauss and Ikeda (1990). The advantage of MPLE in the context of ERGMs is that even more complicated models can be fitted without significant problems and with comparatively little computational effort, as described in Robins et al. (2007a).

Note that logistic regression assumes that the observations (nodes) are independent of each other. However, this assumption is fundamentally not given for network data as these are relational. Consequently, the estimates of the parameters in a MPLE may be biased because the dependence structure of the data is ignored. Only if the variables are conditionally independent for the existence of an edge Y_{ij}, the maximum pseudolikelihood estimator is unbiased. If even only weak dependence structures are present, the estimates could provide reasonably good approximations for the maximum likelihood estimator; see, for instance, Kolaczyk (2009). Subsequently, if, on the other hand, strong dependence structures are present in the network, the maximum pseudolikelihood estimator may be significantly biased.

1.5.2 Markov Chain Monte Carlo Maximum Likelihood Estimation

Maximizing the true log-likelihood (1.7) is, of course, preferable to maximizing the pseudolikelihood since the behavior of the MPLE is hard to evaluate (Hummel et al. 2012b). However, due to the normalization constant, the maximization of the log-likelihood is impossible. Hunter and Handcock (2006) introduced the Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMC-MLE) procedure to tackle this problem, which originates from Geyer and Thompson (1992), where an arbitrary parameter $\boldsymbol{\theta}_0$ is specified, and then the following log-likelihood formulation is considered:

$$\begin{aligned} l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0) &= \boldsymbol{\theta}^T \boldsymbol{s}(y) - \log(\kappa(\boldsymbol{\theta})) - [\boldsymbol{\theta}_0^T \boldsymbol{s}(y) - \log(\kappa(\boldsymbol{\theta}_0))] \\ &= (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(y) - \log\left\{\frac{\kappa(\boldsymbol{\theta})}{\kappa(\boldsymbol{\theta}_0)}\right\} \\ &= (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(y) - \log(\mathbb{E}_{\boldsymbol{\theta}_0}[\exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(Y)\}]) \end{aligned} \tag{1.9}$$

The maximum pseudolikelihood estimator from (1.8) is suitable as an initial value for $\boldsymbol{\theta}_0$. As described in Hunter and Handcock (2006), the log-likelihood (1.9) can be exploited by generating random networks $Y^{(0)}, Y^{(1)}, \ldots, Y^{(m)}$ from the target distribution by the MCMC algorithm to approximate $l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0)$ as follows:

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0) \approx (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(y) - \log \left[\frac{1}{m} \sum_{i=1}^{m} \exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(Y^{(i)})\} \right] \tag{1.10}$$

and for $m \to \infty$ the law of large numbers guarantees the convergence of the expectation term in (1.9) to

$$\frac{1}{m} \sum_{i=1}^{m} \exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{s}(Y^{(i)})\}.$$

Therefore, by simulating m random graphs, the true parameter can be approximated by the sample mean, and thus an approximation for the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is obtained.

As mentioned in Geyer and Thompson (1992), the log-likelihood approximation in (1.9) performs best if $\boldsymbol{\theta}_0$ is close to the true parameter vector $\boldsymbol{\theta}$. On the other hand, if $\boldsymbol{\theta}_0$ is not close enough to $\boldsymbol{\theta}$, this approach yields only inaccurate or even no estimates. This problem can be alleviated by applying the so-called Importance Sampling according to Geyer and Thompson (1992). However, this approach also relies on how well $\boldsymbol{\theta}_0$ is chosen; see, e.g., Hummel et al. (2012b).

1.5.3 Stepping Algorithm

The so-called stepping algorithm, introduced by Hummel et al. (2012a), offers a solution to the significant problem of the MCMC-MLE procedure. As already noted, the MCMC-MLE procedure only provides reasonable estimates if $\boldsymbol{\theta}_0$ is close to the true MLE.

Therefore, the goal of the stepping algorithm is to gradually approximate $\boldsymbol{\theta}_0$ to the unknown MLE $\boldsymbol{\theta}$. Hummel et al. (2012a) define the ERGM parameterization in (1.4) as a canonical parameterization with a mean-value

$$\xi(\boldsymbol{\theta}) = \mathbb{E}(\boldsymbol{s}(y)|\boldsymbol{\theta})$$

to approach the MLE stepwise by pretending that the MLE is not $\boldsymbol{s}(y)$, but some random point between $\boldsymbol{s}(y)$ and the estimate of $\xi(\boldsymbol{\theta}_0)$, which we call pseudo-observation. In doing so, the stepping algorithm repeatedly switches stepwise from a canonical parameterization to a mean-value parameterization, and vice versa. To be more specific, the stepping algorithm consists of several steps. Step one is executed only at the start of the very first run of the

stepping algorithm, where the MPLE is usually used as the initial value for $\boldsymbol{\theta}_t$, with $t = 0$. In step two, random networks $Y^{(0)}, Y^{(1)}, \ldots, Y^{(m)}$ are generated from the target distribution by the MCMC algorithm, calculating the sample mean $\bar{\xi}_t = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{s}(Y^{(i)})$. Consequently, a new pseudo-observation $\hat{\xi}_t$ is determined taking the parameters $\boldsymbol{s}(y)$, $\bar{\xi}_t$, and γ_t into account

$$\hat{\xi}_t = \gamma_t \boldsymbol{s}(y) + (1 - \gamma_t)\bar{\xi}_t \quad \text{with } \gamma_t \in (0, 1]\,.$$

We choose γ_t as close to 1 as possible while not leaving the convex hull of $\boldsymbol{s}(Y^{(1)}), \ldots, \boldsymbol{s}(Y^{(m)})$. This modification allows the stepping algorithm to counteract degeneracy and be stable, whereas other algorithms may be unstable due to potential degeneracy. Finally, the parameter vector $\boldsymbol{\theta}_{t+1}$ is determined by updating (1.10) and we get

$$\boldsymbol{\theta}_{t+1} = \text{argmax}_{\boldsymbol{\theta}}\Big\{(\boldsymbol{\theta} - \boldsymbol{\theta}_t)^T \hat{\xi}_t - \log\Big[\frac{1}{m}\sum_{i=1}^{m} \exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_t)^T \boldsymbol{s}(Y^{(i)})\}\Big]\Big\}.$$

We set $t = t + 1$ and repeat this procedure. However, as soon as it is possible for $\gamma_t = 1$ without leaving the convex hull, the stepping algorithm is considered to be converged.

1.5.4 Stochastic Approximation (Robbins-Monro)

Another way to circumvent the problem of the intractable normalization constant is the so-called stochastic approximation, as suggested by Snijders (2002), which utilizes the Robbins-Monro algorithm (Robbins and Monro, 1951).

This approach aims to center the distribution of the simulated networks' statistics over the observed network's statistics. The distribution of the statistics of the simulated networks is considered centered if $\boldsymbol{s}(y) - \mathbb{E}(\boldsymbol{s}(Y)|\boldsymbol{\theta}) = 0$. Note that, taking the derivative of the log-likelihood in (1.7) along with property (1.5), we obtain the same result

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \boldsymbol{s}(y) - \frac{\partial}{\partial \boldsymbol{\theta}} \log(\kappa(\boldsymbol{\theta})) \\
&= \boldsymbol{s}(y) - \frac{\partial}{\partial \boldsymbol{\theta}} \log\Big(\sum_{\boldsymbol{y} \in \mathcal{Y}} \exp\big\{\boldsymbol{\theta}^T \boldsymbol{s}(y)\big\}\Big) \\
&= \boldsymbol{s}(y) - \sum_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{s}(y) \cdot \mathbb{P}(y|\boldsymbol{\theta}) \\
&= \boldsymbol{s}(y) - \mathbb{E}(\boldsymbol{s}(Y)|\boldsymbol{\theta}).
\end{aligned}$$

We can now exploit $\frac{\partial}{\partial \boldsymbol{\theta}} \log(\kappa(\boldsymbol{\theta}))$ to approximate the score function by simulating random graphs from the target distribution. To do this, we draw an MCMC-based single random network $y^{(t)}$ from the target distribution for each iterative update step t. Accordingly, the

network $y^{(t)}$ is based on the current estimate of $\boldsymbol{\theta}_t$. Furthermore, as in the MCMC-MLE procedure, an initial value must be specified, which can be obtained using, for instance, the MPLE. We now generate a sequence of parameters $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ using a Newton-Raphson-type procedure. An update of the sequence from $\boldsymbol{\theta}_t$ to $\boldsymbol{\theta}_{t+1}$ is done by

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - a_t \mathbf{D}_0^{-1}\Big(\boldsymbol{s}(y^{(t)}) - \boldsymbol{s}(y)\Big),$$

where $\mathbf{D}_0$ is a scaling matrix determined in an initialization phase as the diagonal of an estimated covariance matrix of the network statistics based on network samples using $\boldsymbol{\theta}_0$. The term a_t denotes the so-called gain factor, which controls the size of the update steps. The gain factor is a positive value, which becomes smaller in each iteration step and eventually converges to 0 as discussed in Snijders (2002). Thus, the changes within the sequence become smaller and smaller as $\boldsymbol{s}(y^{(t)})$ approaches $\boldsymbol{s}(y^{(t+1)})$ and a_t converges to 0. Consequently, if $\boldsymbol{\theta}_{t+1}$ is the true MLE, then for a sufficiently large sample of networks we obtain $\mathbb{E}(\boldsymbol{s}(Y)|\boldsymbol{\theta}_{t+1}) = \boldsymbol{s}(y)$. A more detailed discussion on the stochastic approximation is provided in Koskinen and Snijders (2013).

1.5.5 Summary of Estimation Methods

Table 1.1 summarizes the advantages and disadvantages of the estimation methods discussed above. The question of which estimation method is preferable is difficult to answer. We conclude that the performance of simulation-based estimation techniques depends on whether and how well the ERGM captures the structure of the observed network. In the numerous simulation studies, we notice that the ERGM generally faces convergence issues as nodal heterogeneity in the network increases; more on this topic in Section 1.8. That being said, MCMC-MLE is the method most likely to experience

Table 1.1 *Summary of the estimation methods in ERGMs.*

Methods	Advantages	Disadvantages
MPLE	+ Simplicity + Speed, via logistic regression	- Maximizes the pseudo-likelihood, not the likelihood
MCMC-MLE	+ Theoretical guarantee of convergence to MLE	- Highly sensitive to starting point - May require enormous MC samples - May require several iterations
Stepping Algorithm	+ Theoretical guarantee of convergence to MLE + Degeneracy counteraction & stability	- Sensitive to starting point - Requires setup expertise - May require several iterations
Stochastic Approximation	+ Theoretical guarantee of convergence to MLE + Simple updates	- May converge too slowly to be practical - Non-practical implementation of convergence conditions - Requires trial-and-error calibration

convergence problems. These problems occur less frequently with the other simulation-based estimation methods, whereby it tends to be the stepping algorithm for which we observe convergence problems least frequently. Therefore, in our proposed iterative estimation algorithms introduced in Chapters 2 and 3, we use the stepping algorithm as part of our algorithm for estimating the network parameters.

1.6 Degeneracy

A useful stochastic model should, among other things, place most of its probability mass on those networks that have a high probability of being generated by the actual underlying emergence process. On the other hand, if a model allocats an inappropriately large proportion of its probability mass on only a very few of the possible networks, it is called a degenerate model, see, for instance, Snijders et al. (2006), Schweinberger (2011), and Chatterjee and Diaconis (2013). More specifically, in this case a large part of the probability mass is often placed on unrealistic networks such as empty or full networks. While some real-world networks may be nearly empty or nearly full, they are unlikely to be the subject of network analysis, as discussed in Handcock (2003).

Furthermore, some models show a so-called near-degenerate behavior. These models place a large portion of their probability mass on either full, empty, or a mixture of both networks. A near-degenerate model is where $\mathbb{E}(\boldsymbol{s}(Y)|\boldsymbol{\theta})$ lies near the edge of the convex hull of all possible network statistics. In an Exponential Random Graph framework many introductory models are usually appealing because of the easy-to-interpret effects, e.g., models containing only 2-stars or triangles; however, these models often lead to degeneracy. For such models, adding a single edge can enormously increase the value of the change statistics for other edges, causing a chain reaction.

Therefore, most ERGMs are prone to the problem of degeneracy (Snijders et al. 2006, Schweinberger 2011), which leads to a very poor approximation of the true likelihood. This is mainly because degenerate models are unstable, and even minimal changes in parameter values can cause significant changes in the simulated networks. Moreover, degenerate models may prevent the convergence of the simulation-based estimation methods. In such a case, the simulation-based estimation methods fail to find the MLE. Consequently, this can disrupt the algorithm's convergence in that full or empty networks are generated during the simulation of random networks.

Accordingly, stable and far-from-degeneracy models are desired. A model is stable if small changes in parameter values cause only minor changes in the probabilistic structure of the model. In contrast, even minimal changes in parameter values can lead to significant differences in network structures for an unstable model. Thus, very similar parameter values may represent significantly different network structures. Therefore, to mitigate the

degeneracy problem by avoiding large changes in triangles and k-stars ($k \geq 2$) statistics, Snijders et al. (2006) propose a set of new alternative specifications. Hunter and Handcock (2006), Goodreau (2007), Robins et al. (2007a) and Robins et al. (2007b) provide further discussions on the new specifications in the ERGM framework.

1.7 Specifications of Network Statistics

As mentioned before, there is no blanket answer to which statistics must be included to obtain a good model. Among other things, this depends heavily on the nature and structure of the network.

However, a vital statistic in ERGMs is the number of edges in a network. This statistic mimics the role of the intercept, and gives the basic probability of an edge formation in the network. Since the change statistic for the number of edges is always equal to 1, the associated parameter affects each network similarly. The k-stars parameter in the model accounts for the number of k-stars in a network. A k-star is a node that has an edge to k other nodes. It is worth noting that k-star(1) in an undirected network corresponds to the edge statistic. Since all k-stars up to size $(n-1)$ can be modeled by this parameter, it can also be regarded as a parameter for degree distribution. The transitivity effect in the network is captured by the triangle parameter corresponding to three directly connected nodes. Moreover, as already noted, in ERGMs, we can also include explanatory covariates as exogenous network statistics. There are several possibilities to include these covariates, depending, for instance, on the research question and the scale level. For example, suppose we are interested in studying social homophily, i.e., the tendency to be more likely to form friendships with individuals similar to oneself. In that case, the similarity of the actors concerning criteria such as gender, ethnicity, or education level could be of interest.

Nevertheless, the traditional network statistics, such as k-stars, or triangles, despite their easy-to-interpret effects, lead to degeneracy issues, as discussed in Section 1.6. Although the new specifications, such as alternating k-stars, alternating k-triangles, and alternating k-two-paths, handle the degeneracy problem and stabilize the model, their interpretation is challenging.

The following subsections provide an overview of the new specifications for ERGMs in unipartite networks and their extensions to bipartite networks. Pattison and Robins (2002) and Morris et al. (2008) provide a more detailed discussion on this matter.

1.7.1 Alternating Statistics

Alternating k-stars

In ERGMs, we can model k-stars up to size $(n-1)$, which leads to a model putting large weights on big stars or on nodes with a high degree, resulting in model degeneracy. The number of k-stars in a network can be expressed as follows:

$$S_k(y) = \sum_{i=1}^{n} \binom{y_{i+}}{k},$$

where y_{i+} is the degree of node i. The new specification, alternating k-star statistics, short, AKS, introduces a weight parameter λ, $\lambda \geq 1$, which numbs the effect of large changes in the statistics of large stars. This is achieved by alternating the signs of the weights of the k-star statistics so that negative weights of odd-k-star statistics balance the positive weights of even-k-star statistics. This is expressed as:

$$\begin{aligned} AKS &= S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \ldots + (-1)^{n-2}\frac{S_{n-1}}{\lambda^{n-3}} \\ &= \sum_{k=2}^{n-1}(-1)^k \frac{S_k}{\lambda^{k-2}}\,. \end{aligned}$$

As can be seen, the parameter λ is a decreasing weight of the k-star statistics, and this prevents particularly assigning dense networks a higher probability. The alternating k-star statistics are intended for modeling the degree distribution. A positive parameter value for this statistic suggests a skewed degree distribution. On the other hand, a negative parameter value suggests a homogeneous degree distribution.

For a bipartite network Y of size $(n \times m)$, where n and m denote the number of nodes in R (first mode) and C (second mode), respectively, the alternating k-star statistics are defined as:

$$\begin{aligned} AKS_R &= \sum_{k=2}^{n-1}(-1)^k \frac{S_{R_k}}{\lambda^{k-2}} \\ AKS_C &= \sum_{k=2}^{m-1}(-1)^k \frac{S_{C_k}}{\lambda^{k-2}}\,, \end{aligned}$$

which have the same properties as the alternating k-stars in unipartite networks.

Alternating k-triangles

According to Snijders et al. (2006), using statistics such as the triangle statistic, it is nearly impossible to model networks with a high degree of transitivity but still have a relatively low density. In a friendship network, a high degree of transitivity would mean that forming a friendship between two individuals who have several friends in common is more likely than forming a friendship between two individuals with no friends in common. However, networks with a rather low density despite a high degree of transitivity can often be observed.

Usually, triangles in social networks form cliques, where many triangles are formed within a small subgroup of nodes. The new specification of alternating k-triangles is therefore defined as k triangles sharing a common edge $\{i, j\}$, as demonstrated in Figure 1.6. The

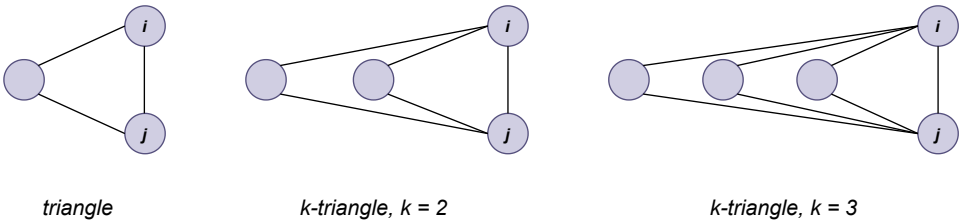


Figure 1.6 *Examples of alternating k-triangles.*

number of k-triangles for $k \geq 2$ can be expressed as

$$T_k = \sum_{i<j} y_{ij} \binom{L_{2ij}}{k},$$

where

$$L_{2ij} = \sum_{h \neq i,j} y_{ih} y_{hj}$$

corresponds to the number of two-paths between i and j. Furthermore, it is evident that a $k = 3$-triangle contains three $k = 2$-triangles. Hence, a $k = 1$-triangle corresponds to a triangle. The number of $k = 1$-triangles in a network is given by:

$$T_1 = \frac{1}{3} \sum_{i<j} y_{ij} L_{2ij} \,.$$

For the alternating k-triangle statistic, decreasing weights with alternating signs are used, as in the case of the alternating k-stars statistic:

$$\begin{aligned} AKT &= 3T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} - \ldots + (-1)^{n-3}\frac{T_{n-2}}{\lambda^{n-3}} \\ &= \sum_{i<j} y_{ij} \sum_{k=1}^{n-2} \left(\frac{-1}{\lambda}\right)^{k-1} \binom{L_{2ij}}{k} \\ &= \lambda \sum_{i<j} y_{ij} \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}. \end{aligned}$$

Triangles cannot be formed in bipartite networks as edges within the two modes are not defined. Therefore, the alternating k-triangles statistic cannot be applied. The most minor local closure in a bipartite network is a four-cycle, a two-path closed by another two-path. A more detailed discussion is given below.

Alternating k-two-paths

A two-path can be considered as the same as a 2-star; consequently, four nodes with two-paths form a four-cycle or 2-two-path. Therefore, a k-two-paths statistic is a pair of nodes $(i,\ j)$ with k shared neighbors, as illustrated in Figure 1.7. The alternating k-two-paths statistic principle is the same as the alternating statistics presented earlier. Accordingly,

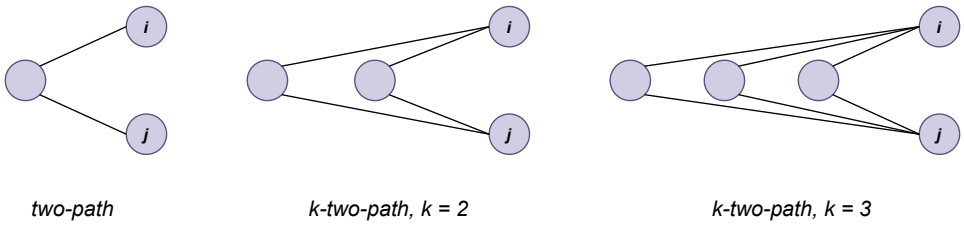


Figure 1.7 *Examples of alternating k-two-paths.*

the number of k-two-paths can be expressed as follows:

$$U_k = \begin{cases} \sum_{i<j} \binom{L_{2ij}}{k} & \text{if } k > 2 \\ \frac{1}{2} \sum_{i<j} \binom{L_{2ij}}{2} & \text{if } k = 2 \text{ due to symmetry}. \end{cases}$$

The resulting alternating k-two-paths statistic can be determined by

$$AKTP = U_1 - \frac{2U_2}{\lambda} + \sum_{k=3}^{n-2} \left(\frac{-1}{\lambda}\right)^{k-1} U_k$$
$$= \lambda \sum_{i<j} \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}.$$

When $\lambda = 1$, this statistic diminishes to the number of dyads that are indirectly connected by at least one two-path.

$$AKTP = \sum_{i<j} I_{L_{2ij}>0} \,,$$

where I is a binary indicator function.

The most minor closure, which is not a dyad, is a four-path in a bipartite network. Therefore, the parameter value for alternating k-two-paths in bipartite networks indicates the likelihood of forming a social circuit. As bipartite networks have two disjoint sets of nodes R and C, two different k-two-paths structures or closures can be formed, as shown in Figur 1.8. The circle and square nodes indicate the nodes in R and C, respectively. In

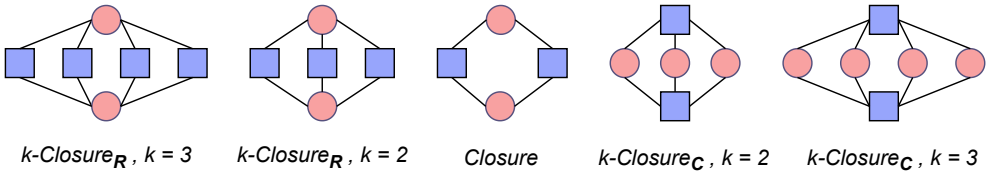


Figure 1.8 *Examples of alternating k-two-paths in bipartite networks.*

bipartite networks, L_{2ij} defines the number of two-paths between nodes i and j, where both belong to the same set of nodes. Accordingly, the alternating k-two-paths statistics can be expressed as

$$AKTP_R = \lambda \sum_{i<j} \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}$$
$$AKTP_C = \lambda \sum_{i<j} \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}.$$

1.7.2 Geometrically Weighted Statistics

Hunter (2007) proposed alternative representations for the alternating statistics and implemented them in the `ergm` package (Hunter et al., 2008a). However, from the modeling

point of view, these proposed representations are equivalent to the alternating statistics (Goodreau, 2007; Hunter, 2007). For the alternating k-stars, Hunter (2007) introduced the geometrically weighted degree (GWD) statistic

$$e^{\theta_s} \sum_{i=1}^{n-1} \left\{ 1 - \left(1 - e^{-\theta_s}\right)^i \right\} \mathrm{D}_i(y),$$

where $\mathrm{D}_i(y)$ denotes the number of nodes with degree i, and the parameter θ_s is a constant, i.e., usually set to a fixed value. For the alternating k-triangles and alternating k-two-paths, geometrically weighted edgewise shared partners (GWESP)

$$e^{\theta_t} \sum_{i=1}^{n-2} \left\{ 1 - \left(1 - e^{-\theta_t}\right)^i \right\} \mathrm{EP}_i(y),$$

and geometrically weighted dyadic shared partners (GWDSP) statistics are introduced, respectively

$$e^{\theta_p} \sum_{i=1}^{n-2} \left\{ 1 - \left(1 - e^{-\theta_p}\right)^i \right\} \mathrm{DP}_i(y),$$

where $\mathrm{EP}_i(y)$ denotes the number of edges with i shared partners, and $\mathrm{DP}_i(y)$ denotes the number of dyads with i shared partners. The parameters θ_t and θ_p are, as before, some fixed values. All these statistics use an exponential down-weighting of the icorporated counts. However, these statistics stabilize the entire model fitting framework, their interpretation proves to be challenging.

The geometrically weighted statistics except for the GWESP statistic are also available for bipartite networks, which are defined as the alternating statistics for each mode of a bipartite network, and are implemented in the `ergm` package.

1.8 Discussion

Conventional ERGMs are static, i.e., suitable for cross-sectional networks. This means that if the underlying network generating process is dynamic, ERGMs cannot be applied. However, extensions are available for modeling longitudinal/dynamic networks, see, e.g., Snijders et al. (2010a), and for a general overview, Snijders and Koskinen (2013). Hanneke et al. (2010), for example, introduced the time-discrete temporal ERGM (TERGM), postulating an exponential family model for the transition probability from a network at time t to a network at time $t + 1$, in which the network at time t is a single draw from an ERGM conditional on the network at time $t - 1$. Krivitsky and Handcock (2014) proposed a separable version of TERGM (STERGM), assuming that the processes of edge formation and dissolution are separable from each other within a time step. Another widely

used modeling approach for longitudinal networks is the stochastic-actor-oriented model (SAOM); see, e.g., Snijders (2001) and Snijders et al. (2010b), which are not edge-oriented like the ERGMs, but focus on the edge changes during two-time steps. However, dynamic extensions will not be discussed any further since the focus of this dissertation is on static ERGMs in both unipartite and bipartite networks.

Another limitation of ERGMs is that they are restricted to modeling only binary edge variables. For instance, in social networks, if we are interested in the strength of social relationships, which can be seen as a function of their duration and intimacy, rather than just whether there is a relationship between two actors, this may cause information loss. Desmarais and Cranmer (2012), Krivitsky (2012) and Krivitsky and Butts (2015) propose some possible approaches for dealing with such valued networks.

As already noted, modeling network data proves to be particularly challenging due to the prevailing complex dependency structure in many real networks. One of the modeling challenges is the assumption of nodal homogeneity. In ERGMs, two networks have the same probability if they coincide in $\boldsymbol{s}(y)$, the vector of network statistics. This means that exogenous covariates, if available, explain all possible node-specific heterogeneity in the network. This assumption is more than questionable. For instance, friendships between individuals in social networks are driven by many factors. Some individuals tend to attract more connections than others, and such phenomena cannot be fully explained by covariates like age and gender of the individuals.

Thiemichen et al. (2016) proposed a Bayesian estimation framework incorporating node-specific heterogeneity to account for nodal heterogeneity in networks; however, it proved to be infeasible for large networks, i.e., networks with more than 100 nodes. This approach forms the basis for our extensions. Note that incorporating an n dimensional vector of random node-specific coefficients $\boldsymbol{u} = (u_1, u_2, \ldots)$ in ERGM, with $\boldsymbol{u} \sim \mathcal{N}(0, \sigma_u^2 \boldsymbol{\mathcal{I}}_n)$, where σ_u^2 is the variance and $\boldsymbol{\mathcal{I}}_n$ the n dimensional identity matrix, leads to a mixed model with fixed and random coefficients. Therefore, the conditional model for a single edge Y_{ij} conditional on the rest of the network Y_{-ij} can be defined as

$$\operatorname{logit}\Big[\mathbb{P}(Y_{ij} = 1 | Y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u})\Big] = \boldsymbol{\theta}^T \, \Delta_{ij} \, \boldsymbol{s}(y) \; + u_i + u_j \; . \tag{1.11}$$

The terms u_i and u_j account for unobserved nodal heterogeneity in the network not captured by $\Delta_{ij}\boldsymbol{s}(y)$, the vector of change statistics. Assuming normality for $\boldsymbol{u}$, the conditional model (1.11) resembles a mixed logistic regression model. In Chapter 2, we dive in more depth into this strategy and introduce an iterative estimation approach for the fixed and random coefficients. Although time-consuming in some cases, the iterative estimation approach ensures stable estimates. The scalability of our approach is demonstrated in Chapter 3, where the iterative estimation approach is extended for bipartite networks, where $\boldsymbol{u} = (u_1, u_2, \ldots)$ and $\boldsymbol{v} = (v_1, v_2, \ldots)$, with $\boldsymbol{u} \sim \mathcal{N}(0, \sigma_u^2 \boldsymbol{\mathcal{I}}_n)$, and $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2 \boldsymbol{\mathcal{I}}_m)$,

are n and m dimensional vectors of random node-specific coefficients for first- and second-mode nodes, respectively, resulting in a conditional model defined as

$$\text{logit}\Big[\mathbb{P}(Y_{ij} = 1|Y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v})\Big] = \boldsymbol{\theta}^T \, \Delta_{ij} \, \boldsymbol{s}(y) \, + u_i + v_j \; .$$

The connection between the predicted nodal random effects and nodal degrees is apparent in our approaches, as demonstrated in Chapters 2 and 3. Hence, the nodal random effects' estimated variance can be interpreted as the unobserved nodal heterogeneity level of the underlying network. Our approach and the model extensions are explained and discussed in detail in the following two chapters.

2 Iterative Estimation of Mixed Exponential Random Graph Models with Nodal Random Effects

Contributed Manuscript

Kevork S. and Kauermann G. (2021). Iterative estimation of mixed exponential random graph models with nodal random effects. *Network Science*, 9(4):478–498.
`https://doi.org/10.1017/nws.2021.22`

Copyright

Author Contributions

The iterative estimation idea of the structural effects and the nodal random effects in the framework of Exponential Random Graph Models came from Göran Kauermann. The model selection idea based on AIC came from Sevag Kevork. Sevag Kevork wrote the algorithmic implementation, conducted the simulation studies, and performed the data analysis. The manuscript was written by Sevag Kevork, and Göran Kauermann. All authors were involved in proofreading the manuscript.

Software

All computations and plots in this chapter have been produced using `R` version 3.6.0 with packages `ergm` 3.7.1, `mgcv` 1.8-31, `network` 1.15, `Matrix` 1.3-4, `GGally` 2.0.0, and `matrixStats` 0.58.0.

The iterative estimation algorithm and the model comparison algorithm are available in the package `mergm` on GitHub (`https://github.com/kevorks/mergm`).

Network Science **9** (4): 478–498, 2022
doi:10.1017/nws.2021.22

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Iterative estimation of mixed exponential random graph models with nodal random effects

S. Kevork* and G. Kauermann

Institut für Statistik, Ludwig-Maximilians-Universität München, München, Germany.
*Corresponding author. Email: sevag.kevork@stat.uni-muenchen.de

Action Editor: Stanley Wasserman

Abstract

The presence of unobserved node-specific heterogeneity in exponential random graph models (ERGM) is a general concern, both with respect to model validity as well as estimation instability. We, therefore, include node-specific random effects in the ERGM that account for unobserved heterogeneity in the network. This leads to a mixed model with parametric as well as random coefficients, labelled as mixed ERGM. Estimation is carried out by iterating between approximate pseudolikelihood estimation for the random effects and maximum likelihood estimation for the remaining parameters in the model. This approach provides a stable algorithm, which allows to fit nodal heterogeneity effects even for large scale networks. We also propose model selection based on the Akaike Information Criterion to check for node-specific heterogeneity.

Keywords: exponential random graph models; random effects; generalized linear mixed models; network data analysis

1. Introduction

The analysis of network data has become a challenging and emerging field in statistics in the last years. Goldenberg et al. (2010), Hunter et al. (2012) and Fienberg (2012) provide comprehensive articles on statistical approaches, challenges and developments in network data analysis. We also refer to Kolaczyk (2009) and Kolaczyk and Csárdi (2014) for a general introduction and the related routines for network data analysis in R. In this study, we concentrate on exponential random graph models (ERGM) originally introduced in Frank and Strauss (1986) and more deeply discussed e.g. in Lusher et al. (2013). Unless node-specific covariates are included in the ERGM, the probability for all possible edges across the graph is assumed to be homogeneous, which also means that any permutation of the node labels will yield the same probability. This is a questionable assumption, in particular in large networks, which also contributes towards stability problems for estimation as discussed e.g. in Schweinberger (2011) or Schweinberger et al. (2017). We follow Thiemichen et al. (2016) and extend the ERGM by incorporating node-specific heterogeneity effects to overcome the homogeneity assumption of ERGM and capture the unobserved heterogeneity in the network.

Consider a network of n actors for which some dyadic relationships have been recorded. These relations can be represented in the form of an $n \times n$ adjacency matrix Y, with elements $Y_{ij} = 1$ if an edge from i to j exists and $Y_{ij} = 0$ otherwise. In undirected networks, we have $Y_{ij} = Y_{ji} \; \forall \, i \neq j$, which we assume throughout this paper. We consider the probability of

observing a given network conditional on a set of (sufficient) network statistics is given by the ERGM

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}\big) = \frac{\exp\{\boldsymbol{\theta}^T\, \boldsymbol{s}(y)\}}{\kappa(\boldsymbol{\theta})} \tag{1}$$

Here, $\boldsymbol{s}(y)$ is the vector of network statistics and $\boldsymbol{\theta}$ is the vector of model coefficients. Vector $\boldsymbol{s}(.)$ includes any structural characteristics of the network and we refer to Snijders et al. (2006) for a general discussion on network statistics, see also Hunter and Handcock (2006). The denominator $\kappa(\boldsymbol{\theta})$ in (1) represents the normalizing factor to ensure that (1) is a legitimate probability mass function. In general, $\kappa(\boldsymbol{\theta})$ is numerically intractable unless for miniature networks. Estimation of $\boldsymbol{\theta}$ in model (1) needs therefore to be carried out simulation based. An early reference for estimation of ERGMs is Snijders (2002). For a general discussion, we refer to Hunter et al. (2012). A numerical stable routine has been proposed in Hummel et al. (2012) using a so-called stepping algorithm. Bayesian estimation is proposed in Caimo and Friel (2011). An important property resulting from equation (1) is that $\boldsymbol{s}(y)$ is a vector of sufficient statistics for the network. This means, two networks, which coincide in $\boldsymbol{s}(y)$, have the same probability. In particular, this means that all possible node-specific heterogeneity in the network is explained by exogenous effects, which may also be included in model (1). This can be seen as questionable assumption. For instance, in a friendship network, we may believe that the formation of friendships (edges) between individuals (nodes) is driven by many factors, observable as well as unobservable. We may suspect that there are quantities, intangible factors specific to each individual (node) that are difficult if not impossible to measure. It seems therefore, advisable to include node-specific heterogeneity to capture possible unobserved heterogeneity of the nodes.

An early model that incorporates node-specific heterogeneity is the so-called p_1 model proposed in Holland and Leinhardt (1981). The model includes parametric sender and receiver effects but no network statistics (except of reciprocity). Random nodal heterogeneity was proposed by Duijn et al. (2004) or Zijlstra et al. (2006) which led to the so-called p_2 model. Thiemichen et al. (2016) combined the approach with ERGMs and proposed Bayesian estimation, which however is infeasible for large networks, i.e. networks with more than about 100 actors. We follow the approach of Thiemichen et al. (2016) and extended the ERGM towards

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}, \boldsymbol{u}\big) = \frac{\exp\big\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{t}(y)\big\}}{\kappa\big(\boldsymbol{\theta}, \boldsymbol{u}\big)} \tag{2}$$

where $\boldsymbol{s}(y) = (s_0(y), s_1(y), \dots)$ is, as above, the p dimensional vector of network statistics with $s_0(y) = \sum_i \sum_{j>i} y_{ij}$ as intercept and $\boldsymbol{t}(y) = \big(\sum_{j\neq 1} y_{1j},\ \sum_{j\neq 2} y_{2j}, \dots\big)$ is the n dimensional vector of node degrees. The normalization now equals $\kappa(\boldsymbol{\theta}, \boldsymbol{u}) = \log \sum_{y\in\mathcal{Y}} \exp\big(\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{t}(y)\big)$, where $\mathcal{Y}$ is the set of n by n networks. Conditional on $\boldsymbol{u}$, we obtain node-specific heterogeneity, which can be seen as follows. We assume now that $\boldsymbol{\theta} \in \mathbb{R}^p$ is a p dimensional parameter vector while $\boldsymbol{u} = (u_1, u_2, \dots)$ is a n dimensional vector of random node-specific coefficients with

$$\boldsymbol{u} \sim \mathcal{N}(0,\ \sigma_u^2\, \mathcal{I}_n) \tag{3}$$

with σ_u^2 as variance and $\mathcal{I}_n$ as n dimensional identity matrix. This leads to a mixed model with fixed and random coefficients, termed in the following as mixed ERGM, or in short mERGM. The reasoning behind the model structure can be seen through the conditional model for a single edge Y_{ij} conditional on the rest of the network denoted as Y_{-ij}. From (2), we obtain

$$\log\left\{\frac{\mathbb{P}(Y_{ij} = 1|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u})}{\mathbb{P}(Y_{ij} = 0|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u})}\right\} = \boldsymbol{\theta}^T\ \Delta_{ij}\, \boldsymbol{s}(y)\ + u_i + u_j \tag{4}$$

Here, $\Delta_{ij}\boldsymbol{s}(y) = \boldsymbol{s}(y_{ij} = 1, y_{-ij}) - \boldsymbol{s}(y_{ij} = 0, y_{-ij})$ is the so-called change statistics where y_{-ij} is the network except of edge y_{ij}. The terms u_i and u_j are the random node-specific coefficients accounting for heterogeneity not captured with $\Delta_{ij}\boldsymbol{s}(y)$. If we assume normality for coefficients u_i, formula (4) resembles a mixed logistic regression model as extensively discussed e.g. in Breslow and Clayton (1993). A similar model to (4) has been proposed by Box-Steffensmeier et al. (2018) taking the coefficients $\boldsymbol{u}$ in (4) as random normal variable with mean zero and unknown variance. For estimation, they apply a pseudolikelihood approach. Though this circumvents the numerical burden of estimation in ERGM, it comes for the price of biased estimation of the paramteric coefficients $\boldsymbol{\theta}$. In other words, their estimation approach is biased even if the random node effects are close or equal to zero. We refer to Schmid and Desmarais (2017) for a general discussion on pseudolikelihood estimation in ERGMs, see also Strauss and Ikeda (1990) or Desmarais and Cranmer (2012).

Model (2) combined with the probability model (3) can also be seen as special case of the exponential-family random network model as proposed in Fellows and Handcock (2012). They propose simulation-based estimation, which is restricted to small network sizes. The restriction to small networks is also pointed out in Thiemichen et al. (2016). Hence, even though model (2) can be considered as ERGM with some prior structure on coefficients $\boldsymbol{u}$ given in (2), estimation based on simulation based methods (see Snijders 2002) becomes infeasible for larger networks. We therefore, propose a different estimation strategy, which is motivated in the next paragraph. At this point, we also note that several approaches to capture unobserved heterogeneity were introduced in network data analysis, for instance, Koskinen (2009) introduced binary latent class indicators, Schweinberger and Handcock (2015) examined local dependence using a Bayesian framework in random graph models and Henry et al. (2020) developed a modeling framework to capture unobserved heterogeneity in the effects of nodal covariates.

The first goal of this paper is to provide an iterative estimation strategy, combining both maximum likelihood and pseudolikelihood estimation techniques. To be more specific, we take model (4) as starting point and make use of pseudolikelihood estimation for the random coefficients $\boldsymbol{u}$ while for estimation of $\boldsymbol{\theta}$ we use the steplength MCMC-MLE approach proposed in Hummel et al. (2012) and the corresponding stepping algorithm Hummel et al. (2012), implemented in the `ergm` package in R (see Hunter et al. 2008). These two steps are used iteratively, leading to feasible estimation. Our estimation strategy allows us to fit large scale networks, and we observe that the inclusion of the nodal effects works towards numerical estimation stability, as demonstrated through examples and simulations. Moreover, as the second goal of this paper, we propose a simple model selection strategy to evaluate nodal heterogeneity. To be specific, we use Akaike's Information Criterion (see Akaike 1974) to select a model with or without nodal effects. The latter is calculated numerically by employing a Laplace approximation.

This paper is organized as follows. In Section 2, we discuss the estimation for the underlying model in detail and introduce our algorithm. Furthermore, in Section 3, we present a simulation-based method for evaluating our model, which allows us to calculate the AIC value for the mERGM and compare it with the AIC value of a corresponding ERGM. In Section 4, we present a simulation-based study with the corresponding results, in Section 5, we then apply our approach to three data examples. Finally, Section 6 closes with a discussion.

2. Model and estimation

We consider model (2) where the nodal heterogeneity effects $u_1, \ldots, u_n$ are assumed to be random with σ_u^2 as variance and $\mathcal{I}_n$ as n dimensional identity matrix. The aim is to fit parameter $\boldsymbol{\theta}$ taking nodal heterogeneity into account. Moreover, we need to estimate σ_u^2, which in fact quantifies the amount of nodal heterogeneity.

In principle, we need to maximize the marginal log-likelihood

$$l(\boldsymbol{\theta},\ \sigma_u^2) = \log \int \frac{\exp\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{t}(y)\}}{\kappa(\boldsymbol{\theta}, \boldsymbol{u})} \cdot \frac{1}{(2\pi\sigma_u^2)^{\frac{n}{2}}} \cdot \exp\Big(-\frac{1}{2}\,\frac{\boldsymbol{u}^T\boldsymbol{u}}{\sigma_u^2}\Big) \prod_{i=1}^{n} du_i$$

$$= \boldsymbol{\theta}^T \boldsymbol{s}(y) - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_u^2) + \log \int \exp\Big(g(\boldsymbol{u}, \boldsymbol{\theta}, \sigma_u^2)\Big) \prod_{i=1}^{n} du_i \tag{5}$$

where

$$g(\boldsymbol{u}, \boldsymbol{\theta}, \sigma_u^2) = \boldsymbol{u}^T \boldsymbol{t}(y) - \log(\kappa(\boldsymbol{\theta}, \boldsymbol{u})) - \frac{1}{2} \cdot \frac{\boldsymbol{u}^T\boldsymbol{u}}{\sigma_u^2} \tag{6}$$

We may approximate the integral in (5) using a Laplace approximation. Let therefore $\hat{u}$ be the maximizer of $g(\boldsymbol{u}, \boldsymbol{\theta}, \sigma_u^2)$, which apparently depends on both, $\boldsymbol{\theta}$ and σ_u^2. This leads to the approximate log-likelihood

$$l(\boldsymbol{\theta}, \sigma_u^2) \approx \boldsymbol{\theta}^T \boldsymbol{s}(y) + \hat{u}^T \boldsymbol{t}(y) - \log(\kappa(\boldsymbol{\theta}, \hat{u})) - \frac{1}{2} \cdot \frac{\hat{u}^T\hat{u}}{\sigma_u^2} - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_u^2)$$

$$- \frac{1}{2}\log\left|\frac{\partial^2 g(\hat{u}, \boldsymbol{\theta}, \sigma_u^2)}{\partial u\, \partial u}\right| \tag{7}$$

Note that the likelihood can be considered as profile likelihood, where $\boldsymbol{u}$ is "estimated" through maximizing (6). If we now treat $\hat{u}$ as given, then maximization of $l(\boldsymbol{\theta}, \sigma_u^2)$ with respect to $\boldsymbol{\theta}$ corresponds to maximizing the likelihood of the probability model

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}, \hat{u}\big) = \frac{\exp\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \hat{u}^T \boldsymbol{t}(y)\}}{\kappa(\boldsymbol{\theta}, \hat{u})}$$

where $\hat{u}^T \boldsymbol{t}(y)$ is fixed as given offset. The terminology offset means here that the quantity $\hat{u}^T \boldsymbol{t}(y)$ is treated as fixed. In other words, setting the random coefficients to $\hat{u}$ simplifies the estimation of $\boldsymbol{\theta}$ to Maximum Likelihood estimation in an ERGM with offset $\hat{u}^T \boldsymbol{t}(y)$. This is numerically available with standard software packages (e.g. Hunter et al. 2008) and the stepping algorithm proposed in Hummel et al. (2012). Let us therefore first look in more detail how to obtain $\hat{u}$ if we keep $\boldsymbol{\theta}$ as fixed. Note that $\hat{u}$ results by solving

$$\frac{\partial\, g(\boldsymbol{u}, \boldsymbol{\theta}, \sigma_u^2)}{\partial \boldsymbol{u}} = 0$$

Apparently, this is numerically problematic, since $\kappa(\boldsymbol{\theta}, \boldsymbol{u})$ is numerically intractable. Differentiation yields

$$\frac{\partial \log \kappa(\boldsymbol{\theta}, \boldsymbol{u})}{\partial \boldsymbol{u}} = \mathbb{E}(\boldsymbol{t}(y)|\boldsymbol{u})$$

which in principle can be approximated using simulation based approaches (see Snijders 2002). However, this is a numerically challenging task, since $\boldsymbol{u}$ is high dimensional, namely n dimensional. We therefore propose to approximate the estimation step of $\boldsymbol{u}$ by pseudolikelihood estimation. To do so, we look at the model for a single edge given the rest of the network. That is to say we take model (4), but now we fix $o_{ij} := \boldsymbol{\theta}^T\, \Delta_{ij}\, \boldsymbol{s}(y)$ as offset and ignore the dependence on

y. This leads to a pseudolikelihood approach as discussed for instance in van Duijn et al. (2009). The pseudo-log-likelihood thereby equals

$$l_{\text{pseudo}}(\boldsymbol{u}) = \sum_i \sum_{j>i} y_{ij}\{o_{ij} + u_i + u_j\} - \log\{1 + \exp(o_{ij} + u_i + u_j)\} \tag{8}$$

where a priori coefficient $\boldsymbol{u}$ follows a normal distribution as given in (3). The pseudo-log-likelihood (8) together with (3) leads to (pseudo) generalized linear mixed effects model, so that we can borrow estimation strategies from this field. In particular, we make use of Breslow and Clayton (1993) who propose to approximate the resulting marginal likelihood using a Laplace approximation, similar to (7) above. This allows to estimate the a priori variance σ_u^2 and predict the random coefficients $\boldsymbol{u}$. Note, if o_{ij} is independent of y_{-ij}, then the Laplace approach (7) is equal to the estimation proposed in Breslow and Clayton (1993). Apparently, this is the case for the so-called p_2 model (see Duijn et al. 2004). To be specific, the Laplace approximated pseudolikelihood with $o_{ij} = (o_{1\,2}, o_{1\,3}, \ldots, o_{n-1\,n})$ as offset and constant terms omitted results through

$$l_{\text{pseudo}}(\sigma_u^2) \approx -\frac{1}{2} \cdot \frac{\check{u}^T \check{u}}{\sigma_u^2} - \frac{n}{2} \log(\sigma_u^2) \tag{9}$$

where $\check{u}$ is the maximizer of

$$\check{g}(\boldsymbol{u}, \sigma_u^2; o) = \boldsymbol{u}^T \boldsymbol{t}(y) - \sum_i \sum_j \log\{1 + \exp(o_{ij} + u_i + u_j)\} - \frac{1}{2} \cdot \frac{\boldsymbol{u}^T \boldsymbol{u}}{\sigma_u^2}$$

We call $\check{u}$ an estimate subsequently, even though of course it is a predictor given that $\boldsymbol{u}$ is considered as random. This is implemented in multiple R packages, see e.g. Faraway (2016). To be specific, for estimation of $\boldsymbol{u}$ we use the `mgcv` package (see Wood 2011). Denoting with $\check{\boldsymbol{u}} = (\check{u}_1, \ldots, \check{u}_n)$ the resulting estimates, we set $\check{\boldsymbol{u}}^T \boldsymbol{t}(y)$ in (2) as offset and estimate parameter $\boldsymbol{\theta}$ using simulation based techniques. For this step we use the `ergm` package (see Hunter et al. 2008). Both estimation steps are used iteratively until convergence. That is we take the current estimate $\hat{\boldsymbol{\theta}}_{(t)}$ and update $\check{\boldsymbol{u}}$ with pseudolikelihood leading to $\check{\boldsymbol{u}}_{(t+1)}$. This in turn allows to update $\hat{\boldsymbol{\theta}}$ after replacing the offset by $\check{\boldsymbol{u}}_{(t+1)}^T \boldsymbol{t}(y)$. Our algorithmic steps work in detail as follows:

Algorithm 1: Fit ERGM with nodal random effect components

Step 0: Obtain a prediction for $\boldsymbol{u}$ and estimate σ_u^2:
- a. Fit the model `logit` $\mathbb{P}(y_{ij} = 1 | y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}) = \theta_{(0)} + u_i + u_j$ to the data, where $1 \le i < j \le n$
- b. extract the vector of the predicted random effects $\check{u}_{(0)}$ as offset and set $t = 0$

Step 1: Estimate $\boldsymbol{\theta}$ with ERGM and take $\check{u}_{(t)}^T \boldsymbol{t}(y)$ as an offset parameter:
- a. Fit the model $\mathbb{P}(Y = y | \boldsymbol{\theta}) \propto \exp\{\boldsymbol{\theta}_{(t+1)}^T \boldsymbol{s}(y) + \underbrace{\check{u}_{(t)}^T \boldsymbol{t}(y)}_{\text{offset}}\}$ using maximum likelihood and simulation based methods
- b. extract $o_{ij} = \boldsymbol{\theta}_{(t+1)}^T \Delta_{ij} \boldsymbol{s}(y)$ as new offset for $1 \le i < j \le n$

Step 2: Update $\check{u}_{(t+1)}$ and $\hat{\sigma}_{u(t+1)}^2$ now taking o_{ij} as offset parameter:
- a. Fit the model `logit` $\mathbb{P}(y_{ij} = 1 | y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}) = o_{ij} + u_i + u_j$ with priori structure (3) to the data
- b. extract the vector of the predicted random effects $\check{u}_{(t+1)}$ as new offset

Set $t = t + 1$ and iterate between step 1 and 2 until $\boldsymbol{\theta}_{(t)}$ converges. Convergence is achieved if $|\boldsymbol{\theta}_{(t)} - \boldsymbol{\theta}_{(t+1)}| \leq \epsilon = 0.05$.

We need to mention that pseudolikelihood estimation in network data analysis is biased, in particular, if dyadic statistics are involved. Our setting here, however, is slightly different since we use pseudolikelihood estimation only for the degree statistics $\boldsymbol{t}(y)$. Moreover, we consider the coefficients $\boldsymbol{u}$ not to be fixed but random so that in general, we are more interested in the variance of coefficients u_i and less interested in their true value. We refer to Schmid and Desmarais (2017) or Cranmer and Desmarais (2011) for further discussion on pseudolikelihood estimation in network data analysis.

3. Inference through model selection and variance estimation

The central question in network data analysis is to explain the dominating factors in the network, i.e. the sufficient statistics describing the network structure. If we allow for node-specific heterogeneity, we are additionally faced with the problem of model selection. In other words, we need to describe whether the network data at hand shows evidence with respect to heterogeneous nodes or whether the homogeneity assumption of ERGM seems valid. We tackle this question by approximate calculation of the Akaike Information Criterion (AIC). To do so, we assume for simplicity that the determinant component in (7) depends only weakly on $\boldsymbol{\theta}$ so that we can ignore it subsequently. This is in line with the arguments proposed in Breslow and Clayton (1993) who suggest the use of Laplace approximation in generalized linear mixed models. Note that

$$\frac{\partial^2 g(\boldsymbol{u}, \boldsymbol{\theta}, \sigma_u^2)}{\partial \boldsymbol{u}\, \partial \boldsymbol{u}^{\boldsymbol{T}}} = -Var(\boldsymbol{t}(Y)\,|\boldsymbol{\theta}, \boldsymbol{u}) - \frac{1}{\hat{\sigma}_u^2}\, \mathcal{I}_n \tag{10}$$

Hence, ignoring the determinant component in (7) is justified if we assume that the variance matrix of the degree vector $\boldsymbol{t}(Y)$ depends only weakly on $\boldsymbol{\theta}$. We refer to Breslow and Clayton (1993) for a deeper discussion and motivation which justifies to pursue this simplification. Generally, the variance is difficult to calculate or even infeasible for large networks, so that we make use of simulations to estimate (10). To do so, we simulate networks in order to obtain a simulation based approximation for $\kappa(\hat{\boldsymbol{\theta}}, \check{u})$. We make further use of the simulated networks to obtain a simulation based approximation of $Var(\boldsymbol{t}(y)|\hat{\boldsymbol{\theta}}, \check{u})$. To be specific, let $y^{\star(1)}, \ldots, y^{\star(N)}$ be a set of (independent) network simulations derived from model (2) with $\boldsymbol{\theta}$ set to $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{u}$ set to $\check{u}$. We estimate $Var(\boldsymbol{t}(y)\,|\,\hat{\boldsymbol{\theta}},\ \check{\boldsymbol{u}})$ through

$$\frac{1}{N} \sum_{j=1}^{N} \left[\boldsymbol{t}(y^{\star(j)}) - \bar{t}(y^{\star})\right] \left[\boldsymbol{t}(y^{\star(j)}) - \bar{t}(y^{\star})\right]^T$$

where $\bar{t}(y^{\star})$ is the arithmetic mean of the simulated values.

With these prerequisites, we can now approximate all quantities in (7). This also holds for the normalization constant, which is estimated through

$$\hat{\kappa}(\hat{\boldsymbol{\theta}}, \check{u}) = \frac{1}{N} \sum_{j=1}^{N} \exp\left(\hat{\boldsymbol{\theta}}^T \boldsymbol{s}(y^{\star(j)}) \ + \ \check{u}^T \boldsymbol{t}(y^{\star(j)})\right)$$

Model comparison can now be carried out with the AIC. Setting p as the number of parameters in $\boldsymbol{\theta}$ the AIC results to

$$AIC_{mERGM} = -2\, l(\hat{\boldsymbol{\theta}}, \hat{\sigma}_u^2) \ + \ 2 \cdot (p + 1) \tag{11}$$

Note that (11) resembles the marginal AIC, that is after integrating out $\boldsymbol{u}$. We refer to Greven and Kneib (2010) or Vaida and Blanchard (2005) for a deeper discussion of applying AIC

in random effects models. In our case, formula (11) serves as approximation, relying on the pseudolikelihood estimation for $\boldsymbol{u}$.

We compare the AIC in the mERGM to the resulting AIC in the case of node homogeneity, that is by setting $\sigma_u^2 = 0$. This is carried out in a similar way, but we set $\boldsymbol{u} = 0$.

In other words, we use the likelihood in (1) by calculating $\kappa(\boldsymbol{\theta})$ simulation based from N draws $y^{\star(1)}, \ldots, y^{\star(N)}$ from model (1) with $\boldsymbol{\theta}$ set to the ML estimate in model (1). We call this

$$AIC_{ERGM} = -2\, l_{ERGM}(\hat{\boldsymbol{\theta}}) \; + \; 2p \tag{12}$$

where l_{ERGM} is the log-likelihood in the ERGM resulting when $\boldsymbol{u} \equiv 0$.

Though the focus of the paper lies on model comparison, we shortly discuss how to calculate the variance of the estimate if the algorithm above is used. In the ERGM, we obtain the Fisher information matrix

$$I(\boldsymbol{\theta}) = Var(\boldsymbol{s}(Y)|\boldsymbol{\theta}) \tag{13}$$

This can be estimated simulation based, that is we simulate from model (1) and calculate $Var(\boldsymbol{s}(Y))$ based on the simulated values. For the mERGM, we need to take into account that coefficients $\boldsymbol{u}$ are considered to be random so that in principle, we need to calculate the (inverse) Fisher information of the log-likelihood. Assuming that the determinant in (7) does depend only weekly on $\boldsymbol{\theta}$ and ignoring for simplicity the dependence of $\breve{u}$ on $\boldsymbol{\theta}$ we obtain again (13). This is, of course, an approximation since we ignored the dependence of $\breve{u}$ and $\boldsymbol{\theta}$ as well as estimation variability of σ_u^2. In other words, exact variance calculation in the mERGM is complicated and we here provide a rough approximation only. However, inference can be carried out by model selection via the AIC, which is what we pursue in simulations and data examples below. Still, we can make use of the simulations from above to obtain an estimate of the Fisher information and hence a variance estimate for the estimates.

4. Simulation study

In our simulation study, we want to explore the estimation results of the model parameters and the model selection step based on two network sizes. Small networks with 50 nodes, and large networks with 500 nodes. For each network size, we use network settings with different levels of nodal heterogeneity σ_u^2. For each network setting and network size, we simulate 50 networks using the simulation routines from the `ergm` package (Hunter et al., 2008). Each network setting has the same structural effects $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\theta_{edges},\ \theta_{gwesp},\ \theta_{2-stars}) = (-1,\ 0.2,\ -0.3)$, but the nodal heterogeneity takes six different levels $\sigma_u^2 = (0,\ 0.1,\ 0.2,\ 0.5,\ 0.8,\ 1)$, where $\boldsymbol{u}$ is randomly drawn for each simulated network and setting from a normal distribution following (3). For each network size of these six heterogeneity levels, we fit an ERGM and a mERGM to the 50 simulated networks. Note that the tuning parameters for the θ_{gwesp} term are the same for both ERGMs and mERGMs with (`decay = 0.8, fixed = TRUE`). Additionally, we provide the results of two more simulation studies as supplementary material, where we perturb the parameter vector $\boldsymbol{\theta}$ with different configurations for both small and large networks. Furthermore, we also provide supplementary material illustrations of how well σ_u^2 is recovered in the simulation study.

Our first focus is on the performance of the estimates. In Tables 1 and 2, we summarize the results of our simulation study for small (50 nodes) and large (500 nodes) networks respectively, distinguishing all six different levels of nodal heterogeneity. Let us first look at the case $\sigma_u^2 = 0$, the fitted parametric coefficients of the ERGMs in both cases, small and large networks, show a stable estimation performance; however, this stability is more evident in large networks. The mERGMs, on the other hand, show some estimation variability and are outperformed by ERGMs concerning parameter estimation. The trend changes increasingly in favour of mERGMs, with

Table 1. Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 50 nodes and for all six σ_u^2 levels

Network size: 50 Nodes								
σ_u^2	Model type	Parameter	Real value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	−1	−0.71	0.79	−1.65	−0.74	0.37
		θ_{gwesp}	0.2	0.06	0.78	−0.06	0.17	0.38
		$\theta_{2-stars}$	−0.3	−0.35	0.13	−0.50	−0.34	−0.20
0	mERGM	θ_{edges}	−1	0.52	1.18	−0.89	0.39	0.31
		θ_{gwesp}	0.2	0.06	0.81	−0.06	0.17	0.38
		$\theta_{2-stars}$	−0.3	−0.58	0.19	−0.86	−0.57	−0.35
0.1	ERGM	θ_{edges}	−1	−1.11	0.77	−2.13	−1.12	−0.15
		θ_{gwesp}	0.2	0.16	0.16	0.03	0.19	0.34
		$\theta_{2-stars}$	−0.3	−0.25	0.12	−0.39	−0.24	−0.10
0.1	mERGM	θ_{edges}	−1	0.06	0.89	−0.96	−0.03	0.98
		θ_{gwesp}	0.2	0.16	0.16	−0.02	0.19	0.34
		$\theta_{2-stars}$	−0.3	−0.45	0.14	−0.60	−0.44	−0.28
0.2	ERGM	θ_{edges}	−1	−1.63	0.52	−2.21	−1.72	−0.88
		θ_{gwesp}	0.2	0.17	0.21	−0.06	0.17	0.39
		$\theta_{2-stars}$	−0.3	−0.17	0.08	−0.31	−0.16	−0.06
0.2	mERGM	θ_{edges}	−1	−0.73	0.61	−1.44	−0.85	0.21
		θ_{gwesp}	0.2	0.18	0.21	−0.07	0.19	0.40
		$\theta_{2-stars}$	−0.3	−0.33	0.11	−0.49	−0.31	−0.20
0.5	ERGM	θ_{edges}	−1	−2.05	0.43	−2.51	−2.10	−1.47
		θ_{gwesp}	0.2	0.21	0.12	0.03	0.21	0.37
		$\theta_{2-stars}$	−0.3	−0.07	0.06	−0.13	−0.06	−0.01
0.5	mERGM	θ_{edges}	−1	−0.71	0.62	−1.35	−0.77	0.03
		θ_{gwesp}	0.2	0.21	0.12	0.03	0.22	0.36
		$\theta_{2-stars}$	−0.3	−0.26	0.08	−0.38	−0.26	−0.17
0.8	ERGM	θ_{edges}	−1	−2.90	0.27	−3.17	−2.93	−2.58
		θ_{gwesp}	0.2	0.21	0.14	−0.01	0.21	0.34
		$\theta_{2-stars}$	−0.3	0.02	0.04	−0.04	0.03	0.07
0.8	mERGM	θ_{edges}	−1	−1.02	0.22	−1.33	−1.04	−0.74
		θ_{gwesp}	0.2	0.20	0.16	−0.03	0.23	0.37
		$\theta_{2-stars}$	−0.3	−0.39	0.14	−0.49	−0.33	−0.29
1	ERGM	θ_{edges}	−1	1123.83	9739.40	−186.45	−12.76	4.27
		θ_{gwesp}	0.2	−1112.36	4651.95	−737.19	−15.23	0.36
		$\theta_{2-stars}$	−0.3	785.38	10336.06	−2118.73	−3.77	104.20
1	mERGM	θ_{edges}	−1	−1.23	0.40	−1.68	−1.23	−0.74
		θ_{gwesp}	0.2	0.26	0.18	0.01	0.30	0.49
		$\theta_{2-stars}$	−0.3	-00.37	0.10	−0.49	−0.35	−0.26

increasing heterogeneity. At a heterogeneity of $\sigma_u^2 = 0.8$, the mERGMs excel with better results in small and large networks than the ERGMs, which show substantial variability in the results. At a heterogeneity of $\sigma_u^2 = 1$, the results indicate severe stability problems of the ERGM estimates, especially in the small networks, but the more the size of the network increases, the less unstable

Table 2. Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 500 nodes and for all six σ_u^2 levels

Network size: 500 nodes								
σ_u^2	Model type	Parameter	Real value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	−1	−0.96	0.29	−1.31	−0.98	−0.55
		θ_{gwesp}	0.2	0.19	0.05	0.12	0.20	0.27
		$\theta_{2-stars}$	−0.3	−0.30	0.03	−0.34	−0.30	−0.27
0	mERGM	θ_{edges}	−1	0.18	0.51	−0.53	0.27	0.81
		θ_{gwesp}	0.2	0.19	0.05	0.13	0.20	0.27
		$\theta_{2-stars}$	−0.3	−0.16	0.04	−0.21	−0.16	−0.11
0.1	ERGM	θ_{edges}	−1	−1.99	0.24	−2.34	−1.99	−1.73
		θ_{gwesp}	0.2	0.19	0.05	0.13	0.20	0.25
		$\theta_{2-stars}$	−0.3	−0.22	0.02	−0.23	−0.21	−0.18
0.1	mERGM	θ_{edges}	−1	−0.13	0.07	−0.22	−0.14	−0.04
		θ_{gwesp}	0.2	0.19	0.05	0.14	0.20	0.25
		$\theta_{2-stars}$	−0.3	−0.16	0.09	−0.28	−0.17	−0.03
0.2	ERGM	θ_{edges}	−1	−2.59	0.29	−2.93	−2.63	−2.16
		θ_{gwesp}	0.2	0.19	0.06	0.12	0.20	0.26
		$\theta_{2-stars}$	−0.3	−0.16	0.02	−0.19	−0.15	−0.13
0.2	mERGM	θ_{edges}	−1	−0.57	0.26	−0.92	−0.54	−0.22
		θ_{gwesp}	0.2	0.19	0.07	0.12	0.19	0.27
		$\theta_{2-stars}$	−0.3	−0.45	0.04	−0.49	−0.45	−0.42
0.5	ERGM	θ_{edges}	−1	−3.65	0.16	−3.82	−3.68	−3.45
		θ_{gwesp}	0.2	0.19	0.04	0.14	0.20	0.25
		$\theta_{2-stars}$	−0.3	−0.06	0.01	−0.08	−0.06	−0.05
0.5	mERGM	θ_{edges}	−1	−0.48	0.28	−0.79	−0.48	−0.19
		θ_{gwesp}	0.2	0.19	0.05	0.14	0.18	0.25
		$\theta_{2-stars}$	−0.3	−0.34	0.02	−0.37	−0.35	−0.32
0.8	ERGM	θ_{edges}	−1	−4.29	0.11	−4.41	−4.29	−4.15
		θ_{gwesp}	0.2	0.19	0.04	0.14	0.20	0.24
		$\theta_{2-stars}$	−0.3	−0.01	0.01	−0.02	−0.01	−0.001
0.8	mERGM	θ_{edges}	−1	−1.45	0.24	−1.77	−1.45	−1.16
		θ_{gwesp}	0.2	0.20	0.04	0.15	0.20	0.24
		$\theta_{2-stars}$	−0.3	−0.26	0.02	−0.28	−0.26	−0.23
1	ERGM	θ_{edges}	−1	−4.58	0.08	−4.67	−4.57	−4.48
		θ_{gwesp}	0.2	0.19	0.04	0.15	0.19	0.23
		$\theta_{2-stars}$	−0.3	0.01	0.01	0.01	0.01	0.02
1	mERGM	θ_{edges}	−1	−1.22	0.21	−1.84	−1.41	−1.29
		θ_{gwesp}	0.2	0.19	0.05	0.14	0.19	0.24
		$\theta_{2-stars}$	−0.3	−0.26	0.02	−0.26	−0.25	−0.23

the estimates of the ERGMs become. This fact is not surprising since a heterogeneity of $\sigma_u^2 = 1$ in a network with 50 nodes is a different claim than in a network with 500 nodes. Nevertheless, the mERGMs in both cases, small and large networks, show an appropriate and stable performance.

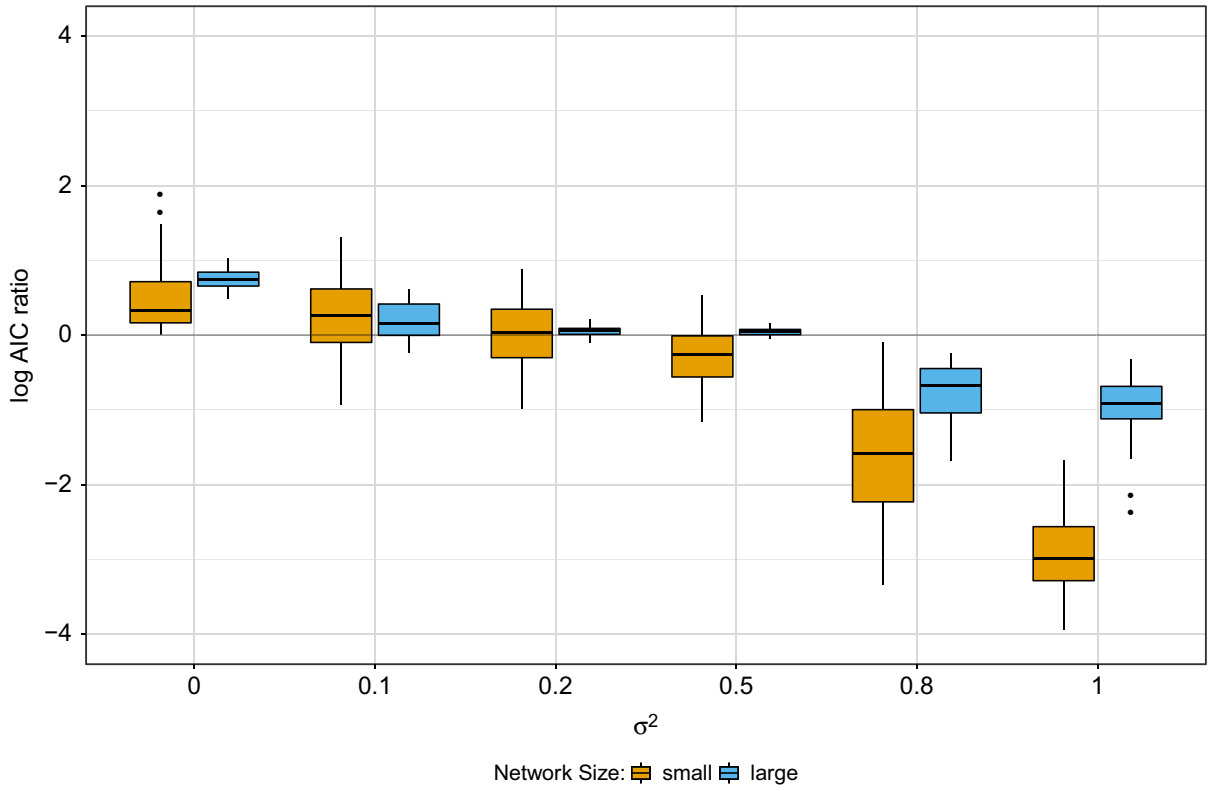


Figure 1. Resulting log AIC ratios of mERGM and ERGM for each network setting and size. Orange boxplots indicate small networks (50 nodes), blue boxplots large networks (500 nodes).

Hence, including nodal heterogeneity in the model increases the stability of the mERGM fit. This is a welcome effect of the model extension from ERGM to mERGM.

As a second point, we consider the performance of the model selection based on the AIC. To do so, we calculate for each of the 50 simulations in each setting the log ratio

$$\log\left(\frac{AIC_{mERGM_k}}{AIC_{ERGM_k}}\right) \quad k = 1, \ldots, 50$$

If the log ratio is positive, it speaks in favour of a model without nodal heterogeneity. In contrast, if the log ratio is negative, there is an indication of model heterogeneity. Figure 1 visualizes the log AIC ratio for different strengths of nodal heterogeneity and both network sizes. In small (50 nodes) and large (500 nodes) networks, the ERGM was correctly preferred in the case of missing nodal heterogeneity, that is $\sigma_u^2 = 0$. With increasing nodal heterogeneity level in the network, the mERGM becomes more appropriate, and from a heterogeneity level of $\sigma_u^2 = 0.5$ and $\sigma_u^2 = 0.8$ for small networks and large networks, respectively, the mERGM gets clearly selected based on the AIC. We, therefore, can conclude that the AIC allows for model selection in case of node-specific heterogeneity.

Additionally, we compare our approach with the approach proposed by Box-Steffensmeier et al. (2018) (FERGM) to illustrate the point of biased estimation of the parametric coefficients $\boldsymbol{\theta}$ we mentioned in Section (1). We apply the same simulation approach as above, distinguishing the size of the networks and compare the results of the mERGM with them of FERGM fitted to the 50 simulated networks with setting $\sigma_u^2 = 1$. The results of our comparison are given in Table 3. As we can conclude from Table 3, regardless of the network size, the performance of FERGM is worse. Nevertheless, we have to note that Box-Steffensmeier et al. (2018) also points out that their approach might only be suitable for certain research questions or certain network types.

Table 3. Resulting estimated means, standard deviations, the medians, 0.1 and 0.9 quantiles of the parameters for network setting with nodal heterogeneity $\sigma_u^2 = 1$

Network size	Model type	Parameter	Real value	Mean	SD	Q 0.1	Median	Q 0.9
50 Nodes	mERGM	θ_{edges}	−1	−1.23	0.40	−1.68	−1.23	−0.74
		θ_{gwesp}	0.2	0.26	0.18	0.01	0.30	0.49
		$\theta_{2-stars}$	−0.3	−0.37	0.10	−0.49	−0.35	−0.26
50 Nodes	FERGM	θ_{edges}	−1	37.99	6.51	35.29	38.89	42.06
		θ_{gwesp}	0.2	0.12	0.19	−0.11	0.10	0.38
		$\theta_{2-stars}$	−0.3	-7.58	1.23	−8.43	−7.65	−7.08
500 Nodes	mERGM	θ_{edges}	−1	−1.22	0.21	−1.84	−1.41	−1.29
		θ_{gwesp}	0.2	0.19	0.05	0.14	0.19	0.24
		$\theta_{2-stars}$	−0.3	−0.26	0.02	−0.26	−0.25	−0.23
500 Nodes	FERGM	θ_{edges}	-[1	-[5.18	0.04	-[5.24	-[5.18	-[5.12
		θ_{gwesp}	0.2	0.18	0.04	0.14	0.19	0.23
		$\theta_{2-stars}$	-[0.3	0.06	0.01	0.05	0.06	0.06

Table 4. Models with sufficient network statistics for the Facebook network data

Model type	Model	θ_{edges}	θ_{gwesp}	$\theta_{2-stars}$	θ_{gwnsp}	Nodal effects
ERGM	1	✓	✓	✓		
	2	✓	✓		✓	
mERGM	3	✓	✓	✓		✓
	4	✓	✓		✓	✓

5. Examples

5.1 Facebook network

As a first data example, we look at Facebook (undirected) network data, which is publicly accessible (McAuley and Leskovec, 2012). The entire network comprises 4039 nodes. We analyze a subset of 250 nodes to demonstrate the performance of our routine, e.g. we take the first 251 nodes of the network and remove the "center" node (the "ego"). Figure 2 gives a visual impression of the network. Just by looking at the network we can easily conclude that an ERGM which assumes nodal homogeneity is more than questionable. The mERGM, therefore, appears as a possible alternative. The aim of our analysis is to evaluate and compare the two models: the mERGM and the standard ERGM, with the intention to quantify the evidence for the presumed favour of the mERGM.

We fitted four models to the data, two mERGMs and two ERGMs. Table 4 describes the models by listing the sufficient network statistics. As network statistics, we included the number of edges, the number of two-stars and two weighted statistics, i.e. geometrically weighted edgewise shared partners (gwesp) and geometrically weighted nonedgewise shared partners (gwnsp). For the exact definitions of the weighted statistics, we refer to Snijders et al. (2006). The number of iterations for the mERGMs was set at 50 to ensure convergence.

Table 5 shows the resulting estimates for the models. Note that the standard deviations given for the mERGM estimates result from the ERGM fit taking the random effects as fixed. As noted above, these estimates are not reliable as they ignore the uncertainty of the estimated random effects. We, therefore, give these values for completeness only but do not interpret them. We see that the gwesp coefficient is always positive, indicating that the probability for an edge between two partners increases with the number of shared partners for the considered edge. The effect is however generally smaller in the mERGM, that is, if node-specific heterogeneity is taken into account. To make the models comparable, we calculated the AIC values for both the ERGMs and

Table 5. Model fitting results for the Facebook network data. Standard errors listed in the mERGM are not accurate since they ignore the variability resulting through node heterogeneity

Model type	Model	Parameter	Estimate	SE	AIC
ERGM	1	θ_{edges}	−7.178	0.091	6049.086
		θ_{gwesp}	1.875	0.046	
		$\theta_{2-stars}$	0.052	0.094	
ERGM	2	θ_{edges}	−5.973	1.698	70977.08
		θ_{gwesp}	2.076	0.881	
		θ_{gwnsp}	−0.034	1.211	
mERGM	3	θ_{edges}	−6.021	*0.468*	4260.723
		θ_{gwesp}	1.186	*0.192*	
		$\theta_{2-stars}$	−0.008	*0.001*	
mERGM	4	θ_{edges}	−3.943	*0.698*	2867.874
		θ_{gwesp}	0.486	*0.305*	
		θ_{gwnsp}	−0.053	*0.008*	

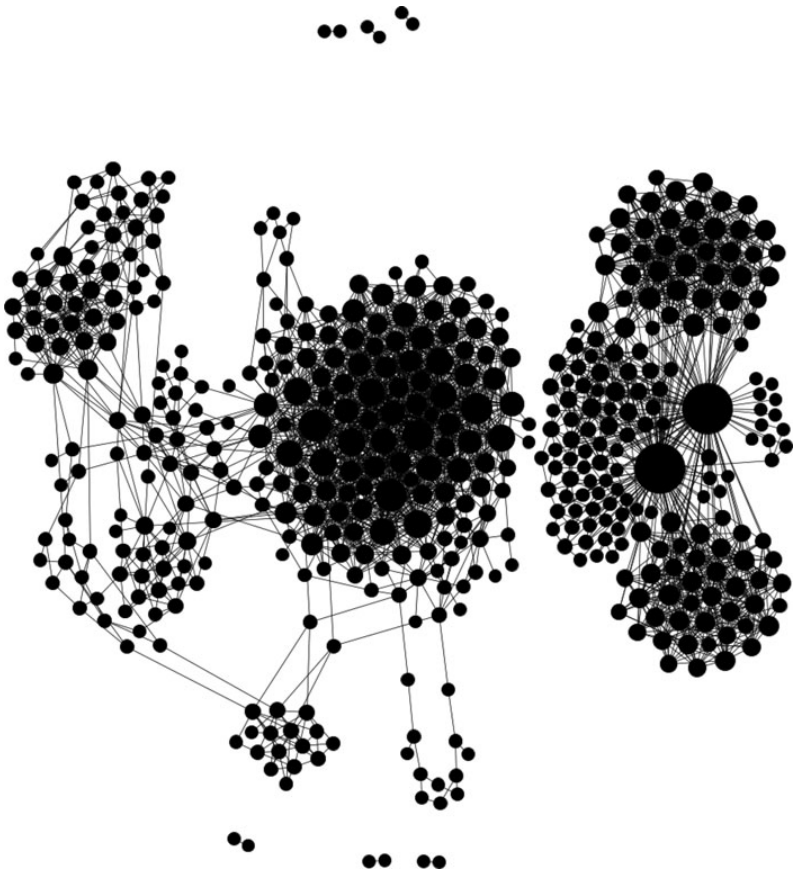


Figure 2. Facebook Network Data. Large nodes indicate nodes with high degrees.

the mERGMs according to the proposed approach as described in Section 3. For the calculation of the AIC values, we used 1000 simulations for both ERGMs and mERGMs, respectively.

Looking at the AIC values of the four models in Table 5, we see that both mERGMs outperform the ERGMs. This gives clear evidence of existing node-specific heterogeneity, and hence the proposed models with nodal random effects are preferable. This is also apparent in the goodness-of-fit plots. For instance, in Figure 3, we can see that model 3 fitted with mERGM gives a better

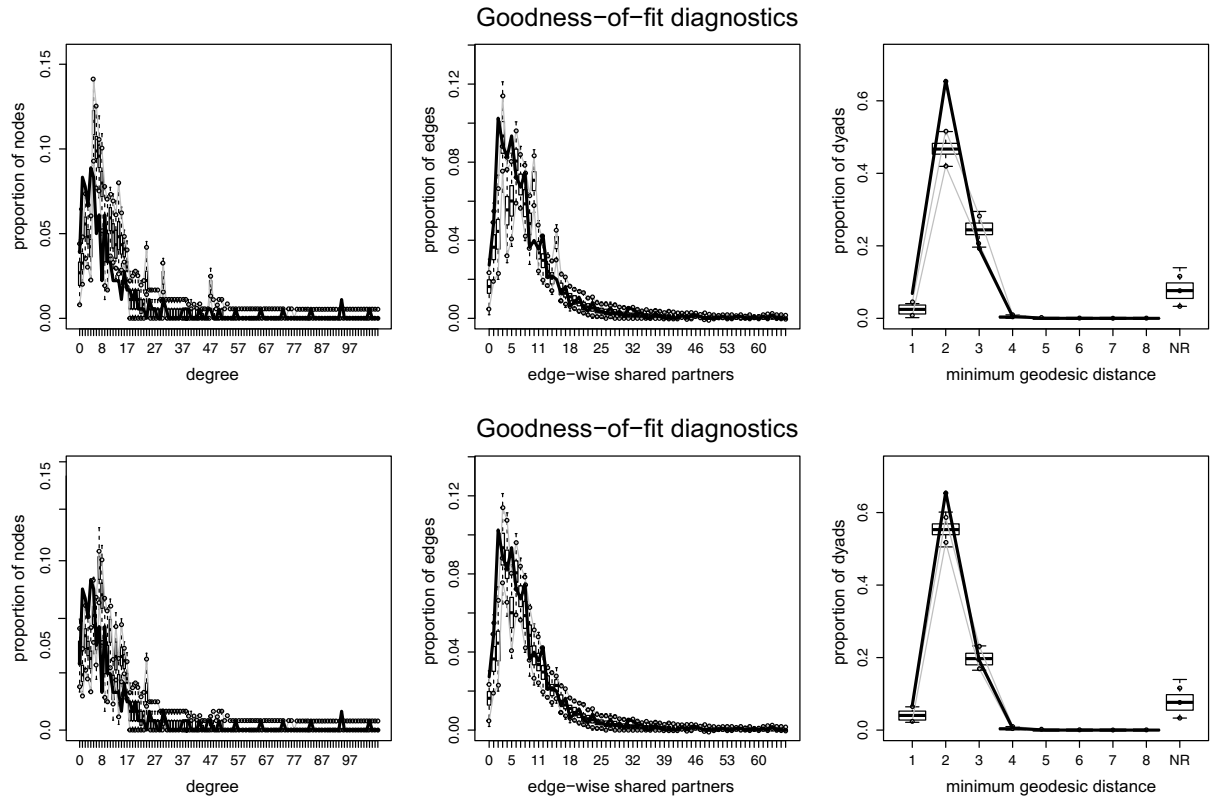


Figure 3. Goodness-of-fit diagnostics for model 1 fitted with ERGM (top row) and for model 3 fitted with mERGM (bottom row) for the Facebook network data.

fit compared to model 1 fitted with ERGM, which is in line with the corresponding AIC values. Furthermore, the AIC value of model 2 in Table 5 fitted with ERGM is exceptionally huge compared to model 4 fitted with mERGM, and we observe this issue well reflected in the goodness-of-fit plots in Figure 4, where model 2 struggles with huge convergence issues. Overall, model 4 appears to be the most suitable among the four fitted models to describe the data.

5.2 Zachary's Karate club network

As a second data example, we look at a well known dataset, the Zachary's karate club (Zachary, 1977). This undirected network data represents the friendship among 34 members of a university karate club. Figure 5 shows the network graph of Zachary's Karate Club. One can easily see that in this network, there are few nodes with high degree, while the remaining nodes have only few edges, so again, we assume that the mERGM should be a suitable approach to capture the unobserved nodal heterogeneity in the network. We fitted three different ERGMs and three mERGMs to this data. To make the models comparable, we included the same network statistics. Table 6 gives an overview of the different models. In Table 7, we summarize the results of our models including the AIC values. The iteration steps for the mERGMs was set to 50. For the calculation of the AIC values, we used 1000 simulations for both ERGMs and mERGMs.

We can see that model 1 fitted with ERGM struggles with convergence issues. This is mirrored in invalid variance estimates, resulting from a badly conditioned Fisher matrix. We, therefore, indicate this as "$\star$" in Table 7, which also means, of course, that the estimate itself is not reliable at all. We refer to Hunter et al. (2008) for further explanations. On the other hand, model 4 fitted with mERGM with the same model parameters as model 1 does not show any convergence issues, which also means that the mERGM can deal with estimation degeneracy issues. The inclusion of node-specific heterogeneity works towards numerical stabilization. To explore this in more depth,

Table 6. Models with sufficient network statistics for the Zachary network data

Model type	Model	θ_{edges}	θ_{gwesp}	$\theta_{2-stars}$	θ_{gwnsp}	$\theta_{gwdegree}$	Nodal effects
ERGM	1	✓	✓	✓			
	2	✓	✓		✓		
	3	✓	✓			✓	
mERGM	4	✓	✓	✓			✓
	5	✓	✓		✓		✓
	6	✓	✓			✓	✓

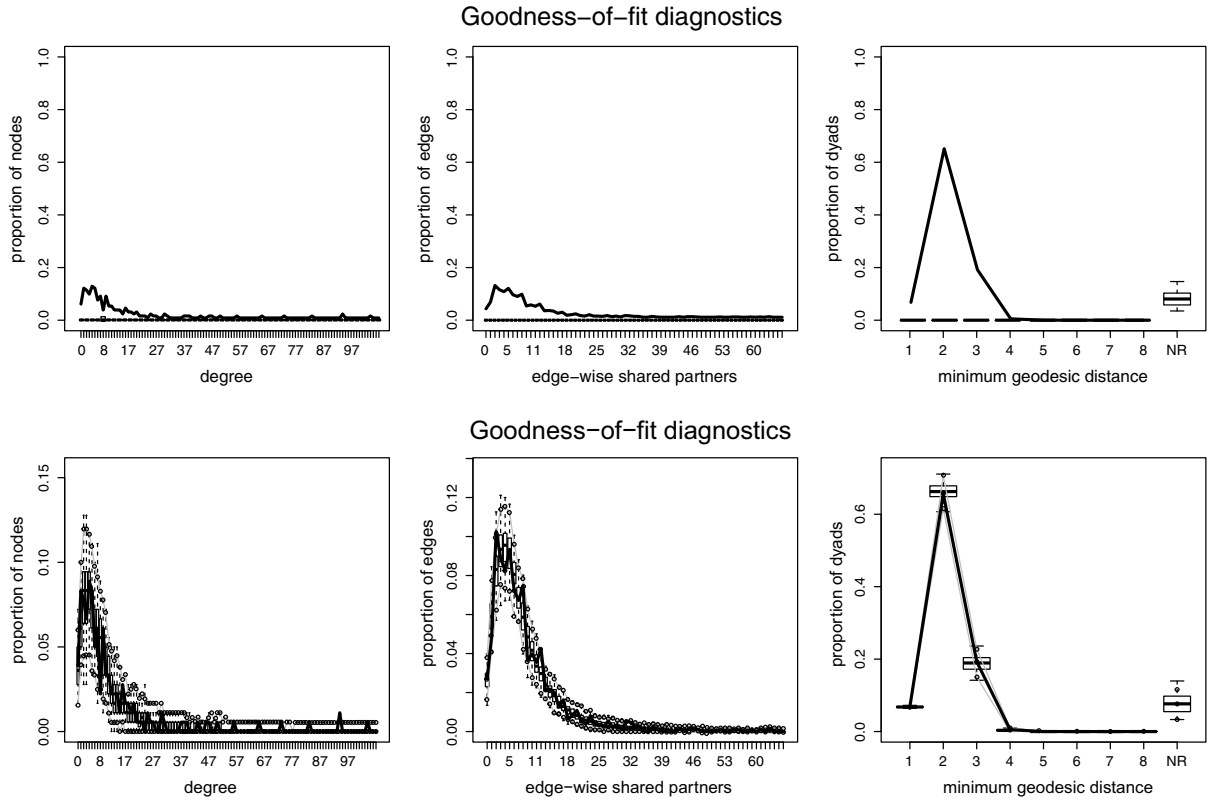


Figure 4. Goodness-of-fit diagnostics for model 2 fitted with ERGM (top row) and for model 4 fitted with mERGM (bottom row) for the Facebook network data.

we look in Figure 6 at the goodness-of-fit plots for model 1 fitted with ERGM. Figure 6 shows the same diagnostics results for model 4 fitted with mERGM. Remember that these two models include the same sufficient network statistics. Boxplots of the distributions of degree, edge-wise shared partners and minimum geodesic distance for the resulting simulated networks are shown in the plots where the bold line indicates the values of the original karate club dataset. In the diagnostics plots of model 1 in Figure 6, we can clearly see that ERGM fails to fit the model, whereas the diagnostics plots of model 4 in Figure 6 gives a good evidence of an appropriate fit.

We extend the model exploration to the other models. Figure 7 shows the goodness-of-fit diagnostics plots of both model 2 fitted with ERGM and model 5 fitted with mERGM, respectively. For model 2, we see some problems in Figure 7 concerning all the three diagnostics, the degree distribution, the edgewise-shared partners distribution and the minimum geodesic distance, which indicate the poorness of the model. For model 4, in contrast, we can see in Figure 7 again a much better performance of the fit.

Finally, Figure 8 shows the diagnostics plots of model 3 fitted with the ERGM, this model is the best ERGM fitted to this data according to the AIC value and also the diagnostics plots

Table 7. Model fitting results for the Zachary network data. Standard errors listed in the mERGM are not accurate since they ignore the variability resulting through node heterogeneity

Model type	Model	Parameter	Estimate	SE	AIC
ERGM	1	θ_{edges}	−4.893	⋆⋆⋆	
		θ_{gwesp}	0.642	⋆⋆⋆	⋆⋆⋆
		$\theta_{2-stars}$	0.689	⋆⋆⋆	
ERGM	2	θ_{edges}	−3.635	0.241	
		θ_{gwesp}	0.596	0.117	496.351
		θ_{gwnsp}	0.153	0.026	
ERGM	3	θ_{edges}	−3.183	0.513	
		θ_{gwesp}	0.716	0.181	442.883
		$\theta_{gwdegree}$	−0.519	0.8990	
mERGM	4	θ_{edges}	−1.214	*0.386*	
		θ_{gwesp}	0.236	*0.174*	311.304
		$\theta_{2-stars}$	−0.159	*0.042*	
mERGM	5	θ_{edges}	−1.776	*0.366*	
		θ_{gwesp}	−0.144	*0.109*	337.894
		θ_{gwnsp}	−0.089	*0.043*	
mERGM	6	θ_{edges}	−4.464	*0.724*	
		θ_{gwesp}	0.213	*0.176*	303.217
		$\theta_{gwdegree}$	4.427	*1.341*	

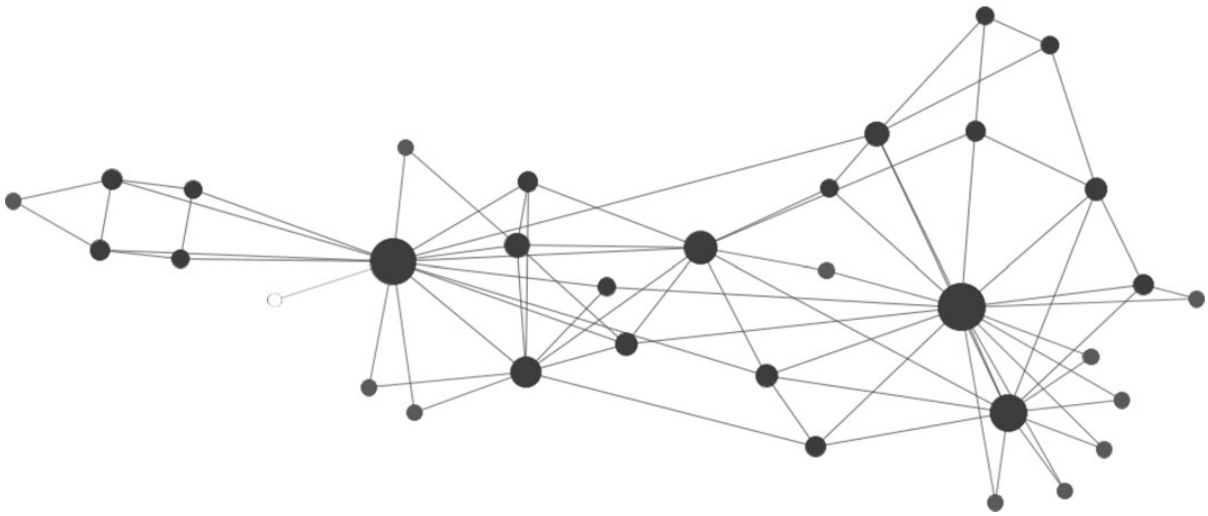


Figure 5. Zachary's Karate Club Network Data. Large nodes indicate nodes with high degrees.

are reasonable. On the other hand, model 6 fitted with mERGM, including the same sufficient network statistics as model 3, is the best mERGM fitted to this data according to the AIC value. However, the goodness-of-fit of model 6 shown in Figure 8 visually looks better than of model 3, which also justifies with a smaller AIC value.

5.3 High school friendship network

As a third data example, we investigate a real-world dataset, a high school friendship network in Marseilles, France, provided by Mastrandrea et al. (2015). This network data represents the friendship among 134 high school students of specific classes. These specific classes, unique to

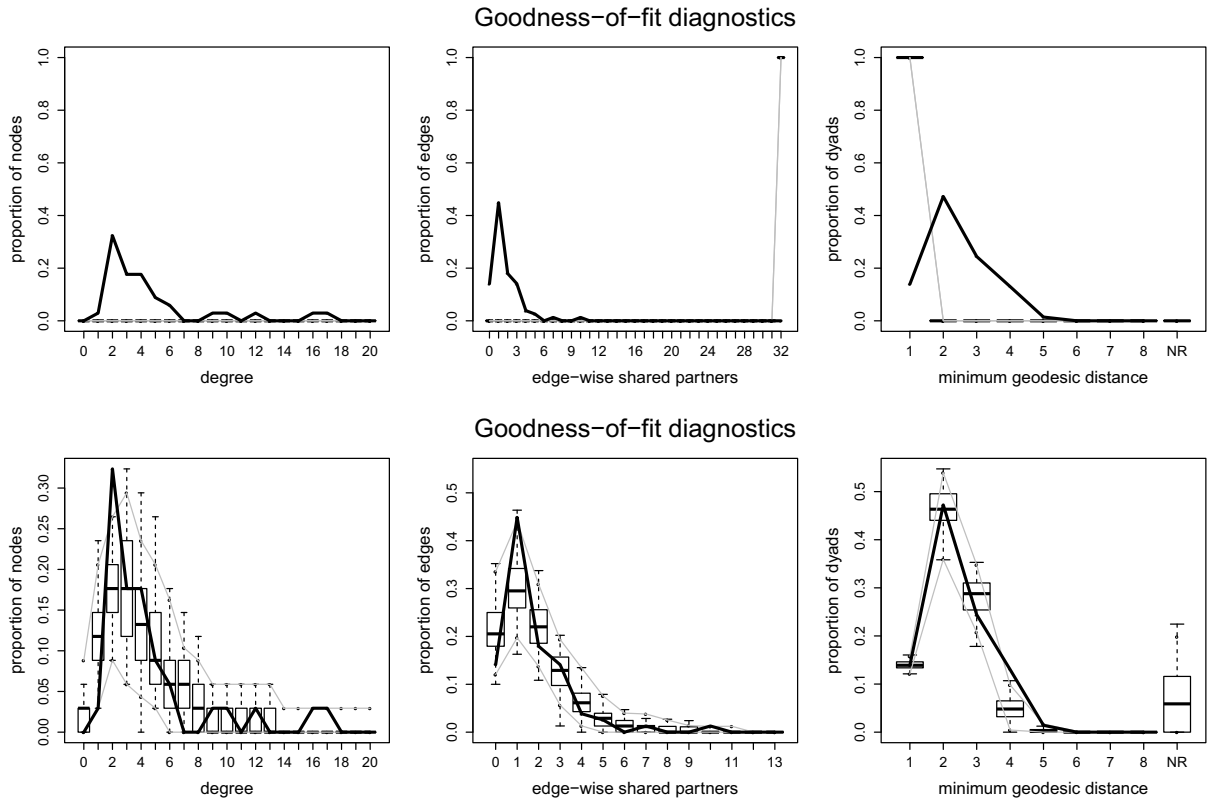


Figure 6. Goodness-of-fit diagnostics for model 1 fitted with ERGM (top row) and for model 4 fitted with mERGM (bottom row) for the karate club data.

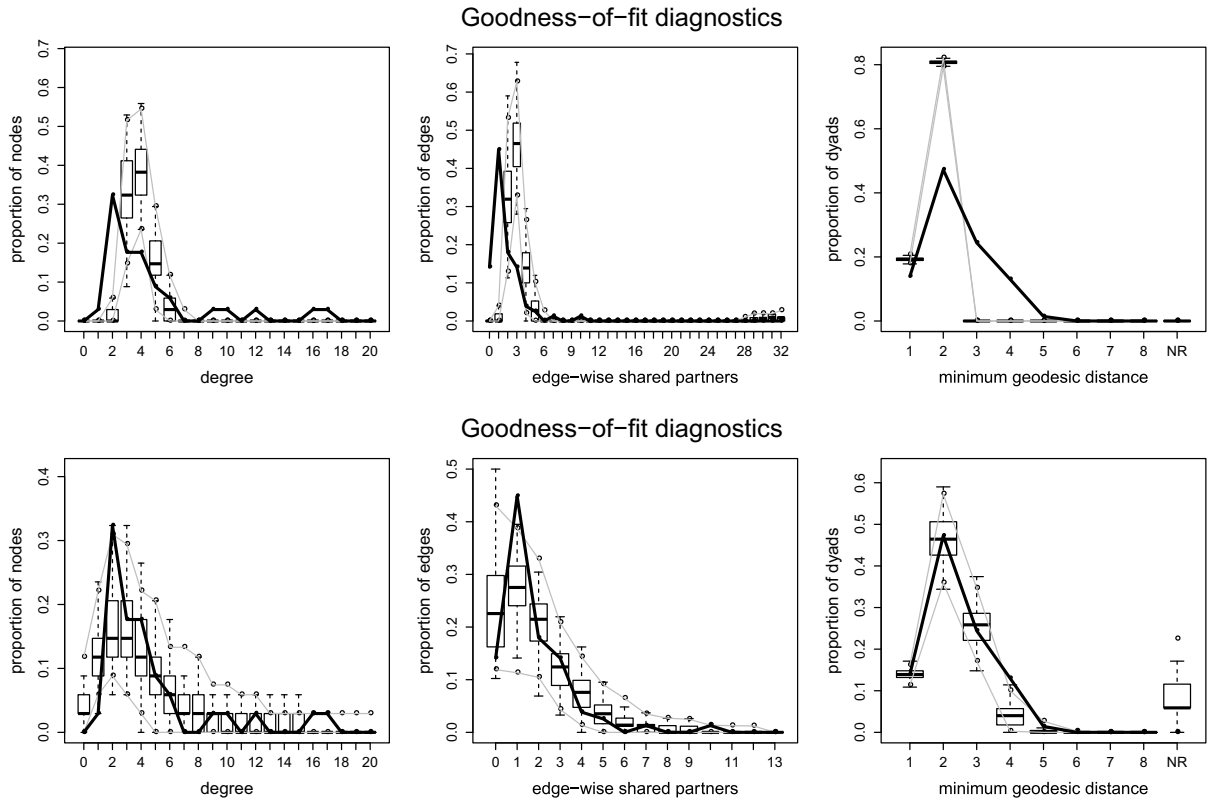


Figure 7. Goodness-of-fit diagnostics for model 2 fitted with ERGM (top row) and for model 5 fitted with mERGM (bottom row) for the karate club data.

Table 8. Number of students according to the types of the classes in the data

Classes	MP	MP*1	MP*2	PC	PC*	PSI*	2BIO1	2BIO2	2BIO3
No. of Students	21	3	7	21	10	15	10	19	28

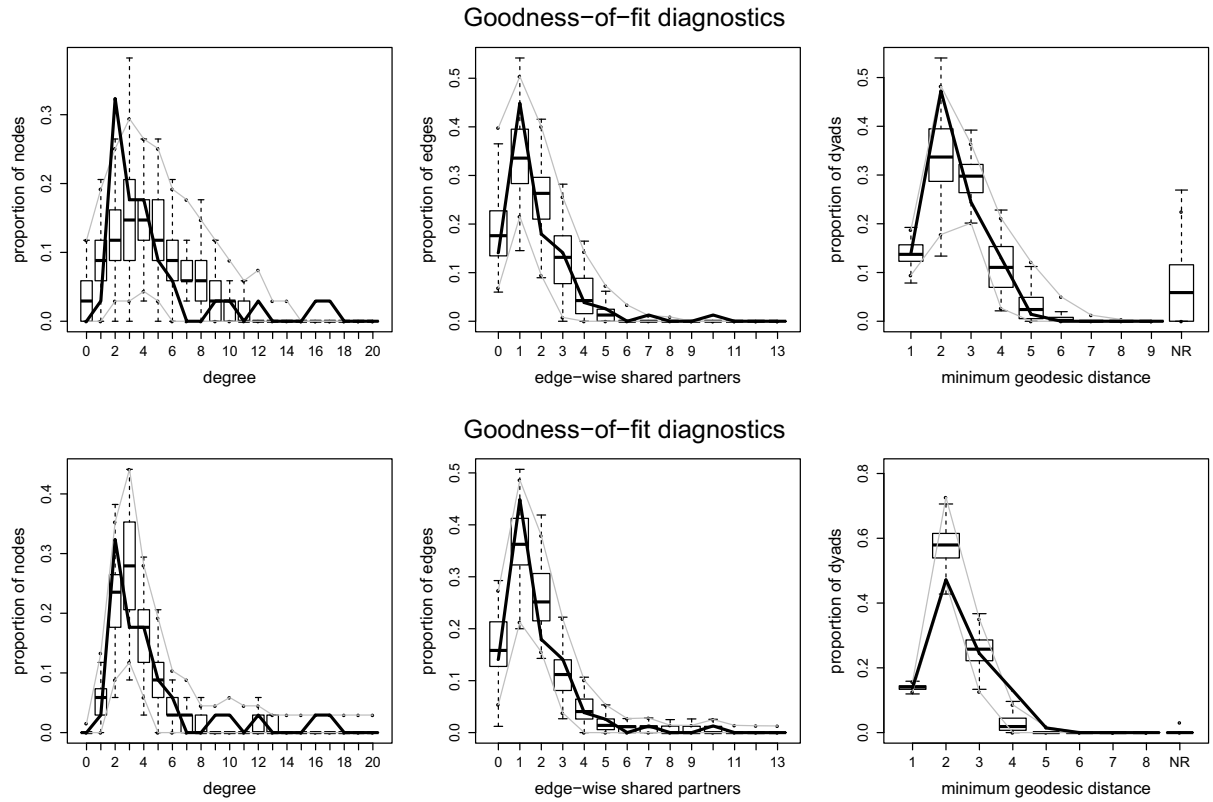


Figure 8. Goodness-of-fit diagnostics for model 3 fitted with ERGM (top row) and for model 6 fitted with mERGM (bottom row) for the karate club data.

the French educational system, gather students for studies that last two years after completing the usual high school studies. These classes are in a different part of the building, so the students are somehow isolated from the "regular" high school students. As a result, they form an almost closed population with little contact with the outside world, at least during schooldays. At the end of these 2 years, these students take competitive exams to gain admission to various higher educational institutions. The classes have different specializations: "MP" classes focus more on mathematics and physics, "PC" classes on physics and chemistry, "PSI" classes on engineering, and "BIO" classes on biology. Furthermore, there are three classes of type "MP", two of type "PC", one of type "PSI" and three of type "BIO". Due to the class sizes in the dataset, as demonstrated in Table 8, we decided to merge the types of classes to get an appropriate fit for our models, e.g., we do not distinguish between the types of the classes.

Additionally, we also have the gender information of the students in the data, with 79 female and 55 male students. Figure 9 shows the friendship network graph, the colour of the nodes represents the different classes, where the shape of the nodes the gender of the students. Assuming that there might be unobserved nodal heterogeneity, which can not be captured exclusively by the nodal covariates, the mERGM could be a reasonable approach.

We fit an ERGM and a mERGM to this data, including dyad-dependent network statistics such as the `GWESP` (geometrically weighted edgewise shared partner distribution) with a fixed decay parameter equal to 0.25 and dyad-independent terms such as the `nodefactor` and `nodematch` parameters. The `nodematch` is the homophily parameter in ERGM, where we also allow for each

Table 9. Model estimates for the high school friendship network. Standard errors listed in the mERGM are not accurate since they ignore the variability resulting through node heterogeneity

Model type	Parameter	Estimate	SE	Model type	Parameter	Estimate	SE
	Edges	−7.606	0.403		Edges	−9.021	*0.415*
	GWESP (0.25)	2.026	0.165		GWESP (0.25)	0.538	*0.215*
	Male	0.157	0.052		Male	−0.494	*0.068*
	Gender match	0.226	0.089		Gender match	0.813	*0.099*
	MP	0.627	0.272		MP	2.355	*0.238*
ERGM	PC	0.309	0.272	mERGM	PC	1.253	*0.238*
	PSI	1.054	0.276		PSI	3.042	*0.221*
	BIO class match	2.562	0.402		BIO class match	4.403	*0.295*
	MP class match	1.547	0.337		MP class match	2.814	*0.349*
	PC class match	2.279	0.393		PC class match	3.572	*0.379*
	PSI class match	1.504	0.379		PSI class match	2.047	*0.355*
AIC	3880.681			AIC	1097.836		

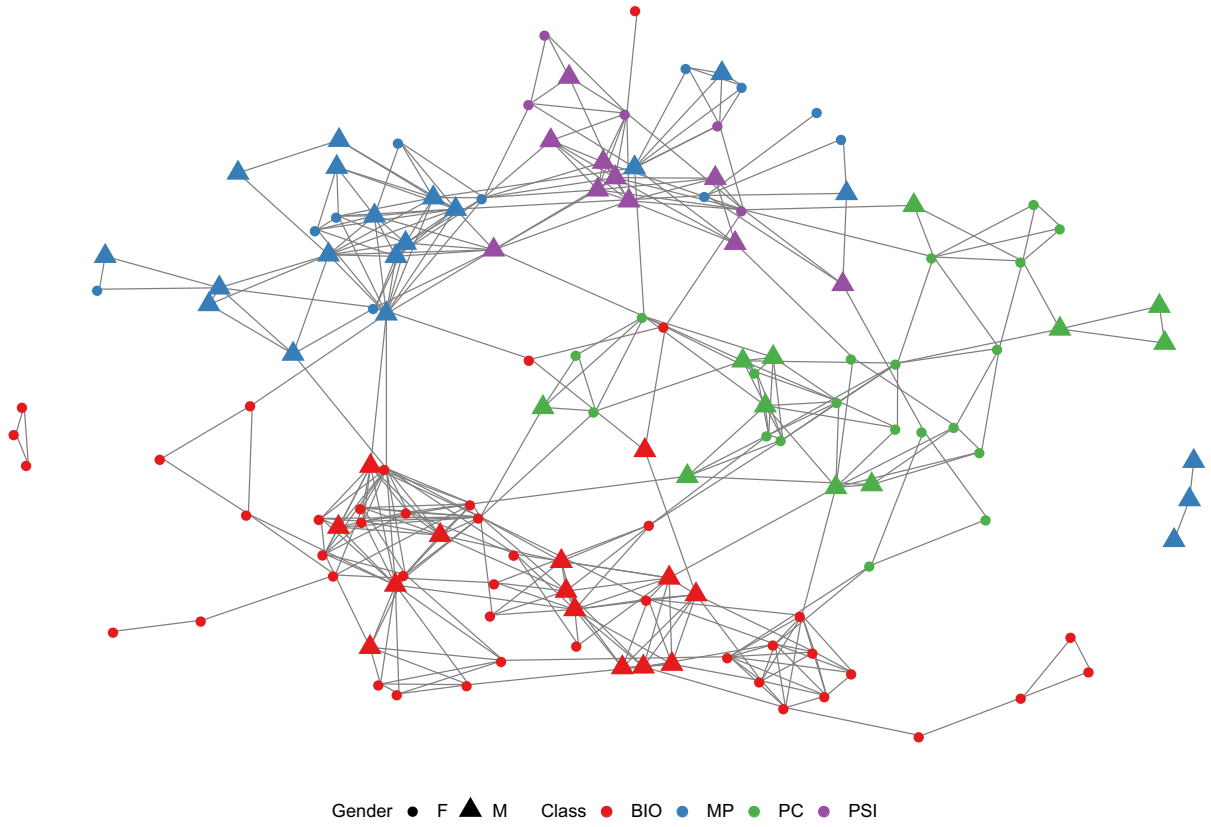


Figure 9. High School Friendship Network Data. The Colour of the nodes represents the different classes. The shape of the nodes describes the gender of the students.

class to have a unique propensity for within-class ties, we refer to Morris et al. (2008) for instance, for more details regarding the ERGM terms and their specifications. The iteration steps for the mERGMs was set to 50. For the calculation of the AIC values, we used 1000 simulations for both ERGMs and mERGMs. In Table 9, we summarize the results of the fitted models. Comparing the

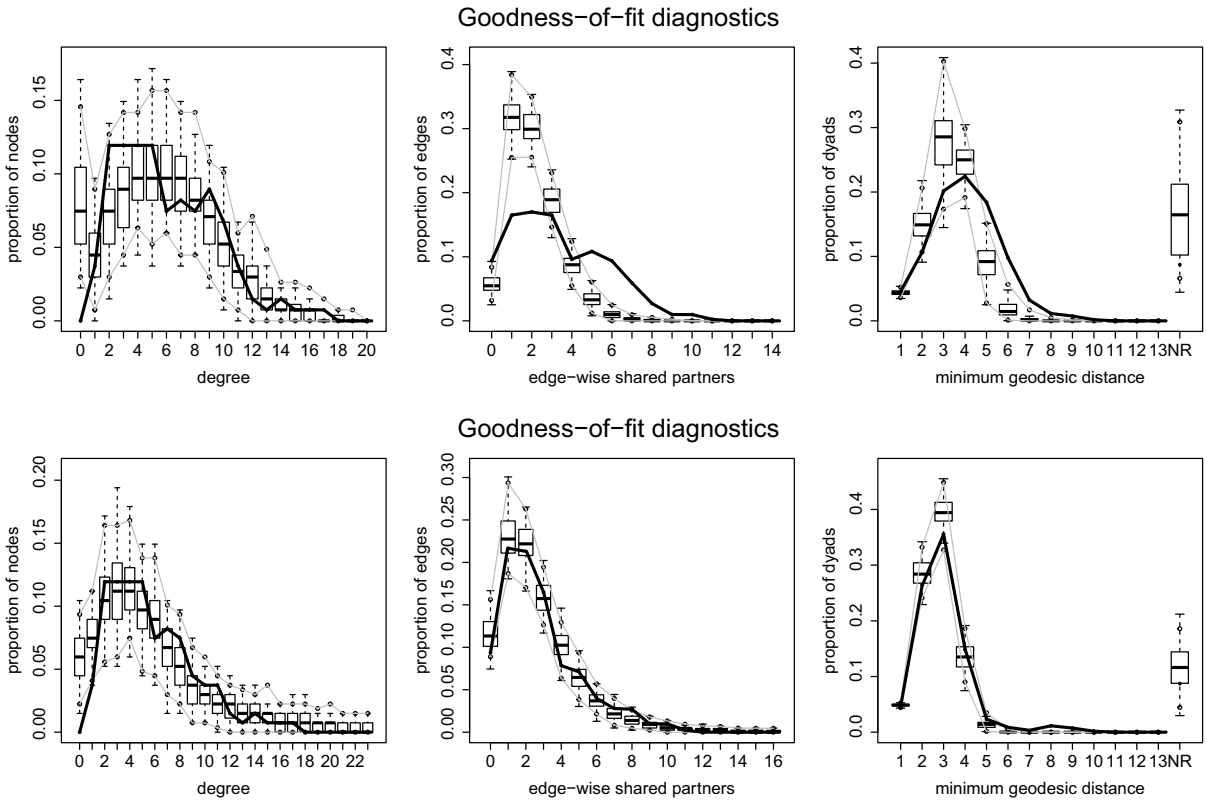


Figure 10. Goodness-of-fit diagnostics for the model fitted with ERGM (top row) and for the model fitted with mERGM (bottom row) for the high school friendship network data.

AIC values of the two models, the model fitted with mERGM is preferred, which is not very surprising since the mERGM additionally takes the "unobserved" nodal-heterogeneity into account, which is not captured by dyad-dependent or independent terms, aka network statistics. For the sake of interpretation, we look, for instance, at the log-odds of a hypothetical tie between two male students attending the MP class that does not close a triangle. In ERGM, the log-odds of such a tie would be -4.264; in mERGM, the interpretation is slightly different since it should be only in the conditional sense. This means, given two male students that attend the MP class, the conditional log-odds of such a tie is -1.671. In Figure 10, we show the diagnostic plots of the two models, we can clearly see that the model fitted with the mERGM performs much better, which is also clear evidence to the corresponding AIC values.

6. Discussion

In most cases, nodal heterogeneity in the network is explained by including known or well-studied nodal covariates, see e.g. Robins et al. (2001). However, the node-specific covariates cannot fully or sufficiently account for unobserved heterogeneity in the network. Our extensions towards mixed ERGMs can therefore be a meaningful approach to model network data by just adding nodal random effects to the model to capture the unobserved nodal heterogeneity.

Our proposed model (4) can in principle also be extended to directed networks, where $u_i^{(s)}$ and $u_j^{(r)}$ would be treated as random sender and random receiver node-specific coefficients, respectively. We consider it beyond the scope of the current paper.

Though the calculation of the AIC value is computationally intensive, our proposed method of estimating and calculating the AIC values allows us to compare the mERGM with the conventional ERGM. Furthermore, as we can see in our simulation study, the mERGM can always be a

reasonable approach for modelling networks even if we observe small nodal heterogeneity in the network. Overall, the mERGM works towards stabilizing the fitting routine without adding too much numerical effort.

In this study, neither in the simulation study nor in the examples, we investigate assortative mixing parameters. Therefore, a more thorough investigation in this regard is worth pursuing in future work.

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

Funding. This work has been funded by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IS18036A. The authors of this work take full responsibility for its content.

Competing interests. None.

Ethical statement. The authors also confirm that each author has seen and approved the contents of the submitted manuscript.

Supplementary. For supplementary material for this article, please visit http://dx.doi.org/10.1017/nws.2021.22.

References

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike* (pp. 215–222). Springer.

Box-Steffensmeier, J. M., Christenson, D. P., & Morgan, J. W. (2018). Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. *Political Analysis*, 26(1), 3–19.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9–25.

Caimo, A., & Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1), 41–55.

Cranmer, S. J., & Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1), 66–86.

Desmarais, B., & Cranmer, S. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4), 1865–1876.

Duijn, M. A. J., Snijders, T. A., & Zijlstra, B. J. H. (2004). p_2: A random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2), 234–254.

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC Press.

Fellows, I., & Handcock, M. S. (2012). Exponential-family random network models. Retrieved from arXiv preprint arXiv:1208.0121.

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., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832–842.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–233.

Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, 97(4), 773–789.

Henry, T. R., Gates, K. M., Prinstein, M. J., & Steinley, D. (2020). Modeling heterogeneous peer assortment effects using finite mixture exponential random graph models. *Psychometrika*, 85(1), 8–34.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33–50.

Hummel, R. M., Hunter, D. R., & Handcock, M. S. (2012). Improving simulation-based algorithms for fitting ergms. *Journal of Computational and Graphical Statistics*, 21(4), 920–939.

Hunter, D. R. & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3), 565–583.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), 1–29.

Hunter, D. R., Krivitsky, P. N., & Schweinberger, M. (2012). Computational statistical methods for social network analysis. *Journal of Computational and Graphical Statistics*, 21(4), 856–882.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Models*. New York: Springer.

Kolaczyk, E. D. & Csárdi, G. (2014). *Statistical analysis of network data with R*. New York: Springer.
Koskinen, J. (2009). Using latent variables to account for heterogeneity in exponential family random graph models.
Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks*. Cambridge: Cambridge University Press.
Mastrandrea, R., Fournet, J., & Barrat, A. (2015). Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9), e0136497.
McAuley, J. J., & Leskovec, J. (2012). In *NIPS*, volume 2012. Citeseer (pp. 548–656).
Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4), 1548.
Robins, G. L., Elliot, P., & Pattison, P. (2001). Network models for social selection processes. *Social Networks*, 23(1), 1–30.
Schmid, C. S. & Desmarais, B. A. (2017). Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE (pp. 116–121).
Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496), 1361–1370.
Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the American Statistical Association*, 77(3), 647.
Schweinberger, M., Krivitsky, P. N., & Butts, C. T. (2017). Foundations of finite-, super-, and infinite-population random graph inference. arXiv preprint arXiv:1707.04800.
Snijders, T. A. B. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1–40.
Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99–153.
Strauss, D. & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409), 204–212.
Thiemichen, S., Friel, N., Caimo, A., & Kauermann, G. (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46, 11–28.
Vaida, F. & Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, 92(2), 351–370.
van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31(1), 52–62.
Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.
Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4), 452–473.
Zijlstra, B. J. H., Duijn, M. A. J., & Snijders, T. A. B. (2006). The multilevel p_2 model: A random effects model for the analysis of multiple social networks. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(1), 42–47.

Cite this article: Kevork S. and Kauermann G. (2022). Iterative estimation of mixed exponential random graph models with nodal random effects. *Network Science* **9**, 478–498. https://doi.org/10.1017/nws.2021.22

3 Bipartite Exponential Random Graph Models with Nodal Random Effects

Contributed Manuscript

Kevork, S. and Kauermann, G. (2022). Bipartite exponential random graph models with nodal random effects. *Social Networks*, 70:90–99.
`https://doi.org/10.1016/j.socnet.2021.11.002`

Copyright

Author Contributions

The extension idea of the iterative estimation approach applied on bipartite networks, including nodal random effects, came from Sevag Kevork and Göran Kauermann. Sevag Kevork wrote the algorithmic implementation, conducted the simulation study, and performed the data analysis. The manuscript was mainly written by Sevag Kevork. All authors were involved in proofreading the manuscript.

Software

All computations and plots in this chapter have been produced using `R` version 3.6.0 with packages `ergm` 3.7.1, `mgcv` 1.8-31, `network` 1.15, `Matrix` 1.3-4, `GGally` 2.0.0, and `matrixStats` 0.58.0.

The estimation algorithm is available in the `R` package `bimergm` on GitHub (`https://github.com/kevorks/bimergm`).

Social Networks 70 (2022) 90–99

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Bipartite exponential random graph models with nodal random effects

Sevag Kevork *, Göran Kauermann

Institut für Statistik, Ludwigs-Maximilians-Universität München, Ludwigstr. 33, 80538 München, Germany

ARTICLE INFO

Keywords:
Exponential random graph models
Bipartite networks
Random effects
Generalized linear mixed models
Network data analysis

ABSTRACT

We examine the inclusion of specific nodal random effects for first- and second-mode nodes towards an ERGM for bipartite networks. The inclusion of such node-specific random effects in the ERGM accounts for unobserved heterogeneity in the bipartite network and ensures stable estimation results, especially for large-scale bipartite networks. Moreover, The predicted nodal random effects deliver reasonable interpretation to understand the network behavior. The estimation is carried out by an iterative estimation technique, iterating between pseudolikelihood estimation for the nodal random effects and maximum likelihood estimation for the network parameters.

1. Introduction

A bipartite network can be represented as a triple $Y = (R, C, E)$ where R and C, also often called first- and second-mode nodes, are two disjoint sets of nodes, and $E \subseteq R \times C$ is the set of edges of the network. This representation can be demonstrated as a $n \times m$ rectangular matrix Y (see Fig. 1), where n and m are the number of nodes in R (mode-1 nodes) and C (mode-2 nodes), respectively, and with elements $Y_{ij} = 1$ if there is an edge between i and j and $Y_{ij} = 0$ otherwise. Bipartite networks are also known as affiliation or two-mode networks. Some of the widely used examples of this type of network are citation networks, collaboration networks, actor–film networks, or patent networks. Borgatti and Everett (1997) provide a general overview of the basic notions of bipartite network analysis and discuss ways of applying and interpreting traditional network analytic techniques to two-mode network data. More recently, Latapy et al. (2008) propose extensions of basic tools to analyze large one-mode networks (the classical case) to the two-mode case. A general survey of different approaches for bipartite network data is provided in Shi et al. (2017).

A central model for unipartite networks is the Exponential Random Graph Models (ERGM) introduced by Frank and Strauss (1986). This model class allows to explain local network structures, see Lusher et al. (2013). The ERGM has been extended in the last years to bipartite, aka two-mode network analysis. Agneessens and Roose (2008) and Wang et al. (2013) proposed a set of new local network configurations (statistics) for ERGM specifications that are relevant to bipartite networks and allow to study or explain bipartite network structures, see also Wang et al. (2013) or Wang et al. (2016). The key ingredient of the model is that the probability of observing a bipartite network depends on a set of (sufficient) network statistics, similar to unipartite networks, and can be defined as

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}\big) = \frac{\exp\{\boldsymbol{\theta}^T \boldsymbol{s}(y)\}}{\kappa(\boldsymbol{\theta})}, \tag{1}$$

where $\boldsymbol{s}(y)$ is the vector of network statistics and $\boldsymbol{\theta}$ is the vector of model coefficients. Vector $\boldsymbol{s}(.)$ includes any structural configurations of the bipartite network, which can be extended to $\boldsymbol{s}(y, X)$ if covariates X are included in the model as well, as discussed in Handcock and Gile (2010). The denominator $\kappa(\boldsymbol{\theta})$ in (1) represents the normalizing factor to ensure that (1) is a legitimate probability mass function. Usually, $\kappa(\boldsymbol{\theta})$ is numerically intractable unless for small-scale networks. Estimation of $\boldsymbol{\theta}$ in (1) needs, therefore, to be carried out simulation-based. An early reference for the estimation of ERGMs in standard networks is Snijders (2002). For a general discussion, we refer to Hunter et al. (2012). A stable numerical routine has been proposed in Hummel et al. (2012) using a so-called stepping algorithm.

A resulting consequence of the model (1) is that the structure of a given network is completely explained by the endogenous network statistics defined in $\boldsymbol{s}(y)$, or, if covariates are included, by $\boldsymbol{s}(y, X)$. For simplicity of notation, we omit exogenous covariates X for the moment and consider model (1). The model formulation implies that networks with the same network statistics $s(y)$ have the same probability. As a consequence, it follows that both first- and second-mode nodes are considered to be exchangeable. In many applications, this assumption is questionable. It seems, therefore, advisable to allow for node-specific heterogeneity in the model. In unipartite network, this can be done by the inclusion of node-specific random effects in the style of a so-called p_2 model (see Duijn et al., 2004 and Zijlstra et al., 2006); see

* Corresponding author.
E-mail addresses: sevag.kevork@stat.uni-muenchen.de (S. Kevork), goeran.kauermann@stat.uni-muenchen.de (G. Kauermann).

https://doi.org/10.1016/j.socnet.2021.11.002

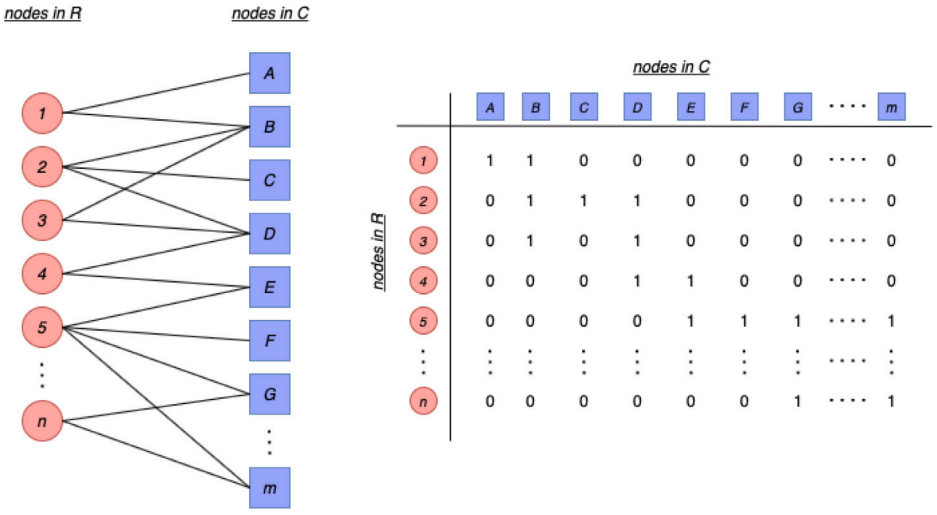


Fig. 1. Schematic representation of a bipartite network. The circles represent the nodes in R, and the squares represent the nodes in C.

also Thiemichen et al. (2016) for Bayesian estimation. A numerically faster estimation was proposed in Box-Steffensmeier et al. (2018) and Kevork and Kauermann (2019) using a combination of pseudolikelihood and maximum likelihood estimation. This paper aims to extend the inclusion of node-specific random effects in ERGMs for bipartite networks for first- and second-mode nodes. Moreover, through our extension, the interpretation of the predicted nodal random effects allows us to explain the bipartite network structure in more depth. We demonstrate both interpretability as well as numerical feasibility of our model extensions with several data examples. The proposed model can also be seen as an extension to a classical Rasch model (Rasch, 1961) by considering the bipartite matrix as an item response matrix for multiple items asked to several units.

This paper is organized as follows. In Section 2 we introduce our model and present our algorithm steps in detail. We also discuss methods of model evaluation. In Section 3 we present a simulation study and introduce the estimation results and a detailed analysis of the random coefficients of our proposed model. In Section 4 we apply our approach to three data examples of different sizes. Finally, Section 5 closes with a discussion.

2. Model estimation and evaluation

2.1. Approximate iterative estimation

We extend model (1) by including bipartite degree statistics. To be specific, we assume

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}\big) = \frac{\exp\big\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{r}(y) + \boldsymbol{v}^T \boldsymbol{c}(y)\big\}}{\kappa\big(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}\big)}, \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^q$ is a q dimensional parameter vector, $\boldsymbol{s}(y) = (s_0(y), s_1(y), \ldots)$ is a q dimensional vector of network statistics with $s_0(y) = \sum_i \sum_j y_{ij}$ and $\boldsymbol{r}(y) = \big(\sum_{j=1}^m y_{1j}, \sum_{j=1}^m y_{2j}, \ldots\big)$ is the n dimensional vector of node degrees in R and $\boldsymbol{c}(y) = \big(\sum_{i=1}^n y_{i1}, \sum_{i=1}^n y_{i2}, \ldots\big)$ is the m dimensional vector of node degrees in C. The normalization now equals $\kappa(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}) = \log \sum_{y \in \mathcal{Y}} \exp\big(\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{r}(y) + \boldsymbol{v}^T \boldsymbol{c}(y)\big)$, where $\mathcal{Y}$ is the set of all n by m bipartite networks. Note that $\boldsymbol{u}$ is n dimensional and $\boldsymbol{v}$ is m dimensional, so that model (2) is parameterized by a $q + n + m$ parameters. This makes estimation rather infeasible, so that some kind regularization is required. We follow a Bayesian viewpoint here and assume that both $\boldsymbol{u} = (u_1, u_2, \ldots)$ and $\boldsymbol{v} = (v_1, v_2, \ldots)$ are unobserved vectors of random coefficients with prior structure

$$\boldsymbol{u} \sim \mathcal{N}(0, \sigma_u^2 \boldsymbol{I}_n) \quad \text{and} \quad \boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2 \boldsymbol{I}_m), \tag{3}$$

where σ_u^2 and σ_v^2 are the corresponding variances and $\boldsymbol{I}_n$ and $\boldsymbol{I}_m$ are n and m dimensional identity matrices, respectively. This leads to a model with fixed and random coefficients. We, therefore, call (2) in combination with (3) in the following a bipartite mixed ERGM or, in short, biMERGM. The reasoning for the model becomes clear by looking at the conditional distribution of a single edge Y_{ij} given the rest of the network Y_{-ij}. This conditional distribution follows a Bernoulli distribution with log-odds resulting from (2) through

$$\log\left\{\frac{\mathbb{P}(Y_{ij} = 1|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v})}{\mathbb{P}(Y_{ij} = 0|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v})}\right\} = \boldsymbol{\theta}^T \, \Delta_{ij} \, \boldsymbol{s}(y) + u_i + v_j. \tag{4}$$

Here, $\Delta_{ij}\boldsymbol{s}(y) = \boldsymbol{s}(y_{ij} = 1, y_{-ij}) - \boldsymbol{s}(y_{ij} = 0, y_{-ij})$ is the so called change statistics. Assuming normality as given in (3), the resulting conditional logit model (4) mimics a generalized mixed regression model (GLMM) as broadly discussed in Breslow and Clayton (1993). We take the conditional model (4) and the full model (2) as starting points for our estimation algorithm and propose to fit the model with an iterative combination of simulation-based routines and pseudolikelihood estimation. To be specific, we make use of pseudolikelihood estimation for predicting the random coefficients $\boldsymbol{u}$ and $\boldsymbol{v}$, while for estimation of $\boldsymbol{\theta}$, we use a MCMC-MLE based approach as suggested in Hummel et al. (2012). These two steps are used iteratively, leading to numerically feasible estimation even for large networks, as will be demonstrated.

To be specific, we aim to maximize the following marginal log-likelihood

$$\begin{aligned} l(\boldsymbol{\theta}, \sigma_u^2, \sigma_v^2) &\propto \log \int \frac{\exp\big\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \boldsymbol{u}^T \boldsymbol{r}(y) + \boldsymbol{v}^T \boldsymbol{c}(y)\big\}}{\kappa\big(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}\big)} \frac{1}{(\sigma_u^2)^{\frac{n}{2}}} \\ &\quad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\Big(-\frac{\boldsymbol{u}^T\boldsymbol{u}}{2\sigma_u^2} - \frac{\boldsymbol{v}^T\boldsymbol{v}}{2\sigma_v^2}\Big) \prod_{i=1}^n \prod_{j=1}^m du_i \, dv_j \\ &= \boldsymbol{\theta}^T \boldsymbol{s}(y) + \log \int \exp\Big(g(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\theta}, \sigma_u^2, \sigma_v^2)\Big) \prod_{i=1}^n \prod_{j=1}^m du_i \, dv_j \end{aligned} \tag{5}$$

where

$$\begin{aligned} g(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\theta}, \sigma_u^2, \sigma_v^2) &\propto \boldsymbol{u}^T \boldsymbol{r}(y) + \boldsymbol{v}^T \boldsymbol{c}(y) - \log(\kappa(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v})) - \frac{\boldsymbol{u}^T\boldsymbol{u}}{2\sigma_u^2} - \frac{\boldsymbol{v}^T\boldsymbol{v}}{2\sigma_v^2} \\ &\quad - \frac{n}{2}\log(\sigma_u^2) - \frac{m}{2}\log(\sigma_v^2). \end{aligned} \tag{6}$$

The integral in (5) can be approximated by Laplace approximation, which requires to maximize (6) with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$. Let $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ denote the corresponding maximizers. If we now treat $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ as given, then the maximization of the log-likelihood corresponds to maximizing

the likelihood of the probability model

$$\mathbb{P}\big(Y = y|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}\big) = \frac{\exp\big\{\boldsymbol{\theta}^T \boldsymbol{s}(y) + \hat{\boldsymbol{u}}^T \boldsymbol{r}(y) + \hat{\boldsymbol{v}}^T \boldsymbol{c}(y)\big\}}{\kappa\big(\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}\big)},$$

where $\hat{\boldsymbol{u}}^T \boldsymbol{r}(y)$ and $\hat{\boldsymbol{v}}^T \boldsymbol{c}(y)$ are fixed as given offsets. This implies that with given $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$, we can simplify the estimation of $\boldsymbol{\theta}$ to (profile) maximum likelihood estimation in an ERGM with two offset parameters. This is numerically feasible, as long as q, the dimension of $\boldsymbol{\theta}$ is reasonably small, which we assume. The computational challenge is that maximizing Eq. (6) with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$ is a numerically demanding if not infeasible due to the high dimension of $\boldsymbol{u}$ and $\boldsymbol{v}$. Bear in mind that these are n and m dimensional so that simulation-based techniques are proposed in Snijders (2002) are not suitable for large networks. We, therefore, propose to replace the estimation of $\boldsymbol{u}$ and $\boldsymbol{v}$ with pseudolikelihood estimation. That is to say, we fix the quantity $o_{ij} := \boldsymbol{\theta}^T \Delta_{ij}\, s(y)$ in model (4) as offset and ignore the dependency on $\boldsymbol{y}$. This leads to a pseudolikelihood approach similar to Duijn et al. (2009) but with random components. To be specific, for estimation of $\boldsymbol{u}$ and $\boldsymbol{v}$ we assume the likelihood

$$\begin{aligned} l(\sigma_u^2, \sigma_v^2; \boldsymbol{o}) &= \iint \prod_{i=1}^{n}\prod_{j=1}^{m} \mathbb{P}(Y_{ij} = y_{ij}|u_i, v_j, o_{ij}) \frac{1}{(\sigma_u^2)^{\frac{n}{2}}} \\ &\quad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\Big(-\frac{\boldsymbol{u}^T\boldsymbol{u}}{2\sigma_u^2} - \frac{\boldsymbol{v}^T\boldsymbol{v}}{2\sigma_v^2}\Big) du_i\, dv_j \\ &= \iint \exp\Big(\tilde{g}(\boldsymbol{u}, \boldsymbol{v}; \sigma_u^2, \sigma_v^2, \boldsymbol{o})\Big) \prod_{i=1}^{n}\prod_{j=1}^{m} du_i\, dv_j, \end{aligned} \tag{7}$$

where $\boldsymbol{o} = (o_{1\,1}, o_{1\,2}, \dots, o_{n\,m})$ is treated as offset with $o_{ij} = \boldsymbol{\theta}^T \Delta_{ij} \boldsymbol{s}(y)$ and

$$\begin{aligned} \tilde{g}(\boldsymbol{u}, \boldsymbol{v}; \sigma_u^2, \sigma_v^2, \boldsymbol{o}) &= y_{ij}(u_i + v_j + o_{ij}) - \log(1 + \exp(u_i + v_j + o_{ij})) \\ &\quad - \frac{n}{2}\log(\sigma_u^2) \frac{m}{2}\log(\sigma_v^2) - \frac{\boldsymbol{u}^T\boldsymbol{u}}{2\sigma_u^2} - \frac{\boldsymbol{v}^T\boldsymbol{v}}{2\sigma_v^2} \end{aligned}$$

is the log-likelihood of a binomial model with normal prior on the coefficients $\boldsymbol{u}$ and $\boldsymbol{v}$.

The pseudolikelihood approach transfers the estimation problem to settings extensively discussed in the context of generalized linear mixed models. To be specific, the integral on (7) can again be approximated by Laplace approximation, as proposed in Breslow and Clayton (1993). To do so, we now denote with $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ the maximizer of $\tilde{g}(\boldsymbol{u}, \boldsymbol{v}, \sigma_u^2, \sigma_v^2, \boldsymbol{o})$ which allows us to approximate the integral and in turn provides estimates for σ_u^2 and σ_v^2 by maximizing the Laplace approximated log-likelihood (7). This corresponds to a well established estimation strategy for generalized linear mixed models.

We can combine the two simulation steps in an iterative manner using implemented routines as for instance in R. In fact, we can use the `mgcv` package (see Wood, 2011) for estimation of $\boldsymbol{u}$ and $\boldsymbol{v}$ and for the estimation of $\boldsymbol{\theta}$ we use the `ergm` package (see Hunter et al., 2008). Our algorithmic steps work as follows:

Algorithm: Fit bipartite ERGM with nodal random effects

Step 0: Obtain a prediction for $\boldsymbol{u}$ and $\boldsymbol{v}$ and estimate σ_u^2 and σ_v^2:

(i) Fit the model `logit` $\mathbb{P}(y_{ij} = 1|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}) = \theta_{(0)} + u_i + v_j$ to the data

(ii) extract the vector of the predicted random effects $\boldsymbol{u}_{(0)}$ and $\boldsymbol{v}_{(0)}$ as offset and set $t = 0$

Step 1: Estimate $\boldsymbol{\theta}$ with ERGM and take $\boldsymbol{u}_{(t)}^T \boldsymbol{r}(\boldsymbol{y})$ and $\boldsymbol{v}_{(t)}^T \boldsymbol{c}(\boldsymbol{y})$ as offset parameters:

(i) Fit the model $\mathbb{P}(Y = y|\boldsymbol{\theta}) \propto \exp\{\boldsymbol{\theta}_{(t+1)}^T \boldsymbol{s}(y) + \underbrace{\boldsymbol{u}_{(t)}^T \boldsymbol{r}(y) + \boldsymbol{v}_{(t)}^T \boldsymbol{c}(y)}_{\text{offset}}\}$ using maximum likelihood and simulation-based methods

(ii) extract $o_{ij} = \boldsymbol{\theta}_{(t+1)}^T \Delta_{ij} \boldsymbol{s}(y)$ as new offset

Step 2: Update $\boldsymbol{u}_{(t+1)}$, $\boldsymbol{v}_{(t+1)}$ and $\hat{\sigma}_{u(t+1)}^2$, $\hat{\sigma}_{v(t+1)}^2$ now taking o_{ij} as offset parameter:

(i) Fit the model `logit` $\mathbb{P}(y_{ij} = 1|y_{-ij}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}) = \underbrace{o_{ij}}_{\text{offset}} + u_i + v_j$ to the data

(ii) extract the vector of the predicted random effects $\boldsymbol{u}_{(t+1)}$ and $\boldsymbol{v}_{(t+1)}$ as new offset

Set $t = t + 1$ and iterate between step 1 and 2 until $\boldsymbol{\theta}_{(t)}$ converges. Convergence is achieved if $|\boldsymbol{\theta}_{(t)} - \boldsymbol{\theta}_{(t+1)}| \leq \epsilon = 0.05$.

2.2. Model evaluation

After fitting the model using the algorithm above, it is advisable to evaluate the model. One of the major aims in modeling networks is to capture specific network structures of the observed data with the specified and estimated model. A common way to evaluate how well a network model fits the data is to sample several networks from the estimated model and compare those sampled networks to the original observed network by looking at specific characteristics and network statistics. In our application case, we follow Hanneke et al. (2010) and fit our proposed model (2) to the observed bipartite network. Based on this model, we sample B networks and calculate various network statistics. We then compare the distribution of the network statistics of the sampled networks with the network statistics on the actually observed network. With this approach, we can evaluate whether the simulated networks from the fitted model yield similar networks, which would infer stable estimations of the model parameters.

Furthermore, to evaluate the estimates obtained from our fitted models in Section 4 we follow Snijders and Van Duijn (2002) and provide the t-ratio of the estimates, which is defined as follows

$$t_k = \frac{s_k(y) - m_k(\hat{\theta})}{\sigma_k(\hat{\theta})}$$

where $s_k(y)$ is the kth observed network statistic, $m_k(\hat{\theta})$ is the mean of the kth network statistic over $B = 1000$ network samples, and $\sigma_k(\hat{\theta})$ is the standard deviation. To be more specific, from these $B = 1000$ network samples we compute the mean and the standard deviation of the kth network statistic and calculate the t-ratio in absolute value. As a criteria for good convergence the t-ratio should be less than 0.1 in absolute value for the kth network statistic included in the model.

3. Simulation study

Before applying our model and the proposed estimation routine to real-world networks, we run some simulation studies to check and evaluate the performance of our algorithm. In the subsequent simulation study, we aim to verify our proposed estimation approach and evaluate the estimation results of the model parameters based on network settings with different nodal heterogeneity levels in both first- and second-mode nodes. For each network setting, we simulate 100 networks using the simulation techniques from the `ergm` package (Hunter et al., 2008), with fixed structural effects and randomly drawn $\boldsymbol{u}$ and $\boldsymbol{v}$ for each simulated network and setting. We define our simulation study configurations as follows:

- $n = 160$ and $m = 100$, denoting the number of nodes in R and C respectively.
- $\boldsymbol{\theta} = (\theta_{edges} = -2,\ \theta_{gwb1dsp} = 0.2,\ \theta_{gwb2dsp} = -0.2)$ where `gwb1dsp` and `gwb2dsp` are bipartite network statistics denoting the geometrically weighted dyadwise shared partner distribution for dyads in the first/second bipartition respectively, see Hunter and Handcock (2006) for more details.
- $\sigma_u^2 \in (0, 0.5, 1)$ and $\sigma_v^2 \in (0, 0.5, 1)$ denoting different heterogeneity levels in R and C respectively. The different heterogeneity levels for R, and C lead to nine possible combinations.

Table 1
Resulting means, medians and the standard deviations of the estimated coefficients for all simulation settings.

Heterogeneity level (σ_u^2, σ_v^2)	Parameter	Real value	Mean	Median	SD
(0, 0)	θ_{edges}	−2	−2.142	−2.161	0.577
	$\theta_{gwb1dsp}$	0.2	0.270	0.257	0.408
	$\theta_{gwb2dsp}$	−0.2	−0.207	−0.217	0.329
(0, 0.5)	θ_{edges}	−2	−2.185	−2.211	0.550
	$\theta_{gwb1dsp}$	0.2	0.161	0.179	0.401
	$\theta_{gwb2dsp}$	−0.2	−0.218	−0.237	0.347
(0, 1)	θ_{edges}	−2	−2.112	−2.104	0.322
	$\theta_{gwb1dsp}$	0.2	0.217	0.225	0.445
	$\theta_{gwb2dsp}$	−0.2	−0.219	−0.232	0.314
(0.5, 0)	θ_{edges}	−2	−2.306	−2.284	0.372
	$\theta_{gwb1dsp}$	0.2	0.172	0.198	0.429
	$\theta_{gwb2dsp}$	−0.2	−0.175	−0.166	0.303
(0.5, 0.5)	θ_{edges}	−2	−1.948	−1.947	0.205
	$\theta_{gwb1dsp}$	0.2	0.216	0.197	0.305
	$\theta_{gwb2dsp}$	−0.2	−0.188	−0.187	0.201
(0.5, 1)	θ_{edges}	−2	−2.108	−2.115	0.213
	$\theta_{gwb1dsp}$	0.2	0.204	0.198	0.165
	$\theta_{gwb2dsp}$	−0.2	−0.197	−0.205	0.213
(1, 0)	θ_{edges}	−2	−2.137	−2.150	0.238
	$\theta_{gwb1dsp}$	0.2	0.190	0.181	0.238
	$\theta_{gwb2dsp}$	−0.2	−0.228	−0.252	0.316
(1, 0.5)	θ_{edges}	−2	−1.981	−1.953	0.179
	$\theta_{gwb1dsp}$	0.2	0.204	0.219	0.103
	$\theta_{gwb2dsp}$	−0.2	−0.190	−0.174	0.327
(1, 1)	θ_{edges}	−2	−2.016	−2.027	0.131
	$\theta_{gwb1dsp}$	0.2	0.198	0.194	0.093
	$\theta_{gwb2dsp}$	−0.2	−0.202	−0.216	0.131

Table 2
Resulting correlation means, medians and the standard deviations between the random effects and the predicted values.

	(σ_u^2, σ_v^2)	Mean	Median	SD		(σ_u^2, σ_v^2)	Mean	Median	SD
$\rho_{u,\hat{u}}$	(0.5, 0)	0.899	0.896	0.028	$\rho_{v,\hat{v}}$	(0, 0.5)	0.881	0.882	0.025
	(0.5, 0.5)	0.906	0.904	0.019		(0.5, 0.5)	0.901	0.902	0.020
	(0.5, 1.0)	0.915	0.913	0.0250		(1.0, 0.5)	0.918	0.917	0.029
	(1.0, 0)	0.907	0.904	0.022		(0, 1.0)	0.916	0.914	0.029
	(1.0, 0.5)	0.917	0.920	0.025		(0.5, 1.0)	0.908	0.916	0.031
	(1.0, 1.0)	0.938	0.942	0.027		(1.0, 1.0)	0.929	0.925	0.030

In Table 1, we give the results of the estimated coefficients for all possible network settings. Overall, we see an appropriate behavior of the estimates for all settings, which becomes more noticeable, especially when the heterogeneity levels in both modes increase.

Additionally, we look in Table 2 at the correlation results between the random effects $\boldsymbol{u}$, $\boldsymbol{v}$ and the predicted values $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$. This is only done for $\sigma_u^2 > 0$ and/or $\sigma_v^2 > 0$ because otherwise, the coefficients are equal to zero. We note a high positive correlation between the random effects and the predicted values in all possible settings. This can be quite an impactful point concerning the interpretability of the coefficients, especially concerning the degrees of the nodes in the two modes.

4. Examples

4.1. Southern Women data

As a first data example, we look at a well-known benchmark dataset in bipartite network analysis, the Southern Women data set (Davis et al., 1941), which is discussed and examined in various topics of bipartite network data analysis, see, for instance, Aitkin et al. (2014), and Wang et al. (2009) in more detail regarding the choice of model parameters in the Southern Women network data. This data set consists of 18 women observed over nine months period. During that period, these women met in a series of 14 informal social events. The data

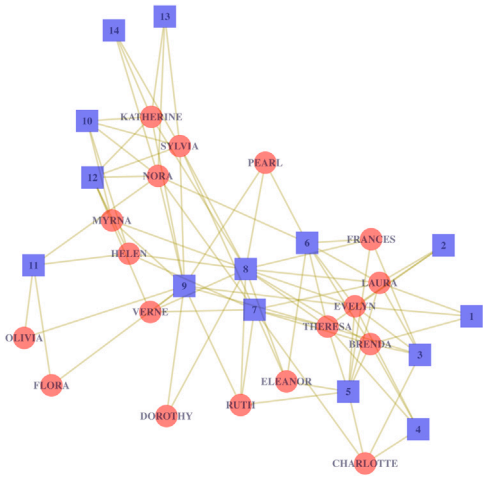


Fig. 2. Southern Women network data. The red (circle) nodes indicate the women in the network, and the blue (square) nodes the 14 different events. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Model fitting results and the t-ratios for the Southern Women network data. Standard errors listed in the biMERGM are not accurate since they ignore the variability resulting through node heterogeneity.

Parameter	Estimate	SE	$\lvert t-\text{ratio}\rvert$
Edge	−2.118	*0.267*	0.073
2-star (1 woman, 2 events)	0.113	*0.082*	0.088
2-star (2 women, 1 event)	0.238	*0.055*	0.009
Closure (2 women, k events)	−0.122	*0.093*	0.095

record which women met during which events. Fig. 2 shows the bipartite network graph of Southern Women network data. We can see that in this bipartite network, there are few blue nodes representing the events that have high degrees, and some red nodes representing the women that have high degrees, while the remaining nodes have few ties. Therefore, the presence of nodal heterogeneity seems to be a reasonable assumption. A typical ERGM for bipartite networks will include local configurations as shown in Fig. 3, which mostly satisfy the dependence assumption as discussed in Pattison and Robins (2004) for bipartite networks specifically and Hunter and Handcock (2006) in general. The labels *2-star* and *3-star* in Fig. 3 represent the "event-stars" and "woman-stars", respectively. The label *closure* denotes the geometrically weighted dyadwise shared partner distribution for actors and events, which can also be seen as local closure configuration. Considering that a dependence structure for the Southern Women bipartite network assumes that the attendance of woman i at event j is conditionally dependent on the attendance of woman i at other events k, as well as on the attendance of other women l at event j, we, therefore, fit a model to reveal or understand the attendance tendency of the southern women taking the attractivity of the women and the events explicitly into account.

In Table 3 we summarize the result of the fitted model to this data along with the associated t-ratios (in absolute value) for the estimated parameters as discussed in Section 2, whereby convergence is reached after six iteration steps. The parameters of the model suggest that while there is a tendency for women to attend multiple events and a tendency for many women to be in attendance at a given event, there is a descending tendency for two women to attend many events if they have already attended one of the events together. Additionally, in Table 4 and in Table 5 we give the prediction results of attractiveness for women and events, respectively. The predicted values can be interpreted as follows: the higher the predicted values are, the more likely is it for a woman to attend and for a meeting to be attended.

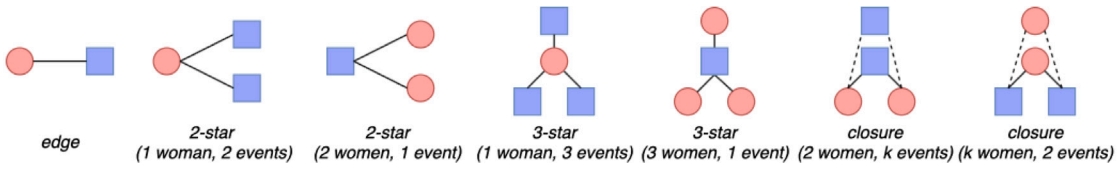


Fig. 3. Configurations for bipartite graph models.

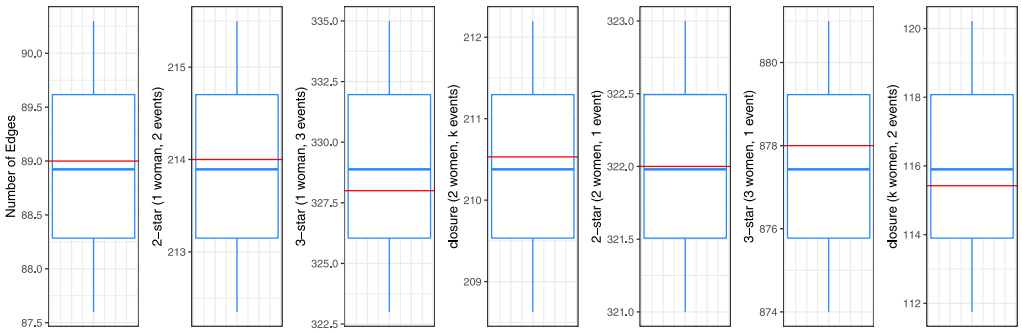


Fig. 4. Distribution of network characteristics and statistics for $B = 1000$ networks, sampled from the estimated model. Network statistics of the observed network in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Predicted attractivity of the women resulting from the model fit.

Woman	Predicted attractivity	Woman	Predicted attractivity
BRENDA	0.251	LAURA	0.251
CHARLOTTE	−0.094	MYRNA	−0.117
DOROTHY	−0.361	NORA	0.367
ELEANOR	−0.129	OLIVIA	−0.305
EVELYN	0.363	PEARL	−0.246
FLORA	−0.304	RUTH	−0.134
FRANCES	−0.110	SYLVIA	0.235
HELEN	0.002	THERESA	0.355
KATHERINE	0.138	VERNE	−0.137

Table 5
Predicted attractivity of the events resulting from the model fit.

Event	Predicted attractivity	Event	Predicted attractivity
1	−0.567	8	1.268
2	−0.573	9	0.973
3	−0.043	10	−0.225
4	−0.373	11	−0.344
5	0.264	12	−0.065
6	0.259	13	−0.572
7	0.579	14	−0.572

There is a strong correlation between the node degrees and the fitted random effects resulting from the model. For instance, we predicted for Dorothy an attractivity of $\hat{u}_{\text{Dorothy}} = -0.361$ and for Nora $\hat{u}_{\text{Nora}} = 0.367$; looking at the graph in Fig. 2, we can see that Dorothy attended only two events, while Nora attended many events. On the other hand, we predicted event 1 as one of the most unattractive events and event 8 as the most attractive one, which is also consistent with the graph. The fitted random women effects $\hat{\boldsymbol{u}}$, and the event effects $\hat{\boldsymbol{v}}$ are provided in Tables 4 and 5, respectively.

Fig. 4 visualizes the results of the described model evaluation approach in Section 2 based on $B = 1000$ network samples. We computed for each of the sampled and the observed networks selected network characteristics and statistics as stated in Fig. 3. The distribution of the statistics of the sampled networks is presented by boxplots; the network statistics on the observed networks are presented by a red line. In general, it seems that sampling from our proposed models produces networks that show a very similar structure to the true, observed ones.

4.2. World City Network

As a second data example, we consider the "World City Network" data introduced in Taylor et al. (2002), consisting of 100 global service firms distributed across 315 cities worldwide. All firms in this network supply advanced producer services (accountancy, advertising, banking/finance, insurance, law, and management consultancy) through offices in at least 15 cities (including at least one in Pacific Asia, western Europe, and northern America). An edge in this network represents the presence of a given firm in a given city. Fig. 5 provides a visual impression of the network. Since additional nodal attributes are available as the type of the firms, we investigate whether an edge formation in this network exhibits homophilic behavior. In other words, we want to fit a model which answers questions like, given the attractiveness of cities and firms, do firms of the same type tend to co-locate in the same cities more or less often than we would expect by chance? This question will be answered with our proposed biMERGM model, for which we take the heterogeneity of the nodes, aka the random effects (which counts for the attractiveness of the cities and firms) and the homophily network statistics, aka the attributes of the firms (which counts for the co-location of the same firm types in the same cities) into account. Additionally, we include a dyad-dependent endogenous network statistic such as the closure of two firms in k different cities as in the example before. In Table 6, we summarize the results of the fitted model, whereby convergence is reached after 11 iteration steps. The network statistic `nodematch` (see Handcock et al., 2017) can be obtained, for instance, by summing $y_{ik}y_{jk}$ for all firm nodes of the same type i and j and all city nodes k.

$$\text{nodematch}(y) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} y_{ik}y_{jk}I\{i \text{ matches } j,\ i \neq j\} \tag{8}$$

The indicator function in the formula above is one iff i and j are from the same mode and share the same type of node; for instance, this is one for firms i, and j of type let us say "Accountancy", and zero otherwise. With some adjustments, we can look at the matching nodes of each firm type separately. Each of these statistics in our model gives half the sum of the number of edges that are part of at least one two-path joining two firms of the same type. For instance, the conditional log-odds of an accountancy firm locating in a particular large city in the world where no closure between firms and cities exists

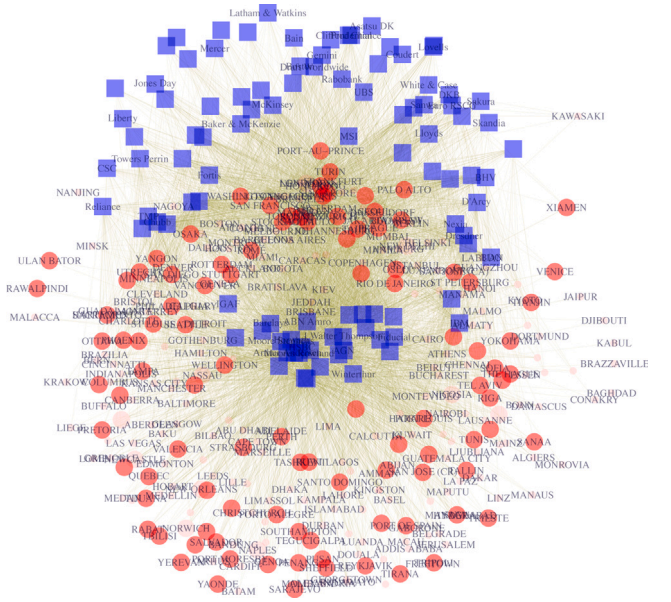


Fig. 5. World City Network graph. The red (circle) nodes indicate the 315 cities in the network, while the blue (square) nodes the 100 firms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Model fitting results and the t-ratios for the World City Network data. Standard errors are not accurate since they ignore the variability resulting through node heterogeneity.

Parameter	Estimate	SE	$\lvert t-\text{ratio}\rvert$
Edge	−1.976	*0.037*	0.0564
Closure (2 firms, k cities)	0.079	*0.018*	0.096
Nodematch (firmtype: Accountancy)	0.274	*0.092*	0.055
Nodematch (firmtype: Advertising)	0.488	*0.101*	0.082
Nodematch (firmtype: Banking/Finance)	0.201	*0.090*	0.066
Nodematch (firmtype: Insurance)	−0.051	*0.111*	0.083
Nodematch (firmtype: Law)	0.441	*0.101*	0.097
Nodematch (firmtype: Management, Consultancy)	0.562	*0.090*	0.084

is $-1.976+(0.274\times 0.5)=-1.839$ if another firm is already there. There is also a slight tendency for firms to locate in different cities if these firms have a pre-existing location in a given city. In Fig. 6 we present the predicted attractivity values $\hat{u}_i$ of the cities visualized on a world map. We can identify significant clusters of attractive cities in central Europe and the eastern USA and some very attractive cities in east Asia and Australia. We demonstrate the predicted attractivity of the firms $\hat{v}_j$ in Fig. 7. Each firm is categorized according to its corresponding type. For instance, some accounting firms have high attractiveness, whereas law firms generally have low attractiveness. Additionally, we provide distribution comparisons of predicted attractivity and the firms' degree, distinguishing between their corresponding categories as supplementary material.

Finally, we run a simulation-based model evaluation as proposed in Section 2. Fig. 8 presents the results based on $B = 1000$ network samples. The performance of our model is convincing across many network statistics and characteristics.

4.3. Patent data

We now consider a rather extensive network resulting from patent data. This data includes patent applications filed to the European Patent Office (EPO) and the German Patent and Trademark Office (DPMA), respectively, in 2015 in the main area of electrical engineering. The data includes patents that listed at least one inventor residing in German territory. We observe a total of 10 251 inventors who have applied to 8206 different patents in the field of electrical engineering, which itself can be divided into eight sub-areas. Table 7 gives the frequency of patents in these eight sub-areas. Despite the computational intensity, we are able to fit the proposed model (see Table 8) to this extensive bipartite network. The computation is carried out on a Linux cluster with eight nodes and 28 tasks per node. Convergence is achieved after 21 iterations. The computation time for one iteration was about three hours, i.e., it took approximately three days for 21 iterations. The research question that we aim to answer is the following: which of the sub-areas is most seclusive in the sense that inventors with multiple patents tend to have them in the same sub-area. We, therefore, include nodematch statistics comparable to (8), but this time the nodematch is the other component, i.e. on the patents, i.e. two patents submitted by one or more joint inventors in the same sub-area. The estimated coefficients are shown in Table 8, although some structural effects have a t-ratio slightly above the 0.1 threshold, we decided to keep these structural effects due to the size of the network and its' computational expensiveness. The model's coefficients point out a strong presence of seclusiveness, which means inventors with more than one patent tend more likely to file their patents in the same sub-area, i.e. all coefficients are positive. The effect is strongest for the sub-area "Electrical Energy" while for "IT Methods", the effect is small and not even significantly positive. Hence, the sub-area of "IT Methods" is hardly seclusive, meaning that inventors having a patent there are likely to have another patent in some other sub-area.

Since we also know the location of the inventors, we can visualize in Fig. 9 the predicted attractivity $\hat{u}_i$ of the inventors according to the location of each inventor. We can conclude that the huge cluster in southern Germany is not surprisingly in the south of Germany, which is the driving force of the economy in Germany with many large firms and excellent universities, and this pattern seems to be valid for inventors who locate in large cities in Germany. Since the patents filed in the main area of electrical engineering are categorized into eight different areas, we visualize the predicted attractiveness of the patents $\hat{v}_j$ according to their area in Fig. 10. We identify the area of "Semiconductors" as one of the most attractive patent areas in the data based on our model prediction. Similar to the examples before, we evaluate our model based on 1000 network samples for different network characteristics and statistics, as in Fig. 11 demonstrated. The overall evaluation of the model is very promising.

Table 7
Frequency of patents according to the areas in the data.

Area	Frequency	Area	Frequency
Audiovisual	631	Communication process	271
Computer technology	937	Digital communication	578
Electrical energy	3969	IT methods	134
Semiconductors	903	Telecommunication	783

Table 8
Model fitting results and the t-ratios for the Patent Network data. Standard errors listed in the biMERGM are not accurate since they ignore the variability resulting through node heterogeneity.

Parameter	Estimate	SE	$\lvert t-\text{ratio}\rvert$
Edge	−7.213	*0.031*	0.097
Closure (2 inventors, k patents)	0.942	*0.082*	0.110
Nodematch (area: Audiovisual)	1.251	*0.099*	0.092
Nodematch (area: Computer Technology)	1.810	*0.068*	0.236
Nodematch (area: Electrical Energy)	2.377	*0.166*	0.096
Nodematch (area: Semiconductors)	1.288	*0.182*	0.092
Nodematch (area: Communication Process)	0.633	*0.101*	0.065
Nodematch (area: Digital Communication)	0.519	*0.173*	0.077
Nodematch (area: IT Methods)	0.129	*0.671*	0.124
Nodematch (area: Telecommunication)	0.969	*0.181*	0.092

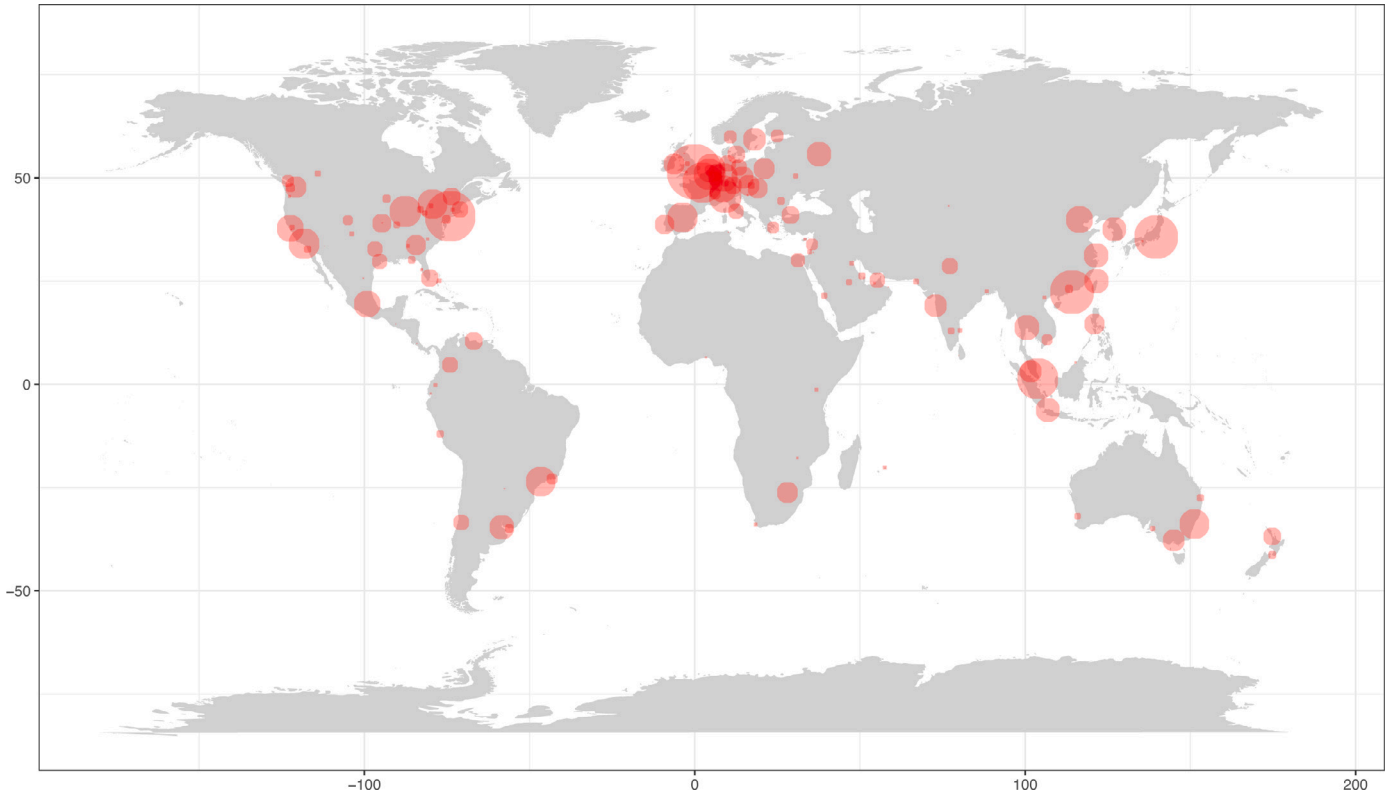


Fig. 6. Predicted attractivity of the cities. Big transparent nodes indicate cities with high attractivity.

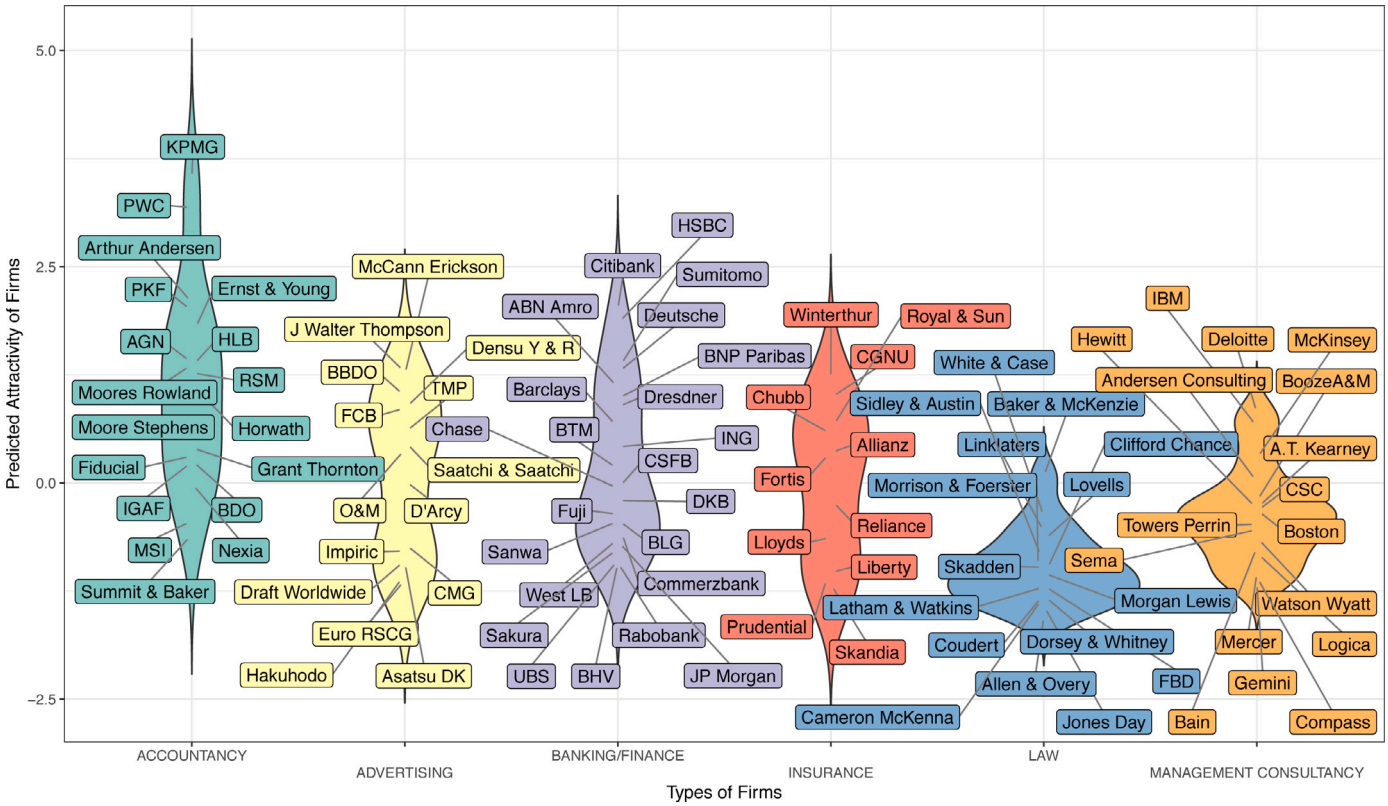


Fig. 7. Predicted attractivity of the firms, categorized according to the type of each firm.

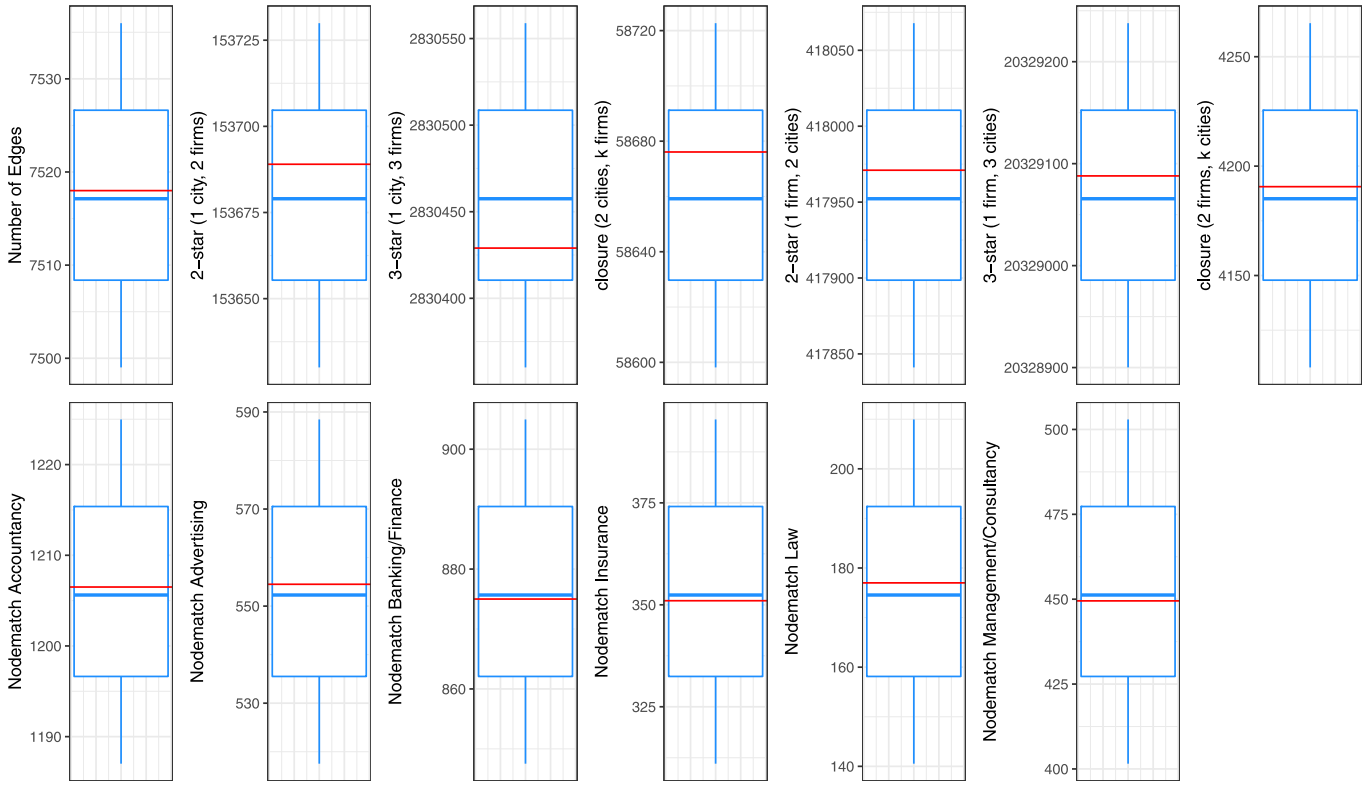


Fig. 8. Distribution of network statistics for $B = 1000$ networks sampled from the estimated model. Network statistics of the observed network in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

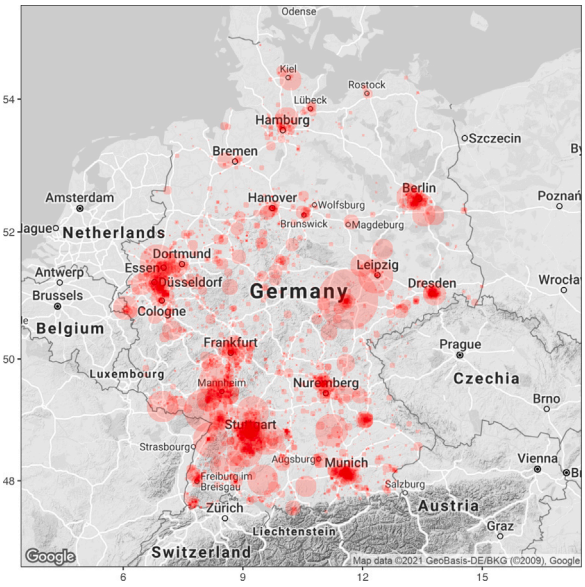


Fig. 9. Predicted attractivity of the inventors. Big transparent nodes indicate inventors with high attractivity.

5. Discussion

The inclusion of node-specific covariates usually explains the heterogeneity in bipartite networks see, e.g., Lusher et al. (2013). However, in this paper, we argue that often heterogeneity cannot be fully explained by covariates. Therefore, our model allows for the inclusion of random heterogeneity effects, and the proposed algorithm is scalable towards the analysis of large bipartite networks, which is demonstrated through simulations and examples.

While the model, as well as its estimation, provide promising results, some open issues remain. As pursued in this paper, model evaluation is simple and could be extended towards more sophisticated routines, for instance, based on the Akaike Information Criterion. This, however, would require several approximations to make it numerically feasible so that we consider this beyond the scope of this paper. Secondly, the standard errors given are approximate based on the model where coefficients $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ are treated as offsets. Exact calculations are theoretically as well as numerically demanding, so that further research is required. For instance, a proper numerically efficient bootstrap would be desirable to obtain appropriate standard errors. Overall, we conclude that the biMERGM proposed and fitted in this paper can provide a valuable and numerically feasible tool for analyzing large bipartite networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors also confirm that each author has seen and approved the contents of the submitted manuscript.

Funding

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

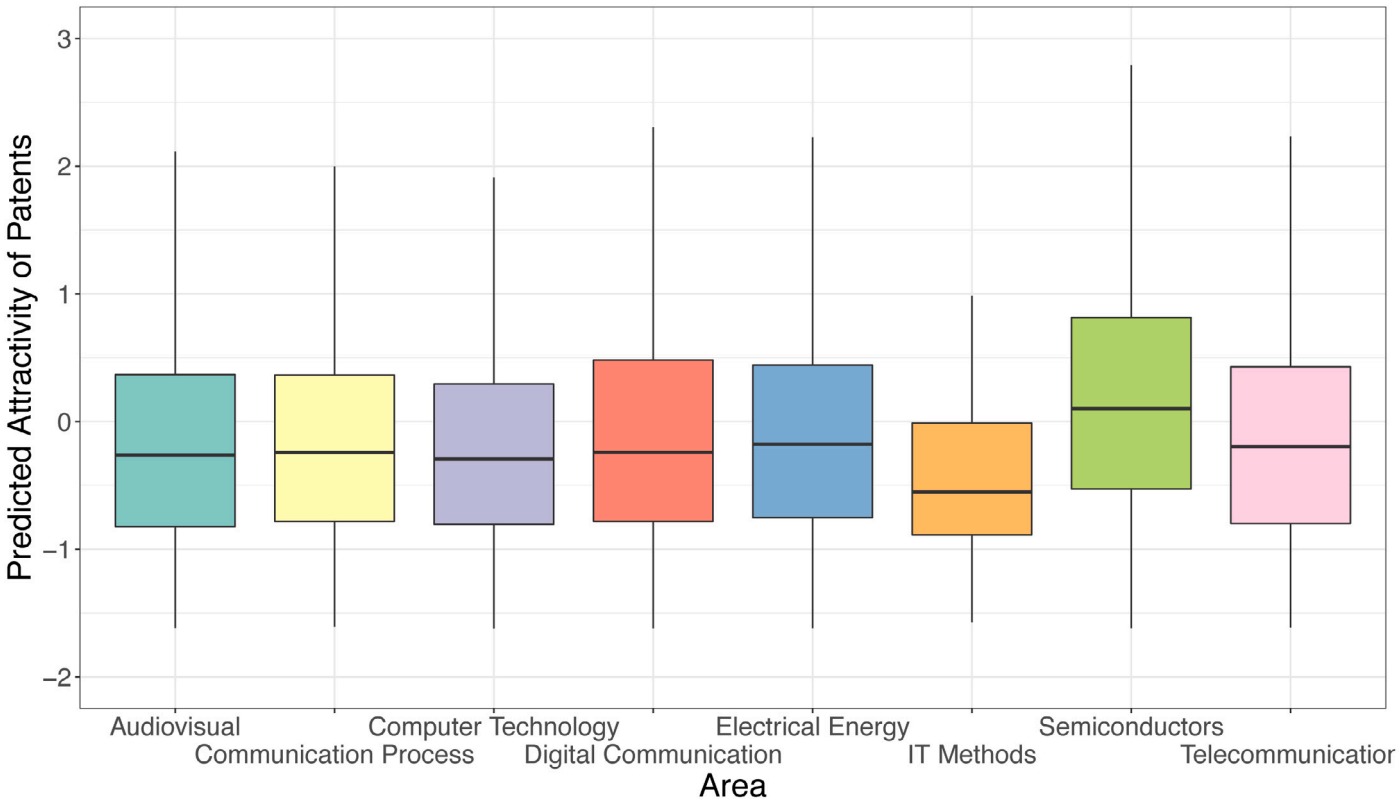


Fig. 10. Predicted attractivity of patents, categorized according to each patent's area.

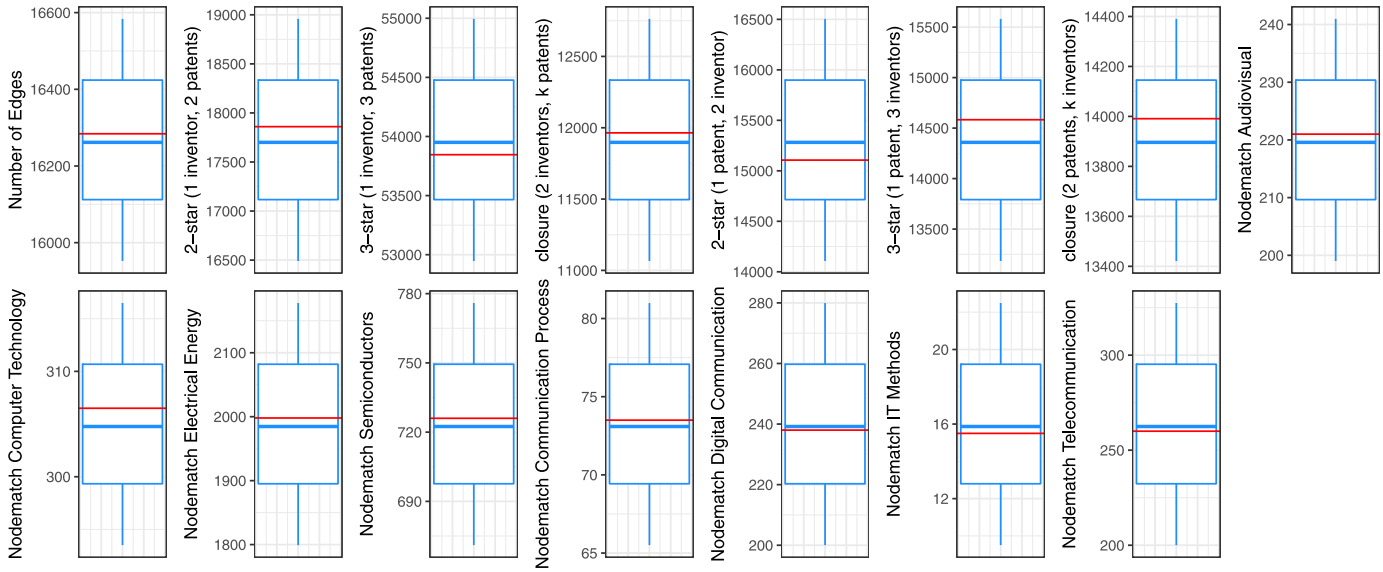


Fig. 11. Distribution of network statistics for $B = 1000$ networks sampled from the estimated model. Network statistics of the observed network in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socnet.2021.11.002.

References

Agneessens, F., Roose, H., 2008. Local structural properties and attribute characteristics in 2-mode networks: p^* models to map choices of theater events. J. Math. Sociol. 32 (3), 204–237.

Aitkin, M., Vu, D., Francis, B., 2014. Statistical modelling of the group structure of social networks. Social Networks 38, 74–87.

Borgatti, S.P., Everett, M.G., 1997. Network analysis of 2-mode data. Social Networks 19 (3), 243–269.

Box-Steffensmeier, J.M., Christenson, D.P., Morgan, J.W., 2018. Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. Polit. Anal. 26 (1), 3–19.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88 (421), 9–25.
Davis, A., Gardner, B.B., Gardner, M.R., Warner, W.L., 1941. Deep South: A Sociological Anthropological Study of Caste and Class. University of Chicago Press.
Duijn, M.A.J., Gile, K.J., Handcock, M.S., 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Social Networks 31 (1), 52–62.
Duijn, M.A.J., Snijders, T.A.B., Zijlstra, B.J.H., 2004. p_2: A Random effects model with covariates for directed graphs. Stat. Neerl. 58 (2), 234–254.
Frank, O., Strauss, D., 1986. Markov graphs. J. Amer. Statist. Assoc. 81 (395), 832–842.
Handcock, M.S., Gile, K.J., 2010. Modeling social networks from sampled data. Ann. Appl. Stat. 4 (1), 5.
Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N., Morris, M., 2017. Ergm: Fit, simulate and diagnose exponential-family models for networks. In: The Statnet Project. R Package Version, 3(0), (http://www.statnet.org).
Hanneke, S., Fu, W., Xing, E.P., et al., 2010. Discrete temporal models of social networks. Electron. J. Stat. 4, 585–605.
Hummel, R.M., Hunter, D.R., Handcock, M.S., 2012. Improving simulation-based algorithms for fitting ERGMs. J. Comput. Graph. Statist. 21 (4), 920–939.
Hunter, D.R., Handcock, M.S., 2006. Inference in curved exponential family models for networks. J. Comput. Graph. Statist. 15 (3), 565–583.
Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M., 2008. Ergm: A package to fit, simulate and diagnose exponential-family models for networks. J. Stat. Softw. 24 (3), 1–29.
Hunter, D.R., Krivitsky, P.N., Schweinberger, M., 2012. Computational statistical methods for social network analysis. J. Comput. Graph. Statist. 21 (4), 856–882.
Kevork, S., Kauermann, G., 2019. Iterative estimation of mixed exponential random graph models with nodal random effects. arXiv preprint arXiv:1911.02397.
Latapy, M., Magnien, C., Del Vecchio, N., 2008. Basic notions for the analysis of large two-mode networks. Social Networks 30 (1), 31–48.
Lusher, D., Koskinen, J., Robins, G., 2013. Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications, Vol. 35. Cambridge University Press.
Pattison, P., Robins, G., 2004. Building models for social space: Neighourhood-based models for social networks and affiliation structures. Math. Sci. Hum. Math. Soc. Sci. (168).
Rasch, G., 1961. On general laws and the meaning of measurement in psychology. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, p. 321–333.
Shi, C., Li, Y., Zhang, J., Sun, Y., Yu, P.S., 2017. A survey of heterogeneous information network analysis. IEEE Trans. Knowl. Data Eng. 29 (1), 17–37.
Snijders, T.A.B., 2002. Markov chain Monte Carlo estimation of exponential random graph models. J. Soc. Struct. 3 (2), 1–40.
Snijders, T.A.B., Van Duijn, M.A., 2002. Conditional maximum likelihood estimation under various specifications of exponential random graph models. Contrib. Soc. Netw. Anal. Inf. Theory Other Top. Statist. 117–134.
Taylor, P., Catalano, G., Walker, D., 2002. Measurement of the world city network. Urban Stud. 39 (13), 2367–2376.
Thiemichen, S., Friel, N., Caimo, A., Kauermann, G., 2016. BayesIan exponential random graph models with nodal random effects. Social Networks 46, 11–28.
Wang, P., Robins, G., Matous, P., 2016. Multilevel network analysis using ERGM and its extension. In: Multilevel Network Analysis for the Social Sciences. Springer, pp. 125–143.
Wang, P., Robins, G., Pattison, P., Lazega, E., 2013. Exponential random graph models for multilevel networks. Social Networks 35 (1), 96–115.
Wang, P., Sharpe, K., Robins, G.L., Pattison, P.E., 2009. Exponential random graph ($p\star$) models for affiliation networks. Social Networks 31 (1), 12–25.
Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Statist. Soc. (B) 73 (1), 3–36.
Zijlstra, B.J.H., Duijn, M.A.J., Snijders, T.A.B., 2006. The multilevel p_2 model: A random effects model for the analysis of multiple social networks. Methodol.: Eur. J. Res. Methods Behav. Soc. Sci. 2 (1), 42–47.

Appendix

Further Simulation Results

In the simulation study of Chapter 2, we construct our simulation setting with fixed $\boldsymbol{\theta}$ values where $\boldsymbol{\theta} = (\theta_{edges}, \theta_{gwesp}, \theta_{2-stars}) = (-1, 0.2, -0.3)$, and looking at six different heterogeneity levels $\sigma_u^2 = (0, 0.1, 0.2, 0.5, 0.8, 1)$ for both network sizes small (50 nodes) and large (500 nodes). However, to demonstrate the scalability/flexibility of the model, we provide the results of two more simulation studies, where first we perturb just the θ_{gwesp} parameter changing its value from 0.2 to 0.5, and second, we perturb the whole parameter vector $\boldsymbol{\theta}$ setting the values to $\boldsymbol{\theta} = (\theta_{edges}, \theta_{gwesp}, \theta_{2-stars}) = (-1.5, -0.3, -0.4)$. Furthermore, we also provide illustrations of how well σ_u^2 is recovered in the simulation study of Chapter 2.

Predicted Attractivity Versus Degree Distribution Comparison

We illustrate distribution comparisons of predicted attractivity and the firms' degree, distinguishing between their corresponding categories from the "World City Network" example in Chapter 3. With this plot, we demonstrate the predicted random effects' performance compared with the actual degrees of the nodes (firms).

Network Size: 50 Nodes								
σ_u^2	Model type	Parameter	Real Value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	-1	-0.79	0.23	-1.06	-0.77	-0.53
		θ_{gwesp}	0.5	0.20	0.19	-0.02	0.19	0.45
		$\theta_{2-stars}$	-0.3	-0.25	0.14	-0.42	-0.26	-0.07
0	mERGM	θ_{edges}	-1	-0.78	0.17	-0.97	-0.81	-0.59
		θ_{gwesp}	0.5	0.13	0.17	-0.05	0.11	0.32
		$\theta_{2-stars}$	-0.3	-0.23	0.22	-0.47	-0.27	-0.08
0.1	ERGM	θ_{edges}	-1	-1.38	0.42	-1.93	-1.39	-0.91
		θ_{gwesp}	0.5	0.33	0.30	-0.08	0.35	0.73
		$\theta_{2-stars}$	-0.3	-0.24	0.27	-0.55	-0.32	0.16
0.1	mERGM	θ_{edges}	-1	-0.68	0.15	-0.87	-0.68	-0.49
		θ_{gwesp}	0.5	0.38	0.25	0.04	0.36	0.67
		$\theta_{2-stars}$	-0.3	-0.21	0.23	-0.52	-0.18	-0.04
0.2	ERGM	θ_{edges}	-1	-1.49	0.41	-1.98	-1.5	-0.93
		θ_{gwesp}	0.5	0.39	0.22	0.12	0.33	0.72
		$\theta_{2-stars}$	-0.3	-0.17	0.16	-0.43	-0.14	0.04
0.2	mERGM	θ_{edges}	-1	-0.75	0.11	-0.89	-0.76	-0.61
		θ_{gwesp}	0.5	0.42	0.11	0.28	0.43	0.56
		$\theta_{2-stars}$	-0.3	-0.35	0.12	-0.49	-0.36	-0.23
0.5	ERGM	θ_{edges}	-1	-1.84	0.38	-2.24	-1.85	-1.31
		θ_{gwesp}	0.5	0.37	0.08	0.24	0.37	0.48
		$\theta_{2-stars}$	-0.3	-0.01	0.14	-0.17	-0.01	0.19
0.5	mERGM	θ_{edges}	-1	-0.86	0.08	-0.96	-0.86	-0.78
		θ_{gwesp}	0.5	0.46	0.05	0.40	0.45	0.52
		$\theta_{2-stars}$	-0.3	-0.31	0.02	-0.34	-0.31	-0.28
0.8	ERGM	θ_{edges}	-1	-3.1	0.53	-3.78	-3.07	-2.45
		θ_{gwesp}	0.5	0.47	0.11	0.36	0.47	0.61
		$\theta_{2-stars}$	-0.3	0.05	0.26	-0.12	0.05	0.41
0.8	mERGM	θ_{edges}	-1	-0.97	0.02	-1.01	-0.98	-0.95
		θ_{gwesp}	0.5	0.49	0.03	0.43	0.48	0.53
		$\theta_{2-stars}$	-0.3	-0.31	0.02	-0.34	-0.31	-0.28
1	ERGM	θ_{edges}	-1	-6385.68	25525.91	-162.93	-21.29	3.61
		θ_{gwesp}	0.5	-14204.17	56808.78	-698.11	-13.01	1.21
		$\theta_{2-stars}$	-0.3	298.89	1678.24	-2099.88	-2.89	98.56
1	mERGM	θ_{edges}	-1	-1.09	0.02	-1.12	-1.09	-1.06
		θ_{gwesp}	0.5	0.51	0.04	0.45	0.50	0.56
		$\theta_{2-stars}$	-0.3	-0.31	0.01	-0.33	-0.32	-0.30

Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 50 nodes and for all six σ_u^2 levels. In this setting, only the GWESP parameter is manipulated compared to the setting in the manuscript.

Network Size: 500 Nodes								
σ_u^2	Model type	Parameter	Real Value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	-1	-1.08	0.19	-1.26	-1.01	-0.75
		θ_{gwesp}	0.5	0.48	0.19	0.23	0.48	0.74
		$\theta_{2-stars}$	-0.3	-0.31	0.01	-0.32	-0.30	-0.29
0	mERGM	θ_{edges}	-1	-0.70	0.15	-0.90	-0.70	-0.51
		θ_{gwesp}	0.5	0.46	0.02	0.43	0.46	0.50
		$\theta_{2-stars}$	-0.3	-0.18	0.14	-0.36	-0.18	-0.001
0.1	ERGM	θ_{edges}	-1	-1.30	0.25	-1.61	-1.31	-0.97
		θ_{gwesp}	0.5	0.45	0.08	0.34	0.45	0.57
		$\theta_{2-stars}$	-0.3	-0.23	0.15	-0.42	-0.23	-0.03
0.1	mERGM	θ_{edges}	-1	-0.77	0.19	-1.01	-0.77	-0.53
		θ_{gwesp}	0.5	0.46	0.03	0.41	0.46	0.50
		$\theta_{2-stars}$	-0.3	-0.20	0.12	-0.35	-0.20	-0.05
0.2	ERGM	θ_{edges}	-1	-1.54	0.21	-1.81	-1.54	-1.27
		θ_{gwesp}	0.5	0.45	0.08	0.33	0.45	0.56
		$\theta_{2-stars}$	-0.3	-0.20	0.09	-0.33	-0.20	-0.07
0.2	mERGM	θ_{edges}	-1	-0.70	0.16	-0.92	-0.70	-0.48
		θ_{gwesp}	0.5	0.44	0.08	0.34	0.44	0.54
		$\theta_{2-stars}$	-0.3	-0.38	0.12	-0.54	-0.38	-0.23
0.5	ERGM	θ_{edges}	-1	-1.88	0.35	-2.34	-1.87	-1.43
		θ_{gwesp}	0.5	0.48	0.07	0.39	0.47	0.56
		$\theta_{2-stars}$	-0.3	-0.09	0.20	-0.34	-0.07	0.18
0.5	mERGM	θ_{edges}	-1	-0.67	0.10	-0.80	-0.68	-0.54
		θ_{gwesp}	0.5	0.51	0.05	0.44	0.51	0.58
		$\theta_{2-stars}$	-0.3	-0.32	0.01	-0.34	-0.32	-0.31
0.8	ERGM	θ_{edges}	-1	-3.90	0.43	-4.47	-3.87	-3.36
		θ_{gwesp}	0.5	0.45	0.17	0.20	0.44	0.68
		$\theta_{2-stars}$	-0.3	-0.001	0.16	-0.21	-0.005	0.20
0.8	mERGM	θ_{edges}	-1	-1.24	0.04	-1.29	-1.24	-1.19
		θ_{gwesp}	0.5	0.48	0.01	0.45	0.48	0.50
		$\theta_{2-stars}$	-0.3	-0.32	0.01	-0.34	-0.32	-0.31
1	ERGM	θ_{edges}	-1	-5.11	0.14	-5.30	-5.10	-4.93
		θ_{gwesp}	0.5	0.43	0.06	0.33	0.42	0.52
		$\theta_{2-stars}$	-0.3	-0.004	0.10	-0.13	-0.006	0.12
1	mERGM	θ_{edges}	-1	-0.98	0.01	-0.99	-0.98	-0.96
		θ_{gwesp}	0.5	0.51	0.03	0.46	0.50	0.55
		$\theta_{2-stars}$	-0.3	-0.32	0.009	-0.33	-0.31	-0.30

Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 500 nodes and for all six σ_u^2 levels. In this setting, only the GWESP parameter is manipulated compared to the setting in the manuscript.

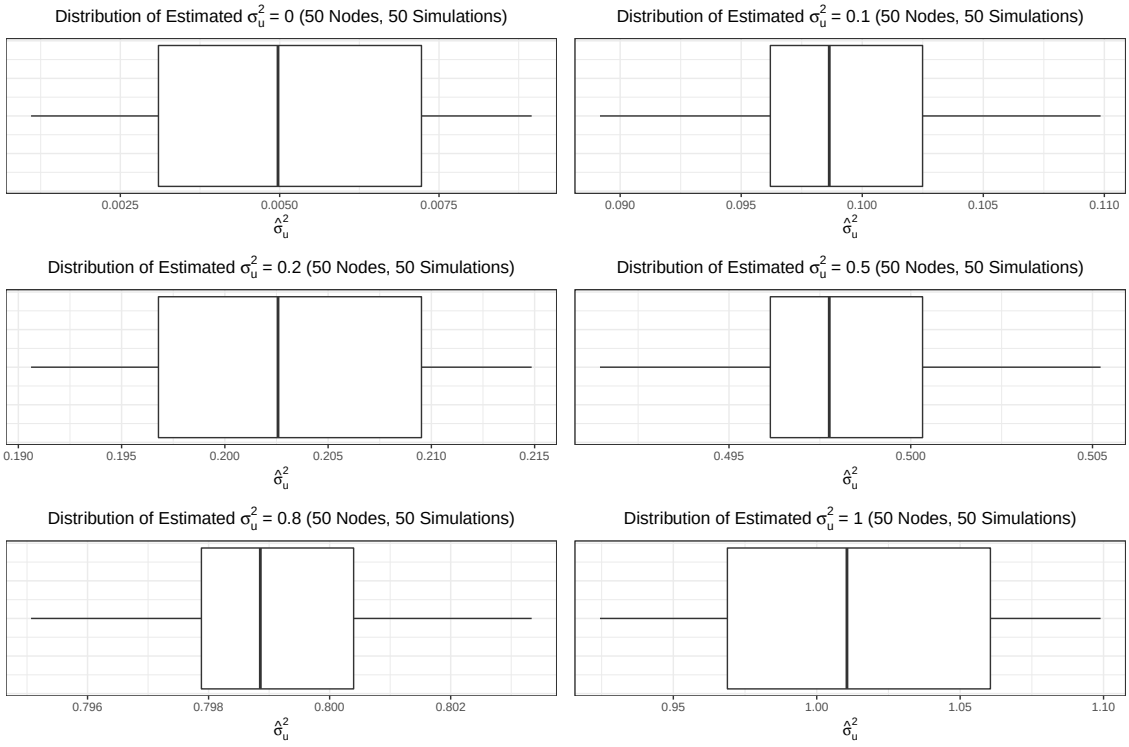
Network Size: 50 Nodes								
σ_u^2	Model type	Parameter	Real Value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	-1.5	-1.34	0.11	-1.48	-1.35	-1.20
		θ_{gwesp}	-0.3	-0.24	0.18	-0.47	-0.26	-0.004
		$\theta_{2-stars}$	-0.4	-0.36	0.13	-0.52	-0.37	-0.17
0	mERGM	θ_{edges}	-1.5	-1.30	0.17	-1.49	-1.33	-1.11
		θ_{gwesp}	-0.3	-0.24	0.25	-0.52	-0.28	0.04
		$\theta_{2-stars}$	-0.4	-0.38	0.22	-0.61	-0.41	-0.14
0.1	ERGM	θ_{edges}	-1.5	-1.18	0.42	-1.72	-1.18	-0.71
		θ_{gwesp}	-0.3	-0.32	0.31	-0.742	-0.31	0.071
		$\theta_{2-stars}$	-0.4	-0.37	0.22	-0.63	-0.44	-0.02
0.1	mERGM	θ_{edges}	-1.5	-1.11	0.23	-1.41	-1.11	-0.81
		θ_{gwesp}	-0.3	-0.22	0.11	-0.36	-0.22	-0.08
		$\theta_{2-stars}$	-0.4	-0.36	0.12	-0.53	-0.34	-0.22
0.2	ERGM	θ_{edges}	-1.5	-1.17	0.41	-1.66	-1.18	-0.61
		θ_{gwesp}	-0.3	-0.27	0.13	-0.43	-0.31	0.07
		$\theta_{2-stars}$	-0.4	-0.17	0.15	-0.40	-0.15	0.02
0.2	mERGM	θ_{edges}	-1.5	-1.61	0.10	-1.74	-1.62	-1.47
		θ_{gwesp}	-0.3	-0.24	0.15	-0.42	-0.24	-0.05
		$\theta_{2-stars}$	-0.4	-0.44	0.12	-0.59	-0.44	-0.31
0.5	ERGM	θ_{edges}	-1.5	-0.56	0.35	-0.93	-0.56	-0.07
		θ_{gwesp}	-0.3	-0.35	0.06	-0.43	-0.35	-0.26
		$\theta_{2-stars}$	-0.4	-0.68	0.13	-0.83	-0.68	-0.49
0.5	mERGM	θ_{edges}	-1.5	-1.62	0.07	-1.72	-1.63	-1.54
		θ_{gwesp}	-0.3	-0.36	0.02	-0.39	-0.36	-0.33
		$\theta_{2-stars}$	-0.4	-0.34	0.03	-0.38	-0.34	-0.30
0.8	ERGM	θ_{edges}	-1.5	0.26	0.51	-0.41	0.26	0.85
		θ_{gwesp}	-0.3	-0.20	0.04	-0.24	-0.20	-0.14
		$\theta_{2-stars}$	-0.4	0.04	0.19	-0.23	0.04	0.30
0.8	mERGM	θ_{edges}	-1.5	-1.54	0.03	-1.58	-1.54	-1.50
		θ_{gwesp}	-0.3	-0.27	0.02	-0.32	-0.27	-0.22
		$\theta_{2-stars}$	-0.4	-0.38	0.01	-0.40	-0.38	-0.37
1	ERGM	θ_{edges}	-1.5	-4637.406	3354.11	-8562.84	-4337.4	-498.63
		θ_{gwesp}	-0.3	24081.52	65130.7	-56617.06	31760.1	99770.83
		$\theta_{2-stars}$	-0.4	19111.89	3275.97	14506.83	19131	23475.23
1	mERGM	θ_{edges}	-1.5	-1.52	0.01	-1.54	-1.52	-1.49
		θ_{gwesp}	-0.3	-0.32	0.01	-0.34	-0.32	-0.30
		$\theta_{2-stars}$	-0.4	-0.39	0.08	-0.51	-0.38	-0.29

Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 50 nodes and for all six σ_u^2 levels. In this setting, the whole parameter $\boldsymbol{\theta}$ is manipulated compared to the setting in the manuscript.

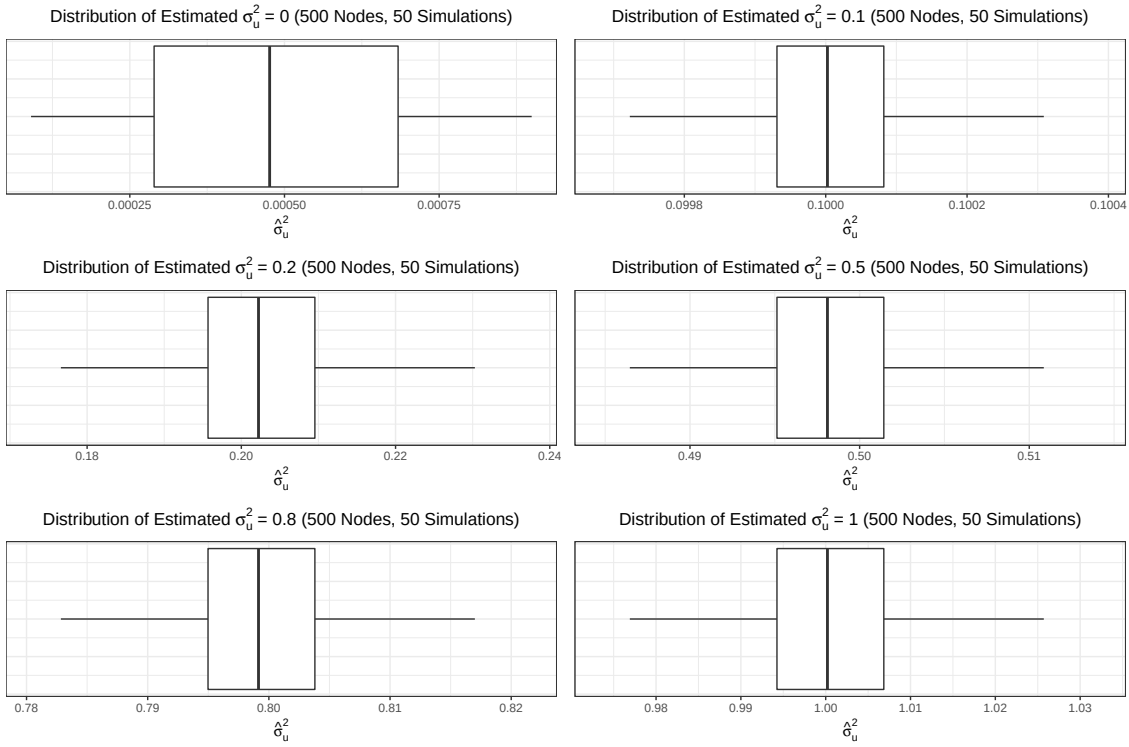
Network Size: 500 Nodes								
σ_u^2	Model type	Parameter	Real Value	Mean	SD	Q 0.1	Median	Q 0.9
0	ERGM	θ_{edges}	-1.5	-1.46	0.08	-1.58	-1.46	-1.35
		θ_{gwesp}	-0.3	-0.27	0.11	-0.42	-0.27	-0.12
		$\theta_{2-stars}$	-0.4	-0.37	0.09	-0.50	-0.37	-0.25
0	mERGM	θ_{edges}	-1.5	-1.38	0.19	-1.63	-1.38	-1.14
		θ_{gwesp}	-0.3	-0.25	0.21	-0.54	-0.25	0.01
		$\theta_{2-stars}$	-0.4	-0.36	0.09	-0.48	-0.35	-0.23
0.1	ERGM	θ_{edges}	-1.5	-1.26	0.32	-1.66	-1.27	-0.84
		θ_{gwesp}	-0.3	-0.27	0.09	-0.39	-0.27	-0.14
		$\theta_{2-stars}$	-0.4	-0.34	0.11	-0.49	-0.35	-0.20
0.1	mERGM	θ_{edges}	-1.5	-1.25	0.24	-1.56	-1.26	-0.95
		θ_{gwesp}	-0.3	-0.22	0.14	-0.39	-0.23	-0.03
		$\theta_{2-stars}$	-0.4	-0.32	0.10	-0.45	-0.33	-0.20
0.2	ERGM	θ_{edges}	-1.5	-1.00	0.31	-1.39	-1.01	-0.60
		θ_{gwesp}	-0.3	-0.26	0.09	-0.39	-0.26	-0.13
		$\theta_{2-stars}$	-0.4	-0.31	0.11	-0.45	-0.30	-0.17
0.2	mERGM	θ_{edges}	-1.5	-1.20	0.11	-1.35	-1.21	-1.05
		θ_{gwesp}	-0.3	-0.23	0.10	-0.36	-0.23	-0.10
		$\theta_{2-stars}$	-0.4	-0.21	0.11	-0.35	-0.20	-0.07
0.5	ERGM	θ_{edges}	-1.5	-2.03	0.41	-2.57	-2.02	-1.50
		θ_{gwesp}	-0.3	-0.33	0.09	-0.45	-0.33	-0.22
		$\theta_{2-stars}$	-0.4	-0.71	0.21	-0.97	-0.69	-0.42
0.5	mERGM	θ_{edges}	-1.5	-1.69	0.11	-1.85	-1.71	-1.56
		θ_{gwesp}	-0.3	-0.33	0.05	-0.40	-0.33	-0.26
		$\theta_{2-stars}$	-0.4	-0.37	0.02	-0.40	-0.37	-0.34
0.8	ERGM	θ_{edges}	-1.5	-3.67	0.49	-4.32	-3.65	-3.07
		θ_{gwesp}	-0.3	-0.21	0.07	-0.31	-0.21	-0.10
		$\theta_{2-stars}$	-0.4	-0.02	0.27	-0.36	-0.02	0.32
0.8	mERGM	θ_{edges}	-1.5	-1.44	0.03	-1.47	-1.43	-1.39
		θ_{gwesp}	-0.3	-0.33	0.02	-0.36	-0.33	-0.29
		$\theta_{2-stars}$	-0.4	-0.36	0.02	-0.39	-0.36	-0.33
1	ERGM	θ_{edges}	-1.5	-4.56	0.61	-5.36	-4.53	-3.82
		θ_{gwesp}	-0.3	-0.34	0.07	-0.45	-0.35	-0.24
		$\theta_{2-stars}$	-0.4	-0.03	0.74	-0.99	-0.05	0.89
1	mERGM	θ_{edges}	-1.5	-1.47	0.09	-1.58	-1.47	-1.35
		θ_{gwesp}	-0.3	-0.31	0.01	-0.33	-0.31	-0.28
		$\theta_{2-stars}$	-0.4	-0.43	0.05	-0.50	-0.42	-0.34

Resulting means, standard deviations, the medians, 0.1 and 0.9 quantiles of the estimated coefficients of network size 500 nodes and for all six σ_u^2 levels. In this setting, the whole parameter $\boldsymbol{\theta}$ is manipulated compared to the setting in the manuscript.

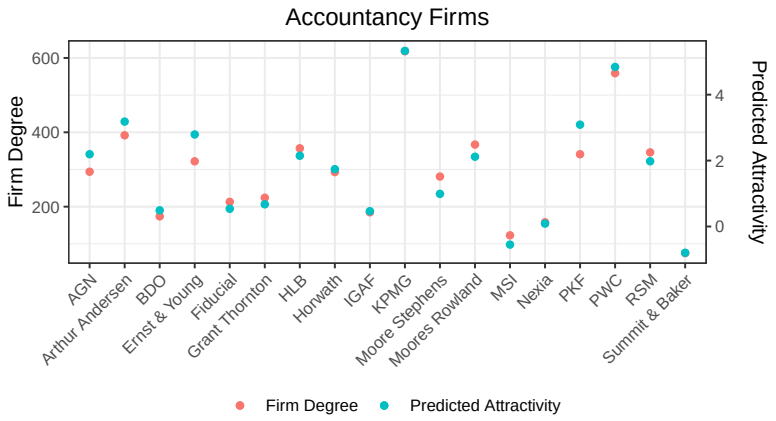
Based on the manuscript's simulation study

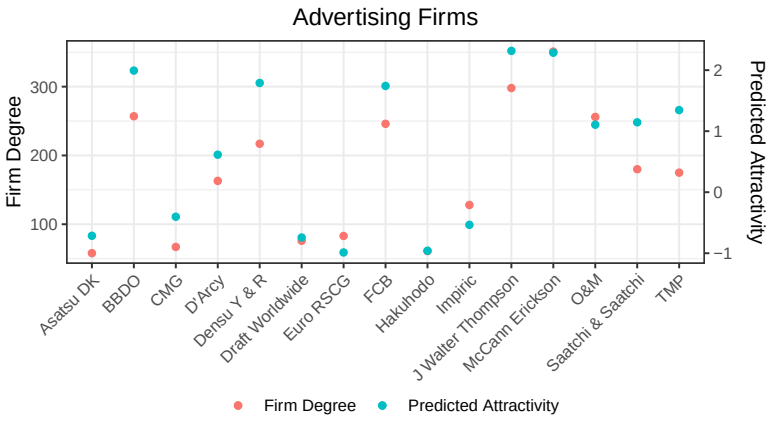


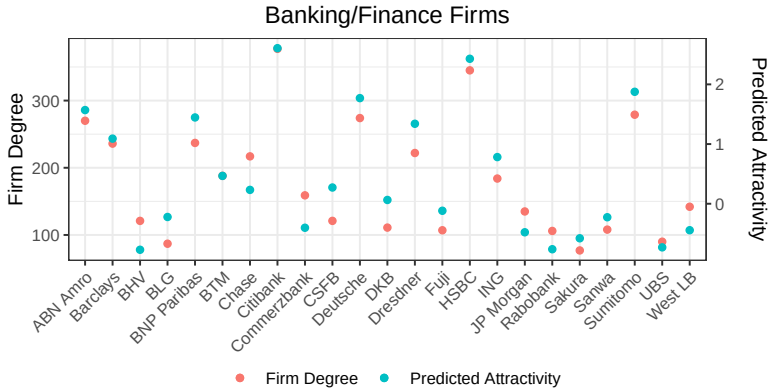
Recovery of σ_u^2 in the simulation study for small (50 nodes) networks.

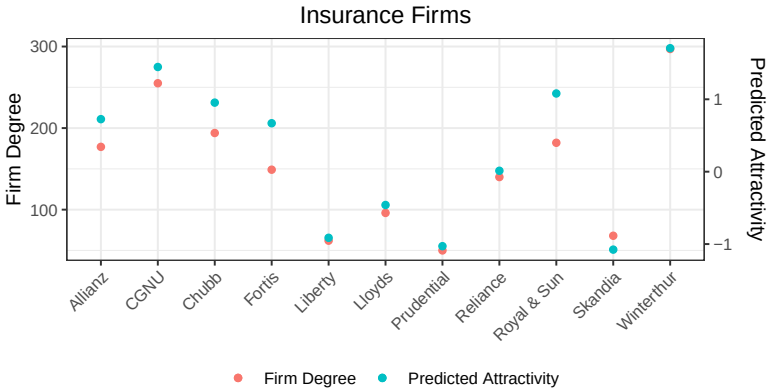


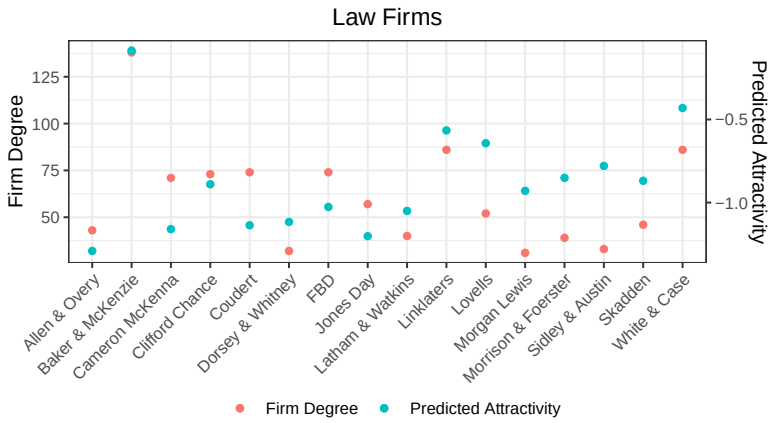
Recovery of σ_u^2 in the simulation study for large (500 nodes) networks.

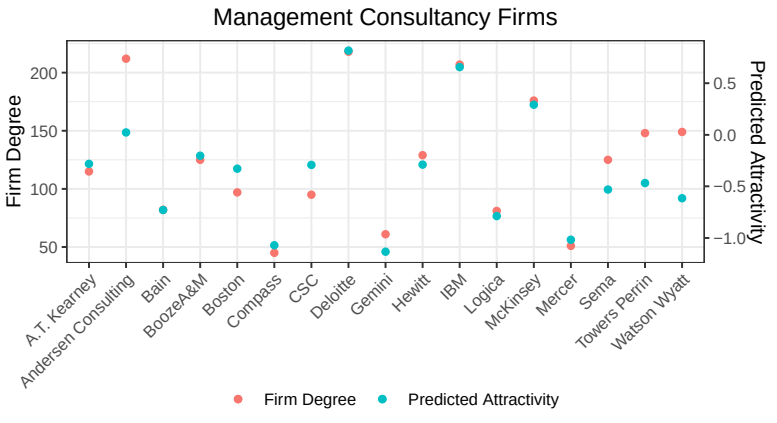

Accountancy Firms
Firm Degree
Predicted Attractivity
600
400
200
4
2
0
AGN
Arthur Andersen
BDO
Ernst & Young
Fiducial
Grant Thornton
HLB
Horwath
IGAF
KPMG
Moore Stephens
Moores Rowland
MSI
Nexia
PKF
PWC
RSM
Summit & Baker
Firm Degree
Predicted Attractivity


Advertising Firms
Firm Degree
Predicted Attractivity
300
200
100
2
1
0
−1
Asatsu DK
BBDO
CMG
D'Arcy
Densu Y & R
Draft Worldwide
Euro RSCG
FCB
Hakuhodo
Impiric
J Walter Thompson
McCann Erickson
O&M
Saatchi & Saatchi
TMP
Firm Degree
Predicted Attractivity


Banking/Finance Firms
Firm Degree
Predicted Attractivity
300
200
100
2
1
0
ABN Amro
Barclays
BHV
BLG
BNP Paribas
BTM
Chase
Citibank
CommerzBank
CSFB
Deutsche
DKB
Dresdner
Fuji
HSBC
ING
JP Morgan
Rabobank
Sakura
Sanwa
Sumitomo
UBS
West LB
Firm Degree
Predicted Attractivity


Insurance Firms
Firm Degree
Predicted Attractivity
300
200
100
1
0
−1
Allianz
CGNU
Chubb
Fortis
Liberty
Lloyds
Prudential
Reliance
Royal & Sun
Skandia
Winterthur
Firm Degree
Predicted Attractivity


Law Firms
Firm Degree
Predicted Attractivity
125
100
75
50
−0.5
−1.0
Allen & Overy
Baker & McKenzie
Cameron McKenna
Clifford Chance
Coudert
Dorsey & Whitney
FBD
Jones Day
Latham & Watkins
Linklaters
Lovells
Morgan Lewis
Morrison & Foerster
Sidley & Austin
Skadden
White & Case
Firm Degree
Predicted Attractivity


Management Consultancy Firms
Firm Degree
Predicted Attractivity
200
150
100
50
0.5
0.0
−0.5
−1.0
A.T. Kearney
Andersen Consulting
Bain
BoozeA&M
Boston
Compass
CSC
Deloitte
Gemini
Hewitt
IBM
Logica
McKinsey
Mercer
Sema
Towers Perrin
Watson Wyatt
Firm Degree
Predicted Attractivity

References

Borgatti, S. P. and Everett, M. G. (1997). Network analysis of 2-mode data. *Social networks*, 19(3):243–269.

Butts, C. T. (2008). network: A package for managing relational data in r. *Journal of Statistical Software*, 24(2):1–36.

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461.

Desmarais, B. A. and Cranmer, S. J. (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PLoS ONE*, 7(1):1–12.

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 Strauss, D. (1986). Markov graphs. *Journal of the american Statistical association*, 81(395):832–842.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3):657–699.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233.

Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, 29(2):231–248.

Handcock, M. S. (2003). Assessing degeneracy in statistical models of social networks. *Technical report, Center for Statistics and Social Sciences, University of Washington*. `http://www.csss.washington.edu/Papers/wp39.pdf`.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1–11.

Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Hummel, R. M., Hunter, D. R., and Handcock, M. S. (2012a). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4):920–939.

Hummel, R. M., Hunter, D. R., and Handcock, M. S. (2012b). Improving simulation-based algorithms for fitting ergms. *Journal of Computational and Graphical Statistics*, 21(4):920–939.

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, 29(2):216–230.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008a). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29.

Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012). Computational statistical methods for social network analysis. *Journal of Computational and Graphical Statistics*, 21(4):856–882.

Kevork, S. and Kauermann, G. (2021). Iterative estimation of mixed exponential random graph models with nodal random effects. *Network Science*, 9(4):478–498.

Kevork, S. and Kauermann, G. (2022). Bipartite exponential random graph models with nodal random effects. *Social Networks*, 70:90–99.

Kolaczyk, E. D. (2009). *Statistical Anaysis of Network Data: Methods and Models*. Springer, New York.

Koskinen, J. and Snijders, T. A. B. (2013). Simulation, estimation, and goodness of fit. In Lusher et al. (2013), pages 141–185.

Krivitsky, P. and Butts, C. T. (2015). Modeling valued networks with statnet. `http://statnet.csde.washington.edu/workshops/SUNBELT/previous/Valued/Valued.pdf`. [Online; accessed: 18 April 2016].

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6:1100–1128.

Krivitsky, P. N. and Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46.

Latapy, M., Magnien, C., and Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48.

Lusher, D., Koskinen, J., and Robins, G. (2013). *Exponential Random Graph Models for Social Networks*. Cambridge University Press, Cambridge.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Moreno, J. L. and Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, pages 342–374.

Morris, M., Handcock, M. S., and Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4):1548.

Pattison, P. and Robins, G. (2002). 9. neighborhood-based models for social networks. *Sociological Methodology*, 32(1):301–337.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407.

Robins, G. L., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29(2):173–191.

Robins, G. L., Snijders, T. A. B., Wang, P., Handcock, M. S., and Pattison, P. (2007b). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks*, 29(2):192–215.

Salter-Townshend, M., White, A., Gollini, I., and Murphy, T. B. (2012). Review of statistical network analysis: models, algorithms, and software. *Statistical Analysis and Data Mining*, 5(4):243–264.

Schweinberger, M. (2011). Instability, Sensitivity, and Degeneracy of Discrete Exponential Families. *Journal of the American Statistical Association*, 106(496):1361–1370.

Shi, C., Li, Y., Zhang, J., Sun, Y., and Yu, P. S. (2017). A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 29(1):17–37.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395.

Snijders, T. A. B. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.

Snijders, T. A. B. and Koskinen, J. (2013). Longitudinal models. In Lusher et al. (2013), pages 130–140.

Snijders, T. A. B., Koskinen, J., and Schweinberger, M. (2010a). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2):567–588.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Snijders, T. A. B., van de Bunt, G. G., and Steglich, C. E. G. (2010b). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212.

Thiemichen, S., Friel, N., Caimo, A., and Kauermann, G. (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46:11–28.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p. *Psychometrika*, 61(3):401–425.

Eidesstattliche Versicherung

(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 03.02.2022 Sevag Kevork