The Consequences of Misspecification in Exponential Random Graph Models

---

Master Thesis in Mathematics (MAT491)

by

Annina Cincera
16-715-070

supervised by

Prof. Dr. Reinhard Furrer
Prof. Dr. Viviana Amati, University of Milano-Bicocca

Zurich, May 2022

The Consequences of Misspecification in Exponential Random Graph Models

Annina Cincera

Version May 19, 2022

Contents

Preface iii

Abstract v

1 Introduction 1
- 1.1 Social Networks . . . 1
- 1.2 Mathematical Representation of a Network . . . 2

2 Exponential Random Graph Models 5
- 2.1 The Model . . . 5
- 2.2 Derivation of the Model Formula . . . 6
- 2.3 Model Specification . . . 13
- 2.4 Simulation . . . 22
- 2.5 Estimation . . . 23
- 2.6 Approximate Wald Test . . . 25
- 2.7 Goodnes of Fit . . . 25
- 2.8 Application: Fitting an ERGM using `statnet` . . . 26

3 Assessing Misspecifications: Motivation and Methods 31
- 3.1 Motivation . . . 31
- 3.2 Setup of the Simulation Study . . . 32
- 3.3 Methods of Evaluation . . . 36
- 3.4 Research Questions . . . 39

4 Results 41
- 4.1 Properties of the Simulated Networks . . . 41
- 4.2 Preliminary Screening . . . 42
- 4.3 The Baseline Models . . . 43
- 4.4 Omitting Gwesp Effects . . . 45

4.5 Omitting Attribute Effects 47
4.6 Specifying the Wrong Functional Form 51
4.7 Results of the MDMC Test 53

5 Discussion **57**

6 Conclusion **61**

A Appendix **63**

Bibliography **71**

Preface

Writing this master thesis has been a very enriching and educational experience. I want to thank Prof. Dr. Reinhard Furrer for being so open to this project, for providing a very helpful and clear structure, and for at the same time enabling me to work very independently. I found the well-structured support and constructive feedback very valuable. I also want to thank Prof. Dr. Viviana Amati for all the work she put into supporting my work on the thesis; for suggesting such interesting topics, for providing me with literature, for answering a whole lot of questions, for the help with structuring the work and finding a systematic approach, for proofreading all first drafts, and for making me feel welcome to ask for help at any time. It was a privilege to have the chance to immerse into the world of social networks and to have such a committed and motivating external supervisor on this journey.
Furthermore, I would like to thank Elia, April, Andreas, and Stefan for proofreading, and our IT support group for repeated rescuing in cases of server or software problems, even on public holidays. I would also like to thank my most amazing flatmates for living through all the ups and downs with me, and my family for the best support one could wish for. Finally, I thank Nathan for supporting me during the whole process of working on this thesis; from writing the first e-mail to placing the last commas.

Annina Cincera
May 2022

Abstract

Social network analysis is a rapidly growing field that is gaining importance in the area of statistics. The most advanced stochastic models for analyzing cross-sectional network data are Exponential Random Graph Models (ERGMs) (Wasserman and Pattison, 1996; Lusher *et al.*, 2013; Robins *et al.*, 2007). ERGMs allow us to understand which network and attribute effects shape the relations in an observed network. The model specification is theory-driven. The choice of the variables included in the model is based on the social processes that are assumed to have generated the observed network. At present, there is a paucity of studies that investigate the consequences of ERGM misspecification. This thesis contributes to investigating the inferential consequences of model misspecification by assessing the consequences of two types of misspecification: omitted variables and wrong functional form. We use a simulation study in which we generate data from a true model with known parameters and fit a postulated model with different model specifications to estimate the parameters based on the generated data. As a baseline situation, we analyze the case where the baseline model is the same as the true model. We then assess the consequences of different misspecifications. We show that there is evidence that omitting attribute effects and specifying a wrong functional form for the homophily effect only have small consequences for the power and robustness of the other network and attribute effects. However, there is also evidence that omitting important network effects from the model may lead to a wrong inference. In particular, the parameter estimates for remaining network effects may be highly biased, leading to misinterpretation of the network processes, especially the processes of network closure.

Chapter 1

Introduction

1.1 Social Networks

Imagine you have a dataset that describes a social network. A network usually consists of nodes representing a group of actors — for example a group of students of a Department of Statistics — some variables with information about the actors — for example the age and the area of specialization of the students — and some information about the relation between the actors — for example which of the students are friends. Imagine, moreover that we are interested in the question of what mechanisms lead to these relations. Why are some of the students friends and others not? Is it the students' area of specialization or some other actor attributes that form a friendship? Or do already existing friendships influence the formation of other friendships, for example, that two students become friends because they already have a common friend? Classical statistical models such as logistic regression do not amount to answer these questions. The main problem is that in general the binary random variables representing the friendship between pairs of students are not independent. Fitting a logistic regression model to a set of observations drawn from dependent binary random variables leads to dependent error terms, which is a violation of a basic assumption of logistic regression. As a consequence, one might over- or underestimate the standard errors and draw wrong inference. We, therefore, need a more suitable model, which allows us to account for and model tie dependencies. This is where network modeling comes into play.

A network consists of a set of nodes and a set of binary relations. The nodes can represent individuals, but also more aggregated entities such as organizations or even countries. The applications of social network modeling are very broad. For example, there is an application where the network represents past exchanges of goods between archaeological sites (Amati *et al.*, 2020) or one where it represents the presence of economic sanctions between countries (Cranmer *et al.*, 2014).

Social network theory is a rapidly growing field and many different methods to analyze networks exist. This thesis concentrates on Exponential Random Graph Models (ERGMs), which are the most advanced stochastic models for the analysis of cross-sectional network data, i.e. one single observation of a network. Even though an extension of ERGMs can be used to model longitudinal models, we restrict ourselves to cross-sectional data, i.e., only consider one network at a time.

1.2 Mathematical Representation of a Network

Mathematically, a social network can be represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, \ldots, n\}$ is the fixed set of *nodes* (also called *vertices*) and $\mathcal{E}$ is the random set of *edges* (or *ties*). For nodes i and j, an edge is denoted (i, j). For directed graphs, the number of possible ties is given by $n \cdot (n-1)$, for undirected graphs, the order of a pair (i, j) is not important and therefore there only exist $0.5 \cdot n \cdot (n-1)$ possible ties. Let X_{ij} be the random variable, called tie-variable, given by

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

X_{ij} is the indicator random variable specifying whether the tie (i, j) is present in the network or not. Note that for an undirected network it holds that $X_{ij} = 1 \Leftrightarrow X_{ji} = 1$. The random variables X_{ij} can be gathered in a stochastic adjacency matrix

$$\boldsymbol{X} = \begin{pmatrix} 0 & X_{12} & \ldots & X_{1n} \\ X_{21} & 0 & \ldots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \ldots & 0 \end{pmatrix}.$$

We do not allow self-ties in the networks and the diagonal elements are therefore set to 0 by default. Consequently, only the off-diagonal elements of the adjacency matrix $\boldsymbol{X}$ are given by random variables. The adjacency matrix of an undirected network is always symmetric. A realization $\boldsymbol{x}$ of $\boldsymbol{X}$ is a particular instance of the stochastic adjacency matrix. We denote by x_{ij} the observed value of X_{ij}. Furthermore, let $\mathcal{X}$ be the space of all networks defined on a fixed set of nodes $\mathcal{N}$. We also define the set of dyads as the set of unordered pairs of nodes.

Besides the adjacency matrix $\boldsymbol{x}$, we might also observe nodal and dyadic attributes. Such covariates give additional information about the nodes and dyads. We denote by $\boldsymbol{v}$ the $n \times p$ matrix, of which each row vector is one of p nodal covariates. We similarly gather the q dyadic covariates in an $n \times n \times q$ array $\boldsymbol{w}$.

Example 1.1. We have a first look at an empirical example of a social network. The network consists of 25 nodes representing pupils of a school class in a Dutch secondary school. There is a tie from pupil i to pupil j if i states that he or she receives and/or gives emotional support to j. We furthermore have two nodal attributes: one indicating the gender and one describing the delinquent behavior of the pupils. The variable delinquency has three levels. The larger the number, the more minor delinquencies (stealing, vandalism, graffiti, or fighting) were committed by the pupil. We, therefore, have $p = 2$ and $\boldsymbol{v}$ is a 25×2 matrix. See Figure 1.1 for a visualization of the network. Furthermore, a dyadic covariate called *primary* is given (here a 25×25 matrix). It is another adjacency matrix $\boldsymbol{w}$ indicating friendship in primary school. From now on, we will refer to this network as the friendship network. The data and the corresponding description are available from www.stats.ox.ac.uk/~snijders/siena/.

The question we want to answer with ERGMs is which mechanisms explain the occurrence of the ties. In statistical terms: we have the goal to model the probability $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x})$, where $\boldsymbol{x}$ is the observed adjacency matrix. The underlying assumption of ERGMs is, that the global structure of the network is determined by local configurations (see Table 1.1). Local configurations, in

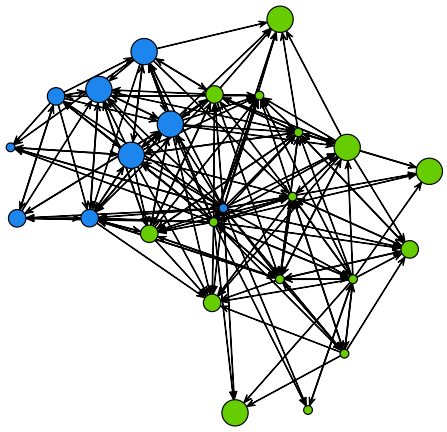


Figure 1.1: Friendship network of a Dutch secondary school class with girls in green, boys in blue and the size of the nodes displaying the three levels of the delinquency variable.

turn, are assumed to be the outcome of social processes. We distinguish between three types of local configurations:

- The first type are so-called *network effects*. They are one of the peculiarities of ERGMs, as they allow to model dependencies between tie-variables X_{ij}. This is motivated by theories about social processes: the presence of certain ties might increase or decrease the probability of some other tie. It for example seems likely that if actor i calls actor j a friend, also j calls i a friend. This effect is called reciprocity and is illustrated in Figure 1.2 (left). Also consider the situation that i calls k a friend and k calls j a friend. We might expect that this increases the probability of i calling j a friend. Such a friendship triangle in a directed network can be seen in Figure 1.2 (right).

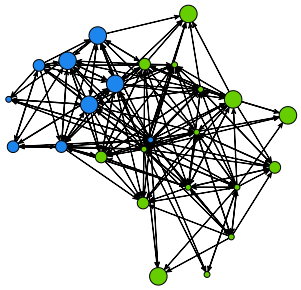
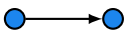
- There are furthermore *nodal attribute effects*, which include the value of a nodal attribute. One important example of such an effect is *homophily*. The underlying social process is the tendency of actors to form relations with actors that have similar attribute values as themselves. We might expect, for example, that friendship is more probable between two pupils of the same gender. This configuration is represented in Figure 1.3, showing a tie between nodes which correspond to two boys of the school class

- The third type of configurations are the *dyadic attribute effects*. An example of such a configuration in the friendship network could be a friendship tie between two nodes which represent pupils that have already been friends in primary school, see Figure 1.3 (right).

Figure 1.2: Illustration of two network effects: a reciprocated tie (left) and a transitive triangle (right).

Figure 1.3: Nodal attribute effect homophily represented by a tie between two nodes standing for two pupils of the same gender (left) and the dyadic attribute effect represented by a tie formed between two pupils that have already been friends in primary school (right).

Table 1.1: Networks are locally emergent. The global structure (left) results from local patterns (right).

Network (global)	**Local configurations**
	network effects nodal attribute effects  dyadic attribute effects

There exist many more local configurations which are the outcome of social processes that might be influencing the global structure of the network (Morris *et al.*, 2008). We will get to know more of them little by little. In Chapter 2, we will see that counts of such local configurations — for example the number of reciprocated ties or the number of transitive triangles — are included in the models as explanatory variables.

The thesis is organized as follows. In Chapter 2 we give an introduction to ERGMs, whereby we pay special attention to the model specification. Model specification is mainly based on theoretical considerations. The researcher includes those effects in the model that correspond to local social processes that are considered to influence the formation of ties. So far, little is known about the consequences of misspecification in ERGMs. In the focus of this thesis lies the question of what the consequences for power and robustness of model misspecifications such as omitted variables and wrong functional forms are. In Chapter 3 we discuss the motivation to assess these model misspecifications and explain the methods used. This is followed by the presentation of the results in Chapter 4, the discussion of the results in Chapter 5 and the conclusion in Chapter 6.

Chapter 2

Exponential Random Graph Models

2.1 The Model

In this section, we state the general form of the statistical network models that we are examining in this thesis. In the subsequent sections, we will then derive the model formula (equation (2.1)) and have a closer look at different model specifications.

Definition 2.1. Exponential random graph models (ERGMs) are a class of probability distributions which take the general form

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{x})\right) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left(\sum_{k=1}^{K} \theta_k z_k(\boldsymbol{x})\right), \tag{2.1}$$

where

- $\boldsymbol{z}(\boldsymbol{x}) = (z_1(\boldsymbol{x}), \ldots, z_K(\boldsymbol{x}))$ is the vector of network statistics. The statistic $z_k(\boldsymbol{x})$ counts the number of times the configuration k is present in the network. In Chapter 1 we have already stated the three types of configurations that might be used as explanatory variables: Network effects, nodal attribute effects, and dyadic attribute effects. For the latter two, the corresponding statistic $\boldsymbol{z}_k(\boldsymbol{x})$ (and thus $\boldsymbol{z}(\boldsymbol{x})$) also depend on the values of $\boldsymbol{v}$ and $\boldsymbol{w}$, but we omit this in the notation for better readability.
- $\boldsymbol{\theta}^\top = (\theta_1, \ldots, \theta_K)$ is the vector of parameters of interest. The parameter θ_k can be seen as a weight that assigns the statistic $z_k(\boldsymbol{x})$ a positive, negative, or no effect. Given some observed network $\boldsymbol{x}$, $\boldsymbol{\theta}$ is unknown and has to be estimated.
- κ is the normalizing constant, making sure that the probabilities over the whole sample space $\mathcal{X}$ sum up to 1. It is given by $\kappa(\boldsymbol{\theta}) = \sum_{x \in \mathcal{X}} \exp\left(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{x})\right)$.

Remark 2.1.

1. We might say that fitting an ERGM consists of three main challenges: the model specification, i.e., choosing the set of statistics, that can explain the global structure of the network; finding a parameter estimate $\widehat{\boldsymbol{\theta}}$ that approximately maximizes the likelihood $L(\boldsymbol{\theta}; \boldsymbol{x}) := \mathrm{P}(\boldsymbol{x}; \boldsymbol{\theta})$; and, finally, interpreting the estimated parameter.
2. For directed networks, there exist $2^{n \cdot (n-1)}$ elements in the sample space $\mathcal{X}$. The normalizing constant $\kappa(\boldsymbol{\theta})$ is therefore not computable except for very small networks. For example, for $n = 10$ the number of possible directed networks is already 2^{90}. We are therefore unable

to compute the exact likelihood, which implies that there is no straightforward way of calculating the maximum likelihood estimator.

3. Note that ERGMs belong to the exponential family of models.

ERGMs were introduced by Wasserman and Pattison (1996), who named them the p^* models and extended the work of Frank and Strauss (1986), who concentrated on Markov models (see Section 2.2.4) and derived equation (2.1) from a dependence assumption using the Hammersley–Clifford theorem as stated by Besag (1974) (compare Section 2.2). A comprehensive description of ERGMs including intuitions and mathematical details can be found in the book by Lusher *et al.* (2013). A concise introduction is provided by Robins *et al.* (2007).

The model is implemented in the R-package `statnet` (Handcock *et al.*, 2018; Hunter *et al.*, 2008b), which is also the one used for all the computations and simulations presented in this thesis. In Section 2.8 we present an example of how an ERGM can be fitted in R using `statnet`. Alternatively, the Java implementation PNet can be used to fit ERGMs (Wang *et al.*, 2009).

2.2 Derivation of the Model Formula

In this section, we explain how equation (2.1) is derived by the Hammersley–Clifford theorem. To apply the Hammersley–Clifford theorem, we need to impose a dependence assumption on the tie-variables. Recall that we assume that the global structure of the network is determined by local configurations which are the outcome of social processes. The social processes (for example reciprocation or the tendency to form triangles) express the dependencies between ties. We, therefore, need to impose a dependence assumption on the tie-variables based on theory about the social processes. Given a specific dependence assumption, we can derive a model formula that has the form of equation (2.1). The different dependence assumptions imply different sets of statistics $\boldsymbol{z}(\boldsymbol{x})$ that may be included in the model. In Section 2.2.1, we state and explain the Bernoulli, Markov, and social circuit dependence assumptions. We then introduce the concept of the dependence graph, state the Hammersley–Clifford theorem and explain how equation (2.1) can be derived in Section 2.2.2. In Sections 2.2.3, 2.2.4, and 2.2.5, we introduce the models implied by the Bernoulli, Markov, and social circuit dependence assumptions respectively.

2.2.1 Dependence Assumptions

As an introductory example, consider the simplest model:

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{\kappa(\theta_1)} \exp\left(\theta_1 \cdot z_1(\boldsymbol{x})\right), \tag{2.2}$$

where $z_1(\boldsymbol{x})$ is the number of ties in $\boldsymbol{x}$. This model assigns the same probability to all realizations $\boldsymbol{x}$ with the same number of ties. The model is known as the homogeneous Bernoulli random graph model (e.g., Frank and Strauss, 1986; Wasserman and Pattison, 1996) and we will look at it in more detail in Section 2.2.3. In the model given by equation (2.2), we have implicitly considered the simplest dependence assumption, the so-called Bernoulli dependence assumption, which assumes that all ties are independent.

Definition 2.2. Under the Bernoulli (in-)dependence assumption, we assume that all tie-variables are independent, i.e., that $X_{ij} \perp X_{kl}$ for all $(i, j) \neq (k, l)$.

For many empirical examples, however, tie-independence is not realistic. In the example of friendship networks, processes such as reciprocation or the formation of friendship triangles are very common. Also for other types of social networks, there exist many social processes that rely on dependencies between the tie-variables. We, therefore, introduce a further dependence assumption — the Markov dependence assumption. As the definition of the Markov dependence assumption includes conditional dependence, we introduce some new notation: for some random network $\boldsymbol{X}$, let $\boldsymbol{X}_{-ij}$ be the set of all tie-variables, but without $\boldsymbol{X}_{ij}$. We are now prepared for the definition of conditional Markov independence.

Definition 2.3. Under the Markov dependence assumption, we assume that
$X_{ij} \perp X_{kl} \mid \boldsymbol{X}_{-\{ij,kl\}} = \boldsymbol{x}_{-\{ij,kl\}}$ for all ties $(i,j),(k,l)$ with $\{i,j\} \cap \{k,l\} = \emptyset$.

Given the rest of the graph, two ties are independent with respect to the Markov dependence, if and only if they do not have a common node.
In Section 2.3.1, we will see that there is evidence that the Markov dependence assumption is still too restrictive and does not adequately represent reality. Pattison and Robins (2002) developed a further dependence assumption that is more realistic. It is called the social circuit dependence assumption and we state it for an undirected network.

Definition 2.4. Under the social circuit dependence assumption, we assume that
$X_{ij} \not\perp X_{kl} \mid \boldsymbol{X}_{-\{ij,kl\}} = \boldsymbol{x}_{-\{ij,kl\}}$ if and only if one of the two following conditions is satisfied:

1. $\{i,j\} \cap \{k,l\} \neq \emptyset$, i.e., they share a node.

2. $x_{jk} = x_{il} = 1$ or $x_{ik} = x_{jl} = 1$, i.e., if the ties $(i,j),(k,l)$ were present in the network, they would create a four-cycle.

Two tie-variables are independent (with respect to the social circuit dependence), given the rest of the network, if and only if they do not share a node (the usual Markov independence) and they do not create a four-cycle when present in the network. The second condition is called a *partial conditional dependence*, as the tie-variables are conditionally dependent given certain observed values of other variables.
It is easy to see that this dependence assumption is more realistic for many empirical settings. Let for example i and k as well as j and l be two pairs of friends. Imagine that i and j meet at a dinner and start to go out with each other. They decide to introduce each other to a good friend and go to a party together. To this party, i brings k and j brings l. At the party k and l get along very well and they also start to go out as well (see Figure 2.1). Note that the relation between k and l in this example was not independent of the relation between i and j. There exist many more examples of social contexts where the social circuit dependence assumption is realistic to be assumed.

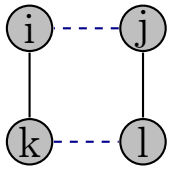


Figure 2.1: Illustration of the second condition of the social circuit dependence. Under the social circuit dependence assumption, X_{ij} and X_{kl} are conditionally dependent.

The social circuit dependence assumption is considered to be the most realistic for many empirical applications and is the one most widely used.
Now that we have introduced the concept of dependence assumptions and have already taken a look at three examples of such assumptions, we are ready to apply the Hammersley–Clifford theorem and derive the model formula. We will derive equation (2.1) from a dependence assumption, where the set of statistics $\boldsymbol{z}(\boldsymbol{x})$ that may be included in the model is subject to the dependence assumption used for the derivation. We will see that given a dependence assumption, this set of statistics that may be included, consists only of ties that are mutually conditionally dependent.

2.2.2 Dependence Graphs and the Hammersley–Clifford Theorem

To be able to apply the Hammersley–Clifford theorem, we have to introduce the notion of the dependence graph.

Definition 2.5. A dependence graph $\mathcal{D}$ for a random graph $\mathcal{G}$ has as nodes $\mathcal{N}_\mathcal{D}$ the tie-variables X_{ij} from $\mathcal{G}$ and the set of edges is given by $\mathcal{E}_\mathcal{D} = \{(X_{ij}, X_{kl}) : X_{ij} \not\perp X_{k,l} \mid \boldsymbol{X}_{-\{ij,kl\}} = \boldsymbol{x}_{-\{ij,kl\}}\}$.

The dependence graph of a directed graph of size n, therefore, has $n \cdot (n-1)$ nodes and, given a certain dependence assumption, its edges indicate, which tie-variables of the original graph $\mathcal{G}$ are conditionally dependent (i.e., dependent conditional on the rest of the network). Note that the dependence graph is a graph for which the nodes are random variables, while for the social networks considered, we take the ties to be the random variables. The random tie-variables of the networks correspond to the random nodes of the dependence graph.
For the Bernoulli dependence assumption, the dependence graph does not contain any edges, as all tie-variables X_{ij} of the network are independent. The Markov dependence assumptions however imply that two ties are conditionally dependent, whenever they have a common vertex. The corresponding dependence graph, therefore, has an edge between node X_{ji} and node X_{kl} as soon as the intersection of $\{i, j\}$ and $\{k, l\}$ is non-empty. See Figure 2.2 for an illustration of the two dependence graphs. The dependence graph corresponding to the social circuit dependence assumption will be discussed in Section 2.2.5.
Note that for any dependence assumption, the dependence graph $\mathcal{D}$ is an undirected graph with random variables as nodes and the property that two nodes are conditionally independent if and only if there is no edge between them, i.e., $X_{ij} \perp X_{kl} \mid \boldsymbol{X}_{-\{ij,kl\}} = \boldsymbol{x}_{-\{ij,kl\}} \Leftrightarrow (X_{ij}, X_{kl}) \notin \mathcal{E}_\mathcal{D}$. This follows directly from Definition 2.5 and it is exactly the definition of a Markov random field.

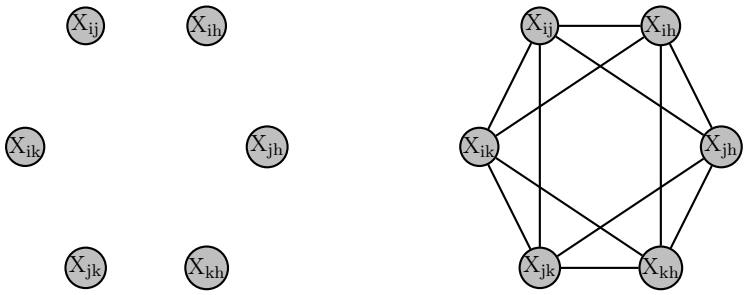


Figure 2.2: The dependence graphs for the Bernoulli dependence assumption (left) and the Markov dependence assumption (right) for an undirected random Graph $\mathcal{G}$ with four nodes.

Definition 2.6. An undirected random graph $\mathcal{D}$ with nodes given by random variables collected in the adjacency matrix $\boldsymbol{X}$ and set of edges $\mathcal{E}$, is a Markov random field if it satisfies the Markov property $X_{ij} \perp X_{kl} \mid \boldsymbol{X}_{-\{ij,kl\}} = \boldsymbol{x}_{-\{ij,kl\}} \Leftrightarrow (X_{ij}, X_{kl}) \notin \mathcal{E}$

The Hammersley–Clifford Theorem 2.1 tells us that under a positivity condition, a Markov random field is equivalent to a Gibbs field.

Definition 2.7. A Gibbs field is a random graph $\mathcal{D}$ with nodes given by the random variables collected in the adjacency matrix $\boldsymbol{X}$ with the property that the probability distribution $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x})$ can be factorized over the cliques $\mathcal{C}_{\mathcal{D}}$ (i.e. the complete subgraphs) of $\mathcal{D}$ as follows:

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{\kappa} \exp\left(\sum_{c \in \mathcal{C}_{\mathcal{D}}} (\theta_c \prod_{\{ij | X_{ij} \in c\}} x_{ij}) \right),$$

where κ is a normalizing constant and $\theta_c \in \mathbb{R}$.

Theorem 2.1. *(Hammersley–Clifford)* A random graph $\mathcal{D}$ with nodes given by the random variables collected in the adjacency matrix $\boldsymbol{X}$ and the property that $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ is a Markov random field if and only if it is a Gibbs field.

Assuming that $\boldsymbol{x}$ represents an observed network, $\prod_{\{ij|X_{ij}\in c\}} x_{ij}$ is equal to 1 if and only if for all tie-variables X_{ij} belonging to the clique c, the value of the observation x_{ij} is equal to 1. Otherwise, $\prod_{\{ij|X_{ij}\in c\}} x_{ij}$ is equal to 0.
Several proofs of Theorem 2.1 exist, see for example Besag (1974) or Grimmett (1973). We continue exploiting the consequences of the Hammersley–Clifford theorem for our probability distribution $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x})$ of our random network $\boldsymbol{X}$.
Note that a clique c of the dependence graph $\mathcal{D}$ corresponds to a set of tie-variables X_{ij} of the original network, represented by the graph $\mathcal{G}$. Assuming that $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) > 0$ for all $\boldsymbol{x}$, we apply Theorem 2.1 and obtain:

$$\begin{aligned} \mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) &= \frac{1}{\kappa} \exp\left(\sum_{c \in \mathcal{C}_{\mathcal{D}}} (\theta_c \prod_{\{ij | X_{ij} \in c\}} x_{ij}) \right) \\ &= \frac{1}{\kappa} \exp\left(\sum_{c \in \mathcal{C}_{\mathcal{D}}} \theta_c z_c(\boldsymbol{x}) \right), \end{aligned} \tag{2.3}$$

where $z_c(\boldsymbol{x}) = \prod_{\{ij|X_{ij}\in c\}} x_{ij}$, i.e., $z_c(\boldsymbol{x})$ is an indicator variable that has value 1 if and only if all tie-variables of clique c take the value 1 in the observed network $\boldsymbol{x}$. Note that equation (2.3) has the form of an ERGM as stated in equation (2.1). Assuming a dependence structure, we derived an ERGM, where the probability distribution of $\boldsymbol{x}$ only depends on the cliques of the dependence graph, which correspond to a set of ties in graph $\boldsymbol{x}$. The dependence assumption thus implies specific network configurations to include in the model. Note, however, that the probability distribution given in equation (2.3) is overparametrized, as every single clique of $\mathcal{D}$ has its own parameter. To reduce the model, we impose so-called *homogeneity assumptions*, suggested by Frank and Strauss (1986). For that, we assume that the parameters θ_c are constant over the cliques c that belong to the same type of configuration.

We now return to the dependence assumptions we have already introduced in Section 2.2.1 and look at the cliques of the corresponding dependence graphs. We explain how to derive the formula in these concrete examples and also familiarize ourselves with how to interpret the parameters.

2.2.3 Bernoulli Random Graph Models

Consider again the dependence graph $\mathcal{D}$ of the Bernoulli (in-)dependence assumption (Figure 2.2). The only cliques in $\mathcal{D}$ are the single nodes. The probability distribution of the Bernoulli random graph model is therefore defined as $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) = (1/\kappa(\boldsymbol{\theta})) \exp(\sum_{\{ij|X_{ij} \in \mathcal{N}_{\mathcal{D}}\}} \theta_{ij} z_{ij})$. Note that every tie $z_{ij} = x_{ij}$ in $\boldsymbol{x}$ has its own parameter θ_{ij}. To obtain the probability distribution (2.2), we need to impose a homogeneity assumption. We hence equate all parameters and set $\theta_{ij} = \theta_e$ for all ties (i, j). It follows that $\sum_{\{ij|X_{ij} \in \mathcal{N}_{\mathcal{D}}\}} \theta_{ij} z_{ij} = \theta_e z_e(\boldsymbol{x})$, where $z_e(\boldsymbol{x})$ is the number of edges in $\boldsymbol{x}$. Now we are left with only one parameter that belongs to the count of edges in $\boldsymbol{x}$ instead of one parameter for every single edge in $\boldsymbol{x}$. Note that the homogeneity assumptions mean that we assume that the network does not depend on its labeling, but only on its structure and that all isomorphic graphs are equally probable.
We therefore end up with $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) = (1/\kappa(\boldsymbol{\theta})) \exp(\theta_e z_e(\boldsymbol{x}))$, the homogeneous Bernoulli random graph model. In this model, all ties are equally likely to be present. Note that it is equivalent to a logistic linear regression model with only the intercept included. To be able to interpret the parameter θ_e, it is useful to have a look at the conditional log-odds of a tie-variable X_{ij}:

$$\begin{aligned}
\log\left(\frac{\mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}{1 - \mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}\right) &= \log\left(\frac{\mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}{\mathrm{P}(X_{ij} = 0 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}\right) \\
&= \log\left(\frac{\frac{1}{\kappa(\boldsymbol{\theta})} \exp(\theta_e z_e(\boldsymbol{x}_{ij}^+))}{\frac{1}{\kappa(\boldsymbol{\theta})} \exp(\theta_e z_e(\boldsymbol{x}_{ij}^-))}\right) \\
&= \theta_e \left(z_e(\boldsymbol{x}_{ij}^+) - z_e(\boldsymbol{x}_{ij}^-)\right) \\
&= \theta_e,
\end{aligned}$$

where $\boldsymbol{x}_{ij}^+$ and $\boldsymbol{x}_{ij}^-$ are the networks where the tie (i, j) is present and absent respectively and equal to $\boldsymbol{x}_{-ij}$ for all other ties. The probability of a tie in the Bernoulli model is therefore given by

$$\mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}) = \mathrm{P}(X_{ij} = 1) = \frac{\exp(\theta_e)}{1 + \exp(\theta_e)}.$$

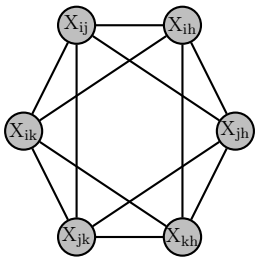
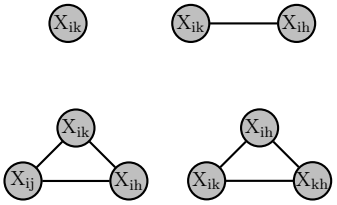
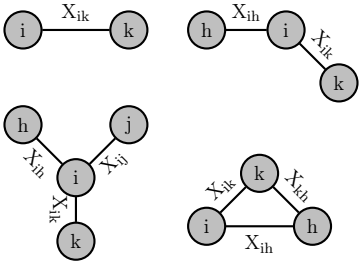
The parameter θ_e is almost always included in the model. It has a similar role as the intercept in regression models. Including the edge count in a model can also be justified using the dependence assumption: the single nodes are cliques in every dependence graph and, after imposing a homogeneity constraint, the number of ties is, therefore, an implied configuration for every dependence assumption. An exception to this rule is for example the case, where we condition on the number of edges, i.e., when we restrict the sample space $\mathcal{X}$ to the set of networks with the same number of edges as the observed graph.

Let us now continue with a first dyadic dependence model: the Markov model.

2.2.4 Markov Random Graph Models

The cliques of the dependence graph representing the Markov dependence assumptions are the single nodes, as well as single edges, triples, or larger cliques composed of ties that have a common vertex (see Table 2.1). In an undirected graph, a single node in the dependence graph represents a tie in the original graph, a single edge represents a so-called 2-star, a triple in the dependence graph is a 3-star or a triangle in the original graph as illustrated in Table 2.1. A clique of

Table 2.1: The dependence graph for the Markov dependence assumption (left), some examples of cliques of the dependence graph (middle), and the corresponding configurations in the original network (right).

Dependence Graph	Cliques	Configurations
		

size k, $k > 3$ represents the k-star configuration. The largest clique that can exist in the Markov dependence graph has size $(n-1)$ because all the nodes of the clique need to have a common vertex. Note that all the configurations of the Markov model must consist of ties that pairwise share a node.

The configurations of the Markov model of an undirected graph, therefore, consist of edges, triangles, and star configurations. After again imposing a homogeneity assumption, i.e., assuming that the labeling of the graph is not important and all edges, triangles, and stars have the same effect, we obtain the model

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp(\theta_e z_e(\boldsymbol{x}) + \theta_t z_t(\boldsymbol{x}) + \sum_{k=2}^{n-1} \theta_{s_k} z_{s_k}(\boldsymbol{x})),$$

where

- $z_e(\boldsymbol{x})$ is again the number of edges,
- $z_t(\boldsymbol{x})$ is the number of triangles,
- and $z_{s_k}(\boldsymbol{x})$ is the number of k-stars for $k \geq 2$.

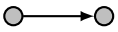
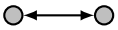
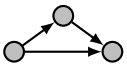
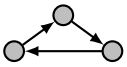
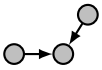
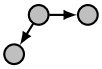
Remark 2.2.

1. Note that the number of 1-stars is the same as the number of edges.
2. The star counts represent the degree distribution of the graph; it can be shown that $z_{s_2}(\boldsymbol{x})$ models the second moment of the degree distribution (the variance) and similarly that $z_{s_k}(\boldsymbol{x})$ models the kth moment (Lusher *et al.*, 2013, Ch. 6). Usually, stars of orders higher than 3 are not included in the model to prevent over-fitting.
3. The number of triangles $z_t(\boldsymbol{x})$ models the clustering of ties in the network, also referred to as *network closure*.

For a directed graph, the set of possible configurations of the Markov model is much larger. Besides the number of ties, it includes reciprocated ties, in-, out-, and mixed stars as well as different types of triangles (for example transitive or cyclic triangles). A researcher can choose, which of the configuration implied by the dependence graph he or she wants to include in the model. This process of model selection is driven by theory about the underlying social processes.

Example 2.1. In Table 2.2, one can find examples of some configurations implied by the Markov dependence assumption for directed networks and their counts in the friendship network.

Table 2.2: Some configurations of the Markov specification and their counts in the friendship network.

Configuration	Count	Illustration
Edges	181	
Reciprocated ties	42	
Transitive triples	724	
Cyclic triplets	118	
In-2-stars	671	
Out-2-stars	1059	

We will see in Section 2.3.1 that for many specifications the Markov models suffer from *degeneracy*. In the next section we have a look at the social circuit models, which are derived from the social circuit dependence assumption. They brought great improvements over the Markov models in terms of degeneracy. Therefore, the social circuit models have overtaken the Markov models and are now the most widely used.

2.2.5 Social Circuit Models

The second condition of the social circuit dependence assumption implies that the tie-variables are conditionally dependent given certain observed values of other variables. Therefore, also the dependence graph depends on the observed values. Assuming we observed a complete graph on four nodes, the corresponding dependence graph is also complete (see Figure 2.3), because any two tie-variables X_{ij} and X_{kl} share a common vertex or belong to a four-cycle.

The configurations implied by the social circuit model are those where all the tie-variables of the configuration pairwise share a node or belong to a four-cycle. Note that all the configurations implied by the Markov model are also implied by the social circuit model. However, the social circuit models allow additional statistics, for example, the *geometrically weighted edgewise shared partner statistics*, which we will get to know in Section 2.3.2.

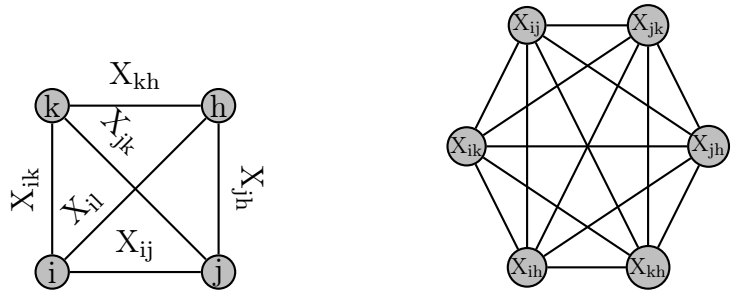


Figure 2.3: A realization of an undirected graph on four nodes (left). The corresponding dependence graph under the social circuit dependence assumption (right).

2.3 Model Specification

The procedure of deriving a model, as introduced in the previous section, consists of the following steps:

1. A dependence assumption is postulated.
2. The corresponding dependence graph and its cliques are determined.
   Applying the Hammersley–Clifford theorem, the factorization of the probability distribution over the cliques of $\mathcal{D}$ is derived.
3. Homogeneity constraints are postulated to reduce the number of parameters.
4. A model is selected based on sociological theory: given the derived set of statistics, the researcher includes only those effects in the model that are considered to influence the social process of tie formation.
   Note that for more complicated dependence assumptions, the set of configurations implied by the Hammersley–Clifford theorem is large and only a subset can be included to prevent overfitting.
5. Additionally to the network effects, nodal and dyadic attribute effects might be included (see Section 2.3.3).

Steps 1 and 4 are the crucial steps in the modeling procedure. In step 1, one has to decide on the dependence assumption for the tie-variable. The social circuit assumption is the most general assumption introduced and also the one considered most realistic for many empirical examples. In step 4, a subset of the implied statistics is chosen and included in the model. This is done based on theories that the researcher has about the social processes that influence the network (compare Lusher *et al.*, 2013, Ch. 2).

There is an important set of statistics, which we have not introduced yet. The set of *alternating* and *geometrically weighted* statistics. They have been proposed as a solution to the problem of degeneracy that was observed for many specifications of the Markov model. In the next section, we will describe the problem of degeneracy and then in Section 2.3.2 introduce the statistics which were proposed as a solution to the problem.

2.3.1 Bimodal Distributions and Degeneracy

We start with an example of a Markov model that illustrates nicely the problem of degeneracy. It is similar to the one given by Lusher *et al.* (2013, Ch. 6), and more details can be found there.

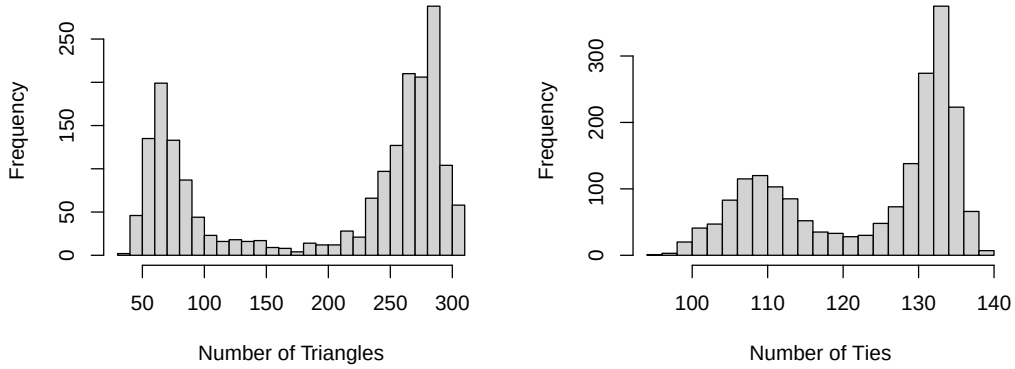


Figure 2.4: Distribution of the triangle statistic (left) and the tie statistic (right). The parameters used for the simulation are $(\theta_e, \theta_{s_2}, \theta_{s_3}, \theta_t) = (-1.8, 0.4, -0.2, 1.2)$.

Example 2.2. Consider an undirected random graph consisting of $n = 35$ nodes and random tie-variables X_{ij} collected in the adjacency matrix $\boldsymbol{X}$. We consider the Markov model, where the variables edge count z_e, number of 2-stars z_{s_2}, the number of 3-stars z_{s_3} and the number of triangles z_t are included. We choose the parameters to be $\boldsymbol{\theta}^\top = (\theta_e, \theta_{s_2}, \theta_{s_3}, \theta_t) = (-1.8, 0.4, -0.2, 1.2)$. The probability distribution of $\boldsymbol{X}$ is therefore given by $\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) = (1/\kappa) \exp\left(-1.8 \cdot z_e + 0.4 \cdot z_{s_2} - 0.2 \cdot z_{s_3} + 1.2 \cdot z_t\right)$. The edge count parameter θ_e models the mean number of ties per node, while the 2-star θ_{s_2} parameter models the variation of the degree distribution. The positive 2-star parameter increases the *centrality* in the network, as a larger variation in the degree distribution increases the existence of *central* nodes, i.e., nodes with more ties than others. The negative 3-star parameter θ_{s_3} prevents the effect of centralization to get out of hand, by penalizing 3-stars and therefore penalizing nodes with high degree (as these induce multiple three stars). The positive triangle parameter θ_t implies *clustering* in the network by rewarding triangles. Given a fixed number of ties, the largest collection of triangles happens in dense regions, where ties belong to several triangles simultaneously. Given the negative edge count parameter, the network will hence reward such dense, clustered regions where many 2-stars are *closed* to triangles. This is why the triangle parameter is said to model network *closure*.
Given the chosen parameter values, we wonder which networks have a high probability. To answer this question, we simulate 2000 networks that are samples of the specified distribution. (How simulation of ERGMs works can be found in Section 2.4.) In Figure 2.4 (left) we take a closer look at the number of triangles of the simulated networks. Note that the distribution has a bimodal shape. There is a collection of networks with about 70 triangles and a collection of networks with about 280 triangles. The distribution of the number of edges statistics looks similar (see Figure 2.4, right).
The probability distribution places a large part of the probability on rather sparse networks with few triangles and another part on denser networks with many triangles. This is not an effect from a probability distribution we wish for. We want the distribution of each statistic to have only one mode. An unimodal distribution is easier to interpret and only then can we find parameters for an observed network such that the observed statistics lie at the center of the probability distribution.

The problem of bimodal distributions for some of the statistics does not only occur with the

parameter and model specification of Example 2.2, but for many other models, especially Markov models (Snijders, 2002).

Another problem that appears often, is that the probability distribution of a model places all its mass on only a very small subset of networks, often on the almost empty network, the almost complete network, or a combination of both (Handcock, 2002). Such a model is called *degenerate* and they are hardly ever useful for modeling, as the realizations of these models do not represent commonly observed networks. Degeneracy can be explained by the *avalanche* effect (Snijders *et al.*, 2006). Consider a model with only an edge count and the number of 2-stars included. Let us have a look at the conditional log-odds of a tie-variable X_{ij} given this model:

$$\begin{aligned}
\log\left(\frac{\mathrm{P}(X_{ij}=1 \mid \boldsymbol{X}_{-ij}=\boldsymbol{x}_{-ij};\boldsymbol{\theta})}{1-\mathrm{P}(X_{ij}=1 \mid \boldsymbol{X}_{-ij}=\boldsymbol{x}_{-ij};\boldsymbol{\theta})}\right) &= \log\left(\frac{\mathrm{P}(X_{ij}=1 \mid \boldsymbol{X}_{-ij}=\boldsymbol{x}_{-ij};\boldsymbol{\theta})}{\mathrm{P}(X_{ij}=0 \mid \boldsymbol{X}_{-ij}=\boldsymbol{x}_{-ij};\boldsymbol{\theta})}\right) \\
&= \log\left(\frac{\frac{1}{\kappa(\boldsymbol{\theta})}\exp(\theta_e z_e(\boldsymbol{x}_{ij}^{+}) + \theta_t z_t(\boldsymbol{x}_{ij}^{+}))}{\frac{1}{\kappa(\boldsymbol{\theta})}\exp(\theta_e z_e(\boldsymbol{x}_{ij}^{-}) + \theta_t z_t(\boldsymbol{x}_{ij}^{-}))}\right) \\
&= \theta_e\left(z_e(\boldsymbol{x}_{ij}^{+}) - z_e(\boldsymbol{x}_{ij}^{-})\right) + \theta_t\left(z_t(\boldsymbol{x}_{ij}^{+}) - z_t(\boldsymbol{x}_{ij}^{-})\right) \\
&= \theta_e\left(\delta_e^{+ij}\right) + \theta_t\left(\delta_t^{+ij}\right),
\end{aligned}$$

where δ_e^{+ij} and δ_t^{+ij} are the so-called *change statistics* given by $z(\boldsymbol{x}_{ij}^{+}) - z_(\boldsymbol{x}_{ij}^{-})$ for the edge and triangle count respectively. The conditional log-odds of the tie-variable X_{ij} is therefore given by $\theta_e \cdot \delta_e^{+ij} + \theta_t \cdot \delta_t^{+ij}$. Assume we have a given network $\boldsymbol{x}$ and we change the value of x_{ij} from 0 to 1, which we assume to close at least one triangle. The added tie does not only increase the number of triangles by δ_t^{+ij}, it might also lead to a large increase in the change statistic δ_t^{+kl} of some other tie-variable X_{kl}. Assumed that θ_t is positive, the conditional log-odds for X_{kl} would therefore also increase. It is therefore likely that the tie (k, l) is present in the network as well. This will further increase the values of other change statistics, which ends up in an avalanche, which leads to a complete graph. Even if the edge parameter θ_e is negative, a too-large positive θ_t will only produce probability distributions that place almost all probability on the complete graph. Similarly, we can end up with only empty networks, when the parameters are negative.

2.3.2 Geometrically Weighted Statistics

A good improvement to the problem of degeneracy brought the introduction of *alternating* star and triangle statistics, developed by Snijders *et al.* (2006). We start with introducing the statistics for undirected graphs and then expand the topic to directed graphs. Consider first the alternating k-star statistic:

$$z_a^s(\boldsymbol{x};\lambda_s) := z_{s_2} - \frac{z_{s_3}(\boldsymbol{x})}{\lambda_s} + \cdots + (-1)^{n-3}\frac{z_{s_{n-1}}(\boldsymbol{x})}{\lambda_s^{n-3}} = \sum_{k=2}^{n-1}(-1)^{k-2}\frac{z_{s_k}(\boldsymbol{x})}{\lambda_s^{k-2}},$$

where $\lambda_s > 0$ is a constant. Instead of adding the k-star statistics individually to the model, we combine them into a single statistic. It was observed in empirical examples that estimates of higher-order star parameters were smaller and had the opposite sign than lower-order star parameters (Robins and Morris, 2007). This observation provided the idea for the new statistic. While adding an edge might cause multiple star statistics to increase, the statistic z_a^s will change only moderately because of the alternating signs, and an *avalanche* effect can be prevented.
To better interpret the new statistics, they were further rewritten and developed by Hunter

(2007). We will now take a closer look at this advancement and new notation. For us it is useful not only because of interpretability but also because the R-package `statnet` uses the definitions and notation of Hunter (2007).

Given an undirected graph $\boldsymbol{x}$, let $z_{d_0}(\boldsymbol{x}), \ldots, z_{d_{n-1}}(\boldsymbol{x})$ be the degree statistics, i.e., $z_{d_k}(\boldsymbol{x})$ denotes the number of nodes in $\boldsymbol{x}$ that have k *neighbors*. Node i has node j as *neighbor* if there exists a tie (i, j). We can now rewrite a k-star statistics z_{s_k} in terms of degree statistics:

$$z_{s_k}(\boldsymbol{x}) = \sum_{i=k}^{n-1} \binom{i}{k} z_{d_i}(\boldsymbol{x}), \qquad \text{for } 2 \leq k \leq n-1.$$

The 1-star count, which is the same as the edge count, is given by

$$z_{s_1}(\boldsymbol{x}) = z_e = \frac{1}{2} \sum_{i=1}^{n-1} i z_{d_i}(\boldsymbol{x}).$$

Using this, we now rewrite the alternating k-star statistics in terms of degree statistics:

$$\begin{aligned}
z_a^s(\boldsymbol{x}; \lambda_s) &= \sum_{k=2}^{n-1} (-1)^{k-2} \frac{z_{s_k}(\boldsymbol{x})}{\lambda_s^{k-2}} \\
&= \sum_{k=2}^{n-1} \sum_{i=k}^{n-1} \left(\frac{-1}{\lambda_s}\right)^{k-2} \binom{i}{k} z_{d_i}(\boldsymbol{x}) \\
&= \lambda_s^2 \sum_{i=2}^{n-1} \left(\sum_{k=2}^{n-1} \left(\frac{-1}{\lambda_s}\right)^{k} \binom{i}{k} \right) z_{d_i}(\boldsymbol{x}) \\
&= \lambda_s^2 \sum_{i=2}^{n-1} \left(\left(1 - \frac{1}{\lambda_s}\right)^i + \frac{i}{\lambda_s} - 1 \right) z_{d_i}(\boldsymbol{x}) \qquad (2.4) \\
&= \lambda_s^2 \sum_{i=1}^{n-1} \left(\left(1 - \frac{1}{\lambda_s}\right)^i - 1 \right) z_{d_i}(\boldsymbol{x}) + 2\lambda_s z_e(\boldsymbol{x}),
\end{aligned}$$

where (2.4) is derived using the binomial theorem $(1 - 1/\lambda_s)^i = \sum_{k=0}^{i} \binom{i}{k} (-1/\lambda_s)^k$.
We now define the *geometrically weighted degree (gwdegree)* statistic:

$$z_{\text{gw}}^{\text{deg}}(\boldsymbol{x}; \beta_s) := \mathrm{e}^{\beta_s} \sum_{k=1}^{n-1} \left(1 - (1 - \mathrm{e}^{-\beta_s})^k\right) z_{d_k}(\boldsymbol{x}).$$

where β_s is a new parameter with $\lambda_s = \exp(\beta_s)$ to ensure that λ_s is positive. The name refers to the geometric series $(1 - \exp(-\beta_s))^i$. The connection between z_a^s and $z_{\text{gw}}^{\text{deg}}$ is given by $z_a^s = \mathrm{e}^{\beta_s} \left(2 z_e(\boldsymbol{x}) - z_{\text{gw}}^{\text{deg}}\right)$ and an ERGM that includes both, z_a^s and z_e, is mathematically equivalent to an ERGM containing $z_{\text{gw}}^{\text{deg}}$ and z_e (Hunter, 2007).

There exists a similar extension of the triangle statistic z_t, which also turned out to be an important improvement of ERGMs. Snijders *et al.* (2006) observed that there were problems with degeneracy especially when the triangles in the observed network were "clumped" and formed dense regions in the network. The triangle count was not capturing transitive closure and such clique-like structures well enough. Again, we will work with the notation introduced by Hunter (2007) and start with introducing the statistics for undirected graphs.

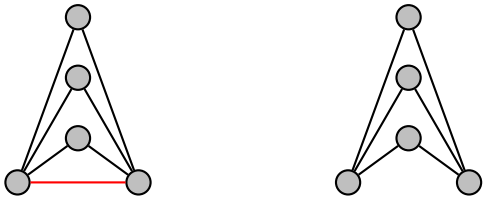


Figure 2.5: A pair of nodes with three edgewise shared partners (left) and three dyadic shared partners (right). In red is the tie that acts as a base for this example.

Let $z_{\text{esp}_0}(\boldsymbol{x}), \ldots, z_{\text{esp}_{n-1}}(\boldsymbol{x})$ be the *edgewise shared partner* statistic, where $z_{\text{esp}_k}(\boldsymbol{x})$ is defined as the number of ties (i, j) of $\boldsymbol{x}$ such that i and j have exactly k common neighbors. We call the tie for whose vertices we count the number of shared partners the *base*.
Similarly, we define the *dyadic shared partner* statistics $z_{\text{dsp}_0}(\boldsymbol{x}), \ldots, z_{\text{dsp}_{n-1}}(\boldsymbol{x})$, where $z_{\text{dsp}_k}(\boldsymbol{x})$ is the number of dyads (i, j) such that i and j have exactly k common neighbors, regardless, of whether the tie (i, j) is present in the network or not, i.e., regardless, of whether $x_{ij} = 0$ or $x_{ij} = 1$. See Figure 2.5 for an illustration of the statistics.
Note that the statistics are closely related to the triangle statistic: it holds that $z_t(\boldsymbol{x}) = 1/3 \cdot \sum_{i=0}^{n-2} i z_{\text{esp}_i}(\boldsymbol{x})$. We need the factor $1/3$ because any of the three edges of a triangle is once considered to be the base edge for which we count the number of edgewise shared partners.
Similarly as for the gwdegree statistics, we now combine the edgewise shared partner statistics to a new geometrically weighted statistic.
The *geometrically weighted edgewise shared partner* statistic $z_{\text{gw}}^{\text{esp}}$, short gwesp, is defined as:

$$z_{\text{gw}}^{\text{esp}}(\boldsymbol{x}; \beta_t) := \mathrm{e}^{\beta_t} \sum_{i=1}^{n-2} \left(1 - (1 - \mathrm{e}^{-\beta_t})^i\right) z_{\text{esp}_i}(\boldsymbol{x}).$$

This new statistic does not model network closure by only counting the number of triangles, but also the clustering of several triangles in "clumps" by adding the different edgewise shared partner statistics, multiplied with a certain weight.
To interpret the parameter value $\theta_{\text{gw}}^{\text{esp}}$ of the gwesp statistic, we consider once again the log-odds of some tie-variable X_{ij}. Note that adding a tie can contribute to the gwesp statistic in two ways: it can add an additional shared partner and it can create a base, i.e. a tie between two nodes that have shared partners. We assume for now that nodes i and j have k shared partners and that adding the tie (i, j) closes those k triangles, i.e., that the tie adds a base. Even though it is unrealistic, we ignore for the moment that adding the tie (i, j) might also create some new shared partner for some other base. Adding the tie (i, j) therefore increases the statistic z_{esp_k} by 1. The log-odds of X_{ij} are given by

$$\log\left(\frac{\mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}{1 - \mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij}; \boldsymbol{\theta})}\right) = \theta_{\text{gw}}^{\text{esp}} \mathrm{e}^{\beta_t} \left(1 - (1 - \mathrm{e}^{-\beta_t})^k\right).$$

For a positive $\theta_{\text{gw}}^{\text{esp}}$ and a positive β_t, the log-odds of the tie (i, j) are positive and increase the more shared partners k the nodes i and j have. This is in line with theory for example about the formation of friendship: the more shared friends the nodes i and j have, the more likely it is that also i and j become friends. However the increase of the log-odds for positive $\theta_{\text{gw}}^{\text{esp}}$ and a positive β_t is not linear but decreases with k. This again makes sense when applied to a friendship network: it seems to be realistic that for example having two instead of one common friend increases the probability for friendship more than having for example six instead of five

friends. Similarly, it can be shown that in a model with positive $\theta_{\text{gw}}^{\text{esp}}$ and β_t, adding a shared partner for some base leads to an increase in probability (the edge has positive log-odds), but that the log-odds are the largest if there is not yet any other shared partner. The more shared partners two nodes have, the smaller the probability of an additional shared partner. In practical examples, we of course also have to take into account the effect on the log-odds of other statistics included in the model.

To be able to distinguish between the transitive closure due to adding a base and transitive closure due to adding additional shared partners, also the statistic *geometrically weighted dyadic shared partners* was introduced. It is a combination of the dyadic shared partners statistics, which count the number of shared partners no matter whether the base is present or not. The *geometrically weighted dyadic shared partner* statistic $z_{\text{gw}}^{\text{dsp}}$, short gwdsp, is defined as:

$$z_{\text{gw}}^{\text{dsp}}(\boldsymbol{x};\beta_d) := \mathrm{e}^{\beta_d} \sum_{i=1}^{n-2} \left(1-(1-\mathrm{e}^{-\beta_d})^i\right) z_{\text{dsp}_i}(\boldsymbol{x}).$$

The decay parameters β_s, β_t and β_d are usually chosen to be positive. They can be chosen manually (Robins and Morris, 2007 name $\log(2)$ to be a reasonable value) or they can be estimated using the methods of Hunter and Handcock (2006).

The new statistics can be easily extended to directed networks. In Table 2.3, one can find the most important ones.
Note that the edgewise and dyadic shared partner statistics $z_{\text{esp}_k}(\boldsymbol{x})$ and $z_{\text{dsp}_k}(\boldsymbol{x})$ do not belong to the framework of the Markov models for $k \geq 2$, as they do not correspond to cliques in the dependence graph of the Markov dependence assumption. However, they belong to the framework of the social circuit models, as any two edges either share an edge or belong to a four-cycle. The gwesp and gwdsp statistics are therefore implied by the social circuit dependence assumption. Even though degeneracy might still occur, these additional statistics have brought a big improvement (Hunter *et al.*, 2008a; Lusher *et al.*, 2013).

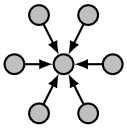
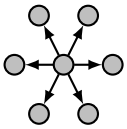
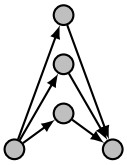
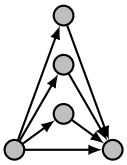
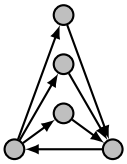
Summing up, the statistics that are most often included in ERGMs to model directed networks are the edge count, the count of reciprocated ties, the geometrically weighted in- and out-degree statistics, the geometrically weighted edgewise shared partner statistic (transitive and cyclic), and the geometrically weighted dyadic shared partner statistic. Of course, there exist many more statistics that could be included (Morris *et al.*, 2008). The researcher has to choose the effects considered to be influential for the particular observed network.

2.3.3 Nodal and Dyadic Covariates

So far, we have focused only on the network effects. However, as listed in Section 2.1, there are also attribute effects, which are related to the tie formation process and which we treat as exogenous explanatory variables. On the one hand, there are the actor attributes effects, that is, counts of configurations that depend on an actor attributes; and on the other hand, there are the dyadic attribute effects, that is, counts of configurations that depend on attributes of pairs of actors. Both, actor and dyadic attributes, can influence the formation of an edge.

Example 2.3. Recall that for the friendship network the two nodal attributes *gender* and *delinquency* are given. In addition, we have the dyadic attribute *primary*, which indicates which

Table 2.3: Geometrically weighted statistics for directed networks.

Configuration	Illustration	Statistic
Geometrically weighted in-degree		$\mathrm{e}^{\beta_s^{\text{in}}} \sum_{k=1}^{n-1} \left(1 - (1 - \mathrm{e}^{-\beta_s^{\text{in}}})^k\right) z_{d_k}^{\text{in}}(\boldsymbol{x})$
Geometrically weighted out-degree		$\mathrm{e}^{\beta_s^{\text{out}}} \sum_{k=1}^{n-1} \left(1 - (1 - \mathrm{e}^{-\beta_s^{\text{out}}})^k\right) z_{d_k}^{\text{out}}(\boldsymbol{x})$
Geometrically weighted dyadic shared partners		$\mathrm{e}^{\beta_d} \sum_{i=1}^{n-2} \left(1 - (1 - \mathrm{e}^{-\beta_d})^i\right) z_{\text{dsp}_i}(\boldsymbol{x})$
Geometrically weighted edgewise shared partners (transitive)		$\mathrm{e}^{\beta_t^{\text{trans}}} \sum_{i=1}^{n-2} \left(1 - (1 - \mathrm{e}^{-\beta_t^{\text{trans}}})^i\right) z_{\text{esp}_i}^{\text{trans}}(\boldsymbol{x})$
Geometrically weighted edgeweise shared partners (cyclic)		$\mathrm{e}^{\beta_t^{\text{cyc}}} \sum_{i=1}^{n-2} \left(1 - (1 - \mathrm{e}^{-\beta_t^{\text{cyc}}})^i\right) z_{\text{esp}_i}^{\text{cyc}}(\boldsymbol{x})$

of the pairs of secondary school pupils were already friends in primary school. Here are some examples of attribute effects we could look at:

- We might expect that friendship is more likely between two students of the same gender. We can therefore include a homophily effect of gender in the model, for example by counting the number of ties between two pupils of the same gender.

- Perhaps we suspect that students of one gender have more friendships than of the other. Since we are dealing with a directed network, we distinguish between two cases: one gender might "send" more friendships (that is, there are more outgoing ties for the gender, compared to the other gender), or it might "receive" more friendships (that is, there are more incoming ties). So we could add two statistics to the model: one counts the outgoing ties and one counts the incoming ties for one gender. Since the edge count statistic is always added to the model, it is sufficient to count the outgoing and incoming ties for one of the genders. Otherwise, the model is overparameterized. The two effects are called sender and receiver effects respectively, or they might be referred to as activity and popularity of *gender*.

- Also for the variable *delinquency*, we can add statistics for homophily, activity, and popularity.

- Furthermore, we could add a statistic to the model that includes the dyadic attribute. We might wonder if there is a connection between friendship in primary school and friendship in secondary school. Let us suppose that the variable *primary* has the value 1 for a dyad (i, j), if i and j have already been friends in primary school, and 0 otherwise. Then we can add an effect to the model that counts the number of ties in the current friendship network that have the value 1 for the dyadic attribute variable.

We now make a small overview of the most important attribute effects and their mathematical formulation. Let $\boldsymbol{x}$ be the realization of a directed random network with n nodes. Let $\boldsymbol{v}^{(b)} \in \mathbb{R}^n$ be a binary nodal attribute variable, $\boldsymbol{v}^{(c)} \in \mathbb{R}^n$ a continuous nodal attribute variable and $\boldsymbol{w} \in \mathbb{R}^{n\cdot(n-1)}$ a dyadic attribute variable (binary or continuous). For the binary nodal attribute variable $\boldsymbol{v}^{(b)}$, we assume that it takes the values 0 and 1 and we say that a node i has the attribute if $v_i^{(b)} = 1$ and does not have it if $v_i^{(b)} = 0$. In Table 2.4 some of the most important statistics including attributes are listed and illustrated. For the binary attribute variable, a node is filled in black if it has the attribute, and not filled if both are possible: that it has the attribute or not. For the continuous attribute variable, a larger node indicates a larger attribute value. A node that is not filled represents a node for which the attribute value is not relevant. The statistics for undirected networks are very similar and further details can be found in Lusher *et al.* (2013) in Chapter 8.

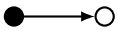
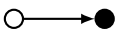
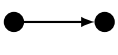
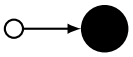
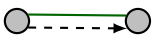
Remark 2.3.

1. As we only have cross-sectional data, we cannot know if the attributes lead to a tie, or if the ties affect the attribute values. Of course, a variable like *gender* can be assumed to be fixed. However, for a variable like *delinquency*, this must not be the case and there are two possibilities for the association: that friendship ties are formed due to the variable, or that the variable is influenced by friendship ties. As in a regression model, a significant effect only means that there is a correlation, not a causal relationship.

2. Note that there are different statistics for the homophily effect. For a binary variable, you can either simply count the ties that are between two nodes with the same attribute (uniform homophily) or you can add two statistics to the model, one that counts the ties between nodes with the attribute and one that counts the ties between nodes without the attribute (differential homophily). In the former, one assumes that the homophily effect is the same for nodes with or without attribute; in the latter, nodes with the attribute can have a different homophily effect than those without. A positive parameter indicates a tendency for homophily in the network and a negative parameter indicates heterophily.

3. For continuous variables, there are also different ways to capture homophily. We can sum up the differences (difference homophily statistic), absolute differences (absolute homophily statistic), or squared differences (squared homophily statistic) of the attribute values of i and j over all ties (i, j). Note that the parameters have a slightly different meaning depending on the statistic. For absolute homophily, a negative parameter (i.e., small absolute differences between the attribute values of the nodes of a tie) indicates homophily and a positive parameter (i.e., larger absolute differences observed) indicates heterophily (because the latter means that the absolute differences between the nodes of a tie tend to

Table 2.4: Some important attribute effects for a directed network.

Configuration	Statistic	Illustration
Binary nodal attribute:		
Sender effect	$\sum_{i,j} v_i^{(b)} x_{ij}$	
Receiver effect	$\sum_{i,j} v_j^{(b)} x_{ij}$	
Homophily effect	$\sum_{i,j} \mathbb{1}_{\{v_i^{(b)}=v_j^{(b)}\}} x_{ij}$ (uniform) or $\sum_{i,j} \mathbb{1}_{\{v_i^{(b)}=v_j^{(b)}=1\}} x_{ij}$ and $\sum_{i,j} \mathbb{1}_{\{v_i^{(b)}=v_j^{(b)}=0\}} x_{ij}$ (differential)	
Continuous nodal attribute:		
Sender effect	$\sum_{i,j} v_i^{(c)} x_{ij}$	
Receiver effect	$\sum_{i,j} v_j^{(c)} x_{ij}$	
Homophily effect	$\sum_{i,j} \mid v_i^{(c)} - v_j^{(c)} \mid x_{ij}$ (absolute) or $\sum_{i,j} (v_i^{(c)} - v_j^{(c)})^2 x_{ij}$ (squared) or $\sum_{i,j} (v_i^{(c)} - v_j^{(c)}) x_{ij}$ (difference)	
Dyadic attribute:		
Dyadic attribute effect	$\sum_{i,j} w_{ij} x_{ij}$	

be large). The parameter of the squared homophily statistic can be interpreted similarly, though we need to consider that the squared homophily statistic amplifies the difference for large values (larger than one) and reduces it for small differences (smaller than one). A positive difference homophily parameter indicates that the tails of ties tend to have larger attribute values than the heads. A negative parameter thus means that the heads tend to have larger attribute values than the tails. Note that only for the difference homophily statistic, the direction of the ties matters. The researcher chooses the mathematical formula of the homophily effect based on theoretical considerations about the underlying social processes.

4. Normally, the homophily effect is added to the model together with the sender and receiver effects. Because, if a sender and receiver effect is significant for an attribute, it is likely that as a consequence there are also many ties between nodes with this attribute. However, this does not necessarily mean that there is a significant homophily effect in the network. If we are interested in homophily, we should include sender and receiver effects as control variables in the model.

5. Note that all attribute effects we consider only affect single dyads at a time and it therefore holds that $\mathrm{P}(X_{ij} = 1 \mid \boldsymbol{X}_{-ij}, \boldsymbol{v}, \boldsymbol{w}) = \mathrm{P}(X_{ij} = 1 \mid \boldsymbol{v}, \boldsymbol{w})$ for every dyad (i, j). The parameter of a model including only the edge count and dyadic independent attribute effects could therefore be estimated using traditional generalized regression methods.

6. We consider only binary and continuous attribute variables in this section, but the statistics can be easily extended, e.g., for categorical variables with more than two levels.

2.4 Simulation

Imagine we have given a probability distribution (our target distribution)

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{x})\right)$$

for a random network $\boldsymbol{X}$ of size n and a fixed parameter $\boldsymbol{\theta}$. Now we would like to draw samples from this distribution and we might also be interested in the properties of the networks coming from this distribution. This will be especially important for parameter estimation. To obtain a sample $\boldsymbol{x}$ from the distribution $\mathrm{P}(\boldsymbol{X}; \boldsymbol{\theta})$, we use Markov chain Monte Carlo (MCMC) methods. For this purpose, a set of networks $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ is generated which is assumed to converge to a sample from our target distribution. The most commonly used algorithm for simulation is the Metropolis algorithm, which consists of the following steps to generate a sample of size m:

[1] Let $\boldsymbol{x}^{(1)}$ be the starting network. It can be the empty, complete, or any arbitrary graph of size n. Set $k = 2$.

[2] Choose at random a pair of nodes i and j. Let the proposal graph $\boldsymbol{x}^*$ be equal to $\boldsymbol{x}^{(k-1)}$, except that $x^*_{ij} = 1 - x^{(k)}_{ij}$. We accept the proposal graph x^* with probability

$$\min\left\{1, \frac{\mathrm{P}(\boldsymbol{x}^*; \boldsymbol{\theta})}{\mathrm{P}(\boldsymbol{x}^{(k-1)}; \boldsymbol{\theta})}\right\} = \min\left\{1, \exp(\boldsymbol{\theta}^\top \cdot \boldsymbol{\delta}^{1-\boldsymbol{x}^{(k-1)}})\right\},$$

where $\boldsymbol{\delta}^{1-\boldsymbol{x}^{(k-1)}}$ is the vector of *change statistics* given by $\boldsymbol{z}\big(\boldsymbol{x}^*_{ij}\big) - \boldsymbol{z}\big(\boldsymbol{x}^{(k-1)}_{ij}\big)$. If the proposal graph is accepted, we set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^*$, and $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ otherwise.

[3] If $k < m$, set $k = k + 1$ and go to step [2], otherwise stop.

Note that a proposal graph is always accepted when $\mathrm{P}(\boldsymbol{x}^*; \boldsymbol{\theta})/\mathrm{P}(\boldsymbol{x}^{(k-1)}; \boldsymbol{\theta}) \geq 1$, i.e., when the conditional log-odds of the change

$$\log\left(\frac{\mathrm{P}\left(X_{ij} = 1 - x^{(k-1)}_{ij} \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}^{(k-1)}_{-ij}\right)}{\mathrm{P}\left(X_{ij} = x^{(k-1)}_{ij} \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}^{(k-1)}_{-ij}\right)}\right) = \boldsymbol{\theta}^\top \cdot \boldsymbol{\delta}^{1-\boldsymbol{x}^{(k-1)}}$$

are larger or equal to 0. We therefore always accept the proposal graph when it is more likely than the current graph and we might still accept it (with probability $\mathrm{P}(\boldsymbol{x}^*;\boldsymbol{\theta})/\mathrm{P}(\boldsymbol{x}^{(k-1)};\boldsymbol{\theta})$), even if the proposal graph is less likely than the current graph.
Furthermore, note that we do not need the normalizing constant $\kappa(\boldsymbol{\theta})$ to calculate the ratio $\mathrm{P}(\boldsymbol{x}^*;\boldsymbol{\theta})/\mathrm{P}(\boldsymbol{x}^{(k-1)};\boldsymbol{\theta})$, which is very important for the practical implementation of the algorithm. The default for m is usually around 10'000 or larger. It is the number of iterations we expect to be needed to ensure that the chain converged and we sample from the target distribution. It is usually called *burn-in*. If we need several samples from the target distribution, we do not need a separate MCMC chain for each sample, but we can use one long MCMC chain. From the long chain, we remove the burn-in, and to ensure that the samples do not depend on each other too much, we use a high thinning, i.e. we discard a high number of networks between the samples we keep. Thinning is usually chosen to be around 1000 or larger. Even with choosing a high thinning, we expect some small autocorrelation in the sample.

Alternative MCMC sampling methods and further explanations about sampling can be found in Lusher *et al.* (2013), Chapter 12.

2.5 Estimation

The goal of estimation is to find reasonable values for the parameter $\boldsymbol{\theta}$ based on data. Recall that the model belongs to the exponential family for which some nice properties about the likelihood are known, some of which we will consider in this section. We use the maximum likelihood method to estimate the parameter. The likelihood is given by

$$L(\boldsymbol{\theta};\boldsymbol{x}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{x})\right).$$

Maximizing the likelihood is equivalent to maximizing the log-likelihood given by

$$\ell(\boldsymbol{\theta};\boldsymbol{x}) = \left(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{x})\right) - \log(\kappa(\boldsymbol{\theta})).$$

As we cannot compute the normalizing constant $\kappa(\boldsymbol{\theta})$, we are not able to compute the likelihood or log-likelihood. We therefore have to approximate the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$. Different methods for the approximation exist. We will have a look at the Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) method by Geyer and Thompson (1992), which is also the default in the R-package `statnet`. Further details about the algorithm and its application to ERGMs can be found in Handcock *et al.* (2003) and Hunter and Handcock (2006). We start by stating some properties of the given likelihood. By conducting some simple computations, one can show that

$$\nabla \ell(\boldsymbol{\theta}) = \boldsymbol{z}(\boldsymbol{x}) - \mathrm{E}_{\boldsymbol{\theta}}\left(\boldsymbol{z}(\boldsymbol{X})\right) \tag{2.5}$$

and

$$\boldsymbol{H}\left(\ell(\boldsymbol{\theta})\right) = -\operatorname{Var}_{\boldsymbol{\theta}}\left(\boldsymbol{z}(\boldsymbol{X})\right), \tag{2.6}$$

where $\boldsymbol{H}\left(\ell(\boldsymbol{\theta})\right)$ is the Hessian matrix of $\ell(\boldsymbol{\theta})$. The right-hand side of (2.5) and (2.6) can be approximated by simulating networks $\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(m)}$ (samples from the distribution $\mathrm{P}(\boldsymbol{X};\boldsymbol{\theta})$) and computing the sample mean and variance. These approximations will prove to be useful in

the estimation process. Note that equation (2.5) proves that maximizing the likelihood equation is equivalent to finding a solution to the first moment equation $\boldsymbol{z}(\boldsymbol{x}) - \mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{z}(\boldsymbol{x})\big) = 0$ over $\boldsymbol{\theta}$. Furthermore, it holds that if a statistic $z_k(\boldsymbol{x})$ is equal to zero, the likelihood will stay constant for any choice of parameter θ_k. In that case, no unique MLE exists and the MCMCMLE algorithm does not work. It is therefore advised to check that none of the statistics included in the model is equal to zero before applying the algorithm.

At the core of the MCMCMLE algorithm lies a Newton–Raphson iteration step: a current parameter $\boldsymbol{\theta}^k$ is updated by

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \big(\boldsymbol{H}(\ell(\boldsymbol{\theta}))\big)^{-1} \cdot \nabla \ell(\boldsymbol{\theta}). \tag{2.7}$$

Applying equations (2.5) and (2.6), we can replace the Hessian and the gradient of $\ell(\boldsymbol{\theta})$ and obtain:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \big(\mathrm{Var}_{\boldsymbol{\theta}^k}(\boldsymbol{z}(\boldsymbol{X}))\big)^{-1} \cdot \big(\mathrm{E}_{\boldsymbol{\theta}^k}(\boldsymbol{z}(\boldsymbol{X})) - \boldsymbol{z}(\boldsymbol{x})\big). \tag{2.8}$$

It is however computationally expensive to draw samples to approximate $\mathrm{E}_{\boldsymbol{\theta}^k}\big(\boldsymbol{z}(\boldsymbol{X})\big)$ and $\mathrm{Var}_{\boldsymbol{\theta}^k}(\boldsymbol{z}(\boldsymbol{X}))$ in every updating step. Therefore, only samples for an initial parameter $\boldsymbol{\theta}^0$ are generated and $\mathrm{E}_{\boldsymbol{\theta}^k}\big(\boldsymbol{z}(\boldsymbol{X})\big)$ and $\mathrm{Var}_{\boldsymbol{\theta}^k}(\boldsymbol{z}(\boldsymbol{X}))$ are approximated using a weighted average and weighted covariance of the sampled statistics $z(\boldsymbol{x}^{(h)})$ respectively. This results in the updating step:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \big(D_h\big)^{-1} \cdot \left(\sum_{h=1}^{H} w_h \boldsymbol{z}(\boldsymbol{x}^{(h)}) - \boldsymbol{z}(\boldsymbol{x})\right), \tag{2.9}$$

where $\boldsymbol{x}^{(h)}$ are samples of $\mathrm{P}(\boldsymbol{X}; \boldsymbol{\theta}^0)$,

$$w_h = \frac{\exp((\boldsymbol{\theta}^k - \boldsymbol{\theta}^0)^\top \boldsymbol{z}(\boldsymbol{x}^{(h)}))}{\sum_{h=1}^{H} \exp((\boldsymbol{\theta}^k - \boldsymbol{\theta}^0)^\top \boldsymbol{z}(\boldsymbol{x}^{(h)}))},$$

and D_h is the weighted covariance matrix is given by

$$\sum_{h=1}^{H} w_h \boldsymbol{z}(\boldsymbol{x}^{(h)})^\top \boldsymbol{z}(\boldsymbol{x}^{(h)}) - \left(\sum_{h=1}^{H} w_h \boldsymbol{z}(\boldsymbol{x}^{(h)})\right)^\top \sum_{h=1}^{H} w_h \boldsymbol{z}(\boldsymbol{x}^{(h)}).$$

This procedure is called *importance sampling*. Step (2.9) is repeated several times and a chain of parameters $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^G$ is created. If $\boldsymbol{\theta}^k$ is is reasonably close to $\boldsymbol{\theta}^0$, the weighted average and weighted covariance are good approximations of $\mathrm{E}_{\boldsymbol{\theta}^k}\big(\boldsymbol{z}(\boldsymbol{X})\big)$ and $\mathrm{Var}_{\boldsymbol{\theta}^k}\big(\boldsymbol{z}(\boldsymbol{X})\big)$ respectively. The approximation becomes better, the larger one chooses H.
Mostly, several iterations of the algorithm are conducted. This is done by setting $\boldsymbol{\theta}^0_{new} = \boldsymbol{\theta}^G$ at the beginning of a new iteration, where $\boldsymbol{\theta}^G$ is the last parameter of a chain of parameters $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^G$ obtained in the previous iteration of the algorithm. In the next iteration, samples of the distribution $\mathrm{P}(\boldsymbol{X}; \boldsymbol{\theta}^0_{new}) = \mathrm{P}(\boldsymbol{X}; \boldsymbol{\theta}^G)$ are used for $x^{(h)}$ in (2.9). In the R-package `statnet`, the maximum number of iterations can be chosen manually and we use as the default a maximum of 40 iterations. For $\boldsymbol{\theta}^0$, the maximum pseudo-likelihood estimator (MPLE) is often used. It is

obtained by setting

$$\mathrm{P}(\boldsymbol{x};\boldsymbol{\theta}) \approx \prod_{(i,j)} \mathrm{P}(X_{ij} = x_{ij} \mid \boldsymbol{X}_{-ij} = \boldsymbol{x}_{-ij};\boldsymbol{\theta}), \tag{2.10}$$

and reading the right-hand side of (2.10) as a likelihood of logistic regression. The explanatory variables are given by the change statistics and the response variables are the off-diagonal elements of the adjacency matrix $\boldsymbol{X}$. The MPLE is then estimated using the same algorithms with which a logistic regression model is estimated. While the MPLE is computationally cheap to estimate, it can be highly biased, especially if there exist complicated dependence structures in the network (van Duijn *et al.*, 2009). While it historically has been used to estimate the parameter $\boldsymbol{\theta}$, it is today only used to obtain the initial value $\boldsymbol{\theta}^0$.

Standard errors of the obtained estimate $\widehat{\boldsymbol{\theta}}$ can be derived from the Fisher information matrix $\boldsymbol{I}(\widehat{\boldsymbol{\theta}})$. It holds that $\boldsymbol{I}(\widehat{\boldsymbol{\theta}}) = \mathrm{Var}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{z}(\boldsymbol{X}))$ and we approximate the asymptotic covariance matrix of the estimator with a sample equivalent of $\left(\mathrm{Var}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{z}(\boldsymbol{X}))\right)^{-1}$.

Furthermore, it is wise to check the convergence at the end of the estimation process. In the R-package `statnet`, there exist a built-in function to do so. With this function, one can check whether the average of statistics of the networks simulated with the estimated parameter fluctuates nicely around the observed statistics.

As protection against degeneracy, it was suggested by Handcock (2002) and Snijders (2002) to condition on the number of edges in the estimation process, i.e., to restrict the distribution to networks that have the same number of edges as the observed network. This can prevent the *avalanche effect*. However, as Handcock (2002) points out, this is an unrealistic assumption, as the number of edges in a graph is the outcome of social processes and cannot be assumed to be fixed. Snijders (2002) furthermore observed, that there remain problems with bimodality. Finding a good model specification, maybe including some geometrically weighted statistics, is a better protection against degeneracy than conditioning on the number of edges.

2.6 Approximate Wald Test

To test the significance of the effects, we use that the MLE asymptotically follows a normal distribution. Hence, approximating $\widehat{\theta}_k/\widehat{\mathrm{SE}}(\widehat{\theta}_k)$ by $\mathcal{N}(0,1)$ is reasonable, where we interpret $\widehat{\theta}_k$ as an estimator. To test $H_0 : \theta_k = 0$ versus $H_1 : \theta_k \neq 0$, we compute the test statistic $z = |\widehat{\theta}_k/\widehat{\mathrm{SE}}(\widehat{\theta}_k)|$ and the p-value given by $2(1 - \Phi(t))$, where Φ is the cumulative distribution function of the standard normal distribution. We reject H_0 at a significance level of 0.05.

2.7 Goodnes of Fit

Hunter *et al.* (2008a) present an approach for the goodness of fit, which compares the observed statistics to the average of statistics of networks simulated from the fitted model. The statistics that are being compared are not (only) the ones used for fitting, but so-called *auxiliary statistics*, which capture the local and global characteristics of the observed network. Typically statistics like the degree distribution or the geodesic distance are used. Auxiliary statistics describe properties of networks that are not explicitly modeled. If the model fits the data well, the observed auxiliary statistics should lie in the center of the distributions of the auxiliary statistics of the

networks simulated from the estimated model. In Section 3.3.4 we will see how the auxiliary statistics can be used to conduct a Mahalanobis distance-based Monte Carlo goodness of fit test.

2.8 Application: Fitting an ERGM using `statnet`

Before we move on to the simulation study, we apply the theory and fit an ERGM to the friendship network using the R-package ***statnet***. It is a collection of packages for statistical network analysis and for our purposes loading the package ***ergm*** (Version 3.11.0), which is part of ***statnet***, is sufficient. In R-Code 2.1, we load the data containing the information about the friendship relations and transform the adjacency matrix to a network object. We also load the nodal and dyadic attribute data. We then add the two vectors containing the information about the nodal attributes *gender* and *delinquency* to the network object. The variable *gender* takes the values 1 (girls) and 2 (boys). We decide on a model that is based on the social circuit dependence assumption. We include the edge count (***edges***), which has the role of an intercept, as well as the two network effects reciprocity (***mutual***) and the transitive geometrically weighted edgewise shared partner effect (***gwesp(decay=0.6, fixed=TRUE)***, where 0.6 is the value chosen for the parameter β_t^{trans} as given in Table 2.3). Furthermore, we include for the nodal attribute *gender* a sender (***nodeofactor("gender")***), receiver (***nodeifactor("gender")***) and uniform homophily (***nodematch("gender")***) effect. The same effects are also included for the variable *delinquency* (***nodeofactor("delinquency")***, ***nodeifactor("delinquency")***, ***nodematch("delinquency")***). Finally, we add ***edgecof("primary")***, which is the dyadic attribute effect that counts the number of friendship ties between pupils that have already been friends in primary school. The function ***summary()*** applied to the model formula prints the counts of the effects in the given network. We check that none of the counts is equal to zero.

R-Code 2.1. Loading the data and print the statistics

```
library(ergm)

# read the adjacency matrix of the friendship network
friendship_net_mat<- as.matrix(read.table("data/friendship_net.dat"))

# transform matrix to network objects
friendship_net <- network(friendship_net_mat)

# load the nodal attribute data
attributes <- read.table("data/attributes.dat",header=TRUE)

# load the dyadic attribute data
primary <- as.matrix(read.table("data/primary.dat"))

# add the nodal attributes to the network object
friendship_net %v% "gender" <- attributes$gender
friendship_net %v% "delinquency" <- attributes$delinquency

# print the statistics
```

```
summary(friendship_net ~ edges +
                         mutual +
                         gwesp(decay=0.6, fixed=TRUE) +
                         nodeofactor("gender") +
                         nodeifactor("gender") +
                         nodematch("gender") +
                         nodeofactor("delinquency") +
                         nodeifactor("delinquency") +
                         nodematch("delinquency") +
                         edgecov(primary))
##                     edges                    mutual
##                  181.0000                   42.0000
##           gwesp.fixed.0.6      nodeofactor.gender.2
##                  282.6775                   77.0000
##      nodeifactor.gender.2          nodematch.gender
##                   61.0000                  125.0000
## nodeofactor.delinquency.2 nodeofactor.delinquency.3
##                   38.0000                   43.0000
## nodeifactor.delinquency.2 nodeifactor.delinquency.3
##                   57.0000                   58.0000
##     nodematch.delinquency           edgecov.primary
##                   67.0000                   44.0000
```

We then fit the ERGM using the function *ergm()* (see R-Code 2.2). Within the function, we specify the seed (26) and the maximal number of iterations in the MCMCMLE algorithm (40). It takes a few seconds to fit this model. One can have a look at the result of the fit by applying *summary()* to the fit object. The output resembles the summary output of a standard linear model fit. The parameter estimates of the two network effects reciprocity and transitive geometrically weighted edgewise shared partners are positive. However, the reciprocity effect is not significant at a 0.05 significance level. The significant and positive transitive gwesp parameter implies that there is a tendency for transitive closure in the network. In particular, it means that an increase in shared friends increases the chances for friendship between two pupils. The positive sender effect of *gender* implies that gender 2 (i.e., the boys) tend to send more friendships than girls. The receiver effect of *gender* is not significant. The positive uniform homophily effect of *gender* provides evidence for the hypothesis that ties are more likely formed between pupils of the same gender. As *delinquency* has three levels and we treat it as a factor, there are two parameters for sender and receiver effect each. They are all significant and give evidence that pupils with delinquency level 2 or 3 send less but receive more ties than pupils with delinquency level 1. The uniform homophily effect of delinquency is not significant. However, the significant and positive dyadic attribute effect implies that friendship between two pupils is more likely if they have already been friends in primary school, as we would expect.

R-Code 2.2. Fitting an ERGM to the friendship network

```
# fit the ERGM
friendship_net_ergm <- ergm(friendship_net ~ edges +
                                             mutual +
                                             gwesp(decay=0.6, fixed=TRUE) +
                                             nodeofactor("gender") +
                                             nodeifactor("gender") +
                                             nodematch("gender") +
                                             nodeofactor("delinquency") +
                                             nodeifactor("delinquency") +
                                             nodematch("delinquency") +
                                             edgecov(primary),
                            control=control.ergm(seed=26, MCMLE.maxit = 40))

# show model summary
print(summary(friendship_net_ergm), print.call = FALSE)
## Monte Carlo Maximum Likelihood Results:
##
##                            Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                       -4.8084     0.4256      0 -11.298  < 1e-04 ***
## mutual                       0.4914     0.3531      0   1.392  0.16404
## gwesp.fixed.0.6              1.3541     0.2006      0   6.750  < 1e-04 ***
## nodeofactor.gender.2         0.9644     0.2037      0   4.735  < 1e-04 ***
## nodeifactor.gender.2        -0.3594     0.2151      0  -1.670  0.09483 .
## nodematch.gender             0.8693     0.1567      0   5.546  < 1e-04 ***
## nodeofactor.delinquency.2   -1.1514     0.2275      0  -5.060  < 1e-04 ***
## nodeofactor.delinquency.3   -0.9864     0.2193      0  -4.497  < 1e-04 ***
## nodeifactor.delinquency.2    0.7738     0.2491      0   3.106  0.00189 **
## nodeifactor.delinquency.3    0.7412     0.2376      0   3.120  0.00181 **
## nodematch.delinquency        0.2423     0.1970      0   1.230  0.21856
## edgecov.primary              0.9670     0.2426      0   3.985  < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 831.8  on 600  degrees of freedom
##  Residual Deviance: 555.9  on 588  degrees of freedom
##
## AIC: 579.9  BIC: 632.7  (Smaller is better. MC Std. Err. = 0.493)
```

With the function *mcmc.diagnostics()* (see R-Code 2.3) we can check whether the algorithm did really converge. For each effect, it displays the statistics of the simulated networks of the final iteration of the MCMCMLE algorithm. On the left, one can find the trace plot, and on the right the distributions of the centered statistics. In Figure 2.6, for example, one can find the convergence diagnostics for the edge, reciprocity, and the transitive gwesp effect. The statistics fluctuate around zero without any evident pattern, which is an indication of convergence.

R-Code 2.3. Convergence diagnostics

```
#convergence diagnostics
mcmc.diagnostics(friendship_net_ergm, which="plots")
```

One could furthermore assess the goodness of fit, either by using the built-in function *gof()* or by conducting a Mahalanobis distance-based Monte Carlo goodness of fit test as described in Section 3.3.4. Of course, one could also modify the model specification and check whether the goodness of fit can be improved.

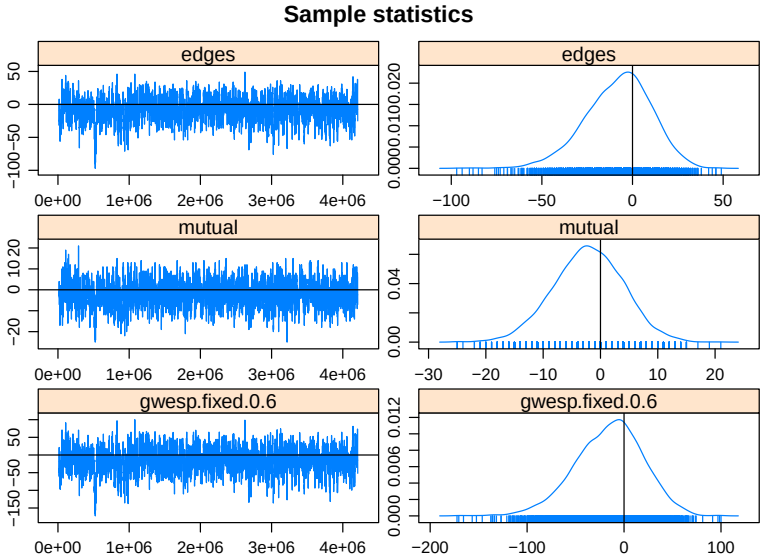


Figure 2.6: Convergence diagnostics of the edges, reciprocity and transitive gwesp effects: traceplots (left) and densities (right) of the centered statistics.

Chapter 3

Assessing Misspecifications: Motivation and Methods

3.1 Motivation

For linear regression models, the types and consequences of model misspecification have been thoroughly investigated. Lewis-Beck *et al.* (2004) name three broad categories of possible misspecifications: i) wrong selection of explanatory variables, in particular omitting an important variable, including a non-influential variable, or misspecifying the functional form of the relationship between explanatory variable and response; ii) errors in measurement; iii) specifying a wrong error term. In this thesis, we are interested in the first category: wrong selection of explanatory variables, or in the language of ERGMs: wrong selection of network statistics $\boldsymbol{z}(\boldsymbol{x})$. This is of special importance for ERGMs as the statistics which are included in the model are chosen based on sociological theory about the processes present in the network and little is known about the consequences of model misspecification so far (Lusher *et al.*, 2013, Ch. 22). At present, there exists no formal study on the consequences of model misspecification. The only consequence that is known is degeneracy when a wrong functional form to model network closure or the degree distribution is used. We will consider two types of misspecifications from the category wrong selection of explanatory variables: omitted variables and specifying the wrong functional form. This thesis is a first attempt to investigate the inferential consequences of model misspecification of these two types. Lewis-Beck *et al.* (2004) state that for linear regression models omitting a relevant variable x from the model has the most severe consequences. If x is not independent of the other explanatory variables, the parameter estimates will be biased and inconsistent, which might lead to incorrect inference and unreliable hypothesis tests. This motivates us to assess in a simulation study the consequences of omitted variables when fitting an ERGM. Considering misspecifications of the type wrongly specified functional form is motivated by the fact that the mathematical formulas of the configurations which represent social processes are chosen based on theories about these social processes. However, the same social process can be translated into different mathematical formulas. We want to assess what are the consequences when we choose the wrong functional form to represent a social process. The process for which we assess this misspecification is homophily, as homophily is an effect that is of interest in many empirical applications and that may be represented by many different mathematical formulas.

3.2 Setup of the Simulation Study

To assess the consequences of the two types of misspecifications, a simulation study is conducted. For that, we choose a set of statistics $\boldsymbol{z}(\boldsymbol{x})$ which specifies the baseline model. We also specify the corresponding vector of parameters $\boldsymbol{\theta}$, which are our true parameters. From this baseline model and true parameters, we simulate a large set of new networks. We then postulate different misspecified models, i.e., models where either variables are omitted or the wrong functional form is used. Next, the baseline model, as well as the misspecified models, are fitted to the simulated networks. In the following sections, we discuss the setup of the simulation study in more detail.

3.2.1 Baseline Models

A reasonable baseline model should include statistics that represent social processes that are considered to be important for many networks, and the parameters should be chosen in a way they might appear in real empirical examples. We also have to pay attention not to specify a degenerate model, as they do not realistically represent observed networks and as the estimation algorithm would not converge for many of the simulated networks. To make sure the baseline model is close to empirical examples, we use real data, i.e., empirically observed networks to which we fit an ERGM, and use the fitted statistics and their estimated parameters as a baseline model. We conduct the simulation study for two different baseline models: one is the model for the friendship network from Section 2.8, and one results from fitting an ERGM to a network that represents the communication between the executives of an organization in the entertainment industry. We, therefore, have two baseline models which come from two different important types of social networks: from a friendship network and a communication network. Both networks are directed, which is more common in empirical examples than undirected networks.
Let us have a closer look at the communication network, which we have not examined so far. The data can be downloaded from `www.melnet.org.au/ergm-book`. The 38 nodes represent executives of an organization. A tie from node i to node j means that actor i stated in a survey that it was important to communicate with j to get work done effectively. The binary nodal attribute *seniority* specifies whether the executive is senior or not. We moreover consider the discrete quantitative nodal attribute *projects* counting the number of projects completed by the corresponding executive. Figure 3.1 displays the network, where the size of the nodes represents the value of the attribute *projects* and the dark nodes correspond to executives which are senior, while the light nodes correspond to executives which are nonsenior.
As for the friendship network, we fit a model and obtain parameter estimates, which we will use for the simulation. The model specification is quite similar to the one used to fit the friendship network, but let us point out some differences. Additionally to the transitive gwesp effect, we add a cyclic gwesp effect. The corresponding estimated parameter is significant and negative, while the parameter of the transitive gwesp effect is significant and positive. The combination of these two parameter estimates implies that there is a tendency towards transitive closure and against cyclic closure. We could describe this as the ties tending to close triangles hierarchically. We also have the quantitative variable *project* for which we include a quantitative homophily effect, namely the absolute homophily effect, which sums up the absolute differences of the attribute values of the nodes of each tie.

In Table 3.1, one can find the model specifications as well as the parameters obtained by fitting the ERGMs. These two baseline models (**B1** and **B2**) and corresponding true parameters are then used to simulate the networks.

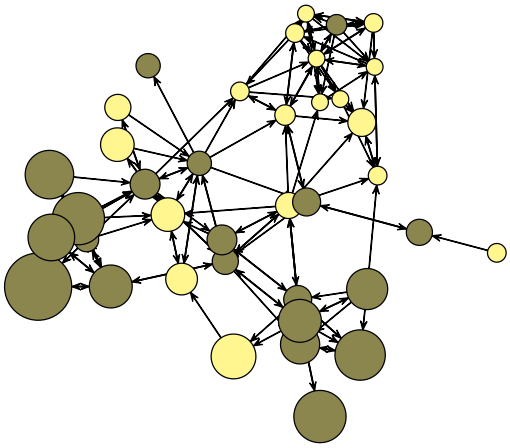


Figure 3.1: Communication network for which the dark nodes correspond to senior actors and the light nodes correspond to actors that are not senior. The larger a node, the more projects the actor has completed.

3.2.2 Simulation of Networks from the Baseline Models

For both baseline models, we simulate 1200 networks with the function *ergm:::simulate.ergm()* from the R-package *statnet* using a thinning of 12'000 and setting a seed (43). High thinning is chosen to reduce the autocorrelation within the sample. We then have a look at the distribution of the statistics in the simulated networks and compute the correlation between the statistics. This is done in Section 4.1. The correlation might help interpret the consequences of model misspecification. Given two highly correlated statistics, we for example expect that there are consequences for the parameter estimates of one statistic when the other statistic is omitted from the model.

3.2.3 Postulating Misspecified Models

For both correct baseline models, we postulate a set of misspecified models. We distinguish between omitted variables and wrong functional form. To easily reference the various misspecifications, we use the following notation: a capital "O" refers to an omitted variable, a capital "F" refers to wrong functional form. We then enumerate the different misspecifications and add "-c" or "-f" to specify whether the misspecification is postulated for the communication or friendship model, respectively. In Table 3.2 we list the misspecifications of the type omitted variables. We distinguish between omitted network effects (green) and omitted attribute effects (blue). Omitting the effects gwesp and homophily is of particular interest to us. There are several reasons for this. Both effects are the variables of interest in many empirical examples. We also expect quite some correlation between the two effects, as clustering among nodes with the same attribute increases the gwesp statistic as well as the homophily statistic. We, therefore, wonder what happens to the homophily effect, when the gwesp effect is omitted and vice versa. For a researcher, it might also be tempting to omit the gwesp effect, as this effect makes the estimation more difficult. The estimation algorithm takes more time, one has to decide on a value for the decay parameter β_t and in general, there is an increased chance to specify a degenerate model, as when leaving out any network effects other than reciprocity. For the communication network, we also have a particular interest in assessing the consequences, when omitting just one of the two gwesp

Table 3.1: Model specifications and parameters used for simulating networks. The parameters are obtained by fitting the two model specifications to the communication and friendship network, respectively.

Effect	**Parameter**
Baseline model 1 (communication network)	
edges	−3.868
reciprocity	3.302
transitive gwesp (decay = log(2))	1.069
cyclic gwesp (decay = log(2))	−0.531
sender *seniority*	0.341
receiver *seniority*	−0.354
homophily *seniority*	0.519
sender *projects*	0.003
receiver *projects*	0.021
homophily *projects*	−0.086
Baseline model 2 (friendship network)	
edges	−4.948
reciprocity	0.597
transitive gwesp (decay = 0.6)	1.390
sender *gender*	0.934
receiver *gender*	−0.381
homophily *gender*	0.842
sender *delinquency (level 2)*	−1.052
sender *delinquency (level 3)*	−0.923
receiver *delinquency (level 2)*	0.890
receiver *delinquency (level 3)*	0.839
homophily *delinquency*	0.281
dyadic attribute *primary*	0.510

effects. In this case, the tendency for hierarchical closure is only partially modeled. We do not specify a misspecification, where we omit the reciprocity effect, as it is common knowledge to include this effect and it is hardly ever omitted.

In Table 3.3, one can find the misspecifications of type wrong functional form. As explained in Section 3.1, we concentrate on the effect of homophily. Recall that for the communication network we have the quantitative nodal attribute *projects* for which the homophily effect in the correct model is measured by summing up the absolute differences of the project values of the two nodes i and j of a tie. Let p_i denote the number of projects completed by node i. The homophily effect of project in the correct model is given by $\sum_{i,j} |p_i - p_j| \, x_{ij}$ (absdiff). We define a misspecification where $\sum_{i,j} (p_i - p_j)^2 \, x_{ij}$ (squared) and one where $\sum_{i,j} (p_i - p_j) \, x_{ij}$ (difference) is used instead of the absdiff statistic. Recall from Section 2.3.3 that the three

Table 3.2: Misspecifications of type omitted variables (green network effects and blue attribute effects).

Name	Misspecification
For the communication networks:	
O1-c	omit transitive gwesp
O2-c	omit cyclic gwesp
O3-c	omit transitive and cyclic gwesp
O4-c	omit homophily of *seniority*
O5-c	omit homophily of *projects*
O6-c	omit sender, reciever and homophily of *seniority*
O7-c	omit sender, reciever and homohpily of *projects*
O8-c	omit all attribute effects
For the friendship networks:	
O1-f	omit transitive gwesp
O2-f	omit homophily of *gender*
O3-f	omit homophily of *delinquency*
O4-f	omit sender, reciever and homophily of *gender*
O5-f	omit sender, reciever and homohpily of *delinquency*
O6-f	omit dyadic attribute *primary*
O7-f	omit all attribute effects

statistics represent slightly different social processes and that the difference statistic is the only one taking the direction of the tie into account. When fitting the difference homophily effect together with sender and receiver effect as well as the edge count, we need to be careful, as the model suffers from colinearity. We, therefore, omit the sender effect of *projects* for this misspecification. One further misspecification concerning the homophily effect of *projects* which we define is using the statistic $\sum_{ij} \mathbb{1}_{\{|p_i - p_j| < c\}} x_{ij}$ (smalldiff) instead of absdiff. The smalldiff statistic counts the number of ties (i, j) for which the absolute difference of the project values p_i and p_j is smaller than some constant c. This effect assesses whether executives with a similar number of projects completed are more likely to state that they have important communication with each other. The values from the covariate *projects* range from 0 to 27 and we decide on the threshold $c = 6$.

For the two binary variables *seniority* and *gender*, we specified the uniform homophily effect in the correct models, which count the number of ties between nodes with the same attribute. We want to assess, what are the consequences when using a *differential* homophily effect instead of a uniform one, i.e., having two statistics instead of one, where one counts the number of ties between nodes with the first attribute value and the other counts the number of ties between nodes with the second attribute value. If the differential homophily effect is included in the model together with sender and receiver effects, as well as the edge count, the model suffers from perfect colinearity. The value of one of these statistics can be calculated analytically given all other statistics. We, therefore, exclude the sender effect of *seniority* in the misspecification

where we specify a differential instead of a uniform homophily effect for *seniority*. We exclude the receiver effect of *gender* in the misspecification where we specify a differential instead of a uniform homophily effect for *gender*. For the variable *delinquency*, which has three levels, we also postulate a misspecification where we specify a differential instead of a uniform homophily effect. As the variable *delinquency* has three levels, this misspecification does not suffer from perfect colinearity and we do not exclude any other effect.

Table 3.3: Misspecifications of type wrong functional form.

Name	Misspecification
For the communication networks:	
F1-c	squared instead of absdiff homophily effect for *projets*
F2-c	difference instead of absdiff homophily effect and omit sender effect for *projets*
F3-c	smalldiff instead of absdiff homophily effect for *projets*
F4-c	differential instead of uniform homophily effect and omit sender effect for *seniority*
For the friendship networks:	
F1-f	differential instead of uniform homophily effect and omit receiver effect for *gender*
F2-f	differential instead of uniform homophily effect for *delinquency*

3.2.4 Fitting the True and Misspecified Models

We fit the correct as well as the misspecified models to the simulated networks. The duration to fit a misspecification to the 1200 simulated networks varies and depends on the misspecification. Since the estimation algorithm does not always converge sample sizes may slightly differ. Recall that the MCMCMLE algorithm is a stochastic algorithm, which makes it necessary to set a seed (we set seed to 28) to make sure that the results are reproducible.

3.3 Methods of Evaluation

The evaluation of the simulation study consists of four parts: we perform a preliminary screening, we assess the power of the models by computing the percentage of parameter estimates that were identified to be significantly different from zero, we assess the Type I error rate by computing the percentage of parameter estimates which are significantly different from the true parameter and we perform a Mahalanobis distance-based Monte Carlo goodness of fit test and assess how well it can detect the misspecifications. In the next sections, we will explain the four parts of the evaluation in more detail.

3.3.1 Components of the Preliminary Screening

The following steps are part of our explanatory analysis:

- For some of the simulated networks and for some of the (mis)specifications, the estimation algorithm did not work due to degeneracy or because the networks were too sparse. For

these networks, we do not get any parameter estimates. For each specification, we count the number of networks for which we did not receive the estimates. We also check whether there are networks for which the estimation did not work for many of the specifications and have a look at the properties of these networks.

- For some of the networks and specifications, the estimation algorithm did not converge after the maximum of 40 iterations of the whole algorithm. For these fits, we have results that are not meaningful. We, therefore, check, for which fits the algorithm needed 40 iterations and have a look at the corresponding convergence diagnostics. If there is evidence that the algorithm did not converge, we remove the fits for further analysis.

- We check whether the estimates contain NA values. They are most often due to colinearity between the model statistics. We remove all fits with NA values for further analysis.

- For the approximate Wald test, we assume that the test statistics — estimate divided by the corresponding standard error — are approximately normally distributed. We create *qq*-plots to check this assumption.

3.3.2 Assessing the Power

The simulated networks are draws from the probability distribution determined by the baseline model. All the effects of the baseline model specification are influential for the simulated networks. We want to assess, whether the correct, as well as the misspecified models, can recognize the effects of the baseline model to be significantly influencing the network, i.e., whether they obtain parameter estimates that are significantly different from zero. For each specification, we, therefore, compute the proportion of estimates, which are significantly different from zero. Recall from Section 2.6 that the test statistic is given by $\widehat{\theta}_k^i / \widehat{\mathrm{SE}}(\widehat{\theta}_k^i)$, where $\widehat{\theta}_k^i$ denotes the parameter estimate of effect k in the simulation number i. We compare the test statistic to the standard normal distribution for every simulation i and compute the proportion of significant test statistics among the simulation results for a given effect. Instead of just using one significance level α for the test and obtaining one value for the proportion, we compute the proportion of significant estimates for many different levels α between 0 and 1. We assemble the resulting proportions into receiver operating characteristic (ROC) curves. For each specification, the ROC curves display the empirically estimated probability of detecting an effect as significant as a function of the significance level α. The higher the curve, the higher the power of the model specification to detect the effect as being significant. We expect the correct model specification to have high power to detect most of the effects as being significant. For some effects with very small true parameters (as the receiver and sender effects of *projects* in the communication networks), the power might be lower than for effects with larger true parameters. We can then compare the power of the misspecified models to the power of the correct model.

3.3.3 Assessing the Type I Error Rate

To asses the bias, we compute the proportion of parameter estimates that are significantly different from the true parameter. This is the Type I error rate under the null hypothesis that the estimated parameter is equal to the true parameter. Similarly as for the power, we draw a ROC curve that displays the Type I error rate as a function of the level α used for the significance testing. For each model specification and each effect, the Type I error ROC curve displays the

empirically estimated probability that the parameter estimate is different from the true parameter of the effect. The proportions are computed by comparing the ratio $\big(\widehat{\theta}_k^i - \theta_k^{\text{true}}\big)/\widehat{\text{SE}}\big(\widehat{\theta}_k^i\big)$ to the standard normal distribution for each simulation i and computing the share of ratios that are significant (significant given a certain level α).

3.3.4 Mahalanobis distance-based Monte Carlo Goodness of Fit Test

The Mahalanobis distance-based Monte Carlo (MDMC) goodness of fit test introduced by Lospinoso and Snijders (2019) can be used to test the hypothesis that a fitted model is equal to the true underlying model. The test employs the approach by Hunter *et al.* (2008a) of using auxiliary statistics to construct a goodness of fit test. The advantage of the MDMC test for our purposes is that it is only one test per set of auxiliary statistics. This enables us to draw a ROC curve with the proportion of significant test results in our simulations as a function of the significance level α. With this ROC curve, we can assess whether the MDMC test can detect the misspecifications. We can further investigate which of the sets of auxiliary statistics work best for most of the types of misspecifications. We first have a closer look at the MDMC test before we explain how we apply the test to our simulations and fitted models.

Let $\boldsymbol{x}$ be an observed network to which a model was fitted and the corresponding vector of parameters $\widehat{\boldsymbol{\theta}}$ was estimated. Furthermore, let $\boldsymbol{a}(\boldsymbol{x})$ be a k-dimensional vector of auxiliary statistics of the network $\boldsymbol{x}$. The auxiliary statistics could for example be given by the in-degree distribution, i.e., consisting of counts of nodes that have a certain number of incoming ties. We now want to test, whether $\boldsymbol{x}$ is a plausible realization of the distribution determined by $\widehat{\boldsymbol{\theta}}$, by comparing $\boldsymbol{a}(\boldsymbol{x})$ to $\boldsymbol{a}(\boldsymbol{X})$, where $\boldsymbol{X}$ follows the distribution implied by $\widehat{\boldsymbol{\theta}}$. If we would know the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}(\boldsymbol{X})$, we could compute the Mahalanobis distance between $\boldsymbol{a}(\boldsymbol{x})$ and $\boldsymbol{a}(\boldsymbol{X})$. The Mahalanobis distance is given by

$$D_{\widehat{\boldsymbol{\theta}}}\big(\boldsymbol{a}(\boldsymbol{x})\big) = \big(\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{\mu}\big)^\top \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{\mu}\big).$$

However, as we cannot compute the mean and covariance matrix of $\boldsymbol{a}(\boldsymbol{X})$ analytically, we simulate a set of networks $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ which are a realization of $\boldsymbol{X}$ and estimate the Mahalanobis distance by using sample estimates of the mean and covariance matrix. Let $\widehat{\boldsymbol{\mu}}$ be the sample mean of $\boldsymbol{a}(\boldsymbol{y}_1), \ldots, \boldsymbol{a}(\boldsymbol{y}_m)$ and let $\widehat{\boldsymbol{\Sigma}}$ be the sample covariance matrix. The estimated Mahalanobis distance is then given by

$$\widehat{D}_{\widehat{\boldsymbol{\theta}}}\big(\boldsymbol{a}(\boldsymbol{x})\big) = \big(\boldsymbol{a}(\boldsymbol{x}) - \widehat{\boldsymbol{\mu}}\big)^\top \widehat{\boldsymbol{\Sigma}}^{-1}\big(\boldsymbol{a}(\boldsymbol{x}) - \widehat{\boldsymbol{\mu}}\big). \tag{3.1}$$

Note that if the covariance matrix is the unit matrix, the Mahalanobis distance is equal to the Euclidean distance between the observation and the sample mean $\widehat{\boldsymbol{\mu}}$. The Mahalanobis distance has the advantage that it takes into account the correlation among the auxiliary statistics. The points with one unit distance from the mean therefore form an ellipsoid around the mean. We now want to test if the distance $\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{x}))$ has the magnitude we expect from a sample from the distribution $\boldsymbol{X}$. The test is constructed by computing the p-value $\text{P}\big(\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{X})) > \widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{x}))\big)$. As we do not know the distribution of $\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{X}))$, we use a parametric bootstrap approach to estimate the p-value. This is done by computing the Mahalanobis distance from each auxiliary statistic $\boldsymbol{a}(\boldsymbol{y}_1), \ldots, \boldsymbol{a}(\boldsymbol{y}_m)$ to the whole sample and then compute the proportion of Mahalanobis

distances that are larger than $\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{x}))$. The estimated p-value is therefore given by

$$\frac{1}{m}\sum_{h=1}^{m}\mathbb{1}_{\left\{\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{y}_h))>\widehat{D}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{a}(\boldsymbol{x}))\right\}}.$$

We incorporate the MDMC test in our evaluation of the simulation study conducting the following steps:

- We select a model (the correct model or a misspecified one). We have given the 1200 networks $\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(1200)}$, simulated from the correct model and the parameter estimates $\widehat{\boldsymbol{\theta}}^{(1)},\ldots,\widehat{\boldsymbol{\theta}}^{(1200)}$, obtained from fitting the selected model. For each of the parameter estimates $\widehat{\boldsymbol{\theta}}^{(h)}$, we simulate 100 networks $\boldsymbol{y}_1^{(h)},\ldots,\boldsymbol{y}_{100}^{(h)}$ which are realizations of the distribution determined by $\widehat{\boldsymbol{\theta}}^{(h)}$.
- We use the four different sets of auxiliary statistics that are suggested by Lospinoso and Snijders (2019): the geodesic distance distribution, the in-degree distribution, the out-degree distribution, and the triad census. The triad census is a vector containing the counts of the 16 possible types of triads in a directed network, where one type is dropped to avoid the linear dependency. For the geodesic distance and the degree distributions, we do not use the full auxiliary statistics that describe the distributions completely, as many of the statistics counting the number of nodes with high degrees or counting large geodesic distances are zero anyway. Therefore, we only use as auxiliary statistics the number of nodes with in-degree or out-degree ranging from 0 to 5 and the number of nodes with geodesic distance ranging from 0 to 5.
- For each of the simulated networks $\boldsymbol{y}_1^{(h)},\ldots,\boldsymbol{y}_{100}^{(h)}$ and for every $h\in\{1,\ldots,1200\}$, we compute the auxiliary statistics of each type.
- We compute the auxiliary statistics of the network $\boldsymbol{x}^{(h)}$ for every $h\in\{1,\ldots,1200\}$.
- We compute the mean and covariance matrix of the auxiliary statistics of $\boldsymbol{y}_1^{(h)},\ldots,\boldsymbol{y}_{100}^{(h)}$ for every h.
- We compute the estimate of the Mahalanobis distance $\widehat{D}_{\widehat{\boldsymbol{\theta}}^{(h)}}(\boldsymbol{a}(\boldsymbol{x}^{(h)}))$ for every h using equation (3.1).
- For every h, for 30 different levels α, we compute whether the Mahalanobis distance $\widehat{D}_{\widehat{\boldsymbol{\theta}}^{(h)}}(\boldsymbol{a}(\boldsymbol{x}^{(h)}))$ is significant at level α using the parametric bootstrap approach.
- We compute the proportion of significant Mahalanobis distances.
- We plot a ROC curve for each of the four types of auxiliary statistics.

3.4 Research Questions

We formulate a set of research questions that we want to answer with the simulation study. We will present those results that are relevant to answering the research questions and the discussion will be conducted based on them.

1. Do the baseline models have low Type I error (Type I error ROC curves close to the diagonal) and high power (power ROC curves close to the upper left corner)? Does the

MDMC test recognize the baseline models as the correct fit (MDMC test ROC curves close to the diagonal)?

2. What are the consequences on power and Type I error of other network effects, when we omit a gwesp effect?

3. What are the consequences on power and Type I error of the homophily effects, when we omit a gwesp effect?

4. What are the consequences on power and Type I error of the network effects, when we omit a homophily effect, omit all effects of one attribute variable, or omit all attribute effects?

5. What are the consequences on power and Type I error of one of the homophily effects, when the other homophily effect or all effects of one of the attribute variables is omitted?

6. What are the consequences on power and Type I error of the network effects, when the dyadic covariate effect is omitted?

7. What are the consequences on power and Type I error of the homophily effects, when the dyadic covariate effect is omitted?

8. What are the consequences on power and Type I error, when specifying a wrong functional form for the homophily effect?

9. How well does the MDMC test detect the different misspecifications and which of the sets of auxiliary statistics work best for most of the misspecifications?

Chapter 4

Results

We present the results of the simulation study starting with the properties of the simulated networks and the preliminary screening. Next, we consider the consequences for power and Type I error under the different model (mis)specifications. Finally, we show the results of the MDMC test. The discussion of the results can be found in Chapter 5.

4.1 Properties of the Simulated Networks

As motivated and explained in Section 3.2.2, we consider the correlations between the statistics of the simulated networks. The correlations between the statistics are quite pronounced. In Figure 4.1 for example, one can see that there is a very strong correlation between the transitive and cyclic gwesp statistics and that the correlation between the two homophily statistics of the nodal attributes *seniority* and *projects* is less pronounced but still clearly present in the simulated communication networks.

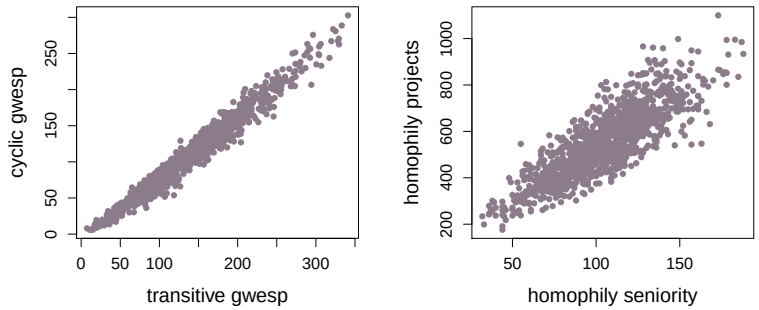


Figure 4.1: Scatter plots of the transitive and cyclic gwesp statistics (left) and the two homophily statistics (right) in the simulated communication networks.

We compute the correlations between our statistics of interest, i.e., between the network effects and the homophily statistics. In Table 4.1, one can find the correlations between the statistics of interest of the simulated communication networks. In Table 4.2, we display the ones of the simulated friendship networks. In the communication networks, the correlations are generally very high and only slightly smaller between homophily *project* and all other statistics listed. In the friendship networks, the correlations are smaller, but still quite high between the gwesp and the reciprocity effect, as well as between the gwesp and the homophily *gender* effect. Correlations

of 0.5 or larger are often observed, due to the fact that the statistics are all counts of configurations of ties present in the observed networks.

Table 4.1: The correlations between the statistics of interest of the simulated communication networks.

	Reciprocity	Transitive gwesp	Cyclic gwesp	Homophily *seniority*	Homophily *projects*
Reciprocity	1	0.96	0.97	0.94	0.84
Transitive gwesp	0.96	1	0.98	0.94	0.83
Cyclic gwesp	0.97	0.98	1	0.93	0.82
Homophily *seniority*	0.94	0.94	0.93	1	0.84
Homophily *projects*	0.84	0.83	0.82	0.84	1

Table 4.2: The correlations between the statistics of interest of the simulated friendship networks.

	Reciprocity	Transitive gwesp	Homophily *gen.*	Homophily *del.*	Dyadic attribute
Reciprocity	1	0.84	0.79	0.63	0.44
Transitive gwesp	0.84	1	0.84	0.72	0.49
Homophily *gen.*	0.79	0.84	1	0.61	0.41
Homophily *del.*	0.63	0.72	0.61	1	0.33
Dyadic attribute	0.44	0.49	0.41	0.33	1

4.2 Preliminary Screening

We have a first look at the outcome of fitting the baseline as well as the misspecified models, as described in Section 3.3.1. First, we check for how many networks per (mis)specification the model fitting did not work. Fitting **B1** to the 1200 simulated communication networks did not work in 27 cases and fitting **B2** to the 1200 friendship networks did not work in 195 cases. It is noticeable that there were no problems fitting the model for specifications when the transitive gwesp effect was omitted. It is well known that modeling network closure can cause degeneracy. Even though the gwesp effects have improved the problems with degeneracy compared to the triangle effect, there are still some specifications with a gwesp effect that result in degenerate models. From the misspecifications fitted to the communication networks, the most problematic one was **O8-c**, where all the attribute effects were omitted. For this misspecification, we did not receive estimates for 373 of the 1200 networks. From the misspecifications fitted to the friendship networks, three were problematic: **O4-f** (omitting all effects involving the attribute *gender*), **O5-f** (omitting all effects involving the attribute *delinquency*), and **O7-f** (omitting all attribute effects). As for these misspecifications, the estimation did not work for more than half of the simulated networks, we do not consider these misspecifications for further analysis.
Considering the edge count of the communication networks for which fitting of at least one of the misspecifications did not work, we observe rather low counts, indicating that the networks are rather sparse. For those networks, the counts of the different configurations might be too small and we fit too many parameters such that it is not possible to disentangle the effects. This is not the case for the friendship networks, where the issue lies with degeneracy.
We also search for the networks for which the algorithm did not converge after 40 iterations (see Section 2.5 for the description of the estimation method) and we discard the parameter estimates resulting from these fits. This happened for at most 14 networks per misspecification.

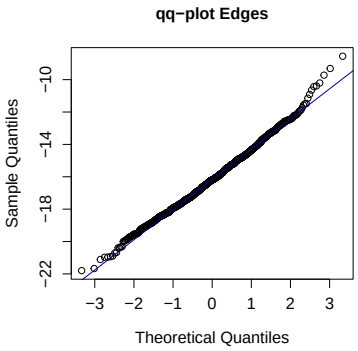


Figure 4.2: *qq*-plot of the parameter estimates divided by the corresponding standard errors of the edge effect in the communication baseline model.

Furthermore, we discard the parameter estimates which contain NA values. This is very rare (for at most 3 fits per misspecification). In the Appendix (Table A.1), one can find the final sample sizes used in the analysis.

We consider the normal *qq*-plots of the ratios $\widehat{\theta}_k/\widehat{\mathrm{SE}}(\widehat{\theta}_k)$ for each effect k, i.e., the *qq*-plots of the parameter estimates divided by the corresponding standard errors to check the normality assumption. We restrict ourselves to the estimates of the baseline models. The *qq*-plots indicate that the ratios are approximately normally distributed. In Figure 4.2, one can find as an example the *qq*-plot for the edge effect in the communication baseline model. The *qq*-plots for the other effects of interest can be found in the Appendix (Figure A.1 and Figure A.2).

4.3 The Baseline Models

We consider the results of fitting the baseline models (i.e., the correct models) to the simulated networks. Figure 4.3 shows the power ROC curves of our effects of interest of the two baseline models. They display the proportions of parameter estimates that are significantly different from zero as a function of the significance level α (see Section 3.3.2 for the explanations of the power assessment). A power ROC curve that is close to the diagonal indicates low power and one that is close to the upper left corner indicates high power to detect the effect as being significant. The power of the communication baseline model is very high for all effects. For the friendship baseline model, the power ROC curves for the homophily effect of *delinquency*, the reciprocity effect, and the dyadic attribute effect are not as close to the upper left corner as they are for the transitive gwesp and homophily *gender* effect. Note, however, that for those three effects the true parameters are rather small, which implies that the effects are not very pronounced in the simulated networks and therefore harder to detect. Recall that the effects reciprocity and homophily *delinquency* were also not significant when fitting the friendship baseline model to the real data from which the parameters were obtained.
Figure 4.4 reports the Type I error ROC curves of the baseline model. They display the proportions of estimates that are significantly different from the true parameters as a function of the significance level α (compare Section 3.3.3). Type I error ROC curves close to the diagonal indicate a low Type I error, i.e., that only few parameter estimates are significantly different from the true parameter. The Type I error ROC curves of the baseline models (**B1** and **B2**) are close to the diagonal.

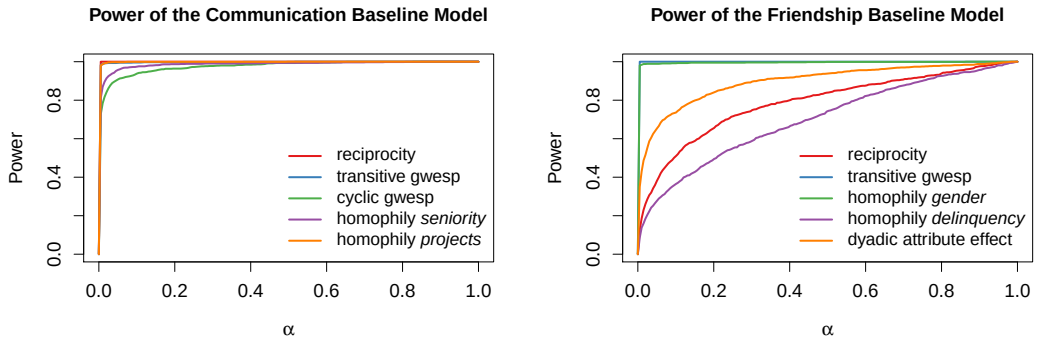


Figure 4.3: Power ROC curves of the effects of interest of the communication baseline model (left) and the friendship baseline model (right).

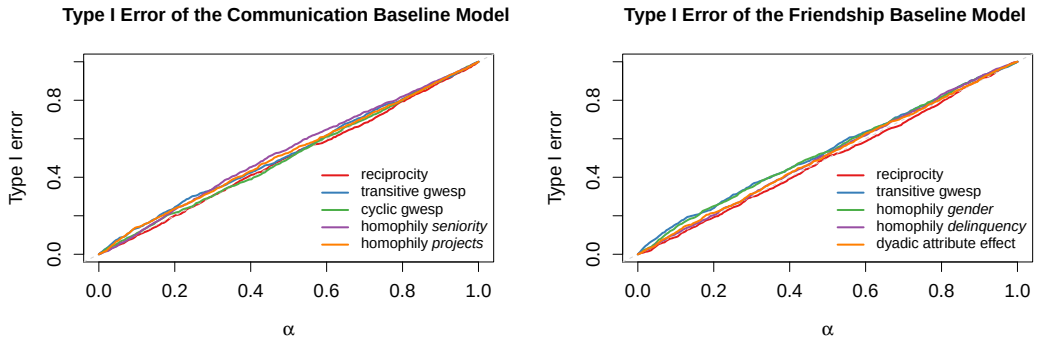


Figure 4.4: Type I error ROC curves of the effects of interest of the communication baseline model (left) and the friendship baseline model (right).

Moreover, we consider the MDMC test ROC curves of the two baseline models. They can be found in Figure 4.5. The ROC curves of the communication baseline model are close to the diagonal. The ROC curves of the friendship baseline model, however, are much closer to the upper left corner, especially for the triad census. That is, for a high number of simulated networks the sets of auxiliary statistics in the simulations are significantly different from the auxiliary statistics we would expect in a sample from the friendship baseline distribution.

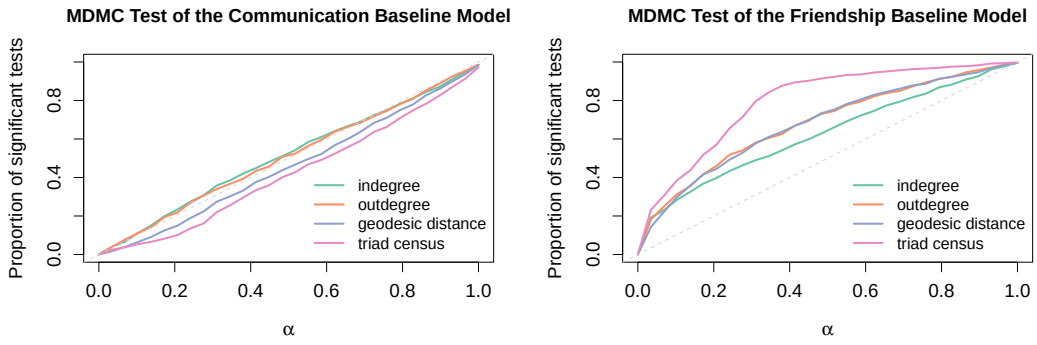


Figure 4.5: MDMC test ROC curves of the four different sets of auxiliary statistics of the communication baseline model (left) and the friendship baseline model (right).

4.4 Omitting Gwesp Effects

We consider the four misspecifications where we omit one or several gwesp effects. These are the misspecifications **O1-c** (transitive gwesp omitted from the communication baseline model), **O2-c** (cyclic gwesp omitted from the communication baseline model), **O3-c** (transitive and cyclic gwesp omitted from the communication baseline model), and **O1-f** (transitive gwesp omitted from the friendship baseline model). We distinguish between the consequences with regard to other network effects and the consequences with regard to the attribute effects. At the end of this section, Table 4.3 provides an overview of the consequences when omitting gwesp effects.

4.4.1 Results Regarding Other Network Effects

The remaining network effects in the communication networks are the reciprocity and cyclic gwesp effects for the misspecification **O1-c** and the reciprocity and transitive gwesp effects for the misspecification **O2-c**. For the misspecification **O3-c**, the only remaining network effect is reciprocity. For the misspecification **O1-f** fitted to the friendship networks, the only remaining network effect is the reciprocity effect as well.

The power to detect the remaining network effects is very high for all misspecifications. All ROC curves are very close to the upper left corner. Recall that the friendship baseline model did not have very large power to detect the reciprocity effect as a significant effect. The power to detect reciprocity is higher for the **O1-f** misspecification than for the baseline model (see Figure 4.6). For all the other misspecifications (**O1-c**, **O2-c**, and **O3-c**), the power to detect the network effects is exactly the same as for the baseline model. The corresponding power ROC curves can be found in the Appendix (Figure A.3).

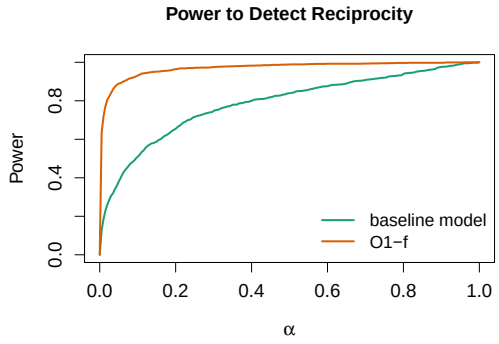


Figure 4.6: Power ROC curve of reciprocity for the baseline model and the misspecification **O1-f** (transitive gwesp omitted).

We consider the Type I error ROC curves misspecification-wise. Figure 4.7 (left) shows the Type I error ROC curves of the misspecification **O1-c** (omitted transitive gwesp). Note that the cyclic gwesp effect has a very large Type I error. That is, almost all parameters for the cyclic gwesp effect are significantly different from the true parameter when we omit the transitive gwesp effect. Figure 4.7 (right) displays the densities of the parameter estimates for the cyclic gwesp effect of the baseline model as well as of the misspecification **O1-c**. The vertical black line displays the true parameter. We can observe that the parameter estimate of the cyclic gwesp effect is positive for almost all simulated networks, while the true parameter is actually negative. Recall that the power to detect the significant cyclic gwesp effect is very high for the misspecification **O1-c**. The

estimated parameter for the cyclic gwesp is therefore significant, however, the estimate has the wrong sign for almost every simulated network. Furthermore, we can observe that the Type I error of reciprocity is small for this misspecification.

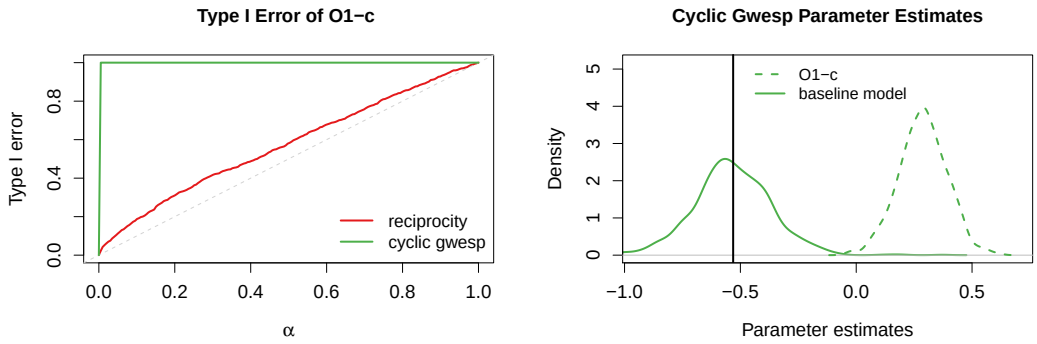


Figure 4.7: Type I error ROC curves of the remaining networks effect of the misspecification **O1-c** (left) and the densities of the parameter estimates for the cyclic gwesp effect of the baseline model and the misspecification **O1-c** (right).

Figure 4.8 (left) illustrates the Type I error ROC curves of the misspecification **O2-c** (omitted cyclic gwesp). Note that the transitive gwesp effect has a very large Type I error, as does the cyclic gwesp effect for the misspecification **O1-c**. Figure 4.7 (right) displays the densities of the parameter estimates for the transitive gwesp effect of the baseline model as well as of the misspecification **O2-c**. The vertical black line displays again the true parameter. We can observe that the parameter estimate of the transitive gwesp effect is underestimated in the misspecification. However, the estimated parameters are all positive and therefore have the correct sign. Recall that the power to detect the significant transitive gwesp effect is very high for this misspecification. Even though the parameter is underestimated, the effect is still detected as significant.
The Type I error of reciprocity is larger than for the misspecification **O1-c**, however, the parameter estimates all have the correct sign, i.e., they are positive, as the correct parameter. The parameter estimates are smaller than the correct parameter, i.e., the reciprocity effect is underestimated. This can be concluded from the density plot of the parameter estimates for the reciprocity effect given the misspecification **O2-c**. The corresponding density plot can be found in the Appendix (Figure A.4). Recall that the power to detect the reciprocity effect is high nonetheless.
For the misspecifications **O3-c** and **O1-f**, the only remaining network effect is the reciprocity effect. For the misspecification **O3-c**, it has a rather small Type I error (see Figure 4.9, left) and a larger Type I error for the misspecification **O1-f** (see Figure 4.9, right). In this case, the reciprocity effect is overestimated. The parameter estimates are positive, as is the true parameter, but they are larger in magnitude than the true parameter. This can be concluded from the density plot of the estimated parameters given in the Appendix (Figure A.4). Recall that the power to detect the reciprocity effect was larger for the misspecification **O1-f** than for the baseline model of the friendship networks.

4.4.2 Results Regarding the Attribute Effects

The power to detect the attribute effects is almost perfectly the same for the four misspecifications **O1-c**, **O2-c**, **O3-c**, and **O4-c** as for the baseline models and in general very high. The power ROC curves can be found in the Appendix (Figure A.5).

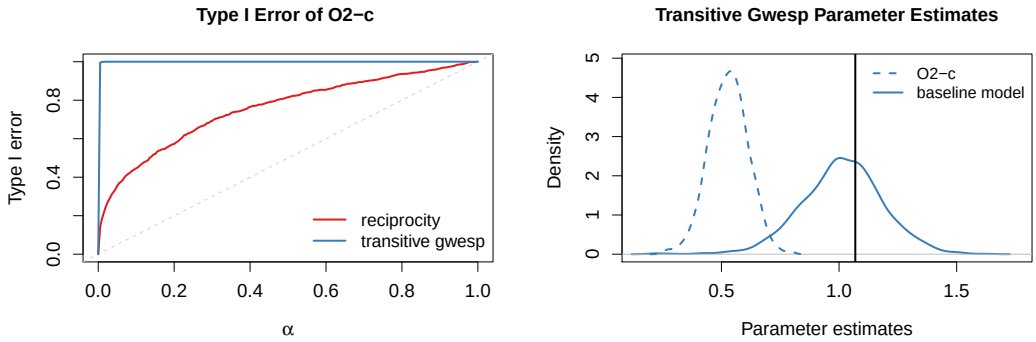


Figure 4.8: Type I error ROC curves of the misspecification **O2-c** (left) and the densities of the parameter estimates for the transitive gwesp effect of the baseline model and the misspecification **O2-c** (right).

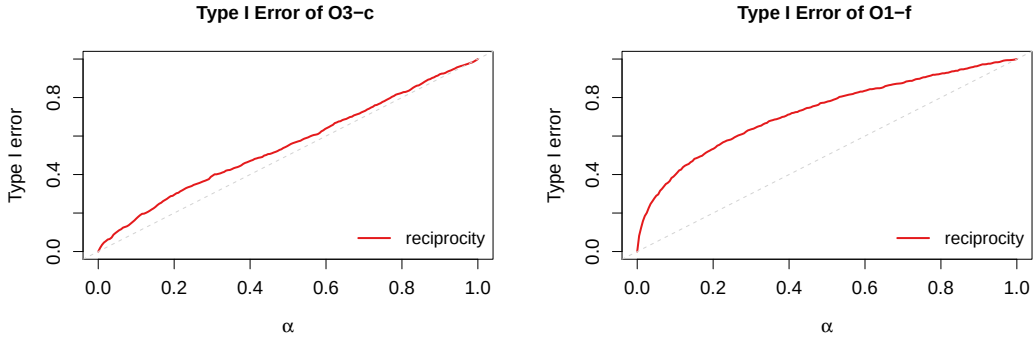


Figure 4.9: Type I error ROC curve of the reciprocity effect for the misspecification **O3-c** (left) and the misspecification **O3-f** (right).

Figure 4.10 shows the Type I error ROC curves of the attribute effects for the four misspecifications. We can observe that the Type I errors of the homophily effects of the communication model are largest for the misspecification **O3-c**, where both gwesp effects are omitted. The homophily effects are overestimated; the parameter estimates have the same sign as the correct parameter, but are larger in absolute values (see density plots in the Appendix, Figure A.6). The Type I error of the attribute effects of the misspecification **O1-f** (omitting transitive gwesp from the friendship baseline model) is the largest for the homophily *gender* effect (see Figure 4.10, bottom right). Also this effect is overestimated by the misspecification (see Figure A.6 in the Appendix).

Table 4.3 summarizes the results of fitting misspecifications where one or several gwesp effects are omitted.

4.5 Omitting Attribute Effects

We consider the eight misspecifications, where we omit one or several attribute effects. These are the misspecifications, where we omit one of the homophily effects (**O4-c**, **O5-c**, **O2-f**, and **O3-f**), the misspecification where we omit the dyadic attribute effect from the friendship baseline model (**O6-f**), the misspecifications where we omit all effects of one of the nodal attribute variables (**O6-c** and **O7-c**) and the misspecifications where we omit all attribute effects from

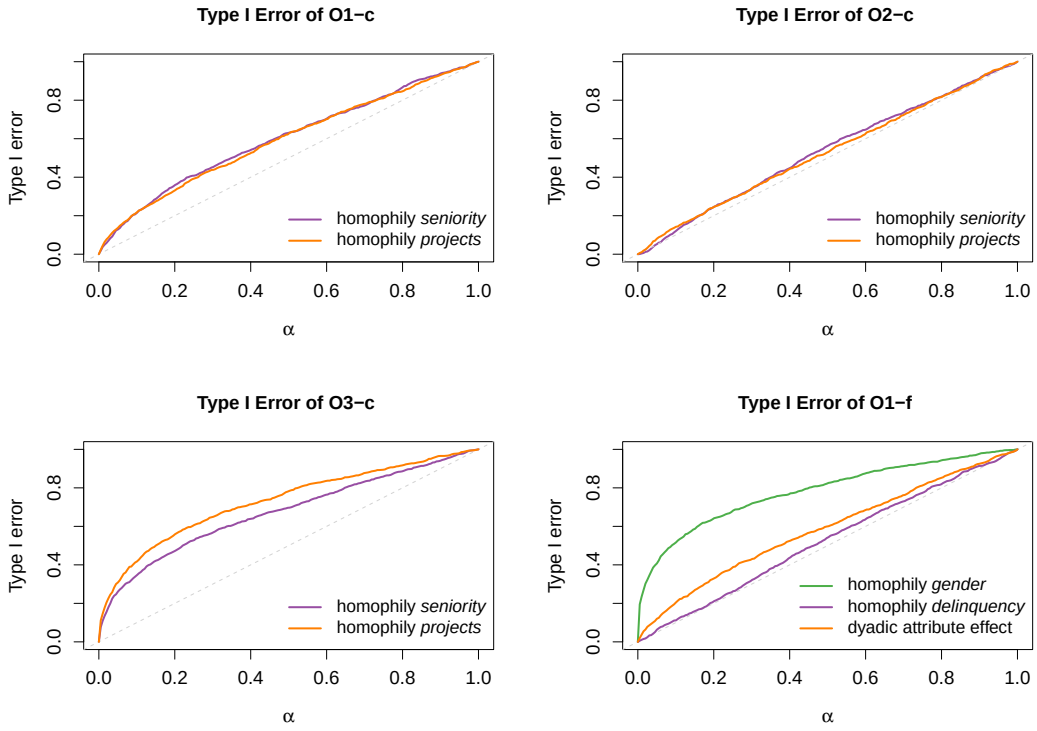


Figure 4.10: Type I error ROC curve of the attribute effects for the misspecification **O1-c** (top left), **O2-c** (top right), **O3-c** (bottom left), and **O1-f** (bottom right).

Table 4.3: Overview of the results of fitting misspecifications where one or several gwesp effects are omitted.

	Results regarding network effects		**Results regarding homophily effects**	
	Power	**Type I Error**	**Power**	**Type I Error**
O1-c	very high, as in **B1**	very large for cyclic gwesp, small for reciprocity	very high, as in **B1**	medium
O2-c	very high, as in **B1**	very large for trans. gwesp, large for reciprocity	very high, as in **B1**	small
O3-c	very high, as in **B1**	small for reciprocity	very high, as in **B1**	large
O1-f	very high; for reciprocity higher as in **B2**	medium for reciprocity	as in **B2**	large for hom. *gender*, small for the others

the communication baseline model (**O8-c**). Recall that we do not consider missepcifications **O4-f**, **O5-f** and **O7-f** for the analysis, as for these misspecifications, fitting the model did not work for a large proportion of networks. Again, we distinguish between the consequences with regard to the network effects and the consequences with regard to the remaining attribute effects of interest. An overview of the consequences when omitting attribute effects can be found in Table 4.4.

4.5.1 Results Regarding the Network Effects

The consequences for power and Type I error of the network effects are quite small when we omit attribute effects. The power to detect the network effect is exactly the same as for the baseline models (and in general very high), except for misspecification **O2-f** (omitting the effect

homophily *gender*). The power to detect the reciprocity effect is higher for this misspecification than for the baseline model (see Figure 4.11). The figures describing all other power ROC curves of the network effects can be found in the Appendix (Figure A.7).

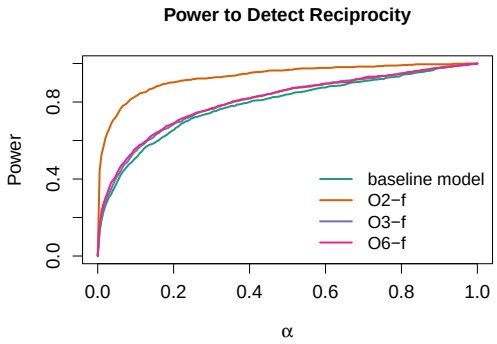


Figure 4.11: Power ROC curves of reciprocity of the friendship baseline model and the misspecifications fitted to the friendship networks.

The Type I errors are generally much smaller when omitting attributes effects, compared to omitting network effects. Since the Type I error ROC curves are all quite close to the diagonal and easy to apprehend, we display the Type I error ROC curves of the network effects and the attribute effects of interest together in one plot (Figure 4.12). The Type I error ROC curves of the network effects lie almost perfectly on the diagonal. The only exceptions are again misspecification **O2-f**, as well as the misspecification **O8-c**. The Type I error ROC curves of the network effects are slightly above the diagonal for **O2-f** (omitting homophily *gender* from the friendship baseline model). **O8-c** is the misspecification, where we omit all attribute effects. For this misspecification, the Type I error ROC curves of all network effects (reciprocity, transitive gweps, and cyclic gweps) are slightly larger than for the communication baseline model.

4.5.2 Results Regarding Other Attribute Effects

The power to detect the remaining attribute effect of interest is again very high and comparable to the baseline models. The only exception is again misspecification **O2-f**. The power to detect the homophily effect of the variable *delinquency* is higher for this misspecification than for the baseline model (see Figure 4.13). The figures describing all other power ROC curves of the attribute effects can be found in the Appendix (Figure A.8).

The Type I error ROC curves of the attribute effects of interest (together with the network effects) can be found in Figure 4.12. We find that the Type I errors of **O4-c** (omitting homophily *seniority*) and **O6-c** (omitting all effects of the nodal attribute *seniority*), as well as **O5-c** (omitting homophily *projects*) and **O7-c** (omitting all effects of the nodal attribute *projects*) follow a similar pattern. For all these misspecifications, the Type I error of the remaining homophily effect is slightly larger than in the communication baseline model. For the misspecifications fitted to the friendship networks, however, the Type I error of the remaining homophily effect is very small. Moreover, the ROC curve of the dyadic attribute effect is very close to the diagonal for these misspecifications. For misspecification **O2-f**, the ROC curve is even somewhat below the diagonal. Looking at the Type I error of **O6-f**, we observe that omitting the dyadic attribute effect also results in Type I errors very close to the diagonal.

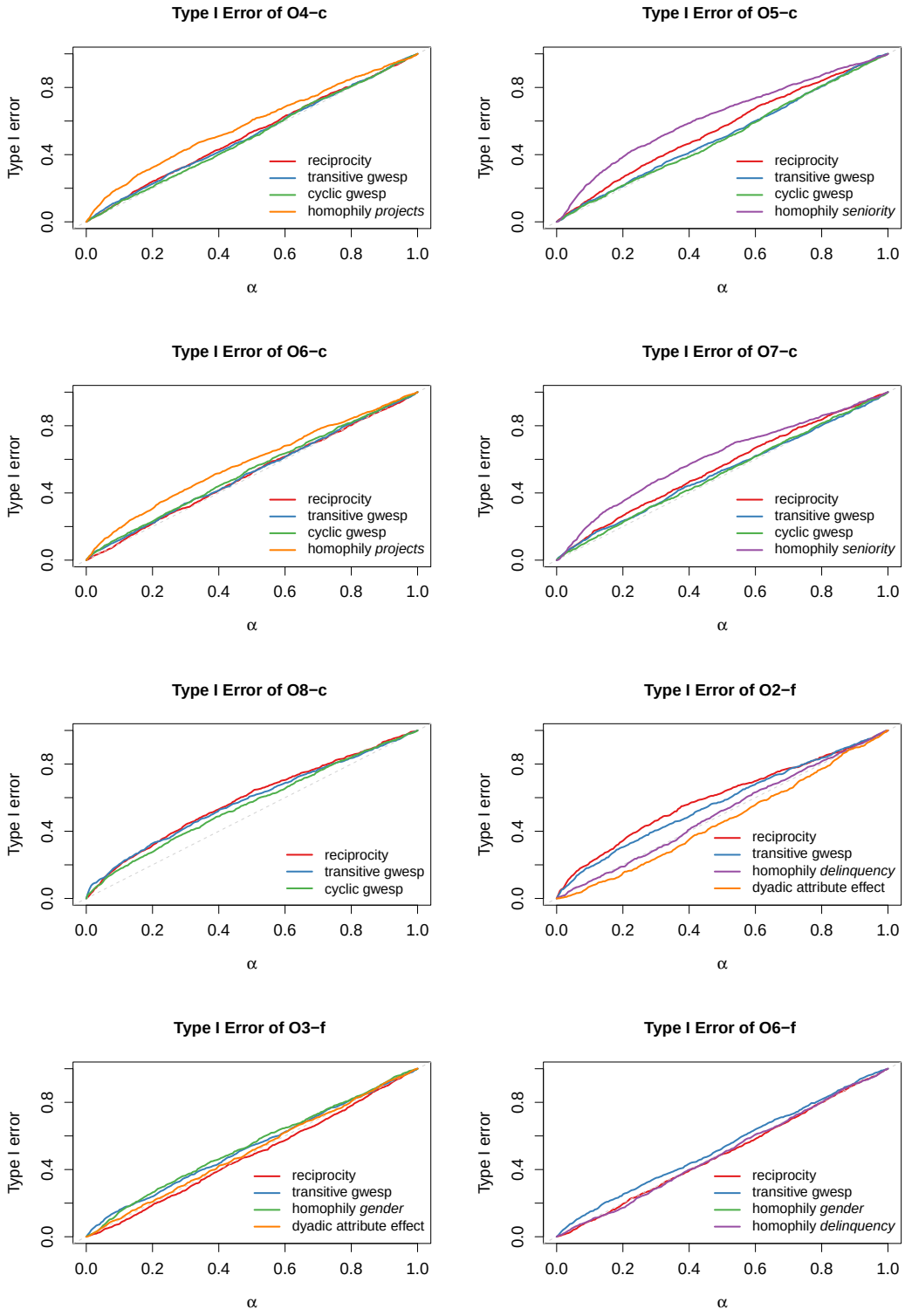


Figure 4.12: Type I error ROC curves of the different misspecifications where one or several attribute effects are omitted.

In general, we can observe that the consequences are much smaller when omitting attribute effects than when omitting network effects. The power and Type I error ROC curves of the misspecified models are very similar to the ones of the baseline models. In Table 4.4, we summarize the consequences for power and Type I error when omitting attribute effects.

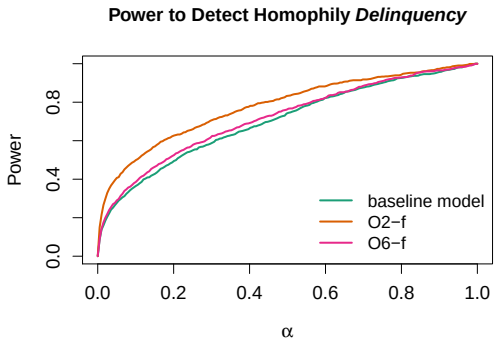


Figure 4.13: Power ROC curves of the effect homophily *delinquency* of the friendship baseline model and of the misspecifications fitted to the friendship networks.

Table 4.4: Overview of the results of fitting misspecifications where one or several attribute effects are omitted.

	Results regarding network effects		**Results regarding homophily effects**	
	Power	**Type I Error**	**Power**	**Type I Error**
O4-c	very high, as in **B1**	small	very high, as in **B1**	medium for hom. *projects*, small for the others
O5-c	very high, as in **B1**	small	very high, as in **B1**	medium for hom. *seniority*, small for the others
O6-c	very high, as in **B1**	small	very high, as in **B1**	very similar to **O4-c**
O7-c	very high, as in **B1**	small	very high, as in **B1**	very similar to **O5-c**
O8-c	very high, as in **B1**	small to medium	very high, as in **B1**	-
O2-f	very high; for reciprocity higher as in **B2**	medium	high; for hom. *delinquency* a bit higher as in **B2**	small
O3-f	high, as in **B2**	small	as in **B2**	small
O6-f	high, as in **B2**	small	as in **B2**	small

4.6 Specifying the Wrong Functional Form

We consider the six misspecifications of type wrong functional form. In the misspecifications **F1-c**, **F2-c**, and **F3-c**, instead of the absdiff homophily effect a squared, difference or smalldiff homophily effect is used, respectively. In the misspecifications **F4-c**, **F1-f**, and **F2-f**, instead of a uniform homophily effect a differential homophily effect is used. The power and Type I error ROC curves resulting from the simulation study are very similar to the ones resulting from the baseline models. Therefore, we omit the figures displaying the power ROC curves. The power to detect the remaining effects (network effects as well as correctly specified attribute effects) is therefore very high in general. The Type I error ROC curves are generally very close to the diagonal, as observed for the baseline models as well. The only two exceptions are the Type I error ROC curves of the homophily *seniority* effect for the misspecifications **F2-c** and **F3-c** (see Figure 4.14). The Type I error of the homophily *seniority* effect is larger for these misspecifications than for the baseline model. The figures displaying the Type I error ROC curves of the misspecifications **F1-c**, **F4-c**, **F1-f**, and **F2-f** can be found in the Appendix (Figure A.9).

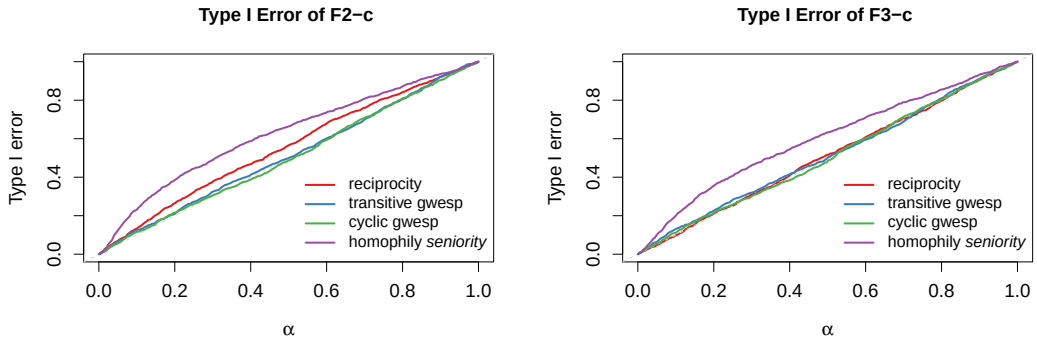


Figure 4.14: Type I error ROC curves of the misspecifications **F2-c** (left) and **F3-c** (right).

For the misspecifications of type wrong functional form, we additionally consider the parameter estimates for the misspecified homophily effects. Even though the effects have the wrong functional form, the parameter estimates might still be significantly different from zero, as there is a high correlation between the homophily statistic used in the baseline model and the homophily statistic corresponding to the misspecified effect. Figure 4.15 shows the power ROC curves corresponding to the misspecified homophily effects. The curves display the proportion of parameter estimates significantly different from zero. We can observe that while the squared and smalldiff homophily effects of *projects* have high power, this is not the case for the difference homophily effect. Recall that the difference homophily effect is the only one taking the direction of the ties into account. The power for the differential homophily effect of *delinquency* is very low, whereas the power for the differential homophily effect of *gender* and *seniority* is high for at least one of the groups.

Table 4.5 summarizes the results of fitting misspecifications of type wrong functional form.

Table 4.5: Overview of the results of fitting misspecifications of type wrong functional form.

	Results regarding network effects		**Results regarding homophily effects**	
	Power	**Type I Error**	**Power**	**Type I Error**
F1-c	very high, as in **B1**	small	very high, as in **B1**	small
F2-c	very high, as in **B1**	small	very high, as in **B1**	medium for hom. *seniority*, small for the others
F3-c	very high, as in **B1**	small	very high, as in **B1**	medium for hom. *seniority*, small for the others
F4-c	very high, as in **B1**	small	very high, as in **B1**	small
F1-f	high, as in **B2**	small	as in **B2**	small
F2-f	high, as in **B2**	small	as in **B2**	small

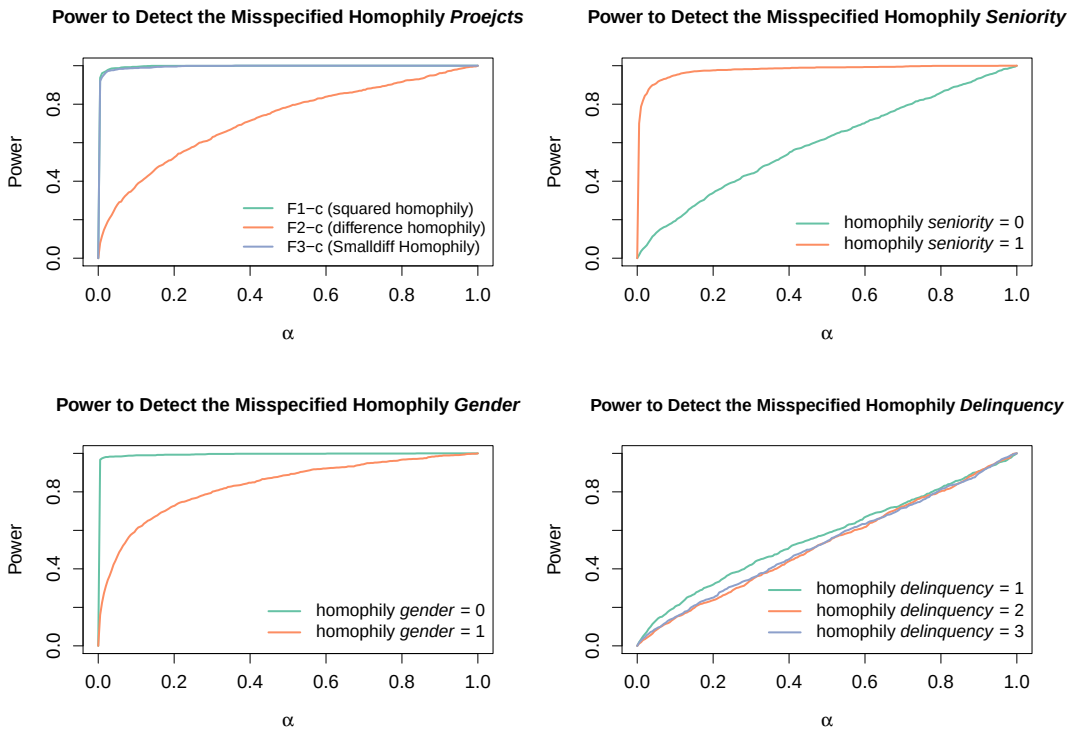


Figure 4.15: Proportion of parameter estimates significantly different from zero for the misspecified homophily effects.

4.7 Results of the MDMC Test

In this section, we consider the ROC curves illustrating the proportion of estimates with significant Mahalanobis distance-based Monte Carlo goodness of fit tests per misspecification as a function of significance level α. Recall that we use the in-degree distribution, out-degree distribution, geodesic distance distribution, and the triad census as auxiliary statistics (as explained in Section 3.3.4). MDMC test ROC curves close to the diagonal indicate that the model fits the network well with respect to the MDMC goodness of fit measure. We restrict ourselves to the MDMC test ROC curves of the misspecification fitted to the communication networks, since the MDMC test for the friendship baseline model has too large proportions of significant estimates and therefore does not perform as expected (recall Figure 4.5, right).

It is remarkable how different the ROC curves are for the different misspecifications fitted to the communication networks. Recall that the MDMC test ROC curves of the communication baseline model were all very close to the diagonal (Figure 4.5). The proportion of significant MDMC test is much larger for the three misspecifications where one or several gwesp effects are omitted, i.e., for **O1-c**, **O2-c**, and **O3-c** (see Figure 4.16). In particular, the proportion of significantly different triad census distributions is very large for these three misspecifications.
The MDMC test ROC curves of the misspecifications where one or several attribute effects are omitted are all very close to the diagonal except for the misspecification **O8-c**, where all attribute effects are omitted (see Figure 4.17). There, we can observe increased proportions of significant MDMC tests compared to the communication baseline model.
The MDMC test ROC curves of the misspecifications of type wrong functional form are all very close to the diagonal and the proportions are very similar to the proportions of the communication

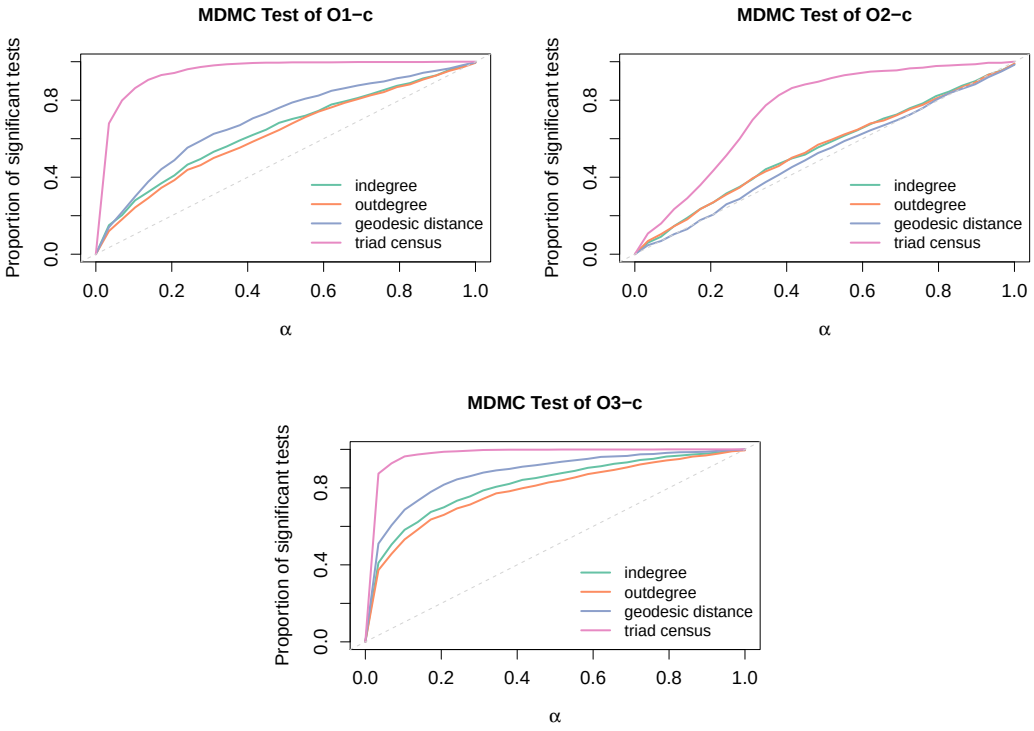


Figure 4.16: MDMC test ROC curves of the misspecifications where one or several gwesp effects are omitted from the communication baseline model.

baseline model (see Figure 4.18).

We can summarize that even though the models **O4-c**, **O5-c**, **O6-c**,**O7-c**, **F1-c**, **F2-c**, **F3-c**, **F4-c** are misspecified, fitting the models to the simulated networks results non-significant MDMC tests, i.e., in good fits. Fitting the models **O1-c**, **O2-c**, and **O3-c** results in a large proportion of significant MDMC tests, especially with respect to the triad census. Fitting the model **O8-c** results in a quite large proportion of significant MDMC tests, especially with respect to the geodesic distance auxiliary statistics.

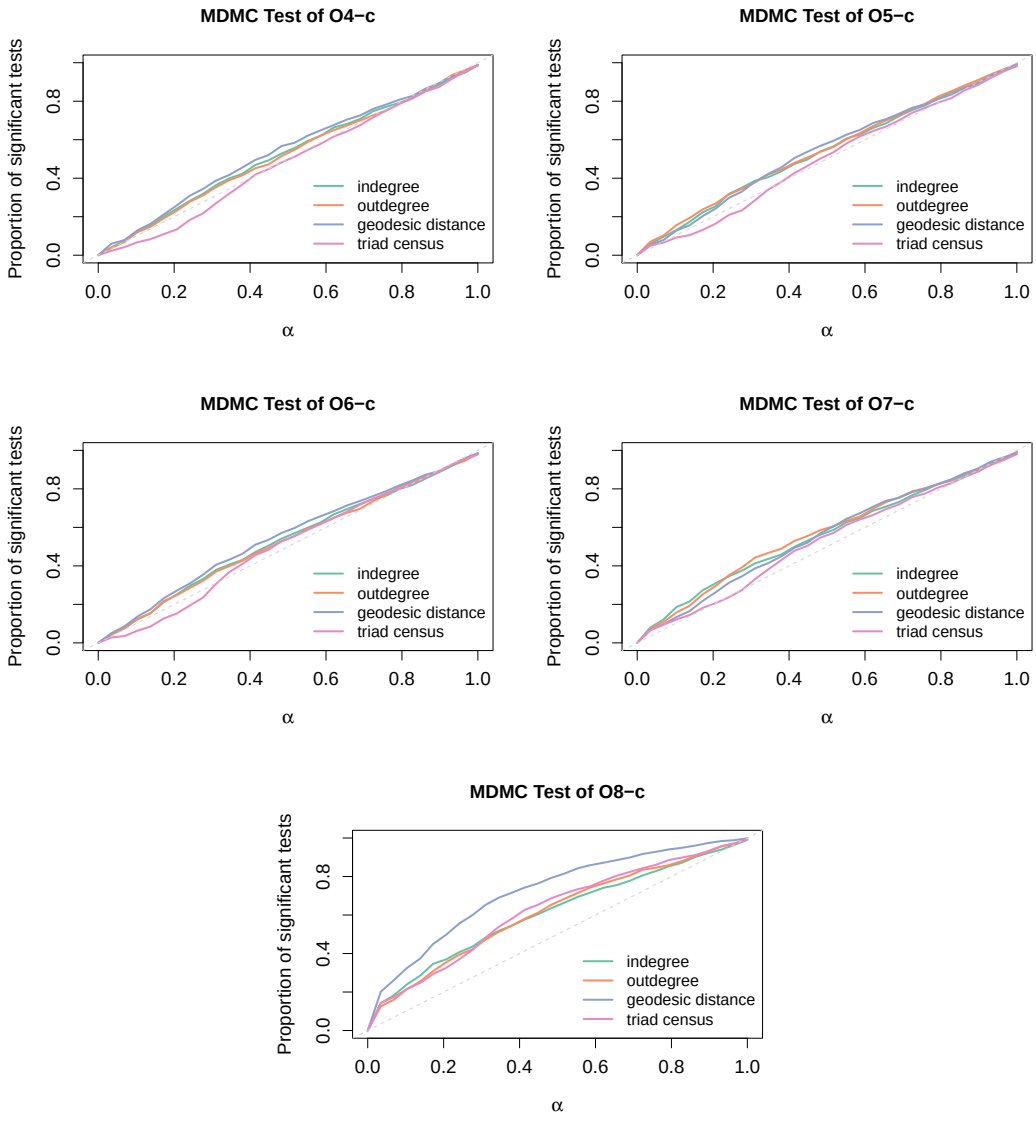


Figure 4.17: MDMC test ROC curves of the misspecifications where one or several attribute effects are omitted from the communication baseline model.

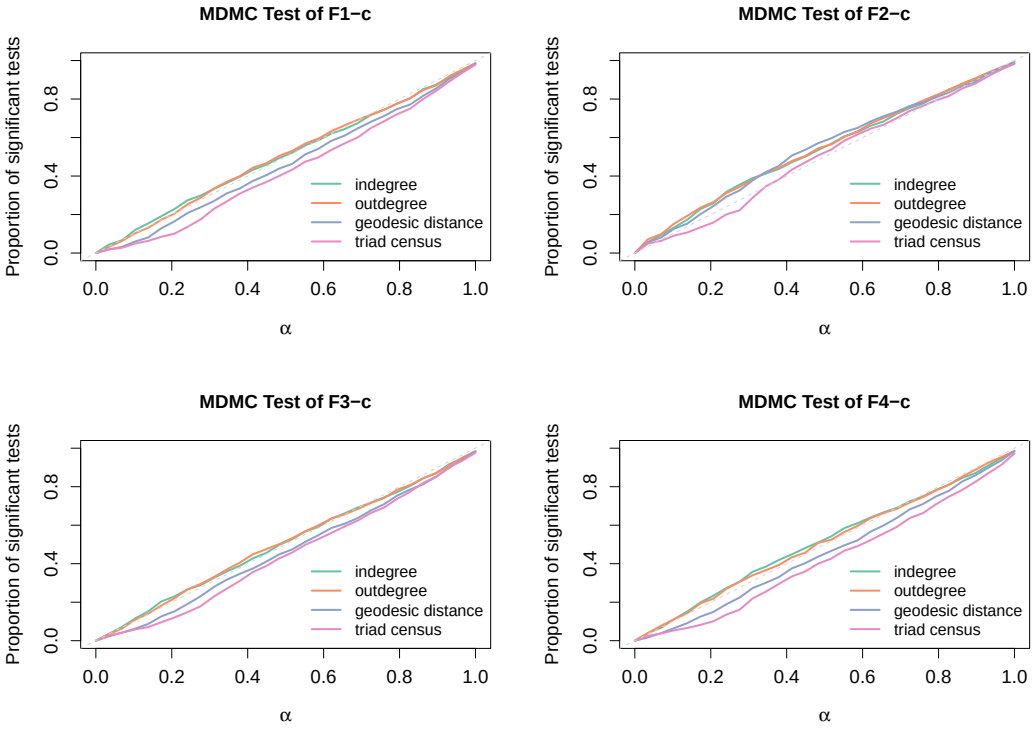


Figure 4.18: MDMC test ROC curves of the misspecifications of type wrong functional form.

Chapter 5

Discussion

In this chapter, we interpret the results reported in Chapter 4. We evaluate the consequences of the different missepcifications and place particular emphasis on the results referring to our research questions from Section 3.4. In parentheses, we refer to the corresponding figures from Chapter 4 which support our statements.

The results from fitting the baseline models to the simulated networks are satisfactory. The power to detect the effects of interest is very high for all effects, except for some of the effects in the friendship baseline model with a small true parameter (Figure 4.3). Recall that the two effects of the friendship baseline model with the lowest power were also the ones that were not significant when fitting the model to the originally observed friendship network. The Type I error of all effects is very small, as we expect (Figure 4.4). The received parameter estimates are therefore unbiased. Also the results of the MDMC test applied to the communication baseline model is encouraging, as the MDMC test ROC curves are close to the diagonal (Figure 4.5, left). However, this is not the case when applying the MDMC test to the friendship baseline model (Figure 4.5, right). We will comment on this in the last paragraph of this chapter.

The most problematic misspecification are the ones where one or several gwesp effects are omitted. This has severe consequences for the parameter estimates of the other network effects, especially for the parameter estimates of other gwesp effects which are significantly influencing the network. We want to highlight that the parameter estimates for the cyclic gwesp effect are estimated positive and significantly different from zero when fitting the misspecification **O1-c** (omitting transitive gwesp), while the true parameter is negative (Figure 4.7). This results in a wrong interpretation of the closure process in the network. It is therefore important to think carefully about which gwesp effects might be influential for the observed network and fit models where all important gwesp effects are included. Obtaining parameter estimates which have a sign that is not inline with the theory about the social process can be an indication of an omitted gwesp effect.
For some of the misspecifications where we omit a gwesp effect, we also have quite a large Type I error for the reciprocity effect (Figure 4.8 and Figure 4.9, right). Note, however, that the reciprocity parameter estimates are positive, as is the true parameter. When interpreting the parameter estimates, we are often mainly concerned about whether a significant effect is positive or negative. Thus, slightly over- or underestimating the reciprocity effects does not have a large impact on the interpretation.
Recall that the transitive and cyclic gwesp statistics, as well as the reciprocity statistics and the (transitive or cyclic) gwesp statistics, are very highly correlated in the simulated networks. This

is why omitting one of the effects has a large influence on the parameter estimates of the other effects.

The results regarding the attribute effects when omitting one or several gweps effects are less severe than with respect to other network effects. The power to detect the important effects is almost perfectly the same as for the baseline models (Figure A.5). However, there is some Type I error for some of the attribute effects (Figure 4.10). The homophily effects are overestimated, especially for the misspecification where we omit both gwesp effects. This is to be expected as the homophily and gwesp effects are highly correlated. Recall that this is the case because clustering among nodes with the same attribute increases the gwesp statistic as well as the homophily statistic. When fitting the misspecified model where both gwesp effects are omitted, the model wrongly assigns the gwesp effect to the homophily effect which results in parameter estimates for the homophily effect which are too large (in absolute value). Recall that in the simulated friendship networks the correlation between the gwesp and the homophily *gender* statistics is larger than between the gwesp and the homophily *delinquency* statistics (Table 4.2). This results in higher Type I errors for the homophily *gender* effect as for the homophily *delinquency* effect, when the gwesp effect is omitted. Note, however, that the parameter estimates of the homophily and the dyadic attribute effects have the same sign as the true parameter for all misspecifications (Figure A.6).

Omitting attribute effects has much less severe consequences than omitting network effects. The parameter estimates of the network effects are more or less unchanged (Figure 4.12 and A.7), except for a small overestimation of the reciprocity effect for the misspecification where the homophily *gender* effect is omitted (Figure 4.11). This small overestimation can be explained by the high correlation between the reciprocity and the homophily *gender* statistics. Reciprocated ties might occur between pupils of the same gender. When fitting the misspecified model where the homophily *gender* effect is omitted, the model wrongly assigns the homophily effect to the reciprocity effect, which is overestimated as a result.
The consequences for the parameter estimates of the remaining attribute effects are rather small as well (Figure A.8 and 4.12). In some of the cases, there is some increased Type I error for the remaining homophily effect when the other homophily effect is omitted. For that, it does not make any difference, whether just the homophily effect of some attribute variable is omitted or whether all effects of that attribute variable are omitted. The consequences for the Type I error are the same.
Furthermore, we can state that omitting the dyadic attribute effect is not problematic. In our simulation study, it did not affect the parameter estimates of the remaining effects.
We would like to recall again that while the consequences for power and Type I error are not very severe when omitting attribute effects, there were many problems when fitting these misspecifications, especially when all attribute effects were omitted at once and especially for the friendship networks. For many of the simulated networks, fitting the misspecification did not work due to degeneracy. This can therefore also be a consequence of misspecification. In situations where fitting the model does not work, it is therefore crucial not to omit a significant gwesp effect (which would probably solve the problem of degeneracy but possibly results in wrong inference), but rather to search for omitted attribute effects.

The least consequences for the parameter estimates may be expected from specifying the wrong functional form for a homophily effect (Figure A.9). We only observed some problems with

estimating the parameter of homophily *seniority*, when the homophily *projects* effect was misspecified using the difference or smalldiff homophily effect instead of the absdiff homophily effect (Figure 4.14). Some of the parameter estimates of *seniority* were significantly different from the true parameter, i.e. the Type I error ROC curve was a bit above the diagonal. This is to expect, as the difference and smalldiff homophily effects describe quite a different homophily processes than the absdiff homophily effect. The difference homophily effect takes the direction of the ties into account, while the absdiff homophily effect does not, and the smaldiff homophily effect only considers differences that are above the specified threshold. This is also reflected in the correlation between the statistics of the different homophily effects. While the correlation between the absdiff and the squared homophily effects is very large (0.94), it is smaller between the absdiff and the smalldiff effect (0.77) and very small between the absdiff and the difference effect (0.06). Hence, it makes sense that for the latter, some parameter estimates of the other homophily effect (homophily *seniority*) are significantly different from the true parameter and that for the difference homophily effect much fewer parameter estimates are significantly different from zero than for the absdiff homophily effect (Figure 4.15, top left). However note, that all the parameter estimates of all the other effects remain more or less unchanged and that the Type I error ROC curve of the homophily *seniority* is only a bit above the diagonal. One problem that one could encounter when interpreting a model fit where a wrong functional form for a homophily effect is used, is that the misspecified homophily effect is not significant while the correct homophily effect would be. The proportion of parameter estimates significantly different from zero for the differential homophily effect of *delinquency* for example is smaller than the ones for the uniform homophily effect of *delinquency* (Figure 4.15). We therefore might wrongly conclude that there is no homophily effect, while in truth there is one.

We would like to emphasize that the observed consequences for power and Type I error depend on the true underlying models. Including further effects in the baseline models or using different types of nodal and dyadic attribute variables might further change the consequences of misspecifications. However, since the baseline models originate from data observed in reality and since we included the most frequently used effects, the results may be generalized. Especially the fact that omitting gwesp effects may result in wrong inference should be taken seriously.

We conclude this section with the interpretation of the MDMC goodness of fit tests. Applying the MDMC tests to the misspecifications fitted to the communication networks is satisfactory (Figure 4.16). Especially the triad census is a good measure to detect the most severe misspecification: omitted gwesp effects. It makes sense that it is exactly the triad census that is a good measure to detect these misspecifications, as the triad census which counts the different types of triangles is highly correlated with the gwesp statistic that is also very related to the triangle counts. It also an interesting finding that the MDMC does not detect the misspecifications where one or several attribute effects are omitted, except for misspecification **O8-c** (omitting all attribute effects), as this is the misspecification with the most severe consequences, and as for the other misspecifications the consequences are not problematic. One concern are the findings of the MDMC test ROC curves received from the friendship baseline model (**B2**) fitted to the friendship networks (Figure 4.5, right). The ROC curves are too far away from the diagonal as we would expect. Reasons for this might be that the test is sensitive to outliers, that we use too few auxiliary statistics, or that the simulated networks used for the MDMC test are not a good sample from the distributions determined by the estimated parameters. One could think about increasing the number of simulated networks as well as the thinning.

Chapter 6

Conclusion

In this thesis, we gave an introduction to ERGMs, paying special attention to the model specification; we explained the benefits of knowing about the consequences of misspecification in ERGM; we presented the methods used to assess the model misspecifications; and we reported and interpreted the results. At the core of this thesis lies a simulation study for which we chose sets of statistics and corresponding parameters which specify the two baseline models. These parameters were obtained by fitting ERGMs to real data. From the baseline models, we simulated a large set of new networks. We then postulated different misspecified models, i.e., models in which either variables were omitted or the wrong functional form for the homophily effect was used. Next, the baseline models, as well as the misspecified models, were fitted to the simulated networks. We assessed the power of the various (mis)specifications by computing the percentage of parameter estimates which were significantly different from zero for different significance levels α. We assembled the resulting proportions into receiver operating characteristic (ROC) curves displaying the computed proportions as a function of the significance level α. We also assessed the Type I error of the different (mis)specifications by computing the percentage of parameter estimates that were significantly different from the true parameter for different significance levels α and displayed the results using ROC curves.

In conclusion, we provided evidence that ERGMs are good models to determine the impact of network and attribute effects in social networks. When fitting the correct models to the simulated networks, the models were able to retrieve the true parameters. Considering the misspecifications, we can state that omitting important network effects from the model may have severe consequences, especially on the parameter estimates of the other network effects. These may lead to wrong interpretation of the network closure effect and an overestimation of the reciprocity effect. The consequences for attribute effects are less severe when omitting important network effects: we observed an overestimation of some of the homophily effects, but in general, the models can detect the important effects with only a small Type I error. Furthermore, we conclude that omitting important attribute effects is less problematic than omitting important network effects. However, a problem that might occur when omitting several important attribute effects is degeneracy. In such a case, it is recommended not to omit a significant gwesp effect, which might lead to wrong inference, but to search for possibly omitted attribute effects.
We can moreover state that specifying the wrong functional form for the homophily effect is in general not problematic for the parameter estimates of the other effects. However, a researcher must be aware that a nonsignificant homophily effect does not necessarily mean that there is no form of homophily effect present in the network. For such a conclusion to be valid, the researcher should consider different functional forms for the homophily effect. We also want to remind the

reader that while it makes sense to pay heed to these recommendations, the above results are limited to the considered case studies.
For the MDMC goodness of fit test, we can conclude that especially the triad census is a good measure to detect misspecifications where one or several gwesp effects are omitted. However, we also found that the MDMC test does not work well for every model. In practice, the MDMC goodness of fit test is not the most common choice. When working with the R-package *statnet*, the built-in function *gof()* is most often used to assess the goodness of fit. It conducts several graphical goodness of fit tests by comparing different auxiliary statistics of the observed network to the auxiliary statistics of networks simulated from the fitted model. If the MDMC goodness of fit test is further improved, by increasing the number of simulated networks used as well as the thinning, it might become a useful alternative to the goodness of fit implemented in the R-package *statnet*.

Up to now, no formal study on the consequences of model misspecification has been conducted. The only consequence that has been investigated thoroughly is the issue of degeneracy. This thesis is a first contribution to investigating the inferential consequences of model misspecification of type omitted variable and wrong functional form. Its results are however limited in several respects. In our simulation study, we have only considered networks of sizes 38 and 25. Further research should include networks of larger sizes. Furthermore, we have included sender and receiver effects, as they are often included together with the homophily effect. However, we declared the homophily effects as our attribute effects of interest and did not consider the consequences of the misspecifications on sender and receiver effects. It might be worthwhile to examine these consequences as well as the consequences for homophily effects when sender and receiver effects are omitted. Moreover, one could examine models which also include geometrically weighted degree or geometrically weighted edgewise shared partners effects, which are also frequently used in empirical research. Further research might also include a simulation study that assesses the consequences of including an irrelevant variable and that investigates the link between the correlation among the variables and the consequences of this misspecification. In general, a natural next step would be to examine the correlation among the variables and investigate its consequences on the estimates of the standard errors.

Appendix A

Appendix

Table A.1: The final sample sizes (n) of the different (mis)specifications, i.e., the number of estimated sets of parameters used for the analysis. Recall that misspecifications **O4-f**, **O5-f** and **O7-f** are not analyzed.

Specification	**n**	**Specification**	**n**
B1	1171	**O1-f**	1200
B2	1000	**O2-f**	745
O1-c	1198	**O3-f**	983
O2-c	1067	**O6-f**	924
O3-c	1199	**F1-c**	1174
O4-c	1171	**F2-c**	1062
O5-c	1068	**F3-c**	1149
O6-c	1153	**F4-c**	1171
O7-c	1000	**F1-f**	1003
O8-c	818	**F2-f**	1005

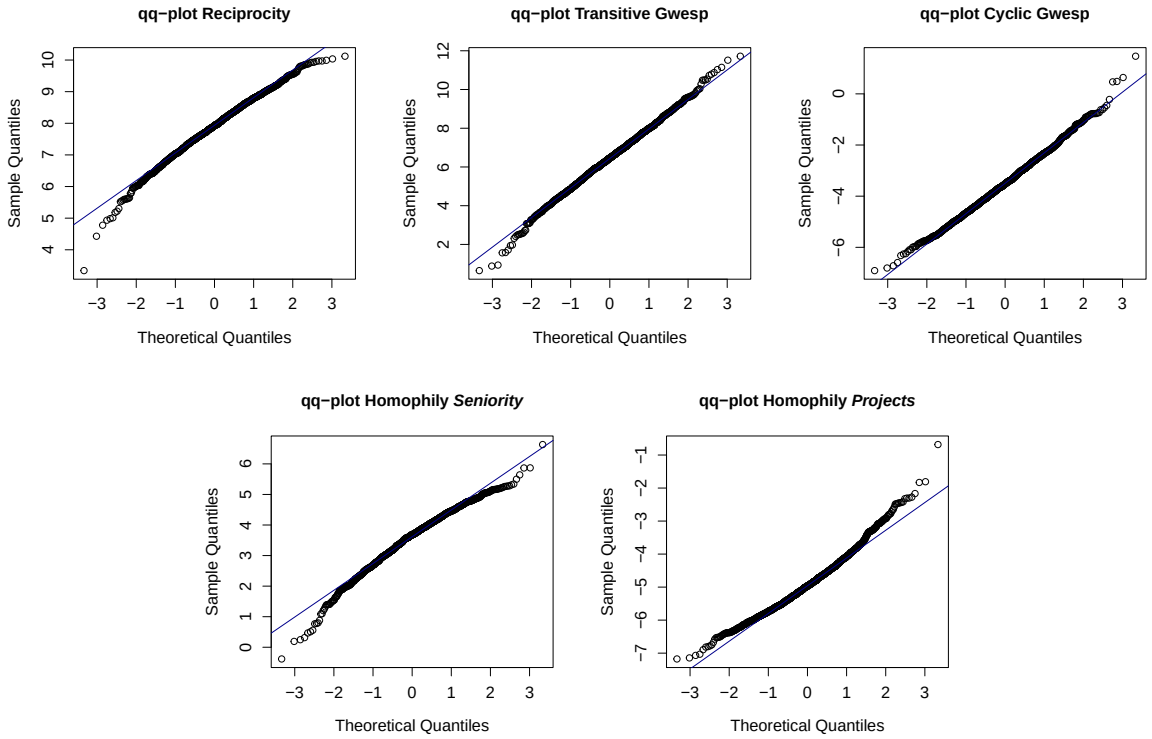


Figure A.1: *qq*-plot of the parameter estimates divided by the corresponding standard errors of the different effects of interest in the communication baseline model.

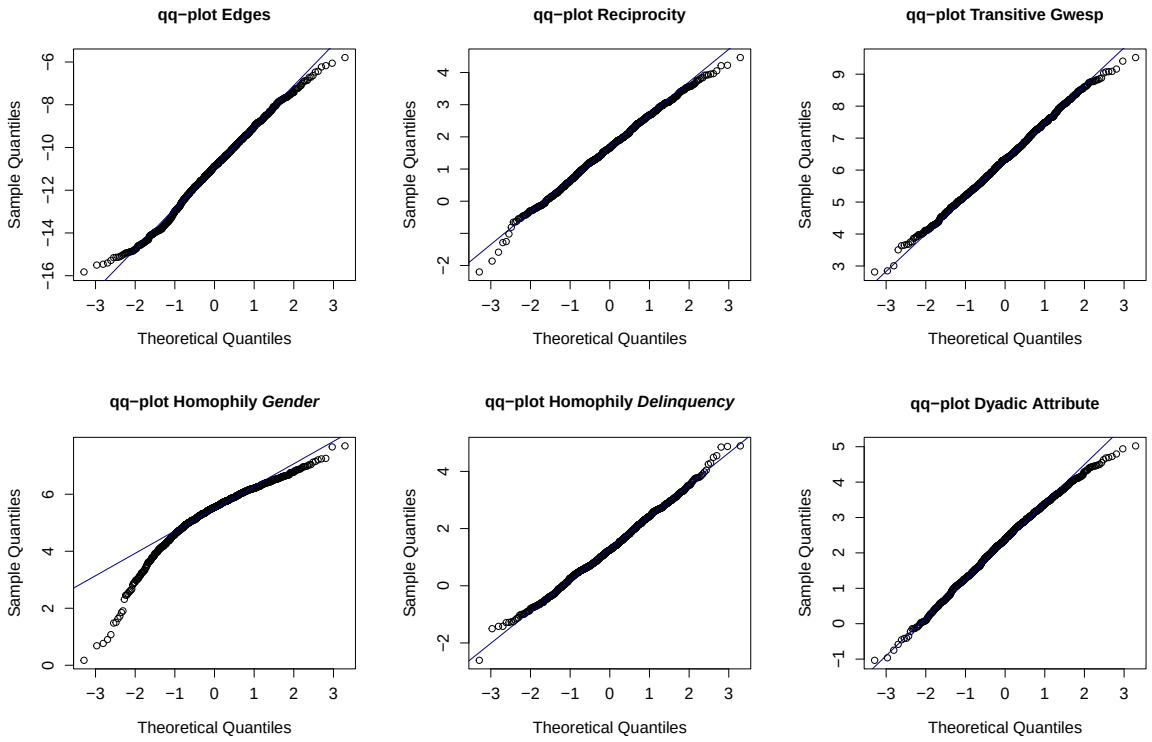


Figure A.2: *qq*-plot of the parameter estimates divided by the corresponding standard errors of the different effects of interest in the friendship baseline model.

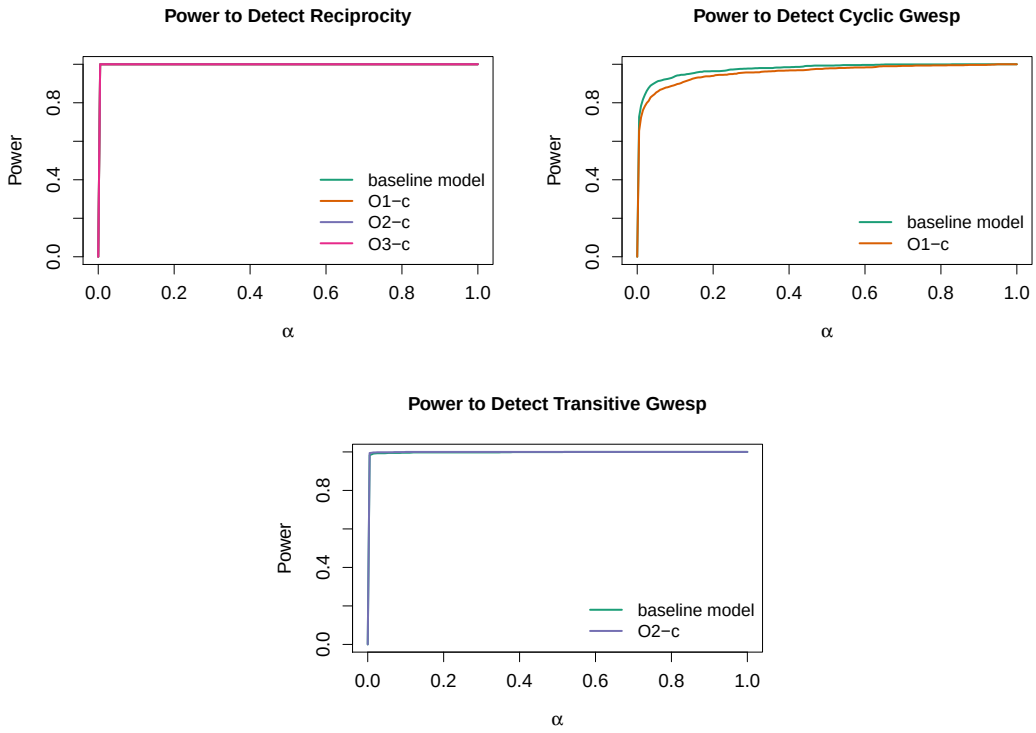


Figure A.3: Power ROC curves of the remaining network effects when omitting a gwesp effect from the communication baseline model. Furthermore, the power ROC curve of the baseline model is included for comparison.

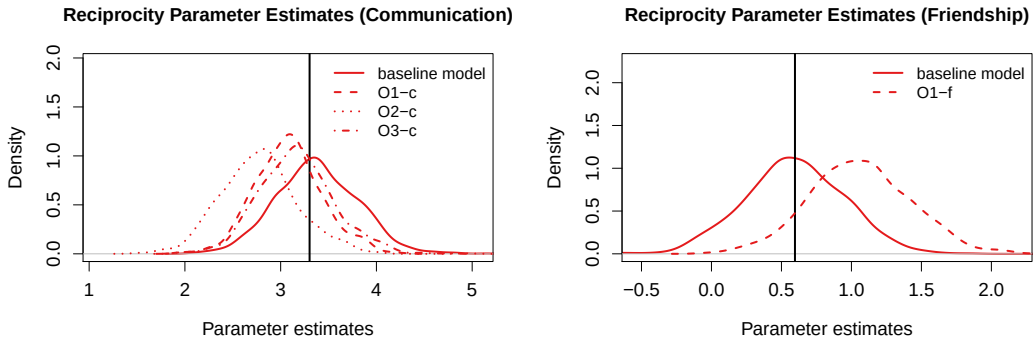


Figure A.4: Densities of the parameter estimates for the reciprocity effect when fitting the baseline model as well as misspecifications to the communication networks (left) and friendship networks (right).

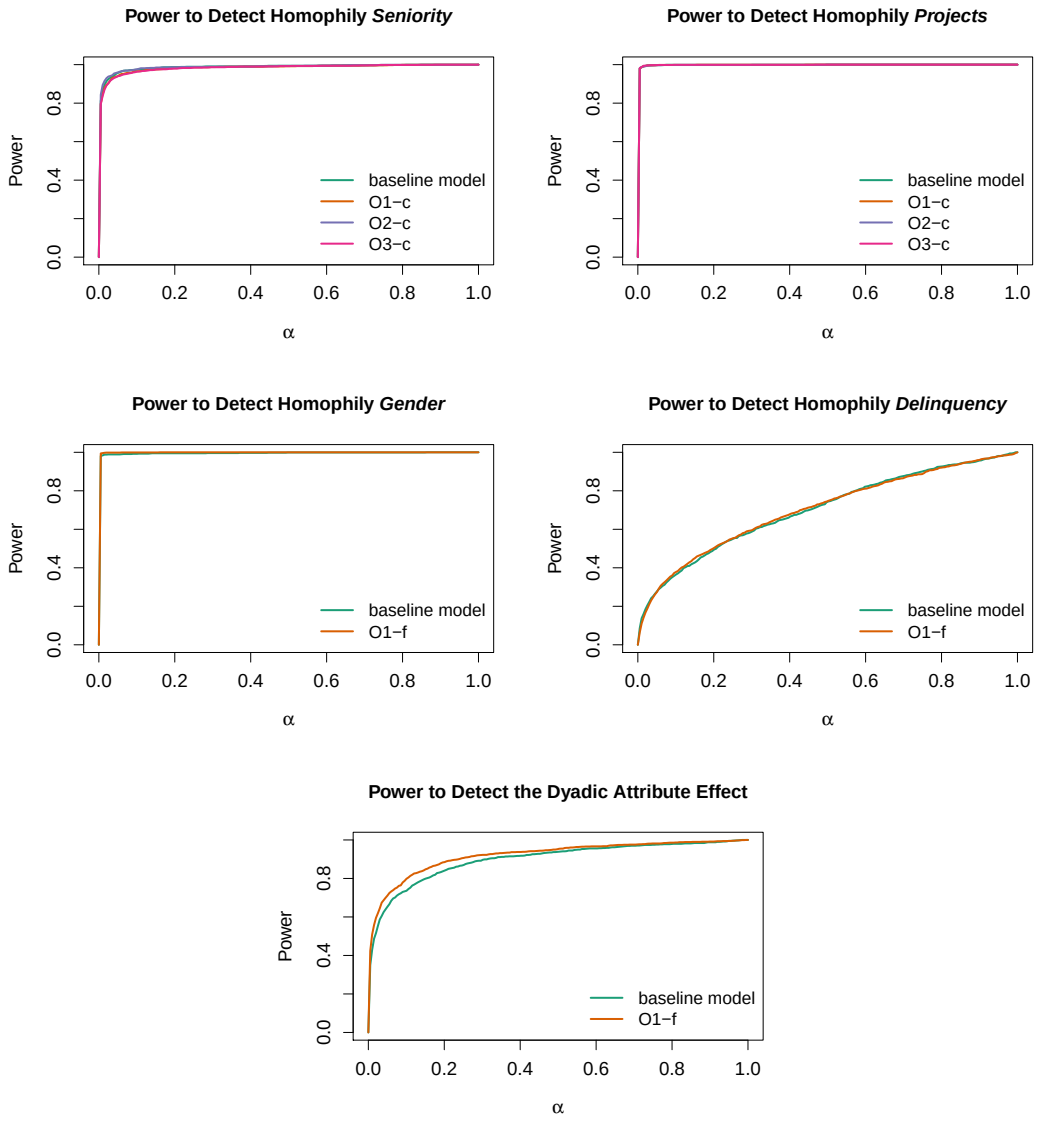


Figure A.5: Power ROC curves of the attribute effects when omitting a gwesp effect from the communication or friendship baseline model. Furthermore, the power ROC curve of the corresponding baseline model is included for comparison.

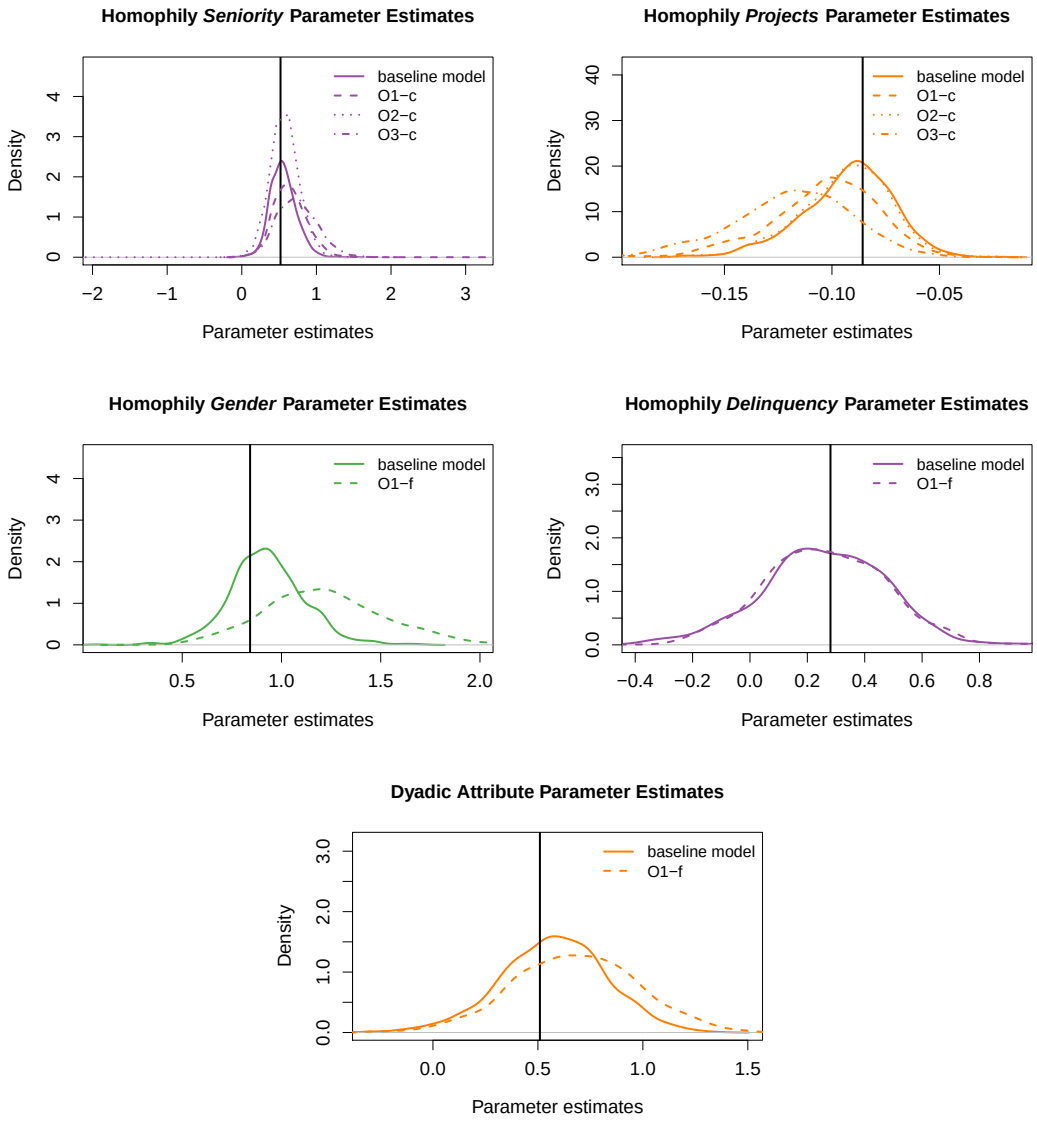


Figure A.6: Densities of the parameter estimates for the attribute effects when omitting gwesp effects from the baseline models.

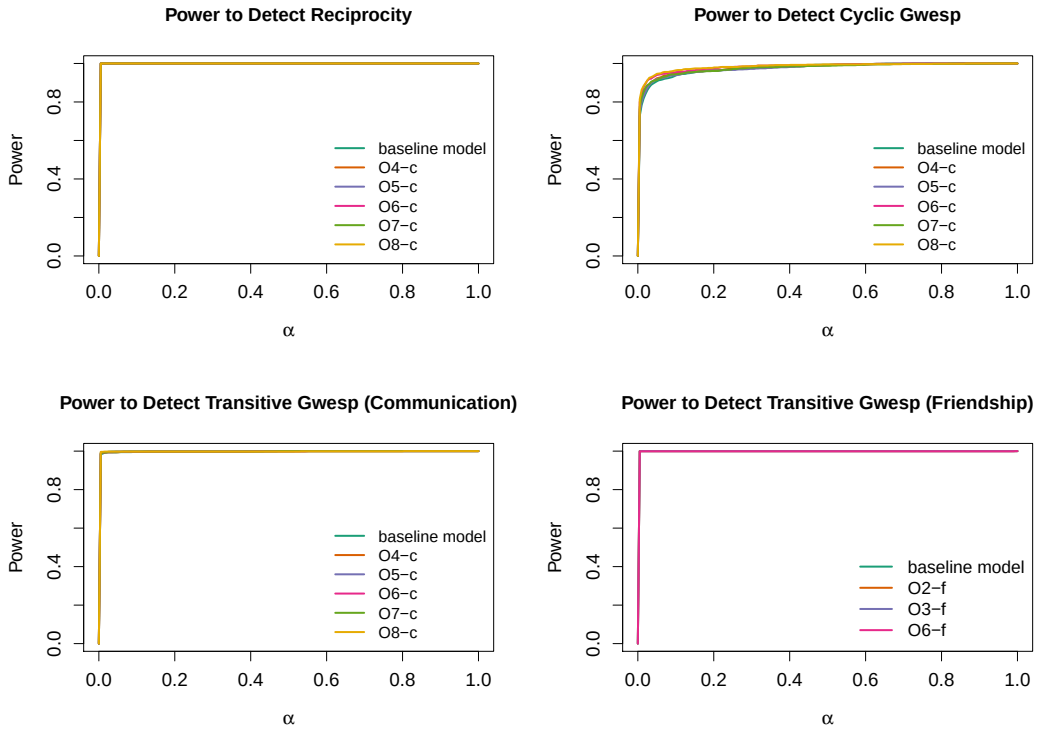


Figure A.7: Power ROC curves of the network effects when omitting attribute effects.

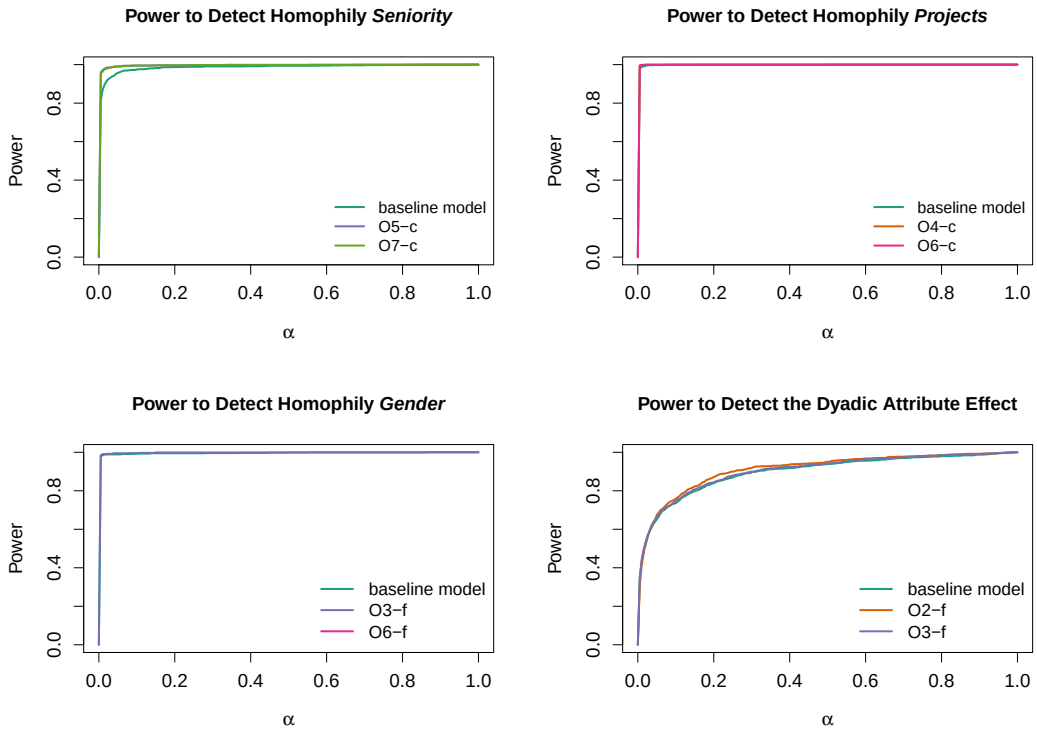


Figure A.8: Power ROC curves of the remaining attribute effects when omitting other attribute effects.

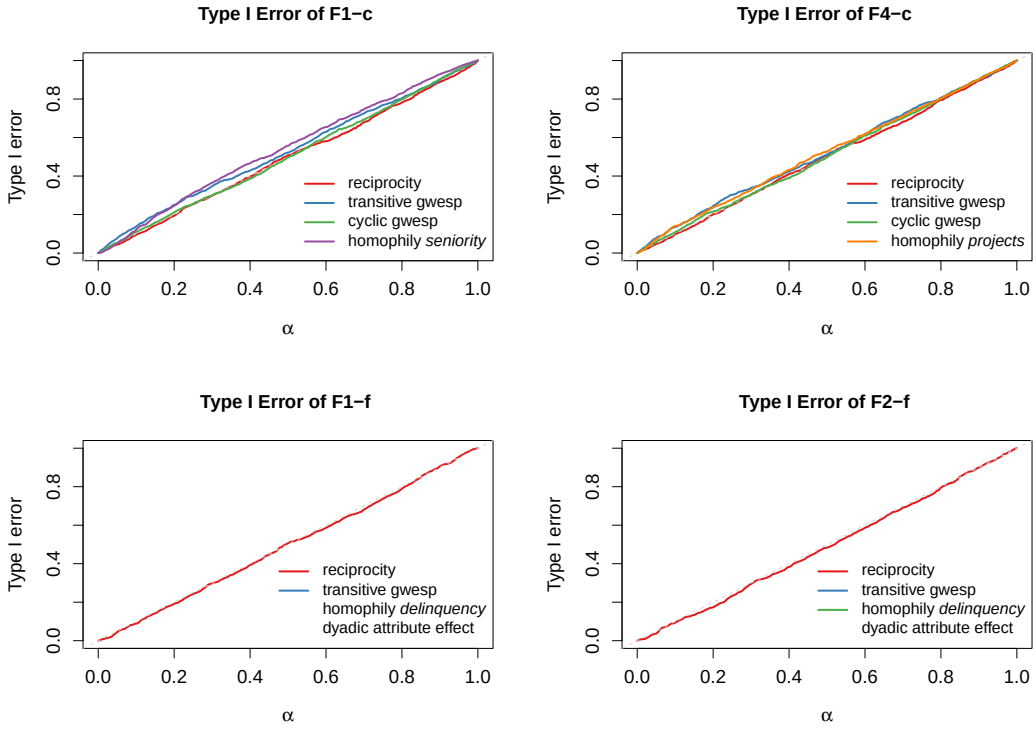


Figure A.9: Type I error ROC curves of the misspecifications **F1-c**, **F4-c**, **F1-f** and **F1-f**.

Bibliography

Amati, V., Mol, A., Shafie, T., Hofman, C., and Brandes, U. (2020). A framework for reconstructing archaeological networks using exponential random graph models. *Journal of Archaeological Method and Theory*, 27. 1

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36 192–225. 6, 9

Cranmer, S. J., Heinrich, T., and Desmarais, B. A. (2014). Reciprocity and the structural determinants of the international sanctions network. *Social Networks*, 36 5–22. 1

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81 832–842. 6, 9

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)*, 54 657–699. 23

Grimmett, G. (1973). A theorem about random fields. *Bulletin of London Mathematical Society*, 5 81–84. 9

Handcock, M. S. (2002). Statistical models for social networks: Inference and degeneracy. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, 229–240. 15, 25

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., and Morris, M. (2018). *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (`http://www.statnet.org`). R package version 3.9.4. 6

Handcock, M. S., Robins, G., Snijders, T., Moody, J., and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. Technical report, Citeseer. 23

Hunter, D. (2007). Curved exponential family models for social networks. *Social Networks*, 29 216–230. 15, 16

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103 248–258. 18, 25, 38

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15 565–583. 18, 23

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 1–29. 6

Lewis-Beck, M., Bryman, A., and Liao, T. (2004). Misspecification. *The SAGE Encyclopedia of Social Science Research Methods*, 650–650. 31

Lospinoso, J. and Snijders, T. (2019). Goodness of fit for stochastic actor-oriented models. *Methodological Innovations*, 12 1–23. 38, 39

Lusher, D., Koskinen, J., and Robins, G., editors (2013). *Exponential Random Graph Models for Social Networks*. Cambridge University Press. v, 6, 11, 13, 18, 20, 23, 31

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 1548–7660. 4, 18

Pattison, P. and Robins, G. (2002). 9. neighborhood-based models for social networks. *Sociological Methodology*, 32 301–337. 7

Robins, G. and Morris, M. (2007). Advances in exponential random graph (p*) models. *Social Networks*, 2 169–172. 15, 18

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29 173–191. v, 6

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3 1–40. 15, 25

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 99–153. 15, 16

van Duijn, M. A., Gile, K. J., and 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 52–62. 25

Wang, P., Robins, G., Pattison, P., and Koskinen, J. (2009). *PNet: program for the simulation and estimation of exponential random graph models*. Melbourne School of Psychological Sciences, The University of Melbourne. 6

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, 61 401–425. v, 6