TECHNISCHE UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR MATHEMATIK
LEHRSTUHL FÜR FINANZMATHEMATIK

Financial Network Models

Assessing Systemic Risk

Janina Engel

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Prof. Dr. Mathias Drton
Prüfer der Dissertation:	1.	Prof. Dr. Matthias Scherer
	2.	Prof. Dr. Wim Schoutens (KU Leuven, Belgien)
	3.	Prof. Dr. Stefano Battiston (Universität Zürich, Schweiz)

Die Dissertation wurde am 02.12.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 24.03.2020 angenommen.

Abstract

Especially since the global financial crisis of 2007–2008, stabilizing and securing our financial system has gained high relevance. To truly identify potential sources of risk and to study the unfolding contagious dynamics, realistic and flexible network models are required. This thesis contributes to these questions by studying two of the most sophisticated network reconstruction techniques, namely ERGMs and fitness models.
A reoccurring problem is the generation of graphs, typically represented by matrices, that satisfy given row and column sums. We derive necessary and sufficient conditions under which the methodology of ERGMs offers a solution to this problem. In addition, the proof to this statement reveals a new and efficient algorithm for parameter calibration. Furthermore, the class of fitness models is usefully extended to cover more flexible degree distributions, thus enabling a more precise reconstruction. Building on these results, a block structured model for the reconstruction of directed and weighted financial networks, spanning multiple countries, is developed. In a first step, the topology of the network is reconstructed via an extended fitness model, allowing a calibration to desired block specific densities and reciprocities. In a second step, weights are allocated via an ERGM, such that desired row and column sums as well as block weights are satisfied. Furthermore, calibrating our model to the EU interbank market, we are able to analyze systemic risk within the EU in detail by applying various prominent contagion models.

Zusammenfassung

Insbesondere seit der globalen Finanzkrise von 2007–2008 hat die Stabilisierung und Sicherung unseres Finanzsystems stark an Relevanz gewonnen. Um mögliche Risikoquellen zu identifizieren und die folgenden kontagiösen Dynamiken zu studieren, werden realistische und flexible Netzwerkmodelle benötigt. Diese Arbeit trägt zur Lösung dieser Probleme bei, indem zwei der fortschrittlichsten Netzwerkrekonstruktionsmethoden, ERGMs und Fitness Modelle, studiert werden.
Ein wiederkehrendes Problem ist die Generierung von Graphen, üblicherweise in Form von Matrizen dargestellt, die gegebene Zeilen- und Spaltensummen erfüllen. Wir leiten notwendige und hinreichende Bedingungen her, unter denen die Methodik der ERGMs eine Lösung für dieses Problem bietet. Zudem zeigt der Beweis dieser Aussage einen neuen und effizienten Algorithmus für die Parameterkalibrierung auf. Des Weiteren wird die Klasse der Fitness Modelle nützlich erweitert, um flexiblere Degree Verteilungen zu erfassen und somit eine akkuratere Rekonstruktion zu ermöglichen. Basierend auf diesen Ergebnissen wird ein blockstrukturiertes Modell für die Rekonstruktion von gerichteten und gewichteten Finanznetzwerken, die mehrere Länder umfassen, entwickelt. In einem ersten Schritt wird die Netzwerktopologie durch ein erweitertes Fitness Modell rekonstruiert, das eine Kalibrierung auf gewünschte blockspezifische Dichten und Reciprocities erlaubt. In einem zweiten Schritt werden Gewichte so allokiert, dass gewünschte Zeilen- und Spaltensummen, sowie Blockgewichte erfüllt werden. Darüber hinaus können wir durch eine Kalibrierung unseres Modells auf das EU Interbankennetzwerk, systemisches Risiko in der EU mittels der Anwendung von prominenten Contagion Modellen im Detail analysieren.

Acknowledgements

First and foremost, I want to thank my supervisor, Matthias Scherer, for his continuous support, his invaluable guidance and his patience. You have honestly been the best Doktorvater that I could have wished for. Without you, this would not have been possible and I wouldn't be where I am today. I am truly and extremely thankful for everything.

Likewise, I am deeply grateful to Francesca Campolongo, Jessica Cariboni, Andrea Pagano, and Wim Schoutens who offered me the unique opportunity to contribute to the highly significant research conducted at the JRC in beautiful Ispra, to provide the necessary scientific advice to the European Commission. Especially, I want to thank Andrea for always taking time for me, his consistent support and encouragement. Furthermore, I am truly happy and thankful for the entire Economics and Finance Unit who made my time at the JRC a unique and cheerful experience that I will never forget.

In particular, I want to thank Nathalie for always being there for me and sharing every aspect of this experience. I am very thankful to Nathalie and Michela N. for our bright and joyous morning coffees, which were the best way to start the day. Many thanks also to Michela R., who always supported me and who still makes me very happy for keeping collaborating on novel and interesting topics.

Also I want to thank the entire Chair of Mathematical Finance at the TUM. Especially Prof. Zagst, who has been my mentor for many years. I am truly grateful to Amelie, Miriam, Gabi, Andreas, and Markus, who I always enjoyed visiting and who helped me whenever I called.

Most of all, I am infinitely grateful for my joyful, funny, and loving family, who I can always trust and build on. This thesis would not have been possible without my one-of-a-kind parents, who have always supported me and fostered my mathematical thinking since my birth. This is equally your achievement, I dedicate this thesis to both of you.

Contents

1 Introduction

The global financial crisis of 2007–2008 highlighted the necessity for a better understanding of our financial markets and an accurate assessment of systemic risk. In 2010, the Basel Committee on Banking Supervision (2010) identified the interconnectedness of financial institutions as a significant source of systemic risk and an important cause for a further amplification of the crisis. Moreover, at the time of the crisis, authorities' stress tests failed to adequately model interlinkages in the banking sector, which in turn led to a dramatic underestimation of the vulnerability of financial systems, see Basel Committee on Banking Supervision (2015a).

Since then, the literature on systemic risk measuring has gained great attention and several sophisticated methodologies modeling the propagation of losses through financial networks have been developed. Prominent examples include Rogers and Veraart (2013), who generalized the clearing mechanism of Eisenberg and Noe (2001) by introducing default costs. The resulting payment vector determines the losses that each institution needs to cover. A different approach is taken by Battiston et al. (2012, 2016), who propose a measure of systemic impact based on the idea of feedback-centrality, the so-called *DebtRank*. Further, Cont et al. (2013) developed the *Contagion Index*, a simulation-based approach that quantifies the expected loss in capital generated by an institution's default given a common adverse shock scenario and in the light of a recovery rate of zero.

For an extensive survey on systemic risk models, see De Bandt and Hartmann (2000) and Hüser (2015). While this strand of literature keeps growing, the problem of constructing realistic models of financial networks remains open. It constitutes a challenging task, because information on bilateral interbank-activities is classified confidential and thus mostly not available. A detailed discussion on the sparsity of consistent bank-level data is provided by Cerutti et al. (2011). The relevance of this issue has also been acknowledged by authorities, who in response launched several initiatives to fill essential data gaps, see, e.g., the G20 Data Gaps Initiative (DGI)[1] and the EU-wide transparency exercise by the European Banking Authority (EBA)[2]. Nonetheless, until today only few data on aggregated levels have been published.

Because of the lack of publicly available information on interbank lending, academics often turn to random graphs or toy networks to apply their developed tools. This allows

[1]For more information, see, e.g., http://www.imf.org/external/np/seminars/eng/dgi/.
[2]For more information, see
http://www.eba.europa.eu/risk-analysis-and-data/eu-wide-transparency-exercise.

to derive theoretical results on systemic risk for very specific and well controlled network structures. For a better understanding of the complex topology of actual financial markets and the underlying mechanisms of shock propagation, however, more realistic network models are needed. This especially concerns policy-makers, who need reliable models to derive adequate and effective regulations. For this reason, this research question has also been raised by the Joint Research Centre (JRC) of the European Commission. The JRC is the European Commission's science and knowledge service and provides independent scientific evidence regarding EU policies throughout the whole policy cycle. In collaboration with the JRC and to support the aim to assess systemic risk within the EU interbank network, this thesis develops tractable network models, that (a) span across multiple countries, (b) allow for fast simulation of sample scenarios, (c) can be calibrated to scarce available information, and thus generate realistic financial markets, and (d) offer the flexibility to easily change particular network characteristics for a detailed analysis.

The thesis and its main contributions are organized as follows.

Chapter 2 introduces the mathematical background on network reconstruction techniques. The most prominent methodologies are exponential random graph models (ERGMs) and fitness models. Both approaches are explained in detail. To complete the picture on network reconstruction an overview over further proposed approaches is provided as well.

A general open problem in the realm of ERGMs is the existence and uniqueness of solutions, which depend strongly on the considered network characteristics that are incorporated in the form of constraints. We contribute to this research question in Chapter 3 by analyzing the class of ERGMs that satisfy given row and column sums. We derive necessary and sufficient conditions under which a solution exists and show that the solution is unique up to certain equivalence classes. The proof is furthermore of special interest as it directly leads to an efficient algorithm for parameter calibration.

The methodology of fitness models is extended in Chapter 4. First, a more generalized version of the fitness model with randomized variables is developed by adjusting it to a more flexible degree distribution. Second, we provide mathematical insight into economic fitness models, which have become very popular as they can easily be applied in light of scarce information. However, the underlying structure and reasons for which these models work well, have not yet been investigated.

Chapter 5 extends the findings in Engel et al. (2019b), where the leading author is also the author of this thesis. Based on the technique of ERGMs, Chapter 5 discusses the reconstruction of domestic interbank markets. More precisely, an ERGM conditioned on the in- and out-degree sequence plus the network reciprocity is considered. The problem of missing available data to estimate the model parameters is tackled by introducing flexible in- and out-degree distributions that are coupled via a Gaussian copula and which can be calibrated in light of scarce information. The performance of the model is demonstrated by reconstructing the German and the Italian interbank market. As many network statistics are closely reproduced, they seem to be a natural consequence of the specific in- and out-degree distribution and the reciprocity.

The following Chapters 6 and 7 are based on our working paper Engel et al. (2019a), where the leading author is also the author of this thesis.

Chapter 6 presents a novel model for the reconstruction of international financial networks, using a block structure of weighted networks and extending the model of Chapter 5. More precisely, in a first step we use an extended fitness model to reconstruct the adjacency matrix of the underlying financial network. The adjacency matrix can be calibrated to desired block-specific densities and reciprocities. This results in a link-probability matrix, which allows to efficiently sample adjacency matrices through bivariate Bernoulli trials. In a second step, the sampled adjacency matrices are weighted, such that interbank assets and liabilities, which are known from the banks' balance sheets, as well as the total weight circulating within and across countries, is met. This is achieved via an ERGM conditioned on the row and column sums as well as on the block weights. Since this model allows to analytically derive the expected weight of each link of a given adjacency matrix, the conditions are fulfilled exactly by the resulting network.

Another contribution of this thesis is the algorithm that was developed for calibrating the parameters of the ERGM presented in Chapter 6. The parameters are determined by a high dimensional non-linear system of equations which is non-trivial to solve. Regarding the reconstruction of the EU interbank market, for example, there are 7,779 parameters, i.e. 7,779 depended equations to be solved. We tackle this problem by exploiting the structure of the system of equations and some findings of Chapter 3.

Last but not least, calibrating the model of Chapter 6 to data on the EU interbank market, we are able to generate realistic networks. The simulated graphs finally enable a detailed assessment of systemic risk. Chapter 7 demonstrates the potential of the model by conducting several analyses on contagion in the EU interbank market. First, a general overview and comparison of the most prominent contagion models is given. Second, the correlation of specific node characteristics and systemic risk as well as the node's vulnerability are analyzed. Third, as global systemically important banks are of special interest, their results on contagion are discussed in more detail. Forth, the influence of the network density, an indicator of risk diversification, on the stability of the network is analyzed and compared to other findings in the literature. The conducted analyses are of special interest as they can pave the way for further improvements on contagion models and systemic risk measures, as well as support the ultimate aim of policy-makers to stabilize financial markets.

Finally, Chapter 8 concludes and points out possible future research questions.

2 Mathematical Preliminaries

This chapter introduces the mathematical concepts this thesis builds upon. In particular, the methodologies of exponential random graph models (Section 2.1) and fitness models (Section 2.2) are thoroughly explained. Novel instances of both model classes form the core of the model for international interbank networks developed in Chapter 6. To complete the picture on network reconstruction techniques, an overview over further well-known approaches is provided in Section 2.3.

2.1 Exponential Random Graph Models

The literature on ERGMs (also called configuration models) traces back to Holland and Leinhardt (1981), who proposed an exponential probability distribution for random networks. Later, Park and Newman (2004) showed that ERGMs in fact constitute the solution to constrained maximum entropy problems and linked them to the field of statistical mechanics in modern physics. The fundamental idea is the construction of a probability distribution over a set of possible graphs, with minimum divergence to the uniform distribution and subject to satisfying desired network statistics in expectation.

To properly introduce the methodology of ERGMs, we start by recalling the fundamentals of general constrained maximum entropy problems. Afterwards, we discuss the adaptation to complex and weighted networks. Furthermore, Appendix B provides an overview over well-known classes of ERMGs.

2.1.1 Constrained Maximum Entropy Problems

Imagine, we are asked to assign a (discrete) probability distribution to a random game with a set Ω of possible outcomes, without any further information. If we have no reason to believe that any outcome $\omega_i \in \Omega$ is more likely to occur than any other outcome $\omega_j \in \Omega$, then we would intuitively, and by the *principle of indifference* by Bernoulli and Laplace, assign an equal probability of $1/|\Omega|$ to each $\omega_i \in \Omega$. This concept is extended by Jaynes' *principle of maximum entropy* (Jaynes (1957a,b)), which states that in light of partial information, the most unbiased distribution is the one which satisfies the given information and is otherwise as close as possible to the uniform distribution.

To make the statement "as close as possible" mathematically precise, we need a measure of divergence. A common measure of similarity between two probability distributions is given by the Kullback–Leibler divergence (Kullback and Leibler (1951); Kullback (1997)), also known as the relative entropy.

Definition 2.1.1 (Kullback–Leibler Divergence)
For two discrete probability distributions with probability mass functions P and Q, defined on the same sample space Ω, the Kullback–Leibler divergence $D_{KL}(P\|Q)$ is defined as

$$D_{KL}(P\|Q) := \mathbb{E}_{X\sim P(X)}\left[\log\left(\frac{P(X)}{Q(X)}\right)\right] = \sum_{\omega\in\Omega} P(\omega)\log\left(\frac{P(\omega)}{Q(\omega)}\right), \tag{2.1}$$

with the continuous extensions $0\log\left(\frac{0}{0}\right) := 0$, $0\log\left(\frac{0}{Q}\right) := 0$, and $P\log\left(\frac{P}{0}\right) := \infty$.

For continuous probability distributions, the Kullback–Leibler divergence is defined analogously. Moreover, the Kullback–Leibler divergence fulfills the following properties.

Lemma 2.1.2 (Properties of the Kullback–Leibler Divergence)
The Kullback–Leibler divergence, defined in Definition 2.1.1, has the following properties:

(i) *non-negativity, i.e.* $D_{KL}(Q\|P) \geq 0$,

(ii) *the minimum* $D_{KL}(Q\|P) = 0$ *is attained if and only if* $Q = P$ *almost everywhere,*

(iii) *asymmetry, i.e.* $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$,

(iv) *convexity, i.e. for probability distributions* P_1, P_2, Q_1, Q_2 *and for* $\lambda \in [0,1]$ *it holds* $D_{KL}(\lambda P_1 + (1-\lambda)P_2\|\lambda Q_1 + (1-\lambda)Q_2) \leq \lambda D_{KL}(P_1\|Q_1) + (1-\lambda)D_{KL}(P_2\|Q_2)$.

Proof
See, e.g., Cover and Thomas (2006). □

It is well-known and straightforward to show, that the probability distribution which minimizes the Kullback–Leibler divergence with respect to the uniform distribution also maximizes the Shannon entropy $S(P) := \sum_{\omega\in\Omega} P(\omega)\log\big(P(\omega)\big)$ (see, e.g., Abbas et al. (2017)). Let the discrete uniform distribution defined on Ω with cardinality $|\Omega|$ be denoted by Q and let $\mathcal{P}$ be the set of all probability mass functions defined on Ω,

$$\begin{aligned}\arg\min_{P\in\mathcal{P}} D_{KL}\left(P\|Q\right) &= \arg\min_{P\in\mathcal{P}} \sum_{\omega\in\Omega} P\left(\omega\right)\log\left(\frac{P\left(\omega\right)}{1/|\Omega|}\right)\\ &= \arg\min_{P\in\mathcal{P}} \log\left(|\Omega|\right)\underbrace{\sum_{\omega\in\Omega} P\left(\omega\right)}_{=1} + \sum_{\omega\in\Omega} P\left(\omega\right)\log\big(P\left(\omega\right)\big)\\ &= \arg\min_{P\in\mathcal{P}} \sum_{\omega\in\Omega} P\left(\omega\right)\log\big(P\left(\omega\right)\big)\\ &= \arg\max_{P\in\mathcal{P}} \underbrace{-\sum_{\omega\in\Omega} P\left(\omega\right)\log\big(P\left(\omega\right)\big)}_{\equiv S(P)}.\end{aligned} \tag{2.2}$$

Maximum entropy problems arise in light of partial available information, when we want to define a probability distribution over a set of possible outcomes, such that certain known characteristics are met in expectation. In most cases, there will exist many probability distributions fulfilling the expected desired characteristics. Therefore, according to the principle of maximum entropy, among these probability distributions, we choose the one that, in addition, is closest to the uniform distribution. For the discrete case, let:

- $\mathcal{P} = \{P : \Omega \to (0,1)\}$ denotes the set of all discrete probability laws defined on Ω,
- the functions $f_i : \Omega \to \mathbb{R}$ describe desired characteristics, for $i \in \{1,\ldots,m\}$, $m \in \mathbb{N}$,
- $\mu_i \in \mathbb{R}$, for $i \in \{1,\ldots,m\}$, denote the given expected values corresponding to the characteristics f_i, (i.e. the partial information).

The ME distribution is formally defined as the solution to the following constrained optimization problem, which minimizes the divergence to the uniform distribution

$$\max_{P\in\mathcal{P}} S\left(P\right) = \max_{P\in\mathcal{P}} -\sum_{\omega\in\Omega} P\left(\omega\right)\log\big(P\left(\omega\right)\big), \tag{2.3}$$

subject to the normalization condition

$$\sum_{\omega\in\Omega} P(\omega) = 1, \tag{2.4}$$

and which satisfies the desired characteristics in expectation,

$$\sum_{\omega\in\Omega} P(\omega) f_i\left(\omega\right) = \mu_i, \qquad \forall i \in \{1,\ldots,m\}. \tag{2.5}$$

Since the Shannon entropy is a strictly concave function, the set $\mathcal{P}$ containing all discrete probability laws defined on Ω is convex, and the constraints are linear functions in P, we know from convex optimization theory that every local solution constitutes a global solution and in case a solution exists it is unique (Boyd and Vandenberghe, 2004, pp. 137–138). Moreover, because the Shannon entropy and the functions of the constraints are continuously differentiable w.r.t. P, it suffices to solve the Karush–Kuhn–Tucker (KKT) conditions (Boyd and Vandenberghe, 2004, p. 244). Hence, let $\mathcal{L} : (0,1)^{|\Omega|} \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ denote the Lagrangian,

$$\mathcal{L}\left(P, \alpha, \lambda\right) = S\left(P\right) + \alpha \left[1 - \sum_{\omega \in \Omega} P(\omega)\right] + \sum_{i=1}^{m} \lambda_i \left[\mu_i - \sum_{\omega \in \Omega} P(\omega) f_i\left(\omega\right)\right]. \tag{2.6}$$

From the KKT conditions it follows, if P^*, λ^*, and α^* satisfy

$$\begin{aligned} &\text{(i)} && \sum_{\omega \in \Omega} P^*(\omega) = 1, \\ &\text{(ii)} && \sum_{\omega \in \Omega} P^*(\omega) f_i\left(\omega\right) = \mu_i, \quad \text{for all } i \in \{1, \ldots, m\}, \\ &\text{(iii)} && \nabla \mathcal{L}\left(P^*, \alpha^*, \lambda^*\right) = 0, \end{aligned} \tag{2.7}$$

then P^* constitutes a solution to the optimization problem defined in Eqs. (2.3) to (2.5), i.e. the unique solution.

Computing the gradient of the Lagrangian and setting it equal to zero (i.e. KKT condition (iii)), we obtain the particular form that the solving probability distribution takes. Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be finite, then for all $\omega_k \in \Omega$, we have,

$$\begin{aligned} 0 = \frac{\partial}{\partial P\left(\omega_k\right)} \mathcal{L}\left(P, \alpha, \lambda\right) \quad &\Leftrightarrow \quad 0 = -\Big[\log\big(P\left(\omega_k\right)\big) + 1\Big] - \alpha - \sum_{i=1}^{m} \lambda_i f_i\left(\omega_k\right) \\ &\Leftrightarrow \quad \log\big(P\left(\omega_k\right)\big) = -1 - \alpha - \sum_{i=1}^{m} \lambda_i f_i\left(\omega_k\right) \\ &\Leftrightarrow \quad P\left(\omega_k\right) = \exp\left[-1 - \alpha - \sum_{i=1}^{m} \lambda_i f_i\left(\omega_k\right)\right]. \end{aligned} \tag{2.8}$$

The findings explained above lead to the following theorem.

Theorem 2.1.3 (Maximum Entropy Distribution)
Let the probability distribution $P^* : \Omega \to (0,1)$ be given by

$$P^*(\omega) = \exp\left[-1 - \alpha^* - \sum_{i=1}^{m} \lambda_i^* f_i(\omega)\right], \tag{2.9}$$

for $\omega \in \Omega$, and where $(\alpha^*, \lambda_1^*, \ldots, \lambda_m^*) \in \mathbb{R}^{m+1}$ are such that the conditions in Eqs. (2.4) and (2.5) are satisfied, then P^* constitutes the unique solution to the optimization problem defined in Eqs. (2.3) to (2.5).

Proof
Analogous to the continuous case considered in Cover and Thomas (2006) Theorem 12.1.1, we show that P^* uniquely maximizes the Shannon entropy over all possible probability distributions in $\mathcal{P}$. Assume there exists a probability distribution $\tilde{P}$ satisfying the constraints in Eqs. (2.4) and (2.5), then

$$\begin{aligned}
S\left(\tilde{P}\right) &= -\sum_{\omega\in\Omega} \tilde{P}(\omega)\log\left(\tilde{P}(\omega)\right) = -\sum_{\omega\in\Omega} \tilde{P}(\omega)\log\left(\frac{\tilde{P}(\omega)}{P^*(\omega)}P^*(\omega)\right) \\
&= -D_{KL}\left(\tilde{P}\|P^*\right) - \sum_{\omega\in\Omega}\tilde{P}(\omega)\log\left(P^*(\omega)\right) \\
&\leq -\sum_{\omega\in\Omega}\tilde{P}(\omega)\log\left(P^*(\omega)\right), \quad \text{by Lemma 2.1.2 (i)} \\
&= -\sum_{\omega\in\Omega}\tilde{P}(\omega)\left[-1-\alpha^* - \sum_{i=1}^{m}\lambda_i^* f_i(\omega_k)\right] \qquad (2.10) \\
&= -(-1-\alpha^*)\underbrace{\left(\sum_{\omega\in\Omega}\tilde{P}(\omega)\right)}_{=1} + \left[\sum_{\omega\in\Omega}\tilde{P}(\omega)\sum_{i=1}^{m}\lambda_i^* f_i(\omega_k)\right] \\
&= -(-1-\alpha^*)\left(\sum_{\omega\in\Omega}P^*(\omega)\right) + \left[\sum_{i=1}^{m}\lambda_i^* \underbrace{\sum_{\omega\in\Omega}\tilde{P}(\omega) f_i(\omega_k)}_{=\mu_i}\right] \\
&= -(-1-\alpha^*)\left(\sum_{\omega\in\Omega}P^*(\omega)\right) + \left[\sum_{i=1}^{m}\lambda_i^* \sum_{\omega\in\Omega}P^*(\omega) f_i(\omega_k)\right] \\
&= -\sum_{\omega\in\Omega}P^*(\omega)\left[-1-\alpha^* - \sum_{i=1}^{m}\lambda_i^* f_i(\omega_k)\right] \\
&= -\sum_{\omega\in\Omega}P^*(\omega)\log\left(P^*(\omega)\right) = S(P^*). \qquad (2.11)
\end{aligned}$$

Note that in the forth line equality only holds if $\tilde{P} = P^*$ almost everywhere. $\square$

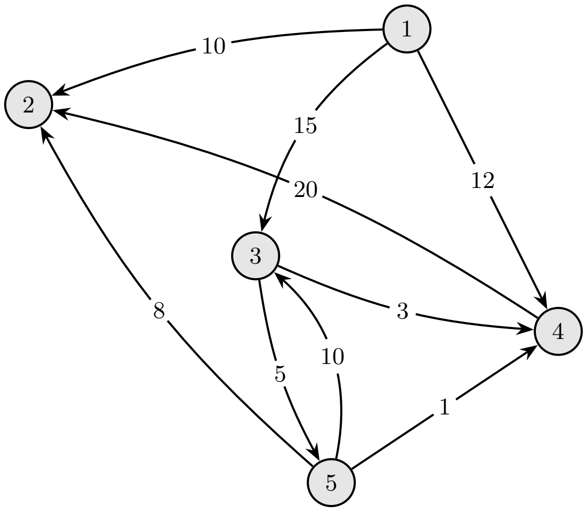
2.1.2 Adaptation to Complex Networks

Having introduced the basic concepts of constrained maximum entropy problems, we now turn to ERGMs. ERGMs constitute a particular family of constrained maximum entropy problems that are adapted to graphs[1] A graph consists of a set of nodes and links. Depending whether the links are assign a direction a graph can be directed or undirected. Furthermore, if the links carry a weight, graphs are said to be weighted, and otherwise unweighted. Figure 2.1 depicts a directed and weighted graph and Fig. 2.2 an undirected and unweighted graph. Commonly, graphs are conveniently represented in form of a matrix, as illustrated in Figs. 2.1 and 2.2.

Definition 2.1.4 (Graph)
A graph G consists of $n \in \mathbb{N}$ nodes and possible links between these nodes. A graph G is conveniently represented by a matrix $W \in \mathbb{R}^{n \times n}$, where the matrix element w_{ij} denotes the link from node i to node j, for all $i, j = 1, \ldots, n$. Commonly, $w_{ij} = 0$ denotes a non-existing link. Furthermore, graphs can have the following characteristics:

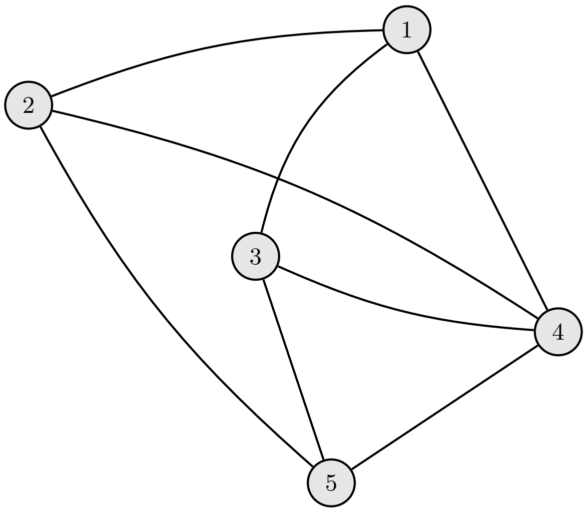
- *If we are only interested whether two nodes i and j are connected, but there is no defined link direction, i.e. $w_{ij} = w_{ji}$ for all $i, j = 1, \ldots, n$, the matrix W is symmetric and the graph G is said to be **undirected**.*
- *Consequently, if we differentiate between a link w_{ij} from node i to node j and a link w_{ji} from node j to node i the graph G is said to be **directed**.*
- *If we are only interested whether a link between two nodes i and j exits, but there is no defined weight of a link, i.e. $w_{ij} \in \{0, 1\}$ for all $i, j = 1, \ldots, n$, the graph G is said to be **unweighted**. In this case, the matrix W is also called the adjacency matrix (commonly denoted by A). Note that there exists an adjacency matrix $A \in \{0, 1\}^{n \times n}$ for every graph W with $a_{ij} := \mathbb{1}_{\{w_{ij} > 0\}}$.*
- *Analogously, if links can carry a weight, i.e. $w_{ij} \in \mathbb{R}$, the graph G is said to be **weighted**.*

[1] The terms ‘graph’ and ‘network’ are used interchangeably throughout this thesis.



$$W = \begin{pmatrix} 0 & 10 & 15 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 20 & 0 & 0 & 0 \\ 0 & 8 & 10 & 1 & 0 \end{pmatrix}$$

Figure 2.1: Example of a directed and weighted graph.



$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Figure 2.2: Example of an undirected and unweighted graph.

Regarding the financial networks studied in this thesis, nodes represent financial institutions and links denote the amount of money lend and borrowed between financial institutions. Regarding the example pictured in Fig. 2.1, the graph indicates that bank 1 lends money of amount 10 to bank 2.

To construct random graphs that fulfill certain desired characteristics in expectation, we turn to ERGMs. In the following let:

- $\mathcal{G}$ denotes a finite set of possible graphs G,

- $\mathcal{P} = \{P : \mathcal{G} \to (0,1)\}$ denotes the set of discrete probability laws defined on $\mathcal{G}$,
- the functions $f_i : \mathcal{G} \to \mathbb{R}$ compute network statistics, for $i \in \{1, \dots, m\}$,
- $\mu_i \in \mathbb{R}$, for $i \in \{1, \dots, m\}$, denote the desired expected values corresponding to the network statistics.

Analogous to Eqs. (2.3) to (2.5), we consider the following constrained maximum entropy problem,

$$\max_{P \in \mathcal{P}} S(P) = \max_{P \in \mathcal{P}} - \sum_{G \in \mathcal{G}} P(G) \log \big(P(G) \big), \tag{2.12}$$

subject to the normalization condition

$$\sum_{G \in \mathcal{G}} P(G) = 1, \tag{2.13}$$

and to the desired network statistics,

$$\sum_{G \in \mathcal{G}} P(G) f_i(G) = \mu_i, \qquad \forall i \in \{1, \dots, m\}. \tag{2.14}$$

Note, in case there are no desired network statistics that are to be met in expectation, i.e. reducing the optimization problem to Eqs. (2.12) and (2.13), the solving probability distribution would assign an equal probability to each graph G in $\mathcal{G}$.

The resulting maximum entropy distribution solving Eqs. (2.12) to (2.14) is given by Theorem 2.1.3 and often reformulated in terms of the *partition function* Z and the *graph Hamiltonian* H. The partition function Z is derived from the normalization condition (i.e. Eq. (2.13)),

$$\begin{aligned} 1 = \sum_{G \in \mathcal{G}} P(G) \quad &\Leftrightarrow \quad 1 = \sum_{G \in \mathcal{G}} \exp\left[-1 - \alpha - \sum_{i=1}^{m} \lambda_i f_i(G) \right], \quad \text{by Theorem 2.1.3} \\ &\Leftrightarrow \quad \exp[1 + \alpha] = \underbrace{\sum_{G \in \mathcal{G}} \exp\left[- \sum_{i=1}^{m} \lambda_i f_i(G) \right]}_{\equiv Z}. \end{aligned} \tag{2.15}$$

The graph Hamiltonian $H : \mathcal{G} \to \mathbb{R}$ is defined as

$$H(G) := \sum_{i=1}^{m} \lambda_i f_i(G). \tag{2.16}$$

It follows

$$P(G) = \frac{e^{-H(G)}}{Z}. \tag{2.17}$$

Moreover, the *free energy* F is defined as

$$F := -\log(Z). \tag{2.18}$$

The free energy F is of interest, because the partial derivatives of F with respect to the Lagrange multipliers constitute a non-linear system of equations, that can be used to calibrate the Lagrange multipliers. More precisely, the partial derivatives yield for all $k = 1, \ldots, m$,

$$\begin{aligned}
\frac{\partial}{\partial \lambda_k} F &= \frac{\partial}{\partial \lambda_k}\left[-\log(Z)\right] = \frac{-1}{Z}\frac{\partial}{\partial \lambda_k}\sum_{G\in\mathcal{G}} \exp\left[-\sum_{i=1}^{m}\lambda_i f_i(G)\right] \\
&= \frac{1}{Z}\sum_{G\in\mathcal{G}} \exp\left[-\sum_{i=1}^{m}\lambda_i f_i(G)\right] f_k(G) = \frac{1}{Z}\sum_{G\in\mathcal{G}} \exp\left[-H(G)\right] f_k(G) \\
&= \sum_{G\in\mathcal{G}} \frac{-e^{H(G)}}{Z} f_k(G) = \sum_{G\in\mathcal{G}} P(G) f_k(G) = \mu_k.
\end{aligned} \tag{2.19}$$

In order to better understand the underlying structures of ERGMs, let's consider an illustrative example. For that purpose and because it will play a central role later, we introduce the degree and strength sequence.

Definition 2.1.5 (Degree Sequence)
Let $W \in \mathbb{R}^{n\times n}$ denote an arbitrary graph and let A with $a_{ij} := \mathbb{1}_{\{w_{ij}>0\}}$ be the corresponding adjacency matrix. The degree sequence $d = (d_1, \ldots, d_n)$ of graph W denotes the number of links that each node is connected to. I.e.,

$$d_i := \sum_{\substack{j=1\\ j\neq i}}^{n} (a_{ij} + a_{ji}), \qquad \forall i = 1, \ldots, n. \tag{2.20}$$

*Furthermore, for a directed graph, the **in-degree sequence** $d^{(\mathrm{in})} = \left(d_1^{(\mathrm{in})}, \ldots, d_n^{(\mathrm{in})}\right)$ and the **out-degree sequence** $d^{(\mathrm{out})} = \left(d_1^{(\mathrm{out})}, \ldots, d_n^{(\mathrm{out})}\right)$ are defined as follows,*

$$d_i^{(\mathrm{in})} := \sum_{\substack{j=1\\ j\neq i}}^{n} a_{ji}, \qquad \text{and} \qquad d_i^{(\mathrm{out})} := \sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij}, \qquad \forall i = 1, \ldots, n. \tag{2.21}$$

Definition 2.1.6 (Strength Sequence)
Let $W \in \mathbb{R}^{n\times n}$ denote an arbitrary weighted graph. The strength sequence $s = (s_1, \ldots, s_n)$

of graph W denotes the sum of weights that the links of each node carry. I.e.,

$$s_i := \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(w_{ij} + w_{ji} \right), \qquad \forall i = 1, \ldots, n. \tag{2.22}$$

Furthermore, for a directed graph, the ***in-strength sequence*** $s^{(\text{in})} = \left(s_1^{(\text{in})}, \ldots, s_n^{(\text{in})} \right)$ *and the* ***out-strength sequence*** $s^{(\text{out})} = \left(s_1^{(\text{out})}, \ldots, s_n^{(\text{out})} \right)$ *are defined as follows,*

$$s_i^{(\text{in})} := \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ji}, \qquad \textit{and} \qquad s_i^{(\text{out})} := \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ij}, \qquad \forall i = 1, \ldots, n, \tag{2.23}$$

i.e. the sum of inflowing and outflowing weight, respectively.

Regarding the illustrative example, let the set of possible graphs $\mathcal{G}$ be given by all unweighted and directed graphs with $n \in \mathbb{N}$ nodes, excluding self-loops, i.e.

$$\mathcal{G} = \left\{ a \in \{0,1\}^{n \times n} : a_{11} = \cdots = a_{nn} = 0 \right\}. \tag{2.24}$$

A short remark on the exclusion of self-loops: Self-loops are edges pointing from a node to itself. For example, in financial networks, a self-loop would mean that a bank is lending money to itself, which makes no sense. In fact, most real-world networks do not exhibit self-loops, which is why the diagonal of the graph matrices is usually set to zero.

Regarding the desired network statistics in our considered example, let's assume we want node i to have in expectation d_i edges, i.e. let $d = (d_1, \ldots, d_n) \in \big(0, 2\,(n-1)\big)^n$ denote the desired degree sequence. The corresponding graph functions $k_i : \mathcal{G} \to \big(0, 2\,(n-1)\big)$, for all $i = 1, \ldots, n$, counting the number of edges of node i in a graph $a \in \mathcal{G}$ are given by

$$k_i\,(a) := \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(a_{ij} + a_{ji} \right). \tag{2.25}$$

This ERGM is formally described by the constrained maximum entropy problem

$$\max_{P \in \mathcal{P}} S\,(P) = \max_{P \in \mathcal{P}} - \sum_{a \in \mathcal{G}} P\,(a) \log \big(P\,(a) \big), \tag{2.26}$$

subject to

$$\begin{aligned} \sum_{a\in\mathcal{G}} P(a) &= 1, \\ \sum_{a\in\mathcal{G}} P(a) k_i\left(a\right) &= d_i, \qquad \forall i \in \{1,\dots,n\}. \end{aligned} \tag{2.27}$$

From Theorem 2.1.3 we know, that if a solution to the optimization problem of Eqs. (2.26) and (2.27) exists, it is unique and the solving probability distribution P takes the following form

$$P(a) = \frac{e^{-H(a)}}{Z} = Z^{-1}\exp\left[-\sum_{i=1}^{n}\lambda_i k_i\left(a\right)\right], \tag{2.28}$$

where H denotes the graph Hamiltonian, as defined in Eq. (2.16), Z the partition function, as defined in Eq. (2.15), and $(\lambda_1,\dots,\lambda_n) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the degree sequence. An essential point is that we can write the Hamiltonian as a simple sum over the elements of the graph matrix:

$$\begin{aligned} H\left(a\right) &= \sum_{i=1}^{n}\lambda_i k_i\left(a\right) = \sum_{i=1}^{n}\lambda_i\Bigg(\sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij} + a_{ji}\Bigg) = \Bigg(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i a_{ij}\Bigg) + \Bigg(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i a_{ji}\Bigg) \\ &= \Bigg(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i a_{ij}\Bigg) + \Bigg(\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n}\lambda_j a_{ij}\Bigg) = \sum_{i\neq j}\left(\lambda_i + \lambda_j\right) a_{ij}. \end{aligned} \tag{2.29}$$

This allows us to compute the partition function Z analytically,

$$\begin{aligned} Z &= \sum_{a\in\mathcal{G}}\exp\left[-H\left(a\right)\right] \\ &= \sum_{a\in\mathcal{G}}\exp\left[-\sum_{i\neq j}\left(\lambda_i+\lambda_j\right)a_{ij}\right] \\ &= \sum_{a\in\mathcal{G}}\prod_{i\neq j} e^{-(\lambda_i+\lambda_j)a_{ij}} \\ &= \sum_{a\in\mathcal{G}} e^{-(\lambda_1+\lambda_2)a_{12}} \prod_{\substack{i\neq j\\ (i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \end{aligned}$$

$$
\begin{aligned}
&= \left[\sum_{\{a\in\mathcal{G}:a_{12}=0\}} e^{-(\lambda_1+\lambda_2)a_{12}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right] \\
&\quad + \left[\sum_{\{a\in\mathcal{G}:a_{12}=1\}} e^{-(\lambda_1+\lambda_2)a_{12}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right] \\
&= \left[e^{-(\lambda_1+\lambda_2)a_{12}}\Big|_{a_{12}=0} \right] \underbrace{\left[\sum_{\{a\in\mathcal{G}:a_{12}=0\}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right]}_{=\star} \\
&\quad + \left[e^{-(\lambda_1+\lambda_2)a_{12}}\Big|_{a_{12}=1} \right] \underbrace{\left[\sum_{\{a\in\mathcal{G}:a_{12}=1\}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right]}_{=\star} \\
&= \sum_{a_{12}^*=0}^{1} e^{-(\lambda_1+\lambda_2)a_{12}^*} \underbrace{\left[\sum_{\left\{a\in\mathcal{G}:a_{12}=a_{12}^*\right\}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right]}_{=\star,\quad \text{constant for all } a_{12}^*} \\
&= \left[\sum_{a_{12}^*=0}^{1} e^{-(\lambda_1+\lambda_2)a_{12}^*} \right] \times \left[\sum_{\{a\in\mathcal{G}:a_{12}=0\}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right] \\
&= \left[\sum_{a_{12}^*=0}^{1} e^{-(\lambda_1+\lambda_2)a_{12}^*} \right] \left[\sum_{a_{13}^*=0}^{1} e^{-(\lambda_1+\lambda_3)a_{13}^*} \right] \times \left[\sum_{\{a\in\mathcal{G}:a_{12}=0,a_{13}=0\}} \prod_{\substack{i\neq j\\(i,j)\neq(1,2),(1,3)}} e^{-(\lambda_i+\lambda_j)a_{ij}} \right] \\
&= \dots \\
&= \prod_{i\neq j} \left[\sum_{a_{ij}=0}^{1} e^{-(\lambda_i+\lambda_j)a_{ij}} \right] = \prod_{i\neq j} \left[1 + e^{-(\lambda_i+\lambda_j)} \right].
\end{aligned}
\tag{2.30}
$$

Note that the complexity of computing Z has been reduced from summing over the $2^{n(n-1)}$ matrices in $\mathcal{G}$ to a product of $\binom{n}{2} = \frac{n!}{(n-2)!\,2} = \frac{n(n-1)}{2}$ factors. This now allows the derivation of a non-linear system of equations, that defines the Lagrange multipliers, by taking partial derivatives of the free energy F as explained in Eq. (2.19). For

$k = 1, \ldots, n$,

$$\begin{aligned} d_k = \frac{\partial}{\partial \lambda_k} F &= \frac{\partial}{\partial \lambda_k} \left[-\log(Z) \right] \\ &= \frac{\partial}{\partial \lambda_k} \left[-\sum_{i \neq j} \log\left(1 + e^{-(\lambda_i + \lambda_j)}\right) \right] \\ &= \frac{\partial}{\partial \lambda_k} \left[-\sum_{\substack{i=1 \\ i \neq k}}^{n} \log\left(1 + e^{-(\lambda_i + \lambda_k)}\right) - \sum_{\substack{j=1 \\ j \neq k}}^{n} \log\left(1 + e^{-(\lambda_k + \lambda_j)}\right) \right] \\ &= -2 \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{-e^{-(\lambda_i + \lambda_k)}}{1 + e^{-(\lambda_i + \lambda_k)}} = 2 \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{e^{-(\lambda_i + \lambda_k)}}{1 + e^{-(\lambda_i + \lambda_k)}}. \end{aligned} \tag{2.31}$$

Moreover, because the graph Hamiltonian constitutes a simple sum over the elements of the graph matrix, we can derive the independent link probabilities for all $k, l \in \{1, \ldots, n\}$ and $k \neq l$,

$$\begin{aligned} & P(A_{kl} = 1) \\ & = \sum_{\{a \in \mathcal{G} : a_{kl} = 1\}} P(a) = Z^{-1} \sum_{\{a \in \mathcal{G} : a_{kl} = 1\}} e^{-H(a)} \\ & = \prod_{i \neq j} \left[1 + e^{-(\lambda_i + \lambda_j)}\right]^{-1} \sum_{\{a \in \mathcal{G} : a_{kl} = 1\}} \prod_{i \neq j} e^{-(\lambda_i + \lambda_j) a_{ij}} \\ & = \prod_{i \neq j} \left[1 + e^{-(\lambda_i + \lambda_j)}\right]^{-1} e^{-(\lambda_k + \lambda_l)} \sum_{\{a \in \mathcal{G} : a_{kl} = 1\}} \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} e^{-(\lambda_i + \lambda_j) a_{ij}} \\ & = \prod_{i \neq j} \left[1 + e^{-(\lambda_i + \lambda_j)}\right]^{-1} e^{-(\lambda_k + \lambda_l)} \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} \left[1 + e^{-(\lambda_i + \lambda_j)}\right], \quad \text{analogous to Eq. (2.30)} \\ & = \frac{e^{-(\lambda_k + \lambda_l)}}{1 + e^{-(\lambda_k + \lambda_l)}}, \end{aligned} \tag{2.32}$$

where A denotes a random matrix defined on the probability space $(\mathcal{G}, \mathfrak{P}(\mathcal{G}), P)$ and $\mathfrak{P}(\mathcal{G})$ refers to the power set of $\mathcal{G}$.

Summing up, the n Lagrange multipliers can be calibrated via the n-dimensional system of equations given by Eq. (2.31). Afterwards, we can easily sample random graphs by drawing independent Bernoulli trials according to the link probabilities given by Eq. (2.32).

As the above discussed example indicates, a critical component of ERGMs is the computation of the partition function Z. In most cases, the set $\mathcal{G}$ contains a substantial number of graphs, which leaves the evaluation of Z, as a sum over all $G \in \mathcal{G}$, computationally infeasible. Fortunately, whenever the desired network statistics, represented by the functions $f_i\left(G\right)$, constitute simple sums over the edges of G, and choosing $\mathcal{G}$ wisely, the partition function Z can be derived analytically. For that purpose, $\mathcal{G}$ is typically set to contain all possible graphs with a given number of nodes. Nevertheless, the flexibility in the particular choice of $\mathcal{G}$ and the constraints f_i give rise to a variety of network models.

We further remark, that while the uniqueness of the solving probability distribution of ERGMs is constituted by Theorem 2.1.3, the existence of a solution depends on the particular considered instance and has, to the best of our knowledge, not yet been solved.

2.1.3 Extension to Weighted Networks

ERGMs can further be extended to weighted networks. Park and Newman (2004) consider an ERGM conditioned on the strength sequence $s = (s_1, \ldots, s_n) \in \mathbb{R}^n_{>0}$, which describes the total weight of each node. Moreover, the set $\mathcal{G}$ of graphs allows all non-negative natural numbers as link weights, i.e. link weights are in $\mathbb{N}_0$. In contrast, Garlaschelli and Loffredo (2009) restrict the allowed link weights to an upper bound $N \in \mathbb{N}$, i.e. link weights are in $\{0, 1, \ldots, N\}$. Moreover, Garlaschelli and Loffredo (2009) consider several ERGMs conditioned on the strength sequence $s = (s_1, \ldots, s_n) \in \mathbb{R}^n_{>0}$ and/ or the degree sequence $d = (d_1, \ldots, d_n) \in \big(0, 2\left(n-1\right)\big)^n$.

The trick to compute the partition function Z is to apply the geometric series, which leads to a simplified closed form expression for Z. Since calculations work mostly analogous in all described settings, here, we introduce a more detailed weighted ERGM by additionally differentiating between the directions 'in-' and 'out-'. As we will see, allowing for unbounded link weights, i.e. in $\mathbb{N}_0$, yields realistic bounded expected weights for all links. Therefore, here, we introduce this case. However, the case of bounded link weights can be derived analogously by applying the geometric series for bounded sums.

We explain the neccessary calculations in the following, by deriving the ERGM which fulfills in expectation

(i) an in-degree sequence $d^{(\mathrm{in})} = \left(d_1^{(\mathrm{in})}, \ldots, d_n^{(\mathrm{in})}\right) \in (0, n-1)^n$, describing the desired number of incoming edges of each node,

(ii) an out-degree sequence $d^{(\mathrm{out})} = \left(d_1^{(\mathrm{out})}, \ldots, d_n^{(\mathrm{out})}\right) \in (0, n-1)^n$, describing the desired number of outgoing edges of each node,

(iii) an in-strength sequence $s^{(\text{in})} = \left(s_1^{(\text{in})}, \ldots, s_n^{(\text{in})}\right) \in \mathbb{R}^n_{>0}$, describing the desired inflowing weight of each node,

(iv) an out-strength sequence $s^{(\text{out})} = \left(s_1^{(\text{out})}, \ldots, s_n^{(\text{out})}\right) \in \mathbb{R}^n_{>0}$, describing the desired outflowing weight of each node.

Let the set of considered graphs $\mathcal{G}$ be given by

$$\mathcal{G} = \left\{ w \in \mathbb{N}_{\geq 0}^{n \times n} : w_{11} = \cdots = w_{nn} = 0 \right\}. \tag{2.33}$$

The ERGM is formally described by the constrained maximum entropy problem

$$\max_{P \in \mathcal{P}} S\left(P\right) = \max_{P \in \mathcal{P}} - \sum_{w \in \mathcal{G}} P\left(w\right) \log\left(P\left(w\right)\right), \tag{2.34}$$

subject to

$$\begin{aligned}
\sum_{w \in \mathcal{G}} P(w) &= 1, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{i=1}^{n} \mathbb{1}_{\{w_{ij} > 0\}} &= d_j^{(\text{in})}, \quad \text{for } j = 1, \ldots, n, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{j=1}^{n} \mathbb{1}_{\{w_{ij} > 0\}} &= d_i^{(\text{out})}, \quad \text{for } i = 1, \ldots, n, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{i=1}^{n} w_{ij} &= s_j^{(\text{in})}, \quad \text{for } j = 1, \ldots, n, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{j=1}^{n} w_{ij} &= s_i^{(\text{out})}, \quad \text{for } i = 1, \ldots, n.
\end{aligned} \tag{2.35}$$

From Theorem 2.1.3 we know, that if a solution to the optimization problem of Eqs. (2.34) and (2.35) exists, it is unique and the solving probability distribution P takes the following form

$$\begin{aligned}
P(w) &= \frac{e^{-H(w)}}{Z} \\
&= Z^{-1} \exp\left[- \sum_{i \neq j} \left(\lambda_i^{(\text{out})} + \lambda_j^{(\text{in})}\right) \mathbb{1}_{\{w_{ij} > 0\}} + \left(\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\right) w_{ij} \right],
\end{aligned} \tag{2.36}$$

where H denotes the graph Hamiltonian, as defined in Eq. (2.16), Z the partition function, as defined in Eq. (2.15), and

(i) $\left(\lambda_1^{(\mathrm{in})}, \ldots, \lambda_n^{(\mathrm{in})}\right) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the in-degree sequence,

(ii) $\left(\lambda_1^{(\mathrm{out})}, \ldots, \lambda_n^{(\mathrm{out})}\right) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the out-degree sequence,

(iii) $\left(\theta_1^{(\mathrm{in})}, \ldots, \theta_n^{(\mathrm{in})}\right) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the in-strength sequence,

(iv) $\left(\theta_1^{(\mathrm{out})}, \ldots, \theta_n^{(\mathrm{out})}\right) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the out-strength sequence.

To simplify notation, in the following we write $\lambda_{ij} := \left(\lambda_i^{(\mathrm{out})} + \lambda_j^{(\mathrm{in})}\right)$ and $\theta_{ij} := \left(\theta_i^{(\mathrm{out})} + \theta_j^{(\mathrm{in})}\right)$.

Since the Hamiltonian constitutes a simple sum over the elements of the graph matrix, the partition function Z can be computed analytically,

$$
\begin{aligned}
Z &= \sum_{w \in \mathcal{G}} \exp\left[-H(w)\right] \\
&= \sum_{w \in \mathcal{G}} \prod_{i \neq j} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \\
&= \sum_{w \in \mathcal{G}} e^{-\left(\lambda_{12} \mathbb{1}_{\{w_{12}>0\}} + \theta_{12} w_{12}\right)} \prod_{\substack{i \neq j \\ (i,j) \neq (1,2)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \\
&= \sum_{w_{12}^*=0}^{\infty} \left[\sum_{\left\{w \in \mathcal{G} : w_{12} = w_{12}^*\right\}} e^{-\left(\lambda_{12} \mathbb{1}_{\left\{w_{12}^*>0\right\}} + \theta_{12} w_{12}^*\right)} \prod_{\substack{i \neq j \\ (i,j) \neq (1,2)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \right] \\
&= \sum_{w_{12}^*=0}^{\infty} e^{-\left(\lambda_{12} \mathbb{1}_{\left\{w_{12}^*>0\right\}} + \theta_{12} w_{12}^*\right)} \underbrace{\left[\sum_{\left\{w \in \mathcal{G} : w_{12} = w_{12}^*\right\}} \prod_{\substack{i \neq j \\ (i,j) \neq (1,2)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \right]}_{\text{constant for all } w_{12}^*} \\
&= \left[\sum_{w_{12}^*=0}^{\infty} e^{-\left(\lambda_{12} \mathbb{1}_{\left\{w_{12}^*>0\right\}} + \theta_{12} w_{12}^*\right)} \right] \times \left[\sum_{\{w \in \mathcal{G} : w_{12} = 0\}} \prod_{\substack{i \neq j \\ (i,j) \neq (1,2)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \right] \\
&= \ldots
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(\star)}{=} \prod_{i \neq j} \left[\sum_{w_{ij}=0}^{\infty} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij}>0\right\}} + \theta_{ij} w_{ij}\right)} \right] \\
&= \prod_{i \neq j} \left[1 + \sum_{w_{ij}=1}^{\infty} e^{-(\lambda_{ij} + \theta_{ij} w_{ij})} \right] \\
&= \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \left(-1 + \sum_{w_{ij}=0}^{\infty} e^{-\theta_{ij} w_{ij}} \right) \right] \\
&= \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \left(-1 + \frac{1}{1 - e^{-\theta_{ij}}} \right) \right], \qquad \text{by the geometric series,} \\
&= \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}} \right],
\end{aligned}
\tag{2.37}
$$

where the same algebraic steps are applied to all elements in $(\star)$ as we applied exemplary to the first element w_{12}. Note that

$$
e^{-\theta_{ij}} < 1 \quad \Leftrightarrow \quad \theta_{ij} > 0 \tag{2.38}
$$

has to hold for all $i \neq j$, since otherwise the value of the partition function Z is infinity, which implies that P from Eq. (2.36) is not a solving probability measure.

Next, we can derive the non-linear system of equations, that defines the Lagrange multipliers, by taking partial derivatives of the free energy F as explained in Eq. (2.19). For $k = 1, \ldots, n$,

$$
\begin{aligned}
d_k^{(\mathrm{in})} &= \frac{\partial}{\partial \lambda_k^{(\mathrm{in})}} F = \frac{\partial}{\partial \lambda_k^{(\mathrm{in})}} \big[-\log (Z) \big] \\
&= \frac{\partial}{\partial \lambda_k^{(\mathrm{in})}} \left[-\sum_{i \neq j} \log \left(1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}} \right) \right] \\
&= \frac{\partial}{\partial \lambda_k^{(\mathrm{in})}} \left[-\sum_{\substack{i=1 \\ i \neq k}}^{n} \log \left(1 + e^{-\lambda_{ik}} \frac{e^{-\theta_{ik}}}{1 - e^{-\theta_{ik}}} \right) \right] \\
&= -\sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{1 - e^{-\theta_{ik}}}{1 - e^{-\theta_{ik}} + e^{-\lambda_{ik}} e^{-\theta_{ik}}} \left(-e^{-\lambda_{ik}} \frac{e^{-\theta_{ik}}}{1 - e^{-\theta_{ik}}} \right) = \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{e^{-\lambda_{ik} - \theta_{ik}}}{1 - e^{-\theta_{ik}} + e^{-\lambda_{ik} - \theta_{ik}}}.
\end{aligned}
\tag{2.39}
$$

Analogously, we get

$$d_k^{(\text{out})} = \sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{e^{-\lambda_{kj}-\theta_{kj}}}{1 - e^{-\theta_{kj}} + e^{-\lambda_{kj}-\theta_{kj}}}. \tag{2.40}$$

Taking partial derivatives of F w.r.t. the Lagrange multipliers of the strength yields,

$$\begin{aligned}
s_k^{(\text{in})} &= \frac{\partial}{\partial\theta_k^{(\text{in})}} F = \frac{\partial}{\partial\theta_k^{(\text{in})}} \left[-\sum_{\substack{i=1 \\ i\neq k}}^{n} \log\left(1 + e^{-\lambda_{ik}} \frac{e^{-\theta_{ik}}}{1-e^{-\theta_{ik}}}\right)\right] \\
&= -\sum_{\substack{i=1 \\ i\neq k}}^{n} \frac{1-e^{-\theta_{ik}}}{1-e^{-\theta_{ik}} + e^{-\lambda_{ik}}e^{-\theta_{ik}}} \left(e^{-\lambda_{ik}} \frac{-e^{-\theta_{ik}}\left(1-e^{-\theta_{ik}}\right) - e^{-\theta_{ik}}e^{-\theta_{ik}}}{\left(1-e^{-\theta_{ik}}\right)^2}\right) \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{n} \frac{e^{-\lambda_{ik}-\theta_{ik}}}{\left(1-e^{-\theta_{ik}} + e^{-\lambda_{ik}-\theta_{ik}}\right)\left(1-e^{-\theta_{ik}}\right)}.
\end{aligned} \tag{2.41}$$

Analogously, we get

$$s_k^{(\text{out})} = \sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{e^{-\lambda_{kj}-\theta_{kj}}}{\left(1-e^{-\theta_{kj}} + e^{-\lambda_{kj}-\theta_{kj}}\right)\left(1-e^{-\theta_{kj}}\right)}. \tag{2.42}$$

Moreover, because the graph Hamiltonian constitutes a sum over the elements of the graph matrix, the probability that a link exists and its expected weight can be derived analytically. We note that Garlaschelli and Loffredo (2009) provide the final expression for the probability that a certain link takes a specific weight, however, the neccessary calculations are not shown. Here, we provide the precise mathematical derivation.

Let W denote a random matrix defined on the probability space $(\mathcal{G}, \mathfrak{P}(\mathcal{G}), P)$, where $\mathfrak{P}(\mathcal{G})$ refers to the power set of $\mathcal{G}$. The independent link probabilities for all

$k, l \in \{1, \dots, n\}$ and $k \neq l$ can be computed as follows,

$$\begin{aligned}
&P\left(W_{kl} > 0\right) \\
&\quad = \sum_{\{w \in \mathcal{G}: w_{kl} > 0\}} P\left(w\right) \\
&\quad = Z^{-1} \sum_{\{w \in \mathcal{G}: w_{kl} > 0\}} e^{-H(w)} \\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1} \sum_{\{w \in \mathcal{G}: w_{kl} > 0\}} \prod_{i \neq j} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij} > 0\right\}} + \theta_{ij} w_{ij}\right)} \\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1} \\
&\qquad \left[\sum_{w_{kl}=1}^{\infty} e^{-(\lambda_{kl} + \theta_{kl} w_{kl})}\right] \sum_{\{w \in \mathcal{G}: w_{kl} = 1\}} \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij} > 0\right\}} + \theta_{ij} w_{ij}\right)} \\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1} \\
&\qquad \left[e^{-\lambda_{kl}} \frac{e^{-\theta_{kl}}}{1 - e^{-\theta_{kl}}}\right] \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right], \quad \text{analogous to Eq. (2.37)} \\
&\quad = \left[1 + e^{-\lambda_{kl}} \frac{e^{-\theta_{kl}}}{1 - e^{-\theta_{kl}}}\right]^{-1} \frac{e^{-\lambda_{kl} - \theta_{kl}}}{1 - e^{-\theta_{kl}}} \\
&\quad = \frac{1 - e^{-\theta_{kl}}}{1 - e^{-\theta_{kl}} + e^{-\lambda_{kl} - \theta_{kl}}} \cdot \frac{e^{-\lambda_{kl} - \theta_{kl}}}{1 - e^{-\theta_{kl}}} \\
&\quad = \frac{e^{-\lambda_{kl} - \theta_{kl}}}{1 - e^{-\theta_{kl}} + e^{-\lambda_{kl} - \theta_{kl}}}.
\end{aligned} \tag{2.43}$$

Similarly, we can derive the expected link weights as follows,

$$
\begin{aligned}
&\mathbb{E}\left[W_{kl}\right] \\
&\quad= \sum_{w\in\mathcal{G}} w_{kl} P\left(w\right) = Z^{-1}\sum_{w\in\mathcal{G}} w_{kl} e^{-H(w)} \\
&\quad= \prod_{i\neq j}\left[1+e^{-\lambda_{ij}}\frac{e^{-\theta_{ij}}}{1-e^{-\theta_{ij}}}\right]^{-1}\sum_{w\in\mathcal{G}} w_{kl}\prod_{i\neq j} e^{-\left(\lambda_{ij}\mathbb{1}_{\left\{w_{ij}>0\right\}}+\theta_{ij}w_{ij}\right)} \\
&\quad= \prod_{i\neq j}\left[1+e^{-\lambda_{ij}}\frac{e^{-\theta_{ij}}}{1-e^{-\theta_{ij}}}\right]^{-1} \\
&\qquad\left[\sum_{w_{kl}=1}^{\infty} w_{kl} e^{-(\lambda_{kl}+\theta_{kl}w_{kl})}\right]\sum_{\{w\in\mathcal{G}:w_{kl}=1\}}\prod_{\substack{i\neq j\\(i,j)\neq(k,l)}} e^{-\left(\lambda_{ij}\mathbb{1}_{\left\{w_{ij}>0\right\}}+\theta_{ij}w_{ij}\right)} \\
&\quad= \prod_{i\neq j}\left[1+e^{-\lambda_{ij}}\frac{e^{-\theta_{ij}}}{1-e^{-\theta_{ij}}}\right]^{-1} \\
&\qquad\left[-\sum_{w_{kl}=1}^{\infty}\frac{\partial}{\partial\theta_k^{(\mathrm{out})}} e^{-(\lambda_{kl}+\theta_{kl}w_{kl})}\right]\prod_{\substack{i\neq j\\(i,j)\neq(k,l)}}\left[1+e^{-\lambda_{ij}}\frac{e^{-\theta_{ij}}}{1-e^{-\theta_{ij}}}\right], \quad \text{analogous to Eq. (2.37)} \\
&\quad= \left[1+e^{-\lambda_{kl}}\frac{e^{-\theta_{kl}}}{1-e^{-\theta_{kl}}}\right]^{-1}\left[-\frac{\partial}{\partial\theta_k^{(\mathrm{out})}}\sum_{w_{kl}=1}^{\infty} e^{-(\lambda_{kl}+\theta_{kl}w_{kl})}\right] \\
&\quad= \left[1+e^{-\lambda_{kl}}\frac{e^{-\theta_{kl}}}{1-e^{-\theta_{kl}}}\right]^{-1}\left[-\frac{\partial}{\partial\theta_k^{(\mathrm{out})}}\left[-1+e^{-\lambda_{kl}}\frac{e^{-\theta_{kl}}}{1-e^{-\theta_{kl}}}\right]\right] \\
&\quad= \left[1+e^{-\lambda_{kl}}\frac{e^{-\theta_{kl}}}{1-e^{-\theta_{kl}}}\right]^{-1}\left[-e^{-\lambda_{kl}}\frac{-e^{-\theta_{kl}}\left(1-e^{-\theta_{kl}}\right)+e^{-\theta_{kl}}e^{-\theta_{kl}}}{\left(1-e^{-\theta_{kl}}\right)^2}\right] \\
&\quad= \frac{1-e^{-\theta_{kl}}}{1-e^{-\theta_{kl}}+e^{-\lambda_{kl}-\theta_{kl}}}\left[e^{-\lambda_{kl}}\frac{e^{-\theta_{kl}}}{\left(1-e^{-\theta_{kl}}\right)^2}\right] = \frac{1}{1-e^{-\theta_{kl}}+e^{-\lambda_{kl}-\theta_{kl}}}\cdot\frac{e^{-\lambda_{kl}-\theta_{kl}}}{1-e^{-\theta_{kl}}} \\
&\quad= \frac{P\left(W_{kl}>0\right)}{1-e^{-\theta_{kl}}}.
\end{aligned}
\tag{2.44}
$$

Furthermore, we can derive the probability that a random link W_{kl} takes a specific weight $w_{kl}^* \in \mathbb{N}_{\geq 0}$,

$$
\begin{aligned}
&P\left(W_{kl} = w_{kl}^*\right)\\
&\quad = \sum_{\left\{w \in \mathcal{G} : w_{kl} = w_{kl}^*\right\}} P\left(w\right) = Z^{-1} \sum_{\left\{w \in \mathcal{G} : w_{kl} = w_{kl}^*\right\}} e^{-H(w)}\\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1} \sum_{\left\{w \in \mathcal{G} : w_{kl} = w_{kl}^*\right\}} \prod_{i \neq j} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij} > 0\right\}} + \theta_{ij} w_{ij}\right)}\\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1}\\
&\qquad e^{-\left(\lambda_{kl} \mathbb{1}_{\left\{w_{kl}^* > 0\right\}} + \theta_{kl} w_{kl}^*\right)} \sum_{\left\{w \in \mathcal{G} : w_{kl} = w_{kl}^*\right\}} \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} e^{-\left(\lambda_{ij} \mathbb{1}_{\left\{w_{ij} > 0\right\}} + \theta_{ij} w_{ij}\right)}\\
&\quad = \prod_{i \neq j} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right]^{-1}\\
&\qquad e^{-\left(\lambda_{kl} \mathbb{1}_{\left\{w_{kl}^* > 0\right\}} + \theta_{kl} w_{kl}^*\right)} \prod_{\substack{i \neq j \\ (i,j) \neq (k,l)}} \left[1 + e^{-\lambda_{ij}} \frac{e^{-\theta_{ij}}}{1 - e^{-\theta_{ij}}}\right], \quad \text{analogous to Eq. (2.37)}\\
&\quad = \left[1 + e^{-\lambda_{kl}} \frac{e^{-\theta_{kl}}}{1 - e^{-\theta_{kl}}}\right]^{-1} e^{-\left(\lambda_{kl} \mathbb{1}_{\left\{w_{kl}^* > 0\right\}} + \theta_{kl} w_{kl}^*\right)}\\
&\quad = \frac{1 - e^{-\theta_{kl}}}{1 - e^{-\theta_{kl}} + e^{-\lambda_{kl} - \theta_{kl}}} \cdot e^{-\left(\lambda_{kl} \mathbb{1}_{\left\{w_{kl}^* > 0\right\}} + \theta_{kl} w_{kl}^*\right)}.
\end{aligned}
\tag{2.45}
$$

Since the links are pairwise independet, we get the following functional form for the probability of a certain graph $w^* \in \mathcal{G}$,

$$
P\left(W = w^*\right) = \prod_{i \neq j} P\left(W_{ij} = w_{ij}^*\right). \tag{2.46}
$$

Summing up, the $4n$ Lagrange multipliers can be calibrated via the $4n$-dimensional system of equations given by Eqs. (2.39) to (2.42). Afterwards, we can easily sample random graphs by drawing independent Bernoulli trials according to the link probabilities given by Eq. (2.43). Weights can be assigned to drawn links either via the expected weight according to Eq. (2.44) or by randomly sampling weights according to their probabilities as given by Eq. (2.45). The probability of a certain graph $w^* \in \mathcal{G}$ can be assessed by Eq. (2.46).

2.2 Fitness Models

Fitness models are based on the idea that the generation of links depends on underlying intrinsic characteristics of the nodes, see Caldarelli et al. (2002); Servedio et al. (2004). For example, in social networks friendships are formed because individuals share common interests, in biological networks proteins interact with each other according to their chemical and physical properties, and the creation of hyperlinks in the world wide web is influenced by the popularity of a webpage. Hence, each node can be thought of being equipped with a hidden variable that controls its probability to create edges.

Formally, fitness models are defined as follows. Let $A \in \{0,1\}^{n \times n}$ denote a random (unweighted) network consisting of n nodes, where each node i is assigned a fitness $x_i \in \mathbb{R}$. The probability of an edge existing between nodes i and j is described by a link probability function $f : \mathbb{R}^2 \to [0,1]$, that takes as input the fitness variables of both nodes, i.e. $P\left(A_{ij} = 1\right) = f\left(x_i, x_j\right)$. This setting obviously gives rise to a variety of network models. In general, we can differentiate between two directions. First, as exemplified above, we might have a clear understanding of which nodes' features drive their tendency of generating edges, and data on these features might be readily available. Second, the fitnesses can be randomized and the structure of the link probability function f can be derived, such that the resulting network exhibits some desired characteristic. In the following, both approaches are discussed in detail.

2.2.1 Empirical Fitness Variables

To properly understand the rationale to exploit nodes' properties, where data is readily available, as fitness variables, we have to take a step back and discuss the calibration of ERGMs. ERGMs, as introduced in Section 2.1, constitute a powerful methodology with a sound mathematical foundation, that can be applied as long as the desired network statistics μ_i (see Eqs. (2.12) to (2.14)) are explicitly available. Let's recall the example of the ERGM discussed in Section 2.1.2 and set up in Eqs. (2.26) and (2.27), that satisfies a given degree sequence $d = (d_1, \dots, d_n) \in (0, 2\,(n-1))^n$. A system of equations was derived in Eq. (2.31), which enables the calibration of the Lagrange multipliers. Note that for the calibration we need to know $(d_1, \dots, d_n)$ explicitly. However, in many real world applications, especially regarding economic and financial networks, the precise degree sequence is not readily available. Very often we have knowledge of a certain heterogeneity across the nodes. For example, financial institutions that are big in terms of total assets or in terms of their total interbank lending volumes, tend to have more financial relationships with other institutions than smaller ones. So we know that there exists some heterogeneity, however, for confidentiality reasons, the explicit numbers of connections that each institution maintains cannot be retrieved.

To tackle this issue, academics have turned to fitness variables. Promising empirical results were found by Garlaschelli and Loffredo (2008) regarding the World Trade Web

(WTW). For clarification, the WTW depicts the trade relationships between all countries. Garlaschelli and Loffredo identified a linear relationship between the exponential function of the negative Lagrange multipliers and the countries (rescaled) GDP. More precisely, again recall the ERGM discussed in Section 2.1.2 and set up in Eqs. (2.26) and (2.27), that satisfies a given degree sequence $d = (d_1, \ldots, d_n) \in \big(0, 2\,(n-1)\,\big)^n$, and let $s = (s_1, \ldots, s_n) \in \mathbb{R}^n_{\geq 0}$ denote the (rescaled) GDP of the n considered countries. Garlaschelli and Loffredo (2008) found the following empirical relationship, for all $i = 1, \ldots, n$,

$$e^{-\lambda_i} \approx z s_i, \tag{2.47}$$

where $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ denotes the corresponding Lagrange multipliers (see Eq. (2.28)) and for some scalar $z \in \mathbb{R}$. Even though at first sight this relationship might be surprising, further similar examples have been published (we will discuss them later) and these results are extremely helpful as they enable the calibration of network models in light of scarce available data. Moreover, in Section 4.2 we provide some first mathematical insight into this phenomenon. The mathematical explanation seems to be that a power law distribution of the fitness variables translates into a power law distribution (with a saturation effect) of the degrees.

The linear relationship found by Garlaschelli and Loffredo (2008) conveniently allows the calibration of the ERGM without knowing the exact degree sequence. Of course a particular degree sequence will not be recovered precisely, but Garlaschelli and Loffredo (2008) show that the degree distribution is reproduced remarkably well. The network model considers unweighted and undirected graphs $A \in \{0,1\}^{n \times n}$ symmetric, i.e. with $A = A^\top$. The link probabilities are given for $i, j = 1, \ldots, n$ and $i \neq j$ by

$$P\left(A_{ij} = 1\right) = \frac{z^2 s_i s_j}{1 + z^2 s_i s_j}. \tag{2.48}$$

There are two aspects to highlight. First, the model incorporates the heterogeneity across the nodes. There is a natural positive correlation between the GDP and the degree of the nodes. The higher a node's GDP the more edges it will have in expectation, which is in line with economic theory. Second, the calibration of the model has been reduced from n parameters to just one, namely z. Commonly in fitness models, z is (numerically) calibrated to a desired density or equivalently to a desired number of edges $L \in (0, n\,(n-1)\,/2)$,

$$L = \sum_{i<j} P\left(A_{ij} = 1\right) = \sum_{i<j} \frac{z^2 s_i s_j}{1 + z^2 s_i s_j}. \tag{2.49}$$

Furthermore, in Garlaschelli and Loffredo (2004a, 2008) the authors report that this model reproduces several network statistics remarkable well, including higher order properties like the average nearest neighbor degree and the clustering coefficient.[2]

A natural and straightforward extension of the undirected fitness model discussed above, is the directed fitness model. In this case, the in- and out-strength of the nodes are commonly used as fitness variables for two reasons. First, in most use cases data on these factors is readily available. For example regarding the interbank market, interbank assets and deposits are publicly listed in the banks' balance sheets. Second, the degree and the strength of nodes are highly correlated, see e.g. Roukny et al. (2014), and hence a natural positive correlation between the fitness variables and the resulting expected degrees is induced. Commonly in the directed fitness model, the link probability function corresponds to that of an ERGM conditioned on the in- and out-degree sequence. Let $A \in \{0,1\}^{n \times n}$ denote a random directed graph matrix, $s^{(\text{out})} = \left(s_1^{(\text{out})}, \dots, s_n^{(\text{out})}\right) \in \mathbb{R}^n_{\geq 0}$ the out-strength of each node and analogously $s^{(\text{in})} = \left(s_1^{(\text{in})}, \dots, s_n^{(\text{in})}\right) \in \mathbb{R}^n_{\geq 0}$ the in-strength, then the link probabilities are defined by

$$P\left(A_{ij} = 1\right) = \frac{z^2 s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z^2 s_i^{(\text{out})} s_j^{(\text{in})}}. \tag{2.50}$$

The only parameter z left to be calibrated is usually chosen such that a desired number of edges $L^{\rightarrow} \in (0, n(n-1))$ is achieved in expectation,

$$L^{\rightarrow} = \sum_{i \neq j} P\left(A_{ij} = 1\right) = \sum_{i \neq j} \frac{z^2 s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z^2 s_i^{(\text{out})} s_j^{(\text{in})}}. \tag{2.51}$$

Moreover, in case the number of links is not known for the entire network, but only for a subset of nodes $I \subseteq \{1, \dots, n\}$, Cimini et al. (2015) suggest to calibrate the parameter z accordingly. Let $L_I^{\rightarrow}$ denote the number of links of the subset I of nodes, then z is calibrated such that

$$\begin{aligned} L_I^{\rightarrow} &= \sum_{i \in I} \sum_{\substack{j=1 \\ j \neq i}}^{n} P\left(A_{ij} = 1\right) + P\left(A_{ji} = 1\right) \\ &= \sum_{i \in I} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{z^2 s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z^2 s_i^{(\text{out})} s_j^{(\text{in})}} + \frac{z^2 s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z^2 s_j^{(\text{out})} s_i^{(\text{in})}} \end{aligned} \tag{2.52}$$

is satisfied. This model has been successfully used to reconstruct the WTW and the e-MID, see Cimini et al. (2015). For clarification, the e-MID is an electronic market for interbank deposits in Europe.

[2] A precise definition of all network statistics used throughout this thesis is provided in Appendix A.

So far, we introduced fitness models that offer a convenient solution for the construction of unweighted graphs in light of limited available data, especially when precise information on the degrees is missing. In a subsequent step, weights can be assigned to the unweighted fitness models according to the strength of the nodes. Several approaches have been proposed and are outlined in the following. We note that although these approaches have been developed for fitness models, they are rather general and can serve for allocating weights to other unweighted network models as well.

Cimini et al. (2015) suggest to use the degrees inferred by a fitness model to construct an ERGM subject to these degrees and the strength. More precisely, consider an undirected random graph $A \in \{0,1\}^{n\times n}$, with $A = A^\top$, a given strength sequence $s = (s_1, \dots, s_n) \in \mathbb{R}^n_{>0}$ and let $d = (d_1, \dots, d_n) \in \big(0, n\,(n-1)\big)^n$ denote the degree sequence deduced from the fitness model,

$$d_i = \sum_{\substack{j=1\\ j\neq i}}^{n} P\left(A_{ij} = 1\right) = \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{z^2 s_i s_j}{1 + z^2 s_i s_j}. \tag{2.53}$$

Deriving the ERGM conditioned on the degree sequence d and the strength sequence s leads to the following link probabilities and expected link weights. Let P^* denote the probability distribution of the ERGM, the parameters $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ the Lagrange multipliers corresponding to the degree constraints, and the parameters $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n$ the Lagrange multipliers corresponding to the strength constraints,

$$P^*\left(W_{ij} > 0\right) = \frac{e^{-\lambda_i - \lambda_j - \theta_i - \theta_j}}{1 - e^{-\theta_i - \theta_j} + e^{-\lambda_i - \lambda_j - \theta_i - \theta_j}}, \tag{2.54}$$

$$\mathbb{E}\left[W_{ij}\right] = \frac{e^{-\lambda_i - \lambda_j - \theta_i - \theta_j}}{\left(1 - e^{-\theta_i - \theta_j}\right)\left(1 - e^{-\theta_i - \theta_j} + e^{-\lambda_i - \lambda_j - \theta_i - \theta_j}\right)}, \tag{2.55}$$

for all $i \neq j$ and where W_{ij} denotes the random weighted link between nodes i and j. The Lagrange multipliers can be calibrated such that for each node the desired degree and strength are satisfied in expectation, i.e. for all $i = 1, \dots, n$,

$$\begin{aligned} d_i &= \sum_{\substack{j=1\\ j\neq i}}^{n} P^*\left(W_{ij} > 0\right), \\ s_i &= \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[W_{ij}\right]. \end{aligned} \tag{2.56}$$

Cimini et al. (2015) demonstrate the potential of this model by reconstructing two real world networks, the WTW and the e-MID. Both networks are reproduced remarkably well, in terms of unweighted as well as weighted network characteristics, such as the average nearest neighbors degree and strength, and the binary and weighted clustering coefficient. As already stated above for the WTW in Eq. (2.47), the reason why the empirical fitness model works well is because of the linear relationship of the exponential function of the negative Lagrange multipliers and the nodes' strengths, which the authors also find for the e-MID network.

Another option to subsequently allocate weights to a network model of unweighted graphs is given by the adjusted gravity model, presented in Cimini et al. (2015). Let the network model for unweighted directed graphs be described by the random binary graph matrix $A \in \{0,1\}^{n\times n}$ and the probability distribution P defining the link probabilities. To ensure that the given in- and out-strength of each node are met in expectation, Cimini et al. (2015) suggest to allocate the following weights $w \in \mathbb{R}_{\geq 0}^{n\times n}$ to a realization a of A,

$$w_{ij} = \frac{s_i^{(\text{out})} s_j^{(\text{in})}}{s^{(\text{total})} P\left(A_{ij} = 1\right)} a_{ij}, \tag{2.57}$$

where $s^{(\text{total})}$ denotes the total weight of the network, i.e. $s^{(\text{total})} = \sum_{i=1}^{n} s_i^{(\text{out})} = \sum_{i=1}^{n} s_i^{(\text{in})}$. It is straightforward to prove that this model satisfies the in- and out-strength of each node in expectation. In combination with the directed fitness model, see Eq. (2.50), Cimini et al. (2015) call this model the “degree-corrected gravity model” and demonstrate that this approach reconstructs the in- and out-strengths of the WTW and the e-MID remarkably well.

A fundamentally different approach for weight allocation is provided by Gandy and Veraart (2017a), who propose exponentially distributed weights, i.e. for all $i, j = 1, \ldots, n$ and $i \neq j$,

$$W_{ij} | W_{ij} > 0 \sim Exp\left(\alpha_{ij}\right). \tag{2.58}$$

Moreover, to generate networks from this distribution Gandy and Veraart (2017a) developed a Markov Chain Monte Carlo (MCMC) sampler. By iteratively selecting cycles of different length across the network matrix and shifting weight along the chosen cycle, the MCMC sampler creates a sequence of weighted graphs that all match the desired row and column sums exactly, and that converges to the exponential distributions of Eq. (2.58). In Gandy and Veraart (2017b), the authors demonstrate the remarkable performance of their model, consisting of an undirected fitness model as described in Eq. (2.50) combined with the MCMC sampler, by reconstructing a number of credit default swap exposure networks.

2.2.2 Randomized Fitness Variables

A second strand of literature takes a more analytic view on fitness models by randomizing the fitness variables. Following Caldarelli et al. (2002); Servedio et al. (2004); Gandy and Veraart (2017a), we start by introducing the general concepts. Subsequently, the degree distribution is fitted to follow a power law, since this class of distributions appears ubiquitously in real world quantities. For particular choices of the link probability function and the distribution of the random fitness variables, closed form expressions for the distributions of interest can then be derived.

Consider an undirected (and unweighted) network with $n \in \mathbb{N}$ nodes, each equipped with a non-negative random fitness variable $X_i \in \mathbb{R}_{\geq 0}$, for $i = 1, \ldots, n$. Moreover, let the random fitness variables X_i be i.i.d. according to some probability density function ρ_X. The probability that a link exists between two nodes i and j depends on the realized fitness variables x_i and x_j and is defined by a function $f : \mathbb{R}^2_{\geq 0} \to [0, 1]$, hence, the link probability is given by $f(x_i, x_j)$. We shortly remark, that the case of undirected networks naturally implies that f should be symmetric in its arguments, i.e. we want $f(x_i, x_j) = f(x_j, x_i)$ to hold.

Next, we define the function $d : \mathbb{R} \to (0, n-1)$ which computes the expected degree of a node. For a node i with realized fitness x, we have

$$\begin{aligned} d(x) &:= \mathbb{E}\left[\sum_{\substack{j=1\\ j\neq i}}^{n} f(x, X_j)\right] \\ &= (n-1)\,\mathbb{E}\left[f(x, X)\right], \quad \text{since } X_j \text{ are i.i.d. } \forall j \neq i \\ &= (n-1)\int_0^{\infty} f(x, z)\,\rho_X(z)\,\mathrm{d}z. \end{aligned} \tag{2.59}$$

Moreover, we define the random variable $Y := d(X)$ as the expected degree of a node with random fitness variable X. For the function d being continuous and strictly monotonic, i.e. the inverse d^{-1} exists, the density of ρ_Y of Y can be derived as described in the transformation theorem of probability densities. The following analyses and discussed models are restricted to d being strictly monotonically increasing. In this case, we get

$$\begin{aligned} \rho_Y(y) &= \frac{\partial}{\partial y} P(d(X) \leq y) = \frac{\partial}{\partial y} P\left(X \leq d^{-1}(y)\right) \\ &= \frac{\partial}{\partial y} F_X\left(d^{-1}(y)\right) = \rho_X\left(d^{-1}(y)\right)\frac{\partial}{\partial y} d^{-1}(y), \end{aligned} \tag{2.60}$$

where F_X denotes the cumulative distribution function of X. In addition, we know from the inverse function theorem that the derivative of $d^{-1}(y)$ is given by $1/d'(x)$, and hence,

$$\rho_Y(y) = \frac{\rho_X(x)}{d'(x)}. \tag{2.61}$$

The density function ρ_Y denotes the degree distribution of the networks generated by the random fitness variables and the link probability function. Since many real world networks have been found to exhibit degree distributions that follow a power law, fitting ρ_Y to power law distributions is of particular interest. Power law degree distributions describe networks where a small number of 'big' nodes maintains many links, while most nodes are 'small' and maintain only few links. In fact, power laws are known to appear ubiquitously in real world quantities, network examples include the Internet, the connectivity of cities and the interbank market, see Clauset et al. (2009). Therefore, we are interested in fitting the degree distribution $\rho_Y(y)$ to follow a power law with exponent $\alpha \in \mathbb{R}_{<0}$ on the finite range $[d_0, d_\infty] \subseteq (0, n-1]$, where $d_s := \lim_{x\to s} d(x)$. Note that for d strictly monotonically increasing, it follows that $d_0 < d_\infty$. Hence, recalling $y = d(x)$, we want

$$\begin{aligned} \rho_Y(y) &= cy^\alpha \\ \Leftrightarrow \quad \rho_Y(d(x)) &= cd(x)^\alpha, \end{aligned} \tag{2.62}$$

to hold. The parameter c is a constant ensuring the normalization condition $\int_{d_0}^{d_\infty} cy^\alpha \mathrm{d}y = 1$. For $\alpha = -1$,

$$\begin{aligned} \int_{d_0}^{d_\infty} cy^{-1}\mathrm{d}y &= 1 \\ \Leftrightarrow \quad c\left[\log(d_\infty) - \log(d_0)\right] &= 1 \\ \Leftrightarrow \quad c &= \left(\log\left(\frac{d_\infty}{d_0}\right)\right)^{-1}. \end{aligned} \tag{2.63}$$

For $\alpha \neq -1$,

$$\begin{aligned} \int_{d_0}^{d_\infty} cy^{-\alpha}\mathrm{d}y &= 1 \\ \Leftrightarrow \quad \frac{c}{\alpha+1}\left[d_\infty^{\alpha+1} - d_0^{\alpha+1}\right] &= 1 \\ \Leftrightarrow \quad c &= \frac{\alpha+1}{d_\infty^{\alpha+1} - d_0^{\alpha+1}}. \end{aligned} \tag{2.64}$$

Note that since $d_0 < d_\infty$ holds, c is positive for all $\alpha \in \mathbb{R}_{<0}$.

Combining Eq. (2.61) and Eq. (2.62) yields

$$\begin{aligned} \frac{\rho_X(x)}{d'(x)} &= c(d(x))^\alpha \\ \Leftrightarrow \quad \rho_X(x) &= c(d(x))^\alpha d'(x). \end{aligned} \tag{2.65}$$

Integrating both sides of Eq. (2.65) from 0 to x gives

$$\begin{aligned} R(x) := \int_0^x \rho_X(z)\,\mathrm{d}z &= \int_0^x c(d(z))^\alpha d'(z)\,\mathrm{d}z \\ &= \begin{cases} c\left(\log(d(x)) - \log(d_0)\right), & \text{if } \alpha = -1 \\ \frac{c}{\alpha+1}\left(d(x)^{\alpha+1} - d_0^{\alpha+1}\right), & \text{if } \alpha \neq -1. \end{cases} \end{aligned} \tag{2.66}$$

Solving for $d(x)$ leads to

$$d(x) = \begin{cases} \exp\left[\frac{1}{c} R(x)\right] d_0, & \text{if } \alpha = -1 \\ \left[\frac{\alpha+1}{c} R(x) + d_0^{\alpha+1}\right]^{\frac{1}{\alpha+1}}, & \text{if } \alpha \neq -1. \end{cases} \tag{2.67}$$

For certain forms of the link probability function f and the distribution ρ_X of the fitness variables, closed form expressions for the quantities of interest can be derived. In the following, we present the examples studied in Caldarelli et al. (2002); Servedio et al. (2004); Gandy and Veraart (2017a).

A natural first choice for the link probability function f is to consider the product of possibly transformed fitness variables, i.e.

$$f(x_i, x_j) = g(x_i) h(x_j). \tag{2.68}$$

As already pointed out above, in the case of undirected graphs, we want f to be symmetric in its arguments, i.e. $f(x_i, x_j) = f(x_j, x_i)$ should hold. For that reason, we set $g(x) \equiv h(x)$.

Theorem 2.2.1 (Fitness model with $f(x_i, x_j) = g(x_i) g(x_j)$, arbitrary ρ_X, and power law degree distribution with exponent $\alpha = -1$)
For any arbitrary probability distribution ρ_X of the fitness variables X, there exists a function g, such that the link probability function $f(x_i, x_j) = g(x_i) g(x_j)$ generates networks with degrees that are power law distributed with exponent (-1) on the finite range $[d_0, d_\infty]$, i.e. $\rho_Y(d(x)) = c d(x)^{-1}$. Moreover, the function g is given by

$$g(x) = \beta \left(\frac{\gamma}{\beta}\right)^{R(x)} \left[\frac{\log(\gamma) - \log(\beta)}{\gamma - \beta}\right]^{1/2}, \tag{2.69}$$

where $\beta = d_0/(n-1)$ and $\gamma = d_\infty/(n-1)$. The domain of the power law distribution, i.e. $0 < \beta < \gamma \leq 1$, can be chosen arbitrarily within the following constraint

$$\gamma^2 \frac{\log(\gamma) - \log(\beta)}{\gamma - \beta} \leq 1. \tag{2.70}$$

Proof
For a node with realized fitness variable x, we get the following expected degree (i.e. by inserting f in Eq. (2.59)),

$$\begin{aligned} d(x) &= (n-1) \int_0^\infty f(x,z) \rho_X(z) \, \mathrm{d}z = (n-1) g(x) \int_0^\infty g(z) \rho_X(z) \, \mathrm{d}z \\ &= (n-1) g(x) \mathbb{E}[g(Z)]. \end{aligned} \tag{2.71}$$

Combining Eq. (2.71) and Eq. (2.67) for $\alpha = -1$, gives

$$(n-1)\, g(x)\, \mathbb{E}\left[g(Z)\right] = \exp\left[\frac{1}{c} R(x)\right] d_0. \tag{2.72}$$

Computing the expected value of both sides of Eq. (2.72) with respect to the random fitness variable X, and recalling $R(x) = \int_0^x \rho_X(z)\,\mathrm{d}z$, yields

$$\begin{aligned}
\int_0^\infty (n-1)\, g(x)\, \mathbb{E}\left[g(Z)\right] \rho_X(x)\,\mathrm{d}x &= \int_0^\infty \exp\left[\frac{1}{c} R(x)\right] d_0 \rho_X(x)\,\mathrm{d}x \\
\Leftrightarrow \qquad (n-1)\, \mathbb{E}\left[g(Z)\right] \mathbb{E}\left[g(X)\right] &= d_0 \left[c \exp\left[\frac{1}{c} R(x)\right]\right]_0^\infty \\
\Leftrightarrow \qquad (n-1) \left(\mathbb{E}\left[g(X)\right]\right)^2 &= d_0 c \left[\exp\left(\frac{1}{c}\right) - 1\right] \\
\Leftrightarrow \qquad \mathbb{E}\left[g(X)\right] &= \left[\frac{d_0 c}{n-1}\left(e^{1/c} - 1\right)\right]^{1/2}.
\end{aligned} \tag{2.73}$$

Substituting c from Eq. (2.63), gives

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] &= \left[\frac{d_0}{n-1}\left(\log\left(\frac{d_\infty}{d_0}\right)\right)^{-1}\left(\exp\left(\log\left(\frac{d_\infty}{d_0}\right)\right) - 1\right)\right]^{1/2} \\
&= \left[\frac{d_0}{n-1}\left(\log\left(\frac{d_\infty}{d_0}\right)\right)^{-1}\left(\frac{d_\infty}{d_0} - 1\right)\right]^{1/2} \\
&= \left[\left(\log\left(\frac{d_\infty}{d_0}\right)\right)^{-1} \frac{d_\infty - d_0}{n-1}\right]^{1/2}.
\end{aligned} \tag{2.74}$$

Moreover, from Eq. (2.59) we see that $d_0 = \beta(n-1)$ and $d_\infty = \gamma(n-1)$ holds for some $\beta, \gamma \in (0,1]$ and $\beta < \gamma$. This reformulation is convenient as it allows to simplify equations.

Equation (2.74) can be rewritten to

$$\mathbb{E}\left[g(X)\right] = \left[\left(\log\left(\frac{\gamma}{\beta}\right)\right)^{-1}(\gamma - \beta)\right]^{1/2} = \left[\frac{\gamma - \beta}{\log(\gamma) - \log(\beta)}\right]^{1/2}. \tag{2.75}$$

By inserting the expected value of $g(X)$ from Eq. (2.75) and c from Eq. (2.63) into

Eq. (2.72), we can derive the functional form of g,

$$\begin{aligned}
& (n-1)\, g(x)\, \mathbb{E}\left[g(Z)\right] = \exp\left[\frac{1}{c} R(x)\right] d_0 \\
\Leftrightarrow \quad & g(x) = \frac{d_0}{n-1} \exp\left[\frac{1}{c} R(x)\right] \mathbb{E}\left[g(Z)\right]^{-1} \\
\Leftrightarrow \quad & g(x) = \frac{d_0}{n-1} \exp\left[\log\left(\frac{d_\infty}{d_0}\right) R(x)\right] \left[\frac{\gamma-\beta}{\log(\gamma)-\log(\beta)}\right]^{-1/2} \\
\Leftrightarrow \quad & g(x) = \beta \left(\frac{\gamma}{\beta}\right)^{R(x)} \left[\frac{\log(\gamma)-\log(\beta)}{\gamma-\beta}\right]^{1/2}.
\end{aligned} \tag{2.76}$$

Note that $g(x) > 0$ holds for all $x \in \mathbb{R}_{\geq 0}$ and g is strictly monotonically increasing in x since $\beta < \gamma$. Since g defines the link probability function $f(x_i, x_j) = g(x_i)\, g(x_j) \in [0, 1]$, we also require that $g(x) \leq 1$ has to hold. Because of the monotonicity, it suffices to ensure that $\lim_{x\to\infty} g(x) \leq 1$ holds,

$$\begin{aligned}
& \lim_{x\to\infty} g(x) \leq 1 \\
\Leftrightarrow \quad & \lim_{x\to\infty} \beta \left(\frac{\gamma}{\beta}\right)^{R(x)} \left[\frac{\log(\gamma)-\log(\beta)}{\gamma-\beta}\right]^{1/2} \leq 1 \\
\Leftrightarrow \quad & \beta \left(\frac{\gamma}{\beta}\right) \left[\frac{\log(\gamma)-\log(\beta)}{\gamma-\beta}\right]^{1/2} \leq 1 \\
\Leftrightarrow \quad & \gamma^2 \frac{\log(\gamma)-\log(\beta)}{\gamma-\beta} \leq 1.
\end{aligned} \tag{2.77}$$

Hence, requiring $g(x) \leq 1$ induces a dependence between β and γ. In other words, we can choose $d_0 \in (0, n-1)$ (or $d_\infty \in (0, n-1)$) freely, while Eq. (2.77) defines a bound for d_∞ (respectively d_0). □

Remark 2.2.2 (Fitnes model with $f(x_i, x_j) = g(x_i)\, g(x_j)$, arbitrary ρ_X, and power law degree distribution with arbitrary exponent α)
Servedio et al. (2004) state that for arbitrary exponents α of the power law distribution, the necessary calculations can be derived analogously. Hence, for any distribution ρ_X of the fitness variables X and a power law distribution cy^α, there exists a function g such that the degrees of the generated networks follow the chosen power law distribution.

While Theorem 2.2.1 allows for arbitrary distributions ρ_X of the random fitness variables X and derives the implied functional form of g, we can also allow for arbitrary functions g and derive the induced distribution ρ_X, as the following theorem shows.

Theorem 2.2.3 (Fitness model with $f(x_i, x_j) = g(x_i)\, g(x_j)$, with g monotonically increasing, and power law degree distribution with arbitrary exponent α)
For any monotonically increasing function $g : \mathbb{R}_{\geq 0} \to (0, 1]$, there exists a probability density function ρ_X for the random fitness variables, such that the link probability function $f(x_i, x_j) = g(x_i)\, g(x_j)$ generates networks with degrees that are power law distributed with exponent α on the finite range $[d_0, d_\infty]$, i.e. $\rho_Y(d(x)) = cd(x)^\alpha$ holds. Moreover, the probability density function ρ_X is given by

$$\rho_X(x) = c\,(n-1)^{\alpha+1}\, \mathbb{E}\left[g(X)\right]^{\alpha+1} g(x)^\alpha\, g'(x)\,. \tag{2.78}$$

The domain of the power law distribution is such that $d_0 \in (0, n-1)$ (or $d_\infty \in (0, n-1)$) can be chosen arbitrarily, while d_∞ (respectively d_0) is defined by

$$\begin{aligned} \frac{d_\infty}{d_0} &= \frac{g_\infty}{g_0}, && \text{for } \alpha = -1, \\ \frac{d_\infty^{\alpha+1} - d_0^{\alpha+1}}{g_\infty^{\alpha+1} - g_0^{\alpha+1}} &= (n-1)^{\alpha+1}\, \mathbb{E}\left[g(Z)\right]^{\alpha+1}, && \text{for } \alpha \neq -1, \end{aligned} \tag{2.79}$$

where $g_s := \lim_{x \to s} g(s)$.

Proof
The implied form of the density function ρ_X can be derived by inserting Eq. (2.71) in Eq. (2.65),

$$\begin{aligned} \rho_X(x) &= cd(x)^\alpha\, d'(x) \\ &= c\Big[(n-1)\, g(x)\, \mathbb{E}\left[g(Z)\right]\Big]^\alpha (n-1)\, g'(x)\, \mathbb{E}\left[g(Z)\right] \\ &= c\,(n-1)^{\alpha+1}\, \mathbb{E}\left[g(Z)\right]^{\alpha+1} g(x)^\alpha\, g'(x)\,. \end{aligned} \tag{2.80}$$

We note that in order for ρ_X being well defined, g has to be monotonically increasing in x and $g(x) > 0$ has to hold for all $x \in \mathbb{R}_{\geq 0}$. Furthermore, because of the link probability function f, also $g(x) \leq 1$ has to be satisfied for all x.

To ensure that ρ_X truly constitutes a probability density function, we analyze the normalization condition. For $\alpha = -1$,

$$
\begin{aligned}
& & 1 &= \int_0^\infty \rho_X(x)\,\mathrm{d}x \\
&\Leftrightarrow & 1 &= \int_0^\infty c\,(n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1} g(x)^\alpha g'(x)\,\mathrm{d}x \\
&\Leftrightarrow & 1 &= \int_0^\infty c g(x)^{-1} g'(x)\,\mathrm{d}x \\
&\Leftrightarrow & 1 &= c\Big[\log\big(g(x)\big)\Big]_0^\infty \\
&\Leftrightarrow & 1 &= c\log\left(\frac{g_\infty}{g_0}\right) \\
&\Leftrightarrow & 1 &= \left(\log\left(\frac{d_\infty}{d_0}\right)\right)^{-1}\log\left(\frac{g_\infty}{g_0}\right), \quad \text{by Eq. (2.63)} \\
&\Leftrightarrow & \frac{d_\infty}{d_0} &= \frac{g_\infty}{g_0},
\end{aligned}
\tag{2.81}
$$

where $g_s := \lim_{x\to s} g(x)$. Hence, we can either choose the lower or the upper bound of the domain of the power law distribution, while the other bound is defined by Eq. (2.81).

For $\alpha \neq -1$,

$$
\begin{aligned}
& & 1 &= \int_0^\infty \rho_X(x)\,\mathrm{d}x \\
&\Leftrightarrow & 1 &= \int_0^\infty c\,(n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1} g(x)^\alpha g'(x)\,\mathrm{d}x \\
&\Leftrightarrow & 1 &= c\,(n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1}\frac{1}{\alpha+1}\left[g(x)^{\alpha+1}\right]_0^\infty \\
&\Leftrightarrow & 1 &= c\,(n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1}\frac{1}{\alpha+1}\left[g_\infty^{\alpha+1} - g_0^{\alpha+1}\right] \\
&\Leftrightarrow & 1 &= \frac{\alpha+1}{d_\infty^{\alpha+1} - d_0^{\alpha+1}}\,(n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1}\frac{1}{\alpha+1}\left[g_\infty^{\alpha+1} - g_0^{\alpha+1}\right], \\
& & & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by Eq. (2.64)} \\
&\Leftrightarrow & \frac{d_\infty^{\alpha+1} - d_0^{\alpha+1}}{g_\infty^{\alpha+1} - g_0^{\alpha+1}} &= (n-1)^{\alpha+1}\,\mathbb{E}\left[g(Z)\right]^{\alpha+1}.
\end{aligned}
\tag{2.82}
$$

Hence, we can either choose the lower or the upper bound for the domain of the power law distribution, while the other bound is defined by Eq. (2.82). □

A more difficult form for the link probability function is a summation of the fitness variables $f(x_i, x_j) = \tilde{f}(x_i + x_j)$. Nevertheless, this case can still be solved for exponentially distributed random fitness variables.

Theorem 2.2.4 (Fitness model with $f(x_i, x_j) = \tilde{f}(x_i + x_j)$, $X \sim Exp(1)$, and power law degree distribution with arbitrary exponent α)
Let the random fitness variables be exponentially distributed with rate parameter 1, i.e. $X \sim Exp(1)$. There exists a link probability function $f(x_i, x_j) = \tilde{f}(x_i + x_j)$, that generates networks with degrees that are power law distributed with exponent α on the finite range $[d_0, d_\infty]$, i.e. such that $\rho_Y(d(x)) = cd(x)^\alpha$ holds. Moreover, the link probability function f is given by

$$f(x_i, x_j) = \frac{d(x_i + x_j) - d'(x_i + x_j)}{n-1}. \tag{2.83}$$

The domain of the power law distribution can be chosen arbitrarily for $\alpha \leq -2$. For $\alpha \in (-2, 0)$, the domain is subject to the constraint

$$\begin{aligned} \frac{d_\infty}{d_0} &\leq e, && \text{for } \alpha = -1, \\ \frac{d_\infty}{d_0} &\leq (\alpha+2)^{\frac{1}{\alpha+1}}, && \text{for } \alpha \in (-2,-1) \cup (-1, 0). \end{aligned} \tag{2.84}$$

Proof
For a node with realized fitness variable x, we get the following expected degree (i.e. by inserting f in Eq. (2.59)),

$$d(x) = (n-1) \int_0^\infty \tilde{f}(x+z)\, \rho_X(z)\, \mathrm{d}z, \tag{2.85}$$

substituting $u = x + z$,

$$d(x) = (n-1) \int_x^\infty \tilde{f}(u)\, \rho_X(u - x)\, \mathrm{d}u. \tag{2.86}$$

For exponentially distributed fitness variables $X \sim Exp(1)$, we get,

$$\begin{aligned} d(x) &= (n-1) \int_x^\infty \tilde{f}(u)\, e^{-(u-x)} \mathrm{d}u \\ &= (n-1)\, e^x \int_x^\infty \tilde{f}(u)\, e^{-u} \mathrm{d}u \\ \Leftrightarrow \quad \frac{1}{n-1} e^{-x} d(x) &= \int_x^\infty \tilde{f}(u)\, e^{-u} \mathrm{d}u. \end{aligned} \tag{2.87}$$

Differentiating both sides w.r.t. x gives

$$\begin{aligned} \frac{1}{n-1} \left[-d(x)\, e^{-x} + d'(x)\, e^{-x} \right] &= -\tilde{f}(x)\, e^{-x} \\ \Leftrightarrow \quad \frac{d(x) - d'(x)}{n-1} &= \tilde{f}(x). \end{aligned} \tag{2.88}$$

Hence, the link probability function is given by

$$f\left(x_i, x_j\right) = \tilde{f}\left(x_i + x_j\right) = \frac{d\left(x_i + x_j\right) - d'\left(x_i + x_j\right)}{n-1}. \tag{2.89}$$

Note that $d(x)$ (and hence also $d'(x)$) is completely defined by Eq. (2.67), since in this example the fitness variables are set to be exponentially distributed, i.e. $R(x) = \int_0^x \rho_X(z)\,\mathrm{d}z = \int_0^x e^{-z}\mathrm{d}z = 1 - e^{-x}$.

Since f determines the link probability, we have to ensure that $f\left(x_i, x_j\right) \in [0,1]$ for all $x_i, x_j \in \mathbb{R}$. Recalling $d(x)$ from Eq. (2.67)

$$d(x) = \begin{cases} \exp\left[\frac{1}{c}R(x)\right] d_0, & \text{if } \alpha = -1, \\ \left[\frac{\alpha+1}{c}R(x) + d_0^{\alpha+1}\right]^{\frac{1}{\alpha+1}}, & \text{if } \alpha \neq -1, \end{cases} \tag{2.90}$$

where $R(x) = 1 - e^{-x}$ and hence $\frac{\partial}{\partial x}R(x) = e^{-x}$. The derivative of d is given by

$$\begin{aligned} d'(x) &= \begin{cases} d(x)\,\frac{1}{c}e^{-x}, & \text{if } \alpha = -1, \\ \frac{1}{\alpha+1}\left[\frac{\alpha+1}{c}R(x) + d_0^{\alpha+1}\right]^{\frac{-\alpha}{\alpha+1}} \frac{\alpha+1}{c}e^{-x}, & \text{if } \alpha \neq -1, \end{cases} \\ &= \begin{cases} d(x)\,\frac{1}{c}e^{-x}, & \text{if } \alpha = -1, \\ d(x)^{-\alpha}\,\frac{1}{c}e^{-x}, & \text{if } \alpha \neq -1, \end{cases} \\ &= \begin{cases} d(x)\log\left(\frac{d_\infty}{d_0}\right)e^{-x}, & \text{if } \alpha = -1, \quad \text{by Eq. (2.63)} \\ d(x)^{-\alpha}\,\frac{d_\infty^{\alpha+1} - d_0^{\alpha+1}}{\alpha+1}e^{-x}, & \text{if } \alpha \neq -1, \quad \text{by Eq. (2.64).} \end{cases} \end{aligned} \tag{2.91}$$

Hence for $d_0 < d_\infty$ it follows $d'(x) > 0$, i.e. $d(x)$ is strictly monotonically increasing.

For $\alpha = -1$, the lower bound of $\tilde{f}$ has to hold for all $x \in \mathbb{R}_{>0}$,

$$\begin{aligned} 0 \leq \tilde{f}(x) \quad &\Leftrightarrow \quad 0 \leq \frac{d(x) - d(x)\,\frac{1}{c}e^{-x}}{n-1} \\ &\Leftrightarrow \quad 0 \leq 1 - \frac{1}{c}e^{-x}. \end{aligned} \tag{2.92}$$

Note that the right hand side is increasing in x. Therefore, it suffices to consider the lower bound of the support of x,

$$\begin{aligned} 0 \leq \tilde{f}(x) \quad &\Leftrightarrow \quad 0 \leq \lim_{x \to 0}\left[1 - \frac{1}{c}e^{-x}\right] \\ &\Leftrightarrow \quad 0 \leq 1 - \log\left(\frac{d_\infty}{d_0}\right) \\ &\Leftrightarrow \quad \frac{d_\infty}{d_0} \leq e. \end{aligned} \tag{2.93}$$

For the upper bound of $\tilde{f}$, we get,

$$\begin{aligned} 1 \geq \tilde{f}(x) \quad &\Leftrightarrow \quad 1 \geq \frac{d(x) - d(x)\frac{1}{c}e^{-x}}{n-1} \\ &\Leftrightarrow \quad n-1 \geq d(x)\left[1 - \frac{1}{c}e^{-x}\right]. \end{aligned} \tag{2.94}$$

Note that the right hand side is increasing in x. Therefore, it suffices to consider the upper bound of the support of x,

$$\begin{aligned} 1 \geq \tilde{f}(x) \quad &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} d(x)\left[1 - \frac{1}{c}e^{-x}\right] \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} d(x) \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} \exp\left[\frac{1}{c}R(x)\right] d_0 \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} \exp\left[\log\left(\frac{d_\infty}{d_0}\right)\left(1 - e^{-x}\right)\right] d_0 \\ &\Leftrightarrow \quad n-1 \geq d_\infty, \end{aligned} \tag{2.95}$$

which is satisfied by the definition of d_∞.

For $\alpha \neq -1$, the lower bound of $\tilde{f}$ has to hold for all $x \in \mathbb{R}_{>0}$,

$$\begin{aligned} 0 \leq \tilde{f}(x) \quad &\Leftrightarrow \quad 0 \leq \frac{d(x) - d(x)^{-\alpha}\frac{1}{c}e^{-x}}{n-1} \\ &\Leftrightarrow \quad 0 \leq d(x) - d(x)^{-\alpha}\frac{1}{c}e^{-x} \\ &\Leftrightarrow \quad 0 \leq 1 - d(x)^{-\alpha-1}\frac{1}{c}e^{-x} \\ &\Leftrightarrow \quad 0 \leq 1 - \left[\frac{\alpha+1}{c}R(x) + d_0^{\alpha+1}\right]^{-1}\frac{1}{c}e^{-x} \\ &\Leftrightarrow \quad 0 \leq 1 - \left[\left(d_\infty^{\alpha+1} - d_0^{\alpha+1}\right)\left(1 - e^{-x}\right) + d_0^{\alpha+1}\right]^{-1}\frac{1}{c}e^{-x} \\ &\Leftrightarrow \quad 0 \leq 1 - \left[\left(d_\infty^{\alpha+1} - d_0^{\alpha+1}\right)\left(e^{x} - 1\right) + d_0^{\alpha+1}e^{x}\right]^{-1}\frac{1}{c} \\ &\Leftrightarrow \quad 0 \leq 1 - \left[d_\infty^{\alpha+1}\left(e^{x} - 1\right) + d_0^{\alpha+1}\right]^{-1}\frac{1}{c}. \end{aligned} \tag{2.96}$$

Note that the right hand side is increasing in x. Therefore, it suffices to consider the lower bound of the support of x,

$$\begin{aligned} 0 \leq \tilde{f}(x) \quad &\Leftrightarrow \quad 0 \leq \lim_{x \to 0} \left[1 - \left[d_\infty^{\alpha+1} (e^x - 1) + d_0^{\alpha+1} \right]^{-1} \frac{1}{c} \right] \\ &\Leftrightarrow \quad 0 \leq 1 - d_0^{-\alpha-1} \frac{d_\infty^{\alpha+1} - d_0^{\alpha+1}}{\alpha+1} \\ &\Leftrightarrow \quad 1 \geq \frac{d_\infty^{\alpha+1} - d_0^{\alpha+1}}{d_0^{\alpha+1}} \frac{1}{\alpha+1} \\ &\Leftrightarrow \quad 1 \geq \left(\left(\frac{d_\infty}{d_0} \right)^{\alpha+1} - 1 \right) \frac{1}{\alpha+1} \quad \Leftrightarrow \quad (\star). \end{aligned} \tag{2.97}$$

For $\alpha \in (-1, 0)$,

$$\begin{aligned} (\star) \quad \Leftrightarrow \quad (\alpha+2) &\geq \left(\frac{d_\infty}{d_0} \right)^{\alpha+1}, \\ \Leftrightarrow \quad (\alpha+2)^{\frac{1}{\alpha+1}} &\geq \frac{d_\infty}{d_0}. \end{aligned} \tag{2.98}$$

For $\alpha \in (-2, -1)$,

$$\begin{aligned} (\star) \quad \Leftrightarrow \quad (\alpha+2) &\leq \left(\frac{d_\infty}{d_0} \right)^{\alpha+1}, \\ \Leftrightarrow \quad (\alpha+2)^{\frac{1}{\alpha+1}} &\geq \frac{d_\infty}{d_0}. \end{aligned} \tag{2.99}$$

For $\alpha \in (-\infty, -2]$,

$$(\star) \quad \Leftrightarrow \quad (\alpha+2) \leq \left(\frac{d_\infty}{d_0} \right)^{\alpha+1}, \tag{2.100}$$

which is always satisfied, because the left hand side is negative and the right hand side positive.

For the upper bound of $\tilde{f}$, we get,

$$\begin{aligned} 1 \geq \tilde{f}(x) \quad &\Leftrightarrow \quad 1 \geq \frac{d(x) - d(x)^{-\alpha} \frac{1}{c} e^{-x}}{n-1} \\ &\Leftrightarrow \quad n-1 \geq d(x) \left[1 - d(x)^{-\alpha-1} \frac{1}{c} e^{-x} \right] \\ &\Leftrightarrow \quad n-1 \geq d(x) \left[1 - \left[d_\infty^{\alpha+1} (e^x - 1) + d_0^{\alpha+1} \right]^{-1} \frac{1}{c} \right]. \end{aligned} \tag{2.101}$$

Note that the right hand side is increasing in x. Therefore, it suffices to consider the upper bound of the support of x,

$$\begin{aligned}1 \geq \tilde{f}(x) \quad &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} d(x) \left[1 - \left[d_\infty^{\alpha+1}\left(e^x - 1\right) + d_0^{\alpha+1}\right]^{-1} \frac{1}{c}\right] \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} d(x) \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} \left[\frac{\alpha+1}{c} R(x) + d_0^{\alpha+1}\right]^{\frac{1}{\alpha+1}} \\ &\Leftrightarrow \quad n-1 \geq \lim_{x\to\infty} \left[\left(d_\infty^{\alpha+1} - d_0^{\alpha+1}\right)\left(1 - e^{-x}\right) + d_0^{\alpha+1}\right]^{\frac{1}{\alpha+1}} \\ &\Leftrightarrow \quad n-1 \geq d_\infty,\end{aligned} \tag{2.102}$$

which is satisfied by the definition of d_∞. □

Remark 2.2.5 (Generalization to Directed Graphs)
As Gandy and Veraart (2017a) show for the fitness model of Theorem 2.2.4, it is straightforward to generalize the models presented in this section to directed graphs. In this case we can fit either the in- or the out-degree distribution to follow a power law. In fact, all calculations stay the same, but the link probabilities $f(x_i, x_j)$ are interpreted as directed links from node i to node j.

Note that it is non-trivial to further generalize fitness models to produce both a desired in- and out-degree distribution simultaneously.

2.3 Further Network Reconstruction Techniques

Besides ERGMs and fitness models, there exist further approaches to reconstruct financial networks based on scarce information. Comparisons of some of the proposed methods are provided by the Basel Committee on Banking Supervision (2015b), Mazzarisi and Lillo (2017), and Anand et al. (2018). A general conclusion from these papers is that each model focuses on the adequate reproduction of a few network characteristics such as the density, but falls short of reproducing others. Hence, the adequacy of a method strongly depends on whether the considered network reflects the characteristics under focus. In the following we give an overview over the main approaches.

Maximum entropy (ME) by Elsinger et al. (2013); Upper (2011); Upper and Worm (2004)

A popular approach for reconstructing networks based only on the row and column sums is the ME model. Without further information, a natural simplifying assumption is to consider the margins, i.e., the relative row and column sums, to be independent. This leads to a network $\tilde{w}$ with link weights $\tilde{w}_{ij} = s_i^{(\text{out})} s_j^{(\text{in})} / \sum s_i^{(\text{out})}$, where $s_i^{(\text{out})}$ denotes the sum of row i and $s_j^{(\text{in})}$ the sum of column j. A problem with $\tilde{w}$ is that weights are also allocated on the diagonal, while most real-world networks do not exhibit self-loops. Therefore, the constrained optimization problem of minimizing the Kullback–Leibler divergence to $\tilde{w}$, while satisfying the row and column sums as well as a zero diagonal, is considered and solved via the RAS algorithm.

A drawback of the ME model is that the resulting graphs are almost completely connected. Typical real-world financial networks, however, are sparse. Moreover, it has been shown that the ME model underestimates systemic risk; see, e.g., Anand et al. (2015); Mistrulli (2011). To address these issues, various modified approaches have been developed. For example, Drehmann and Tarashev (2013) consider randomly perturbed ME networks to relax the independence assumption. In addition, Baral and Fique (2012) replace the independence prior by a fitted Gumbel copula, and thus reduce the density.

Minimum and low density by Anand et al. (2015)

In contrast to the ME model, Anand et al. (2015) propose a minimum density model. Their aim is to create as few links as needed to satisfy the given row and column sums. This can again be formulated as a constrained optimization problem, where the objective function is a cost function scaling with the number of links. Furthermore, Anand et al. (2015) extend their model to reconstruct disassortative networks. Disassortativity is a commonly observed feature and means that nodes with few edges tend to be connected to nodes with many edges and vice versa. This characteristic is included through an additional penalty term in the objective function which measures the divergence (relative entropy) to a probability matrix capturing the disassortativity. Anand et al. (2015) also developed a heuristic procedure to generate such networks. Moreover, they report that their model overestimates systemic risk.

Empirical Bayesian methodology by Gandy and Veraart (2017b)

Gandy and Veraart (2017b) propose an empirical Bayesian methodology comprising a fitness model to reconstruct directed and weighted networks, based on the row and column sums plus an additional network statistic such as the density. Moreover, they include the possibility of some links being known. The basic model is split in two steps. First, independent Bernoulli distributed links are drawn: $P\big(W_{ij} > 0\big) = p_{ij}$, where W_{ij} denotes the random variables representing a weighted link from node i to node j. Second, these links are assigned exponentially distributed weights: $W_{ij}|W_{ij} > 0 \sim Exp\big(\lambda_{ij}\big)$. Gandy and Veraart (2017a) have shown that under mild conditions for any choice of $p \in [0,1]^{n\times n}$, there exists a $\lambda \in (0,\infty)^{n\times n}$, such that the row and column sums are fulfilled in expectation. Furthermore, they consider different choices for p and λ. One option is to choose all entries of p and λ to be identical, which leads to an Erdős–Rényi graph. This model can be calibrated to match a desired network statistic in expectation, e.g., the network density.

In order to achieve a core-periphery structure, Gandy and Veraart (2017b) suggest using two constant link probabilities: one for links connecting a large bank with any other bank and one for links connecting two small banks. In addition, hierarchical models are defined, where the basic model is embedded into a larger Bayesian model that randomizes p and λ. Three examples of such models are studied. The first is defined by p_{ij} following a Beta distribution and λ_{ij} following a Gamma distribution. The second is a fitness model with fitness variables $X_i \sim Exp\,(1)$. For this setting, Gandy and Veraart derive a link probability function $p_{ij}\big(X_i, X_j\big)$ such that the resulting model exhibits a power law degree distribution, see Theorem 2.2.4. As a third example, Gandy and Veraart consider the same link probabilities as in Cimini et al. (2015), see Eq. (2.48), but allocate weights based on an MCMC sampler that generates a sequence of matrices fulfilling the row and column sums exactly. In addition, the MCMC sampler allows deriving the distribution of the conditional network model.

Probability map by Hałaj and Kok (2013)

Based on county level exposures of 89 banks, disclosed by the EBA's EU-wide stress test (see http://www.eba.europa.eu), Hałaj and Kok (2013) construct a map of link probabilities by aggregating the banks' exposures by countries. To sample networks, links are drawn successively at random and are kept with the corresponding probability of the probability map. Once a link is kept, its relative weight of unallocated interbank liabilities is sampled uniformly on $[0, 1]$. This process is repeated until all interbank liabilities are allocated.

3 Notes on the ERGMs with Desired In- and Out-Strength Sequence

As pointed out in the introduction to ERGMs, the existence and uniqueness of the solution to general ERGMs has not yet been solved and depends substantially on the chosen constraints. Here, we start filling this gap by analyzing the existence and uniqueness of the solution to the specific ERGM with desired in- and out-strength sequence. The results are of special interest as they (a) provide new insight into the generated graph structure and (b) imply a novel algorithm for calibrating this class of ERGMs.

3.1 Existence of a Solution

Consider the directed and weighted ERGM that fulfills in expectation

(i) an in-strength sequence $s^{(\mathrm{in})} = \left(s_1^{(\mathrm{in})}, \ldots, s_n^{(\mathrm{in})}\right) \in \mathbb{R}^n_{>0}$, describing the desired inflowing weight of each node,

(ii) and an out-strength sequence $s^{(\mathrm{out})} = \left(s_1^{(\mathrm{out})}, \ldots, s_n^{(\mathrm{out})}\right) \in \mathbb{R}^n_{>0}$, describing the desired outflowing weight of each node.

Let the set of considered graphs $\mathcal{G}$ be given by

$$\mathcal{G} = \left\{ w \in \mathbb{N}_{\geq 0}^{n \times n} : w_{11} = \cdots = w_{nn} = 0 \right\}. \tag{3.1}$$

The ERGM is formally described by the constrained maximum entropy problem

$$\max_{P \in \mathcal{P}} S(P) = \max_{P \in \mathcal{P}} - \sum_{w \in \mathcal{G}} P(w) \log\big(P(w)\big), \tag{3.2}$$

subject to

$$
\begin{aligned}
\sum_{w \in \mathcal{G}} P(w) &= 1, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{i=1}^{n} w_{ij} &= s_j^{(\mathrm{in})}, \quad \text{for } j = 1, \ldots, n, \\
\sum_{w \in \mathcal{G}} P(w) \sum_{j=1}^{n} w_{ij} &= s_i^{(\mathrm{out})}, \quad \text{for } i = 1, \ldots, n.
\end{aligned}
\tag{3.3}
$$

Solving the ERGM as usual via the method of Lagrange multipliers, leads to the following expected link weights, for $i, j = 1, \ldots, n$ and $i \neq j$,

$$
\mathbb{E}\left[W_{ij}\right] = \frac{e^{-\lambda_i^{(\mathrm{out})} - \lambda_j^{(\mathrm{in})}}}{1 - e^{-\lambda_i^{(\mathrm{out})} - \lambda_j^{(\mathrm{in})}}},
\tag{3.4}
$$

where $\left(\lambda_1^{(\mathrm{in})}, \ldots, \lambda_n^{(\mathrm{in})}\right) \in \mathbb{R}^n$ denote the Lagrange multipliers belonging to the constraints on the in-strength sequence, and $\left(\lambda_1^{(\mathrm{out})}, \ldots, \lambda_n^{(\mathrm{out})}\right) \in \mathbb{R}^n$ the Lagrange multipliers belonging to the constraints on the out-strength sequence. To simplify notation and calculations, in the following we write $a_i := \exp\left(-\lambda_i^{(\mathrm{out})}\right)$ and $b_i := \exp\left(-\lambda_i^{(\mathrm{in})}\right)$.

Similar to the weighted ERGM derived in detail in Section 2.1.3 (compare Eq. (2.38)), the following condition has to be satisfied,

$$
a_i b_j < 1,
\tag{3.5}
$$

for all $i, j = 1, \ldots, n$ and $i \neq j$.

The parameters can be calibrated based on the matrix of expected weights,

$$
\mathbb{E}\left[W\right] = \begin{array}{cc}
\begin{pmatrix}
0 & \frac{a_1 b_2}{1-a_1 b_2} & \frac{a_1 b_3}{1-a_1 b_3} & \cdots & \cdots & \frac{a_1 b_n}{1-a_1 b_n} \\
\frac{a_2 b_1}{1-a_2 b_1} & \ddots & \frac{a_2 b_3}{1-a_2 b_3} & \cdots & \cdots & \frac{a_2 b_n}{1-a_2 b_n} \\
\vdots & & \ddots & & & \vdots \\
\vdots & & & \ddots & & \vdots \\
\vdots & & & & \ddots & \frac{a_{n-1} b_n}{1-a_{n-1} b_n} \\
\frac{a_n b_1}{1-a_n b_1} & \cdots & \cdots & \cdots & \frac{a_n b_{n-1}}{1-a_n b_{n-1}} & 0
\end{pmatrix}
&
\begin{matrix}
s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} \\ \vdots \\ \vdots \\ \vdots \\ s_n^{(\mathrm{out})}
\end{matrix}
\\
\begin{matrix}
s_1^{(\mathrm{in})} & s_2^{(\mathrm{in})} & \cdots & \cdots & \cdots & s_n^{(\mathrm{in})}
\end{matrix}
&
\end{array}.
\tag{3.6}
$$

Hence, the question of the existence of a solution to the ERGM with desired in- and out-strength sequence can be reformulated to the question whether and for which $s^{(\mathrm{in})}$ and $s^{(\mathrm{out})}$ there exist vectors $a = (a_1, \dots, a_n) \in \mathbb{R}^n_{>0}$ and $b = (b_1, \dots, b_n) \in \mathbb{R}^n_{>0}$, that satisfy Eq. (3.5) and fulfill the row and column sums of the matrix of Eq. (3.6). The following theorem answers this question.

Theorem 3.1.1 (Existence of a Solution to the ERGM with Desired In- and Out-Strength Sequence)
The directed weighted ERGM that fulfills in expectation an in-strength sequence $s^{(\mathrm{in})} = \left(s^{(\mathrm{in})}_1, \dots, s^{(\mathrm{in})}_n\right) \in \mathbb{R}^n_{>0}$*, and an out-strength sequence* $s^{(\mathrm{out})} = \left(s^{(\mathrm{out})}_1, \dots, s^{(\mathrm{out})}_n\right) \in \mathbb{R}^n_{>0}$*, as specified by the optimization problem given in Eqs.* (3.2) *and* (3.3)*, has a solution if and only if the in- and out-strength sequences satisfy the following two conditions,*

$$\text{(i)} \qquad \sum_{i=1}^{n} s^{(\mathrm{out})}_i = \sum_{j=1}^{n} s^{(\mathrm{in})}_j, \tag{3.7}$$

$$\text{(ii)} \qquad s^{(\mathrm{in})}_j < \sum_{\substack{i=1\\ i\neq j}}^{n} s^{(\mathrm{out})}_i, \quad \textit{for all } j = 1, \dots, n. \tag{3.8}$$

In other words, there exist vectors $a = (a_1, \dots, a_n) \in \mathbb{R}^n_{>0}$ *and* $b = (b_1, \dots, b_n) \in \mathbb{R}^n_{>0}$ *that satisfy Eq.* (3.5) *and fulfill the row and column sums of the matrix of Eq.* (3.6)*.*

Proof
The first direction "⇒" is trivial. Let the vectors $a \in \mathbb{R}^n_{>0}$ and $b \in \mathbb{R}^n_{>0}$ constitute a solution to the ERGM under consideration. The following holds,

$$\sum_{i=1}^{n} s^{(\mathrm{out})}_i = \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{a_i b_j}{1 - a_i b_j} = \sum_{j=1}^{n} \sum_{\substack{i=1\\ i\neq j}}^{n} \frac{a_i b_j}{1 - a_i b_j} = \sum_{j=1}^{n} s^{(\mathrm{in})}_j, \tag{3.9}$$

i.e. the condition of Eq. (3.7) is fulfilled. Similarly, it follows,

$$s^{(\mathrm{in})}_j = \sum_{\substack{i=1\\ i\neq j}}^{n} \frac{a_i b_j}{1 - a_i b_j} < \left(\sum_{\substack{i=1\\ i\neq j}}^{n} \frac{a_i b_j}{1 - a_i b_j} \right) + \sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{\substack{k=1\\ k\neq i,j}}^{n} \underbrace{\frac{a_i b_k}{1 - a_i b_k}}_{>0} = \sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{\substack{k=1\\ k\neq i}}^{n} \frac{a_i b_k}{1 - a_i b_k} = \sum_{\substack{i=1\\ i\neq j}}^{n} s^{(\mathrm{out})}_i, \tag{3.10}$$

i.e. the condition of Eq. (3.8) is fulfilled.

The second direction "⇐" is more complicated to formalize. The proof is based on the observation that by iteratively adjusting the row parameters $(a_1, \dots, a_n)$ such that the row sums are fulfilled, and subsequently adjusting the column parameters $(b_1, \dots, b_n)$ such that the column sums are fulfilled, the vectors a and b converge to the solution.

To better understand what is going on, we start by illustrating the methodology via an example.

Example 3.1.2

Let the desired row and column sums be given by $s^{(\mathrm{out})} = (12, 1, 8, 3)$ *and* $s^{(\mathrm{in})} = (4, 4, 1, 15)$. *We start with a homogeneous matrix where each element is assigned a weight of* $1/(4 \cdot 3)\sum_{i=1}^{4} s_i^{(\mathrm{out})} = 24/12 = 2$, *see Eq.* (3.12). *Moreover, let the row and column parameters all be identical, i.e.* $a_1 = \ldots = a_n = b_1 = \ldots = b_n$ *and hence,*

$$\frac{a_1^2}{1-a_1^2} = 2 \quad \Rightarrow \quad a_1 = \left(\frac{2}{3}\right)^{1/2} = 0.8165. \tag{3.11}$$

The following matrices illustrate the changes taking place when updating the row and column parameters iteratively. The ϵ *variables denote the excess (or missing) weight of the rows and columns. Whenever a row (resp. column) has too much weight, the corresponding row (resp. column) parameter will decrease in the subsequent step. This means all elements of the concerned row (resp. column) will decrease (marked in blue). Analogously, in case a row (resp. column) is missing weight, the corresponding parameter and all elements of the concerned row (resp. column) will increase (marked in green). The magnitude of the change of each element is reflected by the intensity of the color (scaled to 100% in each step). The arrows next to the parameters indicate if the parameter was increased or decreased.*

We start with the homogeneous matrix pictured in Eq. (3.12),

$$\begin{array}{cc}
\begin{pmatrix} 0 & 2 & 2 & 2 \\ 2 & 0 & 2 & 2 \\ 2 & 2 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{pmatrix} &
\begin{array}{ll} a_1{=}0.8165 & \epsilon_1^{(\mathrm{out})}{=}6{-}12{=}{-}6 \\ a_2{=}0.8165 & \epsilon_2^{(\mathrm{out})}{=}6{-}1{=}5 \\ a_3{=}0.8165 & \epsilon_3^{(\mathrm{out})}{=}6{-}8{=}{-}2 \\ a_4{=}0.8165 & \epsilon_4^{(\mathrm{out})}{=}6{-}3{=}3 \end{array} \\
\begin{array}{cccc} b_1{=}0.8165 & b_2{=}0.8165 & b_3{=}0.8165 & b_4{=}0.8165 \\ \epsilon_1^{(\mathrm{in})}{=}6{-}4{=}2 & \epsilon_2^{(\mathrm{in})}{=}6{-}4{=}2 & \epsilon_3^{(\mathrm{in})}{=}6{-}1{=}5 & \epsilon_4^{(\mathrm{in})}{=}6{-}15{=}{-}9 \end{array} &
\begin{array}{ll} & \sum_i|\epsilon_i^{(\mathrm{out})}|{=}16 \\ \sum_i|\epsilon_i^{(\mathrm{in})}|{=}18 & . \end{array}
\end{array} \tag{3.12}$$

Iteration 1: Adjusting the row parameters leads to

$$\begin{array}{cccccc}
\left(\begin{matrix} 0 & 4 & 4 & 4 \\ 0.3333 & 0 & 0.3333 & 0.3333 \\ 2.6666 & 2.6666 & 0 & 2.6666 \\ 1 & 1 & 1 & 0 \end{matrix}\right) & & & & \begin{matrix} a_1{=}0.9798\uparrow \\ a_2{=}0.3062\downarrow \\ a_3{=}0.8907\uparrow \\ a_4{=}0.6124\downarrow \end{matrix} & \begin{matrix} \epsilon_1^{(\text{out})}{=}12{-}12{=}0 \\ \epsilon_2^{(\text{out})}{=}1{-}1{=}0 \\ \epsilon_3^{(\text{out})}{=}8{-}8{=}0 \\ \epsilon_4^{(\text{out})}{=}3{-}3{=}0 \end{matrix} \\
\begin{matrix} b_1{=}0.8165 \\ \epsilon_1^{(\text{in})}{=}4{-}4{=}0 \end{matrix} & \begin{matrix} b_2{=}0.8165 \\ \epsilon_2^{(\text{in})}{=}7.67{-}4{=}3.67 \end{matrix} & \begin{matrix} b_3{=}0.8165 \\ \epsilon_3^{(\text{in})}{=}5.33{-}1{=}4.33 \end{matrix} & \begin{matrix} b_4{=}0.8165 \\ \epsilon_4^{(\text{in})}{=}7{-}15{=}{-}8 \end{matrix} & \sum_i|\epsilon_i^{(\text{in})}|{=}16 & .
\end{array} \tag{3.13}$$

Iteration 1: Adjusting the column parameters leads to

$$\begin{array}{cccccc}
\left(\begin{matrix} 0 & 1.8647 & 0.5754 & 9.8736 \\ 0.3333 & 0 & 0.1288 & 0.3962 \\ 2.6666 & 1.4495 & 0 & 4.7302 \\ 1 & 0.6858 & 0.2958 & 0 \end{matrix}\right) & & & & \begin{matrix} a_1{=}0.9798 \\ a_2{=}0.3062 \\ a_3{=}0.8907 \\ a_4{=}0.6124 \end{matrix} & \begin{matrix} \epsilon_1^{(\text{out})}{=}12.31{-}12{=}0.31 \\ \epsilon_2^{(\text{out})}{=}0.86{-}1{=}{-}0.14 \\ \epsilon_3^{(\text{out})}{=}8.85{-}8{=}0.85 \\ \epsilon_4^{(\text{out})}{=}1.95{-}3{=}{-}1.02 \end{matrix} \\
\begin{matrix} b_1{=}0.8165 \\ \epsilon_1^{(\text{in})}{=}4{-}4{=}0 \end{matrix} & \begin{matrix} b_2{=}0.6643\downarrow \\ \epsilon_2^{(\text{in})}{=}4{-}4{=}0 \end{matrix} & \begin{matrix} b_3{=}0.3728\downarrow \\ \epsilon_3^{(\text{in})}{=}1{-}1{=}0 \end{matrix} & \begin{matrix} b_4{=}0.9268\uparrow \\ \epsilon_4^{(\text{in})}{=}15{-}15{=}0 \end{matrix} & & \sum_i|\epsilon_i^{(\text{out})}|{=}2.32 \\
 & & & & & .
\end{array} \tag{3.14}$$

Note that within each column the magnitude by which each element changes follows the order of the row parameters $(a_1, \dots, a_n)$*. Moreover, note that the sum of absolute errors decreases with each iteration. This is because of two reasons. Firstly, during every step weight of* $\frac{1}{2}\sum_i|\epsilon_i^{(\text{in})}|$ *(resp.* $\frac{1}{2}\sum_i|\epsilon_i^{(\text{out})}|$*) is shifted. Hence, the error induced in the row sums (resp. column sums) is bounded by the size of the current error. Second, decreasing some elements of a row (resp. column) while increasing others leads to part of the error canceling out across the row (resp. column) sums. Consider for example the matrix pictured in Eq.* (3.14)*. Decreasing* b_2 *and* b_3 *induces a negative error of* $(1.8647 - 4) + (0.5754 - 4) = -5.5599$ *in the first row, while increasing* b_4 *induces a positive error of* $9.8736 - 4 = 5.8736$*. Hence, in total the error induced in the first row by updating the column parameters equals only* 0.3137*.*

Iteration 2: Adjusting the row parameters leads to

$$
\begin{array}{cc}
\begin{pmatrix} 0 & 1.8496 & 0.5728 & 9.5776 \\ 0.3884 & 0 & 0.1464 & 0.4652 \\ 2.4565 & 1.3711 & 0 & 4.1724 \\ 1.6040 & 1.0048 & 0.3912 & 0 \end{pmatrix} &
\begin{array}{ll} a_1{=}0.9770\downarrow & \epsilon_1^{(\mathrm{out})}{=}0 \\ a_2{=}0.3426\uparrow & \epsilon_2^{(\mathrm{out})}{=}0 \\ a_3{=}0.8704\downarrow & \epsilon_3^{(\mathrm{out})}{=}0 \\ a_4{=}0.7544\uparrow & \epsilon_4^{(\mathrm{out})}{=}0 \end{array} \\
\begin{array}{cccc} b_1{=}0.8165 & b_2{=}0.6643 & b_3{=}0.3728 & b_4{=}0.9268 \\ \epsilon_1^{(\mathrm{in})}{=}4.45{-}4{=}0.45 & \epsilon_2^{(\mathrm{in})}{=}4.23{-}4{=}0.23 & \epsilon_3^{(\mathrm{in})}{=}1.11{-}1{=}0.11 & \epsilon_4^{(\mathrm{in})}{=}14.22{-}15{=}{-}0.78 \end{array} & \textstyle\sum_i|\epsilon_i^{(\mathrm{in})}|{=}1.57
\end{array}
\quad . \tag{3.15}
$$

Iteration 2: Adjusting the column parameters leads to

$$
\begin{array}{cc}
\begin{pmatrix} 0 & 1.7376 & 0.5116 & 10.2250 \\ 0.3689 & 0 & 0.1347 & 0.4693 \\ 2.1714 & 1.3013 & 0 & 4.3057 \\ 1.4596 & 0.9612 & 0.3538 & 0 \end{pmatrix} &
\begin{array}{ll} a_1{=}0.9770 & \epsilon_1^{(\mathrm{out})}{=}12.47{-}12{=}0.47 \\ a_2{=}0.3426 & \epsilon_2^{(\mathrm{out})}{=}0.97{-}1{=}{-}0.03 \\ a_3{=}0.8704 & \epsilon_3^{(\mathrm{out})}{=}7.78{-}8{=}{-}0.22 \\ a_4{=}0.7544 & \epsilon_4^{(\mathrm{out})}{=}2.77{-}3{=}{-}0.23 \end{array} \\
\begin{array}{cccc} b_1{=}0.7866\downarrow & b_2{=}0.6496\downarrow & b_3{=}0.3464\downarrow & b_4{=}0.9323\uparrow \\ \epsilon_1^{(\mathrm{in})}{=}4{-}4{=}0 & \epsilon_2^{(\mathrm{in})}{=}4{-}4{=}0 & \epsilon_3^{(\mathrm{out})}{=}1{-}1{=}0 & \epsilon_4^{(\mathrm{in})}{=}15{-}15{=}0 \end{array} & \textstyle\sum_i|\epsilon_i^{(\mathrm{in})}|{=}0.95
\end{array}
\quad . \tag{3.16}
$$

Setting the stopping criteria to $\sum_i|\epsilon_i^{(\mathrm{out})}| \leq 10^{-5}$, *the implementation in Matlab stops after 15 iterations with the following result:*

$$
\begin{array}{cc}
\begin{pmatrix} 0 & 1.6505 & 0.4913 & 9.8582 \\ 0.3702 & 0 & 0.1362 & 0.4936 \\ 2.0559 & 1.2959 & 0 & 4.6482 \\ 1.5739 & 1.0536 & 0.3725 & 0 \end{pmatrix} &
\begin{array}{ll} a_1{=}0.9704 & \epsilon_1^{(\mathrm{out})}{=}1.36e{-}06 \\ a_2{=}0.3532 & \epsilon_2^{(\mathrm{out})}{=}{-}1.22e{-}07 \\ a_3{=}0.8796 & \epsilon_3^{(\mathrm{out})}{=}{-}1.21e{-}06 \\ a_4{=}0.7995 & \epsilon_4^{(\mathrm{out})}{=}{-}1.28e{-}06 \end{array} \\
\begin{array}{cccc} b_1{=}0.7648 & b_2{=}0.6417 & b_3{=}0.3395 & b_4{=}0.9356 \\ \epsilon_1^{(\mathrm{in})}{=}{-}1.01e{-}07 & \epsilon_2^{(\mathrm{in})}{=}{-}1.05e{-}07 & \epsilon_3^{(\mathrm{in})}{=}{-}2.34e{-}09 & \epsilon_4^{(\mathrm{in})}{=}{-}1.03e{-}06 \end{array} &
\end{array}
\quad . \tag{3.17}
$$

The formal proof is structured as follows.

(1) First, we show that for arbitrary vectors b, there exists a unique, admissible (in the sense of fulfilling Eq. (3.5)) vector a, such that all row sums are satisfied. Vice versa, for arbitrary vectors a, there exists a unique, admissible vector b, such that all column sums are satisfied.

(2) We proceed to prove that in each iteration (updating the row and column parameters) the sum of absolute errors over all row and column sums is strictly decreasing.

(3) With the exception of a special case, the amount by which the global error decreases in each iteration is bounded from below by $\frac{\epsilon}{(n-1)^2}$, where ϵ denotes the error at the current iteration.

(4) Last, we discuss the special case and show that it eventually also leads to the solution.

(1) We begin by proofing that in each iteration we can indeed find a vector a (resp. b) such that all row sums (resp. column sums) are satisfied. Moreover, the respective vector is unique. For this purpose, we define the following continuous functions for the row and column sums of the matrix of expected weights, given in Eq. (3.6), for $i, j = 1, \ldots, n$,

$$\begin{aligned} f_i^{(\text{out})}(a,b) &:= \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i b_j}{1 - a_i b_j} \right) - s_i^{(\text{out})}, \\ f_j^{(\text{in})}(a,b) &:= \left(\sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{a_i b_j}{1 - a_i b_j} \right) - s_j^{(\text{in})}. \end{aligned} \tag{3.18}$$

Solving the ERGM is equivalent to finding the global root of these functions, which is non trivial. However, for the proposed methodology, it suffices to show that there always exists a root for each of the functions, when considered independently. Note that for every vector $a = (a_1, \ldots, a_n) \in \mathbb{R}^n_{>0}$ and every vector $b = (b_1, \ldots, b_n) \in \mathbb{R}^n_{>0}$, the following limits hold, for $i = 1, \ldots, n$,

$$\begin{aligned} \lim_{a_i \searrow 0} f_i^{(\text{out})}(a,b) &= 0 - s_i^{(\text{out})} < 0, \\ \lim_{a_i \nearrow \min\left\{b_j^{-1} : j=1,\ldots,n \wedge j \neq i\right\}} f_i^{(\text{out})}(a,b) &= \infty - s_i^{(\text{out})} > 0, \end{aligned} \tag{3.19}$$

and analogously, for $j = 1, \ldots, n$,

$$\begin{aligned} \lim_{b_j \searrow 0} f_j^{(\text{in})}(a,b) &= 0 - s_j^{(\text{in})} < 0, \\ \lim_{b_j \nearrow \min\left\{a_i^{-1} : i=1,\ldots,n \wedge i \neq j\right\}} f_j^{(\text{in})}(a,b) &= \infty - s_j^{(\text{in})} > 0. \end{aligned} \tag{3.20}$$

Hence, we know from the intermediate value theorem that for all vectors a and b, there always exist $a_i^* \in \mathbb{R}_{>0}$ and $b_j^* \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} f_i^{(\mathrm{out})}(a,b)\Big|_{a_i=a_i^*} &= 0, \\ f_j^{(\mathrm{in})}(a,b)\Big|_{b_j=b_j^*} &= 0. \end{aligned} \tag{3.21}$$

Since the functions $f_i^{(\mathrm{out})}$ and $f_j^{(\mathrm{in})}$ are strictly monotonically increasing in all variables, the roots a_i^* and b_j^* are unique. Moreover, note that $f_i^{(\mathrm{out})}$ depends only on a_i and b. Hence, for every vector $b = (b_1, \ldots, b_n) \in \mathbb{R}^n_{>0}$ there exists a unique vector $a^* = (a_1^*, \ldots, a_n^*) \in \mathbb{R}^n_{>0}$ such that $f_i^{(\mathrm{out})}(a^*, b) = 0$ holds for all $i = 1, \ldots, n$. Analogously, for every vector $a = (a_1, \ldots, a_n) \in \mathbb{R}^n_{>0}$ there exists a unique vector $b^* = (b_1^*, \ldots, b_n^*) \in \mathbb{R}^n_{>0}$ such that $f_j^{(\mathrm{in})}(a, b^*) = 0$ holds for all $j = 1, \ldots, n$. Computing the roots a^* and b^* numerically is straightforward since the functions are continuous and strictly monotonically increasing. Thus, we conclude that for arbitrary vectors b, there exists a unique, admissible vector a, such that all row sums are satisfies, and vice versa.

(2) Next we prove that in each iteration the global error is strictly deceasing. Since the parameters have to be updated several times, we denote the iteration number in brackets, i.e. $a(t)$ denotes the vector a after t iterations. As initial values any admissible parameters $a(0) \in \mathbb{R}^n_{>0}$ and $b(0) \in \mathbb{R}^n_{>0}$, fulfilling the condition of Eq. (3.5), serve. For example, we can start with homogeneous parameters $a_1(0) = \ldots = a_n(0) = b_1(0) = \ldots = b_n(0)$, and hence,

$$\begin{aligned} & n(n-1)\frac{a_1^2(0)}{1-a_1^2(0)} = s^{(\mathrm{total})} \\ \Leftrightarrow \quad & a_1^2(0) = \frac{s^{(\mathrm{total})}}{n(n-1)}\left(1-a_1^2(0)\right) \\ \Leftrightarrow \quad & a_1^2(0)\left(1+\frac{s^{(\mathrm{total})}}{n(n-1)}\right) = \frac{s^{(\mathrm{total})}}{n(n-1)} \\ \Leftrightarrow \quad & a_1^2(0) = \frac{s^{(\mathrm{total})}}{n(n-1)}\frac{n(n-1)}{n(n-1)+s^{(\mathrm{total})}} \\ \Leftrightarrow \quad & a_1(0) = \left(\frac{s^{(\mathrm{total})}}{n(n-1)+s^{(\mathrm{total})}}\right)^{1/2}, \end{aligned} \tag{3.22}$$

where $s^{(\mathrm{total})} = \sum_{i=1}^n s_i^{(\mathrm{out})}$ denotes the total weight.

According to the result derived above (see (1)), we can adjust the vector $a(0) \rightarrow a(1)$, such that the row sums are fulfilled, i.e.

$$\sum_{i=1}^{n} \left| f_i^{(\text{out})}\Big(a(1), b(0)\Big) \right| = \sum_{i=1}^{n} \left| \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i(1)\, b_j(0)}{1 - a_i(1)\, b_j(0)} \right) - s_i^{(\text{out})} \right| = 0 \tag{3.23}$$

holds. While the row sums are satisfied at the moment, the column sums most probably will not sum up to the desired values. We denote the sum of absolute errors of the column sums by $2\epsilon_{1,0}^{(\text{col})}$, i.e.

$$\sum_{j=1}^{n} \left| f_j^{(\text{in})}\Big(a(1), b(0)\Big) \right| = 2\epsilon_{1,0}^{(\text{col})}. \tag{3.24}$$

Note that because the entire weight $s^{(\text{total})}$ is allocated, the sum of weight that is missing in some columns has to equal the sum of weight that is in excess in other columns. Hence, when adjusting the column sums (by updating b accordingly), weight of amount $\epsilon_{1,0}^{(\text{col})}$ is shifted between the columns. Moreover, for those columns where weight is missing, we have to increase the corresponding column parameter, and for those columns that have too much weight, the corresponding parameters have to be decreased. We group the parameters according to their direction of change,

$$\begin{aligned} B^{\uparrow} &:= \left\{ b_j : f_j^{(\text{in})}\Big(a(1), b(0)\Big) < 0, \text{ for } j = 1, \dots, n \right\}, \\ B^{\downarrow} &:= \left\{ b_j : f_j^{(\text{in})}\Big(a(1), b(0)\Big) > 0, \text{ for } j = 1, \dots, n \right\}, \\ B^{0} &:= \left\{ b_j : f_j^{(\text{in})}\Big(a(1), b(0)\Big) = 0, \text{ for } j = 1, \dots, n \right\}. \end{aligned} \tag{3.25}$$

Next, we note that shifting weight of amount $\epsilon_{1,0}^{(\text{col})}$ between the columns can cause a maximum error in the row sums of exactly $2\epsilon_{1,0}^{(\text{col})}$, as consequently weight of $\epsilon_{1,0}^{(\text{col})}$ could be missing in one row, while another row receives the same amount in excess,

$$
\begin{aligned}
&\sum_{i=1}^{n} \left| f_i^{(\mathrm{out})}\Big(a\left(1\right), b\left(1\right)\Big) \right| \\
&\quad = \sum_{i=1}^{n} \left| \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} \right) - s_i^{(\mathrm{out})} \right| \\
&\quad = \sum_{i=1}^{n} \left| \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} + \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) - s_i^{(\mathrm{out})} \right| \\
&\quad = \sum_{i=1}^{n} \left| \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) + \underbrace{\left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) - s_i^{(\mathrm{out})}}_{=0} \right| \\
&\quad = \sum_{i=1}^{n} \left| \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) \right| \\
&\quad \overset{(*)}{\leq} \sum_{i=1}^{n} \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \left| \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right| \right) \\
&\quad = \sum_{i=1}^{n} \left(\sum_{\substack{b_j \in B^{\uparrow} \\ j \neq i}} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) \\
&\qquad\qquad + \left(\sum_{\substack{b_j \in B^{\downarrow} \\ j \neq i}} \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} - \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} \right) \\
&\quad = \left(\sum_{b_j \in B^{\uparrow}} \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} - \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} \right) \\
&\qquad + \left(\sum_{b_j \in B^{\downarrow}} \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{a_i\left(1\right) b_j\left(0\right)}{1 - a_i\left(1\right) b_j\left(0\right)} - \frac{a_i\left(1\right) b_j\left(1\right)}{1 - a_i\left(1\right) b_j\left(1\right)} \right) \\
&\quad = 2\epsilon_{1,0}^{(\mathrm{col})}.
\end{aligned}
\tag{3.26}
$$

Analyzing $(*)$ in more detail yields,

$$\begin{aligned}
&\left|\left(\sum_{\substack{j=1\\j\neq i}}^{n}\frac{a_i(1)\,b_j(1)}{1-a_i(1)\,b_j(1)}-\frac{a_i(1)\,b_j(0)}{1-a_i(1)\,b_j(0)}\right)\right|\\
&=\Bigg|\underbrace{\left(\sum_{\substack{b_j\in B^{\uparrow}\\j\neq i}}\frac{a_i(1)\,b_j(1)}{1-a_i(1)\,b_j(1)}-\frac{a_i(1)\,b_j(0)}{1-a_i(1)\,b_j(0)}\right)}_{=:B_1}-\underbrace{\left(\sum_{\substack{b_j\in B^{\downarrow}\\j\neq i}}\frac{a_i(1)\,b_j(0)}{1-a_i(1)\,b_j(0)}-\frac{a_i(1)\,b_j(1)}{1-a_i(1)\,b_j(1)}\right)}_{=:B_2}\Bigg|\\
&=\begin{cases}B_1-B_2, & \text{for } B_1\geq B_2\\ B_2-B_1, & \text{for } B_1\leq B_2\end{cases}
\end{aligned} \tag{3.27}$$

and hence equality in $(*)$ holds

$$\begin{aligned}
&\text{for } B_1\geq B_2: \quad B_1-B_2=B_1+B_2 \quad\Leftrightarrow\quad B_2=0 \quad\Leftrightarrow\quad B^{\downarrow}\setminus b_i=\emptyset,\\
&\text{for } B_1\leq B_2: \quad B_2-B_1=B_2+B_1 \quad\Leftrightarrow\quad B_1=0 \quad\Leftrightarrow\quad B^{\uparrow}\setminus b_i=\emptyset.
\end{aligned} \tag{3.28}$$

Thus, equality in $(*)$ in Eq. (3.26) holds if and only if Eq. (3.28) is fulfilled for all i. This means $B^{\downarrow}=\emptyset$ or $B^{\uparrow}=\emptyset$ has to hold. Since the total weight is always allocated, whenever there exists a column with excess weight there is also at least one column that is missing weight (and vice versa), i.e.

$$|B^{\downarrow}|>0 \quad\Leftrightarrow\quad |B^{\uparrow}|>0. \tag{3.29}$$

Hence, Eq. (3.28) can only hold for all i when $B^{\downarrow}=B^{\uparrow}=\emptyset$. This in turn means all row and column sums are satisfied and the current parameters constitute a solution. The results of Eqs. (3.26) to (3.28) hold for every iteration t and also analogously when updating the row parameters a. Therefore, updating iteratively the row and column parameters by solving for the roots of each row function $f_i^{(\mathrm{out})}$ and each column function $f_j^{(\mathrm{in})}$, for $i,j=1,\ldots,n$, strictly decreases the error.

(3) It is left to show that the sequence by which the error decreases converges to zero only if the parameters converge to a valid solution. To better understand the unfolding dynamics when updating the parameters, consider the example pictured in Eq. (3.30). W.l.o.g. let $a(t+1)$ be such that the row sums are satisfied. Hence, next, we update the column parameters $b(t)\to b(t+1)$. Furthermore, let $\epsilon_1^{(\mathrm{in})},\ldots,\epsilon_4^{(\mathrm{in})}$ denote the current error terms of each column.

$$
\begin{array}{ccc}
\begin{pmatrix}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{pmatrix}
&
\begin{array}{c}
s_1^{(\text{out})} \\
s_2^{(\text{out})} \\
s_3^{(\text{out})} \\
s_4^{(\text{out})} \\
s_5^{(\text{out})}
\end{array}
&
\begin{array}{c}
a_1\left(t+1\right) \\
a_2\left(t+1\right) \\
a_3\left(t+1\right) \\
a_4\left(t+1\right) \\
a_5\left(t+1\right)
\end{array}
\\
\begin{array}{ccccc}
s_1^{(\text{in})}+\epsilon_1^{(\text{in})} & s_2^{(\text{in})}-\epsilon_2^{(\text{in})} & s_3^{(\text{in})}+\epsilon_3^{(\text{in})} & s_4^{(\text{in})}-\epsilon_4^{(\text{in})} & s_5^{(\text{in})}
\end{array}
& & \\
\begin{array}{ccccc}
b_1\left(t\right)\downarrow & b_2\left(t\right)\uparrow & b_3\left(t\right)\downarrow & b_4\left(t\right)\uparrow & b_5\left(t\right)
\end{array}
& &
\end{array}
\tag{3.30}
$$

We denote by $j_\oplus$ the index of the column with the highest amount of excess weight and by $j_\ominus$ the index of the column that is missing most weight, i.e.

$$
\begin{aligned}
f_{j_\oplus}^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big) &= \max\Big\{f_j^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big) : j=1,\ldots,n\Big\}, \\
f_{j_\ominus}^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big) &= \min\Big\{f_j^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big) : j=1,\ldots,n\Big\}.
\end{aligned}
\tag{3.31}
$$

In the example in Eq. (3.30), let $j_\oplus = 1$ and $j_\ominus = 2$.

Furthermore, we can define the following lower bound for the absolute error of columns $j_\oplus$ and $j_\ominus$,

$$
\min\left\{f_{j_\oplus}^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big), -f_{j_\ominus}^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big)\right\} \geq \min\left\{\frac{\epsilon_{t+1,t}^{(\text{col})}}{|B^\uparrow|}, \frac{\epsilon_{t+1,t}^{(\text{col})}}{|B^\downarrow|}\right\} \geq \frac{\epsilon_{t+1,t}^{(\text{col})}}{n-1},
\tag{3.32}
$$

where $2\epsilon_{t+1,t}^{(\text{col})} = \sum_{j=1}^{n}\left|f_j^{(\text{in})}\Big(a\left(t+1\right), b\left(t\right)\Big)\right|$ denotes the sum of absolute errors over all columns and the sets $B^\uparrow$ and $B^\downarrow$ are defined as in Eq. (3.25) w.r.t the current iteration.

Updating the column parameters $b\left(t\right) \rightarrow b\left(t+1\right)$, i.e. in the example of Eq. (3.30) this means decreasing b_1 and increasing b_2, induces an error in the rows sums. However, part of the induced error cancels out. Consider the example pictured in Eq. (3.30), decreasing b_1 leads to a decrease in all elements of the first column, marked in blue. Since the first column has excess weight of amount $\epsilon_1^{(\text{in})}$, the parameter b_1 is decreased until the elements of the first column jointly lose weight of amount $\epsilon_1^{(\text{in})}$. This in turn induces a negative error of ϵ_1 in the row sums. Likewise, increasing b_2 leads to an increase in all elements of the second column, marked in green, which induces a positive error of $\epsilon_2^{(\text{in})}$ in the row sums. However, part of the error induced to the row sums cancels out as the decrease of w_{i1} is partly offset by the increase of w_{i2} in the sum of row i, for all $i = 3, \ldots, n$.

Hence, by updating $b_{j\oplus}$ and $b_{j\ominus}$, the overall error reduced in each of the rows $i \in \{1,\ldots,n\} \setminus \{j_\oplus, j_\ominus\}$ by the minimum of the change in the two elements $w_{ij\oplus}$ and $w_{ij\ominus}$, i.e. by

$$\begin{aligned}\min\Bigg\{&\frac{a_i\,(t+1)\,b_{j\oplus}\,(t)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t)} - \frac{a_i\,(t+1)\,b_{j\oplus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t+1)},\\ &\quad\frac{a_i\,(t+1)\,b_{j\ominus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t+1)} - \frac{a_i\,(t+1)\,b_{j\ominus}\,(t)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t)}\Bigg\}.\end{aligned} \tag{3.33}$$

Note, that the error that cancels out in each row is increasing in a_i,

$$\begin{aligned}&\frac{\partial}{\partial a_i}\left(\frac{a_i\,(t+1)\,b_{j\oplus}\,(t)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t)} - \frac{a_i\,(t+1)\,b_{j\oplus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t+1)}\right)\\ &= \frac{b_{j\oplus}\,(t)}{\Big(1-a_i\,(t+1)\,b_{j\oplus}\,(t)\Big)^2} - \frac{b_{j\oplus}\,(t+1)}{\Big(1-a_i\,(t+1)\,b_{j\oplus}\,(t+1)\Big)^2} > 0,\end{aligned} \tag{3.34}$$

where the last inequality holds because $b_{j\oplus}\,(t) > b_{j\oplus}\,(t+1)$ and $0 < \big(1-a_i b_{j\oplus}\,(t)\big) < \big(1-a_i b_{j\oplus}\,(t+1)\big)$; and analogously,

$$\begin{aligned}&\frac{\partial}{\partial a_i}\left(\frac{a_i\,(t+1)\,b_{j\ominus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t+1)} - \frac{a_i\,(t+1)\,b_{j\ominus}\,(t)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t)}\right)\\ &= \frac{b_{j\ominus}\,(t+1)}{\Big(1-a_i\,(t+1)\,b_{j\ominus}\,(t+1)\Big)^2} - \frac{b_{j\ominus}\,(t)}{\Big(1-a_i\,(t+1)\,b_{j\ominus}\,(t)\Big)^2} > 0.\end{aligned} \tag{3.35}$$

Moreover, note that the error $\epsilon_j^{(\mathrm{in})}$ of column j (compare the example pictured in Eq. (3.30)), i.e. the change in weight by updating the column parameter b_j, is distributed along $(n-1)$ rows. Therefore, the element of the row with the highest parameter a_i, for $i \in \{1,\ldots,n\} \setminus \{j\}$, will change by at least $\epsilon_j^{(\mathrm{in})} / (n-1)$. Thus, together with Eq. (3.32), and as long as $\max\{a_i : i = 1,\ldots,n\} \notin \{a_{j\oplus}, a_{j\ominus}\}$ holds, we can define the following lower bound for the reduction of the total error $\epsilon_{t+1,t}^{(\mathrm{col})}$,

$$\min\left\{\left(w_{ij\oplus}^{(\mathrm{old})} - w_{ij\oplus}^{(\mathrm{new})}\right), \left(w_{ij\ominus}^{(\mathrm{new})} - w_{ij\ominus}^{(\mathrm{old})}\right) : a_i = \max\{a_1,\ldots,a_n\}\right\} \geq \frac{\epsilon_{t+1,t}^{(\mathrm{col})}}{(n-1)^2}, \tag{3.36}$$

where

$$\begin{aligned} w_{ij\oplus}^{(\mathrm{old})} &= \frac{a_i\,(t+1)\,b_{j\oplus}\,(t)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t)}, & w_{ij\oplus}^{(\mathrm{new})} &= \frac{a_i\,(t+1)\,b_{j\oplus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\oplus}\,(t+1)},\\ w_{ij\ominus}^{(\mathrm{old})} &= \frac{a_i\,(t+1)\,b_{j\ominus}\,(t)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t)}, & w_{ij\ominus}^{(\mathrm{new})} &= \frac{a_i\,(t+1)\,b_{j\ominus}\,(t+1)}{1-a_i\,(t+1)\,b_{j\ominus}\,(t+1)}.\end{aligned} \tag{3.37}$$

Summing up, in each iteration where the highest row (resp. column) parameter has a different index than the columns (resp. rows) with the highest positive and highest negative error, we know that the global error $\epsilon_{t+1,t}^{(\mathrm{col})}$ (resp. $\epsilon_{t+1,t+1}^{(\mathrm{row})}$) decreases at least by an amount of $\frac{\epsilon_{t+1,t}^{(\mathrm{col})}}{(n-1)^2}$ (resp. $\frac{\epsilon_{t+1,t+1}^{(\mathrm{row})}}{(n-1)^2}$).

Furthermore, we can derive a similar result, whenever there are at least two columns (resp. rows) with positive error and two columns (resp. rows) with negative error. This even holds in case the highest row (resp. column) parameter has the same index as the column (resp. row) with the highest positive or negative error. Equation (3.38) visualizes this case. W.l.o.g. let $\epsilon_1^{(\mathrm{in})} > \epsilon_3^{(\mathrm{in})}$, $\epsilon_2^{(\mathrm{in})} > \epsilon_4^{(\mathrm{in})}$, and $a_1 \gg a_2 \gg a_3, a_4, a_5$ hold. This means that updating the first column happens mainly in w_{21} and updating the second column happens mainly in w_{12}. More precisely, w_{21} is losing weight of at least $\epsilon_1^{(\mathrm{in})}/\,(n-1)$ and w_{12} is gaining weight of at least $\epsilon_2^{(\mathrm{in})}/\,(n-1)$. As these changes do not happen within the same row, the lower bound derived in Eq. (3.36) does not have to hold. However, because column 3 is also updated, which means w_{13} is losing weight of at least $\epsilon_3/\,(n-1)$, this loss is offset by the gain of weight in w_{12}. Hence, the global error $\epsilon_{t+1,t}^{(\mathrm{in})}$ decreases at least by $\min\left\{\epsilon_2^{(\mathrm{in})}, \epsilon_3^{(\mathrm{in})}\right\}/\,(n-1)$.

$$
\begin{array}{cc}
\left(\begin{array}{ccccc}
0 & w_{12} & w_{13} & w_{14} & w_{15}\\
w_{21} & 0 & w_{23} & w_{24} & w_{25}\\
w_{31} & w_{32} & 0 & w_{34} & w_{35}\\
w_{41} & w_{42} & w_{43} & 0 & w_{45}\\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{array}\right)
&
\begin{array}{lc}
s_1^{(\mathrm{out})} & a_1 \gg a_2, a_3, a_4, a_5\\
s_2^{(\mathrm{out})} & a_2 \gg a_3, a_4, a_5\\
s_3^{(\mathrm{out})} & a_3\\
s_4^{(\mathrm{out})} & a_4\\
s_5^{(\mathrm{out})} & a_5
\end{array}
\\
\begin{array}{ccccc}
s_1^{(\mathrm{in})}+\epsilon_1^{(\mathrm{in})} & s_2^{(\mathrm{in})}-\epsilon_2^{(\mathrm{in})} & s_3^{(\mathrm{in})}+\epsilon_3^{(\mathrm{in})} & s_4^{(\mathrm{in})}-\epsilon_4^{(\mathrm{in})} & s_5^{(\mathrm{in})}\\
b_1\downarrow & b_2\uparrow & b_3\downarrow & b_4\uparrow & b_5
\end{array}
&
\end{array}
\tag{3.38}
$$

The analogous case where $a_2 \gg a_1 \gg a_3, a_4, a_5$ would lead to part of the error canceling out in row 2, and amounting to at least $\min\left\{\epsilon_1^{(\mathrm{in})}, \epsilon_4^{(\mathrm{in})}\right\}/\,(n-1)$. In general, we can conclude the following. At the iteration $t \to (t+1)$ let the parameters $a\,(t+1)$ and $b\,(t)$ be such that the row sums are fulfilled. Further let,

$$
\begin{aligned}
a_k\,(t+1) &= \max\left\{a_i\,(t+1): \quad i = 1,\ldots,n\right\},\\
\epsilon_{k,\oplus}^{(\mathrm{in})} &= \max\left\{\epsilon_j^{(\mathrm{in})} : f_j^{(\mathrm{in})}\Big(a\,(t+1)\,, b\,(t)\Big) > 0, \quad \forall j \neq k\right\} > 0,\\
\epsilon_{k,\ominus}^{(\mathrm{in})} &= \max\left\{\epsilon_j^{(\mathrm{in})} : f_j^{(\mathrm{in})}\Big(a\,(t+1)\,, b\,(t)\Big) < 0, \quad \forall j \neq k\right\} > 0.
\end{aligned}
\tag{3.39}
$$

Updating the column parameters leads to weight added and removed on the respective column elements and amounts to $2\epsilon^{(\text{in})}_{t+1,t}$ in absolute terms. Let $\gamma_k 2\epsilon^{(\text{in})}_{t+1,t}$, with $\gamma_k \in (0,1)$ denote the part of the weight change taking place in row k. Hence, the error induced in all other row sums (besides row k) is bounded from above by $(1-\gamma_k)\, 2\epsilon^{(\text{in})}_{t+1,t}$. The global error $\epsilon^{(\text{out})}_{t+1,t+1}$, induced in all row sums, is decreasing by at least $\min\left\{\epsilon^{(\text{in})}_{k,\oplus}, \epsilon^{(\text{in})}_{k,\ominus}\right\} / (n-1)$,

$$
\begin{aligned}
2\epsilon^{(\text{out})}_{t+1,t+1} &= \sum_{i=1}^{n} \left| f_i^{(\text{out})}\Big(a\,(t+1)\,, b\,(t+1)\Big)\right| \\
&\leq (1-\gamma_k)\, 2\epsilon^{(\text{in})}_{t+1,t} + \left| f_k^{(\text{out})}\Big(a\,(t+1)\,, b\,(t+1)\Big)\right| \\
&= (1-\gamma_k)\, 2\epsilon^{(\text{in})}_{t+1,t} + \left| \sum_{\substack{j=1\\ j\neq k}}^{n} \frac{a_k\,(t+1)\, b_j\,(t+1)}{1-a_k\,(t+1)\,, b_j\,(t+1)} - \frac{a_k\,(t+1)\, b_j\,(t)}{1-a_k\,(t+1)\, b_j\,(t)}\right| \\
&\leq (1-\gamma_k)\, 2\epsilon^{(\text{in})}_{t+1,t} + \gamma_k 2\epsilon^{(\text{in})}_{t+1,t} \\
&\qquad - 2\min\left\{\frac{a_k\,(t+1)\, b_{k,\oplus}\,(t)}{1-a_k\,(t+1)\, b_{k,\oplus}\,(t)} - \frac{a_k\,(t+1)\, b_{k,\oplus}\,(t+1)}{1-a_k\,(t+1)\, b_{k,\oplus}\,(t+1)},\right. \\
&\qquad\qquad \left.\frac{a_k\,(t+1)\, b_{k,\ominus}\,(t+1)}{1-a_k\,(t+1)\, b_{k,\ominus}\,(t+1)} - \frac{a_k\,(t+1)\, b_{k,\ominus}\,(t)}{1-a_k\,(t+1)\, b_{k,\ominus}\,(t)}\right\} \\
&\leq 2\epsilon^{(\text{in})}_{t+1,t} - \frac{2\min\left\{\epsilon^{(\text{in})}_{k,\oplus}, \epsilon^{(\text{in})}_{k,\ominus}\right\}}{n-1}.
\end{aligned}
\tag{3.40}
$$

The analogous holds when updating the row parameters.

(4) Hence, there are only two cases left to discuss.

(a) There is only one column (resp. row) with excess weight, which has the same index as the highest row (resp. column) parameter, and there is at least one column (resp. row) missing weight.

(b) There is only one column (resp. row) missing weight, which has the same index as the highest row (resp. column) parameter, and there is at least one column (resp. rows) with excess weight.

Case (4a) is illustrated in Eq. (3.41).

$$\begin{array}{ccc}
\begin{pmatrix} 0 & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & 0 & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & 0 & w_{34} & w_{35} \\ w_{41} & w_{42} & w_{43} & 0 & w_{45} \\ w_{51} & w_{52} & w_{53} & w_{54} & 0 \end{pmatrix} & \begin{matrix} s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} \\ s_3^{(\mathrm{out})} \\ s_4^{(\mathrm{out})} \\ s_5^{(\mathrm{out})} \end{matrix} & \begin{matrix} a_1 \gg a_2, a_3, a_4, a_5 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{matrix} \\
\begin{matrix} s_1^{(\mathrm{in})} + \epsilon_1 & s_2^{(\mathrm{in})} - \epsilon_2 & s_3^{(\mathrm{in})} - \epsilon_3 & s_4^{(\mathrm{in})} & s_5^{(\mathrm{in})} \end{matrix} & & \\
\begin{matrix} b_1 \downarrow & b_2 \uparrow & b_3 \uparrow & b_4 & b_5 \end{matrix} & &
\end{array} \tag{3.41}$$

Increasing the column parameters where weight is missing leads to an increase of weight mostly at the elements of row 1 at the concerned columns, while decreasing the column parameter of the column with excess weight can affect all other rows, excluding row 1. Therefore, we cannot identify a lower bound for the error that is canceling out. In this case the parameters are updated at follows. Let $\epsilon_1^{(\mathrm{in})}$ be fixed to denote the error of the first column sum at this precise iteration. First, only b_1 is decreased, such that the first column sum looses weight of $(1/3)\,\epsilon_1^{(\mathrm{in})}$. Afterwards, $a_2, \ldots, a_n$ are increased until the row sums are met again. This process is visualized in Eq. (3.42) and Eq. (3.43).

$$\begin{array}{ccc}
\begin{pmatrix} 0 & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & 0 & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & 0 & w_{34} & w_{35} \\ w_{41} & w_{42} & w_{43} & 0 & w_{45} \\ w_{51} & w_{52} & w_{53} & w_{54} & 0 \end{pmatrix} & \begin{matrix} s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} \\ s_3^{(\mathrm{out})} \\ s_4^{(\mathrm{out})} \\ s_5^{(\mathrm{out})} \end{matrix} & \begin{matrix} a_1 \gg a_2, a_3, a_4, a_5 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{matrix} \\
\begin{matrix} s_1^{(\mathrm{in})} + \epsilon_1^{(\mathrm{in})} & s_2^{(\mathrm{in})} - \epsilon_2^{(\mathrm{in})} & s_3^{(\mathrm{in})} - \epsilon_3 & s_4^{(\mathrm{in})} & s_5^{(\mathrm{in})} \end{matrix} & & \\
\begin{matrix} b_1 \downarrow & b_2 & b_3 & b_4 & b_5 \end{matrix} & &
\end{array} \tag{3.42}$$

$$
\begin{array}{ccc}
\begin{pmatrix}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{pmatrix}
&
\begin{array}{c}
s_1^{(\text{out})} \\
s_2^{(\text{out})} - \epsilon_2^{(\text{out})} \\
s_3^{(\text{out})} - \epsilon_3^{(\text{out})} \\
s_4^{(\text{out})} - \epsilon_4^{(\text{out})} \\
s_5^{(\text{out})} - \epsilon_5^{(\text{out})}
\end{array}
&
\begin{array}{c}
a_1 \gg a_2, a_3, a_4, a_5 \\
a_2 \uparrow \\
a_3 \uparrow \\
a_4 \uparrow \\
a_5 \uparrow
\end{array}
\\
\begin{array}{ccccc}
s_1^{(\text{in})} & s_2^{(\text{in})} & s_3^{(\text{in})} & s_4^{(\text{in})} & s_5^{(\text{in})} \\
 & & & & \\
b_1 & b_2 & b_3 & b_4 & b_5
\end{array}
& &
\end{array}
\tag{3.43}
$$

Part of the error might cancel out, because the elements of columns 2 and 3, that are missing weight, are now gaining weight from the increase of the row parameters. In the following, we will further discuss what happens, if the error does not decrease, i.e. increasing the row parameters leads mostly to an increase of weight in the other columns sums. This happens for example if $b_4 \gg b_2, b_3$.

Before we start explaining the dynamics of the new updating procedure, we note that the product of the parameters $a_i b_j$ (for $i \neq j$) is always bounded from above by,

$$
\begin{aligned}
\frac{a_i b_j}{1 - a_i b_j} < s^{(\text{total})} \quad &\Leftrightarrow \quad a_i b_j < s^{(\text{total})} - s^{(\text{total})} a_i b_j \\
&\Leftrightarrow \quad a_i b_j \left(1 + s^{(\text{total})}\right) < s^{(\text{total})} \\
&\Leftrightarrow \quad a_i b_j < \frac{s^{(\text{total})}}{1 + s^{(\text{total})}}.
\end{aligned}
\tag{3.44}
$$

Let $a_l = \max\{a_i : i = 1, \ldots, n \wedge i \neq 1\}$ denote the second highest row parameter. When b_1 is upated, the element w_{l1} has to lose weight of at least $\epsilon_1^{(\text{in})} / (3(n-1))$. Therefore, a lower bound for the decrease of parameter $b_1(t) \searrow b_1(t+1)$ can be defined,

$$
\begin{aligned}
& \frac{a_l(t+1)\, b_1(t)}{1 - a_l(t+1)\, b_1(t)} - \frac{a_l(t+1)\, b_1(t+1)}{1 - a_l(t+1)\, b_1(t+1)} \geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \\
\Leftrightarrow \quad & \frac{a_l(t+1)\, b_1(t) - a_l(t+1)\, b_1(t+1)}{\left(1 - a_l(t+1)\, b_1(t)\right)\left(1 - a_l(t+1)\, b_1(t+1)\right)} \geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \\
\Leftrightarrow \quad & \left[b_1(t) - b_1(t+1)\right] a_l(t+1) \geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \left(1 - a_l(t+1)\, b_1(t)\right) \\
& \qquad \left(1 - a_l(t+1)\, b_1(t+1)\right),
\end{aligned}
\tag{3.45}
$$

and

$$\begin{aligned}
\left[b_1(t) - b_1(t+1)\right] a_l(t+1) &\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \left(1 - a_l(t+1) b_1(t)\right)\left(1 - a_l(t+1) b_1(t+1)\right) \\
&\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \left(1 - a_l(t+1) b_1(t)\right)^2 \\
&\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)} \left(1 - \frac{s^{(\text{total})}}{1 + s^{(\text{total})}}\right)^2, \quad \text{with Eq. (3.44)} \\
&\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)\left(1 + s^{(\text{total})}\right)^2} \\
\Leftrightarrow \qquad b_1(t) - b_1(t+1) &\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)\left(1 + s^{(\text{total})}\right)^2 a_l(t+1)}.
\end{aligned} \tag{3.46}$$

Moreover, since

$$a_l(t+1) < \left(\max\{b_j : j = 1, \ldots, n \wedge j \neq l\}\right)^{-1} \leq \left(\max\{b_j : j = 1, \ldots, n \wedge j \neq 1, l\}\right)^{-1}$$

holds, we can further conclude,

$$\begin{aligned}
b_1(t) - b_1(t+1) &\geq \frac{\epsilon_1^{(\text{in})}}{3(n-1)\left(1 + s^{(\text{total})}\right)^2 a_l(t+1)} \\
&> \frac{\epsilon_1^{(\text{in})} \max\{b_j : j = 2, \ldots, n \wedge j \neq l\}}{3(n-1)\left(1 + s^{(\text{total})}\right)^2} \\
&> \frac{\epsilon_1^{(\text{in})} \min\{b_j : j = 2, \ldots, n\}}{3(n-1)\left(1 + s^{(\text{total})}\right)^2}.
\end{aligned} \tag{3.47}$$

Note, that since we do not change $b_2, \ldots, b_n$, the right hand side of Eq. (3.47) defines a constant. This means that the performed iterations lead to a non-negligible decrease in b_1 in each step. As b_1 decreases $a_2, \ldots, a_n$ increase. Note that for $a_l(t+1) = \max\{a_i(t+1) : i = 1, \ldots, n \wedge i \neq 1\}$, after updating the column parameter $b_1(t) \searrow b_1(t+1)$, the element w_{l1} has to carry weight of at least $s_1^{(\text{in})}/(n-1)$. Thus, the following holds,

$$\begin{aligned}
\frac{a_l(t+1) b_1(t+1)}{1 - a_l(t+1) b_1(t+1)} &\geq \frac{s_1^{(\text{in})}}{n-1} \\
\Leftrightarrow \qquad a_l(t+1) b_1(t+1) &\geq \frac{s_1^{(\text{in})}/(n-1)}{1 + s_1^{(\text{in})}/(n-1)} \\
\Leftrightarrow \qquad a_l(t+1) &\geq \frac{s_1^{(\text{in})}}{\left(n - 1 + s_1^{(\text{in})}\right) b_1(t+1)}.
\end{aligned} \tag{3.48}$$

Hence, as the decrease in b_1 is non-negligible, the increase in a_l is also non-negligible. For that reason, after a finite number of iterations one of the following three cases will happen first.

(i) One of the increasing row parameters will exceed a_1. If, furthermore, the new highest row parameter has a different index than the column with the highest amount of excess weight (i.e. column 1) and the highest amount of missing weight, we can return to the standard procedure, updating all column parameters, knowing that the error will decrease by a non-negligible amount. Otherwise, we continue the procedure until (ii) or (iii) occurs.

(ii) The parameter b_1 will fall below two of the others column parameters. Let k_1 and k_2 denote the indices of the new two highest column parameters. At this point, increasing the row parameters leads to additional weight of at least $\epsilon_1^{(\text{in})} / \left(3\,(n-1)^2\right)$ accumulating at column k, with $k \in \{k_1, k_2\}$. If column k is one of the columns that is missing weight, a substantial part of the error is canceling out. Otherwise we end up having two columns with excess weight, column 1 and column k. If, furthermore, there are at least two columns missing weight, we can return to the standard procedure, knowing from part (3) of the proof that the error is decreasing substantially. Otherwise, if there is only one column missing weight, and furthermore, the highest row parameter has a different index than the only column that is missing weight, we can return to the standard procedure, knowing from part (3) of the proof that the error is decreasing substantially. The last case, where the only column that is missing weight has the same index as the highest row parameter equals case (4b) and is treated in the following.

(iii) The error in the first column sum falls below $(1/2)\,\epsilon_1^{(\text{in})}$. The amount of weight which has been removed from the first column sum must have been allocated at one of the other columns. More precisely, there must exist a column $k \neq 1$ that received weight of at least $(1/2)\,\epsilon_1^{(\text{in})} / (n-1)$. If column k is one of the columns that is missing weight, a substantial part of the error is canceling out. Otherwise we end up having two columns with excess weight, column 1 and column k. If, furthermore, there are at least two columns missing weight, we can return to the standard procedure, knowing from part (3) of the proof that the error is decreasing substantially. Otherwise, if there is only one column missing weight, and furthermore, the highest row parameter has a different index than the only column that is missing weight, we can return to the standard procedure, knowing from part (3) of the proof that the error is decreasing substantially. The last case, where the only column that is missing weight has the same index as the highest row parameter equals case (4b) and is treated in the following.

Case (4b) is illustrated in Eq. (3.49).

$$\begin{array}{ccc}
\left(\begin{array}{ccccc}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{array}\right) &
\begin{array}{c} s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} \\ s_3^{(\mathrm{out})} \\ s_4^{(\mathrm{out})} \\ s_5^{(\mathrm{out})} \end{array} &
\begin{array}{c} a_1 \gg a_2, a_3, a_4, a_5 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} \\
\begin{array}{ccccc} s_1^{(\mathrm{in})} - \epsilon_1 & s_2^{(\mathrm{in})} + \epsilon_2 & s_3^{(\mathrm{in})} + \epsilon_3 & s_4^{(\mathrm{in})} & s_5^{(\mathrm{in})} \end{array} & & \\
\begin{array}{ccccc} b_1 \uparrow & b_2 \downarrow & b_3 \downarrow & b_4 & b_5 \end{array} & &
\end{array} \tag{3.49}$$

We proceed similar to case (4a). First, b_1 is increased, such that column 1 gains weight of $\epsilon_1^{(\mathrm{in})}/2$, and subsequently the row parameters $a_2, \ldots, a_5$ are decreased, while a_1 and $b_2, \ldots, b_5$ stay constant, compare the illustration in Eqs. (3.50) and (3.51).

$$\begin{array}{ccc}
\left(\begin{array}{ccccc}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{array}\right) &
\begin{array}{c} s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} \\ s_3^{(\mathrm{out})} \\ s_4^{(\mathrm{out})} \\ s_5^{(\mathrm{out})} \end{array} &
\begin{array}{c} a_1 \gg a_2, a_3, a_4, a_5 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} \\
\begin{array}{ccccc} s_1^{(\mathrm{in})} - \epsilon_1^{(\mathrm{in})} & s_2^{(\mathrm{in})} + \epsilon_2^{(\mathrm{in})} & s_3^{(\mathrm{in})} + \epsilon_3^{(\mathrm{in})} & s_4^{(\mathrm{in})} & s_5^{(\mathrm{in})} \end{array} & & \\
\begin{array}{ccccc} b_1 \uparrow & b_2 & b_3 & b_4 & b_5 \end{array} & &
\end{array} \tag{3.50}$$

$$\begin{array}{ccc}
\left(\begin{array}{ccccc}
0 & w_{12} & w_{13} & w_{14} & w_{15} \\
w_{21} & 0 & w_{23} & w_{24} & w_{25} \\
w_{31} & w_{32} & 0 & w_{34} & w_{35} \\
w_{41} & w_{42} & w_{43} & 0 & w_{45} \\
w_{51} & w_{52} & w_{53} & w_{54} & 0
\end{array}\right) &
\begin{array}{c} s_1^{(\mathrm{out})} \\ s_2^{(\mathrm{out})} + \epsilon_2^{(\mathrm{out})} \\ s_3^{(\mathrm{out})} + \epsilon_3^{(\mathrm{out})} \\ s_4^{(\mathrm{out})} + \epsilon_4^{(\mathrm{out})} \\ s_5^{(\mathrm{out})} + \epsilon_5^{(\mathrm{out})} \end{array} &
\begin{array}{c} a_1 \gg a_2, a_3, a_4, a_5 \\ a_2 \downarrow \\ a_3 \downarrow \\ a_4 \downarrow \\ a_5 \downarrow \end{array} \\
\begin{array}{ccccc} s_1^{(\mathrm{in})} & s_2^{(\mathrm{in})} & s_3^{(\mathrm{in})} & s_4^{(\mathrm{in})} & s_5^{(\mathrm{in})} \end{array} & & \\
\begin{array}{ccccc} b_1 & b_2 & b_3 & b_4 & b_5 \end{array} & &
\end{array} \tag{3.51}$$

As explained in Eqs. (3.45) and (3.46), a lower bound for the change in b_1 can be derived,

$$b_1(t+1) - b_1(t) \geq \frac{\epsilon_1^{(\mathrm{in})}}{2(n-1)\left(1+s^{(\mathrm{total})}\right)^2 a_l(t+1)}, \tag{3.52}$$

where $a_l = \max\{a_i : i = 1, \ldots, n \wedge i \neq 1\} < \left(\max\{b_j : j = 1, \ldots, n \wedge j \neq l\}\right)^{-1}$. Furthermore, analogously to Eq. (3.48) an upper bound for the value of a_i for all $i = 2, \ldots, 5$ can be defined,

$$\frac{a_i(t+1)\, b_1(t)}{1 - a_i(t)\, b_1(t)} \leq s_i^{(\mathrm{out})} \quad \Leftrightarrow \quad a_i(t+1) \leq \frac{s_i^{(\mathrm{out})}}{\left(1+s_i^{(\mathrm{out})}\right) b_1(t)}. \tag{3.53}$$

Hence, as long as the decrease in the error term is negligible, b_1 keeps increasing by a non-negligible amount and all row parameters $a_2, \ldots, a_n$ keep decreasing by a non-negligible amount.

After a finite number of iterations, one of the following cases will occur. Let $b_k = \max\{b_j : j = 1, \ldots, n\}$ denote the highest column parameter.

(i) For $l \neq k$ and column k having excess weight, increasing b_1 and decreasing a_l leads to a reduction of weight of at least $\epsilon_1^{(\mathrm{in})}/2(n-1)^2$ at the element w_{lk}. Hence, the decrease in the global error is non-negligible.

(ii) For $l \neq k$ and column k being without error term, increasing b_1 and decreasing a_l leads to a reduction of weight of at least $\epsilon_1^{(\mathrm{in})}/2(n-1)^2$ at the element w_{lk}. Hence, returning to the standard procedure, updating all column parameters leads to a non-negligible part of the error canceling out in row 1, as increasing b_k leads to an increase in weight of at least $\epsilon_k^{(\mathrm{in})}/(n-1)$ at the element w_{1k}, and decreasing the column parameters b_j, for column j having excess weight, leads to a decrease of weight of at least $\epsilon_j^{(\mathrm{in})}/(n-1)$ at the element w_{1j}.

(iii) For $l = k$, we continue increasing b_1 and decreasing $a_2, \ldots, a_n$. When b_1 is increased, row k perceives the highest increase of weight, hence, in turn a_k will subsequently be decreased more than the other row parameters. Therefore, at some point, a_k will not be the highest row parameter any more and one of the other cases proceeds.

(iv) For $l \neq k$ and $k = 1$, we continue increasing b_1 and decreasing $a_2, \ldots, a_n$. As the column parameters $b_2, \ldots, b_n$ stay constant, there exists a finite number $N \in \mathbb{N}$ of iterations, such that for every $\delta \in \mathbb{R}_{>0}$, it holds that $w_{ij} < \delta$ for all $i, j = 2, \ldots, n$. This means that the entire weight of the row sums $s_2^{(\mathrm{out})}, \ldots, s_5^{(\mathrm{out})}$ is accumulated in the first column. Since the second condition of the theorem, compare Eq. (3.8), requires that

$$s_1^{(\text{in})} < \sum_{\substack{i=1 \\ i \neq 1}}^{n} s_i^{(\text{out})}, \tag{3.54}$$

holds, the error of the first column can at most amount to $\sum_{\substack{i,j=2 \\ i \neq j}}^{n} w_{ij}$. Since this sum converges to zero, the error in the first column sum, consequently, has to converge to zero as well.

In fact, whenever the algorithm does not converge, we know that the given in- and out-strength sequences are infeasible. □

Figures 3.1 to 3.3 demonstrate the performance of the algorithm proposed above. For feasible in- and out strength sequences, the algorithm converges quickly to an error of zero, compare Fig. 3.1. In contrast, if the first condition of Theorem 3.1.1, i.e. Eq. (3.7), is not fulfilled, the error converges exactly to the difference of the sum of the in-strength sequence and the sum of the out-strength sequence, i.e. to the amount that is impossible to satisfy, compare Fig. 3.2. Similarly, if the second condition of Theorem 3.1.1, i.e. Eq. (3.8), is not fulfilled, the error again converges exactly to the amount of weight that is impossible to satisfy, compare Fig. 3.3.

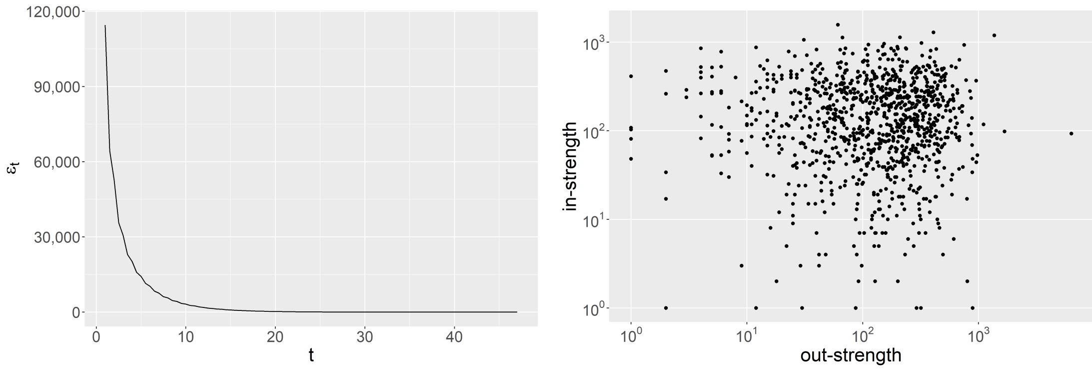


Figure 3.1: Performance of the algorithm proposed in the proof of Theorem 3.1.1 for an exemplary desired in- and out-strength sequence of dimension $n = 1,000$. The plot on the left shows the decrease in the sum of absolute errors ϵ_t, where t denotes the iterations. Here, one iteration includes updating both the row and the column parameters once. The plot on the right shows the desired in- and out-strength sequence.

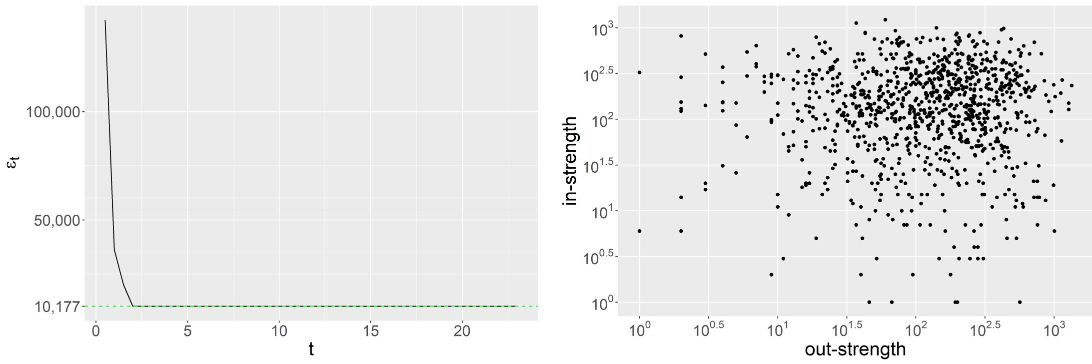


Figure 3.2: Performance of the algorithm proposed in the proof of Theorem 3.1.1 for an exemplary desired in- and out-strength sequence of dimension $n = 1,000$, that do not satisfy the first condition of Eq. (3.7), since $\sum_{i=1}^{1,000} s_i^{(\text{in})} - \sum_{i=1}^{1,000} s_i^{(\text{out})} = 10,177$. The plot on the left shows the decrease in the sum of absolute errors ϵ_t, where t denotes the iterations. Here, one iteration includes updating both the row and the column parameters once. The error converges to the difference between the sum of the in-strength sequence and the sum of the out-strength sequence, that is impossible to meet. The plot on the right shows the desired in- and out-strength sequence.

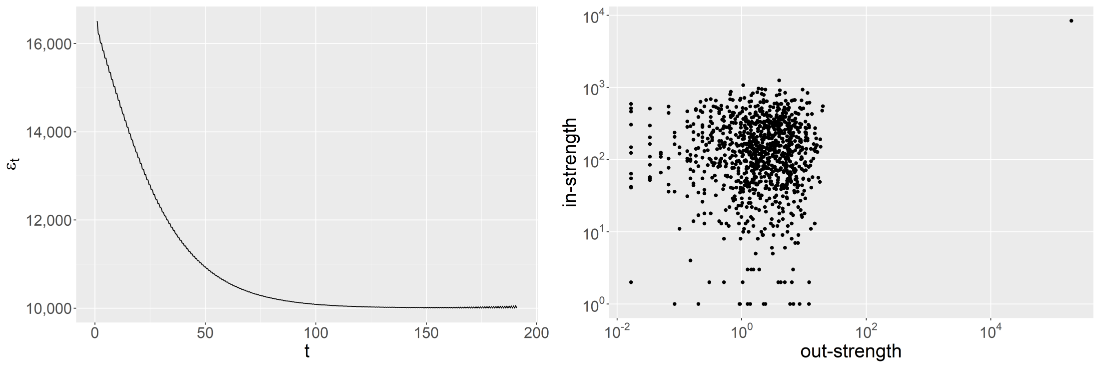


Figure 3.3: Performance of the algorithm proposed in the proof of Theorem 3.1.1 for an exemplary desired in- and out-strength sequence of dimension $n = 1,000$, that do not satisfy the second condition of Eq. (3.8), since $s_1^{(\text{in})} = 8,356 > 3,356 = \sum_{i=2}^{1,000} s_i^{(\text{out})}$ and $s_1^{(\text{out})} = 196,541 > 191,541 = \sum_{i=2}^{1,000} s_i^{(\text{in})}$. The plot on the left shows the decrease in the sum of absolute errors ϵ_t, where t denotes the iterations. Here, one iteration includes updating both the row and the column parameters once. Since the error in the first iteration is extremely high, it is omitted here to provide a clearer picture. The error converges to the following value $\left|s_1^{(\text{in})} - \left(\sum_{i=2}^{1,000} s_i^{(\text{out})}\right)\right| + \left|s_1^{(\text{out})} - \left(\sum_{i=2}^{1,000} s_i^{(\text{in})}\right)\right| = 10,000$, which after updating the row parameters equals the missing weight in the first column plus the excess weight in columns 2 to n. The plot on the right shows the desired in- and out-strength sequence.

3.2 Uniqueness of a Solution

Besides the existence of a solution to the ERGM with desired in- and out-strength sequence, we can also show that the solution is unique up to certain equivalence classes.

Theorem 3.2.1 (Uniqueness of the Solution to the ERGM with Desired In- and Out-Strength Sequence)
The solution to the ERGM with desired in- and out-strength sequence, established by Theorem 3.1.1 and represented by the two vectors $a = (a_1, \dots, a_n) \in \mathbb{R}^n_{>0}$ *and* $b = (b_1, \dots, b_n) \in \mathbb{R}^n_{>0}$ *is unique in the sense that all products* $a_i b_j$ *for* $i, j = 1, \dots, n$ *and* $i \neq j$ *are unique. Hence, for all* $c \in \mathbb{R}_{>0}$, $a \cdot c$ *and* b/c *likewise constitute a solution, describing the same probability distribution that solves the optimization problem given in Eqs.* (3.2) *and* (3.3).

Proof

We have to show that the solution to the ERGM is unique regarding the product of all $a_i b_j$, where $i \neq j$. Recall from Theorem 2.1.3 that the probability distribution P solving the constrained maximum entropy problem is unique. This means that the probability $P(w)$ is unique for all $w \in \mathcal{G}$. Assume there exists a second solution denoted by $\tilde{a}$ and $\tilde{b}$. Moreover, let $w^{(0)}$ denote the graph where all $w_{ij}^{(0)} = 0$. It follows that,

$$
\begin{aligned}
& P_{a,b}\left(w^{(0)}\right) = P_{\tilde{a},\tilde{b}}\left(w^{(0)}\right) \\
\Leftrightarrow \quad & Z_{a,b}^{-1} \underbrace{e^{-H_{a,b}\left(w^{(0)}\right)}}_{=1} = Z_{\tilde{a},\tilde{b}}^{-1} \underbrace{e^{-H_{\tilde{a},\tilde{b}}\left(w^{(0)}\right)}}_{=1} \\
\Leftrightarrow \quad & Z_{a,b} = Z_{\tilde{a},\tilde{b}}.
\end{aligned}
$$

For all $i, j = 1, \ldots, n$ and $i \neq j$, let $w^{(\mathrm{i,j})}$ denote the graph where all elements are 0 and $w_{ij}^{(\mathrm{i,j})} = 1$. This yields,

$$
\begin{aligned}
& P_{a,b}\left(w^{(\mathrm{i,j})}\right) = P_{\tilde{a},\tilde{b}}\left(w^{(\mathrm{i,j})}\right) \\
\Leftrightarrow \quad & Z_{a,b}^{-1} \underbrace{e^{-H_{a,b}\left(w^{(a,b)}\right)}}_{=\exp(-\theta_{ab})} = Z_{\tilde{a},\tilde{b}}^{-1} \underbrace{e^{-H_{\tilde{a},\tilde{b}}\left(w^{(a,b)}\right)}}_{=\exp\left(-\tilde{\theta}_{ab}\right)} \\
\Leftrightarrow \quad & a_i b_j = \tilde{a}_i \tilde{b}_j.
\end{aligned}
$$

Hence, it follows that all products $a_i b_j$ are unique for all $i, j = 1, \ldots, n$ and $i \neq j$.

□

4 Extended Fitness Models

This chapter provides a novel extension to randomized fitness models and mathematical insight into empirical fitness models.

First, we extend the randomized fitness models introduced in Section 2.2.2 to a flexible degree distribution that allows fitting it to real-world networks. The advantage of this model is a more precise reconstruction of scale-free networks, which is achieved by offering more degrees of freedom to fit the desired degree distribution.

Second, the empirical fitness models discussed in Section 2.2.1 are analyzed analytically. The nodes' strength, which are often found to follow a power-law distribution, have been shown to serve remarkably well as fitness variables in the reconstruction of economic and financial networks. We provide mathematical insight into this phenomenon by analyzing the degree distribution induced by combining power-law distributed fitness variables and the commonly used link probability function.

4.1 Fitness Models with Flexible Power-Law Degree Distributions

Many real-world networks seem to exhibit power law distributed degrees, see Clauset et al. (2009). This is especially true for the upper tail of the distribution. The lower tail, however, is often not well described by a pure power law distribution, see e.g. the degree distributions of the German and Italian interbank market, pictured in Figs. 5.1 to 5.3. Another problem with the power law degree distributions of the randomized fitness models derived in Section 2.2.2 is that they offer only two degrees of freedom, the exponent $\alpha \in \mathbb{R}_{<0}$ and either the lower bound d_0 or the upper bound d_∞ of the support. An exception is the fitness model given in Theorem 2.2.4, which for $\alpha \leq -2$ allows three degrees of freedom. It is therefore difficult to fit a given degree distribution to, for example, a desired mean, median and some quantile, as done in Chapter 5 for the German and Italian interbank market. Morover, Chapter 5 shows that fitting the (in- and out-) degree distribution plays a crucial role in network reconstruction, as many further network characteristics seem to follow as a natural consequence.

For these reasons, we suggest a mixture distribution for the degrees, consisting of a uniform distribution in the lower tail and a power law distribution in the upper tail. Let $Y = d(X)$ again denote the random variable representing the degrees of nodes with random fitness variables X. We want the random degrees Y_i, for $i = 1, \dots, n$, of a network with n nodes, to have the following density function,

$$\rho_Y(y) = c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \le y \le L\}} + y^{-\alpha} \times \mathbb{1}_{\{L < y \le d_\infty\}}\right). \tag{4.1}$$

The parameter c acts as normalization constant, α is the exponent of the power law distribution, and L divides the support into the initial uniform and subsequent power law distributed part. In comparison to the degree distribution of the models discussed in Section 2.2.2, the distribution in Eq. (4.1) provides an additional degree of freedom, and hence, facilitates fitting to desired characteristics.

The parameter c can be derived via the normalization condition $\int_{d_0}^{d_\infty} \rho_Y(y)\,\mathrm{d}y = 1$. For $\alpha = 1$,

$$\begin{aligned} & 1 = \int_{d_0}^{d_\infty} c\left(L^{-1} \times \mathbb{1}_{\{d_0 \le y \le L\}} + y^{-1} \times \mathbb{1}_{\{L < y \le d_\infty\}}\right) \mathrm{d}y \\ \Leftrightarrow\quad & 1 = c\left[L^{-1}(L - d_0) + \log(d_\infty) - \log(L)\right] \\ \Leftrightarrow\quad & c = \left[1 - L^{-1} d_0 + \log(d_\infty) - \log(L)\right]^{-1}. \end{aligned} \tag{4.2}$$

For $\alpha \neq 1$,

$$\begin{aligned} & 1 = \int_{d_0}^{d_\infty} c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \le y \le L\}} + y^{-\alpha} \times \mathbb{1}_{\{L < y \le d_\infty\}}\right) \mathrm{d}y \\ \Leftrightarrow\quad & 1 = cL^{-\alpha}(L - d_0) + \frac{c}{1-\alpha}\left[d_\infty^{1-\alpha} - L^{1-\alpha}\right] \\ \Leftrightarrow\quad & c = \left[L^{1-\alpha} - L^{-\alpha} d_0 + \frac{1}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right)\right]^{-1}. \end{aligned} \tag{4.3}$$

Note that for increasing functions d, it follows $d_0 \le L < d_\infty$ and, hence, c is positive for all $\alpha \in \mathbb{R}_{>0}$.

The following theorem extends Theorem 2.2.4 to the more flexible degree distribution of Eq. (4.1).

Theorem 4.1.1 (Fitness model with $f(x_i, x_j) = \tilde{f}(x_i + x_j)$, $X \sim Exp(1)$, and flexible power law distribution with arbitrary exponent α)
Consider a fitness model with exponentially distributed random fitness variables, i.e. $X \sim Exp(1)$. There exists a link probability function $f(x_i, x_j) = \tilde{f}(x_i + x_j)$ that generates networks with degrees distributed according to Eq. (4.1), i.e. such that

$$\rho_Y(d(x)) = c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \le d(x) \le L\}} + d(x)^{-\alpha} \times \mathbb{1}_{\{L < d(x) \le d_\infty\}}\right) \tag{4.4}$$

holds. Moreover, the link probability function f takes the following form

$$f(x_i, x_j) = \frac{d(x_i + x_j) - d'(x_i + x_j)}{\tilde{n}}, \tag{4.5}$$

where $\tilde{n} = n - 1$ for undirected networks and $\tilde{n} = 2(n-1)$ for directed networks. The function d is given by

$$d(x) = \begin{cases} R(x)\, c^{-1} L^{\alpha} + d_0, & \text{if } x \leq t \\ \exp\left[\Big(R(x) - cL^{-1}(L - d_0)\Big)c^{-1} + \log(L)\right], & \text{if } x > t \text{ and } \alpha = 1 \\ \left[\Big(R(x) - cL^{-\alpha}(L - d_0)\Big)\frac{1-\alpha}{c} + L^{1-\alpha}\right]^{\frac{1}{1-\alpha}}, & \text{if } x > t \text{ and } \alpha \neq 1, \end{cases} \tag{4.6}$$

where t denotes the point where $d(t) = L$ holds.
The domain of the power law distribution can be chosen arbitrarily for $\alpha \geq 2$. For $\alpha \in (0,2)$, the domain is subject to the constraint

$$\begin{aligned} \frac{d_\infty}{L} &\leq e, & &\text{for } \alpha = 1, \\ \frac{d_\infty}{L} &\leq (2-\alpha)^{\frac{1}{1-\alpha}}, & &\text{for } \alpha \in (0,1) \cup (1,2). \end{aligned} \tag{4.7}$$

Proof
Recall the function d computing the expected degree of a node with realized fitness x,

$$d(x) = \tilde{n} \int_0^\infty f(x,z)\, \rho_X(z)\, \mathrm{d}z, \tag{4.8}$$

where $\tilde{n} = n - 1$ for undirected networks and $\tilde{n} = 2(n-1)$ for directed networks. As shown in Section 2.2.2, for d strictly monotonically increasing, we know that $d_0 \leq L < d_\infty$ holds and from the transformation theorem of probability densities it follows,

$$\rho_Y(d(x)) = \frac{\rho_X(x)}{d'(x)}. \tag{4.9}$$

Combining Eq. (4.1) and Eq. (4.9) yields

$$\begin{aligned} & \frac{\rho_X(x)}{d'(x)} = c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \leq d(x) \leq L\}} + d(x)^{-\alpha} \times \mathbb{1}_{\{L < d(x) \leq d_\infty\}}\right) \\ \Leftrightarrow \quad & \rho_X(x) = c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \leq d(x) \leq L\}} + d(x)^{-\alpha} \times \mathbb{1}_{\{L < d(x) \leq d_\infty\}}\right) d'(x). \end{aligned} \tag{4.10}$$

Let t be defined such that $d(t) = L$ holds. Integrating both sides of Eq. (4.10) from 0 to x gives

$$
\begin{aligned}
R(x) &:= \int_0^x \rho_X(z)\,\mathrm{d}z \\
&= \int_0^x c\left(L^{-\alpha} \times \mathbb{1}_{\{d_0 \leq d(z) \leq L\}} + d(z)^{-\alpha} \times \mathbb{1}_{\{L < d(z) \leq d_\infty\}}\right) d'(z)\,\mathrm{d}z \\
&= \begin{cases} \int_0^x cL^{-\alpha} d'(z)\,\mathrm{d}z, & \text{if } d_0 \leq d(x) \leq L \\ \int_0^t cL^{-\alpha} d'(z)\,\mathrm{d}z + \int_t^x c\,(d(z))^{-\alpha}\, d'(z)\,\mathrm{d}z, & \text{if } L < d(x) \leq d_\infty \end{cases} \\
&= \begin{cases} \int_0^x cL^{-\alpha} d'(z)\,\mathrm{d}z, & \text{if } x \leq t \\ \int_0^t cL^{-\alpha} d'(z)\,\mathrm{d}z + \int_t^x c\,(d(z))^{-\alpha}\, d'(z)\,\mathrm{d}z, & \text{if } x > t \end{cases} \\
&= \begin{cases} cL^{-\alpha}\left(d(x) - d_0\right), & \text{if } x \leq t \\ cL^{-1}\left(d(t) - d_0\right) + c\left(\log\left(d(x)\right) - \log\left(d(t)\right)\right), & \text{if } x > t \text{ and } \alpha = 1 \\ cL^{-\alpha}\left(d(t) - d_0\right) + \frac{c}{1-\alpha}\left(\left(d(x)\right)^{1-\alpha} - \left(d(t)\right)^{1-\alpha}\right), & \text{if } x > t \text{ and } \alpha \neq 1 \end{cases} \\
&= \begin{cases} cL^{-\alpha}\left(d(x) - d_0\right), & \text{if } x \leq t \\ cL^{-1}\left(L - d_0\right) + c\left(\log\left(d(x)\right) - \log\left(L\right)\right), & \text{if } x > t \text{ and } \alpha = 1 \\ cL^{-\alpha}\left(L - d_0\right) + \frac{c}{1-\alpha}\left(\left(d(x)\right)^{1-\alpha} - L^{1-\alpha}\right), & \text{if } x > t \text{ and } \alpha \neq 1. \end{cases}
\end{aligned}
\tag{4.11}
$$

Solving for $d(x)$ leads to

$$
d(x) = \begin{cases} R(x)\,c^{-1}L^{\alpha} + d_0, & \text{if } x \leq t \\ \exp\left[\left(R(x) - cL^{-1}\left(L - d_0\right)\right)c^{-1} + \log\left(L\right)\right], & \text{if } x > t \text{ and } \alpha = 1 \\ \left[\left(R(x) - cL^{-\alpha}\left(L - d_0\right)\right)\frac{1-\alpha}{c} + L^{1-\alpha}\right]^{\frac{1}{1-\alpha}}, & \text{if } x > t \text{ and } \alpha \neq 1. \end{cases} \tag{4.12}
$$

For exponentially distributed fitness variables $X \sim Exp(1)$, we know from Eqs. (2.85) to (2.88), that

$$
f(x_i, x_j) = \tilde{f}(x_i + x_j) = \frac{d(x_i + x_j) - d'(x_i + x_j)}{\tilde{n}}. \tag{4.13}
$$

Note that $d(x)$ (and hence also $d'(x)$) is completely defined by Eq. (4.12). Since f determines the link probability, we have to ensure that $f(x_i, x_j) \in [0, 1]$ for all $x_i, x_j \in \mathbb{R}_{\geq 0}$. Recalling that $R(x) = \int_0^x \rho_X(z)\,\mathrm{d}z = \int_0^x e^{-z}\mathrm{d}z = 1 - e^{-x}$, we can compute the derivative of d,

$$d'(x) = \begin{cases} e^{-x}c^{-1}L^{\alpha}, & \text{if } x \leq t \\ \exp\left[\Big(R(x) - cL^{-1}(L-d_0)\Big)c^{-1} + \log(L)\right]e^{-x}c^{-1}, & \text{if } x > t \text{ and } \alpha = 1 \\ \frac{1}{1-\alpha}\left[\Big(R(x) - cL^{-\alpha}(L-d_0)\Big)\frac{1-\alpha}{c} + L^{1-\alpha}\right]^{\frac{1}{1-\alpha}-1} e^{-x}\frac{1-\alpha}{c}, & \text{if } x > t \text{ and } \alpha \neq 1. \end{cases}$$
$$= \begin{cases} e^{-x}c^{-1}L^{\alpha}, & \text{if } x \leq t \\ d(x)\, e^{-x}c^{-1}, & \text{if } x > t \text{ and } \alpha = 1 \\ d(x)^{\alpha}\, e^{-x}c^{-1}, & \text{if } x > t \text{ and } \alpha \neq 1. \end{cases} \tag{4.14}$$

Furthermore, t can be derived,

$$\begin{aligned} L &= d(t) \\ L &= R(t)\, c^{-1}L^{\alpha} + d_0 \\ (L-d_0)\, cL^{-\alpha} &= 1 - e^{-t} \\ \Leftrightarrow \qquad t &= -\log\Big(1 - \Big(L - d_0\Big)cL^{-\alpha}\Big). \end{aligned} \tag{4.15}$$

We start with the case of $x \leq t$. The link probability function $\tilde{f}$ is then given by

$$\tilde{f}(x) = \frac{1}{\tilde{n}}\left[d(x) - e^{-x}c^{-1}L^{\alpha}\right]. \tag{4.16}$$

Since d is increasing in x, the same holds for $\tilde{f}$. Therefore, it suffices to consider the lower and upper bound of the support of x to ensure that $\tilde{f}(x) \in [0,1]$ holds. The lower bound yields,

$$\begin{aligned} & 0 \leq \lim_{x\to 0}\tilde{f}(x) \\ \Leftrightarrow\quad & 0 \leq d_0 - c^{-1}L^{\alpha} \\ \Leftrightarrow\quad & 0 \leq \begin{cases} d_0 - \left[1 - L^{-1}d_0 + \log(d_\infty) - \log(L)\right]L, & \text{for } \alpha = 1\text{, by Eq. (4.2)} \\ d_0 - \left[L^{1-\alpha} - L^{-\alpha}d_0 + \frac{1}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right)\right]L^{\alpha}, & \text{for } \alpha \neq 1\text{, by Eq. (4.3)} \end{cases} \\ \Leftrightarrow\quad & 0 \leq \begin{cases} 2d_0 - L - L\log\left(\frac{d_\infty}{L}\right), & \text{for } \alpha = 1 \\ d_0 - L + d_0 - \frac{L^{\alpha}}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right), & \text{for } \alpha \neq 1 \end{cases} \\ \Leftrightarrow\quad & d_0 \geq \begin{cases} \frac{L}{2}\left(1 + \log\left(\frac{d_\infty}{L}\right)\right), & \text{for } \alpha = 1 \\ \frac{L}{2}\left(1 + \frac{L^{\alpha-1}}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right)\right), & \text{for } \alpha \neq 1. \end{cases} \end{aligned} \tag{4.17}$$

Since $d_0 \leq L$ has to hold, Eq. (4.17) implies for $\alpha = 1$

$$2 \geq 1 + \log\left(\frac{d_\infty}{L}\right) \quad \Leftrightarrow \quad eL \geq d_\infty. \tag{4.18}$$

Analogously, Eq. (4.17) implies for $\alpha \in (0,1)$

$$\begin{aligned} 2 \geq 1 + \frac{L^{\alpha-1}}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right) \quad &\Leftrightarrow \quad 1-\alpha \geq \left(\frac{d_\infty}{L}\right)^{1-\alpha} - 1 \\ &\Leftrightarrow \quad (2-\alpha)^{\frac{1}{1-\alpha}} \geq \frac{d_\infty}{L}. \end{aligned} \tag{4.19}$$

Analogously, Eq. (4.17) implies for $\alpha \geq 1$

$$2 \geq 1 + \frac{L^{\alpha-1}}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right) \quad \Leftrightarrow \quad 2-\alpha \leq \left(\frac{d_\infty}{L}\right)^{1-\alpha}, \tag{4.20}$$

which is fulfilled for $\alpha \geq 2$ and which leads to the same condition as in Eq. (4.19) for $\alpha \in (1,2)$.

Regarding the upper bound of $\tilde{f}$, for $x \leq t$, we get

$$\begin{aligned} 1 \geq \lim_{x\to\infty} \tilde{f}(x) &= \lim_{x\to\infty} \frac{1}{\tilde{n}}\left[d(x) - e^{-x}c^{-1}L^{\alpha}\right] \\ \Leftrightarrow \quad \tilde{n} &\geq d_\infty, \end{aligned} \tag{4.21}$$

which is satisfied by the definition of d_∞.

In the case of $x > t$ and $\alpha = 1$, the link probability function $\tilde{f}$ is then given by

$$\tilde{f}(x) = \frac{1}{\tilde{n}}\left[d(x) - d(x)\,e^{-x}c^{-1}\right] = \frac{1}{\tilde{n}}\left[d(x)\left(1 - e^{-x}c^{-1}\right)\right]. \tag{4.22}$$

Since d is increasing in x, the same holds for $\tilde{f}$. Therefore, it suffices to consider the lower and upper bound of the support of x to ensure that $\tilde{f}(x) \in [0,1]$ holds. The lower bound yields,

$$\begin{aligned} 0 \leq \tilde{f}(t) \quad &\Leftrightarrow \quad 0 \leq 1 - e^{-t}c^{-1} \\ &\Leftrightarrow \quad 0 \leq 1 - \left(1 - \left(L - d_0\right)cL^{-1}\right)c^{-1}, \quad \text{by Eq. (4.15)} \\ &\Leftrightarrow \quad 0 \leq 1 - c^{-1} + \left(L - d_0\right)L^{-1} \\ &\Leftrightarrow \quad 0 \leq 1 - \left[1 - L^{-1}d_0 + \log(d_\infty) - \log(L)\right] + 1 - d_0L^{-1}, \quad \text{by Eq. (4.2)} \\ &\Leftrightarrow \quad 0 \leq 1 - \log\left(\frac{d_\infty}{L}\right) \\ &\Leftrightarrow \quad d_\infty \leq eL. \end{aligned} \tag{4.23}$$

Regarding the upper bound of $\tilde{f}$, for $x > t$ and $\alpha = 1$, we get

$$\begin{aligned} 1 &\geq \lim_{x\to\infty} \tilde{f}(x) = \lim_{x\to\infty} \frac{1}{\tilde{n}} \left[d(x) \left(1 - e^{-x} c^{-1}\right)\right] \\ \Leftrightarrow \quad \tilde{n} &\geq d_\infty, \end{aligned} \tag{4.24}$$

which is satisfied by the definition of d_∞.

In the case of $x > t$ and $\alpha \neq 1$, the link probability function $\tilde{f}$ is then given by

$$\begin{aligned} \tilde{f}(x) &= \frac{1}{\tilde{n}} \left[d(x) - d(x)^\alpha e^{-x} c^{-1} \right] \\ &= \frac{1}{\tilde{n}} d(x) \left[1 - d(x)^{\alpha-1} e^{-x} c^{-1} \right] \\ &= \frac{1}{\tilde{n}} d(x) \left[1 - \left[\left(R(x) - cL^{-\alpha}(L - d_0) \right) \frac{1-\alpha}{c} + L^{1-\alpha} \right]^{-1} e^{-x} c^{-1} \right] \\ &= \frac{1}{\tilde{n}} d(x) \left[1 - \left[\frac{1-\alpha}{c} R(x) - (1-\alpha) L^{-\alpha} (L - d_0) + L^{1-\alpha} \right]^{-1} e^{-x} c^{-1} \right] \\ &= \frac{1}{\tilde{n}} d(x) \Bigg[1 - \Bigg[(1-\alpha) \left[L^{1-\alpha} - L^{-\alpha} d_0 + \frac{1}{1-\alpha} \left(d_\infty^{1-\alpha} - L^{1-\alpha} \right) \right] \left(1 - e^{-x}\right) \\ &\qquad\qquad - (1-\alpha) L^{-\alpha} (L - d_0) + L^{1-\alpha} \Bigg]^{-1} e^{-x} c^{-1} \Bigg] \\ &= \frac{1}{\tilde{n}} d(x) \Bigg[1 - \Bigg[\left[(1-\alpha) L^{-\alpha} (L - d_0) + \left(d_\infty^{1-\alpha} - L^{1-\alpha} \right) \right] \left(1 - e^{-x}\right) \\ &\qquad\qquad - (1-\alpha) L^{-\alpha} (L - d_0) + L^{1-\alpha} \Bigg]^{-1} e^{-x} c^{-1} \Bigg] \\ &= \frac{1}{\tilde{n}} d(x) \left[1 - \left[d_\infty^{1-\alpha} - \left[(1-\alpha) L^{-\alpha} (L - d_0) + \left(d_\infty^{1-\alpha} - L^{1-\alpha} \right) \right] e^{-x} \right]^{-1} e^{-x} c^{-1} \right] \\ &= \frac{1}{\tilde{n}} d(x) \left[1 - \left[d_\infty^{1-\alpha} e^{x} - \left[(1-\alpha) L^{-\alpha} (L - d_0) + \left(d_\infty^{1-\alpha} - L^{1-\alpha} \right) \right] \right]^{-1} c^{-1} \right]. \end{aligned} \tag{4.25}$$

Hence, $\tilde{f}$ is increasing in x. Therefore, it suffices to consider the lower and upper bound of the support of x to ensure that $\tilde{f}(x) \in [0,1]$ holds. The lower bound yields,

$$\begin{aligned}
& 0 \leq \tilde{f}(t) = \frac{1}{\tilde{n}} d(t) \left[1 - d(t)^{\alpha-1} e^{-t} c^{-1}\right] \\
\Leftrightarrow\quad & 0 \leq 1 - L^{\alpha-1} e^{-t} c^{-1} \\
\Leftrightarrow\quad & 0 \leq 1 - L^{\alpha-1} \left(1 - \left(L - d_0\right) c L^{-\alpha}\right) c^{-1}, \quad \text{by Eq. (4.15)} \\
\Leftrightarrow\quad & 0 \leq 1 - L^{\alpha-1} \left(c^{-1} - \left(L - d_0\right) L^{-\alpha}\right) \\
\Leftrightarrow\quad & 0 \leq 1 - L^{\alpha-1} \left(\left[L^{1-\alpha} - L^{-\alpha} d_0 + \frac{1}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right)\right] - L^{1-\alpha} + L^{-\alpha} d_0\right) \\
\Leftrightarrow\quad & 0 \leq 1 - L^{\alpha-1} \left(\frac{1}{1-\alpha}\left(d_\infty^{1-\alpha} - L^{1-\alpha}\right)\right) \\
\Leftrightarrow\quad & 1 \geq \frac{1}{1-\alpha}\left(\left(\frac{d_\infty}{L}\right)^{1-\alpha} - 1\right) \quad \Leftrightarrow \quad (\star).
\end{aligned} \tag{4.26}$$

For $x > t$ and $\alpha \in (0,1)$, we get

$$(\star) \quad \Leftrightarrow \quad (2-\alpha)^{\frac{1}{1-\alpha}} \geq \frac{d_\infty}{L}. \tag{4.27}$$

For $x > t$ and $\alpha > 1$, the lower bound of $\tilde{f}$ yields

$$(\star) \quad \Leftrightarrow \quad 2 - \alpha \leq \left(\frac{d_\infty}{L}\right)^{1-\alpha} \tag{4.28}$$

which is always satisfied for $\alpha \geq 2$, since in that case the left hand side is negative and the right hand side positive. For $\alpha \in (1,2)$, we get the same condition as in Eq. (4.27).

Regarding the upper bound of $\tilde{f}$, for $x > t$ and $\alpha \neq 1$, we get

$$\begin{aligned}
& 1 \geq \lim_{x\to\infty} \tilde{f}(x) = \frac{1}{\tilde{n}} d(x) \left[1 - d(x)^{\alpha-1} e^{-x} c^{-1}\right] \\
\Leftrightarrow\quad & \tilde{n} \geq d_\infty,
\end{aligned} \tag{4.29}$$

which is satisfied by the definition of d_∞. □

Remark 4.1.2
Note that by choosing $d_0 = L$ in Theorem 4.1.1, the fitness model of Theorem 2.2.4 is recovered.

4.2 Fitness Models with Power-Law Distributed Fitness Variables

Empirical fitness models, as discussed in Section 2.2.1, have successfully been used to reconstruct economic and financial networks, see Garlaschelli and Loffredo (2004a, 2008); Cimini et al. (2015). However, the underlying mathematical structures have not yet been analyzed. The nodes' strengths, which are typically used as empirical fitness variables, often seem to follow a power law distribution, i.e $\rho_X(x) = cx^{\alpha}$, see Clauset et al. (2009). Moreover, a link probability function corresponding to the ERGM conditioned on the degree sequence is chosen, i.e. $f(x_i, x_j) = \frac{\psi^2 x_i x_j}{\psi^2 x_i x_j + 1}$, where $\psi \in \mathbb{R}_{>0}$ is a constant. Commonly, ψ is calibrated such that the empirical fitness models yields a desired density. Hence, the question arises what kind of degree distribution this model generates.

In this section we make a first step towards an analytical explanation of empirical fitness models by deriving the degree distribution induced by a fitness model with power law distributed fitness variables and a link probability function of the form $f(x_i, x_j) = \frac{\psi^2 x_i x_j}{\psi^2 x_i x_j + 1}$.

The following theorem summarizes our findings for fitness variables that follow a power law distribution with exponent $\alpha = -1$.

Theorem 4.2.1 (Fitness model with $\rho_X(x) = cx^{-1}$ and $f(x_i, x_j) = \frac{\psi^2 x_i x_j}{\psi^2 x_i x_j + 1}$)
Consider a fitness model with power law distributed random fitness variables with exponent $\alpha = -1$ defined on the range $[s_{\text{lb}}, s_{\text{ub}}]$, i.e. $\rho_X(x) = cx^{-1}$. Moreover, let the link probability function be given by $f(x_i, x_j) = \frac{\psi^2 x_i x_j}{\psi^2 x_i x_j + 1}$. The degree distribution ρ_Y of the generated networks will then take the following form

$$\rho_Y(d(x)) = \frac{s_{\text{ub}} - s_{\text{lb}}}{\tilde{n}} \left[s_{\text{lb}} \left(1 - \exp\left(\frac{d(x)}{\tilde{n}c}\right)\right) + s_{\text{ub}} \left(1 - \exp\left(-\frac{d(x)}{\tilde{n}c}\right)\right)\right]^{-1}, \quad (4.30)$$

where $\tilde{n} = n - 1$ for undirected networks and $\tilde{n} = 2(n-1)$ for directed networks.

Proof
The normalization parameter c of the power law distribution ρ_X is given by

$$1 = \int_{s_{\text{lb}}}^{s_{\text{ub}}} cx^{-1} \mathrm{d}x \quad \Leftrightarrow \quad c = \left[\log\left(\frac{s_{\text{ub}}}{s_{\text{lb}}}\right)\right]^{-1}. \quad (4.31)$$

Recall the function d computing the expected degree of a node with realized fitness x,

$$\begin{aligned} d(x) &= \tilde{n}\int_{s_{\mathrm{lb}}}^{s_{\mathrm{ub}}} f(x,z)\,\rho_X(z)\,\mathrm{d}z = \tilde{n}\int_{s_{\mathrm{lb}}}^{s_{\mathrm{ub}}} \frac{\psi^2 xz}{\psi^2 xz+1} c z^{-1}\mathrm{d}z \\ &= \tilde{n}c\int_{s_{\mathrm{lb}}}^{s_{\mathrm{ub}}} \frac{\psi^2 x}{\psi^2 xz+1}\mathrm{d}z = \tilde{n}c\left[\log\left(\psi^2 xz+1\right)\right]_{s_{\mathrm{lb}}}^{s_{\mathrm{ub}}} \\ &= \tilde{n}c\log\left(\frac{\psi^2 x s_{\mathrm{ub}}+1}{\psi^2 x s_{\mathrm{lb}}+1}\right). \end{aligned} \tag{4.32}$$

The derivative of $d(x)$ is given by

$$\begin{aligned} d'(x) &= \tilde{n}c\frac{\partial}{\partial x}\left[\log\left(\psi^2 x s_{\mathrm{ub}}+1\right) - \log\left(\psi^2 x s_{\mathrm{lb}}+1\right)\right] \\ &= \tilde{n}c\left(\frac{\psi^2 s_{\mathrm{ub}}}{\psi^2 x s_{\mathrm{ub}}+1} - \frac{\psi^2 s_{\mathrm{lb}}}{\psi^2 x s_{\mathrm{lb}}+1}\right). \end{aligned} \tag{4.33}$$

Next, we can derive the inverse of $d(x)$,

$$\begin{aligned} && d(x) &= \tilde{n}c\log\left(\frac{\psi^2 x s_{\mathrm{ub}}+1}{\psi^2 x s_{\mathrm{lb}}+1}\right) \\ \Leftrightarrow && \exp\left(d(x)\,\tilde{n}^{-1}c^{-1}\right) &= \frac{\psi^2 x s_{\mathrm{ub}}+1}{\psi^2 x s_{\mathrm{lb}}+1} \\ \Leftrightarrow && \exp\left(d(x)\,\tilde{n}^{-1}c^{-1}\right) - 1 &= \psi^2 x s_{\mathrm{ub}} - \exp\left(d(x)\,\tilde{n}^{-1}c^{-1}\right)\psi^2 x s_{\mathrm{lb}} \\ \Leftrightarrow && x &= \frac{\exp\left(d(x)\,\tilde{n}^{-1}c^{-1}\right) - 1}{\psi^2 s_{\mathrm{ub}} - \exp\left(d(x)\,\tilde{n}^{-1}c^{-1}\right)\psi^2 s_{\mathrm{lb}}}. \end{aligned} \tag{4.34}$$

As shown in Section 2.2.2, for d strictly monotonically increasing, the transformation theorem of probability densities yields,

$$\begin{aligned} \rho_Y(d(x)) &= \frac{\rho_X(x)}{d'(x)} = cx^{-1}\left[\tilde{n}c\left(\frac{\psi^2 s_{\mathrm{ub}}}{\psi^2 x s_{\mathrm{ub}}+1} - \frac{\psi^2 s_{\mathrm{lb}}}{\psi^2 x s_{\mathrm{lb}}+1}\right)\right]^{-1} \\ &= \tilde{n}^{-1}\left[\frac{\psi^2 x s_{\mathrm{ub}}}{\psi^2 x s_{\mathrm{ub}}+1} - \frac{\psi^2 x s_{\mathrm{lb}}}{\psi^2 x s_{\mathrm{lb}}+1}\right]^{-1}. \end{aligned} \tag{4.35}$$

Inserting Eq. (4.34) leads to

$$
\begin{aligned}
\rho_Y\left(d\left(x\right)\right) &= \tilde{n}^{-1}\Big[\psi^2 s_{\mathrm{ub}}\left(\psi^2 s_{\mathrm{ub}} + x^{-1}\right)^{-1} - \psi^2 s_{\mathrm{lb}}\left(\psi^2 s_{\mathrm{lb}} + x^{-1}\right)^{-1}\Big]^{-1} \\
&= \tilde{n}^{-1}\left[\psi^2 s_{\mathrm{ub}}\left(\psi^2 s_{\mathrm{ub}} + \frac{\psi^2 s_{\mathrm{ub}} - \exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right)\psi^2 s_{\mathrm{lb}}}{\exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right) - 1}\right)^{-1}\right. \\
&\qquad \left. -\,\psi^2 s_{\mathrm{lb}}\left(\psi^2 s_{\mathrm{lb}} + \frac{\psi^2 s_{\mathrm{ub}} - \exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right)\psi^2 s_{\mathrm{lb}}}{\exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right) - 1}\right)^{-1}\right]^{-1} \\
&= \tilde{n}^{-1}\left[\psi^2 s_{\mathrm{ub}}\frac{\exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right) - 1}{\exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right)\psi^2 s_{\mathrm{ub}} - \exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right)\psi^2 s_{\mathrm{lb}}}\right. \\
&\qquad \left. -\,\psi^2 s_{\mathrm{lb}}\frac{\exp\left(d\left(x\right)\tilde{n}^{-1}c^{-1}\right) - 1}{\psi^2 s_{\mathrm{ub}} - \psi^2 s_{\mathrm{lb}}}\right]^{-1} \\
&= \frac{s_{\mathrm{ub}} - s_{\mathrm{lb}}}{\tilde{n}}\left[s_{\mathrm{ub}}\left(1 - \exp\left(-\frac{d\left(x\right)}{\tilde{n}c}\right)\right) - s_{\mathrm{lb}}\left(\exp\left(\frac{d\left(x\right)}{\tilde{n}c}\right) - 1\right)\right]^{-1}. \quad \square
\end{aligned}
\tag{4.36}
$$

5 Reconstructing the Topology of Financial Networks from Degree Distributions and Reciprocity

As explained in Chapter 1 the reconstructing of topologies of interbank networks constitutes a challenging task, since information on bilateral interbank activities is classified confidential and therefore mostly not available. However, realistic network models of our financial markets are urgently needed for a proper assessment of systemic risk. To tackle this problem, in this chapter, we use an ERGM coupled with flexible in- and out-degree distributions, that are correlated via a Gaussian copula, to reproduce realistic interbank topologies in light of scarce available information. The performance of the model is demonstrated by a reconstruction of the German and the Italian interbank market. These two networks have been chosen, as their central banks, the German Bundesbank and the Banca d'Italia, published detailed empirical manuscripts on their domestic interbank markets, disclosing a number of aggregated network statistics. This allows us to evaluate the goodness of fit of the simulated networks. This chapter is based on our paper Engel et al. (2019b).

5.1 Reconstruction Problem

We start with a brief introduction to the reconstruction problem. We aim at deriving realistic stochastic models for financial networks, relying only on publicly available information and with a focus on the EU. For this reason, in a first step we scrutinized what information is publicly available. To the best of our knowledge, the empirical works by Bargigli et al. (2015) and Roukny et al. (2014) on the Italian and German interbank market, respectively, are the most detailed descriptions of financial networks of EU member states. The authors of both papers report that most network statistics remain very stable over the considered period of 2008–12 and 2002–12, respectively. This stability serves as justification to rely on the presented numbers for today's banking network. A second source of information that is publicly available are the banks' balance sheets, which disclose, for example, total interbank assets and liabilities. These weights, however, refer to the global interbank market, while the network statistics reported by Bargigli et al. (2015) and Roukny et al. (2014) refer to a single country. Moreover, the binary network topology has been identified to drive systemic risk substantially; see Squartini et al. (2013), and is therefore of special interest. Furthermore, weights can be allocated to sampled adjacency matrices subsequently, based on given row and column sums via existing methods as proposed, e.g., by Gandy and Veraart (2017a).

Consider a financial network G^* with $n \in \mathbb{N}$ banks for which only some network statistics $y_1(G^*), \ldots, y_m(G^*)$ are given. Examples of such statistics are the network density, the degree sequence, the reciprocity and the strength sequence. We are interested in tractable stochastic network models that match these available statistics. To the best of our knowledge, the framework of exponential random graphs is the only methodology currently available that can incorporate multiple network statistics; cf. Section 2.3. To keep the ERGM tractable, we condition only on the in- and out-degrees and the reciprocity. More precisely, we take the following characteristics as input:

(i) mean, median, and upper 1% quantile of the in- and out-degree distribution;

(ii) Pearson's correlation coefficient of the in- and out-degree sequence;

(iii) degree reciprocity.

We use (i) and (ii) to sample coupled in- and out-degree sequences. Let the set of possible adjacency matrices representing the networks be denoted by $\mathcal{G} = \left\{G \in \{0,1\}^{n \times n}\right\}$. The in- and out-degree of each node, as well as the number of reciprocal links, constitutes a given network statistic and is represented, for each $i \in \{1, \ldots, m\}$, by a function $G \mapsto y_i(G)$. The desired output of our model is a discrete probability distribution $P : \mathcal{G} \to [0,1]$ that allows fast network sampling and satisfies all statistics in expectation, i.e. the following holds

$$\sum_{G \in \mathcal{G}} P(G) y_i(G) = y_i(G^*), \qquad \forall i \in \{1, \ldots, m\}.$$

These conditions still leave many degrees of freedom for P. A natural additional objective is to distribute the probability mass as "even as possible" on all graphs $G \in \mathcal{G}$, i.e., to minimize the divergence to the uniform distribution. This reflects the fundamental idea by Shannon and others, that if we have no additional information about a network, then every graph in $\mathcal{G}$ should be assigned the same probability. Exponential random graphs (ERGs) offer an elegant approach to model precisely this situation.

To the best of our knowledge, there exists no general set of conditions under which the existence of a solving probability distribution is guaranteed. This is an interesting and complex open question in the wider realm of the theory on ERGMs, which is left for further research. Since regarding most ERGMs a closed form of the parameters of the probability distribution is not known, the parameters are estimated by minimizing the error of the constraints, and thus a probability distribution with minimal error is always found. Moreover, the construction of a probability distribution that reconstructs the network statistics as closely as possible (and as long as the error is within an acceptable range) already constitutes an advancement in the assessment of systemic risk.

5.2 ERGM Conditioned on the In- and Out-Degree Sequence and Reciprocity

For the reconstruction of interbank networks, we opted for an ERGM that is conditioned on the in- and out-degree sequence and the reciprocity, since this incorporates the maximum amount of publicly available information on the one hand, and remains tractable on the other hand. Furthermore, we assume that no specific links are known. However, in case a financial institution or the regulator has partial knowledge of the network, the ERGM can easily be adapted to contain this information. Similar ERGMs conditioned on the in- and out-degree sequence plus the number of reciprocated links of each node have been studied before; see Squartini et al. (2013); Bargigli et al. (2015). However, their information setting differs from ours, since the authors of both references had access to data on the Dutch and the Italian interbank network, respectively, while we rely solely on publicly available information. For this reason, the ERGM considered here is conditioned on the aggregated number of reciprocal links in the network. Furthermore, in contrast to the mentioned references we do not know the in- and out-degree sequence explicitly, but sample them from fitted distributions that are coupled via a Gaussian copula, as explained in Section 5.4.

Let $\mathcal{G} = \{x \in \{0,1\}^{n\times n} : x_{11} = \cdots = x_{nn} = 0\}$ denote the set of possible adjacency matrices without self-loops. The in- (and out-) degree of a node i in a specific graph x can be computed as a simple sum over the ith column (resp. ith row) of the adjacency matrix x, i.e.,

$$k_i^{(\text{in})}(x) = \sum_{j=1, j\neq i}^{n} x_{ji}, \quad k_i^{(\text{out})}(x) = \sum_{j=1, j\neq i}^{n} x_{ij}.$$

Analogously, the number of reciprocal links in a graph x is calculated by $r(x) = \sum_{j\neq i} x_{ij}x_{ji}$. Thus, the Hamiltonian for this situation is given by

$$\begin{aligned} H(x) &= \theta_r r(x) + \sum_i \theta_i^{(\text{in})} k_i^{(\text{in})}(x) + \theta_i^{(\text{out})} k_i^{(\text{out})}(x) \\ &= \sum_{i<j} 2\theta_r x_{ij}x_{ji} + \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}x_{ij} + \{\theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}x_{ji}, \end{aligned}$$

where $(\theta_r, \theta_1^{(\text{in})}, \ldots, \theta_n^{(\text{in})}, \theta_1^{(\text{out})}, \ldots, \theta_n^{(\text{out})}) \in \mathbb{R}^{2n+1}$ denote the corresponding Lagrange multipliers.

Next, we derive the partition function Z, viz.

$$\begin{aligned}
Z &= \sum_{x\in\mathcal{G}} e^{-H(x)} \\
&= \sum_{x\in\mathcal{G}} \prod_{i<j} \exp[-2\theta_r x_{ij}x_{ji} - \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}x_{ij} - \{\theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}x_{ji}] \\
&= \sum_{(x_{12}^*,x_{21}^*)\in\{(0,0),(1,0),(0,1),(1,1)\}} e^{-2\theta_r x_{12}^* x_{21}^* - \{\theta_1^{(\text{out})}+\theta_2^{(\text{in})}\}x_{12}^* - \{\theta_2^{(\text{out})}+\theta_1^{(\text{in})}\}x_{21}^*} \\
&\qquad \times \underbrace{\left[\sum_{x\in\mathcal{G}:(x_{12},x_{21})=(x_{12}^*,x_{21}^*)} \prod_{\substack{i<j \\ (ij)\notin\{(12)\}}} e^{-2\theta_r x_{ij}x_{ji} - \{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}\}x_{ij} - \{\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}x_{ji}} \right]}_{\text{constant for all } (x_{12}^*,x_{21}^*)} \\
&\overset{(*)}{=} \prod_{i<j} \sum_{(x_{ij},x_{ji})\in\{(0,0),(1,0),(0,1),(1,1)\}} e^{-2\theta_r x_{ij}x_{ji} - \{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}\}x_{ij} - \{\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}x_{ji}} \\
&= \prod_{i<j} \left[1 + e^{-\{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}\}} + e^{-\{\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}+\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}} \right],
\end{aligned}$$

where the same algebraic steps are applied to all pairs in $(*)$ as we applied exemplary to the first pair (x_{12}^*, x_{21}^*). Having Z available in closed form enables us to derive an equation system that defines the Lagrange multipliers by taking the derivatives of $F = -\log(Z)$; see Equation (2.19). We get

$$\frac{\partial}{\partial\theta_r} F = 2\sum_{i<j} \frac{e^{-2\theta_r - \{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}+\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}}}{1 + e^{-\{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}\}} + e^{-\{\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}+\theta_j^{(\text{out})}+\theta_i^{(\text{in})}\}}} = \langle r \rangle, \tag{5.1}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta_i^{(\text{in})}} F &= \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{e^{-\{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}+\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\}}}{1 + e^{-\{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}\}} + e^{-\{\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}+\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\}}} \\
&= \langle k_i^{(\text{in})} \rangle,
\end{aligned} \tag{5.2}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta_i^{(\text{out})}} F &= \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{e^{-\left\{\theta_i^{(\text{out})}+\theta_j^{(\text{in})}\right\}} + e^{-2\theta_r - \left\{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}+\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\right\}}}{1 + e^{-\left\{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}\right\}} + e^{-\left\{\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\right\}} + e^{-2\theta_r - \left\{\theta_i^{(\text{in})}+\theta_j^{(\text{out})}+\theta_j^{(\text{in})}+\theta_i^{(\text{out})}\right\}}} \\
&= \langle k_i^{(\text{out})} \rangle,
\end{aligned} \tag{5.3}$$

for all $i \in \{1, \ldots, n\}$, and where $\langle r \rangle$ denotes the desired number of reciprocal links, $\langle k_i^{(\text{in})} \rangle$ the desired number of incoming edges of node i, and $\langle k_i^{(\text{out})} \rangle$ the desired number of outgoing edges of node i.

The above equations already yield the link probabilities. Let X_{ij} denote the random variable representing a link from node i to node j in the random network, then the probabilities for the four different cases that a dyad can take are given by

$$P(X_{ij} = 1 \wedge X_{ji} = 1) = \frac{e^{-2\theta_r - \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}}}{1 + e^{-\{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}} + e^{-\{\theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}}}, \tag{5.4}$$

$$P(X_{ij} = 1 \wedge X_{ji} = 0) = \frac{e^{-\{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}}}{1 + e^{-\{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}} + e^{-\{\theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}}}, \tag{5.5}$$

$$P(X_{ij} = 0 \wedge X_{ji} = 0) = \frac{1}{1 + e^{-\{\theta_i^{(\text{out})} + \theta_j^{(\text{in})}\}} + e^{-\{\theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}} + e^{-2\theta_r - \{\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_j^{(\text{out})} + \theta_i^{(\text{in})}\}}}. \tag{5.6}$$

These probabilities can be derived analogously to Eq. (2.32). Link probabilities have to be considered in the form of dyads, since dyad dependency was explicitly induced via the condition on link reciprocity. Furthermore, the above dyad probabilities enable us to easily sample networks via bivariate Bernoulli trials.

It can be shown that if the set of probability distributions satisfying the constraints of Equation (2.5) is not empty, the link probabilities are defined uniquely, see Theorem 2.1.3. From this it follows that the Lagrange multipliers are unique, or in case the Lagrange multipliers appear only as sums of two distinct subsets, they are unique up to a constant β that can be added to one subset and subtracted from the other subset, such that the sum of two Lagrange multipliers, one of each subset, stays the same. From Theorem 2.1.3, we know that constrained maximum entropy problems define a unique probability distribution, which means that the maximizing distribution P defined on $\mathcal{G}$ is unique; see Cover and Thomas (2006). Furthermore, for the ERGM considered in this section we have

$$\begin{aligned} P(x) = \prod_{i<j} & P(X_{ij} = 1 \wedge X_{ji} = 1)^{x_{ij}x_{ji}} P(X_{ij} = 1 \wedge X_{ji} = 0)^{x_{ij}(1-x_{ji})} \\ & P(X_{ij} = 0 \wedge X_{ji} = 1)^{(1-x_{ij})x_{ij}} P(X_{ij} = 0 \wedge X_{ji} = 0)^{(1-x_{ij})(1-x_{ji})}, \end{aligned}$$

because of inter-dyad independence. Now assume that there are two vectors of Lagrange multipliers $\theta \neq \lambda$ that solve the ERGM. Further, consider four graphs that are almost identical but differ in one dyad (x_{ab}, x_{ba}) denoted by $x^{(0,0)}, x^{(0,1)}, x^{(1,0)}, x^{(1,1)} \in \mathcal{G}$. From the uniqueness of P we know that

(i) $P_\theta\{x^{(0,0)}\} = P_\lambda\{x^{(0,0)}\}$,

(ii) $P_\theta\{x^{(0,1)}\} = P_\lambda\{x^{(0,1)}\}$,

(iii) $P_\theta\{x^{(1,0)}\} = P_\lambda\{x^{(1,0)}\}$,

(iv) $P_\theta\{x^{(1,1)}\} = P_\lambda\{x^{(1,1)}\}$.

has to hold. Furthermore, dividing (i) by (iii) yields $\theta_a^{(\text{out})} + \theta_b^{(\text{in})} = \lambda_a^{(\text{out})} + \lambda_b^{(\text{in})}$ and dividing (i) by (iv) yields $-2\theta_r - \{\theta_a^{(\text{out})} + \theta_b^{(\text{in})} + \theta_b^{(\text{out})} + \theta_a^{(\text{in})}\} = -2\lambda_r - \{\lambda_a^{(\text{out})} + \lambda_b^{(\text{in})} + \lambda_b^{(\text{out})} + \lambda_a^{(\text{in})}\}$. Thus, it follows that θ_r is defined uniquely and $\{\theta_a^{(\text{out})} + \beta\}$, $\{\theta_b^{(\text{in})} - \beta\}$ are unique up to an additive constant $\beta \in \mathbb{R}$ for all $a, b \in \{1, \ldots, n\}$.

5.3 Data Description

In order to reconstruct the Italian and German interbank networks, we calibrate the ERGM derived in Section 5.2 to the respective network statistics published in Bargigli et al. (2015); Roukny et al. (2014).

Bargigli et al. (2015) have access to bilateral interbank exposures of Italian banks and Italian branches of foreign banks that were transmitted to the Banca d'Italia in the form of supervisory reports. Their data cover the period 2008–12 and were consolidated at group level. The network statistics provided in Bargigli et al. (2015) on the total unweighted Italian interbank market are given in Table 5.1. Furthermore, Bargigli et al. (2015) observe that these statistics stay very stable over the observed period.

Roukny et al. (2014) extract bilateral exposures of the German interbank market from the German large credit register (Millionenkredit Evidenzzentrale). German banks are obliged to report all bilateral liability exposures that surpass a certain threshold to the German large credit register of the Deutsche Bundesbank; see Roukny et al. (2014). The data analyzed in Roukny et al. (2014) comprise the period 2002–12 and were aggregated at the holding corporation level. The network statistics provided in Roukny et al. (2014) on the unweighted German interbank market are given in Table 5.1. Consistent with Bargigli et al. (2015), Roukny et al. (2014) note that the statistics on the interbank market are very stable over the observed period.

Table 5.1: Network statistics of the Italian interbank market provided by Bargigli et al. (2015) and the German interbank market provided by Roukny et al. (2014).

	Italy (2012)	Germany (2012)
# nodes	533	1,700
# edges	3,235	21,318
Density	1.0%	0.74%
Largest weak component	533	NA
Largest strong component	513	$\approx 1{,}670$
Avg undir. path length	2.2	NA
Avg dir. path length	2.4	≈ 2.241
Out-degree assort.	-0.31	NA
In-degree assort.	-0.37	NA
Degree assort.	-0.37	-0.45
Degree reciprocity	0.45	0.31
Avg dir. clustering	0.448	NA
Avg undir. clustering	0.577	≈ 0.81
Degree Herfindhal index (HHI)	NA	≈ 0.011
Avg Betweenness	NA	≈ 0.00064
Avg Closeness	NA	≈ 0.48
Avg Eigenvector	NA	≈ 0.0054
Corr(in-degree, out-degree)	NA	≈ 0.68

In addition, Bargigli et al. (2015) analyze whether certain higher order network statistics are automatically reproduced by explicitly reconstructing lower order statistics. To answer this question they use three different ERGMs constrained on:

(a) the in- and out-degree sequence (also known as the binary directed configuration model (BDCM));

(b) the in- and out-degree sequence plus the sequence of nodes' reciprocity (also known as the reciprocal configuration model (RCM));

(c) the in- and out-degree sequence plus the in- and out-strength (also known as the directed weighted configuration model (DWCM)).

Since Bargigli et al. (2015) have access to the Italian interbank market, they can constrain their models on the exact values. This differs from our setting, as we use only publicly available information. In their analysis they find that the BDCM closely reconstructs the assortativity of a network but fails short at reconstructing triadic motifs. The latter can be improved by the RCM. For the DWCM they conclude that net exposures are reconstructed well, while the strength disassortativity is overestimated.

5.4 Coupled In- and Out-Degree Distribution

The empirical in- and out-degree distribution of the Italian interbank market is displayed in Bargigli et al. (2015). Also, Roukny et al. (2014) provide graphs of the degree distribution of the German interbank market, as well as Pearson's correlation coefficient of the in- and out-degree sequence. To reproduce the reported distributions we suggest to combine a uniform and a power-law distribution that can be fitted to the available information, i.e., the mean of the distribution and the maximum degree. In the following we consider the discrete probability mass function

$$f(d) = \kappa \left(L^{-\alpha} \times \mathbf{1}_{\{d \in \{1,\ldots,L\}\}} + d^{-\alpha} \times \mathbf{1}_{\{d \in \{L+1,\ldots,R\}\}} \right) \tag{5.7}$$

for $d \in \{1, \ldots, R\}$ and $L \in \{1, \ldots, R\}$. Its continuous analog, however, could be used as well. The parameter κ acts as normalization constant, α is the exponent of the power-law distribution, L divides the support into the initial uniform and subsequent power-law distributed part, and R defines the maximum degree. This gives us two degrees of freedom (α, L), as R is fixed by the data. Considering the available information, we decided to calibrate α and L to match the average empirical degree which is reported by Bargigli et al. (2015) and Roukny et al. (2014), the median, and the upper 1% quantile that can be deduced from the graphs in their papers.

As detailed information on network statistics are not available for every country, we decided to construct our model as flexible as possible. This particularly means that we have to define the median and the upper 1% quantile in dependence of a known parameter. We found that the degree distributions of both the Italian and the German interbank market can be well reproduced by choosing the median of the in- (resp. out-) degree distribution to equal $0.002 \times n$ (resp. $0.004 \times n$) and the upper 1% quantile to equal $0.2 \times n$ (resp. $0.12 \times n$).

Furthermore, we couple the in- and out-degree sequence via a Gaussian copula which enables us to adjust Pearson's correlation coefficient. The calibrated parameters are presented in Table 5.2 and the reproduced distributions are illustrated in Figs. 5.1–5.3 in comparison to the empirical distributions of the true networks.

Table 5.2: Parameter values of the fitted in- and out-degree density function f defined in Equation (5.7) for the Italian and the German interbank market.

Parameter	Italy		Germany	
	in-degree	out-degree	in-degree	out-degree
κ	0.5540	1.2055	1.4994	19.0878
α	1.8510	2.0611	1.9523	2.5534
L	1	3	4	9
R	532	532	1,699	1,699

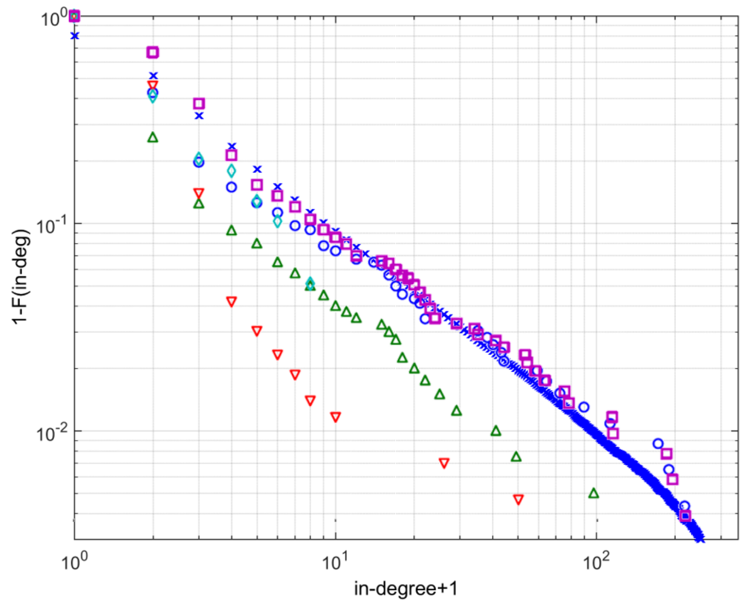


Figure 5.1: Loglog plot comparing the in-degree distribution of the Italian interbank network taken from (Bargigli et al., 2015, p. 11) and our proposed distribution. The total Italian interbank market is represented by purple squares and our calibrated distribution is given by blue crosses. The blue circles, green upward-pointing triangles, red downward-pointing triangles, and turquoise diamonds included in Bargigli et al. (2015) represent resp. the Italian overnight market, unsecured short-term, unsecured long-term, and secured short-term market. In this thesis, however, we focus on the total interbank market.

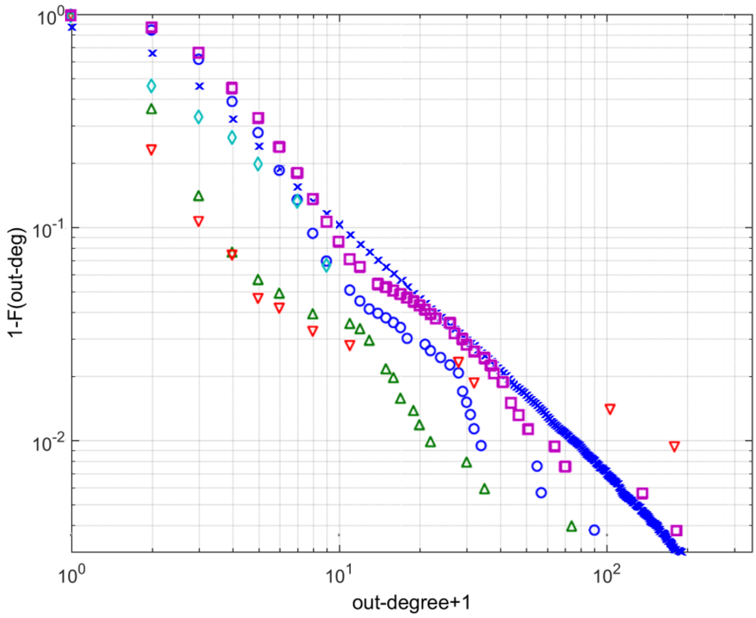


Figure 5.2: Loglog plot comparing the out-degree distribution of the Italian interbank network taken from (Bargigli et al., 2015, p. 11) and our proposed distribution. The total Italian interbank market is represented by purple squares and our calibrated distribution is given by blue crosses. The blue circles, green upward-pointing triangles, red downward-pointing triangles, and turquoise diamonds included in Bargigli et al. (2015) represent resp. the Italian overnight market, unsecured short-term, unsecured long-term, and secured short-term market. In this thesis, however, we focus on the total interbank market.

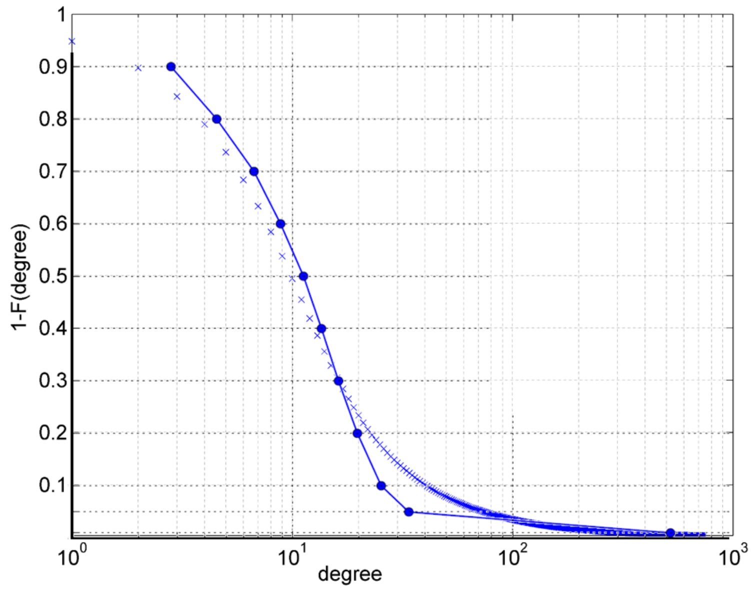


Figure 5.3: Semilog plot comparing the degree distribution of the German interbank network taken from (Roukny et al., 2014, p. 14) and our proposed distribution. The German interbank market is represented by the solid line and filled circles and our calibrated distribution is given by blue crosses.

5.5 Simulation Results

In this section we simulate from the constructed probability laws on $\mathcal{G}$ and compare the output to reported network statistics. The simulation consists of the following three steps.

(1) *Sampling the in- and out-degree sequence:* First, two vectors $u^{(\text{in})}$ and $u^{(\text{out})}$ (realizations of $\big(U^{(\text{in})}, U^{(\text{out})}\big)$) are drawn from a Gaussian copula with correlation parameter ρ, i.e., $\big(U^{(\text{in})}, U^{(\text{out})}\big) \sim C_\rho^{\text{Gauss}}$. Note that $U^{(\text{in})}$ and $U^{(\text{out})}$ are uniformly distributed on $[0, 1]$. We derive realizations of the in- and out-degree sequences via the quantile functions of the degree distributions as defined in Section 5.4, i.e., $k^{(\text{in})} = F^{(\text{in})^{-1}}\big\{u^{(\text{in})}\big\}$ and $k^{(\text{out})} = F^{(\text{out})^{-1}}\big\{u^{(\text{out})}\big\}$. Furthermore, the correlation parameter ρ of the Gaussian copula is set such that $\operatorname{corr}\big(k^{(\text{in})}, k^{(\text{out})}\big) \approx 0.68$. Since for most sampled degree sequences $\sum_i k_i^{(\text{in})} \neq \sum_i k_i^{(\text{out})}$, i.e., the number of sampled incoming links does not equal the number of sampled outgoing links, we add as many shadow banks as needed (usually only one) to absorb the difference.

(2) *Calibration of the ERGM:* The realizations of the in- and out-degree sequence $k^{(\text{in})}$ and $k^{(\text{out})}$ from Step (1) as well as the number of reciprocal links serve as input for the ERGM. More precisely, these values determine the right hand side of the

non-linear equation system defined by Eqs. (5.1)–(5.3). This equation system is solved numerically (e.g., in MATLAB via the FSOLVE function) and returns the values of the Lagrange multipliers.

(3) *Sampling the network:* The Lagrange multipliers specify the link probabilities, as given in Eqs. (5.4)–(5.6). A network can conveniently be sampled via bivariate Bernoulli trials according to these link probabilities.

We remark that it is essential to add shadow banks to absorb the difference in the number of links specified by the sampled in- and out-degree sequence, since otherwise the constraints of the ERGM would be impossible to be satisfied. One could alternatively try to solve such a wrongly specified maximum entropy problem by minimizing the overall error. Attempts to do so, however, increase runtime excessively. This might be due to the existence of various different parameter combinations with a similar absolute error.

The results of 100 simulations of the Italian interbank market for 2012 are presented in Table 5.3 and compared to the available empirical numbers. In addition, Bargigli et al. (2015) report the network statistics of the Italian interbank market for 2008. Based on this information we conducted the same analysis for 2008, to assess the quality of our model over time. As the network statistics reported in Bargigli et al. (2015) remain quite stable over the observed period, our simulation results for 2008 are very similar to those of 2012.

Table 5.3: Comparison of some network statistics for 2012 as reported by Bargigli et al. (2015) and for 100 simulated networks of the Italian interbank market excluding shadow banks. The simulation of 100 Italian interbank networks in Matlab 2017a using parallel computing took 2.5h on an Intel(R) Xeon(R) E5-2687W v3 at 3.1 GHz. Additionally, in brackets the values of an ERGM conditioned only on the in- and out-degree sequence are reported.

Network Statistic	Bargigli et al. (2015)	Mean	Standard Deviation	95% Confidence Interval
Total number of links	$3,235$	$3,168$ ($3,323$)	540 (703)	$[3,062; 3,274]$ ($[3,186; 3,461]$)
Number of reciprocal links	1,476	$1,558$ ($1,161$)	229 (600)	$[1,513; 1,603]$ ($[1,043; 1,279]$)
In-degree assortativity	-0.37	-0.39 (-0.40)	0.10 (0.10)	$[-0.41; -0.37]$ ($[-0.42; -0.38]$)
Out-degree assortativity	-0.31	-0.37 (-0.38)	0.10 (0.09)	$[-0.39; -0.35]$ ($[-0.40; -0.37]$)
Avg. dir. path length	2.4	2.5 (2.4)	0.34 (0.29)	$[2.46; 2.59]$ ($[2.24; 2.50]$)
Avg. undir. path length	2.2	2.3 (2.3)	0.25 (0.22)	$[2.28; 2.38]$ ($[2.26; 2.35]$)
Largest strong component	513	385 (377)	58 (68)	$[374; 397]$ ($[363; 390]$)
Largest weak component	533	505 (512)	23 (18)	$[501; 509]$ ($[508; 516]$)

Overall, our model accurately reproduces most network statistics. The discrepancy in the out-degree assortativity indicates that in the simulations the out-degrees of connected banks are too heterogeneous. In fact, Figure 5.2 shows that very low out-degrees are overrepresented in the simulation. This could be fixed by increasing the median, which is used to calibrate the out-degree distribution. This would, however, also result in a faster decline of the distribution after the upper 1% quantile in order to preserve the expected value, and thus high values would be underrepresented. The same would apply for the German degree distribution, since the median is defined as a function of n, which would deteriorate the goodness-of-fit of the German degree distribution. Furthermore, the error of having too many small out-degrees seems to be minor, considering that most applications focus on a subset of banks that comprise a substantial monetary volume. As one would imagine, Roukny et al. (2014) report that the correlation of volume and degree is approximately 0.91. Thus, reducing the interbank network to a subset of banks of interest eliminates banks with small out-degrees. The same holds for the error in the LSC. A reduction to a subset of banks of interest most likely eliminates banks that are not part of the LCS.

Table 5.3 additionally lists the statistics for an ERGM conditioned solely on the in- and out-degree sequence, to analyze the effect of including the reciprocity in the ERGM. Interestingly, with the exception of the number of reciprocal links, which naturally yields a better fit for the ERGM incorporating reciprocity, all other statistics change only slightly. This means that degree disassortativity, short path lengths, and large connected components are fundamentally steered by the heterogeneity in the in- and out-degree sequence.

The results of 100 simulations of the German interbank market for 2012 are presented in Table 5.4. Based on network statistics provided by Roukny et al. (2014), we also conducted the same analysis on the German interbank market for 2002 and 2008, to assess the quality of our model over time. As the network statistics reported in Roukny et al. (2014) remain quite stable over the observed period, our simulation results for 2002 and 2008 are very similar to those of 2012.

Table 5.4: Comparison of some network statistics for 2012 as reported by Roukny et al. (2014) and for 100 simulated networks of the German interbank market excluding shadow banks. Additionally, in brackets the values of an ERGM conditioned only on the in- and out-degree sequence are reported.

Network Statistic	Roukny et al. (2014)	Mean	Standard Deviation	95% Confidence Interval
Total number of links	21,318	$20,727$ (20,667)	$1,355$ $(1,847)$	$[20,460; 20,994]$ $([20,305; 21,029])$
Number of reciprocal links	6,717	$6,921$ (3,482)	584 $(1,460)$	$[6,805; 7,306]$ $([3,195; 3,768])$
Degree assortativity	-0.45	-0.25 (-0.25)	0.06 (0.07)	$[-0.26; -0.24]$ $([-0.27; -0.24])$
Avg. dir. path length	2.24	2.59 (2.58)	0.27 (0.26)	$[2.54; 2.64]$ $([2.53; 2.63])$
Largest strong component	1,670	$1,424$ $(1,401)$	175 (169)	$[1,389; 1,458]$ $([1,368; 1,434])$
HHI	0.011	0.0070 (0.0069)	0.0020 (0.0019)	$[0.0066; 0.0074]$ $([0.0065; 0.0073])$
Avg. Closeness	0.48	0.3944 (0.3961)	0.0421 (0.0418)	$[0.3861; 0.4027]$ $([0.3879; 0.4043])$
Avg. Eigenvector	0.0054	0.0156 (0.0155)	0.0005 (0.0005)	$[0.0155; 0.0157]$ $([0.0154; 0.0156])$
Corr(in-degree, out-degree)	0.68	0.69 (0.67)	0.10 (0.10)	$[0.67; 0.71]$ $([0.65; 0.69])$

Most of the relevant network statistics are very close to the true values. The discrepancy in the degree assortativity indicates that the degrees of connected banks are too homogeneous. In fact, Figure 5.3 shows that degrees of about 15 to 100 are overrepresented while extreme degrees exceeding 200 are underrepresented. In order to account for this finding, one would have to choose a different degree distribution with heavier tails. Again, the error in the LSC seems to be minor when considering a reduction to a subset of banks of interest that comprise a substantial monetary volume. Furthermore, the centrality measures indicate that the presence of banks acting as intermediaries is stronger in the true network.

Table 5.4 additionally lists the statistics for an ERGM conditioned solely on the in- and out-degree sequence, to analyze the effect of including the reciprocity in the ERGM. Similar to the results on the Italian interbank market, we again observe that all network statistics, with the exception of the number of reciprocal links, barely change. This indicates that these network statistics are essentially steered by the heterogeneity in the in- and out-degree sequence.

5.6 Conclusion and Outlook

In this chapter, we have developed an analytically tractable instance of an ERGM for the reconstruction and simulation of financial networks. Based on available information on the Italian and German interbank market, we have demonstrated that the model adequately reproduces a number of network statistics. Moreover, the model needs only few and rather general input parameters, consisting of the in- and out-degree distribution, their coupling via a Gaussian copula, and the number of reciprocal links. This has two important implications. First, as no further information is required, this simplifies the adaptation of the model to other countries. Second, as the model adequately reconstructs further network characteristics that have not been explicitly incorporated into the model, this indicates that these features are a natural consequence of the degree distribution. Not only is this finding interesting in its own right, it is also welcome news considering the very limited data availability.

Depending on the desired analysis to which the model serves as input, one might still have to weight the adjacency matrices resulting from the model. One possibility to assign weights, such that row and column sums are fulfilled, is given by the algorithm from Gandy and Veraart (2017a). However, while network statistics are available on country level, interbank assets and liabilities published in the banks' balance sheets refer to the international interbank market. Therefore, it seems reasonable to first reconstruct the interbank markets of relevant countries and to connect them to a global network in a second step. Cross border links could, for example, be estimated based on the information provided by the EBA's transparency exercise.

6 A Block-Structured Model for Banking Networks Across Multiple Countries

This chapter extends the model presented in Chapter 5 to international weighted financial networks. In a first step we use an extended fitness model, that can be calibrated to a desired density and reciprocity, to reconstruct the unweighted and directed network of each domestic and each cross-border subgraph. This results in a link-probability matrix from which we can easily sample adjacency matrices through bivariate Bernoulli trials. In a second step, the sampled adjacency matrices are weighted, such that interbank assets and liabilities, which are known from the banks' balance sheets, as well as the total weight circulating within and across countries, is met. This is achieved via an exponential random graph model (ERGM), conditioned on the row and column sums as well as on the block weights. Since this model allows to analytically derive the expected weight of each link of a given adjacency matrix, the conditions are fulfilled exactly by the resulting network.

This model is analytically tractable, allows a calibration on scarce publicly available data, and closely reconstructs known network characteristics of financial markets. Moreover, the model finally enables the application of the proposed contagion mechanisms and systemic risk measures to more realistic and international financial networks, as demonstrated in Chapter 7.

This chapter is based on our manuscript Engel et al. (2019a).

6.1 Problem of Financial Network Reconstruction

In the following, we consider a network consisting of n financial institutions that are located in N countries, which are denoted by $C_1, C_2, \ldots, C_N$; $|C_j|$ denoting the number of banks in country $j \in \{1, \ldots, N\}$. This creates a chessboard with N^2 blocks. The corresponding network can be visualized in form of a matrix w, as illustrated in Fig. 6.1. The element w_{ij} equals the nominal value of loans that bank i lends to bank j. Consequently, the row sum $s_i^{(\text{out})}$ (resp. column sum $s_i^{(\text{in})}$) denotes the total interbank assets (resp. deposits) of bank i.

As pointed out in Chapter 1, the fundamental problem of reconstructing financial networks is the limited data availability. In fact, regarding interbank networks, the only information which is regularly available to the public are interbank assets $s^{(\text{out})}$ and

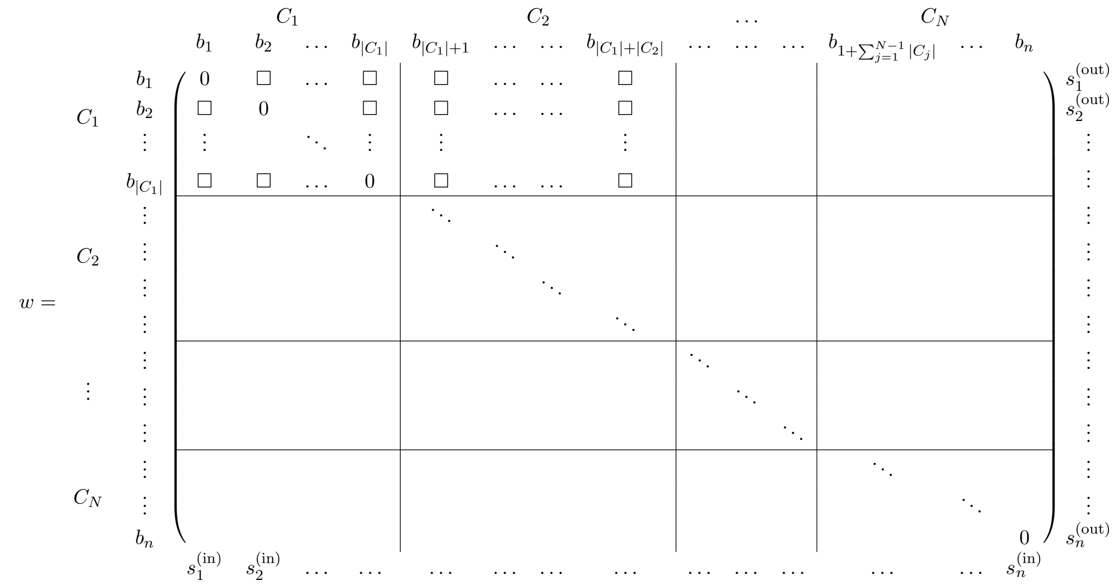


Figure 6.1: Illustration of an international financial network comprising n financial institutions $b_1, \dots, b_n$, grouped by their country of origin $C_1, \dots, C_N$. The variables $s_i^{(\text{out})}$ and $s_i^{(\text{in})}$ denote the i-th row and column sum, respectively.

interbank liabilities $s^{(\text{in})}$ of each bank. However, considering all networks which fulfill the given row and column sums as possible financial networks is misleading, since financial networks are, for example, known to be sparse, see, e.g., Craig and Von Peter (2014). Hence, densely connected networks should be excluded on the grounds of not being realistic. Moreover, financial networks are known to be disassortative, to exhibit a core-periphery structure, to feature short paths, degrees and strength are highly correlated, etc., see, e.g., Craig and Von Peter (2014), Roukny et al. (2014), Bargigli et al. (2015). But precise, regular, and complete information on these characteristics are not publicly available. Thus, to construct realistic networks, a trade-off has to be made on incorporating available data and accuracy of the reconstructed network topology. Furthermore, we aim at a tractable model which also includes the flexibility to change particular network statistics, in order to allow for a detailed assessment of systemic risk and an analyses of potentially more stable network structures. Therefore, we choose to base our model on the following characteristics (i) - (iv) for two reasons. First, the necessary information is either publicly available or can be approximated from publicly available data, as demonstrated in Section 6.5.1 w.r.t. the EU interbank network. This allows the construction of realistic interbank networks. Second, these characteristics can be incorporated in an analytically tractable model. Hence, these characteristics can easily be twisted and network ensembles with different structures can be generated. This creates a flexible playground for a detailed analysis on systemic risk and possible new regulations, which is a key objective of this work. We assume the following information to be given:

(i) The density for each block, or equivalently the number of links $L^{\rightarrow}_{kl}$ for all blocks $k, l = 1, \ldots, N$, i.e.

$$L^{\rightarrow} \in \mathbb{N}_0^{N \times N}; \tag{6.1}$$

(ii) The (degree-) reciprocity for each pair of transposed blocks, or equivalently the number of mutual links $L^{\leftrightarrow}_{kl} = L^{\leftrightarrow}_{lk}$ for all blocks $k, l = 1, \ldots, N$, i.e.

$$L^{\leftrightarrow} \in \mathbb{N}_0^{N \times N}, \text{ and } L^{\leftrightarrow} \text{ symmetric}; \tag{6.2}$$

(iii) Interbank assets $s_i^{(\text{out})}$ and liabilities $s_i^{(\text{in})}$ for each bank $i = 1, \ldots, n$ in the network, i.e.

$$s^{(\text{in})} \in \mathbb{N}_0^n \quad \text{and} \quad s^{(\text{out})} \in \mathbb{N}_0^n; \tag{6.3}$$

(iv) The total weight for each block, i.e.

$$s^{(\text{block})} \in \mathbb{N}_0^{N \times N}. \tag{6.4}$$

The row and column sums, as well as the block weights, are restricted to the set of natural numbers, as this considerably simplifies the required calculations, see Section 6.3.

Following the notation of Gandy and Veraart (2017a), let $w \in \mathbb{N}_0^{n \times n}$ denote an interbank matrix, where the element w_{ij} denotes the nominal value of interbank loans granted from bank i to bank j. Some stakeholders, such as financial institutions, central banks, or regulators might have partial knowledge on the interbank network, i.e. they might know the true value of some elements of w. Therefore, we define $w^* \in \mathcal{W} := \left(\{*\} \cup \mathbb{N}_0\right)^{n \times n}$, where $w^*_{ij} = *$ denotes an unknown matrix element. We are interested in the set of all interbank matrices fulfilling the desired characteristics (i) – (iv), as well as matching all known bilateral interbank elements. Constructing this set of matrices in a tractable and computationally feasible way is a non-trivial task, and to the best of our knowledge such a model is not yet available. To provide a solution, we relax the problem and consider instead a probability space that generates interbank matrices which satisfy the desired characteristics in expectation.

Definition 6.1.1 (Admissible probability space for interbank networks)
Let $\Omega := \left\{w \in \mathbb{N}_0^{n \times n}\right\}$ *be a set of weighted and directed graphs and let* $P : \mathfrak{P}(\Omega) \to [0, 1]$ *be a probability measure defined on the power set* $\mathfrak{P}(\Omega)$ *of* Ω*. The probability space* $(\Omega, \mathfrak{P}(\Omega), P)$ *is called* admissible *w.r.t.* $L^{\rightarrow} \in \mathbb{N}_0^{N \times N}$, $L^{\leftrightarrow} \in \mathbb{N}_0^{N \times N}$ *symmetric,* $s^{(\text{in})} \in \mathbb{N}_0^n$, $s^{(\text{out})} \in \mathbb{N}_0^n$, $s^{(\text{block})} \in \mathbb{N}_0^{N \times N}$*, and* w^* *if the following conditions are met:*

(i)

$$\sum_{w\in\Omega} P(w)\left[\sum_{i\in C_k, j\in C_l} \mathbb{1}_{\{w_{ij}>0\}}\right] = L_{kl}^{\rightarrow}, \quad \forall k,l = 1,\ldots,N, \qquad \text{(directed links)}$$

(ii)

$$\sum_{w\in\Omega} P(w)\left[\sum_{i\in C_k, j\in C_l} \mathbb{1}_{\{w_{ij}>0\}}\mathbb{1}_{\{w_{ji}>0\}}\right] = L_{kl}^{\leftrightarrow}, \quad \forall k,l = 1,\ldots,N, \qquad \text{(reciprocal links)}$$

(iii)

$$\sum_{w\in\Omega} P(w)\left[\sum_{j=1}^{n} w_{ij}\right] = s_i^{(\text{out})}, \quad \forall i = 1,\ldots,n, \qquad \text{(assets)}$$

(iv)

$$\sum_{w\in\Omega} P(w)\left[\sum_{i=1}^{n} w_{ij}\right] = s_j^{(\text{in})}, \quad \forall j = 1,\ldots,n, \qquad \text{(liabilities)}$$

(v)

$$\sum_{w\in\Omega} P(w)\left[\sum_{i\in C_k, j\in C_l} w_{ij}\right] = s_{kl}^{(\text{block})}, \quad \forall k,l = 1,\ldots,N, \qquad \text{(block weights)}$$

(vi)

$$w_{ij} = w_{ij}^{*}, \quad \forall w \in \Omega \textit{ and } w_{ij}^{*} \neq *. \qquad \text{(known links)}$$

In the following, we present a model for generating such admissible ensembles of interbank matrices.

6.2 Reconstructing Unweighted Directed Graphs via an Extended Fitness Model

As explained in Section 2.2.1, fitness models have successfully been used to reconstruct several economic and financial networks. Until now the focus of fitness models has been the reconstruction of realistic degree distributions. Here, we extend this methodology to additionally incorporate the degree reciprocity. More precisely, we use the link probability function as defined by an ERGM conditioned on the in- and out-degree sequences and the number of reciprocal links. For a derivation of this ERGM, see Section 5.2. The parameters of the ERGM can be estimated for a given in- and out-degree sequence and a given number of reciprocal links. For most blocks, however, no information on the in- and out-degrees is publicly available. For this reason, we resort to the idea of fitness models and consider the unknown parameters (i.e. the exponential function of the negative Lagrange multipliers) as hidden variables. More precisely, the hidden variables controlling the link probabilities are specified by the banks' interbank assets $s_i^{(\text{out})}$ and interbank liabilities $s_i^{(\text{in})}$, multiplied by a block specific parameter $z \in \mathbb{R}_{\geq 0}^{N\times N}$ that controls for the network density of each block. This leads to the following link probabilities: Let $A \in \{0,1\}^{n\times n}$ denote the random adjacency matrix, and to simplify notation let z_{ij} denote the z parameter of the corresponding block, then for $i \neq j$,

$$P\left(A_{ij}=1 \wedge A_{ji}=1\right) = \frac{r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z_{ij} s_i^{(\text{out})} s_j^{(\text{in})} + z_{ji} s_j^{(\text{out})} s_i^{(\text{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.5}$$

$$P\left(A_{ij}=1 \wedge A_{ji}=0\right) = \frac{z_{ij} s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z_{ij} s_i^{(\text{out})} s_j^{(\text{in})} + z_{ji} s_j^{(\text{out})} s_i^{(\text{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.6}$$

$$P\left(A_{ij}=0 \wedge A_{ji}=0\right) = \frac{1}{1 + z_{ij} s_i^{(\text{out})} s_j^{(\text{in})} + z_{ji} s_j^{(\text{out})} s_i^{(\text{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.7}$$

where $r \in \mathbb{R}_{>0}^{N\times N}$ with $r_{kl} = r_{lk}$ is a block specific parameter controlling for the number of reciprocal links. Again to simplify notation, r_{ij} denotes the r parameter of the corresponding block. Furthermore, since financial networks do not exhibit self-loops, we set $A_{ii} = 0$ for all $i = 1, \dots, n$.

The parameter r corresponds to the term $\exp(-\lambda_r)$ of the discussed ERGM, with λ_r being the Lagrange multiplier of the constraint on the reciprocal links. Setting this Lagrange multiplier to zero $\lambda_r = 0$, or equivalently $r = 1$, and assuming pairwise independence for all links, results in the classical fitness model. This model has been studied in detail and it has been shown to yield good results, see for example Cimini et al. (2015) and Gandy and Veraart (2017b). Our model extends the classical fitness model by additionally incorporating the number of reciprocal links and therefore introducing a dependence structure between the dyads (A_{ij}, A_{ji}). We decided to include the reciprocity in our model, since it has been shown to constitute an important network characteristic, which for most networks does not come as a natural consequence of the degree sequence, see for example Garlaschelli and Loffredo (2004b) and Bargigli et al. (2015), and because it introduces only one additional parameter. As can be seen from the link probabilities, Eqs. (6.5) to (6.7), this setting correlates the number of links of a node to its weight, i.e. the higher the total incoming and outgoing weight of a node, the higher the number of incoming and outgoing links of a node. This is an essential characteristic that one would expect for financial networks, and which has been shown to hold in many empirical works, e.g. Roukny et al. (2014), Bargigli et al. (2015).

Incorporating available information on the existence of certain links is straightforward. Let $a^* \in \{*, 0, 1\}^{n \times n}$, where $a_{ij} = *$ denotes an unknown link. The link probability matrix A is extended as follows,

$$A_{ij} = a^*_{ij}, \quad \forall a^*_{ij} \neq *. \tag{6.8}$$

In case only one of two reciprocal links is known, i.e. $a^*_{ij} \neq *$ and $a^*_{ji} = *$, the probability distribution of the unknown link is given by

$$P\left(A_{ji} = 0 | a^*_{ij} = 0\right) = \frac{P\left(A_{ji} = 0, A^*_{ij} = 0\right)}{P\left(A_{ij} = 0\right)} = \frac{1}{1 + z_{ji} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.9}$$

$$P\left(A_{ji} = 0 | a^*_{ij} = 1\right) = \frac{P\left(A_{ji} = 0, A^*_{ij} = 1\right)}{P\left(A_{ij} = 1\right)} = \frac{z_{ij} s_i^{(\text{out})} s_j^{(\text{in})}}{z_{ij} s_i^{(\text{out})} s_j^{(\text{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.10}$$

$$P\left(A_{ji} = 1 | a^*_{ij} = 0\right) = \frac{P\left(A_{ji} = 1, A^*_{ij} = 0\right)}{P\left(A_{ij} = 0\right)} = \frac{z_{ji} s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z_{ji} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.11}$$

$$P\left(A_{ji} = 1 | a^*_{ij} = 1\right) = \frac{P\left(A_{ji} = 1, A^*_{ij} = 1\right)}{P\left(A_{ij} = 1\right)} = \frac{r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}{z_{ij} s_i^{(\text{out})} s_j^{(\text{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}. \tag{6.12}$$

The unknown parameters z and r can be calibrated such that a desired number of links $L^{\rightarrow}$ and a desired number of reciprocal links $L^{\leftrightarrow}$ is met in expectation. The following three equations are used to calibrate the three parameters for each pair of transposed blocks $k, l = 1, \ldots, N$,

$$\mathbb{E}\left[\sum_{i \in C_k, j \in C_l} A_{ij}\right] = L^{\rightarrow}_{kl}, \qquad \text{(directed links)} \tag{6.13}$$

$$\mathbb{E}\left[\sum_{i \in C_l, j \in C_k} A_{ij}\right] = L^{\rightarrow}_{lk}, \qquad \text{(directed links)} \tag{6.14}$$

and

$$\mathbb{E}\left[\sum_{i\in C_k, j\in C_l} A_{ij}A_{ji}\right] = \mathbb{E}\left[\sum_{i\in C_l, j\in C_k} A_{ij}A_{ji}\right] = L_{kl}^{\leftrightarrow}. \qquad \text{(reciprocal links)} \quad (6.15)$$

There are many options to solve Eqs. (6.13) to (6.15), for example Matlab's nonlinear least-squares solver. For the blocks on the diagonal, i.e. for $k = l$ (representing domestic interbank markets), Eq. (6.13) and Eq. (6.14) are identical, hence there are only two equations to be solved. Once the parameters are calibrated, sampling adjacency matrices, i.e. unweighted directed graphs, is easy and fast via bivariate Bernoulli trials.

If the desired number of links and reciprocal links are set to feasible numbers there always exists a solution, as the following theorem shows. Since the parameters z and r are independent for all pairs of submatrices $\left(A^{(\mathrm{kl})}, A^{(\mathrm{lk})}\right)$, it suffices to show the existence of a solution for one pair $\left(A^{(\mathrm{kl})}, A^{(\mathrm{lk})}\right)$.

Theorem 6.2.1 (Existence of a solution for the extended fitness model)
Consider four vectors $s^{(\mathrm{out,k})}, s^{(\mathrm{in,k})} \in \mathbb{R}^{n_k}_{>0}$, and $s^{(\mathrm{out,l})}, s^{(\mathrm{in,l})} \in \mathbb{R}^{n_l}_{>0}$. Let the random matrices $A^{(\mathrm{kl})} \in \{0,1\}^{n_k \times n_l}$ and $A^{(\mathrm{lk})} \in \{0,1\}^{n_l \times n_k}$ be defined by the probability function P as given by Eqs. (6.5) *to* (6.7)*. For any feasible number of*

(i) reciprocal links $L_{kl}^{\leftrightarrow} \in [0, n_k n_l)$,

(ii) and links $\tilde{L}_{kl}^{\rightarrow} := L_{kl}^{\rightarrow} - L_{kl}^{\leftrightarrow}$ *and* $\tilde{L}_{lk}^{\rightarrow} := L_{lk}^{\rightarrow} - L_{kl}^{\leftrightarrow}$ *with* $\left(\tilde{L}_{kl}^{\rightarrow} + \tilde{L}_{lk}^{\rightarrow}\right) \in (0, n_k n_l - L_{kl}^{\leftrightarrow})$,

there exist $z_{kl}, z_{lk}, r_{kl} \in \mathbb{R}_{\geq 0}$, such that Eqs. (6.13) *to* (6.15) *are fulfilled.*

Proof
Since the proof takes several pages, it has been moved to the end of this chapter; see Section 6.6.

To prove whether the solution of the extended fitness model is unique is non-trivial. However, we can show that the solution is unique w.r.t. the expected in- and out-degree sequences and the expected number of reciprocal links created by the solution.

Theorem 6.2.2 (Uniqueness of a solution for the extended fitness model)
A solution z_{kl}, z_{lk}, r_{kl} of the extended fitness model, as described in Theorem 6.2.1, yields

the following particular sequences of expected in- and out-degrees,

$$d_j^{(\mathrm{in,kl})} = \sum_{i \in C_k} P\left(A_{ij}^{(\mathrm{kl})} = 1\right), \quad \forall j \in C_l, \tag{6.16}$$

$$d_i^{(\mathrm{out,kl})} = \sum_{j \in C_l} P\left(A_{ij}^{(\mathrm{kl})} = 1\right), \quad \forall i \in C_k, \tag{6.17}$$

$$d_j^{(\mathrm{in,lk})} = \sum_{i \in C_l} P\left(A_{ij}^{(\mathrm{lk})} = 1\right), \quad \forall j \in C_k, \tag{6.18}$$

$$d_i^{(\mathrm{out,lk})} = \sum_{j \in C_k} P\left(A_{ij}^{(\mathrm{lk})} = 1\right), \quad \forall i \in C_l, \tag{6.19}$$

as well as the expected number of reciprocal links $L_{kl}^{\leftrightarrow}$. The solution z_{kl}, z_{lk}, r_{kl} is unique in the sense that it is the only parameter combination which generates the specific expected values $d_j^{(\mathrm{in,kl})}, d_i^{(\mathrm{out,kl})}, d_j^{(\mathrm{in,lk})}, d_i^{(\mathrm{out,lk})}$, and $L_{kl}^{\leftrightarrow}$.

Proof
Consider the ERGM defined by maximizing the Shannon entropy such that the expected particular degree sequences of Eqs. (6.16) to (6.19) and the expected number of reciprocal links $L_{kl}^{\leftrightarrow}$ are satisfied. A solution to this ERGM is given by

$$e^{-\lambda_i^{(\mathrm{out,k})}} := \sqrt{z_{kl}} s_i^{(\mathrm{out,k})}, \quad \forall i \in C_k, \tag{6.20}$$

$$e^{-\lambda_i^{(\mathrm{in,k})}} := \sqrt{z_{kl}} s_i^{(\mathrm{in,l})}, \quad \forall i \in C_l, \tag{6.21}$$

$$e^{-\lambda_i^{(\mathrm{out,l})}} := \sqrt{z_{lk}} s_i^{(\mathrm{out,l})}, \quad \forall i \in C_l, \tag{6.22}$$

$$e^{-\lambda_i^{(\mathrm{in,l})}} := \sqrt{z_{lk}} s_i^{(\mathrm{in,k})}, \quad \forall i \in C_k, \tag{6.23}$$

$$e^{-\lambda_r} := r, \tag{6.24}$$

where the λ's denote the corresponding Lagrange multipliers. From the general theory of maximum entropy problems, we know that the solving probability distribution is unique, see Theorem 2.1.3. This means that all link probabilities, as given by Eqs. (6.5) to (6.7), are unique, see Engel et al. (2019b). From this it follows that z_{kl}, z_{lk}, and r are unique. □

6.3 Allocation of Weights via an ERGM

In a second step, we allocate weights to an adjacency matrix, sampled from the fitness model discussed in Section 6.2, through an exponential random graph model.

Let $a \in \{0,1\}^{n \times n}$ denote a realization of the random adjacency matrix A, as specified in the previous section. We define a set of possible weighted graphs $\mathcal{G}_a$ consistent with a, as the set of all graphs that assign weights in $\mathbb{N}_0$ to existing links of a and zero weight to non-existing links of a:

$$\mathcal{G}_a = \Bigg\{ w \in \mathbb{N}_0^{n\times n} \mid w_{ij} = 0, \forall a_{ij} = 0 \text{ and } w_{ij} \in \mathbb{N}_0, \forall a_{ij} = 1 \Bigg\}. \tag{6.25}$$

Remark 6.3.1 (Bounded link weights)
We could further restrict the set of considered graphs, as the maximum weight that a link can carry is given by the minimum of the corresponding row, column, and block weight, i.e.

$$\begin{aligned}\tilde{\mathcal{G}}_a = \Bigg\{ & w \in \mathbb{N}_0^{n\times n} \mid w_{ij} = 0, \forall a_{ij} = 0 \\ & \textit{and } w_{ij} \in \Big\{1, 2, \ldots, \min\Big\{ s_i^{(\text{out})}, s_j^{(\text{in})}, s_{ij}^{(\text{block})} \Big\}\Big\}, \forall a_{ij} = 1 \Bigg\}.\end{aligned} \tag{6.26}$$

The analytical derivation of this model works analogously to the one considering $\mathcal{G}_a$. However, the resulting expected link weights take a slightly more complex form, which renders parameter estimation more difficult. Since all expected weights, in the setting of $\mathcal{G}_a$, lie in the interval $\Big(0, \min\Big\{ s_i^{(\text{out})}, s_j^{(\text{in})}, s_{ij}^{(\text{block})} \Big\}\Big]$, here we consider the simpler setting of $\mathcal{G}_a$.

Remark 6.3.2 (Partial knowledge of certain weights)
Incorporating available information on the weight of certain links is straightforward. Let $w^* \in \mathcal{W} := \big(\{*\} \cup \mathbb{N}_0\big)^{n\times n}$, where $w^*_{ij} = *$ denotes an unknown matrix element. In this case we simply consider the set of graphs

$$\begin{aligned}\mathcal{G}_{a,w^*} = \Bigg\{ & w \in \mathbb{N}_0^{n\times n} \mid w_{ij} = w^*_{ij}, \forall w^*_{ij} \neq * \textit{ and } w_{ij} = 0, \forall a_{ij} = 0 \\ & \textit{and } w_{ij} \in \mathbb{N}_0, \forall a_{ij} = 1, w^*_{ij} = * \Bigg\}.\end{aligned} \tag{6.27}$$

The analytical derivation of this model works analogously to the one considering $\mathcal{G}_a$.

Further, let $\mathcal{P} := \{p : \mathcal{G}_a \to [0, 1]\}$ denote the set of all probability measures defined on $\mathcal{G}_a$. The most unbiased probability measure $p \in \mathcal{P}$ is the one with the minimum Kullback–Leibler divergence w.r.t. the uniform distribution, or equivalently with maximum Shannon entropy, and which fulfills the desired row sums, column sums, and block weights in expectation. This translates to the following constrained optimization problem,

$$\max_{p\in\mathcal{P}} - \sum_{w\in\mathcal{G}_a} p(w) \log\Big(p(w)\Big) \tag{6.28}$$

subject to

$$\sum_{w \in \mathcal{G}_a} p(w) \left(\sum_{j=1}^{n} w_{ij} \right) = s_i^{(\text{out})}, \qquad \forall i = 1, \ldots, n, \qquad \text{(assets)}$$

$$\sum_{w \in \mathcal{G}_a} p(w) \left(\sum_{i=1}^{n} w_{ij} \right) = s_j^{(\text{in})}, \qquad \forall j = 1, \ldots, n, \qquad \text{(liabilities)}$$

$$\sum_{w \in \mathcal{G}_a} p(w) \left(\sum_{i \in C_k, j \in C_l} w_{ij} \right) = s_{kl}^{(\text{block})}, \quad \forall k, l = 1, \ldots, N, \qquad \text{(block weights)}$$

$$\sum_{w \in \mathcal{G}_a} p(w) = 1. \tag{6.29}$$

Solving this optimization problem as usual by the method of Lagrange multipliers, yields the following graph Hamiltonian $H : \mathcal{G}_a \to \mathbb{R}$

$$H(w) = \sum_{i,j=1}^{n} \left(\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_{ij}^{(\text{block})} \right) w_{ij}, \tag{6.30}$$

where $\theta^{(\text{out})}, \theta^{(\text{in})} \in \mathbb{R}^n$, and $\theta^{(\text{block})} \in \mathbb{R}^{N \times N}$ denote the corresponding Lagrange multipliers. To simplify the notation, let $\theta_{ij}^{(\text{block})}$ denote the Lagrange multiplier of the corresponding block. Next, we derive the partition function Z. For further clarity we abbreviate $\theta_{ij} := \left(\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_{ij}^{(\text{block})}\right)$. W.l.o.g. let $a_{12} = 1$

$$\begin{aligned}
Z &= \sum_{w \in \mathcal{G}_a} e^{-H(w)} = \sum_{w \in \mathcal{G}_a} \prod_{i \neq j} \exp\left(-\theta_{ij} w_{ij}\right) \\
&= \sum_{w \in \mathcal{G}_a} e^{-\theta_{12} w_{12}} \prod_{\substack{i \neq j \\ (ij) \notin \{(12)\}}} e^{-\theta_{ij} w_{ij}} \\
&= \left(\sum_{w \in \mathcal{G}_a | w_{12} = 0} e^{-\theta_{12} w_{12}} \prod_{\substack{i \neq j \\ (ij) \notin \{(12)\}}} e^{-\theta_{ij} w_{ij}} \right) + \left(\sum_{w \in \mathcal{G}_a | w_{12} = 1} e^{-\theta_{12} w_{12}} \prod_{\substack{i \neq j \\ (ij) \notin \{(12)\}}} e^{-\theta_{ij} w_{ij}} \right) + \ldots \\
&= \left(\sum_{\tilde{w}_{12} = 0}^{\infty} e^{-\theta_{12} \tilde{w}_{12}} \right) \underbrace{\left[\sum_{w \in \mathcal{G}_a | w_{12} = \tilde{w}_{12}} \prod_{\substack{i \neq j \\ (ij) \notin \{(12)\}}} e^{-\theta_{ij} w_{ij}} \right]}_{\text{constant for all } \tilde{w}_{12}} \\
&\overset{(\star)}{=} \prod_{(i,j) | a_{ij} = 1} \left(\sum_{w_{ij} = 0}^{\infty} e^{-\theta_{ij} w_{ij}} \right) = \prod_{(i,j) | a_{ij} = 1} \frac{1}{1 - e^{-\theta_{ij}}}, \qquad \text{by the geometric series,}
\end{aligned} \tag{6.31}$$

where the same algebraic steps are applied to all elements in $(\star)$ as we applied exemplary to the first element w_{12}. Note that

$$e^{-\theta_{ij}} < 1 \tag{6.32}$$

has to hold for all (i, j) for which $a_{ij} = 1$, since otherwise the value of the partition function Z is infinity, which implies that p is not a solving probability measure.

From the general theory of ERGMs we known that taking partial derivatives of $F := -\log(Z)$ w.r.t. the Lagrange multipliers yields the expected value of the corresponding constraint, see Eq. (2.19). Thus, we get the following $\left(2n + N^2\right)$-dimensional system of equations, which can serve for calibrating the Lagrange multipliers,

$$\frac{\partial F}{\partial \theta_i^{(\text{out})}} = \sum_{j | a_{ij} = 1} \frac{\exp\left(-\theta_{ij}\right)}{1 - \exp\left(-\theta_{ij}\right)} = s_i^{(\text{out})}, \quad \forall i = 1, \ldots, n, \tag{6.33}$$

$$\frac{\partial F}{\partial \theta_j^{(\text{in})}} = \sum_{i | a_{ij} = 1} \frac{\exp\left(-\theta_{ij}\right)}{1 - \exp\left(-\theta_{ij}\right)} = s_j^{(\text{in})}, \quad \forall j = 1, \ldots, n, \tag{6.34}$$

$$\frac{\partial F}{\partial \theta_{kl}^{(\text{block})}} = \sum_{i \in C_k, j \in C_l | a_{ij} = 1} \frac{\exp\left(-\theta_{ij}\right)}{1 - \exp\left(-\theta_{ij}\right)} = s_{kl}^{(\text{block})}, \quad \forall k, l = 1, \ldots, N. \tag{6.35}$$

Moreover, we can derive the expected link weights as follows. Let W_{bc} denote the random variable representing a link from node b to node c, defined on the probability space $(\mathcal{G}_a, \mathfrak{P}(\mathcal{G}_a), p)$, where p denotes the probability distribution of the ERGM defined in this section. We get,

$$
\begin{aligned}
&\mathbb{E}\left[W_{bc} \mid a_{bc} = 1\right] \\
&\quad = \sum_{w \in \mathcal{G}_a} w_{bc} P(w) \\
&\quad = Z^{-1} \sum_{w \in \mathcal{G}_a} w_{bc} e^{-H(a)} \\
&\quad = \prod_{(i,j)|a_{ij}=1} \left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} \sum_{\{w \in \mathcal{G}_a\}} w_{bc} \prod_{(i,j)|a_{ij}=1} e^{-\theta_{ij} w_{ij}} \\
&\quad = \prod_{(i,j)|a_{ij}=1} \left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} \left[\sum_{\tilde{w}_{bc}=0}^{\infty} \tilde{w}_{bc} e^{-\theta_{bc}\tilde{w}_{bc}} \underbrace{\left(\sum_{w \in \mathcal{G}_a | w_{bc} = \tilde{w}_{bc}} \prod_{\substack{(i,j)|a_{ij}=1 \\ (ij) \notin \{(bc)\}}} e^{-\theta_{ij} w_{ij}}\right)}_{\text{constant for all } \tilde{w}_{bc}}\right] \\
&\quad = \prod_{(i,j)|a_{ij}=1} \left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} \left(\sum_{\tilde{w}_{bc}=0}^{\infty} \tilde{w}_{bc} e^{-\theta_{bc}\tilde{w}_{bc}}\right) \prod_{\substack{(i,j)|a_{ij}=1 \\ (ij) \notin \{(bc)\}}} \left[\frac{1}{1-e^{-\theta_{ij}}}\right], \quad \text{analogous to Eq. (6.31)} \\
&\quad = \left[\frac{1}{1-e^{-\theta_{bc}}}\right]^{-1} \sum_{\tilde{w}_{bc}=0}^{\infty} \tilde{w}_{bc} e^{-\theta_{bc}\tilde{w}_{bc}} \\
&\quad = \left(1-e^{-\theta_{bc}}\right) \sum_{\tilde{w}_{bc}=0}^{\infty} \frac{\partial}{\partial \theta_b^{(\text{out})}} \left[-e^{-\theta_{bc}\tilde{w}_{bc}}\right] \\
&\quad = \left(1-e^{-\theta_{bc}}\right) \frac{\partial}{\partial \theta_b^{(\text{out})}} \left[-\sum_{\tilde{w}_{bc}=0}^{\infty} e^{-\theta_{bc}\tilde{w}_{bc}}\right] \\
&\quad = \left(1-e^{-\theta_{bc}}\right) \frac{\partial}{\partial \theta_b^{(\text{out})}} \left[\frac{-1}{1-e^{-\theta_{bc}}}\right], \qquad \text{by the geometric series} \\
&\quad = \left(1-e^{-\theta_{bc}}\right) \frac{e^{-\theta_{bc}}}{\left(1-e^{-\theta_{bc}}\right)^2}, \\
&\quad = \frac{e^{-\theta_{bc}}}{1-e^{-\theta_{bc}}}.
\end{aligned}
\tag{6.36}
$$

Hence, summing up

$$\begin{aligned}
\mathbb{E}\left[W_{ij} \mid a_{ij}=1\right] &= \frac{\exp\left(-\theta_{ij}\right)}{1-\exp\left(-\theta_{ij}\right)}, && \forall i,j=1,\dots,n,\\
\mathbb{E}\left[W_{ij} \mid a_{ij}=0\right] &= 0, && \forall i,j=1,\dots,n.
\end{aligned}$$

Furthermore, we can derive the probability that a random link W_{bc} takes a specific weight $w^*_{bc} \in \mathbb{N}_{\geq 0}$,

$$\begin{aligned}
&P\left(W_{bc}=w^*_{bc} \mid a_{bc}=1\right)\\
&\quad= \sum_{\left\{w\in\mathcal{G}_a : w_{bc}=w^*_{bc}\right\}} P\left(w\right)\\
&\quad= Z^{-1} \sum_{\left\{w\in\mathcal{G}_a : w_{bc}=w^*_{bc}\right\}} e^{-H(a)}\\
&\quad= \prod_{(i,j)|a_{ij}=1}\left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} \sum_{\left\{w\in\mathcal{G}_a : w_{bc}=w^*_{bc}\right\}} \prod_{(i,j)|a_{ij}=1} e^{-\theta_{ij}w_{ij}}\\
&\quad= \prod_{(i,j)|a_{ij}=1}\left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} \left[e^{-\theta_{bc}w^*_{bc}} \sum_{\left\{w\in\mathcal{G}_a : w_{bc}=w^*_{bc}\right\}} \prod_{\substack{(i,j)|a_{ij}=1\\(ij)\notin\{(bc)\}}} e^{-\theta_{ij}w_{ij}}\right]\\
&\quad= \prod_{(i,j)|a_{ij}=1}\left[\frac{1}{1-e^{-\theta_{ij}}}\right]^{-1} e^{-\theta_{bc}w^*_{bc}} \prod_{\substack{(i,j)|a_{ij}=1\\(ij)\notin\{(bc)\}}}\left[\frac{1}{1-e^{-\theta_{ij}}}\right], \quad \text{analogous to Eq. (6.31)}\\
&\quad= \left[\frac{1}{1-e^{-\theta_{bc}}}\right]^{-1} e^{-\theta_{bc}w^*_{bc}}\\
&\quad= \left(1-e^{-\theta_{bc}}\right) e^{-\theta_{bc}w^*_{bc}}.
\end{aligned} \tag{6.37}$$

Since the links are pairwise independet, we get the following functional form for the probability of a certain graph $w^* \in \mathcal{G}_a$,

$$P\left(W=w^*\right) = \prod_{(i,j)|a_{ij}=1} P\left(W_{ij}=w^*_{ij}\right). \tag{6.38}$$

The question whether a solution for an ERGM exists is non-trivial and to the best of our knowledge still constitutes an open problem in the wider realm of the theory of ERMGs. Regarding the empirical analysis of the EU interbank market, conducted in Section 6.5, our algorithm for parameter estimation, see Section 6.4, was always able to quickly find a solution with minimal error. From the general theory of ERMGs, we know that if the set of solving distributions is non-empty, then all Lagrange parameters are unique up to possible equivalence classes.

Theorem 6.3.3 (Uniqueness of a solution for the ERGM)
If the set of probability measures $p \in \mathcal{P}$ that satisfy all constraints of Eq. (6.29) is non-empty, then the solving distribution function of the ERGM is unique up to certain equivalence classes. More precisely, the sum $\theta_{ij} = \left(\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_{ij}^{(\text{block})}\right)$ is unique for all $i, j = 1, \ldots, n$ where $a_{ij} = 1$. Any set $\theta^{(\text{out})}, \theta^{(\text{in})}, \theta^{(\text{block})}$ that matches the unique sums, defines the same solving probability measure and constitutes an equivalence class.

Proof
From the general theory of maximum entropy problems, we know that the solving probability measure is unique, see Theorem 2.1.3. Let's assume there exists a second set of parameters $\tilde{\theta}$ solving the ERGM, i.e. defining the same probability measure. This especially means that $p_\theta(w) = p_{\tilde{\theta}}(w)$ for all $w \in \mathcal{G}_a$. Let $w^{(0)}$ denote the graph where all $w_{ij}^{(0)} = 0$, then we get

$$\begin{aligned} & p_\theta(w^{(0)}) && = p_{\tilde{\theta}}(w^{(0)}) \\ \Leftrightarrow \quad & Z_\theta^{-1} \underbrace{e^{-H_\theta(w^{(0)})}}_{=1} && = Z_{\tilde{\theta}}^{-1} \underbrace{e^{-H_{\tilde{\theta}}(w^{(0)})}}_{=1} \\ \Leftrightarrow \quad & Z_\theta && = Z_{\tilde{\theta}}. \end{aligned}$$

Let $w^{(a,b)}$ denote the graph where all elements are 0 and $w_{ab}^{(a,b)} = 1$, then we get

$$\begin{aligned} & p_\theta(w^{(a,b)}) && = p_{\tilde{\theta}}(w^{(a,b)}) \\ \Leftrightarrow \quad & Z_\theta^{-1} \underbrace{e^{-H_\theta(w^{(a,b)})}}_{=\exp(-\theta_{ab})} && = Z_{\tilde{\theta}}^{-1} \underbrace{e^{-H_{\tilde{\theta}}(w^{(a,b)})}}_{=\exp\left(-\tilde{\theta}_{ab}\right)} \\ \Leftrightarrow \quad & \theta_{ab} && = \tilde{\theta}_{ab}. \end{aligned}$$

Hence, it follows that all sums $\theta_{ij} = \left(\theta_i^{(\text{out})} + \theta_j^{(\text{in})} + \theta_{ij}^{(\text{block})}\right)$ are unique for all $i, j = 1, \ldots, n$ where $a_{ij} = 1$. □

6.4 Model Calibration

Calibrating the Lagrange multipliers of the ERGM presented in Section 6.3 is demanding, since the system of equations is nonlinear, the number of parameters is (in most cases) very big, and because of the upper bound constraints of certain sums of the Lagrange multipliers (see Eq. (6.32)). For example, the reconstruction of the EU interbank network, conducted in Section 6.5, comprises $(2n + N^2) = (2 \cdot 3,469 + 29^2) = 7,779$ Lagrange multipliers.[1] To solve this problem, we make use of the structural characteristics of the expected link weights and the system of equations.

To simplify notation, we rewrite the problem in terms of $x_i^{(\text{out})} := \exp\left(-\theta_i^{(\text{out})}\right)$, $x_j^{(\text{in})} := \exp\left(-\theta_j^{(\text{in})}\right)$, and $x_{ij}^{(\text{block})} := \exp\left(-\theta_{ij}^{(\text{block})}\right)$, for all $i, j = 1, \ldots, n$. From the previous section we know that the matrix of expected link weights takes the following form, see Section 6.3, for $a \in \{0,1\}^{n \times n}$ a given adjacency matrix,

$$\mathbb{E}\left[W \mid a\right] = \left(\frac{x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})}}{1 - x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})}} a_{ij}\right)_{i,j=1,\ldots,n}. \tag{6.39}$$

The system of equations is essentially given by requiring the row sums, the column sums, and the block weights of $\mathbb{E}\left[W \mid a\right]$ to equal the desired weights $s^{(\text{out})}$, $s^{(\text{in})}$, and $s^{(\text{block})}$, respectively. Moreover, the condition of Eq. (6.32) can be rewritten to $x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})} < 1$ for all (i, j) with $a_{ij} = 1$.

Next, we note that for fixed admissible parameters $x^{(\text{in})} \in \mathbb{R}^n_{>0}$ and $x^{(\text{block})} \in \mathbb{R}^{N \times N}_{>0}$, the equations of the row sums simplify to n independent, univariate non-linear functions. In addition, these functions are on the admissible support continuously differentiable and strictly monotonically increasing. Hence, this subproblem can easily be tackled for example by the univariate Newton's method. The same holds true when considering only the subset of column (block) parameters and column (block) equations. Therefore, we implement an iterative algorithm updating either the row, or the column, or the block parameters in each iteration w.r.t. the row, column, or block equations, respectively.

More precisely, in each iteration we compute three sums of absolute errors consisting of the row constraints, the column constraints, and the block constraints. The subset (row, column, or block) with the highest error is selected for updating the respective subset of parameters by one step of the Newton's method. In addition, we scale the step of the Newton's method by a global *stepsize* parameter, in order to control for the impact on the disregarded equations. As the algorithm moves towards a minimum, the *stepsize* parameter is gradually decreased. To ensure that the bounds of Eq. (6.32) are always satisfied, we adjust a parameter update that would violate the lower (resp. upper) bound, by setting the concerned parameter to the smallest (resp. largest) admissible value. The algorithm terminates, once an acceptable remaining error is reached.

As can be seen from Eq. (6.39), big desired weights $s^{(\text{out})}, s^{(\text{in})}$, and $s^{(\text{block})}$, and thus big expected link weights, mean that the product of the parameters $x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})}$ gets pushed closer to 1. More precisely, the bigger the desired weights $s^{(\text{out})}, s^{(\text{in})}$, and $s^{(\text{block})}$, the more will the parameters cluster just below the value of 1, and hence the

[1] There are 28 EU countries plus a Rest-of-the-World node modeling interbank linkages between EU and non–EU banks. The Rest-of-the-World node gives rise to an additional row and column of blocks.

more difficult the calibration. Therefore, we consider relative instead of absolute weights, i.e. all row, column, and block weights are divided by the total weight of the network. This allows the parameters to spread more broadly and thus facilitates the calibration. Furthermore, as starting parameters we choose $x = s/(1+s)$ which in our experiments works well, but any other starting values can be used likewise.

The pseudo-code of the algorithm is presented in Algorithm 1. Regarding the reconstruction of the EU interbank market, conducted in Section 6.5, the proposed algorithm is reasonably fast, see Table 6.2.

Algorithm 1: Parameter calibration of the ERGM discussed in Section 6.3.

Function `calibrate_ERGM(`*function:* `expectedWeights`, *vector:* $s^{(\text{out})}$, *vector:* $s^{(\text{in})}$, *matrix:* $s^{(\text{block})}$, *matrix:* a`)`

for $i \leftarrow 1$ ***to*** n **do** `// initialize parameters`
    $x_i^{(\text{out})} \leftarrow s_i^{(\text{out})} / \left(s_i^{(\text{out})} + 1\right)$
    $x_i^{(\text{in})} \leftarrow s_i^{(\text{in})} / \left(s_i^{(\text{in})} + 1\right)$
end
for $k \leftarrow 1$ ***to*** N **do**
    for $l \leftarrow 1$ ***to*** N **do**
        $x_{kl}^{(\text{block})} \leftarrow s_{kl}^{(\text{block})} / \left(\max\left\{s^{(\text{block})}\right\}\right)$
    end
end

$w \leftarrow$ `expectedWeights(`$x^{(\text{out})}, x^{(\text{in})}, x^{(\text{block})}$`)` `// compute expected link weights`
$errorRows \leftarrow \left(\sum \mid \texttt{rowSums}(w) - s^{(\text{out})} \mid\right) / \left(\sum s^{(\text{out})}\right)$ `// get errors`
$errorColumns \leftarrow \left(\sum \mid \texttt{colSums}(w) - s^{(\text{in})} \mid\right) / \left(\sum s^{(\text{in})}\right)$
$errorBlocks \leftarrow \left(\sum \mid \texttt{blockSums}(w) - s^{(\text{block})} \mid\right) / \left(\sum s^{(\text{block})}\right)$

$errorAccept \leftarrow 1\%$ `// set acceptable error threshold`
$errorVec \leftarrow [errorRows, errorColumns, errorBlocks]$ `// initialize auxiliary variables`
$stepsize \leftarrow 0.1$
$stepAdj \leftarrow 100$

---

while $\max\{errorRows, errorColumns, errorBlocks\} > errorAccept$ **do**
  switch $\max\{errorRows, errorColumns, errorBlocks\}$ **do**
    case $errorRows$ `// update all` $x_i^{(\text{out})}$
      for $i \leftarrow 1$ ***to*** n **do**
        $primeRowSum_i \leftarrow$
        $\frac{\partial}{\partial x_i^{(\text{out})}} \left(\sum_{j=1}^{n} \frac{x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})}}{1 - x_i^{(\text{out})} x_j^{(\text{in})} x_{ij}^{(\text{block})}} \mathbb{1}_{\{a_{ij}=1\}} - s_i^{(\text{out})} \right) x_i^{(\text{out})} \leftarrow$
        $x_i^{(\text{out})} - stepsize \cdot \left(\texttt{rowSums}(w)_i - s_i^{(\text{out})} \right) / primeRowSum_i$
        if $x_i^{(\text{out})} > admissible\ support$ **then**
          $x_i^{(\text{out})} \leftarrow \max\{\text{admissible support}\}$
        else if $x_i^{(\text{out})} < admissible\ support$ **then**
          $x_i^{(\text{out})} \leftarrow \min\{\text{admissible support}\}$
    case $errorColumns$ `// update all` $x_i^{(\text{in})}$
      analogously
    case $errorBlocks$ `// update all` $x_i^{(\text{block})}$
      analogously
  $w \leftarrow \texttt{expectedWeights}(x^{(\text{out})}, x^{(\text{in})}, x^{(\text{block})})$ `// compute expected link weights`
  $errorRows \leftarrow \left(\sum \mid \texttt{rowSums}(w) - s^{(\text{out})} \mid\right) / \left(\sum s^{(\text{out})}\right)$ `// get errors`
  $errorColumns \leftarrow \left(\sum \mid \texttt{colSums}(w) - s^{(\text{in})} \mid\right) / \left(\sum s^{(\text{in})}\right)$
  $errorBlocks \leftarrow \left(\sum \mid \texttt{blockSums}(w) - s^{(\text{block})} \mid\right) / \left(\sum s^{(\text{block})}\right)$
  $errorVec \leftarrow [errorVec, [errorRows, errorColumns, errorBlocks]]$
  if $\texttt{length}(errorVec) > stepAdj$ **then** `// adjust stepsize parameter`
    if $\texttt{mean}(errorVec(end\text{-}stepAdj{:}end\text{-}stepAdj/2)) - \texttt{mean}(errorVec(end\text{-}stepAdj/2{:}end)) \leq 1e-4$ **then**
      $stepsize \leftarrow stepsize/1.2$
      $errorVec \leftarrow \min\{errorVec\}$
      $x^{(\text{in})} \leftarrow best(x^{(\text{in})})$ `// restart at best solution found so far`
      $x^{(\text{out})} \leftarrow best(x^{(\text{out})})$
      $x^{(\text{block})} \leftarrow best(x^{(\text{block})})$
return $(x^{(\text{out})}, x^{(\text{in})}, x^{(\text{block})})$

6.5 Empirical Case Study: Reconstructing the EU Interbank Network

The model developed in this chapter can be calibrated to reconstruct the EU interbank market. For this purpose, we first discuss how the input variables can be estimated from publicly available data, and subsequently we present the simulation results. The simulated networks allow a detailed assessment of systemic risk, which is demonstrated in Chapter 7.

6.5.1 Data

The network characteristics that are explicitly incorporated in our model and can thus be set as desired, are

(i) interbank assets $s_i^{(\text{out})}$ and liabilities $s_i^{(\text{in})}$ for each bank i in the network;

(ii) the network density of each block, i.e. the number of links $L^{\rightarrow} \in \mathbb{N}^{N \times N}$;

(iii) the reciprocity of each block, i.e. the number of reciprocal links $L^{\leftrightarrow} \in \mathbb{N}^{N \times N}$ symmetric;

(iv) and the weight $s_{kl}^{(\text{block})}$ of each block, $k, l = 1, \dots, N$.

Even though the aggregated data of (ii) - (iv) do not reveal any individual bilateral lending information, these statistics are not readily available. Therefore, we approximate these statistics based on publicly available information. Total interbank assets and liabilities, on the other hand, are published in the banks' balance sheet. We obtain this data from the Bankscope (now Orbis BankFocus) database of Bureau Van Dijk.

Roukny et al. (2014) and Bargigli et al. (2015) are granted access to real data from the German and the Italian central bank and provide detailed empirical analyses on the respective interbank markets. To the best of our knowledge, their works constitute the most extensive descriptions on the topology of financial networks of EU member states. Without further publicly available information, we propose to approximate the network densities within countries based on the information of the German and Italian interbank market. More precisely, for a country with $|C_j|$ banks we suggest a density equal to the average density found in the subgraphs of the German and Italian interbank market, consisting of the $|C_j|$ banks with the highest degree. This idea is motivated by the assumption that the difference in the number of banks in a country is mainly due to the number of small and local banks, while the need for a well connected core of big banks is universal. Hence, we assume that at least the density of the core of the interbank network is similar across countries. In case more information on interbank markets becomes available, the chosen density can easily be adjusted.

The density of the German and Italian interbank market, reduced to a number of best connected banks, can be derived via our earlier work Engel et al. (2019b), see also Chapter 5, on the reconstruction of the unweighted German and Italian interbank market. Fig. 6.2 presents the average densities over 100 simulated German and Italian interbank networks, reduced to subgraphs of banks with the highest degree. Since the degree of a bank highly correlates with its weight, the subgraphs can also be interpreted to contain the biggest banks.

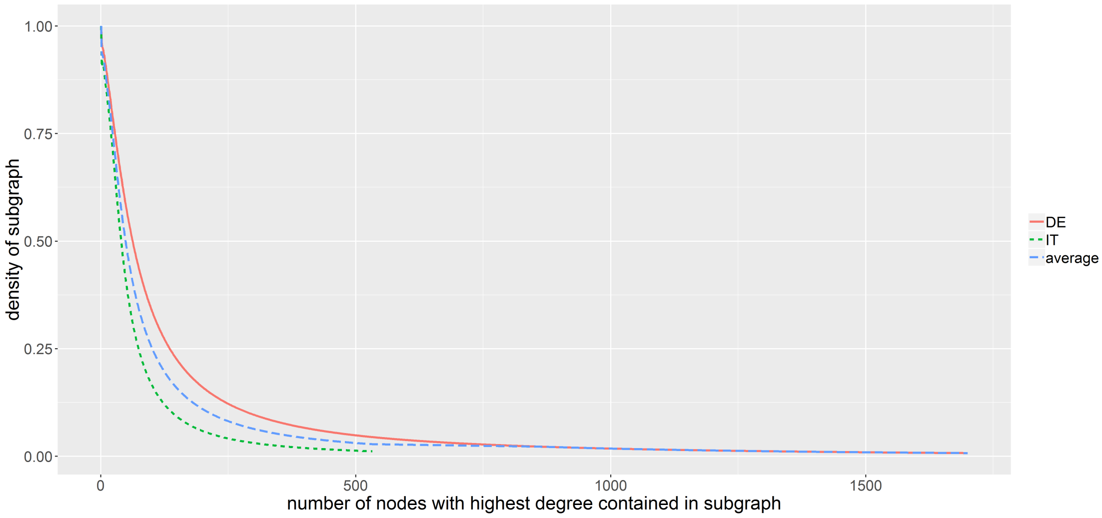


Figure 6.2: Average density over subgraphs of 100 German and 100 Italian simulated interbank networks, as well as the average over both countries. The size of the subgraphs is indicated on the x-axis, and the selection is based on the (descending) degree of the banks.

To every country we assign a network density according to its number of banks and the average density over the German and Italian interbank networks, as presented by the dashed line in Fig. 6.2. A summary over all EU countries, their number of banks and allocated densities, is given in Table 6.1. The densities range from 93% for Malta, a country with only 5 banks, to 1% for Germany, a country with 1415 banks. It seems reasonable that countries comprising only a small number of banks are very well connected, while interbank networks of countries containing a large number of banks are rather sparse.

Next, we discuss the input factor of the block reciprocity. Degree reciprocity ρ_r is defined as the correlation coefficient between the symmetric entries of the adjacency matrix a of a directed graph, i.e. the tendency of nodes to form mutual links,

$$\rho_r := \frac{\sum_{i \neq j} (a_{ij} - \bar{a})(a_{ji} - \bar{a})}{\sum_{i \neq j} (a_{ij} - \bar{a})^2}, \tag{6.40}$$

Table 6.1: Estimated network density for each country (data from 2016).

Country	AT	BE	BG	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	HU
number of banks	527	33	14	22	19	1415	45	4	98	13	247	5	23	12
density	3%	65%	85%	77%	80%	1%	53%	93%	26%	86%	8%	93%	75%	87%

Country	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
number of banks	16	518	6	42	13	5	18	25	108	20	65	12	14	129
density	83%	3%	92%	56%	86%	93%	81%	73%	23%	79%	39%	87%	85%	19%

where $\bar{a}$ denotes the network density. The 'neutral' case $\rho_r = 0$ indicates that the network has exactly as many reciprocal links as expected in a random graph with the same number of vertices and links. Moreover, $\rho_r > 0$ (resp. $\rho_r < 0$) signifies that there are more (resp. less) reciprocal links than expected by chance. For a detailed discussion on reciprocity, see Garlaschelli and Loffredo (2004b). Furthermore, Roukny et al. (2014) and Bargigli et al. (2015) report a reciprocity of 0.31 for the German and 0.45 for the Italian interbank market. Without any further information on the reciprocity of other interbank markets, we decided to set the tendency that banks form reciprocal lending relationships, i.e. the reciprocity of each block, to the average of the two available values, hence, to 0.38.

The estimation of the block weights is non-trivial, since there is no complete and consistent data set on cross border exposures publicly available. In an attempt to fill this gap and *"thus contributing to market discipline and financial stability in the EU"*[2] the European Banking Authority (EBA) conducts a transparency exercise since 2013. As part of this exercise, the EBA discloses interbank credit exposure of 131 (in 2016) European banks disaggregated on country level. We propose to use this data set to construct a prior distribution of credit exposure aggregated on country level. Subsequently, we derive the distribution of the block weights which is as close as possible to the prior distribution, i.e. the EBA data, while fulfilling the given marginals of total interbank assets and liabilities for each country, as given by the BankFocus database. As a measure of divergence we use the Kullback–Leibler divergence, which means we have to solve a simple maximum entropy problem. Furthermore, for the derivation of the block weights, we differentiate between cross-border active banks and domestic banks. As an approximation, we consider those banks as cross-border active which are marked as significant by the ECB[3] or which are classified as global systemically important banks (G-SIBs)[4]. The derivation of the distribution of the block weights is explained step by step in the following.

(1) For those countries for which the EBA data set includes at least one bank with

[2]See https://www.eba.europa.eu/-/eba-transparency-exercise.
[3]See https://www.bankingsupervision.europa.eu/banking/list/criteria/html/index.en.html
[4]See http://www.fsb.org/what-we-do/policy-development/systematically-important-financial-institutions-sifis/global-systemically-important-financial-institutions-g-sifis/

a detailed country level distribution of its interbank credit exposures, we derive the relative distribution of credit exposure, aggregated over all listed banks in a country. The relative distribution is split to all EU countries, the rest of the world, and unallocated, which denotes the difference between total credit exposure and the sum over all listed country exposures.

(2) Countries for which we have a relative exposure distribution from the EBA and which comprise cross-border active banks: We derive block weights by distributing interbank assets of cross-border active banks according to the relative distribution from the EBA data set, and adding interbank assets of the domestic banks to the home country. For 2016 these countries are: AT, BE, CY, DE, ES, FI, FR, GR, IE, IT, LU, LV, MT, NL, PT, SE, SI, UK.

(3) Countries for which we have a relative exposure distribution from the EBA but are missing information on which banks are cross-border active: We approximate the relative amount of interbank assets of cross-border active banks by the mean over all countries with EBA and cross-border active information. This amount is then distributed according to the EBA data, while the corresponding amount of interbank assets of domestic banks is allocated to the home country. For 2016 these countries are: BG, DK, HU.

(4) Countries for which the cross-border active banks are known, but data of the EBA is missing: We approximate the amount of interbank assets of cross-border active banks that is allocated within the home country by the mean over all other domestic distributions of cross-border active banks that are allocated so far. The amount of interbank assets of domestic banks is also allocated to the home country. For 2016 these countries are: CZ, EE, LT, SK.

(5) Countries that are not comprised in the EBA data set and for which information on cross-border active banks is missing: A home bias is added by allocating the mean over all assigned home biases multiplied by total interbank assets of the respective countries. For 2016 these countries are: HR, PL, RO.

(6) Next, we compute how much of the countries interbank assets and liabilities, as given by BankFocus, are still unallocated. For countries for which we have already allocated a higher amount than available according to BankFocus, the value of the unallocated amount is set to zero.

(7) To distribute the unallocated interbank assets to the EU and to the rest of the world in a reasonable way, we allocate an amount of the sum of unallocated interbank assets to the rest of the world, that is proportional to the amount of weight that has been allocated to the rest of the world so far.

(8) The amount of interbank assets which is now still unallocated is spread over all EU blocks according to the unallocated marginals.

(9) In a last step we solve the optimization problem of minimizing the Kullback–Leibler divergence to the thus constructed prior distribution of block weights, subject to the marginal country constraints as given by BankFocus, i.e. total interbank assets and liabilities of each country have to be fulfilled.

The resulting distribution is presented in Fig. 6.3. For most countries, we can identify a clear home bias, visualized by the dark blue colors on the (anti-) diagonal. Furthermore, some countries allocate a substantial amount of their interbank assets to France, the rest of the world, Italy, Spain, UK, and Germany. Regarding interbank liabilities, a substantial amount comes from countries outside of the EU, as well as France and Italy.

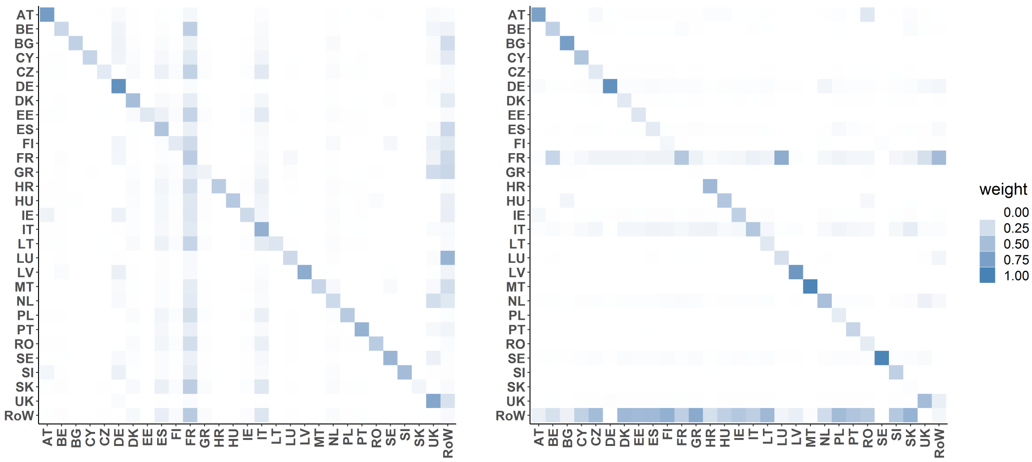


Figure 6.3: Distribution of interbank assets and liabilities, based on the EBA transparency exercise of 2016. The plot on the left shows how each country distributes its interbank assets, i.e. each row sums up to one. The plot on the right shows where the interbank liabilities of each country come from, i.e. each column sums up to one. 'RoW' denotes the 'Rest of the World'.

The last input factor that we need to discuss is the density of cross-border blocks. Since at this point, we have already derived the density of each country and the block weights, we can compute the average weight per link within each country. Without further information on cross-border interbank markets, we propose to take the minimum weight per link of two countries as a proxy for the weight per link of the cross-border block between both countries. This means the number of links in the cross-border matrix of countries k and l is approximated by

$$L_{kl}^{\rightarrow} = \left(\min \left\{ \frac{s_{kk}^{(\text{block})}}{L_{kk}^{\rightarrow}}, \frac{s_{ll}^{(\text{block})}}{L_{ll}^{\rightarrow}} \right\} \right)^{-1} \cdot s_{kl}^{(\text{block})}. \tag{6.41}$$

This section illustrated one approach to estimate the model input factors based on scarce publicly available information. Moreover, these factors only serve for calibration and do not impact the methodological part of the model. Also, in case further aggregated data on financial networks, such as the density, degree distribution, block weights, or reciprocity become available in the future, our model can easily incorporate this information. Actually, policy-makers might already have access to some additional, not publicly available data, which they can use to calibrate the model more accurately.

6.5.2 Simulation Results

This section presents the results of the reconstruction of the EU interbank market. The case study is based on data of 2016, for which BankFocus lists 3,468 unconsolidated EU banks with positive interbank assets and liabilities. Adding a Rest-of-the-World node leads to a network of 3,469 nodes and 29 regions (28 EU countries + Rest-of-the-World), i.e. $29^2 = 841$ blocks. There are two sources of errors that should be differentiated. First, an adjacency matrix might be drawn that has no links in some rows, columns, and blocks, and hence no weight can be allocated. Second, the weight allocation found by Algorithm 1 yields a remaining error. Table 6.2 summarizes the runtime and the error of both parts of the model. The relative error of the row sums (column sums/ block weights) refers to the sum of absolute errors over all rows (columns/ blocks) divided by the sum of all row (column/ block) weights.

Table 6.2: Runtime and error of the fitness model and the ERGM, w.r.t. 100 simulated interbank networks (including the Rest-of-the-World node) and with an acceptable error threshold of 1% in Algorithm 1.

	Fitness Model adjacency matrix		ERGM weight allocation	
	mean	std	mean	std
runtime	36 seconds	NA	2.5 min/ network	NA
relative error:				
row sums	3.86e-04	3.40e-05	9.81e-03	1.42e-04
column sums	3.22e-04	3.05e-05	9.83e-03	1.34e-04
block weights	1.00e-02	1.40e-05	9.85e-03	1.10e-04

To gain some insight into the topology of the simulated networks, Table 6.3 reports the most prominent network statistics. Unfortunately, we do not have access to data on the actual EU interbank market and hence, cannot conduct a detailed assessment of the goodness of fit. However, our model seems to successfully reproduce some commonly reported characteristics of financial networks, such as sparsity, a positive reciprocity, disassortativity, and short paths.

Table 6.3: Mean, standard deviation, and 95% confidence interval of different network statistics, w.r.t. 100 simulated networks (excluding the Rest-of-the-World node).

	mean	std	95% confidence interval
total number of links	69,451	223	[69,408; 69,495]
number of reciprocal links	26,995	163	[26,964; 27,027]
in-degree assortativity	-0.23	0.0026	[-0.23; -0.23]
out-degree assortativity	-0.19	0.0017	[-0.19; -0.19]
directed clustering coefficient	0.66	0.0036	[0.66; 0.66]
undirected clustering coefficient	0.72	0.0032	[0.72; 0.72]
shortest directed path	2.92	0.0079	[2.92; 2.92]
shortest undirected path	2.95	0.0092	[2.94; 2.95]
number of isolated nodes	81	8	[80; 83]
largest strongly connected component	2,828	17	[2,824; 2,831]
largest weakly connected component	3,387	8	[3,385; 3,388]

Since the simulated networks serve as a basis for an assessment of systemic risk, the network similarity within the drawn sample is also of interest. If the location of the links and their weight does not change much across the sample, systemic risk results will be very stable as well. With increasing variation in the sampled networks, however, we expect an increasing variance and uncertainty in the quantification of systemic risk. Therefore, we now analyze the similarity of the sampled adjacency matrices and the allocated weights. The similarity between two realizations a and $\tilde{a}$ of the link probability matrix $A \sim Bin\big(1, (p_{ij}, p_{ji})\big)^{n \times n}$ can be derived analytically. The expected number of links that exist in both adjacency matrices (drawn independent of each other) is given by

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\neq j} \mathbb{1}_{\{a_{ij}=1\wedge\tilde{a}_{ij}=1\}}\right] \\
&= \mathbb{E}\Bigg[\sum_{i<j} \mathbb{1}_{\{a_{ij}=1\wedge a_{ji}=0\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=0\}} + \mathbb{1}_{\{a_{ij}=1\wedge a_{ji}=0\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=1\}} \\
&\qquad + \mathbb{1}_{\{a_{ij}=0\wedge a_{ji}=1\wedge\tilde{a}_{ij}=0\wedge\tilde{a}_{ji}=1\}} + \mathbb{1}_{\{a_{ij}=0\wedge a_{ji}=1\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=1\}} \\
&\qquad + 2\mathbb{1}_{\{a_{ij}=1\wedge a_{ji}=1\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=1\}} + \mathbb{1}_{\{a_{ij}=1\wedge a_{ji}=1\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=0\}} + \mathbb{1}_{\{a_{ij}=1\wedge a_{ji}=1\wedge\tilde{a}_{ij}=0\wedge\tilde{a}_{ji}=1\}}\Bigg] \\
&= \sum_{i<j} \mathbb{P}\left(a_{ij}=1\wedge a_{ji}=0\wedge\tilde{a}_{ij}=1\wedge\tilde{a}_{ji}=0\right) + \ldots + \mathbb{P}\left(a_{ij}=1\wedge a_{ji}=1\wedge\tilde{a}_{ij}=0\wedge\tilde{a}_{ji}=1\right) \\
&= \sum_{i<j} p_{ij}^{(1,0)}\left(p_{ij}^{(1,0)}+p_{ij}^{(1,1)}\right) + p_{ij}^{(0,1)}\left(p_{ij}^{(0,1)}+p_{ij}^{(1,1)}\right) + p_{ij}^{(1,1)}\left(2p_{ij}^{(1,1)}+p_{ij}^{(1,0)}+p_{ij}^{(0,1)}\right).
\end{aligned}
\tag{6.42}
$$

The expected numbers of links that differs and that is absent in a and $\tilde{a}$ can be computed analogously. Table 6.4 summarizes the expected similarity and dissimilarity between two sampled adjacency matrices of the reconstructed EU interbank market. In expectation, almost half of the sampled links in the network will be identical in both realizations, and half of the sampled links will change location. Regarding the sampled zeros in the adjacency matrices, i.e. non-existing links, on average 99.7% of the zeros will be identical in two simulated networks.

Table 6.4: Expected similarities and dissimilarities in the sampled adjacency matrices modeling the EU interbank market (excluding the Rest-of-the-World node) if two independent simulation runs are drawn.

fraction of existing links that is identical in two runs	fraction of existing links that differs in two runs	fraction of absent links that is identical in two runs
$\frac{\mathbb{E}\left[\sum_{i,j}\mathbb{1}_{\{a_{ij}=1\wedge\tilde{a}_{ij}=1\}}\right]}{\mathbb{E}\left[\sum_{i,j}a_{ij}\right]}$ $= 48.81\%$	$\frac{\mathbb{E}\left[\sum_{i,j}\mathbb{1}_{\{a_{ij}=1\wedge\tilde{a}_{ij}=0\}}\right]}{\mathbb{E}\left[\sum_{i,j}a_{ij}\right]}$ $= 51.19\%$	$\frac{\mathbb{E}\left[\sum_{i,j}\mathbb{1}_{\{a_{ij}=0\wedge\tilde{a}_{ij}=0\}}\right]}{\mathbb{E}\left[\sum_{i,j}1-a_{ij}\right]}$ $= 99.70\%$

Next, we analyze the similarity between the allocated weights. Since the parameters of the ERGM are recalibrated for every realization of the adjacency matrix, the similarity between the weights cannot be derived analytically. Therefore, we compute two empirical similarity measures: the relative difference and the cosine similarity. Let w and $\tilde{w}$ denote two realizations of the EU interbank market. The relative difference between w and $\tilde{w}$

is given by

$$\frac{\sum_{ij}|w_{ij} - \tilde{w}_{ij}|}{\sum_{ij} w_{ij} + \sum_{ij} \tilde{w}_{ij}}.$$

Comparing all 100 simulated networks pairwise yields an average relative difference of 0.18 and a standard deviation of 0.002. The cosine similarity is defined as

$$\frac{\sum_{i,j} w_{ij}\tilde{w}_{ij}}{\sqrt{\sum_{i,j} w_{ij}^2}\sqrt{\sum_{i,j} \tilde{w}_{ij}^2}}. \tag{6.43}$$

Interpreting both networks as n^2-dimensional vectors, the cosine similarity gives the cosine of the angle between the two vectors. Since all weights are non-negative, the cosine similarity is bounded by $[0, 1]$ with 1 (resp. 0) signifying the strongest (resp. least) possible similarity. Comparing the sampled networks pairwise, gives an average cosine similarity of 0.98 and a standard deviation of 0.002.

The high cosine similarity together with the high number of changing links and a substantial difference in weight allocation, suggest that links with high weights, connecting big banks, stay quite constant over the set of sampled networks, while links with small weights, involving at least one small bank, vary notably (in existence and weight). This assumption can be verified by plotting the elements of sampled network matrices against each other, see Fig. 6.4.

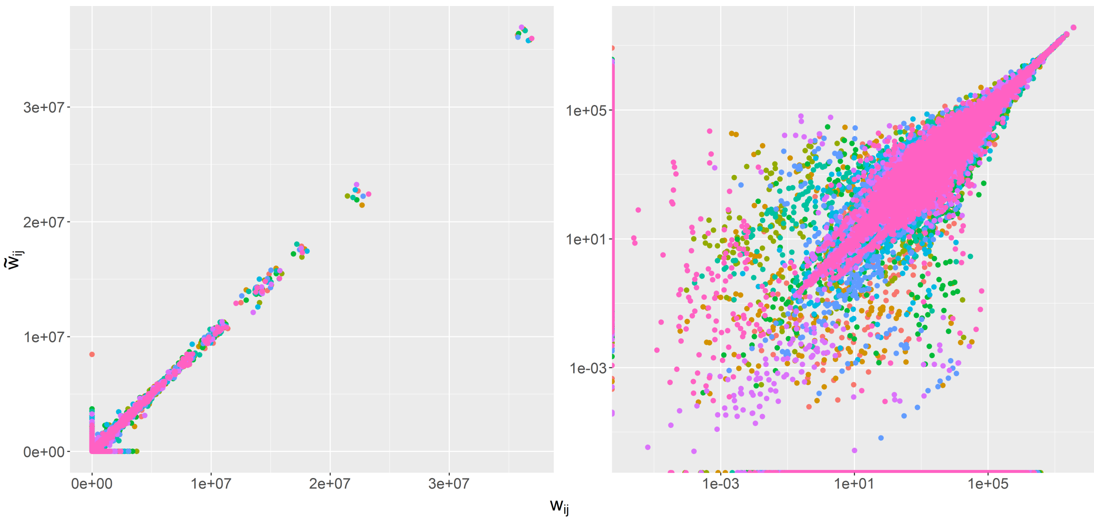


Figure 6.4: Scatterplot of link weights w_{ij} and $\tilde{w}_{ij}$, comparing 10 simulated networks pairwise (i.e. network 1 vs. network 2, network 2 vs. network 3, ..., network 9 vs. network 10). Links that do not exist in neither of the two respectively considered networks are omitted. The figure on the right is in log-log scale.

6.6 Supplementary Information

In this section we provide a detailed proof for Theorem 6.2.1. For the reader's convenience the theorem and the necessary elements are recalled in the following.

Theorem 6.2.1 (Existence of a solution for the extended fitness model) *Consider four vectors $s^{(\mathrm{out,k})}, s^{(\mathrm{in,k})} \in \mathbb{R}^{n_k}_{>0}$, and $s^{(\mathrm{out,l})}, s^{(\mathrm{in,l})} \in \mathbb{R}^{n_l}_{>0}$. Let the random matrices $A^{(\mathrm{kl})} \in \{0,1\}^{n_k \times n_l}$ and $A^{(\mathrm{lk})} \in \{0,1\}^{n_l \times n_k}$ be defined by the probability function P as given by Eqs. (6.44) to (6.46). For any feasible number of*

(i) reciprocal links $L^{\leftrightarrow}_{kl} \in [0, n_k n_l)$,

(ii) and links $\tilde{L}^{\rightarrow}_{kl} := L^{\rightarrow}_{kl} - L^{\leftrightarrow}_{kl}$ and $\tilde{L}^{\rightarrow}_{lk} := L^{\rightarrow}_{lk} - L^{\leftrightarrow}_{kl}$ with $\left(\tilde{L}^{\rightarrow}_{kl} + \tilde{L}^{\rightarrow}_{lk}\right) \in (0, n_k n_l - L^{\leftrightarrow}_{kl})$,

there exist $z_{kl}, z_{lk}, r_{kl} \in \mathbb{R}_{\geq 0}$, such that Eqs. (6.47) to (6.49) are fulfilled.

The probability function P is defined by

$$P\left(A_{ij} = 1 \wedge A_{ji} = 1\right) = \frac{r_{ij}^2 z_{ij} z_{ji} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})}}{1 + z_{ij} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} + z_{ji} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})}}, \tag{6.44}$$

$$P\left(A_{ij} = 1 \wedge A_{ji} = 0\right) = \frac{z_{ij} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})}}{1 + z_{ij} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} + z_{ji} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})}}, \tag{6.45}$$

$$P\left(A_{ij} = 0 \wedge A_{ji} = 0\right) = \frac{1}{1 + z_{ij} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} + z_{ji} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})} + r_{ij}^2 z_{ij} z_{ji} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})}}. \tag{6.46}$$

In expectation the following has to hold

$$\mathbb{E}\left[\sum_{i \in C_k, j \in C_l} A_{ij}\right] = L^{\rightarrow}_{kl}, \qquad \text{(directed links)} \tag{6.47}$$

$$\mathbb{E}\left[\sum_{i \in C_l, j \in C_k} A_{ij}\right] = L^{\rightarrow}_{lk}, \qquad \text{(directed links)} \tag{6.48}$$

$$\mathbb{E}\left[\sum_{i \in C_k, j \in C_l} A_{ij} A_{ji}\right] = \mathbb{E}\left[\sum_{i \in C_l, j \in C_k} A_{ij} A_{ji}\right] = L^{\leftrightarrow}_{kl}. \qquad \text{(reciprocal links)} \tag{6.49}$$

Proof of Theorem 6.2.1
We start with the interior case. For $L_{kl}^{\leftrightarrow}, \tilde{L}_{kl}^{\rightarrow}, \tilde{L}_{lk}^{\rightarrow} \in \mathbb{R}_{>0}$ with $\left(L_{kl}^{\leftrightarrow} + \tilde{L}_{kl}^{\rightarrow} + \tilde{L}_{lk}^{\rightarrow}\right) \in (0, n_k n_l)$, we have to show, that there exist $z_{kl}, z_{lk}, r_{kl} \in \mathbb{R}_{>0}$, such that

$$L_{kl}^{\leftrightarrow} = \sum_{i \in C_k, j \in C_l} \frac{r_{kl}^2 z_{kl} z_{lk} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z_{kl} s_i^{(\text{out})} s_j^{(\text{in})} + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})} + r_{kl}^2 z_{kl} z_{lk} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.50}$$

$$\tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{z_{kl} s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z_{kl} s_i^{(\text{out})} s_j^{(\text{in})} + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})} + r_{kl}^2 z_{kl} z_{lk} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.51}$$

$$\tilde{L}_{lk}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{z_{lk} s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z_{kl} s_i^{(\text{out})} s_j^{(\text{in})} + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})} + r_{kl}^2 z_{kl} z_{lk} s_i^{(\text{out})} s_j^{(\text{in})} s_j^{(\text{out})} s_i^{(\text{in})}} \tag{6.52}$$

hold.

For simplicity we write $s_{ij} := s_i^{(\text{out})} s_j^{(\text{in})}$ and $s_{ji} := s_j^{(\text{out})} s_i^{(\text{in})}$ in the following. Further, we substitute $y := r_{kl}^2 z_{kl} z_{lk}$ in Eqs. (6.50) to (6.52). It suffices to show that there exist $z_{kl}, z_{lk}, y \in \mathbb{R}_{>0}$ such that

$$L_{kl}^{\leftrightarrow} = \sum_{i \in C_k, j \in C_l} \frac{y s_{ij} s_{ji}}{1 + z_{kl} s_{ij} + z_{lk} s_{ji} + y s_{ij} s_{ji}}, \tag{6.53}$$

$$\tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{z_{kl} s_{ij}}{1 + z_{kl} s_{ij} + z_{lk} s_{ji} + y s_{ij} s_{ji}}, \tag{6.54}$$

$$\tilde{L}_{lk}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{z_{lk} s_{ji}}{1 + z_{kl} s_{ij} + z_{lk} s_{ji} + y s_{ij} s_{ji}} \tag{6.55}$$

hold. If z_{kl}, z_{lk}, y exist, then there also exists $r^2 = y/(z_{kl} z_{lk})$.

This can be proved by the Bolzano–Poincaré–Miranda theorem[5].

Theorem 6.6.1 (Bolzano–Poincaré–Miranda)
Let I denote a parallelotope in $\mathbb{R}^n$, i.e. for $a, b \in \mathbb{R}$, let $I := \{\boldsymbol{x} \in [a,b]^n\}$. Let $I_i^{\ominus} := \{\boldsymbol{x} \in I : x_i = a\}$ and $I_i^{\oplus} := \{\boldsymbol{x} \in I : x_i = b\}$. Further, let $f : I \to \mathbb{R}^n$, $f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x}))$ be a continuous map such that $f_i\left(I_i^{\ominus}\right) \subset (-\infty, 0]$ and $f_i\left(I_i^{\oplus}\right) \subset [0, \infty)$ for $i = 1, \ldots, n$, then there exists $\boldsymbol{x}_0 \in I$ such that $f(\boldsymbol{x}_0) = (0, \ldots, 0)$.

We split the proof in two cases, depending on the values of $L_{kl}^{\leftrightarrow}, \tilde{L}_{kl}^{\rightarrow}$, and $\tilde{L}_{lk}^{\rightarrow}$.

[5]See for example Turzański (2012), Theorem 2, and Mawhin (2013), and the references therein.

Case 1: For

$$\begin{aligned} L_{kl}^{\leftrightarrow} &\leq \sum_{i \in C_k, j \in C_l} \frac{s_{ij} s_{ji}}{s_{ij} + s_{ji} + s_{ij} s_{ji}}, \\ \tilde{L}_{kl}^{\rightarrow} &\leq \sum_{i \in C_k, j \in C_l} \frac{s_{ij}}{s_{ij} + s_{ji} + s_{ij} s_{ji}}, \\ \tilde{L}_{lk}^{\rightarrow} &\leq \sum_{i \in C_k, j \in C_l} \frac{s_{ji}}{s_{ij} + s_{ji} + s_{ij} s_{ji}}. \end{aligned} \tag{6.56}$$

For Case 1, define the parameter support by $(y, z_{kl}, z_{lk}) \in \left[m^{-1}, m\right]^3$, for $m \in \mathbb{R}_{>0}$, $m \gg 1$. Furthermore, consider Eq. (6.53) as a function f_1 of y, z_{kl}, and z_{lk}, i.e.

$$f_1\left(y, z_{kl}, z_{lk}\right) := \left[\sum_{i \in C_k, j \in C_l} \frac{y s_{ij} s_{ji}}{1 + z_{kl} s_{ij} + z_{lk} s_{ji} + y s_{ij} s_{ji}}\right] - L_{kl}^{\leftrightarrow}. \tag{6.57}$$

Note that f_1 is continuous, strictly monotonically increasing in y, and strictly monotonically decreasing in z_{kl}, z_{lk}. Thus, we get the following boundary values

$$\begin{aligned} \max_{z_{kl}, z_{lk}} f_1\left(m^{-1}, z_{kl}, z_{lk}\right) &= f_1\left(m^{-1}, m^{-1}, m^{-1}\right) \\ &= \left[\sum_{i \in C_k, j \in C_l} \frac{m^{-1} s_{ij} s_{ji}}{1 + m^{-1} s_{ij} + m^{-1} s_{ji} + m^{-1} s_{ij} s_{ji}}\right] - L_{kl}^{\leftrightarrow} \\ &\overset{m \to \infty}{\rightarrow} -L_{kl}^{\leftrightarrow} \leq 0, \end{aligned} \tag{6.58}$$

and

$$\begin{aligned} \min_{z_{kl}, z_{lk}} f_1\left(m, z_{kl}, z_{lk}\right) &= f_1\left(m, m, m\right) \\ &= \left[\sum_{i \in C_k, j \in C_l} \frac{m s_{ij} s_{ji}}{1 + m s_{ij} + m s_{ji} + m s_{ij} s_{ji}}\right] - L_{kl}^{\leftrightarrow} \\ &\overset{m \to \infty}{\rightarrow} \left[\sum_{i \in C_k, j \in C_l} \frac{s_{ij} s_{ji}}{s_{ij} + s_{ji} + s_{ij} s_{ji}}\right] - L_{kl}^{\leftrightarrow} \overset{Eq.\ (6.56)}{\geq} 0. \end{aligned} \tag{6.59}$$

Analogously, consider Eq. (6.54) as a function f_2 of y, z_{kl} and z_{lk}, i.e.

$$f_2\left(y, z_{kl}, z_{lk}\right) := \left[\sum_{i \in C_k, j \in C_l} \frac{z_{kl} s_{ij}}{1 + z_{kl} s_{ij} + z_{lk} s_{ji} + y s_{ij} s_{ji}}\right] - \tilde{L}_{kl}^{\rightarrow}. \tag{6.60}$$

Note that f_2 is continuous, strictly monotonically increasing z_{kl}, and strictly monotonically decreasing in y, z_{lk}. Thus, we get the following boundary values

$$\begin{aligned}\max_{y,z_{lk}} f_2\left(y, m^{-1}, z_{lk}\right) &= f_2\left(m^{-1}, m^{-1}, m^{-1}\right)\\&= \left[\sum_{i\in C_k, j\in C_l} \frac{m^{-1}s_{ij}}{1+m^{-1}s_{ij}+m^{-1}s_{ji}+m^{-1}s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{kl}}\\&\overset{m\to\infty}{\to} -\tilde{L}_{\overrightarrow{kl}} \leq 0,\end{aligned} \tag{6.61}$$

and

$$\begin{aligned}\min_{y,z_{lk}} f_2\left(y, m, z_{lk}\right) &= f_2\left(m, m, m\right)\\&= \left[\sum_{i\in C_k, j\in C_l} \frac{m s_{ij}}{1+m s_{ij}+m s_{ji}+m s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{kl}}\\&\overset{m\to\infty}{\to} \left[\sum_{i\in C_k, j\in C_l} \frac{s_{ij}}{s_{ij}+s_{ji}+s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{kl}} \overset{Eq.\ (6.56)}{\geq} 0.\end{aligned} \tag{6.62}$$

Analogously, consider Eq. (6.55) as a function f_2 of y, z_{kl}, and z_{lk}, i.e.

$$f_3\left(y, z_{kl}, z_{lk}\right) := \left[\sum_{i\in C_k, j\in C_l} \frac{z_{lk}s_{ji}}{1+z_{kl}s_{ij}+z_{lk}s_{ji}+y s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{lk}}. \tag{6.63}$$

Note that f_3 is continuous, strictly monotonically increasing z_{lk}, and strictly monotonically decreasing in y, z_{kl}. Thus, we get the following boundary values

$$\begin{aligned}\max_{y,z_{kl}} f_3\left(y, z_{kl}, m^{-1}\right) &= f_3\left(m^{-1}, m^{-1}, m^{-1}\right)\\&= \left[\sum_{i\in C_k, j\in C_l} \frac{m^{-1}s_{ji}}{1+m^{-1}s_{ij}+m^{-1}s_{ji}+m^{-1}s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{lk}}\\&\overset{m\to\infty}{\to} -\tilde{L}_{\overrightarrow{lk}} \leq 0,\end{aligned} \tag{6.64}$$

and

$$\begin{aligned}\min_{y,z_{kl}} f_3\left(y, z_{kl}, m\right) &= f_3\left(m, m, m\right)\\&= \left[\sum_{i\in C_k, j\in C_l} \frac{m s_{ji}}{1+m s_{ij}+m s_{ji}+m s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{lk}}\\&\overset{m\to\infty}{\to} \left[\sum_{i\in C_k, j\in C_l} \frac{s_{ji}}{s_{ij}+s_{ji}+s_{ij}s_{ji}}\right] - \tilde{L}_{\overrightarrow{lk}} \overset{Eq.\ (6.56)}{\geq} 0.\end{aligned} \tag{6.65}$$

Hence, it follows from the Bolzano–Poincaré–Miranda theorem, that there exist $(y^*, z^*_{kl}, z^*_{lk}) \in [m^{-1}, m]^3$, for m big enough, such that $f_1(y^*, z^*_{kl}, z^*_{lk}) = 0$, $f_2(y^*, z^*_{kl}, z^*_{lk}) = 0$, and $f_3(y^*, z^*_{kl}, z^*_{lk}) = 0$, and therefore Eqs. (6.53) to (6.55) are fulfilled.

Case 2: If the conditions of Case 1 are not satisfied, i.e. Eq. (6.56), choose $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} L^{\leftrightarrow}_{kl} &\le \sum_{i\in C_k, j\in C_l} \frac{\gamma s_{ij} s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \\ \tilde{L}^{\rightarrow}_{kl} &\le \sum_{i\in C_k, j\in C_l} \frac{\alpha s_{ij}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \\ \tilde{L}^{\rightarrow}_{lk} &\le \sum_{i\in C_k, j\in C_l} \frac{\beta s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \end{aligned} \tag{6.66}$$

holds. Then it follows from Case 1 above that there exist $(y^*, z^*_{kl}, z^*_{lk}) \in [m^{-1}, m]^3$ that fulfill Eqs. (6.67) to (6.69)

$$L^{\leftrightarrow}_{kl} = \sum_{i\in C_k, j\in C_l} \frac{y^* \gamma s_{ij} s_{ji}}{1 + z^*_{kl}\alpha s_{ij} + z^*_{lk}\beta s_{ji} + y^*\gamma s_{ij} s_{ji}}, \tag{6.67}$$

$$\tilde{L}^{\rightarrow}_{kl} = \sum_{i\in C_k, j\in C_l} \frac{z^*_{kl} \alpha s_{ij}}{1 + z^*_{kl}\alpha s_{ij} + z^*_{lk}\beta s_{ji} + y^*\gamma s_{ij} s_{ji}}, \tag{6.68}$$

$$\tilde{L}^{\rightarrow}_{lk} = \sum_{i\in C_k, j\in C_l} \frac{z^*_{lk} \beta s_{ji}}{1 + z^*_{kl}\alpha s_{ij} + z^*_{lk}\beta s_{ji} + y^*\gamma s_{ij} s_{ji}}. \tag{6.69}$$

Hence, $z_{kl} = (z^*_{kl}\alpha)$, $z_{lk} = (z^*_{lk}\beta)$ and $y = (y^*\gamma)$ fulfill the original equations Eqs. (6.53) to (6.55).

It remains to prove, that for any $L^{\leftrightarrow}_{kl}, \tilde{L}^{\rightarrow}_{kl}, \tilde{L}^{\rightarrow}_{lk} \in \mathbb{R}_{>0}$ with $\left(L^{\leftrightarrow}_{kl} + \tilde{L}^{\rightarrow}_{kl} + \tilde{L}^{\rightarrow}_{lk}\right) \in (0, n_k n_l)$, there exist $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$, such that Eq. (6.66) holds. This problem can be simplifying by reformulating it several times. First, we replace the less or equal sign by an equal sign in the second and third equation of Eq. (6.66), i.e. we note that it suffices to show that (the stronger conditions)

$$\begin{aligned}
&\text{(i)} \quad L_{kl}^{\leftrightarrow} \leq \sum_{i \in C_k, j \in C_l} \frac{\gamma s_{ij} s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \\
&\text{(ii)} \quad \tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{\alpha s_{ij}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \\
&\text{(iii)} \quad \tilde{L}_{lk}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{\beta s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}},
\end{aligned} \tag{6.70}$$

hold. Next, we note that the sum of the right hand sides of (i) - (iii) in Eq. (6.70) equals $n_k n_l$ for arbitrary α, β, γ. Therefore, we get

$$\begin{aligned}
&\sum_{i \in C_k, j \in C_l} \frac{\gamma s_{ij} s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}} \\
&= \sum_{i \in C_k, j \in C_l} 1 - \frac{\alpha s_{ij} + \beta s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}} \\
&= n_k n_l - \left(\sum_{i \in C_k, j \in C_l} \frac{\alpha s_{ij}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}} \right) - \left(\sum_{i \in C_k, j \in C_l} \frac{\beta s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}} \right) \\
&= n_k n_l - \tilde{L}_{kl}^{\rightarrow} - \tilde{L}_{lk}^{\rightarrow}, \quad \text{if (ii) and (iii) hold} \\
&\geq L_{kl}^{\leftrightarrow}, \quad \text{since } L_{kl}^{\leftrightarrow} + \tilde{L}_{kl}^{\rightarrow} + \tilde{L}_{lk}^{\rightarrow} \leq n_k n_l.
\end{aligned} \tag{6.71}$$

Hence, if (ii) and (iii) are fulfilled, than (i) automatically holds as well. Thus, it suffices to show that for any $\tilde{L}_{kl}^{\rightarrow}, \tilde{L}_{lk}^{\rightarrow} \in \mathbb{R}_{>0}$, with $\tilde{L}_{kl}^{\rightarrow} + \tilde{L}_{lk}^{\rightarrow} < n_k n_l$, there exist $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ such that

$$\begin{aligned}
&\text{(ii)} \quad \tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{\alpha s_{ij}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}}, \\
&\text{(iii)} \quad \tilde{L}_{lk}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{\beta s_{ji}}{\alpha s_{ij} + \beta s_{ji} + \gamma s_{ij} s_{ji}},
\end{aligned} \tag{6.72}$$

hold. Canceling γ in the fractions of (ii) and (iii) in Eq. (6.72) and with $a := \alpha/\gamma$ and $b := \beta/\gamma$ gives

$$\begin{aligned}
&\text{(ii)} \quad \tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{a s_{ij}}{a s_{ij} + b s_{ji} + s_{ij} s_{ji}}, \\
&\text{(iii)} \quad \tilde{L}_{lk}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{b s_{ji}}{a s_{ij} + b s_{ji} + s_{ij} s_{ji}}.
\end{aligned} \tag{6.73}$$

Moreover, it suffices to show that (ii) and (ii)+(iii) in Eq. (6.73) are fulfilled, i.e. that for any $\tilde{L}_{kl}^{\rightarrow}, \tilde{L} \in \mathbb{R}_{>0}$, with $\tilde{L}_{kl}^{\rightarrow} < \tilde{L} < n_l n_k$, there exist $a, b \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} &\text{(ii)} \quad \tilde{L}_{kl}^{\rightarrow} = \sum_{i \in C_k, j \in C_l} \frac{a s_{ij}}{a s_{ij} + b s_{ji} + s_{ij} s_{ji}}, \\ &\text{(iii)}' \quad \tilde{L} \;\; = \sum_{i \in C_k, j \in C_l} \frac{a s_{ij} + b s_{ji}}{a s_{ij} + b s_{ji} + s_{ij} s_{ji}}, \end{aligned} \tag{6.74}$$

hold.

Consider (ii) from Eq. (6.74) as a function $f_{\tilde{L}_{kl}^{\rightarrow}}$ of a and b, i.e.

$$f_{\tilde{L}_{kl}^{\rightarrow}}(a, b) := \left(\sum_{i \in C_k, j \in C_l} \frac{a s_{ij}}{a s_{ij} + b s_{ji} + s_{ij} s_{ji}} \right) - \tilde{L}_{kl}^{\rightarrow}. \tag{6.75}$$

Note that $f_{\tilde{L}_{kl}^{\rightarrow}}$ is continuously differentiable in all variables, strictly monotonically increasing in a, and strictly monotonically decreasing in b. Furthermore, let $a \in \left[m^{-1}, m^2\right]$ and $b \in \left[m^{-1}, m\right]$, with $m \in \mathbb{R}_{>0}$ big enough. Now, for any b, we get

$$\lim_{m \to \infty} f_{\tilde{L}_{kl}^{\rightarrow}}\left(m^{-1}, b\right) = -\tilde{L}_{kl}^{\rightarrow} < 0, \tag{6.76}$$

and

$$\lim_{m \to \infty} f_{\tilde{L}_{kl}^{\rightarrow}}\left(m^{2}, b\right) = n_l n_k - \tilde{L}_{kl}^{\rightarrow} > 0. \tag{6.77}$$

Thus, it follows from the intermediate value theorem that for any $b_0 \in \left[m^{-1}, m\right]$ there exists $a_0 \in \left[m^{-1}, m^2\right]$ such that $f_{\tilde{L}_{kl}^{\rightarrow}}(a_0, b_0) = 0$. Together with $f_{\tilde{L}_{kl}^{\rightarrow}}$ being continuously differentiable and strictly monotonically increasing in a it follows from the implicit function theorem, that for every $(a_0, b_0) \in \left]m^{-1}, m^2\right[\times \left]m^{-1}, m\right[$ there exist an open neighborhood $V \subseteq \left]m^{-1}, m^2\right[$ of a_0, an open neighborhood $U \subseteq \left]m^{-1}, m\right[$ of b_0, and a unique and continuously differentiable function $g_0 : U \to V$ with $g(b_0) = a_0$ and such that for all $(a, b) \in V \times U$ it holds

$$f_{\tilde{L}_{kl}^{\rightarrow}}(a, b) = 0 \quad \Leftrightarrow \quad a = g_0(b). \tag{6.78}$$

Since there exists such a function g_0 for every $(a_0, b_0) \in \left]m^{-1}, m^2\right[\times \left]m^{-1}, m\right[$, this especially means that there exists a unique and continuously differentiable function $g : \left]m^{-1}, m\right[\to \left]m^{-1}, m^2\right[$ such that for all $(a, b) \in \left]m^{-1}, m\right[\times \left]m^{-1}, m^2\right[$

$$f_{\tilde{L}_{kl}^{\rightarrow}}(a, b) = 0 \quad \Leftrightarrow \quad a = g(b). \tag{6.79}$$

Moreover, from the definition of $f_{\tilde{L}_{kl}^{\rightarrow}}$, see Eq. (6.75), it is obvious that g is strictly monotonically increasing.

Next, we consider (iii)$'$ from Eq. (6.74) as a function $f_{\tilde{L}}$ of b and $g(b)$, i.e.

$$f_{\tilde{L}}(b) := \left(\sum_{i \in C_k, j \in C_l} \frac{g(b)\, s_{ij} + b s_{ji}}{g(b)\, s_{ij} + b s_{ji} + s_{ij} s_{ji}} \right) - \tilde{L}. \tag{6.80}$$

Note that $f_{\tilde{L}^{\rightarrow}}$ is continuously differentiable in all variables and strictly monotonically increasing in $g(b)$ and b. Furthermore, there exists $\epsilon \in \mathbb{R}_{>0}$ small enough, such that

$$\lim_{m \to \infty} f_{\tilde{L}}(m - \epsilon) = n_k n_l - \tilde{L} > 0, \tag{6.81}$$

and

$$\lim_{m \to \infty} f_{\tilde{L}}\left(m^{-1} + \epsilon\right) = \tilde{L}^{\rightarrow}_{kl} - \tilde{L} < 0. \tag{6.82}$$

Thus, it follows from the intermediate value theorem that there exists $b^* \in \left[m^{-1} + \epsilon, m - \epsilon\right]$, such that $f_{\tilde{L}}(b^*) = 0$. Moreover, since $f_{\tilde{L}^{\rightarrow}_{kl}}$ and $f_{\tilde{L}}$ are continuous and strictly monotone in all variables, $\left(g(b^*), b^*\right)$ constitutes the unique solution for (ii) and (iii)$'$ in Eq. (6.74).

Last, we note that the special case of $L^{\leftrightarrow}_{kl} = 0$, can be proved by the Bolzano–Poincaré–Miranda theorem, analogously to the interior case considered above. It follows that $r_{kl} = 0$ has to hold. This leaves us with simplified versions of Eqs. (6.51) and (6.52)

$$\tilde{L}^{\rightarrow}_{lk} = \sum_{i \in C_k, j \in C_l} \frac{z_{kl} s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z_{kl} s_i^{(\text{out})} s_j^{(\text{in})} + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.83}$$

$$\tilde{L}^{\rightarrow}_{lk} = \sum_{i \in C_k, j \in C_l} \frac{z_{lk} s_j^{(\text{out})} s_i^{(\text{in})}}{1 + z_{kl} s_i^{(\text{out})} s_j^{(\text{in})} + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})}}. \tag{6.84}$$

□

For completeness, we will briefly discuss the special case where $L^{\leftrightarrow}_{kl} = n_k n_l$, that is excluded in Theorem 6.2.1. Note that $L^{\leftrightarrow}_{kl} = n_k n_l$ implies that $\tilde{L}^{\rightarrow}_{kl} = 0$ and $\tilde{L}^{\rightarrow}_{lk} = 0$, and hence, all links in the adjacency matrices $A^{(\text{kl})}$ and $A^{(\text{lk})}$ are set to 1. In the special case of $L^{\rightarrow}_{kl} = 0$, it follows that $z_{kl} = 0$ and $L^{\leftrightarrow}_{kl} = 0$, and we simply have to find a z_{lk} such that

$$\tilde{L}^{\rightarrow}_{lk} = \sum_{i \in C_k, j \in C_l} \frac{z_{lk} s_i^{(\text{out})} s_j^{(\text{in})}}{1 + z_{lk} s_j^{(\text{out})} s_i^{(\text{in})}}, \tag{6.85}$$

is satisfied. The existence of a solution to Eq. (6.85) follows from the intermediate value theorem. In the special case of $L^{\rightarrow}_{kl} = n_k n_l$, it follows that $L^{\leftrightarrow}_{kl} = L^{\rightarrow}_{lk}$. All links in $A^{(\mathrm{kl})}$ are set to 1 and for the random matrix $A^{(\mathrm{lk})}$, we consider the simplified problem of identifying z_{lk}, such that

$$L^{\rightarrow}_{lk} = \sum_{i \in C_k, j \in C_l} \frac{z_{lk} s_i^{(\mathrm{out})} s_j^{(\mathrm{in})}}{1 + z_{lk} s_j^{(\mathrm{out})} s_i^{(\mathrm{in})}}, \tag{6.86}$$

holds. Again, the existence of a solution to Eq. (6.86) follows from the intermediate value theorem.

6.7 Conclusion and Outlook

Realistic models of inter-banking networks are necessary for an adequate and flexible assessment of systemic risk. Their construction, however, remains challenging, because of the very limited data availability. In this chapter we contribute to this research topic by presenting a block-structured model that reconstructs inter-banking networks across multiple countries. The advantages of our model are the following. First of all, our model allows to incorporate structural differences in financial networks across countries and offers great flexibility via the block-structure. The density and the reciprocity can be chosen separately for every block. Likewise the constraints on the weights, i.e. row sums, column sums, and block-weights can be set separately. This allows users, like central banks or policy-makers, who might have partial access to additional information to calibrate the model more accurately. Also, in case further information on aggregated level becomes available in the future, it can easily be incorporated in our model. As a trade-off on accuracy of network reconstruction and data availability, our model is calibrated on a small number of input factors, that are able to induce important network characteristics. As shown in Section 6.5.2 the model correctly reproduces known aggregated characteristics of financial networks like sparsity, positive reciprocity, disassortativity, and short paths using only a small set of input factors. Moreover, we show how block-density, block-reciprocity, and block-weights, which might not be available explicitly, can be approximated. Since the calibration of the model is non trivial, we also present an algorithm to handle this task.

7 Assessing Systemic Risk in the EU Interbank Market

Our network reconstruction model enables the application of various contagion mechanisms and systemic risk measures to realistic, international financial networks. To demonstrate this, we conduct a systemic risk analysis on the sample of reconstructed EU interbank networks, presented in Section 6.5. Like the previous chapter, this chapter is as well based on our manuscript Engel et al. (2019a).

The following results are computed by the 'FINEXUS Leverage Network Framework for Stress-testing' software of Gabriele Visentin, Marco D'Errico, and Stefano Battiston Battiston et al. (2016); Visentin et al. (2016), which integrates five of the most popular financial contagion models, namely:

- the 'clearing vector' (EN) by Eisenberg and Noe (2001),
- the 'extended clearing vector' (RV) incorporating default costs by Rogers and Veraart (2013),
- the 'default cascades model' (DC) by Battiston et al. (2012),
- the 'acyclic DebtRank' (aDR) by Battiston et al. (2012),
- and the 'cyclic DebtRank' (cDR) by Bardoscia et al. (2015).

Furthermore, following Battiston et al. (2016); Visentin et al. (2016) we differentiate between two risk dimensions. First, there is the risk that a bank under stress triggers waves of contagious losses throughout the entire system. Second, there is the risk that a bank is vulnerable to other banks in the network being under stress. More precisely, we distinguish:

- **Global Vulnerability** = relative loss in equity that a shock scenario causes to the entire system;
- **Individual Vulnerability** = relative loss in equity that a bank suffers from a shock scenario.

To keep individual vulnerability comparable across banks, in the following we consider instead the absolute loss in equity suffered by each bank.

There are many interesting questions regarding systemic risk that can now be analyzed. Here, we focus on the following four aspects. First, we give a general overview over the network fragility for various shock sizes and according to the different contagion models, see Section 7.1. Subsequently, Section 7.2 describes the correlation of node characteristics and systemic risk. In Section 7.3 we compare the official list of G-SIBs in the EU, provided by the Basel Committee on Banking Supervision, with our results. Last, in Section 7.4, we analyze the question how network density, an indicator of diversification in interbank lending, affects financial stability and compare our findings with the literature.

Throughout this section, a recovery rate of 40% is used for all banks.[1] The contagion model of Rogers and Veraart (RV) additionally considers a recovery rate for external assets, which is fixed to 50%, the default value of the 'FINEXUS Leverage Network Framework for Stress-testing' software. Moreover, 'first Round' effects refer to initial losses, caused solely by external shocks on the banks, disregarding propagation. 'Second Round' effects, computed by the different contagion models, report additional losses due to contagion, excluding first round losses.

7.1 Systemic Risk for Different Shock Sizes

We start the assessment of systemic risk by shocking all banks equally with various shock sizes and by propagating the shocks according to five different contagion models. Figures 7.1 and 7.2 present the resulting loss in equity and the fraction of defaulted banks. Comparing our results to those of Visentin et al. (2016), who analyze systemic risk in reconstructed networks consisting of the 50 largest EU banks, we find that our networks are more stable, but the structure of the curves is very similar. Interestingly, however, we observe a slightly different partial ordering for the global vulnerability H:

$$H^{\mathrm{EN}} \leq H^{\mathrm{RV}} \leq H^{\mathrm{DC}} \leq H^{\mathrm{aDR}} \leq H^{\mathrm{cDR}}. \tag{7.1}$$

In all simulations analyzed by Visentin et al. (2016), the authors find the following partial ordering:

$$H^{\mathrm{EN}} \leq H^{\mathrm{DC}} \leq H^{\mathrm{RV}} \leq H^{\mathrm{aDR}} \leq H^{\mathrm{cDR}}. \tag{7.2}$$

Whether the DC or the RV model yields higher global losses depends on the shock size and on the recovery rates. In both the DC and the RV model banks spread losses only in case of their default. Within the DC model, a defaulted bank triggers distress to its interbank creditors proportionally to the nominal liabilities and adjusted by the recovery rate, as soon as the losses suffered reach the bank's equity. In the RV model, in contrast, only the amount of a bank's obligations (external plus interbank liabilities) that exceeds its assets (external plus interbank assets) is spread proportionally to all creditors (external plus interbank creditors), adjusted by the recovery rate. Hence, if a bank suffers a shock equal to the size of its equity, in the DC model it will propagate

[1] See, for example, www.cdsmodel.com and Altman and Kishore (1996).

the maximum loss that it can spread, while in the RV model, it will simply absorb this shock and no propagation takes place. Moreover, in the RV model part of the distress flows out of the interbank network, since losses are also propagated to external creditors (via external liabilities). However, in the RV model a bank can suffer shocks higher than its equity. Therefore, with increasing shock sizes the losses spreading through the RV model will increase. The DC model, on the other hand, saturates at shock levels equal to the size of the banks' equity. Another important parameter that determines losses in the RV model is the recovery rate on external assets (which is set to 50% in all considered examples).[2]

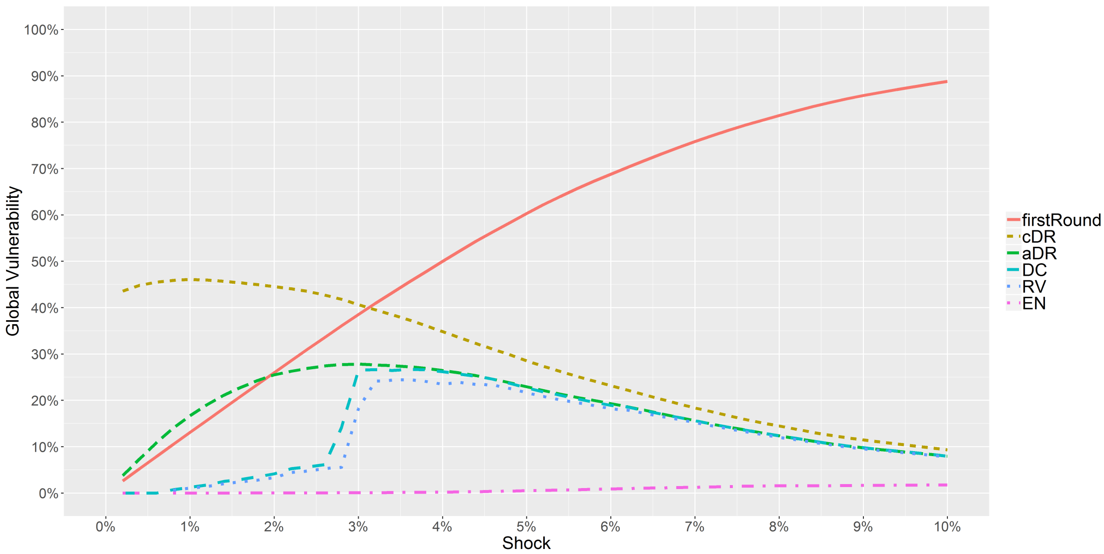


Figure 7.1: Global vulnerability caused by shocking all banks equally with various shock sizes and for different contagion models. Reported values are averages over 100 simulated networks (excluding the Rest-of-the-World node).

[2]For more details on the contagion models, see, e.g., Visentin et al. Visentin et al. (2016) and the references therein.

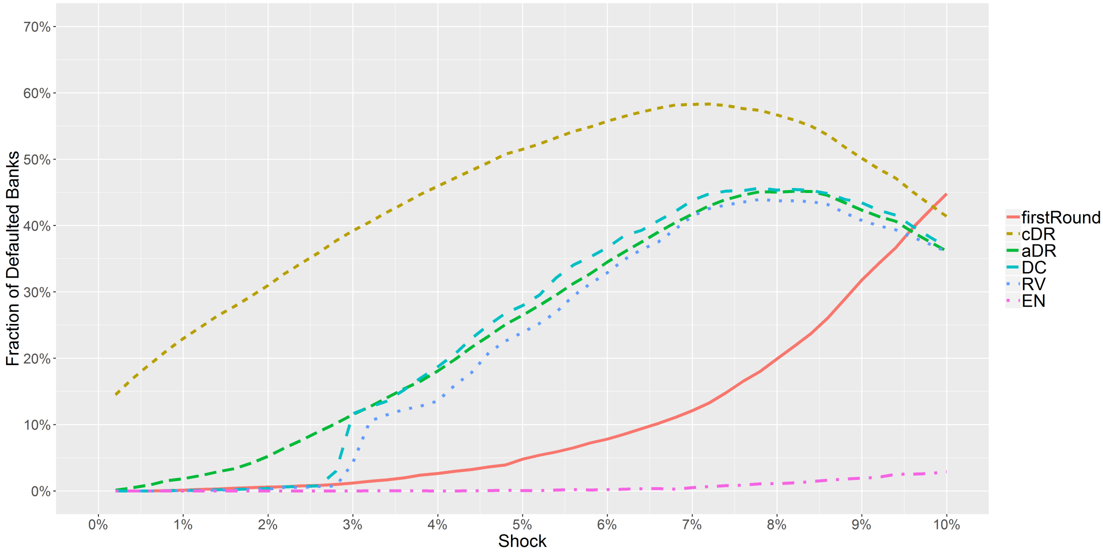


Figure 7.2: Fraction of defaulted banks caused by shocking all banks equally with various shock sizes and for different contagion models. Reported values are averages over 100 simulated networks (excluding the Rest-of-the-World node).

7.2 Correlation of Node Characteristics and Systemic Risk

In this paragraph we investigate:

1) Which network statistics make a node systemically important, in the sense that its default causes a severe shock on the entire banking network?

2) Which network statistics make a node vulnerable towards the default of other banks in the system?

We analyze both questions empirically in our set of reconstructed EU interbank networks, by letting one bank default at a time and computing the impact on the entire system as well as towards each of the other banks individually. As network statistics we consider the nodes' degree and strength and their centrality w.r.t. the (un-) directed and (un-) weighted network. In the framework of DC, aDR, and cDR a bank defaults as soon as the loss suffered reaches the value of its equity. At this point the defaulted bank triggers the maximum loss spread that it can cause (which equals the sum of its outstanding interbank liabilities multiplied by one minus its recovery rate). Even if the loss suffered exceeds the bank's equity level, the triggered loss spread does not increase further. Therefore, in the following, we initiate a bank's default by a shock on its external assets in the size of its equity. However, the EN and RV model assume that losses up to the size of a bank's equity can be absorbed by the bank and only losses exceeding the equity level are spread to the system. Hence, shocking a single bank by a loss equal to the size of its equity causes only the respective bank to default and no further losses occur. So, in this setting, global vulnerability in the EN and RV model equals simply the relative value of the defaulted bank's equity w.r.t. the total equity in the network. Analogously, individual vulnerability (i.e. losses suffered by a bank upon the default of another bank) equals zero in the EN and RV model.

Table 7.1 reports the rank correlation, measured by Kendall's tau, between different network statistics of the nodes and the relative loss in equity on the entire system caused by the nodes' default. Since shocks propagate backwards, from a node to its creditors, we can observe a high correlation between a node's number of incoming links as well as the weight carried by these links and global vulnerability. Regarding the centrality measures, interestingly closeness centrality seems to be the most relevant. Closeness centrality is defined as the inverse of the average distance from a node to the other nodes in the network. Hence, the shorter the paths between a bank and the other banks in the network, the higher its systemic impact, which is exactly what one would expect.

The correlation of nodes characteristics and individual vulnerability is presented in Table 7.2. Overall, we observe a high positive correlation between the number of outgoing links as well as their carried weight and individual vulnerability. Again, in most cases, closeness centrality turns out to be the most relevant centrality measure.

Table 7.1: Kendall's tau for different node characteristics and global vulnerability caused by the default of the respective node. Values of the EN and RV model essentially report the correlation with the banks' equity, since in the considered shock setting no propagation is triggered. All values are statistically highly significant with p-values smaller than 1e-200.

	EN	RV	DC	aDR	cDR
in-degree	0.57	0.57	0.65	0.63	0.68
in-strength	0.61	0.61	0.72	0.84	0.79
closeness undirected unweighted	0.61	0.61	0.67	0.72	0.62
in-closeness unweighted	0.58	0.58	0.67	0.75	0.72
closeness undirected weighted	0.67	0.67	0.75	0.79	0.68
in-closeness weighted	0.58	0.58	0.69	0.85	0.77
betweenness undirected unweighted	0.56	0.56	0.59	0.52	0.51
betweenness directed unweighted	0.55	0.55	0.58	0.53	0.54
betweenness undirected weighted	0.52	0.52	0.51	0.50	0.50
betweenness directed weighted	0.48	0.48	0.54	0.54	0.49
eigenvector centrality undirected unweighted	0.60	0.60	0.65	0.69	0.55
eigenvector centrality undirected weighted	0.45	0.45	0.51	0.61	0.45

Table 7.2: Kendall's tau for different node characteristics and individual vulnerability suffered by the respective node and caused by the default of other nodes. Values for the EN and RV model are not available, since in the considered shock setting no propagation is triggered. All values are statistically highly significant with p-values smaller than 1e-14.

	EN	RV	DC	aDR	cDR
out-degree	NA	NA	0.68	0.63	0.57
out-strength	NA	NA	0.95	0.89	0.82
closeness undirected unweighted	NA	NA	0.55	0.58	0.60
out-closeness unweighted	NA	NA	0.72	0.74	0.76
closeness undirected weighted	NA	NA	0.62	0.65	0.65
out-closeness weighted	NA	NA	0.85	0.83	0.78
betweenness undirected unweighted	NA	NA	0.59	0.56	0.53
betweenness directed unweighted	NA	NA	0.61	0.58	0.55
betweenness undirected weighted	NA	NA	0.50	0.51	0.51
betweenness directed weighted	NA	NA	0.55	0.55	0.55
eigenvector centrality undirected unweighted	NA	NA	0.46	0.49	0.49
eigenvector centrality undirected weighted	NA	NA	0.34	0.38	0.40

7.3 Global Systemically Important Banks (G-SIBs)

The official methodology for identifying global systemically important banks (G-SIBs) was developed by the Basel Committee on Banking Supervision (BCBS). It is essentially a weighted sum over a number of normalized financial positions. The BCBS states: *"The Committee is of the view that global systemic importance should be measured in terms of the impact that a bank's failure can have on the global financial system and wider economy, rather than the risk that a failure could occur. This can be thought of as a global, system-wide, loss-given-default (LGD) concept rather than a probability of default (PD) concept."*, (see https://www.bis.org/publ/bcbs255.pdf).

As our model allows to directly simulate the failure of a single bank and, hence, to compute the LGD in the interbank market, naturally the question arises whether the ranking in terms of global vulnerability is aligned to the one derived by the BCBS methodology. Differences in the two approaches that should be kept in mind are listed in Table 7.3.

Figure 7.3 compares the scores of the BCBS against global vulnerability caused by the single default of each bank. The default of a bank is again initiated by a shock on

Table 7.3: Differences in the assessment of G-SIBs.

	BCBS	Model presented in Chapter 6
coverage	worldwide	EU
aggregation level of banks	consolidated	unconsolidated
LGD w.r.t.	global economy	interbank market

external assets in the size of the bank's equity. Hence, in the EN and RV model global vulnerability equals the relative value of the bank's equity w.r.t. the total equity in the network. We observe that the ranking of the banks differs substantially across the contagion models. A low but at the level of 5% significant rank correlation between the BCBS methodology and global vulnerability can only be identified for the EN and the RV model, see Table 7.4. Thus, the BCBS ranking is more aligned with the equity of the banks, than with the systemic risk computed by the DC, aDR, and cDR in our reconstructed EU interbank networks.

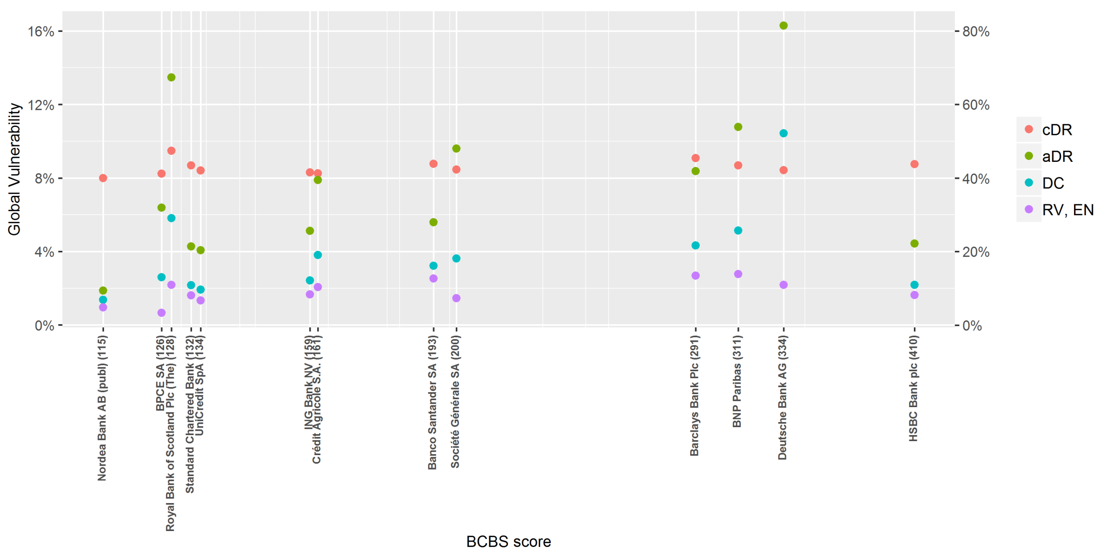


Figure 7.3: Comparison of G-SIBs in the EU as classified by the BCBS' methodology (scores on the x-axis in log scale) and global vulnerability (on the y-axis). The secondary y-axis (right) refers to global vulnerability computed by the cDR. Global vulnerability values are averages over 100 simulated networks (excluding the Rest-of-the-World node). Note that the y-axis denotes the relative loss in equity while the scale of the x-axis is difficult to interpret.

Table 7.4: Kendall's tau of the ranking of the G-SIBs in the EU between the BCBS' classification and global vulnerability. Significance at the level of 5% is marked by '*'.

	EN, RV	DC	aDR	cDR
Kendall's tau	0.46*	0.41	0.38	0.28
p-value	0.03	0.06	0.08	0.20

In addition to the computation of a LGD, our model also allows for a detailed analysis of the consequences of one of the G-SIBs being under stress. For example, we can simulate a shock of arbitrary size to a G-SIB and compute the resulting network-wide loss. Figure 7.4 illustrates these shock scenarios for the Deutsche Bank AG and w.r.t. all five contagion models.[3] The point on the x-axis at which a function turns into a constant marks the shock size at which the bank defaults, i.e. the point at which the absolute value of the shock equals the bank's equity. In the DC, aDR, and cDR this triggers the maximum loss spread that a bank can propagate (which equals the total amount of its outstanding interbank liabilities adjusted by the recovery rate). We observe that the amount of total losses depends heavily on the chosen contagion model.

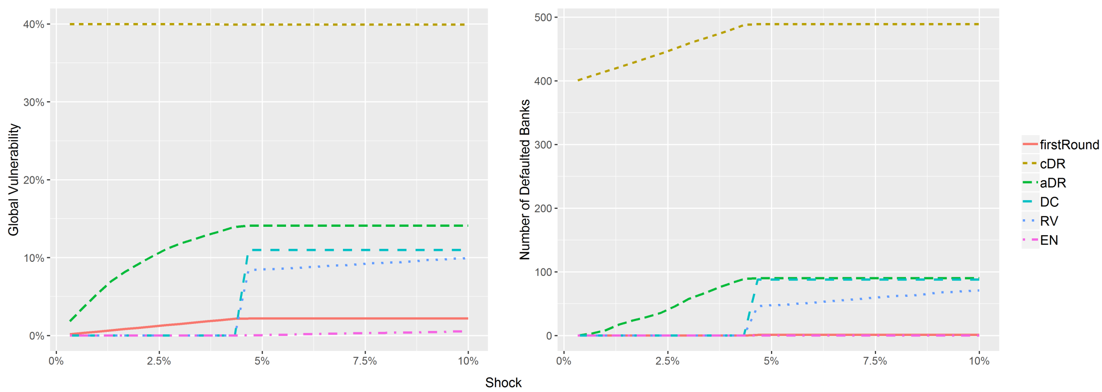


Figure 7.4: Global vulnerability (left) and number of defaults (right) caused by shocking the Deutsche Bank AG with shock sizes ranging from 1% to 10% of external assets, computed by the contagion models: EN, RV, DC, aDR, and cDR. Reported values are averages over 100 simulated networks (excluding the Rest-of-the-World node).

In summary, our model enables a detailed systemic risk analysis. The BCBS's methodology, on the other hand, allows for a fast calculation and incorporates indicators covering

[3]Equivalent plots for the other G-SIBs in the EU are included in Fig. 7.6.

the global economy. Table 7.5 gives an overview over advantages and disadvantages of both methodologies.

Table 7.5: Advantages and disadvantages of the BCBS's methodology and our model

	BCBS methodology	Model presented in Chapter 6
Advantages	• fast calculation • covers the global economy	• resulting score has a monetary interpretation • analysis of arbitrary shock sizes (< default) is possible • analysis of simultaneously shocking several banks is possible • distribution of shock impact can be estimated • new regulations can be tested • impact of network statistics on systemic risk can be analyzed
Disadvantages	• scale of scores has no interpretation	• some input parameters are not known explicitly, i.e. need to be estimated • data availability limits the scope of the model, i.e. determines which countries can be included • systemic risk values are not deterministic across different simulated networks, but seem to have a very low standard deviation • covers only the EU interbank market

7.4 Network Density and Stability

Another interesting question is how the network density, an indicator of diversification in interbank lending, influences financial stability. This aspect has already been analyzed in several papers. Recently, Roncoroni et al. (2018) confirmed earlier results of Acemoglu et al. (2015), stating that densely connected networks are more stable regarding small shocks, but more fragile regarding big shocks. Acemoglu et al. (2015) base their analysis on artificial networks and focus mostly on regular networks (= total claims and liabilities of all banks are equal). Roncoroni et al. (2018), on the other hand, analyze a unique dataset of the European Central Bank, consisting of 26 large EU banks.

Are these results confirmed within our sample of reconstructed EU interbank networks comprising 3,468 banks?

To answer this question, we construct a second network ensemble, where the density of each block was increased to $d^{(\text{new})} = d^{(\text{old})} + (1 - d^{(\text{old})})/2$. The density matrix is presented in form of a heatmap in Fig. 7.5.

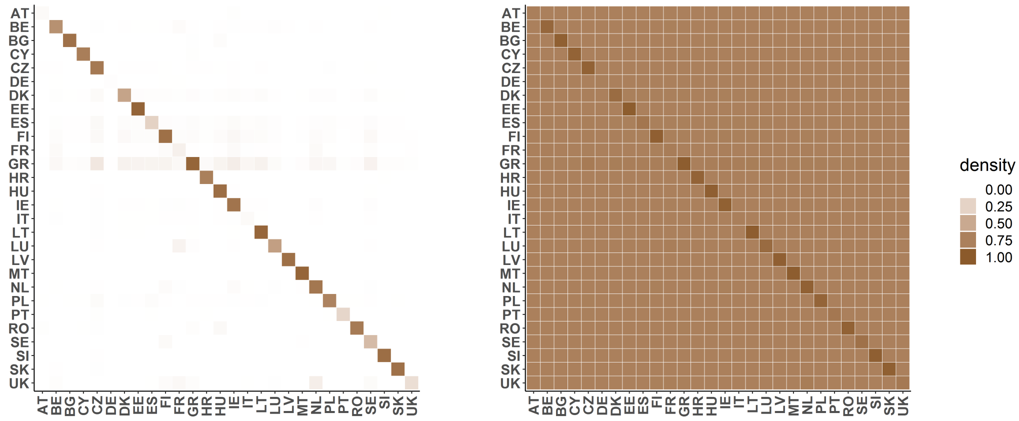
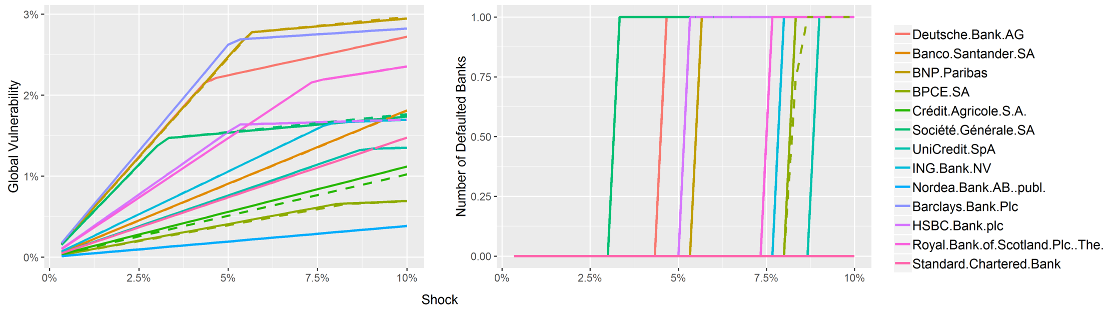
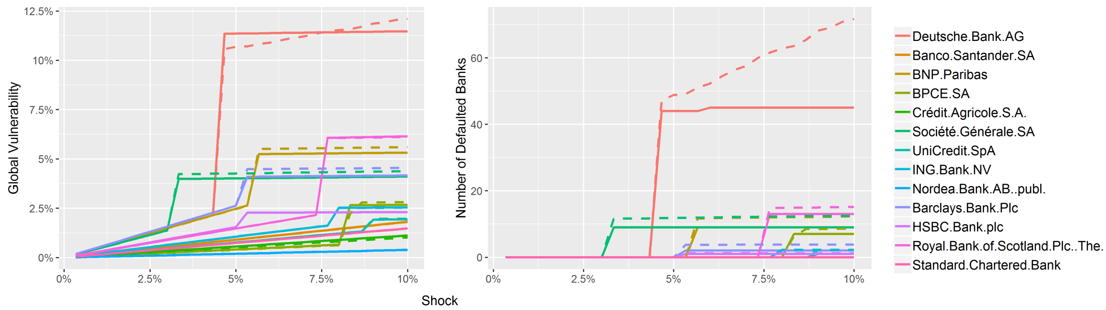


Figure 7.5: Heatmap of original densities (left) and increased densities (right).

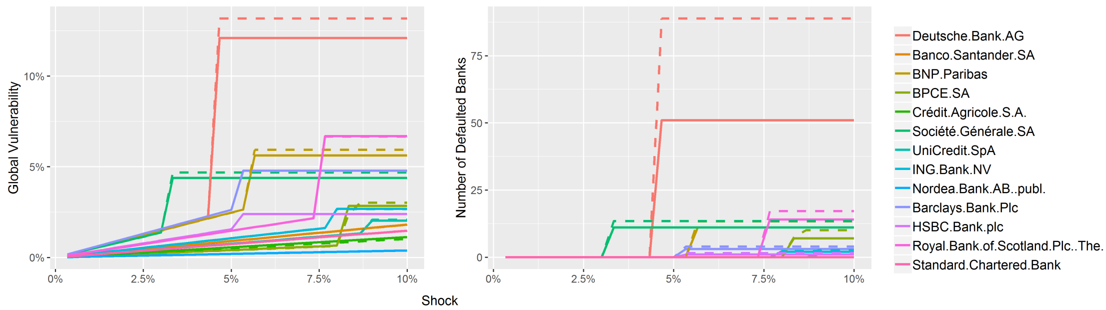
First, we analyze the effect of the increased network density on the G-SIBs. Figure 7.6 presents the results of shocking the G-SIBs separately in the more densely connected networks in comparison to the originally sparse networks, computed by the EN, RV, DC, aDR, and cDR contagion model. In almost all considered cases, global vulnerability and the number of defaulting banks is smaller in the densely connected networks. The magnitude of the difference depends heavily on the applied contagion model.


Global Vulnerability
3%
2%
1%
0%
0%
2.5%
5%
7.5%
10%
Number of Defaulted Banks
1.00
0.75
0.50
0.25
0.00
Shock
Deutsche.Bank.AG
Banco.Santander.SA
BNP.Paribas
BPCE.SA
Crédit.Agricole.S.A.
Société.Générale.SA
UniCredit.SpA
ING.Bank.NV
Nordea.Bank.AB..publ.
Barclays.Bank.Plc
HSBC.Bank.plc
Royal.Bank.of.Scotland.Plc..The.
Standard.Chartered.Bank

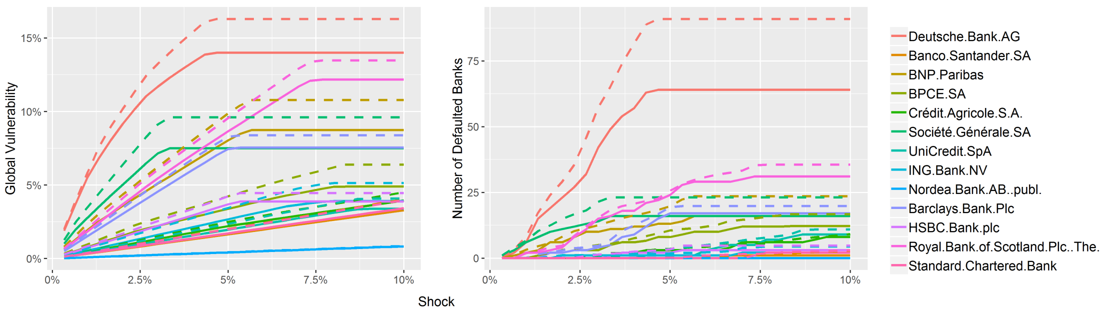
(a) EN


Global Vulnerability
12.5%
10%
7.5%
5%
2.5%
0%
0%
2.5%
5%
7.5%
10%
Number of Defaulted Banks
60
40
20
0
Shock
Deutsche.Bank.AG
Banco.Santander.SA
BNP.Paribas
BPCE.SA
Crédit.Agricole.S.A.
Société.Générale.SA
UniCredit.SpA
ING.Bank.NV
Nordea.Bank.AB..publ.
Barclays.Bank.Plc
HSBC.Bank.plc
Royal.Bank.of.Scotland.Plc..The.
Standard.Chartered.Bank

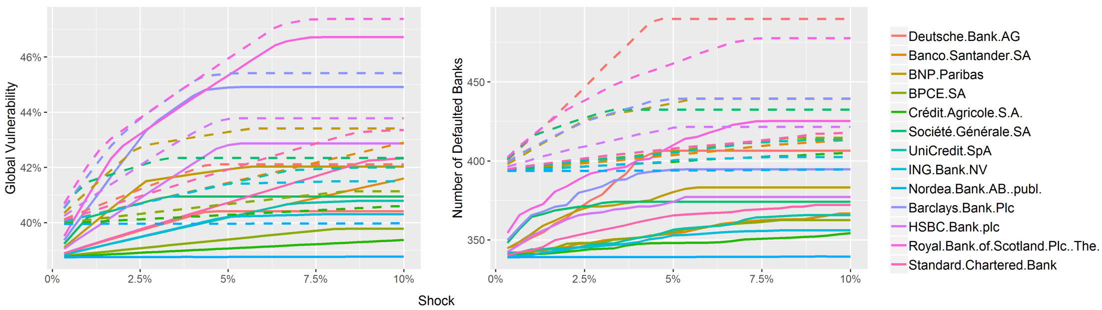
(b) RV


Global Vulnerability
10%
5%
0%
0%
2.5%
5%
7.5%
10%
Number of Defaulted Banks
75
50
25
0
Shock
Deutsche.Bank.AG
Banco.Santander.SA
BNP.Paribas
BPCE.SA
Crédit.Agricole.S.A.
Société.Générale.SA
UniCredit.SpA
ING.Bank.NV
Nordea.Bank.AB..publ.
Barclays.Bank.Plc
HSBC.Bank.plc
Royal.Bank.of.Scotland.Plc..The.
Standard.Chartered.Bank

(c) DC



(d) aDR



(e) cDR

Figure 7.6: Global vulnerability (left) and number of defaults (right) caused by shocking the G-SIBs separately with shock sizes ranging from 1% to 10% of external assets. Values of the networks with the original densities are pictured in dashed lines. Reported values are averages over 100 simulated networks (excluding the Rest-of-the-World node). Legends are ordered according to the respective global vulnerability at a shock size of 10%.

Next, we analyze shock scenarios where all banks are shocked equally with various shock sizes on external assets. The results are presented in Fig. 7.7 in comparison to the results of the same shock scenarios applied to the originally sparse networks. Interestingly, for the considered shock scenarios, global vulnerability does barley differ between the sparse and densly connected networks. However, we can observe an increase in the number of defaulting banks for networks with higher densities. This means that losses are distributed differently in both network sets. In the more densely connected networks contagion seems to flow across a bigger number of small nodes that at some point default and propagate their losses back to big nodes.

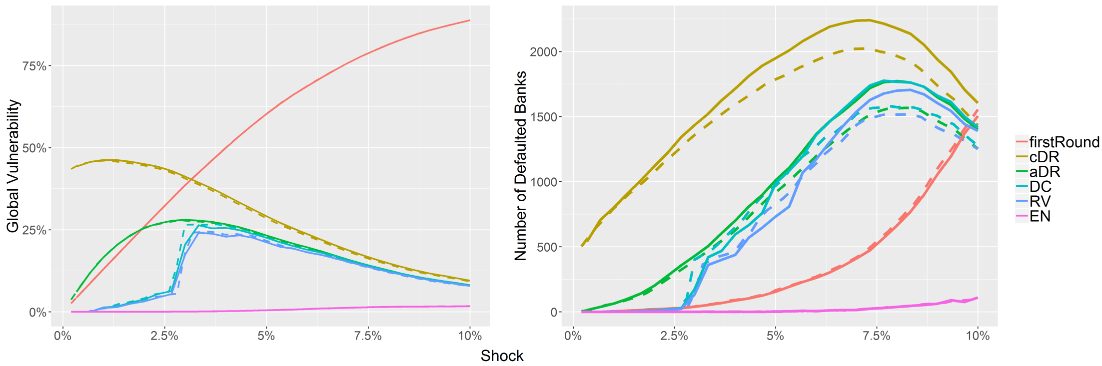


Figure 7.7: Global vulnerability (left) and number of defaults (right) caused by shocking all banks equally with shock sizes ranging from 1% to 10% of external assets. Values of the networks with the original densities are pictured in dashed lines. Reported values are averages over 100 simulated networks (excluding the Rest-of-the-World node).

7.5 Conclusion and Outlook

The simulated EU interbank networks, derived in Section 6.5.2, enable the application of a battery of contagion mechanisms and systemic risk measures. To demonstrate the potential of the model, developed in Chapter 6, this chapter conducts a systemic risk analysis on the reconstructed European interbank market. The results highlight the differences in systemic risk measures along five of the most prominent contagion models. Furthermore, the correlation between node characteristics and systemic risk caused by a bank's default as well the vulnerability suffered from the default of other banks is analyzed. We find that the loss that a bank's default causes on the interbank network is highly correlated with the number of its creditors (in-degree) as well as the amount borrowed (in-strength). Likewise, the vulnerability of a bank is highly correlated with the number of its debtors (out-degree) and the amount lend (out-strength). Among centrality measures, closeness centrality turns out to be the most significant for indicating systemic importance. In addition, the results on systemic risk are compared with the BCBS's ranking of global systemically important banks, which again turns out to depend heavily on the applied contagion model. Lastly, we can confirm earlier conclusions on the effect of the network density on systemic risk. Densely connected networks are more stable regarding small shocks and more fragile regarding big shocks when measured by the number of defaulting banks. The total loss in equity caused by big shocks, on the other hand, is not affected by the density of the network. These outcomes shed new light on systemic risk and its monitoring and can support policy-makers in their aim to stabilize the EU interbank market.

8 Conclusion

The aim of this thesis was to derive tractable, realistic, and flexible models for our highly interconnected banking systems. The relevance of such models was dramatically emphasized by the last global financial crisis of 2007–2008, as the lack of knowledge on the interdependencies between financial institutions lead to a further amplification of the crisis, see Basel Committee on Banking Supervision (2010).

To support the ongoing efforts on stabilizing and securing the financial system, this thesis provides the following contributions.

First, the most prominent network reconstruction techniques, namely ERGMs and fitness models, are analyzed and extended. A main contribution is the derivation of necessary and sufficient conditions under which a solution to the weighted and directed ERGM, satisfying given row and column sums, exists, see Theorem 3.1.1. We show that two quite intuitive conditions on the desired row and column sums suffice to ensure the existence of a solution, which is, moreover, unique up to certain equivalence classes. This is an important result, as it ensures a wide applicability of the model in the present context, based on a sound mathematical foundation. In addition, the proof reveals a new and efficient algorithm for calibrating the parameters. The question of the existence of a solution to other classes of ERGMs still remains an open research topic. Solving these questions could be of interest as this might lead to new and more efficient techniques for parameter calibration, which is sometimes difficult regarding the reconstruction of very big networks.
Furthermore, the randomized fitness model is conveniently extended towards more flexible degree distributions, and thus, enables a more precise calibration to real-world networks.
Empirical fitness models have been found very useful in the light of scarce available information. However, the underlying mathematical structure has not yet been analyzed. We provide a first step towards a better understanding of these models by analytically deriving the degree distribution of the networks, generated by fitness variables which follow a power law with exponent $\alpha = 1$. The resulting degree distributions induced by other exponents would provide further insight into these models, however, the analytical analysis seems considerably more complex and is left for further research. Based on our studies, the reason why empirical fitness models work that well seems to be that the usual power law distribution of the empirical fitness variables likewise induces a power law distribution for the degrees of the generated networks.

Besides the theoretical contributions, this thesis provides important practical results as well. Building on ERGMs, we demonstrate how domestic interbank networks can be realistically reconstructed using only scarce publicly available data. Comparing the sample of reconstructed networks with true network statistics of the German and Italian interbank market leads to the conclusion that many financial networks characteristics constitute a natural consequence of the heterogeneous in- and out-degree distributions and the reciprocity. These findings are especially of relevance, since data on financial interconnections are mostly not available.

The proposed model for domestic and unweighted interbank networks is subsequently extended to cover multiple countries in a weighted and directed graph. The developed block structured model presents another main contribution of this thesis. This model combines the advantages of fitness models and ERGMs and is build on our previously established results. To the best of our knowledge, this model currently constitutes the most detailed, flexible, and analytically tractable approach to reconstruct international financial networks. The model is split in two parts. First, the network topology is reconstructed via a new extended fitness model. In addition, we show that a solution to the extended fitness model is guaranteed to exist under very general conditions. Moreover, the solution is unique w.r.t. the generated expected in- and out-degree sequence and the number of reciprocal links. Weights are allocated to the sampled adjacency matrices in a second step based on a new ERGM. Since the parameter calibration of the ERGM requires to solve a high dimensional and complex system of equations, we also provide a fast and efficient algorithm for this purpose. The potential of the proposed model is demonstrated by reconstructing the EU interbank network, comprising 3,469 banks.

The reconstructed networks finally enable a detailed assessment of systemic risk. Applying five of the most prominent contagion models, we provide a systemic risk analysis on the EU interbank market. We find that certain node based network statistics are highly correlated with the systemic risk that a financial institution bears, as well as its vulnerability. Furthermore, we analysis the GSIBs, which are of special interest, in more detail. An intriguing result is that the ranking of the GSIBs in terms of systemic risk, computed via different contagion models, is not aligned with the official ranking of the BCBS. Investigating the reasons of the discrepancy would be interesting, but is outside the scope of this thesis, and left for further research. Last, we can confirm previous results on the ambiguity of the influence of the network density on systemic risk.

This thesis sheds valuable light on the complex nature of our highly interconnected banking systems and its potential risks. We hope that the derived contributions help the relevant authorities in their efforts to stabilize and secure our financial system.

A Network Statistics

For the sake of completeness and the readers' convenience, we state the definitions of the network statistics used in this thesis. Throughout this section we use the following notation:

- n denotes the number of nodes of the considered graph;
- $a \in \{0,1\}^{n \times n}$ denotes the adjacency matrix of a given graph without self-loops, i.e. $a_{ii} = 0$ for all $i = 1, \dots, n$;
- $d = (d_1, \dots, d_n)$ denotes the degree sequence of a, i.e. $d_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} (a_{ij} + a_{ji})$ for directed graphs and $d_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}$ for undirected graphs.

Assortativity: The degree assortativity (resp. in-degree assortativity and out-degree assortativity) is defined as Pearson's correlation coefficient of the degrees (resp. in-degrees and out-degrees) of connected nodes. Negative assortativity indicates that banks with small degree (resp. small in-degree and small out-degree) tend to connect to banks with large degree (resp. large in-degree and large out-degree) and vice versa. Positive assortativity indicates that banks tend to connect to banks with a similar degree (resp. similar in-degree and similar out-degree).

Average nearest neighbor degree: The average nearest neighbor degree $k_{nn}(k)$, also called average degree connectivity, is defined as the average over the degree of the neighbors of nodes with degree k. Let $\mathcal{I}(k) = \{i = 1, \dots, n : d_i = k\}$ denote the set of nodes with degree k and $\mathcal{N}(i) = \{j = 1, \dots, n : a_{ij} = 1 \vee a_{ji} = 1\}$ the neighbors of node i.

$$k_{nn}(k) = \frac{1}{|\mathcal{I}(k)|} \sum_{i \in \mathcal{I}(k)} \frac{1}{k} \sum_{j \in \mathcal{N}(i)} d_j.$$

Betweenness centrality: We rely on the definition of Roukny et al. (2014). Betweenness is a measure of centrality and is based on the number of shortest paths in the network that pass through a node. For a node i, its betweenness is defined as

$$b_i = \sum_{j} \sum_{h \neq j} s_{jh}(i)/s_{jh},$$

where $s_{jh}(i)$ is the number of shortest paths between node j and h that pass through node i, and s_{jh} is the total number of shortest paths between node j and h.

Closeness centrality: We rely to the definition of Roukny et al. (2014). Closeness is a measure of centrality and is defined as the inverse of the average shortest paths between a node i and all other nodes, i.e.,

$$c_i = (n-1)/\sum_j d_{ij}$$

where d_{ij} denotes the length of the shortest path between node i and j.

Clustering coefficient: The clustering coefficient measures the fraction of triangles in a network. It can be interpreted as the tendency of a node's neighbors to also be connected to each other. The clustering coefficient C_i of a node i is defined as

$$C_i\left(a\right) = \frac{\sum_{j\neq i}\sum_{h\notin\{i,j\}} a_{ij}a_{ih}a_{jh}}{d_i\left(d_i-1\right)} = \frac{\left(a^3\right)_{ii}}{d_i\left(d_i-1\right)}.$$

For more information on the clustering coefficient and its generalized versions for directed (and undirected) weighted graphs, see Fagiolo (2007).

Eigenvector centrality: We rely on the definition of Roukny et al. Roukny et al. (2014). Eigenvector centrality is a measure of the influence of a node in the network. Let λ denote the largest eigenvalue of the adjacency matrix a, and e its corresponding eigenvector. So $\lambda e = ae$ holds. Now the eigenvector centrality of node i is defined as

$$e_i = \sum_j a_{ij}e_j/\lambda.$$

Herndhal Hirschman Index (HHI): We rely on the definition of Roukny et al. (2014). The HHI index is defined as the sum of squared shares of links of all nodes, i.e.,

$$HHI = \sum_{i=1}^{n} s_i^2,$$

where $s_i = k_i/\sum_i k_i$ is the share of node i and k_i its degree. The HHI is a measure of concentration and lies in $[1/n, 1]$. The higher the value of the HHI, the higher is the number of links that is concentrated on few banks acting as hubs. The inverse of the HHI gives a proxy of the number of leading actors in the market.

Largest strong component: The LSC denotes the number of nodes in the largest subset of the network, such that there is a directed path from each node to every other node in the subset.

Largest weak component: The LWC denotes the number of nodes in the largest subset of the network, such that there is an undirected path from each node to every other node in the subset.

Reciprocity: The (degree) reciprocity (introduced by Garlaschelli and Loffredo (2004b)) is defined as Pearson's correlation coefficient of the adjacency matrix a and its transpose $a^{\top}$, i.e.,

$$\rho_r = \sum_{i \neq j} \left(a_{ij} - \bar{a}\right)\left(a_{ji} - \bar{a}\right) / \sum_{i \neq j} \left(a_{ij} - \bar{a}\right)^2,$$

where $\bar{x}$ denotes the average value of the entries of x, i.e., the network density. Reciprocity can also be expressed in terms of the total number of edges $L^{\rightarrow} = \sum_{i \neq j} a_{ij}$ and the number of mutual edges $L^{\leftrightarrow} = \sum_{i \neq j} a_{ij} a_{ji}$,

$$\rho_r = \frac{L^{\leftrightarrow}/L^{\rightarrow} - \bar{a}}{1 - \bar{a}}.$$

B Classes of ERGMs

Table B.1 summarizes well kown classes of ERGMs, that have been proposed and studied in the literature. Although, the corresponding references are listed, we note that these are often not as precise, and Table B.1 actually provides additional information.

Throughout this thesis, unless explicitly stated otherwise, the following notation and setting is used:

- Unweighted random graphs with n nodes are denoted by $A \in \{0,1\}^{n \times n}$. Possible realizations of A are denoted by a.
- We exclude self-loops, i.e. $a_{ii} = 0$ for all $i = 1, \dots, n$ and for all considered realizations a. This implies $P\left(A_{ii} = 0\right) = 1$. However, it is straightforward to adapt ERGMs to include self-loops.
- We focus on directed graphs. However, in all cases where the network statistics allow an equivalent application to undirected graphs, the corresponding ERGMs can be derived likewise.
- Weighted random graphs with n nodes are denoted by $W \in \mathbb{N}_0^{n \times n}$. Possible realizations of W are denoted by w.

Table B.1: Overview over various instances of ERGMs.

ERGM with desired density (Newman (2010); Park and Newman (2004))[1]	
Considered set of graphs $\mathcal{G}$	$\mathcal{G} = \left\{a \in \{0,1\}^{n\times n} : a_{11} = \cdots = a_{nn} = 0\right\}$
Network statistic f	$f(a) = \sum_{i\neq j} a_{ij}$, counts the number of edges in a
Desired density	$\frac{L^{\rightarrow}}{n(n-1)} \in (0,1)$, i.e. $L^{\rightarrow} \in (0, n(n-1))$, denotes the desired number of edges
Hamiltonian H	$H(a) = \lambda \sum_{i\neq j} a_{ij}$
Partition function Z	$Z = \left[1 + e^{-\lambda}\right]^{n(n-1)}$
Calibration	$\lambda = \log\left(\frac{n(n-1)}{L^{\rightarrow}} - 1\right)$
Link probabilities	$P(A_{ij} = 1) = \frac{L^{\rightarrow}}{n(n-1)}, \quad \forall i \neq j \wedge i,j \in \{1,\ldots,n\}$
Graph probabilities	$P(A = a) = \prod_{i\neq j}\left(\frac{L^{\rightarrow}}{n(n-1)}\right)^{a_{ij}}\left(1 - \frac{L^{\rightarrow}}{n(n-1)}\right)^{1-a_{ij}}$

ERGM with desired degree sequence (Newman (2010); Park and Newman (2004))[1]	
Considered set of graphs $\mathcal{G}$	$\mathcal{G} = \left\{a \in \{0,1\}^{n\times n} : a_{11} = \cdots = a_{nn} = 0\right\}$
Network statistics k_i	$k_i(a) = \sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij} + a_{ji}, \quad \forall i \in \{1,\ldots,n\}$, counts the number of edges of node i

[1] The references consider undirected graphs, while we consider directed graphs. However, computations work analogously.

Desired degrees	$d_i \in (0, 2(n-1))$, denotes the desired number of edges of node i
Partition function Z	$Z = \prod_{i \neq j} \left[1 + e^{-\lambda_i - \lambda_j}\right]$
Link probabilities	$P(A_{ij} = 1) = \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}, \quad \forall i \neq j \wedge i, j \in \{1, \ldots, n\}$
Graph probabilities	$P(A = a) = \prod_{i \neq j} \left[P(A_{ij} = 1)\right]^{a_{ij}} \left[1 - P(A_{ij} = 1)\right]^{1 - a_{ij}}$

ERGM with desired in- and out-degree sequence (Newman (2010); Park and Newman (2004))

Considered set of graphs $\mathcal{G}$	$\mathcal{G} = \left\{a \in \{0,1\}^{n \times n} : a_{11} = \cdots = a_{nn} = 0\right\}$
Network statistics $k_j^{(\text{in})}$	$k_j^{(\text{in})}(a) = \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{ij}, \quad \forall j \in \{1, \ldots, n\}$, counts the number of incoming edges of node j
Network statistics $k_i^{(\text{out})}$	$k_i^{(\text{in})}(a) = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}, \quad \forall i \in \{1, \ldots, n\}$, counts the number of outgoing edges of node i
Desired in-degrees	$d_i^{(\text{in})} \in (0, n-1)$, denotes the desired number of incoming edges of node i
Desired out-degrees	$d_i^{(\text{out})} \in (0, n-1)$, denotes the desired number of outgoing edges of node i
Hamiltonian H	$H(a) = \sum_{i \neq j} \left(\lambda_i^{(\text{out})} + \lambda_j^{(\text{in})}\right) a_{ij}$
Partition function Z	$Z = \prod_{i \neq j} \left[1 + e^{-\lambda_i^{(\text{out})} - \lambda_j^{(\text{in})}}\right]$
Link probabilities	$P(A_{ij} = 1) = \frac{e^{-\lambda_i^{(\text{out})} - \lambda_j^{(\text{in})}}}{1 + e^{-\lambda_i^{(\text{out})} - \lambda_j^{(\text{in})}}}, \quad \forall i \neq j \wedge i, j \in \{1, \ldots, n\}$

Graph probabilities	$P\left(A=a\right)=\prod_{i\neq j}\left[P\left(A_{ij}=1\right)\right]^{a_{ij}}\left[1-P\left(A_{ij}=1\right)\right]^{1-a_{ij}}$

ERGM with desired density and reciprocity (Newman (2010))	
Considered set of graphs $\mathcal{G}$	$\mathcal{G}=\left\{a\in\{0,1\}^{n\times n}:a_{11}=\cdots=a_{nn}=0\right\}$
Network statistic f	$f\left(a\right)=\sum_{i\neq j}a_{ij}$, counts the number of edges in a
Network statistic r	$r\left(a\right)=\sum_{i\neq j}a_{ij}a_{ji}$, counts the number of reciprocated edges in a
Desired density	indirectly given by the desired number of edges $L^{\rightarrow}\in\left(0,n\left(n-1\right)\right)$
Desired reciprocity [2]	indirectly given by the desired number of reciprocated edges $L^{\leftrightarrow}\in\left(\max\left\{0,2L^{\rightarrow}-n\left(n-1\right)\right\},L^{\rightarrow}\right)$
Hamiltonian H	$H\left(a\right)=\sum_{i<j}2\lambda_{L^{\leftrightarrow}}a_{ij}a_{ji}+\lambda_{L^{\rightarrow}}\left(a_{ij}+a_{ji}\right)$
Partition function Z	$Z=\left[1+2e^{-\lambda_{L^{\rightarrow}}}+e^{-2\lambda_{L^{\leftrightarrow}}-2\lambda_{L^{\rightarrow}}}\right]^{\binom{n}{2}}$
Calibration	$\lambda_{L^{\rightarrow}}=\log\left(\frac{n(n-1)-L^{\leftrightarrow}}{L^{\rightarrow}-L^{\leftrightarrow}}-2\right)$
	$\lambda_{L^{\leftrightarrow}}=\log\left(\frac{(L^{\rightarrow}-L^{\leftrightarrow})^2}{n(n-1)L^{\leftrightarrow}-2L^{\rightarrow}L^{\leftrightarrow}+(L^{\leftrightarrow})^2}\right)/2$
Link probabilities	$P\left(A_{ij}=0,A_{ji}=0\right)=\frac{1}{1+2\exp(-\lambda_{L^{\rightarrow}})+\exp(-2\lambda_{L^{\leftrightarrow}}-2\lambda_{L^{\rightarrow}})}$,
	$P\left(A_{ij}=1,A_{ji}=0\right)=\frac{\exp(-\lambda_{L^{\rightarrow}})}{1+2\exp(-\lambda_{L^{\rightarrow}})+\exp(-2\lambda_{L^{\leftrightarrow}}-2\lambda_{L^{\rightarrow}})}$,

[2] The domain of $L^{\leftrightarrow}$ is obviously bounded from above by $L^{\rightarrow}$, since the set of reciprocal edges is a subset of all edges. Similarly, a maximum of $n\left(n-1\right)/2$ edges can be allocate as non-reciprocal edges, every additional edge automatically produces two reciprocal edges, hence, the lower bound.

	$P\left(A_{ij}=1, A_{ji}=1\right)=\frac{\exp(-2\lambda_{L\leftrightarrow}-2\lambda_{L\rightarrow})}{1+2\exp(-\lambda_{L\rightarrow})+\exp(-2\lambda_{L\leftrightarrow}-2\lambda_{L\rightarrow})}, \quad \forall i \neq j \wedge i,j \in \{1,\ldots,n\}$
Graph probabilities	$P\left(A=a\right)=\prod_{i<j}\left[P\left(A_{ij}=0, A_{ji}=0\right)\right]^{(1-a_{ij})(1-a_{ji})}\left[P\left(A_{ij}=1, A_{ji}=0\right)\right]^{a_{ij}(1-a_{ji})}$ $\left[P\left(A_{ij}=0, A_{ji}=1\right)\right]^{(1-a_{ij})a_{ji}}\left[P\left(A_{ij}=1, A_{ji}=1\right)\right]^{a_{ij}a_{ji}}$

ERGM with desired strength sequence (Park and Newman (2004))	
Considered set of graphs $\mathcal{G}$	$\mathcal{G}=\left\{w \in \mathbb{N}_0^{n\times n} : w_{11}=\cdots=w_{nn}=0\right\}$
Network statistic f_i	$f_i\left(w\right)=\sum_{\substack{j=1\\ j\neq i}}^{n} w_{ij}+w_{ji}$, computes the weight of node i
Desired strength	$s_i \in \mathbb{R}_{>0}, \quad \forall i \in \{1,\ldots,n\}$, denotes the desired weight of node i
Hamiltonian H	$H\left(w\right)=\sum_{i\neq j}\left(\lambda_i+\lambda_j\right)w_{ij}$
Partition function Z	$Z=\prod_{i\neq j}\frac{1}{1-e^{-\lambda_i-\lambda_j}}$ subject to the condition $e^{-\lambda_i-\lambda_j}<1, \quad \forall i \neq j \wedge i,j \in \{1,\ldots,n\}$
Expected link weights	$\mathbb{E}\left[W_{ij}\right]=\frac{e^{-\lambda_i-\lambda_j}}{1-e^{-\lambda_i-\lambda_j}}, \quad \forall i \neq j \wedge i,j \in \{1,\ldots,n\}$
Probability of link weights	$P\left(W_{ij}=w_{ij}^*\right)=\left(1-e^{-\lambda_i-\lambda_j}\right)e^{-(\lambda_i+\lambda_j)w_{ij}^*}, \quad \forall i \neq j \wedge i,j \in \{1,\ldots,n\}$ and $w_{ij}^* \in \mathbb{N}_0$
Graph probabilities	$P\left(W=w\right)=\prod_{i\neq j}P\left(W_{ij}=w_{ij}^*\right), \quad$ for $w \in \mathcal{G}$

Bibliography

Abbas, A. E., A. H. Cadenbach, and E. Salimi (2017). A Kullback–Leibler View of Maximum Entropy and Maximum Log–Probability Methods. *Entropy 19*(5).

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic Risk and Stability in Financial Networks. *American Economic Review 105*(2), 564–608.

Altman, E. I. and V. M. Kishore (1996). Almost Everything You Wanted to Know about Recoveries on Defaulted Bonds. *Financial Analysts Journal 52*(6), 57–64.

Anand, K., B. Craig, and G. Von Peter (2015). Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance 15*(4), 625–636.

Anand, K., I. van Lelyveld, A. Banai, T. C. Silva, S. Friedrich, R. Garratt, G. Halaj, I. Hansen, B. Howell, H. Lee, S. Martínez Jaramillo, J. Molina-Borboa, S. Nobili, S. Rajan, S. R. Stancato de Souza, D. Salakhova, and L. Silvestri (2018). The missing links: A global study on uncovering financial network structures from partial data. *Journal of Financial Stability 35*, 107–119.

Baral, P. and J. Fique (2012). Estimation of bilateral connections in a network: copula vs. maximum entropy. *Mimeo*.

Bardoscia, M., S. Battiston, F. Caccioli, and G. Caldarelli (2015). DebtRank: A microscopic foundation for shock propagation. *PLoS ONE 10*(6).

Bargigli, L., G. Di Iasio, L. Infante, F. Lillo, and F. Pierobon (2015). The multiplex structure of interbank networks. *Quantitative Finance 15*(4), 673–691.

Basel Committee on Banking Supervision (2010). Basel III: A global regulatory framework for more resilient banks and banking systems Bank for International Settlements.

Basel Committee on Banking Supervision (2015a). Making supervisory stress tests more macroprudential: Considering liquidity and solvency interactions and systemic risk. *BIS Working Paper*.

Basel Committee on Banking Supervision (2015b). Making supervisory stress tests more macroprudential: Considering liquidity and solvency interactions and systemic risk. BCBS Working Papers 29, Bank for International Settlements.

Battiston, S., G. Caldarelli, M. D'Errico, and S. Gurciullo (2016). Leveraging the network: a stress-test framework based on DebtRank. Statistics and Risk Modeling. *Statistics Risk Modeling 33*(3–4), 117–138.

Battiston, S., D. Delli Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz (2012). Default cascades: When does risk diversification increase stability? *Journal of Financial Stability 8*(3), 138–149.

Battiston, S., M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli (2012). DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. *Scientific Reports 2*, 541.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization.* Cambridge University Press.

Caldarelli, G., A. Capocci, P. De Los Rios, and M. Muñoz (2002). Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett. 89*(25).

Cerutti, E., S. Claessens, and P. McGuire (2011). Systemic Risks in Global Banking: What Available Data can tell us and What More Data are needed? Working Paper 11/222, International Monetary Fund.

Cimini, G., T. Squartini, A. Gabrielli, and D. Garlaschelli (2015). Estimating topological properties of weighted networks from limited information. *Phys. Rev. Lett. E 92*(040802(R)).

Cimini, G., T. Squartini, D. Garlaschelli, and A. Gabrielli (2015). Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports 5.*

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-Law Distributions in Empirical Data. *SIAM Rev. 51*(4), 661–703.

Cont, R., A. Moussa, and E. B. Santos (2013). *Network structure and systemic risk in banking systems*, pp. 327–367. Cambridge University Press.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory.* Wiley-Interscience.

Craig, B. and G. Von Peter (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation 23*(3), 322–347.

De Bandt, P. and O. Hartmann (2000). Systemic risk: A survey. Working Paper 35, European Central Bank, Frankfurt, Germany.

Drehmann, M. and N. Tarashev (2013). Measuring the systemic importance of interconnected banks. *Journal of Financial Intermediation 2*, 586–607.

Eisenberg, L. and T. H. Noe (2001). Systemic risk in financial systems. *Management Science 47*(2), 236–249.

Elsinger, H., A. Lehar, and M. Summer (2013). *Network models and systemic risk assessment*, Volume 48, pp. 287–305. Cambridge University Press.

Engel, J., M. Scherer, and A. Pagano (2019a). A block-structured model for banking networks across multiple countries. Working paper.

Engel, J., M. Scherer, and A. Pagano (2019b). Reconstructing the topology of financial networks from degree distributions and reciprocity. *Journal of Multivariate Analysis 172*, 210–222.

Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E 76*(2).

Gandy, A. and L. A. M. Veraart (2017a). A Bayesian methodology for systemic risk assessment in financial networks. *Management Science 63*(2).

Gandy, A. and L. A. M. Veraart (2017b). Adjustable Network Reconstruction with Applications to CDS Exposures.

Garlaschelli, D. and M. I. Loffredo (2004a). Fitness-dependent topological properties of the world trade web. *Physical review letters 93*(18).

Garlaschelli, D. and M. I. Loffredo (2004b). Patterns of link reciprocity in directed networks. *Phys. Rev. Lett. 93.*

Garlaschelli, D. and M. I. Loffredo (2008). Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E 78.*

Garlaschelli, D. and M. I. Loffredo (2009). Generalized Bose-Fermi statistics and structural correlations in weighted networks. *Phys. Rev. Lett. 102.*

Hałaj, G. and C. Kok (2013). Assessing interbank contagion using simulated networks. Working Paper 1506, European Central Bank.

Holland, P. W. and S. Leinhardt (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association 76*(373), 33–50.

Hüser, A.-C. (2015). Too interconnected to fail: A survey of the interbank networks literature. *Journal of Network Theory in Finance 1*(3), 1–50.

Jaynes, E. T. (1957a). Information Theory and Statistical Mechanics. *Physical Review 106*(4), 620–630.

Jaynes, E. T. (1957b). Information Theory and Statistical Mechanics. II. *Physical Review 108*(2), 171–190.

Kullback, S. (1997). *Information Theory and Statistics.* Dover Publications, Mineola, NY, USA.

Kullback, S. and A. Leibler (1951). On Information and Sufficiency. *Ann. Math. Statist. 22*(1), 79–86.

Mawhin, J. (2013). Variations on Poincaré–Miranda's theorem. *Advanced Nonlinear Studies 13*.

Mazzarisi, P. and F. Lillo (2017). *Methods for Reconstructing Interbank Networks from Limited Information: A Comparison*, pp. 201–215. Springer International Publishing, Cham.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance 35*(5), 1114–1127.

Newman, M. E. J. (2010). *Networks An Introduction.* Oxford University Press.

Park, J. and M. E. J. Newman (2004). The statistical mechanics of networks. *Phys. Rev. 70.*

Rogers, L. C. G. and L. A. M. Veraart (2013). Failure and Rescue in an Interbank Network. *Management Science 59*(4), 882–898.

Roncoroni, A., S. Battiston, M. D'Errico, G. Hałaj, and C. Kok (2018). Interconnected Banks and Systemically Important Exposures.

Roukny, T., C.-P. Georg, and S. Battiston (2014). A network analysis of the evolution of the German interbank market. Discussion Paper 22, Deutsche Bundesbank.

Servedio, V. D. P., G. Caldarelli, and A. Buttà (2004). Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys. Rev. E 70*(056126).

Squartini, T., I. van Lelyveld, and D. Garlaschelli (2013). Early-warning signals of topological collapse in interbank networks. *Scientific Reports 3*(3357).

Turzański, M. (2012). The Bolzano–Poincaré–Miranda theorem – discrete version. *Topology and its Applications 159*(13), 3130–3135.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability 7*, 111–125.

Upper, C. and A. Worm (2004). Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review 48*, 827–849.

Visentin, G., S. Battiston, and M. D'Errico (2016). Rethinking financial contagion.

List of Tables

List of Figures