WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

MASTERARBEIT / MASTER'S THESIS

Reconstructing Financial Networks from Global Syndicated Loans Using Valued ERGMs

verfasst von / submitted by

Maroan El Sirfy, BSc. (WU)

angestrebter akademischer Grad / intended degree

Master of Science (WU) / MSc (WU)

Matrikelnummer / student ID number:	11839876
Studienrichtung / degree programme:	Digital Economy
Betreut von / Supervisor:	Prof. Dr. Stefan Sobernig

Wien, August 2025 / Vienna, August 2025

Abstract

The construction and analysis of synthetic and real financial networks have gained attention in recent financial research to understand systemic risk and interbank relationships. This thesis investigates financial network reconstruction through a dual approach. First, a systematic literature review was conducted to evaluate existing methodologies and challenges in financial network modeling. The review revealed, that while theoretical models are predominantly used due to data limitations, there is a growing trend toward fitting empirical data to generative models, especially within the interbank domain. However, cross-validation efforts of the resulting networks remain limited, and results often lack consistency across studies.

Based on these findings, valued Exponential Random Graph Models (ERGMs) were employed to reconstruct an interbank network using global syndicated loan data. The model was fitted to capture key structural properties of global systemically important banks (GSIBs), including directed edge formation, out-going loan weights, and reciprocity. To assess the representativeness, 500 synthetic networks were generated and evaluated against the observed network using aggregated structural and centrality measures, showing that the ERGM closely matches observed density, clustering, small-world geodesics and the heavy-tail of exposures while outperforming Erdős–Rényi, Watts–Strogatz, and Barabási–Albert baseline models. Findings indicate that ERGMs can generate synthetic financial networks that preserve key topological properties, suggesting their potential for approximating real-world financial interdependencies.

Further work is required to validate these findings at a larger scale and establish uniform evaluation frameworks for synthetic network comparison. This study contributes to the ongoing broader discourse on financial network modeling by demonstrating the strengths and limitations of ERGMs in constructing representative synthetic networks.

Zusammenfassung

Der Aufbau und die Analyse synthetischer und realer Finanznetzwerke haben in der jüngeren Finanzforschung an Bedeutung gewonnen, um systemische Risiken und Interbankenbeziehungen besser zu verstehen. Diese Arbeit untersucht die Rekonstruktion von Finanznetzwerken mittels eines dualen Ansatzes. Zunächst wurde eine systematische Literaturrecherche durchgeführt, um bestehende Methoden und Herausforderungen der Finanznetzwerkmodellierung zu evaluieren. Die Auswertung zeigt, dass aufgrund begrenzter Datenverfügbarkeit überwiegend theoretische Modelle eingesetzt werden, jedoch ein wachsender Trend zur Nutzung empirischer Daten für generative Modelle besteht, insbesondere im Interbankbereich. Allerdings bleiben Bestrebungen zur Kreuzvalidierung der resultierenden Netzwerke begrenzt, und die Ergebnisse sind oft über die verschiedenen Studien hinweg nicht konsistent.

Auf der Grundlage dieser Ergebnisse wurden gewichtete Exponential Random Graph Models (valued ERGMs) eingesetzt, um anhand globaler Konsortialkreditdaten ein Interbankennetzwerk zu rekonstruieren. Das Modell wurde so kalibriert, dass zentrale Struktureigenschaften für global systemrelevante Banken (GSIBs) erfasst werden, darunter die Bildung gerichteter Kanten, die Gewichte ausgehender Kreditbeziehungen sowie die Reziprozität. Zur Beurteilung der Repräsentativität wurden 500 synthetische Netzwerke generiert und anhand aggregierter Struktur- und Zentralitätsmaße mit dem beobachteten Netzwerk verglichen. Dabei zeigte sich, dass das ERGM der beobachteten Dichte, Clusterbildung, Small-World-Geodäten, und sowie die endlastige Verteilung der Interbankenforderungen (engl.: "heavy tail") eng abbildet und dabei die Basismodelle von Erdős-Rényi, Watts-Strogatz und Barabás-Albert übertrifft. Die Ergebnisse deuten darauf hin, dass ERGMs synthetische Finanznetzwerke generieren können, die wesentliche topologische Eigenschaften bewahren, was auf ihr Potenzial hindeutet, dass Interdependenzen in realen Finanzsystemen approximiert werden können.

Zur Validierung dieser Ergebnisse in größerem Maßstab und zur Etablierung einheitlicher Bewertungsrahmen für den Vergleich synthetischer Netzwerke ist weitere Forschung erforderlich. Die Studie leistet einen Beitrag zum laufenden breiteren Diskurs der Finanznetzwerkmodellierung, indem sie die Stärken und Grenzen von ERGMs bei der Konstruktion repräsentativer synthetischer Netzwerke herausarbeitet.

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Stefan Sobernig, for his invaluable guidance, support and patience throughout this project.
I am also deeply grateful to my family for their constant encouragement and understanding during my studies.

Contents

List of Figures

List of Tables

1 Introduction

Economic networks and financial networks, both specialized forms of social networks, start their history in the 1990s [1, 2]. As will be eluded to in the definitions, the terms financial and economic network are popularly used for the same concept. This stems from the variety of interdisciplinary fields that combine in the area of economic networks, as well as the diversity of publication journals (more on this in 4.1).

Economic networks, usually referring to networks that focus on global trade, national industrial trade, or even production networks, try to analyze macroeconomic stability from a systemic perspective. Financial networks, on the other hand, focus on monetary exchanges and include, most notably, interbank, transaction, and stock networks. Studying the structure of the financial system rather than the behavior of single banks has gained traction in the last decade [3], especially since this allowed the assessment of systemic financial risk.

The awareness that economic behavior is ideally investigated by considering all the relevant interconnected components coupled with the increasing availability of relevant data through the digitalization allowed for the creation of different types of financial and economic networks. Stabilizing the world economy in times of financial shock is among the main drivers for financial network researchers. Many contributions draw inspiration from preceding financial or societal events, such as a financial crisis, prompting new approaches.

The *4Rs*-repeatable, reproducible, replicable, and reusable [4], set methodological expectations for credible network science. *Repeatable* implies that a workflow yields identical outputs given the same inputs and environment; *reproducible* that an independent researcher can obtain the same results using the shared code, configuration, and data; *replicable* that core findings persist under alternative data samples or implementations; and *reusable* that data, code, and documentation are sufficiently transparent, licensed, and modular to support downstream use. These requirements frequently conflict with the realities of financial data: access is often proprietary or restricted, documentation is uneven, and legal/privacy constraints impede sharing. As a result, many published studies cannot fully satisfy the 4Rs, despite using broadly available methodologies.

While data limitations are not necessarily idiosyncratic to this field, they have an extensive impact on the ability of research in this area to accurately assess real-world financial conditions. The informativeness of a financial contagion study depends on its ability to best model these conditions, where meaningful results usually include insights into macroeconomic fluctuations, risk transmission, shock, stability and resilience, and default. For policymakers and other decision-makers, it is essential that these insights accurately assess the situation and translate well from academic theory to the real world.

Because data are scarce, yet accurate network modelling is highly valued, various approaches attempt to bridge gaps created by missing, noisy, or non-existent observations. Even with relatively rich data, constructing representative real financial networks remains challenging. It is therefore of interest to review approaches that prioritise accuracy in network construction, to reimplement common (or particularly informative) methods, and to evaluate their effectiveness. These approaches range from purely simulation-based, theory-driven procedures that yield *synthetic* networks to data-driven procedures that infer *empirical* (real) networks from observed information. In practice, partial datasets are often combined with theoretical constraints to recover latent structure. Because these terms are used precisely in what follows, explicit definitions are established first.

This thesis covers two approaches: (i) a systematic literature review covering a broad spectrum of financial-network research, and (ii) the construction of empirical networks from syndicated-loan data, benchmarked against baseline models and exponential random graph models.

Structure of the Thesis

Building on the research objectives articulated in Section 3.1, the thesis is organized into seven substantive sections.

Background Section 2 acts as a glossary. It defines and describes important contexts used throughout the thesis, providing insight into the basics, important network measures, the idiosyncrasies of the syndicated loan market, and exponential random graph models (ERGMs).

Methodological design Section 3 outlines the dual methodological framework. First, the protocol for a systematic literature review is specified, including search strategy, screening criteria, and data-extraction procedures. Second, the modeling workflow for constructing both real and synthetic financial networks is detailed. The workflow encompasses data acquisition from syndicated-loan records, extraction of a core subgraph and specification of baseline random-graph families and more advanced generative models.

Systematic literature review Section 4 reports the outcomes of the review, providing a comprehensive map of existing research on construction and generation of financial networks. The analysis shows prevailing data sources, modeling techniques, and evaluation practices. It notes existing common criteria for assessing the representativeness of synthetic networks. Identified methodological gaps and limitations are also discussed.

Network construction results Section 5 presents the empirical results of the network construction pipeline. Descriptive analyses of the empirical (real) network and its synthetic counterparts are compared, focusing on topological and multiplex attributes identified as critical in the literature. The empirical findings reveal generalization challenges associated with current modeling approaches, and favorable structural results using ERGMs.

Discussion Section 6 integrates insights from the literature review and the empirical implementation. Principal findings are interpreted, implications for model selection and diagnostic practice are considered, and results are benchmarked against comparable studies. Strengths, data-driven limitations, and future research directions are considered.

Related Work Section 7 places the study within the closest body of scholarship. Comparisons are drawn to previous systematic reviews, research constructing syndicated loan networks, and studies applying valued ERGMs in financial contexts, thereby clarifying the incremental contribution of the present thesis.

Conclusion Section 8 summarizes the main contributions, reflects on methodological constraints, and proposes avenues for further investigation into the construction and evaluation of synthetic financial networks.

Reproducibility All scripts, code and the networks are made available through a repository, found in Appendix B.2.

2 Background

Clarifying the sometimes largely different vocabulary used in the literature for the same concept may alleviate uncertainty surrounding the terminology. Furthermore, it also serves to define the concepts for the rest of the thesis.

2.1 Foundations in Network Theory

2.1.1 What is a Network?

Throughout modern and classical literature, a graph is the intuitive mathematical representation of a network, where a network is defined as an ordered pair of disjoint sets $G = (V, E)$, with V being the set of nodes (vertices) and E the set of edges (links between vertices) [5]. Any pair of vertices is called a dyad [6].

Vertices connected via edges are also called adjacent vertices. Edges may be binary (present or absent) or valued, for example, weighted by the amount of credit between two banks. The edges can be undirected or directed, a comprehensive basis for these definitions being established in [7, 8, 9]. The term network and graph is used synonymously throughout literature. A network in a financial context, could be nodes being the banks, and directed weighted edges between the nodes showing the flow of money.

Multi-edges refer to the same dyads sharing multiple edges; in the directed network in the same direction.

Although the vast majority of literature focuses on static networks, interest has increased with dynamic networks. These are often also called temporal or evolving networks, and refer to networks where node interactions or "contacts" change with time [10, 11]. Temporal networks are not the focus of this analysis, but important literature will be mentioned to put methodological approaches into context.

2.1.2 Bipartite vs. Unipartite

A bipartite graph is a common structure for economic networks, consisting of two disjoint node sets, often denoted U and V, with edges only between, but never within, the sets [9]. In a financial context, one may let U represent lending banks and V represent borrowers; edges then indicate that a particular bank in U extends credit to a borrower in V.

A unipartite network, by contrast, contains a single node type. A one-mode projection of a bipartite graph can be obtained by connecting two nodes in U whenever they share a neighbor in V. For example, projecting the lender–borrower bipartite graph onto the lender set yields a network in which two banks are adjacent if they both finance the same borrower. Such projections can inflate apparent connectivity and obscure the original two-mode structure unless edges are weighted (e.g., by the number of joint loans or the total amount loaned).

2.1.3 Core–Periphery Structure

Block structures in networks and their methods to identify them have been widely discussed since the early 2000 [12], mentioning several ways to view core-periphery structures. The idea is to partition nodes into two classes, one as the core and the other periphery, or to see the network as just one group, to which all nodes belong, in a physical sense with a "center" of nodes and periphery nodes surrounding it. This is especially relevant in the context of financial networks, where a group of "important", well-connected banks might be the center, and smaller banks, with fewer connections surround the periphery [13, 14, 15].

2.2 Descriptive Network-Analytical Measures

2.2.1 Degree and Degree Distribution

The Degree The degree of a vertex v is the number of edges that are incident to the vertex v [16]. In plain terms, the degree number is the number of neighbors that share an edge with a vertex in an undirected graph.

In a directed graph, because the edges are going in and out of a vertex v, the in-degree is the number of neighbors that share a directed edge towards v, whereas the out-degree is the number of neighbors that share a directed edge away from v. The combined in- and out-degree is the total degree of v.

A vertex with a high degree or the maximum degree of the network may be called a hub. If edges have weights, the degree of v is also called strength, and can also be directional.

Degree Distribution The degree distribution is the probability distribution of degrees of nodes in the network, and shows how many nodes have each possible degree number in the network [17]. In undirected graphs, the degree of a node is the different number of edges each node might have. In directed graphs, separate in- and out-degree distributions are typically considered. Empirically, it is often visualized as a histogram or empirical cumulative distribution to observe the relative occurrence of low-degree versus high-degree nodes. Because the degree distribution captures the heterogeneity of node connectivity showing, for example, whether most nodes have few edges or if a small subset are hubs, it is an important descriptive measure used to characterize and compare network structures.

When the majority of nodes have a low degree, and only a few are high, this is called a skewed, or heavy-tail distribution.

Degree-distribution diagnostics Heavy-tail behaviour can be assessed by (i) the Gini coefficient of the degree sequence [18], and (ii) maximum-likelihood estimation of a discrete power-law tail following [19]. The exponent α and lower cut-off $x_{\min}$ can be reported alongside a Kolmogorov–Smirnov goodness-of-fit p-value [20].

2.2.2 Global and Local Connectivity Measures

Key statistics are commonly computed measures to quantitatively describe a network. To make explanations relevant and context driven, an interbank network, where nodes represent banks and edges denote bilateral lending, is used as a subsequent example.

Connected Components A connected component in an undirected graph is a maximal set of nodes, such that every pair of nodes within the set is joined by at least one path [21]. In a bank network, multiple components correspond to clusters of banks that never interact with banks outside their cluster, which indicates possible market or regional segmentation. If the largest component contains nearly all banks, the network is nearly connected, suggesting that they form a cohesive market. Otherwise, fragmentation may indicate isolated markets or data limitations.

Neighborhoods The 1-step neighborhood of a node i consists of i plus all nodes directly adjacent to i, together with any edges among these neighbors [1]. More formally, the ego-network of i is the induced subgraph on $\{i\} \cup \{j : (i, j) \in E\}$ [1]. One can also define a k-step neighborhood as the set of nodes reachable from i within k hops. In a simple interbank network, where the nodes represent banks and a directed edge $(i \to j)$ indicates that bank i has loaned to bank j, the ego-network of bank i contains all banks that either borrow from or lend to i. Neighborhood measures are building blocks for clustering and centrality statistics and help show whether localized lending communities exist within the broader interbank market.

Clustering Coefficient The global clustering coefficient (also called transitivity) is computed by counting the number of closed triplets (triangles or triadic nodes) divided by the total number of connected triplets [22]. In a simple interbank network, a high clustering coefficient means that if banks A and B both lend to bank C, then A and B also transact directly, forming a triangle. Such triangle closures often indicate tightly connected banking cores or repeated partnerships among a small group of institutions. Because clustering affects how risks and information propagate through the network, it serves as a descriptive measure of local cohesion.

Average Path Length The average path length of a connected component is the mean of the shortest-path distances between all pairs of nodes within that component [23]. If the network is not fully connected, one typically computes it only on the largest component, ignoring infinite distances. The shortest path distance between any dyad is called the geodesic distance. In an interbank network, a small average path length implies that any two banks can be reached through few intermediary credit relationships, reflecting efficient connectivity and rapid propagation.

Assortativity Assortativity quantifies the tendency of nodes to connect to others with similar degree. The degree-based assortativity coefficient r is defined as the Pearson correlation between the degrees d_i and d_j of nodes at either end of each edge [24]. In a bank network, a positive r indicates that highly connected banks tend to lend to other highly connected banks, while a negative r implies that large banks preferentially transact with smaller, less-connected institutions.

Reciprocity For a directed graph, reciprocity is defined as the fraction of directed edges that are mutual. Let L be the total number of directed edges in G, and $L_{\leftrightarrow}$ be the number of edges $(i \to j)$ for which the reverse edge $(j \to i)$ also exists [17, 1, 25], and can be expressed as

$$\text{Reciprocity} = \frac{L_{\leftrightarrow}}{L}.$$

In a financial lending network where direction matters (i.e., bank i lends to bank j), reciprocity quantifies the extent to which banks both lend to and borrow from one another. A high reciprocity value suggests reciprocal credit relationships or bilateral credit lines, indicating strong bilateral trust; low reciprocity implies one-sided lending ties.

Density Network density is the ratio of the number of observed edges E to the maximum possible edges among N nodes. For an undirected graph, Golbeck [17] formulates

$$\text{Density} = \frac{2E}{N(N-1)}$$

In an interbank network, density measures the fraction of all potential credit relationships that are active. A high density indicates a richly interconnected market with many direct links, facilitating liquidity but also enabling rapid contagion; a low density suggests sparse, selective lending ties, and potential segmentation.

2.2.3 Centrality Measures

Degree Centrality Degree centrality of a node i is simply its degree d_i, i.e., the number of edges incident to i. In a bank lending network, a bank with a high degree centrality has credit relationships with many other nodes, showing broad market engagement or diversification. Because degree captures raw connectivity without regard to path lengths or neighbor importance, it is the only straightforward measure of how "active" a node is.

Betweenness Centrality Betweenness centrality of a node i measures the extent to which i lies on shortest paths between other node pairs. Formally, if σ_{jk} denotes the number of shortest paths between nodes j and k, and $\sigma_{jk}(i)$ the number of those paths that pass through i [16], then

$$C_B(i) \;=\; \sum_{j\neq k\neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}}.$$

In an interbank network, a bank with high betweenness centrality serves as a bridge or broker connecting otherwise disparate clusters of banks. Such banks may exert influence over interbank liquidity flows and are critical for maintaining network cohesion; their removal or distress could fragment the network more severely than the removal of a high-degree but locally clustered bank.

Closeness Centrality Closeness centrality of a node i is the reciprocal of the average shortest-path distance from i to all other reachable nodes:

$$C_C(i) \;=\; \frac{N_C - 1}{\sum_{j\in C\setminus\{i\}} d(i,j)},$$

where $d(i,j)$ is the shortest-path distance and N_C is the size of i's connected component C [16].

In a simple interbank network, a bank with high closeness centrality can quickly reach (or be reached by) any other bank through few intermediaries, which could reflect better access to liquidity. Closeness thus captures global integration: banks with higher closeness are centrally located in terms of shortest-path distances.

Eigenvector-Type Centralities Besides centralities purely based on node distance, eigenvector-type centralities assign relative scores to nodes based on the principle that connections to high-scoring neighbors contribute more to a node's importance, than connections to low-scoring neighbors. If A is the adjacency matrix of an undirected network, the eigenvector centrality x_i of node i satisfies

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} \, x_j,$$

where λ is the largest eigenvalue of A [16]. Variants include Katz which is standard for weakly connected graphs using a damping factor, and PageRank, which applies a random-walk interpretation with a damping factor and is well suited for directed networks.

In an interbank context, a bank with high eigenvector centrality is connected to other highly central banks, indicating influence beyond its immediate degree: it sits in a core of well-connected institutions. Such centrality measures capture recursive prestige or systemic importance more effectively than degree alone.

2.3 Random and Synthetic Networks

Descriptive measures reveal what the observed network looks like, but assessing how unusual those features are requires reference models. Three canonical baselines are considered here: the Erdős–Rényi (ER) model, which assigns edges uniformly at random; the Watts–Strogatz (WS) small-world model, which combines high local clustering with short global paths; and the Barabási–Albert (BA) scale-free model, whose preferential attachment produces heavy-tailed degree distributions.

2.3.1 Random Networks & Erdős–Rényi model

Random networks were first discussed in the context of mathematical biology in 1951 [26]. The properties were elaborated by Gilbert in 1959 [27] and made popular in the same year by Erdős and Renyi [28]. Although all discuss the idea of a $G_{n,p}$ model, the specific model is also popularly (and arguably falsely) referred to as the Erdős-Renyi (ER) model in literature.

In general, the term random graph is generally used inconsistently in literature and can also sometimes refer to the $G_{n,p}$ model. Random graphs are also sometimes referred to as stochastic graphs and vice-versa because the process with which they are generated follow a stochastic model for the degree distribution. For the purposes of this thesis, random graphs and random networks will be the overall category. All stochastic graphs that are generated with a specific degree distribution, such as ER and the ones discussed in the following paragraphs, fall under this category. An important note, when a paper calls a random graph an ER model, it will also be called ER model as part of this analysis to be consistent with the literature, although it might not actually follow the model of Erdős and Rényi [28].

Random networks have a homogeneous degree distribution, as the existence of each edge has the same probability. They generally follow a Poisson degree distribution [29, 30], however, it has been known and pointed out since the early 2000s that this distribution for a random network is inadequate for representing real-world networks ([31, 32, 22] as cited in [33]). Many empirical networks exhibit heavy-tailed or skewed degree distributions rather than the Poisson-like pattern of a random graph.

2.3.2 Small-World Networks & Watts–Strogatz Model

Although many large real-world networks can exhibit the property of short distances between vertices, networks generated with the $G_{n,p}$ model also have the property of short vertex distances even with a large number of vertices, a phenomenon called a small-world property. However, the $G_{n,p}$ model often does not have the high clustering attribute, which is often present in many real-world networks. Famous models for combining these properties are the Watts and Strogatz (WS) model [22] and the Newman-Watts variation model [34].

2.3.3 Scale-Free Networks & Barabási Albert Model

Scale Free Networks Another phenomenon that can often be found with real-world networks is that a few nodes in a network will have a higher degree than most other nodes (hubs), which is also called a heavy-tailed distribution. Many real examples can be mentioned, and Albert, Jang, and Barabási [35] famously found this behavior with the Internet. Although they falsely equated their found heavy-tailed distribution with a power-law distribution [36, 37], many heavy-tailed networks do indeed appear to follow a power-law (Pareto) distribution.

Preferential Attachment The corresponding explanatory model for the emergence of a scale-free power-law distribution is often called the preferential attachment (PA) model. As the procedure was suggested by Barabási and Albert [38], it is also often called the Barabási-Albert model (BA model). In Section 4.1 literature popularly uses the BA model, even though Bollobás and Riordan [39] later refined the model offering a more precise procedure allowing systematic mathematical analysis.

2.3.4 Synthetic Networks

The term synthetic is most often synonymous with *generated* or *constructed*. Many random networks, such as the ER/WS/BA model, which are also often called baseline or null models, popularly fall into this term. As in any interdisciplinary field, definitions might not be adhered to in all fields, and most often it is not clear what a synthetic network entails. Lim et al. attempted to define a synthetic network as a "graph generated using a *graph generative model*", while a *realistic synthetic graph* uses "fitted model parameters derived from [...] real-world graphs" [40]. For the purpose of this thesis, a synthetic network will follow this latter definition and be any network that, as opposed to a real network, does not have any real-world data included, but rather the network or its constituent node-vertex matrix is purely generated, while parameters can be derived from or adjusted to real-world networks to attain a matching synthetic graph.

2.4 The Syndicated Loan Market as a Network

2.4.1 Financial Networks

Definitionally what constitutes a financial network varies in literature. Most often, financial networks are associated with interbank networks, which represent liabilities between banks and financial institutions [41, 42], or represent a transaction network of banks or in general monetary flow [43]. Liabilities between banks are often also called interbank exposure and is an important concept in the context of financial contagion [44]. Contagion analysis tries to assess whether the failure of a bank (to satisfy its liabilities and pay its debts, i.e., default) causes others in the network to default on their obligations. If a bank failure causes not one bank but a chain of bank failures, it is generally defined as a default cascade [41, 45].

Literature is not united in their definition of systemic risk and whether a default cascade or contagion is a necessary concept for systemic risk to be present in a network [44, 46]. Jackson and Pernoud [47], who have made an attempt at unification, provide a taxonomy of systemic risk in financial networks.

As is popular, the financial network will be the overarching category and can also include economic networks. A real financial network (empirical network) is based on empirical available data.

2.4.2 Introduction to Syndicated Loans

Syndicated loans are usually issued by multiple lenders to a single borrower. The lenders are often a consortium of financial institutions and investment banks that coordinate, underwrite and sell the loan [48]. This structure allows borrowers to access larger sums and lenders to diversify risk. In rare cases, a syndicated loan can be underwritten by a single entity, a sole lender. Sole lenders in syndicated loans can stem from a variety of reasons, among them loan structuring setup, interim funding, or challenges during the syndication process. Syndicated loans are not standardized financial instruments that are traded like other securities, but rather ad hoc customized arrangements and contracts that institutions partake in.

2.4.3 Key Roles and Terminology

As part of the syndication process in a consortium, financial institutions take on different roles. The name of the role and responsibilities are neither legally nor scientifically agreed on internationally and can differ for each syndicate [49, 50]. Most often, one institution is responsible for forming the syndicate, finding suitable participants and possible lenders. This role is most often called "Mandated Lead Arranger" or "Bookrunner", but other terms can be "Lead Arranger", "Syndication Agent", or "Agent" [49]. Based on the ad hoc nature of syndicated loans, the actual responsibility of the leading roles differs for each tailored syndication contract. Any institution that participates, but with less involvement than the leading role, can be found, among many other possibilities, as "Arranger", "Co-Arranger", "Co-Lender", or simply "Participant". Sometimes, a purely administrative role is highlighted with "Administrative Agent" or "Documentation". In addition, roles in syndicated loans can be sold, with amended agreements [49]. Because of the banks non-standard approaches and intentional information asymmetries, literature has resorted to assigning institutions common roles: "Lead Arranger" (also simply "Lead"), "Agent" and "Participant" [51].

2.5 Statistical Modeling of Networks: ERGMs

2.5.1 Exponential-Family Random Graph Models (ERGMs)

Statistical modeling of networks moves beyond descriptive summaries and synthetic theoretical models by providing a unified inferential framework for tie formation. Exponential-family random graph models (ERGMs) [52, 53] specify the probability of observing a particular network G as

$$P_\theta(G) \;=\; \frac{\exp\{\sum_k \theta_k \, g_k(G)\}}{\kappa(\theta)},$$

where each $g_k(G)$ is a network statistic (e.g. number of edges, triangles, degree-based counts, homophily terms) and θ_k is the corresponding parameter [52, 54]. By fitting multiple statistics simultaneously, ERGMs quantify how configurations such as overall density, clustering, degree heterogeneity, or attribute-based mixing influence the log-odds of each tie, while accounting for their interdependence. Positive values of θ_k indicate that the associated configuration occurs more frequently than under a baseline random graph, whereas negative values imply inhibition. Estimation typically relies on maximum pseudo-likelihood or simulation-based maximum likelihood; the methodological details and model specification adopted in this study are provided in Section 3.

2.5.2 Estimating ERGMs with MCMC Maximum Likelihood

Because the normalizing constant $\kappa(\theta)$ involves a sum over all possible graphs with n nodes, exact maximum likelihood is infeasible. Snijders [55] proposed replacing the intractable likelihood with a *Markov-chain Monte Carlo* (MCMC) approximation [56, 57]: a Metropolis–Hastings chain is run whose stationary distribution is the current ERGM, and network statistics are averaged over the sampled graphs. Wang et al. [58] formalized this MCMC maximum-likelihood estimation (MCMLE) algorithm, which updates θ iteratively until the simulated means of the statistics match their observed values. This is still computationally expensive as the MCMLE must explore a dyad state space of size $2^{n(n-1)/2}$ [59], but the approach yields consistent tractable estimates while accommodating the complex dependence embodied in ERGMs. The procedure is widely used and is implemented in the `ergm` package within the `statnet` suite [60, 61], which automates Markov chain initialization, burn-in, thinning, and convergence diagnostics.

2.5.3 Valued ERGMs (vERGM) and Edge Weights

Many financial networks record not only whether a tie exists (binary) but also its magnitude, for instance the exposure an interbank loan creates. Valued ERGMs incorporate that information by combining an edge weight *reference measure* with the binary ERGM form,

$$\Pr_{\theta}\bigl(G^{(w)}\bigr) = \frac{\exp\bigl\{\sum_k \theta_k \, g_k(G^{(w)})\bigr\}}{\kappa(\theta)},$$

where the reference assigns baseline probabilities to each weight value (e.g. geometric or Poisson for count data) and the statistics $g_k(\cdot)$ now act on those weights [62]. The implementation of ERGMs with weight was intended to *count* the amount of relationships that occur between the nodes [62], but they can be repurposed to represent any weight. Coefficients maintain the interpretation of binary ERGMs, where positive values increase the log-odds of networks exhibiting large total exposure, strong dyadic flow, or other weighted configurations. Exemplary the monetary amount of each interbank loan can be viewed as a non-negative count and adopt a geometric reference, allowing the model to capture both the presence of ties and the heavy-tailed distribution of their sizes.

2.5.4 Common ERGM Statistics and Interpretation

ERGMs are built from statistics (parameters) that each highlight a particular structural feature of the network [53]. Four broad families are repeatedly mentioned in the literature:

1. **Baseline connectivity.** In binary models, this is a simple edge count; in valued models it may be the number of non-zero ties or the total volume of weights, capturing overall density or market size [52, 54].
2. **Heterogeneity in tie strength.** Indicators that flag edges whose weight exceeds one or more thresholds allow the model to reproduce heavy-tailed exposure distributions common in real-world data [62].
3. **Attribute mixing.** Statistics that condition on nodal attributes (e.g. bank size, region) test for homophily, whether similar institutions lend more, or in different amounts, to each other [63, 64].
4. **Dyadic dependence.** Terms that combine weights in both directions of a pair measure reciprocity in valued networks [62, 65], while triadic or closure terms (where implemented) account for clustering [54].

The sign and magnitude of a coefficient indicate whether the corresponding configuration is encouraged (positive) or discouraged (negative) relative to a random baseline. The specific statistics of these families included in the empirical model and how they are implemented are provided in Section 3.

When modeling with directed ERGMs, a reciprocity term captures this mutual-tie propensity beyond degree effects; if the loan data are treated as undirected, then reciprocity is not applicable (and analogous mutuality is reflected instead in clustering or edge weights).

Because the fitted model provides a full probability distribution, it can also be used to generate synthetic networks whose structural features mirror the observed graph; an important property for goodness-of-fit checks and counterfactual simulations.

2.5.5 AIC and BIC

Competing ERGM specifications are compared with information criteria that trade off fit against complexity [61].
The Akaike Information Criterion (AIC; [66]) is

$$\mathrm{AIC} = -2\log L(\hat{\theta}) + 2K,$$

while the Bayesian Information Criterion (BIC; [67]) is

$$\mathrm{BIC} = -2\log L(\hat{\theta}) + K\log N,$$

where $L(\hat{\theta})$ is the maximised likelihood, K the number of free parameters, and N the number of dyads (potential edges). Both reward goodness of fit via the log-likelihood term and penalize over-parameterization, where BIC applies a heavier penalty that grows with network size. Lower values indicate a preferable balance, and lower values are an indication for the model with the smaller criterion, although it must be mentioned that AIC and BIC are often only useful as approximations, as the independence assumptions of samples are not met in ERGMs [61].

2.5.6 ERGM-Specific Diagnostics

Simulation-based goodness-of-fit (GOF) diagnostics provide the standard means to assess whether an ERGM reproduces the structural features of an observed network [60, 61]. Lusher et al. [64] call this heuristic goodness of fit in the ERGM context. After estimating the model parameters, a sample of m networks is generated from the fitted ERGM. For each simulated graph a collection of summary statistics is recorded, typically including the degree distribution, reciprocity, triadic closure, the distribution of geodesic distances, and, when analyzing weighted graphs, the edge-weight distribution [61, 62]. The empirical distribution of each statistic derived from the simulations is then contrasted with its counterpart in the observed network. If the observed value lies within a central quantile range, the model is considered adequate with respect to that statistic. Common analysis employs $100 \leq m \leq 1000$ simulations [60, 61, 64]. Graphical displays such as box plots, or quantile–quantile plots can be used to summarize the comparison and to analyze any systematic lack of fit [61] or lack of fit of a particular feature of the network [64].

2.6 Preparation for a Systematic Literature Review

To ensure rigor and reproducibility, the forthcoming literature review adopts the hybrid search framework developed by Wohlin et al. [68, 69]. This framework combines a structured database search with iterative backward and forward snowballing, to maximize recall while maintaining screening efficiency. Its selection is motivated by two considerations: (i) the approach is widely recognized in empirical software engineering research, which is a methodologically adjacent area to the present study, and (ii) its step-by-step checklists generalize to the financial and network theory inquiries. Controlled vocabulary, inclusion criteria, and stopping rules were tested in Scopus and Google Scholar, proceeding with the full search strategy described in Section 3.

3 Research Design

This section delineates the guiding research questions and the overarching design of the study. Two complementary strands are covered: (i) the protocol for conducting a systematic literature review, and (ii) the methodological framework for constructing financial network models.

3.1 Research Questions

Within the broad and rapidly evolving field of financial network analysis, the thesis concentrates on three interrelated research questions:

RQ1: **Motivations for synthetic data generation**
For what purposes, and to what extent, is synthetic data employed when constructing financial network models? The question identifies typical sources of missing or unobservable information and examines the generative strategies adopted to overcome such limitations.

RQ2: **Prevailing modeling techniques**
Which generative and modeling approaches are most frequently applied to produce financial networks, and where do these approaches fall short in reproducing empirically observed structures? Techniques ranging from baseline random-graph families (e.g., Erdős-Rényi, Watts-Strogatz, Barabási-Albert) to advanced statistical network frameworks are systematically cataloged and critically evaluated for their ability to reproduce important topological and multiplex features observed in empirical financial networks.

RQ3: **Representativeness of generated networks**
To what degree do synthetic financial networks represent the empirical phenomena they seek to emulate? The analysis first consolidates the criteria and diagnostics used in the literature to judge representativeness, then contrasts the empirical fidelity of networks derived from complete data, partially observed data, and wholly synthetic data. Diagnostic metrics will be selected in accordance with best practice identified during the review (e.g., degree distributions, centrality metrics, geodesic distance profiles).

3.2 Design of the Literature Review

A transparent and reproducible search strategy was implemented to minimize publication bias and ensure comprehensive coverage. The design follows the hybrid protocol introduced in Section 2.6: a database search anchors the process, and iterative snowballing expands the corpus.

3.2.1 Literature Data Sources

Two multidisciplinary databases were queried: Scopus and Google Scholar. While Google Scholar is frequently chosen to seed hybrid searches [70, 71, 68], comparative evidence indicates that Scopus retrieves a higher proportion of relevant records [72]. Employing both databases therefore combines breadth of coverage with precision.

Construction of the start set Search strings listed in Table 1 (adapted from Wohlin [69]) were executed sequentially, with each "Listing" substituted into its designated slot, yielding twelve queries per database. Candidate start-set papers were retained only if titles and abstracts satisfied relevance to topic and quality criteria.

Table 1: List of keywords for a search string

Listing 1	Listing 2	Listing 3	Listing 4
financial bank financial transaction	network structure	graph	construction generation synthetic data

Scopus search Queries in Scopus were executed without a publication-year filter and sorted by relevance. To suppress noise, domain-irrelevant keywords identified were excluded (see Appendix A.1). Screening against inclusion criteria continued until two consecutive result pages produced no new eligible papers.

Google Scholar search Google Scholar was queried with the same twelve expressions, without the year filter. Because Google Scholar orders results by perceived relevance, screening ceased once a page yielded no new inclusions, preventing diminishing returns from later, often largely redundant pages.

3.2.2 Study Selection and Search Strategy

All articles retrieved at each iterative stage were first screened on title, abstract, and author-supplied keywords. Provisionally relevant papers were then selectively skimmed by section or, where necessary, in full to verify topical alignment with the research questions. Following Kitchenham [73], both peer-reviewed journal articles and conference proceedings were considered, as excluding the latter could introduce publication bias.

Inclusion The search terms in Table 1 were formulated to mirror the research questions. A paper qualified for inclusion if it satisfied at least one of the following criteria:

(a) The paper provides a review or survey of network construction in a financial or economic context; or

(b) The study addresses at least one term in "Listing 4," i.e., it proposes, evaluates, or applies a method for generating a network.

Foundational work required to understand later contributions was added selectively under the hybrid strategy. Only publications in English or German were retained. This scope limits peripheral analyses (e.g., social-network construction) that do not directly inform the stated research objectives; residual bias is examined in Section 4.4.

Backwards Snowballing Reference lists of all papers in the current set were examined iteratively. Unlike the exhaustive scanning recommended in Wohlin's original guidelines [68], only relevant citations judged in context during full-text reading were traced. This modification (i) integrates information extraction for the literature review with paper retrieval, (ii) operationalizes the guideline that advocates inspecting how and where a cited paper is used, and (iii) incidentally captures forward-citation candidates appearing in similar textual locations.

Forward Snowballing Each included paper was queried in Google Scholar to retrieve forward citations; Scopus's "Cited by" function provided an additional check. Forward links were reviewed using the same title/abstract screen described above. Because full texts were already being examined during backward snowballing, instances where a forward citation appeared in a comparable context could be flagged immediately, reducing duplication across the two snowballing directions.

3.2.3 Literature Extraction

Articles retrieved by database queries, backward snowballing, and forward snowballing were screened manually. The hybrid workflow began with a start set derived mainly from Scopus; complementary Google Scholar queries supplied additional, non-overlapping studies. Each newly identified article entered the set for the next snowballing iteration, ensuring that every eligible paper was both screened and, where appropriate, analyzed in subsequent rounds. In total 202 publications were retrieved; the dataset of the papers can be found in Appendix A.2.

3.2.4 Qualitative Analysis

The qualitative content analysis (QCA) of the extracted papers followed a deductive category-allocation based on Mayring et al. [74] using their QCA software QCAmap. Full text papers were converted to machine-readable format coded in the QCAmap environment based on six predefined categories. Coding rules, and the detailed category definitions are documented in Appendix A.3. Empirical findings for each category are discussed in Section 4.1.

3.2.5 Deviations from the Planned Protocol

The deviation that altered the research design during the procedure can be summarized as follows.

- The initial search term “Bank” was replaced by “Financial Bank” (Table 1) after test runs indicated excessive noise.
- After establishing the start set, backward snowballing was simplified: reference lists were inspected selectively, as described in Section 3.2.2, to balance comprehensiveness with timely results.
- One study provided its core contribution in a supplementary document; that supplement rather than the main article was imported into QCAmap for coding.

3.2.6 Bibliographic Analysis

To analyze the papers found quantitatively, a meta-analysis of the authors and journals helps paint a picture about the landscape of the papers found in this review. The analysis follows the ideas of Silva et al. [75], due to the similarity of papers and the research field. Some features of the bibliometric analysis have been conducted using bibliometrix [76].

In Figure 1, it becomes obvious that the publishing vehicles varied greatly, with nine journals concentrating 48 of the 202 articles (24%). Most journals ("Others") only published one to two article. Even though financial network

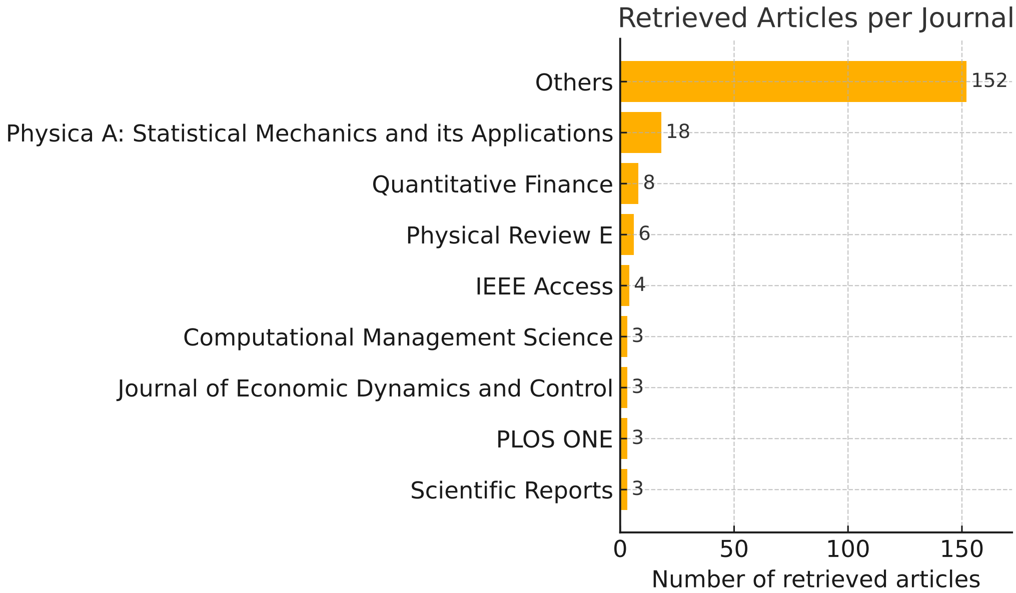


Figure 1: Number of Articles per Journal

research started in the 1980s, most of the articles discussed have been published since the 2010s, with a spike in 2016 and 2021, as seen in Figure 2. It is possible, that the database search results rank newer papers higher to some extent.

In total, 202 documents were collected, from 131 sources, in a publishing time span from the 1960s to 2023, as can be seen in Table 2.

Table 2: Main Bibliometric Information about the Collected Articles

Timespan	1960 : 2023
Sources (Journals, Publishers, etc)	131
Documents (Papers, publications)	202
Annual Growth Rate	3.14
Document Average Age (Years)	10

The papers were published by 521 authors, the vast majority appearing only once, as seen in Table 3. A small amount of papers are single-authored. Although German was a retrievable language, all the analyzed papers are English.

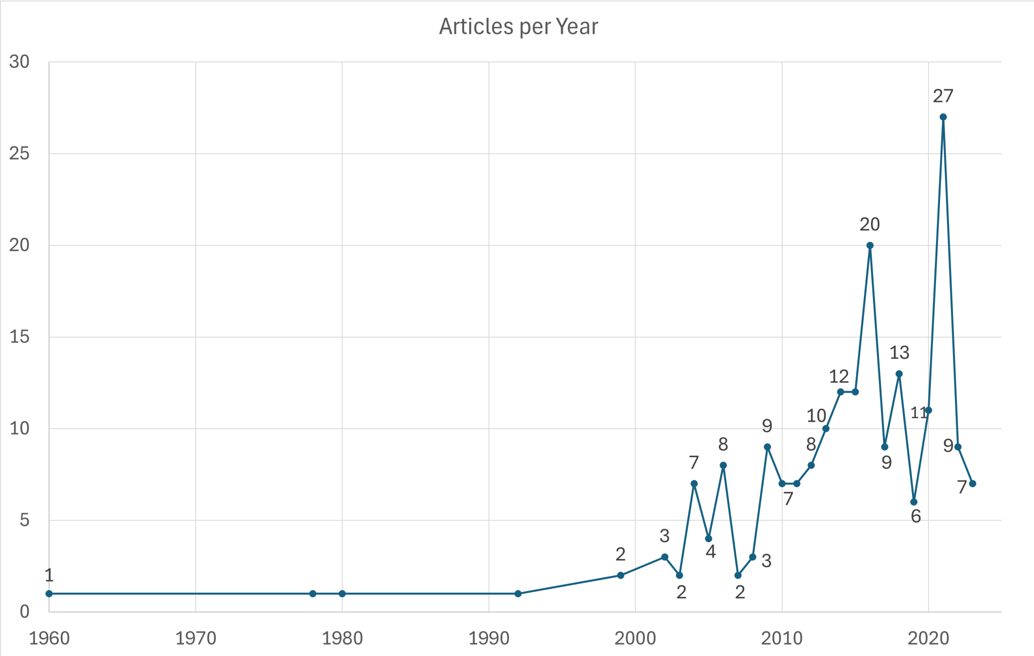


Figure 2: Number of Articles per Year

Table 3: Bibliometric Information about the Authors

Authors	521
Author Appearances	620
Authors of single-authored docs	18
Documents per Author	0.388
Co-Authors per Doc	3.07

3.3 Representativeness of Financial Networks

Representativeness is considered at two levels: (i) an empirical financial network must constitute an informative proxy for the underlying financial market, and (ii) a synthetic network must replicate the empirical network's important structural properties with sufficient closeness.

Representing the financial market Taking different network construction methods into consideration, the question arises what the necessary attributes are, that a constructed real financial network should possess. A real financial network is deemed representativeness of the broader market when it satisfies these principals and conditions:

- **Coverage of systemically important institutions**: The node set includes global or regional systemically important financial institutions (SIFIs), whose balance-sheet size and interconnectedness shape systemic risk [77].

- **Interpretable exposure**: Edges and weights capture exposure channels through which financial distress can propagate (e.g., committed loan amounts) [78].

- **Inclusion of core interdependencies**: Highly interconnected lenders and credit channels (exposure) appear explicitly, making analysis of contagion pathways possible [45, 42, 79].

- **Capacity for stress-testing insights**: The network topology permits counterfactual shock simulations or comparative statics, thereby informing stability assessments [80, 81].

- **Temporal plausibility (when applicable)**: If a temporal network is analyzed, structural shifts align with documented market conditions, under periods of financial stress [82, 83, 84].

Recent studies argue representativeness of financial networks can be validated via topological similarity, such as degree heterogeneity and distribution, centrality analysis, and weight representation [85, 86, 87, 88, 80].

Representativeness of the Data Among several candidate data sources, syndicated loan data provide a compelling empirical basis. Lender participation in large syndicates usually redistributes credit exposure across multiple banks but simultaneously creates overlapping correlated losses if the borrower defaults [89, 90]. Furthermore, syndicated lending is inherently cross-border, linking borrowers from emerging markets with global banks, which is usually not investigated due to missing data [91]. Frequent involvement of large, internationally active institutions suggests, that a network derived from syndicated loans can capture materially relevant exposures [92, 90, 93]. The 2024 estimated market size (measured in notational outstanding USD value) for syndicated loans is 9 trillion $, which is roughly 22% of the fixed income credit market at 41 trillion $ and about 11% of the size of the credit market including corporate bonds at 81 trillion $ [94]. Finally, syndicated loan participation among banking institutions has been shown to mirror interdependence patterns in other markets, such as derivatives and bonds [95].

Representing the real financial network A synthetic network will be deemed representative if it reproduces four categories of diagnostics:

- **Global size and density**: Node count, edge count, and link density fall within the empirical 95 % confidence intervals.
- **Baseline topological class**: High-level properties (e.g., small-world coefficient, heavy-tail exponent) match or improve upon basline graphs such as Erdős–Rényi, Watts–Strogatz, or Barabási–Albert.
- **Distributions of node-level metrics**: Degree, strength, and centrality distributions resemble those observed in the empirical network, acknowledging their role in shock propagation [96].
- **Model-specific goodness-of-fit**: Statistical generators are evaluated by in-sample fit or predictive error (e.g., simulation-based goodness-of-fit tests).

3.4 Network Construction Methodology

This section lays out the procedure used to construct the empirical interbank loan network and the synthetic benchmarks employed for comparative analysis. The workflow proceeds in three stages: (i) assembling the real lender-borrower network from syndicated loan data, (ii) deriving baseline networks that match selected summary statistics, and (iii) fitting Exponential Random Graph Models (ERGMs) to replicate higher-order dependency structures.

This framework of reconstructing financial networks using baseline graphs and likelihood models is motivated by the results of [97] and [98]. Baseline graphs can reproduce global density or small-world characteristics but assume dyad-independence, thereby missing higher-order dependencies such as reciprocity, clustering, and weighted heterogeneity. Among more flexible generators reviewed by Lee et al. [99], stochastic block models capture community structure, and latent-space models encode distance-driven transitivity; however, neither simultaneously matches the heavy-tailed degree, weighted core–periphery, and reciprocity patterns documented in real networks (such as syndicated loan networks). The ERGM framework [55, 52] permits explicit specification of these dependencies through sufficient statistics, and are extended to also model weighted networks [62]. Empirical studies of interbank networks show that ERGMs outperform dyad-independent models in reproducing systemic risk–relevant motifs, degree distributions and reciprocity [100, 83]. Accordingly, ERGMs are employed here to generate synthetic networks that adhere to the representativeness criteria while providing a statistical basis for goodness-of-fit testing.

3.4.1 Data Source

Global syndicated loan data were sourced from the *LSEG Loan Pricing Corporation (LPC)* database (see Appendix B.1). Only loans in which both the borrower and all listed lenders are financial institutions were retained; securities, derivatives, and corporate (non-financial) loans contained in the same package were excluded. For constructing a real financial network, syndicated loans are especially interesting, as the distribution of exposure of big financial institutions is the central idea of assessing financial stability.

Data Properties The extractable dataset contains syndicated loans issued between 1.1.2000 and 19.11.2024, with contractual maturities between 14.05.2024 and 17.02.2061. After initial cleaning, the extracted dataset contains 5149 distinct syndicated loans issued between 25.03.2003 and 19.11.2024. Figure 3 illustrates the four entities extracted from the raw file: *Borrower*, *Syndicated Loan*, *Lender*, and *Lender Participation*. All 29 original columns are shown in the schema to document data provenance, although only a subset is used in the network analysis (see Appendix B.2).

The lenders are from 75 countries, with the country-wise majority of loans issued in the United States 24%, Australia 19%, and Hong Kong 7%. However, region-wise, 44% of all loans are issued in the Asian region, 28% in the EMEA and 26% in North America. The borrowers are from 79 countries, the majority in the US 23%, Australia 17% and France 6%. All loans are also denominated in the US dollar.

After normalizing financial institution names, there are 1612 unique lenders, lending to 2197 unique borrowers. Among these, 302 loans exhibit a sole-lender structure. All 29 globally systemically important banks (GSIBs) identified by the Financial Stability Board (FSB) [101] (Table 4), are present as lenders or borrowers (see also Table 18 in Appendix B.3).

Possible Network Constructions Table 5 shows the possible graph projections that can be constructed from this data. Each projection is classified by node set, semantic interpretation of edges, and graph type. Only the *financial-institution* $\rightarrow$ *financial-institution* projection is pursued in the main analysis.

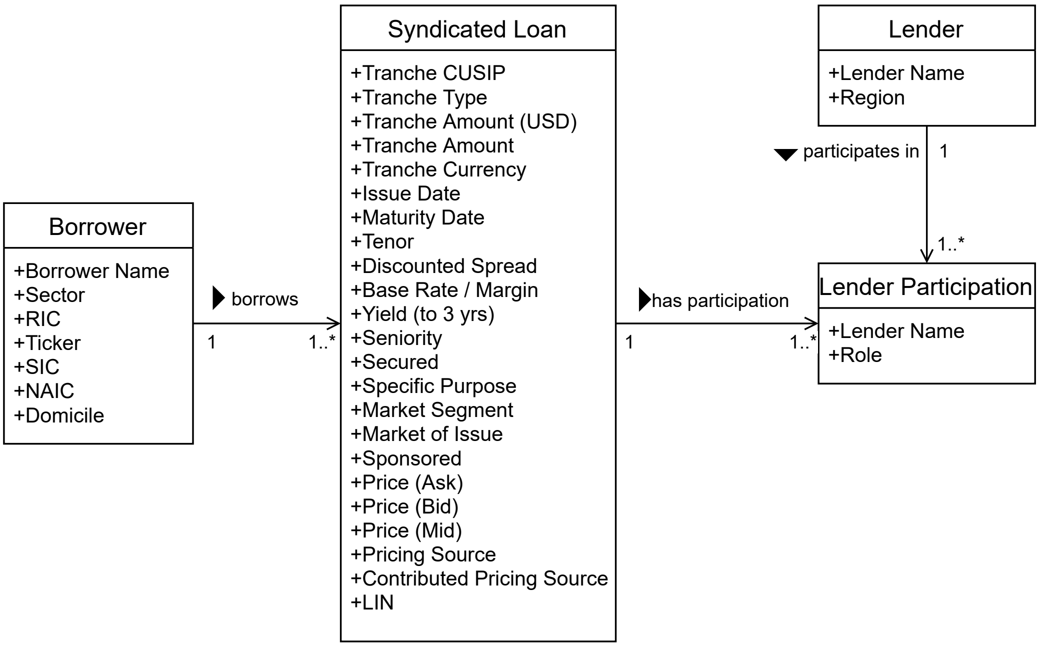


Figure 3: Raw Data Schema of Syndicated Loan Data

Table 4: Globally systemically important banks (G-SIBs) adapted from [101]

Bank	**Bank**
JP Morgan Chase	Citigroup
HSBC	Agricultural Bank of China
Bank of America	Bank of China
Barclays	BNP Paribas
China Construction Bank	Deutsche Bank
Goldman Sachs	Groupe Crédit Agricole
Industrial and Commercial Bank of China	MUFG (Mitsubishi UFJ FG)
UBS	BoCom (Bank of Communications)
Bank of New York Mellon	Groupe BPCE
ING	Mizuho FG
Morgan Stanley	Royal Bank of Canada
Santander	Société Générale
Standard Chartered	State Street
SMBC (Sumitomo Mitsui FG)	Toronto Dominion
Wells Fargo	

Order matches the 2024 FSB release (alphabetical within capital-buffer buckets). Bucket labels are omitted. Role frequency for each bank can be found in Appendix B.3

Table 5: Network projections derivable from the syndicated-loan schema (FI = Financial Institution)

Projection	**Node set(s)**	**Edge meaning**	**Graph type**
FI → FI exposure*	All lenders & borrowers	Directional USD exposure aggregated over tranches	Unipartite directed weighted
FI ↔ FI (co-participation)	All lenders & borrowers	Two institutions appear in the same tranche; count of joint commitments	Unipartite undirected weighted
Borrower ↔ Borrower (shared lenders)	Borrowers	Two borrowers share ≥1 lender; count of shared lenders	Unipartite undirected weighted
Loan ↔ FI	Loan IDs† & all institutions	Incidence of each institution in each loan	Bipartite undirected binary
FI (lender role) → FI (borrower role)‡	Same legal entities, duplicated by role	Credit exposure with roles kept disjoint	Bipartite directed weighted

* Projection used in the main analysis.
† 'LoanID' is a possible generated key for each syndicated loan during processing.
‡ Retains strict bipartiteness even when institutions act in both roles.
Note: Any weighted projection can be made binary by setting $weight = 1$ whenever the weight is positive and 0 otherwise.

Projection adopted for the analysis The study proceeds with the *financial-institution* $\rightarrow$ *financial-institution* projection, which is a unipartite, directed, weighted graph, where every vertex is a financial institution that can act as both lender and borrower. This choice is motivated by four considerations:

1. **Consistency to credit exposure:** edge weights aggregate all loan-level commitments between any ordered pair and therefore mirror the magnitude and direction of bilateral funding dependencies.

2. **Role symmetry:** some entities appear in both lender and borrower roles; enforcing a strict bipartition would either duplicate vertices or conceal cross-role activity. A single vertex set preserves this empirical fact.

3. **Relevance to systemic-risk questions:** the research studies how synthetic networks can be reconstructed in a representative way that would allow systemic risk among financial institutions to be analyzed; the selected projection maps precisely those liabilities.

4. **Completeness of coverage:** this projection retains every institution in the raw file, maximizing the size and connectedness of the empirical network.

Unused projections Although only the first projection feeds the subsequent analysis, the alternative graphs in Table 5 are reported for transparency. Listing them (i) contextualizes the chosen projection within the broader network construction possibilities, (ii) eases future robustness checks and replication studies, and (iii) illustrates that other projections were systematically considered before the scope was narrowed.

3.4.2 Constructing Full Financial Network

The adopted graph $G = (V, E, w)$ is a *directed, simple, weighted* network: vertices $V(G)$ represent financial institutions acting either as lenders or as borrowers; edges $E(G)$ encode loan exposures; and weights $w : E(G) \rightarrow \mathbb{R}_+$ equal the lender's USD commitment, across all syndicated loan edges linking the same ordered pair.

Because LPC's syndicated loans are reported at the loan level with the total value, each loan must be decomposed into lender–borrower dyads. As individual lender shares are absent, the approach in [102] is adopted: commitments are allocated according to the role structure described below.

Role Exposure in Syndicated Loans Syndicated loan contracts disclose the role of each lending bank (e.g., Mandated Lead Arranger, Participant, Agent) but not the exact percentage of the loan each bank keeps. This complicates loan analysis and is widely discussed in syndicated loan literature. Blickle et al. [102] survey three allocation rules commonly applied to approximate lender shares:

(a) **Equal-value allocation**: the full loan amount is assigned to every lender, overstating total exposures and double counting the borrower's liability.

(b) **Equal-count allocation**: the loan amount is divided by the number of lenders, ignoring the roles and potentially misstating individual lender risk in the event of borrower distress.

(c) **Loan-share prorating**: an observed (or estimated) percentage is allocated to the lead bank(s), and the residual estimated percentages are split among other participants. Blickle et al. [102] refine this option with a regression that predicts loan shares when sufficient loan-level disclosures are available.

The present dataset lacks the granular data needed to estimate the regression distribution proposed by [102]. Consequently, the loan-share prorating method (option c) is adopted in its simpler, literature-based form. First, reported roles are categorized as either "Lead", "Participant" and "Agent", consistent with previous studies.

Table 6, summarizes the reported lead shares and residual splits in the literature, while Table 7 converts those figures into the *base percentages* actually applied in this study. During network construction, each syndicated loan is first decomposed into lender-borrower dyads. The base percentages in Table 7 are then scaled by the number of lenders in the syndicate to prorate USD exposure across the participating financial institutions. A complete mapping between LPC role codes and the three groups is provided in the code repository (Appendix B.2).

Lender Share Percentage Using the data from Table 7, every lender receives a role specific base percentage share that is then adjusted for the syndicate size and applied to the loan.

Table 6: Syndicated Role Share

Articles (2009-2024)	Lead Lender Share (Reported in %)	Participant Shares (Reported in %)	Note
Sufi [51]	28		Lead % is average of three samples*
Ivashina [103]	27.17		
Maskara [104]	31.34		Lead % is average of two samples*
Beatty [105]	21.5		
Cai et al.[106]	31.37	20.847	Agent Share 11.679%
Chala [107]	30.01		
Bruche et al. [108]	7.5		Database used measures shares kept after selling the syndicated loan
Chi et al. [109]	37.478	5.689	
Schneider et al. [110]	24.98		

Table self compiled from literature sources.

* Sample refers to samples of different databases.

Table 7: Averaged base percentage each role receives as part of a syndicate

Syndication role	**Base percentage (%)**
Lead Arranger	28.812
Participant	13.268
Agent	11.679

Base percentage is the mean of the reported values for each role. Based on the literature collected in Table 6.

Lender Share Calculation

1. Definitions and Notation

   (a) **Set of Loans**: Let $\mathcal{L}$ be the set of all loans in your dataset.

   (b) **Set of Lenders**: For each loan $\ell \in \mathcal{L}$, let $L(\ell)$ be the set of lenders participating in loan ℓ.

   (c) **Role of a Lender**: Each lender $i \in L(\ell)$ has a "role" r_i, which belongs to a finite set of standardized roles,
   e.g., $\{\text{Lead Arranger}, \text{Agent}, \text{Participant}\}$.

   (d) **Total Loan Amount**: Denote by $V(\ell)$ the total principal (or "face value") of loan ℓ.

(e) **Base Percentage Function**: Define

$$b : \{\text{Roles}\} \to \mathbb{R}_{\geq 0}$$

such that $b(r)$ returns the *typical base percentage* for a given role r. (For example, $b(\text{Lead Arranger}) = 28.812$.)

2. Base Percentage Assignment

   For each loan ℓ and each lender $i \in L(\ell)$:

   (a) Identify the lender's standardized role r_i.

   (b) Assign the base percentage:

$$\text{BasePct}(i) = b(r_i).$$

3. Computing the Scale Factor

   (a) Sum of Base Percentages:

$$S(\ell) = \sum_{i \in L(\ell)} \text{BasePct}(i) = \sum_{i \in L(\ell)} b(r_i).$$

   (b) Scale Factor:

$$\alpha(\ell) = \frac{100}{S(\ell)}.$$

   This ensures that when multiplying each lender's base percentage by $\alpha(\ell)$ and sum over all lenders, the total becomes 100% for the loan ℓ.

   If the sum of base percentages $S(\ell)$ is not 100, scaling them by $\alpha(\ell)$ makes the final distribution add up to 100%.

4. Final Scaled Percentage per Lender

   Define the scaled percentage for each lender $i \in L(\ell)$ by:

$$\text{ScaledPct}(i) = \text{BasePct}(i) \times \alpha(\ell).$$

   Hence,

$$\text{ScaledPct}(i) = b(r_i) \times \frac{100}{S(\ell)}.$$

   With this definition,

$$\sum_{i \in L(\ell)} \text{ScaledPct}(i) = \sum_{i \in L(\ell)} b(r_i) \cdot \frac{100}{S(\ell)} = \left(\frac{100}{S(\ell)}\right) \sum_{i \in L(\ell)} b(r_i) = 100.$$

5. Assigning Dollar Amounts

   Finally, to compute the actual dollar share (or currency share) allocated to lender $i \in L(\ell)$:

$$\text{AssignedShare}(i) \;=\; \frac{\text{ScaledPct}(i)}{100} \;\times\; V(\ell).$$

   Substituting the previous expression:

$$\text{AssignedShare}(i) \;=\; \left(b(r_i) \times \frac{100}{S(\ell)} \Big/ \, 100\right) \times V(\ell) \;=\; \frac{b(r_i)}{S(\ell)} \times V(\ell).$$

   Hence, each lender i receives a fraction of the total loan $V(\ell)$, proportional to $b(r_i)$ relative to the sum of $b(r_j)$ for all $j \in L(\ell)$.

6. Special Case: Sole Lender

   If a particular loan ℓ has exactly one lender i whose standardized role is “Sole Lender,” a separate rule is defined:

$$\begin{aligned} \text{AssignedShare}(i) \;&=\; V(\ell), \quad \text{and} \\ \text{AssignedShare}(j) \;&=\; 0 \quad \text{for} \quad j \in L(\ell) \setminus \{i\}. \end{aligned}$$

   In other words, one lender takes 100% exposure, and no other participant has any share.

Lender Share Processing Example Tying everything together, the processing can be visualized in Figure 4.

1. **Original Data:** From the raw data, Lenders & Role, Tranche Amount in USD and Borrower are selected, and a Loan ID is added for processing.
2. **Lender Cleaning and role split:** The lender names are standardized, the roles are mapped into the categories of "Lead Arranger", "Agent", and "Participant".
3. **Lender Share Calculation:** Based on the base percentage for each role (see Table 7), and the defined lender share calculation, each lender’s part of the loan is calculated, representing their allocated exposure for this loan. Sole lenders would receive 100% of the loan amount, with other lenders in the syndicate receiving a symbolic weight of one to retain edge formation.

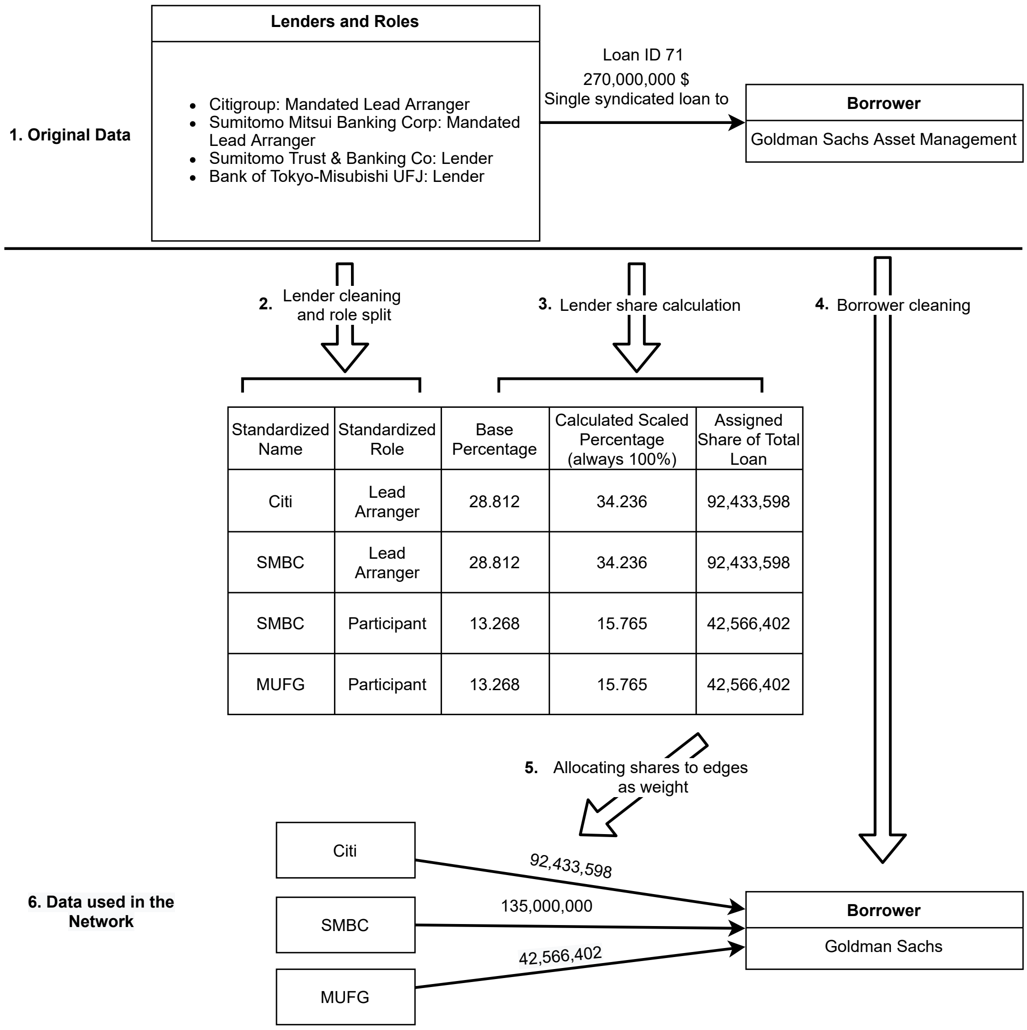


Figure 4: Syndication Loan Data Processing - Single Syndicated Loan Example

4. **Borrower Cleaning:** Borrower names are standardized.

5. **Allocating shares in the network:** The calculated share of the loan is added as a weight in the network. Note the example presented by Figure 4, the Sumitomo Mitsui Banking Corporation (SMBC) is part of two roles in this syndicated loan, so their calculated lender share is summed to reflect that.

6. **Data used for network construction:** The allocated loan shares act as weights between the financial institutions. During network construction all loan shares across all syndicated loans between SMBC and

Goldman Sachs (or any other node pair weight) is summed.

Lender Share Results There are 1613 unique lenders in the processed data. Among all syndicated loans, 1098 distinct lenders appear at least once in a syndicated loan with equal lender share. 1154 distinct lenders appear at least once in a syndicated loan with different lender shares. 173 distinct lenders are lending the full share in 588 syndicated loans (though their role in the raw data is not always "sole lender").

3.4.3 GSIB-W_2 Extraction

The syndicated loan graph contains $n = 3\,674$ vertices (see Section 5, giving $n(n-1) = 13.5$ million dyads, of which only 16 177 (0.12%) carry a positive weight (exist). Valued ERGM estimation relies on Markov-chain Monte Carlo (MCMC) algorithms whose computational cost scales with the total number of dyads rather than the number of observed edges (see 2.5). Studies report that the current implementation becomes prohibitive beyond $\approx$ 10 million binary dyads [111, 112, 113]). The unbounded implementation of valued ERGM models can further destabilize estimation in large search spaces [114, 115, 62], and evidence suggest weight variance might be the main cause of convergence issues in ERGMs [116].

Therefore, regulatory precedent is followed and a set of globally systemically important banks (GSIBs) identified by the FSB [101] is extracted. To ensure a single weakly connected component every intermediary vertex on at least one length-$d \leq 2$ shortest path between each GSIB pair is added.

Node Extraction To obtain a tractable yet contagion-relevant network, the following five-step "wiring" filter is applied to extract a weakly connected subgraph:

1. **Terminal set:** Each of the 29 GSIBs' vertex is flagged in the full graph.
2. **Distance metric:** For reachability analysis every edge is treated as one hop, ignoring the loan-amount attribute. This yields an *unweighted* geodesic distance d_{ij} between any two banks i and j.
3. **Wiring radius:** Exploratory distances show that all G-SIB pairs are reachable in at most two hops, while at least one pair is separated by $d_{ij} = 2$. Hence $k_{\max} = 2$ is the *minimum* radius that renders the G-SIB layer weakly connected; choosing $k_{\max} = 1$ would still leave isolated components, whereas larger k_{max} would add peripheral banks.

4. **Path-based vertex expansion:** For each unordered G-SIB pair (i, j) with $d_{ij} \leq k_{\max}$, every intermediary vertex on *one* length-d_{ij} shortest path is collected. The union of the 29 GSIBs and these intermediaries defines the vertex set $\mathcal{V}^*$.

5. **Induced sub-graph:** The sub-graph $G^* = G[\mathcal{V}^*]$ induced by $\mathcal{V}^*$ is extracted, preserving all observed edges among retained vertices and discarding edges incident to any vertex outside $\mathcal{V}^*$.

Focus on GSIB-W$_2$ The resulting induced subgraph (hereafter GSIB-W$_2$ network) contains 38 nodes and 241 directed weighted edges (see in Section 5.1.3, Figure 7)

GSIBs originate 54% of the total loan exposure, yet only 4% circulates among GSIBs themselves 5.1.3. Despite the small volume, GSIB vertices occupy 9 of the global top-10 weighted betweenness ranks (see Section 5.1.2). Contagion research and regulatory stress tests consistently identify bilateral exposures between GSIBs as key amplifiers of systemic risk [117, 118, 119]. Restricting the valued ERGM to GSIB-W$_2$ therefore targets the network core that concentrates contagion leverage while reducing dyad count well within current MCMC limits.

Limitations Inference pertains only to the GSIB-W$_2$ topology, while dependencies that involve non-GSIB institutions remain unmodeled. A discussion on inference is found in Section 6.2.

3.4.4 Generating Baseline Networks

Baseline networks provide a statistical frame of reference against which empirical and ERGM-generated graphs can be evaluated. Three canonical random-graph families are employed: Erdős–Rényi (ER), Watts–Strogatz (WS), and Barabási–Albert (BA) (see 2.3). Each captures a distinct "stylized fact" of complex networks, random connectivity, small-world clustering, and preferential attachment, while preserving analytical tractability.

Baselines for the full network All baseline graphs are generated with the same vertex count n as the empirical syndicated-loan network.

- **ER model**. An $\mathsf{ER}(n, p)$ graph is sampled where the edge probability p is set such that the expected number of undirected edges equals the empirical edge count m: $p = 2m/[n(n-1)]$.

- **WS model**. A small-world graph $\mathsf{WS}(n, k, \beta)$ is produced by first forming a ring lattice with average degree $k = \lfloor 2m/n \rfloor$ and then rewiring each edge with probability $\beta = 0.5$. The modest rewiring rate reproduces the short average path length typical of financial-exposure networks while retaining clustering.

- **BA model**. A preferential-attachment network $\mathsf{BA}(n, m_0, m_a)$ is grown from an initial seed of $m_0 = 4$ nodes by attaching $m_a = \lfloor 2m/n \rfloor / 2$ new edges per arriving node, matching the empirical mean degree within rounding error.

Because the intent is to isolate topological effects, all baselines are left unweighted; weighted diagnostics are reserved for the ERGM analysis, where edge strengths are explicitly modeled.

Baselines for the (GSIB–$\mathbf{W_2}$) The GSIB–W_2 core subgraph, consists of 38 institutions identified via the k-core procedure in Section 3.4.3. Separate ER, WS, and BA baselines are generated using the same parameter-matching scheme but with $n = n_{\text{GSIB–W}_2}$ and $m = m_{\text{GSIB–W}_2}$. Edge weights are similarly ignored at this stage; weighted statistics will be assessed in Section 5.3.

These baselines serve two purposes: (i) they establish whether the empirical graph exhibits statistically significant deviations from random or small-world behavior, and (ii) they provide a low-complexity benchmark for the representativeness tests in Section 5.2.

3.4.5 Fitting and Constructing ERGM Networks

The construction of synthetic networks using ERGMs follows the approach in [120, 121]. The syndicated loan network is modeled by a valued exponential–random-graph model (vERGM) whose probability–mass function is recalled for completeness (see Section 2.5),

$$P_\theta(Y = y) \;\propto\; \exp\{\theta^\top g(y)\},$$

with a geometric reference distribution [62], that accommodates the heavy-tailed, strictly positive loan amounts observed in the data (see Section 5.1.2. Edge weights are rescaled to millions of USD to improve numerical stability during MCMC sampling.

Candidate Statistic Families Four statistic classes are considered:

(i) **Incidence** terms (`nonzero`, `sum`) to separate the existence of edges from their monetary magnitude;

(ii) **Tail-increasing** operators (`greaterthan`) applied at empirically motivated monetary thresholds to better reproduce extreme exposures;

(iii) **Node-level covariates**, e.g. a binary indicator of "big lender", to capture heterogeneity in lender activity;

(iv) **Structural offsets**, notably a fixed negative coefficient on `mutual`, enforcing the theoretical expectation that bilateral exposures are discouraged in syndicated-loan markets.

Estimation Controls In the current `statnet` implementation of `ergm` [121], after maximum pseudo-likelihood estimation (MPLE) initialization, parameters are refined iteratively by Markov-chain monte carlo MCMC maximum likelihood. Pilot runs are used to select a burn-in period, thinning interval, and retained MCMC sample size such that (i) trace plots show stationarity, (ii) the Geweke diagnostic remains within ± 1.96, and importantly (iii) the Metropolis–Hastings acceptance rate falls within converging estimates [121, 116]. This approach is consistent with previous studies [98, 122, 83]. Chains are run in parallel with a fixed random-seed to ensure reproducibility. The final model settings are reported with the convergence diagnostics in Section 5.3.

Term–Selection Rule A forward–backward stepwise search of models is iterated until no degeneracy or convergence warnings are issued by `ergm`. A converged model is approximated for minimum Akaike information criterion (AIC), and terminated when the improvement between successive models satisfies $\Delta\text{AIC} < 200$, as the goodness-of-fit plots provide finer information on model fit, following the approach in [61].

Goodness-of-Fit Protocol The final specification is deemed acceptable only after simulating $m = 500$ synthetic networks under the fitted parameters (fixed seed), which is deemed sufficient to analyze valued ERGM statistics [61, 64] (see Section 2.5). Observed statistics for degree heterogeneity, and weight distribution, reciprocity are compared with their simulated counterparts via overlay plots and z-scores; a model is accepted when $|z| < 2$ for the desired statistic [61]. Detailed coefficient estimates, convergence diagnostics, and goodness-of-fit results are presented in Section 5.3. Supplementary code material of the fitted model can found in the Appendix B.2.

4 Literature Review Results

The literature review has been conducted based on the methodology designed in Section 3.2.2.

4.1 Literature Review

This review will first establish a basis of important review and survey papers that were found to provide an overview of network theory in a financial context. Detailed information on concepts frequently found in literature will be provided in five case categories. This gives six categories discussed in total. To reiterate, this review aims to answer the research questions relating to which methods are usually used to construct financial networks, and specifically, how synthetic financial networks are constructed, as specified in 3.1.

4.1.1 Overview of Review and Surveys

Among the chosen literature, there are a number of review and survey papers that aim to gather an overview of the research available at the time of publication. Surveys chosen offer context with respect to topics discussed in this thesis, specifically with the inherent problem of lack of empirical data, the issue of constructing a network, or the implications that can be drawn from one.

Reviews of Real Financial Networks A common theme in financial network literature and therefore also in review papers is the research of interbank networks. Summer [44] provided an overview of the methods for constructing such interbank networks 10 years ago. In his survey he discusses insights into the limitations of existing shock models, which often cannot explain the widespread cascade effects and their seemingly inherent amplification seen in financial crisis, while lacking quantitative evidence for interbank exposure. Importantly, he mentions the debate around the cause or effect of financial instability on recession, highlighting the lack of significant contribution to the interaction between financial exposure networks and the real economy. This is underlined by systemic risk research that uses financial networks that play an insignificant role in the economic analysis of financial crisis. These results can also be found in the comprehensive systemic risk survey of Silva et al. [75], whose analysis of that financial risk literature shows that only a small part of network papers are relevant. Emmert-Streib et al. [2] extended on this in a later survey, aiming to categorize the networks by node types.

They mainly discuss the core-periphery structure in a data-based approach to interbank exposure networks and expand the understanding of amplification mechanisms in an economic context. Similarly, Smolyak et al. [123] provide a review of complex interdependent networks in economics and finance. They conclude, that research not only qualitatively but also quantitatively suggests that the networks can correctly reconstruct the inner workings of institutions and differentiate the systemic behavior and therefore risk from that of its constituents.

Emmert-Streib et al. [9] aimed to analyze the elementary articles focusing on the structure of financial networks, summarizing various different stock and interbank networks. They call attention to the majorly fragmented state of network research across finance and economics in terms of research topics and applications, nonconformity of approaches, and diverse topics, the publishing split between various discipline and cross-discipline journals, and the slow adoption of financial network research in finance journals. They argue that improving the ability of researchers to follow up and learn from each other's results could lead to significant closure of gaps. This has also been mentioned in [2], stating that the corresponding fields reflect on their respective angels. The only commonalities between research papers on financial networks across disciplines were the agreement on the meaning of nodes and edges, and visual network exploration preceding quantitative analysis.

Furthermore, Nagurney [124] provides a more recent broad overview of all papers published in the *Networks* journal about economics and finance since its establishment; however, especially in the last decade they also showcase the differing theoretical and data-driven approaches and their varying conclusions about the underlying structure. Importantly, his review seemingly does not include many seminal papers found in other journals.

Reviews of Synthetic Networks From all systematically found research papers and reviews, Lim et al. [40] in 2016 chronologically produced the first survey of synthetically generated graphs, which relates to this analysis. Even if many parts deal with networks in general, specifically, their main findings of gaps are important to reiterate, as they hint at answers also asked as part of this research:

1. To select a generative model, graph fitting mechanisms remain a challenge, as there has been a minimal focus on constructing synthetic graphs based on empirical data. This could be overcome with developing a graph fitting mechanism for each existing graph generative model or developing new construction approaches that can fit synthetic graphs on empirical data.

2. The process of validating graph models is not unified, which makes graph comparison based on current similarity measures also not comparable. They suggested unifying similarity measures, using precision and recall in node correspondence, and cross-validating synthetic graphs with empirical graphs.

3. The scalability of synthetic graphs remains a challenge, and could be solved when graph generative models are scalable and parallelized.

4. Many approaches construct general graphs that may not even be related to the appropriate domain of financial network research, and to overcome this would require a benchmark on the networks on an application level.

5. Most methods for constructing graphs assume homogeneous graphs with single types of nodes or edges, sometimes inappropriately. Generating synthetic heterogeneous graphs could be a viable development.

Overall they note, the biggest challenges are the generative model selection mechanisms, as parameter search strategies and fitting criteria for evaluation are most often missing, combined with a lack of accurate fidelity measures for graph validation.

Assefa et al. [125] analyze synthetic data in finance. In their minor discussion on a network approach, they include random networks such as ER, BA and Kronecker graphs as synthetic, but also the emergence of the maximum-entropy methods. Overall, they refer their discussion to the works by Lim et al. [40] on synthetic graphs.

Lastly, and most recently graph neural networks have been a major interest of researchers discussing construction of financial networks. Wang et al. [126] provide a recent review of graph neural network in financial applications. They also list challenges in graph evaluation methods, the hard-to-compare construction between graphs without unified metrics for graphs. In addition, as the construction models employ neural networks, their structural information is often difficult to interpret, and they do not scale well, especially with large graph use cases common in the real world.

4.1.2 The Case of Random Networks

This section discusses research around random network models, such as the Erdős–Rényi (ER), Watts–Strogatz (WS), and Barabási–Albert (BA) models. As opposed to realistic synthetic networks, random networks generally do not aim to generate specific topologies of financial networks but rather use already existing implementations of random network models.

Stylized For a range of researchers constructing random networks, it is insignificant which type of financial network it represents, i.e. whether it is an interbank or stock market network is inconsequential to their research. These networks provide a basis for theoretical arguments. Exemplary, Denev [127] presents his own probabilities for edge connections in a highly simplified banking graph to construct a directed cyclic graph. Gai and Kapadia [117] construct a simplified network with a Poisson distribution, analyzing a theoretical contagion model. Hurd et al. [128] constructed a random directed assortative configuration graph explored earlier [129], which is based on the well-known random graph model of [130] and [131] to analyze a banking default cascade. Later, Hurd et al. [132] also employed percolation theory to generate a unique random network to analyze global cascades within a stylized banking network. However, their findings do not closely align quantitatively with those from Monte Carlo simulations of networks with finite sizes.

Random Interbank Networks Interbank Networks are often represented using the ER model. Even recent research implements them, such as Wang et al. [133] who use an ER model and a Hub network approach [134], to benchmark their risk spillover estimation model in financially correlated assets. Brummitt [135] also uses an ER model to simulate a debt network between banks, and adapts it to be a scale-free network with a heavy-tailed degree distribution to take earlier results into account [46] (see 4.1.4). Similarly [136] estimates assets and interbank loans between banks with the ER model, but also a small-world network following the algorithm of [137] and a scale-free following [138]. Steinbacher and Jagric [139] also have a similar approach, where they estimate the interbank network with an ER model, small-world and scale-free preferential attachment based on [28, 22, 140]. Amini and Minca [141] interested in default cascades, analyze, among other uniquely constructed random networks, a directed homogeneous random network based on the ER model, suggested earlier by [142]. The Credit Default Swap interbank market network is investigated by [143] generalizing the work of [144, 145, 146] and stylizing the network as a sparse ER network.

Generators Also interesting is the research around generators that are used to reconstruct specific generalizable network data. Researchers in this area build on already generated data or bring forth a new approach. In particular, the industry payment simulation network tool called the FNA payment simulator, used by researchers and banks to simulate interbank payments for Norway [147], the UK [148], and the US Fedwire payment network [149], is used to construct random scale-free networks based on the BA model [38].

4.1.3 The Case of Synthetic Data

When appropriate data is not available, researchers opt to reconstruct networks. Unlike the random networks discussed earlier, the following section provides insights into network construction methods outside of the often-used random network models.

Synthetic Interbank Networks Mastromatteo et al. [80] provided a seminal paper on the synthetic generation of interbank networks proposed a message-passing algorithm and showed that maximum-entropy techniques used by other authors to estimate interbank networks [150, 151, 152, 153] are not well suited to represent the risk found in real interbank markets.

O'Halloran and Nowaczyk [154] study systemic interbank risk and following empirical evidence for Pareto degree distribution [46], construct Pareto distribution sequences synthetically and then follow the erased configuration model to construct the graph first described in [155, 156].

Berardi and Tedeschi [157] simulate an interbank market with an endogenous preferential attachment mechanism based on bank liquidity and its interest rate. Their simulation results are in line with [13, 14] (see also 4.1.4). Other researchers have also proposed a similar approach for determining network topology with intrinsic non-topological properties "fitness" [158, 159].

Nier et al. [78] studying systemic risk, construct an interbank network by introducing stochastic balance sheet data on a base graph with edges created using the ER probability. Their results were later directly contrasted in [160], where an analytical approach to the random network provided a more complete explanation of findings. Sachs [161], on the other hand, constructed the balance sheet data first and then independently generated the liability matrices using the maximum entropy method. Her approach extends on Mistrulli [162] (see 4.1.6).

Von Landesberger et al. [163] wanting to visualize contagion in networks employ the approach of [46] calling it the "Moussa–Cont generator for financial graph structure", using a stylized banking network for their simulations.

Transaction generators Transaction networks are often used to analyze and detect fraudulent transactions such as in money laundering applications. Fraudulent transaction and anti-money laundering (AML) analysis is an ongoing field of research in which some studies propose the usage of network models. In terms of a purely synthetic analysis, this is interestingly commonly accompanied by what researchers call generators, to generate the transactions with prespecified injected fraud behavior. Exemplary, Lopez-Rojas et al. have propagated synthetic data for bank payments with BankSim [164] used by [165] to construct transaction networks for fraud detection. Tian et al. [166] propose their own synthetic graph generator for a large-scale payment network. Base networks are constructed using the BA model following preferential attachment, upon which transaction attributes, accounts, and fraud patterns follow, and then sampled for a dynamic payment graph. Using the GTGraph generator [167], Zhang et al. [168] create among other networks, a directed random network with small world properties, which allows one to create temporal edges for a financial context (they also use other non-relevant networks).

Zhang et al. [169] construct synthetic temporal financial transaction graphs to propose a new machine learning model to learn graph embeddings for AML purposes, similar to [170]. They generate a graph using the AMLSim1, which constructs a graph in two steps. First, data is generated based on degree distributions using the NetworkX python library [171] and then, built on the PaySim transaction generator proposed by Lopez-Rojas et al. [172], generate a distribution-based time series of transactions. AMLSim has also been further developed into AMLWorld by Altman et al. [173], which provides a range of benefits, among others, more robust money laundering patterns in the underlying dataset for the emerging graph.

Godase et al. [174] analyze a temporal synthetic financial fraud network based on the generator works of [175] and a skewed network based on [58].

Soltani et al. [176] also focused on an AML algorithm for graphs and used their own proprietary algorithm to create the synthetic transaction data necessary for the graph; however, no validation of the data or method for real networks is discussed.

A noticeable exception for not using generators, while creating synthetic fraudulent transactions is Zhou et al. [177], who create a transaction network by simulating a core-periphery Kronecker graph with a uniform distribution and injected fraud behavior (they also create other non-financial networks).

4.1.4 The Case of Real Data

There are cases where real data is available to construct a network. These are usually networks created from daily return of stocks, anonymized transaction data published by banks, or data that might represent the interbank network as a whole. Stock and transaction data are usually in a time series (temporal) format; however, the resulting network might be static or dynamic. The following category discusses the construction methods for these networks.

Interbank Interbank networks have been of interest since the beginning of financial network analysis in the early 1990s [178] with network construction from balance sheet data of the federal reserve. One of the most cited papers in financial networks is the Boss et al. [179] analysis of the Austrian interbank network structure. They find it to be closest to a small-world with a low degree of separation, but most importantly it exhibits a power-law distribution. Similarly, Cont et al. [46] findings of the Brazilian interbank network also suggest a heavy-tail distribution, and they note that is is overall similar to the findings of Austria [179]. Soramäki et al. [180] also find a scale-free distribution in the US interbank market, a high clustering coefficient, and the small-world property. Later, Bargigli et al. [15] suggested that in their analysis of the Italian interbank market, higher order topological properties differ from those of random networks and more advanced network models are needed.

Puliga et al. [181] construct an accounting network of bank assets based on balance sheet data using the cosine similarity transformation. Huang and Chen [182] construct the US interbank market based on balance sheet data of bilateral exposures as a bank correlation network. Interestingly, they use the PageRank algorithm [183] to determine node importance and placement, essentially assuming a core-periphery structure of the network. Similarly, Craig and von Peter [14] developed a core-periphery model that can fit the German bilateral interbank market. Fricke and Lux [13] also found a stronger core-periphery structure of the Italian e-MID interbank network but not scale-free or power-law. Indeed, the assumption and modeling of interbank networks along core-periphery estimations has been ongoing since 2000 with Borgatti and Everett [12], see also [184, 185] for the Italian e-MID, [186, 180] for the US interbank, and[187] for Japan. A European-wide interbank network can also be created from balance sheet bilateral exposure data [188] by extracting the portfolio values of sovereign bonds held by the banks, consequently creating a bond network with a bipartite structure. Most recently Hao et al. [189] had a similar approach for a Chinese interbank network, but they first created the bipartite structure based on [190] and then mapped commercial bank loan register data to the network. The Japanese interbank market was also estimated from balance sheet data by Kanno [191], using the RAS algorithm [192] and his results are in line with other estimates of interbank networks. A global temporal interbank market has been analyzed by [193], who use the e-MID interbank market and base much of their work on [194] by constructing the network from overnight loans and who find a non-random structure over an aggregated period. Finger et al. [3] later also revisited this idea and provide evidence suggesting that the network formation process in this temporal context is dependent on historical link formation and relationships of the banks. Giudici and Spelta [195] construct a global interbank network using the Bayesian graphical model for static and dynamic networks and suggest a partial correlation approach to construct the edges of the interbank network in both cases. Battiston et al. [196] estimate interbank exposure following an endogenous fitness model of [197] and along the lines of [184, 198, 97].

The Diebold–Yilmaz framework The framework [199, 200] can be used to create bilateral interaction banking networks. It is used by Ando et al. [201] to create a network of credit default swap risks between 17 countries. This approach to build a risk-spillover network, especially in a global credit default swap market, is based on [202] and [203], which both interestingly do not use real data, but what [201] argues is essentially a synthetic CDS spread network. The Diebold-Yilmaz framework [200] is also used by Demirer et al. [204] to construct a global static and dynamic stock volatility network of the top 150 publicly traded banks. They estimate the links using a modified method to construct a global network, circumventing a high-dimensionality problem, usually arising from global bank networks.

Stocks The approaches to build a network out of stock data are highly varied, consequent of the fact that often time series data is used, where stationarity and auto-correlation can majorly impact the causal inference of anything done with the data. Most often, researchers in financial stock networks are interested in analyzing temporal networks. Li et al. [205] provided a new method called PMIME to construct a directed temporal financial system out of Chinese time series stock data, while alleviating the limitations of bi-variate causality. Gao et al. [206] created a network based on the Granger causality of the time series data.

Deev and Lyócsa [207] connect nodes of European financial institutions, if there is a significant return on the stock, resulting in a cross-quantilogram network. This approach is similar to stock networks of the US and globally in [208, 209, 210, 211].

Nobi et al. [212] construct a temporal Korean stock market network from the cross-correlation coefficients of stock price time series, built on previous correlation threshold approaches [213, 214, 215] for stock networks globally, of the US and Turkey respectively. Correlation-based stock networks have also been constructed and studied in [216] and [217] for the SP500 price changes and [218] for the Chinese stock market, and they show, that the networks follow a power-law model, whereas [219] show for the US closing prices correlation network that it follows a scale-free distribution. Chu and Nadarajah [220] mentioned this for the UK correlation network, who largely follow the approach of the US-based stock correlation network that follows a power-law model [221].

Huang et al. [222] add to the literature on stock price correlation networks by suggesting the use of dynamic conditional correlation for the analysis of systemic risk in the Chinese stock market, and filtering the resulting graph using Planar Maximally Filtered Graph (PMFG). Jiang et al. [223] also study the Chinese stock market but build their network more elaborately. Their so-called multiplex network construction is layered, with a PMFG algorithm used as a filtering method (also used this way in [224, 225, 226]), of stock price volatility correlation, variance decomposition (based on the Diebold-Yilmaz framework [200]) and investor sentiment.

Kaylagin et al. [227], while analyzing optimal methods for constructing a stock correlation network, find that the best method is a sign similarity model [228] as opposed to a Pearson correlation [229]. As an outlier, Zhong et al. [230] propose to create a temporal stock correlation network for price prediction purposes, using the spearman correlation coefficient to achieve favorable results. This is similar to the approach found in Zhang et al. [231] who filter the graph with Minimum Spanning Trees (MST), based on the works of [232, 233].

Opposed to all the aforementioned researchers using correlations to construct stock networks, is Fiedor [234], who argues that the stock markets are inherently non-linear while correlation stock networks showcase linear dependencies. He proposes a novel way, incorporating a mutual information rate and finding networks different from correlation ones. He notes the overwhelming evidence for non-linearity in stock market, index returns and currency exchange rates, which are not captured using correlation networks.

Economic Mandel et al. [235] analyze a US production relationship network of public and private companies, and suggest that it follows a power-law distribution similar to earlier research [236, 237]. Similar results are found in a network of Chinese provinces by Wang et al. [238] in regards to a power-law for economic networks (multi-regional input–output tables) and more generally [239].

Transaction Real payment data of bank clients is rarely available, but [29] provides a payment network for Estonia, which they compare with scaling-free networks that exhibit small-world behavior and a power-law distribution. This is different from results found earlier [30], as the Danish payment network does not seem to indicate a power-law distribution.

Graph Neural Networks Graph neural networks being used in networks with underlying real data is a niche research area. Shumovskaia et al. [240] use a recurrent neural network to include more information in their temporal graph and solve the link prediction problem in large-scale graphs, specifically European bank client transaction graphs. Umaithanu et al. [241] use a graph neural network to learn transaction embeddings and predict the temporal edges in a PayPal transaction network, constructed using a similar bipartite approach to [242]. It is noteworthy, that all these papers discuss transaction networks. As an exception, Cheng et al. [243] propose a neural net that learns to estimate the temporal structure of a guaranteed-loan graph of companies.

4.1.5 Using Synthetic and Real as Benchmarks

This part discusses papers that test network analysis methods or innovative approaches on both synthetic and real networks. This stems from the fact that analysis of certain methods needs known structures ("ground truths") about the network that are known. In other words, certain endogenous and exogenous properties need to be present in the network before network analysis can provide significant results. Important to note is that often a paper may discuss a variety of multiple networks also outside of scope, so in accordance with the search strategy only the relevant ones for this summary will be mentioned. Focus is on the construction method of relevant networks and not on the network analysis methods, that the papers introduce and study. Researchers may try to keep the real and synthetic networks similar, but papers discussed here do not put a focus on it.

Interbank Amini et al. [244] studying interbank contagion construct, among their own random networks, static random networks using the preferential attachment method, and include a high degree of heterogeneity in their interbank networks, to keep them comparable to empirical findings of [179, 180, 46]. Amini et al. [245] also struggled to find complete data on interbank bilateral liabilities for cluster analysis purposes and created synthetic networks that follow a degree Pareto distribution, however they also construct some semi-synthetic networks using incomplete Korean central bank data using the Bayesian methodology proposed earlier by [246]. Moreover, they analyze clustering in reconstructed stock portfolio networks according to the Braverman and Mincas method [247]. Similarly, Marchese et al. [248]

create a range of random networks from a core-periphery structure of [12] (see also 4.1.4), to ER networks and also analyse financial networks of [184, 249].

Bardoscia et al. [250] analyze theoretical systemic risk in synthetic interbank networks created using the ER, scale-free and core-periphery model. Similar to [119] for Japan (see 4.1.4) they use the RAS algorithm to construct exposures of the European interbank market from balance sheet data. Barjašić et al. [251] simulate a range of ER networks, and later create a temporal Croatian creditor-debtor network for a default cascade analysis context. Türkmen et al. [252] have also explored finding an accurate representation of the underlying graph structure in the context of temporal interbank networks, where they employ multidimensional Hawkes processes [253, 254], which can be used to create synthetic networks, in the context of financial market correlations [255]. They compare their approach to empirical data by analyzing a small interbank currency exchange market, with a large scale yet to be validated.

Stocks Seabrook et al. [82] simulate financial networks using the ER model and construct real static and temporal country bilateral trade networks to analyze edge formation importance.

Huffner et al. [256] create a synthetic using the random intersection graph $G_{n,m,p}$ model [257] based on the ER model and a real graph of stock correlations based on Boginski et al.'s work [258] and [221] as discussed earlier (see 4.1.4). Similarly, Zhang et al. [259] create three synthetic graphs with ER random, small-world and BA scale-free model, and a real stock correlation network based on the approach of [260].

Cucuringu et al. [261] for their clustering approach, consider the stochastich block model (SBM) and BA model for their ground truth experiments. For their test on real networks, an SP500 correlation network and a currency exchange correlation network is used.

Natali et al. [262] analyzing temporal graphs, create a series of temporal synthetic graphs using signal processing [263] with varying distributions. For their real network, they propose to create an SP500 temporal stock network with a Gaussian distribution.

Jalaldoust et al. [264] analyzing causal processes in temporal networks, create random networks using a Bernoulli distribution for the creation and a decaying uniform distribution for subsequent sparse-graph scenarios. Their real networks are constructed from the default return volatility of the G7 bond data based on the approach of [204] (see also 4.1.4).

He et al. [265] propose a new way to estimate temporal centrality and create synthetic graphs using the SBM and compare their approach using an SP100 stock network with signals estimated from interest levels on key words.

Shafipour et al. [266] create a series of ER, small-world and scale free graphs, and a stock network of opening and closing prices as signals for their proposed algorithm to recover edges in the graph and reveal interdependencies between firms. It suggests topology similar to a small-world network, however, their results are not as robust for large graphs. Similarly, Saboksayr et al. [267] suggest new graph learning algorithms, create ground-truth networks using the ER model and a Gaussian distribution for the temporal signals, and validate their work by estimating the graph signals on a stock price closing network. Torri et al. [268] study the stress of European bank networks constructed using quantile graphical models expanded upon [269] and compare it with a European bank network constructed from time series stock return data. Saboksayr et al. [270] also estimate the temporal deviation to create a stock price network based on [271, 272], and create many static and temporal synthetic graphs under a Gaussian distribution using the ER and BA model, with varying underlying ground truths. Ying et al. [273] also analyze graph learning with ground-truth BA networks and a chain graph that follows a Gaussian degree distribution, and propose their method in a S&P500 stock network. She et al. [274] construct large-scale dynamic graphs using the Gaussian Graphical Model and then try to learn the underlying edges in a S&P500 and Nasdaq stock correlation network. Tugnait [275] employs an ER graph and a chain graph to analyze a novel graph learning method, and explores the dependency structure of a real graph, constructed from S&P100 stocks. As can be seen and in general, this approach and similar ones of finding the underlying dependencies in real stock networks after creating synthetic graphs with known ground-truths usually constructed from a Gaussian distribution is common in literature [276, 277, 278, 279].

Transaction Li et al. [280] build a synthetic network based on a Poisson distribution with hidden groups and compare it with a real transaction network based on the intensity of the transaction for their weight, but notice that this approach is only viable for small networks.

Huang et al. [281] build both a synthetic and real financial fraud network from the ICIJ Leaked Offshore Database with real financial entities, with the synthetic having limited size and distinct fraud injections, and the real constructed as a large graph with random fraud injections.

Ji et al. [282] propose a new way to detect cash-out fraud and build a real bipartite transaction network of Chinese bank customers and compare it to a randomly generated graph of similar size with injected cash-out fraud behavior and find similar results for both.

Yao et al. [283] build a synthetic and a real fund-raising transaction network using the PA method with a power-law distribution [284] generalized from the classic BA model. In the real network, this approach is used to construct the edges and capital flow, with a power-law distribution.

Harrigan et al. [285] create 50 ER and 50 small-world networks for testing suspicious transaction detection, and test their model on a simply constructed real peer-lending network for clustering analysis.

Tsigkanos et al. [286] suggest a novel guesstimation framework. The approach is based on incomplete real data that has missing features and values of firms and fuses expert domain knowledge to reconstruct static transaction networks, which they argue is desirable as the resulting network respects inherent constraints.

Zhang et al. [287], analyzing a new approach to graph mining, construct a series of static and dynamic temporal transaction graphs. Their baseline static graphs are based on the ER and BA models, while the temporal generation models are TagGen [288] and TGGAN [289], and their real graphs are based on readily available financial fraud transaction networks.

Graph Neural Network Balmaseda et al. [290] construct a range of synthetic and real networks. Though unclear, it can be assumed they employed the maximum-entropy approach with a minimum density estimation aiming for a heavy-tailed distribution of interbank liabilities. For their real network they use S&P Global Bank data for constructing the distribution of European interbank assets, and layering liabilities as weights.

4.1.6 Comparing Synthetic and Real Networks

This section discusses literature, where researchers first construct real or synthetic networks and then compare or attempt to validate it with the opposite. As mentioned earlier in reviews (see 4.1), similarity measures, comparison metrics, and validation of differing graphs remain a challenge, making these papers important to note but generalizing results complex.

Interbank Barucca and Lillo [291] suggested to use a SBM to generate a directed bipartite structure that best describes the Italian e-MID interbank market. Similarly, Ma and Mankad [292] base their real e-MID interbank network construction on [194], similar to [193] (see 4.1.4), and use it to validate a network generated with the stochastic block model and show that it can represent it accurately in a dynamic setting [293]. Dumitrescu et al. [294] include a transaction network of a Romanian bank and try to generate a synthetic network with the stochastic block model that fits the data under certain small-world assumptions in an AML context.

Under the assumption of maximum entropy, banks are considered to spread their interbank claims as uniformly as possible between their counterparties. Maximum entropy is a popular approach to reconstruct interbank networks [295, 296, 297], however, Anand et al. [298] have shown that networks constructed using the maximum entropy approach have undesirable topological properties and underestimate systemic risk, and they propose a minimum density approach. This is further confirmed by Mistrulli [162] who compares the contagion of networks using the maximum entropy method to Italian interbank data and finds that it is underestimating the scope of contagion for most parameters, while overestimating it for some constellations. This result can be significant, as it directly compares the maximum-entropy construction with empirical data.

Researchers have suggested to modify the maximum entropy approach [33, 299, 122] and constrain it, also known as exponential random graph models (ERG models or ERGMs). Engel et al. [100] discuss this approach in detail and reconstruct Italian and German interbank networks using a constrained exponential random graph model similar to the work of [122]. They compare their results with the real Italian [15] and German [300] bilateral exposure interbank networks and find (under specific unclear calibration) a desirable similarity in network statistics between the synthetic and real networks. Cimini et al. [98] analyze the boostrapping method proposed by [97, 301] to find topological properties endogenously with a fitness model in ERGMs. They concluded that it works well for World Trade networks [249] but not so well for the interbank market of [184].

Ezzodin et al. [302] created their own method for constructing networks and measuring the profitability and defaults of single financial groups with different holdings. However, they use interbank network data of [303] to "validate" their model, which limits their results, as these are not validly comparable networks.

4.2 Quantitative Summaries

A majority of the analyzed papers regard the lack or suboptimal availability of real-world data for networks to be the driving force in their attempt to find a method that results in a realistic representation of their desired network. Most of the literature retrieved has a focus on the interbank market, analyzing systemic risk or contagion stemming from the interconnectedness of financial institutions. Another major part are stock networks and industry transaction networks usually with a focus on fraud detection, while other economic networks are a smaller part of the research field.

4.3 Qualitative Summaries

Qualitatively, the literature gathered is from a wide range of authors and publications, and deals with a diverse set of challenges and solutions related to financial network generation. Reviews highlight the fragmented state of network research in finance and economics; however, some commonalities can be found.

The qualitative findings from this literature review with regard to the construction of synthetic and real graphs can be summarized as follows:

1. In recent literature, interbank networks are still often constructed using classical random network models (notably most common is the Erdős–Rényi (ER) model) to approximate or represent interbank networks. Otherwise, they are usually estimated first along a power-law distribution before being constructed either from the maximum-entropy method or stochastic block model (SBM). Some research suggests that synthetic interbank networks constructed from SBM [291, 292] or exponential random graph models (ERGMs) [100, 122] can have similar or equivalent characteristics compared with interbank networks constructed using empirical data. It should be noted again, as many authors point out, that similarity measures are hard to generalize, making comparisons across studies challenging. Real interbank networks are mostly constructed from balance sheet data [188, 189, 191], or if available, interbank transaction data such as the e-MID [15, 13, 184, 185, 193, 194, 291, 292, 162].

2. Stock networks are mainly created from some form of correlation network which results in very large networks, requiring researcher to filter the network using filtering processes such as minimum spanning trees (MST) or Planar Maximally Filtered Graph (PMFG). Most papers focusing on stock networks construct both random networks (ER, WS, BA) and real networks for method analysis and benchmarking purposes, usually in a temporal context, with minimal or no focus on the representativeness of synthetic networks to their real counterparts. In many instances, both the construction of models such as ER, WS, BA, as well as the naive construction of simple stock return correlation networks can be questioned later in the discussion, as there has been strong evidence against using either. This is moreover highlighted by the fact that not one paper in this review, attempts to validate a stock

network with any synthetic network or vice-versa but both are often used in combination to compare other network analysis methods.

3. Transaction networks are usually constructed for AML purposes, where synthetic ones are usually constructed using classic random network models, which research suggests can be sufficiently representative of real transaction networks in many cases.

4. Graph neural networks (GNN) are used for the construction of financial networks is a relatively recent endeavor and seems to be a niche research field. They can be used to learn transaction embeddings, aiding in the construction of temporal networks. Even though this literature search did not yield papers using GNN in the interbank context, a dedicated Scopus search for "interbank graph neural network" revealed 5 papers [304, 305, 306, 307, 308], all published after 2023 and using the same data-source, suggesting newly available data make GNN efforts in the interbank domain possible.

5. The criteria for determining what makes a synthetic network representative of a real one differ among research domains. Most often, network centrality measures, degree distribution and clustering properties are the main metrics for comparing synthetic and real networks.

Although there are several literature reviews [44, 75, 2, 9, 124, 40, 125, 126] regarding the general topic of financial networks and contagion in them, network construction of financial and economic networks and data has been lacking a review contribution which considers the broad spectrum of literature across domains and provides an overview. This review provides a broad but systematic analysis for construction of financial networks both real and synthetic, their comparison across domains, and their combined use cases.

4.4 Threats to Validity

Although, threats to internal validity of a qualitative literature review are restricted, biases can introduce threats to the external validity, mainly the selection of the presented papers and the drawn conclusions.

Researcher and Confirmation Bias Researcher bias or experimenter bias is the result of subjective expectation and feelings of the researcher influencing the research questions and design [309]. It is broad and can affect the study design on multiple fronts with specific selection of research question, and design methodology to favor expectations of the researcher.

Confirmation bias is the selection of search or information based on existing beliefs [310]. In the context of this literature review, these biases are highly similar as it translates to selection of methods, questions, and papers that could favor specific author groups, publications, or network construction methods.

To reduce research design bias, the guidelines of Wohlin et al. [68] provided a framework. The selection criteria for retrieval were pre-specified, and to diversify authors and publications, Scopus and Google Scholar were chosen as search databases. The design was chosen before any prior research about different methods for synthetic network construction had been conducted by the author. To mitigate confirmation bias during the manual search and selection of the results, no article was chosen purely based on the title, but through reading analysis of its relation to the topic, even if, relation to the topic seemed unlikely.

Limitations of the Criteria The decision of which articles to include, as mentioned, was closely guided by Wohlin et al. [68]. Nonetheless, there are limitations related to the implementation of this research. The search terms and study selection were closely guided by the research questions. However, the complexity of finding appropriate search terms for a database is also discussed widely outside of Wohlin's context, where most often the search terms should be iteratively optimized [311], and was done only once. The database search creates limitations in so far as the sorting of results by relevance may change the order of papers of the search if in the future Scopus or Google Scholar decide to change the relevance sorting algorithm. This does not affect the results presented in this thesis, but may limit the reproducibility of the search. This trade-off was taken because the keyword searches did not quantitatively reduce the search results on either database.

Furthermore, during retrieval, the inclusion of only financial-related papers might hinder the adoption of an approach used to construct a network in another field that would be innovative in the field of financial networks. There may also exist favorable methods or new approaches in other domains that would be suitable to at least investigate in a financial context. Finding methods that work cross-domain was out-of-scope, but could be investigated in future research.

Observational Bias Observational bias is related to the selection or perception of data to interpret the results [312]. For this systemic literature search, it relates to the manual analysis, selection and retrieval of the papers. The author may have had pre-conceived ideas about what constitutes methods for network construction or what kind of papers should be included. Because of manual judgment and reading, the collection of articles, as well as the summarized findings from the collected material, can be affected by observer bias.

To limit this, no research has been conducted prior in assessing the different construction methods and their viability, with the exception of basic terminology. Importantly, the implementation part of this thesis which investigates the methods was conducted after the collection of the literature, so expectation bias towards specific methods is limited. To mitigate the selection bias of the observational bias of papers even further, no region or timeframe exclusion was set, to diversify the results as much as possible. To mitigate qualitative bias in summarizing the papers, the QCAmap software has been used post selection to code the text parts which justifies their inclusion into the categories.

Reactivity and Publication Bias Reactivity bias, also called participant bias [313], is minimal for the results of this literature review, as it would apply to the authors of the articles, and who could not have reacted differently because of this study. However, a related bias in this context is publication bias, in which studies with positive results are more likely to be published. This would mean financial network construction papers with favorable results are more likely to be included in this selection. A way to mitigate it in this study was to select papers based on relation to the topic and not analyzing results of the paper during the selection phase. Moreover, as discussed in section 3, considering only peer-reviewed published articles in setting their guidelines might lead to publication bias, which can lead to systematic bias of the literature review [73]. To mitigate this, peer-reviewed journal articles and conference proceedings were considered for inclusion.

5 Network Results

5.1 Real Network descriptives

5.1.1 Observed Graph G_{obs}

The data processed according to 3.4 yields a graph G_{obs} with $N = 3\,719$ institutions linked by $E = 16\,200$ edges. Figure 5 shows the disconnected components, 45 institutions with 23 edges, most of which are sole lendings between a pair of institutes not connected to the rest of the syndicated loan network.

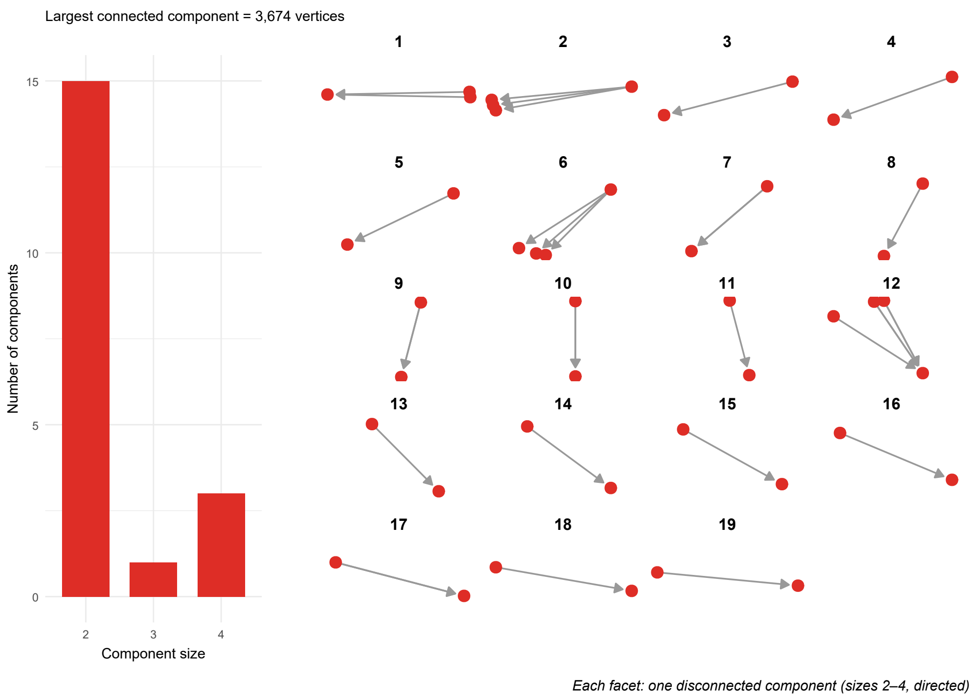


Figure 5: G_{obs} disconnected components
Note. Layout of the components of G_{obs} computed with the circle algorithm[314]; plotted with `ggplot2`[315].

For the subsequent analysis , only the largest connected component (LCC) is taken into consideration.

5.1.2 Largest Connected Component Graph $\mathrm{G}_{\mathrm{LCC}}$

Structural metrics The largest connected component of the observed network comprises $N = 3\,674$ institutions linked by $E = 16\,177$ edges representing total shared value in syndicated loans, forming a single weakly–connected component. Figure 6 shows the graph visualized with a concentric shell layout ordered by k-core (inner rings = higher coreness). Node size is proportional to degree. Edges are drawn with low opacity. The plot reveals a dense, highly embedded core with sparser peripheral shells and many edges into the core. Visually this represents a core-periphery structure.

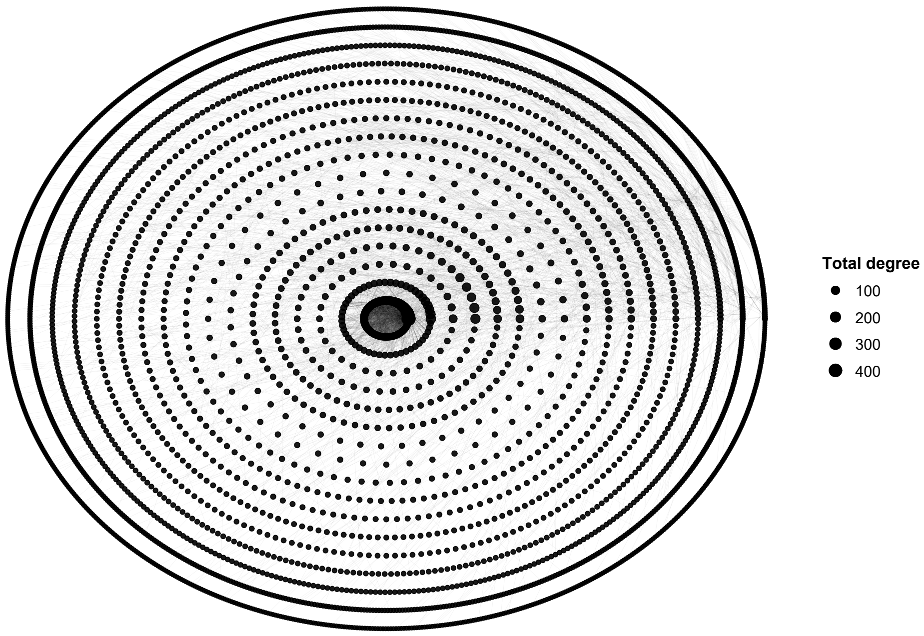


Figure 6: $\mathrm{G}_{\mathrm{LCC}}$ graph
Note. Rings calculated with the coreness algorithm (inner = higher k) using shell layout [314]; Total degree: node size encodes total degree. Edge directions are ignored for layout and drawn with low opacity; plotted with `ggplot2`.

The metrics in Table 8 present the binary structural results.

Despite an extremely low density ($\rho = 0.0012$) and therefore interconnectedness, information and exposures can traverse the graph in at most eight binary steps, with an average geodesic distance of $\langle \ell \rangle = 2.56$, indicating short average connections between a pair of financial institutions. Local cohesion is weak: global transitivity reaches only 0.043, reciprocity is effectively nil

(0.0004), and both edge- and vertex-connectivity are zero, indicating that individual edge or node removals is enough to disconnect the network in theory. Mild disassortative degree mixing ($r = -0.13$) further shows concentrated shortest paths through the hubs, and a slight trend towards lending in a hierarchical relationship (banks with dissimilar degrees interact).

Table 8: Headline structural metrics for $\mathrm{G_{LCC}}$.

Metric	Value
Nodes (N)	3 674
Edges (E)	16 177
Density (ρ)	0.0012
Diameter	8
Mean path length $\langle \ell \rangle$	2.5617
Transitivity	0.0430
Maximum degree $k_{\max}$	461
Skew ratio $k_{\max}/\langle k \rangle$	52.30
Assortativity r	-0.1302
Average betweenness	52.0

Degree and strength distributions are highly skewed: the maximum-to-mean strength ratio is $6.11 \times 10^{10}/7.91 \times 10^{8} \approx 77{:}1$. The most connected institution holds $k_{\max} = 461$ counterparties, 52 times the mean degree. The degree sequence exhibits a Gini coefficient of 0.669, and its upper tail follows a discrete power-law with exponent $\alpha = 2.34$ (xmin $= 14$, $p = 0.49$), confirming a heavy-tailed but finite-variance topology. The weighted degree (strength) vector has a Gini coefficient of $G_k = 0.802$, increasing to 0.954 when only considering out-degree (lending) alone, showing how only few large lenders dominate credit supply.

There are 1500 institutions (40%) with only out-degree (lenders) and 2083 (56,7%) with only in-degree (borrowers). Only 91 institutions (2,5%) both lend and borrow. Taken together, the ten largest lenders by weighted strength concentrate 30.84 % of total exposure in the system of USD 1453.262 billion (see Table 9), underscoring a highly concentrated credit landscape.

Weighted distance metrics Considering the USD face value of each exposure w_{ij} as an edge length produces long geodesics: $\mathrm{diam}(w) = \$4.7$ billion and $\langle \ell(w) \rangle = \$2.06 \times 10^{8}$. Because larger exposures facilitate rather than hinder contagion [45], the analysis below instead adopts the inverse-distance convention $\ell_{ij} = 1/w_{ij}$ (units USD^{-1}) for all distance-based measures (di-

Table 9: Top-10 largest bilateral exposures: lenders vs. borrowers (USD billion)

Highest-exposure banks (lenders)		**Highest-exposure borrowers**	
Institution	USD bn	Institution	USD bn
Bank of America	61.104	Blackstone	26.387
J. P. Morgan Chase	53.347	Saudi United Investment	20.000
SMBC	51.298	SoftBank	19.215
Bank of China	47.915	Bunge Limited Finance	15.215
MUFG	44.777	MOZ LNG1 Financing	14.902
Mizuho	43.206	Iberdrola Financiación	11.355
Citi	41.308	DoggerBank	10.562
HSBC	36.938	SHKP	10.258
BNP Paribas	35.546	BlackRock	10.150
Wells Fargo	32.712	Blue Owl	9.360
Subtotal	448.151	**Subtotal**	147.404

ameter, mean path length, betweenness, closeness). Considering this, the network collapses into an almost one-step structure:

$$\mathrm{diam}(1/w) = 5.1 \times 10^{-6}\ \mathrm{USD}^{-1}, \qquad \langle \ell(1/w) \rangle = 9.2 \times 10^{-8}\ \mathrm{USD}^{-1},$$

and shrinks six orders of magnitude relative to the binary edges. Closeness centrality is ≈ 0 for most nodes because the graph is only weakly connected; distances that traverse the LCC become infinite in directed space.

Centrality profile Table 10 lists the ten most central institutions under four centrality measures. Sumitomo Mitsui Banking Corporation (SMBC) is shown as the universal intermediary: it tops both binary degree ($k_{\max} = 461$) and inverse-weight betweenness ($g_{\max} = 5.9 \times 10^4$). All top ten in these two centrality measures are classified as GSIBs, with the exception of KEB Hana Bank. The availability of institutes with high betweenness centrality affects liquidity channels across the network.

Weighted eigenvector centrality based on Katz centrality [316] highlights Bunge Limited Finance, which borrows heavily (see Table 9) from already influential lenders. By contrast, PageRank elevates TÜRKIYE VARLIK FONU YÖNETIMI AŞ because PageRank redistributes influence along outgoing edges: nodes that only borrow act as terminal sinks for PageRank's random-walk process and therefore accumulate probability mass. Among the top five PageRank nodes (Türkiye Varlık Fonu, JHF Agency, Sonangol Finance, Qatar National Bank, and Ziraat Bankası) the first three are pure borrowers, showing a funding-dependence asymmetry.

Table 10: Top-10 institutions by four centrality measures in G_{LCC}

Rank	Degree*	Betweenness $(1/w)^{**}$	Eigenvector ***	PageRank
1	SMBC	SMBC	Bunge Limited Finance	Türkiye TWF
2	Bank of China	Citi	Saudi United Investment	JHF Agency
3	MUFG	HSBC	Rackspace Finance	Sonangol Finance
4	BNP Paribas	KEB Hana Bank	Blackstone	Qatar National Bank
5	Bank of America	Morgan Stanley	BlackRock	Ziraat Bankası
6	J. P. Morgan Chase	Mizuho	American Honda Finance	SMBC
7	Mizuho	MUFG	Corebridge Financial	Fosun International
8	HSBC	BNY	Schwab Funds	CPF Investment
9	Citi	Goldman Sachs	SBA Senior Finance	Mercuria
10	Crédit Agricole	Standard Bank Group	Perpetual	Africa Finance

* Degree centrality counts both incoming and outgoing edges.
** Betweenness was calculated on inverted edge weights $(1/w)$.
*** Eigenvector centrality computed based on Katz centrality [316].

GSIB footprint and systemic relevance The 29 designated GSIBs constitute $<1\%$ of all institutions yet originate 54 % of aggregate lending exposure (out-strength) and 27.4 % of total bilateral exposure (in + out) in the G_{LCC}. Yet, most of the exposure (96%) is not among GSIBs in the G_{LCC}, only 4% circulates among them. Therefore, they present the majority of access to syndicated loans for other borrowing institutions.

They mediate an even larger fraction of flow of exposure in the network: 73 % of inverse-weight betweenness, and they occupy 10 of the top 30 positions in that centrality ranking, showing they are among the highest intermediary institutions in the network. A cross-check with total degree metrics, 24 (binary degree) and 22 (weighted degree) of the top-30 nodes are GSIBs. Hence GSIBs dominate *size-based* centralities, while intermediary roles in the network are shared with a wider set of institutes.

5.1.3 GSIB-W_2 Subgraph

A 38-node subgraph (GSIB-W_2) comprises the 29 designated GSIBs plus nine nodes connected using $k_{\max} = 2$ (see Section 3.4), as shown in Figure 7. Tables 11 and 12 compare and contrast its descriptive metrics with those of the market-wide G_{LCC}.

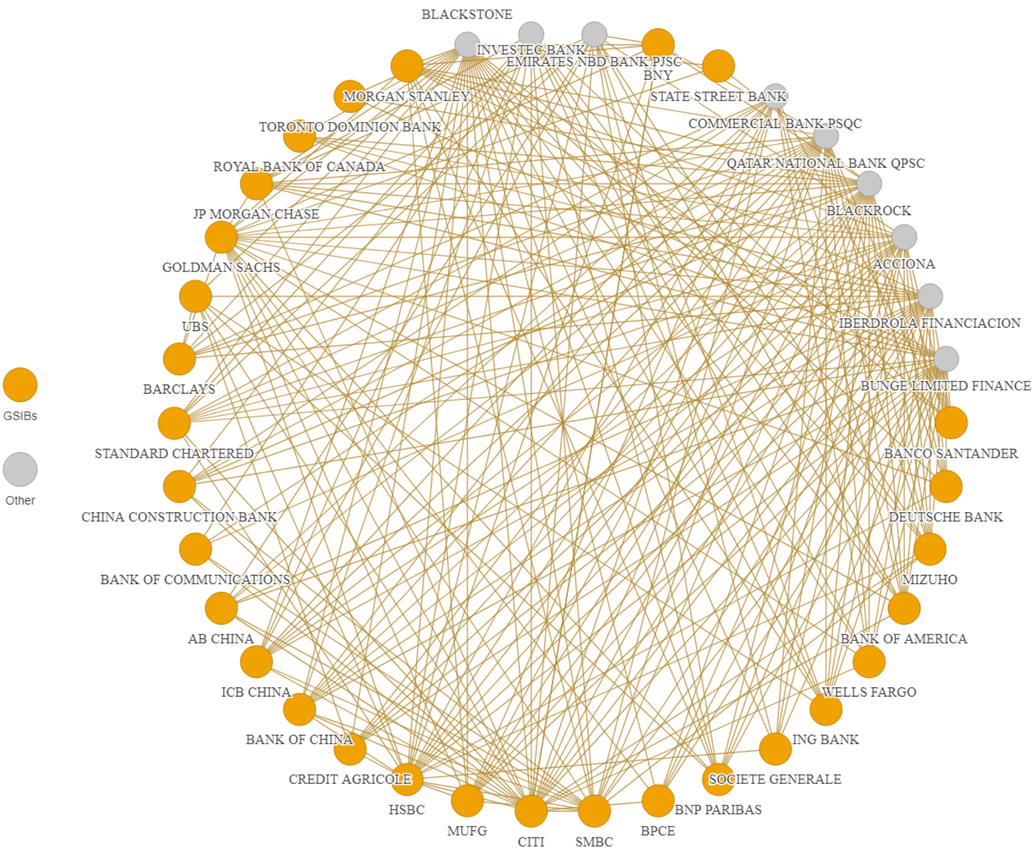


Figure 7: GSIB-W_2
Note. Layout computed with the circle algorithm; plotted with `visNetwork`[317]. Orange are the GSIBs, gray the additional k to make it weakly connected. An interactive html of the plot can be found in the repository (B.2).

Interpretation and systemic relevance GSIB-W_2 is two orders of magnitude denser than G_{LCC} and one order of magnitude more clustered; transitivity rises from 0.043 to 0.334. Binary geodesics shorten accordingly (diameter $8 \to 4$. mean path $2.56 \to 1.36$). The reciprocity is higher ($0.0004 \to 0.008$), but effectively only 2 edges are reciprocal in GSIB-W_2 compared to 6 in G_{LCC}. Using inverse weights as USD compresses both net-

Table 11: $\mathrm{G_{LCC}}$ network versus GSIB-$\mathrm{W_2}$ (descriptive metrics).

Metric	$\mathbf{G_{LCC}}$	**GSIB-$\mathbf{W_2}$**
Nodes	3 674	38
Edges	16 177	241
Density ρ	0.0012	0.1714
Diameter	8	4
Mean path* $\langle \ell \rangle$	2.5617	1.3556
Transitivity	0.0430	0.3344
$k_{\max}$	461	26
$k_{\max}/\langle k \rangle$	52.3	2.05
Assortativity r	−0.1302	−0.2016
Avg. betweenness**	52.0	3.37

* The normalized mean path for $\mathrm{G_{LCC}}$ is 0.036 and for GSIB-$\mathrm{W_2}$ is 0.069.
** The normalized avg. betweenness for $\mathrm{G_{LCC}}$ is < 0.000003 and for GSIB-$\mathrm{W_2}$ 0.0025.

works to near one-step networks, yet the core remains the more efficient (diameter$_{1/w}$ $= 6 \times 10^{-8}$ USD^{-1}). Strength metrics in Table 12 represent total lending + borrowing. Average weighted strength increases in GSIB-$\mathrm{W_2}$ ($3.3 bn vs $0.79 bn), while the maximum exposure is still $22 bn, indicating a heavy-tailed credit weights.

Within the GSIB-$\mathrm{W_2}$ the weight distribution remains skewed, although lesser: the total strength-Gini equals 0.49. The lender exposure inequality falls considerably, from $\mathrm{G_{LCC}}$ 0.95 → 0.52 for GSIB-$\mathrm{W_2}$ showing that GSIBs are more equal in their lending activity to each other than to other institu-

Table 12: Weighted metrics under the inverse-distance convention

Metric	$\mathbf{G_{LCC}}$	**GSIB-$\mathbf{W_2}$**
Diameter$_{1/w}$ (USD^{-1})	5.07×10^{-6}	6.02×10^{-8}
Mean path$_{1/w}$ (USD^{-1})	9.18×10^{-8}	1.27×10^{-8}
Max. strength (USD)	$61.10 bn	$22.00 bn
Mean strength $\bar{s}$ (USD)	$0.79 bn	$3.31 bn
Strength-Gini G_s	0.842	0.49
Weighted transitivity	0.043	0.3344
Weighted assortativity	0.0581	−0.0411
Avg. betweenness$_{1/w}$	55.81	4.62

tions. Borrowing inequality however, stays about the same 0.81 → 0.83.

5.2 Baseline Comparison

5.2.1 G_{LCC} Baselines

The empirical largest connected component (LCC) of the interbank exposure network, G_{LCC}, is benchmarked against three canonical random–graph families—Erdős–Rényi (ER), Watts–Strogatz (WS) and Barabási–Albert (BA). Each baseline graph is generated with $N = 3,674$ vertices and, where possible, an edge count matching the empirical density ρ. Because the baselines are unweighted, the metrics are all kept unweighted for consistency.

Headline descriptors Table 13 lists path-length, clustering and degree distribution metrics for G_{LCC} and its three baseline comparison models. A summary of the important contrasts is provided below.

Table 13: G_{LCC} network (N = 3 674) compared with ER, WS and BA baseline models.

Metric	Real	ER	WS	BA
Nodes	3 674	3 674	3 674	3 674
Edges	16 177	16 177	14 696	14 692
Density ρ	0.0012	0.0012	0.0011	0.0011
Diameter	8	14	13	11
Mean path $\langle \ell \rangle$	2.5617	5.6870	6.1217	3.1235
Transitivity	0.0430	0.0022	0.0110	0.2184
$k_{\max}$	461	21	16	2 941
$k_{\max}/\langle k \rangle$	52.3	2.4	2.0	367.7
Assortativity r	−0.1302	−0.0035	−0.0059	−0.1234
Avg. betweenness	52.0	16 768.5	18 247.8	59.9

Path length and small-world index The average geodesic length is $\langle \ell \rangle = 2.56$ with a diameter $D = 8$. Relative to density-matched nulls, the Humphries–Gurney small-world index [318] reaches $\sigma_{\mathrm{LCC}} = 17.1$ (ER: 1.0; WS: 4.7; BA: 4.8), confirming an ultra-small-world structure.

Clustering Transitivity equals $C = 0.043$, almost an order of magnitude above the ER yet well below the BA counterpart ($C = 0.218$). The empirical network therefore exhibits moderate average clustering without the excessive triangle formation induced by scale-free hubs.

Degree heterogeneity The ratio $k_{\max}/\langle k\rangle \approx 52$ signals a heavy-tailed degree sequence. A Gini coefficient of 0.67 exceeds ER (0.19), WS (0.17) and BA (0.45), while the heavy-tail exponent $\hat{\alpha} = 2.34$ is close to the BA estimate. Preferential attachment thus operates, but is tempered by constraints that curb extreme hub dominance.

Efficiency Global efficiency is $E_{\text{glob}} = 0.28$, higher than ER (0.26) and WS (0.25) yet below BA (0.39). Local efficiency reaches $E_{\text{loc}} = 0.77$, surpassing all three baselines and indicating abundant short alternate paths within neighborhoods.

Mixing pattern Assortativity is mildly negative ($r = -0.13$), midway between WS (-0.01) and BA (-0.12). Large lenders tend to connect to smaller ones rather than to peers.

Interim G_{LCC} combines scale-free path efficiency with moderate clustering and noticeable degree heterogeneity, placing it between WS and BA on the random–small-world–scale-free spectrum. The pattern is best described as a *strongly scale-free* topology [319].

5.2.2 GSIB-W_2 Baselines

GSIB-W_2, consists of 38 globally systemically important banks connected at density $\rho = 0.17$. Table 14 reports the same metric set for the core and three density-matched baselines (ER_{sub}, WS_{sub}, BA_{sub}). Additional tables are found in Appendix B.4. Because the baselines are unweighted, the metrics are all kept unweighted for consistency.

Path length The core exhibits mean distance $\langle \ell \rangle = 1.36$ and diameter 4, notably shorter than ER_{sub}(2.09) and WS_{sub}(2.20) and marginally shorter than BA_{sub} (1.58).

Clustering and small-worldness Transitivity is 0.29, close to ER_{sub} and WS_{sub} (both ≈ 0.32) and higher than BA_{sub} (0.21). Owing to the network's density, the small-world index for GSIB-W_2 is modest: $\sigma = 1.33$ (baseline ER_{sub}: 0.96).

Table 14: GSIB-W_2 (N = 38) compared with ER, WS and BA baseline models.

Metric	GSIB-W_2	ER_{sub}	WS_{sub}	BA_{sub}
Nodes	38	38	38	38
Edges	241	241	228	259
Density ρ	0.1714	0.1714	0.1622	0.1842
Diameter	4	4	6	4
Mean path $\langle \ell \rangle$	1.3556	2.0855	2.2013	1.5758
Transitivity	0.3344	0.3134	0.3331	0.7249
$k_{\max}$	26	23	18	191
$k_{\max}/\langle k \rangle$	2.05	1.81	1.50	14.02
Assortativity r	−0.2016	−0.0337	−0.1410	−0.4549
Avg. betweenness	3.37	39.11	44.45	2.50

Degree heterogeneity Degree inequality is moderate ($Gini = 0.29$; $\hat{\alpha} = 2.74$). ER_{sub} and WS_{sub} possess more equal tails, whereas BA_{sub} overshoots ($G = 0.45$; $\hat{\alpha} = 2.15$).

Efficiency Global efficiency is uniformly high because of density; $E_{glob} = 0.67$ versus 0.65–0.66 in the baselines. Local efficiency equals $E_{loc} = 0.73$, slightly above ER_{sub} and WS_{sub} and below BA_{sub} (0.91) (see Appendix B.4).

Interim GSIB-W_2 approximates a *dense small-world* configuration with restrained scale-free characteristics. Among the baselines, WS_{sub} captures clustering and efficiency most accurately, whereas BA $_{sub}$ best reflects path length; neither matches the empirical degree inequality.

5.3 ERGM of the GSIB-W_2

5.3.1 Model Specification

Estimation Environment & formula The model was estimated using a valued ERGM [62] implemented in the current version `ergm 4.9.0` [320] in `R 4.5` via `RStudio 2025.05.0`. A dedicated Windows 10 workstation (12-core Intel Xeon Gold-6154, 32 GB RAM) was used. End-to-end Monte Carlo maximum likelihood estimates (MCMLE) runtime for four parallel chains was $\approx$ **11h 20m**.

The full script is archived in the GitHub repository (Appendix B.2). A concise verbal representation of the fitted formula can be written as:

$GSIB\text{-}W_2 \sim$ `nonzero` + `sum` + `greaterthan`(400, 1000) + `nodeofactor(big_sender)` + `offset(mutual)`;
response: `w_scale` (edge weights, USD millions);
reference: geometric.

Coefficient Estimates Table 15 reports the MCMLE with standard errors and 95% confidence intervals.

Table 15: ERGM Results

	Monte Carlo Maximum Likelihood Results:			
	Estimate	Std. Error	z value	Pr(> |z|)
nonzero	−6.965	0.103	−67.389	< 1e-04***
sum	−0.009	0.001	−6.976	< 1e-04***
greaterthan.400	−2.074	0.272	−7.621	< 1e-04***
greaterthan.1000	−1.894	0.679	−2.789	0.00529**
nodeofactor.big_sender	0.008	0.001	6.476	< 1e-04***
offset(mutual.min)	−5.000	0.000	−∞	< 1e-04***
Null Deviance				0 (df = 1406)†
Residual Deviance				−13 544 (df = 1400)†
Akaike Inf. Crit.				−13 534
Bayesian Inf. Crit.				−13 507

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$
† df is degrees of freedom

Key interpretative points (log-odds scale unless noted):

- **`nonzero` ($\hat{\beta} \approx -6.97$):** Sets the baseline probability that any dyad carries a nonzero overnight exposure, capturing sparsity.
- **`sum` ($\hat{\beta} \approx -0.009$):** Each additional *USD 1 m* of aggregate outgoing volume apart from dyad-specific effects slightly reduces the odds of creating further ties (about 0.9% per million).
- **Tail increasing terms:** `greaterthan.400` ($\hat{\beta} \approx -2.07$) and `greaterthan.1000` ($\hat{\beta} \approx -1.89$) are a threshold that relax penalty from `nonzero` once cumulative dyadic weight crosses *USD 400 m* and *USD 1 bn*, respectively, allowing a heavy right tail while discouraging

unbounded exposures. Threshold numbers were approximated iteratively until model weight variance was within standard deviation of the weight variance in the GSIB-W_2.

- `nodeofactor("big_sender")` **(**$\hat{\beta} \approx 0.0082$**):** To approximate high out-degree lenders better, nodes carry a binary factor if out-degrees are higher than the mean. Each extra *USD 1 m* of outgoing volume increases the expected log-odds of sending another euro by about 0.008 (roughly 0.8%), capturing the degree heterogeneity.

- **Offset on** `mutual`(-5, fixed): Imposes a $\sim$150-fold reduction in the odds of reciprocal exposures ($e^{-5} \approx 0.0067$), to keep mutual ties bounded.

The model's AIC (-13534) and BIC (-13507) approximate an improvement over other tested model specifications (see repository in Appendix B.2.

MCMC diagnostics Four independent parallel Monte Carlo Markov chains (MCMC) were estimated and 14,250 draws per chain (total $\approx$ 57,000 post-burn-in samples) were retained. The overall *acceptance rate* was $\sim$**24%**, which sits in the usual 0.20–0.40 range for geometric-reference ERGMs [62]. The chains' trace plots fluctuate around stable means with no visible trends, and autocorrelation falls below $\sim$0.3 by lag 31,248, indicating successive MCMC sample draws are not overly dependent [60]. *Geweke tests*, which compare early and late parts of a chain to check that sampling has stabilized [115], are non-significant (joint p-values 0.29–0.84). Taken together, these indicators support that the sampler mixed well (model converged) and that the MCMLE estimates can be used for inference. The `mcmc.diagnostics` plot outputs are found in Appendix B.5.

5.3.2 Goodness-of-fit

Goodness-of-fit is assessed by simulating $n_{sim} = 500$ networks from the fitted ERGM (same nodes and weight scale).

Goodness-of-fit quantities are closely matched, as seen in Table 16: density ($Z = 0.43$), total system weight ($Z = 0.37$), the USD 1 bn tail count ($Z = 0.00$), and mean lender out-strength ($Z = 0.37$). The model underestimates mid-tail exposures at $\geq$ USD 400 m ($Z = 2.20$) and understates dispersion of lender strength ($Z = 1.58$). Exposure concentration in simulations is slightly underestimated (0.467 vs 0.522 observed in GSIB-W_2), with a considerable standard deviation. For completeness, the in-strength Gini

Table 16: Goodness-of-fit (valued ERGM): scalar statistics with simulation-based Z-scores.

Metric	Observed	Sim mean	Sim SD	Z
Density (nonzero)	0.171	0.168	0.007	0.43
Total weight (USDm)	62,858	60,669	5,860	0.37
Tail $\geq$ 400m (count)	37	26	5	**2.20**
Tail $\geq$ 1bn (count)	5	5	2	0.00
Reciprocity rate	0.008	0.000	0.000	—
Mean out-strength (USDm)	1,654.2	1,596.6	154.2	0.37
SD out-strength (USDm)	1,850.4	1,428.2	267.1	1.58
Gini (out-strength)	0.522	0.467	0.309	1.79
Gini (in-strength)	0.832	0.324	0.431	**11.79**
Gini (total-strength)	0.493	0.280	0.035	**6.05**

Z = (Observed − Sim mean) / Sim SD; values in **bold** indicate $|Z| \geq 2$ [321]. Reciprocity is constrained by an offset (-5), yielding zero variance in simulations; Z is therefore not defined.

is 0.324 vs 0.832 observed ($Z = 11.79$), which indicates the fitted specification smooths inequality across borrowers considerably, leading to the total-strength inequality being undervalued. Reciprocity is constrained by model coefficient (`offset`), and reported for transparency but is not assessed via Z. Higher offsets were tested, but increased mutual ties multi-fold, while GSIB-W_2 only has 2 reciprocal ties in the network.

5.3.3 Goodness-of-fit Plots

For each statistic the observed curve is plotted against pointwise 95% simulation boundaries, following the approach in [61]. Degree and geodesic distance are computed on binary graphs, geodesic uses an undirected projection to avoid infinite distances.

Degree distributions (probability mass functions) A probability mass function (PMF) for degree, gives the probability that a randomly chosen node has degree $k = 0, 1, 2, \ldots$, which allows assessing the distribution of the simulation against GSIB-W_2. The out-degree PMF in Figure 8 mostly matches at each observed k within the boundaries. The in-degree PMF shows a slightly heavier right tail than the simulation median (more high-inflow borrowers), likely due to the $> 50\%$ probability of no in-degree in the observed network.

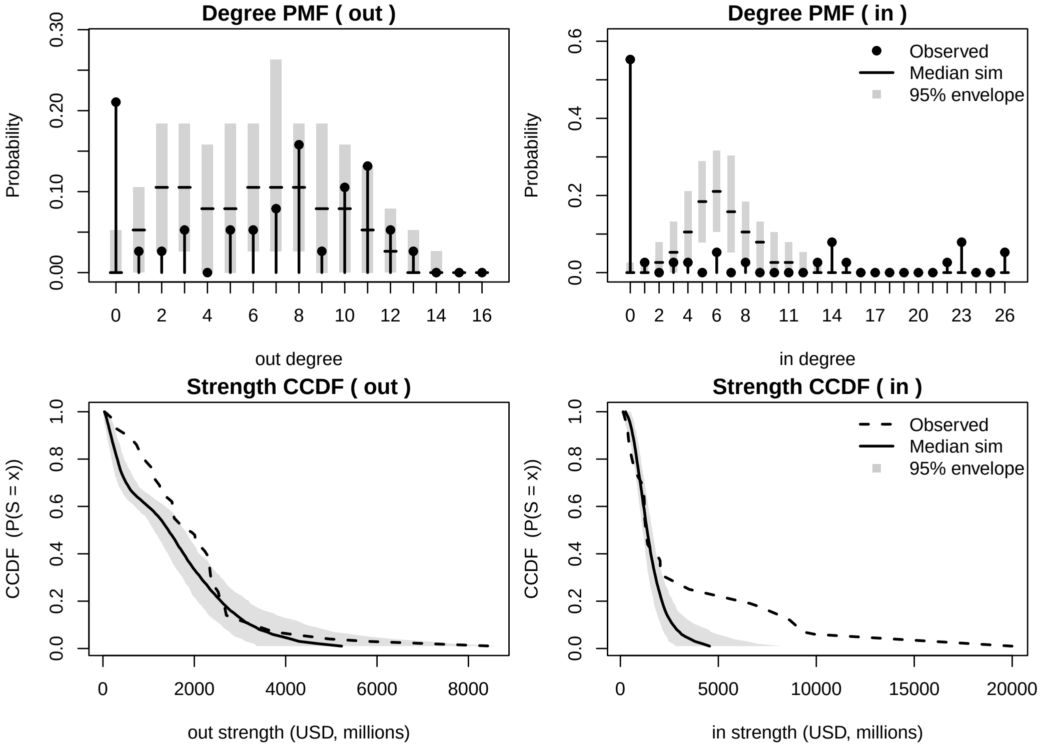


Figure 8: Goodness-of-fit: (top) out-/in-degree PMFs on the binarized graph; (bottom) out-/in-strength CCDFs on the weighted graph. Shaded regions are pointwise 95% envelopes from 500 simulations; solid lines are simulation medians; dashed lines are observed.

Strength distributions (complementary cumulative distribution functions) A complementary cumulative distribution function (CCDF) for a non-negative variable X is $P(X \geq x)$ [322]. Two CCDFs are plotted in Figure 8 as a function of x to highlight tail behavior: (i) out-strength, where X is total outgoing weight of a node; and (ii) in-strength, where X is total incoming weight. The out-strength CCDF tracks the simulation median closely, with a somewhat heavier mid-tail among big lenders around USD 1-3 bn. The in-strength CCDF sits above the envelopes across a wide range, indicating heavier borrower concentration in the data. This aligns with the large Gini Z-score in Table 16.

Geodesic distance (undirected, binary) Geodesic distance of undirected binary observed and simulated networks. Figure 9 shows the distance in GSIB-W_2 concentrates at two hops, with smaller mass at one and a thin tail at three; all observed bars lie within the simulation 95% boundaries.

This shows the model reproduces the small-world connectivity of GSIB-W_2.

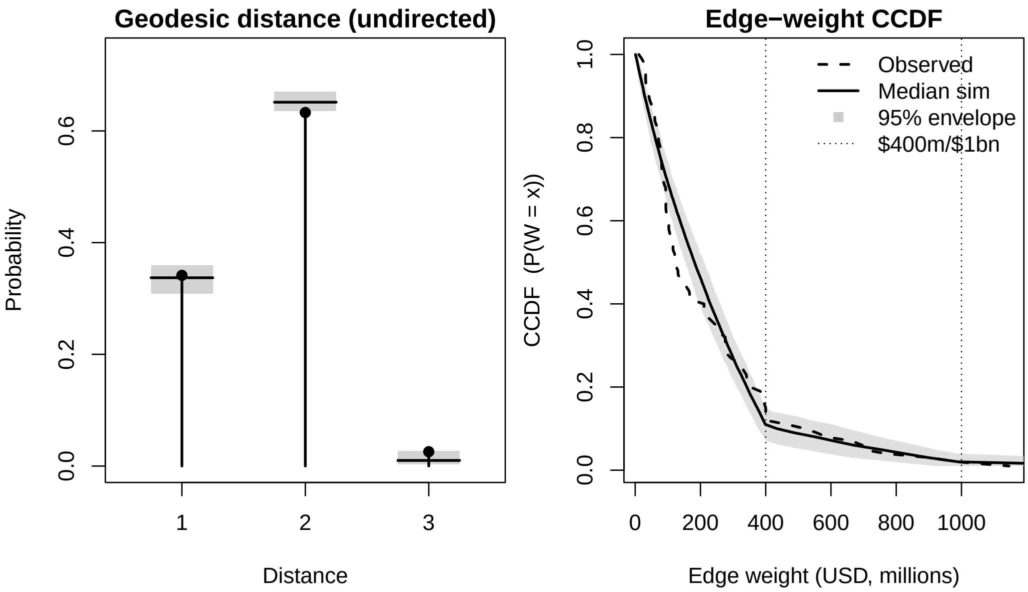


Figure 9: Goodness-of-fit: (left) geodesic distance distribution on the undirected projection of the binarized graph; (right) edge-weight CCDF with vertical markers at \$400m and \$1bn. Shaded areas are pointwise 95% simulation envelopes.

Edge-weight CCDF The model captures the overall tail of total edge weights well. There is a modest under-representation around USD 400 m (the observed CCDF lies above the envelope near that threshold), whereas the far right tail near USD 1 bn is matched closely—consistent with the scalar tail counts (Table 16).

5.3.4 Simulation Insights

Concentration of flows (Lorenz, Gini, top-k) Figure 10 contrasts the *Lorenz curves* [323] for out- and in-strength with pointwise 95% simulation envelopes. While the out-strength curve (left) sits close to the simulation median across the full range with a modest gap the in-strength curve (right) bends further from the line of equality than any typical network simulated, indicating that a very small set of institutions absorbs a disproportionately large share of incoming loans.

Figure 11 quantifies these patterns. For out-strength, the *Gini coefficient* of the data lies near the upper tail of the simulation distribution, and the

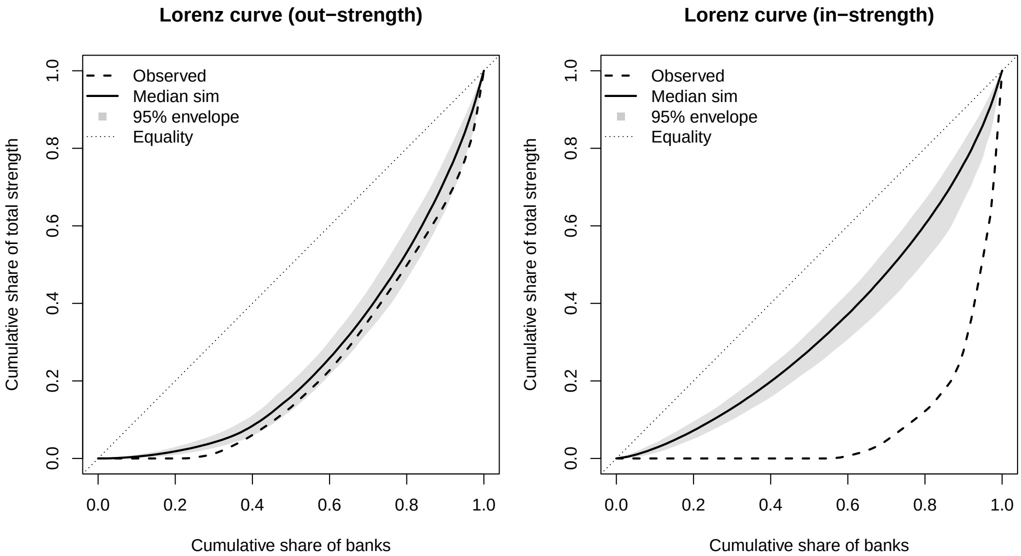


Figure 10: Lorenz curves for out-strength (left) and in-strength (right). Shaded bands are pointwise 95% envelopes from $n_{\text{sim}} = 500$ ERGM simulations; solid line is the simulation median; dashed line is observed; dotted line is equality.

observed *top-5* and *top-10* lender exposure compared to the total exposure in the network is only slightly above the simulated modes, showing mild but tolerable underestimation of lender inequality. For in-strength, however, all three indicators (Gini, top-5, top-10 receiver shares) place the observed network far to the right of the simulation histograms: borrower concentration in the data is much stronger than the model implies, in line with the goodness-of-fit findings.

Tail of bilateral exposures Figure 12 shows the simulated distributions of exceedance counts, so the number of edges at or above the model threshold of \$400m and \$1bn, with the observed counts marked as vertical lines. The model matches the far right tail well (observed $\geq$ \$1bn count near the simulation center), but it understates the mid-tail slightly around \$400m: the observed count lies beyond the upper simulation quantiles. Together with the in-strength results above, this indicates that the fitted specification captures extreme exposures but smooths the mass of "large but not extreme" loans and the inequality on the borrower side.

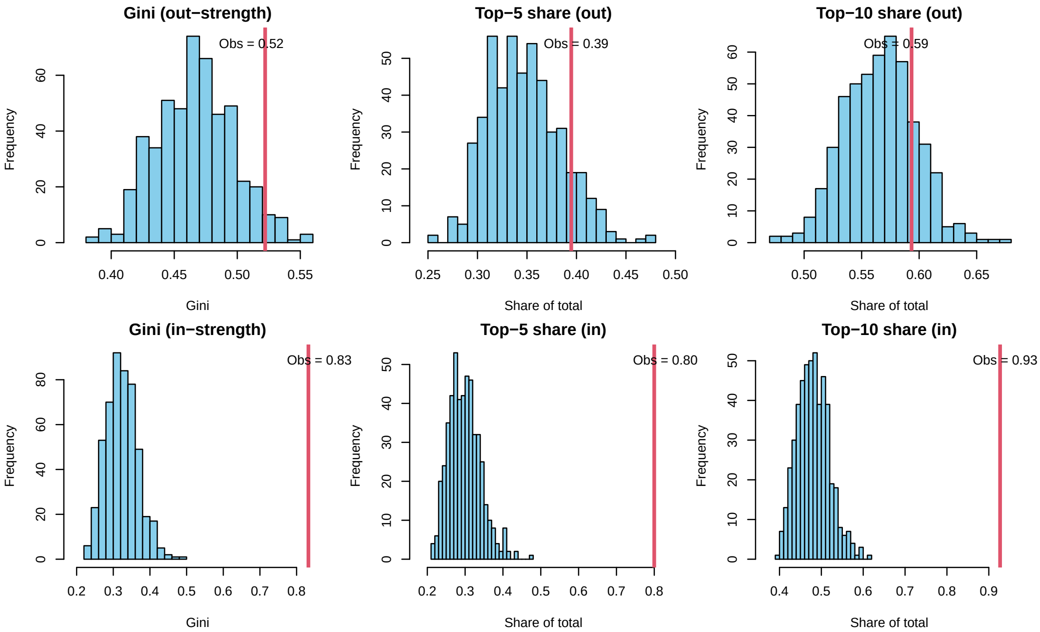


Figure 11: Concentration metrics across simulations with observed marked in red. Top row: out-strength Gini, top-5 share, top-10 share. Bottom row: the same for in-strength.
Gini: The x-axis is the Gini index of in- or out-strength in a given simulation ($0 =$ equal, $1 =$ maximally unequal); the y-axis is its frequency across the 500 simulations.
Top-k shares: For each simulation banks are ranked by strength (total in- or out-flow) and the fraction of total exposure by the top k banks is computed. The x-axis is that share; the y-axis is its frequency across simulations.

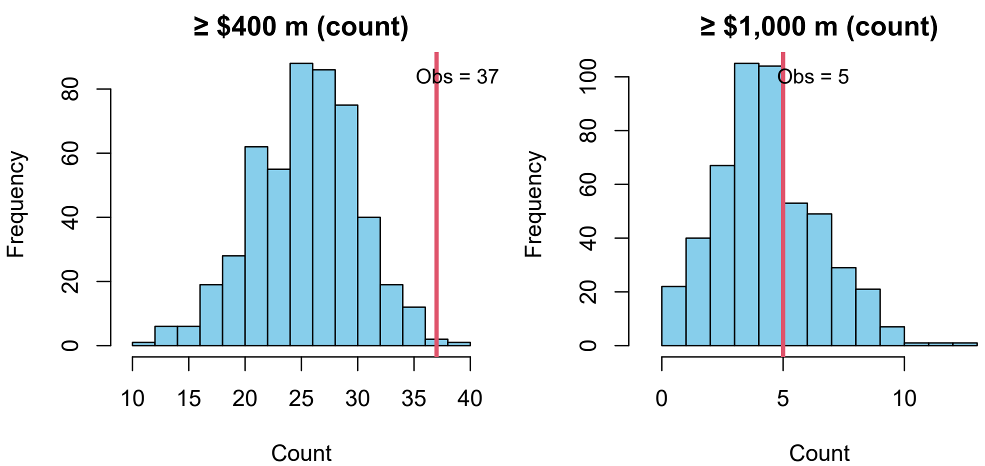


Figure 12: Simulation histograms of exceedance counts for edges $\geq$ \$400m (left) and $\geq$ \$1bn (right); vertical red lines mark the observed counts.
Counts: The x-axis is the number of directed edges whose weight is $\geq$ the stated threshold (e.g., $\geq$ \$400m); the y-axis is how many simulations produced that count.

Table 17: Baseline metrics: GSIB-W_2 vs. ERGM simulations and single-draw ER/WS/BA baselines (directed, binary graphs).

Metric	GSIB-W_2	ERGM mean	ERGM sd	ERGM Z	ER_{sub}	WS_{sub}	BA_{sub}
Nodes	38	38	0.000	—	38	38	38
Edges	241	236	10.000	—	241	228	259
Density	0.171	0.168	0.007	0.50	0.171	0.162	0.184
Diameter	4.000	4.564	0.557	-1.01	4.000	6.000	4.000
Mean path	1.356	2.191	0.059	-14.11	2.085	2.201	1.576
Transitivity	0.339	0.350	0.017	-0.66	0.313	0.333	0.725
$k_{\max}$	26	20	1.280	4.70	23	18	191
$k_{\max}/\bar{k}$	2.050	1.581	0.099	4.72	1.813	1.500	14.012
Assortativity	-0.202	-0.018	0.067	-2.72	-0.034	-0.141	—
Avg. betweenness	3.368	43.443	2.260	-17.73	39.105	44.447	2.500

$Z = (\text{Obs} - \overline{\text{ERGM}})/\text{sd}(\text{ERGM})$ based on ERGM simulations; "—" indicates not applicable.
Metrics computed on directed, binary graphs; transitivity and assortativity use the undirected skeleton for comparability with baselines.

Table 17 shows, the ERGM reproduces first-order scale (density) and clustering (transitivity) well, and its diameter is close ($|Z| \approx 1$). Where it falls short structurally is directed path length ($Z \approx -14$): simulations are markedly longer on average, consistent with the strong negative reciprocity offset reducing two-way shortcuts. The model also understates hub dominance: both k_{max} are smaller than observed ($Z \approx 4.7$), whereas the Barabási-Albert baseline overshoots (BA $k_{max} = 191$). For assortativity, the ERGM is too close to neutral (mean -0.018 vs observed -0.202; $Z \approx -2.7$), while Watts-Strogatz happens to be closer on this single metric. On balance, ERGM improves over ER/WS/BA by capturing realistic clustering and avoiding the extreme hub bias of BA, but it trades this for longer directed paths and milder degree disassortativity.

Interim Implication Structurally, the fitted ERGM reproduces (i) overall scale and lender-side concentration and (ii) the far-tail of exposures, but it *understates borrower inequality and mid-tail frequency.* To improve this model gap without increasing model complexity, a natural extension would be a receiver-heterogeneity term (e.g., an `nodeifactor` on a "big_receiver" indicator or a related in-strength variance-inflating term). This is left as a future exploration rather than part of the fitted model specification.

Relative to ER/WS/BA baselines, the ERGM provides the closest match on the structural metrics (Table 17): density and clustering are accurate ($|Z| \lesssim 1$), hubs (k_{max})) are calibrated without the extreme overshoot of BA, and the far-right tail of degree is more accurately reproduced than in the baseline networks. The main trade-off is longer directed mean paths (likely a consequence 0 mutual ties due to reciprocity offset); WS is occasionally nearer on degree assortativity, but at the cost of inaccurate degree heterogeneity and hub structure.

Interim ERGM Limitations The ERGM scope encodes group heterogeneity via `nodeofactor("big_sender")` but no institution-specific fixed effects. Within groups, nodes are exchangeable, so the model is not designed to reproduce the order names of institutions. Findings about "who is #1/#2" in simulated models is only possible descriptively.

Simulation diagnostics show heavier GSIB-W_2 in-strength concentration and more edges near the mid-tail ($\approx$ \$400m) than the ERGM generates. A plausible extension for future work is a receiver attribute or variance-inflating term to better match borrowing inequality.

The goodness-of-fit methodology is constrained in valued ERGM. The function `ergm::gof()` is not implemented for valued `ergm.count` models

[62]. Therefore a custom pipeline was used, analogue to [61]: scalar Z-scores, degree/strength/geodesic overlays, and edge-weight CCDFs. These results depend on (i) the binary rule for structure (edge exists iff weight > 0), and (ii) the quantile grid for CCDFs. All analysis and goodness-of-fit scripts are in the repository (Appendix B.2).

The MCMLE took ≈11 h 20 m on a dedicated workstation. Heavier model complexity scales roughly linearly with dyad count and may be constrained by available time and CPU. Parallelization of MCMC does not necessarily reduce runtime. Each parallel MCMC is a new chain to search the sample space, which has increased probabilistic not deterministic reduction of convergence time with additional overhead [320].

Results pertain to the GSIB-W_2 network. Inference and ability to generalize are discussed in Section 6.

5.4 Threats to Validity

Threats to validity for selection models and data differ from literature bias; the following threats are adapted from [324].

Experimenter and Confirmation Bias Experimenter (researcher) bias arises when a researcher's expectations influence the research questions or design [309]. Confirmation bias relates to data processing and selection. This risk is important here because preprocessing before model fitting was extensive; for example, financial institution names were normalized manually. Experimenter bias could also occur when choosing model parameters to favor expected outcomes. In this study, the risk is mitigated because the research question seeks the best-fitting model endogenously rather than a specific exogenous result. No hypothesis about particular systemic topologies was posited; instead, networks were simulated and compared using statistical error-reduction metrics, reducing the likelihood of unconscious model-selection bias.

Coverage Bias Coverage bias would arise if the data were unrepresentative, introducing structural bias into the network [324]. Syndicated loans capture a specific form of connectedness: they are neither over-the-counter (OTC), nor exchange-traded, depend on trust among parties, and syndication partners often repeat prior structures [325]. They also lack the strict reporting standards of derivatives markets, which for large banks can be larger and thus pose higher systemic risk [326]. Nonetheless, prior work reports correlation between the syndicated-loan and derivatives markets in participating institutions, size, and interconnectedness [327, 328]. A residual risk remains that unreported entities could materially affect network structure, although syndicated loans were chosen partly because major institutions are highly active in this market.

Sample Bias Sample bias arises when modeling data are not randomized [324]. This study fitted the ERGM on a relatively small sample, rather than the full G_{LCC} network, which may threaten the model's validity. However, as discussed in the design section, this trade-off was weighed against the study's objectives of analyzing a network model capable of reproducing important structural and weighted metrics of financial networks. Consequently, the financial ERGM may not generalize perfectly, but it yields interpretable, structure-consistent results.

6 Discussion

6.1 Literature Review

As part of the systematic literature review of financial network papers, a diverse range of papers has been analyzed. The papers were collected and described in six groups; namely, 1) reviews, 2) random networks, 3) synthetic networks, 4) real networks, 5) synthetic and real networks used as benchmarks, and 6) comparison of synthetic and real networks.

By undergoing a systematic literature review, this thesis puts literature of these categories in relation, and explores related concepts comprehensively.

Random networks, especially in the context of interbank networks often use the base models Erdős–Rényi (ER), Watts–Strogatz (WS), and Barabási–Albert (BA), with ER most often used. Apart from those baseline models, synthetic networks were analyzed for interbank networks and transaction networks. The most common approach to determine network structure is using endogenous properties of the network to determine topology, such as adapting the base models, using exponential random graph models (ERGMs), or unique configurations depending on the use case, such as AML injections into randomly generated transaction networks.

Studies using empiric data have been found constructing interbank, stocks, economic, and transaction networks. It can be seen, that the e-MID interbank data is frequently used for interbank network analysis, not least because its representativeness of the European money market has been stated repeatedly [329, 330]. It is also common to construct the network from balance sheet data. The topology of networks created from balance sheet data most often reportedly follow heavy-tail (or scale-free) degree distribution. Important to note is, that the detection of power-law distributions in empirical data is not trivial, and data gathering processes can introduce bias ([331] as found in [332, 19, 333]). Evidence in real economic networks seems to suggest they also follow a power-law topology. For transaction networks, there is contrasting evidence with regards to the power-law topology, which seems to differ based on countries. Transaction networks are also often analyzed and reconstructed using graph neural networks. Lastly, stock networks are appealing for network researchers, as data is readily and widely available. They are usually constructed using a form of correlation model.

Real networks and synthetic networks can be used in a single paper, for testing a network analysis method or an innovative approach. In this work, this kind of literature was referred to as benchmarking, as they benchmark a network method on different networks, most often against random networks with known network structures. In this instance, stock network highlight an interesting approach of the researchers with regards to any kind of comparison between stock network and the benchmarking model. When stock networks and some form of synthetic random network were used to benchmark network methods, there is no instance where researchers attempted to showcase representativeness in terms of comparing results or network structure between empiric and randomly generated network. Even though creating simple random networks and naive stock correlation was the most common approach to benchmark methods, because both of these are straight-forward procedurally, it opens the question as to the actual validity of the network methods proposed. This is compounded by the fact, that there is counter-evidence that correlation networks of stocks are not representative of stock markets, with regards to their linearity [234].

Lastly, papers found comparing synthetic and real networks directly were only analyzing interbank networks. It is noteworthy, that the maximum entropy assumption (uniform spread of interbank claims) showcased evidence of non-representativeness, and was rightly modified by researchers and extended into the ERGM. However, there has been contrasting evidence found for ERGMs ability to reconstruct network topologies of interbank networks. There was also an overarching theme when comparing synthetic networks to real ones, of non-uniformity in measuring representativeness of a network.

Interestingly, the papers found in this systematic search did not analyze the recent banking failure series [334] in the US in 2021/2022 either empirically or theoretically as part of a network analysis in any detail.

Principal findings Papers found in the systematic literature review discuss random networks with base models, synthetic networks, and real networks. These can also be used to benchmark a network method, or compare the network types respectively. Interbank network research and stock networks are a common field, while transaction, economic and graph neural networks are analyzed less.

Stylized and simple random network models are still used widely, either taking the model at face value or adapting it with different distributions. Synthetic networks are then unique methods that can also incorporate base random networks. Literature focused on synthetic networks in the interbank field and transaction networks. When comparing network methods on a real and synthetic network, base models were used by almost all papers to construct the synthetic network, usually without validating the blanket usage of the base model or methods used to construct the real network. This came especially apparent in the widespread usage of correlation stock networks for benchmarking network analysis techniques. The main field where synthetic and real network were validated against each others were interbank networks. These principal findings are in line with reviews such as [40] (see Section 7).

The reason and extent of data generation used when constructing financial networks has been discussed broadly, with data availability being the main driver for researchers to attempt other means of network construction. Data generation without underlying real data is done widely using theoretical models, and the results suggests that it may be inappropriately done in some instances. This answers the first research question posed, namely the reason and extend in which data generation is used when constructing financial networks.

It is noteworthy, that not a single papers in the literature review part of the thesis, provided empiric data (especially the claimed openly available interbank data) to reproduce the results. Most often, a reproduction packaged lacked entirely.

To answer the question regarding which common methods and techniques are used to generate and model financial networks, the systematic literature review provided broadly discussed, summarized insight into these methods, where results show various methods and often unique techniques used depending on the goal of the analysis.

Strengths and Weaknesses This systemic literature review provides, arguably for the first time, a broad but systematic analysis about the state of network research for constructing both real and synthetic networks, their comparison across different financial domains and their use cases. However, it is not possible to generalize these findings to other network research fields, as only financial networks have been analyzed.

Lessons Learned If the literature review is to be done again, the scope should be more well defined, so that search results of the database search are limited. This would then allow for an automatic extraction, as opposed to an intensive manual labor one. Instead of Google Scholar, Web of Science could be used, in combination with Scopus. This would also make the bibliometric analysis straightforward and more informative, as automatically extracted papers from those databases have meta-data, that can be read by bibliometric analysis software. However, categorizing and marking of relevant passages in the literature during full-text analysis is still favorable. While labor intensive, it made presenting and discussing the results simpler.

6.2 Network Discussion

For the network implementation section of this analysis, first a network G_{LCC} of loans between financial institutions (interbank) using real syndicated loans data was constructed. From this a network of global systemically important banks (GSIBs) GSIB-W_2 was extracted. To put the structure of the these networks into theoretical context, baseline Erdős-Rényi (ER), Watts-Strogatz (WS), and Barabási-Albert (BA) models were implemented. Both networks revealed a sparse topology, with only a few financial institutions holding the majority of exposure of the loans. Almost all of the loan connections went in one way and reciprocity was very low.

Results for ERGM The fitted valued ERGM based on the network GSIB-W_2 generated synthetic syndicated loan networks that are comparable with GSIB-W_2 along the metrics most relevant for risk. It reproduced first-order scale (density), global clustering, and the small-world backbone of the lending connections, while matching the extreme right tail of exposures (edges $\geq$ \$1bn) and weight distribution for outgoing loans. Distributional GOF overlayed showed observed degree and geodesic patterns lying within 95% boundaries of 500 simulations, and scalar diagnostics placed density, tail counts, and clustering within ≈ 1 standard deviation of the model's means.

Relative to Erdős-Rényi, Watts-Strogatz, and Barabási-Albert baselines, the ERGM provides the closest like-for-like match across the metrics: it shows clustering and hub structure without BA's extreme hub structure, and it captures exposure tails that ER and WS miss. WS is occasionally nearer on a single scalar (e.g., assortativity). Further, no baseline achieves the joint fit of structure *and* weights.

ERGM Model Limitations Two model limitations of the implementation studied are noteworthy. First, the 500 simulations showed that the model understated borrower-side inequality: the in-strength CCDF sat above the envelopes and the Gini coefficient is higher in the data, implying that a small set of borrowers concentrates more inflow than the model specification allows. Second, a structural implication is longer directed mean paths in simulations, which is a by-product of the reciprocity being so low in GSIB-W_2 and therefore the strong negative offset on reciprocity in the model that creates negligible probability for two-way shortcuts, despite a good match to undirected geodesics.

As a network model, ERGM's main limitation is it's ability to deal with large dyad counts, and unbounded weight implementation of valued ERGMs [62], as presented in Section 3.4.3. There are different solutions available, to overcome this ERGM limit for binary networks [111, 112]. Furthermore, there is active discussion about ERGM statistical inference of subgraphs remaining consistent with larger graphs [335]. However, there is no similar approach for valued ERGMs. Therefore, these results pertain only to the network of GSIBs (GSIB-W_2), and any risk-based analysis that might want to be done on with network model limits its ability to depict GSIBs to periphery bank connections.

From a modeling perspective, these results support ERGMs as a useful generator of ensembles of plausible financial networks: they encode interpretable mechanisms (sparsity, tail relaxation, degree heterogeneity) and yield results for system features (concentration, tail counts, reachability). The fitted model lacked in its reproduction of borrower-side concentration and the mid-tail; if one wished to tighten fit without increasing model complexity, a targeted borrower heterogeneity or variance-inflating term would be the natural next step. Otherwise, for the purposes of synthetic-network generation and uncertainty-aware stress analysis, the baseline ERGM is preferable to stylized baselines, provided it is either limited in size, or binary.

6.2.1 Comparison of the Real Network with Related Work

By constructing and analyzing a real syndicated loan network, the thesis relates to studies of real interbank networks, and the construction of networks using syndicated loans. Because syndicated loan networks were not explicit part of the literature review, a dedicated Scopus search for related syndicated loan network literature between 2014 and 2024, provides contextual information.

Champagne [325], analyzing similar syndicated loans but for more industries and non-banks shows similar but slightly lower distance network measures in almost all regards, more interconnectedness, and slightly higher clustering. Analogue to [325], small world properties are present in the observed network. Their possible explanation for the low reciprocity in the syndicated loan market, as well as findings in [336] fall short as an explanation to the low reciprocity in the observed network, because the network was constructed without excluding any nodes (banks) based on syndicated loan roles, loan regions or other node properties. It, therefore, remains an open topic of research in the syndicated loan market.

Harris et al. [337] analyze the connectedness in a syndicated loan network but construct their network based on additional balance sheet interest rates data, so result are not comparable.

Similar to Oh and Park [338], banks in the US syndicated loan network have been analyzed according to centrality measures and, in line with their result, exhibit skewed centrality measures in the observed network. They also find evidence of core-periphery behavior in the network, also suggested here. Their network results also show the importance of individual banks in line with results in this study, namely the importance of GSIBs and specifically the United States in the global syndicated loan market. This proposes the idea, that the US syndicated loan market is a subset of the global syndicated loan market, however, conclusiveness needs future research.

Even though contagion analysis was not in-scope, inference can be made based on previous results. There has been evidence, that suggests, more complete networks are less prone to contagion effects [88, 42]. Under this assumption, it can be argued that this observed network implies a high contagion in the financial system because it is very sparse, and the majority of exposure is in few nodes. Gupta et al. [339] corroborate these findings for a syndicated loan networks, showing it is a propagator of shocks reducing financial stability. The structure of the network with skewed exposure also falls in line with core-periphery assumptions made in previous studies about interbank networks, and their consequence of contagion (see section 4.1.4).

6.2.2 Comparison of the Synthetic Networks with Related Work

By creating a series of synthetic networks, this thesis adds to the knowledge on reconstructing interbank networks using ERGMs and analyzing the resulting topology and comparing it with the real network. The study of reconstructed interbank networks using ERGMs is a niche new field, with only a handful of published results, some of them discussed as part of the SLR, specifically Engel et al. [100] and Cimini et al. [98]. In related work there is contrasting evidence found with regards to ERGMs capability to represent interbank networks, which is in line with results of this study. On the one hand, the inadequate scaling of ERGMs to attain fitness on large networks, especially with valued ERGMs, makes representativeness problematic, as interbank networks are ideally modeled to contain as much information of the financial system as possible. On the other hand, with the limited scaling used in this thesis, the results of the ERGM fall in line with [100], where under certain parameters used, it can produce favorable results for interbank network. Similar to their model, the ERGM analyzed here is based on simple parameters, one of them capturing tie formation with difference in in- and out-degrees between nodes. Their suggestion that simple ERGM terms generalize well may hold true even outside of interbank data they used, as similar results were seen here with the syndicated loan network.

During the analysis of this thesis, a related paper was published by Macchiati et al. [83], which extend on the ERGM approach for temporal networks and encode density and reciprocity into the fitness model. It can be noted, that their model importance on fitting density, reciprocity and analysis of eigenvector value, especially in the real network is mirrored in this work. However, it can be argued due to the different data source, parameters used and the temporal network use-case, as well as the scaling down of the syndicated loan networks in this analysis, result comparison remains incompatible.

Lastly, it can be noted that the research in this thesis is unique, as it analyzes the ability to reconstruct the syndicated loan market using an ERGM and create synthetic syndicated loan networks.

6.2.3 Principal Findings

To answer the third and last research question on the representativeness of financial networks, criteria often found in literature have been explored and discussed. Then, to analysis specific representativeness of a reconstructed interbank network and provide empirical findings, a syndicated loan network has been implemented and an ERGM model was used to create synthetic networks.

The network construction part of this research fits topic-wise into the literature niche of comparing synthetic and real interbank networks. The findings show that the real network construction of the syndicated loan market as an indicator for an interbank network, is in line with previous findings. Most of the literature on interbank markets in general and specifically syndicated loan networks construct temporal networks, with analysis of risk contagion. As this analysis attempts a unique reconstruction of an interbank network using syndicated loans with an ERGM, comparing results to existing niche literature is limited.

Findings in the real network show similar results to literature with regards to topology and structural properties of the network. This shows that using different data sources and methods to create the real network, representativeness of the financial system might still be exhibited under differing circumstances.

The implementation of ERGM provided evidence into the representativeness of synthetic financial networks. It suggests that ERGM can reasonably reconstruct a syndicated loan network on a small scale and create synthetic networks, that are, on average and depending on the parameters used, fitting the real network. Findings of ERGMs used in literature to reconstruct financial interbank networks show similar results.

6.2.4 Strengths and Weaknesses

As the focus was on network construction and comparison, this network analysis does not include implementation of network analysis methods such as shock propagation, default cascade, or debt rank, to understand financial stability. Moreover, the data was not used to create a dynamic network, but rather shows a point in time network. This provides evidence of the structure of the debt network in its cumulative current state, but does not showcase information about link formation.

Data Limitations Using data of syndicated loans to create an interbank network assumes that they represent a significant portion of exposure of financial institutions. As mentioned, derivative exposure such as CDS, CLO, and CDO can at times be much larger than loans, especially for large banks [326]. However, as discussed, they often correlate with syndicated loans, not only in size depending on the bank, but also connection to other banks. This would mean, the syndicated loan market can be used as a meaningful indicator. Because of the correlation, the oscillation of instability would arguably be bigger but affect the network the same. The syndicated loan data was not chosen based on banks only, but also includes non-bank financial institutions. Some financial institutions might still be missing, if they do not engage in syndicated loans. The real network suggests, that most are represented, where analysis of under-representation of non-banks would need a separate analysis. The network of syndicated loans also does not capture short-term funding markets (such as direct interbank lending), even though they could provide vital liquidity during instability, and also does not capture granular transactions in the network.

To improve representativeness, syndicated loans could be enriched with other data streams, e.g. derivative exposures (if partly available of the financial institution in the network), market-based indicators such as CDS spreads, granular interbank lending data where available, and temporal data.

6.2.5 Representativeness of Networks

Analogue to the pre-defined criteria outlined in section 3.3, the results can be discussed.

- The real syndicated loan network as an indicator for the interbank financial market is capturing systemically important financial institutions (SIFIs) and also many of their partners and connected market participants outside their geographic region. This is true for both analyzed graphs in this study G_{LCC} and GSIB-W_2, although arguably the sampling of the subgraph was influenced with this representativeness criteria in mind.

- Both the real and synthetic networks include an accurate representation of exposure in the network through approximation of lender share exposure of the syndicated loans. The loans directly determine the structure of the network, with their total value used as weight. The geometric distribution with which the ERGM reconstructs the weights and attempts to recreate outgoing exposure between the nodes was representative, but less representative for the borrower-side.

- The real network does reflect complex interdependencies and highly interconnected institutions. However, as the reciprocity in the real network is quite low, and the network experiences heterogeneity between high in-degree and out-degree nodes, flow of exposure might be less accurate due and better understood using more granular data.

- The network would be suitable for future shock response and contagion analysis, although it was outside the scope of this analysis.

- Insight into market conditions under temporal financial stress are limited, because this is not a temporal construction and analysis of the network. However, as discussed the structure and topology of the observed network suggests it would behave similar to results found in previous and recent literature. This is evidenced by structural assumption on sparsity, scaling of degrees, reciprocity and skewed distribution of exposure in the network.

Moreover, to summarize ERGM results in context of representativeness of synthetic networks to empiric networks, they are argued as follows:

- Structural size, edges, and density are matched.

- The fitted ERGM improves upon the baseline models across the metrics.

- The node level metrics, specifically degree and strength skewness for out-going loans resemble those in GSIB-W_2 within a 95% interval in 500 simulations. It is not representative of the borrower-side obligation in the network, as it was not explicity modeled.

Implications The results argue for a positive evidence that under certain calibration ERGMs can be used to create synthetic representative financial network. The model calibration necessitates that structural features like degree tails or reciprocity, as well as weight heterogeneity, need to be explicitly defined in the model. The syndicated loan networks representativeness opens a new possibility for researchers to analyze financial stability in the interbank market. While syndicated loans are only one layer among many that constitute the full financial ecosystem of interdependence, they do serve as an indicator. This also has implication for policy makers wanting to understand results of exposures in the interbank network. The global syndicated loan market of banks suggests, that majority position loans and, in fact, the majority of loans are underwritten by only few banks worldwide, and that evidence of a balanced global bank network was not found.

The research also has implications on banks, showing that the syndicated loan market, while only indicative of size of interbank loans, shows highly skewed exposure and therefore risk on few financial entities. However, as syndicated loan contracts are not published, drawing inference to the actual risk of certain banks is not valid. Moreover, the syndicated loan splitting in Section 4 specifically showed, that there is no uniformity in reporting of syndicated loans or the banks roles in them, making meaningful analysis challenging.

The literature review revealed that researchers should study the robustness and limitations of using theoretical network models for understanding financial networks, as opposed to disproportionately drawing structural inference of network statistics and financial implications from them. Researchers in this field should also look into improving the reproducibility of their results, as often data used in their studies is unavailable to verify, or compare results against. There was no instance in which interbank data was available to reproduce the results. Many times, parameter model specification is only hinted at, or expressed in mathematical terms instead of providing part of the source code, or parameters themselves that were used to set up the network model.

6.3 Future Work

There are multiple ways the research of this thesis can be extended in the future. To make networks, their data and findings comparable, in line with suggestion of [2], making the constructed syndicated loan graphs available on a unified database that collects and distributes graphs such as [340], after comprehensive robustness checks can be a next direct step.

Future work can also include a deep analysis of the real syndicated loan interbank network, and testing cascade failure on this network, as this was not done with any previous syndicated loan network. Moreover, this network offers the possibility to compare different financial interbank systems at a future point, as previous research often used direct e-MID interbank data. The syndicated loan data can also be suitable for constructing temporal networks, and therefore, temporal ERGM analysis could be interesting.

To improve the ERGM specification presented in this study, the next natural step would be to introduce a targeted borrower heterogeneity or variance-inflating term to improve borrower-side loan heterogeneity.

To build on the underlying ERGM, researchers may work towards implementing better likelihood algorithms available for binary ERGMs [112, 341] into valued ERGMs and finding solutions around weight variance increasing model convergence challenges [116], which would allow a fit of the full syndicated loan network. If this is not possible, an extensive robustness test between a range of down-scaled subgraphs and their generalizability using ERGMs should be conducted.

7 Related Work

While this thesis fills a gap in research on synthetic network construction of a real syndicated loan network, throughout the analysis and discussion related relevant papers were mentioned. Most notably, this work extends on mainly three points, namely the literature reviews, the construction of real syndicated loan networks, and the synthetic network construction using ERGM.

Literature Reviews A seminal paper, which falls in line with and was discussed during the systematic literature review is the review by Lim et al. [40]. While points like graph generative models inability to deal with and fit real data have been somewhat overcome, their main critique of robust graph validation processes falls in line with results of this literature review. As there are no unified measures, even in recent literature as to how networks are assessed, and sometimes cross validation of synthetic graphs with real ones is missing, it remains challenging to compare graphs across publications. Their finding of researchers constructing general random graphs and using them in a financial context was extended in this literature review, as even in recent literature theoretical random models remain used. Lastly, their point on the inability of synthetic graph models to scale still remains true, as discussed with the ERGM. Similarly Emmert-Streib et al. [9] found the fragmented state of the financial network literature across publications, topics and approaches noteworthy, which is in line with the literature results discussed.

Syndicated Loan Networks This thesis provides a unique analysis of reconstructing a syndicated loan network, however similar papers that construct empirical syndicated loan networks are the papers which were used to compare results during the network discussion [325][338][339][342][336]. Oh and Parks [338] analysis of a real network, while more extensive, most closely resembles that in this thesis. As mentioned, their results regarding network metrics largely fall in line with results found in the analyzed empiric network.

ERGM Network Construction As financial network reconstruction using ERGMs is a niche research topic, comparable related studies are rare. Most notable is Engel et al. [100] who reconstruct financial networks using ERGM, however, based only on metrics mentioned in previous primary literature, find promising results in the models ability to reconstruct topological properties. Similarly, Cimini et al. [98] reconstruct an interbank network using the ERGM on e-MID data and find unfavorable results. Most recently similar work was done by Macchiati et al. [83], which find favorable results on e-MID data using an adapted ERGM. Due to a lack of overlapping data and metrics, it is difficult to compare results of syndicated loan interbank networks to interbank networks constructed from e-MID data, however it shows that this analysis is highly relevant to recent discussion in the financial network literature.

8 Conclusion

Adding to previous literature on the construction and comparison of synthetic and real networks, this thesis analyses the construction of financial networks. First, broad primarily literature was investigated using a systematic literature review to understand the current and relevant approaches, challenges and opportunities in financial network research. Secondly, to reconstruct a real and synthetic interbank network, considering most recent developments in literature, an exponential random graph model (ERGM) was fitted and implemented. The real network was constructed using a global syndicated loan network of financial institutions.

The main results unfold from these two parts. First, the literature review provided insight into the need for synthetic network and network reconstruction, namely unavailable or insufficient data of financial institutions. While the review showed, that it is common to use theoretical random network models, some researchers attempted to solve the issue of data sparsity by fitting empiric data to generative models. Literature discussed networks in the domain of stocks, payment and transactions, interbank, and economic networks. Network construction results in literature were only cross validated in the interbank domain, with stock networks usually used in combination with theoretical models to analyze general network analysis methods. The results could rarely be compared against each other, and showed non-uniformity even within the same domain.

As ERGMs are the latest development of fitting and reconstructing interbank networks, as well as some results in cross-validation, it was chosen among the methods discussed. The data of syndicated global loans of financial institutions used, was processed in line with previous studies, and then fitted using the ERGM. Discussing the results of the constructing the real network suggests, that it is able to represent the complex interdependencies of the interbank financial system. A known issue encountered was the challenge to use ERGM on large scale models, as computational needs are high. To converge a model realistically, a subgraph with globally systemically important banks was extracted and used for fitting the ERGM. The best ERGM fit was found with parameters capturing the presence of edges, sum of weight, and explicit coefficients for skewness of out-going loans and fixed low reciprocity. These results are in line with findings of ERGM models in related literature. The resulting model allowed to generate synthetic networks. Specifically, 500 networks were generated and collectively analyzed, and results suggest the ERGM model can be used to reconstruct the financial network of syndicated loans appropriately. The networks moreover seem to represent the real network, in structural properties and centrality measures and improve upon theoretical baseline models. Some results will need to be validated with the full scale network at a future point. Moreover, future research can focus on uniform ways to validate the resulting network with results presented in literature.

References

[1] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, November 1994.

[2] F. Emmert-Streib, S. Tripathi, O. Yli-Harja, and M. Dehmer. Understanding the World Economy in Terms of Networks: A Survey of Data-Based Network Science Approaches on Economic Networks. *Frontiers in Applied Mathematics and Statistics*, 4, 2018.

[3] Karl Finger and Thomas Lux. Network formation in the interbank money market: An application of the actor-oriented model. *Social Networks*, 48:237–249, January 2017.

[4] José Armando Hernández and Miguel Colom. Repeatability, Reproducibility, Replicability, Reusability (4R) in Journals' Policies and Software/Data Management in Scientific Publications: A Survey, Discussion, and Perspectives, December 2023.

[5] Diana Karimova. *Modelling Financial Networks with Kronecker Graph.* PhD thesis, Università Ca' Foscari Venezia, July 2018.

[6] James H. Jones, Elspeth Ready, and Ashley Hazel. Graph concepts and measures. https://eehh-stanford.github.io/SNA-workshop/graphs.html, April 2018.

[7] Jørgen Bang-Jensen and Gregory Z. Gutin. *Digraphs: Theory, Algorithms and Applications.* Springer, December 2008.

[8] Frank Harary et al. *Addison-Wesley.* Reading, MA, 1969.

[9] Frank Emmert-Streib, Aliyu Musa, Kestutis Baltakys, Juho Kanniainen, Shailesh Tripathi, Olli Yli-Harja, Herbert Jodlbauer, and Matthias Dehmer. Computational Analysis of the structural properties of Economic and Financial Networks, October 2017.

[10] Petter Holme. Modern temporal network theory: A colloquium. *The European Physical Journal B*, 88(9):234, September 2015.

[11] Mohammed Saqr. Temporal network analysis: Introduction, methods and detailed tutorial with R, July 2023.

[12] Stephen P Borgatti and Martin G Everett. Models of corerperiphery structures. *Social Networks*, 1999.

[13] Daniel Fricke and Thomas Lux. Core–Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform. *Computational Economics*, 45(3):359–395, March 2015.

[14] Ben Craig and Goetz Von Peter. Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347, July 2014.

[15] L. Bargigli, G. Di Iasio, L. Infante, F. Lillo, and F. Pierobon. The multiplex structure of interbank networks. *Quantitative Finance*, 15(4):673–691, April 2015.

[16] Mark Newman. *Networks*, volume 1. Oxford University Press, 2 edition, October 2018.

[17] Jennifer Golbeck. Chapter 3 - Network Structure and Measures. In Jennifer Golbeck, editor, *Analyzing the Social Web*, pages 25–44. Morgan Kaufmann, Boston, January 2013.

[18] Frank Cowell. Measurement of Inequality (Published in Handbook of Income Distribution, a B Atkinson and F Bourguignon (Eds), 1998), July 1998.

[19] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, November 2009.

[20] Frank J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[21] Reinhard Diestel. *Graph Theory*. Number 173 in Graduate Texts in Mathematics. Springer, Berlin, fifth edition, first softcover printing edition, 2018.

[22] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[23] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, January 2003.

[24] M. E. J. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20):208701, October 2002.

[25] Diego Garlaschelli and Maria I. Loffredo. Patterns of link reciprocity in directed networks. *Physical Review Letters*, 93(26):268701, December 2004.

[26] Ray Solomonoff and Anatol Rapoport. Connectivity of random nets. *The Bulletin of Mathematical Biophysics*, 13(2):107–117, June 1951.

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

[28] Erdős and Renyi. On random graph. *Publicationes Mathematicate*, 6:290–297, 1959.

[29] Stephanie Rendón de la Torre, Jaan Kalda, Robert Kitt, and Jüri Engelbrecht. On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos, Solitons & Fractals*, 90:18–27, September 2016.

[30] Kirsten Bonde Rørdam and Morten L Bech. The topology of Danish interbank money flows. *Banks and Bank Systems*, 4(4), 2009.

[31] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079–1187, June 2002.

[32] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, March 2001.

[33] Juyong Park and M. E. J. Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, December 2004.

[34] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6):7332–7342, December 1999.

[35] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the World-Wide Web. *Nature*, 401(6749):130–131, September 1999.

[36] Robert Meusel, Sebastiano Vigna, Oliver Lehmberg, and Christian Bizer. Graph structure in the web — revisited: A trick of the heavy tail. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 427–432, New York, NY, USA, April 2014. Association for Computing Machinery.

[37] Robert Meusel, Sebastiano Vigna, Oliver Lehmberg, and Christian Bizer. The Graph Structure in the Web – Analyzed on Different Aggregation Levels. *The Journal of Web Science*, 1, July 2015.

[38] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999.

[39] Béla Bollobás and Oliver Riordan. The Diameter of a Scale-Free RandomGraph. *Combinatorica*, 24(1):5–34, January 2004.

[40] Seung-Hwan Lim, Sangkeun Lee, Sarah S Powers, Mallikarjun Shankar, and Neena Imam. Survey of Approaches to Generate Realistic Synthetic Graphs. Technical Report ORNL/TM–2016/3, 1339361, Oak Ridge National Laboratory, October 2016.

[41] Matthew Elliott, Benjamin Golub, and Matthew O. Jackson. Financial Networks and Contagion. *American Economic Review*, 104(10):3115–3153, October 2014.

[42] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic Risk and Stability in Financial Networks. *American Economic Review*, 105(2):564–608, February 2015.

[43] Wei Li, Dror Y. Kenett, Kazuko Yamasaki, H. Eugene Stanley, and Shlomo Havlin. Ranking the Economic Importance of Countries and Industries, August 2014.

[44] Martin Summer. Financial Contagion and Network Analysis. *Annual Review of Financial Economics*, 5(1):277–297, 2013.

[45] Paul Glasserman and H. Peyton Young. Contagion in Financial Networks. *Journal of Economic Literature*, 54(3):779–831, September 2016.

[46] Rama Cont, Amal Moussa, and Edson B. Santos. Network Structure and Systemic Risk in Banking Systems. In Jean-Pierre Fouque and Joseph A. Langsam, editors, *Handbook on Systemic Risk*, pages 327–368. Cambridge University Press, 1 edition, May 2013.

[47] Matthew O. Jackson and Agathe Pernoud. Systemic Risk in Financial Networks: A Survey, December 2020.

[48] John B. Caouette, Edward I. Altman, Paul Narayanan, and Robert Nimmo. *Managing Credit Risk: The Great Challenge for the Global Financial Markets*. Wiley, 1 edition, May 2008.

[49] Gary A Goodman. Special Problems of Syndicated Loans, April 2014.

[50] BBVA. What is a syndicated financing?, May 2024.

[51] Amir Sufi. Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans. *The Journal of Finance*, 62(2):629–668, 2007.

[52] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, May 2007.

[53] Martina Morris, Mark S. Handcock, and David R. Hunter. Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software*, 24:1–24, May 2008.

[54] David R. Hunter, Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, 24:1–29, May 2008.

[55] Tom A B Snijders. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure*, 3(2), June 2002.

[56] Charles J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

[57] David R. Hunter and Mark S. Handcock. Inference in Curved Exponential Family Models for Networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006.

[58] Yi Wang, Yang Zhang, and Yong Wang. Mining Data Streams with Skewed Distribution by Static Classifier Ensemble. In Been-Chian Chien and Tzung-Pei Hong, editors, *Opportunities and Challenges for Next-Generation Applied Intelligence*, Studies in Computational Intelligence, pages 65–71. Springer, Berlin, Heidelberg, 2009.

[59] Angelo Mele. Estimating Network Models using Neural Networks, February 2025.

[60] Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *Journal of Statistical Software*, 24:1–11, May 2008.

[61] David R. Hunter, Steven M. Goodreau, and Mark S. Handcock. Goodness of Fit of Social Network Model. *Journal of the American Statistical Association*, 103(481):248–258, 2008.

[62] Pavel N. Krivitsky. Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6(none), January 2012.

[63] Skyler J. Cranmer and Bruce A. Desmarais. Inferential Network Analysis with Exponential Random Graph Models. *Political Analysis*, 19(1):66–86, January 2011.

[64] Dean Lusher, Johan Koskinen, and Garry Robins, editors. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, Cambridge, 2012.

[65] Pavel N. Krivitsky and Carter T. Butts. Exponential-Family Random Graph Models for Rank-Order Relational Data. *Sociological Methodology*, 47(1):68–112, June 2015.

[66] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.

[67] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, March 1978.

[68] Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, pages 1–10, London England United Kingdom, May 2014. ACM.

[69] Claes Wohlin, Marcos Kalinowski, Katia Romero Felizardo, and Emilia Mendes. Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, 147:106908, July 2022.

[70] Samireh Jalali and Claes Wohlin. Systematic literature studies: Database searches vs. backward snowballing. In *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 29–38, Lund Sweden, September 2012. ACM.

[71] Deepika Badampudi, Claes Wohlin, and Kai Petersen. Experiences from using snowballing and database searches in systematic literature studies. In *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, EASE '15, pages 1–10,

New York, NY, USA, April 2015. Association for Computing Machinery.

[72] Erica Mourão, João Felipe Pimentel, Leonardo Murta, Marcos Kalinowski, Emilia Mendes, and Claes Wohlin. On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and Software Technology*, 123:106294, July 2020.

[73] Barbara Kitchenham. Procedures for Performing Systematic Reviews. *Keele, UK, Keele University*, 33, August 2004.

[74] Philipp Mayring. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*, volume 13. Beltz, Weinheim Basel, 13., überarbeitete auflage edition, 2022.

[75] Walmir Silva, Herbert Kimura, and Vinicius Amorim Sobreiro. An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability*, 28:91–114, February 2017.

[76] Massimo Aria and Corrado Cuccurullo. *Bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4):959–975, November 2017.

[77] Volker Brühl. How to Define a Systemically Important Financial Institution — A New Perspective. *Intereconomics*, 52(2):107–110, March 2017.

[78] Erlend W. Nier, Jing Yang, Tanju Yorulmazer, and Amadeo Alentorn. Network Models and Financial Stability, April 2008.

[79] Galina Hale, Tümer Kapan, and Camelia Minoiu. Crisis Transmission in the Global Banking Network*. *IMF Working Papers*, 2016(091), April 2016.

[80] I. Mastromatteo, E. Zarinelli, and M. Marsili. Reconstruction of financial networks for robust estimation of systemic risk. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(3), 2012.

[81] Giulio Cimini, Tiziano Squartini, Diego Garlaschelli, and Andrea Gabrielli. Systemic Risk Analysis on Reconstructed Economic and Financial Networks. *Scientific Reports*, 5(1):15758, October 2015.

[82] Isobel E. Seabrook, Paolo Barucca, and Fabio Caccioli. Evaluating structural edge importance in temporal networks. *EPJ Data Science*, 10(1):1–23, December 2021.

[83] Valentina Macchiati, Piero Mazzarisi, and Diego Garlaschelli. Interbank network reconstruction enforcing density and reciprocity. *Chaos, Solitons & Fractals*, 186:115279, September 2024.

[84] Nikos Paltalidis, Dimitrios Gounopoulos, Renatas Kizys, and Yiannis Koutelidakis. Transmission channels of systemic risk and contagion in the European financial network. *Journal of Banking & Finance*, 61:S36–S52, December 2015.

[85] Fei Su, Lili Zhai, Yunyan Zhou, Zixi Zhuang, and Feifan Wang. Risk contagion in financial markets: A systematic review using bibliometric methods. *Australian Economic Papers*, n/a(n/a), May 2023.

[86] Marco Bardoscia, Paolo Barucca, Stefano Battiston, Fabio Caccioli, Giulio Cimini, Diego Garlaschelli, Fabio Saracco, Tiziano Squartini, and Guido Caldarelli. The Physics of Financial Networks. *Nature Reviews Physics*, 3(7):490–507, June 2021.

[87] Giulia Iori and Rosario N. Mantegna. Chapter 11 - Empirical Analyses of Networks in Finance. In Cars Hommes and Blake LeBaron, editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, pages 637–685. Elsevier, January 2018.

[88] Franklin Allen and Ana Babus. Networks in finance. *The network challenge: strategy, profit, and risk in an interlinked world*, 367, 2009.

[89] Makoto Nirei, Vladyslav Sushko, and Julián Caballero. Bank Capital Shock Propagation via Syndicated Interconnectedness. *Computational Economics*, 47(1):67–96, January 2016.

[90] Jian Cai, Frederik Eidam, Anthony Saunders, and Sascha Steffen. Syndication, interconnectedness, and systemic risk. *Journal of Financial Stability*, 34:105–120, February 2018.

[91] Greg Nini. THE ROLE OF LOCAL BANKS IN PROMOTING EXTERNAL FINANCE: A STUDY OF SYNDICATED LENDING TO EMERGING MARKET BORROWERS. *Bank of International Settlements BIS*, January 2004.

[92] Seung Jung Lee, Lucy Qian Liu, and Viktors Stebunovs. Risk Taking and Interest Rates: Evidence from Decades in the Global Syndicated Loan Market. *International Finance Discussion Papers*, 2017.0(1188):1–38, January 2017.

[93] William R. McCumber. Global Private Capital: A Network Perspective, July 2015.

[94] Leonid Sopotnitskiy. The size of global markets 2024 in charts. Market Dynamics Insights, LSEG Data & Analytics, 2024.

[95] Yener Altunbaş and Alper Kara. LARGE DEBT FINANCING SYNDICATED LOANS VERSUS CORPORATE BONDS. *European Central Bank Eurosystem*, 1028, March 2009.

[96] Stefan Thurner and Sebastian Poledna. DebtRank-transparency: Controlling systemic risk in financial networks. *Scientific Reports*, 3(1):1888, May 2013.

[97] Nicolò Musmeci, Stefano Battiston, Guido Caldarelli, Michelangelo Puliga, and Andrea Gabrielli. Bootstrapping Topological Properties and Systemic Risk of Complex Networks Using the Fitness Model. *Journal of Statistical Physics*, 151, May 2013.

[98] Giulio Cimini, Tiziano Squartini, Nicolò Musmeci, Michelangelo Puliga, Andrea Gabrielli, Diego Garlaschelli, Stefano Battiston, and Guido Caldarelli. Reconstructing Topological Properties of Complex Networks Using the Fitness Model. In Luca Maria Aiello and Daniel McFarland, editors, *Social Informatics*, Lecture Notes in Computer Science, pages 323–333, Cham, 2015. Springer International Publishing.

[99] Clement Lee and Darren J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1–50, December 2019.

[100] Janina Engel, Andrea Pagano, and Matthias Scherer. Reconstructing the topology of financial networks from degree distributions and reciprocity. *Journal of Multivariate Analysis*, 172:210–222, July 2019.

[101] Financial Stability Board. 2024 List of Global Systemically Important Banks (G-SIBs). Publication, FSB, November 2024.

[102] Kristian Blickle, Quirin Fleckenstein, Sebastian Hillenbrand, and Anthony Saunders. The Myth of the Lead Arranger's Share. *SSRN Electronic Journal*, 2020.

[103] Victoria Ivashina. Asymmetric information effects on loan spreads. *Journal of Financial Economics*, 92(2):300–319, May 2009.

[104] Pankaj Kumar Maskara. Economic value in tranching of syndicated loans. *Journal of Banking & Finance*, 34(5):946–955, May 2010.

[105] Anne L. Beatty, Wei-Yi (Scott) Liao, and Haiwen Zhang. The Effect of Banks' Financial Reporting on Syndicated Loan Structures. *SSRN Electronic Journal*, 2018.

[106] Jian Cai, Frederik Eidam, Anthony Saunders, and Sascha Steffen. Loan Syndication Structures and Price Collusion. *SSRN Electronic Journal*, 2018.

[107] Alemu Tulu Chala. Syndicated Lending: The Role of Relationships for the Retained Share. *EconStor*, 2018(34), 2018.

[108] Max Bruche, Frederic Malherbe, and Ralf R Meisenzahl. Pipeline Risk in Leveraged Loan Syndication. *The Review of Financial Studies*, 33(12):5660–5705, December 2020.

[109] Sabrina Chi, Hengda Jin, Edward Owens, and Karen Ton. What's My Share? Information Acquisition by Loan Syndicate Participants. *SSRN Electronic Journal*, January 2020.

[110] Thomas Schneider, Philip E Strahan, and Jun Yang. Syndicated Lending, Competition, and Relative Performance Evaluation. *The Review of Financial Studies*, 37(12):3802–3834, December 2024.

[111] Alex Stivala, Garry Robins, and Alessandro Lomi. Exponential random graph model parameter estimation for very large directed networks. *PLoS ONE*, 15(1):e0227804, January 2020.

[112] Maksym Byshkin, Alex Stivala, Antonietta Mira, Rolf Krause, Garry Robins, and Alessandro Lomi. Auxiliary Parameter MCMC for Exponential Random Graph Models. *Journal of Statistical Physics*, 165(4):740–754, November 2016.

[113] Jenny Krheel. RPubs - Valued ERGMs. https://rpubs.com/krhee1/720308, January 2021.

[114] Galit Shmueli, Thomas P. Minka, Joseph B. Kadane, Sharad Borle, and Peter Boatwright. A Useful Distribution for Fitting Discrete Data: Revival of the Conway–Maxwell–Poisson Distribution. *Journal of the*

Royal Statistical Society Series C: Applied Statistics, 54(1):127–142, January 2005.

[115] James H. Jones, Elspeth Ready, and Ashley Hazel. Introduction to ERGMs. https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html, February 2020.

[116] Peng Huang and Carter T. Butts. Parameter estimation procedures for exponential-family random graph models on count-valued networks: A comparative simulation study. *Social Networks*, 76:51–67, January 2024.

[117] Prasanna Gai and Sujit Kapadia. Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120):2401–2423, March 2010.

[118] Prasanna Gai, Andrew Haldane, and Sujit Kapadia. Complexity, concentration and contagion. *Journal of Monetary Economics*, 58(5):453–470, July 2011.

[119] Masayasu Kanno. Assessing systemic risk using interbank exposures in the global banking system. *Journal of Financial Stability*, 20:105–130, October 2015.

[120] Pavel N. Krivitsky, David R. Hunter, Martina Morris, and Chad Klumb. **Ergm** 4: New Features for Analyzing Exponential-Family Random Graph Models. *Journal of Statistical Software*, 105(6), 2023.

[121] Pavel N. Krivitsky and Carter T. Butts. Modeling valued networks with statnet. https://statnet.org/workshop-valued/valued.html#311_Valued_ERGMs, June 2024.

[122] Rossana Mastrandrea, Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli. Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics*, 16(4):043022, April 2014.

[123] Alex Smolyak, Orr Levy, Louis Shekhtman, and Shlomo Havlin. Interdependent networks in Economics and Finance—A Physics approach. *Physica A: Statistical Mechanics and its Applications*, 512:612–619, December 2018.

[124] Anna Nagurney. Networks in economics and finance in Networks and beyond: A half century retrospective. *Networks*, 77(1):50–65, 2021.

[125] Samuel A. Assefa, Danial Dervovic, Mahmoud Mahfouz, Robert E. Tillman, Prashant Reddy, and Manuela Veloso. Generating synthetic data in finance: Opportunities, challenges and pitfalls. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, pages 1–8, New York, NY, USA, October 2021. Association for Computing Machinery.

[126] Jianian Wang, Sheng Zhang, Yanghua Xiao, and Rui Song. A Review on Graph Neural Network Methods in Financial Applications, April 2022.

[127] Alexander Denev. FINANCIAL NETWORKS AS DIRECTED CYCLIC GRAPHS -. *Corporate Ownership & Control*, 11(4), 2014.

[128] T. R. Hurd, Davide Cellai, Sergey Melnik, and Quentin H. Shao. DOUBLE CASCADE MODEL OF FINANCIAL CRISES. *International Journal of Theoretical and Applied Finance*, 19(05):1650041, August 2016.

[129] T. R. Hurd. Random Graph Models. In T. R. Hurd, editor, *Contagion! Systemic Risk in Financial Networks*, SpringerBriefs in Quantitative Finance, pages 45–70. Springer International Publishing, Cham, 2016.

[130] Edward A Bender and E. Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, May 1978.

[131] Béla Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1(4):311–316, December 1980.

[132] Thomas R. Hurd, James P. Gleeson, and Sergey Melnik. A framework for analyzing contagion in assortative banking networks. *PLOS ONE*, 12(2):e0170579, February 2017.

[133] Xiaokang Wang, Huiwen Wang, Zhichao Wang, Shan Lu, and Ying Fan. Risk spillover network structure learning for correlated financial assets: A directed acyclic graph approach. *Information Sciences*, 580:152–173, November 2021.

[134] Tongfeng Weng, Haiying Wang, Huijie Yang, Changgui Gu, Jie Zhang, and Michael Small. Representing complex networks without connectivity via spectrum series. *Information Sciences*, 563:16–22, July 2021.

[135] Charles D. Brummitt and Teruyoshi Kobayashi. Cascades in multiplex financial networks with debts of different seniority. *Physical Review E*, 91(6):062813, June 2015.

[136] Yuanying Guan and Micah Pollak. CONTAGION IN HETEROGENEOUS FINANCIAL NETWORKS. *Advances in Complex Systems*, 19(01n02):1650001, February 2016.

[137] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, December 1999.

[138] Christian Borgs, Jennifer Chayes, and Oliver Riordan. Directed Scale-Free Graphs. *Proc. 14th Annual ACM-SIAM Symp. on Discrete Algorithms*, 2003.

[139] Mitja Steinbacher and Timotej Jagrič. Interbank rules during economic declines: Can banks safeguard capital base? *Journal of Economic Interaction and Coordination*, 15(2):471–499, April 2020.

[140] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002.

[141] Hamed Amini and Andreea Minca. Inhomogeneous Financial Networks and Contagious Links. *Operations Research*, 64(5):1109–1120, October 2016.

[142] Mindaugas Bloznelis, Friedrich Götze, and Jerzy Jaworski. Birth of a Strongly Connected Giant in an Inhomogeneous Random Digraph. *Journal of Applied Probability*, 49(3):601–611, September 2012.

[143] S. Heise and R. Kühn. Derivatives and credit contagion in interconnected networks. *The European Physical Journal B*, 85(4):115, April 2012.

[144] J. P. L. Hatchett and R. Kühn. Effects of economic interactions on credit risk. *Journal of Physics A: Mathematical and General*, 39(10):2231, February 2006.

[145] J. P. L. Hatchett and R. Kühn. Credit contagion and credit risk. *Quantitative Finance*, 9(4):373–382, June 2009.

[146] Peter Neu and Reimer Kühn. Credit risk enhancement in a network of interdependent firms. *Physica A: Statistical Mechanics and its Applications*, 342(3-4):639–655, November 2004.

[147] Tor Oddvar Berge. Operational problems in banks – Effects on the settlement of payments in Norges Bank. *ECONOMIC BULLETIN 2012*, 83:36–47, 2010.

[148] Joanna McLafferty and Edward Denbee. Liquidity Saving in CHAPS: A Simulation Study. In *Simulation in Computational Finance and Economics: Tools and Emerging Applications*, pages 120–142. IGI Global, 2013.

[149] Kimmo Soramäki and Samantha Cook. SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems. *Economics*, 7(1), December 2013.

[150] M. Blavarg and Patrick Nimander. Interbank exposures and systemic risk. *Economic Review*, 2:19–45, January 2002.

[151] Hans Degryse and Grégory Nguyen. Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System. *International Journal of Central Banking*, 2007.

[152] Iman van Lelyveld and Franka Liedorp. Interbank Contagion in the Dutch Banking Sector: A Sensitivity Analysis. *International Journal of Central Banking*, 2006.

[153] Simon J. Wells. Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion, September 2004.

[154] Sharyn O'Halloran and Nikolai Nowaczyk. An Artificial Intelligence Approach to Regulating Systemic Risk. *Frontiers in Artificial Intelligence*, 2:7, May 2019.

[155] Mohsen Bayati, Jeong Han Kim, and Amin Saberi. A Sequential Algorithm for Generating Random Graphs. *Algorithmica*, 58(4):860–910, December 2010.

[156] Tom Britton, Maria Deijfen, and Anders Martin-Löf. Generating Simple Random Graphs with Prescribed Degree Distribution. *Journal of Statistical Physics*, 124(6):1377–1397, September 2006.

[157] Simone Berardi and Gabriele Tedeschi. From banks' strategies to financial (in)stability. *International Review of Economics & Finance*, 47:255–272, January 2017.

[158] D. Garlaschelli and M. I. Loffredo. Fitness-dependent topological properties of the World Trade Web. *Physical Review Letters*, 93(18):188701, October 2004.

[159] Diego Garlaschelli, Stefano Battiston, Maurizio Castri, Vito D. P. Servedio, and Guido Caldarelli. The scale-free topology of market investments. *Physica A: Statistical Mechanics and its Applications*, 350(2-4):491–499, May 2005.

[160] Robert M. May and Nimalan Arinaminpathy. Systemic risk: The dynamics of model banking systems. *Journal of The Royal Society Interface*, 7(46):823–838, October 2009.

[161] Angelika Sachs. Completeness, interconnectedness and distribution of interbank exposures—a parameterized analysis of the stability of financial networks. *Quantitative Finance*, 14(9):1677–1692, September 2014.

[162] Paolo Emilio Mistrulli. Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127, May 2011.

[163] Tatiana von Landesberger, Simon Diel, Sebastian Bremm, and Dieter W Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, April 2015.

[164] Edgar Alonso Lopez-Rojas and Stefan Axelsson. BANKSIM: A BANK PAYMENTS SIMULATOR FOR FRAUD DETECTION RESEARCH. In *Proceedings of the European Modeling and Simulation Symposium, 2014*, Genoa, 2014. Blekinge Institute of Technology.

[165] Debachudamani Prusti, Daisy Das, and Santanu Kumar Rath. Credit Card Fraud Detection Technique by Applying Graph Database Model. *Arabian Journal for Science and Engineering*, 46(9):1–20, September 2021.

[166] Xiao Tian, Mahashweta Das, and Chiranjeet Chetia. Generating Large-Scale Synthetic Payment Graphs with Realistic Fraudulent and Money Laundering Patterns. In *KDD-MLF '21: KDD Workshop on Machine Learning in Finance, August 2021, Singapore.*, New York, NY, USA, August 2021. ACM.

[167] David Bader and Kamesh Madduri. GTGraph: A synthetic graph generator suite. College of Computing: Georgia Institute of Technology, March 2006.

[168] Fan Zhang, Lei Zou, Li Zeng, and Xiangyang Gou. Dolha - an efficient and exact data structure for streaming graphs. *World Wide Web*, 23(2):873–903, March 2020.

[169] Shilei Zhang, Toyotaro Suzumura, and Li Zhang. DynGraphTrans: Dynamic Graph Embedding via Modified Universal Transformer Networks for Financial Transaction Data. In *2021 IEEE International Conference on Smart Data Services (SMDS)*, pages 184–191, Chicago, IL, USA, September 2021. IEEE.

[170] F. L. C. Visser. Detection of Money Laundering Transaction Network Structures and Typologies using Machine Learning Techniques. Master's thesis, Erasmus University Rotterdam School of Economics, May 2020.

[171] Aric A Hagberg, Daniel A Schult, and Pieter J Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, 2008.

[172] Edgar Alonso Lopez-Rojas, Ahmad Elmir, and Stefan Axelsson. PAYSIM: A FINANCIAL MOBILE MONEY SIMULATOR FOR FRAUD DETECTION. In *28th European Modeling and Simulation Symposium*, pages 249–255, Dime University of Genoa, September 2016. Dime University of Genoa.

[173] Erik Altman, Jovan Blanuša, Luc von Niederhäusern, Béni Egressy, Andreea Anghel, and Kubilay Atasu. Realistic Synthetic Financial Transactions for Anti-Money Laundering Models, October 2023.

[174] Abhijeet Godase and Vahida Attar. Classifier Ensemble for Imbalanced Data Stream Classification. In *Proceedings of the CUBE International Information Technology Conference*, CUBE '12, pages 284–289, New York, NY, USA, September 2012. Association for Computing Machinery.

[175] Sheng Chen and Haibo He. Towards incremental learning of nonstationary imbalanced data stream: A multiple selectively recursive approach. *Evolving Systems*, 2(1):35–50, March 2011.

[176] Reza Soltani, Uyen Trang Nguyen, Yang Yang, Mohammad Faghani, Alaa Yagoub, and Aijun An. A new algorithm for money laundering detection based on structural similarity. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 1–7, October 2016.

[177] Ruizhi Zhou, Qin Zhang, Peng Zhang, Lingfeng Niu, and Xiaodong Lin. Anomaly detection in dynamic attributed networks. *Neural Computing and Applications*, 33(6):2125–2136, March 2021.

[178] Merritt Hughes and Anna Nagurney. A network model and algorithm for the analysis and estimation of financial flow of funds. *Computer Science in Economics and Management*, 5(1):23–39, February 1992.

[179] Michael Boss, Helmut Elsinger, Martin Summer, and Stefan Thurner 4. Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684, December 2004.

[180] Kimmo Soramäki, Morten L. Bech, Jeffrey Arnold, Robert J. Glass, and Walter E. Beyeler. The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333, 2007.

[181] Michelangelo Puliga, Andrea Flori, Giuseppe Pappalardo, Alessandro Chessa, and Fabio Pammolli. The Accounting Network: How Financial Institutions React to Systemic Crisis. *PLOS ONE*, 11(10):e0162855, October 2016.

[182] Yajing Huang and Feng Chen. Community Structure and Systemic Risk of Bank Correlation Networks Based on the U.S. Financial Crisis in 2008. *Algorithms*, 14(6):162, May 2021.

[183] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117, April 1998.

[184] G. De Masi, G. Iori, and G. Caldarelli. Fitness model for the Italian interbank money market. *Physical Review E*, 74(6):066112, December 2006.

[185] Giulia Iori, Giulia De Masi, Ovidiu Vasile Precup, Giampaolo Gabbi, and Guido Caldarelli. A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1):259–278, January 2008.

[186] Morten L. Bech and Enghin Atalay. The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22):5223–5246, November 2010.

[187] Hajime Inaoka, Takuto Ninomiya, Ken Taniguchi, Tokiko Shimizu, and Hideki Takayasu. Fractal Network derived from banking transaction. *Bank of Japan Working Paper Series*, April 2004.

[188] Martin Keller-Ressel and Stephanie Nargang. The hyperbolic geometry of financial networks. *Scientific Reports*, 11(1):4732, February 2021.

[189] Qingmin Hao, Jim Huangnan Shen, and Chien-Chiang Lee. Risk contagion of bank-firm loan network: Evidence from China. *Eurasian Business Review*, 13(2):341–361, June 2023.

[190] Larry Eisenberg and Thomas H. Noe. Systemic Risk in Financial Systems. *Management Science*, 47(2):236–249, February 2001.

[191] Masayasu Kanno. The network structure and systemic risk in the Japanese interbank market. *Japan and the World Economy*, 36:102–112, November 2015.

[192] Yair Censor and Stavros Andrea Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, 1997.

[193] Samaneh Ebrahimi, Mostafa Reisi-Gahrooei, Kamran Paynabar, and Shawn Mankad. Monitoring sparse and attributed networks with online Hurdle models. *IISE Transactions*, pages 1–14, January 2021.

[194] Karl Finger, Daniel Fricke, and Thomas Lux. Network analysis of the e-MID overnight money market: The informational value of different aggregation levels for intrinsic dynamic processes. *Computational Management Science*, 10(2):187–211, June 2013.

[195] P. Giudici and A. Spelta. Graphical Network Models for International Financial Flows. *Journal of Business & Economic Statistics*, 34(1):128–138, January 2016.

[196] Stefano Battiston, Guido Caldarelli, Marco D'Errico, and Stefano Gurciullo. Leveraging the network: A stress-test framework based on DebtRank. *Statistics & Risk Modeling*, 33(3-4):117–138, December 2016.

[197] Grzegorz Halaj and Christoffer Kok. Modelling the emergence of the interbank networks. *Quantitative Finance*, 15(4):653–671, 2015.

[198] Mattia Montagna and Thomas Lux. Contagion risk in the interbank market: A probabilistic approach to cope with incomplete structural information. *Quantitative Finance*, 17(1):101–120, January 2017.

[199] Francis X. Diebold and Kamil Yilmaz. Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets. *The Economic Journal*, 119(534):158–171, January 2009.

[200] Francis X. Diebold and Kamil Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, September 2014.

[201] Tomohiro Ando, Matthew Greenwood-Nimmo, and Yongcheol Shin. Quantile Connectedness: Modeling Tail Behavior in the Topology of Financial Networks. *Management Science*, 68(4):2401–2431, April 2022.

[202] Viral Acharya, Itamar Drechsler, and Philipp Schnabl. A Pyrrhic Victory? Bank Bailouts and Sovereign Credit Risk. *The Journal of Finance*, 69(6):2689–2739, 2014.

[203] Matthew Greenwood-Nimmo, Jingong Huang, and Viet Hoang Nguyen. Financial sector bailouts, sovereign bailouts, and the transfer of credit risk. *Journal of Financial Markets*, 42:121–142, January 2019.

[204] Mert Demirer, Francis X. Diebold, Laura Liu, and Kamil Yilmaz. Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1):1–15, 2018.

[205] W. Li, W. Liu, L. Wu, and X. Guo. Risk spillover networks in financial system based on information theory. *PLoS ONE*, 16(6 June), 2021.

[206] Xiangyun Gao, Shupei Huang, Xiaoqi Sun, Xiaoqing Hao, and Feng An. Modelling cointegration and Granger causality network to detect long-term equilibrium and diffusion paths in the financial system. *Royal Society Open Science*, 5(3):172092, March 2018.

[207] Oleg Deev and Štefan Lyócsa. Connectedness of financial institutions in Europe: A network approach across quantiles. *Physica A: Statistical Mechanics and its Applications*, 550:124035, July 2020.

[208] Eduard Baumöhl, Evžen Kočenda, Štefan Lyócsa, and Tomáš Výrost. Networks of volatility spillovers among stock markets. *KIER Discussion Paper*, 941, May 2016.

[209] Štefan Lyócsa, Tomáš Výrost, and Eduard Baumöhl. Return spillovers around the globe: A network approach. *Economic Modelling*, 77:133–146, March 2019.

[210] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertész. Dynamic asset trees and Black Monday. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):247–252, June 2003.

[211] Syed Jawad Hussain Shahzad, Jose Areola Hernandez, Mobeen Ur Rehman, Khamis Hamed Al-Yahyaee, and Muhammad Zakaria. A global network topology of stock markets: Transmitters and receivers of spillover effects. *Physica A: Statistical Mechanics and its Applications*, 492:2136–2153, February 2018.

[212] Ashadun Nobi, Seong Eun Maeng, Gyeong Gyun Ha, and Jae Woo Lee. Effects of global financial crisis on network structure in a local stock market. *Physica A: Statistical Mechanics and its Applications*, 407:135–143, August 2014.

[213] Sunil Kumar and Nivedita Deo. Correlation and network analysis of global financial indices. *Physical Review E*, 86(2):026101, August 2012.

[214] A. Namaki, A.H. Shirazi, R. Raei, and G.R. Jafari. Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications*, 390(21-22):3835–3841, 2011.

[215] Ersin Kantar, Mustafa Keskin, and Bayram Deviren. Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods. *Physica A: Statistical Mechanics and its Applications*, 391(7):2342–2352, April 2012.

[216] L. Kullmann, J. Kertész, and K. Kaski. Time-dependent cross-correlations between different stock returns: A directed network of influence. *Physical Review E*, 66(2):026125, August 2002.

[217] Frank Emmert-Streib and Matthias Dehmer. Influence of the Time Scale on the Construction of Financial Networks. *PLOS ONE*, 5(9):e12884, September 2010.

[218] Wei-Qiang Huang, Xin-Tian Zhuang, and Shuang Yao. A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 388(14):2956–2964, July 2009.

[219] Chi K. Tse, Jing Liu, and Francis C. M. Lau. A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667, September 2010.

[220] J. Chu and S. Nadarajah. A statistical analysis of UK financial networks. *Physica A: Statistical Mechanics and its Applications*, 471:445–459, April 2017.

[221] Vladimir Boginski, Sergiy Butenko, and Panos M. Pardalos. Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48(2):431–443, February 2005.

[222] Wei-Qiang Huang, Xin-Tian Zhuang, Shuang Yao, and Stan Uryasev. A financial network perspective of financial institutions' systemic risk contributions. *Physica A: Statistical Mechanics and its Applications*, 456:183–196, August 2016.

[223] Cheng Jiang, Qian Sun, Tanglin Ye, and Qingyun Wang. Identification of systemically important financial institutions in a multiplex financial network: A multi-attribute decision-based approach. *Physica A: Statistical Mechanics and its Applications*, 611:128446, February 2023.

[224] Chun-Xiao Nie and Fu-Tie Song. Constructing financial network based on P M F G and threshold method. *Physica A: Statistical Mechanics and its Applications*, 495:104–113, April 2018.

[225] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426, July 2005.

[226] Longfeng Zhao, Wei Li, and Xu Cai. Structure and dynamics of stock market in times of crisis. *Physics Letters A*, 380(5-6):654–666, February 2016.

[227] V. A. Kalyagin, A. P. Koldanov, P. A. Koldanov, and P. M. Pardalos. Optimal decision for the market graph identification problem in a sign similarity network. *Annals of Operations Research*, 266(1-2):313–327, July 2018.

[228] Grigory A. Bautin, Valery A. Kalyagin, Alexander P. Koldanov, Petr A. Koldanov, and Panos M. Pardalos. Simple measure of similarity for the market graph construction. *Computational Management Science*, 10(2):105–124, June 2013.

[229] Grigory A. Bautin, Valery A. Kalyagin, and Alexander P. Koldanov. Comparative Analysis of Two Similarity Measures for the Market Graph Construction. In Boris I. Goldengorin, Valery A. Kalyagin, and Panos M. Pardalos, editors, *Models, Algorithms, and Technologies for Network Analysis*, pages 29–41, New York, NY, 2013. Springer.

[230] Tao Zhong, Qinke Peng, Xiao Wang, and Jing Zhang. Novel indexes based on network structure to indicate financial market. *Physica A: Statistical Mechanics and its Applications*, 443:583–594, February 2016.

[231] Zhiwei Zhang, Dayong Zhang, Fei Wu, and Qiang Ji. Systemic risk in the Chinese financial system: A copula-based network approach. *International Journal of Finance & Economics*, 26(2):2044–2063, 2021.

[232] Qiang Ji and Ying Fan. Evolution of the world crude oil market integration: A graph theory analysis. *Energy Economics*, 53:90–100, January 2016.

[233] Fei Wu, Dayong Zhang, and Zhiwei Zhang. Systemic Risks and Spillovers in the Stock Market of China: A Sectoral Analysis, September 2018.

[234] Paweł Fiedor. Networks in financial markets based on the mutual information rate. *Physical Review E*, 89(5):052801, May 2014.

[235] Antoine Mandel and Vipin P. Veetil. Monetary dynamics in a network economy. *Journal of Economic Dynamics and Control*, 125:104084, April 2021.

[236] Enghin Atalay, Ali Hortaçsu, James Roberts, and Chad Syverson. Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202, March 2011.

[237] Tomohiko Konno. Network Structure of Japanese Firms. Scale-Free, Hierarchy, and Degree Correlation: Analysis from 800,000 Firms. *Economics*, 3(1), December 2009.

[238] Tao Wang, Shiying Xiao, Jun Yan, and Panpan Zhang. Regional and sectoral structures of the Chinese economy: A network perspective from multi-regional input–output tables. *Physica A: Statistical Mechanics and its Applications*, 581:126196, November 2021.

[239] Louis Kaplow. Pareto Principle and Competing Principles. *The Harvard John M. Olin Discussion Paper Series*, 2005(Discussion Paper No. 519), August 2005.

[240] Valentina Shumovskaia, Kirill Fedyanin, Ivan Sukharev, Dmitry Berestnev, and Maxim Panov. Linking bank clients using graph neural networks powered by rich transactional data. *International Journal of Data Science and Analytics*, 12(2):135–145, August 2021.

[241] Mahesh Balan Umaithanu, Vignesh Ravichandran, M. Rohith Srinivaas, and Venkat Subramanian Selvaraj. DeepGRASS: Graph, Sequence and Scaled Embeddings on large scale transactions data. In *2021 Swedish Workshop on Data Science (SweDS)*, pages 1–6, December 2021.

[242] Anish Khazane, Jonathan Rider, Max Serpe, Antonia Gogoglou, Keegan Hines, C. Bayan Bruss, and Richard Serpe. DeepTrax: Embedding Graphs of Financial Transactions. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 126–133, December 2019.

[243] Dawei Cheng, Zhibin Niu, and Liqing Zhang. Delinquent Events Prediction in Temporal Networked-Guarantee Loans. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4):1692–1704, April 2023.

[244] Hamed Amini, Rama Cont, and Andreea Minca. Resilience to Contagion in Financial Networks. *Mathematical Finance*, 26(2):329–365, 2016.

[245] Hamed Amini, Yudong Chen, Andreea Minca, and Xin Qian. Clustering heterogeneous financial networks. *Mathematical Finance*, n/a(n/a), June 2023.

[246] Axel Gandy and Luitgard A. M. Veraart. A Bayesian Methodology for Systemic Risk Assessment in Financial Networks. *Management Science*, 63(12):4428–4446, December 2017.

[247] Anton Braverman and Andreea Minca. Networks of common asset holdings: Aggregation and measures of vulnerability. *Journal of Network Theory in Finance*, August 2018.

[248] Emiliano Marchese, Guido Caldarelli, and Tiziano Squartini. Detecting mesoscale structures by surprise. *Communications Physics*, 5(1):132, May 2022.

[249] KRISTIAN SKREDE GLEDITSCH. Expanded Trade and GDP Data. *Journal of Conflict Resolution*, 46(5):712–724, October 2002.

[250] Marco Bardoscia, Stefano Battiston, Fabio Caccioli, and Guido Caldarelli. Pathways towards instability in financial networks. *Nature Communications*, 8(1):14416, February 2017.

[251] Irena Barjašić, Hrvoje Štefančić, Vedrana Pribičević, and Vinko Zlatić. Causal motifs and existence of endogenous cascades in directed networks with application to company defaults. *Scientific Reports*, 11(1):24028, December 2021.

[252] Ali Caner Türkmen, Gökhan Çapan, and Ali Taylan Cemgil. Clustering Event Streams With Low Rank Hawkes Processes. *IEEE Signal Processing Letters*, 27:1575–1579, 2020.

[253] Alan G. Hawkes. Point Spectra of Some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, October 1971.

[254] ALAN G. HAWKES. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, April 1971.

[255] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes Processes in Finance. *Market Microstructure and Liquidity*, 01(01):1550005, June 2015.

[256] Falk Hüffner, Christian Komusiewicz, Hannes Moser, and Rolf Niedermeier. Isolation concepts for clique enumeration: Comparison and computational experiments. *Theoretical Computer Science*, 410(52):5384–5397, December 2009.

[257] Michael Behrisch and Anusch Taraz. Efficiently covering complex networks with cliques of similar vertices. *Theoretical Computer Science*, 355(1):37–47, April 2006.

[258] Vladimir Boginski, Sergiy Butenko, and Panos M. Pardalos. Mining market data: A network approach. *Computers & Operations Research*, 33(11):3171–3184, November 2006.

[259] Zhihong Zhang, Dongdong Chen, Lu Bai, Jianjia Wang, and Edwin R. Hancock. Graph Motif Entropy for Understanding Time-Evolving Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4):1651–1665, April 2023.

[260] Filipi N. Silva, Cesar H. Comin, Thomas K. DM Peron, Francisco A. Rodrigues, Cheng Ye, Richard C. Wilson, Edwin Hancock, and Luciano

da F. Costa. Modular Dynamics of Financial Market Networks, July 2015.

[261] Mihai Cucuringu, Andrea Pizzoferrato, and Yves Van Gennip. An MBO scheme for clustering and semi-supervised clustering of signed networks. *Communications in Mathematical Sciences*, 19(1):73–109, 2021.

[262] Alberto Natali, Elvin Isufi, Mario Coutino, and Geert Leus. Learning Time-Varying Graphs From Online Data. *IEEE Open Journal of Signal Processing*, 3:212–228, 2022.

[263] Nathanaël Perraudin, Johan Paratte, David Shuman, Lionel Martin, Vassilis Kalofolias, Pierre Vandergheynst, and David K. Hammond. GSPBOX: A toolbox for signal processing on graphs, March 2016.

[264] A. Jalaldoust, K. Hlaváčková-Schindler, and C. Plant. Causal Discovery in Hawkes Processes by Minimum Description Length. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, volume 36, pages 6978–6987, 2022.

[265] Yiran He and Hoi-To Wai. Estimating Centrality Blindly From Low-Pass Filtered Graph Signals. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5330–5334, May 2020.

[266] Rasoul Shafipour, Santiago Segarra, Antonio Marques, and Gonzalo Mateos. Identifying the Topology of Undirected Networks From Diffused Non-Stationary Graph Signals. *IEEE Open Journal of Signal Processing*, 2:171–189, 2021.

[267] Seyed Saman Saboksayr, Gonzalo Mateos, and Mujdat Cetin. Online Graph Learning under Smoothness Priors. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 1820–1824, August 2021.

[268] Gabriele Torri, Rosella Giacometti, and Tomáš Tichý. Network tail risk estimation in the European banking system. *Journal of Economic Dynamics and Control*, 127:104125, June 2021.

[269] Tobias Adrian and Markus K. Brunnermeier. CoVaR. *American Economic Review*, 106(7):1705–1741, July 2016.

[270] Seyed Saman Saboksayr, Gonzalo Mateos, and Mujdat Cetin. Online discriminative graph learning from multi-class smooth signals. *Signal Processing*, 186:108101, September 2021.

[271] José Vinícius de Miranda Cardoso and Daniel P. Palomar. Learning Undirected Graphs in Financial Markets, November 2020.

[272] David Hallac, Youngsuk Park, Stephen Boyd, and Jure Leskovec. Network Inference via the Time-Varying Graphical Lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 205–213, New York, NY, USA, August 2017. Association for Computing Machinery.

[273] Jiaxi Ying, José Vinícius de M. Cardoso, and Daniel P. Palomar. A Fast Algorithm for Graph Learning under Attractive Gaussian Markov Random Fields. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pages 1520–1524, October 2021.

[274] Yiyuan She, Yuejia He, Shijie Li, and Dapeng Wu. Joint Association Graph Screening and Decomposition for Large-Scale Linear Dynamical Systems. *IEEE Transactions on Signal Processing*, 63(2):389–401, January 2015.

[275] Jitendra K. Tugnait. Sparse Graph Learning Under Laplacian-Related Constraints. *IEEE Access*, 9:151067–151079, 2021.

[276] Patrick Danaher, Pei Wang, and Daniela M. Witten. The Joint Graphical Lasso for Inverse Covariance Estimation Across Multiple Classes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(2):373–397, March 2014.

[277] Eduardo Pavez and Antonio Ortega. Generalized Laplacian precision matrix estimation for graph signal processing. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6350–6354, March 2016.

[278] Eduardo Pavez, Hilmi E. Egilmez, and Antonio Ortega. Learning Graphs With Monotone Topology Properties and Multiple Connected Components. *IEEE Transactions on Signal Processing*, 66(9):2399–2413, May 2018.

[279] Jiaxi Ying, José Vinícius de Miranda Cardoso, and Daniel Palomar. Nonconvex Sparse Graph Learning under Laplacian Constrained

Graphical Model. In *Advances in Neural Information Processing Systems*, volume 33, pages 7101–7113. Curran Associates, Inc., 2020.

[280] Yuhua Li, Dongsheng Duan, Guanghao Hu, and Zhengding Lu. Discovering Hidden Group in Financial Transaction Network Using Hidden Markov Model and Genetic Algorithm. In *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 5, pages 253–258, August 2009.

[281] Dongxu Huang, Dejun Mu, Libin Yang, and Xiaoyan Cai. CoDetect: Financial Fraud Detection With Anomaly Feature Detection. *IEEE Access*, 6:19161–19174, 2018.

[282] Yingsheng Ji, Zheng Zhang, Xinlei Tang, Jiachen Shen, Xi Zhang, and Guangwen Yang. Detecting Cash-out Users via Dense Subgraphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 687–697, Washington DC USA, August 2022. ACM.

[283] Yuan Yao, Yicheng Pan, Shaojiang Wang, Hongan Wang, and Waqas Nazeer. Flow-Based Clustering on Directed Graphs: A Structural Entropy Minimization Approach. *IEEE Access*, 8:152579–152591, 2020.

[284] Fan R. K. Chung and Linyuan Lu. *Complex Graphs and Networks*. American Mathematical Soc., 2006.

[285] Martin Harrigan, Daniel Archambault, Pádraig Cunningham, and Neil Hurley. EgoNav: Exploring networks through egocentric spatializations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 563–570, Italy, May 2012. ACM.

[286] Christos Tsigkanos, Alessio Arleo, Johannes Sorger, and Schahram Dustdar. How Do Firms Transact? Guesstimation and Validation of Financial Transaction Networks with Satisfiability. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 15–22, July 2019.

[287] Shuaicheng Zhang, Yada Zhu, and Dawei Zhou. TGEditor: Task-Guided Graph Editing for Augmenting Temporal Financial Transaction Networks. In *4th ACM International Conference on AI in Finance*, pages 219–226, Brooklyn NY USA, November 2023. ACM.

[288] Dawei Zhou, Lecheng Zheng, Jiawei Han, and Jingrui He. A Data-Driven Graph Generative Model for Temporal Interaction Networks.

In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pages 401–411, New York, NY, USA, August 2020. Association for Computing Machinery.

[289] Liming Zhang, Liang Zhao, Shan Qin, and Dieter Pfoser. TG-GAN: Continuous-time Temporal Graph Generation with Deep Generative Models. In *Proceedings of the Web Conference 2021*, pages 2104–2116, April 2021.

[290] Vicente Balmaseda, María Coronado, and Gonzalo de Cadenas-Santiago. Predicting systemic risk in financial systems using Deep Graph Learning. *Intelligent Systems with Applications*, 19:200240, September 2023.

[291] Paolo Barucca and Fabrizio Lillo. The organization of the interbank network and how ECB unconventional measures affected the e-MID overnight market. *Computational Management Science*, 15(1):33–53, January 2018.

[292] Desheng Ma and Shawn Mankad. CP-Squared: A method for change point detection in core–periphery networks. *Expert Systems with Applications*, 196:116660, June 2022.

[293] K.S. Xu and A.O. Hero III. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal on Selected Topics in Signal Processing*, 8(4):552–562, 2014.

[294] Bogdan Dumitrescu, Andra Baltoiu, and Stefania Budulan. Anomaly Detection in Graphs of Bank Transactions for Anti Money Laundering Applications. *IEEE Access*, 10:47699–47714, 2022.

[295] H. Elsinger, A. Lehar, and M. Summer. Network models and systemic risk assessment. *Handbook on Systemic Risk*, pages 287–305, 2013.

[296] Christian Upper and Andreas Worms. Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849, August 2004.

[297] Christian Upper. Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125, August 2011.

[298] Kartik Anand, Ben Craig, and Goetz Von Peter. Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance*, 15(4):625–636, April 2015.

[299] Tiziano Squartini and Diego Garlaschelli. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):083001, August 2011.

[300] Tarik Roukny and Stefano Battiston. A Network Analysis of the Evolution of the German Interbank Market, 2014.

[301] Guido Caldarelli, Alessandro Chessa, Fabio Pammolli, Andrea Gabrielli, and Michelangelo Puliga. Reconstructing a credit network. *Nature Physics*, 9(3):125–126, March 2013.

[302] Mahboobeh Ezzodin, Alireza Bahiraie, and A. K. M. Azhar. Modelling the Cognitive Financial Group Interconnected Risk Network. *International Journal of Nonlinear Analysis and Applications*, 11(1):259–276, April 2020.

[303] Craig H. Furfine. Interbank Exposures: Quantifying the Risk of Contagion. *Journal of Money, Credit and Banking*, 35(1):111–128, 2003.

[304] Jiangtong Li, Ziyuan Zhou, Jingkai Zhang, Dawei Cheng, and Changjun Jiang. HFTCRNet: Hierarchical Fusion Transformer for Interbank Credit Rating and Risk Assessment. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7):13006–13020, 2025.

[305] Seonkyu Lim, Jeongwhan Choi, Jaehoon Lee, and Noseong Park. FrAug: Enhanced Fraud Detection in Interbank Transfers via Augmented Account Features. In *Proc. IEEE Int. Conf. Big Data Smart Comput., BIGCOMP*, number 2025 in 64353, pages 199–206, Malaysia, 2025. IEEE.

[306] Junyi Liu, Dawei Cheng, and Changjun Jiang. Preventing Attacks in Interbank Credit Rating with Selective-aware Graph Neural Network. In *IJCAI Int. Joint Conf. Artif. Intell.*, volume 2023-August, pages 6085–6093. International Joint Conferences on Artificial Intelligence, 2023.

[307] Junyi Liu, Dawei Cheng, and Changjun Jiang. Preferential Selective-Aware Graph Neural Network for Preventing Attacks in Interbank Credit Rating. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):11414–11427, 2025.

[308] Jiehao Tang, Wenjun Wang, Dawei Cheng, Hui Zhao, and Changjun Jiang. Dual Pairwise Pre-training and Prompt-tuning with Aligned Prototypes for Interbank Credit Rating. In *WWW - Proc. ACM Web*

Conf., pages 5234–5243. Association for Computing Machinery, Inc, 2025.

[309] Simone Romano, Davide Fucci, Giuseppe Scanniello, Maria Teresa Baldassarre, Burak Turhan, and Natalia Juristo. Researcher Bias in Software Engineering Experiments: A Qualitative Investigation. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pages 276–283, August 2020.

[310] Kassiani Nikolopoulou. What Is Confirmation Bias? | Definition & Examples. https://www.scribbr.com/research-bias/confirmation-bias/, September 2022.

[311] Jacqueline M. McGrath, Roy E. Brown, and Haifa A. Samra. Before You Search the Literature: How to Prepare and Get the Most Out of Citation Databases. *Newborn and Infant Nursing Reviews*, 12(3):162–170, September 2012.

[312] Pritha Bhandari. Observer Bias | Definition, Examples, Prevention. https://www.scribbr.com/research-bias/observer-bias/, December 2021.

[313] Francis C. Harris and Benjamin B. Lahey. Subject reactivity in direct observational assessment: A review and critical analysis. *Clinical Psychology Review*, 2(4):523–538, January 1982.

[314] Gábor Csárdi, Tamás Nepusz, Kirill Müller, Szabolcs Horvát, Vincent Traag, Fabio Zanini, and Daniel Noom. Igraph for R: R interface of the igraph library for graph theory and network analysis. Zenodo, January 2025.

[315] Hadley Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer international publishing, Cham, second edition edition, 2016.

[316] Justin Zhan, Sweta Gurung, and Sai Phani Krishna Parsa. Identification of top-K nodes in large networks using Katz centrality. *Journal of Big Data*, 4(1):16, May 2017.

[317] B.V. Almende and Thieurmel Benoit. visNetwork: Network Visualization using 'vis.js' Library, September 2022.

[318] Mark D. Humphries and Kevin Gurney. Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLOS ONE*, 3(4):e0002051, April 2008.

[319] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1):1017, March 2019.

[320] Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, Martina Morris, Li Wang, Kirk Li, Skye Bender-deMoll, Chad Klumb, Michał Bojanowski, Ben Bolker, Christian Schmid, Joyce Cheng, Arya Karami, and Adrien Le Guillou. Ergm, June 2025.

[321] Ruth M. Hummel, David R. Hunter, and Mark S. Handcock. Improving Simulation-Based Algorithms for Fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4):920–939, October 2012.

[322] Douglas C. Montgomery and George C. Runger. *Applied Statistics and Probability for Engineers*. Wiley, New York, 3rd ed edition, 2003.

[323] Joe Hasell. Measuring inequality: What is the Gini coefficient? *Our World in Data*, June 2023.

[324] Google. Fairness: Types of bias | Machine Learning. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias, October 2024.

[325] Claudia Champagne. The international syndicated loan market network: An "unholy trinity"? *Global Finance Journal*, 25(2):148–168, 2014.

[326] Iñaki Aldasoro and Andreas Barth. Syndicated Loans and CDS Positioning. *SSRN Electronic Journal*, 2017.

[327] Maria Basovskaya. European Syndicated Loan Interconnectedness: Constituents and Its Contribution to Systemic Risk. Master's thesis, Erasmus University Rotterdam School of Economics, August 2018.

[328] Miguel A Ferreira and Pedro Matos. Universal banks and corporate control: Evidence from the global syndicated loan market. *European Central Bank ECB-CFS Research Network*, 1066, July 2009.

[329] Renaud Beaupain and Alain Durré. Inferring trading dynamics for an OTC market: The case of the euro area overnight money market. *Quantitative Finance*, 11(9):1285–1295, September 2011.

[330] Renaud Beaupain and Alain Durré. Nonlinear Liquidity Adjustments in the Euro Area Overnight Money Market, November 2012.

[331] Mark Strembeck. Stochastic Graph Models, 2022.

[332] Dimitris Achlioptas, Aaron Clauset, David Kempe, and Cristopher Moore. On the Bias of Traceroute Sampling; or, Power-law Degree Distributions in Regular Graphs, March 2006.

[333] Yogesh S. Virkar. *Power-Law Distributions and Binned Empirical Data*. PhD thesis, University of Colorado at Boulder, 2012.

[334] Serkan Karadas and Nilufer Ozdemir. Does Public Corruption Affect Bank Failures? Evidence from the United States. *Journal of Risk and Financial Management*, 16(10):451, October 2023.

[335] Michael Schweinberger, Pavel N. Krivitsky, Carter T. Butts, and Jonathan Stewart. Exponential-Family Models of Random Graphs: Inference in Finite-, Super-, and Infinite Population Scenarios. *Statistical Science*, 35(4), November 2020.

[336] Jian Cai. Competition or Collaboration? The Reciprocity Effect in Loan Syndication, April 2010.

[337] Jeffrey H. Harris, Ioannis Spyridopoulos, Morad Zekhnini, and Celso Brunetti. The Role of Networks in Loan Syndicate Markets, January 2024.

[338] Gabjin Oh and A.-young Park. Lending Diversification and Interconnectedness of the Syndicated Loan Market. *Frontiers in Physics*, 8, January 2021.

[339] Abhimanyu Gupta, Sotirios Kokas, and Alexander Michaelides. Credit Market Spillovers: Evidence from a Syndicated Loan Market Network. *SSRN Electronic Journal*, 2017.

[340] Aaron Clauset, Ellen Tucker, and Matthias Sainz. The Colorado Index of Complex Networks. https://icon.colorado.edu/, 2016.

[341] Angelo Mele. Estimating Network Models using Neural Networks, February 2025.

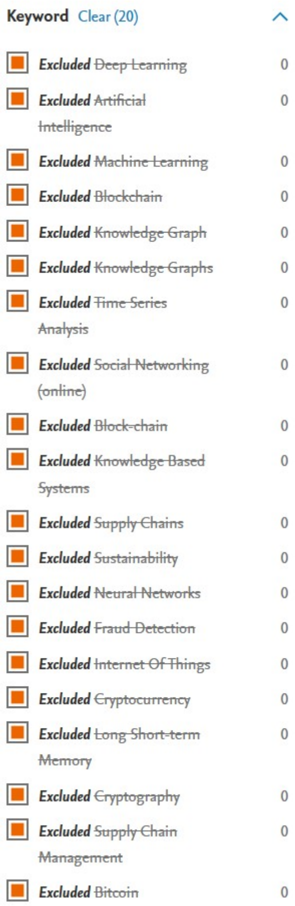
[342] Masayasu Kanno. Exploring risks in syndicated loan networks: Evidence from real estate investment trusts. *Economic Modelling*, 115:105953, October 2022.

A Appendix A

This Appendix A includes additional information for the literature part of this thesis.

A.1 Scopus Keywords

The complete list of scopus keyword excluded during the search



A.2 Literature Table

A list of 100 papers of the set of the literature that was retrieved for the literature review with notes and additional information can be found at the authors Notion. Note that this is not the complete list of references but only details the first 100 systematically collected literature for the literature review section.

The complete dataset used for analysis can be found in the GitHub repository mentioned below.

A.3 Quantitative Content Analysis

This section includes the supplementary information for qualitative analysis, namely the coding. Exceptions to the coding rule of start set papers are written with reference to their correct categorization, as in "(see relevant section)". The six different coding categories of the literature later became the six subsections of 4.

The full Github repository for the literature part of the thesis can be found on Github or Zenodo:

- https://github.com/maroan-ls/financialnetworks_SLR
- https://doi.org/10.5281/zenodo.15106185

It includes a dataset showing the results of the qualitative literature analysis.

B Appendix B

This Appendix Bincludes supplementary notes and materials for the network part of the thesis.

B.1 Data Collection and Processing

LSEG Ex-Revinitiv Eikon The original downloaded data for the network is also available through the closed-source Refinitiv (now LSEG) Eikon database, which is part of the London Stock Exchange Group (LSEG) Data & Analytics software package. More information on Eikon can be obtained here: https://www.lseg.com/en/data-analytics/products/eikon-trading-software.

More information on the Loan Pricing Corporation, from which the original data was gathered can be found here: https://www.lseg.com/en/data-analytics/investment-banking/lpc#overview

Retrieval Replication Steps The data was accessed through a university library terminal. Here are the replication steps to get to the data: - Access the Loans App for Eikon. - Global Loans -> Loans -> Instruments -> Credit Loans -> Loans (This excludes Corporate Bonds, CDS, Securitized Products, and Sukuk). The total Open Loans accessible at the time of request (14.11.24) were 92243. The search was filtered through: *Sector: Include*

*Investment Management & Fund Operators NEC, Banks NEC, Investment Management, Financial & Commodity Market Operators & Service Providers NECS, Hedge Funds, Mutual Funds NEC, Transaction & Payment Services OR Borrower Contains *Bank*, *banc*, *fund*, *financ*.*

This ensures that if a sector is not given or a specific industry sector is applied, but the lender is still a financial entity, it remains in the dataset.

B.2 Supplementary Code Material for Data Processing and Model Fitting

A complete replication package is found on Github (also published on Zenodo):

- https://github.com/maroan-ls/synthetic_networks
- https://doi.org/10.5281/zenodo.15106166

The replication package is covered by the `Creative Commons Attribution 4.0 International` (CC BY 4.0) license.

Data processing is done in Python, with the rest of the network analysis conducted in R. The repository includes the Readme, which explains the order of execution of code files, and alternatively the orchestration file for the important R files to replicate the analysis.

It also includes information on the subset of columns used for the network analysis part in R. The role mapping can be found in the Python data processing file.

B.3 Additional Figures and Tables of Network Design

Table 18 shows the frequency of each mapped role for each GSIB. The lender roles are presented after role mapping has taken place. The repository includes a CSV all 16

B.4 Additional Figures and Tables of the Network Results

The following Figure 13 and Figure 14 are the degree distributions of the baseline graphs.

Table 19 and 20 provide extra metrics for the baseline results.

Figure 13: Degree Distribution and Cumulative Degree Distribution of the ER graph

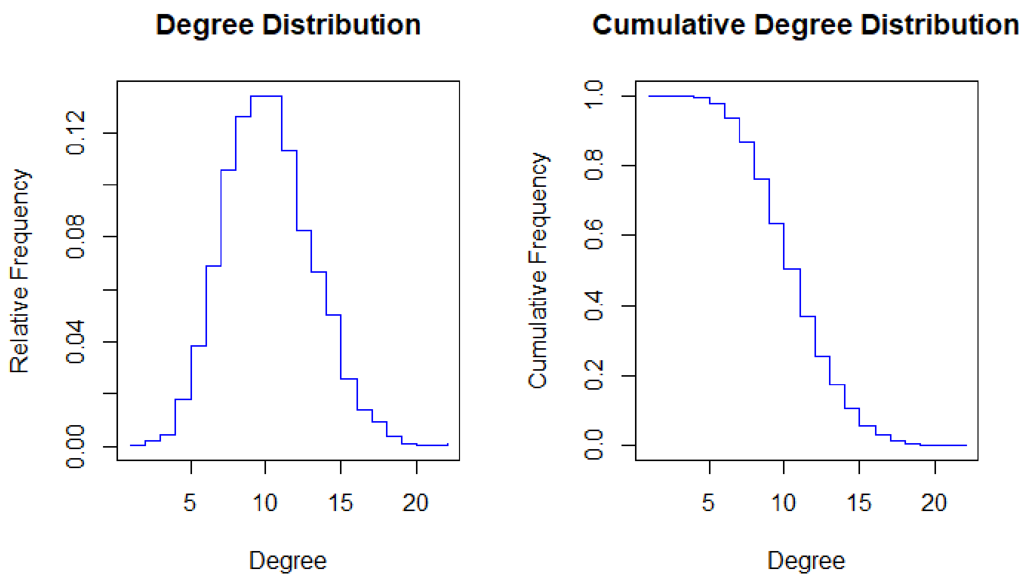


Table 18: GSIBs and their Roles in the Data

Lender	Arranger	Agent	Participant	Sole Lender
JP Morgan Chase	286	258	104	2
Citi	283	153	84	3
HSBC	564	61	119	1
AB China	175	3	30	0
Bank Of America	277	305	88	0
Bank Of China	881	30	103	7
Barclays	221	60	53	0
BNP Paribas	673	38	86	26
China Construction Bank	220	5	31	1
Deutsche Bank	371	28	64	4
Goldman Sachs	202	68	69	0
Credit Agricole	582	32	133	4
ICB China	378	4	55	1
MUFG	728	98	113	21
UBS	88	19	24	0
Bank Of Communications	266	4	27	0
BNY	30	9	56	0
BPCE	282	21	48	2
ING Bank	439	54	41	5
Mizuho	682	62	63	12
Morgan Stanley	188	58	60	0
Royal Bank Of Canada	162	82	69	1
Banco Santander	233	24	36	8
Societe Generale	495	12	95	6
Standard Chartered	370	5	46	1
State Street Bank	8	16	44	0
SMBC	1011	78	100	12
Toronto Dominion Bank	58	66	52	0
Wells Fargo	134	194	61	0

Order matches the 2024 FSB release (alphabetical within capital-buffer buckets). Bucket labels are omitted.
The lender roles are presented after role mapping has taken place. The repository in Appendix B.2 includes a CSV with all non-standard roles.

Figure 14: Degree Distribution and Cumulative Degree Distribution of the SW graph

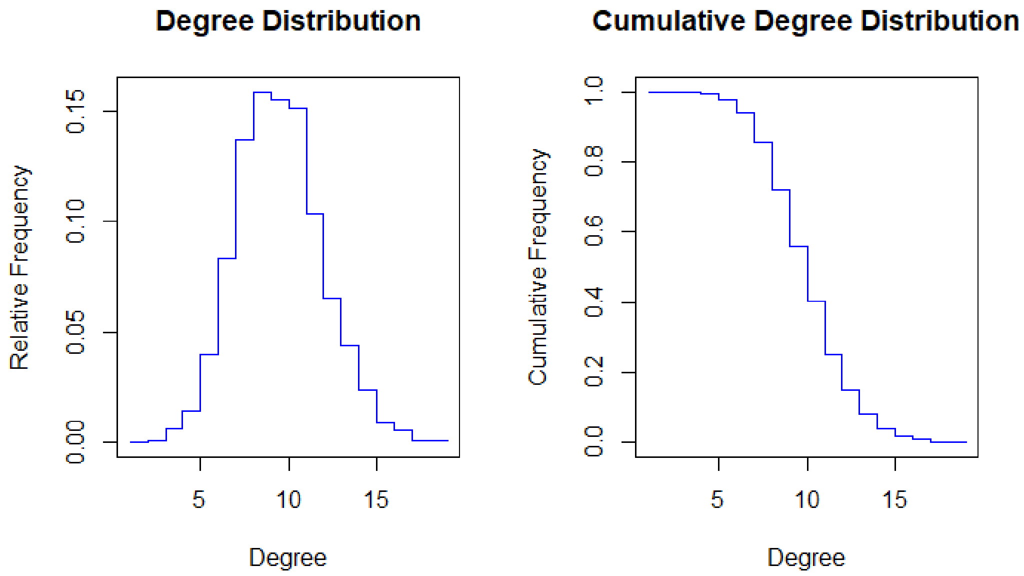


Table 19: Additional structural metrics for the network G_{LCC} ($N = 3\,674$) and density-matched baseline models.

Metric	G_{LCC}	ER	WS	BA
Tail exponent $\hat{\alpha}$	2.342	12.976	25.539	2.320
Kolmogorov–Smirnov D_{KS}	0.023	0.016	0.053	0.027
Gini coefficient G	0.669	0.189	0.171	0.448
Small-world index σ	17.08	1.04	4.73	4.84
Global efficiency E_{glob}	0.284	0.259	0.247	0.394
Local efficiency E_{loc}	0.768	0.614	0.625	0.741

Figure 15: Degree Distribution and Cumulative Degree Distribution of the BA graph

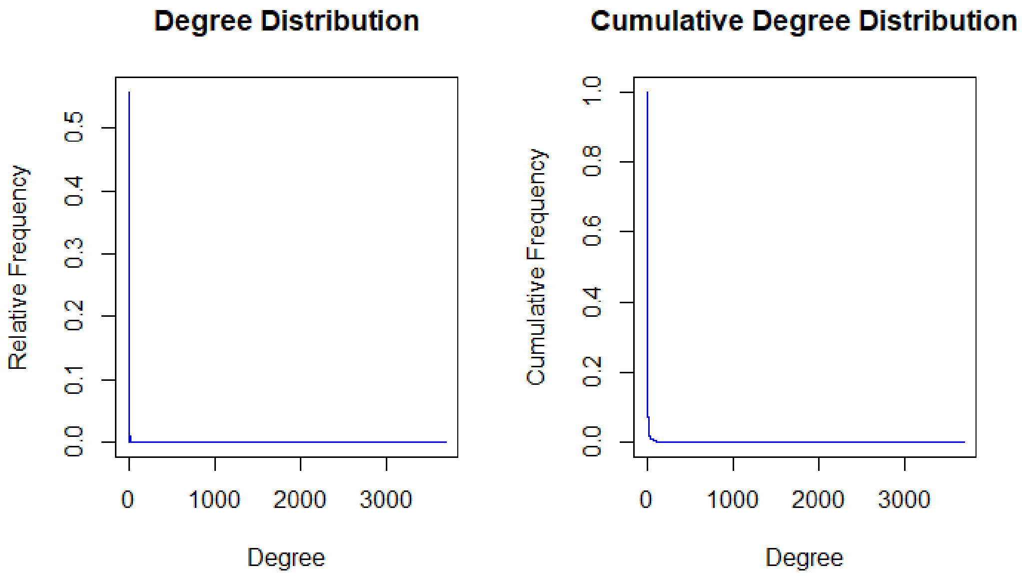


Table 20: Additional structural metrics for the GSIB-W_2 network ($N = 38$) and baseline models.

Metric	GSIB-W_2	ER_{sub}	WS_{sub}	BA_{sub}
Tail exponent $\hat{\alpha}$	2.738	7.156	7.850	2.150
Kolmogorov–Smirnov D_{KS}	0.110	0.068	0.084	0.093
Gini coefficient G	0.288	0.141	0.128	0.450
Small-world index σ	1.34	0.96	1.07	0.85
Global efficiency E_{glob}	0.666	0.654	0.660	0.568
Local efficiency E_{loc}	0.730	0.714	0.720	0.908

B.5 ERGM MCMC Diagnostics

Additional output of `mcmc.diagnostics`: chains are heuristically considered good, if chains in the first column of Figure 16 and 17 are oscillating and not linear, and if the bell-curve shows an approximate normal distribution around 0 for every ERGM coefficient.

Figure 16: MCMC Diagnostics: output of `mcmc.diagnostics(fit)`. Each row is an ERGM coefficient. On the left are the parallel MCMC chains, on the right the approximated distribution

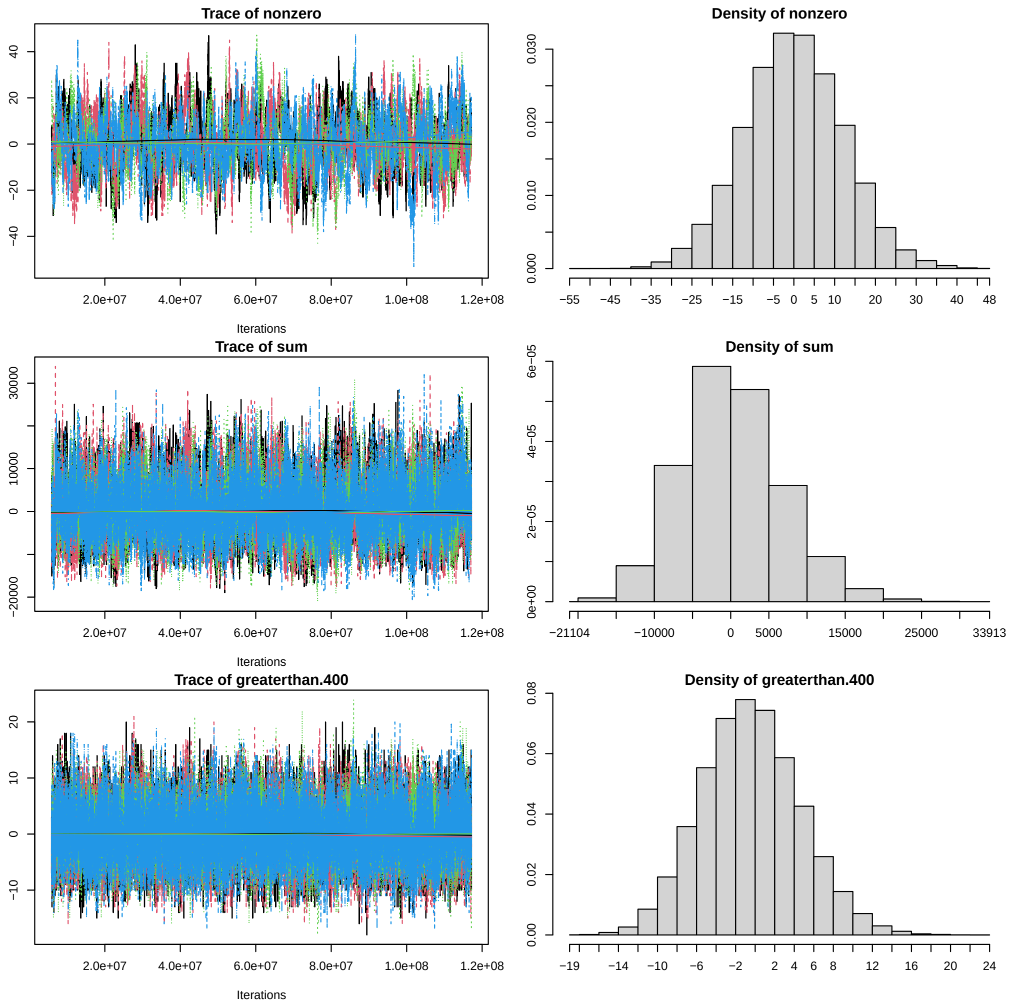


Figure 17: MCMC Diagnostics: output of `mcmc.diagnostics(fit)`. Each row is an ERGM coefficient. On the left are the parallel MCMC chains, on the right the approximated distribution