Journal Pre-proof

Reconstructing and Stress Testing Credit Networks

Amanah Ramadiah, Fabio Caccioli, Daniel Fricke

PII: S0165-1889(19)30212-X
DOI: https://doi.org/10.1016/j.jedc.2019.103817
Reference: DYNCON 103817

To appear in: *Journal of Economic Dynamics & Control*

Received date: 15 October 2018
Revised date: 25 October 2019
Accepted date: 2 December 2019

Please cite this article as: Amanah Ramadiah, Fabio Caccioli, Daniel Fricke, Reconstructing and Stress Testing Credit Networks, *Journal of Economic Dynamics & Control* (2019), doi: https://doi.org/10.1016/j.jedc.2019.103817

Reconstructing and Stress Testing Credit Networks

Amanah Ramadiah*

University College London, Department of Computer Science, United Kingdom

Fabio Caccioli

University College London, Department of Computer Science, United Kingdom
London School of Economics & Political Science, Systemic Risk Centre, United Kingdom

London Mathematical Laboratory, United Kingdom

Daniel Fricke

Deutsche Bundesbank, Directorate General Financial Stability, Germany

University College London, Department of Computer Science, United Kingdom

Abstract

Financial networks are an important source of systemic risk, but often only partial network information is available. In this paper, we use data on bank-firm credit relationships in Japan and conduct a horse race between different network reconstruction methods in terms of their ability to reproduce the actual credit networks. We then compare the different reconstruction methods in terms of their implied levels of systemic risk based on a standard model of price-mediated contagion. We find that the observed credit network displays relatively high levels of systemic risk compared with most reconstruction methods. Lastly, we explore whether different policies can improve the robustness of the system.

Keywords: network reconstruction; stress testing; systemic risk; bipartite credit network; aggregation level.

JEL Classification: G11, G20, G21, G28, G32

*Corresponding author. Email: a.ramadiah@cs.ucl.ac.uk. Address: University College London, Dept. Computer Science, 66-72 Gower St, London WC1E 6EA, United Kingdom.

1 Introduction

The 2007-09 financial crisis has brought the interconnectedness of the financial system to light, and financial networks have been identified as an important source of systemic risk. Accordingly, the regulatory framework has taken a more macroprudential perspective to maintain the stability of the system as a whole. For example, Basel III includes capital surcharges for systemically important financial institutions.

Stress tests are an important tool to assess the vulnerability of a given financial network. To this end, detailed data on (direct or indirect) interactions between individual financial institutions is needed. However, it is difficult to collect such data in full and to make them readily available to researchers (e.g., due to data confidentiality), such that we generally do not have complete information about financial networks. For example, Haldane (2015) suggests that even among the world's largest banks the collection of interbank exposure data is partial, and even regulators often do not have complete information (Glasserman and Young (2016)). In response, several data collection initiatives have been proposed, but granular interaction-specific data generally remain unavailable (Anand et al. (2017)).

Finding accurate reconstruction methods for financial networks from partial information is therefore an important topic. Most of the existing work focuses on the case of interbank credit networks (Squartini et al. (2017); Gandy and Veraart (2017); Anand et al. (2017)). Over the last decade, common asset holdings (or overlapping portfolios) have been identified as an important source of systemic risk via price-mediated contagion (Shleifer and Vishny (2011); Caccioli et al. (2014); Greenwood et al. (2015); Cont and Wagalath (2016); Gualdi et al. (2016); Lillo and Pirino (2015); Fricke and Fricke (2019)). The idea is (that, when they suffer a decline in their investment portfolios, leveraged investors often have to liquidate (parts of) their investments (Adrian and Shin (2010)). Such liquidations can have systemic effects when asset sales are synchronized among many investors, potentially leading to fire sale contagion dynamics. Empirical evidence suggests that fire sales occur in many different markets (see, e.g., Pulvino (1998) for real assets, Coval and Stafford (2007) for equities, and Ellul et al. (2011) for corporate bonds), which can result in contagious dynamics between asset classes (see, e.g., Manconi et al. (2012)).[1] Hence, understanding the structure and dynamics of common asset holdings is important (Fricke (2016)), but often hampered by data availability.

In this paper, we focus on reconstructing and stress testing bipartite credit networks using detailed micro-data on bank-firm credit interactions in Japan for the period 1980 - 2010. We explore the performance of several network reconstruction methods at different aggregation levels along two different dimensions. First, we look at their capability to reproduce the topological features of the observed credit networks. This part of the paper is closest to some recent works on unipartite interbank networks (e.g., BIS (2015), Anand et al. (2017), and Mazzarisi and Lillo (2017)). Different reconstruction methods require different amounts of information as inputs, and we

[1]Fire sales are also dangerous because they provide an incentive for banks to hoard liquidity, a behavior that can potentially lead to a complete freeze of the financial system (Diamond and Rajan (2011); Gale and Yorulmazer (2013)).

aim to understand how adding such information affects a method's performance, since one would expect that methods that take more information into account should be able to reproduce the network more accurately. Interestingly, we find that this is not always the case. Overall, there is no single "best" reconstruction method - it depends on the assumed criterion of interest.

We then test each method's ability to reproduce observed levels of systemic risk. For this purpose, we use the fire-sale stress test of Huang et al. (2013) and apply it to the actual and the reconstructed credit networks. To the best of our knowledge, this is the first paper to conduct a horse race of bipartite network reconstruction methods in terms of their implied levels of systemic risk.[2] Our main findings are as follows: first, we identify a significantly negative time trend for the observed systemic risk levels of the Japanese banking system, suggesting that the system has become less vulnerable to systemic asset liquidations over time. Second, in many instances the actual credit networks display the highest levels of systemic risk, at least for the most disaggregated bank-firm interactions. In other words, many reconstruction methods tend to underestimate systemic risk. This is remarkable given that the reconstruction methods under study here can generate completely different network architectures; for example, the MaxEntropy (MinDensity) approach yields a maximally (minimally) connected credit network. Moreover, we find that the network aggregation level can affect the performance of the different reconstruction methods.

Lastly, given that the observed credit networks tend to display relatively high levels of systemic risk compared to most reconstruction methods, we explore different policies (such as merging or breaking-up banks, or leverage caps) in order to improve the robustness of the system. Our main finding is that no single policy can reduce the systemic risk level of the actual network to that of the most stable reconstruction method. Nevertheless, we find that leverage caps and bank mergers could improve the robustness of the network. This finding is driven by the fact that the largest banks in our sample tend to be less leveraged. Therefore, merging those banks results in a very large, but moderately leveraged bank which is less likely to spread shocks through the system.

Overall, this paper contributes to different strands of literature: first, we add to the growing literature on reconstructing financial networks from partial information (Squartini et al. (2017); Gandy and Veraart (2017); Anand et al. (2017); Squartini et al. (2018)). For the case of bipartite networks we are only aware of the works of Di Gangi et al. (2018) and Squartini et al. (2017). Given that most reconstruction methods have been designed for the case of unipartite credit networks, we adjust some of these methods to the case of bipartite networks. Second, we contribute to the literature on systemic risk assessment by performing stress tests both for the actual and the reconstructed credit networks in Japan. Lastly, we contribute to the literature that explores the effects of aggregation on stress test results. For example, Hale et al. (2015) study the optimal aggregation level for stress testing models on macroeconomic variables, and they find that the aggregation level matters. We obtain a similar conclusion based on a completely different stress testing approach.

[2] Some related papers for the case of unipartite interbank networks are Mistrulli (2011), Anand et al. (2015), and Gandy and Veraart (2017).

The remainder of this paper is structured as follows: Section 2 defines the credit network at different aggregation levels, and section 3 briefly describes the dataset. In section 4, we explore the performance of network reconstruction methods in terms of their ability to match the observed credit network topology. In section 5, we look at the capability of each methods to reproduce the observed levels of systemic risk. In section 6, we analyze different policy measures in order to improve the robustness of the system. Section 7 summarizes the main findings and concludes.

2 The Credit Network

Let us start by defining the credit network at different aggregation levels. The most granular data (disaggregated level) is the credit interaction network between banks and firms. The baseline credit network consists of two distinct sets of nodes, where the first set contains a total number of n^B nodes (banks), and the second set a total of n^F nodes (firms). A link exists between a bank and a firm when there is a credit relationship between the two. The network is bipartite, since links can only arise between banks and firms.

This credit network can be represented as a rectangular matrix of size $(n^B \times n^F)$, which we denote by $\mathbf{W}$. An element w_{ij} of this matrix represents the total value of credit extended by bank i to firm j at a given point in time.[3] The value of w_{ij} can thus be seen as a measure of link intensity. The total loan volume can be calculated as

$$v = \sum_i \sum_j w_{ij}.$$

For what follows, it is also useful to define the *strengths* of banks and firms as their corresponding loan volumes:

$$s_i^B = \sum_j w_{ij}$$

and

$$s_j^F = \sum_i w_{ij}$$

for bank i and firm j, respectively.

We also define the binary adjacency matrix, $\mathbf{B}$, where each element $b_{ij} = 1$ if $w_{ij} > 0$ and zero otherwise. From the binary network matrix, we calculate the total number of links

$$m = \sum_i \sum_j b_{ij}.$$

In addition, we define the *degrees* of banks and firm as their corresponding number of connections:

$$k_i^B = \sum_j b_{ij}$$

[3] We drop time subscripts in the following, but it should be clear that matrix $\mathbf{W}$ changes over time.

and

$$k_j^F = \sum_i b_{ij}.$$

Following Fricke and Roukny (2018), we also look at an aggregated version of the credit (bank-industry) network, which we denote by $\mathbf{W^I}$. In this case, the second set of nodes is defined based on firms' industry affiliations, with a total number of n^I industries. We can represent firms' industry affiliations using a new matrix $\mathbf{A}$ of dimension $(n^F \times n^I)$, where $a_{jk} = 1$ if firm j is affiliated with industry k.[4] Given this, $\mathbf{W^I}$ can be obtained by multiplying $\mathbf{W}$ with $\mathbf{A}$. In line with the definitions for the original bank-firm credit network, we can define the same network indicators (strength and degree sequences, respectively) for the aggregated network.

Note that an important reason for also exploring the aggregated networks is that (at least some rough) information on banks' investments in different industries/asset classes should be more easily available than detailed microdata on asset-specific investments. From this perspective, the analyses based on the aggregated networks are likely to be most relevant for researchers that have only relatively coarse information on banks' asset portfolios.

Finally, we consider an intermediate level in which we apply the network reconstruction methods at the disaggregated level (bank-firm) and then aggregate the network according to firms' observed industry affiliations (thus giving us a different bank-industry credit network). This particular aggregation level is of interest in the case when there is sufficient data to perform network reconstruction at a more granular level (e.g., firm level), but the network needs to be analyzed at a more aggregated level (e.g., sector level), for instance because of confidentiality issues that prevent reporting results associated with individual institutions. We denote the intermediate aggregation level as $\mathbf{W} \to \mathbf{W^I}$ and calculate the same network indicators also as for the other levels. We summarize the three different aggregation levels in Table 1.

Aggregation level	**Network reconstruction**	**Systemic risk analysis**
Disaggregated	disaggregated	disaggregated
Aggregated	aggregated	aggregated
Intermediate	disaggregated	aggregated

Table 1: Summary of the three different aggregation levels. At the intermediate level, we perform the network reconstruction at the disaggregated data, and conduct the systemic risk analysis at the aggregated version of that reconstructed network.

[4] In our dataset, each firm is only affiliated with its major industry. In principle, one could allow for multiple industry affiliations, in which case a_{jk} would represent the fraction of firm j's sales in industry j.

3 Data

In this paper we use historical data on bank-firm credit interactions in Japan from the Nikkei NEEDS database for the period 1980 - 2013.[5] The database provides extensive accounting and loan information for all listed companies in Japan, and since 1996 it also covers firms traded in the JASDAQ (OTC) market. The dataset contains information on firms outstanding loan volumes from each lender at the end of the firms fiscal year, based on survey data (compiled by Nikkei Media Marketing, Inc.). We use the sum of short- and long-term borrowing in everything that follows. Table 2 shows some summary statistics in terms of the size and connectivity of the credit network at different aggregation levels over time.[6] Given that our analyses are computationally intensive, we restrict ourselves to the years of data as shown in the first column of Table 2.[7]

Panel A - Disaggregated

Year	Size	v ($\times 10^{13}$)	Density	$\bar{k}^B$	$\bar{k}^F$	r	C ($\times 10^{-2}$)	NODF
1980	151 × 1386	3.395	0.093	128.377	13.986	-0.299	0.272	0.441
1985	148 × 1443	4.350	0.088	127.770	13.105	-0.290	0.251	0.437
1990	148 × 1443	6.249	0.081	125.762	12.236	-0.306	0.218	0.427
1995	145 × 1734	7.031	0.081	140.938	11.785	-0.302	0.212	0.444
1996	147 × 2523	7.525	0.070	175.782	10.242	-0.292	0.141	0.406
2000	135 × 2607	5.987	0.061	160.304	8.301	-0.273	0.091	0.387
2005	123 × 2569	2.469	0.042	109.423	5.184	-0.272	0.029	0.322
2010	116 × 2296	2.814	0.042	96.474	4.874	-0.215	0.028	0.359

Panel B - Aggregated

Year	Size	v ($\times 10^{13}$)	Density	$\bar{k}^B$	$\bar{k}^I$	r	C	NODF
1980	151 × 33	3.395	0.516	17.033	77.939	-0.336	0.192	0.824
1985	148 × 33	4.350	0.500	16.507	74.030	-0.344	0.181	0.823
1990	151 × 33	6.250	0.498	16.424	75.152	-0.351	0.181	0.810
1995	145 × 33	7.031	0.518	17.090	75.091	-0.341	0.195	0.834
1996	147 × 34	7.526	0.536	18.238	78.853	-0.344	0.206	0.852
2000	135 × 34	5.987	0.508	17.260	68.529	-0.349	0.177	0.839
2005	123 × 34	2.470	0.488	16.585	60.000	-0.340	0.151	0.822
2010	116 × 34	2.814	0.461	15.664	53.441	-0.330	0.134	0.819

Table 2: Properties of the credit networks at different aggregation levels over time. Panel A shows the properties of $\mathbf{W}$. Panel B shows the properties of $\mathbf{W^I}$. $\bar{k}^B$ and $\bar{k}^{F(I)}$ correspond to the average degree of banks and firms (industries) respectively. As defined in the main text, r denotes the assortativity, C denotes the clustering coefficient, and NODF denotes the nestedness.

In Table 2, we present several basic network characteristics of our dataset. Specifically, we show the assortativity, the clustering coefficient, and the nestedness. In the following, we define the measures of those characteristics for the bank-firm network at

[5] See `https://www.nikkeieu.com/needs/needs_data.html` for details.

[6] A detailed explanation of the dataset, summary statistics, and a brief history of the Japanese financial system can be found in Fricke and Roukny (2018).

[7] Given that bank-firm interactions are highly persistent, the structure of the credit network is quite stable over time.

the disaggregated level. In line with these definitions, we can define the same measures for the bank-industry network at the aggregated level.

Assortativity is the tendency of banks to connect to firms (industries) with similar characteristics, and vice versa. We define assortativity, r, as the Pearson correlation coefficient of the degrees of connected banks and firms. Note that r lies in the range $[-1, 1]$ in which positive value indicates an assortative network while negative value denotes a disassortative network. A network is said to be assortative when high degree banks (low degree banks) are connected to other high degree firms (low degree firms) on average. Meanwhile, a network is said to be disassortative when high degree banks (low degree banks) are connected to other low(er) degree firms (high(er) degree firms) on average. Here we find that the networks are generally disassortative, both at the disaggregated level and the aggregated level. This means that low-degree banks and low-degree firms rarely interact with each other.

The *clustering coefficient* measures the degree to which nodes in a network tend to form clusters. In a unipartite network it is usually defined as the number of observed triangles (three closed connected nodes) relative to the maximum possible number of triangles. Since our network is bipartite, links can only exist between different sets of nodes (banks and firms/industries), thus triangles can not be formed. Therefore, following Zhang et al. (2008), we consider squares instead of triangles as the basic cycle here, such that the local clustering coefficient is defined as the ratio between the number of observed squares relative to the maximum possible number of squares,

$$C_{mn}(i) = \frac{q_{imn}}{(k_m - \eta_{imn}) + (k_n - \eta_{imn}) + q_{imn}} \tag{1}$$

where m and n are a pair of neighbors of node i (see Figure 1 for an illustration), q_{imn} is the number of squares which include these three nodes, while $\eta_{imn} = 1 + q_{imn}$.

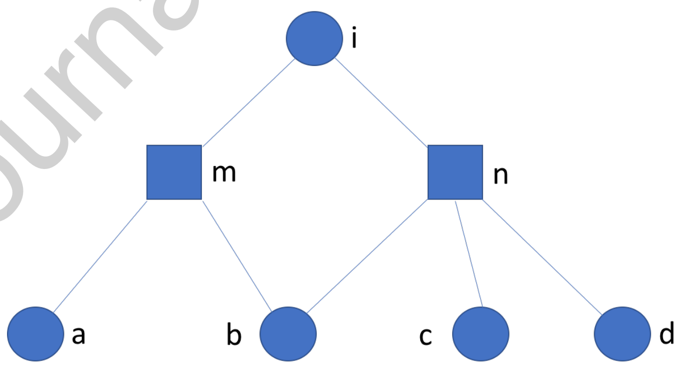


Figure 1: Illustration of calculating the observed and the possible squares in a bipartite network (Zhang et al. (2008)). In this figure, m and n are a pair of neighbors of node i. Here we observe 1 square cycle ($q_{imn} = 1$) that consists of node *imbn*, and 4 possible squares (*iman*, *imbn*, *incm*, *indm*).

Let $C^{row}(i)$ and $C^{col}(i)$ are the average $C_{mn}(i)$ of node i across all possible combination of its pairs of neighbors m and n, we then calculate the global clustering coefficient as,

$$C = \frac{1}{n^B + n^F} \left(\sum_{i=1}^{n^B} C^{row}(i) + \sum_{i=1}^{n^F} C^{col}(i) \right), \tag{2}$$

which ranges between $[0, 1]$; higher values indicate a more clustered network, and a value of 1 corresponds to a perfectly clustered network. Put simply, in our case higher clustering would indicate that banks tend to cluster their investments on the same set of firms (or industries), or equivalently, firms (or industries) tend to borrow from the same banks. Table 2 shows that the networks are clustered at the aggregated level but not at the disaggregated one.

Lastly, *nestedness* quantifies the degree to which low-degree banks (firms/industries) tend to interact with a subset of firms/industries (banks) that the high-degree banks (firms/industries) interact with. We follow Almeida-Neto et al. (2008) and use NODF (***N**estedness metric based on **O**verlap and **D**ecreasing **F**ill*) as our measure of nestedness

$$\text{NODF} = \frac{\sum_{ij} G_{ij}^{row} + \sum_{ij} G_{ij}^{col}}{n^B(n^B-1)/2 + n^F(n^F-1)/2}, \tag{3}$$

where

$$G_{ij}^{row} = \begin{cases} 0 & \text{if } k_i \leq k_j \\ \sum_{d=1}^{n^F} \mathbb{I}\{b_{id} = 1 \text{ AND } \ b_{jd} = 1\}/\min(k_i, k_j) & \text{otherwise.} \end{cases} \tag{4}$$

is the paired overlap of rows i and j, which is simply the fraction of 1's (which denotes to the existence of a link) present in both rows i and j. A similar term G_{ij}^{col} is used to compute the percentage of paired overlap of columns i and j. NODF lies in the range $[0, 1]$; higher values correspond to higher nestedness, and a value of 1 indicates a perfectly nested network. Table 2 shows that all networks are nested at both aggregation levels, suggesting a strong overlap of Japanese banks' loan portfolios (see Fricke and Roukny (2018)).[8]

In summary, Table 2 shows that the disaggregated credit networks are sparse, disassortative, and nested. On the other hand, the aggregated networks are also disassortative and nested, but also dense and clustered. We now aim to find reconstruction methods that are able to reproduce these features.

4 Network Reconstruction

The literature on network reconstruction is concerned with finding appropriate null models (i.e., network randomizations) that replicate certain features of the actual network. In this paper, we look at four different network reconstruction methods that have been found to be of importance for unipartite financial networks (see Anand et al. (2015); Anand et al. (2017); BIS (2015); Gandy and Veraart (2017); Mazzarisi and

[8] Note that these values cannot be used to assess the significance of nestedness. For this, one would have to compare them with what would be expected at random, i.e., using different null models. This is not the aim of this paper, but the results in Table 5 suggest that the actual credit networks indeed tend to show higher NODF values than their random counterparts.

Lillo (2017); Mistrulli (2011)).

As in Anand et al. (2017), we focus on the independent reconstruction of the static credit networks for each year. As such, we do not take into account the existence of previous bank-firm credit relationships and do not explicitly model time variation in the observed network topologies due to certain economic mechanisms. For example, additional models that explicitly take memory and/or preferential lending into account (De Masi and Gallegati (2012); Iori et al. (2015); Hatzopoulos et al. (2015)) could potentially improve network reconstruction. Here we focus exclusively on methods that preserve certain features of the observed network. This choice is justified by the fact that, among others, Hatzopoulos et al. (2015) find that preferential lending in interbank networks is largely driven by the degree distribution.

Existing reconstruction methods can be classified in terms of the inputs needed to reconstruct the network, the desired network features, and the outputs. To reconstruct a given interbank network, for example, all methods use the information of banks' aggregate borrowing and lending positions, respectively. In this way, the total size of the system and the size of each individual market participant are expected to match the actual values. In addition, some methods also use the system's overall connectivity (Squartini et al. (2017)), while others use each bank's individual connectivity (Squartini and Garlaschelli (2011)). In terms of the desired network features, some methods focus on minimizing the total number of connections (Anand et al. (2015)), while others focus on minimizing the exposure with respect to each counterparty (Upper (2011)). Lastly, in terms of their outputs, some methods produce a single network for a given set of partial information, while others generate an ensemble of networks. Several available network reconstruction methods, including some that we explore in this paper, have been compared with each other previously for unipartite interbank networks, but this is one of the first studies to focus on bipartite financial networks.

4.1 Null Models

Let us briefly describe the different null models used in this paper (see Table 3 for an overview).

4.1.1 Details

Maximum Entropy (MaxEntropy). First, we look at the well-known method of Maximum Entropy (MaxEntropy). In the literature on financial networks, MaxEntropy is often considered as the standard approach to derive individual interbank liabilities in the absence of further information. It has been widely used to reconstruct interbank networks of different countries (see Upper (2011); Anand et al. (2015)). The main characteristic of MaxEntropy is that it generates fully connected networks, i.e., it assumes maximum diversification. Di Gangi et al. (2018) show that, in the case of bipartite networks, MaxEntropy implies that all market participants hold the exact same (market) portfolio.

Null model	Required information	Definition and remarks
Configuration Model 1 (CM1)	k^B, $k^F(k^I)$, s^B, and $s^F(s^I)$ sequences	Generates ensemble of networks. Link allocation: based on the approach of Squartini and Garlaschelli (2011), but adjusted for bipartite network. The probability of link existence between every two nodes in the network, $p_{ij} = \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j}$, is calculated by solving: $\sum_j \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j} = k_i^B \quad \forall i, \quad \sum_i \frac{\theta_i \gamma_j}{1+\theta_i \gamma_j} = k_j^F \quad \forall j.$ for θ and γ. Weight is allocated using RAS.
Configuration Model 2 (CM2)	s^B and $s^F(s^I)$ sequences and m	Generates ensemble of networks. Using fitness model. Link allocation: based on the approach of Squartini et al. (2017). The probability of link existence between every two nodes in the network, $p_{ij} = \frac{zV_iV_j}{1+zViVj}$, is calculated by solving $\sum_i \sum_j \frac{zV_iV_j}{1+ViVj}. = m$ for θ and γ. Weight is allocated using RAS.
Maximum Entropy (MaxEntropy)	s^B and $s^F(s^I)$ sequences	Simple implementation of standard maximum entropy approaches. Produces completely connected network. Generates one single network. Economic interpretation: each node is as diversified as possible.
Minimum Density (MinDensity)	s^B and $s^F(s^I)$ sequences	Each bank and industry have the same total loan amounts but we minimize the total number of links. Generates ensemble of networks due to multiple possible solutions. Economic interpretation: each node is as specialized as possible. Based on the approach of Anand et al. (2015), but adjusted for the case of bipartite networks.

Table 3: Summary different network reconstruction methods used in this paper.

Minimum Density (MinDensity). Second, we look at the Minimum Density approach (MinDensity) of Anand et al. (2015). This method was developed to acknowledge the fact that real financial networks tend to be sparse, in which case using MaxEntropy is rather problematic (Mistrulli (2011)). In a sense, MinDensity can be seen as the opposite extreme of MaxEntropy, given that it starts from the premise that establishing/maintaining links is costly, which is in line with the fact that most banking networks are sparse. As a result, banks do not spread their borrowing and lending across the entire system, and MinDensity identifies the network that satisfies the total aggregate positions with the minimum number of links. This assumption is in line with the fact that relationship banking is of the utmost importance in most banking systems. Since the MinDensity-algorithm may yield multiple solutions, we treat this algorithm as an ensemble method. In our specific case, the bank-firm networks are sparse as well (see Table 5). On the other hand, the aggregated bank-industry networks are dense, such that MinDensity is likely to have difficulties in replicating the aggregated networks.

Configuration Models (CM). Lastly, we use two different versions of the popular configuration model (CM). CMs are probably the most popular types of random graph models because they allow to randomize a given network while preserving its degree distribution. As such, CM can be quite restrictive. CMs have been previously explored in different fields, from sociology to biology (see Fosdick et al. (2016) for an overview), and several of them have been applied in financial network settings (Squartini and Garlaschelli (2011); Musmeci et al. (2013); Mastrandrea et al. (2014); Cimini et al. (2015b); Squartini et al. (2017)). We are aware of only one other application that applies the CM to bipartite financial networks (Squartini et al. (2017)).

The first configuration model, CM1, is based on Squartini and Garlaschelli (2011), but adjusted for the case of bipartite networks. In addition to the strength sequences, CM1 requires the degree sequences of all nodes as additional inputs, thus preserving the exact degree distributions. The second configuration model, CM2, is based on Squartini et al. (2017), which extends the reconstruction method for unipartite networks introduced in Cimini et al. (2015b) to the bipartite case. CM2 preserves the degree distribution as well, but only requires the total number of links additional input. Hence, CM2 needs less detailed information compared to CM1.

We should stress that, in contrast to MaxEntropy and MinDensity, both CMs produce binary instead of weighted networks.[9] After obtaining a randomized adjacency matrix, we need to distribute the observed credit volumes across links. There are different approaches for this (see Table A.1 in the Appendices for an overview), but in the following we use the standard RAS algorithm of Blien and Graef (1998).[10]

[9]The original model of Squartini et al. (2017), where CM2 is based on, generates weighted networks. However, here we only consider part of their method to produce binary networks. This part of their method is based on the work of Saracco et al. (2015) where the formalism for the fitness bipartite is first introduced for the world trade web.

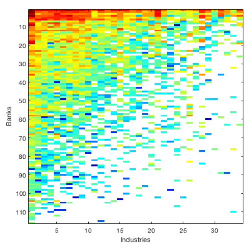
[10]The RAS algorithm generally performed best in our analysis (in terms of the corresponding L_1-error), but we also experimented with the other weight allocation methods mentioned in Table A.1 in the Appendices. The results are qualitatively similar to what is shown here. Details available upon request from the authors.

Method	Input			Output	
	Aggregate positions	Total links	Degree sequence	Single	Ensemble
CM1	v	v	v		v
CM2	v	v			v
MaxEntropy	v			v	
MinDensity	v				v

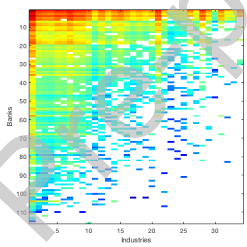
Table 4: Summary classification of the methods based on the input and the output.

Table 3 provides more technical details of our implementation of the four null models. Table 4 summarizes the differences in terms of the required inputs and the outputs. It should be clear that CM1 requires the most detailed information as inputs (followed by CM2), while MaxEntropy and MinDensity require only the strength sequences. Furthermore, CM1, CM2 and MinDensity can produce an ensemble of networks while MaxEntropy generates one single output for any particular input.

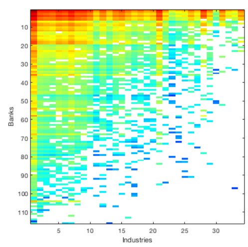
4.1.2 Illustration



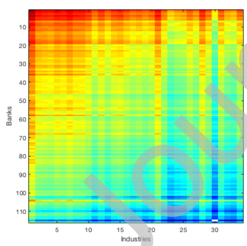
(a) Actual



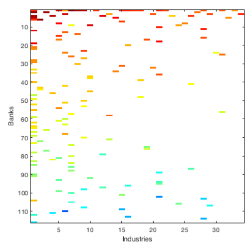
(b) CM1



(c) CM2



(d) MaxEntropy



(e) MinDensity

Figure 2: Weighted credit network bank-industry in 2010 and one realization for each of the four reconstruction methods. Data are log transformed. Warmer colors indicate stronger links, and white dots correspond to the absence of a link.

In order to provide some intuition for the typical outputs of each method, Figure 2 shows the weighted version of the actual aggregated credit network (log-transformed) for the year 2010 and one realization of each corresponding null model. Warmer colors denote stronger relationships, and white dots correspond to the absence of a link. It becomes clear that different reconstruction methods can generate very different network architectures - for example, MaxEntropy produces a fully connected credit network while MinDensity yields a highly compartmentalized and sparse network. In this specific case, MinDensity needs less than 5% of the total links in the actual network to distribute the weight (the actual density is around 46%). The two CMs, on the other hand, tend to produce networks that are visually much closer to the actual one. As such, it is natural to expect that these will perform well.

4.2 Defining Relevant Dimensions of Comparison

In this section we define the different dimensions along which we will compare the actual credit networks and the reconstruction methods.

4.2.1 Network Characteristics

To understand how similar the statistics of the reconstructed networks are to the actual networks, we compare their density, average degree, assortativity, clustering and nestedness (as defined in the previous section) at the different aggregation levels.

4.2.2 Allocation of Links and Weights

In addition to comparing specific network properties, we also look at the performance of each method in terms both of placing links and distributing weights correctly, respectively. In the following, we formally define the network similarity measures for the bank-firm credit network. In line with these definitions, we can define similar measures for the bank-industry credit network.

Link Allocation. In order to understand the ability of a method to reproduce correct links in the network, we calculate the values of Accuracy, Sensitivity, and Specificity. We define the *Accuracy* of a given reconstructed network as

$$\text{Accuracy} = \frac{1}{n^B \times n^F} \sum_{i=1}^{n^B} \sum_{j=1}^{n^F} (\mathbb{I}\{b_{ij} = 0 \textit{ and } \hat{b}_{ij} = 0\} + \mathbb{I}\{b_{ij} = 1 \textit{ and } \hat{b}_{ij} = 1\}), \quad (5)$$

where $\hat{b}_{ij}$ equals 1 if there is a link between nodes i and j in the reconstructed network of a given null model. Put simply, Accuracy tells us the total number of links and non-links that are allocated correctly, relative to the size of the network.

Sensitivity

$$\text{Sensitivity} = \frac{1}{m} \sum_{i=1}^{n^B} \sum_{j=1}^{n^F} (\mathbb{I}\{b_{ij} = 1 \textit{ and } \hat{b}_{ij} = 1\}), \quad (6)$$

measures the number of actual links correctly allocated.

Lastly, *Specificity*

$$\text{Specificity} = \frac{1}{n^B \times n^F - m} \sum_{i=1}^{n^B} \sum_{j=1}^{n^F} (\mathbb{I}\{b_{ij} = 0 \ and \ \hat{b}_{ij} = 0\}, \quad (7)$$

measures the number of non-existing links correctly allocated. These three measures take values in the range $[0, 1]$, with higher values corresponding to greater similarity.

Weight Allocation. We are also interested in quantifying the ability of each null model to reproduce the observed link weights in the credit network. For this purpose, we use three different measures: L_1-error, root-mean-square deviation (RMSE) and cosine similarity (Cos-Sim). L_1-*error* is defined as

$$L_1 = \sum_{i=1}^{n^B} |\hat{s_i^B} - s_i^B| + \sum_{j=1}^{n^F} |\hat{s_j^F} - s_j^F| \quad (8)$$

which allows us to understand how well the reconstructed network is able to satisfy the aggregate positions, which is the total borrowing (lending) of banks (firms/industries), in the actual network. As mentioned previously in Table 3 and Table 4, all null models are expected to reproduce the actual aggregate positions. Therefore, L_1-error measures the degree to which those constraints have been satisfied by a given null model. In everything that follows, we scale the L_1-error by the average lending volume of banks in the actual network.

Additionally, we calculate *RMSE* which is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n^B} \sum_{j=1}^{n^F} (\hat{w}_{i,j} - w_{i,j})^2}{n^B \times n^F}}, \quad (9)$$

where $\hat{w}_{i,j}$ is the allocated credit volume of bank i to firm j in a given reconstructed network. In everything that follows, we scale RMSE by the average exposure of a link in the actual network which makes values comparable over time.

Cosine Similarity as

$$Cos - Sim = \frac{\sum_{i=1}^{n^B} \sum_{j=1}^{n^F} \hat{w_{i,j}} w_{i,i}}{\sqrt{\sum_{i=1}^{n^B} \sum_{j=1}^{n^F} \hat{w}_{i,j}^2} \sqrt{\sum_{i=1}^{n^B} \sum_{j=1}^{n^F} w_{i,j}^2}}. \quad (10)$$

to quantify deviations in the weight allocation across all links in the network.

L_1-error and RMSE have values in the range $[0,\infty]$ with lower values corresponding to greater similarity. Meanwhile, Cos-Sim has values in the range [0,1] and higher values correspond to greater similarity.

4.3 Results on Horse Racing Different Methods

In this section, we show the empirical results on horse racing the different network reconstruction methods. For each year under study and each null model, we generate 100 network realizations for each aggregation level. We then calculate the average of each of the characteristics mentioned previously. For the sake of brevity and also illustrative purposes, in the following we only show results for the year 2010, but the results are qualitatively similar for other years and do not affect our main conclusions.

The main results for the three aggregation levels can be found in Tables 5 (network statistics) and 6 (link/weight similarity). In all cases, the *best* method for each statistic is highlighted using the ⋆ symbol. Let us briefly describe the results for the different aggregation levels.

Network characteristics						
Disaggregated	Density	$\bar{k^B}$	$\bar{k^F}$	r	C $(\times 10^{-2})$	NODF
$\mathbf{W}(116 \times 2296)$	0.042	96.474	4.874	-0.215	0.028	0.359
CM1	0.042	96.601	4.881	⋆-0.205	⋆0.028	⋆0.366
CM2	⋆0.042	⋆96.510	⋆4.876	-0.321	0.062	0.254
MaxEntropy	1.000	2296	116	NaN	1.000	0.000
MinDensity	0.009	20.789	1.050	-0.125	0.000	0.009

Network characteristics						
Aggregated	Density	k^B	k^I	r	C	NODF
$\mathbf{W^I}(116 \times 34)$	0.461	15.664	53.441	-0.330	0.134	0.819
CM1	⋆0.460	⋆15.649	⋆53.392	⋆-0.370	⋆0.136	⋆0.821
CM2	0.461	15.683	53.507	-0.248	0.131	0.704
MaxEntropy	1.000	34.000	116.000	NaN	1.000	0.000
MinDensity	0.038	1.285	4.385	-0.224	0.000	0.044

Network characteristics						
Intermediate	Density	k^B	$k^{F \to I}$	r	C	NODF
$\mathbf{W} \to \mathbf{W^I}$	0.461	15.664	53.441	-0.330	0.134	0.819
CM1	⋆0.482	⋆16.395	⋆55.936	⋆-0.308	⋆0.152	⋆0.798
CM2	0.493	16.771	57.218	-0.289	0.175	0.769
MaxEntropy	1.000	34.000	116	NaN	1.000	0.000
MinDensity	0.178	6.055	20.658	-0.329	0.019	0.442

Table 5: Comparison of the statistics between the actual credit network for year 2010 and the reconstructed networks for different aggregation levels. $\bar{k}^B$ and $\bar{k}^F$ ($\bar{k}^I$) correspond to the average degree, r denotes the assortativity, C indicates the clustering coefficient and NODF denotes the nestedness of the network. We highlight the best reconstruction method for a given statistic (the value closest to the actual network) using the ⋆ symbol.

4.3.1 Disaggregated Level

At the disaggregated level (bank-firm), the top panel of Table 5 shows that the two CMs tend to reproduce the features of the actual network reasonably well: the density,

Disaggregated	Link similarity			Weight similarity		
	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	0.941	0.304	0.969	4.511	18.674	0.442
CM2	0.936	0.241	0.967	2.706	13.850	0.633
MaxEntropy	0.042	⋆1.000	0.000	0.000	⋆13.038	⋆0.681
MinDensity	⋆0.955	0.071	⋆0.994	⋆0.000	27.896	0.278

Aggregated	Link similarity			Weight similarity		
	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	⋆0.781	0.762	0.798	0.015	⋆2.527	⋆0.915
CM2	0.711	0.687	0.732	0.018	2.555	0.914
MaxEntropy	0.461	⋆1.000	0.000	⋆0.000	2.572	0.914
MinDensity	0.558	0.061	⋆0.982	0.000	8.607	0.532

Intermediate	Link similarity			Weight similarity		
	Accuracy	Sensitivity	Specificity	L_1-error	RMSE	Cos-Sim
CM1	⋆0.767	0.771	0.764	4.511	2.675	0.905
CM2	0.738	0.750	0.726	2.706	⋆2.530	⋆0.915
MaxEntropy	0.461	⋆1.000	0.000	0.000	2.572	0.914
MinDensity	0.668	0.333	⋆0.954	⋆0.000	3.676	0.836

Table 6: Link and weight similarity of the reconstruction methods with the actual credit network in 2010 for different aggregation levels. Accuracy, sensitivity, specificity and cosine similarity lie in the range [0,1] and higher values correspond to higher similarity. L_1-error and RMSE lie in the range $[0,\infty]$ with smaller values corresponding to greater similarity. We highlight the best reconstruction method for a given statistic (the value closest to the actual network) using the $\star$ symbol.

average degree, assortativity, clustering, and nestedness are always quite similar to the actual values. On the other hand, MaxEntropy and MinDensity perform rather poorly: for example, in terms of density MaxEntropy (MinDensity) produce much higher (lower) values.

The results for link allocation and weight distribution (top panel of Table 6), are broadly in line with those for the network characteristics: again the two CMs perform relatively well across the different measures. In this case, however, the results are not always consistent. For example, MinDensity achieves the highest Accuracy and the lowest L_1-error, but shows the worst Sensitivity, RMSE, and Cos-Sim. On the other hand, MaxEntropy yields the worst Accuracy but the best RMSE and Cos-Sim. Not surprisingly, MaxEntropy achieves the maximum Sensitivity simply because it predicts a fully connected network.

We should also mention that both CM1 and CM2 generate relatively large L_1-errors, indicating that they do not manage to perfectly allocate the aggregate positions. This is due to the nature of CM1 and CM2 as preserving the degree sequences only in expectation, such that specific realizations can lead to some low-degree nodes being inactive (or unconnected).[11]

4.3.2 Aggregated Level

Similar to the previous results, the center panel of Table 5 shows that the two CMs tend to reproduce the observed network characteristics reasonably well at the aggregated (bank-industry) level. In this particular case, CM1 consistently performs best. For link/weight similarity, the results are also comparable (center panel of Table 6), except for Sensitivity and Specificity which are again dominated by MaxEntropy and MinDensity, respectively.

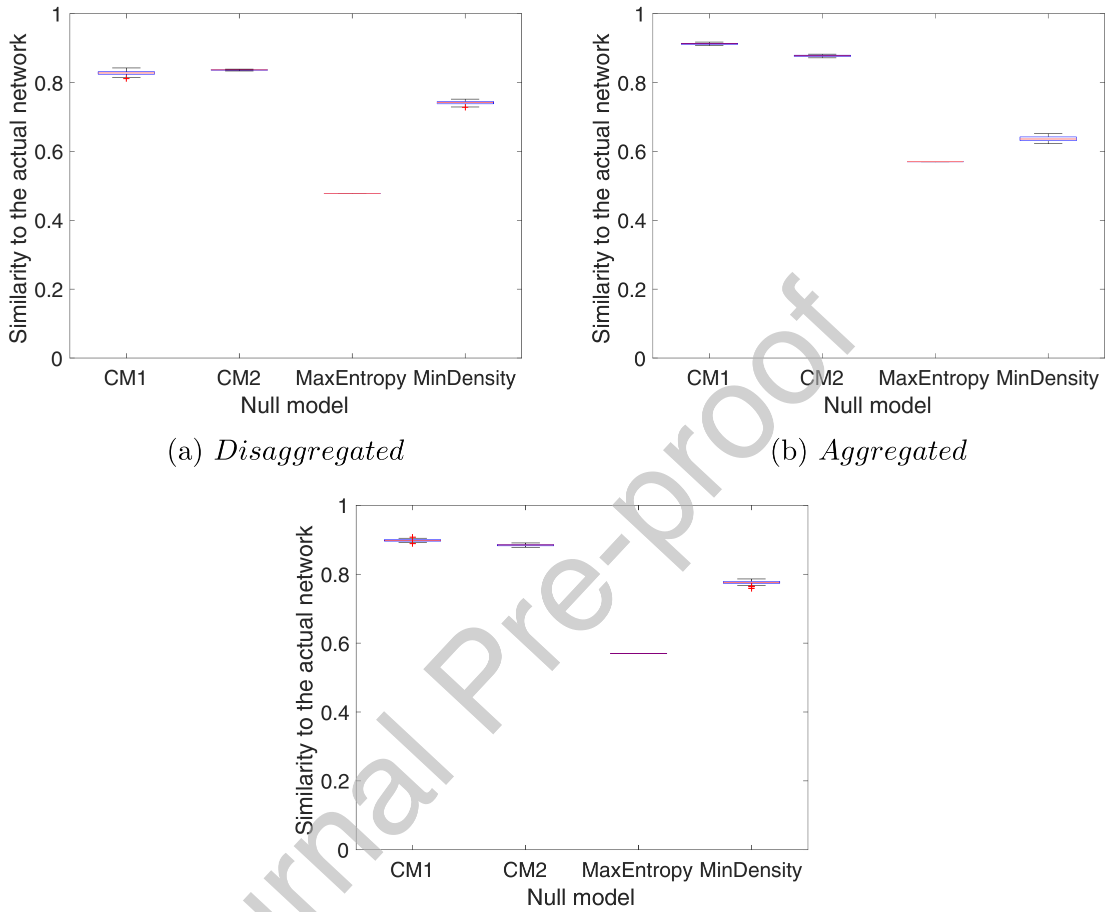
4.3.3 Intermediate Level

Lastly, the two bottom panels of Tables 5 and 6 show the results for the intermediate aggregation level, where we construct synthetic networks for the disaggregated (bank-firm) level and then aggregate these to the industry level using firms' observed industry affiliations. Overall, the statistics shown here are very similar to those at the aggregated level (with the exception of the L_1-error, which is close to the value at the disaggregated level), with CM1 performing best for the network statistics and the Accuracy.

4.4 Summary and Discussion - Network Reconstruction

Previous studies on the reconstruction of interbank networks (e.g., Anand et al. (2017)) suggest that the best reconstruction method depends on the type of network characteristics of interest. Our findings support this conclusion. We see, for example, that if we

[11] We also experimented with a minimum threshold in terms of active nodes' degrees, but observe a similar issue.



(a) *Disaggregated* (b) *Aggregated*

(c) *Intermediate*

Figure 3: Standard box-plots of the (normalized) similarity measures between the actual network and different network reconstruction methods (averaged over many realizations). For each reconstruction method, we consider each of the measures in the three categories of similarity under study: network characteristics, link similarity, and weight distribution. We normalize each of these measures such that a value of 1 (0) indicates that the reconstructed network and the actual network are identical (completely different) in terms of these features. We compute the metrics for each synthetic network, and take the average of these quantities.

focus the horse race on the number of non-existent links in the adjacency matrix that are correctly estimated (Specificity), MinDensity which produces sparse networks is clearly the winner. However, when we look at the number of links correctly estimated (Sensitivity), MaxEntropy, which generates a fully connected network, outperforms all other methods.

Given that our comparison is based on multiple network statistics, Figure 3 summarizes the results by combining statistics for the individual features. For this purpose, we normalize each measure in Table 5 (network characteristics) and Table 6 (link similarity and weight distribution) such that each of them ranges between 0 and 1, where the value of 1 (0) indicates that the reconstructed network and the actual network are identical (completely different) in terms of these features. We compute the metrics for each synthetic network, and take the average of these quantities. Figure 3 shows a standard box-plot of the metrics of each reconstruction method that are averaged over the realizations of the synthetic networks.[12] Overall, we find that the two CMs consistently perform the best, followed by MinDensity and MaxEntropy. We also note that, in general, CM1 and CM2 succeed in reproducing the topological structure related to the heterogeneity of links in the actual network (e.g., assortativity). This is important since heterogeneity plays an important role for systemic risk in financial networks (Iori et al. (2006), Banwo et al. (2016)).

Since CM1 and CM2 require more information relative to the other methods (degree sequence and total degree, respectively), it seems clear that adding such information improves the performance of the reconstruction methods (see also Mastrandrea et al. (2014) and Cimini et al. (2015a)). This finding is in line with Gandy and Veraart (2016), who suggests that using the information on aggregate positions only is not sufficient to reconstruct certain topological properties of the network. Overall, it seems reassuring that, despite the fact that CM1 requires more information than CM2, both methods generate very similar networks (in some cases CM2 even outperforms CM1). This indicates that the degree distribution of the network might indeed, to a certain extent, be inferred without the full knowledge of the degree sequence. An obvious follow-up question is to what extent CM2 would still perform well if we treated the overall density as a free parameter. We leave this question for future research.

5 Systemic Risk Analysis

One of the main reasons why regulators and policymakers are interested in reconstructing financial networks from partial information is because of their potential contribution to financial instability. Therefore, exploring how well different methods are able to reconstruct the observed networks is only the first step. The next step is to compare how well the different network reconstruction methods are able to replicate the levels of systemic risk of the actual credit networks. Clearly, this analysis is not independent from the results of the previous section, in the sense that we would expect a method that closely reproduces the actual networks to also yield similar systemic risk levels.

[12] Note that we ignore the average degree (since it is redundant with density) and assortativity (since it is not defined for MaxEntropy) from the calculation.

To the best of our knowledge, however, such an exercise has not been performed for the case of bipartite financial networks.

5.1 Measuring Systemic Risk

Over the last decade, common asset holdings (or overlapping portfolios) have been identified as an important source of systemic risk and several stress testing models have been introduced (see Table B.1 in the Appendices for a comparison of different models). In this paper, we use the stress testing model of Huang et al. (2013) in order to quantify the vulnerability of the bipartite credit networks to systemic asset liquidations. The model has also been used in a study of the Venezuelan banking system (Levy-Carciente et al. (2015)) and is similar in spirit to the models in Greenwood et al. (2015) and Caccioli et al. (2014).[13]

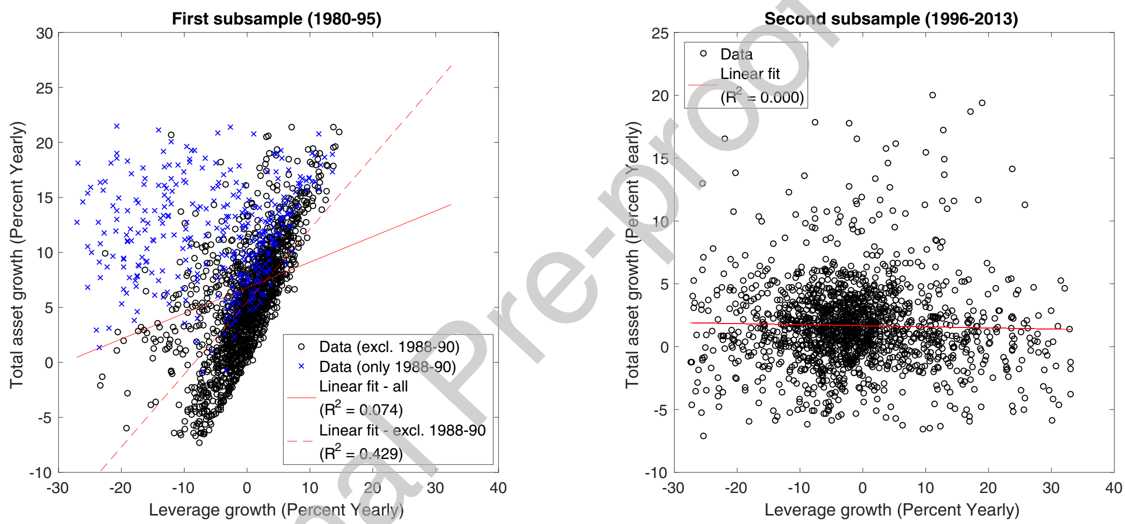


Figure 4: Scatter plot of change in leverage and change in total asset of banks drawn from our dataset. The left panel displays the results for 1980-95, and the right panel shows the results for 1996-2013. Red lines show the best linear fit between the two variables.

The model of Huang et al. (2013) uses a linear market impact function (always yielding positive prices) and, in contrast to several other studies, assumes that banks do not target their leverage. First, choosing a linear impact function can be seen as more conservative, in the sense that we tend to overestimate the resulting price impact of a given asset liquidation. Second, regarding the exclusion of leverage targeting, we checked whether we find similar results as in Adrian and Shin (2010) for our sample of Japanese banks. Figure 4 shows scatter plots of the change in leverage against the change in total asset (both in percent) for two subsamples, with the best linear fits

[13]For the purpose of finding out how the systemic risk analysis might vary if leverage targeting model (as in Greenwood et al. (2015)) and threshold model (as in Cont and Schaanning (2017)) are used, we also performed the same exercise with these other models. We find that the rank ordering of the different methods are generally consistent with those presented in the main text. See Appendices for more details.

shown as red lines. If Japanese banks were targeting fixed leverage values, we would expect most observations to cluster around a vertical line at zero leverage growth. We find that this is not the case for either of the subsamples under study here: for the first subsample (1980-95), we find a positive relationship between the two variables, suggesting that banks tended to use procyclical leverage during the first half of the sample. Note that the left panel also shows results without the noisy 1988-90 data which improves the fit dramatically. For the second subsample (1996-2013), the right panel shows that that assuming no leverage targeting is again a reasonable assumption. In fact, this plot looks similar to the corresponding Figure for non-financial, non-farm corporates in Adrian and Shin (2010). This suggests that Japanese banks appear to manage their leverage to a certain extent, but clearly do not seem to have fixed leverage targets.

Let us briefly sketch the model details: let the total market value of asset j be defined as $\Gamma_j = \sum_i \gamma_{i,j}$, with $\gamma_{i,j}$ the amount of asset j owned by bank i. The basic steps of the model are:

1. We shock a given industry j by reducing its market value to $p \in [0, 1]$ times its original value. Note that a value of $p = x$ would mean that the market value of industry j is reduced by x, or in other words it is a (1-x) shock to industry k. Therefore, a smaller p corresponds to a larger shock.
2. Does any bank default? This occurs if a bank's total assets drop below its liabilities. (Process terminates if no bank defaults.)
3. If a bank i defaults, it liquidates all of its remaining asset holdings. This has an indirect effect on other banks, because the market value of its assets drops proportional to $\alpha \in [0, 1]$ times the bank's current holdings. The unit price of a liquidated asset j becomes a fraction $\frac{\Gamma_j - \alpha\gamma_{i,j}}{\Gamma_j}$ of its original price.
4. Back to step 2 ...

Note that α is a homogeneous (identical across assets) market impact parameter: a value $\alpha = 0$ corresponds to an extremely liquid asset, that is when any sales would not alter the market value of the asset, while $\alpha = 1$ corresponds to an extremely illiquid asset, where sales could potentially push the market price down to 0.[14] We will show results for different values of p and α. In the following, we mainly focus on a specific range of parameters. In particular, we consider loans to be relatively illiquid and therefore focus on the upper range of the market impact parameter ($\alpha \in [0.6, 1]$).[15] Moreover, in line with previous studies on price-mediated contagion, we consider relatively small values of the intial shock ($p \in [0.6, 1]$).[16]

We perform the above exercise separately for each industry j. At the aggregated level, for each iteration we shock one node (industry), while for the disaggregated level

[14] Among other things, the liquidity of a loan might be dependent on its remaining maturity. We leave a detailed calibration of the market impact parameter for future work.

[15] For $\alpha = 0.7$, the asset price drops by 7% when 10% of the asset is liquidated; for $\alpha = 1$, the price drops by 10% when 10% of the asset is liquidated.

[16] Greenwood et al. (2015) consider a 50% write-off on GIIPS debts, while Cont and Schaanning (2017) gradually increase the shock from 0% to 20%.

we shock all the nodes (firms) that belong to the same industry. To quantify the impact of a shock on industry j, we first define *default rate*

$$r_j = \frac{n_j^{Bdefault}}{n^B} \tag{11}$$

as the ratio between the number of failed banks to the total number of active banks in the network. We then define the *probability of default*

$$P_d = \frac{\sum_{j=1}^{n^I} r_j}{n^I} \tag{12}$$

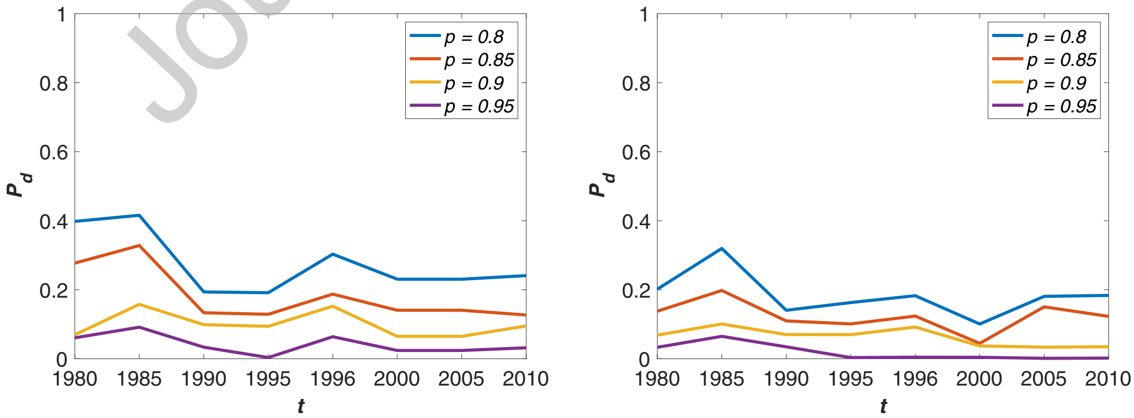
as the average of r_j across all industries. This is our systemic risk measure and in the following we use the terms systemic risk and P_d interchangeably. Finally, for a given reconstructed network $\tilde{W}$, we also define *relative difference* between the actual P_d and the null model P_d as

$$D_r = \frac{P_d^W - P_d^{\tilde{W}}}{P_d^W}. \tag{13}$$

A positive (negative) value of D_r indicates that a given null model underestimates (overestimates) the actual P_d.

5.2 Time Dynamics of Systemic Risk

Before going into the details regarding the different reconstruction methods, we first quantify the level of systemic risk, P_d, over time. Figure 5 plots the P_d over time, both for the disaggregated (left panel) and the aggregated level (right panel), respectively. As a benchmark, we use a market impact parameter $\alpha = 0.7$ and different values of the initial shock p. The plots in Figure 5 suggest that P_d is substantially smaller in 2010 compared with the values earlier in the sample. In other words, in many instances the level of systemic risk appears to have been reduced over time.



(a) **W**: Disaggregated level (b) $\mathbf{W^I}$ Aggregated level

Figure 5: P_d over time for the disaggregated (left panel) and the aggregated level (right panel). We use $\alpha = 0.7$ and various values of p.

We also test for a significant trend in P_d for different values of α and p.[17] We then plot the corresponding p-value of the estimated trend as a heatmap in Figure 6, where darker colors correspond to smaller p-values (i.e., significance) of the estimated trends. The Figure shows that we obtain a significant trend for most values of p (except for very large values) whenever α is relatively small.[18]

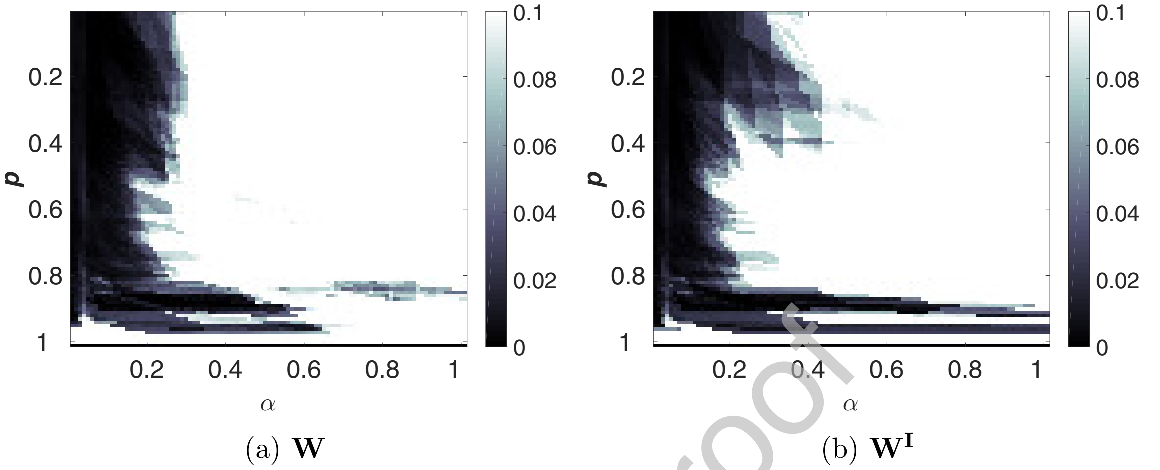


Figure 6: Trend analysis. p-value of regression analysis of P_d against a constant and a time variable (year), for different combinations of p and α. Darker color denotes a smaller p-value.

5.3 Results on Horse Racing Different Methods

We now turn to a detailed analysis of the different null models and their implied levels of systemic risk. As before, we focus our presentation on the results for one particular year of data, namely 2010. We will see that the results shown here are again broadly consistent over time. We will show three sets of results: first, Figures 7-11 show heatmaps of D_r for all possible combinations of p and α. Second, Figure 8 allows us to take a closer look at the systemic risk levels, P_d, for a specific choices of α as a function of p in the range $p \in [0.6, 1]$. Third, to illustrate that our findings are robust over time, Figure 9 shows the P_d's over time for specific choices of α and p.

As for the network reconstruction part in section 4, we briefly discuss the results separately for the three different aggregation levels. Table 7 then summarizes these results.

[17] Technically, for a given combination of α and p, we regress the resulting P_d on a constant and a time variable (year).

[18] For relatively large values of α the absence of a time trend in P_d is easily explained by the fact that in these cases all banks will tend to default in every single year. Hence, P_d will be roughly constant over time.

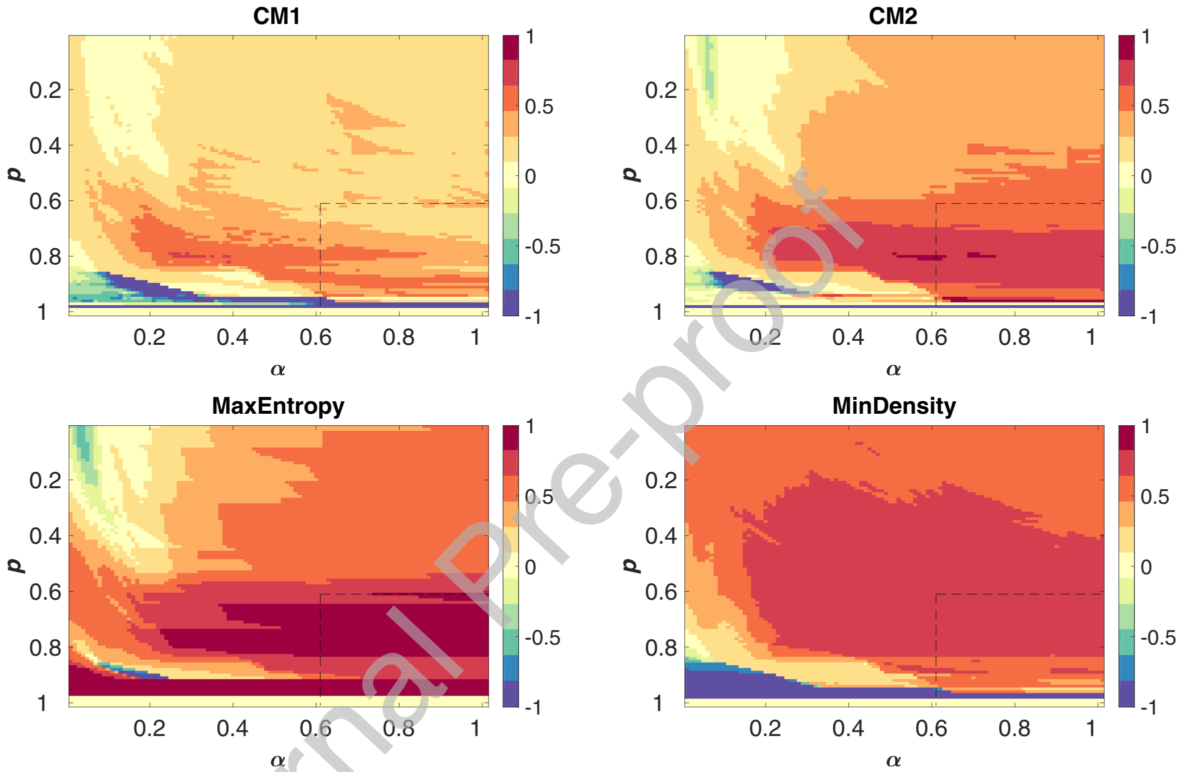
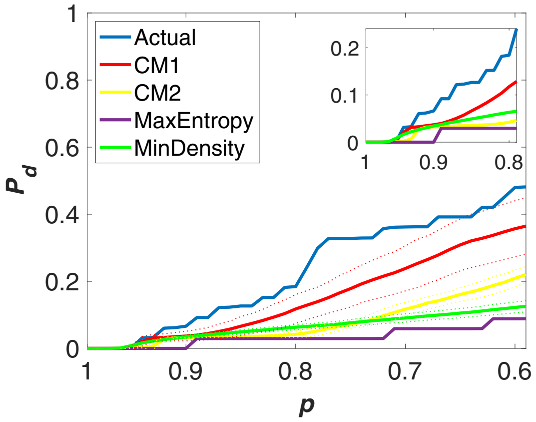
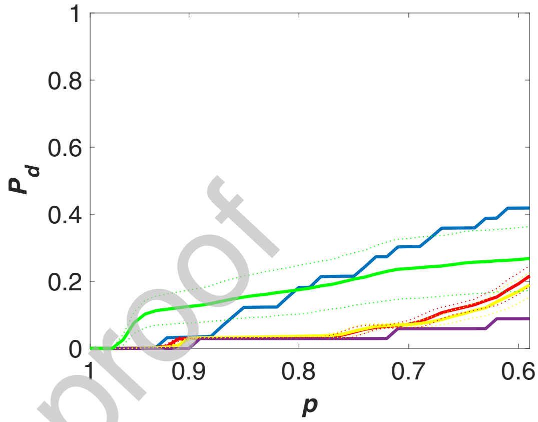


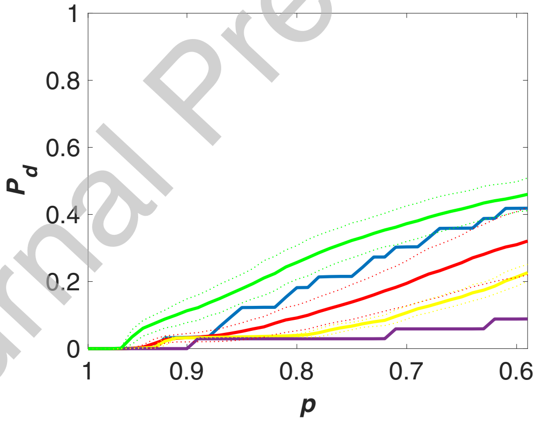
Figure 7: Relative difference of the probability of default between actual network and the null models (D_r) at the **disaggregated level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation. Our main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.



(a) Network at disaggregated level

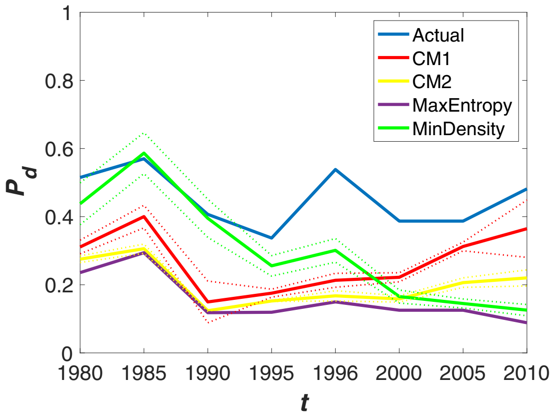


(b) Aggregated level

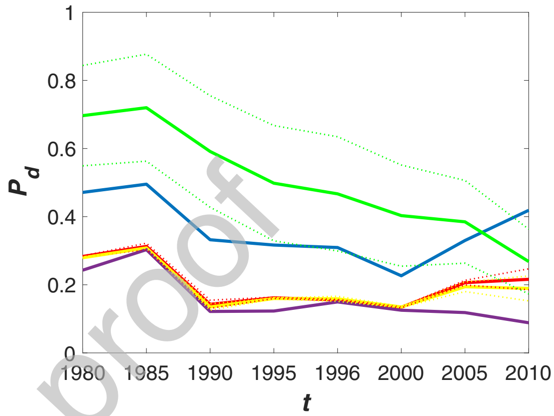


(c) Intermediate level

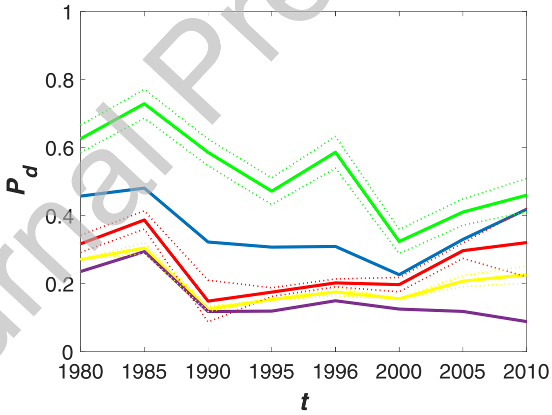
Figure 8: P_d for initial shock $p \in [0.6, 1]$ and $\alpha = 0.7$. Data for year 2010. Dotted line indicates the value within one standard deviation. Inset: P_d for $p \in [0.8, 1]$ and $\alpha = 0.7$ for network at disaggregated level.



(a) Network at disaggregated level



(b) Aggregated level



(c) Intermediate level

Figure 9: P_d over time for illiquid market impact $\alpha = 0.7$ and initial shock $p = 0.6$ for data of different years. Dotted line indicates the value within one standard deviation.

5.3.1 Disaggregated Level

Figure 7 shows that the actual network tends to be the riskiest, because all null models underestimate the actual P_d for most values of p and α. We observe that this underestimation is consistent for the range of parameters that we consider here (small initial shock and high market impact), which is shown by an area inside the black dashed line in Figure 7. Moreover, all null models overestimate the actual P_d only in a small region of the parameter space, for example when $p = 0.9$ (small initial shock) and $\alpha = 0.1$ (small market impact). Figure 7 also shows that the magnitude of the underestimation gets larger as α increases.

Panel (a) in Figure 8 shows the performance of the null models for a specific value of $\alpha = 0.7$ as a function of specific values of $p \in [0.6, 1]$. All null models underestimate the actual P_d, with CM1 being the closest match, followed by CM2, MinDensity and MaxEntropy. We note that, for relatively smaller values of the initial shock (inset in Figure 8, Panel (a)), MinDensity instead produces higher values of P_d than CM2. This result is driven by the fact that MinDensity produces very sparse networks and allocates very few assets to each banks (high concentration levels). Banks are therefore vulnerable to idiosyncratic shocks in this case, and some banks can default due to the initial shock. However, given that the portfolio overlap between banks is very low in the MinDensity model, shocks cannot spread easily through the system and the number of banks defaulting through fire sale cascades is rather small. The opposite is true for CM2: when the initial shock is large enough to cause banks to default, the shock can propagate to other banks. Therefore, for larger initial shocks, CM2 produces higher values of P_d than MinDensity. Lastly, Panel (a) of Figure 9 shows the P_d over time for specific parameters ($\alpha = 0.7$ and $p = 0.6$). As for the 2010 data, the actual network tends to be the most risky one.

5.3.2 Aggregated Level

The results for the aggregated networks are shown in Figure 10 and in Panel (b) of Figures 8 and 9, respectively. With the exception of MinDensity, the actual network tends to be the riskiest (at least for the 2010 data) and all other reconstruction methods tend to underestimate the actual P_d for most values of p and α. As for MinDensity, it overestimates (underestimates) the actual P_d for relatively small (large) initial shocks. However, it should be noted that for the aggregated networks, MinDensity produces the riskiest networks in terms of P_d in most years (see also Figure 9, Panel (B)). The intuition for this finding is similar to our explanations for the disaggregated networks, with the important difference that MinDensity tends to produce denser networks at the aggregate level. Therefore, the initial shock can lead to bank defaults which then spread the shock trough the system.

The results also suggest that the other reconstruction models (CM and MaxEntropy) tend to produce values of P_d closer to each other. This suggests that data aggregation may reduce differences among P_d's of different null models. This is mainly because aggregating data will result in a smaller number of nodes (assets) in the network, so reconstruction models have fewer links to allocate. This reduces differences

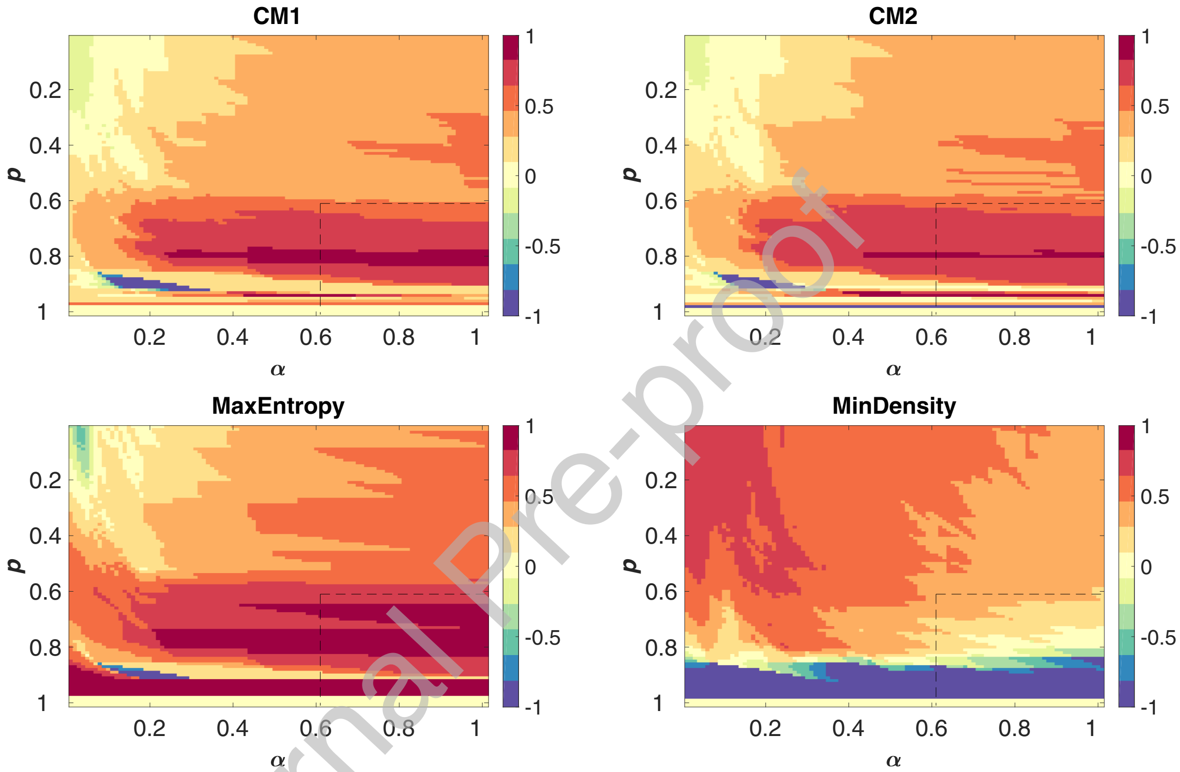


Figure 10: Relative difference of the probability of default between actual network and the null models (D_r) at the **aggregated level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation. Our main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.

between the reconstructed networks.

5.3.3 Intermediate Level

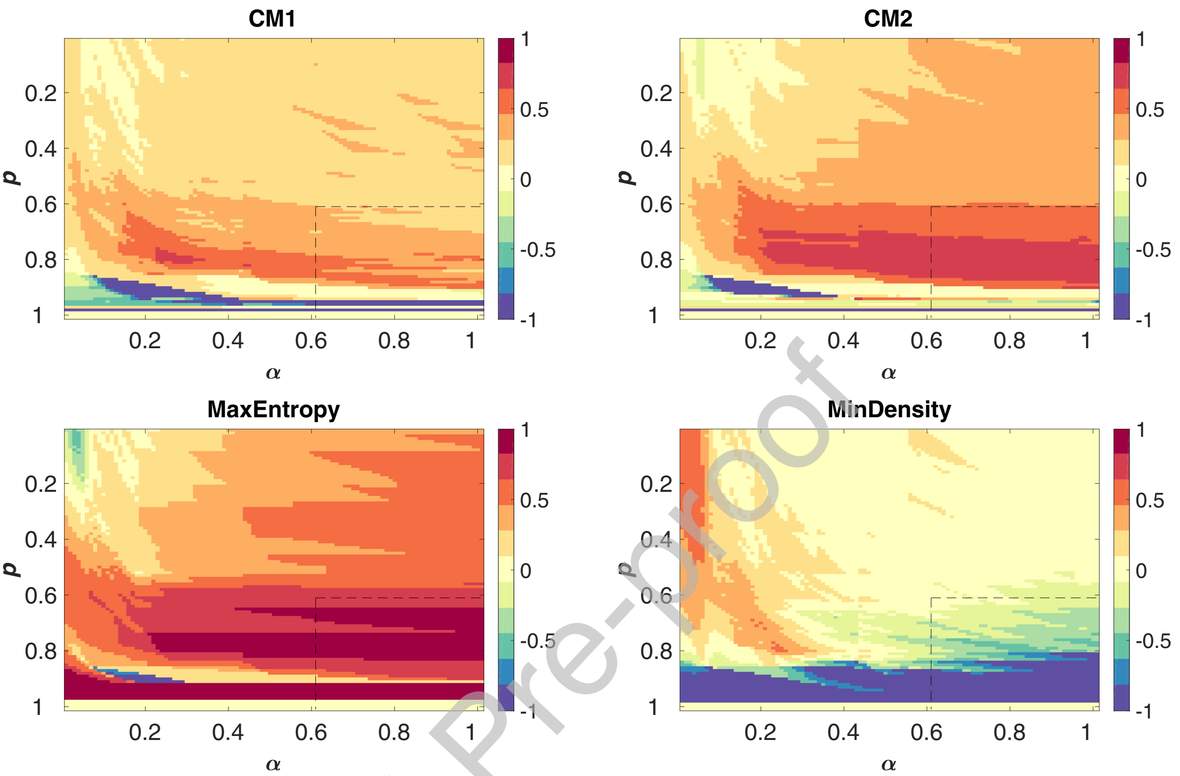


Figure 11: Relative difference of the probability of default between actual network and the null models (D_r) at the **intermediate level** for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation. Our main analysis focuses on small values of the initial shock and large values of market impact, which is shown by the area inside the black dashed line square.

For the intermediate aggregation level, the black dashed line in Figure 11, and Panel (c) of Figure 8 show that MinDensity heavily overestimates the actual P_d. Hence, MinDensity yields the most risky networks at this aggregation level. The Panel (c) of Figure 9 also shows that these results are consistent over time. The reasoning for this finding is again similar to what we saw at the other aggregation levels. The main difference here is that the aggregation takes place after the MinDensity bank-firm networks have been generated, which increases connectivity between banks and thus allows shock to propagate more easily.

5.4 Summary and Discussion - Systemic Risk Analysis

As for the network reconstruction part, Table 7 summarizes the results from the systemic risk analysis. We rank the different methods, along with the actual networks, based on the average P_d (standard deviations in parentheses) for the restriced parameter ranges ($p \in [0.6, 1]$ and $\alpha \in [0.6, 1]$).[19]

Rank	Disaggregated		Aggregated		Intermediate	
	Null model	$\overline{P_d}$	**Null model**	$\overline{P_d}$	**Null model**	$\overline{P_d}$
1	Actual	0.236 (0.164)	Actual	0.195 (0.148)	MinDensity	0.275 (0.159)
2	CM1	0.163 (0.128)	MinDensity	0.193 (0.085)	Actual	0.195 (0.148)
3	CM2	0.079 (0.067)	CM1	0.061 (0.060)	CM1	0.133 (0.109)
4	MinDensity	0.069 (0.041)	CM2	0.057 (0.051)	CM2	0.076 (0.067)
5	MaxEntropy	0.035 (0.026)	MaxEntropy	0.035 (0.027)	MaxEntropy	0.034 (0.027)

Table 7: Rank of the actual networks and the corresponding null models at different aggregation levels for the 2010 data. Rank 1 corresponds to the most risky network. $\overline{P_d}$ denotes the average. We also show the standard deviation of P_d in brackets, which is calculated using the P_d across the restricted parameter range ($p \in \{0.60, 0.61, 0.62, ... ,1\}$ and $\alpha \in \{0.60, 0.61, 0.62, ... ,1\}$) .

First, we find that the actual network tends to display the highest levels of systemic risk in many instances, at least for the disaggregated networks. This is remarkable, given that some of the reconstruction methods generate very different network architectures; for example, MaxEntropy (MinDensity) yields a maximally (minimally) connected credit network. This finding also suggests that even the null models that preserve the degree distribution, like CM1 and CM2, fail to accurately reproduce the actual P_d.[20] However, our result contrasts Anand et al. (2015) which indicates that MinDensity yields an upper bound of the actual risk. Here we find that MinDensity in many instances underestimates the actual P_d, in particular for the disaggregated networks, but overestimates systemic risk at the intermediate aggregation level.

Second, with regards to the performance of each null model, we find that CM1, followed by CM2 and MaxEntropy, has the closest behavior to the actual network overall, while MinDensity shows an inconsistent performance across different aggregation levels. Given that the different null models require different inputs, we propose CM2 as the most appealing model as it requires less information than CM1.[21]

[19] We also compute the rank for all possible combinations of parameters, including those within and outside restricted range, in Table D.1. We find that the main results are qualitatively similar to those in the main text. We also formally test whether the difference between each network P_d is significant. Specifically, we run a two-sided Wilcoxon signed rank test on each pair of the actual network and the null model (see Tables E.1 and Table E.2 in the Appendices for the test results).

[20] This finding is related to previous studies on interbank networks (Mistrulli (2011) and Anand et al. (2015)) which suggest that MaxEntropy underestimates the actual risk.

[21] Wilcoxon tests (see Appendix) indicate that the P_d results from CM1 and CM2 are not significantly different from each other (as opposed to the results from the other reconstruction methods).

Lastly, the choice of aggregation level of financial networks matters for stress testing. Certain models can change their behavior at different aggregation levels, most notably MinDensity. On the other hand, configuration models generally behave rather well at all aggregation levels. However, the ranking of each null model in term of reconstructing the topological features of the actual network is not necessarily consistent with that of reproducing actual systemic risk level (see the comparison between Figure 3 and Table 7). Future research should explore which network characteristic is most important for reproducing the actual systemic risk levels.

6 Policy Exercise

Our findings suggest that, with respect to the null models we considered, the actual network displays the highest level of P_d in many instances (at least for the 2010 data). This implies that it is possible to make the network more stable by changing its structure. With this in mind, we now explore different policies in order to increase the robustness of the actual credit network.

6.1 Policies

To this end, we use a similar approach as Greenwood et al. (2015) and explore three different sets of policies (see Table 8 for an overview):

1. merging banks with certain characteristics;
2. breaking up banks with certain characteristics;
3. imposing a leverage cap.

First, we explore the effect of merging banks. In this context, we consider four different scenarios in which we merge a group of large or small banks that are chosen on the basis of their size or leverage.We sort the banks according to their total assets (or their leverage ratios), and merge the top and bottom 15% of them into a single bank.

Second, we study the effect of breaking up banks. Specifically, we split a large bank into two smaller banks. Moreover we assume that one of the smaller banks connects only to the group of relatively connected firms, while the other bank connects only to a group of relatively unconnected firms. (Here we use the number of bank relationships per firm as a proxy of firms' connectedness). Additionally, we assume that the leverage ratio of both banks is identical to the leverage of the original bank.

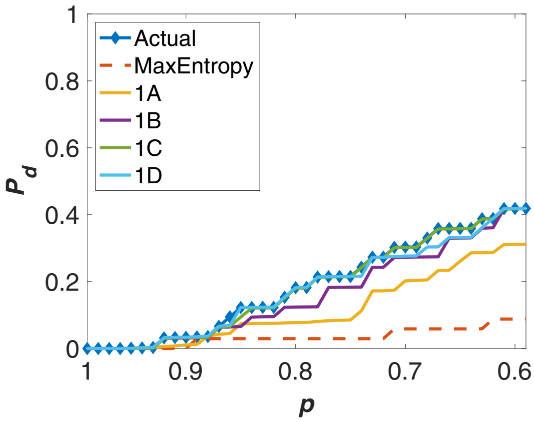
Third, we explore the effect of a leverage cap, i.e. we limit the maximum ratio between debt to equity of a bank. In this case, we assume that banks that breach the limit need to raise new equity to satisfy the cap (without changing the size of the credit portfolio).

Policy choice	Observable outcome	
1 - Bank merger	Number of banks merged	Total assets of a new merged bank
A) Top 15% (total assets)	17	¥23.58 Tr
B) Top 15% (leverage)	17	¥2.99 Tr
C) Bottom 15% (total assets)	17	¥0.03 Tr
D) Bottom 15% (leverage)	17	¥3.18 Tr
2 - Bank break-up	Number of banks split	Total assets of impacted banks
A) Split each of the top 15% banks (total assets) into one that connects to top 15% industries (connectedness), while the other connects to the bottom 85% industries (connectedness)	17	¥23.58 Tr
B) Split each of the top 15% banks (leverage) into one that connects to top 15% industries (connectedness) and the other connected to the bottom 85% industries (connectedness)	17	¥2.99 Tr
3 - Leverage cap	Equity issue	Number of banks capped
A) max debt/equity = 15	¥354.6 Bn	107
B) max debt/equity = 20	¥79.6 Bn	64
C) max debt/equity = 25	¥34.4 Bn	31
D) max debt/equity = 30	¥18.5 Bn	11

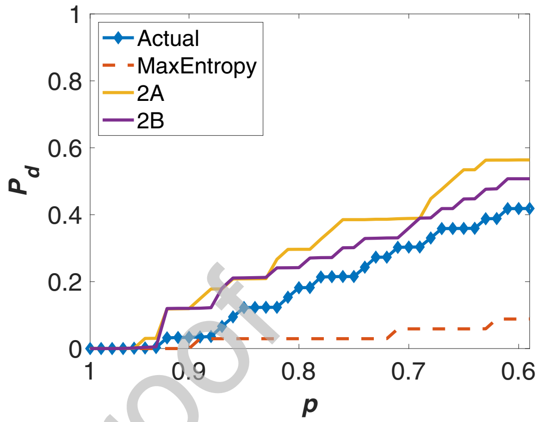
Table 8: Different policy exercises applied to the actual network in 2010. Tr and Bn stand for trillion and billion (in ¥).

6.2 Results

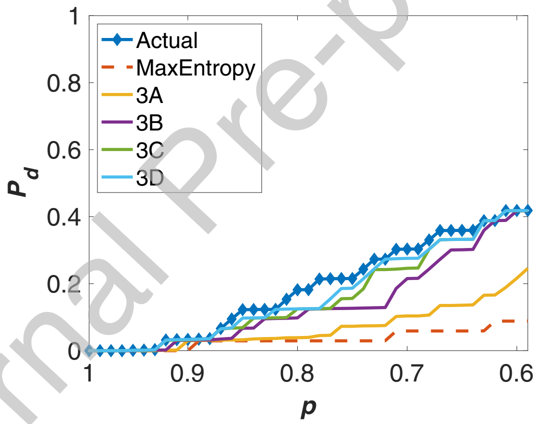
We apply each policy separately to the actual network and then conduct the systemic risk analysis on these modified networks. For this exercise, we explore the aggregated network in 2010, with $\alpha = 0.7$. We compare P_d of the modified networks to that of the actual network, and to MaxEntropy (the least risky network in this case).



(a) Bank merger



(b) Bank break-up



(c) Bank leverage cap

Figure 12: Effect of different policy exercises on P_d, relative to the actual network. MaxEntropy serves as the lower bound as it is the least risky network in this case. Here we use the data for 2010 with a market impact parameter of $\alpha = 0.7$.

Figure 12 shows the results for the three sets of policies. First, we find that merging the largest banks based on their total assets (1A) decreases P_d.[22] Our specific merging procedure yields a large but moderately leveraged bank. We illustrate this in Figure 13, where the merged bank (red colored bar) ends up holding 84% assets in the system and with a leverage ratio of 18. Moreover, by holding a majority of the assets in the system, it becomes less vulnerable to other banks' asset liquidations.[23] Figure 12

[22]This is different to the results of Greenwood et al. (2015) where the merger may lead to an even more leveraged bank.

[23]Contrary to Greenwood et al. (2015), we assume that banks only sell assets when they default. Under alternative assumptions (such as leverage targeting), the results may differ from what is reported here.

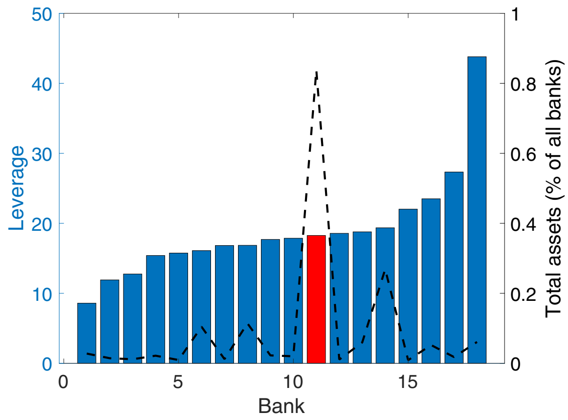


Figure 13: Leverage of the top 15% banks (total assets) for the 2010 data. There are 17 (out of 116) banks that belong to this category. The red bar refers to the leverage of a merged bank (resulted from merging the largest 15% banks), while the dashed black line refers to the total assets of the corresponding banks. Overall, the merger results in a very large but moderately leveraged bank.

shows that other merging procedures (1B, 1C and 1D) do not lower P_d as effectively. These results are mainly driven by the relatively insignificant total assets of the merged banks obtained from procedures 1B, 1C and 1D as shown in Table 8.[24] Moreover, we note that 1B is better than 1D to reduce the actual P_d. The intuition is similar to that of procedure with 1A: merging highly leveraged banks yields a moderately leveraged bank that is more stable during distress.

Figure 12 shows that breaking-up banks (2A and 2B) can increase systemic risk. Intuitively, one would expect that this policy should reduce the possibility of shock propagation since we break-up a given bank into two smaller banks that connect to different sets of firms/industries. However, this policy also leads to relatively concentrated banks that are more vulnerable to idiosyncratic shocks. As we only split the top 15% banks and keep the other 75% as they are, banks are still sufficiently interconnected to propagate the shock in the network.

Lastly, Figure 12 shows that a leverage cap can lead to substantially more stable networks, with tighter constraints yielding lower values of P_d. However, the results show that for modest leverage caps (such as scenario 3D) P_d remains largely unaffected. Hence, a substantial part of the observed vulnerability of the system is driven by banks' size and their portfolio overlap.

Overall, we find that neither of the three different policy exercises is able to bring down P_d to the values of the least risky network for this particular network (Max-Entropy). We find that merging banks and introducing a leverage cap may improve the robustness of the system, while splitting banks does not. These results can be dependent on the specific choice of the initial shock scenario.

[24]The leverage of banks in our datasets does not correlate to their size where highly and lowly leveraged banks might consist of both large and small banks.

7 Conclusions

There is widespread interest in finding accurate reconstruction methods for financial networks from partial information. In this paper, we focus on reconstructing and stress testing bipartite credit networks using detailed micro-data on bank-firm credit interactions in Japan for the period 1980 - 2010.

We find that there is no single "best" network reconstruction method - it depends on the assumed criterion of interest. This is also true when we look at each method's ability to reproduce observed levels of systemic risk. In fact, in many instances the actual credit networks display the highest levels of systemic risk, at least for the most disaggregated data. Hene, many reconstruction methods tend to underestimate systemic risk. Lastly, we find that the network aggregation level affects the individual performance of the different reconstruction methods.

Our findings suggest several interesting paths for future research. First and foremost, it is important to perform similar analyses for other datasets. Secondly, another important follow-up question is whether there are other reconstruction methods that are able to replicate the actual systemic risk levels more closely. In this paper, we only include a small number of popular reconstruction methods, but other methods may work better. Lastly, different stress tests can lead to different results. We therefore aim to generalize the modeling framework proposed here and test the robustness of the results in future research.

Acknowledgements

We thank Tiziano Squartini and Giulio Cimini for useful comments. We also thank participants of the 2018 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics, and 3rd workshop on Statistical Physics for Financial & Economic Networks at NetSci 2018. A.R. acknowledges a PhD scholarship of the Indonesia Endowment Fund for Education (LPDP). F.C. acknowledges support of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (ES/K002309/1 and ES/R009724/1). The views expressed in this paper represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

REFERENCES

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437.

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239.

Anand, K., Craig, B., von Peter, G., and Von Peter, G. (2015). Filling in the blanks: network structure and interbank contagion. *Quantitative Finance*, 15(4):625–636.

Anand, K., Van Lelyveld, I., Banai, Á., Friedrich, S., Garratt, R., Hałaj, G., Fique, J., Hansen, I., Jaramillo, S. M., Lee, H., Molina-Borboa, J. L., Nobili, S., Rajan, S., Salakhova, D., Silva, T. C., Silvestri, L., and De Souza, S. R. S. (2017). The missing links: A global study on uncovering financial network structures from partial data. *Journal of Financial Stability*.

Banwo, O., Caccioli, F., Harrald, P., and Medda, F. (2016). The effect of heterogeneity on financial contagion due to overlapping portfolios. *Advances in Complex Systems*, 19(08):1650016.

BIS (2015). Making supervisory stress tests more macroprudential: considering liquidity and solvency interactions and systemic risk. *BCBS Working Papers No 29.*

Blien, U. and Graef, F. (1998). Entropy optimizing methods for the estimation of tables. In *Classification, Data Analysis, and Data Highways*, pages 3–15.

Caccioli, F., Shrestha, M., Moore, C., and Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, 46:233–245.

Cimini, G., Squartini, T., Gabrielli, A., and Garlaschelli, D. (2015a). Estimating topological properties of weighted networks from limited information. *Physical Review E*, 92(4):1539–3755.

Cimini, G., Squartini, T., Garlaschelli, D., and Gabrielli, A. (2015b). Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports*, 5:15758.

Cont, R. and Schaanning, E. (2017). Fire sales, indirect contagion and systemic stress testing. *Norges Bank Working Paper*, 2.

Cont, R. and Wagalath, L. (2016). Fire sales forensics: measuring endogeneous risk. *Mathematical Finance*, 26(4):835–866.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms Third Edition.* The MIT Press.

Coval, J. and Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512.

De Masi, G. and Gallegati, M. (2012). Bank-firms topology in italy. *Empirical Economics*, 43:851–866.

Di Gangi, D., Lillo, F., and Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control*, 94:117–141.

Diamond, D. W. and Rajan, R. G. (2011). Fear of fire sales, illiquidity seeking, and credit freezes. *The Quarterly Journal of Economics*, 126(2):557–591.

Ellul, A., Jotikasthira, C., and Lundblad, C. T. (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics*, 101(3):596–620.

Fosdick, B. K., Larremore, D. B., Nishimura, J., and Ugander, J. (2016). Configuring random graph models with fixed degree sequences. *ArXiv e-prints:1608.00607v1*.

Fricke, C. and Fricke, D. (2019). Vulnerable Asset Management? The Case of Mutual Funds. *Journal of Financial Stability (forthcoming)*.

Fricke, D. (2016). Has the banking system become more homogeneous? Evidence from banks' loan portfolios. *Economics Letters*, 142:45–48.

Fricke, D. and Roukny, T. (2018). Generalists and specialists in the credit market. *Journal of Banking and Finance*, page doi: 10.1016/j.jbankfin.2018.04.01.

Gale, D. and Yorulmazer, T. (2013). Liquidity hoarding. *Theoretical Economics*, 8:291–324.

Gandy, A. and Veraart, L. A. M. (2016). A bayesian methodology for systemic risk assessment in financial networks. *Management Science*, pages 1–20.

Gandy, A. and Veraart, L. A. M. (2017). Adjustable network reconstruction with applications to cds exposures.

Glasserman, P. and Young, H. P. (2016). Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831.

Greenwood, R., Landier, A., and Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics*, 115(3):471–485.

Gualdi, S., Cimini, G., Primicerio, K., Clemente, R. D., and Challet, D. (2016). Statistically validated network of portfolio overlaps and systemic risk. *Scientific Reports*, 6:39467.

Haldane, A. G. (2015). On microscopes and telescopes. speech available at http://www.bankofengland.co.uk.

Hale, G., Krainer, J., and Mccarthy, E. (2015). Aggregation level in stress testing models. *FRBSF Working Paper*.

Hatzopoulos, V., Iori, G., Mantegna, R. N., Miccichè, S., and Tumminello, M. (2015). Quantifying preferential trading in the e-mid interbank market. *Quantitative Finance*, 15(4):693–710.

Huang, X., Vodenska, I., Havlin, S., and Stanley, H. E. (2013). Cascading failures in bi-partite graphs: model for systemic risk propagation. *Scientific Reports*, 3:1219.

Iori, G., Jafarey, S., and Padilla, F. G. (2006). Systemic risk on the interbank market. *Journal of Economic Behavior & Organization*, 61(4):525–542.

Iori, G., Mantegna, R. N., Marotta, L., Miccichè, S., Porter, J., and Tumminello, M. (2015). Networked relationships in the e-mid interbank market: A trading model with memory. *Journal of Economic Dynamics and Control*, 50:98–116.

Klincewicz, J. G. (1989). Implementing an exact newton method for separable convex transportation problems. *Networks*, 19(1):95–105.

Levy-Carciente, S., Kenett, D. Y., Avakian, A., Stanley, H. E., and Havlin, S. (2015). Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance*, 59:164–181.

Lillo, F. and Pirino, D. (2015). The impact of systemic and illiquidity risk on financing with risky collateral. *Journal of Economic Dynamics and Control*, 50:180–202.

Manconi, A., Massa, M., and Yasuda, A. (2012). The role of institutional investors in propagating the crisis of 20072008. *Journal of Financial Economics*, 104(3):491–518.

Mastrandrea, R., Squartini, T., Fagiolo, G., and Garlaschelli, D. (2014). Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics*, 16(4):043022.

Mazzarisi, P. and Lillo, F. (2017). Methods for reconstructing interbank networks from limited information: A comparison. In *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 201–215. Springer, Cham.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127.

Mohr, M. and Polenske, K. R. (1987). A linear programming approach to solving infeasible ras problems. *Journal of Regional Science*, 27(4):587–603.

Musmeci, N., Battiston, S., Caldarelli, G., Puliga, M., and Gabrielli, A. (2013). Bootstrapping topological properties and systemic risk of complex networks using the fitness model. *J Stat Phys*, 151(3-4):720–734.

Pulvino, T. C. (1998). Do asset fire sales exist? an empirical investigation of commercial aircraft transactions. *Journal of Finance*, 53(3):939–978.

Saracco, F., Di Clemente, R., Gabrielli, A., and Squartini, T. (2015). Randomizing bipartite networks: the case of the world trade web. *Scientific Reports*, 5(10595).

Shleifer, A. and Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, 25(1):29–48.

Squartini, T., Almog, A., Caldarelli, G., Van Lelyveld, I., Garlaschelli, D., and Cimini, G. (2017). Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks. *PHYSICAL REVIEW E*, 96.

Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A., and Garlaschelli, D. (2018). Reconstruction methods for networks: the case of economic and financial systems. *arXiv preprint arXiv:1806.06941*.

Squartini, T. and Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):83001.

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

Zhang, P., Wang, J., Li, X., Li, M., Di, Z., and Fan, Y. (2008). Clustering coefficient and community structure of bipartite networks. *Physica A*, 387(27):6869–6875.

Appendices

A Weight Allocation Methods

We use RAS (Blien and Graef (1998)) method to distribute the observed credit volumes across links for the generated adjacency matrix of CM1 and CM2. Previously, we experimented with different weight allocation approaches defined below and finally find that RAS generally performed best in our analysis (in term of corresponding L_1-error).

Weight allocation method	Definition
RAS (Blien and Graef (1998))	Column constraint $$\hat{w}_{i,j}(t+1) = \frac{\hat{w}_{i,j}(t)}{\hat{s}(t)^B_i,} \times s^B_i,$$ Row constraint $$\hat{w}_{i,j}(t+1) = \frac{\hat{w}_{i,j}(t)}{\hat{s}(t)^F_i,} \times s^F_i,$$ where t is the respective iteration step.
Linear Programming (Mohr and Polenske (1987))	$$Maximize \sum_{i=1}^{n^B} \sum_{j=1}^{n^F} c_{ij}, \hat{w}_{i,j}$$ $$subject\, to \sum_{i=1}^{n^B} \hat{w}_{i,j} = s^F_j \qquad (j = 1, \ldots, n)$$ $$\sum_{j=1}^{n^F} \hat{w}_{i,j} = s^B_i \qquad [i = 1, \ldots, (m-1)]$$ $$\hat{w}_{i,j} > (c_{i,j})(\epsilon)$$ where $b_{i,j} > 0 \rightarrow c_{i,j} = 1$, $b_{i,j} = 0 \rightarrow c_{i,j} = 0$
Convex transportation problem (Klincewicz (1989))	$$\hat{w} = (\hat{w}_{1,1}, \hat{w}_{1,2}, \ldots, \hat{w}_{2,n^F}, \ldots, \hat{w}_{n^B,1}, \hat{w}_{n^B,2}, ..., \hat{w}_{n^B,n^F})^T$$ $$s = (s^B_1, s^B_2, \ldots, s^B_{n^B}, s^F_1, s^F_2, \ldots, s^F_{n^F})^T$$ $$B\hat{w} = s$$
Maximum Flow (Cormen et al. (2009))	See Gandy and Veraart (2016) for the discussion on how to transform this into a maximum flow problem.

Table A.1: Summary of different weight allocation methods for the bank-firm network. Note that we can define the same methods for the bank-industry network.

B Systemic Risk Models

To quantify the vulnerability of the bipartite credit networks to systemic asset liquidations, we use the stress testing model of Huang et al. (2013) which uses a linear market impact, and assumes that banks do not target their leverage. This model is related with other models that have been recently introduced.

Here we compare the models based on the type of market impact function and whether it assumes some form of leverage targeting. First, we note that the model of Caccioli et al. (2014) uses a non-linear market impact and neglects leverage targeting, but then the leverage targeting is incorporated in the extended version of that model. Similar to the extended version of Caccioli et al. (2014), the model of Greenwood et al. (2015) incorporates leverage targeting, but assumes a linear market impact function. Cont and Schaanning (2017) do not include pure leverage targeting, but assume that banks have some regulatory constraint regarding their maximum leverage and banks will only liquidate when they exceed that maximum threshold. Another distinction between the two models is that even though the model of Cont and Schaanning (2017) also assumes a linear market impact for small volumes, they use a non-linear impact function with heterogeneous price impacts for each asset class.

		Market impact	
		linear	*non-linear*
Leverage targeting	*not-included*	Huang et al. (2013)	Caccioli et al. (2014)
	included with threshold		Cont and Schaanning (2017)
	included	Greenwood et al. (2015)	Caccioli et al. (2014) (extended)

Table B.1: Comparison between different stress testing model for bipartite credit network based on the type of market impact function used and whether leverage targeting is included or not.

C Additional Results: Systemic Risk Analysis on Other Models

For the purpose of finding out how the systemic risk analysis might vary if leverage targeting model (as in Greenwood et al. (2015)) and threshold model (as in Cont and Schaanning (2017)) are used, we also performed the same exercise with these other models. We find that the rank ordering of the different methods are generally consistent with those presented in the main text.

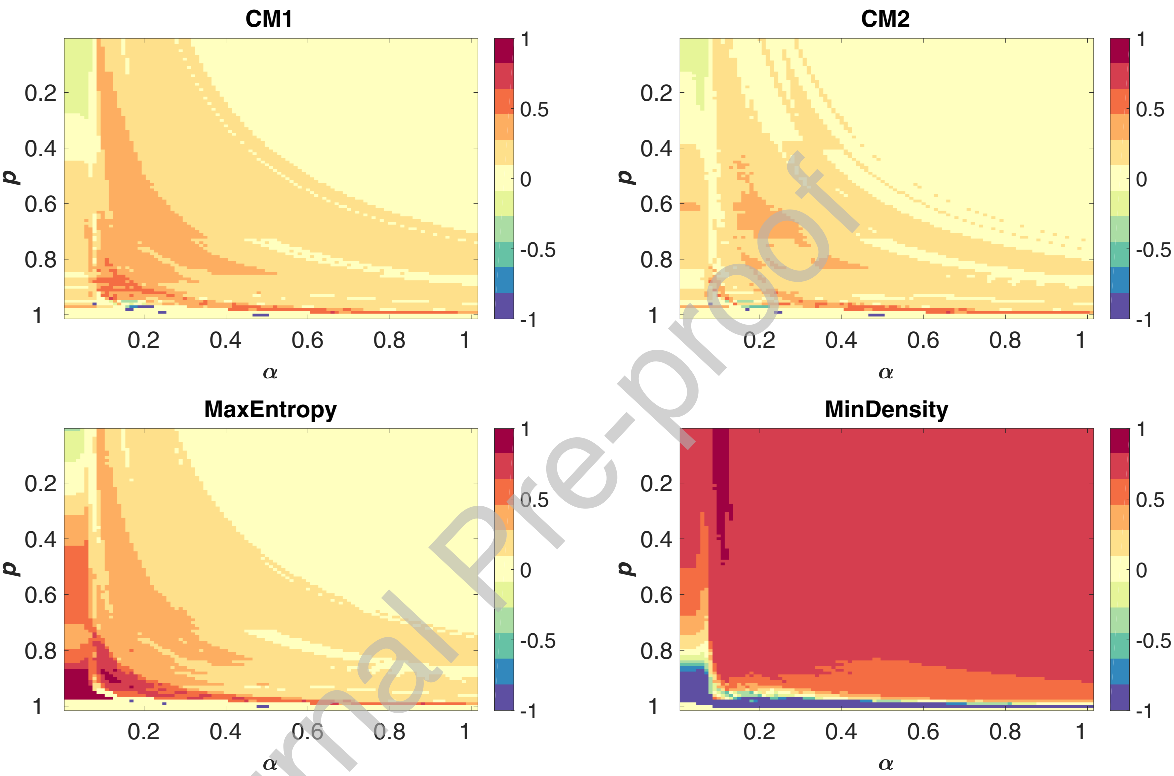


Figure C.1: Relative difference of the probability of default between actual network and the null models (D_r) at the aggregated level for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Leverage targeting model Greenwood et al. (2015)is used. Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation.

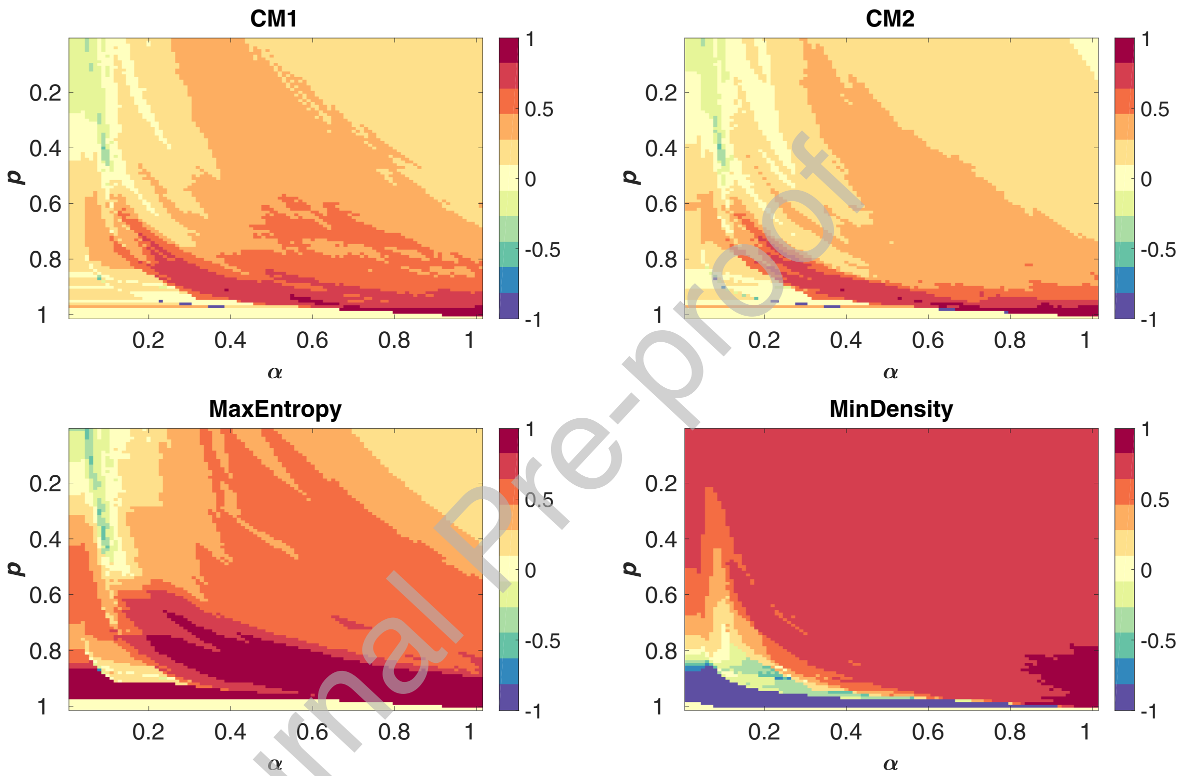


Figure C.2: Relative difference of the probability of default between actual network and the null models (D_r) at the aggregated level for $\alpha \in [0,1]$ (small to large market impact) and $p \in [0,1]$ (large to small initial shock). Threshold model (Cont and Schaanning (2017))is used. Data for year 2010. Warm color corresponds to an underestimation of the actual network, while cold color indicates an overestimation.

D Additional Results:

In Table 7, we previously computed the average of P_d across a range of restricted values of parameters that we defined as $p \in [0.6, 1]$ and $\alpha \in [0.6, 1]$. In the following, we again compute the average of P_d but for all possible values of parameters (including those that we define as restricted and not restricted).

Rank	Disaggregated		Aggregated		Intermediate	
	Null model	$\overline{P_d}$	**Null model**	$\overline{P_d}$	**Null model**	$\overline{P_d}$
1	Actual	0.393 (0.254)	Actual	0.360 (0.230)	Actual	0.360 (0.230)
2	CM1	0.301 (0.202)	CM1	0.218 (0.156)	MinDensity	0.358 (0.217)
3	CM2	0.243 (0.176)	CM2	0.217 (0.157)	CM1	0.275 (0.182)
4	MaxEntropy	0.190 (0.149)	MinDensity	0.202 (0.122)	CM2	0.241 (0.174)
5	MinDensity	0.140 (0.096)	MaxEntropy	0.190 (0.149)	MaxEntropy	0.190 (0.149)

Table D.1: Rank of the actual networks and the corresponding null models at different aggregation levels as in Table 7 for the 2010 data. However, as opposed to Table 7, where we consider only the restricted values of parameters, here we calculate the average of P_d across all possible parameter combinations: $p \in \{0, 0.01, 0.02, ... ,1\}$ and $\alpha \in \{0, 0.01, 0.02, ... ,1\}$. Rank 1 corresponds to the most risky network. $\overline{P_d}$ denotes the average, and the value inside the bracket is its corresponding standard deviation.

E Additional Results: Wilcoxon Signed Rank Test on the Networks

We formally test whether the difference between each network P_d is significant by running a two-sided Wilcoxon signed rank test on each pair of the actual network and the null model. In the tables below, we show the corresponding p-values of each test for different range of p and α. In Table E.1, we consider a range of parameters **within** the restricted values that we defined in the main text ($p \in [0.6, 0.8]$ and $\alpha \in [0.6, 0.8]$). Meanwhile, in Table E.2, we consider a range of parameters **outside** the restricted values that we previously defined. In particular, we consider all possible values of the initial shock ($p \in [0, 1]$) and relatively liquid assets ($\alpha \in [0, 0.5]$). Overall, we find that the difference between each network P_d is significant, except for CM1 and CM2 in some instances.

Disaggregated	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Aggregated	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		◇0.389	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Intermediate	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Table E.1: P-value of a two-sided Wilcoxon signed rank test on each pair of the network. A sufficiently small p-value indicates that the test rejects the null hypothesis that the difference between the pairs follow a symmetric distribution around zero, thus the two networks have significantly different P_ds. Meanwhile, a large p-value indicates that the test fails to reject the null hypothesis, thus the difference between the two networks P_ds is not significant. Here we test the P_d value of each network for $p \in [0.6, 0.8]$ and $\alpha \in [0.6, 0.8]$. We highlight the p-value above 0.05 using the $\diamond$ symbol.

Disaggregated	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Aggregated	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		⋄0.158	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Intermediate	CM1	CM2	MaxEntropy	Min-Density
Actual	0.000	0.000	0.000	0.000
CM1		0.000	0.000	0.000
CM2			0.000	0.000
MaxEntropy				0.000

Table E.2: P-value of a two-sided Wilcoxon signed rank test on each pair of the network as in Table E.1. However, here we consider $p \in [0, 1]$ and $\alpha \in [0, 0.5]$.