Axel Gandy and Luitgard A. M. Veraart

Adjustable network reconstruction with applications to CDS exposures

Article (Accepted version) (Refereed)

Adjustable network reconstruction with applications to CDS exposures

Axel Gandy[a], Luitgard Anna Maria Veraart[b,*]

[a] *Imperial College London, Department of Mathematics, London, SW7 2AZ, UK*
[b] *London School of Economics and Political Science, Department of Mathematics, Houghton Street, London WC2A 2AE, UK*

Abstract

This paper is concerned with reconstructing weighted directed networks from the total in- and out-weight of each node. This problem arises for example in the analysis of systemic risk of partially observed financial networks. Typically a wide range of networks is consistent with this partial information. We develop an empirical Bayesian methodology that can be adjusted such that the resulting networks are consistent with the observations and satisfy certain desired global topological properties such as a given mean density, extending the approach by Gandy and Veraart (2017). Furthermore we propose a new fitness-based model within this framework. We provide a case study based on a data set consisting of 89 fully observed financial networks of credit default swap exposures. We reconstruct those networks based on only partial information using the newly proposed as well as existing methods. To assess the quality of the reconstruction, we use a wide range of criteria, including measures on how well the degree distribution can be captured and higher order measures of systemic risk. We find that the empirical Bayesian approach performs best.

Keywords: Bayesian methods, calibration, matrix balancing, random graphs, systemic risk
2010 MSC: 62P05, 62F15

1. Introduction

Financial institutions are linked in various ways. It is natural to use network models to assess the stability of financial systems [3, 26, 28]. A key hindrance for using such models is that the networks are often not directly observable. As mentioned by Haldane in a 2015 speech (see p. 14 in [27]): "Even among the world's largest banks, data on their bilateral exposures to one another remains partial and patchy, especially for off balance sheet positions and securities holdings. That means large parts of the core of the international banking map remain, essentially, uncharted territory." Furthermore, as mentioned on p. 31 of Bank of England's document [5], there are "big gaps in the data on interlinkages between different parts of the financial system and common exposures across the financial system. The lack of data makes it difficult to build up a point-in-time picture of the interlinkages between different parts of the financial system and calibrate quantitative models." This is despite efforts by regulators to collect more data [23]. We develop an empirical Bayesian approach to obtain a distribution of networks from partial information.

A financial network can be characterized in terms of a matrix where each entry corresponds to a financial link, e.g., a loan from one market participant to another one in the system or an exposure due to the sale of products such as a Credit Default Swap (CDS). We consider the classical network reconstruction problem, in which the row and column sums of the network matrix are the only available information. This is, e.g., the case for interbank assets and liabilities, where the aggregate information is available from balance sheet information. Sometimes additional entries of the matrix might be known to the regulator, since they correspond to (large) exposures within the regulator's jurisdiction [4, 33].

The classical network reconstruction problem in which approximately n^2 unknowns (the entries of the matrix) are estimated from around $2n$ observations seems to be ill-posed. Indeed, usually a large variety of matrices is consistent with the marginals. This is somewhat comparable to problems in high-dimensional statistics in which a large number p of parameters is estimated from a small number n of observations with $n \ll p$, where useful results

*Corresponding author

can be obtained by assuming a structure of the solution, such as sparsity. Similarly, in the matrix reconstruction problem, such assumptions will have to be made.

Recently, the performance of six different approaches to network reconstruction has been compared on 23 different financial networks [4]. A slightly extended and updated comparison can be found in [2]. We briefly outline these and some other approaches in Section 2. The study finds that there is no single method that is best for all types of networks. For example, the degree of sparsity influences which method works best. This then leads to recommendations as to which methods work best for different types of networks depending on their density/structure.

The most important conclusion from this study, which was not actually made, is that it is crucial for a network reconstruction method to be adjustable to topological characteristics such as the density of the network, which we define as the proportion of links existing divided by the total number of possible links. It is in general not possible to infer the density of the network from the marginals alone [24]. Hence, one can construct matrices of (almost) any desired density that are consistent with the observed marginals. Any network reconstruction technique that does not allow for calibration to additional information or target characteristics cannot be successful over a wide range of different types of networks.

Recently, Gandy and Veraart [24] proposed a Bayesian approach to network reconstruction that is very flexible in terms of the underlying structure of the network. The Bayesian approach assumes a prior distribution for the financial network which is then conditioned on the available information. This posterior distribution is approximated using a Markov Chain Monte Carlo (MCMC) sampler. The sampled matrices, i.e., the financial networks, are consistent with the row and column sums (and potentially some additional observed entries). Several well-known modeling assumptions can be included in the prior model, such as, e.g., the Erdős–Rényi random graph model, core-periphery structures or power laws for both the degree distribution and the distribution of the weights. Gandy and Veraart [24] do not give a method for adjusting the method to desirable topological properties. They also do not test their approach on empirical data in which the full network is actually observable.

Our paper makes three main contributions. First we develop an empirical Bayesian approach to network reconstruction by developing a calibration methodology of the Bayesian approach proposed in [24]. In particular, we show how networks can be reconstructed that have a specified target density and are consistent with the observed marginals, and possibly additional entries; see Section 4.3.

Second we introduce in Section 4.4 an empirical fitness model within the Bayesian framework and show how this can be calibrated to topological information about the network density. This new model accounts for heterogeneity in the nodes of the network while still being parsimonious in the number of parameters and hence easy to calibrate.

Third we evaluate in Section 5 the performance of the new methods as well as of existing methods on real data. For this, we use 89 fully observed networks of Credit Default Swaps (CDS) exposures described in Section 3, assume that we do not observe the full networks and attempt a reconstruction. Based on a wide range of measures, the (empirical) Bayesian approach performs very well in absolute terms and also compared to alternative reconstruction techniques. When looking at criteria that measure systemic risk or check whether the probability distribution from which the reconstructed networks are sampled has the true observed weights of the links within their support or not, we find that the (empirical) Bayesian approach performs significantly better than all competitor models considered in this paper. This is particularly remarkable because we included a competitor in the comparison that was considered the "winner" among the probabilistic approaches to network reconstruction considered in the comparison study by Anand et al. [2], which did not consider the Bayesian approach by Gandy and Veraart [24] and our extension. Furthermore, the (empirical) Bayesian approach can be calibrated quite easily to match important quantities of interest. We provide specific results for one example network in Section 5.3 and results for all 89 networks in Section 5.4.

We compare theoretical properties of different network reconstruction methods in Section 6, and find that the (empirical) Bayesian approach has more desirable properties than all the alternatives considered. We provide an implementation of the methods introduced in this paper as part of the R-package `systemicrisk`.

2. Literature review

The problem of reconstructing networks characterized as matrices from their row and column sums has been considered in a wide range of fields and is also referred to as matrix balancing problem; see, e.g., [41] for an overview. Specifically, it is related to the problem of constructing tests for independence in contingency tables; see [8, 37] for

some recent work on this. A key difference is that the contingency tables are usually binary or integer-valued, whereas we do not assume any such restriction in our approach. Furthermore, a uniform distribution given the row and column sums is usually of interest — we assume more structure.

It is also related to the inference of origin-destination matrices for a transit system, where only the number of passengers entering and leaving a vehicle at stops is observed. One specific example of such a problem is considered in [29], where a Bayesian approach for modeling the system is used. A key difference is that in this application the unknown matrices are integer-valued triangular matrices.

As long as only the aggregates are available, no actual network data such as edge weights are observed, and hence the substantial body of literature on statistical inference of network data cannot be used directly; see, e.g., [32] for a good overview and [22] for some recent challenges. What can be used, however, are models for random graphs that have been studied in detail in the context of the statistical analysis of network data, probability and graph theory; see, e.g., Chapter 6 in [32] and [10].

There are two classes of reconstruction methods. First, methods that return only one matrix, usually a solution of a constrained optimization problem. Second, methods that return an ensemble of matrices, often based on random graph models.

The most widely used approach of the first class is to minimize the Kullback–Leibler divergence between the network matrix and a previously specified input matrix subject to the linear constraints imposed by the row and column sums [48]. The classical input matrix contains in row i and column j the product of the row sum of i times the column sum of j divided by the sum of all row sums. This optimization problem can be solved efficiently using, e.g., the RAS algorithm [41]. The RAS algorithm is also known as iterative proportional fitting algorithm and has been discovered and rediscovered in many fields; see Section 2.1 in [41] for some history and references.

We will consider this approach in our case study. Empirical studies based on this method have been conducted for banking networks of several countries, e.g., for Germany [48], for the UK [20, 49] and for Belgium [16]. The main drawback of this approach is that based on the classical input matrix, the reconstructed network is usually a complete network which is inconsistent with the sparsity of many real financial networks [14]. Mistrulli [38] has shown in an empirical study of the Italian interbank market that the Kullback–Leibler method can underestimate systemic risk. One could try to at least partially overcome this problem by choosing different input matrices for the Kullback–Leibler method [6, 17].

Another example of the first class of methods is the minimum density method [1], which finds the network with the minimum number of edges that is consistent with the aggregated interbank assets and liabilities. Anand et al. [1] also discuss an extension that would result in a less sparse network.

In the second class of methods, there are several sampling approaches which generate an ensemble of networks that match in expectation a certain real-world property of a network while maximizing the entropy of the ensemble [40]. For instance, one might want to match a specific degree sequence of a network. Based on this idea, so-called fitness models have been developed. Every node is equipped with a fitness that models, e.g., the size of the node and link existence probabilities are functions of the corresponding fitnesses; see [11].

There are general results which show that using additional information on topological information provided by information linked to the degree distribution results in much better reconstruction results than looking at the in- and out-strength alone [35, 44–46].

An application of fitness models and additional use of topological information for reconstructing financial networks is provided by Cimini et al. [13]. They assume that in- and out-strengths are observable and in addition that there is a subset of nodes for which the in- and out-degrees are known. The link-existence probabilities are functions of the fitnesses and a parameter that is chosen such that the expected in- and out-degrees coincide with the observed in-and out-degrees of the observed subnetwork. While a rather advanced methodology is used here to sample the adjacency matrix, the corresponding weights are assigned based on a degree-corrected gravity model. This implies that each entry in the matrix can only take two values: zero and another (cell-specific) nonnegative value. Hence, while these models can generate a wide range of structures for the adjacency matrix, the range of the corresponding weights is very limited. All networks in the sample satisfy the row and column sums only in expectation. We will use this approach in our comparative study of performance of different network reconstruction techniques. Other applications of fitness models for reconstructing financial networks can, e.g., be found in [7, 9].

Another simulation-based approach based on random graph models was considered by Moussa [39]. The weighted network is generated from an externally specified random graph model and not just the adjacency matrix. These

networks do not satisfy the row and column sum constraints. Therefore a suitable weight is chosen for every sample network such that the sum of the weighted samples satisfies the constraints.

Hałaj and Kok [25] propose to simulate financial networks based on a probability map that they construct from additional empirical data and use as the model for link existence probabilities in the sampling. Weights are then assigned to the links in a rather ad hoc way, but they do satisfy the row and column constraints exactly and not just in expectation.

A conceptually different approach from all previous approaches is the Bayesian approach by Gandy and Veraart [24]. It allows to incorporate stylized facts about both the degree distribution and the distribution of weights as a prior model in a Bayesian context. In particular, several well-known random graph models can be used. It then generates samples through an MCMC sampler that satisfy the row and column constraints exactly and not just in expectation. To the best of our knowledge, the Bayesian approach is the only currently available approach that generates an ensemble of network matrices that match the constraints on the row and column sums exactly and at the same time incorporates stylized facts of financial networks in a rigorous way.

The literature on network reconstruction has paid little attention so far to the way in which network reconstruction methods can be adjusted or calibrated in a parsimonious way to deal with different types of networks. Exceptions are [36] and [13]. Mastromatteo et al. [36] have proposed a message-passing algorithm for estimating interbank exposures. They choose a global level of sparsity for the financial network first and then once the adjacency matrix is constructed consistent with this sparsity level, they distribute weights along existing edges similarly to the Kullback–Leibler method. Cimini et al. [13] assume that in addition to the row and column sums, they observe the in- and outdegrees of a subset of nodes in the network. They use this information to calibrate a parameter in their fitness model. Gandy and Veraart [24] demonstrate how different a priori assumptions lead to different types of reconstructed networks, but did not develop a calibration methodology.

3. Data

Throughout the article, we use a snapshot of roughly 134,000 outstanding positions in credit default swaps referencing 89 different UK institutions, taken in the second half of 2011. The data come from the Depository Trust & Clearing Corporation's (DTCC) Trade Information Warehouse (TIW) and were supplied to us by the Bank of England with anonymized counterparties.

For each reference entity, the data record both counterparties of a position (buyer and seller), the notional amount and whether the counterparties are a dealer or a non-dealer. The notional amounts are quoted in EUR — we ignore the few (less than 100) entries that are quoted in other currencies. As mentioned on p. 3 of [18], the notional amount "represents the par amount of credit protection bought or sold, equivalent to debt or bond amounts, and is used to derive the coupon payment calculations for each payment period and the recovery amounts in the event of a default."

We construct 89 networks as follows. For every UK institution being referenced, we construct a network between buyers and sellers describing the total outstanding positions in credit default swaps referencing this particular institution. Sometimes there are multiple entries, i.e., for a given reference entity, a pair of buyer and seller is listed more than once. These multiple entries correspond to different maturities. When this occurs, we add up all the multiple entries to obtain the total weight for such an edge.

Figure 1 shows one of these networks, which we will call example network. Thirty-two counterparties are active in this network. The sum of all weights has been normalized to 1. Rows refer to sellers and columns to buyers. Counterparties are ordered first such that non-dealers are before dealers and second by the total size of their position (sum of credit protection bought and sold).

Table 1: Summary statistics for the size and density of the 89 networks.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Size n	4	30	40	45	57	117
Density	0.035	0.098	0.133	0.154	0.191	0.429

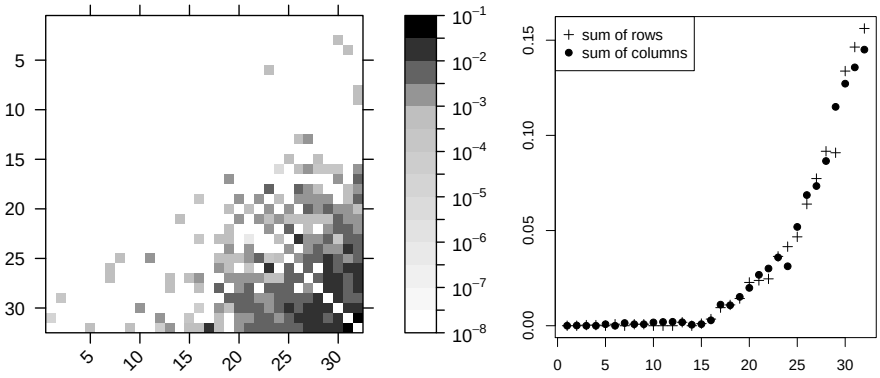

Figure 1: Left: Normalized exposure network for the example reference entity (row: protection seller, column: protection buyer). Right: Row and column sums of the matrix.

Table 1 reports summary statistics for the 89 networks. Their size varies between 4 and 117 nodes with a mean number of 45 nodes. The density varies between 3.5% and 42.9% with mean density of 15.4%.

4. The empirical Bayesian approach

4.1. The problem

This section recaps the network reconstruction problem as described in [24]. We consider a financial network that consists of $n \in \mathbb{N} = \{1, 2, \ldots\}$ nodes with indices in $\{1, \ldots, n\}$ and weighted directed edges given by a nonnegative matrix $L \in [0, \infty)^{n \times n}$. An edge from node i to note j exists if $L_{ij} > 0$ and then L_{ij} defines the weight of this edge.

We assume that we have two types of information about the matrix L. First, we assume that we know the row and the column sums of L given by vectors $\ell, a \in [0, \infty)^n$, i.e.,

$$\forall_{j \in \{1,\ldots,n\}} \sum_{i=1}^{n} L_{ij} = a_j, \quad \text{and} \quad \forall_{i \in \{1,\ldots,n\}} \sum_{j=1}^{n} L_{ij} = \ell_i.$$

Second, we assume that some elements of L are known, by requiring

$$\forall_{i,j \in \{1,\ldots,n\}} L^*_{ij} \neq * \quad \Rightarrow \quad L_{ij} = L^*_{ij}.$$

where $L^* \in \mathcal{L}^* = (\{*\} \cup [0, \infty))^{n \times n}$. The symbol $*$ signifies that it is not known whether the corresponding edge exists and if so what weight it has.

A matrix L that satisfies these conditions is called admissible for a, ℓ respecting L^*. Conditions on a, ℓ and L^* for the existence of such admissible matrices are known; see Theorem 2.3 in [24].

Often, one assumes that there are no links between a node and itself, which is specified by setting $L^*_{ii} = 0$ for $i \in \{1, \ldots, n\}$ and $L^*_{ij} = *$ for $i \neq j$. We refer to this as the minimal observation setting.

Consider the example network given on the left panel of Figure 1. The reconstruction of the network would be only based on the row and column sum of this matrix, displayed on the right panel of Figure 1. In this example the row and column sums are reasonably similar — the methods, however, do not rely on this.

4.2. Overview of the Bayesian approach to network reconstruction

Gandy and Veraart [24] approach the network construction as a Bayesian problem, by specifying a generative model for L and then conditioning on the observations. In other words, the network reconstruction is given through the posterior distribution of L.

As underlying model, they used the following model. First it constructs an unweighted graph based on a generalized version of the Erdős–Rényi model [21]: directed edges from i to j are generated via independent Bernoulli trials with success probabilities $p_{ij} \in [0, 1]$ for all $i, j \in \{1, \ldots, n\}$. Second, weights sampled from an exponential

distribution (denoted by $\mathcal{E}(\lambda)$, where λ denotes the rate) are assigned to the existing edges. Formally, the model is as follows:

$$\Pr(L_{ij} > 0) = p_{ij}, \quad L_{ij}|L_{ij} > 0 \sim \mathcal{E}(\lambda_{ij}).$$

The model parameters consist of two matrices: $p \in [0,1]^{n\times n}$, where p_{ij} is the probability of the existence of a directed edge from i to j, and $\lambda \in (0,\infty)^{n\times n}$, which governs the distribution of the weights given that an edge exists.

In Section 4.3, we will work with the basic model where all entries of p and λ are identical. In Section 4.4, we discuss a fitness based model, where the entries of p depend on the size of the institutions involved; for simplicity, all entries of λ will still be the same.

To obtain the network reconstruction, one needs the probability distribution of L conditional on being admissible for a, ℓ, respecting L^*. Since this probability distribution is not available in closed form, Gandy and Veraart [24] have constructed an MCMC sampler for this distribution. We will use this sampler for examples in this paper.

4.3. Calibrating the basic model

Choosing the model parameters, i.e., the two matrices λ and p, is not straightforward. Often, there would be little information about this even if one assumed that the observed network was actually generated by a generalized Erdős–Rényi model with exponential weights. Proposition 3.1 in Gandy and Veraart [24] implies that under some mild conditions, one can find for any choice of p a corresponding λ such that the expected row and column sums coincide with the observed ones. In particular, this implies that characteristics such as the density of the network cannot be observed from the marginals alone. This is a principal problem of network reconstruction and not only a problem of the Bayesian approach.

In the following we show how to calibrate the Bayesian approach. By doing so we move away from a purely Bayesian methodology towards an empirical Bayesian approach in which prior modeling assumptions are derived from empirical data. Our main goal is to find a choice of parameters that results in a good reconstruction of the network. We will see that a suitable choice of parameters might still be able to produce good reconstruction results even if the underlying modeling assumptions are not completely satisfied.

In this section we consider the basic model, where all entries of p and λ are identical. We also denote these individual entries by p and λ. We use this simple model because it is a natural starting point in practice if one does not have any additional information about the network that one wants to reconstruct. Furthermore, we will see that it performs well in our empirical tests.

4.3.1. Parameter of the conditional distribution

We first demonstrate that the posterior $\mathcal{L}$ of L, i.e., the conditional distribution of L given row and column sums and potentially some known entries, does depend on the two one-dimensional parameters p and λ only indirectly through the one-dimensional parameter $\kappa = p\lambda/(1-p)$. Indeed, if L is an admissible network matrix for a, ℓ respecting L^*, we have

$$\begin{aligned}\mathcal{L}(L|\ell, a, L^*) \propto \prod_{i,j} p^{\mathbf{1}\{L_{ij}>0\}} \lambda^{\mathbf{1}\{L_{ij}>0\}} e^{-\lambda L_{ij}} (1-p)^{\mathbf{1}\{L_{ij}=0\}} &= \left(\frac{p\lambda}{1-p}\right)^{\sum_{i,j}\mathbf{1}\{L_{ij}>0\}} e^{-\lambda A}(1-p)^K \\ &\propto \left(\frac{p\lambda}{1-p}\right)^{\sum_{i,j}\mathbf{1}\{L_{ij}>0\}} = \kappa^{\sum_{i,j}\mathbf{1}\{L_{ij}>0\}},\end{aligned}$$

where $\propto$ means that both sides are proportional when considered as function of the unknown entries of L, products and sums are over these unknown entries, K is the number of these unknown entries and $A = \sum_{i,j} L_{ij}$. The second line follows because both K and A are the same for all admissible matrices. For non-admissible matrices, $\mathcal{L}(L|\ell, a, L^*) = 0$.

Thus, for a reconstruction with the basic model, only the choice of κ matters. Figure 2 illustrates the effect of this choice. It shows the true example network in the left most picture, whose empirical density is roughly 0.21. Additionally, it shows three randomly selected sample matrices that were generated from models where κ was chosen such that the density of the network has specific values (0.1, 0.21 and 0.5). Even if the parameter choice is not optimal, the overall shape of the network, namely that most of the edges occur in the lower right corner is reproduced. Not surprisingly, the example matrix is best matched by the matrix from the sampler calibrated to the true density.

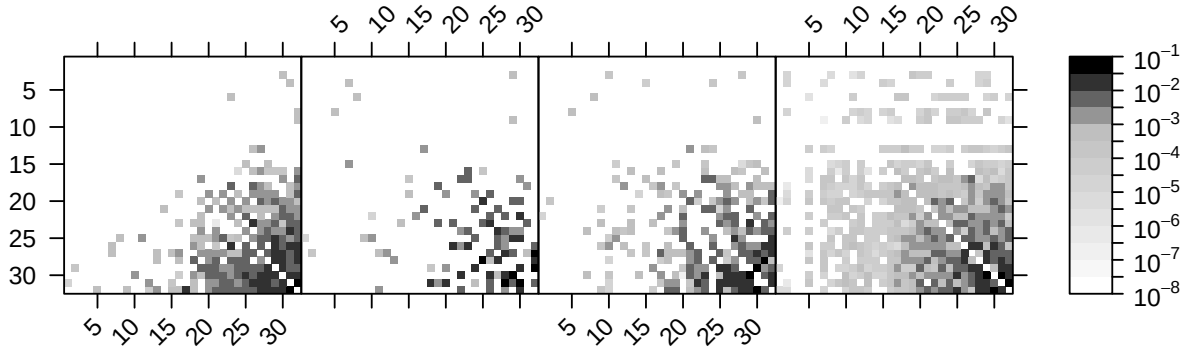

Figure 2: Example network (left) and reconstructions of the example network calibrated to several densities (from left to right: 0.1, 0.21, 0.5).

4.3.2. Calibration to a property of interest

Motivated by this example, we suggest to choose κ to ensure that the reconstructed network matches a certain desirable property. For example, we might want to ensure that the mean density of the network, i.e., the proportion of existing links, takes on a certain value; this is what was done in Figure 2. Other desirable properties could include a certain fixed mean in-degree or a fixed mean out-degree. Alternatively, one might want to match certain quantiles of the (in- or out-) degree distribution. For example, one could specify the (in- or out-) degree of one of the most connected nodes by targeting, e.g., the 95% quantile of the degree distribution.

To put this in more formal terms, let $H : \mathbb{R}^{n\times n} \to \mathbb{R}$ be a function that computes a property of a matrix such as the proportion of positive entries. Suppose we are interested in choosing κ such that, for a given τ,

$$\tau = \mathrm{E}\{H(L)|\ell, a, L^*, \kappa\}.$$

The existence of a solution to this equation will very much depend on the choice of the function H. For any fixed κ, we can approximate the right-hand side by running the MCMC sampler. Thus, one can solve the above equation to a reasonable precision (provided a solution exists) using a one-dimensional numerical root search. Even though every step of the root search requires new MCMC samples, this is computationally feasible, if one does not require a high precision.

4.3.3. Computational considerations

For numerical reasons, it is sensible to vary p and to fix λ to a reasonable value. This is because λ appears in exponents in quantities computed by the MCMC sampler and should thus not be too extreme. In the minimal information setting, we could for example set $\lambda = \lambda(p) = pn(n-1)/A$, which ensures that the expected total entries of the matrix (which is $n(n-1)p/\lambda$) is equal to the observed sum A of all entries. This choice is independent of the choice of the function H used in the calibration. In the example in which we calibrate to a certain mean density given by p^*, we would then set λ to be $p^*n(n-1)/A$.

To obtain the reconstructions in Figure 2, we use the implementation of [24] for sampling matrices within the standard R root finding method to calibrate the models to the desired density, with the convergence tolerance set to 0.01. In each of these runs, a fixed amount of thinning ($200 \times n^2$ steps) producing 100 matrices is used. Additionally, 5% of the effort for sampling the matrices is used as burn in. For example, for calibrating to a density of 0.21, the root finding algorithm uses seven steps, leading to a total of $n^2 \times 200 \times 100 \times 7 \times 1.05 \approx 1.5^8$ basic update steps. These choices were made so that the resulting density will be within ± 0.02 of the target density. The central part of the code is written in C++, allowing for fast execution. In the example for calibrating to a density of 0.21, the execution time on one core of a Intel i7-6700 CPU with 3.40 GHz is about 32s. Once the density has been calibrated, the MCMC chain will be run once more, with the amount of thinning chosen such that a good effective sample size is reached.

4.3.4. Using additional information to estimate the density

Thus the choice of the model parameter does matter and using the density of the network to calibrate, it appears reasonable. In general, however, as shown in Proposition 3.1 of [24], it is not possible to infer the density of a network

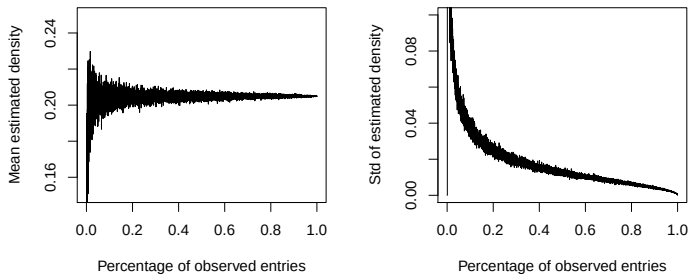

Figure 3: Estimating the density from different percentages of randomly selected observed entries in the example network given in Figure 1.

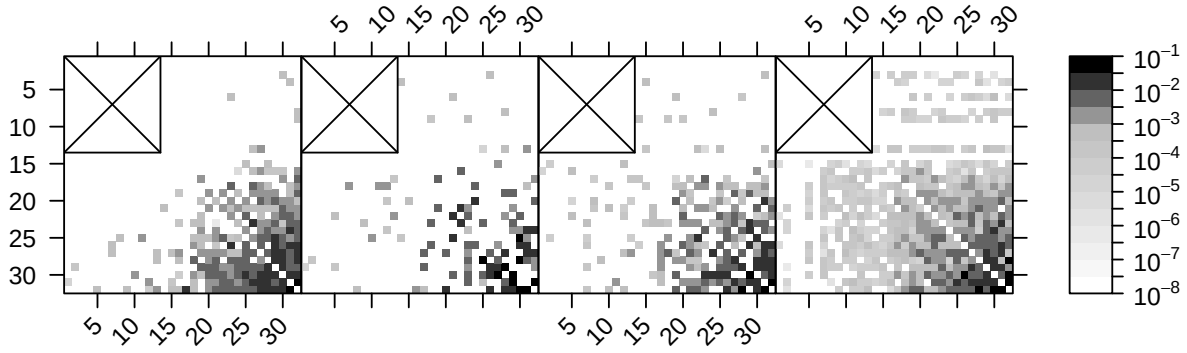

Figure 4: Matrix of example network (left) and reconstructions of the example network calibrated to several densities (from left to right: 0.1, 0.21, 0.5) with additional information that non-dealers do not trade with non-dealers — these fixed values are crossed out.

from the marginals alone. In practice, one might have additional information available that can be used to choose the target density for the calibration. For example, one might have historical data available about (parts of) the network that can be used.

Alternatively, one might observe additional entries of the network and can use these to estimate the target density. In Figure 3 we show that if one selects a certain number of elements at random from the underlying true network and uses these elements to estimate the density, this estimate converges to the true density quite quickly.

However, when estimating network properties from observable entries, one needs to be careful about a potential bias. Often there are special reasons why certain entries are observable and this might lead to a biased estimation of the density.

Furthermore, there is a substantial body of statistical research on the problem of whether statistical inference conducted on a subnetwork generalizes to the full network and hence is consistent under sampling; see, e.g., Chapter 5 in [32] and [43]. As shown in [43], several popular random graph models such as, e.g., some of the class of exponential random graph models are not consistent under sampling. Hence, as soon as additional information on data from a subnetwork is available, one needs to be very careful on how it is used in the reconstruction problem.

4.3.5. Additional known entries

One feature in the data is not well captured by the reconstructions in Figure 2: the top left corner does not contain any link. This corner consists only of non-dealers and there are no trades between non-dealers and non-dealers for this particular reference entity (in the whole data sets there are less than 50 trades between non-dealers). We can include the additional information that there are no such trades by setting the corresponding entries of L^* to zero. Doing this gives samples as shown in Figure 4 that respect this special structure.

4.4. Empirical fitness model

Figure 1 (the observed example network) demonstrated a tendency of small nodes (size measured in terms aggregated row and column sums) to link up with larger nodes and not with other smaller nodes. The classical Erdős–Rényi

model in which $p_{ij} = p$ for all i, j does not allow for such a behavior. Gandy and Veraart [24] have proposed a fitness model which can account for heterogeneity of the nodes by relying only on a small number of parameters. The idea is that each node is equipped with a random fitness x_i (with probability density function denoted by ρ) and the probability that there exists a directed edge between two nodes i and j is then a suitable function $f(x_i, x_j)$ of the corresponding two fitnesses. In the literature on large networks such functions are called graphons; see, e.g., [34]. Gandy and Veraart [24] have shown that suitable fitness models can mimic some characteristics observed in the data quite well. Their fitness model was a purely Bayesian model, in which also the fitness is assumed to be random and is learned from the observed row and column sums.

In a non-Bayesian context random fitness models have been considered in, e.g., [11, 42] to model undirected non-weighted networks. Different choices for f and ρ have been studied and they result in different degree distributions for the generated network. In particular, Servedio et al. [42] have shown how one can generate power laws for the degree distribution using suitable fitness models.

Non-random fitness models have been used, e.g., in the context of analysing financial networks [15] and also in the context of network reconstruction [13].

4.4.1. The model

In the following we will assume that every market participant i is assigned a deterministic fitness $x_i \in [0, \infty)$. We discuss specific choices later. For the link existence probability we set, for all $i, j \in \{1, \dots, n\}$, $p_{ij} = \alpha x_i x_j/(1 + \alpha x_i x_j)$, and $\lambda_{ij} = \lambda$, where α and λ are some given constants. In other words the rates are the same for all observations but the link existence probabilities can be influenced by the fitnesses.

The link existence probabilities are nondecreasing functions in the fitnesses. Often fitnesses are used to represent the size of the node. Such a monotonic relationship would then capture the effect that larger nodes are connected to each other with higher probabilities.

We also have the symmetry $p_{ij} = p_{ji}$ for all i, j. As we have seen in Figure 1, there are many situations in which nodes in financial networks are not net buyers or net sellers but are usually involved in both types of transactions. In particular, it is often observed in CDS markets that dealers are mainly doing intermediation, which is also sometimes referred to as "hot potato trading". These dealer-dealer trades account for a large amount of the total volume observed in our data. In such situations, the symmetry assumption above exactly reflects the passing of the hot potato.

One could extend this model by assigning a two-dimensional fitness vector $(x_i^{\text{in}}, x_i^{\text{out}})$ to every node i and then setting, for all $i, j \in \{1, \dots, n\}$,

$$p_{ij} = \alpha x_i^{\text{out}} x_j^{\text{in}}/(1 + \alpha x_i^{\text{out}} x_j^{\text{in}}). \tag{1}$$

Then, the matrix p is not necessarily symmetric anymore.

This functional form of the fitness (1) has, e.g., been used by Cimini et al. [13] in their approach to reconstructing financial networks. It can be interpreted as a "fitness-induced configuration model"; see [13] for a detailed discussion. In the random graph literature these types of link existence probabilities are considered in the so-called β-model for undirected (non-weighted) graphs (using a slightly different parametrisation and assuming that what we call fitness here are parameters that need to be estimated) in [12] and for directed (non-weighted) graphs in [30].

Despite having a similar functional form our approach is still fundamentally different. Cimini et al. [13] use this fitness model to model the adjacency matrix and then attach weights to existing links based on the "degree-corrected gravity model". The weights that are assigned to existing links are deterministic and hence the weight for a particular edge is always the same for all samples for which this link exists. Hence, the individual samples will usually violate the constraints imposed by the observed row and column sums and the observed row and column sums are only matched in expectation. Our (empirical) Bayesian approach is designed, such that the observed marginals are matched exactly by every single sample. Furthermore, the Bayesian approach allows for random weights.

4.4.2. Parameter of the conditional distribution

In the basic model, conditional on row and column sums, the distribution of the matrix L has only one parameter. This is also the case for our fitness model. Indeed, the posterior of L in our fitness model is

$$\mathcal{L}(L|\ell, a, L^*) \propto \prod_{i,j} p_{ij}^{\mathbf{1}\{L_{ij}>0\}} \lambda^{\mathbf{1}\{L_{ij}>0\}} e^{-\lambda L_{ij}} (1-p_{ij})^{\mathbf{1}\{L_{ij}=0\}} = \left(\prod_{i,j:L_{ij}>0} \frac{p_{ij}\lambda}{1-p_{ij}} \right) e^{-\lambda \sum_{i,j} L_{ij}} \left\{ \prod_{i,j} (1-p_{ij}) \right\}$$
$$\propto \prod_{i,j:L_{ij}>0} \frac{p_{ij}\lambda}{1-p_{ij}} = (\alpha\lambda)^{\sum_{i,j} \mathbf{1}\{L_{ij}>0\}} \prod_{i,j:L_{ij}>0} x_i x_j.$$

This depends on the two parameters α and λ only via their product. Thus, for this empirical fitness model, in the reconstruction we still have only one parameter to choose. Hence we can again calibrate the model to a desired property (such as a mean density). Indeed, we can accomplish this by fixing λ to a reasonable constant and by running a numerical search with the parameter α.

4.4.3. Choosing the fitness

We now discuss possible approaches to determining the fitness parameters $x_1, \ldots, x_n$. Gandy and Veraart [24] have already discussed the full Bayesian extension in which also fitnesses are assumed to be random and have a prior distribution. Now we look into using empirical data to assign fitnesses.

In the absence of any additional information about the underlying network other than its row and column sums, we can use these aggregates to determine fitnesses. In general, the observed row and column sums are outcomes of a Bayesian model and should thus not be used as direct inputs into the model selection. However, if the focus is more on a simple empirical reconstruction method than on having a fully coherent Bayesian approach, then we can attempt to use this information. This could be seen as an empirical Bayesian approach. In particular we can then assume that the fitness of a node is a function of its row and column aggregates.

More specifically, in this paper we consider the following model. Every market participant i will be assigned a fitness x_i which we set to be $x_i = \ell_i + a_i$. Since we use empirical data to choose p we refer to this model as empirical fitness model.

In the non-Bayesian context, Cimini et al. [13] have considered two-dimensional fitness vectors with $(x_i^{\text{in}}, x_i^{\text{out}}) = (a_i, \ell_i)$ for all i. We do not pursue this extension in the present paper because in our data each row aggregate ℓ_i is similar to the corresponding column aggregate a_i; see Figure 1.

If more information than the row and column aggregates is available, the fitnesses could for example be estimated using external covariates in a suitable regression model. If historical data of the network are available, these could also be used to estimate fitnesses.

4.4.4. Application to the example network

We now reconstruct our example network with the empirical fitness model, calibrated to mean densities 0.1, 0.21 and 0.5; see Figure 5. Again, the reconstruction with the density matched to the observed density of 0.21 seems to be looking most similar to the real network. The tendency of small nodes (where size is measured in terms of their aggregated row and column sums) to link up with larger nodes and not with other smaller nodes seem to be much better captured than in the reconstructions using the basic Erdős–Rényi model in Figure 4.

4.4.5. Discussion

The empirical fitness model could be generalized by using as fitness $x_i = (\ell_i + a_i)^\beta$ for some $\beta \geq 0$. This allows dampening ($\beta < 1$) or amplifying ($\beta > 1$) the difference in fitnesses. This empirical fitness model can be seen as logit model for the link existence probability p_{ij}. Indeed, as $p_{ij} = 1/[1 + \exp[-\{\ln(\alpha) + \ln(x_i) + \ln(x_j)\}]]$, the only change required is to re-parametrize by taking the log of the fitnesses $x_1, \ldots, x_n$ and the parameter α.

5. Assessing the quality of the reconstruction

We now assess the quality of different reconstruction methods on the empirical data of 89 fully observed networks of CDS exposures. To do so, we compute the row and column sums for each network and assume that we do not observe the individual entries and reconstruct them.

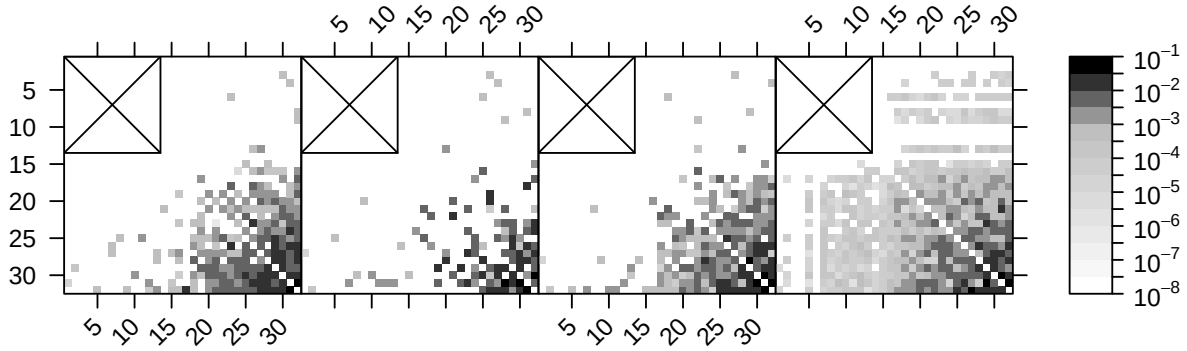

Figure 5: Matrix of example network (left) and reconstructions of the example network using the empirical fitness model calibrated to several densities (from left to right: 0.1, 0.21, 0.5) with no deals between non-dealers.

We use several different measures to quantify the quality of a network reconstruction. Often only univariate measures are used; see, e.g., [4, 13]. Many of these are based only on the corresponding adjacency matrix and not the weighted network. We will look at multivariate measures of both the adjacency matrices and the weighted network matrices as well; in particular we will investigate how well we can reconstruct the in- and out-degree distributions and the distribution of the weights.

In practice, the choice of the criteria will depend on the specific application. One important application is the assessment of systemic risk in financial networks. Since there is no clear understanding yet on how systemic risk depends on certain characteristics of the underlying network, we report a wide range of criteria that might be useful in this context.

We first discuss the reconstruction of the example network in Section 5.3. This gives already good insights, as the network, consisting of 32 counterparties, has $32^2 - 32 = 992$ unknown entries. When looking at multivariate measures such as the in- or out-degree distribution, however, there is only one observation for the in- and out-degree distribution. Therefore we also look at the quality of fit over all 89 networks (Section 5.4).

5.1. Reconstruction methods used in comparison

We consider three different classes of reconstruction methods in our empirical comparison.

First, we consider the calibrated Bayesian approach. To denote various settings, we use the following notation: CBayes(model, density), where density is the target density to which the model has been calibrated (we choose target densities 0.1, 0.21 and 0.5, where 0.21 is the mean density of the observed network) and model will be either the basic Erdős–Rényi model (ER), the empirical fitness model (abbreviated by Fit) or these two models modified to prohibit trades between non-dealers (ERD and FitD). The D stands for "use of the dealer information".

We also use as reconstruction the posterior mean of these Bayesian models, denoted by E{CBayes(model,density)}. It is approximated by the average of the reconstructed matrices produced by the MCMC sampler.

Second, as competitors we use the degree corrected gravity model of [13], calibrated to the same densities as the Bayesian models, with the diagonal being set to 0. We will denote this by StatPhys(density) as the method is inspired by methods from statistical physics. As before, density denotes the density the method is calibrated to. In our study, it represents the class of models that use random graph models to generate matrices that match the row and column constraints in expectation. The fact that this model can also be calibrated to topological properties makes it a strong contender. Indeed, this model has performed very well in a recent study that compares several network reconstruction techniques but did not consider the Bayesian approach in [24] and our extension presented here. Quoting from p. 17 in [2]: "if our focus is on probabilistic methods, we find that... [the model by [13]] is the clear winner across all measures of interest".

Third, we use the reconstruction method that is based on minimizing the Kullback–Leibler divergence suggested in [48], based on the code of [47]. It represents the optimization based reconstruction methods in our comparison and has been chosen since it has frequently been used in practice.

We generated $k = 1000$ samples for each model, except for the Kullback–Leibler method and the posterior mean methods, which only return one fixed reconstruction ($k = 1$). We used suitable settings for the MCMC algorithm and suitable thinning to reduce autocorrelation.

5.2. Criteria used in comparison

Suppose $L^1, \ldots, L^k$ are the reconstructions for the matrix L provided by a given method (with $k = 1$ for the posterior mean reconstructions and the Kullback–Leibler method). We use the following criteria to evaluate such a reconstruction.

We use standard criteria for the adjacency matrices: the accuracy, which is the proportion of edges correctly identified, the sensitivity which is the proportion of truly existing links present in the reconstruction and the specificity which is the proportion of truly missing links missing in the reconstruction. Formally they are given as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{1}{k}\sum_{\nu=1}^{k}\left\{\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left(\mathbf{1}\{L_{ij}=0 \text{ and } L^\nu_{ij}=0\}+\mathbf{1}\{L_{ij}>0 \text{ and } L^\nu_{ij}>0\}\right)\right\},\\
\text{Sensitivity} &= \frac{1}{k}\sum_{\nu=1}^{k}\left\{\frac{1}{\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{L_{ij}>0\}}\sum_{i=1}^{n}\sum_{j\neq i}\left(\mathbf{1}\{L_{ij}>0\}\mathbf{1}\{L^\nu_{ij}>0\}\right)\right\},\\
\text{Specificity} &= \frac{1}{k}\sum_{\nu=1}^{k}\left\{\frac{1}{\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{L_{ij}=0\}}\sum_{i=1}^{n}\sum_{j\neq i}\left(\mathbf{1}\{L_{ij}=0\}\mathbf{1}\{L^\nu_{ij}=0\}\right)\right\}.
\end{aligned}$$

We also use three criteria comparing the weighted network matrices, namely the root mean square error (RMSE), the L_1-error (L1) and a a criterion which we call proportion of truth in the support (PTS). It is designed to capture whether positive entries of the matrix are within the support of the reconstructed matrix. For every matrix element with a truly positive entry L_{ij}, we check if any of the reconstructions L^ν_{ij} are within 10% of the true value of L_{ij}. The three criteria are formally defined as follows:

$$\begin{aligned}
\text{RMSE} &= \frac{1}{k}\sum_{\nu=1}^{k}\sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(L^\nu_{ij}-L_{ij})^2}, \quad \text{L1} = \frac{1}{k}\sum_{\nu=1}^{k}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|L^\nu_{ij}-L_{ij}|,\\
\text{PTS} &= \sum_{i,j}\mathbf{1}\{L_{ij}>0 \text{ and } \exists_{\nu\in\{1,\ldots,k\}}\ 0.9L_{ij}\le L^\nu_{ij}\le 1.1L_{ij}\}\Big/\sum_{i,j}\mathbf{1}\{L_{ij}>0\}.
\end{aligned}$$

Next, we consider criteria for the degree distribution. For this let $d_i = \sum_j \mathbf{1}\{L_{ij} > 0\}$ and $d^\nu_i = \sum_j \mathbf{1}\{L^\nu_{ij} > 0\}$ be the number of positive entries in the rows of observed/reconstructed matrix (the out-degree of the network). We use the bias of the mean out-degree (bias mean deg) and the bias of the maximum out-degree (bias max deg) as criteria:

$$\text{bias mean deg} = \left(\frac{1}{k}\sum_{\nu=1}^{k}\frac{1}{n}\sum_i d^\nu_i\right)-\frac{1}{n}\sum_i d_i, \quad \text{bias max deg} = \frac{1}{k}\sum_{\nu=1}^{k}\max_i d^\nu_i - \max_i d_i.$$

Furthermore, the root mean squared error of a specified quantile of the out-degree distribution (RMSEQD(x)), is formally defined as

$$\text{RMSEQD(x)} = \left[\frac{1}{k}\sum_{\nu=1}^{k}\left\{\text{quant}(d^\nu_1,\ldots,d^\nu_n;x)-\text{quant}(d_1,\ldots,d_n;x)\right\}^2\right]^{1/2}.$$

In the above quant(. . . ;x), computes the empirical x quantile of the first arguments; we use the default implementation in R which is Definition 7 in [31].

Finally, we consider criteria that assess the reconstructions for use in systemic risk measurement through the clearing approach of Eisenberg and Noe [19]. This approach requires a matrix of liabilities. For simplicity we assume that the credit event underlying our data occurred and that the full notional protection amounts in L are now liabilities.

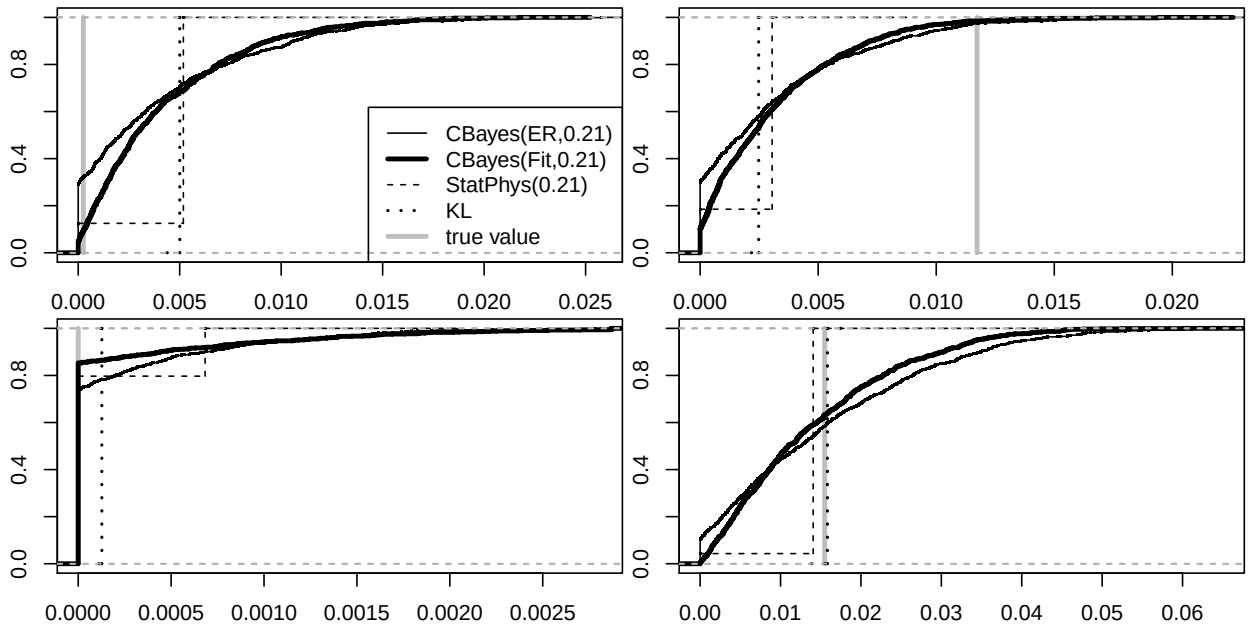

Figure 6: CDF of the reconstruction of 4 different entries of the matrix by various methods.

Furthermore, the approach needs information about capital buffers of the market participants. We do not have such information in our anonymized data set, so we use the following ad hoc choice. We equip all nodes in the financial network with a cash buffer b: for a given node i, we set the buffer to be $b_i = \max(-e_i, 0) \times 1.1$, where $e_i = a_i - \ell_i = \sum_{j=1}^n L_{ji} - \sum_{j=1}^n L_{ij}$ is the initial net worth of node i. With this buffer no defaults occur, hence we need to reduce the buffer of at least one node to trigger a potential contagion process. We choose the node which is the largest net-seller, i.e., we set $i^* = \operatorname{argmin}_{i\in\{0,\dots,n\}} e_i$ and set $b_{i^*} = \max(-e_{i^*}, 0) \times 0.5$.

Then we apply the clearing method of [19] assuming that L plays the role of the liabilities matrix and b the role of the exogenous operating cash flow. This allows us to determine which node defaults. Let $D : [0,\infty)^{n\times n} \to \{0,1\}^n$ be the function that computes for any liabilities matrix the vector of indicator which nodes defaults (1 = default, 0 = no default) using the clearing method of [19] with exogenous operating cash flow b. Applying this to the matrices of a reconstructions allows us to estimate the probability of default for this method, which we then compare to the actual default event as follows:

$$\text{RMSEDP} = \sqrt{\frac{1}{n}\sum_{i=1}^n \{\pi_i^{\text{rec}} - D(L)_i\}^2}, \quad \pi_i^{\text{rec}} = \frac{1}{k}\sum_{\nu=1}^k D(L^\nu)_i, \quad \text{bias DP} = \frac{1}{n}\sum_{i=1}^n \{\pi_i^{\text{rec}} - D(L)_i\}$$

5.3. The example network

We now assess the quality of the reconstruction for the example network shown in Figure 1. Table 2 displays the results for the individual criteria. Before discussing these, we will look at some descriptive plots of the reconstruction.

5.3.1. Individual entries

Figure 6 shows the distribution of individual entries L_{ij} of the reconstructed matrix for several reconstruction method as well as the true value. The specific entries, i.e., the i and the j, were picked to show a spread of possible situations.

The Kullback–Leibler method only gives a fixed value, the degree corrected gravity model (denoted "StatPhys 0.21") returns either 0 or a fixed value in the reconstruction (which is determined by the algorithm such that row and column sums are matched in expectation). In contrast to this, the Bayesian methods (ER-model, empirical fitness model) give a continuous distribution on the positive values.

Table 2: Evaluation of reconstruction of the example network. Best method in bold.

	accuracy	sensitivity	specificity	RMSE [10^-3]	L1 [10^-3]	bias mean deg	bias max deg	PTS	RMSEQD(0.25)	RMSEQD(0.5)	RMSEQD(0.75)	RMSEDP	bias DP
CBayes(ER,0.1)	0.817	0.30	0.96	4.46	1.20	–3.5	–10.5	0.92	**0.0**	2.2	7.0	0.23	–0.10
CBayes(ER,0.21)	0.832	0.58	0.90	3.29	0.93	–0.3	–4.3	0.98	**0.0**	1.8	1.4	0.15	–0.06
CBayes(ERD,0.21)	0.826	0.59	0.89	3.27	0.93	–0.0	-3.9	**0.99**	0.4	2.0	1.1	0.18	–0.04
E[CBayes(ER,0.21)]	0.531	**1.00**	0.41	1.61	0.52	14.5	9.0	0.15	20.2	23.5	15.8	0.18	0.03
CBayes(ER,0.5)	0.694	0.93	0.63	2.92	0.82	8.5	5.5	0.90	6.8	15.2	10.4	0.20	–0.00
CBayes(Fit,0.1)	0.838	0.35	0.97	3.79	1.08	–3.5	–7.8	0.85	**0.0**	3.4	7.4	0.22	–0.08
CBayes(Fit,0.21)	**0.882**	0.71	0.93	2.91	0.84	–0.1	–2.2	0.94	**0.0**	0.8	**0.7**	0.13	–0.05
CBayes(FitD,0.21)	0.880	0.71	0.92	2.91	0.84	–0.0	**–1.5**	0.91	**0.0**	**0.7**	0.7	0.19	–0.04
E[CBayes(Fit,0.21)]	0.677	0.96	0.60	**1.56**	0.51	9.5	5.0	0.11	**0.0**	20.5	12.8	**0.00**	**0.00**
CBayes(Fit,0.5)	0.703	0.98	0.63	2.91	0.82	8.9	7.1	0.90	4.2	15.3	12.5	0.20	0.00
StatPhys(0.1)	0.851	0.38	**0.98**	2.42	0.79	–3.5	–8.4	0.06	0.8	3.3	6.6	0.41	0.12
StatPhys(0.21)	0.869	0.69	0.92	1.79	0.57	**0.0**	–1.9	0.15	0.7	1.3	0.8	0.36	0.12
StatPhys(0.5)	0.701	0.97	0.63	1.66	0.52	8.9	7.2	0.09	5.9	14.5	12.3	0.22	0.03
KL	0.531	**1.00**	0.41	1.57	**0.51**	14.5	9.0	0.13	20.2	23.5	15.8	0.18	0.03

Figure 6 also shows the true value of the corresponding entry. If the true value is 0 (as in the diagram in the bottom left corner), then the true value is in the support of all reconstructions, with the exception of the Kullback–Leibler model. If the true value is positive, then it is usually only within the support of the reconstructed distribution for the Bayesian models (ER, Fitness). Both the Kullback–Leibler and the statistical physics model only produce one possible positive value for a given entry, which then may or may not be close to the truth.

5.3.2. Evaluating the adjacency matrix

We first consider the three criteria that consider the adjacency matrix only. Table 2 shows that the empirical fitness model is doing best in terms of accuracy, particularly if calibrated to the correct density. This is further supported by the numbers for the sensitivity and specificity.

The use of the dealer information does not seem to have an appreciable effect on these criteria. The ER-model has a slightly weaker performance than the empirical fitness model throughout. The statistical physics model is also performing well according to these criteria and is only marginally worse than the empirical fitness model, but better than the ER-model. Its quality of fit is equally sensitive to the calibration used for the density as both the empirical fitness model and the ER-model.

The models calibrated to a high density, the posterior mean reconstructions as well as the Kullback–Leibler reconstruction are not doing particularly well — not surprisingly as they include (almost) all possible links, leading to a high sensitivity, but a poor specificity (the specificity is not 0 for the Kullback–Leibler method and the posterior mean model because some row or column sums being equal to 0 lead to these rows/columns being equal to 0 also for these methods).

5.3.3. Evaluating the weights

Next we look at the classical error criteria for the weighted samples, the RMSE and the L1 error. Both of these have been scaled by a factor of 10^3 in Table 2. In this case the posterior mean reconstructions as well as the Kullback–Leibler method are doing best. This is due to them attempting to give some approximation of an average weight for each possible link, which is good for the RMSE and the L1 error. The individual samples drawn from the posterior distribution have, almost by necessity, a higher error according to these criteria, as the introduction of 0s in the matrix forces more randomness. Generally speaking, the higher the density, the lower the RMSE and the L1 error, with an

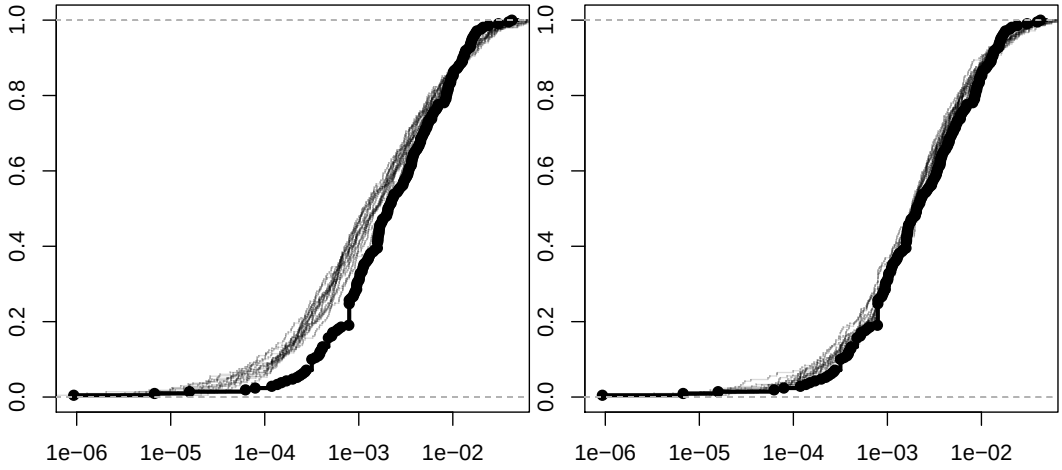

Figure 7: Empirical cdf of the positive weights of the example network underlaid by reconstructions. Left: CBayes(ERD,0.21), right: CBayes(FitnessD,0.21).

indication that the empirical fitness model is doing better than the basic Erdős–Rényi model. The statistical physics model is worse than the posterior mean and Kullback–Leibler models but considerably better than the individual samples in the Bayesian models. This is mainly due to the fact that the weights in the statistical physics model conditional on the existence of a link are deterministic and therefore have a lower variance than the individual Bayesian samples.

The Bayesian models are doing well according to criterion that measures the proportion of truth in the support (PTS). Not surprisingly, the posterior mean reconstructions and the Kullback–Leibler method do not work well as they only give one reconstruction value for each element. Similarly, the statistical physics model does not work well — this is because every element can be reconstructed as two values — either 0 or another fixed value; which then will most likely not be close to the true value.

Figure 7 shows the empirical cdf of the positive weights of the example network, underlaid by reconstruction using the ER-model as well as the empirical fitness model, calibrated to the true density. Both show a good performance, but the empirical fitness model seems to be doing better for the smaller entries of the matrix.

5.3.4. Degree distribution

The bias of the mean degree reflects the calibration and the fact that the posterior mean reconstructions and the Kullback–Leibler model construct very full matrices. Not surprisingly, calibration to the true density does best — after all, the density is the same as the average degree divided by the number of nodes — so calibrating to the true density is the same as calibrating the bias of the mean degree to be 0.

The fitness models calibrated to the true density have the best match to the degree distribution overall. Importantly, they are doing better than the ER-models calibrated to the same density, mainly because they better reproduce the higher parts of the degree distribution, as evidenced by the lower RMSE for these parts of the quantile distribution. The statistical physics model calibrated to the correct density is competitive, with a slightly worse match to the overall degree distribution than the empirical fitness model.

The models calibrated to an incorrect density, and the reconstructions using the posterior mean or the Kullback–Leibler method are not doing well on these criteria.

To reinforce the point that the empirical fitness model captures the degree distribution very well, Figure 8 shows the empirical cumulative distribution function (empirical cdf) of the true degree distribution underlaid by the empirical cdf of 25 reconstructed networks using the ER-model as well as the empirical fitness model, both calibrated to the true density of 0.21. Both seem to be good, but the empirical fitness model produces samples that match more closely the highest degrees observed in the real network.

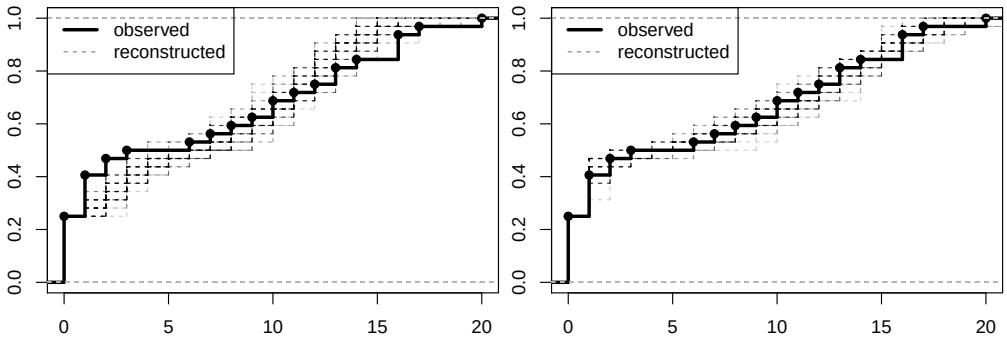

Figure 8: Row degree distribution of the example network underlaid by the reconstructions CBayes(ERD,0.21) (left) and CBayes(FitnessD,0.21) (right).

5.3.5. Systemic risk

When looking at the two criteria that assess how well the reconstruction methods can reproduce the default probabilities associated with the network (RMSEDP and biasDP), we see that all Bayesian methods considered here perform significantly better than the statistical physics model. The Bayesian models that are calibrated to the true density also significantly outperform the Kullback–Leibler method. Using the empirical fitness model as a priori assumption in the Bayesian framework offers slight advantages in most cases over using the Erdős–Rényi model as a prior model. The posterior mean model in the empirical fitness model calibrated to the true density even results in a perfect fit for the default probabilities for this particular network.

The Kullback–Leibler method performs significantly better than the statistical physics model but is worse than most of the Bayesian methods.

The interesting feature of the systemic risk criteria is that edges are not considered individually but default probabilities depend on the weights of several edges at the same time. In that sense the systemic risk criteria are higher order criteria and show a strong dominance of the Bayesian approach.

The Kullback–Leibler method substantially outperforms the statistical physics model which indicates that even though the statistical physics model clearly captures the degree distribution much better than the Kullback–Leibler method it pays a heavy price for this feature because the corresponding weights assigned to the edges do not seem to reproduce the observed weights well.

When looking at the bias in the default probabilities (biasDP), we see that the statistical physics model seems to overestimate the default probabilities quite substantially. The bias for both the Kullback–Leibler method and the Bayesian methods seems rather small.

5.4. Across networks

We now consider all 89 networks. We assume that we only know the row and column sums and reconstruct the networks from these partial information. We do not use any additional information on dealer — non-dealer in this subsection. We consider models calibrated to the true density of the network to be reconstructed as well as models calibrated to a density of 15.4%, which is the average of the densities of the 89 networks. Figure 9 reports the results for all networks for three selected measures (accuracy, proportion of truth in the support and the root mean square error). Table 3 reports the mean value of the measures discussed in Section 5.3 over the 89 networks. Overall, the relative performance of the methods stays similar to the performance in the example network.

When looking at performance measures linked to the adjacency matrix, such as accuracy, sensitivity and specificity, and measures related to the degree distribution the Bayesian empirical fitness model performs best (in particular when calibrated to the correct density), followed closely by the statistical physics model and then the Bayesian ER-model. The left plot of Figure 9 shows that calibrating the model improves performance.

For criteria that consider average error measures from the true weighted matrix (RMSE, L1), the posterior mean models (both for the fitness and the ER-model) the Kullback–Leibler and the statistical physics models outperform the individual samples. The middle figure of Figure 9 shows that for the Kullback–Leibler and the statistical physics

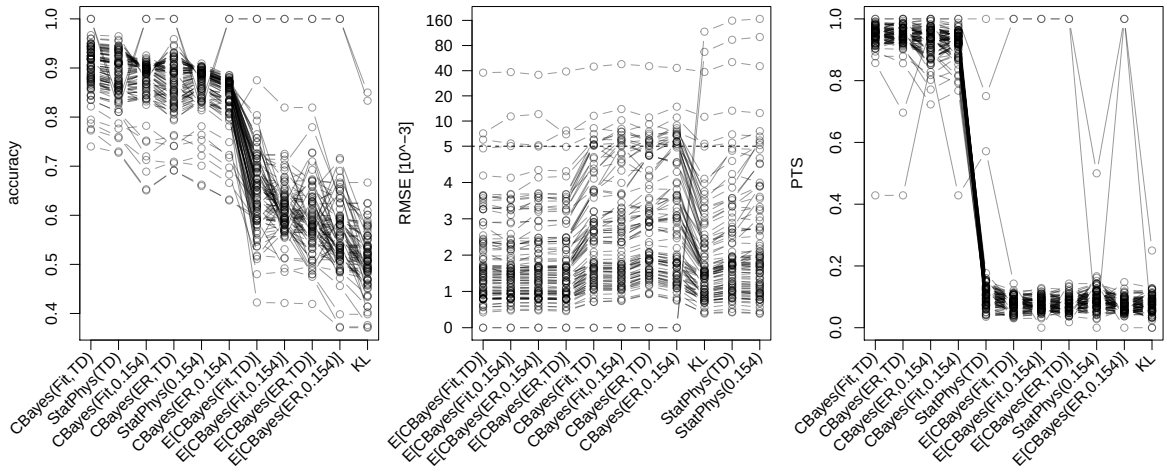

Figure 9: Evaluation of reconstruction of the 89 networks — accuracy, RMSE and proportion of truth in support. The plot of the RMSE has a split scale on the y-axis (linear up to 5, log from 5 onwards). Methods ordered by mean performance.

models, the reconstruction of two networks is particularly poor (whereas it is perfect for the other models) — these are small networks with only few positive entries.

When looking at the criterion that measures whether there exist samples that are close to the true values (PTS), see also the right panel of Figure 9, we have two clear winners: the Bayesian empirical fitness and the Bayesian ER-model which reach scores of around 95%. We would expect that these scores can even be higher, when we increase the sample size. The posterior mean model due to the averaging property performs rather poorly according to this criteria. Both the Kullback–Leibler method and the statistical physics model also perform poorly here, since the weights they assign to the matrix are purely deterministic and there is no guarantee that this weight is in any way close to the true weight.

The criteria looking at measures of systemic risk in the network are in line with the results obtained for the single network. The Bayesian approach performs best according to these criteria (with again the empirical fitness model performing slightly better than the Erdős–Rényi model as a priori modeling assumptions), the Kullback–Leibler method comes second and the statistical physics model comes last. The statistical physics model significantly overestimates default probabilities. The posterior mean models in the Bayesian approach slightly overestimate the default probabilities, but they still have a smaller positive bias than the Kullback–Leibler method. The Bayesian methods using the individual samples (in contrast to the posterior mean) has only a very small (negative) bias, indicating that default probabilities can be slightly underestimated.

Figure 10 evaluates the quality of the calibration for the CBayes(ER,TD) method, i.e., using the true density for calibration, across all 89 networks. The calibration seems to work very well for most networks — with small discrepancies for some networks. This is due to a limited computational effort used in the calibration.

Importantly, Figure 10 shows that the density of the individual reconstructions stays close to the calibrated reconstructions — this shows that without calibration it is impossible to get a good reconstruction with the ER model. A similar picture emerges when we look at reconstructions from the empirical fitness model (not reported here). Also, for the statistical physics model, the densities are very close to the calibrated for medium to large networks — only for small networks do they scatter more widely.

6. Discussion and conclusion

The (empirical) Bayesian approach performed well on empirical data and outperformed all competitors that we considered. We now show that the Bayesian approach is also very appealing from a theoretical point. For this, we discuss desirable properties for a network reconstruction technique and whether they are satisfied by the methods studied in the present paper. Table 4 gives an overview.

Table 3: Evaluation of reconstruction of the networks — mean performance over 89 networks with standard error in brackets

method	accuracy	sensitivity	specificity	RMSE [10^-3]	L1 [10^-3]	bias mean deg	bias max deg	PTS	RMSEQD(0.25)	RMSEQD(0.5)	RMSEQD(0.75)	RMSEDP	bias DP
CBayes(ER,TD)	0.863	0.49	0.92	3.50	1.02	-0.4	-6.6	0.94	0.47	1.48	1.41	0.212	-0.013
	(0.006)	*(0.01)*	*(0.01)*	*(0.52)*	*(0.18)*	*(0.0)*	*(0.6)*	*(0.01)*	*(0.09)*	*(0.09)*	*(0.07)*	*(0.01)*	*(0.01)*
CBayes(ER,0.154)	0.840	0.53	0.91	3.70	1.13	1.0	-3.5	0.93	0.70	4.13	5.54	0.213	-0.008
	(0.006)	*(0.02)*	*(0.00)*	*(0.52)*	*(0.21)*	*(0.4)*	*(0.4)*	*(0.01)*	*(0.13)*	*(0.33)*	*(0.55)*	*(0.01)*	*(0.01)*
E[CBayes(ER,TD)]	0.599	0.97	0.52	2.28	0.75	17.3	12.6	0.10	11.60	26.78	23.24	0.252	0.025
	(0.010)	*(0.00)*	*(0.01)*	*(0.44)*	*(0.17)*	*(0.8)*	*(0.8)*	*(0.01)*	*(1.35)*	*(1.56)*	*(1.60)*	*(0.01)*	*(0.01)*
E[CBayes(ER,0.154)]	0.566	0.98	0.49	2.28	0.79	19.7	13.9	0.09	13.97	30.97	24.92	0.258	0.027
	(0.010)	*(0.00)*	*(0.01)*	*(0.41)*	*(0.20)*	*(1.2)*	*(1.0)*	*(0.01)*	*(1.65)*	*(1.92)*	*(1.85)*	*(0.01)*	*(0.01)*
CBayes(Fit,TD)	**0.899**	0.64	**0.94**	3.15	0.93	-0.0	-2.0	**0.94**	**0.12**	**0.59**	**1.15**	**0.205**	-0.009
	(0.005)	*(0.01)*	*(0.00)*	*(0.51)*	*(0.17)*	*(0.0)*	*(0.4)*	*(0.01)*	*(0.04)*	*(0.05)*	*(0.06)*	*(0.01)*	*(0.01)*
CBayes(Fit,0.154)	0.873	0.66	0.93	3.41	1.04	1.2	**1.7**	0.92	0.22	2.58	5.69	0.209	**-0.006**
	(0.006)	*(0.02)*	*(0.00)*	*(0.56)*	*(0.20)*	*(0.4)*	*(0.8)*	*(0.01)*	*(0.07)*	*(0.23)*	*(0.48)*	*(0.01)*	*(0.01)*
E[CBayes(Fit,TD)]	0.655	0.97	0.59	**2.22**	**0.73**	14.2	11.7	0.10	9.44	20.13	21.79	0.253	0.024
	(0.010)	*(0.00)*	*(0.01)*	*(0.43)*	*(0.16)*	*(0.6)*	*(0.8)*	*(0.01)*	*(1.00)*	*(0.90)*	*(1.44)*	*(0.01)*	*(0.01)*
E[CBayes(Fit,0.154)]	0.619	0.98	0.55	2.25	0.77	17.0	12.4	0.10	10.39	26.76	23.86	0.251	0.026
	(0.008)	*(0.00)*	*(0.01)*	*(0.44)*	*(0.19)*	*(1.0)*	*(0.9)*	*(0.01)*	*(1.18)*	*(1.57)*	*(1.81)*	*(0.01)*	*(0.01)*
StatPhys(TD)	0.882	0.59	0.93	5.47	1.93	**-0.0**	-1.9	0.12	0.58	0.97	1.25	0.382	0.212
	(0.005)	*(0.01)*	*(0.00)*	*(2.10)*	*(0.79)*	*(0.0)*	*(0.3)*	*(0.01)*	*(0.05)*	*(0.05)*	*(0.05)*	*(0.00)*	*(0.01)*
StatPhys(0.154)	0.861	0.62	0.92	5.71	2.10	1.2	1.9	0.10	0.68	3.00	5.22	0.376	0.204
	(0.005)	*(0.02)*	*(0.00)*	*(2.20)*	*(0.86)*	*(0.4)*	*(0.8)*	*(0.01)*	*(0.09)*	*(0.23)*	*(0.44)*	*(0.01)*	*(0.01)*
KL	0.517	**1.00**	0.42	4.30	1.81	21.3	14.6	0.08	18.66	32.14	26.03	0.255	0.043
	(0.008)	*(0.00)*	*(0.01)*	*(1.54)*	*(0.76)*	*(1.1)*	*(0.9)*	*(0.00)*	*(1.68)*	*(1.87)*	*(1.80)*	*(0.01)*	*(0.01)*

TD = true density of the network to be reconstructed; best method in bold.

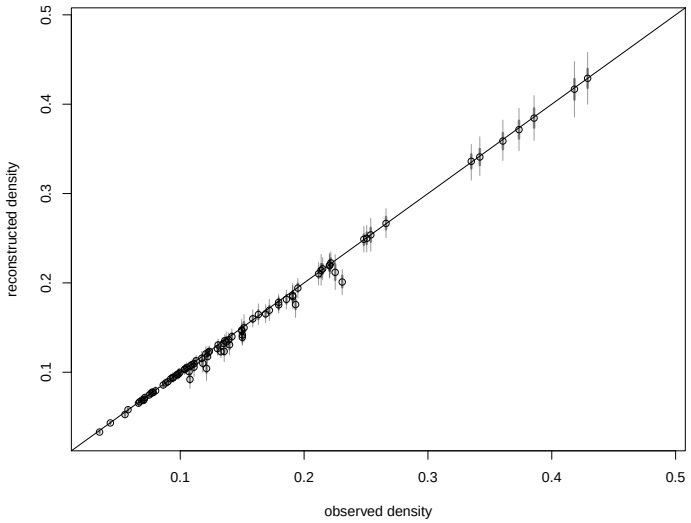

Figure 10: Evaluation of density calibration for the CBayes(ER,TD) model. True density versus mean density of the reconstructed network. Vertical lines: spread of reconstruction in the posterior [thin: 5% and 95% quantile of the posterior density, thick: lower and upper quartile]

Table 4: Summary of properties of different reconstruction methods.

Properties	Kullback–Leibler	Statistical Physics	Bayesian
Creates sample of matrices	no	yes	yes
Range of samples can in principle include true matrix	no	no	yes
Matches row and column sums exactly	yes	no	yes
Allows for different types of degree distributions	no	depends	yes
Allows for different types of weight distributions	no	no	yes
Degree and weight distributions can be modeled independently	no	no	yes
Can be calibrated with small number of parameters	no	yes	yes

All methods, that only return one reconstructed network do not allow for an assessment of estimation error. It seems problematic to just pick one network out of the usually infinite number of networks consistent with the observations and draw all conclusions just from this network. In classical statistics this would correspond to using only a point estimator and ignoring concepts such as confidence intervals, etc. Reconstruction methods that return an ensemble of matrices at least can give some indication of measurement errors; how meaningful they are, however, depends very much on the sampling methodology.

When looking at methods that return an ensemble of network matrices, the obvious question is whether the sampling methodology is designed such that the range of the samples can in principle contain the true (observed) network. Often this is not the case. Statistical physics type models generate samples to match the observed row and column sums in expectation, since every single sample will usually not satisfy these. As we have discussed, they only allow for one deterministic entry per matrix cell and 0 and therefore cannot generate samples whose range can in principle contain the true matrix. The Bayesian approach, however, which is based on a full stochastic model for both the existence of edges and their weights, can in principle generate a range that contains the true matrix.

Usually methods based on the Kullback–Leibler approach or other optimization approaches that return only one reconstruction for the network satisfy the row and column sums. The Bayesian approach also returns samples that all satisfy the row and column constraints. The individual samples in the statistical physics approach do not satisfy the row and column constraints.

The Bayesian approach is the only approach which allows for the specification of a wide range of degree and weight distributions; moreover, they can be specified independently. In the statistical physics models, these two choices are interlinked since they need to be chosen such that in expectation the row and column sums are matched. The Kullback–Leibler method can be modified by choosing different input matrices in the optimization procedure. It still remains unclear what the resulting degree and weight distributions are.

The statistical physics and the Bayesian approach can both be calibrated well, based on a small number of parameters to account for additional information such as the density of the network. The Kullback–Leibler method cannot easily be calibrated with a small number of parameters.

From a practical point of view, the statistical physics model is the easiest to implement from scratch, followed by the Kullback–Leibler model and then the Bayesian approaches. However, software for the latter two approaches is available, making them easy to use in practice.

The (empirical) Bayesian approach is the most flexible and has more desirable properties than the alternative methods considered in this paper.

Acknowledgments

Parts of this research were carried out while Luitgard Veraart was a George Fellow at the Bank of England whose hospitality and funding by the Houblon-Norman Fund are gratefully acknowledged. Both authors would like to thank the Bank of England for providing the data for this research project. In particular, we would like to thank Nicholas Vause (Bank of England) for regular discussions and feedback. We also would like to thank the Editor-in-Chief (C. Genest), the Guest Editor-in-Chief (J.G. Nešlehová), the Guest Associate Editor, and the anonymous referees for their thoughtful and helpful comments and suggestions.

References

[1] K. Anand, B. Craig, G. Von Peter, Filling in the blanks: Network structure and interbank contagion, Quant. Finance 15 (2015) 625–636.

[2] K. Anand, I. van Lelyveld, A. Banai, S. Friedrich, R. Garratt, G. Hałaj, J. Fique, I. Hansen, S.M. Jaramillo, H. Lee, J.L. Molina-Borboa, S. Nobili, S. Rajan, D. Salakhova, T.C. Silva, L. Silvestri, S.R.S. de Souza, The missing links: A global study on uncovering financial network structures from partial data, J. Financial Stability 35 (2018) 107–119.

[3] Bank for International Settlements, Recent Advances in Modelling Systemic Risk Using Network Analysis, 2010. Available at `https://www.ecb.europa.eu/pub/pdf/other/modellingsystemicrisk012010en.pdf`.

[4] Bank for International Settlements, Making Supervisory Stress Tests more Macroprudential: Considering Liquidity and Solvency Interactions and Systemic Risk, 2015. BCBS Working Papers No 29.

[5] Bank of England, The Bank of England's Approach to Stress Testing the UK Banking System, 2015. Available at `http://www.bankofengland.co.uk/financialstability/Documents/stresstesting/2015/approach.pdf`.

[6] P. Baral, J. P. Fique, Estimation of bilateral exposures — A copula approach, 2012. Available at: `http://www.cirano.qc.ca/conferences/public/pdf/networks2012/02-BARAL-FIQUE-Estimation_of_Bilateral_Exposures-A_Copula_Approach.pdf`.

[7] M. Bardoscia, S. Battiston, F. Caccioli, G. Caldarelli, Debtrank: A microscopic foundation for shock propagation, PloS one 10 (2015) e0130406.

[8] A. Barvinok, Matrices with prescribed row and column sums, Lin. Algebra Appl. 436 (2012) 820–844.

[9] S. Battiston, G. Caldarelli, M. D'Errico, S. Gurciullo, Leveraging the network: A stress-test framework based on DebtRank, Available at arXiv:1503.00621 (2016).

[10] B. Bollobás, Modern Graph Theory, Springer-Verlag, New York, 2013.

[11] G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz, Scale-free networks from varying vertex intrinsic fitness, Phys, Rev. Lett. 89 (2002) 258702.

[12] S. Chatterjee, P. Diaconis, A. Sly, Random graphs with a given degree sequence, Ann. Appl. Probab. (2011) 1400–1435.

[13] G. Cimini, T. Squartini, D. Garlaschelli, A. Gabrielli, Systemic risk analysis on reconstructed economic and financial networks, Scientific Reports 5 (2015).

[14] B. Craig, G. Von Peter, Interbank tiering and money center banks, J. Financial Intermediation 23 (2014) 322–347.

[15] G. De Masi, G. Iori, G. Caldarelli, Fitness model for the Italian interbank money market, Phys. Rev. E 74 (2006) 066112.

[16] H. Degryse, G. Nguyen, Interbank exposures: An empirical examination of contagion risk in the Belgian banking system, Internat. J. Central Banking 3 (2007) 123–171.

[17] M. Drehmann, N. Tarashev, Measuring the systemic importance of interconnected banks, J. Financial Intermediation 22 (2013) 586–607.

[18] DTCC, Trade Information Warehouse explanation of TIW reports, 2015. Available from `https://www.dtccdata.com/~/media/data-products-portal/files/resources/tiw/tiw-market-report-guide.pdf`.

[19] L. Eisenberg, T. H. Noe, Systemic risk in financial systems, Manag. Sci. 47 (2001) 236–249.

[20] H. Elsinger, A. Lehar, M. Summer, Using market information for banking system risk assessment, Internat. J. Central Banking 2 (2006).

[21] P. Erdős, A. Rényi, On random graphs I, Publ. Math-Debrecen 6 (1959) 290–297.

[22] S.E. Fienberg, A brief history of statistical models for network analysis and open challenges, J. Comput. Graph. Statist. 21 (2012) 825–839.

[23] Financial Stability Board, International Monetary Fund, The Financial Crisis and Information Gaps, Sixth Progress Report on the Implementation of the G-20 Data Gaps Initiative, 2015. Available at http://www.imf.org/external/np/g20/pdf/2015/6thprogressrep.pdf.

[24] A. Gandy, L.A.M. Veraart, A Bayesian methodology for systemic risk assessment in financial networks, Manag. Sci. 63 (2017) 4428–4446.

[25] G. Hałaj, C. Kok, Assessing interbank contagion using simulated networks, ECB Working Paper Series, No 1506, 2013.

[26] A.G. Haldane, Rethinking the financial network, 2009. Speech available at `http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2009/speech386.pdf`.

[27] A.G. Haldane, On microscopes and telescopes, 2015. Speech available at `http://www.bankofengland.co.uk/publications/Pages/speeches/2015/812.aspx`.

[28] A.G. Haldane, R.M. May, Systemic risk in banking ecosystems, Nature 469 (2011) 351–355.

[29] M.L. Hazelton, Statistical inference for transit system origin-destination matrices, Technometrics (2010) 221–230.

[30] P. W. Holland, S. Leinhardt, An exponential family of probability distributions for directed graphs, J. Amer. Statist. Assoc. 76 (1981) 33–50.

[31] R.J. Hyndman, Y. Fan, Sample quantiles in statistical packages, Amer. Statist. 50 (1996) 361–365.

[32] E.D. Kolaczyk, Statistical Analysis of Network Data, Methods and Models, Springer, New York, 2009.

[33] S. Langfield, Z. Liu, T. Ota, Mapping the UK interbank system, J. Banking Finance 45 (2014) 288–303.

[34] L. Lovász, Large Networks and Graph Limits, Amer. Math. Soc., Providence, RI, 2012.

[35] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees, New J. Phys. 16 (2014) 043022.

[36] I. Mastromatteo, E. Zarinelli, M. Marsili, Reconstruction of financial networks for robust estimation of systemic risk, J. Statist. Mech. Theory Exper. 2012 (2012) P03011.

[37] J.W. Miller, M.T. Harrison, Exact sampling and counting for fixed-margin matrices, Ann. Statist. 41 (2013) 1569–1592.

[38] P.E. Mistrulli, Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns, J. Banking Finance 35 (2011) 1114–1127.

[39] A. Moussa, Contagion and Systemic Risk in Financial, Ph.D. thesis, Columbia University, New York, 2011.

[40] J. Park, M. E. Newman, Statistical mechanics of networks, Phys. Rev. E 70 (2004) 066117.

[41] M.H. Schneider, S.A. Zenios, A comparative study of algorithms for matrix balancing, Oper. Res. 38 (1990) 439–455.

[42] V.D.P. Servedio, G. Caldarelli, P. Buttà, Vertex intrinsic fitness: How to produce arbitrary scale-free networks, Phys. Rev. E 70 (2004) 056126.

[43] C.R. Shalizi, A. Rinaldo, Consistency under sampling of exponential random graph models, Ann. Statist. 41 (2013) 508–535.

[44] T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade, I: A binary network analysis, Phys. Rev. E 84 (2011) 046117.

[45] T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade, II: A weighted network analysis, Phys. Rev. E 84 (2011) 046118.
[46] T. Squartini, I. van Lelyveld, D. Garlaschelli, Early-warning signals of topological collapse in interbank networks, Scientific Reports 3 (2013).
[47] U. Temurshoev, R.E. Miller, M.C. Bouwmeester, A note on the GRAS method, Econom. Systems Res. 25 (2013) 361–367.
[48] C. Upper, A. Worms, Estimating bilateral exposures in the German interbank market: Is there a danger of contagion?, Europ. Econom. Rev. 48 (2004) 827–849.
[49] S.J. Wells, Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion, Bank of England Working Paper, 2004.