RECONSTRUCTION OF FINANCIAL NETWORKS FOR ROBUST ESTIMATION OF SYSTEMIC RISK

IACOPO MASTROMATTEO, ELIA ZARINELLI, AND MATTEO MARSILI

Abstract. In this paper we estimate the propagation of liquidity shocks through interbank markets when the information about the underlying credit network is incomplete. We show that techniques such as Maximum Entropy currently used to reconstruct credit networks severely underestimate the risk of contagion by assuming a trivial (fully connected) topology, a type of network structure which can be very different from the one empirically observed. We propose an efficient message-passing algorithm to explore the space of possible network structures, and show that a correct estimation of the network degree of connectedness leads to more reliable estimations for systemic risk. Such algorithm is also able to produce maximally fragile structures, providing a practical upper bound for the risk of contagion when the actual network structure is unknown. We test our algorithm on ensembles of synthetic data encoding some features of real financial networks (sparsity and heterogeneity), finding that more accurate estimations of risk can be achieved. Finally we find that this algorithm can be used to control the amount of information regulators need to require from banks in order to sufficiently constrain the reconstruction of financial networks.

1. Introduction

The estimation of the robustness of a financial network to shocks and crashes is a topic of central importance to assess the stability of an economic system. Recent dramatic events evidenced the fragility of many economies, supporting the claim that "the worlds financial system can collapse like a row of dominoes" [1]. As a result, governments and international organizations became increasingly concerned about systemic risk. The banking system is thought to be a fundamental channel in the propagation of shocks to the entire economy: the economic distress of an insolvent bank can be transmitted to its creditors by interbank linkages, thus a shock can easily propagate to the whole network. Unfortunately detailed data on banks bilateral exposures is not always available, and institutions are often left with the problem of assessing the resilience of a system to financial shocks by exploiting an incomplete information set. In this framework the reconstruction of bilateral exposures becomes a central issue for the estimation of risk, and requires the application of sophisticated inference schemes to obtain reliable estimations. Among several methods, a commonly used tool for this task is the so called *entropy maximization method* [2, 3, 4, 5]. The main limitation of this procedure is that it assumes a market structure which can

The authors wish to thank F. Caccioli, F. Krzakala, Y. Sun, L. Zdeborova for very useful discussions. I.M. acknowledges support from GDRE 224 GREFI-MEFI CNRS-INdAM.

be quite different from the actual one: it tends to spread the debt as evenly as possible, without assuming any heterogeneity in the structure for the network [6]. Unfortunately these assumptions lead to an undervaluation of the extent of contagion, as the measure of the vulnerability to financial contagion depends crucially on the pattern of interbank linkages. Stress-tests used to quantitatively analyze this dependence confirm this results both for simulated and real data, as shown in figures 2, 3 and in reference [6].
In this paper we will introduce a message-passing algorithm to overcome this limitation, and to sample efficiently the space of possible structures for the network. This method can be used to propose plausible candidates for the real network structure, and to produce worst case scenarios for the spread of financial contagion. We remark that despite the huge number of possible network structures ($\sim 2^{N^2}$) we are able to sample configuration from this space in a time which scales quadratically in the number of unknown entries of the liability matrix.

In section 2 we introduce the main concepts and define the problem of network reconstruction, while in 3 we present the Maximum Entropy (ME) algorithm, a commonly used procedure to infer credit networks from incomplete datasets. In section 4 we show the idea which allows our algorithm to explore the space of network structures and extend the validity of ME. Section 5 describes the stress-test which we employ to analyze the robustness of financial networks, and in section 6 we apply all these ideas to synthetic datasets. In section 7 we discuss the reliability of the reconstruction algorithm as a function of the policy adopted by regulatory institutions. Finally in section 8 we draw our last conclusions.

2. Framework

Let us consider an ensemble $\mathcal{B} = \{b_0, \ldots, b_{N-1}\}$ of N banks, in which each bank in $\mathcal{B}$ may borrow to or lend money from other banks in $\mathcal{B}$. This structure is encoded in the so-called liability matrix L, an $N \times N$ non-symmetric matrix describing the instantaneous state of a credit network. Each element L_{ij} denotes the funds that bank $j \in \mathcal{B}$ borrowed from bank $i \in \mathcal{B}$ (regardless of the maturity of the debt). We fix the convention that $L_{ij} \geq 0$ $\forall (i,j) \in \mathcal{B} \times \mathcal{B}$, $L_{ii} = 0 \; \forall i \in \mathcal{B}$. With this definition, the expression $L_i^{\rightarrow} = \sum_j L_{ij}$ represents the total credit which the institution i possesses against the system, while $L_j^{\leftarrow} = \sum_i L_{ij}$ represents the total debt owed by the institution j to the environment.[1] This matrix contains information about the instantaneous state of a credit network, and it is sufficient to estimate the risk of contagion in many cases of practical relevance. Indeed one is often unable to obtain from empirical data the complete expression for the matrix L. Data are typically extracted by a bank balance sheets or by institutional databases [7], and partial informations have to be coherently integrated into a list of plausible liability matrices. In the following discussion, we will suppose that three different types of informations about L are available, as typically reported in the literature [8]:

[1]Without loss of generality we consider a closed economy ($\sum_i L_i^{\rightarrow} = \sum_j L_j^{\leftarrow}$), by using bank b_0 as a placeholder to take into account flows of money external to the system.

(1) All the debts larger than a certain threshold θ are known. This allows us to rescale all the elements of L by θ, so that we consider without loss of generality liability matrices for which all the unknown elements are bound to be in the interval [0,1]. We assume to have at most order N elements exceeding such threshold.
(2) We assume a certain set of entries (which we take to be of order N) to be known. This corresponds to banks or bank sectors for which some particular position needs to be disclosed by law.
(3) The total credit $L_i^{\rightarrow}$ and the total debit $L_j^{\leftarrow}$ of each bank are known. Acceptable candidates for liability matrices need to satisfy a set of $2N$ linear constraints, whose rank is in general $\mathcal{R} \leq 2N - 1$.

We remark that we have defined a set of constraints of order N elements, which is too small to single out a unique candidate for the true unknown liability matrix. The possible solutions compatible with the observations define a space Λ, whose members we denote with $\hat{L}$. Let U be the set of not directly known (i.e. non-fixed by to constraints of type (1) and (2)) entries of the liabilities matrix. Then those entries of the liability matrix (whose number is $M = |U|$) are real numbers subject to domain constraints (they must be in $[0, 1]$) and linear algebraic constraints (the sum on the rows and on the columns must be respected). The ratio $M/\mathcal{R} \geq 1$ controls the degree of underdetermination of the network, and is typically much larger than one.

3. Dense reconstruction

A possible procedure to study the robustness of a financial network when the complete information about the liability matrix is not uniquely specified, is to pick from the set of candidate matrices Λ a representative matrix, and to test the stability uniquely for the network specified by such $\hat{L}$. In this case a criteria has to be chosen to select a particular matrix out of the Λ space, by doing some assumptions about the structure of the true L_{ij}. A choice which is commonly adopted [2, 3, 4, 5] is based on the maximum entropy criteria, which assumes that banks spread their lending as evenly as possible. The problem becomes in this case to finding a vector $\vec{L} = \{L_\alpha\}_{\alpha \in U}$ (the unknown entries of the liability matrix) whose entries respect the algebraic and domain constraints and minimize the distance with the uniform vector $\vec{Q} = \{Q_\alpha\}_{\alpha \in \mathcal{U}}$ (such that $\forall \alpha \quad Q_\alpha = 1$), where the distance is quantified by the Kullback-Leibler divergence:

$$D_{KL}(\vec{L}, \vec{Q}) = \sum_\alpha L_\alpha \log \frac{L_\alpha}{Q_\alpha}$$

The minimization of such function is a standard convex optimization problem, that can be solved efficiently in polynomial time. In financial literature this algorithm is known with the name of Maximal Entropy (ME) reconstruction. We remark that by using this algorithm no entry is exactly put to zero unless it is forced by the algebraic constraints.[2]

4. Sparse reconstruction

ME might not be a particularly good description of reality since the number of counterparties of a bank is expected to be limited and much smaller than N, while ME tends to produce completely connected structures. In the case of real networks the degree of market concentration can be higher than suggested by ME. This systematically leads to an underestimation of risk, as a structure in which the debt is distributed homogeneously among the nodes is generally known to be able to absorb shocks more effectively than a system in which few nodes dominate the network [6]. In order to be closer to reality and to estimate more accurately the risk contagion it is than necessary to reconstruct liability matrices whose degree of sparsity can be tuned, and eventually taken to be as big as possible. We present in this section an algorithm which, given the fraction λ of entries which are expected to be exactly zero, is able to reconstruct a sample of network structures compatible with this requirement, and to find a λ_{max} which bounds the maximum possible sparsity. The issues that we want to solve are: (i) is it possible to fix a fraction λ of the unknown entires to zero without violating the domain and the algebraic constraints? (ii) For a given fraction λ, how many possible reconstructed liability matrices do exist? The algorithm solves this problems by sampling, for each given λ, the space of all possible supports for the reconstructed liability matrix such that the constraints are not violated, and by evaluating the volume of such support space. As one can easily expect, there will be a range of $[\lambda_{min}, \lambda_{max}]$ of fractions of fixed zeros compatible with the constraints: trivially $\lambda_{min} = 0$ corresponds to the dense network, which always admits a compatible solution, but we are able to find a non-trivial λ_{max} which corresponds to the maximally sparse network of banks. A plot of the logarithm of the number of possible supports as a function of λ is given in figure 1 ($\times$ signs) for a network as the ones described in section 6. Once a support is given, the liability matrix elements can easily be reconstructed via ME.

We briefly sketch here the idea behind the algorithm, relegating to the appendix the more technical parts. Suppose that a liability matrix with unknown entires is given, together with of the vectors of total credit ($L_i^{\rightarrow}$) and the one of total liabilities ($L_i^{\leftarrow}$). Then without loss of generality one can assume the known entries to be equal to zero, as the values of the known entries always can be absorbed into a rescaled value of the $L_i^{\rightarrow}$ and $L_i^{\leftarrow}$, and restrict the problem just to the unknown entries of the matrix. Under this assumption we can define the network $\mathcal{G}$ which is the support of the unknown entries of the liability matrix. Each node of $\mathcal{G}$ is a bank and the directed edges are the elements of U. For each node i of

[2]This algorithm is not the only possible choice to extract a representative matrix out from the set Λ. Indeed existing algorithms share with the ME the property of returning solutions located in the *interior* of Λ. On the other hand, when choosing a point at random in a compact set in very high dimension d, it is very likely that the point will be very close to the boundary (i.e. at a distance of order $1/d$). Hence, it is reasonable to expect that typical feasible liability matrices are located on or close the boundaries of Λ.

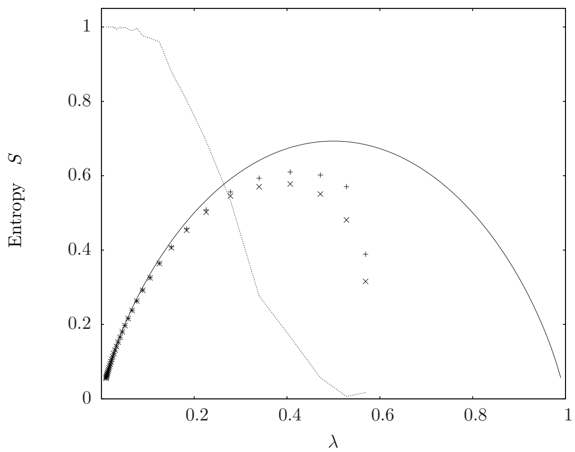

FIGURE 1. Entropy S of the space of compatible configurations $a_{i,j}$ at fixed magnetization $\hat{\lambda}$ with the energy $\mathcal{H}(a_{i,j})$ (+ sign) and true energy $\mathcal{H}_0(a_{i,j})$ ($\times$ sign) for the examples discussed in the text. S is defined as the logarithm of the number of configurations $\{a_{i,j}\}$ with $\mathcal{H} = 0$ (or $\mathcal{H}_0$), divided by the number M of possibly non-zero entries $a_{i,j}$. The solid line plotted for comparison is the entropy of a system of independent links $a_{i,j}$ with the same density (i.e. number of non-zero links). The probability for a solution of $\mathcal{H}_0(a_{i,j})$ to be also a solution of $\mathcal{H}(a_{i,j})$ is also plotted on the same graph (dashed line).

$\mathcal{G}$ the sum of the incoming entries $L_i^{\rightarrow} = \sum_j L_{ij}$ and of the outgoing entries $L_j^{\leftarrow} = \sum_i L_{ij}$ is known. Let $k_i^{\leftarrow}$ ($k_i^{\rightarrow}$) be the number of incoming (outgoing) links in the subset of edges where $L_{i,j} > 0$. Since $L_{i,j} \leq 1$, the number $k_i^{\leftarrow}$ ($k_i^{\rightarrow}$) of incoming (outgoing) links is at least the integer part of $L_i^{\leftarrow}$ ($L_i^{\leftarrow}$) plus one. Therefore, one can define a cost function[3]

$$\mathcal{H}\{a_{i,j}\} = \sum_i \left[\theta\left(L_i^{\rightarrow} - k_i^{\rightarrow}\right) + \theta\left(L_i^{\leftarrow} - k_i^{\leftarrow}\right)\right] \tag{1}$$

over the dynamical variables $a_{i,j} = 0, 1$ which identify the subset of edges, with

$$k_i^{\rightarrow} = \sum_j a_{i,j}, \qquad k_i^{\leftarrow} = \sum_j a_{j,i}.$$

All sub-graphs $a_{i,j}$ with $\mathcal{H} = 0$ are feasible candidates for the support of solutions $L_{i,j} > 0$ to the problem. In general, the constraints are $2N$ linear equations and, as long as the number on non-zero elements $L_{i,j}$ is larger than $2N$ solutions exist, but it is not granted that they have $L_{i,j} \in [0, 1]$ for al i, j. In other words, all the compatible solutions have to satisfy the constraint $\mathcal{H} = 0$, but the converse is not true (as shown in figure 1), because some support $a_{i,j}$ may not admit a solution with $L_{i,j} \in [0, 1]$ for al i, j. We distinguish

[3]Here $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ otherwise is the Heaviside step function.

these two cases by formally introducing a different cost function $\mathcal{H}_0\{a_{i,j}\}$ which vanishes only on the supports $a_{i,j}$ for which an admissible solution ofr $L_{i,j}$ exists. This cost function involves constraints that the approximate $\mathcal{H}$ is not able to capture.
Message passing algorithms can be derived along the lines of Refs. [9, 10] to solve efficiently the problem of sampling the space of solutions of (1) as described in detail in appendix. An additional variable (analogous to a chemical potential in physics) can be introduced in order to select denser or sparser sub-graphs (i.e. tuning the λ parameter). In particular, this allows one to find the maximally sparse subgraph compatible with the constraints.

5. Furfine stress-test

The aim of this section is to show that some measures of vulnerability of a banking system to financial contagion, also known with the name of stress-tests, are sensitive to the way in which the liability matrix is reconstructed. In particular the dense ME reconstruction typically underestimates the risk of contagion, while more realistic results are found if one employs a sparsification parameter λ controlling the density of links in a financial system.

A widely used measure of vulnerability in financial literature is the stress-test introduced by Furfine [11], which is a sequential algorithm to simulate contagion. Suppose that the liability matrix L is given and let us define C_z the initial capital of a bank z in the system $\mathcal{B}$. The idea of the algorithm is simple: suppose that a bank z of the ensemble $\mathcal{B}$ fails due to exogenous reasons. Then it is assumed that any bank $i \in \mathcal{B}$ loses a quantity of money equal to its exposure versus z (L_{iz}) multiplied by an exogenously given parameter $\alpha \in [0, 1]$ for loss-given-default. Than if the loss of the bank i exceeds its capital C_i, bank i fails. This procedure is then iterated until no more banks fail, and the total number of defaults is recorded.
The procedure described above can be formally rephrased in the following steps:

Step 0: A bank $z \in B$ fails for external reasons. Let us define $D_0 = \{z\}$, $S_0 = \mathcal{B} \backslash \{z\}$. For the banks $i \in S_0$ we set $C_i^0 = C_i$.

Step t: The capital C_i^{t-1} at step $t-1$ of banks $i \in S_{t-1}$ is updated according to

$$C_i^t = C_i^{t-1} - \alpha \sum_{j \in D^{t-1}} L_{ij}$$

with $\alpha \in [0, 1]$. A bank $i \in S_{t-1}$ fails at time t if $C_i^t < 0$. Let us define D_t, the ensemble of all the banks $i \in S_{t-1}$ that failed at time t and $S_t = S_{t-1} \backslash D_t$ the ensemble of banks survived at step t.

Step t_{stop}: The algorithm stops at time t_{stop} such that $D_{t_{stop}} = \varnothing$.

We remark that the capital C_i of each bank is exogenously given, and in principle it is not linked to the liability matrix L. The same holds for α, so that the result of a stress-test is understood as a curve quantifying the number of defaults as a function of the α parameter. Finally, the results of the stress-test depend on the first bank $z \in \mathcal{B}$ which defaults. Then one may choose either to consider the results of the stress-test dependent on the z which has been chosen or to average the outcome on all the banks in the system $\mathcal{B}$; we adopt this second type of measure, and consider the default of all the banks to be equally likely.

6. Application to synthetic data

In this section we will show how our algorithm of reconstruction of the liability matrix L_{ij} (presented in section 4) gives more realistic stress-test results if compared with ME reconstruction algorithm (presented in section 3).
We choose to present the results obtained for specific ensembles of artificial matrices, whose structure should capture the relevant features of real credit networks[4]. The first case that we analyze is the simplest possible network with a non-trivial topology, namely the one in which every entrance of the liability matrix L_{ij} with $i \neq j$ is set to zero with probability λ_{tr}, and otherwise is a random number uniformly chosen in $[0, 1]$. We set the banks initial capital C_i to random numbers uniformly chosen in $[C_{min}, C_{max}]$. We impose the threshold $\theta = 1$, which means that all the entrance of the liability matrix are unknown (a worst-case scenario). We than reconstruct the liability matrix via ME algorithm and via our algorithm trying to fix the fraction λ of zeroes equal to λ_{tr}. Then we stress-test via the Furfine algorithm the three liability matrices: the true one, the one reconstructed via ME algorithm one and the reconstructed by means of our message-passing algorithm, varying the loss-given-default α in $[0, 1]$. The results of our simulations are shown in figure 2. We clearly show that the ME algorithm underestimates the risk of contagion, while more realistic results are obtained if the correct degree of sparsity λ_{tr} is assumed.

Notice, that even with the correct estimate of the sparsity, stress tests on the reconstructed matrix still underestimate systemic risk. This is because the weights L_α on the reconstructed sub-graph are assigned again using the ME algorithm. This by itself produces an assignment of weights which is much more uniform than a random assignment of L_{ij} on the sub-graph, which satisfies the constraints (see footnote 2). As a result, the propagation of risk is much reduced in the ME solution.

The second ensemble that we consider is a simple extension of the first one, in which the only modification that we have introduced implements heterogeneity in the size of the liabilities L_{ij}. In particular we consider matrix elements distributed according to:

$$p(L_{ij}) \sim (b + L_{ij})^{-\mu-1}$$

[4]Our attempts to obtain data on real financial networks, such as those in Refs. [6, 7], from central banks were unsuccessful. We focus on ensembles of homogeneous networks (i.e. non-scale free). This is appropriate since the unknown part of the financial network concerns small liabilities, and there is no a priory reason to assume a particularly skewed distribution of degrees for the unknown part of the financial newtork.

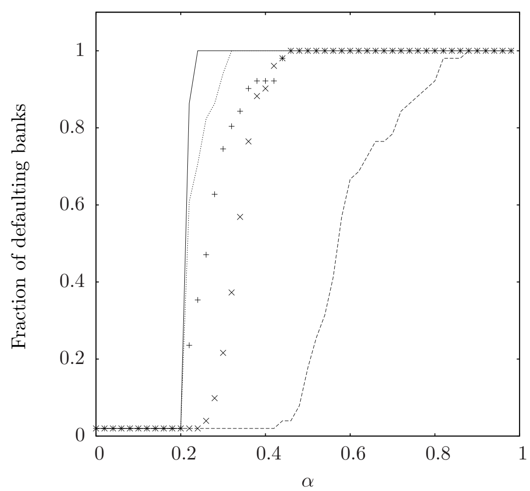

Figure 2. Plot of mean fraction of failed banks vs loss-given-default parameter α. The mean is done by averaging over the defaulting bank which starts the contagion. Results are obtained by considering: true liability matrix (solid line), reconstructed via ME algorithm liability matrix (thick dashed line) and the maximally sparse matrix (soft dashed line). Plots were obtained for a network of $N = 50$ banks with entries uniform in [0,1], where the link probability was fixed to 1/2 and the initial capital was set to $C_i = C = 0.2$. One can easily see that a better estimation of the true risk of contagion is obtained if the reconstruction of the liability matrix is done by enforcing the correct sparsity of the network rather than with the ME algorithm: the results obtained by putting the *correct* support ($\times$ signs) are also plotted, as well as the ones obtained by using a *typical* support with a correctly tuned sparsity parameter (+ signs).

Also in this case we can show (figure 3) that a more correct estimation of the default probability is achieved by enforcing the sparsity parameter of the reconstructed network to be the correct one. In this case the maximally sparse curve is less informative than in the uniform case. This is easily understood as due to the fact that the typical element $L_{ij} \sim 10^{-2}$ is much smaller than the threshold $\theta = 1$, so that a number of zero entries substantially larger than the correct one can be fixed without violating the hard constraints.

In both cases, when the true sparsity of the network is unknown, focusing on the sparsest possible graph likely over-estimates systemic cascades, thereby providing a conservative measure for systemic risk.

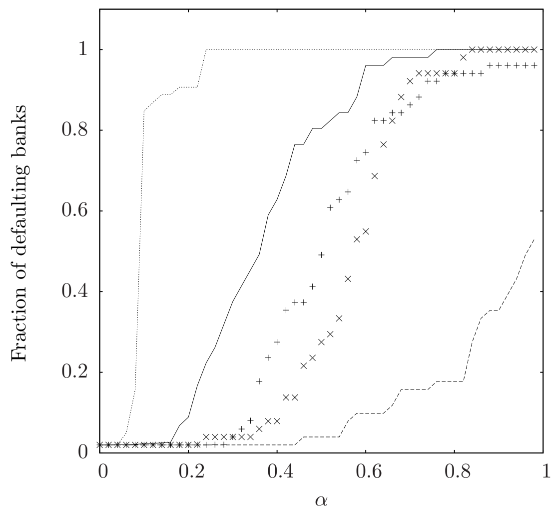

FIGURE 3. A plot analogous to the one in figure 2 for the case of power-law distributed entries of the liability matrix. This plots was obtained for a network of size $N = 50$, where the link probability was fixed to $1/2$. The parameters for the distribution of the entries were set to $b = 0.01$ and $\mu = 2$, while the capital of each bank was fixed to $C_i = C = 0.02$.

7. The role of the threshold

In the discussion above we disregarded the role of the threshold θ above which an exposure L_{ij} has to be made publicly available to regulators by setting it equal to 1. Indeed the problem of setting such threshold is a central problem to build a regulatory policy, hence the discussion of the reliability of the reconstruction algorithm varying θ while keeping fixed the true L is in practice particularly relevant. An appropriate way to address this issue is the following: given a network ensemble (such as the ones described in previous section) and a threshold θ, how many network structures are there with a compatible support? In particular, we remark that among all such compatible supports the maximally sparse one can be used to bound from above the maximum amount of risk given a policy for the thresholding. In particular for each value of θ, we empirically find that $\lambda_{max}[\theta]$ enjoys the following properties:

(1) The maximum sparsity $\lambda_{max}(\theta)$ is a decreasing function of θ. In particular for $\theta \to 0$ one has $\lambda_{max}(\theta) \to \lambda_{tr}$;
(2) The entropy $S(\lambda(\theta)) \to 0$ when the threshold goes to 0.

An example of this behavior for an ensemble of networks with power-law distributed weights is represented in figure 4, while in 5 we plot the entropy $S(\lambda_{max})$ structures as a function of θ. Therefore the algorithm described in section 4 provides quantitative measures for the uncertainty induced by the choice of a given threshold θ on network reconstruction.

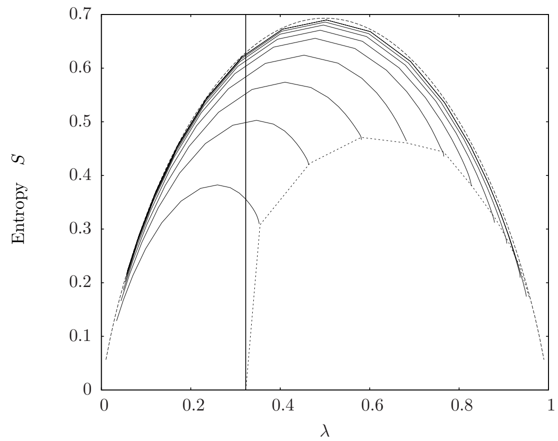

Figure 4. We plot the entropy of the space of compatible distributions (i.e. of the solutions of $\mathcal{H}\{a_{i,j}\}$) as a function of the sparsity parameter λ by varying the threshold θ from 1 (top curve) to 0.01 (bottom curve). The dashed line signals the transition point where solutions cease to exist. We consider power-law distributed entries for the true network ($D = 30$, $\lambda_{tr} \approx 0.3$, $b = 0.01$ and $\mu = 2$). This shows how the volume of the space is reduced by a change of the threshold and how λ_{max} gets closer to λ_{tr} by lowering θ.

Ideally θ should be chosen so that maximally sparse structures are close to the true ones, and that the space of compatible structures is not too large (small entropy).

8. Conclusions

We have shown how it is possible to estimate the robustness of a financial network to exogenous crashes by using partial information. We confirm [6] that systemic risk measures depend crucially on the topological properties of the underlying network, and we show that the number of links in a credit network controls in a critical manner its resilience: connected networks tend to absorb the response to external shocks more homogeneously than sparse ones. We have also proposed an efficient message-passing algorithm for the reconstruction of the topology of partially unknown credit networks, in order to estimate with more accuracy their robustness. Such algorithm allows (i) to sample the space of possible network structures, which is assumed to be trivial in Maximal Entropy algorithms commonly employed for network reconstruction, and (ii) to produce typical credit networks, respecting the topological constraint on the total number of links. Finally, we test our algorithms on ensembles of synthetic credit networks which incorporate some of the main features of real credit networks (sparsity and heterogeneity), and find that the quality of the stress-test when only partial information is available critically depends on the assumptions which are done about the network topology. In particular, we find that ME underestimates

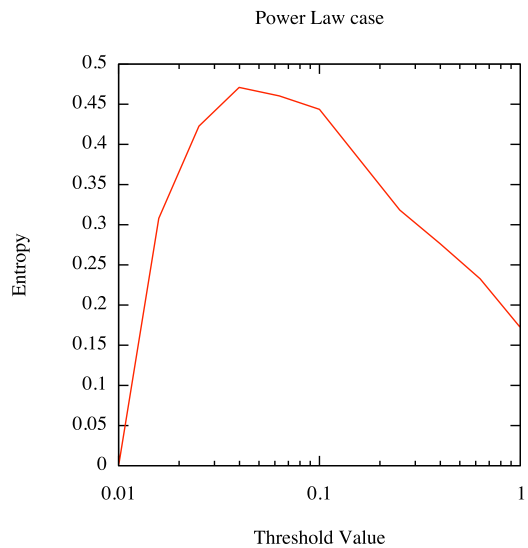

FIGURE 5. The entropy of the space of solutions $\mathcal{H}\{a_{i,j}\}$ as a function of the threshold for the same network as the one depicted in figure 4.

the risk of contagion if the sparsity of the real ensemble is big enough, while our algorithm provides less biased estimates. We remark that a worst case analysis of the topology is possible using the proposed algorithm, as we are able to produce the maximally sparse (hence, maximally fragile) possible structure for the network. Further developments of this work are indeed possible, in particular the identification and the reconstruction of other relevant topological features of credit networks would be relevant for a more accurate estimation of the contagion risk.

REFERENCES

[1] Schwarcz S. L., *Systemic Risk*, Duke Law School Legal Studies Paper No. 163

[2] Blavarg, M. and P. Nimander, Interbank Exposures and Systemic Risk, *Economic Review*, 2002, 2, 1945

[3] Wells S., Financial interlinkages in the United Kingdom's interbank market and the risk of contagion, 2004 *Bank of England* Working Paper No. 230

[4] Degryse H. and Nguyen G., Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System, 2007, *International Journal of Central Banking*, 3, 123-171.

[5] van Lelyveld I. P. P. and Liedorp F.R.,Interbank Contagion in the Dutch Banking Sector, 2006, *International Journal of Central Banking*, 2, 99-134.

[6] Mistrulli P.E., Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns, 2010 *Journal of Banking & Finance*

[7] Boss, M. and Elsinger, H. and Summer, M. and Thurner, S., Network topology of the interbank market, *Quant. Financ.*

[8] Upper C., Using counterfactual simulations to assess the danger of contagion in interbank markets, *BIS working papers* No 234
[9] Pretti, M. and Weigt, M., Sudden emergence of q-regular subgraphs in random graphs, *Europhysics Letters*, 75, 2008
[10] L. Zdeborová, M. Mézard, Constraint satisfaction problems with isolated solutions are hard *J. Stat. Mech.* 2008 P12004.
[11] Furfine, C. H., Interbank exposures: Quantifying the risk of contagion, *Journal of Money, Credit and Banking*, 2003

Appendix A. Message-passing algorithm

We describe here the algorithm which we use to sample the solution space of the energy function:

$$\mathcal{H}\{a_{i,j}\} = \sum_i \left[\theta\left(L_i^{\rightarrow} - k_i^{\rightarrow}\right) + \theta\left(L_i^{\leftarrow} - k_i^{\leftarrow}\right)\right]$$

which we derived along the line of [9]. The structure of the problem admits a graphical representation as a factor graph, in which $|U|$ variable nodes are associated to the $a_{i,j}$ degrees of freedom, while the constraints are represented as factor nodes. In particular, there are $2N$ function nodes, labeled $a \in \{i \rightarrow, \leftarrow i,\ i = 1, \dots, N\}$ each with k_a variable nodes attached. Let the variables be denoted $x_{a,b} = x_{b,a} = 0, 1$ with a, b and let ∂a be the set of neighbors of node a. Let $M = \frac{1}{2}\sum_a |\partial a|$ be the total number of variables. The messages can be written as

$$\mu_{a\to b} = P\{x_{a,b} = 1 |\ \not{b}\}$$

where $\not{b}$ means "when the node b is removed". The BP equations are written in terms of the statistical weights[5]

$$V^m_{S\to a} = \sum_{U\in S:|U|=m} \prod_{b\in U} \mu_{b\to a} \prod_{c\in S\setminus U} (1 - \mu_{c\to a})$$

and they read:

$$\begin{aligned} \mu_{a\to b} &= \frac{\sum_{m=L_a-1}^{k_a-1} z^{m+1} V^m_{\partial a\setminus b\to a}}{\sum_{m=L_a-1}^{k_a-1} z^{m+1} V^m_{\partial a\setminus b\to a} + \sum_{m=L_a}^{k_a-1} z^m V^m_{\partial a\setminus b\to a}} \qquad (2)\\ &= \frac{V^{L_a-1}_{\partial a\setminus b\to a} + zW_{\partial a\setminus b\to a}}{V^{L_a-1}_{\partial a\setminus b\to a} + (1+z)W_{\partial a\setminus b\to a}} \end{aligned}$$

$$W_{\partial a\setminus b\to a} = \sum_{m=L_a}^{k_a-1} z^{m-L_a} V^m_{\partial a\setminus b\to a}. \qquad (3)$$

[5] Since k_a can be as large as N, the direct computation of $V^m_{S\to a}$ involved in principle 2^{k_a} terms, which may be very large. A faster way to compute it is to use the recursion relation

$$V^m_{S\to a} = (1-\mu_{b\to a})V^m_{S\setminus b\to a} + \mu_{b\to a}V^{m-1}_{S\setminus b\to a},\ \forall b \in S.$$

In practice this allows one to build $V^m_{S\to a}$ adding one at a time the nodes in S. This procedure involves of order $m^2 \le k_a^2$ operations.

Here z is the fugacity of links, and controls the degree of sparsity λ of the typical supports in the solution space. For $z \to 0$ we obtain the equation for the sparsest possible graph

$$\mu_{a\to b} = \frac{V_{\partial a\setminus b\to a}^{L_a-1}}{V_{\partial a\setminus b\to a}^{L_a-1} + V_{\partial a\setminus b\to a}^{L_a}}$$

whereas for $z \to \infty$ we recover the maximally connected graph $\mu_{a\to b} = 1$ for all a and $b \in \partial a$.

Once the fixed point of Eqs. (2,3) is found by iteration, for a given z, one can compute the probability

$$p_{a,b} = P\{x_{a,b} = 1\} = \frac{\mu_{a\to b}\mu_{b\to a}}{\mu_{a\to b}\mu_{b\to a} + (1-\mu_{a\to b})(1-\mu_{b\to a})}$$

that link (a, b) is present, and the entropy

$$S(z) = \sum_a \log \sum_{m=L_a}^{k_a} V_{\partial a\to a}^m - \frac{1}{2}\sum_a \sum_{b\in\partial a} \log\left[\mu_{a\to b}\mu_{b\to a} + (1-\mu_{a\to b})(1-\mu_{b\to a})\right]$$

To plot the number of solutions (or of different supports) as a function of the sparsity parameter λ, and the associated entropy $\Sigma(\lambda)$ one should use the fact that:

$$e^{MS(z)} = \int_0^1 d\lambda\, e^{M\Sigma(\lambda)+M(1-\lambda)\log z}$$

and hence perform the back-Legendre transform.

Iacopo Mastromatteo, *International School for Advanced Studies, via Beirut 2/4, 34014, Trieste, Italy*

Elia Zarinelli, *LPTMS, CNRS and Université Paris-Sud, UMR8626, Bât. 100, 91405 Orsay, France*

Matteo Marsili, *The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*