

Nicolò Musmeci · Stefano Battiston ·
Guido Caldarelli · Michelangelo
Puliga · Andrea Gabrielli

Bootstrapping topology and systemic risk of complex network using the fitness model

Received: date / Accepted: date

Abstract We present a novel method to reconstruct complex network from partial information. We assume to know the links only for a subset of the nodes and to know some non-topological quantity (fitness) characterising every node. The missing links are generated on the basis of the latter quantity according to a fitness model calibrated on the subset of nodes for which links are known. We measure the quality of the reconstruction of several topological properties, such as the network density and the degree distribution as a function of the size of the initial subset of nodes. Moreover, we also study the resilience of the network to distress propagation. We first test the method on ensembles of synthetic networks generated with the Exponential Random Graph model which allows to apply common tools from statistical mechanics. We then test it on the empirical case of the World

N. Musmeci
Dipartimento di Fisica Università “Sapienza” P.le A. Moro 5, 00185 Rome, Italy.
E-mail: nik.musmeci87@gmail.com

S. Battiston
Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, 8092, Zurich, Switzerland. E-mail: sbattiston@ethz.ch

G. Caldarelli
IMT Alti Studi Lucca, Piazza S. Ponziano 6 55100, Lucca, Italy. ISC-CNR UdR Sapienza, P.le A. Moro 5, 00185 Rome, Italy. London Institute for Mathematical Sciences 35a South Street Mayfair London W1K 2XF, UK. E-mail: guido.caldarelli@imtlucca.it

M. Puliga
Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, 8092, Zurich, Switzerland. E-mail: puligam@ethz.ch

A. Gabrielli
ISC-CNR UdR Sapienza, P.le A. Moro 5, 00185 Rome, Italy. IMT Alti Studi Lucca, Piazza S. Ponziano 6 55100, Lucca, Italy. London Institute for Mathematical Sciences 35a South Street Mayfair London W1K 2XF, UK. E-mail: andrea.gabrielli@roma1.infn.it

Trade Web. In both cases, we find that a subset of 10 % of nodes is enough to reconstruct the main features of the network along with its resilience with an error of 5%.

Keywords Complex Networks · Financial Systems

PACS 89.75.Da · 02.50.Le · 89.65.Ef

1 Introduction

The reconstruction of the system from partial information is one of the outstanding and unresolved problems in the field of statistical physics of networks [4, 15]. Indeed, there are several real-world economic and financial contexts where knowledge of the entire network structure would be crucial to assess the resilience of the system to both exogenous and endogenous shocks while, at the same time, only limited information on that structure is available. An example is the case of financial networks where nodes represent financial institutions and links represent financial ties of various types such as loans or derivative contracts. These ties result in many cases in dependencies among institutions and constitute the ground for the propagation of financial distress across the network. The resilience of the whole system to the default or the distress of one or more institutions depends on the topological structure of the network [1, 2]. Unfortunately, due to confidentiality issues banks do not disclose their mutual exposures.

Typically the analysis of systemic risk is done by reconstructing the network using the so-called Maximum Entropy (ME) algorithm. This method assumes that the network is fully connected (for this reason this class of approaches is called "dense reconstruction methods"). The weights of the links are then obtained via a "maximum homogeneity" principle. This means that each node is assumed to bear a similar level of dependence from all other nodes. The second step consists in finding a matrix that, while satisfying certain constraints (imposed in this case by the budget of the individual banks), minimizes the distance from the uniform matrix in which each entry has the same value. Such a matrix is found using the Kullback-Leibler divergence as the objective function to minimize [6, 14]. The hypothesis that the network is fully connected is a strong limitation of the ME algorithm, since empirical networks are characterized by heterogeneous degree. Moreover, [15] has shown how the "dense reconstruction" leads to an underestimation of the systemic risk. They also provide a new algorithm that allows to minimize Kullback-Leibler divergence obtaining a matrix with an arbitrary level of heterogeneity under a maximum value depending on constraints. Their algorithm provides a "sparse reconstruction" that is more reliable than the dense one. Nevertheless it leaves open the question of what value of heterogeneity would be appropriate to choose, since the density of connections must be specified ex-ante and it is not recovered by the algorithm.

In this paper we introduce a new general method, the Bootstrapping Method (BM), to reconstruct the topology of the whole network starting from the knowledge of a subset of nodes. This method overcomes some of

the limitations described above. To validate the method we use both synthetic networks as well as an example of real economic systems. In both cases we compare the reconstruction with the known total structure. This is the so-called World Trade Web (WTW), i.e. the network of countries where links represent the financial flows corresponding to the trade volumes among them. In our method, the allocation of the links among nodes is carried out using the fitness model [3, 11]. Differently from other network generation models, the fitness model generates a network structure starting from a non-topological variables (fitness) associated to the nodes. This approach has been used in the past to reproduce the topological properties of several empirical economical networks, including the network of equity investments in the stock market [10], the interbank market [5], the currency market [9], and the WTW [12]. We investigate how well it is possible to recover both the topological properties of the network and its resilience to distress propagation, as we vary the size of the subset of nodes for which information is available. Among the topological properties, we focus on those that play an important role in contagion processes and in the propagation of distress, i.e., the network density [1], the degree distribution [17], the k-core structure [13].

We find that having information on a relatively small fraction of nodes is sufficient to recover with good approximation the above properties. For instance, with only about 7% of the nodes (10 out of 185) we have a relative error of about: 7% on the density, 10% on the average degree of the main core, 7% on the size of the main core. For the resilience, we focus on a recently introduced notion, debtrank [2], which measures the systemic impact of the initial distress of a subset of nodes, whenever the links in the network represent the dependencies among nodes. Similarly to the above results, we find that with about 7% of the nodes the resilience is recovered with a relative error within 10%.

At a first thought, it can be surprising that a small fraction of nodes enables to reconstruct so well global emerging properties of the network. However, one should bear in mind that in the method, while the links are known only for a subset of nodes, the fitness is always known for all the nodes. Thus, the method would probably require much higher fraction of nodes in order to reconstruct networks with special topologies such as strong community structure or networks where the fitness is not a strong factor in driving the connectivity. The investigation of these effects is left for future research. Overall, our method can be applied in principle to any network representing a set of dependencies among components in a complex system and it is thus of general interest in the field of complex networks and statistical physics.

2 Exponential random graph and fitness model

In this paper we propose a Bootstrap Method (BM) to build the network using a fitness model. The method is described in Section 3. Here we briefly describe the ERGMs and the associated fitness model. In order to generate ensembles of complex networks both dynamic and static approaches can be

utilized. In the dynamic case, nodes and/or links are added step by step using for instance a "preferential attachment" algorithm. In the static case, instead, the number of nodes is fixed the links are assigned at once according to some statistical or deterministic criterion. Exponential random graph models (ERGM) are one of the most studied class of network models[16, 7]. They can be described using the powerful mathematical formalism of the equilibrium statistical mechanics [16].

As a specific example, we will consider the so-called *fitness* or *hidden variables* models, where the network topology is determined by an intrinsic property (called *fitness*) associated with each node of the network[3]. Through this scheme we can define a framework to investigate those networks where the topology is driven, at least in part, by non-topological properties of the nodes. With the fitness model it is possible to study several economical networks, ranging from the WTW (where the fitness of the model are the GDP of the various countries) [12], to the financial networks (where fitness are, for instance, the market capitalization of each institution) [10, 5].

Given a set of network properties, $\{C_a\}$ the Exponential Random Graph Model (ERGM) is defined as the ensemble Ω of maximally random networks with $\{C_a\}$ constrained to some statistical properties. More specifically, let us suppose that the ensemble averages of $\{C_a\}$ are fixed:

$$\langle C_a \rangle_{\Omega} \equiv \sum_G P(G) C_a(G) = C_a^* \quad \forall a \quad (1)$$

It has been shown that Ω can be defined through a set of control parameters $\{\theta_a\}$, the values of which depend on a set of constraining values $\{C_a^*\}$ [16, 7]. Furthermore the probability $P(G)$ of a network G to occur in Ω is given by $P(G) = e^{-H(G)}/Z$, where we introduced the graph Hamiltonian $H(G) \equiv \sum_a \theta_a C_a(G)$, and the partition function $Z \equiv \sum_G \exp(-H(G))$. $\{\theta_a\}$ is the set of Lagrange multipliers associated to the constraints $\{C_a^*\}$. The fitness model can be seen as a specific case where the set of properties $\{C_a\}$ is the degree sequence $\{k_i\}_{i=1, \dots, N}$ of the nodes of the network. In this case $H = \sum_i \theta_i k_i$, the partition function is exactly computable and each node can be identified by its control parameter (or Lagrange multiplier) θ_i . Fixing the values of $\{\theta_i\}$ is equivalent to fix the mean values of $\{k_i\}$. In order to further clarify the role of $\{\theta_i\}$ in controlling the topology, let us define $x_i \equiv e^{-\theta_i}$. It is possible to show that the ensemble is such that for each network in Ω two nodes i and j are connected with a probability given by:

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j}. \quad (2)$$

Therefore x_i can be considered as the fitness of the node i and it is related to the ability of i to create links with other nodes.

The average in Ω of several topological properties of the network can be expressed in terms of appropriate compositions of the linking probabilities

p_{ij} for every i and j . For instance, we can write the degree k_i as

$$\langle k_i \rangle = \sum_{j(\neq i)=1}^N p_{ij}; \quad (3)$$

the average nearest neighbor degree K_i^{nn} as

$$\langle k_i^{nn} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq j} p_{ij} p_{jk}}{\langle k_i \rangle}; \quad (4)$$

and the clustering coefficient C_i as

$$\langle C_i \rangle = \frac{\sum_{j \neq i} \sum_{k \neq j, i} p_{ij} p_{jk} p_{ki}}{\langle k_i \rangle [\langle k_i \rangle - 1]}. \quad (5)$$

In the limit of small values of fitnesses (and therefore small connectivity), x_i is proportional to the *desired degree* of the node i . Indeed, in this limit we can assume $\langle k_i \rangle \simeq \sum_j x_i x_j \propto x_i$.

3 Bootstrapping Method

The estimation of the linking probability p_{ij} between node i and node j , p_{ij} is the initial step in order to develop a network Bootstrapping method. Let us suppose to have incomplete information about the topology of a real network (say G_0). In particular, we assume to know the links of only a subset I of the nodes. Moreover, we assume to know, for all the nodes, a non-topological property, denoted as y_i , that is correlated to some statistical properties of the degree k_i of the nodes by a known relation as below clarified. For instance, in the World Trade Web y_i could be the country GDP, while in financial networks it can be the operating revenue of the firm i . We would like to estimate the value $x(G_0)$ of a topological property $X(G_0)$ of the network G_0 . We make two hypotheses:

1. the network G_0 has been drawn from an ensemble of ERGM, that we call Ω . From the statistical mechanics of networks we know that the value $x(G_0)$ of the property X in G_0 , varies within the range $\langle x \rangle_\Omega \pm \sigma_x^\Omega$ where σ_x^Ω is the standard deviation, and $\langle x \rangle_\Omega$ the average of the property X estimated on the whole ensemble Ω .
2. each known value of the non-topological property y_i is assumed to be proportional to the fitness, denoted as x_i (because a generic property of the network can be used as a fitness variable) of the node i in the ensemble Ω , through a universal unknown parameter z : $\sqrt{z} y_i = x_i$. Therefore Eq. (2) becomes:

$$p_{ij} = \frac{z y_i y_j}{1 + z y_i y_j} \quad (6)$$

With these hypotheses, we map the problem of evaluating $x(G_0)$ into the one of choosing an ERGM ensemble Ω compatible with the constraints given by the fitness. Once Ω is determined (it is univocally defined by the set of $\{x_i\}$), we can use the average $\langle X \rangle_\Omega$ as a good estimation for $x(G_0)$. Within this framework, the question is which ensemble Ω , belonging to the class ERGM, is the most probable to extract the real network G_0 , knowing only the partial information $\{y_i\}$. Since we know $\{y_i\}$, i.e. the rescaled fitness values (a non topological property of the network), the problem becomes to find the most likely value of z . For this reason we use the notation $\Omega(z)$ for the desired ensemble.

Notice that if we knew not only $\{y_i\}$ but the entire topology of the network, z could be found by means of a maximum likelihood argument (Ref. [12]) comparing the average number of links in the ensemble networks with the total number of links L_0 in G_0 :

$$\langle L \rangle = \frac{1}{2} \sum_{i=1}^N \langle k_i \rangle = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i} p_{ij} = L_0 \quad (7)$$

where p_{ij} contains the unknown parameter z through Eq. (6). Therefore we can evaluate z as L_0 is known and by definition of $\sqrt{zy_i} = x_i$ go back to x_i that is our desired output. Let us call z_0 the estimation calculated in this way, and $\Omega(z_0)$ the respective ERGM ensemble. But note that we assume to know only the degrees of the nodes in a subset I and not the entire topology. Let be n the number of nodes of I . In this case the relation we have to apply in order to use the maximum likelihood principle in the estimation of z is:

$$\sum_{i \in I} \langle k_i \rangle = \sum_{i \in I} \sum_{j \neq i} p_{ij} = \sum_{i \in I} k_i \quad (8)$$

where the degrees k_i are calculated in the original network G_0 . For a subset I of the entire set of nodes of the network the estimation is less precise than the one in Eq. (7). However even with just the knowledge of the degree of a single node, the Eq. (8) estimates z , and finally $X(G_0)$.

The network bootstrap of a network G_0 is defined by the above equations using the following procedure. Let us assume to know the non topological property y_i of all N nodes of the system and the links of a subset I of $n < N$ nodes.

- Given the topological information of the links in the subset I , we compute the sum of all degrees of these n nodes in G_0 : $\sum_{i \in I} k_i$.
- This sum is substituted into the Eq. 8 to obtain the relative value of z , denoted as z' , that is an approximation of the z_0 .
- With the value of z' and the knowledge of every y_i we assign all the links in the network according to the linking probability of Eq. 6.

We want to estimate the accuracy of the network bootstrap method for both topological and non-topological properties. To this end, we first apply the method to a synthetic network generated using the fitness model (see Section 4). We then apply the method to an empirical case, i.e. the WTW (see Section 5). In the second case, we test also how well we can reconstruct

a global and non-topological property such as the resilience of the network to distress propagation (see Section 6).

4 Test of BM: synthetic networks

Let be $\{I_\alpha\}$, with $\alpha = 1, \dots, M$, an ensemble of subsets of the network G_0 each of them containing n nodes. In order to test how much our BM estimate of the property X is precise, we will proceed in the following way:

- evaluate z for each subset I_α from Eq. (8), let us call it z_α
- use the value z_α to estimate, through the relation $\sqrt{z}y_i = x_i$, the average property $\langle X \rangle_\alpha \equiv \langle X \rangle_{\Omega(z_\alpha) \equiv I}$
- repeat the calculation for all other sets I'_α , accumulate the values of $\langle X \rangle_\alpha$ and compute the average $\overline{\langle X \rangle}$ of this quantity and its associated root mean square deviation with respect to the real value of $X(G_0)$ across all the realizations of I_α , for fixed n .

The property X is then estimated by averaging the $\langle X \rangle$ computed for each subset I . Notice that each value $\langle X \rangle$ is by itself an estimation of the true, unknown, property X .

In order to study the accuracy of the reconstruction, we study how the root mean square error varies as a function of the size n of the subset of nodes for which information is available. Using the fitness model and all the available information, we generate an ensemble of networks G each one of size N and we compute several properties like the network density, the size of the main core and the average degree of the main core. These values will be our benchmarks to test how good is the network reconstruction with the BM.

We test the BM by using the following three topological quantities because they have been found to play a role in the distress propagation and contagion processes and therefore are relevant to the resilience of the network to systemic risk (see Section 1):

1. **density** D ;
2. **degree of the main core**, k^{main} . In a network, the k -core is defined as the “largest subgraph whose nodes have at least k connections (within this subgraph, of course)” [7]. The main core is k -core with the highest possible degree, k^{main} ;
3. **size of the main core**, S^{main} , i.e. its number of nodes.

Each of these measures will play the role of the property X in the previous notation. In order to use a real-world fitness, we take as reference the WTW (in year 2000) which contains 185 nodes. We thus generate networks of size $N=185$ and we use as fitness y_i the GDP from the WTW. For each of these properties we will carried out the procedure described here below.

1. choose a value for the variable z_0 (compatible with the fitness model for WTW, where the fitness is the GDP of a country); we start with $z_0 = 10^4$

2. using as fitness the GDP of a country create 50 networks. Let be Ω_N this ensemble. Compute on the this set the average link density D_e
3. use a 51th network from the Ω_N ensemble as reference network, call it G_0 , this will be the network to reconstruct
4. starting from network with a single node $n = 1$ with known degree k and GDP y_i use this information to compute an estimation of z , say z'
5. from the new value of z' create a new ensemble of 50 networks (say I_α because is referring to one particular set of random chosen nodes)
6. choose another set of n nodes, generate 50 networks from this set, repeat this operation 100 times each time with a different set I_α of n nodes
7. in each of the 100 ensembles of 50 networks, I_α estimate the average density $\langle D_\alpha \rangle$
8. compute the root mean square error: $\sigma_d = 1/100 \sum_\alpha \sqrt{(\langle D_\alpha \rangle - D_e)^2}$ the difference is between the reconstructed networks $\langle D_\alpha \rangle$ and the original average link density D_e
9. compute and plot σ_d/D_e
10. repeat the points from 4 to 9 using a different values of n

The entire procedure is repeated for the quantities S^{main} and k^{main} , and the results are shown in Fig.1 for 3 different values of z_0 , corresponding to different values of density. We observe that in all cases there is a rapid decrease of the relative error as the number of nodes n , used to reconstruct the topology, increases. This is a good indication of the goodness of the method. Even with a single node, plus the information on the fitness y_i (GDP of the countries), we are able to estimate the topological properties of the network with a relative error of about 13% for the main core average degree k^{main} , about 18% for the network density D , and about 10% for the size of main core S^{main} .

As expected, if we have a denser network (Fig.1d-f) the relative error is smaller because the network has more links from with the BM can reconstruct the topology. The same trend in the decrease of the relative error is verified for all the examined topological quantities.

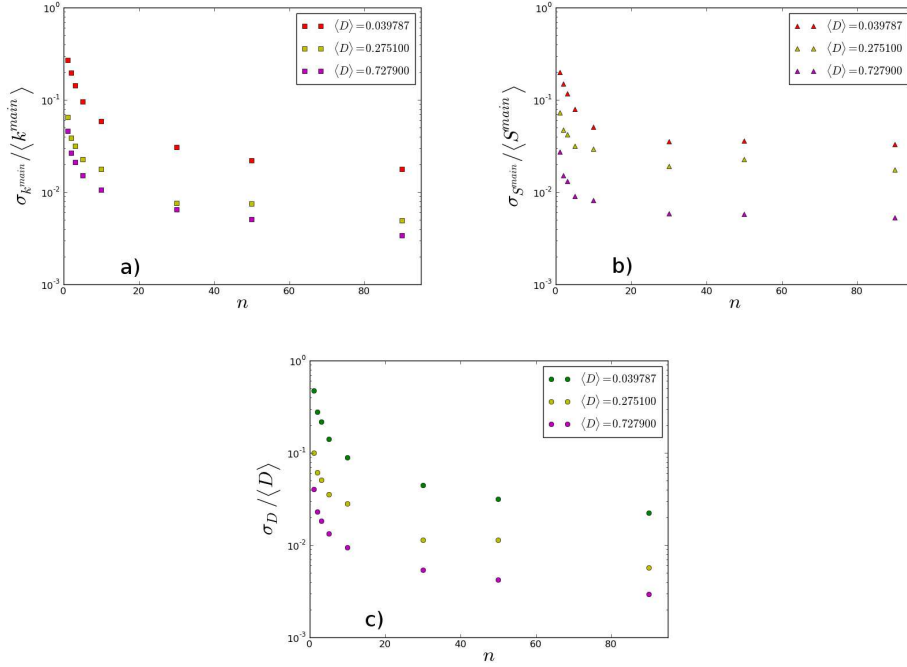


Fig. 1: The pictures from top left represent respectively: a) using three values of z_0 related to three different link average densities $\langle D \rangle$ (estimated numerically) compute the relative error $\sigma_k^{main}/k_e^{main}$ for the three cases b) same as (a) but for the relative error of the S -main core size, c) same as in (a) but for the density of the links D . In all the 3 plots it is evident how the quality of the reconstruction increases with the number of nodes used to generate the network ensemble.

5 Test of BM: World Trade Web

We now test the empirical network of the WTW for the same topological properties of the previous case. The main difference is that now instead of using a reference network generated with the fitness model and an average measure of this network class (generated in the ensemble Ω_N), the reference is now the empirical WTW network. We perform the test with the following similar procedure:

1. compute the variable z_0 from the WTW using the GDP of the countries and all the links of the original network
2. from the WTW network compute the density D_{WTW} of the links
3. starting from a network with a single node $n = 1$ with known degree k and GDP y_i use this information to compute an estimation of z , say z'

4. from the new value of z' create a new ensemble of 50 networks (say I_α because is referring to one particular set of random chosen nodes)
5. choose another set of n nodes, generate 50 networks from this set, repeat this operation 100 times each time with a different set I_α of n nodes
6. in each of the 100 ensembles of 50 networks, I_α estimate the average density $\langle D_\alpha \rangle$
7. compute the root mean square error: $\sigma_d = 1/100 \sum_\alpha \sqrt{(\langle D_\alpha \rangle - D_{WTW})^2}$ the difference is between the reconstructed networks $\langle D_\alpha \rangle$ and the original WTW link density D_{WTW}
8. compute and plot σ_d/D_{WTW}
9. repeat the points from 4 to 9 using a different value of n

the same test is carried out for the other quantities k^{main} and S^{main} . We expect to see more error than the previous case. We plot the quantity

6 Test of BM: DebtRank a measure of systemic risk

With the WTW network we can use a novel measure of systemic risk: the DebtRank [2] that represents the expected distress of the nodes in case of financial events. In the WTW case a financial event can be, for instance, the default of a country and the subsequent impossibility to pay the traded goods: this *shock* generate a distress propagation in the network causing losses to the other countries. The DR captures the impact I_i of the shock to each node i .

We compute the DebtRank DR of a single node (due to a shock hitting a single node at a time), and the group debtrank GDR of the ensemble (due to an initial shock hitting all the nodes simultaneously) using an algorithm described in the Methods section of [2] that consists in computing a feedback centrality from the matrix of the weights, given an initial shock ψ (to one or more nodes) that is carried out by a variable h_i *impact* specific to the method. During the calculation of this feedback centrality we use several values of the impact rescaling factor $0 < \alpha < 1$ to propagate the shocks in the network.

To compute the DR of each node we use the procedure:

1. choose an impact rescaling factor α , the greater is this factor the greater will be reverberation effect on the network
 2. assign an initial shock $0 < \psi < 1$ to a node i
 3. run the DR algorithm (as described in the Methods section of [2])
 4. save the values of the impact at the end and the beginning of the simulation: the DR of the node i will be the difference in the impact on all nodes after the propagation of the distress
 5. repeat for a different node
1. impact rescaling factor α
 2. assign an initial shock $0 < \psi < 1$ to *all* nodes
 3. run the DR algorithm

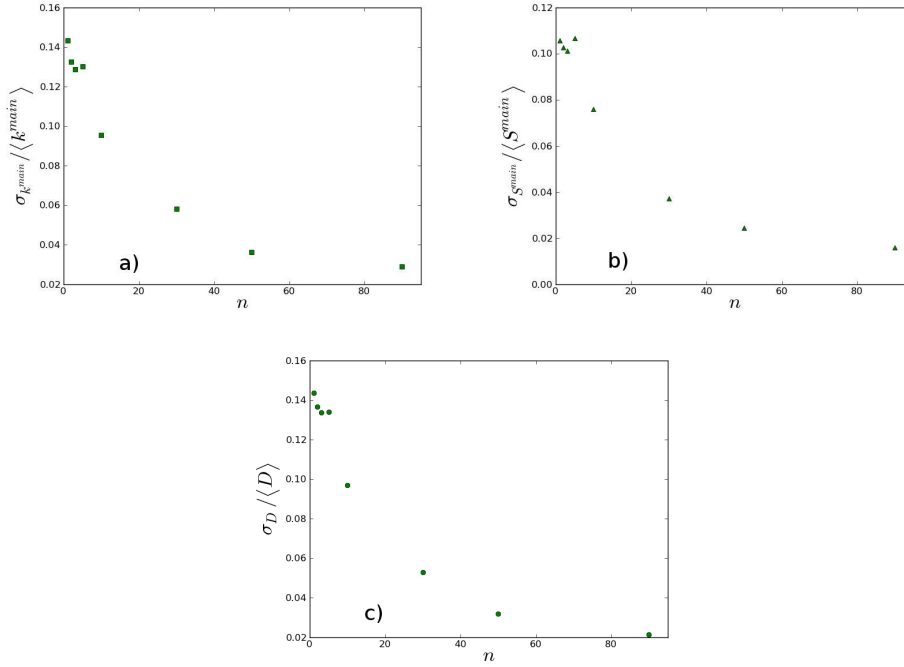


Fig. 2: The pictures from top left represent respectively: a) the relative error in the estimation of average degree of the main core $\sigma_k^{main}/k_{WTW}^{main}$ computed with real WTW network following the procedure described in the previous paragraph b) same as in (a) but for the relative error in the size of main core c) same as in (a) but for the density of the links D . In all the 3 plots it is evident how the goodness of the reconstruction of the WTW network increases with the number of nodes used to generate the network ensemble.

4. save the values of the impact at the end and the beginning of the simulation. The GDR will be the difference in the impact on all nodes after the propagation of the distress: $GDR = \sum_j h_j(T)v_j - \sum_j h_j(1)v_j$, where h_i is the impact on each node, T indicates the end of the simulation (when the distress propagates to the entire network).

Our goal is to test how well the GDR is computed with the network bootstrap. We make several tests for different values of initial impact ψ and impact rescaling factor α . The DebtRank is strongly dependent by the weights (value of the elements of the adjacency matrix) that are instead unknown during the simulations. In fact the fitness model reconstructs the degree sequence, not the weight of the nodes we then use a value for each link with two rules:

- Compute the average weight averaging the elements of the W_{ij} matrix associated to the $n < N$ nodes. Use this value as *homogeneous* weight for all nodes
- Assign to each node an weight similarly to a gravity model (see [8]) where the link l_{ij} has a weight proportional to the product of the GDPs $GDP_i \cdot GDP_j$

We want to consider the impact in case a country defaults (it is not paying) while the actual adjacency matrix represents the economic value of the goods (the links are in the opposite direction). For this reason we transpose the WTW matrix and follow this procedure and we normalize imposing a row stochastic condition $\sum_j w_{ij} = 1$. The procedure for computing the GDR is the following (results on Fig. 3)

1. Compute the reference Group Debt Rank on the original WTW network with an initial shock $\phi = 0.1$ keep this value as reference (green dashed line in the plots)
2. Create a new network with the same topology of the WTW but with the weights replaced by the average weight of the WTW links (this value will be used in the simulation for all links)
3. Compute the GDR on a network with the same topology of the WTW and weight imposed homogeneous. Our goal is to study how close to this value (blue dashed line in the plot), using homogeneous values for the weights, the bootstrap method can go
4. Bootstrap the networks with networks of size $n < N$ using homogeneous weights (computed as average of the weights of the only nodes that we start from in the simulation), compute the average GDR on 50 bootstrapped networks with homogeneous weights. Repeat the operation 100 times changing the starting set of nodes during the generation of the 50 networks, obtaining for each n an average value with error (blue dots).
5. In the non homogeneous case (green dots) bootstrap the networks using weights according to a gravity model, where the weight of the link is the product of the GDPs of each node. To add an error, a "perturbation", on a such network we estimate empirically from the plot of W_{ij} vs $GDP_i \cdot GDP_j$ the average variation of the weight W_{ij} in function of the GDPs product. We then alter the corresponding adjacency matrix W'_{ij} imposing, for each weight, a random normal error: $w'_{ij} = w_{ij} + \sigma N(0, 1)$ where σ is a standard deviation computed on w_{ij} for the corresponding fixed value of $GDP_i \cdot GDP_j$. The new perturbed weight matrix is then transformed to maintain the row stochasticity.

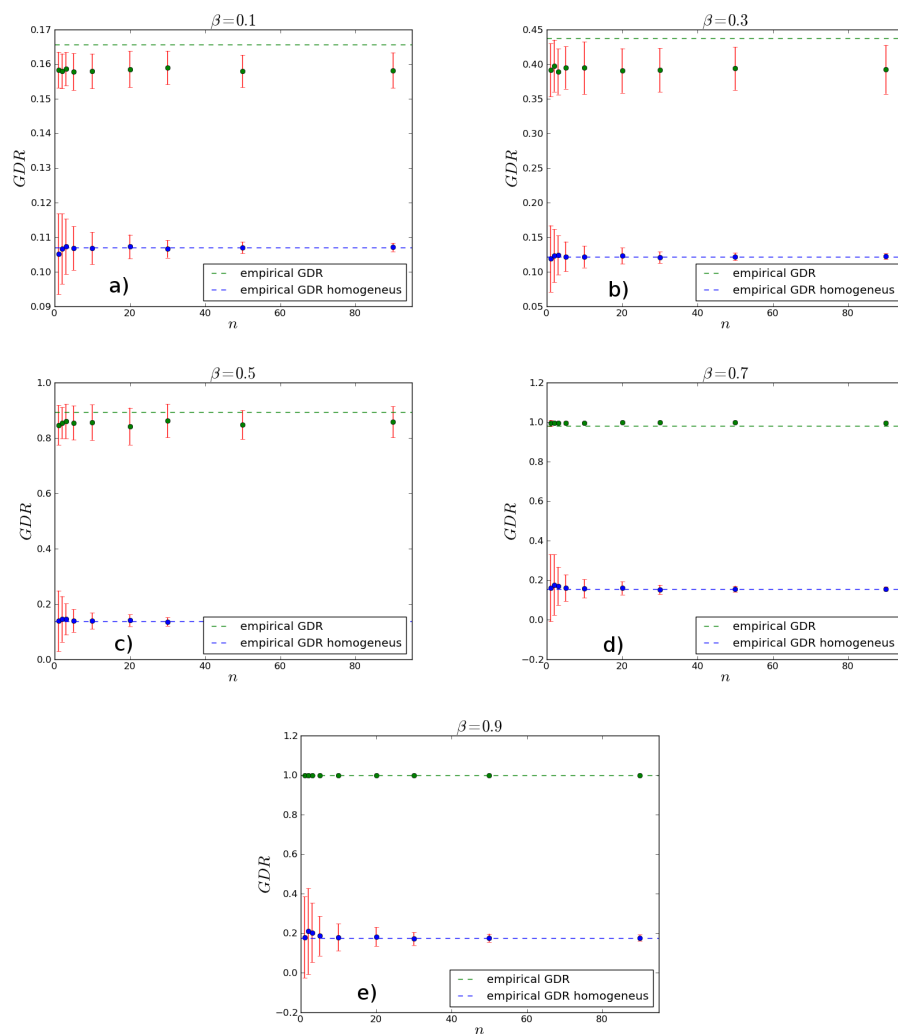


Fig. 3: The pictures from top left represent respectively: a) using the impact rescaling factor $\alpha = 0.1$, compute the Group Debt Rank on the original WTW network with empirical weights (green dashed line), the average Group Debt Rank on the 100 bootstrapped networks with weights obtained using gravity model (green dots) and respective errors, the Group Debt Rank on the original WTW network with homogeneous weights (blue dashed line), and finally the average Group Debt Rank on the 100 bootstrapped networks with homogeneous weights (blue dots) and respective errors. b) same as in (a) but for $\alpha = 0.3$, c) same as in (a) but for $\alpha = 0.5$, d) same as in (a) but for $\alpha = 0.7$, e) same as in (a) but for $\alpha = 0.9$

In Fig. 3 we plot the GDR for various α values, ranging from 0.1 to 0.9, in all cases the initial shock is $\psi = 0.1$, 10% of the value of the trade of every country. From the pictures we can draw the following conclusions:

- There is a significant difference when using homogeneous weights or heterogeneous weights. Using a constant value for the weight of each link lead to underestimating the value of systemic risk as measured by DebtRank
- The reconstruction of the DebtRank values is already good for small subsets of nodes in the network.
- Using a gravity model (even if simplified) improves the estimate of the GDR of the WTW network
- The gap between the homogeneous GDR and the empirical one increases for larger values of the impact rescaling factor α . This can be interpreted as follows: when the network effects are important the use of homogeneous weights in the dynamics leads to a larger error. Conversely when the network effects (reverberation) are less important the homogeneous weights are not so far from the *true* value of the GDR.

From this analysis we can conclude that the BM is good in reconstructing a non topological property such as the DebtRank but to achieve this goal one has to chose careful the weights of links because the use of an average value leads to inaccurate estimates, especially if the network effects are relevant.

The impact of the single countries (DR) on the WTW network is shown in Tab. 1 where the impact rescaling factor $\alpha = 0.5$ and the initial shock $\psi = 0.1$. As expected the biggest is the GDP the biggest is the corresponding DebtRank but with some variation due to network effects. Consider for instance a country like Canada, a big exporter of oil and minerals, its impact on the WTW will be larger than Germany that is a strong exporter of final goods. This analysis shows as the DebtRank measure is important to assess the distress (losses) propagation giving results that are not trivially expressed by the size of the countries.

Country	DebtRank	GDP rank(2000)
USA	0.48	1
JPN	0.32	2
CAN	0.26	8
CHN	0.23	6
DEU	0.23	3
MEX	0.18	10
GBR	0.17	4
FRA	0.16	5
ITA	0.12	7
NLD	0.10	15
KOR	0.09	12
TWN	0.09	16
BEL	0.08	20
ESP	0.08	11
SGP	0.07	39
MYS	0.05	40
CHE	0.05	18
BRA	0.05	9
IRL	0.04	38
AUS	0.04	14

Table 1: Table showing the DebtRank and the GDP rank (year 2000) for the 20 biggest countries in the WTW network. Notice that DebtRank is only in part respecting the same ranking given by the GDP of countries. Depending on the size of the export each country can be more or less affected by a shock on the default of the others countries. The values are computed using impact rescaling factor $\alpha = 0.5$ and $\psi = 0.1$, this numbers can interpreted in the following way: if the US will not pay the 10% of their obligations to the rest of the world, the size of the trade is so big to cause a total loss of 48% of the total WTW volumes. The amplification effect due to the network structure appears evident with a tool like DebtRank.

7 Conclusions

In this paper we have proposed a new method to reconstruct the topology of a network using only partial information from its connections and an auxiliary non-topological property: the fitness associated to each node. This method is particularly useful to overcome the lack of topological information for several financial networks whose systemic risk must be measured. Our approach allows to reconstruct the network using the topological information from one fraction of the nodes (i.e. their links) and a non topological property of each node derived from a fitness model.

We tested the network Bootstrap Method (BM) on the World Trade Web network, where we can use an accurate fitness model to describe its topology starting from the a non topological property (the GDP of the countries). We studied how well are reconstructed the following topological properties: the average density, the size, and the average degree of the main core. All these measures are related to systemic risk for financial networks as briefly presented in the introduction of this paper.

We found that the density of the links, the size of the main core, the average degree of the main core are reconstructed with an error varying from 1% to 10%, depending from the property examined, using with 5% of the nodes (10 over the 185 nodes of the WTW network). An interesting finding is that the denser is a network the better is the reconstruction. The goodness of the reconstruction increases with the number of the nodes used as initial information and it is strongly dependent by the accuracy of the fitness model (that for the WTW is an accurate model describing how the links form across countries depending on their GDP and geographical distance).

The BM method was checked with another non topological property: the DebtRank a novel measure of systemic risk. We discovered that the method was really effective in evaluating this property also with a small number of starting nodes. We carried out a test using link weights derived from the gravity model of the WTW (the weight of a link is proportional to the product of the GDPs of each node) or using homogeneous (averaged) weights.

In the case of homogeneous weights the BM estimates a value of Group Debt Rank (a measure of the DebtRank in a set of nodes) that is lower than the real one. This means that when the network is simulated using an average value for the weight there is a systematic bias in the evaluation of Systemic Risk measures such as the DebtRank. Conversely, in the case of non homogeneous weights, imposing more realistic values from the WTW fitness (gravity) model we obtain a more accurate estimation of the Group Debt Rank. This result stresses the importance, in the study of network systemic risk, to use a correct estimation of the weights and of the topological properties. Finally we notice that the bigger is the impact factor α in rescaling the nodes the greater is the distress propagation in the network captured by the Group Debt Rank.

We highlight that, for systemic risk, the network effects are responsible of an amplification of the distress. In fact the losses in the system due to a node failure (a default or a partial impossibility to pay) are bigger than the size of a country in terms of its ratio of GDP over the total market. In the paper we showed that countries that are not so big for GDP can have a significant impact on the WTW network depending on the size of their connections with others.

For what concerns possible future development, our work opens several challenges. To start with, we plan to test BM on other socio-economical networks, mainly financial ones. As written above, BM precision depends on how well fitness model describes real network. With WTW the fitness model works surprisingly well and it reproduces topological properties of any order [18]. We therefore need to test BM in more general cases, for example with networks where fitness model is less accurate and reproduces just some property.

Acknowledgements We thank support from FET Open project FOC (255987) and Italian PNR project CRISIS-Lab

References

1. Battiston, S., Gatti, D., Gallegati, M., Greenwald, B., Stiglitz, J.: Liasons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *J. of Economic Dynamics and Control* **36**(8), 1121–1141 (2012). URL <http://www.nber.org/papers/w15611>
2. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Caldarelli, G.: Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports* **2** (2012)
3. Caldarelli, G., Capocci, A., De Los Rios, P., Muñoz, M.: Scale-free networks from varying vertex intrinsic fitness. *Physical review letters* **89**(25), 258,702 (2002)
4. Clauset, A., Moore, C., Newman, M.: Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191), 98–101 (2008)
5. De Masi, G., Iori, G., Caldarelli, G.: Fitness model for the italian interbank money market. *Physical Review E* **74**(6), 066,112 (2006)
6. Degryse, H., Nguyen, G.: Interbank exposures: An empirical examination of contagion risk in the belgian banking system. *International Journal of Central Banking* **3**(2), 123–171 (2007)
7. Dorogovtsev, S.: Lectures on complex networks. *Physik Journal* **9**(11), 51 (2010)
8. Fagiolo, G.: The international-trade network: gravity equations and topological properties. *Journal of Economic Interaction and Coordination* **5**(1), 1–25 (2010)
9. Galluccio, S., Caldarelli, G., Marsili, M. and Zhang, Y.: Scaling in currency exchange. *Physica A* **245**(423) (1997)
10. Garlaschelli, D., Battiston, S., Castri, M., Servedio, V., Caldarelli, G.: The scale-free topology of market investments. *Physica A: Statistical Mechanics and its Applications* **350**(2), 491–499 (2005)
11. Garlaschelli, D., Capocci, A., Caldarelli, G.: Self-organized network evolution coupled to extremal dynamics. *Nature Physics* **3**(5), 813–817 (2007)
12. Garlaschelli, D., Loffredo, M.: Fitness-dependent topological properties of the world trade web. *Physical review letters* **93**(18), 188,701 (2004)
13. Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H., Makse, H.: Identification of influential spreaders in complex networks. *Nature Physics* **6**(11), 888–893 (2010)
14. van Lelyveld, I., Liedorp, F.: Interbank contagion in the dutch banking sector. *International Journal of Central Banking* **2**, 99–134 (2006)
15. Mastromatteo, I., Zarinelli, E., Marsili, M.: Reconstruction of financial network for robust estimation of systemic risk. Arxiv preprint arXiv:1109.6210 (2011)
16. Park, J., Newman, M.: Statistical mechanics of networks. *Physical Review E* **70**(6), 066,117 (2004)
17. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. *Physical review letters* **86**(14), 3200–3203 (2001)
18. Squartini, T., Garlaschelli, D.: Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* **13**, 083,001 (2011)