

Bargigli, Leonardo; Gallegati, Mauro

**Article**

## Finding communities in credit networks

Economics: The Open-Access, Open-Assessment E-Journal

**Provided in Cooperation with:**

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

*Suggested Citation:* Bargigli, Leonardo; Gallegati, Mauro (2013) : Finding communities in credit networks, Economics: The Open-Access, Open-Assessment E-Journal, ISSN 1864-6042, Kiel Institute for the World Economy (IfW), Kiel, Vol. 7, Iss. 2013-17, pp. 1-39, <https://doi.org/10.5018/economics-ejournal.ja.2013-17>

This Version is available at:

<https://hdl.handle.net/10419/73462>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/3.0/>

# Finding Communities in Credit Networks

*Leonardo Bargigli and Mauro Gallegati*

## Abstract

In this paper the authors focus on credit connections as a potential source of systemic risk. In particular, they seek to answer the following question: how do we find densely connected subsets of nodes within a credit network? The question is relevant for policy, since these subsets are likely to channel any shock affecting the network. As it turns out, a reliable answer can be obtained with the aid of complex network theory. In particular, the authors show how it is possible to take advantage of the "community detection" network literature. The proposed answer entails two subsequent steps. Firstly, the authors verify the hypothesis that the network under study truly has communities. Secondly, they devise a reliable algorithm to find those communities. In order to be sure that a given algorithm works, they test it over a sample of random benchmark networks with known communities. To overcome the limitation of existing benchmarks, the authors introduce a new model and test alternative algorithms, obtaining very good results with an adapted spectral decomposition method. To illustrate this method they provide a community description of the Japanese bank-firm credit network, getting evidence of a strengthening of communities over time and finding support for the well-known Japanese main "bank" system. Thus, the authors find comfort both from simulations and from real data on the possibility to apply community detection methods to credit markets. They believe that this method can fruitfully complement the study of contagious defaults. Since network risk depends crucially on community structure, their results suggest that policy maker should identify systemically important communities, i.e. those able extend the initial shock to the entire system.

Published in Special Issue [Coping with Systemic Risk](#)

**JEL** C49 C63 D85 E51 G21

**Keywords** Credit networks; communities; contagion, systemic risk

## Authors

*Leonardo Bargigli*, ✉ Scuola Normale Superiore, Pisa, Italy, [leonardo.bargigli@gmail.com](mailto:leonardo.bargigli@gmail.com)  
*Mauro Gallegati*, Università Politecnica delle Marche, Dipartimento di Scienze Economiche e Sociali, Ancona, Italy

**Citation** Leonardo Bargigli and Mauro Gallegati (2013). Finding Communities in Credit Networks. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 7, 2013-17. <http://dx.doi.org/10.5018/economics-ejournal.ja.2013-17>

Received August 13, 2012 Published as Economics Discussion Paper August 30, 2012

Revised March 8, 2013 Accepted April 9, 2013 Published April 18, 2013

© Author(s) 2013. Licensed under the [Creative Commons License - Attribution 3.0](#)

## 1 Introduction

Since the outbreak of the global crisis, policy makers have been haunted by the nightmare of a global financial meltdown breaking out of uncontrollable feedbacks spreading across financial markets. In order to avoid this scenario, it became imperative to identify the potential sources of system-wide shocks, i.e. the sources of systemic risk.

Although many transmission channels for systemic risk have been listed by the literature (Brunnermeier, 2009), in this paper we will focus on those determined by the reciprocal claims between institutions over credit markets. The interconnectiveness of credit institutions is a source of counter-party risk on interbank credit markets, which has been addressed recently by a number of theoretical models tackling the problem of contagious defaults (Gai and Kapadia, 2010; Amini et al., 2010, 2012; Battiston et al., 2012). These models, which go beyond previous simulation based works (Nier et al., 2007; Elsinger et al., 2006), rely on complex network theory, which has become a prominent tool in this field. In fact, network theory provides a rigorous representation of complex interactions, as well as the possibility to predict their static and dynamic behavior in many cases.<sup>1</sup>

Regarding the case in point, it's natural to conceive credit markets as networks in which nodes represent agents and links represent credit claims and liabilities. Most works in this field focus specifically on the interbank market, since the latter is particularly relevant for financial stability and, at the same time, well suited for a representation with basic network theory. While earlier contributions (Allen and Gale, 2000) stressed the benefits of increasing diversification, suggesting that the more connections, the better for financial stability, more recently a growing number of works have challenged this view, showing that diversification is not always beneficial for stability. For instance, the recent model of Battiston et al. (2012) shows that, if market-related effects are considered along with credit-related effects by introducing a financial accelerator mechanism, then a potential trade-off between individual risk and systemic risk may exist for increasing connectivity of the network. Similar results are provided by Gai and Kapadia (2010), who show that financial systems exhibit a robust-yet-fragile tendency: while the probability

---

<sup>1</sup> For an economics oriented introduction see Vega-Redondo (2007) or Jackson (2008).

of contagion may be low, once a default cascade is started its spread may be quite large. This effect is non monotonic in connectivity: for a given range of values, connectivity increases the chances that institutions surviving the effects of the initial default will be exposed to more than one defaulting counterpart after the first round of contagion, thus making them more vulnerable to a second-round default.

The empirical support for the relevance of contagious defaults in the interbank market is mixed. This is not surprising at all, since empirical works in this field rely on a variety of simulation based approaches and diverse behavioral assumptions.<sup>2</sup> For instance, those works which examine the effects of idiosyncratic shocks affecting a single bank, come to the conclusion that the scope of contagion is limited (Elsinger et al., 2006; Upper and Worms, 2004; Mistrulli, 2011). By adopting a more realistic setting, e.g. taking into account correlated market shocks and short-term 100% losses for creditors, quite different results have been obtained (Cont et al., 2012). Notwithstanding this uncertainty, central banks are getting more and more interested in network analysis, supporting network-related research and dissemination, although most empirical work in this direction still looks merely descriptive (Soramäki, 2007; Castrén and Kavonius, 2009; ECB, 2010).

In this paper, instead of dealing directly with contagious defaults, we wish to answer the following related question: how do we find densely connected subsets of nodes within a given credit network? As it turns out, we can provide a rigorous answer by drawing from a growing body of literature devoted to community detection in complex networks (Fortunato, 2010). This answer is very important for any analysis of contagious defaults. In fact, the dynamics of any contagion process depends crucially on the topology of the network at the moment of the initial shock. This fact agrees with the following simple intuition: whenever a shock affects a node of a financial network, it will be transmitted to her neighbors with a probability that is proportional to the strength of their linkage to the shocked node. Thus, finding strongly connected subsets of nodes allows us to identify those regions of the network which are most likely to be hit when a specific subset of nodes is shocked initially. In this sense, community detection allows to outline the general “risk maps” invoked by policy makers, and in particular to identify the

---

<sup>2</sup> For a survey see Upper (2011).

systemically important communities of financial institutions, i.e. those which may transmit an initial shock to the entire system.

The rest of the paper is organized as follows. As a first step, we shortly review the related literature on complex networks and community detection (Section 2). Subsequently (Section 3) we focus on a benchmark model endowed with a significant community structure as defined at the end of Section 2. Then we test alternative community detection algorithms over the benchmark model examining both the case of strong, i.e. non overlapping, and weak, overlapping, communities (Section 4). After having verified that a spectral decomposition algorithm provides the best results, we apply the same algorithm to real data in order to provide a community description of the Japanese bank-firm credit network (Section 5). Finally, the conclusive Section summarizes the main results of the paper.

## 2 Related literature

In this paper we adopt a statistical equilibrium approach to the representation of credit markets.<sup>3</sup> Statistical means that the claim  $w_{ij}$  of agent  $i$  towards agent  $j$  is viewed as the realization of a random variable, defined over a discrete nonnegative domain. The market as a whole is nothing more than the collection of all these variables, that can be represented as a random matrix  $W$  with entries which are statistically independent but non necessarily equally distributed. Each realization of  $W$  represents a possible state or configuration of the market, and the collection of all these market states, together with a probability distribution over states  $P(W)$ , is called a *statistical ensemble*. Equilibrium means instead that, if the market is allowed to relax without external disturbances, it will converge to the stable probability distribution  $P^*(W)$  which is obtained by solving the model itself.<sup>4</sup> As explained below, when the model is adequately constrained, it's possible to make

---

<sup>3</sup> For the general idea of using statistical equilibrium as a key tool in macroeconomics see Aoki and Yoshikawa (2006)

<sup>4</sup> There is a bijective relationship between a given model and the corresponding ensemble, since solving a given model means actually to find the particular  $P^*$  which is consistent with the constraints of that model. For this reason the terms “model” and “market ensemble”, “network ensemble” or simply “ensemble” can be used as equivalents.

the ensemble reflect some desirable economic property, like the expected balancing of supply and demand (see Note 7).

Against this backdrop, the transition to complex networks theory appears very natural. In fact, the representation of markets outlined above is nothing different from the matrix representation of a directed network  $G$ , according to which the strength of the link  $i \rightarrow j$  is given by  $w_{ij}$ .<sup>5</sup> In particular, while the interbank market is a simple (weighted and directed) network, credit relationships between banks and firms are represented by a bipartite (weighted and directed) network. The latter is defined in network theory as a network whose nodes can be divided into two disjoint sets  $A$  and  $B$  (called bipartite sets) such that every link connects an element of  $A$  with an element of  $B$ . A matrix representation of a bipartite network is given by a rectangular matrix  $W \in \mathbb{N}_+^{n \times m}$ , where  $n = |A|$  and  $m = |B|$ . The analysis of these networks is addressed either with specific tools or by generalizing the standard tools in various ways. In this paper we follow the latter route, because we wish to provide a general method which may be readily applied both to interbank networks and to bank-firms credit networks. Indeed, most of the tools we provide can be applied to any interaction occurring among economic agents or entities, provided that this interaction can be represented by a nonnegative matrix of arbitrary shape.

As mentioned above, we are interested in network ensembles displaying desirable properties. Park and Newman (2004) have proposed a general methodology for building ensembles of networks, with a fixed number  $n$  of nodes, satisfying linear and non linear constraints over the expected values of network observables:

$$E \left[ \vec{F}(W) \right] = \vec{c} \quad (1)$$

$\vec{F}(W)$  is a function returning a vector of observables, which are dependent on the network state represented by  $W$ , and  $\vec{c}$  is a vector of arbitrary values acting as constraints for those observables. In analogy with equilibrium statistical physics, it's possible to show that  $P^*$  in this case is a Boltzmann-Gibbs probability

---

<sup>5</sup> For this reason, in this paper the terms “market” and “network” are used as equivalents even if, strictly speaking, a market is a directed weighted network. If  $G$  instead is a binary network, its links can take only binary values, and thus its matrix representation is given by the adjacency matrix  $A$  with binary entries.

distribution over network configurations. Since the Boltzmann-Gibbs distribution belongs to the exponential family of probability distributions, networks belonging to the resulting ensemble are labeled as *exponential networks*.

Park and Newman (2004) have provided a solution of this model when the constraints are represented by the strength distribution  $\vec{w}$  of the network.<sup>6</sup> Given our previous market description, these constraints represent in credit networks the expected lending and borrowing of the agents. By allowing for arbitrary strengths, this approach provides a decisive improvement with respect to current economic models, which usually adopt the unrealistic assumption of agents with identical strengths (Allen and Gale, 2000), introducing instead heterogeneity in the modeling of market interactions.<sup>7</sup>

Exponential networks don't represent the only ensemble which may comply with the heterogeneity of strengths. In the already mentioned community detection literature, a different model is usually employed as a benchmark. This model is related to the modularity function  $Q$ , originally introduced by Newman and Girvan (2004). The definition of  $Q$  reads

$$Q(W) \propto \sum_{ij} b_{ij} \delta_{ij} \quad (2)$$

where  $\delta_{ij} = 1$  if  $i$  and  $j$  belong to the same community, and the  $b_{ij}$  are the elements of the modularity matrix  $B = W - E[W]$ . The goal of community detection

<sup>6</sup> By strength of a node  $i$  in a weighted network we define the sum  $w_i = \sum_{j \neq i} w_{ij}$ . If  $W$  is asymmetric, i.e.  $G$  is directed, we need to distinguish between the out-strength and in-strength of the node  $i$ , and thus we have two distinct distributions  $\vec{w}_{out}$  and  $\vec{w}_{in}$ . The degree of a node  $i$ , instead, is defined over the binary matrix  $A$  as the sum  $d_i = \sum_{j \neq i} a_{ij}$ . If  $A$  is asymmetric, again we need to distinguish between the out-degree and in-degree of the node  $i$ . In the case of weighted networks, it's useful to define  $v = \sum_i w_i$ , where  $v$  is said to be the *volume* of  $G$ .

<sup>7</sup> Incidentally, we observe that heterogeneity doesn't necessarily imply disequilibrium in this context. In fact, for any two out- and in-strength distributions we have that  $\sum_i w_i^{out} = \sum_i w_i^{in} = v$ , i.e. that the market is expected to clear.

is to choose the  $\delta_{ij}$  in such a way as to maximize  $Q$ . The elements of  $E[W]$  are defined, in the directed case, as follows:

$$E[w_{ij}] = \frac{w_i^{out} w_j^{in}}{\nu} \quad (3)$$

Squartini and Garlaschelli (2011) have showed that this expectation is different from the one we obtain in the exponential model using the constraints on strength distributions. Alternatively, we introduce a different ensemble whose members can be labeled as *binomial networks* (Bargigli and Gallegati, 2011). In this ensemble the  $w_{ij}$  are binomially distributed with parameters  $\nu$  (see Note 6) and  $p_{ij}$ :

$$w_{ij} \sim B(x|p_{ij}, \nu) = \binom{\nu}{x} p_{ij}^x (1 - p_{ij})^{\nu-x} \quad (4)$$

In the directed case, the parameters  $p_{ij}$  are obtained by solving the following maximum entropy (hereafter denoted as ME) problem:

$$\max_{\mathbf{p}} g(\mathbf{p}) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \ln p_{ij} \quad (5)$$

$$\sum_{j=1}^m p_{ij} = r_i \equiv w_i^{out} / \nu \quad i = 1 \dots n$$

$$\sum_{i=1}^n p_{ij} = c_j \equiv w_j^{in} / \nu \quad j = 1 \dots m$$

$$p_{ij} > 0$$

Using the linear constraints of the problem, we can obtain the following explicit solution (Bargigli and Gallegati, 2011), from which we immediately derive the expectation (3):

$$p_{ij} = \frac{w_i^{out} w_j^{in}}{\nu^2} \quad (6)$$



By construction we have, of course, that  $Q(E[W]) \equiv 0$ . On the other hand, for a binomial network we generally have that  $Q(W) > 0$ , if the  $\delta_{ij}$  are chosen to maximize  $Q$ , since networks in the ensemble are obviously different from  $E[W]$  because of statistical fluctuations. At the same time, it's clear that these fluctuations cannot provide a real community structure. We may systematize these observations with the following three definitions:

- $E[W]$  is the *community free network*
- the communities of networks  $G \in \mathcal{G}$ , where  $\mathcal{G}$  stands for the binomial ensemble, are *statistically non significant*, since they are the outcome of a random fluctuation around  $E[W]$ .
- the communities of networks  $G \notin \mathcal{G}$  are *statistically significant*.

Thus, to be sure that we detect significant communities, we need to verify the null hypothesis  $H_0 : G \in \mathcal{G}$ . In the next section, we show how it is possible to obtain this result.

Finally, we may wonder why we are so interested in an expectation of the form (3). From the very nature of the ME problem (5) we see that (3) provides the maximally diversified expectation which satisfies the constraint over strength distributions. By construction,  $p_{ij}$  is given by the product of  $r_i$  and  $c_j$ , the relative strengths of the agents involved. No special relationships are allowed for. Thus (3) represents a convenient benchmark in so far as diversification is beneficial for the reduction of risk. On the other hand, we need to stress that although binomial networks contain no risky exposures in the sense just explained, they may be nonetheless financially fragile because of an excessive level of lending or borrowing.

Following the definitions given above, communities are implicitly defined as a significant deviation from diversification. As such, they represent an additional source of risk alongside fragility. By comparing a given real network with a sample of binomial networks with the same expected strength distributions, we are able to detect the risk in excess of the risk level observed on a maximally diversified ensemble. In other words, we may define *network risk* as the default rate of a real network which is in excess of the average default rate observed on a sample

of binomial networks with the same strength distribution.<sup>8</sup> From a systemic risk perspective, this is convenient since we focus on the specific network risk of the market under observation *conditioned to* the distribution of debt and credit within the same market.

To sum up: according to our framework, strength distributions represent an independent source of risk, which needs to be clearly separated from network risk itself. By construction, network risk, if it exists, is due to communities. This is the reason why detecting communities becomes important. In particular, it is worth to detect the systemically important communities mentioned in Section 1, since if these are severely affected by a shock, the effects are likely to spill over to other communities.

### 3 The modular binomial model

Artificial networks endowed with a community structure represent a very precious tool for testing community detection algorithms (Lancichinetti and Fortunato, 2009a). In these networks the assignment of nodes to communities is specified in advance by the researcher and the algorithm's task becomes to recover the known community subdivision of agents in the network. Then we can compare how good are alternative algorithms at this job with the help of some suitable measure, like normalized mutual information (see below).

We can induce a community structure in the binomial network of the previous section by fixing the weights between agents belonging to different arbitrary defined communities. Instead of providing different weight values for each combination of communities, like in Bianconi (2009, 2008), we simply add the following constraint to the problem (5):

$$\sum_{ij} (1 - \delta_{ij}) p_{ij} = s \tag{7}$$

---

<sup>8</sup> As it turns out, maximally diversified networks are usually found to underestimate the extent of contagion (van Lelyveld and Liedorp, 2006; Mistrulli, 2011), which means that real networks are indeed an independent source of risk in the sense just explained. But it is also conceivable that real networks are less risky in some cases.

Here  $\delta_{ij}$  becomes a fixed parameter such that  $\delta_{ij} = 1$  if  $i$  and  $j$  are assigned to the same community, and  $s$  is specified in eq. (9) below. By solving for  $p_{ij}$  in the first order conditions we obtain

$$p_{ij} = \exp(\lambda_i + \mu_j + \theta(1 - \delta_{ij})) = x_i y_j z^{1 - \delta_{ij}} \quad (8)$$

Here  $\lambda_i$ ,  $\mu_j$  and  $\theta$  are Lagrange multipliers. From this expression we derive a system of  $n + m + 1$  coupled equations

$$\begin{aligned} x_i \sum_j y_j z^{1 - \delta_{ij}} &= r_i & i = 1 \dots n \\ y_j \sum_i x_i z^{1 - \delta_{ij}} &= c_j & j = 1 \dots m \\ \sum_{ij} (1 - \delta_{ij}) x_i y_j z^{1 - \delta_{ij}} &= s \end{aligned}$$

This system, once solved numerically, delivers what we call the modular binomial model (hereinafter denoted as MB model). In practice, it is convenient to specify  $s$  as follows

$$s = \mu \sum_{ij} (1 - \delta_{ij}) \bar{p}_{ij} \quad (9)$$

Here  $\mu \in (0, 1]$  and  $\bar{p}_{ij}$  is specified as in eq. (6). In this way we assign to inter-modular connections a fraction  $\mu$  of the ME probability obtained in the binomial model. Thus, it becomes possible to produce a continuum of networks ensembles with increasingly significant community structure, ranging from a (significant) community-free ensemble coincident with the binomial model ( $\mu = 1$ ) to a quasi-disconnected union of (significant) communities ( $\mu = \varepsilon$  for a small  $\varepsilon$ ). The results of this procedure for a sample of 100 randomly generated MB networks with  $q = 10$ , for each value of  $\mu$  and different size distributions of communities, are depicted in Figure 1. While agents are randomly assigned to communities, the strength sequence used for the simulations is taken from real data, namely from

the Japanese credit network data analyzed below (see Section 5). In particular, we employ the strength distribution of the 211 Japanese banks recorded during the year 2000 to build a binomial ensemble of symmetric (undirected) networks.

The lines of Figure 1 represent the sample average<sup>9</sup> of the largest singular values of the normalized matrix  $K$ , defined as follows:

$$K = D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \quad (10)$$

Here  $D$  is a diagonal matrix with elements  $\{w_1, w_2, \dots, w_n\}$ .<sup>10</sup> The normalization is operated with respect to the matrix 2-norm, i.e. we have that  $\sigma_0(K) \equiv 1$ , where  $\sigma_0$  is the largest singular value of  $K$ .

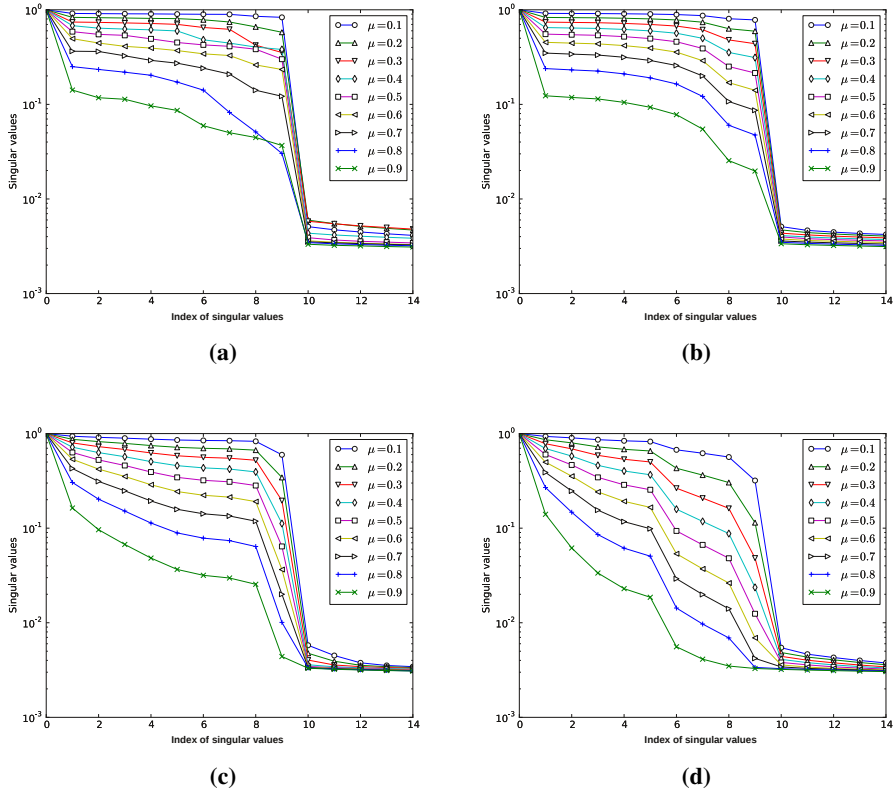
As underlined in many contributions (e.g. Chauhan et al. (2009) or Mitrovi and Tadi (2009)), the spectral properties of different network-related matrices are connected to the community structure of the network itself. In particular, the spectral or singular gap, i.e. the largest difference between eigenvalues or singular values, clearly identifies the number of communities (Ng et al., 2001). We can verify from Figure (1) that in our samples the number of non unitary singular values which are bounded away from zero is equal to  $q - 1$ , as expected, although  $q$  may become harder to detect when the size distribution of communities becomes more uneven (see Panels (c) and (d)), since in this case we have very small communities which are more difficult to identify.

It is instructive to verify the null hypothesis  $H_0 : G \in \mathcal{G}$  for MB networks (see end of Section 2). Of course, we expect  $H_0$  to be rejected if  $\mu$  is not too close to unity. We can employ for this task the statistical test devised in Bargigli and Gallegati (2011):

$$P(\Sigma^2(K) \geq \lambda) \leq \frac{nm}{\lambda v} \quad (11)$$

<sup>9</sup> Regarding the statistics collected from MB simulations, we stress that we observed always a low variability within samples, resulting in very small standard errors. Thus we don't need to build larger samples in order to obtain robust results.

<sup>10</sup> In case of directed networks, normalization becomes  $K = D_{out}^{-\frac{1}{2}}WD_{in}^{-\frac{1}{2}}$  with obvious adaptation of the diagonal matrices.



**Figure 1:** Average singular values of  $K$  for randomly generated MB networks with 10 communities of different size distributions: (a) uniform; (b) power law ( $\alpha = 1.4$ ) with sizes ranging from 10 to 50 nodes; (c) power law ( $\alpha = 1.4$ ) with sizes ranging from 5 to 100 nodes; (d) power law ( $\alpha = 1.4$ ) with unconstrained size

**Table 1:** p-values for the null hypothesis that MB networks belong to the binomial ensemble

$\mu$	max p-value	min p-value	$\mu$	max p-value	min p-value
0.1	0.0010	0.0010	0.91	0.0959	0.0954
0.2	0.0012	0.0012	0.92	0.1211	0.1205
0.3	0.0014	0.0014	0.93	0.1580	0.1569
0.4	0.0020	0.0020	0.94	0.2142	0.2127
0.5	0.0030	0.0030	0.95	0.3066	0.3035
0.6	0.0047	0.0047	0.96	0.4729	0.4672
0.7	0.0085	0.0085	0.97	0.8156	0.8050
0.8	0.0193	0.0193	0.98	0.9999	0.9999
0.9	0.0777	0.0773	0.99	0.9999	0.9999

Here  $v$  is the volume of the network (see Note 6) and  $\Sigma^2(K) = \sum_{ij}(k_{ij} - E[k_{ij}])^2$ . Following the definitions of Section 2, if  $H_0$  is rejected the network contains statistically significant communities. In order to compute the test of eq. (11), we can employ the singular values of  $K$ . In fact, using the Eckart-Young low rank approximation theorem it's possible to show that  $\Sigma^2(K) = \sum_{k=1}^{r-1} \sigma_k^2$ , where the sigmas now stand for the singular values of  $K$  and  $r$  is the rank of  $K$ . In Table (1) we reproduce the maximal and minimal p-value obtained on samples of 100 networks, for each value of  $\mu$ , using eq. (11). The p-values are quite stable within samples, and increasing in  $\mu$  as expected. In particular, the null hypothesis is clearly rejected for  $\mu \leq 0.9$ .

We can employ the same test in order to recover the number of communities in MB networks in a more rigorous way than by visual inspection. The basic idea is simple: if we order the singular modes of  $K$  by the decreasing magnitude of the associated singular values, and subtract them (with the exception of the first one) from  $K$  in the same order, we let  $K$  converge to its expectation under the binomial model.<sup>11</sup> Then, if we compute the test (11) after each subtraction and record the resulting p-value, at some step we obtain that the null hypothesis is not rejected anymore, which means that all the significant communities have been subtracted away from the network.

<sup>11</sup> The singular modes are the terms of the sum  $K = \sum_{i=0}^{r-1} \sigma_i u_i v_i^T$ , where the  $\sigma_i$  are the (decreasing) singular values of  $W$ , and the  $u_i$  and  $v_i$  are the corresponding left and right singular vectors. Under appropriate conditions it's possible to show that  $u_0 v_0^T = E[K]$  (Bargigli and Gallegati, 2011).

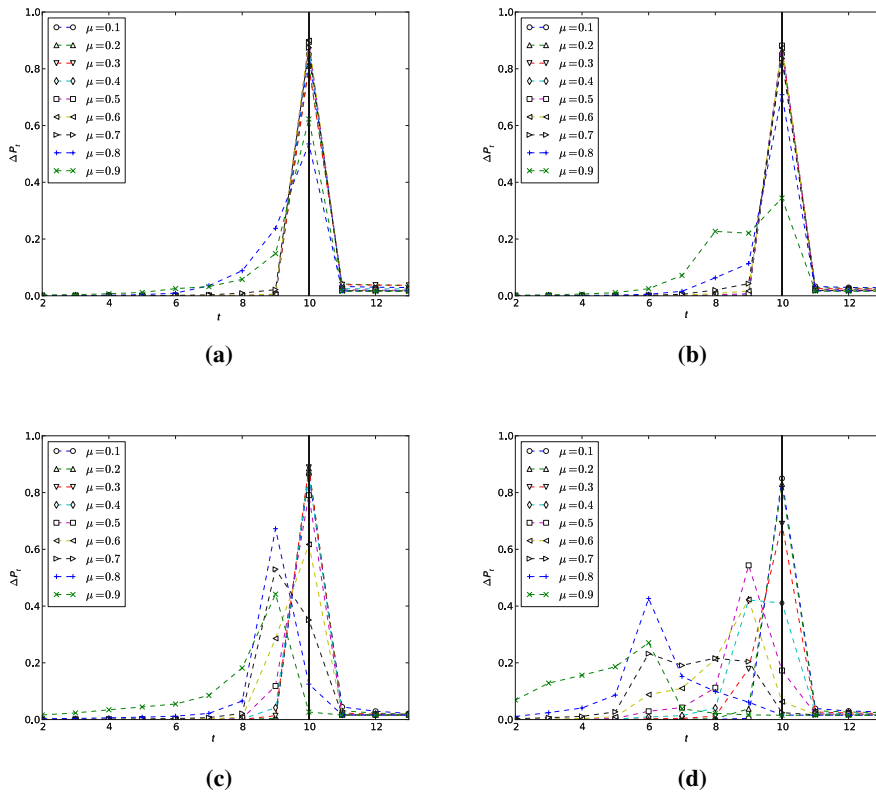
In practice, it turns out that  $q$  is best detected by observing the increments of the p-values, since the maximum of the increments is associated with the singular gap. In order to identify this maximum, it is convenient to introduce the quantity  $\Sigma_t^2(K) = \sum_{k=t}^{r-1} \sigma_k^2$ . Then we compute the following expression iteratively for  $t = 2, \dots, r-1$ :

$$\Delta P_t = \left( \frac{1}{\Sigma_t^2(K)} - \frac{1}{\Sigma_{t-1}^2(K)} \right) \frac{nm}{v} \quad (12)$$

As we see from Figure 2, this method works pretty well when the community structure is clear enough (2a and 2b) and for  $\mu \leq 0.5$ .

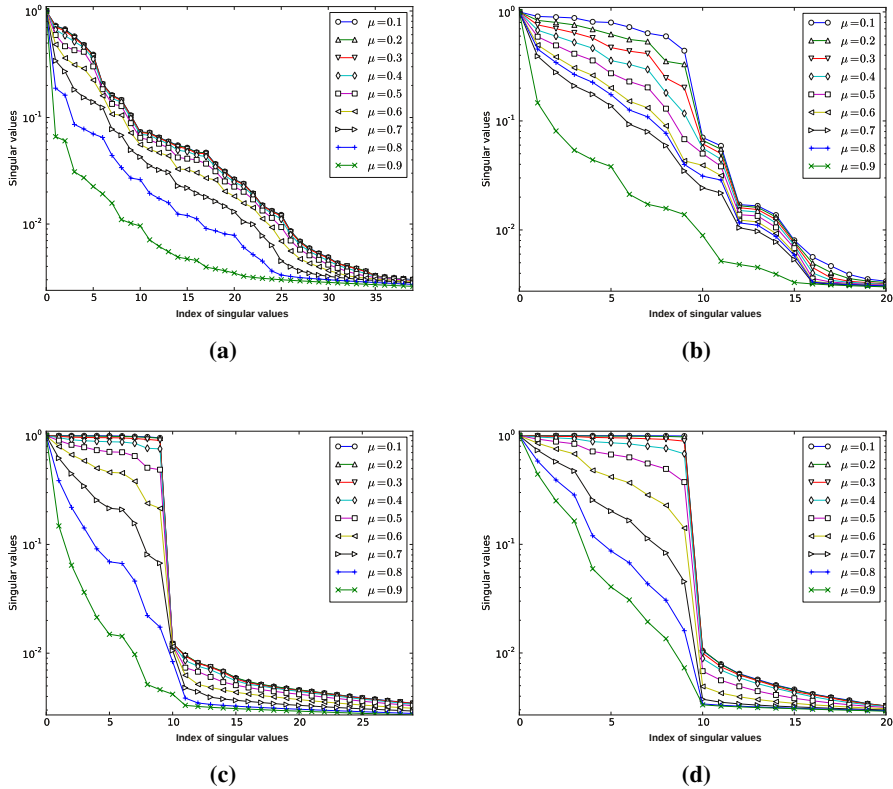
From the analysis of real data (see Section 5), it turns out that the constraint of non-overlapping community membership of nodes may not describe adequately real credit networks. By relaxing this assumption, we introduce *weak* (i.e. overlapping) communities instead of *strong* (i.e. non overlapping) communities. From Figure 3, we see that the behavior of simulated networks changes drastically: in our example, as the fraction  $\beta$  of nodes belonging to 2 different communities increases, the community structure of the network becomes less clear.

From a geometrical point of view, the reason is intuitive. With strong communities  $W$  can be partitioned into orthogonal submatrices, corresponding to the communities, which are associated to clearly separated singular modes. With weak communities orthogonality is violated, and the stronger it is this violation the harder it is to detect clearly separated singular modes. In the limit case when all nodes belong to all communities, we simply recover binomial networks, which don't display significant singular modes. Thus, we may conclude that weak communities can cohabit with a clear community structure only if the fraction of overlapping nodes is low. Not surprisingly, the more blurred structure of weak communities makes it harder to recover  $q$  with the help of eq. (12). From Figure 4 we see that  $q$  is correctly recovered only for low overlapping rates (4c and 4d).

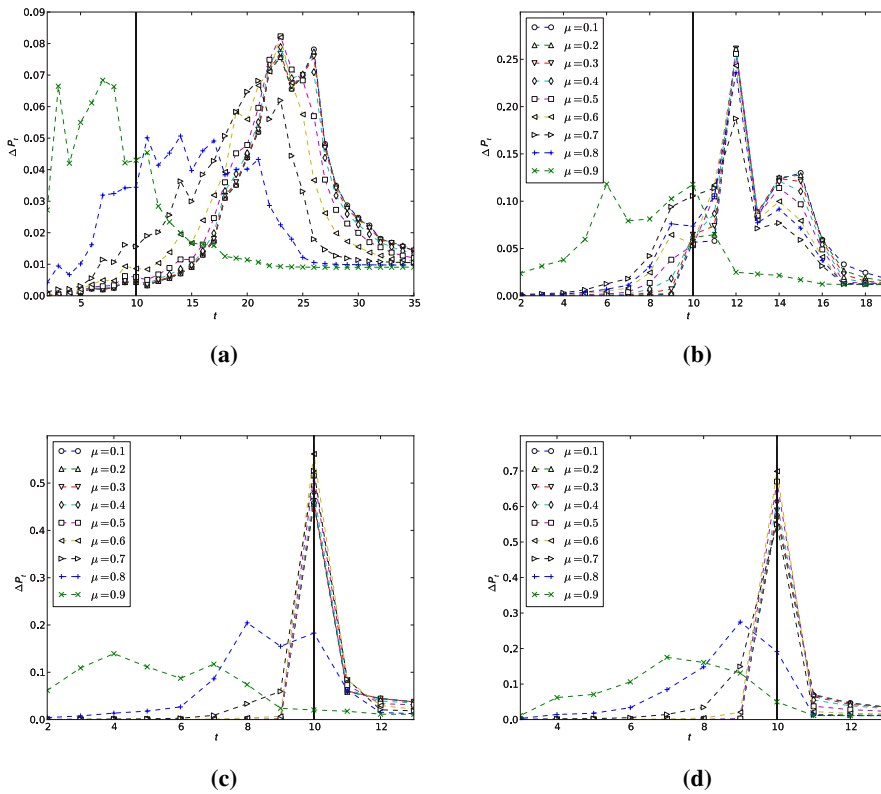


**Figure 2:** Average  $\Delta P$  computed from (12) for MB networks with 10 communities of different size distributions (see Figure 1). The true number of communities  $q$  is signaled by the black vertical line





**Figure 3:** Average singular values of  $K$  for MB networks with 10 bounded (max: 100; min: 5) power-law distributed communities for different overlapping rates: (a)  $\beta = 1$ ; (b)  $\beta = 0.1$ ; (c)  $\beta = 0.05$ ; (d)  $\beta = 0.01$



**Figure 4:** Average  $\Delta P$  computed from (12) for MB networks with 10 bounded (max: 100; min: 5) power-law distributed communities for different overlapping rates (see Figure 3)

## 4 Test of community detection algorithms

As underlined above, the availability of artificial modular benchmarks is of utmost importance for developing reliable community detection methods. In fact, given the complexity of the task, we are never sure of the result we obtain with a given algorithm unless we can trace back this result to a previously known correct answer (Fortunato, 2010). From this perspective, MB networks are particularly effective because they deliver the same features of existing benchmarks (Lancichinetti and Fortunato, 2009a) by allowing for weighted, directed and even bipartite networks, as well as for both strong and weak communities (see Section 3), while they don't require to make hypotheses on the strength distributions. Thus, the confidence in the tests performed on MB networks is bolstered by the fact that we build artificial modular networks which display important properties, such as strength distributions, that are equal on average to those of some real network under study.

In this section we don't aim to perform a systematic comparison of the main community detection procedures, since such comparison is already available in the literature (Lancichinetti and Fortunato, 2009b). Instead, we will focus on two different algorithms. The first one is Infomap (Rosvall and Bergstrom, 2008), which is widely held as the most reliable method since modularity-based methods have been challenged by the discovery of a resolution bias.<sup>12</sup> The second one is an adaptation of the spectral method proposed by Donetti and Muñoz (2005).

The main motivation to perform this comparison is that spectral methods can be applied also to bipartite networks, while this is not possible for Infomap. In fact, this method employs the properties of Markov chains by defining a random walk over a directed network by means of the stochastic matrix  $S$  obtained by suitably normalizing  $W$ . But the convergence of the random walk over a directed network requires the latter to be strongly connected and aperiodic, a condition which, by construction, cannot be fulfilled by bipartite networks. On the other hand, the

---

<sup>12</sup>  $Q$  was introduced with the double purpose to evaluate the assignment of nodes to a given number  $q$  of communities and to provide a criterion for changing  $q$  itself (Fortunato, 2010). The maximization of  $Q$  has become a popular method for community detection, until Fortunato and Barthélemy (2007) have showed that it is plagued by a resolution bias, i.e. it is unable to detect correctly small communities. Because of this limit, modularity maximization is unfit to recover the true  $q$  of artificial benchmarks, and consequently to assign correctly nodes to communities.

available community detection methods for bipartite networks fail to deliver very good results or otherwise are subject to severe computational limits (Sawardecker et al., 2009), so that improvements are strongly needed in this area.

We complement these algorithms with the approach of statistically validated networks (hereinafter denoted as SV networks) proposed by Tumminello et al. (2011). This approach evaluates the probability to observe a given strength value  $w_{ij} = x$ , e.g. in a symmetric network, against the hypergeometric probability distribution:

$$w_{ij} \sim H(x|w_i, w_j, v) = \frac{\binom{w_i}{x} \binom{v-w_i}{w_j-x}}{\binom{v}{w_j}} \quad (13)$$

The intuition is clear: the probability is obtained by enumerating the possible extraction sequences (without replacement) of length  $w_j$  in which we obtain exactly  $x$  successes out of  $w_i$  favorable cases contained in the urn with  $v$  elements. The probability of the event  $\{w_{ij} \geq x\}$  is

$$P\{w_{ij} \geq x\} = 1 - \sum_{w_{ij}=0}^{x-1} H(x|w_i, w_j, v) \quad (14)$$

The link between  $i$  and  $j$  is validated only if it is highly unlikely when evaluated using eq. (14), i.e. when the null hypothesis that  $w_{ij}$  is distributed according to  $H(x|w_i, w_j, v)$  is rejected. The link validation procedure involves multiple hypothesis testing. Therefore the statistical threshold must be corrected for multiple comparisons by employing Bonferroni correction, i. e. by setting the rejection threshold for the null hypothesis at  $\frac{p}{T}$  where  $T$  is the number of hypotheses to be tested ( $T = n^2$  in the case of a network with  $n$  nodes) and  $p$  is the chosen confidence level,<sup>13</sup>. Once the procedure is completed, we obtain a binary, possibly directed, validated network which can be useful to solve the community detection problem

<sup>13</sup> In principle, we can employ other corrections for multiple testing, like FDR (False Discovery Rate). We have verified that our results are unaffected if we employ FDR instead of Bonferroni.

in two ways: firstly, by looking at the components of the SV network itself<sup>14</sup> which may reflect, to some extent, the underlying community assignment; secondly, by letting the algorithms work more effectively than on the original network. Next we are going to pursue both strategies, in order to verify the contribution of this approach over MB networks.

By adopting the hypergeometric distribution the SV approach refers to the so-called configuration model (Molloy and Reed, 1995; Newman et al., 2001), in which the strength distributions are fixed exactly and not on average. In order to adapt this approach to the MB model, we replace the hypergeometric distribution with the binomial distribution (4). If we evaluate the strengths of a MB network using eq. (4), we expect only within-community connections to be validated. Thus, as claimed above, the components of the SV network should reflect communities.

Regarding the spectral algorithm, we introduce the following two modifications with respect to Donetti and Muñoz (2005): 1. the number  $q$  of communities is determined by means of eq. (12) and not by means of modularity optimization; 2. spectral decomposition is performed over  $K$  and not over the Laplacian  $L$ . Except for these differences, we perform the same steps: 1. the eigenvectors associated to the  $q$  largest eigenvalues are computed; 2. each node in the network is treated like a point in the  $q - 1$  - dimensional space with coordinates given by the node's projections onto the first  $q - 1$  nontrivial eigenvectors; 3. hierarchical clustering is applied over the nodes' coordinates in order to obtain  $q$  clusters. In particular, we observed the best performance using complete-linkage clustering over angular distances.

To sum up, the methods employed in the tests are listed as follows: “Infomap” denotes the algorithm introduced in Rosvall and Bergstrom (2008); “Natural partition” denotes the partition corresponding to the components of the SV network; “Spectral” denotes our adaptation of the algorithm introduced in Donetti and Muñoz (2005); the suffix “on SV net” denotes the fact that the (Infomap or spectral) algorithm is applied to the SV network. In order to verify the effectiveness of

---

<sup>14</sup> The components of a network are defined as its maximal connected subsets of nodes.

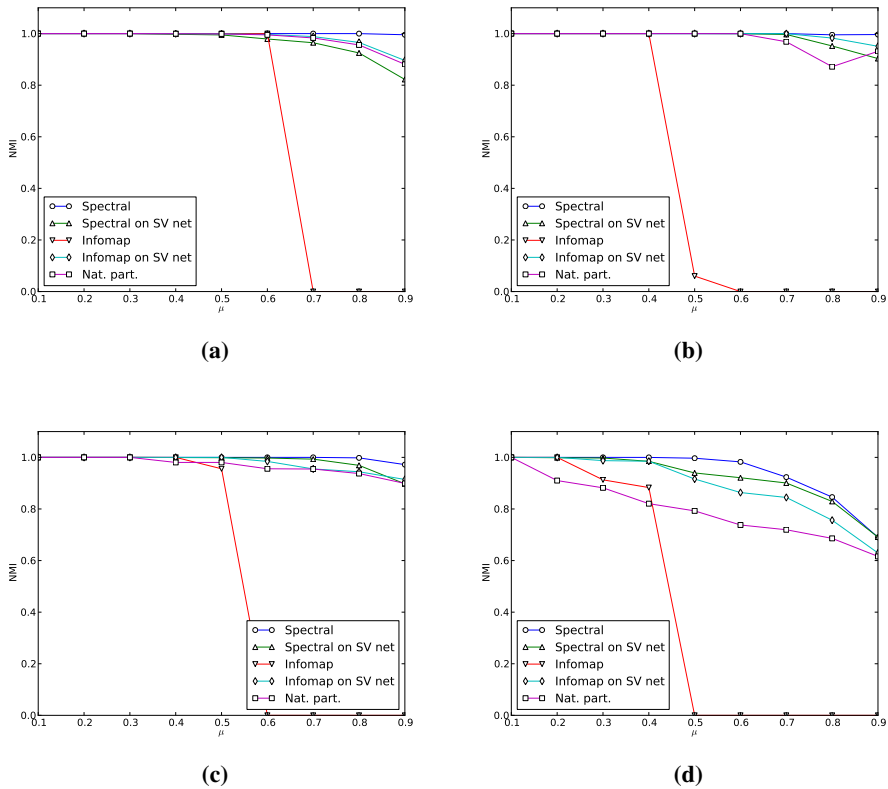
all these algorithms, we adopt the widely used measure of *normalized mutual information* between two random variables (Fortunato, 2010):

$$I_{norm}(X, Y) = \frac{I(X, Y)}{\sqrt{|H(X)H(Y)|}} \quad (15)$$

Here  $H$  stands for the entropy of the random variable and  $I(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$ . In our case,  $I_{norm}$  is used to compare the partition obtained employing some community detection algorithm with the one originally employed to solve the MB problem or, as we can say more shortly, to compare the *recovered* partition  $P$  with the *true* partition  $P^*$ . In practice, for each node the assignment to a community is considered as a random variable taking values in  $\{0, 1, \dots, q - 1\}$ . If the true and recovered partition are equal, we obtain that  $I_{norm}(P, P^*) = 1$ .

The results of the tests performed over MB networks are depicted in Figure 5. The lines represent the average values of  $I_{norm}(P, P^*)$  over samples of 100 artificial networks. Again, we remark that within-sample variability is very low, so that we don't need to build larger samples. Our results provide a unequivocal indication: the best performance is obtained with the spectral algorithm applied to the original network. Further, we observe that SV networks help improve significantly the performance of Infomap for low  $\mu$ , while they generally worsen the performance of the spectral algorithm. Also, the natural partition is showed to represent faithfully the underlying community structure, except for uneven community size distributions (Figure 5d).

In order to extend our testing framework to the case of weak communities, we need to complement our partition methods with a procedure to detect the overlapping nodes between two or more different communities. For this purpose we adapt the method proposed by Lancichinetti et al. (2009), which evaluates the probability that a given node external to a given community has  $k$  links (or equivalently a link with weight  $k$ ) with that community. In particular, the procedure runs as follows: 1. we start with a preliminary partition of the network into  $q$  communities; 2. for each community we compute, either on a binary version of the original network or on the SV network, the number  $k$  of links running between that community and each external node; 3. we evaluate the probability of observing  $k$



**Figure 5:** Tests over samples of 100 MB networks: (a) 422 nodes with 10 communities of uniform size distribution; (b) 211 nodes with 10 communities of uniform size distribution; (c) 211 nodes with 10 bounded (max: 50; min: 10) power-law distributed communities; (d) 211 nodes with 10 bounded (max: 100; min: 5) power-law distributed communities.

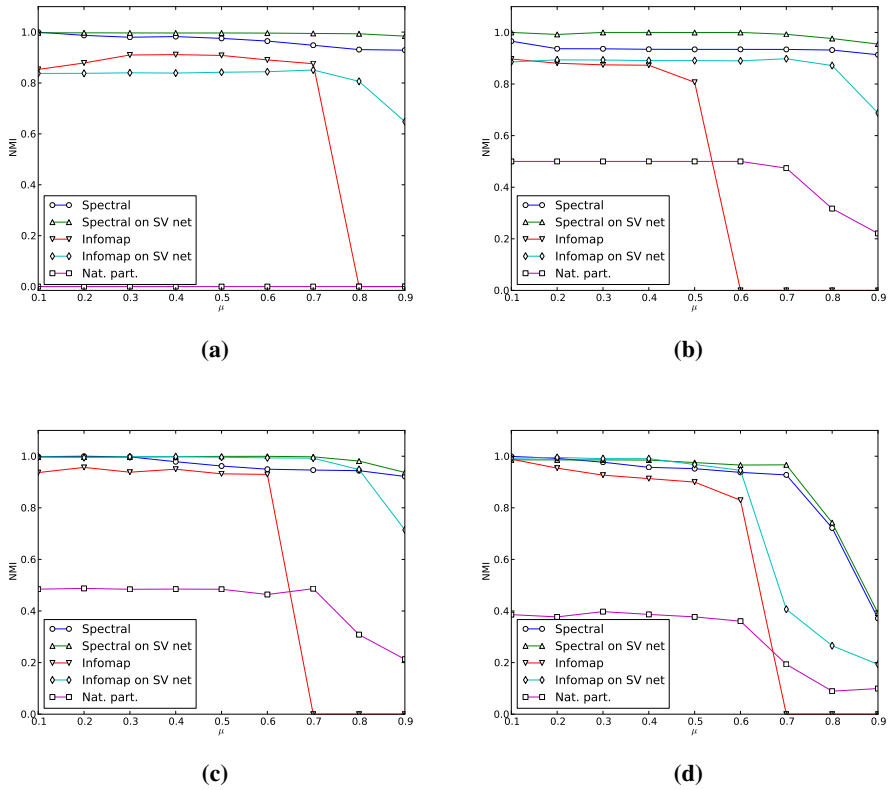
links under the null model using the binomial distribution (4) with suitably adapted parameters; 4. we correct for multiple hypothesis testing by requiring that, in order to validate the inclusion of  $r$  nodes into the  $q$  communities, the overall probability of their links doesn't exceed a threshold  $p$ , i. e. we employ again Bonferroni correction, lowering the acceptance probabilities as  $r$  increases. As a final step, we adapt  $I_{norm}(P, P^*)$  as suggested by Radicchi et al. (2011) in order to compare correctly the resulting assignment.

The results obtained are summarized in Figure 6. In this context the suffix “on SV net”, when referred to the spectral algorithm, has a different meaning than before. Since we know that performing spectral partition over the SV network has a negative effect, we proceed as follows: we apply the spectral algorithm to the original network, while the procedure for the recovery of overlapping nodes is performed either on the binarized version of the original network or on the SV network. Instead, in the case of Infomap, both the partition procedure and the recovery of overlapping nodes are applied either on the former or on the latter. We observe that the best performance is obtained with spectral partition, and particularly when the recovery of overlapping nodes is performed over the SV network.

## 5 Real credit networks

In this section we are going to apply the methods devised above to a dataset describing credit relationships between Japanese firms and banks. This dataset, which has been analyzed under different perspectives in previous works (De Masi et al., 2011; Fujiwara, 2009; Fujiwara and Aoyama, 2008; Bargigli and Gallegati, 2011), includes firms listed in the Japanese stock-exchange markets. Data are compiled from firms' financial statements, integrated by a survey of *Nikkei Media Marketing, Inc.* in Tokyo. They include the indication of the amount of borrowing obtained from each financial institution, subdivided in short-term debt and long-term debt. Financial institutions, which for sake of simplicity are referred to as “banks”, consist of long-term credit banks, city banks, regional banks (primary





**Figure 6:** Tests over samples of 100 MB networks with overlapping rate  $\beta = 0.05$  (see Figure 5).

and secondary), trust banks and insurance companies, all of which represent the universe of financial institutions in Japan.<sup>15</sup>

**Table 2:** Descriptive statistics of the Dataset

Year	Firms	Banks	Links
2000	2,629	211	27,389
2001	2,714	204	26,597
2002	2,739	202	24,555
2003	2,700	192	22,585
2004	2,700	190	21,919
2005	2,674	182	21,811

We already know that bank-firm relationships are represented by a (weighted and directed) bipartite network, where the strengths are given by the amounts of credit/debt and the bipartite sets are represented by firms  $F$  and banks  $B$  respectively. In this context, a correct assignment requires that each community contains at least one element from each bipartite set. In fact, two firms (banks) may belong to the same community only if both are strongly connected to the same banks (firms). Thus, the smallest bipartite community is made of a single firm and a single bank. For this reason, we follow Barber (2007) in classifying both bipartite sets simultaneously.

In our analysis, we work directly with  $K$ , as defined in Section 3, which is of shape  $|F| \times |B|$  and has entries equal to the normalized amounts of debt/credit between each firm  $j \in F$  and bank  $i \in B$ . Since  $K$  is rectangular, we employ its singular vectors as inputs for the spectral decomposition method described in section 4. We resort to spectral decomposition since, as explained in the same section, Infomap cannot be applied to bipartite networks. Finally, we detect multi-community nodes by means of the probabilistic procedure described in the final part of the previous section.

<sup>15</sup> For a more extensive description of this network the reader can refer to De Masi et al. (2011).

The first step of the analysis is to determine the number of communities by means of the procedure based on eq. (12). In Figure 7 we draw the squared singular values and  $\Delta P$  for total credit data as we did for artificial networks (we obtained equivalent graphs for partial data). We observe that the tail of the ranked singular values line is regular like in Figure 1. This fact suggests that Japanese real networks have a clear community structure, with a very limited overlap of communities. On the other hand, the p-values grow in a very regular way and display a very clear peak, as we observed in Figure 2. Thus the number of communities can be safely detected from singular values according to our previous tests.

By comparing the complete results displayed in Table 3 with Table 2, we see that the number of communities tracks quite closely the number of banks in the dataset. This suggests that, while Japanese quoted firms may entertain relationships with a number of banks, at the same time they tend to cluster either around a single bank or, possibly, around a limited number of banks as their “special” partners. This result is consistent with previous analyses performed on the same dataset, as well as with the widely known Japanese “main bank system” (De Masi et al., 2011).

**Table 3:** Number of communities

Year	Short term	Long term	Total
2000	169	192	199
2001	164	184	192
2002	158	182	188
2003	155	175	181
2004	149	174	176
2005	144	171	173

In Tables 4, 5, 6 we report some summary statistics computed as a result of our community detection procedure. From the second line of each table we see that our hypothesis of a limited community overlap is confirmed, as the fraction of multi-community nodes never exceeds significantly the 3% threshold. The large majority of nodes within these subsets are represented by banks, as expected.

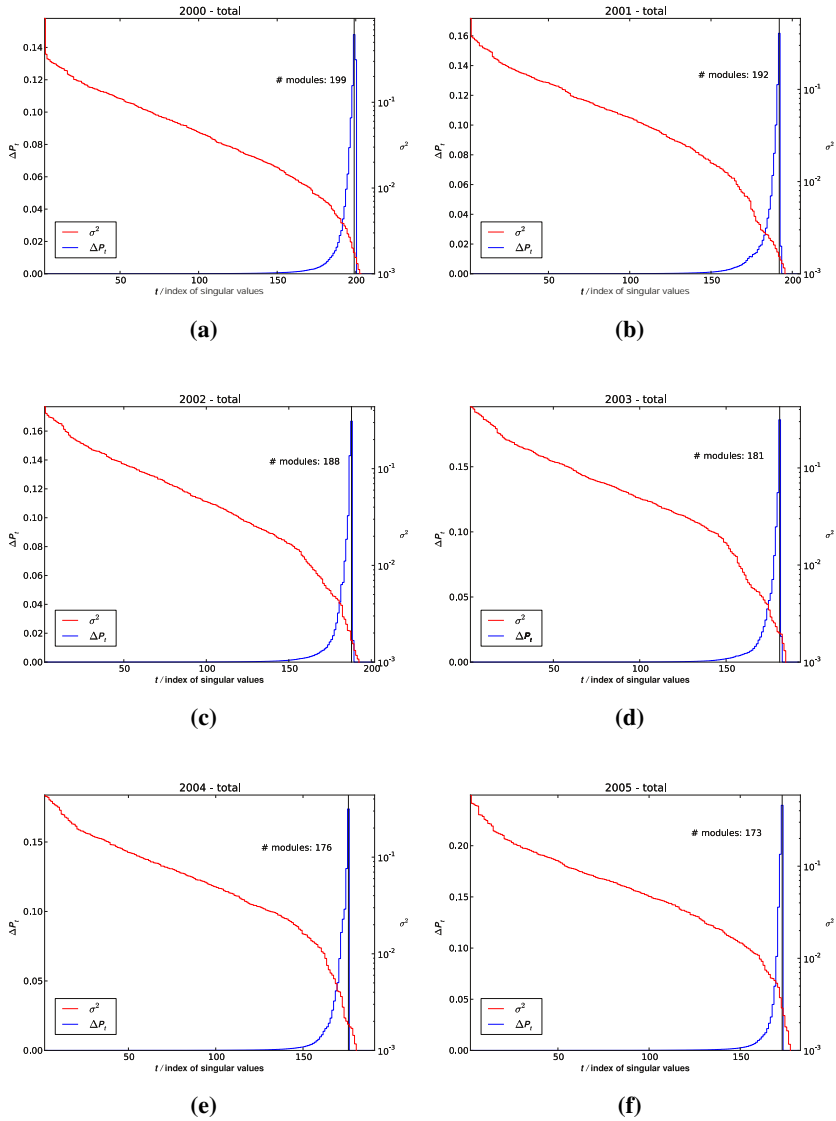


Figure 7: Number of communities in the Japanese credit networks - total credit data.

Further, we see that the total fraction is mostly higher than the partial ones, which would suggest that the overlapping patterns for different maturities tend to be complementary, i.e. multi-community nodes at one level don't act as such at a different level. It's noteworthy that the very limited degree of overlap coexists with high connectivity, since almost all of the nodes in each network are included in the largest component of the SVnet, as we observe from the first line of each table. Since a comprehensive planning of connections at the network level is excluded, here we find a clear sign of self-organization, as nodes behave spontaneously in such a way as to obtain a (quasi) optimal state in which connectivity is maximized while community overlap is minimized.

**Table 4:** Statistics: total data

Stats	2000	2001	2002	2003	2004	2005
Size of largest valid component	2,836	2,914	2,936	2,889	2,882	2,850
Fraction of multi-community nodes	0.0303	0.0260	0.0282	0.0239	0.0242	0.0280
Fraction of internal links	0.1359	0.1423	0.1584	0.1728	0.1736	0.1748
Fraction of internal weight	0.1681	0.1828	0.1905	0.2144	0.2229	0.2103
Mean $Q$ per internal link	2,992.4	3,022.8	2,866.3	2,887.1	2,895.9	2,676.4
Mean $-Q$ per external link	-468.6	-502.3	-540.4	-602.6	-608.4	-566.5

**Table 5:** Statistics: long-term data

Stats	2000	2001	2002	2003	2004	2005
Size of largest valid component	2,372	2,457	2,495	2,467	2,474	2,449
Fraction of multi-community nodes	0.0231	0.0252	0.0244	0.0243	0.0221	0.0232
Fraction of internal links	0.1526	0.1769	0.1907	0.1885	0.2081	0.2020
Fraction of internal weight	0.1748	0.2113	0.2076	0.2237	0.2408	0.2131
Mean $Q$ per internal link	2,497.2	2,514.2	2,399.4	2,641.0	2,491.9	2,320.7
Mean $-Q$ per external link	-450.0	-542.3	-568.8	-613.6	-655.0	-587.3

Links running between nodes in the same community carry only a minor portion of the total weight of the network. This is not surprising since “internal” links themselves represent a minority of total links. The latter is an interesting outcome: the participants of Japanese credit networks are likely to entertain a large

**Table 6:** Statistics: short-term data

Stats	2000	2001	2002	2003	2004	2005
Size of largest valid component	2,564	2,595	2,585	2,503	2,378	2,324
Fraction of multi-community nodes	0.0230	0.0227	0.0224	0.0231	0.0231	0.0236
Fraction of internal links	0.1640	0.1722	0.1920	0.2036	0.2145	0.2228
Fraction of internal weight	0.1914	0.2158	0.2491	0.2984	0.3010	0.2879
Mean $Q$ per internal link	1,788.0	1,811.9	1,859.8	1,989.1	1,708.7	1,544.0
Mean $-Q$ per external link	-350.6	-377.0	-442.9	-509.5	-469.6	-442.6

number of relationship which may be defined as “weak” instead of focusing only on preferential relationships, and cutting all the remaining ties. This behavior is likely to be explained either by an implicit assurance against the eventuality of a disruption of preferential relationships or by complementary factors like, for instance, information gathering (De Masi et al., 2011). At any rate, we can detect a first clue of the effectiveness of our community assignment in the fact that the fraction of internal weight is always higher than the fraction of internal links. A more cogent measure is obtained by rewriting Newman’s modularity as follows:

$$Q = \text{Tr}(D) = \text{Tr}(R^T B S) \quad (16)$$

Here  $R$  and  $S$  are the community matrices obtained for the bipartite sets of firms and banks respectively<sup>16</sup> and  $B$  is the Newman’s modularity matrix (see Section 3). In order to obtain comparable values across different networks, we divide  $Q$  by the number of internal links. Since  $\sum_{ij} d_{ij} = 0$ , the sum of the off-diagonal cells of  $D$  is equal to  $-Q$  by construction. Thus we divide the latter value by the (higher) number of external links in the last line of each table. The resulting values confirm the argument, advanced in Bargigli and Gallegati (2011), that in the Japanese credit network preferential connections with a small number of nodes are traded against the relatively modest downsizing of a large number of non preferential connections, instead of being traded against a reduction in the overall number of links.

<sup>16</sup> A community matrix is a matrix  $M \in \{0, 1\}^{n \times q}$  with elements  $m_{ij}$  such that  $m_{ij} = 1$  if node  $i$  belongs to community  $j$  and  $m_{ij} = 0$  otherwise.

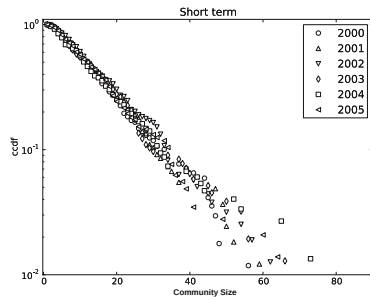
Our community statistics are also informative of the temporal evolution of the networks. For instance, we see that the fraction of internal links and internal weight is increasing over time, especially for short-term data. The tendency towards a more pronounced community structure is consistent with the results of Bargigli and Gallegati (2011), where it was found that the distance between the observed networks and their expected configuration under the binomial model was growing over time. Curiously, this tendency is associated more to a decrease of external  $Q$  than to an increase of internal  $Q$ .

As a final step, we plot the distributions of communities by their size (Figure 8) and of nodes by the number of communities they belong to (Figure 9). From the figures we see no sign of fat tails, since all distributions appear linear on a semilog scale. Regarding community size, we don't find in our data very large communities, although the tail of the distribution tends to become higher over time. This result is consistent with the previous remarks on the high number of external links (which is reflected in the relatively small size of communities), as well as with the evolution towards a more pronounced community structure over time. It's also interesting to observe that very small communities are detected with our procedure (down to the theoretical minimum), so that, as expected, we discover no sign of a resolution bias in our results.

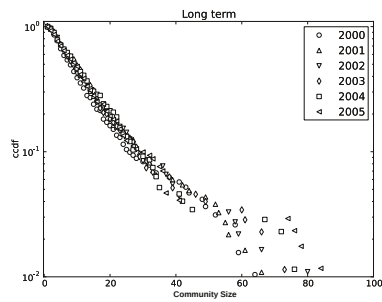
Finally, regarding Figure 9, we remark that some nodes are likely to be part not just of two, but of many communities. The role of these small subsets could turn out to be qualitatively and/or quantitatively relevant within the overall system, especially with respect to the transmission of shocks between communities. For this reason, it might be worth to carry out a more detailed analysis of this phenomenon.

## 6 Conclusions

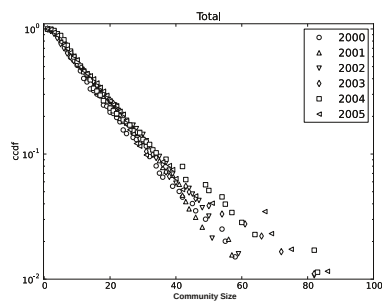
It is worth to summarize the main findings of our analysis. As a first step, we have shortly presented, in Section 2, a variety of techniques for building artificial network models satisfying sets of constraints on network observables. Then, in Section 3 we have applied one of these techniques to obtain ensembles of random networks endowed with a community structure, which we called modular binomial (MB) networks. We showed that it's possible to detect the number of



(a)



(b)



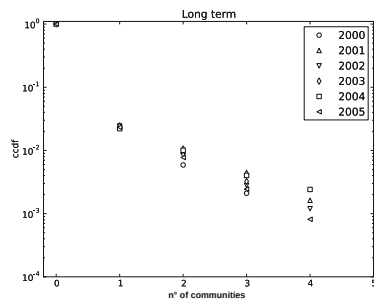
(c)

Figure 8: Community size distribution

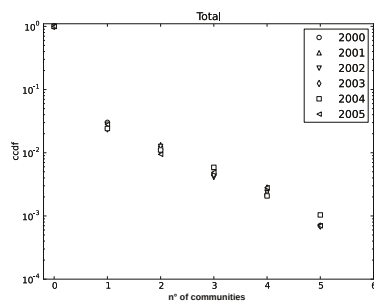




(a)



(b)



(c)

Figure 9: Distribution of nodes by the number of their communities

communities contained in MB networks by looking at their singular values. This result helps us to overcome the limitations of modularity as a device to detect the correct number of communities (Note 12). Instead, in Section 4 we resort to alternative algorithms, such as Infomap and spectral decomposition, complemented by the “Statistically validated networks” approach. Further, since we cannot exclude the possibility that communities in real networks overlap, we introduce the notion of *weak* communities by letting some nodes be part of more than a single community. In order to detect the overlapping nodes, we adopt the solution proposed by Radicchi et al. (2011) of adding single nodes to an initial partition with the help of statistical inference. Both in the case of strong and of weak communities, we obtain the best results by applying the spectral decomposition algorithm to the original (non validated) network.

In order to illustrate our method, in Section 5 we have applied the spectral decomposition algorithm to the Japanese credit dataset. Firstly, we have obtained the number of communities by using singular values, and observed that Japanese firms tend to cluster around a single bank as their major partner, a result which is consistent with the Japanese “main bank” system (De Masi et al., 2011). Secondly, we showed that the real networks under study are highly connected (since the largest valid component spans most of the nodes), while at the same time the fraction of overlapping nodes between communities is quite limited (and mostly represented by banks). Thirdly, we observed that the majority of links is external to communities. Taken together, these findings show that the connectivity of Japanese credit network is much higher of the one we could expect by taking into account only within-community (internal) connections. In other terms, Japanese firms and banks entertain a high number of links which we could define as “weak” since they connect those nodes to other nodes belonging to different communities. Lastly, we confirm the findings of Bargigli and Gallegati (2011) in that we find evidence of a strengthening of the community structure over time, with an increasing number of internal links and a growing size of communities themselves.

Thus, we find comfort both from simulations and from real data regarding the possibility to apply a suitably adapted community detection method to credit markets. As explained in section 2, the existence of statistically significant communities is assessed by comparing real markets with the expectation of the binomial network ensemble. This expectation is maximally diversified under the constraints

over strength distributions, i.e. over the expected lending or borrowing of the agents. Thanks to this property, binomial networks allow us to define, for a real credit market, *network risk* as the default rate in excess of the default rate which is expected in the binomial ensemble with the same strength distributions. The comparison between with real markets has already showed, in some cases, that the latter are indeed a source of network risk (van Lelyveld and Liedorp, 2006; Mistrulli, 2011).

Starting from these premises, we believe that our method can fruitfully complement the study of contagious defaults, which are currently aimed at the identification of systemically important banks (or, more generally, nodes). As explained in Section 3, network risk depends crucially on the community structure of the network. Our results suggest that policy makers should identify systemically important communities, i.e. those able extend the initial shock to the entire system. In this paper we have made the first step in this direction providing a robust algorithm to detect communities in the first place. The next step is to estimate the role and weight of intracommunity contagion under different shock scenarios. While leaving this task for future research, we wish to say a final word of caution with respect to interbank credit markets. In fact, there is evidence that the latter is a tiered system, with a clear distinction between a core and a periphery (Boss et al., 2004; Craig and von Peter, 2010). If this is true, it becomes necessary to introduce the more general notion of role (Reichardt and White, 2007) alongside the notion of community when analyzing these kinds of networks.

**Acknowledgements:** Authors acknowledge the financial support from the European Community Seventh Framework Programme (FP7/2007-2013) under Socio-economic Sciences and Humanities, grant agreement no. 255987 (FOC-II). We thank one anonymous referee for the helpful comments and suggestions. We are especially thankful to Hideaki Aoyama and Yoshi Fujiwara for sharing Japanese network data. The usual disclaimers apply.

## References

- Allen, F., Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33. <http://ideas.repec.org/a/ucp/jpolec/v108y2001i1p1-33.html>
- Amini, H., Cont, R., Minca, A. (2010). Resilience to contagion in financial networks. *SSRN eLibrary*. <http://ideas.repec.org/p/arx/papers/1112.5687.html>
- Amini, H., Cont, R., Minca, A. (2012). Stress testing the resilience of financial networks. *International Journal of Theoretical and Applied Finance*, 15(1). <http://ideas.repec.org/a/wsi/ijtafx/v15y2012i01p1250006-1-1250006-20.html>
- Aoki, M. , Yoshikawa, H. (2006). *Reconstructing Macroeconomics*, Cambridge University Press.
- Barber, M.J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76:066102, December.
- Bargigli, L., Gallegati, M. (2011). Random digraphs with given expected degree sequences: A model for economic networks. *Journal of Economic Behavior & Organization*, 78(3):396–411. <http://ideas.repec.org/a/eee/jeborg/v78y2011i3p396-411.html>
- Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B.C., Stiglitz, J.E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics & Control*, 36(8):1121–1141. <http://dx.doi.org/10.1016/j.jedc.2012.04.001>
- Bianconi, G. (2008). The entropy of randomized network ensembles. *EPL (Europhysics Letters)*, 81(2):28005.
- Bianconi, G. (2009). Entropy of network ensembles. *Physical Review E*, 79(3):036114, March. <http://www.citeulike.org/group/15022/article/4236393>
- Boss, M., Elsinger, H., Summer, M., Thurner, S. (2004). The Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684. <http://ideas.repec.org/a/taf/quantf/v4y2004i6p677-684.html>

- Brunnermeier, M.K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives*, 23(1): 77–100, Winter. <http://ideas.repec.org/a/aea/jecper/v23y2009i1p77-100.html>
- Castrén, O., Kavonius, I.K. (2009). Balance sheet interlinkages and macro-financial risk analysis in the euro area. Working Paper Series 1124, European Central Bank, December. <http://ideas.repec.org/p/ecb/ecbwps/20091124.html>
- Chauhan, S., Girvan, M., Ott, E. (2009). Spectral properties of networks with community structure. *Physical Review E*, 80(5):056114, November. <http://www.citeulike.org/user/dpf/article/6218765>
- Cont, R., Moussa, A., Santos, E.B. (2012). Network Structure and Systemic Risk in Banking Systems, in Fouque, J.P., & Langsam, J. (eds.), *Handbook of Systemic Risk*, Cambridge University Press.
- Craig, B., von Peter, G. (2010). Interbank tiering and money center banks. BIS Working Papers 322, October. <http://www.bis.org/publ/work322.pdf>
- De Masi, G., Fujiwara, Y., Gallegati, M., Greenwald, B.C., Stiglitz, J.E. (2011). An analysis of the Japanese credit network. *Evolutionary and Institutional Economics Review*, 7(2): 209–232. <http://ideas.repec.org/p/arx/papers/0901.2384.html>
- Donetti, L., Muñoz, M. A. (2005). Improved spectral algorithm for the detection of network communities. *AIP Conference Proceedings*, 779:104–107, July. <http://arxiv.org/abs/physics/0504059>
- Elsinger, H., Lehar, A., Summer, M. (2006). Systemically important banks: An analysis for the European banking system. *International Economics and Economic Policy*, 3(1):73–89. <http://ideas.repec.org/a/kap/iecepo/v3y2006i1p73-89.html>
- ECB (European Central Bank) (2010). *Recent advances in modelling systemic risk using network Analysis*, Frankfurt am Main, January 2010. European Central Bank.

- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5):75–174.  
<http://www.sciencedirect.com/science/article/pii/S0370157309002841>
- Fortunato, S., Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41.  
<http://www.jstor.org/stable/25426046>
- Fujiwara, Y., Aoyama, H. (2008). Large-scale structure of a nation-wide production network. *The European Physical Journal B*, 77(4): 565–580.
- Fujiwara, Y., Aoyama, H., Ikeda, Y., Iyetomi, H., Wataru, S. (2009). Structure and temporal change of the credit network between banks and large firms in Japan. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 3, 2009–7.  
<http://dx.doi.org/10.5018/economics-ejournal.ja.2009-7>
- Gai, P., Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 466(2120): 2401–2423.  
<http://rspa.royalsocietypublishing.org/content/466/2120/2401.abstract>
- Jackson, Matthew O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.
- Lancichinetti, A., Fortunato, S. (2009a). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1):016118, July. <http://arxiv.org/abs/0904.3940>
- Lancichinetti, A., Fortunato, S. (2009b). Community detection algorithms: A comparative analysis. *Physical Review E*, 80(5):056117, November.  
<http://arxiv.org/abs/0908.1062>
- Lancichinetti, A., Fortunato, S., Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015. <http://arxiv.org/abs/0802.1218>

- Mistrulli, P.E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127. <http://ideas.repec.org/a/eee/jbfina/v35y2011i5p1114-1127.html>
- Mitrović, M., Tadić, B. (2009). Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. *Physical Review E*, 80(2): 026123, August. <http://arxiv.org/abs/0809.4850>
- Molloy, M., Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–179. <http://dl.acm.org/citation.cfm?id=1943750>
- Newman, M.E.J., Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, February. <http://arxiv.org/abs/cond-mat/0308217>
- Newman, M. E. J., Strogatz, S. H., Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 026118. <http://arxiv.org/abs/cond-mat/0007235/>
- Ng, A. Y., Jordan, M. I., Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, MIT Press.
- Nier, E., Yang, J., Yorulmazer, T., Alentorn, A. (2007). Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060. <http://ideas.repec.org/a/eee/dyncon/v31y2007i6p2033-2060.html>
- Park, J., Newman, M. E. J. (2004). Statistical mechanics of networks. *Phys. Rev. E*, 70:066117. <http://arxiv.org/abs/cond-mat/0405566>
- Radicchi, F., Lancichinetti, A., Ramasco, J., Fortunato, S. (2011). Finding statistically significant communities in networks. *PLoS ONE*, 6(4):e18961, 04. <http://arxiv.org/abs/1012.2363>
- Reichardt, J., White, D. R. (2007). Role models for complex networks. *European Physical Journal B*, 60(2), 217-224. <http://arxiv.org/abs/0708.0958>

- Rosvall, M., Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123. <http://www.jstor.org/stable/25451237>
- Sawardecker, E. N., Amundsen, C. A., Sales-Pardo, M., Amaral, L. A.N. (2009). Comparison of methods for the detection of node group membership in bipartite networks. *European Physical Journal B*, 72(4):671–677, 2009. <http://ideas.repec.org/a/spr/eurphb/v72y2009i4p671-677.html>
- Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333. <http://adsabs.harvard.edu/abs/2007PhyA..379..317S>
- Squartini, T. Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):083001, 2011. <http://arxiv.org/abs/1103.0701>
- Tumminello, M., Miccichè, S., Lillo, F., Piilo, J., Mantegna, R. N., Ben-Jacob, E. (2011). Statistically Validated Networks in Bipartite Complex Systems. *PLoS ONE*, 6, March. <http://arxiv.org/abs/1008.1414>
- Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111 – 125. <http://ideas.repec.org/a/eee/finsta/v7y2011i3p111-125.html>
- Upper, C., Worms, A. (2004). Estimating bilateral exposures in the german interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849. <http://ideas.repec.org/a/eee/eecrev/v48y2004i4p827-849.html>
- van Lelyveld, I., Liedorp, F. (2006). Interbank contagion in the dutch banking sector: A sensitivity analysis. *International Journal of Central Banking*, 2(2), May. <http://ideas.repec.org/a/ijc/ijcjou/y2006q2a4.html>
- Vega-Redondo, F. (2007). *Complex Social Networks*. Cambridge: Cambridge University Press.



Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

<http://dx.doi.org/10.5018/economics-ejournal.ja.2013-17>

The Editor