Fitness model for the Italian interbank money market

G. De Masi,[1] G. Iori,[2] and G. Caldarelli[3]
[1]*Dipartimento di Fisica, Università di L'Aquila, Via Vetoio, 67010 Coppito (AQ), Italy and Dipartimento di Fisica, Università di Roma "La Sapienza," Piazzale Moro 5, 00185 Rome, Italy*
[2]*Department of Economics, City University, Northampton Square, EC1 V 0HB London, United Kingdom*
[3]*INFM-CNR Centro SMC and Dipartimento di Fisica, Università di Roma "La Sapienza," Piazzale Moro 5, 00185 Rome, Italy, and Centro Studi e Museo della Fisica Enrico Fermi, Compendio Viminale, 00185 Rome, Italy*
(Received 22 May 2006; published 21 December 2006)

We use the theory of complex networks in order to quantitatively characterize the formation of communities in a particular financial market. The system is composed by different banks exchanging on a daily basis loans and debts of liquidity. Through topological analysis and by means of a model of network growth we can determine the formation of different group of banks characterized by different business strategy. The model based on Pareto's law makes no use of growth or preferential attachment and it reproduces correctly all the various statistical properties of the system. We believe that this network modeling of the market could be an efficient way to evaluate the impact of different policies in the market of liquidity.

DOI: 10.1103/PhysRevE.74.066112 PACS number(s): 89.65.Gh, 02.50.−r, 05.45.Df

Coevolution and interaction between different agents is known to be one of the ingredients of the so-called complex systems. Several examples can be found in social [1,2], biological [3–6], economical [7], and technological systems [8]. Any of these systems is composed by a set of agents competing and sometimes receiving reciprocal advantage interacting each other. In the above situation both coalition and competition are at the basis of the process of co-evolution and self-organization of the system. While this class of problems has been traditionally studied in game theory, more recently it has been introduced an approach based on graph theory [9,10] By using networks [11,12], we can characterize quantitatively the interaction between agents by means of a series of topological quantities. The case of study presented here is composed by banks operating in the Italian market [13]. Banks try to maximize their returns given some constraints from the European Central Bank. This complex interaction results in a differentiation of the strategies that is well described by means of graph cliques. More specifically banks of the same size tend to form a cluster and to adopt a similar business strategy.

A network is a mathematical object composed by vertices and edges joining them. Different measures can be made, from the degree distribution (the degree is the number of edges per vertex) to the diameter (i.e., the maximum of the distances between every couple of vertices). It is interesting to note that different real world networks (ranging from social to biological ones), display a scale-free distribution of degrees and a "small-world" character, that is to say the diameter is usually very small [14]. More complicated measures determine also the presence of communities in a network. In this case, some methods have been proposed [15–17] but no general approach is available. The set of banks with their internal loans and debts has a structure than can be naturally described by means of a network. In this case the vertices are the different banks. For every pair of banks i and j we draw an oriented edge from i to j, if bank j borrows liquidity from bank i. The number of incoming and outgoing edges of a vertex is called, respectively, the in degree k_{in} and the out degree k_{out} of the vertex (their sum gives the total degree k). The loans are originated by the fact that every bank needs liquidity in order to satisfy demands of customers. To buffer liquidity shocks the European Central Bank requires that on average 2% of all deposits and debts owned by banks are stored in national central banks. Given this constraint, banks can exchange excess reserves on the interbank market with the objective to satisfy the reserve requirement and in order to minimize the reserve implicit costs [18–20]. The data set analyzed is the electronic broker market for interbank deposits (e-MID) (reference dataset) [21]. This data set is composed by 586 007 overnight transactions (i.e., payments of loans must be done in 24 hours) concluded from January 1, 1999 to December 31, 2002. The network is composed by a set of N banks (the average number of $\langle N \rangle$ banks daily active is 140) connected by an average number of links $\langle L \rangle = 200$ (in case of multiple transactions among banks i and j, we count just one link). As in many other complex networks we find here a fat tail distribution. By fitting these data with a power law we obtain for the total degree a frequency distribution $F(k) \propto k^{-2.3}$ and a similar behavior for the in/out degree with exponents $F(k_{\text{in}}) \propto k_{\text{in}}^{-2.7}$ and $F(k_{\text{out}}) \propto k_{\text{out}}^{-2.15}$. Regardless the precise form of the fit, the fat tail indicates that banks have a highly heterogeneous behavior, since the number of their partners varies very widely. We also measure the assortativity and the clustering coefficient of the network. The first one is defined as the average value $K_{nn}(k)$ of the neighbors of a vertex whose degree is k. We find $K_{nn}(k) \propto k^{-0.5}$. This means that banks with few partners interact with banks with many partners. Conversely (on average) banks with many partners interact with banks with few or one. The clustering coefficient instead accounts for the number of triangles a vertex of degree k belongs to. Also this quantity has a power law behavior of the kind $c(k) \propto k^{-0.8}$. All these measurements refer to daily networks resulting from composing all transactions of every day. In fact, the system is characterized by a typical time scale of the system, the month. This time scale arises from the above-mentioned requirement from European Central Bank. The 2% to be deposited in national central banks are computed

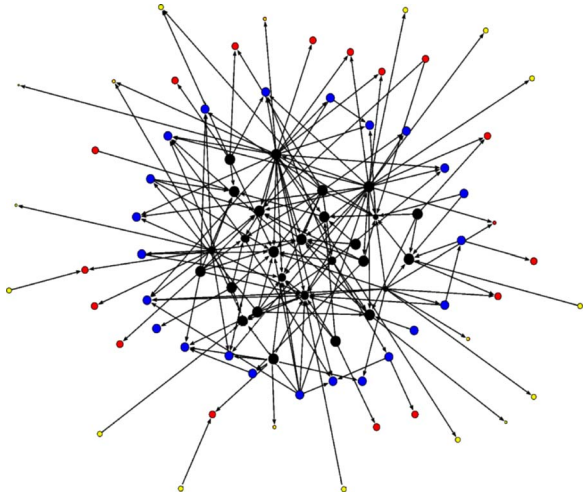

FIG. 1. (Color online) A plot of the interbank network. The group of the vertices (banks) goes from light (group of small banks) to dark (large banks) (the color codes for the various groups are the following: 1=yellow, 2=red, 3=blue, 4=black). Note that the darkest vertices (bank of group 4) form the core of the system.

every month (on the 23rd day). The day in which this happens (also indicated as end of month or EOM) witnesses a frantic activity of the banks. Interestingly, regardless of the change in volumes all the above topological measurements remain similar when computed in different days of the month.

We try to understand if there are some banks with similar behavior and if they have some properties in common. We have been able to identify specific features for banks of different capital size. In fact for each bank we know only its category (small, medium, large, very large) based on the capital of the banks (as recorded by Bank of Italy). Nevertheless we observe that this classification is strongly correlated with the total amount of daily volume of transactions: we use this latter quantity as it is strictly related to capital size. Using this quantity we can divide banks into four groups (same number of classes of the Bank of Italy classification): *Group 1* with volume in the range 0–23 million Euro per day, *group 2* in the range 23–70 million Euro per day, *group 3* in the range 70–165 million Euro per day, *group 4* over 165 million Euro per day. In this way we find an overlap of more than 90% between the two classifications.

Using this information we realized a picture of the system as an oriented network whose size and color of the vertices represent the different groups that play the role of communities when described by means of a network. As evident from Fig. 1 we find that the core of the structure is composed by banks of the last groups (very large). The edges in Fig. 1 represent the net amount of money exchanged in a whole day. As mentioned above the measurements in different days give similar results. A more quantitative measure of the different behavior of banks from different groups is given in Table I, where for every pair of groups we reported the mean percentage of the total number of transactions between banks of those groups. This result is confirmed by the first two plots of Fig. 2, where we represented in-degree frequency distribution (number of borrowing edges) and the out-degree

TABLE I. The number of daily interactions between the banks of different groups. Data have been averaged during 1 month.

Group	1	2	3	4
1	0	6	4	8
2	6	3	8	17
3	4	8	5	27
4	8	17	27	22

frequency distribution (number of lending edges) in the network (experimental distributions are obtained on an ensemble of daily networks). It is possible to compute the group of the banks whose degree is k. We represented this information by coloring accordingly the plot. We have separated informations about degree and volumes of different banks. Interestingly we note that the degree and the volume are correlated [22], since $v(k) \sim k^{1.1}$ (see Fig. 3).

With respect to the scale of colors in Fig. 1, we also added some intermediate colors to account for the values between one group and another. The tail of the two distributions is black, i.e., it is mainly composed by banks of group 4. We again find that banks of groups 1 and 2 are the leaves of the network, staying at periphery of the structure and not interacting with each other. This particularity together with the experimental evidence that they are more lenders on average means that banks of these groups are the lenders for the whole system.

The role of the different groups is shown in Fig. 4. Another measure of the clustering of banks in different groups is given by the volume-volume correlation $v_{nn}(v)$, that is the average value v_{nn} of the neighbors of a vertex whose volume is v, In fact we find that $v_{nn}(v)$ is the superposition of a

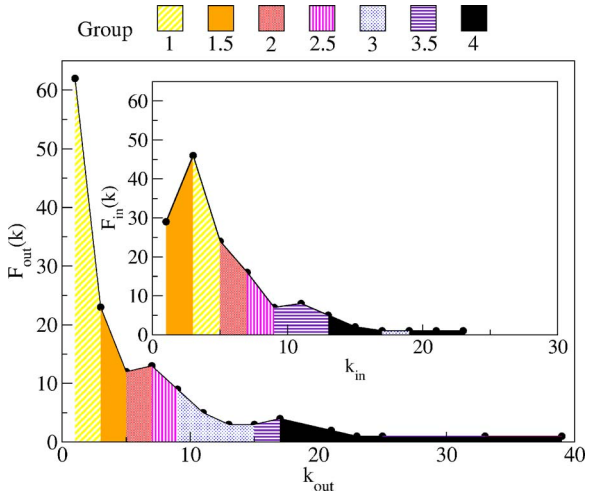

FIG. 2. (Color online) A plot of the out-degree and in-degree (in the inset) distributions, respectively. As already noticed, the contribution to the tail of frequency distribution emerges from the banks of group 4. Using the division in four groups, we determine the average group of each bin of $F(k)$. Every bin is then represented with a particular color and/or texture according to the value of the group obtained. For noninteger value of this average we introduced intermediate colors.

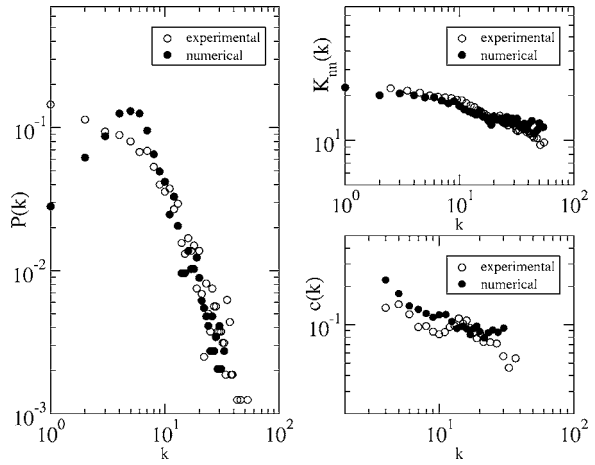

FIG. 3. Left-hand side: Probability distribution $P(k)$ for the degree k. With empty circles we have the experimental results to be compared with simulation of our model (filled circles). Right-hand side: Above comparison between experiment (empty circles) and results obtained with simulation of our model (filled circles) for the assortativity $\langle K_{nn}(k)\rangle$ and below for the clustering coefficient $c(k)$.

power-law function $v_{nn}(v) \sim v^{-0.3}$ with a function peaked around volume values of banks of group 1.

In order to reproduce the topological properties we define a model whose only assumption is that a vertex is solely determined by its size (as measured by its capital or equivalently by its group). Therefore, the idea is that the vertices representing the banks are defined by means of an intrinsic character corresponding to the size of the bank [23,24]. Since this information is not available we use the total daily volume of transactions as a good measure of the size of banks (we stated above that this is a good approximation). We call this quantity *fitness* of the bank; this is the main quantity driving the network formation in our model.

Following Pareto's law (confirmed in this data analysis) we assume that the distribution of sizes v in the model is a power law $P(v) \propto v^{-2}$, where the value of the exponent correspond to that of the data (see Fig. 5).

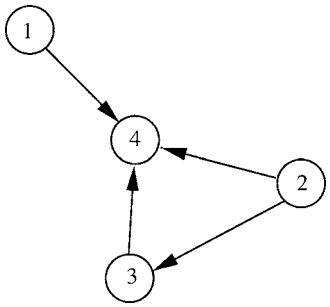

FIG. 4. The division on classes of vertices permits to represent in a very easy way the organizational principles of the network. Following results of Table I we draw a link among two groups when the number of links between banks belonging to them is bigger than the average value. Using the net volumes as weight of links, we can represent the directed interactions among classes of nodes: class 4 appear to be clearly a borrower and class 1 is a lender.

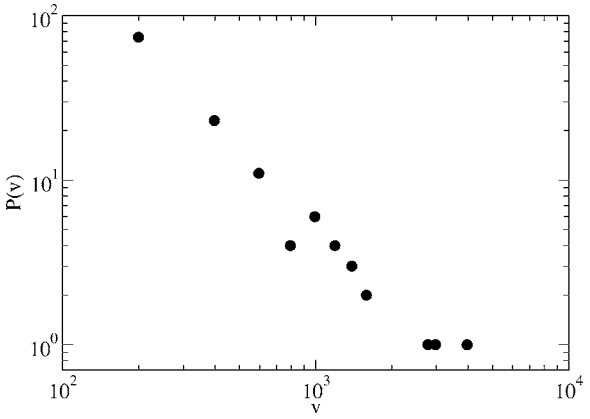

FIG. 5. Distribution of the total daily volume of transaction per bank. This quantity is used as fitness in our model.

We assign to the N nodes (N is the size of the system) a value drawn from the previous distribution. Vertices origin and destination for one edge are chosen with a probability p_{ij} proportional to the sum of respective sizes v_i and v_j. In formulas

$$p_{ij} = \frac{(v_i + v_j)}{\sum_{i,j>i} (v_i + v_j)}, \tag{1}$$

$$\sum_{i,j>i} (v_i + v_j) = \frac{1}{2} \sum_{i,j\neq i} (v_i + v_j) = (N-1) V_{\text{tot}}, \tag{2}$$

where

$$V_{\text{tot}} = \sum_j v_j. \tag{3}$$

We obtain in this way $p_{ij} = \frac{v_i + v_j}{(N-1)V_{\text{tot}}}$. This choice of probability reproduces the fact that big banks are privileged in transactions among themselves while two little banks are very unlikely to interact. We produce an ensemble of 100 statistical realizations of the model and then we calculate average statistical distributions. In Fig. 2 we compare experimental and simulated $P(k)$, $c(k)$, and $k_{nn}(k)$: here the distributions are also averaged on all EOM days of 2002. The simulation of the model reproduces remarkably well the considered topological properties of the interbank market $P(k)$, $c(k)$, and $k_{nn}(k)$. The real and simulated networks disclose disassortative behavior: this phenomenon has already been observed in other systems and it has been called *rich club phenomenon*, referring to the fact that in many real networks hubs are often connected each other[2]. Fitness models on the other hand are known to produce disassortative networks, even if with different fitness distributions [24].

It is interesting to note that this model does not consider preferential attachment rules. With the term "preferential attachment" it is indicated a specific procedure in which a vertex receives more edges according to the value of its degree. Note that this procedure must be very precise because if the probability of growth is proportional to the degree raised to a power different from 1, the scale invariance is

destroyed. Therefore, preferential attachment has a precise definition different from "rough proportionality." When considering instead a fitness algorithm, it is true that the largest the fitness the largest the degree, but the microscopic procedure is different. A large degree is a consequence of an intrinsic quality, not the cause of the improvement of site connectivity. This is an important point since in this way the search for the origin of scale invariance in networks can be explained by means of the ubiquitous presence of Pareto's law in economics and finance.

To quantify the agreement between experimental and simulated networks we also define an overlap parameter m specifying how good is the behavior of the model in reproducing the observed clustering.

To quantify the agreement between experimental and simulated networks, we proceed in the following way. We define a matrix E, that is a weighted matrix 4×4, where the weights represent the number of connections between groups. In order to measure the overlap between the matrices obtained by data and by computer model, we define a distance based on the differences between the elements of the matrices,

$$d=\sum_{g,k\geqslant g}|E_{g,k}^{\mathrm{expt}}-E_{g,k}^{\mathrm{num}}|. \tag{4}$$

The sum of all elements, $\Sigma_{g,k\geqslant g}E_{g,k}^{\mathrm{expt}}$ and $\Sigma_{g,k\geqslant g}E_{g,k}^{\mathrm{num}}$, is equal to E_{tot} in both cases. Therefore the maximum possible difference is $2E_{\mathrm{tot}}$. This happens when all the links are between two groups in one case and in other two groups in the other. We use this maximum value to normalize the above expression and we than define the overlap parameter m, $m=1-d/2E_{\mathrm{tot}}$.

A natural way to define groups in the model is to obtain a similar number c of banks for each class, i.e., $c=N_{\mathrm{banks}}/N_{\mathrm{classes}}$. It is useful nevertheless to pass to continuous form. Using the previously introduced $P(v)$ giving the probability distribution of the size v of one bank. Banks of the same group g are in the range $[v_g,v_g+\Delta v_g]$,

$$\int_{v_g}^{v_g+\Delta v_g}P(v')dv'=c. \tag{5}$$

In our case, since the average number of banks is 140, we obtain $c\simeq35$. Then $\Delta v=cv^2/(N-cv)$. We now compute the number $E_{g,k}$ of links going from one group of banks g_g to another one g_k, for every possible pair of banks,

$$E_{g,k}=\sum_{i,j}a_{i,j}\delta[g_g-g(i)]\delta[g_k-g(j)], \tag{6}$$

where $g(i)$ represent the group of bank i and $a_{i,j}$ is the element of the adjacency matrix. In the continuous approximation, defining $E_{v'v''}$ the number of edges from vertices of fitness v' to vertices of fitness v'', $E_{g,k}$ is given by

$$\begin{aligned}E_{g,k}&=\int_{v_g}^{v_g+\Delta v_g}\int_{v_k}^{v_k+\Delta v_k}E_{v'v''}dv'dv''\\&=(N/2)\int\int P(v')P(v'')p(v',v'')dv'dv'',\end{aligned} \tag{7}$$

where N is the number of vertices, $p(v',v'')$ is the linking probability, $P(v)$ is the fitness distribution and the formula is obtained integrating the expression for the average degree [24] (the integration domains are the ranges of volumes of groups g and k, respectively). To evaluate the relevance of division in classes, we must compare the value of $E_{g,k}$ with the corresponding quantity $E_{g,k}^{\mathrm{null}}$ for a network where there is not a division in classes (null hypothesis). The analytical expression for the null case is $E_{g,k}^{\mathrm{null}}=E_{\mathrm{tot}}/10$ where 10 is the number of possible couplings between the four groups. The comparison between the two networks evidences that in the real case emerges the division in groups: in Table. I for each possible combination of groups is reported the value $E_{g,k}/E_{\mathrm{tot}}$. In the null case, each element of the same matrix should be equal to 10. In our case the overlap m is very good (98%).

In conclusion we present here a network representation of a financial market that in a natural way allows to measure the presence of clustering. By means of a suitable chosen model of network formation we can also understand the mechanism driving the formation of such clusters. The agreement between the model and experimental results is remarkably good; this seems to suggest that the network formation is not due to the growth mechanism of preferential attachment. Since the effects of European Central Bank policies are under debate [19], graph theory can help to understand the system behavior under change of external conditions.

One of the authors (G.C.) acknowledges support from European Project DELIS.

[1] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, Phys. Rev. E **74**, 036116 (2006).
[2] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nat. Phys. **2**, 110 (2006).
[3] D. Garlaschelli, G. Caldarelli, and L. Pietronero, Nature (London) **423**, 165 (2003).
[4] J. Memmott, N. D. Martinez, and J. E. Cohen, J. Anim. Ecol. **69**, 1 (2000).
[5] M. Vendruscolo and C. M. Dobson, Proc. Natl. Acad. Sci. U.S.A. **102**, 5641 (2005).
[6] J. Hou *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 3651 (2005).
[7] G. Caldarelli, M. Marsili, and Y.-C. Zhang, Europhys. Lett. **40**, 479 (1997).
[8] J. Kleinberg, ACM Comput. Surv. **46**, 604 (1999).
[9] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, Phys. Rev. E **68**, 046130 (2003).
[10] J.-P Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A.

Kanto, Phys. Rev. E **68**, 056110 (2003).
[11] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[12] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[13] G. Iori, G. De Masi, O. Precup, G. Gabbi, and G. Caldarelli, J. Econ. Dyn. Control (to be published).
[14] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[15] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[16] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, Physica A **352**, 669 (2005).
[17] J. Reichardt and S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004).
[18] P. Hartmann, M. Manna, and A. Manzanares, J. Int. Money Finance **20**, 895 (2001).
[19] E. Barucci, C. Impenna, and R. Reno, Res. Bank. Finance **4**, 321 (2004).
[20] M. Boss *et al.*, Quant. Finance **4**, 677 (2004).
[21] Data are taken from the Italian electronic broker market MID (Market for Interbank Deposits) is run by e-MID S. p. A. Società Interbancaria per l'Automazione (SIA), Milan. The Italian electronic broker Market for Interbank Deposit (e-MID) covers the entire overnight deposit market in Italy. The information about the parties involved in a transaction allows us to perform an accurate daily analysis of the connectivity among banks and its change over time.
[22] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[23] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[24] V. D. P. Servedio, G. Caldarelli, and P. Buttà, Phys. Rev. Lett. **70**, 056126 (2004).