Communities in Italian corporate networks

Carlo Piccardi a,*, Lisa Calatroni a, Fabio Bertoni b

a Department of Electronics and Information, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
b Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy

ARTICLE INFO

Article history:
Received 28 October 2009
Received in revised form 15 March 2010
Available online xxxx

Keywords:
Networks
Communities
Modularity
Board network
Ownership network

ABSTRACT

The community structure of two real-world financial networks, namely the board network and the ownership network of the firms of the Italian Stock Exchange, is analyzed by means of the maximum modularity approach. The main result is that both networks exhibit a strong community structure and, moreover, that the two structures overlap significantly. This is due to a number of reasons, including the existence of pyramidal groups and directors serving in several boards. Overall, this means that the “small world” of listed companies is actually split into well identifiable “continents” (i.e., the communities).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Complex networks have been one of the most extensively studied subjects, in the last decade, in the field of applied mathematics. Researchers have dealt with a wide variety of real systems composed of a large number of highly interconnected units, such as the internet and the world wide web, social and communication networks, scientific citations and collaborations, epidemic spreading mechanisms, food webs, biochemical networks of metabolic reactions or protein interactions, neural networks, and many others. Surveys of most of the theoretical and applied results can be found in several papers and textbooks (e.g., Refs. [1–6]). Economics and finance have also been deeply involved in this research effort. Topics that have been considered include ownership relationships and corporate control, pyramidal groups’ evolution and cross shareholding dynamics [7–11], financial architecture of corporations in national or global economies [12–14], interlocking directorships and their consequences on the decisional processes [15–18], decision problem decomposition into distinct tasks, distributed among the different units of an organization, both in formal (top management, business units) and informal structures (community of practice) [19,20].

One of the topics on which network researchers have recently been more active is community analysis [21–24], which is aimed at revealing possible partitions of the network into communities (i.e., subsets of nodes) with dense intra- but sparse inter-community connections. Finding and analyzing such partitions can provide invaluable help in capturing some fundamental properties of the network. For instance, it has often been observed that groups of nodes belonging to the same community are likely to share common properties or play a similar role. Or, identifying communities and their boundaries allows one to classify vertices according to their topological position. Community analysis is a computationally hard task, however, because the “best” partition must be found in a set whose cardinality grows exponentially with the number of nodes. Past work on methods for discovering groups in graphs has mainly exploited standard clustering techniques, such as graph partitioning and hierarchical clustering. In the last decade, together with the growing interest on complex networks, several other approaches have been put forward. A thorough, critical survey can be found in the review paper [24].

* Corresponding author.
E-mail address: carlo.piccardi@polimi.it (C. Piccardi).

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.06.038

Please cite this article in press as: C. Piccardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038
where the “modern” approaches to community analysis are classified into divisive algorithms; modularity-based methods; spectral algorithms; dynamic algorithms; and methods based on statistical inference. The approach followed in this paper is modularity optimization, which follows from the seminal work of Newman and Girvan [21]. Among the methods for community analysis, it is probably the one that has encountered the largest popularity, both in terms of algorithms and in applications to real-world data (see, e.g., Refs. [25–29,24]), although some critical aspects and limitations have been evidenced [30–32].

We analyze and compare the community structure of two real-world financial networks, namely the board network and the ownership network based on the firms listed in the Italian Stock Exchange (“Borsa Italiana”) at the end of 2008. In both networks, each node represents a firm. In the board network, a connection is established if one or more directors sit in the boards of two companies. The ownership network describes instead direct ownership connections among firms. Akin to previous works, we find a high degree of interconnection in both networks mainly as a consequence of pyramidal groups (i.e., sets of legally independent firms subject to a unitary control), cross-shareholdings (i.e., mutual ownership ties between companies), and “busy” directors (i.e., directors serving in several boards). The main contribution of this work to the existing literature is that we find that both networks exhibit a strong community structure and, moreover, that the two structures overlap significantly and capture some crucial aspects of the Italian financial system. The paper is organized as follows: in Section 2, the essential notions on networks and communities are recalled. In Sections 3 and 4, respectively, the board and ownership networks are defined and their community structure is analyzed. Then, in Section 5, the partitions obtained, in the two networks, by community analysis are compared by means of three different similarity indexes. A thorough economic interpretation of the results is given in Section 6, and some concluding remarks are in Section 7.

2. Networks and communities

In this section, we briefly introduce the terminology on networks and communities that will be used throughout the paper. Detailed surveys can be found in Refs. [4,5,23,24].

We consider networks composed of \( N \) nodes (\( N = \{1, 2, \ldots, N\} \) is the set of nodes) and \( L \) links, and denote by \( A = [a_{ij}] \) the \( N \times N \) connectivity matrix, where \( a_{ij} = 1 \) if there exists the link \( i \rightarrow j \) and \( a_{ij} = 0 \) otherwise. \( A \) is symmetric if the network is undirected (i.e., \( i \rightarrow j \) implies the existence of \( j \rightarrow i \)). In the latter case, the degree of node \( i \), i.e., the number of links incident to \( i \), is given by \( k_i = \sum_j a_{ij} = \sum_j a_{ji} \), and the degree sequence is the list \( \{k_1, k_2, \ldots, k_N\} \) of the node degrees. In a directed network, on the other hand, we distinguish between the in-degree \( k_i^\text{in} = \sum_j a_{ij} \), the out-degree \( k_i^\text{out} = \sum_j a_{ji} \), and the total degree \( k_i = k_i^\text{in} + k_i^\text{out} \). The degree distribution \( p(k) \) specifies, for each \( k \), the fraction of nodes having (total) degree \( k = k_1 \), whereas the cumulative degree distribution \( P(k) \) is defined as \( P(k) = \sum_{h=k}^{\infty} p(h) \). The average degree of the network is given by \( \langle k \rangle = 2L/N = \sum_k kp(k) \).

A network is connected if, for every pair \( (i, j) \) of distinct nodes, there exists a path from \( i \) to \( j \). The distance \( d_{ij} \) is the length of the shortest path from \( i \) to \( j \), while the diameter is the largest distance in the network. If the network is not connected, the set \( \mathbb{N} \) of nodes can be partitioned in components \( \mathbb{X}^1, \mathbb{X}^2, \ldots, \mathbb{X}^m \) having, without loss of generality, \( N_1 \geq N_2 \geq \cdots \geq N_m > 0 \) nodes, respectively \( (\sum N_i = N) \). Each component is a maximally connected sub-network (i.e., it is connected and it is not part of a larger connected sub-network). In our study, we will typically find that the largest component \( \mathbb{X}^1 \) has a dimension \( N_1 \) which is of the same order of magnitude as \( N \) and, on the other hand, it is much larger than all the other components: such a sub-network is called a giant component. Network components can be identified by means of standard algorithms of graph analysis [33].

Consider now an undirected, connected network (or, if not connected, its giant component). Roughly speaking, a subset \( \mathbb{C}_h \subset \mathbb{N} \) is called a community if the density of links internal to \( \mathbb{C}_h \) is much larger than the density of links connecting \( \mathbb{C}_h \) to the rest of the network. A precise, quantitative formulation of this notion has been put forward by Newman and Girvan [21], and it relies on the notion of modularity \( Q \). Given a network with nodes \( \mathbb{N} \) and a partition \( \mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_q \) (i.e., \( \bigcup_h \mathbb{C}_h = \mathbb{N} \) and \( \mathbb{C}_h \cap \mathbb{C}_k = \emptyset \) for all \( h, k \)), the modularity is given by

\[
Q = \frac{1}{2L} \sum_{h=1}^{q} \sum_{i,j \in \mathbb{C}_h} \left[ a_{ij} - \frac{k_i k_j}{2L} \right].
\]

(1)

\( Q \) is obtained by summing up, through all the sets \( \mathbb{C}_h \), the difference between the actual number of links internal to the set \( \left( \frac{1}{2} \sum_{i,j \in \mathbb{C}_h} a_{ij} \right) \) and the value expected if links were created at random but preserving the node degrees \( \left( \text{which can be proved to be} \frac{1}{2} \sum_{i,j} \frac{k_i k_j}{2L} \right) \). Thus \( Q \) is large (i.e., it tends to 1, due to the proper normalization) when the density of links internal to the sets \( \mathbb{C}_h \) (the communities) is surprisingly large with respect to a random distribution of links in the network.

Analyzing the community structure of a network amounts, first of all, at finding the partition \( \mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_q \) to which the largest modularity \( Q = Q_{\text{max}} \) is associated. At this point, a high value of \( Q_{\text{max}} \) denotes the existence of a true, significant community structure in the network, whereas a small value of \( Q_{\text{max}} \) means that the link distribution in the network is not far from being random. In order to assess whether the obtained value of \( Q_{\text{max}} \) is significantly high, one can consider the ensemble of networks having the same degree sequence as the original one, and extract a large number \( M \) of random networks.

Please cite this article in press as: C. Piccardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038
from this ensemble (see Ref. [34] for a comparative analysis of generation methods). Then, the maximum modularity $Q_i$, $i = 1, 2, \ldots, M$, is computed for each one of them. At this point, denoting by $\mu$ and $\sigma$ the mean and standard deviation of the $Q_i$s, a large value of the $z$-score

$$z = \frac{Q_{\text{max}} - \mu}{\sigma}$$

(2)

indicates that the maximum modularity obtained for the original network is significantly high. Although this technique may yield questionable results in some specific cases (see Ref. [24] for a discussion), we will see in the next sections that, for the networks we study, the $z$-scores we obtain are so large to safely guarantee the significance of the community partition.

The notion of modularity can be extended to weighted networks [35]. If we denote by $w_{ij}$ the weight of the link $i \to j$, then the modularity $Q$ of a given partition is defined as

$$Q = \frac{1}{2w} \sum_{h=1}^{q} \sum_{i,j \in C_h} \left[ w_{ij} - \frac{w_i w_j}{2w} \right],$$

(3)

where $w_i = \sum_j w_{ij}$ is the strength of node $i$, i.e., the total weight of the links incident to $i$, and $w = \frac{1}{2} \sum_i w_i$ is the total weight of the links in the network. Eq. (3) is the natural extension of (1) in the case of integer weights, since $w_{ij}$ can be interpreted as the number of links connecting $i$ to $j$ (in this case the network is a multigraph). Otherwise, one can (approximately) transform the weights of the original network into integers by discretization, namely by measuring them with respect to a sufficiently small unit (“discretized network”). This procedure is required, for instance, to generate a set of random networks in order to compute the $z$-score (2). Notice, however, that a smaller and smaller unit yields larger and larger (integer) weights in the discretized network. This artificially enlarges the ensemble of multigraphs having the same strength sequence as the discretized network, including graphs that tend to be more and more uniformly connected. The ultimate result is a bias toward 0 of the sampled $Q$-scores, resulting in an artificially large $z$-score. Taking the unit equal to the minimum weight in the network, as suggested in Ref. [36], does not help if the weights span many orders of magnitude. We heuristically cope with this problem by adopting a somehow opposite strategy, namely we discretize the $w_{ij}$s with a unit as large as possible, with the constraint, however, that the $Q_{\text{max}}$ values of the original and discretized network differ by no more than $10^{-2}$.

After Ref. [21], the modularity approach to analyze network communities has been widely applied and proved to be very effective in capturing the structure of many real-world networks (see Refs. [22,25–29,23], just to mention a few). In parallel, a great effort has been devoted to devising efficient algorithms for finding the best network partition (i.e., $Q = Q_{\text{max}}$). Since it has been proved that the exhaustive optimization of $Q$ is a computationally hard problem [37], a large number of practical, sub-optimal methods have been proposed. We have used three of them: the spectral method by Newman [22,25], the greedy algorithm based on extremal optimization [38], and the aggregative, hierarchical method devised by Blondel and coworkers [39]. The latter one outperformed the others both in terms of $Q_{\text{max}}$ (i.e., in the capability of finding a partition with higher modularity) and in computational requirements. Therefore, the results presented in the paper are those obtained by using this algorithm.

We end this section by recalling the notion of the (undirected) bipartite network, that is a network composed of two distinct sets of nodes, $N_A = \{1, 2, \ldots, N_A\}$ and $N_B = \{1, 2, \ldots, N_B\}$, and whose links can only connect one node in $N_A$ with one node in $N_B$. The projection of the bipartite network onto the set $N_A$ is an ordinary undirected network with nodes $N = \{1, 2, \ldots, N\}$, where the link $i \leftrightarrow j$ exists (i.e., $a_{ij} = a_{ji} = 1$ in the connectivity matrix) if and only if $i, j \in N_A$ have at least one common neighbor in $N_B$. The projection can also be weighted, the simplest choice being $w_{ij} = n_j$ if $i, j \in N_A$ have exactly $n_j$ common neighbors in $N_B$. A more sophisticated weighting scheme will be proposed in the next section.

### 3. The Italian corporate board network

It is common fact that, in a financial system composed of several companies, some directors sit in more than one board. Thereby, many boards are connected by shared directors; more specifically, a board interlock takes place when two firms have at least one director in common. Thus, a corporate board network consists of boards connected through common directors. Equally, a corporate director network is obtained considering directors as nodes connected through links if they serve together in one or more boards. Board network and director network are projections of a bipartite network, where directors and boards correspond to the sets $N_A$ and $N_B$ (see Section 2) and the link $h \leftrightarrow k$ exists when the director $h \in N_A$ sits in the board $k \in N_B$. Several studies have recently dealt with corporate governance issues by a graph theoretic approach (e.g., Refs. [40,17,18,41]): the aim of this section is to analyze the Italian corporate board network for those companies listed in the Italian Stock Market at the end of 2008 (data are publicly available at [http://www.consob.it](http://www.consob.it)). For that, we first create the bipartite network, which is composed of 2365 directors and 292 boards. A visual representation of the network is shown in Fig. 1, and some basic statistics are reported in Table 1 and Fig. 2.

The projection of the bipartite network onto the set of boards generates the board network, which is fully described by the symmetric $292 \times 292$ connectivity matrix $A$. It turns out that this network has 68 components: one of them is a giant

---

1 All the network pictures of the paper have been produced with Pajek [42].
Fig. 1. The bipartite network of boards (yellow nodes) and directors (red nodes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Statistics on the bipartite network of boards and directors.

<table>
<thead>
<tr>
<th></th>
<th>Boards</th>
<th>Directors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>292</td>
<td>2365</td>
</tr>
<tr>
<td>Average degree</td>
<td>9.82</td>
<td>1.21</td>
</tr>
<tr>
<td>Min degree</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Max degree</td>
<td>25</td>
<td>7</td>
</tr>
</tbody>
</table>

Fig. 2. Degree distribution of boards and directors in the bipartite network.

Component, as it has 217 nodes, namely 74.3% of the total. The other components just contain one or two nodes, thus from now on we restrict our attention to the giant component. Some related statistics are in Table 2 and Fig. 3. Note that the degree distribution is approximately exponential in a large range of $k$.

We now introduce weights in the board network by a scheme which, with the aim of reinforcing strong connections and blunting the weak ones, relates the co-membership to the size of the boards. The weight of the link $i \equiv j$ is defined as

\[ w_{ij} = \frac{2n_{ij}}{n_i + n_j}, \]  

(4)

Please cite this article in press as: C. Piccardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038
Table 2
Statistics on the giant component of the board network.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>217</td>
</tr>
<tr>
<td>Average degree</td>
<td>5.14</td>
</tr>
<tr>
<td>Min degree</td>
<td>1</td>
</tr>
<tr>
<td>Max degree</td>
<td>34</td>
</tr>
<tr>
<td>Average distance</td>
<td>3.88</td>
</tr>
<tr>
<td>Diameter</td>
<td>8</td>
</tr>
</tbody>
</table>

![Distance distribution](image1)

**Fig. 3.** Distance distribution (above) and cumulative degree distribution (below) of the giant component of the board network. The red line is the best linear fitting (in log-linear coordinates) for $1 \leq k \leq 23$.

where $n_{ij}$ is the number of directors shared by boards $i$ and $j$, and $n_i$, $n_j$ are the dimensions (i.e., the total number of directors) of the two boards. The weight $w_{ij}$ ranges from 0 (no link, namely no interlock between the boards $i, j$) to the limit case of 1 (the two companies have exactly the same board composition).

Community analysis applied to the giant component of the board network reveals a rather strong community structure, as it identifies 12 communities with a modularity $Q_{\text{max}} = 0.66$. The z-score (2), computed as described in Section 2 on a sample of $M = 100$ random networks, is as high as $z = 51.7$. Thus it can safely be claimed that communities are relevant topological features of the Italian stock exchange board network. This appears to be evident just by looking at Fig. 4, and noticing that the density of links within most of the groups is clearly higher than among them. Indeed, even if we neglect the weights (4) and consider the unweighted network corresponding to Fig. 4, we still obtain a respectable $Q_{\text{max}} = 0.54$ with $z = 12.0$. In other words, the pure topology of the board co-memberships displays, per se, an important community structure, which is even reinforced if, in addition, the strength of these co-memberships is accounted for.

4. The Italian corporate ownership network

Interconnections among firms can assume several forms, and one of them is represented by shareholdings. In recent years, several studies have been carried out to model shareholding markets by graph theory (e.g., Refs. [43,14]). We consider the companies of the Italian Stock Exchange (292 altogether) at the end of 2008, taking shareholding data from the Consob database. We consider all ownership connections publicly disclosed by Consob. Italian regulation is particularly strict about the ownership threshold which entails mandatory disclosure: if any individual or company owns more than 2% in a listed company it should send a filing to Consob which, in turn, will make the information public. As a result we have a very fine-grained description of ownership interconnections between listed companies. Moreover, since the disclosure threshold refers to shares being held both directly and indirectly, the exclusion from the analysis of unlisted firms engenders no loss of information about ownership links among listed companies: if a listed company A owns a stake larger than 2% in another listed company B which is partly or wholly held by a controlled unlisted entity C (a case which occurs frequently for fiscal reasons, since C can be located in a country with lower taxation on dividends and capital gains), company A (the ultimate shareholder) will have to declare to Consob the whole stake it holds, specifying which fraction is held directly and which is instead held by its controlled vehicles. Eventually, we obtain an oriented, weighted network, where nodes are companies...
and a link $i \rightarrow j$ is drawn from the shareholder $i$ to the owned company $j$. We take into account both immediate and ultimate ownership whereas we do not consider any shares held by companies which are not listed in the Stock Exchange (such a network is called “restricted”). Links are weighted according to the proportion of ordinary shares held by the shareholder. Overall, the network has 292 nodes, but approximately half of them are isolated. After removing them, we obtain a “cleaned-out network” with 141 nodes, which is represented in Fig. 5.

The directed graph describing the cleaned-out network contains many nodes with a large in-degree or large out-degree, but just a few of them hold both kinds of link. Thus the network has a bow-tie structure [44], and the community analysis, although feasible [29], would give poor results. In this circumstance, it is reasonable to consider the links as a sign of connection between two firms, regardless of their role as shareholder or owned company. By connection here we refer to the fact that ownership links influence firms’ decision processes by providing an incentive to cooperate (and collude). When company A owns shares in company B, it can influence B’s decisions by voting in its shareholder meetings and, possibly, by
Fig. 6. Distance distribution (above) and cumulative degree distribution (below) of the giant component of the undirected ownership network. The red line is the best linear fitting (in log-linear coordinates) for $1 \leq k \leq 17$.

electing members in its board. As a consequence B’s decision process is influenced by A’s. At the same time, the economic performance of B will affect firm A’s profits through the stake that A owns in B. As a consequence, A’s outcome is influenced by B’s. This creates a strategic interconnection between the two firms. Suppose for instance that one of the two companies has to decide upon something which could hamper the economic performance of the other (e.g., introducing a new product which can put the other company out of business). The existence of an ownership link makes this event less likely regardless of the identity of the decision maker. If the decision has to be taken by A, its ownership stake will translate into an additional cost due to the internalization of a fraction of the economic harm caused to B. If the decision has to be taken by B, A will exercise all the influence it has thanks to its stake to vote against the decision. Eventually, the result is that an ownership link between A and B affects the decision process of both companies and facilitates cooperation (collusion being one possible way cooperation materializes). This ultimately leads to ignoring the directions in the network, i.e., making the graph undirected and symmetrizing the connectivity matrix. We note that this approach is consistent with the work of other authors on, e.g., biological or bibliographic networks [45,46].

The new connectivity matrix is therefore symmetric and weighted: its entries $w_{ij} = w_{ji} \geq 0$ represent the percentage of equity capital hold by company $i$ or $j$. Altogether, 14 components are found out. Most of the network is connected in a giant component composed of 108 nodes, that is 76.6% of the cleaned-out network. The other components contain no more than four nodes, so that we restrict our community analysis to the giant component, whose main statistics are presented in Table 3 and Fig. 6. Analogously to the board network, the degree distribution turns out to be approximately exponential in a wide range of $k$.

The analysis of this network reveals a strong community structure, namely it identifies 15 communities with a modularity as high as $Q_{max} = 0.82$, and a z-score of $z = 29.2$. Analogously to the board network, it emerges that communities are important features of the Italian stock exchange ownership network. Here, the role of the weights appears to be even more important. Indeed, if we neglect the weights (i.e., set them all equal to 1) we obtain a modularity $Q_{max} = 0.59$ (with $z = 3.82$), still respectable but definitely less important than the 0.82 obtained for the weighted network. It is a rather obvious result: the strength of the ownership (the value of $w_{ij}$) is a much richer information than the sole existence of the ownership tie (the link $i \leftrightarrow j$).

Fig. 7 displays the obtained community structure: network connections are dense within groups, and the links with the highest weight are all internal to groups.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics on the giant component of the undirected ownership network.</td>
</tr>
<tr>
<td>Number of nodes</td>
</tr>
<tr>
<td>Average degree</td>
</tr>
<tr>
<td>Min degree</td>
</tr>
<tr>
<td>Max degree</td>
</tr>
<tr>
<td>Average distance</td>
</tr>
<tr>
<td>Diameter</td>
</tr>
</tbody>
</table>

Please cite this article in press as: C. Picardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038
5. Similarity of the board and ownership communities

In Sections 3 and 4, two partitions have been obtained for two distinct networks defined on the same set \( N \) of nodes, namely the 292 companies of the Italian stock exchange\(^2\). The two networks, although technically distinct, are however conceptually related since board members are ultimately elected by shareholders. This leads us to expect similarities between the two partitions. In Section 6, a detailed discussion of the relationships between the community partitions and the existing business groups will be presented. Here, we concentrate instead on a quantitative analysis aimed at assessing whether the overlap between the two partitions is significant. To this end, three different indexes of comparison are considered.

Most of the criteria for comparing different partitions of a set \( N = \{1, 2, \ldots, N\} \) (the network nodes, in our case) can be grouped into three main categories: pair counting, set matching, and information-theory based criteria (see Refs. \([47, 24]\) for detailed expositions). We will consider one index for each category. Generally speaking, we have to compare two different partitions \( A_1 \cup A_2 \cup \cdots \cup A_A = N \) and \( B_1 \cup B_2 \cup \cdots \cup B_B = N \) or, equivalently, two different labelings \( L_1 = \{a_1, a_2, \ldots, a_N\} \) and \( L_2 = \{b_1, b_2, \ldots, b_N\} \) of \( N \), where \( 1 \leq a_i \leq A \) and \( 1 \leq b_i \leq B \) are the labels of node \( i \) in the first and second partition, respectively. Specifically, we consider, for the board network and the ownership network (both with \( N = 292 \) nodes), the partition obtained by first dividing the network into connected components (i.e., \( K_1 \cup K_2 \cup \cdots \cup K_m = N \), see Section 2), and then by identifying the communities within the giant component, thus obtaining a partition of the form \( C_1 \cup C_2 \cup \cdots \cup C_q \cup K_2 \cup \cdots \cup K_m = N \), whose elements – be they components or communities – will be denoted as clusters in the remainder of this section.

The **Rand index** \( R \) \([48]\) is a pair counting index: it is the fraction of node pairs that are treated in the same way in both partitions. More precisely:

\[
R = \frac{N_{11} + N_{00}}{N(N - 1)/2},
\]

where \( N_{11} \) (resp., \( N_{00} \)) is the number of pairs that are classified in the same cluster (resp., in different clusters) in both partitions, and \( N(N - 1)/2 \) is the total number of pairs. In our case, we obtain \( R = 0.938 \), which is apparently very large (notice that \( 0 \leq R \leq 1 \), and \( R = 1 \) if and only if the partitions are coincident). A well known drawback of \( R \), however, is that its values tend to concentrate near 1 \([47]\), a fact that suggested the use of adjusted versions of the index. We assess the significance of our value by comparing it with the results obtained from an ensemble of randomized partitions. More precisely, we randomly shuffle the node labels (see above) in both partitions for \( 10^3 \) times, recompute (5) each time, and calculate the mean \( \mu_R \) and the standard deviation \( \sigma_R \) of the \( 10^3 \) values obtained through shuffling. Then, we evaluate \( z_R = (R - \mu_R)/\sigma_R \), i.e., the distance from \( R \) to \( \mu_R \) in units of the standard deviation. The result is \( z_R = 9.87 \), which seems to testify in favor of the significance of the overlapping between the two original partitions. Notice, however, that we find \( \mu_R = 0.935 \), which is also very large and, most notably, strikingly close to \( R \) (in fact, randomization yields values strictly

\(^2\) The complete list of the companies, with the partition in communities and components of the board and ownership networks, can be downloaded at http://home.dei.polimi.it/piccardi/misc/ListCompaniesCommunities.zip.
concentrated around $\mu_R$). Thus the indications given by the Rand index seem to be inconclusive, and other indexes are needed.

The (normalized) van Donsel distance $D$ is a set matching index, defined as follows:

$$D = \frac{1}{2N} \left( 1 - \sum_{i=1}^{A} \max_{j=1}^{B} n_{ij} - \sum_{j=1}^{B} \max_{i=1}^{A} n_{ij} \right),$$

where $n_{ij}$ is the number of nodes which are classified in the $i$th cluster in the first partition and in the $j$th cluster in the second one. It basically consists in finding, for each cluster of a given partition, the “best matching” in the other partition, i.e., the cluster with the largest overlapping. It satisfies $0 \leq D < 1$, with $D = 0$ if and only if the partitions are coincident. In our case, we obtain $D = 0.437$, and the randomization procedure gives $z_D = -10.7$. Although the van Donsel distance may lead to partially distorted results in some specific cases [47], this result is in favor of a non casual overlapping between the two partitions.

Finally, we consider the variation of information $V$, which is an information-theory based distance first introduced by Meilă [47] and rapidly adopted by network scholars (e.g., Refs. [28,24]). Let us define the joint probability distribution $\pi_{ij} = n_{ij}/N$, $i = 1, 2, \ldots, A, j = 1, 2, \ldots, B$ and then $\pi_A^i = \sum_j \pi_{ij}$ and $\pi_B^j = \sum_i \pi_{ij}$, which are the fraction of nodes having label $i$ (resp., $j$) in the first (resp., second) partition. Then, once defined the Shannon entropies $H_A = -\sum_i \pi_A^i \log(\pi_A^i)$ and $H_B = -\sum_j \pi_B^j \log(\pi_B^j)$, and the mutual information

$$I = \sum_{i=1}^{A} \sum_{j=1}^{B} \pi_{ij} \log \frac{\pi_{ij}}{\pi_A^i \pi_B^j},$$

the variation of information is defined as

$$V = H_A + H_B - 2I.$$

If we rewrite (8) as $V = (H_A - I) + (H_B - I)$, the variation of information can be interpreted as the sum of two symmetrical terms: the first one can be regarded as the loss of uncertainty on the partition $A$ due the mutual information provided by $B$, and vice versa for the second term. The index $V$ turns out to be a proper distance in the space of the partitions, and has many other useful properties [47]. It satisfies $0 \leq V \leq \log N$, with $V = 0$ if and only if the two partitions are coincident, and $V = \log N$ if they are “maximally different”, i.e., one partition has $N$ clusters and the other only 1. For the sake of homogeneity with the previous indexes, we normalize the range to $0 \leq V \leq 1$, i.e., we replace (8) with $V = (H_A + H_B - 2I)/\log N$. Returning to our case, we obtain $V = 0.450$, and the randomization procedure gives $z_V = -10.1$, which is another indication in favor of a significant overlapping between the two partitions obtained for board interlocking and shareholding.

6. Business groups, alliances, and corporate networks

By analyzing the composition of the communities, it turns out that they reflect some well-known feature of the Italian shareholding structure (as investigated, e.g., in Refs. [49,8,50]). The underlying economic reason why board and ownership networks exhibit such a high level of similarity is that they are ultimately the consequence of the same phenomenon: the existence of common interests between companies. These common interests give rise to more or less formalized and stable structures of which cross-ownership links and interlocks constitute the “glue”. The strongest of these aggregations is the business group, that is a set of legally independent companies which are subject to a unitary control. Often business groups take a pyramidal structure, with a holding company (i.e., a company whose main business is to hold financial securities and, especially, shares in other companies) on top, and operative companies at the bottom of the hierarchy. An example for this is the Agnelli group: the Agnelli family controls IFI (an holding company), which in turns controls IFIL (another holding company) which ultimately controls FIAT (an holding company which, among other things, owns FIAT Auto, a car maker) and Juventus (a football team). Since IFI, IFIL, FIAT and Juventus are all listed companies included in our sample, we would expect to find them in the same community in both the board and ownership networks. More generally, since business groups are a form of aggregation which is relatively easy to identify, we can use it to check ex-post the consistency of the partitions we obtain. To identify business groups we need a formal definition for the concept of control. We do so by referring to the International Accounting Standard (IAS) 27, which defines control as the “power to govern the financial and operating policies of an entity” [51]. Control is normally, but not necessarily, obtained by holding the majority of a firm’s voting shares. Control can also be obtained by means of shareholder agreements and statutory clauses and, in general, applies to all cases in which an investor can appoint or remove the majority of the members of the board or casts the majority of votes at a meeting of the board of directors. When one company controls another company according to IAS 27, it has to issue a consolidated financial statement in which the two companies are amalgamated (i.e., present joint results, the net of intra-company business). For instance, FIAT and Juventus are fully consolidated in the accounting statement of IFIL. In turn, IFIL (and indirectly FIAT and Juventus) is consolidated in the accounting statement of IFI. Failing to consolidate controlled companies, as well as consolidating non-controlled ones, can lead to serious legal consequences for the company itself and its directors. Eventually, by combining information on the consolidation perimeter of listed companies and Consob filings about ownership we can identify listed companies belonging to the same business group (i.e., subject to unitary control).
We identify 13 business groups including between 2 and 5 listed companies each, for a total of 39 companies. All these 39 companies are included in the giant component of the board network. With only two exceptions all companies which belong to the same business group also belong to the same community in the board network. The giant component of the ownership network includes 22 of the 39 companies, representing 10 business groups. All companies which belong to the same business group are also included in the same ownership community.

Overall, both community structures identify consistently the existence of business groups. Yet, several other “softer” forms of coalition among firms exist, which are more difficult to identify than business groups and that the community analysis might help to highlight. Control, indeed, is only the most extreme form of the more general concept of influence, that is the power to steer the decisions of a company towards a direction. Due to the difficulty in detecting these softer forms of coalition we do not perform any quantitative test but we limit ourselves to cite two relevant cases. The Pirelli group includes two companies, Pirelli & C. and Pirelli Real Estates; Pirelli & C. is an holding company which controls Pirelli Real Estates and whose largest shareholder is, with 26.15% of its voting rights, CAMFIN, another listed holding. The ownership stake of CAMFIN over Pirelli & C. does not fall into the definition of control given by IAS 27 (and Pirelli & C. is, accordingly, not fully consolidated in CAMFIN’s financial statement). Yet, the influence of CAMFIN over the Pirelli group is extremely relevant and Marco Tronchetti Provera, a well-known Italian financier who controls CAMFIN, is the chairman of Pirelli. Interestingly, CAMFIN is in the same community with the Pirelli group in both the board and the ownership networks.

Another interesting example is the succession battle at Assicurazioni Generali, the largest Italian insurer, which will result in the election of the new chairman in April 2010. The Financial Times summarizes quite effectively how Mediobanca and its president, Cesare Geronzi, are influencing this process: “[...] it will be within the walls of Mediobanca, the Milan-based investment bank, that the future leadership of Generali will most likely be decided. Tucked away behind La Scala opera house, Mediobanca [...] retains one last vestige of its former status in the shape of its 15% stake in Trieste-based Generali. In the run-up to Generali’s April shareholders’ meeting at which the insurers board is up for renewal, [...] Mediobanca will momentarily resume its place in the limelight as the leadership of one of Italy’s great names comes up for grabs” (Financial Times, February 3, 2010, “French connection could decide Generali’s succession”, by Paul Betts). The influence which Mediobanca is recognized to exert over Generali is not due to a control which the former has over the latter (the stake is only 15%), but rather the outcome of a complex equilibrium of powers (with a strong historic, or path-dependent, component) which is very hard to measure quantitatively. It is interesting to observe that in both the board and ownership network, Mediobanca shares its community with one or both companies of the Generali group.

Summarizing, the community structure which we find in both the ownership and the board networks is the consequence of the tangling-up of interests which characterizes Italian companies and which takes different forms, from business groups down to softer forms of reciprocal influence. Interestingly, while strong forms of coalition (like business groups) can be identified pretty easily by applying objective rules, softer coalitions are far more difficult to measure quantitatively and community analysis proves to be a very effective method to spot them.

7. Concluding remarks

The Italian Stock Exchange is an ideal environment to test the application of community analysis. First, we can find in Italy the whole spectrum of interconnections among companies, including fully-fledged family-controlled pyramidal groups, State-owned companies, and “softer” coalitions such as those around Mediobanca (for a description of corporate ownership in Italy see Ref. [50]). Second, despite some specificities, Italy is somewhat representative of the complex tangling up of interests which also characterizes other European countries like Germany, France and Belgium (see, for instance [52] for a cross-country comparison of director networks). Finally, Italy has the advantage of being the European country with the lowest threshold for disclosure of ownership in a listed company: whoever comes to possess directly or indirectly more than 2% of the voting capital of a listed company has to send a filing to Consob (the Italian SEC) which in turn makes it available to the public. This allows us to have a very precise description of the ownership structure of listed companies which is the fundamental input of the ownership network.

Our results add to the growing literature on corporate networks by showing that both the ownership network and the board network exhibit a structure which is organized in communities and, in addition, that communities overlap significantly between the two networks. The most well-known form of corporate interdependence which drives this result is that of pyramidal groups. Control in pyramidal groups is obtained by means of hierarchical shareholdings (i.e., companies are organized in layers with a holding company usually at the top of the pyramid) and is enforced by means of strongly

---

3 We provide here the full list of the business groups we identified; companies included in each group are reported in brackets; we assign to each group the name which is normally used to identify it in the press and by practitioners (which is often the name of the controlling family, if any): De Benedetti (CIR, COFIDE, L’Espresso, SOGEFI), Pirelli (Pirelli & C, Pirelli Real Estates), Agnelli (IFIL, IFIL, FIAT, Juventus), Pesenti (Italmobiliare, Italcementi), Ligresti (PREMAFIN, Fondiaria-SAI, Milano Assicurazioni), Caltagirone (Caltagirone, Caltagirone Editori, Cementir, Vianini Industrie, Vianini Lavori), Telecom (Telecom Italia, Telecom Italia Media), Fininvest (Mediaset, Mediolanum, Mondadori), Benetton (Autogrill, Atlantia, Benetton Group, Autostrade Meridionali), Burani (Antichi Pellettieri, Burani Group, Greenvision, Bioera), Intek (Intek, KME group), Aurelia (SIAS, Autostrade Torino–Milano), Generali (Assicurazioni Generali, Banca Generali).

4 The exceptions are the Telecom group which is split between communities 6 and 11, and the Generali group which is split between communities 9 and 11.

---

Please cite this article in press as: C. Piccardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038
overlapping boards. This form of hard interdependence is quite common worldwide, especially in continental Europe [8] and Asian countries [53]. It is usually relatively simple to decide whether a firm belongs to a pyramidal group thanks to the notion of control; consequently we use this prior to test the extent to which the partitions we obtain with community analysis are empirically sound. As was shown in Section 6, the giant component of the board (resp., ownership) network contains 39 (resp., 22) companies belonging to 13 (resp., 10) business groups, and, as expected, community analysis assigns to the same community the companies belonging to the same business group, with only two exceptions (no exceptions in the ownership network).

The most interesting aspect of community analysis, however, is its ability to spot “softer” forms of coalitions. Board as well as ownership links can be used by companies to strengthen alliances. A very special case is that of two competitors which use interlocks or cross-ownership ties as means for sustaining a collusive behavior. To avoid this, regulators usually limit the possibility of interlocks and cross-ownership between competitors.\(^5\) Another, less detectable, form of coalition is due to the desire of a group of companies to form an alliance to lobby in favor of some law or obtain some political influence. Examples about the importance of political connections abound: Faccio [54] finds that political connections are widespread worldwide (they are present in 35 out of 47 countries in her sample) and are especially common in countries with higher perceived corruption.\(^6\) There are also many circumstances in which interlocks and cross-ownership can strengthen more constructive coalitions, like strategic alliances [55], or serve as mechanisms for co-optation and monitoring (see Ref. [56] for an extensive review of different theories of interlocks formation). Due to the difficulty in detecting these softer forms of coalition we do not perform any quantitative test but we cite two relevant cases (the chairmen of Pirelli and Generali) to illustrate that non-controlling influence is extremely relevant and, notably, is captured by community analysis.

In order to understand the nature and consequences of corporate coalitions, whatever they are, it is clear that it is essential to have a method for clearly identifying them. To this extent community analysis can prove to be a substantial methodological contribution to the literature.

Acknowledgements

The authors are grateful to two of the anonymous reviewers for their comments and suggestions.

References


\(^5\) The view that interlocks could sustain collusion was first pointed out by in the US by the Pujo Committee in the early twentieth century. Interlocks between competitors were later banned by the Clayton Act in 1914.

\(^6\) See Ref. [50] for some examples about political connections in Italy since the early 1900s and their importance.

Please cite this article in press as: C. Piccardi, et al., Communities in Italian corporate networks, Physica A (2010), doi:10.1016/j.physa.2010.06.038