POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Network Analysis of Listed Italian Companies

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Alessandro Schiraldi

Advisor: Prof. Francesca Grassetti

Co-advisor: Davide Stocco

Academic year: 2020-2021

---

1. Introduction

The thesis focuses on the ownership structure and presence of interlocks among several companies listed in the italian stock exchange. It aims to study the relationship between those features across firms by introducing a multi layer network representation of the market. This is done by creating and analysing different graphs, each one build from either data regarding ownership composition of firms or their Board of Directors structure. Those graphs show the different interconnetions taking place between companies in the italian market, and each of them gives a different representation of the overall market and of its players, by displaying the degrees of connections tighting it up and the importance of each company in it. In particular, four different graphs are created: two of them consider interlocks for their creations, while the others focus on the presence of ownership ties or on the degree of similarity between companies'ownership structure. These graphs are first presented independently with the aim of identifying their general topology. Then, they are then studied together, allowing for the comparison between layers. Measures as edge overlap, graph distance and VN entropy are used for understanding the relationships between the layers and the degree of similarity between them and therefore of the dataset from which they stem.

2. Mathematical Framework

Let $G = (V, E)$ be a graph where V represents the set of nodes of G and $E \subset V \times V$ the set of edges of G. Consider undirected graphs, that is edges connecting nodes without having a specific direction.
Consider a graph $G = (V, E)$ of N nodes and indicate them as $V = \{v_1, \dots, v_N\}$. Write $A = (a_{ij})_{i,j=1}^N$ as the adjacency matrix of G, that is a matrix whose entry is 1 if nodes v_i and v_j are linked by and edge and 0 otherwise. It is possible to attach positive real numbers to edges representing their importance; these are called weights of the edges. Indicate with $W = (w_{ij})_{i,j=1}^N$ the weight matrix of G, where $w_{ij} \in [01]$ is equal to the weight of the edge connecting nodes v_i and v_j if this one exists and 0 otherwise. The degree and strength of node v_i are respectively $k_i = \sum_{j=1} a_{ij} = \sum_{j=1} a_{ji}$ and $s_i = \sum_{j=1} w_{ij} = \sum_{j=1} w_{ji}$ A multilayer network $\mathcal{M}$ of order M is a pair $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ where $\mathcal{G} = \{G_1, \dots, G_M\}$ is a set of M graphs $G_\alpha = (V_\alpha, E_\alpha)$ and $\mathcal{C}$ is a set of interconnections between nodes of different layers G_α and G_β. When comparing the layers, but also for graphs

in general, consider for simplicity graphs to insist on the same set of nodes. For a network $\mathcal{M}$ this is equal to say that $V_1 = \cdots = V_M = V$. Consider moreover in this case $\mathcal{C}$ to be the set of edges simply connecting the same nodes over the different layers.
For each $\alpha = 1, \ldots, M$, indicate with $A^{[\alpha]} = a_{ij}^{[\alpha]}$ and $W^{[\alpha]} = w_{ij}^{[\alpha]}$ the $N \times N$ adjacency and weight matrices of layer G_α. Indicate with $k_i^{[\alpha]} = \sum_{j=1}^N a_{ij}^{[\alpha]}$ and $s_i^{[\alpha]} = \sum_{j=1}^N w_{ij}^{[\alpha]}$ respectively the degree and strength of node v_i in layer α.
For comparing different layers across a network, we introduce, among others, the following measures: *edge overlap*, *graph distance* and *VN entropy*. Considering a network $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ and two two subnetworks $\mathcal{M}_1$ and $\mathcal{M}_2$ whose layers are the graphs $\mathcal{G}_1 = (G_{\alpha_1}, \ldots, G_{\alpha_{M_1}})$ and $\mathcal{G}_2 = (G_{\beta_1}, \ldots, G_{\beta_{M_2}})$, with $G_j \in \mathcal{G} \forall j$ and $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{G}$, the edge overlap EO between the two subnetworks is:

$$EO(\mathcal{M}_1, \mathcal{M}_2) = \frac{\sum_{i,j=1}^N a_{ij}^{[\alpha_1]} \cdot \ldots \cdot a_{ij}^{[\alpha_{M_1}]}}{\sum_{i,j=1}^N a_{ij}^{[\beta_1]} \cdot \ldots \cdot a_{ij}^{[\beta_{M_2}]}}$$

where N is the number of nodes. The edge overlaps computes the fraction of edges which are presented in all layers of $\mathcal{G}_2$ that are also present in all layers of $\mathcal{G}_1$. It can take values from 0 to 1.
For what concerns the definition of a distance $d : \mathbb{G} \times \mathbb{G} \to \mathbb{R}$ between graphs (here $\mathbb{G}$ represents the set of graphs of N nodes), this can be done in several ways. The important thing is that the operator d satisfies the usual properties that a distance measure is required to satisfy. These are being positively defined, taking null value only when valuating the distance of an element with itself and satisfying the triangular inequality. For unweighted graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$, we define d as :

$$d(G_1, G_2) = \sum_{\{u,v\} \in E_1 \cap E_2} |d_{G_1}(u,v) - d_{G_2}(u,v)| + \sum_{\{u,v\} \in E_1 \setminus E_2} |N - d_{G_1}(u,v)| + \sum_{\{u,v\} \in E_2 \setminus E_1} |N - d_{G_2}(u,v)|$$

where d_{G_i} is the usual distance among nodes in graphs G_i (i.e. it corresponds to the length of the shortest path connecting them. The length of a path is the number of edges composing the path). This distance can be normalised so to fit in the interval [0 1].
VN entropy on the other hand requires first the definition of the projection operator on a network $\mathcal{M}$ in order to be introduced. In this direction, indicate with $MN(N, M)$ the set of (possibly weighted) multi networks composed by N nodes and M layers. Then, given a network $\mathcal{M} = (\mathcal{G}, \mathcal{C}) \in MN(N, M)$, the *projection* of order $m \in 1, \ldots, M$ of the sub network $\mathcal{S} \subset \mathcal{G}$ composed by layers $(S_{\alpha_1}, \ldots, S_{\alpha_m})$ is the operator π_S such that:

$$\pi_S : \begin{array}{c} MN(N, M) \to MN(N, 1) \\ \mathcal{M} \mapsto \pi_S(\mathcal{M}) \end{array}$$

where the 1 layer network $\pi_S(\mathcal{M})$ is completely identified by the adjacency matrix $\mathcal{A}_{\pi_S}$ whose entries are

$$a_{ij}^{\pi_S} = \begin{cases} 1 \;\; if \;\; \exists \alpha \in \{\alpha_1, \ldots, \alpha_m\} : a_{ij}^{[\alpha]} = 1 \\ 0 \;\; otherwise \end{cases}$$

and, in case of a weighted networks, by the matrix $\mathcal{W}_{\pi_S}$ whose entries are:

$$\mathcal{W}_{\pi_S}(i,j) = w_{ij}^{\pi_S} = \sum_{\alpha \in \{\alpha_1, \ldots, \alpha_m\}} w_{ij}^{[\alpha]}$$

Given this notation, considering a network $\mathcal{M}$ of N nodes and M layers, the aggregate *projection matrix* $\mathcal{A}_p = (a_{ij}^p)_{i,j=1}^N$ is defined as the adjacency matrix of the 1 layer network $\pi_p(\mathcal{M}) := \pi_{\mathcal{M}}(\mathcal{M})$.
Now, given a graph $G = (V, E)$ of N nodes, the Von Neuman entropy h_{VN} of the graph is given by:

$$h_{VN} = -Tr(\mathcal{L}_G \cdot log_2(\mathcal{L}_G))$$

where the operator Tr is the trace operator, that is $Tr(B) = \sum_{i=1}^N b_{ii}$ with $B = (b_{ij})_{i,j=1}^N$, and $\mathcal{L}_G$ is the rescaled laplacian matrix associated with the graph G. $\mathcal{L}_G = c \cdot (D - A)$ where D is a diagonal matrix, having as diagonal entrances the degrees of the nodes, A is the adjacency matrix and $c = 1/\sum_{i,j=1}^N a_{ij}$ is a rescaling factor. Assuming now to consider a network $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ of M layers across N nodes, the Von Neumann entropy for $\mathcal{M}$ is

$$H_{VN}(\mathcal{M}) = \frac{1}{M} \sum_{\alpha=1}^M h_{VN}^{[\alpha]}$$

VN entropy can be computed for every representation $\mathcal{R}$ of the network. A representation $\mathcal{R}$ is a subnetwork of $\mathcal{M}$ whose layers are either original layers of $\mathcal{M}$ or stem from the projection of some of them onto one single layer. In the notation just introduces the $m \leq M$ layer network $\mathcal{R}$ is composed by layers $(R_1, \ldots, R_m)$ such that for each i, $R_i = G_j \wedge R_i = \pi_{\mathcal{S}}(\mathcal{M})$ for some j and some $\mathcal{S}$. Moreover each layer G_j belonging to $\mathcal{M}$ must contribute to the construction of *one and only one* of the layers R_i. Structural reducibility calculates the VN entropy between a given network and its aggregate, projected representation. Structural reducibility aims to find the best representation $\mathcal{R}$ of the network $\mathcal{M}$ that maximises the quantity:

$$q(\mathcal{R}) = 1 - \frac{H_{VN}(\mathcal{R})}{H_{VN}(\pi_{\mathcal{M}}(\mathcal{M}))}$$

3. Layers Construction

The dataset considered regards companies listed in the italian stock exchange, collecting information for 234 companies for which data is avaliable regarding either ownership structure or BoD composition. Based on this data four different weighted layers are constructed: two regarding interlocks (a direct and an indirect version), one considering possible ownership ties and one linking firms with similar onwhersip structure.

3.1. Direct Interlocking

In the DI layer, companies are linked when they have at least one director in common. DI graph turns out to have a relatively low number of edges (269). Out of 201 considered firms, 38 of the are not connected to any other, while 143 of them, around 73% of the sample, compose a giant component. Weights are proportional to the number of interlocks. The average degree and strength are 2,67 and 0,33, showing how the layer is far from being highly connected.

3.2. Indirect Interlocking

In the II layer, a link is created among two companies not only if they share a common director, but also if there exist a couple of directors, sitting on the boards of the respective firms, that also sit, together, as members of another listed BoD. When considering also this scenario, the final layer structure presents many more edges (1020). Interestingly enough, the topology of the graph remains the same; indeed, the same 38 companies stay isolated, while the giant component is again including the same 143 firms. What changes are the mean values for degree and strength (respectively 10,15 and 2,19). This is a consequence of the increment in the number of edges.

3.3. Ownership Ties

In the OT layer, links are born when either companies have a common shareholder or one is investing in the other. Links are undirected, so the information about which firm is investing in which is discarded. The final structure is composed by 232 firms and 239 edges. In this framework, companies are roughly divided in three categories: isolated nodes, firms belonging to the main giant component and companies ingaging in few links outside of the main component. The giant component is composed by 72 firms, that is a little more than 31% of the sample, while the number of isolated nodes is 130. As fewers links are present, also mean degree and strength witness a degrease, reaching values of 2,06 and 0,74.

3.4. Ownership Similarity

The OS layer is build by means of a *similarity index* s. This score assigns to each couple of companies a number that takes values in the range [0 1]. The higher the score, the more companies are regarded as similar. This index is based on ownership data; it takes into account the percentage of shares held by the main shareholder as well as the aggregate shares owned by the relevant shareholders (A shareholder is relevant if he holds at least 2% of the total equity). It considers the number of relevant shareholders and finally also the category to which the companies' main shareholders belong to. If, for two companies, the score s is higher than a certain threshold t, then a link is constructed between the two firms. t is chosen as 0,875. The final structure results in a disconnected graph, composed by 232 firms and 879 edges. It decomposes in two main giant components and other smaller disconnected ones. Overall, the number of connected components is 62, when including also isolated nodes. It turns out that firms belonging to the main giant components are heavily controlled by their first shareholder who owns more than 50% of total shares. Mean degree and strength are respectively 7,58 and 2,81.

4. Network Analysis

The layers are studied together in order to compare the results. Two 3-layer networks are considered, one composed by layers (DI, OT, OS) and the other one by (II, OT, OS). This to understand the effect of including indirect interlockings into the picture.

4.1. Network 1

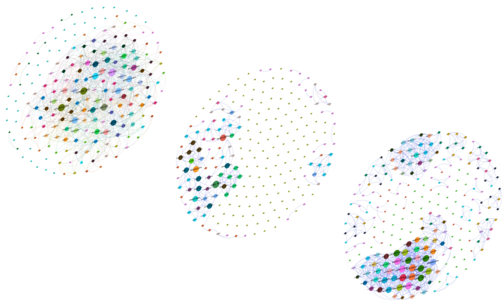

Figure 1: Network (II,OT,OS)

The first network $\mathcal{M}_1$ is composed by a 234 companies and a total number of 1387 edges. The OS layer is the most relevant one in terms of number of edges as more than 65% of them belong to this layer. Beside 6 isolated firms and two other only connected to each other, all companies are connected in a giant component.

$\mathbf{EO}_A$	(DI, OT)	(DI, OS)	(OT, OS)	(DI, OT, OS)
DI	0,1152	0,0260		0,0186
OT	0,1297		0,0837	0,0209
OS		0,0080	0,0228	0,0057
(DI, OT)				0,1613
(DI, OS)				0,7143
(OT, OS)				0,2500

Table 1: Edge overlap

Edge overlap is very low, especially when taking into account the OS layer. OT and DI are the layer overlapping the most, given the fact that $EO(DI, (DI, OT)) = 0,12$ and $EO(OT, (DI, OT)) = 0,13$. Looking at the dataset more in depth shows how companies for which overlapping happens are linked mainly by owners that are either a physical person or a public investor.

distances	DI	OT	OS
DI	0	0,3679	0,4288
OT	0,3679	0	0,1774
OS	0,4288	0,1774	0

Table 2: Distances across graphs

The distance d on the other hand indicates OT and OS as the most similar graphs. The reason behind this is because both layers present many same couple of nodes that are disconnected, and are thus regarded as equally distancing between the two layers. DI on the other hand identifies a main connected component to which the majority of nodes belongs to. VN entropy and structural reducibility state that the network representation $\mathcal{R}$ departing the most from the projected network representation $\pi_p(\mathcal{M}_1)$ is the original 3-layer network.

$\mathcal{R}$	**VN entropy**	**q**
(DI,OT,OS)	6,79	0,0928
((DI,OT),OS)	7,05	0,0581
((DI,OS),OT)	6,82	0,0881
(DI,(OT,OS))	7,18	0,0410
((DI,OT,OS))	7,48	0

Table 3: VN entropy

Because of these results and because of the very low nomber of overlaps, it turns out that all layers tend to complement each other with respect to their projected layer.

4.2. Network 2

The second network $\mathcal{M}_2$ is composed by layers (II, OT, OS).

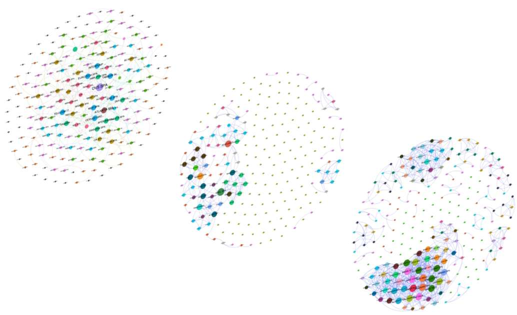

Figure 2: Network (II,OT,OS)

The network present a total of 2147 eges. The topology is the same as in $\mathcal{M}_1$, in the sense that there are the same 6 isolated nodes and the same giant component.

EO	*(II,OT)*	*(II,OS)*	*(OT,OS)*	*(II,OT,OS)*
II	0,0486	0,0233		0,0058
OT	0,2092		0,0837	0,0251
OS		0,0273	0,0228	0,0068
(II,OT)				0,12
(II,OS)				0,25
(OT,OS)				0,3

Table 4: Edge overlap

Edge overlap between interlocks and ties increases with respect to layer OT as $EO(OT,(II,OT)) = 0,21$. The distance among layers remain practically the same.

distance	*II*	*OT*	*OS*
II	0	0,3709	0,4319
OT	0,3709	0	0,1774
OS	0,4319	0,1774	0

Table 5: Distances between graphs

VN entropy appoints again the original representation $\mathcal{M}_2$ as the best one for describing the network.

$\mathcal{R}$	**VN entropy**	**q**
(DI,OT,OS)	6,706	0,0987
((DI,OT),OS)	7,0518	0,0598
((DI,OS),OT)	6,8682	0,0843
(DI,(OT,OS))	7,1760	0,0433
((DI,OT,OS))	7,5000	0

Table 6: VN entropy

In general, even if by including also indirect interlocks into the picture leas to many more edges to be born, the overall structure of the network does not change significantly.

5. Conclusions

What emerges from the study is that there is a strong level of separation between the layers, meaning that considering different data leads to very different final structures. All layers complement each other, and there are very few companies engaging in more than one link. Because of the low level of edge overlap, no relevant correlation is found between strongly concentrated ownership and interlocks. As the dataset considers companies that, on average, are controlled by few investors, this could lead to the conclusion that interlocks are less likely to happen for those firms. On the other hand, the majority of ties between companies happens across firms with a more dispersed ownership structure. Interlocks and ownership ties are the layers that tend to overlap the most, but still present a very complementary feature. The overlaps between those are driven by small clusters of firms, all mainly linked by either a single person or a common public investor. Considering indirect, rather than only directs interlocks leads to more connected networks, but does not change meaningfully the relationships between the layers.

Bibliography

[1] Brandes U. and Erlebach T., Network Analysis, *Springer*, 2005

[2] De Domenico M, Arenas A, Nicosia A and Latora V, Structural reducibility of multilayer networks, *Nature Comunication*, 2015

[3] Brandes U. and Erlebach T., Network Analysis: Methodological Foundations, *Springer*, 2005

[4] Fazlollha M., An Introduction to Information Theory, *Dover Publications Inc.*, 1994