Investigating the relation between companies with topological analysis of a network of Stock Exchange in Brazil

Carlos Magno G. Barbosa, Lucas Gabriel da S. Felix, Carolina R. Xavier, Vinicius da F. Vieira

Federal University of São João del-Rei, Brazil
vinicius@ufsj.edu.br

Abstract. B3 (Brasil, Bolsa, Balcão) is the official stock exchange in Brazil and plays a key role in the world financial market. Stock exchange allows people and companies to relate through the shareholding and the purchase and sale of shares. The study of the relationship between people and companies can reveal valuable information about the operation of the stock exchange and, consequently, the financial market as a whole. In this work, the relations in B3 are modeled as a network, in which the vertices represent companies and people and the edges represent shareholdings. From the built network, several analyzes are performed with the objective of understanding and characterizing the patterns found in relationships. Investigation on the topology of the network is performed under different perspectives, such as the centrality of the vertices, organization of vertices in communities, the robustness and the diffusion of influence. The results show a strong community structure in the B3 network and, even though the network is fragile for the removal of vetices, the definition of the criterion of vertices to be chosen as a target can be determinant in the characterization of the robustness.

Categories and Subject Descriptors: E.1 [**Data**]: Graphs and networks; H.2.8 [**Database Applications**]]: Data Mining

Keywords: Graph Mining, data mining, stock exchange, B3

1. INTRODUCTION

Modeling complex systems as networks, where nodes represent elements and edges represent their relations, can bring revealing and valuable information in many contexts, such as sociology, biology, transportation and economy. Specifically, in the economic context, many works have been concerned in proposing analysis of the stock market with complex network based models [Huang et al. 2009; Xia et al. 2018; Tabak et al. 2010].

Several papers can be found in the literature with the objective of investigating networks in the stock market, most of them focused on the analysis of the network generated from the stock price correlation [Huang et al. 2009; Wang et al. 2018; Esmaeilpour Moghadam et al. 2019]. Wang et al. [2018] investigate the evolution of the stock market from a global perspective, considering a wide time range of daily stock prices of 57 markets by constructing a correlation-based network and observe that correlation between pairs of stock markets is highly affected by other markets. A more local perspective for the analysis of financial networks is considered by Esmaeilpour Moghadam et al. [2019] and Xu et al. [2017]. Esmaeilpour Moghadam et al. [2019] exploits the correlation of stock proce returns in Iran and the roles of elements by analyzing the centrality of the stocks. The authors observe that stocks with higher capitalization, liquidity and transaction volume are more central. The relation of stock prices is also considered by Xu et al. [2017], however the authors investigate the connectivity between the financial time series, showing its ability in understanding the complexity of financial networks. Stock correlation is also exploited by Huang et al. [2009] in a work in which the authors present a statistical approach to construct a network of stock price correlation, presenting a topological and statistical analysis of the stock market by analyzing the Chinese stock market in order to identify the behavior of the network and its robustness. Tabak et al. [2010] explore the stock market as a graph to identify the most important sectors through analysis of the minimum spanning tree.

Considering the great importance of the Brazilian stock market in Latin America and its influence on economic growth [Caporale et al. 2004], this work uses B3, the official stock exchange of Brazil. This work presents the construction of a weighted directed network considering the common share relations in the stock exchange. The use of common shares, i.e., shares whose owners deliberate on the company's activity, are justified because shares of this class allow a potential influence on the company's decisions, as opposed to preferred shares, whose holders have priority in receiving dividends. Analyzes from different perspectives are conducted, taking into account the centrality of the vertices, the robustness of the network, the division of the network into communities and the propagation of influence.

The main objective of this work is to investigate patterns of organization of the company from the analysis of the relations between companies and shareholders in B3. To do this, the elements investigated are studied from different perspectives, considering several tools of the theory of complex networks, such as vertex centrality analysis, network organization in communities, influence of vertices and the way the network resists with the absence of specific elements. For the definition of the importance of the vertices, classical approaches are considered and in addition, other approaches are proposed, more appropriate to the context of the present work, namely extension and assets. Moreover, the definitions of power, influence and bargain, adapted from the work of Verona et al. [2017], originally proposed for the context of political campaign financing, are considered. The centrality rankings are used to define the seed nodes to simulate the propagation of influence among the elements, considering the Independent Cascade Model (ICM) for diffusion. This study allows us to contrast two notions of influence: as a static measure, intrinsic to the relations present in the network and as a dynamic stochastic process of diffusion on the network.

This research may complement the existing bibliography presenting an analysis focused on the voting shareholders, allowing the formation of more diverse and planned financial portfolios. Therefore, in order to add existing bibliography knowledge about the relationship and control between these relevant public and private agents and the characteristics of the network formed by them, this work presents a topological analysis on the subject.

Another contribution of this work is the analysis of companies and shareholders from the perspective of their influence and the bargain prevailing their relations. The social influence, as defined by Rashotte[Rashotte 2007] is the change in thinking, feeling, acting or behaving of an individual, due to interaction with another person or group. Social influence takes many forms and can be easily seen on online social networks. Nowadays, influence can be seen in cascades of shares, "likes"and retweets, implying that the vast connectivity of social networks lets its users exposed to several sources of influence from different types. In the field of large-scale data analysis, many applications, such as viral marketing, recommendation systems and information diffusion are related to the study of social influence, as Heidari *et. al.* exploit in their study [Heidari et al. 2015].

Assuming that an individual preferences and knowledge have influence on others, a company can promote a new product by focusing on influential individuals, which have the potential to promote the adoption of this product by a lot of users on a network. The viral marketing strategy is based on the selection of a small subset of influential members on a network, and the application of strategies to convince them to adopt a new product. Their social influence makes them to, directly or indirectly, recommend the product to their friends, and may trigger a large cascade of adoptions. Analogously, the same can happen with ideas and behaviors. Therefore, considering the context of this work, one can argue that, since companies are affected by the decisions of common shareholders, it is affected by the relation among them.

This work is an extension of Barbosa et al. [2018], incorporating an analysis of the robustness of the B3 network, enriching the understanding of the relations in the potential situation of drastic lost of nodes in the network. The present work also deepens the discussion about the obtained results and the concepts involved in the proposed methodology, also presenting a wider range of related works.

The remaining of this paper is organized as follows. Section 2 presents the basic concepts related to the work, allowing a better understanding of the results obtained and the discussions conducted. In Section 3, the methodology proposed in the present work is presented for the construction of the B3 network and for the realization of the experiments. Section 4 presents the experiments performed and discusses the results obtained from their execution. Finally, Section 5 presents some conclusions obtained from the realization of this work and points some future directions to the work.

2. BASIC CONCEPTS

This section presents some basic concepts, useful for a better understanding of the analysis conducted in this work. First, some formal characterization is necessary. Consider a network $G = (V, E)$ where V represents the set of nodes and E represents the set of edges, such that $n = |V|$ and $m = |E|$. G can be represented by an adjacency matrix A, where an element $A_{uv} = 1$, if a node u is connected to a node v and $A_{uv} = 0$, otherwise. Vector k represents the degree of the nodes and the degree of a node v is denoted by k_v.

2.1 Communities

One of the most important topological properties in a network is the organization of its nodes into communities, a division of nodes into groups with high internal density and low external density. Community research plays a key role in many contexts, such as sociology, economics and *marketing*. This work considers the partitioning problem, i.e., a vertex belongs to one and only one community. The notion of community becomes more evident as the difference between the number of internal and external edges increases, even though there is no consensual notion for a community. Often, the quality of a community structure is measured with modularity (Q), proposed by Newman and Girvan [2004], that considers the difference between the existing edges in a community and the edges that would exist in a network with the same degree distribution, but with randomly positioned edges, which can be defined by

$$Q = \frac{1}{2m} \sum_{uv} \left(A_{uv} - \frac{k_u k_v}{2m} \right) \delta(c_u, c_v), \tag{1}$$

where k is the degree vector of the network, m is the number of edges, $c(u)$ represents the community of vertex u and $\delta(.,.)$ is the Kronecker delta, which returns 1 if the operands are the same and 0, otherwise.

Many works can be found in the literature with methodologies to identify the community structure that maximizes modularity. In this work, the the method of Newman and Girvan [2004] is used for community identification, more specifically, a high-performance implementation of the Newman method that uses efficient data structures and reduces unnecessary operations in the method, as proposed by Vieira et al. [2014].

2.2 Centrality and influence

An important issue that arises in the analysis of the topological structure of the networks is: which are the most important - or central - nodes in a network? The definition of centrality can be considered from different perspectives [Moore and Newman 2000]. This work considers a wide range of network centralities, encompassing the most popularly definitions described and used in the literature [Newman

2010] and more specifically, in financial network analysis [Esmaeilpour Moghadam et al. 2019; Wang et al. 2018].

The degree is possibly the most basic measure to evaluate the centrality of the vertices of a network and takes into account their number of neighbors. In directed networks, each nodes present an indegree and an outdegree, and both notions may be considered for the characterization of the centrality.

PageRank, proposed by Brin and Page [1998], is an extension of the degree centrality that considers not only the number of connections of a vertex, but also the quality of those connections. Thus, a vertex have a high PageRank if it is connected to other vertices of high PageRank. The PageRank $x_u^{(PageRank)}$ of a node u can be defined as

$$x_u^{(PageRank)} = \alpha \sum_v A_{uv} \frac{x_v}{k_v^{(out)}} + \beta \tag{2}$$

where the nodes v are neighbors of u, $k_v^{(out)}$ is the outdegree of node v, α is a parameter set as 0.85 and β is a parameter set as 0.15.

Yet considering the importance of links in the definition of centrality, HITS was proposed by Kleinberg [1998], using two components to define the importance of a node: *Authorities* and *Hubs*, so that good *Hubs* are pointed out by good *Authorities* and good *Authorities* point to good *Hubs*. The authority $x_u^{(authority)}$ and hub $x_u^{(hub)}$ centralities of a node u can be defined, respectivey, as

$$x_u^{(authority)} = \sum_v x_v^{(hub)} \tag{3}$$

and

$$x_u^{(hub)} = \sum_v x_v^{(authority)} \tag{4}$$

where the nodes v are neighbors of u.

In a different perspective, the measure of *Betweennes* defines the centrality of a vertex as its ability to connect the other vertices of a network. The vertices of greater centrality are those that interconnect many others in a smaller way. In order to define betweenness centrality, we must consider $g_u^{(st)}$ as the number of geodesic paths that start from an initial node s to a final node t and pass through node u. Yet considering that n_{st} is the total number of geodesic paths between s and t, betweenness centrality $x_u^{(betweenness)}$ of an node u can be defined as

$$x_u^{(betweenness)} = \sum_{st} \frac{g_u^{(st)}}{n_{st}}. \tag{5}$$

In addition to the classical measures of centrality, this paper also considers the power measure as proposed by Verona et al. [2017], which aggregates measures of influence and bargain. The authors present an ample discussion, based on Social and Political Sciences concepts to understand the relations of power between entities. One of the authors used as base for this discussion is Castells [Castells 2011], which presents power as the “ability of an agent or organization to control access to strategic resources and impose rules on the other elements”. In a context of systems modeled like networks, this imposition of rules and the accumulation of power, according to Castells, is realized through the use of two mechanisms that are:

(1) The ability to form, control and define a general goal to be achieved by the network;
(2) The ability to connect and secure strategic cooperations of various networks to combine resources and achieve a common goal.

In the work in which they propose the measure of power in networks, Verona et al. [2017] evaluate relations in the Brazilian National Congress through a network that represents campaign financing in the Federal Senate. For this, the authors define the relation of power as the difference between two measures: influence and bargain. Considering that W is a matrix, which represents the edge weights in a directed network, of which an element w_{ij} indicates the weight of an edge (i, j), k^{out} is the vector representing the output degrees of the vertices of the network, and k^{in} is the vector representing the input degrees of the vertices of the network. The measures of influence and bargain between two vertices i and j can be defined, respectively by

$$influence_{ij} = \frac{W_{ij}}{k_b^{in}}, \tag{6}$$

$$bargain_{ji} = \frac{W_{ij}}{k_a^{out}}. \tag{7}$$

Power can, thus, be defined as the difference between influence and bargain:

$$power_{ij} = influence_{ij} - bargain_{ji}. \tag{8}$$

A power relationship, as discussed by Verona *et al.*, becomes more evident from the imbalance between influence and bargain in a relation. A positive value of power indicates that the source vertex has power over the target vertex, while a negative value indicates the inverse. A value close to or equal to zero indicates a balance in the power relationship.

2.3 Influence as a diffusion process

One can expect that the approach proposed by Verona et al. [2017] for the definition of influence in the relations of the National Congress is quite adequate for the definition of relations in the stock exchange, justifying its adoption in the present work. However, traditionally in the literature, the notion of influence in networks has a different connotation than that presented by the authors, being more often associated with the idea of the ability of a vertex to diffuse something through the network. In this way, influence is commonly associated with the idea of propagating some idea or message in a network, not a property that a node holds. In this sense, the present work compares the notions of centrality and relational influence with the traditional notion of influence literature as the effect of a diffusion process that, according to Ghosh et al. [2011], is a dynamic and stochastic process that propagates a particular element over the network.

As previously discussed, this paper investigates the relations of power, influence and bargain proposed by Verona *et al.* [Verona et al. 2017], based on the concepts presented by Castells [Castells 2011]. For this, a comparison of the notions discussed in the cited works with other notions of influences, often explored in the literature of networks as models of diffusion, is performed. The diffusion process in networks, according to Ghosh *et al.*[Ghosh et al. 2011], is a dynamic and stochastic process that spreads a particular element through the network. It is important to note that, in the scope of this work, there may be an ambiguity in the interpretation of the term "influence", which may refer to the measure of characterization of relations, as proposed by Verona *et al.* to the energy propagated in the diffusion process.

In the context of online social networks, social influence can take many forms and be easily noticed in the cascades of likes and retweets, implying that the great connectivity of social networks enables its users exposed to diverse sources influence of different types. Suposing that the preferences of an individual and his ideas may influence others, a new product can be promoted through this individual, influencing others in ther network. In the context of the present work, influence can represent the flow of a guiding idea that is propagated by companies through their shareholders.

In a broader view, to understand spreading processes has a fundamental role in many contexts. From a biological perspective, epidemiological processes for diseases propagation, such as flu and tuberculosis can be analyzed with spreading models. In digital systems, computer viruses propagation has a spreading pattern similar to diseases and can be also considered an epidemic process [Pastor-Satorras and Vespignani 2001].

In the context of the present work, the construction of the network is based on the relations betweens the companies and an edge exists if a shareholding of a company in another can be observed. The shareholding may define a potential influence of a company in another and, in this work, a spreading model is explored in this sense. Chen et al. [2013] states that propagation models consider two basic elements: a network $G(V, E)$ and a function $p : E \rightarrow [0, 1]$ that associates a parameter p_{uv} which captures the influence of u over v. The diffusion process occurs in discrete time steps, and each node may be active or inactive in a certain time step. The activation of a node u indicates that it has adopted a new idea or a new information. The seed set is a group of nodes that is initially activated and the diffusion model defines a randomized process in which the seeds can activate the rest of the nodes of the network.

Two main classes of models for the diffusion of information can be indicated: Independent Cascade Models (ICM) [Kempe et al. 2003] and Threshold Models (TM) [Granovetter 1978]. In Independent Cascade Models the diffusion through each edge is independent from others. On the other hand, Threshold Models considers that the probability of activation of a node is increased from multiple sources.

In this work, the propagation of influence is modeled with *Independent Cascade Model* (ICM), as described by Goldenberg et al. [2001]. The choice of ICM was made due to the fact that this kind of model is frequently used to investigate the propagation of information and ideas [Chen et al. 2013]. Additionally, the network model proposed in this work considers the ammount of common shares that different entities hold and, since the holders participate in important decisions of the companies, this process can be seen from the perspective of information and idea propagation.

The ICM implemented in this work is based on the activation of nodes and the propagation of enegy from vertices and can be briefly described as follows. It starts from an initial set of active vertices, called seeds, chosen according to some convenient criterion. If, at time t, a vertex u becomes active, it has a unique possibility of trying to influence each of its previously uninitiated neighbors v in time $t + 1$. The successful activation of v has a probability $p_{u,v}$. In addition, if multiple neighbors of v become active in time t, their attempts to activate v are sequenced in an arbitrary order, and if one of them is successful in time t, then v will be active in time $t + 1$; however, being u successful or not, it will not be able to make further attempts in the following time steps. The process ends when there are no more activations possible. By the way that the influence propagates in this model, one can notice that its idea is that the vertex is influenced passively, since it is not attentive to the environment, sometimes needing only one attempt to be activated successfully.

Influence maximization is one of the most fundamental problems in the study of social influence and considers that people are susceptible to be influenced by decisions of their friends and colleagues. Marketing researchers have been investigating social influence and its word-to-mouth effect in the promotion of new products, aiming to improve marketing strategies. The problem of influence maximization arises with the following question: "which individuals must be chosen as promising seeds in order to maximize the influence spreading in a network?". In this work, the pproblem of influence maximization is considered from the perspective of the evaluation of the effectiveness of different centrality measures to define seed nodes.

3. METHODOLOGY FOR THE CONSTRUCTION OF B3 NETWORK

This section presents the methodology proposed for the construction of the B3 network and adaptation of the measures of assets, extension, power and bargain for the context of the present work. The values considered as parameters for the realization of the experiments are presented in Section 4.

In this work, the analysis of the relations of companies operating in B3, Brazil's official stock exchange, is performed from the perspective of a network model. For this, it is considered that each vertex i represents an individual or legal entity, that holds common shares in the stock exchange. The edges (i, j) represent the shareholdings between the entities i and j. Because the relation is unidirectional, the edges are directed. The weights $w_{i,j}$ for edges (i, j) represent the volume of the shareholding in the transaction between i and j. The network was generated using the stock composition listing available on the B3 website [1] in December 2017 and the price history between August and December 2017.

All the companies and shareholders present in the listing available on the stock exchange website were transformed into vertices and their shareholding relationships were transformed into edges, and the percentage of participation combined to the average price of a period of five months were used in the definition of the weights. The percentage of participation, considered for the definition of edge weights, was defined by the average share price of the company in the observed period, in order to consider the magnitude of the financial volume traded. In cases where stock history was not available, an arbitrary weight of very small value was defined to indicate that the stockholder relationship exists.

The work considers the use of common shares, represented by code 3 on the stock exchange, because holders of shares of this class can participate in relevant decisions in the company. Preferred shares were disregarded due to the fact that in most corporations shares of this class do not directly influence the decisions of the same.

The weights used to quantify the relationships between companies and shareholders were generated based on three terms: a) Total issued common shares; b) Percentage of common shares controlled; and c) average of the price of the period collected.

Equation 9 describes the weight W_{ij} of a relation between vertices i and j:

$$W_{ij} = \left(\frac{P_{ij}^{on}}{100} * total_j^{on} \right) * proce_j^{on}, \tag{9}$$

where P_{ij}^{on} indicates the fractionof common shares controled by shareholder i in the company j, $total_j^{on}$ indicates the maximum quantity of common shares issued and $price_j^{on}$ indicates the average price of shares of company j in the observed period.

3.1 Power, influence and bargain as metrics for the evaluation of vertices

As discussed by Verona et al. [2017], power is a definition that applies to each relation, i.e., each relation between two elements i and j can be characterized by the influence performed by i on j and by the bargain that j has with i. Power can be observed by analyzing the difference between influence and bargain in that relationship. On the other hand, the methodology proposed by the present is based on the vertices, which represent the shareholders in the stock exchange. The analysis on the relations is made in an indirect way, from the observation of patterns of behavior of the elements. Therefore, the measures of power, influence and bargain have been adapted so that they can be applied on the vertices.

In the context of the present work, the measures of influence and bargain of a relationship between a shareholder i and a company j can be interpreted, respectively, as the influence of a shareholder on the company and the bargain of the company on a shareholder. In the analysis of the stock exchange network, the measure of influence of each vertex was treated as the influence that a vertex has in all its relations. Thus, it can be interpreted as the influence that an element has on the stock exchange as a whole. Similarly, the bargain of a particular individual was treated as the average bargain that

[1] http://www.bmfbovespa.com.br/pt_br/produtos/listados-a-vista-e-derivativos/renda-variavel/empresas-listadas.htm

other individuals have with him. For a company, it can then be interpreted as the bargain of councils in relation to the company. Power remains being calculated as the difference between influence and bargain, but can be interpreted as the imbalance between shareholder influence and bargain of boards on the network as a whole.

In order to use the metrics of power and bargain to analyze the network of the stock market, they were adapted to identify the relation of power in this market. The metrics of political influence and bargain were transformed into the metrics of influence of the shareholder and influence of the board. The shareholder influence component is intended to evaluate the direct influence that a shareholder has on the company board, and the influence component of the board evaluates the inverse influence, i.e. the influence of the board on the shareholder.

$$ShareholderInfluence_{comp.a,comp.b} = \frac{EdgeWeight_{comp.a,comp.b}}{TotalShareholders_{comp.b}}. \quad (10)$$

$$BoardInfluence_{comp.b,comp.a} = \frac{EdgeWeitght_{comp.a,comp.b}}{TotalInvestment_{comp.a}}. \quad (11)$$

$$Power_{comp.a,comp.b} = ShareholderInfluence - BoardInfluence. \quad (12)$$

A positive value in Equation 12 indicates that the investor has some influence over the company in which he is investing, while a negative value indicates that the company board has some influence over the investor.

3.2 Other metrics for vertices evaluation

In addition to the classical measures (degree, PageRank, HITS and betweenness) and the measures of power, influence and bargain, other measures, more specific to the context investigated were considered in this work: asset value and extension.

The asset value quantifies the importance of a vertex based on the market value of the common-class securities traded for that company. The market value of a vertex is then calculated by multiplying the average quotation by the total number of shares controlled by the shareholder.

The extension aims to evaluate the capillarity of the relation of the organizations present in the stock exchange and reflects the size of the closure of a vertex, that is, the number of vertices reached directly or indirectly by it. The application of this measure was motivated by the desire to investigate the number of vertices potentially affected by a possible decision or transfer of resources in a company, represented as a node of the network. It can also be used to identify the depth of a company's relationships within the stock market, i.e. how deep their financial roots are within the stock market.

4. EXPERIMENTS AND DISCUSSION

This section presents a series of experiments carried out with the objective of analyzing the relationship of the companies that operate in B3 through a network built with the methodology presented in Section 3 and discussions based on the results obtained. After a description of the general characteristics of the network constructed, the centrality measures are compared through a correlation analysis. Next, a study of the community structure identified for the nodes of the network is carried out. Considering different measures of centrality, ranks are generated, used as basis for the analysis of the robustness of the network and for a study on the propagation of influence, using a diffusion model.

The construction of the network was computationally implemented in Python, as well as most network analysis. For the general description of the network and the centrality analysis, the igraph library [Csardi and Nepusz 2006] was considered. The robustness analysis, the cascade model and the community detection algorithm were implemented by the authors. It is worth to mention that the community detection consideras a high performance implementation of Newman's spectral method, described in [Vieira et al. 2014].

4.1 General description of the network

The network constructed to model the relations in B3 is directed and weighted. Based on the complete network, it is possible to observe, $n = 2438$ vertices and $m = 2347$ edges. The network has 186 connected components and, in order to make the analysis and the conclusions obtained more significant, the rest of the work has considered only the giant component, that is, the component with the highest number of vertices.

The giant component, considered in the remaining of the work as the network B3, shows $n = 1142$ vertices and $m = 1239$ edges, with an average degree $\hat{k} = 1.08$. Thus, a very striking feature of the B3 network is its high sparsity and the absence of cycles, so that the network can be seen as a tree if the directions of the edges are ignored. The absence of cycles indicates a lack of cross-shareholding, that is, there is no situation in which a Company is a shareholder of Company j and Company j is a shareholder of Company i. One possible explanation for this effect is the fact that these crosses are in practice performed through the use of subsidiary companies that have different names and shares.

4.2 Descriptive analysis of the network

Some data regarding a descriptive statistical analysis of the network degrees is presented in Table I. This data show that the network has a great variation considering average degree and a great difference between the highest and lowest degree in the network.

Table I. Descriptive statistical analysis of the degrees.

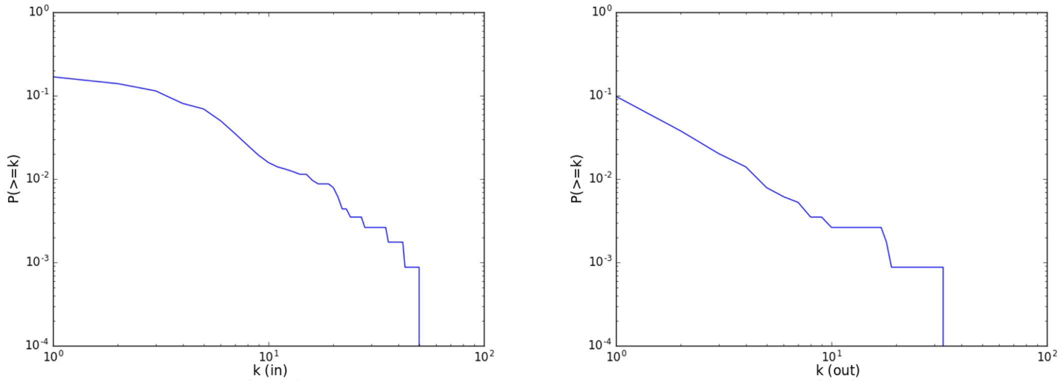
Measure	Value
Mean	2.16
Median	1.0
Maximum degree	51.00
Minimum degree	1.00
Amplitude	50.00
Standard deviation	3.66

The degree distribution is illustrated in Figure 1, which shows the Complementary Cumulative Distribution Function (CCDF) of the indegree (Figure 1(a)) and outdegree (Figure 1(b)).

4.3 Community structure of the B3 network

In order to analyze the community structure of the B3 network, the directed edges were transformed into undirected edges, only eliminating their directions. The community structure was identified by the spectral method of Newman and Girvan [2004], more specifically, using the methodology proposed by Vieira *et al.* [Vieira et al. 2014]. We identified 34 communities, which are organized in a very modular way, with the modularity $Q = 0.8556$. Figure 2 presents a graphical representation of the communities found.

A very clear pattern in the organization of the communities can be noticed, with some nodes more central and others more peripheral, that relate to the central nodes. In order to allow a deeper analysis of the community structure in the B3 network, Figure 3 presents closer snippets of two isolated arbitrary communities, with annotations of company names and stockholders.



(a) Indegree. (b) Outdegre.

Fig. 1. Complementary Cumulative Distribution Function (CCDF) of nodes degrees.

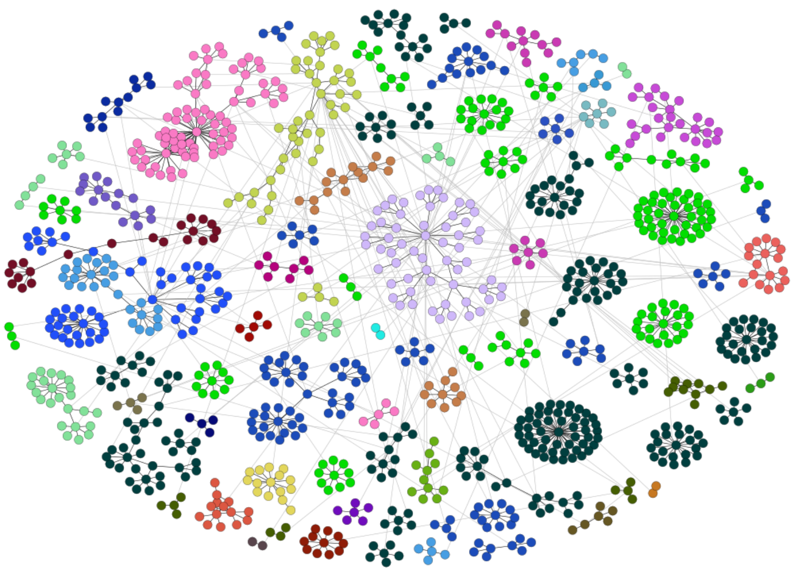
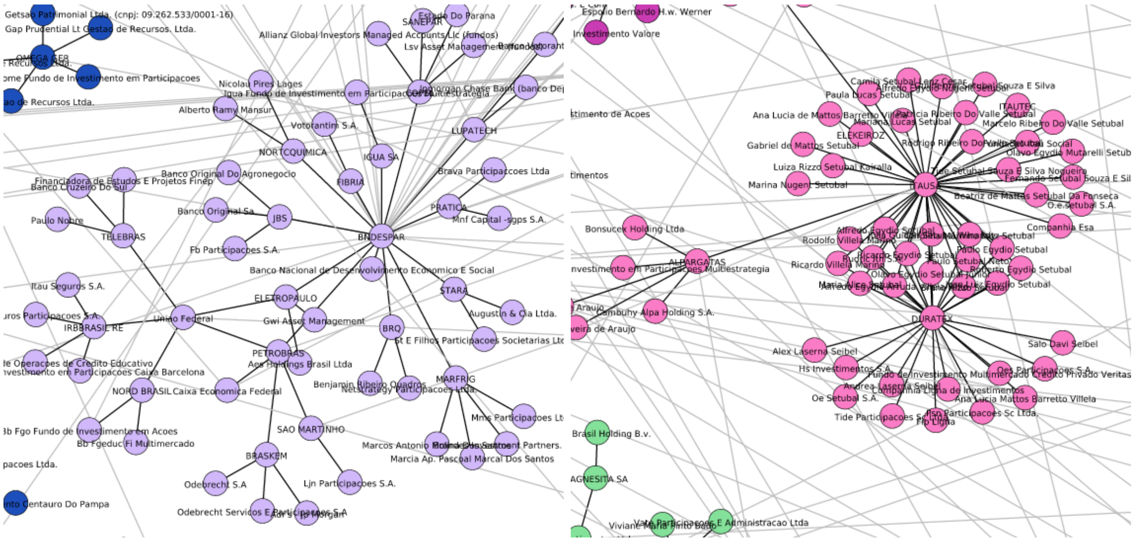


Fig. 2. Division of the network in communities.

In the community represented by Figure 3(a), BNDES can be identified as an important element, serving as the center. Other companies, such as Petrobras, JBS and the Federal Government itself orbit the BNDES and are subcenters of their respective peripheries. In the community represented by Figure 3(b), there is a bipolarization of the centers, which is divided between ITAUSA and DURATEX, located in the same community due to a large number of common shareholders. The vertex that represents the ALPARGATAS is also highlighted and has a series of vertices around itself.

4.4 Characterization of the importance of B3 network vertices

The importance of the vertices in the B3 network, which represent companies and holders of company shares, was assessed using each of the measures presented in Section 3. Some of the most relevant results obtained from this analysis are presented and discussed in this section.



(a) Isolated community 1. (b) Isolated community 2.

Fig. 3. Examples of communities in B3 network.

Considering the concentration of financial assets, it was observed that the financial organizations followed by the federal government and state-owned enterprises are the most outstanding. Among them, Banco Itaú Unibanco has more than R$ 100 Billions in asset value and the Union has more than R$ 140 Billions.

When the relationships are taken into account, the most direct way of assessing the importance of a vertex is through its degree. In the ingoing direction, CEDRO, from the textile sector, which has among its 51 shareholders, 50 individuals. CTC, on the other hand, has, among the majority of its 43 shareholders, companies related to agribusiness. In the outgoing direction, the top of the rank is dominated by companies linked to the financial market. A clear highlight is the BNDES which holds shares of 34 companies, including some of great relevance, such as JBS, Petrobras, Eletrobras and Oi.

Regarding PageRank, it can be observed that the most prominent companies were financial companies such as ITAUSA, ITAUUNIBANCO, MONT ARANHA followed by companies from the logistics and agro-sugar industry. These companies, besides receiving a large volume of investments, also receive investments from important companies. The fact that financial firms are well ranked according to PageRank, it is quite consistent with the intuitive notion that investing in such companies can be strategically critical for other financial firms and other companies, such as mining. The strong investment received by companies in the logistics sector is also quite reasonable, when we observe the strong interest of companies from the most varied sectors. On the other hand, agro-sugar companies, such as the CTC, are well-positioned to be an important technology center on sugar cane, the largest in the world, receiving shareholdings from many other companies related to alcohol production, of fundamental importance in the Brazilian economy.

The application of HITS originates two measures: hub and authority. Analyzing the obtained results, it is noticed that the first rank positions of the best hubs are basically dominated by financial companies such as IUPAR, ITAUSA and Companhia Ligna de Investimentos, which is quite reasonable when considering the nature of the activities of the companies and the mechanism of operation of the measure. Already the top positions of the rank of better authorities present companies like Duratex, Alpargatas and TOTVS, focused on the productive sector. In addition to allowing the relevance of these companies to be seen on the stock exchange, this result also provides a notion of the importance of investments for the operation of these companies.

Applying the classic Hubs and Authorities metrics, financial industry companies and equity and investment companies are prominent. However, there were companies from the agricultural sector, various industries, mining, energy, cellulose and companies in the area of software development. We can highlight that these areas were represented by the companies traded with the appropriate acronyms ITAUUNIBANCO, DURATEX, ALGARPARTAS, ELEIKEROZ, PETROBRAS, BRF S.A, VALE and TOTVS.

The betweenness rank featured, at the top, financial, energy and logistics companies. Using this measure, it was also possible to observe a large participation of state and mixed capital organizations, among which are ELETROBRAS, BNDESPAR and CEMIG. This indicates that such companies can be considered good bridges in the B3 network, having a great relevance as a link between different companies.

In addition, the observation of the capillarity of the companies, quantified by the extension measure, showed that investment banks, financial organizations and the Union itself are the organizations with the greatest reach within the stock exchange.

Some observations can be made from the power, influence, and bargain ranks. First, it is noted that Companhia Paulista de Força e Luz is the one that stands out most in influence and bargain. When the measure of influence is considered in isolation, other energy-related companies, such as Light, and the MG and ES states stand out, showing that the measures taken by these agents have a great potential to have a great impact on the network. On the other hand, the bargain shows in its first positions companies of very diversified fields, such as BIOSEV (agribusiness), IOCHPE-MAXION (automotive equipment), MULTIPLAN (shopping malls) and PRUMO (logistics) . Already the first positions of the power rank are dominated by COSAN and COSAN LTD, with diversified activities in the areas of logistics, energy, financial and infrastructure. In addition, the State of Minas Gerais and other electric companies, such as NEOENERGIA, appear in the top positions.

The generated ranks are quite distinct, although some fields of business are recurrent in the top ranks of the ranks, mainly financial, energy and logistics. In order to compare the centrality measures in the B3 network, Figure 4 presents a correlation matrix between the ranks generated for the different measures. For the construction of the matrix, we considered the Spearman coefficient, quite adequate for the comparison of ranks. Values closer to $1(-1)$ indicate a stronger positive(negative) correlation. Values closer to 0 indicate absence of correlation. The values at each position in the array indicate the Spearman coefficient obtained for each pair of ranks.

In fact, most pairs of ranks show fairly weak correlation. The strongest correlation observed in Figure 4 occurs between the PageRank and Hub measures (0.73), but a more interesting result can be observed in the comparison between the measure of bargain and the PageRank and hub measures (0.55 and 0.59). These correlations, although not very strong, indicate that a vertex that points to important vertices has a good bargain in relations, possibly due to the diversification of their fields of activities, which is quite consistent with the intuitive notion that one can have of the scenario studied.

4.5 Robustness analysis

Considering different measures of centrality, the analysis of network robustness against direct attacks or failures was performed. The purpose of this investigation is to understand the ability of the B3 network to remain operative in the event of, for example, corporate bankruptcies. It is to be expected that, because it is a sparse network, the B3 network is very fragile in relation to failures. Figure 5 gives a summary of the results obtained. The x-axis represents the percentage of nodes removed and the y-axis represents the percentage of nodes that remain in the giant component, thus ensuring the operation of the B3 network. Each row represents the result obtained for each node selection criterion.

From Figure 5, it is noted that, for any criterion of choice of nodes to be attacked, the network rapidly deteriorates. This result is not surprising considering the low density of the B3 network.

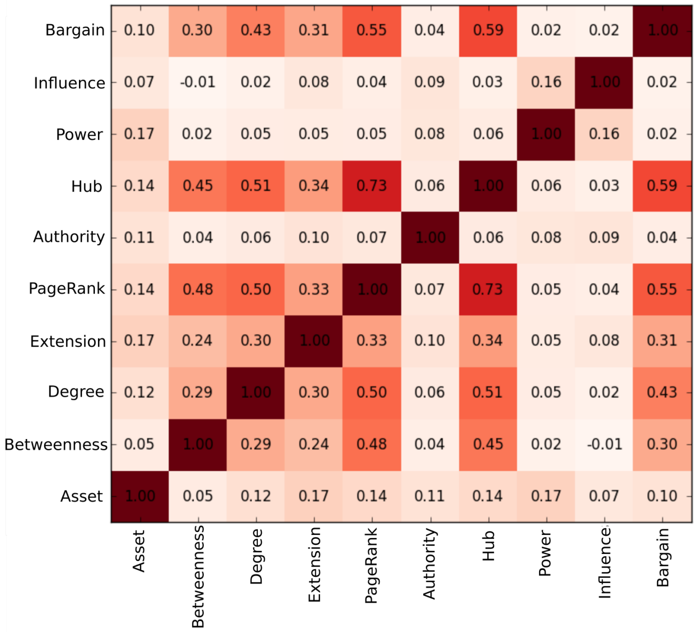


Fig. 4. Correlation of the centrality measures investigated.

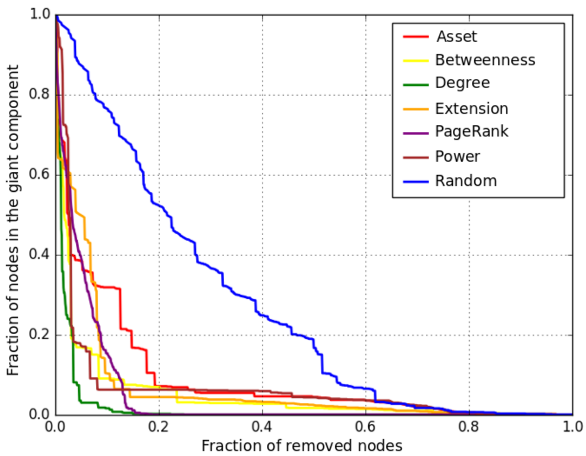


Fig. 5. Robustness analisys on B3 network.

However, it is observed that the attack of higher degree vertices is more effective in characterizing the network than other measures, which can be observed by the rapid reduction of the size of the giant component in relation to the removal of edges when the degree criterion is considered. In contrast, the failure of the most active vertices is the one that has the least potential to cause a network collapse. As a basis for comparison, the random removal of vertices causes a very slow degradation of the network, even though it is extremely sparse.

4.6 Influence analysis on B3 network

The centralities of the nodes were used as a criterion for seed selection in an influence propagation model, in the sense of a diffusion problem: the Independent Cascade Model (ICM). Thus, each measure can be evaluated for its potential in spreading some element, which could be an injection of capital, valuation in shares values or adoption of strategic actions. Figure 6 gives a summary of the results obtained at this stage of the analysis. The x-axis represents the number of seeds considered for activation in ICM. The y-axis represents the number of nodes reached at each execution. Each row represents the result considering a criterion. As it is a non-deterministic algorithm, the ICM was executed 100 times for each criterion for seed selection and each number of seeds considered. It is also important to say that the parameter that controls the probability of activation of a node i in a node j was defined as the weight of the edge (i, j), that is, the share of i in j.

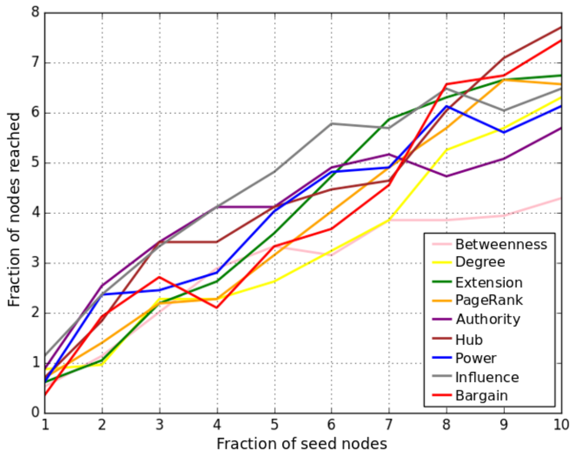


Fig. 6. Influence analysis on B3 network.

Figure 6 shows that, for a seed percentage close to 10, hub and bargain are efficient in defining good seeds. In addition, the influence measure is quite efficient for the definition of good seeds, corroborating the observations of Verona et al. [2017] for its definition.

5. CONCLUSIONS AND FUTURE WORKS

This work presents an investigation of the relationship between companies and holders of shares on the Brazilian stock exchange. The analysis was performed from the point of view of complex networks, with the elements (companies and share holders) being represented by vertices and the shareholdings represented by edges. The development of the work was based on the definition and comparison of different measures of classical centrality (degree, PageRank, HITS and betweenness) and other less conventional ones (power, influence, bargain, extension and assets).

The relationships occurred in the B3 network were investigated from different points of view, considering the direct comparison between several centrality ranks, the community structure, the robustness and the influence propagation. The results obtained evidenced a series of structural patterns in the network, which can serve to better understand the Brazilian stock market.

The results show that there is an enormous importance of investment funds in the stock exchange, regardless of the perspective used to define this importance. In addition, it can be seen that, although the generated ranks always point to companies in the same industry in the top positions, depending on the criteria used, companies may vary.

It is also noted that the indication of the main elements in network B3 is very sensitive to the definition that is adopted to define it. In addition, a strong structure of communities in the B3 network is observed, with a very pronounced pattern of formation: a few central elements and several other elements orbiting around them. It is also noticed that, even though it is intuitively fragile, due to its sparsity, the definition of the criterion of vertices to be chosen as a target can be determinant in the characterization of the robustness.

As future work, it is intended to further investigate the performance of companies in the market, with the aim of better understanding the impact of their decisions on the economy of the country. The intention is also to relate the performance of Brazilian congressmen to the B3 network in order to understand how votes in the National Congress can influence and be influenced by the stock market.

Acknowledgment

The authors thank the agencies FAPEMIG and CNPq for the financial support.

REFERENCES

Barbosa, C. M. G., Felix, L. G. d. S., Xavier, C. R., and Vieira, V. d. F. Investigação da relação entre empresas através da análise topológica de uma rede na bolsa de valores do brasil. In *Proceedings of the 6th Symposium on Knowledge Discovery, Mining and Learning - KDMile 2018*. pp. 28–35, 2018.

Brin, S. and Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* 30 (1-7): 107–117, Apr., 1998.

Caporale, G., Howells, G., Soliman, A. M., et al. Stock market development and economic growth: the causal linkage. *Journal of economic development* 29 (1): 33–50, 2004.

Castells, M. Network theory— a network theory of power. *International Journal of Communication* vol. 5, pp. 15, 2011.

Chen, W., Lakshmanan, L. V. S., and Castillo, C. *Information and Influence Propagation in Social Networks*. Morgan & Claypool Publishers, 2013.

Csardi, G. and Nepusz, T. The igraph software package for complex network research. *InterJournal* vol. Complex Systems, pp. 1695, 2006.

Esmaeilpour Moghadam, H., Mohammadi, T., Feghhi Kashani, M., and Shakeri, A. Complex networks analysis in Iran stock market: The application of centrality. *Physica A Statistical Mechanics and its Applications* vol. 531, pp. 121800, Oct, 2019.

Ghosh, R., Lerman, K., Surachawala, T., Voevodski, K., and Teng, S.-H. Non-conservative diffusion and its application to social network analysis. *arXiv preprint arXiv:1102.4639*, 2011.

Goldenberg, J., Libai, B., and Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* 12 (3): 211–223, 2001.

Granovetter, M. Threshold models of collective behavior. *The American Journal of Sociology* 83 (6): 1420–1443, 1978.

Heidari, M., Asadpour, M., and Faili, H. SMG: Fast scalable greedy algorithm for influence maximization in social networks. *Physica A Statistical Mechanics and its Applications* vol. 420, pp. 124–133, Feb, 2015.

Huang, W.-Q., Zhuang, X.-T., and Yao, S. A network analysis of the chinese stock market. *Phys A: Statistical Mechanics and its Applications* 388 (14): 2956 – 2964, 2009.

Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press, pp. 137–146, 2003.

Kleinberg, J. M. Authoritative sources in a hyperlinked environment. In In Proceedings of the ACM-SIAM symposium on discrete algorithms. *Journal of the ACM* vol. 46, pp. 668–677, 1998.

Moore, C. and Newman, M. E. J. Epidemics and percolation in small-world networks. *Phys. Rev. E* vol. 61, pp. 5678–5682, 2000.

Newman, M. *Networks: An Introduction*. Oxford Univ. Press, USA, 2010.

Newman, M. E. J. and Girvan, M. Finding and evaluating community structure in networks. *Physical Review E: Statistical, nonlinear and soft matter physics* 69 (2): 026113, 2004.

PASTOR-SATORRAS, R. AND VESPIGNANI, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86 (14): 3200–3203, Apr., 2001.

RASHOTTE, L. In , *Social Influence*. American Cancer Society, 2007.

TABAK, B. M., SERRA, T. R., AND CAJUEIRO, D. O. Topological properties of stock market networks: The case of brazil. *Physica A: Statistical Mechanics and its Applications* 389 (16): 3240 – 3249, 2010.

VERONA, L. V., OLIVEIRA, J. O., AND CAMPOS, M. L. Métricas para aná lise de poder em redes sociais e sua aplicação nas doaçõ es de campanha para o senado federal brasileiro. In *Anais do VI Brazilian Workshop on Social Network Analysis and Mining*. SBC, Porto Alegre, RS, Brasil, 2017.

VIEIRA, V., XAVIER, C., EBECKEN, N., AND EVSUKOFF, A. Performance evaluation of modularity based community detection algorithms in large scale networks. *Mathematical Problems in Engineering* vol. 2014, pp. 1–15, Dec., 2014.

WANG, G.-J., XIE, C., AND STANLEY, H. Correlation structure and evolution of world stock markets: Evidence from pearson and partial correlation-based networks. *Computational Economics* vol. 51, pp. 607–635, 03, 2018.

XIA, L., YOU, D., JIANG, X., AND GUO, Q. Comparison between global financial crisis and local stock disaster on top of chinese stock network. *Physica A: Statistical Mechanics and its Applications* vol. 490, pp. 222 – 230, 2018.

XU, R., WONG, W.-K., CHEN, G., AND HUANG, S. Topological characteristics of the hong kong stock market: A test-based p-threshold approach to understanding network complexity. *Scientific Reports* vol. 7, pp. 41379, 02, 2017.