

Using Social Network Analysis to Unveil Cartels in Public Bids

Ademir Cristiano Gabardo*, Heitor Silvério Lopes*

*Federal University of Technology - Paraná - UTFPR

ademirgabardo@gmail.com, hslopes@utfpr.edu.br

Abstract—In recent years, the study of complex networks has attracted great attention. Several fields of science have used techniques of social network analysis and complex networks to represent a wide range of structures such as; social networks, political influence, communication, epidemics and several other aspects of human behavior.

Most of the complex networks show community structures. Revealing these communities is highly relevant to understanding several social phenomena such as the organizing of groups, the flow of information and the strength of the influence of some members over the group.

In this article, we use techniques of social network analysis and complex networks to represent the relationship between companies that are participating in public bids to unveil community structures analog to cartels.

Several nations are facing injuries trough the misuse of public money caused by the formation of cartels, which are groupings of companies aiming to defraud the free competition. Our main goal in this work is to present a methodology for identifying these communities. Furthermore, we aim to address wheter companies that have high success rates in public bids are grouped and identify whether they are taking advantage of their influence in the network.

Keywords—Complex networks; graphs; social network analysis; community detection; clustering; public bids;

I. INTRODUCTION

Misuse of public money is a sensitive issue that generates huge losses to nations. Violation of the rules of free competition hinders or prevents the proper use of public resources. The formation of cartels is one way to ensure that a particular group of competitors gains illicit advantages in public procurement processes. Tracking these groups is a complex task that involves analyzing large amounts of data.

Many structures can be represented by means of complex networks and graphs [1]. We approach the problem of identifying cartels by representing public bids by means of graphs and complex networks. Graphs are mathematical abstractions of networks that can be defined as: $G = (V, E)$ where:

- V is a set of vertices.
- E is a set of edges.

The Brazilian government stores most of its data from public bids with the support of computer programs such as databases and searchable and indexable digitalized documents. Only in 2009, the Federal Integrated System of Financial Administration (SIAFI) registered one billion financial transactions in twenty-four thousand administrative units [2].

It is possible to compute the participation of competitors in public bidding processes to generate complex networks, however due to the large amount of transactions this will result in large networks with thousands of vertices and millions of edges. This characteristic leads to computing intensive tasks. Adopting an appropriate methodology to analyse these networks is critical for obtaining clear results.

These complex networks have communities or clusters, which are cohesive groups or modules of a complex network. Communities are closely attached members of a group that share interests, common relationships or a high level of similarity. The Social Network Analysis classifies clusters as a collection of individuals with dense friendship patterns within the group and sparse friendships outside it.

Several approaches have been developed to pursue optimal results, high accuracy or capacity for dealing with large amounts of data. Despite the large number of algorithms aiming to detect communities in complex networks, we focused our work on the following metrics related to community detection:

- 1) **Clustering coefficient:** Clustering coefficient is a metric used to evaluate the degree to which vertices tend to cluster together. There are two clustering coefficient metrics: global and local clustering coefficients. The first is based on triplets of nodes and measures the number of closed triplets or triangles [3]. The Clustering Coefficient is shown in Equation 1:

$$C = \frac{3 \times nt}{ct} \quad (1)$$

where C is the Clustering coefficient, nt is the number of triangles in the graph, and ct is the number of connected triplets of vertices. The local clustering coefficient is relative to the number of connections to a particular vertex in a graph, the proportion between the number of connections to a vertex and the total number of possible connections between the vertex and its neighbours [4].

Directed and undirected graph clustering coefficients are distinct. In this work we are interested only in undirected graphs and thus only the local undirected clustering coefficient formula is presented in Equation 2:

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{K_i(K_i - 1)} \quad (2)$$

where C_i is the local clustering coefficient, K_i is the number of vertices in the graph, v_j and v_k represents the vertices in the graph from j to k , and N_i represents the neighbourhood of a vertex which is defined by its immediately connected neighbours.

- 2) **Modularity:** Modularity is a quality index for clusterings. Its objective is to evaluate the division of a network into modules [5], [6]. Modules will have dense connections between the nodes within modules but sparse connections between nodes in different modules. The modularity of a graph can be computed using equation 3:

$$q(\zeta) = \sum_{C \in \zeta} \left[\frac{|E(C)|}{m} - \left(\frac{\sum_{v \in C} \deg(v)}{2m} \right)^2 \right] \quad (3)$$

where $q(\zeta)$ is the modularity of a clustering C , $m := |E|$ represent the edges, $\deg(v)$ represents the degree for a vertex and the term $\frac{|E(C)|}{m}$ is known as coverage.

Clustering coefficient and Modularity are complex network metrics based on network topology used to reveal the presence of communities. We have used these metrics in addition to data dimensions, such as the number of participations and number of victories, to depict the success of these groups of companies that can be identified as analogous to cartels.

II. METHODOLOGY

A. Using Social Network Analysis to detect cartels in public bids

It is possible to consider the organization of the relationships between companies that participated in public bids as a complex network, and the interrelationships between participants can also be seen similar to a social network. These complex networks are organized as follows:

- 1) Each company represents a vertex of the graph.
- 2) If two or more companies share the same bid, these connections are turned into edges.
- 3) The repetition of connections are computed as edge weights.

In our approach, we track the connection repetitions over time to determine the edges weights. The concept is very straightforward, based on how many times a group or community appears fully or partially in the network. This metric provides an index of strength for the cohesion of the group. If a group receives a large number of members or loses members frequently the connections will be noticed as weak; on the other hand, if the same members are constantly clustered, the connections will be seen as strong.

Whenever two different companies participated in the same bidding process, this was taken into consideration for building this metric. The value in the adjacency matrix is set as 1 for participants; otherwise it is set at 0. Self loops are always set at 0. An example of an adjacency matrix for public bids from 1 to 4 and companies from A to D is shown in Table I.

TABLE I
ADJACENCY MATRIX OF PUBLIC BIDS 1 TO 4 FOR COMPANIES A TO D

	Bid 1	Bid 2	Bid 3	Bid 4
Company A	0	1	1	1
Company B	1	1	1	1
Company C	0	0	0	1
Company D	1	1	1	0

The total number of participations between companies is then added to obtain an undirected weighted graph $G = (V, E, w)$, where w means the weight of the edges. The resulting adjacency matrix with the weights for the presented example is shown in Table II.

TABLE II
RESULTING ADJACENCY MATRIX.

	Company D	Company C	Company B	Company A
Company A	2	1	3	0
Company B	3	1	0	
Company C	0	0		
Company D	0			

Of the topological aspects of the complex networks, two dimensions of data were attributed to the vertices to build a success rate. By computing the number of victories and the number of participations, it is possible to obtain the success rate as shown in equation 4:

$$Sr = \frac{\sum_{i=1}^{\rho} \vartheta}{\rho} \quad (4)$$

where Sr is the success rate, ρ is the total number of participations in public bids and ϑ is the number of victories in public bids. This leads to values between 0 and 1 which correspond to the percentage of victories for each company. For simplification, the success rate Sr was discretized according to the criteria shown in Table III.

TABLE III
DISCRETIZED VALUES FOR SUCCESS RATES

Continuous Values	Discrete Values	Colour	Sample
$\geq 70\%$	Winner	Black	●
$30\% > 70\%$	Average	Gray	●
$\leq 30\%$	Loser	White	○

The success rate and the degree of the vertices are used to evaluate the assortativity, or assortative mixing, of these

complex networks. The assortativity demonstrates that network vertices tend to connect to vertices that are somehow similar [7] [3]. In social networks, vertices with a high degree tend to connect with high degree vertices [8]. By evaluating the assortativity of the network, we aim to answer practical questions such as:

- Are the winners taking advantage of those that always lose?
- Are the winners clustered together?
- Does a company with more participation enjoy higher success rate?

In addition to these concepts, for visualization purposes, the size of the vertices reflects the number of participations. The success rate is represented by the color, and the number of repetitions represents the strength of the connection between two companies. This representation was chosen to provide visual clues about the organization of the complex network for audiences not familiar with concepts of complex networks. The resulting graph for data presented in Table II is shown in Figure 1.

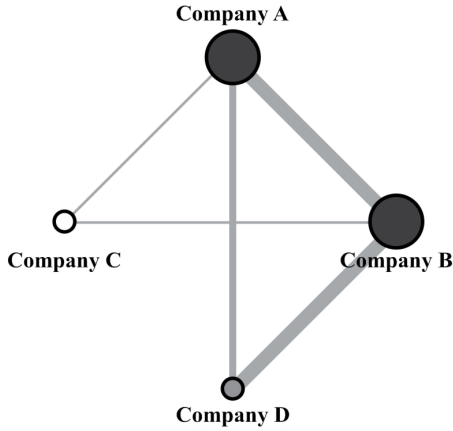


Fig. 1. The resulting undirected weighted complex network relative to the data in Table II.

Although this representation is driven by very simple concepts, the final result provides an the opportunity to visualize the complex network in a pleasant aesthetic way. It is intuitively simple to observe that the connections between Companies A, B and D are stronger that the connections with Company C. The vertex degree, reflected by the vertex sizes suggests that the most relevant vertices are A and B. Therefore, the final perception is that the most relevant interaction in this complex network is between Companies A and B.

B. A strength measure based on repetitions over time

This section provides further details of the strength metric based on repetitions over time. Most techniques employed to study and analyse complex networks are mainly designed for static networks and generally fail to capture the evolution of phenomena and their dynamical properties and temporal dimension, focusing instead on structural or statistical aspects

of the system [9]. However, it is important to consider that social networks are dynamic and that connections can change over time.

Correlations of strength and intensity in complex networks overt time have been studied for different purposes [10]. In this work we propose the following strategy to capture the evolution of a complex network. Given an undirected weighted graph $G(V, E, w)$, a formula for calculating the weight of a connection between the vertices (x, y) over time is shown in Equation 5:

$$w(x, y) = \frac{\sum_{i=1}^t w(x, y)}{t}, \quad (5)$$

where t is the number of changes in the whole graph and not only for a specific vertex. Therefore, the number of interactions of each vertex will be re-computed to calculate the edge weights. For the groups of vertices that remains connected over time, the weights will remain unchanged. For the groups that lose connections, the weight will decrease.

The variable t can be set as other metrics, such as time, days or recurrence. As a result of this technique, communities which keep connections over time will be noticed more than communities that receive new members or that frequently lose members. Figure 2 illustrates the process of measuring strength based on repetitions over time.

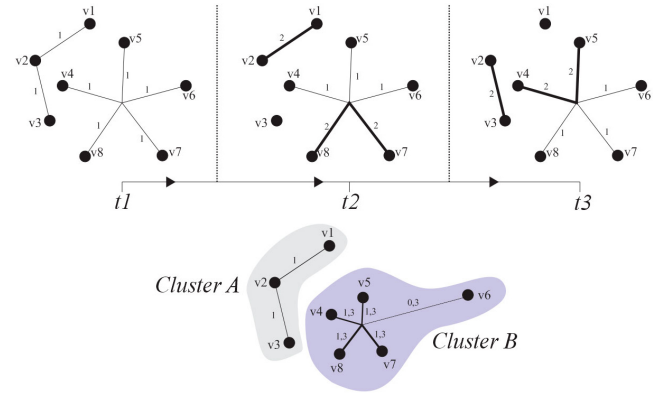


Fig. 2. Clusters highlighted within a complex network based on strength metrics of repetitions over time.

In Figure 2, three different moments in time are shown where some of the connections between vertices have changed. By applying the concept of strength the communities, *Cluster A* and *Cluster B*, were observed.

After following these steps an undirected weighted complex network resulting from N interactions in N time periods is available for clustering.

This task is done in two steps. The first is simply to group the individuals of the complex network according to a criterion such as the success rate or number of participations. This simple clustering can be done with the K-means algorithm [11] or one of its variants. Since the number of partitions is

already known is easy to cluster the vertices, for instance, by the groups of winners, average and losers.

The second step is to detect the communities according to the network topology by computing the Clustering Coefficient and the Modularity of the network.

Finally, we used the Fruchterman and Reingold [12] force-direct layout algorithm to produce an aesthetically and comprehensive graphic representation of the complex networks.

III. EXPERIMENTS AND RESULTS

In 2011, Paraná State in Brazil conducted 21,878 public bids, with a total of 41,385 participations by 15,955 distinct companies. The state of Paraná is a unit of the Federation of Brazilian States, composed of 399 municipalities. This state is also divided geographically into 10 meso-regions. For the case study we selected the public bids for construction and engineering services in the metropolitan region of Curitiba, which is also the capital of the state. The resulting complex network for these public bids has 544 vertices and 3129 edges, as shown in Figure 3.

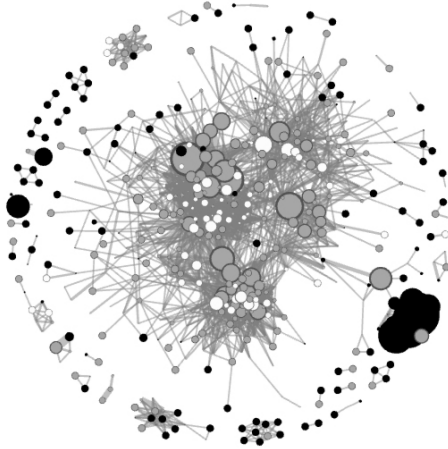


Fig. 3. Complex network from public bids for construction and engineering services in the metropolitan region of Curitiba in 2011.

The degree distribution of a complex network provides a glimpse of how a network is organized in terms of topology. Real world networks often follow a power law degree distribution [8]. Figure 4 shows the degree distribution for the complex network from public bids for construction and engineering services in the metropolitan region of Curitiba in 2011.

In addition to degree distribution, we also analysed the clustering coefficient, computing the number of triplets within a complex network. A high clustering coefficient suggests the community structures in complex networks. The complex network shown in Figure 4 has an average clustering coefficient equal to 0.521. The clustering coefficient was computed using the algorithm of M. Latapy [13].

It is already known that cartels are small, cohesive and strongly connected groups of companies [14] [2] [15]. By computing the modularity with the algorithm of Blondel et al

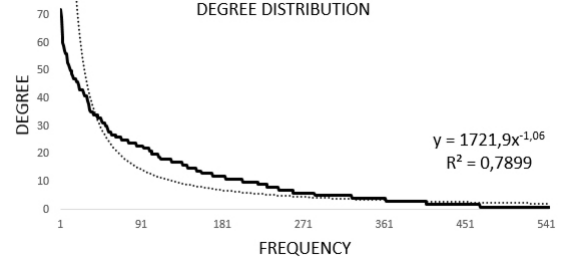


Fig. 4. Degree distribution for complex network shown in Figure 3.

[16] we were able to identify 44 distinct modularity classes. There are four modularity classes within the network that are too large to be considered a cartel, representing 26.10%, 15.62%, 14.34% and 12.32% of the network dimension, respectively. The rest of the complex network is made up of 40 smaller groups, as shown in Figure 5.

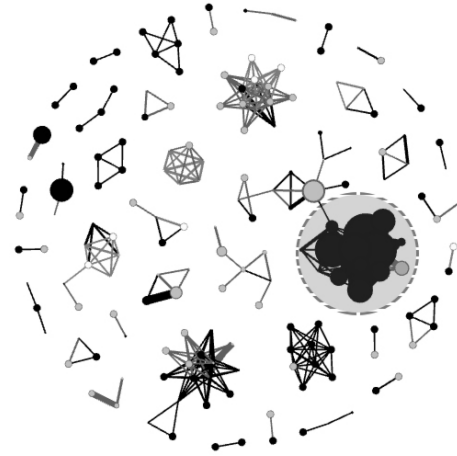


Fig. 5. The forty small community structures in the complex network shown in Figure 3, with one group highlighted.

We took the community highlighted in the complex network shown in Figure 5 and labeled its members as A, B, C, D, E and F. For this community, we analysed the number of participations, number of victories and the success rate. These data are shown in Table IV.

TABLE IV
TOTAL DATA FOR A PARTICULAR CLUSTER

Company	Participations	Victories	Success Rate
A	2	0	0%
B	7	2	29%
C	8	2	25%
D	2	1	50%
E	5	0	0%
F	5	0	0%
Average			17%

Taking into account that companies B, C and D won public bids in 2011 this triad was selected for further analysis. Hereafter, the time frame was increased for all available data, from 2005 to 2012 in search of all public bids where the B, C and D triad were present. The whole group participated in 26 public bids, with Companies A, E and F partially present. Table V contains the sum of all the information regarding these 26 public bids ¹.

TABLE V
TOTAL DATA WITH NUMBER OF VICTORIES FOR COMPANIES A TO F FROM 2005 TO 2012

	Victories
Company A	0
Company B	5
Company C	5
Company D	11
Company E	4
Company F	3
Sum of victories	28
Number of public bids where at least one member of the cluster wins	19
Success rate for the group	73%

Compared to the data shown in Table IV, it is possible to observe a significant increase from 17% to 73% for the success rates of the companies in this group in comparison with the rates observed for the same companies individually. The company labeled as A never won, suggesting that its participation is favored others within the group. The most successful companies were B, C and D. Several similar groups were analysed with similar results.

IV. CONCLUSION

A recurring and sensitive problem for governments is how to ensure fair competition in public bids. In this article we used our methodology to detect possible cartels operating in public auctions in Brazil. We employed techniques of social network analysis to detect groups of companies that indicating the possible formation of cartels.

This study focused on two major points. The first was the development and use of complex network approach to deal with the relationships between companies participating in public auctions, thus offering a platform for analyse these data through Social Network Analysis. The second was the creation of establishment of a methodology with the proper technique and algorithms that provide a comprehensive framework for complex network analysis.

With our experimentations, we found several groups of companies whose composition and actions suggest the formation of cartels. The same methodology can be applied to trace political influence, social organization and several other

aspects of human behavior. The main innovative aspect of our study is related to approaching complex networks respecting their dynamic features, computing several moments over time instead of focusing on static snapshots.

ACKNOWLEDGMENT

The authors would like to thank to the Court of Accounts of the State of Paraná - Brazil (TCE/PR) for providing the data for this study.

We also wish to thank our colleagues at the Bioinformatics Laboratory at the Federal University of Technology of Paraná who shared valuable knowledge and many cups of coffee.

The authors would also like to thank the Federal University of Technology of Paraná for its support.

REFERENCES

- [1] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [2] C. V. S. Silva and C. G. Ralha, "Detection of cartel formation in government biddings using data mining agents," *Revista Eletrônica de Sistemas de Informação*, vol. 10, no. 1, pp. 1–8, 2011.
- [3] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [4] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [5] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.
- [6] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On modularity clustering," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 2, pp. 172–188, 2008.
- [7] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.
- [8] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [9] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, "Time-varying graphs and dynamic networks," *Computing Research Repository*, vol. abs/1012.0009, no. 1, 2010.
- [10] M. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, p. 1, 1973.
- [11] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Procedures of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1, 1967, pp. 281–297.
- [12] T. M. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.
- [13] M. Latapy, "Main-memory triangle computations for very large (sparse (power-law)) graphs," *Theoretical Computer Science*, vol. 407, no. 1, pp. 458–473, 2008.
- [14] Brazil, *Fighting Cartels: Brazil's Leniency Program*, 3rd ed. Bureau of Economic Law, Antitrust Division Council for Economic Defense, Ministry of Justice, 2009, vol. 1.
- [15] M. C. Levenstein and V. Y. Suslow, "What determines cartel success?" *Journal of Economic Literature*, vol. 44, no. 1, pp. 43–95, 2006.
- [16] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

¹Note that a single public bid allows for multiple winners, which is why the sum of victories is higher than the number of bids.