NOVA IMS Information Management School

MDSAA

Master's Degree Program in

Data Science and Advanced Analytics

Communities' susceptibility to collusion in the Portuguese Procurement System

An analysis of communities generated in a co-bidding firm-to-firm network

Diogo Bulhosa da Cruz Pinto

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Communities' susceptibility to collusion in the Portuguese Procurement System
An analysis of communities generated in a co-bidding firm-to-firm network

by

Diogo Bulhosa da Cruz Pinto

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics

Supervised by

Flávio L. P. Pinheiro (PhD) NOVA IMS - Information Management School

Bruno M. P. Damásio (PhD) NOVA IMS - Information Management School

November, 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 01/11/2023

ACKNOWLEDGEMENTS

I would like to thank Prof. Dr Flávio Pinheiro for his guidance and my family and friends for all the support given during the development of this thesis.

ABSTRACT

Public procurement plays a crucial role in economic systems, by shaping infrastructure development and delivering essential public services. However, the susceptibility of these systems to corruption poses a grave threat, potentially leading to substantial economic losses. Firms can exploit these vulnerabilities by orchestrating cartels to manipulate and control the market dynamics. Recognizing the importance of such risks, we acquired data from contracts within the Portuguese public procurement system. Using this dataset, a firm-to-firm co-bidding network was constructed, to which the Louvain Algorithm was applied to identify distinct communities. Subsequently we analyse these communities employing diverse methods such as Diversity Simpson's Index, a set of Network Measures such as Modularity and Average Clustering Coefficient, Topological Metrics like Coherence and Exclusivity, and Performance Metrics related to contract awards and final values. With this analysis we aim to characterize and evaluate the susceptibility of the Portuguese Procurement System to collusion.

KEYWORDS

Public Procurement; Network Science; Louvain Algorithm; Portugal

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF FORMULAS

1. INTRODUCTION

Public Procurements are defined as the process by which public authorities purchase goods, services, and processes from companies, representing an important economic activity of governments. According to the OECD (2017), the ageing European populations and the increase in dependency ratios will affect the demand for public services, turning the capacity to deliver them into something to be concerned about. Public sector productivity has a significant impact on the performance of the national economy and societal well-being and can only be improved by either creating more resources or by increasing productivity through the public sector (OECD, 2017).

According to data from the Annual Report of Public Procurements 2021, the contractual value from this past year was of 10 thousand million euros, representing 5.36% of the national GDP. This significant value is not only witnessed in Portugal, but everywhere else in the world. In 2018, Public Procurements represented around 12% of the Global GDP.

Taking into consideration the importance of these procurements to society, and also the amount of public money involved, there is a demanding concern to avoid favouritism in their allocation. Within a well-functioning democracy, particular groups of citizens should not improperly influence neither legislators, legislation nor the specific allocation of public funds (Gilens, 2012; Hodler and Raschky, 2014). When such undue conditioning is allowed, spending allocations are detached from their regulating principles, allowing for lower social welfare and an increase in inequalities across social groups (Bramoullé and Sanjeev, 2016).

The main idea behind this paper is to create a firm-to-firm co-bidding network using social network science and raise flags for particular suspicious activities that can be linked to cartels inside the procurement system or corruption activities used by companies in order to be more successful.

The remainder of this paper is organized as follows. Section 2 thoroughly explains the concept of Public Procurements, Big Data and Network Science, and lastly Related Studies. Section 3 presents the obtained Data and displays its exploratory analysis and cleaning. It also demonstrates how the co-bidding network is created and filtered. Section 4 elaborates on the analysis of the network and the subsequent communities created, and in the end presents a little discussion crossing the results. Lastly section 5 offers a conclusion and suggests venues that could be explored in future works.

2. LITERATURE REVIEW

2.1. PUBLIC PROCUREMENT

As previously stated, Public Procurements can be a critical factor to the development of a society. Not only Portugal but several countries have converged into the usage of Public Contracts to provide better conditions and an economic growth of their businesses. Winning government contracts can boost firm growth in the short and medium-long term. This is demonstrated by Ferraz et al., (2015), who found that firms in Brazil that won government contracts experienced higher employee growth, were more likely to participate and win future auctions, and expanded their business into new markets. This tendency was mainly noticed in highly developed countries, which can be justified by the needs of building and supporting big public infrastructures (Esteban Ortiz-Ospina & Max Roser, 2016). Public Procurement not only helps large entities completing large state projects but also small entities in participating in them (or other small ones). By addressing smaller businesses, the necessary financial help, it opens a new market for their development and promotes their innovation (Preuss, 2011). Taking this into account Portugal has also invested in this system, and in just one year, from 2020 to 2021 there was a 3.3% increase in public procurements. (Public Procurement Indicators 2018)

The intricate interplay between public and private interests, coupled with the abundance of resources involved, makes public contracting processes susceptible to various risks, such as collusion and corruption. Corruption is a transnational phenomenon that affects all societies in deep and multiple ways, at their political, economic, ecological, and social fronts (United Nations Convention Against Corruption, n.d.). According to the 2014 OECD Foreign Bribery Report there is an estimation that bribery consumes 10.9 per cent of the total transaction value in public procurement globally (OECD Foreign Bribery Report, 2014).

Considering the critical importance of public procurements and the mandatory need to combat corruption in them, governments are continuously striving to improve their procurement systems, with the aim to achieve greater transparency and efficiency as well as attain a broader socio-economic policy (Llano et al., 2022). These factors have caused a significant evolution in public procurement, with the most significant change being the transition from paper-based to digital systems. Although it is hard to expose corruption, it has been shown that e-procurement has great potential in helping uncover cases by improving transparency in various contexts (Bertot et al., 2010). Transparency leads to a more open and exhaustive exploration of these contracts, which leads to a reduction in inefficiency and corruption (Lewis-Faupel et al., 2016) .In 2004, e-procurement was introduced to make public procurement more accessible. This shift was driven by the rapid pace of technological advancement, which necessitated the modernization of public contracts. In 2009, Portugal's version for e-procurements was released, with the name of base portal- https://www.base.gov.pt/. The portal is a central repository for information about public procurement procedures in Portugal which are mandatorily digital under the Portuguese Code of Public Procurement (CCP). It provides up-to-date information on the formation and execution of contracts, making it easy to track and monitor them (IMPIC, n.d.). Although e-procurement platforms have limitations that constrain the evaluation of the impact of public procurement on the creation of public value (Cunha et al., 2019), the creation of BASE has shown a positive impact in the overall openness to research and development of the public procurement system in Portugal (Roriz & Ministro, n.d.).

2.2. BIG DATA AND NETWORK SCIENCE

Big Data has been a prominently growing area of technology since the beginning of the 21st century. Philip Chen & Zhang, (2014) defined Big Data using the three V's: volume, velocity, and variety, while other sources add one or more V's. A relevant addition to this was the V for value, defined by the World Economic Forum, which defended that value was, in the end, what everyone was striving to obtain from this technology (Bilbao-Osorio et al., 2014). The relevance of Big Data can be easily explained by the number of areas it can be applied to with great advantages. Several studies have surfaced in distinct social and economic subjects that use this technology such as biological science and health (Fassio et al., 2020; Herland et al., 2014; Zheng et al., n.d.), urban mobility and development (Neves et al., 2020; Szell, 2018) , resource management (Li et al., 2023), government, and others.

Just like the research on public procurement policies remains at an early stage (Patrucco et al., 2017), the usage of Big Administrative Data to understand and act on corruption is also a very recent complement to this field. Despite being recent, it became a key factor to deviate public procurement studies from empirical studies of corruption to an actual exploratory analysis of real-world data.

Throughout the past years different approaches to this newly available data have been explored to fight bid-rigging, one of the most influential methods of corruption in the Public Procurement space. Collusion (or bid-rigging) in auctions, occurs when businesses that would otherwise compete between each other to win tenders, compromise this competition by secretly raising prices or lowering the quality of goods or services for buyers. By doing so, companies save or free up resources for other uses. (Capobianco, n.d.). This undermines the effectiveness of this system, which causes a roadblock to economic growth (Gallego et al., 2021; Mauro, 1995) and political development (Mungiu-Pippidi, 2015).

As it is argued by Reeves-Latour & Morselli (2017), the field of bid-rigging can be grouped into three general approaches: Firstly, there is the approach by industrial economists and auction theorists who immerse themselves in the complexities of market structures and procurement dynamics, delving into the nuanced underpinnings of bid-rigging schemes. Simultaneously, they engage in the development and refinement of indicators aimed at enhancing the efficacy of cartel detection mechanisms. Secondly, criminologists and sociologists frequently direct their attention toward the fundamental institutional and structural factors that give rise to and sustain antitrust offenses within private organizations and industries. Their focus extends beyond the surface manifestations, aiming to comprehend the deeper societal and systemic forces that contribute to the occurrence and continuity of such offenses within these sectors. Thirdly, political scientists focus on the more theoretical side, by focusing on illegal public contracts and unravelling the intricacies of the regulatory frameworks and processes that may inadvertently facilitate or be exploited in instances of corruption within public contracting.

For this study we will mostly focus on the first two categories, as these are highly correlated to building a social-network model of firms.

Before tackling specific studies that contributed to the Public Procurement development and improvement using Network Science it is important to highlight how these networks have impacted several other research fields.

A network emerges from the interaction between nodes. These nodes can be associated to friends, species or, like in this study, firms. Their interactions develop patterns, statistics and cues of observation that allows us to analyse and better understand communities that can emerge from them. Network Science has been used to create co-citation networks (Gmür, 2003; Nai et al., 2022), language networks (Ronen et al., 2014), disease networks (Shirazi et al., 2019), a movie actor collaboration network (Newman et al., 2001), and others.

2.3. Related Studies

The use of network science to valuably understand corruption is not new (Lauchs et al., 2011; Rostami & Mondani, 2015). Several studies have been made in order to explore the inner circles formed inside the Public Procurement system.

It is argued by Lyra et al., 2021 that there are two main groups of studies using network science upon the e-procurement system.

In the first group, we have works that focus mainly on the bipartite networks, created by the relationships between Firms and Tenders or Contracts. Wachs et al. (2019) explores several datasets of public procurement systems across Europe, by creating a bipartite tender-firm network. Observing the network behaviour by using descriptive statistics, such as the network density, Robins-Alexander clustering and other methods, this study finds local correlations in the markets and shows that their network diverges from a random behaviour. Another insightful study exploring bipartite networks is the work of Tóth et al. (2015). By exploring a bipartite network of Firms-Tenders based on the Hungarian procurement system, it tries to present an applicable toolkit of collusion risk indicators.

The second group, focus mostly on firm-to-firm unipartite networks (Cheng et al., 2017; Leiva et al., 2020; Lyra et al., 2021; Reeves-Latour & Morselli, 2017a; Wachs & Kertész, 2019; Xiong et al., 2021; Zhu et al., 2020).

In the Wachs et al. (2019) study, after exploring the bipartite network, they focus on the unipartite nature of firm-to-firm relationships in the Hungarian procurement system. Using a K-Shell approach (Garas et al., 2012), a Core-Periphery analysis is developed, which allows them to better understand the core firms of the procurement systems and take the side of the debate claiming that centralization facilitates corruption.
Reeves-Latour & Morselli, (2017b) also takes a similar approach, focusing on Core-Periphery analysis. Here they add a continuous mode that allows for a semi-peripheral core of bidding

actors (Rombach et al., 2012). In this study it is also considered that this Core-Peripheral analysis can be complemented by introducing other measures in order to screen irregular bidding patterns more accurately. Contract share and successful bidding rate of the core firms were called performance measures. Relying on the Coefficient of Variation, it was assessed how these measures, were distributed among bidders and across time. The combination of these two strategies, allowed this study to describe two types of markets: competitive markets (low heterogeneity and more dispersion in performance features across core bidders (high CVs)) and closed markets (higher level of bid concentration among the same subset of firms (high heterogeneity), similarities in winning patterns (low CVs)).

Besides Core-Periphery analysis, other methods have been involved in trying to understand how communities (not only consisting in core members) are formed inside these networks. Zhu et al. (2020) demonstrates how there are communities evolving the Chinese Construction Projects. Using a fast modularity optimization method referred to as BGLL (Blondel et al., 2008) and a Cluster Percolation Method in combination with network measurements, this study creates a three coloured code to warn of potential levels of bid collusions and formulates a differentiated response strategy for government regulatory bodies to employ in industry collusion monitoring.

Other studies such as Cheng et al. (2017) and Lyra et al. (2021), approach community formation using modularity-based algorithms. The first, uses the Givan-Newman algorithm (Girvan & Newman, n.d.) to explore a co-bidding firm-to-firm network with data from the Tianjin Water Conservancy Transaction Management Centre. The second uses the Louvain algorithm to explore and identify communities of firms in the State of Ceára, Brazil. The later, also shares an implicit strategy to measure the common behaviour of companies in a network with Wachs & Kertész (2019). Both studies use the Jaccard Similarity index in order to distinguish strong and weak links between companies.

3. DATA

This data was obtained by accessing the previously mentioned public procurement portal, BASE. Using web scraping techniques, information was obtained on all contracts from 2009 to 2022, resulting in a total of 1.399.315 contracts.

To guarantee a quality dataset for the further analysis of the Portuguese procurement system, we began by carrying out a thorough exploratory analysis and preparation of the data.

3.1. Data Quality and Cleaning

Composing this dataset, there was a great number of details about the contracts. The Contract Id, the firms that contested the contract (detailed with NIFS and name) and the respective winner, the organiser, values, dates, and other more specific details (i.e. reasons for a contract to have the delivery date delayed). From this large number of columns, only the ones considered important were left out: Contract ID; Contestant NIF; Organizer NIF; Contract Winner; Procedure Type; Initial and Final Price; Award Per Winner; Contract Year; Location and CPV Division.

Table 1. Summary of Variables

Variable Name	Description
Contract ID	Contract identifier
Contestant Name	Company name
Contestant NIF	Company identifier
Name Contracted	Name of contracted company
NIF Contracted	Contracted company identifier
Contract Type	Specifies the type of contract
Invitees	Companies invited to participate in the contract
Centralized Procedure	A single contracting authority is responsible for the procurement
Procedure Type	Specifies the type of awarding procedure

Name Contracting Agency	Name of the contracting authority
NIF Contracting Agency	Contracting agency identifier
Publication Date	Date when procurement notice or invitation to tender (ITT) is published
Close Date	Date when the contract was finished
CPVS	Product Code of the given tender
Execution Deadline	Until when should the contract be terminated
Execution Location	Location of contract competition
Initial Price	Initial price agreed in the contract
Final Price	Final price that the contract ended up costing
Signing Date	Date when the company was hired
Environmental Criteria	If they have environmental criteria or not.
Material Criteria	If they have material criteria or not.
Causes Deadline Change	Justification of deadline change
Causes Price Change	Justification of price change
Document ID	Some contracts can have documentation annexed. Identifier and Name of those documents.
Document Name	
Contract Year	Year the contract

Although the database had no missing values in these main columns some data preparation was still necessary.

Firstly, we started by looking at what composed each categorical data. In several categories, there were several variable-level problems. Some of these were easily solved as there was a large discrepancy in terms of the number of contracts in which they occurred, resulting always in a small percentage of contracts eliminated from the dataset.

One of the main problems of the dataset is how data is stored. In most of the categorical values it is easy to observe that these are very dense in terms of categories. It was urgent to dilute these categories into more specific and less spread ones.

For Procedure Types a straightforward approach was used. We were presented with a high number of types, but this time, each contract was only associated to one procedure. The discrepancy between the number of contracts associated with each category was very high, so in the end we kept the most present types in the dataset:

Table 2. Summary of Procedure Types

Procedure Type in Data	Procedure Type English	Description
Ajuste Direto Regime Geral	Direct Adjustment General Regime	The contracting authority directly invites a chosen entity to submit a proposal.
Concurso público	Public Tender	Economic operators submit proposals, eliminating a prior qualification phase for technical and/or financial capacity assessment of competitors.
Ao abrigo de acordo-quadro (art.º 259.º)	Under a framework agreement (Art. 259)	Inviting co-contractors to submit proposals based on specified terms and aspects.

Consulta Prévia	Prior Consultation	The contracting authority directly invites at least three chosen entities to submit proposals.
Ao abrigo de acordo-quadro (art.º 258.º)	Under a framework agreement (Art. 258)	Direct Adjustment where the awardee is chosen based on specified criteria without reopening competition, and detailed specifications are not required if they align with the framework agreement conditions.

Location and CPV are very important variables as they provide crucial information for the exploratory analysis of our network. Despite their importance, the way they are stored in contracts for our objective is not the best. Location is stored with information about where the contract was withheld, and although this seems reasonable, it is stored with information about the municipality, district, and country. For this study, this information is unnecessarily narrow, and we will cut it to a broader descriptive location. The same happens with the CPVs.

3.1.1. Location

In order to deal with the location of the procurements, it was decided that the best strategy was to divide the country according to NUTS II. NUTS is the acronym for "Nomenclature of Territorial Units for Statistics", a hierarchical system for dividing the territory into regions (PORDATA, n.d.) NUTS II is the second level of this division, and it consists in the division of Portugal by: North, Centre, Metropolitan Area of Lisbon, Alentejo, Algarve and both Autonomous Regions of Azores and Madeira (these two were combined into Autonomous Regions).

We associated each municipality with the NUTS II region according to the PORDATA information and by doing so we created six specific markets.

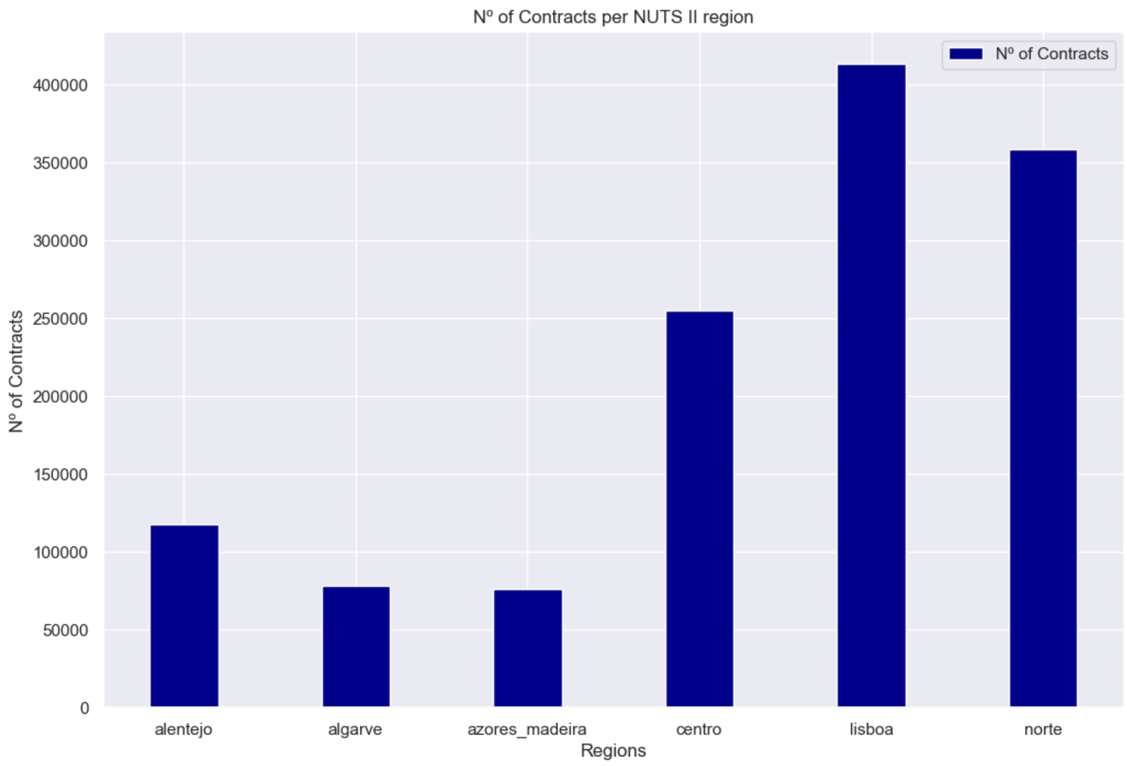


Figure 1. Distribution of contracts by NUTS II

It is clear that most of the contracts are heavily concentrated in Lisbon (which was expected, being the capital). It is important to say that although the north seems to hold up a high number of contracts, this can be slightly misleading, since just the district of Porto, holds 57% of these contracts. Independently of this we decided to keep Porto and North as the same regions. The same could be said about the district of Lisbon, which holds 87% of the contracts in the Metropolitan Area of Lisbon.

Fig.2 represents the visual distribution of the number of contracts across the country and autonomous regions.

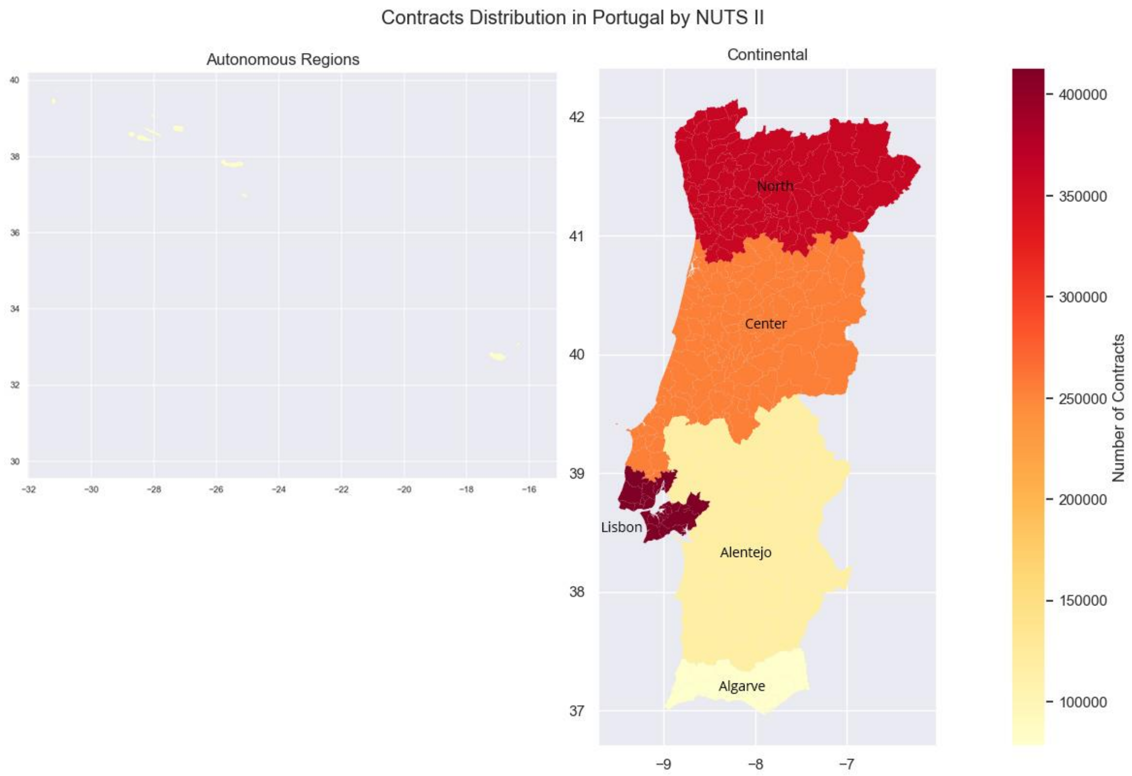


Figure 2. Heatmap of Portugal with number of contracts

3.1.2. Common Procurement Vocabulary

The CPV code is a classification system for defining the object of public procurement contracts. The main vocabulary consists in a code structure of nine digits with the purpose of specifying additional information about the contract, such as the quality or environmental requirements (https://simap.ted.europa.eu/pt/web/simap/cpv).

Table 3. How the CPV code works

CPV Code	Description
XX000000-Y	The first two digits identify the divisions
XXX00000-Y	The first three digits identify the groups
XXXX0000-Y	The first four digits identify the classes
XXXXX000-Y	The first five digits identify the categories

For our case study, we considered that a certain generalisation of the types of contracts would be better than a very specific division. Considering that the first two digits already provide 45 different categories, this was more than enough for a later analysis of the network. To translate the first two digits into a category, a dataset made available by the European Union was used, providing the division to which each contract is associated. It is important to mention that the CPV codes were updated in 2003, but this wasn't a problem considering that the data of all the contracts available in our dataset are post 2009.

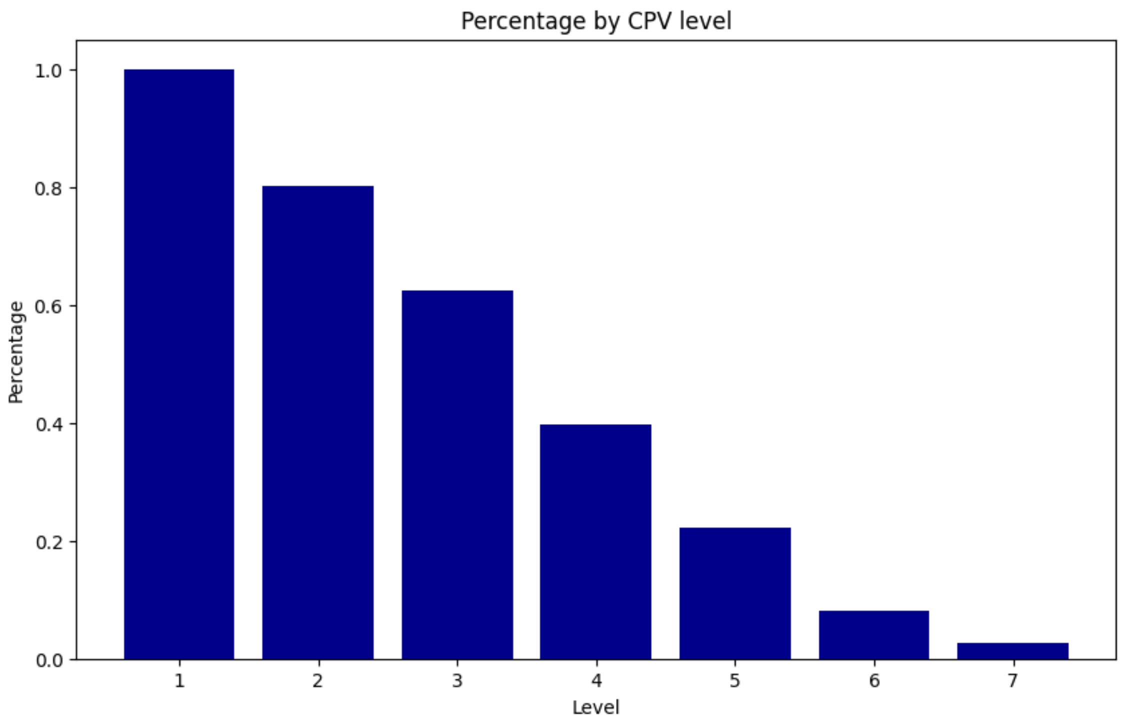


Figure 3. Percentage of CPV l Level of information in the contracts

As we can observe in Fig.3, the dataset is only consistent at the first level. As it was previously said, since the first level already as a fair share of types (45), this was the only level that was kept.

3.1.3. Initial, final, and awarded price

Succeeding the exploratory analysis of the categorical data, we decided to dive into the information related to the contract prices. In this area we have three remaining columns, the initial price, final price and the award per winner. The award per winner is the final price divided by the number of participants in the contract.

As we can see in Fig.4, in the end of every year since the e-procurement data is available, the changes in price during a contract time, as always been higher than the initial price. Usually not by much, except in 2011 where there is a huge discrepancy between the initial price of contracts and the final price.

Figure 4. Difference between initial and final prices available in the contracts

While looking for other inconsistencies in the prices data we observed that there were a lot of contracts that had an initial and final price bellow one euro. We observed that in some contracts the unit price was the initial price, and the final price was the total purchase. This shows inconsistency in data registration. In the end, we ended up clearing every contract that had a final price below one euro. (Curado et al., 2021).

3.1.4. Bidders and Sporadic activity

Succeeding the analysis of the price data, we decided to dive into the information related to the bidders. Due to the nature of the dataset provided, there is very little information about the bidders as this data is mostly about the contracts.

In the data every NIF has a name associated and although there are some irregularities in the names, we figured that since we are using the NIF as the main identifier, this wouldn't be a problem. There are also some errors in the data, for example, contracts where the winning company has two different NIFs, one NIF as the contestant, and one as the winning contestant. These were probably typing errors, which although consisting of minor mistakes, it shows a concerning lack of care handling the registration of contracts in the e-procurement network. Some of these irregularities were spotted and taken care of, but there is no guarantee that others might have gone unnoticed.

A very important aspect to further create the Firm-to-Firm bipartite network, was to make sure that we were solely focusing on firms that had a significant and consistent impact on the market. To do so, we discarded all the firms that didn't bid at least once every year. By doing so we expected to extract the core active firms of the dataset, revealing the underlying structure of the network, and identifying the most important and interconnected firms. (Lyra et al., 2021).

3.2. NETWORK DEVELOPMENT

To use the available dataset for a network analysis, it was firstly necessary to go through some steps to create a valid firm-to-firm projection. The original dataset, as it is previously stated, comes in the form of Contract ID, firm, and organiser. This was converted into a Contract-Firm Bipartite Network where each firm singularly connects to a Contract. The result is obviously immense, as the network is very large, and not only do we have a high number of firms participating but also contracts available.

The Contract-Firm network is composed by 2246 firms, 674200 contracting authorities and 1177058 edges.

Because of the large number stated in the previous paragraph, we proceeded to perform an analysis of which links are statistically significant. For this we use the standard methods of Φ correlation and t statistic (Ronen et al., 2014), which are defined as follows.

Considering matrix M_{ij}which represents the number of participations of companies i and j. The Φij correlation between company i and j is given by:

Formula 1. Φij

$$\Phi_{ij} = \frac{M_{ij}N - M_i M_j}{\sqrt{M_i M_j (N - Mi)(N - Mj)}}$$

Where M_i represents the total number of appearances of company i in the network (same for M_j) and N the total number of contracts in the dataset. The value of Φ_{ij} will be positive if the pair of companies appears together more often than expected based on their individual occurrence rates in the dataset, and negative otherwise. To assess the statistical significance of these correlations, we use the t-statistic, which is defined as follows:

Formula 2. t_{ij}

$$t_{ij} = \frac{\Phi_{ij}\sqrt{D-2}}{\sqrt{1-\Phi_{ij}^{2}}}$$

In this t-statistic $D-2$ represents the degrees of freedom. Here, to obtain a stricter criterion to find the result, we considered $D = \max(M_i, M_j)$. To conclude the t-test, we implement the Student's t cumulative distribution function. We then estimate the one-tailed p-value associated with Φij by calculating the upper tail probability of obtaining a value equal or greater than Φij from the cumulative frequency of the null-distribution. We set $\alpha = 0.05$ and discard all edges with a p-value higher than this. In order to guarantee the minimum of false positives due to small statistics, we also proceeded to eliminate all links that had the number of participations lower than the median ($M_{ij} \geq 3$).

In the end, this resulted in a final Firm-Firm co-bidding network with a total of 1107 nodes (firms) and 7353 (participations). In this final unipartite network, we treat the relationships as weighted and undirected (Kinkhabwala, 2015).

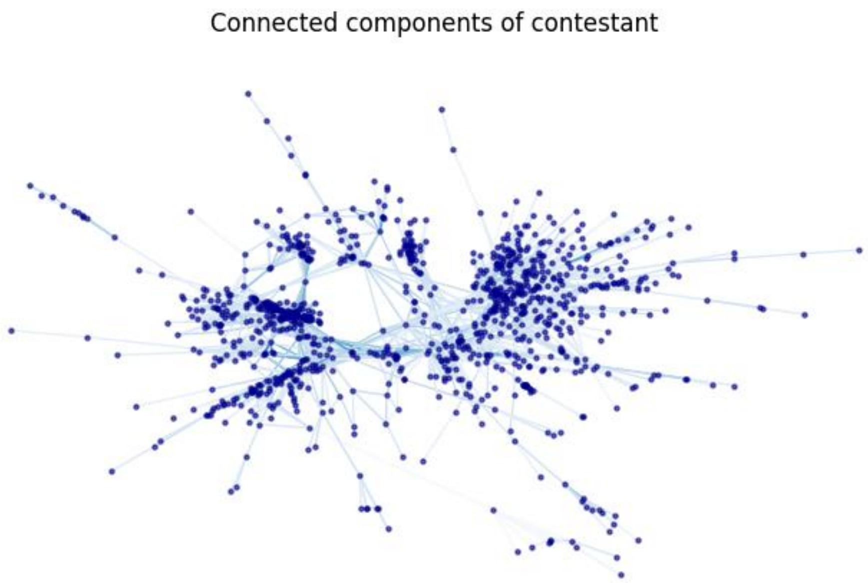


Figure 5. Visualization of the created network. The edges weights were converted to log scale for visualization purposes.

3.3. NETWORK VISUALIZATION

Before jumping right away to the next chapter, it is important to mention how we are able to observe this network and plot it in a 2D visualization.

In order to create the networks for this study, it was used the Python library NetworkX. This package allows its users to effortlessly create, manipulate and study complex networks. It comes with several tools for these purposes including different visualization techniques. Our chosen method to visualize the firm-to-firm co-bidding network was the Spring layout. This layout uses the Fruchterman-Reingold force-directed algorithm (Fruchterman & Reingold, 1991). The algorithm emulates a force-directed layout of the network by modelling edges as springs that draw nodes together, while nodes act as repulsive entities, akin to an anti-gravity force. The simulation iterates until the positions converge to a state of equilibrium.(NetworkX Documentation, n.d.)

4. RESULTS AND DISCUSSION

4.1. NETWORK ANALYSIS

Opening the results chapter, we will start by analysing the results obtained by a first glance macro analysis. As it was stated in the previous section the final bidder network is composed by 1107 nodes and 7353 edges.

As it can be seen in Table 4 our network presents high average degree and weighted degree values. The degree of a node refers to the number of edges connected to the node, while the weighted degree refers to the sum of the weights of the edges connected to it. In our network these two measures are associated with the co-bidding between two firms, where the degree refers to the extent of bidders' involvement in joint bidding, while weighted degree signifies the frequency of bidders' engagement in joint bidding activities. It is considered that the higher these two values are, the higher are the chances of a firm participating on co-bidding schemes. We can also observe in Fig.6B) how there is an exponential decay of the Degree Distribution. This indicates that the underlying mechanism of co-bidding can be modelled using a random attachment process (Albert & Barabási, 2001).

The density statistic represents the proportion of how many of the possible connections of nodes in the network are present. A network is considered to be dense when the values are closer to one and sparse when the values are closer to zero. Considering our network yields a density value of 0.012, we conclude that this is a sparse network. (https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=networks-network-density).

When it comes to Cluster Coefficient our network shows a value of 0.634. The cluster coefficient reflects the degree to which nodes in a network tend to cluster together. According to Duncan J. Watts* & Steven H. Strogatz (1998) and Kaiser (2008) a network is considered to be a small-world network when the average clustering is higher than the average clustering coefficient of a random network of the same scale, which indicates that the network is highly concentrated. This is the case of our firm-to-firm co-bidding network. The clustering coefficient serves as a metric to assess the extent to which a bidder and its neighbouring bidders engage in joint bidding activities. The average clustering coefficient across all bidders defines the bidder network's overall clustering coefficient. Consequently, a higher clustering coefficient indicates a stronger connection among bidders, signifying closer relationships and an increased likelihood of collusion.

Finally, our network yields a modularity (Q) value of 0.33. The maximum value of Q is 1 and according to Newman & Girvan (2003) a real-world network presents a significant community structure when its modularity value surpasses the 0.3 margin. Although our network is just above this threshold, we can consider that there is an implied community structure in our network.

These features combined, allow us to conclude that our network presents a structure where the bidders are prone to collude.

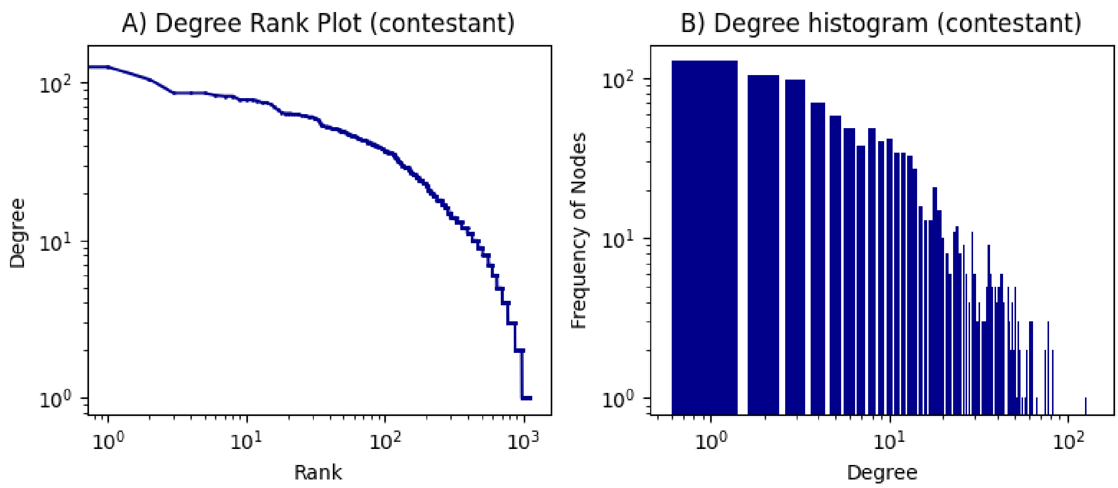


Figure 6.

Table 4. Information related to the firm-to-firm co-bidding network.

Nodes	Edges	Average Degree	Average Weighted Degree	Average clustering coefficient	Network diameter	Network density	Modularity
1107	7353	36.7	46.6	0.634	15	0.012	0.33

4.2. COMMUNITY DETECTION

After the analysis of our general network, we moved on to find communities in our network. In order to perform this, we used the Louvain Algorithm.

The Louvain Algorithm is a popular community detection algorithm for large networks. It was introduced by Blondel et al. (2008) and it is a greedy, modularity-based algorithm. This algorithm has two phases, its first phase starts by associating every node a different community. It then proceeds to evaluate the possibility of one node joining one of his neighbours' communities by addressing the modularity ΔQ of this possibility. Subsequently the node is jointed into the community where there is more modularity gain. This is repeated for every node and in the end the result is a set of first-level partitions, after this, the second phase begins. The second part consists in building a network whose nodes are now the partitions found in the first part. The weights of the edges between the new nodes are given by the sum of the weight of the edges in the corresponding two partitions. These two steps of the algorithm are then repeated, providing new hierarchical levels and bigger partition graphs. Eventually the algorithm stops once the communities become stable and there is no gain in modularity across the network.

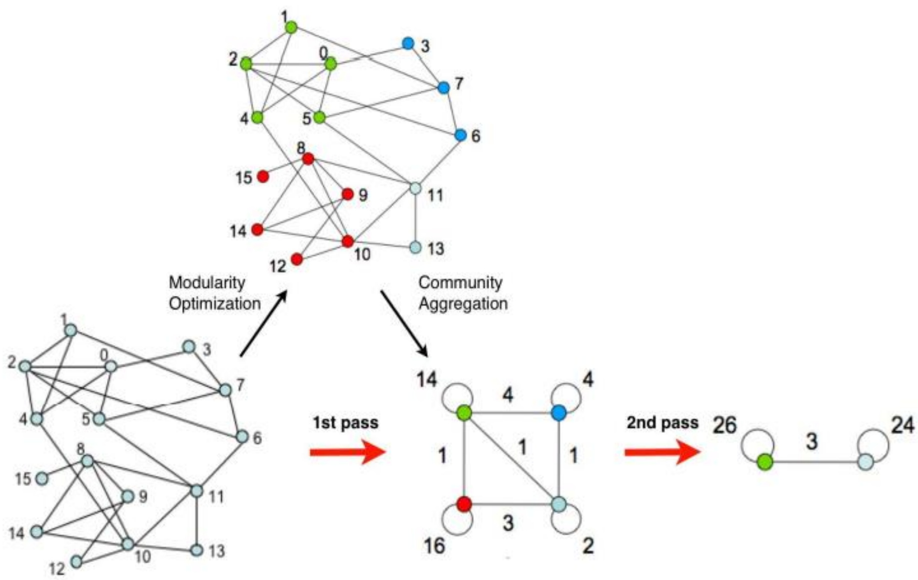


Figure 7. Visualization of how the Louvain Algorithm works.

One of the advantages of this algorithm is that it is very computationally fast. Due to its time complexity of $O(n \log n)$, it provides the chance to analyse big sized data. This algorithm tends to converge very quickly and identifies communities in very few iterations (Elhadi & Agam, 2013; Shirazi et al., 2019; Xiong et al., 2021).

After applying this algorithm, our network was divided into nineteen communities.

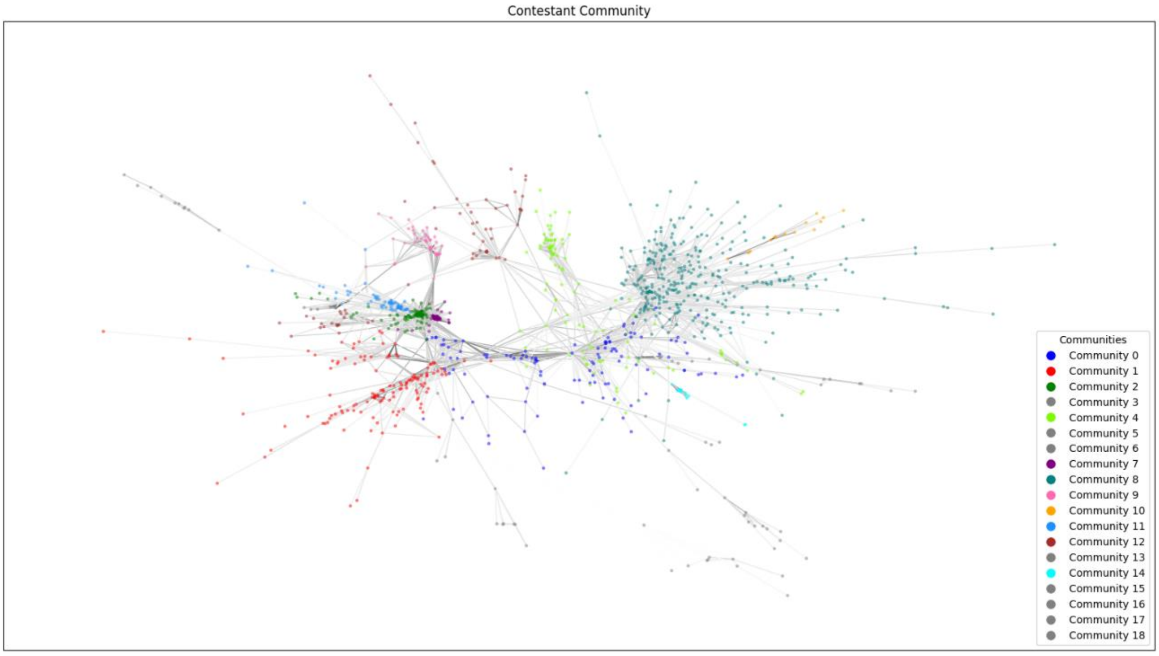


Figure 8. Firm-to-Firm co-bidding network divided into communities by the Louvain Algorithm.

Table 5. Network information

Community	Nº of Nodes	Nº of Edges	Average Degree	Average Weight	Average Shortest Path	Network Diameter	Network Density	Average Clustering	Modularity
0	112	432	15.217774	18.296727	3.271236	9	0.069498	0.624	0.408784
1	131	548	17.605108	21.699152	2.963829	8	0.064357	0.584	0.49428
2	101	1594	45.616919	51.478867	1.831881	5	0.315644	0.737	0
3	13	18	4.108333	4.862127	2.269231	5	0.230769	0.344	0.036955
4	104	421	14.528485	17.261274	3.31124	7	0.078603	0.674	0.382479
5	8	24	6.497619	6.87722	1.142857	2	0.857143	0.889	0
6	8	12	3.538889	3.911507	1.714286	3	0.428571	0.636	0
7	59	939	39.083723	41.867564	1.467563	3	0.548802	0.828	0
8	325	1569	20.099007	23.8726	3.464862	10	0.029801	0.604	0.491394
9	44	208	15.140434	16.80429	1.895349	4	0.219873	0.751	0
10	14	17	4.492222	5.622388	2.32967	5	0.186813	0.29	0.053716
11	64	473	22.125899	25.603131	2.028274	5	0.234623	0.675	0
12	64	194	9.35321	10.141925	3.340774	8	0.09623	0.572	0.609743
13	9	20	6.067262	6.873098	1.444444	2	0.555556	0.746	0
14	15	70	11.24347	12.043944	1.333333	2	0.666667	0.775	0
15	10	16	4.637897	4.547214	1.822222	3	0.355556	0.451	0
16	11	17	4.222222	4.364589	2.090909	4	0.309091	0.514	0.254791
17	10	18	4.906746	4.930276	1.822222	4	0.4	0.574	0.081535
18	5	4	2.5	2.5	1.6	2	0.4	0	0

4.3. COMMUNITIES DESCRIPTION

As we can see in Fig.8, our network has been divided into nineteen communities. In this figure only the communities with a significant value of contracts and firms are coloured. Complementing this image there is also Table 5 with information about every community.

Looking at this table we can clearly say that there major and minor communities. Some communities such as Community 2, 7, 9 and 11 show a very high level of clustering coefficient, which can be a sign that there are companies working together to control this market. Paying special attention to community 2, we can see that this community despite a relatively large number of nodes and intense activity in the market (edges) shows not only high levels of clustering coefficient, but also a very small network diameter and very high levels of degree and weighted degree. These are all indicators of a strong probability of existing bidding collusion.

4.3.1. Activity Diversity

One important factor when it comes to co-bidding is whether certain groups of firms can or cannot control specific markets. For example, if a company focuses its activities selling a service/product, or in a specific region, this has no harm because this might just mean that the company has a narrow activity scope. However, if a community of firms shows that their activity has low diversity in terms of service/product or region where they act this can be a worrying sign. This low diversity can indicate that cartels are being formed, and that companies of these communities might be coordinating their bids/wins in order to take advantage of the procurement system.

In order to understand and evaluate if this is happening in the Portuguese procurement space, we used the communities formed in our network and estimated the Simpson's Diversity Index for each one of them. (Lyra et al., 2021).

Simpson's Diversity Index (S) is a quantitative measure of the diversity of a community. It considers both the number of characteristics present (richness) and the relative abundance of each characteristic (evenness). This index measures the probability of two randomly sample individuals from a community share a given characteristic (Kim et al., 2017). In that sense a higher Simpson's Diversity Index ($S = 1$) indicates a less diverse community, while a lower index ($S = 0$) indicates a more diverse community.

In order to calculate the Simpson's Diversity Index, we used the following formula:

Formula 3. Simpsons' Diversity Index

$$S_{cat}^{C_i} = \sum\nolimits_{t \in \gamma} (p_{C_i}^{t})^2$$

In this formula C_i indicates Community i and cat a specific category (CPV, Regional or Procedure Type). t represents a type in the universal set (γ) of a category (for example, North in the NUTS II). So $p_{C_i}^{t}$ is the faction of bids done for each type in each community. The quantity of $p_{C_i}^{t}$ is normalized so that $\sum_t p_{C_i}^{t} = 1$.

Table 6. Simpsons' Diversity Index Values

Community	NUTSII Region	CPV Type	Procedure Type
Community 0	0.24785	0.462931	0.344339
Community 1	0.282497	0.532838	0.288621
Community 2	0.223705	0.885978	0.313258
Community 3	0.208755	0.5042	0.414062
Community 4	0.238689	0.500382	0.350175
Community 5	0.22628	0.650492	0.391165
Community 6	0.231872	0.967294	0.408779
Community 7	0.238379	0.986557	0.49059
Community 8	0.23586	0.834892	0.451475
Community 9	0.2586	0.956313	0.411028
Community 10	0.216396	0.688277	0.259777
Community 11	0.263863	0.883298	0.443794
Community 12	0.234841	0.56403	0.374406
Community 13	0.307594	0.961389	0.281479
Community 14	0.249027	0.909728	0.311662
Community 15	0.220996	0.987442	0.456568
Community 16	0.305301	0.495249	0.389451
Community 17	0.825027	0.935253	0.518254
Community 18	0.392665	0.544318	0.393054

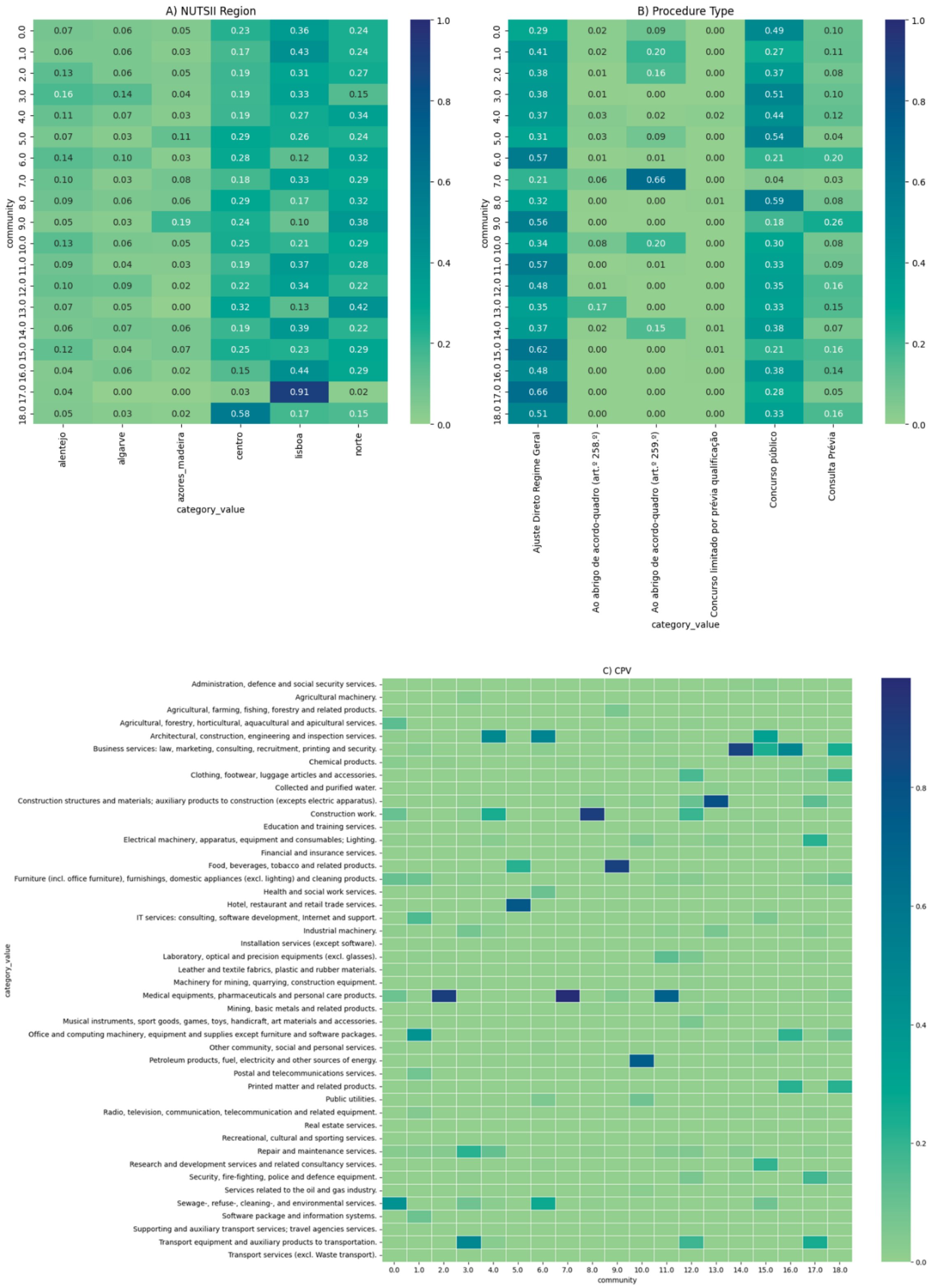


Figure 9.Simpsons' Diversity Index p values (faction of bids done for each type in each community) across categories.

As we can see from Fig.9B), Procedure Type is a characteristic that is highly distinguishable in each community. Most of the communities tend to focus their contracts on a specific type of procedure. This can be interesting as other types of procedures besides public tenders can be indicators of collusion. For example, looking at Community 7, it shows that this community tends to go for contracts that are Under a framework agreement (Art. 259). As outlined in section 3.1, within these contracts, companies extend invitations to other entities to collaborate in achieving the contract's objectives. This is obviously highly susceptible to corruption.

On the one hand, when it comes to diversity in terms of location, our network shows a good diversity pattern, with relatively low values for the Simpson's Diversity Index. As we can see from Fig.9A), the values of p are always very low and disperse across each region for each category, so it is very hard to characterize our communities in this sense. In spite of this relatively low values of p, it is important to notice that 10 out of 18 communities have Lisbon as the main location for their activities. Community 13, 16,17 and 18 are the only ones that break the 0.3 barrier in the Simpson's Diversity Index. This makes sense since these communities are among the smallest communities in the network.

On the other hand, when it comes to diversity in CPVs the story is not quite the same. Although some communities have healthy indices of diversity, half of them are highly focused on a specific type of CPV. Taking a look at the most populated networks, we can see that, for example, despite being a large community, Community 7 shows very low diversity in terms of CPV areas. This can mean that the companies in this community, tend to control this specific market.

Fig.10 complements this analysis by comparing the regional and CPV indices. Here it is important to highlight community 17 (top right point in the plot, specially coloured for this visualization). This community composed by 10 companies focus their activity on a specific type of contracts (where they are dominant) and in a specific region. These factors increase the probability of collusion and co-bidding activities.

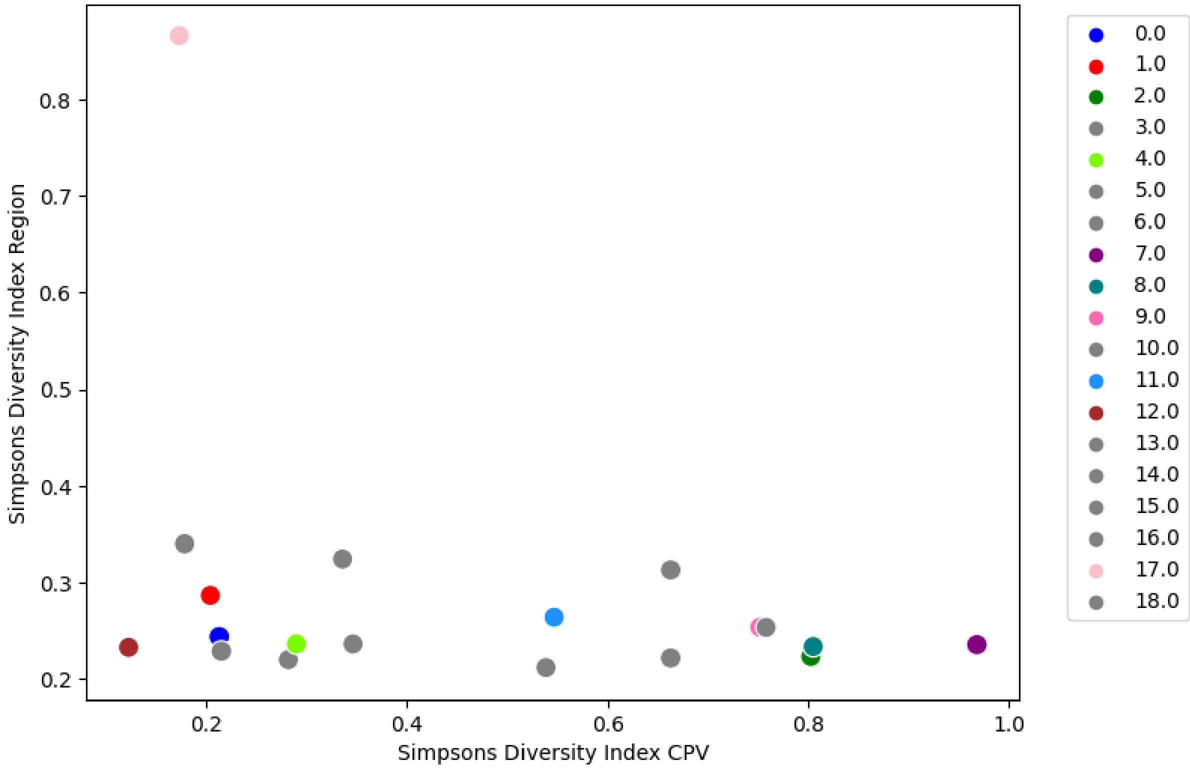


Figure 10.Compares the communities Simpsons' Diversity Index by their diversity in preferable region and area of activity.

4.3.2. Coherence and Exclusivity

Two other indicators than can highlight the susceptibility of our communities to corruption are Coherence and Exclusivity (Wachs & Kertész, 2019). Coherence is quantified as the ratio of the geometric mean to the arithmetic mean of the edge weights among group members. This measurement serves to assess the balance and overall frequency of interactions among the members within the group (community), providing insights into the distribution and intensity of connections among them. To calculate this topological measure, we used the following formula:

Formula 4. Coherence

$$C_G = \frac{\left(\prod_{l_G} w_l\right)^{\frac{1}{|l_G|}}}{\frac{\sum_{l \in G} wl}{|l_G|}}$$

Exclusivity is quantified as the ratio of strength within the group (community) over the total strength of the group, excluding on edges to non-group members, measuring the group's relative isolation in the broader market:

Formula 5. Exclusivity

$$E_G = \frac{S_{in}^G}{(S_{in}^G + S_{out}^G)}$$

Within the framework of community of firms, coherence functions as a metric capturing the consistent and intense nature of interactions among the firms within the group, while exclusivity serves to quantify the extent to which interactions within the group transpire in isolation from the broader market, delineating the degree of independence from other firms operating in the same market landscape.

In the context of this study, our proposition is that in communities that show high coherence and exclusivity rates, collusion is more susceptible to happen.

Looking at Table 7 and Fig.11, we can see that most of our communities' present significant high values of exclusivity and a decent rate of coherence. One interesting take away from these topological measures is the behaviour of the small communities. We can see that all smaller communities (except one) show a very high level of exclusivity, but when it comes to coherence, they are divided. Half presents low coherence levels and the other half very high coherence levels.

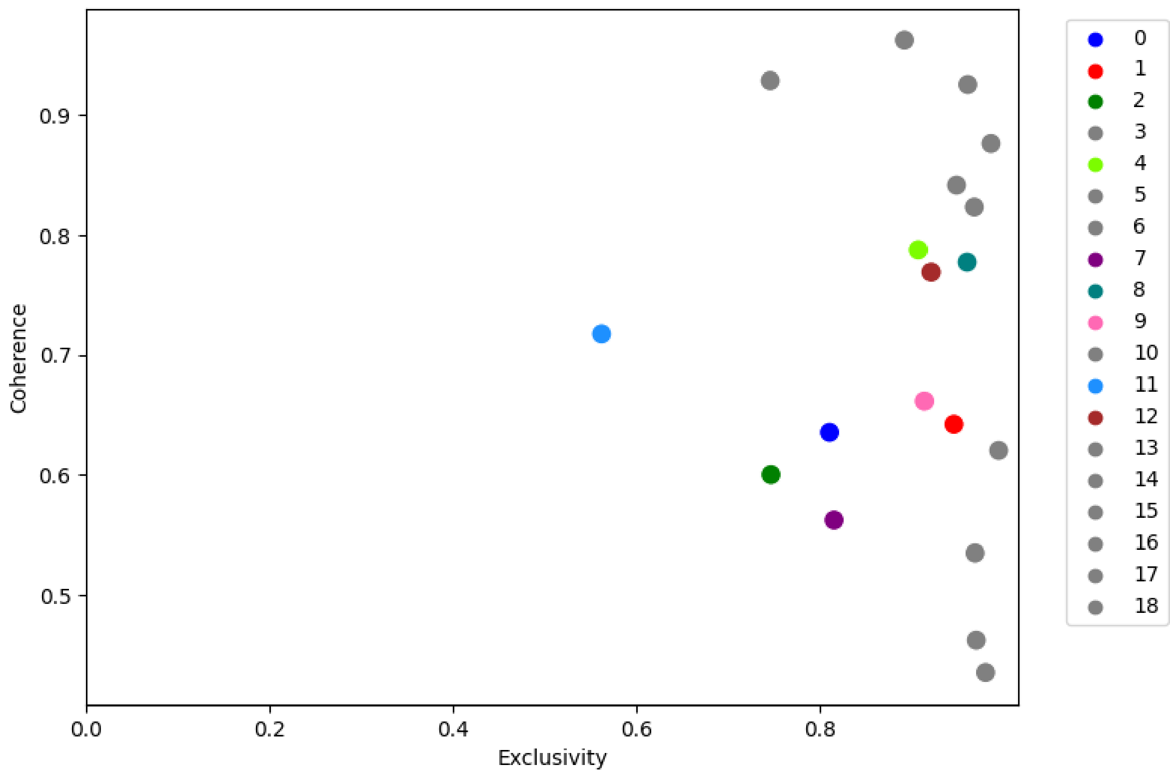


Figure 11.Coherence and Exclusivity values of each community.

~

Table 7. Exclusivity and coherence values of each community.

Community	Exclusivity	Coherence
Community 0	0.810684	0.635539
Community 1	0.946165	0.642314
Community 2	0.746608	0.600247
Community 3	0.968478	0.822906
Community 4	0.907655	0.787504
Community 5	0.970591	0.462341
Community 6	0.949025	0.841221
Community 7	0.815589	0.562281
Community 8	0.960585	0.777292
Community 9	0.914195	0.661618
Community 10	0.980792	0.435335
Community 11	0.561917	0.717344
Community 12	0.921455	0.768976
Community 13	0.969275	0.535022
Community 14	0.995102	0.620431
Community 15	0.96129	0.925234
Community 16	0.892575	0.96195
Community 17	0.986695	0.876031
Community 18	0.745763	0.928387

4.3.3. Single-Bidding Patterns and Collusion Behaviour

When it comes to bidding patterns and collusion one of the main indicators of these behaviours is single bidding. The use of single bidding has been stated to be a signal of corruption risk, or “undetected fraud” (Jungclaus Bundesverband Materialwirtschaft & Karin Attström, n.d.). However, independently of the results, one cannot straightly assume that this is a sign of corruption. Sometimes governments may need a very specific product, and this might lead to some companies having a high number of single bids as they are the only suppliers of this service/product.

To begin exploring how communities behave when it comes to single bidding in comparison to the rest of the network, we can a look at Fig 12A). In this figure, we observe how the average number of single bidders in communities is scattered (as previously, the coloured nodes refer to the bigger communities in the network). At a community level, a low level of single bidding can be an indicator that firms in this network might be coordinating their bids in this market, while a high level can mean that this market is either not very competitive or that it can exist an informal advantage in the tendering process. The single bidding rate of the Portuguese procurement network is 40%, meaning that usually companies tend to enter in contracts that are not very competitive. This information is complementary to the information obtained in the Simpson’s Diversity Index of Procurement Type, most of the communities tend to specify their market activity in Direct Adjustments which is a type of procurement where competitiveness tends to be low. In Fig.12A) we can observe that some communities show an average single bidding much higher than the network. In community 7 the average of single bidding by companies in this community is almost four times higher than the network average. Considering that this community is composed by 59 firms (which is just above the average of nodes of communities in our network), this demonstrates the possibility of a cartel operating by dominating a market where there is no competition.

Besides this, we took our analysis to a greater extent and tried to understand if companies exploit coordination in their bids in order to rank up their chances of winning a contract. One way to detect this is by studying how companies inside communities choose their contracts when it comes to its number of participants. To this end, we calculated the average number of competitors in contracts that companies enter and then normalized this value by the community size to which they belong. This normalization is necessary, as it would be expected that communities with a higher number of firms, would also have a higher number of participants in their contracts. In Fig.12B) we can see that this pattern is observed in our network.

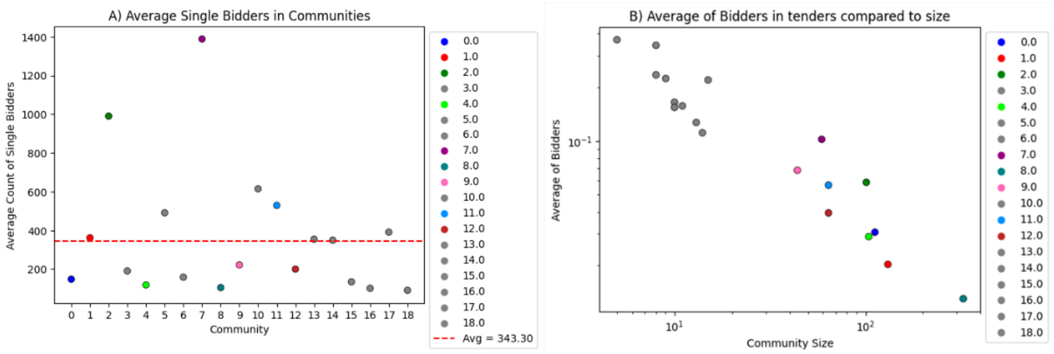


Figure 12.A) Average number of times a company belonging to a community enters a tender where they are the single bidder. B) Average number of bidders in the tenders that a company belonging to a community participates.

To further complement our analysis, we also observed how companies inside communities usually deal with competitors in contracts. Although our network shows low levels of competitiveness, in the case of competition, it is important to address if companies deal with it well or not. In order to observe this, we calculated the percentage of contracts that firms won when their competitors were belonged to their community, and when there was one or more competitors from other communities.

As we can see in Figure 13, some communities still do very well when they have competition from firms belonging to other communities, however, most of them show a negative reaction this competition. This raises the flag that cooperative behaviour inside communities can exist.

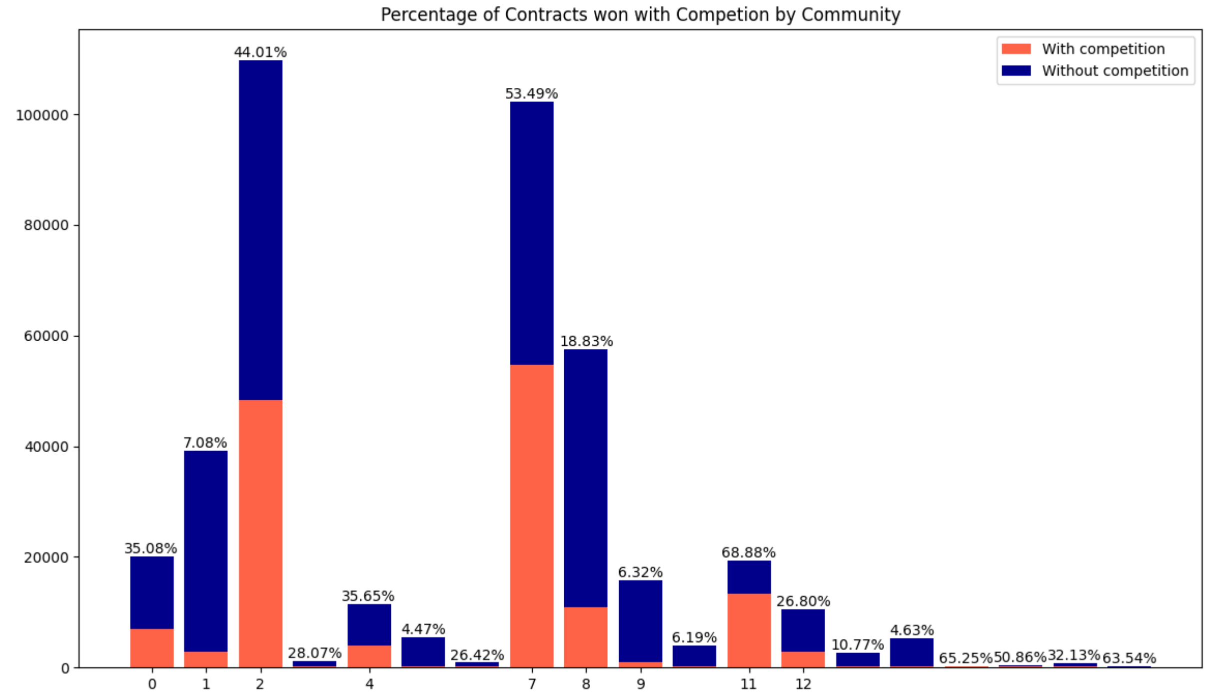


Figure 13.Communities' reaction to competition in contracts. The percentage values represent the percentage of contracts won by companies of a community when they compete against firms from other communities.

4.3.4. Performance Measurements and Central Nodes

For our final analysis we want to understand how the percentage of winning bids and the final price of contracts that companies bid on are distributed across the network. To achieve this purposed we observed two performance measures, winning bids over applied bids, and fluctuation of final prices of the contract's companies bid on, we then calculated the coefficient of variation that determines how these are distributed. Proceeding this, we also took the opportunity to check if core firms can have a more suspicious behaviour in comparison to the other firms.

The Coefficient of Variation (CV) is a fundamental descriptive metric previously employed in cartel-detection studies (Abrantes-Metz et al., 2006). The CV, expressed as the standard deviation divided by the mean, provides a relative measure of data dispersion. It measures the similarity or dissimilarity in performance measures among core firms across time in relation to the mean of the distribution. In the context of our study, when it comes to final prices, a low CV is considered bad because this can mean that the prices are set collectively, eliminating actual competition. In terms of winning bids, a low CV may suggest a tendency towards market-sharing. In a competitive market we should expect more dispersion and irregularity (high heterogeneity) in performance measures across bidders.

We say that a group of firms has a low bidding coefficient of variation if it is less than one standard deviation below the market average. (Wachs & Kertész, 2019)

To separate the core firms from the rest of the firms we used four centrality measures, averaged their values, and considered that the most important companies of the network, were the nodes with the highest values.

When it comes to the centrality's indexes, we explored four types of centralities. The degree centrality is referring to the number of edges a node has. The eigenvector centrality of a node is directly related to the cumulative centrality of its neighbouring nodes. Closeness evaluates the proximity of a node to other nodes. Higher values indicate shorter contact distances with other bidders. At last, Betweenness Centrality quantifies the number of shortest paths a node appears between others. The higher the betweenness centrality, the easier it is for the node to establish contact with other nodes. Considering these factors, it is clear that high values of centrality imply a greater importance of a firm in their community. This translates to an easier contact to other firms and closer relationships, which raise the suspicion of collusion or other advantage forms.

We can examine the CV values generated by our network in the provided Table 8. Here, we note that both values fall below the established threshold in terms of bidding performance measurement. Additionally, it is worth highlighting that, although by a slight margin, the core firms exhibit a lower CV compared to the other firms in the network. As mentioned earlier, a lower CV suggests higher levels of heterogeneity and may signal potential collusion.

In relation to the performance measure accounting the final prices of bids that companies bid on, it is hard to come to a certain conclusion. The threshold shows a negative value which indicates that the standard deviation is bigger than the mean. This corrupts these results and makes it difficult to understand the relevance of final price distributions. This can be cause because some contracts have very high values while others very low. Despite this we can still observe that core firms show, once again lower CV values then the rest of the members.

After observing these values in the general network, we took our analysis further and explored some of our communities. We limited our observations to the largest communities of the networks. The results are displayed in Table 9.

As we can see, the same trends we identified in the general network are also observable in communities. However, certain groups don't conform to this typical pattern. Regarding the performance measure for winning bids, even though all groups still exhibit low CVs, it's worth noting that in some of these groups, the main firms have less variability than the other firms. This implies that in these particular groups, there's a heightened likelihood of collaborative bidding activities when a firm holds a less central position in the group compared to a highly central one. In Annex A we can find plots of our main communities when the Louvain Algorithm is applied to them, and how the central nodes are positioned in them.

Table 8. Network CV Values. (Threshold for Avg. Winning CV is 0.78 and the threshold for Final Price is -1461272)

Core	Avg. Winning CV	Final Price CV
Is core	0.196534	2.406861
Is not core	0.215604	3.394843

Table 9. CV Values for Communities

Community	Avg. Wins Core	Avg. Wins Others	Wins Thershold	Final Price Core	Final Price Others	Final Price Threshold
Community 0	0.44	0.25	0.68	0.67	1.52	-72453.4
Community 1	0.21	0.19	0.75	0.73	1.05	-3126.85
Community 2	0.11	0.16	0.78	0.50	0.90	2482.85
Community 4	0.18	0.23	0.72	1.06	2.74	-353722
Community 7	0.14	0.10	0.84	0.45	0.77	13801.77
Community 8	0.26	0.25	0.73	1.68	2.32	-928205
Community 9	0.11	0.17	0.81	0.22	0.75	4607.20
Community 11	0.05	0.14	0.79	0.52	0.70	7801.02
Community 12	0.10	0.20	0.76	0.54	0.90	6613.25

4.4. DISCUSSION

To conclude this chapter, we provide a concise review of the previously obtained results, aiming to interconnect these findings to outline how the applied analytical methods can contribute to defining the potential risk of corruption in the Portuguese public procurement system.

Starting by the Diversity Simpson's Index, we discerned a prevailing trend where companies concentrate their bidding efforts either within the Direct Adjustment General Regime or Public Tenders. Specifically, Communities 6, 8, 11, 15, and 17 exhibit a pronounced inclination towards Direct Adjustments, while Communities 0, 3, 5, and 8 demonstrate a substantial focus on Public Tenders.

Amidst the diversity assessment values of the communities, one stands out. Community 7 exhibits a distinctive behaviour, with its most frequently employed procedure being Under a framework agreement (Art. 258) (very similar to Direct Adjustments), followed closely by Direct Adjustments. When we juxtapose this with our observation of single bidders in the network, it becomes apparent that this community is the most notable, displaying substantially higher values than the overall network average. It becomes evident that within this community, there may be a lack of competitiveness, potentially paving the way for coordinated bidding opportunities among the member companies.

Delving into the topological metrics of Coherence and Exclusivity, we find that all communities exhibit a high degree of exclusivity, a predictable outcome given the distinct focus areas within each. However, communities 4, 8, and 12 not only exhibit high values of exclusivity but also coherence. This indicates that these communities not only tend to isolate themselves from the rest of the network but also demonstrate a high level of consistency and intensity in their interactions. These are signs of a close-market dynamic. By cross-referencing these values with those of the Average Clustering Coefficient and Modularity, we can further solidify this theory. These three communities rank among the highest in both network measures, indicating a propensity to easily form groups. Once again, this serves as a potential warning flag for the risk of collusion.

Concluding this final analysis, we examine the distribution of the two performance measures. As previously mentioned, the first performance measure pertains to the average number of contracts won per bids submitted, and the second relates to the final value of contracts in which a company participates. The distribution of these measures is quantified through the coefficient of variation, which indicates their stability in the network. A lower variation suggests higher stability in the network, often reflecting bid coordination or market coordination. After observing the respective values of the communities, we can infer strong indications of corruption for the first measure, as all of them fall below the considered threshold. Notably, communities 4 and 11 stand out, displaying values significantly distant from their threshold, for both core and other firms within their community. Considering the performance comparison between core bidders and other firms, communities 2, 4, 9, 11, and 12 exhibit a clear advantage for the former. This might imply a market dominance by these companies, leveraging their significance to establish standards within their respective communities. Regarding the second measure, as previously stated, it is challenging to interpret the real value of its variation. However, it is observable once again that for all communities, the variation in the final prices of contracts where a bid is submitted by core companies tends to be lower than that of the other firms. This observation underscores the potential influence of central firms in shaping the final contract prices within their communities.

5. CONCLUSION AND FUTURE WORKS

This study aims to analyse the susceptibility of communities within a network that connects participating companies in the Portuguese public procurement system through their joint participation in contracts. We convert a set of contracts into a bipartite network and then convert this bipartite tender-firm into a firm-to-firm co-bidding network. While studies on co-bidding networks to detect cartel formation are common, we sought to combine methods that haven't previously been used together and elevate the analysis to a more specific level by observing potential collusion indicators within the communities themselves. Our final network consists of 1107 nodes and 7353 edges, forming a total of 19 communities with diverse characteristics.

In order to understand not only the overall network but also the networks formed by the communities generated by the Louvain Algorithm, we employed various methods to obtain the most concrete answer possible to the question: is the Portuguese public procurement service susceptible to collusion? The results of this analysis provide empirical evidence that the Portuguese public procurement network indeed exhibits patterns typically associated with corruption risks. Additionally, we highlight the susceptibility of certain communities, such as Community 4, which is predominantly present in the northern market and Lisbon and focus its activity on construction. This community shows pronounced indicators pointing to suspicious behaviour among its member companies. Community 7 also conveys indicators, such as their preferable procurement type, that can be an indicator of lack of competition in this community.

However, we acknowledge limitations in our thesis, particularly in the performance measurements section. This section includes errors observed in the average of final contract prices that constrain our analysis. We also considered that a better methodology to determine the qualification of core and non-core companies could have been used. For future studies in this section, we recommend conducting a Core-Periphery Analysis.

Furthermore, for future research, we propose investigating the evolution of the network over annual periods. This analysis would help assess whether core companies tend to dominate the market during certain periods and how events such as elections may impact the market, generating more or fewer indications of collusion among companies. Another potentially valuable study would be the analysis of bids within a contract and how companies might submit fake bids to coordinate their market entries, forming a clear cartel. Currently, such a study is impossible due to limited information on each company's bids in a contract. Thus, this proposal is not just a suggestion but also an appeal to the need to update and improve the public procurement network to facilitate more studies and enhance corruption detection, thereby improving the public procurement system.

To the best of our knowledge, despite studies exploring similar techniques to those in this paper, we are not aware of any similar study applied to the Portuguese public procurement network.

BIBLIOGRAPHICAL REFERENCES

Abrantes-Metz, R. M., Froeb, L. M., Geweke, J., & Taylor, C. T. (2006). A variance screen for collusion. *International Journal of Industrial Organization*, *24*(3), 467–486. https://doi.org/10.1016/j.ijindorg.2005.10.003

Albert, R., & Barabási, A.-L. (2001). *Statistical Mechanics of Complex Networks*.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, *27*(3), 264–271. https://doi.org/10.1016/j.giq.2010.03.001

Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014). *The Global Information Technology Report 2014 Rewards and Risks of Big Data*.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10). https://doi.org/10.1088/1742-5468/2008/10/P10008

Capobianco, A. (n.d.). *GUIDELINES FOR FIGHTING BID RIGGING IN PUBLIC PROCUREMENT Helping governments to obtain best value for money*. www.oecd.org/competition

Cheng, T., Liu, T., Meng, L., & Wang, C. (2017). *The Analysis of Water Project Bid Rigging Behavior Based on Complex Network*.

Cunha, S., Ferreira, I., Amaral, L. A., & Camões, P. J. (2019). Public Procurement Electronic Platforms Assessment: Legal Requirements. In *Lecture Notes in Information Systems and Organisation* (Vol. 31, pp. 81–95). Springer Heidelberg. https://doi.org/10.1007/978-3-030-14850-8_6

Curado, A., Damasio, B., Encarnacão, S., Candia, C., & Pinheiro, F. L. (2021). Scaling behavior of public procurement activity. *PLoS ONE*, *16*(12 December). https://doi.org/10.1371/journal.pone.0260806

Duncan J. Watts*, & Steven H. Strogatz. (1998). *Collective dynamics of 'small-world' networks*. https://doi.org/doi:10.1038/30918

Elhadi, H., & Agam, G. (2013). Structure and attributes community detection: Comparative analysis of composite, ensemble and selection methods. *Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNA-KDD 2013*. https://doi.org/10.1145/2501025.2501034

Esteban Ortiz-Ospina, & Max Roser. (2016). *Government Spending*. Https://Ourworldindata.Org/Government-Spending.

Fassio, A. V., Santos, L. H., Silveira, S. A., Ferreira, R. S., & De Melo-Minardi, R. C. (2020). NAPOLI: A Graph-Based Strategy to Detect and Visualize Conserved Protein-Ligand Interactions in Large-

Scale. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17*(4), 1317–1328. https://doi.org/10.1109/TCBB.2019.2892099

Ferraz, C., Finan, F., & Szerman, D. (2015). *NBER WORKING PAPER SERIES PROCURING FIRM GROWTH: THE EFFECTS OF GOVERNMENT PURCHASES ON FIRM DYNAMICS Procuring Firm Growth: The Effects of Government Purchases on Firm Dynamics*. http://www.nber.org/papers/w21219

Fruchterman, T. M. J., & Reingold, E. M. (1991). *Graph Drawing by Force-directed Placement* (Vol. 21, Issue 1).

Gallego, J., Rivero, G., & Martínez, J. (2021). Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting*, *37*(1), 360–377. https://doi.org/10.1016/j.ijforecast.2020.06.006

Garas, A., Schweitzer, F., & Havlin, S. (2012). A κ-shell decomposition method for weighted networks. *New Journal of Physics*, *14*. https://doi.org/10.1088/1367-2630/14/8/083030

Girvan, M., & Newman, M. E. J. (n.d.). *Community structure in social and biological networks*. https://www.pnas.org

Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, *57*(1), 27–57. https://doi.org/10.1023/A:1023619503005

Herland, M., Khoshgoftaar, T. M., & Wald, R. (2014). A review of data mining using big data in health informatics. In *Journal of Big Data* (Vol. 1, Issue 2). http://www.journalofbigdata.com/content/1/1/2

IMPIC. (n.d.). *BASE contractos públicos online. Retrieved from BASE contractos públicos online: http://www.base.gov.pt/Base/pt/*.

Jungclaus Bundesverband Materialwirtschaft, M., & Karin Attström Robert Kröber Jan Christoph Jähne Jost Bergschneider Jakob Thomä Tristan Swysen Felix-Sebastian Scholzen Thomas Trautenmüller, A. (n.d.). *FINAL REPORT REVIEW OF THE FUNCTIONING OF THE CPV CODES/SYSTEM*. www.ramboll-management.com

Kaiser, M. (2008). Mean clustering coefficients: The role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, *10*. https://doi.org/10.1088/1367-2630/10/8/083042

Kim, B. R., Shin, J., Guevarra, R. B., Lee, J. H., Kim, D. W., Seol, K. H., Lee, J. H., Kim, H. B., & Isaacson, R. E. (2017). Deciphering diversity indices for a better understanding of microbial communities. *Journal of Microbiology and Biotechnology*, *27*(12), 2089–2093. https://doi.org/10.4014/jmb.1709.09027

Kinkhabwala, A. (2015). Implications of network topology on stability. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0122150

Lauchs, M., Keast, R., & Yousefpour, N. (2011). Corrupt police networks: Uncovering hidden relationship patterns, functions and roles. *Policing and Society*, *21*(1), 110–127. https://doi.org/10.1080/10439463.2010.540656

Leiva, F. H., Torres, R., Nicolis, O., & Fe, R. S. (2020). Characterization of the Chilean Public Procurement Ecosystem Using Social Network Analysis. *IEEE Access*, *8*, 138846–138858. https://doi.org/10.1109/ACCESS.2020.3011947

Lewis-Faupel, S., Neggers, Y., Olken, B. A., & Pande, R. (2016). Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia. *American Economic Journal: Economic Policy*, *8*(3), 258–283. https://doi.org/10.1257/pol.20140258

Li, D., Guan, X., Tang, T., Zhao, L., Tong, W., & Wang, Z. (2023). The clean energy development path and sustainable development of the ecological environment driven by big data for mining projects. *Journal of Environmental Management*, *348*. https://doi.org/10.1016/j.jenvman.2023.119426

Llano, Q., Fernanda, M., & Aoki, N. (2022). *A framework for analyzing the design of public procurement policies A thesis submitted in partial fulfillment for the Master's Degree in Public Policy*.

Lyra, M. S., Curado, A., Damásio, B., Bação, F., & Pinheiro, F. L. (2021). Characterization of the firm–firm public procurement co-bidding network from the State of Ceará (Brazil) municipalities. *Applied Network Science*, *6*(1). https://doi.org/10.1007/s41109-021-00418-y

Mauro, P. (1995). Corruption and Growth. In *Source: The Quarterly Journal of Economics* (Vol. 110, Issue 3).

Mungiu-Pippidi, A. (2015). The quest for good governance: How societies develop control of corruption. In *The Quest for Good Governance: How Societies Develop Control of Corruption*. Cambridge University Press. https://doi.org/10.1017/CBO9781316286937

Nai, R., Sulis, E., & Meo, R. (2022). *Public Procurement Fraud Detection and Artificial Intelligence Techniques: a Literature Review*. http://ceur-ws.org

Neves, F. T., de Castro Neto, M., & Aparicio, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. *Cities*, *106*. https://doi.org/10.1016/j.cities.2020.102860

Newman, M. E. J., & Girvan, M. (2003). *Finding and evaluating community structure in networks*.

Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *64*(2), 17. https://doi.org/10.1103/PhysRevE.64.026118

OECD Foreign Bribery Report. (2014). OECD. https://doi.org/10.1787/9789264226616-en

Patrucco, A. S., Luzzini, D., & Ronchi, S. (2017). Research perspectives on public procurement: Content analysis of 14 years of publications in the journal of public procurement. *Journal of Public Procurement*, *17*(2), 229–269. https://doi.org/10.1108/jopp-17-02-2017-b003

Philip Chen, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Preuss, L. (2011). On the contribution of public procurement to entrepreneurship and small business policy. *Entrepreneurship and Regional Development*, *23*(9–10), 787–814. https://doi.org/10.1080/08985626.2010.546433

Reeves-Latour, M., & Morselli, C. (2017a). Bid-rigging networks and state-corporate crime in the construction industry. *Social Networks*, *51*, 158–170. https://doi.org/10.1016/j.socnet.2016.10.003

Reeves-Latour, M., & Morselli, C. (2017b). Bid-rigging networks and state-corporate crime in the construction industry. *Social Networks*, *51*, 158–170. https://doi.org/10.1016/j.socnet.2016.10.003

Rombach, M. P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2012). *Core-Periphery Structure in Networks*. http://arxiv.org/abs/1202.2684

Ronen, S., Gonçalves, B., Hu, K. Z., Vespignani, A., Pinker, S., & Hidalgo, C. A. (2014). Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), E5616–E5622. https://doi.org/10.1073/pnas.1410931111

Roriz, C., & Ministro, P. (n.d.). *Title: Public Procurement in Portugal 2015*. www.base.gov.pt

Rostami, A., & Mondani, H. (2015). The complexity of crime network data: A case study of its consequences for crime control and the study of networks. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0119309

Shirazi, S., Baziyad, H., Ahmadi, N., & Albadvi, A. (2019). A New Application of Louvain Algorithm for Identifying Disease Fields Using Big Data Techniques. In *Journal of Biostatistics and Epidemiology J Biostat Epidemiol* (Vol. 5, Issue 3). http://jbe.tums.ac.ir

Szell, M. (2018). Crowdsourced quantification and visualization of urban mobility space inequality. *Urban Planning*, *3*(1), 1–20. https://doi.org/10.17645/up.v3i1.1209

Tóth, B., Fazekas, M., Czibik, Á., & Tóth, I. J. (2015). *Toolkit for detecting collusive bidding in public procurement With examples from Hungary*. http://www.crcb.eu/

UNITED NATIONS CONVENTION AGAINST CORRUPTION. (n.d.). www.unodc.org

Wachs, J., Fazekas, M., & Kertész, J. (2019). *Corruption Risk in Contracting Markets: A Network Science Perspective*. http://arxiv.org/abs/1909.08664

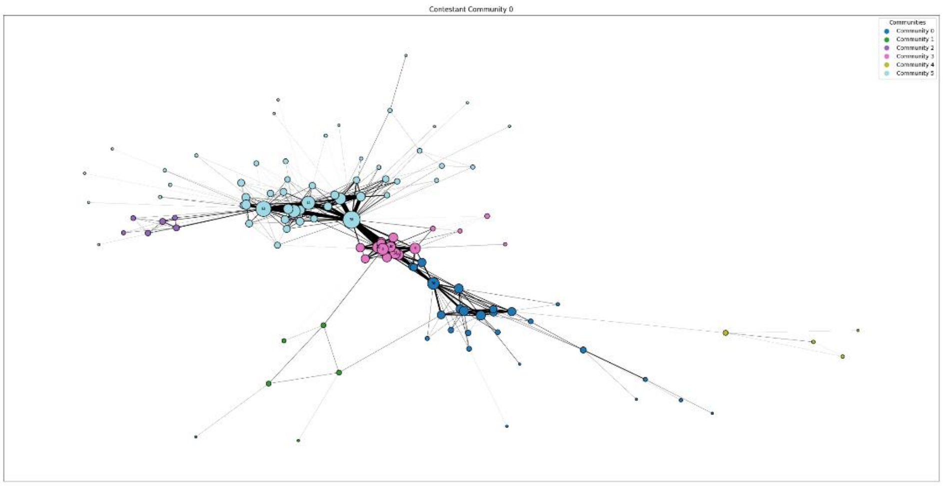
Wachs, J., & Kertész, J. (2019). A network approach to cartel detection in public auction markets. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-47198-1

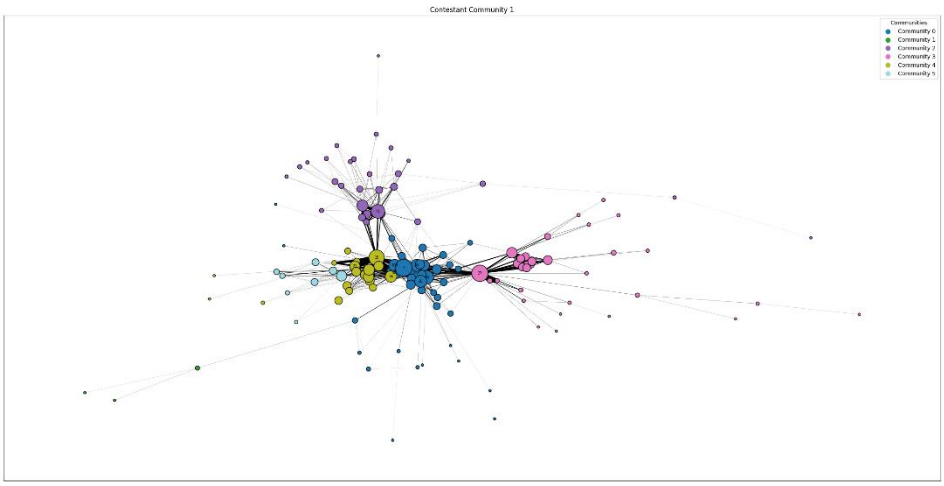
Xiong, P., Ping, W., & Chen, H. (2021). Comparison of Community Detection Algorithms on Contracts Networks. *Chinese Control Conference, CCC*, *2021-July*, 7474–7479. https://doi.org/10.23919/CCC52363.2021.9549569

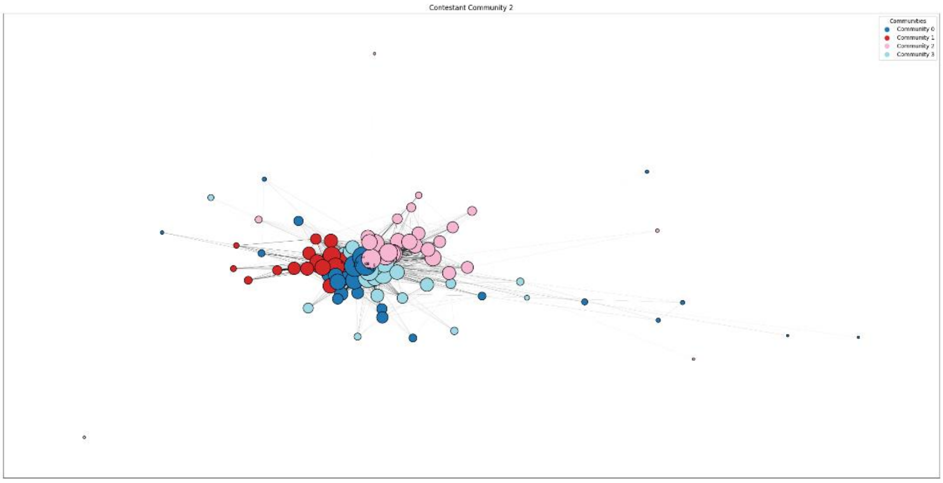
Zheng, H., Ulster University, Universidad de Granada, Universidad Carlos III de Madrid, Institute of Electrical and Electronics Engineers, IEEE Computer Society, & IEEE Computer Society. Technical Committee on Computational Life Sciences. (n.d.). *Proceedings, 2018 IEEE International Conference on Bioinformatics and Biomedicine : 3-6 Dec. 2018, Madrid, Spain*.

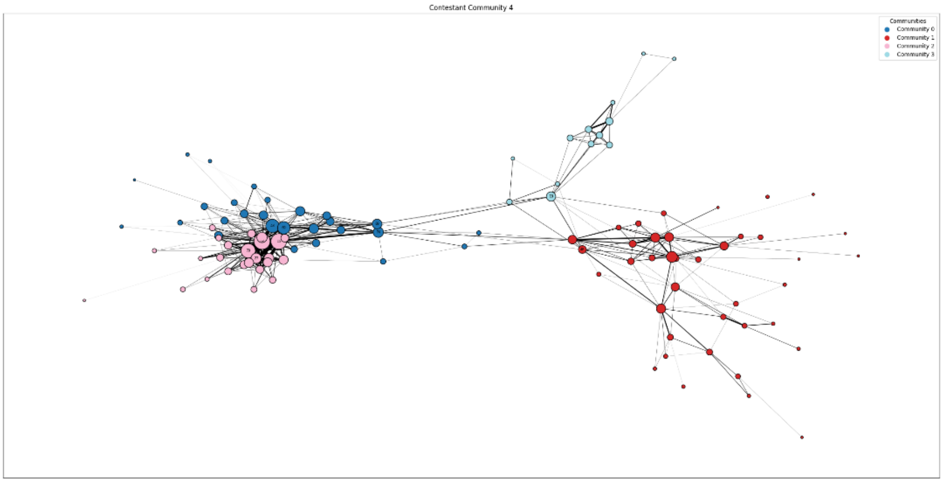
Zhu, J., Wang, B., Li, L., & Wang, J. (2020). Bidder Network Community Division and Collusion Suspicion Analysis in Chinese Construction Projects. *Advances in Civil Engineering*, *2020*. https://doi.org/10.1155/2020/6612848

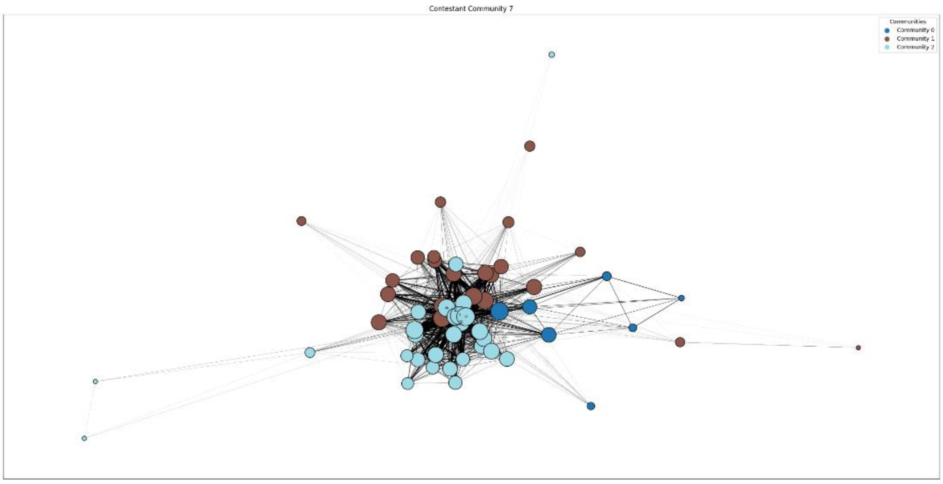
APPENDIX A

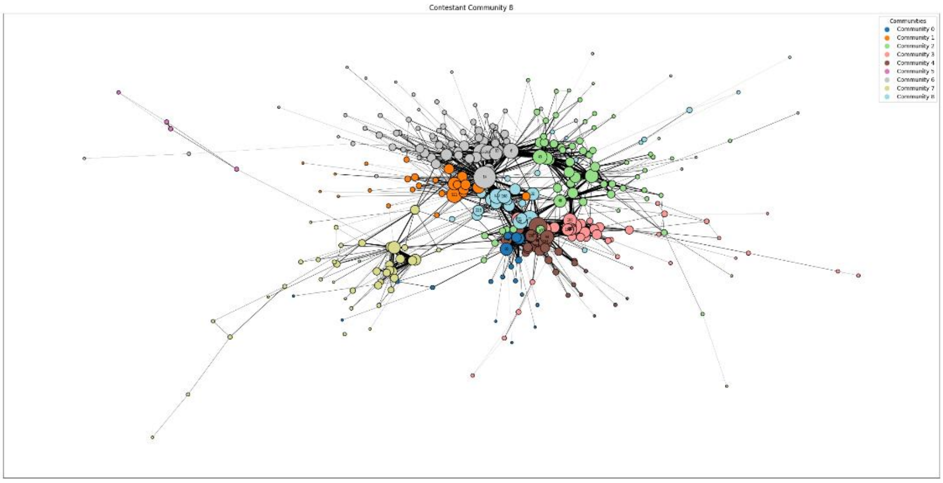

Contestant Community 0
Communities
Community 0
Community 1
Community 2
Community 3
Community 4
Community 5

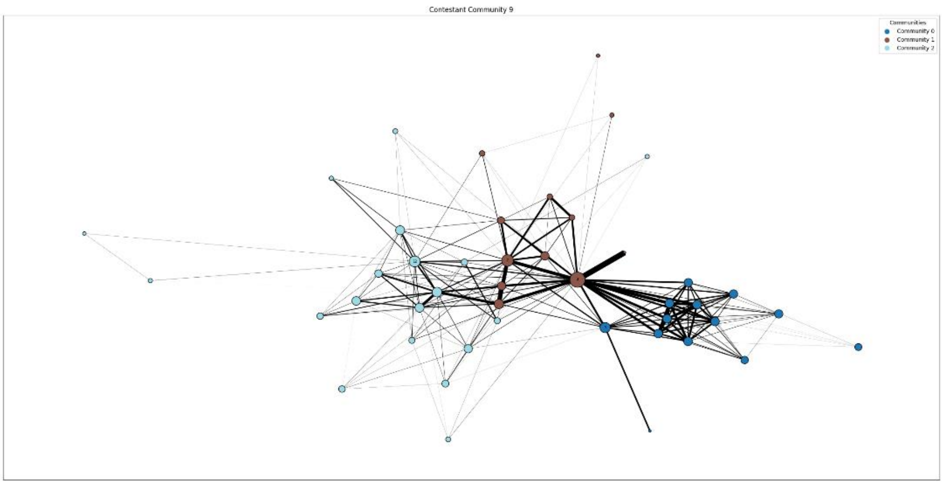

Contestant Community 1
Communities
Community 0
Community 1
Community 2
Community 3
Community 4
Community 5

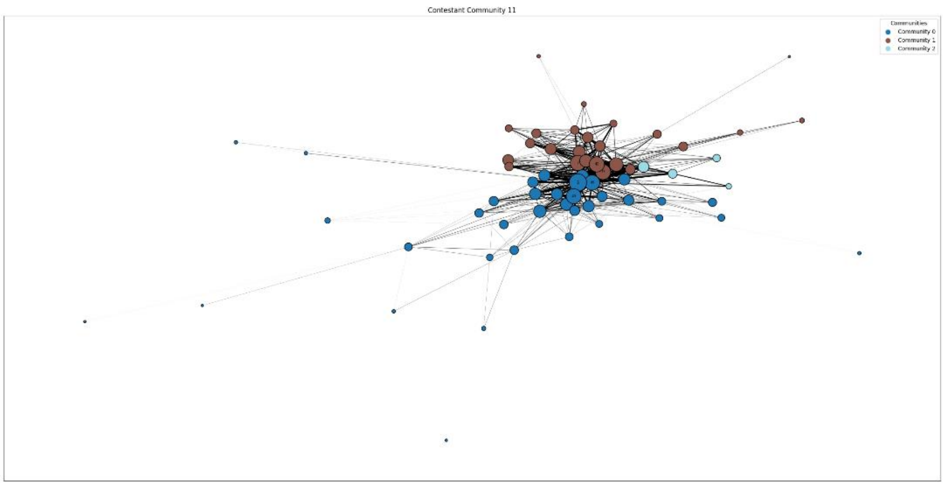

Contestant Community 2
Communities
Community 0
Community 1
Community 2
Community 3

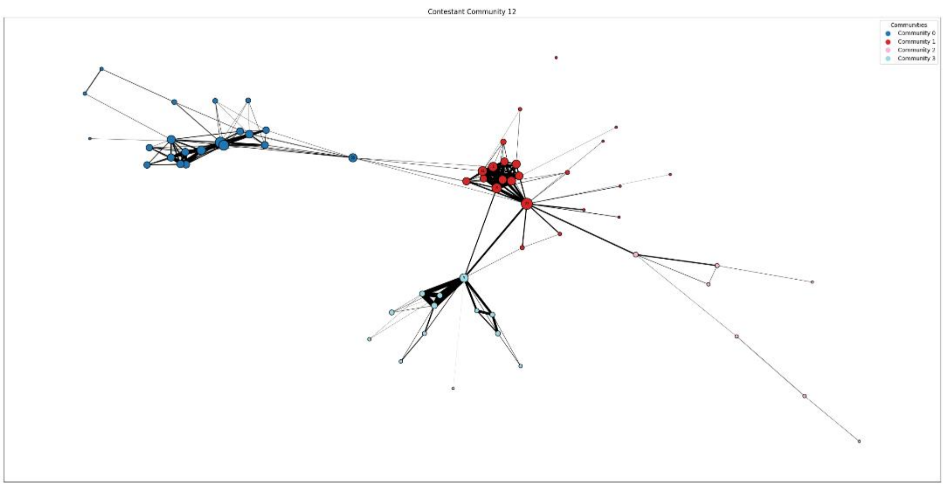

Contestant Community 4
Communities
Community 0
Community 1
Community 2
Community 3


Contestant Community 7
Communities
Community 0
Community 1
Community 2


Contestant Community 8
Communities
Community 0
Community 1
Community 2
Community 3
Community 4
Community 5
Community 6
Community 7
Community 8


Contestant Community 9
Communities
Community 0
Community 1
Community 2


Contestant Community 11
Communities
Community 0
Community 1
Community 2


Contestant Community 12
Communities
Community 0
Community 1
Community 2
Community 3

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa