Characterization of the firm–firm public procurement co-bidding network from the State of Ceará (Brazil) municipalities

Marcos S. Lyra¹², António Curado¹, Bruno Damásio¹, Fernando Bação¹ and Flávio L. Pinheiro¹*

Abstract
Fraud in public funding can have deleterious consequences for societies’ economic, social, and political well-being. Fraudulent activity associated with public procurement contracts accounts for losses of billions of euros every year. Thus, it is of utmost relevance to explore analytical frameworks that can help public authorities identify agents that are more susceptible to irregular activities. Here, we use standard network science methods to study the co-bidding relationships between firms that participate in public tenders issued by the 184 municipalities of the State of Ceará (Brazil) between 2015 and 2019. We identify 22 groups/communities of firms with similar patterns of procurement activity, defined by their geographic and activity scopes. The profiling of the communities allows us to highlight organizations that are more susceptible to market manipulation and irregular activities. Our work reinforces the potential application of network analysis in policy to unfold the complex nature of relationships between market agents in a scenario of scarce data.

Keywords: Public procurement, Network analysis, Data mining, Co-bidding network, Brazil

Introduction
Despite the weight of public procurement in governmental budgets (OECD.Stat 2017), procurement activity is still one of the most vulnerable vehicles open to corruption (Murray 2014; OECD 2015). In that context, corruption can have many forms (Angulo Garzaro 2018) and occur at any point in the procurement cycle—from the pre-tendering to the tendering and the past-award phases—making it difficult to detect and measure (Mufutau and Mojisola 2016; Rustiarini et al. 2019; Whiteman 2019). In Brazil alone, corruption in procurement contracts can represent an additional cost of 20% to 30% over the expected price, which represents losses of around 200 billion Reais annually (Zeferino 2020). Likewise, in Europe, it is estimated that losses figure around 5 billion Euros annually (Hafner et al. 2016). Naturally, these losses undermine the ability of governments and public authorities to forge ahead with essential investments in health, education, infrastructure, security, housing, and social services (Søreide 2002; Beittel et al. 2019). Unsurprisingly, there is a considerable effort to develop analytical solutions to understand and mitigate the effects of corruption...
in the public procurement process (Fazekas et al. 2018). The increasing availability of open data concerning public administration activities (Curado et al. 2020) has recently renewed the scientific community’s efforts to uncover hidden connections between participating agents and how their relationships can link to fraudulent activities (Herrera 2019; Kertész and Wachs 2021).

One of the most challenging aspects of identifying corruption in public procurement contracts is the lack of labeled data. Indeed, it is largely impossible to know which instances stem from corruption (Wachs and Kertész 2019). However, a fundamental principle in public procurement is that of transparency in bidding (Adjei-Bamfo et al. 2019; Angulo Garzaro 2018; Nowrousian 2019; Spagnolo 2012), and efficiency can be obtained through independent and open competition between firms. In that sense, past works have approached this problem from an unsupervised learning perspective, meaning that they look to extract information about the relationships between the involved parties and, thus, flag groups of agents with patterns that might be linked to a high risk of corrupting activities. Indeed, firms can achieve leverage to manipulate the tender process by establishing the right relationships among themselves and coordinate their activity (Hanák and Serrat 2018).

An open issue remains, can communities of firms obtained from co-bidding patterns allow us to highlight groups that are more susceptible to collusion and market manipulation? The use of network analysis for the study of corruption is not new (Lauchs et al. 2011; Chang 2018; Grassi et al. 2019). In the context of public procurement, past studies can be divided into two main groups: (1) works that explore bipartite relationships between public bodies and firms (Fazekas and Tóth 2016; Wachs et al. 2019); and (2) studies that explore firm–firm co-bidding relationships in public tenders (Toth et al. 2014; Reeves-Latour and Morselli 2017; Morselli and Ouellet 2018; Wachs et al. 2020). Both approaches have their merits, and each is suitable to identify different mechanics underlying the manipulation of the procurement process. For instance, bipartite relationships are suitable for identifying fraud stemming from bribes and influence ties, while firm–firm relationships are more suited to identifying cartels and collusion. Despite these, the use of network analysis to study the relationship between firms in procurement bids is a relatively new venture (Reeves-Latour and Morselli 2017). More evidence is required to understand the universality of the existing patterns and mechanics across cultural and socio-economic contexts.

Here, we use network science and complexity sciences methods to map and characterize the co-bidding network (Wachs and Kertész 2019; Piccolo et al. 2018; Ramalho et al. 2020; Reeves-Latour and Morselli 2017) between firms that participated in public tenders issued by the 184 municipalities of the state of Ceará (Brazil). In that sense, we characterize of the relationships between competing firms and identify the major communities of firms that often compete for tenders with a similar scope. Moreover, we argue that some such communities have characteristics that place them at a higher risk of market manipulation and irregular activities often associated with corruption.

Data

We used data from the State of Ceará Audit court authority—Tribunal de Contas do Estado Ceará (Brazil) covering public tenders issued by the 184 municipalities of the State of Ceará between 2015 to 2019. Each observation informs about a firm’s bid to a tender and whether the bid was one of the winning bids. It also includes information
about the municipality that issued the tender, and whether a firm won a contract. Hence, the data is naturally represented through a bipartite nature (Fierăscu 2017), which connects firms to tenders (see Fig. 1a). The data set contains 196,608 observations that account for the bids of 45,502 firms to 84,835 tenders.

Information about the firms and tenders is anonymized, and bidding values are not available. Moreover, the data set does not contain information about which contracts/firms have been investigated for irregularities in the past.

Network inference
Since we are interested in studying the relationships between firms we focus on the Firm–Firm projection. We estimated the projection from the co-bidding patterns of firms (Piccolo et al. 2018) using the Jaccard similarity coefficient (Veech 2013; Mainali et al. 2017; Chung et al. 2019; Wachs and Kertész 2019). Figure 1a shows a graphical illustration of the data structure and depicts the steps conducted to infer the Firm–Firm network from the original Tender-Firm bipartite structure.

In order to infer the Firm–Firm co-bidding network, we started by discarding all firms that did not bid at least once during each year under analysis. By doing so, we were able to extract the core of active firms, while removing firms with sporadic activity. Figure 1b,
c compare the original (\(P_{ALL}\)) with the filtered data set (\(P_{Sample}\)). In particular, it shows that filtering removes excess participants from tenders while not affecting the distribution of the number of bids done by each firm. Likewise, we refer to firms present in the firm–firm co-bidding network as Established firms. The final working data set includes 1906 firms, which account for 72,078 bids to 39,523 tenders.

Hence, next, we compute the centered Jaccard/Tanimoto coefficient (Chung et al. 2019) between each pair of firms. The centered Jaccard coefficient measures the similarity between the bidding of two firms, accounting for the occurrence probabilities of each firm. Formally, it can be computed as:

\[
J_{ij}^c = \frac{\sum_k b_{it} b_{jt}}{\sum_k (b_{it} + b_{jt} - b_{it} b_{jt})} - \frac{p_i p_j}{p_i + p_j - p_i p_j},
\]

where \(b_{it}\) is one if firm \(i\) made a bid to tender \(t\), being zero otherwise; and \(p_i\) is the fraction of tenders in which firm \(i\) participated (\(p_i = \sum_k b_{it}\)). The second term in Eq. (1) provides the expected number of observations when the bids from both firms are independent and identically distributed through a Bernoulli process (Chung et al. 2019). Hence, the centered Jaccard coefficient allowed us to distinguish between positive and negative associations between firms, accounting for their individual level of activity.

Finally, we estimate the significance of the observed \(J_{ij}^c\) to test the hypothesis that \(J_{ij}^c > 0\). To that end, we bootstrapped a null distribution (\(\hat{J}_{ij}\)) of centered Jaccard coefficient for each pair of firms by generating an ensemble of 1000 randomizations of the initial bipartite network. Data was randomized in order to ensure that the number of bids observed per firm and per year remained constant while preserving the number of firms bidding to each tender. Then, we estimate the one-tailed \(p\)-value associated with \(J_{ij}^c\) by calculating the upper tail probability of obtaining a value equal or greater than \(J_{ij}^c\) from the cumulative frequency of the null-distribution \(\hat{J}_{ij}\) (Gotelli 2000). Links with a \(p\)-value greater than 0.05 were discarded.

The resulting firm–firm co-bidding network contains 1529 nodes and 12,892 edges. Relationships are treated as undirected and unweighted. The network exhibits an average degree of 16.86, with a cluster coefficient of 0.52 (Newman et al. 2020), and 56 connected components. Figure 2 shows the Degree Distribution (Fig. 2a) decays exponentially with the degree, which suggests that the underlying mechanics of co-bidding can be approximated by a random attachment process (Albert and Barabási 2002). Alternative distribution (e.g., power-law) has been tested but shown a worst likelihood given data. However, the average clustering coefficient shows a power-law inverse relationship with the degree (Fig. 2b), suggesting the existence of some level of hierarchy in the structure of the network. It is noteworthy to mention that the largest connected component contains 1141 nodes, 10,630 edges, and a clustering coefficient of 0.43.

**Results and discussion**

Figure 3 presents the giant component of the firm–firm co-bidding network. Using the Louvain algorithm (Blondel et al. 2008) we identified 22 communities with a modularity of 0.66. We refer to these communities as \(C_1, C_2, \ldots, C_{22}\), and they are indexed in descending order in respect to their size. For readability, we have colored the eight largest communities in Fig. 3. However, the high modularity of the network is unsurprisingly
and can be explained by the fact that the network primarily represents competing firms that are specialized in supplying different services—works, goods, services, etc—in different regions. Hence, the most prominent communities divide the network into two major groups of firms that operate mainly in the northern (Red, Blue, and Green) and southern (Purple and Yellow) regions of the state of Ceará, but also on contracts that supply Food services (Blue and Purple) or construction works (Red and Yellow). Interestingly,
the remaining communities operate at a state-wide level (Pink and light Blue), and there is one particular community (Violet) that operates exclusively in the mesoregion of Jaguaribe only supplies Food services.

In some cases, firms form densely connected sub-graphs (e.g., $C_{14}$). These structures can be a first indicator to flag groups of firms that present a high risk of collusion and procurement manipulation. As such, next, we explore possible additional metrics to classify each community of firms through their activity in order to further narrow which groups of firms might deserve a more profound investigation by audit officials.

**Activities diversity**

We started by looking at the regional diversity on which firms performed their activities (e.g., bid on tenders in order to supply services) and the diversity of the type of contracts for which they bid. While a firm with low diversity in both regional reach and contract-type can simply indicate a firm that is narrow in both scope and domain; the existence of groups of connected firms (i.e., a community of firms) that share a low diversity in both dimensions can highlight a more troublesome scenario. In particular, it can indicate the conditions for firms to coordinate and cooperate to control a specific market and regional context, and should be investigated with further discernment.

To that end, we estimate the Simpson’s diversity index$^3$ for each community. The Simpson’s diversity index ($\lambda$) measures the probability that two randomly sampled elements from a set share a given characteristic in common. In that sense, $\lambda = 0$ is associated with the highest diversity possible, and $\lambda = 1$ with the lowest diversity. Formally we estimate the Simpson’s index for each community as

$$\lambda^{C_i}_{\text{cat}} = \sum_{t \in \gamma} (p_{tC_i}^{\gamma})^2$$  \hspace{1cm} (2)

where $p_{tC_i}^{\gamma}$ corresponds to the fraction of bids done in a procurement contract type $\gamma$ :{Consumables Health, Services, Construction, Events, Food, Fuel,...} or mesoregion $\gamma$ :{Metropolina, Norte, Sul, Noroeste,...} by the firms in community $C_i$, $\forall i \in \{1, 2, \ldots, 21, 22\}$. The quantity $p_{tC_i}^{\gamma}$ is normalized per community, so that $\sum_t p_{tC_i}^{\gamma} = 1.0$. We estimated $\lambda^{C_i}_{\text{cat}}$ independently for each community ($C_i$), and for contracts according to the region that issue the tender and the tender contract type (e.g., services, food, tenancy, construction, etc). Our choice of the Simpson’s index over other alternatives (e.g., entropy) is due to its straightforward interpretation in our context: the probability that two bids made by firms within the same community share the same characteristic (e.g., region or contract type).

Figure 4a illustrates the empirical distributions ($p_{tC_i}^{\gamma}$) of procurement activity for the ten most prominent communities. We show the results for both the Regional distribution of activities and by Contract Type. Blue colors denote a low relative frequency of bids, while red identifies a high frequency. These indicators allowed us to infer the degree of specialization and agglomeration of a community. In particular, we found that Community 8 ($C_8$) activities are agglomerated in a single region (Jaguaribe) and firms

$^3$ In some fields, the Simpson’s index is also known as the Herfindahl index.
specialize in one type of contract (Food). The same conclusion can also be inferred from the high levels of $\lambda_{\text{cat}}$, which means that Community 8 has low diversity of activity distribution. Figure 4b compares all the 22 communities in terms of the two diversity indicators defined above. We find a clustering of communities in the bottom left quadrant—a low level of agglomeration and specialization—that we associate with healthy markets composed of firms that, on average, have a diversified portfolio of activities and regional distribution. In contrast, in the top right quadrant, we found communities that relied on procurement contracts of a single type and agglomerated in a small number of regions.

The combination of these two diversity indicators, at the community level, provides a powerful feature to identify groups of firms that can dominate over a niche market or, in the worst case, develop undesirable leverage, as a group, in negotiating procurement contracts. Hence, lowering the desirable efficiency that public procurement aims at achieving in the tendering process. However, it is important to stress that these metrics are just indicative of potential problems, and thus the true nature of the activities of the firms in each community should be carefully investigated by the corresponding local authorities.

**Bidding coordination**

To further investigate the risk/susceptibility of market manipulation by firms, we next looked at the propensity that each community has in participating in “single bidder” contracts. Another pattern often associated with corruption and loss of efficiency. Hence, what is the susceptibility of each community to such practice? To answer this question, we started by investigating the average number of times, per community, that a firm is the single bidder of a tender. Figure 5a shows the results for all 22 communities in the most significant component of the Firm–Firm network. Traditionally, a high
level of single bids can be an indicator of firms acting with some level of informal advantage in the tendering process or due to lack of competition in a specific market. At the community level, such an indicator can be indicative of unusual activity from a group of firms. Hence, low levels of single bidding indicate the risk of coordination (e.g., firms participating coherently in the same contracts) while high levels can sign the prevalence of less competitive markets or informal advantage in the tendering process. Overall, of the largest ten communities, only Community 8 exhibits low levels of single bidders, a pattern that extends to Communities 14 and 21 as well. In contrast, we saw that community 12 strongly deviates from the baseline with an average value of single bidding that is roughly four times that of a typical firm.

In addition, we looked at the average number of bidders per tender in order to assess the potential existence of coherent behavior, that is, coordination between the firms in a community. To that end, we estimated the average number of bidders per tender for each community, which we normalized by the size of the community (i.e., the number of firms in a community). Interestingly, Fig. 5b shows that in Community 8, firms tend to participate in tenders with several firms that match almost exactly the community’s size. While, in some cases—Communities 14 and 19—firms tend to bid to tenders that are several times larger than their communities. Noteworthy to mention that this analysis is biased by the size of the communities, so the expectation would be to see a smoothly increasing relationship, with the largest community achieving the smallest value, and in the limiting case of a community with a single firm we would obtain the maximum. However, it is clear that in some cases—Communities 8, 14, and 19—there are apparent deviations.

Conclusions
In this manuscript, we explored the potential of mining a large data set of public tenders collected from firms’ activity to compete for procurement contracts issued by the municipalities of the State of Ceará (Brazil). By matching firms with similar bidding patterns, we have inferred a firm–firm network comprising a total of 1141 nodes and 10,630 edges.

We showed that we were able to identify communities of firms with similar bidding patterns. The network exhibits a high modular structure partitioned in 22 communities.
These communities cluster firms that have a similar scope in procurement activities both in the nature of the contracts they celebrated and in the regional reach of their activities. Moreover, we looked at two diversity indicators—regional diversity and procurement contract nature diversity—as a sign of the potential of certain communities to develop leverage over the procurement process. In other words, in affecting the expected efficiency of the market. Finally, we looked at the sizes of the tenders, first by looking at the abundance of single bidders in communities, and secondly by looking at the average number of bidders in each tender. Overall we identified a particular community (Community 8) that combines several undesirable properties. Community 8 involves a group of firms that offers Food services in the region of Jaguaribe. They have an unusually low number of single bids; the average number of participating firms per tender matches the number of firms in the community, and they exhibit a high specialization and agglomeration in their activities. Nevertheless, having said that, such an odd combination of characteristics can be useful to narrow down the activities of audit officials, but it does not allow us to conclude much about the true nature of the activities of firms involved in Community 8.

Finally, it is essential to highlight some shortcomings in our analysis and future working directions. The lack of pre-labeled data on past corruption cases significantly limits our ability to make any causal link between the network structure, its motifs, and the location of firms in the network with irregular procurement behavior. In that sense, our results are merely exploratory and show the potential of combining network science methods with descriptive statistics to highlight relevant groups of firms according to their activity pattern in a data-scarce environment. Future works should look at the evolution of the network, that is, if a larger temporal window is available, to capture the evolution and segregation of communities of interest but also of their parametric path in terms of the diversity of their activities.

Acknowledgements
The authors are thankful to Tribunal de Contas do Estado do Ceará for sharing the data for this study. The authors are thankful to Cristian Candia for the useful discussions and insights. The findings, interpretations, and conclusions expressed by the authors in this work do not necessarily reflect the views of the Tribunal de Contas of Ceará.

Authors’ contributions
MSL and AC performed the data extraction, cleaning, and analysis. FLP, FB, and BD supervised the study and conceived the idea. All authors participated in the writing of the manuscript. All authors read and approved the final manuscript.

Funding
MSL was financially supported by Tribunal de Contas do Estado do Ceará through a PhD scholarship. BD, FLP, and FB acknowledge the financial support provided by FCT Portugal under the project UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC). FB and FLP acknowledge the financial support provided by FCT Portugal under the project DSAIPA/DS/0116/2019.

Availability of data and materials
The dataset supporting the conclusions of this article is available in the figshare repository, https://figshare.com/s/0366318e1cBB81f5a2ce0.

Declarations
Competing interests
The authors declare that they have no competing interests.

Author details
1Information Management School (IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal. 2Tribunal de Contas do Estado do Ceará, Rua Sena Madureira, 1047, Fortaleza 60055-080, Brazil.

Received: 20 April 2021 Accepted: 30 September 2021
Published online: 15 October 2021
References

Adjei-Bamfo P, Maloeh-Nyamekye T, Ahenkan A (2019) The role of e-government in sustainable public procurement in developing countries: a systematic literature review. Resour Conserv Recycl 142:189–203

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47

Angulo Garzao N (2018) Eu competition law and public procurement: competition-driven limits imposed to public bodies when they source goods, works and services. Works and Services (April 9, 2018)

Beittel JS, Meyer P, Seelk C, Taft-Morales M, Gracia E (2019) Combating corruption in Latin America: congressional considerations. Report Number R45733

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008(10):10008

Chang Z (2018) Understanding the corruption networks revealed in the current Chinese anti-corruption campaign: a social network approach. J Contemp China 27(113):735–747

Chung NC, Missalovedov B, Startek M, Gamban A (2019). Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. BMC Bioinform 20(15):1–11

Curado A, Damásio B, Encarnação S, Cândia C, Pinheiro FL (2020) Scaling behavior of public procurement activity. arXiv preprint arXiv:2007.15276

Fazezakas M, Toth U (2016) From corruption to state capture: a new analytical framework with empirical applications from Hungary. Poli Res Q 69(2):320–334

Fazezakas M, Cingolani L, Tóth B (2018) Innovations in objectively measuring corruption in public procurement. In: Anheier HK, Haber M, Kayser MA (eds) Governance indicators: approaches, progress, promise. Oxford University Press, Oxford, pp 154–185

Fierúsco S (2017) The networked phenomenon of state capture: network dynamics, unintended consequences, and business-political relations in Hungary. Central European University, Budapest

Gotelli NJ (2000) Null model analysis of species co-occurrence patterns. Ecology 81(9):2606–2621

Grass R, Calderoni F, Bianchi M, Terrizzio A (2019) Betweenness to assess leaders in criminal networks: new evidence using the dual projection approach. Soc Netw 56:23–32

Hafner M, Taylor J, Disk E, Theibes S, Barberi M, Stepanek M, Levi M (2016) The cost of non-Europe in the area of organised crime and corruption: annex II-corruption. RAND Corporation, Santa Monica, CA

Hanák T, Serrat C (2018) Analysis of construction auctions data in Slovak public procurement. Adv Civ Eng 2018:1–13

Herrera-MD14 (2019) Using social network analysis in open contracting data to detect corruption and collusion risks. Eindhoven University of Technology - Department of Mathematics and Computer Science, Eindhoven

Kertész J, Wachs J (2021) Complexity science approach to economic crime. Nat Rev Phys 3(2):70–71

Lauchs M, Keast R, Youssefpoour N (2011) Corrupt public networks: uncovering hidden relationship patterns, functions and roles. Poli Soc 21(1):10–127

Mainali KP, Bewick S, Thielen P, Mehoke T, Breitwieser FP, Paudel S, Adhikari A, Wolfe J, Slud EV, Karig D et al (2017) Statistical community detection. Phys Rev E 101(4):042304

Morselli C, Ouellet M (2018) Network similarity and collusion. Soc Netw 55:21–30

Murray JG (2014) Procurement fraud vulnerability: a case study. EDPACS EDP Audit Control Secur Newsl 49(5):7–17

Nowrousian B (2019) Combatting public procurement criminality or simple rules for complex cases. J Financ Crime 26:203–210

OECD (2015) Preventing corruption in public procurement. https://www.oecd.org/gov/public-procurement/integrity/

OECD (2017) Government at a glance—2017 edition: public procurement. https://stats.oecd.org/Index.aspx?QueryId=78413. Accessed: 14 March 2021

Piccio S, Lehmann S, Maier A (2018) Design process robustness: a bipartite network analysis reveals the central importance of people. Des Sci 4:e1

Ramalho HMB, Almeida ATC, Fraga AA (2020) Detection of suspected collusion cases in public biddings: an application of the a priori machine learning algorithm for the State of Paraiba, 2nd edn, vol 10, pp 5–22

Reeves-Latour M, Morselli C (2017) Bid-rigging networks and state-corporate crime in the construction industry. Soc Netw 56:154–185

Rustianni NW, Sutrisno T, Nurkholis N, Andayani W (2019) Why people commit public procurement fraud? The fraud diamond view. J Public Procure 19(4):345–362

Sareide T (2002) Corruption in public procurement: Causes, consequences and cures. Chr. Michelsen Institute, Bergen

Spagnolo G (2012) Reputation, competition, and entry in procurement. Int J Ind Organ 30(3):291–296

Toth B, Fazezakas M, Czibik A, Toth IJ (2014) Toolkit for detecting collusive bidding in public procurement with examples from Hungary. Corruption Research Center Budapest (CRC-WP/2014/02)

Vechea JA (2013) A Pareto-abistitial model for analysing species co-occurrence. Glob Ecol Biogeogr 22(2):252–260

Wachs J, Kertész J (2019) A network approach to cartel detection in public auction markets. Sci Rep 9:10818

Wachs J, Yassen T, Lengyel B, Kertész J (2019) Social capital predicts corruption risk in towns. R Soc Open Sci 6(4):182103

Wachs J, Fazezakas M, Kertész J (2020) Corruption risk in contracting markets: a network science perspective. Int J Data Sci Anal 12:1–16

Whiteman R (2019) Fraud and corruption tracker. The Chartered Institute of Public Finance and Accountancy—CIPFA

Zeferino LF (2020) A corrupção na construção de edifícios públicos no brasil: análise de instrumentos inibidores e facilitadores na etapa de projeto arquitetônico. Master's thesis

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.