Journal on Interactive Systems, 2024, 15:1, doi: 10.5753/jis.2024.3836

Exploring Irregularities in Brazilian Public Bids: An In-depth Analysis on Small Companies

Camila S. Braz [**Universidade Federal de Minas Gerais** | ***camilabraz@ufmg.br***]
Marco Túlio Dutra [**Universidade Federal de Ouro Preto** | ***marco.dutra@aluno.ufop.edu.br***]
Gabriel P. Oliveira [**Universidade Federal de Minas Gerais** | ***gabrielpoliveira@dcc.ufmg.br***]
Lucas G. L. Costa [**Universidade Federal de Minas Gerais** | ***lucas-lage@ufmg.br***]
Mariana O. Silva [**Universidade Federal de Minas Gerais** | ***mariana.santos@dcc.ufmg.br***]
Michele A. Brandão [**Instituto Federal de Minas Gerais** | ***michele.brandao@ifmg.edu.br***]
Anisio Lacerda [**Universidade Federal de Minas Gerais** | ***anisio@dcc.ufmg.br***]
Gisele L. Pappa [**Universidade Federal de Minas Gerais** | ***glpappa@dcc.ufmg.br***]

Computer Science Department, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG, 31270-010, Brazil.

Received: 09 November 2023 • **Accepted:** 28 March 2024 • **Published:** 04 April 2024

Abstract In Brazil, bidding processes constitute the main method through which the Public Administration acquires goods and services, and they aim to select the best proposal between several bidding companies. Analyzing public bids can reveal several negotiating characteristics between companies and the public sector, including alerts of fraudulent activities involving such businesses. This article presents two approaches for detecting irregularities within small companies using data extracted from public bids in the Brazilian state of Minas Gerais. For each approach, we perform exploratory and geospatial analysis to better understand specific characteristics of the companies with irregularity alerts. Furthermore, we execute a network analysis to examine the underlying connections between such companies. Our findings reveal the efficacy of both approaches in indicating small companies that may be involved in fraudulent activities. Our methodology and results represent a significant advance for the public sector as they have the potential to enhance mechanisms for overseeing and preventing fraud within bidding processes.

Keywords: public bids, fraud detection, e-government, small businesses, network analysis

1 Introduction

In Brazil, public bids are mechanisms used by the public sector to acquire goods or contract services from private companies. In short, bidding processes aim to select the proposal that offers the best result for the Public Administration, ensuring equality, fair competition between bidders, adequate prices, and stimulating innovation and the country's sustainable development[1]. Such processes apply to the Public Administrations of the Union, the 26 States, the Federal District, and cities, and they are currently regulated by Law No. 14,133, of April 1, 2021[2], which establishes a sequence of phases and the modalities in which public bids happen.

Participating in public bids represents an excellent opportunity for companies (i.e., bidders), as the public sector can acquire goods and/or services offered by the private sector, ensuring a stable revenue [Costa *et al.*, 2022]. To compete in a public bid, companies must comply with the legislation, possess all the necessary documents, and not be involved in the basic contracting project or be part of the contracting entity[3]. This competitive process not only presents an opportunity for businesses to secure government contracts but also emphasizes the importance of transparency, fairness, and adherence to stringent criteria in pursuing public projects [Luna and Figueiredo, 2022].

Indeed, several laws regulate the operation and functioning of companies, especially when participating in public bids [Pereira *et al.*, 2022]. In this context, seminal works in the analysis of public bids focus on distinct aspects of fraud detection, including cartels between bidders [Gabardo and Lopes, 2014], bid characterization [Luna and Figueiredo, 2022], alerts based on network analysis [Costa *et al.*, 2022; Pereira *et al.*, 2022], and inconsistencies between bidder activities and bidding items [Oliveira *et al.*, 2022a] Furthermore, Silva *et al.* [2022] also propose LiPSet, a dataset of public bid documents, and mention fraud detection as one of its applications.

In this work, our research goal is to analyze public bids to generate alerts regarding bidders who are small-sized companies. Our main contribution is to propose two auditing approaches that assess different aspects of bidding companies. The first approach analyzes public bids with small-sized bidding companies with annual revenue above the allowed limit. The second approach analyzes bids with small-sized bidding companies linked to legal entities. Both approaches are part of the scope of the Analytical Capabilities Program and were defined in collaboration with experts from the Prosecution Service of the State of Minas Gerais (in Por-

[1] Bidding and Contracts - Transparency Portal: `https://www.portaltransparencia.gov.br/entenda-a-gestao-publica/licitacoes-e-contratacoes`, access on 28 March 2024.
[2] `https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14133.htm`, access on 28 March 2024.
[3] Supplier Selection Criteria: `http://www.tcu.gov.br/arquivosrca/001.003.011.048.htm`, access on 28 March 2024.

tuguese, *Ministério Público do Estado de Minas Gerais*, or simply MPMG). Identifying bids with these characteristics is crucial, as it raises fraud alerts, allowing experts to analyze them in more detail.

This article extends a full paper from the 11th Workshop on Computing Applied to E-Government (WCGE 2023) [Braz *et al.*, 2023]. As new contributions, we first update our analyses with a new data load to solve quality problems [Oliveira *et al.*, 2022b]. Next, we execute a geospatial analysis to identify whether there are areas where the companies with an irregularity alert are concentrated. Furthermore, we perform a network analysis on such companies to investigate the connections between them, including detecting communities of companies with such alerts. Overall, our new results reveal meaningful insights into the behavior of the companies with alerts in public bids, including how they connect with each other and how they are organized.

This article is structured as follows. First, Section 2 presents the basic concepts on network science. Then, Section 3 describes related work. Section 4 presents the main steps for conducting this work. Section 5 characterizes various bids considered as input in the two auditing approaches and Section 6 presents the network analysis over the companies with irregularity alerts. Section 7 details real bids captured by the auditing approaches defined in this work. Finally, Section 8 concludes this work, outlining possible limitations and suggesting future research.

2 Background on Network Science

In this section, we introduce the basic network science concepts used throughout this work. In short, network modeling is a computational technique for representing and analyzing relationships between entities using graphs. Such graphs consist of nodes (or vertices) representing entities and edges representing connections or relationships between such entities. They provide a visual and mathematical framework to study the dynamics, behavior, and structure of these complex interconnections [Barabási, 2016]. Regarding its applications, network analysis can offer several benefits to public bidding data, enhancing transparency, efficiency, and accountability in the process [Costa *et al.*, 2022].

Here, we use well-established network science metrics to assess relationships within the context of public bids. Such metrics focus on the topological characteristics of the network, specifically addressing the structure of nodes and edges. The metrics used are briefly described as follows[4].

Degree and Weighted Degree. Measure the connectivity of individual nodes within the network. The degree of a node is the number of edges incident upon it, whereas the weighted degree is the sum of the weights of such edges.

Density. Measures the proportion of actual connections in a network relative to the total number of possible connections.

Path Length. The number of steps or edges required to travel from one node to another in a network.

Diameter. The longest shortest path between any pair of nodes in a network. It indicates the network's overall size and connectivity.

Clustering Coefficient. The tendency of neighboring nodes to be interconnected among themselves. Higher values of such a coefficient indicate a more interconnected neighborhood for the given nodes.

Besides building and characterizing networks with such metrics, we tackle the *community detection* problem, which is a task that involves identifying groups of nodes in a network that are more densely connected to each other than to nodes outside those groups. It plays an important role in understanding the structure and function of complex systems, such as social networks, biological networks, and information networks. Detecting communities helps uncover hidden patterns, relationships, and substructures within networks, which are essential for various applications, including recommendation systems, anomaly detection, and understanding the dynamics of real-world networks [Barabási, 2016].

3 Related Work

Since the enactment of the Access to Information Act in Brazil (Law No. 12,527, of November 18, 2011[5]), several works have emerged to analyze Brazilian public data to obtain relevant information and/or knowledge related to public administration. For instance, Lyra *et al.* [2021] perform a characterization of a network of co-bidding entities, whereas Coelho *et al.* [2022] focus on text classification of legal opinions. Moreover, the use of artificial intelligence in the public sector is addressed by Monteiro *et al.* [2023], who analyze pitfalls in governmental chatbots.

In addition, the Brazilian Act for the Protection of Personal Data (in Portuguese, *Lei Geral de Proteção de Dados*, LGPD – Law No. 13,709, of August 14, 2018[6]) establishes the basis for data protection by regulating the usage of personal data [Beppu *et al.*, 2021; Reis *et al.*, 2023]. Along with such legislation, there is also the movement of Open Government Data (OGD), which occurs in various countries to promote transparency, accountability, and value creation by releasing government data to all citizens [Buryakov *et al.*, 2022].

In this context, Nai *et al.* [2022] present a literature review of recent research addressing fraud detection in public organizations. In this work, the authors observed that machine learning algorithms have been widely used in these studies. In the Brazilian context, Lima *et al.* [2020] use deep neural networks to analyze fraudulent collusion risk. Moreover, Oliveira *et al.* [2022a] propose a heuristic model to detect inconsistencies between the bidder activity and the bidding item, and Galvão Júnior *et al.* [2023] combine set theory with a classification task to identify fraud related to collusion.

Besides machine learning techniques, network science has been widely used in investigating fraud in the public sector. Specifically, Figueiredo *et al.* [2020] propose a model

[4]For formal definitions and more information on such metrics, see Barabási [2016].

[5]`https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm`, access on 28 March 2024.

[6]Brazilian Act for the Protection of Personal Data: `https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`, access on 28 March 2024.

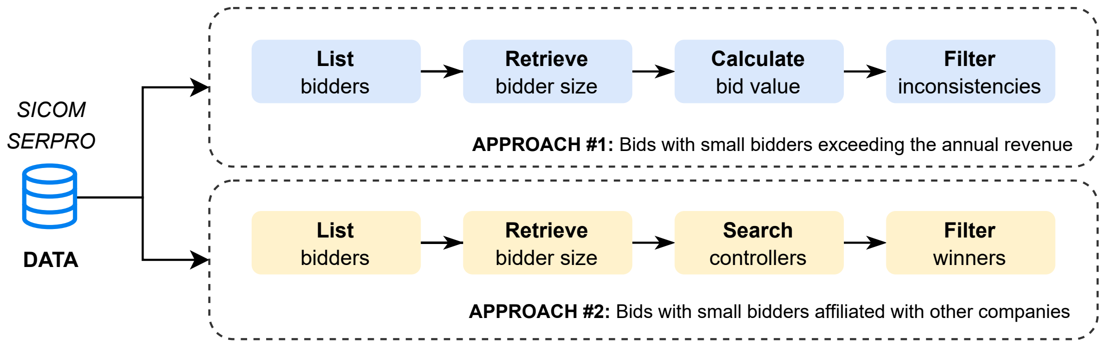

Figure 1. Modeling of the two auditing approaches created to detect irregularities in small companies.

for identifying relevant nodes in a network using Operation Phantom Bidder as a case study. Furthermore, Costa *et al.* [2022] propose a model for generating fraud alerts from a social network. This work introduces the concept of audit trails, defined as a sequence of steps followed to identify signs of specific types of irregularities in bidding processes.

Going further, Pereira *et al.* [2022] and Luna and Figueiredo [2022] also use social network metrics and characteristics of public bids to identify signs of corruption and fraud risk. The main difference between these two works is that the former proposes an approach to identify a company as involved in fraud or not, while the latter characterizes public bids and proposes a metric that can raise a fraud alert. More recently, Brandão *et al.* [2023] propose a semi-automated pipeline focusing on public bids. Such a pipeline consists of a classification and a data quality module, and its applications include audit trails for fraud detection and overpricing identification.

The work presented here is similar to those discussed in this section, which analyze data from the Brazilian public sector to evaluate potential frauds. However, the main difference lies in the type of analysis conducted on these bids to raise fraud alerts. In particular, this work presents two auditing approaches to identify possible fraud in small-sized bidding companies. Besides the paper from which this article is extended [Braz *et al.*, 2023], to the best of our knowledge, there is no existing work in the field of computer science that specifically examines fraud alerts in such a context.

4 Methodology

This section presents the methodology used for this work, illustrated by Figure 1. First, we describe the dataset of public bids and bidders on which irregularities will be investigated (Section 4.1). Next, we describe the modeling of the two auditing approaches created to detect irregularities in small companies: "Bids with small bidders with annual revenue above the limit" (first approach, Section 4.2) and "Bids with small bidders affiliated with other companies" (second approach, Section 4.3).

4.1 Data

In this work, we consider data provided by the Prosecution Service of the State of Minas Gerais (MPMG) through its Analytical Capabilities Program. Such data are available in a data warehouse that aggregates heterogeneous information from various sources. Some data contain confidential information and will be briefly described here. Despite the wide variety of available information, only data related to bids and bidding companies are used.

Regarding bidding data, we consider processes that occurred in the state of Minas Gerais at both city and state levels. For city-level bids, we use data from the Computerized System of Municipal Accounts (SICOM), a technology developed by the State Audit Court of Minas Gerais (in Portuguese, *Tribunal de Contas do Estado de Minas Gerais*, or simply TCE-MG) that aggregates information from the transparency portals of the 853 municipalities in Minas Gerais. As for state bids, the data is obtained directly from the Transparency Portal of the State of Minas Gerais.[7]

Additional information about bidding companies is required to implement our two proposed approaches, including registration data and corporate information. For this purpose, private data from the Federal Service of Data Processing (SERPRO) is used. Aggregating such information with public bidding data enriches the dataset for better application in investigating possible irregularities.

It is important to highlight that all the aforementioned information was directly obtained from a structured data warehouse composed of tables provided by MPMG. Therefore, our data integration process is done by joining the required tables, and inconsistent data are not considered after joining. In this work, each proposed approach follows a specific integration process according to its objectives. We detail such processes in the following sections. Moreover, due to the combined nature of public and private information, we cannot provide more details on the schema of our dataset.

Overall, our final dataset comprises a total of 763,566 public bids, of which 403,932 are municipal and 359,634 are state bids. These bids occurred from 2007 to 2022 and are divided into 18 modalities, including in-person and electronic auctions, invitations, and contests. In addition, the dataset contains information about 110,387 bidders, including individuals and legal entities. However, since the proposed approaches focus specifically on companies, only bidders that are legal entities are considered.

[7]Transparency Portal of the State of Minas Gerais: `https://www.transparencia.mg.gov.br/licitacoes-e-contratos/compras-e-contratos`, access on 28 March 2024.

Table 1. Revenue limit considered for each type of company.

Type	Period	Anual limit	Legislation
MEI	2009–2011	R$ 36,000	Complementary Law No. 128, of December 19, 2008
	2012–2017	R$ 60,000	Complementary Law No. 139, of November 10, 2011
	2018–2022	R$ 81,000	Complementary Law No. 155, of October 27, 2016
ME	2009–2010	R$ 240,000	Complementary Law No. 123, of December 14, 2006
	2011–2022	R$ 360,000	Complementary Law No. 139, of November 10, 2011
EPP	2009–2010	R$ 2,400,000	Complementary Law No. 123, of December 14, 2006
	2011–2017	R$ 3,600,000	Complementary Law No. 139, of November 10, 2011
	2018–2022	R$ 4,800,000	Complementary Law No. 155, of October 27, 2016

4.2 First Approach: Bids with Small Bidders Exceeding the Annual Revenue Limit

In this section, we provide a detailed explanation of the implementation of the first approach, which aims to identify "Bids with small-sized bidders exceeding the annual revenue limit." Such an approach focuses on finding bids in which small companies have exceeded the annual revenue limit defined by their size. To do so, we analyze bids held from 2009 to 2022, and companies classified as Individual Microentrepreneur (*Microempreendedor Individual,* – MEI), Microenterprise (*Microempresa* – ME), and Small-Sized Company (*Empresa de Pequeno Porte* – EPP).

Table 1 presents the maximum annual gross revenue for each company size and year. To conduct the audit, data specified in Section 4.1 needed to be cross-referenced. Initially, the bid data extracted from SICOM and SERPRO were used, and they were joined by the identification number of the bidder (in Brazil, this is made through CPF and CNPJ, which are unique identification numbers for people and companies, respectively). This join allowed us to determine the size of CPF/CNPJs authorized to participate in a bid within a time window when their registration was active in the Federal Revenue of Brazil (in Portuguese, *Receita Federal do Brasil*). To do this, we ensured that the reference date is greater than or equal to the start date of the company's activity and the date of the registration status.

Furthermore, by analyzing SICOM data, relevant information about bids, such as the number of items approved per bidder and the unit value of the items, was retrieved. Based on this data, the approved value of each bid was calculated, which is equivalent to the product of the unit value by the approved quantity. Subsequently, the total approved values were grouped by the year of the bid process and the CPF/CNPJ number of the bidder.

After the grouping, it was necessary to cross-reference SERPRO data with SICOM data, but only for those resulting from the grouping of CPF/CNPJ, annual approved value, and year of the process, using CPF/CNPJ and the year of the process as keys. Then, a restriction was applied to filter bids where the bidder exceeded the revenue limit for their size in the year of the bid process, using the information presented in Table 1.

4.3 Second Approach: Bids with Small Bidders affiliated with other Companies

This section details the implementation of the second approach, which aims to investigate "bids with small-sized bidders linked to legal entities". This approach seeks to identify indications that small-sized companies and micro-enterprises winning bids, although formally constituted, do not carry out their activities. To obtain the benefits granted to them in the bidding process, they are controlled by another organization (i.e., the controller).

To implement this auditing approach, we conducted a data merge of the bid data from SICOM with SERPRO data (which contains the company's size), using the CPF/CNPJ number of the bidder as the common identifier. This data integration was necessary to retrieve the CPF/CNPJs authorized to participate in a bid within a time window when their registration was active in the tax authority.

Due to the temporal nature of SERPRO data, the reference date was used to identify the time window in which the bid falls. Additionally, we ensured that the reference date is greater than or equal to the start date of the company's activity. Finally, to identify only the winners of the bid process, the data was filtered according to this condition.

5 Characterization of Public Bids Identified by Audit Approaches

From the bidding data set and the development of auditing approaches, it is possible to select the bids with irregularity alerts related to bidders that are small-sized companies. Sections 5.1 and 5.2 characterize the bids that triggered alerts for approaches 1 and 2, respectively.

5.1 First Approach: Bids with Small Bidders Exceeding the Annual Revenue Limit

In this section, we present a characterization of the results of the first approach. A total of 226,435 bids were identified, of which 216,385 were won by bidders with annual revenue above the allowed limit for their size. Additionally, 18,932 bidders with revenue restrictions were identified. Figures 2A and 2B display the distribution of these bids by mesoregions[8] and modalities, featuring only the winning bidders. State-level bids were grouped into a single category labeled "State

[8]Mesoregions are subdivisions of the Federative Units of Brazil classified by the Brazilian Institute of Geography and Statistics (IBGE).

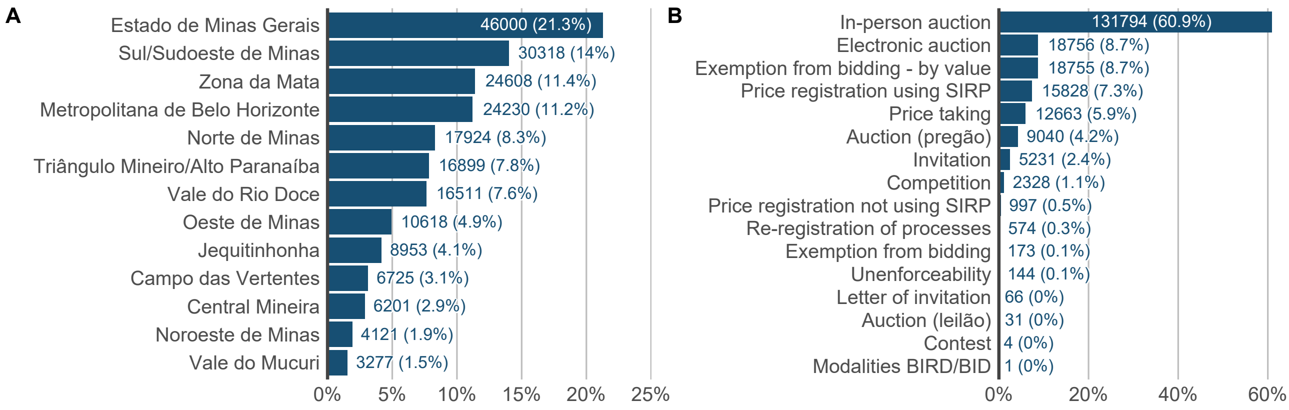

Figure 2. Distribution of bids falling under the first approach by (A) mesoregion and (B) modality. In parentheses is the percentage of the quantity of the total number of bids.

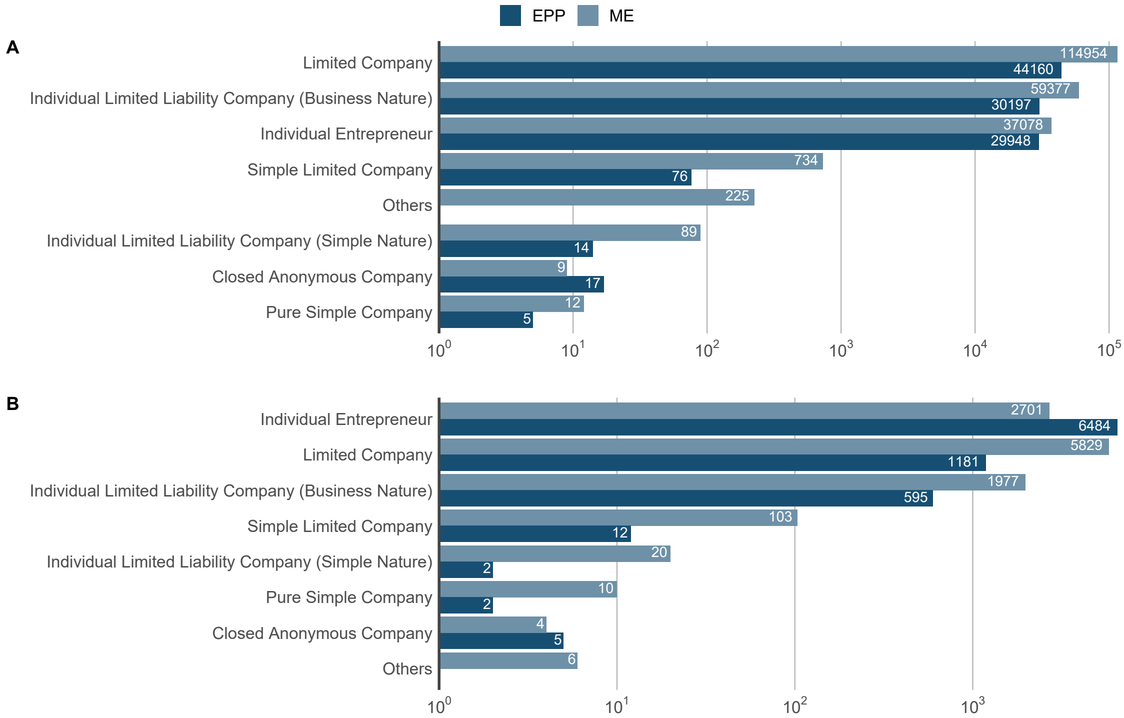

Figure 3. Distribution of (A) bids and (B) bidders falling under the first approach, categorized by legal nature.

of Minas Gerais". In total, 46,000 state-level bids were identified, representing 12.8% of the total. It is important to note that 78.7% of the bids identified by this approach are municipal, totaling 170,385 bids distributed across 12 mesoregions, representing 42.2% of the total municipal bids.

Regarding modalities, we analyze the values of the "modality" field for each bid in SICOM. Overall, bids from 16 modalities were identified: in-person auction, electronic auction, exemption from bidding (by value or not), price registration (using SIRP or not), price taking, other auctions (*pregão* or *leilão*), invitation, competition, re-registration of processes, unenforceability, letter of invitation, contest, modalities BIRD/BID. Bids with invalid, unavailable, and unspecified modalities were also identified. Among these, the in-person bidding modality was the most frequent, totaling 131,794 bids, accounting for 60.85% of the total. It's important to highlight that many bids in this modality may be related to the ease of participation for companies, which may favor businesses controlled by other organizations.

We also analyzed the legal nature of the bids and bidders. Since each bid can have more than one bidder, categorized as Individual Microentrepreneur (MEI), Microenterprise (ME), and Small-sized Company (EPP), for more precise analysis, the number of bids and bidders was evaluated according to each size category. Figures 3A and 3B respectively demonstrate the number of bids and bidders by legal nature, a classification by which a company's categorization and structure are determined. Bidders of ten different legal nature types were identified. No MEI companies were identified by this approach, while for ME and EPP, the most predominant legal nature is "Limited Company".

Lastly, Figure 4 presents the distribution of the difference between the total value and the size limit value for each type of bidder. We perform a statistical test (T-Test) to verify

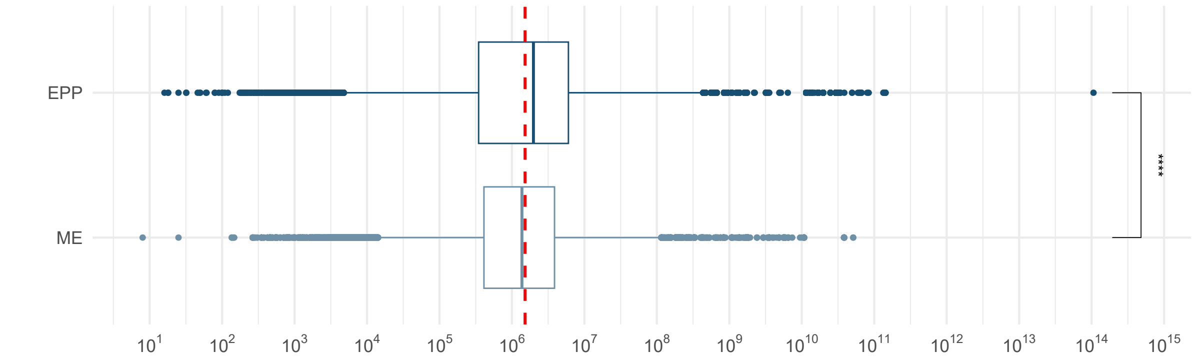

Figure 4. Distribution of the difference between the total value and the size limit value for each type of bidder. The dashed red line represents the overall median. The *p-value* levels are symbolized as (1) ns: $p > 0.05$, (2) ***: $p < 0.001$, (3) ****: $p < 0.0001$.

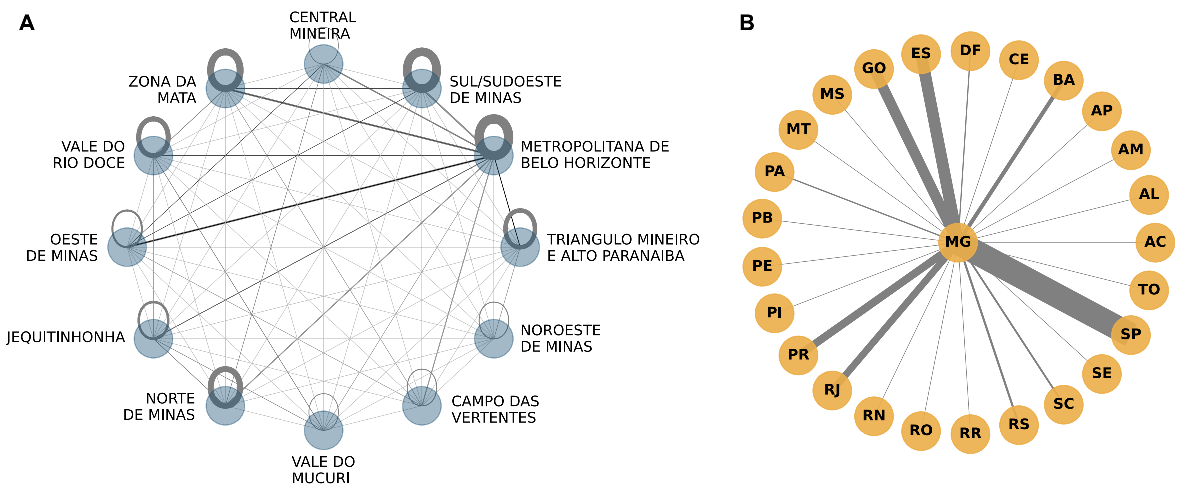

Figure 5. Connections between the bidders' locations and the bidding cities from the first approach: bidders from Minas Gerais by mesoregion (A) and from outside Minas Gerais by state (B).

whether the distributions differ significantly. The results suggest considerable variation in the difference between the total value and the size limit value for each type of bidder, with the most significant difference observed for small-sized companies (EPP). Furthermore, most companies showed a difference between these values of approximately R$1,516,549, which may indicate a tendency to submit values close to the maximum size limit. However, around 10% of entries have a difference exceeding R$11 million, which may indicate potential irregularities in bids involving these companies.

Geographical Analysis. We further extend our analysis to include a geographic dimension considering the findings of the first approach. Such an analysis aims to verify whether there are specific regions in which companies flagged for irregularities tend to concentrate. To do so, we compare the company's location (as it is registered in the Federal Revenue of Brazil) and the city where the public bid happened. Overall, our geographical analysis is divided into two distinct segments: the first segment evaluates companies situated within the boundaries of Minas Gerais, whereas the second segment focuses on companies located in other Brazilian states, that is, those outside of Minas Gerais.

Figure 5 presents network visualizations of the connections between the bidders' locations and the cities where the bidding occurred. Specifically, Figure 5A only considers companies in Minas Gerais. Thus, the lines connect the mesoregions of the bidding company and the bidding city, and their width represents the number of these relationships (i.e., the number of bidders from one mesoregion who participate in bids in another mesoregion). The figure reveals that all 12 mesoregions of the state have connections between them. However, some links are stronger than others. For example, the link between the mesoregions "Metropolitana of Belo Horizonte" with "Oeste de Minas", "Zona da Mata" and "Sul/Sudoeste de Minas" is greater than with other regions. This means that it is more common for bidders with an irregularity alert to come from these mesoregions, which contain some of the largest cities in the state, such as Belo Horizonte (the capital), Divinópolis, Juiz de Fora, and Poços de Caldas, respectively.

Furthermore, there are loops in the nodes of all mesoregions, representing cases in which bidding companies and bidding cities are located in the same region. In some mesoregions, such links have the greatest width, indicating that most companies bidding with alerts are in the same mesoregion as the cities whose bids generated the alerts. Once again, the "Metropolitana de Belo Horizonte", "Sul/Sudoeste of Minas" and "Zona da Mata" mesoregions appear prominently in this analysis. Besides concentrating cities with large populations, such regions present large industrial and commercial hubs, which naturally result in many companies.

Figure 5B presents a similar visualization with the connec-

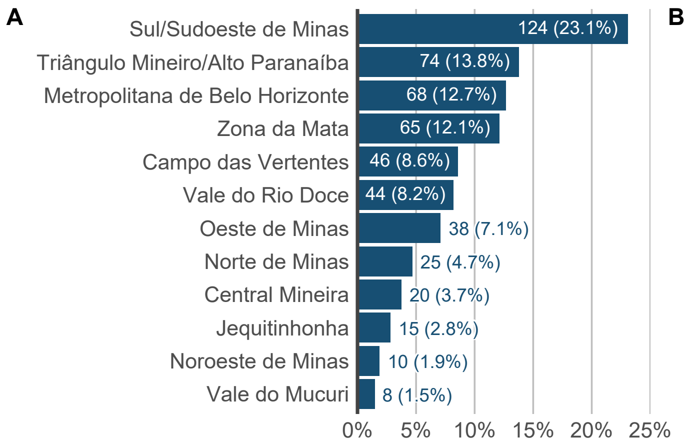

Figure 6. Distribution of bids falling under the second approach by (A) mesoregion and (B) modality. In parentheses is the percentage of the quantity in relation to the total number of bids.

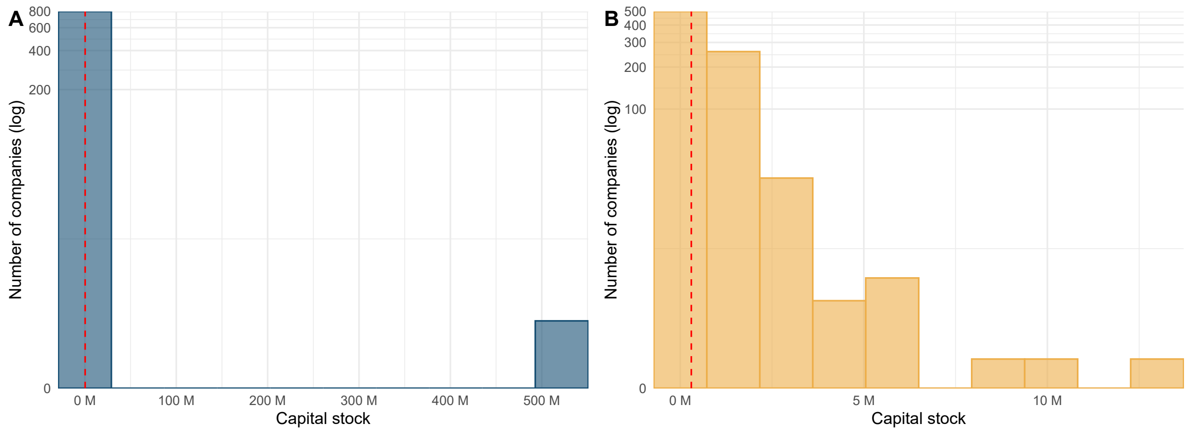

Figure 7. Capital stock of (A) controlling companies and (B) controlled companies. The dashed line indicates the median of the analyzed values.

tions between bidding companies with alerts that are located outside of Minas Gerais (MG). Such companies are spread across 24 states and the Federal District (DF), representing almost all of Brazil's federative units (the State of Maranhão is the exception). Among them, the state of São Paulo (SP) has the most significant number of bidders with alerts, followed by Espírito Santo (ES), Goiás (GO), Paraná (PR), Rio de Janeiro (RJ), and Bahia (BA). This can be explained in part by the fact that all of these states are geographically close to Minas Gerais (except for Paraná, all of them have boundaries with it), which would facilitate the participation of companies from these states in bids from Minas Gerais.

5.2 Second Approach: Bids with Small Bidders affiliated with other Companies

This section presents the analysis of the results obtained by the second approach. In total, 537 bids were identified in which small-sized bidders were affiliated with other legal entities, which corresponds to 0.07% of the total bids analyzed. Figures 6A and 6B depict the distribution of these bids by mesoregions, considering state and municipal levels, as well as the distribution by type of modality. All state-level bids were grouped into a category labeled "State of Minas Gerais". However, in this approach, no state-level bid was identified. Therefore, all bids are city-level, distributed across the 12 mesoregions of Minas Gerais.

Furthermore, in the analysis of the modalities of bids in Figure 6, it can be noted that bids of five different modalities were identified. Among them, the "In-person auction" modality was the most frequent, totaling 412 bids, accounting for 76.7% of the total. The other modalities had lower frequencies, ranging from 1.1% to 16.0% of the total.

The capital stock of both controlling and controlled companies was also analyzed. Capital stock is a measure of a company's financial power, representing the initial investment made by the partners at the time of its founding. Figures 7A and 7B display the distribution of capital stock for controlling and controlled companies, respectively. Note that the vertical axis, representing the number of companies, is on a logarithmic scale. Note that the horizontal axis has different scales. The results indicate that controlling companies have a higher capital stock than the controlled companies, which is expected since controlling companies are generally larger and more established in the market.

As a final analysis, Figure 8 presents the number of bids categorized by the legal nature of controlling and controlled companies. The results indicate that most companies falling under this approach have the legal nature of "Limited Company". In other words, they are companies with well-structured organizational levels. Furthermore, this legal nature is one of Brazil's most common forms of company con-

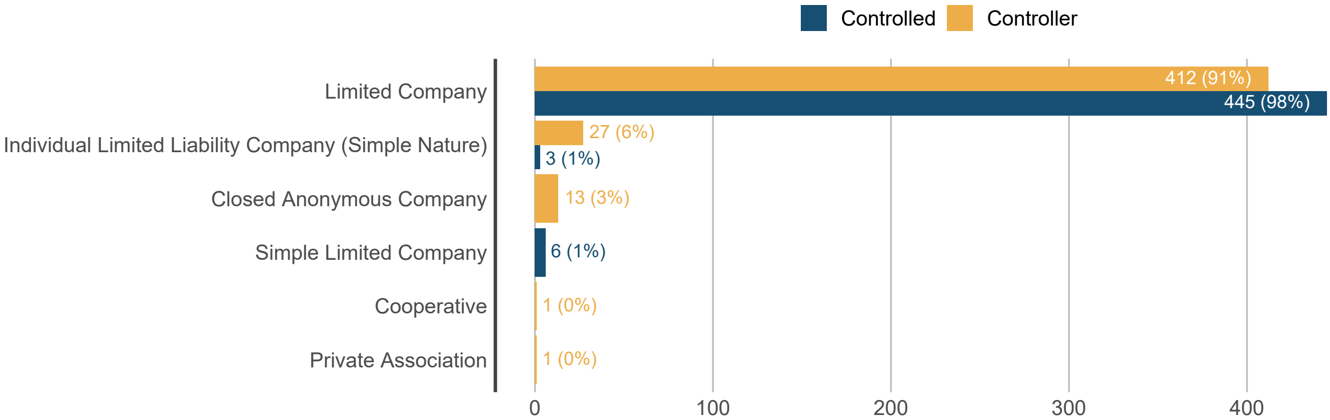

Figure 8. Quantity of bids by legal nature.

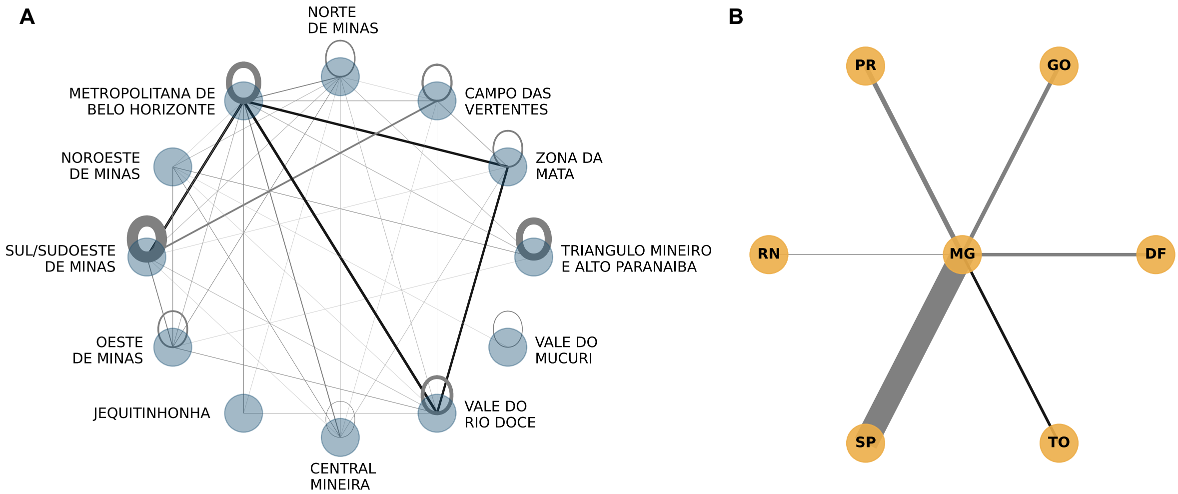

Figure 9. Connections between the bidders' locations and the bidding cities from the second approach: bidders from Minas Gerais by mesoregion (A) and from outside Minas Gerais by state (B).

stitution, which may explain its prevalence in the analyzed sample. Hence, most companies are expected to comprise the controlling/controlled relationship.

On the other hand, private associations and closed anonymous companies are the minority of companies categorized in this approach, as it is not expected for companies with more superficial organizational structures to either be controlled or act as controlling entities. Therefore, companies with this legal nature may indicate atypical situations or areas that warrant closer scrutiny in analyzing potential bid frauds or irregularities.

Geographical Analysis. Similar to the previous section, we perform a geographic analysis with the results of the second approach to check whether there is a concentration of alerted bidders in a given region. In Figure 9A, we specifically analyze the alert bidding companies located in Minas Gerais and relate them to the location of the bidding cities. Although the number of connections is not similar to that found in the first approach, there are still some significant relationships. First, the "Sul/Sudoeste de Minas", "Triângulo Mineiro e Alto Paranaíba" and "Metropolitana de Belo Horizonte" mesoregions are those with the widest loop edges; that is, they have the largest number of companies with alert that supply to cities in the same mesoregion. Furthermore, the "Jequitinhonha" mesoregion does not have a loop edge, indicating that all alerted bidders who participate in bids for cities in this mesoregion are from outside it.

Furthermore, there is a strong connection between the "Metropolitana de Belo Horizonte", "Zona da Mata" and "Vale do Rio Doce" mesoregions. As mentioned in the previous section, such regions comprise important commercial and industrial hubs in the state, with many companies installed. For example, the three mesoregions comprise six of the ten most populous cities (Belo Horizonte, Contagem, Juiz de Fora, Betim, Ribeirão das Neves, and Governador Valadares) and six of the ten cities with the higher GDP in Minas Gerais (Belo Horizonte, Contagem, Betim, Juiz de Fora, Nova Lima, and Ipatinga)[9].

Finally, Figure 9B shows the interstate connections between bidders that are located outside of Minas Gerais (MG). The second approach produced alerts for companies from only six federative units: Distrito Federal (DF), Goiás (GO), Paraná (PR), Rio Grande do Norte (RN), São Paulo (SP) and Tocantins (TO). Again, the state of São Paulo appears with the highest number of cases, which can be related to its geographic proximity to Minas Gerais and also because it is the most populous state and has the highest GDP in the country[10],

[9]The data on population and GDP of the cities were extracted from IBGE: `https://cidades.ibge.gov.br/`, access on 28 March 2024.

[10]`https://www.en.investe.sp.gov.br/why-sao-paulo/diversified-economy/gdp/`, access on 28 March 2024.

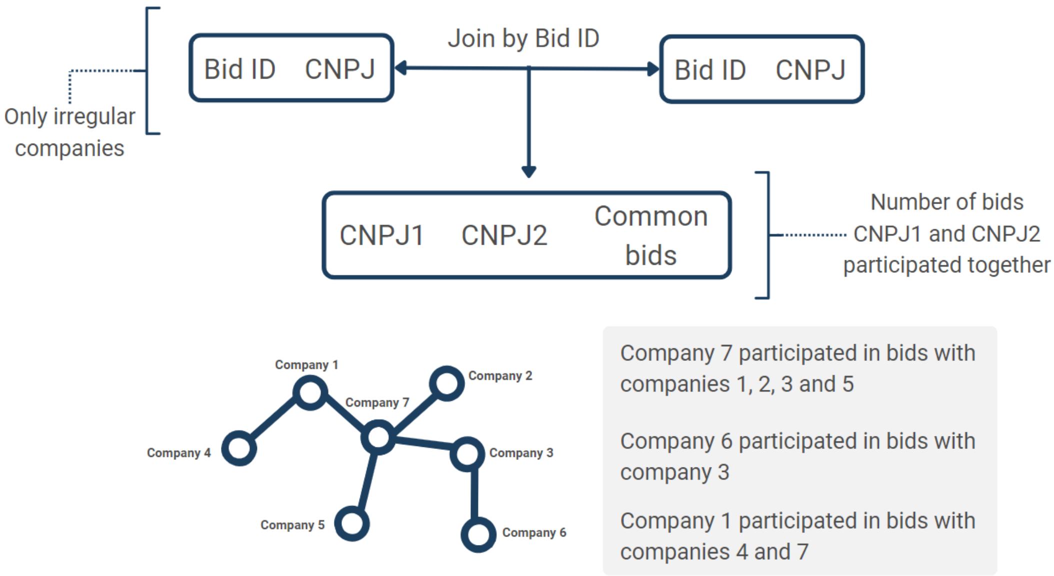

Figure 10. Network building process. Such a network connects companies with alerts that participate together in public bids.

being Brazil's largest industrial and commercial hub.

6 Network Analysis of Irregularities

This section delves into using network analysis techniques (see Section 2) to uncover relevant patterns within companies that participated in public bids. Here, we focus on the network that connects the companies with irregularity alerts detected by our two approaches. First, in Section 6.1, we present the network building process. Then, Section 6.2 contains a network characterization with traditional network metrics. Finally, in Section 6.3, we apply a community detection method to uncover association patterns within our network.

6.1 Network Building

In this work, we build a network of bidding companies with alerts generated in the two approaches detailed in Section 4. Specifically, each node within the network represents a company flagged with alerts, while edges denote connections between companies participating in the same public bid.

We define the network as weighted, and the weight of each edge corresponds to the frequency of joint participation in bids by the connected companies. Furthermore, the edges are undirected, as the order of companies is not important for our analytical framework. The building process of our network is illustrated in Figure 10.

6.2 Network Characterization

In this section, we characterize the company network that emerges from the alerts generated by the two approaches proposed in this study. Table 2 presents the metric values for the complete network and its giant component (i.e., the largest connected component). Such metrics are well-established indicators in the network science field and are important descriptors of the network structure, revealing relevant information about how nodes are connected. The complexity of this network is revealed by its size, with a total of 12,777 nodes interconnected by 95,659 edges. Besides having over 700 connected components, our network is well connected, as the giant component of the network contains approximately 81.2% and 95.8% of the network's nodes and edges, respectively. The dynamics of the connections highlight the richness of relationships between bidders, offering a detailed and comprehensive view of the scenario explored in our analysis.

Table 2. Network statistics (complete and giant component).

Metric	Complete	Giant
Nodes	12,777	10,380
Edges	95,659	91,648
Average Degree	15.0	17.7
Average Weighted Degree	48.4	58.5
Density	0.001	0.002
Diameter	–	20
Average Path Length	–	4.7
Average Clustering Coefficient	0.610	0.574
Connected Components	701	1

Furthermore, the average degree of the nodes is considerable (15.0 for the complete network and 17.7 for the giant component), which can be explained by the existence of central companies highly connected with others that compete in many public bids. The number of these bids (i.e., the weight of the edges) is also relevant since the values for the weighted average degree are much higher. Since our analysis focuses exclusively on companies with alerts, the average degree would be even higher in a network containing all bidding companies.

Although the average degree values (weighted or not) are considerable, the density values reveal that both the network and its giant component are sparse (0.001 and 0.002, respectively). In short, the density of a network is the ratio between the number of edges existing in a network and the total number of possible edges in that same network (i.e., a clique, which is a graph where every vertex is directly connected to all the others). Even so, the average length of the network's paths reveals that the connections between nodes are well distributed, allowing easy access to its ends. The longest path

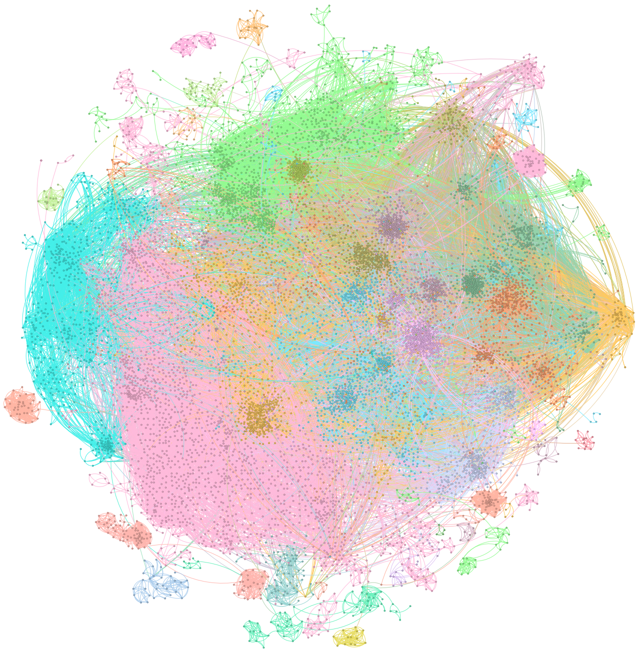

Figure 11. Overview of the communities in the giant component.

in the network (i.e., its diameter) is 20, which indicates that the bidding companies are at most 20 nodes apart. This high diameter value reveals that many bidding companies are not working together in a public bid. This result is reinforced by the low density value, 0.001 for all nodes in the network and 0.002 for the giant component, i.e., the number of direct connections between bidding companies is not large.

However, the insights derived from the characterization may not be easily applied to all individual cases in the network. In other words, nodes within the network may not fit the general pattern identified during characterization. Therefore, it is crucial to conduct a more detailed and specific analysis. Next, we perform community detection to identify nodes that share similar behaviors, allowing a more refined understanding of the dynamics present in the network.

6.3 Community Detection

In this work, we use the Louvain algorithm to detect communities based on the concept of modularity, a measure of the quality of a community structure within a network [Barabási, 2016]. We run such a method because it has a low computational cost and is suitable for larger networks [Blondel *et al.*, 2008]. In short, the algorithm aims to maximize the modularity score, quantifying how much the network can be divided into densely connected communities with sparse connections between them. The algorithm operates through a two-step iterative process: the first step optimizes local modularity within small communities, and the second step aggregates these communities to form a new network, over which the process is repeated.

We applied the community detection algorithm only to the giant component of the network, as each small component would be a community in the complete network. Thus, the results reveal 25 communities in the giant component, illustrated in Figure 11 and summarized in Table 3. The largest community found corresponds to 2323 nodes (22.38%) in the giant component. Furthermore, the nine largest communities

Table 3. Largest communities in the network giant component.

Community	Nodes	%	Community	Nodes	%
1	2323	22.38%	6	649	6.25%
2	1609	15.50%	7	633	6.10%
3	267	9.32%	8	523	5.04%
4	803	7.74%	9	484	4.66%
5	698	6.72%	10 to 25	1691	16.29%

correspond to more than 83% of the nodes analyzed. Such communities comprise companies that participate together in bids and can indicate irregularities since they stand out due to their size and connections within the network. The largest communities may suggest potential collusion or cooperation among these companies, which could warrant further investigation by regulatory agencies or organizations looking to maintain a fair and competitive marketplace.

Moreover, examining the smaller communities can offer insights into niche interactions or specialized partnerships that may exist within the network. These smaller communities could indicate specific industry segments or regional associations, which might also be relevant for understanding market dynamics and potential irregularities. Following findings from previous work [Costa *et al.*, 2022; Brandão *et al.*, 2023], all computational techniques (including our audit approaches and network analysis) do not replace the manual investigation phase by content specialists. Still, they certainly can act as a guide for such inspection.

7 Case Studies in Real Bids

This section presents two case studies of real bids categorized under both approaches described in this article. Although there are bids with alerts for only one of the approaches, using the approaches in conjunction enhances the subsequent analysis of such bids by experts. This allows for prioritizing some bids in the investigation based on the alerts. As the bids presented here still require more detailed investigation, the data that allows for the identification of the entity (i.e., municipality or state) and the bidders have been anonymized.

The first bid took place in the year 2020 and had only one participating bidder. At the time, the company was a microenterprise (ME), and according to the current legislation, the annual revenue limit was R$360,000. However, the total of the values approved for the bidder in that year was approximately R$4.6 million, which is more than twelve times the allowed limit. Furthermore, this company has another controlling entity in its shareholder structure, raising a second alert of possible irregularity.

The second bidding process took place in the year 2021 and had 27 participants. Among these, 14 were identified as microenterprises or small-sized companies with revenue exceeding the allowed limit. Table 4 presents the list of these bidders, their type (i.e., microenterprise or small-sized company), the revenue limit for the year 2021, and the total values approved for these companies in that year. Additionally, the column "Legal Entity Link" indicates the presence of a legal entity in the shareholder structure of the companies. Concerning the approved values, some companies have values up to 43 times greater than the limit established by law.

Despite the significant results, it is not possible to con-

Table 4. Bidders with irregularities in a 2021 bidding process.

Bidder	Type	Revenue Limit	Homologation Total	Legal Entity Link
Company A	EPP	R$ 4,800,000	R$ 15,216,523.13	No
Company B	EPP	R$ 4,800,000	R$ 9,358,924.64	No
Company C	EPP	R$ 4,800,000	R$ 8,565,064.25	No
Company D	EPP	R$ 4,800,000	R$ 6,001,121.19	No
Company E	EPP	R$ 4,800,000	R$ 5,264,032.57	No
Company F	EPP	R$ 4,800,000	R$ 5,141,139.90	No
Company G	ME	R$ 360,000	R$ 15,678,546.65	No
Company H	ME	R$ 360,000	R$ 10,847,238.81	No
Company I	ME	R$ 360,000	R$ 9,604,140.71	No
Company J	ME	R$ 360,000	R$ 8,475,730.18	No
Company K	ME	R$ 360,000	R$ 6,698,865.65	No
Company L	ME	R$ 360,000	R$ 4,036,121.40	No
Company M	ME	R$ 360,000	R$ 3,109,644.00	No
Company N	ME	R$ 360,000	R$ 1,837,392.23	Yes

clude that these bids have irregularities based only on these findings. Audit approaches are essential tools for prioritizing the investigation of bids, but manual analysis by experts is still required. Only the manual analysis can solve potential data-related issues, including typing errors and loading process problems. However, since there are hundreds of thousands of bids in our dataset and many of these processes are created every day, the use of automated methods to generate alerts can simplify the investigation process for experts by raising possible indications of irregularities.

8 Conclusion

Considering the principles of Open Government Data and the importance of building automated solutions to deal with the massive volume of governmental data, this article presented two audit approaches that can be used to identify irregularities in small-sized companies participating in public bidding processes. The analysis of these approaches was conducted using data extracted from public bids in the Brazilian state of Minas Gerais. The first approach involved investigating bids with small-sized bidders with annual revenue exceeding the limit, while the second approach focused on analyzing bids with small-sized bidders affiliated with legal entities.

The results of both approaches revealed their capability to identify small-sized companies suspected of being involved in fraud. In particular, for the first approach, bids with several different legal natures were captured for the bidders, and most of the Microenterprises (ME) and Small-sized companies (EPP) are classified as "Limited Company". As for the second approach, it was found that the companies falling under it are mostly well-structured in terms of their organizational level. It is worth noting that despite the seemingly superficial nature of the proposed approach, this work is crucial for better filtering bids among the many that may indicate some form of fraud in both the public and private sectors.

Furthermore, the geographic analysis of bidders and the detection of communities in the network reveal important insights into how alerted bidders behave in bidding processes. In other words, such approaches enable a deeper and more comprehensive understanding of the relationship dynamics between these companies. For example, identifying geographic concentrations of companies and bidding communities allows government agencies to optimize their public procurement processes, strengthening the integrity and transparency of public procurement procedures. As a result, our results play a fundamental step in promoting electronic governance and building a more effective, responsible public administration aligned with the needs of society.

Limitations and Challenges. In this work, we did not conduct a comprehensive validation with all the bids identified by both approaches. Additionally, using private data and sensitive information may affect the reproducibility of the experiments. As a result, the execution of such approaches would be limited to institutions that have access to such information, including MPMG itself. Furthermore, due to the vast scope of SICOM, which aggregates data from all 853 municipalities in Minas Gerais, it is possible that problems may occur during the process of inserting this data into the system. Such potential difficulties can directly impact the accuracy and reliability of the results of our proposed approaches.

Future Work. We plan to present the bids identified by the approaches to experts to indicate whether there was indeed fraud. This validation may serve as a ground-truth for the proposed approaches to be used as attributes in a learning model capable of automatically classifying a bid as potentially fraudulent. In addition, we plan to refine our geographical analysis for a more accurate understanding of the concentrations of companies and their implications for public bids. As for the network analysis, we plan to enhance our modeling to identify connections not only at the business level but also in terms of personal or political relationships.

Declarations

Acknowledgements

The authors thank Bárbara M. A. Mendes, who collaborated on the previous version of this work.

Funding

This work was funded by the Prosecution Service of State of Minas Gerais (in Portuguese, *Ministério Público do Estado de Minas Gerais*, or simply MPMG) through its Analytical Capabilities Project and by CNPq, CAPES, FAPEMIG and the partnership project between AWS and CNPq.

Competing interests

The authors declare that they have no competing interests.

References

Barabási, A.-L. (2016). *Network science*. Cambridge University Press.

Beppu, F. R., Maciel, C., and Viterbo, J. (2021). Contributions of the Brazilian Act for the Protection of Personal Data for treating Digital Legacy. *Journal on Interactive Systems*, 12(1):112–124. DOI: 10.5753/jis.2021.1654.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Brandão, M. A., Reis, A. P. G., Mendes, B. M. A., Almeida, C. A. B., Oliveira, G. P., Hott, H., Gomide, L. D., Costa, L. L., Silva, M. O., Lacerda, A., and Pappa, G. L. (2023). PLUS: A Semi-automated Pipeline for Fraud Detection in Public Bids. *Digital Government: Research and Practice*, 5(1):1–16. DOI: 10.1145/3616396.

Braz, C. S., Mendes, B. M. A., Oliveira, G. P., Costa, L. L., Silva, M. O., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2023). Análise de irregularidades em licitações públicas com foco em empresas de pequeno porte. In *Proceedings of the 11th Workshop on Computing Applied to E-Government*, pages 94–105, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wcge.2023.230254.

Buryakov, D., Kovács, M., Kryssanov, V. V., and Serdült, U. (2022). Using open government data to facilitate the design of voting advice applications. In *Electronic Participation - 14th IFIP WG 8.5 International Conference*, volume 13392 of *Lecture Notes in Computer Science*, pages 19–34. Springer. DOI: 10.1007/978-3-031-23213-8_2.

Coelho, G. M. C., Ramos, A. C., de Sousa, J., Cavaliere, M., de Lima, M. J., Mangeth, A., Frajhof, I. Z., Cury, C., and Casanova, M. A. (2022). Text classification in the brazilian legal domain. In *Proceedings of the 24th International Conference on Enterprise Information Systems, ICEIS*, pages 355–363. SCITEPRESS. DOI: 10.5220/0011062000003179.

Costa, L. L., Reis, A. P. G., Bacha, C. A., Oliveira, G. P., Silva, M. O., Teixeira, M. C., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2022). Alertas de fraude em licitações: Uma abordagem baseada em redes sociais. In *Proceedings of the 11th Brazilian Workshop on Social Network Analysis and Mining*, pages 37–48, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/brasnam.2022.223175.

Figueiredo, B., Nakamura, F., Felix, G., and Nakamura, E. (2020). Usando análises sociais na identificação de nós relevantes em um cenário multi-redes: Operação licitante fantasma, um estudo de caso. In *Anais do VIII Workshop de Computação Aplicada em Governo Eletrônico*, pages 108–119, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wcge.2020.11262.

Gabardo, A. C. and Lopes, H. S. (2014). Using social network analysis to unveil cartels in public bids. In *European Network Intelligence Conference, ENIC*, pages 17–21. IEEE Computer Society. DOI: 10.1109/ENIC.2014.11.

Galvão Júnior, D., Filho, G. S., and Cabral, L. (2023). Classificação de fraudes em licitações públicas através do agrupamento de empresas em conluios. In *Anais do XI Workshop de Computação Aplicada em Governo Eletrônico*, pages 13–24, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wcge.2023.229519.

Lima, M. C., Silva, R., de Souza Mendes, F. L., de Carvalho, L. R., Araújo, A. P. F., and de Barros Vidal, F. (2020). Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach. In *Findings of the Association for Computational Linguistics: EMNLP*, volume EMNLP 2020 of *Findings of ACL*, pages 1580–1588. ACL. DOI: 10.18653/V1/2020.FINDINGS-EMNLP.143.

Luna, R. and Figueiredo, D. (2022). Caracterização das licitações públicas no estado do rio de janeiro: Diversidade, licitantes Únicos e redes. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 145–156. SBC. DOI: 10.5753/wcge.2022.222675.

Lyra, M. S., Curado, A., Damásio, B., Bação, F., and Pinheiro, F. L. (2021). Characterization of the firm-firm public procurement co-bidding network from the State of Ceará (Brazil) municipalities. *Appl. Netw. Sci.*, 6(1):77. DOI: 10.1007/S41109-021-00418-Y.

Monteiro, M. S., Batista, G. O. S., and Salgado, L. C. C. (2023). Investigating usability pitfalls in Brazilian and Foreign governmental chatbots. *Journal on Interactive Systems*, 14(1):331–340. DOI: 10.5753/jis.2023.3104.

Nai, R., Sulis, E., and Meo, R. (2022). Public procurement fraud detection and artificial intelligence techniques: a literature review. In *Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*, volume 3256 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Oliveira, G. P., Reis, A. P. G., Freitas, F. A. N., Costa, L. L., Silva, M. O., Brum, P. P. V., Oliveira, S. E. L., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2022a). Detecting inconsistencies in public bids: An automated and data-based approach. In *Brazilian Symposium on Multimedia and Web*, pages 182–190. ACM. DOI: 10.1145/3539637.3558230.

Oliveira, G. P., Reis, A. P. G., Mendes, B. M. A., Bacha, C. A., Costa, L. L., Canguçu, G. L., Silva, M. O., Caetano, V., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2022b). Ferramentas open-source de qualidade de dados para licitações públicas: Uma análise comparativa. In *Proceedings of the 37th Brazilian Symposium on Databases*, pages 116–127, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbbd.2022.224351.

Pereira, A. K. S., Vita, Y. M., Felix, G. S., Gimaque, J. M. F., Damasceno, M. L. S., and Figueirêdo, B. C. B. (2022). Usando redes complexas na identificação de empresas fraudulentas em licitações públicas. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 13–24. SBC. DOI: 10.5753/wcge.2022.222704.

Reis, V. Q., Rabello, M. E. R., Lima, A. C., Jardim, G. P. S., Fernandes, E. R., and Brefeld, U. (2023). Data practices in apps from brazil: What do privacy policies inform us about? *Journal on Interactive Systems*, 14(1):1–8. DOI: 10.5753/jis.2023.2954.

Silva, M. O., Paula, A. F., Oliveira, G. P., Vaz, I. A. D., Hott, H., Gomide, L. D., Reis, A. P. G., Mendes, B. M. A., Bacha, C. A., Costa, L. L., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2022). LiPSet: Um conjunto de Dados com Documentos Rotulados de Licitações Públicas. In *Proceedings of the 4th Dataset Showcase Workshop*, pages 13–24. SBC. DOI: 10.5753/dsw.2022.224925.