

REVIEW

Open Access



Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods

Marcos S. Lyra^{1,2}, Bruno Damásio¹, Flávio L. Pinheiro¹ and Fernando Bacao^{1*}

*Correspondence:
bacao@novaims.unl.pt

¹ Information Management School (IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

² Tribunal de Contas Do Estado Do Ceará, Rua Sena Madureira, 1047, Fortaleza 60055-080, Brazil

Abstract

Fraud, corruption, and collusion are the most common types of crime in public procurement processes; they produce significant monetary losses, inefficiency, and misuse of the public treasury. However, empirical research in this area to detect these crimes is still insufficient. This article presents a systematic literature review focusing on the most contemporary data-driven techniques applied to crime detection in public procurement. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) methodology was adopted to identify typical elements that lead to crimes in public contracting. We collected scientific papers and analyzed the selected research using the Scopus repository. We evaluated and summarized findings related to crime detection techniques based mainly on machine learning and network science, as well as studies using fraud risk indices. Some methodologies presented promising results in identifying crimes, especially those using labeled data and machine learning techniques. However, due to the frequent unavailability of pre-labeled data on past cases, analysis through network science tools has become more evident and relevant in exploratory research.

Keywords: Corruption, Fraud detection, Public sector, Procurement, Network, Machine learning, SNA, PRISMA

Introduction

Public procurement is the process by which the public sector purchases services, goods, and construction works from third-party companies. It represents a substantial share of the government's public expenditure (OECD 2021), amounting to 12% of global GDP (Bosio et al. 2022). According to the Organization for Economic Cooperation and Development (OECD), in 2017, public procurement represented 29.1% of the general government expenditures among OECD countries.

The reliability of these processes is key for an effective act of governance (Silva Filho 2017). However, while governments seek the best cost–benefit regarding

the price and quality of their procured services (Costa et al. 2020), public procurement is also well known to be quite permeable to irregularities.

Indeed, Transparency International Organization estimates that losses in public contracts due to corruption account for 20% to 25% of the value of the contract and up to 50% in some cases (Volosin 2015).

Among public administration, public procurement is the process that contains the highest risks of illegality that can arise virtually at any stage of the process of acquiring goods, services, or construction works (Rustiarini et al. 2019). Crimes in this context are elaborated and refined, involving schemes such as abuse of power, corruption, peculate, bribery, favoritism, stunt, nepotism, and misappropriation (Padhi & Mohapatra 2011). In that sense, these crimes are difficult to detect and measure (Mufutau & Mojisola 2016), particularly after awarded procurement contracts (Whiteman 2019). As such, well-defined legislation is not enough to circumvent criminality in public procurement, and mechanisms to prevent it are frequently inadequate (Wensink & Vet 2013).

In that sense, it is crucial to understand the phenomena and establish which methods and practices have been explored by the academic community to help identify actors and objects linked to fraudulent activities in the public procurement process. In particular, we focus on actors who act through the practice of Corruption (the relational activity between corrupting agents and corrupted entities, always with the participation of the public body) (Paganetto & Scandizzo 2015), Collusion (an agreement between third parties to defraud operations of the public administration, without its participation) (Mvelase 2015), and Fraud (crimes of forgery of documents, fraudulent proposals, use of shell companies, the introduction of false assessment of competitor, identity fraud, and data theft) (Mlondo 2013) to impair the public administration from acting efficiently.

Here, we use the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to identify and review the use of data-driven methods in academic literature to study fraud, collusion, and corruption in public procurement processes. We look to identify the problems and contexts in which mainly network science and machine learning methods have been applied. Overall, we identify in our work 48 relevant manuscripts and discuss the methods used, the problems they explored, and their framing in the academic literature.

Recent works have leveraged the increasing availability of public data on public procurement, which resulted from adopting electronic procurement (e-procurement) bidding processes/modalities. These have been introduced to increase transparency and competitiveness and decrease the bureaucracy associated with the procurement processes (Borras 2015) (Weigel 2017). E-procurement is, thus, a modality that offers a speed-up process at reduced costs (Motta 2003). Although e-procurement has been around for a long time and has been widely adopted, there is limited research on its impact on reducing/mitigating crime in the public sector (Kartika, 2020). Therefore, we extend our analysis to procurement activity in the private sector. This dimension has seen more research and explored more advanced techniques to detect and mitigate crimes in the procurement process (Wu Chebili, 2022). It is the case, in our opinion, that the public sector can learn from the best practices of the private sector, in particular, given the similarities in the procurement processes in both cases.

This manuscript is structured as follows: “**Methodology**” section announces the method used for identifying relevant research and conducting a bibliometric analysis. In “**Results**” section, we raise the results, introducing the keywords, citations, occurrences,

methods, and research objectives. In “[Discussion](#)” section, we discuss the outcomes and show the research limitations. Finally, “[Conclusion and further studies](#)” section concludes the research, identifies some research gaps, and gives directions for further work.

Methodology

We perform a systematic literature review using the PRISMA methodology. PRISMA is an auditable methodology based on a checklist of items often depicted in a flow diagram. The checklist includes 27 topics distributed throughout seven sections: title, abstract, introduction, methods, results, discussion, and other information. PRISMA is used to enhance the meta-analysis of systematic review reports and promote their reproducibility (Fig. 1).

The PRISMA 2020 template for systematic reviews can be summarized by a three-step flowchart that includes the identification, screening, and selection of manuscripts.

Here, we proceed with the PRISMA methodology to answer the following question: *“How have data-driven methods contributed to detecting, characterizing, and predicting fraud, collusion, and corruption in public procurement activities?”*

We begin by identifying relevant scientific manuscripts. In that sense, we followed a systematic process of iterative query in the research repository Scopus to identify publications with the scope defined by the search string. The Scopus repository was chosen due to its comprehensive coverage of scientific venues, more immediate indexing methods, and the availability of more articles than other repositories.

Keyword tracking strategy and bibliometric analysis

Bibliometric analysis is part of an interdisciplinary science concerning a quantitative study. It uses mathematical and statistical methods to capture specific knowledge and expertise from various authors (Merigó, 2016). It is a methodology usually used to identify the development of a particular scientific field (Merigó et al. 2015; Železnik et al. 2017). Based on this, we conducted an iterative tracking process with an assembled search string, shown below, encompassing words and phrases logically connected by Boolean operators in the titles, abstracts, and keywords. We chose the words aiming to answer the research question, serve the research objectives, and whose frequency, in most various articles, on the topic was high in the abstracts, keywords, and texts.

Search String: (fraud* OR corrupt* OR bribe OR rigg* OR cartel* OR collusi*) AND (predict* OR detect* OR SNA OR network* OR statistic*) AND (“public sector “ OR “public administration “ OR procurement OR bidd* OR auction OR tender OR government).

To sort, extract, organize and store metadata (publications, citations, authors, and references), we used the Software Mendeley, which provides a free reference manager that assists in academic work, manages electronic files (PDF format), helps in the normalization of citations, and generates references automatically. To assemble the network that describes the relations among authors and topics, we used the open-source software VOSviewer (Van_Eck & Waltman, 2014). It is an engine for constructing and visualizing bibliometric networks, developed by Nees Jan van Eck and Ludo Waltman at the Center for Science and Technology Studies at Leiden University. These networks are built based on co-citation, bibliographic coupling, or co-authorship ratios. VOSviewer

also offers text mining capabilities that allow building and viewing simultaneous occurrence networks of important terms extracted from the scientific literature. Furthermore, we investigate the citation's framework to capture the conjectural basis of the research (Ding & Yang 2020).

Using measures like degree centrality (displays how many times writings are cited, illustrated by the nodes' size) and connection weight (symbolizes how many times two papers are mentioned jointly in different articles—characterized by the thickness of the link), we applied VOSviewer to create graphs founded on data gathered from the Scopus repository and analyzed them employing network Science methods. Furthermore, we examined the data gathered from the simultaneous occurrence of keywords and authors, which indicates that the greater the node size, the more frequently the keyword emerges in writings. On the other hand, the thicker the connections between nodes, the grander the number of times these keywords appear in diverse papers. The same idea was established to article citations, i.e., the bigger the node size, the more cited the issued articles in the specified database, and the wider the linkage between them, the more frequently they operate together in publications.

Results

The data collection began in April 2021 and was later updated in December 2021 and January 2022. At first, we applied a broader search string without setting the specific area of interest included in the last "AND" to capture the dimension of the study being researched. As a result, we identified 39,445 documents. Thus, we applied the entire search string assembled, obtaining 2,914 articles, representing 7.4% of the scientific production in the sector. In Fig. 2, we see the distribution of these articles by annual publication and annual citation amounts presented on a 2-year moving average. In Fig. 2a, we noticed that one of the first articles presented by Scopus in the sector dates back to 1991, with a more significant and constant increase from the year 2000 on. From 2016 onwards, we see a steeper curve slope, indicating a greater interest in the subject over time. On the other hand, in Fig. 2b, we note that the peak of citations occurs in articles from the 2007/2008 biennium since those articles are available for longer for research and, therefore, are better able to receive more citations.

Among the 2914 studies identified, Table 1 presents the ten institutions that most contributed to the research using the specified string. Table 2 shows how these articles were distributed in the ten countries that published the most.

Henceforward, we followed with the extraction of relevant publications through the PRISMA stages. We took all available publications with the author's identification, discarding 98 without a definition of authorship and 712 publications prior to 2011 to avoid picking up outdated articles, so we entered the screening phase with 2104 documents.

During this phase, we carried out suppressions based on the three sets of words: network*, some "public expressions" (public administration, public sector, tender*, auction, bid*, procurement), which were not presented in the title, abstract or keyword of the articles. Then we excluded 1969 publications and maintained 135 of them. Next, we analyzed the abstracts of the 135 documents and abandoned 81 that were beyond the scope, keeping 54. After making a complete reading of the remaining articles, we excluded six

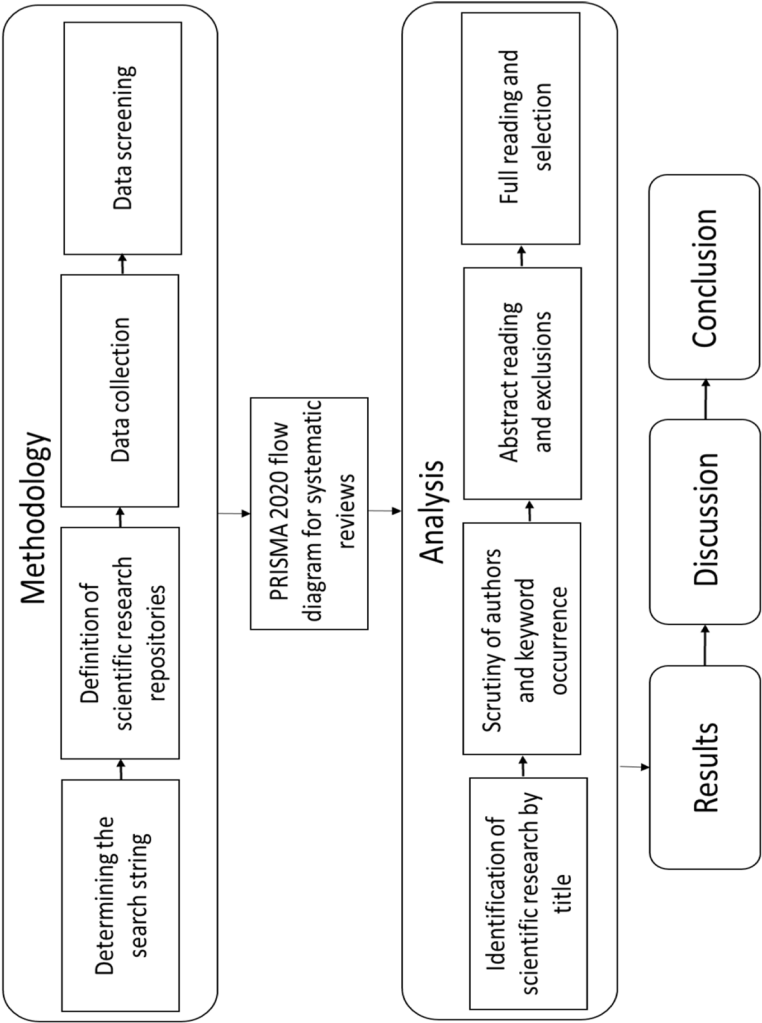


Fig. 1 Research workflow

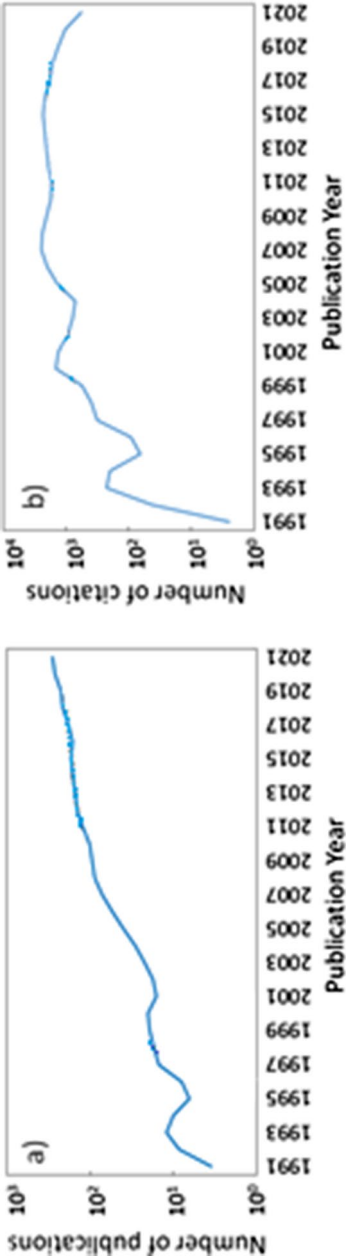


Fig. 2 Articles regarding corruption detection in the public sector – 2-year moving average Log_{10} **a** Evolution of the number of publications. **b** Evolution of the number of citations

Table 1 Main institutions

Institution	Publicat	Citations
The Hong Kong Polytechnic University	5	68
University of Maryland	4	255
London School of Economics and Political Science	4	133
Illinois State University	4	87
University of Massachusetts Dartmouth	4	60
Australian National University	4	53
Universiti Teknologi Mara—Malaysia	4	32
Princeton University	4	32
Islamic University of Indonesia	4	31
University of Manchester	4	21

Table 2 Main countries

Country	Publicat	Citations
USA	732	19,015
UK	270	4741
China	193	1310
India	160	976
Australia	136	1900
Canada	97	1273
Germany	82	996
Brazil	74	518
Italy	72	900
Nigeria	71	135

more publications, keeping 48 papers (Fig. 3). Of the 48 articles, 36 were issued by scientific journals, while 13 were presented in conference procedures.

Scientific articles published in journals and conferences

The 48 selected publications are distributed over a decade in quantities ranging from two publications in 2011 to a maximum of 10 in 2021. It indicates that the topic has been gaining notoriety and being more investigated over time. The documents gathered from the Scopus from diversified journals and conference annals encompass a vast variety of scientific work, from business & economics to computer science, passing by law & government, and science administration. The survey of scientific writings is fundamental for identifying pertinent material for our area of investigation. We selected papers mentioned up to 47 times, deriving from the list of publications in Table 3.

The material acquired from the repositories represents the most recent and updated studies currently used to detect fraud in the public sector. The more cited articles and authors are shown in Table 3. We verify that article no. 1 has the largest volume with 47 citations, followed by article no. 2 with 41, and article no. 3 with 35 citations, all related to the Scopus platform. The citations were examined in depth in the Author and Co-Authorship Citation Occurrences section, guided individually.

Table 3 Articles selected from the Scopus repository (ordered by number of citations)

No	Title/Author/Year	Citation	Journal/Conference
1	An Objective Corruption Risk Index Using Public Procurement Data (Fazekas et al. 2016)	47	European Journal on Criminal Policy and Research
2	From Corruption to State Capture: A New Analytical Framework with Empirical Applications from Hungary (Fazekas & Tóth 2016)	41	Political Research Quarterly
3	Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data (Fazekas & Kocsis 2017)	35	British Journal of Political Science
4	Internet auction fraud detection using social network analysis and classification tree approaches (Chiu et al. 2011)	34	International Journal of Electronic Commerce
5	Bid-rigging networks and state-corporate crime in the construction industry (Reeves-Latour & Morselli 2017)	32	Social Networks
6	The impact of inter-organizational relationships on contractors' success in winning public procurement projects: The case of the construction industry in the Veneto region (Sedita & Apa 2015)	28	International Journal of Project Management
7	Combining ranking concept and social network analysis to detect collusive groups in online auctions (Lin et al. 2012)	23	Expert Systems With Applications
8	Finding the needle: A risk-based ranking of product listings at online auction sites for non-delivery fraud prediction (Almendra 2013)	19	Expert Systems With Applications
9	Corruption and complexity: a scientific framework for the analysis of corruption networks (Luna-Pla & Carlock 2020)	18	Applied Network Science
10	Fuzzy rule optimization for online auction frauds detection based on genetic algorithm (Yu & Lin 2013)	15	Electronic Commerce Research
11	A network approach to cartel detection in public auction markets (Wachs & Kertesz 2019)	14	Scientific Reports
12	Social capital predicts corruption risk in towns (Wachs et al. 2019)	13	Royal Society Open Science
13	Leveraging social networks to combat collusion in reputation systems for peer-to-peer networks (Li et al. 2013)	13	IEEE Transactions on Computers
14	Distinguishing Characteristics of Corruption Risks in Iranian Construction Projects: A Weighted Correlation Network Analysis (Hosseini et al. 2019)	12	Science and Engineering Ethics
15	Network similarity and collusion (Morselli & Ouellet 2018)	12	Social Networks
16	Machine learning with screens for detecting bid-rigging cartels (Huber & Imhof 2019)	9	International Journal of Industrial Organization
17	Online Detection of Shill Bidding Fraud Based on Machine Learning Techniques (Ganguly & Sadaoui 2018)	8	31st International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE)
18	A decision support system for fraud detection in public procurement (Velasco et al. 2021)	7	International Transactions in Operational Research
19	Prediction of public procurement corruption indices using machine learning methods (Rabuzin & Modrusan 2019)	7	11th International Conference on Knowledge Management and Information Systems (KMIS)
20	Comparison of ANN Classifier to the Neuro-Fuzzy System for Collusion Detection in the Tender Procedures of Road Construction Sector (Anysz et al. 2018)	7	3rd World Multidisciplinary Civil Engineering, Architecture, Urban Planning Symposium (WMCAUS)

Table 3 (continued)

No	Title/Author/Year	Citation	Journal/Conference
21	Using Self-Organizing Maps for fraud prediction at online auction sites (Almendra & Enachescu 2014)	7	15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)
22	Detection of feedback reputation fraud in Taobao using social network theory (Zhu et al. 2011)	7	Proceedings—2011 International Joint Conference on Service Sciences, IJCSS 2011
23	Corruption risk in contracting markets: a network science perspective (Wachs, Fazekas, & Kertész, 2021)	6	International Journal of Data Science and Analytics
24	Corruption and the network structure of public contracting markets across government change (Fazekas & Wachs 2020)	5	Politics and Governance
25	Mapping Corruption Risks in Public Procurement: Uncovering Improvement Opportunities and Strengthening Controls (Sharma et al. 2019)	5	Public Performance & Management Review
26	Big data system for analyzing risky procurement entities (Dhurandhar et Al. 2015)	5	Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
27	Online hybrid model for online fraud prevention and detection (Mundra & Rakesh 2014)	5	Advances in Intelligent Systems and Computing
28	Uncovering the structure of public procurement transactions (Popa 2019)	4	Business and Politics
29	Preventing rather than punishing: An early warning model of malfeasance in public procurement (Gallego et al. 2021)	3	International Journal of Forecasting
30	Clustering and Labeling Auction Fraud Data (Alzahrani & Sadaoui 2020)	3	International Conference on Data Management, Analytics and Innovation (ICDMAI)
31	Trust model based on Islamic business ethics and social network analysis (Lei et al. 2018a, b)	3	International Journal on Advanced Science, Engineering and Information Technology
32	How do strategic networks influence awarding contract? Evidence from French public procurement (Mamavi et al. 2017)	3	International Journal of Public Sector Management
33	Detecting fraudsters in online auction using variations of neighbor diversity (Khomnotai & Lin 2015)	3	International Journal of Engineering and Technology Innovation
34	Incremental collusive fraud detection in large-scale online auction networks (Dadfarnia et al. 2020)	2	Journal of Supercomputing
35	Bidder Network Community Division and Collusion Suspicion Analysis in Chinese Construction Projects (Zhu et al. 2020)	2	Advances in Civil Engineering
36	Detecting the collusive bidding behavior in below average bid auction (Lei et al. 2018a, b)	2	13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)
37	Analysis on bidding behaviours for detecting shill bidders in online auctions (Majadi et al. 2017)	2	Proceedings—2016 16th IEEE International Conference on Computer and Information Technology, CIT 2016, 2016 6th International
38	Network Analysis for Fraud Detection in Portuguese Public Procurement (Carneiro et al. 2020)	1	International Conference on Intelligent Data Engineering and Automated Learning
39	Improving fraudster detection in online auctions by using neighbor-driven attributes (Lin & Khomnotai 2016)	1	Entropy
40	Characterization of the firm–firm public procurement co-bidding network from the State of Ceará (Brazil) municipalities (Lyra et al. 2021a, b,2)	0	Applied Network Science

Table 3 (continued)

No	Title/Author/Year	Citation	Journal/Conference
41	Tackling corruption in urban infrastructure procurement: Dynamic evaluation of the critical constructs and the anti-corruption measures (Owusu et al. 2021)	0	Cities
42	A social network-based examination on bid riggers' relationships in the construction industry: A case study of China (Xiao et al. 2021)	0	Buildings
43	Conspiracy of Corporate Networks in Corruption Scandals (Nicolás-Carlock & Luna-Pla 2021)	0	Frontiers in Physics
44	The extra-legal governance of corruption: Tracing the organization of corruption in public procurement (Fazekas et al. 2021)	0	Governance
45	Networked Corruption Risks in European Defense Procurement (Czibik et al. 2021)	0	Understanding Complex Systems
46	Corruption in the Implementation of Public Procurement from Small and Medium Businesses (Perevezentceva, et al. 2021)	0	Lecture Notes in Networks and Systems
47	Real-Time Shill Bidding Fraud Detection Empowered with Fused Machine Learning (Abidi et al. 2021)	0	IEEE Access
48	Adaptation of cluster analysis methods in respect to vector space of social network analysis indicators for revealing suspicious government contracts (Davydenko et al. 2017)	0	IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)

Publishing journals and conferences

Regarding the publication of articles in Journals, 36 studies were distributed in 32 distinct Journals and published by 26 editors. Based on Scimago Journal & Country Rank, 21 Journals (65%) are classified in the first quartile Q1, five journals (16%) in the second quartile Q2, four journals (13%) in quartile Q3, and two journals (6%) are classified in quartile Q4. The leading areas of action of the 32 journals are computer science, business, Social Science, and some multidisciplinary areas, see Table 4.

The selected papers were issued by publishers belonging to eleven nations. The most recurring were Netherlands, Switzerland, and the USA, publishing six works each. In turn, the most commonly used editors were Springer and Springer Netherlands.

Conference proceedings

The conferences were responsible for editing 12 documents published by seven editors belonging to five countries of the selected articles. The main field of investigation once again was computer science, Table 5.

Of the 12 conference proceedings, 11 were classified as computer science. The publishers were also varied; the IEEE published papers, particularly in Conferences on Fuzzy Systems, Artificial Intelligence, Internet of Things, Information Systems and Technologies, and Internet Computing. Springer Nature issued articles on Intelligent Systems and Technologies, CISTI. The other five editors published articles on Intelligent Data Engineering, Data Management, Multidisciplinary Civil Engineering, and Data Mining.

Table 4 Published journals

Journal	Qt	Rank	Publisher	Country	Field
Social Networks	2	Q1	Elsevier	Netherlands	Anthropology; Sociology
Applied Network Science	2	Q1	Springer Open	Switzerland	Computer Science
Expert Systems With Applications	2	Q1	Elsevier Science LTD	England	Computer Science
Lecture Notes in Networks and Systems	1	Q4	Springer International Publishing AG	Switzerland	Computer Science
British Journal of Political Science	1	Q1	Cambridge University Press	UK	Social Science
Business and Politics	1	Q1	Cambridge University Press	UK	Social Science
Cities	1	Q1	Elsevier Ltd	UK	Social Science
Electronic Commerce Research	1	Q1	Springer	Netherlands	Economics
European Journal on Criminal Policy and Research	1	Q1	Springer Netherlands	Netherlands	Social Sciences
Governance	1	Q1	Wiley-Blackwell Publishing Ltd	UK	Social Sciences
IEEE Access	1	Q1	Engineers Inc	USA	Computer Science
IEEE Transactions on Computers	1	Q1	IEEE Computer Society	USA	Computer Science
International Journal of Electronic Commerce	1	Q1	M.E. Sharpe Inc	USA	Business, Management & Accounting
International Journal of Forecasting	1	Q1	Elsevier	Netherlands	Business, Management & Accounting
International Journal of Industrial Organization	1	Q1	Elsevier Inc	USA	Business, Management & Accounting
International Journal of Project Management	1	Q1	Elsevier Science LTD	England	Business & Economics
International Transactions in Operational Research	1	Q1	Wiley	USA	Business & Economics
Policing & Society	1	Q1	Routledge Journals, Taylor & Francis LTD	England	Criminology & Penology
Political Research Quarterly	1	Q1	Sage Publications Inc	USA	Government & Law
Public Performance & Management Review	1	Q1	SAGE Publications Inc	USA	Business, Management & Accounting
Royal Society Open Science	1	Q1	Royal Soc	England	Science & Technology
Science and Engineering Ethics	1	Q1	Springer	Netherlands	Multidisciplinary Sciences
Scientific Reports	1	Q1	Nature Publishing Group	England	Science & Technology
International Journal of Public Sector Management	1	Q2	Emerald Group Publishing LTD	England	Business & Economics; Public Administration
Politics and Governance	1	Q2	Cogitatio Press	Portugal	Government & Law
Buildings	1	Q2	MDPI Multidisciplinary Digital Publishing Institute	Switzerland	Building and Construction
Entropy	1	Q2	MDPI	Switzerland	Physics, Multidisciplinary
Frontiers in Physics	1	Q2	Frontiers Media S.A	Switzerland	Materials Science
International Journal of Data Science and Analytics	1	Q2	Springer	Switzerland	Computer Science
Advances in Civil Engineering	1	Q3	Hindawi LTD	England	Construction & Building Technology
Advances in Intelligent Systems and Computing	1	Q3	Springer Verlag	Germany	Computer Science
International Journal on Advanced Science, Engineering and Information Technology	1	Q3	-	Indonesia	Agricultural and Biological Sciences

Table 4 (continued)

Journal	Qt	Rank	Publisher	Country	Field
Journal of Supercomputing	1	Q3	Springer Netherlands	Netherlands	Computer Science
International Journal of Engineering and Technology Innovation	1	Q4	Taiwan Association of Engineering and Technology Innovation	Taiwan	Civil and Structural Engineering
Understanding Complex Systems	1	Q4	Springer Verlag	Germany	Computer Science

Table 5 Conference Annals

Conference	Qt	Publisher	Country	Field
International Conference on Intelligent Data Engineering and Automated Learning	1	Springer International Publishing AG	Switzerland	Computer Science
International Conference on Data Management, Analytics and Innovation (ICDMAI)	1	Springer Verlag Singapore PTE LTD	Singapore	Computer Science
11th International Conference on Knowledge Management and Information Systems (KMIS)	1	SCITEPRESS	Portugal	Computer Science
13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)	1	IEEE	USA	Computer Science
31st International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE)	1	Springer Nature Switzerland AG	Switzerland	Computer Science
3rd World Multidisciplinary Civil Engineering, Architecture, Urban Planning Symposium (WMCAUS)	1	IOP Publishing LTD	England	Architecture; Engineering; Urban Studies
Proceedings—2016 16th IEEE International Conference on Computer and Information Technology, CIT 2016, 2016 6th International	1	IEEE	USA	Computer Science
12th Iberian Conference on Information Systems and Technologies (CISTI)	1	IEEE	USA	Computer Science
Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	1	ASSOC COMPUTING MACHINERY	USA	Computer Science
15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)	1	IEEE	USA	Computer Science; Engineering
Proceedings—2011 International Joint Conference on Service Sciences, IJCSS 2011	1	IEEE	USA	Computer Science
IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)	1	IEEE	USA	Computer Science; Telecommunications

Articles classification

To classify the articles, we took the specific themes addressed in each one. It is worth remembering that the public procurement articles referred to various types of contracts; however, not all papers offered a more comprehensive detailing of the contracted object. Furthermore, E-procurement* was presented as a resource for the Public E-procurement surveys. They are broader research and have more outstanding fraud detection capabilities, which may assist the public sector with solutions.

Table 6 shows that general public procurement dominates the researched occurrences due to the greater availability of open data in this segment, followed by the construction industry, which represents a large part of the expenditure associated with procurement contracts. On the other hand, the three types of crime studied perceived the same magnitude order in terms of monetary values. The most common crimes identified were corruption in public procurement and collusion in construction. Fraud appeared more in online auctions or e-procurement; however, this topic was better explored in the private sector.

Keyword occurrences

To examine in detail sensitive themes, we performed a co-occurrence assessment by employing keywords occurrence based on the count of the individual or combined words in portions of scientific research (title, keyword, and abstract). Besides, the investigation is extended to various documents concomitantly, helping researchers review and exploit the writings statistically (Horvat, 2015).

The analysis of keywords occurrence has been performed employing VOSviewer to assemble and view bibliometric networks. The examination was carried out through the complete counting method, containing ten keywords, in which the lowest occurrence boundary was fixed as six. The ten keywords classified in descending order of occurrence are displayed in Table 7.

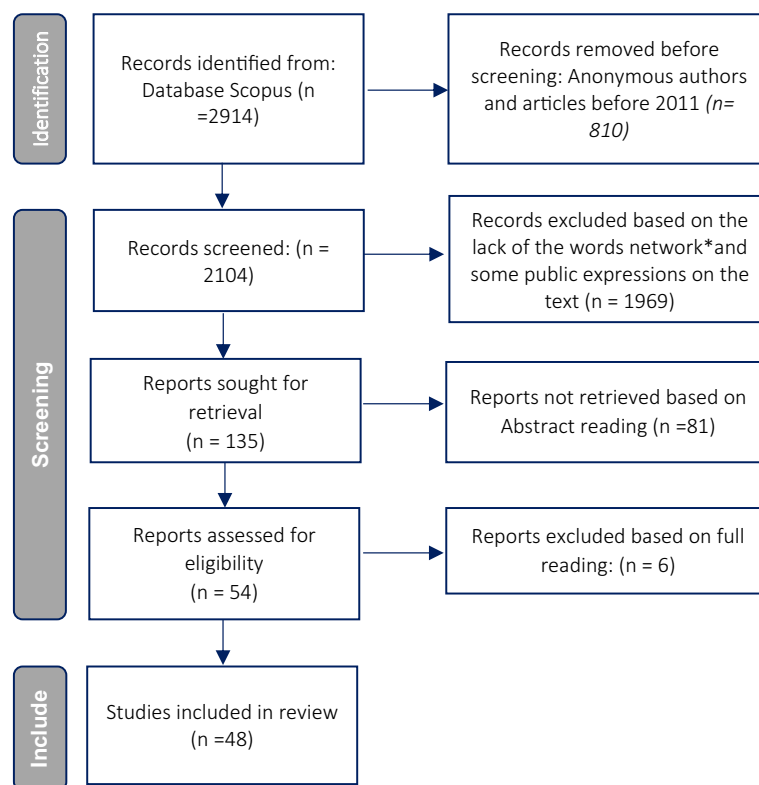
We can also see the average year of keyword usage in the selected documents. We perceive that the most relevant keywords, crime, public procurement, and corruption, have their average launch years between 2017 and 2019, and social network analysis (SNA) and social networking more between 2013 and 2014 when this expression was still widely used for network science tools. (Table 7).

Table 6 Article classification

Article Classification	Number of articles	Corruption cases	Fraud cases	Collusion cases
General Public Procurement	18	17	3	5
Construction industry	9	5	0	8
Public online auction	1	1	1	0
Shell companies contracting	1	1	0	0
Bidding markets	1	0	0	1
Defense	1	1	0	0
*Private online auction	17	0	16	4
TOTAL	48	25	20	18

Table 7 Major keywords occurrence

Keyword	Link strength	Occurrences	Avg. pub. Year
crime	276	19	2017
public procurement	146	15	2018
corruption	81	14	2019
fraud detection	144	11	2017
social network analysis	63	8	2013
construction industry	82	7	2018
social networking (online)	93	6	2014
learning systems	92	6	2018
data mining	76	6	2017
machine learning	71	6	2017

**Fig. 3** PRISMA 2020 flowchart showing three phases, where n denotes the number of publications retrieved in each level

The identified keywords appearing the most had the following link strength and occurrence, respectively: crime (276, 19), public procurement (146, 15), corruption (81, 14), fraud detection (144, 11), and social network analysis. In Fig. 4, we see the co-occurrence connections between keywords in various scientific papers collected. Here we perceive the formation of two communities in the keyword co-occurrence network with ten nodes and 30 links.

Keyword nodes are linked by their concurrent existence in separate documents, where keyword communities are formed based on the similarity between their connections. The node size denotes the occurrence of keywords; the bigger the node, the greater the

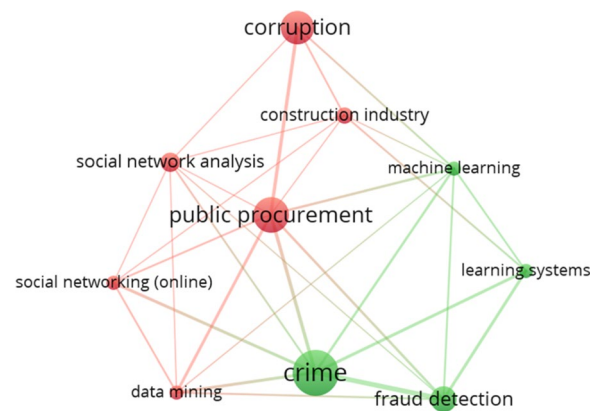


Fig. 4 Keyword occurrence network—community visualization

keyword occurrence. Figure 4 shows the previously demonstrated statistics, where we perceive that the nodes representing public procurement, corruption, and crime keywords are the largest, meaning that they are the ones that appear the most. We named the two communities based on their most representative and centralized nodes, Public procurement (red) and Crime (green).

The network emphasizes "public procurement" as the most studied area. On the other hand, the keyword "corruption", "social network analysis", and "data mining" appears in the same community as the problem being investigated and the techniques used for its identification. In turn, the community Crime presents machine learning and learning systems as techniques for fraud detection.

Author and Co-authorship citation occurrences

Once again, we used the VOSviewer tool to investigate the author's and co-author's citation occurrence and the link strength. We employed the full counting method and established the minimum link strength as six. The documentation brought in 11 authors presented in Table 8 by link strength order.

The authors and co-authors that stand out the most for their Link strength and citations are Fazekas M. with seven articles and 134 citations and Wachs J. with five documents and 38 citations. We observed that these two authors are very interconnected. Of the 12 articles they wrote, three they did together; Fazekas was the author in two and Wachs in one. The average publication period of the articles from the 11 authors is very recent, varying from 2019 to 2021.

As it is an emerging and incipient community, the network representation assembled by VOSviewer is still very disconnected and dispersed, but it is a community with high potential for future collaboration and should coordinate efforts to systematize techniques and approaches.

Figure 5 presents the giant component of the citation network. Here we see the formation of 5 clusters with 21 nodes (articles) and 41 links, where Fazekas and Wachs appear in two communities. Fazekas has the most articles written, mainly in the blue community, and Wachs is the central link between the different communities and probably the most influential author in this subnet.

Table 8 noteworthy citations of authors and co-authors

Author	Strength	Documents	Citations	Avg. pub. Year
Fazekas, Mihály	12	7	134	2019
Wachs, Johannes	10	5	38	2020
Chan, Albert.P.C	7	2	12	2020
Kertész, János	6	3	33	2019
Abidi, Wajhe Ul Husnian	6	1	0	2021
Alzahrani, Ahmad M	6	1	0	2021
Alyas, Tahir	6	1	0	2021
Daoud, Mohammad S	6	1	0	2021
Fatima, Areej	6	1	0	2021
Ihnaini, Baha	6	1	0	2021
Khan, Muhammad.Adnan	6	1	0	2021

Methods, connectivity, and investigation type of the articles

In the list of scientific articles collected, we identified two most popular types of techniques, together or separately, to detect crimes against public administration: machine learning and network science. Furthermore, we verified the connectivity of the data presented in the papers, classifying them as connected networks if they had these characteristics. The status of connected networks was determined by a route between any random pair of nodes, even if intermediate nodes are needed to establish the course (Dong, 2020). The high density of a network considerably improves the potential for collaborative action and cooperation between actors (Olsson et al. 2004), as well-connected networks foster communication, promote common confidence, and help contain or control conflicts (Bodin and Norberg 2005). Furthermore, the high density of links promotes joint action, particularly when there are many connections between different agents (Sandström and Carlsson 2008). In our study, 34 or 71% of articles were related to connected networks. It allowed researchers to explore the relationships between agents and uncover hidden connections.

Finally, we examined the type of investigation carried out, if exploratory or using labeled data. Labeled data research was used to confirm predictions and train models based on patterns of the information used. If the data are representative of the population, these will confirm or not the results of a finding, or a machine learning model will make predictions when using new data (Xiao et al. 2018). On the other hand, exploratory data analysis sought to research little-known problems with no confirmatory data and investigated information to develop hypotheses (Robson 2002). It aimed to develop knowledge that is still new (Kuncoro 2013).

In our research, we have 52% reports using labeled data and 48% exploratory research. It allowed researchers to explore the data and confirm the results when using labeled data, and in the absence of confirmatory data, the exploratory study could draw significant conclusions. Notably, 85% of machine learning research used pre-labeled data in their analyses, while network science used it in 44% of its research. The information is presented in more detail for each article in Appendix 1.

Objectives of the works

This section presents the primary purposes of each study grouped into higher-order objectives, Table 9.

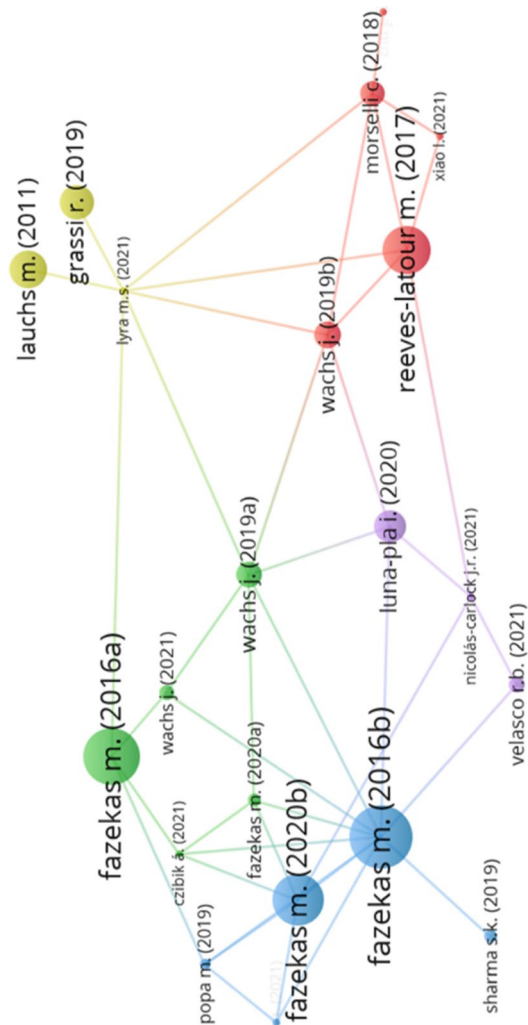


Fig. 5 Citation network – giant component, community visualization

We divided the articles into ten macro-objective groups to better understand the proposals presented. A significant part of the articles falls under the detection of fraud, corruption, and collusion in public procurement, with 11 articles. Ranking second to detect identical crimes, we have the construction industry with eight articles. We also present two macro objectives studied from the private* segment referring to online auctions, which are widely used in the public domain but little explored in scientific research to detect crimes in this segment. To have a more specific glimpse of the objectives of each scientific article, a synopsis of the purposes of the 48 topics is presented in Appendix 2.

Discussion

The present study uses the PRISMA methodology and network analysis to spot the most recent research and techniques for crime identification in public procurement processes. This section introduces a straightforward assessment comparing the outcome of the Scopus repository. It analyzes how we answer the research question and addresses the limitations of the research as follows:

RQ: How have data-driven methods contributed to detecting, characterizing, and predicting fraud and corruption in public procurement activities?

Most commonly used techniques using labeled data

Concerning the main techniques applied, 85% of the approaches used in machine learning papers employed labeled data alone or in conjunction with another method for detecting crimes in public contracts. Different techniques were applied: Artificial Neural Network—ANN (Articles no. 16, 20, 47); Support Vector (Articles no. 17, 19, 34, 47); Random Forest—RF (Articles no. 16, 28, 44); Logistic Regression (Articles no. 19, 44); Lasso Regression (Articles no. 16, 29); Gradient Boosting Classification Model (Article no. 29); Naive Bayes (Article no. 19); Boosted trees (Article no. 8); binary classification (Article no. 21).

Random Forest was employed to find bid-rigging cartels, Identify close connections between government and private firms in repeated interaction and geographic dispersion, and detect single bidding and government spending concentration. It reached an adequate performance in finding competitive tenders and classifying collusive ones.

Table 9 Central purpose of the articles

Research objective	Articles number
Identify the risks of fraud, corruption, and collusion in public procurement	1, 2, 3, 19, 23, 24, 25, 26, 28, 29, 32, 38, 41, 43, 44, 46
Detect the risks of fraud, corruption, and collusion in the construction industry	5, 6, 14, 15, 16, 20, 35, 42
Uncover scams in co-bidding relationships	11, 40
Recognize fraud in product listings, and fraudulent dealers on online public auctions	36, 48
Identify potential crimes in suspected shell companies	9, 18
Detect the risks of corruption and collusion in public defense contracts	45
Recognize collusion in the public administration	12
*Identify fraud in product listings, fraudulent dealers, or bidder reputation on private online auctions	4, 7, 8, 10, 13, 21, 22, 17, 31, 33, 34, 39
*Detect shill bidding fraud in private online auctions	17, 30, 37, 47

It performed well in both collusive and non-collusive periods, appropriate for detecting bid-rigging cartels. RF also identified the emergence of geographically close and repeated connections between private entities and public authorities, suggesting corrupt or socially undesirable behavior and favoring illicit schemes. When used to detect single bidders or government spending concentration, it showed encouraging results with internal data but only moderate when externally validated. Nonetheless, it achieved higher explanatory power than Logistic Regression due to a more flexible parameterization.

Logistic Regression helped find suspicious tenders with text-mining techniques and detect single bidding and government spending concentration. Applied jointly with text mining to recognize contract corruption, it showed intermediate precision, recall, and Receiver Operating Characteristic (ROC) results. It had better results than SVM but lower than Naive Bayes. Alike RF, Logistic Regression for detecting single bidders and concentration of government spending showed significant results when using internal data; however, only reasonable outcomes when validated with external information. On the other hand, the Logistic Regression obtained less explanatory strength than the RF due to a less flexible parameterization.

The Lasso Regression was intended to detect bid-rigging cartels and predict inefficiency and corruption in public procurement. The lasso performed slightly better in collusive than non-collusive periods, but the results were quite satisfactory in detecting bid-rigging. When used to predict inefficiency and corruption in tenders, it was outperformed by Gradient Boosting Machine in all scenarios.

Regarding network science techniques using labeled data, the most employed were cluster analysis (community detection/clustering) used by eight studies (Articles No: 24, 31, 35, 36, 39, 40, 43, 48), where clustering was formed when the similarity between nodes is high, and nodes are separated into different communities when similarity is below the threshold defined by the algorithm. Centrality measures were used by four studies (Articles No: 33, 38, 42, 48) which usually reflect the prominence, status, prestige, or visibility of a node and often explain behaviors on the network (Marsden 2005). One (Articles No: 40) applied bipartite network projection to generate a co-participation network and capture the hidden relationships between players.

They examined the endurance of fraudsters in a weak competitive community over the years after a government turnover to identify corrupt connections between agents. The centrality measures were used to help identify the most central, active, influential players and those presumed of fraudulent connections. The degree centrality, for example, excels in revealing the most linked players of the net. Nevertheless, while it is able to assist in indicating network hubs, criminal interfaces are not always intermediated by them (Morselli 2008); therefore, other measures and analyses were necessary to identify suspicious connections. On the other hand, regardless of employing closeness centrality, some criminals were conscious of the existence of risk indicators and were persevering to bypass them by decreasing the mean closeness values through data tampering. An analysis of clusters demonstrated that the most effective indicators to expose suspect features were based on contract expenditures, contract title, closeness weights, and eigenvector mean values, demonstrating, via these measures, that a considerable portion of government deals held infractions.

Some other methods that deserve further comments used combinations of models. The labeled data based on hierarchical Clustering Using REpresentatives (CURE) is a clustering model that identifies bidding strategies. This technique delivers an adequate number of clusters, and k-mean clustering supplies elements concerning the players' methods in each cluster. Furthermore, a combination of approaches via Machine Learning techniques (hybrid Decision Support System, decision trees, Support Vector Machine, and Neural Networks) and network science algorithms (PageRank, density, and centralities) was used to reveal collusion and corruption associated with defrauding conduct on networks, and together with the concept of neighbor diversity to differentiate fraudsters from standard actors.

Most applied techniques using exploratory research

Through network science techniques, community detection was the most applied mechanism to verify the formation of clusters, particularly cartels. It was chosen to examine collusive trades in significant markets, in which market shares with appropriate prerequisites for cartel appearance were uncovered in previously hidden connections. Nevertheless, the approach to detecting cartels was sometimes merely representative and could not always convincingly indicate crimes without employing labeled data.

Community detection also allowed witnessing organizations making extensive agreements with each other rather than making alliances beyond their communities. It pointed to communities practicing transgressions in their contracts, spotting groups where no agreements could provide the awarded objects. The diversification of co-bidding communities formed in different regions was also evaluated, in which alleged cartel formation was identified based on low regional and contractual diversification of the communities, suggesting collusion of companies in some regions to supply specific products.

In turn, it was analyzed to what extent the volume of single bids differed in diverse communities within a network. The diagnosis was that fraud is not erratically disseminated in procurement processes; the higher the clustering of single bidders, the more likely the analysis of their neighbors would bring up more fraud cases. Besides, communities with extra bonding social capital stood at a greater risk of fraud, and communities with a more assorted extrinsic connectivity were less susceptible to it. Furthermore, some networks were split into different clusters, categorizing intersections as new communities and classifying them as conspiracy cases when there was a high degree of nodes overlaying.

Over time, an interchangeable and nonlinear association between a corruption risk index and the closeness centrality was pointed out. It indicated the closeness measure effectively revealing fraud risk after government changes.

Betweenness was used to estimate the likelihood of agents having weak links with each other but still being a necessary mediator in trades and partnership management. They were able to underline the relational effects of organizations that could control the award of contracts.

It should be noted that, regardless of using some measures like network diameter, mean path length, and degree measures, these metrics stayed unchanged in certain studies, not supplying sufficient facts to predict the risk of crimes. Besides, in networks of bipartite nature, these measures alone were not sufficient to demonstrate evidence of a crime.

Based on the facts exposed above, we can realize that some mechanisms of network science are not always capable of providing accurate sole outcomes. Therefore, they are usually applied with different frameworks, such as neighbor diversity, single bidder, diversity of winners, victory rate of a bidder, or even creating indexes from intrinsic process parameters.

Research gaps

We found three crucial gaps during this research; the first is related to the online auction in the context of the public sector. Although the process is very similar in the public and private domains, it has been more researched in the private sector, contributing to developing fraud detection techniques in this segment. This type of bidding has a substantial share of public spending and should be considered when studying crimes in public contracts. A good example is electronic trading accounts which represent more than 90% of public contracting expenditures in Brazil's federal government (Governo Federal do Brasil 2022).

The second gap is related to the limited use of objective indicators to measure corruption risk in tenders (Lyra et al. 2021a, b.1). Corruption indices are usually obtained based on qualitative research and tend to reflect individual researchers' perceptions and biases. Objective microdata such as the number of bidders, market share of winners, bidding modality, time-lapse for awarding a contract, amendment of contractual clauses during its regular term, duration of the contract, and extension of the contract period after the end of the regular term are intrinsic to biddings operations. They promote greater transparency and accuracy in developing a fraud indicator, especially when well-calibrated and weighted. Although We collected 5 articles from the same group of authors and co-authors, using data related to EU countries. Currently, corruption indicators are mostly subjective and somewhat unreliable (Fazekas & Kocsis 2017). Different red flags in public procurement processes could be better explored and taken as elements of a composite indicator, using regression coefficients to estimate the weights.

Third, it is essential to continue looking for solutions that allow the use of unlabeled data more accurately to detect crimes against the public administration. Despite the increase in the amount of data available from the public sector, the vast majority of those with access are not classified. In this sense, a methodology that proved to be more feasible in the absence of pre-classified data is using network science tools.

Research limitations

The methods applied in this systematic literature review contain limitations for identifying papers beyond the keywords defined in the search string. The chosen essays may not comprise extensive studies in crime detection in public procurement since some keywords or expressions may have been neglected, such as Red Flag, Hoax, Trick, Blow, and Fake. Besides, Boolean connectors might have restricted the search query. Furthermore, only the repository Scopus were investigated, and maybe vaster results could have been reached using different platforms.

However, concerning the research period between 2011 and 2021, we believe it is not demanding to search periods prior to 2011 since the search query for the period before it returned 745 documents, representing 25% of the total. In addition, the fraud detection methodology in the public sector, mainly regarding procurement, has begun to attain visibility more lately. Taking a more specific search string regarding the area of interest, for example, such as ("network analysis" OR SNA OR "Network science") AND ("public procurement") AND (corruption OR fraud), we found only eight articles in Scopus, where the publications start to appear only in 2016 with one paper, 2017 with another one, 2020 with three and 2021 with another three articles. It suggests that the specific subject is recent and very little investigated.

Conclusion and further studies

Many existing methods for detecting fraud in the public sector rely on rules established by an expert. However, the access to an increasing number of data sources and the ability to analyze large volumes of data through machine learning, network science, and other methods offers the possibility to understand and predict fraudulent activities.

Here, we used the PRISMA methodology to identify the most relevant contributions to the topic. We characterize the research efforts in the field, identifying the most active authors, venues, methodologies, and application contexts. Finally, we emphasized the increasing importance of network science and machine learning in this domain.

We highlighted 48 contributions chosen based on their relevance to respond to the research question regarding the methods used to detect crimes in public procurement. The survey allowed us to investigate deeper into machine learning methods, most of them using supervised techniques and network science approaches to detect crimes in public contracts.

Machine learning techniques have proved to be very efficient in identifying instances of inadequate public spending use. However, there is no specific best tool to detect all types of crime. We found that the most effective and most used approaches rely on Random Forest algorithms. At the same time, ANN was particularly adequate to detect shill bidding in real-time, SVM in finding fraudsters through neighbor diversity features, Naive Bayes detecting inefficiency and corruption in tenders, and Random Forest in uncovering bid-rigging cartels.

Network science methods have been used to estimate centrality measures and relevant communities among the participating actors, particularly to verify the existence of cartels. The use of network science is particularly relevant when studying unlabeled data. That is when there is no clear identification of which contracts have been flagged as subject to illegal procedures (e.g., fraudulent activity). The absence of confirmatory information was addressed through exploratory investigation and identifying hidden relationships between network agents. That said, we perceive that network science is a central framework for developing more sophisticated methods for studying fraud, collusion, and corruption activities lacking confirmatory data. Future work should direct efforts to investigate the issues addressed in the uncovered gaps, especially identifying fraudulent actions in electronic public auctions, since this is becoming the most used means by world governments in public acquisition.

Appendix 1 Articles classification

See Table 10.

Table 10 Method, connectivity, and investigation type

Article No	Method	Connected network	Investigation type
1	Statistics	No	Exploratory
2	Network science	Yes	Exploratory
3	Statistics	No	Exploratory
4	Network science and machine learning	Yes	Labeled data
5	Network science	Yes	Exploratory
6	Network science	Yes	Exploratory
7	Network science	Yes	Labeled data
9	Machine learning	No	Labeled data
9	Network science	Yes	Exploratory
10	Network science and machine learning	Yes	Labeled data
11	Network science	Yes	Labeled data
12	Network science	Yes	Labeled data
13	Network science	Yes	Labeled data
14	Network science	Yes	Exploratory
15	Statistics	Yes	Exploratory
16	Machine learning	Yes	Labeled data
17	Machine Learning	No	Labeled data
18	Network science and statistics	Yes	Exploratory
19	Machine learning	No	Labeled data
20	Machine Learning	Yes	Labeled data
21	Machine learning	No	Labeled data
22	Network science	Yes	Labeled data
23	Network science	Yes	Labeled data
24	Network science	Yes	Labeled data
25	Machine learning and statistics	No	Exploratory
26	Network science and machine learning	Yes	Exploratory
27	Network science and statistics	Yes	Exploratory
28	Machine learning	Yes	Labeled data
29	Machine learning	No	Labeled data
30	Clustering Using REpresentatives (CURE)	No	Labeled data
31	Network science	Yes	Exploratory
32	Network science and statistics	Yes	Exploratory
33	Network science	Yes	Labeled data
34	Machine learning	Yes	Labeled data
35	Network science	Yes	Exploratory
36	Network science	Yes	Exploratory
37	Machine learning and statistics	No	Exploratory
38	Network science and machine learning	No	Labeled data
39	Network science	Yes	Exploratory
40	Network science	Yes	Exploratory
41	Network science	No	Exploratory
42	Network science	Yes	Labeled data
43	Network science	Yes	Labeled data
44	Machine learning	Yes	Labeled data
45	Statistics	Yes	Labeled data
46	Statistics	No	Exploratory
47	Machine learning	No	Labeled data
48	Network science	Yes	Labeled data

Table 10 presents the methods, data connectivity, and type of investigation carried out in each article.

From Table 10, we can see that the most frequent approaches when using label data are generally related to machine learning, representing 100% of articles exclusively concerning it. They represent 50% of these studies regarding network science using label data. On the other hand, studies with label data using other methods represent 53% of articles. That said, it emerges that in studies where label data is available, the use of machine learning as a preferred method prevails. In contrast, approaches such as network science are evenly distributed with the use of label data and exploratory research.

Appendix 2 Articles' purpose

See Table 11

Table 11 Objective of the articles (ordered by number of citations)

N _o	Author/Year	Objective
1	(Fazekas, Toth, & King, 2016)	Intend to measure grand corruption and state capture at the micro-level of individual contracts and tenders based solely on "objective" behavioral data
2	(Fazekas & Tóth 2016)	Identify clusters at high-corruption risk in public procurement in a contractual network of procuring authorities and suppliers
3	(Fazekas & Kocsis 2017)	Contemplate the absence of feasible corruption indicators to estimate the level of fraud in public procurement
4	(Chiu et al. 2011)	Proposes a model to discover fraudsters based on Internet auction transaction records using data collected from Yahoo! Auctions
5	(Reeves-Latour & Morselli 2017)	Outlines the appearance of bidding processes and corporate criminal networks in the building industry. It describes the development of corrupt networks around dubious and continued actions of criminality, collusion, and bribery and analyzes relations among players
6	(Sedita & Apa 2015)	Perform an empirical investigation into a network with relationships between companies involved in public procurement schemes in the construction sector in the Veneto region between 2008 and 2012
7	(Lin et al. 2012)	Detect irregular conduct and supply a classification method to evaluate the level of risk of defrauding online auctions by collusive groups
9	(Almendra 2013)	Recognize fraudulent dealers in auctions and combat non-delivery of non-existent goods before the scammers can act
9	(Luna-Pla & Carlock 2020)	Conduct an empirical process for modeling corruption practices concerning a network of hundreds of shell firms
10	(Yu & Lin 2013)	Uncover fraud cases on online auction sites and detect the accounts of the most active fraudsters on the network, in addition to helping buyers identify them
11	(Wachs & Kertesz 2019)	Intend to detect the potential cartels involved in bidding processes
12	(Wachs et al. 2019)	Investigate the social factors of corruption and evaluates the risk of scams due to suppressed competition and the absence of clearness in the public contracts awarded to the deal. It compares the risk of fraud in agreements with fragmented social networks and excess compulsory social capital
13	(Li et al. 2013)	Seeks to measure and improve the predictive capacity of reputation systems in the fight against collusion in peer-to-peer networks
14	(Hosseini et al. 2019)	Determine the character of the endemic corruption threats in the Iranian construction initiative, its grade of occurrence, and the strength of its effect

Table 11 (continued)

N _o	Author/Year	Objective
15	(Morselli & Ouellet 2018)	Investigate whether the companies' market share can be identified through the co-bidding similarity in bidding patterns and whether this measure forecasts their market share in processes with recognized conspiracy
16	(Huber & Imhof 2019)	Predict collusion through bid-rigging cartels in tenders within the Swiss construction sector
17	(Ganguly & Sadaoui 2018)	Identify and characterize Shill Bidding in e-auctions employing real data from eBay
18	(Velasco et al. 2021)	Identify the principal patterns of corruption risk in public expenses, such as stunts, conflicts of interest, and the participation of shell companies in bidding processes
19	(Rabuzin & Modrusan 2019)	Compare prediction models of suspicious bids and develop a model to detect suspicious one-bid in public procurement processes
20	(Anysz et al. 2018)	Compare two machine learning techniques for detecting collusion between bidders in road construction tenders
21	(Almendra & Enachescu 2014)	Identify fraud in product listings on online auction sites for fair operators, customers, and vendors
22	(Zhu et al. 2011)	Seeks to extract characteristic traits in the behaviors of fraudsters from the feedback reputation system, who manipulate their feedback scores by engaging in various legitimate sales
23	(Wachs et al. 2021)	Aim to investigate the risk of fraud in public procurement agreements of the Member States of the European Union
24	(Fazekas & Wachs 2020)	Explore the association between corruption and market anatomy in public procurement approaches
25	(Sharma et al. 2019)	Identify the risks of corruption in public procurement and discover the contribution of suspected actors and their relationships in the network
26	(Dhurandhar et al. 2015)	It analyzes corporation loss due to fraud and proposes tools and processes to quickly and cheaply identify fraud/risks related to public and private procurement and potential wrongdoers
27	(Mundra & Rakesh 2014)	Investigates how a contractor's network position affects its success in winning public procurement projects through its ability to partner and influence the network
28	(Popa 2019)	It analyzes a set of public procurement data from European countries in order to identify the emergence of close links between bidders and public administration, defined in terms of repeated interaction and geographical dispersion
29	(Gallego et al. 2021)	Determine trades that can become troublesome and forecast inefficiency and fraud in public procurement by employing a dataset with about two million public procurement processes in Colombia
30	(Alzahrani & Sadaoui 2020)	Identify shill bidding fraud in online auctions
31	(Lei et al. 2018a, b)	It intends to measure the user reputation (trust score) by considering the feedback reliability status for all transactions in the e-commerce market, such as e-auction
32	(Mamavi et al. 2017)	Explore conniving networks deriving from collaborative associations among firms. It suggests a protocol to investigate networks within the public sector based on a quantitative method. It also emphasizes the interactive effects of companies that might impact the granting of contracts in decision-making, besides the influence of weak and strong bonds between firms on the contracts
33	(Khomnotai & Lin 2015)	Recognize scams on online auctions by exploring the concept of neighborhood diversity as an effective resource to identify competitiveness and joint participation between companies that may be acting together to defraud bids regularly
34	(Dadfarnia et al. 2020)	Detect fraudulent users in collusive acts in online auctions
35	(Zhu et al. 2020)	Design a social networking standard to detect possible bidder conspiracy in the construction sector
36	(Lei et al. 2018a, b)	Detect the formation of scheming groups in auctions through collusive activities or manipulation of the average price of the bids towards a higher or even unrealistic winning price

Table 11 (continued)

N _o	Author/Year	Objective
37	(Majadi et al. 2017)	Provides a brief overview of leading research on bidding patterns to detect shill bidders in online auctions, illustrates the characteristics of such bidding patterns, and presents case studies identifying shill bidding behaviors in eBay's datasets
38	(Carneiro et al. 2020)	Detect fraud in public tenders of Portugal, investigate the clearness of the deals between network actors and promote access to relevant data
39	(Lin & Khomnotai 2016)	It seeks to determine the change in the reliability perception of competitors within the bidder's reputation system on online auction sites when creating false transactions
40	(Lyra et al. 2021a, b,2)	Explore analytical structures that can help public controls recognize agents more sensitive to irregular activities in Public Procurement in Ceará (Brazil)
41	(Owusu et al. 2021)	It examines the effects of corruption at the stages of procurement processes and develops a framework for modeling the impacts of corruption and improving anti-corruption measures
42	(Xiao et al. 2021)	Examine the features of conniving bidding networks and the types of collusive bidding groups in the construction industry
43	(Nicolás-Carlock & Luna-Pla 2021)	Understand the criminal collusion of companies implicated in corruption scandals in public tenders through exclusive data from reported corruption cases in Mexico, where several companies were manipulated to misappropriate billions of dollars
44	(Fazekas et al. 2021)	Explores through empirical data the role of governance played by criminal groups organized in corruption networks, facilitating corrupt transactions by reducing the costs of searching, negotiating, and enforcing public purchases
45	(Czibik et al. 2021)	It studies the corruption risks in EU defense procurement that present a significant potential for corruption and state capture. It uses a large set of contract data covering ten years of investigation
46	(Perevezentceva et al. 2021)	It studies fraud in public procurement and the development of the control role in the bidding system and suggests introducing an anti-corruption investigation of the records of candidates participating in competitive processes
47	(Abidi et al. 2021)	Detect Shill Bidding (SB), which is when the seller presents false bidders to increase the final price of a bid
48	(Davydenko et al. 2017)	Detect likely risks of corruption in public tenders competition using empirical data from procurement processes and labeled data encompassing the registration of suppliers with lawsuits in fraudulent cases of public acquisitions

Table 11 presents the objectives of the selected writings to disclose supposed fraud, corruption, or collusion in public contracting. Here we can glance at the scope of each survey, its instructions, and the outcomes presented. We verified that the most adopted methods to capture crimes in public acquisition vary widely, ranging from network science tools (social network analysis and complex network techniques) to statistical analysis and machine learning approaches.

Out of 48 articles, 22 (45%) are related to Network Science (Articles no. 2, 5, 6, 7, 9, 11, 12, 13, 114, 22, 23, 24, 31, 33, 35, 36, 39, 40, 41, 42, 43, 48); 11 (23%) to Machine Learning (Articles no. 8, 16, 17, 19, 20, 21, 28, 29, 34, 4447); five (10%) statistical methods (Articles no. 1, 3, 15, 45, 46); and 10 (21%) used combined methods like “network science & machine learning” (Articles no. 4, 10, 26, 38), “network science & statistics” (Articles no. 18, 27, 32), “machine learning & statistics” (Articles no. 25, 37) and “Clustering Using REpresentatives – CURE, (Article no. 30).

From the selected articles 34 (71%) used connected Network in their analysis (Articles no. 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 24, 26, 27, 28, 31, 32, 33, 34, 35, 36, 39, 40, 42, 43, 44, 45, 48), and 29 used labeled data (Articles no. 4, 7, 8, 10, 11, 12, 13, 16, 17, 19, 20, 21, 22, 23, 24, 25, 29, 31, 33, 34, 38, 42, 43, 44, 45, 47, 48).

Acknowledgements

We are very grateful for the suggestions given in the revision carried out by the referee, in which we were able to clarify some points and enrich the research.

Author contributions

MSL performed the research selection in the Scopus repository and the 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 Ph.D. scholarship. FLP, FB and BD 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

All the documents supporting the conclusions of this article were collected using the repository of scientific research, Scopus.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 February 2022 Accepted: 28 November 2022

Published online: 15 December 2022

References

- Abidi W, Daoudm M, Ihnaini B, Khan M, Alyas T (2021) Real-time shill bidding fraud detection empowered with fused machine learning. *IEEE Access* 9:113162
- Akoglu L, Tong H, Koutra D (2015) Graph based anomaly detection and description: a survey. *Data Mining Knowl Discov* 39(3):626
- Almendra V (2013) Finding the needle: a risk-based ranking of product listings at online auction sites for non-delivery fraud prediction. *Expert Syst Appl* 40(2):4805
- Almendra V, & Enachescu D (2014) Using self-organizing maps for fraud prediction at online auction sites. 15th international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)
- Alzahrani A, & Sadaoui S (2020) Clustering and labeling auction fraud data. *International conference on data management, analytics and innovation (ICDMAI)*
- Anysz H, Foremny A, & Kulejewski J (2018) Comparison of ANN classifier to the neuro-fuzzy system for collusion detection in the tender procedures of road construction sector. 3rd world multidisciplinary civil engineering, architecture, urban planning symposium (WMCAUS)
- Bodin O, & Norberg J (2005) Information network topologies for enhanced local adaptive management. *Environ Manag* 35(2):175–193. <https://doi.org/10.1007/s00267-004-0036-7>
- Borras J, Cousins B, Liu J, McKay B, & Milhorance De Castro C (2015) Emerging trends in global commodities markets: the role of Brazil and China in contemporary agrarian transformations. *International Institute of Social Studies, The Hague*
- Bosio E, Djankov S, Glaeser E, Shleifer A (2022) Public procurement in law and practice. *Am Econom Rev* 112(4):1091–1117
- Carneiro D, Veloso P, Ventura A, Palumbo G, & Costa J (2020) Network analysis for fraud detection in portuguese public procurement. *Conference on intelligent data engineering and automated learning*
- Chiu C, Ku Y, Lie T, Chen Y (2011) Internet auction fraud detection using social network analysis and classification tree approaches. *Int J Electron Comm* 15(3):123
- Costa GA, Machado D, Martins V (2020) The efficiency of social control in municipal bidding: a study in social observatories. *Soc Contab Gestão* 14(4):112
- Czibik Á, Fazekas M, Sanchez A, J, W, (2021) Networked corruption risks in European defense procurement understanding complex systems. Springer, USA

- Dadfarnia M, Adibnia F, Abadi M, Dorri A (2020) Incremental collusive fraud detection in large-scale online auction networks. *J Supercomput* 76(9):7416
- Davydenko V I, Morozov N V, & Burmistrov M I (2017) Adaptation of cluster analysis methods in respect to vector space of social network analysis indicators for revealing suspicious government contracts. *IEEE 5th international conference on future*
- Dhurandhar A, Graves B, Ravi R, & Maniachari G (2015) Big data system for analyzing risky procurement entities. *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*
- Ding X, & Yang Z (2020) Knowledge mapping of platform research: a visual analysis using VOSviewer and CiteSpace. *Electron Commer Res*
- Dong S, Wang H, Mostafizi A, Song X (2020) A network-of-networks percolation analysis of cascading failures in spatially co-located road-sewer infrastructure networks. *Physica A* 538:122971
- Fazekas M, Kocsis G (2017) Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data. *British J Polit Sci* 50(1):155
- Fazekas M, Tóth IJ (2016) From corruption to state capture: a new analytical framework with empirical applications from Hungary. *Polit Res Q* 69(2):320–334
- Fazekas M, Wachs J (2020) Corruption and the network structure of public contracting markets across government change. *Polit Governance* 8(2):153
- Fazekas M, Toth IJ, King LP (2016) an objective corruption risk index using public procurement data. *European J Crim Pol Res* 22(3):369
- Fazekas M, Sberna S, Vannucci A (2021) The extra-legal governance of corruption: tracing the organization of corruption in public procurement. *Governance* 35(4):1139
- Gallejo J, Rivero G, Martinez J (2021) Preventing rather than punishing: an early warning model of malfeasance in public procurement. *Int J Forecast* 37(1):360
- Ganguly S, & Sadaoui S (2018) Online detection of shill bidding fraud based on machine learning techniques. *31st international conference on industrial engineering and other applications of applied intelligent systems (IEA/AIE)*
- Governo Federal do Brasil (2022). <https://www.gov.br/compras/pt-br/aceso-a-informacao/noticias/decreto-aprimora-regras-do-pregao-eletronico>. Accessed 08/04/2022
- Grassi R, Calderoni F, Bianchi M, Torriero A (2019) Betweenness to assess leaders in criminal networks: new evidence using the dual projection approach. *Social Networks* 56:23–32
- Horvat T, Havaš L, & Logozar R (2015) The analysis of keyword occurrences within specific parts of multiple articles — the concept and the first implementation. *SSN* 1846–6168 UDK 001.8.
- Hosseini M, Martek I, Banihashemi S (2019) Distinguishing characteristics of corruption risks in iranian construction projects: a weighted correlation network analysis. *Sci Eng Ethics* 26(1):205
- Huber M, Imhof D (2019) Machine learning with screens for detecting bid-rigging cartels. *Int J Indus Organ* 65:277
- IDB (2018) Better spending for better lives: how latin america and the caribbean can do more with less. *Inter-American Development Bank*
- Kartika D, Kebijakan L, Barang P, Pemerintah J (2020) The impact of e-procurement implementation on public procurement's corruption cases – evidences from Indonesia and India. *J Kajian Wilayah* 11:193–212
- Khomnotai L, Lin J (2015) Detecting fraudsters in online auction using variations of neighbor diversity. *Int J Eng Technol Innov* 5(3):153
- Kuncoro M (2013) *Research methods for business and economics* edition 4. Indonesia Erlangga Publisher, Jakarta
- Lauchs M, Keast R, Yousefpour N (2011) Corrupt police networks: uncovering hidden relationship patterns, functions and roles. *Polic Soc* 21(1):110–112
- Lei M, Yin Z, Li S, & Li H (2018a) Detecting the collusive bidding behavior in below average bid auction. *13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)*.
- Lei M, Yin Z, Li S, & Li H (2018b) Trust model based on Islamic business ethics and social network analysis. *Int J Adv Sci, Eng Inform Technol*
- Li Z, Shen H, Sapra K (2013) Leveraging social networks to combat collusion in reputation systems for peer-to-peer networks. *IEEE Trans Comput* 62(9):1745
- Lin J, Khomnotai L (2016) Improving fraudster detection in online auctions by using neighbor-driven attributes. *Entropy* 18(1):18010011
- Lin S, Jheng Y-Y, Yu C (2012) Combining ranking concept and social network analysis to detect collusive groups in online auctions. *Expert Syst Appl* 39(10):9079
- Luna-Pla I, Carlock N (2020) Corruption and complexity: a scientific framework for the analysis of corruption networks. *Appl Net Sci*. <https://doi.org/10.1007/s41109-020-00258-2>
- Lyra M, Bacao F, & Pinheiro F (2021a) Public procurement fraud detection – a review using network analysis. *Complex Network* 2021a
- Lyra, M., Curado, A., Damásio, B., Bacao, F., & Pinheiro, F. (2021b). Characterization of the firm-firm public procurement co-bidding network from the state of Ceara (Brazil) Municipalities. *Appl Net Sci* <https://doi.org/10.1007/s41109-021-00418-y>
- Majadi N, Trevathan J, & Bergmann N (2017) Analysis on bidding behaviours for detecting shill bidders in online auctions. *Proceedings - 2016 16th IEEE international conference on computer and information technology, cit 2016, 2016 6th international*
- Mamavi O, Meier O, Zerbib R (2017) How do strategic networks influence awarding contract? Evidence from French public procurement. *Int J Publ Sec Manage* 30(4):357
- Marsden P (2005) Network analysis. In *encyclopedia of social measurement*
- Merigó J (2016) Academic research in innovation: a country analysis. *Scientometrics* 108:559–593
- Merigó J, Mas-Tur A, Roig-Tierno N, Ribeiro-Soriano D (2015) A bibliometric overview of the journal of business research between 1973 and 2014. *J Bus Res* 68:2645–2653
- Miralles-Pechuán L, Jiménez F, & García J (2022) An alternative auction system to generalized second-price for real-time bidding optimized using genetic algorithms. *Lecture Notes in Networks and Systems*

- Mlondo N J (2013) Effectiveness of knowing your customer policy in combating money laundering in commercial banks in Tanzania: a case of bank of Africa (T) limited (Doctoral dissertation, Mzumbe University)
- Morselli C (2008) Inside criminal networks studies of organized crime. Springer, New York
- Morselli C, Ouellet M (2018) Network similarity and collusion. *Soc Net* 55:21
- Motta FCP (2003) Teoria das Organizações: evolução e crítica. 2nd ed. Pioneira Thomson Learning, SP
- Mufutau G, Mojisola O (2016) Detection and prevention of contract and procurement, fraud Catalyst to organization profitability. *J Business Manage* 18:9
- Mundra A, & Rakesh N (2014) Online hybrid model for online fraud prevention and detection. *Adv Intell Syst Comput*
- Mvelase T O (2015) The impact of non-compliance with Eskom procurement policies (Doctoral dissertation, Nelson Mandela Metropolitan University)
- Nicolás-Carlock J, & Luna-Pla I (2021) Conspiracy of corporate networks in corruption Scandals. *Front Phys*
- OECD (2021) Government at a Glance 2021. OECD Publishing, Paris
- Olsson P, Folke C, Berkes F (2004) Adaptive comanagement for building resilience in social-ecological systems. *Environ Manag* 34(1):75–90. <https://doi.org/10.1007/s00267-003-0101-7>
- Owusu E, Chan A, Wang T (2021) Tackling corruption in urban infrastructure procurement: dynamic evaluation of the critical constructs and the anti-corruption measures. *Cities* 119:103379
- Padhi S, Mohapatra P (2011) Detection of collusion in government procurement auctions. *J Purch Supply Manag* 17:207–221
- Paganetto L, Scandizzo PL (2015) Governance, moral and economic values in achieving dynamism in an anaemic Europe. Springer, Cham
- Pati D, Lorusso L (2018) How to write a systematic review of the literature. *Health Environ Res Des* 11:15–30
- Perevezentceva E, Osipova K, & Skrynnikova K (2021) Corruption in the implementation of public procurement from small and medium businesses. *Lecture notes in networks and systems*
- Popa M (2019) Uncovering the structure of public procurement transactions. *Business Politics* 21(3):351
- Rabuzin K, & Modrusan N (2019) Prediction of public procurement corruption indices using machine learning methods. 11th international conference on knowledge management and information systems (KMIS)
- Reeves-Latour M, Morselli C (2017) Bid-rigging networks and state-corporate crime in the construction industry. *Soc Net* 51:158
- Robson C (2002) Real world research: a resource for social scientists and practitioner-researchers. Wiley-Blackwell
- Rustiarini N, Sutrisno T, Nurkholis N, Andayani W (2019) Why people commit public procurement fraud? The fraud diamond view. *J Publ Procure* 19(4):345–362
- Sandström A, Carlsson L (2008) The performance of policy networks: the relation between network structure and network performance. *Policy Stud J* 36(4):497–524. <https://doi.org/10.1111/j.1541-0072.2008.00281.x>
- Sedita S, Apa R (2015) The impact of inter-organizational relationships on contractors' success in winning public procurement projects: the case of the construction industry in the Veneto region. *Int J Project Manage* 33(7):1548
- Shaffril, H. A., Samsuddin, S. F., & Samah, A. A. (2020). The ABC of systematic literature review: The basic methodological guidance for beginners. *Quality & Quantity*, 1–28.
- Silva Filho J B (2017) A eficiência do controle social nas licitações e contratos administrativos. Master's thesis - Universidade Nove de Julho, São Paulo.
- Sharma S, Sengupta A, Panja S (2019) Mapping corruption risks in public procurement: uncovering improvement opportunities and strengthening controls. *Publ Perform Manage Rev* 42(4):924
- Van Eck, N., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: methods and practice* (pp 285–320). Springer.
- Van Erven G C, Carvalho R N, & de Holanda e a (2017) Graph database: a case study for detecting fraud in acquisition of Brazilian government. 12th Iberian conference on information systems and technologies (CISTI)
- Velasco R, Carpanese I, Interian R, et, a. (2021) A decision support system for fraud detection in public procurement. *Int Trans Operat Res* 28(1):27
- Volosin N (2015) Datos abiertos, corrupción y compras públicas. RIGA - ILDA.
- Wachs J, & Kertész J (2019) A network approach to cartel detection in public auction markets. *Scientific reports*
- Wachs J, Yasseri T, Lengyel B, Kertész J (2019) Social capital predicts corruption risk in towns. *Royal Soc Open Sci* 6(4):18210
- Wachs J, Fazekas M, Kertész J (2021) Corruption risk in contracting markets: a network science perspective. *Int J Data Sci Anal* 12(1):45
- Weigel U, & Ruecker M (2017) e-procurement. In: the strategic procurement practice guide (pp 179–209)
- Wensink W, Vet M (2013) Identifying and reducing corruption in public procurement in the EU. European Commission, Bruxelles
- Whiteman R (2019) Fraud and corruption tracker. The chartered institute of public finance and accountancy – CIPFA
- Wu Z, Lu Z, Han W, Long C, Ma Z, & Huang K (2021) Research on identification technology of encircled serial tags based on big data and semantic analysis. 2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering, ICBAIE
- Wu Chebili B, La Cascia H, Collineau F, Salomon A, Calvet B, & Moreau Y (2022). Electronic government procurement implementation types
- Xiao Q, Chang HH, Geng G, & Liu Y (2018) An ensemble machine-learning model to predict historical PM2. 5 concentrations in China from satellite data. *Environ Sci Technol* 52(22):13260–13269
- Xiao L, Ye K, Zhou J, Ye X (2021) A social network-based examination on bid riggers' relationships in the construction industry: a case study of China. *Buildings* 11(8):363
- Yu C, Lin S (2013) Fuzzy rule optimization for online auction frauds detection based on genetic algorithm. *Electron Commerce Res* 13(2):169
- Železnik D, Vošner H, Kokol P (2017) A bibliometric analysis of the journal of advanced nursing, 1976–2015. *J Adv Nurs* 73:2407–2419

- Zhu Y, Zhang W, & Yu C (2011) Detection of feedback reputation fraud in Taobao using social network theory. *Proceedings - 2011 international joint conference on service sciences, IJCSS*
- Zhu J, Wang B, Li L, et al (2020) Bidder network community division and collusion suspicion analysis in chinese construction projects. *Adv Civil Eng* 2020:1

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
