

Usando redes complexas na identificação de empresas fraudulentas em licitações públicas

Ane Karoline da Silva Pereira¹, Yuran Moreira Vita¹, Gardenya da Silva Felix²,
João Mateus Freitas Gimaque¹, Marcos Luan Sousa Damasceno¹,
Bruno César Barreto de Figueirêdo^{1,2}

¹Departamento de Computação - Universidade Estadual de Roraima - UERR
R. Sete de Setembro, 231 - Canarinho, Boa Vista - RR, 69306-530

²Tribunal de Contas do Estado de Roraima
Rua Prof. Agnelo Bittencourt nº 126 - Centro - CEP: 69301-430

pereiraanekaroline@gmail.com, yuran.vita@uerr.edu.br,

{brunocesar, gardenya}@tce.rr.leg.br, joaomateus568@gmail.com,

luansousa9141@gmail.com

Abstract. *Detecting fraudulent schemes in public tenders is hard work, especially when such acts are common, and it is not possible to do so manually. This article proposes a new approach to identifying companies that have engaged in collusion, using complex networks and measures of centrality such as Betweenness (BW), Eigenvector centrality (EV), PageRank (PR), and Weighted Degree (WD). The analysis period was from January 2021 to June 2022, reaching a precision of more than 71% and an accuracy of 68%. In addition to proving the relationship between the companies' earnings and the classifications provided by the centralities through the Pearson and Spearman correlations.*

Resumo. *Detectar esquemas fraudulentos em licitações públicas é um trabalho árduo, principalmente quando tais atos são costumazes, sendo inviável de forma manual. Este artigo propõe uma nova abordagem na identificação de empresas que se envolveram em conluíus, utilizando redes complexas e medidas de centralidade tais como: Betweenness (BW), Eigenvector centrality (EV), PageRank (PR) e Weighted Degree (WD). O período de análise foi de janeiro de 2021 a junho de 2022, onde obteve-se uma precisão superior a 71% e uma acurácia de 68%, além da comprovação da relação entre os ganhos das empresas e as classificações fornecidas pelas centralidades por meio da aplicação das correlações de Pearson e Spearman.*

1. Introdução

Define-se “licitação pública” como o processo de concorrência entre fornecedores, deflagrado pelo setor público, com vistas a auferir as melhores condições na aquisição de bens e serviços [Mankiw and Taylor 2011]. Apesar de existir legislação específica de combate às fraudes em licitações (Lei 14.133, DE 1/4/2021), a administração pública ainda se encontra bastante vulnerável a esse tipo de prática. Nesse diapasão a associação indevida de empresas, com o objetivo de formação cartéis, onde, supostos concorrentes, simulam uma disputa fictícia, com ganhadores escolhidos previamente. Essa prática tem como

objetivo a majoração de preços e eliminação da concorrência. A grande frequência e diversidade de esquemas de fraude inviabiliza os processos manuais de detecção e combate [Figueiredo et al. 2020].

Estratégias para determinar as entidades mais relevantes em um cenário são temas recorrentes em estudos científicos. A lista vai de técnicas de mineração de dados [Ghedini Ralha and Sarmento Silva 2012, Costa and Aparicio 2011], inteligência artificial [Cunha and Bugarin 2014, Silva and Ralha 2010], redes complexas [Scabini et al. 2021], entre outros.

Este artigo propõe a utilização de redes complexas no mapeamento de processos licitatórios, em um período estabelecido, visando a detecção de conluios. Esse mapeamento será realizado por meio de uma métrica estabelecida como critério para a criação das entidades e dos seus relacionamentos. Uma vez mapeadas as licitações o uso de medidas de centralidade clássicas será capaz de apontar quais as empresas com maior chance de pertencer a um esquema de associação ilícita, as classificando segundo sua relevância.

A organização do trabalho é a seguinte: a seção 2 discute os trabalhos relacionados à detecção de entidades mais relevantes num contexto, com foco especial na detecção de fraudes e de seus participantes; a seção 3 apresenta a metodologia utilizada, destacando a métrica proposta na elaboração das redes complexas e as fontes de dados utilizadas; a seção 4 descreve as contribuições da pesquisa e a seção 5 traz as conclusões e sugere direções para futuros trabalhos.

2. Trabalhos Relacionados

Uma rede complexa consiste em um grafo onde os nós representam as entidades e as arestas os relacionamentos entre elas. Sua estrutura topológica é, em geral, a representação, ou modelagem, de um cenário real ou fictício, que pode evoluir ao longo do tempo [Boccaletti et al. 2006], sendo compostas de estruturas que não são completamente regulares ou aleatórias [Carneiro 2016]. Em outras palavras, a sua estrutura é mutável ao longo do tempo, por representar a descrição de cenários também mutáveis que evoluem de forma temporal, não sendo regulares ou aleatórias justamente por representar tais cenários.

Este artigo utiliza a notação tensorial, na representação das matrizes de adjacência utilizando álgebra de alta ordem [De Domenico et al. 2013]. Uma vantagem em utilizar tensores é a sua forma compacta na representação das matrizes de adjacência. Em notação tensorial, uma linha do vetor $i \in N$ é dada pelo vetor covariante i_α ($\alpha = [1, N]$), pelo correspondente vetor contravariante i^α , ou seja, seu vetor dual, que é um vetor de coluna no espaço euclidiano. Um vetor canônico é atribuído a cada nó e um tensor de adjacência representando as conexões de uma rede complexa.

As centralidades fazem uso de métodos que, por meio de observação de casos específicos, inferem sobre o funcionamento das interações entre os nós, principalmente com relação à disseminação de informações em um grupo [Bavelas 1950]. Dessa forma, as medidas de centralidade, cada uma com uma estratégia, tentam resolver o problema de classificar a relevância dos nós de uma rede. Exemplos de centralidades clássicas incluem: *Betweenness* [Freeman 1978, Otte and Rousseau 2002], *Eigenvector centrality* [Bonacich 1972], *PageRank* [Bonacich 2007], *Weighted Degree* [Beveridge and Shan 2016] e *closeness* [Freeman 1978].

Além das chamadas centralidade clássicas, novos trabalhos têm proposto novas estratégias de centralidades, em especial utilizando as redes complexas multicamadas. Numa rede complexa multicamadas, cada camada pode representar um aspecto diferente de um mesmo cenário, representado pela rede [Bianconi 2018]. Em notação tensorial [De Domenico et al. 2013], um vetor canônico é atribuído a cada nó e um tensor de adjacência representa a rede multicamadas interconectada correspondente. Neste caso, um tensor $A_{ij}^{\alpha\beta}$ pode representar a intensidade da relação (que pode não ser simétrica) de um nó i na camada α para um nó j na camada β . Um caso específico diz respeito às redes Multiplex, que são um caso específico das redes multicamadas, onde um nó só pode se relacionar com outros nós pertencentes à mesma camada. Dessa forma, um tensor A_{ij}^{α} representa a intensidade da relação de um nó i com um nó j na camada α .

O uso da centralidade multiplex GCMN na detecção de empresas fraudulentas em um esquema de conluio em licitações públicas chamado “Operação Licitante Fantasma”, alcançou uma precisão de 92%, revocação de 94%, F_1 de 93% e uma acurácia de 54%. Esses resultados foram reforçados pela coerência da correlação entre os valores fraudados e a classificação fornecida pela centralidade, além da validade da proposta de divisão das empresas em grupos, proposta pelo artigo [De Figueirêdo et al. 2021].

Trabalhos utilizando redes complexas multicamadas, fazem uso de medidas de centralidade adaptadas a essas redes, tendo sempre como objetivo de apontar as entidades mais relevantes em um contexto. Nesse contexto surgem propostas de modelos para abordagens específicas como o COMplexVID-19 [Scabini et al. 2021], ou novas centralidades de propósito geral como: “Novel Multiplex PageRank in Multilayer Networks” [Tu et al. 2018], “Random walk centrality in interconnected multilayer networks” [Solé-Ribalta et al. 2016], “Random Walks on Multiplex Networks: Supplementary Information for Navigability of Interconnected Networks under Random Failures” [De Domenico et al. 2014] e a GCMN [De Figueirêdo et al. 2021].

Uma outra abordagem de sucesso trata do emprego de mineração de dados associada a regras de auditoria. O uso de tabelas de preços padrão, onde discrepâncias que extrapolem um percentual quantificável de acordo com as características de cada licitação, pode indicar a existência de associações ilícitas em licitações públicas. É a prática do chamado *overpricing*, ou extrapolação de preços, típica dos carteis de fornecedores [Silva and Ralha 2010, Costa and Aparicio 2011].

O uso de heurísticas e de agentes inteligentes, associado a técnicas de mineração de dados, tem sido testado com sucesso em bases de dados voltadas à auditoria como uma ferramenta de combate à formação de carteis em licitações públicas [Cunha and Bugarin 2014, Silva and Ralha 2010]. São encontradas abordagens semelhantes no desenvolvimento de Agentes de Mineração de Dados (AGMI) que, utilizando base de dados reais da Controladoria Geral da União (CGU), atuam como uma ferramenta de predição de corrupção e formação de carteis, conseguindo, nessa tarefa, uma acurácia de 90% de casos detectados [Ghedini Ralha and Sarmiento Silva 2012].

3. Metodologia

A Lei de acesso à informação (Lei no 12.527, de 18/11/2001) exige que todos os dados de licitações públicas estejam disponíveis ao público em geral. Em cumprimento a essa norma, o site de transparência de compras do Governo Federal disponibilizou uma API

(*Application Programming Interface*) para permitir de consultas às bases de dados de forma automatizada (<http://compras.dados.gov.br/docs/home.html>). Utilizou-se essa API na extração dos dados de todas as licitações federais entre janeiro de 2020 e junho de 2021. Esses dados subsidiaram a construção das redes complexas analisadas neste estudo de caso. O esquema de geração da rede complexa e das consequentes classificações pela aplicação das centralidades é descrito na Figura 1.

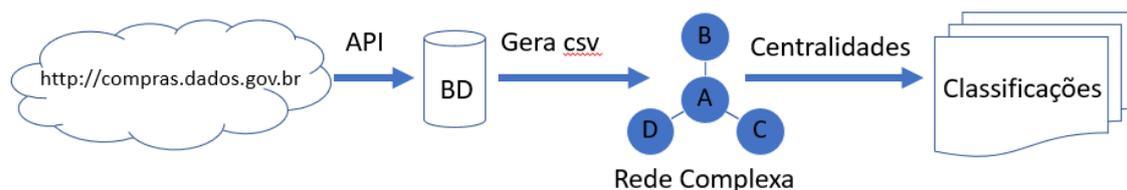


Figura 1. Esquema de geração das classificações por meio das centralidades.

Foi realizada a implementação dos algoritmos de extração de dados em linguagem de programação Java [Arnold et al. 2000] e, para persistência, foi utilizado o SGBD PostgreSQL (www.postgresql.org). Para os cálculos relativos às redes complexas, tais como: modularidade e das medidas de centralidade clássicas foi utilizado o software Gephi, versão 0.9.2 (www.gephi.org). Para as centralidades não disponíveis no Gephi, utilizou-se a linguagem Java ou a PLPGSQL, nativa do PostgreSQL, para as suas implementações. Para os cálculos das métricas *precision*, *recall*, *accuracy* e F_1 e das correlações propostas por Pearson e Spearman, foi utilizado o software livre JASP, versão 0.14.1.

3.1. Precisão, Revocação, Acurácia e F_1 Score

Uma maneira típica de quantificar a qualidade da classificação e agrupamentos é usando métricas como precisão e revocação [Carneiro 2016, Hilary 2015]. A precisão corresponde à fração de elementos relevantes entre os elementos recuperados, enquanto a revocação avalia a fração da quantidade total de elementos relevantes que foram recuperados. Ambas métricas ajudam a medir a relevância dos nós classificados. A F_1 Score é a média harmônica da precisão e da revocação e resume a qualidade do agrupamento em um valor. A acurácia indica o grau de concordância que há entre o resultado da medição e o dito valor verdadeiro (aquele que é aceito, desde que estabelecido por uma definição ou consenso) da grandeza [Carneiro 2016, Hilary 2015].

3.2. Correlações de Pearson e Spearman

O coeficiente de Pearson [Pearson 1905] r mede a força da correlação entre duas variáveis de escala métrica na estatística descritiva. O uso da correlação de Pearson deve obedecer ao seguinte: a escala de medição deve ser uma escala ou razão de intervalo; a distribuição das variáveis deve ser aproximadamente uniforme; a associação deve ser linear e não deve haver *outliers* nos dados.

Para entender a correlação de Spearman [Spearman 1904] (r_s) é necessário saber o que é uma função monotônica. Uma função monotônica preserva (ou inverte) a relação de ordem. O coeficiente de correlação de Spearman pode analisar a intensidade e a direção de uma relação monotônica entre duas variáveis contínuas ou ordinais. Em um relacionamento monotônico, as variáveis tendem a se mover na mesma direção

relativa, mas não necessariamente a uma taxa linear. Essas relações monotônicas podem ser estritamente crescentes ($\forall x, y \in A, (x > y \Rightarrow f(x) > f(y))$) ou decrescentes ($\forall x, y \in A, (x > y \Rightarrow f(x) < f(y))$).

Os coeficientes de correlação de Pearson e Spearman são medidas estatísticas da força de uma relação entre dados pareados e têm valores na faixa de -1 a 1 . O sinal de cada coeficiente indica o sentido da relação. Se ambas as variáveis tendem a aumentar ou diminuir juntas, o coeficiente é positivo; caso contrário, se uma variável aumenta à medida que a outra diminui, o coeficiente é negativo. Assim, quanto mais perto r ou r_s se tornam de 1 ou -1 , mais significativo é o relacionamento.

Dessa forma, o uso das correlações de Pearson e Spearman acrescenta uma análise relevante ao ranqueamento fornecido pelas centralidades analisadas na medida que avalia as correlações entre essas classificações e métricas propostas para os estudos de caso [Carneiro 2016, Hilary 2015].

A análise dos coeficientes de Pearson e Spearman deve também considerar a significância (valor p). Para determinar se a correlação entre as variáveis é significativa, comparando-se o seu valor com seu nível de significância. Em geral, um nível de significância (denotado como α) de $0,05$ é aceitável, indicando que o risco de concluir que existe uma correlação, quando na verdade não existe correlação, é de 5% . O valor p indica se o coeficiente de correlação é significativamente diferente de 0 . (Um coeficiente de 0 indica que não há relacionamento linear). Se o valor p -value $\leq \alpha$, a correlação é estatisticamente significativa.

3.3. Métrica de Avaliação Proposta

Dada a extensão da base de dados, considerou-se como relevantes as licitações com valores acima de R\$ 1mi. A modelagem proposta atribui aos nós das redes as empresas participantes dos processos licitatórios e a sua participação conjunta na mesma licitação como sendo as relações entre essas empresas. A quantidade de licitações em comum foi utilizada como parâmetro para o estabelecimento do peso (*weight*) da relação. Dessa forma, uma empresa i que tenha participado de licitações onde a empresa j também tenha participado terá o tensor A_{ij}^α representando essa participação em comum, onde α representa a rede complexa.

No processo de classificação foram utilizadas as centralidades clássicas *Betweenness* (BW) [Freeman 1978, Otte and Rousseau 2002], *Eigenvector centrality* (EV) [Bonacich 1972], *PageRank* (PR) [Bonacich 2007] e *Weighted Degree* (WD) [Beveridge and Shan 2016]; na tentativa de indicar quais as empresas mais relevantes no cenário. Após essa classificação foram tomadas as 100 empresas mais bem classificadas em cada centralidade. Para essas empresas foi feita uma ampla pesquisa de escândalos associados a fraudes que as envolvesse. Essa pesquisa considerou sites governamentais, tribunais de contas, diários oficiais, notícias em sites de agências de notícias, entre outros. Dessa forma, se dividiu o universo de empresas em dois grupos, as que não tinham escândalos associados e as que foram citadas como partícipes de algum escândalo envolvendo fraude. Foi então estabelecido um “ponto de corte” onde as 50 empresas mais bem classificadas em cada centralidade foram consideradas como sendo as mais relevantes.

Considerando, como premissa, para que uma empresa representada, por um nó i , seja considerada como válida ter a sua presença em, pelo menos, uma licitação. Tem-se

que $\sum A_{ij}^\alpha > 0$, onde α representa a rede complexa que modela o cenário das licitações analisadas. Dessa forma, tem-se quatro conjuntos disjuntos:

$$VP = \{i \mid C_i \leq 50 \wedge i \in E\}, \quad (1)$$

$$FP = \{i \mid C_i \leq 50 \wedge i \notin E\}, \quad (2)$$

$$FN = \{i \mid C_i > 50 \wedge i \in E\}, \quad (3)$$

$$VN = \{i \mid C_i > 50 \wedge i \notin E\}, \quad (4)$$

onde E é o conjunto das empresas envolvidas em escândalos e C_i é a classificação do nó i na centralidade C .

Dessa forma, foi possível a formatação da matriz de confusão (Tabela 1) para o cálculo da precisão, revocação e acurácia. Estabelecendo critérios para a determinação dos *verdadeiros positivos* - VP, *verdadeiros negativos* - VN, *falsos positivos* - FP e *falsos negativos* - FN. As colunas consideram o envolvimento das empresas em escândalos ou não.

Tabela 1. Matriz de confusão utilizada nas centralidades clássicas

	Envolvidas	Não Envolvidas
Classificação ≤ 50	VP	FP
Classificação > 50	FN	VN

Um outro critério utilizado foi utilizar as correlações de Pearson e Spearman para determinar a força da relação entre as classificações e o valor das licitações. Dessa forma, se à medida em que as empresas são melhor classificadas pelas centralidades também cresça o valor auferido por elas nas licitações, e sendo essas empresas acusadas de envolvimento em ações ilegais, tem-se a constatação de que a metodologia proposta aponta com precisão, idealmente linear, as empresas fraudulentas em ordem de relevância.

4. Resultados e Discussão

4.1. Análise da Rede Gerada

A rede complexa gerada (Figura 2) é composta de quatro sub-redes, sendo duas completamente separadas (1 e 4) e duas com poucas relações entre si (2 e 3). Isso demonstra a segmentação das empresas e licitações por área de atuação, o que é esperado.

Verifica-se que as sub-redes 1, 2 e a parte em azul da sub-rede 3, têm formatos semelhantes. O mesmo ocorre entre a parte em verde da sub-rede 3 e a sub-rede 4. Em todos os casos encontra-se um número significativo de nós periféricos, nas bordas das sub-redes, e alguns nós centrais. Foi verificado que esses nós centrais foram os maiores

vencedores nas licitações, sempre acompanhados dos nós periféricos em um esquema de revezamento.

Essa análise não pode concluir, de forma tácita, a existência de um esquema de conluio, porém se habilita a ser um indício da presença de uma associação indébita, que pode ser objeto de uma auditoria. Assim se conclui que a simples análise visual das redes já se constitui em indícios de formação de cartéis.

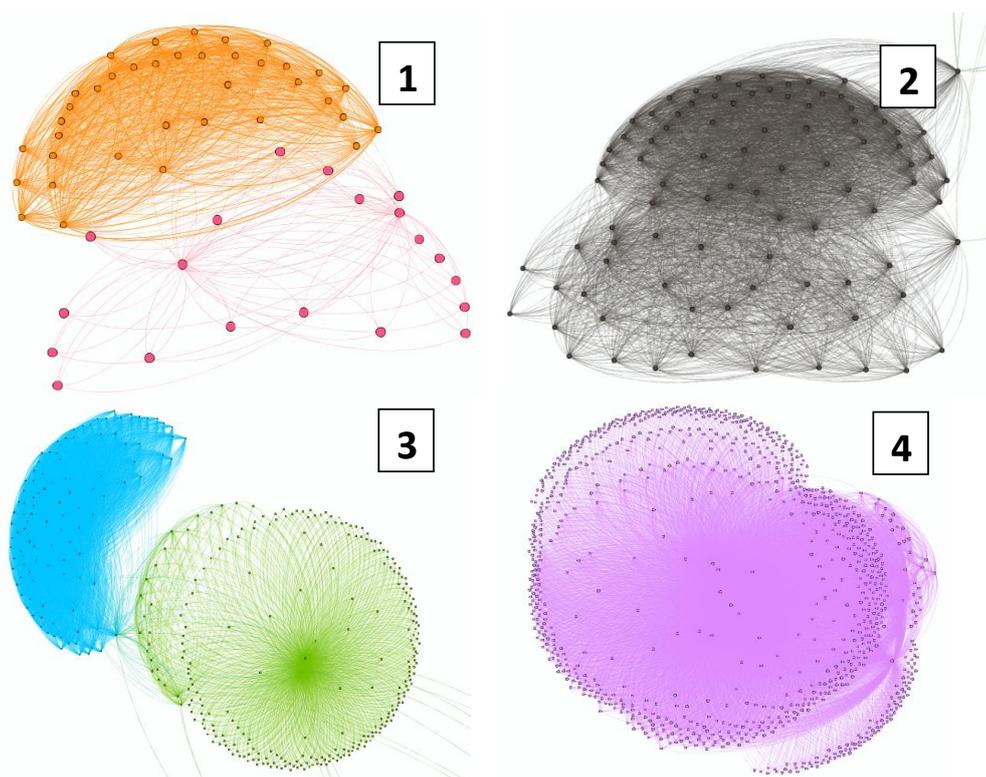


Figura 2. Rede Complexa Gerada.

4.2. Análise Estatística dos Resultados

Quanto à aferição da Precisão, Revocação, Acurácia e F_1 Score, verifica-se que a *Eigenvector centrality* (EV) [Bonacich 1972] obteve os melhores resultados, atingido uma precisão de superior a 70%, além de uma revocação de 60%, um F_1 Score de 65% e uma acurácia de 68%.

No estudo de caso, o uso da análise estatística é uma forma de demonstrar a coerência do uso da distribuição por grupos para classificar os nós de acordo com sua suposta relevância. Pretende-se demonstrar a correlação entre os valores alcançados pelas empresas e as classificações obtidas pelo uso das centralidades. A força dessa relação é capaz de demonstrar a coerência da classificação. Sempre considerando que as empresas utilizadas nessa análise tiveram seus nomes envolvidos ou, no mínimo, foram suspeitas de atividades ilegais.

Uma vez que a correlação proposta por Spearman [Spearman 1904] trata de funções monotônicas ela é aplicável ao estudo de caso, por se tratar de uma função monotônica de crescimento estrito, ou seja, $\forall x, y \in A, (x > y \Rightarrow f(x) \geq f(y))$, onde x

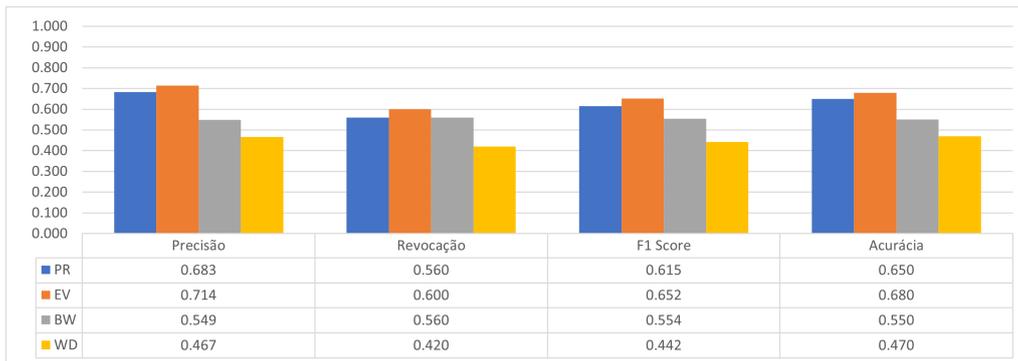


Figura 3. Análise consolidada de Accuracy, Precision, Recall e F_1 Score.

pode representar o faturamento das empresas e y o ranqueamento fornecido pelas centralidades analisadas. Dessa forma, o coeficiente de correlação de Spearman é capaz analisar a intensidade e a direção dessa relação monotônica. Ressalta-se que em um relacionamento monotônico, as variáveis tendem a se mover na mesma direção relativa, mas não necessariamente a uma taxa linear.

Como os coeficientes de correlação de Pearson e Spearman são medidas estatísticas da força de uma relação entre dados pareados e, considerando que, nesse estudo de caso, lida-se com duas variáveis que deveriam, na situação ideal, ter uma correlação linear e monotônica, ou seja, à medida que uma cresce, a outra também deveria crescer linear e continuamente, utiliza-se ambos os coeficientes para avaliar essas relações.

A tabela 2 traz os valores das correlações de Pearson e Spearman para as quatro centralidades analisadas. Verifica-se que as centralidades WD e EV obtiveram os melhores resultados, sendo próximos a -1. É interessante ressaltar que os resultados negativos nas correlações é o esperado uma vez que à medida que os valores de faturamento crescem as classificações tendem a decrescer, isso por que as classificações menores indicam justamente as empresas mais importantes no contexto. Isso significa que elas se comportaram na direção esperada; ou seja, eles tiveram uma concentração de valores inversamente proporcional ao crescimento do faturamento.

Considerando que a análise dos coeficientes de Pearson e Spearman deve considerar a significância (valor p) para validar a análise das correlações, verifica-se que, no estudo de caso, a correlação de Pearson atingiu um p -value <0.001 para as centralidades WD e EV; ou seja, isso indica que a correlação é estatisticamente significativa. A correlação de Spearman, com p -value <0.001 , para a WD, e p -value $=0.006$, no caso da EV, também é estatisticamente significativa.

A Figura 4 traz os coeficientes de correlação de Pearson e Spearman para cada centralidade analisada. Verifica-se, visualmente, a coerência do comportamento entre o faturamento e as classificações fornecidas pelas centralidades WD e —EV. Dessa forma, fica claro que as duas centralidades são as que melhor representaram a correlação entre o faturamento e o ranqueamento.

4.3. Aplicabilidade

A aplicabilidade do uso de medidas de centralidade na classificação dos nós mais relevantes de uma rede complexa se estende por uma ampla gama de aplicações. Ou seja, sempre

Tabela 2. Análise de correlações de Pearson e Spearman.

Centralidade	Pearson		Spearman	
	r	p -value	r_s	p -value
WD	-0.890	<0.001	-0.952	<0.001
BW	-0.067	0.566	0.111	0.342
PR	-0.172	0.139	-0.083	0.478
EV	-0.505	<0.001	-0.314	0.006

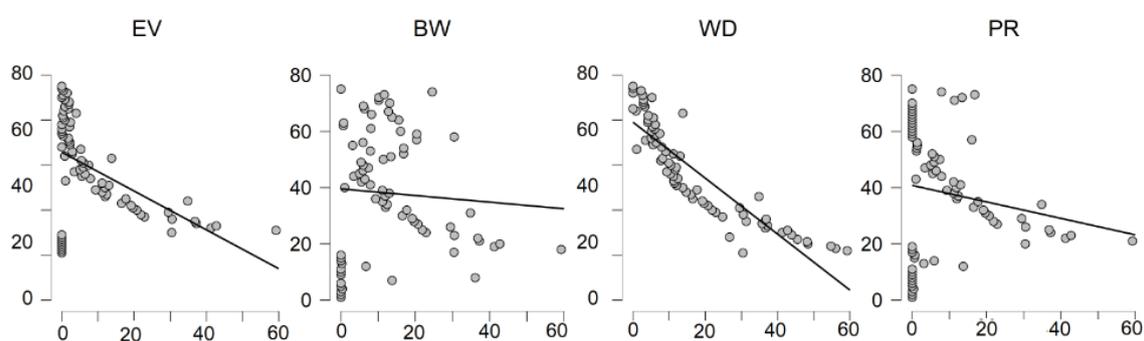


Figura 4. Análise Gráfica das Correlações de Pearson e Spearman.

que seja importante descobrir quais são as entidades mais relevantes num cenário o uso dessa tecnologia se mostra aplicável.

O uso de medidas de centralidade na descoberta de empresas fraudulentas tem uma aplicação direta no processo de auditoria desenvolvido pelos Tribunais de Contas. Nesses cenários é comum que um número reduzido de auditores se depare com um número significativo de empresas suspeitas de cometimento de fraude, sendo inviável que o processo de auditoria desenvolva sobre o total das empresas. Dessa forma, um instrumento que indique quais auditar com maiores chances de sucesso uma ferramenta útil na otimização dos recursos de auditoria, viabilizando, muitas vezes a própria auditoria [Figueiredo et al. 2020].

Um outro exemplo é a descoberta de indivíduos suspeitos com maior chance de ter cometido crime. Na análise de depoimentos da Operação Lava Jato, o uso de medidas de centralidade demonstrou ser capaz de apontar esses indivíduos com uma precisão de 90% [Almeida et al. 2017, De Figueirêdo et al. 2021]. Logo, o Ministério Público e a Polícia têm a sua disposição uma ferramenta que pode ajudar na precisão e minorar o uso de recursos da investigação.

5. Conclusão

Este artigo propõe o uso de redes complexas e medidas de centralidade na detecção de esquemas de fraude, por meio de conluio, entre empresas participantes de licitações públicas. Os resultados obtidos demonstraram ser viável essa abordagem. Lembrando que, a simples detecção dessas empresas não implica na sua efetiva participação em es-

quem fraudulentos, sendo apenas um indicativo de que há maiores chances dessas empresas estarem envolvidas numa associação ilegal. A apuração e confirmação dos fatos deverá ser obrigatoriamente feita por meio de um processo de auditoria. Dessa forma, o que se propõe neste artigo trata-se de uma ferramenta de auxílio ao processo de auditoria, normalmente realizado, no âmbito governamental, pelos Tribunais de Contas, Controladorias e Ministério Público.

A análise gráfica das redes (2) foi capaz de apontar as empresas mais importantes em cinco contextos distintos, sendo essas empresas suspeitas de participação em esquemas de conluio. Dessa forma, verificou-se que a utilização da tecnologia das redes complexas é uma alternativa viável na detecção dessas empresas.

O uso das correlações de Pearson e Spearman foi capaz de confirmar a existência de uma relação inversamente proporcional entre os ganhos das empresas e as classificações fornecidas pelas centralidades *Eigenvector centrality* (EV) e *Weighted Degree* (WD). Ambas as correlações têm significância estatística e demonstram que o uso de centralidades pode indicar as empresas com maiores ganhos (Tabela 2 e Figura 4).

A *Eigenvector centrality* (EV) foi obtida uma precisão de superior a 71%, uma revocação de 60%, um F_1 Score de 65% e uma acurácia de 68% na detecção de empresas participantes de esquemas de conluio. Esse resultado demonstra ser viável o uso dessa centralidade como instrumento de auxílio ao processo de auditoria de órgãos governamentais, indicando quais as empresas com maior probabilidade de estar participando de um esquema de conluio.

Quanto às limitações do trabalho o uso de uma quantidade maior de medidas de centralidade, inclusive multi camadas, seria capaz de proporcionar resultados mais precisos. Além disso, a falta de uma base de dados oficial de empresas fraudulentas leva à necessidade de uma pesquisa mais árdua e menos precisa de empresas envolvidas em esquemas de conluio.

Como proposta de trabalho futuro tem-se o uso das classificações fornecidas pelas centralidades clássicas formatando uma rede multiplex para o uso de uma centralidade multicamadas, como e.g. a GCMN [De Figueirêdo et al. 2021], na tentativa de obtenção de resultados superiores quanto à precisão, revocação e acurácia.

Os dados suplementares, bancos de dados, redes complexas, scripts e imagens do artigo estão disponíveis em <https://data.mendeley.com/datasets/tyjmr9tyb4/draft?a=c4e27d01-3328-4466-bbf6-c947bb45984c>.

Referências

- Almeida, T., Nakamura, F. G., and Nakamura, E. F. (2017). Uma abordagem baseada em redes complexas para análise de depoimentos legais. In *XXXVII Congresso da Sociedade Brasileira de Computação*, pages 2482–2491.
- Arnold, K., Gosling, J., and Holmes, D. (2000). *The Java Programming Language*. Addison Wesley Publishing Company.
- Bavelas, A. (1950). Communication patterns in task-oriented groups. *J Acoust Soc Am*.
- Beveridge, A. and Shan, J. (2016). Network of thrones. *Math Horizons*.

- Bianconi, G. (2018). *Multilayer Networks: Structure and Function*. Oxford University Press.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *J Math Sociol.*
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Soc Networks*.
- Carneiro, M. G. (2016). *Redes complexas para classificação de dados via conformidade de padrão, caracterização de importância e otimização estrutural*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- Costa, C. and Aparicio, M. (2011). Using data mining to help auditors. In *17th International Business Information Management Association Conference, IBIMA 2011; Milan; Italy; 14 November 2011 through 15 November 2011; Code 106712*, volume 4, pages 1864–1868.
- Cunha, R. and Bugarin, M. S. (2014). Lei de benford e auditoria de obras públicas: Uma análise de sobrepreço na reforma do maracanã. *Revista TCU*, pages 48–53.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S., and Arenas, A. (2013). Mathematical formulation of multilayer networks. *Phys. Rev. X*, 3:041022.
- De Domenico, M., Sole-Ribalta, A., Gomez, S., and Arenas, A. (2014). Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*, 111(23):8351–8356.
- De Figueirêdo, B. C. B., Nakamura, F. G., and Nakamura, E. F. (2021). A group-based centrality for undirected multiplex networks: A case study of the brazilian car wash operation. *IEEE Access*, 9:81946–81956.
- 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.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Soc Networks*.
- Ghedini Ralha, C. and Sarmiento Silva, C. V. (2012). A multi-agent data mining system for cartel detection in brazilian government procurement. *Expert Systems with Applications*, 39(14):11642–11656.
- Hilary (2015). *Centrality measures in multilayer networks*. PhD thesis, University of Oxford.
- Mankiw, N. G. and Taylor, M. P. (2011). *Principles of Economics*. 2th edition.
- Otte, E. and Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *J Inf Sci*.

- Pearson, K. (1905). On the general theory of skew correlation and non-linear regression. *Dulau and Co.*
- Scabini, L. F., Ribas, L. C., Neiva, M. B., Junior, A. G., Farfán, A. J., and Bruno, O. M. (2021). Social interaction layers in complex networks for the dynamical epidemic modeling of covid-19 in brazil. *Physica A: Statistical Mechanics and its Applications*, 564:125498.
- Silva, C. V. S. and Ralha, C. G. (2010). Utilização de técnicas de mineração de dados como auxílio na detecção de cartéis em licitações. pages 1–14. XXX Congresso Da Sociedade Brasileira de Computação.
- Solé-Ribalta, A., De Domenico, M., Gómez, S., and Arenas, A. (2016). Random walk centrality in interconnected multilayer networks. *Physica D: Nonlinear Phenomena*, 323-324:73–79. Nonlinear Dynamics on Interconnected Networks.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.
- Tu, X., Jiang, G., Song, Y., and Zhang, X. (2018). Novel multiplex pagerank in multilayer networks. *IEEE Access*, 6:12530–12538.