

Alertas de fraude em licitações: Uma abordagem baseada em redes sociais

Lucas L. Costa¹, Arthur P. G. Reis¹, Clara A. Bacha¹,
Gabriel P. Oliveira¹, Mariana O. Silva¹, Matheus C. Teixeira¹,
Michele A. Brandão^{1,2}, Anísio Lacerda¹, Gisele L. Pappa¹

¹Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG, Brasil

²Instituto Federal de Minas Gerais (IFMG) – Ribeirão das Neves, MG, Brasil

{lucas-lage, arthurpetrocchi, clarabacha}@ufmg.br
{gabrielpoliveira, mariana.santos, matheus.candido}@dcc.ufmg.br
michele.brandao@ifmg.edu.br, {anisio, glpappa}@dcc.ufmg.br

Abstract. *In Brazil, public bids must guarantee transparency and free competition between bidders. However, monitoring irregularities is complex because it involves a huge volume of data and a small number of specialists. In this context, this work proposes the use of a methodology based on the concepts of audit trails and social networks to raise fraud alerts, to assist in the fight against corruption. The characterization and analysis of a real social network, associated with a case study of a possible fraudulent bid, reveal that the methodology presented is able to identify suspicious bids, identified by a set of audit trails.*

Resumo. *No Brasil, licitações públicas devem garantir a transparência e a livre concorrência entre licitantes. No entanto, o monitoramento de irregularidades é complexo por envolver um enorme volume de dados e uma quantidade reduzida de especialistas. Nesse contexto, este trabalho propõe o uso de uma metodologia baseada em conceitos de trilhas de auditoria e redes sociais para criar alertas de fraude em licitações, de forma a auxiliar no combate à corrupção. A caracterização e a análise de uma rede social real, associada a um estudo de caso com uma possível licitação fraudulenta, revelam que a metodologia apresentada consegue identificar licitações suspeitas, em geral, identificadas por mais de uma trilha de auditoria.*

1. Introdução

Uma licitação é definida como um conjunto de procedimentos por meio do qual a Administração Pública realiza compras ou contratações de produtos e serviços. Ela possui três objetivos principais: selecionar a proposta mais vantajosa para a Administração Pública, garantir igualdade de condições a todos que queiram contratar com o Poder Público e promover o desenvolvimento nacional sustentável.¹

No Brasil, embora a legislação imponha mecanismos para impedir fraudes em licitações públicas, escândalos de corrupção acabam acontecendo. Isso porque, geralmente, a checagem de licitações públicas é realizada através de um monitoramento em tempo real e de forma manual. Ou seja, todo o processo licitatório é acompanhado —

¹Lei nº 8.666/93: http://www.planalto.gov.br/ccivil_03/Leis/L8666cons.htm

desde a publicação do edital até que seja firmado e executado o contrato, e verificado o cumprimento do objeto licitado — por pessoas. Essa inspeção manual requer tempo e grande mobilização de recursos humanos. Assim, combinar perícia humana com sistemas computacionais é essencial.

Além disso, a detecção de fraudes em licitações é um processo complexo, uma vez que essas fraudes não são pontuais e isoladas, mas envolvem interações diretas, indiretas e até temporais entre as entidades envolvidas. Dentre essas entidades podemos citar os licitantes, representados por fornecedores, sejam eles pessoas físicas ou jurídicas, interessados em fornecer o objeto da licitação ao licitador, ou seja, um órgão público. Portanto, formas automáticas de monitoramento, análise e cruzamento dos dados à procura de infrações são necessárias para auxiliar na identificação de fraudes.

O principal objetivo deste trabalho é desenvolver uma metodologia capaz de identificar indícios de fraudes em licitações públicas por meio da combinação de conceitos de trilhas de auditoria e redes sociais. Segundo a lei, uma fraude é definida como o ato de “frustrar ou de fraudar, mediante ajuste, combinação ou qualquer outro expediente, o caráter competitivo do procedimento licitatório, com o intuito de obter, para si ou para outrem, vantagem decorrente da adjudicação do objeto da licitação”. Além disso, “a configuração da fraude à licitação não está associada ao seu resultado, ou seja, ao sucesso da empreitada”. Portanto, a fraude ocorre independentemente do fraudador vencer a licitação e causar prejuízo aos cofres públicos.

Para isso, a partir de um conjunto de dados de licitações públicas municipais e estaduais de Minas Gerais, foi definido e implementado um conjunto de trilhas de auditoria em parceria com especialistas do Ministério Público de Minas Gerais (MPMG). Uma trilha de auditoria é uma sequência de passos seguidos para identificar indícios de tipos específicos de irregularidades encontrados em fraudes em licitações. Essas trilhas ajudam a selecionar os dados de interesse das bases de dados de licitações.

Em seguida, esses dados são pré-processados e modelados como uma rede, que captura a interação entre os licitantes. Com a modelagem na forma de grafo, o arcabouço da análise de grafos pode ser empregado para a verificação de indícios de fraudes. Com os resultados obtidos, é possível tanto utilizar a metodologia para filtrar os dados, quanto reduzir o volume de dados a serem analisados por especialistas e também gerar subsídios para elaboração de algoritmos capazes de classificar uma licitação como fraude ou não.

Este trabalho está organizado da seguinte maneira. Os trabalhos relacionados são apresentados na Seção 2 e a base de dados utilizada no trabalho é descrita na Seção 3. A Seção 4 apresenta a metodologia adotada, descrevendo o processo de seleção e pré-processamento dos dados e modelagem do grafo. Já na Seção 5, são apresentadas várias caracterizações do grafo que podem ser usados como indicadores de ilicitude. Na Seção 6, é apresentado um estudo de caso de uma licitação fraudulenta. Por fim, na Seção 7, são apresentadas as conclusões e as principais limitações deste trabalho.

2. Trabalhos Relacionados

Diferentes abordagens na área de mineração de dados e inteligência artificial têm sido propostas no contexto geral da detecção automática de fraudes [Ganguly and Sadaoui 2018, Elshaar and Sadaoui 2020]. Por exemplo, Velasco et al. propõem uma ferramenta que incorpora algoritmos de mineração de dados para quantificar padrões de risco de corrupção,

levando a melhorias na qualidade dos gastos públicos e à identificação de casos de fraude [Velasco et al. 2021]. Com uma abordagem baseada em um algoritmo de aprendizado de máquina supervisionado, Anowar and Sadaoui desenvolveram um classificador de fraude eficiente que permite distinguir entre licitantes legítimos e não legítimos licitantes [Anowar and Sadaoui 2019]. Além de classificadores clássicos, alguns estudos exploraram técnicas mais robustas baseadas em redes neurais para tratar a tarefa de detecção de fraudes [Pereira and Murai 2021, Abidi et al. 2021].

Apesar de serem eficazes, a maioria das abordagens citadas anteriormente necessitam de dados rotulados para o treinamento ou de técnicas mais complexas de mineração de dados. Como alternativas mais simples e eficientes, abordagens baseadas em redes sociais vêm sendo propostas para a tarefa de detecção de fraudes [Araújo et al. 2021]. Vlasselaeer et al. apresentam uma nova abordagem que usa inferência ativa através de um algoritmo baseado em rede para melhor detectar fraudes em redes sociais que variam no tempo [Vlasselaeer et al. 2015]. Em outro estudo, os autores introduzem uma nova abordagem que define e extrai recursos de uma rede ponderada no tempo, explorando e integrando recursos intrínsecos na detecção de fraudes [Vlasselaeer et al. 2015].

Independente da abordagem utilizada, ainda são poucos os estudos tratando especificamente da análise ou detecção de fraude em licitações públicas [Ralha and Silva 2012, Grace et al. 2016, Andrade et al. 2020, Lima et al. 2020], provavelmente, devido à falta de dados adequados para apoiar tal tarefa. Nesse contexto, este trabalho busca preencher as lacunas de pesquisa existentes ao analisar alertas de fraudes em licitações no estado (e nos municípios) de Minas Gerais através da análise da rede social construída a partir de trilhas de auditoria.

3. Base de Dados

Neste trabalho, foram utilizados os dados fornecidos pelo Ministério Público de Minas Gerais (MPMG) por meio do Programa de Capacidades Analíticas. Esses dados são sigilosos e, por isso, pouco detalhados. Os dados foram disponibilizados em um Sistema de Gerenciamento de Bancos de Dados (SGBD), que possui informações de licitações municipais e estaduais, bem como de seus licitantes e sócios.

Para as licitações, foram utilizados dados públicos. As municipais são provenientes do Sistema Informatizado de Contas dos Municípios (SICOM) do TCE-MG², enquanto as estaduais vieram dos dados do Portal da Transparência do TCE-MG³⁴. Também foi considerada a base de dados do CEIS (Cadastro de Empresas Inidôneas e Suspensas),⁵ que apresenta empresas com sanções vigentes em todo o país. Além dos dados públicos, foram utilizados dados sigilosos advindos do SERPRO (Serviço Federal de Processamento de Dados) para enriquecimento das informações das empresas licitantes.

Ao todo, foram consideradas 14.565 licitações estaduais e 363.572 licitações municipais. As licitações estão divididas entre 17 modalidades, incluindo concorrência, convite, concurso, pregão, entre outras. Por fim, nosso conjunto de dados possui informações de 103.858 empresas licitantes distintas que participaram de tais processos licitatórios.

²<http://dadosabertos.tce.mg.gov.br/>

³<https://www.transparencia.mg.gov.br/compras-e-patrimonio/compras-e-contratos>

⁴<https://fiscalizandocomtce.tce.mg.gov.br/#/inicio>

⁵<https://www.portaldatransparencia.gov.br/sancoes/ceis>



Figura 1. Metodologia para alerta de fraudes em licitações.

Tabela 1. Definição das trilhas de auditoria a serem aplicadas nas licitações.

#	Trilha de Auditoria	Regra
<i>Trilhas de Licitante (Nós)</i>		
1	Licitante licitando antes de registro	Verificar as licitações que contenham participantes que estão licitando antes da empresa iniciar suas atividades
2	Licitante licitando com sanção ativa	Verificar as licitações que contenham licitantes com alguma sanção ativa na base do CEIS (Cadastro de Empresas Inidôneas e Suspensas)
3	Licitante com CNPJ inativo	Verificar licitações com a presença de licitantes com o CNPJ inativo
4	Licitante perdedor frequente	Identificar licitações com a presença de licitantes com alto percentual de derrotas
5	Licitante vencedor frequente	Identificar licitações com a presença de licitantes com alto percentual de vitórias
6	Licitante único	Identificar licitações com licitante único
<i>Trilhas de Vínculo (Arestas)</i>		
7	Licitantes com sócios em comum	Identificar as licitações com licitantes distintos que possuem pelo menos um sócio em comum
8	Licitantes com e-mails em comum	Verificar se a licitação possui licitantes distintos com e-mails em comum
9	Licitantes com telefones em comum	Verificar se a licitação possui licitantes distintos com número de telefone em comum
10	Licitantes com endereços em comum	Identificar as licitações com licitantes distintos com endereço em comum

Os dados incluem informações relevantes para detecção de vínculo em comum, incluindo endereço, telefones, e-mail e sócios.

4. Metodologia para Identificação de Fraudes

Esta seção apresenta a metodologia para identificação de possíveis fraudes em licitações. As principais etapas dessa metodologia são baseadas na construção de trilhas de auditoria, cujo conceito foi discutido na Seção 1. Essas etapas são apresentadas na Figura 1 e incluem: a filtragem dos dados, o pré-processamento dos dados filtrados, a modelagem de uma rede social a partir de tais dados e a análise dessa rede para levantamento de alertas em fraudes de licitação. É importante destacar que as etapas da metodologia foram realizadas tendo como base a definição de dez trilhas de auditoria, apresentadas na Tabela 1. As trilhas 1 a 6 pretendem investigar se há alguma irregularidade levando em conta alguma características do licitante, portanto são denominadas de *Trilhas de Licitante (Nós)*. Enquanto que as trilhas 7 a 10 investigam se há irregularidade analisando os vínculos entre os licitantes, portanto são *Trilhas de Vínculo (Arestas)*.

4.1. Filtragem dos Dados

Para realizar a tarefa de identificação de possíveis fraudes em licitações, alguns filtros foram criados. Esses filtros consideraram as trilhas de auditoria descritas na Tabela 1, e as entidades de interesse envolvidas.

Conforme descrito na Seção 3, os dados de licitações utilizados são dados públicos, portanto os licitantes pessoa física tiveram seus CPFs mascarados. Isso impossibilita qualquer processamento sobre esses dados, então eles foram desconsiderados da análise. Sendo assim, consideramos 3 tipos de entidades de interesse: licitações, licitantes pessoa jurídicas (Empresas) e seus sócios (Pessoas).

Após a análise dos dados disponíveis e das trilhas de interesse, foram elaboradas consultas para obter informações sobre licitações, empresas licitantes e seus sócios de forma que possam auxiliar na identificação de possíveis fraudes em licitações.

4.2. Pré-processamento dos Dados

A partir dos dados retornados pelas consultas e das trilhas de auditoria de interesse, dois tipos de pré-processamento foram executados: padronização de dados e construção de atributos referente as entidades. Em relação à padronização, foram padronizados os campos e-mail, endereço e telefone, que passarão a fazer parte da entidade licitante. Para o campo e-mail, verificou-se se o formado era válido (**conter xx@xx.xxx.xx**) e foram convertidos em letras minúsculas.

Para a padronização do campo de endereço, utilizou-se a função *soundex* do SQL (*Structured Query Language*) que converte uma string em um código de quatro caracteres baseando-se no som da cadeia de caracteres no idioma inglês. Para os números de telefone, removeu-se a máscara e a formatação dos valores. Em seguida, foram retirados todos os caracteres não numéricos e verificado o tamanho do número de telefone resultante, sendo um tamanho entre 9 e 11 caracteres considerado válido.

4.3. Modelagem da Rede Social

A modelagem do problema de geração de alertas de fraudes em licitações foi feita a partir de uma rede social representada por um grafo G , que possui dois tipos de nós: as empresas licitantes (V) e os sócios dessas empresas (\bar{V}). Há duas tipos de arestas: entre nós do tipo empresa que participaram de uma mesma licitação (E) e entre nós empresa e pessoa (\bar{E}), onde a pessoa é uma sócia da empresa. No caso de nós entre empresas, uma função $f(m, n)$ determina o peso da aresta de acordo com a quantidade de vínculos que uma empresa licitante tem com outra. Os vínculos levados em conta foram sócios, e-mail, telefones ou endereço em comum, i.e., o domínio de $f(m, n)$ está no intervalo $[1, 4]$. No caso de relações entre empresas e pessoas, a relação só existe quando a pessoa é sócia da empresa. Formalmente:

$$G = (\{V \cup \bar{V}\}, \{E \cup \bar{E}\}), \text{ onde}$$

$$V = \{v \mid v \in \text{Empresas}\}, \bar{V} = \{\bar{v} \mid \bar{v} \in \text{Pessoas}\},$$

$$E = \{(m, n) \mid ((m, n) \in V^2) \Leftrightarrow f(m, n) > 0\},$$

$$f(m, n) = \text{número de vínculos em comum de } m \text{ com } n, \text{ onde}$$

$$\text{vínculo} \in \{\text{sócio}, \text{endereço}, \text{telefone}, \text{email}\}$$

$$\bar{E} = \{(m, \bar{o}) \mid (m \in V \wedge \bar{o} \in \bar{V}) \Leftrightarrow \bar{o} \text{ é sócio de } m\}$$

Além disso, os nós da rede social proposta apresentam atributos, utilizados para construir as arestas do grafo e definir as regras que precisam ser seguidas pelas trilhas de auditoria (que serão apresentadas na Seção 4.4). Para o licitante (nó do tipo empresa),

temos a data de registro da empresa (*data_registro*), CNPJ, *status* da empresa na Receita Federal (ex: ativa ou não), *endereço*, *e-mail*, *telefone*, *sócios* e *sanções*. Para o sócio (no pessoa), temos o CPF e o tipo de *vínculo* com a empresa (ex: sócio, representante legal).

Dado o grafo G , uma licitação é representada pelo grafo $\hat{G}_i = (\{\hat{V}_i \cup \hat{V}_i\}, \{\hat{E}_i \cup \hat{E}_i\})$, onde $\hat{G}_i \subset G$ e $\hat{V}_i = \{v \mid v \in \text{Empresas que participaram da licitação } i\}$, $\hat{V}_i = \{\bar{v} \mid \bar{v} \in \text{Pessoas sócias das empresas que participaram da licitação } i\}$. As arestas que refletem os vínculos entre as empresas licitantes de \hat{G}_i são representadas pelo conjunto $\hat{E}_i = \{(m, n) \in \text{Arestas entre empresas participantes da licitação } i\}$. Já as arestas empresa-pessoa da licitação \hat{G}_i são representadas pelo conjunto $\hat{E}_i = \{(m, \bar{o}) \in \text{Arestas entre empresas da licitação } \hat{G}_i \text{ e pessoas sócias dessas empresas}\}$.

Além disso, temos a data de cada licitação e construímos dois novos atributos: a lista de licitantes perdedores e vencedores frequentes. Para a identificação de empresas vencedoras frequentes, foram analisadas empresas que, entre os anos de 2014 e 2021, somaram um valor total homologado maior ou igual a R\$400.000,00 (quatrocentos mil reais). Essas empresas também deveriam ter participado de pelo menos dois processos licitatórios com mais de uma empresa participante nas modalidades pregão eletrônico, pregão presencial ou dispensa de licitação por valor. Para uma empresa ser considerada vencedora de uma licitação, é necessário que ela tenha vencido pelo menos 70% do valor total desta, sendo considerada uma vencedora frequente quando houver uma taxa de vitórias de pelo menos 70% das licitações participadas.

Da mesma forma, para a identificação de empresas perdedoras frequentes, foram analisadas empresas que, entre os anos de 2014 e 2021, tenham perdido pelo menos 70% das licitações com uma participação mínima de pelo menos seis licitações por ano. As modalidades dessas licitações são pregão eletrônico, pregão presencial ou dispensa de licitação por valor, resultando em um mínimo de 42 participações como licitante. Para que uma empresa seja considerada perdedora em uma licitação, ela não pode ter vencido em nenhum item.

Por fim, a Figura 2 ilustra e exemplifica a modelagem descrita nessa seção, nela é possível observar todos os conjuntos e sub-conjuntos aqui apresentados.

4.4. Trilhas de Auditoria para Alertas de Fraudes

O conceito de trilha de auditoria, definido na Seção 1, pode ser entendido, dentro do contexto da rede social proposta, como uma série de regras (x) para identificar alertas de fraude em licitações. Dessa forma, uma trilha pode ser modelada como um conjunto de licitações que descumpriram essas regras. Portanto, o conjunto $T_x = (\hat{G}_1, \hat{G}_2, \dots, \hat{G}_j)$, é composto pelas licitações que descumpriram as regras x . Ou seja, a trilha é um conjunto de licitações, portanto é um conjunto de subgrafos da rede social.

A Tabela 2 descreve as regras que $\hat{G}_i = (\{\hat{V}_i \cup \hat{V}_i\}, \{\hat{E}_i \cup \hat{E}_i\})$ precisa atender para que \hat{G}_i seja enquadrada em cada uma das dez trilhas de auditoria apresentadas na Tabela 1.

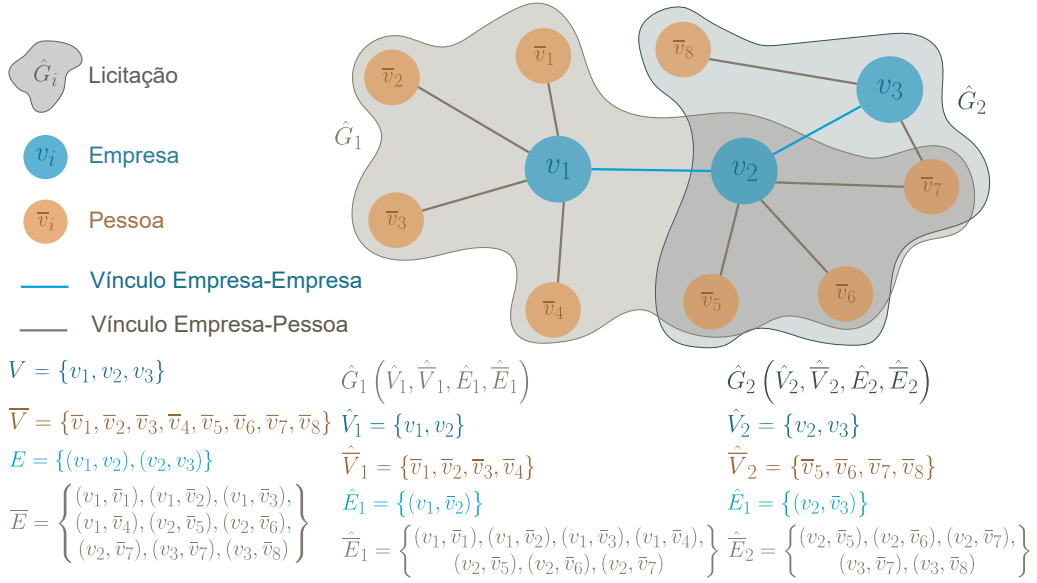


Figura 2. Rede social heterogênea entre empresas licitantes e seus sócios.

Tabela 2. Regras que $\hat{G}_i = (\{\hat{V}_i \cup \hat{V}_i\}, \{\hat{E}_i \cup \hat{E}_i\})$ precisa atender para que \hat{G}_i seja enquadrada em cada uma das trilhas de auditoria.

#	Definição Formal da Trilha de Auditoria
T_1	$\hat{V}_i = \{v \mid \exists v \rightarrow v[data_registro] < \hat{G}_i[data_licitacao]\}$
T_2	$\hat{V}_i = \{v \mid \exists v \rightarrow v[sanções] > 0\}$
T_3	$\hat{V}_i = \{v \mid \exists v \rightarrow e[status] \neq 'ATIVO'\}$
T_4	$\hat{V}_i = \{v \mid \exists v \in vencedores_frequentes\}$
T_5	$\hat{V}_i = \{v \mid \exists v \in perdedores_frequentes\}$
T_6	$ \hat{V}_i = 1$
T_7	$\hat{E}_i = \{(m, n) \in V^2 \mid \exists(m, n) \rightarrow m[socios] \cap n[socios] > 0\}$
T_8	$\hat{E}_i = \{(m, n) \in V^2 \mid \exists(m, n) \rightarrow m[email] = n[email]\}$
T_9	$\hat{E}_i = \{(m, n) \in V^2 \mid \exists(m, n) \rightarrow m[telephones] \cap n[telephones] > 0\}$
T_{10}	$\hat{E}_i = \{(m, n) \in V^2 \mid \exists(m, n) \rightarrow m[endereço] = n[endereço]\}$

5. Caracterização e Análise da Rede Social Real

Após modelar e construir a rede social entre licitantes e sócios, pode-se utilizar tal estrutura para analisar os relacionamentos entre os indivíduos e verificar possíveis alertas de fraude. Na Seção 5.1, é apresentada uma visão geral da rede a partir de métricas topológicas que a descrevem. Em seguida, a Seção 5.2 analisa a cocorrência de alertas de fraude através da análise de correlação entre as trilhas de auditoria.

5.1. Visão Geral da Rede Social

Nesta seção, são analisados os relacionamentos da rede social para aprofundar o entendimento da dinâmica das conexões entre empresas licitantes e seus sócios. A complexidade da rede já se revela em seu tamanho: são mais de 144 mil nós, entre empresas e pessoas, que possuem mais de 110 mil arestas entre si. Ilustrar tal rede de forma adequada é uma tarefa desafiadora. Portanto, a Figura 3 apresenta somente a visualização do componente gigante da rede, que compreende o maior conjunto conectado de nós. É importante res-

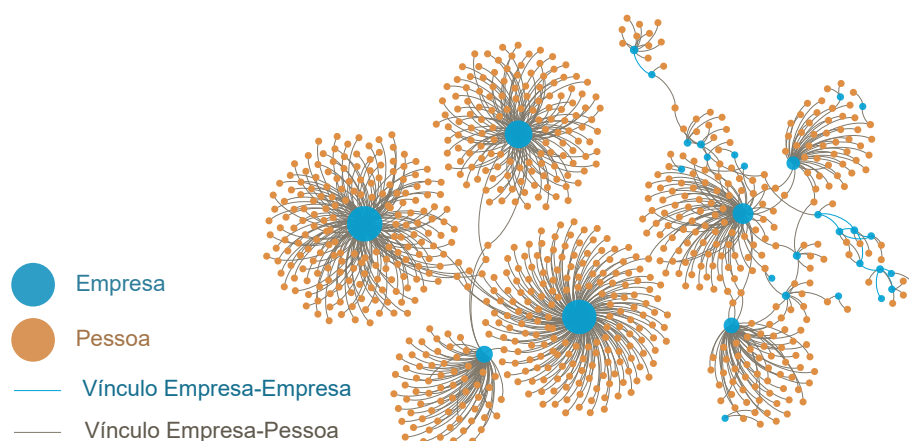


Figura 3. Componente gigante da rede social entre empresas licitantes e seus sócios. Empresas estão conectadas quando possuem vínculos em comum (e.g., endereço, sócios) e participam de um mesmo processo licitatório.

Tabela 3. Estatísticas gerais da rede social de licitantes e seus sócios. São apresentados os resultados tanto para a rede completa quanto para seu componente gigante, i.e., o maior componente conectado.

	Completa	C. Gigante		Completa	C. Gigante
Empresas (CNPJ)	54.310	32	Densidade (10^{-4})	0,106	23,690
Pessoas (CPF)	90.347	827	Grau médio	1,531	2,033
Nós (CNPJ e CPF)	144.657	859	CC médio (10^{-1})	0,226	0,008
Arestas	110.750	873	Componentes conectados	43.955	1

CC: Coeficiente de Clusterização

saltar a natureza heterogênea da rede social, que contém nós de diferentes tipos que se conectam em um mesmo contexto.

Em seguida, foram utilizadas métricas topológicas (i.e., obtidas a partir da própria estrutura da rede) para analisar as principais características da rede social: número de nós e arestas, densidade (razão entre a quantidade de arestas existentes e todas as arestas possíveis), grau médio e coeficiente de clusterização (mede a tendência dos nós de se conectarem uns aos outros). Conceitos e definições formais das métricas podem ser acessados em [Barabási 2016].

A Tabela 3 apresenta os resultados das métricas topológicas tanto para a rede completa quanto para seu componente gigante. O número de nós e arestas da rede completa revela a complexidade da rede construída. Como esperado, a maior parte dos nós são de pessoas físicas que são sócias de empresas licitantes, uma vez que empresas de médio e grande porte normalmente possuem vários representantes legais. Tal comportamento se repete no componente gigante, que contém 32 empresas e 827 sócios no total.

Ao analisar métricas como o número de arestas, coeficiente de clusterização médio e densidade da rede, percebe-se que a rede social construída é esparsa (valor da densidade próximo a zero), i.e., existem poucas arestas na rede comparadas à quantidade possível. Assim, o fato de relacionamentos na rede serem relativamente raros torna sua análise ainda mais relevante, pois eles podem evidenciar alertas de fraude mais facilmente, além de revelar possíveis esquemas de associações entre licitantes.

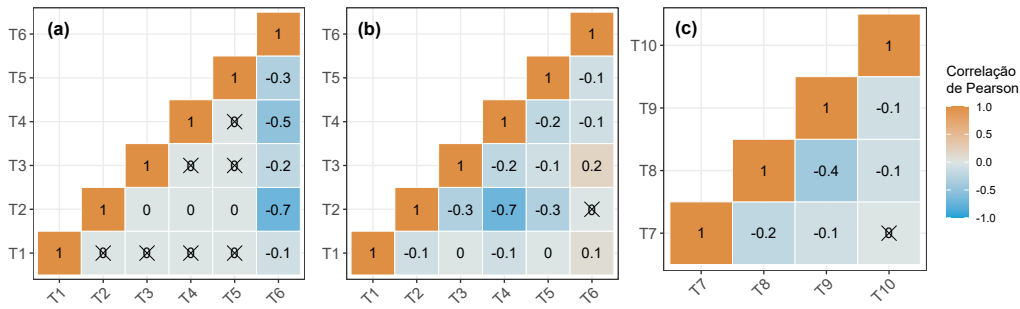


Figura 4. Correlação entre as trilhas relacionadas aos (a-b) licitantes e aos (c) vínculos. A matriz (b) é uma versão filtrada da (a), onde foram removidas as ocorrências em que a licitação se enquadra apenas na trilha T_6 . O símbolo \times indica uma correlação estatisticamente não significativa.

Por fim, o número de componentes conectados na rede completa revela uma alta fragmentação entre licitantes. Isso pode significar que as trilhas de auditoria não conseguiram revelar grandes grupos de licitantes que possuam vínculos em comum. Se a rede possuísse poucos componentes conectados, poderia-se imaginar uma grande associação entre empresas licitantes, necessitando uma posterior análise por parte de especialistas. Ainda assim, os relacionamentos dentro dos pequenos grupos existentes na rede devem ser analisados a fim de verificar os alertas de possíveis fraudes. Portanto, análises mais aprofundadas nos resultados das trilhas são necessárias.

5.2. Análise de Correlação entre as Trilhas de Auditoria

Nesta seção, realizamos uma análise de correlação entre as trilhas de auditoria, tanto relacionadas aos vínculos (arestas), quanto aos licitantes (nós). Para isso, geramos matrizes de correlação para avaliar a coocorrência entre todas as trilhas consideradas. Ao todo, foram avaliadas 11.343 licitações que se enquadram em pelo menos uma das trilhas descritas na Tabela 2. As Figuras 4(a) e 4(c) mostram as matrizes de correlação de Pearson entre as trilhas referentes aos licitantes e aos vínculos, respectivamente. A Figura 4(b) apresenta uma versão filtrada da matriz (a), onde foram removidos as licitações que se enquadram apenas na trilha T_6 , que investiga licitantes únicos em uma licitação.

Em relação às trilhas baseadas nos licitantes (i.e., trilhas T_1 à T_6), na Figura 4(a), a falta de correlação significativa entre as trilhas, exceto entre a trilha de licitante único (T_6), é causada pelo volume superior de licitações que se enquadram apenas nessa trilha. Comparando as Figuras 4(a) e 4(b), após a filtragem dessas licitações, fica evidente o impacto que a trilha T_6 provoca na análise de correlação. Com essa filtragem, podemos notar que a grande maioria das correlações é estatisticamente significativa e apresentam uma relação negativa fraca a moderada, variando entre -0,1 a -0,3. A única correlação forte (-0,7) identificada acontece entre as trilhas T_2 e T_4 , que analisam licitações que contenham licitantes com alguma sanção ativa na base do CEIS e licitantes com alto percentual de derrotas em licitações, respectivamente. Tal resultado é esperado, visto que licitantes cadastrados na base do CEIS não poderiam concorrer em licitações.

Analisando as trilhas referentes aos vínculos (i.e., T_7 a T_{10}), nota-se novamente que todas as correlações estatisticamente significativas apresentam uma relação negativa fraca a moderada, variando entre -0,1 a -0,4. De fato, a correlação mais evidente acontece

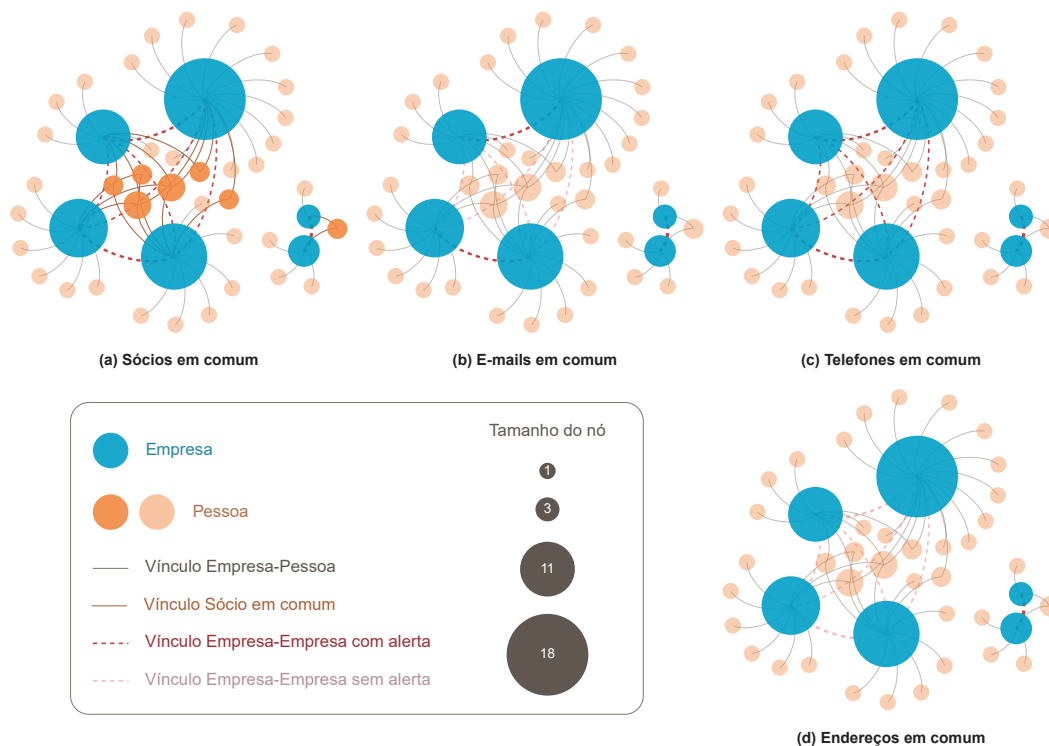


Figura 5. Sub-grafo da rede social real de uma licitação com visualizações diferentes por trilha. O tamanho do nó é baseado no grau do nó.

entre as trilhas T_8 e T_9 , que investigam licitações contendo licitantes com e-mails e telefones em comum, respectivamente. Tal resultado indica que grande parte das licitações enquadradas na trilha de endereços não se enquadram na de telefone, e vice-versa.

Por fim, a falta de correlação geral entre a maioria das trilhas (de vínculo e licitantes) revela a inexistência de um padrão entre os alertas, evidenciando o quão desafiador é a tarefa de análise de possíveis fraudes em licitações.

6. Estudo de Caso de uma Possível Licitação Fraudulenta

Esta seção apresenta um estudo de caso de uma licitação que foi identificada por quatro das trilhas de auditoria descritas neste artigo: T_7 , T_8 , T_9 e T_{10} . A Figura 5 apresenta diferentes perspectivas de visualização da rede social real dessa licitação de acordo com as trilhas em que ela foi identificada: T_7 : sócios em comum na Figura 5(a); T_8 : e-mails em comum na Figura 5(b); T_9 : telefones em comum na Figura 5(c); e T_{10} : endereços em comum na Figura 5(d). Essa rede é formada por um total de 44 nós, sendo seis nós de empresas (13,64%) e 38 nós de pessoas (86,36%). As arestas representam quatro tipos de vínculos: sócios em comum (sete arestas), e-mail (três arestas), telefone (sete arestas) e endereço (uma aresta). É possível perceber que, como a modelagem da rede é baseada nas empresas licitantes, elas tendem a possuir um maior grau na rede em relação às pessoas. Vale destacar que dezoito é o valor do nó de uma empresa com maior grau, e um é o valor do nó de uma empresa e/ou pessoa com menor grau. Neste estudo de caso, empresas com maior grau foram identificadas por mais trilhas de auditoria.

Em relação à trilha T_7 de sócios em comum, observa-se que seis empresas possuem pelo menos um sócio em comum, na cor laranja em destaque na Figura 5(a). Já para

a trilha T_8 , percebe-se três pares de empresas com vínculo Empresa-Empresa com alerta por terem o mesmo e-mail. Para a trilha T_9 , nota-se que quatro empresas, com maiores graus na rede, possuem telefone em comum entre elas e um par de empresas, com menor grau na rede, também possuem telefone em comum. Por fim, para a trilha T_{10} , apenas um par de empresas, com grau menor na rede, possuem o mesmo endereço. Portanto, essa licitação foi enquadrada nas quatro trilhas devido aos vínculos entre as duas empresas com os dois nós de menor grau. Isso levanta um alerta de fraude na licitação avaliada e, então, recomenda-se um estudo mais aprofundado dos dados referentes a essa licitação.

O estudo de caso revela que a identificação de licitações fraudulentas não é trivial, visto que é necessário a análise de múltiplos aspectos. Apesar da metodologia proposta facilitar na identificação de possíveis fraudes, o papel de especialistas de conteúdo na auditoria de licitações ainda é necessário. Entretanto, frente a tantas licitações, a metodologia permite reduzir os documentos que precisam ser avaliados por especialistas.

7. Conclusões e Trabalhos Futuros

Neste artigo, foram combinados conceitos de redes sociais com trilhas de auditoria para identificar indícios de licitações públicas fraudulentas. Especificamente, foi definido um conjunto de trilhas de auditoria capaz de verificar indicadores de fraudes, a partir de dados que incluem informações de licitações e os respectivos licitantes e sócios. Para isso, tais dados foram processados e modelados como uma rede social, que captura a interação entre os licitantes. Os resultados obtidos revelam ser possível utilizar a metodologia para filtrar os dados e reduzir o volume de dados a serem analisados por especialistas. Além disso, tais descobertas geram subsídio para elaboração de algoritmos capazes de classificar uma licitação como fraude ou não, de forma a auxiliar no combate à corrupção. Como trabalhos futuros, pretendemos implementar trilhas de auditoria adicionais, seguindo a metodologia proposta. Por exemplo, implementar uma trilha que detecta licitações onde os sócios das empresas licitantes são funcionários públicos (ou parentes) envolvidos no processo licitatório, visto que esse tipo de vinculação é um caso de fraude.

Limitações e desafios. Para as trilhas $T_3, T_7, T_8, T_9, T_{10}$ existe a limitação da falta de dados históricos, pois não há informações de sócios, telefones, endereços, etc., no momento em que a licitação aconteceu. Há apenas os valores que constavam na base do Serpro em agosto de 2021. Essa falta de dados pode resultar em falsos positivos nos resultados das trilhas. Algumas empresas que usam serviços de escritório de contabilidade cadastram, junto a receita federal, o e-mail, telefone e endereço do escritório de contabilidade ao invés dos dados da empresa. Isso gera ruído nos resultados das trilhas T_8, T_9, T_{10} .

Agradecimentos. Ao Ministério Público de Minas Gerais (MPMG) pelo apoio através do Projeto Capacidades Analíticas. Ao CNPq, CAPES e Fapemig pelo apoio aos pesquisadores envolvidos.

Referências

- Abidi, W. U. H. et al. (2021). Real-time shill bidding fraud detection empowered with fussed machine learning. *IEEE Access*, 9:113612–113621.
- Andrade, P. H. M. A. et al. (2020). Auditing government purchases with a multicriteria anomaly detection strategy. *J. Inf. Data Manag.*, 11(1).

- Anowar, F. and Sadaoui, S. (2019). Multi-class ensemble learning of imbalanced bidding fraud data. In *Canadian AI*, volume 11489 of *Lecture Notes in Computer Science*, pages 352–358. Springer.
- Araújo, J. L. et al. (2021). Caracterização de caminhos mais prováveis em uma rede complexa de processos jurídicos. In *BraSNAM*, pages 44–54, Porto Alegre, Brasil. SBC.
- Barabási, A.-L. (2016). *Network science*. Cambridge University Press.
- Elshaar, S. and Sadaoui, S. (2020). Detecting bidding fraud using a few labeled data. In *ICAART*, pages 17–25. SCITEPRESS.
- Ganguly, S. and Sadaoui, S. (2018). Online detection of shill bidding fraud based on machine learning techniques. In *IEA/AIE*, volume 10868 of *Lecture Notes in Computer Science*, pages 303–314. Springer.
- Grace, E. et al. (2016). Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system. In *IEEE BigData*, pages 1444–1453. IEEE Computer Society.
- Lima, M. et al. (2020). Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach. In *EMNLP*, pages 1580–1588, Online. ACL.
- Pereira, R. and Murai, F. (2021). Quão efetivas são redes neurais baseadas em grafos na detecção de fraude para dados em rede? In *BraSNAM*, pages 205–210, Porto Alegre, Brasil. SBC.
- Ralha, C. G. and Silva, C. V. S. (2012). A multi-agent data mining system for cartel detection in brazilian government procurement. *Exp. Syst. Appl.*, 39(14):11642–11656.
- Velasco, R. B. et al. (2021). A decision support system for fraud detection in public procurement. *Int. Trans. Oper. Res.*, 28(1):27–47.
- Vlasselaer, V. V. et al. (2015). AFRAID: fraud detection via active inference in time-evolving social networks. In *ASONAM*, pages 659–666. ACM.