

# Ranqueamento de Licitações Públicas a partir de Alertas de Fraude

Gabriel P. Oliveira<sup>1</sup>, Bárbara M. A. Mendes<sup>1</sup>,  
Camila S. Braz<sup>1</sup>, Lucas L. Costa<sup>1</sup>, Mariana O. Silva<sup>1</sup>,  
Michele A. Brandão<sup>1,2</sup>, Anísio Lacerda<sup>1</sup>, Gisele L. Pappa<sup>1</sup>

<sup>1</sup>Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG, Brasil

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

gabrieloliveira@dcc.ufmg.br, {barbaramit, camilabraz}@ufmg.br  
lucas-lage@ufmg.br, mariana.santos@dcc.ufmg.br  
michele.brandao@ifmg.edu.br, {anisio, glpappa}@dcc.ufmg.br

**Abstract.** *Fraud detection is a complex task in various scenarios. This work proposes an approach to rank public tenders suspected of fraud. For this, we have created 19 audit trails, which were modeled as a social network, and a strategy to consider the alerts they raised in the ranking. The results reveal that the proposed ranking approach can correctly identify bids suspected of fraud.*

**Resumo.** *A detecção de fraudes é uma tarefa complexa nos diversos cenários. Este trabalho representa uma contribuição para essa tarefa ao propor uma abordagem para ranquear licitações públicas suspeitas de possuírem algum tipo de fraude. Para isso, foram elaboradas 19 trilhas de auditoria, que foram modeladas como uma rede social, e uma estratégia para considerar os alertas levantados por elas no ranqueamento. Os resultados revelam que a abordagem de ranqueamento foi capaz de identificar licitações suspeitas de fraude.*

## 1. Introdução

Dados são recursos valiosos para organizações públicas e privadas, permitindo a extração de informações para auxiliar na tomada de decisão em diversas áreas [Velasco et al. 2021]. No setor público, vários países têm firmado acordos para incentivar e fortalecer a transparência governamental, uma política apoiada pelas Tecnologias de Informação e Comunicação (TICs), que também ajudam no combate à corrupção [Park and Kim 2020]. Dentre os dados abertos brasileiros, encontram-se as licitações públicas, que são um conjunto de procedimentos administrativos utilizados pela Administração Pública para comprar ou contratar produtos e serviços<sup>1</sup>.

Apesar dos mecanismos impostos pela legislação para prevenir fraudes em licitações públicas, o Brasil ainda enfrenta escândalos de corrupção. A checagem manual em tempo real é comumente utilizada para acompanhar todo o processo licitatório, desde a publicação do edital até o cumprimento do objeto licitado. Contudo, essa inspeção manual é um processo demorado e que exige grande mobilização de recursos humanos. Nesse sentido, é essencial combinar perícia humana com sistemas computacionais. A detecção de fraudes em licitações é complexa, já que essas práticas envolvem interações

---

<sup>1</sup>Lei nº 8.666/93: [http://www.planalto.gov.br/ccivil\\_03/Leis/L8666cons.htm](http://www.planalto.gov.br/ccivil_03/Leis/L8666cons.htm)

diretas, indiretas e temporais entre as entidades envolvidas. Dessa forma, é necessário o uso de formas automáticas de monitoramento, análise e cruzamento dos dados para auxiliar na identificação de infrações.

Este trabalho tem como principal objetivo aplicar conceitos de trilhas de auditoria e Redes Sociais (RS) para identificar indícios de fraudes em licitações públicas, facilitando a análise desses dados por meio de tecnologias da informação. Para alcançar essa meta, um conjunto de trilhas de auditoria foi definido em colaboração com especialistas do Ministério Público de Minas Gerais, a partir de dados de licitações públicas municipais e estaduais de Minas Gerais. Essas trilhas são sequências de passos que ajudam a identificar tipos específicos de irregularidades encontradas em fraudes de licitações, selecionando os dados relevantes das bases de dados de licitações. Por exemplo, é possível identificar licitações com apenas um licitante ou licitantes que possuem sócios em comum.

Assim como em análises anteriores [Costa et al. 2022, Costa et al. 2023], neste trabalho, utilizamos uma abordagem baseada em grafos para a detecção de fraudes em licitações públicas. Como principal novidade, incluímos sete novas trilhas de auditoria ao modelo de ranqueamento de licitações baseado no número de alertas de fraude. Tal metodologia, além de reduzir o volume de dados a serem analisados por especialistas e indicar uma lista de prioridades de investigação, também fornece subsídios para a elaboração de algoritmos capazes de classificar automaticamente as licitações.

## **2. Trabalhos Relacionados**

As teorias de redes sociais têm sido aplicadas em diversos contextos, como detecção de emoções e identificação de suspeitos de crimes [Kansaon et al. 2019, Florentino et al. 2022]. No contexto de detecção de fraudes, as redes sociais têm sido utilizadas para obter informações sobre colaborações entre suspeitos de fraude [Óskarsdóttir et al. 2022]. Além disso, técnicas de mineração de dados e aprendizado de máquina têm sido empregadas na detecção automática de fraudes, incluindo a quantificação de padrões de risco de corrupção [Velasco et al. 2021] e o desenvolvimento de classificadores para identificação de fraudes [Anowar and Sadaoui 2019].

Outra abordagem utilizada é a mineração de processos, que combina mineração de dados e análise de processos para a detecção de fraudes. Essa abordagem foi aplicada na auditoria de fraudes em pedidos de pagamento da União Europeia [Santoro et al. 2020] e no monitoramento do comportamento de usuários em uma plataforma de *e-commerce* [Yu et al. 2023]. Embora as abordagens baseadas em dados rotulados e mineração de processos sejam comuns, o uso de redes sociais tem se mostrado uma alternativa mais simples e efetiva para a detecção de fraudes [Araújo et al. 2021, Li et al. 2023].

Apesar das diversas abordagens existentes, poucos estudos têm se dedicado especificamente à detecção de fraudes em licitações públicas [Grace et al. 2016, Andrade et al. 2020, Lima et al. 2020]. Este trabalho visa preencher essa lacuna ao analisar alertas de fraudes em licitações públicas em Minas Gerais por meio da análise da rede social construída a partir das trilhas de auditoria. Além disso, uma abordagem baseada nos alertas gerados pelas trilhas é aplicada para ranquear as licitações públicas.



**Figura 1. Metodologia para alerta de fraudes em licitações.**

### 3. Conjunto de Dados

As trilhas de auditoria construídas consomem dados relacionados principalmente a licitações ocorridas em Minas Gerais e seus licitantes. Neste trabalho, consideramos informações de 767.188 licitações públicas ocorridas em Minas Gerais de 2007 a 2022, compreendendo um total de 110.387 licitantes, incluindo pessoas físicas e jurídicas. Os dados foram fornecidos pelo Ministério Público de Minas Gerais (MPMG) no âmbito do Programa de Capacidades Analíticas e são disponibilizados em um armazém de dados que agrega informações de diferentes fontes públicas e privadas.

Os dados de licitações são provenientes do Sistema Informatizado de Contas dos Municípios (SICOM)<sup>2</sup> e do Portal da Transparência do Estado<sup>3</sup> e contêm informações como data, modalidade e itens de licitações, além da lista de licitantes participantes e vencedores. As informações sobre os licitantes são complementadas com dados sigilosos provenientes do Serviço Federal de Processamento de Dados (SERPRO), que incluem dados cadastrais e quadro societário de empresas. Além disso, algumas trilhas específicas consideram informações de outras fontes, como o registro de ligações de energia elétrica da Companhia Energética de Minas Gerais (CEMIG), dados de filiação partidária e de doações e serviços prestados a campanhas eleitorais do Tribunal Superior Eleitoral (TSE) e do Tribunal Regional Eleitoral de Minas Gerais (TRE-MG), e dados do Instituto de Previdência dos Servidores do Estado de Minas Gerais (IPSEMG).

### 4. Metodologia para Identificação de Fraudes

Esta seção apresenta a metodologia proposta para identificação de possíveis fraudes a partir de alertas gerados para licitações públicas. Todas as etapas da metodologia foram pensadas a partir da definição das trilhas de auditoria (Seção 4.1). A Figura 1 resume a metodologia utilizada, que consiste em: filtragem e pré-processamento dos dados (Seção 4.2), modelagem de uma rede social para representar as trilhas (Seção 4.3) e, por fim, um modelo de ranqueamento de licitações com base nos alertas gerados pelas trilhas.

#### 4.1. Definição das Trilhas de Auditoria

Neste trabalho, foram definidas em conjunto com especialistas uma série de regras (i.e., trilhas) para gerar alertas que podem indicar possíveis fraudes em licitações públicas. A Tabela 1 apresenta as 19 trilhas consideradas neste trabalho, sendo que 12 trilhas foram definidas em [Costa et al. 2023] e as sete novas são propostas neste trabalho (marcadas em negrito na tabela). As primeiras doze trilhas visam investigar se existem irregularidades na licitação considerando características do licitante (e.g., licitante com CNPJ

<sup>2</sup>SICOM: <https://portalsicom1.tce.mg.gov.br/>

<sup>3</sup>Portal da Transparência do Estado de Minas Gerais: <https://www.transparencia.mg.gov.br/compras-e-patrimonio/compras-e-contratos>

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

#	Trilha de Auditoria	Regra
<i>Trilhas de Licitante</i>		
$T_1$	Licitante licitando antes de registro	Verificar as licitações que contenham participantes que estão licitando antes da empresa iniciar suas atividades
$T_2$	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)
$T_3$	Licitante com CNPJ inativo	Verificar as licitações com a presença de licitantes com o CNPJ inativo
$T_4$	Licitante perdedor frequente	Identificar licitações com licitantes com alto percentual de derrotas
$T_5$	Licitante vencedor frequente	Identificar licitações com licitantes com alto percentual de vitórias
$T_6$	Licitante único	Identificar licitações com licitante único
$T_7$	Licitante com CNAE incongruente	Identificar licitações com licitantes cujo código CNAE (Classificação Nacional de Atividades Econômicas) é incongruente com a descrição dos itens licitados [Oliveira et al. 2022, Costa et al. 2023]
$T_8$	<b>Licitante que prestou serviço para campanhas políticas</b>	Identificar licitações com licitantes que prestaram serviço para campanhas políticas
$T_9$	<b>Licitante cujo lance foi o menor e não venceu</b>	Identificar licitações com licitantes que deram o menor lance, mas o lance não foi o vencedor
$T_{10}$	<b>Licitante sem registro de ligação de energia elétrica em MG</b>	Identificar licitações com a presença de licitantes baseados em Minas Gerais sem registro de ligação de energia elétrica no estado
$T_{11}$	<b>Licitante de pequeno porte com faturamento anual acima do limite</b>	Identificar licitações com licitantes que são empresas de pequeno porte, mas que possuem o faturamento acima do permitido por lei [Braz et al. 2023]
$T_{12}$	<b>Licitante de pequeno porte vinculado a outra pessoa jurídica</b>	Identificar licitações com licitantes que são empresas de pequeno porte e possuem outras pessoas jurídicas como sócias [Braz et al. 2023]
<i>Trilhas de Sócio</i>		
$T_{13}$	Licitante cujos sócios são ou têm vínculo com servidores públicos	Identificar as licitações que apresentam licitantes cujos sócios são servidores públicos ou dependentes de servidores da entidade que realiza a licitação [Costa et al. 2023]
$T_{14}$	<b>Licitante cujos sócios são doadores de campanha política</b>	Identificar licitações com licitantes cujos sócios fizeram doações para campanhas políticas
$T_{15}$	<b>Licitante cujos sócios são filiados a partidos políticos</b>	Identificar licitações com licitantes cujos sócios são filiados a partidos políticos
<i>Trilhas de Vínculo</i>		
$T_{16}$	Licitantes com sócios em comum	Identificar as licitações com licitantes distintos que possuem pelo menos um sócio em comum
$T_{17}$	Licitantes com e-mails em comum	Verificar se a licitação possui licitantes distintos com e-mails em comum
$T_{18}$	Licitantes com telefones em comum	Verificar se a licitação possui licitantes distintos com telefone em comum
$T_{19}$	Licitantes com endereços em comum	Identificar as licitações com licitantes distintos com endereço em comum

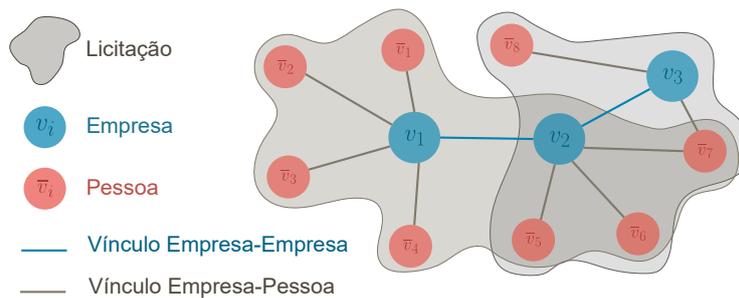
\*Trilhas em negrito foram propostas neste trabalho.

– Cadastro Nacional de Pessoas Jurídicas – inativo), sendo denominadas *Trilhas de Licitante*. As três trilhas seguintes verificam irregularidades relacionadas a pessoas sócias dos licitantes, sendo denominadas *Trilhas de Sócio*. Já as quatro trilhas restantes investigam irregularidades considerando os vínculos entre os licitantes (e.g., empresas com sócios em comum), sendo chamadas *Trilhas de Vínculo*.

É importante ressaltar que uma trilha analisada separadamente não necessariamente indica fraude. Por exemplo, a trilha  $T_8$  visa identificar licitações com licitantes que doaram para campanhas políticas. De acordo com a legislação vigente no Brasil, tal ato não é ilegal. No entanto, ao considerar este alerta em conjunto com alertas gerados por outras trilhas, é possível ter uma indicação de investigação por parte de especialistas. Ou seja, a construção e execução das trilhas não exclui a necessidade da análise manual das licitações públicas, mas otimiza e torna o trabalho dos analistas mais eficiente.

#### 4.2. Filtragem e Pré-processamento dos Dados

Conforme mencionado na Seção 3, são utilizados dados públicos de licitações, e dessa forma os licitantes que são pessoas físicas estão com o CPF (CADASTRO DE PESSOA



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

FÍSICA) mascarado no conjunto de dados considerado. Isso impossibilita qualquer processamento sobre esses dados e, portanto, eles são removidos das análises. Sendo assim, são considerados apenas três tipos de entidades de interesse neste trabalho: licitações, licitantes do tipo pessoa jurídica (Empresas) e seus sócios (Pessoas).

Após a filtragem dos dados, o próximo passo é lidar com os dados brutos retornados pelas consultas. Para isso, são executadas duas etapas adicionais de pré-processamento dos dados: padronização de dados e construção de atributos referente às entidades. A primeira visa normalizar informações de contato (i.e., e-mail, endereço e telefone) e também utiliza expressões regulares para verificar o formato de cada campo. Já o segundo tipo de pré-processamento é necessário devido às múltiplas fontes de dados utilizadas para construir o conjunto final de dados (Seção 3). Por exemplo, as licitações estaduais só possuem informações dos licitantes vencedores. Por isso, para essas licitações, apenas tais licitantes são considerados.

### 4.3. Modelagem da Rede Social

O problema para levantar alertas de fraudes em licitações pode ser modelado como uma rede social representada por um grafo  $G$ , com dois tipos de nós: as empresas licitantes ( $V$ ) e os sócios dessas empresas ( $\bar{V}$ ). Nesse grafo, há dois 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, cada aresta possui um peso representado pela quantidade de vínculos que uma empresa licitante tem com outra. Os vínculos levados em conta foram sócios, e-mail, telefones ou endereço em comum. No caso de relações entre empresas e pessoas, a relação só existe quando a pessoa é sócia da empresa. A Figura 2 ilustra a modelagem dessa rede social.

Além disso, os nós da rede social proposta apresentam atributos, que apesar de já utilizados para construir as arestas do grafo, são necessários para definir as regras que devem ser seguidas pelas trilhas de auditoria (apresentadas na Seção 4.1). Para o licitante (nó do tipo empresa), temos a data de registro da empresa (*data\_registro*), CNPJ, CNAE (Classificação Nacional de Atividades Econômicas), *status* da empresa na Receita Federal (e.g., ativa ou não), *endereço*, *e-mail*, *telefone*, *sócios* e *sanções*. Para o sócio (nó pessoa), temos o CPF e o tipo de *vínculo* com a empresa (e.g., sócio, representante legal).

É importante destacar que com a data de cada licitação, foram construídos dois novos atributos: vencedor frequente e perdedor frente, que assumem valor verdadeiro ou falso. Para a identificação de empresas vencedoras frequentes, foram analisadas empresas que somaram um valor total homologado maior ou igual a R\$400.000,00 (quatrocentos

mil reais). Para uma empresa ser considerada vencedora de uma licitação, é necessário que ela tenha vencido pelo menos 70% das licitações concorridas. Para a identificação de empresas perdedoras frequentes, foram analisadas empresas que tenham perdido pelo menos 70% das licitações com uma participação mínima de pelo menos seis licitações por ano. Para uma empresa ser perdedora em uma licitação, ela não pode ter vencido em nenhum item.

#### 4.4. Ranqueamento de Licitações por Alertas

Para realizar o ranqueamento de licitações, este artigo apresenta uma abordagem que gera uma pontuação para cada licitação e define o grau de risco das trilhas de auditoria. A abordagem é baseada em [Costa et al. 2023] e foi atualizada para incluir as novas trilhas e aprimorar limitações da versão anterior.

**Definição da Abordagem de Ranqueamento.** Essa abordagem considera dois aspectos: (i) a definição empírica de pesos diferentes para cada trilha de acordo com seu potencial risco de fraude; e (ii) o uso do valor total licitado, uma vez que licitações de maior valor são consideradas mais relevantes e têm prioridade na alocação de recursos humanos. O indicador de risco de fraude de uma licitação pode ser calculado pelo produto escalar do vetor contendo os alertas gerados pelas trilhas para a licitação em questão com o vetor de pesos de cada trilha. Este resultado é depois multiplicado pelo valor normalizado da licitação. Tal indicador é dado pela função:  $R_i(Q_{i*}, P, v_i) = (Q_{i*} \cdot P) \log(v_i)$ , onde:

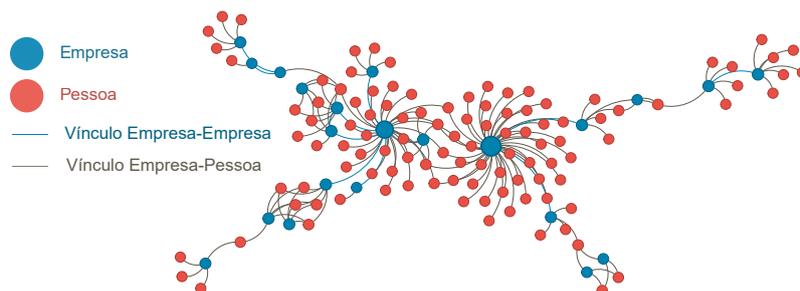
- $Q_{I \times X} = \{q_{ix} \in \mathbb{N}\}$  é a matriz com a quantidade de alertas das licitações e trilhas, onde as linhas são as licitações e as colunas são as trilhas. A célula dessa matriz,  $q_{ix}$ , representa a quantidade de alertas que a licitação  $i$  possui na trilha  $x$ .
- $Q_{i*}$  é uma linha da matriz  $Q$ , que representa os alertas da licitação  $i$ , ou seja,  $Q_{i*} = \{q_{i1}, q_{i2}, \dots, q_{i19}\}$ ;
- $P$  é o vetor dos pesos de cada trilha, e o peso atribuído para a trilha  $x$  é representado por  $p_x$ , ou seja,  $P = \{p_x \in \mathbb{R}\}$ ;
- $v_i$  é o valor total da licitação  $i$ .

O cálculo da quantidade de alertas  $q_{ix}$  depende do tipo de trilha (Tabela 1). Quando a trilha é de licitantes, a quantidade de alertas que descumprem a regra da trilha é o número de nós do tipo empresa; já para trilha de sócio, a quantidade de alertas é o número de nós do tipo pessoa; e para trilhas de vínculo, essa quantidade é o número de arestas.

**Grau de Risco das Trilhas.** Foi atribuído um grau de risco para cada trilha e um valor de peso para cada grau de risco. Esse grau de risco foi atribuído por auditores especializados de acordo com a gravidade do alerta gerado pela trilha de auditoria. Em resumo, as trilhas de grau de risco *alto* ( $T_1, T_2, T_3, T_{16}$ ) vão ter peso 1, as trilhas de grau *médio* ( $T_7, T_9, T_{10}, T_{13}, T_{19}$ ) vão ter peso 0,8 e as trilhas de grau *baixo* ( $T_4, T_5, T_6, T_8, T_{11}, T_{12}, T_{14}, T_{15}, T_{17}, T_{18}$ ) vão ter peso 0,6. Esses valores de peso foram definidos empiricamente em conjunto com auditores especialistas do MPMG.

## 5. Caracterização e Análise da Rede Social Real

A análise da rede social entre licitantes e seus sócios possibilita a identificação de possíveis alertas de fraude. Na Seção 5.1, é fornecida uma visão geral da rede por meio de métricas topológicas que a descrevem. Posteriormente, na Seção 5.2, é realizada uma



**Figura 3. Exemplo de um componente conectado da RS entre empresas licitantes e seus sócios. Empresas estão conectadas quando possuem vínculos em comum (e.g., endereço) e participam de um mesmo processo licitatório.**

**Tabela 2. Estatísticas da rede social de licitantes e seus sócios para a rede completa e seu componente gigante.**

	Completa	C. Gigante		Completa	C. Gigante
Empresas (CNPJ)	26.045	2	Densidade ( $10^{-2}$ )	0,0030	0,7484
Pessoas (CPF)	30.129	284	Grau médio	1,6948	2,1329
Nós (CNPJ e CPF)	56.174	286	CC médio	0,4195	0
Arestas	47.603	305	Componentes conectados	18.212	1

CC: Coeficiente de Clusterização

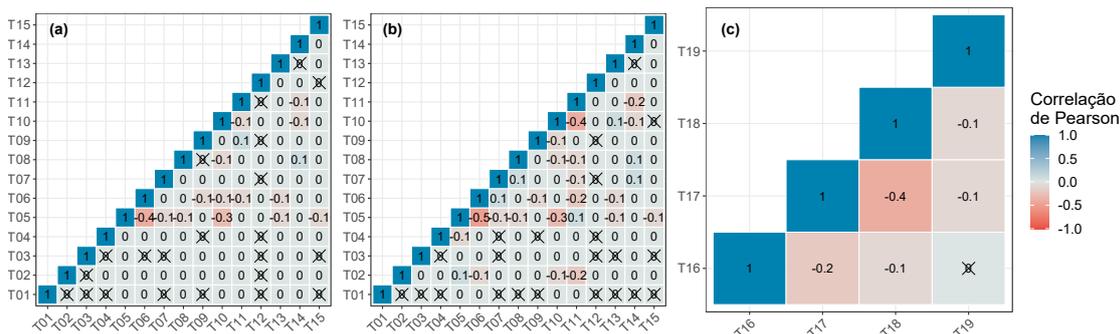
análise de correlação entre as trilhas de auditoria para verificar a coocorrência de alertas de fraude. Essa abordagem permite identificar padrões e conexões entre os diferentes indícios de fraude, contribuindo para uma compreensão mais completa do cenário e auxiliando na detecção e investigação de potenciais irregularidades nas licitações.

### 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 fica evidente em seu tamanho, uma vez que ela é composta por mais de 56 mil nós, representando empresas e seus sócios, e quase 48 mil arestas que representam as conexões entre esses nós. Essa vasta quantidade de nós e arestas reflete a riqueza de interações e vínculos presentes na rede, proporcionando uma visão abrangente dos relacionamentos entre os participantes das licitações públicas. Ilustrar tal rede de forma adequada é uma tarefa desafiadora. Portanto, a Figura 3 apresenta somente a visualização de um dos maiores componentes conectados da rede.

Em seguida, foram aplicadas métricas topológicas para analisar as principais características da rede social. Foram considerados o número de nós e arestas presentes na rede, a densidade, que representa a razão entre a quantidade de arestas existentes e todas as arestas possíveis, o grau médio, que indica a média de conexões de cada nó, e o coeficiente de clusterização, que mede a tendência dos nós de se conectarem uns aos outros [Barabási 2016]. A Tabela 2 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 estrutura construída. A maioria dos nós são representados por pessoas físicas que atuam como sócias das empresas licitantes, o que é esperado, considerando que empresas de médio e



**Figura 4. Correlação entre as trilhas de (a-b) licitantes e aos (c) pares de licitante. A matriz (b) é uma versão filtrada da (a), onde foram removidas as ocorrências em que o licitante se enquadra em apenas uma das trilhas. × indica uma correlação estatisticamente não significativa (p-valor  $\geq 0,05$ ).**

grande porte geralmente possuem vários representantes legais. Essa distribuição também é observada no componente gigante da rede, que abrange 2 empresas e 284 sócios no total. Ao analisar métricas como o número de arestas, coeficiente médio de clusterização e densidade da rede, é possível perceber que a rede social construída é esparsa, i.e., apresenta uma baixa densidade (valor próximo a zero) de conexões em relação ao total de conexões possíveis. 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.

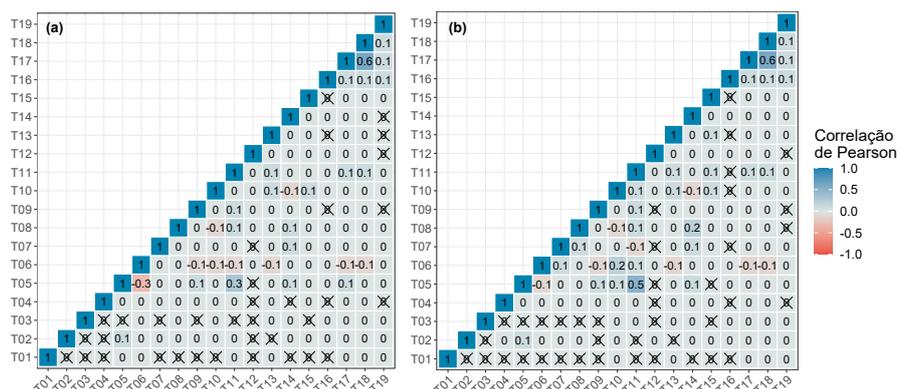
Por fim, o número de componentes conectados na rede completa revela uma alta fragmentação entre licitantes. Isso pode significar que as trilhas de auditoria não conseguiram revelar grandes grupos de licitantes que possuam vínculos em comum. Se a rede possuísse poucos componentes conectados, poderia-se 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, é realizada uma análise de correlação entre as trilhas de auditoria, tanto relacionadas às arestas (pares de licitantes), quanto aos nós (licitantes ou sócios). Para isso, foram consideradas um total de 583.631 licitações que se enquadram em pelo menos uma das trilhas descritas na Tabela 1. Visando avaliar a coocorrência entre essas trilhas, foram geradas matrizes de correlação.

**Correlação em nível de licitante.** As Figuras 4(a) e 4(c) mostram as matrizes de correlação de Pearson entre as trilhas referentes aos licitantes e aos pares de licitante, respectivamente. Para facilitar a visualização e interpretação dos resultados, a Figura 4(b) exibe uma versão filtrada da matriz (a), na qual foram removidas as licitações que são exclusivas de uma única trilha. Essa filtragem permite identificar padrões mais relevantes de coocorrência entre as trilhas, eliminando associações isoladas e menos significativas.

Ao comparar as Figuras 4(a) e 4(b), é possível identificar padrões mais claros



**Figura 5. Correlação entre as trilhas de auditoria, em nível de licitação. A matriz (b) é uma versão filtrada da (a), onde foram removidas as ocorrências em que a licitação se enquadra apenas em cada uma das trilhas. × indica uma correlação estatisticamente não significativa ( $p$ -valor  $\geq 0,05$ ).**

e relevantes na matriz (b), uma vez que ela destaca as correlações mais consistentes e recorrentes entre as trilhas. Após a filtragem, podemos notar que a maioria das correlações estatisticamente significativas apresenta uma relação negativa fraca a moderada, variando entre -0,1 a -0,4. A única correlação forte (-0,5) identificada acontece entre as trilhas  $T_5$  e  $T_6$ , indicando uma coocorrência negativa. Ou seja, licitações que possuem licitantes com alto percentual de vitórias ( $T_5$ ) tendem a não possuir um único licitante ( $T_6$ ).

Analisando as trilhas referentes aos pares de licitantes (i.e.,  $T_{16}$  a  $T_{19}$ ), 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 entre as trilhas  $T_{17}$  e  $T_{18}$ , que investigam licitações contendo licitantes com e-mails e telefones em comum, respectivamente. Tal resultado indica que grande parte dos pares de licitantes enquadrados na trilha de e-mails não se enquadram na de telefone, e vice-versa. Ou seja, no geral, empresas licitantes que possuem e-mails em comum, não compartilham o mesmo telefone.

**Correlação em nível de licitação.** Na Figura 5, foi analisada a correlação entre todas as trilhas, agora em nível de licitação. Em outras palavras, foi analisada a ocorrência simultânea das trilhas em uma mesma licitação. Mais uma vez, foi feita a filtragem das licitações que são exclusivas de uma única trilha (Figura 5(b)). No geral, a grande maioria das correlações não é estatisticamente significativa ou não apresentam nenhuma relação, mesmo após a filtragem. A correlação mais evidente é a relação positiva forte (0,6) entre as trilhas  $T_{17}$  e  $T_{18}$ , indicando que se uma licitação apresentar licitantes distintos com e-mails em comum, há uma grande chance de também apresentar licitantes distintos com número de telefone em comum, e vice-versa.

Após a filtragem, outra correlação positiva (0,5) aparente ocorre entre as trilhas  $T_5$  e  $T_{11}$ . Tal resultado indica que se uma licitação tem licitantes com alto percentual de vitórias ( $T_5$ ), ela tem mais chances de ter licitantes que são empresas de pequeno porte, mas que possuem o faturamento acima do permitido por lei ( $T_{11}$ ). Essa correlação pode indicar uma prática fraudulenta em que empresas de pequeno porte estão inflando seus resultados para obter vantagens competitivas.

**Tabela 3. Top 5 licitações com maiores valores para o Indicador de Risco.**

Pos.	Trilhas (Qtd. Alertas)	Modalidade	Valor Licitado	Risco
1 <sup>a</sup>	$T_{10}$ (88), $T_{17}$ (2), $T_{18}$ (3), $T_{19}$ (1)	Concurso	R\$ 548.500,00	980,55
2 <sup>a</sup>	$T_{10}$ (57), $T_{11}$ (11), $T_{17}$ (5), $T_{18}$ (2), $T_{19}$ (7)	Pregão Presencial	R\$ 3.432.317,50	933,02
3 <sup>a</sup>	$T_9$ (1), $T_{10}$ (55), $T_{11}$ (6), $T_{17}$ (5), $T_{18}$ (2), $T_{19}$ (4)	Pregão Presencial	R\$ 2.710.163,40	826,54
4 <sup>a</sup>	$T_{10}$ (51), $T_{11}$ (6), $T_{17}$ (4), $T_{18}$ (2), $T_{19}$ (3)	Pregão Presencial	R\$ 2.947.456,20	750,78
5 <sup>a</sup>	$T_2$ (1), $T_5$ (39), $T_8$ (2), $T_{11}$ (18), $T_{14}$ (5), $T_{16}$ (1), $T_{17}$ (1), $T_{18}$ (1), $T_{19}$ (1)	Pregão Eletrônico	R\$ 19.411.624,20	711,53

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. Além disso, é importante destacar que a presença de uma correlação não implica necessariamente em uma relação causal entre as trilhas. A interpretação dessas correlações requer uma análise mais aprofundada e contextualizada, considerando as particularidades do domínio das licitações públicas e a interação entre as diferentes trilhas de auditoria.

## 6. Análise das Licitações no Ranqueamento

Esta seção apresenta os resultados da abordagem de ranqueamento das licitações de acordo com seu indicador de risco. No total, 566.539 licitações foram classificadas com indicador de risco maior que zero e foram incluídas no ranqueamento. A Tabela 3 apresenta as top 5 licitações ranqueadas. A primeira coluna indica a posição da licitação no ranqueamento, seguida pela coluna “Trilhas”, que mostra a quantidade de alertas das 19 trilhas associadas à licitação. Nas colunas subsequentes, são apresentados o preço total dos objetos descritos na licitação (valor licitado) e o indicador de risco correspondente.

A análise do ranqueamento revela que as licitações posicionadas no topo do ranking apresentam um elevado número de alertas nas trilhas  $T_5$ ,  $T_{10}$  e  $T_{11}$ , mesmo essas trilhas possuindo pesos baixos ou médios, conforme mencionado na Seção 4.4. Essas licitações foram bem ranqueadas devido à grande quantidade de alertas nessas trilhas. Por exemplo, as quatro primeiras licitações do ranking possuem mais de 50 alertas na trilha relacionada a licitantes sem registro de ligação de energia elétrica em Minas Gerais ( $T_{10}$ ), sendo que a primeira posição apresenta um total de 88 alertas nessa trilha. Em relação às trilhas de peso alto, apenas a quinta licitação possui um alerta referente a  $T_2$ , sobre licitantes com sanção ativa.

Esses resultados sugerem que, embora as trilhas de peso alto sejam consideradas importantes para a detecção de fraudes, as licitações com um elevado número de alertas nas trilhas de peso médio e baixo também merecem atenção. Isso pode indicar que, mesmo que um indicador específico possua um peso menor, a sua ocorrência em muitas licitações pode ser um sinal de potencial risco.

Para avaliar o ranqueamento, foi realizada a comparação com uma pequena amostra de 117 licitações já investigadas pelo MPMG e que comprovadamente tiveram algum tipo de fraude. Dessas 117, 113 (96,58%) possuem pelo menos um alerta identificado e, portanto, possuem um indicador de risco maior que zero e fazem parte do ranqueamento. A Tabela 4 apresenta as top 5 licitações fraudulentas investigadas por especialistas do MPMG e que possuem um indicador de risco. Observa-se que nenhuma delas ficou no

**Tabela 4. Posição no ranqueamento de licitações já investigadas pelo MPMG.**

Pos.	Trilhas (Qtd. Alertas)	Modalidade	Valor Licitado	Risco
10324 <sup>a</sup>	T <sub>9</sub> (5), T <sub>10</sub> (3), T <sub>11</sub> (3)	Pregão Presencial	R\$ 299.007,48	103,39
14542 <sup>a</sup>	T <sub>5</sub> (4), T <sub>10</sub> (3), T <sub>11</sub> (3)	Pregão Presencial	R\$ 780.299,50	89,54
16693 <sup>a</sup>	T <sub>10</sub> (4), T <sub>11</sub> (5)	Pregão Presencial	R\$ 792.506,84	84,21
19692 <sup>a</sup>	T <sub>9</sub> (3), T <sub>10</sub> (3), T <sub>11</sub> (2)	Pregão Presencial	R\$ 439.300,00	77,96
23033 <sup>a</sup>	T <sub>5</sub> (2), T <sub>10</sub> (3), T <sub>17</sub> (1), T <sub>18</sub> (1)	Pregão Presencial	R\$ 3.407.281,40	72,20

top 5 do ranqueamento gerado pela abordagem proposta, sendo que a licitação de posição mais alta ficou em 10324<sup>a</sup>.

A justificativa para esse resultado é que existem licitações com uma quantidade maior de alertas no topo do ranqueamento, conforme mostrado na Tabela 3. No entanto, é importante destacar que as cinco primeiras licitações da Tabela 4 ficaram acima do percentil 95 no ranqueamento, o que representa uma posição relativamente alta. Isso indica que a abordagem proposta conseguiu identificar licitações fraudulentas que estavam entre as mais suspeitas, embora não tenham alcançado o topo do ranqueamento.

Esses resultados reforçam a importância de uma análise mais detalhada e complementar por parte dos especialistas do MPMG, que consideram outros fatores além do ranqueamento, como evidências adicionais e informações específicas de cada licitação, para confirmar a presença de fraudes. O ranqueamento serve como uma ferramenta preliminar de triagem, que direciona a atenção para as licitações mais suspeitas e auxilia no processo de investigação. Portanto, embora as licitações fraudulentas não tenham alcançado o topo do ranqueamento, a abordagem proposta mostrou-se útil na identificação de licitações com indicadores de risco relevantes.

## 7. Conclusões e Trabalhos Futuros

As trilhas de auditoria definidas neste artigo combinam conceitos de redes sociais para identificar indícios de fraudes em licitações públicas. Os dados das licitações, licitantes e sócios foram modelados como uma rede social, permitindo a análise das interações entre os participantes. Os resultados obtidos demonstram a eficácia da metodologia na filtragem e redução do volume de dados a serem analisados por especialistas. Além disso, as descobertas fornecem subsídios para o desenvolvimento de algoritmos de classificação de licitações fraudulentas, contribuindo para o combate à corrupção. Como trabalhos futuros, planeja-se criar um painel interativo para facilitar a visualização e interpretação dos resultados obtidos através das trilhas.

**Limitações e desafios.** A maior limitação deste trabalho está na disponibilidade de dados. Por exemplo, não existem dados históricos que informem a situação cadastral e societária da empresa no momento em que a licitação aconteceu. Além disso, algumas empresas que usam serviços de contabilidade cadastram o e-mail e telefone dos escritórios de contabilidade na Receita Federal. Por fim, algumas fontes externas não possuem dados recentes, o que pode afetar o resultado das trilhas para licitações mais atuais.

**Agradecimentos.** Ao Ministério Público de Minas Gerais (MPMG) pelo apoio através do Projeto Capacidades Analíticas. Ao CNPq, CAPES e FAPEMIG pelo apoio aos pes-

quisadores envolvidos. Por fim, a Arthur P. G. Reis, Clara A. Bacha e Matheus C. Teixeira pela participação nos trabalhos anteriores.

## Referências

- 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 *LNCS*, 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.
- Braz, C. et al. (2023). Análise de irregularidades em licitações públicas com foco em empresas de pequeno porte. In *WCGE*. SBC. [em submissão].
- Costa, L. L. et al. (2022). Alertas de fraude em licitações: Uma abordagem baseada em redes sociais. In *BraSNAM*, pages 37–48. SBC.
- Costa, L. L. et al. (2023). Identificação de licitações suspeitas de fraude por meio de trilhas de auditoria. *iSys*. [aceito para publicação].
- Florentino, É. S. et al. (2022). Identificando suspeitos de crimes por meio de interações implícitas no youtube. *iSys*, 15(1):3–1.
- 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.
- Kansoon, D. P. et al. (2019). Análise de algoritmos de classificação para detecção de emoções em tweets em português brasileiro. *iSys*, 12(3):116–138.
- Li, P. et al. (2023). LGM-GNN: A local and global aware memory-based graph neural network for fraud detection. *IEEE Transactions on Big Data*.
- 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.
- Oliveira, G. P. et al. (2022). Detecting inconsistencies in public bids: An automated and data-based approach. In *WebMedia*, pages 182–190.
- Óskarsdóttir, M. et al. (2022). Social network analytics for supervised fraud detection in insurance. *Risk Analysis*, 42(8):1872–1890.
- Park, C. H. and Kim, K. (2020). E-government as an anti-corruption tool: Panel data analysis across countries. *Intl. Review of Administrative Sciences*, 86(4):691–707.
- Santoro, F. M. et al. (2020). Process mining techniques in internal auditing: A stepwise case study. *iSys*, 13(4):48–76.
- Velasco, R. B. et al. (2021). A decision support system for fraud detection in public procurement. *Int. Trans. Oper. Res.*, 28(1):27–47.
- Yu, W. et al. (2023). A multiperspective fraud detection method for multiparticipant e-commerce transactions. *IEEE Transactions on Computational Social Systems*.