

INTELIGÊNCIA ARTIFICIAL · UNIFOR · 26 DE JUNHO DE 2026

Supervised learning on graph-derived structural and weighted metrics for tax fraud detection

Por Daniel Sabóia, Rilder S. Pires e Erneson A. Oliveira

Fonte: OpenAlex · Unifor · [Publicação original](#)

INOVAÇÃO

8/10

APLICABILIDADE

9/10

POTENCIAL
ECONÔMICO

9/10

MATURIDADE

Média

O estudo valida o uso de métricas estruturais de grafos combinadas com aprendizado de máquina (XGBoost) para detecção de fraude fiscal, superando modelos de deep learning (GCN) em interpretabilidade e performance, revelando padrões de compactez em redes fraudulentas.

NÚCLEO DA ANÁLISE

Ideia de startup ou produto

Desenvolvimento de uma solução SaaS B2G/B2B de 'Graph Risk Scoring' que classifica o risco fiscal de entidades baseando-se na topologia de suas redes de relacionamentos.

Aplicações práticas

Sistemas de auditoria tributária preditiva, plataformas de compliance corporativa para detecção de risco de terceiros e ferramentas de inteligência financeira forense.

Potencial de mercado

Alto potencial no setor de RegTech e GovTech, dada a necessidade global de modernizar arrecadação e combater a elisão fiscal com técnicas avançadas de dados.

FUNDAMENTO CIENTÍFICO

Problema abordado

Dificuldade em identificar esquemas complexos de evasão fiscal que utilizam redes de relacionamento empresariais intrincadas para ocultar beneficiários reais e comportamentos econômicos coordenados.

Metodologia

Extração de características em nível de grafo (compactez, coesão, influência ponderada) de redes fiscais ego-centradas e avaliação comparativa entre classificadores supervisionados (XGBoost) e redes neurais em grafos (GCN, GraphSAGE).

Principais descobertas

Descritores estruturais de grafos capturam informações preditivas significativas; XGBoost obteve o maior F1-score e AUC-ROC; existe um eixo estrutural claro que distingue redes fraudulentas (mais compactas) de legítimas.

LEITURA PARA GESTÃO PÚBLICA

Integração da metodologia aos sistemas de inteligência da Receita Federal e SEFAZ para priorização de auditorias e cruzamento automatizado de dados cadastrais e transacionais.

PARCERIA UNIVERSIDADE × EMPRESA

Parceria tecnológica entre a UNIFOR e órgãos de fiscalização ou grandes firmas de auditoria para validar o algoritmo em bases de dados reais e produção de software sob medida.

OPORTUNIDADES DERIVADAS

3 direções estratégicas identificadas

PRODUTO CORPORATIVO · IMPACTO ALTO · CIÊNCIA DE DADOS

Módulo de Due Diligence Relacional

Plugin para ERPs corporativos que alerta sobre conexões suspeitas de parceiros comerciais utilizando análise de grafos.

POLÍTICA PÚBLICA · IMPACTO ALTO · GOVTECH

Auditoria Tributária Orientada por Grafos

Sistema estadual para identificação de cartéis e fraudes em licitações e notas fiscais baseado na topologia de redes.

STARTUP · IMPACTO MÉDIO · STARTUPS & INOVAÇÃO

FraudGraph Analytics

Spin-off focada em oferecer análise de risco fiscal como serviço para escritórios de contabilidade e departamentos financeiros.

ABSTRACT ORIGINAL

Structured tax fraud schemes frequently rely on complex relational arrangements designed to obscure beneficial ownership and coordinated economic behavior. This study investigates whether graph-derived structural and weighted descriptors can support supervised tax fraud detection in ego-centered fiscal relationship networks. Graph-level features capturing compactness, local cohesion, path organization, and weighted relational influence were extracted from institutional investigation graphs and evaluated using multiple supervised classifiers. Additionally, graph convolutional networks (GCN) and GraphSAGE (graph sample and aggregate) were evaluated as graph-native benchmark models operating directly on the

original relational structures. Exploratory analyses revealed a dominant structural axis separating compact and distance-driven configurations between fraudulent and non-fraudulent networks. Among the evaluated models, XGBoost achieved the highest cross-validated F1-score and strong ROC-AUC discrimination, while GraphSAGE achieved competitive performance among graph-native approaches. Results suggest that interpretable graph-derived descriptors capture substantial predictive relational information embedded in the investigated fiscal networks. These findings highlight the potential of graph-based descriptors as auditable tools for tax fraud detection in institutionally sensitive environments.

MATÉRIA PARA LEIGOS

Para leigos: Usando grafos para desvendar fraudes fiscais complexas

O cenário atual

Esquemas estruturados de fraude fiscal frequentemente dependem de arranjos relacionais complexos. O objetivo dessas estruturas é ofuscar a propriedade beneficiária real e ocultar comportamentos econômicos coordenados. Identificar esses padrões é um desafio para as autoridades.

O que os pesquisadores fizeram

Os pesquisadores investigaram se descritores estruturais e ponderados, derivados de grafos, podem apoiar a detecção supervisionada de fraudes fiscais. O estudo focou em redes de relacionamentos fiscais ego-centradas. Foram extraídas características em nível de grafo e avaliadas usando múltiplos classificadores supervisionados. Além disso, modelos nativos de grafo, como redes neurais convolucionais de grafos (GCN) e GraphSAGE, foram testados como benchmarks.

Como funciona na prática

A equipe extraiu métricas que capturam a compacidade, a coesão local, a organização de caminhos e a influência relacional ponderada. Essas informações vieram de grafos de investigação institucional. Em seguida, esses dados foram utilizados para treinar algoritmos de aprendizado de máquina (supervised learning) a fim de distinguir entre redes fraudulentas e não fraudulentas.

Resultados e evidência

Análises exploratórias revelaram um eixo estrutural dominante que separa configurações compactas de configurações guiadas por distância. Essa separação ocorreu entre as redes fraudulentas e as não fraudulentas. Entre os modelos avaliados, o XGBoost alcançou o maior F1-score validado cruzadamente e uma forte discriminação ROC-AUC. Entre as abordagens nativas de grafo, o modelo GraphSAGE apresentou desempenho competitivo.

Implicações práticas

Os resultados sugerem que descritores baseados em grafos, que são interpretáveis, capturam uma quantidade substancial de informação relacional preditiva. Isso destaca o potencial dessas métricas como ferramentas auditáveis para a detecção de fraudes fiscais, especialmente em ambientes institucionalmente sensíveis onde a transparência dos métodos é necessária.

Limitações e próximos passos

O paper não detalha limitações específicas do estudo ou indica explicitamente quais serão os próximos passos da pesquisa além dos resultados apresentados.

QUEM SÃO OS PESQUISADORES

Quem são os pesquisadores

O estudo foi conduzido por Daniel Sabóia, Rilder S. Pires e Erneson A. Oliveira, vinculados à Universidade de Fortaleza (UNIFOR). O material fornecido não detalha a formação acadêmica específica dos autores, suas trajetórias profissionais anteriores ou papéis distintos (como orientador ou orientando) dentro desta pesquisa específica.

PALAVRAS-CHAVE

Benchmark (surveying) · Graph · Sample (material) · Exploratory analysis · Exploratory research · Convolutional neural network