

Governança Digital e Investigação Patrimonial: Uma Heurística para Identificação de Grupos Econômicos

Francisco G. de Araújo Filho¹, Edison Ishikawa¹

¹Programa de Pós-Graduação em Computação Aplicada (PPCA)
Universidade de Brasília (UnB), DF – Brazil

fgaraujofh@gmail.com, ishikawa@unb.br

Abstract. *The identification of de facto economic groups is a relevant challenge in asset recovery, especially in cases of asset concealment through complex corporate structures. The lack of formal links hinders joint liability and compromises the effectiveness of judicial decisions. This paper proposes an investigative heuristic to identify hidden relationships between companies, based on real-world patterns observed by experts. The approach aims to support the Judiciary's efforts, aligning with the National Judiciary Strategy 2021–2026 and the goals of SDG 16. Preliminary results indicate that the tool reduces analysis time and enhances digital governance in asset investigations.*

Resumo. *A identificação de grupos econômicos de fato é um desafio relevante na recuperação de ativos, especialmente em contextos de ocultação patrimonial por meio de estruturas societárias complexas. A ausência de vínculos formais dificulta a responsabilização solidária e compromete a efetividade das decisões judiciais. Este artigo propõe uma heurística investigativa que identifica relações ocultas entre empresas, com base em padrões reais observados por especialistas. A abordagem visa apoiar a atuação do Poder Judiciário, alinhando-se à Estratégia Nacional 2021-2026 e aos objetivos do ODS 16. Resultados preliminares indicam que a ferramenta reduz o tempo de análise e aprimora a governança digital nas investigações patrimoniais.*

1. Introdução

A recuperação de ativos no Brasil enfrenta desafios significativos, sobretudo na identificação de grupos econômicos de fato — arranjos empresariais compostos por pessoas jurídicas formalmente independentes, mas que operam de forma coordenada e com unidade econômica —, que frequentemente ocultam patrimônio por meio de estruturas societárias complexas [Santos et al. 2024]. Como essas conexões nem sempre estão formalmente documentadas, torna-se difícil a responsabilização solidária e a aplicação de medidas como a desconsideração da personalidade jurídica. Esse cenário impacta diretamente a fase de execução de processos judiciais, um dos maiores gargalos do Poder Judiciário, dificultando a efetivação de decisões que buscam recuperar valores devidos¹.

A Estratégia Nacional do Poder Judiciário 2021-2026, instituída pelo Conselho Nacional de Justiça (CNJ), reconhece a necessidade de aprimorar a gestão administrativa

¹Conselho Nacional de Justiça. Justiça em Números 2024. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2024/05/justica-em-numeros-2024.pdf>. Acesso em: 5 mar. 2025.

e a governança judiciária, promovendo maior eficiência e transparência na condução de investigações patrimoniais. O Objetivo de Desenvolvimento Sustentável (ODS) 16 da Agenda 2030 das Nações Unidas reforça esse compromisso ao incentivar o fortalecimento institucional e a implementação de mecanismos que aprimorem a justiça e combatam fraudes financeiras².

Neste contexto, este trabalho propõe uma heurística investigativa para a identificação automatizada de grupos econômicos de fato, utilizando padrões estruturais recorrentes, como sobreposição de gestores, similaridade de endereços, interdependência de atividades econômicas e vínculos societários indiretos. O detalhamento desses critérios e a metodologia utilizada para inferi-los serão apresentados na Seção 4.3.

A proposta inclui o desenvolvimento de uma interface visual amigável, que facilita a identificação intuitiva das conexões detectadas. Após a aplicação da heurística, espera-se que os especialistas consigam visualizar as relações entre entidades investigadas, destacando indícios concretos da formação de um grupo econômico de fato, o que otimiza tanto o processo decisório quanto a condução da investigação.

Para isso, a abordagem utiliza dados do portal de dados abertos do Governo Federal, em especial o Cadastro Nacional de Pessoas Jurídicas (CNPJ).

2. Revisão da Literatura

A identificação de grupos econômicos é um desafio crítico tanto para fins de investigação patrimonial quanto para a execução de obrigações legais e fiscais, uma vez que sua caracterização exige análise de vínculos não evidentes entre pessoas jurídicas [Lopes and Frattari 2015, Santos et al. 2024].

A literatura jurídica e econômica define grupos econômicos como estruturas empresariais compostas por múltiplas entidades juridicamente independentes, mas interligadas por relações de controle, administração ou coordenação [Lopes and Frattari 2015]. No Brasil, o reconhecimento de um grupo econômico pode ocorrer de duas formas: *de direito*, quando formalmente constituído por meio de contratos e registros empresariais, e *de fato*, quando empresas atuam de maneira coordenada sem um vínculo societário formal, sendo essa última categoria mais desafiadora para identificação [Santos et al. 2024].

Com o avanço das técnicas de mineração de dados e modelagem de grafos, novas abordagens surgiram para identificar grupos econômicos de fato. A análise de grafos tem se mostrado uma ferramenta eficaz na detecção de conexões ocultas entre empresas, permitindo a construção de redes de relacionamentos e inferência de vínculos indiretos. [de Carvalho and Holanda Filho 2024] destaca a aplicabilidade da análise de grafos para detectar fraudes em aquisições públicas no Brasil, demonstrando que grupos econômicos podem ser identificados por meio da sobreposição de sócios e padrões de comportamento anômalos nas licitações públicas.

Além disso, técnicas computacionais têm sido aplicadas para identificar padrões indicativos de grupos econômicos de fato a partir de dados empresariais. Estudos como

²Conselho Nacional de Justiça. Estratégia Nacional do Poder Judiciário 2021-2026. Disponível em: <https://www.cnj.jus.br/gestao-estrategica-e-planejamento/estrategia-nacional-do-poder-judiciario-2021-2026/>. Acesso em: 5 mar. 2025.

o de Brandão et al. [de Carvalho and Holanda Filho 2024] exploram o uso de algoritmos de detecção de comunidades e previsão de ligações para mapear redes de empresas interligadas.

A correta identificação de grupos econômicos de fato desempenha um papel essencial na fundamentação da desconsideração da personalidade jurídica. No contexto jurídico, a comprovação de vínculos ocultos entre empresas pode fornecer evidências mais robustas para o redirecionamento de execuções fiscais e trabalhistas, especialmente em casos de confusão patrimonial e abuso da personalidade jurídica. [Santos et al. 2024] abordam como esse mecanismo tem sido utilizado pelo Fisco para redirecionar execuções fiscais a empresas de um mesmo grupo econômico quando há indícios de confusão patrimonial e fraude. Da mesma forma, [Brandão et al. 2024] analisa a responsabilidade trabalhista do sócio no grupo econômico, ressaltando que a caracterização de grupos econômicos pode ser um fator determinante para a responsabilização solidária das empresas envolvidas.

Esses estudos demonstram que a identificação automatizada de grupos econômicos de fato tem implicações tanto na investigação patrimonial quanto na responsabilização fiscal e trabalhista. Dado o impacto dessas abordagens na governança fiscal e nas investigações patrimoniais, torna-se essencial explorar metodologias que combinem técnicas computacionais avançadas com critérios jurídicos bem fundamentados. O presente trabalho propõe uma heurística investigativa que busca preencher essa lacuna, fornecendo uma solução automatizada para a identificação de grupos econômicos de fato.

3. Visão Geral do Projeto

Este estudo tem como objetivo central automatizar a identificação de grupos econômicos de fato no contexto da investigação patrimonial, proporcionando maior eficiência na responsabilização solidária de pessoas físicas e jurídicas associadas ao devedor. A abordagem proposta reduz o tempo e o esforço empregados por especialistas na localização dessas estruturas empresariais interconectadas, além de ampliar a precisão e a abrangência das investigações patrimoniais. Para atingir esse propósito, são empregadas técnicas avançadas de mineração de dados, modelagem de grafos e aprendizado de máquina.

A base de dados utilizada é o Cadastro Nacional de Pessoas Jurídicas (CNPJ), cujos registros são extraídos, transformados e carregados em uma estrutura de grafos. Esse processo é realizado por meio de um pipeline de *ETL* (Extração, Transformação e Carga), no qual os dados brutos, originalmente em arquivos CSV, são processados utilizando Apache Spark para permitir sua modelagem eficiente utilizando técnicas de processamento paralelo. A transformação inclui a padronização dos registros e a criação de relações explícitas entre empresas e indivíduos, representando-os como *nós* e seus vínculos societários como *arestas*. Essa modelagem possibilita não apenas a identificação de conexões formais entre entidades, mas também a inferência de relações implícitas.

Além das conexões documentadas no CNPJ, uma técnica essencial para a identificação de grupos econômicos de fato envolve a similaridade entre denominações empresariais. Utiliza-se *Sentence Transformers*, um modelo de aprendizado profundo pré-treinado, para calcular a proximidade semântica entre nomes empresariais e nomes fantasia. Essa abordagem permite identificar padrões que indicam possíveis relações econômicas entre empresas, mesmo quando não há vínculos societários formais regis-

trados. A combinação desse modelo com a estrutura de grafos amplia significativamente a capacidade de detectar conglomerados empresariais informais.

A definição dos critérios para identificação de grupos econômicos foi baseada na experiência de juízes especializados em recuperação de ativos, tanto na Justiça do Trabalho quanto na Justiça Federal. A heurística construída considera um conjunto de características frequentemente observadas em casos de desconsideração da personalidade jurídica, tais como a identidade de gestores, a formação de quadro societário com indivíduos recorrentes, a semelhança entre denominações sociais, a atuação complementar no mercado, o compartilhamento de infraestrutura e a independência meramente formal entre empresas [Lopes and Frattari 2015]. Esses critérios foram diretamente correlacionados com os dados do CNPJ, permitindo a formulação de um modelo automatizado para sugerir a existência de grupos econômicos de fato.

Para viabilizar a exploração eficiente dessas informações, foi desenvolvida uma interface interativa voltada a especialistas em investigação patrimonial. A ferramenta permite a visualização gráfica das conexões entre empresas e sócios, além de oferecer funcionalidades de busca avançada e filtragem para facilitar a navegação entre os dados. A interface também incorpora sugestões automáticas baseadas na heurística desenvolvida, permitindo que os investigadores patrimoniais explorem as relações econômicas de forma intuitiva e eficaz. Dessa maneira, a ferramenta contribui para a celeridade e fundamentação das análises, possibilitando uma investigação mais precisa e aprofundada sobre a formação de grupos econômicos e seus impactos na recuperação de ativos.

4. Metodologia Proposta

A metodologia proposta para a identificação de Grupos Econômicos de Fato e a desconsideração da personalidade jurídica é baseada em um entendimento do domínio do problema, utilizando dados do Cadastro Nacional de Pessoas Jurídicas (CNPJ) e um modelo heurístico desenvolvido com a orientação de especialistas – um Juiz Federal do Tribunal Regional Federal da 1^a Região e um Juiz do Trabalho do Tribunal Regional do Trabalho da 14^a Região.

A desconsideração da personalidade jurídica ocorre em casos de **abuso de personalidade jurídica**, caracterizado por elementos como **identidade de gestores comuns, quadro societário compartilhado, denominação social aproximada, atuação complementar, compartilhamento de bens e infraestrutura e independência meramente formal**. Esses fatores indicam um vínculo econômico substancial entre empresas, justificando a desconsideração e a responsabilização solidária.

A heurística para a identificação de Grupos Econômicos de Fato foi formulada com base nesses critérios, relacionando-os com os dados disponíveis no CNPJ. O objetivo é identificar padrões de interdependência entre entidades formalmente independentes, mas que operam como uma única estrutura econômica. O mapeamento desses critérios foi estruturado na Tabela 1, que apresenta a relação entre as características da desconsideração da personalidade jurídica e os respectivos dados do CNPJ.

4.1. Transformação dos Dados e Organização em Grafo

O processo de estruturação dos dados foi realizado através de um fluxo de **ETL (Extract, Transform, Load)**, permitindo a extração, transformação e carga dos dados em um for-

Tabela 1. Características para Identificação de Grupos Econômicos de Fato

Característica da Desconsideração	Dado do CNPJ
Identidade de gestores comuns	Representantes legais em comum
Formação de quadro societário comum	Sócios em comum
Denominação social aproximada	Semelhança entre nomes fantasia ou empresariais
Atuação complementar ou semelhante	CNAE similar
Compartilhamento de bens e infraestrutura	Endereço, e-mail, telefone em comum
Independência meramente formal	Combinação dos critérios acima

mato adequado para análise de grupos econômicos. Os dados foram inicialmente obtidos do *Portal de Transparência*, no formato **CSV (Comma-Separated Values)**.

Os arquivos CSV utilizados são os seguintes: **Empresas** (Contém informações cadastrais de cada pessoa jurídica); **Sócios** (Armazena dados sobre os sócios de cada empresa, incluindo tanto Pessoas Físicas quanto Jurídicas); **Estabelecimentos** (Representa unidades produtivas das empresas, incluindo matrizes e filiais). A Junção entre Empresas e Estabelecimento tem cerca de 65 milhões de linhas, enquanto a entidade Sócios tem menos de 25 milhões de registros.

Após a extração, os dados foram transformados para permitir uma modelagem mais intuitiva das conexões entre empresas e sócios. A organização foi realizada em uma estrutura de grafos, facilitando consultas e visualizações das inter-relações societárias. Cada nó representa uma entidade e as arestas simbolizam relações de sociedade.

Para representar matematicamente a estrutura descrita, adotamos a modelagem de um grafo híbrido, definido da seguinte forma:

$$G = (V, E) = (V_{PF} \cup V_{PJ}, E) \quad (1)$$

Onde:

O conjunto V representa os nós do grafo, composto por:

- V_{PF} : conjunto de nós que representam sócios pessoas físicas.
- V_{PJ} : conjunto de nós que representam pessoas jurídicas, resultantes da junção das tabelas *Empresa* e *Estabelecimento*.

O conjunto $E \subseteq V \times V$ representa as arestas do grafo, que correspondem às relações societárias entre os nós (vínculos entre sócios e empresas).

Essa estrutura permite representar tanto conexões formais quanto inferências indiretas entre entidades, facilitando a aplicação de técnicas de análise de grafos sobre os dados do CNPJ.

O processo de ETL foi conduzido em três etapas conforme demonstrado na figura 1.

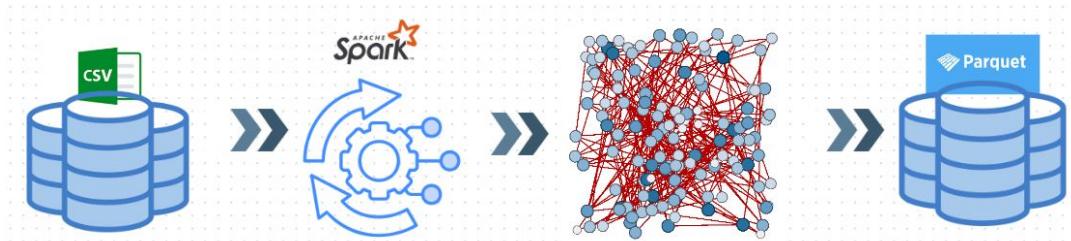


Figura 1. Fluxo do Processo ETL para estruturação dos dados

A escolha do *Apache Spark* e de sua ferramenta *GraphX* para a realização das atividades de extração e transformação dos dados foi motivada pela necessidade de lidar com grandes volumes de informações e executar computações paralelas em grafos de forma eficiente. O *Spark* é uma estrutura de processamento distribuído que oferece recursos robustos para escalabilidade e tolerância a falhas, aspectos essenciais ao trabalhar com os dados do CNPJ, que envolvem milhões de registros interconectados.

Para a persistência dos dados transformados, foi adotado o formato *Parquet*, que armazena os dados em colunas, proporcionando um armazenamento mais eficiente e uma leitura otimizada para consultas futuras. O *Parquet* é especialmente vantajoso para grandes volumes de dados, pois permite compressão eficiente e acesso rápido a colunas específicas. Essa abordagem é essencial para análises iterativas e exploratórias no grafo, garantindo um desempenho otimizado tanto na consulta quanto na recuperação dos dados.

4.2. Embeddings de Nome Empresarial

Um dos critérios adotados pela Heurística Investigativa é a localização de nomes empresariais semelhantes. Durante a fase de preparação e modelagem dos dados, utilizamos a biblioteca *SentenceTransformers* [SentenceTransformers 2024] para processar os embeddings de todos os nomes fantasia e, na ausência destes, do nome empresarial, armazenando-os em arquivos *Parquet*.

Os embeddings armazenados são utilizados para a avaliação de similaridade, com a biblioteca *Cosine Similarity* da *scikit learn* [Scikit-learn 2024], aplicada durante o processo de investigação das empresas-alvo. Essa abordagem permite a identificação eficiente de relações entre empresas que compartilham nomes semelhantes, facilitando a detecção de possíveis vínculos ocultos entre entidades jurídicas.

Para essa tarefa, foi selecionado o modelo de transformação de sentenças *paraphrase-multilingual-MiniLM-L12-v2* [Reimers and Gurevych 2019], devido à sua capacidade de gerar representações vetoriais densas e altamente discriminativas para textos curtos, como nomes empresariais. Esse modelo é eficaz em aplicações que envolvem busca e comparação de similaridade semântica, pois mapeia frases para um espaço vetorial onde textos com significados próximos tendem a estar localizados em regiões próximas, aprimorando a qualidade das inferências sobre relações empresariais.

4.3. Heurística Investigatória

A Heurística Investigatória formula um fluxo operacional com três estágios para a identificação de Grupos Econômicos de Fato, simulando as ações conduzidas por investigadores. Esse fluxo baseia-se exclusivamente na análise de dados abertos do CNPJ, garantindo que a investigação seja realizada de forma transparente e replicável.

Os investigadores seguem um conjunto de diretrizes estruturadas para localizar similaridades cadastrais entre entidades jurídicas. Essas diretrizes são fundamentadas nas características apresentadas na Tabela 1, que define os critérios adotados para estabelecer possíveis relações entre empresas e sócios.

A heurística desenvolvida permite identificar padrões de interdependência entre empresas que, formalmente, apresentam estruturas separadas, mas que, potencialmente, operam de maneira coordenada, seja por meio de vínculos societários, endereços comuns, gestão compartilhada ou nomes empresariais semanticamente similares.

Para dar início à execução do fluxo operacional da heurística, o investigador insere um ou mais números de CNPJ para consulta. Essas Pessoas Jurídicas, juntamente com todas as demais empresas formalmente vinculadas ao seu quadro societário, são classificadas como empresas-alvo, tornando-se o foco central da investigação.

4.3.1. Primeiro Estágio: Construção do Cluster Inicial

No primeiro estágio da heurística, os *DataFrames PySpark* contendo nós (empresas e sócios) e arestas (relações societárias) são carregados e estruturados na forma do grafo da equação 1. A construção inicial desse grafo permite identificar diretamente todas as conexões existentes entre a empresa-alvo e seus respectivos sócios.

A partir desse ponto, um segmento do universo do CNPJ é formado por meio da filtragem de empresas que compartilham características cadastrais com as empresas-alvo. As características de filtragem, que podem ser combinadas ou não são: **Endereço** (um ou mais CEP); **E-mail**; **número telefônico**; **CPF do responsável legal**.

Essa filtragem cria um subconjunto de empresas com maior probabilidade de pertencerem ao mesmo Grupo Econômico de Fato, reduzindo o universo de investigação para o próximo estágio.

4.3.2. Segundo Estágio: Análise de Similaridade Semântica

O segundo estágio expande a análise para identificar empresas que compartilham nomes empresariais, ou fantasia, semanticamente similares. Para isso, os *embeddings* de nomes previamente gerados são comparados utilizando a métrica de *Cosine Similarity*. A matriz de similaridade obtida compara o nome da empresa-alvo com os nomes das empresas do *cluster* gerado na fase anterior.

Empresas cujo escore de similaridade ultrapassa um determinado limiar (por exemplo, 90%) são mantidas na investigação, pois apresentam indícios de pertencer ao mesmo grupo econômico, mesmo que seus nomes não sejam semelhantes graficamente.

4.3.3. Terceiro Estágio: Consolidação dos Resultados

No estágio final, as empresas selecionadas nos estágios anteriores são organizadas em tabelas de comparação, permitindo uma análise estruturada das semelhanças e dos vínculos identificados. Esse formato facilita a visualização e interpretação das conexões entre empresas, fornecendo evidências mais robustas para a caracterização de Grupos Econômicos de Fato.

49/0001-62 - CONSTROI										
	CEP	Logradouro	CPF Responsável	Telefone 1	Email	CNAE	Sócios Comuns	Nome Sim	Nome Fantasia	Detalhe
	✓	✓	✓	✓	✓	✓	✓	100.00%	CONSTROI	▼
	✗	✗	✓	✓	✓	✓	✓	15.47%		▼
	✓	✓	✓	✓	✗	✓	✓	30.92%		▼
	✓	✓	✓	✓	✓	✓	✓	81.54%	IMOLAIT	▼
	✗	✗	✓	✓	✓	✓	✓	30.73%		▼
	✓	✓	✓	✓	✓	✓	✓	44.17%	GSA	▼
	✓	✗	✓	✓	✓	✓	✓	24.63%	GSA INCORPORACOES 03	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✗	✗	✓	✓	✓	✓	✓	68.58%	CINCOL	▼
	✓	✗	✓	✓	✓	✗	✓	23.32%	GSA INCORPORACOES 02	▼
	✓	✗	✓	✓	✓	✓	✓	16.95%	GSA INCORPORACOES 04	▼

Figura 2. Tabela Comparativa de Empresas

A apresentação gráfica das características utilizadas para identificar Grupos Econômicos de Fato é essencial para facilitar a análise comparativa entre as **Empresas-Alvo** e as empresas localizadas na investigação. A interface gera uma tabela estruturada que destaca todas as características relevantes do CNPJ alvo em relação às empresas identificadas, proporcionando uma visão clara e objetiva das semelhanças e conexões existentes. Esse formato visual, exemplificado na Figura 2, representa um resultado real da aplicação da heurística sobre dados do CNPJ e permite que especialistas identifiquem padrões e relações de forma mais intuitiva, agilizando a interpretação dos dados.

5. Resultado

Para avaliar a eficácia da heurística proposta na identificação de Grupos Econômicos de Fato, foi formado um grupo de 12 especialistas em recuperação de ativos, composto por juízes do trabalho e servidores públicos federais. Esses especialistas conduziram a avaliação da ferramenta ao longo de 14 dias, período em que foram realizadas 160 investigações (consultas).

5.1. Avaliação da Heurística

Cada investigação consistiu na análise de um conjunto de entidades jurídicas e suas respectivas relações, utilizando a ferramenta proposta. Após cada consulta, os especialistas foram convidados a classificar os resultados com base nos seguintes critérios, que foram associados à matriz de confusão:

1. Localizou empresas que não faziam parte do quadro societário formalizado, mas pertenciam ao Grupo Econômico de Fato (Verdadeiro Positivo - TP), indicando que a ferramenta conseguiu identificar corretamente empresas que fazem parte de um grupo econômico sem registro societário formal.
2. Localizou empresas que não faziam parte do quadro societário formalizado, mas não pertenciam ao Grupo Econômico de Fato (Falso Positivo - FP), apontando que a ferramenta sugeriu incorretamente empresas que não deveriam pertencer ao grupo.
3. Não localizou empresas que não faziam parte do quadro societário, e efetivamente não existiam empresas formadoras de grupos econômicos de fato (Verdadeiro Negativo - TN), demonstrando que a ferramenta corretamente evitou sugerir entidades irrelevantes.
4. Não localizou empresas que não faziam parte do quadro societário, mas havia empresas formadoras de grupos econômicos de fato (Falso Negativo - FN), evidenciando falhas na identificação de entidades que deveriam ser retornadas como parte do grupo econômico.

Os resultados obtidos a partir das avaliações dos especialistas foram:

- **Precisão (Precision):** 66.67%. Representa a proporção de empresas corretamente identificadas como parte do Grupo Econômico de Fato em relação ao total de empresas sugeridas, mostrando que, embora boa, ainda há espaço para melhorar a assertividade.
- **Revocação (Recall):** 85.71%. Mede a capacidade da ferramenta de encontrar corretamente todas as empresas que pertenciam ao Grupo Econômico de Fato, mostrando que a maior parte das entidades relevantes foram identificadas.
- **F1-Score:** 75%. Reflete o equilíbrio entre precisão e recall, fornecendo um indicador geral de qualidade da classificação.

Os resultados indicam que a heurística utilizada pela ferramenta apresenta alta capacidade de recuperação de grupos econômicos de fato, evidenciada pelo recall de 85.71%. No entanto, a precisão de 66.67% sugere que algumas empresas classificadas como pertencentes ao grupo não correspondem a essa realidade, o que pode indicar a necessidade de refinamento nos critérios da heurística. O F1-Score de 75% reforça que há um equilíbrio razoável entre precisão e revocação, mas ainda há espaço para melhorias, especialmente na redução de falsos positivos, para aumentar a confiabilidade das investigações patrimoniais.

Embora a heurística tenha demonstrado alta capacidade de recuperação (Recall de 85,71%), a precisão de 66,67% indica a presença relevante de falsos positivos. Esse fator pode impactar negativamente a análise investigativa, ao incluir empresas que não pertencem ao grupo econômico de fato, exigindo maior esforço dos especialistas para triagem manual dos resultados. Ainda que o ambiente de uso envolva verificação humana, a redução dessa taxa é desejável para minimizar distrações e evitar a diluição de indícios mais robustos.

5.2. Avaliação da ferramenta

Os especialistas foram convidados ainda a responderem a um total de 12 perguntas, divididas em quatro categorias principais. As sentenças e respostas utilizaram uma escala Likert de 1 a 5, onde 1 representa discordância total e 5 representa concordância total:

- **Precisão e Eficácia:** Avalia se a ferramenta melhora a identificação de grupos econômicos e a descoberta de conexões patrimoniais ocultas.
- **Usabilidade e Eficiência:** Mede a facilidade de uso da interface, o tempo de resposta e a disposição visual dos resultados.
- **Impacto na Investigação Patrimonial:** Examina o quanto a ferramenta contribui para a execução processual e ampliação da análise patrimonial.
- **Comparação com Outras Ferramentas:** Avalia a relevância e confiabilidade das informações em relação a outras abordagens investigativas disponíveis.

As respostas foram consolidadas em uma média por categoria e apresentadas na Figura 3) com barras e erro padrão para indicar variação estatística.

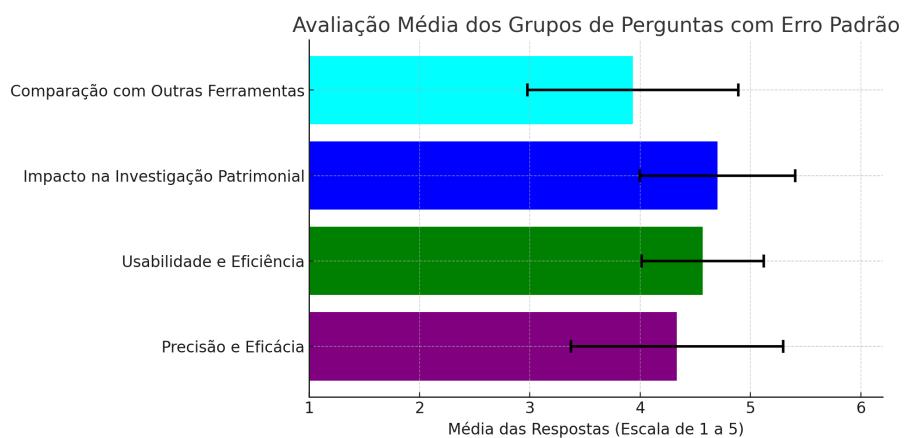


Figura 3. Avaliação Média dos Grupos de Perguntas com Erro Padrão

A categoria **Impacto na Investigação Patrimonial** obteve a maior média entre as respostas (**4,70**), indicando uma percepção altamente positiva sobre a capacidade da ferramenta em aprimorar as investigações patrimoniais. Além disso, seu intervalo de confiança (IC 95%) demonstra que a maioria das respostas está concentrada em valores elevados.

A **Usabilidade e Eficiência** também apresentou uma média elevada (**4,57**), com a menor variabilidade entre todas as categorias, sugerindo que os especialistas tiveram uma experiência bastante uniforme ao interagir com a ferramenta.

Já a categoria **Precisão e Eficácia** mostrou uma média positiva (**4,33**), mas com uma variabilidade um pouco maior, o que pode indicar que alguns especialistas tiveram experiências diferentes quanto à exatidão da ferramenta na identificação de grupos econômicos.

Por fim, a **Comparação com Outras Ferramentas** teve a menor média (**3,93**) e maior erro padrão, sugerindo que os especialistas podem ter dúvidas sobre a superioridade da ferramenta em relação a outras soluções investigativas existentes.

Além das perguntas de avaliação categorizadas, os especialistas responderam à métrica Net Promoter Score (NPS) [Reichheld 2003], que mede a probabilidade de recomendação da ferramenta para outros profissionais da área. O cálculo do NPS resultou em um valor de 80, o que indica um alto nível de satisfação e lealdade dos usuários,

sugerindo que a maioria dos especialistas que testaram a ferramenta estaria disposta a recomendá-la para outros profissionais da área.

Esse valor indica que a maioria dos especialistas que testaram a ferramenta não apenas aprovam seu uso, mas também estariam dispostos a recomendá-la a outros profissionais. Esse resultado reforça a percepção positiva observada nas respostas anteriores e sugere um alto nível de confiança na ferramenta para fins investigativos.

6. Conclusão

A análise dos resultados indica que a ferramenta investigativa proposta apresenta um desempenho promissor na identificação de grupos econômicos de fato. A heurística utilizada alcançou um Revocação (*Recall*) de 85,71%, evidenciando sua capacidade de recuperar a maior parte das entidades relevantes para a investigação patrimonial. No entanto, a Precisão ((*Precision*) de 66,67% indica que algumas empresas foram classificadas erroneamente como pertencentes ao grupo econômico, sugerindo a necessidade de ajustes nos critérios de inferência para aprimorar a assertividade da ferramenta. O F1-Score de 75% reforça que há um equilíbrio razoável entre *precision* e *recall*, permitindo que a ferramenta atue de forma confiável no suporte a investigações patrimoniais.

Além do desempenho técnico da heurística, a avaliação dos especialistas revelou uma percepção positiva da ferramenta. A categoria de maior destaque foi o impacto na investigação patrimonial, com uma média de 4,70 na escala (*Likert*, indicando que a ferramenta é vista como um recurso valioso para otimizar a análise de ativos e fortalecer os processos investigativos. A usabilidade e eficiência também foram bem avaliadas, sugerindo que a interface desenvolvida permite uma interação intuitiva e eficaz.

Outro aspecto relevante foi a aceitação geral da ferramenta pelos especialistas, medida pelo *Net Promoter Score (NPS)*, que atingiu um valor de 80. Esse resultado reflete um alto nível de satisfação e confiança, indicando que a maioria dos especialistas recomendaria a ferramenta para outros profissionais da área. Essa aprovação sugere que a solução proposta pode ser amplamente adotada como um recurso útil no contexto da recuperação de ativos e identificação de fraudes.

A ausência de uma base de dados consolidada com resultados de investigações patrimoniais anteriores limitou a possibilidade de uma avaliação mais ampla da heurística. Dessa forma, a validação da ferramenta foi conduzida com base na operação real do sistema por especialistas, que utilizaram seus conhecimentos e experiência para interpretar os resultados obtidos. Esse tipo de avaliação, embora qualitativa, fornece uma perspectiva prática sobre a aplicabilidade da ferramenta no contexto investigativo.

Embora esta pesquisa tenha focado na aplicação da heurística para investigação patrimonial no Poder Judiciário, a metodologia proposta possui ampla aplicabilidade dentro do Executivo Federal e de outras esferas governamentais. Órgãos de fiscalização e controle, como a Receita Federal, podem utilizá-la para detecção de fraudes fiscais e ocultação de patrimônio por meio de estruturas societárias complexas. Além disso, o Conselho Administrativo de Defesa Econômica (CADE) pode empregar a abordagem para validar as informações apresentadas por empresas ao submeterem formulários de notificação, bem como auxiliar na dosimetria de multas e penalidades aplicadas em casos de infração à ordem econômica.

Para trabalhos futuros, a heurística será aprimorada com o objetivo de reduzir a taxa de falsos positivos, incorporando características adicionais sugeridas pelos especialistas, incluindo a realização de clusters baseados em contadores das empresas. Essa abordagem permitirá uma inferência mais precisa das conexões entre entidades, refinando a detecção de grupos econômicos de fato e aumentando a confiabilidade da ferramenta. Além disso, pretende-se explorar técnicas avançadas de descoberta de ligações (Link Prediction) em Redes Neurais em Grafos (GNNs), amplamente aplicadas em redes sociais. Especificamente, a abordagem Graph Attention Networks (GAT) será investigada, permitindo a ponderação adaptativa das conexões entre entidades, priorizando relações mais relevantes.

Por fim, a contribuição da ferramenta para a Computação Aplicada em Governo Eletrônico é evidente. A automação da identificação de grupos econômicos de fato aborda um dos principais gargalos processuais do judiciário, possibilitando uma atuação mais ágil e fundamentada na recuperação de ativos. Ao oferecer um meio estruturado para a identificação de relações empresariais ocultas, a ferramenta fortalece a governança digital, promovendo maior transparência e eficiência na atuação do sistema judiciário.

Referências

- Brandão, M. A., Reis, A. P., Mendes, B. M., De Almeida, C. A. B., Oliveira, G. P., Hott, H., Gomide, L. D., Costa, L. L., Silva, M. O., Lacerda, A., et al. (2024). Plus: A semi-automated pipeline for fraud detection in public bids. *Digital Government: Research and Practice*, 5(1):1–16.
- de Carvalho, Á. A. B. and Holanda Filho, R. (2024). Detecting fraud in public acquisition of brazilian government with an analytical approach. *Procedia Computer Science*, 238:248–254.
- Lopes, C. S. B. and Frattari, R. (2015). Solidariedade tributária e grupos econômicos. *Revista de Direito Tributário e Financeiro*, 1(1):581–603.
- Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, 81(12):46–55.
- Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Santos, S. G. d. N., Elias, V. C. d. S., and Capelletti, L. R. (2024). Desconsideração da personalidade jurídica na execução fiscal. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 10(5):5422–5431.
- Scikit-learn (2024). Scikit-learn: Machine learning in python. Último acesso em 7 de março de 2025.
- SentenceTransformers (2024). SentenceTransformers: Pretrained models for sentence embeddings. Último acesso em 7 de março de 2025.