

**Desafios na identificação automática de grupos econômicos para aprimorar a gestão
de riscos e controle de crédito.**

Gustavo de Paula Aguiar Araújo ^{1*}; Diego de Oliveira Da Cunha

¹ Universidade de São Paulo - Bacharelado em Sistemas da informação. Rua Arlindo Bétio, 1000 - Ermelino
Matarazzo, São Paulo - SP, 03828-000

¹ Faculdade de Informática e Administração Paulista – MBA Engenharia de Software aplicado em SOA. Faculdade
de Informática e Administração Paulista.

Desafios na identificação automática de grupos econômicos para aprimorar a gestão de riscos e controle de crédito.

Resumo

A gestão de risco de crédito é essencial para a estabilidade das instituições financeiras e para a economia como um todo. O trabalho propõe a criação de um algoritmo que permite identificar estruturas de Grupos Econômicos (GEs) a partir de vínculo societário, com o objetivo de melhorar e ampliar a análise de risco de crédito nas instituições financeiras [IFs], partindo de uma análise mais profunda dos envolvidos na gestão do crédito.

Utilizando conceitos da teoria dos grafos, o algoritmo mapeia as relações entre empresas e sócios, e disponibiliza a visão das conexões que existem entre as empresas e seus sócios.

Os resultados mostram que a abordagem permite uma avaliação mais precisa e abrangente do risco de crédito, facilitando a gestão de carteiras de crédito de forma proativa e possibilitando aumentar a capacidade das [IFs] na tomada de decisões ao disponibilizar e gerir o crédito.

Palavras-chave: carteira de crédito; vínculo entre empresas; grafos por relação de empresas;

Introdução

Há muito tempo, instituições financeiras [IFs] e não financeiras têm se preocupado com a gestão eficaz do risco de crédito (IPEA, 2006). Os eventos complexos que resultaram em impactos significativos nas economias e sociedades, tais como, recessões e crises financeiras, imprimiram desafios que forçaram a necessidade de controlar e desenvolver estratégias para garantir uma menor exposição ao risco de crédito, maximização dos lucros e manter a sustentabilidade no mercado (Jassé, 2020).

Uma das principais atividades bancárias é a concessão de crédito, a qual possibilita a expansão natural do mercado através da troca de recursos financeiros entre provedor e tomador. O fato da exposição a inadimplência, ou seja, do descumprimento do acordo de retorno dos recursos financeiros emprestados ao tomador pelo provedor, exige ao provedor conhecer melhor o tomador, antes de assumir o risco da transferência do recurso financeiro. Dessa forma, a necessidade de controlar e garantir o menor risco de concessão de crédito é fundamental e essencial para garantir a saúde do ambiente financeiro atual e futuro. (Van Gestel e Baesens, 2008)

Comentado [DD1]: O resumo deixamos para o final.

Comentado [DD2]: Quem afirmou isso?

Comentado [GA3R2]: Ajustado com a referência

Comentado [DD4]: Não tem referências mais recentes?

Comentado [GA5R4]: Nova referencia que faz menção ao conteúdo e é mais atual

A gestão de risco de crédito, ao longo do tempo, teve diferentes marcos que contribuíram para seu amadurecimento. Até o início do século XX, a análise e aprovação de crédito ainda era feita subjetivamente, dependendo somente do julgamento de analistas. Esse método, além de não utilizar critérios objetivos, era moroso e não considerava uma análise ampla, com todas as variáveis da exposição ao risco de crédito para as IFs, tornando-se de certa forma subjetivo (Camargos, 2012). Com o advento de novas tecnologias e modelos estatísticos, foi possível desenvolver maneiras mais eficientes para controlar carteiras de créditos, criando mecanismos para conhecer melhor seus clientes.

É nesse contexto em que é introduzido o conceito de grupo econômico [GE]. Essa terminologia é dada ao conjunto de empresas que, estão interligadas por relações contratuais, cuja propriedade pertence a indivíduos ou instituições, que exercem o controle efetivo sobre essas empresas (Gonçalves, 1991). A identificação e correlação dos indivíduos de um GE é de grande importância para aprimorar os modelos de precificação de risco de crédito, garantindo uma maior pluralidade nas informações dos envolvidos na concessão de crédito.

Desta forma, em 2017 o Banco Central Brasileiro [BACEN] propõe a regulamentação do controle de contrapartes para fins de gerenciamento de risco. A resolução propõe e exige que as IFs realizem o controle de contrapartes conectadas que compartilhem risco de crédito, documentando os critérios utilizados para identificação de cada indivíduo pertencente ao GE (BACEN, 2017, Art. 22). A norma é prerrogativa para definições das diretrizes e requisitos para a estrutura de gerenciamento de riscos e apetite por riscos nas instituições financeiras, visando a promover uma gestão prudente e eficaz das suas carteiras.

É neste contexto que o presente trabalho tenta demonstrar que é possível criar maneiras de identificar melhor os vínculos entre empresas e sócios, desenvolvendo um algoritmo que auxilie a identificação de clientes envolvidos na concessão de crédito. Foca na identificação de possíveis estruturas de GEs para facilitar a identificação de relações entre empresas e seus sócios, o que possibilita uma melhor condução na análise de risco de crédito nas IFs. Ao identificar essas relações de maneira automatizada, é possível tomar decisões mais informadas e reduzir a exposição ao risco de crédito. Em outras palavras, o artigo descreve como é possível criar esse modelo e com isso facilita identificar as ligações entre as contrapartes envolvidas na concessão de crédito

Material e Métodos

Nesta sessão do trabalho, é apresentado o processo de criação e organização da construção do algoritmo de agrupamento de GEs, assim como as ferramentas e desafios do processo. Em seguida, são apresentados alguns dos resultados do algoritmo.

Comentado [DD6]: Mas antes, qual seu objetivo geral?

Comentado [GA7R6]: Ajustei o parágrafo abaixo, com a ideia do projeto. Por avalie se ficou claro, se não me ajude a ajustar esse ponto?

Metadados

Os dados privados de cada indivíduo foram preservados e somente foram utilizados dados públicos de participação societária de empresas disponibilizados pelo Ministério da Economia [ME]. Os dados utilizados para a realização das análises do trabalho, são de propriedade pública disponibilizados pela Secretaria Especial da Receita Federal do Brasil [RFB], os quais podem ser acessados através do canal Dados Abertos (Governo do Brasil, 2024).

Devido a quantidade de informação disponível, foi necessário a obtenção de uma ferramenta a qual pudesse disponibilizar as informações de forma massiva. Desta forma, foi utilizado as informações centralizadas pelo site *base dos dados* ([BASE DOS DADOS](#)~~base dos dados~~, 2024), o qual, de forma independente, centraliza a ingestão dos dados públicos da receita federal e disponibiliza os dados publicamente na ferramenta *BigQuery* do *Google*. Essa ferramenta, disponibilizada, possibilita a análise de dados gerenciadas em larga escala ([GOOGLE](#)~~Google~~ *BigQuery*~~GQUERY~~, 2023). A utilização da ferramenta foi essencial para o processo de Data Wrangling, etapa fundamental para organizar os dados que foram utilizados para a modelagem do algoritmo, visto que foi necessário a análise de toda composição societário disponível.

Data Wrangling- tratamento de dados

Para obtenção dos dados foi necessário criar uma estrutura de análise espelho das bases fornecidas pela RFB, a fim de estudar as estruturas e vínculos entre as bases e suas respectivas observações (Imagem 1 do Apêndice A).

A obtenção dos dados em larga escala trouxe maior complexidade para realizar a gestão dos dados, pois com a existência de milhares de observações disponíveis a execução ficou prejudicada devido ao tratamento dos dados em hardware não escalonável. Sendo assim, para o projeto, se optou a utilização da ingestão de dados diretamente do *BigQuery* com a persistência de dados em uma base de dados local MySQL.

Como mostrado na Figura 1, com a utilização dos dados mascarados para os sócios, foi necessário a criação de uma chave de referência entre as empresas e sócios, para que fosse possível criar um vínculo único entre as observações. Para a visão das empresas, foi concatenado o número básico do CNPJ com o nome da empresa, já para os sócios, seguiu-se a mesma lógica, concatenando o nome do sócio com um algoritmo de mascaramento e a

Comentado [DD8]: Tem misturando formatação nova com a antiga

Comentado [GA9R8]: Ajustei as referencias nesse ponto. Havia incluído uma referencia ao meu proprio trabalho e achie q não fazia sentido.

base do CPF ou CNPJ mascarado. Essas chaves possibilitam alimentar os grafos e criar as referências e vínculos entre os integrantes do GE.

empresas	socios
08169126-FR. COMERCIO DE CALCADOS E ACESSORI...	ACMDO*****.***364671**
08169126-FR. COMERCIO DE CALCADOS E ACESSORI...	AJM*****.***599401**
24705950-RESTAURANTE CANTINHO DO SABOR LTDA	AJM*****.***599401**
24705950-RESTAURANTE CANTINHO DO SABOR LTDA	DM*****.***643061**

Figura 1 - Chaves únicas de referência entre empresas e sócios

Fonte: Resultados originais da pesquisa

Grafos

A teoria dos grafos é uma área da matemática que se dedica ao estudo de diagramas, que são estruturas matemáticas utilizadas para representar relações entre indivíduos. Um diagrama é composto por nós (também conhecidos como vértices) ligados por linhas (também chamadas de arestas ou conexões). Os diagramas podem ser não direcionados, nos quais as linhas conectam os nós simetricamente, ou direcionados, nos quais as linhas conectam os nós assimetricamente (WILSON, 1996).

Em grafos direcionados, as arestas possuem uma direção clara, indicando um ponto de partida e um ponto de chegada. Isso é útil para representar relações unidirecionais, ou seja, relações que partem sempre de um nó para outro, demonstrando um fluxo contínuo de ligação entre nós de acordo com o que é demonstrado na Figura 2. Observa-se que o nó A se liga ao nó B e o B ao nó C, porém não existe nenhuma ligação ou vínculo do B para o A, do C para o B e nem do C para o A, ou seja, são vínculos unidirecionais. Rotas de entregas em aplicativos de entrega, pode ser considerado um exemplo da utilização desse tipo de abordagem.

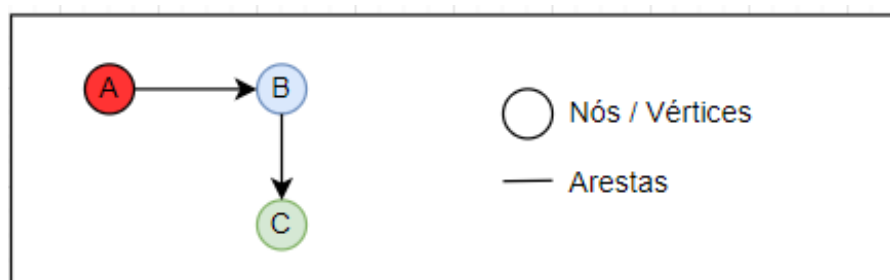


Figura 2. Grafo direcionado

Fonte: Resultados originais da pesquisa

Já em grafos não direcionados, as arestas não têm direção, indicando uma relação bidirecional ou mútua entre os vértices. Esse tipo de grafo é usado por exemplo, para representar redes sociais ou conexões físicas, onde a relação entre os nós é simétrica.

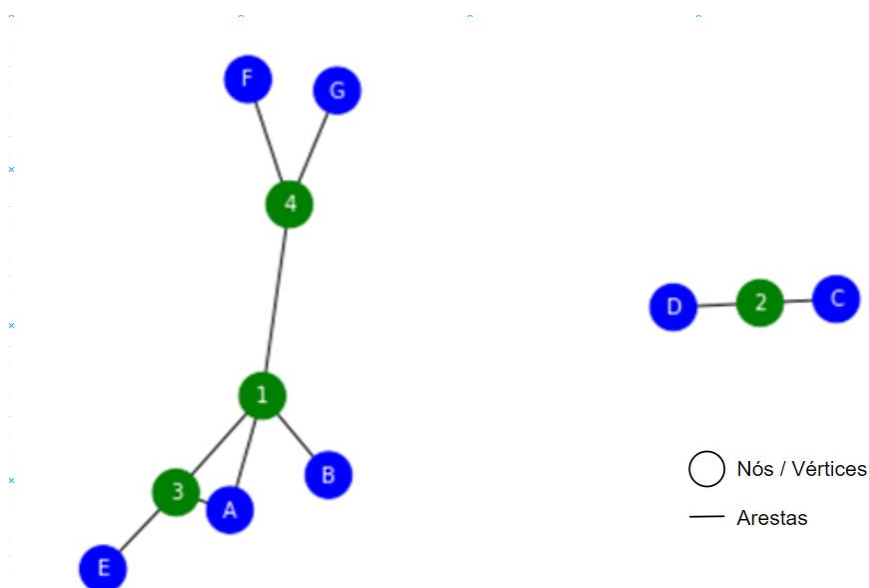


Figura 3. Estrutura de Grado não direcionado

Fonte: Resultados originais da pesquisa

Método

Após a modelagem dos dados, foi desenvolvido métodos para criação de subgrafos gerados pela interação das relações entre empresas e sócios. Na figura 1, pode-se avaliar um exemplo de como são identificadas as relações entre a empresa 'FR Comercio' com o 'Restaurante Catinho' através do sócio 'AJM'. A execução parte de três Dataframes¹, descritos

¹ DataFrame: Estrutura de dados bidimensional utilizada principalmente em bibliotecas como o 'pandas' no Python, que permite a manipulação de dados tabulares de forma eficiente. Um DataFrame é comparável a uma tabela em uma base de dados ou a uma planilha, onde os dados são organizados em linhas e colunas, podendo conter diferentes tipos de dados.

como empresas, sócios e relação empresa e sócio. Em seguida é executado massivamente a validação de relacionamentos entre as empresas e sócios, criando e associando as arestas dos subgrafos. A figura 4, ilustra como parte dos conjuntos, denominados GE, foram gerados após a execução do algoritmo.

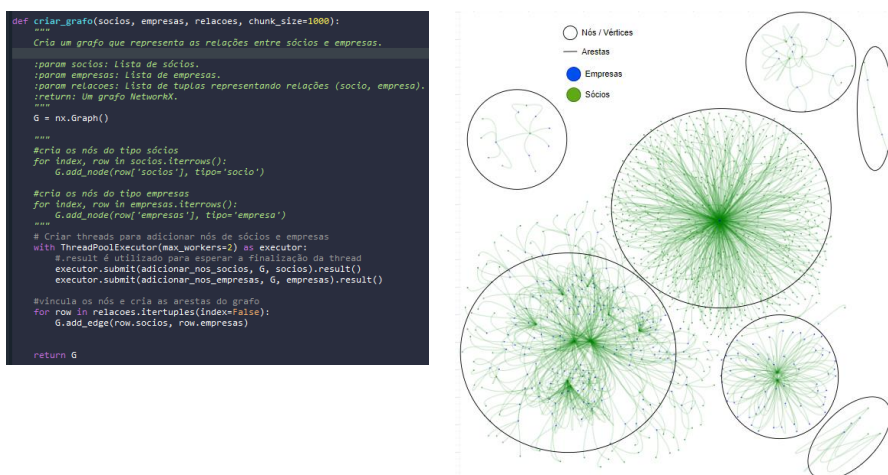


Figura 4. Grafo de referência

Fonte: Resultados originais da pesquisa

Resultados Preliminares

A seguir são apresentados os resultados do trabalho já com as imagens exportadas pelo algoritmo e como é possível identificar os vínculos entre sócios e empresas.

Essas estruturas de grafos podem ser abstraídas a uma relação entre empresas e sócios, pois sempre uma empresa possui um ou mais de um sócio responsável pela gestão da empresa. Desta forma, é possível dizer que a relação entre os indivíduos em um grafo não direcionado pode formar uma estrutura de GE, conforme visto na figura 5.

Comentado [DD10]: Algum autor corrobora com essa afirmação?

Comentado [GA11R10]: Aqui ajustei para tentar deixar claro q essa é uma ideia nova de identificação de grupos economicos.
Não consigo fazer referente a outros projetos.

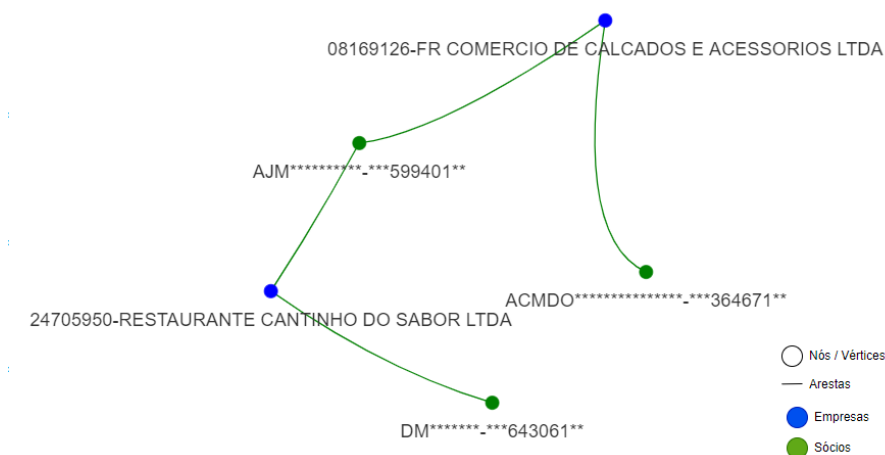


Figura 5. Grupo Econômico – Vínculo Societário

Fonte: Resultados originais da pesquisa

Na figura 5, pode-se visualizar as referências entre as empresas e sócios. É possível identificar um vínculo entre as empresas 'FR Comercio' e a 'Restaurante Cantinho' através do nó 'AJM*'. Esse vínculo entre as empresas, por sua vez, demonstra que a análise isoladamente das empresas e sócios pode ser insuficiente ao avaliar o risco de crédito envolvido. Os vínculos entre indivíduos trás maior visibilidade na identificação das características dos envolvidos na análise de crédito, já que dessa forma é possível identificar o poder de inferência positiva ou negativa entre os envolvidos.

Por exemplo, supondo que a empresa 'Restaurante Cantinho', visto na figura 5, seja uma empresa que possua alguma restrição bancária, tal como uma desonra de contrato. O fato de um dos indivíduos do grupo não possuir boa relação com o crédito, pode criar uma situação negativa para todo o GE, ou seja, ao disponibilizar crédito para empresa 'FR Comercio' é necessário ter maior cautela já que existe alguma restrição ativa dentro do GE.

Essa abordagem deve ser relevante com a adição de outras variáveis, ou seja, a identificação das relações entre os indivíduos, não deve ser necessariamente considerado como a única forma da análise do risco de crédito envolvido, mas já garante maior visibilidade na avaliação do perfil do GE.

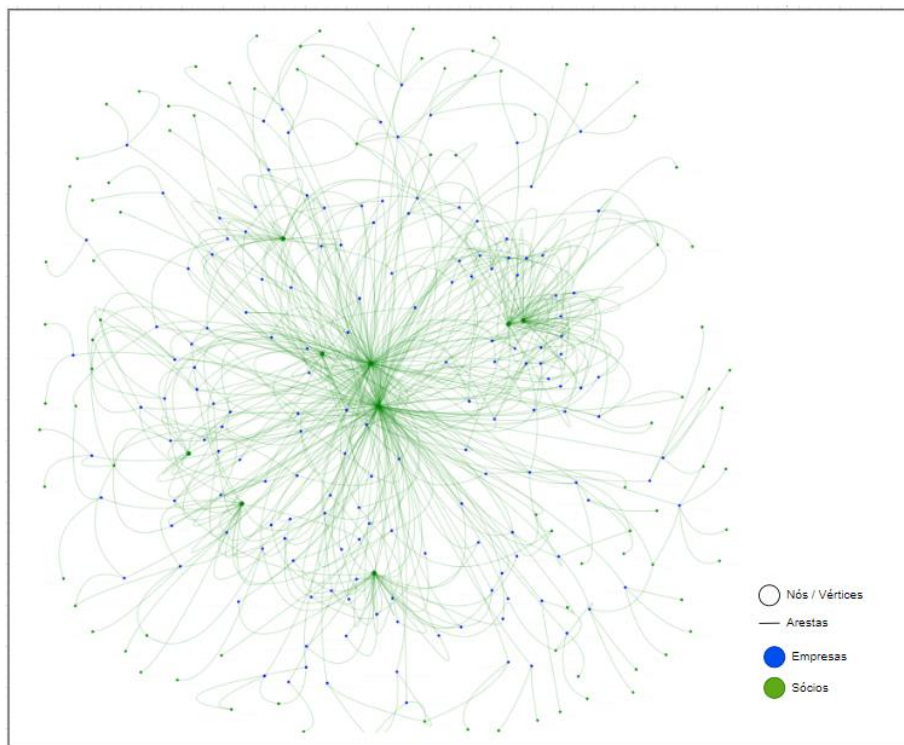


Figura 6. Grupo Econômico – Vínculo Societário com grande ramificação

Fonte: Resultados originais da pesquisa

A relação dos sócios está diretamente relacionada com a capacidade de tomada de decisão na empresa e por isso sócios com pouca participação societária, pode não representar grande impacto ao risco envolvido. Percebe-se que ao considerar somente o vínculo entre as empresas e sócios em um GE, podemos ter grupos com muitos vínculos como visto na figura 6 e 7, trazendo talvez uma visão ofuscada em relação ao risco agregado, já que mesmo havendo vínculos entre as empresas e sócios, a distância entre os nós, pode ser muito grande e com pouca relevância para avaliação do risco de crédito.

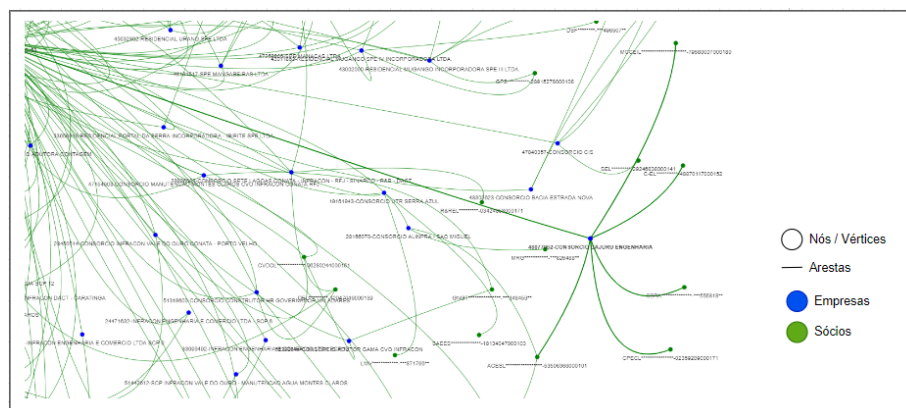


Figura 7. Grupo Econômico – Vínculo Societário – Zoom vínculo com muitas arestas e distante

Fonte: Resultados originais da pesquisa

Existe também uma abordagem para empresas compostas muitas arestas, as quais podem ter uma grande capacidade de criação de novos vínculos, como visto na figura 6 e 7. Nesses casos podemos considerar que a identificação de forma rápida e proativa dos novos vínculos dá maior dinamismo as IFs ao avaliar o risco de crédito inferido permitindo uma reavaliação do nível de risco proposto para cada GE.

Conclusão

Concessão de crédito e a gestão de risco são fundamentais para assegurar a estabilidade e a sustentabilidade das IFs, pois são basicamente o núcleo de seus negócios. A evolução das abordagens na avaliação de crédito, impulsionada por inovações associadas ao incremento de tecnologias e de novos modelos, tem proporcionado uma compreensão mais aprofundada das interações econômicas e sociais que afetam a concessão de crédito. Ao introduzir o conceito de GE na análise de risco de crédito, percebemos uma otimização nos métodos de avaliação de risco, acrescentam maior segurança para atender as diretrizes definidas pelo BACEN.

Neste estudo, com o desenvolvimento do algoritmo que simplifica a identificação das estruturas dos GEs, evidenciando os laços entre empresas e seus sócios partindo das informações sobre participações societárias, foi possível criar visões das conexões entre os indivíduos, trazendo maior capacidade e clareza na avaliação de risco de crédito, ao identificar melhor os clientes. Essa afirmação é corroborada indiretamente com a análise demonstrada

por Jesse (2020), que discorre sobre como medir e controlar o risco de carteiras de crédito, utilizando métodos quantitativos e qualitativos. Segundo o autor, ao utilizar o método quantitativo, é possível identificar variáveis como renda, taxas de juros aplicadas, prazos, histórico de inadimplência, liquidez do tomador, que podem indicar que as contrapartes envolvidas no contrato podem não respeitar os acordos firmados. Já em uma visão qualitativa, é inferido um perfil ao tomador, baseado em suas características, tais como, capacidade de honra, vínculos empregatícios legais (ex: trabalho análogo a escravidão), outros contratos em dia ou em atraso etc.

Assim, as IFs precisam investir no controle da gestão de crédito e riscos. Essa gestão é realizada utilizando modelos estatísticos, algoritmos de aprendizado de máquina e técnicas avançadas para prever o risco associado a diferentes tipos de clientes e operações de crédito. É nesse contexto que o trabalho apresentou uma abordagem para identificar automaticamente a formação de Grupos Econômicos. Esses grupos, por estarem interligados economicamente, compartilham responsabilidades e riscos. Portanto, uma análise eficaz de GEs pode revelar potenciais exposições ao risco de crédito que, de outra forma, passariam despercebidas em uma avaliação isolada das variáveis quantitativas e qualitativas de cada empresa.

A automatização da identificação de GEs pode trazer inúmeros benefícios para a gestão de riscos em instituições financeiras. Em primeiro lugar, a automatização acelera o processo de análise, permitindo que grandes volumes de dados sejam processados em tempo real, possibilitando decisões mais rápidas e precisas. O reconhecimento automático de relações entre empresas e sócios dentro de um GE melhora a visibilidade sobre as possíveis interdependências financeiras, evitando que bancos concedam crédito a empresas cujo grupo pode estar comprometido, reduzindo assim a exposição ao risco e melhorando a alocação de capital.

A utilização e análise dos grafos, tanto dirigidos quanto não dirigidos, para a construção do algoritmo permitiu uma representação mais clara e objetiva dessas relações, ressaltando a importância de levar em conta os vínculos indiretos/indiretos e a influência mútua dos envolvidos no processo de concessão de crédito. Por outro lado, se a identificação de grupos econômicos não for precisa, há o risco de sobrecarregar a análise de crédito com informações irrelevantes, gerando confusão sobre quais conexões realmente afetam o risco financeiro. Isso pode levar a decisões inadequadas, seja recusando crédito a empresas financeiramente sólidas ou concedendo a grupos de alto risco sem o devido cuidado.

Portanto, a identificação de GE por grafos, não pode ser a única abordagem, pois existe a possibilidade de que o algoritmo, por focar somente na identificação de relações entre sócios e empresas, apresente uma visão incompleta dos reais envolvidos na operação de crédito. Para aprimorar esse processo, sugere-se o desenvolvimento de algoritmos que

incorporem variáveis adicionais e métodos de análise preditiva, como análise de redes complexas, com o objetivo de gerar insights mais precisos sobre as estruturas de risco e a identificação de sócios com influência decisiva no grupo. Um exemplo disso é visto quando existem sócios minoritários presentes na formação proposta pelo algoritmo e esses sócios podem não possuir poder de decisão nas empresas e não representam positivamente ou negativamente no cálculo de variáveis da concessão e risco de crédito.

Além disso, é possível vislumbrar novas direções para pesquisas futuras que podem complementar o algoritmo atual. A integração de dados alternativos, como redes sociais corporativas ou transações financeiras em tempo real, pode aumentar a precisão da identificação de GEs. O uso de algoritmos de aprendizado de máquina para prever a formação de novos grupos, ou ainda a inclusão de variáveis qualitativas sobre a reputação de sócios, são algumas direções que poderiam tornar essa ferramenta mais robusta.

Em resumo, a análise automatizada de grupos econômicos, quando bem implementada, pode otimizar a gestão de riscos em instituições financeiras, proporcionando uma visão mais completa e eficiente das relações econômicas entre os tomadores de crédito e seus sócios. Ou seja, a implementação de modelos de grafos não pode ser considerada como o único meio de identificação de GE, mas sim um ponto de partida identificar de forma proativa as conexões entre empresas e sócios, que possibilita as IFs se antecipar e tomarem melhores decisões na concessão e gestão de crédito.

Agradecimentos

Agradeço e dedico o esforço deste trabalho aos meus familiares e amigos que sempre me incentivaram a crescer profissionalmente e pessoalmente.

Referências

Base dos Dados. 2023. Quadro Societário CNPJ. Disponível em: <https://basedosdados.org/dataset/e43f0d5b-43cf-4bfb-8d90-c38a4e0d7c4f?table=81272674-f522-4e43-a70b-05bf46f0a163>. Acesso em: 10 mar. 2024.

Banco Central do Brasil [BACEN]. 2017. Resolução nº 4557, de 23 de fevereiro. Dispõe sobre a estrutura de gerenciamento de riscos e apetite por riscos nas instituições financeiras. Disponível em: https://normativos.bcb.gov.br/Lists/Normativos/Attachments/50344/Res_4557_v8_L.pdf. Acesso em: 10 jun. 2024.

Camargos, M.A. 2012. A inadimplência em um programa de crédito de uma instituição financeira pública de minas gerais: uma análise utilizando regressão logística. REGE -

Comentado [DD12]: Precisa rever, pois estão fora de formatação.

Formatado: Fonte parágr. padrão

Formatado: Recuo: À esquerda: 0,63 cm

Formatado: Fonte: Não Negrito

Revista de Gestão. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1809227616303204><https://www.sciencedirect.com/science/article/pii/S1809227616303204>. Acesso em: 10 mar. 2024.

Governo Federal [GOV]. 2024. Cadastro Nacional da Pessoa Jurídica – CNPJ. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica-cnpj>. Acesso em: 24 mar. 2024.

Gonçalves, Reinaldo. 1991. Grupos econômicos: uma análise conceitual e teórica. Disponível em: <https://periodicos.fgv.br/rbe/article/view/534>. Acesso em: 10 mar. 2024.

Google BigQuery. 2024. Serviço de armazenamento de dados. Disponível em: <https://cloud.google.com/bigquery>. Acesso em: 10 jun. 2024.

Instituto de Pesquisa Econômica Aplicada [IPEA]. 2006. Risco de Crédito: Desenvolvimento de modelo Credit Scoring para a gestão da inadimplência de uma instituição de microcrédito. Disponível em: https://www.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006_Profissional_02lugar_tema03.pdf. Acesso em: 30 jun. 2024.

Jassé, Pedro. 2020. Gestão Do Risco De Crédito Bancário: Estudo Empírico. Disponível em: <https://comun.rcaap.pt/handle/10400.26/31674>. Acesso em: 30 jun. 2024.

Van Gestel, T; Baesens, B. 2008. Credit Risk Management: Basic Concepts financial risk componentes, rating analysis, models, economic and regulatory capital. Disponível em: https://www.academia.edu/37069057/Credit_Risk_Management_Basic_Concepts. Acesso em: 01 abr. 2024.

Wilson, R. J. 1996. Introduction to Graph Theory. 4. ed. Harlow: Pearson Education.

Código de campo alterado

Formatado: Fonte parág. padrão

Formatado: Fonte: Não Negrito

Formatado: Fonte: Não Negrito

Formatado: Fonte parág. padrão

Formatado: Fonte: Não Negrito

Formatado: Fonte parág. padrão

Formatado: Fonte: (Padrão) Arial, 11 pt, Cor da fonte: Automática

Formatado: Fonte parág. padrão

Formatado: Fonte: Não Negrito

Formatado: Português (Brasil)

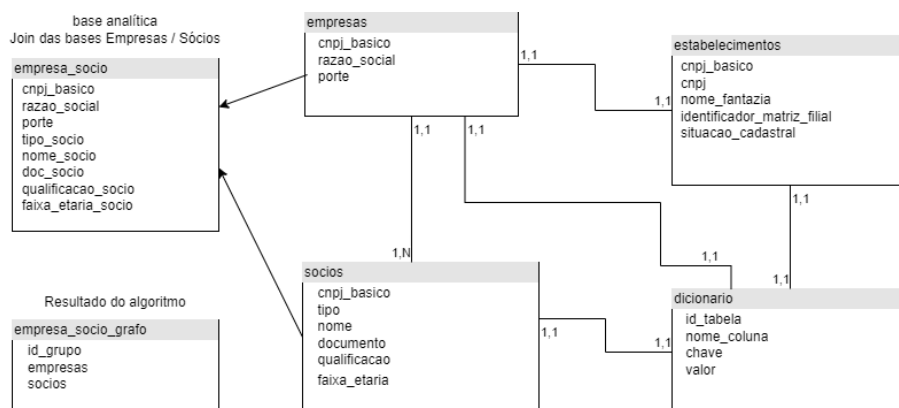
Formatado: Fonte parág. padrão

Formatado: Fonte: Não Itálico, Português (Brasil)

Formatado: Português (Brasil)

Apêndice

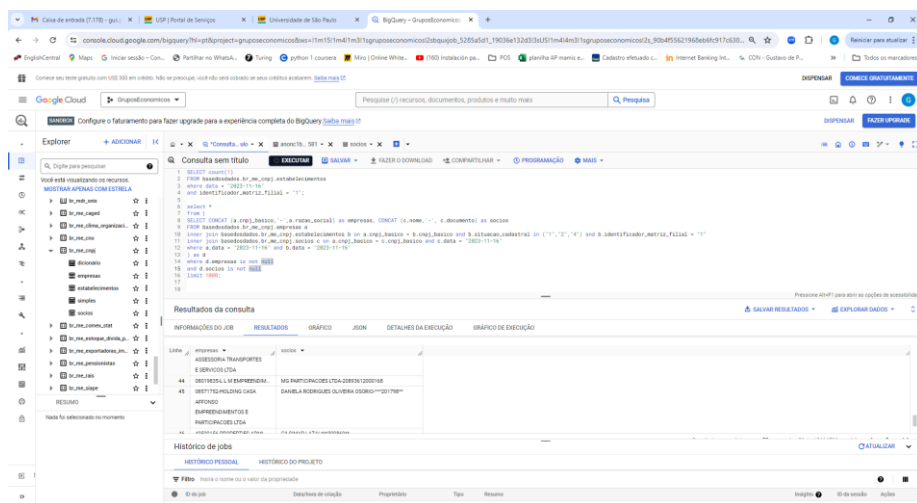
Apêndice A – Imagem 1. Modelo de dados proposto - Data Wrangling



Legenda: Modelagem espelho bases RFB e resultado projeto.

Fonte: Resultados originais da pesquisa

Apêndice A – Imagem 2 – Data Wrangling – Plataforma Google Bigquery



Legenda: BiqQuery – Comand Line.

Fonte: Resultados originais da pesquisa