**Aplicação de Grafos na Identificação e Análise de Grupos Econômicos com a identificação de vínculos entre empresas e sócios**

Gustavo de Paula Aguiar Araújo ¹\*;Diego de Oliveira Da Cunha

1 Universidade de São Paulo - Bacharelado em Sistemas da informação. Rua Arlindo Béttio, 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.

**Aplicação de Grafos na Identificação e Análise de Grupos Econômicos com a identificação de vínculos entre empresas e sócios**

**Resumo**

O trabalho aborda a aplicação da teoria dos grafos para identificar e analisar estruturas de Grupos Econômicos (GEs) com foco na gestão de risco de crédito em instituições financeiras (IFs). Um algoritmo foi desenvolvido para mapear relações entre empresas e sócios, utilizando dados públicos de participações societárias. Essa abordagem permitiu uma análise mais detalhada das interconexões econômicas, promovendo maior precisão na avaliação de riscos de crédito e decisões estratégicas.

Os resultados evidenciam a eficácia dos grafos para visualizar e automatizar a identificação de vínculos entre os integrantes de um GE. Além disso, destaca-se a relevância de incorporar técnicas modernas, como aprendizado de máquina e grafos dinâmicos, para ampliar a capacidade de adaptação às mudanças das redes econômicas e identificar padrões ocultos. Essa solução contribui para decisões mais informadas e para a mitigação de riscos financeiros, assegurando a estabilidade e sustentabilidade das IFs.

Por fim, o trabalho propõe a expansão futura do modelo, incluindo fontes externas de dados e variáveis qualitativas, além de explorar redes sociais corporativas e transações financeiras em tempo real, consolidando sua importância para o setor financeiro.

**Palavras-chave:** Grupos Econômicos, teoria dos grafos, análise de risco de crédito, aprendizado de máquina.

**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, impulsionaram o desenvolvimento de estratégias para reduzir a exposição ao risco de crédito, maximizar os lucros e manter a sustentabilidade no mercado (Jassé, 2020). Esses desafios levaram as IFs a desenvolver estratégias para mitigar riscos e garantir a sustentabilidade.

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.   
Controlar e mitigar riscos na concessão de crédito é essencial para a saúde financeira atual e futura. De acordo com Santos (2020), a gestão de risco de crédito consiste em processos e ferramentas que permitem às instituições financeiras, identificar, medir, monitorar e mitigar os riscos associados às suas operações de crédito. Essa gestão é fundamental para assegurar a sustentabilidade do sistema financeiro, incluindo o uso de métodos quantitativos e qualitativos para avaliar a capacidade dos tomadores de crédito em honrar suas obrigações, bem como para identificar fatores externos, como as condições macroeconômicas, que podem influenciar diretamente o risco de inadimplência. Além disso, políticas de gestão preventiva, como as descritas no modelo de gestão de risco Engenheiro (2015), têm se mostrado eficazes para minimizar perdas financeiras e melhorar a resiliência das instituições em cenários econômicos adversos

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 era realizada de forma subjetiva, baseada exclusivamente no julgamento de analistas. Esse método, além de não utilizar critérios objetivos, era demorado e não abrangia todas as variáveis de exposição ao risco de crédito para as IFs, tornando-se impreciso e limitado (Camargos, 2012). Segundo Santos (2020), a gestão eficaz do risco de crédito requer o uso integrado de métodos quantitativos e qualitativos, incluindo a análise de variáveis como renda, histórico de crédito e liquidez do tomador. Além disso, é fundamental conhecer melhor as contrapartes envolvidas para mitigar riscos e tomar decisões informadas. Com o avanço de novas tecnologias e modelos estatísticos, tornou-se possível desenvolver mecanismos analíticos mais precisos e eficientes para controlar carteiras de crédito.

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. Conglomerados familiares são exemplos comuns de GE no Brasil, onde empresas distintas compartilham sócios ou gestores, influenciando a análise de crédito e o risco sistêmico.

Em 2017 o Banco Central Brasileiro [BACEN] propõe a regulamentação do controle de contrapartes para fins de gerenciamento de risco. A Resolução nº 4.557 estabelece diretrizes claras para a estrutura de gerenciamento de riscos e a definição do apetite ao risco nas instituições financeiras. Esta regulamentação visa assegurar que as instituições financeiras mantenham controle detalhado sobre as contrapartes conectadas que compartilhem risco de crédito, documentando os critérios utilizados para identificação de cada indivíduo pertencente a um GE (BACEN, 2017, Art. 22). Exige também que as IFs implementem processos e sistemas que garantam a identificação e mitigação de riscos associados à concentração de crédito, promovendo uma visão integrada e abrangente dos riscos financeiros. Esses requisitos protegem as IFs contra perdas e promovem a estabilidade financeira. A norma também inclui requisitos específicos para monitoramento contínuo, análise de cenários e divulgação transparente das metodologias empregadas na avaliação de riscos.

A adoção dessas diretrizes representa um marco importante na regulação do setor financeiro brasileiro, alinhando-se às melhores práticas internacionais e às demandas do Comitê de Basileia. Tais medidas incentivam uma abordagem proativa na gestão de riscos, reforçando a confiança do mercado e protegendo a integridade das operações financeiras.

Este trabalho busca demonstrar a viabilidade de identificar de forma mais eficaz os vínculos entre empresas e sócios, utilizando a teoria dos grafos, que oferece uma abordagem para mapear conexões e interdependências entre indivíduos.

Enquanto estudos, fornecem a base teórica fundamental para o entendimento da teoria dos grafos, abordando conceitos estruturais, propriedades e algoritmos essenciais. Pesquisas mais recentes ampliaram significativamente o escopo dessa área, integrando-a a técnicas de aprendizado de máquina para atender a demandas modernas, como a análise de grafos dinâmicos. Esses avanços são particularmente importantes para modelar sistemas em constante mudança, como sistemas financeiros. Por exemplo, Chami et al. (2020) propõem organizar os métodos de aprendizado de representação para dados estruturados em grafos, unificando diferentes abordagens e destacando a aplicabilidade dessas técnicas em diversos contextos. Outras pesquisas, destacam que grafos dinâmicos permitem mapear informações complexas de redes temporais para representações vetoriais de baixa dimensão. Isso é essencial para tarefas como previsão de eventos futuros (ex.: transações financeiras entre entidades) e classificação de vínculos (ex.: identificação de novos vínculos entre empresas ou sócios) (Barros et. al., 2021).

Essa integração entre grafos e aprendizado de máquina não apenas expande as capacidades analíticas, mas também promove maior eficiência no processamento de grandes volumes de dados, alinhando-se às necessidades modernas de automação e precisão em áreas críticas como gestão de risco, análise de redes econômicas e identificação de interdependências ocultas.

Desta forma grafos foram escolhidos como base para este estudo devido à sua capacidade de modelar e analisar relações complexas entre entidades. Essa abordagem permite representar conexões, como as entre sócios e empresas, de forma a identificar padrões estruturais, como agrupamentos e redes de influência.

O uso de grafos não apenas facilita a visualização das relações, mas também possibilita o desenvolvimento de algoritmos eficientes para a análise automatizada de grandes volumes de dados e ao identificar essas relações de maneira automatizada, é possível tomar decisões mais informadas e reduzir a exposição ao risco de crédito. Em outras palavras, este trabalho demonstra como modelos baseados em grafos automatizam a análise de vínculos entre contrapartes, promovendo uma gestão de risco de crédito mais eficiente.

**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.

**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 à obtenção de uma ferramenta a qual pudesse disponibilizar as informações de forma massiva. Desta forma, foram utilizadas as informações centralizadas pelo site *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*. A escolha da ferramenta foi definida devido aos dados já estarem disponíveis na conta do site base dos dados e por possibilitar a análise de dados gerenciados em larga escala (Google BigQuery, 2023). A utilização 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 necessária a análise de toda composição societária 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, optou-se à utilização da ingestão de dados diretamente do *BigQuery* com a persistência de dados em uma base de dados local, MySQL.

O MySQL é uma ferramenta de gerenciamento de banco de dados relacional (RDBMS[[1]](#footnote-1)) de código aberto, amplamente adotada para realizar o armazenamento, a organização e a consulta de dados de maneira eficiente. Como um software open source, destaca-se por oferecer acesso gratuito, possibilidades de customização e compatibilidade com diversas aplicações. Baseando-se em um modelo relacional, permite estruturar dados de forma organizada, assegurando consistência, integridade e desempenho em operações complexas. Essas características tornou o MySQL uma opção para a estruturação das bases para construção do algoritmo do projeto.

Como mostrado na Figura 1, com a utilização dos dados mascarados para os sócios, foi necessária a criação de uma chave de referência entre as empresas e sócios, para possibilitar a criação de 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 base do CPF ou CNPJ mascarado. Essas chaves possibilitam alimentar os grafos e criar as referências e vínculos entre os integrantes do GE.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figura 1 - Chaves únicas de referência entre empresas e sócios

Fonte: Resultados originais da pesquisa

**Grafos**

A teoria dos grafos teve sua origem formalizada por Euler no século XVIII com o problema das Pontes de Königsberg[[2]](#footnote-2), sendo posteriormente expandida para diversas aplicações em redes, logística e análise de dados. Segundo Feofiloff et al. (2009), grafos desempenham papéis fundamentais ao representar relações complexas em sistemas sociais e econômicos, fornecendo uma base matemática robusta para o estudo de interações entre elementos.

Estudos mais recentes, como o de Karrer e Newman (2010), ampliam as aplicações da teoria dos grafos, explorando como grafos dinâmicos podem capturar mudanças temporais em redes diversas, incluindo sociais, financeiras e biológicas. Esses trabalhos demonstram que redes dinâmicas são ferramentas eficazes para modelar fluxos de transações financeiras, interações entre investidores e flutuações de mercado, permitindo análises mais detalhadas de sistemas em constante transformação.

A teoria dos grafos, definida como o estudo de estruturas matemáticas que representam relações entre entidades (Feofiloff et al., 2011), é utilizada neste trabalho para mapear as conexões entre sócios e empresas em GEs. Essa abordagem permite identificar padrões estruturais e vínculos críticos que influenciam a avaliação de risco de crédito.

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, 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, podem ser consideradas um exemplo da utilização desse tipo de abordagem.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente com confiança média

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.

Gráfico

Descrição gerada automaticamente

Figura 3. Estrutura de Grado não direcionado

Fonte: Resultados originais da pesquisa

O uso de grafos direcionados e não direcionados permite modelar diversas aplicações em áreas como ciências sociais, informática e economia, ampliando a compreensão sobre relações complexas entre elementos de um sistema. Essa afirmação corrobora com a ideia do trabalho, ao qual se utiliza da teoria para conseguir identificar correção entre empresas e sócios, de modo a identificar possíveis estruturas de GEs.

**Método**

Após a modelagem dos dados, foram desenvolvidos 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 Comércio’ e o ‘Restaurante Catinho’ através do sócio ‘AJM’. A execução parte de três *Dataframes*[[3]](#footnote-3), descritos como empresas, sócios e relação empresa e sócio. Em seguida, é executada 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.

|  |  |
| --- | --- |
| Texto  Descrição gerada automaticamente |  |

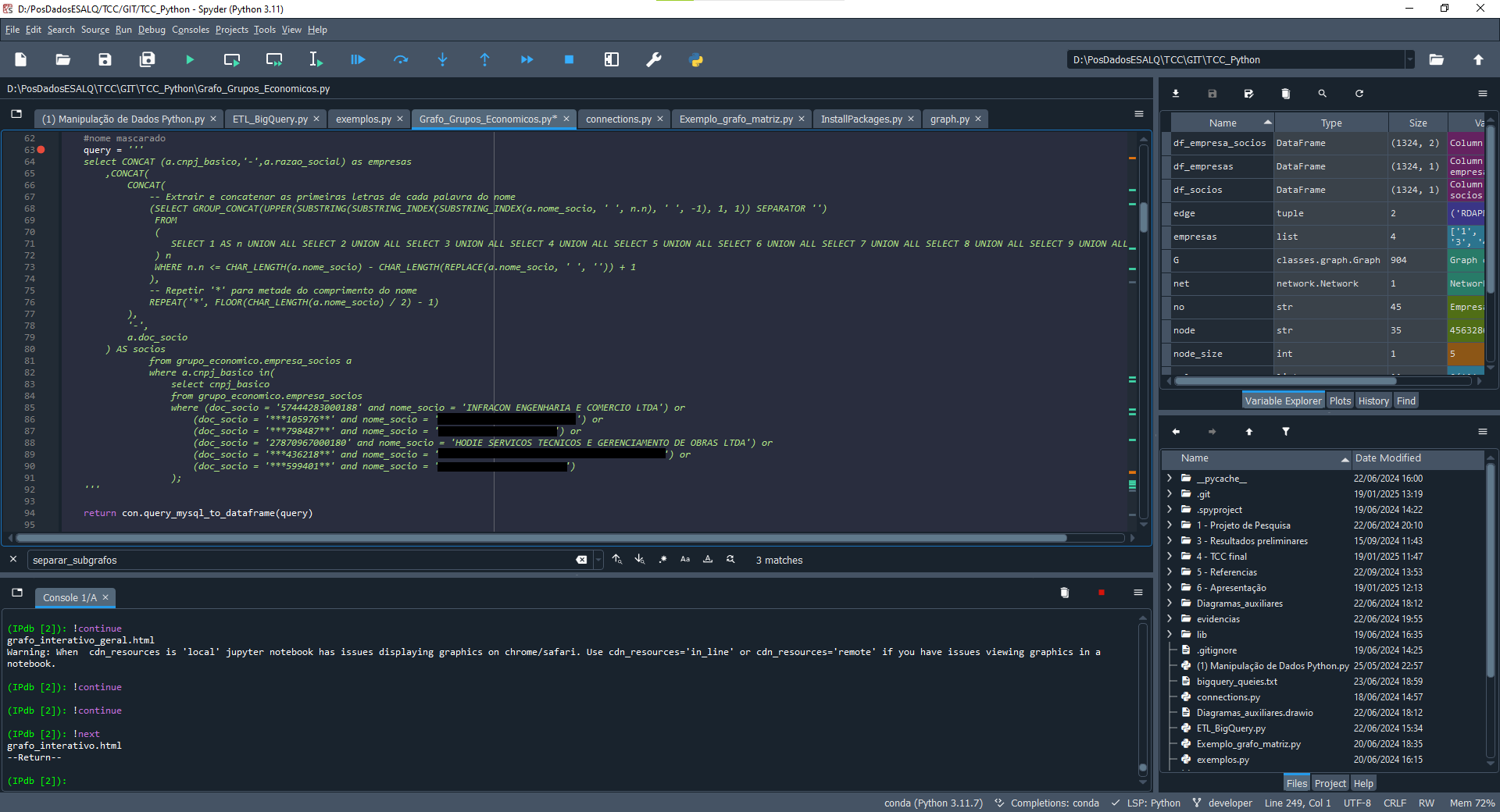
Figura 4. Grafo de referência

Fonte: Resultados originais da pesquisa

**Funcionamento do algoritmo**

O sistema é baseado em algumas etapas. Inicialmente existe a preparação das bases de entradas, tratamento dos dados, preparação do grafo, resultado gráfico já apresentado anteriormente no trabalho e por fim a criação de uma matriz referencial para facilitar a análise via banco de dados.

Para o case, elegemos um possível GE, com o intuito de realizar a análise pontual dos vínculos entre as empresas e seus sócios. No caso, já foi incluído o tratamento e criptografia dos nomes dos envolvidos e em seguida é gerado um dataframe utilizando a lib sqlalchemy e pandas, visualizado na figura 8.

  
Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Figura 8. Case - Consume de dados

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/Grafo_Grupos_Economicos.py>

Percebe-se a geração de um dataframe descrito como “df\_empresas\_socios”, que representa os vínculos de participação societária entre sócios e empresas. Ainda nesse ponto não é possível identificar a ligação entre os sócios e as diversas empresas, pois conta somente o vínculo único daquele indivíduo sócio com uma ou N empresas.

Em seguida, o algoritmo separa os dados de empresas “df\_empresas” e sócios “df\_socios” em dataframes distintos, os quais servirão de entrada como Vértices e Arestas no grafo proposto, demonstrado na figura 9.

Interface gráfica do usuário

Descrição gerada automaticamente

Figura 9. Case – Tratamento dos dados do grafo

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/Grafo_Grupos_Economicos.py>

A partir desse ponto o sistema consegue prover a formação do grafo, o qual criará massivamente os subgrafos propostos, partindo-se das conexões entre as empresas e os sócios recuperados previamente e incluídos nos distintos dataframes*.*

A criação do grafo é bastante custosa, ao se utilizar a base total de empresas e sócios da RFB e por isso optou-se a inclusão de chunks[[4]](#footnote-4), processamento 1000 nós por vez, paralelamente. Nesse ponto do projeto, foi separado a vinculação dos nós de empresas e sócios do grafo. Isso permitiu a criação do grafo com 55 milhões de interligações, utilizando pouca infraestrutura. Figura 10.

A forma de criação do grafo, está dividida em duas etapas, inicialmente são incluídos os nós na memória e em seguida é realizado a vinculação entre os nós, conforma ilustrado na figura 10, através do método add\_edge, o qual utiliza um laço para a criação dos vínculos entre os nós utilizando-se da biblioteca networkx para a criação do grafo.

Tela de computador

Descrição gerada automaticamente

Figura 10. Case – Criando o grafo de referencias entre empresas e sócios

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/Grafo_Grupos_Economicos.py>

Desse ponto em diante, já é possível identificar possíveis formações de GEs, visto que os subgrafos já foram criados e estão disponíveis em memória, porém visualmente não seria possível, pois todos os subgrafos e formações padrões criadas estariam em memória e não há um padrão visual.

Na próxima etapa o sistema separa os subgrafos, possibilitando distinguir os públicos e assim permitindo a visualização de padrões de agrupamento de indivíduos distintos. Neste ponto, para efeito de demonstração, utilizou-se a biblioteca pyvis.network a qual cria interativamente as populações de subgrafos, distinguindo o tamanho dos nós para empresas e sócios, como visto na figura 11, linha 215.

**Tela de computador com jogo

Descrição gerada automaticamente**

Figura 11. Case – Visualizando o grafo.

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/Grafo_Grupos_Economicos.py>

A visualização do grafo gerado, geralmente é confusa para grupos de indivíduos que se conectam muito, como demonstrado na figura 12. Para esse exemplo, a unificação dos nós é muito extensa e precisaria de maior aprofundamento em cada cliente para uma tomada de decisão de crédito, porém é sabido que todos os integrantes do grafo, de alguma forma direta ou indiretamente possui uma ligação por participação societária entre eles e com isso possibilitaria ampliar o controle e análise de risco de crédito para os envolvidos.

**Diagrama

Descrição gerada automaticamente**

Figura 12. Case – Grafo – Grupo sugerido.

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/grafo_interativo.html>

Para validar as interrelações entre empresas e sócios, separou-se uma pequena parte do subgrafos para demonstrar os vínculos entre empresas e sócios distintos. Conforme visualizamos na figura 13, que é um recorte do subgrafos gerado anteriormente (figura 12), pode-se perceber que a empresa ‘CONSORCIO TRIX INFRACON SOROCABA’, possui um de seus sócios ‘TECLERJ’, que é sócio direto da empresa ‘CONSORCIO TRIX INFRACON CONCORDIA’, que por sua vez possui sociedade com o sócio ‘MZT’. A interligação entre o sócio ‘FDDO’ e o ‘MZT’, nunca seria disponibilizada se não houvesse uma análise aprofundada das relações entre os sócios e empresas. Já com a criação de um grafo, pode-se obter essa ligação de forma sistêmica, possibilitando aprofundar a análise dos envolvidos em um contrato de crédito, uma vez que pode haver alguma ligação positiva ou negativa entre os sócios descritos, na hora de conceder crédito, por exemplo.

A possibilidade de identificar correlação entre os sócios, pode trazer insumos para qualificação das empresas e consequentemente diminuir ou aumentar o apetite de crédito e consequentemente melhorar a qualidade e o risco envolvido.

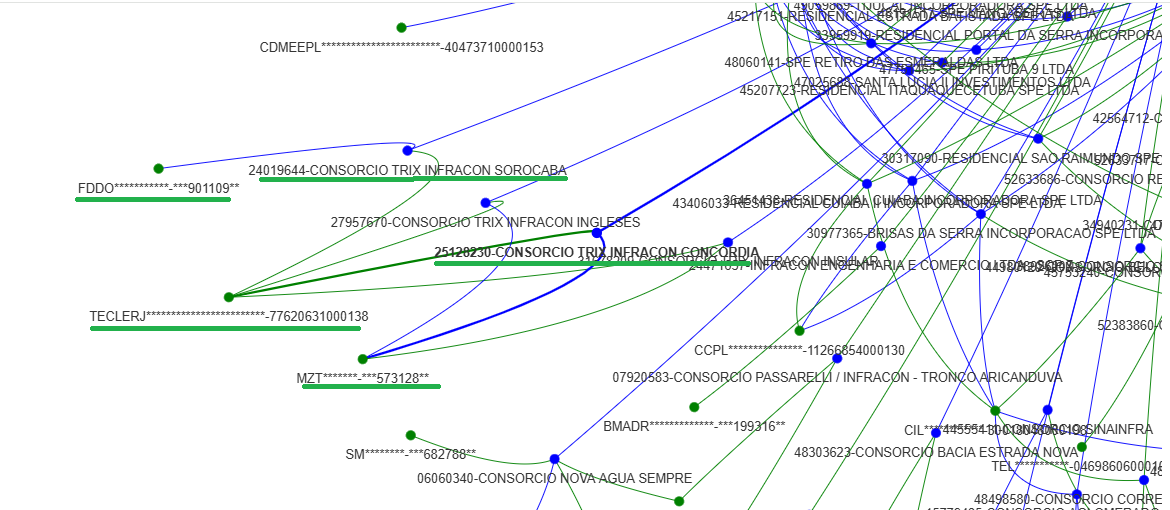


Figura 13. Case – Grafo – Zoom grupo sugerido.

Fonte: Resultados originais da pesquisa

Por fim, o algoritmo propõe incluir a visão dos subgrafos gerados em formados de matrizes, para que seja possível armazenar a informação obtida em tabelas relacionais. Essa opção foi realizada para utilização da mesma infraestrutura proposta desde então, porém é possível aprofundar mais o algoritmo incluindo os subgrafos em bancos de dados de grafos, tais como Neptune da Amazon, Microsoft Azure Cosmos DB, ArangoDB etc. Para fim de análise, o algoritmo disponibiliza todas as relações obtidas armazenadas na matriz, para uma possível análise de interpelação societária, assim como demonstrado no console na figura 14.

**Interface gráfica do usuário, Texto

Descrição gerada automaticamente**

Figura 14. Case – Grafo – Zoom grupo sugerido.

Fonte: <https://github.com/gupauag/TCC_Python/blob/developer/Grafo_Grupos_Economicos.py>

**Resultados Preliminares**

Uma das formas de se descobrir e analisar a formação de GE é a investigação da composição acionária e os sócios de uma empresa, identificando vínculos entre empresas por meio de acionistas comuns. Com a utilização do algoritmo proposto, foi possível montar estruturas complexas instantaneamente, demonstrando a relevância da ideia.

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 inferir uma estrutura de GE, conforme visto na figura 15.

Diagrama

Descrição gerada automaticamente

Figura 15. Grupo Econômico – Vínculo Societário

Fonte: Resultados originais da pesquisa

Na figura 15, pode-se visualizar as referências entre as empresas e sócios. É possível identificar um vínculo entre as empresas ‘FR Comércio’ 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 trazem 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’, vista na figura 15, 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 a empresa ‘FR Comércio’ 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 considerada como a única forma da análise do risco de crédito envolvido, mas já garante maior visibilidade na avaliação do perfil do GE.

Desenho de uma flor

Descrição gerada automaticamente com confiança média

Figura 16. 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 podem não representar grande impacto ao risco de crédito 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 16 e 17, 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.

Interface gráfica do usuário

Descrição gerada automaticamente com confiança baixa

Figura 17. Grupo Econômico – Vínculo Societário – Zoom vínculo com muitas arestas e distante

Fonte: Resultados originais da pesquisa

Existe também uma abordagem para empresas compostas por muitas arestas, as quais podem ter uma grande capacidade de criação de novos vínculos, como visto na figura 16 e 17. Nesses casos, podemos considerar que a identificação de forma rápida e proativa dos novos vínculos dá maior dinamismo às 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. Ao introduzir o conceito de GE na análise de risco de crédito, observamos uma otimização dos métodos de avaliação, aumentando a segurança para cumprir as diretrizes definidas pelo BACEN.

O algoritmo desenvolvido simplifica a identificação das estruturas dos Grupos Econômicos (GEs), evidenciando os laços entre empresas e sócios a partir das participações societárias. Isso possibilitou a criação de visões claras das conexões entre os indivíduos, aumentando a capilaridade na avaliação de risco de crédito. Essa abordagem é corroborada por Paula et al. (2019), que discutem a aplicação de métodos quantitativos na gestão de riscos em projetos. Segundo os autores, a utilização de técnicas quantitativas permite identificar variáveis como renda, taxas de juros aplicadas, prazos, histórico de inadimplência e liquidez do tomador, que podem indicar a probabilidade de as contrapartes não cumprirem os acordos firmados. Além disso, a integração de métodos qualitativos fornece uma compreensão mais abrangente dos riscos, considerando aspectos que não são facilmente quantificáveis. A combinação dos métodos com a descoberta dos vínculos entre os integrantes do GE, traz maior clareza no controle de risco já que 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 à escravidão), outros contratos em dia ou em atraso etc.

Sendo assim, as IFs precisam controlar de forma mais eficiência a gestão do crédito e riscos. Esse controle pode ser realizado 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 GEs. Esses grupos, por possuírem integrantes que estão 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, 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.

O uso de grafos dirigidos e não dirigidos no algoritmo proporcionou uma representação mais clara dessas relações, destacando a importância dos vínculos diretos e indiretos e da influência mútua no processo de concessão de crédito. Por outro lado, se a identificação de GEs não for precisa, há o risco de sobrecarregar a análise de crédito com informações irrelevantes, gerando confusão sobre quais conexões afetam realmente 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.

A integração de aprendizado de máquina com grafos, como explorado por Wang et al. (2022), destaca a evolução das técnicas computacionais na análise de dados estruturados e não estruturados. Essa abordagem permite não apenas capturar relações complexas entre elementos, mas também adaptá-las a contextos dinâmicos e em tempo real. Wang et al. (2022) enfatizam o uso de ferramentas automatizadas, como redes neurais gráficas (GNNs), para explorar e otimizar modelos baseados em grafos. Essas técnicas são particularmente úteis para incorporar grafos dinâmicos, onde os relacionamentos entre os nós mudam com o tempo, permitindo análises que refletem a evolução das interações no sistema, assim como é identificado nas estruturas societárias as quais não são estáticas e imutáveis.

A análise em tempo real, possibilitada por essas integrações, amplia o escopo das aplicações, permitindo monitorar redes financeiras, detectar padrões anômalos, como fraudes, e prever eventos futuros com maior precisão. Por exemplo, em sistemas financeiros, a capacidade de capturar mudanças nas relações entre empresas e sócios em tempo real é essencial para a mitigação de riscos. Além disso, a combinação de aprendizado de máquina e grafos dinâmicos possibilita modelar interações temporais complexas, como fluxos de capital ou transações comerciais, que são frequentemente difíceis de analisar usando métodos convencionais.

Desta forma, a identificação de GE por grafos não pode ser a única abordagem, pois o algoritmo foca somente na identificação de relações entre sócios e empresas e pode fornecer uma visão incompleta dos 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, visando gerar novas premissas mais precisas 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 positiva 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 das direções que poderiam tornar essa ferramenta mais robusta.

Portanto, ao combinar aprendizado de máquina, novas variáveis e grafos, o modelo proposto neste trabalho pode ser expandido para incorporar análise preditiva em redes dinâmicas, aumentar a precisão de detecção de padrões ocultos e melhorar a capacidade de adaptação às mudanças nas interações entre elementos do sistema financeiro. Esses avanços consolidam a importância de novas abordagens na gestão e risco e na tomada de decisões estratégicas.

Em resumo, a análise automatizada de GEs, 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 para identificar de forma proativa as conexões entre empresas e sócios, possibilitando às 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, cujo apoio constante e incentivo foram fundamentais para meu crescimento profissional e pessoal.

**Referências**

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_v1_O.pdf>>. Acesso em: 10 jun. 2024.

Barros, D. T. C.; Mendonça, M. R. F.; Vieira, A. B.; Ziviani, A. 2021. A survey on embedding dynamic graphs. Journal of Data Science and Engineering. Disponível em: <<https://arxiv.org/abs/2101.01229>>. Acesso em: 18 jan. 2025.

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.

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 - Revista de Gestão. Disponível em: <[https://www.sciencedirect.com/science/article/pii/S180 9227616303204](https://www.sciencedirect.com/science/article/pii/S180%209227616303204)>. Acesso em: 10 mar. 2024.

Chami, I.; Abu-El-Haija, S.; Perozzi, B.; Re, C.; Murphy, K. 2020. Machine learning on graphs: A model and comprehensive taxonomy. Disponível em: <<https://arxiv.org/abs/2005.03675>>. Acesso em: 18 jan. 2025.

Engenheiro, A. S. L. 2015. O Crédito Bancário: A Prevenção do Risco e Gestão de Situações de Incumprimento. Disponível em: <<https://run.unl.pt/bitstream/10362/16176/1/Engenheiro_2015.pdf>>. Acesso em: 19 jan. 2025.

Feofiloff, P.; Kohayakawa, Y.; Wakabayashi, Y. 2011. Uma Introdução Sucinta à Teoria dos Grafos. Disponível em: <<https://www.ime.usp.br/~yw/publications/books/TeoriaDosGrafos.pdf>>. Acesso em: 19 jan. 2025.

Gestel, T. V; 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.

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

Governo do Brasil. Cadastro Nacional da Pessoa Jurídica – CNPJ. 2024. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>>. Acesso em: 24 mar. 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, P. 2020. Gestão Do Risco De Crédito Bancário: Estudo Empírico. Disponível em: <<https://comum.rcaap.pt/handle/10400.26/31674>>. Acesso em: 30 jun. 2024.

Karrer, B.; Newman, M. E. J. 2010. Random Graphs Containing Arbitrary Distributions of Subgraphs. Oxford University Press, v. 14, n. 8, p. 1-15. Disponível em: <<https://arxiv.org/abs/1005.1659>>. Acesso em: 19 jan. 2025.

Paula, C. P.; Cordeiro, G. A.; Rampasso, I. S.; Ordoñez, R. E. C.; Anholon, R. 2017. Métodos Quantitativos para Gestão de Risco em Projetos: Uma Revisão da Literatura. Revista Gestão da Produção Operações e Sistemas. Revista GEPROS. Disponível em: <<https://revista.feb.unesp.br/gepros/article/view/2210>>. Acesso em: 19 jan. 2025.

Santos, V. A. C**.** 2020. Gestão de Risco de Crédito Bancário: Caso do Banco Comercial e de Investimentos (BCI, S.A.). Disponível em: <<https://repositorio.iscte-iul.pt/bitstream/10071/22452/1/master_vanda_carvalho_santos.pdf>>. Acesso em: 19 jan. 2025.

Wang, X.; Zhang, Z.; Li, H.; Zhu, W. 2022. Automated Graph Machine Learning: Approaches, Libraries, and Benchmarks. Journal of LaTeX Class Files, v. 14, n. 8, p. 1-15. Disponível em: <<https://arxiv.org/abs/2201.01288>>. Acesso em: 19 jan. 2025.

**Apêndice**

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

Diagrama

Descrição gerada automaticamente

Legenda: Modelagem espelho bases RFB e resultado projeto.

Fonte: Resultados originais da pesquisa

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

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Legenda: BiqQuery – Comand Line.

Fonte: Resultados originais da pesquisa

1. RDBMS significa Relational Database Management System (Sistema de Gerenciamento de Banco de Dados Relacional) [↑](#footnote-ref-1)
2. Problema matemático que queria responder a seguinte questão: É possível atravessar todas as sete pontes da cidade passando por cada uma delas exatamente uma vez? [↑](#footnote-ref-2)
3. 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. [↑](#footnote-ref-3)
4. Chunks. No python referem-se a partições ou divisões de dados em partes menores para processamento. O termo é amplamente utilizado em contextos como manipulação de arquivos, listas, strings, fluxos de dados, entre outros. Dividir dados em chunks é útil para lidar com grandes volumes de informações de forma eficiente, especialmente quando os dados não cabem inteiramente na memória. [↑](#footnote-ref-4)