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

Gustavo de Paula Aguiar Araújo;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

2 Nome da Empresa ou Instituição (opcional). Titulação ou função ou departamento. Endereço completo (pessoal ou profissional) – Bairro; 00000-000 Cidade, Estado, País

\*autor correspondente: nome@email.com

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

**Resumo**

Tópico obrigatório para o depósito do TCC, porém opcional para a etapa dos Resultados preliminares.

**Palavras-chave:** (inserir até cinco palavras diferentes das contidas no título, separadas por ponto-e-vírgula).

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Introdução**

Há muito tempo, instituições financeiras [IF]s e não financeiras têm se preocupado com a gestão eficaz do risco de crédito. 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. (Van Gestel e Baesens, 2008)

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 conceção de crédito é fundamental e essencial para garantir a saúde do ambiente financeiro atual e futuro. (Van Gestel e Baesens, 2008)

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). Desde então, através do advento de novas tecnologias e modelos estatísticos, as IFs vêm desenvolvendo maneiras mais eficientes para controlar sua carteira de crédito, criando mecanismos para conhecer melhor seus clientes.

É nesse contexto em que é introduzido o conceito de grupo econômico [GE]. GE é o termo dado 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.

Neste contexto, este trabalho apresenta um algoritmo focado 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. Em outras palavras, o artigo descreve como é possível criar um modelo que identifique de forma automática relação entre os sócios e empresas, apresentando uma visão na avaliação e identificação das contrapartes envolvidas na conceção 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.

**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, 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 BIGQUERY,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**

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, assim como descrito no código fonte disponível no GitHub (ARAÚJO, 2024).

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

Figura 1. Chaves únicas de referência entre empresas e sócios

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Autor, 2024

**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 (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  
Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente com confiança média

Fonte: Autor

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

Figura 3. Estrutura de Grado não direcionadoGráfico

Descrição gerada automaticamente

Fonte: Autor

**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 Dataframes[[1]](#footnote-1), descritos como empresas, sócios e relação empresa e sócio. Em seguida é executado massivamente a validação de relacionamentos entre as empresas e sócios, associando as arestas dos subgrafos. A figura X, ilustra como parte dos subgrafos, denominados GE, foram gerados após a execução do algoritmo.

Figura X. Grafo de referência

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

Fonte: Autor

**Resultados Preliminares**

A seguir são apresentados os resultados do trabalho já com as imagens exportadas pelo algoritmo e como é possível identificar estruturas complexas ou não de vínculo entre as 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.

Figura 4. Grupo Econômico – Vínculo Societário

Diagrama

Descrição gerada automaticamente

Imagem gerada com o auxílio da biblioteca pyvis.network. Fonte: Autor

Na figura 4, pode-se visualizar as referências entre as empresas e sócios. Também 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’ (Figura 4), 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 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.

Figura 5. Grupo Econômico – Vínculo Societário

Desenho de uma flor

Descrição gerada automaticamente com confiança média

Imagem gerada com o auxílio da biblioteca pyvis.network. Fonte: Autor

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 se considerar somente o vínculo entre as empresas e sócios em um GE, podemos ter grupos com muitos vínculos, 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.

Figura 6. Grupo Econômico – Vínculo Societário – Zoom vínculo com 1 aresta e distante

Interface gráfica do usuário

Descrição gerada automaticamente com confiança baixa

Imagem gerada com o auxílio da biblioteca pyvis.network. Fonte: Autor

Outra característica que se pode visualizar com a geração das visões de GE criadas por vínculos entre empresas e sócios é de que empresas que são compostas por consórcios, tais como empresas de engenharia civil, por exemplo, possuem uma grande capacidade de criação de novos CNPJs (Figura 5) o que demonstra a necessidade de identificação de forma rápida dos vínculos entre os indivíduos de um GE.

**Conclusão(ões) ou Considerações Finais**

Tópico obrigatório para o depósito do TCC, porém opcional para a etapa dos Resultados preliminares. A seção deve conter frases curtas, apresentando as conclusões e inferências elaboradas a partir da discussão dos resultados obtidos até o momento. Os autores não devem, em hipótese alguma, mencionar, citar ou reproduzir resultados de outros estudos na(s) Conclusão(ões) ou Considerações Finais do TCC. Além disso, esta seção não deve conter tabelas ou figuras, sendo redigida de forma sucinta.

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Agradecimentos** (opcional, 1 parágrafo sucinto)

O título da seção Agradecimentos deve ser alinhado à esquerda e grafado em negrito, com a primeira letra da palavra grafada em letra maiúscula. Trata-se de uma seção opcional, de no máximo três linhas, na qual o autor agradece aqueles que contribuíram de maneira relevante para o desenvolvimento do trabalho e elaboração do TCC, mas que não tiveram o envolvimento intelectual necessário à atribuição de coautoria do mesmo, abstendo-se totalmente da menção ou citação de nomes de empresas, instituições ou pessoas que permitiram ou contribuíram com o desenvolvimento do trabalho, a menos que esteja documentalmente autorizado a fazê-lo.

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Referências**

**Araújo, G.** Código-fonte do projeto de TCC. [online]. 2024. Disponível em: https://github.com/gupauag/TCC\_Python. Acesso em: junho de 2024.

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

**Banco Central do Brasil.** Resolução nº 4557, de 23 de fevereiro de 2017. Dispõe sobre a estrutura de gerenciamento de riscos e apetite por riscos nas instituições financeiras. Diário Oficial da União, Brasília, DF, 23 fev. 2017. Disponível em: https://normativos.bcb.gov.br/Lists/Normativos/Attachments/50344/Res\_4557\_v7\_L.pdf. Acesso em: junho de 2024.

**Camargos**, M.A.; Araújo, E.A.T.; Camargos, M.C.S. 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, 19(3): 473-492. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1809227616303204>. Acessado em: março de 2024

**Dados.gov**. Cadastro Nacional da Pessoa Jurídica – CNPJ. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>. Acessado em: 24 de março de 2024.

**Gonçalves**, Reinaldo. 1991. Grupos econômicos: uma análise conceitual e teórica. Disponível em: <https://periodicos.fgv.br/rbe/article/view/534)>. Acessado em: março de 2024.

**Governo do Brasil.** Cadastro Nacional da Pessoa Jurídica - CNPJ. Portal gov.br. Disponível em: https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj. Acesso em: 01 jun. 2024.

**Google BigQuery.** Serviço de armazenamento de dados. Disponível em: <https://cloud.google.com/bigquery?utm_source=google&utm_medium=cpc&utm_campaign=latam-BR-all-es-dr-SKWS-all-all-trial-e-dr-1707800-LUAC0020236&utm_content=text-ad-none-any-DEV_c-CRE_688140022991-ADGP_Hybrid+%7C+SKWS+-+EXA+%7C+Txt_Usecases-Big+Query-KWID_43700079279114517-kwd-12297987241&utm_term=KW_big+query-ST_big+query&gad_source=1&gclid=CjwKCAjwydSzBhBOEiwAj0XN4K9w8u9qXUaA3D4hFCZ8jGFAxZ2XMXRlheefDbxvLFeR2VrcfsiE6xoCQNUQAvD_BwE&gclsrc=aw.ds&hl=pt>. Acessado em: junho de 2024.

**IME**. Grafos. Disponível em: <https://www.ime.usp.br/~pf/algoritmos_para_grafos/aulas/graphs.html>. Acessado: março de 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>

Acessado em: 01 de abril de 2024.

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

**Apêndice**

Apêndice A – Imagem 1.

Diagrama

Descrição gerada automaticamente

Legenda: Modelagem espelho bases RFB e resultado projeto. Fonte: Autor, 2024

Apêndice A – Imagem 2.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Legenda: BiqQuery – Comand Line. Fonte: Autor, 2024

**ou Anexo** (opcional)

Apêndices são textos e/ou documentos que foram elaborados pelo autor e que são importantes para complementar a argumentação do trabalho. Anexos são textos ou documentos que ilustram, mas que não foram elaborados pelos autores. Apêndices deverão seguir as mesmas normas de formatação do restante do texto, inclusive para figuras e tabelas.

O TCC deverá conter no máximo 30 páginas, incluindo o(s) Apêndice(s) e/ou Anexo(s).

**Atenção:** antes de enviar o arquivo para o Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

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. [↑](#footnote-ref-1)