

DESCOBERTA DE RELAÇÕES ENTRE ESTADOS BRASILEIROS A PARTIR DE DADOS FINANCEIROS DE OPERAÇÕES DE CRÉDITO DISPONÍVEIS EM DADOS ABERTOS DO BANCO CENTRAL

Discovery of relations between brazilian states from financial data of credit operations available in open data of the Central Bank of Brasil

Juliana Rodrigues de Lima Meirelles¹, Henrique Monteiro Cristovão², Daniela Lucas da Silva Lemos³

(1) Universidade Federal do Espírito Santo, Vitória-ES, juliana.meirelles@edu.ufes.br

(2) Universidade Federal do Espírito Santo, Vitória-ES, henrique.cristovao@ufes.br

(3) Universidade Federal do Espírito Santo, Vitória-ES, daniela.l.silva@ufes.br

Resumo:

No sistema bancário brasileiro mudanças têm impactado a forma como a sociedade lida com suas finanças, e cada vez mais o uso dos dados é enfatizado como sendo o fator de competitividade de um produto para o outro. O objetivo geral é investigar e evidenciar relacionamentos existentes entre os estados brasileiros sobre aspectos do consumo de produtos de crédito bancário. Para tanto, foram utilizados dados financeiros e de operações de crédito a partir de dados abertos do Banco Central. Consiste em pesquisa aplicada, de caráter exploratório e descritivo, que visa não só relacionar as variáveis de análise central, mas também apresentar informações que possam servir de diretrizes para ações de transformação da realidade. Foi feita uma modelagem conceitual sobre um grupo de variáveis que representassem o problema de pesquisa. Utilizou-se técnicas de análise de redes de informação para revelar relacionamentos não aparentes entre as variáveis selecionadas. Condensando os resultados, pôde-se observar os estados de SP, RS, SC, PR e MG como sendo relevantes perante à sociedade, nos aspectos relativos ao consumo de produtos bancários, bem como o incentivo econômico gerado pelos mesmos, alavancando também o consumo de outros estados dependentes. O desenvolvimento deste estudo, bem como o exame da literatura mostrou-nos que o cenário atual encontra-se promissor para aplicação de técnicas de análise de redes de informação, pois possibilita gerar insights com execuções interessantes principalmente para a área de gestão da informação e do conhecimento.

Palavras-chave: análise de redes de informação; dados abertos; rede monopartida; modelagem de dados; mercado financeiro.

Abstract:

In the Brazilian banking/financial system, changes have impacted the way society deals with its finances and the use of data is increasingly emphasized as a factor of competitiveness from one product to another (VALENTIM, 2006). The present study sets out a research problem: What are the most relevant relationships between Brazilian states on aspects of the consumption of bank credit products? Thus, the general objective becomes to investigate and bring relevant relationships between Brazilian states on aspects of the consumption of bank credit products. For this purpose, financial data and credit operations based on open data from the Central Bank will be used. It consists of applied research, of an exploratory and descriptive nature, which aims not only to relate the variables of central analysis, but also to present information that can serve as guidelines for actions to transform reality. Condensing the results, one can observe the states of SP, RS, SC, PR and MG as being relevant to society, in the aspects related to the consumption of banking products, as well as the economic incentive generated by them, also leveraging the consumption of other dependent states. The development of this study, as well as the examination of the literature, showed us that the current scenario is promising for network science. With the study, we were able to observe that the practical applicability of this technique in the financial area is satisfactory and promising, as it makes it possible to generate insights with interesting executions, mainly for the area of information and knowledge management.

Keywords: analysis of information networks; open data; one-party network; data modeling; financial market.

1 Introdução

O mundo passa por grandes transformações no que tange à gestão e uso da informação. No sistema bancário/financeiro brasileiro não é diferente, pois mudanças têm impactado a forma como a sociedade lida com suas finanças e cada vez mais o uso dos dados é enfatizado como sendo o fator de competitividade de um produto para o outro (VALENTIM, 2006). Gonzalez (2016) argumenta que a demanda por serviços financeiros mais eficientes e apropriados ainda é grande. Isso explica a crescente criação de *startups*, considerado por Gitahy (2011) como um grupo de pessoas à procura de um modelo de negócios repetível e escalável. Inclui-se aí as *FinTechs* que são orientadas pela tecnologia, não estando vinculados a sistemas legados (JUENGERKES, 2016). Outro fator importante e atual do setor financeiro é o surgimento do *Open Banking*, ou sistema financeiro aberto. Ele promete revolucionar a forma como se lida com produtos, serviços e dados financeiros dos cidadãos. Em uma primeira fase do Open Banking, houve o compartilhamento dos dados das instituições financeiras de forma padrão, pelo Banco Central. Essa nova realidade traz consigo a inovação, o surgimento de novos modelos de negócios, difusão de informação, favorecendo a inclusão e a educação financeira da população, possibilitando, por exemplo, o surgimento de aplicativos de comparação de preços de produtos financeiros (OPEN BANKING BRASIL, 2022).

Segundo Valentim (2008), deve-se refletir sobre os dispositivos que transformam dados em informação. Saber lidar com o alto volume, diferentes origens e formatos de dados é uma das habilidades mais valorizadas das últimas décadas (DAVENPORT et al., 2012 apud ISOTANI, 2015). A utilização da modelagem conceitual de dados é um conhecimento já consolidado no mercado de trabalho, principalmente para o desenvolvimento de softwares, e na Ciência da Informação encontra-se presente na organização e representação da informação como forma de fornecer uma descrição estável e coerente dos dados (SAYÃO, 2001). Segundo Guizzardi (2005), a modelagem está associada à representação

das coisas do mundo, independente de escolhas tecnológicas, e ela ainda objetiva entender, comunicar e raciocinar sobre uma porção da realidade. No mesmo intuito, a visualização da informação visa aumentar a compreensão do usuário sobre algum conhecimento, trazendo assim a inclusão informacional dos usuários (DIAS, 2007).

Nossa sociedade está cada vez mais dependente de redes e se pudéssemos resumir em uma palavra a atual sociedade esta palavra seria “conectada” (CAVALCANTE, 2009) e ainda, de acordo com Barabási (2013), várias redes surgem e evoluem baseados em leis e mecanismos que são a base para a nova ciência chamada de Ciência das Redes.

A análise de redes sociais preocupa-se com o entendimento das ligações entre entidades sociais e as implicações destas ligações (WASSERMAN; FAUST, 1994). Para a análise de redes de informação, também conhecida por análise de redes complexas, utilizam-se praticamente as mesmas técnicas da análise de redes sociais. Considerada como um dos métodos de apoio à descoberta de conhecimento na Ciência de Dados, a análise de redes de informação é uma prática que tem ajudado muitos pesquisadores a descobrirem relacionamentos não aparentes entre variáveis de uma base de dados.

Uma rede denominada de monopartida é composta por nós advindos de apenas uma variável e é útil para revelar relacionamentos existentes entre os grupos de nós sobre essa variável. Normalmente ela é obtida de projeções bipartidas, que é um processo de eliminação de variáveis de uma rede.

O Banco Central do Brasil divulga mensalmente informações agregadas das operações de crédito realizadas no âmbito nacional pelas instituições autorizadas, recebidas através do Sistema de Informações de Créditos – SCR. São disponibilizados dados agregados de Carteira Ativa, Inadimplência e Ativo Problemático com possibilidade de detalhamento por tipo de cliente (PF/PJ), modalidade de crédito, unidade da federação, natureza da ocupação (PF), porte/rendimento dos clientes, origem de recursos e indexador das operações (Painel

de Operações de Crédito, 2022). Esses dados são disponibilizados no Portal de Dados Abertos do Banco Central onde é possível fazer os downloads das séries em diversos formatos.

Questões relativas à gestão da informação, como acessá-las e difundi-las, tornaram-se áreas discutidas por estudiosos da Ciência da Informação, pois é sabido que os sistemas de informação, por si só, não são suficientes para responder às demandas informacionais das pessoas e organizações. Tendo em vista a relevância dos impactos das mudanças no sistema bancário/financeiro brasileiro para as sociedades, e a necessidade da incorporação definitiva do tema nas agendas dos governos locais em todas as partes do mundo, o presente estudo estabelece como problema de pesquisa: existem relacionamentos relevantes entre os estados brasileiros sobre aspectos do consumo de produtos de crédito bancário?

Dessa forma, há necessidade de um direcionamento da oferta no setor de crédito, que poderá ser utilizado socialmente em forma de incentivos direcionados pelo governo a nichos específicos de consumidores de produtos de crédito, como forma de alavancar o consumo e consequentemente a economia. Segundo Matta (2010), em se tratando de aspectos informacionais, é fato que todos os envolvidos em um processo de mudança de comportamento necessitam, buscam e utilizam informações para alcançar os objetivos almejados.

2 Objetivos

O objetivo geral é investigar e evidenciar os relacionamentos existentes entre os estados brasileiros sobre aspectos do consumo de produtos de crédito bancário. Para tanto, serão utilizados dados financeiros e de operações de crédito a partir de dados abertos do Banco Central¹.

¹ Portal de dados abertos do Banco Central do Brasil, disponível em: Disponível em: https://dadosabertos.bcb.gov.br/dataset/scr_data. Acesso em: 12 ago. 2022.

3 Procedimentos Metodológicos

O presente estudo consiste em uma pesquisa de natureza aplicada, de caráter exploratório e descritivo, que visa não só relacionar as variáveis de análise central, mas também apresentar informações que possam servir de diretrizes para ações de transformação da realidade.

Os resultados são apresentados sob a forma qualitativa, a partir da coleta de informações de fontes primárias e secundárias, incluindo revisão bibliográfica. A planificação da pesquisa inclui o levantamento dos dados e a revisão da literatura. A análise está direcionada ao contexto que configura o objeto de estudo, a saber, o sistema bancário brasileiro. Para a pesquisa bibliográfica foram utilizados autores relevantes para a Ciência da Informação. A coleta de dados foi realizada utilizando dados financeiros e de operações de crédito a partir de dados abertos do Banco Central.

Foi utilizada a base de dados denominada “SCR.data - Painel de Operações de Créditos”. O Banco Central do Brasil divulga mensalmente informações agregadas das operações de crédito recebidas através do Sistema de Informações de Créditos – SCR. São disponibilizados dados agregados de Carteira Ativa, Inadimplência e Ativo Problemático com possibilidade de detalhamento por tipo de cliente (PF/PJ), modalidade de crédito, unidade da federação, Classificação Nacional de Atividades Econômicas – CNAE (PJ), natureza da ocupação (PF), porte/rendimento dos clientes, origem de recursos e indexador das operações. A publicação compreende o período de junho de 2012 até a atualidade e é baseada no documento 3040 (SCR), com informações detalhadas de todas as operações de crédito cursadas no país de valor superior a R\$ 1.000 até a data-base de maio/16 e de valor superior a R\$ 200 a partir da data-base junho/16. O total nacional mais 27 unidades da federação estão presentes no relatório.

A presente pesquisa usou 840.114 registros de crédito, que correspondem ao mês de maio de 2022. Das 22 variáveis (campos) existentes na base de dados,

foram selecionadas sete para viabilizar a realização da pesquisa, para um primeiro momento. Nessa seleção das variáveis usou-se o critério daquelas que mais se aproximavam do problema de pesquisa. Para tanto foi consultado o dicionário de dados denominado "metodologia" disponível no site.

A modelagem conceitual foi construída sobre as variáveis selecionadas, usando-se um modelo de entidade e relacionamento (MER)² e a ferramenta de diagramação DBDesigner³. Foi realizada identificando duas entidades centrais, a saber, Clientes e Créditos. Seguindo os princípios de normalização dos dados, identificamos outras entidades relacionadas aos Clientes: UF, Ocupação, e Porte; e aos Créditos: Indexador e Modalidade.

A ferramenta Open Refine⁴ foi usada para limpeza e preparação de dados, realizando a exportação dos dados para o formato GML⁵ para, posteriormente, serem abertos na ferramenta Gephi⁶.

Finalmente, utilizando-se de técnicas de análise de redes complexas, formatações adequadas e projeções bipartidas sobre as redes obtidas, pôde-se revelar os relacionamentos que são apresentados na seção de resultados.

² MER é um modelo de dados para descrever os dados ou aspectos de informação de um domínio de negócio ou seus requisitos de processo. Disponível em: https://pt.wikipedia.org/wiki/Modelo_entidade_relacionamento.

³ A DBDesigner é uma ferramenta CASE para a modelagem de dados que trabalha com o modelo lógico. Disponível em: <https://www.devmedia.com.br/dbdesigner-modelagem-de-dados>.

⁴ OpenRefine é um software para limpeza, preparação e reconciliação de dados. Disponível em: <https://openrefine.org/https://www.devmedia.com.br/dbdesigner-modelagem-de-dados>.

⁵ GML (*graph modelling language*) é um formato universal para representação de grafos. Disponível em: https://en.wikipedia.org/wiki/Graph_Modelling_Language.

⁶ Gephi é um software para visualização, exploração e análise de vários tipos de redes e sistemas complexos, grafos dinâmicos e hierárquicos. Disponível em: <https://gephi.org/>.

4 Resultados

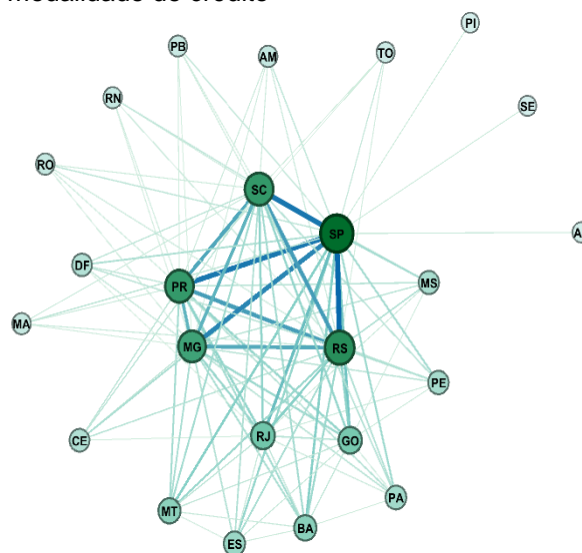
Na modelagem conceitual, sobre as sete variáveis escolhidas da base de dados, Figura 4 do Apêndice A, foram identificadas as principais entidades do modelo: Cliente e Crédito. Em seguida identificou-se os atributos de cada uma das entidades, a normalização dos dados, e os relacionamentos entre os atributos dos consumidores (Cliente) e do Crédito.

Após a preparação dos dados e mapeamento em formato de rede GML, foram construídas três redes monopartidas de estados brasileiros para realização de análise comparativa pela modalidade de crédito, pelo porte salarial e pela ocupação.

A Figura 1 apresenta a rede monopartida de estados obtida por uma projeção bipartida pela eliminação da variável modalidade de crédito. As modalidades das operações de crédito são agrupadas em:

- PF - Cartão de crédito
- PF - Empréstimo com consignação em folha
- PF - Empréstimo sem consignação em folha
- PF - Habitacional
- PF - Outros créditos
- PF - Rural e agroindustrial
- PF - Veículos

Figura 1 - Rede monopartida de estados, por modalidade de crédito



Fonte: autoria própria, com apoio do software Gephi

Na inspeção visual da rede da Figura 1, os estados que mais se destacaram no

com volumes semelhantes e mais fortes em relação aos outros estados.

Condensando os resultados, cabe portanto salientar o quanto os estados SP, RS, SC, PR e MG são relevantes perante à sociedade, nos aspectos relativos ao consumo de produtos bancários, bem como o incentivo econômico gerado pelos mesmos, alavancando também o consumo de outros estados mais dependentes.

Os grafos obtidos através da análise dos dados abertos do banco central possuem grande potencial para abstrair informações relevantes para a área financeira. Segundo Passos (2020) a teoria de redes complexas pode ser utilizada ainda para descrever e modelar sistemas de direcionamento de produtos e serviços das empresas, criando um sistema de recomendação. Com a projeção de rede bipartida é possível comparar o método com outras formas de recomendação. São inúmeras as possibilidades diante das bases abertas do Banco Central, selecionando por exemplo as tags relativas às tarifas teria-se um arcabouço para a criação de um comparador de instituições financeiras em diversos aspectos que trariam informações relevantes do ponto de vista da sociedade e sem conflitos de interesses devido à natureza dos dados utilizados.

4 Considerações Finais

Como a Ciência das Redes se configura como área relativamente nova no âmbito das ciências, as pesquisas estão se ajustando e se desenvolvendo em conjunto com novas tecnologias. A exploração de sua interdisciplinaridade, a análise da aplicabilidade de ferramentas, e a descoberta de relações com outros saberes e informações, emerge, então, como uma área promissora para pesquisa e para a Ciência da Informação.

O desenvolvimento deste estudo, bem como o exame da literatura mostrou que o cenário atual encontra-se promissor para aplicação das técnicas empregadas. Com o estudo pode-se observar que a aplicabilidade prática dessa técnica na área financeira é satisfatória e promissora, pois possibilita gerar insights com execuções interessantes

principalmente para a área de gestão da informação e do conhecimento.

A partir dos dados abertos sobre operações de crédito, disponibilizados pelo Banco Central, foi realizada uma modelagem conceitual e a construção de três redes monopartidas de estados considerando-se a modalidade de crédito, o porte salarial e a ocupação. Contudo, utilizou-se um número reduzido de variáveis como forma de viabilizar o experimento.

Nas três redes monopartidas destacaram-se os estados SP, RS, SC, PR e MG, observando-se que eles são relevantes perante à sociedade, nos aspectos relativos ao consumo de produtos bancários.

O objetivo proposto, a saber, de investigar, e trazer relacionamentos relevantes entre os estados brasileiros sobre aspectos do consumo de produtos de crédito bancário foi alcançado e evidenciado com a modelagem conceitual e, principalmente, pela análise das redes geradas por meio de inspeção visual.

Os resultados da pesquisa podem orientar a aplicação de incentivos direcionados a nichos específicos de consumidores de produtos de crédito, para incentivar o consumo em determinada área ou estado, e consequentemente estimular a economia.

A inspeção visual s redes demonstraram que existem relacionamentos relevantes entre os estados brasileiros sobre aspectos do consumo de produtos de crédito bancário, e que podem ser ainda explorados sob diversos aspectos da base de dados utilizada, visando trabalhos futuros com a mesma base ou outras subjacentes que venham tratar com questões interessantes sobre por exemplo a inadimplência da população.

Acredita-se que este trabalho seja um passo inicial para contribuir com os diversos usuários do sistema financeiro, mais especificamente do mercado de crédito bancário, no sentido de auxiliar sobre o direcionamento de investimentos a partir da constatação de que existem relacionamentos relevantes entre os estados brasileiros no que tange a produtos de crédito bancário, porte salarial, e natureza da ocupação dos usuários.

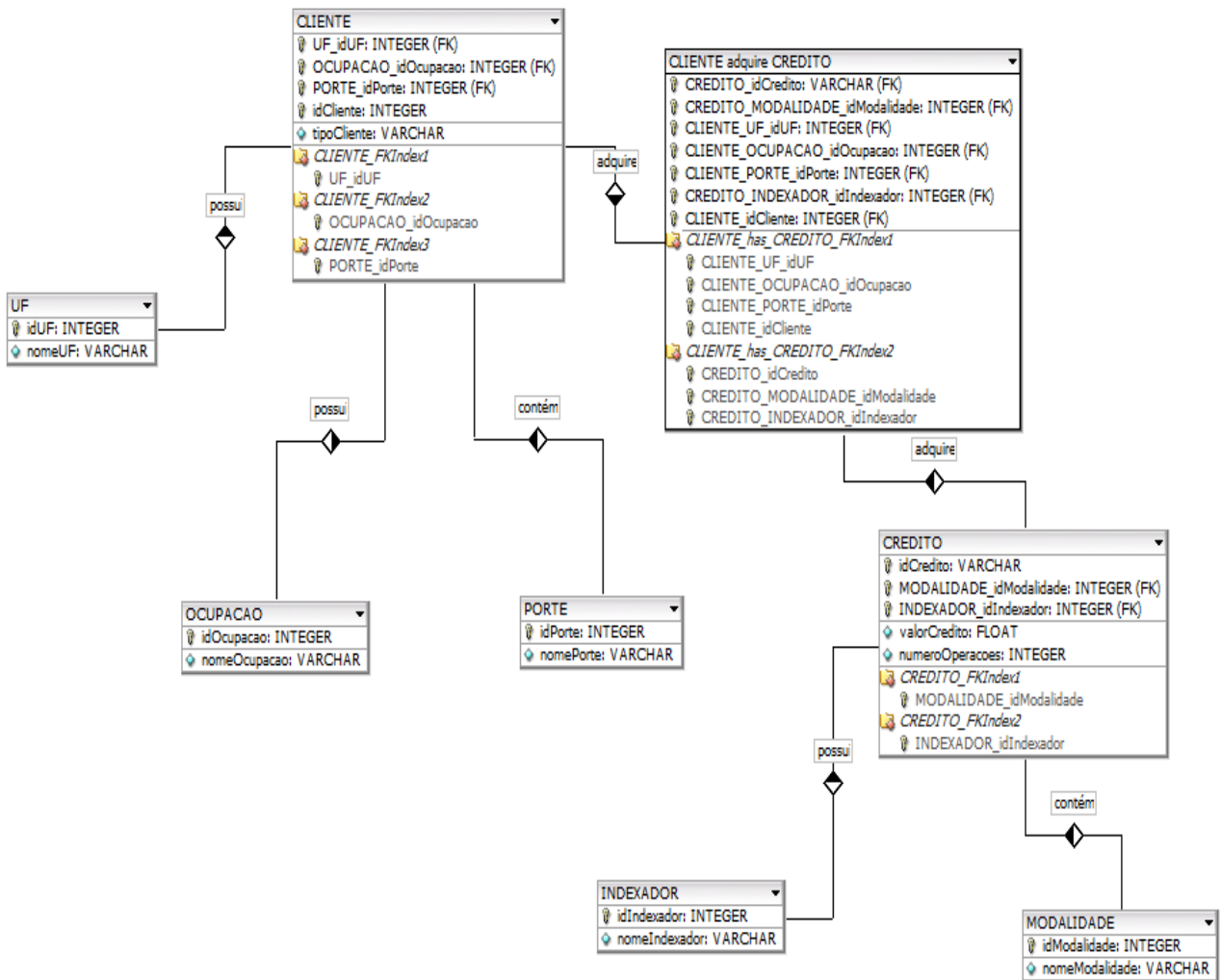
Nesse sentido, um exemplo de aplicação poderia ser a partir de órgãos públicos estaduais ou governamentais que tenham interesse no direcionamento de créditos de incentivo financeiro para a população, incentivo fiscal para a indústria e comércio da região que tenham características comuns, ou mesmo políticas de educação financeira da população. Podem ser determinadas políticas de incentivo ao consumo de determinado produto de crédito como redução de juros à medida que fique evidenciada a necessidade latente da população, como exemplo, os créditos rural, agro e de habitação de dada região. Podem ser verificados também perfis comportamentais de consumo dos habitantes, em função da faixa salarial e de natureza da ocupação de determinado estado ou região que aparecem nos grafos de forma semelhante.

Referências

- BARABÁSI, Albert-László. Ciência da rede. *Transações Filosóficas da Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 371, n. 1987, p. 20120375, 2013.
- CAVALCANTE, Gustavo Vasconcellos. *Ciência das Redes: aspectos epistemológicos*. 2009.
- DIAS, Mateus Pereira; CARVALHO, José Oscar Fontanini de. *A Visualização da Informação e a sua contribuição para a Ciência da Informação*. *DataGramZero*, Rio de Janeiro, v. 8, n. 5, p. 1-16, 2007.
- GITAHY, Yuri. *O que é uma startup. Empreendedor Online—Empreendedorismo na Internet e negócios online*, 2011.
- GONZALEZ, L.; CERNEV, A. K.; DINIZ, E. O. *Desafio da Inclusão Financeira e a Promessa das Fintechs. Sistema Financeiro Nacional: o que fazer*, 2016
- GUIZZARDI, Giancarlo. *Fundamentos ontológicos para modelos conceituais estruturais*. 2005.
- ISOTANI, Seiji; BITTENCOURT, Ig Ibert. *Dados abertos conectados: em busca da web do conhecimento*. Novatec Editora, 2015.
- JUENGERKES, Bjoern Erik. *FinTechs and Banks—Collaboration is Key. The FinTech Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries*, p. 179-182, 2016
- Mais liberdade de escolha e serviços para você. *Open Banking Brasil*, 2022. Disponível em: <https://openbankingbrasil.org.br/?cookie=true>. Acesso em: 18 set. 2022.
- MATTA, Rodrigo Octávio Beton. *Modelo de comportamento informacional de usuários: uma abordagem teórica. Gestão, mediação e uso da informação*, p. 127-142, 2010.
- Painel de Operações de Créditos. 22 jul. 22. Disponível em: https://dadosabertos.bcb.gov.br/dataset/scr_data. Acesso em: 12 ago. 2022.
- PASSOS, Gabriela. *Redes Bipartidas para recomendação de produtos financeiros*. 2020. Tese de Doutorado. Universidade de São Paulo.
- VALENTIM, Marta Lígia Pomim. *Capítulo 1 PROCESSO DE INTELIGÊNCIA COMPETITIVA ORGANIZACIONAL. Informação, conhecimento e inteligência organizacional*, p. 9, 2006.
- VALENTIM, Marta Lígia Pomim. *Gestão da Informação e Gestão do Conhecimento em ambientes organizacionais. Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 1, n. 1, 2008.
- SAYÃO, Luís Fernando. *Modelos teóricos em ciência da informação-abstração e método científico. Ciência da informação*, v. 30, p. 82-91, 2001.
- WASSERMAN, Stanley; FAUST, Katherine. *Social network analysis: methods and applications*. Cambridge, England; New York: Cambridge University Press, 1994.

Apêndice A

Figura 4 - Modelagem conceitual sobre sete variáveis da base de dados



Fonte: autoria própria, com apoio do software DBDesigner.