



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

JULIANA RODRIGUES DE LIMA MEIRELLES

**O FENÔMENO OPEN FINANCE COM FOCO NA SOCIALIZAÇÃO DA
INFORMAÇÃO ANALISADO A PARTIR DOS DADOS ABERTOS DO BANCO
CENTRAL DO BRASIL SOBRE OPERAÇÕES DE CRÉDITO COM SUPORTE DE
ANÁLISE DE REDES DE INFORMAÇÃO**

Vitória-ES

2024

JULIANA RODRIGUES DE LIMA MEIRELLES

**O FENÔMENO OPEN FINANCE COM FOCO NA SOCIALIZAÇÃO DA
INFORMAÇÃO ANALISADO A PARTIR DOS DADOS ABERTOS DO BANCO
CENTRAL DO BRASIL SOBRE OPERAÇÕES DE CRÉDITO COM SUPORTE DE
ANÁLISE DE REDES DE INFORMAÇÃO**

Dissertação de mestrado apresentada ao
Programa de Pós-Graduação em Ciência
da Informação da Universidade Federal
do Espírito Santo PPGCI-UFES como
requisito para obtenção do título de
Mestre em Ciência da Informação.

Orientador: Prof. Doutor Henrique
Monteiro Cristovão

Linha de pesquisa 2: Memória,
Representação e Informação

VITÓRIA-ES

2024

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

R696f RODRIGUES DE LIMA MEIRELLES, JULIANA, 1986-
O FENÔMENO OPEN FINANCE COM FOCO NA
SOCIALIZAÇÃO DA INFORMAÇÃO ANALISADO A PARTIR
DOS DADOS ABERTOS DO BANCO CENTRAL DO
BRASIL SOBRE OPERAÇÕES DE CRÉDITO COM SUPORTE
DE ANÁLISE DE REDES DE INFORMAÇÃO / JULIANA
RODRIGUES DE LIMA MEIRELLES. - 2024.
108 p. : il.

Orientador: Henrique Monteiro Cristovão.
Dissertação (Mestrado em Ciência da Informação) -
Universidade Federal do Espírito Santo, Centro de Ciências
Jurídicas e Econômicas.

1. Análise de redes de informação. 2. Visualização de
informação. 3. Descoberta de conhecimento. 4. Open Finance. 5.
Operações de crédito. 6. Socialização da informação. I. Monteiro
Cristovão, Henrique. II. Universidade Federal do Espírito Santo.
Centro de Ciências Jurídicas e Econômicas. III. Título.

CDU: 02

JULIANA RODRIGUES DE LIMA MEIRELLES

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo (PPGCI/UFES) como requisito para a obtenção do título de Mestra em Ciência da Informação.

Linha de pesquisa 2: Memória, Representação e Uso da informação

Aprovada em 27 de junho de 2024.

Comissão examinadora:

[assinatura digital]

Prof. Dr. Henrique Monteiro Cristovão
Orientador (PPGCI/UFES)

[assinatura digital]

Profa. Dra. Daniela Lucas da Silva Lemos
PPGCI/UFES

[assinatura digital]

Prof. Dr. José Eduardo Santarem Segundo
Universidade de São Paulo (USP)





Folha de aprovação - Juliana

Data e Hora de Criação: 27/06/2024 às 20:47:11

Documentos que originaram esse envelope:

- Folha de aprovação - Juliana.pdf (Arquivo PDF) - 1 página(s)



Hashs únicas referente à esse envelope de documentos

[SHA256]: dd565cdbf1058a9da2100b37dcaceec65ee6440bb60c756786c22a74763e2c50

[SHA512]: 969cee47ca5f7ab39aef525659c6121e384122a3ed5f7d8e7e2412bea768315c6798bcbf3a05e5ab8e10073ef67c3932a558abf3cddc4672fae077036f72b462

Lista de assinaturas solicitadas e associadas à esse envelope



ASSINADO - Henrique Monteiro Cristovão (henrique.cristovao@ufes.br)

Data/Hora: 27/06/2024 - 20:54:19, IP: 177.104.235.128, Geolocalização: [-20.258216, -40.264282]

[SHA256]: a7c8543f9fe1bc90695be253109ee8c75920cb9a6eabea58e309b6a334f7742c



ASSINADO - Daniela Lucas da Silva Lemos (daniela.l.silva@ufes.br)

Data/Hora: 30/06/2024 - 10:14:12, IP: 187.36.170.78, Geolocalização: [-20.296932, -40.292621]

[SHA256]: 28f7576a7afda6c38e063e1d48d25b1e36fc202e07c74b6064896776ddb34a42



ASSINADO - José Eduardo Santarem Segundo (santarem@usp.br)

Data/Hora: 30/06/2024 - 18:42:09, IP: 177.94.244.93

[SHA256]: cc0ef2163b961ebf67455be3bf97176e384e67e57233a56890aec246088d8b66

Histórico de eventos registrados neste envelope

30/06/2024 18:42:09 - Envelope finalizado por santarem@usp.br, IP 177.94.244.93

30/06/2024 18:42:09 - Assinatura realizada por santarem@usp.br, IP 177.94.244.93

30/06/2024 18:42:04 - Envelope visualizado por santarem@usp.br, IP 177.94.244.93

30/06/2024 10:14:12 - Assinatura realizada por daniela.l.silva@ufes.br, IP 187.36.170.78

30/06/2024 10:14:01 - Envelope visualizado por daniela.l.silva@ufes.br, IP 187.36.170.78

27/06/2024 20:54:19 - Assinatura realizada por henrique.cristovao@ufes.br, IP 177.104.235.128

27/06/2024 20:54:14 - Envelope visualizado por henrique.cristovao@ufes.br, IP 177.104.235.128

27/06/2024 20:49:26 - Envelope registrado na Blockchain por edma.jantorno@ufes.br, IP 200.137.65.102

27/06/2024 20:49:25 - Envelope encaminhado para assinaturas por edma.jantorno@ufes.br, IP 200.137.65.102

27/06/2024 20:47:11 - Envelope criado por edma.jantorno@ufes.br, IP 200.137.65.102

RESUMO

No cenário bancário e financeiro do Brasil ocorrem transformações que estão influenciando a maneira como as pessoas lidam com suas finanças. O uso de dados emerge como um elemento crucial de competitividade entre produtos, destacando-se como fator essencial na dinâmica atual do sistema financeiro, onde a sociedade busca adaptar-se às novas realidades. Nesse contexto, a problemática da pesquisa é delineada nas bases de dados abertas do Banco Central com a questão de como esses dados podem fornecer insights para benefícios sociais. Sendo assim, esta pesquisa tem como objetivo geral investigar variáveis presentes nas bases de dados abertas do Banco Central do Brasil, relacionadas às operações de crédito, no contexto do Open Finance, para a revelação de insights relevantes, considerando possibilidades de socialização da informação para benefício social. Sendo uma pesquisa qualitativa de natureza aplicada, de caráter exploratório e descritivo, a pesquisa se utiliza, principalmente, de visualização de informação com técnicas de análise de redes complexas. Como resultados, tem-se uma modelagem de dados baseada em uma ontologia de referência de domínio; elementos visuais estatísticos sobre os dados da base; redes de informação resultantes de mapeamentos realizados sobre os dados. Esses resultados podem ser utilizados para uma compreensão de como diferentes grupos de clientes interagem com o mercado de crédito nesse contexto de mudanças. Essas informações desempenham um papel importante no desenvolvimento de políticas públicas mais inclusivas. Ao identificar comportamentos de pagamento ou necessidades de crédito específicas em determinadas regiões ou grupos semelhantes, as autoridades podem direcionar políticas de incentivo fiscal para facilitar o acesso ao crédito para comunidades historicamente desfavorecidas. Os resultados sinalizam a possibilidade de benefícios sociais, como a educação financeira da população com informações que não são disponibilizadas pelas instituições, garantindo assim o legítimo interesse da coletividade.

Palavras chave: Análise de redes de informação. Visualização de informação. Descoberta de conhecimento. Open Finance. Operações de crédito. Socialização da informação.

ABSTRACT

In the Brazilian banking and financial scenario, transformations are occurring that are influencing the way people deal with their finances. The use of data is emerging as a crucial element of competitiveness between products, standing out as an essential factor in the current dynamics of the financial system, where society seeks to adapt to new realities. In this context, the research problem is outlined in the open databases of the Central Bank with the question of how these data can provide insights for social benefits. Therefore, this research has the general objective of investigating variables present in the open databases of the Central Bank of Brazil, related to credit operations, in the context of Open Finance, to reveal relevant insights, considering possibilities of socializing information for social benefit. Being a qualitative research of an applied nature, of an exploratory and descriptive nature, the research uses, mainly, information visualization with complex network analysis techniques. As results, there is a data modeling based on a domain reference ontology; statistical visual elements on the database data; information networks resulting from mappings performed on the data. These results can be used to understand how different groups of customers interact with the credit market in this changing context. This information plays an important role in developing more inclusive public policies. By identifying specific payment behaviors or credit needs in certain regions or similar groups, authorities can target fiscal incentive policies to facilitate access to credit for historically disadvantaged communities. The results indicate the possibility of social benefits, such as financial education of the population with information that is not made available by institutions, thus ensuring the legitimate interest of the community.

Keywords: Information network analysis. Information visualization. Knowledge discovery. Open Finance. Credit operations. Information socialization.

SUMÁRIO

1 INTRODUÇÃO.....	6
1.1 PROBLEMA DE PESQUISA.....	8
1.2 OBJETIVO GERAL.....	9
1.3 OBJETIVOS ESPECÍFICOS.....	10
1.4 JUSTIFICATIVA.....	10
1.5 ORGANIZAÇÃO DA DISSERTAÇÃO.....	13
2 ESTADO DA ARTE DA ANÁLISE DE REDES DE INFORMAÇÃO NO CONTEXTO DAS OPERAÇÕES DE CRÉDITO NO OPEN FINANCE.....	14
3 REFERENCIAL TEÓRICO.....	25
3.1 CIÊNCIA DA INFORMAÇÃO E ORGANIZAÇÃO DO CONHECIMENTO.....	25
3.1.1 Socialização da Informação.....	26
3.1.2 Sistemas de Organização de Conhecimento.....	27
3.1.3 Ontologias.....	28
3.1.4 Ferramenta para modelagem de dados.....	31
3.2 CIÊNCIA DE DADOS.....	31
3.2.1 Dados, <i>Data Driven</i> e <i>Data Literacy</i>	32
3.2.2 Mineração de Dados e Descoberta de Conhecimento.....	34
3.2.3 Ferramenta para manipulação de dados.....	36
3.2.4 Visualização de Informação.....	36
3.2.5 Ferramenta para visualização de informação.....	38
3.3 CIÊNCIA DAS REDES.....	39
3.3.1 Redes Complexas.....	39
3.3.2 Análise de Redes de Informação.....	40
3.3.3 Redes Bipartidas e Projeção Bipartida.....	43
3.3.3 Ferramenta para análise de redes de informação.....	46
3.3.4 Ciência das Redes na Ciência da Informação.....	46
3.4 OPEN FINANCE.....	47
3.4.1 Surgimento.....	48
3.4.2 Funcionamento e Tendências.....	49
3.4.3 Open Finance na Ciência da Informação.....	52
4 PROCEDIMENTOS METODOLÓGICOS.....	54
4.1 CLASSIFICAÇÃO DA PESQUISA.....	54
4.2 LEVANTAMENTO BIBLIOGRÁFICO.....	55
4.2.1 Levantamento bibliográfico do Estado da Arte.....	56
4.3 CENÁRIO DA PESQUISA.....	57

4.4 ETAPAS PROCEDIMENTAIS DO DESENVOLVIMENTO.....	60
4.4.1 Etapa 1: Investigação de Bases de Dados.....	62
4.4.2 Etapa 2: Criação do Modelo Conceitual.....	62
4.4.3 Etapa 3: Preparação e Transformação de Dados.....	62
4.4.4 Etapa 4: Modelo de Dados e Mapeamento dos Dados.....	64
4.4.5 Etapa 5: Mineração e Visualização de Dados.....	64
4.4.6 Etapa 6: Análise e Interpretação dos Dados.....	64
5 RESULTADOS E DISCUSSÃO.....	66
5.1 Variáveis selecionadas e modelo conceitual.....	66
5.2. Pré-processamento, modelo de dados e mapeamento para rede de informação.....	74
5.3. Visualização da informação.....	78
5.4. Discussão geral.....	86
6 CONSIDERAÇÕES FINAIS.....	90
REFERÊNCIAS.....	93
ANEXOS.....	107

1 INTRODUÇÃO

O mundo passa por grandes transformações no que tange à gestão e uso da informação. Diante do avanço das tecnologias, o fenômeno da *datafication* se consolida cada vez mais em diversas áreas. O *datafication* vai além da simples digitalização, converte em dados todos os tipos de comportamento de uma sociedade transformando-os em análises, gerando previsões por meio de processamento algorítmico, servindo para os mais diversos objetivos, transformando inclusive as formas de sociabilidade de toda uma sociedade (Lemos, 2021).

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 fator de competitividade de um produto para o outro (Valentim, 2002). Mais cedo ou mais tarde acontecerá a adesão a autorizações, aceites e concordâncias com contratos pré determinados por instituições financeiras. Gonzalez (2016) argumenta que a demanda por serviços financeiros mais eficientes ainda é significativa, impulsionando a criação crescente de *startups*, como as *FinTechs* (do inglês: *finance and technology*). Os *FinTechs* são orientados pela tecnologia e desvinculados de sistemas legados, buscam modelos de negócios repetíveis e escaláveis (Gitahy, 2011; Juengerkes, 2016).

Outro acontecimento importante e atual no setor financeiro é o surgimento do Open Finance, ou sistema financeiro aberto. Ele promete revolucionar a forma como se lida com produtos, serviços e dados financeiros dos cidadãos. Essa nova realidade traz consigo a inovação, o surgimento de novos modelos de negócios e difusão de informação, favorecendo a inclusão e a educação financeira da população (Open Banking Brasil, 2022). Segundo Valentim (2008), deve-se refletir sobre os dispositivos que transformam dados em informação. Investigar o domínio da informação sob diferentes enfoques é parte integrante do processo de conhecimento da realidade. Organizar e representar a informação é um processo por meio do qual a realidade é aprendida e sistematizada (Araújo, 2013). O autor enfatiza ainda a importância dos estudos em representação da informação em domínios específicos tem aumentado. A construção de sistemas de representação como ontologias, contribui para uma melhor compreensão e organização da

informação nesses contextos específicos. Essa estruturação possibilita uma abordagem mais clara e precisa na interpretação de dados complexos.

Saber lidar com o volume, diferentes origens e formatos de dados é uma das habilidades mais valorizadas das últimas décadas (Davenport et al., 2012). Em paralelo, dados em formato aberto são dados disponibilizados de forma livre e democrática a qualquer cidadão. Segundo Isotani (2015), os dados abertos podem ser acessados, utilizados e redistribuídos por qualquer pessoa, desde que estejam de acordo com as exigências de licença. O Banco Central, por meio de sua Política de Dados Abertos¹, busca estimular a interoperabilidade de dados e sistemas governamentais, promover a produção de conhecimento, fomentar a gestão pública participativa e incentivar a inovação em setores que prestem serviços aos cidadãos. Além disso, fortalece a cultura de transparência, contribuindo indiretamente para a missão de promover o bem-estar econômico da sociedade.

Nesse contexto, a área da descoberta de conhecimento, também denominada de *knowledge discovery in databases* (KDD), se propõe a identificar possíveis relacionamentos interessantes entre uma gama de variáveis selecionadas (Santos, 2001). Essa área está relacionada à visualização de informação quando evidencia os relacionamentos, antes implícitos, através de objetos cognitivos que facilitam o entendimento do usuário final. Segundo Dias (2007), um dos principais objetivos da visualização de informação é justamente a inclusão informacional dos indivíduos.

Vive-se hoje em uma sociedade altamente tecnológica e conectada (Cavalcante, 2009) onde as conexões têm sido difundidas principalmente com o advento da internet, mais especificamente pela World Wide Web. Várias redes surgem e evoluem, baseados em leis e mecanismos que são a base para uma ciência denominada de Ciência das Redes (Barabási, 2013). Inclui-se na Ciência das Redes a análise de redes sociais (ARS). Essa área preocupa-se com o entendimento das ligações entre entidades sociais e as implicações destas ligações

¹ A Política de Dados Abertos do Poder Executivo Federal foi instituída pelo Decreto nº 8.777, de 11 de maio de 2016, tem por objetivo a disponibilização na internet – por parte de órgãos e entidades da administração pública federal direta, autárquica e fundacional – de dados e informações acessíveis ao público que possam ser livremente lidos por máquina, usados, cruzados e reutilizados, fomentando-se o controle social, o desenvolvimento tecnológico, o aprimoramento da cultura de transparência pública e a inovação nos diversos setores da sociedade. Disponível em: <<https://www.bcb.gov.br/acessoinformacao/dadosabertos>>. Acesso em: 20 jan. 2024.

(Wasserman; Faust, 1994). Para a análise de redes de informação, também conhecida por análise de redes complexas, utilizam-se praticamente as mesmas técnicas da ARS.

Segundo Cavalcante (2008), a Ciência da Informação vem colaborando para o desenvolvimento da Ciência das Redes. 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.

A informação deve ser estruturada e organizada de alguma forma para que se torne utilizável (McGarry, 1999). A organização do conhecimento vem atuando como precursor da recuperação de informação e trazendo as bases para que ela seja realizada de forma eficaz. A organização do conhecimento está preocupada em descrever, classificar, e indexar informações e preocupa-se também com a natureza e qualidade de tais processos (Hjorland, 2016), destacando sua importância na busca eficiente por conhecimento em Ciência de Dados.

1.1 PROBLEMA DE PESQUISA

Em um mundo onde a tecnologia proporciona novidades a cada dia, é crucial que as pessoas não sejam levadas apenas pela novidade, mas que busquem compreender a fundo essas transformações tecnológicas. Sobretudo ao fazer escolhas, livres das sugestões automáticas, tenham plena consciência da responsabilidade perante a consequência de suas escolhas.

Nesse contexto, o sistema bancário/financeiro brasileiro passa por mudanças tecnológicas que não apenas afetarão a maneira como a sociedade gerencia suas finanças, mas também têm o potencial de redefinir significativamente o mercado de crédito. Cada vez mais a experiência do cliente é enfatizada como sendo o fator de distinção de um produto para o outro, gerando o chamado “empoderamento do cliente bancário” (Pereira, Oliveira, Ceci, 2022). Surge então uma nova categoria de empresas no ramo financeiro: os *FinTechs* (do inglês: *finance and technology*). Os *FinTechs* são orientados pela tecnologia e desvinculados de sistemas legados, buscam modelos de negócios repetíveis e escaláveis (Gitahy, 2011; Juengerkes, 2016).

O consumidor não obtém informações a respeito de produtos e serviços financeiros com a mesma velocidade com que as inovações surgem (Serasa, 2023). A falta dessas informações pode impactar diretamente em escolhas desfavoráveis por parte dos clientes do sistema bancário/financeiro. Em muitos casos os clientes são meramente seduzidos por sugestões dos sistemas sem ter com clareza as consequências de suas escolhas.

Diante dessa realidade, o sistema bancário tem importância fundamental no desenvolvimento sustentável da sociedade e é de suma relevância que o consumidor esteja bem informado ao relacionar-se com esse mercado. Más escolhas por parte dos consumidores em um ambiente global podem resultar em diversas consequências negativas, impactando não apenas os indivíduos, mas também a economia, o meio ambiente e a sociedade como um todo. Algumas das implicações podem incluir endividamento excessivo, levando a problemas financeiros significativos, instabilidade financeira decorrente da falta de compreensão sobre investimentos, impactos ambientais por escolhas de consumo não sustentáveis, perpetuação de desigualdades sociais e riscos para a saúde associados a decisões inadequadas. Sendo assim, os dados desempenham um papel fundamental na busca pela consecução das chamadas Finanças Sustentáveis, que buscam respeitar as pessoas e o meio ambiente (Fia, 2024).

Uma definição proposta por Norbert Wiener, fundador da cibernética, citado em McGarry (1999), salienta que a vida, de fato, carece de informação. A sociedade está diante do início de um novo Mercado Financeiro (Embratel, 2021). O Open Finance, ao prometer revolucionar a forma como se lida com produtos, serviços, dados financeiros e, em particular, o mercado de crédito dos cidadãos (Open Finance Brasil, 2022), surge como um catalisador de transformações significativas no cenário financeiro, abrindo novas possibilidades e dinâmicas para a interação entre consumidores, instituições financeiras e os diversos elementos que permeiam o mercado de crédito (Open Finance Brasil, 2022).

Dessa forma, o presente estudo estabelece a seguinte questão de pesquisa: como extrair informações das bases de dados abertas do Banco Central do Brasil, relacionadas às operações de crédito, no contexto do Open Finance, para a revelação de insights que possam auxiliar a socialização deste tipo de informação para benefício social?

1.2 OBJETIVO GERAL

A pesquisa tem como objetivo geral investigar variáveis presentes nas bases de dados abertas do Banco Central do Brasil, relacionadas às operações de crédito, no contexto do Open Finance, para a revelação de insights relevantes, considerando possibilidades de socialização da informação para benefício social.

1.3 OBJETIVOS ESPECÍFICOS

São objetivos específicos da presente pesquisa:

i. Analisar à luz da literatura, no campo da Ciência da Informação, os conceitos de visualização de informação, modelagem conceitual, análise de redes complexas e Open Finance.

ii. Investigar bases de dados abertas do Banco Central do Brasil relativas a dados sobre operações de crédito, no contexto do Open Finance, e selecionar variáveis promissoras e relevantes para a pesquisa.

iii. Criar um modelo conceitual para representar e facilitar o entendimento do contexto que envolve as variáveis selecionadas no objetivo específico (ii).

iv. Mapear os dados referentes às variáveis selecionadas no objetivo específico (ii) e modeladas no objetivo específico (iii) para o formato de redes de informação.

v. Aplicar técnicas de análise de redes e visualização de informação para revelar relacionamentos relevantes entre as variáveis selecionadas no objetivo específico (ii) e modeladas no objetivo específico (iii) sobre a rede de informação criada no objetivo específico (iv), com foco em possíveis contribuições sociais.

vi. Inferir possíveis contribuições sociais, por meio da socialização da informação, que os relacionamentos revelados no objetivo específico (v) poderão fomentar.

1.4 JUSTIFICATIVA

A disponibilização dos dados, essencial no contexto contemporâneo de inteligência coletiva, desempenha um papel crucial no ambiente competitivo delineado pelo Open Finance, especialmente no mercado de crédito. Conforme

destacado por Lévy (1998), a capacidade de grupos se tornarem coletivos inteligentes está diretamente relacionada à adaptação eficaz nesse ambiente dinâmico. Ele ressalta que a prosperidade das nações está intrinsecamente ligada à habilidade de gerenciar eficientemente todo esse conhecimento. Dessa maneira, além da hiperpersonalização proporcionada por esse novo arranjo do sistema financeiro, o cliente ganha uma perspectiva mais abrangente ao adquirir produtos ou serviços financeiros, inserindo-se em um ecossistema de dados abertos no qual ele é o ator central (Waard, Cousijn, Aalbersberg, 2016).

Ferreira (2019) traz pistas do que seria um cidadão do mundo global: necessariamente informado, que percebe o mundo digital não por intuição, mas por percepção do que está à sua frente. Ele enfatiza a democratização dos dados como caminho para interoperabilidade e consequente digitalização de toda uma sociedade. E assegura que abrindo caminho à digitalização, tem-se como resultado poupar tempo, reduzir custos, aumentar a transparência e melhorar a qualidade dos dados e serviços. Conceito semelhante a esse, é visto nas instituições que estão em processo de transformação digital, por um fenômeno denominado *Data Driven* (Mason, Patil, 2021). Esse processo ratifica a necessidade emergente de sermos “orientados por dados”, e traz ainda como parte do processo o conceito de Alfabetização em dados (*Data literacy*) (Shields, 2005).

Cabe ressaltar as evidências de que o tema escolhido, dados abertos do Banco Central do Brasil, relacionados às operações de crédito, no contexto do fenômeno Open Finance tem relação direta com os estudos da Ciência da Informação e a linha de pesquisa, a saber, “Memória, Representação e Informação”. Inicialmente a relação com a Ciência da Informação se dá a partir do momento em que a essência dessa pesquisa preocupa-se com a análise, coleta, classificação, manipulação, armazenamento, recuperação e disseminação da informação, incluindo a pesquisa sobre representação da informação (Borko, 1968). A pesquisa em questão é categorizada por uma profunda conexão com a Ciência da Informação, evidenciando uma íntima relação com a tecnologia, conforme destacado por Saracevic (1996) e de sua raiz de caráter tecnológico (Capurro, 2017). Ela dedica-se ainda ao estudo dos fenômenos ligados à informação relacionados à geração de produtos e serviços por meio da organização da informação utilizando, para isso, tecnologias de recuperação da informação (Rabello, 2012). Essa proximidade entre pesquisa e tecnologia é uma característica fundamental da

Ciência da Informação, tornando-a uma participante ativa na evolução da Sociedade da Informação.

Um aspecto distintivo dessa pesquisa é a abordagem que propõe uma reflexão sobre eventos e suas influências, considerando-os como uma tecnologia intelectual. Essa perspectiva, conforme expressa por Levy (1993), sugere que a essência da pesquisa não está apenas no objeto em si, mas na maneira como evoluímos como um todo, analisando o contexto e os objetos associados. Além disso, destaca-se a ideia de que as tecnologias intelectuais não estão restritas apenas aos objetos, mas também estão presentes nos sujeitos por meio da imaginação e da aprendizagem. Essa concepção ampla permite uma compreensão mais abrangente do papel da pesquisa no desenvolvimento intelectual e na evolução da sociedade como um todo, proporcionando uma valiosa contribuição para a compreensão do papel da pesquisa na sociedade da informação.

Essa nova dinâmica do mercado financeiro tem o potencial de impactar diretamente as informações presentes na base de dados abertas do Banco Central do Brasil sobre operações de crédito. A competição entre instituições financeiras pode refletir nas informações sobre oferta e demanda, bem como no tipo de crédito disponível. O acesso a dados mais diversificados e personalizados possibilita ao consumidor tomar decisões mais embasadas, alinhadas às suas necessidades específicas e preferências. Dessa forma, o Open Finance não apenas empodera o consumidor, mas também contribui para uma maior eficiência e transparência no mercado financeiro.

O resultado da análise dos relacionamentos existentes nas bases de dados abertos, especificamente no contexto do mercado de crédito, do Banco Central do Brasil no âmbito do fenômeno Open Finance, podem trazer não só trazer benefícios significativos à educação financeira da população, mas proporcionar políticas públicas eficazes. Tais políticas podem incluir incentivos fiscais, como isenção de impostos para serviços de crédito direcionados a determinados perfis de clientes, alinhando-se com as demandas da política pública em vigor. Além disso, ao fornecer informações que não são disponibilizadas pelas instituições, contribui para a transparência e o empoderamento dos consumidores, garantindo assim o legítimo interesse da coletividade (Silva, 2005). Entende-se com isso que esta pesquisa cumpre com essas características e tem foco na satisfação das necessidades do usuário. Ressalta-se ainda o caráter sociológico dessa pesquisa, como citada por

Araújo (2018), que caracteriza esse modelo de pesquisa como sendo valorizador do contexto onde a informação reside e retrata ainda que sua manifestação se dá, entre outras, na análise de domínio.

Por fim, cabe relatar a socialização da informação, que pretende trazer ao usuário comum informações que não estão explícitas. A área de Interação humano-dados (IHD), área análoga a IHC (Interação humano-computador), se propõe a incluir os humanos no centro do fluxo de dados. Esta abordagem não é exclusiva da Ciência da Computação, mas é também de interesse da Ciência da Informação e evidencia sua importância na recuperação da informação (Carvalho, 2003). A Ciência da Informação caracteriza-se como ciência social a partir do momento que percebe o sujeito como principal ator dos sistemas de informação (Araújo, 2018). Segundo o autor, esse modelo de pesquisa valoriza o contexto onde a informação reside e retrata ainda que sua manifestação se dá, entre outras, na análise de domínio.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

A dissertação está organizada em 7 capítulos. O Capítulo 2 apresenta o estado da arte na análise de redes de informação no contexto do sistema financeiro, trazendo trabalhos mais recentes e correlatos à presente pesquisa. O Capítulo 3 trata da referencial teórico da pesquisa abordando temas como a Ciência da Informação e a Organização do Conhecimento, a Ciência de Dados, a Ciência das Redes, e o Open Finance. O Capítulo 4 trata da fundamentação metodológica com abordagens conceituais e ferramentais da pesquisa para fornecer subsídios ao entendimento dos procedimentos metodológicos que são abordados nos Capítulos 5 e 6. O Capítulo 5, por sua vez, trata dos procedimentos metodológicos, incluindo o levantamento bibliográfico, o cenário da pesquisa e as etapas procedimentais do desenvolvimento da pesquisa. Os resultados da pesquisa são apresentados no Capítulo 6, sendo uma ontologia de referência de domínio, dois elementos visuais estatísticos, duas redes monopartidas e uma breve conclusão descritiva da interpretação desses elementos por meio de uma inspeção visual. Finalmente, o Capítulo 7, possui as referências bibliográficas.

2 ESTADO DA ARTE DA ANÁLISE DE REDES DE INFORMAÇÃO NO CONTEXTO DAS OPERAÇÕES DE CRÉDITO NO OPEN FINANCE

Foi realizada uma investigação em trabalhos mais atuais sobre a aplicação de análise de redes de informação, ou análise de redes sociais (ARS), no contexto do sistema financeiro. Os procedimentos metodológicos sobre o levantamento bibliográfico do estado da arte são abordados de forma detalhada na Seção 4.2.1.

Os textos escolhidos enfatizam técnicas de análise de redes que foram efetivamente utilizadas nas pesquisas. Posteriormente tais textos foram classificados em temas de acordo com o assunto do trabalho realizado, conforme mostra o Quadro 1. Os principais temas abordados nesses textos foram: tratamento de dados e ARS com 3 trabalhos, fraudes e crimes de lavagem de dinheiro, com um total de 4 trabalhos, identificação de rede de relações, com 4 trabalhos, e análise de risco e comportamento com 4 trabalhos.

Quadro 1 - Referências selecionadas para o estado da arte.

Referências selecionadas na base de dados	Foco do trabalho
CRUZ, Wilton Vicente Gonçalves da. Rede proprietária das maiores corporações brasileiras: centralidade, intermediação e grupos. 2021.	Tratamento de dados; ARS
CLAUDINO, Leandro; BALTORÉ, Pablo Piorno. Visualização analítica do sentimento negativo em opiniões acerca de serviços de telecomunicações emitidas por usuários do Twitter. 2015	Tratamento de dados
SILVA, Maria Fernanda Borsatto de Oliveira. Da internet para as ruas: estudo sobre o contágio social entre os usuários da rede twitter nas manifestações de junho de 2013 no	Tratamento de dados; ARS

Brasil. Trabalho de Conclusão de Curso (Graduação em Comunicação-Habilitação em Publicidade e Propaganda)-Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013	
PEREIRA, João Arami Martins; ALVES, Vinicius França; BARBOSA, Letícia Cristina Bizarro. LAVAGEM DE DINHEIRO: MÉTODOS, ANÁLISES E SUAS REPERCUSSÕES NA ECONOMIA. SEPE-Seminário de Ensino, Pesquisa e Extensão da UFFS, v. 11, 2022.	Crimes de lavagem de dinheiro
CAMPOS, Bianca Mariana Migliorini de. Uso de grafos no auxílio da identificação de fraudes em cartões de crédito. 2022	Fraudes
LIMA, RAFAEL SOUSA; SERRANO, ANDRÉ LUIZ MARQUES; CUPERTINO, CESAR MEDEIROS. Contabilidade Forense e Grafos no Combate à Lavagem de Dinheiro.	Crimes de lavagem de dinheiro
ZAINA, Roberto et al. Identificação de entidades destaque na análise de relatórios de inteligência financeira. 2020.	Crimes de lavagem de dinheiro
PARREIRAS, Pedro Henrique Schneider et al. Hegemonia e transnacionalização: uma análise	Redes de relações

exploratória da rede de relações do sistema financeiro brasileiro. 2021.	
BORGES, Edi Carlos. Um estudo da atuação das cooperativas de crédito brasileiras através de redes complexas. 2022. Tese de Doutorado. Universidade de São Paulo	Redes de relações
RIBEIRO, Henrique César Melo. PRODUÇÃO CIENTÍFICA INTERNACIONAL DO TEMA BITCOIN À LUZ DA SOCIOMETRIA. SINERGIA-Revista do Instituto de Ciências Econômicas, Administrativas e Contábeis, v. 24, n. 2, p. 61-74, 2020.	Redes de relações
TESSMANN, Mathias Schneid et al. Ensaio em precificação e volatilidade de commodities. 2022.	Redes de relações
DIAS, Eduardo de Souza. Estudo do risco sistêmico em redes interbancárias pela abordagem de sistemas complexos. 2015. Tese de Doutorado. Universidade de São Paulo.	Análise de riscos
COSTA, Ewerton da Silva. Análise de séries temporais do mercado financeiro	Análise de risco e comportamento

utilizando teoria de redes complexas. 2023	
CARVALHO, João Vinícius. E se o sistema falhar? Avaliação de risco sistêmico da prática bancassurance no Brasil	Análise de risco
SCHWAIDA, Ana Paula Lara Rabelo. Risco sistêmico nos bancos brasileiros: uma abordagem por meio das redes complexas. 2022.	Análise de risco

Fonte: autoria própria.

Sobre os trabalhos relacionados a tratamento de dados e análise de redes sociais, Cruz (2021) analisou as redes de propriedade, que permitem entender quem são os controladores dos ativos, ações e decisões estratégicas para o estudo do controle corporativo no setor empresarial brasileiro. O trabalho procura responder a questão de quem são os donos das maiores empresas brasileiras. Analisou as medidas de densidade, grau médio, modularidade e o número de componentes conectados. Analisou também as medidas de centralidade do grau de diferença ponderado, intermediação e vetor próprio. Identificou e analisou comunidades a partir de medidas de classe de modularidade. As medidas calculadas foram em nível de rede, centralidade em nível de nó, comunidades e fluxo de controle. Os resultados mostraram que a rede resultante é de baixa densidade, é esparsa e altamente modularizada. As medidas de centralidade revelam que as empresas financeiras (principalmente fundos de investimento passivos) e as empresas vinculadas ao Estado Brasileiro apresentam os maiores valores. Em relação à detecção de comunidades, foram identificados grandes grupos de multinacionais, firmas financeiras e firmas ligadas ao Estado Brasileiro e essas comunidades apresentam os maiores valores de fluxo de controle. Essa configuração converge para uma forma hierárquica de capitalismo. Desses resultados, é possível inferir a financeirização do controle das maiores corporações brasileiras e também a força do Estado em setores-chave da economia.

Claudino (2015) realizou o tratamento de grande volume de dados extraídos da rede social Twitter, de modo a revelar padrões na quantidade de reclamações dos usuários de serviços relacionados a telecomunicações, criando um ambiente de visualização em tempo real dos resultados obtidos a partir de um sistema web com uma interface acolhedora. Desenvolveu com o software Gephi gráficos relacionados a grandes volumes de dados para auxiliar a interpretação humana e a visão global da informação gerada a partir do estudo de Big Data e mineração de dados em redes sociais. A partir dos dados coletados no Twitter, foi gerado um estudo sobre a percepção por parte dos usuários dos serviços de telecomunicações das principais empresas atuantes no mercado brasileiro. Este estudo ilustrou quantitativamente e de forma comprovada, através de um tratamento de inteligência artificial, uma análise do possível sentimento de insatisfação vivenciado por alguns usuários, não sendo, no entanto, capaz de acertar na totalidade dos casos.

Ainda sobre o tema de tratamento de dados, Silva (2013) tratou da lógica social de contágio e influência no nível microssociológico para engajamento de indivíduos em ações coletivas. A abordagem teórica parte da sociologia de Gabriel Tarde, atualizando sua perspectiva através da Teoria do Ator-rede de Bruno Latour e de autores que desenvolvem estudos empíricos em torno das redes sócio-técnicas, como Mark Granovetter e Duncan Watts, entre outros, a partir da perspectiva de ARS. Para a coleta de dados, foram utilizados os softwares Node XL e Topsy para extrair as conversações entre os usuários no Twitter em torno do tema “Mídia Ninja” e para visualizar a difusão da informação em grafos de rede entre o período de 20 de julho a 20 de agosto de 2013. O objetivo desta análise foi compreender a relação entre as informações compartilhadas na rede e comportamento coletivo nas manifestações. Os resultados apontam que as redes sociais online apresentam estrutura e dinâmica centralizada, destacando a importância dos hubs no processo de influência e contágio social.

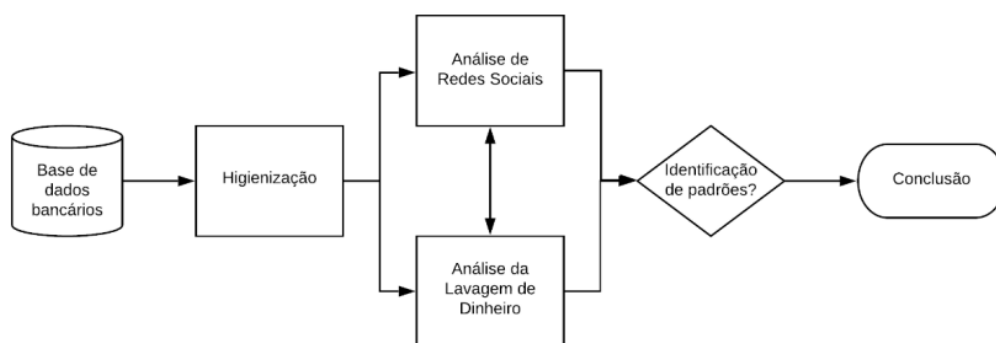
Sobre os trabalhos relacionados a crimes de lavagem de dinheiro, Pereira (2022) realizou uma ARS no relatório COAF com duas variáveis, os Sinais de Alertas e as Atividades Econômicas. Houve a aplicação do método de ARS que permitiu a análise do relatório do COAF. Os resultados obtidos foram que o crime de lavagem de dinheiro, caracterizado como a tentativa de ocultar a origem do ativo ilícito, passando pela colocação, ocultação e integração na economia, chegou-se as considerações de que são utilizados os seguintes métodos para lavar o capital:

paraísos fiscais como as *offshore*; mercado negro de obra de artes; pessoas interpostas “laranjas”; empresas de fachadas; utilização do sistema financeiro nacional; organizações não governamentais; mercado de câmbio e criptomoedas.

Campos (2022) estudou sobre as fraudes em cartão de crédito nas instituições financeiras, de forma a tentar identificar de maneira mais ágil e assertiva indicando que uma dada operação pode ser uma possível fraude. Foi realizada revisão da literatura abrangendo conceitos de Fraudes, Open Banking e Teoria de Redes. A partir deste levantamento, foi proposta uma abordagem de redução de escopo das milhares de operações de cartão, divididas por setor de atuação das empresas que possuem maior probabilidade de ocorrer dentro o universo de operações. Como resultado foram obtidos alguns grafos que, dentro do universo de operações fraudulentas, além de fraudes eventuais causadas por fraudadores de fora dos estabelecimentos, o caso de empresas com alto potencial de serem comparsas na aplicação das fraudes dentro dos estabelecimentos.

Lima (2021) realizou uma pesquisa para compreender como o contador forense pode analisar transações bancárias suspeitas de envolvimento com o crime de lavagem de dinheiro por meio da técnica de grafos. Os dados foram analisados por meio da abordagem de análise de rede social. A ferramenta computacional escolhida foi o software IBM i2 Analyst's Notebook. A análise dos dados acompanhou o fluxo demonstrado na Figura 1 que expressa o resultado obtido. A figura apresenta de forma clara o fluxo de análise de dados para detectar transações bancárias suspeitas de lavagem de dinheiro. Inicialmente, ocorre a higienização dos dados bancários, seguida pela aplicação da abordagem de ARS através da técnica de grafos. Essa metodologia visualiza as transações como nós e suas relações como arestas, facilitando a identificação de padrões complexos. A figura culmina na conclusão, onde os resultados da análise são interpretados para determinar a presença de atividades suspeitas, destacando a importância da integração de métodos avançados na prevenção de crimes financeiros. A pesquisa colaborou para o desenvolvimento desse campo da contabilidade, favorecendo a efetividade no combate aos crimes financeiros.

Figura 1 - Fluxo da Análise de Dados



Fonte: Lima (2021)

Zaina (2020) propôs o método de identificação automática de elementos relevantes em Relatórios de Inteligência Financeira. Estes relatórios contêm importantes informações para a investigação de crimes de lavagem de dinheiro, pois indicam operações suspeitas ocorridas no sistema financeiro nacional. Na pesquisa, Zaina decidiu verificar a possibilidade de usar programas para identificar automaticamente elementos relevantes e destacá-los em grafos. Estas métricas foram configuradas e processadas em programa de *business intelligence*. Após o estudo de publicações contendo tais expressões, chegou-se à conclusão de que o termo mais adequado ao contexto dessa pesquisa é o *social network analysis*, ou ARS. Foi desenvolvida uma metodologia inédita de análise de RIF na Polícia Federal combinando o processamento de dados em programa de *business intelligence* e a geração de grafos com destaque de entidades em programa de análise de vínculos. Os resultados foram mostrados em ferramentas de análise de vínculos por meio de grafos com destaque de entidades, sendo que a análise de rede social é o mapeamento e a medição de relacionamentos e fluxos entre pessoas, grupos, organizações, computadores, sites da Web e outras entidades de processamento de informações. Para o estudo foi usado o IBM i2 Analyst's Notebook, o mesmo usado na pesquisa de Lima (2021). Ainda nessa pesquisa de Zaina (2020), foi usado o programa Qlikview Personal adotada por vários analistas da Polícia Federal do Brasil para análises e cruzamentos de dados estruturados, como extratos bancários e extratos telefônicos, porém ainda não utilizada para análise de RIF gerado em arquivo do tipo PDF, que é objeto de estudo do trabalho de Zaina. Isto facilitou sobremaneira a análise do grande volume de dados e diminuiu a complexidade das informações de Relatórios de Inteligência Financeira testados.

Já no campo da análise de redes de relações, Parreiras (2021) propôs uma ARS de forma exploratória para compreender as redes – nacionais e transnacionais - entre os diretores e entre as instituições que atuaram no sistema financeiro brasileiro entre os anos de 2010 e 2019. Em síntese, utilizou-se a ARS de maneira exploratória - ou seja, sem a mobilização de uma modelagem estrutural - para mapear e analisar as redes de relações do SFB. Após as análises foi possível identificar uma rede coesa de relações que se organiza, principalmente em torno de associações de classe – como a FEBRABAN e a CNF.

Borges (2022) analisou as características das cooperativas de crédito no Brasil, a fim de verificar relações existentes entre cooperativas e municípios. O arcabouço de análise de redes foi a base. Foi utilizado o software Gephi na confecção do grafo bipartido. Os resultados obtidos foram que na expansão de locais de atendimento, as cooperativas dão preferência em instalar-se em municípios já atendidos pelo cooperativismo, elevando o grau na rede complexa. Foi evidenciado o número de relações entre os diferentes segmentos de cooperativas.

Ribeiro (2020), investigou o perfil e as características de colaboração da produção científica internacional do tema bitcoin durante o período de 2008 a 2017, à luz da ARS. Metodologicamente, utilizaram-se as técnicas de ARS mensuradas com apoio do software UCINET e visualizados pelo software NetDraw. Como principal resultado, nas redes *one-mode*, ou monopartidas, foi constatada uma baixa densidade. No que tange aos atores influentes, realçam-se Glenn Parry e Beth Kewell, enfocando os pesquisadores; University of the West of England, Holy Spirit University of Haslik e The University of Surrey no que tange às instituições; e os Estados Unidos e o Reino Unido em relação aos países. No contexto das redes *two-mode*, ou bipartidas, averiguou-se que as publicações sobre Bitcoin começaram a ascender a partir de 2012;

Tessman (2022) buscou aferir as transmissões de volatilidade que ocorrem inter e intra classes de ativos entre commodities metálicas e de energia, utilizando métricas da mecânica estatística de redes complexas. Os resultados demonstram que para ambas commodities a hipótese da lei do preço único não é satisfeita. Por outro lado, encontrou uma forte evidência de relação dos preços local e internacional da soja, inclusive com uma relação de longo prazo que é importante para tomar decisões futuras. Acredita que esse resultado seja devido à sua maior integração entre as cadeias globais. Segundo o autor, estes achados são úteis para a literatura

que investiga os derivativos de commodities, para investidores e gestores de portfólios, assim como para *policy makers* que consideram esses grãos em suas decisões.

Com o tema de análise de riscos, Dias (2015) realizou o estudo do risco sistêmico em redes interbancárias pela abordagem de sistemas complexos. Testes foram elaborados com cenários hipotéticos e chegou-se à conclusão de que no aumento do índice de cluster da rede interbancária, os bancos conseguiram maior crescimento, porém sem os efeitos indesejáveis causados pelo aumento do risco.

Costa (2023) fez a análise de séries temporais do mercado financeiro utilizando teoria de redes complexas e relatou que embora a área de Sistemas Econômicos seja tradicionalmente estudada por economistas e matemáticos, é crescente o número de estudos nessa área realizada por físicos. Consequentemente, ferramentas de Sistemas Complexos e Ciência de Redes tem se provado úteis na modelagem de mercados financeiros. Em particular, conceitos como distância de correlação e *minimal spanning tree* (MST) tem sido largamente utilizados na literatura para estudar a dinâmica de mercados financeiros, através da similaridade de medidas, análise de risco e comportamento em períodos de crise financeira. Neste trabalho, foram utilizados dados do mercado financeiro composto por índices de diversos países e regiões, moedas, commodities, títulos públicos, entre outros. Foi realizada uma análise de como diferentes grupos financeiros se relacionam levando em conta sua correlação.

Carvalho (2022) avaliou o potencial e a suscetibilidade à ruína de seguradoras que praticam o modelo de *bancassurance*² ao risco de contágio. O mercado ressegurador local foi mapeado usando a abordagem de redes complexas, na qual os agentes são vértices, conectados por meio de operações de cosseguro, resseguro e retrocessão, usando as bases de dados oficiais da SUSEP, entre 2015 a 2020. O papel de cada resseguradora foi avaliado por meio de uma análise topológica baseada em métricas de centralidade. Como resultados, pode-se afirmar que a prática de *bancassurance* não apresenta potencial ou vulnerabilidade

² Expressão em francês que se refere à oferta combinada de produtos bancários e de seguros, realizada por bancos e caixas econômicas, ou seja, trata-se da distribuição de seguros através da rede de agências de uma instituição financeira. *bancassurance* (bank insurance). Disponível em: <<https://www.fundacionmapfre.com.br/publicacoes/diccionario-mapfre-seguros/bancassurance/>>. Acesso em: 20 jun. 2023.

sistêmica. Carvalho conclui afirmando que tal evidência tem repercussão internacional, uma vez que pelo tamanho e pelo modelo baseado em conglomerados financeiros, as medidas de risco de contágio do bancassurance no Brasil pode ser visto como internacionalmente conservadora.

O trabalho de Schwaida (2022) sobre análise de riscos baseou-se no objetivo de demonstrar o risco sistêmico presente no sistema financeiro brasileiro utilizando a metodologia de redes complexas. A autora enfatiza a importância da utilização da teoria de redes complexas visto que sua aplicação no estudo do risco sistêmico permite identificar quais integrantes são mais centrais para a rede, pois as interações aferidas por meio de medidas de centralidade apontam para a importância de um integrante na rede em que faz parte. Os resultados da pesquisa de Schwaida foram obtidos por meio das variáveis contábeis Ativo, Passivo, Patrimônio Líquido e Lucro Líquido. Além disso, foram selecionadas as maiores instituições financeiras, baseado em informações públicas da base de dados do Banco Central do Brasil. Foram também apuradas medidas de centralidade da rede de cada variável no intuito de identificar quais instituições são mais influentes no sistema e que podem contribuir para um possível risco sistêmico.

Os trabalhos analisados apresentam correlações em sua abordagem de tratamento de dados e análise de redes. Cruz (2021) explora a análise de redes para compreender a estrutura de propriedade no setor empresarial brasileiro, enquanto Claudino (2015) utiliza ARS para estudar aplicações de usuários de telecomunicações no Twitter. Silva (2013) também se vale da ARS, examinando a influência nas manifestações online. Pereira (2022) e Campos (2022) concentram-se em análises de redes para identificar padrões e métodos relacionados a crimes financeiros, como lavagem de dinheiro e fraudes em cartões de crédito.

Lima (2021) e Zaina (2020) abordam a análise de transações bancárias suspeitas por meio de técnicas de grafos, utilizando como IBM i2 Analyst's Notebook. Ambos os trabalhos apresentados para o entendimento das características de lavagem de dinheiro. A pesquisa de Parreiras (2021) destaca a análise de redes no sistema financeiro, fornecendo insights exploratórios sobre relações entre diretores e instituições. Além disso, os estudos de Borges (2022), Ribeiro (2020), Tessman (2022) e Tessmann et al. (2022) também contribuem significativamente para o campo, cada um trazendo perspectivas distintas e enriquecedoras sobre temas relacionados à análise de redes de relações no

contexto financeiro. Esses estudos em conjunto fornecem uma compreensão mais ampla e aprofundada das dinâmicas das redes de relações no sistema financeiro contemporâneo. Além disso, a proposta da pesquisa compartilha elementos com os estudos de Dias (2015), Costa (2023), Carvalho (2022) e Schwaida (2022) ao abordar aspectos de análise de risco no sistema financeiro. Eles utilizam a teoria de redes complexas para avaliar interconexões entre entidades financeiras, destacando a importância das instituições na rede. Esses estudos enriquecem a compreensão sobre as dinâmicas e interconexões no sistema financeiro, fornecendo insights valiosos para a gestão de riscos e a tomada de decisões.

Em resumo, os trabalhos realizados fornecem insights valiosos para a proposta de pesquisa, compartilhando a abordagem de ARS, o tratamento de grandes volumes de dados e a compreensão de padrões e riscos no sistema financeiro. Essas correlações fortalecem a fundamentação teórica e metodológica da pesquisa proposta.

Embora os artigos analisados no presente capítulo tragam informações de importantes aspectos do sistema financeiro, a investigação realizada nessas pesquisas evidenciou, para alguns trabalhos, a necessidade de um aprofundamento relativo ao tratamento dos dados e as metodologias e ferramentas utilizadas na análise de redes de informação.

Além disso, a revisão da literatura revelou que há uma escassez de estudos específicos que abordaram a análise de redes de informação com o objetivo de fornecer informações diretamente relacionadas aos benefícios sociais.

3 REFERENCIAL TEÓRICO

O referencial teórico apresenta elementos fundamentais para o entendimento do desenvolvimento da pesquisa. Aborda a Organização do Conhecimento e a Ciência de Dados de forma contextualizada na Ciência da Informação; a Ciência das Redes nas Ciências Sociais com foco maior sobre a Análise de Redes de Informação e as Redes Bipartidas com suas Projeções Bipartidas; e o Open Finance, seu funcionamento, tendências inter-relação com a Ciência da Informação.

3.1 CIÊNCIA DA INFORMAÇÃO E ORGANIZAÇÃO DO CONHECIMENTO

A Ciência da Informação (CI) embora com pouco tempo de existência, estabeleceu-se no campo científico tendo seus primórdios alinhados ao fim da segunda guerra mundial, em uma competição tecnológica e científica da guerra fria (Araújo, 2020). A CI, com uma natureza multidisciplinar, identifica-se com várias outras áreas como as neurociências, ciências sociais em geral, as comunicações, a biblioteconomia, informática e alguns outros campos (Robredo, 2003). Ela é a disciplina que investiga as propriedades, o comportamento, as forças e os fluxos que regem a informação. Se preocupa com os meios de processamento da informação para um máximo de acessibilidade e uso. Esses meios de processamento incluem produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação (Capurro, 2017).

Partindo do conceito de que informação é um registro em suporte material chamado documento, este se torna o objeto da organização (Le coadic, 2004). Urdaneta (1992) apresenta uma definição mais detalhada a respeito da informação e afirma que ela pode ser classificada em quatro níveis que são representados por meio da pirâmide informacional, a saber: dado, informação, conhecimento e inteligência. A informação traz a qualidade de um significado (Matta, 2010). Segundo Liew (2013), quando os dados são processados e analisados tornam-se informação, que são transformadas em conhecimento na mente das pessoas. Liew explicita ainda que o conhecimento, quando passa por um processo mental de reestruturação, transforma-se em inteligência, que é a capacidade de se retirar do conhecimento uma atitude ou intervenção que seja vantajosa perante a realidade, ou seja, é o conhecimento aplicado. E finalmente a tão almejada sabedoria,

considerada por Liew , como bom senso ou verdade universal. O indivíduo sábio usa o conhecimento adquirido da melhor forma de acordo com as necessidades de cada momento, e tem grande conexão com a experiência provocada pelo tempo dedicado à aplicação do conhecimento.

É interessante considerar que não existe um consenso na literatura sobre essa hierarquia piramidal (dado, informação e conhecimento). Por exemplo, Fricke (2019) considera que há um processo interativo entre dado e conhecimento, assumindo que ambos influenciam-se mutuamente.

Sendo assim, a informação precisa ser descrita e classificada para a sua organização e o produto desse processo descritivo é denominado representação (Araújo, 2018). A representação é utilizada então para conduzir à uma recuperação eficaz das informações previamente organizadas (Alvarenga, 2003).

A Organização do Conhecimento (OC) está preocupada em descrever, classificar, e indexar informações e preocupa-se com a natureza e qualidade de tais processos (Hjorland, 2016). Ela diz respeito à representação de uma estrutura, um arcabouço de conceitos, sendo uma réplica distorcida de um conhecimento acumulado (Barité, 2021).

Existem diferentes abordagens históricas e teóricas sobre a OC. Em um sentido mais restrito ela diz respeito a atividades como descrição de documentos, indexação e classificação. No sentido mais amplo vai se ater a como o conhecimento é organizado socialmente e ainda se preocupa em como a realidade é organizada. Na CI, a OC é a disciplina central no sentido mais restrito da palavra. No entanto, não há como desenvolver o sentido restrito sem considerar o sentido amplo, pois não há como estudar um “universo de conhecimento” separado de outras realidades ou ciências (Hjorland, 2016). O desenvolvimento do conhecimento deve, portanto, preservar a diversidade e assegurar a integração das diversas partes de um todo (Mazzocchi, 2018).

3.1.1 Socialização da Informação

A socialização da informação desempenha um papel fundamental em diversas áreas do conhecimento, uma vez que qualquer prática depende do conhecimento como seu ponto de partida (Christovão; Braga, 1997). Este fenômeno revela-se crucial em práticas organizacionais, como tomada de decisão e inovação, desempenhando um

papel essencial no contexto contemporâneo (Valentim, 2002). Contudo, mais do que apenas um conhecimento prático, a socialização da informação busca estabelecer um equilíbrio na Sociedade da Informação.

Nesse contexto, a Sociedade da Informação emerge como um agente transformador, reconfigurando as estruturas de organização e produção em escala global e redesenhando o posicionamento dos países na sociedade internacional e no sistema econômico mundial (Takahashi, 2000). Dentro desse cenário dinâmico, as ações de socialização da informação e do conhecimento são percebidas como meios cruciais para capacitar os indivíduos, fornecendo-lhes a capacidade necessária para desempenhar um papel cidadão de maneira crítica e autônoma (Saeger, et al., 2018).

É relevante salientar que, no contexto deste trabalho, a socialização da informação visa fornecer informações não triviais ao cidadão comum, com o propósito de capacitá-lo para uma participação mais efetiva em uma sociedade mais justa. O processo de socialização da informação, portanto, vai além da mera transmissão de conhecimento prático, buscando promover uma compreensão crítica e autônoma que contribua para a construção de uma sociedade mais equitativa.

3.1.2 Sistemas de Organização de Conhecimento

É importante esclarecer as diferenças entre a Organização da Informação e do Conhecimento. De maneira breve, a organização da Informação está para representação de objetos físicos e seu produto final é um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. Já a OC está relacionada ao mundo das idéias, das cognições, das abstrações de uma realidade. Seu produto é um modelo de abstração do mundo real, construído para determinada finalidade (Alvares, 2023).

A OC em seu sentido mais restrito faz uso dos sistemas de organização do conhecimento (SOC), ou *knowledge organization system* (KOS). Eles são organizados de acordo com a complexidade de suas estruturas e principais funções (Hjorland, 2008). Segundo Zeng (2008), os SOCs podem ser explicados de acordo com quatro grandes grupos, não mutuamente exclusivos: listas de termos, modelos

semelhantes a metadados, classificação e categorização, e modelos de relacionamento. As quatro categorias possuem peculiaridades (Zeng, 2008):

- Listas de termos são conjuntos limitados de termos em alguma ordem sequencial. Como exemplo, os dicionários são listas alfabéticas de termos e suas definições fornecem sentidos variantes para cada termo, quando aplicável.
- Os modelos de metadados podem ser entendidos como diretórios, que são listas de nomes e suas informações de contato associadas. Como exemplo do campo da classificação e categorização, as taxonomias dividem itens em grupos ordenados ou categorias com base em determinadas características.
- Modelos para classificação e categorização, se enquadram, por exemplo as taxonomias.
- Tesouros, redes semânticas e ontologias são classificados como modelos de relacionamento. Os tesouros são conjuntos de termos que representam conceitos e as relações hierárquicas entre eles. As redes semânticas são conjuntos de termos que representam conceitos, modelados como nós de uma rede de tipos de relacionamento variáveis. Ontologias são modelos de conceitos específicos que representam os relacionamentos complexos entre objetos, incluindo as regras que faltam nas redes semânticas. O SOC ontologia será tratado em mais detalhes em subseção própria por ser um dos elementos que fundamenta a presente pesquisa.

Algumas funções consideradas comuns entre os sistemas de organização do conhecimento são: eliminar ambiguidades, controlar sinônimos, estabelecer relacionamentos (hierárquicos e associativos).

3.1.3 Ontologias

A definição de ontologia ainda tem sido muito debatida, porém uma amplamente aceita é que ela é uma especificação explícita e formal para uma conceituação compartilhada (Gruber, 1993). Isso significa que ela é capaz de abstrair e descrever ambientes de forma inteligível e ser aceita por diversos atores, sejam eles homem ou máquina, nos sistemas de informação. Ela abrange a estrutura classificatória de taxonomia e tesouros e apresenta a característica única

de apresentar propriedades de cada classe dentro de sua estrutura classificatória (Zeng, 2008). Amplamente reconhecida como instrumento conceitual bastante útil, a ontologia se destaca principalmente na área da Web semântica.

Um dos usos mais comuns da ontologia tem sido o compartilhamento de informações entre pessoas ou agentes de software, com os objetivos de reutilização de informação e análise de conhecimento de domínio (Noy, 2001). O autor enfatiza ainda o uso da ontologia como base de conhecimento, explicitando a linha tênue que há entre esses dois artefatos. Uma ontologia estabelece um vocabulário comum que facilita a comunicação entre pesquisadores ao compartilhar informações dentro de um determinado domínio. Segundo Noy (2001), desenvolver uma ontologia é motivado por diversas razões, tais como:

- Facilitar o entendimento comum da estrutura da informação entre indivíduos ou agentes de software.
- Possibilitar a reutilização eficiente do conhecimento do domínio.
- Explicitar as suposições inerentes ao domínio em questão.
- Separar o conhecimento do domínio do conhecimento operacional.
- Analisar de maneira mais efetiva o conhecimento dentro do domínio em foco.

Um ponto importante é a diferença de conceituação da ontologia entre as áreas dentro da própria computação. Na área de modelagem conceitual ela é definida como um sistema de categorias formais independente de domínio e filosoficamente bem fundamentado, que pode ser usado para enunciar modelos da realidade específicos de domínio (Guizzardi; Falbo; Guizzardi, 2008). Segundo os autores, em outras áreas é considerada um artefato concreto de engenharia projetado para um propósito específico, sem dar muita atenção para questões de fundamentação. Mas fato é que uma ontologia é fundamental para a recuperação, compartilhamento e representação do conhecimento no contexto das organizações (Oliveira; Tenório; Bortolozzi, 2022).

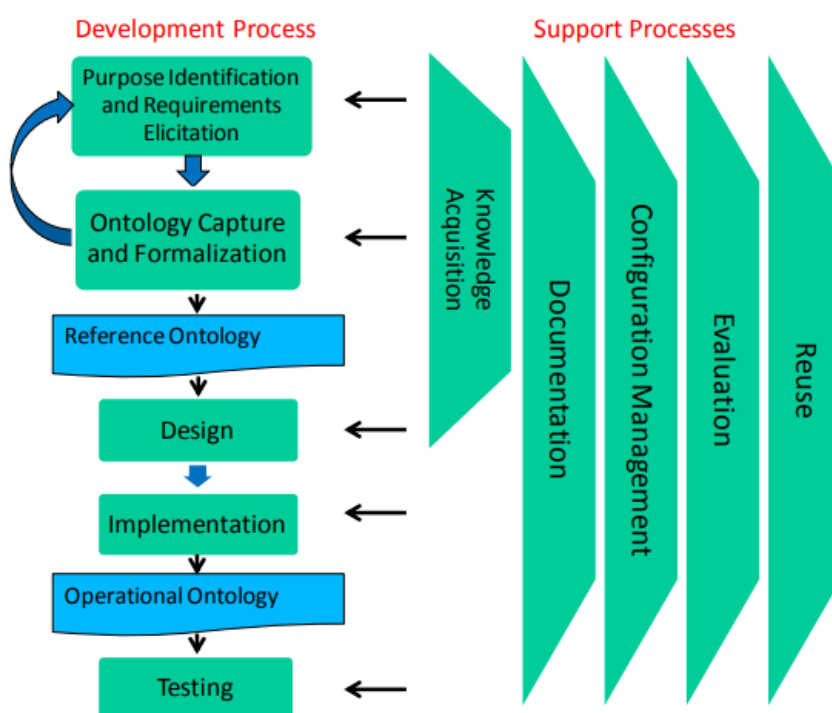
Na Ciência da Informação ela exerce papel preponderante da área de representação da informação, visto que é capaz de representar não só o objeto documento, mas todo um domínio (Marcondes, 2008). Segundo o autor, essa definição assemelha-se muito ao objetivo da Ciência da Informação em termos de conceito e objetivo.

As ontologias vêm sendo desenvolvidas desde a época de Aristóteles sendo propostas mais recentemente nas Ontologias Aplicadas sob o nome de ontologias de fundamentação. Em especial merece destaque a UFO (*Unified Foundational Ontology*). As ontologias de fundamentação têm papel relevante na construção de ontologias de domínio. Ela tem sido aplicada de forma satisfatória para avaliar, reprojeter e integrar modelos de linguagem de modelagem conceitual (Guizzardi, Falbo, Guizzardi, 2008). A OntoUML é uma linguagem de modelagem orientada a ontologia baseada na UFO. Ela é ontologicamente bem fundamentada e tem sido utilizada para fazer modelos de simulação conceitual (Guizzardi, Wagner, 2012). Desse modo, as ontologias têm sido propostas como um método para representação de uma realidade (Almeida, 2014).

A construção de ontologias requer métodos apropriados, como qualquer processo de desenvolvimento de sistemas de informação. Reconhece-se a metodologia SÁBIO (*Systematic Approach for Building Ontologies*, ou Abordagem Sistemática para a Construção de Ontologias) como um modelo de referência para construção de ontologias. A metodologia SABIO foi concebida para apoiar o desenvolvimento de ontologias de referência de domínio. Por ontologia de referência de domínio entende-se uma ontologia construída com o objetivo de fazer a melhor descrição possível do domínio. É uma especificação independente de solução (modelo conceitual) para fins de comunicação, aprendizado e solução de problemas.

O processo de desenvolvimento da metodologia SABIO compreende cinco principais fases, conforme mostra a Figura 2: (i) Identificação do propósito e levantamento de requisitos; (ii) captura e formalização da ontologia; (iii) projeto; (iv) implementação; e (v) teste.

Figura 2 - Processo da metodologia SABIO



Fonte: Falbo (2014)

Entende-se portanto que ontologias são reconhecidamente importantes como instrumento de modelagem de dados (Guizzardi, 2008). Fato esse que corrobora para que também possa ser utilizada na Ciência de Dados como artefato de modelagem.

3.1.4 Ferramenta para modelagem de dados

Existem ferramentas que auxiliam na criação de modelos conceituais e ontologias, tal como o software Visual Paradigm³. O Visual Paradigm é considerado de grande usabilidade, possui versão de uso gratuita com vários recursos gráficos para a construção de diagramas diversos de modelagem.

Além disso, o Visual Paradigm, usado na modelagem conceitual juntamente com o plugin gUFO, permite a diagramação utilizando-se a linguagem de modelagem orientada a ontologia, OntoUML, baseada na ontologia de fundamentação UFO.

³ Visual Paradigm, disponível em: <https://www.visual-paradigm.com/>.

3.2 CIÊNCIA DE DADOS

Sabe-se que a capacidade humana em gerar informação é superior à capacidade de armazenar as informações de forma organizada (Bush, 1945), porém o fato de gerar e armazenar informações em grande quantidade não torna os indivíduos mais sábios (Gleick, 2011).

A Ciência de Dados é uma área interdisciplinar com um campo de atuação bem amplo. Pode responder a perguntas que ainda não foram feitas e também trabalhar com base de dados criadas e alimentadas para atender a propósitos já preestabelecidos, ou que serviram para outra finalidade diferente da mineração de dados (Cristovão, 2023). Segundo Aalst (2016), ela visa transformar dados em valor real.

Atualmente, os termos *Data Science* (Ciência de Dados) e *big data* são utilizados como se fossem conceitos novos, mas se referem a fases referentes à preparação, apresentação e descrição dos dados de trabalhos estatísticos realizados há mais de dois séculos (Morettin, 2020). *Big data* é um termo utilizado para designar essa grande quantidade de dados disponíveis. Rautenberg (2020), vai referir-se ao Big Data como conjunto de dados cujo tamanho é fator impeditivo de tratamento por ferramentas tradicionais, ou seja, requer formas inovadoras de tratamento de grande volume de dados, fomentando assim uma nova área de atuação: a Ciência de Dados.

A Ciência de Dados requer diversos conhecimentos e habilidades de diferentes disciplinas, entre os quais está a CI (Virkus; Garoufallou, 2021). Segundo Rautenberg (2020), a CI tem papel fundamental na consolidação de ecossistemas de Big Data, que diz respeito à utilização salutar de dados, informação e conhecimento. Principalmente com relação às competências de organização e representação de dados e metadados e da Curadoria Digital dos recursos mantidos nesses ecossistemas.

3.2.1 Dados, *Data Driven* e *Data Literacy*

Dados são, segundo Davenport (1998), “[...] matéria prima essencial para a criação da informação [...]”, ou ainda, “[...] em informática, dado é a representação codificada de uma informação em uma forma que permita submetê-la a processamento eletrônico” (Le coadic, 2004).

Conforme Hjørland (2018), no âmbito da Ciência da Informação, é essencial considerar diferentes domínios quando lidamos com questões relacionadas a dados na era atual do big data. Nesse contexto, a demanda é por especialistas em dados específicos desses domínios, em vez de apenas cientistas generalistas em dados. A epistemologia social, nesse sentido, reconhece que um sistema de organização do conhecimento não pode ser neutro, tendendo sempre a destacar certas perspectivas em detrimento de outras. Hjørland (2018) destaca que existem diversos tipos de representações de conhecimento na epistemologia social, vinculadas a interesses distintos relacionados a diferentes domínios e grupos sociais.

Ainda conforme Hjørland (2018), os dados nunca existem de forma independente de contexto, sendo sempre produzidos a partir de perspectivas específicas com critérios próprios de relevância. A coleta e o processamento de dados devem ocorrer com a compreensão dos interesses e objetivos que orientam sua captura, seleção e processamento, especialmente quando utilizados em novos contextos que podem envolver objetivos conflitantes ou conhecimentos implícitos. O tratamento adequado de dados em big data, nos contextos de domínios específicos, requer o uso de metadados ou padrões de metadados cuidadosamente selecionados e/ou desenvolvidos para apoiar os objetivos particulares desse domínio específico.

No universo dos dados, o chamado *big data*, encontram-se as organizações que tentam atualizar-se e tirar proveito da quantidade de informação produzida a cada minuto pelos usuários e potenciais clientes. Essa corrida das organizações em busca de ter decisões baseadas em dados e não apenas suposições é denominada *data driven*, ou orientado a dados. Essencialmente trata-se de estabelecer a cultura de toda uma organização para que ela atinja níveis de maturidade de tratamento de dados (Anderson, 2015). Dentro do objetivo *data driven* encontra-se *data literacy*. A Gartner, empresa mundialmente reconhecida pelo processo de implantação da cultura de dados em empresas de grande porte (Wikipedia, 2023), define de forma interessante o termo *data literacy* como “a habilidade de ler, escrever e comunicar dados” ou informalmente, em inglês, ‘...do you speak data?’. Os termos mencionados tratam de uma alusão à existência de uma nova forma de comunicação: a linguagem dos dados.

No mundo corporativo o termo também é conhecido por alfabetização de dados, dando a entender a importância dos indivíduos de serem “letrados” em

dados. Na prática, é a capacidade de selecionar tudo aquilo que é útil ao negócio, em meio a uma imensa variedade de dados, tanto estruturados (tabelas e planilhas), como não-estruturados (Santos; Santana, 2013). *Data literacy* atua em conjunto com o *Big Data*, extraindo o máximo possível de *insights* e conhecimentos de valor para os negócios. Portanto, esse letramento em dados é interpretado como uma habilidade crucial para todos os membros de uma organização e inclusive da sociedade como um todo (Becker, Diniz, 2019).

Tratando-se de termo extremamente novo no contexto social, *data literacy* é definido como conhecimento subjacente ao *information literacy* que já é um tema tratado de forma mais pragmática pela Ciência da Informação (Belluzzo, 2004). Muito embora o conhecimento no tratamento aos dados seja inerente ao processo de formação de conhecimento em qualquer área, no sentido mais literal da palavra através do pensamento crítico (Wolff, 2016), o processo intitulado *data literacy* diz respeito não só a habilidades de apropriação da informação (Rasteli, Cavalcante, 2014), mas de aprendizados mais técnicos relacionados à manipulação de ferramentas usadas em análise de dados (Schield, 2004). Nesse sentido, o termo é conhecido internacionalmente como *data literacy* e vem sendo mencionado por alguns autores da computação como literacia em dados (Silva; Brito, 2022), inclusive vem sendo discutida pela área de Interação humano-dados (IHD) que se propõe a incluir os humanos no centro do fluxo de dados promovendo mecanismos para que os cidadãos interajam com os sistemas de dados. O termo também é encontrado na literatura como alfabetização em dados (Yamauti; Tsunoda, 2019) e ainda competência em dados (Reis; Balbinotti, 2022).

3.2.2 Mineração de Dados e Descoberta de Conhecimento

A mineração de dados é a análise de um conjunto de dados, normalmente grande, para encontrar relações não esperadas sintetizando-os de forma a serem compreensíveis para o seu proprietário (Hand et al., 2001). A maior parte dos métodos de mineração de dados são baseados em técnicas experimentadas e testadas de aprendizado de máquina (machine learning), reconhecimento de padrões e estatísticas tais como classificação, armazenamento em cluster, regressão entre outras (Fayyad et al., 1996).

Segundo Carvalho (2005) a mineração de dados possui as seguintes fases:

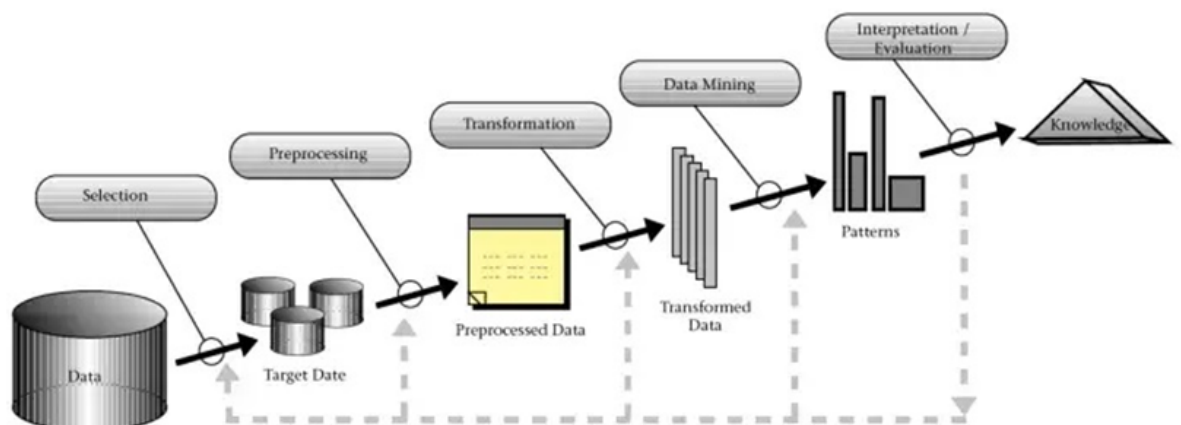
- Identificação de um problema ou definição de um objetivo;
- Descoberta de novas relações por técnica de data mining;
- Análise humana das novas relações descobertas;
- Uso racional das novas relações descobertas;
- Avaliação dos resultados.

A descoberta de conhecimento em base de dados, conhecida por KDD (*knowledge discovery in databases*) é um campo de pesquisa da inteligência artificial que envolve a mineração de dados (Hand et al., 2001).

Fayyad et al. (1996) resume o processo de descoberta de conhecimento em cinco fases, que são representadas na Figura 3:

- Seleção: os dados são selecionados de uma ou mais fontes de dados;
- Pré-processamento: preparação, limpeza de dados, e ajustes diversos etc.;
- Transformação: os dados são transformados de acordo com a fase posterior. Para uma base que seja aceita pelos algoritmos de data mining;
- Mineração de Dados: Mineração de dados de Máquinas ou mesmo da Estatística. É aqui onde os padrões escondidos e valiosos são encontrados;
- Interpretação/Avaliação: Nessa etapa, os padrões escondidos e os modelos inteligentes são avaliados e interpretados, normalmente em parceria com os especialistas do domínio e com o cliente.

Figura 3: Processo KDD

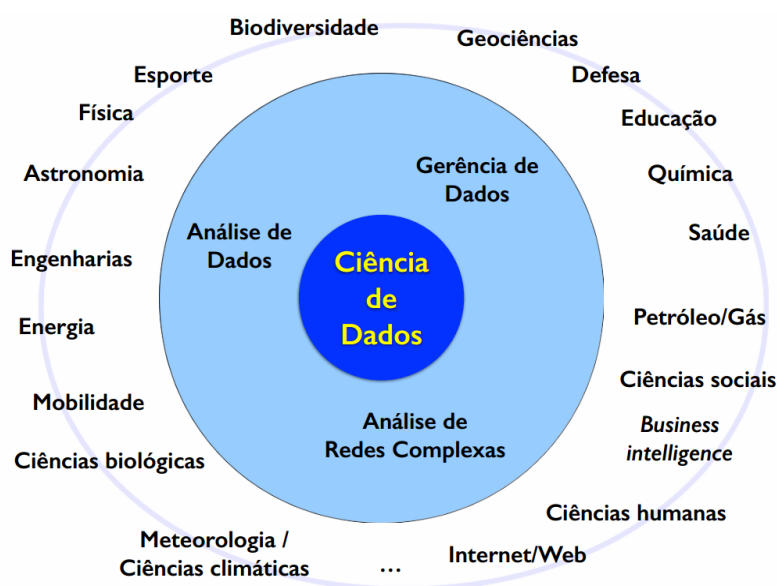


Fonte: Fayyad et al. (1996)

Segundo Porto e Ziviani (2014) os desafios de pesquisa em Ciência de Dados partem de três grandes áreas, Figura 4:

- Análise de Dados,
- Gerência de Dados,
- Análise de Redes Complexas.

Figura 4 - Desafios de pesquisa para a Ciência de Dados



Fonte: Porto e Ziviani (2014)

Porto e Ziviani, na Figura 4, destacam que a análise de redes complexas é um desafio em Ciência de Dados, área central para a presente pesquisa, abordada na Subseção 3.3.2. Em consonância, a mineração de dados faz parte das etapas da descoberta do conhecimento em análise de redes complexas. Desde a seleção dos dados até a interpretação e avaliação dos padrões, a mineração de dados é essencial para extrair conhecimento valioso, estando intrinsecamente vinculada às diversas etapas do processo.

3.2.3 Ferramenta para manipulação de dados

Softwares de manipulação de dados são recursos importantes para o processo de descoberta de conhecimento em base de dados, KDD, conforme abordado na Seção 3.2.2. Esses softwares têm a função de auxiliar no processo de

ETL (Extract Transform Load), reconhecidamente como o processo que é responsável por 80% do tempo gasto em um projeto de KDD (Mason; Patil, 2015).

Para que dados sejam preparados para o processo de transformação, existem ainda softwares específicos para a limpeza e preparação, como o Open Refine⁴, UW Interactive Data Lab⁵.

Um dos softwares ETL atualmente conhecidos é o Open Refine, que é capaz de operar em grandes intervalos diferenciando-se das planilhas eletrônicas comuns, sendo classificado como ITD (Interactive Data Transformation). O Open Refine possui a capacidade de trazer uma visão mais analítica dos dados ao usuário. É uma poderosa ferramenta gratuita de código aberto para trabalhar com dados confusos, conseguindo limpar, transformar de um formato em outro e estender com serviços da Web e dados externos.

3.2.4 Visualização de Informação

A visualização de dados também é considerada visualização de informação ou visualização científica (Sadiku, 2016). A visualização de informação visa aumentar a compreensão do usuário sobre algum conhecimento, trazendo assim a inclusão informacional dos usuários (Dias, 2007). No *big data*, após a organização e representação dos dados, desenvolvem-se ações de recuperação e visualização de informação. A atividade de visualização facilita a interpretação pois privilegia elementos de representação gráfica da informação (Rautenberg; Carmo, 2020).

Correspondendo a uma das áreas de estudo da Ciência de Dados, a visualização de informação pode parecer em um primeiro momento apenas uma modificação da representação gráfica, no entanto, a visualização permite interagir com a informação exposta. Trata-se de uma nova forma de representação da informação possibilitada pelo avanço tecnológico (Silva, 2019). Possibilita ainda aos usuários a exploração de grande quantidade de dados e obtenção de novos insights.

A visualização de dados não é um processo inato, mas sim um processo que exige aprendizado e prática, e carece de um processo de letramento semelhante ao letramento textual tanto para a leitura quanto para a construção de visualizações de dados. Há a necessidade de seguir etapas necessariamente para extrair insights dos

⁴ OpenRefine, disponível em: <https://openrefine.org/>

⁵ UW Interactive Data Lab, disponível em: http://datalab.epina.at/en_download.html

dados. Muito embora não haja um consenso nessas metodologias, existem algumas metodologias que podem ser utilizadas tanto na construção quanto no consumo dos resultados das análises. Especificamente para a análise de conjuntos de dados antes de gerar a visualização, podemos nos concentrar em alguns tipos de análises que são comumente utilizados nesse tipo de pré-processamento: A análise estatística é uma ferramenta poderosa para ordenar e classificar dados, proporcionando uma compreensão mais profunda da distribuição e variabilidade dos mesmos. A análise temporal responde à pergunta crucial sobre "quando", permitindo a identificação de tendências ao longo do tempo e padrões temporais essenciais. A análise geoespacial, por sua vez, oferece insights valiosos sobre a distribuição espacial dos dados, sendo vital para identificar padrões e variações em diferentes localidades. Por fim, a análise relacional, conforme destacado por Börner et al. (2019), foca nas conexões entre diferentes entidades, fornecendo uma visão detalhada dos relacionamentos presentes nos dados. Esses tipos de análises combinados são fundamentais para preparar o terreno antes da visualização, garantindo que as representações visuais sejam robustas, significativas e capazes de contar histórias envolventes a partir dos dados analisados. Para análise de redes em especial, o autor enfatiza que há uma dificuldade maior na interpretação dessas visualizações e dá um direcionamento do tipo de evidência criada com os nós: a partir das cores, do tamanho e do posicionamento.

A visualização de informação tem colaborado principalmente com as áreas de mineração de dados, interação humano-computador e design de interfaces, e é precursora da visualização do conhecimento (Burkhard, 2005). Diferentemente da visualização de informação que concentra-se em tarefas exploratórias de fatos e números, a visualização do conhecimento está voltada para a transferência de conhecimentos entre os indivíduos por meio de representações visuais, que buscam responder questões como “porque”, “quem” e “como” através de suposições, instruções, e insights.

A visualização adquiriu importante valor na área de análise exploratória de redes e torna-se cada vez mais relevante em áreas diversas. Como exemplo disso, é possível citar a IHD (interação humano-dados), área análoga à IHC (Interação Humano-Computador) que vem abordando questões sobre acesso da população leiga a dados e visualizações (Brito, 2022). Não sendo de exclusividade da Ciência

da Computação, mas é também de interesse da CI e evidencia sua importância na recuperação da informação (Carvalho, 2003).

3.2.5 Ferramenta para visualização de informação

Alguns dos softwares mais utilizados na atualidade para visualização de informação e de dados são o Looker Studio⁶ e o Power BI⁷. Ambas as ferramentas são disponibilizadas gratuitamente até determinado nível de interação.

A versão Power BI Desktop é gratuita e é capaz de fornecer desde visualizações interativas e recursos de BI (*business intelligence*). Com uma interface simples para a criação de relatórios e *dashboards*, colaborando assim com a área da visualização de informação.

Contudo, diferentemente do Power BI Desktop, o Looker Studio permite publicar os resultados dos *dashboards* construídos de forma gratuita na Web.

3.3 CIÊNCIA DAS REDES

As Redes se constituem de campo interdisciplinar, com aplicabilidades múltiplas e distintas. Suas aplicações vão desde banco de dados, a internet, campos sociais, e biologia (Newman, 2018). Muitas redes emergem e evoluem, baseadas em um conjunto fundamental de leis e mecanismos, e estas são a base de uma nova ciência, denominada de Ciência das Redes (Barabási, 2013). Segundo Cavalcante (2009), a Ciência das Redes, encontra-se ainda em seus primórdios como ciência, mas com potencial para se tornar uma ciência robusta epistemologicamente falando. Apesar de ser tão antiga quanto a história da humanidade, somente muito recentemente teve seu reconhecimento como uma ferramenta organizacional (Marteletto, 2001).

Nas subseções, abordaremos diversos tópicos relacionados às redes complexas, análise de redes de informação, redes bipartidas e projeção bipartida, além de explorar a aplicação da ciência das redes na área da Ciência da

⁶ Looker Studio, disponível em: <https://lookerstudio.google.com/>

⁷ Power BI Desktop, disponível em: <https://powerbi.microsoft.com/pt-br/>

Informação. Cada ponto destacado oferecerá uma compreensão sobre os conceitos associados a essas temáticas, proporcionando uma visão abrangente do papel das redes na análise e organização da informação no contexto da Ciência da Informação.

3.3.1 Redes Complexas

Redes, podem ser definidas em termos simples, por um conjunto de nós (vértices) e ligações ou conexões (arestas). Estudar suas propriedades nos dão uma noção simplificada de um sistema, de uma estrutura ou topologia. Cientistas têm desenvolvido há anos modelos matemáticos e computacionais para analisar, modelar e entender as redes devido o seu potencial uso em diferentes áreas (Newman, 2010). Nossa sociedade está cada vez mais dependente de redes e se fosse possível resumi-la em uma palavra, esta seria “conectada” (Cavalcante, 2009). Essas conexões têm sido difundidas principalmente com o advento da internet, mais especificamente pela Web (World Wide Web) que é um subconjunto da internet.

A teoria dos grafos tem sido a principal ferramenta da matemática para a descrição das redes. Segundo Cavalcante (2009), Kircher é considerado o primeiro pesquisador a investigar redes, porém o estudo das redes começou realmente a partir de Leonhard Euler publicando um artigo sobre o problema das sete pontes de Königsberg. A grande contribuição de Euler foi o fato de ver o problema como um grafo. Mas foi somente no século XX que ocorreu o grande desenvolvimento de estudos nessa área, principalmente para a análise de redes sociais (ARS) e a migração de conceitos da teoria dos grafos para as ciências sociais (Murty et al., 1976 apud Cavalcante, 2009).

Embora decorrentes da teoria dos grafos, as redes complexas tornaram-se importante campo de pesquisa pela possibilidade de poder representar desde sistemas no campo da biologia até a internet. A diferenciação que se faz entre uma rede simples e uma rede complexa pode ser expressa da seguinte forma: nas redes simples pode-se ter uma ideia de sua conectividade somente por meio do seu grau, que diz respeito ao número de conexões de um nó (Costa, 2018). Segundo o autor, as redes complexas exigirão a especificação de muito mais propriedades ou recursos adicionais.

Inclui-se na Ciência das Redes a ARS. Essa área 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 ARS. Considerada como um dos métodos de apoio à descoberta de conhecimento e uma das áreas promissoras, conforme Porto e Ziviani (2014), 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. Segundo Cavalcante (2008), a CI vem colaborando para o desenvolvimento da Ciência das Redes.

A utilização do conceito de redes na sociologia está ligada à estrutura da rede, no que diz respeito aos nós e suas ramificações (Matheus, 2006). O autor retrata que o conceito foi relatado de modo precursor pelo sociólogo Jacob Moreno, sendo o mais citado pela utilização do uso de grafos para representar a estrutura de redes sociais.

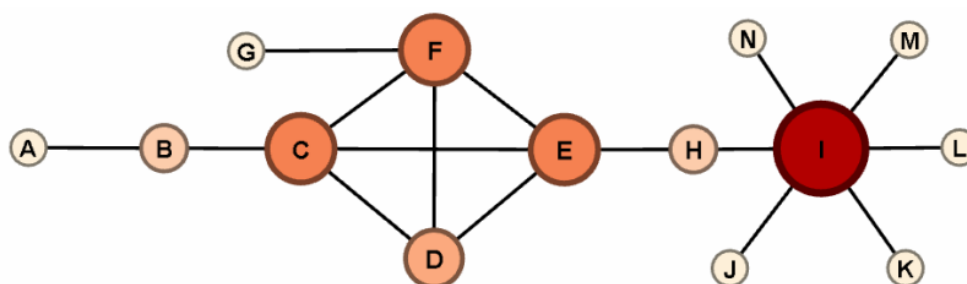
3.3.2 Análise de Redes de Informação

Uma rede social pode ser considerada um tipo de rede de informação, pois contém, sobretudo, informações sobre entidades sociais. As redes de informação conhecidas também por redes de conhecimento, possuem características e objetivos similares às identificadas nas redes sociais, exercendo papel fundamental na sociedade para o fortalecimento de ações em prol da mesma. Basicamente uma rede de informações trata de um grupo de unidades voltadas para o interesse comum, que podem ser obtidas por meio da compilação de bases de dados de conhecimento formal, um sistema cooperativo de catalogação, uma rede de citação de artigos e podem ainda ser utilizados como base para estudos em informetria (Tomaél, 2016).

Por meio da análise da topologia de uma rede pode-se ter a identificação da importância de um nó ou de um grupo de nós em uma rede (Newman, 2010). Para isso foram desenvolvidas métricas a fim de identificar e quantificar essas relações. A presente Seção descreve, nos próximos parágrafos, três métricas que são muito usuais em trabalhos de análise de redes de informação.

A centralidade de grau mede a importância de um nó baseado na quantidade de ligações que possui. A Figura 5 mostra uma rede com 14 nós contendo destaque de tamanho e cores mais avermelhadas nos nós de maior grau de centralidade. Por exemplo, o nó 'I', com grau 6, é o que tem maior grau na rede. Em seguida os nós 'C', 'E' e 'F' possuem grau igual a quatro.

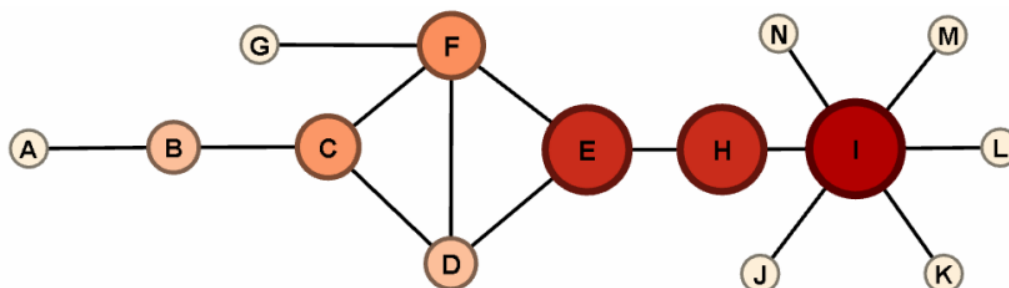
Figura 5 - Rede com destaque pela centralidade de grau



Fonte: Cristovão (2016)

A centralidade de intermediação, também conhecida por *betweenness centrality*, mede o potencial dos nós que servem de intermediários, ou o quanto o nó atua como ponte facilitadora do fluxo de informações em determinada rede (Marteleto, 2001). A Figura 6 mostra uma rede com 14 nós contendo destaque de tamanho e cores mais avermelhadas nos nós de maior grau de intermediação. De forma comparativa com a Figura 5, sobre centralidade de grau, os nós 'E' e 'H', Figura 6, passam a ter grande destaque pela intermediação que exercem na rede como um todo.

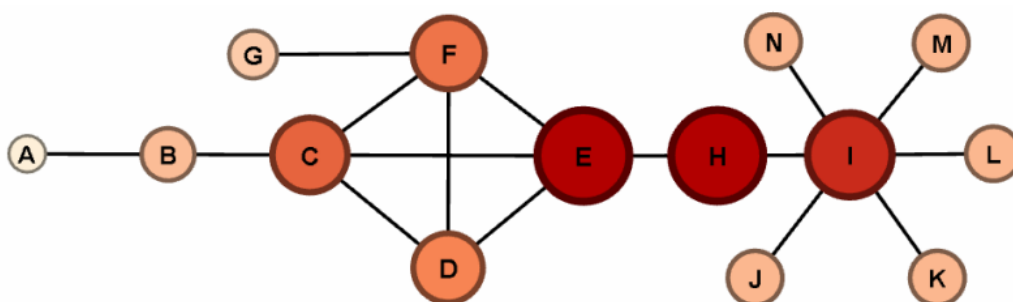
Figura 6 - Rede com destaque pela centralidade de intermediação (*betweenness centrality*)



Fonte: Cristovão (2016)

A centralidade de proximidade, também conhecida por *closeness centrality*, significa que um ator é tão mais central quanto menor a distância que precisa percorrer para alcançar os outros nós da rede. A Figura 7 mostra uma rede com 14 nós contendo destaque de tamanho e cores mais avermelhadas nos nós de maior proximidade. De forma comparativa com a Figura 5, sobre centralidade de grau, os nós 'C', 'D', 'E', 'F', 'H' e 'I', Figura 7, passam a ter maior destaque por estarem mais próximos aos outros nós da rede.

Figura 7 - Rede com destaque pela centralidade de proximidade (*closeness centrality*)



Fonte: Cristovão (2016)

A inspeção visual é parte muito importante na análise de redes. Nessa técnica, que se utiliza da visualização de informação, são direcionadas investigações visuais sobre redes devidamente formatadas para facilitar a revelação de relacionamentos, nós e *cluster* de nós importantes.. A sentença "*what you see drives your search*", traduzido para "o que você vê direciona sua pesquisa", representa bem a ideia

desse tipo de investigação para extração de informações (Penna; Magazzeni, Orefice, 2012). No entanto há uma forte dependência de uma boa formatação das redes de informação geradas para que se consiga usufruir dos benefícios da inspeção visual.

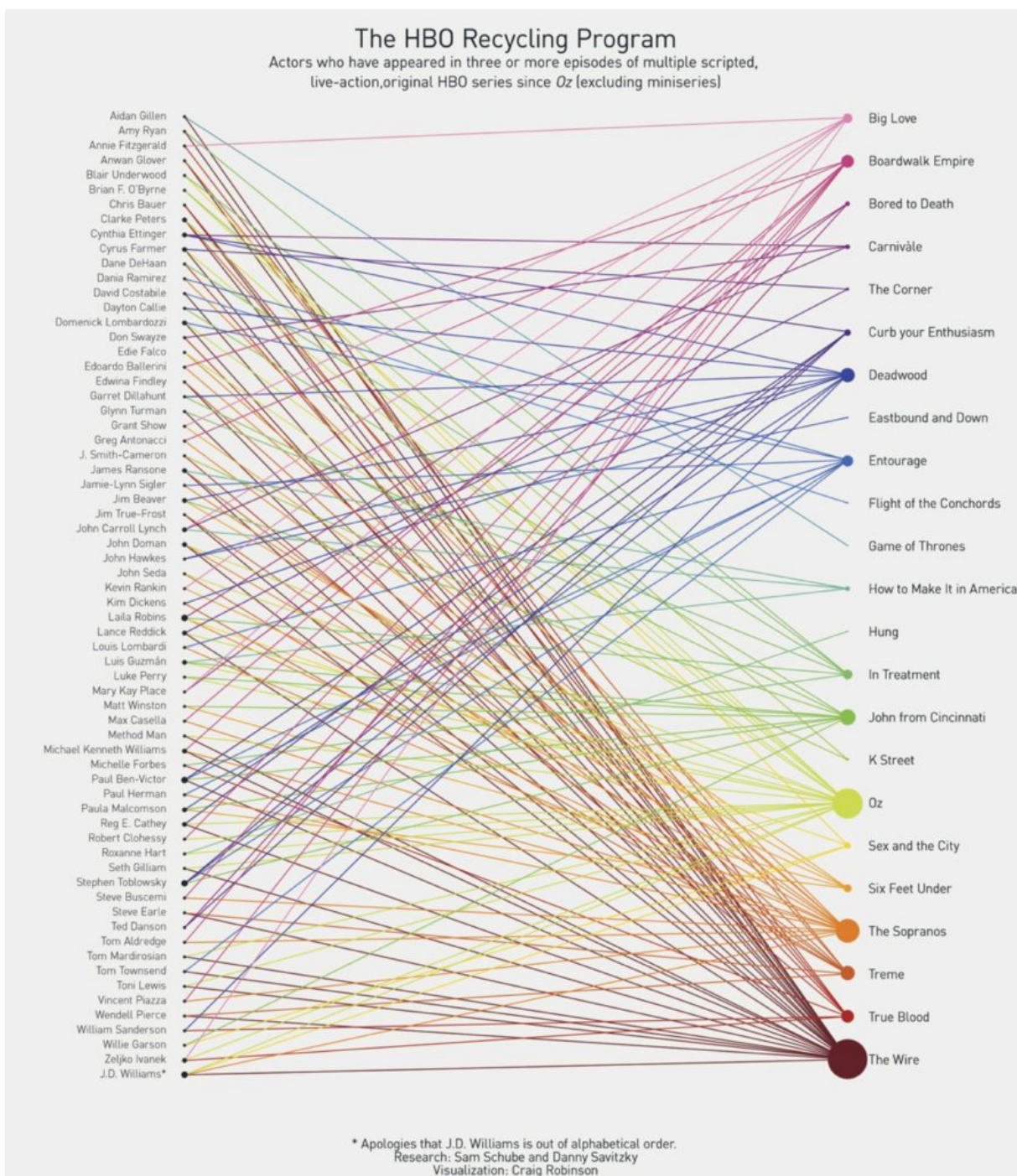
3.3.3 Redes Bipartidas e Projeção Bipartida

As redes bipartidas, ou redes de afiliação, ou ainda redes de dois modos ou *two-mode*, são considerados importantes meios para explicar agrupamentos ou clusterizações (Newman, 2002). Esse tipo de rede é formado por dois grupos de nós, sendo que as conexões só são permitidas entre os nós de grupos diferentes, isto é, não são permitidas conexões entre nós do mesmo grupo.

Um exemplo clássico de rede bipartida é uma rede de atores e suas participações em produções cinematográficas. A rede da Figura 8 mostra esse exemplo, citado por Barabási (2014), aplicado ao contexto de atores que participam de programas da HBO⁸ em uma rede de atores de Hollywood, na qual um conjunto de nós corresponde a filmes e o outro a atores.

Figura 8: Rede bipartida: Conjunto de atores e conjunto de programas da HBO que eles atuaram.

⁸ Home Box Office (HBO) é um rede de televisão por assinatura estadunidense, de propriedade da Warner Bros. Discovery (WIKIPÉDIA).



Fonte: <http://social-dynamics.org/projecting-bipartite-network-gephi/>

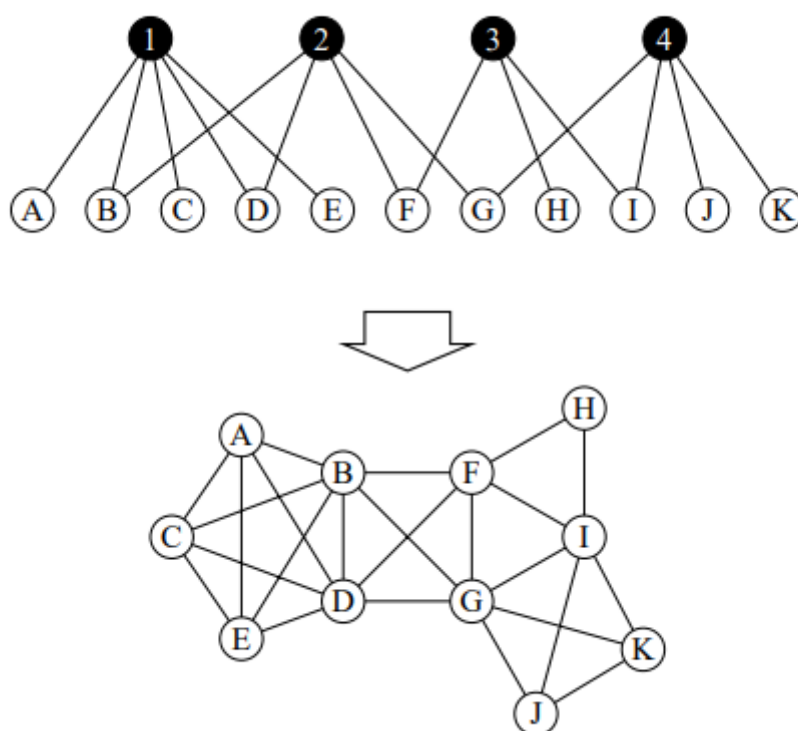
Existem também redes tripartidas e n-partidas, que possuem a mesma linha de raciocínio, isto é, possuem três ou n grupos distintos de nós.

Uma projeção bipartida, da rede bipartida da Figura 8, mostraria como resultado uma rede de atores, na qual dois nós são conectados entre si se tiverem participado do mesmo programa. Essa rede resultante da projeção bipartida,

composta por apenas uma categoria de nós, é denominada de rede monopartida, ou rede *one mode*. Outra possibilidade de projeção bipartida para essa mesma rede, Figura 8, seria a rede monopartida dos programas, na qual dois programas são conectados se compartilharem pelo menos um ator em seu elenco.

A técnica da projeção bipartida é utilizada para evidenciar relacionamentos para facilitar uma inspeção visual, tendo o potencial de revelar informação adicional dos dados de uma rede bipartida (Everett; Borgatti, 2012).

Figura 9 - Exemplo de uma projeção bipartida



Fonte: Newman (2002)

A parte de cima da Figura 9 expressa, hipoteticamente, uma rede bipartida de cientistas (nós de A a K) e de artigos científicos (nós de 1 a 4) com ligações entre o cientista ao artigo no qual seu nome apareceu como coautor. A parte de baixo da Figura 9, mostra o resultado da projeção bipartida, eliminando-se os artigos, para obter, como resultante, a rede monopartida de coautorias entre os cientistas.

Em uma análise crítica realizada por Everett e Borgatti (2012) a respeito da eficácia no uso da técnica de projeção bipartida, os autores concluíram que por meio da técnica, as redes resultantes tiveram sua análise facilitada.

3.3.3 Ferramenta para análise de redes de informação

Existem diversos softwares que têm a finalidade de auxiliar a tarefa da análise de redes de informação.

O software Gephi⁹ é um software interativo para visualização, exploração e análise de vários tipos de redes complexas, grafos dinâmicos e hierárquicos. Ele trabalha com vários formatos de arquivos que configuram as características de uma rede, além do seu próprio formato. Um desses formatos é o GML¹⁰ (*graph modelling language*). Esse formato é especialmente interessante pois possibilita a integração com o software OpenRefine.

O Gephi possui um plugin, denominado MultiMode Networks¹¹, para a geração de redes por meio da projeção bipartida.

3.3.4 Ciência das Redes na Ciência da Informação

Aliada à representação e organização dos dados, a análise de redes apresenta-se como tema emergente nos últimos anos, e abre perspectivas para questões de grande interesse na atualidade, como a ARS associadas a diversos contextos, principalmente como estratégia para tomada de decisão. Surge então uma nova ciência chamada de Ciência das Redes (BARABÁSI, 2013). A Ciência da Informação veio estabelecer ligações com a Ciência das redes e tornou-se importante no campo das ciências sociais ganhando destaque com o surgimento da internet, especialmente por meio da World Wide Web, que constitui uma parte integrante desse vasto ambiente digital. A internet tem a capacidade de estabelecer conexões sociais entre as pessoas, enriquecendo seu capital social. É essencial criar ferramentas que auxiliem as pessoas na navegação e na busca de

⁹ Gephi, disponível em: <https://gephi.org/>

¹⁰ GML (graph modelling language) é um formato universal para representação de grafos. Disponível em: https://en.wikipedia.org/wiki/Graph_Modelling_Language.

¹¹ Plugin do Gephi utilizado para geração de redes de projeção bipartida.

conhecimento em sociedades complexas, fragmentadas e interconectadas, como destacado por Wellman (2001).

Considerado como um dos autores de renome nas áreas de análise de redes, Patrick Doreian está entre os que têm maior influência especialmente na ARS. Junto a Barry Wellman são os cientistas mais centrais no campo, publicando vários artigos em revistas de Ciência da informação. A ARS têm interferências além da sociologia, nos campos da matemática e da ciência da computação (Otte et al., 2002). O uso da ARS vem crescendo significativamente nos últimos anos, especialmente pela evolução da tecnologia e consequente aumento da quantidade de dados disponíveis (Matheus, 2006). Borgatti e Foster (2003) enfatizam o fato do crescimento da pesquisa em ARS se dar em várias áreas do conhecimento especialmente a partir do início dos anos 70.

Dentre os problemas de pesquisas associados à Ciência da Informação e que podem ser resolvidos por meio da abordagem de ARS podemos incluir o presente trabalho nos temas relacionados ao Fluxo de Informação dentro das organizações, Gestão do conhecimento e tomada de decisão (Matheus, 2006).

Este estudo fundamenta-se na estratégia conhecida como "rede dinâmica", conforme proposta por Deroy-Pineau, que envolve abordar as redes como uma estratégia de ação com o propósito de gerar instrumentos para mobilização de recursos. Em um contexto mais amplo, essa pesquisa, denominada eficácia das redes, apresenta uma dupla aplicação: a "utilização estática" e a "utilização dinâmica". Diferenciando-se da abordagem estática, que busca explorar as redes para compreender a sociedade ou um grupo social por meio de sua estrutura, nós e ramificações (MARTELETO, 2001), o foco desta investigação concentra-se na perspectiva dinâmica da rede.

3.4 OPEN FINANCE

O Open Finance é um modelo que visa tornar o mercado financeiro mais aberto e acessível, a partir do compartilhamento de informações financeiras autorizadas entre instituições. Ele envolve o uso de tecnologias como APIs (Interface de Programação de Aplicativos) para permitir que empresas e indivíduos acessem

dados financeiros de diversas fontes em um único lugar (Open Finance Brasil, 2024). Os aplicativos financeiros se comunicam, oferecendo uma experiência mais integrada, segura e personalizada. Todo esse processo é regulado pelo Banco Central do Brasil. Essa Seção trata do surgimento do Open Finance, seu funcionamento, tendências e sua interconexão com a Ciência da Informação.

3.4.1 Surgimento

Diante de uma revolução dos serviços bancários e financeiros, Paravela e Domingues (2021) sinalizaram que o Open Banking surgiu no Reino Unido e tem desenvolvido-se no mundo contando com o apoio da OCDE¹². Austrália, Cingapura, México, entre outros países já contam com o sistema financeiro aberto. No Brasil, iniciou-se por meio do Comunicado 33.455, de 24 de abril de 2019, no qual, o Banco Central do Brasil divulgou premissas para a implementação no país, definindo objetivos e estratégias para a implementação.

O conceito do Open Finance surge estrategicamente para fomentar a competição, inovação e inclusão no setor financeiro. Sob essa perspectiva, os consumidores ganham controle sobre seus dados, decidindo quais compartilhar e com quem. Com uma gama expandida de produtos financeiros, a comparação e contratação tornam-se mais acessíveis. A segurança e transparência são prioritárias, com o compartilhamento de dados condicionado à autorização do titular, seguindo as diretrizes normativas do Banco Central, assegurando a conformidade e confiança no sistema. O Open Finance é a ampliação do Open Banking. O projeto mudou de nome para mostrar sua maior abrangência, incluindo além de informações sobre produtos e serviços financeiros mais tradicionais (contas e operações de crédito), serviços de câmbio, investimentos, seguro e previdência (Open Finance, 2024).

O Open Finance iniciou com as instituições participantes disponibilizando informações padronizadas sobre os seus canais de atendimento e as características dos produtos comercializados pelas instituições e serviços bancários tradicionais

¹² A OCDE é uma organização internacional com 38 estados-membros, fundada em 1961 com o objetivo de promover políticas fiscais, comerciais e de bem-estar entre seus membros e o resto do mundo e estimular o progresso econômico mundial. Seus atuais membros são: Alemanha, Austrália, Áustria, Bélgica, Canadá, Chile, Colômbia, Coreia do Sul, Costa Rica, Dinamarca, Eslováquia, Eslovênia, Espanha, Estados Unidos da América, Estônia, Finlândia, França, Grécia, Hungria, Irlanda, Islândia, Israel, Itália, Japão, Letônia, Lituânia, Luxemburgo, México, Nova Zelândia, Noruega, Países Baixos, Polônia, Portugal, Reino Unido, República Tcheca, Suécia, Suíça e Turquia (WIKIPEDIA).

oferecidos. Mais tarde, foram acrescentados a essa lista dados sobre produtos referentes a câmbio, seguros, previdência e investimento. Esses dados foram publicados em formato aberto possibilitando que qualquer desenvolvedor de programas utilize gratuitamente e tenha a possibilidade de construir soluções que comparem diferentes ofertas de produtos e serviços financeiros. Exemplos que podem surgir são: comparadores de tarifas bancárias, de tipos de contas e de cartões de crédito. Essas ferramentas poderão auxiliar as pessoas a escolherem a opção mais adequada ao seu perfil e necessidades (Open Finance, 2023). Essa realidade incentivará a inovação e o surgimento de novos modelos de negócios e difusão de informação, favorecendo a inclusão e educação financeira da população (Open Finance Brasil, 2022).

3.4.2 Funcionamento e Tendências

O Open Finance, ou sistema financeiro aberto, é a possibilidade de clientes de produtos e serviços financeiros permitirem o compartilhamento de suas informações entre diferentes instituições autorizadas pelo Banco Central do Brasil, e a movimentação de suas contas bancárias a partir de diferentes plataformas.

O tipo de informação acessada no Open Finance depende dos produtos e serviços financeiros que o usuário utiliza e das instituições autorizadas a compartilhar os dados. Por exemplo, no caso de uma conta corrente, um cartão de crédito, um seguro e um plano de previdência, são compartilhados dados como saldo, limite, apólice e extrato. Esses dados podem ser usados por outras instituições para oferecer produtos e serviços mais adequados ao perfil, como taxas de juros mais baixas, coberturas mais amplas ou rentabilidades mais altas. Sendo assim, pessoas e empresas podem solicitar o compartilhamento de seus dados cadastrais, de informações sobre transações em suas contas, cartão de crédito e produtos de crédito contratados. O compartilhamento ocorre através de um aplicativo ou site da instituição escolhida para o compartilhamento das informações ocorrendo apenas entre instituições participantes do Open Finance, sempre para finalidades determinadas e por um prazo específico. É necessária a autorização da pessoa cliente, sendo possível o cancelamento da autorização a qualquer momento nas instituições envolvidas no compartilhamento.

Com o Open Finance as instituições podem se conectar às plataformas de outras instituições e acessar exatamente os dados autorizados pelos clientes. O cliente passa a ter um papel preponderante em relação a seus dados financeiros e os bancos têm condições de desenvolver serviços de maneira personalizada e adequada ao perfil de cada cliente, bem como as suas necessidades (Banco do Brasil, 2022). Um exemplo simples de utilização pode ser o acesso a um extrato multibanco em um só lugar. Outra aplicação prática seria na contratação de um empréstimo pessoal: Ao acessar o aplicativo de uma fintech que oferece esse serviço e autorizar o compartilhamento dos dados bancários, a fintech pode usar esses dados para analisar o perfil e oferecer uma proposta de empréstimo com condições mais vantajosas do que o banco de origem. Existe a possibilidade de aceitar a proposta e receber o dinheiro na conta corrente, sem precisar abrir uma conta na fintech. O Open Finance tem relevância significativa na concessão de crédito, beneficiando tanto pessoas físicas quanto jurídicas. Rodrigo Cabernite, co-CEO da Gyra+, destaca que o Open Finance facilita a obtenção de informações essenciais para tomada de decisões de crédito. Ao proporcionar acesso a dados anteriormente restritos a uma única instituição financeira, essa iniciativa concede às pessoas a propriedade de seus próprios dados, possibilitando a transferência independente entre as instituições participantes (Capital, 2024).

A principal vantagem do Open Finance no setor de crédito é a ampla variedade de ofertas disponíveis. Com as instituições financeiras tendo acesso aos dados compartilhados via Open Finance, surge uma maior competitividade, levando as instituições a disputarem o cliente com as ofertas mais adequadas ao seu perfil. Um levantamento divulgado pelo Serasa revela a desigualdade social nas tentativas de contratação de crédito no país. A pesquisa aponta que os bancos negam aproximadamente 44% das solicitações de acesso a empréstimos, financiamentos e outras concessões financeiras para aqueles que recebem menos de cinco salários mínimos por mês (R\$5500). Esse índice reduz para 18% entre os que possuem rendimentos acima desse valor (Correio Brasiliense, 2024).

Naturalmente, surgem nesse contexto, novos comportamentos dos usuários no mercado financeiro, como por exemplo, a escolha de produtos da conta corrente, entre diversos bancos, ou diferentes *players*. Percebe-se que há uma liberdade nunca antes vista nesse cenário. Esse fenômeno promete então revolucionar a forma como se lida com produtos, serviços e dados financeiros dos cidadãos.

Encontra-se disponível também a possibilidade de realizar pagamentos no ambiente de Open Finance. O Pix é a primeira modalidade a ser disponibilizada. A iniciação de pagamentos no ambiente Open Finance possibilita pagamentos por meio de canais mais convenientes para o cliente, preservando a segurança do processo. Pagar em lojas virtuais, por exemplo, torna-se mais rápido e cômodo. Existem também outros projetos que fazem parte desse ecossistema, como o Open Insurance, que é o projeto de sistema de seguros aberto, e o Open Capital Market denominado também como PIX do mercado de capitais: a portabilidade de investimentos instantânea e sem custos de uma corretora para outra, que poderá ser feita por meio de um app no celular (B3, 2023). Fazer com que esses ecossistemas sejam interoperáveis significa possibilitar que esses ambientes possam “conversar” entre si. Ou seja, dados e informações poderão transitar entre eles, trazendo maior comodidade e benefícios às pessoas que os utilizam. Essa é uma etapa prevista na agenda evolutiva do Open Finance e está sendo construída pelo Banco Central conjuntamente com as respectivas estruturas de governança.

A indústria de serviços financeiros pode ser considerada uma indústria de serviços baseada em dados, pois grande parte do valor gerado decorre de melhorias nos processos baseados em dados (Banco Central, 2023). Na agenda de inovação do sistema financeiro feita pelo Banco Central, o Open Finance encaixa-se em um contexto em que o *big data* é condição necessária para que tecnologias com a inteligência artificial possam atuar de modo preponderante. Ele está aliado a outras inovações como o PIX e suas variadas formas e o DREX, moeda digital brasileira. Por conta disso, gigantes da tecnologia como o Google estão investindo prioritariamente em serviços financeiros, é o caso do estudo FINFACTS do Google, enfatizando que o bom uso da tecnologia é fator decisivo na performance das instituições financeiras (FINFACTS, 2023). O desdobramento de todas essas mudanças se consolida em uma única solução denominada de “Super App” do Banco Central. Trata-se de uma aplicação de soluções integradas ou agregador de serviços financeiros reunindo todas as informações financeiras em um único aplicativo (Banco Central, 2023).

Está claro que o Open Finance trará um contexto de mudanças. Segundo Matta (2010), todos os envolvidos em um processo de mudança de comportamento necessitam, em se tratando de aspectos informacionais, buscar e utilizar informações para alcançar os objetivos almejados. Portanto, questões relativas à

gestão da informação, como acessá-las e difundi-las, tornaram-se áreas discutidas por estudiosos da CI, pois é sabido que os sistemas de informação, por si só, não são suficientes para responder às demandas informacionais das pessoas e organizações.

3.4.3 Open Finance na Ciência da Informação

Segundo Cherobim (2017) todos os empreendimentos necessitam da atividade financeira, seja na captação ou aplicação de recursos. E ainda Gonzalez (2016) defende que a necessidade por serviços financeiros mais eficazes e adequados continua substancial. A atividade de informar é considerada um negócio pois a informação é um produto, um bem comercial (Borges, 2018).

Le Coadic (2004) afirma que a informação consiste em um conhecimento inscrito em forma escrita, oral ou audiovisual, em um suporte. Dentro desse cenário encontra-se o usuário que necessita de informação que lhe transmita segurança para o compartilhamento de seus dados. Nesse contexto, Wandscheer, Jarude e Vita (2020) frisam que as empresas deverão obter consentimento dos clientes para o uso da informação dos mesmos. Faz-se necessário criar um ambiente de maior segurança, que transgrida a informação técnica e dê ao usuário a segurança do conhecimento. Como a informação é o oposto da incerteza (Le coadic, 2004; McGarry, 1999), medir o efeito de uma mensagem é justamente mensurar a quantidade de incerteza que reduziu. Não se tratando apenas de uma medição, mas de uma condição essencial da condição humana (McGarry, 1999). Entende-se portanto que, ao informar-se o usuário tem uma compreensão mais clara do ambiente em que está operando, contribuindo para sua segurança, transparência e tomada de decisões financeiras mais conscientes. Além disso, a democratização do acesso à informação está inerentemente ligada ao caminho para interoperabilidade e consequente digitalização de toda a sociedade (Ferreira, 2019).

Estudos têm avançado de forma progressiva nos domínios das Ciências da Informação e da Computação, direcionando-se para investigações sobre o desafio do volume e organização das informações. O propósito é aprimorar a eficácia dos sistemas de recuperação de informação (Da Silva; Souza, 2014). Fazer com que os ecossistemas do Open Finance sejam interoperáveis significa possibilitar que dados

e informações possam transitar entre eles. Na perspectiva científica, a interoperabilidade no contexto do Open Finance é intrinsecamente vinculada à importância dos metadados (De Andrade, 2017). Estes desempenham um papel essencial ao oferecer informações contextuais cruciais para a compreensão e integração eficiente de dados entre os diversos ecossistemas financeiros. Agindo como descritores detalhados sobre o conteúdo, origem e formato dos dados, os metadados são fundamentais para a interpretação apropriada e para facilitar a interação entre dados compartilhados. Em suma, a gestão precisa de metadados constitui uma peça-chave para o sucesso da interoperabilidade no Open Finance, proporcionando uma base científica sólida para a interação eficaz entre os ecossistemas financeiros interconectados (Prado, 2019). Outro ponto importantíssimo que conta com a contribuição da Ciência da Informação é justamente sobre a especificação de informações sem ambiguidades em busca da interoperabilidade semântica entre sistemas. A solução pode estar na representação da informação por meio das ontologias de domínio, embora haja uma lacuna existente entre os métodos de construção e sua utilização na prática (Farinelli; Elkin, 2017). Para analisar o contexto do Open Finance é essencial ter em mente que a informação resulta da interseção dessas duas perspectivas: o que o dado "indica" e aquilo que o conhecimento "possibilita" entender a partir do dado. Isso estabelece um cenário de compreensão da informação mais elaborado (Araújo, 2013).

Dessa forma, profissionais da Ciência da Informação se comportam como observadores desse fenômeno denominado Open Finance, a fim de trazer soluções e informações que beneficiem os usuários, satisfazendo necessidades essenciais e contribuindo sociologicamente com os indivíduos, trazendo objetivos de pesquisa inerentes à Ciência da Informação.

4 PROCEDIMENTOS METODOLÓGICOS

Este capítulo classifica a pesquisa metodologicamente, descreve os procedimentos usados no levantamento bibliográfico do referencial teórico e do estado da arte, descreve o cenário da pesquisa e apresenta os procedimentos metodológicos das etapas do desenvolvimento.

4.1 CLASSIFICAÇÃO DA PESQUISA

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 CI. A coleta de dados foi realizada utilizando dados financeiros e de operações de crédito a partir de dados abertos do Banco Central do Brasil.

Uma abordagem qualitativa busca a compreensão do objeto de pesquisa. Segundo Minayo (2003) a pesquisa qualitativa lida com uma variedade de conceitos, inspirações, concepções, princípios e condutas relacionados aos procedimentos e ocorrências que não são quantificáveis. O que possibilita uma análise em todo o processo. A subjetividade é um aspecto relevante na pesquisa qualitativa, pois permite uma interpretação dos dados pesquisados (Teis, 2022).

A pesquisa bibliográfica é embasada nos conhecimentos teóricos da pesquisa, e tem a função de possibilitar ao pesquisador acesso aos documentos produzidos e registrados sobre o tema da pesquisa (Pádua, 2010). A pesquisa bibliográfica faz um levantamento de documentos publicados nos mais diversos formatos e suportes, para o acesso do pesquisador.

Sendo assim, entende-se que a pesquisa bibliográfica e a abordagem qualitativa foram essenciais para o desenvolvimento deste estudo.

Sem a pretensão de estabelecer um discurso conclusivo sobre as questões pesquisadas, buscou-se analisar os conceitos chave tratados, contribuindo com

novas reflexões e perspectivas de estudo.

4.2 LEVANTAMENTO BIBLIOGRÁFICO

A presente pesquisa, relativa ao referencial teórico, foi fundamentada em estudos em língua portuguesa e inglesa sem definição de lapso temporal, e contidos nas seguintes bases de dados: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD – IBICT)¹³, Google Acadêmico¹⁴ e Portal de Periódicos da CAPES¹⁵.

Em linhas gerais, o exame das pesquisas sobre Open Finance revela que a maioria dos estudos se concentra na proteção e privacidade dos dados pessoais dos usuários do Open Finance. Há também trabalhos que se concentram em realizar um percurso histórico e conceitual em relação ao Sistema Financeiro Nacional Brasileiro, apresentando seu contexto, estrutura e evolução.

Nas buscas realizadas, no contexto do Open Finance, foram recuperados seis documentos relevantes para a pesquisa. Foram utilizados os termos de busca “Open Banking” e “Sistema Financeiro”. Dos seis documentos recuperados, foram utilizados: Paravela e Domingues (2021), Wandscheer, Jarude e Vita (2020). Os documentos "Paravela e Domingues (2021)" e "Wandscheer, Jarude e Vita (2020)" foram selecionados devido à sua pertinência temática e atualidade. O primeiro aborda a perspectiva do consumidor na implementação do Open Banking, destacando questões como transparência e segurança. Enquanto isso, o segundo documento analisa a interseção entre o Open Banking, a regulação bancária e a Lei Geral de Proteção de Dados (LGPD), fornecendo insights sobre os desafios regulatórios e jurídicos associados. Ambos contribuem significativamente para a compreensão do Open Banking no contexto brasileiro, enriquecendo a pesquisa com perspectivas valiosas e atualizadas.

Para o capítulo sobre Ciência da Informação e Organização do Conhecimento, foram utilizados autores da CI indicados em aulas de disciplinas cursadas pela autora desta pesquisa no Programa de Pós-graduação em Ciência da Informação (PPGCI) da Universidade Federal do Espírito Santo (UFES). Dentre

¹³ Biblioteca Digital Brasileira de Teses e Dissertações (BDTD – IBICT), disponível em: <https://bdtd.ibict.br/>

¹⁴ Google Acadêmico, disponível em: <https://scholar.google.com.br/?hl=pt>

¹⁵ Portal de Periódicos da CAPES, disponível em: <https://www.periodicos.capes.gov.br/>

dessas, as que tratam de aspectos informacionais no contexto de ambientes dinâmicos: Matta (2010), McGarry (1999), e Le Coadic (2004), onde a mudança é prática previsível, e que tem a informação como subsídio para sanar as incertezas de tais ambientes. Outras no contexto dos aspectos epistemológicos da CI: Araújo (2018) e Borko (1968) e autores que tratam dos sistemas de organização da informação e organização do conhecimento: Hjørland (2008) , Barité (2021) e Zeng (2008).

Para o capítulo sobre Ciência das Redes, utilizou-se autores importantes tais como: Cavalcante (2008), Buchanan (2003), Barabási (2013), Newman (2002).

A Subseção 4.2.1 descreve os procedimentos metodológicos específicos do levantamento bibliográfico realizado para a construção do estado da arte do Capítulo 2.

4.2.1 Levantamento bibliográfico do Estado da Arte

Para o estado da arte foram recuperados quinze documentos relevantes para a pesquisa, utilizando os termos "análise de redes" e "sistema financeiro" em português e inglês. Como resultado obteve-se o retorno de 547 resultados. Configurado o período para 2013-2023, obteve-se 402 resultados. Trabalhou-se os seguintes critérios para o recorte temático:

- 1- Teorias que tratam sobre aplicação do método de análise de redes no contexto do sistema financeiro nacional ou
- 2-Pesquisas com objetivos de analisar resultados com o método de análise de redes ou
- 3- Pesquisas que tratam sobre a análise de redes sociais no geral.

O critério de seleção utilizado para a escolha das pesquisas inicialmente foi o título das pesquisas que tratam especificamente sobre a utilização das técnicas de análise de redes para resolução de algum problema no contexto do sistema financeiro. Depois os resumos das obras foram analisados para a verificação mais apurada dos objetivos da pesquisa.

Passando nesse trâmite, os quinze trabalhos resultantes foram descritos no Capítulo 2: o seu objetivo, a metodologia, o que conseguiram de resultados e a correlação com o presente trabalho.

4.3 CENÁRIO DA PESQUISA

O Portal Brasileiro de Dados Abertos do Banco Central¹⁶ é o meio utilizado pelo BC para disponibilizar dados e informações públicas. O principal objetivo é o de ampliar a transparência, com eficiência e qualidade, de forma a contribuir para reforçar a credibilidade e o cumprimento da missão institucional do BC, bem como fomentar o controle social, o aperfeiçoamento da integridade, a redução de custos, e a participação social. (Banco Central, 2023).

As informações são disponibilizadas assim como foram recebidas, sendo elas de responsabilidade das respectivas instituições/empresas, de acordo com a regulamentação em vigor.

Dada a significativa importância do Open Finance no cenário de crédito, a base de dados denominada "SCR.data - Painel de Operações de Créditos" tem potencial para revelar informações cruciais para os consumidores. O Banco Central do Brasil divulga mensalmente informações agregadas das operações de crédito recebidas através do Sistema de Informações de Créditos – SCR¹⁷. 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 pesquisa em questão empregou duas bases de dados: uma contendo 636.812 linhas de registros de crédito referentes ao mês de março de 2019, e outra

¹⁶ 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.

¹⁷ Sistema de Informações de Créditos (SCR): É um sistema mantido pelo Banco Central do Brasil que reúne informações sobre operações de crédito contratadas por pessoas físicas e jurídicas. O objetivo principal do SCR é fornecer dados relevantes para auxiliar as instituições financeiras na análise de risco de crédito e na concessão de empréstimos. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/scr>. Acesso em: 22 jan. 2024.

contendo 903.435 linhas de registros de crédito referentes ao mês de março de 2023. Ambas as bases foram utilizadas com o propósito de conduzir uma análise comparativa temporal, abrangendo o período prévio ao advento do Open Finance e o mês mais recente disponível na base de dados aberta do banco central. O Open Finance teve seu início por meio do Comunicado 33.455, datado de 24 de abril de 2019, conforme mencionado na Seção 3.4.1 do estudo. Para viabilizar a realização da pesquisa, inicialmente, foram utilizadas as 23 variáveis disponíveis na base de dados. Posteriormente, dentre essas variáveis, foram selecionadas quatorze para análise mais aprofundada. A seleção das variáveis foi conduzida com base na sua proximidade com o problema de pesquisa, seguindo os critérios estabelecidos no dicionário de dados¹⁸. Foi priorizada a inclusão daquelas que demonstraram relevância tanto teórica quanto empírica para o problema em questão. Este processo visou garantir que as variáveis escolhidas fossem fundamentais para a análise do fenômeno em estudo, ao mesmo tempo em que mantinham a integridade e a representatividade das informações necessárias para a pesquisa. As variáveis selecionadas serão apresentadas e analisadas no capítulo de resultados e discussão, contribuindo para uma compreensão abrangente do fenômeno investigado.

Segue descrição das variáveis conforme dicionário de dados::

- Ativo Problemático: Essa variável representa uma soma de diferentes tipos de créditos considerados problemáticos. Isso inclui créditos vencidos há mais de 15 dias e parcelas de créditos inadimplidos há mais de 90 dias.
- Carteira Ativa: Refere-se ao conjunto total de operações de crédito ativas em um determinado momento.
- Carteira inadimplida arrastada: Representa o somatório das operações de crédito que estão em atraso há mais de 90 dias, incluindo parcelas vencidas.
- CNAE Seção: O CNAE (Classificação Nacional de Atividades Econômicas) é um sistema utilizado para classificar as atividades econômicas das empresas. A variável "CNAE Seção" indica a Seção a

¹⁸ https://www.bcb.gov.br/content/estabilidadefinanceira/scr/scr.data/scr_data_metodologia.pdf

que pertence a atividade econômica da empresa, como agricultura, indústria, comércio, etc.

- CNAE Subseção: Similar à variável "CNAE Seção", mas oferece uma classificação mais detalhada, indicando a subseção da atividade econômica da empresa.
- Cliente: Essa variável representa os clientes envolvidos nas operações de crédito, podendo ser tanto pessoas físicas quanto jurídicas.
- Data Base: Refere-se à data de referência utilizada para a análise das operações de crédito.
- Indexador: Indica o tipo de indexador utilizado para calcular os juros das operações de crédito, como taxa prefixada, pós-fixada, flutuante, etc.
- Modalidade: Representa a categoria ou tipo de operação de crédito realizada, como empréstimo pessoal, financiamento habitacional, capital de giro para empresas, entre outros.
- Número de Operações: Indica a quantidade total de operações de crédito consideradas em uma análise.
- Ocupação: Refere-se à ocupação principal dos clientes no caso de pessoas físicas, indicando se são servidores públicos, autônomos, empresários, entre outros.
- Origem: Indica se os recursos utilizados nas operações de crédito têm uma destinação específica ou não.
- Porte: Representa o tamanho das empresas no caso de pessoas jurídicas e o rendimento ou faixa de salário no caso de pessoas físicas.
- TCB (Tipo de Consolidado Bancário): Agrupa as instituições financeiras de acordo com sua natureza, como bancos comerciais, instituições de pagamento, cooperativas de crédito, entre outros.
- UF: Refere-se à unidade federativa onde estão localizados os clientes ou sedes das empresas envolvidas nas operações de crédito.
- Valores a vencer até 90 dias: Representa o somatório dos créditos que devem ser quitados em até 90 dias.
- Valores a vencer de 91 a 360 dias: Representa o somatório dos créditos que devem ser quitados entre 91 e 360 dias.
- Valores a vencer de 361 até 1080 dias: Representa o somatório dos

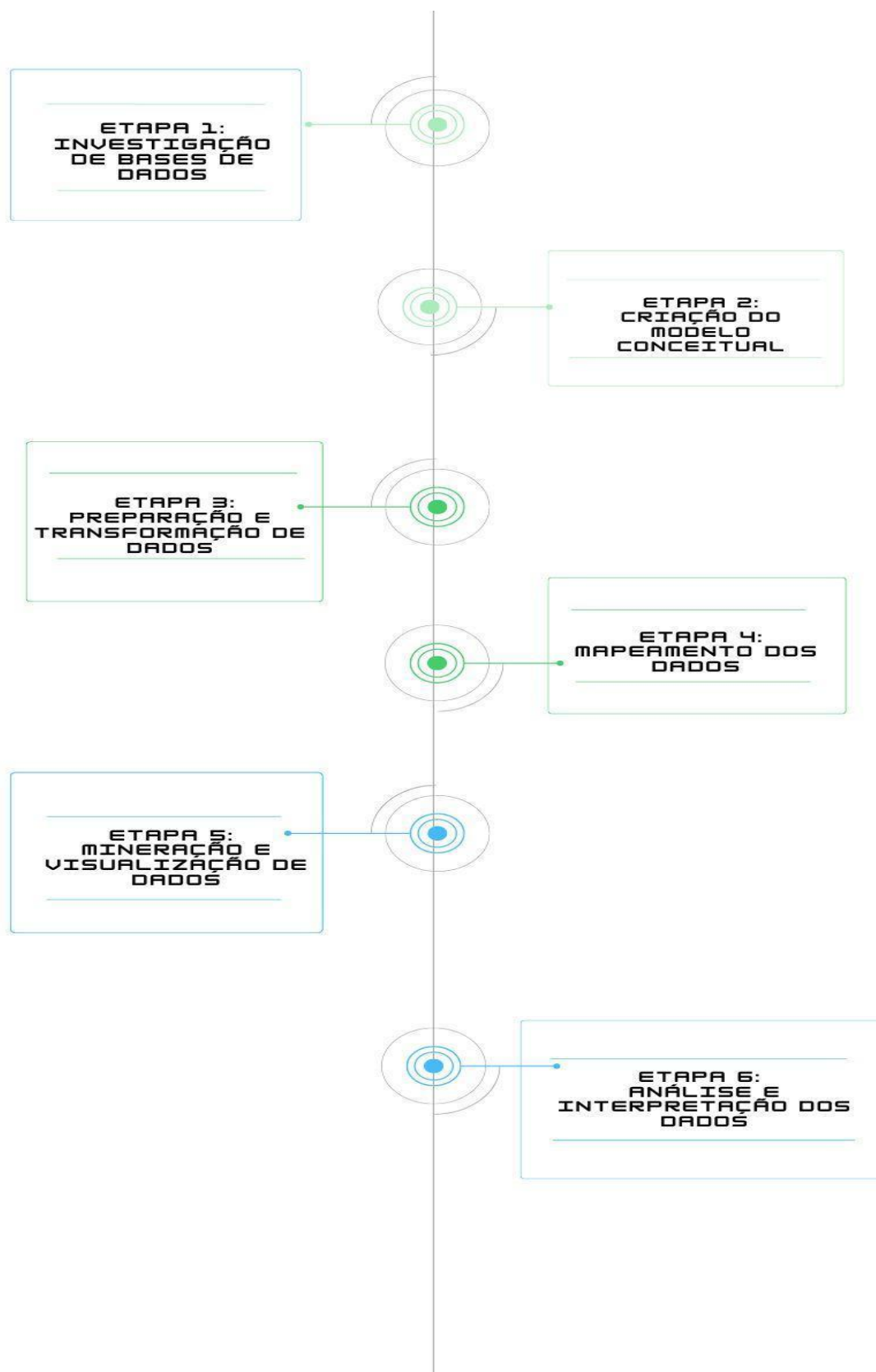
créditos que devem ser quitados entre 361 e 1080 dias.

- Valores a vencer de 1081 até 1800 dias: Representa o somatório dos créditos que devem ser quitados entre 1081 e 1800 dias.
- Valores a vencer de 1801 até 5400 dias: Representa o somatório dos créditos que devem ser quitados entre 1801 e 5400 dias.
- Valores a vencer acima de 5400 dias: Representa o somatório dos créditos que devem ser quitados em mais de 5400 dias.
- Valores vencidos acima de 15 dias: Representa o somatório dos créditos que já venceram e estão em atraso há mais de 15 dias.

4.4 ETAPAS PROCEDIMENTAIS DO DESENVOLVIMENTO

As etapas apresentadas nas próximas subseções estão, de certa forma, sintonizadas com os objetivos específicos da subseção 1.3 e com a metodologia de descoberta de conhecimento proposta por Fayyad et al. (1996), mostrada na subseção 3.2.2. A Figura 10 mostra de forma esquemática essas etapas.

Figura 10 - Etapas procedimentais do desenvolvimento



Fonte: elaboração própria.

As próximas subseções descrevem cada uma das etapas mostradas na

Figura 10.

4.4.1 Etapa 1: Investigação de Bases de Dados

Essa etapa é referente ao objetivo específico (ii) e cuida da investigação de bases de dados abertas do Banco Central do Brasil relativas a dados sobre operações de crédito. Ela também tem como objetivo selecionar variáveis relevantes para a pesquisa.

Na investigação da base de dados, fez-se necessário um estudo com o apoio do dicionário de dados com os metadados para um entendimento do contexto dos dados e o objetivo do uso dos mesmos ajustado ao objetivo geral da pesquisa, que tem o foco na investigação e revelação de relacionamentos entre variáveis de bases de dados abertas do Banco Central relativas a dados sobre operações de crédito.

A etapa utiliza-se da ferramenta OpenRefine para que haja uma melhor exploração das facetas de cada coluna e identificação das entidades para a modelagem.

4.4.2 Etapa 2: Criação do Modelo Conceitual

Nessa etapa, as entidades, seus atributos e os relacionamentos entre elas já foram identificados para a construção da modelagem conceitual. O ponto central desta fase foi a elaboração de uma ontologia de referência de domínio, conforme a metodologia SABIO (Falbo, 2014). Para auxiliar na diagramação da ontologia, foi utilizada a ferramenta Visual Paradigm.

4.4.3 Etapa 3: Preparação e Transformação de Dados

A preparação, limpeza e transformação dos dados foram realizadas tanto na ferramenta Power BI quanto na ferramenta OpenRefine, visando melhorar a usabilidade dos dados. Esses procedimentos envolveram:

- Criação da variável derivada "uf" na tabela de população (Anexo 1): Essa etapa foi necessária para facilitar o relacionamento entre a tabela principal e a tabela de população, permitindo uma análise integrada dos dados por unidade federativa (UF).

- Transformação da variável população em população percentual: A conversão da variável de população em população percentual possibilitou uma análise comparativa, permitindo avaliar a distribuição da população em relação ao total de habitantes.
- Criação da tabela população percentual: A criação desta tabela específica foi importante para separar e organizar os dados de população transformados, facilitando sua utilização para as análises realizadas posteriormente.
- Conversão da variável população estimada percentual em número decimal: Essa conversão foi realizada para tornar os dados mais compreensíveis e fáceis de interpretar, facilitando a análise estatística.
- Criação de variável para o fator de correção: Foi criada uma variável com a expressão $(1 - [\text{população estimada percentual}])$, considerando a proporção da população residente de cada estado. O cálculo do fator de correção foi essencial para ajustar os dados em função das diferenças de tamanho populacional entre os estados, garantindo uma análise mais equitativa e precisa.
- Criação de tabela integrada com estados no Power BI: Foi criada uma tabela que integra os dados dos estados, incluindo as variáveis uf, população estimada, população estimada percentual, fator de correção, modalidade, ocupação, porte, número de Operações, Carteira Ativa e Ativo Problemático. A integração dos dados dos estados em uma única tabela no Power BI proporcionou uma análise mais organizada, concentrando-se apenas nas variáveis relevantes para o estudo em questão.
- Limpeza dos dados e preparação para exportação no OpenRefine: Utilizando o OpenRefine, foram realizadas as transformações necessárias, como a alteração do formato para camelCase e a preparação dos dados para exportação no formato GML. Essas ações

foram realizadas visando a padronização e compatibilidade dos dados para análises posteriores.

Durante esse processo, também foi necessário obter os dados relativos à quantidade de habitantes por Unidades de Federação, disponíveis no IBGE (2024), para fazer uma proporcionalidade demográfica das relações dos estados com os outros elementos da base em função do tamanho da população (Anexo 1).

4.4.4 Etapa 4: Modelo de Dados e Mapeamento dos Dados

O modelo de dados é um diagrama que representa com mais precisão e clareza os relacionamentos entre as entidades de dados, atributos e chaves. Ele tem a capacidade de mostrar em detalhes as variáveis de uma base de dados e os seus relacionamentos. Foi utilizada a ferramenta Power Bi para a diagramação do modelo de dados.

Utilizando-se do serviço de exportação do OpenRefine, foi realizado o mapeamento das variáveis selecionadas para o formato de redes complexas registrada como um arquivo do tipo GML. Essa exportação dos dados para o formato GML obedeceu a uma linguagem própria do OpenRefine, sendo construído um script para cada rede de informação gerada. Posteriormente, as redes em formato GML puderam ser abertas na ferramenta Gephi.

4.4.5 Etapa 5: Mineração e Visualização de Dados

Foram utilizadas técnicas de análise de redes complexas, formatações adequadas e projeções bipartidas por meio do software Gephi, munido do plugin MultiMode Networks, para a geração de redes de informação monopartidas por meio de projeções bipartidas. Além disso, para colaborar com a visualização de informação e complementar os resultados das redes de informação geradas pelo Gephi, foram elaborados gráficos e dashboards por meio da ferramenta Power BI Desktop.

4.4.6 Etapa 6: Análise e Interpretação dos Dados

Após analisar e interpretar os resultados obtidos na Etapa 5, chegou-se a conclusões acerca das possibilidades de contribuições sociais que os

relacionamentos revelados traziam, sem perder o foco na socialização da informação para benefício social.

5 RESULTADOS E DISCUSSÃO

O presente capítulo apresenta e discute os resultados obtidos mediante a análise da base de dados “SCR.data - Painel de Operações de Créditos” do Banco Central do Brasil, conforme objetivos propostos para a pesquisa.

Os resultados são apresentados de forma sintonizada com as etapas procedimentais elencadas na Seção 4.4 do capítulo dos procedimentos metodológicos. Dessa forma, ele é organizado em seções para tratar dos seguintes resultados discutidos à luz do referencial teórico: as variáveis selecionadas e modelo conceitual construído; o pré-processamento e mapeamentos dos dados para o formato de redes de informação; e os elementos da visualização de informação. Uma última Seção ainda faz algumas discussões mais gerais sobre os resultados obtidos na pesquisa.

5.1 Variáveis selecionadas e modelo conceitual

A Etapa 1, Seção 4.4.1, denominada de Investigação de Bases de Dados, foi realizada com o objetivo de selecionar a fonte de dados mais adequada para atender ao propósito da pesquisa. Considerando o objetivo geral de investigar variáveis presentes nas bases de dados abertas do Banco Central do Brasil relacionadas às operações de crédito, no contexto do Open Finance, a escolha criteriosa da base de dados foi fundamental para o desenvolvimento da pesquisa.

Conforme detalhado na Seção 4.3, a base de dados selecionada, o SCR.data - Painel de Operações de Créditos, foi reconhecida como uma fonte robusta e abrangente de informações, fornecendo uma visão detalhada das operações de crédito no Brasil. Abrangendo o período de junho de 2012 até a atualidade, esta base oferece dados abrangentes que englobam uma variedade de segmentos de clientes, modalidades de crédito e unidades da federação, proporcionando uma visão completa e detalhada do panorama das operações de crédito no país.

No contexto da base de dados escolhida, foram analisadas as variáveis da base de dados que tratam especificamente da classificação das operações de crédito agregadas, contemplando variáveis como Carteira Ativa 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, número e indexador das operações. Essas variáveis estão representadas na ontologia de referência de domínio da Figura 11, que representa o resultado da Etapa 2, Seção 4.4.2, denominada de Criação do Modelo Conceitual.

Foram empregadas 23 variáveis disponíveis na base de dados para a seleção. Destas, quatorze foram escolhidas devido à sua relevância para o problema de pesquisa em questão. A lista de variáveis selecionadas inclui:

- Ativo Problemático
- Carteira Ativa
- CNAE Seção
- CNAE Subseção
- Cliente
- Data Base
- Indexador
- Modalidade
- Número de Operações
- Ocupação
- Origem
- Porte
- TCB
- UF

As variáveis descartadas não representam de forma significativa o objeto de estudo e, de certa forma, já estão consolidadas em outras. Essa decisão foi tomada para simplificar a análise e aprimorar a eficácia da pesquisa, assegurando a integridade e a representatividade das informações pertinentes. As variáveis descartadas foram:

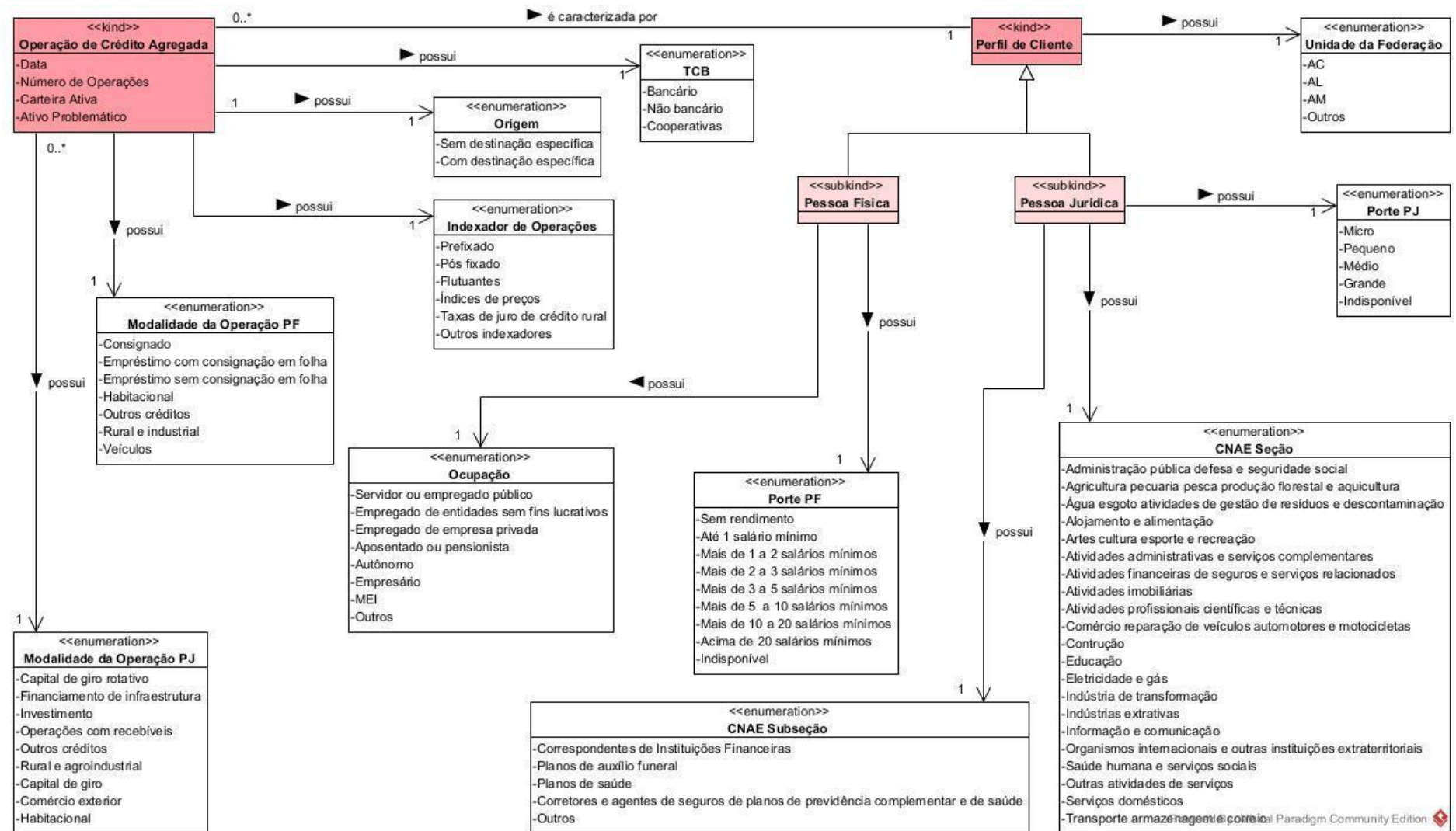
- Carteira inadimplida arrastada
- Valores a vencer até 90 dias
- Valores a vencer de 91 a 360 dias
- Valores a vencer de 361 até 1080 dias
- Valores a vencer de 1081 até 1800 dias
- Valores a vencer de 1801 até 5400 dias
- Valores a vencer acima de 5400 dias
- Valores vencidos acima de 15 dias

O descarte dessas variáveis trouxe um foco mais especializado e factível de ser operacionalizado considerando-se o tempo disponível para a pesquisa.

A Figura 11 apresenta uma modelagem conceitual da base de dados, a qual é representada por uma ontologia de referência de domínio fundamentada em UFO (Guizzardi, Falbo e Guizzardi, 2008), e diagramada na linguagem OntoUML. Esta linguagem é reconhecida como uma ferramenta eficaz para representar a realidade (Almeida, 2014), e neste contexto específico, foi empregada para facilitar o entendimento conceitual e relacional entre os elementos da base de dados de operações de crédito.

Segundo Zeng (2008) as ontologias são classificadas como bons modelos de relacionamento. Dessa forma, a ontologia da Figura 11 representa o relacionamento de duas classes principais: Operação de Crédito Agregada e Perfil de Cliente, sendo que essa última se deriva nas subclasses Pessoa Física e Pessoa Jurídica. Como apoio a vários atributos dessas classes, foram criadas 11 listas de termos, SOCs segundo Zeng (2008), representadas na modelagem conceitual da Figura 11 por enumerações, indicadas nas caixas com a marca <<enumeration>>.

Figura 11 - Modelagem conceitual por meio de uma ontologia de referência de domínio



Fonte: elaboração própria com apoio do software Visual Paradigm - versão com melhor definição disponível em: <https://github.com/infojuliana/ontologia/blob/main/Ontologia.jpg>

As operações de crédito agregadas são aquelas agrupadas com base em características específicas, como o perfil dos clientes que realizaram essas transações. Tais perfis podem abranger tanto pessoas físicas quanto jurídicas e estão relacionados a diferentes estados brasileiros. No contexto específico das pessoas físicas, que foi o foco da pesquisa, esses perfis podem ser delineados pelas seguintes características, apresentados na lista de termos <<enumeration>> e conectadas à classe Pessoa Física, Figura 11:

- **Ocupação:** Esta característica diz respeito à atividade profissional principal do cliente. Por exemplo, um cliente pode ser classificado como servidor público, autônomo, aposentado, entre outros. A ocupação oferece insights sobre a estabilidade financeira, padrões de renda e propensão ao risco do cliente.
- **Porte:** Refere-se ao tamanho da renda ou patrimônio líquido do cliente, categorizado em faixas de valores variando de 1 a 20 salários mínimos. O porte financeiro também pode servir como um indicador da capacidade de endividamento e da propensão ao consumo do cliente.

Quanto às pessoas jurídicas, representadas pela classe Pessoa Jurídica e os relacionamentos com as listas de termos Porte PJ, CNAE Seção e CNAE Subseção, na Figura 11, observa-se:

- **Porte da PJ:** Representa o tamanho das empresas, podendo ser categorizado em microempresa, pequena empresa, média empresa ou grande empresa.
- **CNAE Seção:** O CNAE (Classificação Nacional de Atividades Econômicas) é um sistema utilizado para classificar as atividades econômicas das empresas.

As seções do CNAE incluem:

- Administração Pública, Defesa e Seguridade Social
- Agricultura, Pecuária, Produção Florestal, Pesca, Agricultura, Água, Esgoto, Atividades de Gestão de Resíduos e Descontaminação
- Alojamento e Alimentação
- Artes, Cultura, Esporte e Recreação
- Atividades Administrativas e Serviços Complementares
- Atividades Financeiras, de Seguros e Serviços Relacionados
- Atividades Imobiliárias

- Atividades Profissionais, Científicas e Técnicas
- Comércio, Reparação de Veículos Automotores e Motocicletas
- Construção
- Educação
- Eletricidade e Gás, Indústria de Transformação
- Indústrias Extrativas
- Informação e Comunicação
- Organismos Internacionais e Outras Instituições Extraterritoriais
- Outras Atividades de Serviços
- Saúde Humana e Serviços de Assistência Social
- Serviços Domésticos
- Transporte, Armazenagem e Correio
- Essas seções do CNAE fornecem informações detalhadas sobre as atividades econômicas das empresas, auxiliando na compreensão dos perfis dos clientes e no direcionamento das análises de crédito.
- CNAE Subseção: Similar à variável "CNAE Seção", mas oferece uma classificação mais detalhada, indicando a subseção da atividade econômica da empresa, com cerca de 1200 opções.

Além das características individuais dos clientes, as operações de crédito agregadas também são definidas pelas modalidades de operação de crédito, que podem ser tanto para pessoa física quanto para pessoa jurídica. Essas modalidades abrangem uma variedade de produtos financeiros, conforme as listas de termos na modelagem conceitual da Figura 11:

- Modalidades de Operação de Crédito para Pessoa Física:
 - Cartão de Crédito
 - Empréstimo com Consignação em Folha
 - Empréstimo sem Consignação em Folha
 - Crédito Habitacional
 - Outros Créditos
 - Crédito Rural e Agroindustrial
 - Crédito para Veículos
- Modalidades de Operação de Crédito para Pessoa Jurídica:

- Capital de Giro
- Cheque Especial e Conta Garantida
- Comércio Exterior
- Financiamento de Infraestrutura/Desenvolvimento
- Crédito Habitacional
- Investimento
- Operações com Recebíveis
- Outros Créditos
- Crédito Rural e Agroindustrial

As condições de contratação, taxas de juros e demais aspectos podem variar de acordo com cada modalidade, conforme as listas de termos na modelagem conceitual da Figura 11.

- Origem dos Recursos:
 - Sem Destinação Específica: Indica que os recursos utilizados nas operações de crédito não têm uma destinação específica pré-determinada, permitindo maior flexibilidade ao cliente no uso dos recursos.
 - Com Destinação Específica: Indica que os recursos têm uma destinação pré-determinada, como financiamento de imóvel ou compra de veículo, com condições específicas relacionadas à finalidade do crédito.
- Tipo de Consolidado Bancário (TCB):
 - Bancário: Agrupa instituições financeiras de natureza bancária, como bancos comerciais e instituições de pagamento.
 - Não Bancário: Agrupa instituições financeiras não consideradas bancos, como cooperativas de crédito e outras entidades financeiras.

Além disso, as operações de crédito podem estar vinculadas a indexadores como taxa básica de juros (Selic) ou Índice de Preços ao Consumidor Amplo (IPCA), afetando as condições de reajuste das parcelas ao longo do tempo, conforme lista de termos Indexador de Operações na modelagem conceitual da Figura 11:

- Prefixado

- Pós-fixado
- Flutuante
- Outros Indexadores (como TCR/TRFC e índices de preços)

Adicionalmente, as informações das operações de crédito agregadas incluem dados como o número de operações, a carteira ativa e os ativos problemáticos, Figura 11:

- Número de Operações: Indica a quantidade total de operações de crédito consideradas em uma análise.
- Carteira Ativa: Refere-se ao conjunto total de operações de crédito ativas em um determinado momento.
- Ativo Problemático: Essa variável representa uma soma de diferentes tipos de créditos considerados problemáticos, incluindo créditos vencidos há mais de 15 dias e parcelas de créditos inadimplidos há mais de 90 dias.

É importante ressaltar que tanto as operações de crédito quanto os perfis de clientes podem ser classificados por estado, refletindo as características econômicas, sociais e demográficas locais e influenciando na oferta e demanda por crédito. Esse relacionamento é indicado na ontologia da Figura 11 entre as classes Perfil do Cliente e Unidade da Federação.

Conforme Hjørland (2018), citado na Seção 3.2.1, os dados nunca existem de forma independente de contexto. A coleta e o processamento de dados devem ocorrer com a compreensão dos interesses e objetivos que orientam sua captura, seleção e processamento, assumindo principalmente que possa haver conhecimentos implícitos. Noy (2001), citado na Seção 3.1.2, descreve uma ontologia como sendo um vocabulário comum que facilita a comunicação entre pesquisadores ao compartilhar informações dentro de um determinado domínio. Desenvolver uma ontologia pode ser motivado por diversas razões, incluindo facilitar o entendimento comum da estrutura da informação. A utilização da ontologia de referência de domínio da Figura 11 foi uma abordagem que trouxe solidez e clareza, permitindo uma modelagem semântica consistente dos dados. Essa abordagem facilitou a análise de insights e informações e a adequada identificação dos relacionamentos.

5.2. Pré-processamento, modelo de dados e mapeamento para rede de informação

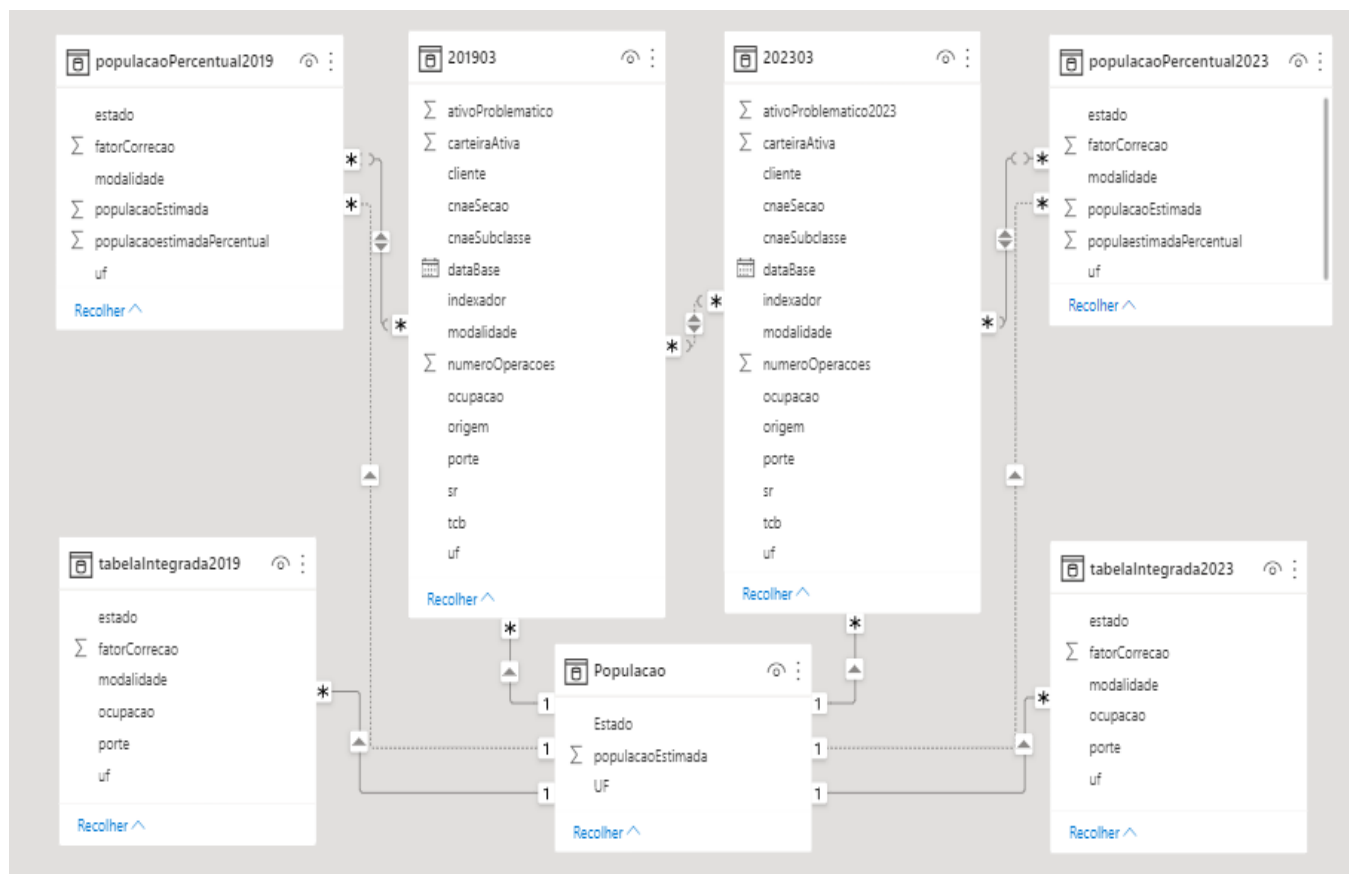
Na Etapa 3, conforme Seção 4.4.3, foram realizadas as atividades de preparação, limpeza e transformação dos dados, inicialmente utilizando o Power BI e posteriormente o OpenRefine. Executados com o objetivo de melhorar a qualidade e a usabilidade dos dados, tais processos foram essenciais para corrigir inconsistências, remover duplicatas e formatar os dados adequadamente.

O modelo de dados, Figura 12, construído baseado no modelo conceitual e na base de dados pré-processada, demonstra com mais clareza os relacionamentos estabelecidos e a criação de variáveis de dados para a pesquisa, destacando também as variáveis existentes em cada entidade. A seção 4.4.3 traz mais detalhes sobre a metodologia.

Nesse modelo é importante destacar:

- Tabela população
 - estado: nome dos estados por extenso
 - população estimada: valor da população por estado
 - uf: siglas dos estados
- Tabela população percentual
 - fator correção: $(1 - [\text{população estimada percentual}])$
 - população estimada percentual: porcentagem do total geral da população estimada
 - uf: siglas dos estados
- Tabela integrada
 - fator correção: $(1 - [\text{população estimada percentual}])$
 - modalidade: tipos de operação de crédito
 - ocupação: ocupação principal dos clientes PF
 - porte: faixa salarial dos clientes PF
 - uf: siglas dos estados

Figura 12 - Modelo de dados



Fonte: elaboração própria com apoio do software Power BI

Softwares de manipulação de dados são essenciais para o KDD, facilitando o processo ETL, que consome cerca de 80% do tempo do projeto (Mason; Patil, 2015), citado na Seção 3.2.3. Essa etapa foi fundamental para garantir a qualidade e confiabilidade dos dados utilizados na pesquisa. A preparação, limpeza e transformação dos dados são processos essenciais para preparar os dados de forma adequada antes de realizar qualquer tipo de análise ou interpretação. Isso garante que as conclusões obtidas sejam baseadas em informações precisas e confiáveis, possibilitando uma tomada de decisão informada.

Na Etapa 4, conforme a Seção 4.4.4 intitulada "Mapeamento dos Dados", o serviço de exportação do OpenRefine foi empregado para converter a base de dados relativa às variáveis selecionadas para o formato GML. Essa conversão para

o formato GML usou uma linguagem específica do OpenRefine, para a qual foi criado um script individual para cada rede de informação gerada.

Quadro 1 - Script do mapeamento dos dados para o formato de redes GML

```
graph [
directed 0

node [ id {{jsonize(cells.estado.value)}} categoria "estado" grupo "estado"
]
node [ id {{jsonize(cells.modalidade.value)}} categoria "modalidade" grupo
"condição" ]
node [ id {{jsonize(cells.ocupacao.value)}} categoria "ocupacao" grupo
"condição" ]
node [ id {{jsonize(cells.porte.value)}} categoria "porte" grupo "condição"
]

edge [ source {{jsonize(cells.modalidade.value)}} target
{{jsonize(cells.estado.value)}} weight
{{jsonize(cells.fatorCorrecao.value)}} ]

edge [ source {{jsonize(cells.ocupacao.value)}} target
{{jsonize(cells.estado.value)}} weight
{{jsonize(cells.fatorCorrecao.value)}} ]

edge [ source {{jsonize(cells.porte.value)}} target
{{jsonize(cells.estado.value)}} weight
{{jsonize(cells.fatorCorrecao.value)}} ]
```

Fonte: elaboração própria com apoio do software Open Refine

O script de mapeamento de dados do projeto oriundo do OpenRefine para a rede de informação, em formato de GML, é apresentado no Quadro 1. Ele foi desenvolvido com o objetivo de criar nós e conexões na rede de informação, utilizando as variáveis oriundas da base de dados pré-processada das operações de crédito. Por meio dessa linguagem, representada pelo script do Quadro 1, foi possível definir os nós que compõem a rede, utilizando-se o comando **node**. Esses nós representam diferentes entidades: estados, modalidades de operação de crédito, ocupações e portes. Cada nó possui uma categoria atribuída que indica o seu tipo: estado, modalidade, ocupação e porte. Os nós relacionados a estados têm um grupo atribuído como "estado", indicando que representam estados brasileiros. Os nós

relacionados a modalidades, ocupações e portes têm um grupo atribuído como "condição", indicando que representam diferentes condições relacionadas às operações de crédito.

Ainda no script do Quadro 1, o comando **edge** definiu as arestas que fazem a conexão entre os valores das entidades, nesse caso, entre estado e modalidades de operação de crédito, estados e ocupações, e estados e portes. Para cada uma dessas arestas foi definido um peso, por meio do comando **weight**. Esse peso representa o fator de correção calculado de forma proporcional pela população de cada estado brasileiro, indicado na tabela **Populacao** do modelo de dados, Figura 12.

Os nós e arestas estão agrupados de acordo com suas categorias e grupos para facilitar a interpretação do gráfico. As entidades escolhidas (estados, modalidades, ocupações e portes) estão relacionadas ao contexto de operações de crédito, representando análises sobre as condições de crédito em diferentes estados com base em diferentes variáveis.

A seleção das entidades, como estados, modalidades de operação de crédito, ocupações e portes, foi de grande importância para entender e analisar o contexto das operações de crédito de forma abrangente:

- **Representatividade Regional:** Os estados brasileiros fornecem uma divisão geográfica que permite entender como as operações de crédito são distribuídas regionalmente. Isso é crucial para identificar disparidades regionais e direcionar políticas ou estratégias específicas para cada região.
- **Diversidade de Modalidades:** As diferentes modalidades de operação de crédito oferecem *insights* sobre os diversos tipos de serviços financeiros disponíveis e como eles são utilizados em diferentes contextos e regiões.
- **Segmentação Ocupacional:** As ocupações dos tomadores de crédito podem revelar informações sobre os setores econômicos predominantes em cada região, bem como sobre a capacidade de pagamento e o risco associado a diferentes grupos profissionais.
- **Segmentação Socioeconômica:** O porte das pessoas físicas que buscam crédito pode fornecer *insights* sobre sua capacidade

financeira, comportamento de gastos e padrões de consumo. Isso é crucial para entender as necessidades e demandas dos diferentes estratos sociais e direcionar produtos financeiros adequados a cada grupo.

5.3. Visualização da informação

A presente seção mostra resultados relativos à Etapa 5 dos procedimentos metodológicos, Subseção 4.4.5, denominada Mineração e Visualização de Dados.

Conforme citado na Subseção 3.2.4, a atividade de visualização da informação facilita a interpretação, pois privilegia elementos de representação gráfica (Rautenberg; Carmo, 2020), aumentando a compreensão dos conhecimentos inerentes aos dados (Dias, 2007). Dessa forma, os resultados interpretativos obtidos e apresentados na presente subseção são embasados e auxiliados pelos seguintes elementos visuais:

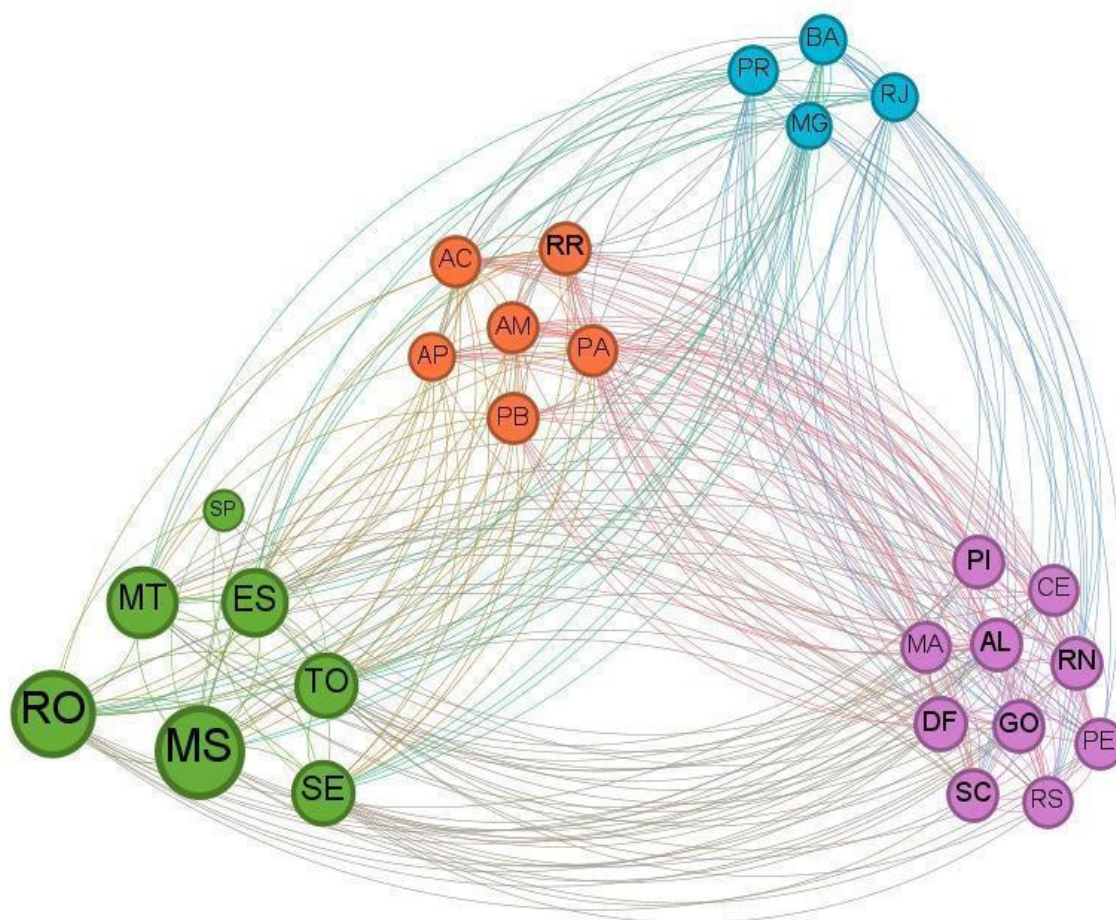
- Redes de informação monopartidas de estados brasileiros, referentes aos anos de 2019 e 2023, representados respectivamente nas Figuras 12 e 13;
- Gráfico comparativo entre grupos de estados dos anos de 2019 e 2023, no Gráfico 1;
- Gráfico comparativo do percentual de número de operações, ativo problemático e carteira ativa por modalidade de crédito, no Gráfico 2;
- Gráfico dos ativos problemáticos por ocupação, no Gráfico 3.

Conforme descrito na Seção 4.3, foram utilizadas duas bases de dados. A primeira base de registros de crédito refere-se ao mês de março de 2019, enquanto a segunda base de registros de crédito é relativa ao mês de março de 2023. As bases foram empregadas com o propósito de realizar uma análise comparativa temporal, abrangendo o período prévio ao advento do Open Finance e o mês mais recente disponível na base de dados aberta do Banco Central. Foram realizadas análises do grupo em relação ao todo, a centralidade de grau do nó, com relação à localização geográfica dos clusters, a estrutura de clusterização analisando características semelhantes e a consistência temática dos grupos. Foi possível identificar mudanças significativas nos padrões de relacionamento ao longo do

tempo, o que pode fornecer insights sobre tendências e evoluções no cenário estudado.

Uma rede bipartida foi construída inicialmente com os estados agrupados de acordo com as modalidades de crédito por ocupação e porte de todas as operações de crédito. Posteriormente, uma projeção bipartida foi realizada para remover as variáveis independentes, ou seja, as variáveis explicativas - modalidade de crédito, ocupação e porte - permanecendo apenas a variável dependente, ou seja, os estados (as relações entre eles). As variáveis independentes representam as conexões de modalidade, porte e ocupação com os estados. Após essa etapa, foi aplicada uma formatação nos nós da rede considerando o grau ponderado e a modularidade. Assim, obteve-se uma rede monopartida de estados, na qual foram identificadas 4 comunidades com base no grau de modularidade. Essas etapas foram aplicadas na base de dados de março de 2019 e de março de 2023, apresentadas respectivamente por meio das Figura 12 e Figura 13.

Figura 12 - rede monopartida de estados, por modalidade de crédito, ocupação e porte salarial de 2019/03



Fonte: elaboração própria.

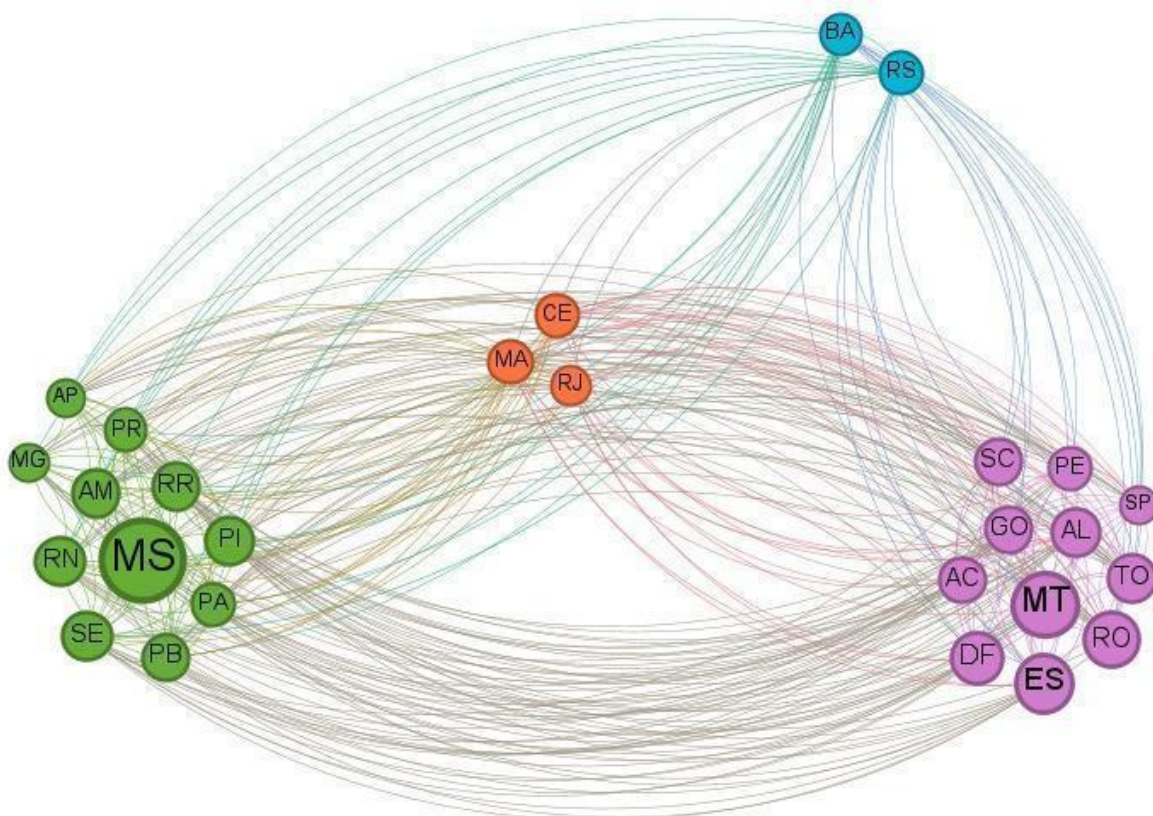
Na análise visual sobre a rede da Figura 12, nota-se que a comunidade verde, abrangendo os estados de RO, MS, SE, MT, ES, TO e SP, se evidencia de maneira significativa especialmente nos estados de RO e MS, com exceção do estado de SP. Importante ressaltar que as proporções foram ajustadas conforme a população de cada estado (Anexo 1).

Além disso, outra comunidade que se destaca, representada na cor roxa, engloba uma maior quantidade de estados e, portanto, exibe uma maior densidade ou número de conexões com outras regiões. Esta comunidade abrange os estados: PI, CE, AL, RN, MA, DF, GO, SC, RS e PE.

Além das comunidades verde e roxa, foi possível identificar outras duas comunidades na rede. A comunidade azul abrangendo os estados de MG, PR, BA e RJ, e a comunidade laranja composta por AC, RR, AM, PA, PB e AP. Embora não

seja possível fazer observações significativas sobre essas comunidades, em especial, foi possível notar que na comunidade azul existe uma diversidade geográfica considerável em relação à região em que estão situados os estados.

Figura 13 - rede monopartida de estados, por modalidade de crédito, ocupação e porte salarial de 2023/03



Fonte: elaboração própria.

Na análise visual da rede representada na Figura 13, é notável que a comunidade verde mantém sua proeminência com relação à rede de 2019 apresentada na Figura 12, abrangendo os estados MS, SE, PB, PA, PI, RN, RR, AM, PR, MG e AP. Especialmente os estados MS e SE surgem novamente na rede, e o MS, como relatado na Figura 12, continua a se destacar pela sua maior centralidade de grau. Isso sugere uma continuidade no padrão de relacionamentos desses estados em comparação com a análise visual anterior da Figura 12. Além disso, observa-se uma integração de outros estados na comunidade, aumentando seu

tamanho. Essa constatação sugere consistência na composição e destaque dessa comunidade em comparação com os dados anteriores da Figura 12.

Na comunidade roxa, composta por MT, ES, RO, TO, DF, AC, GO, AL, SP, PE e SC, os estados de AL, DF, GO, PE e SC reaparecem, indicando uma considerável consistência temática, enquanto os novatos MT e ES se destacam em termos de centralidade de grau. É importante ressaltar que SP continua sendo o estado de menor tamanho em comparação com os outros nós. Em relação ao tamanho do grupo, permanece quase o mesmo, em termos de quantidade de estados.

A comunidade laranja, que inclui RJ, CE e MA, surge com uma nova formatação em relação à Figura 12, e seu tamanho é reduzido pela metade. Enquanto isso, a comunidade azul, representada por BA e RS, também perde membros na Figura 13 em comparação com a Figura 12, mas mantém a consistência com pelo menos o estado da BA.

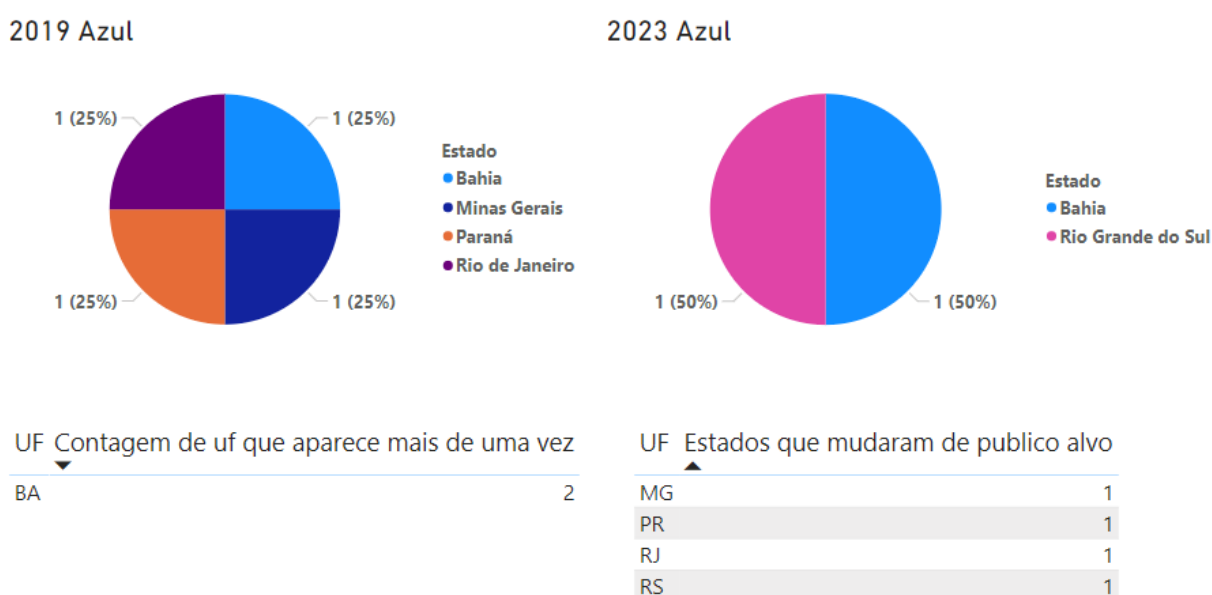
Conforme delineado na Seção 3.3.2, a análise da topologia de uma rede foi fundamental para identificar a importância de um nó ou grupo de nós dentro da rede (Newman, 2010). Paralelamente, a mineração de dados desempenhou um papel crucial na pesquisa, permitindo analisar um conjunto extenso de dados para revelar relações não previstas e torná-las compreensíveis. Seguindo as etapas propostas por Carvalho (2005), como mencionado na Seção 3.2.3, a mineração de dados facilitou a identificação de problemas, a descoberta de novas relações, a análise humana das relações descobertas, a utilização racional dessas relações e a avaliação dos resultados obtidos, fornecendo uma base sólida para a investigação e uma possível tomada de decisões. Além disso, conforme evidenciado na Seção 3.2.4, a visualização da informação desempenhou um papel crucial ao melhorar a compreensão do usuário sobre o conhecimento, promovendo, assim, a inclusão informacional (Dias, 2007). Isso ocorre porque a visualização da informação transforma dados complexos e abstratos em representações gráficas e visuais intuitivas, facilitando a assimilação de conceitos e padrões que seriam difíceis de entender apenas através de texto ou números. A visualização permite também que pessoas com diferentes níveis de alfabetização e habilidades cognitivas acessem e compreendam a informação de maneira mais eficiente. Por exemplo, gráficos, mapas e infográficos podem tornar evidente tendências e correlações que não seriam facilmente perceptíveis em tabelas ou descrições textuais, reduzindo barreiras cognitivas e promovendo um acesso mais equitativo ao conhecimento.

Dessa forma, a visualização da informação não só melhora a compreensão individual, mas também contribui para uma inclusão mais ampla e democrática no acesso à informação.

Na Seção 3.1, ressalta-se a ausência de um consenso na literatura sobre a hierarquia piramidal entre dados, informações e conhecimento. Por exemplo, Fricke (2019) propõe um processo interativo entre dados e conhecimento, sugerindo uma influência mútua entre ambos. Essa perspectiva enfatiza a dinâmica e complexidade da relação entre esses conceitos, destacando a necessidade de considerar diversas abordagens na compreensão dessa interação. A interação entre dados e sua representação visual reforça a ideia de um processo contínuo de retroalimentação e aprimoramento mútuo, ampliando as possibilidades de interpretação e análise. Esse fenômeno foi observado de maneira específica na análise temporal das duas redes, onde a evolução das relações entre os dados e sua visualização destacou a importância dessa interdependência para uma compreensão aprofundada dos fenômenos estudados.

Na Etapa 6, conforme descrito na Seção 4.4.6, destinada à Análise e Interpretação dos Dados, foram examinados e interpretados os resultados obtidos na Etapa 5. A partir dessa análise, foram extraídas conclusões sobre as possíveis contribuições sociais dos relacionamentos revelados, mantendo sempre o foco na socialização da informação para benefício social.

Gráfico 1 - Comparação entre grupos de estados de 2019 e 2023 (Grupo Azul)



Fonte: elaboração própria.

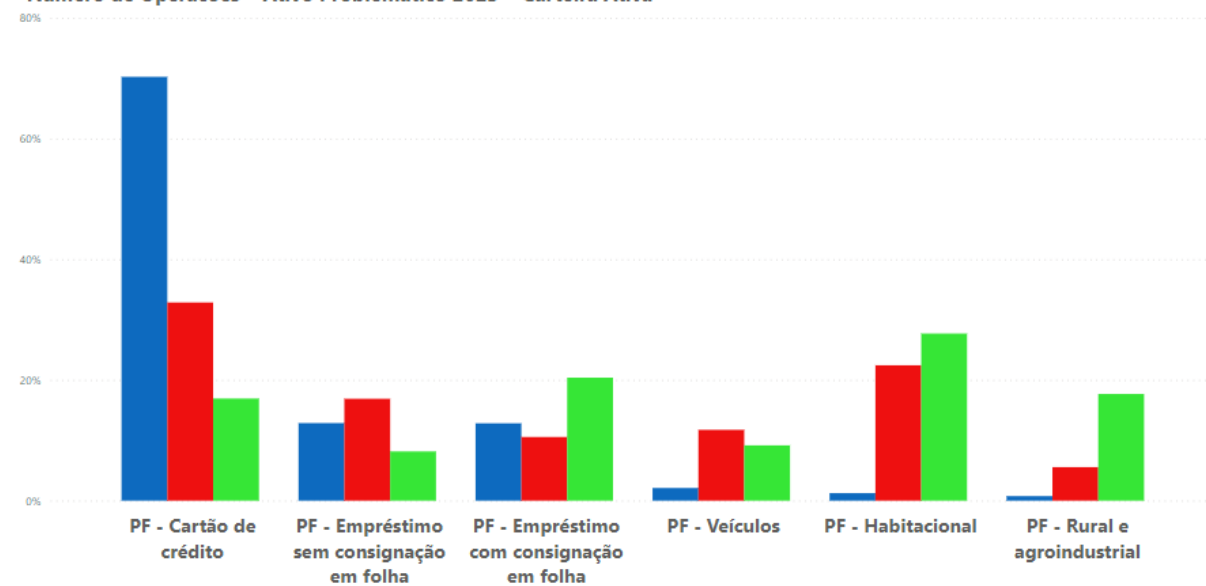
Notou-se que em certos grupos os estados permaneceram consistentes em seus respectivos grupos, em especial na Comunidade Azul, o estado da BA permaneceu constante, conforme demonstrado no Gráfico 1. Essa análise foi realizada em todos os outros grupos para verificar a consistência dos estados em relação ao seu público-alvo. Na comunidade verde, os estados MS e SE mantiveram sua presença. No grupo roxo, que apresentou o maior número de estados reincidentes, os estados AL, DF, GO, PE e SC repetiram-se e permaneceram consistentes. No entanto, na comunidade laranja, não houve estados reincidentes.

É interessante observar que a região não parece ter exercido uma influência direta nessas relações. Por outro lado, a mudança nos estados que migraram de grupo, sugere uma mudança em seu público-alvo a partir de 2019. A maioria dos estados mudando de grupo sugere uma dinâmica no mercado de crédito como um todo, que pode ser considerado volátil. Assim como no mercado financeiro em geral, o mercado de crédito está sujeito a mudanças rápidas e imprevisíveis devido a uma variedade de fatores, como condições econômicas, políticas governamentais, taxas de juros, comportamento dos consumidores e até mesmo eventos externos. Esses insights destacam a complexidade das dinâmicas de mercado e a importância de uma análise abrangente para compreender as relações entre os clientes e as operações de crédito.

Gráfico 2 - Percentual de Número de Operações, Ativo Problemático e Carteira Ativa por Modalidade de Crédito

Percentual de N° de operações, Ativo problemático e Carteira ativa por Modalidade de Crédito

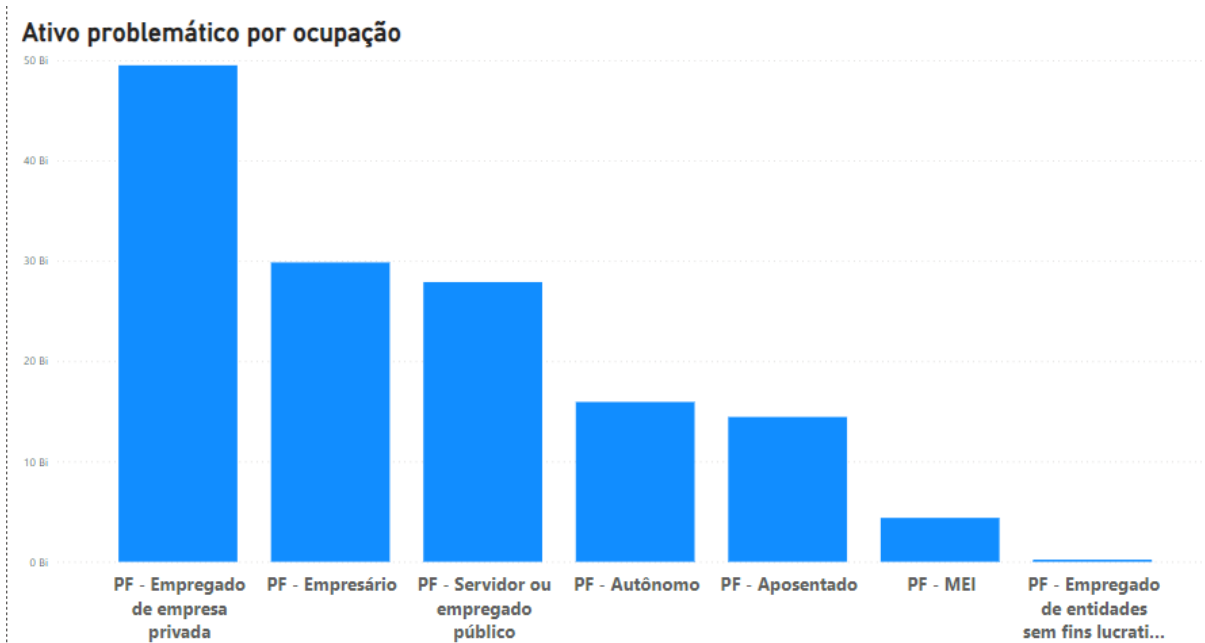
● Número de Operacoes ● Ativo Problematico 2023 ● Carteira Ativa



Fonte: elaboração própria.

No Gráfico 2, três variáveis são destacadas: o número de operações de crédito, a carteira ativa e o ativo problemático da base de dados de 2023. O número de operações representa quantas transações de crédito foram realizadas. A carteira ativa é a soma dos valores devidos, tanto a vencer quanto vencidos, para diferentes modalidades de crédito. Já o ativo problemático engloba créditos em situação de risco, operações reestruturadas e dívidas em atraso. Em modalidades de crédito onde o número de operações se destaca, isso indica pulverização ou diversificação na modalidade. Quando o ativo problemático é maior, significa que há problemas com os pagamentos, levando a atrasos nas operações e, possivelmente, juros mais altos (Giolo, 2009). Quando a carteira ativa é maior do que o número de operações, isso sugere uma concentração de crédito. Essas informações são cruciais para entender o cenário de crédito e podem influenciar na definição das taxas de juros aplicadas.

Gráfico 3 - Ativo problemático por ocupação



Fonte: elaboração própria.

No Gráfico 3, foi destacada a relação entre o ativo problemático e a ocupação da população pessoa física, com base nos dados de 2023. Observou-se que a população com maior atraso de pagamento é composta principalmente por empregados de empresas privadas, seguidos por empresários - pessoas físicas e servidores públicos. Esta análise evidencia a importância de considerar o tipo de ocupação ao avaliar o risco de inadimplência e a necessidade de implementar estratégias específicas para lidar com as dificuldades financeiras enfrentadas por esses grupos.

Ao analisar de maneira geral os diversos objetos de visualização apresentados tanto nas redes de informação (Figuras 12 e 13), quanto nos gráficos para complementação do entendimento do cenário (Gráficos 1, 2 e 3), pode-se obter *insights* importantes para a escolha de produtos de crédito no Open Finance:

- Identificar quais estados possuem comportamento semelhante em relação a produtos de crédito.
- Observar quais estados apresentam consistência ao longo do tempo, indicando solidez no comportamento.
- Avaliar quais produtos são mais suscetíveis ao aumento da taxa de juros.

- Analisar quais ocupações têm maior número de ativos problemáticos, o que pode resultar em maiores taxas de juros devido ao histórico de inadimplência

5.4. Discussão geral

A abordagem de ARS da Ciência da Informação pode colaborar significativamente com temas como Fluxo de Informação, Gestão do Conhecimento e Tomada de Decisão, como mencionado por Matheus (2006). Isso pode ser alinhado aos resultados da pesquisa no contexto do Open Finance, onde a análise dos dados abertos do Banco Central fornece *insights* valiosos sobre os padrões de uso dos produtos financeiros pela população, colaborando assim com a gestão do conhecimento e tomada de decisão com relação a produtos de crédito por parte dos consumidores.

Conforme Subseção 3.1, a organização do conhecimento em seu sentido mais restrito faz uso dos Sistemas de Organização do Conhecimento (SOC), que estão organizados de acordo com suas estruturas e principais funções (Hjørland, 2008). Ontologias são classificadas como modelos de relacionamento, apresentando a característica única de apresentar propriedades de cada classe dentro de sua estrutura classificatória, reconhecidamente utilizadas como instrumento conceitual (Zeng, 2008), sendo propostas como um método útil para representação de uma realidade (Almeida, 2014). Nesse intuito foi apresentada uma ontologia da base de dados mencionada, buscando representar a realidade e com isso facilitar a compreensão dos dados e suas relações.

Tratando ainda desse mesmo contexto, e no objetivo de garantir acesso a produtos de crédito de qualidade, a modelagem, a mineração e a visualização da informação, que são áreas cruciais na representação de dados, contribuíram significativamente na compreensão dos dados, auxiliando na obtenção de *insights* (Hand et al., 2001) com vistas a melhorar as condições de crédito e impulsionar o desenvolvimento econômico sustentável. Sendo assim, é interessante notar como os SOC, em especial a ontologia de referência, apresentada na Figura 11 e o modelo de dados da Figura 12, tiveram um papel crucial na compreensão dos padrões de dados evidenciando a interação dos clientes e o mercado de produtos de crédito.

Outro conceito abordado pela presente pesquisa é definido como *data literacy* (Belluzzo, 2004), tratando não só de habilidades de apropriação da informação

(Rasteli, Cavalcante, 2014), mas de aprendizados mais técnicos usados em análise de dados (Schield, 2004), conforme subseção 3.2.1. No contexto da presente pesquisa, esse entendimento permitiu um amadurecimento da competência, por parte da autora desta pesquisa, no uso da informação de maneira crítica e técnica.

Além disso, a mineração de dados proposta trabalha em conjunto à análise de redes complexas, que tornou-se importante campo de pesquisa pela possibilidade de poder representar diversos sistemas (Costa, 2018).

Condensando resultados, pode-se observar que, apesar da volatilidade do mercado de crédito, certos estados mantiveram uma consistência notável em relação ao seu mercado e público-alvo. Essa consistência sugere uma maior capacidade desses estados em oferecer taxas de juros competitivas, devido ao seu conhecimento aprofundado de nicho de mercado. Em um ambiente caracterizado por mudanças frequentes, como o Open Finance, o acesso a essas informações desempenha um papel crucial, pois a condição do crédito oferecido é influenciada diretamente pelo risco percebido, o que, por sua vez, afeta as taxas de juros aplicadas. Assim, essas informações tornam-se fundamentais para decisões financeiras mais bem informadas como destacado na Seção 3.1.1 por Valentim (2002).

A análise dos dados abertos do Banco Central, no contexto do Open Finance, forneceu também *insights* valiosos sobre os padrões de uso dos produtos financeiros pela população, e de como diferentes grupos de clientes interagem com o mercado de crédito. Essas informações desempenham um papel crucial no desenvolvimento de políticas públicas mais inclusivas. Por exemplo, ao identificar comportamentos de pagamento ou necessidades de crédito específicas em determinadas regiões ou grupos semelhantes, as autoridades podem direcionar políticas de incentivo fiscal para facilitar o acesso ao crédito. Isso torna-se ainda mais relevante considerando a diversidade de necessidades e perfis de tomadores, conhecida como pulverização do crédito. Ao reconhecer o impacto das altas taxas de juros enfrentadas pela população menos favorecida ou com dificuldades de pagamento, as políticas públicas podem ser ajustadas para mitigar esses efeitos, criando condições mais justas e acessíveis para todos os cidadãos. Fundamentando suas decisões em análises informadas, as políticas públicas têm o potencial de reduzir as disparidades sociais e fomentar um sistema financeiro mais equitativo e inclusivo.

A disseminação da informação desempenha um papel essencial em várias

esferas do conhecimento. Considerando que toda prática depende do conhecimento como base fundamental (Christovão; Braga, 1997), pode-se dizer que orientações sobre produtos de crédito possibilitam a oferta de produtos de crédito mais acessíveis, levando a uma maior adesão de produtos financeiros pelos usuários, com custos reduzidos.

Embora os dados tenham por si só um valor econômico significativo, é essencial não apenas ter acesso a eles, mas também saber como utilizá-los de forma eficaz. Na justificativa dos resultados, pode-se destacar o argumento de que proporcionar acesso ao crédito é crucial para a melhoria sustentável das condições de vida, conforme mencionado por Paulo Picchetti, Diretor de Assuntos Internacionais e de Gestão de Riscos Corporativos do BC, durante uma coletiva de imprensa no âmbito do G20. Ele enfatizou que a inclusão financeira deve abordar tanto a quantidade quanto a qualidade, visando promover o crescimento das pessoas (Agência Brasil, 2020).

Contrastando os resultados obtidos com as pesquisas mostradas no estado da arte, Capítulo 2, pôde-se observar adjacências mais proeminentes com os trabalhos relativos à análise de redes de relações, como o trabalho de Parreiras (2021) que propôs a ARS de forma exploratória para compreender as redes entre os diretores e entre as instituições que atuaram no sistema financeiro brasileiro, utilizando assim a ARS de maneira exploratória para mapear e analisar as redes de relações do Sistema Financeiro Brasileiro. Nesse mesmo tema de análise de redes de relações, Borges (2022) analisou as características das cooperativas de crédito no Brasil. Ele verificou a existência de relações entre cooperativas e municípios. Também foi utilizado o software Gephi na confecção do grafo bipartido. Os resultados mostraram que as cooperativas dão preferência em instalar-se em municípios já atendidos pelo cooperativismo, elevando o grau na rede complexa. Sendo assim, o presente trabalho trouxe resultados semelhantes a essas pesquisas no sentido de avaliar as redes de relações entre os produtos de crédito e as variáveis selecionadas e principalmente trouxe também a questão da territorialidade como objeto de estudo.

Com relação a aspectos observados nas pesquisas do estado da arte e tratados na presente pesquisa, observou-se que embora os artigos analisados no estado da arte tenham trazido a análise de importantes aspectos do sistema financeiro, a investigação realizada nas pesquisas evidenciou em alguns trabalhos a

necessidade de um tratamento mais aprofundado com relação ao tratamento dos dados e as metodologias e ferramentas utilizadas na análise de redes de informação. Foi verificado também que são poucos os trabalhos que trataram especificamente da análise de redes de informação que objetivaram trazer alguma informação diretamente associada a benefícios sociais.

6 CONSIDERAÇÕES FINAIS

Quanto aos objetivos específicos elencados na pesquisa, Seção 1.3, o objetivo específico (i) analisar à luz da literatura da Ciência da Informação os conceitos de visualização de informação, modelagem conceitual, análise de redes complexas e Open Finance foi cumprido de forma abrangente em diferentes seções do trabalho:

- No Capítulo 3.2, intitulado Ciência de Dados, na seção 3.2.4, abordou-se o conceito de visualização da informação, explorando sua relação com outras áreas e aplicações dentro do contexto da Ciência da Informação.
- A modelagem conceitual foi discutida na seção 3.1.3, intitulada Ontologias, onde se explorou seu uso para representar conceitos e relações em domínios específicos. Ela foi aplicada de forma satisfatória para representar o relacionamento das variáveis da base de dados, evidenciando a importância reconhecida das ontologias como instrumento de modelagem de dados, enfatizando que na Ciência da Informação, elas desempenham um papel preponderante na área de representação da informação, visto que são capazes de representar todo um domínio.
- As seções 3.3.1 (Redes Complexas) e 3.3.2 (Análise de Redes de Informação) abordaram o tema da análise de redes complexas, discutindo suas características, técnicas e métodos de análise aplicados à Ciência da Informação.
- Por fim, a seção 3.4 tratou do surgimento do Open Finance, seu funcionamento, tendências e sua relação com a Ciência da Informação, atendendo ao objetivo proposto de analisar o conceito à luz da literatura relevante.

O objetivo específico (ii) investigar bases de dados abertas do Banco Central do Brasil sobre operações de crédito no contexto do Open Finance e selecionar variáveis relevantes para a pesquisa, foi cumprido conforme a etapa procedimental descrita na seção 4.4.1 e a Subseção de resultados 5.1. Durante a etapa de investigação de bases de dados, houve uma seleção criteriosa da fonte de dados. A base escolhida, o SCR.data - Painel de Operações de Crédito, foi reconhecida como

uma fonte robusta e abrangente, fornecendo uma visão detalhada das operações de crédito no Brasil ao longo do tempo. Para a análise comparativa temporal, foram utilizadas duas bases de dados: uma referente ao mês de março de 2019 e outra referente ao mês de março de 2023. Essas bases permitiram analisar o período prévio ao advento do Open Finance, e de um momento mais recente, correspondente ao mesmo mês, disponível na base de dados. Foram selecionadas variáveis, conforme Subseção 5.1, e essas forneceram uma base sólida para análises posteriores no contexto do Open Finance.

O objetivo específico (iii) criar um modelo conceitual para representar e facilitar o entendimento do contexto que envolve as variáveis selecionadas no objetivo específico (ii), foi cumprido conforme demonstrado na subseção 5.1. Foi apresentada uma ontologia de referência de domínio, que representa de forma estruturada as variáveis selecionadas no objetivo específico (ii). A Figura 11 ilustra esse modelo conceitual, fornecendo uma visualização clara das relações entre as variáveis e seu contexto no domínio das operações de crédito.

O objetivo específico (iv) mapear os dados referentes às variáveis selecionadas no objetivo específico (ii) e modeladas no objetivo específico (iii) para o formato de redes de informação, foi alcançado conforme evidenciado na subseção 4.4.4. Nesta etapa, intitulada Mapeamento dos Dados, utilizou-se o serviço de exportação do OpenRefine para realizar o mapeamento das variáveis selecionadas para o formato de redes complexas. O script utilizado, juntamente com sua explicação, está apresentado na subseção 5.2, no Quadro 1 - Script do mapeamento dos dados para o formato de redes GML.

O objetivo específico (v) aplicar técnicas de análise de redes e visualização de informação para revelar relacionamentos relevantes entre as variáveis selecionadas no objetivo específico (ii) e modeladas no objetivo específico (iii) sobre a rede de informação criada no objetivo específico (iv), foi cumprido conforme relatado na Subseção 4.4.5 e Subseção 5.3. Nesta etapa, intitulada Mineração e Visualização de Dados, foram utilizadas técnicas de análise de redes complexas, formatações adequadas e projeções bipartidas por meio do software Gephi, munido do plugin MultiMode Networks, para a geração de redes de informação monopartidas por meio de projeções bipartidas. Além disso, para auxiliar na visualização de informação e complementar os resultados das redes de informação geradas pelo Gephi, foram elaborados gráficos e *dashboards* por meio da ferramenta Power BI

Desktop. Os elementos, resultantes de mapeamentos realizados sobre os dados e a modelagem conceitual, foram apresentados na subseção 5.3, denominada Visualização da Informação.

O objetivo específico (vi) inferir possíveis contribuições sociais por meio da socialização da informação, que os relacionamentos revelados no objetivo específico (iv) poderão fomentar, foi cumprido conforme mencionado na Subseção 4.4.6 e Subseção 5.4 que evidenciou questões sociais e financeiras que podem ser alcançadas por meio do estudo apresentado.

Portanto, o objetivo geral de investigar variáveis presentes nas bases de dados abertas do Banco Central do Brasil, relacionadas às operações de crédito, no contexto do Open Finance, para a revelação de insights relevantes, considerando possibilidades de socialização da informação para benefício social, foi cumprido. Reunindo os resultados dos objetivos específicos, foi possível alcançar o objetivo geral e trazer informações significativas para contribuição social relativas ao ambiente de mudanças trazido pelo Open Finance.

Conforme evidenciado na Discussão Geral, Seção 5.4, a abordagem de análise de redes pode colaborar significativamente com o ambiente de gestão do conhecimento e tomada de decisão. No cerne da pesquisa encontra-se o usuário, capacitado em termos de análise de dados, enriquecendo a capacidade de lidar com informações de maneira crítica e técnica. Assim, a presente pesquisa traz reflexões sobre possíveis benefícios sociais advindos não somente da revelação dos dados em questão, mas do conhecimento e da utilização dos mesmos como fonte de informação relevante. Dessa forma, a pesquisa evidencia na análise de dados uma ambiência social e sustentável, que possibilita o fomento da inclusão social e capacitando os indivíduos a serem protagonistas de suas próprias escolhas financeiras.

Outro aspecto em destaque da pesquisa foi o uso de ontologias para modelagem em busca de uma melhor compreensão do contexto do ambiente financeiro, impulsionado pelas mudanças promovidas pelo Open Finance. Conforme já relatado na subseção 3.4.3, denominada Open Finance na Ciência da Informação, um ponto importante que conta com a contribuição da Ciência da Informação é sobre a especificação de informações sem ambiguidades em busca da interoperabilidade entre sistemas, e a solução pode estar na representação da informação por meio das ontologias. Portanto, como continuidade do trabalho, os estudos relativos à

interoperabilidade no Open Finance, surgem como uma opção promissora.

REFERÊNCIAS

AALST, Wil Van der. Data Science in Action. Em: AALST, Wil Van der (org.). Process Mining. Berlin, Heidelberg: Springer, 2016. p. 3–23. Disponível em: https://doi.org/10.1007/978-3-662-49851-4_1. Acesso em: 3 jul. 2023.

ALMEIDA, Maurício Barcellos. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. Perspectivas em Ciência da Informação, v. 19, p. 242-258, 2014.

ALVARENGA, Lídia. Representação do conhecimento da perspectiva da Ciência da Informação em tempo e espaço digitais. Enc. Bibli: R. Eletr. Bilbiotecon. Ci. Inf., Florianópolis, v. 8, n. 15, p. 18-40, 2003.

ANDERSON, Carl. Creating a data-driven organization: Practical advice from the trenches. " O'Reilly Media, Inc.", 2015.

Apresentação do Presidente Roberto Campos Neto no Fórum Brasileiro de Inteligência Artificial. Disponível em: <https://www.bcb.gov.br/detalhenoticia/17957/nota>. Acesso em: 23 ago. 2023.

ARAÚJO, C. A. Á. O que é Ciência da Informação? **Informação & informação**, v. 19, n. 1, p. 01, 2013.

ARAÚJO, Carlos Alberto Ávila. O que é Ciência da Informação. Belo Horizonte: KMA, 2018.

ARAÚJO, Carlos Alberto Ávila. Movimentos epistemológicos da Ciência da Informação. Códices, Bogotá, v. 14, n. 1, p. 61-78, enero-junio. 2018. Disponível em: <https://cnb.gov.co/codices/online/Vol14-2018/VI.pdf>. Acesso em: 3 mar. 2023.

Área do Desenvolvedor - Open Finance Brasil - Área do Desenvolvedor.

Disponível em: <<https://openfinancebrasil.atlassian.net/wiki/spaces/OF/overview>>.

Acesso em: 23 jan. 2024.

BALBINOTTI, Stheve et al. Competência em dados: uma necessidade contemporânea para pesquisadores e para a sociedade. *AtoZ: novas práticas em informação e conhecimento*, v. 11, p. 1-11, 2022.

BARABÁSI, Albert-László. Network science: graph theory. Web page: Barabasi site, 2014. Disponível em: <http://barabasi.com/f/625.pdf>. Acesso em: 30 abr. 2023.

BARITÉ, Mario. Sistemas de organização do conhecimento: uma tipologia atualizada. *Informação & Informação*, [S. l.], v. 16, n. 2, p. 122–139, 2011. DOI: 10.5433/1981-8920.2011v16n2p122. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/9952>. Acesso em: 23 mar. 2023.

BECKER, João Luiz; DINIZ, Eduardo Henrique. Como disseminar a linguagem dos dados. *GV-EXECUTIVO*, v. 18, n. 4, p. 12-15, 2019.

BELLUZZO, Regina Célia Baptista. Formação contínua de professores do ensino fundamental sob a ótica do desenvolvimento da information literacy, competência indispensável ao acesso à informação e geração do conhecimento. *Transinformação*, v. 16, p. 17-32, 2004.

BORGATTI, S. P.; FOSTER, P. C. The network paradigm in organizational research: a review and typology. *Journal of Management*, Miami, v. 29, n. 6, p.991-1013, 2003.

BORGATTI, Stephen P.; HALGIN, Daniel S. Analyzing affiliation networks. Em: SCOTT, John; CARRINGTON, Peter (eds.). *The SAGE Handbook of Social Network Analysis*. London, United Kingdom: SAGE Publications Ltd, 2014. p. 417–433. Disponível em: <https://methods>.

sagepub.com/book/the-sage-handbook-of-social-network-analysis/n28.xml. Acesso em: 3 jul. 2023.

BORKO, Harold. Ciência da Informação: o que é isto. *American Documentation*, v. 19, n. 1, p. 3-5, 1968.

BÖRNER, K.; BUECKLE, A.; GINDA, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 6, p. 1857–1864, 2019.

BRITO, Luciana et al. Workshop Investigações em Interação Humano-Dados–WIDE. In: *Anais Estendidos do XXI Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais*. SBC, 2022. p. 5-8.

BURKHARD, Remo Aslak. Towards a framework and a model for knowledge visualization: synergies between information and knowledge visualization. Em: TERGAN, Sigmar-Olaf; KELLER, Tanja (org.). *Knowledge and Information Visualization. Lecture Notes in Computer Science[s.l.]* : Springer Berlin Heidelberg, 2005. p. 238–255. DOI: 10.1007/11510154_13. Disponível em: http://link.springer.com/chapter/10.1007/11510154_13. Acesso em: 4 jul. 2023.

CAPITAL, V. **Open Finance: iniciativa facilita a concessão de crédito para pessoa física e jurídica.** Disponível em: <<https://www.vert-capital.com/noticia/open-finance-iniciativa-facilita-a-concessao-de-credito-para-pessoa-fisica-e-juridica>>. Acesso em: 24 jan. 2024.

CARVALHO, José Oscar Fontanini de. O papel da interação humano-computador na inclusão digital. *Transinformação*, v. 15, p. 75-89, 2003.

CHRISTOVÃO, Heloisa Tardin; BRAGA, Gilda Maria. Ciência da informação e sociologia do conhecimento científico: a intertematicidade plural. **Transinformação**, v. 9, n. 3, 1997.

Ciência da Informação, [S. l.], v. 7, n. 2, 2006. Disponível em: http://www.dgz.org.br/abr06/Art_03.htm. Acesso em: 24 fev. 2023.

CLARA, M. Palavras, palavras, palavras.... Disponível em: <https://humanidadesdigitais.org/2013/05/10/palavras-palavras-palavras/>. Acesso em: 22 abr. 2023.

COSTA, Luciano da Fontoura. What is a Complex Network? [S. l.], 2018. DOI: 10.13140/RG.2.2.10450.04804/2. Disponível em: <http://rgdoi.net/10.13140/RG.2.2.10450.04804/2>. Acesso em: 25 fev. 2023.

CRISTOVÃO, Henrique Monteiro. Um modelo híbrido de recuperação de informação e conhecimento baseado na síntese de mapas conceituais obtidos por operações de transformação de redes complexas orientadas por busca de relacionamentos entre termos de consulta em bases de dados ligados. 2016. Tese - (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2016. Disponível em: http://bit.ly/tese_cristovaohm. Acesso em: 28 jan. 2023.

DA COSTA SANTOS, Plácida LV Amorim; SANTANA, Ricardo César Gonçalves. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. *Ciência da informação*, v. 42, n. 2, 2013.

DA SILVA, Daniela Lucas; SOUZA, Renato Rocha. Representação de documentos multimídia: dos metadados às anotações semânticas. **Tendências da pesquisa brasileira em ciência da informação**, v. 7, n. 1, 2014.

DataGramZero - Revista de Ciência da Informação - Artigo 03. Disponível em: http://eprints.rclis.org/7470/1/Art_03.htm. Acesso em: 8 jan. 2024.

DAVENPORT, Thomas H. Conhecimento empresarial. Elsevier Brasil, 1998.

DAVENPORT, T. H. Dominando a gestão da informação. Porto Alegre: Bookman, 2004. Disponível em: <<http://lillianalvares.fci.unb.br/phocadownload/Fundamentos/Representacao/Aula12Ol.pdf>>. Acesso em: 23 fev. 2023.

DE ANDRADE, Morgana Carneiro. Interoperabilidade semântica entre repositórios digitais científicos: um perfil de aplicação de metadados baseado nas necessidades de informação em oncologia. 2017. Tese de Doutorado. Universidade do Minho (Portugal).

DE MEDEIROS TRAVASSOS SAEGER, M. M.; DE PINHO NETO, J. A. S.; LOUREIRO, J. Socialização da informação e do conhecimento: uma proposta emancipadora para a efetivação de instrumentos de democracia participativa. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 28, n. 1, 2018.

EVERETT, M. G.; BORGATTI, S. P. The dual-projection approach for two-mode networks. *Social Networks*, v. 35, n. 2, p. 204–210, 2012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0378873312000354>. Acesso em: 3 jul. 2023.

FARINELLI, F.; ELKIN, P. L. Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico. **Ciência da Informação**, v. 46, n. 1, 2017.

FERNANDES', 'FERNANDA. **Pessoas de baixa renda têm menos acesso a crédito no Brasil.** Disponível em: <<https://www.correiobraziliense.com.br/economia/2021/07/4940108-pessoas-de-baixa-renda-tem-menos-acesso-a-credito-no-brasil.html>>. Acesso em: 24 jan. 2024.

FIA. **Finanças sustentáveis: o que é, objetivo, importância e exemplos.** Disponível em: <<https://fia.com.br/blog/financas-sustentaveis/>>. Acesso em: 25 jan. 2024.

FRICKÉ, Martin. Knowledge pyramid: the DIKW hierarchy. **Knowledge Organization**, [S. l.], v. 49, n. 1, p. 33–46, 2019. Disponível em:

https://www.ergon-verlag.de/isko_ko/downloads/ko_46_2019_1_d.pdf. Acesso em: 10 maio. 2023.

GIOLO, Aguiar. A formação da taxa de juros nas instituições financeiras. 2009.

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.

Google Cloud - FinFacts 2023. Disponível em: <https://inthecloud.withgoogle.com/google-cloud-finfacts-2023/Register.html?utm_source=linkedin&utm_medium=smarpshare&utm_campaign=FY23-Q3-latam-LATM174-physicalevent-er-google-cloud-finfacts-2023&utm_content=linkedin_finfacts&utm_term=linkedin>. Acesso em: 29 ago. 2023.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, [S. l.], v. 5, n. 2, p. 199–220, 1993. DOI: 10.1006/knac.1993.1008. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1042814383710083>. Acesso em: 16 set. 2023.

GUIZZARDI, Giancarlo; WAGNER, Gerd. Conceptual simulation modeling with ONTO-UML advanced tutorial. In: Proceedings of the 2012 Winter Simulation Conference (WSC). IEEE, 2012. p. 1-15.

GUIZZARDI, Giancarlo. Ontology, Ontologies and the “I” of FAIR. Data Intelligence, [S. l.], v. 2, n. 1–2, p. 181–191, 2020. DOI: 10.1162/dint_a_00040. Disponível em: https://doi.org/10.1162/dint_a_00040. Acesso em: 26 jul. 2023.

HAND, David J.; MANNILA, Heikki; SMYTH, Padhraic. Principles of data mining. Cambridge, MA: MIT Press, 2001.

HJØRLAND, Birger. Data (with big data and database semantics). **Knowledge Organization**, [S. l.], v. 45, n. 8, p. 643–652, 2018. DOI: [10.5771/0943-7444-2018-8-643](https://doi.org/10.5771/0943-7444-2018-8-643). Disponível em: <https://www.isko.org/cyclo/data>. Acesso em: 28 jul. 2021.

HJØRLAND, Birger. What is knowledge organization (KO)? **Knowledge organization. International journal devoted to concept theory, classification, indexing and knowledge representation**, [S. l.], 2008. Disponível em: <http://hdl.handle.net/10150/106183>. Acesso em: 15 nov. 2023.

Inclusão financeira e renegociação de dívidas são debatidas no G20. (2024, fevereiro 29). Agência Brasil. <https://agenciabrasil.ebc.com.br/economia/noticia/2024-02/inclusao-financeira-e-rene-gociacao-de-dividas-sao-debatidas-no-g20>

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

LÉVY, Pierre. A inteligência coletiva. São Paulo: Loyola, p. 22, 1998.

LÉVY, P. As tecnologias da inteligência. O futuro do pensamento na era da informática. Rio de Janeiro: 34, 1993.

LE COADIC, Yves-François. A ciência da informação. Briquet de Lemos, 2004.

LEMOS, André. Dataficação da vida. *Civitas-Revista de Ciências Sociais*, v. 21, p. 193-202, 2021.

LIEW, Anthony. DIKIW: data, information, knowledge, intelligence, wisdom and their interrelationships. **Business Management Dynamics**, [S. l.], v. 2, n. 10, p. 49–62,

2013. Disponível em: http://bmdynamics.com/recent_issue.php?id=25. Acesso em: 31 mar. 2023.

Mais liberdade de escolha e serviços para você. Open Banking Brasil, 2022. Disponível em: <https://openbankingbrasil.org.br/?cookie=true>. Acesso em: 18 set. 2023.

MARCONDES, Carlos Henrique; DE ALMEIDA CAMPOS, Maria Luiza. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. PontodeAcesso, v. 2, n. 1, p. 107-136, 2008.

MARTELETO, Regina Maria. Análise de redes sociais – aplicação nos estudos de transferência da informação. Ciência da Informação, v. 30, n. 1, 2001. Disponível em: <http://www.scielo.br/pdf/ci/v30n1/a09v30n1.pdf>. Acesso em: 3 jul. 2023.

MASON, Hilary; PATIL, Dj. Data Driven. Sebastopol, CA, USA: O'Reilly Media, Inc., 2015. . Acesso em: 4 set. 2023.

MATHEUS, Renato Fabiano; SILVA, Antonio Braz de Oliveira E. Análise de redes sociais como método para a Ciência da Informação. DataGramaZero - Revista de Ciência da Informação, v. 7, n. 2, 2006. Disponível em: http://www.dgz.org.br/abr06/Art_03.htm. Acesso em: 3 jul. 2023.

MATTA, Rodrigo Octávio Beton. Modelo de comportamento informacional de usuários: uma abordagem teórica. Gestão, mediação e uso da informação, p. 127-142, 2010.

MAZZOCCHI, Fulvio. Knowledge organization system (KOS). Knowledge Organization, [S. l.], v. 45, n. 1, p. 54–78, 2018. Disponível em: <https://www.isko.org/cyclo/kos.htm>.

MCGARRY, Kevin. O contexto dinâmico da informação: uma análise introdutória. In: O contexto dinâmico da informação: uma análise introdutória. 1999. p. vi, 206-vi, 206.

MORETTIN, Pedro A.; SINGER, Julio M. **Introdução à Ciência de Dados: fundamentos e aplicações**. São Paulo: USP, 2020. Disponível em: <https://www.ime.usp.br/~jmsinger/MAE0217/cdados2020mar.pdf>.

NETO, R. C. **Cenário econômico e Agenda BC#**. Disponível em: <https://www.bcb.gov.br/conteudo/home-ptbr/TextosApresentacoes/Apresentação_RCN_IEDI_VPUB.pdf>. Acesso em: 29 ago. 2023.

NEWMAN, Marcos EJ. A estrutura e a função das redes. Computer Physics Communications, v. 147, n. 1-2, p. 40-45, 2002.

NEWMAN, Mark. Networks. Oxford university press, 2018.

NEWMAN, M. E. J. Networks: an introduction. Oxford ; New York: Oxford University Press, 2010.

No novo sistema financeiro, o consumidor é quem manda. Disponível em: <<https://valor.globo.com/patrocinado/embratel/juntos-no-proximo-nivel/noticia/2021/11/03/no-novo-sistema-financeiro-o-consumidor-e-quem-manda.ghtml>>. Acesso em: 10 nov. 2023.

OData - a melhor maneira de REST. Disponível em: <<https://www.odata.org/>>. Acesso em: 7 abr. 2023.

Open Finance. Disponível em: <<https://www.bb.com.br/site/open-finance/>>. Acesso em: 20 jan. 2024.

Open Finance. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/openfinance>>. Acesso em: 2 nov. 2023.

Open Finance. (2022, janeiro 26). Portal BB. <https://www.bb.com.br/site/open-finance/>

OLIVEIRA, Mariana; TENÓRIO, Nelson; BORTOLOZZI, Flávio. A compreensão de reporte de bugs no desenvolvimento e uso de software: uma representação do conhecimento por meio de ontologia. **Revista Tecnologia e Sociedade**, v. 18, n. 51, p. 244-259, 2022.

OTTE, Evelien; ROUSSEAU, Ronald. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, v. 28, n. 6, p. 441–453, 2002. Disponível em: <https://doi.org/10.1177/016555150202800601>. Acesso em: 3 jul. 2023.

OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. **Journal of information science**, v. 28, n. 6, p. 441–453, 2002.

Panorama do Censo 2022. Disponível em: <https://censo2022.ibge.gov.br/panorama/>. Acesso em: 10 mar. 2024.

PARAVELA, Tatyana Chiari; DOMINGUES, Juliana Oliveira. Open Banking: a implementação do sistema financeiro aberto no Brasil na perspectiva do consumidor. *Revista da Procuradoria-Geral do Banco Central*, v. 15, n. 2, p. 81-92, 2021.

População residente estimada. Disponível em: <https://sidra.ibge.gov.br/tabela/6579>. Acesso em: 11 mar. 2024.

PORTO, Fábio; ZIVIANI, Arthur. Ciência de Dados. Em: 2014, Rio de Janeiro. Anais [...]. . Em: SEMINÁRIO DE GRANDES DESAFIOS DA COMPUTAÇÃO NO BRASIL. Rio de Janeiro: SBC, 2014. Disponível em: <https://www.lncc.br/~ziviani/papers/III-Desafios-SBC2014-CiD.pdf>.

PRADO FILHO, H. R. DO. A Qualidade dos metadados dos documentos de arquivo. *Revista Digital AdNormas*, 29 out. 2019.

Publicação. Disponível em: https://www.linkedin.com/posts/banco-central-do-brasil_dar-%C3%A0s-pessoas-acesso-a-cr%C3%A9dito-%C3%A9-um-passo-activity-7170826928822677507-Cb14/. Acesso em: 5 mar. 2024.

Quem Participa. Disponível em: <<https://openfinancebrasil.org.br/quem-participa/>>. Acesso em: 2 nov. 2023.

RABELLO, Rodrigo. A Ciência da Informação como objeto: epistemologias como lugares de encontro. *Perspectivas em Ciência da Informação*, v. 17, p. 2-36, 2012.

RASTELI, Alessandro; CAVALCANTE, Lúcia Eugênia. Mediação cultural e apropriação da informação em bibliotecas públicas. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 19, n. 39, p. 43–58, 2014. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2014v19n39p43>. Acesso em: 3 jul. 2023.

RAUTENBERG, Sandro; CARMO, Paulo Ricardo Viviurka Do. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. *Brazilian Journal of Information Science*, [S. l.], v. 13, n. 1, p. 56–67, 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6983493>. Acesso em: 8 maio. 2023.

REIS, Makson de Jesus. *Ciência de dados e ciência da informação: guia para alfabetização de dados para bibliotecários*. 2019.

ROBREDO, Jaime. *Da Ciencia Da Informacao Revisitada..* Thesaurus Editora, 2003.

SADIKU, Mateus e cols. Visualização de dados. *Revista Internacional de Pesquisa em Engenharia e Tecnologia Avançada (IJERAT)*, v. 2, n. 12, pág. 11-16, 2016.

SARACEVIC, Tefko. *Ciência da informação: origem, evolução e relações*. *Perspectivas em ciência da informação*, v. 1, n. 1, 1996.

SCHIELD, Milo. Information literacy, statistical literacy and data literacy. Em: 2004, Madison, Wisconsin, USA. *Anais [...]*. Em: INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE & TECHNOLOGY, IASSIST. Madison, Wisconsin, USA Disponível em: https://iassistquarterly.com/public/pdfs/iqvol282_3shields.pdf. Acesso em: 3 jul. 2023.

Serviços financeiros caminham para a “hiper personalização”. (2022, outubro 25). Valor Investe. <https://valorinveste.globo.com/produtos/servicos-financeiros/noticia/2022/10/25/servicos-financeiros-caminham-para-a-hiper-personalizacao.ghtml>

SILVA, Fabiano Couto Corrêa da. Visualização de dados: passado, presente e futuro. LIINC em revista. Rio de Janeiro, RJ. Vol. 15, n. 2 (nov. 2019), p. 205-223, 2019.

SILVA, Erico Verissimo et al. Memória, privacidade e o legítimo interesse da coletividade: o Sesc Memórias e a Lei Geral de Proteção de Dados Pessoais. 2022. SHIELDS, Milo. Information literacy, statistical literacy, data literacy. IASSIST quarterly, v. 28, n. 2-3, p. 6-6, 2005.

SIMÕES, Maria da Graça; LIMA, Gercina ngela De. Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências. [s.l.] : Coimbra University Press, 2020. DOI: 10.14195/978-989-26-1944-6. Disponível em: <https://directory.doabooks.org/handle/20.500.12854/45342>. Acesso em: 15 dez. 2021.

Sobre o Portal - Portal de Dados Abertos do Banco Central do Brasil. ([s.d.]). Gov.br. Recuperado em maio de 2022, de <https://dadosabertos.bcb.gov.br/pages/sobre-o-portal>

TAKAHASHI, Tadao. **Sociedade da informação no Brasil: livro verde**. Ministério da Ciência e Tecnologia (MCT), 2000.

TEIS, Denize Terezinha; TEIS, Mirtes Aparecida. A abordagem qualitativa: a leitura no campo de pesquisa. 2006. Disponível em: <http://bocc.ufp.pt/pag/teis-denize-abordagem-qualitativa.pdf>. Acesso em: 5 dez. 2023.

TOMAÉL, Maria Inês. Redes de informação: o ponto de contato dos serviços e unidades de informação no Brasil. Informação & Informação, [S. l.], v. 10, n. 1–2, p. 5–30, 2005. DOI: 10.5433/1981-8920.2005v10n1-2p5. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1611>. Acesso em: 4 abr. 2023.

URDANETA, I. P. Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo informacional: retos y oportunidades. Caracas: Instituto del Conocimiento de la Universidad Simon Bolívar, 1992.

VALENTIM, Marta Lígia Pomim et al. Inteligência competitiva em organizações: dado, informação e conhecimento. DataGramaZero, Rio de Janeiro, v. 3, n. 4, p. 1-13, 2002

VALENTIM, M. L. P. Processos de compartilhamento e socialização do conhecimento em ambientes empresariais. **Ciência da Informação**, v. 45, n. 3, 2016.

Vem aí o open capital market. Presidente da CVM explica o que é. Disponível em:

<<https://borainvestir.b3.com.br/noticias/mercado/vem-ai-o-open-capital-market-presidente-da-cvm-explica-o-que-e/>>. Acesso em: 29 ago. 2023.

VIRKUS, Sirje; GAROUFALLOU, Emmanouel. Data science from a library and information science perspective. Data Technologies and Applications, v. 53, n. 4, p. 422–441, 2019. Disponível em: <https://doi.org/10.1108/DTA-05-2019-0076>. Acesso em: 3 jul. 2023.

WAARD, Anita De; COUSIJN, Helena; AALBERSBERG, IJsbrand Jan. **10 aspects of highly effective research data.** web page. 2015. Disponível em: <https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data>. Acesso em: 2 nov. 2023.

WIKIPEDIA CONTRIBUTORS. Estado da arte. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Estado_da_arte&oldid=63582245>.

WIKIPEDIA CONTRIBUTORS. **Gartner.** Disponível em: <https://en.wikipedia.org/w/index.php?title=Gartner&oldid=1139789021>.

WIKIPEDIA CONTRIBUTORS. Organização para a Cooperação e Desenvolvimento Econômico. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Organiza%C3%A7%C3%A3o_para_a_Coopera%C3%A7%C3%A3o_e_Developolvimento_Econ%C3%B4mico&oldid=64949085>.

Wisconsin, USA Disponível em: https://iassistquarterly.com/public/pdfs/iqvol282_3shields.pdf.

WOLFF, Annika; GOOCH, Daniel; MONTANER, Jose J. Caverio; RASHID, Umar; KORTUEM, Gerd. Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics, [S. l.], v. 12, n. 3, 2016. DOI: 10.15353/joci.v12i3.3275. Disponível em: <https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3275>. Acesso em: 23 fev. 2023.

WANDSCHEER, Lucelaine dos Santos Weiss; JARUDE, Jamile Nazaré Duarte Moreno; VITA, Jonathan Barros. O SISTEMA FINANCEIRO ABERTO (OPEN BANKING) SOB A PERSPECTIVA DA REGULAÇÃO BANCÁRIA E DA LEI GERAL DE PROTEÇÃO DE DADOS. Revista Brasileira de Filosofia do Direito, v. 6, n. 1, p. 78-95, 2020.

WASSERMAN, Stanley; FAUST, Katherine. Social network analysis: methods and applications. Cambridge, England; New York: Cambridge University Press, 1994.

WELLMAN, Barry. Computer networks as social networks. Science, v. 293, n. 5537, p. 2031-2034, 2001.

WIKIPEDIA CONTRIBUTORS. **Dicionário de dados**. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Dicion%C3%A1rio_de_dados&oldid=63545652#Nota%C3%A7%C3%A3o>.

WIKIPEDIA CONTRIBUTORS. **Patrick Doreian**. Disponível em: <https://en.wikipedia.org/w/index.php?title=Patrick_Doreian&oldid=1161482401>.

YAMAUTI, Rogerio Koji; TSUNODA, Denise Fukumi. Análise bibliométrica das publicações sobre Alfabetização em Dados. Pesquisa em GESTÃO e ORGANIZAÇÃO da Informação, p. 127, 2019.

ZENG, Marcia Lei. Knowledge organization systems (KOS). **KNOWLEDGE ORGANIZATION**, [S. l.], v. 35, n. 2–3, p. 160–182, 2008. DOI: 10.5771/0943-7444-2008-2-3-160. Disponível em: https://www.ergon-verlag.de/isko_ko/downloads/ko35200823h.pdf. Acesso em: 21 set. 2023.

ANEXOS

Anexo 1: Tabela População

Estado	População Estimada
Acre	906876
Alagoas	3365351
Amapá	877613
Amazonas	4269995
Bahia	14985284
Ceará	9240580
Distrito Federal	3094325
Espírito Santo	4108508
Goiás	7206589
Maranhão	7153262
Mato Grosso	3567234
Mato Grosso do Sul	2839188
Minas Gerais	21411923
Pará	8777124
Paraíba	4059905
Paraná	11597484
Pernambuco	9674793
Piauí	3289290
Rio de Janeiro	17463349
Rio Grande do Norte	3560903
Rio Grande do Sul	11466630
Rondônia	1815278
Roraima	652713
Santa Catarina	7338473
São Paulo	46649132
Sergipe	2338474
Tocantins	1607363

Fonte: IBGE