



GT-8 – Informação e Tecnologia

ISSN 2177-3688

PRÁTICAS EM ANÁLISE DE REDES SOCIAIS PARA DATA LITERACY EM CIÊNCIA DE DADOS NO ÂMBITO DA FORMAÇÃO NA ÁREA CIÊNCIA DA INFORMAÇÃO

SOCIAL NETWORK ANALYSIS PRACTICES FOR DATA LITERACY IN DATA SCIENCE IN THE FIELD OF INFORMATION SCIENCE EDUCATION

Henrique Monteiro Cristovão - Universidade Federal do Espírito Santo (UFES)

Lucas dos Santos do Vale - Universidade Federal do Espírito Santo (UFES)

Juliana Rodrigues de Lima Meirelles - Universidade Federal do Espírito Santo (UFES)

Stella Schwanz Dias de Assis - Universidade Federal do Espírito Santo (UFES)

Modalidade: Trabalho Completo

Resumo: Em um mundo cada vez mais orientado a dados, investir em tecnologia não tem sido suficiente, pois há demandas por competências para lidar com dados, ou *data literacy*, em Ciência de Dados (CD). Pesquisadores sinalizaram a importância de *data literacy* por profissionais de áreas da Ciência da Informação (CI). Nesse cenário, e considerando uma abordagem com atividades factíveis para o contexto de estudantes não pertencentes à área tecnológica, o objetivo da pesquisa é discutir a importância de experiências práticas de estudo de técnicas de análise de redes sociais (ARS) enquanto possibilidade para o desenvolvimento de *data literacy* em CD na formação de estudantes de áreas da CI. Com natureza qualitativa, tipo descritiva-exploratória, a pesquisa utilizou-se de duas práticas de ensino de ARS em disciplinas de graduação em Arquivologia. O contexto da primeira prática foi em um ambiente de fóruns de discussão virtuais onde os alunos debateram temáticas contemporâneas e polêmicas. O contexto da segunda prática foi em uma base de dados sobre informações pessoais dos próprios alunos da turma. Com o apoio de ferramentas tecnológicas adequadas, foram criadas redes sociais e aplicadas técnicas de ARS para a descoberta de conhecimento enquanto procedimentos inerentes à CD. Considerando a inserção de atividades de ARS em ambientes de ensino-aprendizagem habituais, e sendo os próprios alunos os atores das redes sociais analisadas, houve um bom entendimento de aspectos da ARS enquanto processo de mineração de dados em CD, entendimento do funcionamento de algoritmos de inteligência artificial em mídias sociais e a percepção de melhoria de *data literacy* nos estudantes.

Palavras-chave: análise de redes sociais; data literacy; formação em ciência da informação; ciência de dados; descoberta de conhecimento.

Abstract: In an increasingly data-driven world, investment in technology has not been enough, as there is a demand for skills to deal with data, or data literacy, in Data Science (DS). Researchers signaled the importance of data literacy for professionals in the areas of Information Science (IS). In this scenario, and considering an approach with feasible activities for the context of students not belonging to the technological area, the objective of the research is to discuss the importance of practical study experiences of studying social network analysis techniques (SNA) as a possibility for the development of data literacy on DS in the training of students in IS areas. With a qualitative, descriptive-exploratory nature, the research used two SNA teaching practices in undergraduate courses in Archival Science. The context of the first practice was in an environment of virtual discussion forums where students debated contemporary and controversial themes. The context of

the second practice was in a dataset of personal information of the students in the class. With the support of appropriate technological tools, social networks were created and SNA techniques were applied to discover knowledge as procedures inherent to DS. Considering the inclusion of SNA activities in usual teaching-learning environments, and the students themselves being the actors of the analyzed social networks, there was a good understanding of aspects of SNA as a process of data mining on DS, understanding of the functioning of algorithms of artificial intelligence in social media and the perception of data literacy improvement in students.

Keywords: information network analysis, data literacy, training in information science, data science, knowledge discovery.

1 INTRODUÇÃO

Parte considerável do conhecimento explícito hoje no mundo, organizado ou não, está em redes como, por exemplo, bancos de dados e repositórios interconectados, páginas Web e mídias sociais. Nesse contexto, o mundo tem se movimentado em direção à organização desse conhecimento utilizando-se de uma granularidade cada vez menor e de forma orientada a dados, ou *data driven*, para possibilitar a realização de tarefas antes impossíveis ou impensadas. Muitas técnicas têm sido empregadas para viabilizar a realização de *insights*, análises, recuperação de informação e descoberta de conhecimento com demandas de usuários cada vez mais exigentes e sofisticadas. Por exemplo, nas mídias sociais, muitas organizações utilizam-se da análise de redes sociais para obter aprimoramento em seus serviços de atendimento ao cliente, seja nas vendas, no suporte pós-vendas ou mesmo no apoio a tomada de decisões estratégicas para melhoria da sua gestão e dos fluxos informacionais.

Lyon et al. (2015) fizeram um estudo sobre os requisitos e funções solicitadas para a contratação de *data archivist* (arquivista de dados), *data librarian* (bibliotecário de dados) e *data steward/curator* (administrador/curador de dados) nos Estados Unidos. Nessas três tipologias constatou-se fortes demandas por competências associadas ao tratamento de dados e metadados. Alguns destaques de competências foram indicados, onde arquivista de dados é fortemente ligado a questões da preparação e documentação dos dados, bibliotecário de dados no treinamento do uso e na disseminação dos dados, e administrador/curador de dados com ênfase em processos gerenciais da administração dos dados.

Nessa mesma linha, a pesquisa de Madeiro e Dias (2020), em sites de bancos internacionais de trabalho, verificou demandas por profissionais da informação, em especial arquivistas, na área do gerenciamento e curadoria de dados, descrição de metadados e ainda

o domínio de pacotes de software estatísticos e linguagens computacionais. Ainda na função de arquivista, Cristovão, Ferrari e Moraes (2022) indicaram que esse profissional pode gerir dados, classificando-os em categorias, utilizando-se de metadados adequados para garantir o seu registro preciso, a proveniência e a organicidade com foco na melhoria de sua recuperação e na manutenção de sua autenticidade.

Sayão e Sales (2016) também indicaram competências no entorno do tratamento de dados por profissionais da informação, sendo que a função de bibliotecário é destacada por Silva (2016) como tendo o perfil necessário para o desenvolvimento de projetos de gestão de dados ou ainda, indicado por Lage, Losoff e Maness (2011), na curadoria de dados. No âmbito de museus, Markus et al. (2021) sinalizaram a importância do desenvolvimento de habilidades e conhecimentos sobre ferramentas de manipulação de dados pela equipe interna dessas instituições para que houvesse uma integração eficaz com os profissionais especialistas em Ciência de Dados (CD) que integram o grupo de trabalho.

Contudo, a aplicação adequada de técnicas que lidam com dados para descoberta de conhecimento, como os processos de análise de redes, normalmente é restrita a profissionais com formação específica em tecnologia. Por outro lado, observa-se que hoje em dia a compreensão desses processos, dependendo da abordagem, não é complexo e pode empoderar profissionais diversos a participarem mais ativamente em suas instituições.

Enquanto o investimento em tecnologia, por parte das empresas e instituições, torna-se necessário, observa-se que somente esse tipo de aporte não é suficiente. Há carência de profissionais com capacidade em aproveitar de forma plena as potencialidades das tecnologias por meio de competências em *data literacy* que, segundo Wolff et al. (2016), é a capacidade de fazer e responder perguntas associadas a problemas do mundo real a partir de bases de dados utilizando-se de processos de investigação. Por outro lado, não basta formar profissionais especialistas em CD, pois a demanda por tarefas que exigem competências com dados é crescente. Pothier e Condon (2020) propõem a solução de criar condições para que estudantes de várias áreas do conhecimento sejam capacitados em *data literacy* em suas próprias formações, e não somente em formações na área de tecnologia.

Uma das consequências positivas do movimento de *data literacy* é a apropriação de conhecimentos relativos à CD que é uma área interdisciplinar com um campo de atuação bem amplo, cujas possibilidades para obtenção de benefícios sociais são concretas. As conexões entre a CD e a Ciência da informação (CI) estão cada vez mais aparentes na

pesquisa científica. Matos, Conduru e Benchimol (2022) fizeram uma análise bibliométrica de artigos científicos que apresentaram correlação entre a CI e a CD, no período de 1996 a 2021, e constataram um aumento significativo da quantidade, principalmente nos últimos 6 anos. Os autores também perceberam maior interesse da CI nos temas relacionados à parte prática da CD e em áreas específicas como a mineração de dados e a competência em informação. Para Lemos, Martins e Carmo (2022) a CI possui papel fundamental na qualidade dos dados e metadados que são usados pelo campo da CD em processos de descrição, catalogação, classificação e indexação.

Além disso, tem sido observado maior incidência do ensino de CD em cursos da área da CI. Zakaria (2023) desenvolveu uma pesquisa sobre a atual situação do ensino de CD em graduações e pós-graduações em áreas da CI e fez indicações sobre a importância da inclusão de competências de *data literacy* em CD nos currículos para a formação dos estudantes dessas áreas. Cristovão, Ferrari e Moraes (2022) relataram a experiência de ensino de CD em curso de graduação em Arquivologia na UFES. Curty e Serafim (2016) citaram o caso de um curso de graduação em CI da UFSC com uma formação voltada ao desenvolvimento de habilidades e competências de CD.

Contudo, segundo Martins (2022), a CD é muito nova no contexto da CI e a "[...] compreensão de suas especificidades ainda é um problema de pesquisa em aberto suscitando discussão, experiências e análise de propostas em andamento". Martins complementa afirmando que o ensino da CD no Brasil e no contexto da CI, ainda está em sua fase inicial de apropriação pelos pesquisadores e professores da área, apesar da importância desse tipo de ensino para melhoria da formação dos alunos. Zakaria (2023) também sinaliza que o ensino de CD em cursos de áreas da CI ainda é incipiente, porém ele indica a sua incorporação nos currículos.

Uma das temáticas estudadas na CD é a análise de redes de informação, mais conhecida por análise de redes sociais (ARS), é considerada como uma das técnicas de mineração de dados para descoberta de conhecimento. A sua utilização é normalmente associada ao uso de ferramentas com interface gráfica com bom nível de usabilidade. Essas ferramentas possibilitam executar ações de descoberta de conhecimento por meio de formatações, cálculo de métricas, análise e a execução de algoritmos de distribuição e inspeção visual. Utilizando-se da técnica da inspeção visual, a observação atenta do usuário

no reconhecimento de padrões visuais é mais importante do que conhecimentos avançados em cálculos estatísticos.

Diante desse cenário de necessidade de melhoria em *data literacy* e considerando uma abordagem com possibilidades de atividades factíveis para o contexto dos estudantes de áreas da CI, o objetivo da pesquisa é discutir a importância de experiências práticas de estudo de ARS enquanto possibilidade para o desenvolvimento de *data literacy* em CD na formação de estudantes de áreas da CI.

Espera-se contribuir com o debate sobre possibilidades de uso de tecnologias em processos de ensino-aprendizagem para aprimoramento de *data literacy* na formação de estudantes de áreas da CI e com foco em CD. Para isso, a pesquisa baseou-se em dois estudos de caso de ensino de ARS, em situações distintas, envolvendo disciplinas de um curso de graduação em Arquivologia.

2 DATA LITERACY E CIÊNCIA DE DADOS

Em um mundo que tende a ser orientado a dados, ou *data driven*, tem-se a preocupação não só em armazenar os dados de forma organizada para posterior recuperação, conforme orientam os princípios FAIR¹ (WILKINSON et al, 2016), mas também o desafio em descobrir conhecimentos, a partir de inferências sobre esses dados, para orientar e prever ações em vários contextos e situações organizacionais.

Dentro do contexto do movimento *data driven*, a ação denominada de *datafy*, segundo Mayer-Schonberger e Cukier (2013), é aplicada a um fenômeno para colocá-lo em um formato quantificado com o objetivo de ser tabulado e analisado. Nesse contexto, a *datafication* é uma tendência atual e tecnológica capaz de transformar aspectos da vida cotidiana em dados que possam ser transformados em valores e insights.

Ainda no contexto do movimento *data driven*, o fenômeno denominado de *dataism* foi primeiramente descrito por David Brooks em 2013 como uma vertente filosófica para denotar que "[...] tudo que pode ser medido deve ser medido, e os dados são uma lente transparente e confiável que permite filtrar o emocionalismo e a ideologia, ajudando a fazer coisas notáveis como prever o futuro" (BROOKS, 2013, tradução nossa). O autor também destaca que os dados podem iluminar padrões de comportamento difíceis de serem

¹ FAIR é o acrônimo de *findable, accessible, interoperable e reusable* (encontrável, acessível, interoperável e reusável) aplicável a dados.

observados pois, diferentemente dos computadores, nossos cérebros geralmente não percebem padrões sutis. Harari (2016) complementa afirmando que, para o *dataism*, todas as estruturas políticas, sociais e econômicas podem ser interpretadas como sistemas de processamento de dados e, portanto, podem ser mensuradas e registradas como dados.

Considerando a contemporaneidade do movimento *data driven* e os fenômenos *datafication* e *dataism*, observa-se a necessidade por melhoria nas competências associadas a *data literacy* pelos profissionais da informação. *Data literacy*, ou literacia em dados, é como um conhecimento subjacente à *information literacy* no contexto da CI (BELLUZZO; FERES; BASSETTO, 2011).

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, por meio do pensamento crítico (WOLFF *et al.*, 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).

No âmbito da *data literacy*, Pothier e Condon (2020) identificaram sete competências fundamentais para serem desenvolvidas: (1) Organizar e armazenar dados; (2) Entender os dados usados enquanto elementos fundamentais para o negócio; (3) Avaliar a qualidade das fontes de dados; (4) Interpretar dados; (5) Tomar decisões baseada em dados; (6) Comunicar e apresentar utilizando-se de dados; e (7) Agir com ética e segurança com dados. Muitas dessas competências indicadas por Pothier e Condon (2020) são fundamentais para a CD. Segundo Virkus e Garoufallou (2021), a CD requer diversos conhecimentos e habilidades de diferentes disciplinas, entre os quais está a CI. Assim, o desenvolvimento da *data literacy* pode cumprir um importante papel na aproximação de profissionais que não são originalmente da tecnologia em campos de pesquisa da CD.

A CD visa transformar dados (estruturados ou não estruturados, grandes ou pequenos, estáticos ou *streaming*) em valor real (na forma de previsões, decisões automatizadas, modelos baseados em dados ou visualizações de dados que ofereçam informações úteis) (AALST, 2016). Dentre as metodologias e técnicas usadas na CD, destaca-se a descoberta de conhecimento, conhecida na literatura mundial por KDD (*knowledge discovery in databases*), como sendo um campo de pesquisa da inteligência artificial que envolve a *data mining* (mineração de dados) (HAND; MANNILA; SMYTH, 2001).

A CD pode responder a perguntas que ainda não foram formuladas 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 (HAND; MANNILA; SMYTH, 2001).

No contexto da CD, a apropriação de competências de *data literacy* não só abrirá possibilidades no tratamento direto com os dados, pelos profissionais da informação, como também irá permitir a esses profissionais a participação e diálogo com equipes interdisciplinares envolvendo contextos mais tecnológicos e *data driven*.

3 ANÁLISE DE REDES SOCIAIS NA CIÊNCIA DA INFORMAÇÃO E CIÊNCIA DE DADOS

No âmbito das Ciências Sociais, a ARS permite abordagens capazes de responder a questões de pesquisa que remetem a uma definição formal e precisa para aspectos do ambiente estrutural político, econômico ou social (WASSERMAN; FAUST, 1994). A ARS não constitui um fim em si mesma, mas um meio para realizar uma análise estrutural com objetivo de elucidar fenômenos onde a forma (ou topologia) da rede consegue explicar os fenômenos analisados (MARTELETO, 2001).

A ARS é considerada uma ferramenta metodológica (MATHEUS; SILVA, 2006) que permite enxergar o que outras abordagens não permitem (WASSERMAN; FAUST, 1994; HIGGINS, RIBEIRO; 2018), e possui "[...] origem multidisciplinar (psicologia, sociologia, antropologia, matemática, estatística) cuja principal vantagem é a possibilidade de formalização gráfica e quantitativa de conceitos abstraídos a partir de propriedades e processos característicos da realidade social" (SOUZA; QUANDT, 2008), além de ser considerada uma estratégia para investigar estruturas sociais e sendo fortemente vinculada à área da CI (OTTE, ROUSSEAU, 2002). A CI pode colaborar fortemente com a ARS cuidando da parte da "[...] organização do conhecimento, de informações e de dados, entregando um conteúdo mais bem tratado e semanticamente bem identificado para o processamento da parte específica da análise de dados, tratada por técnicas da CD" (VALE; CRISTOVÃO, 2023).

Porto e Ziviani (2014) destacam a análise de redes complexas (que se utiliza das técnicas de ARS), como sendo um dos três grandes desafios de pesquisa em CD, junto com a análise e a gerência de dados. A ARS é considerada como um dos processos para a mineração de dados na descoberta de conhecimento na área da CD.

A ARS gera padrões visuais que são tratados pela área da visualização de informação, conhecida por *information visualization*. Essa área tem como objetivos a revelação de padrões invisíveis a partir de dados abstratos e a possibilidade de obtenção de novas percepções, e não apenas imagens bonitas (CHEN, 2013). A representação desses dados abstratos pode ser feita em formato de gráficos e imagens diversas que favorecem a leitura de seus significados. Normalmente, aplicam-se processos de mineração de dados para potencializar a revelação de relações não esperadas (HAND; MANNILA; SMYTH, 2001).

Parte considerável das redes tratadas na ARS advém de redes denominadas de bipartidas, conhecidas também por redes de afiliação ou redes de dois modos. Elas são redes com dois conjuntos de nós, onde os nós de um conjunto só se conectam com os nós do outro conjunto. Por exemplo, uma rede com um conjunto de nós representando autores de artigos publicados e outro conjunto de nós representando as palavras-chave de seus artigos. Nessa rede bipartida, um autor somente pode se relacionar com palavras-chave, mas não com outro autor.

Um dos processos usados em ações de mineração de dados e descoberta de conhecimento em CD, é a projeção bipartida que transforma uma rede bipartida em uma monopartida. Nesse caso, os nós de uma rede monopartida podem se conectar entre si, sem restrição. Por exemplo, quando a rede bipartida de autores e palavras-chave sofre uma projeção bipartida para eliminação, por exemplo, dos nós que representam as palavras-chave, tem-se como resultante uma rede monopartida de autores. Essa rede social resultante pode revelar grupos de interesse de pesquisa entre os autores que foram conectados entre si por intermédio dos nós das palavras-chave coincidentes.

Redes bipartidas podem levar à descoberta de conexões subjacentes e facilitar a interpretação com a aplicação de projeções bipartidas (BORGATTI, HALGIN; 2014; GAO et al., 2017). Apesar de parte considerável de uma rede, que sofreu uma projeção bipartida, ser eliminada, essa ação viabiliza a análise por meio de inspeção visual, uma vez que as redes ficam menores e podem evidenciar relacionamentos escondidos. Embora alguns pesquisadores, como Newman (2010) e Kadushin (2004), recomendarem que uma rede deve ser trabalhada em sua totalidade para não haver perda de informações sobre os relacionamentos entre as entidades, a projeção bipartida pode evidenciar mais facilmente relações entre variáveis.

Everett e Borgatti (2012) analisaram criticamente a suposição de que projeções bipartidas poderiam levar a perdas de informações estruturais da rede. Os autores concluíram que análises realizadas sobre as redes resultantes não provocaram perdas de informação e ainda tinham a sua análise facilitada. Nessa mesma linha, Melamed (2014) também fez estudos comparativos e concluiu que os resultados de análises realizadas em redes resultantes de projeções bipartidas são, em muitos casos, melhores do que aqueles realizados na rede bipartida original.

4 PROCEDIMENTOS METODOLÓGICOS

A pesquisa possui natureza qualitativa, é do tipo descritiva-exploratória. O caráter exploratório da pesquisa foi delineado pelo fato de não terem sido encontradas publicações com experiências similares ao contexto das práticas relatadas, isto é, com a utilização de ARS com o propósito de desenvolver *data literacy* em CD para estudantes da área de CI.

As duas práticas, com cenários distintos, foram iniciadas no ano de 2021 e continuam sendo aplicadas até o ano atual de 2023 a estudantes de graduação em Arquivologia da Universidade Federal do Espírito Santo (UFES). Detalhes descritivos e os procedimentos metodológicos, utilizados em cada prática, são apresentados nos próximos parágrafos.

Todos os arquivos utilizados para gerar os resultados, inclusive os exemplos da seção de procedimento metodológicos, estão disponíveis no repositório de dados da pesquisa². Para atender a LGPD³, os dados dos participantes foram pseudonimizados que, segundo a LGPD é o "[...] tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro".

O cenário da primeira prática ocorre em um ambiente de fóruns de discussão virtuais onde os alunos debatem temáticas contemporâneas e polêmicas associadas ao conteúdo das disciplinas: Comunicação na Sociedade da Informação, obrigatória do 2º período, e Raciocínio Lógico, obrigatória do 1º período.

O ambiente virtual de aprendizagem (AVA), baseado no Moodle⁴, dá apoio às atividades presenciais e oferece um serviço de criação e administração de fóruns e de

² Repositório de dados da pesquisa disponível em: https://github.com/henrique-cristovao/2023_enancib.

³ LGPD (Lei Geral de Proteção de Dados) disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

⁴ Moodle é uma plataforma para criação de cursos e salas virtuais para apoio ao ensino-aprendizagem. Disponível em: <https://moodle.org/>.

enquete. Os fóruns podem ser respondidos livremente, criando-se um novo tópico como resposta a alguma das questões, ou simplesmente respondendo um tópico existente de algum colega. As enquetes de cada fórum possuem o intuito de verificar o posicionamento dos estudantes quanto à questão formulada: Opção 1 - totalmente favorável; Opção 2 - mais à favor do que contrário, Opção 3 - mais contrário do que à favor, e Opção 4 - totalmente contrário.

A participação na atividade é organizada em uma sequência cronológica com três fases: 1ª fase - os estudantes manifestam o seu posicionamento inicial sobre a questão na enquete; 2ª fase - os estudantes participam livremente das discussões no fórum; e 3ª fase - os estudantes manifestam o seu posicionamento final na enquete após o debate.

Os dados são coletados diretamente dos fóruns do AVA, por meio de um software de planilha eletrônica e organizados em quatro tabelas: (1) **Nodes Table** - é a tabela de nós da rede - cada nó possui um identificador, um label, e um tipo (posicionamento e participante); (2) **Adjacency List/posicionamento inicial** - possui a lista de participantes para cada um dos quatro posicionamentos possíveis, antes das discussões; (3) **Adjacency List/discussão**: cada linha representa um fio de discussão contendo uma lista com os seus participantes; (4) **Adjacency List/posicionamento final**: possui a lista de participantes para cada um dos quatro posicionamentos possíveis, depois das discussões.

As quatro tabelas são importadas para o software Gephi⁵ a fim de formar uma rede bipartida de participantes e posicionamentos. Após a aplicação de uma projeção bipartida, tem-se uma rede monopartida de participantes que, por meio de um algoritmo de clusterização do Gephi, são revelados os grupos de afinidade entre os alunos.

O cenário da segunda prática ocorre na disciplina optativa Ciência de Dados onde os alunos constroem uma planilha coletiva de coleta de dados sobre diversas informações de sua vida acadêmica e/ou particulares. A planilha é importada no software OpenRefine⁶ para limpeza e padronização dos dados. Em seguida, os dados são mapeados para o formato de rede GML⁷. O mapeamento feito no OpenRefine é transportado para o software Gephi para

⁵ GEPHI é um software de código aberto utilizado para visualização, análise e manipulação de redes e grafos. Disponível em <https://gephi.org/>.

⁶ O software OpenRefine é uma ferramenta de código aberto utilizada para a limpeza e transformação de dados. Disponível em <https://openrefine.org/>.

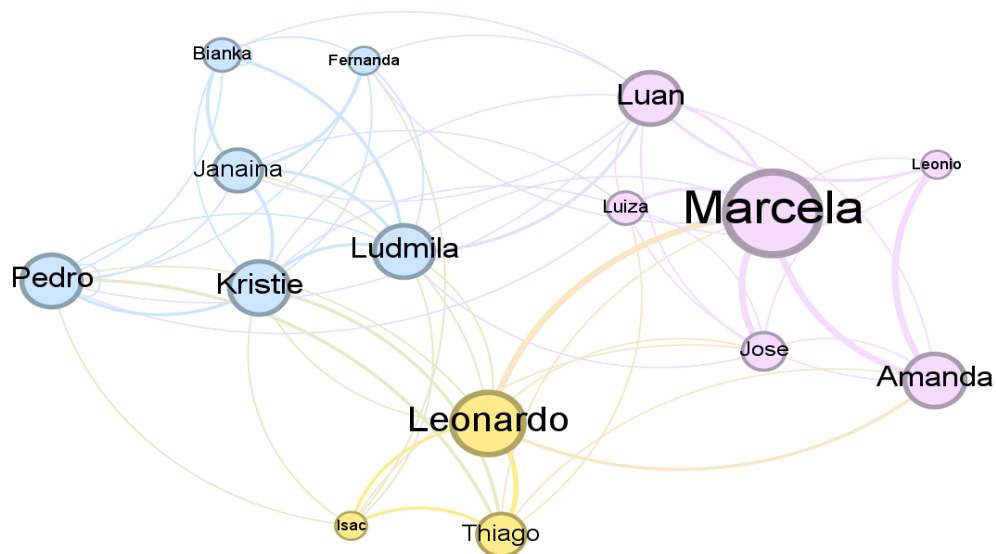
⁷ GML (*graph modelling language*) é um formato para representação de grafos de fácil leitura por humanos e com uma capacidade semântica razoável para configurar as características da rede, dos nós e das arestas. Disponível em: https://en.wikipedia.org/wiki/Graph_Modelling_Language/.

as formatações e projeções bipartidas. Finalmente, aplica-se um algoritmo de clusterização do Gephi sobre a rede monopartida de participantes para revelação de grupos de afinidade.

5 RESULTADOS E DISCUSSÃO

Um dos fóruns aplicados levantou a questão "Você é favorável à ideia de que ações de combate à desinformação podem limitar a liberdade de expressão?". Após um tempo delimitado de participação dos estudantes da turma, houve a coleta de dados das discussões e das sondagens inicial e final. Em seguida a rede bipartida de posicionamentos e participantes foi formada. Logo em seguida, com a projeção bipartida, obteve-se a rede monopartida de participantes com destaque de três grupos de afinidades dando ênfase no tamanho do nó aos participantes que mais participaram com postagens, como mostra a Figura 1.

Figura 1 – Rede monopartida de participantes com grupos de afinidades.

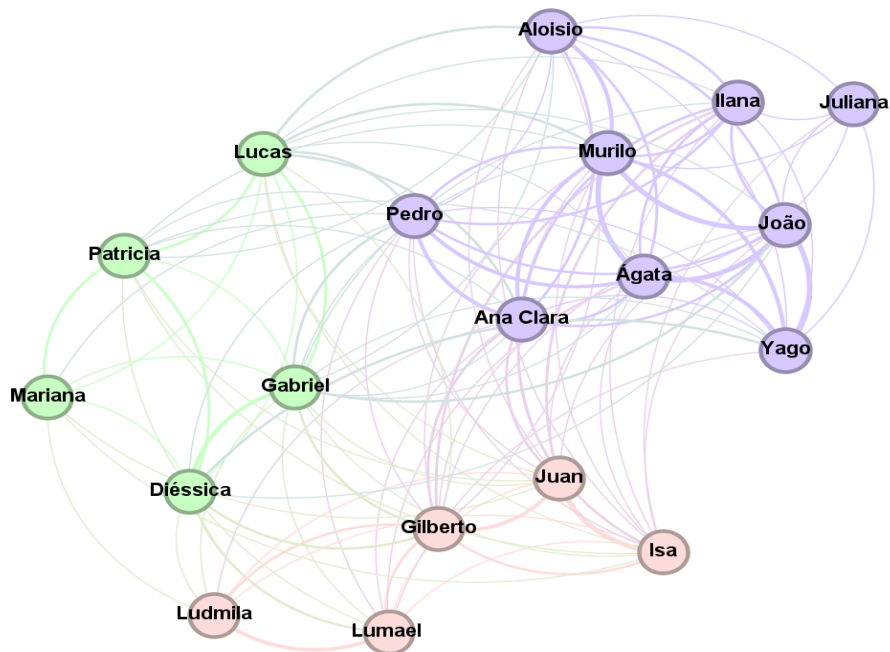


Fonte: Elaborado pelos autores

O resultado de uma aplicação da segunda prática de ARS pode ser visto na Figura 2. Essa ação foi desenvolvida a partir da coleta dos seguintes dados dos alunos: município de moradia, curso, período do curso, interesse de pesquisa e gênero de filme que mais gosta. A rede monopartida revelou três grupos de afinidades.

Observou-se que as práticas proporcionaram aos alunos uma vivência com a ARS de uma forma diferenciada, pois eles foram os próprios atores nas redes sociais analisadas. Esse sentimento de pertencimento e a curiosidade de se achar em um dos clusters formados nas redes trouxe uma motivação extra para maior familiaridade com as ferramentas e, conseqüentemente, o desenvolvimento da *data literacy* em CD.

Figura 2 – Rede monopartida e clusterizada obtida pela segunda prática de ARS



Fonte: Elaborado pelos autores

Pode-se destacar que, entre as sete competências sinalizadas por Pothier e Condon (2020), no âmbito da *data literacy*, houve: (1) desenvolvimento de atividades de organização e preparação para armazenagem de dados, (2) o entendimento da importância dos dados coletados e o quanto eles são determinantes para a descoberta dos grupos de afinidades entre os alunos, (3) avaliação da qualidade dos dados coletados e a sua correção, (4) interpretação de dados por meio de inspeção visual nas redes sociais formadas, (5) ensaio de tomada de decisão, a partir dos dados interpretados nas redes resultantes, sobre possíveis orientações aos participantes quanto aos seus posicionamentos perante as questões polêmicas, (6) apresentação dos dados finais com apoio de *storytelling*, e (7) discussão sobre a ética no que tange ao respeito sobre a diversidade de posicionamentos e quanto a segurança com dados por meio de processo de pseudonimização.

Além disso, o conhecimento descoberto por meio da ARS, enquanto técnica da mineração de dados na CD, isto é, os grupos de afinidades entre os alunos, possibilitou uma percepção, ainda que simples, do funcionamento de parte dos algoritmos de IA (inteligência artificial) quanto a criação de perfis de usuários pelas mídias sociais para sugestionamento automático de postagens, anúncios e conexões potenciais com outros membros da rede.

6 CONSIDERAÇÕES FINAIS

Foi realizado um levantamento bibliográfico das relações entre *data literacy*, CD, ARS e CI, como ponto de partida para a discussão proposta no objetivo da pesquisa. A condução das duas práticas de ARS, que não necessitam de conhecimento avançado em tecnologia, proporcionaram um desenvolvimento factível de *data literacy* em CD aos estudantes de áreas da CI. Além disso, foi possível verificar benefícios da prática de ARS para a melhoria da *data literacy* em CD pelos estudantes.

Sendo os próprios alunos os atores nas redes sociais analisadas, houve um bom entendimento de aspectos da ARS enquanto processo de mineração de dados em CD. Houve também um entendimento básico do funcionamento de algoritmos de inteligência artificial quanto a criação de perfis de usuários pelas mídias sociais usado no sugestionamento automático de postagens, anúncios e novos membros, e ainda a percepção de melhoria de *data literacy* pelos estudantes.

Destaque para a primeira prática, baseada na coleta de dados dos fóruns de discussão, por poder ser inserida em ambientes de ensino-aprendizagem habituais em diversas disciplinas de currículos em áreas da CI, necessitando apenas de conhecimento básico em tecnologia e o uso de fóruns virtuais como apoio ao ensino presencial. Essa prática possibilita, com uma carga horária pequena, o desenvolvimento de *data literacy* abrindo a discussão para vários aspectos contemporâneos da CD.

As práticas relatadas podem ser aplicadas em outros contextos educacionais além da graduação, por exemplo na pós-graduação em Ciência da Informação. Nesse caso, a ARS pode ser benéfica para auxiliar pesquisas diversas na CI, desde levantamentos bibliométricos de referenciais teóricos, até na análise de dados em redes de informação oriundas de coleta de dados de formulários *online* ou de bases de dados abertas no contexto da pesquisa.

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.

BELLUZZO, Regina Célia Baptista; FERES, Glória Georges; BASSETTO, Clemilton Luis. A competência em informação como um fator crítico de sucesso para a pesquisa na área de ciência da informação: transferência de princípios para reflexão. **Revista EDICIC**, v. 1, n. 1, 2011. Disponível em: <http://www.edicic.org/revista/index.php?journal=RevistaEDICIC&page=article&op=view&path%5B%5D=21>. Acesso em: 3 jul. 2023.

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.

BROOKS, David. The Philosophy of Data. **The New York Times**, 2013. Disponível em: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>. Acesso em: 3 jul. 2023.

CHEN, Chaomei. **Mapping scientific frontiers: the quest for knowledge visualization**. 2. ed. London: Springer Science & Business Media, 2013.

CRISTOVÃO, Henrique Monteiro; FERRARI, Luciana Itida; MORAES, Margarete Farias De. Ciência de Dados na graduação em Arquivologia. **ÁGORA: Arquivologia em debate**, v. 32, n. 65, p. 1–21, 2022. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/1141>. Acesso em: 3 jul. 2023.

CURTY, Renata Gonçalves; SERAFIM, Jucenir da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. **Informação & Informação**, Londrina, PR, v. 21, n. 2, p. 307–331, 2016. Disponível em: [10.5433/1981-8920.2016v21n2p307](https://doi.org/10.5433/1981-8920.2016v21n2p307). Acesso em: 3 jul. 2020.

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.

GAO, Man; CHEN, Ling; LI, Bin; LI, Yun; LIU, Wei; XU, Yong-cheng. Projection-based link prediction in a bipartite network. **Information Sciences**, v. 376, p. 158–171, 2017. Disponível em: <https://doi.org/10.1016/j.ins.2016.10.015>. Acesso em: 3 jul. 2023.

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

HARARI, Yuval Noah. **Homo Deus**. 1. ed. São Paulo: Companhia das Letras, 2016.

HIGGINS, Silvio Salej; RIBEIRO, Antonio Carlos Andrade. **Análise de redes em Ciências Sociais**. Brasília: Enap, 2018.

KADUSHIN, C. Introduction to social network theory. **Boston, MA**, 2004. Disponível em: <http://melander335.wdfiles.com/local--files/reading-history/kadushin.pdf>. Acesso em: 3 jul. 2023.

LAGE, Kathryn; LOSOFF, Barbara; MANESS, Jack. Receptivity to library involvement in scientific data curation: a case study at the University of Colorado Boulder. **Libraries and the Academy**, [S. l.], v. 11, n. 4, p. 915–937, 2011. Disponível em: [10.1353/pla.2011.0049](https://doi.org/10.1353/pla.2011.0049). Acesso em: 3 jul. 2023.

LEMOS, Daniela Lucas da Silva; MARTINS, Dalton Lopes; CARMO, Danielle Do. Padrões de qualidade para dados e metadados endereçados a aplicações em Ciência de Dados.

Advanced Notes in Information Science, v. 2, p. 161–170, 2022. Disponível em: <https://pub.colnes.org/index.php/anis/article/view/116/127>. Acesso em: 3 jul. 2023.

LYON, L.; MATTERN, E.; ACKER, A.; LANGMEAD, A. Applying translational principles to data science curriculum development. *Em: IPRES 2015, PROCEEDINGS 2015, Chapel Hill, North Carolina. Anais [...]. Em: IPRES 2015. Chapel Hill, North Carolina: University of Pittsburgh, 2015.* Disponível em: <http://d-scholarship.pitt.edu/27159/>. Acesso em: 3 jul. 2023.

MADEIRO, Andre França; DIAS, Guilherme Ataíde. Arquivista de dados: análise do perfil profissional em sites de bancos internacionais de trabalho. **ÁGORA: Arquivologia em debate**, v. 30, n. 61, p. 649–664, 2020. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/924>. Acesso em: 3 jul. 2023.

MARKUS, Gregory et al. **AI in relation to GLAMs Task Force Report: report and recommendations**. [s.l.] : europeana network association, 2021. Disponível em: <https://pro.europeana.eu/project/ai-in-relation-to-glams>. Acesso em: 3 jul. 2023.

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.

MARTINS, Dalton Lopes. Data science teaching and learning models: focus on the Information Science area. *Em: RODRIGUES DIAS, Thiago Magela (org.). Advanced Notes in Information Science*. [s.l.] : ColNes Publishing, 2022. v. 2. Disponível em: <https://pub.colnes.org/index.php/anis/article/view/100>. Acesso em: 3 jul. 2023.

MATHEUS, Renato F.; SILVA, Antonio B. 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.

MATOS, Maurício Torres De; CONDURÚ, Marise Teles; BENCHIMOL, Alegria Celia. Interseções na produção científica da ciência da informação e ciência de dados. **Acervo**, Rio de Janeiro, RJ, v. 35, n. 2, p. 1–18, 2022. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1804/>. Acesso em: 3 jul. 2023.

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. **Big Data: the essential guide to work, life and learning in the Age of Insight**. [s.l.] : Hachette UK, 2013.

MELAMED, David. Community Structures in Bipartite Networks: A Dual-Projection Approach. **PLOS ONE**, v. 9, n. 5, p. e97823, 2014. Disponível em: <https://doi.org/10.1371/journal.pone.0097823>. Acesso em: 3 jul. 2023.

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

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.

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>.

POTHIER, Wendy Girven; CONDON, Patricia B. Towards data literacy competencies: business students, workforce needs, and the role of the librarian. **Journal of Business & Finance Librarianship**, v. 25, n. 3–4, p. 123–146, 2020. Disponível em: <https://doi.org/10.1080/08963568.2019.1680189>. Acesso em: 3 jul. 2023.

RASTELI, Alessandro; CAVALCANTE, Lídia 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.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital e dados de pesquisa. **Atoz: novas práticas em informação e conhecimento**, v. 5, n. 2, p. 67–71, 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 3 jul. 2023.

SILVA, Fabiano Couto Corrêa Da. O papel dos bibliotecários na gestão de dados científicos. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 14, n. 3, p. 387–406, 2016. Disponível em: [10.20396/rdbci.v14i3.8646333](https://doi.org/10.20396/rdbci.v14i3.8646333). Acesso em: 3 jul. 2023.

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/iquvol282_3shields.pdf. Acesso em: 3 jul. 2023.

SOUZA, Queila; QUANDT, Carlos. Metodologia de análise de redes sociais. *Em*: DUARTE, F.; QUANDT, Carlos; SOUZA, Queila (eds.). **O Tempo das redes**. São Paulo: Perspectiva, 2008. p. 31–63. Disponível em: https://www.academia.edu/257818/Metodologia_De_An%C3%A1lise_De_Redes_Sociais. Acesso em: 3 jul. 2023.

VALE, Lucas; CRISTOVÃO, H. M. Visualização de informação sobre preços de medicamentos da base de dados abertos da ANVISA com auxílio de análise de redes de informação. **Revista Ibero-Americana de Ciência da Informação**, v. 16, n. 1, p. 206–225, 2023. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/47582>. Acesso em: 3 jul. 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.

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

WILKINSON, Mark D. et al. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, p. 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 3 jul. 2023.

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**, v. 12, n. 3, 2016. Disponível em: <https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3275>. Acesso em: 3 jul. 2023.

ZAKARIA, Mahmoud Sherif. Data science education programmes in middle eastern institutions: a survey study. **IFLA Journal**, v. 49, n. 1, p. 157–179, 2023. Disponível em: <https://doi.org/10.1177/03400352221113362>. Acesso em: 3 jul. 2023.