CIÊNCIA EM EVIDÊNCIA
MINERAÇÃO DE DADOS NO ÂMBITO DOS FILTROS BOLHA NO COMPARTILHAMENTO DE FAKE NEWS: REVISÃO DA LITERATURA E PROPOSTA DE MECANISMOS DE PREVENÇÃO
ISSN 2763-5457, V.3, N.1, E022004, 2022.

MINERAÇÃO DE DADOS NO ÂMBITO DOS FILTROS BOLHA NO COMPARTILHAMENTO DE FAKE NEWS: REVISÃO DA LITERATURA E PROPOSTA DE MECANISMOS DE PREVENÇÃO

Marco Antônio G. Carvalho
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP, Capivari, SP, Brasil.
marcogarciacarvalho@gmail.com

Profª. Drª. Bruna C. R. Cunha
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP, Capivari, SP, Brasil.
bruna.rodrigues@ifsp.edu.br

Resumo

Este trabalho buscou identificar como a mineração de dados (MD) e algoritmos de filtros bolha são utilizados a favor de notícias falsas e maliciosas, denominadas *fake news*. Além de traçar uma ligação entre a MD e a propagação de notícias falsas no âmbito das mídias sociais, buscou-se identificar mecanismos de prevenção e proteção. Para este fim, foi conduzida uma revisão sistemática pois esta possibilita identificar e sintetizar publicações relevantes. A mineração de dados pode ser utilizada para identificar relacionamentos sistemáticos entre variáveis, detectando subconjuntos de dados. As mídias sociais fazem uso desses subconjuntos para fragmentar seus usuários em "bolhas" nas quais eles têm contato somente com assuntos pré-determinados por seus interesses. Neste ambiente, a maioria dos usuários não possui conhecimentos necessários para discernir quando se trata de uma notícia maliciosa, visto que não recebem conteúdos divergentes. Entre os recursos identificados para o combate á desinformação destaca-se a necessidade de uma conscientização sistemática do público sobre o consumo de notícias nas mídias online. Alternativas tecnológicas envolvem o reconhecimento automático de padrões em *fake news*. Conclui-se que é imprescindível o fortalecimento e a reconstrução das conexões dos usuários com fontes confiáveis.

Palavras-chave: Filtros Bolha; Mineração de Dados; Notícias Falsas; Revisão Sistemática.

DATA MINING IN THE SCOPE OF FILTER BUBBLES FOR SHARING FAKE NEWS: LITERATURE REVIEW AND PROPOSAL OF PREVENTION MECHANISMS

Abstract

In this work our goal was to identify how data mining (DM) and filter bubble algorithms are used in favor of false and malicious news, denominated fake news. In addition to drawing a link between DM and the propagation of fake news within the scope of social media, we aimed to identify prevention and protection mechanisms. For this purpose, we conducted a systematic review, as it enables to identify and synthesize relevant publications. Data mining can be used to identify systematic relationships between variables, thus detecting subsets of data. Social media makes use of these subsets to fragment their users into "bubbles" in which they only have contact with subjects predetermined by their interests. In this environment, most users do not have the necessary knowledge to discern when it comes to malicious news, as they do not receive divergent content. Among the resources identified for combating fake news, many highlight the need for a systematic awareness of the public about the consumption of news in online media. Technological alternatives involve automatically recognition of patterns used within fake news. Finally, it is essential to strengthen and rebuild users' connection with trustful sources.

Keywords: Data Mining; Fake News; Filter Bubble; Systematic Review.

1. INTRODUÇÃO

Com o advento da tecnologia, novos métodos de compartilhamento de informações vêm emergindo, dentre eles, as tecnologias mediadas pela Internet. Por meio dessa, usuários podem compartilhar notícias com facilidade. Enquanto, por um lado, isso facilita o acesso de pessoas à informação, por outro, tal cenário é propício para geração e disseminação de desinformação (PETERS, 2017). Em 2013, o *Global Risks Report* (WORLD ECONOMIC FORUM, 2013) já alertava sobre os riscos da massiva desinformação digital, a qual poderia causar impactos negativos na sociedade (KORTA, 2018). O entendimento do funcionamento e do efeito das redes sociais trata-se de um assunto de grande relevância para a sociedade, sendo que alguns autores apontam que o uso intensivo das redes sociais favorece a formação de perfis marcados pela intolerância e pelo radicalismo, os quais podem ser interpretados como sinais de ameaça ao estado democrático de direito (QUADRADO; FERREIRA, 2020).

Uma das principais formas de compartilhar informações na Internet é por meio das redes sociais. Como exposto por Recuero, Zago e Soares, (2019), o Brasil utiliza o *Facebook*

e o *Twitter* como principais meios de compartilhamento de notícias. Ambas as redes sociais utilizam filtros para designar e filtrar conteúdo para seus usuários, como descrito por Sastre e Belda (2018). Segundo Pariser (2011), "filtro bolha" é o termo utilizado para definir os algoritmos das redes sociais que funcionam como filtros no ambiente virtual, os quais atuam como motores de previsão para redirecionar conteúdos baseados em perfil, hábitos, lista de amigos interações pertinentes de usuários. Esses dados, quando trabalhados, tornam-se valiosos para organizações, seja para o planejamento, controle, previsão e tomada de decisões, ou para a segmentação de pessoas e comportamentos. A mineração de dados é uma técnica que permite explorar esses dados, extraindo informações ao encontrar padrões e anomalias (PAULA AVELAR, DE; NALDI, 2017). Do ponto de vista do usuário, a rede social gera a sensação de eficiência no direcionamento de conteúdo e informações, porém, ao mesmo tempo, a mesma restringe de forma enviesada a entrega de conteúdo. Métodos de *big data* podem tornar mais fácil encontrar evidências para apoiar qualquer ponto de vista que se deseja propagar (CLEVERLEY, 2017), por causa desse fenômeno, esses algoritmos são chamados de "filtros bolha" pois isolam os usuários de outros conteúdos e reforçam tópicos já pesquisados e enraizados pelos grupos sociais aos quais pertencem.

Notícias deliberadamente falsas e criadas com o intuito de enganar o público foram cunhadas pela mídia e sociedade de *fake news*, um termo relativamente recente (GELFERT, 2018) porém, com impactos profundos na atual conjuntura da sociedade. A associação entre política e a disseminação de notícias falsas é notória, até mesmo no contexto da pandemia de COVID-19, em que notícias nocivas, como a promoção de medicamentos sem eficácia comprovada cientificamente (CERON; DE-LIMA-SANTOS; QUILES, 2021) e a propaganda de argumentos anti-vacina (MONTAGNI et al., 2021), foram deliberadamente introduzidas nos ambientes de socialização virtual. Por se tratar de um meio de influenciar a realidade política de forma geral, é essencial que o funcionamento desses algoritmos seja de conhecimento público e acessível para a comunidade. Assim, a fim de analisar a aplicação de algoritmos de mineração de dados em redes sociais de forma crítica e identificar possíveis alternativas, sociais e tecnológicas, para sua prevenção, foi realizada uma revisão sistemática da literatura cujos resultados são expostos neste artigo.

2. METODOLOGIA

A metodologia utilizada para desenvolvimento foi a revisão sistemática pois, como evidenciado por De-la-Torre-Ugarte-Guanilo, Takahashi e Bertolozzi (2011), trata-se de um

método que possibilita identificar as melhores evidências de forma estruturada e sintetizá-las para fundamentar propostas de mudanças. Caiado *et al*. (2016) propõem que uma Revisão Sistemática da Literatura (RSL) deve compreender sete etapas: (i) formular o problema, (ii) localizar e selecionar os estudos, (iii) avaliar a qualidade dos estudos, (iv) coletar dados, (v) analisar e apresentar os resultados, (vi) interpretar os resultados e (vii) melhorar e atualizar as revisões.

Considerando os procedimentos de RSL mencionados (CAIADO et al., 2016), esse trabalho buscou responder a seguinte questão: "Como a mineração de dados e os algoritmos de filtros bolha são usados a favor das *fake news* e quais são os mecanismos de prevenção?". A base de busca utilizada para o desenvolvimento dessa pesquisa foi o Portal de Periódicos da CAPES. A chave de busca utilizada para a pesquisa dos artigos considerou os termos "Data *Mining*", "*Filter Bubble*" e "*Fake News*", sendo que os três termos precisavam estar presentes no texto dos artigos, sem restrição de data de publicação. Para a seleção dos artigos foi utilizado uma metodologia de cinco passos para identificar o conteúdo e determinar se o artigo deveria ser incluído ou excluído da seleção. A metodologia seguiu os seguintes passos:

1. Leitura do resumo
2. Identificação de pontos de interesse
3. Leitura da conclusão
4. Pesquisa dos termos de busca no texto
5. Leitura completa

Caso algum trabalho não respeitasse os critérios de exclusão e inclusão, em qualquer etapa, era feita sua remoção. Foram retornados 57 artigos e, após a análise, foram selecionados 13 artigos, conforme Quadro 1. Após isso, um resumo sobre os principais temas abordados por cada artigo aceito foi realizado para apoiar a etapa de discussão.

3. RESULTADOS

Nesta seção, os resultados da revisão, listados no Quadro 1, são utilizados para fundamentar uma discussão sobre como a mineração de dados e os algoritmos de filtros bolha favorecem a propagação de *fake news* e quais são os mecanismos de prevenção propostos pelos autores supracitados. Outros artigos relacionados também foram incluídos a fim de enriquecer a argumentação. São discutidos os impactos da mineração de dados no consumo de conteúdo nas redes sociais e como pessoas mal-intencionadas utilizam os mecanismos dessas redes para polarizar a opinião de usuários por meio da disseminação de *fake news*. Por fim,

será abordado como as redes sociais podem implementar mecanismos para se prevenir e como eles funcionam. Destaca-se que este artigo foca no âmbito tecnológico, de modo que métodos de construção de narrativas que compõem notícias falsas e persuasivas não são abordados (BAPTISTA, 2020). No entanto, o entendimento da construção de *fake news* é fundamental para a construção de mecanismos de análise e prevenção automática, como discutido nas subseções.

Quadro 1 - Resumo dos trabalhos resultantes da Revisão Sistemática.

Autor(es)/Ano	Mídias Investigadas	Solução Proposta
BOZDAG (2013)	Facebook e Twitter	Adição de filtro nos algoritmos para que não gerem engajamento caso uma notícia falsa seja identificada como falsa.
CLEVERLEY (2017)	Google	Reestruturação do algoritmo de busca para evitar difusão de *fake news*.
PETERS (2017)	Facebook e Twitter	Conscientização social e educação dos usuários.
KATSIREA (2018)	Facebook e Twitter	Criminalização em casos de compartilhamento proposital.
KORTA (2018)	Facebook, Twitter e YouTube	Conscientização dos usuários no uso das mídias digitais.
SASTRE; DE OLIVEIRA; BELDA (2018)	Facebook	Não propõe solução.
MASIP; RUIZ-CABALLERO; SUAU (2019)	Facebook e Twitter	Conscientização social para que os usuários saibam identificar notícias falsas.
RECUERO; ZAGO; SOARES (2019)	Twitter	Conscientização dos usuários e, principalmente, dos influenciadores digitais sobre suas práticas.
REVIGLIO (2019)	Facebook e YouTube	Não propõe solução.
GUARINO et al. (2020)	Twitter	Os autores propuseram um *framework* sintático para análise da origem das *fake news*.
KAUFHOLD et al. (2020)	Facebook, Twitter e YouTube	Sistema de análise de qualidade de publicações para emissão de alertas em situações de crise.
MANFREDI-SÁNCHEZ (2020)	Twitter	Reformulação da distribuição das notícias dentro das mídias sociais e separação de

		notícias de publicações sociais para facilitar o processo de reconhecimento.
PIÑEIRO-OTERO; ROLÁN (2020)	Facebook, Twitter, YouTube, Instagram e TikTok	Conscientização dos usuários e um combate a propagandas políticas online.

3.1. A mineração de dados e seu impacto nas redes sociais

A mineração de dados é uma técnica que auxilia superar as dificuldades de manipulação e transformação de dados em informações relevantes (PAULA AVELAR, DE; NALDI, 2017). Quando corretamente analisados, dados tornam-se valiosos para organizações, seja para planejamento, controle, previsão ou tomada de decisão, servindo também para segmentação de pessoas e comportamentos (PAULA AVELAR, DE; NALDI, 2017). De acordo com Sastre, De Olivera e Belda (2018), mecanismos de busca de grandes empresas, como o *Facebook* e o *Twitter*, utilizam a mineração de dados para criar algoritmos que filtram os resultados de suas pesquisas e preveem futuros interesses, direcionando o acesso ao conteúdo baseado em perfil, hábitos e consumos de dados dos usuários. Esses algoritmos são denominados de "filtros bolha".

Com o intuito de gerar a sensação de eficácia para o usuário, que a todo momento recebe conteúdos direcionados, dados são extraídos e trabalhados em tempo real nas mídias sociais (GUARINO et al., 2020). Essa extração de dados ocorre principalmente durante o uso das redes sociais, considerando as publicações com as quais o usuário mais interage, e através de dados fornecidos pelo próprio usuário em seu perfil (KATSIREA, 2018). Ainda, segundo acusações da mídia, como em uma reportagem do BBC de 2016 (KLEINMAN, 2016), outra forma de extração de dados poderia ocorrer por meio do áudio do microfone dos *smartphones* para entregar publicidades sobre determinados assuntos mencionados em conversas particulares. Apesar das acusações não possuírem evidências acadêmicas empíricas, alguns autores afirmam que não é possível descartar a possibilidade de implementação de tais mecanismos de escuta (KRÖGER; RASCHKE, 2019).

A maior parte dos usuários não possui conhecimentos sobre a existência dessa extração e processamento de dados, por conta disso, não ficam cientes do efeito do "filtro bolha". A falta de contato com publicações antagônicas e opinião de conexões sociais mais diversificadas limita o julgamento dos usuários. Como resultado, a sociedade torna-se mais polarizada e menos consciente de opiniões divergentes (MACHADO; MISKOLCI, 2019). Em uma perspectiva polarizada, o pluralismo da mídia acaba se enfraquecendo por causa dessa

personalização, tornando as pessoas politicamente intransigentes, além de vulneráveis às censuras, propagandas e notícias maliciosas (REVIGLIO, 2019).

3.2. *Fake news*: desinformação, notícias maliciosas e a polarização social

De acordo com a organização norte-americana Freedom House, o termo *fake news* pode ser definido como "informações intencionalmente falsas que foram projetadas para parecerem com notícias legítimas e atrair o máximo de atenção." (KELLY et al., 2017). Em 2013, o *Global Risk Report* (Relatório Global de Riscos publicado pelo Fórum Econômico Mundial) citou as *fake news* como prejudiciais para a sociedade e como um meio de causar potenciais impactos negativos na vida de indivíduos e sociedade (WORLD ECONOMIC FORUM, 2013). Como é exemplificado por Masip, Ruiz-Caballero e Suau (2019), a função dessas notícias falsas é ocasionar uma polarização entre usuários nas mídias sociais e, com isso, provocar campanhas de desinformação e, assim, reforçar uma ideia, um projeto e até mesmo um lado político. Porém, conforme relatado por um artigo recente publicado na revista Science (LAZER et al., 2018), a geração e a difusão viral de notícias falsas permanecem em grande parte um problema em aberto, sem uma campanha efetiva de solução.

Todo esse ambiente, propenso ao compartilhamento de desinformação, é ainda mais acentuado por algoritmos de "filtros bolha" – os quais isolam os usuários de receberem informações diversas e antagônicas – pois perfis que compartilham notícias falsas não condizem com aqueles que compartilham a notícia opositora e verdadeira (GUARINO et al., 2020). Nesse âmbito, Reviglio (2019) discute a ideia de que indivíduos podem reduzir seu empoderamento informacional ("*informational empowerment*") e as sociedades se tornam politicamente mais polarizadas. Uma sociedade polarizada enfraquece o pluralismo da mídia e torna as pessoas mais vulneráveis à censura e propaganda ou, melhor, à autocensura e autopropaganda (SUNSTEIN, 2017).

A eleição presidencial dos Estados Unidos de 2016 marcou verdadeiramente a transição de uma idade de "pós-confiança" para uma era de "pós-verdade" (LIBERINI et al., 2020). Pós-verdade é o termo que qualifica a circunstância em que processo de formação de opinião ignora fatos objetivos, sendo diretamente influenciado apelo emocional e crenças pessoais. O *Facebook* e outras plataformas de mídia social negaram que seus algoritmos de publicidade desempenharam um papel no resultado da eleição presidencial. Contudo, após a comprovação e o agravamento das acusações, seus representantes admitiram que uma possível interferência eleitoral pode ter ocorrido por meio de "falhas" das plataformas. Essa interferência ocorreu, principalmente, por meio de *fake news* que serviram para perpetuar

boatos, rumores e desinformação com o intuito de acentuar a polarização política dentro do país (KORTA, 2018). Esses boatos se espalharam intensamente, pois foram direcionados aos usuários mais propensos a compartilhar informações falsas e polarizadas rapidamente, especialmente quando relacionadas à política (VOSOUGHI; ROY; ARAL, 2018).

Ademais, o termo "*fake news*" tornou-se uma representação de dúvida, ceticismo e falta de confiança dos cidadãos em relação às informações, bem como uma ferramenta utilizada pela elite econômica para minar os fatos e abusar de seu poder (KORTA, 2018). Essa desconfiança faz com que indivíduos não se preocupem tanto com a veracidade da informação que estão compartilhando quando essas reforçam os seus ideais, ao mesmo tempo que notícias verídicas, muitas vezes publicadas pela grande mídia, são referidas como *fake news* simplesmente por não serem consoantes com as convicções de seus consumidores (MASIP; RUIZ-CABALLERO; SUAU, 2019). Reviglio (2019) também aponta que há uma linha tênue entre "notícias falsas" e "sátira" e que essas podem ser muitas vezes confundidas, podendo lançar dúvidas sobre formas legítimas de expressão e tornando mais grave a crise de confiança na comunicação jornalística.

3.3. Métodos de combater e evitar *fake news*

A propagação de notícias tendenciosas funciona bem dentro das mídias sociais pois as conexões entre diversos sites externos são fracas e criam um falso senso de legitimidade para conteúdos ilegítimos. Assim, um dos métodos para combater essas notícias é enfraquecer o algoritmo de compartilhamento de publicações de fontes não verificadas e fortalecer o de fontes verificadas, de forma que os usuários teriam acesso às notícias de fontes confiáveis primeiramente e reconstruiriam a confiança no que se diz respeito à informação (notícias e pesquisas) e expressão (mídia social) (KORTA, 2018).

Existe uma variedade de métodos para detecção de eventos ou agrupamento de mensagens e para extrair informações situacionais em grandes fluxos de informação (KAUFHOLD et al., 2020). Um desses métodos é a utilização de um *framework* sintático de baixo nível voltado para a detecção de padrões dentro das *fake news* por meio de verificações de cobertura de palavras mais utilizadas durante a propagação de desinformação (GUARINO et al., 2020). Notícias maliciosas tendem a seguir uma linha de raciocínio comum, de forma que o reconhecimento de padrões pode auxiliar no processo de identificação. Baptista (2020), por exemplo, argumenta que esse tipo de notícia têm uma ação persuasiva heurística, priorizando o apelo emocional por meio de sentimentos que se sobrepõem à razão e ao raciocínio. Ainda segundo o autor, *fake news* utilizam uma linguagem simples e textos curtos,

ao contrário das notícias que validam os seus argumentos com fatos, evidências e citações. Tais características facilitam um processo de reconhecimento. Contudo, para que o *framework* possa funcionar, é necessário que, em um primeiro momento, os usuários denunciem os conteúdos tendenciosos, para assim criar uma base confiável em que o algoritmo possa aprender e definir um conjunto de palavras e estilo de escrita. Por necessitar dessa base de treinamento, essa solução não é efetiva de imediato (GUARINO et al., 2020), além de não resolver o problema de confiabilidade dos usuários, sendo aplicada apenas como uma medida preventiva.

Outra ferramenta que poderia ser adotada pelas mídias sociais, proposta por Kaufhold et al. (2020), é um algoritmo para verificar todo conteúdo postado nas redes e, durante eventos de crise e emergência, gerar alertas aos usuários com as publicações de maior relevância, qualidade e confiabilidade. O mesmo método poderia ser aplicado de forma "invertida", ou seja, a abordagem poderia ser adaptada para emitir alertas ao identificar notícias falsas de grande alcance, de forma a entregar informações que se contrapõem a bolha gerada pela rede social de cada indivíduo. A abordagem foi projetada para momentos em que há sobrecarga de informações e apresenta oportunidades para melhoria de sua performance, de modo que, segundo os autores, sua replicação para o contexto de *fake news* em redes sociais apresenta viabilidade.

Manfredi-Sánchez (2020), por outro lado, defende que é necessário separar notícias (*i.e.*, links externos) de publicações comuns, geradas pelos usuários pois, assim, a rede social propiciaria uma maior facilidade de verificação e garantiria aos usuários um ambiente de maior confiabilidade e qualidade no que se refere ao consumo de notícias.

Cleverley (2017) apresenta uma solução mais drástica e propõe uma reestruturação do algoritmo de compartilhamento de notícias para garantir que os usuários possam receber primeiramente conteúdos de fontes confiáveis e verificadas e somente depois receber as notícias de outras fontes. O objetivo é fornecer aos usuários um contato prioritário com fontes confiáveis para que possam formar sua opinião com base em um jornalismo profissional e ético.

Ainda nesse sentido, Kaufhold et al. (2020) alegam que é necessário realizar uma verificação de todas as postagens dentro das mídias sociais a fim de garantir que notícias falsas e maliciosas sejam impedidas de serem propagadas. Os autores sugerem que, dentro das redes sociais, todas as postagens e comentários sejam analisados por um filtro para garantir se há autenticidade nos fatos expostos. A aplicabilidade legal e viabilidade computacional da proposta não foram estudadas.

Dentro das soluções encontradas na pesquisa, um ponto fortemente destacado por vários autores, é que os usuários necessitam de uma conscientização sobre o consumo de informação dentro das redes sociais e como isso os impacta. Korta (2018) e Peters, (2017) defendem que os usuários deveriam ser ensinados sobre os impactos das mídias sociais antes mesmo de usá-las.

Korta (2018) argumenta que um fator de suma importância é reestruturar nosso sistema educacional e a conscientizar os usuários, especialmente no que se refere ao consumo de fontes online, para que assim haja uma maior eficácia no combate à desinformação. Bozdag (2013) destaca que as pessoas devem ser educadas no uso das tecnologias da informação e ter acesso suficiente aos meios de informação, para que assim possam participar da vida comum em sociedade. Os autores argumentam que o governo e as grandes corporações usam a mídia de massa para ajudar a informar e doutrinar a população em geral sobre seus ideais. Uma sociedade mais consciente e ciente dos reais impactos da polarização social e dos problemas causados pela disseminação das *fake news* garantirá uma maior consciência dentro das mídias, inclusive para novos usuários (REVIGLIO, 2019).

4. DISCUSSÃO

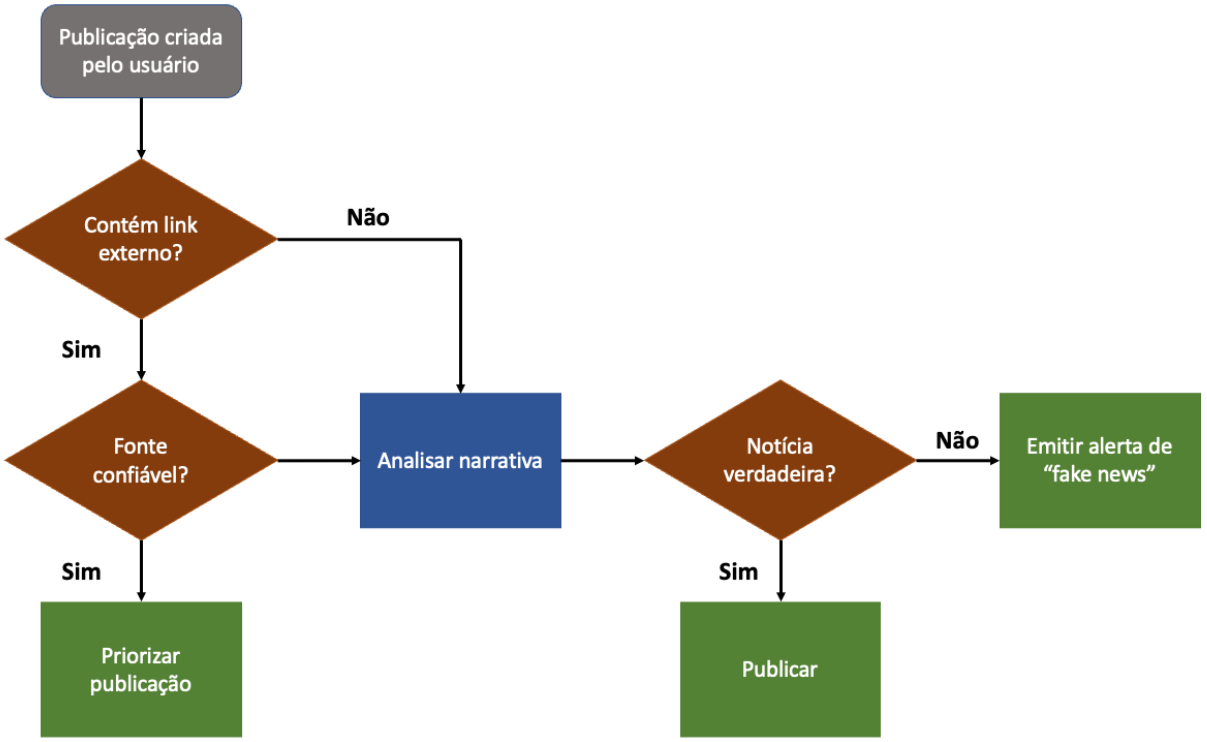
A mineração de dados dentro das mídias sociais tem como objetivo extrair e analisar informações de seus usuários, como interesses, *hobbies*, estilo de vida, religião, tendências políticas e intenções de compras, para, assim, entregar conteúdo personalizado e de acordo com suas convicções. A interface apresentada gera uma sensação de eficiência ao utilizar a rede social. Todavia, esse processamento para entrega de informação personalizada, denominado de algoritmo filtro bolha, aprisiona os usuários em grupos bem definidos, dificultando o contato não somente com opiniões antagônicas, mas com notícias e conteúdos que contrastem com o conteúdo consumido por sua "bolha" social.

Nesse ambiente, notícias falsas se proliferam de forma alarmante pois, visto que os usuários não recebem notícias e informações consideradas irrelevantes pelos filtros bolha, não há contato com notícias que contestem a inverdade propagada. As *fakes news* são propositalmente impactantes (VOGT; PICCININ, 2019) a fim de atrair atenção e suscitar reações emocionais, as quais são convertidas em compartilhamentos. O aspecto de legitimidade pouco importa, dado que o processo persuasivo se dá pela reafirmação dos ideais de indivíduos, motivados pelos aspectos emocionais da "notícia" (BAPTISTA, 2020). A entrega contínua de conteúdos similares, resultado dos algoritmos das redes sociais, reforça a

crença em notícias maliciosas e inverídicas, ao mesmo tempo que anula a contraposição às mesmas. A consequência mais alarmante desse processo é a polarização e radicalização de grupos de pessoas menos informados politicamente e socialmente.

Para evitar a difusão da desinformação nas redes sociais, métodos preventivos podem ser utilizados. Um desses métodos é o fortalecimento de algoritmos para promover o engajamento de fontes verificadas para que os usuários possam receber as informações de fontes confiáveis primeiramente, para então receber notícias de fontes secundárias (CLEVERLEY, 2017). Outro método é a utilização de um *framework* sintático de baixo nível para reconhecimento de padrões utilizados dentro de *fake news* e, assim, combatê-las de forma sistemática antes ou logo após seu compartilhamento (GUARINO et al., 2020). Esse método implica na necessidade de uma extensa base de treinamento bem construída, composta por notícias verídicas e maliciosas devidamente identificadas. Outra solução defendida nessa linha é a separação do consumo de notícias verificadas de publicações comuns geradas pelos usuários (MANFREDI-SÁNCHEZ, 2020).

Figura 1 - Fluxograma: proposta de algoritmo para controle de *fake news* em redes sociais.



Considerando as principais soluções supracitadas, a Figura 1 demonstra, por meio de um fluxograma, um algoritmo para controle de publicações em redes sociais. Em particular, o algoritmo unifica proposições dos autores Cleverley (2017), Guarino et al., (2020), Kaufhold

et al., (2020) Manfredi-Sánchez (2020), ao considerar os seguintes requisitos: (a) análise de origem de links para verificação de fontes de notícias verificadas e confiáveis; (b) favorecimento e impulsionamento de publicações com fontes confiáveis; (c) análise de narrativa de conteúdos gerados pelo usuário ou advindos de links não confiáveis; (d) emissão de alertas em caso de suspeita ou confirmação de notícia falsa. A abordagem representada pelo fluxograma pode alavancar o contato dos usuários com notícias verídicas e de teor jornalístico de maior qualidade, ao mesmo tempo que os informa sobre possíveis manipulações e falácias em notícias e publicações das redes sociais. Além disso, o algoritmo não impede a publicação do conteúdo, de forma que a liberdade de expressão é garantida e avaliações imprecisas do algoritmo podem ser identificadas e corrigidas. No entanto, compreende-se que a aplicação de um algoritmo nesse sentido depende do interesse de empresas privadas, as quais lucram com o engajamento e compartilhamento de conteúdos por parte dos usuários. Assim, a previsibilidade de sua aplicação estaria sujeita a determinações judiciais, as quais, se não ponderadas, podem afetar a dinâmica da liberdade de expressão na rede.

Ademais, é importante considerar que medidas tecnológicas podem ser ineficientes se não forem associadas à conscientização e à educação da população para o consumo racional e esclarecido de informações. É necessária uma forte sensibilização do público sobre o consumo de notícias nas mídias online para que entendam o impacto social que elas causam e como as redes sociais os afetam. É evidente a falta de conhecimento da população geral sobre o processo científico e jornalístico. Seu entendimento é, também, essencial para que fontes confiáveis deixem de ser descredibilizadas por investidas articuladas de desmoralização de trabalhos que seguem abordagens sistemáticas e imparciais. Em tempos de pós-verdade, a educação é fundamental para que usuários se protejam da manipulação a qual estão susceptíveis no ambiente digital.

5. CONSIDERAÇÕES FINAIS

Este trabalho buscou responder a seguinte questão: “Como a mineração de dados e os algoritmos de filtros bolha são usados a favor das *fake news* e quais são os mecanismos de prevenção?”. Para esse fim, foi conduzida uma revisão sistemática da literatura. Trabalhos resultantes do processo de seleção fundamentaram a discussão realizada a fim de responder à questão de pesquisa. Destaca-se que, apesar de responder adequadamente a questão, o produto da revisão foi limitado pela base de busca e pelos termos estabelecidos. Trabalhos futuros

devem considerar um número maior de bases de pesquisa, assim como estender os termos de busca. No entanto, os resultados aqui apresentados estabelecem uma base preliminar para que pesquisadores de diferentes áreas compreendam o processo de disseminação de *fake news* por meio dos filtros bolha, assim como suas proposições de solução. Essa base, sustentada por um método sistemático de revisão, é, também, relevante para que outros trabalhos acadêmicos possam respaldar suas justificativas e motivações.

Com base nas propostas de solução de tecnológica identificadas, é apresentado um fluxograma que unifica aquelas que foram consideradas viáveis e efetivas a curto prazo. Além da necessidade de implementação de novas estratégias de filtragem de publicações em redes sociais, problema que só pode ser sanado pelos responsáveis por sua gestão, a conscientização e educação da população são ações essenciais para garantir sua proteção contra notícias falsas e maliciosas.

Além dos métodos tradicionais de ensino, e considerando o impacto emocional das *fake news* em seu público-alvo, pesquisadores e educadores devem considerar a proposição de alternativas de conscientização originais e sensíveis ao atual contexto da sociedade. No âmbito da computação, jogos sérios, cujo objetivo é informar e treinar usuários na identificação de notícias falsas, têm se mostrado uma ferramenta promissora, principalmente entre o público jovem. O design de soluções motivadoras, que eduquem e gerem engajamento, é um dos instrumentos de enfrentamento às iniciativas mal-intencionadas de alienação da população por meio dos ambientes digitais.

REFERÊNCIAS

BAPTISTA, J. **Ethos, pathos e logos. Análise comparativa do processo persuasivo das (fake) news**. *Eikon*, v. 1, nº 7, p. 43–54, 2020. DOI: http://10.20287/eikon

BOZDAG, E. **Bias in algorithmic filtering and personalization**. *Ethics and information technology*, v. 15, nº 3, p. 209–227, 2013. DOI: https://doi.org/10.1007/s10676-013-9321-6

CAIADO, R. et al. **Metodologia de revisão sistemática da literatura com aplicação do método de apoio multicritério à decisão SMARTER**. Em: *Anais XII Congresso Nacional de Excelência em Gestão e III - Inovarse - Responsabilidade Social Aplicada*. Rio de Janeiro. 2016.

CERON, W.; DE-LIMA-SANTOS, M.-F.; QUILES, M. G. **Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content**. *Online Social Networks and Media*, v. 21, p. 100116, 2021. DOI: https://doi.org/10.1016/j.osnem.2020.100116

DE-LA-TORRE-UGARTE-GUANILO, M. C.; TAKAHASHI, R. F.; BERTOLOZZI, M. R. **Systematic review: general notions**. *Revista da Escola de Enfermagem da USP*, v. 45, nº 5, p. 1260–1266, 2011. DOI: https://doi.org/10.1590/S0080-62342011000500033

GELFERT, A. **Fake news: A definition**. *Informal Logic*, v. 38, nº 1, p. 84–117, 2018. DOI: https://doi.org/10.22329/il.v38i1.5068

GUARINO, S. et al. **Characterizing networks of propaganda on twitter: a case study**. *Applied Network Science*, v. 5, nº 1, p. 1–22, 2020. DOI: https://doi.org/10.1007/s41109-020-00286-y

KATSIREA, I. **"Fake news": reconsidering the value of untruthful expression in the face of regulatory uncertainty**. *Journal of Media Law*, v. 10, nº 2, p. 159–188, 2018. DOI: https://doi.org/10.1080/17577632.2019.1573569

KAUFHOLD, M.-A. et al. **Mitigating information overload in social media during conflicts and crises: design and evaluation of a cross-platform alerting system**. *Behaviour & Information Technology*, v. 39, nº 3, p. 319–342, 2020. DOI: https://doi.org/10.1080/0144929X.2019.1620334

KELLY, S. et al. **Manipulating Social Media to Undermine Democracy | Freedom House**. 2017. Disponível em: <https://freedomhouse.org/report/freedom-net/2017/manipulating-social-media-undermine-democracy>. Acesso em: 04/maio/22.

KLEINMAN, Z. **Is Your Smartphone Listening to You**. *BBC News*. 2016. Disponível em: <https://www.bbc.com/news/technology-35639549>. Acesso em: 16/set./22.

KORTA, S. M. **Fake news, conspiracy theories, and lies: an information laundering model for homeland security**. - Naval Postgraduate School Monterey, 2018. Disponível em: <https://www.hsdl.org/?abstract&did=811312>. Acesso em: 16/set./22.

KRÖGER, J. L.; RASCHKE, P. **Is my phone listening in? On the feasibility and detectability of mobile eavesdropping**. Em: *IFIP Annual Conference on Data and Applications Security and Privacy*. 2019. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-030-22479-0_6>. Acesso em: 16/set./22.

LAZER, D. M. J. et al. **The science of fake news**. *Science*, [s.l.], v. 359, nº 6380, p. 1094–1096, 2018. DOI: https://doi.org/10.1126/science.aao2998

LIBERINI, F. et al. **Politics in the Facebook Era-Evidence from the 2016 US Presidential Elections**. *SSRN Electronic Journal*, 2020. DOI: http://dx.doi.org/10.2139/ssrn.3584086

MACHADO, J.; MISKOLCI, R. **Das Jornadas de junho à cruzada moral: o papel das redes sociais na polarização política brasileira**. *Sociologia & Antropologia*, v. 9, p. 945–970, 2019. DOI: https://doi.org/10.1590/2238-38752019v9310

MANFREDI-SÁNCHEZ, J.-L. **Globalization and power: the consolidation of international communication as a discipline. Review article**. *Profesional de la Información*, v. 29, nº 1, 2020. DOI: https://doi.org/10.3145/epi.2020.ene.11

MASIP, P.; RUIZ-CABALLERO, C.; SUAU, J. **Active audiences and social discussion on the digital public sphere. Review article**. *El profesional de la información (EPI)*, v. 28, nº 2, 2019. DOI: https://doi.org/10.3145/epi.2019.mar.04

MONTAGNI, I. et al. **Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy**. *Journal of Public Health*, v. 43, nº 4, p. 695–702, 2021. DOI: https://doi.org/10.1093/pubmed/fdab028

PARISER, E. **The filter bubble: What the Internet is hiding from you**. 1 ed. London: Penguin UK, 2011. DOI: https://doi.org/10.22456/2175-2745.65827

PAULA AVELAR, G. DE; NALDI, M. C. **Comparação entre abordagens escaláveis para o processamento de conjuntos de dados textuais**. *Revista de Informática Teórica e Aplicada*, v. 24, nº 1, p. 121–149, 2017.

PIÑEIRO-OTERO, T.; ROLÁN, L. X. M. **Para comprender la política digital – Principios y acciones**. *Vivat Academia*, [s.l.], nº 152, p. 19–48, 2020. DOI: https://doi.org/10.15178/va.2020.152.19-48

QUADRADO, J. C.; FERREIRA, E. da S. **Ódio e intolerância nas redes sociais digitais**. *Revista Katálysis*, v. 23, p. 419–428, 2020. DOI: https://doi.org/10.1590/1982-02592020v23n3p419

RECUERO, R.; ZAGO, G.; SOARES, F. **Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter**. *Social Media+ Society*, v. 5, nº 2, 2019. DOI: https://doi.org/10.1177/2056305119848745

REVIGLIO, U. **Serendipity as an emerging design principle of the infosphere: challenges and opportunities**. *Ethics and Information Technology*, v. 21, nº 2, p. 151–166, 2019. DOI: https://doi.org/10.1007/s10676-018-9496-y

SASTRE, A.; OLIVEIRA, C. S. P. DE; BELDA, F. R. **A influência do "filtro bolha" na difusão de Fake News nas mídias sociais: reflexões sobre as mudanças nos algoritmos do Facebook**. *Revista GEMInIS*, v. 9, nº 1, p. 4–17, 2018. DOI: https://doi.org/10.4322/2179-1465.0901001

SUNSTEIN, C. **#Republic: Divided Democracy in the Age of Social Media**. *Ethical Theory and Moral Practice*. 1 ed. Princeton: Princeton University Press, 2017. 328 p.

VOGT, V. A.; PICCININ, F. **Análise da anatomia narrativa das fake news**. *XXV Seminário de Iniciação Científica - UNISC*, p. 101, 2019.

VOSOUGHI, S.; ROY, D.; ARAL, S. **The spread of true and false news online**. *Science*, v. 359, nº 6380, p. 1146–1151, 2018. DOI: https://doi.org/10.1126/science.aap9559

WORLD ECONOMIC FORUM. **Global Risks 2013 Eighth Edition: An Initiative of the Risk Response Network**. *Global Risks 2013*. Geneva: World Economic Forum, 2013.