

Automatização no *fact-checking*: pensando a prática de verificação a partir de agrupamentos semânticos de frases verificadas pelo Projeto Comprova

Automation in fact-checking: assessing the
verification practice from semantic groupings
of verified claims by the Comprova Project

Taís Seibt^[*]
seibt.tais@gmail.com

Didier A. Vega-Oliveros^[**]
didiervega@gmail.com

Anderson Rocha^[**]
anderson.rocha@gmail.com

RESUMO

O artigo promove uma reflexão teórica sobre os limites da prática de checagem de fatos (*fact-checking*) frente a um cenário de desinformação, tanto pela perspectiva computacional quanto pelo princípio da objetividade jornalística. O estudo analisa 2031 mensagens originais de usuários enviadas ao **Projeto Comprova**, coalizão entre dezenas de veículos de mídia brasileiros, utilizando técnicas de aprendizado de máquina, mineração de textos e modelagem de redes complexas. O modelo computacional investiga padrões linguísticos que possam auxiliar no desenvolvimento de ferramentas de automatização para identificar conteúdos potencialmente enganosos na rede a partir de agrupamentos semânticos. Os achados servem à problematização de mudanças estruturais do jornalismo em um ecossistema midiático que exige respostas multidisciplinares.

Palavras-chave: Fact-checking. Jornalismo de verificação. Desinformação. Pandemia. Coronavírus.

ABSTRACT

The paper promotes a theoretical reflection on the limits of the fact-checking practice in a disinformation context, both from the computational perspective and the journalistic principle of objectivity. The study analyzes 2031 original messages sent from users to the Comprova Project, a coalition of dozens of Brazilian media outlets, using machine learning techniques, text mining and complex network modeling. The computational model investigates linguistic patterns that can help in the development of automation tools to identify potentially misleading content on the web based on semantic groupings. The findings serve to problematize structural changes in journalism in a media ecosystem that requires multidisciplinary responses.

Keywords: Fact-checking. Verification journalism. Disinformation. Pandemic. Coronavirus.

^[*] Universidade do Vale do Rio dos Sinos (Unisinos). Av. Unisinos, 950 - Cristo Rei, São Leopoldo (RS).

^[**] Universidade Estadual de Campinas (Unicamp). Cidade Universitária Zeferino Vaz - Barão Geraldo, Campinas (SP).

Introdução

A pandemia do novo coronavírus acirrou a tensão nas relações entre política, saúde pública, desinformação e checagem de fatos (*fact-checking*) nos discursos compartilhados em plataformas de rede social e aplicativos de troca de mensagens. A disputa de narrativas converteu uma crise epidemiológica em disputa ideológica em várias nações, a tal ponto que até mesmo a Organização Mundial da Saúde (OMS) adotou o conceito de infodemia para tratar do fenômeno: “Um excesso de informações, algumas precisas e outras não, que tornam difícil encontrar fontes idôneas e orientações confiáveis quando se precisa” (Organização Panamericana de Saúde, 2020, p. 2).

Para a organização, plataformas de rede social e aplicativos de mensagens possibilitaram grande aumento no volume de informações associadas à pandemia, muitas com base em rumores e manipulação de informações “com intenção duvidosa”, cuja velocidade de propagação é comparável à de um vírus. Assim, a infodemia poderia potencializar a pandemia. Uma das apostas para responder à viralização desses conteúdos é a verificação de fatos. Agências independentes de *fact-checking* creditadas pela *International Fact-checking Network (IFCN)*, que já vinham recebendo incentivos financeiros e tecnológicos para verificar conteúdos que apresentam engajamento desproporcional em sites como *Facebook*, *Instagram* e *Twitter*, pelo menos desde 2018, no caso brasileiro, ficaram ainda mais pressionadas diante da infodemia. Surgiram novas iniciativas e coalizões, incluindo instâncias internacionais de cooperação entre organizações jornalísticas e de checagem de fatos, mas a disseminação de desinformação é exponencialmente maior do que a capacidade de resposta de verificadores humanos.

Encontrar maneiras de automatizar processos para identificar, sinalizar e até antever desinformação nos meios digitais é a próxima fronteira das iniciativas de checagem. Neste artigo, apresentamos um estudo preliminar que se utilizou de técnicas de aprendizado de máquina (AM), mineração de textos e modelagem de redes complexas para investigar padrões linguísticos que possam servir ao desenvolvimento de ferramentas de automatização para identificar conteúdos potencialmente enganosos na rede. A partir da análise de 2031 mensagens originais de usuários

enviadas ao **Projeto Comprova**¹, coalizão de dezenas de veículos de mídia brasileiros para verificação de conteúdos suspeitos, identificamos 16 agrupamentos semânticos. Cada grupo compartilha informações relevantes para tópico e contexto específico, onde sentenças estão fortemente conectadas entre si e apresentam conexões com outros grupos contextual e semanticamente adjacentes. Por esse método, além das palavras mais frequentes, abordagem classicamente utilizada, é possível evidenciar elementos-chave que interagem em cada grupo e a semântica geral dos grupos, inovação de nossas investigações em relação a outros projetos que buscam modelos automatizados para responder ao desafio global de combater a desinformação.

Ainda que iniciais para *insights* preditivos, os achados evidenciam a relação estreita entre conteúdos suspeitos e mensagens com vieses ideológicos, o que desafia verificadores, cujas práticas são alicerçadas no método da objetividade jornalística baseado na busca de evidências factuais para comprovar ou contestar a veracidade dos conteúdos. Muitas das mensagens analisadas não fazem referência a dados numéricos, aspectos legais, cronologia de acontecimentos e outros elementos verificáveis pela metodologia de *fact-checking*, o que as tornaria descartáveis para os verificadores. Por essa razão, é pungente adaptar métodos e critérios de *fact-checking*, em direção a um discurso que associe “jornalismo de verificação”, contextualização e transparência de métodos (Seibt, 2019), de forma a reforçar o papel social do jornalismo no novo ecossistema midiático. Tais práticas podem impulsionar uma mudança paradigmática (Charron e Bonville, 2016), desde que essas funções sejam reconhecidas socialmente e compartilhadas pelos demais atores envolvidos nos processos comunicacionais contemporâneos, onde já não se reconhecem fronteiras entre jornalistas e não-jornalistas (Kovach e Rosenstiel, 2014), entre informação e desinformação (Wardle e Derakhshan, 2017).

Com essas premissas, o artigo propõe refletir teoricamente sobre a prática a partir da própria prática, numa experiência multidisciplinar que permite tensionar as mudanças em curso no jornalismo a partir do desenvolvimento de um modelo computacional pensado para cooperar com jornalistas no trabalho de verificação. A metodologia de Inteligência Artificial desenvolvida pode ser útil na organização de notícias diversas pos-

1 – O projeto teve a primeira edição nas eleições de 2018 e foi retomado durante a pandemia, em 2020, reunindo 24 veículos de mídia tradicional no Brasil numa prática chamada de *crosschecking*, onde profissionais de diferentes redações cooperam no processo de verificação e revisão das checagens. Na versão mais recente da parceria, em 2021, colaboram 33 veículos, incluindo mídias locais. Informações em <https://projecomprova.com.br>

sibilitando que especialistas encontrem rapidamente conteúdos semanticamente relacionados para apontar *hubs* de conexão, ou seja, mensagens altamente influentes em um determinado contexto. A descoberta de que uma mensagem *hub* é potencialmente falsa pode levar à identificação de outras mensagens derivadas, qualificando a verificação e a apuração jornalística no geral.

Assim, os achados proporcionados pelo modelo computacional desenvolvido neste estudo preliminar servem, sobretudo, à problematização da metodologia de *fact-checking*, enquanto prática jornalística emergente fortemente alicerçada em referenciais de objetividade. Ao mesmo tempo, expressa limitações de soluções tecnológicas para verificação automatizada, já que os agrupamentos mostraram certas convicções políticas, criando, inclusive, falsas oposições, como saúde e economia, como se essas fossem dimensões adversárias. Nossas conclusões, portanto, reiteram que a resposta à desinformação exige intervenções multidisciplinares.

Contextualização

A origem do *fact-checking* como formato jornalístico — a verificação é reconhecida pelos profissionais como um elemento do processo jornalístico desde o século XIX² — remete ao início dos anos 1990, quando o âncora da CNN Brooks Jackson rebatia declarações equivocadas dos candidatos à Casa Branca em um programa de televisão. O próprio Jackson, em 2003, fundou o projeto *FactCheck.org* para sinalizar imprecisões no discurso de políticos dos Estados Unidos. Mas foi a partir do prêmio *Pulitzer* recebido em 2009 pelo site *PolitiFact*, fundado dois anos antes por Bill Adair, que o método de checagem de fatos avançou pelo território estadunidense, por meio de franquias e cursos de formação para verificadores, conforme recupera Lucas Graves (2016) ao estudar as iniciativas pioneiras de checagem nos Estados Unidos. Para ele, não se tratava apenas de uma nova prática, mas de um “movimento de reforma” do jornalismo constituído por “[...] organizações jornalísticas

estritamente adaptadas ao ecossistema contemporâneo de notícias em rede” (Graves, 2016, p. 9) e fortemente calcado no princípio jornalístico da objetividade.

As práticas de verificação ganharam o mundo em poucos anos, até que uma comunidade internacional ganhou forma a partir de 2014, quando a *International Fact-checking Network* (IFCN), mantida pelo Poynter Institute, realizou o primeiro encontro da rede. No Brasil, foi justamente em 2014 que as primeiras iniciativas surgiram, com o *Truco*, da *Agência Pública*, e o blog *Preto no Branco*, do jornal *O Globo*, editado por Cristina Tardáguila, que depois criou a *Agência Lupa*. O *Aos Fatos* veio logo depois, em 2015. Já o projeto *Comprova*, que forneceu o material para iniciarmos esta investigação, foi criado em 2018, reunindo veículos de mídia tradicional que apropriaram os métodos de verificação em suas redações para ampliar o alcance de conteúdos verificados durante a eleição presidencial, já que tinham audiências mais consolidadas em diversos espaços de mídia.

Sendo conteúdos com maior potencial de se tornarem “virais” os mais lucrativos no modelo de publicidade digital baseado em engajamento, houve pressão sobre as plataformas sociais quanto ao impulsionamento de conteúdos de campanha, muitos deles contaminados por falsidades e distorções, durante as eleições brasileiras de 2018. Foi nessa disputa eleitoral que surgiu o **Projeto Comprova**, coordenado pela Abraji - Associação Brasileira de Jornalismo Investigativo, a partir de uma iniciativa do *First Draft*, projeto do Centro Shorenstein da Harvard Kennedy School, dos Estados Unidos. Reuniram-se 24 redações brasileiras para monitorar um debate político altamente polarizado, especialmente no segundo turno, disputado entre Jair Bolsonaro, à época no PSL, e Fernando Haddad, do PT, partido do ex-presidente Lula, que havia sido impedido de concorrer após ser condenado em segunda instância na Operação Lava-Jato³. Para além dos questionamentos sobre o uso político da operação que revelou esquemas de corrupção envolvendo agentes do governo, estatais e empreiteiras, o que mais chamou atenção foi a ascensão de um candidato com baixa presença na propaganda eleitoral

2 – É no século XIX que se estabelece o chamado “jornalismo de informação” (Charron e Bonville, 2016), período em que o discurso jornalístico se caracteriza pela notícia, “produto histórico da institucionalização e da legitimação do papel do jornalista” (Alsina, 2009, p. 48) nas organizações jornalísticas.

3 – A Operação Lava-Jato teve início em março de 2014 e se estendeu nos anos seguintes, resultando no julgamento e condenação de dezenas de agentes públicos, membros de partidos políticos, diretores e funcionários de estatais, além de sócios e dirigentes de empreiteiras, por desvio de dinheiro público, superfaturamento de obras e outros crimes, principalmente envolvendo a Petrobras. A operação deu visibilidade ao então juiz Sergio Moro, que tornou-se ministro de Bolsonaro, mas ficou poucos meses no cargo. O ex-presidente Luiz Inácio Lula da Silva tornou-se inelegível em março de 2018, após ser condenado em segunda instância em um dos processos. A condenação de Lula foi revertida pelo Supremo Tribunal Federal em 2021.

gratuita, devido ao pouco espaço de seu partido, e que sequer compareceu a debates nos principais canais de televisão. Em contrapartida, Bolsonaro apostou no engajamento em plataformas de rede social, sendo eleito presidente.

Mecanismos de recomendação das plataformas de rede social se destacam no novo “mercado de hiperconcorrência de informação” (Charron e Bonville, 2016). Hiperconcorrência é um termo do vocabulário da “nova economia” apropriado pelos autores para desenvolver um tipo ideal⁴ que permitisse investigar “[...] como e por meio de que mediações as novas condições do jogo de concorrência influenciam na prática jornalística” (Charron e Bonville, 2016, p. 341). Como os que competem pela atenção dos leitores crescem mais do que o estoque de atenção disponível, “desvios” no discurso informativo tornaram-se mais comuns nas últimas décadas do século XXI, processo que se intensificou na virada do milênio. É nesse regime de hiperconcorrência que os pesquisadores canadenses situam o “jornalismo de comunicação”, um tipo ideal que caracteriza um discurso jornalístico mais próximo da comunicação interpessoal (Charron e Bonville, 2016, p. 380).

O que os autores não poderiam prever quase 20 anos atrás, quando concluíram seus estudos⁵, era que esse processo de subjetificação do discurso jornalístico viria acompanhado de plataformas que borram fronteiras entre jornalismo, publicidade, entretenimento e comunicação interpessoal. Pelos mesmos canais que se consome conteúdo jornalístico atualmente, chegam propagandas políticas, atualizações familiares, produções culturais, anúncios de produtos, convites para eventos, piadas e ainda comunicados falsos. Reafirmar seu compromisso deontológico com a verdade, a partir do qual se deve informar com base em evidências factuais, seria uma saída ao jornalismo para recompor seu papel social nesse novo ecossistema midiático, sendo o formato de *fact-checking* sua melhor representação.

É a partir das características dessa prática que surge outro tipo ideal, o “jornalismo de verificação” (Seibt, 2019), como referencial para estudar as mudanças do jornalismo.

A construção dos tipos ideais de Charron e Bonville (2016) serve à compreensão das mudanças estruturais do jornalismo ao longo da história. Numa estrutura, há “mudanças normais”, ou seja, transformações internas à estrutura, e há mudanças mais profundas, capazes de descaracterizar a estrutura por completo, a ponto de ela já não ser mais reconhecível (Charron e Bonville, 2016, p. 104). Essas são as mudanças estruturais. No caso do jornalismo, segundo os autores, tais mudanças se cristalizam no texto jornalístico, que por sua vez reflete uma série de transformações em 14 parâmetros que compõem a estrutura - o paradigma jornalístico - no respectivo contexto, são eles: texto jornalístico, texto *jornálico*⁶, prática jornalística, jornalistas, organização jornalística, produção midiática, meios de comunicação, fontes de informação, fontes de financiamento, público, práticas culturais e valores, instituições socioculturais, direito e sistema político, economia.

Alterações em quaisquer desses parâmetros, de forma direta ou indireta, promoveriam mudanças perceptíveis no texto jornalístico, ou seja, a manifestação discursiva do jornalismo em seu contexto, pois “[...] não são as suas condições de trabalho, sua organização profissional, sua ideologia, etc., que distinguem o jornalismo de outras profissões ou atividades, mas sua função, que consiste em produzir uma certa categoria de mensagens de interesse público” (Charron; Bonville, 2016, p. 127). Com essas premissas, Charron e Bonville (2016) reconhecem quatro paradigmas⁷ jornalísticos desde o surgimento da imprensa. No século XVII, os primeiros periódicos serviam à publicação de comunicados oficiais, registros náuticos e outros avisos, o “jornalismo de transmissão”; depois vieram folhetos de cunho político em defesa de partidos

4 – Tipo ideal é um método desenvolvido pelo sociólogo alemão Max Weber, considerado um dos fundadores da Sociologia, para compreender os sentidos das ações sociais. Trata-se da concepção de um “caso limite”, uma elaboração teórica que serve de modelo de comparação com a realidade empírica. (Weber, 2014)

5 – A obra original de Charron e Bonville, publicada em francês no Canadá, é de 2004.

6 – Os autores diferenciam o texto jornalístico, que pode ser entendido como o conteúdo noticioso, do texto *jornálico*, termo usado para descrever os demais conteúdos que compõem o jornal, mas que não são jornalísticos, como quadrinhos, horóscopo, palavras cruzadas e outras seções tradicionais em edições impressas.

7 – Segundo Thomas Kuhn (1975, p. 219), paradigma é “[...] aquilo que os membros de uma comunidade partilham”. Charron e Bonville (2016) transpõem essa noção de paradigma para o jornalismo, sendo o paradigma jornalístico definido como “[...] um sistema normativo criado por uma prática fundamentada no exemplo e na imitação, constituído de postulados, de esquemas de interpretação, de valores e de modelos exemplares com os quais se identificam e se referem os membros de uma comunidade jornalística em um dado âmbito espaço-temporal, que unem os integrantes à comunidade e servem para legitimar a prática” (Charron e Bonville, 1996, p. 58 in Charron e Bonville, 2016, p. 68).

ou causas sociais, no espírito da Revolução Francesa, no final do século XVIII, o “jornalismo de opinião”; até que as empresas jornalísticas se consolidaram nas sociedades capitalistas e regimes democráticos, em meados do século XIX, sendo a notícia o texto jornalístico mais representativo deste paradigma, o “jornalismo de informação”.

Quando a linguagem da notícia começa a ser flexibilizada, diante da segmentação do mercado e da necessidade de entreter até mesmo para informar, abrindo margem a subjetividades e até se assemelhar à propaganda em alguns conteúdos, os autores veem surgir um novo paradigma na década de 1980, o “jornalismo de comunicação”, no qual a ligação entre emissor e destinatário é o principal trunfo: “certas mensagens ou partes de mensagem, semanticamente pobres [...] visam essencialmente a estabelecer e manter [...] a comunicação entre emissor e destinatário” (Charron e Bonville, 2016, p. 35). No contexto sócio-histórico estudado pelos pesquisadores canadenses, esse discurso estava presente em canais de TV a cabo e publicações customizadas. Atualmente, é possível aproximar essa percepção com o abuso de técnicas de *SEO*⁸ e *clickbait*⁹ nos meios digitais.

Ainda que tenham sido concebidos no contexto midiático e social canadense, o que impede a generalização dos paradigmas em termos cronológicos simultaneamente em todas as partes do mundo, pode-se perceber a ocorrência desses mesmos paradigmas ao longo da história da imprensa brasileira, inclusive na contemporaneidade. Como os próprios autores enfatizam: “Transmissão, opinião, informação e comunicação são dimensões presentes em graus diversos em todas as formas de jornalismo; é sua importância relativa que varia” (Charron; Bonville, 2016, p. 34). Pode-se considerar que a importância relativa das características discursivas do jornalismo, dependentes dos demais parâmetros estruturais do paradigma jornalístico, podem variar tanto ao longo do tempo quanto nos diferentes países.

Enquanto formato jornalístico, o *fact-checking* resgata as características fundacionais do “jornalismo de informação”, porém é preciso situá-lo em um contexto sócio-histórico diferente. As mesmas práticas discursivas

e até mesmo técnicas do “jornalismo de comunicação” foram apropriadas por outros atores no novo ecossistema midiático. O discurso jornalístico busca na checagem de fatos uma forma de diferenciação nesse ambiente, ao mesmo tempo em que se torna urgente como resposta social ao fenômeno da desinformação, e também às transformações do mercado.

O *fact-checking* surge num contexto em que instituições jornalísticas tradicionais perdem receita e participação nas discussões sociais. “Se quiserem manter ou mesmo aumentar sua relevância, terão de explorar novos métodos de trabalho e processos viabilizados pelas mídias digitais” (Anderson et al., 2013, p. 38). Com “[...] suas próprias regras, rotinas e ‘melhores práticas’, propagadas em suas conferências e *mailing lists*”¹⁰ (Graves, 2016, p. 8), o *fact-checking* estabelece um formato específico de texto jornalístico, constituído pelo processo de apuração em si. O que o diferencia da notícia não é tanto sua produção, e sim sua apresentação. Nos conteúdos de checagem, o jornalista deve indicar as fontes, explicar os procedimentos, para que o leitor possa validar o processo, como prevê o código de conduta da IFCN¹¹: transparência em relação à metodologia; na escolha das fontes; quanto ao financiamento; política pública de correções; apartidarismo.

Outra diferença definitiva é que a checagem de fatos se debruça sobre discursos que já estão públicos, diferentemente da notícia, que busca exclusividade - o “furo” de reportagem, no jargão da área. Ao mesmo tempo, as mudanças no sistema de financiamento do jornalismo, que é um dos parâmetros estruturais do paradigma, incluindo a maior participação das plataformas de mídias sociais como patrocinadoras da checagem de fatos, impulsiona mudanças no texto jornalístico então produzido.

Por esses elementos, a nova prática, que representa “a ideia nova que precisa construir-se um primeiro nicho, antes de poder fortalecer-se, tornar-se uma tendência reconhecida e, finalmente, triunfar” (Morin, 2011, p. 286) como um novo paradigma, serve à problematização de uma mudança estrutural emergente, observada a partir do “jornalismo de verificação” como tipo ideal.

8 – Sigla para *Search Engine Optimization*, que significa, em tradução livre, Otimização para Mecanismos de Buscas. São técnicas para posicionar um conteúdo entre os primeiros resultados de pesquisa em buscadores online.

9 – Conteúdo com manchetes sensacionalistas e/ou imagens apelativas para atrair cliques e incentivar o compartilhamento para obter lucros com publicidade digital. Também chamado em português de “caça-clique”.

10 – Increasingly, fact-checkers have their own rules, routines, and “best practices”, hammered out in their own conferences and mailing lists.

11 – *International Fact-checking Network* (IFCN) é a rede que congrega iniciativas de checagem no mundo todo. Baseada no Instituto Poynter, na Flórida, a IFCN certifica seus signatários no intuito de garantir que sigam os protocolos compartilhados internacionalmente.

Se o paradigma, entendido como estrutura que subentende um sistema de regras – de edição, redação, semânticas e procedimentais – a emergência de um “jornalismo de verificação” promoveria mudanças nas regras de redação e edição, uma vez que se alteram os critérios de seleção – não mais o conteúdo exclusivo, e sim algo que já está público – e de interpretação – o fact-checking obriga o jornalista a tirar uma conclusão quanto ao grau de veracidade das informações, por meio da atribuição de uma etiqueta. Também são identificáveis alterações nas regras procedimentais, já que o processo de verificação de informações prioriza fontes documentais, dados estatísticos e outros tipos de bases não-humanas, diferentemente do processo convencional de reportagem, fortemente embasado em entrevistas. (SEIBT, 2019, p. 93)

É nos critérios de seleção e classificação dos conteúdos verificáveis pelo método de *fact-checking* que residem as tensões da investigação ora analisada, com base nos conteúdos suspeitos enviados ao **Projeto Comprova**. Na descrição de sua metodologia, as agências de checagem delimitam como verificáveis discursos que contenham números, dados estatísticos, comparações, ou que remetam a fatos históricos ou jurídicos¹². O intuito é afastar ao máximo opiniões e interpretações, tornando possível etiquetar uma declaração como verdadeira ou não a partir de critérios objetivos. Na linha do que Wilson Gomes (2009) trata pela norma da veracidade, o jornalismo se compromete “a trabalhar objetiva e metodicamente para afastar o risco do engano ou do erro” (Gomes, 2009, p. 11) — e o *fact-checking* parece perseguir justamente esse compromisso.

A objetividade surge como norma adjacente, como um critério prático, traduzido por alguns autores como o “rigor do método”, que consistiria em “ir ao fundo das investigações materialmente realizáveis” e recolher “todos os fatos confirmados disponíveis” para uma correta descrição dos acontecimentos (Cornu, 1994, p. 391). Para Phillip Meyer (1973), o rigor do método só é possível se houver transparência. Na definição de um “jornalismo de precisão”, o autor sustenta apropriação de mecanismos do método científico ao método jornalístico, de modo a permitir que este seja testado e validado.

Lippmann (2008) já falava muito antes da necessidade de um “espírito científico” para uma prática jornalística baseada no estudo da prova e da verificação — no método

— já que “os fatos não são simples e nem tão óbvios” (Lippmann, 2008, p. 294). Liriam Sponholz (2009) também trata do princípio da objetividade a partir de um método transparente: “Isto significa que as fontes e o método utilizados devem ser transparentes para que outros jornalistas ou experts possam repetir o mesmo procedimento dentro das mesmas condições” (Sponholz, 2009, p. 153).

Na descrição da “disciplina da verificação”, Kovach e Rosenstiel (2004, p. 116) dizem que “[...] no conceito geral, o método seria objetivo, não o jornalista” e recomendam que o jornalista “[...] seja o mais transparente possível sobre seus métodos e motivos” (Kovach e Rosenstiel, 2004, p. 123), pois “[...] só explicando como sabemos o que sabemos podemos fazer com que o público possa, queira, reproduzir a informação. É isso o que significa objetividade de método na ciência, ou no jornalismo” (Kovach e Rosenstiel, 2004, p. 128).

Esses elementos são exacerbados nos conteúdos de checagem, redigidos de forma a demonstrar as evidências de veracidade que embasam a avaliação do jornalista sobre um discurso ao atribuir uma etiqueta ao fim do processo. Mas não é no procedimento de verificação que reside a problemática da checagem de fatos. O problema está em que a tarefa de checagem de fatos depende principalmente de jornalistas e especialistas em mídias, ou seja, recursos humanos. No entanto, muitos dos conteúdos difundidos a cada hora nas mídias sociais passarão despercebidos e não serão checados, dado o volume e a velocidade em que circulam nas redes, além dos critérios editoriais de factualidade para classificar o que é ou não checável.

Delimitação do problema

Determinar os níveis de veracidade de um conteúdo é uma tarefa complexa que nenhum algoritmo ou técnica de inteligência artificial consegue dominar neste momento (Mele, 2017). Ao mesmo tempo abrangente e vago, o termo “fake news” pode se referir a informações não corrigidas, fabricadas, descontextualizadas, distorcidas ou tendenciosas, uma mistura entre informações verdadeiras, falsas e opiniões, ou sensacionalismo com objetivo de atrair a atenção de usuários para clicar no link (Berezow, 2019). Assim, é preciso encontrar uma simbiose entre o poder de processamento e detecção de padrões que as máquinas oferecem junto com a orientação supervisionada e crítica dos especialistas em verificação de fatos - em geral, jornalistas.

12 – Trata-se de uma síntese a partir das metodologias de Agência Lupa, Aos Fatos e Truco (Seibt, 2019).

Diversas abordagens computacionais têm sido propostas para a detecção de conteúdos falsos. Os trabalhos relacionados à tarefa de caracterização de notícias para a checagem de fatos podem ser divididos em: métodos artesanais e os automatizados baseados em técnicas de mineração de texto. Os artesanais extraem características como o estilo e a linguagem do texto, fazendo uso de técnicas de PLN, entre outras. Nesse ponto, as técnicas consideram como hipótese que os conteúdos falsos apresentam estruturas gramaticais simples, alguns com problemas e erros ortográficos, e que são mais similares a um estilo de escrita específico, como sátira, e não ao de jornalismo (Horne, 2017; 2020).

No caso dos métodos automatizados, temos os baseados em grafos ou redes complexas, e os relacionados a modelos estatísticos ou de AM. Os trabalhos na linha de redes complexas têm buscado fazer análise qualitativa do texto. Por exemplo, foi visto que os nós centrais nas comunidades apresentam os tópicos principais das redes de palavras e, ao mesmo tempo, as palavras-chave do texto são as mais centrais na rede (Sibel, 2014; Vega-Oliveros, 2019). Outra vertente em redes está em analisar usuários e postagens para entender as dinâmicas de propagação de conteúdos falsos, propagadores influentes, comunidades, entre outros padrões (Zhou, 2019).

Por outro lado, os métodos baseados em estatística ou AM têm reduzido o problema a um espectro binário entre conteúdos falsos ou verdadeiros. Técnicas de *embedding* de palavras para um texto, como *WordToVec* (Mikolov, 2013) ou *FastText* (Bojanowski, 2017), ou modelos de *embeddings* como BERT - os quais são treinados a partir de um *corpus* de documentos, têm atraído a atenção dos pesquisadores na área. No entanto, existem indícios de que esses modelos treinados não conseguem generalizar para cenários do mundo real, e que técnicas simples como as artesanais podem atingir melhores resultados (Gruppi, 2021).

Recentemente, surgiram algumas abordagens que combinam modelos de AM com redes complexas, considerando, por exemplo, as características do texto de uma notícia junto com as informações do contexto dos usuários que fizeram comentários nela (Schwarz, 2020). Outro exemplo interessante é a combinação de características artesanais do estilo da escrita dos artigos junto com a rede de veículos de difusão desses conteúdos, ou seja, revistas e jornais (Gruppi, 2021). O método estima o nível de veracidade da notícia segundo seu conteúdo e a conexão com os meios que a difundem. No entanto, as abordagens anteriores não propõem uma fusão direta entre abordagens de AM para representação

dos textos junto com a teoria de redes complexas.

Assim, nosso estudo inova ao propor um método de filtragem onde os conteúdos sejam interconectados em camadas semânticas de acordo com modelos robustos de *embedding* da literatura. A partir da fusão das camadas, obtivemos uma rede na qual podemos encontrar os agrupamentos semânticos entre os conteúdos, sem a necessidade de explicitar parâmetros como o número de grupos, tamanhos, densidades, como acontece com outros métodos de agrupamento.

Boa parte dos experimentos de automatização em jornalismo está mais voltada a operações de raspagem de dados, por meio da criação de *bots* (robôs) capazes de processar grandes volumes de dados em bases públicas ou buscar expressões de interesse jornalístico em arquivos para auxiliar no processo de apuração de informações (Dalben, 2020). Em *fact-checking*, uma das experiências documentadas é o robô Fátima¹³, do site *Aos Fatos*, que é basicamente um *chatbot* responsivo que alerta usuários quando compartilham uma potencial desinformação no Twitter ou quando querem saber sobre alguma checagem no WhatsApp. O *Radar Aos Fatos* é um investimento mais avançado da mesma organização, pois a ferramenta atribui notas de 1 a 10 a conteúdos e adiciona ao Radar aqueles que recebem 5 pontos ou menos, sendo considerados suspeitos. Jornalistas da agência produzem relatórios com base nas informações do Radar. (Franciscato; Gonçalves, 2021).

Em comparação ao Radar, talvez a experiência mais avançada em termos de aplicação de inteligência artificial desenvolvida no jornalismo brasileiro para detectar desinformação, o principal diferencial do estudo ora proposto é o agrupamento semântico de mensagens identificadas no mesmo cluster, o que permitiria a sinalização conjunta de um volume maior de conteúdo com características discursivas praticamente idênticas. Muitos dos conteúdos sugeridos para investigação na amostra analisada neste estudo são redundantes e abundantes, onde aqueles mais centrais dentro de cada comunidade são mais relevantes, já que eles agrupam os tópicos e resumizam o conteúdo das mensagens suspeitas dentro sua comunidade, oferecendo aos jornalistas mais agilidade na detecção de ondas de desinformação.

Metodologia

As mensagens analisadas submetidas como suspeitas pelos usuários do Comprova foram coletadas entre os dias 25 de março a 8 de junho de 2020. O conjunto de

13 – <https://www.aosfatos.org/fatima/>

dados recebidos para análise conta com 2031 registros, os quais contêm os seguintes campos: data e hora do registro; tipo do conteúdo; descrição do conteúdo suspeito; plataforma ou fonte do conteúdo; e link de acesso ao conteúdo. O tipo de conteúdo determina se o material foi um vídeo, uma imagem ou meme, um texto ou notícia, um áudio, entre outros. O campo de descrição é onde o usuário descreve e comenta sobre o conteúdo suspeito, ou cola parte desse conteúdo. A plataforma ou fonte informa se o material foi visto no Twitter, WhatsApp, Facebook, escutado de um conhecido, ou outro tipo de fonte. Finalmente, temos um campo com o link ao conteúdo. Porém, menos de 30% das denúncias apresentam links. Dessas, mais da metade não são mais acessíveis, como é o caso de tuítes antigos que são apagados.

A base de dados foi solicitada pelos pesquisadores aos coordenadores do projeto Comprova diante do indicado em eventos públicos de que era do interesse do projeto que suas informações pudessem servir ao desenvolvimento de pesquisas na área. Foi compartilhado um arquivo em formato de planilha, gerado a partir de sugestões de verificação enviadas por leitores diretamente por um número de WhatsApp ou por formulários online disponibilizados nos canais do projeto. Essa base de sugestões é compartilhada entre todos os veículos integrantes da coalizão, que analisam os conteúdos sugeridos e decidem pela verificação daqueles que são considerados editorialmente adequados pelos critérios jornalísticos tradicionais, como relevância, impacto e factualidade do tema em questão, além das condições de checagem com base em evidências factuais, conforme os critérios de *fact-checking* já mencionados.

Dessa forma, boa parte dos conteúdos sugeridos na base não passa pela análise de verificadores. No período da pesquisa, para se ter uma ideia, o total de conteúdos avaliados por verificadores do **Comprova** era de 286: 128 classificados como falsos, 148 como enganosos e 10 comprovados¹⁴. O processo de verificação envolve, no mínimo, dois profissionais de duas redações diferentes, de modo a garantir o processo de dupla checagem e colaboração, chamado de *crosschecking*. A discrepância entre o volume de sugestões de checagem recebidas e o número de conteúdos efetivamente verificados (286

contra 2031), no entanto, reforça a necessidade de se desenvolver soluções tecnológicas que possam otimizar o processo de seleção e análise de veracidade de conteúdos digitais, o que é parte dos objetivos deste estudo.

A partir da base recebida, optamos por analisar o campo de descrição do conteúdo suspeito pelos usuários, já que apresentava maior informação e potencial para auxiliar no processo de checagem de fatos. Assim, realizamos uma etapa de pré-seleção e limpeza no conjunto de mensagens, removendo aquelas que apenas eram endereços *web*, e-mails, ou espaços em branco. Também, foram removidas mensagens pequenas, menores que 81 caracteres, o que corresponde a 45 percentil do tamanho entre todas as mensagens. Desse modo, obtivemos 1059 mensagens válidas a serem analisadas.

Para analisar os conjuntos de mensagens e suas similaridades semânticas, empregamos as técnicas de mineração de textos conhecidas como *word embeddings* (Bojanowski, 2017; Devlin, 2018). Nessas técnicas, as sentenças/palavras são transformadas de forma algébrica em vetores multidimensionais de números reais. Os vetores representam em coordenadas as informações sobre o contexto e os significados contidos nas palavras ou sentenças. Também, os vetores produzidos por *embeddings* são densos e de tamanho fixo chamado de espaço de *embedding* ou espaço de projeção.

Os valores em cada uma das posições do espaço de *embeddings* são obtidos através de um modelo de AM, o qual passou por um processo de treinamento em um *corpus* de milhões de documentos. Por exemplo, todas as páginas da Wikipédia de alguma língua, para o caso de *embeddings* monolíngues, ou várias línguas, para *embeddings* multilíngues. Como vantagem das abordagens com *embeddings*, temos que palavras com significados semelhantes, como rei, rainha, reinado, ou sinônimos, possuem representação muito próximas no espaço vetorial do *embedding*. Desse modo, podemos agrupar palavras e sentenças em grupos com significados e contextos semelhantes. Com essa representação multidimensional das palavras e sentenças, também é possível realizar outras tarefas de Processamento de Linguagem Natural (PLN), como análise de sentimento e emoções, autoria, entre outros.

Na literatura, existem várias técnicas para a trans-

14 – O **Projeto Comprova** usa quatro etiquetas de classificação de conteúdos. Enganoso é o conteúdo retirado do contexto original, que usa dados imprecisos ou que induz a uma interpretação diferente da intenção de seu autor; Falso é o conteúdo inventado ou que tenha sofrido edições de modo deliberado para espalhar uma mentira; Sátira é a etiqueta usada para classificar memes, paródias e imitações publicadas com intuito de fazer humor que possam estar sendo tomados por verdadeiros; e Comprovado é o fato verdadeiro, evento confirmado ou conteúdo original publicado sem edição. As explicações estão no site: <https://projetocomprova.com.br/about/>

formação de sentenças em vetores de palavras. Entre algumas mais recentes e consideradas como o estado da arte na tarefa de caracterização textual e semântica temos, BERT, RoBERTa, XLM-RoBERTa, SciBERT, distiluse-multilingual, entre outras. Neste trabalho, transformamos as sentenças dos usuários em vetores empregando as duas seguintes técnicas: A BERTimbau¹⁵ — um modelo BERT (Devlin, 2018) (*Bidirectional Encoder Representations from Transformers*) treinado exclusivamente para o Português Brasileiro com o BrWaC (*Brazilian Web as Corpus*), um corpus em português coletado de mais de 60 milhões de páginas (Wagner, 2019); o Distilbert-nli-stsb-quora¹⁶ — a versão multilíngue do distilbert-base-nli-stsb-quora, treinado com dados paralelos de mais de 50 idiomas das páginas da rede de consultas Quora (Reimers e Gurevych, 2020). Dado que não existe um modelo universal e ótimo para todas as modalidades, tópicos e tarefas, selecionamos estes dois modelos buscando a maior abrangência entre os possíveis temas, expressões e características das mensagens.

As mensagens são então representadas por um conjunto de vetores multidimensionais, os quais podemos analisar de forma matemática no espaço de projeção calculando a distância ou similaridade dos vetores das mensagens e outras propriedades topológicas. Para isso, fazemos uso da Teoria de Redes Complexas (Silva e Zhao, 2016), na qual temos uma rede, ou grafo, $G(V, E)$ composto por um conjunto de vértices V e um conjunto de arestas ou conexões E . Neste caso, cada mensagem, representada por seu vetor no espaço de *embedding*, se torna um vértice na rede. As conexões são estabelecidas entre mensagens que sejam similares no seu conteúdo e contexto. A ideia é utilizar cada uma das representações de *embeddings* dos modelos treinados, em português e multilíngue, para caracterizar o conteúdo das mensagens. Assim, para cada modelo construímos um grafo direcionado usando o método de K-vizinhos mais próximos (k-NN do inglês *k-Nearest Neighbors*) (Berton, 2018). Para cada vértice da rede x , estabelece-se uma conexão direcionada com os seus K vértices mais similares ou próximos a x , e assim sucessivamente para todos os vértices da rede. O nível de proximidade ou similaridade entre duas mensagens é obtido calculan-

do a distância cosseno entre os vetores de *embedding* dessas mensagens: quanto mais similares o valor tende a 1.0, e vai para zero quanto mais dissimilares.

Desse modo obtivemos duas camadas ou sub-redes de conexões dos vértices para cada modelo de *embedding* adotado. A rede final consiste na fusão de ambas as camadas em uma única rede, isto é, as arestas são agregadas e cada vértice passa a ter os vizinhos ou conexões existentes em ambas as camadas. Note-se que este é um processo inicial e simples de representação e construção das camadas e a fusão entre elas. Outras abordagens de construção, como redes regulares, \square -vizinhança, b-matching (Berton, 2018), podem ser consideradas. Também, o mecanismo de fusão pode considerar diferentes pesos para as conexões entre as camadas, intersecção do conjunto de arestas, entre outras características.

A vantagem da representação em redes é que podemos encontrar quais são as mensagens mais centrais e identificar padrões de conexão entre as mensagens. As medidas de centralidade buscam identificar quais são os vértices mais importantes ou influentes segundo a topologia da rede (Vega-Oliveros, 2019). Por exemplo, a centralidade de grau determina o número de conexões que possui cada vértice, e os *hubs* são os vértices que têm o maior número de conexões em comparação com a média de conexões dos vértices da rede. Outra característica importante presente nas redes do mundo real é a alta concentração de arestas dentro de grupos de vértices, formando estruturas modulares locais chamadas de comunidades (Clauset 2004; Fortunato e Hric, 2016). As comunidades são grupos de vértices densamente interconectados mas pouco conectados com o restante dos vértices da rede¹⁷.

Desse modo e no problema aqui tratado, vértices que pertencem à mesma comunidade compartilham contextos comuns e características semânticas semelhantes (Rossetti e Cazabet, 2018). Assim, as comunidades determinam grupos de mensagens que são altamente similares e próximas no conteúdo e tópicos tratados. Na nossa proposta, é possível usar qualquer algoritmo de detecção de comunidades, sendo esta uma escolha aberta e cabível de análise em trabalhos futuros. Algoritmos como FastGreedy (Clauset, 2004), Louvain (Blondel, 2008), entre outros, são alternativas clássicas adotadas na literatura.

15 – <https://github.com/neuralmind-ai/portuguese-bert>

16 – https://www.sbert.net/docs/pretrained_models.html

17 – Para mais detalhes sobre o problema e algoritmos propostos para a detecção de comunidades em redes complexas, recomendamos a revisão feita por Fortunato e Hric, 2016 (ver referências finais).

Resultados

Na **Figura 1**, observamos as redes construídas para cada modelo de *embedding*, sendo o da esquerda (a) produzida pelo modelo multilingue e a direita (b) pelo modelo treinado em português. As comunidades são representadas por cores diferentes e os *hubs* são os maiores vértices na figura, onde quanto maior o vértice, maior seu grau de centralidade. Podemos ver que os resultados não são robustos quando se consideram as redes por separado, com os *hubs*

ou as comunidades mudando um pouco dependendo do método de *embedding*. Por essa razão, propomos a união de ambas as redes em uma só para calcular a estrutura de comunidades das mensagens. Para isso, usamos o algoritmo Leiden de detecção de comunidade (Traag, 2019). Adotamos o método de Leiden por se tratar de um algoritmo recente com uma grande aceitação acadêmica e que apresenta ótimos resultados. O algoritmo é uma proposta que melhora o método de Louvain, em termos de otimização e divisão local das comunidades. Leiden pode

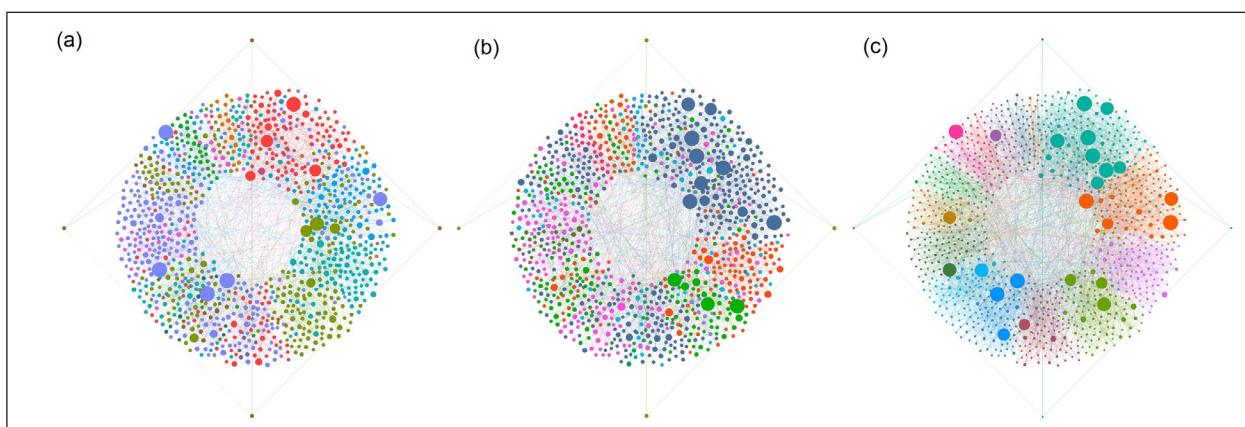


Figura 1. Representação circular das redes junto com as comunidades (cores) e os *hubs* (vértices maiores) para: (a) modelo Distilbert multilingue; (b) modelo BERTimbau em Português; e (c) a união de ambos os modelos.

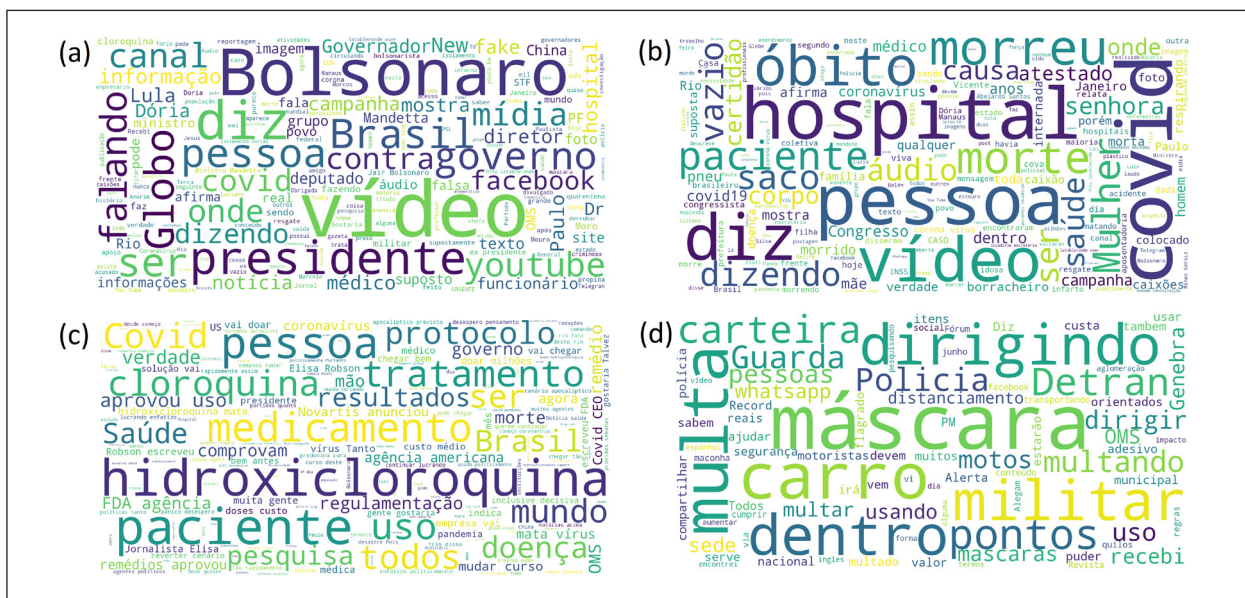


Figura 2. Nuvens de palavras da rede da união de ambos os modelos de *embeddings*, sendo: (a) da maior comunidade; (b) da segunda maior comunidade; (c) da terceira maior comunidade; (d) da menor comunidade da rede.

ser executado em redes de milhões de nós e requer pouco poder computacional. Isto é devido ao fato de ser uma versão otimizada do algoritmo de Louvain, e possuir garantia de convergência para encontrar o conjunto de comunidades ótimas conectadas localmente (Traag, 2019).

Foram identificados 16 agrupamentos semânticos. Cada grupo compartilha informações relevantes para tópicos e contexto específico, onde sentenças estão fortemente conectadas entre si e apresentam conexões com outros grupos, contextual e semanticamente adjacentes. Por esse método, além das palavras mais frequentes, abordagem classicamente utilizada, é possível evidenciar elementos-chave em cada grupo analisando a topologia da rede (Vega-Oliveros, 2019), isto é, os *hubs* ou mensagens altamente influentes em um determinado contexto, e o tópico geral dos grupos, inovação de nossas investigações.

Na **Figura 2**, temos as nuvens de palavras-chave das três maiores e a menor comunidade da rede combinada. Podemos observar que o principal tema dos conteúdos suspeitos em cada uma das comunidades é diferente. A maior comunidade está mais relacionada ao atual presidente do Brasil e seus adversários políticos. A segunda maior comunidade agrega conteúdos relacionados à pandemia de Covid-19, ao sistema de saúde pública e à gravidade do vírus. A terceira comunidade trata sobre protocolos e tratamentos para Covid-19, focado principalmente na hidroxicloroquina. A menor comunidade trata sobre multas de trânsito e ao uso obrigatório de máscaras. Na tabela a seguir, mostramos as 5 palavras-chave por frequência e as 5 frases-chave mais significativas de cada grupo.

Apesar de os resultados serem preliminares, consideramos muito informativas as palavras-chave e frases-chave detectadas por nosso método não supervisionado. As frases centrais nas comunidades podem ajudar como um filtro agregador das milhares de solicitações de verificação que a equipe recebe para checagem de fatos. É importante observar que muitas dessas informações são, em resumo, semelhantes. Um método automático de agrupamento de entradas semelhantes e recuperação/ordenação das mais significativas a serem analisadas primeiro tem potencial para melhorar e otimizar o tempo de verificação dos profissionais de *fact-checking*.

Discussão

Os agrupamentos semânticos encontrados na análise preliminar empreendida neste estudo, a partir de mensagens suspeitas enviadas ao **Projeto Comprova**, exemplificam a complexidade de se combater a interpretação equivocada

de fatos no contexto contemporâneo. Se observarmos as frases-chave do maior agrupamento semântico encontrado neste estudo, veremos que estão repletas de juízos de valor e ideias conspiratórias que não cabem nos critérios de verificação calçados no método de objetividade jornalística: um “ex-presidente comunista” que “mandou matar” Bolsonaro; o “alerta vermelho” aceso na Rede Globo por “queda drástica” de audiência; a “traíção” do ex-juiz Sérgio Moro. Dados em contexto equivocado e estatísticas exageradas, que são fatores considerados verificáveis pelos critérios jornalísticos de *fact-checking*, já não parecem ser os principais problemas de qualidade no discurso público. Esses elementos estão presentes nos agrupamentos da menor comunidade (**Tabela 1**): é possível desmentir ou confirmar se o Detran está aplicando multas a motoristas que conduzem sem máscara, como também é possível contrapor as mensagens de incentivo ao uso de cloroquina com base em critérios científicos. Mesmo que esses dados sejam corrigidos ou se mostrem verdadeiros, a checagem não elimina a distorção do contexto. Ainda, a influência de narrativas negacionistas e preconceituosas, muitas vezes vindas de autoridades que se opõem às evidências, dificulta o consenso em relação ao trabalho dos jornalistas. Associada a princípios de objetividade, a metodologia de *fact-checking* deixaria de fora muitas dessas mensagens por não oferecerem dados objetivos, e sim crenças, opiniões e juízos de valor.

A construção dessas narrativas não se dá ao acaso. Conforme Da Empoli (2020, p. 88), os “engenheiros do caos” estão bem conscientes de que “a indignação, o medo, o preconceito, o insulto, a polêmica racista ou de gênero se propagam nas telas e proporcionam muito mais atenção e engajamento que os debates enfadonhos da velha política”. Wardle e Derakhshan (2017, p. 4) consideram preocupantes as “implicações a longo prazo das campanhas de desinformação concebidas especificamente para semear a desconfiança e aguçar a divisão dos interesses socioculturais existentes por meio da criação de tensões nacionalistas, étnicas, raciais e religiosas”.

Diversos autores têm problematizado a relação entre a desordem informacional e o ressurgimento do populismo nos últimos anos. Para Prior (2019, p. 124), a crise das democracias representativas favorece o “discurso simplista sobre a realidade social, baseando-se na vontade popular para alcançar o poder e para combater um inimigo do povo, geralmente identificado com a elite”. A estratégia do “*nós* contra *eles*” é analisada por Cesarino (2019), com inspiração na análise crítica do discurso, como forma de “fortalecer o carisma do candidato; manter a audiência mobilizada através de mensagens alarmistas e conspiratórias; canibalizar o

	Palavras frequentes	Frases-chave
Primeira comunidade	+ Vídeo + Bolsonaro + Presidente + Brasil + Globo	+ 'ex presidente comunista brasil lula mandou matar presidente brasil bolsonaro segundo katia arbex' + 'Globo entra em 'alerta vermelho' após queda drástica de audiência e promove mudança de emergência' + 'BOMBA ATÔMICA - o exército interceptou a traição de Sérgio Moro e Bolsonaro já sabia de tudo' + 'aécio neves supostamente fazendo campanha eleitoral' + 'encontrou manipulação coordenada generalizada '
Segunda comunidade	+ Hospital + Covid + Pessoas + Vídeo + Óbito	+ 'familiares levam mãe pro hospital abelardo santos' + 'foto dessas pessoas carregando caixão sozinho' + 'psl filippe poubel flagra hospital' + 'chama luzia maria medeiros menezes' + 'irregularidades em estatísticas sobre coronavirus'
Terceira comunidade	+ Hidroxicloroquina + Paciente + Medicamento + Tratamento + Uso	+ 'novartis se compromete a doar até 130 milhões de doses de cloroquina' + 'pleno news': senegal usa cloroquina desde 1º caso e tem apenas 5 mortes' + 'Médica e cientista Nise Yamaguchi reúne-se com o presidente Bolsonaro e propõe a adoção do tratamento precoce com cloroquina em todo o Brasil' + 'os resultados de pesquisas que comprovam que a hidroxicloroquina mata o vírus' + 'FDA, a agência americana de regulamentação de remédios aprovou o uso de hidroxicloroquina em todos os pacientes com covid-19'
Menor Comunidade	+ Carro + Multa + Militar + Máscara + Dirigindo	+ 'multa custa \$ 128 reais' + 'todos devem usar máscara dentro', + 'carro ! 🚗 multa vem' + 'polícia militar estarão multando' + 'detran irá multar motoristas'

Tabela 1. Três grupos mais relevantes encontrados e o menos relevante ao final com as palavras e frases-chave mais relevantes.

opponente e desqualificar fontes de conhecimento padrão como a mídia e a academia”. Conforme Viscardi (2020, p. 1153): “*Nós* poderia ser entendido como os homens comuns (“homens de bem” no contexto brasileiro?); *eles* seria entendido tanto como os estrangeiros quanto os indivíduos que pensam e/ou se comportam de maneira diferente do estabelecido/previsto por aquele grupo político”.

O discurso anti-sistema, ou *anti-establishment*, também é citado por Levitsky e Ziblatt (2018) ao pontua-

rem os quatro indicadores de comportamento autoritário identificados em líderes político que desestabilizaram democracias pelo mundo nos últimos anos: rejeição das regras democráticas, negação da legitimidade de oponentes políticos, tolerância ou encorajamento da violência, e propensão a restringir liberdades civis, inclusive de mídia são comportamentos comuns a políticos autoritários, como Fujimori, no Peru, e Chávez, na Venezuela, além de Donald Trump, nos Estados Unidos,

segundo os autores. Sendo Trump claramente uma inspiração para Jair Bolsonaro, não é de estranhar repetições desses comportamentos no contexto brasileiro.

[...] o populismo raivoso de Bolsonaro ganha força na exploração de um cenário de crise, em que os discursos de ódio ao Partido dos Trabalhadores (do candidato Fernando Haddad) e de combate à corrupção refletem a insatisfação de uma parte da população com os políticos brasileiros e os aproxima. Na mobilização desses afetos, Bolsonaro cria e fortalece a identificação com esse povo que, embora possa ter inúmeras diferenças, «se encontra» na equivalência da insatisfação crescente. Assim, a raiva de Jair Bolsonaro é uma marca que define e molda não só a sua mensagem, que se vê corporificada na persona do atual presidente, mas também define o próprio populismo, que se vale dessa dimensão para criar um sentido de união de um grupo e estabelecer a identificação desse grupo com seu líder. (VISCARDI, 2020, p. 1153)

Viscardi (2020) escreve tendo como objeto empírico de análise a campanha eleitoral de 2018, mas os elementos discursivos que a autora identifica naquele contexto se mantêm ou até se intensificam durante o enfrentamento da pandemia. Segundo ela, direcionar a outros atores as críticas a eles dirigidas é outra característica recorrente no discurso populista que pode contribuir para o reforço da falsa oposição entre “nós” e “eles”, dado que essas mensagens são também carregadas de informações de cunho conspiratório (Viscardi, 2020, p. 1154). Sem um oponente claro, como na disputa eleitoral, Bolsonaro e seus apoiadores parecem direcionar seu “populismo raivoso” a outros “adversários”, opondo-se a líderes de outras instâncias do poder, sejam eles governadores ou ministros do Supremo Tribunal Federal (STF), ou então voltam sua indignação à mídia e à ciência, não raro incitando ódio e violência. Todos esses elementos estão presentes nos agrupamentos semânticos mais frequentes na amostragem deste estudo preliminar, em mensagens que referem, por exemplo, o “ex presidente comunista”, o “alerta vermelho na Globo”, as supostas “irregularidades em estatísticas sobre coronavírus” e a constante propaganda da cloroquina e hidroxiclороquina como cura para Covid-19, enquanto sucessivas pesquisas científicas provam o contrário.

Adaptar métodos e critérios de *fact-checking*, incluindo

soluções tecnológicas e ferramentas de inteligência artificial capazes até mesmo de antever o crescimento de ondas conspiratórias de desinformação é uma urgência, mas também seria necessário, de certa forma, flexibilizar critérios de seleção do que é ou não verificável. Isso implicaria em assumir de forma inequívoca o compromisso com a desconstrução de narrativas que, além de desinformar, reforçam preconceitos, vieses negacionistas e princípios antidemocráticos, os quais costumam ser enquadrados como “juízos de valor” e, portanto, não objetivamente verificáveis pelos *fact-checkers*.

Nesse sentido, a transparência como princípio normativo distintivo do “jornalismo de verificação” deveria ser compartilhada em toda a produção jornalística, não apenas nos conteúdos declaradamente de *fact-checking*. É importante diferenciar o *fact-checking* enquanto manifestação empírica, que se constitui de um novo formato ou tipo de texto jornalístico, do “jornalismo de verificação” enquanto tipo ideal representativo de uma mudança em curso. Conforme Kovach e Rosenstiel (2014, p. X)¹⁸, o que define o jornalismo são seus princípios, não suas práticas, pois é possível reproduzir as mesmas práticas sob outros princípios. “Uma ‘notícia falsa’ é exatamente isso: algo que imita, que aparenta ser notícia, mas não é, porque está em desacordo com os princípios do jornalismo” (Seibt, 2019, p. 240). Sendo possível reproduzir as mesmas práticas sob outros princípios, a transparência, praticada não só em conteúdos do formato *fact-checking*, além de posicionar o jornalismo como um discurso qualificado no novo ecossistema, “[...] ajudará, a longo prazo, no surgimento de um público mais perceptivo. Um público que pode prontamente ver a diferença entre o jornalismo de princípio e a imitação descuidada e de interesse próprio” (Kovach e Rosenstiel, 2004, p. 131).

Seria preciso, no entanto, abdicar de outras práticas que correm em paralelo à checagem de fatos, muitas vezes comprometendo a credibilidade dos conteúdos verificados. Em um jogo de palavras com a “pós-verdade”, Zamith (2019, p. 147) se refere a esses fragmentos como “pós de verdade”. São “restos sem valor” de verdade que contaminam o discurso público, incluindo a publicação apressada sem confirmação, os títulos “caça-cliques”, o uso abusivo de fontes não identificadas, a reprodução acrítica e sem contexto de declarações e a publicação de conteúdos patrocinados ao lado de notícias (Seibt, 2020, p. 265). Por essa razão, não se pode dar por consolidada uma eventual mudança estrutural do jornalismo na contemporaneidade, ainda que se considere razoável conceber a configuração de um período “pré-paradigmático”

18 – Trata-se de uma versão revisada da obra original (ver referências finais). A paginação em números romanos é usada no prefácio da nova publicação.

(Kuhn, 1975). Marcado por instabilidade e experimentação, o limiar entre alterações gradativas e mudanças estruturais nem sempre é perceptível aos contemporâneos de uma etapa de transição, uma vez que “a revolução paradigmática ameaça não apenas conceitos, ideias e teorias, mas também o estatuto, o prestígio, a carreira de todos os que vivem material e psiquicamente da crença estabelecida” (Morin, 2011, p. 286).

Seja pela verificação, seja pela interpretação, incluindo a automatização de processos de seleção e apuração de fatos, as transformações em curso no ecossistema midiático pressionam a prática jornalística, o que afeta também a prática de *fact-checking*, alicerçada no princípio da objetividade. Ao mesmo tempo, é nessa manifestação empírica, como nova prática jornalística, que se encontram elementos úteis à problematização de mudanças nos elementos estruturais do paradigma jornalístico na contemporaneidade.

Considerações finais

Ainda que os resultados obtidos por nosso modelo computacional sejam promissores para a identificação de temas que podem propagar desinformação no ambiente digital, as ligações semânticas encontradas são insuficientes para criar um sistema de resposta automatizado. Isso porque não se pode sinalizar de forma preditiva um conteúdo potencialmente enganoso quando a imprecisão não está nos dados ou nos fatos, e sim no contexto discursivo. Por outro lado, o modelo testado neste estudo preliminar serve para dar mais elementos aos jornalistas e demais interessados em identificar conteúdos potencialmente enganosos com mais agilidade.

Nesse sentido, a novidade neste estudo está em combinar duas ou várias representações vetoriais de modelos de AM (*embeddings*) das mensagens suspeitas recebidas pelo **Comprova**, transformá-las para o domínio de redes complexas, e apresentando um mecanismo simples de acoplamento ou fusão das redes que representam cada modelo de AM, encontrar as comunidades e os elementos-chave que representam os temas principais tratados nas mensagens suspeitas. Tratando-se de um estudo preliminar, existem pontos que podem ser melhorados e explorados com maior profundidade, como o próprio mecanismo de combinação e construção da rede.

Além de rever certos processos e estratégias de comunicação, seria necessário também que o monitoramento e a identificação prévia de ondas de desinformação não ficasse restrito somente aos veículos ou setores jornalísticos especializados em verificação de fatos. Conteúdos desprezados pelos *fact-checkers*, por não atenderem aos critérios de seleção e avaliação próprios de sua metodologia, podem ser abordados em outros espaços e em outros formatos jornalísticos. Desde

que restabelecer os fatos no debate público e corroborar ao reposicionamento da verificação como elemento diferencial do jornalismo não seja algo restrito a um certo grupo de veículos de mídia (agências de checagem) ou editorias específicas em cada veículo (como ocorre com o **Projeto Comprova**), pode se consolidar uma mudança em direção a um “jornalismo de verificação”. Para isso, no entanto, é necessário que outras crises contemporâneas do jornalismo sejam superadas, seja em relação a suas fontes de financiamento seja na reconquista da credibilidade jornalística como um valor sociocultural, apenas para citar dois parâmetros estruturantes do paradigma jornalístico segundo Charron e Bonville (2016).

Por essa perspectiva, entendemos que nosso estudo apresenta alternativas promissoras para a atividade jornalística de forma mais ampla, não apenas no trabalho de checagem de fatos. Além disso, como toda a abordagem é baseada em aprendizado de máquina, quanto mais informações são analisadas, maior é a chance de o sistema aprender a identificar melhor os grupos semânticos e relacioná-los com casos passados, dando ao jornalista mais elementos para a tomada de decisão na difícil tarefa de verificar e comunicar informações. Contudo, pelas características das mensagens agrupadas no nosso modelo, seguirá sendo necessário que a tarefa de verificação seja compartilhada por mecanismos automatizados e verificadores humanos, de modo que se possa cruzar os indícios de imprecisão sugeridos por modelos computacionais, como os testados neste estudo preliminar, com as nuances discursivas que não são tão objetivas quanto a metodologia de *fact-checking* previa inicialmente.

Nossa conclusão central aponta para a impossibilidade de se desenvolver métodos computacionais preditivos para a verificação de fatos. Não se trata, portanto, de determinar o que é verdade, e sim de auxiliar o trabalho de seleção e identificação de ondas de desinformação, o que se torna cada vez mais central para se fazer jornalismo no ambiente comunicacional contemporâneo. Ao testar técnicas automatizadas de detecção de conteúdos suspeitos, identificamos um cenário em que, ao menos de forma isolada, nem o método da objetividade jornalística nem a automatização de mecanismos de verificação dão conta da complexidade do fenômeno contemporâneo da desinformação. É preciso avançar tanto no desenvolvimento de modelos automatizados de rápida detecção de conteúdos suspeitos quanto na capacidade interpretativa de verificadores e usuários em geral, através do letramento não apenas midiático, mas também científico e até mesmo cívico, diante do acirramento de tensões entre política, imprensa e desinformação num ecossistema midiático em transformação.

Referências

- ADALI, S.; LU X.; MAGDON-ISMAIL, M. 2014. *Local, community and global centrality methods for analyzing networks*. Social Network Analysis and Mining 4, no. 1 (2014): 210.
- ALSINA, M. 2009. *A construção da notícia*. Petrópolis, Vozes, 2009, 352p.
- ANDERSON, C.W.; BELL, E.; SHIRKY, C. 2013. Jornalismo Pós-Industrial: adaptação aos novos tempos. *Revista ESPM*, 5(2): 30-89.
- BELL, E.; OWEN, T. 2017. The platform press: how Silicon Valley reengineered journalism. *Tow Center for Digital Journalism*. Nova York, Columbia Journalism School, online. Disponível em: http://towcenter.org/wp-content/uploads/2017/03/The_Platform_Press_Tow_Report_2017.pdf. Acesso em: 23/07/2017
- BEREZOW A. 2019. *Debunkers Debunked: Who Fact-Checks the Fact-Checkers?* American Council on Science and Health, 6 Nov. 2019, www.acsh.org/news/2019/11/04/debunkers-debunked-who-fact-checks-fact-checkers-14378.
- BERTON, L.; LOPES, A.; VEGA-OLIVEROS, D. A. 2018. A Comparison of Graph Construction Methods for Semi-Supervised Learning. *In: International Joint Conference on Neural Networks*, Rio de Janeiro, 2018. Anais do International Joint Conference on Neural Networks, Rio de Janeiro, IJCNN, 1:1-8.
- BLONDEL V. D.; GUILLAUME J. L.; LAMBIOTTE R.; LEFEBVRE E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- BOJANOWSKI, P.; GRAVE, E.; JOULIN, A.; & MIKOLOV, T. 2017. *Enriching word vectors with subword information*. Transactions of the Association for Computational Linguistics, v. 5, p. 135-146.
- CESARINO, L. 2019. *On Digital Populism in Brazil*. PoLAR: Political and Legal Anthropology Review. Disponível em: https://polarjournal.org/2019/04/15/on-jair-bolsonaros-digital-populism/?fbclid=IwAR2JNsLh-hs jYwEwPFx6s11s4NAC0udsh0x6FY7Gy0nlU71UW_C-8DX1krA. Acesso em 19/04/ 2021.
- CHARRON, J.; BONVILLE, J. 2016. *Natureza e transformação do jornalismo*. Florianópolis, Insular, 400p.
- CLAUSET, A.; NEWMAN, M. E. J.; MOORE, C. 2004. *Finding community structure in very large networks*. Physical Review E. 70 (6): 066111.
- DA EMPOLI, G. 2020. *Os engenheiros do caos*. São Paulo: Vestígio, 192p.
- DALBEN S. 2020. *Jornalismo automatizado no Brasil: análise de três robôs no Twitter*. Brazilian Journalism Research. 3(16). Disponível em: https://bjr.sbpjor.org.br/bjr/article/download/1305/pdf_1. Acesso em: 2 fev. 2022
- DEVLIN, J.; CHANG, M. W.; LEE, K.; TOUTANOVA, K. 2018. *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.
- FORTUNATO S.; HRIC D. 2016. Community detection in networks: A user guide. *Physics Reports*, 659:1-44.
- FRANCISCATO C.; GONÇALVES A.L.F. 2021. *A inovação aberta no desenvolvimento das agências de fact-checking durante a pandemia*. Comunicação & Inovação. 22(50): 63-79. Disponível em: https://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/8074/3547 Acesso em 2 fev. 2022
- GOMES, W. 2009. *Jornalismo, fatos e interesses: ensaios de teoria do jornalismo*. Florianópolis, Insular, 112p.
- GRAVES, L. 2016. *Deciding what's true: the rise of political fact-checking in american journalism*. New York, Columbia University Press, 324p.
- GRUPPI, M., HORNE B., ADALI S. 2021. *Tell Me Who Your Friends Are: Using Content Sharing Behavior for News Source Veracity Detection*. arXiv preprint arXiv:2101.10973
- HORNE, B.; NØRREGAARD, J.; ADALI, S. 2019. Robust Fake News Detection Over Time and Attack. *ACM Trans. Intell. Syst. Technol.* 11, 1, Article 7 (February 2020), 23 pages. DOI:<https://doi.org/10.1145/3363818>
- _____. 2019. *Different Spirals of Sameness: A Study of Content Sharing in Mainstream and Alternative Media*. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 257-266.
- KOVACH, B.; ROSENSTIEL, T. 2004. *Os elementos do jornalismo*. São Paulo, Geração Editorial, 302p.
- _____. 2014. *The elements of journalism: revised and updated third edition*. New York, Three Rivers Press, 352p.
- KUHN, T. 1975. *A estrutura das revoluções científicas*. São Paulo, Perspectiva, 264p.
- LEVITSKY, S.; ZIBLATT, D. 2018. *Como as democracias morrem*. Rio de Janeiro, Brasil: Zahar, 270p.
- LIPPMAN, W. 2008. *Opinião Pública*. São Paulo, Vozes, 2008, 352p.
- MELE, N.; LAZER, D.; BAUM, M.; GRINBERG, N.; FRIEDLAND, L.; JOSEPH, K.; HOBBS, W.; MATTSSON, C. 2017. *Combating fake news: An agenda for research and action*. Conference Organized by Matthew Baum (Harvard), David Lazer (Northeastern) and Nicco Mele (Harvard). Sponsored by Northeastern and Harvard University, February 17-18.

- MEYER, P. 1973. *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Indiana, Indiana University Press, 304p.
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. 2013. *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781.
- MORIN, E. 2011. *O método 4: as ideias: habitat, vida, costumes, organização*. Porto Alegre: Sulina, 319p.
- Organização Panamericana de Saúde. 2020. Entenda a infodemia e a desinformação na luta contra a Covid-19. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52054/Factsheet-Infodemic_por.pdf?sequence=14. Acesso em: 04/01/2021
- PRIOR, H. 2019. *Em nome do povo: o populismo e o novo ecossistema mediático*. In: Figueira J., Santos S. (ed.). *As fake news e a nova ordem (des)informativa na era da pós-verdade*. Coimbra, Imprensa da Universidade de Coimbra, p. 123-47.
- REIMERS, N.; GUREVYCH I. 2020. Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation. In: International Conference on Empirical Methods in Natural Language Processing, 2020. Anais do 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 1:4512–4525.
- ROSSETTI G.; CAZABET R. 2018. Community discovery in dynamic networks: A survey. *ACM Computing Survey*, 51(2):1-37.
- SCHWARZ, S.; THEÓPHILO, A.; ROCHA, A. Emet: Embeddings from multilingual-encoder transformer for fake news detection. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.: s.n.], 2020. p. 2777–2781
- SEIBT, T. 2019. *Jornalismo de verificação como tipo ideal: a prática de fact-checking no Brasil*. Porto Alegre, RS. Tese de doutorado. Universidade Federal do Rio Grande do Sul, 265p.
- _____. 2020. Uma coletânea para alargar o olhar sobre a 'nova ordem (des)informativa na era da pós-verdade'. *Reciis*. 1(14): 261-7.
- SILVA T. C.; ZHAO L. 2016. *Machine Learning in Complex Networks*. 1ed, Springer International Publishing, Cham, 331 p.
- SPONHOLZ, L. 2009. *Jornalismo, Conhecimento e Objetividade: Além do Espelho e das Construções*. Florianópolis, Insular, 2009, 192p.
- TRAAG V. A.; WALTMAN L.; VAN ECK N. J. 2019. From louvain to leiden: guaranteeing well connected communities. *Scientific reports*, 9(1):1–12.
- VEGA-OLIVEROS, D. A.; GOMES S. P.; MILIOS, E. E.; BERTON, L. 2019. A multi-centrality index for graph-based keyword extraction. *Information Processing & Management*, 56(6):1020-63.
- VISCARDI, J. M. 2020. *Fake news, verdade e mentira sob a ótica de Jair Bolsonaro no Twitter*. Trabalhos em Linguística Aplicada, vol. 59, no. 2. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/tla/article/view/8658477>. Acesso em 26/03/2021.
- WAGNER FILHO, J. A.; WILKENS, R.; IDIART, M.; VILLAVICENCIO, A.; 2018. The brwac corpus: A new open resource for brazilian portuguese. In International Conference on Language Resources and Evaluation, nº 11, 2018. Anais do the Eleventh International Conference on Language Resources and Evaluation, LREC, 1:4339–44.
- WARDLE, C.; DERAKHASHAN, H. 2017. *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe Report. 27 set. 2017. Disponível em: <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>. Acesso em 10/07/2018.
- WEBER, M. 2014. *Economia e Sociedade: fundamentos da sociologia compreensiva*. Brasília, Editora UnB, 584p.
- ZAMITH, F. 2019. *Pós de verdade: quando o (ciber)jornalismo se contenta com pouco*. In: FIGUEIRA, J.; SANTOS, S. (ed.). *As fake news e a nova ordem (des)informativa na era da pós-verdade*. Coimbra, Imprensa da Universidade de Coimbra, p. 147-66.
- ZHOU, X.; AND ZAFARANI, R. 2019. Network-based Fake News Detection: A Pattern-driven Approach. *SIGKDD Explor. Newsl.* 21, 2 (December 2019), 48–60. DOI: <https://doi.org/10.1145/3373464.3373473>