

Edição v. 44
número 2 / 2025

Contracampo e-ISSN 2238-2577
Niterói (RJ), 44 (2)
mai/2025-ago/2025

A Revista Contracampo é uma revista eletrônica do Programa de Pós-Graduação em Comunicação da Universidade Federal Fluminense e tem como objetivo contribuir para a reflexão crítica em torno do campo midiático, atuando como espaço de circulação da pesquisa e do pensamento acadêmico.

DOSSIÊ

Como as plataformas produzem
desinformação: interfaces, algoritmos e
conteúdos falsos

How platforms produce disinformation:
interfaces, algorithms, and false content

FREDERICO OLIVEIRA

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) – Brasília,
Distrito Federal, Brasil.
E-mail: freddroliveira@gmail.com
ORCID: <https://orcid.org/0000-0001-5653-5715>

ANDRÉ LEMOS

Universidade Federal da Bahia (UFBA) – Salvador, Bahia, Brasil.
E-mail: almlmos@gmail.com
ORCID: <https://orcid.org/0000-0001-9291-6494>

AO CITAR ESTE ARTIGO, UTILIZE A SEGUINTE REFERÊNCIA:

OLIVEIRA, Frederico; LEMOS, André. Como as plataformas produzem desinformação: interfaces, algoritmos e conteúdos falsos. **Contracampo**, Niterói, v. 44, n. 2. 2025.

Submissão em: 02/09/2024. Revisor A: 22/02/2025; Revisor B: 17/07/2025. Aceite em: 04/08/2025.

DOI – <http://dx.doi.org/10.22409/contracampo.v44i2.64045>

Resumo

O artigo demonstra a relação entre a interface e algoritmos das plataformas e a produção, consumo e refutação de conteúdos falsos relacionados ao *kit gay*, à Vaza Jato e à Covid-19. Para tanto, reúne o diário do campo dos pesquisadores, com incursões em grupos públicos sobre política no Telegram e no WhatsApp entre 2018 e 2022; realiza análise de interface do WhatsApp e a coleta e discussão de tuítes sobre o *kit gay*. Os resultados apontam que as interfaces das plataformas: a) produzem tipos específicos de desinformação; b) determinam estratégias para que a circulação de um conteúdo falso seja ampliada; e c) definem condições para a refutação de conteúdos falsos. Sendo assim, no atual contexto de platformização, dataficação e performatividade algorítmica, não é possível compreender a desinformação sem analisar as materialidades das plataformas.

Palavras-chave

Plataformas; Interface; Desinformação.

Abstract

The article demonstrates the relationship between platform interfaces and algorithms and the production, consumption, and refutation of false content related to the “gay kit,” Vaza Jato, and COVID-19. To this end, it gathers the researchers’ field diaries, with incursions into public groups on Telegram and WhatsApp about politics between 2018 and 2022. It conducts an interface analysis of WhatsApp and collects and discusses tweets about the “gay kit.” The results indicate that platform interfaces a) produce specific types of misinformation, b) determine strategies to expand the circulation of false content, and c) define conditions for refuting false content. Thus, in the current context of platformization, datafication, and algorithmic performativity, it is impossible to understand misinformation without analysing the materialities of platforms.

Keywords

Platforms; Interfaces; Disinformation.

Introdução

As plataformas participam da produção, da circulação, do consumo e da refutação de desinformação¹ (Lemos; Oliveira, 2021), seja pela intervenção de suas interfaces e algoritmos; pela influência na seleção noticiosa de agências de checagem; ou por serem proprietárias de tecnologias adotadas em sistemas automatizados de verificação. Em um contexto de plataformização, dataficação e performatividade algorítmica - PDPA (Lemos, 2020), não é possível compreender a desinformação sem analisar a materialidade das plataformas. Afinal, “as *affordances* das ferramentas se tornam o horizonte da verdade”, como destaca Crawford (2018, p. 133, tradução nossa).² As interfaces balizam as formas de veiculação de desinformação específicas (Oliveira, 2019). Por sua vez, os algoritmos desenvolvem a curadoria de conteúdos e uma sociabilidade (Van Djick; Poell; De Wall, 2018, Bucher, 2018, Gillespie, 2018), sendo dispositivos de visibilidade e de reconhecimento (Crawford, 2021, Amoore, 2020).

O fenômeno da desinformação não é, portanto, a simples transposição da mentira ou da informação falsa. É resultado de uma maquinação que tem nos algoritmos e nas interfaces das plataformas agentes constitutivos. A desinformação é um fenômeno que se transforma com a cultura digital e a atual sociedade de plataformas (as *fake news* são um fenômeno do digital, resultado direto da plataformização). Entender as suas dinâmicas e materialidades é uma forma de criar mecanismos técnicos, legais, educacionais, informacionais para combatê-la.

O objetivo deste artigo é demonstrar a agência material das plataformas na constituição contemporânea do fenômeno da desinformação, por meio da análise de três casos e das interfaces que os mediarão.³ Analisa-se a relação entre a interface e algoritmos das plataformas e a produção, consumo e refutação de conteúdos falsos relacionados ao *kit gay*, à Vaza Jato e à Covid-19. É importante ressaltar a difícil tarefa de distinguir a interface dos algoritmos dessas plataformas, já que ambos se correlacionam e compõem o mesmo dispositivo.

A análise reúne o diário do campo dos pesquisadores, com incursões em grupos públicos sobre política no Telegram e no WhatsApp entre 2018 e 2022; baseia-se na análise de interface do WhatsApp; assim como na coleta e análise de tuítes sobre o kit gay. O artigo discute a literatura sobre desinformação, a relação entre as interfaces e a desinformação, relata os procedimentos metodológicos adotados e os resultados desta investigação, que demonstram que interfaces: *a)* produzem tipos específicos de desinformação; *b)* determinam estratégias para que a circulação de um conteúdo falso seja ampliada; e *c)* definem condições para a refutação de conteúdos falsos.

Plataformas e desinformação

A cultura digital contemporânea pode ser definida a partir da plataformização, da dataficação e da performatividade algorítmica (Lemos, 2020). A plataformização é “a penetração de infraestruturas, processos econômicos e estruturas governamentais de plataformas em diferentes setores econômicos e esferas da vida”, um arranjo técnico, social, político e econômico específico, como explicam Poell, Nieborg

1 Fake news são mensagens falsas que parecem verdadeiras, simulando cadeias de referências (Lemos; Oliveira, 2021). Distribuídas intencionalmente para enganar grupos ou com objetivos econômicos e políticos, circulam em cascatas de informação nas plataformas, que as coproduzem – sendo, portanto, resultado da PDPA. Nesse artigo usaremos o termo “desinformação”, mais abrangente.

2 “The affordances of the tools become the horizon of truth” (Crawford, 2018, p. 133).

3 Por “agência” entendemos as mediações originárias de uma relação entre objetos (humanos e não humanos) produzindo uma ação específica. Essa noção encontra-se desenvolvida por Latour (2005) e Barad (2007). Agência não é exclusividade de uma subjetividade ou intencionalidade racional, não pertence a um indivíduo pois todo indivíduo é formado por nexos. Agência é intra-ação, performance (*enactment*) distribuída. Trata-se de reconhecer, nesse artigo, as ações desencadeadas pelas interfaces e algoritmos das plataformas como “intra-ações” mobilizando outras plataformas, informações e usuários.

e Van Djick (2020). A dataficação, por sua vez, envolve não somente a pervasiva extração de dados, mas a política de produção e extração, além de um discurso sobre objetividade dos dados (Lemos, 2021). Por fim, a performatividade dos algoritmos produz novas sociabilidades (Bucher, 2018), assim como define a experiência dos usuários nas plataformas.

As plataformas são arquiteturas informacionais que não somente organizam a interação do usuário, mas são pensadas para maximizar a coleta e processamento de dados de uso. Também são definidas por acordos econômicos e políticos, que medeiam seu funcionamento e a experiência do usuário (Van Djick; Poell; De Wall, 2018). Podem ser entendidas como:

(re) programmable digital infrastructures that facilitate and shape personalized interactions between end users and complementarians organized through systematic collection, algorithmic processing, monetization, and circulation of data (van Dijck; Neiborg; Poell, 2020, p. 4).

Considera-se, como interface, a estrutura de comunicação e captura de informação que permite que usuários – inclusive outras plataformas – possam interagir com uma determinada plataforma. Pode se tratar de uma “entrada” de dados na plataforma, uma ferramenta que recebe requisições de outros sistemas (como as interfaces de programação de aplicações, as APIs), assim como estar relacionada ao modo como a informação é apresentada ao usuário, a interface gráfica de usuário (GUI). O conceito de interface também abarca as possibilidades de interação entre dados e metadados pelos algoritmos de aprendizado de máquina.

Importa destacar que essas maneiras de interação não pertencem ao usuário. Este deve se adaptar ao que a plataforma oferece como possibilidade e permissividade (*affordance*). Nesse sentido, a escolha de uma determinada interface tem em si uma estratégia de produção, uso, processamento e distribuição da informação, geralmente invisível ao usuário. Quando uma interface permite, por exemplo, que alguém possa seguir outra pessoa ou uma instituição como “amigos” (não há outra forma de nomeá-la), produz ontologicamente, para efeitos de informação, “amigos”.

No atual contexto da sociedade de plataformas (Van Djick; Poell; De Wall, 2018), a desinformação é produzida e disseminada pelas plataformas digitais: a plataforma não é um simples canal de difusão, mas um produtor e acionador de cadeias desinformativas pela sua interface e algoritmos performativos. Isto porque sua materialidade (as *affordances* das interfaces e algoritmos) constitui as condições técnicas para a produção, para a circulação, consumo e refutação de conteúdo falso on-line. Estudos apontam essa agência material discursiva na ação dos *bots* no compartilhamento, na intervenção algorítmica dos “filtro-bolha”, nos *dark patterns* das plataformas que potencializam a distribuição de conteúdo nocivo, por exemplo (Lemos; Bitencourt; Santos, 2021, Bakir; Mcstay, 2017, Tandoc; Lim; Ling, 2017, Bacarella *et al.*, 2018, Torres; Gerhart; Negahban, 2018, Bovet; Makse, 2019). A literatura destaca a relação entre interfaces, performatividade algorítmica e desinformação (Bakir, Mcstay, 2017, Bacarella *et al.*, 2018, Tandoc; Lim; Ling, 2017, Torres; Gerhart; Negahban, 2018, Santaella, 2018, 2021b, Lee; Wei, 2022, Lemos; Bitencourt; Santos, 2020, Melo *et al.*, 2020, Lemos; Oliveira, 2021, Cesarino, 2022, Lee; Wei, 2022).

Um conteúdo desinformativo não pode, portanto, ser considerado fora de seu contexto de circulação, não sendo possível ignorar a rede de atores (objetos – neles incluídos interfaces e algoritmos –, discursos, políticas regulatórias e outros) que dele participa. Tendo em vista tais contextos e, partindo da classificação de Rahmanian (2022), busca-se sistematizar, no Quadro 1, um inventário de algumas formas de desinformação, como se fundamentam (inscrições) e sua dependência das plataformas.

Quadro 1 – Proposta de classificação de formas de desinformação

Tipo	Definição	Inscrições	Dependência das plataformas
<i>Alt-science</i>	Revisão conspiracionista da ciência, que nega a produção do conhecimento científico; acusa a ciência de aliar-se a interesses capitalistas; propõe soluções relacionadas ao empirismo ou à religião.	Evidência parca, mas que não nega totalmente o método científico, quer subvertê-lo (Cesarino, 2022).	Prescinde das plataformas, mas nela sua produção e distribuição são fortalecidas.
<i>Clickbait</i>	Uso de títulos que geram a sensação de urgência, curiosidade ou exclusividade. Geralmente o conteúdo não está relacionado ao título, que é pensado a partir de técnicas de SEO.	Não há relação entre conteúdo e título, de forma que as inscrições se rompem.	Envolve técnicas de SEO e o algoritmo de ranqueamento de buscadores.
Conteúdo fabricado	Conteúdo criado para defesa de uma proposição, passando-se por declaração ou publicação legítima.	Não se sustentam, já que o documento é forjado.	Pode aproveitar ferramentas das plataformas, mas não há uma dependência delas.
Conteúdo impostor	Se passa pelo conteúdo postado por uma página, pessoa ou instituição com credibilidade. Pode até ser conteúdo verdadeiro, mas a usurpação da identidade é problemática.	Neste caso, a cadeia de referências pode ser extensa, mas não se sustenta quando se considera a autoria do conteúdo.	Pode aproveitar ferramentas das plataformas, mas não há uma dependência delas.
Conteúdo manipulado	Conteúdos modificados para a defesa de uma proposição.	Têm a aparência de verdadeiros em função das inscrições, mas a falsificação resulta no desmonte dessas provas.	As manipulações podem se aproveitar das ferramentas das plataformas, mas delas não dependem.
Erros jornalísticos	Ocorrem quando a metodologia de apuração jornalística não é suficiente para produzir <i>móveis imutáveis</i> que relatem o fato discutido.	A cadeia de referências pode até ser extensa, mas alguma deficiência na apuração pode levar ao seu desmonte.	Prescindem das plataformas, que podem ser espaço para correção e sanção pública aos erros jornalísticos.
<i>Fake news</i>	Termo guarda-chuva que reúne conteúdos verificadamente falsos, cuja distribuição é pensada para ampliar seus danos e circulação. Sua produção, circulação, consumo e refutação dependem das plataformas.	Por tratar-se de termo “guarda-chuva, pode-se encontrar número de inscrições variáveis e estratégias distintas. Pretensão factual (Dourado, 2020).	Conteúdo falso produzido no âmbito das plataformas, cuja produção, circulação, consumo e refutação delas dependem.
Falso contexto	Conteúdos verdadeiros apresentados em contextos distintos do original. Desse modo, passam a defender aspectos que não fundamentam.	Pode apresentar evidências fortes, mas seu uso fora de contexto coloca sua validade em xeque.	Prescinde das plataformas.

Tipo	Definição	Inscrições	Dependência das plataformas
Rumor	Iniciado por propagadores conscientes, que acreditam ou não no que dizem. São espalhados em cascatas e pela polarização de grupos (Sustein, 2009). Materialidade está relacionada ao corpo e à produção do discurso. Embora o rumor seja materialmente produzido, sua duração é curta. São seus efeitos materiais que têm maior durabilidade (Burrell, 2012).	Um rumor não é necessariamente falso, mas faltam inscrições que permitam sua verificação.	Os rumores prescindem das plataformas. No entanto, elas podem potencializar a produção e distribuição de rumores.
Teorias da conspiração	“Conspiracionismos são bem difundidos no espaço e no tempo, e tipicamente se associam a contextos de crise e dissonância, onde trabalham ao mesmo tempo o sintoma e a cura” (Cesarino, 2022, p. 230).	Evidência parca. Promove o desmonte das cadeias de referência da ciência, afirmando que são falsificadas. Haveria como provar a proposição conspiracionista, mas interesses escusos impediriam tal verificação pela ciência.	Também prescindem das plataformas. No entanto, elas podem potencializar a produção e distribuição de conteúdos conspiratórios.

Fonte: Adaptado a partir de Oliveira (2023)

No quadro 1, são destacadas inscrições específicas relacionadas às formas de desinformação. Há uma pretensão factual (Dourado, 2020), ou seja, a desinformação simula cadeias de referências⁴ sólidas (Lemos; Oliveira, 2021). As interfaces das plataformas, como se aponta adiante, contribuem na montagem dessas inscrições, ao facilitar o envio em massa de conteúdos; ao fazerem parecer que não há mediação; ou quando apagam metadados de imagens, por exemplo. As interfaces apoiam a distribuição da desinformação: nas plataformas de redes sociais, ela se dá por meio de cascatas de desinformação (Sustein, 2009; Recuero; Grunzt, 2019; O’Connor; Weatherall, 2019; Calvo; Aruguete, 2020). O Quadro 1 aponta, também, a relação entre plataformas e formas específicas de desinformação.

A intervenção das interfaces (e dos algoritmos)

Como se apontou, a plataformação se alimenta de uma ampla coleta de dados (dataficação) em todos os setores da vida social (Lemos, 2021). Resultado da plataformação e da dataficação, a curadoria informacional promovida pelos algoritmos (Bucher, 2018; Amore, 2020; Crawford, 2021) baliza formas de sociabilidade por consumo e circulação de determinados conteúdos informativos. Na cultura digital, a desinformação é uma perturbação que resulta do uso abusivo da gramática e da lógica de monetização das plataformas (Lemos, 2024). Aponta-se a predileção das plataformas por conteúdos falsos, que viralizam com maior facilidade do que textos jornalísticos – gerando grande engajamento (Bounegru *et al.*, 2017; Vosogui; Roy; Aral, 2018; Lemos; Bitencourt; Santos, 2020; Oliveira, 2020; Lemos; Oliveira, 2021).

Tome-se, por exemplo, o WhatsApp. Canavilhas, Colussi e Moura (2019) analisaram grupos do app nas eleições de 2018 e perceberam que os usuários preferem enviar imagens (48,5%) e vídeos (32,6%). Os *links* também têm grande importância, com 8,7%. Tais imagens pouco duram no aplicativo: 80% delas

⁴ São “redes de rastros documentais que atestam veracidade de um evento” (Lemos; Oliveira, 2021, p. 74).

são compartilhadas por até dois dias, sendo que apenas 20% podem viralizar (Melo *et al.*, 2020). Nesse processo, a interface do mensageiro promove o anonimato, já que o WhatsApp apaga os metadados de um conteúdo quando o comprime para enviá-lo (Arun, 2019) – esta é uma forma concreta na qual a materialidade da plataforma agencia e potencializa a desinformação. As informações são apagadas e conteúdos antigos que são repostados parecem atuais, já que nem sempre possível ver sua data de criação (Mukherjee, 2020).

Por sua vez, o elevado grau de clusterização (agrupamentos) do WhatsApp aponta redes mais densas, com maior número de conexões entre os membros (Melo *et al.*, 2020a). Isso ocorre porque a interface do aplicativo permite grupos com, no máximo, 250 membros, o que limita a conversação a *clusters* reduzidos:⁵ ou seja, nos grupos, o usuário sabe (e geralmente confia) com quem conversa. A literatura ainda destaca a relação entre a confiança de um usuário em um conteúdo que acessa nas redes sociais e o conhecimento prévio de quem encaminhou aquele conteúdo (Shafi; Ravikumar, 2018), o que também é consequência do número limitado de participantes de grupos.

Para além disso, chamadas de vídeo, áudios e a instantaneidade das mensagens fazem parecer que o *app* oferece acesso imediato e íntimo ao real, não mediado (Cesarino, 2022). No WhatsApp, portanto, prevalece a constante reatualização de conteúdos já publicados – já que o aplicativo apaga os metadados das mídias postadas – e a homofilia. No *app*, também é mais fácil inflar a dimensão de um debate (*astroturfing*) – em função da densidade da rede e das ferramentas do aplicativo que fazem parecer que não há mediação e que todos concordam entre si. Observe-se, portanto, que a interface do mensageiro intervém no modo como os conteúdos são percebidos e compartilhados.

Diante do uso do mensageiro para disseminar desinformação, a Meta implantou em 2018 um limite de encaminhamentos concomitantes de um conteúdo e retirou, da interface do aplicativo na Índia, o botão de compartilhamento rápido: isso impede que um usuário seja um *hub* de conteúdos falsos, embora apenas atrase a viralização de mensagens (Melo *et al.*, 2020). Lee e Wei (2022) criticam as restrições promovidas pela plataforma, já que o compartilhamento ativo de informações entre usuários é limitado. Os autores acreditam ser mais eficiente que se incentive a participação responsável e que o usuário verifique a informação que deseja encaminhar. Além disso, para driblar a limitação imposta pelo *app* basta copiar a mensagem e colar em outro *chat*. Trata-se, inclusive, de “estratégia [desinformativa] para a) manter o tom de conversa, sem que se pareça reproduzir informações ou conteúdos; e b) ampliar o compartilhamento da informação a partir da reputação de quem a envia” (Oliveira, 2019, p. 10-11).

O público que essas plataformas produzem – pela ação de suas interfaces e algoritmos – está envolvido em uma sugestibilidade sem contato (Tarde, 2005), cuja subjetividade é produzida pelo reforço das suas crenças. Como destaca Cesarino (2022), são públicos antiestruturais, marcados pela economia da atenção. Um exemplo: a centralidade do WhatsApp no ecossistema da desinformação brasileiro (Machado *et al.*, 2020), a homofilia promovida pelo aplicativo e a constante atualização de conteúdos fazem emergir um ambiente em que o *call to action* é uma estratégia de desinformação válida e muito eficiente (Cesarino, 2022). São públicos efêmeros e disruptivos, que aderem aos espaços de homofilia e confiança que são os grupos. Isso não se opera sem as materialidades das plataformas envolvidas.

Contudo, a desinformação não está restrita ao mensageiro, mas perpassa todo o ecossistema de plataformas. Machado e colegas (2020) apontam uma *pipeline* entre YouTube e WhatsApp, enquanto Melo e outros pesquisadores (2020b, 2022) inventariam os *links* compartilhados no mensageiro – geralmente, para outras plataformas como o Telegram, o Twitter/X e o YouTube. Editores do Jornal da Cidade Online e de outras páginas compartilham suas publicações nos grupos públicos de debate político no Telegram acompanhados durante a pesquisa, buscando garantir maior número de acesso às páginas. Há de se considerar, também, as cascatas de informação, a postagem constante de conteúdos sobre um

5 A introdução da ferramenta “Comunidades” permitiu a criação de *chats* com maior número de membros, o que pode ampliar o número de *clusters* e reduzir a densidade da rede.

mesmo tema (Sustein, 2009, Recuero; Grunzt, 2019, O'Connor; Weatherall, 2019, Calvo; Aruguete, 2020), que fazem parecer que um debate é mais discutido do que de fato é – *astroturfing*.

Nesse contexto, a agência das interfaces e dos algoritmos das plataformas importa. Embora o WhatsApp e o Telegram não façam ranqueamento de conteúdos por meio de algoritmos de relevância pública (Gillespie, 2018), a distribuição de tais conteúdos no YouTube e no Twitter/X é resultado da curadoria algorítmica. A partir de seu repertório, o usuário escolhe que conteúdos envia para um mensageiro: esse repertório, contudo, é composto também por conteúdos filtrados pelos algoritmos em plataformas como o YouTube (Machado *et al.*, 2020).

Interface e algoritmos são uma amálgama, mas, por suas características, podem ser investigadas por meio de estratégias distintas. Enquanto as interfaces podem ser analisadas por sua descrição – com ferramentas como o método passo a passo (Light; Burgess; Duguay, 2017) –, a performatividade algorítmica pode ser percebida a partir da seleção de eventos específicos e o inventário de seus efeitos. Embora os algoritmos sejam considerados uma caixa-preta, é possível identificar o que fazem-fazer, que atores convocam – o que coloca em discussão o quão opacos são (Bucher, 2018, Amoores, 2020, Crawford, 2021). Não se pode olvidar, porém, que interfaces e algoritmos se interrelacionam e que a visibilidade oferecida pelas plataformas resulta dos algoritmos, mas, também, dos *softwares* e interfaces em que rodam. Embora interfaces e algoritmos sejam elementos distintos que compõem uma plataforma (Van Dijk; Poell, De Wall, 2018), eles operam em rede, formam um dispositivo e intra-ações são performadas.

Assim, enquanto a interface das plataformas formata conteúdos falsos, os empacota em determinada estrutura específica (Lemos; Oliveira, 2021), os algoritmos definem as condições de circulação: o uso de palavras específicas pode implicar em *shadowban* ou ampliar a distribuição de uma mensagem (estratégias de SEO se baseiam nisto), por exemplo. As interfaces buscam padronizar a ação do usuário e a mediação dos algoritmos produz a dataficação (extração de dados e produção de visões sobre o mundo) de acordo com os interesses da plataforma. As diversas formas da desinformação, como destaca Lemos (2024), exploram a lógica algorítmica das plataformas. Sua distribuição é marcada por uma relação afetiva, emocional. Portanto, elas não resultam de erros ou falhas no funcionamento das plataformas, mas geram efeitos ético-valorativos perturbadores no uso dentro do *modus operandi* rotineiro das interfaces e dos algoritmos. Nessa perspectiva, garantir maior transparência algorítmica e pensar a regulação é necessário para o enfrentamento desta perturbação (Lemos, 2024).

Para evitar esses efeitos da desinformação, adota-se estratégias de moderação de conteúdo. Dentre estas estão: remoção de conteúdo e usuários; limitação de acesso por idade ou localização geográfica, além de bloqueios temporários; etiquetas de checagem e alertas sobre conteúdos inadequados; desmonetização e derrubada de redes. No entanto, as plataformas geralmente optam pela redução da visibilidade dessas postagens, solução menos agressiva (Gillespie, 2022), que não impede a circulação de conteúdos de desinformação. Outra estratégia é adoção de programas de *fact-checking*, também marcados pela lógica da plataformização, assimetrias financeiras, políticas e técnicas (Oliveira, 2023).

Assim, até mesmo na refutação dos conteúdos falsos a lógica de mediação técnica aparece: uma estratégia das plataformas no combate à desinformação era a realização de convênios com agências de checagem, marcados por constrangimentos e assimetrias – financeiras, políticas etc. (Jurno, 2020, Whittaker, 2021). No caso do Programa de Verificação de Fatos Independente da Meta (3PFC)⁶, por exemplo, a seleção noticiosa promovida pelas agências de checagem participantes é promovida pela plataforma (Lemos; Oliveira, 2021). No programa da Meta, um *feed* específico é oferecido às agências, que ali podem selecionar conteúdos que desejam checar, assim como usar ferramentas oferecidas pelas plataformas – a exemplo do Crowdtangle, descontinuado em agosto de 2024. Observa-se, contudo, que

6 Em janeiro de 2025, a Meta anunciou que o 3PFC seria descontinuado nos Estados Unidos e que realocaria a equipe de moderação de conteúdo para o Texas. Em seu Centro de Transparência, a plataforma afirma que essas mudanças se restringem àquele país (Meta, 2025).

a checagem jornalística depende das plataformas e suas políticas de acesso às interfaces de programação de aplicações (API), que são interfaces de comunicação estratégicas para se entender o funcionamento das plataformas. Mesmo iniciativas próprias de monitoramento, como o *Radar Aos Fatos*, dependem da coleta de dados por meio de APIs. Ou seja, novas interfaces são necessárias para analisar as ações e performances de interfaces do usuário e performatividade algorítmica.

A partir de tais considerações, tem-se alguns aspectos que guiam esta análise de como as plataformas, interfaces e algoritmos produzem desinformação: *a)* considerar os conteúdos falsos em seu contexto de circulação, partindo da compreensão que as condições de produção, circulação e consumo participam da definição da “mensagem”; *b)* compreender interface-algoritmo como rede, sendo importante uma análise contextual e situada; e *c)* entender que as *fake news* são perturbações que resultam do pleno funcionamento das plataformas, a partir de um uso abusivo de sua gramática e lógica de monetização (Lemos, 2024). Esses princípios orientam os procedimentos metodológicos de análise dos casos, nos quais se observam a participação das plataformas na produção, circulação, consumo e refutação da desinformação.

Procedimentos metodológicos

O interesse desta pesquisa é a compreensão de como as plataformas, através da agência de suas interfaces e algoritmos, produzem e disseminam desinformação,⁷ definida neste artigo como conteúdos falsos que não se sustentam por meio de cadeias de referências sólidas (Lemos; Oliveira, 2021, Lemos; Reis, 2021). Buscando seguir os princípios destacados na seção anterior, optou-se por uma pesquisa básica, descritiva, com abordagem qualitativa, a partir da análise de três casos: o *kit gay*, a Vaza Jato e a Covid-19. Para identificar a intervenção da rede interface-algoritmo, optou-se pela descrição de eventos específicos (Bucher, 2018) e das interfaces envolvidas (Light; Burgess; Duguay, 2017), em uma perspectiva pragmática fundamentada na metodologia proposta por Lemos (2020).

Tais procedimentos envolvem quatro etapas, a saber: *a)* identificação dos modos, quando se observa a questão de interesse e seu contexto; *b)* inventário de humanos e objetos que participam da questão; *c)* descrição das mediações que produzem o fenômeno; e *d)* reunião dos dados, reagregando-os a partir de uma proposição que explique como humanos e objetos coproduzem tal contexto (Lemos; 2020). Tal estratégia permite a descrição das interfaces, garantindo as condições para sua análise, enquanto também oferece insumos para o registro dos efeitos da performatividade algorítmica – permitindo, então, entender o *output* dos algoritmos.

A seleção de casos pautou-se na relevância de tais eventos para o debate político on-line nos últimos anos, sendo marcados pela distribuição de conteúdos falsos. Além disso, foram temas amplamente discutidos nas plataformas, sendo possível a coleta de postagens e o registro, em diário de campo, de características das plataformas e da discussão on-line. Indica-se ao leitor que, em função da exiguidade do espaço, optou-se por destacar aspectos específicos de cada caso, a fim de apontar como as plataformas coproduziram a desinformação em cada um dos eventos – não há, portanto, tratamento igual aos casos, mas se destaca aspectos mais relevantes e se aponta no que eles se distinguem.

Há uma questão de interesse e um contexto bem demarcados (modo): buscou-se analisar a desinformação on-line e sua relação com as intervenções das plataformas, suas interfaces e algoritmos. Para tanto, é necessário inventariar as interfaces e indicar as formas de mediação, ou seja, como interferem na produção, circulação, consumo e refutação de desinformação. Também é importante selecionar os

7 A análise restringe-se à desinformação on-line. Considerando a tipologia apresentada no Quadro 1, o artigo especificamente as *fake news*, que são produzidas para se apropriar da gramática das plataformas. No entanto, outras estratégias / formas de desinformação se misturam, até mesmo aquelas que prescindem das plataformas.

eventos – para, então, entender a ação dos algoritmos –, buscando fontes “síncronas” - documentos publicados enquanto havia uma controvérsia aquecida, como tuítes – “assíncronas” – o diário de campo do pesquisador, a análise da interface etc.

Ainda no inventário, também se participou de grupos públicos de política no WhatsApp entre 2018 e 2022, assim como foi realizada a coleta de tuítes sobre os casos analisados⁸ – *kit gay*, Vaza Jato e Covid-19. Tais eventos foram selecionados por sua relevância na política brasileira apontando como desinformação e interfaces se relacionam. Mais uma vez, considerando limitações éticas, apenas tuítes de pessoas públicas são reproduzidos. Ressalta-se que tais procedimentos foram desenvolvidos após a aprovação de comitê de ética em pesquisa. A apresentação das fontes, estratégias de coleta, tratamento e análise (inventário) é feita no Quadro 2:

Quadro 2 – Fluxograma da metodologia

Fonte	Coleta e tratamento	Análise	Resultados esperados
Tweets em português com o termo “ <i>kit gay</i> ” (289 mil) e a <i>hashtag</i> #showdopavao	Coleta por meio do <i>twint</i> (Python)	Análise de Redes Sociais	Compreensão da distribuição de tuítes sobre os casos.
Conteúdos falsos compartilhados no WhatsApp sobre os casos	Coleta por meio do Monitor de WhatsApp da UFMG.	Visão computacional.	Rede de páginas que distribuíram tais conteúdos.
Interface de aplicativos da Meta	Captura de tela e relato pessoal.	Método passo a passo.	Compreensão das interfaces e aspectos que ampliam/promovem a desinformação.
Diário de campo	Relato de impressões sobre grupos públicos de política em aplicativos de mensageria instantânea. ⁹	Descrição dos casos até a saturação.	Compreensão das características da circulação da desinformação em <i>apps</i> de mensageria.

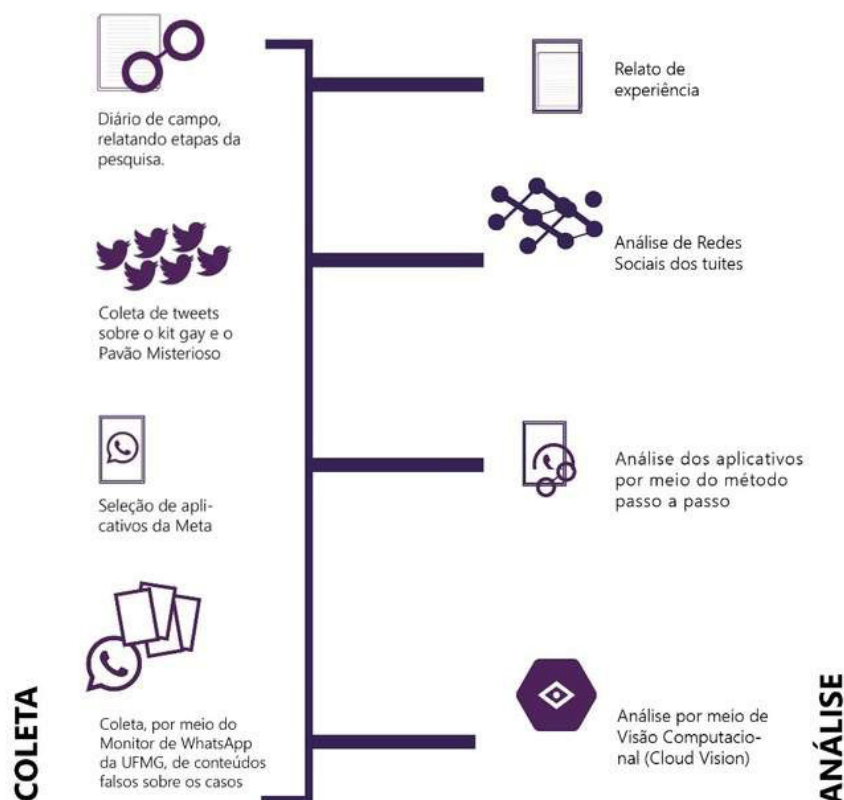
Fonte: Adaptado a partir de Oliveira (2023)

8 A coleta foi realizada por meio de raspagem dos termos “*kit gay*” e “#showdopavao”, com o uso do *twint*, pacote em Python hoje descontinuado que fazia buscas no Twitter a partir de determinados termos e, então, copiava as postagens relacionadas. Considerou-se somente os casos *kit gay* e Vaza Jato, pois tais controvérsias já haviam se estabilizado quando a análise foi desenvolvida. Deste modo, foi possível identificar as estratégias de viralização e os atores envolvidos quando tais assuntos movimentavam o debate público. No caso do *kit gay*, reuniu-se tuítes de dezembro de 2010 a junho de 2020, incluindo mais de 289 mil tuítes. No artigo, contudo, destaca-se apenas a distribuição em 2010, o que envolveu cerca de 3 mil tuítes. Já o #showdopavao, *hashtag* que acompanhou as primeiras denúncias do Pavão Misterioso, retornou cerca de 16 mil tuítes em 16 de junho de 2019. Enfoca-se nas contas verificadas, o que reduziu para 2 mil tuítes. O elevado volume de dados – e a disparidade do número de tuítes em cada caso – demandou a seleção de eventos (recortes), que permitiram a adequada análise.

9 Foram mais de dez grupos de Telegram e WhatsApp, acompanhados desde 2019. Como descrito em Oliveira (2023, p. 249-249), “A simples análise de mensagens publicadas no grupo Exército Bolsonaro, do Telegram, horas após a saída de Sérgio Moro do governo, reuniu mais de 40 mil postagens”. A análise, então, pauta-se na descrição de eventos específicos, das interfaces e em anotações e impressões dos pesquisadores.

A Imagem 1 apresenta o fluxograma da pesquisa:

Imagem 1 – Fluxograma dos procedimentos metodológicos



Fonte: Produzido pelos autores

A descrição das mediações é feita a partir de diferentes instrumentos, detalhados a seguir: para compreender a interface das plataformas, adotou-se o método passo a passo. Adotou-se o método passo a passo para analisar a materialidade das plataformas (Light; Burgess; Duguay, 2017) a partir da descrição das interfaces. A seleção de eventos e sua descrição, juntamente a análise dos efeitos dos algoritmos (Bucher, 2018), por sua vez, permitiu compreender a intervenção algorítmica das plataformas no fenômeno. A seleção de eventos destacou ocasiões em que a materialidade das plataformas se mostrava evidente (quando a viralização de um conteúdo era acompanhada de uma *hashtag*, como na Vaza Jato).

Por sua vez, o diário de campo relata, em notas livres, incursões em grupos de debate político no WhatsApp e no Telegram. Buscou identificar percepções dos pesquisadores sobre os aplicativos, os grupos analisados e sua interface. É uma descrição qualitativa, que foi posteriormente apresentada até sua saturação. A entrada nos grupos públicos, que precedeu a descrição, envolveu as seguintes etapas: a) pesquisa on-line por *links* de convite de grupos públicos; b) tentativas de entrada nos grupos; e c) extração dos dados (Resende *et al.*, 2019). Foram utilizados buscadores on-line, a partir de termos relacionados aos casos analisados. A partir de tais procedimentos, foi possível acompanhar mais de dez grupos – para garantir o anonimato dos autores de postagens ali presentes, não é possível indicar o nome de tais grupos. Além disso, a análise dos tuítes, por técnicas de análises de redes sociais¹⁰ (Recuero, 2020), compõe a etapa de descrição das mediações. Por fim, destaca-se que uma descrição ampla da metodologia e de

10 Considerou a análise de métricas locais e globais das redes de menções nos tuítes. As tabelas produzidas pelo *twint* não registravam retuítes (opção que nem existia em 2010). Ademais, considerando que se objetivava identificar como a desinformação sobre os casos circulou na plataforma, optou-se por identificar que usuários mencionam outros.

parte do diário de campo, além das limitações da pesquisa, é apresentada por Oliveira (2023).

Limitações da pesquisa

Este artigo também é resultado da rede interfaces-algoritmos, pois a coleta e a análise de dados passaram pela intervenção das plataformas. Para se acompanhar os grupos on-line, fez-se a pesquisa por convites por meio do Google. A partir da busca on-line por convites públicos, pode-se adentrar grupos e analisar seu conteúdo (Resende *et al.*, 2019). Por sua vez, a coleta dos tuítes também é restringida pela formatação promovida pela interface do Twitter e pelo pacote *twint*. Observa-se, ainda, que o presente texto não seria capaz de aferir todas as mensagens trocadas nos grupos on-line, de forma que se optou pela seleção e descrição de eventos específicos (Bucher, 2018). Tais estratégias permitem que o relato não se perca na miríade de postagens, enquanto garantem a que a rede descrita permita compreender a produção, a circulação e o consumo de conteúdos falsos – e, claramente, não abraça todos os atores, mas dá um panorama que explicita sua agência.

Além disso, aponta-se o desafio de separar a ação das interfaces e seus algoritmos. Em aplicativos de mensageria instantânea é uma tarefa mais fácil, já que não há sistemas de ranqueamento. No entanto, quando se considera a *pipeline* YouTube – WhatsApp, por exemplo, observa-se que se trata de um ecossistema de desinformação no qual as interfaces e os efeitos dos algoritmos podem ser descritos, mas nem sempre separados.

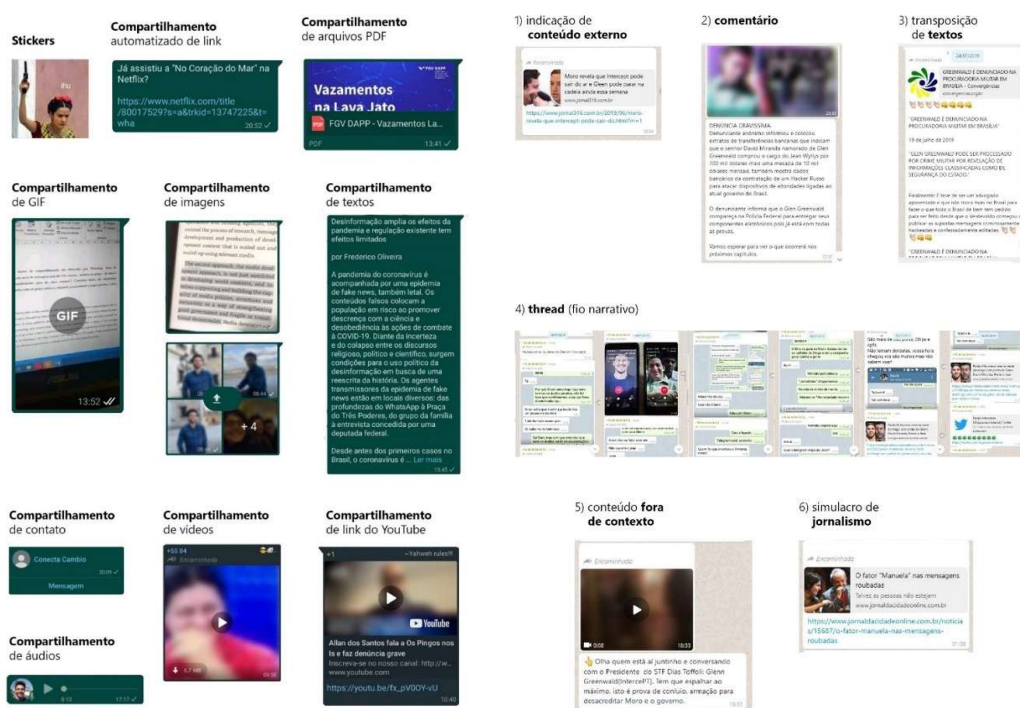
A pesquisa enfoca a descrição das interfaces por considerar sua relevância na formatação de conteúdos falsos: são elas que definem “gêneros e formatos” de desinformação. São suas limitações que estabelecem o modo como conteúdos são moderados. Tome-se, por exemplo, os *stories* do Instagram, que são moderados após a publicação: tal estratégia de moderação torna possível postar nudez desde que se coloque um *emoji* tapando parte dos genitais e mamilos ou esses não estejam totalmente enquadrados; ou inserir conteúdos falsos sobre vacinas substituindo o termo por “v4c1n4”. Embora o artigo enfoque as interfaces, compreende-se sua relação com os algoritmos das plataformas, de modo que, em alguns momentos, é preciso reconhecer sua mediação. Tendo em vista as limitações da pesquisa, parte-se para a descrição dos resultados e sua análise (reagregação).

Interfaces e tipos específicos de desinformação

As interfaces de uma plataforma determinam que tipo de conteúdo será bem distribuído, a aparência e o teor dessas mensagens. “O Facebook e suas aplicações produzem e formatam os conteúdos falsos” (Lemos; Oliveira, 2021, p. 77), seja pelo modo como os sistemas processam imagens neles publicadas ou pelo que as interfaces permitem. Como o WhatsApp apaga os metadados de imagens, por exemplo, os autores de conteúdos difamatórios se beneficiam de um certo “anonimato”. Além disso, o arquivo sempre é “novo”, visto que a data de criação é aquela em que foi enviado pelo servidor ao “destinatário”. Por sua vez, uma postagem no Facebook ou no Instagram com determinadas expressões – “vacina”, “eleições” e “guerra”, por exemplo – pode ser acompanhada de uma etiqueta, inserida pela plataforma, ou ter a distribuição reduzida.

A análise da interface do WhatsApp demonstra que o aplicativo intervém na apresentação dos conteúdos que nele circulam, como explicita a Imagem 2. *Links* externos são apresentados em uma miniatura, que indica apenas o título da página (*meta title tag*) e parte da descrição. Considere-se que a redação digital jornalística e publicitária é definida por estratégias de otimização para buscas (SEO). Um usuário que não tem um plano de dados que permite acessar páginas fora do aplicativo, por exemplo, precisa se contentar com tal miniatura – que pode conter um título exagerado, distorcido ou sensacionalista (*clickbait*). O aplicativo também formata outros conteúdos, o que resulta em formas específicas de desinformação que nele circulam: o WhatsApp formata os conteúdos falsos.

Imagem 2 – Interface do WhatsApp formata apresentação de conteúdos na conversa



Fonte: adaptado de Oliveira (2020).

Observa-se que a materialidade do aplicativo definia como conteúdos sobre o caso eram apresentados – conteúdos eram manipulados e retirados do seu contexto; a apresentação de miniaturas quando se compartilha URLs no mensageiro criava *clickbaits*; assim como rumores eram espalhados rapidamente (nos grupos, as pessoas se conhecem). No caso da Vaza Jato, as mensagens em grupos eram distribuídas em cascatas de informação (Recuero; Grudz, 2019), na forma de correntes e *threads* – emulando o que acontecia na conta do Pavão Misterioso no Twitter. Considerando, isso, propôs-se em pesquisa anterior sobre a Vaza Jato (Oliveira, 2020) uma tipologia¹¹ de desinformação no WhatsApp. Com as atualizações na interface, novas formas de apresentar conteúdos surgem.

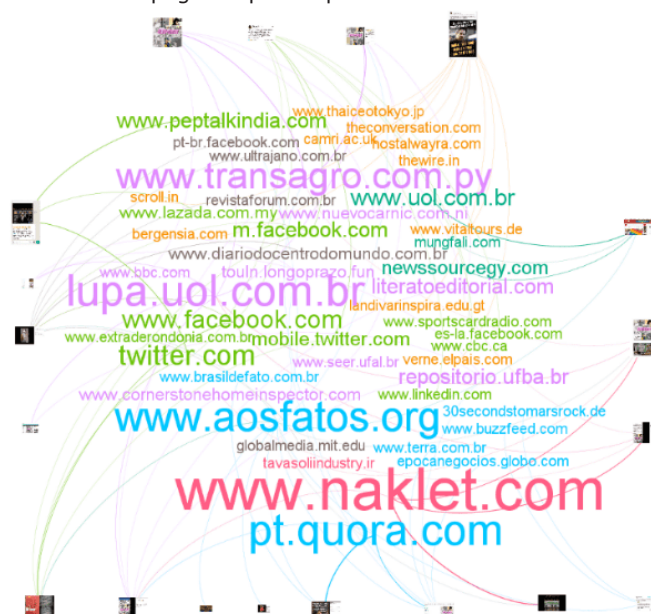
Faz-se necessário, então, analisar as peças de desinformação no contexto em que circulam, ou

11 A proposta aponta: "1) *Indicação de conteúdo externo*: são postagens que apenas indicam URLs para conteúdos que não estão hospedados no WhatsApp. Geralmente não acompanham nenhum texto, mas podem reproduzir o primeiro parágrafo de uma matéria ou apontar pequenos comentários sobre a temática. Caracteriza esse tipo de conteúdo seu caráter sumário: busca-se apenas indicar uma fonte externa, sem que se faça análises a seu respeito. 2) *Comentários*: análises ou denúncias acerca de um tema específico, que podem ou não ser acompanhados de conteúdos que os fundamente. Quando se utiliza imagens, vídeos ou links para conteúdo externo, tais mensagens são mais perigosas, já que parecem estar fundamentadas em fatos. 3) *Transposição de textos*: textos escritos em uma ordem inspirada na técnica de redação jornalística (pirâmide invertida, com uso de lead) ou análises postadas em sites supostamente noticiosos são transpostos para o WhatsApp. Importa destacar que algumas marcas nas postagens identificam a fonte original da informação. 4) *Fio narrativo* (thread): caracterizam as threads o uso de conteúdos de diversas fontes, em sequência, com o objetivo de contextualizar o fato e, nos casos analisados, caluniar alguém. Muito comum no Twitter, tal formato apresenta uma sequência de mensagens sobre uma mesma temática. 5) *Conteúdo fora do contexto*: imagens, vídeos, textos e links também foram retirados de seu contexto original para fortalecer uma narrativa específica. Em si, necessariamente, não trazem mensagens falsas, mas, ao circularem juntamente com outros conteúdos falsos, acabam fortalecendo uma narrativa. 6) *Simulacro de jornalismo*: conteúdos que a literatura internacional taxa como fake news. Produzidos na estrutura do texto jornalístico e simulando o método de apuração do jornalismo, tais postagens podem simular a diagramação de portais de imprensa tradicionais, ou serem produzidas por blogs específicos. No caso dos grupos, em que há uma grande desconfiança com a imprensa, chamada de extrema imprensa, circulam diversos conteúdos de páginas como Jornal da Cidade Online, Terça Livre, entre outros" (Oliveira, 2020, p. 8, grifo no original).

seja, na plataforma em que são apresentadas – e dentro das intervenções feitas por aquela interface específica. Quando alguém faz a captura de tela de uma postagem do Twitter e compartilha tal conteúdo no WhatsApp, promove uma espécie de remediação (Bolter; Grusin, 2000): a postagem, que já havia sido formatada em 240 caracteres, é reformatada pela interface do mensageiro. Tal prática é muito comum em grupos de política, tanto no WhatsApp quanto no Telegram, sendo percebida em todos os casos analisados.

Enquanto as interfaces restringem / formatam um conteúdo, os algoritmos garantem sua credibilidade e distribuição. Considere-se a Imagem 3, que aponta *sites* que hospedam desinformação sobre o *kit gay*¹² checados por Aos Fatos e Lupa: dentre eles, o Repositório Institucional da UFBA, especificamente em uma tese de doutorado. Supõe-se que tal biblioteca digital esteja bem ranqueada pelo *crawler* do buscador do Google, que prioriza a autoridade, a relevância, a confiança e a experiência do usuário¹³ em um site (Google, 2022), o que implica em maior distribuição desses conteúdos falsos (ainda que dentro de uma pesquisa em que são refutados). Ou seja, o algoritmo que ranqueia os resultados pode ampliar a distribuição de um conteúdo falso caso esteja hospedado em uma URL confiável.

Imagem 3 – Rede de páginas que hospedam conteúdos falsos sobre o *kit gay*



Fonte: Produzido pelos autores

Interfaces e estratégias de viralização

As restrições determinadas pela interface implicam diretamente nas estratégias adotadas para ampliar a circulação de uma postagem específica. Um dos perfis usados pelo *Pavão Misterioso*, conta do Twitter que questionava as revelações da série de reportagens do The Intercept Brasil sobre a Lava Jato, já foi uma *fan account* de Justin Bieber. Quando ganhou muitos seguidores, mudou de nome e passou a disseminar conteúdos falsos (Oliveira, 2019). Trata-se de prática comum nas plataformas – e somente a Meta possui recursos para avisar o usuário que determinada conta já teve outros nomes, embora não indique quais foram os *usernames*.

Ainda neste caso, vale destacar que as denúncias do *Pavão Misterioso* foram acompanhadas de *hashtags* específicas, que rapidamente chegavam ao *trending topics*. A ação de contas automatizadas,

12 Ressalta-se que a rede de páginas foi produzida a partir da visão computacional, analisando imagens falsas sobre o *kit gay* que circularam no WhatsApp.

13 *Clickbait*s são redigidos buscando a otimização para motores de busca, sendo sua redação definida pelo que interfaces (limite de caracteres apresentados na interface de busca do Google, por exemplo) e o algoritmo de ranqueamento fazem-fazer.

o interesse dos usuários em conteúdos conspiracionistas e a própria agência da plataforma explicam tal processo. Quando se analisa os tuítes do *kit gay*, diferentes usos de *hashtag* são percebidos – o que também permite analisar a interface da plataforma.

A disseminação do *kit gay* em 2010 demonstra, por exemplo, como cascatas de informação também são possíveis em função da interface das plataformas. Em 15 de dezembro de 2010 – quando o *feed* do Twitter era cronológico – as contas @cotidianogospel, @cristaos_brasil, @crivellaemfoco, @direitacarioca, @enfocuerio, @evangelicos_rj, @foco_cristao, @holofotenet, @irmaosonline, @pauloteixeirarj e @transpar_crista publicaram 319 tuítes sobre o *kit gay*. Naquele dia, haviam sido publicadas 354 postagens públicas sobre o tema. Dentre os 319 tuítes, 66 eram versões de um mesmo texto, como demonstra a Imagem 5. Cada uma das contas tuitou seis vezes a frase em um intervalo de menos de dois minutos – isso por meio do *Splitweet*, ferramenta que permitia a postagem em múltiplas contas. Sem a API da plataforma, tal modelo de disseminação não seria possível. Em um Twitter com *feed* cronológico, quanto mais menções a uma *hashtag*, maior a probabilidade de essa ser elencada nos *trending topics* (ver Imagem 4).

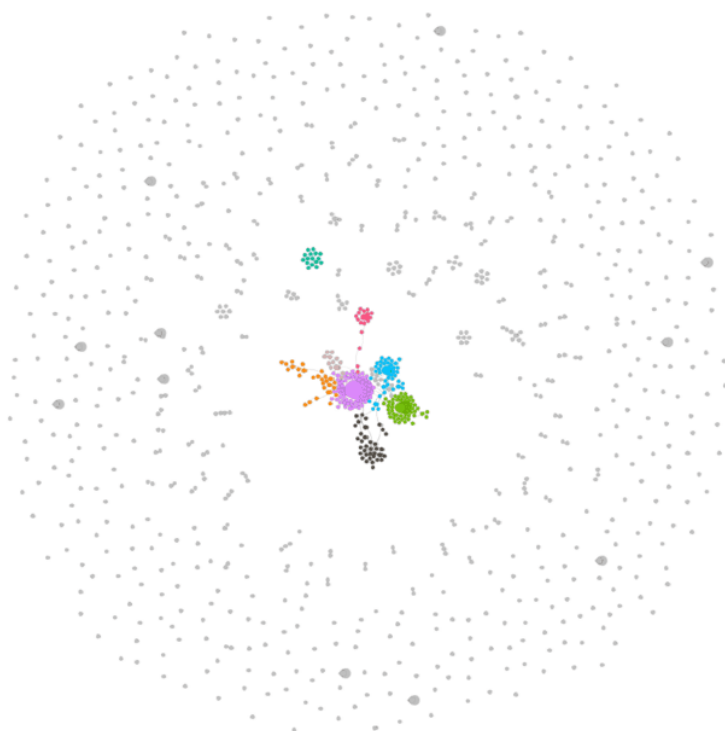
Imagem 4 – Variações do mesmo tuíte buscam ampliar alcance



Fonte: Produzido pelos autores

A análise da rede do *kit gay* no Twitter também aponta outra estratégia adotada com o objetivo de ampliar a disseminação do tema: mencionar o maior número de contas possível, *floodar a timeline*. A Imagem 5 apresenta a rede de tuítes sobre o caso em dezembro de 2010 – quando ainda não havia sido abraçado pela bancada evangélica, sendo uma cruzada de Bolsonaro. Ainda há poucos *clusters* – tratava-se de fato recente –, sendo que diversas contas postam sobre o tema, mas interagem pouco umas com as outras. Na imagem, cada círculo é um nó (uma conta) e seu diâmetro é diretamente proporcional à quantidade de tuítes sobre o *kit gay* de sua autoria. No centro, em rosa, a conta @cOnversadOr, que mencionou mais de 70 perfis de personalidades evangélicas – sendo que o cantor Adhemar de Campos foi o primeiro a engajar na desinformação sobre o *kit gay*. Na imagem, ainda é possível observar círculos na cor cinza, que embora tenham postado, não participam da rede, contatam um menor número de contas.

Imagem 5 – Tuítes que mencionaram o termo kit gay em dezembro de 2010

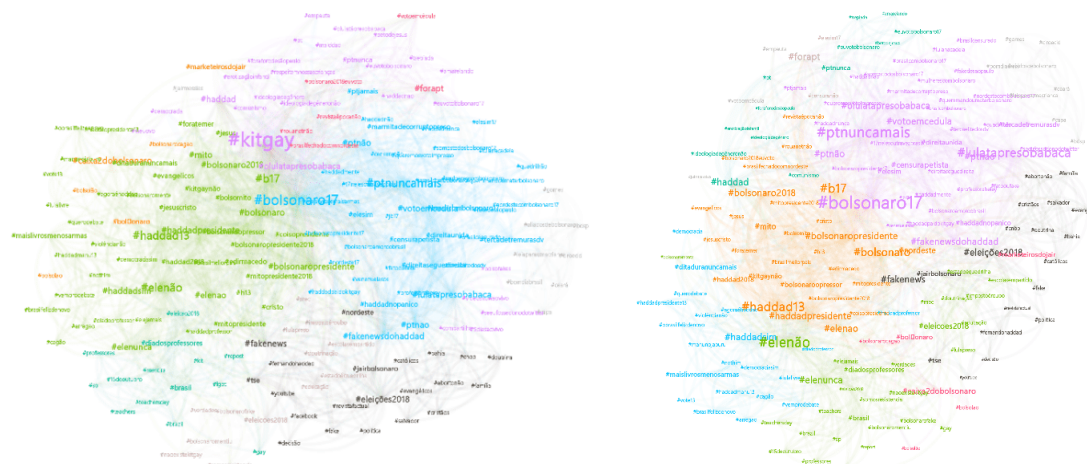


Fonte: Produzido pelos autores

Já em 2018, o uso de *hashtags* parece se relacionar mais a comunidades de sentido, agrupando diferentes posicionamentos políticos. Não deixa, contudo, de ser uma estratégia para viralização. A Imagem 6 apresenta a prática de *co-tagging* – a associação de diversas *hashtags* a um mesmo tuite – nas postagens de 16 de outubro de 2018 (pico de postagens sobre o *kit gay* naquele ano eleitoral). Além de ser uma estratégia de viralização, o *co-tagging* também vincula comunidades semânticas e produz a rede estruturalmente. Mas, para demonstrar isso, é necessária certa intervenção durante a análise.

A *hashtag* mais utilizada naquele dia foi a *#kitgay*. Para observar os *clusters* com mais clareza, optou-se por plotar uma versão da rede sem esse marcador. Isso, contudo, altera a modularidade, já que modifica o conjunto de nós e arestas – afetando, portanto, toda a rede. Optou-se por apresentar as duas versões da rede, que demonstram como uma *hashtag* organiza comunidades de sentido estruturalmente, alterando sua modularidade (cores semelhantes reúnem agrupamentos de mesma modularidade).

Imagem 6 – Hashtags apontam comunidades que debatiam o kit gay



Fonte: Produzido pelos autores

Ambos os exemplos ilustram como configurações específicas das interfaces e algoritmos das plataformas determinam a adoção de diferentes estratégias para tentar viralizar um conteúdo. No caso de desinformação, marcadas por sua distribuição em cascatas, postar diversas vezes e pelo maior número de contas possível é fundamental. As restrições de cada plataforma determinam como tal objetivo será alcançado, seja pelo copiar-colar no WhatsApp, escrevendo *v4c1n4* no Instagram, para evitar a identificação da palavra, ou outra estratégia. Campanhas de desinformação sobre a pandemia, por exemplo, passaram pela lógica de monetização das plataformas, adotando ainda estratégias como *firehosing* e *astroturfing* (Brasil. Senado Federal, 2021a, Cesarino, 2022).

Nos autos do Inquérito 4828/DF do Supremo Tribunal Federal (Brasil, 2020), que investigou os atos antidemocráticos, há diversas indicações do papel de políticas de monetização de conteúdo, *crowdfunding* e financiamento da desinformação. Aponta-se, por exemplo, que *live* do canal *Folha Política* em 3 de maio de 2020, em que Bolsonaro foi entrevistado, garantiu remuneração entre US\$ 6 e 11 mil aos donos. Uma *live* presidencial no Dia do Exército, retransmitida pelo canal *Foco do Brasil*, garantiu entre US\$ 7,5 e 18,8 mil. Mais uma vez, a rede interface e algoritmo define a distribuição de conteúdos falsos: sistemas de *crowdfunding* e a própria política de monetização envolvem limitações definidas pelas próprias interfaces, assim como a lógica algorítmica que define a distribuição e o pagamento.

Chama a atenção também o papel do WhatsApp como espaço de disseminação de conteúdos presentes em outras plataformas: algo que não ocorria quando o *kit gay* surgiu em 2010, mas se manifestou posteriormente no caso. Vídeos desinformativos do YouTube são encaminhados para o *app* e ali podem ser assistidos em miniatura e reenviados, por exemplo (Melo *et al.*, 2020). Os relatórios do Monitor de WhatsApp da UFMG, desenvolvidos nas eleições de 2018 e 2022, que indicam os vinte *links* mais compartilhados no *app*, corroboram tal dinâmica (Imagem 7). Vale ressaltar a ausência de *links* para o TikTok e para o Kwaii, o que é explicável, já que tais aplicativos permitem baixar os vídeos e compartilhá-los no WhatsApp.

Imagem 7 – Links mais compartilhados no WhatsApp nas eleições de 2018 (esquerda) e de 2022 (direita)

Domínio	# Total	# Distintas	# Usuários	Domínio	# Total	# Distintas	# Usuários
whatsapp.com	55468	372	850	google.com	977	151	8
gazetabrasil.com.br	35031	10444	813	youtube.com	662	160	174
bit.ly	23473	7708	1344	gazetabrasil.com.br	361	111	5
youtu.be	20952	6003	2917	financas-web.club	334	3	260
dunapress.org	19115	1505	317	terrabrasilnoticias.com	314	28	185
terrabrasilnoticias.com	17896	3029	1681	facebook.com	248	81	62
facebook.com	13206	5587	2508	whatsapp.com	240	38	79
t.me	12706	453	1066	instagram.com	219	101	71
uol.com.br	12193	6346	2798	t.me	214	15	64
globo.com	12077	6124	3248	novaiguacu24h.com.br	158	2	28
instagram.com	9040	2696	1506	pensandodireita.com	137	3	10
twitter.com	8298	3053	1455	globo.com	135	45	57
jornaldacidadeonline.com.br	6803	1116	1839	bit.ly	124	16	8
portalnovoeste.com.br	6453	587	1168	twitter.com	96	27	29
timesbrasil.com.br	5350	1498	43	uol.com.br	92	44	29
blogspot.com	5069	2594	640	portalcidade.news	85	10	16
youtube.com	4833	1513	1126	creditoparacarta.com	83	1	74
conexaopolitica.com.br	4792	1524	1175	noticiasdebrasil.com	71	3	1
ultimasnoticiasdobrasil.com	4662	490	60	brasil247.com	59	35	23
t.co	4310	2097	799	apostagem.com.br	57	8	29

Fonte: Melo e colegas (2020b, 2022).

Interfaces e refutação de fake news

A pandemia foi marcada por ações de moderação de conteúdo e refutação de *fake news* por parte das plataformas (nem sempre eficientes). Seja pela inserção de etiquetas em conteúdos falsos a partir de projetos de *fact-checking* ou pela identificação automatizada de conteúdos, a intervenção das plataformas na moderação e refutação da desinformação se fez presente. Naquele contexto, parece ter-se optado por apagar conteúdos inadequados, em vez de simplesmente reduzir sua visibilidade. Diante deste cenário, “influenciadores passaram a hospedar seus vídeos mais explícitos no Telegram e em sites alternativos como o Rumble e a Covidflix, cuja estética “estranho-familiar” mimetizava plataformas como YouTube e Netflix” (Cesarino, 2022, p. 264). Ademais, outras estratégias foram adotadas por quem desejava espalhar desinformação.

Vale lembrar que são as plataformas que definem o conteúdo a ser refutado (Lemos; Oliveira, 2021) e também limitam as condições dessa refutação. Além da intervenção do Programa de Verificação de Fatos Independente da Meta (3PFC) na noticiabilidade dos checadores, já descrita em Oliveira (2021), ressalta-se que as ferramentas (formulários – interface – para usuários indicarem conteúdos para serem checados e sistemas de recomendação – algoritmos – para os checadores) oferecidas pelas plataformas possibilitam um tipo de checagem específica.

A partir do *Google Vision*, foram extraídas etiquetas e entidades (Imagem 8) relacionadas a conteúdos falsos sobre o *kit gay* encontrados nos grupos e checados por Aos Fatos e Lupa. Como destacam Omena e colegas (2021), etiquetas são descrições textuais que a inteligência artificial produz, considerando um grau de confiança – qual a chance de a imagem conter aquele elemento –, sendo pouco específicas. Já as entidades são textos recuperados pelo Google Imagens que estão associados à figura: ou seja, seu contexto é definido por autores humanos e compilado pelo sistema.

Percebe-se que a análise de etiquetas permite observar elementos que compõem a desinformação (formatos mais comuns); enquanto a de entidades deixa analisar as temáticas, atores políticos envolvidos e outras características. Os algoritmos das plataformas, operando conjuntamente às interfaces, são sistemas de visibilidade e por consequência, de invisibilidade, que escolhem o que ocultam ou reconhecem – além de reduzir diversos *inputs* em um único *output*, o mais “provável” estatisticamente (Bucher, 2018, Gillespie, 2018, Amore, 2020, Crawford, 2021). A visão computacional se apresenta como uma estratégia útil para identificar e descrever o que há em comum em uma grande quantidade de imagens falsas, mas seus resultados limitam-se ao que é “provável”.

Sistemas próprios de monitoramento, como o *Radar Aos Fatos*, por exemplo, não estão livres da dependência das políticas das APIs. Essa ferramenta usava o Crowdtangle – descontinuado pela Meta –, fazia coleta de tuítes – hoje impossibilitada pela nova política do X – e acessa postagens no WhatsApp. Assim, as estratégias existentes para verificar a desinformação dependem também das próprias plataformas – o que apresenta riscos relacionados à independência dos checadores. Passam, muitas vezes, pela adoção de sistemas de raspagem que não são permitidos pelos termos de uso dessas plataformas, o que também pode resultar em bloqueios e outras sanções – além de questões éticas e legais sobre a coleta de dados e a privacidade em *apps* de mensageria instantânea, por exemplo.

Entender a desinformação na contemporaneidade demanda o reconhecimento de como as materialidades das plataformas agenciam a sua disseminação. A análise das interfaces e do processamento algorítmico dos dados é fundamental para entender as plataformas como agentes importantes na produção do fenômeno e não apenas como veículo para pessoas mal-intencionadas. Esse reconhecimento é inclusive um argumento importante para se pensar a necessidade de regulamentação das plataformas.

Dentre as limitações da metodologia, está a estratégia de coleta de dados, refém das limitações das APIs (D'Andréa, 2021). Vale apontar, contudo, que é justamente o interesse da pesquisa investigar como tais interfaces produzem desinformação – e, também, a pesquisa científica sobre o tema. Ainda, pode-se apontar que o número de grupos do WhatsApp que o pesquisador acompanhou não é representativo, mas ressalta-se que a análise aqui desenvolvida enfoca como a interface intervêm nas conversações agenciando os envolvidos (usuários, softwares, interfaces, IA, decisões políticas) em uma intra-ação que deve ser politizada. Pode-se questionar também o recorte dos dados apresentados, que não reproduz os dados coletados. Foram mais de 100 mil tuítes sobre o *kit gay*, por exemplo. Sendo assim, adotou-se a estratégia de recorte por eventos, proposta por Bucher (2018), que permite realizar análises mais eficientes e efetivas.

19

que ressalta a importância de compreender a desinformação em seu contexto de distribuição. Também ressaltam as diversas intervenções das interfaces, que promovem formas distintas de interação do usuário (Van Djick; Poell, De Wall, 2018). As plataformas definem as estratégias adotadas para que um conteúdo se torne viral. A existência de um *feed* cronológico no Twitter, por exemplo, fez com que diversas versões da *#kitgay* fossem adotadas para alçar o tema aos *trending topics*.

Observa-se, na investigação, que a rede interface-algoritmo formata os conteúdos falsos, dá as condições de circulação, participa da disseminação e pode garantir credibilidade. Enquanto as interfaces formatam e restringem condições (definem o tipo de conteúdo falso, como é apresentado), os algoritmos das plataformas cancelam tal conteúdo (que é avaliado por sistemas de moderação, pode ser compartilhado para mais pessoas e, ainda, pode ser entendido como conteúdo de qualidade). Essa amálgama, essa rede interface-algoritmo, essa intra-ação (Barad, 2009) demonstra que desinformação não é uma erro do sistema, mas sua forma de gerar monetização e engajamento, própria da lógica das plataformas (Lemos, 2024).

A pesquisa ainda se debruçou sobre as intervenções das plataformas na refutação de conteúdos falsos, assim como o papel de ferramentas de monetização e *crowdfunding*. Os pontos aqui discutidos contribuem para a compreensão do fenômeno, destacando a relevância das plataformas. Analisar as interfaces é sua construção é necessário para entender como conteúdos falsos são um fenômeno da PDPA, um reflexo das plataformas e seu *design*. Pesquisas futuras podem apropriar-se da metodologia e das propostas aqui descritas para identificar como diferentes interfaces criam distintas *fake news*, como a circulação de conteúdos falsos é potencializada em cada uma das plataformas e que ações resultam no efetivo combate à desinformação.

Referências

- AMOORE, Louise. **Cloud ethics**: algorithms and the attributes of Ourselves and Others. Durham; London: Duke University Press, 2020.
- ARUN, Chinmayi. On WhatsApp, rumors and lynchings. **Economic & Political Weekly**, [S.l.], v. LIV, n. 6, p. 30-35, 9 fev. 2019.
- BAKIR, Vian; MCSTAY, Andrew. Fake news and the economy of emotions. **Digital Journalism**, [S.l.], v. 6, n. 2, p. 154-175, 2017.
- BACARELLA, Christian *et al.* Social media? It's serious! Understanding the dark side of social media. **European Management Journal**, [S.l.], v. 36, n. 4, p. 431-438, ago. 2018.
- BRASIL. Supremo Tribunal Federal. Inquérito. **Inq 4828/DF**. Relator: Ministro Alexandre de Moraes. Autor: Ministério Público Federal; sob sigilo. Procuradores: Procuradoria-Geral da República; sob sigilo. 2020-atual. Disponível em: <<https://portal.stf.jus.br/processos/detalhe.asp?incidente=5895367>>. Acesso em 23 jan. 2023.
- BOLTER, Jay David; GRUSIN, Richard. **Remediation**: understanding new media. Cambridge: The MIT Press, 2000.
- BOUNEGRU, Liliana; CHAMBERS, Lucy; GRAY, Jonathan. **The data journalism handbook**. [S.l.: s.n.], 201?. Disponível em: <<https://datajournalism.com/read/handbook/one>>. Acesso em 25 jan. 2020.
- BOVET, A.; MAKSE, H. A. Influence of fake news in Twitter during the 2016 US presidential election. **Nature Communications**, v. 10, 2019.
- BUCHER, Tania. **If... then**: algorithmic power and politics. New York: Oxford University Press, 2018.
- D'ANDRÉA, Carlos. Para além dos dados coletados: políticas das APIs nas plataformas de mídias digitais.

MATRIZES, São Paulo, Brasil, v. 15, n. 1, p. 103–122, 2021

CALVO, Ernesto; ARUGUETE, Natalia. **Fake news, trolls y otros encantos: cómo funcionan (para bien y para mal) las redes sociales**. [S.l]: Siglo XXI Editores, 2020.

CANAVILHAS, João; COLUSSI, Juliana; MOURA, Zita-Bacelar. Desinformación em las elecciones presidenciales 2018 em Brasil: un análisis de los grupos familiares en WhatsApp. **El profesional de la información**, [S.l], v. 28, n. 5, 2019.

CESARINO, Letícia. **O mundo do avesso: verdade e política na era digital**. São Paulo: Ubu, 2022.

DOURADO, Tatiana Maria Silva Galvão. **Fake news na eleição presidencial de 2018 no Brasil**. 2020. 308f. Tese Tese (Doutorado em Comunicação e Cultura Contemporâneas) – Universidade Federal da Bahia, Salvador, 2020.

GILLESPIE, Tarleton. A relevância dos algoritmos. **Parágrafo**, São Paulo, v. 6, n. 1, p. 95-121, jan.-abr. 2018.

GILLESPIE, Tarleton. Do not recommend? Reduction as a form of content moderation. **Social Media + Society**, [S.l], p. 1-13, jul.-set. 2022

GOOGLE. **Search Quality Evaluator Guidelines**. 2022. Disponível em: <https://static.googleusercontent.com/media/guidelines.raterhub.com/pt-BR//searchqualityevaluatorguidelines.pdf>. Acesso em 27 jul. 2022.

JURNO, Amanda Chevtchouk. **Facebook e a plataformação do jornalismo: uma cartografia das disputas, parceiras e controvérsias entre 2014 e 2019**. 2020. 225f. Tese (Doutorado em Comunicação Social) – Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

LEE, Kate Sangwon; WEI, Huaxin. Design factors of ethics and responsibility in social media: a systematic review of literature and expert review of guiding principles. **Journal of Media Ethics**, [S.l], v. 37, n. 3, p. 156-178, 2022.

LEMONS, André. Plataformas, Dataficação e Performatividade Algorítmica (PDPA): desafios atuais da cibercultura. In: PRATA, Nair; PESSOA, Sônia Caldas (Orgs.). **Fluxos comunicacionais e crise da democracia**. São Paulo: Intercom, 2020a, p. 107-116

LEMONS, André. Epistemologia da comunicação, neomaterialismo e cultura digital. **Galáxia**, São Paulo, n. 43, p. 54-66, jan.-abr. 2020b.

LEMONS, ANDRÉ. Erros, falhas e perturbações digitais em alucinações das IA generativas: Tipologia, premissas e epistemologia da comunicação. **MatriZES**, [S.l], v. 18, p. 75-91, 2024.

LEMONS, André; BITENCOURT, Elias; SANTOS, João Guilherme Bastos dos. Fake news as fake politics: the digital materialities of YouTube misinformation videos about Brazilian oil spill catastrophe. **Media, Culture & Society**, [S.l], v. 43, n. 3, p. 886-905, 2020.

LEMONS, André; OLIVEIRA, Frederico. Fake news e cadeias de referência: a desinformação sobre Covid-19 e o projeto de verificação do Facebook. **Fronteiras: estudos midiáticos**, [S.l], v. 23, n. 2, p. 73-88, mai.-ago. 2021.

LIGHT, Ben; BURGESS, Jean; DUGUAY, Stefanie. The walkthrough method: an approach to the study of apps. **New Media & Society**, [S.l], v. 20, n. 3, p. 881-900, 2018.

MACHADO, Caio C. Vieira *et al.* **Ciência contaminada: analisando o contágio de desinformação sobre coronavírus via YouTube**. Salvador: LAUT; INCT.DD; CEPEDISA, 2020.

MELO, Philipe de Freitas *et al.* Can WhatsApp counter misinformation by limiting message forwarding? In: CHERIFI, Hofine *et al.* (Eds). **Complex networks and their applications VIII: Volume 1: Proceedings of the Eighth International Conference on Complex Networks and their application**. [S.l]: Springer, 2020a.

MELO, Phillipe *et al.* **Relatório Eleições Presidenciais de 2018**: relatório Monitor de WhatsApp. Monitor de WhatsApp da UFMG, 2020b. Disponível em: < http://150.164.214.48/monitor-de-whatsapp/data/reports/pdfs/2018_06_report_geral_eleicao_presidencial.pdf>. Acesso em 25 jan. 2023.

MELO, Phillipe *et al.* **Report WhatsApp TSE**: Eleições. Monitor de WhatsApp da UFMG, 2022. Disponível em: < http://150.164.214.48/monitor-de-whatsapp/data/reports/pdfs/whatsapp_report_tse_eleicoes.pdf>. Acesso em 25 jan. 2023.

META. Como funciona a verificação de fatos. Centro de Transparência. **Meta**, 7 de abril de 2025. Disponível em: <<https://transparency.meta.com/pt-br/features/how-fact-checking-works/>>. Acesso em 26 abr. 2025.

MUKHERJEE, Rahul. Mobile witnessing on WhatsApp: vigilante virality and the anatomy of mob lynching. **South Asian Popular Culture**, [S.l.], 2020. DOI: 10.1080/14746689.2020.1736810

O'CONNOR, Callin; WEATHERALL, James Owen. **The misinformation age**: how false beliefs spread. New Haven; London: Yale University, 2019.

OLIVEIRA, Frederico. **As fake news e a produção jornalística de referências**. 2023. 382f. Tese (Doutorado em Comunicação e Cultura Contemporâneas) – Universidade Federal da Bahia, Salvador, 2023.

OLIVEIRA, Frederico. Fake news no WhatsApp: relações entre interface e formatos de circulação de conteúdos falsos. In: ENCONTRO VIRTUAL DA ABCIBER, 1, 2020. **Anais [...]**, São Paulo: Abciber, 2020, p.1-16.

OLIVEIRA, Frederico Ramos. Os voos do Pavão Misterioso: circulação de fake news no WhatsApp. In: ENCONTRO NORTE-NORDESTE DA ABCIBER, 2., 2019, Aracaju. Redes educativas e os desafios atuais da Cibercultura. **Anais [...]**. Aracaju: Unit, 2019

OMENA, Janna Joceli *et al.* The potentials of Google Vision API-based networks to study natively digital images. **Diseña**, [S.l.], v. 19, p. 1-25, ago. 2021.

RECUERO, Raquel. **Introdução à análise de redes sociais online**. Salvador: Edufba, 2017. Disponível em: < <https://repositorio.ufba.br/bitstream/ri/24759/4/AnaliseDeRedesPDF.pdf>>. Acesso em 18 jan. 2023.

RECUERO, Raquel; GRUZD, Anatoliy. Cascatas de fake news: um estudo de caso do Twitter. **Galáxia**, São Paulo, n. 41, p. 31-47, mai.-ago. 2019.

RESENDE, Gustavo *et al.* Analyzing textual (mis)information shared in WhatsApp groups. In: ACM CONFERENCE ON WEB SCIENCE, 2019, 11, Boston. **Proceedings [...]**. Boston: WebSci, 2019a, v. 11.

SANTAELLA, Lucia. **A pós-verdade é verdadeira ou falsa?**. Barueri, SP: Estação das Letras e Cores, 2018.

SANTAELLA, Lucia. **De onde vem o poder da mentira?**. Barueri, SP: Estação das Letras, 2021b.

SHAFI, Shuaib; RAVIKUMAR, Madhavi. Dynamics of Fake News Dissemination: A Case Study in the Indian context. **Media Watch**, [S.l.], v. 9, n. 1, p. 131-140, 2018.

SUSTEIN, Cass R. **On rumors**: how falsehoods spread, why we believe them; what can be done. New York: Farrar, Straus and Giroux, 2009.

TANDOC, Edson C.; LIM, Zheng Wei; LING, Richard. Defining “Fake News”: a typology of scholarly definitions. **Digital Journalism**, [S.l.], v. 6, n. 2, p. 137-153, 2017.

TARDE, Gabriel de. **A opinião e as massas**. São Paulo: Martins Fontes, 2005.

TORRES, Russel; GERHART, Natalie; NEGAHBAN, Arash. Combating fake news: an investigation of information verification behaviors on social networking sites. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 51, 2018, Waikoloa Village. **Proceedings... HICSS**, 2018, v. 51.

VAN DIJCK, José; POELL, Thomas; DE WAAL, Martijn. **The platform society: public values in a connective world.** New Yourk: Oxford University Press, 2018.

VAN DIJCK, José. Seeing the forest for the threes: visualizing platforms and the governance. **New Media & Society**, [S.l.], v. 23, n. 9, p. 2801-2819, 2020.

VOSOGUI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. **Science**, [S.l.], v. 359, n. 6380, p. 1146-1151, 2018.

WHITTAKER, Jason. **Tech Giants, Artificial Intelligence, and the Future of Journalism.** New York: Routledge, 2019.

Frederico Oliveira é doutor em Comunicação e Cultura Contemporâneas pela UFBA. É tecnologista no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Neste artigo, contribuiu com a concepção do desenho da pesquisa; desenvolvimento da discussão teórica; interpretação dos dados; apoio na revisão de texto e redação do manuscrito.

André Lemos é doutor em sociologia pela Université René Descartes, Paris V, Sorbonne (1995). É professor titular da Faculdade de Comunicação da UFBA. Neste artigo, contribuiu com a concepção do desenho da pesquisa; desenvolvimento da discussão teórica; interpretação dos dados; apoio na revisão de texto; redação do manuscrito e revisão da versão em língua estrangeira.