



XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

GT-5 – Política e Economia da Informação

**BOTS E REDES SOCIAIS: UMA REVISÃO DE LITERATURA SOBRE PROPAGANDA
COMPUTACIONAL E SEU IMPACTO NA POLÍTICA**

***BOTS AND SOCIAL MEDIA: A LITERATURE REVIEW ON COMPUTATIONAL PROPAGANDA
AND ITS POLITICAL IMPACT***

Rose Marie Santini – Universidade Federal do Rio de Janeiro

Débora Salles e Giulia Tucci – Universidade Federal do Rio de Janeiro

Modalidade: Trabalho Completo

Resumo: Este artigo baseia-se em uma revisão sistemática da literatura para investigar o estado da arte da pesquisa sobre o uso de perfis falsos em redes sociais para manipular a opinião pública a respeito de questões políticas. O objetivo desta revisão é discutir como a propaganda computacional impacta a construção de consenso, autoridade, legitimidade e representatividade de políticos, partidos, grupos hegemônicos e minoritários nas redes sociais. No total, 369 artigos foram obtidos após busca em quatro diferentes bases de dados de indexação de literatura científica (IEEE, Web of Science, Scopus e EBSCO). Destes, 16 foram considerados relevantes para a questão de pesquisa. Para identificar tendências gerais e resultados específicos, os artigos pertinentes foram analisados descritiva e criticamente. Cinco abordagens principais foram observadas: 1) Políticas de uso e neutralidade das plataformas; 2) Definição, detecção e caracterização dos agentes de manipulação; 3) Atuação política dos agentes nas redes sociais; 4) Tentativas de previsão de resultado eleitoral através das redes sociais; e 5) Dinâmicas da rede e estratégias de propagação da informação. Na conclusão, os resultados da revisão de literatura são articulados em uma análise crítica sobre o estágio atual da pesquisa científica sobre o tema.

Palavras-Chave: Bots; Manipulação; Política; Redes sociais; Revisão sistemática da literatura.

Abstract: This article draws on a systematic literature review to investigate the state of the art of research on the use of fake social media profiles to manipulate public opinion on political issues. The purpose of this review is to discuss how computational propaganda impacts on constructing of consensus, authority, legitimacy and representativeness of politicians, parties, hegemonic and minority groups on social media. In total, 369 articles were obtained following a bibliographic search on four different scientific citation indexing databases (IEEE, Web of Science, Scopus and EBSCO). Of these, 16 were considered relevant to the research question. To identify general trends and specific outcomes, relevant articles were analyzed descriptively and critically. Five main approaches were observed: 1) Usage policies and platform neutrality; 2) Definition, detection and description of the manipulation agents; 3) Political role of agents on social networks; 4) Attempted electoral predictions via social networks; and 5) Network dynamics and strategies for disseminating information. In the conclusion, the results of the literature review are articulated in a critical analysis about the current stage of scientific research on the subject.

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Keywords: Bots; Manipulation; Politics; Social Media; Systematic literature review

1 INTRODUÇÃO

As discussões sobre estratégias, técnicas e graus de manipulação da opinião pública nas redes sociais passaram a ocupar espaço significativo na imprensa após os resultados do referendo do Brexit e da eleição presidencial norte-americana em 2016. Durante a campanha vitoriosa de Donald Trump, 20% das mensagens políticas no Twitter foram postadas por *bots*; no Facebook, perfis controlados por russos foram responsáveis pela compra de anúncios e pela postagem de conteúdo polêmico com opiniões polarizadas que viralizaram pela ação de *bots*. O Facebook estima que o conteúdo produzido por russos, incluindo posts e anúncios pagos, atingiu 126 milhões de americanos, ou seja, algo em torno de 40% da população (THE ECONOMIST, 2017).

De acordo com Woolley e Howard (2017), o conceito de propaganda computacional se refere ao uso de algoritmos, automação e curadoria humana com a intenção de disseminar informações enganosas nas redes sociais. Entretanto, a presença de boatos, desinformação, discurso de ódio e ruído nas redes sociais, assim como o uso de perfis falsos para ampliar a propagação de *fake news*, é anterior ao escândalo da eleição norte-americana. Bradshaw e Howard (2017) detectaram o uso de propaganda computacional para manipular a opinião pública em 28 países, entre 2010 e 2017. Dentre as diferentes estratégias empregadas, os autores destacam as tropas cibernéticas, que são equipes formadas por pessoas, *bots* ou mistas dedicadas a manipular a opinião pública nas redes sociais.

Bots sociais são um fenômeno que surgiu com as plataformas de redes sociais, apesar de robôs estarem presentes na Internet desde a década de 1990 (WOOLLEY, 2018). *Bots* sociais são algoritmos que podem produzir e/ou disseminar conteúdo de modo automático e interagir com humanos (FERRARA *et al.*, 2016), criando uma “opinião pública artificial” nas redes sociais. Estes algoritmos conseguem simular a comunicação humana, podendo reproduzir, em alguns casos, padrões temporais de produção e difusão de conteúdo e de expressão de sentimentos. Segundo Ferrara *et al.* (2016), *bots* sociais podem ser usados para distorcer debates políticos, manipular o mercado financeiro, roubar dados de usuários, espalhar *fake news* e gerar desinformação online.

Para Bradshaw e Howard (2017), a forma de organização das tropas cibernéticas vem evoluindo: antes envolvia a elaboração de estratégias e experimentos na Internet por grupos militares, hoje foca na tentativa direta de manipulação da opinião pública.

Bolsover e Howard (2017) argumentam que Estados têm se aproveitado do anonimato da Internet para camuflar a produção e disseminação de propaganda. Assim, empresas de comunicação estratégica são contratadas por organizações de diferentes tipos, principalmente governos, para a elaboração e execução de campanhas de manipulação em redes sociais.

Além disso, não há dados oficiais disponibilizados pelas empresas de redes sociais, que podem se beneficiar com a presença de *bots*, pela ampliação da quantidade de usuários em suas plataformas (DAYEN, 2017). É preciso considerar que a transparência em relação à presença de *bots* poderia comprometer a reputação e credibilidade destas plataformas diante de seus usuários e impactar seu faturamento e valor de mercado.

Neste cenário, cabe questionar: como identificar e rastrear os *bots*? Quais são as diferenças e semelhanças entre personagens falsos, sejam humanos ou perfis automatizados? Como operam estes diferentes agentes e a quem beneficiam? Quem desenvolve os *bots* e quem os comercializa? Quando e como geram conteúdo? Que fontes utilizam? Quais são seus públicos-alvo? Quais são seus efeitos na formação da opinião pública e nos processos políticos? Muitas destas questões ainda estão em aberto. Há estudos robustos sobre o tema, porém são escassos e dispersos. A automatização de perfis em redes sociais é um tema novo, cujo campo de pesquisa é caracterizado por problemas conceituais e metodológicos.

No nível conceitual, há diversas dificuldades na definição e na distinção dos agentes e das ações de manipulação: há perfis falsos que são *bots* se comportando como um exército de humanos, há outros que são pessoas se comportando como um exército de *bots* (*click farms*). Há ainda *trolls* e usuários fanáticos que produzem e difundem mensagens de forma automática, com o objetivo de viralizar o que publicam. *Bots* conseguem se infiltrar em redes de conexões reais, produzir conteúdo com credibilidade e adotar padrões temporais humanos, o que dificulta sua detecção. Podem agregar conteúdo automaticamente de várias fontes predeterminadas, postar material em tempo real, estabelecer conversação verossímil, comentar em posts e responder a perguntas. Além disso, é crescente a tendência de combinar ação humana, *big data* e automação para sofisticar possibilidades de aplicação (FERRARA *et al.*, 2016), o que aproxima o conceito de *bot* ao de *cyborg*.

Em relação aos aspectos metodológicos, são significativos os desafios de pesquisa, considerando que o uso de perfis automatizados pode ser uma atividade questionável, legal e moralmente. Fabricantes de *bots* o fazem de forma sigilosa, dificultando sua identificação e coleta de informações. Na academia, pesquisadores enfrentam dilemas éticos para conduzir experimentos obstrutivos. Outros métodos qualitativos de investigação, como entrevistas, observação participante e grupos focais, enfrentam obstáculos: operadores de *bots*, empresas que os comercializam e contratantes destes serviços não desejam ser associados à atividades suspeitas. No que tange a identificação automática de *bots*, a aplicação de métodos quantitativos é insuficiente, especialmente para *cyborgs*. Os sistemas de identificação baseados na análise de grafos costumam fracassar, porque os *bots* se conectam facilmente com pessoas reais para se camuflar. A identificação humana costuma ser precisa e eficaz (FERRARA *et al.*, 2016), porém é um método inviável econômica e temporalmente, diante da quantidade de dados de usuários de redes sociais.

O objetivo deste trabalho é mapear as pesquisas existentes sobre o uso de perfis nas redes sociais para manipulação da opinião pública na discussão de assuntos políticos. Para isso, foi realizada uma Revisão Sistemática da Literatura (RSL), com a finalidade de apresentar um diagnóstico sobre as pesquisas científicas neste campo emergente. Diante dos trabalhos pertinentes recuperados da literatura, indicamos os tipos de casos estudados, as abordagens e os métodos utilizados e as conclusões práticas e teóricas alcançadas até o momento da condução desta revisão.

O panorama sistematizado de trabalhos produzidos sobre um fenômeno possibilita o maior entendimento de suas dimensões geopolíticas. A realização de uma RSL contribui para a área, pois são evitadas generalizações específicas (DACOMBE, 2018), e há redução de conclusões enviesadas, não permitindo que a perspectiva do autor direcione o leitor, já que previnem a impressão da perspectiva do autor, diferente das revisões de literatura tradicionais (PETTICREW; ROBERTS, 2006). É importante enfatizar que revisões de literatura tradicionais podem ter viés e seguir a narrativa criada pela perspectiva do autor (PETTICREW; ROBERTS, 2006). No atual contexto sócio-político, se torna importante compreender de forma panorâmica os impactos da propaganda computacional na construção de consenso, autoridade, legitimidade e representatividade de diferentes agentes para discutir seus efeitos nos processos políticos contemporâneos.

As pesquisas em Ciência da Informação, assim como o presente artigo, buscam uma melhor compreensão dos princípios que sustentam a criação, organização, armazenamento, comunicação e utilização eficazes de recursos de informação e conhecimento. Como uma disciplina ampla, tem como objetivo entender como as políticas e práticas na área podem ser construídas sobre bases teóricas mais sólidas para obter um impacto maior no contexto prático mundial. Nesse sentido, esta revisão de literatura dialoga com o campo ao fornecer um espaço para conectar debates sobre as emergentes práticas de manipulação computacional nas redes sociais e como elas estão reconfigurando as relações acadêmicas, sociais, comerciais, governamentais e informacionais.

2 MÉTODO

Com o objetivo de mapear a pesquisa existente, indicando as abordagens metodológicas e identificando lacunas e agendas de pesquisa, conduzimos uma RSL sobre o uso de perfis em redes sociais para manipular a opinião pública a respeito de questões políticas. Esta abordagem metodológica é indicada para responder uma determinada questão ou testar uma hipótese específica, identificando, avaliando e sintetizando estudos relevantes (PETTICREW; ROBERTS, 2006). A RSL baseia-se na definição clara da questão de pesquisa, da estratégia de busca, dos critérios de seleção dos documentos recuperados e dos procedimentos para análise.

De acordo com a discussão apresentada na seção anterior, duas questões de pesquisa foram definidas: 1) Como perfis de redes sociais são usados para manipulação da opinião pública no contexto da política? e 2) Quais são as consequências sociais e políticas destes agentes de manipulação nas redes sociais? Desenvolvemos a expressão de busca para aplicá-la em quatro bases de dados científicas, selecionadas devido a reputação, escopo e cobertura temática: Scopus, Web of Science, EBSCO e IEEE Xplore. A busca foi realizada em novembro de 2017 e resultou em 467 documentos, dos quais restaram 369 após a exclusão dos itens duplicados.

Definimos critérios de seleção para garantir a relevância dos artigos em relação às nossas questões de pesquisa. Com o objetivo de capturar o debate internacional, apenas artigos científicos, trabalhos publicados em anais de congressos, livros ou capítulos de livros escritos em inglês foram considerados (PETTICREW; ROBERTS, 2006). Excluímos

artigos que se concentravam na produção de *fake news* e na descrição de rotinas computacionais. Com base nos critérios de inclusão, selecionamos as pesquisas focadas em agentes de manipulação (por exemplo, *bots*, perfis falsos e *trolls*) e no contexto político de atuação destes agentes. A seleção de estudos primários resultou em 45 artigos e foi baseada na codificação de título, palavras-chave e resumos realizada por dois autores. Com a validação de um terceiro autor, 34 artigos permaneceram no *corpus* de análise.

Posteriormente, os artigos foram avaliados em termos de qualidade metodológica e científica, a fim de reduzir o viés e fortalecer as evidências fornecidas pela revisão (PETTICREW; ROBERTS, 2006). Adaptamos os critérios de Najafabadi e Mahrin (2016) e classificamos os artigos de acordo com a clareza dos objetivos, métodos, resultados e limitações. Durante a avaliação de qualidade, cinco documentos foram considerados como não-pertinentes ao escopo da pesquisa, cinco foram excluídos por baixa qualidade metodológica e dois não puderam ser avaliados, uma vez que não conseguimos acesso aos textos na íntegra. Depois da leitura dos textos completos, outros seis artigos foram desconsiderados, por não se adequarem ao escopo desta revisão. Ao final do processo, restaram 16 artigos pertinentes.

Então, sintetizamos as evidências primárias por meio de uma meta-síntese, em um processo iterativo e indutivo de agrupamento de temas em categorias baseadas nas semelhanças, diferenças e relações entre as pesquisas (PETTICREW *et al.*, 2013). A combinação de dados primários pode, ao lidar com pesquisas com variáveis semelhantes, generalizar e redefinir os limites de aplicabilidade dos resultados de cada estudo primário. Esta meta-síntese, baseada na identificação de tópicos proeminentes ou recorrentes na literatura (DIXON-WOODS *et al.*, 2005), resultou em cinco abordagens temáticas que serão apresentadas e discutidas nas próximas seções.

3 RESULTADOS E ANÁLISE DESCRITIVA

Após a seleção dos 16 artigos pertinentes, realizamos análises descritivas, com base nos metadados extraídos e informações sobre as pesquisas. Do total de 50 autores, 31 são homens (62%) e 19, mulheres (38%), com uma concentração de pesquisadores filiados a instituições norte-americanas (38%). Dois artigos foram publicados em 2015,

oito em 2016, e seis, em 2017. A respeito da nacionalidade dos periódicos, nove são norte-americanos, um alemão e um espanhol.

Duas grandes áreas do conhecimento estão pesquisando sobre o tema: comunicação (cinco trabalhos) e computação (onze trabalhos)¹. Onze pesquisas são artigos publicados em periódicos e cinco em anais de congressos, todos apresentados na área de computação. Os cinco trabalhos da área da comunicação foram publicados em periódicos científicos, sendo quatro reunidos em uma edição especial do *International Journal of Communication*.

Como podemos observar no Quadro 1, verificamos que três artigos não consideram processos políticos de países específicos. Quatro estudos têm como contexto de pesquisa processos políticos ocorridos nos Estados Unidos, dois na Espanha e um nos seguintes países: Brasil, Canadá, Equador, França, Indonésia e Reino Unido. Dos dezesseis trabalhos analisados, seis são estudos primários mistos, cinco são estudos primários qualitativos, dois são estudos primários quantitativos e três são revisões de literatura, porém nenhuma sistemática.

Considerando as redes sociais analisadas, doze estudos investigam dados extraídos do Twitter, um extrai informações do Facebook, dois lidam com dados de mais de uma rede social e um não especifica a plataforma pesquisada. A respeito do objeto de pesquisa, onze artigos analisam *bots*, três investigam *trolls*, três estudam perfis falsos, três examinam *spammers* e nove verificam campanhas de cascata de informação. Sobre os processos políticos investigados, dez pesquisas abordam disputa eleitoral; oito discutem cargos eleitos ou personagens políticos; cinco envolvem partidos políticos; três abordam política de Estado (construção de imagem de governo) e três analisam discussões envolvendo grupos minoritários (feminismo, xenofobia e etc.).

Quadro 1: Quadro sinóptico dos artigos empíricos incluídos na RSL

Artigo	País estudado	Rede social	Questão política estudada	Agente de manipulação estudado	Abordagem de Pesquisa
Sadiq et al. (2017)	EUA	Twitter	Disputa eleitoral; personagem político	Bot; perfil falso; campanha de cascata de informação	Estudo primário misto
Woolley (2016)	Análise global	N/A	Política de estado; disputa eleitoral	Bot	Revisão de literatura
Murthy et al. (2016)	Reino Unido	Twitter	Disputa eleitoral	Bot; campanha de cascata de informação	Estudo primário misto

¹ Classificação utilizada pela base de dados Scopus.

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Ibrahim et al. (2015)	Indonésia	Twitter	Disputa eleitoral	Campanha de cascata de informação	Estudo primário misto
Cerón-Guzmán; León (2015)	Colômbia	Twitter	Disputa eleitoral	Spammer	Estudo primário misto
Ferrara (2017)	França	Twitter	Disputa eleitoral; personagem político	Bot; campanha de cascata de informação	Estudo primário misto
Stieglitz et al (2017)	EUA	Twitter	Disputa eleitoral; personagem político	Bot; perfil falso	Estudo primário misto
Ben-David; Fernández (2016)	Espanha	Facebook	Disputa ideológica	Trolagem; campanha de cascata de informação	Estudo primário qualitativo
Ford; Dubois; Puschmann (2016)	Canadá	Twitter	Personagem político	Bot	Estudo primário qualitativo
Bessi; Ferrara (2016)	EUA	Twitter	Disputa eleitoral; personagem político; partidos políticos	Bot; campanha de cascata de informação; spammer	Estudo primário quantitativo
Mustafaraj; Metaxas (2017)	EUA	Twitter	Política de Estado; disputa eleitoral; personagem político; partidos políticos	Bot; campanha de cascata de informação; perfil falso; trolagem; spammer	Revisão de literatura
Oliveira et al. (2016)	Brasil	Twitter	Política de Estado; personagem político; partidos políticos; disputa ideológica	Bot; campanha de cascata de informação	Estudo primário misto
Ferrara et al. (2016)	N/A	Twitter; Facebook	N/A	Bot	Revisão de literatura
Abril (2017)	Espanha	Twitter	Disputa eleitoral; partidos políticos	Trolagem	Estudo primário misto
Maréchal (2016)	N/A	Twitter; Facebook ; Reddit; Telegram	Disputa ideológica	Bot	Estudo primário qualitativo
Recalde et al. (2017)	Equador	Twitter	Campanha de cascata de informação	Cargo eleito ou personagem político; partidos políticos	Estudo primário qualitativo

Fonte: Elaborado pelos autores.

4 ANÁLISE CRÍTICA

Esta seção apresenta uma análise crítica dos artigos pertinentes, introduzindo uma discussão dos temas relevantes de cada trabalho. Os artigos foram classificados de acordo com cinco abordagens (1) Políticas de uso e neutralidade das plataformas; (2) Definição, detecção e caracterização dos agentes de manipulação; (3) Atuação política dos agentes nas redes sociais; (4) Tentativas de previsão de resultado eleitoral com uso de dados das redes sociais; (5) Dinâmicas da rede e estratégias de propagação da informação.

4.1 Políticas de uso e neutralidade das plataformas

Esta temática engloba as políticas de uso e as respostas das plataformas de redes sociais às iniciativas online de manipulação da opinião pública. Os estudos de Ben-David e Matamoros-Fernandéz (2016) e de Marechal (2016) descrevem duas lacunas de regulação: (1) categorização e monitoramento de discurso de ódio; e (2) definição e uso de *bots*.

A partir da análise de páginas do Facebook de partidos de extrema direita espanhóis, Ben-David e Matamoros-Fernandéz (2016) descrevem a circulação online de discurso de ódio, favorecida pelas características técnicas e pelas limitações dos termos de uso da plataforma. As autoras indicam que, em geral, discurso de ódio circula por meio da publicação de *links* para *sites* externos, de conteúdo compartilhado e de comentários em páginas.

Marechal (2016) analisa os termos de uso do Facebook, Twitter, Reddit e Telegram e argumenta que estas empresas não garantem plenamente os direitos dos usuários. Assim, a autora defende que seja desenvolvida uma regulamentação geral para o uso de *bots* nestas plataformas, a partir de três parâmetros: (1) transparência, *bots* devem se identificar como contas automatizadas; (2) consentimento, o contato e o envio de mensagens para usuários comuns deve ocorrer somente com autorização; e (3) limitação do uso secundário de dados, dados capturados por *bots* só devem ser utilizados para fins consentidos explicitamente pelos usuários.

Tanto Ben-David e Matamoros-Fernandéz (2016) quanto Marechal (2016) argumentam que a regulamentação das plataformas enfrenta obstáculos uma vez que as empresas priorizam as receitas em detrimento dos direitos dos usuários. A perspectiva de regulamentação das plataformas poderá passar no futuro pela definição de marcos legais mais amplos (BEN-DAVID; FERNÁNDEZ, 2016) ou pela concorrência de novas empresas que consideram a adequação às normas propostas pela sociedade civil como uma vantagem competitiva (MARÉCHAL, 2016).

4.2. Definição, detecção e caracterização dos agentes de manipulação

Nesta seção, reunimos os artigos que procuram, por meio de diferentes métodos e formas de análise, detectar e identificar os agentes de intervenção e manipulação, assim

como caracterizar suas atividades nas mídias sociais. A respeito da definição dos agentes automatizados, Stiegliz *et al.* (2017) usam o termo *bot* para descrever um *software* projetado para automatizar uma tarefa em ambiente computacional. Para Murthy *et al.* (2016), *bots* são contas em mídias sociais controladas total ou parcialmente por *softwares*. Ferrara (2017) se refere aos *bots* como contas automatizadas em mídias sociais, que têm o intuito de enganar e manipular. Na definição de Bessi e Ferrara (2016), *bots* são contas controladas por algoritmos que emulam o comportamento de usuários humanos, mas operam com maior velocidade. Ferrara *et al.* (2016) definem o termo *social bot* como um algoritmo que produz conteúdo automaticamente e interage com humanos nas redes sociais tentando emular e, possivelmente, influenciar seu comportamento.

Sadiq *et al.* (2017) definem o termo *chatbot* como um *software* rudimentar com automação mínima e habilidades básicas de conversação, cujo *script* é direcionado a redes sociais para atacar, manipular fatos, ou causar ruído. Ford *et al.* (2016) examinam as relações dos *WikiEdits bots* com jornalistas, políticos e o governo canadense. Os *WikiEdit bots* são bots de transparência, que divulgam no Twitter quando determinadas instituições e organizações, principalmente do governo, editam ou publicam artigos na Wikipedia.

Os *bots* são os agentes de intervenção e manipulação mais investigados, porém outros personagens surgem como objeto de estudo, como *trolls*, *buzzers* e *spammers*. Estes agentes podem também ser automatizados e podem ser utilizados com a finalidade de causar ruído. O trabalho de Abril (2016) avalia a atuação de *trolls* durante as eleições parlamentares de 2012 na Catalunha. *Trolls* são perfis que geram mensagens racistas, xenófobas, homofóbicas, misóginas, classistas ou de conteúdo similar (Abril 2016 *apud* Tabachnik 2012) com o intuito de perturbar a discussão online.

Ibrahim *et al.* (2015) definem *buzzers* como contas que dão apoio a candidatos durante uma disputa eleitoral, enquanto difamam os adversários, podendo ser *bots*, apoiadores pagos ou usuários comuns. Já Cerón-Guzmán e León (2015) definem *spammers* como contas criadas para infiltrar redes sociais sem serem detectadas por sistemas de segurança, mimetizar o comportamento humano e ganhar a confiança de pessoas reais. Os autores descrevem dois principais objetivos dos *spammers*: (1) fabricar base de seguidores de um usuário, aumentando sua popularidade e (2) difamar determinado perfil, beneficiando indiretamente outro usuário.

Com o objetivo de entender o comportamento de *bots* no Twitter, Murthy *et al.* (2016) adquiriram rotinas computacionais à venda na Internet para automatizar contas criadas por usuários voluntários e estudar suas possíveis atividades: produção de conteúdo, disseminação de *spam* e manipulação de discussões. Os programadores de *bots* vêm desenvolvendo rotinas computacionais cada vez mais sofisticadas, tornando a detecção automática destas contas mais complexa.

Stiegliz *et al.* (2017) definem uma série de hipóteses a partir da literatura na tentativa de diferenciar comportamento de *bots* do comportamento de humanos. Ferrara *et al.* (2016) sugerem uma taxonomia para detecção de *bots*: (i) sistema baseado na rede de amigos; (ii) sistema baseado em *crowdsourcing*, no qual um mesmo perfil é analisado por diversas pessoas; (iii) métodos de *machine learning*, que permitem identificar diferenças entre *bots* e humanos; e (iv) métodos mistos, que possibilitam analisar múltiplas dimensões do comportamento dos usuários. A partir de dados do Twitter, Cerón-Guzmán and León (2015) criaram técnicas de *machine learning* e identificaram que 22% do total de contas analisadas era *spammers*. Ferrara (2017) também usa técnicas de *machine learning* para detectar contas automatizadas e conclui que possivelmente há um mercado ilegal de reutilização de *bots* para disseminar desinformação política. Bessi and Ferrara (2016) usam o BotOrNot², ferramenta desenvolvida com o uso de *machine learning*, para fazer a detecção de *bots*, e apontam que os *bots* sociais são ferramentas de manipulação do debate político. Sadiq *et al.* (2017) utilizam técnicas de *machine learning* para analisar a postura de usuários reais no Twitter através de seus posts.

4.3. Atuação política dos agentes nas redes sociais

A atuação e o impacto político dos agentes de manipulação nas redes sociais foi foco de estudo de sete artigos desta revisão. Esse tópico trata de entender como atuam essas contas quanto têm por objetivo influenciar discussões políticas, interferir na formação da opinião dos usuários, disseminar propaganda de campanha, criar ruído no debate público, endossar e difamar candidatos. Ferrara *et al.* (2016) conduziram uma revisão de literatura sobre a influências de agentes automatizados nos ambientes online e offline. Os autores reforçam que a presença de *bots* na Internet pode ser neutra e até

² O Botometer (anteriormente BotOrNot) verifica a atividade das contas do Twitter e dá a elas uma pontuação baseada na probabilidade de elas serem bots.

positiva para o usuário. No entanto, tanto *bots* inofensivos quanto maliciosos podem criar cascatas de informação a partir de rumores ou conteúdo falso, causando ruído nas redes e inflando artificialmente as bases de seguidores de determinados perfis. Woolley (2016) destaca a eficácia dos *bots* em atacar, sequestrar e alterar discursos em redes sociais, identificando três formas de atuação política: (1) apoiar um político específico; (2) desmobilizar iniciativas adversárias; (3) impulsionar o número de seguidores de certos perfis. As duas primeiras estratégias tendem a ser mais utilizadas em países mais autoritários e a terceira, em países mais democráticos.

Bessi e Ferrara (2016) observaram a atuação de *bots* sociais e sua influência na discussão política online durante as eleições presidenciais norte-americanas de 2016. Após realizarem análise de sentimento de *tweets* postados por humanos e *bots*, os autores identificaram que, durante eventos externos à rede, como debates televisionados, usuários comuns tiveram participação mais ativa que *bots*. Os autores explicam que *bots* são programados para divulgar informação de forma sistemática e contínua, mas usuários comuns tendem a se engajar em eventos específicos ou inesperados. Além disso, *bots* produziram mais conteúdo positivo em apoio a Donald Trump, que ganhou as eleições. Isto pode ter influenciado a percepção de sua popularidade na Internet, por sugerir a existência de apoio orgânico. Dessa forma, *bots* sociais podem falsear popularidade, polarizar o debate político e aumentar a disseminação de desinformação.

Muitos *bots* são programados para simular o comportamento de usuários humanos, a fim de influenciar a opinião pública. Stieglitz *et al.* (2017) indicam que *bots* podem se comunicar de forma tão eficaz quanto usuários reais. A resposta afetiva das pessoas a conteúdos publicados por *bots* é equiparada às respostas a conteúdos publicados por usuários comuns.

A respeito da dinâmica de disseminação de mensagens no Twitter, Oliveira *et al.* (2016) investigaram a difusão de *retweets* publicados por *bots* durante dois protestos ocorridos no Brasil em 2015, um contra o governo e outro a favor. Os autores argumentam que é necessário considerar a atuação dessas contas automatizadas, por representarem um risco de superestimar o impacto da disseminação de mensagens. O potencial de publicação de conteúdo dos *bots* é maior que o dos usuários comuns, o que

aumenta a capacidade das contas automatizadas de gerar empatia, engajamento e impacto social.

Recalde *et al.* (2017) investigam perfis definidos como sequestradores de *hashtags*, contas que esvaziam o sentido original de *hashtags* de oponentes políticos para criar ruído nas redes. Partindo de um corpus de postagens do Twitter sobre as manifestações de 2015 e 2016 no Equador, os autores analisaram semanticamente mensagens e classificaram 17% dos usuários da amostra como sequestradores de *hashtags*. Os autores argumentam que o Twitter deve incluir uma identificação de sequestradores no seu algoritmo de recomendação de pessoas a serem seguidas, uma vez que esse comportamento distorce a percepção do posicionamento político de usuários.

Dentro do espectro de atuação políticas de contas nas redes sociais, pesquisadores também investigaram os chamados “*bots de transparência*” (Ford *et al.* 2016). Os WikiEdits *bot* são programados para monitorar e divulgar no Twitter quaisquer edições feitas na Wikipedia por IPs vinculados ao governo. Apesar da vigilância ser direcionada e limitada pelos programadores desta ferramenta, a atuação destas contas abre o debate sobre valores como transparência e sigilo na disputa de um campo político mais amplo.

4.4. Tentativas de previsão de resultado eleitoral com uso de dados das redes sociais

Os artigos analisados nesta seção discutem a possibilidade de previsão dos resultados de eleições a partir da extração e análise de dados das redes sociais. Coloca-se, portanto, a necessidade de separar a influência de perfis automatizados das tendências de opinião exclusivamente humanas em análises eleitorais nas redes sociais. Os trabalhos de Ibrahim *et al.* (2015) e de Sadiq *et al.* (2017) procuram não só medir o peso dessa influência automatizada, como propor correlações entre a escala de apoio real a cada candidato nas redes e o resultado em número de votos ao final da eleição.

Ibrahim *et al.* (2015) propuseram um método de previsão de resultado baseado no Twitter para a eleição presidencial na Indonésia em 2014, cuja taxa de erro de 0,61% foi inferior às das pesquisas de opinião tradicionais. A partir de uma análise de sentimento, as mensagens foram classificadas como positivas ou negativas em relação aos candidatos mencionados. Essa pontuação foi utilizada para definir a intenção de voto dos usuários. Para estudar a eleição norte-americana de 2016, Sadiq *et al.* (2017) também

atribuíram aos *tweets* sentimentos positivo, negativo ou neutro em função de cada candidato e tiveram 80% de eficácia em detectar afinidades eleitorais de contas humanas.

As duas pesquisas mencionadas testam a eficiência de modelos de análise emocional aplicadas tanto a contas de *bots* quanto a contas humanas para calcular a probabilidade de cada usuário apoiar um candidato e, em consequência, medir o apoio total na rede a cada um dos políticos (Ibrahim *et al.* 2015; Sadiq *et al.* 2017). Excluído o apoio de *bots* da amostra analisada, projeta-se a proporção de apoios positivos e negativos a cada candidato no total de votos esperados, obtendo-se uma previsão da distribuição de intenções e, portanto, do resultado da eleição (Ibrahim *et al.* 2015). É importante questionar a eficiência alegadamente superior da previsão eleitoral por meio da análise de redes sociais, uma vez que ela abre uma discussão epistemológica nas ciências políticas sobre métodos de pesquisa de opinião e projeções demográficas de voto.

4.5. Dinâmicas da rede e estratégias de propagação da informação

Dos artigos pertinentes, cinco apontam para uma tendência de pesquisa relacionada a possíveis estratégias de propagação da informação. Tais estudos investigam estratégias para criar cascatas informacionais, isto é, ferramentas computacionais para manipular tomadas de decisão sequenciais, buscando viralizar comportamentos ou dados. Na pesquisa realizada durante as eleições parlamentares de 2012 na Catalunha, Abril (2017) investigou a existência e o comportamento de *trolls* no Twitter. Apesar de ter identificado baixa taxa de sucesso nas atividades de trolagem, a autora destacou que, quando bem sucedidas, estas podem gerar propagação artificial da informação e, conseqüentemente, um impacto na opinião pública.

Murthy *et al.* (2016), realizaram um experimento no Twitter em que voluntários criaram contas automatizadas na plataforma para interagir com os eventos midiáticos relacionados às eleições gerais do Reino Unido em 2015. Os *bots* fracassaram ao tentar influenciar as lógicas de difusão da informação na rede, já que as contas criadas para a pesquisa não conseguiram adquirir capital social. Os *bots* não foram capazes de ganhar centralidade, permanecendo distantes dos influenciadores previamente existentes na rede.

Ben-David e Matamoros-Fernandéz (2016) monitoraram a propagação do discurso de ódio no Facebook para testar a hipótese de que a dinâmica da rede favorece a disseminação deste tipo de conteúdo. A lógica dos algoritmos da plataforma cria “filtros invisíveis” (PARISER, 2011) tendo como consequência mais evidente a recomendação de conteúdo redundante. Isto é, usuários com comportamento racista ou discriminatório no Facebook recebem recomendações de conteúdo igualmente intolerante dentro da plataforma.

Ferrara (2017) investiga as campanhas de desinformação operadas por *bots* sociais no Twitter durante as eleições presidenciais francesas de 2017. O estudo analisa o comportamento de *bots* e humanos, além das interações entre estes perfis, concluindo que *bots* atuaram na propagação de desinformação, ainda que sem grande efetividade. Na pesquisa realizada por Mustafaraj *et al.* (2017) durante as eleições presidenciais norte-americanas de 2016, o objetivo foi avaliar os esforços organizados para difusão de desinformação na campanha eleitoral. Os resultados apontam que o passo mais importante para a propagação de desinformação é a inserção de usuários falsos em redes de usuários reais, para que estes repassem o conteúdo dentro de suas próprias redes.

5 CONCLUSÃO

A partir da revisão sistemática de literatura realizada neste artigo, percebe-se que a propaganda computacional é um fenômeno duplo, tanto tecnológico como social. No âmbito acadêmico, é um tema emergente na área da Computação, mas ainda incipiente na Ciência da Informação, o que indica uma importante agenda de pesquisa para o campo, e para também para a grande área das Ciências Sociais. Por um lado, um amplo conhecimento técnico é necessário para extrair dados, lidar com big data, criar algoritmos, produzir *bots* sofisticados e avaliá-los. Por outro, considerar a automação somente sob uma perspectiva técnica nos impede de conhecer seus possíveis efeitos sociais, como por exemplo, seu impacto na democracia, na disseminação de *fake news*, na propagação de discurso de ódio e na percepção da realidade.

As pesquisas ainda estão focadas na criação de ferramentas precisas para detecção de *bots*, mas pouco se sabe sobre como atuam e quais são suas consequências. Não encontramos estudos que apresentem estratégias e técnicas para identificar quem cria, quem financia, quem comercializa e para quem trabalham os robôs ou *cyborgs*. Há

uma demanda latente por métodos eficazes de trabalho de campo para compreender profundamente os atores e as intenções por trás da propaganda computacional. As pesquisas reunidas nesta revisão enfrentam obstáculos na identificação dos responsáveis pelas contas automatizadas, por causa do anonimato dos *bots* e da falta de regulamentação e fiscalização das plataformas. Não há, por ora, transparência e informação suficientes sobre o funcionamento interno das plataformas de redes sociais, sobre seus algoritmos e sobre o comportamento dos agentes.

Há necessidade de desenvolver um arcabouço conceitual e metodológico capaz de definir comportamentos artificiais, como por exemplo: *bots*; usuários fanáticos; contas pessoais administradas por empresas; *trolls*; *buzzers* e *spammers*. Em que medida estes agentes de intervenção diferem quanto ao seu impacto na opinião pública? Ou representam diferentes meios para o mesmo fim? É possível considerar que eles têm um papel de complementaridade entre si diante da propaganda computacional? Estas são perguntas em aberto que exigirão o desenvolvimento de conceitos rigorosos e posicionamento crítico dos pesquisadores diante dos fenômenos tecnológicos, políticos e informacionais.

Outro problema levantado a partir da varredura da literatura é a concentração de trabalhos empíricos baseados em dados extraídos do Twitter. Há poucas pesquisas sobre a presença de *bots* em outras redes sociais. O Twitter possui especificidades que aprofundam ainda mais o viés dos resultados destes estudos: por exemplo, os mais assíduos produtores de conteúdo no Twitter são empresas, marcas, pessoas famosas, micro-celebridades, políticos e pessoas influentes de diferentes tipos e nichos. Além disso, seus 330 milhões de usuários equivalem somente a 16,5% do público do Facebook, que conta com mais de dois bilhões de contas (MOLINA, 2017), o que significa uma baixa representatividade diante das audiências das redes sociais online.

Portanto, cabe questionar de forma mais profunda o papel do Twitter na formação da opinião pública. Teria o Twitter uma função de *two-step flow*, de enfatizar o papel dos formadores de opinião na influência dos usuários comuns, porém de forma distribuída, ou seja, em escala micro? Será que, apesar da baixa penetração, a plataforma tem o poder de *agenda setting* da mídia tradicional? Ou a plataforma funcionaria principalmente como segunda tela, como em transmissões ao vivo, complementando o consumo das

mídias de massa? Neste cenário, é preciso discutir como as teorias clássicas da comunicação e da informação podem ser revisitadas e atualizadas.

Sobre os aspectos metodológicos, é comum que as análises de redes sociais fiquem restritas ao número de amigos, de seguidores, de posts e de compartilhamentos, e as mais sofisticadas conseguem avaliar semanticamente os conteúdos publicados. O Twitter não fornece dados sobre o perfil sócio-demográfico de seus usuários, o que limita muito as possibilidades de pesquisa. Localização geográfica, afiliação religiosa, preferências políticas, gênero, nível educacional e outras variáveis relacionadas com o comportamento social são dados quase impossíveis de extrair das redes sociais de forma automática. Pouco se sabe sobre como estas variáveis afetam a circulação da propaganda computacional e sua capacidade de influenciar a opinião pública online. É preciso recorrer a métodos qualitativos e mistos para considerar estes aspectos sociológicos, o que pode tornar as pesquisas mais lentas e custosas.

Na era digital, a Ciência da Informação se apresenta como um espaço interdisciplinar amplo que trata do impacto das técnicas de geração, organização, preservação, disseminação, acesso e recuperação da informação em diversas esferas da atividade humana. Uma vez que esta revisão de literatura trata de práticas e estudos sobre os aspectos sociais da informação, assim como seus desenvolvimentos tecnológicos, dialogamos com a literatura da Ciência da Informação uma vez que estudamos questões éticas e políticas da informação. Ao adotarmos uma perspectiva crítica sobre os usos das tecnologias de informação e comunicação, atualizamos os fundamentos teóricos da pesquisa sobre implicações políticas, sociais e econômicas da informação.

Dada a atual importância das redes sociais para as democracias contemporâneas, esta revisão pode substanciar debates sobre a regulamentação destas plataformas, por oferecer uma síntese sobre este fenômeno para instituições públicas, pesquisadores, legisladores e organizações da sociedade civil. A compreensão das fraudes e estratégias controversas usadas nas redes para influenciar a opinião pública em processos políticos pode apontar para uma agenda de ações, providências e políticas públicas que visem proteger os princípios democráticos dos maus-usos das plataformas de mídias sociais no Brasil e no mundo. Portanto, é crucial que a comunidade acadêmica compartilhe dados e

informação diante dos desafios éticos, sociais e políticos impostos pelas iniciativas de manipulação online.

REFERÊNCIAS

BOLSOVER, G.; HOWARD, P. Computational Propaganda and Political Big Data: Moving Toward a More Critical Research Agenda. **Big Data**, 2017. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042537361&doi=10.1089%2fbig.2017.29024.cpr&partnerID=40&md5=1bd3f6cee7514f6a1c483cb8b9433257>

BRADSHAW, Samantha; HOWARD, Philip N. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. In: WOOLLEY, Samuel; HOWARD, Philip N. (Eds.). **Working Paper 2017.12**. Oxford, UK: Project on Computational Propaganda., 2017. p. 37.

DACOMBE, Rod. Systematic Reviews in Political Science: What Can the Approach Contribute to Political Research? **Political Studies Review**, v. 16, n. 2, p. 148–157, 2018.

DAYEN, David. **How Twitter Secretly Benefits From Bots and Fake Accounts**The **Intercept**, 2017. Disponível em: <https://theintercept.com/2017/11/06/how-twitter-secretly-benefits-from-bots-and-fake-accounts/>. Acesso em: 9 mar. 2019.

DIXON-WOODS, Mary et al. Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods. **Journal of health services research & policy**, v. 10, p. 45–53, 2005.

MOLINA, Brett. **Twitter overcounted active users since 2014**. 2017. Disponível em: <https://www.usatoday.com/story/tech/news/2017/10/26/twitter-overcounted-active-users-since-2014-shares-surge/801968001/>. Acesso em: 30 jul. 2018.

NAJAFABADI, Maryam Khanian; MAHRIN, Mohd Naz'ri. A Systematic Literature Review on the State of Research and Practice of Collaborative Filtering Technique and Implicit Feedback. **Artif. Intell. Rev.**, v. 45, n. 2, p. 167–201, 2016.

PARISER, Eli. **The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think**. [s.l.] : Penguin, 2011.

PETTICREW, Mark et al. Synthesizing evidence on complex interventions: How meta-analytical, qualitative, and mixed-method approaches can contribute. **Journal of clinical epidemiology**, v. 66, 2013.

PETTICREW, Mark; ROBERTS, Helen. Systematic Reviews in the Social Sciences. p. 354, 2006.

THE ECONOMIST. Do social media threaten democracy? **The Economist**, 2017.
Disponível em: <https://www.economist.com/leaders/2017/11/04/do-social-media-threaten-democracy>. Acesso em: 15 maio. 2019.

WOOLLEY, S. C.; HOWARD, P. N. Computational Propaganda Worldwide: Executive Summary. In: WOOLLEY, Samuel; HOWARD, Philip N. (Eds.). **Working Paper 2017.11**. Oxford, UK: Project on Computational Propaganda, 2017. p. 14.

WOOLLEY, Samuel. The Political Economy of Bots: Theory and Method in the Study of Social Automation. In: **The political economy of robots**. International Political Economy Series[s.l.] : Springer Berlin Heidelberg, 2018. p. 127–155.

REFERÊNCIAS - RESULTADOS EMPÍRICOS

ABRIL, Eulàlia Puig. Unmasking Trolls: Political Discussion on Twitter During the Parliamentary Election in Catalonia. **Trípodos**, v. 0, n. 39, p. 53-69, 2017.

BEN-DAVID, Anat; FERNÁNDEZ, Ariadna Matamoros. Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain. **International Journal of Communication**, v. 10, n. 0, p. 27, 2016.

BESSI, Alessandro; FERRARA, Emilio. Social bots distort the 2016 U.S. Presidential election online discussion. **First Monday**, v. 21, n. 11, 2016. Disponível em: <http://firstmonday.org/ojs/index.php/fm/article/view/7090>. Acesso em: 19 jul. 2018.

CERÓN-GUZMÁN, Jhon Adrián; LEÓN, Elizabeth. Detecting Social Spammers in Colombia 2014 Presidential Election. In: (Obdulía Pichardo Lagunas, Oscar Herrera Alcántara, Gustavo Arroyo Figueroa, Eds.) **ADVANCES IN ARTIFICIAL INTELLIGENCE AND ITS APPLICATIONS 2015, Anais...** : Springer International Publishing, 2015.

DE OLIVEIRA, Éric Tadeu et al. The Influence of Retweeting Robots During Brazilian Protests. In: 2016 49th Hawaii International Conference on System Sciences (HICSS) **Anais...** Hawaii, 2016.

FERRARA, Emilio et al. The rise of social bots. **Communications of the ACM**, v. 59, n. 7, p. 96–104, 2016.

FERRARA, Emilio. Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. **arXiv:1707.00086 [physics]**, 2017. Disponível em: <http://arxiv.org/abs/1707.00086>. Acesso em: 18 jun. 2018.

FORD, Heather; DUBOIS, Elizabeth; PUSCHMANN, Cornelius. Keeping Ottawa Honest-One Tweet at a Time? Politicians, Journalists, Wikipedians, and Their Twitter Bots. **INTERNATIONAL JOURNAL OF COMMUNICATION**, v. 10, p. 4891–4914, 2016.

IBRAHIM, Mochamad et al. Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation. In: 2015, **Anais...** [s.l: s.n.]

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

MARÉCHAL, Nathalie. Automation, algorithms, and politics | when bots tweet: Toward a normative framework for bots on social networking sites (feature). **International Journal of Communication**, v. 10, p. 10, 2016.

MURTHY, D. et al. Bots and political influence: A sociotechnical investigation of social network capital. **International Journal of Communication**, 2016. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85017188067&partnerID=40&md5=1ee0eae3e3f7c1a89d08f66959c1e4ce>

MUSTAFARAJ, Eni; METAXAS, Panagiotis Takis. The Fake News Spreading Plague: Was it Preventable? In: 2017, **Anais...** : ACM Press, 2017. Disponível em: <http://dl.acm.org/citation.cfm?doid=3091478.3091523>. Acesso em: 19 jun. 2018.

RECALDE, Lorena et al. Who You Should Not Follow: Extracting Word Embeddings from Tweets to Identify Groups of Interest and Hijackers in Demonstrations. **IEEE Transactions on Emerging Topics in Computing**, v. PP, p. 1–1, 2017.

SADIQ, Saad et al. AAFA: Associative Affinity Factor Analysis for Bot Detection and Stance Classification in Twitter. In: (Zhang, C and Palanisamy, B and Khan, L and Sarvestani, SS, Ed.) 2017 IEEE 18TH INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION (IEEE IRI 2017) 2017, 345 E 47TH ST, NEW YORK, NY 10017 USA. **Anais...** 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2017.

STIEGLITZ, Stefan et al. Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with Those of Humans. In: MEISELWITZ, Gabriele (Ed.). **Social Computing and Social Media. Human Behavior**. Cham: Springer International Publishing, 2017. v. 10282p. 379–395.

WOOLLEY, S. C. Automating power: Social bot interference in global politics. **First Monday**, 2016. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963548307&doi=10.5210%2ffm.v21i4.6161&partnerID=40&md5=71562c073d93ff8ad2846292692c9089>.