




DOSSIÊ

DEMOCRACIA SOB ATAQUE: POLARIZAÇÃO POLÍTICA E PRODUÇÃO DE CONTEÚDOS HOSTIS NO TWITTER NAS ELEIÇÕES DE 2022

*DEMOCRACY UNDER ATTACK:
POLITICAL POLARIZATION AND
PRODUCTION OF HOSTILE CONTENT
ON TWITTER DURING THE 2022
ELECTIONS*

Patrícia Dias dos Santos* 

Claudio Luis de Camargo Penteado** 

Laura Damaceno de Almeida*** 

Denise Hideko Goya**** 

* Universidade Federal do ABC, Programa de Pós-Graduação em Ciência da Computação (PPGCCM), Santo André/SP, Brasil.

E-mail: patricia.santos@ufabc.edu.br.

** Universidade Federal do ABC, Centro de Engenharia e Ciências Sociais (CECS), São Bernardo do Campo/SP, Brasil.

E-mail: claudio.penteado@ufabc.edu.br.

*** Universidade Federal do ABC, Programa de Pós-Graduação em Ciência da Computação (PPGCCM), Santo André/SP, Brasil.

E-mail: l.damaceno@ufabc.edu.br.

**** Universidade Federal do ABC, Centro de Matemática, Computação e Cognição (CMCC), Santo André/SP, Brasil. email: denise.goya@ufabc.edu.br.

RESUMO

Este artigo analisa o rendimento teórico do clássico ensaio de Marcel Mauss na discussão a respeito da ideia de cidadania em Robert Putnam. Autores concentrados em pensar os fundamentos não formalmente expressos do contrato, seus pressupostos remontam ao debate seminal nas ciências sociais sobre obrigatoriedade versus espontaneidade. Associado aos estudos sobre cidadania, em Comunidade e democracia, Putnam investiga os diferentes graus de engajamento cívico e sua relação com as práticas governamentais, variável capaz de indicar a qualidade dos regimes democráticos pela via de um sistema de prestações. No plano desse debate, cabe explorar, através da revisão bibliográfica, como o Ensaio sobre a dádiva contribui para a compreensão do papel da cultura no desempenho das democracias. Consideramos a hipótese do diálogo entre essas duas versões da aliança na ação cotidiana e sua fecundidade aos estudos sobre o modo pelo qual os indivíduos vivenciam a regra nas democracias modernas. Trata-se de um esforço de interlocução entre os campos da antropologia e da ciência política, em específico o debate sobre o engajamento na vida pública.

Palavras-chave: Teoria Social. Reciprocidade. Democracia. Confiança.

ABSTRACT

This article analyzes the theoretical performance of the classic essay by Marcel Mauss in the discussion about the idea of citizenship in Robert Putnam. Authors focused on thinking about the non-formally expressed foundations of the contract and their assumptions go back to the seminal debate in the social sciences about obligatoriness versus spontaneity. Associated with studies on citizenship, in Making democracy work, Putnam investigates the different degrees of civic engagement and its relationship with government practices, a variable capable of indicating the quality of democratic regimes through a system of benefits. In terms of this debate, it is worth exploring, through a bibliographic review, how the essay The gift contributes to the understanding of the role of culture in the performance of democracies. We consider the hypothesis of the dialogue between these two versions of the alliance in everyday action and its fecundity in studies on the way in which individuals experience rules in modern democracies. It is an effort of dialogue between the fields of anthropology and political science, specifically the debate on engagement in public life.

Keywords: Social Theory. Reciprocity. Democracy. Confidence.

INTRODUÇÃO

A crise da democracia liberal tem possibilitado a emergência de discursos antissistema que questionam a credibilidade e legitimidade das instituições, contexto que favorece o surgimento de lideranças autoritárias (LEVITSKY e ZIBLATT, 2018) e a possibilidade de uma ruptura democrática (CASTELLS, 2018).

A transformação do ecossistema de comunicação com a expansão da internet e a popularização do acesso por meio das plataformas digitais, ao mesmo tempo que possibilitou uma comunicação mais democrática – ao abrir espaço para que novos atores sociais possam expressar suas demandas e opiniões, como mostram os estudos das grandes mobilizações do ciclo de protestos de 2010 a 2013 (CASTELLS, 2017) –, por outro lado permitiu a entrada no debate público de atores políticos radicais (SUNSTEIN, 2018), que criam dificuldades para a formação de consensos democráticos mínimos em torno dos valores da democracia liberal (RASMUSSEN, 2013).

No Brasil, a pesquisa A cara da democracia (IDDC, 2019) mostra que, apesar da maioria da população (64,8%) ainda ser favorável a esse modelo de governo em relação a outras formas, o nível de satisfação é muito baixo, com apenas 33% de pessoas satisfeitas, contra 65% de insatisfeitos. Em relação ao Poder Judiciário, a maioria da população (63,4%) confia pouco ou não confia nesse poder. Esse sentimento de descrédito nas instituições é estimulado pela estratégia discursiva antissistema de Jair Bolsonaro, que constantemente utiliza seus canais em plataformas digitais para tecer críticas ao Supremo Tribunal Federal (STF) e à credibilidade das urnas eletrônicas (NOBRE, 2020).

A desconfiança no Judiciário e a insatisfação com a democracia colocam em risco a credibilidade do sistema eleitoral brasileiro. Essa descrença é alimentada pela difusão de desinformação nas plataformas digitais, com notícias falsas como, por exemplo, a da existência de fraude nas urnas nas eleições de 2018, difundida por perfis influentes (RECUERO, 2020), o que contribui para um processo de deslegitimação das instituições da democracia liberal (CASTELLS, 2018).

Além da existência de um processo massivo de desinformação, o debate on-line também tem se caracterizado pelo conflito discursivo entre perfis com posicionamentos políticos antagônicos (PENTEADO, GOYA e DE FRANÇA, 2021; VINHAS, SAINZ e RECUERO, 2020) e pela expressão de discursos de ódio, produção

de conteúdos hostis e difusão de conspirações que contribuem para a radicalização política (RIEBE et al., 2018).

Dentro desse contexto, o artigo tem por objetivo analisar os ataques ao sistema eleitoral brasileiro nas eleições de 2022 no Twitter, por meio da identificação e classificação de conteúdos hostis relacionados às urnas eletrônicas, ao Tribunal Superior Eleitoral (TSE) e aos ministros desse tribunal. Assim, as questões de pesquisa que este trabalho busca responder são: (P1) Quais os principais tipos de conteúdos hostis? (P2) Qual o posicionamento político dos perfis que publicam conteúdos hostis?

O artigo apresenta uma abordagem teórica e metodológica interdisciplinar, que combina elementos da comunicação política e da ciência da computação, e está organizado da seguinte forma: primeiramente são apresentados o referencial teórico e trabalhos relacionados, subdivididos em discurso de ódio e debate eleitoral, e detecção de toxicidade e o uso de técnicas computacionais, enquanto na seção seguinte é discutida a metodologia proposta. Na sequência, descrevem-se e se discutem os resultados encontrados, e, finalmente, na última seção, são discutidas as conclusões encontradas e possíveis trabalhos futuros.

REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

As mídias sociais, dada a sua usabilidade e popularização entre pessoas conectadas, fornecem uma plataforma para que todos os usuários se expressem livremente; porém, a identificação de conteúdos tóxicos e ofensivos não é rara e pode impactar severamente a experiência do usuário e até mesmo a civilidade de uma comunidade (YIN e ZUBIAGA, 2021). Essa característica tem chamado a atenção de pesquisadores, que, como apresentado a seguir, têm realizado estudos que abordam o uso de conteúdos ofensivos em contexto de debates eleitorais em plataformas digitais com auxílio de técnicas computacionais que possibilitam automatizar a identificação de mensagens com conteúdos hostis em meio a grandes volumes de dados.

DISCURSO DE ÓDIO E O DEBATE ELEITORAL

As pesquisas sobre o debate em plataformas digitais em torno do sistema eleitoral, em geral, abordam o tema da desinformação, como o trabalho de Recuero (2020)

sobre as estratégias no Twitter para legitimar o discurso sobre “fraude nas urnas” nas eleições de 2018, ou a circulação de conteúdos que insinuam a existência de fraude e manipulação eleitoral no Brasil, distribuídos nas campanhas de YouTube entre os anos de 2014 e 2020, conforme pesquisa conduzida por Ruediger e Grassi (2020).

Dadas as características de incivilidade nas conversações políticas em plataformas digitais (ROSSINI, 2020), diversas pesquisas têm se dedicado a estudar os efeitos dos discursos políticos para os processos eleitorais, em especial o discurso de ódio e produção de conteúdos tóxicos. O estudo de Siegel et al. (2019) sobre o discurso de ódio on-line e a retórica nacionalista branca no Twitter da campanha de Trump em 2016, por meio de análise de aprendizado de máquina (machine learning) das publicações na plataforma, aponta que, embora os eventos de campanha estimulassem picos de discurso de ódio, esse tipo de conteúdo tende a se dissipar no decorrer do tempo.

Outra forma de análise dos discursos de ódio é por meio do uso de análise crítica do discurso, o que possibilita uma interpretação dos sentidos mobilizados pelos usuários, como o estudo de Trajkova e Neshkovska (2018) sobre propaganda com discurso de ódio no Facebook e Instagram durante as eleições de 2016 na Macedônia.

Ainda, existem pesquisas que visam verificar se há correlação entre discursos de ódio e formação de câmaras eco-ideológicas (SUNSTEIN, 2018). A radicalização da política nas mídias sociais pode levar à produção de polarização afetiva, que desloca a disputa entre projetos políticos antagônicos para um conflito em torno de afetos e emoções mobilizados pelos grupos políticos em disputa (IYENGAR et al., 2019) e contribui para a disseminação do discurso de ódio no debate político on-line.

Outra forma de estudar o discurso de ódio em plataformas digitais é através da abordagem de análise de redes sociais, que destaca as interações entre os perfis, como o estudo de Mercuri e Lima-Lopes (2020) sobre o discurso de ódio como elemento de uma estratégia discursiva populista de mobilização que favorece a expressão de conteúdos hostis e ataques a grupos estigmatizados.

Uma importante contribuição é o artigo de Pascual-Ferrá et al. (2021), que associa o discurso contra o uso de máscaras na pandemia, durante o segundo semestre de

2020 nos Estados Unidos, com o uso de linguagem tóxica pelos usuários no debate público on-line, realizado por meio da verificação automatizada de conteúdos.

Contudo, inexistem estudos que combinem a análise de discurso de ódio contra as instituições da democracia e a credibilidade do sistema eleitoral, com foco na produção de conteúdos tóxicos, e polarização afetiva ideológica nas plataformas digitais. Desta forma, este artigo visa contribuir com os estudos sobre os ataques ao sistema eleitoral e suas instituições democráticas, por meio da detecção de conteúdos tóxicos e polarização política ideológica, com uso de técnicas computacionais, no debate on-line no Twitter nas eleições de 2022 no Brasil.

DETECÇÃO DE CONTEÚDO HOSTIL COM O APOIO DE TÉCNICAS COMPUTACIONAIS

O uso de técnicas de inteligência artificial é bastante útil para identificar se determinados textos contêm discurso de ódio. De acordo com um levantamento realizado por Fortuna e Nunes (2018), grande parte dessas técnicas se baseia em aprendizado de máquina supervisionado, que requer a intervenção de pessoas para "ensinar" a máquina, por meio de exemplos, quais mensagens caracterizam ou não falas ofensivas e hostis. Os autores discutem várias definições de discurso de ódio, relacionando-as com conceitos similares (porém distintos), como linguagem abusiva, tóxica, discriminatória ou extremista, com o objetivo de demonstrar como a eficiência de uma abordagem computacional é sensível à definição e ao contexto empregados.

Guimarães et al. (2020) apontam que a ferramenta que é considerada estado da arte em detecção de discurso de ódio é a Perspective API (PERSPECTIVE, ©2021), que adota como definição central a toxicidade em redes sociais, que pode ser descrita como “[...] um comentário rude, desrespeitoso ou irracional que provavelmente fará as pessoas deixarem uma discussão” (WULCZYN, THAIN e DIXON, 2017). As linguagens ofensivas e abusivas são utilizadas como termos genéricos no contexto de estudos de detecção automática de toxicidade. Porém, enquanto “fortemente indelicado, rude” e o possível uso de palavrões são vistos em ambas as definições, a linguagem abusiva contém um forte componente de intencionalidade. O discurso de ódio se enquadra em ambas as categorias (YIN e ZUBIAGA, 2021). A Perspective API é uma ferramenta desenvolvida pela Jigsaw, subsidiária do Google, para detectar automaticamente comentários “tóxicos” on-

line. Ela utiliza algoritmos de processamento de linguagem natural e aprendizado de máquina para identificar conteúdos abusivos ou ameaças escritos em diversos idiomas, incluindo o português.

ElSherief et al. (2018) fizeram um estudo comparativo entre os instigadores de discurso de ódio e os usuários-alvo no Twitter utilizando as métricas de toxicidade da ferramenta Perspective API. Eles descobriram que os instigadores escolhem usuários famosos ou muito populares no Twitter, o que aumenta sua própria visibilidade on-line.

Fortuna, Soler e Wanner (2020) analisaram e compararam seis diferentes datasets de discurso de ódio manualmente rotulados em relação à sua semelhança e compatibilidade. Após tratamento para padronizar os rótulos das base de dados, eles submeteram os datasets escolhidos à ferramenta Perspective API, obtendo desempenhos diferentes dependendo das categorias e conjuntos de dados. Os autores concluíram que muitas definições diferentes estão sendo usadas para conceitos equivalentes, o que torna a maioria dos dados disponíveis publicamente incompatíveis.

Guimarães et al. (2020) utilizam a ferramenta Perspective API para caracterizar comentários tóxicos de usuários brasileiros no Facebook no contexto da saída da prisão do ex-presidente Lula. Eles observaram que páginas de mídia recebem mais toxicidade que páginas de figuras públicas, mas os comentários tóxicos são mais prevalentes quando figuras públicas se tornam o tópico de uma postagem.

Desta forma, a Perspective API se mostra como um instrumento importante para o estudo de comentários hostis em mídias sociais, mas há a necessidade de fazer ajustes metodológicos de forma a garantir que as análises e categorias sejam contextualizadas de acordo com o objeto de estudo, no caso conteúdos hostis contra elementos do sistema eleitoral brasileiro, conforme se apresenta na metodologia, a seguir.

METODOLOGIA

Nesta seção são apresentadas as etapas da metodologia proposta para este estudo, incluindo a descrição da coleta de dados, a estratégia adotada para selecionar e agrupar usuários participantes na discussão do tema estudado de acordo com o

seu posicionamento, e como foi inferida a hostilidade dos textos compartilhados por eles no Twitter.

COLETA DE DADOS

Foram coletados tuítes, retuítes e respostas relacionados ao debate sobre o sistema eleitoral brasileiro, em tempo real, entre julho e agosto de 2022, conforme pode ser observado no Quadro 1. Os termos gerais estão relacionados às acaloradas discussões que ocorreram no Brasil em 2021 e 2022, envolvendo argumentos a favor e contra a utilização da urna eletrônica e a adoção do voto impresso. Também foram selecionados nomes de alguns dos ministros que atuam ou atuaram no TSE e STF por eles serem responsáveis por tomar decisões importantes relacionadas às eleições e questões políticas do país, incluindo aqueles mais alinhados ao bolsonarismo. Esses juízes, incluindo o ministro Alexandre de Moraes, uma figura central no debate, foram incluídos juntamente com seus apelidos, tais como Xandão, Xandao, e Alexandre de Moraes. A coleta foi realizada utilizando a API do Twitter (©2023)¹.

Quadro 1. Dados gerais da coleta

Termos Gerais	TSE, Tribunal Superior Eleitoral, urna eletrônica, urna eletrônica, fraude na urna, voto impresso, voto auditável, voto auditavel, tsejusbr
Nomes dos Ministros do TSE e STF	Luís Roberto Barroso, Luís Barroso, Luis Barroso, Fachin, Alexandre de Moraes, Alexandre de Moraes, Xandão, Xandao, Mauro Luiz Campbell Marques, Mauro Campbell, Benedito Gonçalves, Benedito Goncalves, Sérgio Silveira Banhos, Sérgio Banhos, Sergio Banhos, Carlos Bastide Horbach, Horbach, Ricardo Lewandowski, Cármen Lúcia, Carmen Lucia, Raul Araújo, Raul Araujo, Nunes Marques, André Mendonça, Andre Mendonca, Paulo Sanseverino, Maria Cláudia Bucchianeri, Maria Claudia Bucchianeri, Maria Bucchianeri
Hashtags	votoimpresso, votoimpressoauditaveljá, urnaeletronica, fraudenasurnas, fraudeeleitoral, alexandredemoraes, foraalexandredemoraes, EuApoioVotoAuditavel, UrnasForamInvasadas, Brasilpelovotoauditavel
Período de coleta	18-07-2022 a 22-08-2022
Quantidade total de <i>tweets</i>	2.881.263

Fonte: Elaboração própria

¹Disponível em: <https://developer.twitter.com/en/docs/twitter-api>. Acesso em: 24 mar. 2023.

A análise de dados foi feita em três níveis: (i) identificação de posicionamento de clusters polarizados de usuários mais vocais (conforme definição na subseção a seguir), (ii) análise de hostilidade do conteúdo compartilhado por cada cluster, e uma (iii) análise qualitativa do conteúdo das publicações com ataques às instituições e ao sistema democrático.

DETECÇÃO DE POSICIONAMENTO DE CLUSTERS POLARIZADOS

A detecção de posicionamento visa determinar de forma automática se o autor de um texto é favorável, contrário ou neutro em relação a um tópico ou proposição (MOHAMMAD, SOBHANI e KIRITCHENKO, 2017), importante técnica para estudos sobre debates políticos que envolvem necessariamente grupos em conflitos e com posicionamentos diferentes.

Para responder à pergunta de pesquisa sobre qual o posicionamento político dos perfis que publicam conteúdo tóxico, foi aplicado um método de detecção de posicionamento não supervisionado que busca agrupar os usuários em clusters polarizados a partir das contas de usuários que eles retuitam.

A fundamentação dessa abordagem parte de estudos que sugerem que os usuários da web tendem a polarizar sua opinião (CONOVER et al., 2011) e a formar comunidades a partir de afinidades políticas, seguindo o princípio da homofilia discutido por McPherson, Smith-Lovin e Cook. (2001).

Os três passos a seguir foram realizados com base em estudo anterior conduzido por Santos e Goya (2022), que adapta parte da técnica desenvolvida por Darwish et al. (2020):

1. Seleção de características: foi calculado o quartil da quantidade de retweets por usuário, com o objetivo de selecionar 25% dos valores mais elevados. Desse modo, foi possível selecionar os usuários mais vocais, ou seja, aqueles com maior participação na discussão em cada base. Após isso, foi construída uma representação algébrica (vetor de categorias) para cada usuário selecionado, com base nas contas que ele retuitou, e calculada a proximidade entre cada par de usuários (similaridade do cosseno).
2. Redução de dimensionalidade: após o cálculo da similaridade, os usuários foram projetados em um espaço bidimensional utilizando o algoritmo UMAP

(uniform manifold approximation and projection), com o objetivo de identificar de forma objetiva a posição relativa entre os usuários.

3. Clusterização: feita a projeção dos usuários em um espaço bidimensional, eles foram agrupados utilizando o algoritmo mean shift. Isso permitiu a obtenção de dois grandes clusters polarizados.

ANÁLISE DE HOSTILIDADE DO CONTEÚDO COMPARTILHADO

Para inferir a hostilidade dos textos compartilhados pelos usuários pertencentes aos clusters polarizados, foi utilizada a Perspective API. Essa ferramenta utiliza uma escala de 0 a 1 para indicar a probabilidade de um comentário ser tóxico. Adicionalmente, fornece métricas que apontam a probabilidade de um texto conter especificamente, por exemplo, ameaça, insulto ou ataque identitário, conforme descrição do Quadro 2. Segundo Guimarães et al. (2020), a maioria dos resultados dessas métricas “[...] está altamente correlacionada entre si (o coeficiente de correlação de Pearson maior que 0,8 ($p < 0,05$))”, fato confirmado neste presente estudo por observação direta das postagens que receberam pontuação maior do que 0,9 nos índices indicados no Quadro 2. Por esse motivo, optou-se por empregar o termo “hostilidade” para referenciar genericamente qualquer desses tipos, e “toxicidade” quando se analisa especificamente os dados relacionados com a métrica de toxicidade.

Quadro 2. Dados gerais da coleta

Métricas	Descrição
TOXICIDADE	Um comentário rude, desrespeitoso ou irracional que provavelmente fará com que as pessoas abandonem uma discussão.
TOXICIDADE SEVERA	Esse atributo é uma versão da toxicidade menos sensível a formas mais leves de toxicidade, como comentários que incluem usos positivos de palavras.
ATAQUE IDENTITÁRIO	Comentários negativos ou de ódio direcionados a alguém por causa de sua identidade.
INSULTO	Comentário ofensivo, inflamatório ou negativo direcionado a um indivíduo ou grupo.
PROFANIDADE	Palavras, xingamentos ou outra linguagem obscena ou profana.
AMEAÇA	Descreve a intenção de infligir dor, lesão ou violência contra um indivíduo ou grupo.

Fonte: Adaptado de Perspective (©2023), tradução nossa.

ANÁLISE QUALITATIVA DE CONTEÚDO

Quatro codificadores aplicaram manualmente a análise de conteúdo às mensagens finais selecionadas para classificar suas características gerais e identificar ataques às instituições e ao sistema eleitoral brasileiro. A categorização apresentada pode ser suscetível a preconceitos humanos, o que os autores reconheceram e tentaram mitigar ao longo das etapas de brainstorming, iteração e anotação final. A categorização foi refinada com base nas discussões entre os autores. Algumas das anotações foram alteradas e outras foram mantidas com base na maioria dos votos na iteração final.

RESULTADOS E DISCUSSÃO

Foram analisados perto de 2,3 milhões de retweets (cerca de 79,78% dos tweets coletados), compartilhados por cerca de 175 mil usuários. No primeiro nível de análise dos dados, com o emprego da técnica de detecção automática de posicionamento, foi possível identificar e rotular cerca de 16,5% desses usuários, que provaram ser os mais engajados. Tais dados são reportados na Tabela 1. É importante destacar que esses usuários rotulados foram responsáveis por cerca de 71,1% dos retweets feitos. Esses números indicam que, apesar de os usuários vocais serem minoria, são eles que movimentam e participam ativamente das discussões.

Tabela 1. Dados gerais da coleta

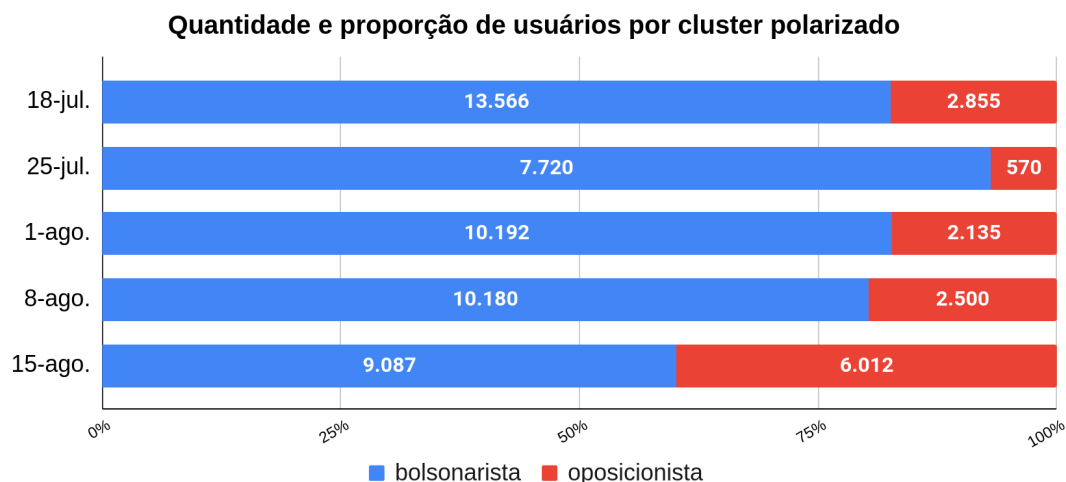
Dado	Quantidade	%
Número de usuários únicos	175.591	100%
Número de usuários únicos rotulados	29.039	16,5%
Total de <i>retweets</i>	2.298.789	100%
Total de <i>retweets</i> realizados por usuários rotulados	1.634.946	71,1%

Fonte: Elaboração própria.

Em todas as cinco semanas de coleta, formaram-se dois agrupamentos fortemente polarizados. O primeiro, composto por perfis de apoiadores do governo Bolsonaro, foi rotulado como "bolsonarista". O segundo grupo, por promover o confronto oposicionista, foi nomeado "oposicionista". Observa-se pela Figura 1 que o primeiro grupo foi majoritário em todas as semanas e, em quatro delas, superou os 75% de prevalência. Apenas na semana de 15 de agosto,

há um expressivo crescimento do cluster oposicionista. Isso ocorreu devido à forte repercussão entre os usuários do Twitter em torno da cerimônia de posse do ministro Alexandre de Moraes à presidência do TSE.

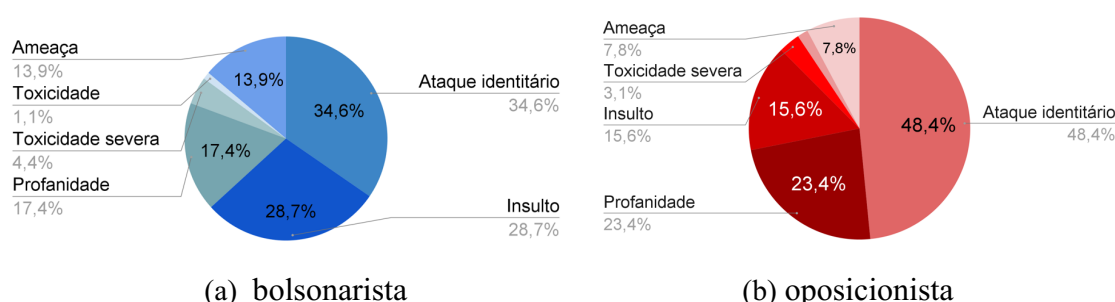
Figura 1. Distribuição dos usuários mais vocais por cluster polarizado nas cinco semanas analisadas



Fonte: Elaboração própria

O segundo nível de análise realizado envolveu o processamento computacional para classificação dos retweets com a ferramenta Perspective, seguido de uma verificação manual pelos quatro pesquisadores para selecionar somente as postagens que estavam relacionadas com o tema e que continham teor hostil.

Para responder à questão de pesquisa (P1) Quais os principais tipos de conteúdos hostis?, foram analisados os retweets que receberam pontuação de toxicidade superior a 0,9 em ao menos uma das seis categorias da ferramenta Perspective selecionadas (ver Quadro 2). Por meio da Figura 2, pode-se perceber que, tanto no cluster bolsonarista quanto no oposicionista, “ataque identitário” foi a métrica que se mostrou mais presente nos retweets; entretanto, os clusters apresentaram diferenças na segunda métrica mais frequente: no cluster bolsonarista (Figura 2a) “insulto” esteve em 28,7% dos retweets, seguido por “profanidade” com 17,4%. No cluster oposicionista (Figura 2b), “profanidade” esteve em 23,4% dos retweets, seguido por “insulto”, com 15,6%.

Figura 2. Distribuição dos tipos de hostilidade identificados nos retweets extraídos de cada cluster

Fonte: Elaboração própria

Alguns exemplos de textos compartilhados por usuários de um dos clusters e que obtiveram pelo menos 0,9 em ao menos uma das métricas da Perspective (ou seja, 90% de probabilidade de ser conteúdo tóxico) podem ser vistos no Quadro 3.

Quadro 3. Exemplos de conteúdos hostis e algumas métricas da Perspective API

Texto	Métrica
Não acreditamos neste supremo o tse que se dane com esta conversinha de merda é só por urnas auditáveis e tá tudo certo	Profanidade 0,96
@RodrigoMaia Tão comprovado que um hacker invadiu a porra toda por 8 meses usando a senha de um ministro. Tão comprovado que o hacker está preso e a PF abriu inquérito a pedido do TSE. Você é idiota ou negacionista?	Insulto 0,96
"@TSEjusbr 🗨️👤🔪 Seus Ditadores ministros do @STF_oficial Vem aqui na minha cidade q o Povo sabe dar Facada. Cuidado! @alexandre @TSEjusbr @LRobertoBarroso @gilmarmendes @MinAMendonca @liberal22_rs https://t.co/dYuaQF9ach "	Ameaça 0,93
ATENÇÃO: DENTRO DO TSE JÁ FALAM ABERTAMENTE QUE IRÃO IMPUGNAR A CANDIDATURA DO NOSSO PRESIDENTE, MORAES CABEÇA DE PIROCA NEM ESCONDE SUAS MÁS INTENÇÕES, DIA 7SETEMBRO O POVO TEM QUE LOTAR AS RUAS DO BRASIL, SUPREMO É O POVO CARALHO!!!	Profanidade 0,94
E o TSE, o mesmo que deu de ombros pras denúncias da campanha bolsonarista de 2018, segue oferecendo condições pra levar uma facada, ou tiro, nas próprias costas... "Brasil 247: TSE avalia ceder a militares para aliviar tensão às vésperas da eleição." https://t.co/ChfpNvMAUQ	Ameaça 0,91
@TSEjusbr No Japão as privadas são eletrônicas mas as urnas não. No Brasil 60% das residências não tem esgoto mas as urnas são infalíveis... 🤔🤔🤔 Só um idiota para acreditar 🤡 #STFVergonhaNacional	Insulto 0,91
Vamos destruir esse braço da esquerda instalado no pt\$tfcc e no t\$é... ! BRASIL UNIDO JAMAIS SERÁ VENCIDO! CHEGA DE ATURAR ESSA CORJA DE COMUNISTAS... General Paulo Sérgio dá ultimato no TSE	Ataque Identitário 0,95

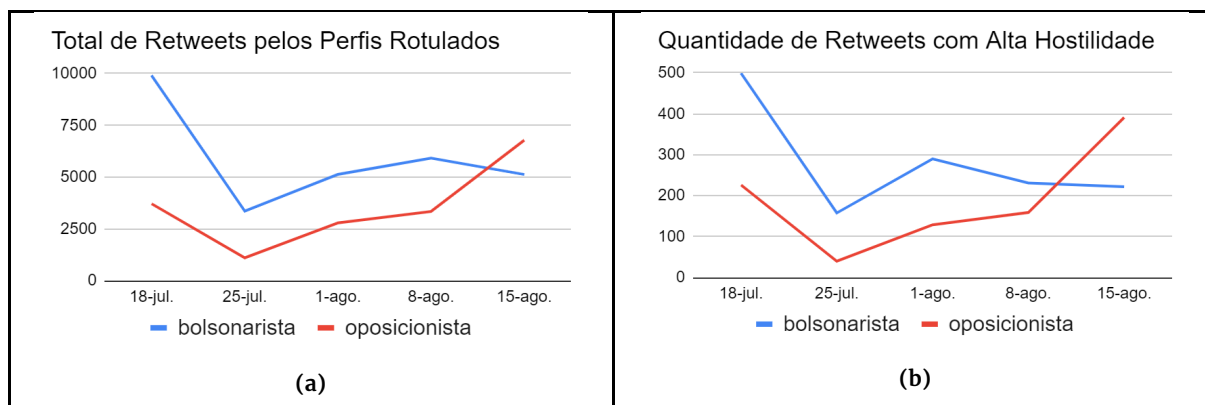
Texto	Métrica
para que Forças Armadas tenham acesso ao código-fonte das urnas	
Tribunal Superior Eleitoral reduz acesso a informações sobre bens de candidatos! Qual é o propósito disso a não ser os urubus de toga protegerem os filhos da puta que militam na política? O tribunal é sócio de picaretas que emporalham o Brasil. Ministros são igualmente bandidos!	Ataque Identitário 0,94
Minha gente !!! Nunca vimos dois ministros STF tão escrachados como Nunes Marques e Cássio com k !!! Os caras são escrotos , bandidos nojentos, cães a serviço do crime !!!	Insulto 0,95
Não dá pra confiar, são todos canalhas ! Xandão e outros ministros do STF, vão ao encontro de Bolsonaro, pra convidá-lo pra posse de ministro. São covardes, bajuladores de miliciano bandido, são farinha do mesmo saco. Canalhas canalhas canalhas.	Ataque Identitário 0,92

Fonte: Elaboração própria

Na Figura 3a, é possível observar a quantidade de retweets produzidos pelos perfis mais vocais, que puderam ser rotulados como pertencentes a um dos dois clusters. E, na Figura 3b, um subconjunto desses retweets que alcançaram pontuação alta para hostilidade (maior que 0,9 em alguma das métricas). Dentre os retweets que circularam no cluster bolsonarista, em média 4,72% foram considerados muito hostis, frente a 4,97% no cluster oposicionista. Pela Figura 3, nota-se que os perfis do cluster bolsonarista foram os que produziram maior volume de retweets e de retweets com teor hostil. No total das cinco semanas, o grupo bolsonarista foi responsável por cerca de 59,7% dos retweets considerados tóxicos, contra 40,3% do grupo opositor.

Durante a semana de 15 de agosto, ocorreu manifestação maior entre os perfis de opositoristas, por conta da posse do ministro Alexandre de Moraes como presidente do TSE, o que gerou uma forte onda de comentários hostis contra o presidente Bolsonaro por parte desse cluster.

Figura 3. Quantidade de retweets: (a) dos usuários vocais (b) com alta hostilidade

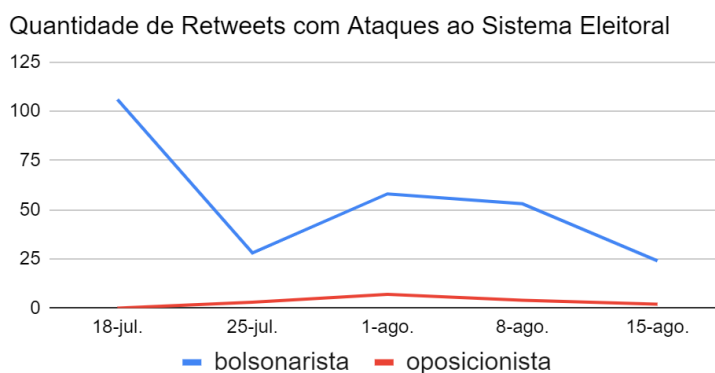


Fonte: Elaboração própria

ATAQUES AO SISTEMA ELEITORAL

O terceiro e último nível de análise teve por objetivo fazer um estudo em relação aos tweets tóxicos com ataques ao sistema democrático, mais especificamente ao sistema eleitoral. Foram selecionados os tweets com hostilidade acima de 0,9 e que contivessem uma das seguintes palavras-chave: “TSE”, “STF”, “ministro”, “urna”, “voto impresso”, “sistema eleitoral”, “eletrônicas”, “auditável”, “urnas”, “eletrônica”, “ministros”, “supremo”, “votos”, “forças armadas”. O volume de tais mensagens ao longo das semanas e em cada agrupamento pode ser observado na Figura 4, em que se torna evidente a ocorrência majoritária de mensagens propagadas por perfis do cluster bolsonarista.

Figura 4. Volume de mensagens que continham ataques ao sistema eleitoral



Fonte: Elaboração própria

Para apoiar uma análise qualitativa sobre as postagens com ataques ao sistema democrático, adotou-se duas estratégias: (i) verificação da ocorrência de termos mais frequentes em nuvens de palavras e (ii) identificação dos assuntos predominantes e suas palavras-chave. Antes de proceder com ambas as estratégias, aos textos dos retweets selecionados, foi aplicado um filtro para remover preposições e palavras que revelam baixo significado (de, para, com, está, etc), pontuações, hashtags, links/urls contidos no texto, nomes de contas e caracteres especiais, a fim de obtermos visualizações mais assertivas dos conteúdos compartilhados pelos usuários.

Na Figura 5, pode-se visualizar as palavras mais frequentes usadas em ataques por cluster polarizado, onde as palavras que estão com o tamanho maior são as que apareceram com maior frequência. Nota-se que há um padrão dos ataques do cluster bolsonarista (Figura 5a) com alvos bem determinados, direcionados às urnas, aos ministros Alexandre de Moraes e Luís Roberto Barroso (representado

pela expressão “boca de veludo”) e a alguns partidos de esquerda. Já o cluster oposicionista (Figura 5b) apresenta uma dispersão bem maior em relação às palavras mais frequentes nos ataques que o cluster anterior, contendo críticas ao atual presidente Bolsonaro (representado pela palavra “genocida”) e aos ministros Nunes Marques e André Mendonça (representado pela expressão “terrivelmente evangélico”), indicados pelo político.

Foram geradas visualizações das palavras mais frequentes em cada uma das semanas, mas não foi identificada uma mudança expressiva de comportamento em cada cluster ao longo do tempo. Assim, levantou-se a hipótese de que cada grupo de usuários preserva uma característica ou estratégia comunicacional (por exemplo, mais organizada ou coordenada no cluster bolsonarista, e mais orgânica e reativa no cluster oposicionista). Optou-se, então, por exibir, na Figura 5, a representação global das discussões ocorridas nas cinco semanas coletadas.

Figura 5. Nuvem de palavras mais frequentes nos ataques de cada cluster



Fonte: Elaboração própria

No Quadro 4, são apresentadas postagens selecionadas e relacionadas a cada palavra mais frequente encontrada nas nuvens de palavras (Figura 5). No cluster bolsonarista foram selecionadas as expressões: “voto impresso”, “alexandre moraes”, “boca veludo” e “urnas eletrônicas”. E, no cluster oposicionista, foram selecionadas as palavras: “genocida referir”, “nunes marques”, “terrivelmente evangélico” e “stf não”.

Quadro 4. Exemplos de conteúdos hostis com ataque ao sistema eleitoral, por cluster

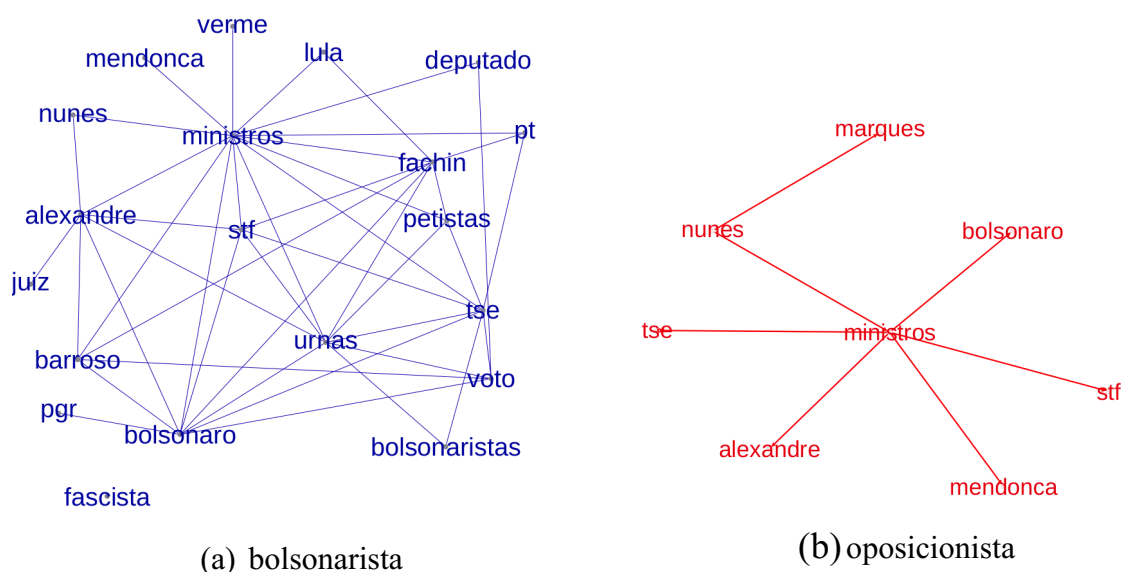
Texto	Cluster
As URNAS não são confiáveis. O STF e TSE foram aparelhados pelo PT e o TWITTER CENSURA conservadores. Mais de um mês bloqueado sem poder seguir ninguém, por dizer que as URNAS eletrônicas são passíveis de FRAUDES. Sem engajamento. Podem cancelar essa porraaaaa. Cansei	Bolsonarista

Texto	Cluster
Esse safado desse Alexandre de Moraes era pra estar preso e não ocupando uma cadeira no STF....esse supremo é o cancer qe não deixar o Brasil progredir!	Bolsonarista
O boca de veludo ataca novamente, Ministro Barroso, além de nos chamar de imbecis e incivilizados, agora, fascistas!? Por estar ministro do STF, ele acha que tem o direito de ofender milhões de brasileiros?? 7 de setembro vai ser grande vamos quita esta fatura no primeiro turno.	Bolsonarista
Será que eles são burros os se fazem? Ele fala como se quisemos o voto em cédula, nós queremos o voto impresso, é diferente. Porque é tão difícil esse povo entender que transparência nunca é demais	Bolsonarista
o ministro Nunes Marques revogou a condenação dele o colocou nas ruas esse é o ministro terrivelmente evangélico que o bolsonaro colocou no stf e não só ele esse vagabundo desse ministro Nunes Marques anda colocando nas ruas más bandidos vagabundos	Oposicionista
Nunes Marques já devolveu cargos aos conselheiros do TCE do Rio envolvidos em corrupção, já liberou bicheiro mafioso, agora libera o Arruda. É esse tipo de ministro que os Bolsonazistas querem no STF?	Oposicionista
O mesmo ministro do TSE bolsominion que tentou censurar o Lollapalooza com 1 decisão ILEGAL agora censura Lula proibindo que ele use a palavra GENOCIDA pra se referir ao GENOCIDA. Entao já sabe, não pode falar GENOCIDA pra se referir a quem ajudou a matar 700 mil pessoas!	Oposicionista
STF não é independente? Então por que raios um de seus ministros ameaça "romper" com o Executivo? Que porra é essa, Banalândia?	Oposicionista

Fonte: Elaboração própria

Dados de redes sociais podem ser visualizados por meio de grafos, com nós representando perfis ou conteúdos, e arestas que os ligam para representar suas interações ou associações (PHYU e MIN, 2019). Na Figura 6, é possível observar as associações (arestas) entre os assuntos abordados (nós) mais recorrentes durante o período analisado, em cada agrupamento polarizado.

Percebe-se que, apesar de o cluster bolsonarista ter investido forte intensidade sobre os ataques centrados em “urnas”, “alexandre de Moraes” e “voto impresso” (Figura 5a), houve outros assuntos relevantes, porém em menor volume de mensagens compartilhadas, como, por exemplo, os associados a hostilidades ao ex-presidente Lula, ao seu partido, PT, ou a deputados (Figura 6a), relacionando-os aos ministros e ao sistema eleitoral. No grafo nota-se que alguns assuntos estão conectados, por exemplo: retweets que falaram sobre as urnas podem citar TSE, STF, o atual presidente Bolsonaro e o ministro Luís Roberto Barroso (Fig. 6a). Ao se observar conjuntamente as Figuras 5a e 6a, é possível identificar, de forma geral, confrontação às instituições democráticas brasileiras e aos seus representantes. Um ponto interessante a se destacar na Figura 6a é o nó isolado “fascista”, termo muito usado por opositores para se referir a bolsonaristas, que apareceu em alguns retweets do cluster bolsonarista como forma de ironia.

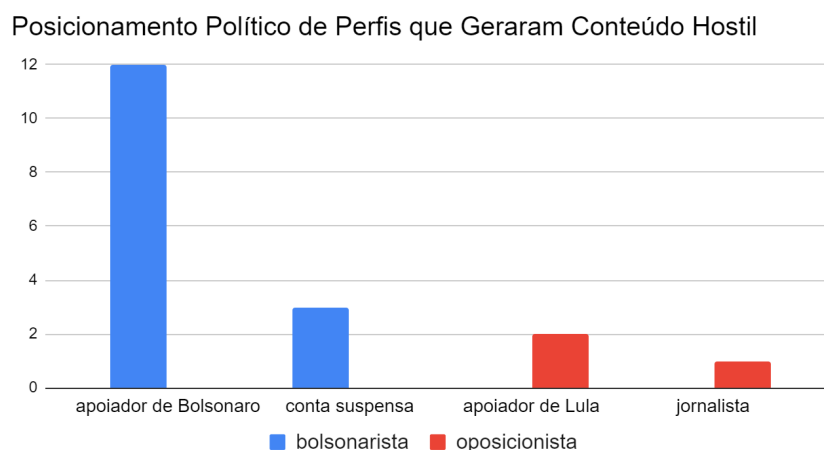
Figura 6. Rede dos assuntos contidos nos conteúdos hostis por cluster polarizado

Fonte: Elaboração própria

Já no cluster oposicionista, os assuntos associados aos conteúdos hostis (Figura 6b) condizem com as palavras mais frequentes apresentadas na Figura 5b. Por apresentar um grafo mais simples, esse cluster demonstra ter uma narrativa mais dispersa com relação à hostilidade, em que poucos termos são recorrentes (os da Figura 5b).

Para responder à questão de pesquisa (P2) Qual o posicionamento político dos perfis que publicam conteúdos hostis, foram selecionados usuários que tiveram seus tweets compartilhados por usuários rotulados em pelo menos duas semanas de coleta. Foram encontrados 18 perfis nessas condições. Desses, 15 foram compartilhados por usuários do cluster bolsonarista, e três do cluster oposicionista. Após analisar os perfis dos usuários bolsonaristas, verificou-se que três contas haviam sido suspensas, possivelmente por violarem regras da plataforma, e, das 12 contas restantes, todas traziam mensagens de apoio ao governo ou elementos comuns do bolsonarismo, tais como o uso das cores da bandeira nacional, fotos do presidente Jair Bolsonaro e descrições clássicas da extrema direita, como “cristão”, “conservador”, “pró-armas” etc. No caso dos três perfis oposicionistas, dois deles declararam apoio à candidatura do ex-presidente Lula, e um era de uma jornalista. Esses dados são representados na Figura 7. É interessante destacar que a maioria desses perfis é de pessoas relativamente populares, possuindo dezenas de milhares de seguidores; o mais popular possui mais de 300 mil seguidores, e o menos popular, cerca de 660.

Figura 7. Posicionamento e quantidade de perfis que produziram conteúdo hostil que foi compartilhado por usuários vocais em pelo menos duas semanas



Fonte: Elaboração própria

CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma análise sobre conteúdos hostis disseminados no Twitter, entre julho e agosto de 2022, relacionados à credibilidade do sistema eleitoral brasileiro. Por meio de uma abordagem interdisciplinar que uniu técnicas da ciência da computação e da comunicação política, foram coletados e processados cerca de 2,8 milhões de tweets que tratavam do tema, com foco na identificação de conteúdos tóxicos e dos perfis que se engajaram neles.

Com base no princípio de homofilia e de formação de conglomerados por afinidade política ou ideológica, os perfis mais participativos nas discussões puderam ser agrupados em dois clusters antagônicos que, por inspeção humana, foram identificados como de posicionamento bolsonarista ou oposicionista e foram assim rotulados. Aproximadamente 16,5% de todos os usuários coletados puderam ser rotulados dessa forma. Essa relativa minoria de perfis foi responsável por mobilizar cerca de 71% do total dos retuítés e, também por esse motivo, foi aqui adjetivada como usuários vocais. Houve predominância de perfis bolsonaristas e de suas mensagens no somatório das semanas.

Apesar de não ser um dos objetivos deste artigo, esses dados representam um importante achado para o campo de estudos sobre o debate público no Twitter, pois indicam que, apesar de haver a possibilidade de uma ampla participação dos usuários na produção de conteúdos, somente uma pequena parte participa desse

processo, sendo responsável pelos sentidos mobilizados na conversação. No caso estudado, sobre o sistema eleitoral brasileiro, o debate teve maior participação de perfis apoiadores de Bolsonaro.

Com o objetivo de identificar e categorizar os tipos de conteúdos hostis compartilhados na rede social, adotou-se uma ferramenta computacional para processamento de linguagem natural e aprendizado de máquina, a Perspective API, que mensura a probabilidade de cada postagem retuitada conter hostilidade. Por meio de inspeção humana, quatro pesquisadores convergiram para a seleção de aproximadamente 5% do total de retuítes selecionados que foram considerados como tóxicos, dos quais cerca de 59,7% circularam entre apoiadores do bolsonarismo, enquanto cerca de 40,3% no grupo opositor. A maior parcela dos conteúdos hostis foi qualificada como xingamentos, ataques à identidade de ministros ou comentários ofensivos sobre as urnas.

O estudo também identificou o posicionamento político dos perfis que publicaram conteúdos hostis que tiveram repercussão entre os usuários rotulados, fato repetido em pelo menos duas semanas. Prevaleram os perfis reconhecidos como de apoiadores de Bolsonaro (12 em 18, ou 66,7%), seguidos por partidários de Lula (2 em 18, ou 11,1%) e uma jornalista rotulada como oposição.

Como trabalhos futuros, pretende-se ampliar a análise para um período maior de coleta e investigar quais são as características topológicas das redes formadas, para analisar comparativamente propriedades relativas à organização comunicacional e sua robustez para ambos agrupamentos político-ideológicos. A metodologia empregada também pode ser estendida para investigar, dentro da base de dados coletada, a existência de uma possível relação entre produção de conteúdo hostil e disseminação de desinformação.

SOBRE OS AUTORES

Patrícia Dias dos Santos: Doutoranda em Ciência da Computação, mestre em Engenharia da Informação e bacharel em Ciência da Computação pela Universidade Federal do ABC (UFABC).

Claudio Luis de Camargo Penteado: Mestre e doutor em Ciências Sociais pela PUC/SP. Professor do Programa de Pós-Graduação em Ciências Humanas e Sociais e do Bacharelado em Políticas Públicas da Universidade Federal do ABC. Pesquisador do Laboratório de Tecnologias Livres (LabLivres/UFABC), Núcleo DATAS/UFABC e do NEAMP/PUCSP.

Laura Damaceno de Almeida: Mestranda em Ciência da Computação pela Universidade Federal do ABC (UFABC). Cientista de dados na Bayer e graduada em Ciência da Computação pela Universidade Anhembi Morumbi (UAM), em São Paulo.

Denise Hideko Goya: Mestre e doutora em Ciência da Computação pela Universidade de São Paulo. Professora do Programa de Pós-Graduação e do Bacharelado em Ciência da Computação da Universidade Federal do ABC. Pesquisadora do Laboratório de Informações em Rede e Tecnologias Educacionais (Lirte/UFABC) e Núcleo DATAS/UFABC.

REFERÊNCIAS

1. CASTELLS, Manuel. Ruptura: a crise da democracia liberal. Rio de Janeiro: Zahar, 2018.
2. CASTELLS, Manuel. Redes de indignação e esperança: movimentos sociais na era da internet. São Paulo: Schwarcz; Companhia das Letras, 2017.
3. CONOVER, Michael et al. Political polarization on Twitter. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 2011. Proceedings of the International AAAI Conference on Web and Social Media. 2011. p. 89-96.
4. DARWISH, Kareem et al. Unsupervised user stance detection on Twitter. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 2011. Proceedings of the International AAAI Conference on Web and Social Media. 2011. p. 141-152.
5. ELSHERIEF, Mai et al. Hate lingo: A target-based linguistic analysis of hate speech in social media. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 2018. Proceedings of the International AAAI Conference on Web and Social Media. 2018.
6. FORTUNA, Paula; NUNES, Sérgio. A survey on automatic detection of hate speech in text. ACM Computing Surveys (CSUR), v. 51, n. 4, p. 1-30, 2018. DOI: <https://doi.org/10.1145/3232676>
7. FORTUNA, Paula; SOLER, Juan; WANNER, Leo. Toxic, hateful, offensive or abusive? what are we really classifying? an empirical analysis of hate speech datasets. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 12., 2020. Proceedings of the 12th language resources and evaluation conference. 2020. p. 6786-6794.
8. GUIMARÃES, Samuel S. et al. Characterizing toxicity on facebook comments in Brazil. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 2020. Proceedings of the Brazilian Symposium on Multimedia and the Web. 2020. p. 253-260.
9. INSTITUTO DA DEMOCRACIA E DA DEMOCRATIZAÇÃO DA COMUNICAÇÃO (IDDC). A cara da Democracia 2019. 2019. Disponível em: https://docs.wixstatic.com/ugd/a46f9a_05967934746d4ba2b0ef032921bde80c.pdf. Acesso em 22 mar. 2023.
10. IYENGAR, Shanto et al. The origins and consequences of affective polarization in the United States. Annual Review of Political Science, v. 22, p. 129-146, 2019. DOI: <https://doi.org/10.1146/annurev-polisci-051117-073034>
11. LEVITSKY, Steven; ZIBLATT, Daniel. Como as democracias morrem. São Paulo: Schwarcz; Companhia das Letras, 2018.
12. MCPHERSON, Miller; SMITH-LOVIN, Lynn; COOK, James M. Birds of a feather: homophily in social networks. Annual Review of Sociology, v. 27, n. 1, p. 415-444, 2001. DOI: <https://doi.org/10.1146/annurev.soc.27.1.415>
13. MERCURI, Karen Tank; LIMA-LOPES, Rodrigo Esteves de. Discurso de ódio em mídias sociais como estratégia de persuasão popular. Trabalhos em Linguística Aplicada, v. 59, p. 1216-1238, 2020. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/tla/article/view/8658475>. Acesso em: 20 mar. 2023.
14. MOHAMMAD, Saif M.; SOBHANI, Parinaz; KIRITCHENKO, Svetlana. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), v. 17, n. 3, p. 1-23, 2017. DOI: <https://doi.org/10.1145/3003433>
15. NOBRE, Marcos. Ponto-final: a guerra de Bolsonaro contra a democracia. São Paulo: Todavia, 2020.
16. PASCUAL-FERRÁ, Paola et al. Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. Big Data & Society, v. 8, n. 1, 2021. DOI: <https://doi.org/10.1177/20539517211023533>
17. PENTEADO, Claudio Luis; GOYA, Denise Hideko; DE FRANÇA, Fabrício Olivetti. Discursive conflicts around the impeachment of Dilma Rousseff (Brazil) on Twitter. Perspectivas da Contemporaneidade, n. 1, 2021. Disponível em: <http://www.perspectivas.periodikos.com.br/article/600715bc0e88256c7652383e/pdf/perspectivas-1-1-49.pdf>. Acesso em: 28 mar. 2023.
18. PERSPECTIVE. Attributes & Languages. ©2023. Disponível em: https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US. Acesso em: 18 mar. 2023.

19. PERSPECTIVE. How it works. ©2021. Disponível em: <https://perspectiveapi.com/how-it-works/>. Acesso em: 18 mar. 2023.
20. PHYU, Kaing Sabai; MIN, Myat Myat. Graph-based community detection in social network. In: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCE (ICIS), 18., 2019. Proceedings of the IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS). IEEE, 2019. p. 12-17.
21. RASMUSSEN, Terje. Internet-based media, Europe and the political public sphere. *Media, Culture & Society*, v. 35, n. 1, p. 97-104, 2013. DOI: <https://doi.org/10.1177/0163443712464563>
22. RECUERO, Raquel. #FraudenasUrnas: estratégias discursivas de desinformação no Twitter nas eleições 2018. *Revista brasileira de linguística aplicada*, v. 20, p. 383-406, 2020.
23. RIEBE, Thea et al. From conspiracies to insults: a case study of radicalisation in social media discourse. In: DACHSELT, Raimund; WEBER, Gerhard. *Mensch und Computer 2018 –Workshopband*. Bonn: Gesellschaft für Informatik, 2018. DOI: <http://doi.org/10.18420/muc2018-ws12-0449>
24. ROSSINI, Patrícia. Beyond toxicity in the online public sphere: understanding incivility in online political talk. In: DUTTON, William. *A research agenda for digital politics*. Elgar Research Agendas, 2020. p. 160-170.
25. RUEDIGER, Marco A.; GRASSI, Amaro (Coords.). *Desinformação on-line e processos políticos: a circulação de links sobre desconfiança no sistema eleitoral brasileiro no Facebook e no YouTube (2014-2020)*. Policy paper. Rio de Janeiro: FGV DAPP, 2020.
26. SANTOS, Patricia D.; GOYA, Denise H. Detecção de posicionamento e rotulação automática de usuários do Twitter: estudo sobre o embate científico-político no contexto da CPI da covid-19. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 11., 2022. *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*. SBC, 2022. p. 49-60.
27. SIEGEL, Alexandra A. et al. Trumping hate on Twitter? Online hate speech in the 2016 US election campaign and its aftermath. *Quarterly Journal of Political Science*, v. 16, n. 1, p. 71-104, 2021. DOI: <https://doi.org/10.1561/100.00019045>
28. SUNSTEIN, Cass R. *#Republic: divided democracy in the age of social media*. Princeton University Press, 2018.
29. TRAJKOVA, Zorica; NESHKOVSKA, Silvana. Online hate propaganda during election period: the case of Macedonia. *Lodz Papers in Pragmatics*, v. 14, n. 2, p. 309-334, 2018. DOI: <https://doi.org/10.1515/lpp-2018-0015>
30. TWITTER. Twitter API. ©2023. Disponível em: <https://developer.twitter.com/en/docs/twitter-api>. Acesso em: 24 mar. 2023.
31. VINHAS, Otávio; SAINZ, Nilton; RECUERO, Raquel. Antagonismos discursivos nas hashtags #marqueteirosdojair e #bolsolão no Twitter nas eleições de 2018 no Brasil: contribuições da análise de redes sociais à sociologia digital. *Estudos de Sociologia*, v. 25, n. 48, 2020. DOI: <https://doi.org/10.52780/res.13433>
32. WULCZYN, Ellery; THAIN, Nithum; DIXON, Lucas. Ex machina: personal attacks seen at scale. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 26., 2017. Proceedings of the 26th International Conference on World Wide Web. 2017. p. 1391-1399.
33. YIN, Wenjie; ZUBIAGA, Arkaitz. Towards generalisable hate speech detection: a review on obstacles and solutions. *PeerJ Computer Science*, v. 7, p. e598, 2021. Disponível em: <https://peerj.com/articles/cs-598/>. Acesso em: 20 mar. 2023.

Recebido em 31 de janeiro de 2023.
Aprovado em 23 de fevereiro de 2023.

