# Das milícias digitais ao comportamento coordenado: métodos interdisciplinares de análise e identificação de *bots* nas eleições brasileiras

**João Guilherme Bastos dos Santos[1], Arthur Ituassu[2], Sérgio Lifschitz[2], Thayane Guimarães[3], Diego Cerqueira[3], Debora Albu[3], Redson Fernando[3], Julia Hellen Ferreira[3], Maria Luiza Mondelli[3,4]**

[1]Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT.DD)
Salvador, BA – Brasil

[2]Departamento de Comunicação Social e Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ)
Rio de Janeiro, RJ – Brasil

[3]Instituto de Tecnologia e Sociedade (ITS Rio)
Rio de Janeiro, RJ – Brasil

[4]Laboratório Nacional de Computação Científica (LNCC)
Petrópolis, RJ – Brasil

santos.jgb@gmail.com,ituassu@puc-rio.br,demotech@itsrio.org

***Abstract.*** *To support automated behavior detection on social networks in the Brazilian scenario, this work applies five methods of bot identification - analysis of (i) networks, (ii) sentiment, (iii) profile, (iv) temporal, and (v) crossing of these analyses - to possible bots in the comments of tweets from candidates for the mayors of the capitals Rio de Janeiro, São Paulo, Recife, Porto Alegre and Fortaleza in 2020. After identification, we followed two analysis fronts (namely, networks and lexical), to identify clusters of bots and similarity of vocabularies used by them, as well as different types of bots. Among the preliminary results obtained, we highlight the multiplicity of types of automation, understanding of regional variations and the role of bots in the Brazilian electoral scenario.*

***Resumo.*** *Para apoiar a detecção de comportamento automatizado em redes sociais no cenário brasileiro, o presente trabalho aplica cinco métodos de identificação de bots - análises de (i) redes, (ii) sentimento, (iii) perfil, (iv) temporal e (v) cruzamento dessas análises - a possíveis bots nos comentários de tweets dos candidatos à prefeitura das capitais Rio de Janeiro, São Paulo, Recife, Porto Alegre e Fortaleza em 2020. Após identificação, seguimos duas frentes de análise (a saber, de redes e lexical), para identificação de clusters de bots e similaridade de vocabulários utilizados por eles, bem como diferentes tipos de bots. Dentre os resultados preliminares obtidos, destacamos a multiplicidade de tipos de automatização, compreensão das variações regionais e atuação de bots no cenário eleitoral brasileiro.*

## 1. Introdução

O papel das plataformas de redes sociais no fomento a democracia é paradoxal: muitas vezes utilizadas como um canal que viabiliza, por exemplo, maior participação popular

nos processos e no debate político, elas também tem sido utilizadas como forma de influenciar a opinião pública [Almeida et al. 2020] em direção a pautas pouco democráticas. A atuação de contas com comportamento automatizado, os *bots*, é uma das técnicas que agravam esse tipo de influência indevida. A discussão acerca desse tema ganhou mais atenção com as eleições presidenciais nos Estados Unidos e o Brexit, ambos em 2016,bem como temores sobre operações de influência entre diferentes Estados, impulsionando o desenvolvimento de ferramentas e técnicas para identificar esse tipo de comportamento.

Considerando o fato de diferentes países possuírem padrões de uso, contextos e gramáticas diversas, existe um desafio em avaliar se os métodos desenvolvidos para outro país são capazes de cobrir as especificidades do cenário brasileiro. A opacidade de parte dos identificadores de *bots* com relação aos métodos de detecção e a dificuldade em aplicar ao cenário nacional critérios feitos para outras realidades limita as análises de jornalistas e pesquisadores envolvidos nas investigações sobre o tema.

Sendo assim, este trabalho se propõe a analisar a existência de comportamento automatizado em perfis do Twitter que interagiram textualmente (i.e. comentários) com candidatos à prefeitura das capitais em que houve segundo turno durante as eleições brasileiras de 2020. Essas capitais compreendem Rio de Janeiro, São Paulo, Recife, Porto Alegre e Fortaleza. A partir dos resultados obtidos, realizamos dois outros tipos de análises: análise de redes e lexical. Isso nos ajuda a entender, tanto em termos de conteúdo quanto em coordenação em rede, as características a respeito dos tipos de comportamento automatizado no Brasil. O artigo é parte de um esforço mais amplo, que inclui posterior inclusão de outras capitais e análise de *retweets*. Além desta seção de introdução, o presente trabalho está organizado como segue: a Seção 2 apresenta trabalhos relacionados, a Seção 3 descreve a metodologia utilizada e a Seção 4 descreve os resultados preliminares obtidos. Por fim, a Seção 5 conclui o trabalho com algumas considerações finais.

## 2. Trabalhos relacionados

Contas seguindo comandos automatizadas que compartilham conteúdo nas redes sociais são também conhecidas como *bots* sociais. Esses *bots* são controlados por aplicações (*software*), responsáveis por gerar conteúdo de forma artificial, imitando comportamento humano e estabelecendo interações com usuários autênticos [Ruediger et al. 2017]. Discussões sobre a existência e comportamento de *bots* em redes sociais se popularizaram nos últimos anos com o surgimento de estudos apontando a interferência dessas contas na criação de movimentos não espontâneos, principalmente para influenciar o debate político.

Sobre as eleições nos Estados Unidos de 2016, por exemplo, Heredia et al. [Heredia et al. 2018] fazem uma análise sobre a influência que *bots* exerceram na opinião pública, considerando mais de 700 mil usuários no Twitter. Os autores utilizam a ferramenta Botometer para identificação de *bots* associando-a à análise de sentimento através de redes neurais convolucionais. Dentre os resultados, os autores mostram o papel dos *bots* influenciando positiva ou negativamente os candidatos analisados. Também dentro do tema eleições, Morstatter et al. [Morstatter et al. 2018] analisam o papel exercido por *bots* nas eleições de 2017 na Alemanha. Usando dados do Twitter, eles identificaram que pelo menos 11% dos usuários analisados eram *bots*. Os autores também fazem uma análise de redes para identificação de comunidades, a fim de entender como ocorre

o fluxo de informação entre elas. Rofrío et al. [Rofrío et al. 2019] apresenta um estudo sobre a presença de *bots* nas eleições de 2017 no Equador. Analisando 30 mil *bots* eles evidenciam como essas contas foram utilizadas para favorecer determinados candidatos.

No contexto brasileiro, Ruediger et al. [Ruediger et al. 2017] analisa a presença de *bots* no Twitter a partir de um método que considera o intervalo de publicação de *tweets* e a utilização de plataformas de automatização para publicação de conteúdo. No relatório, os autores apresentam estudos de caso que incluem temas relacionados, por exemplo, às eleições de 2014 e debates políticos. Através de mapas de interações entre usuários, eles identificam como ocorre a concentração de contas que publicam *tweets* de forma automatizada.

Em comparação às demais abordagens mencionadas, a abordagem proposta neste trabalho busca incluir múltiplos critérios para a identificação de *bots* no contexto brasileiro, de forma mais transparente. Buscamos entender as especificidades de cada método de identificação de *bots* a partir de dados extraídos das contas analisadas do Twitter, bem como as características dos comportamentos coordenados no país. Propomos a realização tanto de uma análise estrutural de rede, quanto da análise lexical que considera o conteúdo que é compartilhado por contas com comportamento total ou parcialmente automatizado. Considerando também o cenário das eleições em diferentes capitais, entendemos que esses dois tipos de análise nos permitem, por exemplo, explorar sobre as diferenças na interação de *bots* com os perfis analisados e os reflexos do comportamento coordenado nos possíveis *clusters* de *bots* e nos padrões de vocabulário em torno de cada eleição.

## 3. Metodologia

Para o presente trabalho, a metodologia utilizada foi dividida em três etapas: (i) coleta de dados, (ii) aplicação de algoritmos de identificação de comportamento automatizado utilizados pelo PegaBot e (iii) cruzamento, análise de rede e lexical. Cada uma das etapas é descrita com mais detalhes nas subseções a seguir.

### 3.1. Coleta de dados

Para a coleta dos dados acerca dos *tweets*, foi utilizada a ferramenta ePOCS Twitter Crawler (eTC)[1], desenvolvida pelos laboratórios de P&D BioBD e COMP da PUC-Rio. A eTC permite a extração de dados históricos do Twitter, coletando dados e metadados envolvidos em *tweets* sobre determinados assuntos dentro de um período de tempo especificado. A eTC é composta por uma aplicação Web e uma aplicação *standalone*, responsável por fazer o *crawling* e extrair os dados. Por meio da interface Web, os usuários agendam a extração de informações sobre *tweets*, especificando para isso o termo da busca e as datas.

Para o presente trabalho, foram realizadas duas buscas para cada candidato, feitas entre 26 de setembro e 30 de novembro de 2020. A busca foi realizada para os candidatos à prefeitura (2º turno) das capitais Rio de Janeiro, São Paulo, Recife, Porto Alegre e Fortaleza em 2020. Assim, para cada *tweet* destas buscas, a ePOCS implementa um algoritmo que seleciona uma amostra de até 100 usuários que comentaram postagens dos candidatos e, em seguida, agrupa todos estes usuários contando a frequência com a qual cada um apareceu. Vale mencionar que os dados necessários à análise, processados

[1] `https://etc.biobd.inf.puc-rio.br/`

apenas para fins de pesquisa, foram agregados, ou seja, qualquer dado referente a usuários específicos foi anonimizado e descartado após uso, garantindo a proteção aos dados e adequação às exigências da Lei Geral de Proteção de Dados (LGPD).

### 3.2. Identificação de comportamento automatizado

Para a detecção de comportamento automatizado, foi utilizada a nova versão do PegaBot[2], projeto do Instituto do Tecnologia e Sociedade do Rio de Janeiro (ITS Rio) e do Instituto Equidade & Tecnologia desenvolvido em 2018. O algoritmo do PegaBot utiliza como fonte de dados as informações públicas dos perfis no Twitter e critérios específicos para reconhecer padrões comportamentais. O objetivo da análise é identificar características que ajudem a determinar se o perfil apresenta comportamentos similares ao de contas automatizadas. Abaixo estão descritos um resumo dos critérios adotados pela ferramenta.

**Perfil do usuário:** algumas das informações públicas dos perfis consideradas são o nome do perfil do usuário, e quantos caracteres ele possui, quantidade de perfis seguidos (*following*) e seguidores (*followers*), texto da descrição do perfil, número de postagens (*tweets*) e favoritos.

**Rede:** é feita a coleta de uma amostra de até 200 *tweets* mais recentes publicados na linha do tempo do usuário, identificando *hashtags* e menções utilizadas. O índice de rede busca compreender se o usuário está, por exemplo, encaminhando mensagens de *spam* para uma *hashtag* específica. Perfis com grande quantidade de interações e comportamento monotemático quanto às *hashtags* utilizadas tendem a ter uma pontuação superior nesse tipo de análise.

**Análise de sentimentos:** a amostra de até 200 *tweets* mais recentes publicados pela conta é usada para identificar a neutralidade do perfil. Isso é feito a partir de uma pontuação, que é atribuída a cada uma das palavras dos *tweets* coletados. A classificação se baseia em um dicionário pré-estabelecido de palavras e pontuações, a partir do qual é calculada a pontuação média para a quantidade de palavras positivas, negativas e neutras utilizadas pelo usuário. Quanto mais neutro, menor a chance de ser considerado um *bot*.

**Análise temporal:** considera a frequência de postagem de *tweets* para a amostra coletada, a data de criação da conta e o total de *tweets* publicados pelo perfil a fim de verificar se o usuário tem uma alta atividade de publicações.

Após coletar as informações, o PegaBot processa e transforma os dados recebidos em variáveis que compõem o cálculo final de probabilidade. O resultado final indica o percentual de probabilidade de um perfil do Twitter ser um *bot*. Quanto maior o percentual, maior a chance da conta não ser de um ser natural.

### 3.3. Cruzamento e análise de rede e lexical

Os dados obtidos nas etapas anteriores alimentam três tipos de análise de comportamento coordenado: análise de redes envolvendo (a) *clusters* de *hashtags*, (b) *clusters* de perfis com citações mútuas e (c) análise lexical. Através de uma expressão regular para raspagem, retiramos e registramos as *hashtags* e perfis citados em cada *tweet*, identificando que outras postagens compartilham as mesmas características e construímos um grafo de rede. Nos *clusters* identificados nesse grafo, podemos definir quais *hashtags* e perfis são

[2] `https://pegabot.com.br/`

utilizados de modo coordenado por perfis com maior probabilidade de serem *bots*, qual é a centralidade de cada um deles na rede identificada (algoritmos de centralidade) e quais são as diferenças entre os grupos identificados (algoritmos de modularidade/Louvain). Como base da comparação entre léxicos, utilizamos a combinação entre critérios, separando possíveis bots identificados nos critérios de sentimento e rede, de bots que combinam identificações nas análises temporal e de usuário (todas as combinações possiveis foram consideradas, buscando a pluralidade dos tipos de bots). O conteúdo textual das mensagens passa por uma análise lexical (*scripts* em linguagem R) centrada na modelagem de tópicos e identificação de vocabulários, visando identificar qual é o vocabulário dos diferentes grupos de *bots* participando ativamente das eleições de 2020 nas prefeituras selecionadas. O resultado final desses três níveis de análise mutuamente independentes passa por uma etapa de validação cruzada, ou seja, os resultados obtidos através de um nível (lexical, por exemplo) são validados através de resultados similares obtidos em outras frentes (análise de redes). Esses resultados podem aprimorar a detecção e análise de comportamento coordenado automatizado e diferenciar os grupos de atores envolvidos.

## 4. Resultados preliminares

Com os dados coletados e processados de acordo com as duas primeiras etapas da metodologia apresentada na Seção 3, seguimos com as análises propostas na terceira etapa. Como parte da primeira análise, foi possível construir uma rede bipartite composta por *hashtags* e os perfis que as utilizam. Essa rede possui 272.807 vértices, sendo 92.600 usuários e 180.207 *hashtags*, conectados por 1.006.341 arestas (cada utilização de *hashtag* conta como uma única aresta entre os vértices envolvidos). Utilizando algoritmo de detecção de comunidades segundo o método de Louvain (interface Gephi 0.9.1, Java) identificamos 1.958 *clusters*. A distribuição dos perfis é altamente concentrada em dois *clusters* (ids 174 e 1020) seguindo *hashtags* temáticas: o primeiro é marcado por ataques ao presidente da Câmara, à imprensa e a candidatos de esquerda ou identificados como oponentes do presidente Jair Bolsonaro; o segundo traz mensagens positivas de apoio a campanhas de candidatos de esquerda. Embora essas duas comunidades sejam grandes, há uma clara concentração de comportamento automatizado em uma delas: nas análises temporal, usuário e total, a comunidade 174 apresenta mais que o dobro do número de *bots* encontrados na comunidade 1020. Nas análises rede e sentimento, a quantidade de *bots* na comunidade 1020 corresponde, respectivamente, a 72% e 66% dos encontrados na 174. Há portanto, não apenas diferentes *clusters* de *bots*, mas tipos de *bots* diferentes em cada um desses *clusters*.

Considerando o texto das postagens para a análise lexical, uma análise fatorial de correspondência aponta para pluralidade de léxicos nos *clusters* identificados, embora os grandes *clusters* concentrando a maior parte dos vértices da rede apresentem vocabulários similares. A utilização de clusterização hierárquica descendente (método Reinert aplicado através da interface para linguagem R IRaMuTeQ) para análise lexical aponta nove vocabulários predominantes: cinco vocabulários relacionados a postagens sobre candidatos envolvidos nas disputas das capitais, dois relacionados a *gifs* e *emojis*, um relacionado a ataques e xingamentos e um último com vocabulário genérico.

## 5. Conclusões

O presente trabalho avança ao analisar e viabilizar a comparação sobre o uso de *bots* e propaganda computacional nas eleições em diferentes capitais brasileiras, auxiliando na

superação de vieses regionais das pesquisas sobre *bots* no país. Identificamos não apenas a participação massiva de *bots* em algumas capitais pouco exploradas (notadamente Porto Alegre) mas também o modo como a utilização de diferentes critérios para identificação de *bots* poderia descrever esse cenário de modos consideravelmente diferentes. Isso confirma a necessidade de uma perspectiva que una essas diferentes análises e possibilite definições mais precisas sobre tipos de *bots* e não apenas sua identificação como tal. A maior clareza quanto às diferenças entre os resultados de cada uma dessas análises de comportamento automatizado e as relações dos *bots* com a discussão eleitoral brasileira no Twitter são ganhos fundamentais para proteção de nossos processos eleitorais. Embora diferentes identificadores de *bots* apresentem leituras diversas sobre a probabilidade, quantidade e atuação indevida desses perfis automatizados no cenário político, investigadores ainda não têm acesso a dados que indiquem o motivo dessas divergências e, portanto, a confiabilidade de suas próprias análises baseadas em identificações feitas por critérios igualmente opacos. Dessa forma, este trabalho busca trazer maior segurança e confiabilidade às conclusões dos atores envolvidos na discussão sobre o tema. Vale mencionar que este é um trabalho em andamento e que, além de nos aprofundarmos nas análises aqui apresentadas, vislumbramos a ampliação do estudo para todas as capitais do país, bem como a possibilidade de recomendações para o combate ao uso indevido de propaganda computacional na eleição presidencial de 2022 e de um mecanismo que apoie o monitoramento desse tipo de propaganda computacional durante o pleito.

## Agradecimentos

Agradecemos a União Europeia pelo financiamento do projeto Pegabot e o suporte da CAPES.

## Referências

Almeida, Y. L., Rubin, F. S., de Faria Alvim, A. C., Dias, V. M. F., and dos Santos, R. P. (2020). O uso das redes sociais para interferir nas democracias: Um mapeamento sistemático da literatura. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 178–183. SBC.

Heredia, B., Prusa, J. D., and Khoshgoftaar, T. M. (2018). The impact of malicious accounts on political tweet sentiment. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 197–202. IEEE.

Morstatter, F., Shao, Y., Galstyan, A., and Karunasekera, S. (2018). From alt-right to alt-rechts: Twitter analysis of the 2017 german federal election. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 621–628, Republic and Canton of Geneva, CHE.

Rofrío, D., Ruiz, A., Sosebee, E., Raza, Q., Bashir, A., Crandall, J., and Sandoval, R. (2019). Presidential elections in ecuador: Bot presence in twitter. In *2019 Sixth International Conference on eDemocracy eGovernment (ICEDEG)*, pages 218–223.

Ruediger, M. A., Grassi, A., Freitas, A., Contarato, A. d. S., Taboada, C., Carvalho, D., Ferreira, H., Silva, L. R. d., Lenhard, P., Bastos, R., et al. (2017). Robôs, redes sociais e política no brasil: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018.