

Analysis of bots political activity in non-electoral period

Análise da atividade política de *bots* em período não eleitoral

Lucca B. R. Barreto,¹ Adolfo P. Guimarães¹

¹Universidade Tiradentes (UNIT)
Aracaju – SE – Brasil

{lucca.bortoloso,adolfo.pinto}@souunit.com.br

Abstract. *The political strategists have begun to explore new means of mass propagation of information, given the high availability of access to information from social media. This allowed these networks to be used to automate the sending and sharing of political ideals composed of messages of support or defamation. Thus, the so-called bots become increasingly frequent on social networks, especially on Twitter. For this reason, this work aims to analyze the political activity generated and shared by bots in a non-electoral period, to understand possible strategies and behaviors that influence on the Brazilian political debate. From hashtags identified as automated or suspicious account activities, it was possible to show many bots in the political debate on social networks and the presence of politicians at the center of interactions.*

Resumo. *Os estrategistas políticos têm começado a explorar novos meios de propagação de informações em massa, visto a grande disponibilidade de acesso às informações advindas das redes sociais. Isso possibilitou que essas redes fossem utilizadas com o intuito de automatizar o envio e compartilhamento de ideais políticos compostos de mensagens de apoio ou de difamação. Sendo assim, os chamados bots tornaram-se cada vez mais frequentes nas redes sociais, principalmente no Twitter. Por esse motivo, este trabalho visa analisar a atividade política gerada e compartilhada por bots em um período não-eleitoral, a fim de entender possíveis estratégias e comportamentos que tenham influência no debate político brasileiro. A partir de hashtags identificadas como atividades de contas automatizadas ou suspeitas, foi possível mostrar um grande número de bots no debate político nas redes sociais e a presença de políticos no centro de interações das redes automatizadas.*

1. Introdução

O advento da internet proporcionou uma ampla interação entre usuários. Por consequência, as redes sociais ganharam uma enorme notoriedade por seu uso acessível, rápido e efetivo, conectando diversas pessoas ao redor do mundo. Atualmente, é comum que grande parte da população brasileira tenha contato com as redes sociais, visto que seu uso não é restrito a determinados grupos sociais ou de faixa etária definida. Mesmo assim, é notoriamente sabido que a internet brasileira não alcança uma parcela da população. Prova disso é que, em um estudo realizado pelo Centro Regional para o Desenvolvimento de Estudos sobre a Sociedade da Informação (Cetic.br) [Núcleo de Informação e Coordenação do Ponto BR (NIC.br) 2020], um em cada quatro

brasileiros não possui acesso à internet, o que corresponde a 25% da população brasileira. Dos usuários que possuem acesso, 92% utilizam para trocar informações em aplicativos de mensagem como o WhatsApp e o Facebook Messenger e 76% utilizam para acessar redes sociais. Nota-se que esses softwares sociais possuem ampla utilização e criam uma rede de comunicação muito vasta de troca de informações e opiniões que vem sendo aproveitada principalmente no âmbito político. Por consequência, surge o debate sobre o uso político das redes sociais, visto que sua influência pode impactar diretamente a democracia de um país.

O uso automatizado das redes sociais tende a crescer nos próximos anos, possibilitando assim uma maior presença dos chamados *bots* no ambiente dessas redes. Sendo assim, o Twitter como uma mídia social amplamente utilizada para criação e aplicação de robôs, serve hoje como meio de observação para estudos que buscam entender como seu conteúdo impacta diretamente no meio político de um país. Além do mais, nos Estados Unidos, cerca de 19% dos *tweets* publicados sobre as eleições presidenciais do país, em 2016, foram produzidos por *bots*, enquanto na Suécia a porcentagem foi de 5% [Bessi and Ferrara 2016], [J. Fernquist and Schroeder 2018]. Em ambos os estudos, os resultados mostraram que os robôs agem de forma semelhante mesmo que em contextos políticos diferentes. Sendo assim, o entendimento de como esses *bots* agem é de imensa importância. Ademais, é fundamental que a investigação leve em consideração os momentos pré-eleitorais e não somente o período eleitoral do país.

Este trabalho tem como objetivo estudar as atividades políticas de *bots* em período não eleitoral no Brasil. Entende-se que atividade política é toda ação de disseminação de conteúdo político que vise alterar o debate público do país. Para isso, foram utilizadas *hashtags* identificadas com atividades de contas automatizadas ou suspeitas, para coleta de *tweets* que contenham temas da política brasileira. Dessa forma, serão analisados os perfis que fizeram a publicação, identificando a probabilidade destes serem *bots*, para que em seguida sejam feitas as análises da estrutura de rede criada a partir da interação entre diferentes perfis. Em síntese, este estudo identificou uma grande quantidade de *bots* presentes nos debates políticos do Brasil (84,13%), sendo que muitos deles seguem padrões já relatados por outros estudos de diferentes países. Além disso, também foi identificada a presença de políticos no centro de interações das redes automatizadas. Entende-se que esse trabalho será importante para futuros estudos sobre a influência de *bots* na política brasileira contribuindo para futuras pesquisas sobre o comportamento de *bots* nas eleições presidenciais que ocorrerão em 2022.

O estudo segue quatro seções principais. Na seção 2 serão apresentados os trabalhos que estão relacionados com o projeto. A seção 3 descreve a metodologia utilizada no estudo, em relação às etapas que foram aplicadas. A seção 4 apresenta os resultados obtidos e, por fim, a seção 5 apresenta as conclusões do presente estudo, sugerindo também futuros estudos.

2. Trabalhos Relacionados

No trabalho [J. Fernquist and Schroeder 2018] os autores treinaram um modelo de aprendizagem de máquina para reconhecer comportamentos automáticos e, dessa forma, estudar as contas de Twitter que estão comentando as eleições suecas. Para esse trabalho, os autores incluíram todas as formas de automatização, sejam provocadas por *bots* ou por

humanos que agem de forma automática no Twitter. No estudo de caso, os pesquisadores fizeram uma análise do número de robôs que estavam comentando sobre as eleições suecas e que tipo de mensagem esses *bots* estavam compartilhando.

O trabalho [L. Hagen and Vasquez 2020] analisou as atividades de *bots* no meio político em período não eleitoral. Sendo assim, os pesquisadores examinaram o impacto do conteúdo gerado por robôs no Twitter. Para isso, utilizaram como estudo de caso a investigação da interferência da Rússia nas eleições presidenciais dos EUA de 2016. Neste caso, os autores focaram em fazer uma análise a respeito de quatro pontos principais: estrutura de rede, conteúdo/mensagens, sentimento e influência. Para isso, os pesquisadores utilizaram o Twitter API para coletar dados entre julho e agosto de 2017, selecionando *tweets* que continham palavras-chave como: “TrumpRussia”, “Donald Trump Jr.”, “trump”, “Russia”, etc. Como resultado, eles montaram uma base de dados no qual avaliaram a quantidade de *bots* incluídos e puderam analisar os quatro pontos principais anteriormente selecionados.

O trabalho [Bessi and Ferrara 2016] investigou como a presença de *bots* nas redes sociais afetaram as eleições presidenciais dos EUA no período de 16 de setembro a 21 de outubro de 2016. Os autores elaboraram uma lista com 23 *hashtags*, escolhidas manualmente, e utilizaram-as como forma de busca e coleta de *tweets* através do Twitter Search API. Para descobrir a quantidade de usuários humanos e robôs, os pesquisadores utilizaram o *framework* BotOrNot que extrai e analisa a possibilidade de um usuário ser um *bot* ou não. Após isso, foram feitas análises sobre as influências dos *bots* no conteúdo gerado nas redes sociais em relação à eleição presidencial estadunidense.

Este trabalho analisou a presença de *bots* nas discussões políticas brasileiras que aconteceram no Twitter durante o período de maio e junho de 2021. Os trabalhos relacionados serviram de base para a escolha de ferramentas como o BotOrNot e o Twitter Search API, utilizados nos trabalhos de [L. Hagen and Vasquez 2020] e [Bessi and Ferrara 2016]. Uma das diferenças deste projeto está na seleção das *hashtags*, ao qual teve o auxílio de uma ferramenta que revela as *hashtags* mais impulsionadas por contas automatizadas durante um período de escolha do usuário, o que possibilitou focalizar o estudo em *tweets* que possuíam vínculos com *bots*.

3. Metodologia

Nesta seção, serão descritas as etapas executadas durante o estudo, informando e detalhando todos os processos. Todas as configurações das ferramentas utilizadas neste trabalho estão descritas no Github do projeto¹.

3.1. Seleção de *hashtags*

As *hashtags* passaram por um critério de seleção definido pelos seguintes parâmetros: 1) ter sido propagada por contas automatizadas, 2) a *hashtag* deve estar estritamente relacionada à política brasileira e 3) ter sido impulsionada dentro do período de estudo. Logo, *hashtags* como, por exemplo, “#FakeNews” não podem ser selecionadas, pois embora essa seja uma *hashtag* propagada no meio político, ela não é estritamente política. Como as *hashtags* possuem papel fundamental na captura de *tweets*, foi necessário o auxílio de

¹https://github.com/adolfoguimaraes/analise_bots_twitter

ferramentas que contribuíram com a execução dos critérios de seleção. Dessa forma, foi decidido pela utilização do Bot Sentinel², que teve papel fundamental para a execução do primeiro e terceiro critério.

O Bot Sentinel disponibiliza em seu site um *dashboard* com o ranqueamento das *hashtags* mais disseminadas por contas suspeitas em um determinado período escolhido pelo usuário. Neste estudo, foram selecionados três horários (3:00 AM, 11:00 AM e 7:00 PM) de análise de *hashtags* em um período de 01 de abril de 2021 à 13 de maio de 2021, considerando o fato de existirem atividades momentâneas de *bots*, ou seja, uma atividade que ocorre em um momento específico e que muitas vezes não voltam a ocorrer.

Durante a seleção, foram encontradas 72 *hashtags* que passaram pelos critérios definidos anteriormente. A fim de reduzir a quantidade de *hashtags* que foram estudadas, foram adotados outros critérios de filtragem das *hashtags*, que consistiam em filtrar as dez *hashtags* mais disseminadas e mais frequentes, de forma manual, levando em consideração o período de estudo, além de não selecionar *hashtags* que continham erros de português ou que foram disseminadas em uma data específica. Por fim, foram selecionadas 10 *hashtags* que são descritas pela Tabela 1. A tabela também mostra o quantitativo de *tweets* coletados por *hashtag*. Detalhes desta coleta estão na seção 3.2.

Tabela 1. Quantidade de *tweets* coletados por *hashtag* e grupo

<i>Hashtag</i>	Quantidade de <i>tweets</i>	
	Grupo 1	Grupo 2
#EuAutorizoPresidente	202	1000
#MamataConnection	298	1000
#STFVergonhaNacional	500	1000
#FechadoComBolsonaro	500	1000
#STFVergonhaMundial	500	1000
#RenanVagabundo	285	1000
#BolsonaroTemRazaoDeNovo	500	1000
#BolsonaroAte2026	500	1000
#VotoAuditavelJa	500	1000
#ForaBolsonaro	500	1000

Vale ressaltar que todas as *hashtags* selecionadas foram com base em uma análise prévia dos dados obtidos no *dashboard* do Bot Sentinel, sem qualquer manifestação pessoal para a escolha destas, conforme os critérios descritos nesta seção.

3.2. Coleta dos dados

A ferramenta para a coleta dos dados foi o Tweepy³, que facilitou o uso dos *endpoints* da API fornecida pelo Twitter. Neste estudo, foram necessários alguns filtros na coleta de *tweets* como o idioma em português, a retirada de *retweets*, a existência de textos no corpo do *tweet* (por exemplo *hashtags*) e a periodicidade dos *tweets*, que são disponibilizados em três períodos: (1) *recent* (retorna os *tweets* mais recentes da data da coleta);

²<https://botsentinel.com/>

³<https://www.tweepy.org/>

(2) *popular* (retorna os *tweets* mais populares, ou seja, de maior interação) e (3) *mixed* (retorna tanto os *tweets* mais populares quanto os mais recentes).

Neste caso, como o objetivo é atingir o maior número de usuários possível, foi utilizado o filtro *mixed*, visto que a escolha de um dos dois outros filtros poderia comprometer a análise do estudo. No caso do filtro *recent*, os *tweets* coletados poderiam apresentar uma maior prática de *spamming*, enquanto no filtro popular, a coleta atingiria um grupo seletivo de influenciadores na rede, deixando de lado a maior parte dos usuários do Twitter. Em contraponto, o filtro *mixed* consegue mesclar os dois conjuntos de dados e assim reduzir os problemas que cada um dos dois outros filtros possuem.

Neste estudo, foram coletados dados do dia 17 de maio de 2021 à 27 de maio de 2021 e, posteriormente, também foram coletados dados do dia 04 de junho de 2021, sendo que no primeiro período de coleta não foram coletados *retweets* e o segundo período coletou *retweets*. Foram reunidos dois grupos de dados, sendo que o primeiro grupo, o qual chamaremos de Grupo 1, contém informações de $N = 4285$ *tweets*, que foram utilizados para o estudo da identificação dos *bots*. No segundo grupo, que será chamado de Grupo 2, foram coletados $N = 10000$ *tweets* que complementam as análises das estruturas de redes baseadas nas detecções de *bots*. A Tabela 1 descreve a quantidade de *tweets* coletados por cada *hashtag*, em cada grupo.

Inicialmente foi decidido pela coleta de 500 *tweets* por *hashtag* para compor o Grupo 1 e 1000 *tweets* por *hashtag* para o Grupo 2. Porém, como demonstrado pela tabela, em três das dez *hashtags* selecionadas não foi possível a coleta dos 500 *tweets* do Grupo 1. Isso pode ter acontecido pois na coleta do Grupo 1 foi utilizado um filtro que retirava *retweets*, portanto para melhor análise, os *tweets* do Grupo 2 não ficaram sujeitos a esse filtro e dessa forma, atingiu-se o objetivo de coleta de $N = 10000$ *tweets*, sendo que desses dez mil, $N = 4798$ foram *retweets*.

3.3. Detecção de *bots* sociais

Detectar os usuários que agem de maneira automatizada é fundamental para que este trabalho atinja o objetivo final. Para isso, foi necessário utilizar uma ferramenta que auxiliasse na detecção de possíveis *bots* sociais. Neste caso, o Botometer⁴ (BotOrNot), ferramenta desenvolvida pelo Observatório de Mídias Sociais (OSoMe) que checa a atividade de uma conta de Twitter e retorna a probabilidade da conta ser automatizada ou não. A probabilidade é baseada em 7 tipos (*scores*) de *bots*. Para esse trabalho vale destacar 3 deles, devido a importância desses três tipos para o estudo: (1) *Astroturf*: *bots* políticos que foram marcados manualmente e contas envolvidas em esquema de *followtrain*, prática esta que consiste em comentários em publicações indicando que se a pessoa seguir o autor do comentário, o mesmo irá seguir o usuário de volta; (2) *Spammer*: contas rotuladas como *spambot*, ou seja, que praticam o envio em massa de mensagens e publicações; (3) *Other*: outros *bots* obtidos de anotação manual, *feedback* do usuário, etc.

O Botometer utiliza um modelo *machine learning* treinado que compara uma conta a dezenas de milhares de outras previamente rotuladas. Após a checagem do perfil, a ferramenta analisa a parte pública do perfil, como *tweets* e menções públicas, usando a API do Twitter. As informações são analisadas pela API do Botometer, que

⁴<https://botometer.osome.iu.edu/>

extraí diversos recursos para caracterizar o perfil da conta, tais como: rede de amigos, estrutura da rede social, padrões temporais de atividade, linguagem e sentimento. Por último, o Botometer utiliza o algoritmo supervisionado *Random Forest* para classificar o usuário como *bot* ou não, utilizando esse método em cada uma das sete pontuações e ao final gerar uma pontuação geral do usuário analisado, conforme descrito por [C. A. Davis and Menczer 2016]. Como retorno, a ferramenta fornece a pontuação de cada tipo de *score* anteriormente explicado. O estudo de [S. Wojcik and Hitlin 2018] sugere que o critério de classificação de *bots* sejam pontuações acima de 0.43, no entanto, no estudo de [Bessi and Ferrara 2016] o critério adotado foram pontuações superiores à 0.5. Baseado nisso, neste estudo adotou-se como critério pontuações iguais ou superiores à 0.5.

Vale ressaltar que, neste trabalho, foi utilizado o plano *Basic* da API do Botometer, visto que o plano é uma versão gratuita e disponível para testes da API. Com isso, ocorreram limitações em relação a quantidade de solicitações por dia, limitadas em 500 requisições por dia. Por fim, o Botometer não é a única ferramenta disponível para classificação de *bots* no Twitter. Outras ferramentas como o projeto brasileiro Pegabot⁵ atuam de forma semelhante a fim de classificar um usuário em automatizado ou não. Contudo, optou-se somente pelo uso do Botometer para manter o mesmo critério de classificação de *bots*.

3.4. Estrutura de Rede

A análise de uma estrutura de rede é fundamental para a compreensão dos *clusters* que se formam durante as interações dos softwares sociais. Dessa forma, para facilitar a análise dessas redes, neste estudo foi utilizada a ferramenta Gephi⁶, um software *open source* de exploração e manipulação de redes [M. Bastian and Jacomy 2009].

As estruturas de rede seguem medidas estatísticas para identificação de comunidades, também conhecidas como *clusters*, que interagem entre si com base em características semelhantes. No estudo dos *bots*, compreender como essas comunidades interagem é fundamental, visto que uma publicação em redes sociais é impulsionada pelas interações dos usuários com aquela publicação. Dessa forma, neste projeto buscou-se estudar como estão organizadas as interações entre *bots* com base nas *hashtags* coletadas.

Neste estudo foram analisadas as relações de *tweet-retweet* entre usuários que compõem os dados coletados no Grupo 2. Este grupo possui $N = 10000$ *tweets*, mas como a proposta é analisar relações de *tweet-retweet*, foi necessário retirar dessa base aqueles *tweets* que não foram “retweetados”. Dessa forma, ao retirar esses dados, sobram informações de $N = 4798$ *tweets* que continham a relação entre o usuário que fez o *retweet* e o usuário da publicação original. Com isso, foi possível criar uma estrutura de rede que, com posterior análise de detecção de contas automatizadas, servirá de base para o estudo de como os *bots* estão se relacionando nas redes sociais.

4. Análise da Atividade Política Brasileira

Inicialmente foi feita uma análise em cima da detecção de *bots*, a fim de compreender as quantidades absolutas dos dados coletados no Grupo 1, ou seja, saber exatamente

⁵<https://pegabot.com.br/>

⁶<https://gephi.org/>

quantos *tweets* foram gerados por *bots*, quantos não foram e quantos desses não foram possíveis calcular a probabilidade do usuário ser automatizado. No total foram 3605 *tweets* gerados por *bots*, 630 gerados por não *bots* e 50 *tweets* gerados por contas sem classificação. É notável a diferença da quantidade de publicações que foram feitas por *bots* em comparação com a quantidade de publicações geradas por contas não automatizadas. Os números mostram que dentro do Grupo 1, os *bots* correspondem a 84,13% das publicações, enquanto os não *bots* totalizam 14,70% das publicações totais.

Partindo dessa análise, o próximo passo foi observar como estão distribuídas as pontuações *CAP*, que calculam a probabilidade de um usuário ser uma conta automatizada ou não. Apesar da pequena quantidade de dados coletados e analisados, devido, principalmente, ao curto período de tempo do estudo, foi notado que a grande quantidade de usuários se encontra em pontuações próximas de 0.8, diferentemente do que foi visto em [Bessi and Ferrara 2016], sendo que essa análise foi gerada com base nas pontuações de $N = 2805$ usuários únicos.

Se analisarmos a quantidade de *tweets* coletados por cada *hashtag* no Grupo 1, descrita pela Tabela 1, percebemos que há uma grande quantidade de *tweets* publicados por *bots*, como por exemplo a *hashtag* “EuAutorizoPresidente” que contém 202 *tweets* coletados no total e quase 200 foram gerados por *bots*. Além disso, a média de pontuação desses usuários está bem próxima de 0.8, como já era esperado, visto que a densidade desses dados estava concentrada exatamente nesse intervalo entre 0.6 e 0.8. Outro ponto para se observar é a porcentagem de contas automatizadas em cada *hashtag*, como descrito na Tabela 2. A única *hashtag* que fica abaixo da taxa de 60% é a *hashtag* “ForaBolsonaro”, que possui uma taxa de *bots* de 58,2%.

Tabela 2. Porcentagem da presença de *bots* em cada *hashtag*.

<i>Hashtag</i>	%
EuAutorizoPresidente	95,0%
MamataConnection	92,3%
BolsonaroTemRazaoDeNovo	91,8%
RenanVagabundo	90,9%
FechadoComBolsonaro	88,0%
STFVergonhaNacional	87,2%
VotoAuditavelJa	87,2%
BolsonaroAte2026	85,0%
STFVergonhaMundial	78,4%
ForaBolsonaro	58,2%

A partir disso, uma importante questão foi levantada: será que muitas das contas, que são notavelmente não *bots* estão sendo pontuadas como *bots*? Para responder essa questão, foram analisados os perfis verificados que se encontram na base de dados do Grupo 1. A verificação de Twitter é um importante parâmetro para responder essa questão, pois somente são autenticados os perfis que são notáveis, autênticos, ativos e com alto interesse público. Ao analisar a quantidade de perfis verificados na base de dados do Grupo 1, foram encontrados 29 perfis únicos que corresponderam a 46 *tweets* presentes na base de dados. Em seguida, foi analisada a distribuição de frequência das pontuações *CAP* desses perfis, a fim de extrair alguma informação que possa responder a questão

anteriormente levantada.

A análise permitiu verificar que os perfis verificados estão sendo pontuados como *bots*, visto que o intervalo de pontuações destes perfis varia de 0.6 à, aproximadamente, 0.85. Dessa forma, podemos concluir que mesmo as contas não sendo propriamente robôs, os comportamentos analisados nos perfis se assemelham muito a ações automatizadas. Além da pontuação *CAP*, outras pontuações são bastante importantes para conseguirmos classificar esses *bots*. Para isso, foram analisadas a frequência e a média de pontuação das seguintes categorias: *Astroturf*, *Spammer* e *Other*.

Ao analisar a pontuação *Astroturf*, a quantidade de registros de *bots* que foram classificados nesta categoria varia entre 100 à 175 registros em oito das dez *hashtags*. Porém, em duas *hashtags* ("EuAutorizoBolsonaro" e "ForaBolsonaro") o número de registros é consideravelmente baixo (43 e 28, respectivamente) comparado aos demais, mostrando uma tendência de que nessas *hashtags* não ocorram tantas marcações manuais ou esquemas de *follow-train*. Ademais, analisando a porcentagem correspondente desses registros é notável a diferença, visto que a quantidade de *bots* políticos na *hashtag* "ForaBolsonaro" corresponde a 5,4%, enquanto a "EuAutorizoPresidente" refere-se a, aproximadamente, 23,76% do total.

Em relação aos *spammers*, metade das *hashtags* não continham contas que foram classificadas como *bots* e das *hashtags* que tiveram registros, notamos que a quantidade de contas detectadas foi extremamente baixa, quase nulas se compararmos com a quantidade de *bots* detectados no geral. Podemos então concluir que nos dados coletados, a presença dos *bots* que praticam *spamming* é quase irrelevante na análise, visto que somente cinco *hashtags* continham essa categoria de conta automatizada, sendo que nessas *hashtags*, o número de registros únicos é inferior ou igual a 6.

Em relação a pontuação *Other*, nota-se que a média de pontuação dos *bots* é levemente inferior às outras duas categorias. Essa é uma pontuação que é baseada em *feedbacks* de usuários, isso pode indicar que é possível que o Twitter esteja excluindo contas que possuam uma pontuação bastante elevada. Porém, com base nas informações coletadas, não podemos afirmar isso, seria necessária uma análise mais profunda da situação.

Vale ressaltar o porquê da decisão de se analisar somente essas três categorias. Essas categorias possuem grande relevância na discussão política de um país, por exemplo os *bots spammers* que ganharam grande visibilidade na eleição presidencial de 2018, no Brasil. Destaca-se a grande repercussão na imprensa brasileira sobre o envio em massa de mensagens falsas que visavam difamar adversários políticos, através de contratações de empresas que prestavam o serviço de *spamming* em redes sociais, como o Whatsapp.

Analisar como esses *bots* interagem entre si e com a comunidade das redes sociais é de suma importância para este estudo. Essa análise pode trazer como resultado os principais autores que estão diretamente ligados à divulgação de publicações que foram feitas ou compartilhadas pelas contas automatizadas. Dessa forma, para o estudo da estrutura de rede gerada pelos *bots*, foram analisados os dados do Grupo 2, conforme destacado na seção 3.4. Ressalta-se que no banco de dados do Grupo 2, existem $N = 10000$ *tweets* coletados, sendo que $N = 4798$ *tweets* possuem a relação que será estudada (relação *tweet-retweet*) e destes *tweets*, foram analisadas $N = 4199$ contas únicas, que com a análise de detecção de *bots* foram comprovadas a existência de $N = 3804$ contas automatizadas.

Portanto, para gerarmos a estrutura de redes, pela ferramenta Gephi, foi necessário criar uma lista contendo todas as $N = 4798$ interações. Além disso, outra lista foi criada contendo apenas a identificação e o nome de usuário das contas automatizadas, para que fosse possível identificar nominalmente os *bots* com maior interação na rede, conforme mostra a Figura 1. Ao analisar a estrutura da rede, notamos que os usuários com maiores interações estão localizados nas bordas, mas não necessariamente isso significa que o centro da rede são de poucas interações, foi apenas uma forma visual de apresentação do grafo.

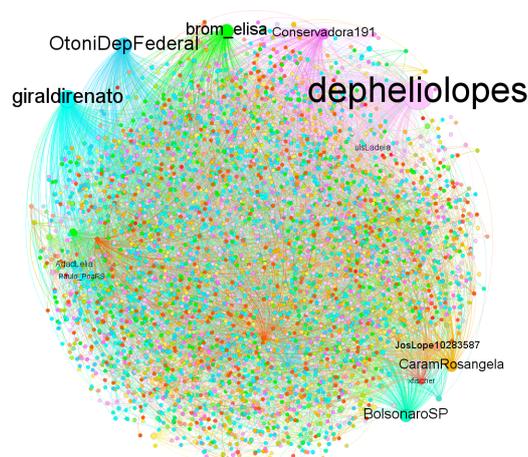


Figura 1. Estrutura de rede gerada pela coleta de tweets presentes no Grupo 2.

A partir dessa rede podemos identificar alguns usuários de maior interação nas redes, além de vários *clusters* formados de acordo com a modularidade dessa rede. Com isso, uma importante análise que podemos fazer é a presença de políticos brasileiros entre os maiores influenciadores da rede gerada pelos dados do Grupo 2. Nessa estrutura, foram identificadas 12 contas automatizadas e 2 contas não automatizadas que mais interagiram com a rede, de acordo com o intervalo de índice de centralidade estabelecido (150 mil até o valor máximo). A construção da rede mostra que podemos extrair informações relevantes da relação dos *tweets* e *retweets*. No entanto, essa análise precisa de um estudo mais amplo e aprofundado das relações encontradas para gerar conclusões mais assertivas.

5. Conclusão

Os resultados demonstraram que é notória a participação de contas automatizadas no debate político brasileiro em um período não eleitoral. Visto que, nos dados extraídos e estudados, os *bots* tiveram uma presença de 84,13%. A porcentagem de contas automatizadas se mostrou bem maior do que em alguns estudos relacionados a este trabalho, porém, assim como no caso dos estudos [L. Hagen and Vasquez 2020] e [Woolley 2016], foi possível demonstrar que os *bots* podem modificar o debate público sobre a política de um país. Neste estudo, além de comprovar a alta porcentagem de presença de *bots* nas publicações relativas à política brasileira, foram também constatadas a elevada pontuação de análise de detecção de *bots*, o que pode significar que há sim um comportamento automatizado suficientemente elevado das discussões políticas nas redes sociais, mais especificamente, no Twitter.

De início, quando pensamos em *bots* sociais lembramos dos *spammers*, pois são os tipos de contas automatizadas mais frequentes e que possuem grande visibilidade nas mídias. Porém, o estudo mostrou que mesmo com a alta presença de *bots* nos dados coletados, os *spammers* eram praticamente nulos quando comparados, por exemplo, aos *bots* políticos. Isso demonstra que mesmo com um alto comportamento automatizado, a ação de *spamming* não é frequente entre os usuários observados. Além disso, pode-se constatar que há uma parcela de usuários, que são verificados pelo Twitter e que agem de forma automatizada, mesmo que o usuário não seja propriamente um *bot*. Foi mostrada a existência de pessoas da política pública no centro da rede de interação entre contas automatizadas no debate político brasileiro no Twitter.

Por fim, este trabalho apresentou alguns padrões de comportamentos dos *bots* no debate político brasileiro em período não eleitoral. Portanto, o estudo pode servir de base para futuras análises em relação ao período eleitoral brasileiro em 2022. Faz-se necessário também estudos de caso que visem ampliar o conhecimento em relação aos resultados apresentados neste estudo, como por exemplo uma análise mais detalhada das entidades presentes nos textos disseminados pelos *bots* e qual o teor das mensagens em relação à essa entidade. Outros estudos podem ser possíveis visando a melhora dos resultados obtidos nas ferramentas, como por exemplo na detecção de *bots* do Botometer, visto que apresentou-se um elevado número de porcentagem de *bots* nos dados coletados, mesmo que as contas não fossem necessariamente robôs, como foi o caso das contas verificadas de Deputados Federais e outras pessoas públicas.

Referências

- Bessi, A. and Ferrara, E. (2016). Social bots distort the 2016 us presidential election online discussion. volume 21. First Monday.
- C. A. Davis, O. Varol, E. F. A. F. and Menczer, F. (2016). Botornot: A system to evaluate social bots. pages 273–274. Proceedings of the 25th international conference companion on world wide web.
- J. Fernquist, L. K. and Schroeder, R. (2018). Political bots and the swedish general election. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 124–129. IEEE.
- L. Hagen, S. Neely, T. K. R. S. and Vasquez, F. E. (2020). Rise of the machines? examining the influence of social bots on a political discussion network. *Social Science Computer Review*.
- M. Bastian, S. H. and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. volume 3. Proceedings of the International AAAI Conference on Web and Social Media.
- Núcleo de Informação e Coordenação do Ponto BR (NIC.br), S. (2020). Pesquisa sobre o uso das tecnologias de informação e comunicação: pesquisa tic domicílios 2019. Cetic.br, <http://cetic.br/pt/arquivos/domicilios/2019/domicilios/>. Acesso em: 16 de jun. de 2021.
- S. Wojcik, S. Messing, A. S. L. R. and Hitlin, P. (2018). Bots in the twittersphere.
- Woolley, S. C. (2016). Automating power: Social bot interference in global politics. First Monday.