

# RELATÓRIO DE ANÁLISE

Compartilhamento de hashtags de apoio ao voto impresso

10 de agosto de 2021

No dia 29/07, quinta-feira, o presidente Jair Bolsonaro convocou a imprensa e a população para assistir a *live* que realiza semanalmente, anunciando que apresentaria provas sobre suposta fraude nas eleições de 2014 e 2018. Durante a *live*, que teve duração de aproximadamente 2 horas e foi transmitida em canal aberto na televisão, o presidente tratou de diversos assuntos e apresentou uma série de notícias inverídicas sobre o sistema eleitoral brasileiro e que já foram desmentidas por órgãos oficiais em diversos momentos. Na ocasião, o presidente admitiu não ter provas sobre risco de fraude nas urnas eletrônicas. Analisamos neste relatório a movimentação de tweets em apoio ao voto impresso antes e depois da transmissão feita pelo presidente.

**575826**

TWEETS ANALISADOS

**5499**

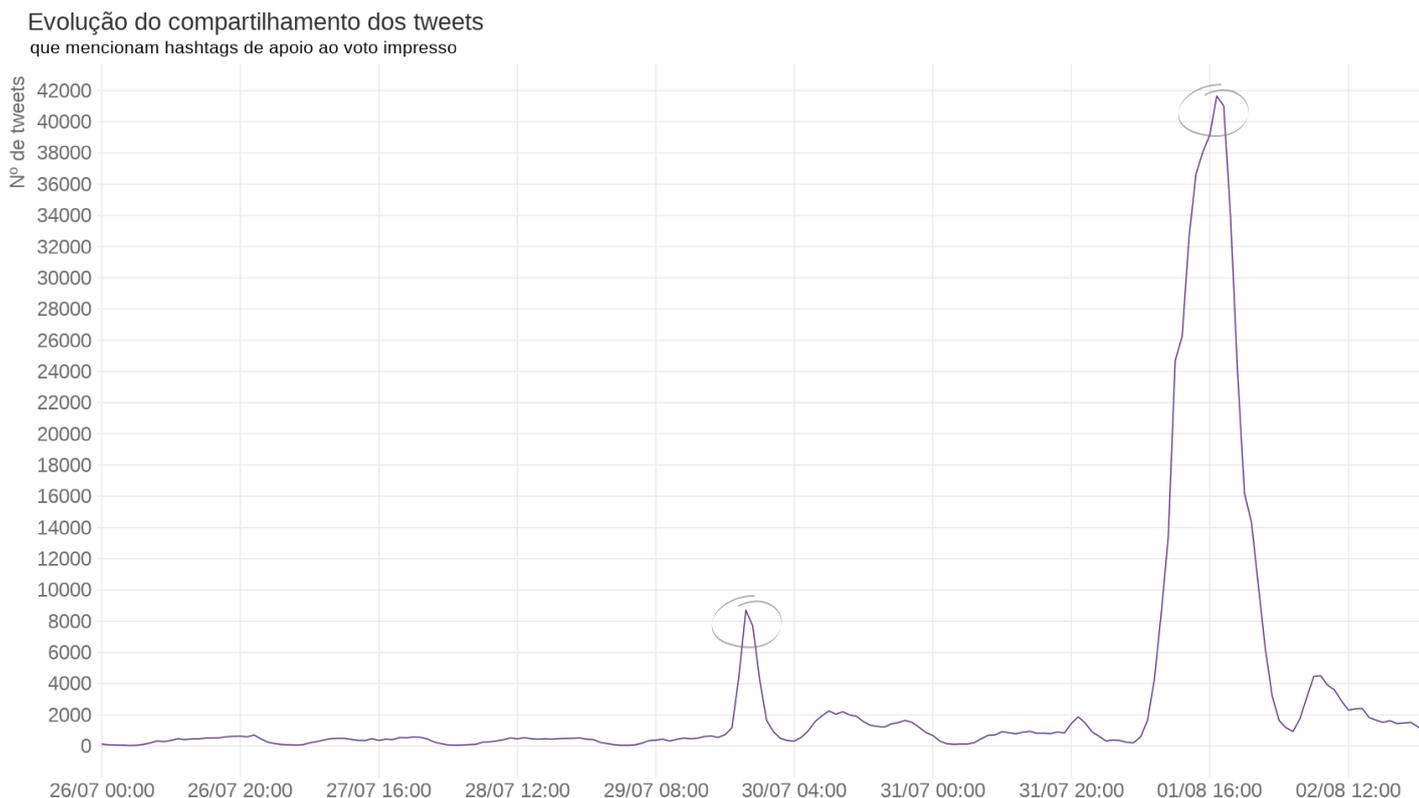
PERFIS VERIFICADOS

## Os principais pontos que você precisa saber

- 1** Coletamos tweets publicados em um período de 8 dias que mencionam ao menos uma das hashtags #VotoAuditavel, #EuApoioVotoAuditavel, #EuApoioVotoImpresso, #VotoImpressoAuditavelJa, #VotoImpressoAuditavelJa2022, #BrasilPeloVotoAuditavel e #UrnasForamInvadidas.
- 2** Verificamos um desbalanceamento expressivo entre o volume de tweets de RTs: das publicações coletadas, 78% são RTs.
- 3** Analisamos 5499 dos perfis que tiveram uma alta frequência de postagens e identificamos, com o Pegabot, que 25% deles apresentaram alta probabilidade de comportamento automatizado.
- 4** Verificamos que esses 5499 perfis foram responsáveis por 60% do volume de tweets publicados durante o período analisado. Alguns chegaram a publicar mais de 500 vezes.
- 5** Observamos um aumento no volume dos tweets publicados após a *live*, com um volume muito expressivo principalmente no dia 01/08, dia em que aconteceram manifestações em apoio ao governo.

## Sobre os tweets analisados

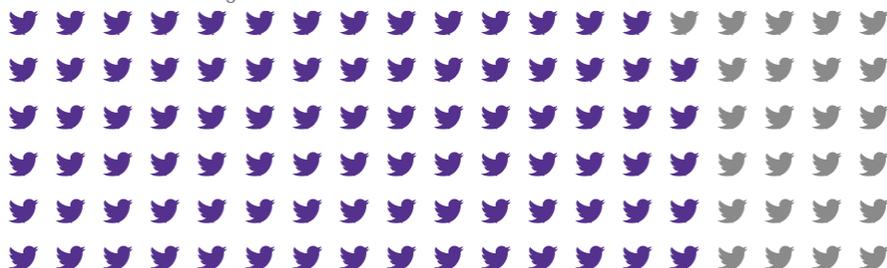
O conjunto de tweets analisados reúne publicações feitas entre os dias 26/07 e 02/08 e que mencionam ao menos uma das seguintes hashtags: #VotoAuditavel, #EuApoioVotoAuditavel, #EuApoioVotoImpresso, #VotoImpressoAuditavelJa, #VotoImpressoAuditavelJa2022, #BrasilPeloVotoAuditavel e #UrnasForamInvadidas. Esse conjunto totaliza 575826 tweets e RTs, compartilhados por 60385 usuários. A seguir apresentamos a evolução do compartilhamento desses tweets ao longo do período analisado:



O primeiro pico de compartilhamento aconteceu no dia 29/07, dia em que aconteceu a *live* realizada pelo presidente Bolsonaro. O intervalo de maior atividade relativo à esse pico ocorreu entre 19h e 00h, reunindo 27921 publicações. No dia seguinte, podemos observar um aumento no volume de publicações contendo as hashtags, se comparado aos dias que antecedem a *live*. O segundo pico, muito mais expressivo, aconteceu no dia 01/08, dia em que aconteceram manifestações em apoio ao governo. A mobilização em torno dessa manifestação aconteceu principalmente através da hashtag #Dia01VaiSerGigante, mas observamos o reflexo disso para as hashtags analisadas. O dia 01/08, sozinho, totalizou 409411 publicações contendo ao menos uma das hashtags monitoradas, o que representa 71% do volume de tweets coletados e analisados neste relatório.

### Proporção de tweets e RTs

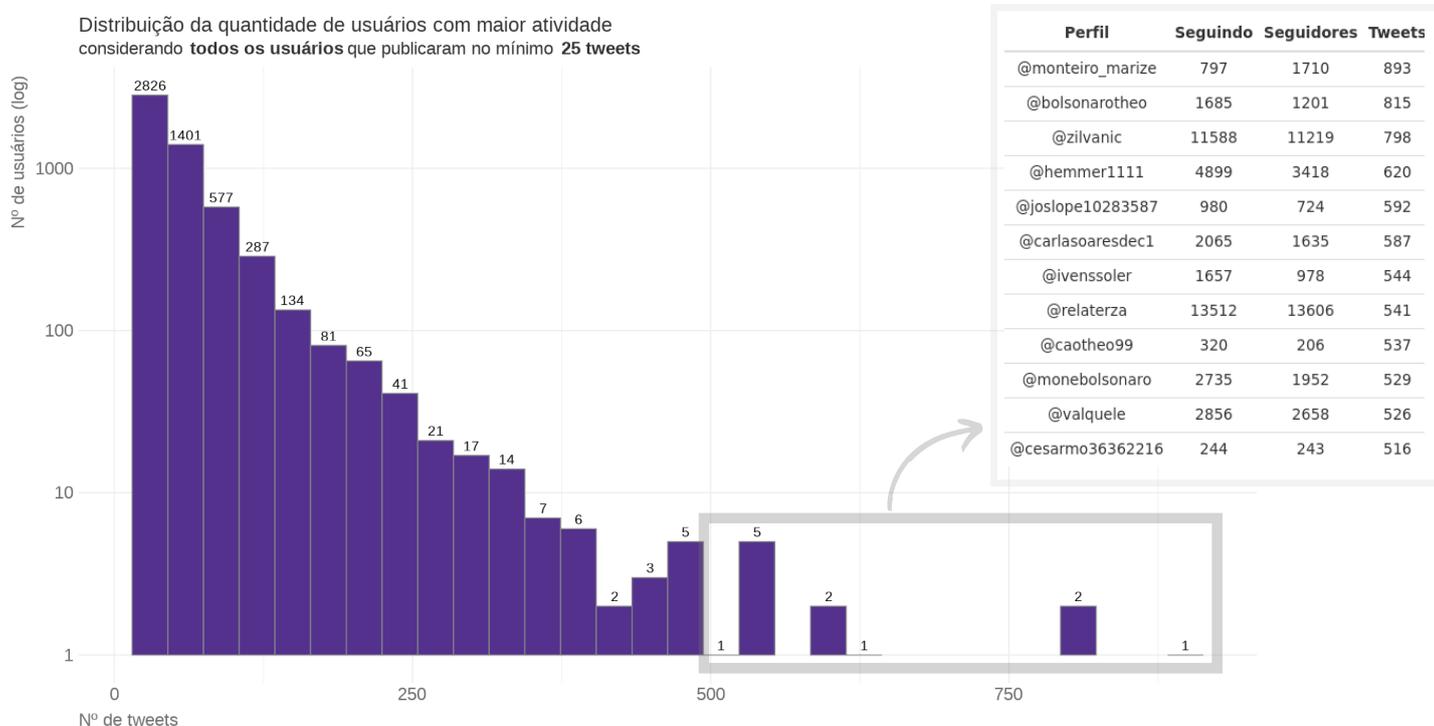
cada unidade = 5000 registros



Com frequência observamos nas nossas análises uma diferença na proporção de tweets e RTs compartilhados. No entanto, aqui o desbalanceamento entre o volume de tweets e RTs chama bastante a atenção: 78% dos registros são RTs e 22% são tweets.

## Sobre a frequência de publicações

O número de publicações que um usuário realiza no Twitter pode variar por uma série de motivos, como a finalidade da conta e número de seguidores, por exemplo. Para entender melhor sobre o comportamento dos usuários no que diz respeito ao volume de postagens, olhamos aqui para a frequência de postagens por usuários. Consideramos aqui todos os 60835 usuários que publicaram tweets mencionando as #s analisadas. Numa primeira análise, notamos um desbalanceamento na distribuição da quantidade de usuários em relação ao número de tweets que eles publicam: 79,2% dos usuários analisados fizeram até 9 publicações, valor da média de publicações por usuário de acordo com os dados coletados. No gráfico a seguir apresentamos a distribuição, onde contabilizamos o total de usuários responsável por publicar um determinado número de tweets. Filtramos os usuários que tiveram uma frequência alta de publicações, considerando o valor do desvio padrão (25) como critério para identificá-los. Verificamos, a partir dessa análise, usuários que chegaram a fazer mais de 500 publicações durante os 8 dias monitorados:



# 9%

(5499) dos usuários compartilharam

# 60%

(345896) das publicações coletadas

Ou seja, um pequeno grupo de usuários foi responsável por um volume grande de postagens. Isso reforça o desbalanceamento da distribuição de tweets compartilhados entre os usuários. Considerando ainda que o período total da coleta foi de 8 dias, a concentração de publicações em um conjunto pequeno de usuários demonstra ser bastante expressiva.

Verificamos que, desse volume de 60% das publicações, 83% são RTs. Isso é um indicador importante sobre o fato de que esses perfis estão mais voltados para a propagação de conteúdo de outros usuários. Levando em consideração todos os tweets coletados, o volume de RTs compartilhados por esse pequeno grupo de usuários corresponde à 50% do volume de publicações.

## Perfis com alta frequência de publicações têm comportamento automatizado?

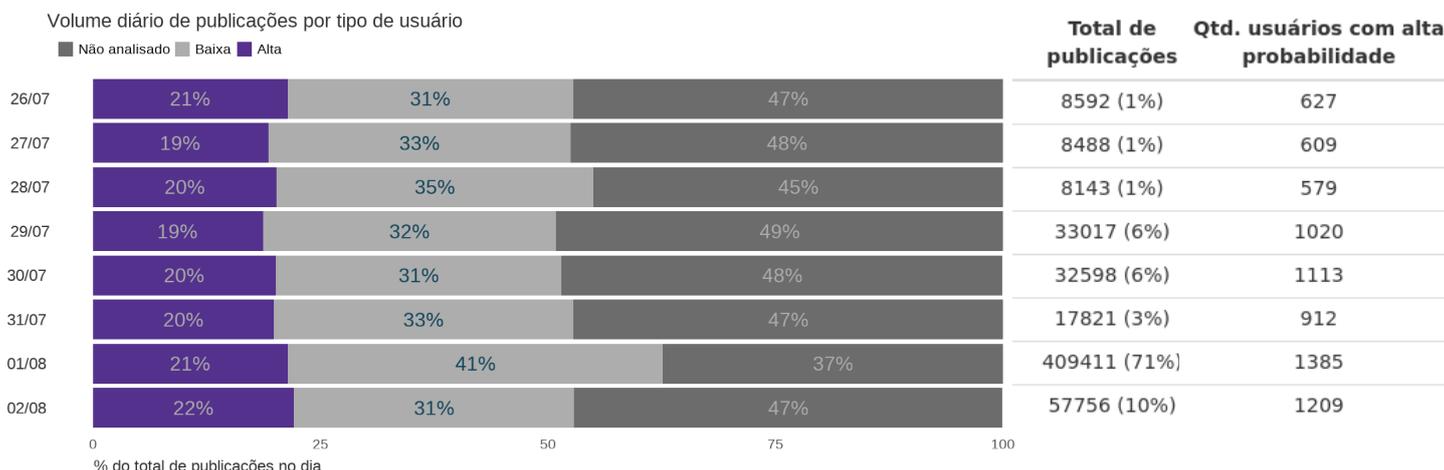
Utilizamos o Pegabot para verificar sobre a existência de comportamento automatizado nos 5499 usuários identificados com alta frequência de publicações. O Pegabot analisa, individualmente para cada perfil, os dados públicos das contas disponíveis por meio de consultas à API do Twitter. Esses dados incluem, por exemplo, nome do perfil, descrição, quantidade de perfis seguidos e seguidores, número de postagens, além de uma amostra dos tweets da linha do tempo para identificar hashtags e menções. Com base nessas informações, o Pegabot estabelece um conjunto de 4 critérios, temporal, usuário, rede e sentimento, que juntos indicam a probabilidade de comportamento automatizado de uma conta, dando uma pontuação de 0 a 100. Quanto mais alto o valor, maior a chance do perfil ser automatizado.

Do total de  
**5499**  
 usuários únicos verificados  
**25%**  
 (1401) apresentaram alta probabilidade de existência de comportamento automatizado

Esse resultado é referente a usuários que pontuaram com valor de, no mínimo, 70% de probabilidade na análise total do Pegabot. Esses perfis foram responsáveis por um volume de 21% das publicações analisadas, volume relevante considerando a quantidade de usuários.

Alguns estudos [1,2] apontam sobre o fato de que a regularidade de postagens é um importante aspecto a ser considerado no processo de identificação de perfis automatizados. Nesse sentido, um olhar focado no critério temporal, analisado pelo Pegabot, indica que a quantidade de perfis com comportamento automatizado chega a 45% dos usuários verificados. Juntos eles reúnem 32% do volume de publicações.

Buscando entender mais sobre a participação de usuários que foram analisados e que têm probabilidade de comportamento automatizado, apresentamos no gráfico a seguir o volume de publicações feito por cada tipo de usuário ao longo das semanas. A categoria 'Não analisado' diz respeito aos usuários que não passaram pelo Pegabot, por não terem sido identificados como usuários com alta frequência de publicações, e as categorias 'Baixa' e 'Alta' compreendem os usuários com baixa e alta probabilidade de comportamento automatizado, respectivamente. Listamos ao lado a quantidade total de publicações na semana, o quanto isso representa do volume total de registros e a quantidade de usuários únicos automatizados que participaram do compartilhamento de ao menos uma publicação contendo alguma das hashtags monitoradas.



Como resultados, podemos observar uma regularidade no volume de publicações feitas por usuários com probabilidade de comportamento automatizado, mesmo nos dias em que não ocorrem picos de compartilhamento. O total de publicações por dia reforça uma maior movimentação em torno do tema após a *live* realizada no dia 29. Da mesma forma, observamos quantidades maiores de usuários com probabilidade de comportamento automatizado na publicação de tweets a partir do dia 29.

## Quais outros temas aparecem junto às publicações



Verificamos a partir dos tweets publicados que outras hashtags mais frequentes nas publicações incluem ainda mais variações de hashtags em apoio ao voto impresso. Podemos observar também hashtags de ataque à imprensa (#globolixo, #imprensaliço), de apoio ao governo e às manifestações realizadas no dia 01, que impulsionaram as publicações em apoio ao voto impresso nesse dia.

## Tweets que mais repercutiram

Listamos a seguir os 10 tweets que foram mais retuitados, de acordo com os dados coletados. Esses 10 tweets que tiveram mais RTs totalizam 6,5% do volume de publicações. Além disso, 87% dos usuários analisados e que possuem alta probabilidade de automação de acordo com a análise total do Pegabot participaram do compartilhamento desses RTs.

Total de RTs	Perfil	Link
5367	@nikolas_dm	<a href="https://twitter.com/nikolas_dm/status/1421864981302288386">https://twitter.com/nikolas_dm/status/1421864981302288386</a>
4826	@CarlaZambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1421929398308155392">https://twitter.com/CarlaZambelli38/status/1421929398308155392</a>
4813	@taoquei1	<a href="https://twitter.com/taoquei1/status/1421874214055854085">https://twitter.com/taoquei1/status/1421874214055854085</a>
3495	@PATRIOTAS	<a href="https://twitter.com/PATRIOTAS/status/1421938308016443392">https://twitter.com/PATRIOTAS/status/1421938308016443392</a>
3447	@CarlaZambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1421846008284291074">https://twitter.com/CarlaZambelli38/status/1421846008284291074</a>
3291	@Rconstantino	<a href="https://twitter.com/Rconstantino/status/1421921671154438146">https://twitter.com/Rconstantino/status/1421921671154438146</a>
3277	@MinLuizRamos	<a href="https://twitter.com/MinLuizRamos/status/1421933789786300429">https://twitter.com/MinLuizRamos/status/1421933789786300429</a>
3163	@CarlaZambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1421875188833177601">https://twitter.com/CarlaZambelli38/status/1421875188833177601</a>
3127	@BrunoEnglerDM	<a href="https://twitter.com/BrunoEnglerDM/status/1421837365732122627">https://twitter.com/BrunoEnglerDM/status/1421837365732122627</a>
2841	@CarlaZambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1421902675612950531">https://twitter.com/CarlaZambelli38/status/1421902675612950531</a>

## Considerações finais

Trouxemos neste relatório um levantamento sobre a existência de comportamento automatizado na veiculação de tweets que mencionam hashtags de apoio ao voto impresso. Consideramos para análise no Pegabot os perfis com alta frequência de postagens e que foram responsáveis por um volume expressivo das publicações coletadas. Destacamos que, assim como qualquer ferramenta de detecção de comportamento automatizado, o Pegabot tem limitações especialmente por conta das técnicas utilizadas para automação de perfis no Twitter serem atualizadas com frequência, impedindo que determinadas contas sejam identificadas pelo algoritmo. Da mesma forma, o Pegabot tem passado por um processo constante de avaliação e ajustes, para se manter atualizado frente às novas estratégias adotadas por esses perfis automatizados. Por esse motivo, procuramos não apontar sobre esses perfis de forma direta, mas sim entender sobre o uso de automação de forma mais ampla, identificando comportamentos que fogem do esperado ou observado.

## Referências

- [1] Freitas, Carlos, et al. "Reverse engineering socialbot infiltration strategies in twitter." 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2015.
- [2] Chu, Zi, et al. "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?." IEEE Transactions on dependable and secure computing 9.6 (2012): 811-824.

## Glossário

Frequência de tweets: indica quantos usuários publicaram um determinado número de tweets.

RT: republicação de um tweet próprio ou de outra pessoa que, para o relatório, contenha algum dos termos consultados na coleta de dados.

Reply: resposta dada à um tweet que, para o relatório, contenha algum dos termos consultados na coleta de dados.