

# RELATÓRIO DE ANÁLISE

Compartilhamento da hashtag #Dia7VaiSerGigante

01 de setembro de 2021

No próximo dia 7 de setembro, Dia da Independência, estão previstas manifestações em apoio ao governo federal em todo o país. Apoiadores do presidente Jair Bolsonaro têm utilizado as redes para realizar convocações e mobilizar os usuários através da hashtag #Dia7VaiSerGigante e algumas variações. Para analisar sobre a existência de comportamento automatizado entre os perfis que têm se posicionado a favor à essas manifestações, coletamos e analisamos neste relatório tweets publicados entre os dias 22/08 e 30/08, seguindo o mesmo molde do relatório anterior para fins de comparação e acompanhamento.

**354.256**

TWEETS ANALISADOS

**48.586**

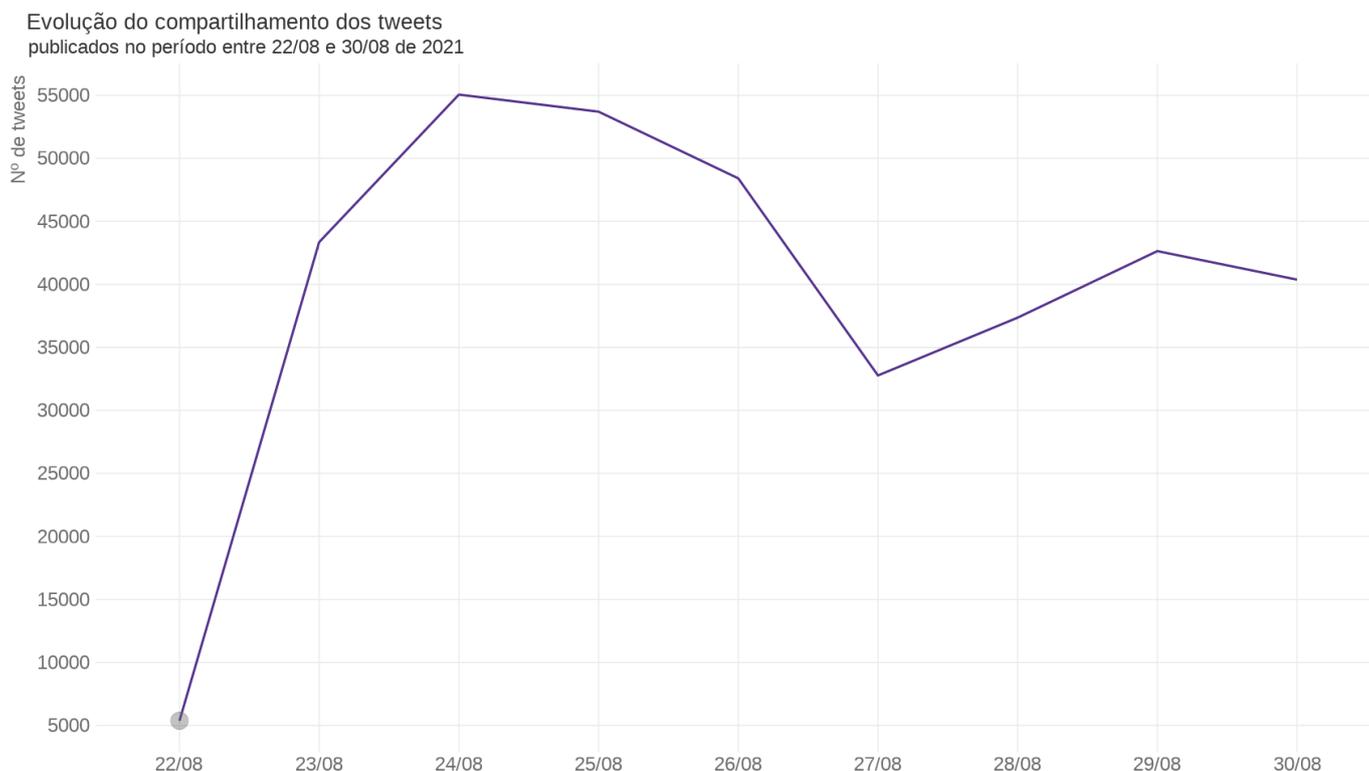
PERFIS VERIFICADOS

## Os principais pontos que você precisa saber

- 1** Verificamos um desbalanceamento entre o volume de tweets de RTs um pouco menor se comparado ao relatório anterior, mas ainda expressivo: das publicações coletadas, 67% são RTs.
- 2** Verificamos que 9% dos perfis foram responsáveis por 55% do volume de tweets publicados durante o período analisado. Um deles chegou a publicar mais de 1400 vezes.
- 3** Analisamos os perfis usando o Pegabot e identificamos que 5% deles apresentaram alta probabilidade de comportamento automatizado. Juntos, eles publicaram 23% dos tweets.
- 4** Considerando a data de criação dos perfis analisados, verificamos um aumento na criação de contas com comportamento automatizado nos últimos anos: 86,3% das contas identificadas com comportamento automatizado foram criadas a partir de 2017.
- 5** Identificamos que, entre os tweets coletados e sem considerar RTs, apenas 57% possuem conteúdo único e original. Os demais 43% são utilizados apenas para publicação das hashtags monitoradas com marcações à outros usuários ou com repetição no conteúdo publicado.

## Sobre os tweets analisados

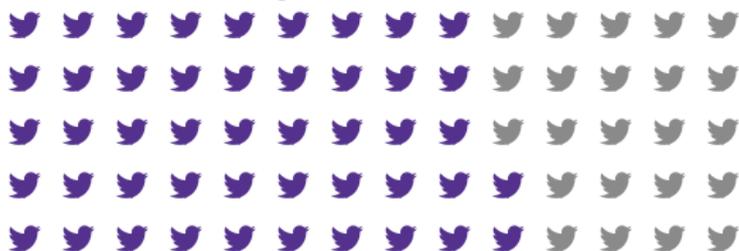
O conjunto de tweets analisados reúne publicações feitas entre os dias 22/08 e 30/08 e que mencionam ao menos uma das seguintes hashtags: #Dia07VaiSerGigante, #Dia7VaiSerGigante, #Dia07VaiSerMaior e #Dia07VaiSerMaior. Esse conjunto totaliza 354.256 tweets e RTs, compartilhados por 48.586 usuários. A seguir apresentamos a evolução do compartilhamento desses tweets ao longo do período analisado:



Para o dia 22 não foi possível recuperar um volume maior de tweets por uma limitação do mecanismo de coleta, e os dados para esse dia cobrem parte dos tweets publicados a partir das 18h10. Podemos observar um volume maior de publicações no dia 24/08. Diferente do primeiro relatório que produzimos sobre as mesmas hashtags em um período anterior, em que os picos de compartilhamento estavam associados à episódios envolvendo o cantor Sérgio Reis e o caminhoneiro Zé Trovão, aqui não identificamos nenhum evento específico associado ao dia 24/08 que tenha mobilizado uma movimentação maior. No dia 25/09 a leve alta nos compartilhamentos foi impulsionada por um tweet feito pelo ex-ministro do Meio Ambiente, Ricardo Salles e que reforça sobre o apoio aos atos pró-governo previstos para o dia 7/09.

## Proporção de tweets e RTs

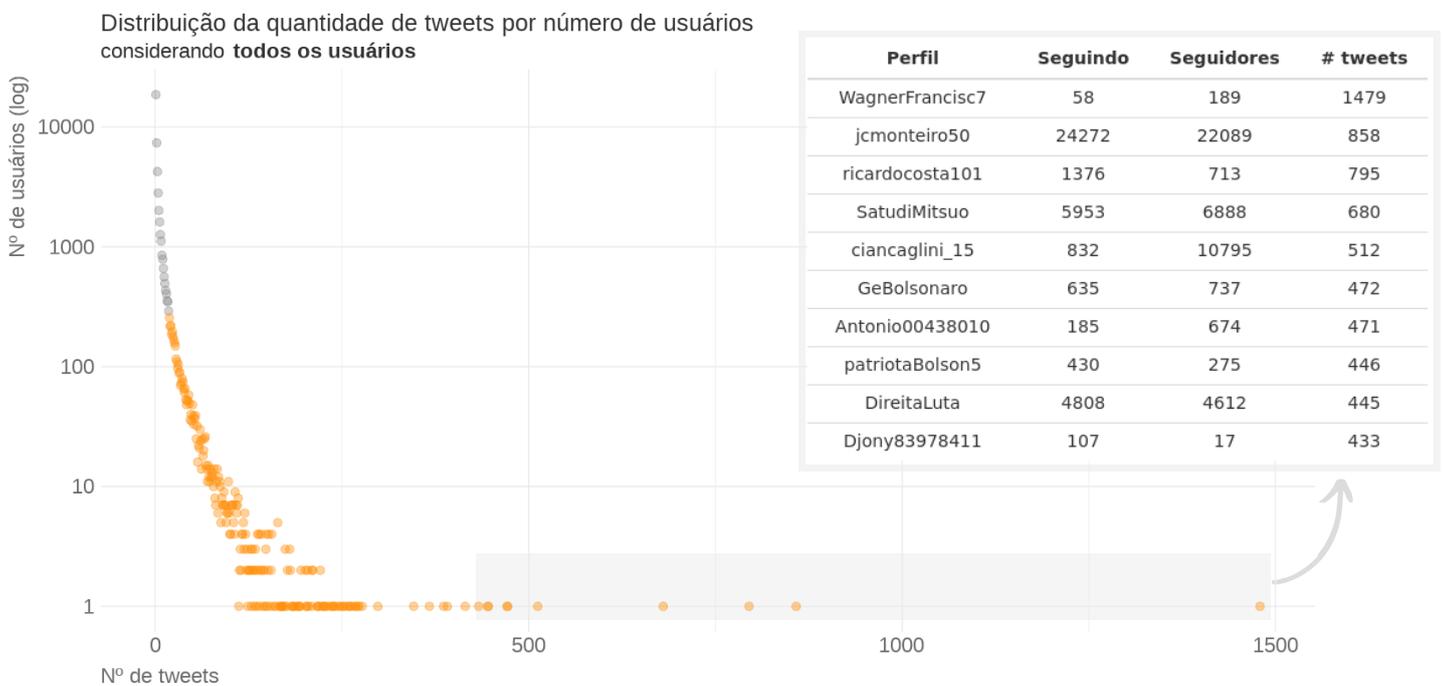
cada unidade = 5000 registros



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 foi um pouco menor em relação ao relatório anterior, mas continua sendo expressivo: 67% dos registros são RTs e 33% 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 48.586 usuários que publicaram tweets mencionando as #s analisadas. Numa primeira análise, notamos um desbalanceamento na distribuição do número de tweets em relação a quantidade de usuários: 78,1% dos usuários analisados fizeram em média até 7 publicações. No gráfico a seguir apresentamos essa distribuição. Estabelecemos o desvio padrão como um valor de corte onde, usuários que publicaram acima deste valor (19 tweets) são considerados usuários com alta frequência de postagens, destacados em laranja. Verificamos, a partir dessa análise, usuários que chegaram a fazer mais de 400 publicações durante os 8 dias monitorados:



# 9%

(4326) dos usuários compartilharam

# 55%

(194408) 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 55% das publicações, 63% são RTs. Isso é um indicador de que esses perfis estão mais voltados para a propagação de conteúdo de outros usuários, volume próximo ao identificado também no primeiro relatório sobre essas #s. Levando em consideração todos os tweets coletados, o volume de RTs compartilhados por esse pequeno grupo de usuários corresponde à 35% do volume de publicações.

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

Utilizamos o [Pegabot](#) para verificar sobre a existência de comportamento automatizado por parte dos usuários que participaram das 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.

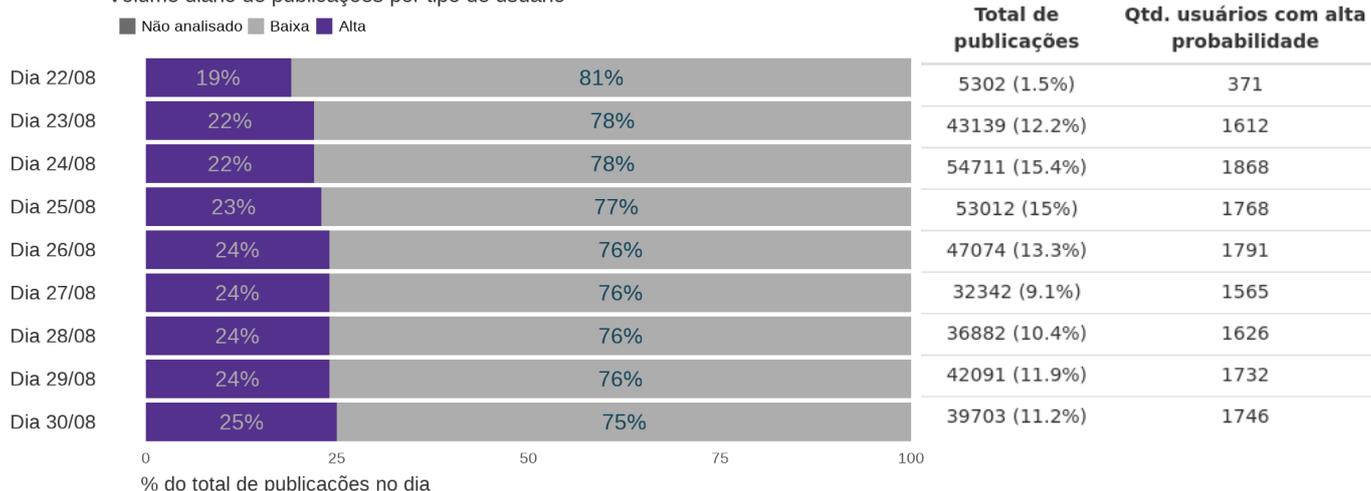
Do total de  
**48.586**  
 usuários únicos verificados  
**5%**  
 (2621) apresentaram alta  
 probabilidade de existência de  
 comportamento automatizado

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 23% das publicações analisadas, volume relevante considerando a quantidade de usuários. 78,3% desse volume corresponde a RTs.

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 13,4% dos usuários verificados. Juntos eles reúnem 37% 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.

Volume diário de publicações por tipo de usuário





## O que mais encontramos no conteúdo dos tweets

Além de identificar quais hashtags aparecem junto às hashtags monitoradas, procuramos entender mais sobre o conteúdo compartilhado nos tweets e o quão diverso é esse conteúdo, ou se ele é sempre muito repetitivo. Para isso, desconsideramos os RTs e aplicamos um processo de limpeza em todos os tweets para remoção de menções, hashtags, urls e emojis. Com isso verificamos que, dos 118.231 registros coletados que compreendem tweets nativos, 57% são tweets com conteúdo único e original, ou seja, não ocorrem mais de uma vez entre os tweets analisados. Além disso, após a limpeza identificamos que 26% dos tweets não contém nenhum outro tipo de conteúdo a não ser menção com marcação de alguma hashtag, ou seja, que não contém nenhum outro tipo de texto.

Conteúdo	Repetições
tmj	400
eu vou	376
com certeza	265
eu apoio o presidente bolsonaro	261
seguindo	255
bora	235
fora stf lixos	185
bom dia	159
deus pátria família	137

Os demais 17% compreendem tweets cujo conteúdo aparece mais de uma vez, podendo ser considerado um indicador importante de uso de técnicas de automação. Listamos ao lado os 10 conteúdos que mais se repetem e a quantidade de vezes em que eles ocorrem.

Identificamos ainda, com a lista mais extensa de repetições, alguns tweets com conteúdos em inglês, indicando uma tentativa de impulsionar o tema fora do país. Parte desses tweets incluem um link para o YouTube, com um vídeo também elaborado em inglês.

## 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 6,3% do volume de publicações. Além disso, 59,6% 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
3351	@rsallesmma	<a href="https://twitter.com/rsallesmma/status/1432091049938366465">https://twitter.com/rsallesmma/status/1432091049938366465</a>
2507	@carlazambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1430219676819283968">https://twitter.com/CarlaZambelli38/status/1430219676819283968</a>
2355	@liberdade_ldr	<a href="https://twitter.com/liberdade_ldr/status/1430870773397397505">https://twitter.com/liberdade_ldr/status/1430870773397397505</a>
2333	@carlazambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1429887303930032138">https://twitter.com/CarlaZambelli38/status/1429887303930032138</a>
2217	@bernardokuster2	<a href="https://twitter.com/bernardokuster2/status/1430746436698841088">https://twitter.com/bernardokuster2/status/1430746436698841088</a>
2169	@carlazambelli38	<a href="https://twitter.com/CarlaZambelli38/status/1430154297392734211">https://twitter.com/CarlaZambelli38/status/1430154297392734211</a>
2007	@kellypstorres	<a href="https://twitter.com/kellypstorres/status/1430314565380812803">https://twitter.com/kellypstorres/status/1430314565380812803</a>
1843	@chiconeto12345	<a href="https://twitter.com/chiconeto12345/status/1431427847273451525">https://twitter.com/chiconeto12345/status/1431427847273451525</a>
1832	@Laurinhalronic	<a href="https://twitter.com/Laurinhalronic/status/1430216353621893124">https://twitter.com/Laurinhalronic/status/1430216353621893124</a>
1796	@mjmacul_lima	<a href="https://twitter.com/mjmacul_lima/status/1429947338706100228">https://twitter.com/mjmacul_lima/status/1429947338706100228</a>

## Considerações finais

Trouxemos neste relatório um segundo levantamento sobre a existência de comportamento automatizado na veiculação de tweets que mencionam hashtags sobre a manifestação em apoio ao governo federal, que vai acontecer no dia 7 de setembro. Consideramos para análise no Pegabot todos os perfis que participaram das publicações coletadas. Em relação ao relatório anterior, observamos um aumento no número de tweets mencionando as hashtags e também de usuários que participaram desses compartilhamentos. A partir dos resultados do Pegabot, identificamos 326 novos usuários com comportamento automatizado. O volume de tweets mobilizado por eles é proporcionalmente bastante similar ao identificado no primeiro levantamento, embora aqui tenhamos analisado um volume geral maior. Além disso, observamos entre os dias um aumento no número de usuários com comportamento automatizado que publicaram tweets em comparação ao relatório anterior.

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.