

Estudos de caso de análise de perfis de usuários agrupados por *hashtags* no Twitter

Pedro H. E. C. Costa¹, Jefferson R. Lima¹, Ronaldo A. Marques¹,
Daniel R. Trindade¹, Karin S. Komati¹

¹Coordenação de Informática, Campus Serra do IFES

pedroh-costa@hotmail.com, {ronaldomarques, jlima, kkomati}@ifes.edu.br

Abstract. *The main goal of this work was to identify the presence of bot profiles in the dissemination of a specific hashtag. This work uses two hashtags that were shared on the social network Twitter. The first one has strong evidence of being disseminated by bots. The second one has no indication of such usage and is used for the purpose of comparison. It was possible to evaluate that the hashtag still presents profiles of bots, even months after it was on trend topics.*

Resumo. *O objetivo principal do estudo foi analisar os perfis de usuários, se são ou não bots, na disseminação de uma determinada hashtag na rede social Twitter. Este trabalho apresenta o estudo de caso de duas hashtags que foram compartilhadas: uma que possui indícios de serem disseminadas por bots, e a outra não, sendo utilizada com o objetivo de comparação. Foi possível avaliar que a hashtag com indícios de ter sido impulsionada por robôs ainda apresenta perfis de usuários bots, mesmo meses depois da mesma ter sido uma trend.*

1. Introdução

O termo *bot* se refere a uma conta automatizada, que pode ter usos benignos ou malignos [Araújo et al. 2018]. Dentre as várias redes sociais, os *bots* dominam mensagens no Twitter. O Twitter é uma rede social que permite aos usuários criar um perfil público e interagir com outras pessoas por meio das mensagens publicadas [Suter et al. 2019]. O Twitter utiliza as *hashtags* (formada pelo símbolo '#' seguido do termo em questão sem espaços) para organizar informações, como uma forma de tornar possível o resgate de tuites sobre um determinado tópico, facilitando a busca de comentários sobre o tema. Os temas mais tuitados são inseridos nos *Trending Topics*, que é uma lista de assuntos mais comentados. Um relatório do instituto independente Pew Research Center¹ indica que 66% das mensagens do Twitter que contêm *links* são publicadas por *bots*. O esforço do próprio Twitter de remover as contas falsas fez com que 70 milhões de contas fossem suspensas² entre os meses de maio e junho de 2018. Mesmo assim, avalia-se que cerca de 15% dos usuários do Twitter, dentre os cerca de 335 milhões de usuários ativos por mês³, são *bots*.

Este trabalho propõe a avaliação de perfis de usuários por *hashtag* na rede social Twitter. A contribuição desta ferramenta é que a análise é feita por *hashtag* e não por

¹<https://www.pewresearch.org/internet/2018/04/09/bots-in-the-tweetsphere/>

²<https://variety.com/2018/digital/news/twitter-locked-accounts-removed-followers-1202870341/>

³<https://kinsta.com/pt/blog/marketing-no-twitter/>

perfil de usuário, mesmo usando uma ferramenta já existente que analisa cada perfil. Nos estudos de caso deste trabalho foram usadas duas *hashtags*: #BolsonaroDay e #DiaDaTolha. A primeira *hashtag* foi escolhida por já ter estudo que reportou o impulsionamento por *bots* [Kalil and Santini 2020]. O segundo estudo de caso é uma *hashtag* de controle, em que não se identificou algum tipo de ganho pelo seu uso, seja ele de cunho político ou econômico. Foram coletados os tuítes de cada *hashtag* de um determinado período de tempo e foram analisados os perfis que a usaram/compartilharam. Posteriormente, a análise dos perfis foi feita com a ferramenta PEGABOT⁴, que fornece uma probabilidade desse perfil ser um *bot*.

2. Trabalhos Correlatos

O trabalho de Nobre, Almeida e Ferreira [Nobre et al. 2019] analisou os usuários do Twitter durante as eleições presidenciais de 2018 no Brasil. O trabalho coletou cerca de 40 milhões de tuítes que continham o nome de um dos quatorze candidatos à presidência e classificou mais de 100 mil usuários da base coletada via Botometer. Uma das conclusões é que foram identificados mais de 2 mil perfis robôs. Embora percentualmente baixo (cerca de 2%), tais perfis foram eficientes em influenciar usuários comuns.

A *hashtag* #BolsonaroDay surgiu no dia 15 de março de 2020, dia marcado por manifestações pró-governo. Um levantamento analisou 66 mil contas que, sozinhas, foram responsáveis por cerca de 1,2 milhão de postagens de apoio aos atos pró-governo. Pelo menos 55% dos tuítes com a *hashtag* #BolsonaroDay publicados no dia das manifestações pró-governo vieram de robôs [Kalil and Santini 2020]. A pesquisa aponta que os robôs que usaram a *hashtag* #BolsonaroDay publicaram, em média, 700 tuítes apenas no dia 15 de março. As contas automatizadas mais ativas chegaram a postar mais de 1,2 mil tuítes nesta data. No trabalho foi usado o software Gotcha.

O trabalho de Paulino, Empinotti e Ventura [Paulino et al. 2020] usou Técnicas de Análise de Redes Sociais para analisar tuítes associados à pergunta “Por que Fabrício Queiroz depositou R\$ 89 mil na conta de Michelle Bolsonaro?”. A coleta de dados foi feita com a ferramenta Netlityc, usando inicialmente a pesquisa com as palavras-chave “Michelle Bolsonaro OR Bolsonaro” do dia 24/08/2020. Após esta pesquisa inicial, os pesquisadores descobriram uma alta correlação positiva com a *hashtag* #BolsonaroOrgulhoDoBrasil, e portanto aumentaram a base de dados com estes novos tuítes. Com a base de dados de tuítes, fizeram a análise de rede com o foco nos *clusters* e centralidade de graus usando a ferramenta Gephi. Foram identificados 4 nós que se destacaram pela quantidade de tuítes e pelo número de seguidores, que foram analisados pelo sistema Botometer que evidenciou a baixa atividade humana nestes perfis.

3. Materiais e Métodos

Para construção da base de dados de cada *hashtag*, as seguintes etapas foram desenvolvidas: (i) desenvolvimento de código para coleta de tuítes (ii) análise via ferramenta PEGABOT (uma versão brasileira inspirada no Botometer) e (iii) análise geral por *hashtag*. Para a coleta de dados, o sistema recebe como entrada: a *hashtag* e datas de início e fim de busca dos tuítes. Para cada perfil de usuário, é feita a análise com a ferramenta PEGABOT, armazenando sua resposta.

⁴<https://pegabot.com.br/>

O PEGABOT analisa o histórico de postagens de um perfil (o nome do perfil, texto da descrição do perfil, idade do perfil, foto do perfil, dentre outros) e se baseia em padrões de comportamento (quantidade de perfis seguidos e seguidores, número de tuítes, linguagem usada nos tuítes e favoritos) para identificar se é mais provável que o perfil pertença a um humano ou a um robô. O PEGABOT indica a probabilidade do perfil, no intervalo de 0 a 100, dividido em três faixas: provavelmente humano (até 40), provavelmente robô (maior que 60) e uma faixa neutra (entre 40 e 60). Com os resultados, são analisadas: a distribuição por faixa dos perfis, a distribuição dos perfis com mais de 10 publicações e a distribuição dos perfis com maior frequência de retuíte.

4. Resultados dos estudos de caso

Nesta seção são apresentados e discutidos os resultados obtidos através da coleta dos tuítes e do seu processamento. A coleta de dados se deu no segundo semestre de 2020.

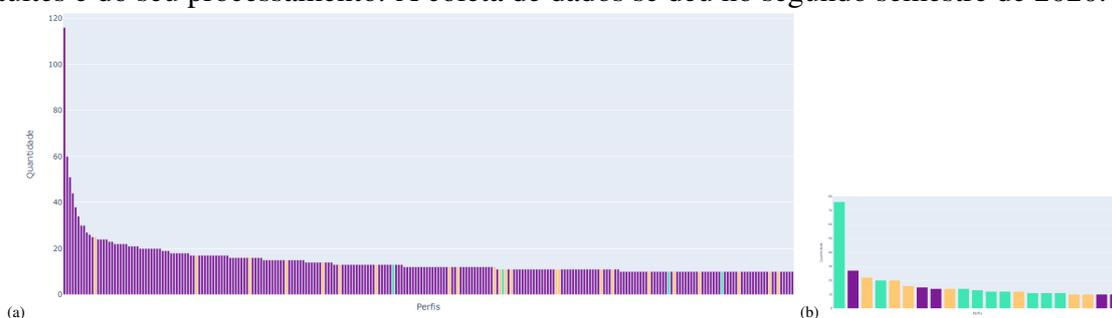


Figura 1. Perfis com mais de 10 publicações (a) Estudo de caso 1 (b) Estudo de caso 2.

Tabela 1. Tabela com os 10 perfis com maior frequência de retuítes (a) Estudo de Caso 1 (b) Estudo de Caso 2

Perfil – Caso 1	Função	Status	PEGABOT	Perfil – Caso 2	Função	Status	PEGABOT
Denise13duncan	DireitaSegueDireita	Ativo	53%	stephenfry	Ator	Ativo	17%
KARL.DAVID.2010	Armamentista	Ativo	66%	erico.borgo		Ativo	21%
carlagcc1967	AliançaPeloBrasil	Ativo	88%	FelipoZarudo	Criador de conteúdo	Ativo	27%
AyurvedaPoa	SomosTodosBOLSONARO	Ativo	79%	cookout	Hambúrgueres	Ativo	13%
MClima9	Bolsonarista	Ativo	97%	ArthurIstari		Ativo	22%
Betobraga22	Direita conservadora	Ativo	62%	ri.science	Instituição de caridade	Ativo	25%
Jouberth19	Cristão, Conservador e Patriota	Ativo	38%	BabuSantana	Ator e Cantor	Ativo	28%
PauloCzardasil2		Conta suspensa		omelete	Site de entretenimento	Ativo	18%
Marcuspitter		Conta suspensa		DannyDutch	Apaixonado por arte	Ativo	21%
Beatriz62890784	BrasilacimadetudoeDeusacimadetodos	Ativo	68%	vivicamilizer		Conta Inexistente	

Para o Estudo de Caso 1, os parâmetros de entrada foram: #BolsonaroDay e data de início e final em “15/03/2020”. Foram coletados 28.988 tuítes e deles extraídos 10.816 perfis não repetidos. No geral, os perfis considerados como robôs foram de 45,9% do total, como humanos 16,2% e neutros 38,0%. O gráfico da Figura 1(a) apresenta os perfis com mais de 10 publicações. A cor roxa representa os robôs, a amarela os neutros e verde para os humanos. Em termos percentuais: 91,3% foram classificados como robôs, apenas 1,16% como humanos e 7,52% de perfis neutros. A Tabela 1(a) resume os 10 perfis que mais retuitaram: 2 contas estão suspensas (PauloCzardasil2 e Marcuspitter), 6 são robôs, apenas 1 é humano (Jouberth19) e apenas 1 é neutro (Denise13duncan).

O Estudo de Caso 2 é a hashtag #DiaDaToalha, que é usada em comemoração ao dia 25 de maio pelos fãs da série de ficção científica “Guia do Mochileiro das Galáxias”, escrita por Douglas Adams. Para a data de início em “25/05/2019” e data final em “25/05/2020” foram coletados 13.298 tuítes e deles extraídos 10.896 perfis não repetidos. No geral, os perfis classificados como robôs foram de 3,45%, humanos 81,7% e

neutros de 14,9%. O gráfico da Figura 1(b) tem a cor predominante verde, o que indica que a maioria dos perfis que publicaram mais são perfis de humanos: 50,0% são humanos, 21,1% robôs e 28,9% neutros. Dos 10 usuários que mais retuitaram, Tabela 1(b), apenas um não existe mais, e dos 9 ativos, todos são considerados de humanos.

5. Considerações Finais

O trabalho apresentou dois estudos de casos de perfis no Twitter para identificar a atuação de *bots* na propagação de uma determinada *hashtag*, mesmo após o assunto deixar de figurar entre os assuntos mais comentados. Uma *hashtag* de cunho político com indícios de ser impulsionada por robôs e a outra não. Os resultados mostraram que mesmo meses depois do dia em que a *hashtag* #BolsonaroDay foi tendência, ainda é possível confirmar as informações dos estudos feitos na época [Kalil and Santini 2020]. De forma geral, mais de 45% dos perfis analisados têm indicativo de serem robôs, de acordo com a ferramenta PEGABOT. Ao analisar os perfis que tinham mais de 10 tuítes com a mesma *hashtag*, esse percentual sobe para mais 91%. Dentre esses perfis que mais retuitaram, 6 dentre 10 foram indicados como robôs. Já a *hashtag* #DiaDaToalha apresentou mais de 81% de perfis humanos, bem como metade dos perfis que tinham mais de 10 tuítes com a mesma *hashtag*. Dentre os perfis que mais retuitaram, 9 dentre 10 foram indicados como humanos. Como trabalhos futuros, pretende-se publicar a ferramenta de forma *online* e pública, pretende-se usar outras ferramentas de análise de perfil e validar alguma amostra dos resultados para confirmar a efetividade da ferramenta.

Referências

- Araújo, Y., Chares, A. C., and de Oliveira Sampaio, J. (2018). Identificação de fake news: uma abordagem utilizando métodos de busca e chatbots. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*, Porto Alegre, RS, Brasil. SBC.
- Kalil, I. and Santini, R. M. (2020). Coronavírus, pandemia, infodemia e política. Technical report. Relatório de pesquisa. Divulgado em 01 de abril de 2020. São Paulo/ Rio de Janeiro: FESPSP/ UFRJ.
- Nobre, G., Almeida, J., and Ferreira, C. (2019). Caracterização de bots no twitter durante as eleições presidenciais no Brasil em 2018. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 107–118, Porto Alegre, RS, Brasil. SBC.
- Paulino, R. d. C. R., Empinotti, M. L., and Ventura, M. P. (2020). A pergunta que o Brasil fez, mas não obteve resposta sob a ótica das mídias sociais¹. In *Anais do 43º Congresso Brasileiro de Ciências da Comunicação*, UFBA, BA. Sociedade Brasileira de Estudos Interdisciplinares da Comunicação.
- Suter, J., Nogueira, R., Tozzi, T., Anderle, D., and Speroni, R. (2019). Um data warehouse baseado no twitter para análise de sentimento em língua portuguesa: Estudo de caso das eleições de 2018. In *Anais da XV Escola Regional de Banco de Dados*, pages 41–50, Porto Alegre, RS, Brasil. SBC.