

# Uma análise da rede dos bots no Twitter no contexto das eleições presidenciais brasileiras de 2018

Gustavo Germano {gustavog@dcc.ufmg.br}

Fabricio Murai {murai@dcc.ufmg.br}

Marcus Abilio Pereira {marcusabilio@fafich.ufmg.br}

## Resumo

A utilização e o impacto dos bots nas redes sociais no contexto eleitoral tem crescido nos últimos anos. Para analisar esse impacto e entender melhor o *modus operandi* dos bots no contexto eleitoral brasileiro de 2018, analisamos as principais características dos bots que *tweetam* sobre política brasileira e de sua rede de influência no Twitter.

## MOTIVAÇÃO

A utilização de bots nas redes sociais como ferramenta política se tornou padrão nas últimas eleições. Estudos mostram influência direta dessa estratégia nas eleições presidenciais estadunidenses de 2016 e no referendo do *Brexit*. Assim, buscamos caracterizar neste trabalho os bots que *tweetaram* sobre as eleições presidenciais brasileiras de 2018 e, principalmente, suas rede de influência.

## COLETA DE DADOS

01

Coleta de hashtags de política no Twitter.

- Observação de hashtags populares relacionadas com as eleições presidenciais brasileiras no Twitter.
- Coleta dessas hashtags.

02

Coleta de usuários e classificação entre bots/não-bots.

- Coleta de usuários que *twittaram* com as hashtags escolhidas.
- Classificação dos usuários entre bots/não-bots através da API *Botometer*.

03

Construção da rede e amostragem através do Frontier Sampling.

- Utilização dos bots como vértices iniciais e seus seguidores como arestas de uma rede.
- Obtenção de uma amostra de usuários dessa rede através do algoritmo Frontier Sampling.

### Hashtags coletadas:

#Bolsonaro17, #bolsonaroCagao, #BolsonaroNao, #Eleições2018, #BolsonaroSim, #EleNao, #EleSim, #ViraViraCIRO, #DebateNaBand, #DebateSBT, #ForaPT, #DebateNaGlobo, #HaddadÉLula, #DebateNaRecord, #HaddadSim, #VemProDebate.

### Frontier Sampling:

#### Algorithm 1: Frontier Sampling (FS).

```

1:  $n \leftarrow 0$  { $n$  is the number of steps}
2: Initialize  $L = (v_1, \dots, v_m)$  with  $m$  randomly chosen vertices (uniformly)
3: repeat
4:   Select  $u \in L$  with probability  $\deg(u) / \sum_{v \in L} \deg(v)$ 
5:   Select an outgoing edge of  $u$ ,  $(u, v)$ , uniformly at random
6:   Replace  $u$  by  $v$  in  $L$  and add  $(u, v)$  to sequence of sampled edges
7:    $n \leftarrow n + 1$ 
8: until  $n \geq B - mc$ 

```

Onde  $L$  é o conjunto de vértices iniciais.

Com isso, coletamos cerca de 2000 usuários, que foram utilizados nas análises a seguir:

## ANÁLISE DOS DADOS

### Abordagem utilizada:

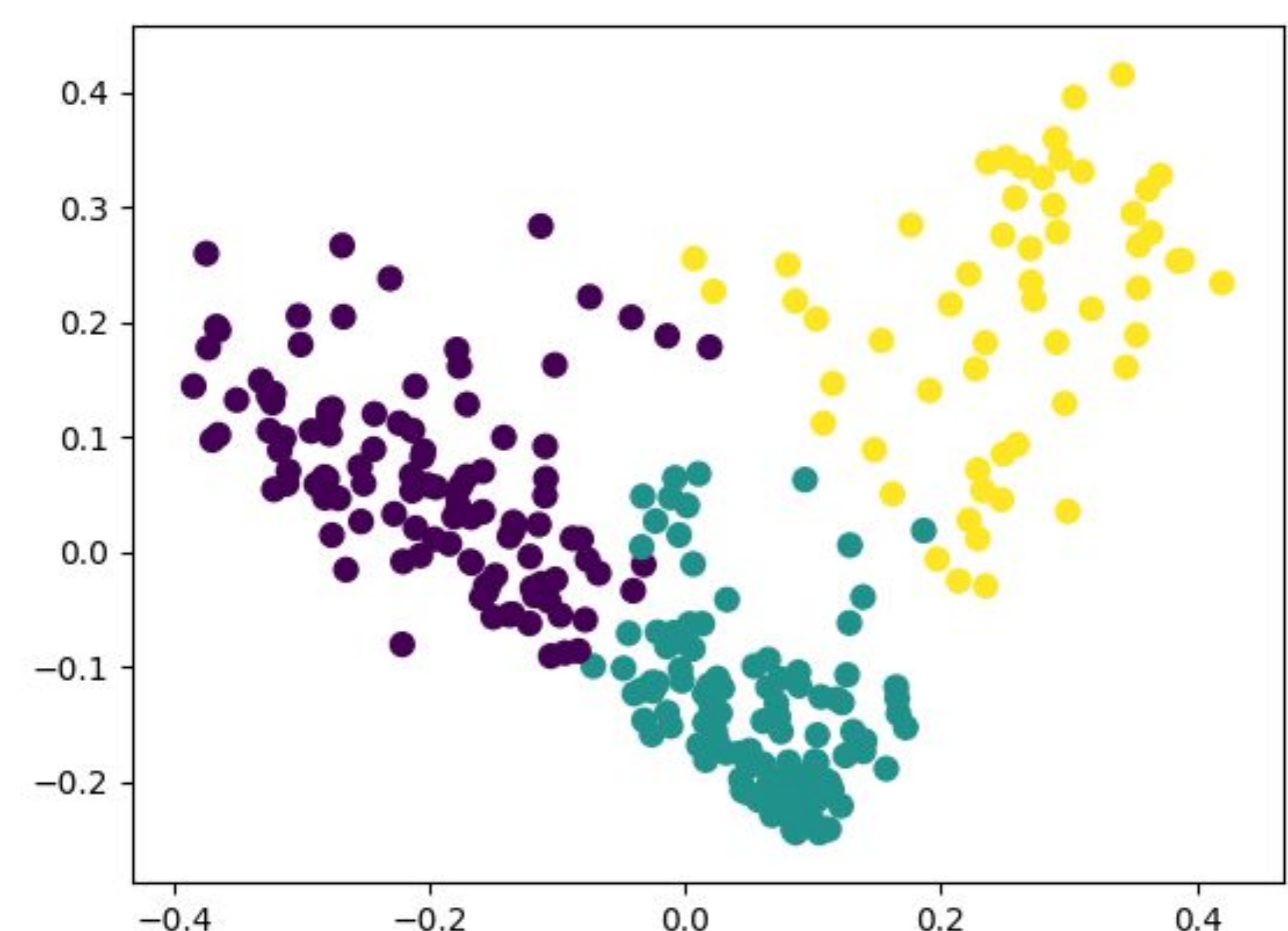
Dos 2000 usuários foram coletadas informações como número de seguidores, tweets, etc, além do corpus dos tweets. A análise se restringiu a tweets realizados durante o período eleitoral.

### Técnicas utilizadas na análise:

Utilizamos o algoritmo LDA para extração de tópicos e clusterizamos os usuários através do k-means. Utilizamos a biblioteca Vader para realizar uma análise de sentimento e a API HateSonar para avaliar agressividade e discurso de ódio.

## RESULTADOS

### Clusterização dos usuários via k-means



### Tópicos por cluster:

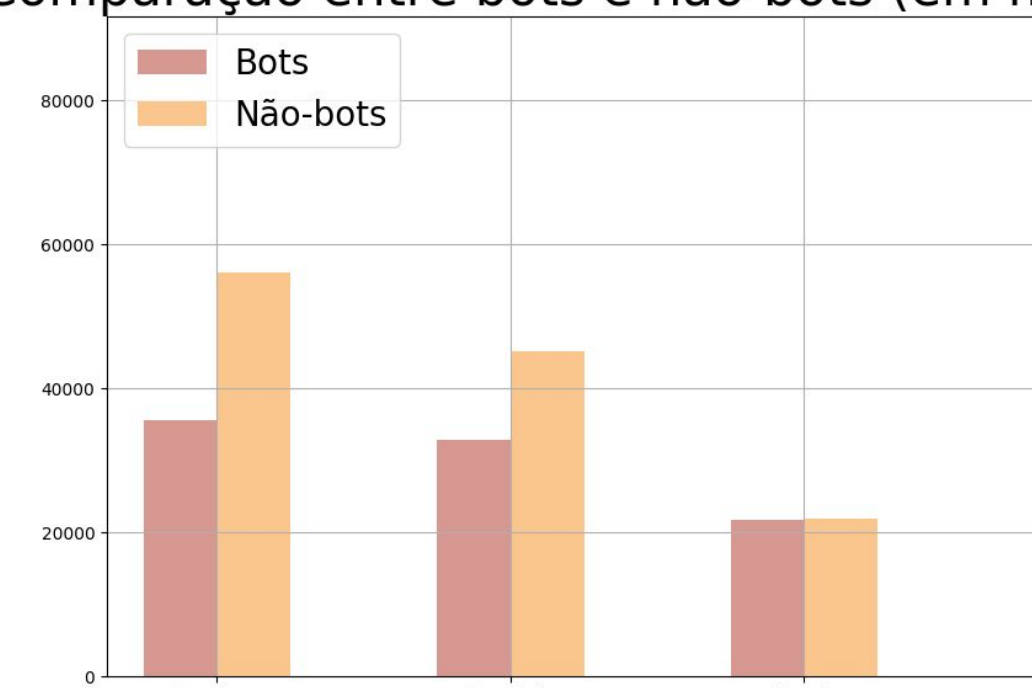
**Cluster roxo:** eu nao vida pra ficar alguem ja pessoa ter

**Cluster amarelo:** link nao pra voce vai ser dia brasil ta

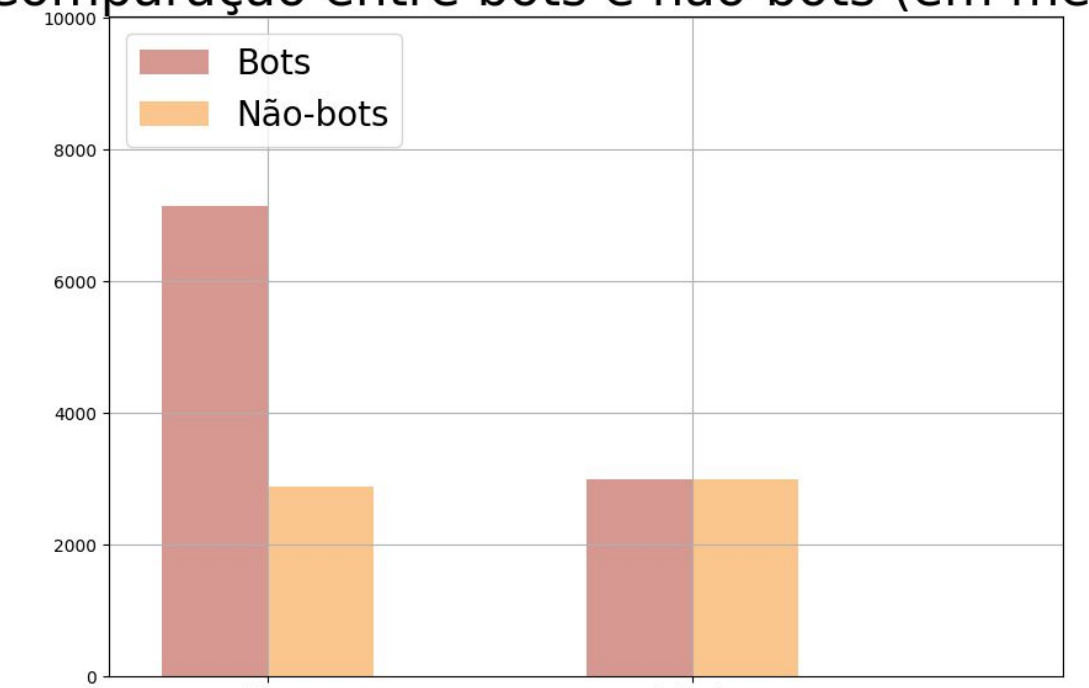
**Cluster ciano:** bolsonaro pt presidente candidato contra pais haddad campanha povo

### Métricas gerais:

Comparação entre bots e não-bots (em média)



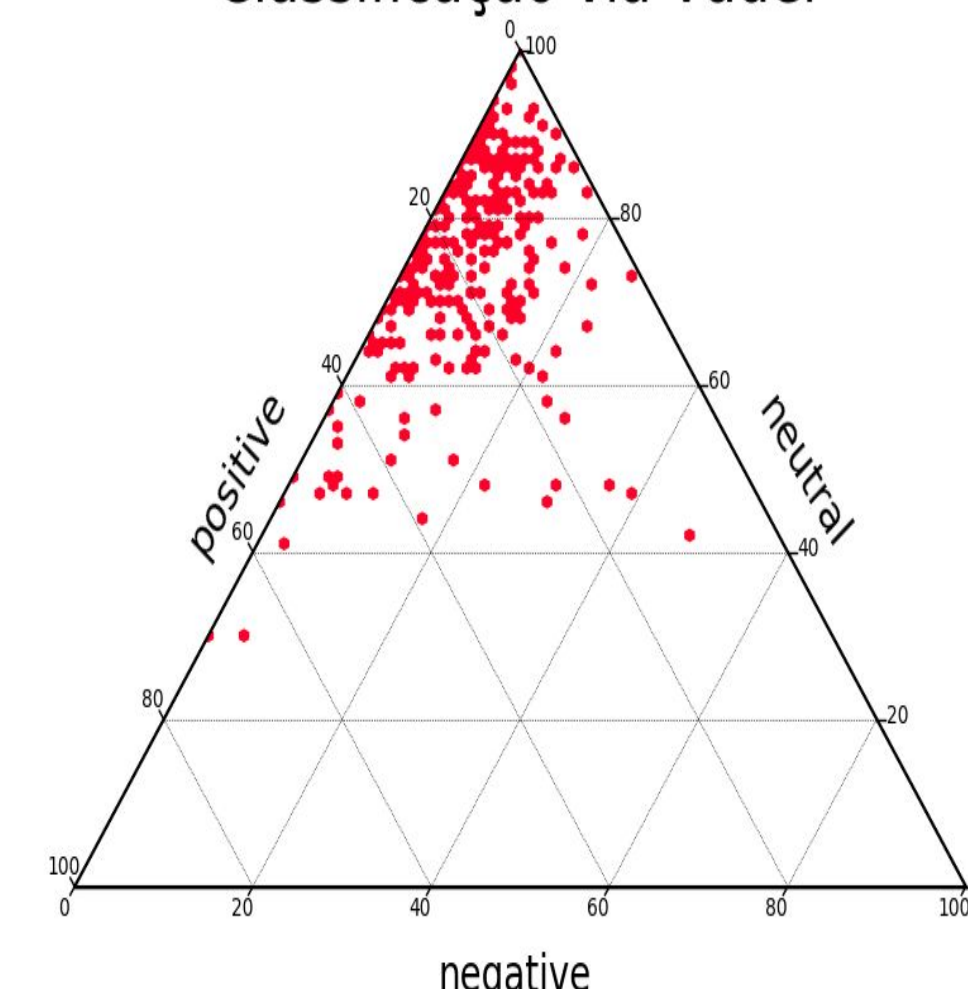
Comparação entre bots e não-bots (em média)



### Análise do discurso:

O texto foi classificado de acordo com três classes:

#### Classificação via Vader



#### Classificação via HateSonar

