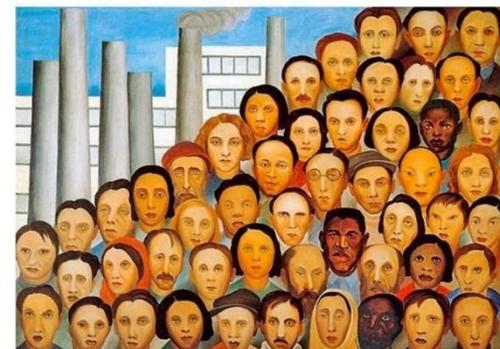
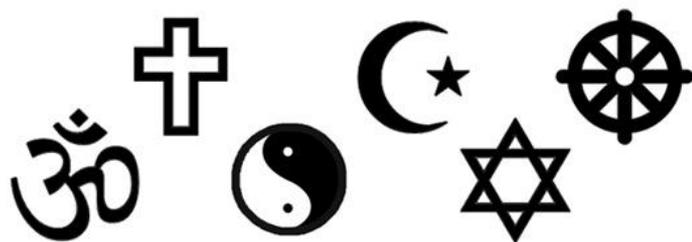
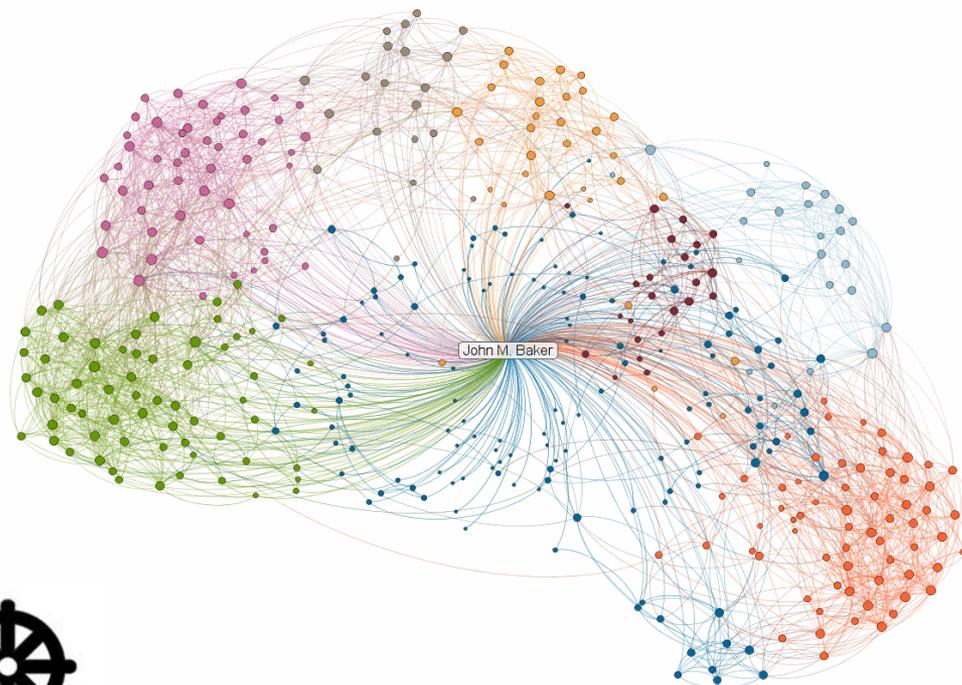


UTILIZANDO ANÁLISE DE SENTIMENTOS PARA DEFINIÇÃO DA HOMOFILIA POLÍTICA DOS USUÁRIOS DO TWITTER DURANTE A ELEIÇÃO PRESIDENCIAL AMERICANA DE 2016

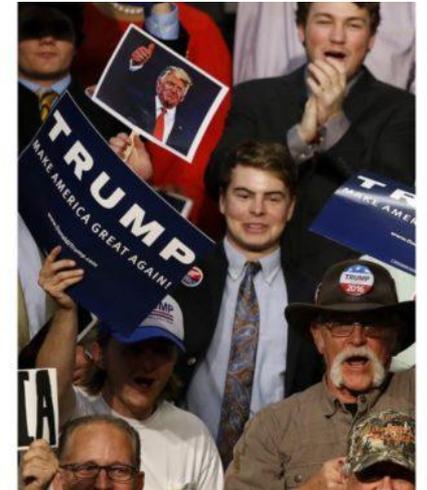
**Josemar Alves Caetano, Hélder Seixas Lima,
Mateus Freira dos Santos, Humberto Torres Marques-Neto**

PROPPG/PPG – Programa de Pós-Graduação em Informática
LEASDLE - Laboratório de Engenharia e Arquitetura de Sistemas Distribuídos em Larga Escala
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)

Introdução



Introdução



Motivação



Donald J. Trump ✓

@realDonaldTrump

 Follow

How long did it take your staff of 823 people to think that up--and where are your 33,000 emails that you deleted?

twitter.com/hillaryclinton...

5:40 PM - 9 Jun 2016

  168,810  297,356



Hillary Clinton ✓

@HillaryClinton

 Follow

"I never said that." —Donald Trump, who said that. [#debatenight](#)

twitter.com/realDonaldTrump...

10:19 PM - 26 Sep 2016

  101,193  160,848

Motivação


Donald J. Trump 
 @realDonaldTrump Follow

How long did it take your staff of 823 people to think that up--and where are your 33,000 emails that you deleted?
twitter.com/hillaryclinton...

5:40 PM - 9 Jun 2016


 168,810  297,356

17 milhões de seguidores
35 mil tweets publicados


Hillary Clinton 
 @HillaryClinton Follow

"I never said that." —Donald Trump, who said that. #debatenight
twitter.com/realDonaldTrump...

10:19 PM - 26 Sep 2016


 101,193  160,848

12 milhões de seguidores
9 mil tweets publicados

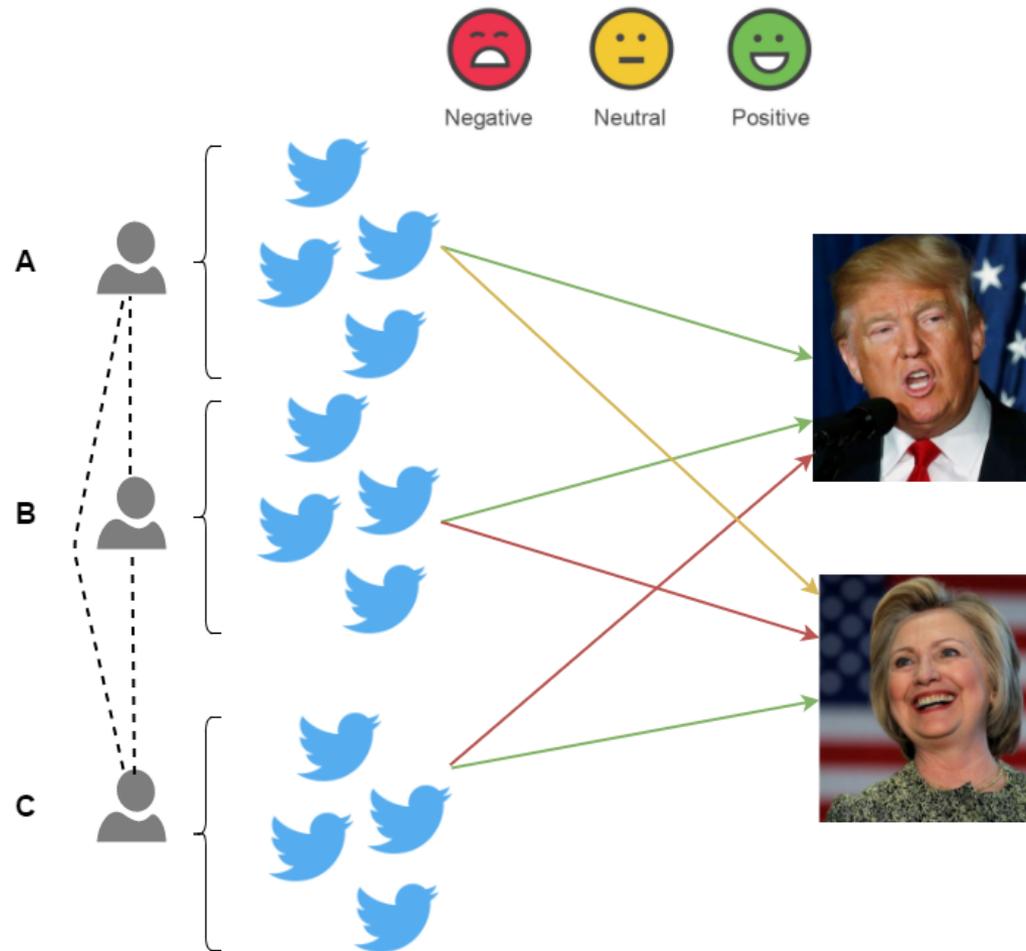
Objetivo

Identificar e analisar a homofilia política entre usuários do Twitter ao longo da campanha presidencial americana de 2016

Contribuição

Propor uma nova abordagem para a análise da homofilia política no Twitter utilizando a **análise de sentimentos** como característica para a classificação dos usuários

Visão Geral



Referencial Teórico

- **Homofilia**

$$H_i = \frac{s_i}{s_i + d_i}$$

- s_i representa o número de conexões que ligam indivíduos da mesma classe i
- d_i representa o número de conexões que ligam indivíduos da classe i com indivíduos de outras classes

Referencial Teórico

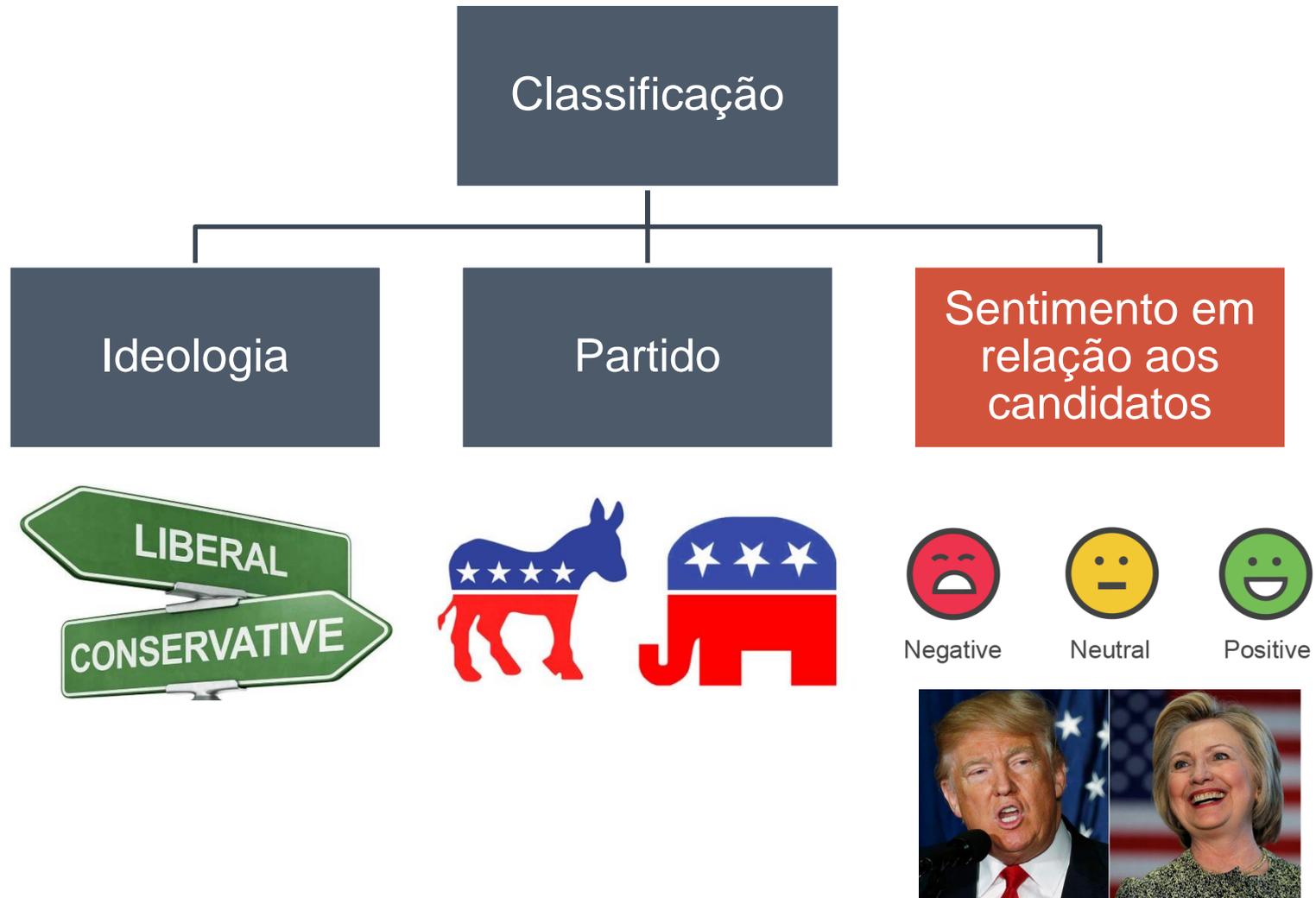
- ***Inbreeding Homophily***

$$IH_i = \frac{H_i - w_i}{1 - w_i}$$

- IH_i é o indicador de *inbreeding homophily*
- w_i representa a probabilidade da ocorrência de indivíduos da classe i na população

$$\begin{cases} IH_i > 0 \rightarrow \text{homofilia} \\ IH_i < 0 \rightarrow \text{heterofilia} \end{cases}$$

Trabalhos Relacionados

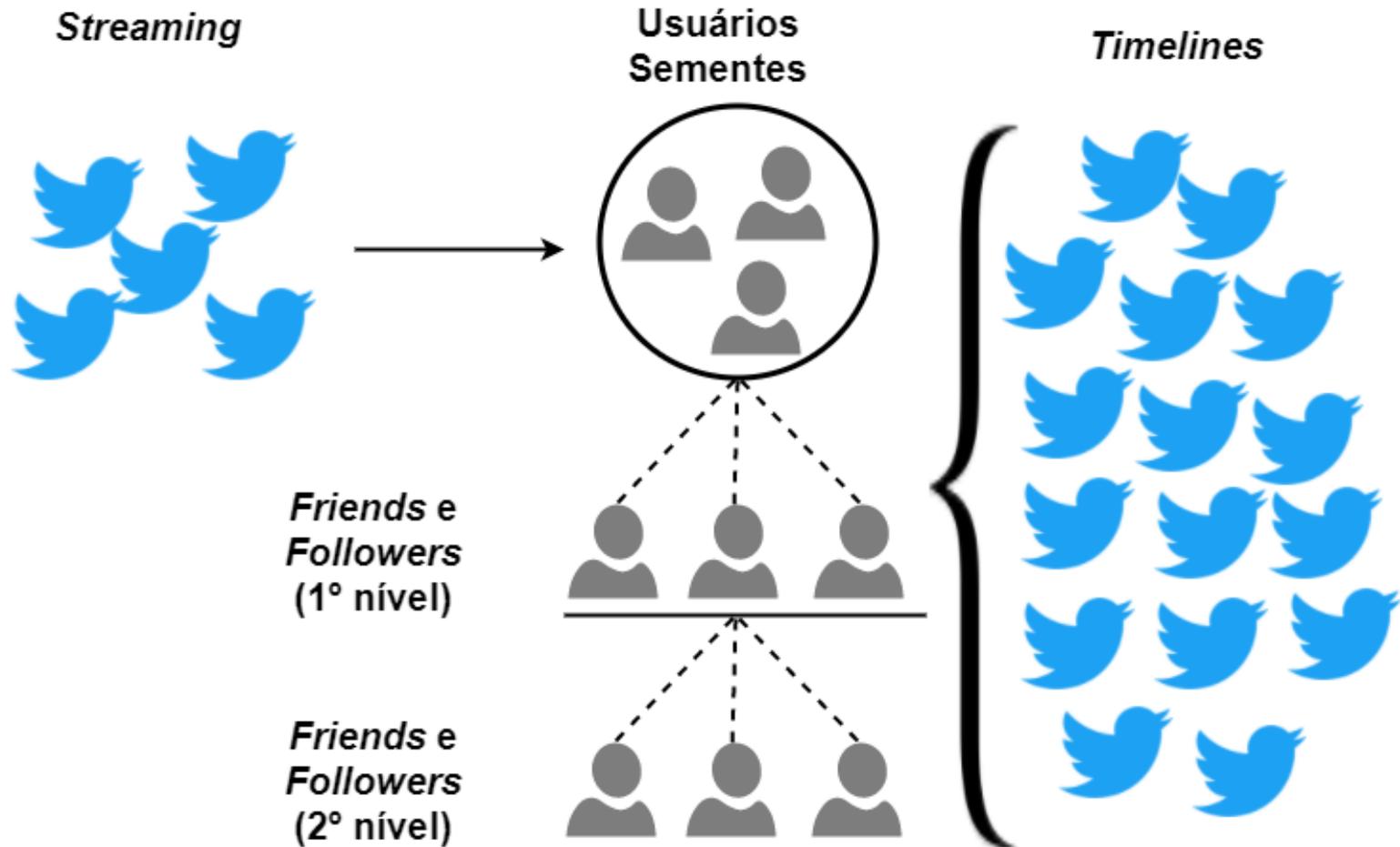


Metodologia

- Etapas



Coleta dos Dados



Coleta dos Dados

- Período de coleta: agosto a novembro de 2016



- *Tweets* coletados: **3,6 milhões**
- Escritos por **18 mil** usuários distintos

Análise de Sentimentos dos *Tweets*



Análise de Sentimentos dos *Tweets*

- Como o sentimento do texto foi calculado?
 - Ferramenta **SentiStrength**
 - Abordagem baseada em dicionário de lexemas



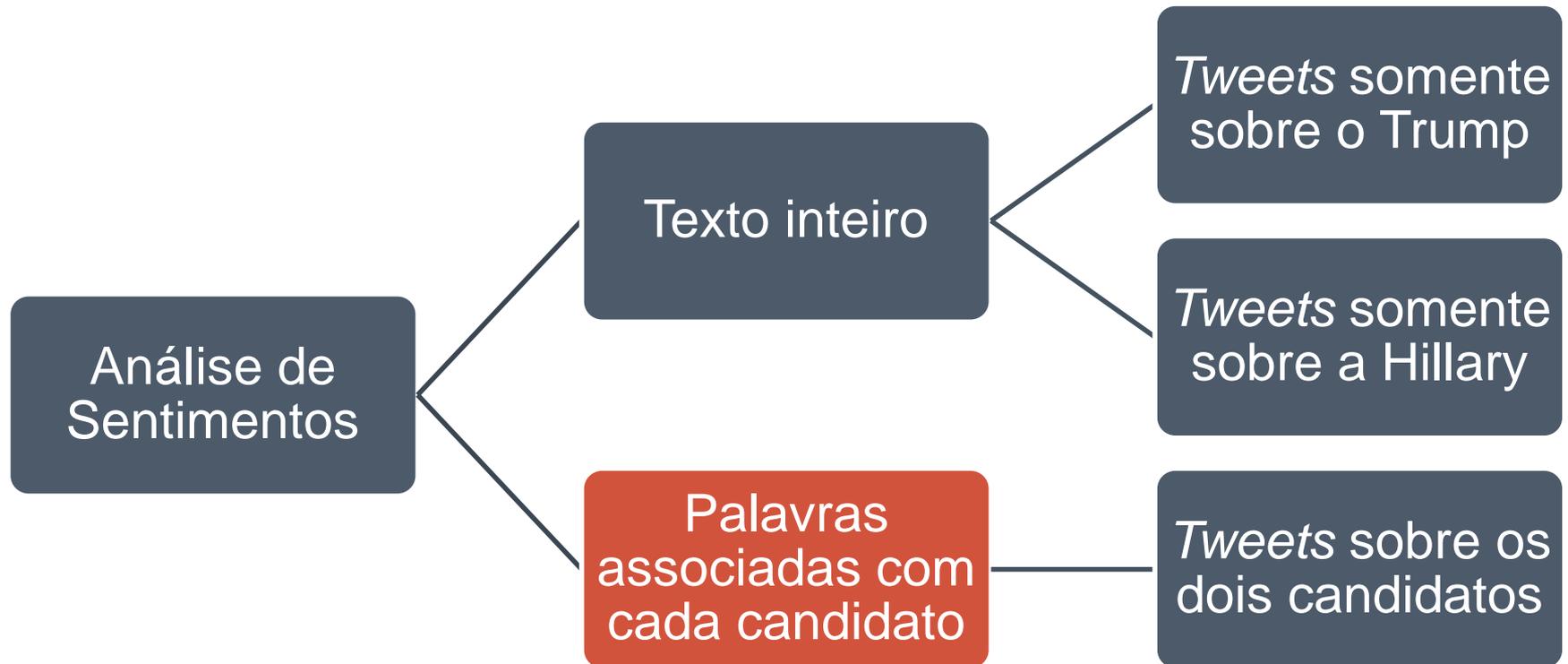
Análise de Sentimentos dos *Tweets*

- Como o sentimento do texto foi calculado?
 - Ferramenta **SentiStrength**
 - Abordagem baseada em dicionário de lexemas



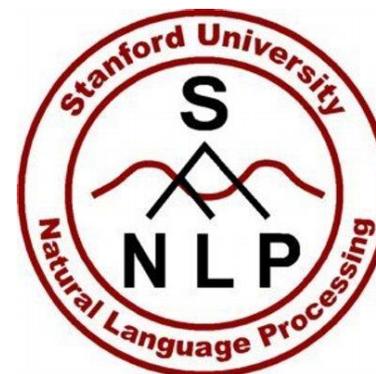
- O sentimento do texto é a soma do **maior valor positivo** com o **menor valor negativo** associados a palavras do texto

Análise de Sentimentos dos *Tweets*



Análise de Sentimentos dos *Tweets*

- Como os candidatos e palavras associadas foram identificados?
 - Ferramenta **Stanford Parser**
 - Processador de linguagem natural



Análise de Sentimentos dos *Tweets*

- Como os candidatos e palavras associadas foram identificados?
 - Ferramenta **Stanford Parser**
 - Processador de linguagem natural

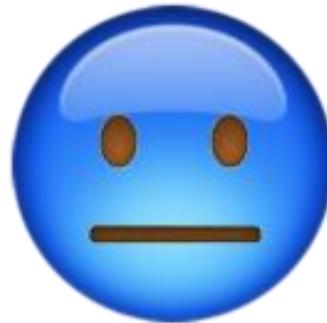


Candidato	Sujeitos associados	Pronomes
Donald Trump	@realDonaldTrump, Trump, DT	he, him
Hillary Clinton	@HillaryClinton, Hillary, HC	she, her

Classes do 1º Experimento (Hillary)



65%



15%

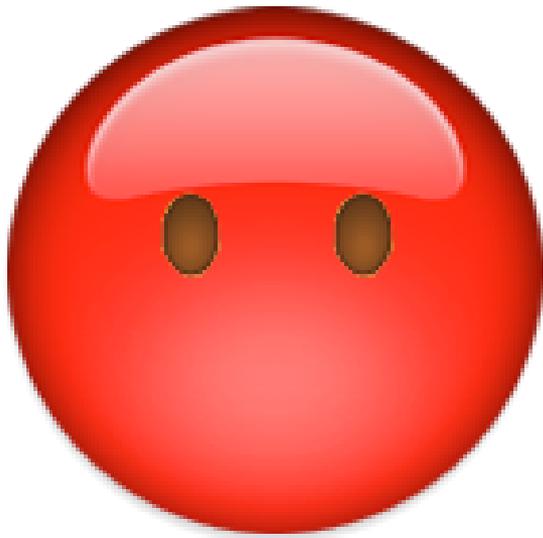


13%



7%

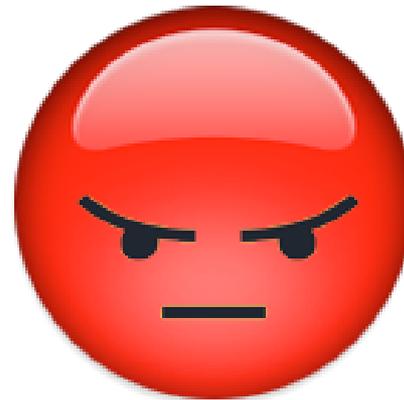
Classes do 1º Experimento (Trump)



55%



13%

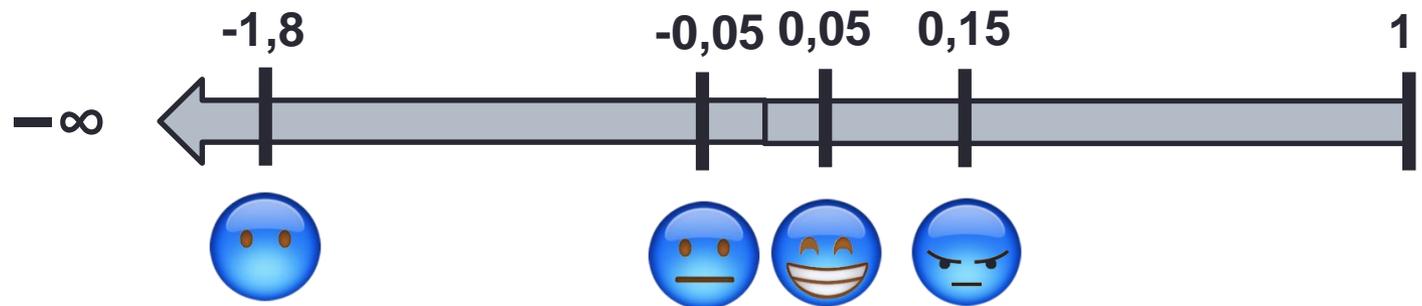
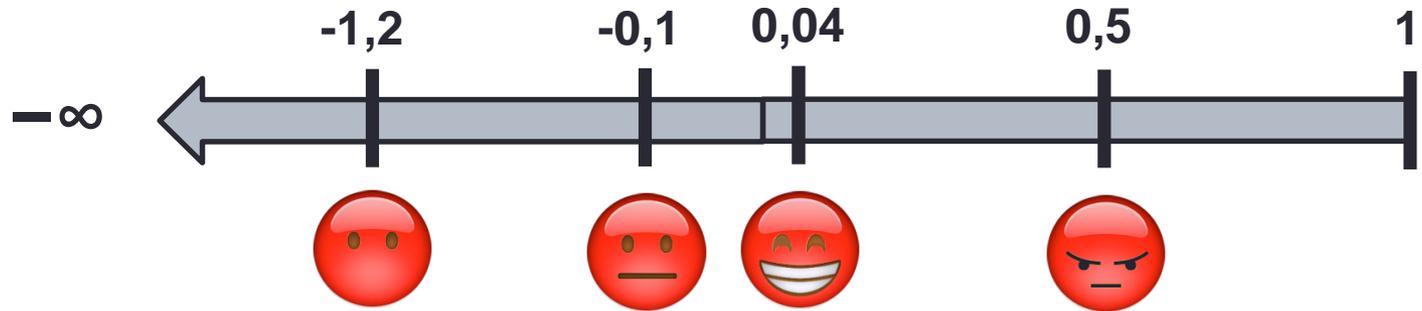


27%

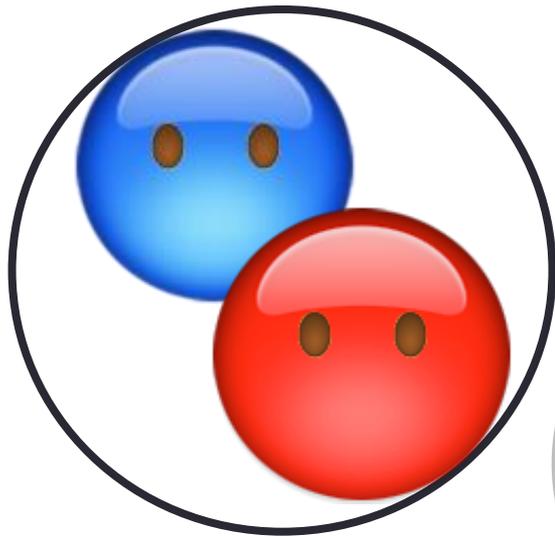


5%

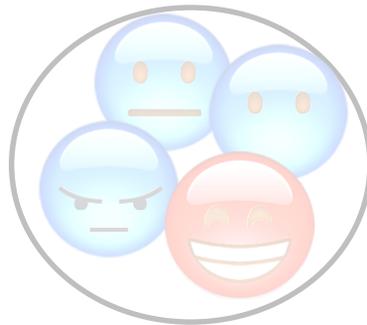
Análise da Homofilia (1º Experimento)



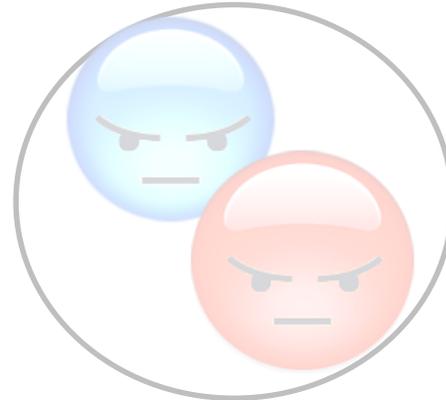
Classes do 2º Experimento



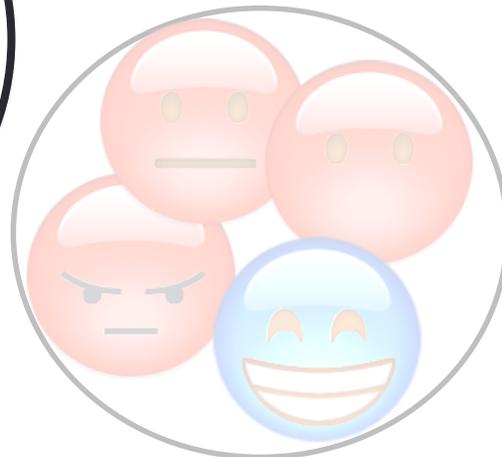
52%
Whatever



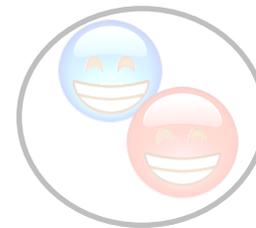
5%
Trump supporter



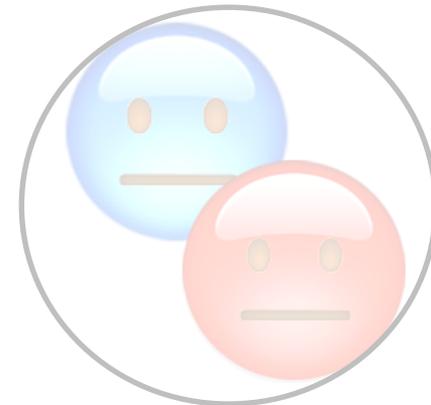
10%
Hater



19%
Hillary supporter

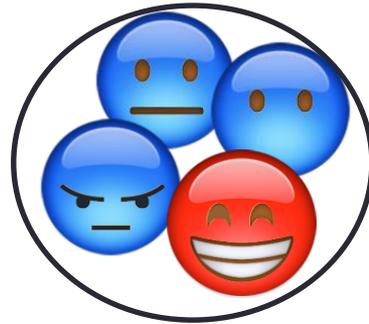


1%
Yes-man



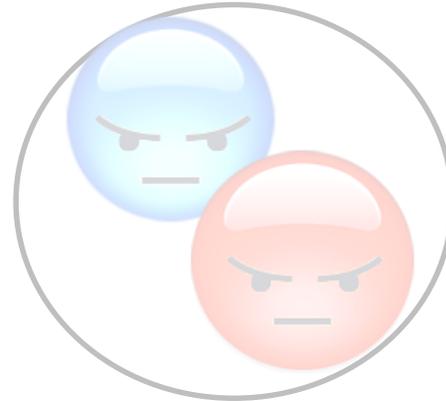
13%
Diplomat

Classes do 2º Experimento

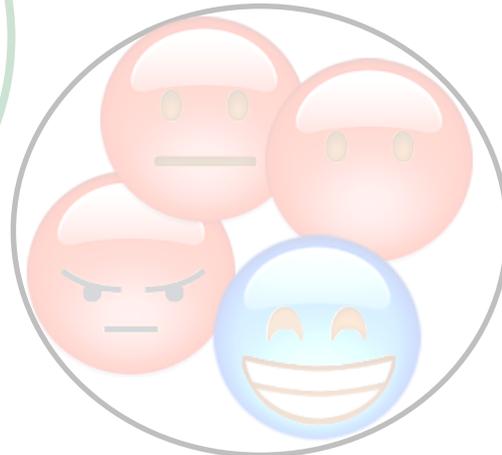


5%

Trump supporter

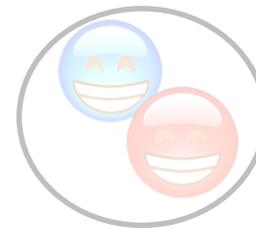


10%
Hater



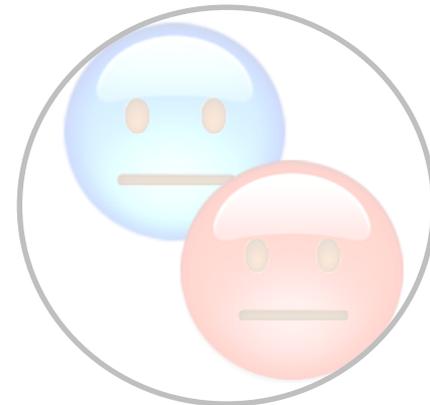
19%

Hillary supporter

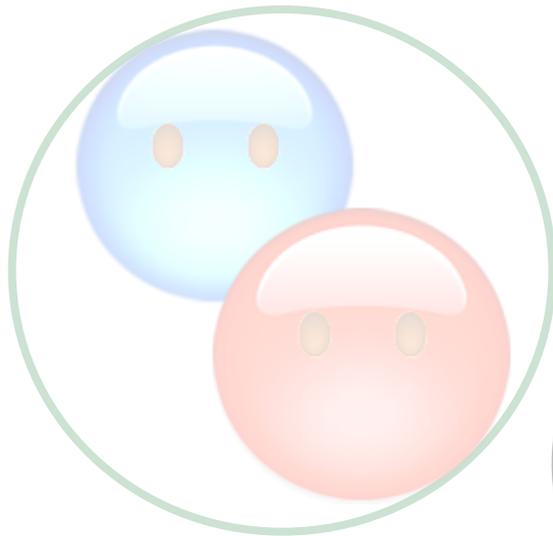


1%

Yes-man

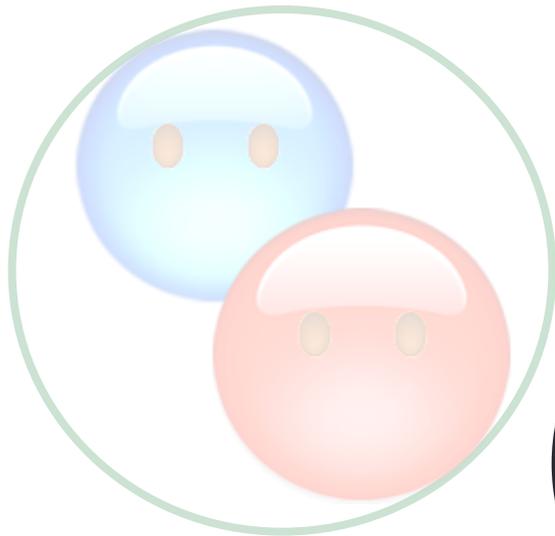


13%
Diplomat

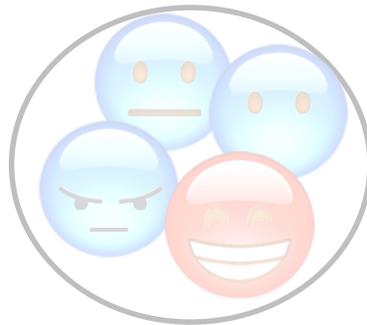


52%
Whatever

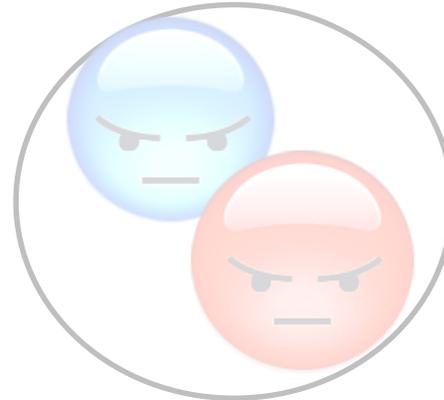
Classes do 2º Experimento



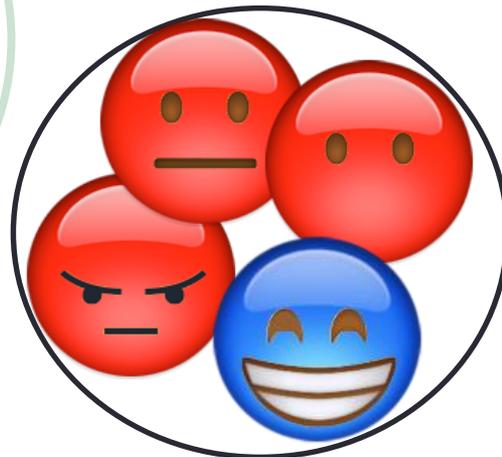
52%
Whatever



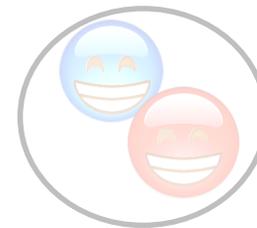
5%
Trump supporter



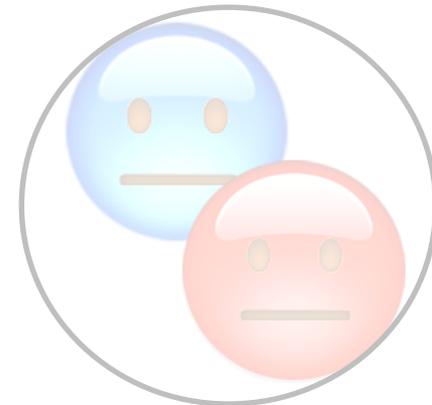
10%
Hater



19%
Hillary supporter

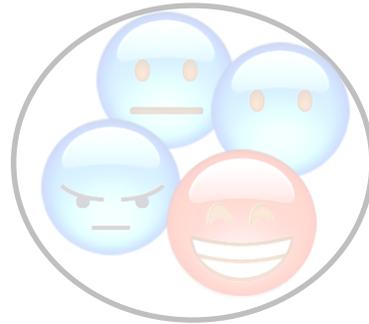


1%
Yes-man



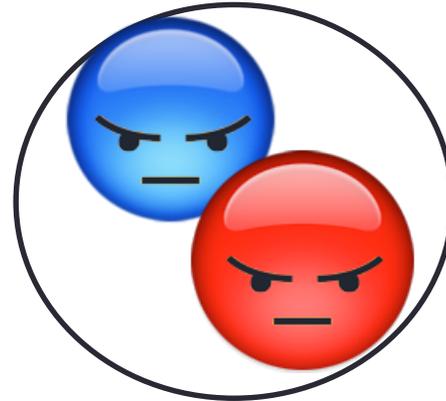
13%
Diplomat

Classes do 2º Experimento

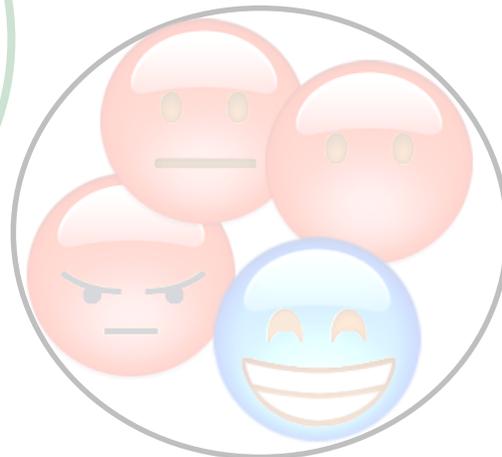


5%

Trump supporter

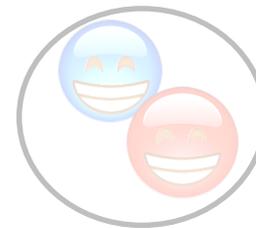


10%
Hater



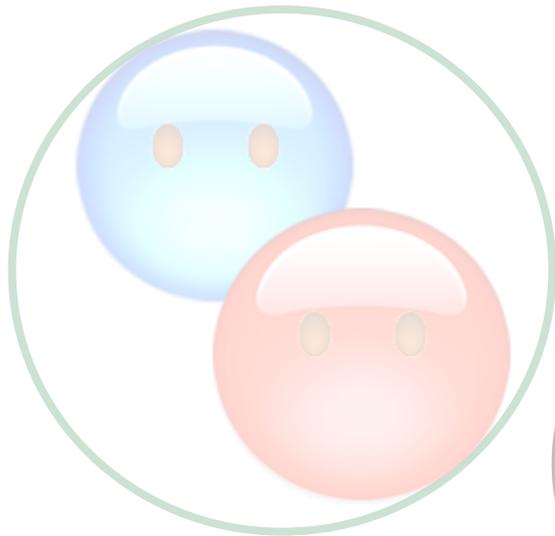
19%

Hillary supporter



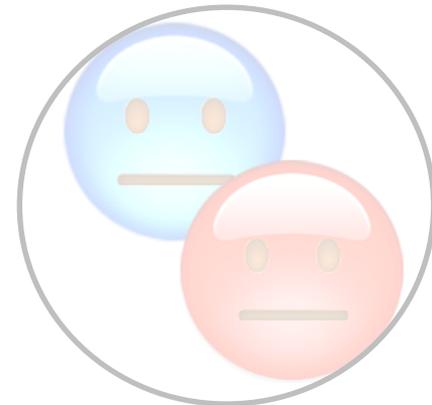
1%

Yes-man



52%

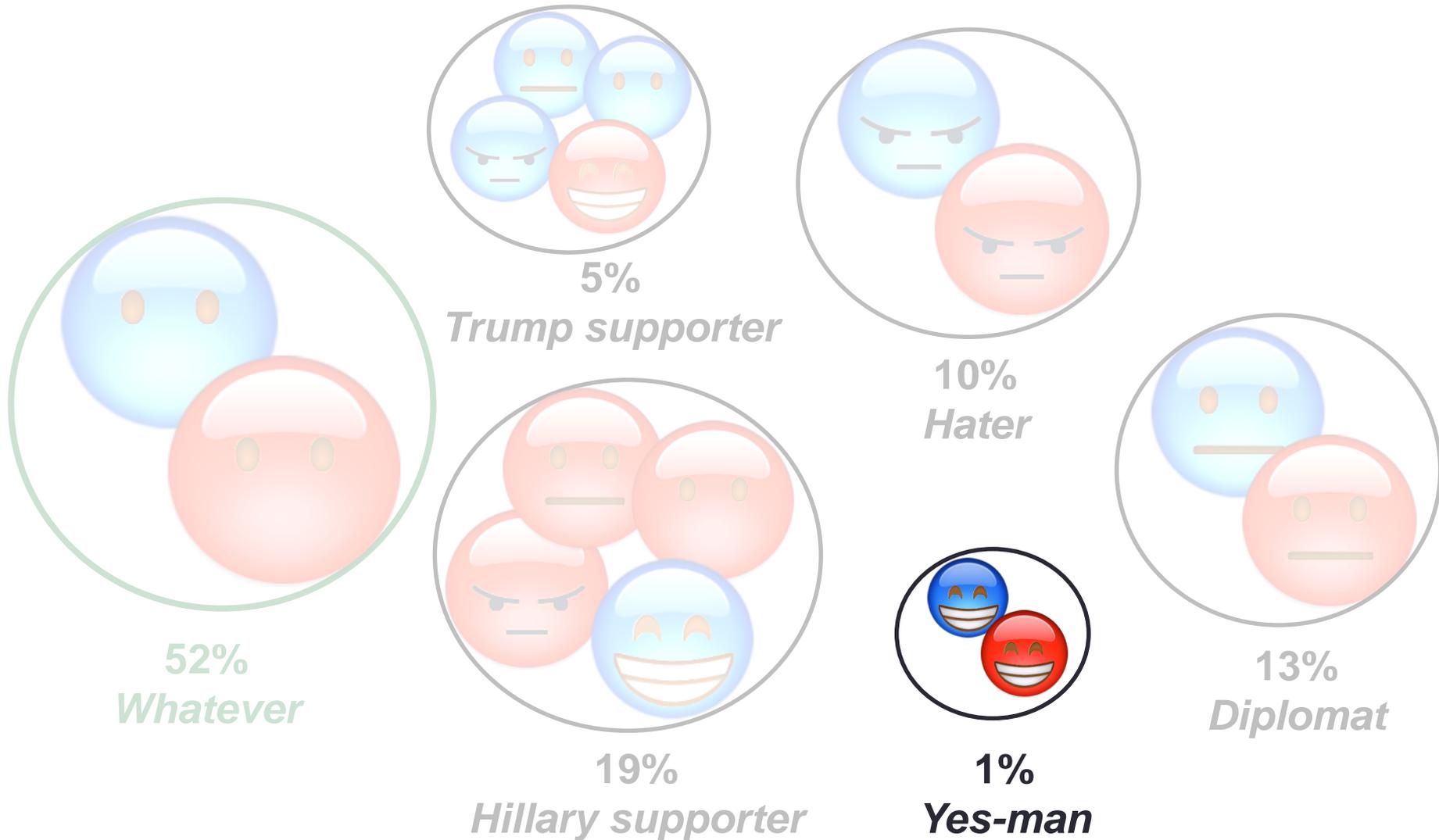
Whatever



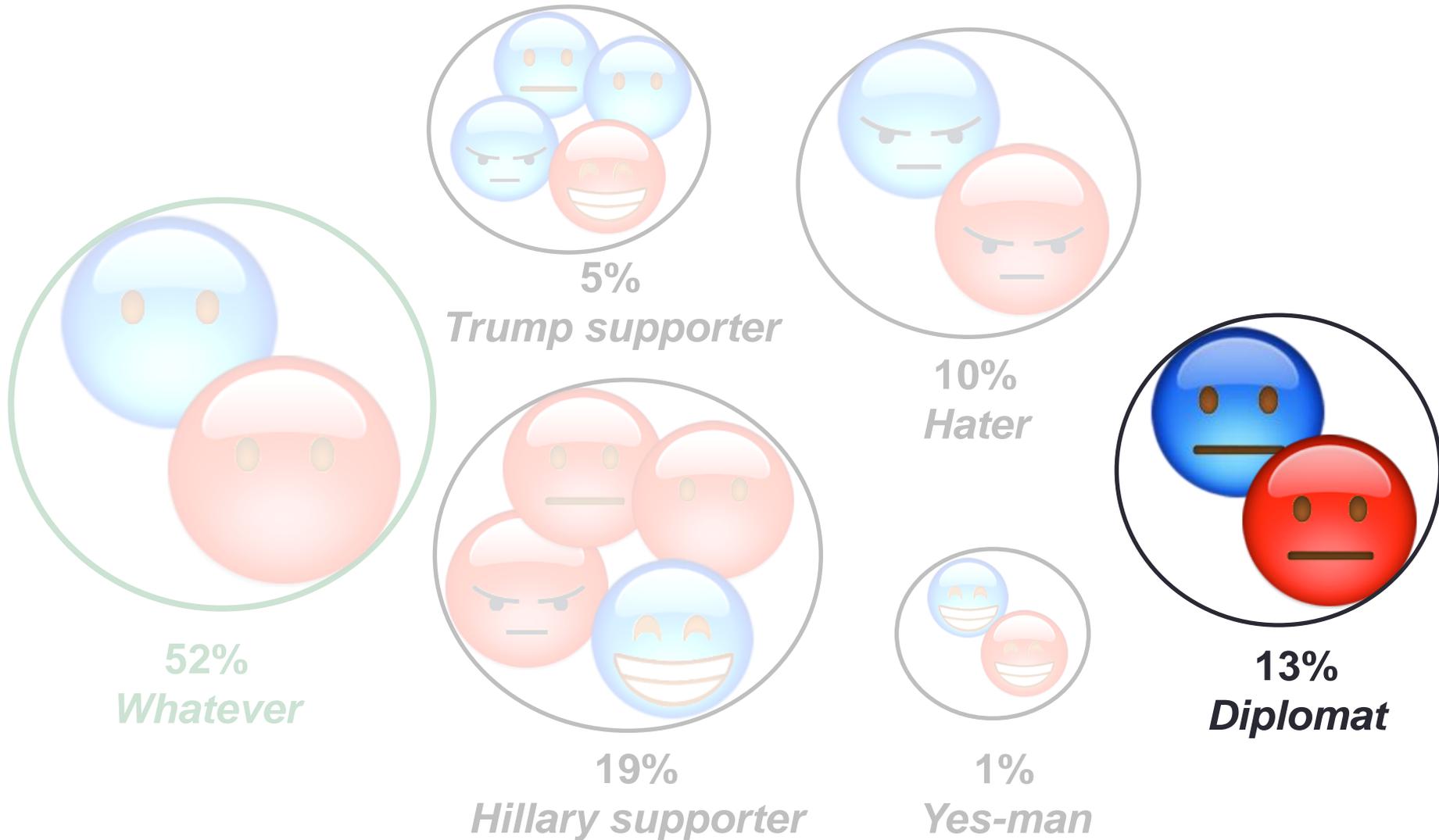
13%

Diplomat

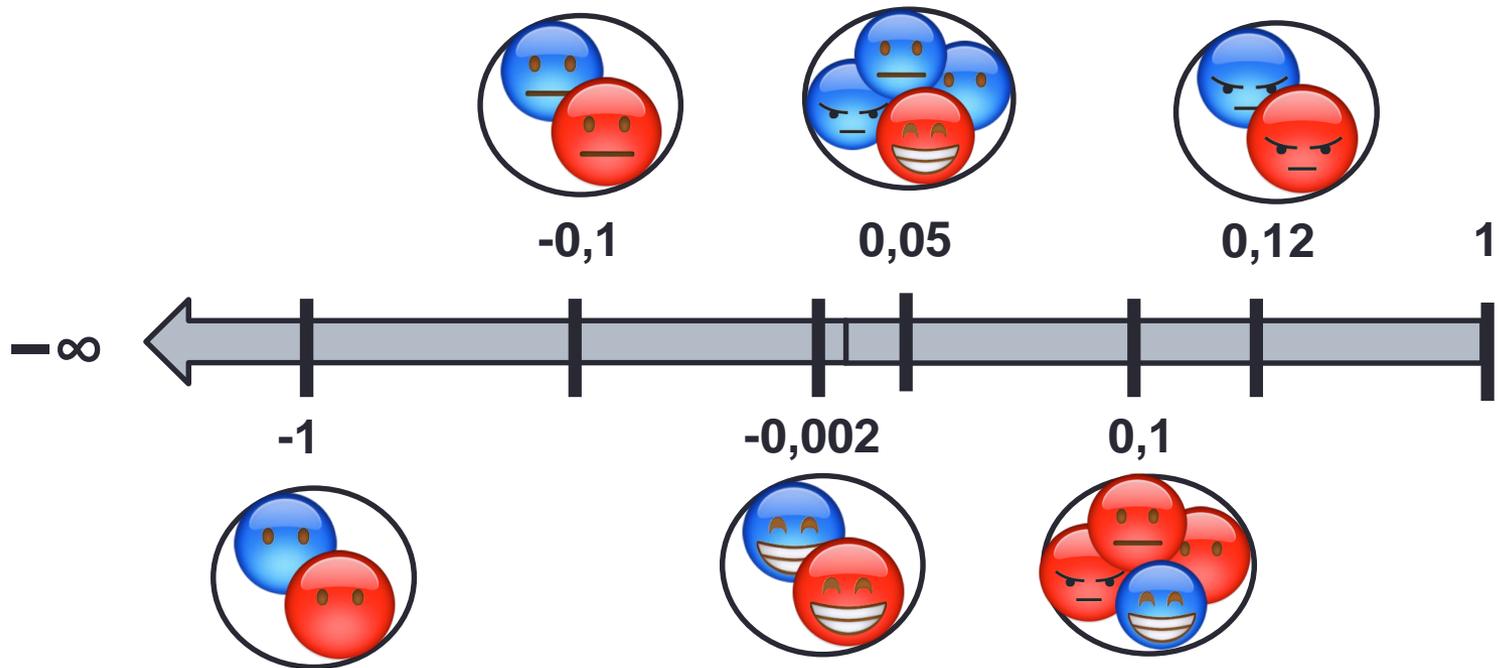
Classes do 2º Experimento



Classes do 2º Experimento

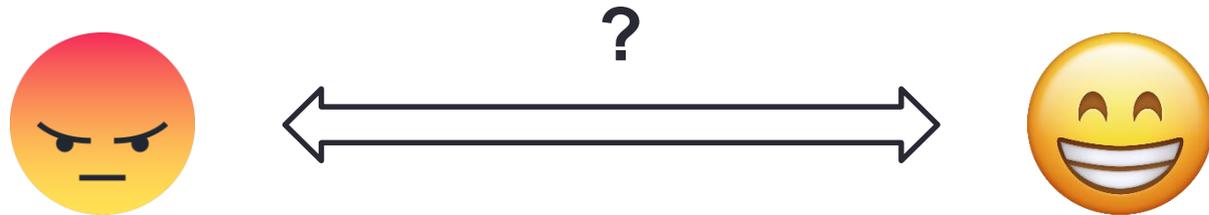


Análise da Homofilia (2º Experimento)

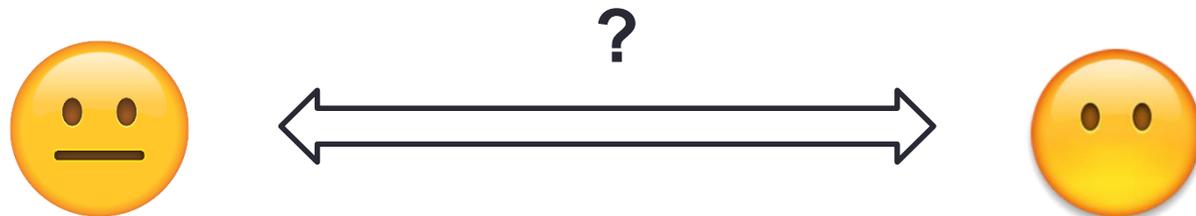


Limitações

- *Tweets* com conteúdo sarcástico



- Limitação da API do Twitter (no máximo os 200 últimos *tweets*)



Conclusões

- Usuários que compartilham **sentimento negativo** em relação a um ou a ambos candidatos tendem a se **relacionarem mais entre si**
- Usuários que **não manifestaram sentimento** ou que manifestaram **sentimento neutro** em relação a um ou a ambos candidatos tendem a se **relacionarem mais com usuários de outras classes**

Trabalhos Futuros

- Caracterizar cada classe (uso de *hashtags*, métricas de redes complexas, usuários influentes)
- Análise temporal (como homofilia política evoluiu ao longo da campanha)
- Eliminar *outliers* (*bots*, *fake accounts*)
- Aplicar mesma metodologia em outras eleições (França 2017)

Agradecimentos



josemar.caetano@sga.pucminas.br