# ESTUDO DE CASO SOBRE COLETA E ANÁLISE DE DADOS DO TWITTER





#### Bruno Tatsuya Masunaga Santos, Claudio Luis de Camargo Penteado, Denise Hideko Goya

Centro de Matemática, Computação e Cognição (CMCC), Universidade Federal do ABC

Av. dos Estados, 5001 - Bangú, Santo André – SP

{bruno.masunaga@aluno.ufabc.edu.br, claudio.penteado@ufabc.edu.br, denise.goya@ufabc.edu.br}

**Resumo:** Com o objetivo de demonstrar a importância de ferramentas computacionais no ambiente acadêmico, este projeto utiliza mecanismos de coleta, tratamento e análise de dados da rede social Twitter, baseando-se num estudo de caso sobre as Eleições Municipais de 2016 nas cidades de São Paulo e do Grande ABC paulista. A partir destes dados, efetuou-se análises de interesse, visando identificar tendências de preferência entre os candidatos e até mesmo prever acontecimentos ou curso das eleições.

Palavras-chave: Ciência das Redes, Redes Sociais de Internet, Eleições Municipais.

## INTRODUÇÃO

A rápida popularização das Redes Sociais de Internet tem produzido importantes transformações nas relações sociais, criando novas formas de sociabilidade mediadas por computadores. Um evento político como as Eleições Municipais de 2016 promove a produção de uma grande quantidade de dados públicos nestas redes. Tais dados podem ser analisados de diversas maneiras, de forma a abranger diversos interesses de estudo.

### **OBJETIVOS**

Demonstrar a importância de ferramentas computacionais no âmbito da pesquisa, dando ênfase à contribuição de caráter interdisciplinar para outras áreas do conhecimento. Para tanto, faz-se necessário identificar tendências de preferência entre os candidatos e até mesmo inferir acontecimentos ou o curso das eleições, utilizando, para isso, dados públicos capturados através das ferramentas.

# **MÉTODO**

As coletas foram efetuadas a partir de *keywords*, compostas de variações textuais dos nomes de cada candidato das cidades estudadas. As análises se basearam em estatísticas representativas com base em valores numéricos associados às informações extraídas nas etapas anteriores. Foram capturados cerca de 266 mil tweets no período de 31/09 a 02/10 de 2016, isto é, período anterior à apuração dos votos do primeiro turno.

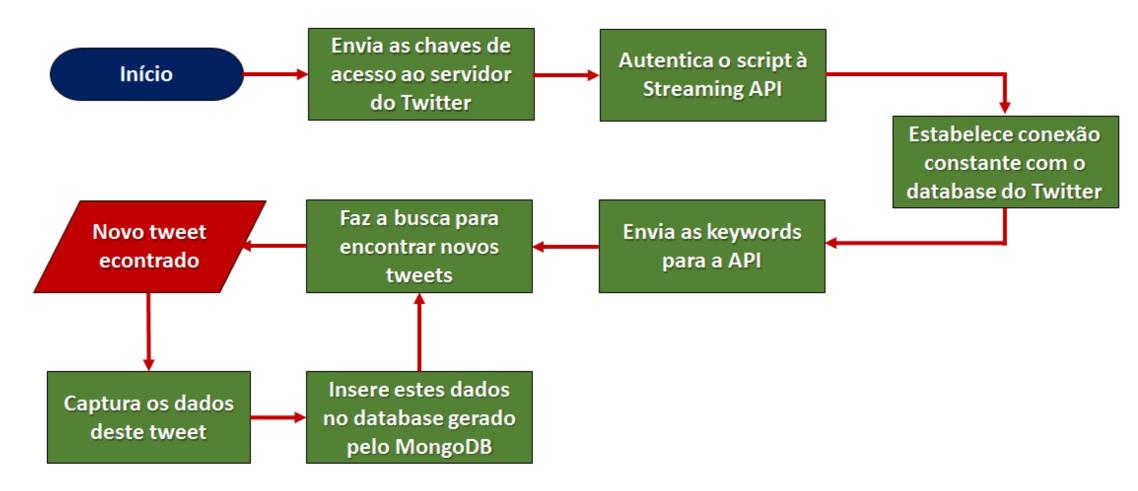


Figura 1: Procedimento de Coleta de Dados utilizado

Fonte: Elaborada pelos autores

#### RESULTADOS E DISCUSSÃO

Análises efetuadas produziram resultados de natureza comparativa. As figuras 2 e 3 representam os termos e hashtags mais utilizados e as proporções de tweets coletados por cidade, respectivamente.



Figura 2: Termos mais utilizados (direita) e hashtags mais utilizadas (esquerda)

Fonte: Elaborada pelos autores

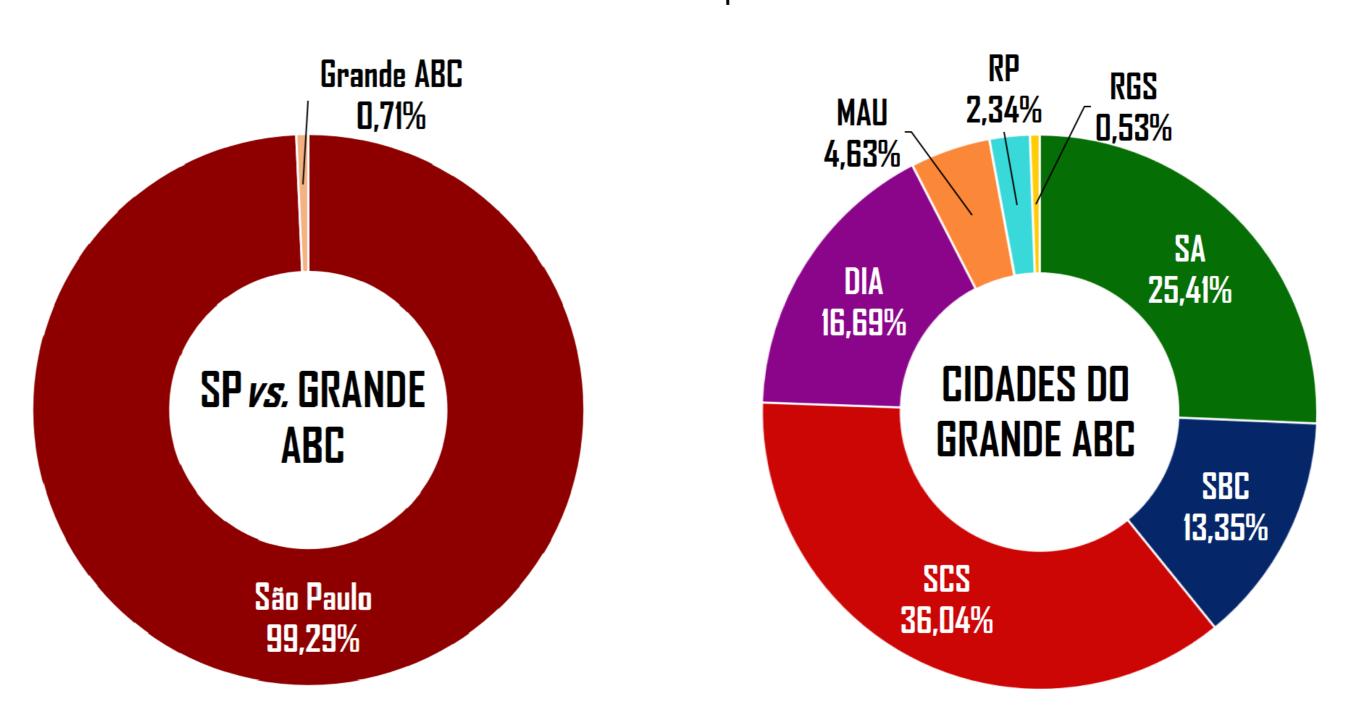


Figura 3: Proporções de tweets coletados por cidade no período estudado

Fonte: Elaborada pelos autores

Estes dados, comparados com os resultados retirados do Tribunal Superior Eleitoral (TSE) após as eleições, apresentaram um índice de concordância de cerca de 88%, indicando que os números obtidos foram condizentes com o resultado das eleições. Além disso, em todas as cidades, o candidato que mais foi citado no estudo de caso também foi aquele que mais recebeu votos no primeiro turno do processo eleitoral.

## CONCLUSÃO

O método aplicado no projeto foi satisfatório para alcançar os objetivos propostos, uma vez que demonstrou um alto índice de concordância quando comparou-se os dados coletados com os dados do TSE. No entanto, existem limitações, como acesso restrito a 1% do total global dos tweets publicados na rede e a necessidade de análise do sentimento das postagens. Apesar disso, os resultados demonstram que o projeto permite novas abordagens, que podem auxiliar o avanço de pesquisas em demais áreas do conhecimento.

## **AGRADECIMENTOS**

Este trabalho foi financiado pelo Programa de Iniciação Científica da UFABC (PIC/UFABC) na projeto Pesquisando Desde o Primeiro Dia (PDPD) de 2016.



