

Investigando a Influência de Tweets em Programas de Votação Popular no Brasil

Igleson F. Figueredo, Leandro B. Marinho, Leonardo A. Santos

¹Departamento de Sistemas e Computação
Universidade Federal de Campina Grande (UFCG) – Campina Grande-PB – Brazil

{igleson.figueredo, leonardo.santos}@ccc.ufcg.edu.br, lbmarinho@dsc.ufcg.edu.br

Abstract. *Twitter has been used as a data source by researchers for the investigation and prediction of various aspects of human behaviour in social networks. We investigate in this work the interplay between the performance of participants of popular vote programs in Brazil and their popularity in Twitter. As study case, we collected data from Twitter about the latest editions of two popular vote programs, Superstar and Big Brother Brasil, and found out that the popularity of the participants in Twitter is an important indicator of their respective performances in these programs. Another interesting discovery is the observation that the participants who have better performances, have more positive tweets than the other competitors during the program.*

Resumo. *O Twitter tem sido utilizado por pesquisadores como uma importante fonte de dados para investigação e predição de diversos aspectos do comportamento humano em redes sociais. Investigamos neste trabalho a sinergia entre o desempenho de participantes de programas de votação popular no Brasil e a sua popularidade no Twitter. Como estudo de caso, nós coletamos dados do Twitter acerca das edições mais recentes de dois programas de votação popular de grande audiência no Brasil, são eles o Superstar e o Big Brother Brasil, e descobrimos que a popularidade dos participantes destes programas no Twitter é um importante indicador dos seus respectivos desempenhos. Outra descoberta interessante foi o fato dos participantes de melhor desempenho terem mais tweets positivos que os demais competidores durante o programa.*

1. Introdução

O Twitter¹ é uma rede social em formato de microblog onde os usuários escrevem livremente e interagem por meio de *retweets*, *likes* e mensagens privadas. Dado o grande fluxo de postagens (aka tweets) e de usuários, o Twitter tem se tornado um instrumento importante na caracterização e predição do comportamento humano em redes sociais. Além disso, ele tem sido usado para elucidar opiniões acerca de diversas entidades do mundo real, tais como empresas, artistas, políticos (eleitos e candidatos à eleição), eventos e muitos outros [Gayo-Avello et al. 2011], [He et al. 2013], [Gomide et al. 2011].

Neste trabalho nós investigamos a sinergia entre o desempenho de participantes de programas de votação popular do Brasil e a sua popularidade no Twitter. Como estudo de caso, coletamos dados do Twitter acerca de dois programas de votação popular

¹<http://www.twitter.com>

de grande audiência no Brasil, o Superstar 2014 e do Big Brother Brasil (BBB) 2015, e descobrimos que a popularidade dos participantes no Twitter é um importante indicador dos seus desempenhos nesses programas, podendo até mesmo ser usada para prever com alguma antecedência o resultado da votação. Além disso, utilizando análise de sentimento dos tweets no período da votação constatamos que, em geral, o candidato de maior desempenho possui mais tweets positivos do que os competidores durante o programa.

O artigo está organizado da seguinte forma. Na Seção 2 apresentamos brevemente os trabalhos relacionados. Na Seção 3 apresentamos a metodologia empregada na coleta dos dados. Na Seção 4 apresentamos os resultados desta pesquisa, discutimos os resultados e, finalmente na Seção 5, concluímos o artigo e apresentamos direções para trabalhos futuros.

2. Trabalhos Relacionados

Não encontramos na literatura revisada nenhum trabalho sobre a influência de tweets nos programas de votação populares brasileiros. Todavia, há uma grande quantidade de trabalhos que analisam dados do Twitter para a predição de outros aspectos do comportamento humano, como, por exemplo, tráfego em uma cidade [He et al. 2013], incidência de dengue [Gomide et al. 2011], eleições políticas [Gayo-Avello et al. 2011] e quão popular um tweet irá se tornar no futuro [Szabo and Huberman 2010], [Bandari et al. 2012], [Hong et al. 2011]. Diferentemente destes, este trabalho tem como objetivo principal investigar se a popularidade, assim como o sentimento, acerca de uma dada entidade (e.g. uma banda) no Twitter é útil para prever seu desempenho em programas de votação popular no Brasil.

3. Coleta e Preparação dos Dados

Por limitações da API do Twitter a coleta de dados foi feita de forma semi-manual. Para isso foi usada a página de busca avançada. Como a página restringe o número de respostas por página, foi criado um script em Javascript para que fosse gerado todas as requisições até a data desejada. Assim o html da página foi baixado e suas informações extraídas via Jsoup (um parser de html em Java). Os dados extraídos foram: *id*, *texto*, *data* e o sentimento do tweet (positivo ou negativo) quando disponível. O sentimento foi extraído por meio da própria ferramenta de busca do Twitter, utilizando emoticons (exemplos são ':)') para positivo e ':(' para negativo).

Para o Superstar foram coletados tweets de Fevereiro até Outubro de 2014, período que precede o início e sucede o término do programa em exatos três meses respectivamente. Consideramos neste estudo somente as quatro bandas finalistas (em negrito), cujos tweets foram coletados por meio dos seguintes conjuntos de *hashtags*: **Suricato**={*Suricato*, *SuricatoBanda*, *BandaSuricato* e *suricatooficial*}, **Jamz**={*Jamz*, *JamzBanda*, *BandaJamz*, *Jamerz* e *jamzoficial*}, **Luan Estilizado**={*LuanEstilizado*, *LuanE-OforroEstilizado*, *LuanEForroEstilizado*}, **Malta**={*AvanteMalta*, *ExercitoAzul*, *MaltaCampea*, *Malta*, *BandaMalta*, *MaltaBanda*, *malta_oficial*}. Já para o BBB foram coletados tweets no período de 01 de Janeiro a 10 de Abril de 2015, que corresponde a vinte dias antes do início do programa e três dias após o término do programa respectivamente. A busca foi feita utilizando o nome do participante seguido de 'BBB', por exemplo '*Fernando BBB*'. No total foram coletados 193.415 tweets sobre o BBB e 9.938 sobre o Su-

perstar. No caso do Superstar foram coletados 73 tweets com data anterior, 5.930 durante e 3.935 com data posterior ao programa.

Como os períodos de exibição dos programas são diferentes, o BBB é um programa diário enquanto o Superstar é semanal, nossas análises levam em consideração que os dados do Superstar são agrupados por semana e os do BBB por dia.

Outra observação importante é que ao analisar manualmente o texto dos tweets com sentimentos do Superstar, é possível notar que a grande maioria dos tweets classificados como negativos tinham na verdade teor positivo em relação à banda. A negatividade do tweet se dava por outros motivos, como, por exemplo, alguma impossibilidade de votar a favor da banda. O tweet a seguir coletado sobre a banda *Malta* exemplifica essa situação: '*Gente, a Banda Malta tem twitter? Não achei :(((*'. Por esse motivo, consideramos todos os tweets negativos do Superstar como sendo na realidade positivos.

4. Resultados e Discussão

Nesta seção nós conduzimos uma ampla análise dos dados coletados com o objetivo principal de responder as seguintes perguntas:

- P1. *A popularidade dos participantes em termos de quantidade de tweets apresenta alguma relação com o desempenho dos mesmos no programa?*
- P2. *A quantidade de tweets positivos dos participantes apresenta alguma relação com o desempenho dos mesmos no programa?*

4.1. Análise de tweets do Superstar

A Figura 1.a mostra a série temporal da quantidade de tweets para cada banda. Note que todas as bandas apresentam a mesma tendência, isto é, possuem poucos tweets antes do programa, durante o programa tem um pico de popularidade e após o programa tem uma queda porém continuam mais populares que anteriormente ao programa. Na final do programa (transição entre o fundo rosa e verde no gráfico) a banda *Malta* tem uma quantidade muito maior de tweets que as outras bandas. É interessante notar que a quantidade total de tweets durante o programa, casa exatamente com a ordem do resultado final do programa, isto é, a banda *Malta* em primeiro com 1738 tweets, seguida da bandas *Jamz* com 429, *Luan Estilizado* com 265 e *Suricato* com 123. A Figura 1.b mostra a mesma série temporal para os tweets positivos, que apresenta basicamente a mesma tendência da Figura 1.a.

Observando apenas a banda campeã, nota-se que desde o início do programa ela já vinha sendo a mais comentada, aumentando a diferença para as outras bandas na metade do programa. Isso indica que nesse momento já teríamos fortes indícios de que ela seria a campeã. Por volta de três quartos do programa a quantidade de tweets de cada banda se encontra na mesma ordem do resultado final do programa.

4.2. Análise de tweets do BBB

No caso do BBB, concentramos nossa análise nos últimos quatro paredões (eliminação por votação do público). A Figura 2 mostra as distribuições da quantidade de tweets positivos para cada paredão. Note que, em todos os paredões, o participante que teve menor número de tweets sempre foi o eliminado, o que nos leva a conjecturar que quanto mais tweets positivos maior a chance do participante continuar no programa.

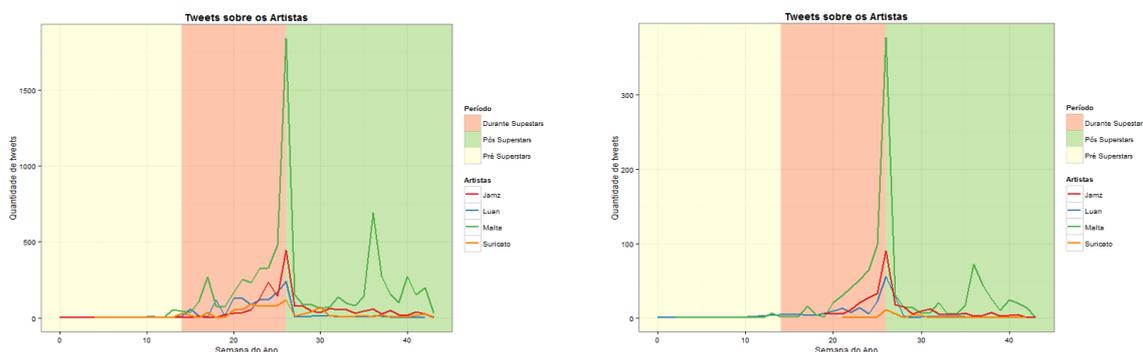


Figura 1. (a) Série temporal dos tweets dos finalistas do Superstar & (b) Série temporal dos tweets positivos dos finalistas do Superstar

A Figura 3 contrasta as porcentagens totais de tweets, tweets positivos e votação real² nos quatro últimos paredões do programa, considerando somente o período que compreende a definição do paredão e dia de votação (regiões cinza da Figura 2). Para uma melhor visualização a quantidade de menções está representada em porcentagem. É interessante notar que no BBB os eliminados do programa são aqueles que tem menos tweets positivos.

4.3. Discussão

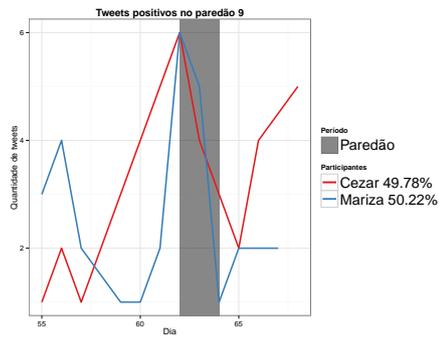
É importante ressaltar que o formato dos programas investigados é diferente. O Superstar é um programa semanal onde cada banda se apresenta e tem basicamente o mesmo tempo de exposição. Dessa forma, os tweets basicamente tendem a endossar as preferências do público pelas bandas. Já os tweets do BBB tratam de vários assuntos muitos dos quais não estão relacionados com intenções de voto. Por exemplo, a existência de um triângulo amoroso envolvendo os participantes do BBB *Fernando, Amanda e Aline*, acabou gerando vários tweets sobre esses participantes³.

Outra observação importante é o fato de que no Superstar todas as bandas se esforçavam para conseguir popularidade no Twitter, por meio de hashtags oficiais, enquanto apenas alguns participantes do BBB fizeram isso, por exemplo o participante Fernando por meio da hashtag *#FernandoProtagonistaDoBBB*.

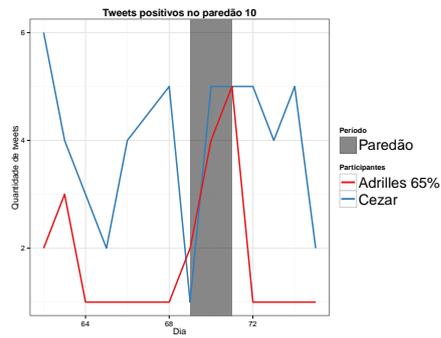
Temos as seguintes considerações às perguntas elencadas no início desta Seção. Para a pergunta P1, conforme foi mostrado e discutido durante toda a seção, a quantidade de tweets tem de fato uma forte relação com o desempenho dos participantes em ambos os programas, especialmente nos dias de votação. Em relação à pergunta P2, observamos que a quantidade de tweets positivos está de fato diretamente relacionada ao desempenho dos participantes em ambos os programas.

²A votação diz respeito à eliminação, ou seja, sai do programa quem recebe mais votos. Com exceção da final, que é o contrário.

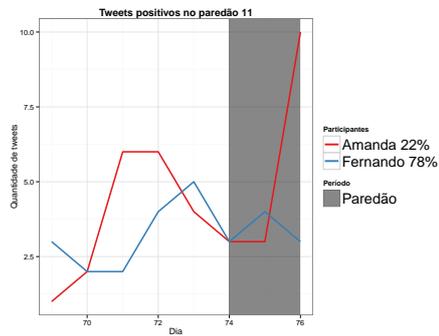
³<http://www.bbb-2015.com/2015/04/votacao-final-cezar-ou-amanda-quem-vai-ganhar.html>



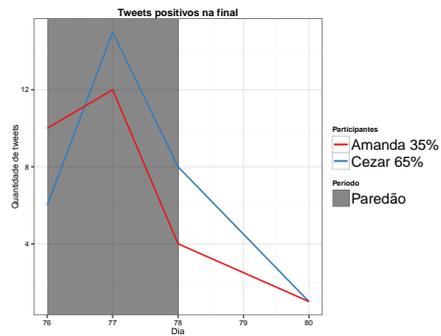
(a) Paredão 9



(b) Paredão 10

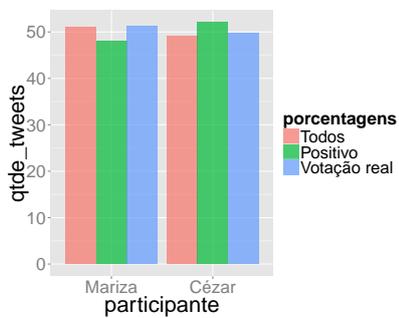


(c) Paredão 11

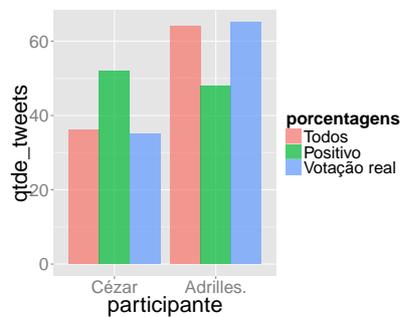


(d) Final

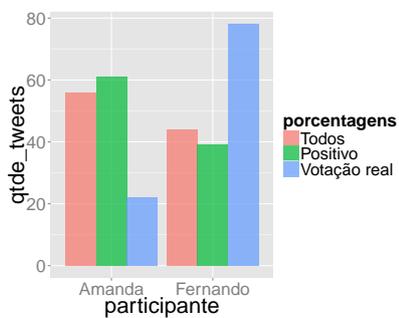
Figura 2. Análise temporal dos tweets positivos para os quatro últimos paredões



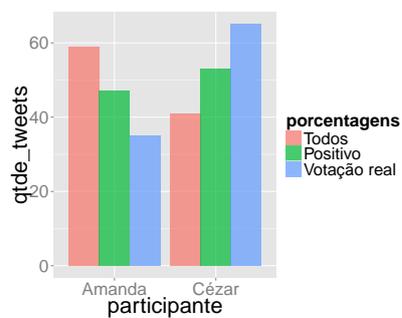
(a) Paredão 9



(b) Paredão 10



(c) Paredão 11



(d) Final

Figura 3. Quantidades de tweets totais e positivos em porcentagem

5. Conclusões e Trabalhos Futuros

Neste artigo nós investigamos a sinergia entre o desempenho de participantes de programas de votação popular da grande mídia no Brasil e a sua popularidade no Twitter durante o programa. Para isso, nós coletamos dados do Twitter acerca das edições mais recentes do Superstar e Big Brother Brasil, dois programas de grande audiência na mídia brasileira.

Notamos que a quantidade total de tweets tem uma forte relação com o desempenho dos participantes dos programas. Enquanto no Superstar a quantidade total de tweets está relacionada a quem fica no programa, no BBB está relacionada a quem é eliminado. Já a quantidade de tweets no dia da votação, nos dois programas, é mais discriminativa para determinar quem receberá mais votos do que os tweets agregados ao longo de todo o programa. Em ambos os programas, a quantidade de tweets positivos está diretamente relacionada ao desempenho dos participantes do programa.

Como trabalhos futuros pretendemos repetir os experimentos do presente artigo para as próximas edições do Superstar e BBB de forma a verificar se as conclusões aqui tiradas continuam válidas. Também pretendemos, a partir dos resultados aqui apresentados, construir modelos de predição para estes programas. Finalmente, também pretendemos aplicar/desenvolver métodos mais sofisticados de análise de sentimento para os tweets acerca desses programas.

Agradecimentos

Gostaríamos de agradecer a Ana Carolina Cabral de Paiva e Matheus Santos Sampaio por terem contribuído com a coleta e análises preliminares dos dados. Este artigo foi desenvolvido em cooperação com a Hewlett-Packard Brasil Ltda., no contexto do projeto HP/UFCG Analytics 2015, e utilizou incentivos da Lei de Informática brasileira (lei n. 8.248 de 1991). Este trabalho também foi apoiado (em parte) pelo Instituto Nacional de Ciência e Tecnologia para Engenharia de Software (INES⁴), financiado pelo CNPq e FACEPE, processos 573964/2008-4 e APQ-1037-1.03/08.

Referências

- Bandari, R., Asur, S., and Huberman, B. A. (2012). The pulse of news in social media: Forecasting popularity. *CoRR*, abs/1202.0332.
- Gayo-Avello, D., Metaxas, P. T., and Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In *Proc. of ICWSM*.
- Gomide, J., Veloso, A., Meira, Jr., W., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proc. of Web Science Conference*, pages 3:1–3:8.
- He, J., Shen, W., Divakaruni, P., Wynter, L., and Lawrence, R. (2013). Improving traffic prediction with tweet semantics. In *Proc. of IJCAI*, pages 1387–1393.
- Hong, L., Dan, O., and Davison, B. D. (2011). Predicting popular messages in twitter. In *Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11*, pages 57–58, New York, NY, USA. ACM.
- Szabo, G. and Huberman, B. A. (2010). Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88.

⁴www.ines.org.br