

# **Análise dos Influenciadores dos Protestos Brasileiros de 2013 via *Twitter***

**Ingrhid Theodoro, Fabio Rangel, Tiago Cruz França, Jonice Oliveira**

Universidade Federal do Rio de Janeiro (UFRJ) Rio de Janeiro, RJ – Brasil

{indytheodoro, fabiorangel, tcruz.franca, jonice}@ufrj.br

***Abstract.** This paper describes an approach of searching by influential users on Twitter, related to protests which took place in Brazil in 2013. The analyzed social network was constructed from the most referenced hashtags and retweets. This paper shows the results of this study, as well as exemplifies how events affect in the communication of Twitter. Moreover, we describe some future works.*

***Resumo.** Este artigo apresenta uma abordagem na busca por usuários influenciadores no Twitter, no tema das manifestações ocorridas no Brasil em 2013. A rede social analisada foi construída a partir das hashtags mais referenciadas no assunto e dos retweets. Este artigo mostra os resultados do estudo, bem como exemplifica como acontecimentos influenciam na comunicação do Twitter. Além da discussão dos resultados, são apontados trabalhos futuros.*

## **1. Introdução**

O *Twitter* é um site de *microblogging* onde se tornou comum publicar mensagens sobre acontecimentos/eventos quase em tempo real [Naaman e Boase 2010]. Em 2013, vários protestos ocorreram no Brasil – onde milhões de pessoas foram às ruas em diferentes dias e localidades - e tiveram repercussão mundial [Globo TV 2013]. Nessas ocasiões as mídias sociais (ou rede social *online*<sup>1</sup>) se tornaram um importante meio de comunicação entre os manifestantes.

O objetivo deste trabalho é a identificação de usuários influentes nas manifestações. As métricas utilizadas basearam-se na definição de ‘prestígio’ detalhada em [Wasserman e Faust 1994]. Neste caso, definimos como ‘influência’ o poder de propagação e repercussão das mensagens emitidas por um usuário. Quanto maior a frequência, alcance e rapidez das mensagens, maior a influência deste usuário sobre a sua rede social.

Este trabalho está organizado em 5 seções. Os trabalhos relacionados são discutidos na Seção 2. Na Seção 3 é descrita a metodologia adotada. A Seção 4 apresenta as análises utilizadas e os resultados obtidos. Na Seção 5 é apresentada a conclusão e trabalhos futuros.

## **2. Trabalhos Relacionados**

O *Twitter* se tornou uma mídia importante para organização e mobilização de eventos [Sakaki et al. 2010]. A população mexicana utilizou o *Twitter* e o *Youtube* para

---

<sup>1</sup> Neste trabalho utilizamos dois conceitos. O primeiro é ‘rede social online’ que se refere a alguma ferramenta de auxílio à interação (como por exemplo, o *Twitter*). O segundo conceito se refere à ‘rede social’, que é o conjunto de interações feitas entre pessoas e que pode ser representada por um sociograma.

demonstrar sua insatisfação contra a ocultação de informações relacionadas às eleições mexicanas [Almazan e Garcia 2013]. Eltantawy e Wiest [2011] realizaram uma análise sobre como o *Twitter*, *Facebook*, *Youtube*, *blogs* e até mesmo aplicativos de mensagens de celular têm exercido influência e auxílio na propagação da informação, bem como na organização dos movimentos sociais contemporâneos. Destacaram também nomes que influenciaram e encorajaram a sociedade, entretanto, as métricas utilizadas não foram detalhadas.

Alguns trabalhos mostraram que as características topológicas da rede podem ser utilizadas na classificação de usuários do *Twitter*. Kwak et al. [2010] aplicou o *PageRank*, na rede social criada a partir dos seguidos e seguidores, comparando com o resultado do número de seguidos e número de *retweets*. Embora todos os trabalhos lidem com o tema de manifestações, apenas Recuero [2013] abordou as manifestações brasileiras. Em sua análise, palavras e *hashtags* foram agrupadas em *clusters* sobre as principais demandas.

Este trabalho diferencia-se dos demais por utilizar mais métricas na análise topológica e pela construção da rede social através dos *retweets* (ao contrário de seguidos e seguidores), sendo mais significativa ao representar a propagação da informação e a concordância com o conteúdo de uma mensagem.

### 3. Metodologia

Os *tweets* foram coletados do dia 23 de junho a 02 de agosto de 2013. Os dados coletados se deram por meio do uso de um *crawler*, que obtém os *tweets* acessando a API oficial de busca do *Twitter*<sup>2</sup>. As principais *hashtags* utilizadas foram: #acordaBrasil, #vemPraRua, #oGiganteAcordou, #anonymousBrazil, #MPL, #passeLivre, #Pec37, #mudaBrasil, #changeBrazil, #anonymousBrasil, #protesto, #foraDilma, #protestoRJ, #protestaBrasil, #primaveraBrasileira, #foraFeliciano, #ocupa, #copaPraQuem, #protest, #Pec33, #Pec99.

A base de dados brutos, formada com os dados obtidos no formato JSON, contém a mensagem (*tweet*) e outros campos de dados (como id, data de publicação, etc.). No próximo passo foi realizado um pré-processamento de todo o conteúdo dessa base. Com a nova base foi possível criar grafos direcionados, onde seus vértices representam usuários do *Twitter* e as arestas representam o fluxo da informação. A direção da aresta é: usuário “retweetado” ao usuário que “retweetou”. A fim de obter uma análise temporal desta rede, construiu-se 20 grafos em intervalos de dois dias.

A definição das métricas utilizadas nas análises foi baseada na definição de ‘prestígio’ feita por Wasserman e Faust [1994]. Utilizamos este conceito como análogo ou correlato à ‘influência’. Os autores definem três tipos de prestígio: de grau, de proximidade e de *status* ou *rank*. O prestígio de grau está relacionado à quantidade de associações diretas. Quanto mais relacionamentos uma pessoa tiver, maior o seu prestígio. Análogo a isto, utilizamos o grau de saída do nó e o seu grau médio. O prestígio de proximidade considera que os vértices de maior prestígio são aqueles mais próximos de todos os demais. Neste caso, utilizamos a Centralidade de Autovetor. O prestígio de *status* ou *rank* considera como vértices de maior prestígio aqueles que são pontes. Para nós, consideramos como maior *status* o nó mais referenciado na rede,

---

<sup>2</sup> <https://dev.twitter.com/docs/api/1.1>, acessado em 07/04/2014.

utilizando a métrica de *PageRank*.

A evolução da rede e as interações existentes nela foram analisadas por meio da verificação do diâmetro e quantidade de nós. Os grafos e cálculos das métricas foram feitos com o *Gephi* versão 0.8.2.

## 4. Análises dos Resultados

Esta seção apresenta a descrição das métricas utilizadas e os resultados obtidos.

### 4.1. Diâmetro da Rede

O diâmetro de um grafo direcionado é o comprimento do caminho geodésico mais longo entre qualquer par de nós [Wasserman e Faust 1994]. No presente trabalho, o diâmetro foi definido pela maior distância dentro de uma mesma componente conexa.

Observando a Figura 1a, percebe-se que o diâmetro da rede diminui com o passar do tempo, variando à medida que o tempo avança. Relacionando os gráficos da Figura 1a e 1b, é possível perceber que existe uma relação direta entre a queda do número de usuários com a queda do diâmetro. É possível cogitar que alguns usuários continuaram a manter tópicos daquele tema ativos, mesmo com a quantidade de usuários interessados diminuindo, dando origem a um grupo mais conectado. Há também a possibilidade do tema ter se modificado, adotando outras *hashtags* que não estavam sendo buscadas pelo *crawler* descrito na metodologia.

### 4.2. Grau de Saída e Grau Médio

O grau de saída de um vértice é obtido pela quantidade de arestas (*retweets*) adjacentes a esse vértice, cuja direção seja a de saída do mesmo. A relação dos *retweets* é: quando o usuário A é “retweetado” pelo usuário B, existirá uma aresta no grafo que sai do vértice A para o B. Usuários influenciadores foram identificados por meio da verificação do seu grau de saída, ou seja, aqueles que são “retweetados”. Quanto mais *retweets* um usuário recebe, maior será sua influência.

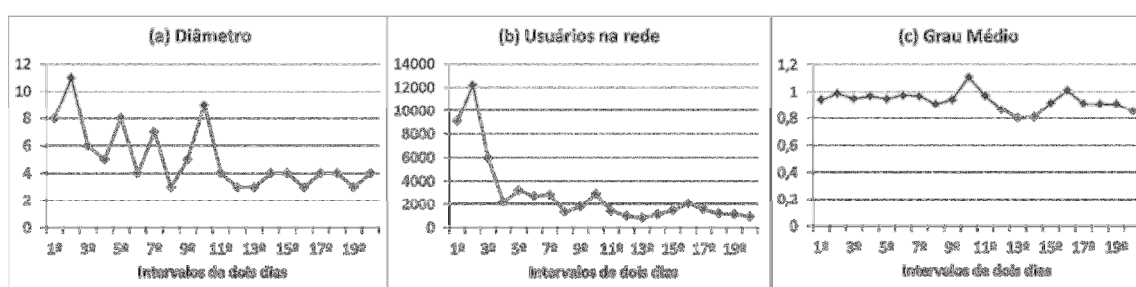


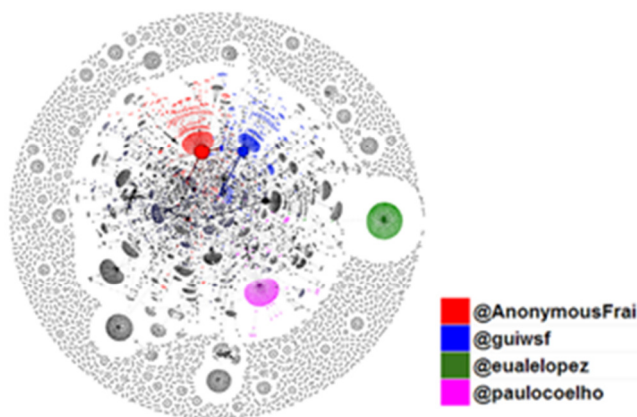
Figura 1. Diâmetro (gráfico a), quantidade de usuários (gráfico b) e grau médio da rede (gráfico c) por intervalos de dois dias

A média dos graus fornece a média de *retweets* por usuário. Como a análise foi realizada com base em *retweets*, o grau médio pôde ser obtido através da divisão do número de arestas pelo número de vértices. É possível observar a variação do grau médio através da Figura 1c. Nota-se que essa variação não foi alta, o que indica que a média de *retweets* por usuários da rede sofreu uma leve variação independente do intervalo.

### 4.3. Centralizadores e Influenciadores

O grafo do primeiro intervalo (23 e 24 de junho) pode ser visualizado na Figura 2. Neste intervalo, a rede contava com 9.107 vértices. Alguns *clusters* (comunidades) foram destacados como exemplo e, na legenda, o nome do usuário núcleo. Nessa figura, cada vértice aumenta de tamanho conforme o seu grau de saída.

Ainda na Figura 2, o vértice de maior destaque, que possui alto grau de saída, é o @AnonymousFrai que teve 432 *retweets* nesse intervalo. Esse usuário é considerado um influenciador da rede, já que seus *tweets* são disseminados (“retweetados”) por muitos usuários. Também existem *clusters* onde o núcleo é um vértice que não se destaca e possui um grau de saída baixo. O alto grau de entrada é uma indicação de disseminadores do tema. Por exemplo, o *cluster* cujo núcleo é o usuário @paulocoelho (nome de usuário do *Twitter* do escritor Paulo Coelho) possui grau de entrada elevado, indicando o interesse do escritor sobre o tema. Considerando sua quantidade de seguidores (mais de nove milhões), pode-se dizer que ele disseminou bastante o tema, porém, como o seu grau de saída na rede é zero, ele não seria um influenciador no tema abordado.



**Figura 2. Visualização do grafo gerado com dados do primeiro intervalo.**

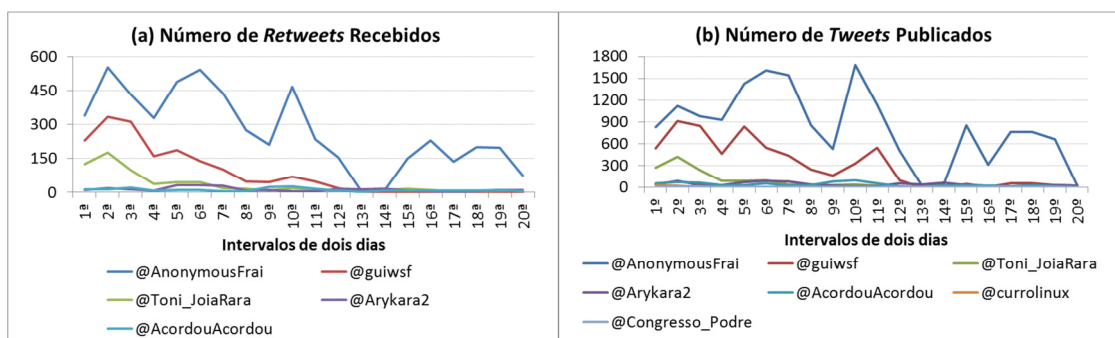
Uma avaliação durante o período em que os dados coletados foram publicados permite que seja analisado o grau de influência (número de *retweets*) de cada usuário ao longo do tempo. A Figura 3a apresenta o gráfico da quantidade de *retweets* por intervalos de dois dias dos usuários mais “retweetados” apresentados na Tabela 1.

**Tabela 1. Os 10 usuários que receberam a maior média de *retweets*.**

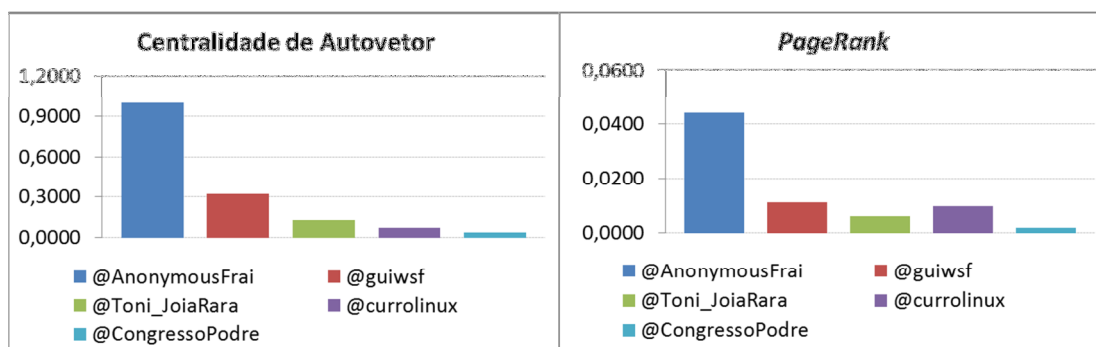
Usuário	Média dos Graus de Saída	Usuário	Média dos Graus de Saída
@AnonymousFrai	272,35	@Sangue_Latino	8,5
@guiwsf	83,9	@MariahBretzke	8,3
@Toni_JoiaRara	32,8	@xandreIm	5,55
@Arykara2	11,85	@BR_Indignado_	5,4
@AcordouAcordou	10,1	@venusaquario	4,9

O gráfico da Figura 3b apresenta o número de *tweets* por intervalo dos usuários mais influentes da rede. Analisando o gráfico, percebe-se que houve uma queda significativa na quantidade de publicações dos usuários @AnonymousFrai e @guiwsf no 13º e 14º intervalos. Não há como informar o motivo da queda, porém é perceptível que os outros usuários também tiveram quedas acentuadas. Porém, observando os dois gráficos da Figura 3, percebe-se ainda que a quantidade de *retweets* que o usuário @AnonymousFrai recebe é relativamente proporcional à quantidade de *tweets* que ele publica, reforçando a influência desse usuário.

Pode ser notado que os usuários @currolinux e @Congresso\_Podre foram adicionados ao gráfico da Figura 3b, embora eles não estivessem contando na Tabela 1. Esses usuários foram destacados graças à outra análise feita na topologia da rede, aplicando-se duas métricas de centralidade em redes complexas: Centralidade de Autovetor [Bonacich 2007] e *PageRank* [Page et al. 1999].



**Figura 3. Relação de retweets e tweets dos usuários mais influentes da rede por intervalos de 2 dias.**



**Figura 4. Comparação entre Centralidade de Autovetor e PageRank**

Para o cálculo do *PageRank* e da Centralidade de Autovetor, foi preciso inverter a direção das arestas. Esse motivo se justifica quando a direção da aresta, no primeiro momento, estava na direção do fluxo da informação, e desta forma, usuários muito influentes seriam os que possuíssem alto grau de saída. No entanto, no cálculo dos nós centrais da rede, as métricas levam em consideração o grau de entrada dos nós como sendo um parâmetro importante (e no nosso caso representaria a referência a um usuário). Assim, foi tomado como medida para contornar esse problema a inversão da direção das arestas. A métrica foi calculada utilizando o grafo formado por todos os intervalos de tempo. Os resultados podem ser observados nos dois gráficos da Figura 4.

Era esperado que o usuário mais influente fosse o @AnonymousFrai seguido pelo usuário @guiwsf, o que se confirmou após a métrica. Porém, outros dois novos usuários que nem haviam sido identificados surgiram graças à topologia das redes. Sendo assim, há duas formas de se pensar. A primeira identificaria o usuário @Toni\_JoiaRara como um usuário de maior influência que o usuário @currolinux, já que a média de *retweets* dele é maior. Essa teoria se confirma no cálculo da centralidade de autovetor. Outra forma de se pensar seria que, mesmo o usuário @currolinux possuindo menos *retweets* que o @Toni\_JoiaRara, @currolinux foi “retweetado” por usuários de maior *PageRank*, logo, melhor contabilizado no cálculo desta métrica.

## 5. Considerações Finais

Este trabalho teve como objetivo identificar os usuários influenciadores no *Twitter* durante as manifestações ocorridas em 2013 no Brasil. Foram analisados 20 intervalos de tempo, compostos por dois dias. Para trabalhos futuros, espera-se identificar os usuários influenciadores por subtemas. Também seria possível identificar os usuários influenciadores por regiões do país, utilizando do *Twitter* a informação de *geocode* pertencente a um *tweet*. O *geocode* permite identificar o local de origem onde foi criado o *tweet*. Como limitação deste trabalho destacou-se a ausência de pesquisa qualitativa que comprove, através de pesquisa de opinião com os membros da rede social traçada, a importância dos membros identificados como influenciadores. Tal análise qualitativa será realizada futuramente.

## Agradecimentos

Ao CNPq e à FAPERJ.

## Referências

- Almazan, R. S. and Garcia, J. R. G. (2013), Cyberactivism through Social Media: Twitter, YouTube, and the Mexican Political Movement “I’m Number 132”, 46th Hawaii International Conference on System Sciences.
- Bonacich, P. (2007), “Some unique properties of eigenvector centrality”, *Social Networks*, Volume 29, Issue 4, October 2007, Pages 555–564.
- Eltantawy, N. and Wiest, J. B. (2011), Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory, *International Journal of Communication* 5, p.1207–1224.
- Globo TV (2013), “Jornais de todo o mundo repercutem os protestos no Brasil”, <http://globo.com/rede-globo/jornal-hoje/v/jornais-de-todo-o-mundo-repercutem-os-protestos-no-brasil/2641286>, acessado em 03/04/2014 às 22:30 (horário de Brasília).
- Naaman, C.-H. L. Mor., and Boase, J. (2010), “Is it all About Me? User Content in Social Awareness Streams”, *ACM Conference on Computer Supported Cooperative Work*, 2010.
- Page, L., Brin, S., Motwani, R. and Winograd, T. (1999), “The PageRank Citation Ranking: Bringing Order to the Web”, Technical Report. Stanford InfoLab.
- Recuero, R. (2013), “Os discursos das Hashtags”, <http://www.raquelrecuero.com/arquivos/2013/06/os-discursos-das-hashtags.html>, acessado em 07/04/2014.
- Wasserman, S. and Faust, K. (1994), *Social Network Analysis. Methods and Applications* Cambridge University Press, Cambridge, England.