Bots Sociais: Como robôs podem se tornar pessoas influentes no Twitter?

Johnnatan Messias[1], **Fabrício Benevenuto**[2], **Ricardo Oliveira**[1]

[1]Departamento de Computação – Universidade Federal de Ouro Preto, UFOP, Brasil

[2]Departamento de Ciência da Computação - Universidade Federal de Minas Gerais, UFMG, Brasil

{johnnatan20@gmail.com, fabricio@dcc.ufmg.br, rrabelo@gmail.com}

Abstract. *Systems like Klout and Twitalyzer were developed as an attempt to measure the influence of user within social networks. Although the algorithms used by these systems are not known, they have been widely used to rank users according to their influence in the Twitter social network. As media companies might base their viral marketing campaigns on influence scores, in this paper, we investigate if these systems are vulnerable and easy to manipulate. Our approach consists of developing Twitter robot accounts able to interact with real users in order to verify strategies that can increase their influence scores according to different systems. Our results show that it is possible to become influential using very simple strategies, suggesting that these systems should review their influence score algorithms to avoid accounting with automatic activity.*

Resumo. *Sistemas como o Klout e Twitalyzer foram desenvolvidos como uma tentativa de medir a influência de usuários dentro de redes sociais. Embora os algoritmos utilizados por estes sistemas não sejam conhecidos eles têm sido amplamente utilizados para classificação de usuários de acordo com a sua influência na rede social Twitter. Como as empresas de mídia podem basear suas campanhas de marketing viral baseando-se no ranking de influência, neste trabalho, investigamos se esses sistemas são vulneráveis e de fácil manipulação. A nossa abordagem consiste em desenvolver contas robot capazes de interagir com os usuários reais a fim de verificar as estratégias que podem aumentar a sua pontuação de influência de acordo com os diferentes sistemas. Nossos resultados mostram que é possível tornar-se influente usando estratégias simples, sugerindo que esses sistemas devem rever seus algoritmos de influência para evitar a contabilização de contas com atividades automáticas.*

1. Introdução

As redes sociais têm adquirido bastante destaque e crescente importância na sociedade moderna. Dentre elas, o *Twitter* tem conquistado grande quantidade de adeptos. É uma das redes sociais que mais movimenta usuários e uma das que mais geram troca de informações. Em sistemas como o *Twitter*, usuários podem influenciar e serem influenciados por outros, o que tem atraído grande interesse político e de empresas relacionadas ao *marketing*.

Neste contexto, várias empresas tem se especializado em medir influência no *Twitter* e em outras redes sociais. Dentre os sistemas mais populares estão o *Klout*[1] e o *Twi-*

talyzer[2], que utilizam abordagens de medições de influência simples e cujos detalhes não são revelados ao público.

Surpreendentemente, vários jornais, revistas e até mesmo artigos científicos [3], [4] tem utilizado essas ferramentas, e divulgado *rankings* de usuários influentes. Como exemplo, uma matéria do jornal *The New York Times* apresentou um estudo das pessoas mais influentes do mundo, baseado no *Twitalyzer* [5]. Entre os mais influentes, o estudo indicou o humorista brasileiro Rafinha Bastos, o *rapper* americano Snoop Dogg, o presidente dos Estados Unidos Barack Obama e o apresentador brasileiro Luciano Huck.

Baseado nisso, queremos investigar que estratégias podem tornar usuários comuns em usuários influentes, segundo *rankings* do *Klout* e *Twitalyzer*. Criamos algoritmos *bots* simples capazes de interagir, através de contas no *Twitter*, como se fossem usuários na rede, trocando informações, seguindo e conquistando novos seguidores, durante um período de 90 dias. Os resultados das análises nos mostraram que as ferramentas *Klout* e *Twitalyzer* não apresentaram métricas ideais para classificar a influência de um indivíduo na rede. Constatamos que *spammers* espalhados na rede podem facilmente se passar por celebridades e pessoas populares, com um alto nível de influência baseado no *Klout Score* e *Twitalyzer Impact*.

Este trabalho foi realizado pelo aluno Johnnatan Messias, estudante de Ciência da Computação na Universidade Federal de Ouro Preto e responsável por todas as etapas de sua execução. O artigo é fruto de um projeto da disciplina Análise de Mídias Sociais ministrada pelo Prof. Fabrício Benevenuto (orientador) e também contou com a colaboração do Prof. Ricardo Oliveira (co-orientador).

2. Trabalhos relacionados

Alguns estudos apontam que um número enorme de seguidores de um usuário no *Twitter* é suficiente para determinar o seu grau de relevância. Outros estudos indicam que a influência de uma pessoa no *Twitter* está muito mais ligada ao conteúdo que ela compartilha do que com o número de pessoas que a seguem. Houveram várias tentativas e métodos para medir, de forma adequada, a influência dos usuários no *Twitter* [6], [7]. Em [8] é proposto o *TwitterRank*, um algoritmo baseado no *PageRank*, que usa tanto o grafo de conexões do *Twitter* quanto informações de *tweets* publicados para identificar usuários influentes. Por outro lado [9] usou agrupamento e classificação de mais de 15 características extraídas do grafo do *Twitter* e dos *tweets* postados por usuários para identificar os mais influentes.

Na referência [10] foi investigado a atividade de *link farms* no *Twitter* e propuseram formas para dissuadir tal atividade. Constataram que um pequeno grupo de usuários legítmos, populares e ativos no *Twitter* são responsáveis pela maioria dos *link farms*. Buscam acumular capital social seguindo diversos usuários e conquistando novos seguidores. Com isso, *spammers* exploram este tipo de comportamento para conquistar seguidores e reputação no *Twitter*. Como forma de minimizar a influência deste tipo de comportamento, propuseram um esquema de classificação, conhecido como *Collusionrank*, onde os usuários são penalizados por seguir *spammers*, reduzindo a influência de *spammers* e seus seguidores.

Na referência [11] foi verificado como as *OSN's* são vulneráveis devido a infiltração em larga escala pelos *socialbots*: programas de computador capazes de controlar contas nas redes sociais e imitar usuários reais. Para isso, adotaram uma tradicional *bot-*

net baseada na *web* e construíram uma *Socialbot Network*. Foram coletados dados dos comportamentos dos usuários do Facebook, utilizando os *socialbots*. Acreditam, ainda, que infiltração em larga escala nas *OSN's* é somente uma das muitas outras ameaças na rede.

Uma versão mais detalhada desse trabalho pode ser encontrada em [12].

3. Construção dos bots

Antes de explicar o funcionamento dos *bots*, primeiramente vamos apresentá-los. Foram criados dois algoritmos *bots*, usando a *API Python* do *Twitter*. O primeiro *bot*, denominado ***fepessoinha*** (https://twitter.com/fepessoinhas2) apenas segue automaticamente outros usuários e conquista novos seguidores. Para seguir os usuários, utilizamos o método de busca em largura, ou seja, a partir de um usuário inicial aleatório o algoritmo o inclui em uma lista, seguindo-o e selecionando aleatoriamente 30 usuários de sua lista de seguidos, que também são incluídos na lista. Feito isso, o processo será repetido de modo a seguir o próximo usuário, até o *bot* alcançar o limite de 2000 usuários seguidos, que é um limite imposto pelo *Twitter*. A Figura 1 mostra uma foto do perfil do *bot* ***fepessoinha***[1].

Figura 1. Perfil bot fepessoinha - Twitter

O segundo, denominado ***scarina*** (https://twitter.com/scarina91), é capaz de interagir, como se fosse um usuário na rede, trocando informações, seguindo e conquistando novos seguidores. Para seguir os usuários, este *bot* utiliza do mesmo princípio utilizado pelo *bot* ***fepessoinha***, também conquistando consequentemente novos seguidores. Porém, este possui duas diferenças: (1)após seguir os 2000 usuários o algoritmo exclui os usuários que não seguem o *bot* e (2)é capaz de postar *tweets*. Para isso, uma sequência do algoritmo lê um dicionário de palavras relacionadas ao tema "Rede Globo". Colocamos no dicionário as seguintes palavras: rede, programação, xuxa, novela, filme, audiência, melhor, reportagem, jornalismo, jornal, nacional, huck, angélica, willian, bonner, bernardes, malhação, notícia, destaques, bbb. Assim, o algoritmo prossegue solicitando uma busca no *Twitter* através do trecho "Globo + palavra", sendo a palavra uma das selecionadas aleatoriamente no dicionário. Em seguida, capta quatro dos *tweets* retornados na pesquisa e os posta em um intervalo de tempo aleatório. Conforme as postagens aumentam, as atividades e interações entre os usuários com o *bot* ***scarina*** também aumentam. A Figura 2 mostra uma foto do perfil do *bot* ***scarina***[1].

Para conseguir automatizar os *bots*, de forma a respeitar as regras impostas pela API do *Twitter*. Os *bots* respeitam as seguintes restrições:

- Número máximo de requisições por hora = 350;
- Cada usuário pode seguir no máximo 2000 usuários, via API;
- Um usuário não pode seguir mais de 1000 usuários por dia.

[1]As fotos utilizadas no perfil de ambos os *bots* são fotos de pessoas que autorizaram o uso de suas fotos nos *bots*.

Figura 2. Perfil bot scarina - Twitter

4. Experimentos com os bots

Para analisar como os *bots* criados adquirem influência segundo o *Klout* e *Twitalyzer*, foram criados quatro cenários experimentais de forma a testar a vulnerabilidade e facilidade de manipulação dos classificadores de influência.

4.1. Cenário 1: Seguindo usuários (*Bot* fepessoinha)

O *bot* ***fepessoinha*** seguiu o limite máximo de usuários permitido pelo *Twitter* (2000), via API, na tentativa de obter seguidores.

4.2. Cenário 2: Seguindo e excluindo usuários (*Bot* scarina)

O *bot* ***scarina*** também seguiu o limite máximo de usuários permitido pelo *Twitter* (2000), via API. Posteriormente, o algoritmo do *bot* exclui alguns usuários, mantendo apenas os que também a seguiram. Neste cenário, o método utilizado para seguir usuários é o mesmo apresentado no cenário 1, porém exclui posteriormente os usuários que não seguem o *bot*.

4.3. Cenário 3: Seguindo, excluindo usuários e postando *tweets* (*Bot* scarina)

O *bot* ***scarina*** posta *tweets* de forma automática, após realizar a estratégia do **Cenário 2**. Tal cenário permite verificar o impacto nos *rankings* de influência devido à postagem de conteúdo. Para isso utilizou um dicionário de dados para postagens de temas relevantes. Conforme as postagens aumentam, as atividades e interações entre os usuários com o *bot* ***scarina*** também aumentam.

4.4. Cenário 4: Seguindo, excluindo usuários, postando *tweets* e com interrupções na execução (*Bot* scarina)

O *bot* ***scarina*** segue e ganha novos seguidores além de postar *tweets* utilizando também o dicionário de dados. Neste cenário também verificamos e analisamos algumas interrupções e reinícios na execução do algoritmo do *bot*. Com o intuito de analisar o que acontece com a influência do *bot* quando este fica períodos de tempo inativo.

5. Coleta de dados

Para coletar os dados pertinentes para este artigo, na criação dos *bots* ***scarina*** e ***fepessoinha*** configuramos suas contas do *Twitter* a fim de enviar todas as informações aos *e-mails* cadastrados. Assim, teríamos um *log* de todas as ações, informações e atividades realizadas durante o período de 90 dias de experimento. Posteriormente, coletamos os *e-mails*, em arquivos ".eml", do servidor do *Yahoo* (serviço de *e-mail* das contas) e fizemos um algoritmo *parser* para filtrar as informações desejadas. Com isso, durante os

90 dias, conseguimos obter as seguintes informações do *Twitter*: *tweets* favoritos, *tweets* mencionados, mensagens, *tweets* respondidos, *retweets* e usuários seguidores.

Após o processo de execução, acessamos os sites do *Klout* e *Twitalyzer* para coletar os resultados de todas as interações. No site da *Klout* foi possível coletar os resultados dos 90 dias de experimento, através de gráficos. Já no site da *Twitalyzer*, apenas conseguimos um resultado da influência final para cada usuário, desde o dia de sua criação. Não conseguimos dados dos dias anteriores ao de consulta.

Os dados coletados através dos *e-mails* servem para complementar os resultados obtidos pelos sistemas *Klout* e *Twitalyzer* e com isso achamos conveniente mostrar os resultados finais, para comprovar a vulnerabilidade e fácil manipulação dos sistemas *Klout* e *Twitalyzer*.

6. Resultados

Para se chegar ao *Klout Score*, o *Klout* trabalha com mais de 25 variáveis divididas em 3 grupos: *Network Impact* (Influência na rede), *Amplification Probablity* (Probabilidade de amplificação) e *True Reach* (Alcance Verdadeiro). O *Klout Score* vai de 0 à 100, e, teoricamente, é o resultado final do grau de influência de um usuário. No caso do *Twitalyzer*, este possui o número conhecido como *Twitalyzer Impact*. Este número também vai de 0 à 100 e baseia-se em 15 variáveis para se chegar ao valor final.

6.1. Cenário 1

O gráfico correspondente à Figura 3 mostra detalhadamente a quantidade de seguidores que o *bot* ***fepessoinha*** conquistou, resultando ao final do experimento o total de 417 seguidores. A quantidade de seguidores foi mais alta no momento em que o *bot* ***fepessoinha*** executou o processo de seguir os 2000 usuários (durante os primeiros 9 dias).

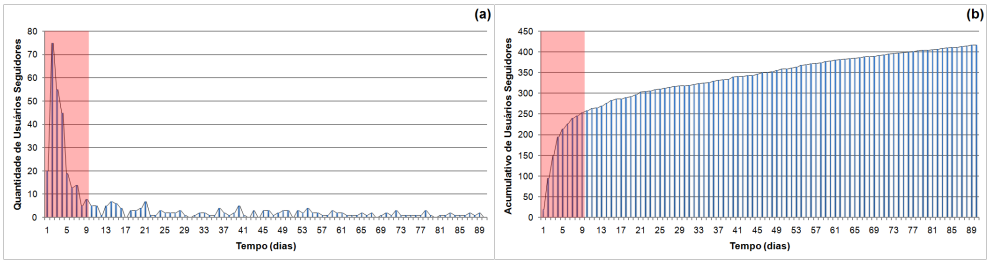

Figura 3. (a) Gráfico de seguidores do bot fepessoinha (b) Gráfico do acumulativo de seguidores do bot fepessoinha

Com isso, podemos ver no gráfico correspondente à Figura 4, o resultado do *Klout Score* do *bot* ***fepessoinha***. O *Klout Score* mostra o resultado final da influência de um usuário. É possível ver que nos 2 primeiros dias há uma grande crescente, devido aos novos usuários que passaram a seguir o *bot*, alcançando o valor 18 de *Klout Score*. Ao fim da execução do algoritmo do *bot* (após 9 dias) é possível ver uma queda constante do *Klout Score* até alcançar o valor 12,3 ao fim dos 90 dias de experimento.

Já no *Twitalyzer*, o *bot* conseguiu o valor 9, durante os 90 dias de experimento. Além desses resultados, o *bot* ***fepessoinha*** obteve durante os 90 dias de experimento: 0

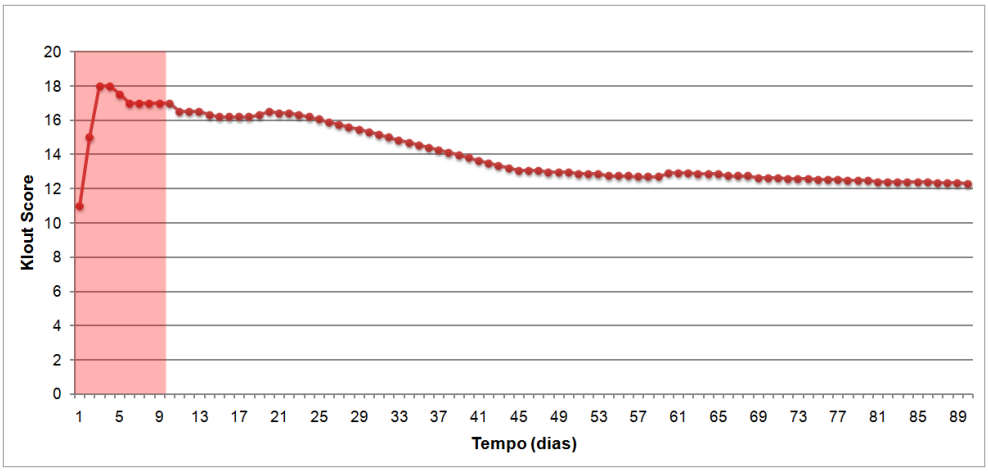

Figura 4. Resultados Klout Score do bot fepessoinha

tweets favoritados e 0 *retweets*, devido ao fato de ter publicado apenas 7 *tweets*. Desses 7, apenas 4 *tweets* foram respondidos. Conforme é mostrado na Figura 5 o *bot* obteve um valor considerável de *tweets* mencionados e mensagens recebidas resultando ao fim do experimento o total de 24 e 21 respectivamente. Observamos que grande parte das menções e mensagens foram feitas nos 9 primeiros dias, nos dias em que o algoritmo de coleta de usuários do *bot* estava em execução. A maioria das menções e mensagens foram de questionamentos sobre quem era o *bot* ***fepessoinha*** e de onde os usuários se conheciam, sendo questionamento comum entre usuários que iniciam um relacionamento em uma rede social.

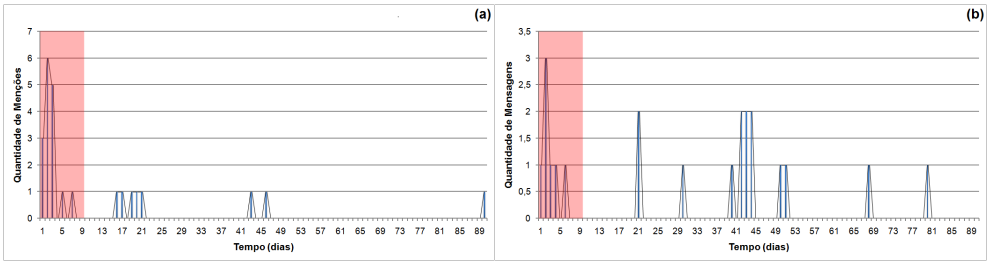

Figura 5. (a) Gráfico de *tweets* mencionados do bot fepessoinha (b) Gráfico de mensagens do bot fepessoinha

6.2. Cenário 2

O gráfico correspondente à Figura 6 é o que expressa melhor o resultado deste experimento. Fazendo uma comparação com o *bot* ***fepessoinha***, o *bot* ***scarina*** alcançou valores menores de influência. É possível ver que o fato do algoritmo do *bot* ***scarina*** excluir os usuários que não o seguem, sua influência acaba sendo menor do que o *bot* ***fepessoinha***, que não exclui usuários, durante os 9 dias em que o algoritmo de coleta e exclusão de usuários foi executado. Após os 9 dias, ao fim do primeiro processo de coleta e exclusão de usuários, é iniciado o processo de postagem de *tweets*, iniciando o processo do **Cenário 3**.

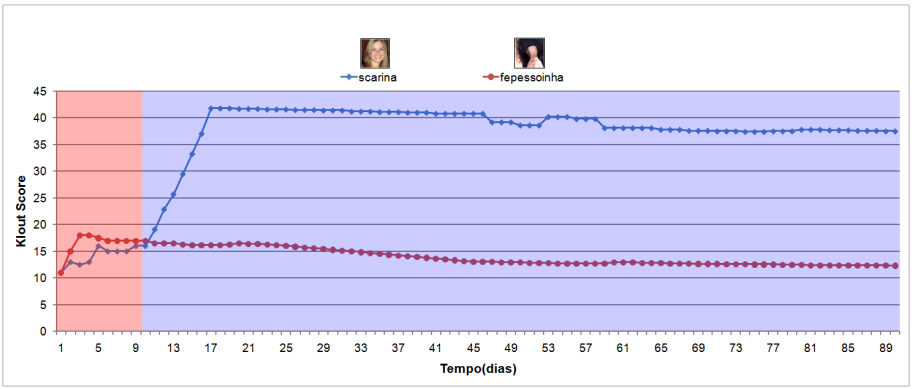

Figura 6. Comparativo do Klout Score entre os bots scarina e fepessoinha

6.3. Cenário 3

A partir do 10º dia, apresentado pelo gráfico correspondente à Figura 6, podemos ver o início de uma discrepância brusca entre os *bots* ***scarina*** e ***fepessoinha***. O *bot* ***scarina***, inicia a publicação de *tweets* a partir do 10º dia, e no 17º dia já atinge o valor 41,8 de *Klout Score*. Observamos no *bot* ***scarina***, através do **Cenário 3**, que o *bot* passa a conquistar novos seguidores mesmo sem seguir outros usuários, apenas pela relevância do conteúdo que posta. O conteúdo que o *bot* posta acaba sendo de interesse de outros usuários que consequentemente passam a seguir o *bot*. Com isso, é possível ver que apenas seguir usuários, como realizado no **Cenário 1**, não é suficiente para atingir um alto grau de influência. Logo, o sistema *Klout*, considera também as publicações de um usuário, para elevar o nível de influência. Enfim, o *bot* ***scarina*** chegou ao valor máximo de 41,8 no *Klout*, enquanto o *bot* ***fepessoinha*** alcançou o valor máximo de 18 no *Klout*.

Além do valor do *Klout*, o *bot* ***scarina*** conseguiu o valor 86 no *Twitalyzer*. Uma comparação precisa, dos níveis de influência mais altos, pode ser verificada no gráfico correspondente à Figura 7.

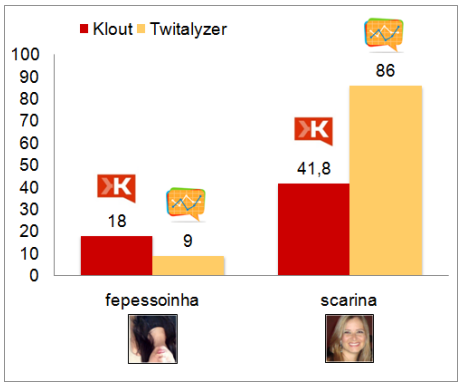

Figura 7. Comparativo Klout e Twitalyzer - fepessoinha x scarina

6.4. Cenário 4

Os resultados para o **Cenário 4** podem ser vistos claramente no gráfico correspondente à Figura 8. É possível ver grandes vales e picos que correspondem a interrupções na execução do algoritmo de postagem de *tweets*. Foram duas interrupções. A primeira interrupção

se deu no 46º dia. Mantivemos o algoritmo de postagem de *tweets* desativado por 6 dias. Reiniciamos o processo completo do *bot* (coleta/exclusão de usuários e postagem) e mantivemos por mais 6 dias a execução, até realizarmos uma nova interrupção. Reiniciamos então apenas o processo de postagem de *tweets*. Através do gráfico, fica evidente que a não publicação de *tweets* (com a interrupção do processo) diminuiu bruscamente o grau de influência do *bot*, baseado no *Klout Score*.

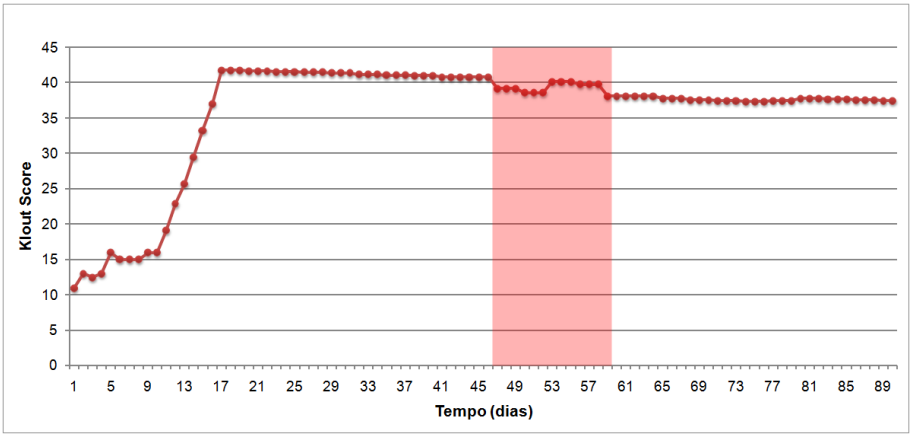

Figura 8. Resultados Klout Score do bot scarina

Conseguimos ver o mesmo comportamento nos gráficos que correspondem aos seguidores, menções e mensagens, principalmente após a primeira interrupção. No gráfico correspondente à Figura 9 podemos ver, a partir do dia 46 que o *bot* conquista diversos novos seguidores, por executar todo o processo inicial, após a primeira interrupção. Ao final de 90 dias de experimento, o *bot* conquistou o total de 691 seguidores.

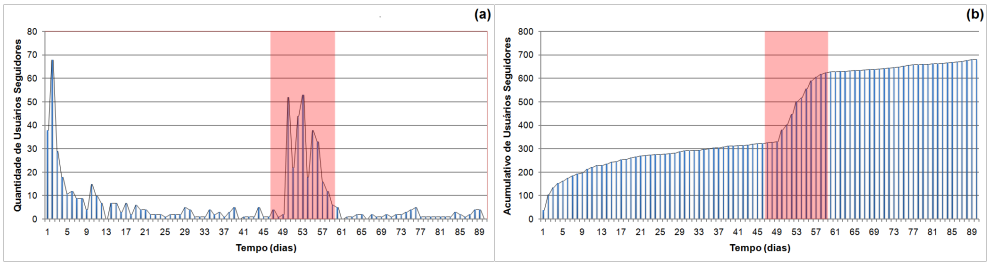

Figura 9. (a) Gráfico de seguidores do bot scarina (b) Gráfico do acumulativo de seguidores do bot scarina

No gráfico correspondente à Figura 10 também podemos ver, a partir do dia 46 que o *bot* recebe novas menções e mensagens. Como o processo inicial foi executado novamente, conquistando novos seguidores, a maioria das menções e mensagens foram de questionamentos sobre quem era o *bot* ***scarina*** e de onde os usuários se conheciam. Ao final de 90 dias de experimento o *bot* ***scarina*** obteve 52 menções e 45 mensagens.

Além dos resultados já apresentados, o *bot* ***scarina*** postou 4997 *tweets*, durante os 90 dias de experimento. Deste total, tivemos apenas 6 *tweets* entre os favoritos pelos usuários seguidores. Além disso, obteve ao fim dos 90 dias de experimento o total de 94 *retweets* e 109 *tweets* respondidos.

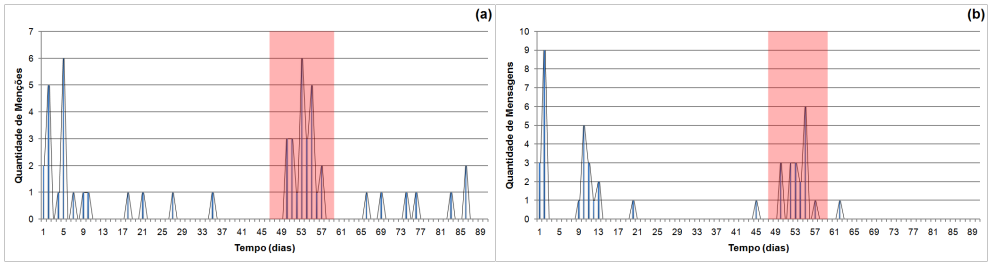

Figura 10. (a) Gráfico de *tweets* mencionados do bot scarina (b) Gráfico de mensagens do bot scarina

6.5. Resultados finais

No final dos 90 dias de experimento, coletamos os valores de *Klout* e *Twitalyzer* das contas mais influentes do *Twitter* e fizemos uma comparação final com os *bots* ***fepessoinha*** e ***scarina***, conforme é mostrado no gráfico correspondente à Figura 11. Entre os mais influentes, estão o humorista brasileiro Rafinha Bastos, o *rapper* americano Snoop Dogg, o presidente dos Estados Unidos Barack Obama e o apresentador brasileiro Luciano Huck. Podemos ver que mesmo de forma automatizada, os *bots* alcançaram níveis iguais ou acima, de outras pessoas com alta reputação, nos classificadores de influência *Klout* e *Twitalyzer*.

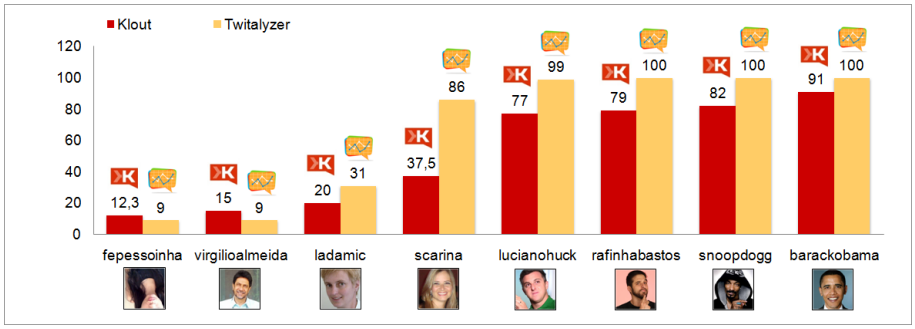

Figura 11. Comparação Klout e Twitalyzer - os mais influentes

7. Trabalhos Futuros

Com a apresentação e comparação de algumas estratégias utilizadas por *bots* na rede social *Twitter*, este trabalho servirá como base para o estudo e comparação de outras estratégias de influências nas redes sociais. Um incremento interessante seria utilizar o *bot* ***scarina*** com o dicionário de dados baseado nos *trends topics* do próprio *Twitter*, de forma dinâmica. Outra alternativa, seria utilizar estratégias de PLN (Processamento de Linguagem Natural) visando a geração e compreensão automática de línguas humanas naturais, no tratamento dos *tweets* postados pelos usuários. Com isso, o *bot* seria capaz de interagir, de forma automática postando conteúdos relacionados ao contexto em que for inserido. Enfim, avaliar mais detalhadamente as melhores estratégias que os *bots* podem usar para manipular *rankings* de influencias em redes sociais.

8. Conclusões

Neste artigo criamos *bots* simples capazes de interagir, através de contas no *Twitter*, como se fossem usuários na rede, trocando informações, seguindo e conquistando novos seguidores, durante o período de 90 dias. Constatamos que durante este período, os *bots* conseguiram resultados significantes em sistemas de classificadores de influência como *Klout* e *Twitalyzer*.

Artigos científicos e a mídia em geral têm usado esses sistemas, exageradamente, como padrão para avaliar a influência de um usuário nas redes sociais. Ao decorrer do trabalho, conseguimos constatar que suas métricas são vulneráveis e facilmente manipuláveis quando utilizamos contas automatizadas, conhecidas como *bots*. Com isso, mostramos que simples *bots* no *Twitter* podem conseguir altos níveis de influência, segundo as métricas do *Klout* e *Twitalyzer*, confirmando nossa hipótese. Além disso, verificamos de forma experimental e prática algumas estratégias para os *bots*. A estratégia de apenas seguir usuários e ser seguido por outros usuários não é o suficiente para um alto grau de influência. Como melhor estratégia, verificamos que o *bot* também deve interagir postando conteúdos relevantes na rede.

Finalmente, ao fim do experimento, para comprovar a eficácia dos nossos algoritmos, em um curto período, nossos *bots* conquistaram níveis próximos à níveis de personalidades e pessoas com alta reputação, que são consideradas as mais influentes no *Twitter*. Com isso, mostramos que processos automáticos com *bots* podem conquistar altos níveis de influência.

Agradecimentos

Parte desse trabalho foi financiado através de recursos do CNPq, FAPEMIG e CAPES.

Referências

[1] Klout. `http://www.klout.com`, March 2012.

[2] Twitalyzer. `http://www.twitalyzer.com/`, March 2012.

[3] Purohit H., Ruan Y., Joshi A., Parthasarathy S., and Sheth A. Understanding user-community engagement by multi-faceted features: A case study on twitter. In *Conference of Social Media Analytics Workshop at World Wide Web (WWW)*, 2011.

[4] Yan J. L. S. and Kaziunas E. What is a tweet worth?: measuring the value of social media for an academic institution. In *Proceedings of the 2012 iConference*, iConference '12, pages 565–566, New York, NY, USA, 2012. ACM.

[5] David Leonhardt. A better way to measure twitter influence. *The New York Times*, 2011.

[6] Cha M., Haddadi H., Benevenuto F., and Gummadi K. P. Measuring User Influence in Twitter: The Million Follower Fallacy. In *AAAI ICWSM*, May 2010.

[7] Romero D. M., Galuba W., Asur S., and Huberman B. A. Influence and passivity in social media. In *ACM WWW*, pages 113–114, 2011.

[8] Weng J., Ee-Peng Lim, Jiang J., and He Q. Twitterrank: finding topic-sensitive influential twitterers. In *ACM WSDM*, 2010.

[9] Pal A. and Counts S. Identifying topical authorities in microblogs. In *ACM WSDM*, pages 45–54, 2011.

[10] Ghosh S., Viswanath B., Kooti F., Sharma N. K., Gautam K., Benevenuto F., Ganguly N., and Gummadi K. Understanding and Combating Link Farming in the Twitter Social Network. In *Proceedings of the 21st International World Wide Web Conference (WWW'12)*, Lyon, France, April 2012.

[11] Boshmaf Y., Muslukhov I., Beznosov K., and Ripeanu M. The socialbot network: When bots socialize for fame and money. In *ACSAC'11 Proceedings of the 27th Annual Computer Security Applications Conference*, pages 93–102, New York, NY, USA, 2011.

[12] Messias J., Schmidt L., Oliveira R., and Benevenuto F. Sigam-me os bons! transformando robôs em pessoas influentes no twitter. *BraSNAM*, 2012.