



GUSTAVO MONTEIRO DA SILVA

**ESTUDO E DISCUSSÃO DA LITERATURA
SOBRE AS PRINCIPAIS TÉCNICAS PARA
DETECÇÃO DE INFLUÊNCIA DE
USUÁRIOS NO TWITTER: PERÍODO 2010 A
2013**

LAVRAS – MG

2013

GUSTAVO MONTEIRO DA SILVA

**ESTUDO E DISCUSSÃO DA LITERATURA SOBRE AS PRINCIPAIS
TÉCNICAS PARA DETECÇÃO DE INFLUÊNCIA DE USUÁRIOS NO
TWITTER: PERÍODO 2010 A 2013**

Monografia de Graduação apresentada ao
Departamento de Ciência da Computação para
obtenção do Título de Bacharel em Sistemas de
Informação

Orientador

Prof. DSc.. Marluce Rodrigues Pereira

LAVRAS – MG

2013

GUSTAVO MONTEIRO DA SILVA

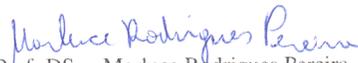
**ESTUDO E DISCUSSÃO DA LITERATURA SOBRE AS PRINCIPAIS
TÉCNICAS PARA DETECÇÃO DE INFLUÊNCIA DE USUÁRIOS NO
TWITTER: PERÍODO 2010 A 2013**

Monografia de Graduação apresentada ao
Departamento de Ciência da Computação para
obtenção do Título de Bacharel em Sistemas de
Informação

APROVADA em 21 de agosto de 2013.

Prof. DSc. Ahmed Ali Abdalla Esmin UFLA

Prof. DSc. André Luiz Zambalde UFLA


Prof. DSc.. Marluce Rodrigues Pereira
(Orientador)

LAVRAS – MG

2013

RESUMO

Twitter é uma rede social abrangente contando atualmente com 200 milhões de usuários ativos, e também uma rede social de postagem em tempo real, com facilidade de espalhar conteúdo para os demais usuários. A ideia de analisar os usuários em relação ao seu comportamento dentro do Twitter identificando o número de *retweets*, número de seguidores, menções e outros aspectos auxilia na detecção de influência dos usuários. As organizações podem utilizar esse conhecimento para aplicar a seus negócios, gerando vantagem competitiva. Por exemplo, o relacionamento com o cliente, as campanhas de marketing podem ser aperfeiçoados, e a obtenção de *feedback* dos usuários pode auxiliar no lançamento de novos produtos. O objetivo principal deste trabalho consiste em estudar os trabalhos existentes na literatura, identificando como é realizado o processo geral de descoberta de usuários influentes no Twitter, os principais algoritmos utilizados e os contextos de aplicação. Este estudo permitirá que novos trabalhos sejam desenvolvidos, aplicando os algoritmos já existentes sobre novas bases de dados, propondo novos algoritmos, otimizando os algoritmos já existentes ou propondo processos diferentes de detecção de usuários influentes.

Palavras-Chave: Influência, Usuários Influentes, Twitter.

ABSTRACT

Twitter is a social networking comprehensive currently with 200 million active users, and also a social network posting in real time, with ease of spreading content to other users. The idea of analyzing the users in relation to their behavior within the Twitter identifying the number of retweets, number of followers, mentions and other aspects helps detecting the influence of users. Organizations can use this knowledge to apply to their business, generating competitive advantage. For example, the customer relationship, marketing campaigns can be optimized, and getting feedback from users can help in launching new products. The main objective of this work is to study the works in the literature, as is done by identifying the overall process of discovering influential users on Twitter, the main algorithms used and the application contexts. This study will allow new works to be developed by applying the existing algorithms on new databases, proposing new algorithms, optimizing existing algorithms or proposing different processes of detecting influential users.

Keywords: Influence, Influential Users, Twitter.

SUMÁRIO

1	Introdução	9
1.1	Contextualização	9
1.2	Motivação	11
1.3	Objetivos	11
1.4	Estrutura do Trabalho	12
2	Twitter e Influência	13
2.1	Twitter	13
2.2	Influência	14
2.3	Influência aplicada em Outros Contextos	19
2.3.1	Sina Weibo	20
2.3.2	Klout	22
3	Revisão da Literatura	24
3.1	Buscas Efetuadas nas Bibliotecas Digitais	24
3.2	Seleção dos Artigos e Leitura dos Resumos	25
3.3	Formulário do Documento de Artigos	25
3.4	Elaboração do Documento de Artigos	26
3.5	Seleção dos Principais Artigos	26
4	Identificação de Usuários Influentes	28
4.1	Pré-Processamento	28
4.2	Extração de Características	29
4.3	Definição de Influência	30
5	Algoritmos de Influência	32
5.1	PageRank	32
5.1.1	Algoritmos que se basearam no PageRank	33

5.1.2	TwitterRank	35
5.1.3	SaID	36
5.2	HITS	38
5.3	H-index	39
6	Áreas de Aplicação	42
6.1	Contexto de Aplicação	42
7	Discussão	46
8	Conclusões	50
A	Apêndices	55

LISTA DE FIGURAS

2.1	Exemplo de um <i>Tweet</i>	14
2.2	Exemplo de <i>Tweet</i> de propaganda baseado nas relações de contato do usuário.	22
3.1	Formulário do Documento de Artigos.	26
4.1	Síntese das etapas de identificação de usuários influentes no Twitter, adaptada das etapas do SaID de (BIGONHA, 2012).	28
5.1	Cálculo simplificado do algoritmo PageRank desenvolvido por (PAGE <i>et al.</i> , 1998).	34
5.2	Quadro da técnica proposta do TwitterRank. Traduzido de (WENG <i>et al.</i> , 2010).	36
5.3	Visão Geral SaID. Traduzido de (BIGONHA, 2012).	37
5.4	Visão Geral do Processo apresentado por (RAO; NAGPAL, 2011).	39
5.5	Processo de descoberta de usuários influentes no contexto da reforma do Serviço Nacional de Saúde da Inglaterra.	40

LISTA DE TABELAS

5.1	Resumo dos principais algoritmos estudados e suas aplicações.	41
6.1	Resumo dos contextos de aplicação.	45
7.1	Resumo dos principais trabalhos de detecção de influência de usuários no Twitter.	47

1 INTRODUÇÃO

Neste capítulo serão apresentadas a contextualização, motivação, objetivos e a estrutura textual deste trabalho.

1.1 Contextualização

O conhecimento extraído de uma base de dados é importante para a tomada de decisão em uma organização. As opiniões emitidas por usuários são fonte de informações tanto para outros usuários quanto para as organizações. De acordo com (PANG; LEE, 2008), saber "o que outras pessoas pensam" consiste em uma fonte de informação, tanto que pedimos amigos para recomendar um determinado profissional, por exemplo, um candidato à determinada vaga de emprego. Como as informações são essenciais, atualmente os dados foram disponibilizados também pela *Web*, facilitando o acesso a estas informações e opiniões.

A *Web* é uma fonte de informações que vem crescendo com a adesão de várias organizações e pessoas. O crescimento das redes sociais e blogs, acessíveis também via dispositivos móveis auxiliou na popularização do uso da *Web*. Com isso, as redes sociais também aumentaram seu número de usuários, pois é uma plataforma que facilita a disponibilização de informações, ou seja, qualquer usuário pode gerar uma informação sobre determinado produto e/ou serviço, gerando assim a possibilidade de expressar suas opiniões sobre determinados produtos que consomem e sobre as organizações prestadoras de serviços. (ZABIN; JEFFERIES, 2008) citaram que as grandes organizações estão cada vez mais percebendo que essas "vozes dos clientes" podem exercer uma influência nas opiniões de outros consumidores, como em lealdade a marca, decisões de compra e defesa da marca.

Analisar as informações geradas na *Web* tem uma devida importância para as organizações, que de acordo com (LO; POTDAR, 2009) a coleta e classificação destas informações da *Web* são importantes para um fabricante do produto, uma

vez que fornece informações sobre eventuais defeitos de um novo produto, auxiliando a organização na gestão de riscos. Também afirmaram que os usuários podem influenciar significativamente nas decisões de clientes existentes e em possíveis clientes na aquisição de um novo produto ou contratação de um serviço inovador. (ZABIN; JEFFERIES, 2008) afirmaram que as organizações podem responder às percepções dos clientes que são geradas através do monitoramento e análise de mídias sociais, modificando as mensagens de *marketing*, posicionamento da marca, e desenvolvimento de produtos e outras atividades de acordo com as informações coletadas.

Este trabalho terá enfoque na ferramenta de *micro-blogging* Twitter¹ que é uma rede social com intuito de gerar informação em tempo real. Nesta rede, os usuários postam mensagens textuais que são denominadas *tweet* com tamanho máximo de 140 caracteres. Devido a este limite de tamanho das mensagens, de acordo com (BIGONHA *et al.*, 2011) isto torna a rede social simples e diversificada. Sendo assim, qualquer usuário pode criar conteúdo e consumir conteúdo. Em tempo real, um usuário que acabou de comprar um produto pode postar um *tweet*, definindo suas características, mostrando vantagens de sua aquisição ou mesmo problemas. Dessa forma, as organizações podem obter um *feedback* destes usuários como conhecimento válido e de relevância para um determinado produto.

(BAO *et al.*, 2013) citaram que o *micro-blogging* está se tornando uma plataforma de mídia social cada vez mais popular em relação ao número de usuários, que podem descobrir informações interessantes sobre qualquer contexto. As organizações podem se aproveitar desta característica para analisar os conteúdos produzidos pelos usuários, a fim de obter informações a respeito do perfil dos mesmos e da influência que é exercida por um determinado grupo de usuários influentes.

¹<https://twitter.com/>

1.2 Motivação

O Twitter apresenta um grande número de usuários ativos estimados em 200 milhões, e 400 milhões de *tweets* por dia - segundo o seu próprio blog² em Março de 2013. Com isso, este ambiente dinâmico que os usuários compartilham experiências e sugestões sobre diversos assuntos gera um grande volume de conteúdo. A detecção de usuários influentes no Twitter auxilia as organizações na identificação dos usuários que lideram as conversas e ações em um determinado contexto. Auxiliando as organizações a encontrar os usuários que lideram as conversas e ações em um determinado contexto. (VALIATI *et al.*, 2012) cita que as redes sociais, por terem se mostrado como um importante mecanismo para atingir um número alto de pessoas, podem influenciar na opinião pública, publicidade de novos produtos e adoção de inovações.

A identificação de usuários influentes também é uma forma de adquirir conhecimento para área de relacionamento com o cliente, pois as organizações poderão encontrar os usuários chave que espalham tanto opiniões positivas quanto opiniões negativas, podendo a organização melhorar seus produtos ou serviços diante desta informação obtida do Twitter. A identificação de usuários influentes pode ser aplicada a outros contextos, tais como fonte de informação em desastres, acidentes de grandes proporções, em manifestações fornecendo informações às autoridades, população em geral e órgãos governamentais responsáveis pela segurança.

1.3 Objetivos

O objetivo principal deste trabalho consiste em estudar os trabalhos existentes na literatura, identificando como é realizado o processo geral de descoberta de usuários influentes no Twitter, os principais algoritmos utilizados e os contextos de aplicação.

²<http://blog.twitter.com/2013/03/celebrating-twitter7.html>

Os objetivos específicos deste trabalho serão apresentados a seguir:

- Busca de trabalhos na área de detecção de influência no Twitter;
- Revisão da Literatura na área de detecção de influência no Twitter;
- Análise das principais características de comportamento dos usuários influentes no Twitter;
- Análise dos principais algoritmos de detecção de influência apresentados na literatura;
- Identificação dos principais contextos adotados nos trabalhos publicados na área de detecção de usuários influentes;
- Tabela com o resumo dos principais trabalhos estudados na área de detecção de influência no Twitter;
- Sugestão de novos trabalhos que podem ser desenvolvidos nesta área.

1.4 Estrutura do Trabalho

O restante deste texto está organizado da seguinte forma. O capítulo 2 descreve a definição de influência de uma forma geral e a influência aplicada no Twitter, as características do Twitter, e a influência aplicada em outros contextos. O capítulo 3 descreve a revisão da literatura que foi utilizada como base para a aplicação da revisão contida neste trabalho. O capítulo 4 mostra como foi efetuada a identificação de usuários influentes nos trabalhos estudados. O capítulo 5 demonstra os principais algoritmos utilizados nestas técnicas, com suas definições e metodologias envolvidas. O capítulo 6 mostra as áreas de aplicação das técnicas de detecção de influência. O capítulo 7 apresenta uma discussão sobre os estudos de técnicas para detecção de influência de usuários no Twitter. Finalmente, o capítulo 8 apresenta as conclusões deste trabalho.

2 TWITTER E INFLUÊNCIA

Nesta seção serão apresentadas as características do *micro-blogging* Twitter, definição de influência de uma forma geral e influência aplicada no Twitter, e a influência aplicada em outros contextos.

2.1 Twitter

A facilidade atual de disponibilização de conteúdo na *Web* ajudou as organizações a obterem mais uma forma de *feedback* dos seus clientes. Com isso, a ferramenta Twitter se destacou pela simplicidade de disponibilização de informações e pela sua grande quantidade de informações geradas (BIGONHA, 2012). Portanto, esta ferramenta consiste como fonte das informações para os estudos das técnicas de detecção de usuários influentes nesta rede social, que é o foco deste trabalho.

O Twitter foi lançado em 2006, e é uma rede social baseada em textos postados com limite de 140 caracteres denominados *tweets*. A Figura 2.1 apresenta o exemplo de um *tweet* postado pelo usuário Ricardo Amorim. A postagem destes *tweets* é realizada por pessoas comuns e demais organizações (BIGONHA, 2012). As suas postagens são exibidas em tempo real de um determinado usuário para os seus seguidores (usuários que assinaram um pedido para receber atualizações) com intuito de informá-los através de um texto definido como *tweet*. Estes *tweets* podem conter também links para outros sites, fotos e vídeos, e demais variações que contêm *hashtag*¹ que são palavras-chave antecedidas pelo símbolo "#" que designam o assunto que está sendo discutido naquele *tweet*.

Cada usuário do Twitter pode acessar sua página de perfil, que contém os seus dados pessoais e mostra em sua linha do tempo (*timeline*) os *tweets* dos usuários que ele está seguindo em uma ordem cronológica. O usuário deve em algum momento optar por aceitar receber os *tweets* desses outros usuários.

¹Disponível em: <http://pt.wikipedia.org/wiki/Hashtag>



Figura 2.1: Exemplo de um *Tweet*.

Um *retweet* ocorre quando um usuário republica um determinado *tweet* de outro usuário. Outro comportamento são as menções que acontecem quando um usuário menciona outro usuário no seu *tweet* usando como referência o nome do usuário na rede social precedido do símbolo "@" para começar uma conversa. Essas menções ocorrem de duas maneiras e com diferentes finalidades: "@Replies" é a resposta a um usuário, e "@Mentions" ocorre quando um usuário menciona outro usuário no meio de um *tweet*. Estas interações têm grande valor na análise de usuários influentes, pois demonstra que os usuários que têm muitas menções e *retweets* são usuários ativos, e podem ser considerados influentes dentro de uma rede social.

2.2 Influência

A definição de pessoas influentes foi apresentada no trabalho de (KATZ; LAZARUS-FELD; RESEARCH, 1955), que descreve os chamados "molecular leaders" (líderes moleculares) que são pessoas que influenciam em seus ambientes locais, mas não necessariamente têm destaque em uma comunidade maior.

Em (WENG *et al.*, 2010) definiu-se que um usuário pode ter uma influência maior, se a soma de influência dos seus seguidores é alta e ao mesmo tempo, a sua influência sobre cada seguidor é determinada pela quantidade de usuários que receberam seus *tweets*. Também é mostrado que a influência de um usuário do Twitter pode ser em diferentes temas.

Na detecção de usuários influentes no Twitter necessita-se analisar primeiramente o contexto, pois determinado usuário pode ser influente em diversas áreas. Na análise de (CHA *et al.*, 2010) mostrou-se que os usuários mais influentes podem influenciar significativamente sobre uma variedade de tópicos. Os influentes podem ser ou não usuários dos produtos, serviços ou marcas, porém estes usuários só se tornam influentes se publicarem suas opiniões. Porém, nem todos que publicam opiniões podem ser considerados influentes. Dependendo do alcance do seu *tweet* a sua postagem não fica visível a um número considerável de pessoas, tornando a sua postagem, *tweet* irrelevante no sentido de influenciar. Para analisarmos estas técnicas de influência no Twitter, foram estudados alguns trabalhos que já tenham resultados que comprovam a eficácia de detecção de influentes dentro de redes sociais .

Na literatura existem diferentes técnicas que mostram como deve ser realizado o estudo de detecção de influência no Twitter que são apresentadas a seguir.

No trabalho de (CHA *et al.*, 2010) foram utilizadas três atividades para representar os diferentes tipos de influência de um usuário:

1. *Indegree Influence* - mede a influência do usuário de acordo com o número de seguidores que o usuário tem, indicando o tamanho de público deste usuário.
2. *Retweet Influence* - mede sua influência através do número de *retweets*, que indica a capacidade do usuário gerar conteúdo de valor, que é republicado pelos outros usuários.
3. *Mention Influence* - mede sua influência através do número de menções do usuário no Twitter, indicando a capacidade do usuário de envolver outros usuários em uma conversa.

Porém, (CHA *et al.*, 2010) concluíram em sua análise que o usuário ter um grande número de seguidores - *Indegree Influence*, não comprova que este usuário

será um usuário considerado influente, ou seja, não necessariamente este usuário que tem um grande número de seguidores gera muitos *retweets* ou menções. Com isso, este usuário é considerado popular mas não está relacionado com outras noções importantes de influência, como o envolvimento do público. Menciona também que a influência não é adquirida de forma espontânea ou acidentalmente, mas através de esforço concentrado.

No estudo de propagação de informações dentro do Twitter, (ROMERO *et al.*, 2011) revelaram que a maioria dos usuários agem como consumidores passivos de informações e não transmitem o conteúdo na rede. Desta forma, um usuário para se tornar influente deve obter a atenção, ser popular e também superar a passividade dentro da rede social. É proposto um algoritmo que determina a influência e a passividade dos usuários do Twitter com base nas suas atividades de encaminhamentos de informação. O estudo realizado mostrou que a correlação entre a popularidade de um usuário e a sua influência é considerada fraca. Isto mostra que para uma informação se propagar dentro do Twitter, os usuários precisam transmitir a informação a outros usuários, o usuário deve envolver-se ativamente, ao invés de apenas ler o *tweet* e não agir, tornando-o mais influente com a sua participação. Assim, um usuário com alta popularidade não implica em um usuário com alta influência, e vice-versa.

A técnica de (BIGONHA, 2012) utiliza a Análise de Sentimentos e propõe um método para a identificação de usuários influentes em três perspectivas: rede, polaridade e qualidade de conteúdo. Portanto, para encontrar os usuários influentes, classifica-os quanto a sua polaridade definindo estes usuários também como usuários que falam bem da marca/produto/serviço chamados de evangelistas, ou usuários que falam mal da marca/produto/serviço chamados de detratores, e analisa também a qualidade do conteúdo do seu *tweet*. E por final fornece uma lista dos Tops - x Evangelistas Influentes e Tops - x Detratores influentes. Neste trabalho (BIGONHA, 2012) conclui que o Gráfico de Interação - (Relacionado a

interações do usuário através de *tweets* como menções, *retweets* e resposta) produziu melhores resultados e realizou-se com menor custo computacional, diante do Gráfico de Ligação - (Relacionado a conexões entre os usuários dentro do Twitter).

De acordo com (CHEN *et al.*, 2012) a influência nas redes sociais tem o problema de medir e quantificá-la, pois o conceito de influência é relativamente subjetivo e não tem uma definição universalmente reconhecida. Para diferentes redes sociais o significado de influência e as métricas para calcular os influentes são distintos. Algumas métricas abordam a interação do usuário, e o seu comportamento dentro da rede. Há uma probabilidade das pessoas poderem definir um usuário ativo como sendo um usuário influente, mas não necessariamente o usuário ativo é um usuário influente, pois a quantidade de conteúdo que é publicado pode não conseguir engajar demais usuários da rede.

(CHEN *et al.*, 2012) dividiram a definição de influência a partir de perspectiva relativa e perspectiva global :

1. Usuários de Influência Relativa: mostra a capacidade de um determinado usuário despertar a ação de outro usuário próximo. Por exemplo, "um usuário A envia um *tweet*, um dos seus seguidores lê este *tweet* e decide retuitar, ou comentar sobre ele, ou então iniciar uma conversa com o usuário A". Este conceito reflete a proximidade dos usuários e classifica estes usuários como amigos individuais. Esta métrica pode ser usada no campo de maximização de influência para medir a probabilidade de difusão de informação.
2. Usuários da Rede de Influência Global: nesta definição a autoridade e o prestígio de um usuário refletem em toda a rede social, ou seja, a influência do usuário desperta ações em vários outros usuários. O exemplo dado por (CHEN *et al.*, 2012) é, "um *tweet* é postado por um usuário A, e é retuitado por um usuário B, e alguns seguidores de B novamente retuítam o *tweet*. Se o processo é repetido constantemente entre todos os seguidores e os seguidores dos seguidores, a influência do usuário A é constantemente

ampliada até o término da difusão da informação". Com isso, a influência global é medida pela amplitude de disseminação da informação dentro de toda a rede social, e os influenciadores são pessoas que podem maximizar a disseminação destas informações.

Com estas definições (CHEN *et al.*, 2012) concluíram que estas ações dos usuários auxiliam na medição de influência. Portanto, existem outras métricas consideradas importantes pelo autor na definição de influentes, que são:

- Número de Seguidores - métrica que indica a popularidade do usuário e também define a cobertura inicial de influência. A popularidade de um usuário não é obtida acidentalmente, e *tweets* com maior audiência se espalham mais facilmente pela rede.
- Qualidade de Seguidores - a difusão de um *tweet* depende também de uma retransmissão secundária de qualidade, ou seja, de que o usuário que retuitou seja influente em relação aos seus seguidores.
- Qualidade de *Tweets* - o método preciso para identificar a qualidade de um *tweet* é a análise semântica, porém utiliza-se uma técnica indireta para refletir a qualidade, onde foram considerados números de *retweets* e comentários que determinado *tweet* gerou.
- Semelhança de Interesse - o modelo de similaridade de interesses do usuário proposto com base nas características do conjunto de dados fornecido. Além disso, a semelhança dos interesses dos usuários é importante para refletir comunidades pessoais, que são enfatizadas na maximização de influência dos usuários.

Na definição de influência alguns autores consideram características diferentes para classificar um usuário como influente dentro do Twitter. A principal diferença no trabalho do (RAO; NAGPAL, 2011) utiliza-se uma técnica que considera

o fato de localização e limitações de tempo do usuário, para classificar influentes para uma determinada área/cidade em tempo real. Os demais trabalhos (WENG *et al.*, 2010), (SONG; LI; ZHENG, 2012), (VALIATI *et al.*, 2012), (KING *et al.*, 2013) não consideraram estas características de localização e limitação de tempo dos usuários do Twitter para definição de influência.

Diante dos estudos mencionados nesta seção, pode-se definir a influência de um usuário com base em algumas características principais, como o envolvimento do usuário dentro da rede social. O envolvimento e engajamento do usuário podem ser avaliados de acordo com suas interações com outros usuários. As principais interações dos usuários que têm uma relevância significativa para determinar influentes são: menções, *retweets*, e conversas geradas a partir de um *tweet*. Outras características relevantes são quantidade de seguidores, qualidade textual dos *tweets*, qualidade de seguidores, qual interesse ou assuntos que o usuário participa, localização dos usuários. Os usuários para serem classificados como influentes dependem de diversas ações para manterem-se como influentes dentro de uma rede social.

2.3 Influência aplicada em Outros Contextos

Na literatura, são encontrados alguns trabalhos que realizam a detecção de influência em outras redes sociais. E também será mostrado uma ferramenta de medida de influência do usuário denominada KLOUT. (KLOUT, 2013) é uma ferramenta que mede a influência do usuário para rede sociais em geral.

O trabalho de (LIU *et al.*, 2012) analisou a influência para descobrir o comportamento do usuário em quatro gêneros diferentes de redes sociais: Twitter, Digg², Renren³ e Citation⁴. A sua técnica consiste em um modelo probabilístico

²<http://digg.com/> Site que reúne *links* para notícias.

³<http://www.renren.com/> Rede social chinesa similar ao Facebook.

⁴Rede de citação bibliográfica.

para a influência direta entre os nós, e diferentes tipos de métodos de propagação de influência para a influência indireta e influência global.

(CHEN *et al.*, 2012) em seu estudo calculou a influência dos usuários considerando a qualidade dos seus seguidores, e também outras métricas. Portanto, o contexto utilizado consiste em um conjunto de dados públicos da rede social Tencent Weibo⁵. (CHEN *et al.*, 2012) propõe o algoritmo InfluenceRank que utiliza da ideia do PageRank (BRIN; PAGE, 1998) estendida, onde fala que o PageRank calcula a soma da influência dos seguidores por iteração, mas ignora a qualidade dos seguidores. Com isso, a sua técnica considera a qualidade dos seguidores como uma das métricas para detectar influentes, ou seja, os seguidores também possuem uma influência em relação a difusão do *tweet*. Considerando que um *tweet* pode ser muito difundido dependendo da retransmissão secundária. (CHEN *et al.*, 2012) calcula a influência dos usuários da rede social Tencent Weibo considerando a qualidade dos seus seguidores, e outras métricas.

Outra rede social bastante estudada na área de detecção de usuários influentes é a Sina Weibo⁶, que será apresentada na próxima subseção com maiores detalhes. Há uma necessidade de destacá-la, pois se utilizou como base de dados para dois trabalhos (LIANG; LU; XU, 2012), (BAO *et al.*, 2013) de detecção de usuários influentes.

2.3.1 Sina Weibo

Sina Weibo é uma rede social similar a rede social Twitter lançada pela SINA Corporation⁷ que é uma empresa de mídia on-line da China em agosto de 2009, com aproximadamente 503 milhões⁸ de usuários registrados no final de 2012. Após alguns tumultos em 2009 no país houve o bloqueio de diversos serviços de *micro-*

⁵<http://t.qq.com/> Rede social similar ao Twitter.

⁶<http://weibo.com/> Rede social similar ao Twitter.

⁷ <http://www.sina.com/>

⁸Disponível em: <http://thenextweb.com/asia/2013/02/21/chinas-sina-weibo-grew-73-in-2012-passing-500-million-registered-accounts/>

blogging como Twitter e Facebook. No entanto, neste momento a empresa lançou o seu serviço similar de *micro-blogging* Sina Weibo, com as mesmas funções básicas de enviar mensagem de 140 caracteres, mensagem privada, comentários, repostagem de mensagem semelhante ao *retweet* do Twitter, uso de *hashtags* e dentre outras características de seguir pessoas e linha do tempo semelhante ao Twitter. De acordo com um relatório de março de 2011 Sina Weibo tinha 56,5% do mercado de *micro-blogging* da China.

Com a importância relativa desta rede social alguns trabalhos (LIANG; LU; XU, 2012), (BAO *et al.*, 2013), se basearam nos dados disponíveis pelo Sina Weibo para calcular a influência dos usuários desta rede social.

(LIANG; LU; XU, 2012) baseou-se no algoritmo PageRank (BRIN; PAGE, 1998) para análise de influência dos usuários do Sina Weibo. Para classificar um usuário com maior influência sobre outros usuários, os autores criaram um valor de PageRank para os usuários chamado "User PR". Este valor é calculado conforme o PageRank, se o usuário A seguir o usuário B, mostra que o usuário A considera o usuário B valioso. Então o usuário B recebe um voto de A. Com o maior número de votos que usuário B recebe, mais usuários seguem o usuário B, e quanto maior o número de seguidores mais usuários podem ver as postagens do usuário B, então, o usuário B terá maior influência sobre os demais usuários. Além disso, analisou-se o número de fãs e juntamente com a sua classificação do *ranking* de "User PR" caracterizou a influência dos usuários do Sina Weibo.

Outro trabalho que também utilizou como base de dados o Sina Weibo, consiste no trabalho de (BAO *et al.*, 2013), que propôs um novo modelo para prever os interesses dos usuários de *micro-blogging* por meio da integração dos usuários seguidor/seguinte e o histórico de interesse dos usuários. Com isso, auxiliar sistemas de *micro-blogging* a fornecerem informações mais personalizadas e serviços de publicidade aos usuários. (JAMALI; ESTER, 2010) cita uma aplicação deste modelo que assume uma rede social entre os usuários, e faz recomendações para um

usuário com base na classificação das relações sociais diretas ou indiretas que ele tem. De acordo com (JAMALI; ESTER, 2010), as empresas de redes sociais utilizam da abordagem baseada nas relações de contato dos usuários para construir sistemas de recomendação para seus usuários. Um exemplo da rede social Twitter será mostrado na Figura 2.2 que tem um *tweet* com conteúdo de propaganda de acordo com o que o usuário segue, oferecendo um produto diretamente para este usuário de acordo com os seus contatos.



Figura 2.2: Exemplo de *Tweet* de propaganda baseado nas relações de contato do usuário.

2.3.2 Klout

Klout (KLOUT, 2013) é uma ferramenta que mede a influência social online do usuário, utilizando uma "Pontuação Klout" que é um valor numérico de 1 a 100, sendo que quanto maior este número, mais influência social o usuário tem. A metodologia utilizada por Klout citada por (BIGONHA, 2012) é baseada em três conceitos: *True Reach* - que é o tamanho da audiência verdadeira de um usuário analisado a partir da contagem de seguidores/seguintes, total de *retweets*, menções, listas que o usuários estão incluídos, quantos spam/contas paradas estão seguindo usuário; *Amplification Probability* - probabilidade que o conteúdo vai se espalhar analisado a partir da interação entre os usuários e seus seguidores, a relação entre o número de *retweets* e o números de seguidores; e *Network Influence* que é a análise da influência considerando a interação entre os seguidores. A partir destes dados retirados da conta do Twitter do usuário, pode-se conectar mais contas

como Facebook⁹, Google+¹⁰, LinkedIn¹¹, Instagram¹² e do próprio Klout (KLOUT, 2013), para calcular a influência social do usuário. O usuário analisado pode ser um usuário comum ou também uma empresa que tem perfis nas redes sociais.

⁹<https://www.facebook.com/> Um site e serviço de rede social.

¹⁰<https://plus.google.com> Rede social desenvolvida pela Google.

¹¹<http://www.linkedin.com/> Rede de relacionamentos profissionais.

¹²<http://instagram.com/> Aplicativo móvel que permite os usuários tirar fotos e compartilhar.

3 REVISÃO DA LITERATURA

Neste capítulo é apresentado como foi realizada a revisão da literatura sobre detecção de usuários influentes no Twitter.

De acordo com (GIL, 2002) revisão da literatura é a contextualização teórica do problema e o relacionamento com o que tem sido investigado a respeito. Em uma revisão da literatura, deve ser esclarecido pressupostos teóricos que dão fundamentação ao estudo, e não pode ser constituída apenas por referências ou sínteses dos estudos feitos. Portanto, a revisão elaborada neste trabalho consiste em uma revisão da literatura dos principais estudos na área, e também abrange as principais características dos estudos, os principais algoritmos utilizados, os contextos de aplicação mostrando como está o estado atual dos estudos de detecção de usuários influentes no Twitter.

Para realizar a revisão da literatura, foram realizadas buscas em bibliotecas digitais, seleção dos artigos e leitura dos resumos, criação de um formulário do Documento de Artigos, elaboração do Documento de Artigos e seleção dos principais artigos para realização desta revisão. Os detalhes de como foi realizada a revisão são apresentados a seguir.

3.1 Buscas Efetuadas nas Bibliotecas Digitais

A revisão realizada por este trabalho consistiu em uma pesquisa nas bibliotecas digitais Scopus¹, IEEEExplorer², ACM Digital Library³ e ScienceDirect⁴ utilizando as palavras-chaves ("influence twitter"). Para realização destas pesquisas foi considerado as restrições de busca apenas em título, palavras-chave e resumo e também do ano de publicação dos artigos pelo período de 2010 à 2013. Foram encontrados

¹<http://www.scopus.com/home.url> Base de dados bibliográfica.

²<http://ieeexplore.ieee.org/Xplore/home.jsp> Banco de dados de pesquisa acadêmica.

³<http://dl.acm.org/> Base de dados bibliográfica.

⁴<http://www.sciencedirect.com/> Base de dados bibliográfica.

3.668 artigos. Desse total, 290 foram obtidos no Scopus, 81 no IEEEExplorer, 1.629 no ACM Digital Library e 1.668 no ScienceDirect, sendo realizada a leitura dos títulos de parte destes resultados, onde os principais artigos foram organizados por relevância dentro de cada biblioteca digital.

3.2 Seleção dos Artigos e Leitura dos Resumos

A partir das buscas realizadas nas bibliotecas digitais foram selecionados 33 artigos para leitura dos resumos, e busca de palavras-chave como "influence" e "twitter" dentro de cada um destes artigos com o objetivo de selecionar os principais trabalhos que envolvessem o tema de detecção de usuários influentes no Twitter, e para auxiliar nesta seleção foi desenvolvido um formulário.

3.3 Formulário do Documento de Artigos

O Formulário do Documento de Artigos foi desenvolvido para auxiliar na seleção dos principais artigos da área de detecção de usuário influentes no Twitter e é mostrado na Figura 3.1. A estrutura deste Formulário do Documento de Artigos contém os seguintes itens a serem preenchidos:

- Autor(es), Título, Publicação/Ano, Número de Citações;
- Abstract, Resumo do artigo elaborado em um parágrafo;
- Responder as questões "O artigo faz alguma classificação para ranquear os influentes ?", "Faz algum tipo de recomendação ?", "Fala sobre algoritmos utilizados ?", e "Foi aplicado a algum contexto específico (saúde, tráfego, política, empresas, etc) ?".

Documento de Artigos

Autor(es):

Título:

Publicação/Ano:

Número de Citações:

Abstract

Resumo (em um parágrafo contendo a principal contribuição)

Verificar se o artigo faz alguma classificação para ranquear os influentes?

Faz algum tipo de recomendação?

Fala sobre algoritmos utilizados?

Foi aplicado a algum contexto específico (saúde, tráfego, política, empresas, etc)?

Figura 3.1: Formulário do Documento de Artigos.

3.4 Elaboração do Documento de Artigos

Este Documento de Artigos foi elaborado para todos os 33 artigos, e pode ser visualizado em detalhes uma página deste documento no Apêndice A. E a partir da elaboração deste documento foram feitas algumas discussões para então selecionar os principais artigos na área de detecção de influência de usuários no Twitter. Com isso, a busca por trabalhos relacionados foram estendidas à partir destes 33 artigos estudados.

3.5 Seleção dos Principais Artigos

Foram encontrado seis artigos à partir das buscas em bibliotecas digitais e elaboração do Documento de Artigos que apresentavam um maior número de citações e abrangiam melhor o tema pesquisado. E finalmente, foram selecionados outros dois trabalhos que não estavam disponíveis nas bibliotecas consultadas que são: (VALIATI *et al.*, 2012) e (BIGONHA, 2012), mas são estudos focados na área de de-

teção de usuários influentes no Twitter. Portanto, no total foram selecionados 8 artigos relacionados as principais técnicas de detecção de influência de usuários no Twitter no período de publicação de 2010 até 2013.

A revisão da literatura efetuada neste trabalho teve características específicas para determinar os principais artigos na área de detecção de usuários influentes no Twitter. Pois a seleção dos principais artigos foi efetuada a partir de algumas bibliotecas digitais e de estudos selecionados previamente.

4 IDENTIFICAÇÃO DE USUÁRIOS INFLUENTES

A identificação de usuários influentes no Twitter possui diversas características particulares e cada trabalho estudado define suas etapas de uma maneira diferente. Portanto, foi desenvolvida uma síntese das etapas estudadas com base nas etapas do algoritmo SaID do trabalho de (BIGONHA, 2012). Na figura 4.1 são apresentadas as etapas desta síntese dos trabalhos de descoberta de usuários influentes no Twitter. As etapas são: Pré-Processamento, Extração de Características e Definição de Influência.

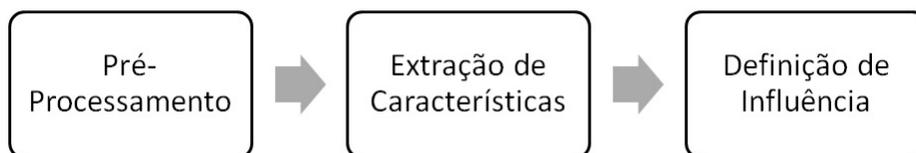


Figura 4.1: Síntese das etapas de identificação de usuários influentes no Twitter, adaptada das etapas do SaID de (BIGONHA, 2012).

Cada uma destas etapas serão descritas nas próximas seções mostrando como foi efetuada a detecção de usuários influentes no Twitter pelos trabalhos estudados.

4.1 Pré-Processamento

Nesta etapa é selecionado o tema sobre o qual se quer identificar os usuários influentes e um período de tempo para coletar os dados no Twitter. Para realizar esta coleta de dados geralmente usa-se a API¹ do próprio Twitter. As definições de contexto dos trabalhos estudados, podem ocorrer de diversas maneiras. O trabalho de (WENG *et al.*, 2010) definiu os tópicos por uma técnica de aprendizagem supervisionada utilizando Processamento de Linguagem Natural para identificar os tópicos

¹Disponível em: <https://dev.twitter.com/>

relevantes que os usuários comentam no Twitter, como por exemplo: Tecnologia da Informação (TI), estilo de vida, *gadgets*, dentre outros. Com isso, definiu os usuários influentes nestes contextos. (CHA *et al.*, 2010) aplicou aos temas populares de 2009 coletados a partir dos Trending Topics de 2009 do Twitter. (ROMERO *et al.*, 2011) analisou *tweets* que continha URL's no seu conteúdo, buscando pela palavra "http". (RAO; NAGPAL, 2011) determinou a influência coletando dados dos usuários de acordo com a localização geográfica em que o usuário se encontrava. (SONG; LI; ZHENG, 2012) coletou *tweets* e propôs um método de encontrar temas populares no Twitter com base nos 5 tópicos de cada dia em um período de 26 dias. (VALIATI *et al.*, 2012) definiu os tópicos mais populares de forma manual dentro do Twitter optando por coletar somente *tweets* no contexto de automóveis e política. (BIGONHA, 2012) analisou o que os usuários do Twitter estavam comentando sobre uma marca de refrigerante, marca de eletrodomésticos e hipermercado, e além disso, coletou as relações de seguidor/seguinte. Finalmente, (KING *et al.*, 2013) aplicou o seu estudo no contexto de reformas do Serviço Nacional de Saúde da Inglaterra dentro do Twitter, coletando os *tweets* que continham os termos "NHS" e "reform(s)" e a hashtag "#NHSreform".

Esta etapa consiste também da filtragem e armazenamento de *tweets* que são de interesse para a identificação dos usuários influentes. No trabalho (BIGONHA, 2012) o pré-processamento consiste em quatro etapas. A primeira é a determinação do tema e o intervalo de tempo do conteúdo a ser analisado, a segunda etapa é a coleta dos *tweets*, a terceira é a filtragem e armazenamento desses *tweets* e a quarta etapa é a extração de dados dos usuários como nome, URL do perfil, lista de seguidores e usuários que ele segue.

4.2 Extração de Características

Nesta etapa ocorre o tratamento da base de dados, para auxílio nos processos seguintes de detecção de influência de usuários no Twitter. Como exemplo, o tra-

balho de (BIGONHA, 2012) definiu esta etapa como extração de características de conteúdo das mensagens do Twitter. E é realizado através da classificação dos *tweets* quanto a polaridade: positivo, negativo e neutro. Somente após essa classificação é que são aplicados algoritmos para detectar influência. Por exemplo, um determinado *tweet* contendo opinião positiva "Gostei muito deste novo produto/serviço X ", um exemplo com opinião negativa "O produto/serviço X não atendeu as expectativas propostas." e um exemplo com opinião neutra "Estou utilizando o produto/serviço X". As análises e classificações destes *tweets* são realizadas para posteriormente efetuar processamentos com softwares específicos em extrair conhecimento desta base de dados de mensagens.

Nesta etapa ocorre a aplicação de diferentes algoritmos e diferentes técnicas de detecção. O algoritmo PageRank (BRIN; PAGE, 1998) consiste no principal algoritmo de base para os trabalhos de (WENG *et al.*, 2010), (SONG; LI; ZHENG, 2012), (VALIATI *et al.*, 2012) e (BIGONHA, 2012). Outro algoritmo denominado HITS (KLEINBERG, 1999) serviu como base para o trabalho de (RAO; NAGPAL, 2011). E também temos o algoritmo H-index que serviu como base para (KING *et al.*, 2013). Os outros trabalhos utilizaram algoritmos desenvolvidos pelos próprios autores como (CHA *et al.*, 2010) que desenvolveu seu algoritmo que mede a influência dos usuários pelo número de seguidores, *retweets* e menções. E (ROMERO *et al.*, 2011) que desenvolveu seu algoritmo para determinar a influência e a passividade de um usuário à partir das atividades efetuadas por estes usuários no Twitter.

4.3 Definição de Influência

Nesta etapa ocorre a definição de influência dos usuários e aplicação deste conhecimento. O uso deste conhecimento pode beneficiar diversas organizações, no aperfeiçoamento do *marketing*, obtenção de *feedback* para auxiliar o lançamento de novos produtos e monitoramento da marca ou produto dentro do Twitter. Além

disso, pode-se ter diversas aplicações deste conhecimento como fonte de informações em grandes movimentos sociais como manifestações e em acidentes de grandes proporções.

A identificação dos usuários influentes pode ser efetuada aplicando diversas técnicas, e utilizando quantas etapas forem necessárias. Sendo assim, a elaboração de um trabalho nesta área não precisa necessariamente seguir todas etapas, podendo haver ainda mais etapas ou então redução de etapas.

5 ALGORITMOS DE INFLUÊNCIA

Neste capítulo são apresentados os algoritmos que medem a influência de um usuário em redes sociais. Com as suas definições e metodologias envolvidas é possível auxiliar o leitor no desenvolvimento dos trabalhos que buscam de alguma forma mostrar os usuários influentes no Twitter.

O trabalho de (CHA *et al.*, 2010) utilizou um algoritmo que mediu a influência pelo número seguidores, *retweets* e menções. Seu principal foco é investigar a dinâmica de influência dos usuários em tópicos no contexto de temas populares no Twitter: eleição iraniana, surto de gripe H1N1 e morte de Michael Jackson.

O trabalho de (ROMERO *et al.*, 2011) analisa a influência e a passividade dos usuários do Twitter. O algoritmo utilizado determina a influência e a passividade com base nas atividades dos usuários. E, o seu trabalho coletou através da API do Twitter os *tweets* que continham uma URL para realização do seu estudo onde a pesquisa destacou o valor dos usuários altamente conectados como elementos-chave na propagação de informações através do Twitter.

Nas próximas subseções temos os algoritmos de maior destaque na área de detecção de influência de usuários no Twitter. Os algoritmos são PageRank, TwitterRank, SaID, HITS e H-index, e serão apresentados com maiores detalhes a seguir.

5.1 PageRank

O PageRank (BRIN; PAGE, 1998) é um algoritmo citado por alguns dos trabalhos que medem a influência de um usuário, tais como: (WENG *et al.*, 2010), (SONG; LI; ZHENG, 2012), (FRAHM; SHEPELYANSKY, 2012), (CHEN *et al.*, 2012), (VALIATI *et al.*, 2012). O propósito principal do PageRank (BRIN; PAGE, 1998) é a classificação

de páginas nas consultas efetuadas dentro do buscador Google¹, porém utilizou-se como base para construção de outros algoritmos que medem a influência dos usuários em uma rede social.

Portanto, uma breve descrição de como é realizada esta classificação de páginas será explicada abaixo, para auxiliar no entendimento do porque utilizá-lo como base para outros algoritmos e técnicas de detecção de usuário influentes no Twitter.

A classificação das páginas efetuadas pelo algoritmo PageRank (BRIN; PAGE, 1998) atribui um valor de importância as páginas da Web, usando sua vasta estrutura de *links* como um indicador do valor de uma página individual. Na definição de (PAGE *et al.*, 1998) o algoritmo PageRank fornece um método mais sofisticado para fazer o ranking de páginas através de *links*. A descrição intuitiva de PageRank mostra que uma página tem alto valor, se a soma das fileiras de seus *backlinks*² é alta. Isso abrange tanto no caso em que uma página têm muitos *backlinks* ou quando uma página tem poucos *backlinks* porém altamente classificados. Na Figura 5.1 demonstra a propagação de classificação do PageRank de um par de páginas para outra página, mostrando os valores de PageRank recebidos por cada uma das páginas.

5.1.1 Algoritmos que se basearam no PageRank

O PageRank que é um algoritmo que classifica páginas da *web* com base na pontuação que uma página recebe através de *links*. (SONG; LI; ZHENG, 2012) propôs um modelo semelhante ao PageRank para detecção dos temas mais populares em *micro-blogging* com base na relação de interesse dos usuários. Portanto, a conclusão que os autores obtiveram consiste que a influência de ambos os tópicos e usuários podem ser calculados através da criação de uma rede de interesse entre

¹ <https://www.google.com.br/> Motor de busca da Google onde é possível efetuar pesquisas na internet.

²<http://en.wikipedia.org/wiki/Backlink> Qualquer ligação recebida por uma página web.

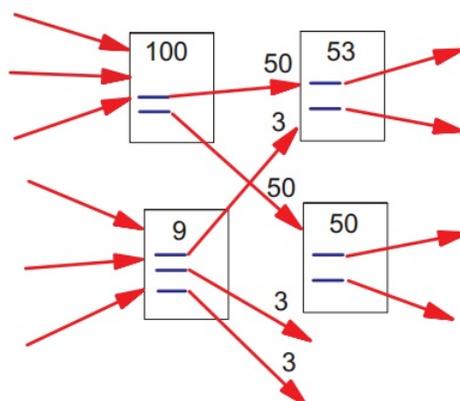


Figura 5.1: Cálculo simplificado do algoritmo PageRank desenvolvido por (PAGE *et al.*, 1998).

os usuários e os tópicos. E a popularidade dos tópicos, reflete no número e na influência dos usuários que estão interessados neles. Com isso, os usuários e os tópicos influentes são os que têm grande número de interessados sobre eles.

Outro trabalho que utilizou o modelo do PageRank, consiste no trabalho de (VALIATI *et al.*, 2012) onde foi elaborada uma técnica para ordenar usuários e conteúdo de acordo com a sua influência e relevância. A definição da técnica foi modelada como uma extensão do problema de PageRank e implementada eficientemente, uma vez que não requer a obtenção da rede de seguidores e determina as ordenações através de uma estratégia randomizada. De acordo com (VALIATI *et al.*, 2012) esta técnica baseou-se na definição intuitiva e circular de relevância e influência, ou seja, os usuários influentes tendem a disseminar conteúdo relevante e conteúdo relevante é em geral disseminado por usuários influentes. Nesta técnica avaliou-se o contexto de dois cenários amplamente discutidos na rede social Twitter que são: política e automóveis. E, este estudo não concentrou em coletar um grupo restrito e pequeno de conteúdos, mas todo o tráfego disseminando no Twitter.

Nas próximas subseções serão apresentadas mais dois trabalhos que tomaram como base o algoritmo PageRank para mostrar os usuários influentes no Twitter, porém serão apresentados com maior número de detalhes pelo destaque no número de citações que obtém, e por ser um trabalho de mestrado bem elaborado, respectivamente.

5.1.2 TwitterRank

O estudo de (WENG *et al.*, 2010) mostrou que existe uma relação entre os usuários no contexto do Twitter, pois observou-se dois comportamentos dos usuários dentro desta rede social. Primeiro, o usuário tem comportamento de seguir outros usuários de forma aleatória, e aqueles que estão sendo seguidos, seguem de volta apenas por cortesia. Outro comportamento observado que caracteriza a "homofilia de rede social" no Twitter é quando um usuário segue o outro por interesse no que aquele determinado usuário posta, e a relação é recíproca onde o outro usuário também segue de volta supondo que eles compartilham de um tema semelhante, mostrando que a relação de seguidor-seguinte é um forte indicador de similaridade entre os usuários. De acordo com (WENG *et al.*, 2010) alguns métodos medem a influência dos usuários do Twitter pelo número de seguidores. Este método é utilizado pelo Twitter e demais serviços. A diferença da técnica TwitterRank para a PageRank mesmo atuando em contextos diferentes é que o TwitterRank identifica usuários influentes em determinados tópicos, e o PageRank identifica páginas importantes. Podemos considerar, que a técnica do PageRank aborda somente a estrutura de *links* e ignora os interesses dos usuários.

A técnica proposta do TwitterRank é mostrada na figura 5.2, e descreve como efetuar o estudo dos usuários influentes neste trabalho. Primeiro, temos o Tema Específico Separado - onde ocorre a separação automática dos temas que os usuários do Twitter estão interessados através da análise do conteúdo dos seus *tweets*. Segundo, Redes de Relacionamentos Construídas em Tema Específico -

baseado nos temas separados, redes de relacionamentos com um tema específico entre os usuários são construídas. Finalmente, o *Ranking* de Influência do Usuário em Tema Específico - que com a aplicação do algoritmo TwitterRank, mede-se a influência dos usuários levando em conta a semelhança de tópico e também a estrutura de *link*.

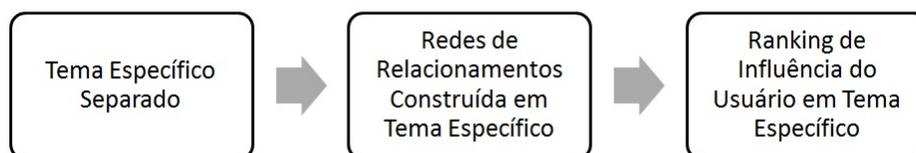


Figura 5.2: Quadro da técnica proposta do TwitterRank. Traduzido de (WENG *et al.*, 2010).

Portanto, o TwitterRank estuda as relações entre os usuários do Twitter e os seus comportamentos com intenção de trazer uma ordem de usuários no Twitter e quantificar a influência dos usuários. No TwitterRank foram analisados os temas que os usuários comentam, a influência que um determinado desperta em outros usuários a comentar do mesmo tema, a quantidade de *tweets* relacionados ao tema específico, se o usuário participa de mais de um tema e também a estrutura de *link*.

5.1.3 SaID

No trabalho de (BIGONHA, 2012) propôs-se um método chamado SaID - *Sentiment-based Influence Detection on Twitter* (Detecção de Influência Baseado em Sentimento no Twitter) para identificação de usuários influentes no Twitter. Na figura 5.3 podemos visualizar a visão geral deste método. O método consiste na divisão em três fases principais: Pré-processamento, Extração de Características e Pontuação de Influência. A primeira fase do SaID consiste nas definições do tópico e de consulta, rastreamento/coleta dos *tweets*, filtragem e armazenamento dos *tweets*, identificação dos usuários e extração de dados dos usuários. A segunda fase corresponde ao tratamento das métricas utilizadas onde analisa o conteúdo dos *tweets*,

as interações e conexões entre os usuários e analisa o sentimento dos *tweets*, onde uma das métricas baseia-se no PageRank. Finalmente, a terceira fase combina essas métricas em uma pontuação única de influência.

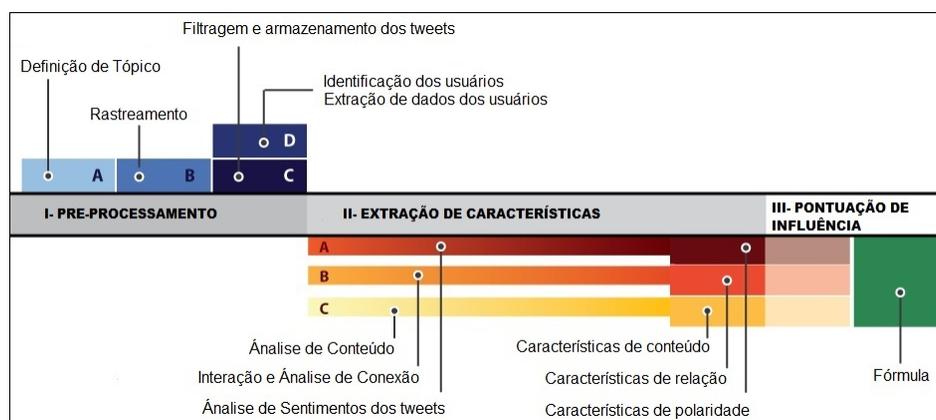


Figura 5.3: Visão Geral SaID. Traduzido de (BIGONHA, 2012).

De acordo com (BIGONHA, 2012) o desenvolvimento do método SaID auxilia os analistas de redes sociais no monitoramento de marcas/produtos, onde a sua principal função é fornecer uma lista de TOP-X evangelistas³ e TOP-X detratores⁴. O valor de X pode ser um intervalo de $10 \leq X \leq 150$, pois este intervalo escolhe-se com base no tamanho de uma lista aceitável para analisar possíveis usuários influentes. O método SaID concentra-se no comportamento dos usuários do Twitter classificando-os como influente ou não. Para efetuar esta classificação extrai as características dos usuários, tais como a polaridade e legibilidade dos seus *tweets* e a centralidade de interações dos usuários via *tweets* e as conexões destes usuários dentro do Twitter. Os resultados obtidos por SaID foram comparados com uma classificação feita por uma equipe de analistas de *marketing*, e por final conclui que os resultados das abordagens manuais e automáticas são muito semelhantes entre si, mas para alguns resultados a abordagem manual se sobressaiu melhor.

³Evangelistas - Usuários que postam principalmente conteúdo positivo sobre marca/produto

⁴Detratores - Usuários que postam principalmente conteúdo negativo sobre marca/produto

5.2 HITS

O algoritmo proposto por (KLEINBERG, 1999) utiliza o conceito de Autoridades e *Hubs*. *Hubs* são páginas com diversos diretórios para outras páginas da web, mas serviam para direcionar os usuários a páginas importantes. Páginas importantes que são autoridades são páginas apontadas por diversos *hubs*. Com isso, são atribuídas para cada página duas pontuações onde, uma pontuação é calculada para a sua autoridade que estima o valor do conteúdo da página, e a outra calcula o valor de *hub* que estima o valor de suas ligações para outras páginas. O autor propôs no seu algoritmo a técnica para a construção de subgrafos que são projetados para produzirem pequenas coleções de páginas que possam conter as páginas com maior autoridade para um determinado tópico.

O trabalho de (RAO; NAGPAL, 2011) baseou-se no algoritmo HITS, onde a analogia usada propõe o conceito de pesagem de um *retweet*, com valores entre 0 e 1, tomando como base o tempo e os fatores geográficos dos usuários do Twitter. O algoritmo proposto encontra a influência de um usuário ponderando os *retweets* que representam a influência relativa do usuário no Twitter. Além disso, é calculada a influência de um usuário considerando a localização e as limitações de tempo, onde a proposta é descobrir como a informação é transmitida rapidamente por um determinado usuário e em uma determinada área. Com isso, classifica a influência dos usuários com base em quão rápido seus *tweets* são respondidos em um determinado lugar específico. Um exemplo citado pelos autores deste modelo é quando em uma campanha publicitária de uma cidade, não será de grande importância um usuário obter *retweets* de usuário de uma outra cidade. Na Figura 5.4 podemos visualizar o processo realizado para encontrar usuários influentes, onde ocorre a contagem e pesagem dos *retweets*, a criação de um gráfico que representa a influência relativa de um usuário na rede, e finalmente o cálculo da influência deste usuário com base no tempo e na sua localização geográfica do usuário.

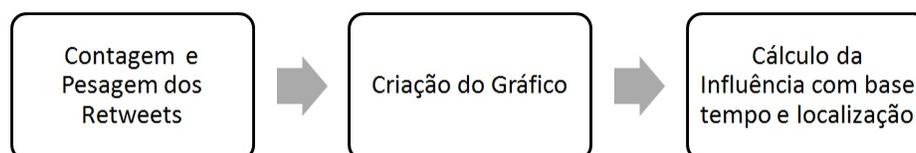


Figura 5.4: Visão Geral do Processo apresentado por (RAO; NAGPAL, 2011).

5.3 H-index

O indicador H-index proposto por (HIRSCH, 2007) mede o impacto de um pesquisador acadêmico com base no seu número de publicações e o no número de citações que cada uma das publicações gera. Por exemplo, um H-index de valor cinco significa que um pesquisador publicou cinco artigos e que cada um destes artigos tem pelo menos cinco citações.

Com base no H-index (KING *et al.*, 2013) propuseram o T-index como uma unidade de medida de influência de um usuário no Twitter em um determinado contexto. Sendo assim, um T-index de valor T é o número de *tweets* em um determinado contexto analisado que tem, pelo menos, T *retweets*. O T-index mostra que o *tweet* de um usuário em um determinado contexto tem influência, não significa que as suas opiniões em outras áreas são igualmente influentes. A influência específica de um usuário proposta pelo T-index pode estar muito relacionada com a popularidade (número de seguidores) deste usuário no Twitter, pois o *tweet* com maior influência no contexto estudado é de um usuário que tem mais de 6 milhões de seguidores no Twitter. Na Figura 5.5 podemos visualizar o processo realizado para encontrar usuários influentes no debate sobre a reforma do Serviço Nacional de Saúde (NHS) da Inglaterra, onde ocorreu a coleta dos *tweets* que continham as palavras-chaves 'NHS' e 'reform(s)', foi examinado o sentimento dos *tweets* quanto a reforma por dois revisores, e por final o cálculo de influência é calculado pelo número de *retweets* mostrando o valor de T-index dos usuários influentes.

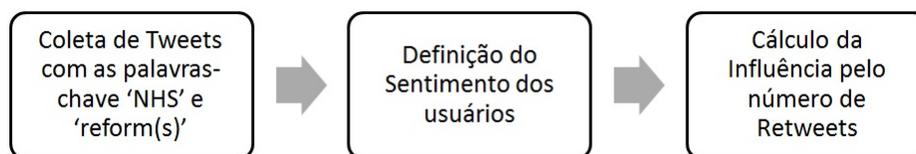


Figura 5.5: Processo de descoberta de usuários influentes no contexto da reforma do Serviço Nacional de Saúde da Inglaterra.

Os principais algoritmos de estudo das técnicas de detecção de influência de usuários no Twitter apresentados podem servir como base para demais estudos. Utilizando estes mesmos algoritmos em novas bases de dados, otimizando os algoritmos existentes, desenvolvendo novos algoritmos e analisando outras características dos usuários, permitindo que novos estudos sejam desenvolvidos nesta área. Pode ser realizada uma comparação de desempenho destes algoritmos, onde os trabalhos possam ser aplicados nas mesmas bases de dados e comparar os resultados obtidos na classificação de usuários influentes. E também pode ser combinado alguns destes algoritmos para criação de um novo algoritmo com melhores resultados.

Na Tabela 5.1 serão apresentados o resumo dos principais algoritmos estudados neste trabalho e suas respectivas aplicações no contexto de usuários influentes no Twitter.

Tabela 5.1: Resumo dos principais algoritmos estudados e suas aplicações.

Publicação	Algoritmo/Técnica	Aplicação
(WENG <i>et al.</i> , 2010)	Algoritmo TwitterRank baseado no PageRank.	Contexto de TOP-1000 usuários do Twitter da Cingapura.
(CHA <i>et al.</i> , 2010)	Algoritmo que mede a influência pelo número de seguidores, <i>retweets</i> e menções.	Contexto de temas populares no Twitter: eleição iraniana, surto gripe H1N1 e morte Michael Jackson.
(ROMERO <i>et al.</i> , 2011)	Algoritmo que determina a influência e a passividade com base nas atividades dos usuários.	Contexto para todos os <i>tweets</i> que continha uma URL.
(RAO; NAGPAL, 2011)	Algoritmo baseado no HITS.	Contexto de classificação da influência dos usuários em uma determinada área/cidade.
(SONG; LI; ZHENG, 2012)	Algoritmo baseado no PageRank para classificar os mais influentes nos tópicos relevantes.	Contexto aplicado em 5 tópicos relevantes de cada dia no período estudado.
(VALIATI <i>et al.</i> , 2012)	Algoritmo baseado no PageRank.	Contexto de dois cenários amplamente discutidos em um determinado período no Twitter: política (eleições presidenciais americanas para presidente) e automóveis.
(BIGONHA, 2012)	Algoritmo PageRank para analisar uma das métricas do SaID.	Contexto apresentou os usuários influentes de três marcas: refrigerante, eletrodomésticos e hipermercado.
(KING <i>et al.</i> , 2013)	Algoritmo baseado no H-index.	Contexto baseou-se em discussões sobre as reformas do Serviço Nacional de Saúde da Inglaterra dentro do Twitter.

6 ÁREAS DE APLICAÇÃO

Neste capítulo são descritas as principais áreas de aplicação dos estudos de usuários influentes no Twitter. Com isso, pode-se verificar a aplicação e os benefícios que os estudos de influência podem trazer para organizações, órgãos governamentais e demais instituições.

6.1 Contexto de Aplicação

Em (WENG *et al.*, 2010) aplicou-se o algoritmo TwitterRank de identificação de usuários influentes em um contexto dos TOP-1000 usuários do Twitter da Cingapura extraídos em abril de 2009 no site <http://twitaholic.com/>¹.

A partir de uma quantidade de dados coletados no Twitter (CHA *et al.*, 2010) optaram por temas mais populares em 2009 que foram: a eleição iraniana, surto da gripe H1N1 e morte de Michael Jackson. Neste trabalho investigou-se a dinâmica de influência dos usuários em tópicos. Nesse trabalho foi mostrado que a influência de um usuário sobre outro usuário é um conceito crucial para sociologia² e *marketing* viral.

(ROMERO *et al.*, 2011) aplicaram o seu estudo em uma base de dados do Twitter, onde obteve 22 milhões de *tweets* contendo URL's, utilizando a palavra chave "http" no ano de 2009, encontrando os usuários altamente conectados como elementos-chave na propagação de informações através do Twitter.

A proposta de (RAO; NAGPAL, 2011) descobriu como a informação é transmitida rapidamente por um determinado usuário e em uma determinada área. Utilizou-se da localização e das limitações dos usuários do Twitter para classificação da influência dos usuários em uma determinada área/cidade.

¹ Twitaholic encontra os principais usuários do Twitter com base no número de seguidores.

² Disponível em: <http://pt.wikipedia.org/wiki/Sociologia>

O estudo de (SONG; LI; ZHENG, 2012) classifica os tópicos mais relevantes e encontra os usuários mais influentes nestes tópicos. Foram classificados os 5 tópicos relevantes de cada dia no período de 04 de setembro de 2009 à 30 de setembro de 2009 com base no interesse dos usuários pelos tópicos, ou seja, tópicos com maior número de comentários feitos pelos usuários do Twitter. Além disso, o seu trabalho mostra o *ranking* dos usuários influentes.

A técnica proposta por (VALIATI *et al.*, 2012) consiste na avaliação no contexto de dois cenários amplamente discutidos na rede social Twitter que são : política (eleições presidenciais americanas para presidente) e automóveis. O foco do trabalho é ordenar usuários e conteúdos, de acordo com sua influência e sua relevância, respectivamente. Um exemplo citado neste trabalho, é o problema de otimizar a disseminação de conteúdo e explorá-lo para fins de *marketing* e divulgação de empresas.

O contexto aplicado por (BIGONHA, 2012) propõe um método para ordenar os usuários influentes de acordo com a sua polaridade em um determinado tópico no Twitter. Portanto, o trabalho apresenta os TOP-X Evangelistas e TOP-X Detratores de uma determinada marca/produto. A sua aplicação principal pode ser na área de *marketing*, área comercial, relacionamento com o cliente e também para diversas outras aplicações. O conteúdo gerado pelos usuários dentro de uma rede social se tornou estrategicamente importante para empresas interessadas em obter *feedback* da comunidade. O seu estudo analisou um conjunto de dados que contém *tweets* nos contextos a seguir: marca de refrigerante, marca de eletrodomésticos e hipermercado.

O contexto aplicado por (KING *et al.*, 2013) baseou-se em discussões sobre as reformas do Serviço Nacional de Saúde da Inglaterra dentro do Twitter. A sua base de dados consiste em 120.000 *tweets*, onde a intenção do seu trabalho investiga o papel do Twitter para informar, debater e influenciar a opinião em uma área específica de política de saúde.

Dentro de uma rede social como o Twitter que tem grande número de usuários postando, comentando, dando suas opiniões e sugestões, conseguir coletar diante destas inúmeras informações disponíveis uma forma de obter um *ranking* dos usuários mais influentes, pode ser de grande valor para uma determinada organização para um contexto definido. As organizações podem utilizar esse conhecimento para aplicar a seus negócios, obtendo vantagem competitiva com as informações dos usuários. Por exemplo, efetuando um levantamento de como a sua imagem e seus produtos estão sendo comentados pelos usuários mais influentes do Twitter, melhorar o relacionamento com o cliente, as campanhas de marketing podem ser aperfeiçoados, e a obtenção de *feedback* dos usuários pode auxiliar no lançamento de novos produtos. E também podendo ser útil em outros contextos como na identificação de usuários influentes que podem ser a fonte de informações em desastres e manifestações.

Os contextos de aplicação são os mais variados, pois as informações no Twitter não têm limites de contextos a serem discutidos pelos usuários. Alguns contextos de aplicação destas técnicas podem ser aplicadas no monitoramento da popularidade de programas de TV, filmes e seriados com intenção de verificar a influência dos usuários neste contexto. Pode-se aplicar também estas técnicas desenvolvidas em dois trabalhos distintos no mesmo contexto, para verificação do desempenho de cada técnica com intuito de comparar os resultados destes trabalhos em uma mesma base de dados.

Na Tabela 6.1 pode-se visualizar os contextos aplicados pelos trabalhos resumidamente, onde mostra a diversidade dos contextos de aplicações dos trabalhos estudados.

Tabela 6.1: Resumo dos contextos de aplicação.

Publicação	Aplicação
(WENG <i>et al.</i> , 2010)	Contexto de TOP-1000 usuários do Twitter da Cingapura.
(CHA <i>et al.</i> , 2010)	Contexto de temas populares no Twitter: eleição iraniana, surto gripe H1N1 e morte Michael Jackson.
(ROMERO <i>et al.</i> , 2011)	Contexto para todos os <i>tweets</i> que continha uma URL, com intenção de encontrar os usuários influentes na propagação de informações.
(RAO; NAGPAL, 2011)	Contexto de classificação da influência dos usuários em uma determinada área/cidade.
(SONG; LI; ZHENG, 2012)	Contexto aplicado em 5 tópicos relevantes de cada dia no período de 26 dias em 2009 extraídos do Twitter.
(VALIATI <i>et al.</i> , 2012)	Contexto de dois cenários amplamente discutidos em um determinado período no Twitter que são: política (eleições presidenciais americanas para presidente) e automóveis.
(BIGONHA, 2012)	Contexto apresentou os usuários influentes que são denominados TOP-X Evangelistas e TOP-X Detratores de três determinadas marcas: refrigerante, eletrodomésticos e hipermercado.
(KING <i>et al.</i> , 2013)	Contexto baseou-se em discussões sobre as reformas do Serviço Nacional de Saúde da Inglaterra dentro do Twitter.

7 DISCUSSÃO

Neste capítulo é realizada uma discussão em relação aos estudos de detecção de influentes no Twitter e também é apresentado um quadro resumindo os estudos sobre influência e as contribuições dos autores.

As principais características que foram observadas na identificação dos usuários influentes foram:

- Uso de *retweets*;
- Uso do número de seguidores;
- Uso de interações dentro do Twitter como: menções, respostas;
- Uso de algoritmos em destaque como o PageRank;
- Uso de temas específicos que estão em destaque no Twitter.

Além destas características a influência está relacionada ao período de tempo em que os dados foram coletados no Twitter.

Na tabela 7.1 é apresentado um resumo dos principais trabalhos publicados na área de detecção de influência de usuários no Twitter. E também, é mostrado o propósito, o algoritmo e o contexto de aplicação de cada trabalho. Com esta tabela tem-se o resumo das pesquisas realizadas na literatura. O principal objetivo de estudar as técnicas de detecção de influência dos usuários no Twitter é mostrar como está a pesquisa acadêmica nesta área. Portanto, este trabalho pode auxiliar os demais pesquisadores mostrando os principais pontos desta área.

Tabela 7.1: Resumo dos principais trabalhos de detecção de influência de usuários no Twitter.

Publicação	Propósito	Algoritmo/Técnica	Aplicação
(WENG <i>et al.</i> , 2010)	Identificar os usuários influentes no Twitter.	Algoritmo TwitterRank baseado no PageRank.	Contexto de TOP-1000 usuários do Twitter da Cingapura.
(CHA <i>et al.</i> , 2010)	Investigar a dinâmica de influência dos usuários em tópicos.	Algoritmo que mede a influência pelo número de seguidores, <i>retweets</i> e menções.	Contexto de temas populares no Twitter: eleição iraniana, surto gripe H1N1 e morte Michael Jackson.
(ROMERO <i>et al.</i> , 2011)	Analisar a influência e a passividade dos usuários.	Algoritmo que determina a influência e a passividade com base nas atividades dos usuários.	Contexto para todos os <i>tweets</i> que continha uma URL, com intenção de encontrar os usuários influentes na propagação de informações.
(RAO; NAGPAL, 2011)	Calcular influência do usuário considerando à localização e limitações de tempo.	Algoritmo baseado no HITS.	Contexto de classificação da influência dos usuários em uma determinada área/cidade.
(SONG; LI; ZHENG, 2012)	Classificar os tópicos mais relevantes e encontrar os usuários mais influentes.	Algoritmo baseado no PageRank para classificar os mais influentes nos tópicos relevantes.	Contexto aplicado em 5 tópicos relevantes de cada dia no período de 26 dias em 2009 extraídos do Twitter.
(VALIATI <i>et al.</i> , 2012)	Detectar os usuários influentes e conteúdo relevante.	Algoritmo baseado no PageRank.	Contexto de dois cenários amplamente discutidos em um determinado período: política (eleições presidenciais americanas para presidente) e automóveis.
(BIGONHA, 2012)	Desenvolver um método chamado SaID que propõe ordenar os usuários influentes de acordo com um determinado tópico.	Algoritmo PageRank para analisar uma das métricas do SaID.	Contexto apresentou os usuários influentes que são denominados TOP-X Evangelistas e TOP-X Detratores de três determinadas marcas: refrigerante, eletrodomésticos e hipermercado.
(KING <i>et al.</i> , 2013)	Analisar os usuários com maior influência em um contexto específico.	Algoritmo baseado no H-index.	Contexto baseou-se em discussões sobre as reformas do Serviço Nacional de Saúde da Inglaterra dentro do Twitter.

Após realização deste estudo foi possível identificar que ainda há outras oportunidades de pesquisa na área de detecção de usuários influentes, que são apresentadas a seguir.

- Os resultados gerados pelo algoritmo TwitterRank podem ser melhorados pelo acréscimo de um estudo do comportamento dos seguidores para identificar aqueles usuários que realizam postagens com melhor qualidade textual.
- O TwitterRank leva em consideração o número de *tweets* que um usuário publica, mas isso faz com que o cálculo dos usuários influentes se torne suscetível a manipulações se um usuário publicar um grande número de *tweets*. Isso pode ser melhorado considerando outras interações entre os usuários como respostas/menções entre dois usuários, por exemplo (WENG *et al.*, 2010).
- Estudos podem ser realizados utilizando dados de outras redes sociais visando verificar como é a influência de um mesmo usuário no contexto de mais de uma rede social.
- Pode-se avaliar o desempenho dos algoritmos utilizando bases de dados maiores.
- Um estudo comparativo dos resultados gerados pelos principais algoritmos, tais como, TwitterRank, H-Index, HITS e demais algoritmos baseados no PageRank, pode ser realizado.
- Pode-se realizar estudos que considerem a mudança dos temas mais populares e dos usuários mais influentes de acordo com o período de tempo determinado e com os eventos de grande importância.
- Avaliar se a recomendação realizada por um usuário influente no Twitter para um determinado produto realmente produz o marketing desejado para a organização.

- Estudar melhor a qualidade do conteúdo postado pelos usuários influentes visando identificar a credibilidade destes usuários.
- Avaliar os efeitos da influência positiva e da influência negativa para as organizações no Twitter.

8 CONCLUSÕES

Este trabalho apresentou uma revisão nos estudos de detecção de usuários influentes no Twitter. A pesquisa mostra os principais trabalhos nesta área, apresentando o propósito de cada trabalho, o algoritmo utilizado, e o contexto de aplicação destes trabalhos.

A principal contribuição deste trabalho é resumir de forma concisa os trabalhos que aplicaram as técnicas de detecção de usuários influentes no Twitter. As principais características dos usuários como: número de *retweets*, número de seguidores, número de interações, foram características destacadas por fazer parte de vários trabalhos estudados. Além disso, os algoritmos estudados como PageRank, TwitterRank, SaID, HITS e H-index podem servir como base para novos algoritmos, ou otimizando os algoritmos existentes. E também pode ser realizada uma comparação de desempenho destes algoritmos, onde os trabalhos possam ser aplicados nas mesmas bases de dados, e comparar os resultados obtidos na classificação de usuários influentes. Com isso, permitir que novos trabalhos sejam desenvolvidos, aplicando os algoritmos sobre novas bases de dados ou propondo processos diferentes de detecção de usuários influentes.

REFERÊNCIAS BIBLIOGRÁFICAS

BAO, H.; LI, Q.; LIAO, S. S.; SONG, S.; GAO, H. A new temporal and social pmf-based method to predict users' interests in micro-blogging. *Decision Support Systems*, v. 55, n. 3, p. 698 – 709, 2013. ISSN 0167-9236. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167923613000663>>.

BIGONHA, C. A. S. Sentiment-based influence detection on twitter. 2012.

BIGONHA, C. A. S.; CARDOSO, T. N. C.; MORA, M. M.; GONÇALVES, M. A.; ALMEIDA, V. A. F. Sentiment-based influence detection on twitter. *Journal of the Brazilian Computer Society*, v. 18, p. 169–183, 2011.

BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. In: *Proceedings of the seventh international conference on World Wide Web 7*. Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V., 1998. (WWW7), p. 107–117. Disponível em: <<http://dl.acm.org/citation.cfm?id=297805.297827>>.

CHA, M.; HADDADI, H.; BENEVENUTO, F.; GUMMADI, K. P. Measuring user influence in twitter: The million follower fallacy. In: *In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. [S.l.: s.n.], 2010.

CHEN, W.; CHENG, S.; HE, X.; JIANG, F. Influencerank: An efficient social influence measurement for millions of users in microblog. In: *Cloud and Green Computing (CGC), 2012 Second International Conference on*. [S.l.: s.n.], 2012. p. 563–570.

FRAHM, K.; SHEPELYANSKY, D. Google matrix of twitter. *The European Physical Journal B*, Springer-Verlag, v. 85, n. 10, p. 1–7, 2012. ISSN 1434-6028. Disponível em: <<http://dx.doi.org/10.1140/epjb/e2012-30599-6>>.

GIL, A. C. *Como Elaborar Projetos de Pesquisa*. 4. ed. [S.l.]: Atlas, 2002.

HIRSCH, J. E. Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, v. 104, n. 49, p. 19193–19198, 2007. Disponível em: <<http://www.pnas.org/content/104/49/19193.abstract>>.

JAMALI, M.; ESTER, M. A matrix factorization technique with trust propagation for recommendation in social networks. In: *Proceedings of the fourth ACM conference on Recommender systems*. New York, NY, USA: ACM, 2010. (RecSys '10), p. 135–142. ISBN 978-1-60558-906-0. Disponível em: <<http://doi.acm.org/10.1145/1864708.1864736>>.

KATZ, E.; LAZARFELD, P.; RESEARCH, C. U. B. of A. S. *Personal Influence: The Part Played by People in the Flow of Mass Communication*. Free Press, 1955. (Foundations of Communications Research). Disponível em: <<http://books.google.com.br/books?id=PgWxAAAAIAAJ>>.

KING, D.; RAMIREZ-CANO, D.; GREAVES, F.; VLAEV, I.; BEALES, S.; DARZI, A. Twitter and the health reforms in the english national health service. *Health Policy*, v. 110, n. 2–3, p. 291 – 297, 2013. ISSN 0168-8510. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168851013000456>>.

KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. *J. ACM*, ACM, New York, NY, USA, v. 46, n. 5, p. 604–632, set. 1999. ISSN 0004-5411. Disponível em: <<http://doi.acm.org/10.1145/324133.324140>>.

KLOUT. 2013. Disponível em: <<http://klout.com/home>>.

LIANG, H.; LU, G.; XU, N. Analyzing user influence of microblog. In: *Advanced Computational Intelligence (ICACI), 2012 IEEE Fifth International Conference on*. [S.l.: s.n.], 2012. p. 15–22.

LIU, L.; TANG, J.; HAN, J.; YANG, S. Learning influence from heterogeneous social networks. *Data Mining and Knowledge Discovery*, Springer US, v. 25, n. 3,

p. 511–544, 2012. ISSN 1384-5810. Disponível em: <<http://dx.doi.org/10.1007/s10618-012-0252-3>>.

LO, Y. W.; POTDAR, V. *A review of opinion mining and sentiment classification framework in social networks*. Ieee, 2009. 396–401 p. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5276705>>.

PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. The pagerank citation ranking: Bringing order to the web. In: *Proceedings of the 7th International World Wide Web Conference*. Brisbane, Australia: [s.n.], 1998. p. 161–172.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, Now Publishers Inc., Hanover, MA, USA, v. 2, n. 1-2, p. 1–135, jan. 2008. ISSN 1554-0669. Disponível em: <<http://dx.doi.org/10.1561/1500000011>>.

RAO, T.; NAGPAL, S. Real-time geo influence in social networks. In: *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*. [S.l.: s.n.], 2011. v. 1, p. 246–250.

ROMERO, D. M.; GALUBA, W.; ASUR, S.; HUBERMAN, B. A. Influence and passivity in social media. In: *Proceedings of the 20th international conference companion on World wide web*. New York, NY, USA: ACM, 2011. (WWW '11), p. 113–114. ISBN 978-1-4503-0637-9. Disponível em: <<http://doi.acm.org/10.1145/1963192.1963250>>.

SONG, S.; LI, Q.; ZHENG, X. Detecting popular topics in micro-blogging based on a user interest-based model. In: *Neural Networks (IJCNN), The 2012 International Joint Conference on*. [S.l.: s.n.], 2012. p. 1–8. ISSN 2161-4393.

VALIATI, H.; SILVA, A.; GUIMARAES, S.; JR., W. M. Detecção de conteúdo relevante e usuários influentes no twitter. 2012. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/brasnam/2012/008.pdf>>.

WENG, J.; LIM, E.-P.; JIANG, J.; HE, Q. Twiterrank: finding topic-sensitive influential twitterers. In: *Proceedings of the third ACM international conference on Web search and data mining*. New York, NY, USA: ACM, 2010. (WSDM '10), p. 261–270. ISBN 978-1-60558-889-6. Disponível em: <<http://doi.acm.org/10.1145/1718487.1718520>>.

ZABIN, J.; JEFFERIES, A. *Social Media Monitoring and Analysis: Generating Consumer Insights from Online Conversation*. January 2008. Aberdeen Group Benchmark Report.

A APÊNDICES

Documento de Artigos

Autor(es): DING Zhaoyun, JIA Yan, ZHOU Bin e HAN Yi

Título: Mining Topical Influencers Based on the Multi-Relational Network in Micro-Blogging Sites

Publicação/Ano: China Communications January 2013

Número de Citações: 1.

Abstract

No contexto de micro-blogging, como o Twitter, o número de produtores de conteúdo podem facilmente chegar a dezenas de milhares de pessoas, e muitos usuários podem participar na discussão de um determinado tópico. Enquanto muitos usuários podem introduzir diversidade, nem todos os usuários são igualmente influente, torna-se difícil de identificar os verdadeiros formadores de opinião, que são geralmente classificados como sendo interessante e de autoridade sobre um determinado assunto. Neste estudo, a influência dos usuários é medido através da realização de caminhadas aleatórias dos dados multi-relacionais em micro-blogging: Retweet, resposta, reintroduzir, e ler. Devido à incerteza dos reintroduzir as operações de leitura e, de um novo método é proposto para determinar a probabilidade de transição de redes relacionais incertos. Além disso, propomos um método para realizar os passeios aleatórios combinados para a rede de influência multi-relacional, considerando tanto as probabilidades de transição para intra-e inter-networking. Os experimentos foram conduzidos em um conjunto de dados reais Twitter contendo cerca de 260 000 usuários e 2,7 milhões de tweets, e os resultados mostram que o nosso método é mais eficaz do que TwitterRank e outros métodos utilizados para descobrir influenciadores.

Resumo (em um parágrafo contendo a principal contribuição)

Muito relacionado. Verificar se ele faz alguma classificação para ranquear os influentes.

Faz o ranking de precisão dos métodos, porém nada relacionado a ranking de usuários.

Faz algum tipo de recomendação?

Neste trabalho, para medir a influência, ignoramos atributos pessoais, tais como a novidade de tweets e a sensibilidade de tweets. No futuro, vamos identificar as pessoas influentes com base em características diferentes dos usuários. Além disso, os atributos pessoais e da rede multi-relacional deve ser considerado de forma abrangente para medir a influência de usuários.

Fala sobre algoritmos utilizados? Fala do método multi-relacional MultiRank2, mas nada específico com relação a processamento.

Foi aplicado a algum contexto específico (saúde, tráfego, política, empresas, etc)?

Não. Foi coletado de 260.000 usuários e 2,7 milhões de tweets na China.