



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Ciência da Computação

Influência e passividade em redes sociais

Juliana Medeiros de Lucena

Trabalho de Graduação

Recife

13 de dezembro de 2011

Universidade Federal de Pernambuco

Centro de Informática

Juliana Medeiros de Lucena

Influência e passividade em redes sociais

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: *Tsang Ing Ren*

Co-orientador: *Alex Sandro Gomes*

Recife

13 de dezembro de 2011

Às crianças cujo direito à educação digna foi negado

Agradecimentos

Inicialmente, agradeço aos meus amados pais por todos os esforços feitos para que eu pudesse ter ótimas oportunidades e experiências. Ao meu pai, garoto do interior que estudou até a 4ª série e mostrou que não é só de livros que se faz o homem, não posso deixar de agradecer pela dica de que “a vida é um livro aberto”, porque por mais clichê que seja, a vida foi seu maior livro. À minha querida mãe eu agradeço a sua preocupação, dedicação e, principalmente, paciência demonstradas incansavelmente durante todos os dias; e, acima de tudo, aos seus ensinamentos de como cuidar das pessoas. À minha avó eu tenho que agradecer pela sua compreensão em minha ausência e por sempre ter acreditado no meu potencial sem pestanejar. À todos os meus familiares, tios e primos, eu agradeço todo o apoio e vivência.

Aos meus professores e amigos do Mazzarello e do CPI agradeço pelos anos de participação na minha vida e por tudo que aprendi e vivi, porque estes anos influenciam na minha vida desde então e os eu carrego com muito orgulho.

Aos meus amigos de graduação e principalmente àqueles que vieram comigo do CPI e me fizeram sentir em casa desde o primeiro instante que pisei na Universidade, leia-se Carol Martiniano, Edilson Ferreira, Tiago Ferreira e Filipe Wanderley, agradeço por tornarem estes anos tão divertidos, agradáveis e proveitosos. Não posso esquecer os momentos maravilhosos que passei com o grupo PET e todos os amigos que lá fiz; nossos momentos foram muito estimulantes e tornaram a minha graduação mais brilhante. A todos que fazem o Centro de Informática agradeço por todo o trabalho e preocupação em nos oferecer uma formação de ótima qualidade.

Ao meu amado companheiro, Tiago Ferreira, agradeço por sua adorável companhia não só na graduação, mas na vida, pois ele soube me compreender e aconselhar como ninguém. A nova família que ganhei, Emanuel Soares, Ivanete Ferreira, Iraci Ferreira, Taísa Ferreira e Tácio Ferreira, e à nossa fiel piadista Vilma Silva, não posso deixar de agradecer pelo apoio, ensinamentos e por todos os momentos felizes que partilhamos.

Aos meus irmãos de sonho, Tiago Ferreira, André Diniz, Guilherme Cavalcanti, Alex Gomes, Cassio Melo, Anderson Duarte e Sérgio Fontes, agradeço por terem me mostrado uma nova graduação e por não me deixarem fraquejar ao me inspirar todos os dias.

Agradeço também ao meu orientador, Tsang Ing, por ter aceitado este desafio comigo e ter me guiado durante este trabalho. Ao meu co-orientador, Alex Gomes, agradeço pelo mesmo e também pelo seu ânimo inabalável. Às minhas amigas Carol Martiniano, Laís Sales e Laís Rodrigues agradeço por terem me ajudado de algum modo neste trabalho, seja corrigindo ou me dando forças quando precisei.

Por fim, agradeço a todos por entenderem a minha ausência em alguns momentos e por terem participado destes anos de algum modo, o apoio de todos vocês é inestimável. Claramente, não posso esquecer dos músicos que não me deixaram na mão em nenhum momento.

*Develop a passion for learning.
If you do, you will never cease to grow.*
—ANTHONY J. D'ANGELO

Resumo

Os fenômenos sociais da disseminação da informação e formação de opinião são estudados há décadas em diversas áreas do conhecimento, porém suas teorias não puderam ser confirmadas devido a dificuldade de coleta das informações necessárias, bem como da complexidade que envolve estes fenômenos.

Este trabalho visa colaborar para o entendimento destes fenômenos sociais através da observação do comportamento geral dos indivíduos em uma rede social de informações. Para esta análise, foram utilizadas técnicas de Análises de Redes Sociais alinhadas com teorias da comunicação. Os resultados obtidos mostram que observar outras variáveis do comportamento, como a passividade dos usuários, são abordagens válidas para a identificação de usuários influentes em um determinado tópico.

Palavras-chave: redes de informação; difusão de informação; influência; *word of mouth* marketing; SNA

Abstract

The social phenomena of information dissemination and opinion formation has been studied for decades in many areas of knowledge, nevertheless, its theories could not be confirmed due to the difficulty to collect the necessary information, as well as the complexity that those phenomena involve.

This work aims to help with the understanding of those social phenomena by observing the general behavior of individuals in an information social network. For this analysis, techniques of Social Network Analysis aligned with communication theories were used. The obtained results show that observing other variables of behavior, like user passivity, is a valid approach to identify influentials in a determined topic.

Keywords: information networks; information difusion; influence; word of mouth marketing; SNA

Sumário

1	Introdução	1
1.1.	Motivação	2
1.2.	Objetivos	2
1.3.	Estrutura do documento	3
2	Análise da influência em redes sociais	4
2.1.	Análise de Redes Sociais	4
2.2.	Influência em redes sociais	5
2.3.	Influência no Twitter	6
2.3.1.	Influência em tópico específico	7
3	Identificação dos influentes da rede em tópicos específicos	8
3.1.	Aspectos da transmissão de informação	8
3.2.	Algoritmo <i>Influence-Passivity</i> (IP)	8
3.2.1.	Grafo de entrada	9
3.2.2.	Premissas, métricas e funcionamento	9
4	Experimentos e Resultados	12
4.1.	A rede de informações Twitter	12
4.2.	Base de Dados	13
4.3.	Aplicação do <i>Influence-Passivity</i> (IP)	13
4.3.1.	Implementação	13
4.3.2.	Resultados Obtidos	14
4.4.	Quantidade de <i>Retweets</i>	18
4.4.1.	Resultados Obtidos	19
4.5.	Considerações finais	20
5	Conclusão	21
5.1.	Principais Contribuições	21
5.2.	Trabalhos Futuros	21
	Referências Bibliográficas	24

Lista de Figuras

<i>Figura 1 - Modelo de influência two-step flow (Watts e Dodds 2007)</i>	1
<i>Figura 2 - Modelo de rede de influência de Watts e Dodds (2007)</i>	1
<i>Figura 3 - Crescimento da análise de redes sociais (Otte e Rousseau 2002)</i>	5

Lista de Tabelas

<i>Tabela 1 - Legenda das cores utilizadas nas tabelas seguintes.....</i>	<i>14</i>
<i>Tabela 2 - Usuários mais influentes pelo IP (base-97).....</i>	<i>14</i>
<i>Tabela 3 - Usuários mais influentes pelo IP (base-50).....</i>	<i>15</i>
<i>Tabela 4 - Usuários mais passivos pelo IP (base-97).....</i>	<i>16</i>
<i>Tabela 5 - Usuários mais passivos pelo IP (base-50).....</i>	<i>17</i>
<i>Tabela 6 - Usuários mais influentes pela quantidade de retweets (base-97).....</i>	<i>19</i>
<i>Tabela 7 - Usuários mais influentes pela quantidade de retweets (base-50).....</i>	<i>19</i>

Introdução

Nas décadas de 40 e 50, a teoria da formação de opinião obteve um grande avanço através dos estudos de Katz, Lazarsfeld, Berelson e Gaudet (Katz e Lazarfeld, 1955; Lazarfeld, Berelson e Gaudet, 1968). Estas pesquisas formularam uma nova teoria ao tentar conciliar o papel da mídia como influenciadora e a crescente percepção de que as pessoas podem ser mais influenciadas umas pelas outras do que pela mídia. Katz e Lazarfeld (1955) criaram o modelo de comunicação two-step flow (Figura 1), no qual uma pequena minoria (chamada de líderes de opinião) exerce a posição de intermediário entre a mídia e a grande maioria da sociedade (Watts e Dodds, 2007). Deste modo, a informação é passada da mídia para os seguidores dos líderes de opinião através destes últimos.

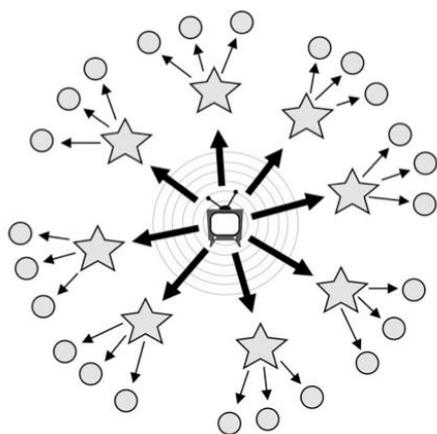


Figura 1 - Modelo de influência *two-step flow* (Watts e Dodds, 2007)

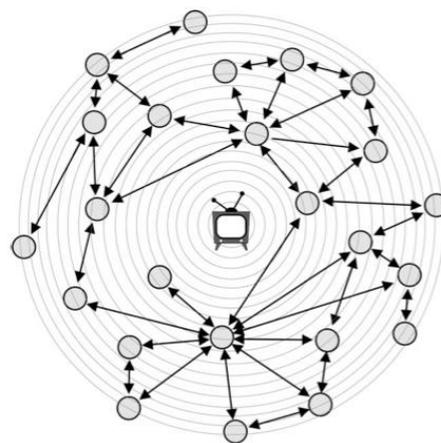


Figura 2 - Modelo de rede de influência de Watts e Dodds (2007)

O modelo de influência two-step flow ficou em evidência nas literaturas de difusão de inovação, marketing, entre outros durante décadas seguintes; de tal modo que Burt (1999) afirmou que o modelo, em tradução livre: "se tornou um guia na pesquisa de difusão e marketing".

Em um segundo momento, Watts and Dodds (2007) criaram um modelo de influência (Figura 2), no qual a influência: (1) pode fluir tanto dos influentes para os seus seguidores como na via contrária; (2) pode propagar por vários passos (através de vários indivíduos). Estas características diminuem o papel dos influentes da rede e colocam como os principais fatores da influência: a relação interpessoal entre usuários comuns e a disposição da sociedade para adotar inovação (Cha *et al.*, 2010). Ou seja, esta nova

abordagem destaca a cultura a frente dos influentes, pois segundo Domingos e Richardson (2001), devido ao grande compartilhamento de informação as pessoas se baseiam na opinião de seus próximos e não na dos influentes.

O fenômeno da influência na formação de opinião ainda possui várias lacunas, pois a rede pela qual a influência espalha-se, na maioria das vezes, não podem ser observadas (Cha *et al.*, 2010; Bakshy *et al.*, 2011). Porém, com a Era da Informação e o surgimento de redes sociais, a observação do processo de difusão de informação se tornou mais palpável, porque estes ambientes virtuais são permeados de relacionamentos interpessoais reais.

O Twitter (2011) intitula-se, em tradução livre: "uma rede de informação em tempo-real que conecta você às últimas informações sobre o que você acha interessante". Este serviço de micro-blogging associado à uma rede social se mostra um ambiente promissor para a análise de influência (Bakshy, 2011), pois ele possibilita a disseminação de informação entre redes de contatos reais em um ambiente monitorável.

1.1. Motivação

Online Word-of-mouth (OWOM) é a transmissão online de informação de pessoa para pessoa (Jansen *et al.*, 2009), em tradução livre: "boca-a-boca". Este processo é utilizado como tática de marketing pelas empresas com o objetivo de influenciar os consumidores. A sua efetividade se baseia no fato de que as pessoas confiam mais nos indivíduos da sua rede social para opiniões comerciais (Jansen *et al.*, 2009). Portanto, OWOM é utilizada pelas empresas para formação de opinião positiva com relação aos seus produtos e a sua marca, por exemplo.

Segundo Milstein *et al.* (2010), o Twitter também pode ser encarado como uma plataforma de marketing. Isto também significa que a estrutura do micro-blogging possibilita a constituição de uma rede de influência, através das trocas de informações, que resulta na formação de opinião. Ou seja, para ocorrer uma maior disseminação da inovação e aumento da eficiência de uma campanha de marketing, é interessante alcançar os usuários influentes da rede, pois isso causará o aumento dos efeitos do OWOM, ao passo que estes usuários irão influenciar na opinião de outros indivíduos e assim por diante (Kempe, Kleinberg e Tardos, 2005). Um melhor direcionamento da campanha pode ser feito ao visar usuários mais influentes em tópicos específicos, pois a identificação destes usuários pode melhorar as opiniões coletadas (Weng *et al.*, 2010).

1.2. Objetivos

Os objetivos principais deste trabalho de graduação são: (1) estudar as características da formação de opinião em um ambiente virtual, (2) identificar o comportamento dos usuários em uma rede de influência, bem como quais são os usuários que iniciam a propagação da influência sobre determinado tópico. Será utilizada a rede de informação Twitter em conjunto com técnicas de análise de rede social para melhor compreender os fatores determinantes no fenômeno social complexo que é a formação de opinião. Deste modo, os processos de marketing poderão compreender melhor os fenômenos sociais e serem mais efetivos em suas ações.

1.3. Estrutura do documento

No capítulo que segue será feita uma revisão das pesquisas mais relevantes para este trabalho no tocante a análise de redes sociais, influência e formação de opinião.

No terceiro capítulo será proposta a utilização do algoritmo *Influence-Passivity* (Romero *et al.*, 2010) para identificação dos usuários influentes em determinado tópico. Esta abordagem será contraposta à abordagem que utiliza apenas a quantidade de mensagens de um usuário que foram repassadas pelos seus seguidores. Então, no quarto capítulo serão apresentados os experimentos e resultados obtidos. Por fim, no capítulo 5, serão discutidas as principais contribuições deste trabalho e os possíveis trabalhos futuros.

Análise da influência em redes sociais

Neste capítulo será dada uma visão geral das pesquisas mais relevantes para este trabalho. Para melhor compreensão, serão apresentadas inicialmente as pesquisas relacionadas à análise de redes sociais; em seguida, aquelas relacionadas à identificação de usuários influentes na rede; por fim, serão explanados os trabalhos referentes à identificação de usuários influentes em tópicos específicos dentro da rede.

2.1. Análise de Redes Sociais

A análise de redes sociais, segundo Otte e Rosseau (2002), em livre tradução: "não é uma teoria formal, mas sim uma ampla estratégia para investigar as estruturas sociais". Embora tanto as relações entre os indivíduos, como as características individuais sejam determinantes para entender um fenômeno social, as estratégias de análise de redes sociais visam priorizar as relações entre os indivíduos (Knoke e Kuklinski, 1982). Essa priorização é feita, pois as regularidades das estruturas influencia o comportamento dos indivíduos (Otte e Rosseau, 2002).

A larga utilização da análise de redes sociais, também conhecida por *Social Network Analysis* (SNA), foi evidenciada na pesquisa de Otte e Rosseau (2002), onde foi verificado o crescimento do seu uso em publicações científicas. A Figura 3 mostra este crescimento, o qual se mostrou expressivo a partir da década de 80; além disso, também revela que vários campos utilizam SNA, pois as bases de dados utilizadas são provenientes de vários campos de estudo. Alguns exemplos de uso são: análise de redes de colaboração de eventos científicos (Medeiros, 2010), análise de como as organizações interagem, compreensão da influência dos padrões de contato humano na disseminação de doenças, recomendação de relacionamentos e observação da difusão de inovação nas redes sociais (Silva, 2010).

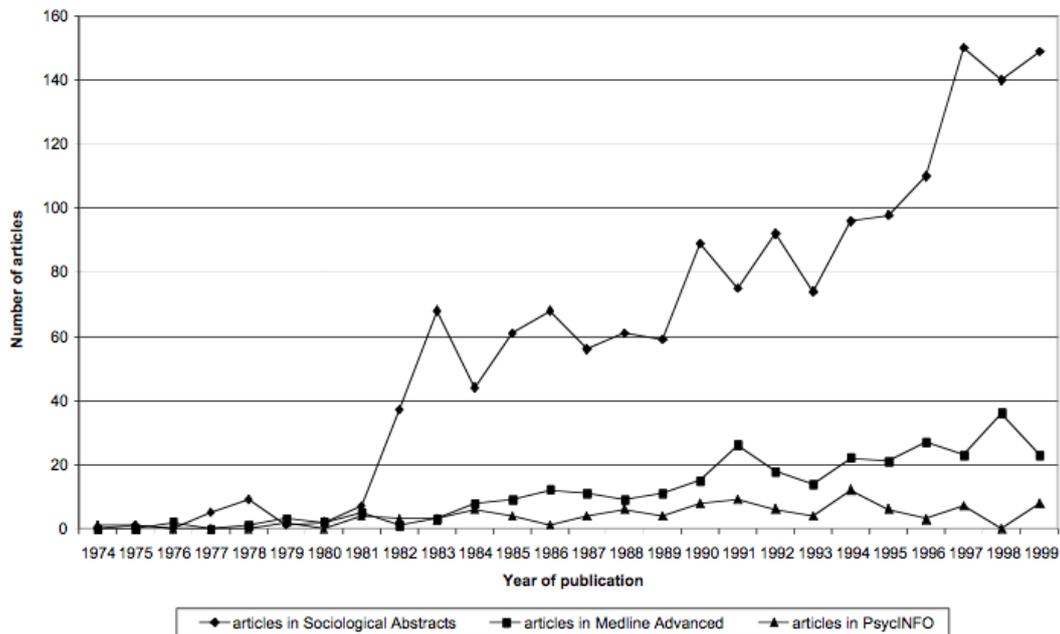


Figura 3 - Crescimento da análise de redes sociais (Otte e Rousseau 2002)

2.2. Influência em redes sociais

Várias áreas do conhecimento como ciências políticas, marketing, comunicação e sociologia estudam há alguns anos a propagação da informação e da influência. Estes estudos geraram teorias sobre como as tendências surgem, porém muitas teorias são conflitantes entre si.

A visão tradicional, apoiada por Katz e Lazarsfeld (1955), Rogers (1962) e Gladwell (2002), assume que na sociedade existem pessoas que possuem certas características que as fazem conduzir as tendências. Estas pessoas influenciam as massas (usuários comuns) através da disseminação de informações e ideias, as quais ganham relevância por estarem sendo veiculadas por estas pessoas ditas "influentes".

Em contrapartida, Domingos e Richardson (2001), Watts e Dodds (2007) e Watts (2007) apoiam que estes usuários excepcionalmente persuasivos não são necessários nem suficientes para a geração das tendências. Estes defendem a importância da cultura acima dos usuários super influentes, pois devido a grande possibilidade de troca de informações que a Web 2.0 proporciona, as pessoas dessa era levam em conta a opinião baseando-se na confiança. Ou seja, o papel dos usuários comuns devem ser levados em conta, pois eles também participam da troca de informações e tem a capacidade de influenciar outras pessoas (mesmo que em menor escala). Outro fator defendido nestas pesquisas é que a concretização de uma tendência também depende do estado atual que a sociedade vive, ou seja, da capacidade de receber esta nova tendência. Isto significa que para uma tendência ser espalhada, basta o ambiente certo e a iniciativa de um usuário comum.

As visões citadas não foram de fato confirmadas por motivos como: dificuldade na coleta de dados e, principalmente, a definição do que a influência realmente é (Cha,

2010). Porém, com o aumento do uso das redes sociais, vários estudos começaram a ser feitos com os dados destas redes, por exemplo: Sun et al. (2009) avaliou a difusão das *fan pages* do Facebook¹, Bakshy et al. (2009) estudou o compartilhamento e adoção de conteúdo no Second Life (Rymaszewski, 2007) e Agarwal et al. (2008) descobriu que os blogueiros mais ativos não são consequentemente os mais influentes ao identificar os blogueiros mais influentes da blogosfera.

2.3. Influência no Twitter

Muitas pesquisas escolheram o Twitter² como estudo de caso devido a sua natureza de disseminação de informação, o que o torna um ótimo ambiente para estudar a difusão de influência e informação.

Bakshy et al. (2011) avaliou a influência de todos os usuários, não apenas dos já conhecidos por serem influentes. A análise foi feita baseando-se na capacidade de espalhamento de um *tweet*, tanto de usuários ligados diretamente à fonte (usuário criador do *tweet*), como dos indiretamente ligados. Foi concluído que para se fazer uma campanha de marketing viral com um bom custo-benefício, basta utilizar-se dos usuários comuns, enquanto que para se ter um resultado mais efetivo, porém mais caro, o foco nos usuários influentes é altamente indicado. Jasen et al. (2009) investigou o Twitter como uma forma de OWOM, neste estudo foram analisadas as postagens, os tipos das expressões e a mudança de sentimento. Foi confirmado que o Twitter é uma ferramenta de OWOM e discutido os impactos gerados pelo uso de um serviço de *microblogging* na estratégia de marketing das corporações.

Huberman et al. (2008) verificou que as estruturas ligadas nas redes sociais não representam de fato interações entre as pessoas, pois o ritmo da vida e do trabalho aliados escassez de atenção faz com que as pessoas interajam apenas com um grupo reduzido. Este grupo reduzido forma uma escondida e esparsa rede de contatos, ao invés do grupo oficialmente indicado no sistema. Em consonância com isso, Cha et al. (2010) fez uma comparação entre três medidas: *indegree*, *retweets* e *mentions*, e concluiu que usuários populares não são necessariamente influentes. Ou seja, o número de seguidores pode não ser uma boa medida de influência, enquanto que *retweets* e *mentions* mostram uma boa correlação.

Por outro lado, Weng et al. (2010) propôs a utilização da estrutura social da rede, pois identificou que ocorre homofilia no Twitter, visto que a reciprocidade dos usuários no tocante a um seguir o outro é alta. Porém, Cha et al. (2010) afirmou que a reciprocidade no Twitter é baixa, deste modo, contrapondo o trabalho de Weng et al. (2010). Aral et al. (2009) encontrou que a homofilia também é responsável pela difusão em redes e distinguiu os efeitos desta dos efeitos da influência no tocante a propagação da informação.

Galuba et al. (2010), assim como Bakshy et al. (2011), utilizou o conceito de cascatas para entender o fluxo de informação. Neste trabalho, foi proposto um modelo de propagação que prevê qual usuário irá criar um *tweet* sobre uma URL baseado na

¹ <http://www.facebook.com/>

² <http://www.twitter.com/>

atividade passada do usuário. Romero *et al.* (2010) também se utilizou de URLs para fazer um estudo da influência dos usuários do Twitter, porém este também levou em conta a passividade dos outros usuários para definir quão influente um usuário é. Em suma, um usuário é tão influente quanto mais usuários passivos ele conseguir tirar da inércia. Esta pesquisa, assim como Cha *et al.* (2010), teve a pouca relação entre popularidade e influência evidenciada.

2.3.1. Influência em tópico específico

Algumas pesquisas foram feitas com o objetivo de identificar quais usuários são mais influentes em tópicos determinados e não de um modo geral. Weng *et al.* (2010) identificou os usuários mais influentes utilizando similaridade em tópicos de interesse entre os usuários e a estrutura social da rede. Isto foi feito ao adicionar a similaridade de tópicos ao *PageRank* (Page *et al.*, 1999). Este trabalho fez a identificação dos tópicos de interesse dos usuários através da análise de conteúdo dos *tweets* e defendeu que as medidas de *retweet*, *reply* e *mention* não levam em conta a estrutura global da rede entre os usuários.

Um trabalho similar foi feito por Haveliwala (2003), porém este trabalho propaga a influência de um usuário em um único tópico para todos os usuários que o seguem em vários tópicos com probabilidades diferentes, enquanto que Weng *et al.* (2010) aplica probabilidades de transições diferentes para tópicos diferentes. Com esta abordagem, Weng *et al.* (2010) mostrou capturar a influência em um tópico específico melhor do que a outra pesquisa.

Romero *et al.* (2010) utilizou a passividade de outros usuários para calcular a influência de um indivíduo na rede. Para isso, ele se utilizou de *retweets* para verificar o fluxo de URLs citadas nos *tweets* criados. Nesta pesquisa também foi verificada a pouca relação entre popularidade e influência. Apesar de não ter apresentado um experimento a respeito, Romero *et al.* (2010) afirmou que se o algoritmo apresentado for aplicado a um grafo que possui *tweets* que tratam de um tópico em específico, os usuários mais influentes neste tópico serão determinados. A pesquisa também deixa em aberto se o algoritmo será igualmente preciso em grafos, restritos a um tópico, de diferentes escalas.

Identificação dos influentes da rede em tópicos específicos

Este trabalho propõe a utilização do algoritmo *Influence-Passivity* (IP) (Romero *et al.*, 2010) para a identificação dos usuários mais influentes na rede social em relação a um certo tópico ou assunto. Visto que o fenômeno da formação de opinião é bastante complexo, neste capítulo serão comentados alguns aspectos da transmissão de informação e quais deles serão levados em conta para a identificação dos usuários mais influentes. Por fim, será explicada a dinâmica do algoritmo IP.

3.1. Aspectos da transmissão de informação

O algoritmo IP foi baseado em dois aspectos da transmissão de informação: a influência e a passividade dos membros da malha social. A definição de influência considerada foi a capacidade de propagar conteúdos através da rede, enquanto que a passividade é a aptidão de colocar uma barreira no fluxo da informação. A popularidade dos usuários, a qual é determinada pela quantidade de atenção recebida, também é um dos aspectos de transmissão de informação. Porém, a correlação entre esta e a influência é baixa, visto que os usuários mais populares não necessariamente são os mais influentes (Cha *et al.*, 2010; Romero *et al.*, 2010).

A influência, de acordo com Cha *et al.* (2010), em livre tradução é: "o potencial de um indivíduo em levar os outros a se engajarem em um certo ato". Este potencial também depende dos usuários que são influenciados, pois a passividade destes é uma barreira a ser quebrada. Quanto maior for a passividade de um usuário, ou seja, quanto menor for a sua participação na disseminação da informação, maior a dificuldade de influenciar este usuário ao ponto dele repassar a mensagem para a sua rede. Deste modo, a quantidade de usuários influenciados não é o fator determinante para definir a influência de um indivíduo, pois esta vai depender da dificuldade em influenciar este grupo (Romero *et al.*, 2010).

3.2. Algoritmo *Influence-Passivity* (IP)

O algoritmo IP é similar ao HITS (Kleinberg, 1999), este foi criado para identificar as páginas da Web com maior autoridade e os *hubs* que apontam para várias páginas relacionadas. Na abordagem do IP, o grau de influência é semelhante ao grau dos *hubs*, enquanto que o grau de passividade é semelhante ao grau das páginas autoritárias.

Este algoritmo associa para cada usuário um grau de influência e passividade, os quais são encontrados ao levar em conta as propriedades estruturais da rede, como também o comportamento de difusão de informação entre os usuários.

3.2.1. Grafo de entrada

O grafo de entrada $G = (N, E, W)$, onde N são os nós, E são os arcos e W os pesos. Os nós são os usuários da rede social a ser analisada, já os arcos representam a taxa de influência que um usuário exerce sobre o outro, em relação à influência total que o primeiro tentou exercer sobre o segundo. De modo mais prático, o arco $e = (i, j)$ representa a taxa de mensagens de i que j repassou para a sua rede (S_{ij}) em relação a quantidade total de mensagens que i publicou e chegou à j (Q_{ij}). Portanto, o peso do arco e pode ser visto da seguinte forma: $e = \frac{S_{ij}}{Q_{ij}}$.

3.2.2. Premissas, métricas e funcionamento

Como dito anteriormente, a influência de um indivíduo depende da quantidade de pessoas que ela influencia e quão passivas são estas pessoas. Já a passividade, depende de quão influentes são as pessoas que tentam influenciar um indivíduo, mas não conseguem. O fato de uma pessoa aceitar ou não a influência de outra, é denominada **taxa de aceitação** e **taxa de rejeição**, respectivamente. Estas taxas são definidas e obtidas das seguintes formas:

- A taxa de aceitação indica quanta consideração ou lealdade um usuário tem pelo outro, ou seja, quanta atenção um indivíduo dá ao outro em relação a atenção que ele despense com todos os outros. Deste modo, cada arco $e = (i, j) \in E$ possui uma taxa de aceitação u_{ij} (ver Equação 1), a qual indica o quanto de influência j aceitou de i normalizado pelo total de influência que j aceitou de todos os outros usuários;
- A taxa de rejeição indica quanta rejeição um usuário tem pelo outro, ou seja, quão rejeitado um membro foi por outro em relação a rejeição deste outro para com os demais indivíduos. Neste caso, cada arco $e = (j, i) \in E$ possui uma taxa de rejeição v_{ji} (ver Equação 2), a qual indica quanta influência i rejeitou de j ($1 - w_{ji}$) normalizado pelo total de influência que j rejeitou dos demais usuários.

$$u_{ij} = \frac{W_{i,j}}{\sum_{k:(k,j) \in E} W_{k,j}}$$

Equação 1 - Cálculo da taxa de aceitação

$$v_{ji} = \frac{1 - w_{ji}}{\sum_{k:(j,k) \in E} (1 - w_{jk})}$$

Equação 2 - Cálculo da taxa de rejeição

O cálculo do grau de influência e passividade de cada membro são calculados simultaneamente, visto que um depende do outro. Para efetuar este cálculo, o algoritmo IP recebe um grafo no formato especificado na seção 3.2.1. Em seguida, como pode ser visto no Algoritmo 1, serão gerados m valores de influência e passividade para cada usuários, pois m é a quantidade de iterações do algoritmo. Inicialmente, todos os usuários possuem valor 1 para as duas pontuações (I_0, P_0). Em conseguinte, para cada iteração, serão calculados os valores de passividade para todos os usuários baseados nos valores de influência calculados na iteração anterior (linha 4). Então, na linha seguinte, os valores de influência são calculados baseados no valores de passividade obtidos na linha anterior (Equação 3). Após estes cálculos, é feita a normalização da influência e passividade de cada usuário em relação ao total de influência/passividade de todos os outros. Ao final de todas as iterações, os valores obtidos na última iteração são as estimativas de influência e passividade para cada usuário (I_m, P_m).

```

1    $I_0 \leftarrow (1, 1, \dots, 1) \in \mathfrak{R}^{|N|}$ 
2    $P_0 \leftarrow (1, 1, \dots, 1) \in \mathfrak{R}^{|N|}$ 
3   para  $i = 1$  até  $m$  faça
4     Calcule  $P_i$  utilizando  $I_{i-1}$ ;
5     Calcule  $I_i$  utilizando  $P_i$ ;
6     para  $j = 1$  até  $|N|$  faça
7        $I_j = \frac{I_j}{\sum_{k \in N} I_k}$ ;
8        $P_j = \frac{P_j}{\sum_{k \in N} P_k}$ ;
9     fim
10  fim
11  retorne  $(I_m, P_m)$ ;

```

Algoritmo 1 - Algoritmo Influence-Passivity

Os cálculos para obter os valores de influência (I) e passividade (P) para cada usuário podem ser vistos a seguir:

$$I_i \leftarrow \sum_{j:(i,j) \in E} u_{ij} P_j$$

$$P_i \leftarrow \sum_{j:(j,i) \in E} v_{ji} I_j$$

Equação 3 - Cálculo da influência (I) e Passividade (P)

Fica claro, portanto, a dependência existente entre as duas medidas. Visto que a influência de um indivíduo depende da passividade e da taxa de aceitação dos demais indivíduos, enquanto que a passividade do mesmo depende da influência e taxa de rejeição de todos os outros.

Experimentos e Resultados

Neste capítulo, serão explicados os detalhes relativos à execução dos experimentos e apresentados os resultados obtidos. Inicialmente, a rede de informações Twitter será apresentada, visto que ela foi utilizada como objeto de estudo. Em seguida, será explicado como a base foi obtida e quais são as suas características. Então, os detalhes da implementação do algoritmo IP e os resultados obtidos com a sua aplicação serão mostrados. Por final, os resultados baseados na segunda abordagem (quantidade de repasses da mensagem) serão vistos.

4.1. A rede de informações Twitter

O Twitter se define como uma rede de informações e o seu objetivo é que as pessoas tenham acesso, em tempo real, às informações que elas têm interesse. A unidade básica do Twitter é o *tweet*, este representa uma atualização feita por um usuário, uma simples mensagem de 140 caracteres. O *tweet* aparece na linha do tempo de todos os *followers* do usuário que o publicou. Os *followers* são os usuários que se inscreveram para ver as atualizações de um dado usuários, enquanto que os *friends* são os usuários que um dado indivíduo segue. Para ser criado um relacionamento entre dois usuários no Twitter, basta que um deles deseje isso. Portanto, o grafo social do Twitter é direcionado.

O *tweet* pode possuir objetos associados como URLs, imagens e vídeos. Normalmente, estas mensagens possuem conteúdos que são tão interessantes para um usuário que ele repassa esta informação para sua rede de *followers*, isto é feito através da funcionalidade chamada *retweet*. O *retweet* repassa o *tweet* original e mantém os créditos do autor, está é a forma explícita de propagar informações relevantes através da rede.

Também é possível referenciar pessoas em um *tweet*, para isso existe o recurso de *mention*. Esta também é uma forma de dar créditos a um usuário, porém Romero *et al.* (2010) obteve melhores resultados ao utilizar o *retweet* para construir o grafo de entrada do algoritmo IP. Além disso, Cha *et al.* (2010) identificou que, em livre tradução: "*retweets* representam o valor do conteúdo de um *tweet*; e *mentions* representando o valor do nome de um usuário". Portanto, *retweets* serão utilizados para o cálculo da influência dos usuários.

4.2. Base de Dados

O Twitter dispõe de três APIs¹ (*Application Programming Interface*): Search API, REST API e Streaming API. Para este experimento, a API de *streaming* foi utilizada, pois ela captura os *tweets* em tempo real.

Como o objetivo é identificar os influentes em um dado tópico, a API capturou todos os *tweets* que falavam sobre "Startups", ou seja, os que possuíam a palavra "startup" em seu conteúdo durante 12h. Para isso, foi utilizada a linguagem Ruby², um *bind* desta linguagem para a API de *streaming* (TweetStream³) e a base de dados orientada a documentos MongoDB⁴. Com estas ferramentas, foram capturados 4752 tweets, 1070 retweets e 4573 usuários, porém devido ao tempo computacional necessário ao algoritmo IP, esta base foi reduzida para 232 tweets, 78 retweets e 97 usuários. A redução foi feita ao remover os usuários que possuem poucos *tweets* e *retweets*, visto que é difícil determinar a influência destes, e foram mantidos os usuários mais ativos (Cha *et al.*, 2010).

Como a base é relativamente pequena, para comprovar a eficiência da aplicação do IP, a base foi reduzida a quase a metade de usuários com o objetivo de mostrar o desempenho do IP em uma base reduzida a metade. Para esta redução, foram removidos os usuários com menor movimentação na rede (aqueles que possuíam menor valor segundo $\frac{(\# \text{ tweets} + \# \text{ retweets})}{2}$). Esta nova base, para fins de citação, denominada **base-50** possui 167 tweets, 50 retweets e 50 usuários. A base citada anteriormente, a qual possui 97 usuários, será referenciada por **base-97**.

4.3. Aplicação do *Influence-Passivity* (IP)

O primeiro experimento contou com a aplicação do algoritmo IP na base-97, em seguida, sua aplicação foi feita na base-50 para verificar o desempenho com a drástica redução do conjunto. O algoritmo foi executado com 2 e 10 iterações, os resultados mostrados são relativos à execução com 10 iterações, pois a cada iteração o algoritmo converge para uma melhor solução. Nas seções a seguir, serão comentadas algumas particularidades da implementação e apresentados os resultados.

4.3.1. Implementação

O algoritmo IP foi implementado com a linguagem Ruby. Devido ao tempo despendido no cálculo das taxas de aceitação e rejeição de cada par de usuários, e visto que estas taxas não se modificam em uma dada base, estes valores foram calculados inicialmente e guardados em uma base de dados. Destes modo, a aplicação do algoritmo IP se tornou

¹ <https://dev.twitter.com/start>

² Linguagem de programação focada na simplicidade e produtividade (<http://www.ruby-lang.org/>)

³ <https://github.com/intridea/tweetstream>

⁴ <http://www.mongodb.org/>

mais rápido, pois não era mais necessário recalcular várias vezes as taxas citadas anteriormente para experimentar diferentes números de iterações.

4.3.2. Resultados Obtidos

Os resultados serão apresentados e cada linha possuirá uma coloração característica para indicar a coincidência ou não entre duas listas (sempre em relação às listas do mesmo aspecto de transmissão, por exemplo, influência da base-97 com influência da base-50). Isto foi feito para facilitar visualmente, a comparação entre listas. O significado de cada cor está representado na Tabela 1.

Tabela 1 - Legenda das cores utilizadas nas tabelas seguintes

Usuário coincidiu a posição nas duas listas
Usuário presente nas duas listas, mas não na mesma posição
Usuário ausente na outra lista
Usuário removido devido a redução da base

A aplicação do algoritmo IP na base-97 gerou um valor de influência e outro de passividade para cada usuário. Na Tabela 2, podem ser vistas as 10 posições dos considerados mais influentes no tema Startup, bem como uma breve descrição de cada.

Tabela 2 - Usuários mais influentes pelo IP (base-97)

Posição	Usuário	Sobre
1°	techcrunch	Mídia voltada para tecnologia
2°	startupweekend	Evento mundial de Startups
3°	sw_mexico	Evento de Startups no México
4°	robin_ahn	Organizadora de eventos voltados para Startups
	nicklongo	Empreendedor
	techberlin	Blog focado em tecnologia da internet e Startups
5°	snowinri	Estrategista de mídias sociais
6°	growvc	Plataforma de <i>Crowdfunding</i> para Startups
	swvampire	Evento de Startups na Romania
7°	guykawasaki	Escritor, investidor anjo, empreendedor muito respeitado
8°	trikro	Envolvido com muitos projetos que lidam com Startups
9°	bmull	Criador de várias Startups e mentor
	cmyworldnow	Startup tentando ganhar o <i>Breakout Startup of the Year</i> (<i>spammer</i>)

	rollpokercom	Alguém tentando vender um domínio. (<i>spammer</i>)
	ninja_startup	Blog referência em inovação, marketing e comunicação
	guy	Empreendedor
	startupaus	Evento de Startups na Austrália
	dottavi	Empreendedor, jornalista e blogueiro.
	dharmesh	Empreendedor e blogueiro (http://onstartups.com)
	getmoneyfarm	Startup Italiana
	danmartell	Empreendedor
10°	mattlauzon	Empreendedor

Como pode ser observado, as pessoas que estão entre as 10 posições estão engajadas nos movimentos de Startups de algum modo, seja por ter uma, apoiar o movimento ou ser um empresário de sucesso. Por exemplo, o TechCrunch¹, o primeiro colocado é um meio de comunicação muito respeitado e abertamente dedicado a coleta de informações sobre Startups, em sua página existe uma seção dedicada ao tema. O segundo e terceiro colocados são de organizações de eventos de Startups, o Startup Weekend². Os outros também possuem relação com o tema, como *robin_ahn* e *bmull* que são pessoas envolvidas no cenário das startups e *growvc*, o qual é um fundo para estes tipos de negócio.

Em um segundo momento, o algoritmo IP foi aplicado na base-50, a lista das 10 posições dos mais influentes pode ser vista na Tabela 3.

Tabela 3 - Usuários mais influentes pelo IP (base-50)

Posição	Usuário
1°	techcrunch
2°	startupweekend
3°	sw_mexico
4°	growvc
	swvampire
5°	robin_ahn
6°	cmyworldnow
	bmull
7°	hightechexec

¹ <http://techcrunch.com/>

² Evento mundial de Startups (<http://startupweekend.org/>)

8°	sw_denver
9°	bbjpublisher
	bbjnewsroom
10°	marcpicornell

Nota-se que as três primeiras posições coincidiram nas duas listas, mesmo com a redução do conjunto. Além disso, apenas a partir da 7ª posição foi que os usuários indicados como influentes não estavam entre as 10 posições da lista gerada pelo IP aplicado à base-97.

A lista de passividade obtida com a base-97 pode ser vista na Tabela 4 com uma pequena descrição para cada usuário. Como pode ser observado, boa parte dos usuários representam serviços voltados para startups, como coworking, marketing e programas de empreendimento. Alguns são indivíduos pouco envolvidos com o assunto que apenas repassaram uma ou outra mensagem dos usuários influentes. Também pode ser visto que a base possui diversos usuários com a passividade muito alta, tanto que nas 10 primeiras posições estão 34 usuários. Isto provavelmente se deu devido ao curto tempo de coleta (12h).

Tabela 4 - Usuários mais passivos pelo IP (base-97)

Posição	Usuário
1°	galenmoore
2°	allthingsleads
	greenspacesny
	fritsalma
	theyec
	to2
	phisab
	officedivvy
	3dsxparis
	deskaway
	michaelmorett
	socialadopter
	entrecenter
	hellopirates
	jasonhjh

	spbattle
3°	tizianapollio
4°	arepavalley
	anca-undscr-foster
5°	aaronkwhite
6°	carlaweishale
7°	socialleslie
8°	swsanjose
9°	jonrossi
10°	the-undscr-new-undscr-harley
	darionocerino
	pminordest
	lrogerwoodward
	cassandrasangel
	scarylalla
	aboutourwork
	sageeb
	mickybarisciano

Ao aplicar o algoritmo IP a base-50, a lista de passividade gerada está contida na tabela 5. Alguns usuários como *anca_foster*, *arepavalley*, *jonrossi*, *cassandrasangel* e *lrogerwoodward* mantiveram suas posições com relação à lista gerada com a base-97. Outros 11 estão em ambas as listas e 12 apareceram na nova lista, mas não na lista da base-97. Porém, provavelmente, isso ocorreu, pois a redução do conjunto foi feita ao remover os usuários menos ativos da base. Portanto, os mais passivos foram excluídos da base, isso pode ser visto porque da lista gerada na base-97, 18 usuários que estavam nas 10 primeiras posições dos mais passivos foram removidos após a redução para 50 usuários (base-50) (elementos em vermelho na Tabela 5). Portanto, novos usuários passaram a fazer parte da lista dos mais passivos da base-50 no lugar dos que foram excluídos.

Tabela 5 - Usuários mais passivos pelo IP (base-50)

Posição	Usuário
1°	scarylalla
	sageeb
	to2

	3dsxparis
	vc4africa
	swinindia
	guykawasaki
	officedivvy
	techberlin
	aboutourwork
	rollpokercom
	nicklongo
	buildmaster
	allthingsleads
	trikro
	pminordest
	fritsalma
2°	galenmoore
3°	twilli2861
4°	anca-undscr-foster
	arepavalley
5°	contextio
6°	socialleslie
7°	gweber
8°	aaronkwhite
9°	jonrossi
10°	cassandrasangel
	lrogerwoodward

4.4. Quantidade de *Retweets*

Outra métrica bastante utilizada (Cha *et al.*, 2010; Suh *et al.*, 2010; Asur *et al.*, 2011) para verificar a influência dos usuários é a quantidade de *retweets*, pois ela é um indicativo de quão espalhadas se tornam as mensagens de um usuário. Asur *et al.* (2011), por exemplo, concluiu que os *retweets* são determinantes na criação de tendências no Twitter.

Portanto, foi verificado quantos *retweets* cada usuário recebeu através do somatório de todos os *retweets* recebidos por seus *tweets*. Esta avaliação foi feita nas duas

bases: base-97 e base-50, para fazer um comparativo com os resultados obtidos pelo algoritmo IP.

4.4.1. Resultados Obtidos

Devido a vários usuários terem a mesma quantidade de *retweets*, não foi possível gerar uma lista com 10 posições, visto que vários usuários ficariam na mesma posição. Deste modo os usuários mais influentes da base-97 podem ser vistos na Tabela 6.

Tabela 6 - Usuários mais influentes pela quantidade de retweets (base-97)

Posição	Usuário	# <i>retweets</i>
1°	sw_mexico	10
2°	hightechexec	6
	techcrunch	6
3°	startupweekend	4
4°	snowinri	3

Pode-se observar que os usuários indicados como mais influentes são eventos relacionados a Startups (*sw_mexico*¹ e *startupweekend*) e um meio de comunicação muito respeitado (*techcrunch*). Isso se dá por serem usuários populares, ou seja, por possuírem uma grande audiência, porém não é levada em conta a qualidade desta audiência. Também são indicados como usuários influentes duas pessoas (*hightechexec* e *snowinri*), as quais são um empreendedor e um estrategista de marketing, respectivamente.

A aplicação da métrica de *retweets* na base reduzida (base-50) não se mostrou muito diferente e insto é bastante intuitivo, visto que foram removidos os usuários que possuíam poucos *tweets* e *retweets*. Apenas o usuário *snowinri* foi removido da lista de usuários mais influentes (Tabela 7), isso se deu porque a base de dados é pequena (*snowinri* possuía apenas 2 *retweets*).

Tabela 7 - Usuários mais influentes pela quantidade de retweets (base-50)

Posição	Usuário	# <i>retweets</i>
1°	sw_mexico	9
2°	hightechexec	6
	techcrunch	6
3°	marcpicornell	2
	startupweekend	2

¹ <http://www.startupweekenddf.com/>

sw_denver	2
twilli2861	2

Não foi gerada uma lista dos usuários mais passivos, visto que a técnica de utilizar a quantidade de *retweets* não possibilita esta análise.

4.5. Considerações finais

As duas abordagens indicaram usuários bem relacionados ao tema como os mais influentes. Além disso, as primeiras posições das listas de mais influente foram semelhantes, porém com algumas nuances, as quais se deram justamente devido às considerações feitas por cada abordagem.

Pode-se notar que o *sw_mexico* obteve mais *retweets* do que todos os outros usuários, porém o algoritmo IP indicou o *techcrunch* e o *startupweekend* como mais influentes do que o primeiro. Isso ocorreu, pois o *sw_mexico* teve mais *retweets* devido a proximidade do evento, o qual se iniciou no dia 11 de novembro (apenas 9 dias após a coleta). O IP também levou em conta a qualidade da audiência, portanto a audiência do *techcrunch* e *startupweekend* se mostrou mais fiel e por isso estes usuários ocuparam o topo da lista.

Esta diferença de abordagem também deu à usuária *robin_ahn* um posto mais importante do que a *hightechexec* e *snowinri*, ao contrário da segunda abordagem que classificou *hightechexec* como um dos mais influentes. Novamente, isto ocorreu devido ao fato de que a quantidade de *retweets* por si só não é um indicativo completo (Romero *et al.*, 2010).

Conclusão

É de grande importância o entendimento da formação de opinião em ambientes virtuais, pois estes são reflexos da sociedade real. Portanto, este entendimento ajudará teorias que há décadas não puderam ser efetivamente comprovadas devido a dificuldade de se monitorar o fenômeno.

Este trabalho verificou a eficiência do algoritmo IP, proposto por Romero *et al.* (2010) na identificação dos usuários mais influentes e mais passivos de um grafo que só possui informações relacionadas a um dado tópico (Startups). Também foi feita uma classificação baseada na métrica de *retweets*, a qual se mostrou menos efetiva e corroborou com as conclusões de Romero *et al.* (2010) em sua pesquisa.

5.1. Principais Contribuições

As principais contribuições deste trabalho se deram a respeito da identificação dos usuários mais influentes e passivos em um tópico específico no ambiente de rede social. Esta contribuição é bastante importante em vários campos como: marketing, sociologia e ciências políticas, pois o modo como se dá a transmissão da informação e a formação de opinião, mesmo após anos de estudo, ainda é bastante nebulosa.

A identificação destas pessoas colabora para o entendimento do fenômeno social de transmissão da informação, pois uma rede social é um ambiente bem próximo ao real. Visto que nelas são encontrados relacionamentos, discussões e informações que não são meramente virtuais, são de fato reais. O entendimento do fluxo da informação e como isso leva à formação de opinião fortalecerá os campos citados anteriormente e tornará mais efetivas atividades como campanhas de marketing viral, campanhas políticas e até identificar problemas em uma organização.

5.2. Trabalhos Futuros

É interessante aplicar o modelo à uma base mais densa para validar a aplicação do algoritmo IP em grafos de diferentes escalas. Também é importante verificar a possibilidade de melhorar a performance do mesmo, visto que ele se mostrou relativamente lento mesmo em bases pequenas. Além disso, levar em conta o comportamento durante diferentes espaços de tempo para identificar os usuários mais passivos e mais ativos com o objetivo de diminuir o efeito de influentes ocasionais no resultado final também é bastante relevante, visto que este efeito ocorreu neste trabalho. Por fim, é de suma im-

portância aplicar este modelo de análise, baseado na passividade dos usuários, em outras redes sociais como Facebook e LinkedIn¹.

¹ <http://www.linkedin.com/>

Referências Bibliográficas

- AGARWAL, N.; LIU, H.; TANG, L.; YU, P. S. **Identifying the influential bloggers in a community**. Proceedings of the international conference on Web search and web data mining. *Anais...*, WSDM '08. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1341531.1341559>>. Acesso em: 22 nov. 2011. , 2008
- ARAL, S.; MUCHNIK, L.; SUNDARARAJAN, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. **Proceedings of the National Academy of Sciences**, v. 106, n. 51, p. 21544 -21549, 22 dez 2009.
- ASUR, S.; HUBERMAN, B. A.; SZABO, G.; WANG, C. Trends in Social Media : Persistence and Decay. **arXiv:1102.1402**, 7 fev 2011.
- BAKSHY, E.; HOFMAN, J. M.; MASON, W. A.; WATTS, D. J. **Everyone's an influencer: quantifying influence on twitter**. Proceedings of the fourth ACM international conference on Web search and data mining. *Anais...*, WSDM '11. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1935826.1935845>>. Acesso em: 19 nov. 2011. , 2011
- BAKSHY, E.; KARRER, B.; ADAMIC, L. A. **Social influence and the diffusion of user-created content**. Proceedings of the 10th ACM conference on Electronic commerce. *Anais...*, EC '09. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1566374.1566421>>. Acesso em: 22 nov. 2011. , 2009
- BURT, R. S. The Social Capital of Opinion Leaders. **Annals of the American Academy of Political and Social Science**, v. 566, p. 37-54, 1 nov 1999.
- CHA, M.; HADDADI, H.; BENEVENUTO, F.; GUMMADI, K. P. Measuring user influence in Twitter: The million follower fallacy. **IN ICWSM '10: PROCEEDINGS OF INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL**, 2010.
- DOMINGOS, P.; RICHARDSON, M. **Mining the network value of customers**. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. *Anais...*, KDD '01. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/502512.502525>>. Acesso em: 22 nov. 2011. , 2001
- GALUBA, W.; ABERER, K.; CHAKRABORTY, D.; DESPOTOVIC, Z.; KELLERER, W. **Outtweeting the twitterers - predicting information cascades in microblogs**. Proceedings of the 3rd conference on Online social networks. *Anais...*, WOSN'10. Berkeley, CA, USA: USENIX Association. Disponível em: <<http://dl.acm.org/citation.cfm?id=1863190.1863193>>. Acesso em: 23 nov. 2011. , 2010

- HAVELIWALA, T. H. Topic-sensitive PageRank: a context-sensitive ranking algorithm for Web search. **IEEE Transactions on Knowledge and Data Engineering**, v. 15, n. 4, p. 784- 796, ago 2003.
- HUBERMAN, B.; ROMERO, D.; WU, F. Social networks that matter: Twitter under the microscope. **ArXiv e-prints**, 2008.
- JANSEN, B. J.; ZHANG, M.; SOBEL, K.; CHOWDURY, A. **Micro-blogging as online word of mouth branding**. Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. **Anais...**, CHI EA '09. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1520340.1520584>>. Acesso em: 20 nov. 2011. , 2009
- JANSEN, B.; ZHANG, M.; SOBEL, K.; CHOWDURY, A. Twitter power: Tweets as electronic word of mouth. **J. Am. Soc. Inf. Sci.**, v. 60, n. 11, p. 2169-2188, 2009.
- KATZ, E.; LAZARFELD, P. F. **Personal influence: the part played by people in the flow of mass communications**. Glencoe: Free Press, 1955.
- KEMPE, D.; KLEINBERG, J.; TARDOS, É. Influential Nodes in a Diffusion Model for Social Networks. In: CAIRES, L.; ITALIANO, G. F.; MONTEIRO, L.; PALAMIDESSI, C.; YUNG, M. (Eds.). **Automata, Languages and Programming**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. v. 3580p. 1127-1138.
- KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. **J. ACM**, v. 46, n. 5, p. 604–632, set 1999.
- KNOKE, D.; KUKLINSKI, J. H. **Network analysis**. Sage: Sage University Paper Series on Quantitative Applications in the Social Sciences, 1982. v. 7
- LAZARFELD, P.; BERELSON, B.; GAUDET, H. **The people's choice how the voter makes up his mind in a presidential campaign**. 3. ed. New York: Columbia University Press, 1968.
- MEDEIROS, V. B. DE O. **Análise comparativa de redes de colaboração de eventos científicos brasileiros**. Recife: Universidade Federal de Pernambuco, 2010.
- MILSTEIN, S.; LORICA, B.; MAGOULAS, R. *et al.* **Twitter and the Micro-Messaging Revolution: Communication, Connections, and Immediacy--140 Characters at a Time**. [S.l.]: O'Reilly Media / Radar, 2008.
- OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. **Journal of Information Science**, v. 28, n. 6, p. 441 -453, 1 dez 2002.
- PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. The PageRank Citation Ranking: Bringing Order to the Web. **STANFORD INFOLAB**, p. 17, 1999.
- ROMERO, D. M.; GALUBA, W.; ASUR, S.; HUBERMAN, B. A. Influence and Passivity in Social Media. **arXiv:1008.1253**, 6 ago 2010.
- RYMASZEWSKI, M. **Second life: the official guide**. [S.l.]: Wiley-Interscience, 2007.
- SILVA, N. B. DA. **Recomendação de relacionamentos em redes sociais baseada em grafos**. Recife: Universidade Federal de Pernambuco, 2010.

- SUH, B.; HONG, L.; PIROLI, P.; CHI, E. H. **Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network**. . [S.l.]: IEEE. Disponível em: <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5590452>. Acesso em: 1 dez. 2011. , ago 2010
- SUN, E.; ROSENN, I.; MARLOW, C.; LENTO, T. **Gesundheit! Modeling Contagion through Facebook News Feed**. International AAAI Conference on Weblogs and Social Media. **Anais...** [S.l.: s.n.]. Disponível em: <http://www.stanford.edu/~esun/ICWSM09_ESun.pdf>. Acesso em: 22 nov. 2011. , 2009
- Twitter**. Disponível em: <<http://twitter.com/about>>. Acesso em: 19 nov. 2011.
- WATTS, D. J.; DODDS, P. S. Influentials, Networks, and Public Opinion Formation. **Journal of Consumer Research**, v. 34, n. 4, p. 441-458, 1 dez 2007.
- WENG, J.; LIM, E.-P.; JIANG, J.; HE, Q. **TwitterRank: finding topic-sensitive influential twitterers**. Proceedings of the third ACM international conference on Web search and data mining. **Anais...**, WSDM '10. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1718487.1718520>>. Acesso em: 20 nov. 2011. , 2010