Computers in Human Behavior 52 (2015) 184–189

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

# Measuring influence in online social network based on the user-content bipartite graph

Zhiguo Zhu [a,b,*], Jingqin Su [a], Liping Kong [c]

[a] Faculty of Management and Economics, Dalian University of Technology, 116023 Dalian, PR China
[b] School of Management Science and Engineering, Dongbei University of Finance and Economics, 116025 Dalian, PR China
[c] Vocational and Technical College, Dongbei University of Finance and Economics, 116025 Dalian, PR China

ARTICLE INFO

*Article history:*

*Keywords:*
Online social network
Bipartite directed graph
Influence measurement
Markov model

ABSTRACT

With the rising of online social networks, influence has been a complex and subtle force to govern users' behaviors and relationship formation. Therefore, how to precisely identify and measure influence has been a hot research direction. Differentiating from existing researches, we are devoted to combining the status of users in the network and the contents generated from these users to synthetically measure the influence diffusion. In this paper, we firstly proposed a directed user-content bipartite graph model. Next, an iterative algorithm is designed to compute two scores: the users' Influence and boards' Reach. Finally, we conduct extensive experiments on the dataset extracted from the online community Pinterest. The experimental results verify our proposed model can discover most influential users and popular broads effectively and can also be expected to benefit various applications, e.g., viral marketing, personal recommendation, information retrieval, etc.



## 1. Introduction

The idea of studying a person's behavior in the context of his/her social connections is quite old as can be seen in Stanley Milligram's experiments which lead him to conclude that the average shortest path length between any two people in the world is about six (Milgram, 1967). However, nowadays online social networks have rapidly become very important hubs of social activity and conduits of information. Popular social sites such as Facebook, Pinterest and Twitter have undergone explosive growth and are turning into community spaces, where users interact with their friends and acquaintances. With the numbers of active users on these sites numbering in the millions or even tens of millions, how to identify and measure the influential users and topics becomes an important problem with applications in marketing (Kempe, Kleinberg, & Tardos, 2003), information dissemination (Gruhl & Liben-nowell, 2004; Leskovec et al., 2007), social relation visualization (Kim, Ji, & Park, 2014), and expertise discovery (Davitz, Yu, Basu, Gutelius, & Harris, 2007).

Given this widespread generation and consumption of content in online social networks, it is desirable to target one's messages to highly connected people who will propagate them further in the social network. In spite of the seemingly chaotic fashion with which all these interactions take place, certain topics manage to get an inordinate amount of attention, thus bubbling to the top in terms of popularity and contributing to new trends and to the public agenda of the community. There is considerable consensus on the fact that two aspects of information transmission seem to be important in determining the influence in social network.

One aspect is the popularity and status of given users in these social networks, which can be measured by the level of attention they receive in the form of followers who create links to their accounts to automatically receive the content they generate. This can also be generally viewed as the degree of user (Bonchi, Castillo, Gionis, & Jaimes, 2011; Mislove, 2009; Valente, 2010). The other aspect is the influence of contents that these users wield, which is determined by the actual propagation of these contents through the network. This influence is determined by many factors, such as the novelty and quality of these contents, and the frequency at which they generate these contents. Therefore, we are interested in combining the status of users in the network and the contents generated from these users to synthetically measure the influence diffusion existed in online social network.

In this paper, we employ the user-content bipartite graph to construct the model and then develop an iterative algorithm to quantify the influence in the network. With the example of Pinterest, which is one of the fastest growing social networks on

* Corresponding author at: Faculty of Management and Economics, Dalian University of Technology, No.2 Linggong Road, Ganjingzi District, Dalian, PR China. Mobile: +86 1594088193.

*E-mail address:* zhuzg0628@126.com (Z. Zhu).

http://dx.doi.org/10.1016/j.chb.2015.04.072
0747-5632/

the Internet and is becoming the focus of advertising companies and brands eager to exploit this vast new medium, we try to understand how the influence is determined by analyzing the propagation of 'pins' on Pinterest. In a word, our influence measure utilizes both the structural properties of the network and the contents the users published.

The rest of this article is organized as follows. The literature background and related work are investigated in Section 2. Section 3 formalizes the problem of measuring influence in mathematical terms and then proposes a directed bipartite graph model. Next, we design an iterative algorithm for the model to discover influential users and popular broads in Section 4. Section 5 presents experimental results that validate the effectiveness of our methodology. Last Section 6 concludes the paper and discusses the possible directions in the future research.

## 2. Related work

Since the advent of online social networks, especially in the past decade or so, there has been a lot of interest in measuring influence and modeling information flows in this new platform. Kleinberg et al. have studied the problem for quite some time and many of the results and concepts from their research are explained in Crandall, Cosley, Huttenlocher, Kleinberg, and Suri (2008), Backstrom, Huttenlocher, Kleinberg, and Lan (2006). Using a dataset of Wikipedia edits, Kleinberg et al. showed that influence (action in a user is triggered by one of his/her friends' recent actions) and selection (people prefer friends who share similar interests and hence they perform similar actions) often mutually enhance each other. Holme and Newman (2006) later improved upon this Kleinberg's study and suggested a probabilistic model for modeling this interaction. Another study that addresses the problem of interplay of selection and influence is presented in Christakis and Fowler (2007) where Fowler and Christakis study the spread of obesity in a network of around 12,000 people over a period of three decades. Kumar, Mahdian, and Anagnostopoulos (2008) have devised simple tests to identify social influence at play in an online community by tests such as the Shuffle Test and the Edge reversal test (also used by Fowler and Christakis).

While it assumes that influence exists as a real phenomenon in online social network, questions have been raised on whether there is evidence of genuine influence in real social network data. Anagnostopoulos, Kumar, and Mahdian (2008) have developed techniques for showing that influence may not be genuine: while there is substantial social correlation in tagging behavior, it cannot be attributed to influence. Another work highlighting the importance of separating influence-based contagion from homophily-driven diffusion is Aral (2010), where it is observed that the former can be overestimated if not measured correctly. Moreover, the strength of the different factors affecting the propagation of a piece of information may vary depending on what type of information (e.g., news, or discussion topic) is being propagated (Aral, Muchnik, & Sundararajan, 2009).

On the other hand, many researchers have designed some algorithms for quantitatively calculating the influence existed in online social networks. "Centrality" (Newman, 2010) is a fairly well studied concept in the context of networks and most measures of influence are more refined versions of one or more of these centrality measures. "PageRank", developed by Page and Brin (1998), has been one of the most influential measures developed in this area. Although PageRank is primarily used to rank sources of information in information networks (such as ranking web pages on the WWW), the algorithm has inspired various modifications to be applied specifically to social networks such as SimRank (Widom & Glen, 2002), Topic sensitive PageRank (Haveliwala, 2003) and TwitterRank (He, Weng, Jiang, & Lim, 2010). Jiawei Han's group (Han, Ming, & Danilevsky, 2011; Han, Sun, & Yu, 2009; Sun, Han, Zhao, Yin, & Cheng, 2009) has published several papers in the area of ranking nodes in heterogeneous networks, which after some adaptations, might be used to measure influence. Tang, Sun, Wang, and Yang (2009) introduce the novel problem of topic-based social influence analysis. They propose a topical-affinity propagation approach to describe the problem. Zhu (2013) build a model of information diffusion oriented for viral marketing and propose a dynamic algorithm of discovering the influential users in the process of information diffusion. Another popular class of algorithms has been improvements on the "HITS" algorithm proposed around 1998 by Kleinberg (1998). Kleinberg himself proposed a possible application of HITS to identify "important" people in a social network (Kleinberg, 1999). He suggested that HITS can be used to calculate "standing, impact or influence" of a node in such a network. This idea was developed further by Romero, Galuba, Asur, and Huberman (2011a) who came up with the IP algorithm and tested it on a Twitter dataset.

In this paper, we consider that the influence of a user thus depends not only on the size of the influenced audiences, but also on their generated contents. So, our will proposed a model based on an iterative algorithm to quantify the influence in the network, which is inspired by HITS (Kleinberg, 1998) and the Influence Passivity algorithm (Romero et al., 2011a). However, there is an important different thought: the prior algorithms treated networks as homogeneous whereas we consider heterogeneous networks.

## 3. Problem statement and model construction

### 3.1. Background about the community of Pinterest

Pinterest is a pinboard-style content sharing platform that allows users to exhibit collections of images or videos. To better present the proposed model, we briefly describe key terminologies in Pinterest below.

- Pin/Repin: Each image/video is called as a pin, and the act of posting a pin is referred to as pinning. If a posted pin is shared by someone, the shared pin is called as a repin. Users who posts and shares (i.e., repins) a pin are the original pinner and repinner, respectively.
- Like/Comment: Similar to Facebook, a user can push a like button for a pin that she likes and leave a comment on a pin.
- Board/Category: A board is a collection of pins organized by a user. Each board belongs to one of the categories in Pinterest.
- Following/Follower: Like many social platforms, the relation between two users in Pinterest is not symmetric. The fact that user A follows user B does not necessarily mean B follows A. If A follows B, A can see the updated news (e.g., the act of posting a new pin) of B.

### 3.2. A bipartite directed graph model

In this section, we formalize our problem of measuring influence in mathematical terms. We define precisely what do we mean by various terms, what data we have and what do we attempt to calculate by our algorithm.

We propose a bipartite directed graph model $G\ (V, E, W)$ based on the dataset described above with the following properties:

(1) The vertex set $V$ has two types of vertices representing users and boards (representing contents) respectively. Mathematically $V = U \cup B$. $U = \{u_1, u_2, \ldots, u_m\}$ is the set of vertices representing users, in which $m$ is the number of users in this online community and $B = \{b_1, b_2, \ldots, b_n\}$ is

the set of vertices representing boards, in which $n$ is the number of broads. In this case, these broads can also be viewed as the contents users published in the Pinterest.

(2) The edge set $E$ has two types of edges – ones going from a user to a board and the other ones going from a board to a user. Mathematically, we can get $E = E_{U \to B} \cup E_{B \to U}$.

(3) The weights $W$ can be defined for two kinds of edges with different directions as follows:

(i) For any edge $e_{u_i \to b_j} = (u_i, b_j) \in E_{U \to B}$, its weight $w(e_{u_i \to b_j})$ can be defined as:

$$w(e_{u_i \to b_j}) = \left( \frac{success(u_i, b_j)}{\sum_{b_k \in B \text{ and } u_i \text{ follows } b_k} success(u_i, b_k)} \right) \tag{1}$$

(ii) For any edge $e'_{b_r \to u_s} = (b_r, u_s) \in E_{B \to U}$, its weight $w(e'_{b_r \to u_s})$ can be defined as:

$$w(e'_{b_r \to u_s}) = \left( \frac{number \;\; of \;\; pins(b_r)}{\sum_{b_k \in B \text{ and } b_k \text{ created by } u_s} number \;\; of \;\; pins(b_k)} \right) \tag{2}$$

Thus, our model $G = (V, E, W)$ forms a bipartite directed graph, with the vertex being divided into two independent sets ($U$ and $B$), which is shown in Fig. 1. Note that a user $u_i$ following another user $u_j$ can be incorporated into this model as $u_i$ following all boards created by $u_j$.

### 3.3. Some definitions

Next, we will give some definitions related to the proposed model.

**Definition 1** (*Success*). We define a quantity called '*Success*' of a board with respect to a user as follows:

$$Success(u_i, b_j) = \left( \frac{number \;\; of \;\; repins \;\; from \;\; b_j \;\; by \;\; u_i}{number \;\; of \;\; pins \;\; on \;\; b_j} \right) \tag{3}$$

Eq. (3) denotes how successful a board is in capturing a user's attention. The maximum number of pins a user can re-pin from this board is equal to the total number of pins that exist on this board. So the equation denotes, as a fraction, how successful the board really is, in comparison to this the maximum pins this broad could have been repined.

**Definition 2** (*Engagement*). We define the engagement of a user with a particular board as follows:

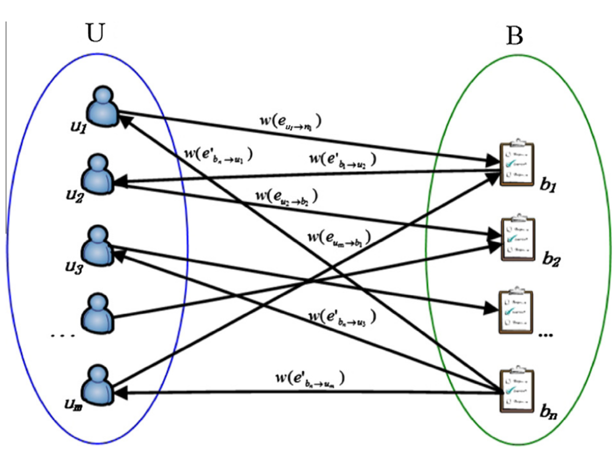


**Fig. 1.** The example of bipartite directed weighted graph.

$$Engagement(u_i, b_j) = \left( \frac{Success(u_i, b_j)}{\sum_{b_k \in B \text{ and } u_i \text{ follows } b_k} Success(u_i, b_k)} \right) \tag{4}$$

This equation above denotes how much does a user engage with a particular board in comparison to all other boards that he/she follows. There are a number of boards eyeing for a user's share of attention. A user pays attention to a board by re-pinning the pins from that board. In the base case, if the user was following just one board, this board will get all the user's attention. However, as the user starts following more and more boards, his/her attention is divided among these various boards. Eq. (4) just denotes, as a fraction, how much does a particular board succeed in capturing a user's attention in comparison to other boards that it is competing with. The reader might have noticed that $Engagement(u_i, b_j) = w(e_{u_i \to b_j})$ for $e_{u_i \to b_j}$.

**Definition 3** (*Activity coefficient*). Activity coefficient is defined as:

$$Act(b_k, u_s) = \left( \frac{number \;\; of \;\; pins(b_r)}{\sum_{b_k \in B \text{ and } b_k \text{ created by } u_s} number \;\; of \;\; pins(b_k)} \right) \tag{5}$$

Eq. (5) simply represents the probability that the creator of a board will pin an image to this board, given that he/she pins an image. Intuitively, this quantity tells us how active the creator of the board is on this board as compared to all other boards. Similarly, $Act(b_k, u_s) = w(e'_{b_k \to u_s})$ for $e'_{b_k \to u_s}$.

**Definition 4 and 5** (*Reach and Influence*). These are the most important definitions for our algorithm. We define the '*Reach*' of a board and the influence '*I*' of a user as follows:

$$Reach(b_j) = \sum_{u_i \in U \text{ and } u_i \text{ follows } b_j} Engage(u_i, b_j) * I(u_i) \tag{6}$$

$$I(u_s) = \sum_{b_k \in B \text{ and } b_k \text{ created by } u_s} Act(b_k, u_s) * Reach(b_j) \tag{7}$$

These definitions have been formulated to mathematically capture the following idea: "successful boards draw engagement from influential users. Influential users also create successful boards."

## 4. The algorithm description and theoretical analysis

### 4.1. An iterative algorithm for proposed model

Given the input directed weighted graph $G = (V, E, W)$ as described in the previous section, we design an algorithm to compute iteratively the users' *Influence* scores and boards' *Reach* scores. The description of the algorithm is shown as follow. Before describing the steps of the algorithm, we summarize here the four main ideas discussed in the previous section:

(1) A user's *Influence* score will depend on the number of boards created by this user, as well as each board's *Reach*.
(2) How active the user is in posting pins to a board he/she created will affect how much of the board's success transfers into the user' influence (as opposed to other boards created by the same user).
(3) A board's *Reach* will depend on the number of followers as well as the influence of each of the followers.
(4) How engaged the follower of a board is, with respect to all other boards followed by him/her will affect how much of the user's influence transfers into the board's reach. Our Eqs. (6) and (7) exactly capture these intuitions as explained earlier.

| |
|---|
| **The algorithm**: computing the users' *influence* and boards' *Reach* |
| **1.** $I^{(0)} \leftarrow (1, \quad 1, \quad \ldots, \quad 1) \in R^{\lvert U \rvert}$ |
| **2.** $Reach^{(0)} \leftarrow (1, \quad 1, \quad \ldots, \quad 1) \in R^{\lvert B \rvert}$ |
| (thus, *Reach* and *I* are vectors of dimension (1×\|U\|) and (1×\|B\|) respectively where $I^i(k)$ denotes influence score of user $u_k$ after '*i*' iterations. Similarly for *Reach* vector). |
| **3.** Compute $engagement(u_i, b_j)$ and $Act(b_k, u_s)$ for all (user, board) pairs using Eqs. (4) and (5). |
| **4.** For *i=1* to *m* |
| *Update* $Reach^{(i)}$ and $I^{(i)}$ scores |
| $Reach^{(i)}(k) = \sum_{u_j \in U \text{ and } u_j \text{ follows } b_k} engage(u_j, b_k) * I^{(i-1)}(u_j)$ |
| (for all *k=1* to \|*B*\|) |
| $I^{(i)}(k) = \sum_{b_s \in B \text{ and } b_s \text{ created by } u_k} Act(b_s, u_k) * Reach^{(i)}(b_s)$ |
| (for all *k=1* to \|*U*\|) |
| Now normalize vectors: |
| For *j*=1 to \|*U*\| |
| $I(j) = \frac{I(j)}{\lvert U \rvert + \lvert B \rvert}$ |
| End |
| For j=1 to \|*B*\| |
| $Reach(j) = \frac{Reach(j)}{\lvert U \rvert + \lvert B \rvert}$ |
| End |
| End |
| **5.** Return($Reach^{(m)}$ and $I^{(m)}$) |

### *4.2. Theoretical analysis for the algorithm*

In this algorithm, we run the outer loop for '*m*' times. What is this '*m*'? How its value is decided and how can we be sure that the values of *I* and *Reach* vectors will converge eventually? We explore a few concepts of Markov chains from Mathematics. Our iterative algorithm can be viewed as a random walk on the input graph $G = (W, E, V)$ and so, we can apply some results about Markov chains to prove that values for *I* and *Reach* vectors converge.

**Theorem.** *A finite state space Markov chain has a unique stationary distribution if it is irreducible and aperiodic.*

This is an important result in the analysis of finite Markov chains. We will use this result to prove that this algorithm indeed converges. It is easy to see why our input graph is a finite state Markov chain (since the number of nodes is finite). A chain *P* is irreducible if for any two states *x*, *y* there is an integer *t* such that $P^t(x, y) > 0$ i.e. it is possible to reach any state from any other state using the probability transitions as given in matrix *P* (in our case, the weights on the edges can be viewed these probability transitions). Thus, the given graph *G* should be strongly connected for it to be irreducible. This clearly holds in our case since during our data collection phase, we explored only a strongly connected component of a graph (even in case this did not hold, we could easily have created a dummy node and connected all nodes to this dummy node, thus ensuring a path from every node *x* to every other node *y*).The second condition is that *G* needs to be aperiodic. Formally, we define a set $T(x) = \{t > 1 : P^t(x, x) > 0\}$, which is the set of times when it is possible for a chain to return to it starting position '*x*'. Period of '*x*' is the GCD of all elements in $T(x)$. The chain will be aperiodic if all states have period 1 otherwise it is periodic.

After some examination, we realize that our input instance *G* is indeed periodic (with period = 2). In fact, all bipartite graphs are periodic. Fortunately, a simple modification can repair this periodicity problem. Given out input graph *G* (with the respective transition matrix *P*), we define $Q = \frac{P+I}{2}$ where *I* is an identity matrix of size equal to size of *P*. Intuitively, we can imagine *Q* as a "lazy version" of *P*: at each step, flip a coin. If it is head, go to another node using transition probabilities of *P*, else stay at the same node. Since $Q(x, x) > 0$ for all *x*, *Q* is aperiodic.

Thus, with this simple modification, we are ensured that our algorithm is guaranteed to converge to the correct answer.

## 5. Experiments and evaluations

### *5.1. An overview of dataset*

In this section, we describe our experimental results obtained from running the algorithm on a real dataset extracted from Pinterest community. Our initial dataset consisted of a community of 4013 users, which together had created 88793 boards. After cleaning and filtering the data (users with <5 followers or <2 boards created, boards with <10 pins), we were left with around 3897 users which together had created 79121 boards (thus an average of around 20 boards per user). The average number of followers per board was 112 (of course not all followers of all boards were included in our dataset).

The behaviors of users in our dataset in terms of the numbers of pins, boards and categories are investigated in Fig. 2. A rather surprising statistic is that Pinterest drives a lot more user engagement with more than 55% of users having more than 100 pins, as shown in Fig. 2a. It is just our major motivations to choose a Pinterest data set, since a lot more information diffusion happens within the Pinterest community as opposed to other social networks. Fig. 2b shows around 55% of users have fewer than 10 boards, while top 1% of users have more than 100 boards. Fig. 2c shows the number of categories on which the user has posted pins. While 23% of users have only one category, top 10% of users are interested in more than 10 categories.

### *5.2. Some experiment results*

#### *5.2.1. The most influential users*

We successfully detect ten most influential users in the dataset by utilizing our algorithm, as show in Table 1. In Table 1, the user accounts 'janew' and 'ben' with '*' are two of the co-founders of Pinterest and the symbol ' + ' represents the most popular user on Pinterest (by number of followers). Furthermore, from Table 1 it is easy to see that while generally, influential users have a very high follower count (compared to the average number of followers for a user which is 112 for our dataset), there is no specific correlation between the ranking of influence and the follower count. For example, as you can see, user 'Perfect Palette' with about 1/5 follower count of user 'maria_mcdonald', has a higher influence score. These results are in line with many previous results from references (Romero, Galuba, Asur, & Huberman, 2011) and Widom & Glen, 2002.

#### *5.2.2. Boards with the most "Reach"*

Table 2 shows the list of ten boards with highest reach as detected by our algorithm. Not surprisingly, most of them were started users who had high influence scores. Furthermore, we observe that many of the boards ranked in top 10 boards with

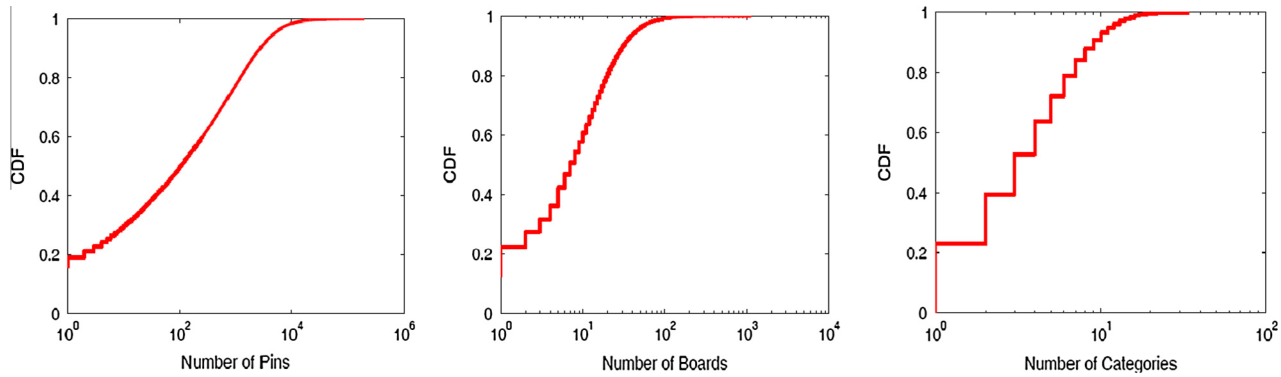


**Fig. 2.** The cumulative distribution functions (CDF) of the number of (a) pins, (b) boards and (c) categories for each user.

**Table 1**
The list of 10 most influential users.

| Rank | User | Boards with high reach | Number of followers |
|---|---|---|---|
| 1 | janew* | Delicious, Happy, Easy hacks | 2912144 |
| 2 | chrisem | Words to Live by, Sweets and Treats | 982303 |
| 3 | Perfect Palette+ | Color Palette Library | 254476 |
| 4 | levato | Funny, Sarcasm | 661462 |
| 5 | maia_mcdonald | Glorious Food, Food- Dessert and Bakery | 1393968 |
| 6 | Mashable | Tech and Gadgets, Social Media | 25618 |
| 7 | ben* | Preparations for baby | 883644 |
| 8 | therealmurphy | Spaces | 869102 |
| 9 | eatsleepwear | Fashion, Foodie | 704470 |
| 10 | chicuolino | Tutorials, Recipes | 139694 |

**Table 2**
Top 10 boards with highest "Reach".

| Rank | Board | Influential followers |
|---|---|---|
| 1 | Delicious | chrisem, danielhunley, firsten, happymundane |
| 2 | Words I live by | janew, levato, stylemepretty |
| 3 | Color Palette Library | chaserskaysers, janew |
| 4 | Funny | freshome, jcards, maryannrizzo |
| 5 | Food- Dessert and Bakery | chrisem, babble, slavingia, wholefoods |
| 6 | Fashion | maia_mcdonald, kaleb_willis, efreedman |
| 7 | Tutorials | levato, everythingfab, dangerorange |
| 8 | We're used to reusing | danielhurley, poppytalk, stacyofksw |
| 9 | Adventures in drink | flowercharlotte, chicoulino |
| 10 | Punboard | janew, bhg, insomniac |

highest reach have followers from among the users ranked in top 10 influential users shown in Table 1.

### 5.2.3. Topic specific influencers

Typically, boards are created by users centered on specific topic as can be seen from some of the boards in the previous table. For example, 'funny' is a board with pins which are humorous, 'Food-Dessert and Bakery' is a board with pins containing photographs and recipes for specific types of food and so on. From doing a preliminary topic classification using names of boards, we came up with topic-specific influencers, as shown in Table 3.

**Table 3**
The list of topic-specific influencers.

| No. | Category | Influential users |
|---|---|---|
| 1 | Food | janew, wholefoods |
| 2 | Fashion | Daniel, chicoulino |
| 3 | Lifestyle | jessicasimko, KayH |
| 4 | Tech/Gadgets | Mashable, willywei |
| 5 | Travel | BBC Travel, Travel Inspiration |
| 6 | Wedding | alliearhart, levato |
| 7 | Cars | danielhunley |
| 8 | Infographics | bestinfographics, weddinggraphics |

## 6. Conclusion and future work

It would be fair to say that detecting and measuring influence in a network of people is a very hard and important work before the advent of online social networks. But nowadays we have data to carry out studies and test hypotheses on a large scale. To differentiate from existing studies, in this paper, we employ the user-content bipartite graph to construct the model and then develop an iterative algorithm to quantify the influence of all the users in the network. With the experiments on dataset from Pinterest, we verify the influence of users is determined by both the structural properties of the network as well as the contents the users published. Furthermore, as a useful byproduct of the algorithm, we get the 'Reach' of various boards. From the point of view of the users and the social business companies, the achievements of this work can be valuable in such applications as:

### 6.1. Topic-Based and group-based influence forecast

The proposed algorithm can be run on a subgraph of the full graph or on the subset of the user activity data. For example, if only users creating and following boards about a certain topic are part of the graph, the algorithm can forecast the most influential users in that topic. Such topic specific influence and group specific influence are invaluable to marketing and advertising agencies that are looking to take advantage of viral marketing to fast improve brand visibility based on social network platforms.

### 6.2. Content ranking

Through the "Reach" calculation of various boards, we get to know how successful a board is at capturing user attention. Boards with more "Reach" tend to be more engaging and thus, can be used to rank user generated content. Thus, given a large enough data set, it might also be a good idea to generate topic based ranking of the boards, which help users quickly find the board with greatest reach on a certain topic.

### 6.3. Personalized recommendation services

By measuring the quantities of success and engagement of a board with respect to a user, we can determine what categories of content a user prefers, which can be utilized for providing personalized recommendation service by the social networks's managers.

But any measure of influence is necessarily a subjective one and depends heavily on the data we are dealing with and our end goal. In the future, we would like to extend our work in two major areas:

(1) Extend a generalized measure of influence to datasets other than Pinterest. In the future we would like to generalize our measure of influence and test it with datasets outside of Pinterest. It would be interesting to compare influencers across networks and see if the medium/user interface plays a role in the spread of influence.
(2) Explore further the idea of "topic-specific" influence. This is an area that is very useful in the commercial aspect of studying influence-brands/companies are always looking for influencers in their respective industry or category to effectively advertise their new products, offers, etc.

## Acknowledgements

This work is supported and funded by the Program for Liaoning Excellent Talents in University in 2014 (LNET, No. WJQ2014040), the National Natural Science Foundation of China (No. 71301021), the Humanities and Social Sciences Youth Project of Ministry of Education P.R.C. (Nos. 13YJC790061 and 12YJCZH321) and China Postdoctoral Science Foundation (No. 2015M570249) funded project. The author also appreciates the valuable comments and suggestions of the editor and anonymous reviewers.

## References

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). "Influence and correlation in social networks". In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*. LAS VEGAS, USA (August 24–27).
Aral, S. (2010). Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 217–223.
Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences, 106*, 21544–21549. 51.
Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). "Group formation in large social networks: Membership, growth, and evolution". *Proceeding of SIGKDD*. Philadelphia, USA (August 20–23).
Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011). "Social network analysis and mining for business applications", *ACM Transactions on Intelligent Systems and Technology*, 2(3).
Christakis, N., & Fowler, J. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine, 357*(4), 3700.
Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). "Feedback Effects between similarity and social influence in online communities". *Proceeding of SIGKDD*. LAS VEGAS, USA (August 24–27).
Davitz, J., Yu, J., Basu, S., Gutelius, D., & Harris, A. (2007). "ilink: Search and routing in social networks". In *KDD '07: Proc. of the 13th ACM SIGKDD int. conf. on knowledge discovery and data* (pp. 931–940). San Jose, USA.
Gruhl, D., & Liben-nowell, D. (2004). "Information diffusion through blogspace". In *Proc. int. world wide web conference (WWW)* (pp. 491–501).
Han, J., Sun, Y., & Yu, Y. (2009). Ranking-based clustering of heterogeneous networks with star-network schema. In *Proceedings of the 15th ACM SIGKDD*. Paris, France (June 28–July 1st).
Han, J., Ming, J., & Danilevsky, M. (2011). "Ranking based classification of heterogeneous Information Networks". In *Proceedings of the 17th ACM SIGKDD*. San Diego, CA (August 21–24).
Haveliwala, T. (2003). Topic-sensitive PageRank: A context-sensitive algorithm for Web Search. *IEEE Transactions on Knowledge and Data Engineering*.
He, Q., Weng, J., Jiang, J., & Lim, E.-P. (2010). TwitterRank: Finding topic-sensitive influential Twitterers. *WSDM*.
Holme, P., & Newman, J. (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review, 74*.
Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network". In *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 137–146). Washington, USA.
Kim, Y.-I., Ji, Y.-K., & Park, S. (2014). Social network visualization method using inherence relationship of user based on cloud. *International Journal of Multimedia and Ubiquitous Engineering, 9*, 13–20. 4.
Kleinberg, J. (1998). "Authoritative sources in a hyper-linked environment". In *Proceedings of the annual ACM-SIAM symposium on Discrete Algorithms*.
Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*.
Kumar, R., Mahdian, M., & Anagnostopoulos, A. (2008). "Influence and correlation in social networks". *Proceeding of SIGKDD*. LAS VEGAS, USA (August 24–27).
Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., Vanbriesen, J., & Glance, N. (2007). "Maximizing the spread of influence through a social network". In *KDD '07: Proc. of the 13th ACM SIGKDD int. conf. on knowledge discovery and data* (pp. 420–429). San Jose, USA.
Milgram, S. (1967). The small word problem. *Psychology Today*, 61–67. 1,1.
Mislove, A. E. (2009). *Online social networks: Measurement, analysis and applications to distributed information systems*. Texsas: Rice University.
Newman, M. (2010). *Networks: An introduction*. Oxford University Press.
Page, L., & Brin, S. (1998). "The anatomy of a large-scale hypertextual Web search engine" In *Proc. 7th international world wide web conference*.
Romero, D., Galuba, W., Asur, S., & Huberman, B. (2011). "Influence and passivity in social media". In *The European conference on machine learning and principles and practice of knowledge discovery in databases*. Athens, Greece (September 5–9).
Romero, D., Galuba, W., Asur, S., & Huberman, B. (2011a). Influence and passivity in social media. *PKDD*.
Sun, Y., Han, J., Zhao, P., Yin, Z., & Cheng, H. (2009). "RankClus: Integrating clustering with ranking for heterogeneous information network analysis". In *12th international conference on extending database technology*. Saint-Petersburg, Russia (March 24–26).
Tang, J., Sun, J., Wang, C., & Yang, Z. (2009). "Social influence analysis in large-scale networks". In *Proceedings of the 15th ACMSIGKDD international conference on knowledge discovery and data mining*. Paris, France (June 28–July 1st).
Valente, T. W. (2010). *Social networks and health: Models, methods, and applications*. New York: Oxford University Press.
Widom, J., & Glen, J. (2002). SimRank: A measure of structural-context similarity. *SIGKDD*.
Zhu, Z. G. (2013). Discovering the influential users oriented to viral marketing based on online social networks. *Physica A: Statistical Mechanics and Its Applications, 392*, 3459–3469.