# Modelling Influence in a Social Network: Metrics and Evaluation
# SCS Technical Report: TR-11-09


Behnam Hajian
School of Computer Science
Carleton University, Ottawa, Canada
Email: bhajian@scs.carleton.ca

Tony White
School of Computer Science
Carleton University, Ottawa, Canada
Email: arpwhite@scs.carleton.ca



***Abstract*—Social recommender systems are a recently introduced type of decision support system. One of the issues to be resolved in social recommender systems is the identification of opinion leaders in a network. Social Network Analysis is generally known for measuring network location metrics, centrality of nodes, and accessibility of users to one another in a network. While these measures are important in providing insight for defining roles in networks or defining different communities in a network, such measures miss important behavioral aspects of the network. The focus of this paper is the analysis of a network based on the interactions between users called behavioral analysis. One such measure is Influence Rank. The hypothesis explored in this paper is that this rank can be quantified based on the interaction between users and their behavior. The Influence Rank for a node is defined as the average Influence Rank of its neighborhoods combined with another index called Magnitude of Influence. The correlation between the proposed indices is analyzed in this paper. This combined measure is calculated by a recursive formula whose calculation complexity in the worst case yields non-polynomial time. However, this measure can be estimated by using the PageRank algorithm. The algorithm proposed here is based on estimating PageRank by eigenvector which is compared with an algorithm based on the limited iteration by depth of a spanning tree of the network. Results supporting the utility of the measure and the accuracy of its estimation using the PageRank approximation are presented.**



**Keywords- Social Influence; Social Network; Modeling; Behavioural Analysis;**


## I. INTRODUCTION

The notion of influence can help us in better understanding how we can help advertisers and marketing professionals design more effective campaigns. According to previous research, a relatively small number of people are able to persuade, or influence, others in their society. By targeting these influential people, we will be able to achieve large-scale influence with a small marketing cost thereby maximizing our ability to sell products or services within the society. This method has been mainly utilized by large companies to promote their products using high profile people, such as singers, actors, members of parliament and political leaders. In other words, companies hire influencers as local opinion leaders with different expertise to promote their products or services [1], [2], [3].

This theory assumes interpersonal relationships among users and the receptiveness of a society to opinion change as key factors of influence. The technique of collaborative filtering is a direct outcome of this theory as a marketing strategy. These theories remained largely unproven prior to the emergence of social networks, such as Twitter and Facebook, due to lack of empirical data for validation. Furthermore, this type of study is not straightforward because we do not have specific criteria for quantification of influence in a society [4], [1].

In Webster's dictionary the word *influence* is defined as "The power or capacity of a person or things in causing an effect in indirect or intangible ways". The study of influence dates back over a century. The role of influence and its effects is studied extensively in sociology, communication, marketing and political science and in understanding peer pressure, obedience and leadership. Three broad categories of social influence were introduced by Hernert Kelman [5]: (i) Compliance, which is defined as agreement among people while keeping their dissenting opinions private. (ii) Identification, in which people are being influenced by someone who is liked and respected, such as a famous celebrity. (iii) Internalization, in which people accept a belief or behaviour and agreement is made publicly and privately.

Morton Deutch and Harold Ferald [6] described two psychological needs; namely, the need to be right and the need to be liked, which are associated with informational and normative social influence respectively. Accepting information from another person is called informational influence as evidence about reality. In normative influence people try to conform to the positive expectation of others. Normative influence leads to public compliance, whereas informational influence leads to private acceptance. Informational influence is applicable when people are uncertain because of social disagreement or stimuli are ambiguous. Some of the factors in being influential are considered to be: charisma, bully pulpit, peer pressure, psychological manipulation, psychotherapy, reputation, emotion, social trends and social structure.

In modern theories, influence is considered to be a subjective topic, which adds the dimension of topics and expertise to this domain. This arises as a consequence of influence having a root in trust, and trust is subjective [7].

In order to identify opinion leaders, Social Network Analysis (SNA) evaluates the location of actors in the network as well as analyzing their interactions [8]. The class of network analysis which concentrates on measuring centrality of users defining their roles in a network is called structural analysis.

The term *Social Network* has been used for more than a century to denote patterns of ties among members of a society. Simply put, a social network is a social structure, which is modeled by a graph constructed from individuals represented as nodes and edges between nodes denoting ties or relations between pairs of nodes. In different social networks these ties connote different concepts; from friendship, and sexual relationships, to physical links between two network routers in a telecommunications network [9]. In SNA we concentrate on the mapping of relations and flows between users, groups, organizations, computers, people or any other entities, to a formal model [10]. This model then facilitates the measurement of criteria demonstrating different roles and properties of different entities in their society as well as defining roles and properties in computer networks.

Network analysis has found applications in many domains beyond social sciences; for example, the study of food chains in different ecosystems [11], identifying criminal and terrorist networks from traces of collected communications [10], [12], and understanding the interaction of proteins in metabolic pathways [13]. Most recently, the intersection of control theory and network analysis has led to a greater understanding of how control of complex networks might be achieved [14].

### A. Motivation, Contribution and Paper Structure

It is shown in [7] that consumers tend to get recommendations from the people they know and trust, such as friends or relatives rather than from automated recommendation systems used in e-commerce websites. This observation motivates e-commerce websites to incorporate social recommender systems as part of their websites. Furthermore, it is said that social networking sites – such as Facebook and Twitter – are influencing an increasing amount of the traffic seen by online retailers. These websites allow promoting of products by targeting influential people in their networks and allowing consumers to make comments on products. These trust-based recommender systems are the new alternative to collaborative filtering. One advantage of this socially-based approach is that it does not suffer from the problem of sparseness and cold-start found in collaborative filtering [15]. Using a social network to propagate information can significantly affect customer decision making [16], [17]; specifically, whether or not to visit an e-commerce website, buying or not buying a product from a particular retailer, and in writing and rating a review about a product. The growing importance of social recommender systems is one of the motivations for us in developing a model for quantifying and modeling influence in social networks.

Previous research has proposed methods for identifying key nodes by quantifying influence based on structural analysis of a network by criteria such as degree, betweenness and closeness centrality [8], [18], [19], [20]. In contrast with structural analysis, there is another class of methods for social network analysis, one that investigates the behavior of individuals in the network and their interactions. This latter class of methods analyzes how people interact with each other and how it affects information diffusion in a network; e.g., how followers of a specific node propagate or retweet postings; how followers are engaged in a conversation or what percentage of followers like a specific posting of a particular user.

In this paper, a social network is formally represented and a model proposed to measure influence in social networks based on not only the structure of the network but also the interaction between nodes in that network. The model proposes an index called the Magnitude of Influence (MOI) which is measured based on the size of the effect that postings of a user have on his neighborhood. MOI is combined with the influence rank of his/her neighborhoods forming another measure called Influence Rank (IR) which is recursively calculated for a node in social networks such as Twitter. The computational complexity of this rank calculation is non-polynomial. This paper proposes the use of the PageRank algorithm [21], [22] for calculating this index in order to find an opinion leader in a social network by an iterative process. The proposed method is evaluated using the FriendFeed dataset [23] and the results are compared in terms of completeness, performance and running time against a recursive method applied to a spanning tree of the network.

The rest of the paper is organized as follows. Section 2 provides background information about modelling social networks and methods for their analysis. In the third section, we model and formulate influence in social networks and provide details of the methods used for measuring social influence rank. Following this, the experimental results and the details regarding implementation are discussed. Finally, conclusions and future work are described in Section 5.

## II. Background and Related Work

As stated earlier, social network analysis has been studied in several domains such as identifying terrorist networks, optimizing wireless and cable network operators, and epidemiology. SNA helps us to understand how contacts among humans could affect the spread of diseases and rumor in social networks as well as preventing the spread of computer worms in a network [24], [25]. A central question in network analysis is understanding the relative importance of each node in propagating information through the network. Therefore, we need metrics to measure the importance of the role of each node in the network.

One measure representing the importance of a node in a network is closeness centrality. This metric is the mean length of all shortest paths from a node to all the other nodes in the network [26]. A node with higher closeness centrality is more visible in the network since this node has quicker accessibility to the entire network. In order to compute this metric, we have to compute all the shortest paths from a specific node to all the other nodes in the network.

Boundary Spanners are the nodes which connect their group to other clusters rather than locally connected nodes. Boundary spanners are well-positioned to combine different ideas. This characteristic makes them good innovators.

Eigenvector centrality is the other metric for measuring a node's popularity in a network. A node's eigenvector centrality is proportional to the sum of the eigenvector centralities of all nodes directly connected to it [26]. This metric is recursively calculated in the network according to the value of other node's eigenvectors centrality values. This metric is useful in defining the nodes which are connected to other well-connected nodes in the network. This is what is traditionally known as Google PageRank. The actual value of the eigenvector is calculated in non-polynomial time, but is often estimated by approximation algorithms such as PageRank.

Simply put, the aforementioned metrics concentrate on the structure of the network rather than the behavior of nodes and their interactions. However, some researchers consider other aspects of analysis such as analysis of users' behavior in a network. Some researchers consider a weight for each tie that shows the importance and strength of a link between two users based on their historical interactions.

Meeyoung Cha et al. [3] studied the measurement of influence in Twitter based on the following three metrics:

- In-degree: is the number of followers of a user. In-degree represents the popularity of a user in a social network.
- Retweets: which is defined as the number of times that followers of a specific node pass-along a posting from a tweeter. Retweeting causes propagation of a posting or news in a network. This metric is important as it shows how an advertisement can propagate across the network using influential users.
- Mentions: this means the number of times that the name of a user is mentioned in his followers' postings. This metric has been observed to follow a power law distribution.

Meeyoung Cha et al. [3] used Spearman's rank correlation coefficient for comparing users' influence. The research compared the three measures mentioned above to analyze topics of the most influential people in Twitter according to the aforementioned analysis and retweets. Meeyoung Cha et al. also revealed that influence is not gained accidentally but requires that users need to be consistently active in the network, an observation corroborated by Khrabrov and Cybenko [28].

Afrasiabi and Benyoucef [27] studied influence as a combination of link strength and incoming and outgoing clustering value defined for each node in the network. The link strength is measured according to the volume of interactions among users while the clustering value is measured by the closeness of a node to highly interconnected communities. They filter the spam and inactive nodes according to their activities and their interaction with other users.

Kempe et. al. [2] modelled influence by two basic diffusion models; i.e., Linear Thresholds and Independent cascade in which nodes are categorized as active or inactive nodes. The former model lies at the core of most subsequent generalizations. In the Linear threshold model, a node v is influenced by each neighbourhood w with a weight $b_{v,w}$ such that $\sum_{\forall w neighbor of v} b_{v,w} \leq 1$. In order for v to become active, the sum of the weights of its active neighbours should be more than a given threshold $\theta_v$. The Independent cascade model is based on interacting particle systems from probability theory in which a node w may become active in the time t+1 according to the probability $p_{w,v}$ if v had become active in the time t. This process runs until no more activations are possible. in this paper they have proved that both of the mentioned models are sub-modular. Finally it is proven that influence maximization problem is NP-hard for the aforementioned models. They have also used approximation methods to estimate influence maximization.

Dynamic Graph Analysis has been studied by Khrabrov and Cybenko [28] in which the number of daily mentions for each user is considered as a indicator for computing different ranks such as PageRank, drank, and starrank for influence analysis of each node in a network. For example, starrank considers user importance with respect to his or her neighborhoods. These researchers used several primitive indices in combination, such as Contiguous Longest Increasing Subsequences (CLIS) and GrowFall for analysis of influence ranks during a period of time. These indices show how the influence rank of a user changes with time. Khrabrov and Cybenko also analyzed the rate of increase in the number of mentions for influencing users in a network for consecutive days.

## III. Modelling and Formulation of Influence in Social Networks

In this paper we consider influence to be the ability of a person to convince others to change their mind about making a specific decision. This characteristic depends on several factors, such as: charisma, reputation, social personality and psychological manipulation abilities. In small communities in a social network the behaviour of each person is important in influencing his friends or followers. For example, the high number of postings in social networks such as Facebook decreases attention to one's postings. On the one hand, the quality of a posting is important for the followers of a specific user as well as the subject of a posting or a tweet. A subject might be interesting for a group of followers while it is bothersome or boring for another group of people. On the other hand, the quality of the personality of a person is another aspect of being influential in a network. The latter is more commonly known as trust which is established between two individuals.

### A. Social Network Modelling definitions

In order to analyze a network, the network needs to be formally defined. Here, a social network, $SN = (G, P, L, C)$, is a tuple in which G is a graph representing the network structure, $P$ is the set of postings posted by the individuals in the network, $L$ is the set of Likes about the postings and $C$ is the set of comments posted by individuals about the postings. In the graph $G =< V, E >$, $V$ is the set of vertices representing individuals and $E$ is the set of edges representing ties between individuals. In some models, $e = \{a, b\}$ just represents a tie and the direction of the relation is not important. Since in this model the direction of edges is important, every edge is represented by an ordered set, $e =< a, b >$ representing a directed edge from the node $a$ to $b$. In the network $SN$, $a$ is called a follower of $b$ if in the graph of their network, there exists an edge $< a, b >$ in $E$. Stated more formally:

$$follow(a, b) \longleftrightarrow \exists e =< a, b > \in E$$

The following definitions are necessary in order to model influence in the network $SN = (G, P, L, C)$:

- $\mid X \mid$ returns the cardinality of the set X and Po(X) is the power set of the set X.
- F(v): is a mapping function from V to set of vertices $F : V \rightarrow Po(V)$ (Po(V) is the power set of the set V). This function returns the set of nodes which are following the node $v \in V$ in a social network.
- P(v): is a mapping function from V to the power set of postings $P : V \rightarrow Po(P)$. This function returns the set of postings posted by the node $v \in V$ in the SN.
- L(p) is a function returning the set of individuals who liked the posting $p \in P$. $L : P \rightarrow Po(V)$.
- C(p) is a function returning the set of individuals who commented on the posting $p \in P$. $C : P \rightarrow Po(V)$.
- RT(p) is a function returning the set of individuals who retweeted the posting $p \in P$ in the network. $RT : P \rightarrow Po(V)$.
- LCRT is a function that determines whether a particular user has liked, commented on or retweeted a particular posting in the network. $LCRT : V \times P \rightarrow \{1, 0\}$.

$$LCRT(v, p) = \begin{cases} 1 & if(v \in L(p) \cup C(p) \cup RT(p)) \\ 0 & Otherwise \end{cases} \tag{1}$$

### B. Information Diffusion Model in a Social Network

Information diffusion in a social network can be thought of as moving and scattering information from one node in the network to its neighbours with the process repeated over and over again. In other words, diffusion of an idea or information explains how a piece of information spreads through the nodes in a society or a network. In this paper, a piece of information is said to be diffused across the network if the number of nodes who shared this information in their social network exceeds a threshold $\theta$. Formally, if the function $diff(i, v)$ returns the number of nodes which share a piece of information $i$ starting from node $v$, $i$ is diffused iff: $diff(i, v) \geq \theta_v$.

An opinion leader in a network is a node $v$ which is capable of starting the sharing of an idea $i$ in the network in a way that the idea get diffused across the network. In this paper, we are focusing on maximizing the probability of diffusing the piece of information $i$ starting from a diffuser node $v$. Hence, we are looking for such node that is most probably capable of that. The following equation shows that the probability of diffusion of information $i$ from $v$ to $w$ is a function of influence of node $v$ on $w$ and the intention of node $w$ for propagating $i$ in the network plus the degree of trust between $v$ and $w$. More formally:

$$Pr(Ret(w, i) \mid Ret(v, i)) = \lambda \times inf(v, w) + \beta \times intentionOfRetweeting(w) + \gamma \times trust(w, v)$$

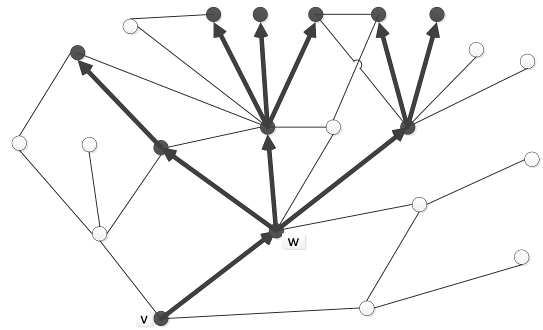


Fig. 1. Example

Figure 1 visually captures the process of information diffusion. The more that a particular idea is shared, the greater the diffusion of that idea.

### C. Modeling Influence in Social Network

In a manner similar to the research work described in Section II we assume influence to be represented as a combination of several distinct metrics, which are dependent on one another. In some of the previous research (e.g., [3]) these metrics were studied independently. However, in the proposed model, metrics are combined with each other to measure the influence of a user in a social network. These metrics are:

- *Number of followers* is a metric which represents the popularity of a node in a network. In certain research [3] this measure is also called the in-degree of a node in the graph representing the social network. The number of followers represents the number of connections that a user has. As a result, this number indicates the number of people who may be affected by his/her postings. In general, this measure does not represent the influence of a person on his society, but an influential person who has more connections is more valuable to us than an influential person whose followers are limited to a small community. Therefore, a non-linear function from the ratio of the in-degree of a node to the maximum in-degree in a network indicates the popularity of a node in the network. The following function $\delta : V \rightarrow R$, normalizes the number of followers in the range of $[0, 1]$ with a non-linear function.

$$\delta(v) = \frac{Ln(\mid F(v) \mid - min_{v' \in V} \mid F(v') \mid)}{Ln(max_{v'' \in V} \mid F(v'') \mid - min_{v' \in V} \mid F(v') \mid)} \tag{2}$$

- *Ratio Of Affection* (ROA) is the main index showing the ratio of followers which were affected by a particular posting. This ratio is proportional to the number of followers who commented on, liked or propagated/retweeted a posting in a network. In previous research this is called $mentions$ [28]. In other words, this number indicates the rate of impression that followers of a particular user get from a particular posting posted by the user. The following function $ROA : V \times P \rightarrow R$ returns the ROA value for a particular posting posted by a specific node.

$$ROA(v,p) = \frac{\sum_{v' \in F(v)} LCRT(v',p)}{\mid F(v) \mid} \tag{3}$$

- *Magnitude of Influence* (MOI) is the root mean square of ROA for all of the postings which were posted by a particular user. The root mean square is often used for measuring the magnitude of a varying quantity. This index indicates the magnitude of the ratio of affection for different postings of a particular user in the social network. The function $MOI : V \rightarrow R$ returns this value for a specific node in the network.

$$MOI(v) = \sqrt{\frac{\sum_{p' \in P(v)} (ROA(v,p'))^2}{\mid P(v) \mid}} \tag{4}$$

MOI indicates the total impression that a user makes on his society. This index depends on the quality of postings as well as characteristics of the user's personality. In fact, a specific posting might be liked or commented on by a high number of followers of a user while the same posting may not affect the same followers if it was posted by another user. On the other hand, a posting might be effective for one community while it may not be interesting for another community in which it was shared. However, each user might be a member of different communities and his postings are visible to all the communities in which he/she participated.

- *Influence Rank* (IR) is another index considering an influential person as an individual who is connected to other influential individuals. An influential person with a high value of IR is called an opinion leader who affects other influential individuals in propagating an idea or an advertisement across the network. The following equation shows that the degree of influence for a user is proportional to the degree of influence of his/her followers.

$$IR(v) = (1 - \delta(v)) \frac{\sum_{v' \in F(v)} IR(v')}{\mid F(v) \mid} + \delta(v) \times MOI(v) \tag{5}$$

In the above equation, the function $\delta(v)$ is a dynamic damping factor which itself is a function of the number of followers for the node. The importance of MOI is dependant on the number of followers. According to the IR index, a user with a large number of followers should have a high value of MOI to be an influential person. On the other hand, for a user with a low number of followers, the influence rank of his/her followers is more important than his/her MOI. The IR value is calculated by approximating algorithms such as the PageRank algorithm [21] described in the next section.

The above indices show that the influence of a particular node in the network is based on its interactions in that society. In order to find opinion leaders in the whole network, we have to consider the structure of the network as well as the magnitude of influence for individuals. Influence Rank as defined in this paper has the property that users with influential neighbourhoods get higher rank. In the next section, two different methods for estimating IR are introduced.

### D. PageRank Algorithm

In order to calculate an IR index while avoiding non-polynomial computational complexity, this index is approximated by Algorithm 1.

PageRank is a link analysis algorithm that is used by Google search engine. Google uses PageRank in order to sort search results. In PageRank, a page is considered important if it gets more votes from other pages. The hyperlinks from other pages to a specific page are considered as votes from other pages to that page. In PageRank, a page that is linked to by many pages with high PageRank receives a high PageRank itself.

Simply stated, the PageRank is a probability distribution representing the likelihood of arriving at a particular webpage starting from a webpage and randomly clicking on a series of links. The probability is a real number in the range of 0 and 1. In the PageRank algorithm, a user who is randomly crawling web pages finally stops surfing. This fact is usually modeled by a damping factor which is a probability of continuing clicking on the next page. In the simplified algorithm the PageRank is defined as follow:

$$PR(u) = (1 - \alpha)/N + \alpha \sum_{v \in M(u)} \frac{PR(v)}{L(v)} \tag{6}$$

The PageRank value of each node is approximated by the eigenvector iterative equation $Av = \lambda v$. Considering t as the number of iterations the above equation is written as follows:

$$PR(P_i; t+1) = (1 - \alpha)/N + \alpha \sum_{P_j \in M(P_i)} \frac{PR(P_j; t)}{L(P_j)} \tag{7}$$

The iterative PageRank algorithm calculates the eigenvector $PR$ using following equation:

$$PR = \begin{bmatrix} PR(p_1) \\ \vdots \\ PR(p_N) \end{bmatrix}$$

$$PR_{t+1} = \begin{bmatrix} \frac{1-\alpha}{N} \\ \vdots \\ \frac{1-\alpha}{N} \end{bmatrix} + \alpha \begin{bmatrix} l(p_1, p_1) & ... & l(p_1, p_N) \\ \vdots & ... & \vdots \\ l(p_N, p_1) & ... & l(p_N, p_N) \end{bmatrix} PR_t$$

$l(p_i, p_j)$ in the above matrix is the element of the adjacency matrix of the social network graph. The PageRank algorithm iteratively calculates the above equation until $Avg(\mid PR_{t+1} - PR_t \mid) \leq \varepsilon$. The details of this algorithm are described in [21]. This algorithm is modified to estimate the Influence Rank as follows:

**Algorithm 1** PageRank algorithm for calculating IR

**Proc PageRank-IR**(double threshold)
**while** error $>$ threshold **do**

$$IR_{t+1} = \begin{bmatrix} \delta(p_1) \times MOI(p_1) \\ \vdots \\ \delta(p_N) \times MOI(p_N) \end{bmatrix}$$

$$+\alpha \begin{bmatrix} l(p_1, p_1) & \dots & l(p_1, p_N) \\ \vdots & \dots & \vdots \\ l(p_N, p_1) & \dots & l(p_N, p_N) \end{bmatrix} IR_t$$

error = $IR_{t+1} - IR_t$
**end while**
**return** $IR_{t+1}$

In this algorithm $l(p_i, p_j)$ is zero if there is no edge between the corresponding edges, and is equal to $Adjacency(p_i, p_j) \times (1 - \delta(p_i))/(\mid F(p_i) \mid)$.

The run time complexity of this algorithm is polynomial. The most expensive calculation in this algorithm is the matrix multiplication which has the order of O($n^{2.38} \times m$) in which $m$ is the number of iterations and $n$ is the number of nodes in the network.

### *E. Limited Recursive Algorithm (LRA)*

Algorithm 2 estimates Influence Rank by traversing neighbourhoods for each node to a given depth. This algorithm recursively explores the neighbourhood of the node for which the influence rank is estimated.

**Algorithm 2** LRA for calculating IR

**Proc LRA-IR**(Node $u$, int level, int max)
**if** level = max **then**
  **return** $\delta(u) \times MOI(u)$
**else**
  **for all** $v \in N(u)$ **do**
    sum $\leftarrow$ sum + **LRA-IR**(v, level+1, max)
  **end for**
**end if**
AvgIR $\leftarrow$ sum / $\mid F(u) \mid$
**return** $(1 - \delta(u))AvgIR + \delta(u) \times MOI(u)$

The run time complexity of this algorithm is non-polynomial with recursion limited by a predefined value as the input parameter; i.e., max in Algorithm 2.

## IV. Analysis of Results

This section focuses on the analysis of the results generated from both estimation algorithms. The above algorithms are evaluated using the FriendFeed 2010 dataset [23]. The dataset contains 665,000 users (nodes), 27 million subscriptions (edges), 8 million tweets, 3 million comments and 1 million records indicating likes. Figure 2 demonstrates the results of the execution of both algorithms for the 200 most influential users. The results comparing the two algorithms are promising since the average relative error is 0.13 with 0.95 correlation using a Pearson Correlation Coefficient. Furthermore, Spearman's Rank Coefficient is 0.90 which indicates a very high degree of correlation between the rankings provided by the two methods. This is particularly important in that the question of most interest in this paper is the identification of individuals of high influence. We are less interested in the absolute value of the influence rank and more interested in the sorted values. Finally, Figure 2 clearly shows that the slopes of the trend graphs for the rankings produced by the two algorithms are almost equal.

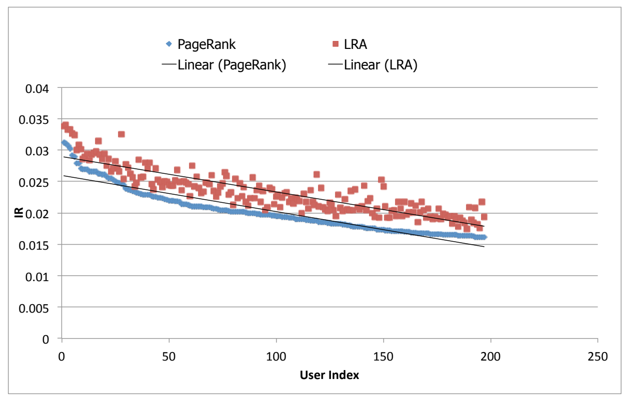


Fig. 2. LRA vs. PageRank

Figure 3 clearly shows that there is no correlation between number of followers and influence rank in our model; i.e., structural information alone cannot be used to predict influence. This is to be expected as IR in our proposed model takes into account the IR of followers. However, the most influential users with a high number of followers are more interesting for advertisers since there is a greater probability for highly connected and influential users to diffuse a piece of information or an advertisement. In the proposed model the number of followers affect the rank in a way that the rank is increased for those nodes which have more connections with high MOI or fewer connections with other nodes that have high influence rank. Hence, the nodes on the envelope bounding the influence rank vs. number of followers scatter plot are of most interest to potential advertisers. For example, in Figure 3, the user with 524 followers and an IR of 0.029 may well represent a promising target for advertisers.

The following graph, Figure 4, shows the correlation between Magnitude Of Influence and Influence Rank. This graph shows a strong linear relationship between the two influence-based indices.

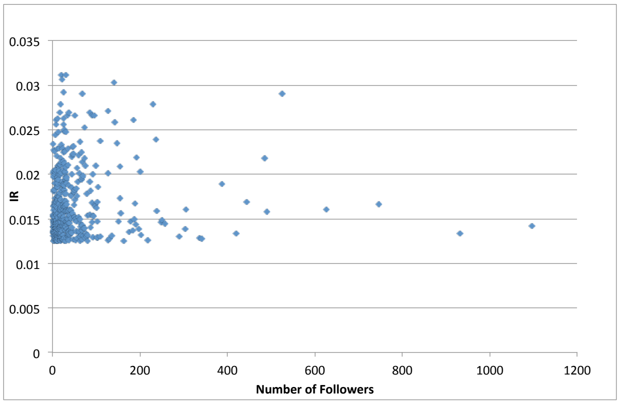


Fig. 3. Correlation between number of followers and Influence Rank for each node

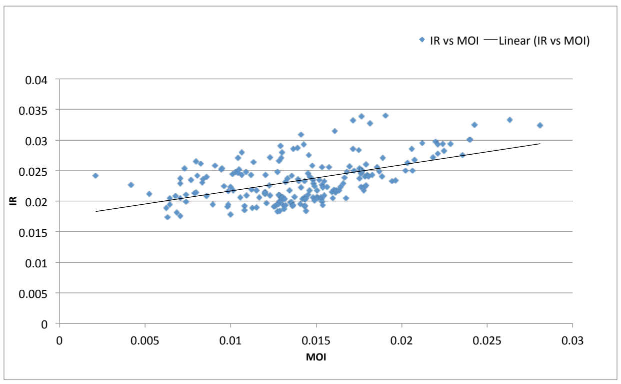


Fig. 4. Correlation between Influence Rank and Magnitude of Influence

## V. Conclusion and Future Works

This paper proposes a formal model for measuring influence in a social network. The number of comments or likes on a posting – or in other words the number of hits for a posting – is formulated as the Magnitude of Influence for a node. Furthermore, in order to be an influential user, the subscribers to (or followers of) the user should be influential nodes. As a consequence of this hypothesis, another index is formulated; a recursive concept of being influential, namely Influence Rank. The recursive notion of Influence Rank implies non-polynomial time complexity for the index calculation. Sections III.D and III.E proposed two algorithms to estimate this measure; i.e., the limited recursive algorithm and a modified PageRank. The limited recursive algorithm results in a value closer to the true value. Assuming LRA to represent the ground truth – the real influence rank – we have shown that a modified PageRank algorithm can be used to accurately compute it. Both the trend in the influence rank values and the rankings themselves correspond closely to the values computed using LRA. The PageRank-generated values are typically within 10% of those computed using LRA. Most importantly, the ranking of the two algorithms is very similar. Given that the question of most interest in this paper is that of the identification of the most influential nodes, the accuracy of the ranking is of crucial importance. Moreover, the complexity of PageRank is polynomial and this algorithm computes influence rank very quickly on the FriendFeed dataset while computing results close to the ground truth.

In our future work we will consider a probabilistic diffusion model of information in a social network that incorporates the semantics of proposed topics of transmission. Our motivation for considering this research is that such models would be useful in finding users with the highest capability for diffusing specific information according to their connections in a network and the trust of their followers in them. It is our belief that multi-dimensional trust models will contribute to this research. Such models may be useful in predicting crises in a network by modelling diffusion of information and the probability of occurrence of crises as the consequence of a rumour or critical news in a network.

## References

[1] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 88–97.

[2] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[3] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proceedings of the 4th International Conference on Weblogs and Social Media*, 2010.

[4] D. Cosley, D. Huttenlocher, J. Kleinberg, X. Lan, and S. Suri, "Sequential influence models in social networks," in *Proc. 4th International Conference on Weblogs and Social Media*, 2010.

[5] H. Kelman, "Compliance, identification, and internalization: Three processes of attitude change," *The Journal of Conflict Resolution*, vol. 2, no. 1, pp. 51–60, 1958.

[6] M. Deutsch and H. Gerard, "A study of normative and informational social influences upon individual judgment." *The journal of abnormal and social psychology*, vol. 51, no. 3, p. 629, 1955.

[7] A. Salehi-Abari and T. White, *The Impact of Naive Agents in Heterogeneous Trust-Aware Societies*. Springer, March 2010, vol. LNAI 5683, pp. 110–122.

[8] M. Fink and J. Spoerhase, "Maximum betweenness centrality: approximability and tractable cases," *WALCOM: Algorithms and Computation*, pp. 9–20, 2011.

[9] B. Wellman and S. Berkowitz, *Social structures: A network approach*. Cambridge Univ Pr, 1988, vol. 2.

[10] L. Freeman, *The development of social network analysis*. Empirical Press, 2004.

[11] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. White, "Economic networks: The new challenges," *science*, vol. 325, no. 5939, p. 422, 2009.

[12] S. Borgatti, A. Mehra, D. Brass, and G. Labianca, "Network analysis in the social sciences," *science*, vol. 323, no. 5916, p. 892, 2009.

[13] B. Balasundaram, S. Butenko, I. Hicks, and S. Sachdeva, "Clique relaxations in social network analysis: The maximum k-plex problem," *Operations Research*, p. 26, 2009.

[14] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabasi, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, May 2011, 10.1038/nature10011. [Online]. Available: http://dx.doi.org/10.1038/nature10011

[15] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.

[16] Y. Kim and J. Srivastava, "Impact of social influence in e-commerce decision making," in *Proceedings of the ninth international conference on Electronic commerce*. ACM, 2007, pp. 293–302.

[17] H. Ma, H. Yang, M. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," in *Proceeding of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 233–242.

[18] K. Goh, E. Oh, B. Kahng, and D. Kim, "Betweenness centrality correlation in social networks," *Physical Review E*, vol. 67, no. 1, p. 017101, 2003.

[19] R. Puzis, Y. Elovici, and S. Dolev, "Finding the most prominent group in complex networks," *AI communications*, vol. 20, no. 4, pp. 287–296, 2007.

[20] N. Suri and Y. Narahari, "Determining the top-k nodes in social networks using the shapley value," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 1509–1512.

[21] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[22] B. Csáji, R. Jungers, and V. Blondel, "Pagerank optimization in polynomial time by stochastic shortest path reformulation," in *Algorithmic Learning Theory*. Springer, 2010, pp. 89–103.

[23] F. Celli, F. M. L. D. Lascio, M. Magnani, B. Pacelli, and L. Rossi, "Social network data and practices: The case of friendfeed," in *SBP*, ser. Lecture Notes in Computer Science, S.-K. Chai, J. J. Salerno, and P. L. Mabry, Eds., vol. 6007. Springer, 2010, pp. 346–353.

[24] S. Eubank, V. Kumar, M. Marathe, A. Srinivasan, and N. Wang, "Structure of social contact networks and their impact on epidemics," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 70, p. 181, 2006.

[25] D. Shah and T. Zaman, "Detecting sources of computer viruses in networks: theory and experiment," *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 1, pp. 203–214, 2010.

[26] S. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[27] A. Rad and M. Benyoucef, "Towards detecting influential users in social networks," in *E-Technologies: Transformation in a Connected World: 5th International Conference, MCETECH 2011, Les Diablerets, Switzerland, January 23-26, 2011, Revised Selected Papers*, vol. 78. Springer, 2011, p. 227.

[28] A. Khrabrov and G. Cybenko, "Discovering influence in communication networks using dynamic graph analysis," *Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust, 2010 IEEE International Conference on*, vol. 0, pp. 288–294, 2010.