IJICTR
International Journal of Information & Communication Technology Research

Volume 10- Number 4 – Autumn 2018 (11 -21)

On the Reselection of Seed Nodes in Independent Cascade Based Influence Maximization

Ali Vardasbi
School of Electrical and Computer Engineering, College of Engineering, University of Tehran
Tehran, Iran
a.vardasbi@ut.ac.ir

Heshaam Faili*
School of Electrical and Computer Engineering, College of Engineering, University of Tehran
Tehran, Iran
hfaili@ut.ac.ir

Masoud Asadpour
School of Electrical and Computer Engineering, College of Engineering, University of Tehran
Tehran, Iran
asadpour@ut.ac.ir

Received: 22 April 2018 - Accepted: 3 September 2018

Abstract— **Influence maximization serves as the main goal of a variety of social network activities such as viral marketing. The independent cascade model for the influence spread assumes a one-time chance for each activated node to influence its neighbors. On the other hand, the manually activated seed set nodes can be reselected without violating the model parameters or assumptions. This view divides the influence maximization process into two cases: the simple case where the reselection of the nodes is not considered and the reselection case. In this study we will analyze real world networks in the reselection case. First we will show that the difference between the simple and the reselection cases constitutes a wide spectrum of networks ranging from the reselection-free to the reselection-friendly ones. Then we will experimentally show a significant entanglement between this and influence spread dynamics as well as other structural parameters of the network. Specifically, we show that under a realistic condition, the reselection gain of a network has a correlation of 0.73 to a newly introduced influence spread dynamic. Furthermore, we propose a measure for detecting star-like networks and experimentally show a significant correlation between our proposed measure and the reselection gain in real world networks with different edge weight models.**

Keywords-Influence Maximization; Network Structure; Independent Cascade; Maximization over Integer Lattice; Core Decomposition

I. INTRODUCTION

The focus on the influence maximization and influence propagation has grown increasingly in the social network studies [1]–[5]. The fundamental question concerning the influence maximization problem is that what group of nodes, when selected as the initial influencers, can spread the desired influence to the highest extent possible [6]. The selection of a node as an initial influencer practically means spending a reasonable amount of budget such as money, time, reputation, etc. in order to activate it. An active node then tries to influence its neighbors and hopefully the cascade of influence would be triggered.

There are different theoretical models for the influence spread in a social network, amongst which the

* Corresponding Author

linear threshold (LT) and independent cascade (IC) models are the most used ones. In the LT model each node is considered to have a threshold and it is activated when the number of its active neighbors goes above that threshold [7], [8]. The IC model, on the other hand, deals with the influence probabilities of the links [9]. According to the IC model, each directed link (u, v) is associated with a probability p_{uv} that indicates the power of u at influencing v. Once u is influenced (either as an initial node or during the influence spread), it has a *one-time chance* to activate v and is successful to do so with probability p_{uv}. During the influence spread process, giving the node u a second chance to influence its neighbor v will increase the influence probability from p_{uv} to $1-(1-p_{uv})^2$ and the parameters of the IC model will be violated. However, when a previously influenced node is manually re-activated the scenario will be different. The difference between these two cases is more clarified in the following example.

Consider a social network for which the influence of individuals on their connections has been estimated from their activity. More specifically, in this example, influence has the form of clicking on the link that one has posted on the network. Furthermore, once a user has clicked on a posted link, his connections will be notified as if he has re-posted the link. Suppose that we have a web page and we desire to increase the number of our page views via advertising it on the mentioned social network. Our budget determines the number of initial users to whom we afford to introduce our page and ask them to post a link of it on the network. During the cascade of influence through the network, naturally a user will not re-post our link twice. Therefore, the connections of an active user will see his post once. But assume that we have paid one of our initial users double and asked him to post our link twice. Since the second chance has been given to him forcefully, the natural process of influence spread in the IC model has not been violated. Furthermore, if the time interval between the two posts of the same user is selected appropriately, his influence power will be nearly doubled.

Alon et al. [10] introduced several budget allocation models for influence maximization in social networks. In [11] the proposed framework of [10] is extended and it is proved that the underlying spread function is submodular over the integer lattice. The main shortcoming in these models is that they do not consider the influence propagation. Avigdor-Elgrabli et al. [12] address this issue by introducing a generalized model for the budget allocation that captures the influence propagation in the network as well. Then, they theoretically study the model in both offline and online settings and identify a family of monotone submodular influence functions over the integer lattice.

In this paper, we will experimentally study a practical budget allocation model and analyze the different behaviors of real-world networks towards such a model. We will call the situation where a node is selected more than once during the influence maximization process, the *reselection* of that node. It is worth noting that the reselection approach is quite common in the real-world advertising. Usually, based on the budget of the company as well as the capacity of an advertising hub, the hub is paid more than once to popularize a specific product. Reselecting a hub to maximize the influence spread demonstrates the fact that when a node has a large number of important connections, a one-time attempt does not saturate its capacity and even if a fraction of its connections has been influenced at the first try, the hub's importance is still more than many other nodes in the network.

We study the dynamics of networks concerning the reselection of seed set nodes in an influence maximization process. Since in the reselection model the seed nodes are not necessarily unique, we use the term *seed multiset* instead of *seed set*. We first evaluate the behavior of different real-world networks against the reselection possibility of the seed nodes. It is shown that different networks respond differently to this new feature. In some networks, there is hardly a duplicate in the seed multiset. This means that, in the aforementioned networks, introducing a new node to the seed multiset usually has a better performance compared to reselecting a previous seed node. On the other hand, in a number of other networks, only a small percent of the seed multiset nodes are unique. These networks have a considerably higher influence spread in presence of the reselection mechanism.

The main question of this study is about the cause of the above observation in social networks. To tackle this question, first it is shown that the reselection gain is correlated to another influence dynamics, the *influence saturation*. Roughly speaking, the influence saturation measures the extent of degradation in the marginal influence spread during the expansion of the seed multiset nodes. Then, using the correlation between the influence saturation and the reselection gain, an entanglement between these dynamics and the network structure will be shown. The significance of this result is most understood for the large networks on which performing the influence maximization algorithms is time consuming. In such cases, our results can be used to identify, in negligible time, whether or not a given network is reselection-friendly. Upon identification, suitable influence spread policies can be adapted accordingly.

Another practical point of this paper is that its results can be used to detect the origin of reselection-aware behavior of different networks. This knowledge is useful for the organization who wants to maximize the influence spread in the network. For example, it gives insights on how to manipulate the network, by adding new nodes or building new links, in order to change its reselection-aware behavior and increase their benefit.

The structure of the consequent sections is as follows. In the following two subsections a brief overview of the influence maximization research and the budget allocation models as well as the definitions and parameters required for the following parts of the paper are presented. In Section II we will discuss the saturation dynamics in the influence maximization and propose a L-curve based parameter for measuring it. In order to be able to present our observations in the real-world networks, we first explain our experimental setup in Section III. In that section we also introduce a new model for the edge weights which considers the

transitivity behavior in social networks. After that, in Section IV, we first show the different behaviors of networks to the reselection possibility. Then, we argue why the reselection gain is supposed to correlate to the saturation behavior and show such a correlation for a class of networks. Finally, we will show a correlation between the percentage of low degree nodes and the reselection gain. We will conclude the paper and propose possible future work in Section V.

A. Related Work

The formal definition of influence maximization is given in [6] as:

Definition 1 (Influence Maximization) Given a directed graph G as a social network and a diffusion model for the influence; determine the set of influential targets of size at most k whose activation will cause the largest number of activated nodes in G.

Kempe et al. showed that the influence spread function is a submodular function and hence proposed a greedy $(1-1/e)$ -approximation to the above problem. The high time complexity of the greedy algorithm commenced a new stream of research on the scalable influence maximization proposals. In this paper the CELF++ algorithm of [13] is referred to as the simple greedy algorithm. However, CELF++ and other speed ups such as [14], [15] did not scale acceptably for the networks of millions of nodes. As the social networks grow larger and larger, the need to scalable algorithms with promising performances becomes more realized. That is why a considerable number of scalable influence maximization algorithms have been published in recent years [16]–[21].

The budget allocation in influence maximization models the fact that the probability of an influencer to influence its neighbors depends on the budget allocated to it. Conveniently, the literature deals with the discrete budgets in this context [10]. As such, the allocation of k units of budget to someone means giving her k chances to influence her neighbors (instead of a one-time chance). Earlier budget allocation models such as [10], [11] did not consider the propagation of influence in the network. These models are consisted of a bipartite graph connecting the source and the target nodes. Based on their allocated budgets and according to a given influence model, the source nodes influence a number of their target neighbors and no spread of influence happens.

In [12] the influence propagation is introduced to previous budget allocation models, yielding a rather complete model for the influence maximization budget allocation in social networks. They consider the influence spread as a two-stage process:

1. Influence of seed nodes based on their allocated budget on their neighbors,
2. Influence propagation initiated by the influenced seed neighbors from the previous stage.

Based on the above assumption, they propose the *budgeted triggering* model whose combined influence function is a monotone submodular function over the integer lattice.

Hatano et al. considered the adaptive allocation of budgets based on the responses from the previous campaigns [22].

B. Our Contribution

We use a practical version of the model by [12] to analyze the behavior of real world networks on the influence spread dynamics when the reselection is possible. Unlike [12] which deals with the theoretical bounds and approximation algorithms of the budget allocation in the influence maximization problem, we try to study the practical consequences of this generalization. Our contributions can be listed as follows:

- The two currently models for the edge weights, namely Weighted Cascade and Trivalency, do not consider the transitivity behavior of the social networks. We present the transitive multi-valency model to address this issue.
- We show that in all of the three models, the networks have a wide range of responses to the reselection possibility. The reselection-friendly networks demonstrate their friendly behavior even for the fading parameters as low as 0.6.
- We detect a high entanglement between the reselection gain and influence spread dynamics for a class of network with a specific structure.
- Finally, it is demonstrated that the star-like networks with a high portion of low degree nodes have a high probability of being reselection-friendly.

C. Parameters and Definitions

Considering the possibility of reselection at the influence maximization seed set nodes is equivalent to substitute *set* into its generalized concept *multiset*. A multiset is a collection of elements that can have multiple instances of elements [23]. The number of instances of an element in a multiset is called the element's *multiplicity*. For example, in the multiset $\{a, a, a, b\}$ the elements a and b have multiplicity 3 and 1 respectively. A set is a special case of a multiset for which all the elements have multiplicity 1. Multisets are sometimes represented by elements of $\mathbb{Z}_+^m$, a vector of non-negative integers where m is the size of the elements space and each field of the vector represents the multiplicity of the corresponding element.

Consequently, the *reselection possible influence maximization* is defined with the help of the multisets.

Definition 2 (reselection possible influence maximization) Given a directed graph G as a social network and a diffusion model for the influence; determine the seed multiset of influential targets of size at most k whose activation will cause the largest number of activated nodes in G. Each node of the seed multiset with multiplicity m has a m times chance at influencing its neighbors.

One may argue that the reselection of a seed node has less influence compared to its selection as the first time. To address this issue, we define a more general setting that models the possible fading effect caused by reselection.

Definition 3 (reselection possible influence maximization with fading) Given a directed graph G as a social network, a diffusion model for the influence and a fading parameter $0 \leq \alpha \leq 1$; determine the seed multiset of influential targets of size at most k whose activation will cause the largest number of activated nodes in G. Each node of the seed multiset with multiplicity m has a m times chance at influencing its neighbors; but its influence at the ω^{th} chance is faded by a factor of $\alpha^{\omega-1}$. The extreme cases where $\alpha = 0$ or $\alpha = 1$ respectively correspond to the simple influence maximization case (Definition 1) and the reselection possible influence maximization without fading (Definition 2).

Submodular functions play an important role in influence maximization as well as a great number of computer science optimization problems. A submodular function is mostly known by the diminishing return property.

Definition 4 (Submodular function) A set function $f: 2^V \rightarrow \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq V$ and $e \in V \backslash B$ it holds that

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B) \qquad (1)$$

When the reselection of nodes is possible and we are dealing with the multiset rather than set, the set function can be extended to a function over the integer lattice; i.e. non-negative integer vectors over the Euclidean space. A submodular function over the integer lattice is characterized as follows:

Definition 5 (Submodular function over integer lattice) A function $f: \mathbb{Z}_+^m \rightarrow \mathbb{R}$ is submodular if for every $x, y \in \mathbb{Z}_+^m$ it holds that

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \qquad (2)$$

Where $x \vee y$ and $x \wedge y$ represent the coordinate-wise maxima and minima, respectively.

Through the rest of this paper, the influence spread function of a set S and a multiset M on a network G is shown by $\sigma_G(S)$ and $\sigma_G^m(M)$, respectively. The superscript m on the latter function denotes the multiset domain of the function. To compute the spread of M, each node of the multiset is given as many chances as its multiplicity within M.

Finally, we define the *reselection gain* (RG) to be the ratio of the influence spread in the reselection case to the simple case. Formally, for a given graph G and seed size k, the reselection gain is defined to be:

$$RG_G(k) = \frac{\max_{|M|=k} \sigma_G^m(M)}{\max_{|S|=k} \sigma_G(S)} \qquad (3)$$

II. Influence Saturation

Suppose that for each k the maximum influence spread on graph G caused by activating k nodes of G is shown by $\tau_G(k)$. The submodularity of the spread function implies that $\tau(k)$ is a concave function of k. As such, for every graph G there is a saturation threshold k_G^* after which the positive slope of the $\tau_G(k)$ function will be insignificant; i.e. the graph *saturates* by the influential seed set nodes of size k_G^*.

Observations on the behavior of the $\tau_G(k)$ function for real world networks G reveals an interesting saturation dynamics. For a number of networks the saturation threshold is 1. In other words, the influence spread of the most influential node is such that the marginal gain of the next seed set nodes becomes negligible. We call this behavior as the *sharp saturation*. Figure 1 shows two sets of networks with different saturation behaviors. The y-axis of these plots is the $\tau_G(k)$ normalized by the node size of graph $|G|$ for simplicity of comparison.

We define the *influence saturation* (IS) parameter to entail the saturation dynamics of different networks. The problem of finding a saturation measure for a concave function has a resemblance to the L-curves and using them to solve ill-posed problems through regularization [24]. One method for locating the elbow in a L-curve is to find the point with maximum distance from the line obtained by connecting the two ending

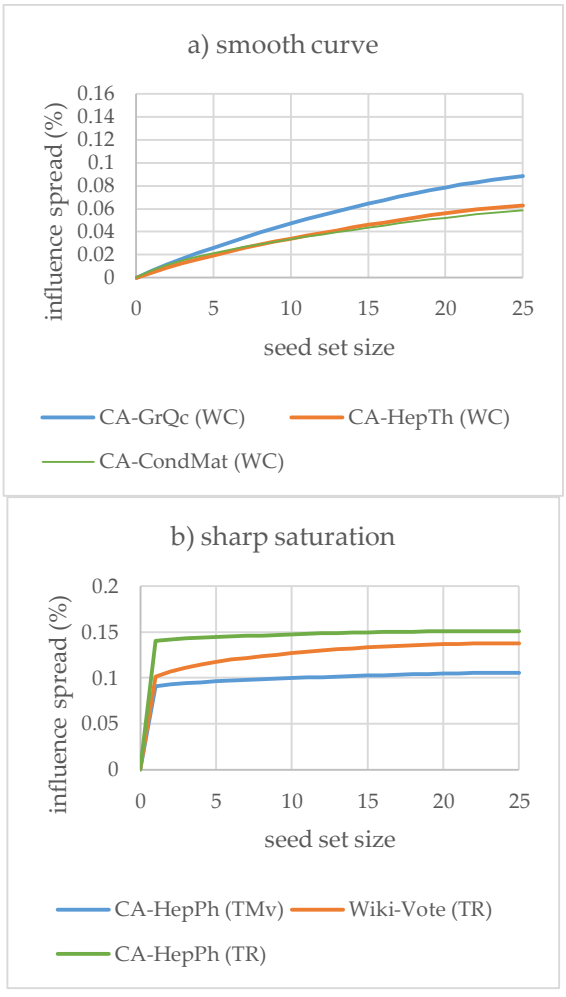

Figure 1. The plot of $\tau_G(k)/|G|$ versus k. The network in (a) have a smooth curve; while the networks in (b) have a sharp saturation

points of the curve. Using this method, we define another metric to measure the sharpness of saturation in a graph:

$$LCIS(G) = \frac{|k_{max} \cdot \tau(k_{min}) - k_{min} \cdot \tau(k_{max})|}{\sqrt{k_{max}^2 + \tau^2(k_{max})}} \qquad (4)$$

The L-curve Influence Saturation defined in (4) is simply the distance of the τ function at point k_{min} from the line connecting the origin to the last point of τ at k_{max}. A high $LCIS$ parameter is a sign of a network with sharp saturation behavior. As will be shown in our experiments, the two parameters proposed in this section are highly correlated.

The LCIS and RG parameters are expected to be high in a star like network. By the star like network we mean a network whose nodes can be decomposed into two components:

- Core nodes: a small set of nodes which are connected to a considerable fraction of network. These nodes are highly influential.
- Loosely connected nodes: a large number of nodes which are weekly connected to each other but strongly connected to the core nodes.

In a star like network, selecting one of the core nodes will spread the influence to a large section of the network and causes a sharp saturation. On the other hand, reselecting the core nodes instead of non-core nodes is likely to increase the influence spread which means a high reselection gain.

In our experiments we will test the hypothesis that "are all the networks with a high RG, star like?"

III. Experimental Setup

We use the Pearson correlation to show the entanglement between different parameters on different networks. In order to provide a confidence level for the reported correlations, we perform the permutation test on the data [25]. We construct the correlation on the randomized data 10^6 times and report the confidence level with a precision of three significant figures.

In the proceeding sections the statistics of the experimented networks as well as the models used as the edge weights are explained.

A. Networks

The experiments of this paper are conducted on the real world networks obtained from [26]. The node and edge sizes of the networks range from 4k to 317k and 28k to 2M respectively. All the networks in this paper are directed. In the cases where the original network was undirected, we have considered two directed edges for each undirected edge, making the edge size of the network twice its original. The networks are described below:

- **Facebook**: The Facebook dataset consists of friend lists from Facebook. The data is collected from survey participants [27]. In our experiments we only used the graph of friendship.
- **Wiki-Vote**: The network contains all the Wikipedia adminship voting data until January 2008. The nodes represent Wikipedia users and a directed edge from node i to node j indicates that user i has voted for the adminship of user j [28], [29].
- **Email-Enron**: This dataset contains the email communications of Enron. The nodes represent the Enron email addresses and an undirected link between i and j indicates that either of them has sent an email to the other [30], [31].
- **Epinions**: This graph is a who-trusts-whom online social network of a general consumer review site Epinions.com [32].
- **Slashdot**: Slashdot is a technology-related news website known for its specific user community. The network contains friend/foe links between the users of Slashdot [31].
- **DBLP**: The DBLP computer science bibliography provides a comprehensive list of research papers in computer science. This graph is a co-authorship network where two authors are connected if they publish at least one paper together [33].
- **CA-GrQc, CA-HepTh, CA-HepPh, CA-Astro, CA-CondMat**: These graphs are the collaboration network from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category, High Energy Physics Theory, High Energy Physics Phenomenology, Astro Physics and Condense Matter categories, respectively [34].
- **Cit-HepPh**: The citation graph from the e-print arXiv that covers all the citations of High Energy Physics Phenomenology papers. A directed link from paper i to j indicates that paper i cites paper j [35], [36].

The network statistics are shown in Table 1.

For the seed (multi)set size, for a network with n nodes we perform the intended influence maximization algorithm with seed (multi)set sizes up to $\frac{n}{\log n}$.

Table 1. Network statistics

Network	#nodes	#edges
Facebook	4,039	176,468
Wiki-Vote	7,115	103,689
Email-Enron	36,692	367,662
Epinions	75,879	508,837
Slashdot	77,360	905,468
DBLP	317,080	2,099,732
CA-GrQc	5,242	28,980
CA-HepTh	9,877	51,971
CA-HepPh	12,008	237,010
CA-AstroPh	18,772	396,160
CA-CondMat	23,133	186,936
Cit-HepPh	34,546	421,578

B. Edge Weight Models

As is common in the influence maximization research on the IC model, for the edge weights we use the following two models:

- **Weighted Cascade (WC) model:** In the WC model, the influence probability of each edge is assigned to $P_{uv} = 1/d_v$, where d_v is the in-degree of v [6].
- **Trivalency (TR) model:** This model assigns a randomly selected probability from {0.1, 0.01, 0.001} to each directed link [15].

The above models for the edge weights do not consider the transitivity behavior observed in real world social networks. In what follows we propose a new *Transitive Multi-valency* model that does so.

C. Transitive Multi-valency Model

Triadic closure [37] and clustering coefficient [38] in the social network theory are two strongly related concepts that demonstrate the transitive behavior in social networks. In the context of influence spread, the transitivity of nodes influence on their neighbors can be stated as follows:

Influence Transitivity: The influence of a node u on a neighbor v is dependent to the portion of v directly influencing nodes who are themselves directly influenced by u.

In other words, let the set of v directly influencing nodes (also known as its in-neighbors) is shown by N_v^- and the set of nodes directly influenced by u (also known as its out-neighbors) is shown by N_u^+. The influence transitivity states that

$$e_{uv} = f\left(\frac{|N_v^- \cap N_u^+|}{|N_v^-|}\right), \quad (5)$$

where e_{uv} is the weight of the edge connecting u to v.

In this study we use our new edge weight model based on the influence transitivity which sets the edge weights as follows

$$e_{uv} = \left(\frac{|N_v^- \cap N_u^+|}{|N_v^-|}\right)^{1.5} \cdot R_{uv} \quad (6)$$

where R_{uv} is a random multi-valency attenuator chosen uniformly at random from the set $\{3^{-2}, 3^{-3}, 3^{-4}, 3^{-5}\}$. This attenuator together with the 1.5 exponent are included to avoid the influence saturation due to the large edge weights. In our experiment results, this model is represented by TMv.

IV. Experiments

The experiments performed in this study are discussed in this section. First, the impact of reselection possibility on different networks and different edge weight models is analyzed. Then, the correlation between RG and influence saturation measures is studied. Finally, a significant entanglement between RG and network structural parameters is identified.

A. Reselection Impact

In this section, before studying the relation between the previously defined parameters, we show the impact of the reselection with varying fading values on different networks. Based on their influence spread behavior in response to the reselection possibility, we categorize the network into the three following cases:

- **Reselection-friendly networks**: When the reselection gain in a network without any fading (α=1) is more than 1.5 we call it a reselection friendly network. In these networks the possibility of the reselecting the nodes increases the influence spread more than 50% compared to the simple case. A simple example of a reselection friendly network is a star graph consisting of a core node and a number of pairwise disjoint nodes connected only to the core node. Obviously, reselecting the core node multiple of times has an outstanding gain compared to the simple case where the core node can only be selected once.
- **Reselection-aware networks:** In the absent of fading (α=1) when the reselection gain of a network lies between 1.05 and 1.5, the network is called to be reselection aware. The impact of reselection on these networks is not as impressive as the previous case; but it is noticeable.
- **Reselection-free networks:** These networks have a reselection gain less than 1.05. In the reselection free networks the multiset obtained by solving the reselection possible influence maximization hardly differs from the solution of the simple influence maximization case. A good example of such networks is a clique with uniform influence probabilities. In a fully connected network all the nodes share the same set of neighbors and reselection of a node has almost the same influence as selecting a new node.

Figure 2 plots the changes of the reselection gain in terms of the fading parameter α when the influence probabilities are derived from the WC model. As can be seen in this figure, Facebook and Wiki-Vote networks are reselection friendly networks (Figure 2-a), CA-AstroPh, CA-CondMat, CA-HepPh and Email-Enron networks are reselection aware (Figure 2-b) and Cit-HepTh, CA-GrQc and CA-HepTh networks are reselection free (Figure 2-c). It is interesting to note that the reselection gain in the reselection friendly networks, even with a fading value as low as α=0.6 is still non-negligible.

When the TMv model (section III.C) is used, no tested network is reselection free. The rise of the reselection gain as a result of increasing α in the TMv model is demonstrated in Figure 3. Similar to the WC model, the reselection friendly networks show meaningful RG values even at $\alpha = 0.6$.

Surprisingly, no one of the tested networks in the TR model are reselection friendly. Figure 4 illustrates the change of reselection gain in terms of fading value α for the TR model. A comparison between Figure 2, Figure 3 and Figure 4 shows that the behavior of the networks is totally dependent to the influence probability model.

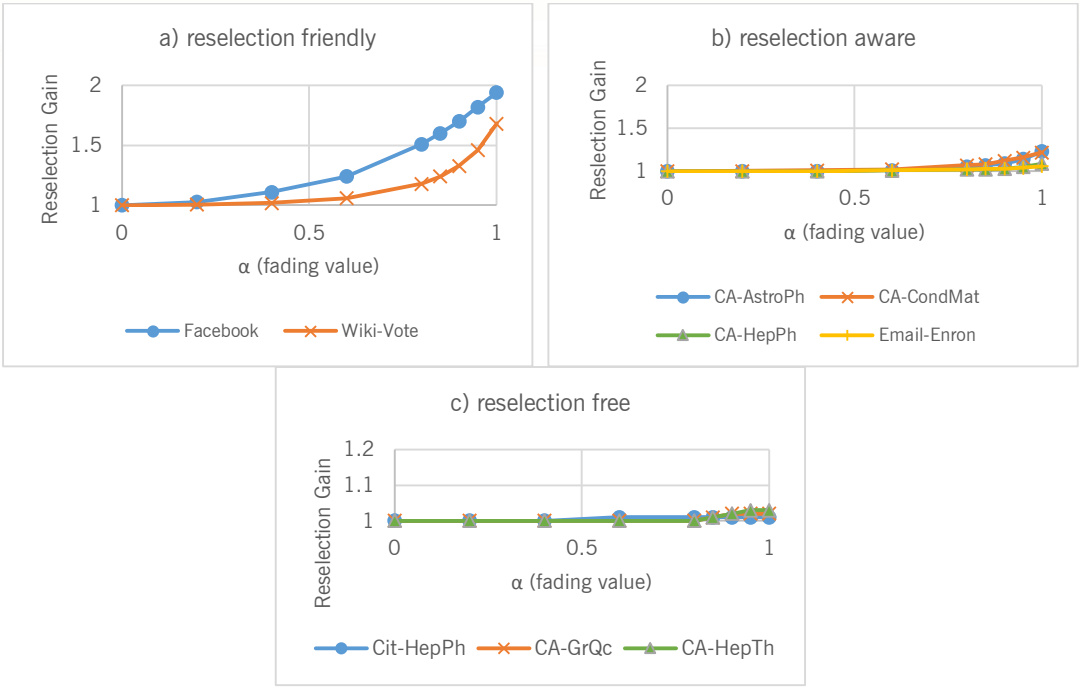

Figure 2. Reselection gain of different networks with WC model in terms of the fading value (α)

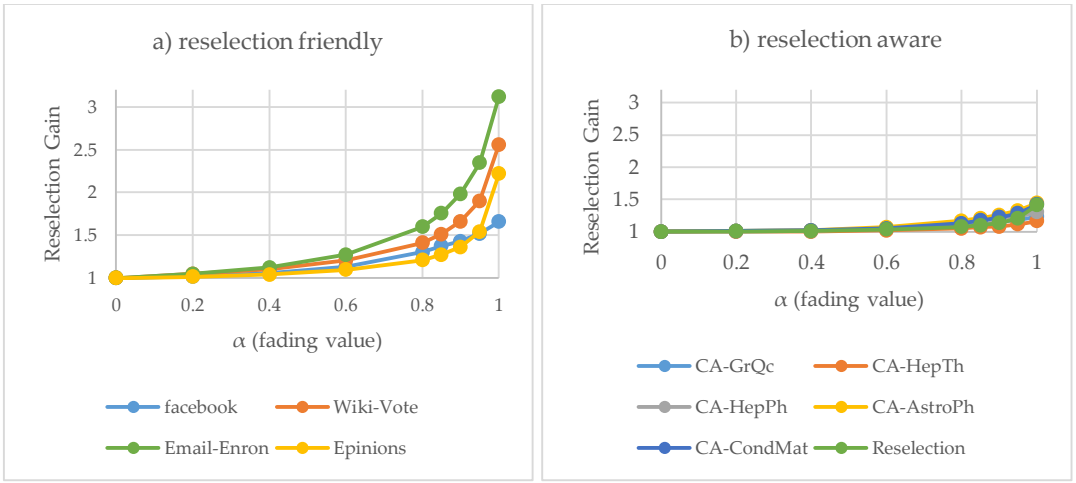

Figure 3. Reselection gain of different networks with TMv model in terms of the fading value (α)

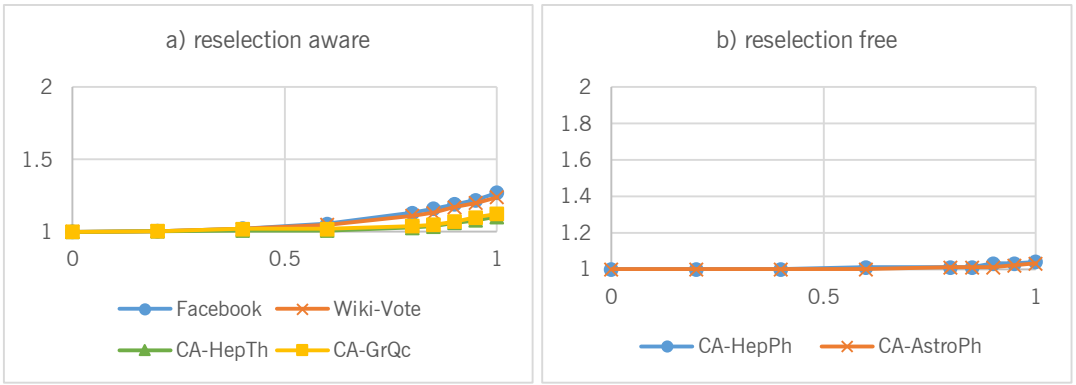

Figure 4. Reselection gain of different networks with TR model in terms of the fading value (α)

B. *Influence Spread Dynamics*

In this section, the relation between the reselection gain and the influence spread dynamics in the networks is experimentally studied. To do so, we have constructed the correlation between the influence saturation (LCIS) (Section II) and the reselection gain (RG) on the real-world networks.

When the LCIS parameter is high in a network, it means that the influence spread of the first seed node is considerably higher than the marginal influence spread of the next seed nodes. The structural interpretation of this dynamics is that the network contains a dense core with two important properties: (I) the density of the core is such that an influential node within the core can influence a great portion of the core; and (II) the strength of the connections from the in-core nodes to the out-core nodes is such that the activated core nodes can influence a great number of outer nodes.

On the other hand, a high RG ratio suggests the presence of strong hubs in the network. In the context of influence maximization, a hub usually has two properties: (I) it has a significant number of strong connections; and (II) its connections, when activated, can in turn influence a considerable number of nodes.

Even though the above situations for the cause of a high LCIS and a high RG does not necessarily translate to each other, they have a positive correlation in real world networks with the WC model. On the contrary, when the TR or TMv models are considered, no meaningful correlation is observed between the LCIS and RG.

Figure 5 shows the influence spread in the simple and reselection cases in a number of our tested networks in the WC model. It also contains the linear approximation of the $\tau(k)$ function. As can be seen in this figure, networks such as Facebook and Email-Enron with a sharp saturation have a high RG ratio, while CA-GrQc and CA-HepTh with a smooth saturation have a RG ratio near the unity.

Table 2 shows $LCIS$ and RG parameters of the networks. Using the values presented in Table 2, the RG ratio has a significant correlation of about 0.74 to the LCIS parameter.

Table 2. The LCIS and RG parameters of the networks with WC model

Network	*LCIS*	*RG*
Facebook	3.6	1.99
CA-GrQc	0.61	1.08
Wiki-Vote	1.76	1.58
CA-HepTh	0.59	1.08
CA-HepPh	1.51	1.16
CA-AstroPh	1.75	1.2
CA-CondMat	1.76	1.17
Cit-HepPh	2.29	1.12
Email-Enron	2.56	1.30
Epinions	2.84	1.18
Slashdot	3.35	1.67
DBLP	0.38	1.04

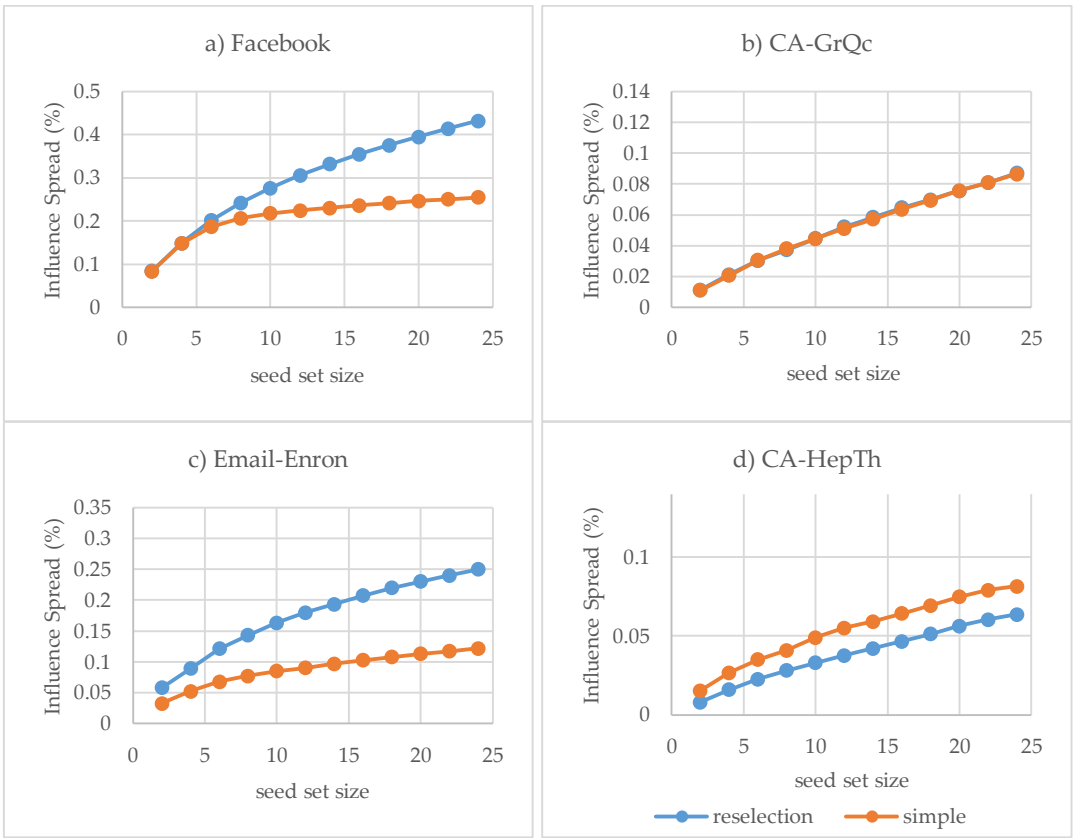

Figure 5. Influence spread in the simple case versus the reselection case (WC model)

This relatively high correlation value demonstrates that usually the reselection-friendly networks are the ones with sharp saturation, while the reselection-free networks usually have a smooth saturation.

As stated earlier in this section, unlike the WC model, the TR and TMv models do not exhibit a meaningful correlation between the LCIS and RG parameters. Several network parameters such as the degree distribution, maximum weighted degree, stepped w-core index distribution [39] and maximum stepped w-core index have been examined to identify the source of this difference. The construction of the WC model for the edge weights (i.e. the sum of input weights to all of the nodes equals to the unity) leads to a maximum stepped w-core index very close to one in almost all the tested networks. This observation led us to the following hypothesis.

Hypothesis: The LCIS and RG parameters are significantly correlated in the networks whose w-core index is close to one.

In order to test the above hypothesis we collected the networks with w-core index within the interval of 1 ± 0.15 as the gold set. The statistics of the gold set is shown in Table 3. The correlation between LCIS and RG in the gold set is 0.73 with a confidence of 1. This partially verifies our hypothesis on the tested networks. It remains to show that the networks outside the criteria of the gold set (i.e. w-core outside the 1 ± 0.15 interval) damage the correlation between LCIS and RG. For each of the networks outside the gold set, we have constructed the correlation between LCIS and RG on the union of the gold set and that specific network. The results are expressed in Table 4. This table contains the statistics of networks outside the gold set. Furthermore, for each network the correlation is constructed on the union of that network and the gold set and the result is included in the table. The last column of Table 4 contains the difference of the correlation after and before (i.e. 0.73) the insertion. As is shown in this table, the insertion of all networks but the Facebook with TMv model significantly damage the correlation on the gold set. This verifies our hypothesis on the tested networks.

Table 3. Tested networks with maximum w-core index within 1 ± 0.15 (gold set)

Network	Model	w-core	LCIS	RG
Facebook	WC	0.858	3.60	1.99
CA-GrQc		0.995	0.61	1.08
CA-HepTh		0.988	0.59	1.09
CA-HepPh		0.998	1.51	1.16
CA-AstroPh		1	1.76	1.20
CA-CondMat		0.99	1.77	1.17
Cit-HepPh		0.998	2.29	1.12
Email-Enron		0.998	2.56	1.30
Epinions		0.99	2.85	1.18
Slashdot		0.914	3.35	1.67
DBLP		0.982	0.38	1.04
CA-GrQc	TMv	1.122	3.14	1.22
CA-HepTh		0.863	2.25	1.17

Table 4. Impact of inserting networks to the gold set on the correlation between LCIS and RG

Network	Model	w-core	LCIS	RG	Correlation after insertion	Impact on the gold set
Facebook	TMv	2.869	9.04	2.20	0.86	+0.13
Wiki-Vote		0.03	1.49	2.56	0.31	-0.42
CA-HepPh		6.476	20.47	1.28	0.17	-0.56
CA-AstroPh		1.318	5.17	1.43	0.67	-0.06
CA-CondMat		0.511	0.65	1.40	0.63	-0.10
Cit-HepPh		0.06	0.42	1.41	0.60	-0.13
Email-Enron		0.213	0.80	3.10	0.05	-0.68
Epinions		0.314	2.67	2.20	0.63	-0.10
Facebook	TR	4.17	18.16	1.31	0.22	-0.51
CA-GrQc		1.227	6.29	1.28	0.51	-0.22
Wiki-Vote		0.564	17.40	1.42	0.33	-0.40
CA-HepTh		0.797	0.97	1.29	0.69	-0.04
CA-HepPh		7.312	22.32	1.16	0.03	-0.70

In this section we have experimentally observed that the reselection gain is correlated to the influence saturation on the networks with w-core index close to unity. In the following section a degree distribution related parameter is introduced and shown to have a high entanglement with the reselection gain.

C. Network Structure

Earlier in Section II it was discussed that the star-like networks are candidates of reselection friendly networks. A star-like network can be characterized by two properties:

- A small number of highly connected core nodes;
- A great number of weakly connected border nodes;

After a thorough investigation of network parameters concerning the first property, no meaningful relation between the tested parameters and the reselection gain were found. But for the second property, a simple weighted degree test has been identified to have a significant correlation with the reselection gain.

We simply set a threshold for the weighted out-degree of the nodes and compute the percentage of the nodes with a weighted out-degree less than the threshold value. In what follows this quantity is shown by δ. One may argue about the selection of the threshold value. Figure 6 shows that the correlation between RG and δ is hardly sensitive to the threshold value and all the threshold values in the range from 3.0E-04 to 3.0E-03 can be chosen safely. For all of the correlations reported in Figure 6, the confidence level is greater than 0.994.

This experiment is performed over all of the 27 networks listed in Table 3 and Table 4 and shows that for a wide range of thresholds, the portion of the nodes with a weighted degree below the threshold value has a correlation greater than 0.63 with a confidence level above 0.994.

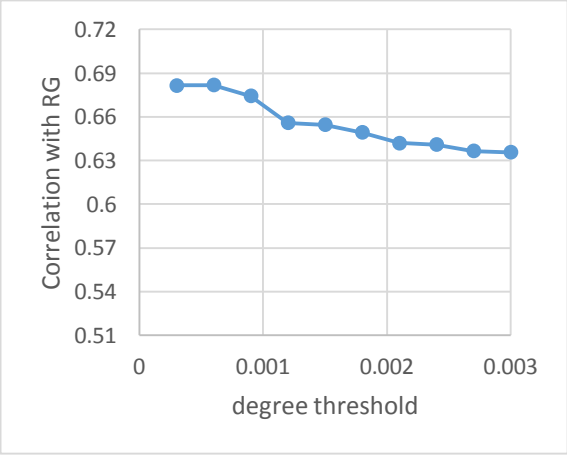

Figure 6. Sensitivity of the correlation between RG and δ to the degree threshold value

V. Conclusion

In this paper we have seen that considering the possibility of node reselection in the influence maximization, or equivalently targeting multiset of seeds instead of set of seeds can have diverse impacts on the influence spread in a number of networks. Based on the reselection gain of the influence spread, we have divided the networks into three groups, namely the reselection-free, reselection-aware and reselection-friendly networks. Our experiments have shown that the reselection gain can vary from 1 to 3.1 in different real-world networks.

We have correlated the reselection gain of networks to another influence maximization dynamics, called the influence saturation. We have shown experimentally that there is a 0.73 correlation between the reselection gain and the influence saturation in our tested networks in the WC model. More generally, we have experimentally verified the hypothesis that the networks with a w-core index close to unity have a high correlation between their reselection gain and influence saturation.

In order to make the propagating models more consistent with reality, we have introduced the transitive multi-valency (TMv) model which also considers the transitivity structures in the network. Consequently, our experiments were performed on the networks with three models for the edge weight: WC, TR and TMv.

Finally, in a search for detecting the star-like networks, we have shown a correlation of at least 0.63 between the reselection gain and the percentage of low degree nodes on a set of 27 networks with WC, TR and TMv models. We think that there are still room for analyzing the reselection gain dynamics in different networks. Finding a stronger entanglement between this dynamics and the network structure enables us to distinguish the reselection-friendly networks and choose our advertising strategies accordingly.

References

[1] C. Aslay, W. Lu, F. Bonchi, A. Goyal, and L. V. S. Lakshmanan, "Viral Marketing Meets Social Advertising : Ad Allocation with Minimum Regret," *Vldb*, no. Ces, pp. 814–825, 2015.

[2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.*, vol. 87, no. 3, pp. 925–979, Aug. 2015.

[3] A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, and M. Tiwari, "Global Diffusion via Cascading Invitations," in *Proceedings of the 24th International Conference on World Wide Web - WWW '15*, 2015, pp. 66–76.

[4] S. A. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*, 2012, p. 33.

[5] F. Morone and H. a. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[6] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the Spread of Influence through a Social Network," *Theory Comput.*, vol. 11, no. 1, pp. 105–147, 2015.

[7] M. Granovetter, "Threshold Models of Collective Behavior," *Am. J. Sociol.*, vol. 83, no. 6, pp. 1420–1443, May 1978.

[8] M. W. Macy, "Chains of Cooperation: Threshold Effects in Collective Action," *Am. Sociol. Rev.*, vol. 56, no. 6, p. 730, Dec. 1991.

[9] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Mark. Lett.*, vol. 12, no. 3, pp. 211–223, 2001.

[10] N. Alon, I. Gamzu, and M. Tennenholtz, "Optimizing budget allocation among channels and influencers," *Proc. 21st Int. Conf. World Wide Web - WWW '12*, p. 381, 2012.

[11] T. Soma, N. Kakimura, K. Inaba, and K. Ken-ichi, "Optimal Budget Allocation: Theoretical Guarantee and Efficient Algorithm," *Proc. 31st ...*, vol. 32, 2014.

[12] N. Avigdor-Elgrabli, G. Blocq, and I. Gamzu, "Offline and Online Models of Budget Allocation for Maximizing Influence Spread," pp. 1–15, 2015.

[13] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELF++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web - WWW '11*, 2011, pp. 47–48.

[14] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "SIMPATH: An Efficient Algorithm for Influence Maximization under the Linear Threshold Model," in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 211–220.

[15] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, 2010, p. 1029.

[16] K. Jung, W. Heo, and W. Chen, "IRIE: Scalable and robust influence maximization in social networks," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 918–923, 2012.

[17] S. Cheng, H. Shen, and J. Huang, "StaticGreedy: solving the scalability-accuracy dilemma in influence maximization," *Proc. 22nd ...*, p. 10, 2013.

[18] E. Cohen, D. Delling, T. Pajor, and R. F. Werneck, "Sketch-based Influence Maximization and Computation: Scaling up with Guarantees," *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manag. (CIKM '14)*, no. Ic, pp. 629–638, 2014.

[19] Y. Tang, X. Xiao, and Y. Shi, "Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency," *In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, pp. 75-86. ACM, 2014.*.

[20] Y. Tang, Y. Shi, and X. Xiao, "Influence Maximization in Near-Linear Time," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD '15*, 2015, pp. 1539–1554.

[21] S. Cheng, H.-W. Shen, J. Huang, W. Chen, and X.-Q. Cheng, "IMRank: Influence Maximization via Finding Self-Consistent Ranking," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. ACM*, 2014.

[22] D. Hatano, T. Fukunaga, and K. I. Kawarabayashi, "Adaptive budget allocation for maximizing influence of advertisements," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-January, pp. 3600–3608, 2016.

[23] W. D. Blizard, "Multiset theory.," *Notre Dame J. Form. Log.*, vol. 30, no. 1, pp. 36–66, Dec. 1988.

[24] P. C. Hansen and D. P. O'Leary, "The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.

[25] M. J. Anderson and J. Robinson, "Permutation Tests for Linear Models," *Aust. New Zeal. J. Stat.*, vol. 43, no. 1, pp. 75–88, Mar. 2001.

[26] J. Leskovec, "Stanford Large Network Dataset Collection," 2016. [Online]. Available: https://snap.stanford.edu/data/index.html.

[27] J. Leskovec and J. Mcauley, "Learning to discover social circles in ego networks," *Adv. neural Inf. Process. ...*, pp. 1–9, 2012.

[28] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed Networks in Social Media," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 1361–1370, 2010.

[29] J. Leskovec, D. Huttenlocher, and J. M. Kleinberg, "Predicting Positive and Negative Links in Online Social Networks," *Conf. World Wide Web (WWW '10)*, pp. 641–650, 2010.

[30] B. Klimt and Y. Yang, "Introducing the Enron Corpus," *Mach. Learn.*, vol. stitutep1, p. wwceasccaers2004168, 2004.

[31] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," Oct. 2008.

[32] M. Richardson, R. Agrawal, and P. Domingos, "Trust Management for the Semantic Web," 2003, pp. 351–368.

[33] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, Jan. 2015.

[34] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 2, pp. 1–39, 2007.

[35] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations," *KDD*, pp. 177–187, 2005.

[36] J. Gehrke, P. Ginsparg, and J. Kleinberg, "Overview of the 2003 KDD Cup," *ACM SIGKDD Explor. Newsl.*, vol. 5, p. 149, 2003.

[37] M. S. Granovetter, "The Strength of Weak Ties," *Am. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, May 1973.

[38] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[39] A. Vardasbi, H. Faili, and M. Asadpour, "SWIM: Stepped Weighted Shell Decomposition Influence Maximization for Large-Scale Networks," *ACM Trans. Inf. Syst.*, vol. 36, no. 1, pp. 1–33, Aug. 2017.

AUTHORS' INFORMATION

Ali Vardasbi finished his Bachelor and Master of Science in Commnunications in Sharif University of Technology and His Ph.D. in Software Engineering at University of Tehran. His research area is mainly Natural Language Processing and Social Networks.

Heshaam Faili, had finished his Bachelor and Master of Science in Software Engineering in Sharif University of Technology and his Ph.D. in AI at the same university. Upon joining to university of Tehran in 2008, he had established Natural Language and Text Processing Laboratories in ECE department. His main research interests are related to Text Processing, focused on Natural Language Processing, Social Network and Text Mining.

Masoud Asadpour received his PhD in Machine Learning and Collective Robotics from EPFL, Switzerland, in 2006. He has been a researcher at Institute for Studies on Theoretical Physics and Mathematics (IPM), Iran, from 1998 to 2001 and again from the beginning of 2010 up to now. He has been a postdoc researcher in Biologically Inspired Robotics Group (BIRG), EPFL, Switzerland, in 2007. In 2008, he joined the Robotics and AI group in the Faculty of Electrical and Computer Engineering, University of Tehran as a faculty member. He is the head of Social Networks Lab there and his research interests are Social Networks, Data Mining, Big Data Analysis, Bioinspired Algorithms and Machine Learning.