

A Holistic Approach to Influence Maximization

Nireshwalya Sumith, Basava Annappa, and Swapan Bhattacharya

Abstract A social network is an Internet-based collaboration platform that plays a vital role in information spread, opinion-forming, trend-setting, and keeps everyone connected. Moreover, the popularity of web and social networks has interesting applications including viral marketing, recommendation systems, poll analysis, etc. In these applications, user influence plays an important role. This chapter discusses how effectively social networks can be used for information propagation in the context of viral marketing. Picking the right group of users, hoping they will cause a chain effect of marketing, is the core of viral marketing applications. The strategy used to select the correct group of users is the influence maximization problem.

This chapter proposes one of the viable solutions to influence maximization. The focus is to find those users in the social networks who would adopt and propagate information, thus resulting in an effective marketing strategy. The three main components that would help in the effective spread of information in the social networks are: the network structure, the user's influence on others, and the seeding algorithm. Amalgamation of these three aspects provides a holistic solution to influence maximization.

Keywords Algorithm • Diffusion • Influence maximization • Social networks • Viral marketing

1 Introduction

With the advent of Web 2.0 came a range of applications that are used in many ways by people across different sections of the society. The social network is one such application that plays a very important role across the world. It is not just a platform for sharing ideas, it is also seen to play an important role in the economic growth of

N. Sumith (✉) • B. Annappa
National Institute of Technology Karnataka, Surathkal, Mangaluru, India
e-mail: s.nireshwalya@gmail.com; annappa@ieee.org

S. Bhattacharya
Jadavpur University, Kolkata, India
e-mail: bswapan2000@yahoo.co.in

the country. The term social economics reflects the importance of social networks in economic transactions. In the era of cloud computing, social media has proved to be a more effective business-related strategy [36].

Often, influence among friends plays an important role in product adoption decisions. An individual's choice to adopt or reject a product is often linked to his/her peers' choices. The term network externalities embodies such choices. Undoubtedly, this trend is seen in social networks and is used to popularize a product in the network to increase sales. An early attempt to model network structures, with some perspective on their impact on economic outcomes, is seen in the cooperative game theory literature. The game theory relies on the premise that users can cooperate only when they are connected. To understand these connections, graphs are employed. People who can communicate can cooperate and generally cooperation leads to higher production or utility than separate efforts [37]. Thus, graph representations became an important part of game theory and social network analysis.

In this chapter, the role of social networks in the context of viral marketing is discussed. The success of viral marketing depends on the strategy used to select initial adopters, network structure, and Influence among users. These aspects are discussed in detail in this chapter. An outline of the chapter is as follows. Section 2, discusses various cases in which viral marketing is employed. Section 3 introduces the influence maximization problem. Section 4 analyzes social networks to obtain a new evaluation metric, followed by details of the proposed approach in Sect. 5. Section 6 summarizes results and a conclusion constitutes Sect. 7.

2 Viral Marketing in the Real World

With billions of users of social network sites, they have become the most powerful tool for marketing. User involvement has made viral marketing, tailored to social networks, to be more dominant than the traditional marketing approaches. The strategy where individuals forward the message to others, creating a vast spread of information and influencing others to adopt it and spread further is popularly referred to as viral marketing. The brand awareness thus created by viral marketing is cost-effective and generates requests for products. The practice of viral marketing in the digital era has been around for more than a decade. The early adopters of the viral marketing strategy were Hotmail, which grew to 12 million users in 18 months, and the John West salmon bear advertisement, to name a few. Although these campaigns were accidentally successful, they were not well planned. The low expenditure on popularizing products is the main reason for enterprises to adopt this strategy. In the following sections, three popular cases across various domains, in which viral marketing created a success story are discussed.

2.1 Case Study 1: Fiesta Ford Movement

Ford had made several attempts to market a small car since the discontinuation of the Aspire in 1997, but without much success. In 2009, Ford Motors launched the Fiesta Movement campaign [34] to promote sales. For 6 months, Ford gave 100 people a car to use and asked them to write about their experiences on social media. Consumers used their Fiestas for various activities and some went for adventures. These consumers wrote about their experiences on YouTube, Flickr, Facebook, and Twitter. The social media audience took great interest in these blogs and soon it resulted in massive sales of the Fiesta.

The Fiesta Movement was the most successful social media marketing experiment for the automotive world. The campaign news was all over the social media with 6.5 million YouTube views alone and 50,000 queries about the car from new customers. In first week of the campaign, Ford sold nearly 10,000 cars. The Fiesta Movement cost the company only a small amount compared with the typical traditional TV campaign. In 2014, Ford used this strategy to introduce their latest Fiesta.

2.2 Case Study 2: Why So Serious?

In 2008, the “Why So Serious?” campaign, an augmented reality game (ARG) was launched to promote the movie, *The Dark Knight* [52]. Over 10 million people participated in this campaign, which was launched 15 months before the release of the movie. Various games and rewards were available all over social media and participants took great interest in these. The ARG was thus able to maintain fan interest up to the release of the movie. Millions of blog posts were seen on social media, resulting in success of the ARG and leading to the success of the film, which made over US\$ 1 billion in box office collections.

The Dark Knight Rises promotion also used a similar campaign. This time the participants were given graffiti to help the Gotham City Police Department find *Batman*. For every piece of graffiti found and tagged on social media, a frame of the trailer would be released. This marketing strategy, because of the massive fan interest, led to completion of the task within a few hours.

2.3 Case Study 3: Ice Bucket Challenge

In 2014, to promote awareness of Amyotrophic Lateral Sclerosis (ALS), the “The Ice Bucket Challenge” campaign was designed [14]. In this challenge, a person needed to pour a bucket of iced water over their head, film it and upload it. A person who did not accept the challenge had to donate to ALS cause within 24 h. Once the participant had either been soaked or had donated, this challenge had to be passed to three friends.

This campaign was popular on Facebook and Twitter, with over 2.4 million tagged videos and 2.2 million Tweets respectively, about the challenge. The views per month grew from 0.16 million views, to over 2.89 million per month in August 2014. Because of this, huge donations to ALS were received. The ALS fund had received over \$40 million from seven hundred thousand donors within 30 days. The ALS association declared the total donation received to be around \$100 million.

There are a number of similar successful cases where a social network was used to effectively promote information for various causes. User involvement in social networks is the driving force behind these successful campaigns. In the following sections, viral marketing is presented as an optimization problem and a solution is proposed.

3 Influence Maximization Problem

There are many cases in which enterprises have created a success story with a viral marketing strategy. The keys to these viral marketing campaigns are those first few users who started the campaign. These initial users were picked by the enterprise based on various criteria. Whatever the strategy was, the outcome was aimed at creating successful results. Therefore, these individuals should be picked with proper planning. Picking these individuals is referred to as the influence maximization problem. Figure 1 shows the process of information propagation by the initial user. Influence maximization is aimed to obtaining a good-quality seed set to maximize the spread of information in the social network. Formally, the problem discussed in this chapter is defined as follows.



Fig. 1 Information spread phenomenon in social network

Influence maximization [57]: Given a cost k and a social network, which is represented as a directed graph $G = (V, E)$, the goal is to find a seed set of k users such that by initially targeting them, the expected influence spread (in terms of expected number of adopted users) can be maximized.

Social networks play a fundamental role as a medium for the diffusion of information and ideas. This diffusion of information can be modelled to understand and answer many of the questions that arise in the real-world application. The independent cascade model (ICM) is a popular model used to understand the diffusion process in the network and is explained as follows.

Independent Cascade Model[24]: Suppose that node u is influenced (i.e., becomes active) at a time t . Then, u has an opportunity to influence every one of its neighbors v with probability $p(u, v)$. If u succeeds in activating v , then v is active from time $t+1$ onwards. If not, u can never try to influence v in subsequent attempts. This process continues until no new node becomes active at the end of the diffusion process.

4 Analyzing the Social Network

The initial part of the section discusses various existing approaches to evaluating a user to rank him/her to be the probable initiator in a viral marketing campaign. In the later part, a new metric for evaluating social network users is introduced.

4.1 Existing Centrality Measures for Evaluating Users

The centrality measures are commonly used approaches to picking up information initiators for applications that include viral marketing and recommendation systems. In this section, popular centrality approaches are discussed. The most popular centrality measures to measure the importance of a node are degree centrality, closeness centrality, betweenness centrality [43]. The degree centrality assumes that a node that has many direct connections is at the center of the network and plays an important role in information spread. In the context of social networks, degree centrality represents the number of contacts of a user. The closeness centrality focuses on how close a node is to all other nodes in the network. This metric refers to the number of friends separating two individuals. An individual may be

linked to a larger portion of the network through a few popular direct friends. This individual himself may have a small degree centrality. The betweenness centrality assumes that if a node is more frequently in the shortest paths between other nodes, it is more important to the network. This metric indicates the power to forward or delay requests between two unfamiliar individuals. 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 [12]. This metric indicates the popularity of an individual in the context of social networks.

There are also other metrics such as PageRank [39], hyperlink-induced topic search(HITS) [28], Birnbaum's component importance (BCI) [1] that rank the nodes individually based on their importance. In their basic form, PageRank and HITS value a node merely according to the graph topology [59]. The concept of a hub is prevalent in identifying key users. Users who are in a hub position are characterized by a great potential for communication and interaction within a network[19]. However, in the real-world networks, users who are connected to the most number of users do not show high interaction rates. The concept of the hub also fails to understand the diffusion mechanism.

Centrality measures are suitable for identifying initial information propagators in typical computer networks. In these networks, every receiving node functions as a sender to all its neighbors, those that meet the stated conditions. However, in a social network, it is more of an individual's choice to spread information to certain neighbors. Therefore, these centrality measures may not be suitable neither for evaluating users nor for selecting initiators for information diffusion in social networks.

4.2 Interaction Rate as a Metric to Evaluate Users

In this section, a new metric to evaluate users in a social network is discussed. To understand the need to introduce this new metric, four standard datasets High energy physics (HEP),¹ Physics -Theory (PHY) (see footnote 1), Wikivote (see footnote 1) and YouTube,² whose description is in Table 1, are analyzed. For the HEP and PHY datasets, interactions were not available and were synthesized on a power law distribution pattern, which can be produced using MATLAB or a similar tool.³ The pseudocode for synthesizing such data is in Algorithm 1. The function *randpower*(1, *n*) generates *n* random numbers on power law distribution. To understand the role of users in the network, their friend count, popularly known

¹<https://www.microsoft.com/en-us/research/people/weic/#publications>

²<http://socialcomputing.asu.edu>

³<http://tuvalu.santafe.edu/~aaronc/powerlaws>

Table 1 Dataset description

| Name | No. of nodes | No. of edges | No. of interactions |
|---------------------|--------------|--------------|---------------------|
| High energy physics | 15,233 | 58,891 | 588,136 |
| Physics-theory | 37,154 | 231,584 | 2,315,840 |
| YouTube | 15,088 | 76,765 | 2,239,440 |
| Wikivote | 8275 | 103,689 | 1,057,868 |

```

Input: List power, social graph G(V,E)
Output: List powerlist
Initialize  $power = \emptyset, powerlist = \emptyset;$ 
 $power\{\} = randpower(1,n)$  // where  $|V| = n$ 
for  $i = 0$  to  $i < n$  do
    |  $edgelist[i] = e(u_i, v_i)$ 
end
for each  $r \in power$  do
    |  $powerlist = powerlist \cup edgelist[r]$ 
end
    
```

Algorithm 1: Pseudocode to synthesize interactions

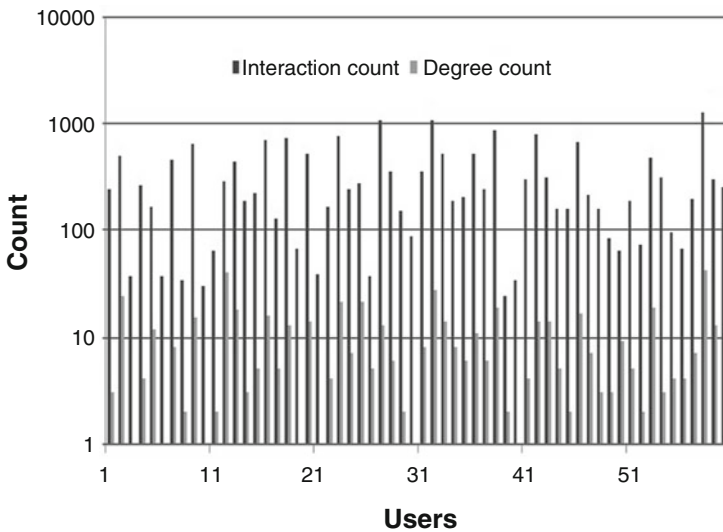


Fig. 2 High energy physics dataset

as “degree of the node”, and their interaction count are analyzed. Figures 2, 3, 4, and 5 show the degree count and interaction count of users for the chosen datasets.

When the activities of the users were analyzed, it was observed that a very large portion of the users did not actively take part in the network activities. Instead, a very small portion of this network was involved in these activities. To this end, it is evident that a high number of interactions come from users who maintained a low friend count (small degree). Also, a striking observation can be made that

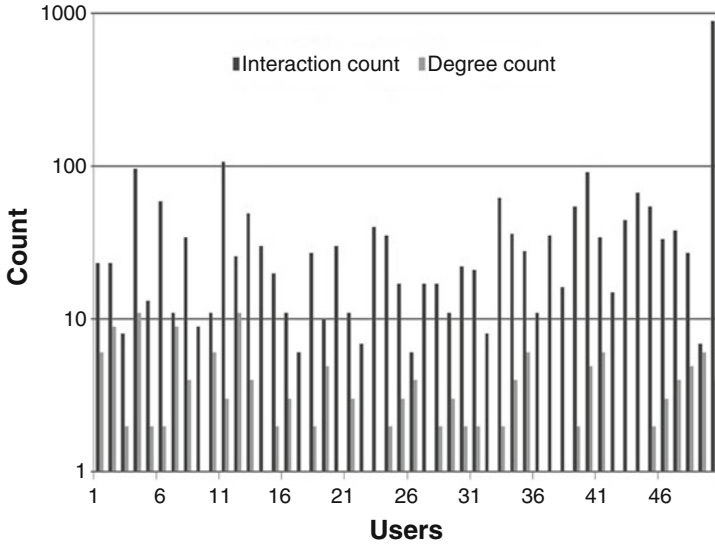


Fig. 3 Physics-theory dataset

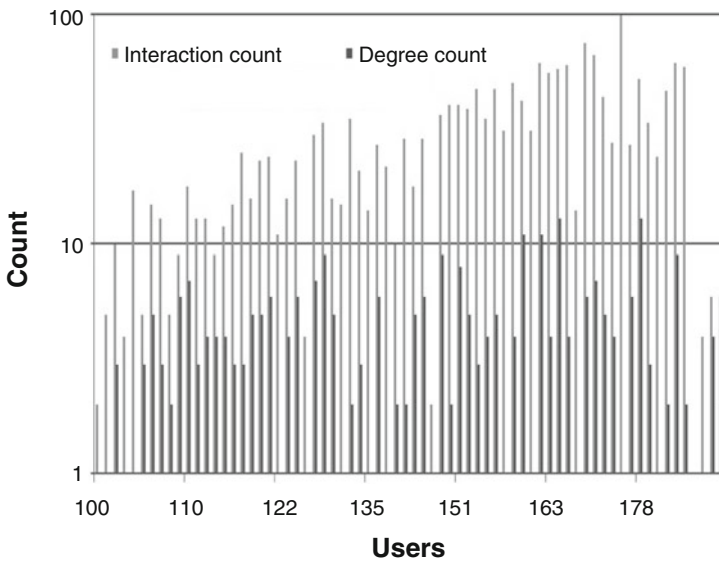


Fig. 4 YouTube dataset

the users with a high degree did not interact well enough among their friends. This observation can also be seen in various other content-based social networks. Conclusively, the traditional approach of evaluating a node with regard to its degree (as discussed in the previous subsection) would not produce accurate results. On

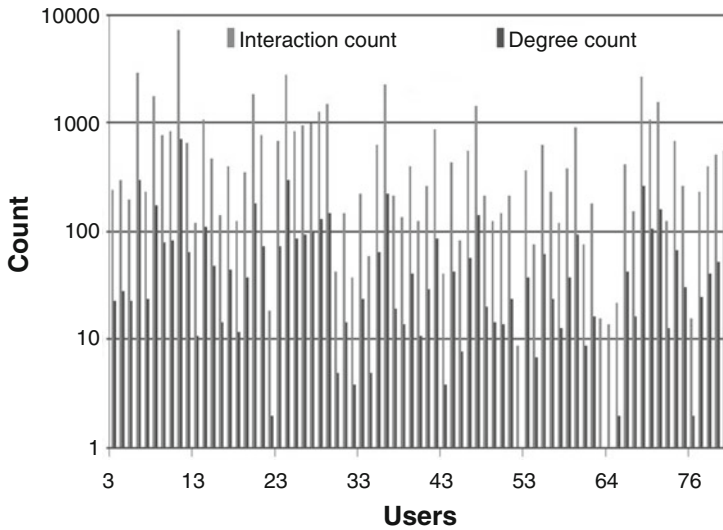


Fig. 5 Wikivote dataset

account of this observation, the interaction count of the users is considered to be an appropriate evaluation metric for applications that rely on user involvement, such as viral marketing.

5 Solving Influence Maximization in a Holistic Approach

Interactions among the users is an important attribute of the network. As most of the social networks, blogs, and forums are content-based models, interactions among users cannot be overlooked. Active users play a vital role in the spread of information or marketing. Based on this premise, a solution to influence maximization is developed that is based on user involvement.

The previous works solve influence maximization either by heuristics or by estimating parameters such as influence. However, the spread of information depends mainly on three factors: network structure, user influence, and seeding strategy. These aspects are explored and a three-stage solution is proposed to solve influence maximization. The growing size of social networks is a major hindrance to analyzing the effectiveness of any algorithm in general. The complexity of an NP-hard problem such as influence maximization increases drastically with an increase in the size of the social network. Therefore, in the first stage of the solution, the scalability issue is tackled by pruning the social network to ascertain the real contributors in the diffusion process. In the second stage, as peer influence is a major factor in the adoption of information or products in social networks, an approach to

estimating user influence is proposed. Finally, a seeding strategy is suggested that selects top influential users to initiate the diffusion process to have an effective spread of information. Thus, the holistic approach is an amalgamation of these aspects and is discussed in detail in the following sections.

5.1 Stage 1: Pruning the Social Network

Popular social networking sites such as Google+, Friendster, Flickr, Facebook, Yahoo, Twitter, etc., have grown from a few users to billions of users. Statistics reveal that the number of social network users has increased from 0.97 billion in 2010 to 1.82 billion in 2014 [48]. These numbers are sure to rise in the coming years, clearly showing evidence for the fact that social networks are growing rapidly. With this rapid growth, comes the gigantic amount of data in various forms, posing a great challenge to data analysis and implementation of an influence maximization solution. Therefore, there is a need to simplify the social networks.

5.1.1 Existing Network Simplification Approaches

Various network pruning strategies, such as to maintain connectivity[60], shortest path[58], source to link flow[35], triangular inequality [42], modularity [2], and cut sparsifiers [13] are seen in the literature. Serrano et al. [47] and Foti et al. [11], focus on weighted networks and select edges that represent statistically significant deviations with respect to a null model. An application of a pruning process to connectivity constraint is also seen in [33].

Although a large amount of literature is available on network simplification, these approaches are not suitable for simplifying the social network. There is a possibility of a decline in accuracy, with the increasing erroneous removal of nodes and edges[3]. In most cases, previous works use the structural properties of the graph during the pruning process without understanding whether a link is used for communication. Removing a connection edge from the social graph may lead to disturbance of its structural properties. Therefore, it is important that any metric sought out for pruning the social graph at the edge and node level should retain the properties required for efficient information propagation.

5.1.2 New Approach for Social Network Simplification

The social network is represented as graph $G(V, E)$, where V is the set of users and E is the set of edges that defines the underlying relationship. A link formed between the users is not a random link and indicates that two users are well connected in terms of similar interests and ideas. It is observed that a few of these links are used more often than others and these are the strong links that keep two individuals firmly

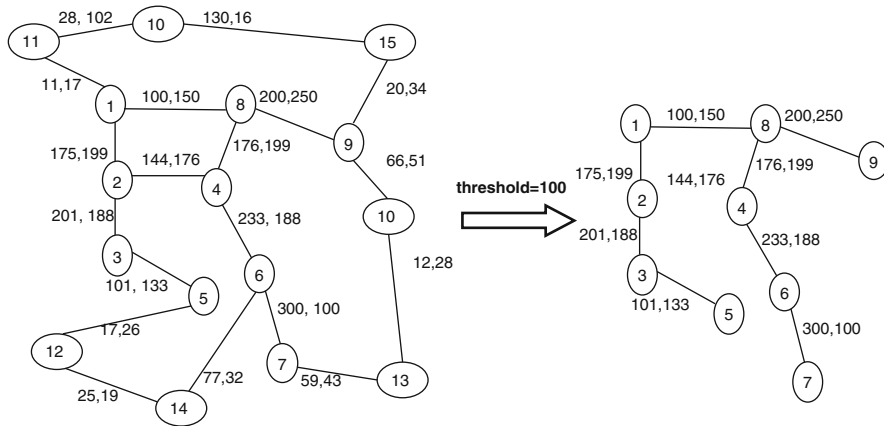


Fig. 6 Pruning of social graph

connected. These individuals have a greater than average potential to influence each other than any two randomly selected users. Therefore, there is a need to distinguish between contact edges and interaction edges. An immediate question to be answered is whether there is a need to keep all those contact edges that have been used at least once for communication. If it is decided to keep all these edges that have been used at least once, the pruning process will not be beneficial. Moreover, there may be a particular pair of users who have not been interacting or who have had very few interactions for a long time. In such a case, retaining this connection is not beneficial. Therefore, there is a need to find the minimum number of interactions needed to classify the user as a contributor or not. A threshold parameter, say *minimum interaction rate*, will remove all those contact edges that have very few interactions in a certain given time window. This threshold prunes the social network to its valuable components. Figure 6 shows the simplification of the original social graph $G(V, E)$ to produce a pruned graph $G_c(V_c, E_c)$.

Interactions play an important role in the dissemination of information. It is observed that every social network has one interaction that stands out over the other interactions. On Facebook, activity is the number of posts, on Twitter it is the number of tweets, on Flickr it is the photo uploads and on YouTube it is the uploads and downloads of videos. Determining the minimum activity rate of users is an open question. For this purpose, the distribution patterns of the interactions of the users in the networks are analyzed. These distribution patterns are shown in Figs. 7, 8, 9, and 10. These distribution patterns follow power law distribution [8]. When any distribution takes such a pattern, the statistical dispersion method of the central tendency, such as mean, mode and standard deviation, cannot be used to get the measure of dispersion. A more robust technique is required to be able to obtain a useful measure of dispersion in the presence of outliers. In the power law distribution pattern, the mean is much larger than the median and the mode, i.e., $mode < median < mean$. Hence, the mean, cannot be used as the measure of dispersion.

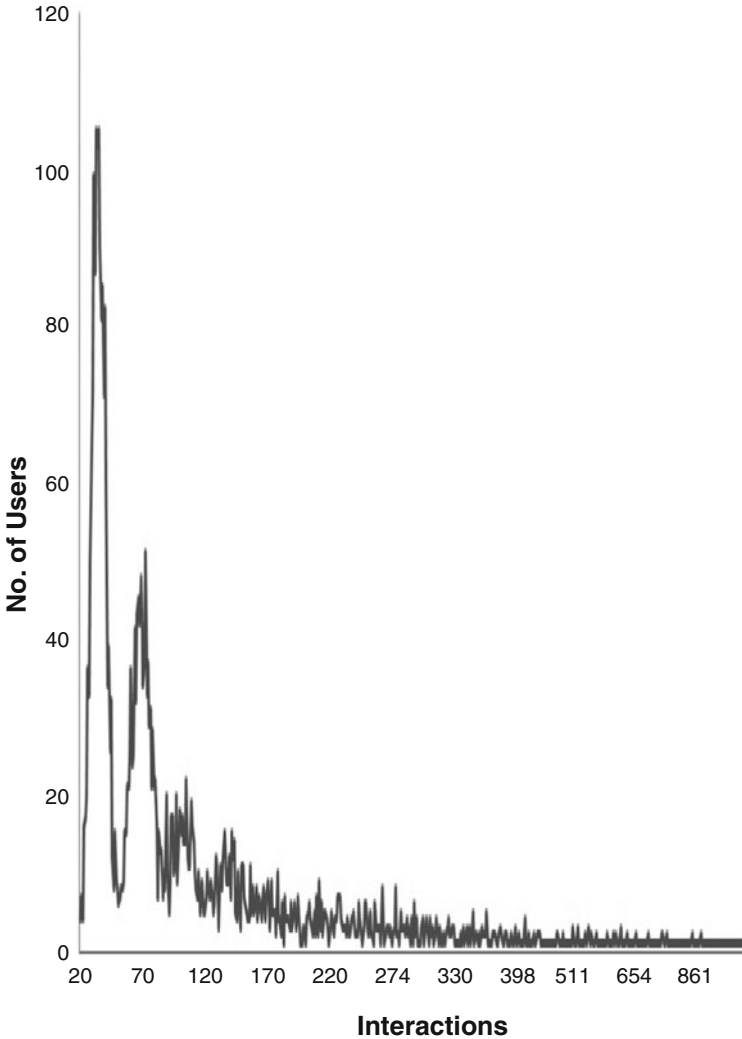


Fig. 7 HEP

For such a distribution, a robust dispersion measure would be median absolute deviation (MAD) [40]. The median absolute deviation approach effectively discards the outliers in the data compared with the standard deviation. Thus, it avoids the need to specifically remove outliers, making the approach less time consuming.

To compute the median absolute deviation, the median for a given population is determined first. Next, the absolute value of the distance between each separate observation and the median is computed. Finally, the median absolute deviation is obtained by computing the median of the values computed in the previous step.

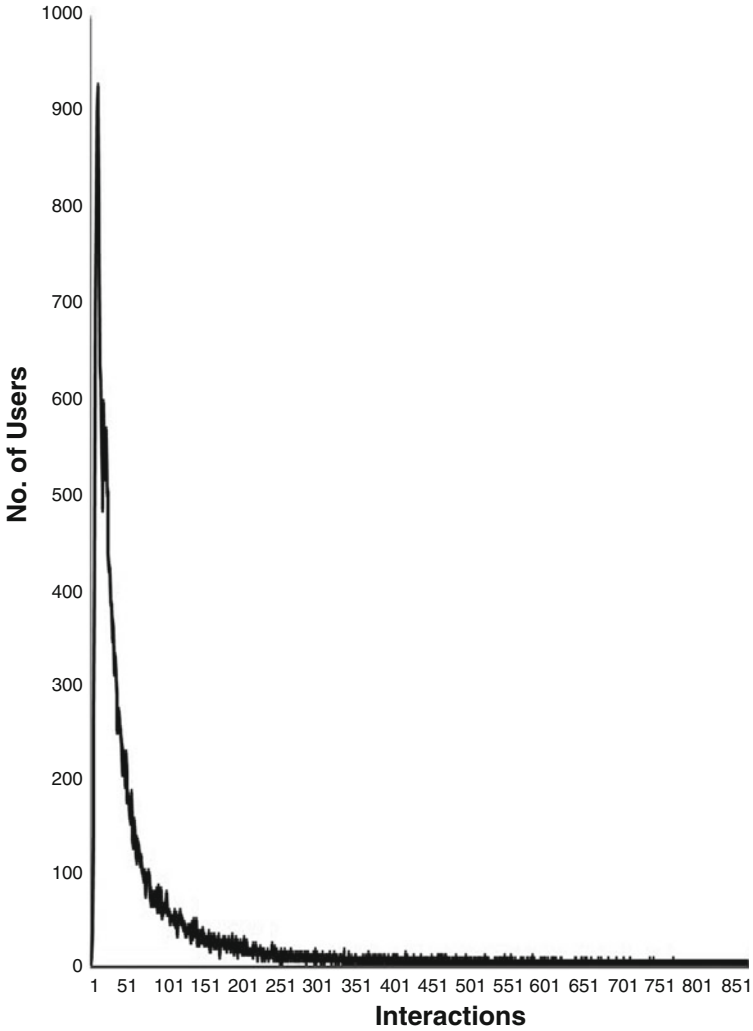


Fig. 8 PHY

More precisely, for a univariate data set X_1, X_2, \dots, X_n , the MAD is defined as the median of the absolute deviations from the median is given as in Eq. (1).

$$MAD = M(|x_i - M(x_j)|), \tag{1}$$

Social network user data are usually grouped where a value is repeated many times. For example, in the HEP dataset, there are 38,757 users who have interacted only once. Therefore, the frequency of one interaction is 38,757. Such a pattern is seen in almost all the social networks. This is a characteristic feature of power

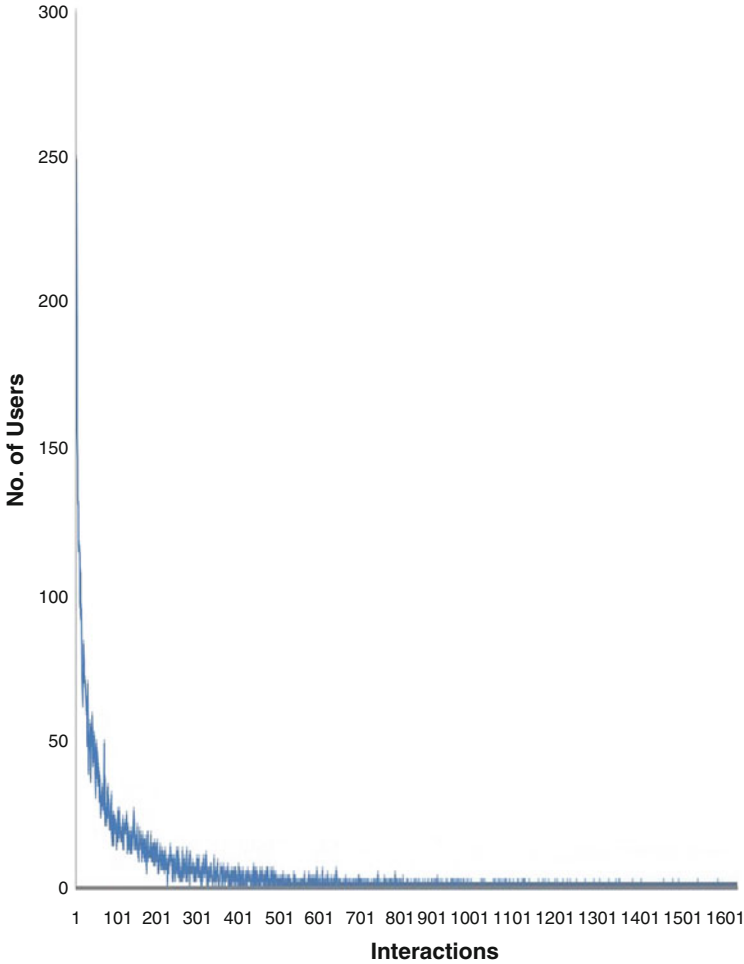


Fig. 9 YouTube

law distribution. For such datasets, the formula for calculating MAD is as given in Eq. (2)

$$\text{MAD} = M(f_i * (|x_i - M(x_i)|)), \quad (2)$$

In the context of the interaction, this would be the ideal count of interactions that would be used to identify a contributor from the rest of the network. As long as the distribution pattern matches the power law distribution, the choice of choosing MAD to define the minimum activity rate is justified.

The approach proposed here reflects the dynamic nature of the social network. Users who may have a high number of interactions within the chosen time period

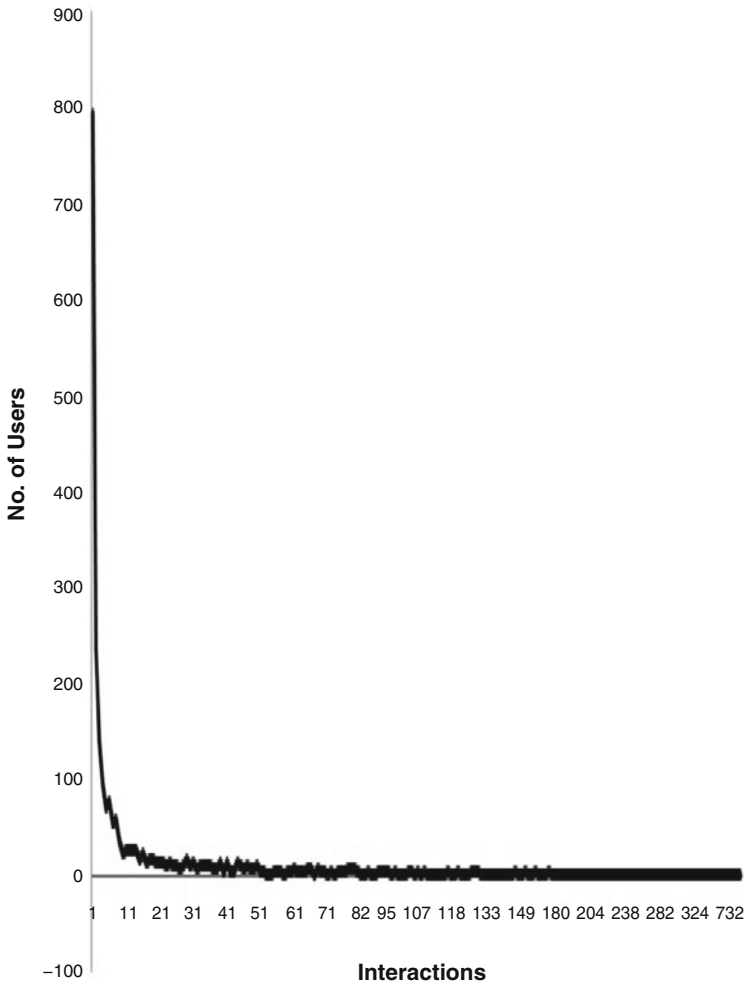


Fig. 10 Wikivote

are included in the pruned graph. By varying the granularity of the time period a close look at the interaction pattern in the social network can be obtained. Details of this approach are available in our previous work [50].

5.2 Stage 2: Estimating User Influence

The importance of influence is quite evident in viral marketing. There are many instances in which a group of people buy products because they were recommended

to do so by their friends. This attribute of a user influencing another user is used to advantage in viral marketing. One of the definitions of influence is as follows.

Social influence is the change in an individual's thoughts, feelings, attitudes, and behaviors that results from interaction from other people or group [31].

5.2.1 Background

The solution to the influence maximization problem starts with the weighted undirected social graph $G(V, E)$, where V is the set of users and E represents the set of edges. The weight on the edge represents the influence probability and is given a pre-assumed uniform value. In reality, the social graph is readily available, whereas the edge weight, i.e., influence probability, is not. There are two reasons for this, the use of such a pre-assumed value may not be an ideal setup in the solution. First, assuming information diffusion to be uniform among all contacts results not only in the overestimation of spread, but also leads to the non-optimal selection of influential users [54] resulting in biased outcome. Second, influence is a behavioral attribute that changes over time. Hence, this parameter should not be made constant.

Estimating influence and obtaining the most influential users are not separate issues. To predict information spread and to evaluate the performance of seed selection algorithms, it is important to estimate influence probability. There have been several attempts made in this direction, such as [10, 16, 20, 23, 27, 29, 33, 44, 45, 51, 53–56]. However, these approaches are resource-expensive and require accurate and in-depth user profile details, which in most of the cases are unavailable. Hence, the proposed approach to estimating influence has been developed, keeping in mind the privacy concerns of the users. The proposed approach uses data related to user activities that are readily available in the action log repositories.

5.2.2 Influx: An Approach to Estimating User Influence

In a social network, various interactions such as posts, likes, recommendations, etc., are seen. In the proposed approach, the type of interaction is not considered, but instead the focus is on knowing whether a pair of connected users interact well enough to influence each other. Further, the approach does not distinguish between a positive influence and a negative influence. However, these could be carried out as separate research.

An approach is developed to estimate the probability of influence of a node using the interaction count of a user. Consider a scenario in which a user A has five contacts B, C, D, E, and F. For illustration, let us assume that the numbers of interactions A has with B, C, D, E, and F are 30, 40, 50, 60, and 30 respectively. As the numbers of interactions of A with his neighbors are different, the probability that A can influence his/her neighbors is also different. Hence, the metric of influence probability of a user should include an aspect of user interaction with his/her neighbors.

The strength of an edge reflects the intensity of the interactions through the tie. The strength of the edge is represented as y_{ij} , which is the number of interactions from v_i to v_j . The number of interactions from user v_i to user v_j is not the same as from user v_j to v_i , i.e., $y_{ij} \neq y_{ji}$. Therefore, the probability of influence is not symmetrical either, i.e., $p_{ij} \neq p_{ji}$.

Thus, the probability of influence is quantified using the interaction count of a user. The probability of influence is calculated as in Eq. (3).

$$p_{u,v} = \frac{y_{u,v}}{S_{s=\{n \in N\}y_{u,s}}} \tag{3}$$

where, N is the set of nodes incident on node u and $y_{u,s}$ is the number of interactions of the node u to the incident node n . The normalization process, sets the value of P within the range (0,1], according to the definition of probability. With this approach, in the given example, $p_{A,B}$ is 0.147. Values on other edges are similarly obtained. This scenario is shown in Fig. 11.

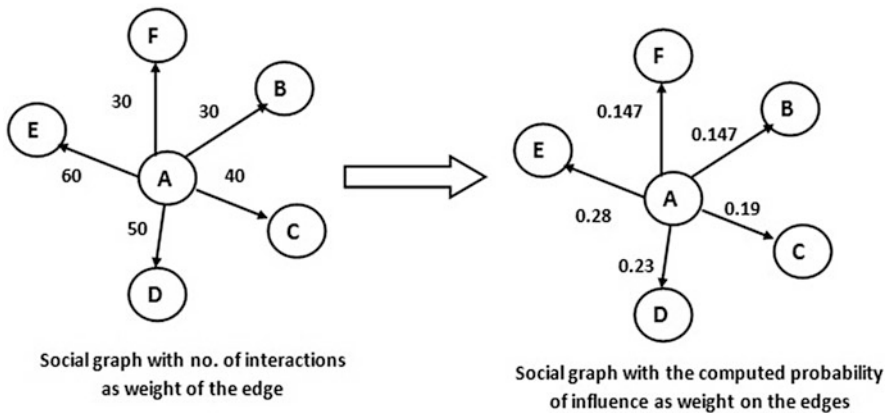


Fig. 11 Social network with influence probability on the edges

5.3 Stage 3: Obtaining the Seed Set

We are now in the position to obtain a good set of initial users, hoping that they will spread the information to a vast population and result in effective sales. In this section, prominent works towards influence maximization are discussed, followed by the proposed approach of ranking users.

5.3.1 Hardness of Influence Maximization

Influence maximization comes under the HP-hard category of problem complexity. Finding a solution in real time, when the input grows exponentially, is impossible for NP-hard problems. Kempe et al. proved that influence maximization under popular diffusion models is NP-hard [24].

Theorem 1 *The influence maximization problem is NP-hard for the independent cascade model and the linear threshold model[24].*

5.3.2 Existing Works

The study of information diffusion in social networks is first proposed by Domingos et al. [9] by identifying the key users. Kempe et al. [24, 25] classified it as NP-hard and proposed a greedy heuristic, which is 66% optimal. However, the running time for the worst case of this algorithm is $O(n^2(m + n))$, making its usage impractical for large scale networks. To reduce the run time, Cost effective lazy forwarding (CELF) [30] and CELF++ [17], MixGreedy and NewGreedy[4] are proposed. In spite of these attempts, the runtime of Greedy could not be reduced; therefore, heuristics were designed that reduced runtime, but did not provide an approximation guarantee. Initial efforts in this direction are the degree discount heuristic [4], coverage under maximum influence paths [5] and a directed acyclic graph [6]. In the direction of diffusion models, works such as the susceptible-infected-recovered (SIR) model [21, 26], the susceptible-infected-susceptible (SIS) model [27], and the continuous time SIS model [46] are seen.

Recent approaches are the incorporation of negative influence [22], the belief propagation model on a directed acyclic graph [38], the time constraint influence spreading paths[32], the combinatorial model of influence spread under time window constraint [15], influence maximization in dynamic networks [61], InFlowMine [49], and the three-step cascade model[41]. Recently, He and Kempe [18] and Chen

et al. [7] came up with a new variant for influence maximization, known as robust influence maximization. He and Kempe found the top influential users in the setting in which multiple influence functions are used for the same model [18]. On the other hand, Chen et al. discussed the solution to influence maximization, given an uncertainty in the parameter input. They propose the LUGreedy algorithm to improve the existing greedy algorithm.

5.3.3 User Ranking on Interaction Rate: Outdegree Rank Algorithm

The formalization of influence maximization and the runtime issue of the greedy approach paved the way for various other alternatives. However, several improvements to the original greedy approach are still not efficient. Therefore, heuristics are designed to solve runtime issues, but compromise optimality. Of these, the degree centrality heuristic has proved to be efficient and close to an optimal solution [5]. The degree concept obtains users with the highest degree (a.k.a. number of contacts), with the belief that such a user will trigger a vast outbreak of information, leading to adoption. However, in the real world, a user interacts with only a small percentage of his/her contacts, raising suspicion regarding the viability of the degree heuristic. Also, other variations of the degree heuristic arouse similar concerns. To make the degree concept viable in the real world, the outdegree rank heuristic for influence maximization is proposed in this work. Unlike existing works, the outdegree rank heuristic considers the user attribute, i.e., interaction count, to obtain the most influential users.

Before getting into the details of the outdegree rank heuristic, three terms: contact degree, interaction graph and interaction degree are introduced to understand the concept.

Definition 1 For a node v , its contact degree is referred to in the social network graph as the number of edges incident on it and is denoted as $Cd(v)$.

Definition 2 The interaction graph is a multi-edge social graph of the contributor graph, where edges represent the interaction between each pair of nodes.

Definition 3 For a node v , its interaction degree is referred to in the interaction graph as the number of edges incident on it and is denoted as $Id(v)$.

User activity types are unique to every social network, for example, wall posts, comments, likes, following, status updates, etc. However, not all users may be actively involved in these activities. Most of the content usually comes from a very small percentage of network users. In a scenario where a user v maintains interactions among few friends, $Id(v) \ll Cd(v)$. Therefore, in this case, the degree centrality approach, which uses $Cd(v)$ as a metric to rank users, may not be optimal.

In this view, degree centrality needs further investigation and the solution is explored in the interaction graph.

It is true that there does not exist any proven correlation between the degree of the node, $Cd(v)$, and interaction degree, $Id(v)$, but an attempt to formalize a relation is made here based on the following reason. When a user's interaction degree is very large compared with his degree, it can be concluded that he/she is interacting actively with at least a few of his/her friends. Also, if $Id(v)$ is very much less than $Cd(v)$, it can safely be concluded that the user is not interacting with all his/her contacts. With $Id(v)$ broken into outdegree and indegree, the ranking of users can be further improved.

The $Id(v)$ can be further specified in terms of *indegree*, defined as the number of edges leading to that node and *outdegree*, defined as the number of edges leading away from that node. In an interaction graph, the *Indegree*(v) represents the popularity index and *Outdegree*(v) represents the participation index. The concept of indegree is already explored in the PageRank heuristic for finding popular web pages [39] and identifying key users in social networks [19]. To solve influence maximization, the final stage is aimed to ranking users on their outdegree.

5.3.4 Outdegree Rank Heuristic for Influence Maximization

For a node, if $outdegree(v) \gg Cd(v)$, it shows that node v is highly interactive. Such a node has greater potential to spread information in the network than other nodes. Based on this surmise, the $outdegree(v)$ is used to rank the users in the social network. This approach is referred to as the outdegree rank heuristic. Unlike the degree heuristic, in which the $Cd(v)$ is almost static, the $outdegree(v)$ is frequently varying according to the changes in the interaction rate of the user. In this way, the outdegree rank reflects the dynamic changes occurring in the social network. Thus, the outdegree rank is a viable solution in the real world.

5.4 Algorithm

Algorithm 2 is a consolidated algorithm for the three steps discussed previously. The input to the algorithm is the social network $G(V, E)$, activity log A , activity record AR , which maintains the number of activities of each user, and k , the number of seeds.

Input: $G(V,E)$, Activity log $A(u, v)$, $AR(u, count), k$
Output: Seed set S of k nodes
Initialize $u.count = 0$, $G_c(E_c, V_c) = \emptyset$, $Total_{u,s} = 0$, $S = \emptyset$;
for Each u in activity log $A(u, v)$ **do**
 if $u.isadded = FALSE$ **then**
 | $u.isadded = TRUE$;
 end
 $u.count++$;
end
Update $AR(u, count)$
Compute the threshold $alpha$ using Eq. (2)
for Each edge $e(u, v) \in E$ **do**
 if $u.count \geq alpha \ \&\& \ v.count \geq alpha$ **then**
 | $E_c = E_c \cup e$
 end
end
for each $e(u, v) \in E_c$ **do**
 $S = \text{neighbourset}(u)$
 for each $s \in S$ **do**
 | $Total_{u,s} += y_{u,s}$
 end
 $p(u, v) = \frac{y_{u,v}}{Total_{u,s}}$
end
for each vertex $v \in V_c$ **do**
 | compute its outdegree out_v
end
for $i=1$ to k **do**
 | $u = \text{argmax}_v \{out_v | v \in V_c - S\}$
 | $S = S \cup u$
end
Output S

Algorithm 2: Influence maximization algorithm

6 Results

6.1 Structure Pruning to Find Contributors

The original social networks those described in Table 1, are pruned to fetch the contributor graph whose description is shown in Table 2. The results show that the pruned graph is now scalable to analyze the effectiveness of influence maximization algorithms. Further, the pruned network should have information propagation properties. To analyze these properties, the pruned network is evaluated on small world metrics, which includes, a higher clustering coefficient, a smaller diameter, smaller average path lengths, lower modularity and fewer components. Tables 3 and 4 summarize the results. The pruned network is compared with the original social network and with the pruned graph obtained by removing those edges that do not constitute the shortest paths. The pruned graph obtained by the proposed method exhibits small world properties significantly better. Thus, the pruned graph

Table 2 Value of α , V_c and E_c of $G_c(V_c, E_c)$

| Name | $G(V)$ | $G(E)$ | α | $G_c(V_c)$ | $G_c(E_c)$ |
|----------|--------|---------|----------|------------|------------|
| HEP | 15,233 | 58,891 | 403 | 205 | 405 |
| PHY | 37,154 | 231,584 | 637 | 310 | 2934 |
| YouTube | 15,088 | 76,765 | 1544 | 545 | 4846 |
| Wikivote | 8275 | 103,689 | 660 | 410 | 11,726 |

Table 3 Structural properties of the social graph and contributor graph of HEP and PHY

| Attribute | HEP | | | PHY | | |
|------------|----------------|----------------------|-------------------|----------------|----------------------|-------------------|
| | Original graph | Shortest path method | Proposed approach | Original graph | Shortest path method | Proposed approach |
| Nodes | 15,233 | 7159 | 205 | 37,154 | 37,149 | 310 |
| Edges | 58,891 | 25,914 | 405 | 23,1584 | 174,161 | 2934 |
| ACC | 0.261 | 0.253 | 0.189 | 0.403 | 0.7 | 0.482 |
| Diameter | 31 | 19 | 7 | 23 | 19 | 11 |
| APL | 10.137 | 5.424 | 2.33 | 7.63 | 6.259 | 3.736 |
| Components | 1779 | 4662 | 9 | 3878 | 3456 | 14 |
| Modularity | 0.852 | 0.712 | 0.722 | 0.927 | 0.923 | 0.7 |

Table 4 Structural properties of the social graph and contributor graph of YouTube and Wikivote

| Attribute | YouTube | | | Wikivote | | |
|------------|----------------|----------------------|-------------------|----------------|----------------------|-------------------|
| | Original graph | Shortest path method | Proposed approach | Original graph | Shortest path method | Proposed approach |
| Nodes | 15,233 | 7159 | 205 | 37,154 | 37,149 | 310 |
| Edges | 58,891 | 25,914 | 405 | 23,1584 | 174,161 | 2934 |
| ACC | 0.261 | 0.253 | 0.189 | 0.403 | 0.7 | 0.482 |
| Diameter | 31 | 19 | 7 | 23 | 19 | 11 |
| APL | 10.137 | 5.424 | 2.33 | 7.63 | 6.259 | 3.736 |
| Components | 1779 | 4662 | 9 | 3878 | 3456 | 14 |
| Modularity | 0.852 | 0.712 | 0.722 | 0.927 | 0.923 | 0.7 |

is an ideal substitute for the original social graph with regard to applications that are expensive in terms of resource usage. The graph size in terms of edges and nodes is drastically reduced, yet the properties the of pruned graph are well suited to information propagation compared with those of the original social network and can thus can be used in place of the original network in applications such as viral marketing.

6.2 *Estimating User Influence*

In this section, the impact of using an estimated influence probability on the spread is discussed. Influence probabilities developed from various other approaches are used to predict spread in various setups. The results are then compared to demonstrate the effectiveness of the proposed approach. The following four cases are used to predict information spread.

1. Trivalency model (TVM) where $p = 0.01$.
2. Random numbers with uniform distribution probability(RNUDp): The influence probabilities are generated from a uniform distribution.
3. Random numbers with normal distribution probability (RNNDp): The influence probabilities are generated from a normal distribution.
4. Influx.

The diffusion process is observed under an independent cascade model run for 1000 simulations for accuracy. The model uses the probability of influence obtained from the above approaches and the seed set size of 50 is considered, which are obtained from the standard algorithms for the seed selection process for influence maximization found in the literature such as highest degree [9], distance [9], single discount [4], and degree discount [4], and are used to compare the outcomes of the proposed approach. The spread estimate is measured as the count of the number of nodes that are activated at the end of the process. The results are shown in Figs. 12, 13, 14, and 15 respectively.

For the HEP dataset, the results, when compared with the approaches that use RNNDp and RNUDp, show an increase of 20% and 18% for the highest degree, 15% and 14% for the distance heuristic, 27.4% and 24.75% for the single discount, and 30% and 26.8% for the degree discount. Moreover, when the value of influence is fixed at 0.01, the influx approach shows an increase of 20%, 15.6%, 28.5%, and 30% for the highest degree, distance, single discount, and degree discount approaches respectively.

For the PHY dataset, too, there is an increase of 8.9%, 14%, 18.43% and 23.17% compared with RNNDp for the highest degree, distance, single discount, and degree discount approaches respectively, and 4.8%, 8.9%, 9.8%, and 12.6% compared with RNUDp for each of the highest degree, distance, single discount, and degree discount approaches respectively. In addition, there is an increase of 8.6%, 14%, 20%, and 25.53% compared with the TVM approach for each of the highest degree, distance, single discount, and degree discount approaches respectively.

For the YouTube dataset, there is also an increase of 9.6%, 15.7%, 10.5% and 10.6% compared with RNNDp for the highest degree, distance, single discount, and degree discount approaches respectively, and 3.15%, 6.3%, 6.3%, and 6.1% compared with RNUDp for the highest degree, distance, single discount, and degree discount approaches respectively. In addition, there is an increase of 19%, 24.21%, 22.1%, and 19.6% compared with the TVM approach for each of these approaches respectively.

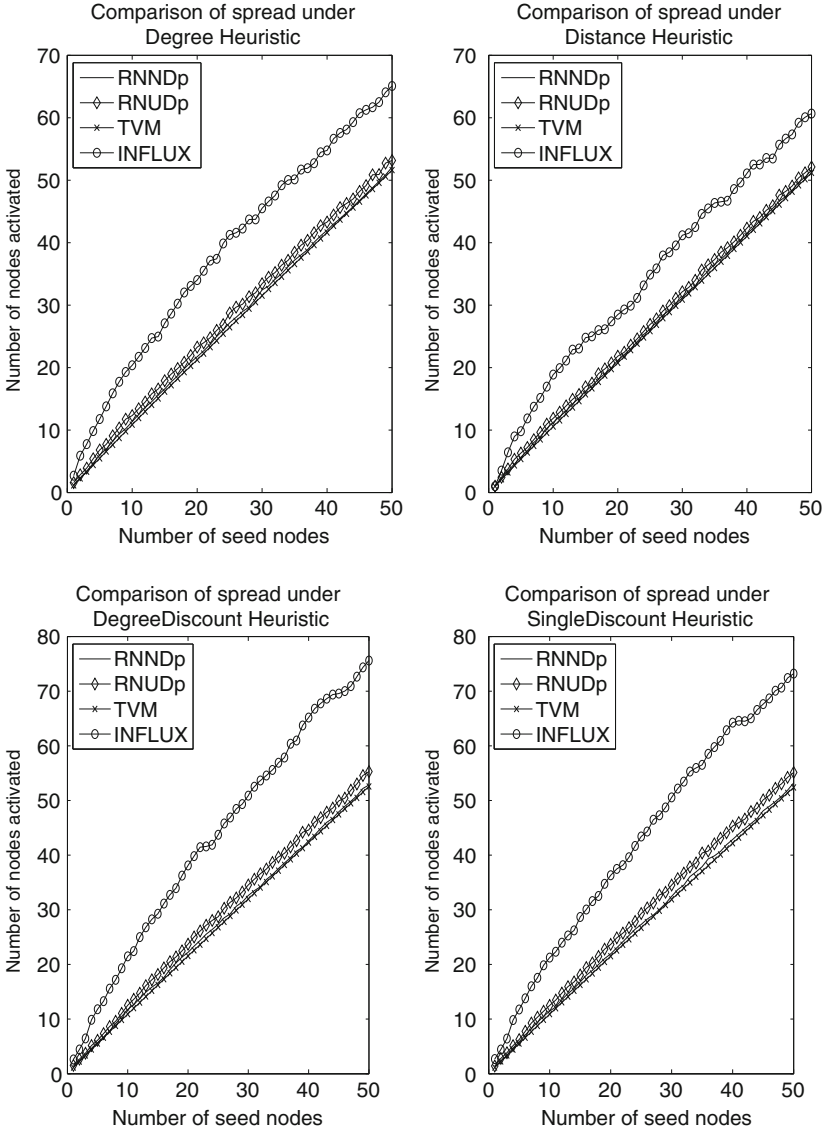


Fig. 12 Comparison of influx with other approaches in HEP

Finally, for the Wikivote dataset, there is an increase of 15%, 13.15%, 13.2% and 15.11% compared with RNNDp for the highest degree, distance, single discount, and degree discount approaches respectively and 40.4%, 35.52%, 39.7%, and 40.6% compared with RNUDp for the highest degree, distance, single discount, and degree discount approaches respectively. In addition, there is an increase of 34.47%, 28%,

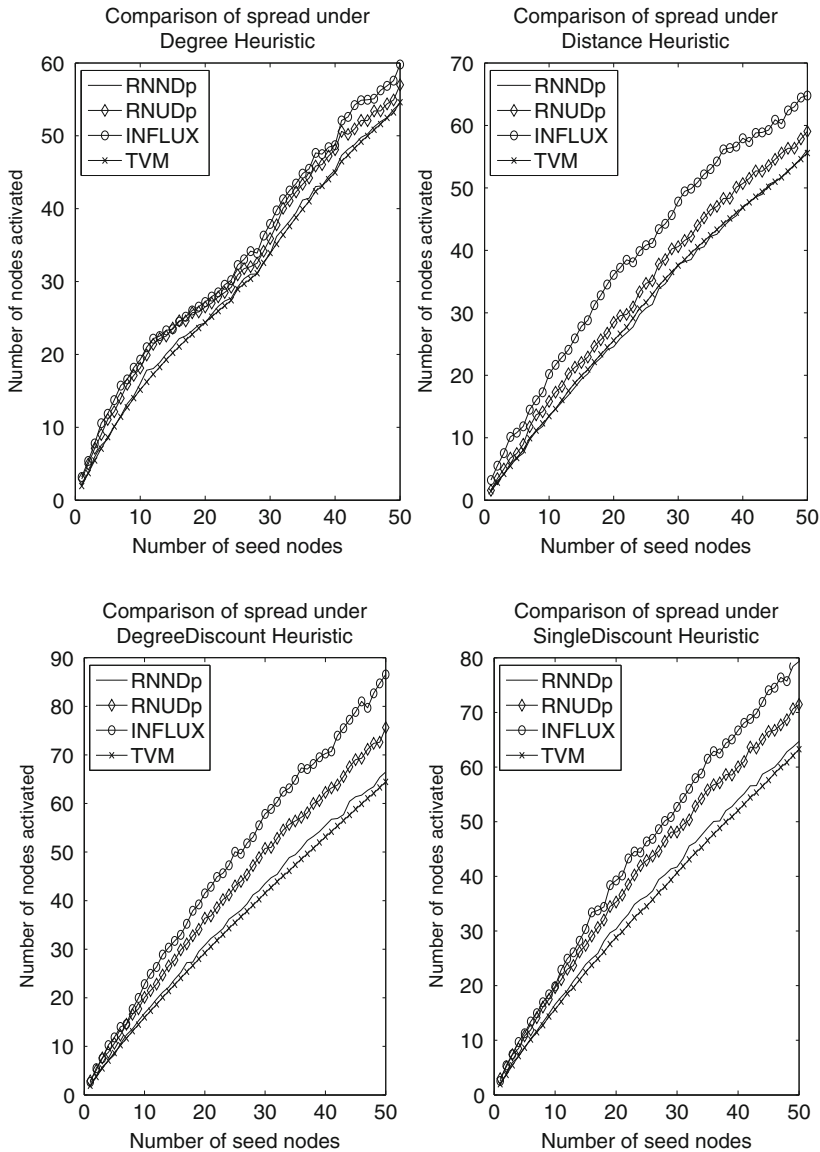


Fig. 13 Comparison of influx with other approaches in PHY

33.26%, and 35.8% compared with the TVM approach for each of these approaches respectively.

Overall, the information spread obtained by using the influx approach is 5–30% higher than RNNDp, RNUDP, and pre-assumed $p=0.01$ for each of the standard algorithms such as highest degree, distance, single discount, and degree discount. It

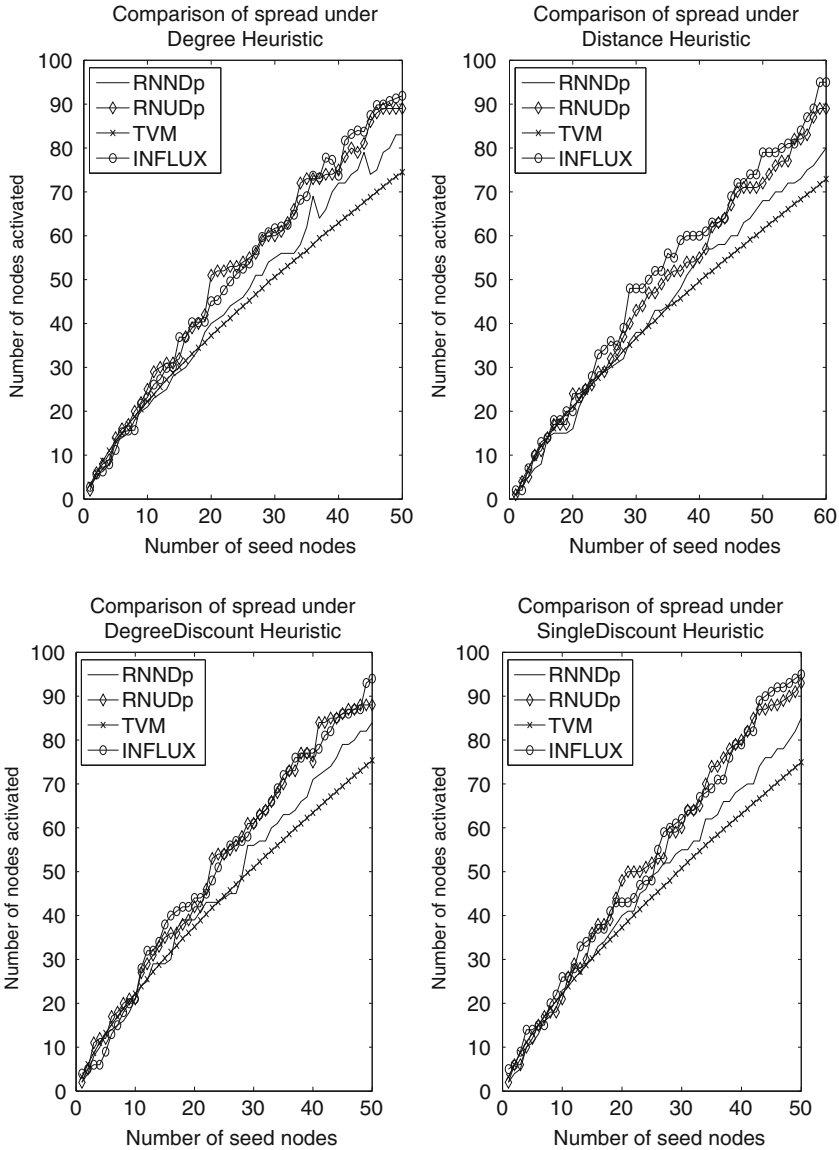


Fig. 14 Comparison of influx with other approaches in YouTube

is clear that the influx approach yields a better outcome as it reflects the interactive nature of the users. whereas the existing approach of using the TVM does not account for user inclination when predicting the spread.

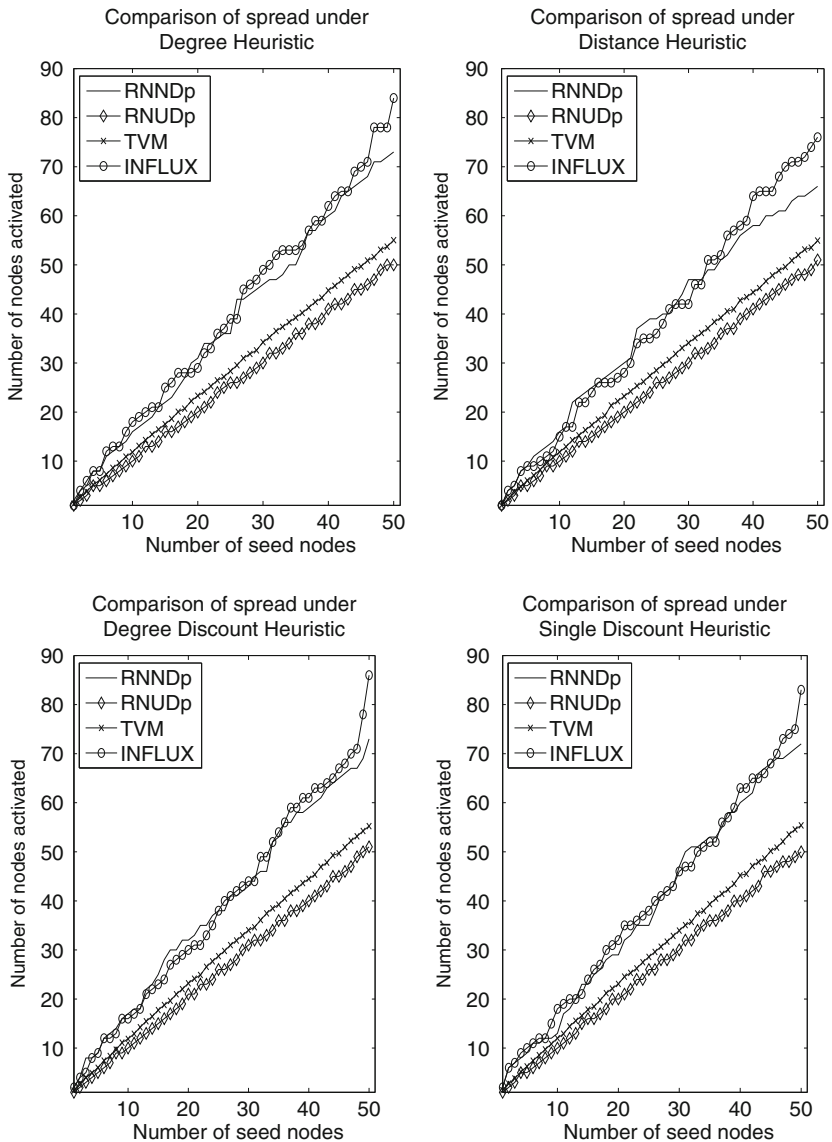


Fig. 15 Comparison of influx with other approaches in Wikivote

6.3 Outdegree Rank with Estimated Influence

Finally, in this section, the performance gain obtained using the outdegree rank with influence estimate (ORIE) is highlighted. The diffusion process is observed under an independent cascade model run for 1000 simulations for accuracy. The model uses

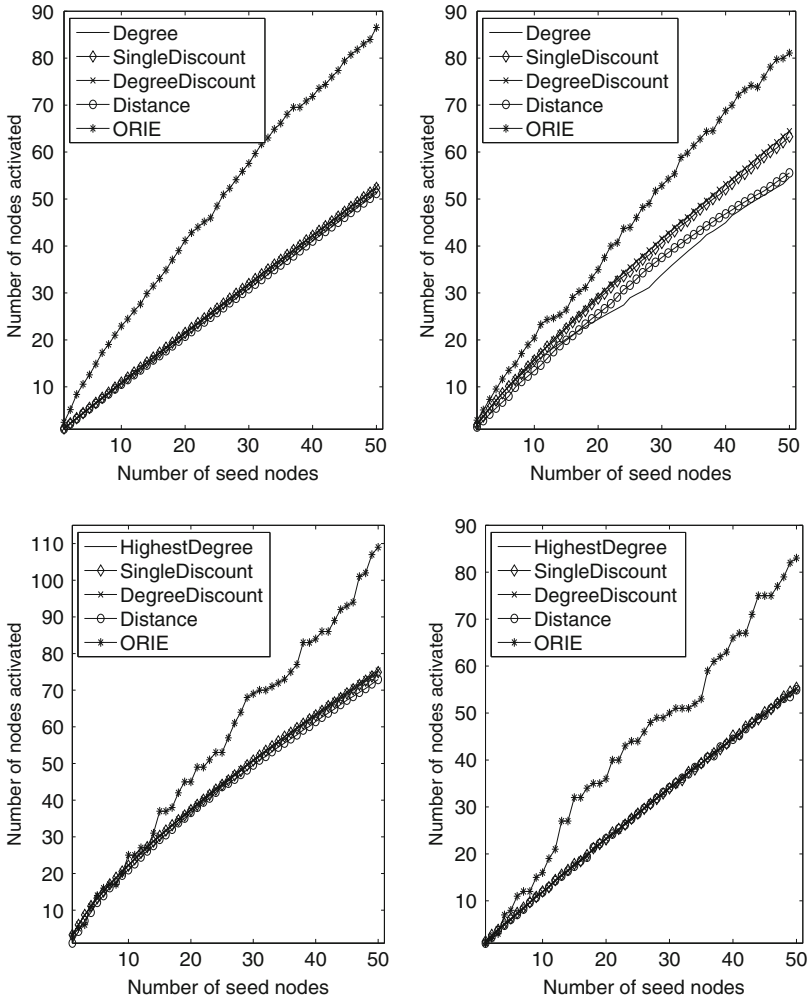


Fig. 16 Performance of outdegree rank with influence estimate on the HEP, PHY, YouTube, and Wikivote datasets

the probability of the influence-obtained influx approach and the seed set size of 50 is considered, which is obtained from standard algorithms used in Sect. 5.2. The spread estimate is measured as the count of the number of nodes that are activated at the end of the process. The performance gain of ORIE compared with other state-of-the-art (SOA) approaches is shown in Fig. 16.

In the HEP dataset, there is an increase of 40.5% compared with the degree, 39.47% compared with the degree single discount, 39.2% compared with the degree discount, and 41% compared with the distance heuristics. Similarly, in the PHY dataset, there is an increase of 32.6% compared with the degree, 21.9% compared

with the degree single discount, 20.5% compared with the degree discount, and 31.5% compared with the distance heuristics. In the YouTube dataset, there is an increase of 31.7% compared with the degree, 31.2% compared with the degree single discount, 30.8% compared with the degree discount, and 33.1% compared with the distance heuristics. Finally, in the Wikivote dataset, there is an increase of 33.6% compared with the degree, 33.2% compared with the degree single discount, 33.4% compared with the degree discount, and 33.8% compared with the distance heuristics.

With the new ranking strategy, there is a gain of up to 41% in information spread compared with the SOA approaches. Conclusively, the outdegree rank strategy outperforms standard approaches that are based on degree centrality.

7 Conclusion

The aim of the chapter is to introduce a new perspective with regard to solving influence maximization. The influence maximization and information diffusion processes are like two faces of the same coin, and should be studied in each other's context. In this chapter, a holistic, an effective, and a scalable solution to influence maximization is proposed. The influence maximization solution is achieved under a combination of aspects of structure, heuristic, and user influence. The work presents three novel approaches to these three aspects, namely structure pruning, influx to estimate influence, and outdegree rank. Finally, the amalgamation of these aspects is a contribution to influence maximization. The experiments on various cases support the claim that user attributes determine the diffusion process; thus, the approach contributes to a new direction of influence maximization solutions. As a future work, the three aspects can be dealt with in detail, adding more features. When the constraints on the privacy of user data for applications is flexible, more user features can be used in each of the phases to provide a better working model. Also, another direction is verifying the validity of the proposed approaches on other diffusion models. Although this work is based on the surmise that all users are equally popular, it can however be extended to include the celebrity aspect.

References

1. Andrews, J.D., Beeson, S.: Birnbaum's measure of component importance for noncoherent systems. *IEEE Trans. Reliab.* **52**(2), 213–219 (2003)
2. Arenas, A., Duch, J., Fernandez, A., Gomez, S.: Size reduction of complex networks preserving modularity. *CoRR* (2007). [abs/physics/0702015](https://arxiv.org/abs/physics/0702015)
3. Borgatti, S.P., Carley, K.M., Krackhardt, D.: On the robustness of centrality measures under conditions of imperfect data. *Soc. Netw.* **28**(2), 124–136 (2006)
4. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: *Proceedings of the 15th ACM SIGKDD, KDD '09*, pp. 199–208. ACM, New York, NY (2009). doi:10.1145/1557019.1557047

5. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD, KDD '10, pp. 1029–1038. ACM, New York, NY (2010) doi:10.1145/1835804.1835934
6. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10, pp. 88–97 (2010)
7. Chen, W., Lin, T., Tan, Z., Zhao, M., Zhou, X.: Robust influence maximization. CoRR (2016). abs/1601.06551. <http://arxiv.org/abs/1601.06551>
8. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. *SIAM Rev.* **51**(4), 661–703 (2009). doi:10.1137/070710111. <http://dx.doi.org/10.1137/070710111>
9. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD, KDD '01, pp. 57–66. ACM, New York, NY (2001). doi:10.1145/502512.502525
10. Fang, X., Hu, P.J.H., Li, Z., Tsai, W.: Predicting adoption probabilities in social networks. *Inf. Syst. Res.* **24**(1), 128–145 (2013)
11. Foti, N.J., Hughes, J.M., Rockmore, D.N.: Nonparametric sparsification of complex multiscale networks. *PLoS One* **6**(2), 16431 (2011). doi:10.1371/journal.pone.0016431
12. Freeman, L.C.: Centrality in social networks conceptual clarification. *Soc. Netw.* **1**(3), 215–239 (1978)
13. Fung, W.S., Hariharan, R., Harvey, N.J.A., Panigrahi, D.: A general framework for graph sparsification. In: Fortnow, L., Vadhan, S.P. (eds.) *STOC*. pp. 71–80. ACM, New York, NY (2011)
14. Ganesan, K.: Case study on ripple effects of ice bucket challenge on social media channels (2016). <http://www.digitalvidya.com/blog/>
15. Gargano, L., Hell, P., Peters, J., Vaccaro, U.: Influence diffusion in social networks under time window constraints. In: *Structural Information and Communication Complexity: 20th International Colloquium, SIROCCO 2013, Ischia, July 1–3, 2013. Revised Selected Papers*
16. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Learning influence probabilities in social networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10, pp. 241–250. ACM, New York, NY (2010)
17. Goyal, A., Lu, W., Lakshmanan, L.V.: 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, pp. 47–48. ACM, New York, NY (2011). doi:10.1145/1963192.1963217
18. He, X., Kempe, D.: Robust influence maximization. CoRR (2016). abs/1602.05240 <http://arxiv.org/abs/1602.05240>
19. Heidemann, J., Klier, M., Probst, F.: Identifying key users in online social networks: a pagerank based approach. In: Sabherwal, R., Sumner, M. (eds.) *ICIS*, p. 79. Association for Information Systems (2010)
20. Jiang, J., Wilson, C., Wang, X., Sha, W., Huang, P., Dai, Y., Zhao, B.Y.: Understanding latent interactions in online social networks. *ACM Trans. Web* **7**(4), 18 (2013)
21. Johnson, T.: Mathematical modeling of diseases: susceptible-infected-recovered (sir) model (2009). <http://op12no2.me/stuff/tjsir.pdf>
22. Jung, K., Heo, W., Chen, W.: IRIE: a scalable influence maximization algorithm for independent cascade model and its extensions. CoRR (2011). abs/1111.4795
23. Kasthurirathna, D., Harre, M., Piraveenan, M.: Influence modelling using bounded rationality in social networks. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pp. 33–40. ACM, New York, NY (2015)
24. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD, KDD '03, pp. 137–146. ACM, New York, NY (2003). doi:10.1145/956750.956769
25. Kempe, D., Kleinberg, J., Tardos, E.: Influential nodes in a diffusion model for social networks. In: Proceedings of the 32Nd International Conference on Automata, Languages and Programming, ICALP '05, pp. 1127–1138. Springer, Berlin, Heidelberg (2005)

26. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics. ii. The problem of endemicity. *Proc. R. Soc. Lond.* **138**(834), (1932). doi:10.1098/rspa.1932.0171
27. Kimura, M., Saito, K., Motoda, H.: Efficient estimation of influence functions for sis model on social networks. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI'09*, pp. 2046–2051. Morgan Kaufmann, San Francisco, CA (2009)
28. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *J. ACM* **46**(5), 604–632 (1999)
29. Kutzkov, K., Bifet, A., Bonchi, F., Gionis, A.: Strip: stream learning of influence probabilities. In: *Proceedings of the 19th ACM SIGKDD, KDD '13*, pp. 275–283. ACM, New York, NY (2013)
30. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: *Proceedings of the 13th ACM SIGKDD, KDD '07*, pp. 420–429. ACM, New York, NY (2007)
31. Lisa, R.: Social influence. In: *The Blackwell Encyclopedia of Sociology*, pp. 4426–4429. Oxford Blackwell, Malden, MA (2008)
32. Liu, B., Cong, G., Xu, D., Zeng, Y.: Time constrained influence maximization in social networks. In: *2012 IEEE 12th International Conference on Data Mining (ICDM), IEEE*, pp. 439–448 (2012)
33. Mathioudakis, M., Bonchi, F., Castillo, C., Gionis, A., Ukkonen, A.: Sparsification of influence networks. In: *Proceedings of the 17th ACM SIGKDD, KDD '11*, pp. 529–537. ACM, New York, NY (2011)
34. McCracken, G.: How ford got social marketing right (2010). <https://hbr.org/2010/01/ford-recently-wrapped-the-firs/>
35. Misiulek, E., Chen, D.Z.: Two flow network simplification algorithms. *Inf. Process. Lett.* **97**(5), 197–202 (2006)
36. Mullaney, T.: Social media is reinventing how business is done (2012). <http://www.usatoday.com/money/economy/story/2012-05-14/social-media-economy-companies/55029088/1/>
37. Myerson, R.: Graphs and cooperation in games. In: Dutta, B., Jackson, M. (eds.) *Networks and Groups, Studies in Economic Design*, pp. 17–22. Springer, Berlin, Heidelberg (2003). doi:10.1007/978-3-540-24790-6_2
38. Nguyen, H., Zheng, R.: Influence spread in large-scale social networks—a belief propagation approach. In: *Machine Learning and Knowledge Discovery in Databases*, pp. 515–530. Springer, Berlin (2012)
39. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. *Stanford Infolab* (1999)
40. Pham-Gia, T., Hung, T.: The mean and median absolute deviations. *Math. Comput. Model.* **34**(7–8), 921–936 (2001)
41. Qin, Y., Ma, J., Gao, S.: Efficient influence maximization under TSCM: a suitable diffusion model in online social networks. *Soft Comput.* 1–12 (2016). doi:10.1007/s00500-016-2068-3
42. Quirin, A., Cordan, O., Santamara, J., Vargas-Quesada, B., Moya-Anegón, F.: A new variant of the pathfinder algorithm to generate large visual science maps in cubic time. *Inf. Process. Manage.* **44**(4), 1611–1623 (2008)
43. I, 215–239 (2008) Robert, H.: Applicability of graph metrics when analyzing online social networks. *Curr. Issues IT-Manage.* **1**, 215–239 (2008)
44. Romero, D.M., Galuba, W., Asur, S., Huberman, B.A.: Influence and passivity in social media. In: *Machine learning and knowledge discovery in databases*, pp. 18–33. Springer, Berlin (2011)
45. Saito, K., Nakano, R., Kimura, M.: Prediction of information diffusion probabilities for independent cascade model. In: Lovrek, I., Howlett, R., Jain, L. (eds.) *Knowledge-Based Intelligent Information and Engineering Systems. Lecture Notes in Computer Science*, vol. 5179, pp. 67–75. Springer, Berlin, Heidelberg (2008). doi:10.1007/978-3-540-85567-5_9
46. Saito, K., Kimura, M., Ohara, K., Motoda, H.: Efficient estimation of cumulative influence for multiple activation information diffusion model with continuous time delay. In: *PRICAI 2010: Trends in Artificial Intelligence, Daegu*, pp. 244–255 (2010)

47. Serrano, M.A., Bog, M., Vespignani, A.: Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.* **106**(16), 6483–6488 (2009)
48. Smith, C.: How many people use the top social media, apps & services (2014). <http://expandedramblings.com>
49. Subbian, K., Aggarwal, C., Srivastava, J.: Mining influencers using information flows in social streams. *ACM Trans. Knowl. Discov. Data* **10**(3), 26:1–26:28 (2016). doi:10.1145/2815625
50. Sumith, N., Annappa, B., Bhattacharya, S.: Social network pruning for building optimal social network: a user perspective. *Knowl. Based Syst.* **117**, 101–110 (2017)
51. Teng, Y.W., Tai, C.H., Yu, P.S., Chen, M.S.: Modeling and utilizing dynamic influence strength for personalized promotion. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 57–64. ACM, New York, NY (2015)
52. Treagus, P.: The dark knight: a case study of viral marketing (2014). <http://philtreagus.com/the-dark-knight-a-case-study-of-viral-marketing/>
53. Wang, Z., Qian, Z., Lu, S.: A probability based algorithm for influence maximization in social networks. In: *Proceedings of the 5th Asia-Pacific Symposium on Internetware, Internetware '13*, pp. 12:1–12:7. ACM, New York, NY (2013)
54. Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P., Zhao, B.Y.: User interactions in social networks and their implications. In: *Proceedings of the 4th ACM European Conference on Computer systems*, pp. 205–218. ACM, New York, NY (2009)
55. Xiang, R., Neville, J., Rogati, M.: Modeling relationship strength in online social networks. In: *Proceedings of the 19th International Conference on World Wide Web*, pp. 981–990. ACM, New York, NY (2010)
56. Yang, J., Leskovec, J.: Modeling information diffusion in implicit networks. In: *2010 IEEE 10th International Conference on Data Mining (ICDM)*, pp. 599–608. IEEE, New York, NY (2010)
57. Zhang, H., Mishra, S., Thai, M.T., Wu, J., Wang, Y.: Recent advances in information diffusion and influence maximization in complex social networks. *Oppor. Mobile Soc. Netw.* **37** (1.1) (2014)
58. Zhou, F., Malher, S., Toivonen, H.: Network simplification with minimal loss of connectivity. In: *2010 IEEE 10th International Conference on Data Mining (ICDM)*, pp. 659–668 (2010). doi:10.1109/ICDM.2010.133
59. Zhou, F., Mahler, S., Toivonen, H.: Review of bisonet abstraction techniques. In: *Bisociative Knowledge Discovery*, pp. 166–178. Springer, Berlin (2012)
60. Zhou, F., Mahler, S., Toivonen, H.: Simplification of networks by edge pruning. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery. Lecture Notes in Computer Science*, vol. 7250, pp. 179–198. Springer, Berlin (2012)
61. Zhuang, H., Sun, Y., Tang, J., Zhang, J., Sun, X.: Influence maximization in dynamic social networks. In: *2013 IEEE 13th International Conference on Data Mining (ICDM)*, pp. 1313–1318. IEEE, New York (2013)