Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2013 Completed Research | ECIS 2013 Proceedings

7-1-2013

On The Role Of Centrality In Information Diffusion In Social Networks

Anastasia Mochalova
Katholische Universität Eichstätt-Ingolstadt, Ingolstasdt, Germany, Anastasia.Mochalova@ku.de

Alexandros Nanopoulos
Katholische Universität Eichstätt-Ingolstadt, Ingolstadt, Germany, alexandros.nanopoulos@ku.de

ON THE ROLE OF CENTRALITY IN INFORMATION DIFFUSION IN SOCIAL NETWORKS

Mochalova, Anastasia, Katholische Universität Eichstätt-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany, Anastasia.Mochalova@ku.de

Nanopoulos, Alexandros, Katholische Universität Eichstätt-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany, Alexandros.Nanopoulos@ku.de

Abstract

Towards understanding how people use social media to interact with each other, it becomes important to investigate the influence mechanism that emerges in online social networks. Due to it, information in social networks often diffuses "virally". In this paper, we recognize the relation between the influence of the seed members that trigger information diffusion and the attitude of the rest members towards the information. To characterize the influence of seed members, we examine their centrality within the network structure, based on the premise that more central members can reach many other members. We examine a comprehensive set of centrality scores for identifying the influence of seed members and conduct a thorough experimental evaluation with data from a real social network. Our experimental results provide insights into the interplay of the various centrality scores with the members' attitude and how this interplay affects the outcome of information diffusion. The first contribution of the proposed approach is, thus, a solution to the influence maximization problem that is based only on information about the structure of a social network and does not require any additional knowledge about the quantification of influence between its members. The second contribution is a thorough investigation of the performance of the various centrality-based seed-selection methods with respect to factors such as the seed size and members' attitude, indicating that different centrality scores are suitable in different cases.

Keywords: Online Social Networks, Influence, Information Diffusion, Centrality.

1 Introduction

Social networks were always part of the human society. Nonetheless, the development of social media automated and accelerated the social signals that pulse through our societyat a global scale and on a daily basis (Aral 2012). With social media, it also became easier to collect and analyse large-scale data from social networks.Interdisciplinary research in social-network analysisbenefits from the existence of such data,in a pursuitof determining the role of social media in our lives.

A fundamental issue in social-network analysis is the investigation of the *influence* mechanism that emerges with increased interconnectedness. This mechanism determines several aspects of our behaviour; e.g., what we consume, which ideas we adopt, etc.Due to this influence mechanism, information that is often triggered by a small seed of members, diffuses "virally" in social networks, because it conveys an implied endorsement from social connections (Jurvetson 2000). The outcome can be a collective action expressing the wide acceptance of the propagated information by a large fraction of the members of the network. This fact has already started to be leveraged for several commercial applications in online social networks, which enable individuals to share (electronic) word-of-mouth and create the potential for exponential growth of spread of information (Bonchi et al. 2011). Therefore, investigating the influence mechanism and how it affects information *diffusion*is an important step towards understanding how people interact with each otherin social media.

Information diffusion in social networks is affected by three factors (Bampo et al. 2008): i) network structure, ii) behavioural characteristics of network members, and iii) characteristics of the information. Especially in online social networks, where information spreads more rapidly than in any other network topology, network structurehas been reported to be very significant (Borgatti et al. 2009). The reason is the virtuous interaction between few members with many connections and the large number of members with few connections (Doer, Fouz & Friedrich 2012). Thus, influential members can be identified according to their connectivity, which determines the impact of such members on the spreadof information through the network. Influential members are exploited in applications such asviral marketing, which detect them based on their *centrality* in the network structure (Newman 2010) and use them as seeds for initiating marketing campaigns (Hinz et al. 2011). Nevertheless, existing approaches (described in Section 2) lack the thorough examination of a wide range of centrality scores for the identification of influential members. More importantly, they do not consider the role of the other two factors, i.e., behavioural characteristics of members and the characteristics of the diffused information.

In this paper, we provide a thorough examination of a comprehensive set of centrality scores used for the purpose of identifying influential members in online social networks. Additionally, we investigate the factor of centrality in relation to the behavioural characteristics of network members, which express the *attitude* of the members towards the diffused information.We conducta thorough experimental evaluation with data from a real social network and examine both the impact of centrality and of attitude. Our results show that the effectiveness of centrality scores in identifying influential membersis related to themembers' attitude. This helps better understanding the properties of centrality when applied in the context of information diffusion in social networks. Therefore, our results can become useful towards:

- Understanding the way that information is diffused in social networks in relation to the influence of the seed members that initiated the diffusion;
- Recognize the role of members' attitude in the coverage that the diffused information is going to attain;
- Identifying when it becomes possible that a given piece of information will be propagated to a large fraction of a social network, although it may be initiated by a very small seed of its members.

The rest of this paper is organized as follows: in Section 2 we present the related work and explain in more detail the motivation of our study. In Section 3 we first formally describe the

examined problem and then we describe the proposed methodology. The organization of our experimental evaluation is presented in Section 4, whereas the experimental results and their discussion are presented in Section 5. Finally, Section 6 concludes the paper.

2 Related Work

Models of Information Diffusion in Social Networks: Users of social media are nowadays rapidly producing and disseminating vast amounts of real-time information. Developing models about the diffusion of information in social networks has been characterized as a challenging task (Leskovec, Adamic & Huberman 2007). The most widely applied models are the Independent Cascade (IC) model (Goldenberg, Libai& Muller 2001) and the Linear Threshold (LT) model (Granovetter 1978). IC considers each interaction between two connected members in a social network as being independent. LT focuses on the threshold behaviour, having information being diffused to a member of a social network when enough of other members connected to it have already adopted the information.

Influence maximization: The task of influence maximization is about selecting the set of members in a social network who will initiate the diffusion of a piece of information and will influence the largest possible number of other members by activating them to adopt this information (Kempe, Kleinberg & Tardos 2003). The selected set of members is called *seeds* and has a predefined size, which represents the cost to initiate the spread of information using the seeds (i.e., the largest the seed set size, the higher the cost). Influence maximization has been shown to be an NP (non-deterministic polynomial time) problem (Kempe, Kleinberg & Tardos 2003). This means that an exact solution has an exponential worst-case time complexity with respect to the number of members, which makes it prohibitive for large social networks. For this reason, Kempe, Kleinberg and Tardos proposed an approximate, greedy hill-climbing algorithm that finds the seed set which activates at least about 63% as much nodes as the optimal seed set. Their approach is based on non-negative, monotone submodular functions (Nemhauser, Wolsey & Fisher 1978), whose main property is that the difference in the value of such a function that a single member makes when added in the seed set, decreases as the size of the seed set increases. This property can be informally described as a kind of diminishing returns, which makes them suitable for an approximate solution to the influence maximization problem. Although the guaranteed performance of the greedy algorithm is advantageous, it requires for each connected pair of members u and v in the network, the knowledge of the strength of their connection. This strength is denoted as *influence factor* and expresses how much u can influence v and vice versa. Apriori knowledge of influence factors is not possible in most real-world cases, thus the direct applicability of the greedy algorithm is limited. To overcome the aforementioned problem, recent research proposed to estimate influence factors based on actions previously performed by members of the network (Goyal, Bonchi & Lakshmanan 2012). However, this approach can be used to select suitable seeds for spreading a new piece of information only in the case that the recorded actions are relevant to this information. Otherwise, the estimated influence factors will not be accurate.

Centrality-based Influence maximization: As mentioned above, acquiring knowledge about influence factors presents several problems. For this reason, recent research has proposed an alternative approach to select the seed members with a central position in the network structure, motivated by the observation that through such members, the diffused information may have better chances to reach a larger part of the network (Hinz et al. 2011). To select members according to how central their position is, research in social-network analysis has proposed a large set of *centrality scores* (Newman 2010). The study of Hinz et al. focuses on two: i) Degree centrality, which identifies members, called hubs, with large number of connections; and ii) Betweenness centrality, which identifies members, called bridges, that are part of a large number of paths connecting the rest of the

members.[1] Although hubs are intuitively good candidates for seed selection, their large number of connections can expose them to information overload, which might render them less likely to propagate information (Porter & Donthu 2008). On the other hand, bridges enable the diffusion of information into parts of a social network that would otherwise be unconnected (Granovetter 1973; Easley & Kleinberg 2010). The experimental study of Hinz et al. compared hubs and bridges based on two applications of information diffusion in two real social networks. In both cases, the two centrality scores presented comparable performance that was better than random seed selection. The advantage of the approach proposed by Hinz et al. is that centrality scores only require knowledge of the network structure, which is not difficult to obtain in online social networks, and do not need the estimation of influence factors.

Motivation for our approach: In this paper, we use models of information diffusion, in particular IC and LT, for implementing diffusion processes over social networks and experimentally studying the performance of various seed-selection methods (see Section 4 for more details). Our focus is, thus, not on developing new models about information diffusion. Instead, our work focuses on the problem of influence maximization. To avoid the difficulties related to acquiring knowledge about influence factors, similar to the work by Hinz et al., our approach is based on centrality-based seed selection. However, Hinz et al. examined only two centrality scores, whereas the literature of social-network analysis provides a much richer set (Newman 2010). Moreover, the experimental results of their study are not indicative about which score should perform better, as they found them to perform equally well. In our work, our motivation is to examine several centrality scores and obtain a more representative comparison. Additionally, in the case of LT model, we consider the impact of members' attitude on the performance of centrality-based seed selection. In summary, our findings help drawing clearer conclusions about the performance of various centrality scores in the context of seed selection problem, which can better guide their practical applications.

3 Centrality Scores for Seed Selection

3.1 Problem Formulation

We consider online social networks that allow their members to add other members of the same network to their list of connections and share with them various types of information, e.g., status updates, tweets, likes, shares, etc. (Kaplan & Haenlein 2010). What distinguishes online social networking platforms (e.g., Facebook, Twitter, LinkedIn) from other, more generic social media for online communities (e.g., Epinions, Qype, Rotten Tomatoes) is that electronic word-of-mouth generally takes place between members who have some kind of personal relationship with one another (Coulter and Roggeveen 2012). Hence, the structure of an online social network is specified by its members and the relationships (linkages) among them (Bampo et al. 2008).

We represent the structure of a social network as a graph $G(V,E)$, where V is the set of nodes (each node corresponds to a member) and E is the set of edges. Due to the symmetric nature of social connections in most social networks, we consider the edges in E to be undirected. Nevertheless, our approach can be extended to consider directed edges (representing asymmetric social connections), by extending the centrality scores of Section 3.2 to their counterparts for directed graphs. An example of a graph G representing the structure of a social network is illustrated in Figure 1a.

For the above explained reasons, the graph G of Figure 1a is undirected. Additionally, it is also unweighted, i.e., its edges have no weights attached to them. The reason is that we focus only on the structure of the network without assuming any knowledge about the strength of its connections.

[1]Hubs and bridges are described in more detail together with the rest examined centrality scores in Section 3.2.

(a) (b)

Figure 1. *(a) Example of an undirected graph representing the structure of a social network. (b) The corresponding directed and weighted latent graph that contains the influence factors (with values between 0 and 1).*

Nevertheless, each edge of *G* represents a connection that has an existing but unknown strengthexpressed through an influence factor (for a description of influence factors see Section 3). Therefore, each such graph *G* has a corresponding graph *G'* whose edges are weighted with influence factors. For the graph *G* of Figure 1a, a corresponding graph *G'* is illustrated in Figure 1b. It has to be noticed that *G'* is both weighted and directed, since the influence factors between two nodes connected with an edge may in general not be equal. In the example of *G'* depicted in Figure 1b, the influence factor of node *A* to node *C* is 0.3, whereas of node *C* to node *A* is 0.1. We have to emphasize, however, that we consider as known only the undirected and unweighted graph *G* that represents the structure of a social network (like the graph of Figure 1a), whereas the directed and weighted graph *G'* with the influence factors (like the graph of Figure 1b) is considered as unknown. Actually, *G'* is latent and realized only during the spread of information through the network. We will utilize this assumption whenwe will describe our experimental framework in Section 6.

Influence maximization problem: The problem of influence maximization seeks for a seed set, *S,* of *k* members of a social network that will be targeted initially and will maximize the expected spread of a given piece of information through the network. We assume that the size of *S* is controlled by the number, *k*, of members in it. As described in Section 2, *k* is predefined and represents the cost to target the members of *S*. The total number of members of the network that will beactivated during the spread initiated by *S*, is denoted as $A_k(S)$. Given the graph $G(V,E)$ that represents the structure of the network, the influence maximization problem can be defined as the problem of finding the seed *S* of size *k* with the maximum possible $A_k(S)$ value; i.e.:

$$S = \underset{U \subseteq V}{\operatorname{argmax}} A_k(U), s.t. |S| = k$$

As described in Section 2, Kempe, Kleinberg and Tardos (2003) showed that this is an NP problem. For this reason, in Section 3.2, we will present a method based on the centrality of members in *G*, in order to obtain an approximation of the optimal solution.

3.2 Seed Selection based on Centrality Scores

The centrality of a vertex in a graph determines the relative importance of the vertex. In the context of a social network, centrality determines how influential a member is within the network. The idea behind using centrality for seed selection is based on the premise that, the more central a member is in the structure of a social network, the higher its influence is going to be on other members, because it is easier for this member to reach all other members. For seed selection based on centrality, we have as input the structure of the social network. Thus, given the undirected and unweighted graph $G(V, E)$ that represents the structure of a social network, we can define for each node $v \in V$, its centrality score $c(v)$. We assume that centrality scores are normalized in the [0, 1] range and that the

higher the score of a node, the more important the node is.[2] Having determined for each node $v \in V$ its centrality score $c(v)$, we then rank the nodes according to decreasing order of their score and select the top-k nodes with the highest scores. Finally, the selected top-$k$nodes comprise the seed.

Different algorithms have been proposed to compute centrality scores (Borgatti & Everett 2006; Newman 2010). In the reminder of this section we will give a brief overview of the main existing algorithms for computing centrality scores:

- *Degree centrality*: computes the number of paths of length one that emanate from a node. The nodes with high degree centrality usually have increased activity and thus more likely to engage in diffusion of information through the network. However, these nodes might be overloaded with information and harder to activate (Hinz et al. 2011).
- *Betweenness centrality*: computes the share of times that one node need another node to reach the third node via the shortest path. Nodes with high betweenness centrality are usually the nodes that connect otherwise unconnected parts of the network. Thus they allow the access and propagation of the idea in several parts of the network at the same time. Disadvantage of this method is that it is harder to compute than, for example, degree centrality.
- *Closeness centrality*: computes the sum of distances from all other nodes where the distance from a node to another is defined as the length of the shortest path from one to the other (since graph G is unweighted, each edge has length equal to 1). For the nodes with high closeness centrality, it is usually easier and faster to reach other nodes. Disadvantage of this method is as with betweenness centrality – difficulty to compute. Another disadvantage is that this score is mostly suitable for connected graphs.
- *Eigenvector centrality*: computes relative scores to all nodes in the network based on the principle that connections to nodes having a high score contribute more to the score of the node in question, i.e., not only the connectedness of the node in questions is analysed but also the connectedness of the nodes it is connected to is taken into account too. PageRank centrality is a variant of Eigenvector centrality that we use in our experimental evaluation.

These four centrality scores are the most widely used (Borgatti & Everett 2006). Existing research (e.g., Hinz et al. 2011; Kempe, Kleinberg & Tardos 2003) focuses mostly on degree and betweenness centrality. In the following we will consider all the scores listed above and compare their performance in different social network settings.

4 Experimental Design

In this section, we describe the design of our experimental study that comparesthe centrality scores presented in Section 3.2.

Data set: We used real data on Facebook users belonging to New Orleans regional network. This data set is publically available at "Online Social Networks Research"[3]and contains 60,290 users connected together by 1,545,686 links in the social network with an average node degree of 25.3. Based on this data set we are able to construct the graph that represents the structure of this network (i.e., an undirected and unweighted graph similar to the example in Figure 1a). According to this graph we can compute the centrality score of each node and select in the seed set those with the top-k scores, based on the method described in Section 3.2. Additionally, in order to be able to implement the task of information diffusion through this network, we need to realize the latent graph with information about the influence factors (i.e., a directed and weighted graph similar to the example of Figure 1b).

[2] There exist some algorithms that compute centrality scores with inverse order, i.e., important nodes have scores closer to 0. However, we can trivially reverse the monotonicity of such scores in order to satisfy our assumption.

[3]http://socialnetworks.mpi-sws.org/

We have to note that this latent graph is not used during the seed selection, but only to implement the information diffusion that will measure the performance of each selected seed (please refer to the discussion in Section 3.1). To obtain the required influence factors in the latent graph, the data set also provides information about the interaction of users in the social network through the so called wall posts, which represent a broadcast-style messaging service between users connected in Facebook. The data set provides 838,092 wall posts, for an average of 13.9 wall posts per user. Following the approach commonly used in related research (Kempe, Kleinberg & Tardos 2003), we consider the number of messages (wall posts) that user u sends to user v as an indicator of the influence that u has on v. Also following existing approaches (Kempe, Kleinberg & Tardos 2003), we normalized the influence factor that u has on v, by dividing the number of messages sent from u to v by the total number of messages sent to v. This way, all influence factors are in the range between 0 and 1, and the total sum of influence factors on each node is equal to 1.

Methodology and Performance Measures: All centrality scores described in Section 3.2 were implemented in Java using the JUNG framework[4]. The performance of each seed S with k members is measured according to $A_k(S)$, i.e., the total number of members of the network that will be activated during the spread initiated by S (see Section 3.1), where higher $A_k(S)$ value denotes better performance. To be able to make this measurement, we implemented diffusion processes over the latent graph of the examined social network using the two widely applied diffusion models:

- *Independent Cascade (IC) model* (Goldenberg, Libai & Muller 2001). IC starts by activating the nodes in the seed set and then the diffusion process unfolds in discrete steps. When a node v becomes active in a step t, it has a single chance to activate each of the nodes w connected to it (w is called the neighbour of v) that are currently inactive. Node v succeeds in activating w with probability $p_{v,w}$ equal to the (normalized) influence factor that v has on w. If v succeeds, then w becomes active in a step t+1 and recursively tries to activate its neighbours. Otherwise, v makes no further attempts to activate$w$. The process runs until no more node activation is possible. Therefore, IC models the individual influence each node has on all its neighbours.
- *Linear Threshold (LT) model* (Granovetter 1978). Each node w is influenced by each of its neighbours v with a weight $b_{v,w}$ that is equal to the (normalized) influence factor that v has on w. Additionally, each node v has a threshold θ_v in the range between 0 and 1, which corresponds to the attitude of v. With the term attitude we refer to the predisposition of a member in a social network to respond positively or negatively towards a diffused piece of information. Therefore, the higher θ_v is, the harder it is to activate v. The diffusion process of LT unfolds in discrete steps. In each step all nodes that were active before, remain active. An inactive node v is activated only if the total weight of its active neighbours is equal or more than its threshold θ_v. The process runs until no more node activation is possible. Therefore, LT models the collective influence that each node receives from all its neighbours.

IC and LT differ with respect to the activation process. According to LT, a member becomes activated when enough members connected to it become active, which represents a kind of "social pressure". IC focuses on the pair-wise relationships between members: the higher the influence of one member on another, the higher the probability of the second member becoming activated. Therefore, as explained above, the attitude of a member v in a social network corresponds to the threshold θ_v that is used in LT and quantifies the "resistance" of v to "social pressure". In contrast, IC model does not consider a parameterisation that corresponds to members' attitude. However, this cannot be consider as a limitation, since IC and LT model information diffusion in social networks from different perspectives, as described above, and are both widely used in related research.

Parameterization: In the case of LT, we consider the threshold θ_v of each node v as a random variable that follows Beta distribution, i.e., $\theta_v \sim Beta(\alpha, \beta)$. The reason for this is that Beta distribution is very

[4]http://jung.sourceforge.net/

flexible and, by tuning its parameters α, β, it can represent social networks with entirely different overall attitudes of their members. Therefore, within the same network, members have various attitudes and the Beta distribution allows us to choose the tendency of the varying attitude scores. In our results we examine the following set of values: i) $\alpha = 2, \beta = 2$ representing a network with most members having neutral attitude, as the mean value is 0.5, and some members having positive and some having negative attitude; ii) $\alpha = 5, \beta = 2$ representing a network with most members having negative attitude, because the mean value is high, thus making them harder to activate; iii) $\alpha = 0.5, \beta = 0.5$ representing a network with members that either have positive or negative attitude and almost no members with neutral attitude.

Since each application of IC and LT involves a probabilistic element (i.e., in each trial IC and LT activate nodes with some probability determined by the influence factors in the way that has been described above), each measurement is repeated 10,000 and we report the averages. This way, we can also compute the standard deviation and statistically check the differences between the examined centrality scores. Regarding the centrality scores, we used the following parameter for the PageRank variant of the Eigenvector centrality: we used uniformed edge weight 0.1. Finally, regarding the seed size, we examine values less than 1% of the total number of network members, because "viral" information diffusion in social networks often starts from a very small seed.

5 Experimental Evaluation

In this section, we first present the results of our experimental evaluation (in Section 5.1), which aims at comparing the various centrality scores. Next, we provide a discussion of the presented results (in Section 5.2).

5.1 Experimental Results

In our first experiment, we focused on the case of information diffusion based on the IC model. We examined all centrality scores listed Section 3.2 as well as the random selection as a baseline. (Please note that, because we focus on the case where influence factors cannot be considered as available, the greedy hill-climbing algorithm (Kempe, Kleinberg & Tardos 2003) is not possible to be used as baseline.) Figure 2 shows the number of activated nodes (in thousands) for varying seed size. All the examined centrality scores clearly surpass random selection. Betweenness centrality presents the best performance, followed by Eigenvector centrality that shows comparable performance. Degree centrality performs favourably only for very small seed sizes. Finally, closeness centrality performs the worst among all other centrality scores. We have to note that the differences in the reported averages (out of 10,000 repetitions) among all examined methods were larger than one standard deviation. Since in all our rest experiments random selection is consistently and by large outperformed, to make our charts more readable, we avoid farther presentation of results for random selection.

Next we move on to examine diffusion based on the LT model. We first focus on the case where members have neutral attitude (i.e., according to the parameterization described in Section 4, this case uses $\alpha = 2, \beta = 2$). The results are shown in Figure 3a. In contrast to the previous measurement, Degree centrality is outperforming all other centrality scores. Betweenness and Eigenvector centrality scores present almost identical performance, whereas Closeness centrality is again outperformed by all others. The differences between the reported averages were larger than one standard deviation except among Betweeness and Eigenvector centrality.

Continuing with the LT model, we next examine the case where members either have positive or negative attitude and there are almost no members with neutral attitude (i.e., according to the parameterization described in Section 4, this case uses $\alpha = 0.5, \beta = 0.5$). The results are shown in Figure 3b. In this case, since there is an adequately large number of members with positive attitude, all

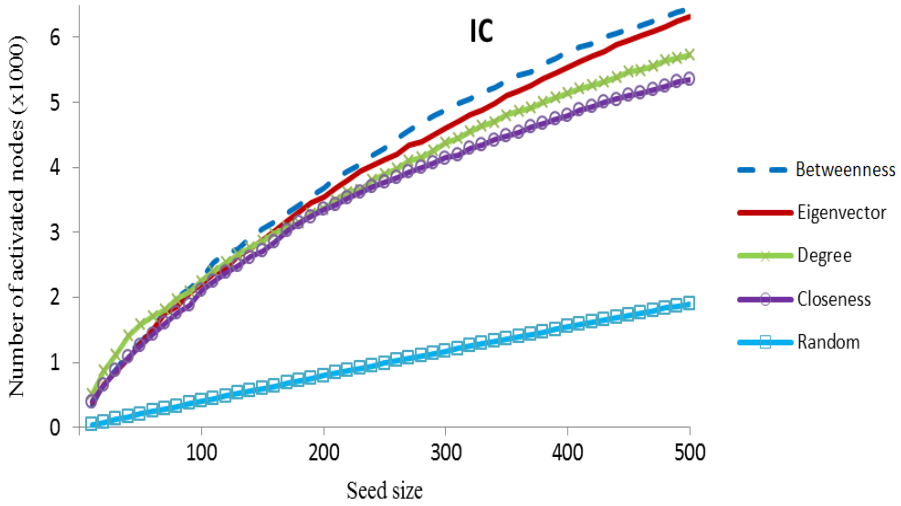

Figure 2. Number of activated nodes (in thousands) vs. seed size for the IC model.

centrality scores attain to activate a much larger number of members compared to the previously examined cases. Focusing on their relative performance, unlike the previous case with normal attitudes, Degree centrality is outperformed by Betweenness and Eigenvector centrality. The differences between the reported averages for Degree centrality and both Betweeness and Eigenvector centrality were larger than one standard deviation.

Finally, we examine the most challenging case where members have a negative attitude (i.e., according to the parameterization described in Section 4, this case uses $\alpha = 5, \beta = 2$). The results are shown in Figure 3c. Interestingly, for this challenging case, it is the Eigenvector centrality that clearly outperforms all other methods, with Betweenness centrality being the second best. The reported averages among all examined methods were larger than one standard deviation.

5.2 Discussion

The main conclusion from the experimental results of Section 5.1 is that the diffusion of information, which is initiated by a small seed set, depends on both the centrality of members in the seed set and on the attitude of all other network members. Although this conclusion seems intuitive, recent research in this area (Hinz et al. 2011) did not thoroughly investigate the interplay between the two aforementioned factors, i.e., centrality and attitude. Our experimental study indicated that the outcome of information diffusion in a social network varies with varying members' attitude and varying centrality of the seed members. In particular, when the members' attitude is mostly negative, Eigenvector centrality, which has not been considered at all by Hinz et al. (2011), is very effective in selecting influential seed members. Regarding Betweenness and Degree centralities, which were considered by Hinz et al. as performing almost equal, our results showed that their performance differs: Degree centrality is more effective when the members attitude is mostly neutral, whereas in all other cases Betweenness centrality outperforms Degree centrality. Another conclusion from our experimental evaluation is that information spreads faster with increasing seed size as well as with increasing number of members with positive attitude (this is evident, for instance, in case of Figure 3b where the network has a large number of members with positive attitude in relation to the other cases in Figure 3).

Therefore, with a more thorough evaluation that takes into account more centrality scores, as well as the attitude of the members, our results offer more insights into the influence mechanism that emerges in social networks. In particular, our findings can help towards the following directions:

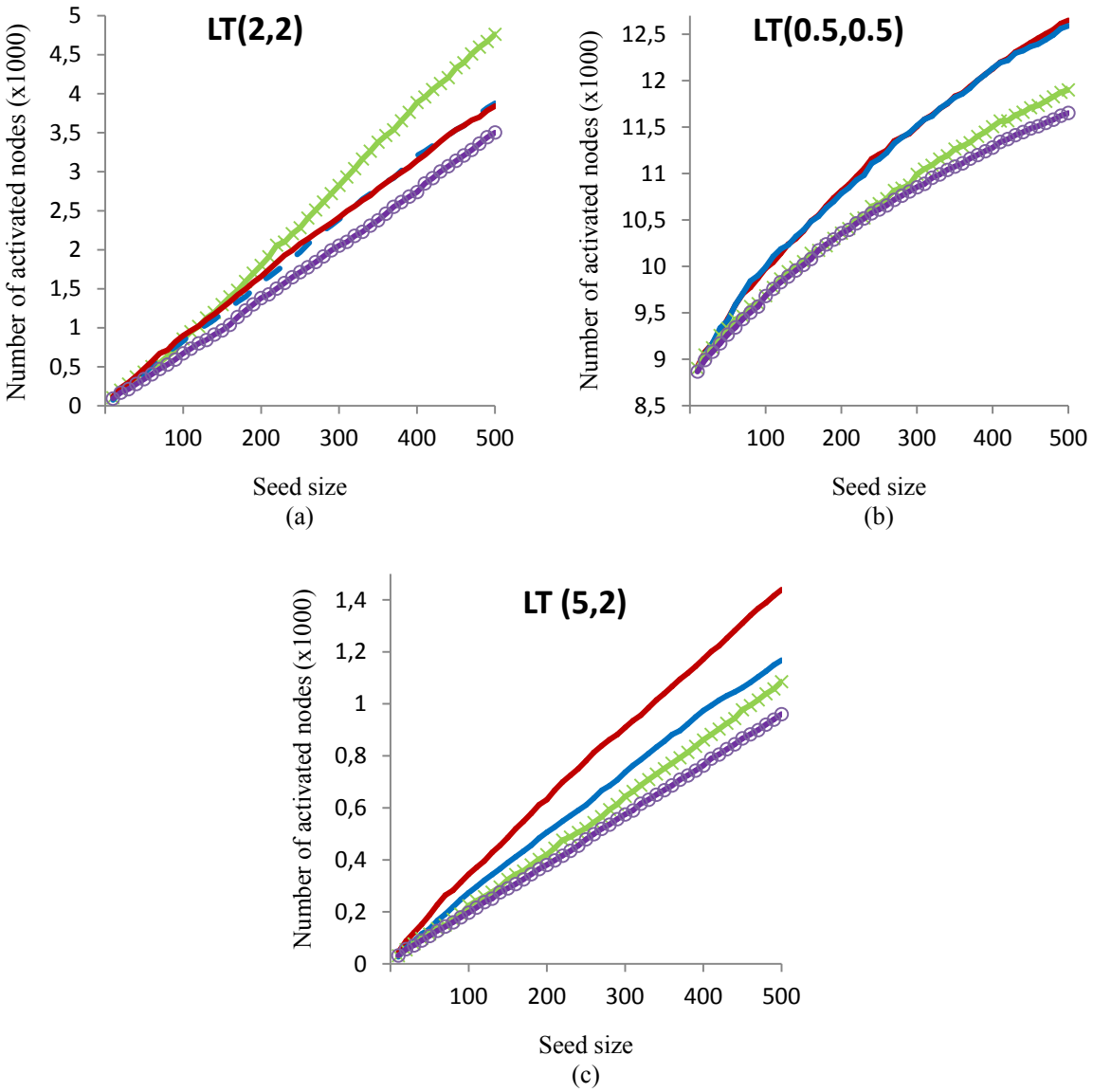

Figure 3. Number of activated nodes (in thousands) vs. seed size for the LT model. (The same legend is used as in Figure 2.)

- The diffusion of information in social networks is often triggered by a small seed of members. The increased interconnectedness in social networks and the resulting influence among their members can have as a consequence a collective action expressed through a wide acceptance of the diffused information, even if it was initiated by a small seed set. Based on our study, we can understand in which cases will a piece of information be spread widely or not, based on the centrality of seed members and the attitude of other members. Specifically, for varying attitudes, we can determine the outcome of information diffusion according to the centrality scores of the seeds in each case.
- Centrality scores can be used for effective seed selection in cases where influence factors (see Section 2) are either not known or not related to the piece of information that will be diffused. As shown by our experimental results, by using appropriate centrality measures, the total amount of activated members can be increased significantly.
- A first application of our findings can be in the case we want to characterize how much accessible will become a diffused piece of information bythe members of a social network. Since information

is often spread "virally", initiated by a small seed of members, we can determine the chances that the diffused information will reach a large part of the network or not.

- A second application of our findings can be in the case we want to coordinate the diffusion of information in order to affect the behaviour (e.g., consumer behaviour) of the network members. This is the case in commercial applications, such as "viral" marketing, where we want to optimize the diffusion by carefully selecting the seeds. Our results can allow practitioners to select seeds based on different centrality scores, by taking into account the attitude of the members (which can be inferred with various methods, e.g., either explicitly through questionnaires or implicitly through responses to similar campaigns in the past).
- Finally, another application can be in the case we would like to coordinate a counter action against 'negative' or dangerous information that has started to propagate in a social network. This can be the case of a so called "firestorm" against an organization, or the deliberate spread of other malicious information. In such cases, we can first have an insight into the chances that the 'negative' information will have to widely spread or not, by investigating the centrality of the members that initiated it. Additionally, we can also get an indication about which are the most suitable members that we can use in a new seed that will start diffusing 'positive' information, which will try to compete against the spread of the 'negative' information will try to constraint its spread as much as possible.

We believe that our study can help towards: i) better understanding the influence mechanism in social networks and how it affects information diffusion; ii) providing a methodology that can find various practical applications, in order to coordinate the spread of information in social networks.

6 Conclusions

Towards understanding the role of social media in our lives and how people use them to interact with each other, in this study we focused on the influence mechanism that emerges with the ever increasing interconnectedness that we find nowadays in online social networks. This influence mechanism can cause the "viral" diffusion of information in social networks.

We examined the factors that affect information diffusion in social networks, by investigating them in the context of the influence maximization problem. A key motivation in our approach is that existing approaches either assume complete knowledge about influence factors (Kempe, Kleinberg & Tardos 2003) or about related actions of members of a social network that can help to estimate the influence factors (Goyal, Bonchi & Lakshmanan 2012). However, these approaches may not be easily applicable, since knowledge about both the influence factors or related actions that can help their estimation may not be known. In our study we aimed at recognizing the relation between the influence of the seed members that trigger information diffusion and the attitude that the rest members have towards the diffused information. To characterize the influence of seed members in a social network, we measured their centrality within the structure of the network, based on the premise that more central members will be able to reach a larger part of the network. We provided a comprehensive study of several centrality scores that were usedto identify the influence of seed members.

We conducted a thorough experimental evaluation of the examined scores with data from a real online social network. Our experimental results contribute to the existing research, by recognizing the interplay of the various centrality scores, which characterize the influence of seed members, with the members' attitude, and how this interplay affects the outcome of information diffusion.

In our future work, we will investigate a larger variety of social network topologies. This will allow us to further analyze the effect of centrality measures with respect to the existence of various types of communities within the network. Additionally, we will examine the patterns with which information is being diffused and whether the various centrality measures result into different patterns, by causing diffusion to propagate through paths whose nature is more in depth or in width.

Acknowledgements

Work supported by Katholische Universität Eichstätt-Ingolstadt with the PRO FORschung (PROFOR) project "Innovative Campaign Management System for Viral Marketing in Social Networks".

References

Aral, S. (2012). Social science: Poked to vote. Nature, 489 (7415), 212-214.

Bampo, M. Ewing, M. T., Mather, D. R., Stewart, D., and Wallace, M. (2008). The effects of the social structure of digital networks on viral marketing performance. Information Systems Research, 19 (3), 273-290.

Bonchi, F., Castillo, C., Gionis, A. and Jaimes, A. (2011). Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2 (3), 22.

Borgatti, S. P. and Everett, M. G. (2006). A graph-theoretic perspective on centrality. Social networks, 28 (4), 466-484.

Borgatti, S. P., Mehra, A., Brass, D. J. and Labianca, G. (2009). Network Analysis in the Social Sciences. Science, 323 (5916), 892-895.

Coulter, K. S. and Roggeveen, A. (2012). Deal or no deal?: How number of buyers, purchase limit, and time-to-expiration impact purchase decisions on group buying websites. Journal of Research in Interactive Marketing, 6 (2), 78-95.

Doer, B., Fouz, M. and Friedrich, T. (2012). Why rumours spread so quickly in social networks. Communications of the ACM, 55 (6), 70-75.

Easley, D. and Kleinberg, J. (2010). Networks, crowds, and markets. Cambridge University Press.

Goldenberg, J., Libai, B. and Muller, E. (2001). Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. Academy of Marketing Science Review, 9 (3), 1-18.

Goyal, A., Bonchi, F. and Lakshmanan, L. V. (2012). A data-based approach to social influence maximization. Proceedings of the VLDB Endowment, 5 (1), 73-84.

Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 1360-1380.

Granovetter, M. (1978). Threshold models of collective behavior. American journal of sociology, 1420-1443.

Hinz, O., Skiera, B., Barrot, C. and Becker, J. U. (2011). Seeding strategies for viral marketing: An empirical comparison. Journal of Marketing, 75 (6), 55-71.

Jurvetson, S. (2000). What exactly is viral marketing. RedHerring, 78, 110-112.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53 (1), 59-68.

Kempe, D., Kleinberg, J. and Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, August, 2003, pp. 137-146.

Kleinberg, J. (2007). Cascading behavior in networks: Algorithmic and economic issues. Algorithmic game theory, 613-632.

Leskovec, J., Adamic, L. and Huberman B. (2007).The Dynamics of Viral Marketing. ACM Transactions on the Web, 1 (1).

Nemhauser, G. L., Wolsey, L. A. and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. Mathematical Programming, 14 (1), 265-294.

Newman, M. E. J. (2010). Networks: An Introduction. Oxford, UK: Oxford University Press.

Porter, C. E. and Donthu, N. (2008). Cultivating trust and harvesting value in virtual communities. Management Science, 54 (1), 113-128.