Two Evidential Data Based Models for Influence Maximization in Twitter

Siwar Jendoubi, Arnaud Martin, Ludovic Liétard, Hend Ben Hadji, Boutheina Ben Yaghlane

Two Evidential Data Based Models for Influence Maximization in Twitter

Siwar Jendoubi[a,b,d,*], Arnaud Martin[b], Ludovic Liétard[b], Hend Ben Hadji[d], Boutheina Ben Yaghlane[c]

[a] *Université de Tunis, ISG Tunis, LARODEC*
[b] *Université de Rennes I, IRISA*
[c] *Université de Carthage, IHEC Carthage, LARODEC*
[d] *Centre d'Etude et de Recherche des Télécommunications*

Abstract

Influence maximization is the problem of selecting a set of influential users in the social network. Those users could adopt the product and trigger a large cascade of adoptions through the "word of mouth" effect. In this paper, we propose two evidential influence maximization models for Twitter social network. The proposed approach uses the theory of belief functions to estimate users influence. Furthermore, the proposed influence estimation measure fuses many influence aspects in Twitter, like the importance of the user in the network structure and the popularity of user's tweets (messages). In our experiments, we compare the proposed solutions to existing ones and we show the performance of our models.

Keywords: Influence maximization, Theory of belief functions, Twitter social network, Influence measure.

1. Introduction

Social influence is defined by "changes in an individual's thoughts, feelings, attitudes, or behaviors that result from interaction with another individual or a group" [1]. Identifying influencers in online social networks has received great attention from researchers in many research fields like sociology, marketing, psychology, computer science, etc. For example, marketers look for influencers to promote their marketing campaign. In fact, influencers are able to make the product propaganda goes viral through the social network, therefore, we are on the brink of defining the problem of influence maximization.

The problem of influence maximization has been widely studied since its introduction [2, 3, 4, 5]. Its purpose is to select a set of k influential users in the social network that could adopt the product and trigger a large cascade of adoptions through the "word of mouth" effect. Consider the following example;

*Corresponding author
Email address: jendoubi.siwar@yahoo.fr (Siwar Jendoubi)

a startup company produces a new product and wants to market it and it has a small budget for that purpose. A good solution may be to get profit from online social networks (OSN) through the "word of mouth" effect. Hence, it can select a small number of initial users, it encourages them to adopt the product, for example by giving them gifts or discounts. Selected users start spreading what through the network to influence their friends and their friends influence their friend's friends until achieving reaching instead as large individuals as possible. To that end a question arises: how do the startup company select the initial set of users to maximize the awareness of the product. The main goal of the influence maximization is to find solutions to that problem.

The authors in [2] were the firsts to introduce the problem of identifying influencers for a marketing campaign as a learning problem. They modeled the customer's network value, *i.e.* "the expected profit from sales to other customers he may influence to buy, the customers those may influence, and so on recursively" [2] , also they considered the market as a social network of customers. Motivated by the work of [2], Kempe *et al.* [3] formulated the problem as an optimization problem which is proven to be NP-Hard. They assumed having the social network and the influence probabilities extent to which each individual influence one another. Their issue is to find/choose a set of influential individuals that maximize the spread of the marketing message within the network.

Real world is full of imprecision and uncertainty and this fact will necessarily reflect on OSN data. The imprecision of the information is characterized by its content. In fact, it is related to the information or to the source. It measures a quality issue of the knowledge. The uncertainty of the information characterizes the degree of its conformity to the reality. Therefore, an uncertain information describes a partial knowledge of the reality. In fact, social interactions can not be always precise and certain, also, OSNs allow only limited access to their data which generates more imprecision and uncertainty for the social network analysis fields. The uncertainty is due to the partial knowledge we have about the user. For example, we do not have all the user's tweets or all his relations in the network. It leads to imprecise measurements. For example, it is not possible to obtain a precise information about the user's opinion because we do not have all his tweets. Then, if we ignore this imperfection of the data, we may be confronted to obtain erroneous analysis results. In such a situation, the theory of belief functions [6, 7] has been widely applied, and it is able to well characterize the uncertain (ignorant) information and reduce the errors. We find it used, for example, in some related research fields like pattern clustering [8, 9] and classification [10]. Furthermore, this theory was used for analyzing social networks [11, 12, 13, 14, 15, 16].

In this paper, we propose two evidential influence maximization models for Twitter social network. We use the theory of belief functions to estimate the influence taking into account data imperfection. The proposed approach benefits from the performance of the mathematical framework of belief functions especially the fusion of the information. In fact, our influence measure consider many influence aspects, like the importance of the user in the network structure

and the popularity of user's tweets (messages), then, this theory manages all these aspects, fuses them and deals with uncertainty and imprecision that characterize social network data. The resulting measure is obviously more refined and more precise than taking each influence aspect separately.

Our work achieves the following contributions:

1. We propose a new evidential influence measure for the Twitter social network[1], the proposed measure considers many influence aspects which makes it more refined and precise than existing measures that are proposed for Twitter.
2. We use the theory of belief functions to combine influence indicators like the number of followers, the number of retweets and the number of mentions, that are generally considered separately while the analysis of the influence on Twitter.
3. We introduced a new influence aspect which is summarized by the fact that "I am more influencer if I am connected to influencer users" and we considered it in our measure.
4. We maximized the influence based on the proposed influence measure.
5. We show that influence maximization under our model is NP-Hard. Nevertheless, we show that the function defining the influence propagation is monotone and sub-modular. Consequently, we develop a greedy based algorithm that guarantees a good approximation to the optimal solution.
6. We conduct our experiments on real word data set that we collected from Twitter.

The remainder of this paper is organized as follows: section 2 introduces some basic concepts of the theory of belief functions, section 3 discusses some related works, section 4 presents the proposed evidential influence maximization model and section 5 provides results from our experiments.

2. Background: Theory of belief functions

In this section, we introduce some basic concepts of the theory of belief functions that were used in the proposed approach. Dempster [6] proposed the *Upper and Lower probabilities* that are considered as the first ancestor of the evidence theory, also called Dempster-Shafer theory or belief functions theory. Then [7] introduced the *Mathematical theory of evidence* and defined the basic mathematical framework of the evidence theory, often called *Shafer model*. The main goal of the Dempster-Shafer theory is to achieve more precise, reliable and coherent information.

Let $\Omega = \{s_1, s_2, ..., s_n\}$ be the *frame of discernment*. The basic belief assignment (BBA), m^Ω, represents the agent belief on Ω, it is defined as:

$$\begin{array}{rcl} 2^\Omega & \rightarrow & [0,1] \\ A & \mapsto & m^\Omega(A) \end{array} \quad (1)$$

[1] www.twitter.com

where $2^{\Omega} = \{\emptyset, \{s_1\}, \{s_2\}, \{s_1, s_2\}, ..., \{s_1, s_2, ..., s_n\}\}$ is the *power set* (set of all subsets) of Ω. $m^{\Omega}(A)$ is the mass value assigned to $A \subseteq \Omega$. The mass function m^{Ω} must respect:

$$\sum_{A \subseteq \Omega} m^{\Omega}(A) = 1 \tag{2}$$

In the case where we have $m^{\Omega}(A) > 0$, A is called *focal element* of m^{Ω}. The mass value given to the set Ω, $m^{\Omega}(\Omega)$, is the mass that cannot be given to its subsets and it is called total ignorance. When we compare a BBA distribution to a probability distribution, we notice that the BBA allows a subset of Ω to be a focal element when we have some doubt about the decision, while the probability theory forces the equiprobability in such a case.

Combination rules in the evidence theory are the main tool that can be used for information fusion. *Dempster's rule* was the first defined in this theory [6], it is used to fuse two distinct BBAs, defined on Ω, that come from two different sources describing the same event. It is defined as:

$$m^{\Omega}_{1\oplus 2}(A) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A} m^{\Omega}_1(B)\, m^{\Omega}_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} m^{\Omega}_1(B)\, m^{\Omega}_2(C)}, & A \subseteq \Omega \setminus \{\emptyset\} \\ 0 & \text{if } A = \emptyset \end{cases} \tag{3}$$

To make a decision with the belief functions framework, we can use the pignistic transformation [17] to get a probability distribution from a BBA distribution. Then, for each element in the frame of discernment Ω, we compute its pignistic probability as follows:

$$\mathrm{BetP}^{\Omega}(s_i) = \sum_{A \in 2^{\Omega}, s_i \in A} \frac{m^{\Omega}(A)}{\mid A \mid (1 - m(\emptyset))}, \; s_i \in \Omega \tag{4}$$

3. Related work

Influence maximization (IM) is the problem of finding a set of k seed nodes that are able to influence the maximum number of nodes in the social network. In the literature, we find many solutions for the IM problem. In this section, we present an overview of the state of the art. First, we introduce the influence maximization basic models that use a diffusion model, then we present data based models. After, we talk about influence in Twitter and finally we present measures that use the theory of belief functions to estimate influence.

3.1. Diffusion model based influence maximization

Given a social network $G = (V, E)$, V is a set of vertices, E is a set of edges and a diffusion model M, the influence maximization (IM) problem is to select a set S of k influential users (called seed set) that maximizes the awareness of the "product" over the social network G [3]. In other words, it is the problem

of choosing S seed nodes that maximize the expected number of influenced nodes, $\sigma_M(S)$, that will adopt the "product". Maximizing $\sigma_M(S)$ is a NP-Hard problem. Authors in [3] prove that $\sigma_M(S)$ is monotone, *i.e.*

$$\sigma_M(S) \leq \sigma_M(T) \tag{5}$$

whenever $S \subseteq T \subseteq V$ and sub-modular, *i.e.*

$$\sigma_M(S \cup \{x\}) - \sigma_M(S) \geq \sigma_M(T \cup \{x\}) - \sigma_M(T) \tag{6}$$

whenever $S \subseteq T \subseteq V$ and $x \in V$, hence, they used the greedy algorithm with the Monte Carlo simulation to extract the seed set. To estimate $\sigma_M(S)$, [3] propose two propagation simulation models which are the *Linear Threshold Model (LTM)* [18] and the *Independent Cascade Model (ICM)* [19]. In these models we suppose that we have a social graph $G = (V, E)$, a vertex v is said to be *active* if it received the information and accepted it. It is said to be *inactive* if it did not receive the information or rejected it. An inactive node can become active. In the LTM model we associate a *weight* to each edge $\omega(u, v)$ and a *threshold* θ_u to each vertex. A vertex u will be activated if the total weight, between it and its activated neighbors, is at least θ_u, *i.e.*

$$\sum_v \omega(u, v) \geq \theta_u \tag{7}$$

The threshold θ_u is a random uniform variable chosen from $[0, 1]$, it "intuitively represents the different latent tendencies of nodes to adopt the innovation when their neighbors do" [3]. In the ICM model each newly activated node is given only one chance to activate its inactive neighbors. For instance, at the step t, a newly activated node u will try to activate its inactive neighbor v, the success probability of u to activate v is given by $p(u, v)$ (parameter of the system). A special case of ICM is *Weighted Cascade (WC)* where

$$p(u, v) = \frac{1}{D_u} \tag{8}$$

such that D_u is the overall degree of the vertex u.

In the literature, many works were conducted to improve the running time when considering ICM and LTM. The work of [20] introduced the Cost Effective Lazy Forward (CELF) algorithm. It exploited the sub-modularity property of the function to be maximized and proved to be 700 times faster than the solution of [3]. Authors in [21] introduced SPIN (Sparcification of influence network). It is an instance of the ICM, that reduces the complexity by network simplification. It starts by selecting the k edges that are most likely to explain the propagation, then it applies the greedy algorithm in order to select arcs that increase the likelihood. After network simplification, it applies the ICM algorithm to detect users that maximize influence. The work of [22] presented the "Continuously activated and Time-restricted IC (CT-IC) model" that generalizes the ICM. CT-IC model gives to each active node many chances to activate its neighbors and these chances are processed until a given time. Other works tried to consider other parameters to improve the quality of the selected seed nodes. Among these parameters we find the topic [23, 24], trust [25, 26], time [27], etc.

3.2. Data based influence maximization

In the literature, we find that influence probabilities are either uniform *i.e.* $p(u,v) = 0.01$, or selected uniformly at random from the set $\{0.01, 0.001, 0.0001\}$ or computed as in the weighted cascade *i.e.* $p(u,v) = \frac{1}{D_u}$ (D_u is the overall degree of the vertex u). The work of [4] introduces many data based models to learn influence probabilities. In their paper they consider the static case, the continuous time case and the discrete time case, for more details the reader can refer to [4].

The credit distribution (CD) [5] is, also, a data based approach that investigates past propagation to detect users of influence. It uses past propagation actions to associate an influence credit to each user in the network. The influence spread is defined as the total influence credit given to a set of users S from the whole network. The idea behind this algorithm is that; when an action a propagates from a user u to a user v, a direct influence credit, $\gamma(u,v)(a)$, is given to u. Also, a credit amount is given to predecessors of u in the propagation graph. The first step of the credit distribution algorithm consists on scanning the action log L (a data structure that is defined as the set of tuples $(User, Action, Time)$ such that $(u,a,t) \in L$ means that the user u performed the action a at time t) to compute the total credit given to u for influencing its neighbor v for the action a, $\Gamma(u,v)(a)$. S the set of seed nodes is initialized to $\emptyset$. In the second step, the algorithm runs up the CELF algorithm to select the node with the maximum marginal gain and so on until getting all needed seed nodes. For more details the reader can refer to [5].

As the work of [5], our work is data based. However, our approaches differ from it in the following ways. First, we propose an influence measure that considers many influence aspects like the structure of the network, the influence of the user's friends, the user's popularity etc. Second, we use the theory of belief functions to combine all pieces of information from each influence aspect in order to model uncertainty and imprecision and to manage the conflict between the pieces of information.

3.3. Influence in Twitter

Actually, Twitter is one of the most popular micro-blogging services. It allows its users to follow updates from each other via the "follow" relationship. For example, let A and B be two Twitter users, then, if A is interested by updates from B, A can simply "follow" it and A will receive all the messages (called tweets) from B in its actuality time-line, also, Twitter users can have access to public tweets that appear in a public time-line. The follow relationship can either be reciprocated or one way. Twitter enables its users to send and read short 140-character messages called "tweets". Besides, Twitter users can spread tweets from others and share them with their own followers using the "retweet" mechanism. Furthermore, users are able to send tweets directly to other users by mentioning their username prefixed with an "@" sign.

In the literature, influence in Twitter was widely studied. In [28] the authors present an in-depth comparison of three influence measures which are indegree

(follow), retweets and mentions. Authors in [29] measure the user influence in Twitter using the K-Shell decomposition algorithm that takes as input the followership network and gives as a result an influence value for each user, also, they modify the basic algorithm to assign to each user a logarithmic K-Shell influence value. The work of [30] proposes InterRank measure that improves the PageRank measure by considering not only the follower relationship of the network but also the topical similarity between Twitter users. In [31] the authors define Twitter influencers as "active actors who have the ability to spread information and inspire other people in the network" and they propose InfRank algorithm that identifies influencers according to their retweet activity. In [32] authors compare six influence metrics (like indegree, eigenvector centrality and clustering coefficient) that are commonly used to identify influential users in Twitter. The work of [33] studies the influence of the "information value" of the tweet (content criteria) and the agent awareness (context criteria) on the retweeting decision.

To sum up, in this section we presented works that searched to estimate influence in Twitter. All these works used only one criterion in each proposed measure, *i.e.* [29] used the follow relationship, [31] used the retweet relationship. Some of these works tried to compare influence measures separately like the work of [28] and [32]. To the best of our knowledge, our work is the first that introduces an influence measure that fuses many influence aspects in Twitter and we used the proposed measure to maximize influence.

3.4. Influence and evidence theory

In the literature, we find two works [11, 12] that talk about identifying influence nodes under the framework of the evidence theory. The work of [11] defines an evidential centrality (EVC) measure that works in a weighted network, *i.e.* network centrality is a measure that searches to identify important nodes in the network according to one or more criteria. They define two BBAs distributions on the frame $\{high, low\}$, *i.e.* high influence and low influence, for each node in the network. The first BBA is used to measure the degree centrality and the second one is used to measure the strength centrality of the node. Finally, the proposed centrality measure is the result of the combination of these BBAs.

Authors in [12] propose another centrality measure that avoids some drawbacks of the measure of [11]. In fact, they modify the EVC measure according to the actual degree of the node instead of following the uniform distribution, also, they extend the semi-local centrality measure [34] to be used with weighted networks. Their centrality measure is the result of the combination of the modified EVC and the modified semi-local centrality measure. Authors in [12] and in [11] use the same frame of discernment, their measures are structure based and they choose the influential nodes to be top-1 ranked nodes according to the proposed centrality measure.

In this paper, we use the BBA estimation mechanism of [11, 12] in our approach. Nevertheless, our influence measure fuses more influence parameters.

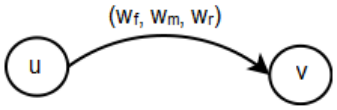


Figure 1: Weight vector between u and v.

4. Proposed evidential influence maximization models

In this paper, we propose two new models to maximize the influence on twitter social network. We use the theory of belief functions to overcome the problem of data imperfection. In fact, twitter API does not allow the access to all Twitter data, in fact there are a limited number of data requests by hour which causes the imperfection of the data: uncertainty, imprecision, lack of data, *etc*, and to fuse many influence aspects in Twitter that were studied separately in several works in the literature like the work of [28] and [32]. Furthermore, we assume that an influencer on Twitter has to be: active by tweeting frequently, followed by several users in the network that are interested by his actuality, frequently mentioned in other users' tweets and his tweets are retweeted many times. In this section we present two evidential influence maximization models that consider our assumption while measuring user's influence.

4.1. Weights computation

In Twitter social network there are three possible relations: the first one is explicit which is the follow relation, *i.e.* the follow relation is created when a given user follows another user, the second and the third relations are implicit which are the mention and the retweet, *i.e.* we obtain these implicit relations when we collect the user's tweets, then when a user mentions or retweets another user we create a new implicit link or we update an existing one. Another property of Twitter, is that between two given users u and v we can have a follow, a mention and/or a retweet relation. We assign to each link (u, v) a vector of three weights, *i.e.* follow weight, mention weight and retweet weight, that has the form (w_f, w_m, w_r) as shown in Figure (1). The follow weight w_f measures the strength of the followership between u and v, *i.e.* when the direct followership relation is broken, w_f measures the fact that u still receives v's tweets via intermediary users between them. The mention weight w_m weights information exchange between users u and v, indeed, when u mentions v in a tweet then this second (v) will receive directly the message in his notification tab. This behavior emphasizes direct communication between twitter users. The retweet weight w_r represents the information diffusion and influence weight between users, in fact, more v retweets from u more it is influenced by u [31].

Let $G(V, E)$ be the social network where V is the set of nodes and E is the set of links. Let $S_u \subseteq V$ be the set of immediate successor of $u \in V$, $P_v \subseteq V$ the set of immediate predecessor of $v \in V$, T_u the set of tweets of u, $R_u(v)$ the set of tweets of u that were retweeted by $v \in V$, $M_u(v)$ the set of tweets of u

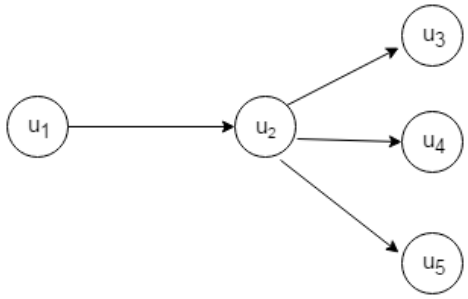


Figure 2: Follow weight example

in which v was mentioned and M_u the set of tweets in which u mentions any user in the network except himself. Ben Jabeur *et al.* [31] define weights of the follow relation, mention relation and retweet relation respectively as follows with $v \in S_u$:

$$w_f(u,v) = \frac{|S_u \cap P_v| + 1}{|S_u|} \quad (9)$$

$$w_m(u,v) = \frac{|M_u(v)|}{|M_u|} \quad (10)$$

$$w_r(u,v) = \frac{|R_u(v)|}{|T_u|} \quad (11)$$

These measures propose to estimate the link weights at a local level, *i.e.* relatively to the source of the link. We noticed that Ben Jabeur *et al.* weights are not suitable to our case. Indeed, in the case where u, the source of the link, has few successors, *i.e.* small S_u, then its out links will get high follow weights and the same goes for mention and retweet weights. This fact causes erroneous results. In fact, we may be confronted to obtain users that have high influence value but they are not active *i.e.* with small $|S_u|$, small $|M_u|$ and small $|T_u|$. Let's take the example in Figure 2. If we use the equation 9 to estimate $w_f(u_1,u_2)$, we will obtain $w_f(u_1,u_2) = 1$. The value of $w_f(u_1,u_2)$ means that the relationship between u and v is very strong, *i.e.* if the direct link between them is broken v still receive news from u, which is not the case.

To remedy this problem, we modify the definitions of [31] and we estimate the weights with respect to the whole network as follows:

$$w_f(u,v) = \frac{|S_u \cap P_v| + 1}{|S_{max}|} \quad (12)$$

$$w_m(u,v) = \frac{|M_v(u)|}{|M_{max}|} \quad (13)$$

$$w_r(u,v) = \frac{|R_u(v)|}{|T_{max}|} \quad (14)$$

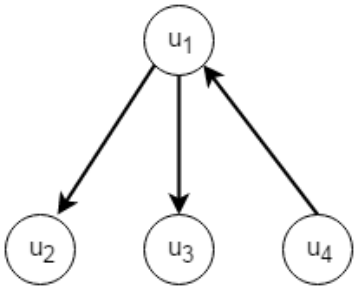


Figure 3: Network example

Link	w_f	w_m	w_r
(u_1, u_2)	0.3	0.4	0.2
(u_1, u_3)	0.4	0.3	0.1
(u_4, u_1)	0.5	0.4	0.3

(a) Links weights

Node	w_f	w_m	w_r
u_1	0.7	0.7	0.3
u_2	0	0	0
u_3	0	0	0
u_4	0.5	0.4	0.3

(b) Nodes weights

Table 1: Links and nodes weights

such that $S_{max} = \max_{u \in V} S_u$, $M_{max} = \max_{u \in V} M_u$, $T_{max} = \max_{u \in V} T_u$ and $M_v(u)$ is the set of tweets of v in which u was mentioned.

In the next step, we need to estimate the weights in the node level. In fact, we compute the three weights for each node in the network, *i.e.* we compute a follow weight, a retweet weight and a mention weight. Thus, for each node in the network we sum its out links weights as:

$$w_x(u) = \sum_{v \in V} w_x(u, v) \tag{15}$$

where $w_x(u) \in \{w_f(u), w_r(u), w_m(u)\}$ and
$w_x(u, v) \in \{w_f(u, v), w_r(u, v), w_m(u, v)\}$.

Let's take the network example given in Figure 3, in this example, we have a social network of four users related to each other by three links. Suppose that after applying the process of link weights estimation described above for each link, we obtain weights given in Table 1a. To compute each node weights, we sum up its outlinks weights, then the follow weight of the node u_1 is $w_f(u_1) = w_f(u_1, u_2) + w_f(u_1, u_3) = 0.3 + 0.4 = 0.7$. Nodes weights are given in Table 1b.

We attract the reader's attention to the fact that we use the sum function to aggregate users weights for its simplicity, but it is possible to use other aggregation functions like the mean for example. In the next section, we focus on the influence estimation and we present step by step our method for estimating influence.

4.2. Influence estimation

In this section, we present our method to estimate the influence. We introduce a new influence measure for Twitter users, the novelty of this measure is that it is an evidential measure and it contracts many influence aspects. Let $\Omega = \{I, P\}$ be our frame of discernment: I models the user's influence and P the user's passivity, a user cannot be influencer and passive at the same time, and let $G = (V, E, W)$ be a directed graph where V is the set of nodes, *i.e.* $v \in V$, $u \in V$ are nodes in G, E is the set of links, *i.e.* $(u, v) \in E$ is the link that has u as a source and v as a destination and W is the set of weights vectors, *i.e.* $(w_f(u,v), w_m(u,v), w_r(u,v)) \in W$ is the weight vector associated to (u, v). The influence estimation process contains three basic steps: in the first step, we estimate a BBA distribution for each node in the network, this BBA summarizes many influence aspects that are related to the node. In the second step, for each node u we use its estimated BBA (the result of step one) to update its in-links weights, *i.e.* links having u as destination. In the last step, we use the updated weights to estimate a BBA distribution that contracted many influence aspects.

Step 1: Node level. Let $N_{min_x} = \min_{u \in V} w_x(u)$ and $N_{max_x} = \max_{u \in V} w_x(u)$. Then, for each node in the network, we estimate a mass distribution for each variable, *i.e.* Follow, Mention and Retweet, using their weights. For each $u \in V$, and for each weight $w_x(u) \in \{w_f(u), w_r(u), w_m(u)\}$, we estimate a mass distribution as follows [11, 12]:

$$m^{\Omega}_{x_{(u)}}(I) = \frac{w_x(u) - N_{min_x}}{\gamma_x} \tag{16}$$

$$m^{\Omega}_{x_{(u)}}(P) = \frac{N_{max_x} - w_x(u)}{\gamma_x} \tag{17}$$

$$m^{\Omega}_{x_{(u)}}(\{I,P\}) = 1 - \left(m^{\Omega}_{x_{(u)}}(I) + m^{\Omega}_{x_{(u)}}(P)\right) \tag{18}$$

where $\gamma_x = N_{max_x} - N_{min_x} + \alpha$, $\alpha \in [0, 1]$. The mass value given to the set $\Omega = \{I, P\}$ is the total ignorance mass value that cannot be given to singletons. In fact this mass models the uncertainty and the imprecision. At the end of this step, we have three BBA distributions defined on Ω, *i.e.* follow BBA, mention BBA and retweet BBA, for each node in the network. Then, we combine all these BBAs using the Dempster's rule of combination (equation (3)), *i.e.* $m^{\Omega}_{(u)} = \left(m^{\Omega}_{f_{(u)}} \oplus m^{\Omega}_{r_{(u)}}\right) \oplus m^{\Omega}_{m_{(u)}}$. After this step, we apply the pignistic transformation on the resulting combined BBA $m^{\Omega}_{(u)}$ (equation (3)). We obtain a pignistic probability distribution $BetP^{\Omega}_{(u)}$ (equation (4)). At the end of this step, we have a probability value for each node that reflects many influence aspects such that:

1. The importance of the user in the network structure. Indeed, the number of user's followers in the Twitter network reflects his structural importance.
2. The popularity of user's tweets that we measure using the number of times where user's tweets are retweeted.

3. The popularity of the user that can be measured by the number of times the user was mentioned in other user's tweet. In fact, we assume that more the user is mentioned more he is popular in the network.

Step 2: Updating weights. The main contribution of this second step is to take into account the following assumption: "*I am more influencer if I am connected to influencer users*". It means that when a given user is connected to other influencers, his personal influence increases. To consider this assumption, we update weights vector of each link in the network using the estimated pignistic probability distributions defined on the link destination node:

$$w_x^{'}(u,v) = w_x(u,v).BetP_{(v)}^{\Omega}(I) \tag{19}$$

where $w_x(u,v) \in \{w_f(u,v), w_r(u,v), w_m(u,v)\}$ and $w_x^{'}(u,v) \in \left\{w_f^{'}(u,v), w_r^{'}(u,v), w_m^{'}(u,v)\right\}$ is the vector of updated link weights. In this equation, we ponder the weight value given to the influence link between u and v by the influence pignistic probability of the destination node v, $BetP_{(v)}^{\Omega}(I)$. Using this step, the node v propagates its influence to its in-neighbors, *i.e.* neighbors having v as destination. Then, if the influence of v is high, the weights of its in-links will maintain a high value from their original amount and if the influence of v is low, the weights of its in-links will maintain only a low value from their influence before the updating. Therefore, if a user u is connected to many influencer users, then, his own influence will be consolidated using the proposed equation.

Step 3: Link level. In this step, we estimate a mass distribution for each weight value and for each link $(u,v) \in E$ as follows:

$$m_{x_{(u,v)}}^{\Omega}(I) = \frac{w_x^{'}(u,v) - L_{min_x}}{\delta} \tag{20}$$

$$m_{x_{(u,v)}}^{\Omega}(P) = \frac{L_{max_x} - w_x^{'}(u,v)}{\delta} \tag{21}$$

$$m_{x_{(u,v)}}^{\Omega}(\{I,P\}) = 1 - \left(m_{x_{(u,v)}}^{\Omega}(I) + m_{x_{(u,v)}}^{\Omega}(P)\right) \tag{22}$$

where $L_{min_x} = \min_{(a,b)\in E} w_x^{'}(a,b)$, $L_{max_x} = \max_{(a,b)\in E} w_x^{'}(a,b)$ and $\delta = L_{max_x} - L_{min_x} + \beta$, $\beta \in [0,1]$ is used to model an imprecise knowledge adding belief on ignorance, *i.e.* the set $\{I,P\}$, to model our uncertainty. Consequently, we obtain three BBA distributions defined on Ω, *i.e.* follow BBA, mention BBA and retweet BBA, for each link in the network. We combine them using the Dempster's rule of combination (equation (3)). As a result, we get a mass distribution, $m_{(u,v)}^{\Omega}$, for each link. The novelty of this BBA is that it fuses many influence aspects:

1. The strength of the link between u and v in the network structure that is measured by the mean of the follow weight.

2. Information exchange and propagation activities between users that is considered through the mention and the retweet weights respectively.
3. The fact of being more influencer if you are connected to influencer users.

Finally, the influence of u on v is defined as the amount of mass given to $\{I\}$ as:

$$Inf(u,v) = m^{\Omega}_{(u,v)}(I) \tag{23}$$

Next, we define the amount of influence given to a set of nodes $S \subseteq V$ for influencing a user $v \in V$. We present two estimation ways; in the first one, we consider the influence on the directly connected nodes to S and in the second one, we consider also nodes that are connected to neighbors of S. The work of Chen *et al.* [35] justifies our formulas. They affirm that when the product have some quality issues, it is more adaptable to choose influencers that have many immediate neighbors. In fact, when the influence propagates in many hops in the network, it may fall on a user that dislikes the product. Besides, when the product have a high quality, we can choose users that have a large reachable set. We estimate the influence of S on a user v as follows:

$$Inf(S,v) = \begin{cases} 1 & \text{if } v \in S \\ \sum_{u \in S} Inf(u,v) & \text{Otherwise} \end{cases} \tag{24}$$

$$Inf(S,v) = \begin{cases} 1 & \text{if } v \in S \\ \sum_{u \in S} \sum_{x \in D_{IN}(v) \cup \{v\}} Inf(u,x).Inf(x,v) & \text{Otherwise} \end{cases} \tag{25}$$

such that $Inf(v,v) = 1$ and $D_{IN}(v)$ is the set of nodes in the indegree of v. Finally, we define the influence spread $\sigma_{Bel}(S)$ under the evidential model as the total influence given to $S \subseteq V$ from all nodes in the social network:

$$\sigma_{Bel}(S) = \sum_{v \in V} Inf(S,v) \tag{26}$$

In the spirit of the IM problem, as defined by [3], $\sigma_{Bel}(S)$ is the objective function to be maximized.

4.3. Evidential influence maximization

In this section, we present the evidential influence maximization model. Its purpose is to find a set of nodes S that maximizes the objective function $\sigma_{Bel}(S)$. Given a directed social network $G = (V, E)$, an integer $k \leq |V|$, a tweet table, T, that contains user's tweets that are published in a period of time t (t is a week for example) and an activity table, A, that contains mentions and retweets that are made in t. The goal is to find a set of users $S \subseteq V$, $|S| = k$, that maximizes $\sigma_{Bel}(S)$.

Theorem 1. *$\sigma_{Bel}(S)$ is monotone and sub-modular.*

PROOF. $\sigma_{Bel}(S)$ is monotone, *i.e.* $\sigma_{Bel}(S) \leq \sigma_{Bel}(T)$, $S \subseteq T$. In fact, $\sum_{v \in V} Inf(S,v) \leq \sum_{v \in V} Inf(T,v)$. $\sigma_{Bel}(S)$ is sub-modular if and only if $\sigma_{Bel}(S \cup \{x\}) - \sigma_{Bel}(S) \geq \sigma_{Bel}(T \cup \{x\}) - \sigma_{Bel}(T)$, $S \subseteq T$, *i.e.* the marginal gain of x with respect to T is no more than the marginal gain of x with respect to S. In the case were $x \in S$, we have
$\sigma_{Bel}(S \cup \{x\}) - \sigma_{Bel}(S) = \sigma_{Bel}(T \cup \{x\}) - \sigma_{Bel}(T) = 0$, $S \subseteq T$. If $x \notin S$ we have two alternatives; if we use the formula (24), we proven that

$$MG_S(x) = 1 + \sum_{v \in V \setminus S} Inf(x,v) \tag{27}$$

$$MG_T(x) = 1 + \sum_{v \in V \setminus T} Inf(x,v) \tag{28}$$

Where $MG_S(X) = \sigma_{Bel}(S \cup \{x\}) - \sigma_{Bel}(S)$ and $MG_T(x) = \sigma_{Bel}(T \cup \{x\}) - \sigma_{Bel}(T)$. In the second case, *i.e.* the case of the formula (25), we have

$$MG_S(x) = 1 + \sum_{v \in V \setminus S} \sum_{a \in D_{IN}(v) \cup \{v\}} Inf(x,a).Inf(a,v) \tag{29}$$

$$MG_T(x) = 1 + \sum_{v \in V \setminus T} \sum_{a \in D_{IN}(v) \cup \{v\}} Inf(x,a).Inf(a,v) \tag{30}$$

In the two cases we have $S \subseteq T$ then $|V \setminus S| \geq |V \setminus T|$ which proves the sub-modularity of $\sigma_{Bel}(S)$. □

Theorem 2. *Influence maximization under the evidential model is NP-Hard.*

PROOF. To demonstrate the hardness of our approach, we show that the function given by the equation (25) can be seen as a particular case of the function of the CD model [5] that was shown to be NP Hard. If we suppose that we have one action a then

$$\gamma(u,v)(a) = \Gamma(u,v)(a) = Inf(u,v) \tag{31}$$

and

$$\Gamma(S,v) = \begin{cases} 1 & \text{if } v \in S \\ \sum_{x \in D_{IN}(v)} Inf(S,x).Inf(x,v) & \text{Otherwise} \end{cases} \tag{32}$$

$Inf(S,v)$ can be seen as $\Gamma(S,v)$ of the CD model by considering only two hops between neighbors while estimating influence. Then we prove that "2 Levels" model is NP Hard. Also, we can write the function given by (24) of the "1 Level" model by a sum of the product of two functions. Then, we show that the "1 Level" model is NP Hard. □

We showed that the influence maximization under the evidential model is NP-Hard, besides, the influence spread function is monotone and sub-modular. Therefore, the greedy algorithm performs good approximation for the optimal solution especially when we use it with formula (27) or formula (29) that computes the marginal gain of a candidate node x. We choose the cost effective lazy-forward algorithm (CELF) [20] which is a two pass modified greedy algorithm that is proved to be about 700 time faster than the basic greedy algorithm. CELF exploits the sub-modularity property of the function to be maximized, in fact, sub-modularity guarantees that marginal benefits decrease with the solution size, hence, instead of computing the marginal benefit of each expected node at each iteration, CELF computes it in the first iteration and keeps an ordered list of nodes according to their marginal benefits value for the next iteration. In the next iteration, it re-evaluates the marginal benefit for the top node then it resorts the node list. If the top node maintains its position, it will be chosen elsewhere CELF re-evaluates the marginal benefit for the new top node and so on. Algorithm (1) shows steps of the CELF based evidential influence maximization algorithm.

```
1  begin
2      S = ∅;
3      // S: the set of seed nodes
4      Q = ∅;
5      // Q: sorted list in decreasing order according to the
           marginal gain of nodes
6      for each node ∈ V do
7          marginalGain(node);
8          // marginalGain() estimate the marginal gain of the
               node
9          Q.add(node);
10     end
11     nodeMax ← Q.pop();
12     S.add(nodeMax);
13     while | S |≤ k do
14         nodeMax ← Q.pop();
15         updateMarginalGain(nodeMax);
16         // We use formula 27 or 29 to update the marginal gain
17         if nodeMax.MG ≥ Q.getFirst().MG then S.add(nodeMax);
18         else Q.add(nodeMax);
19     end
20 end
```

Algorithm 1: CELF based evidential influence maximization algorithm

Table 2: Statistics of the data set

#User	#Tweet	#Follow	#Retweet	#Mention
36274	251329	71027	9789	20300

5. Experiments and results

In this section, we conduct some experiments on real world data to compare the proposed models with existing ones. We used the library Twitter4j[2] which is a java implementation of the Twitter API to collect Twitter data. We crawled the Twitter network for the period between 08/09/2014 and 03/11/2014 and we filtered our data by keeping only tweets that talk about smartphones and users that have at least one tweet in the data base. Table (2) shows some statistics about the collected data and Figure (4) displays users' ranks based on follow, mention, retweet and tweet across our data.

To study the accuracy of the proposed influence maximization models, we use a generated dataset. In fact, we generated data in such a way one can know the influencers. Then, we obtain a useful dataset to study the accuracy of the proposed influence maximization models. Social network structure has some special characteristics that differentiate it from ordinary graphs like the small world assumption [36]. For this reason, we chose to use a real world structure. Then, we selected a random sampling of the collected network from Twitter. The sampled network contains 1010 nodes and 6906 directed links between them. In a second step, we selected a set of users that have at least 15 outlinks. As a result, we have got a set of 108 users. Next, we define, randomly, the influence on each link in the network and the selected 108 users are defined as influencers by setting maximum influence values in their outlinks. The minimum value of influence given to an influencer is a parameter of the random process.

5.1. Experiments configuration

In our experiments, we compare the proposed evidential influence maximization model with:

- Credit distribution (CD) model that we find the closest in its principle to our model.
- Independent cascade model with uniform edge probabilities (UN ICM) equals to 1%.
- ICM with trivalency edge probabilities (TV ICM), *i.e.* chosen randomly from {10%, 1%, 0.1%}.

[2]Twitter4j is a java library for the Twitter API, it is an open-sourced software and free of charge and it was created by Yusuke Yamamoto. More details can be found in http://twitter4j.org/en/index.html.

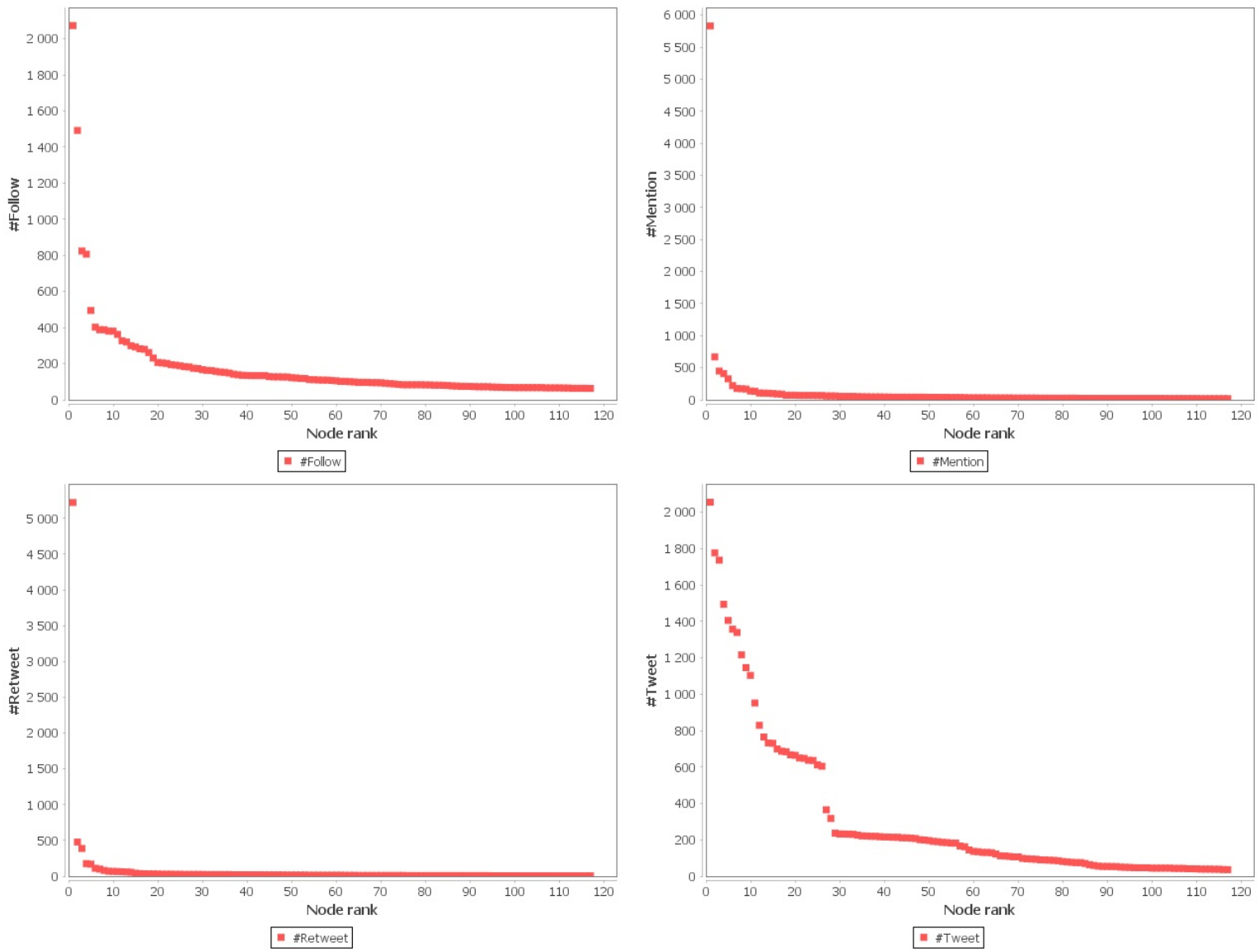


Figure 4: Data distributions

- Weighted cascade (WC ICM) *i.e.* ICM with edge probability of (u, v) equals to $\frac{1}{D_u}$.
- Linear threshold model (LTM) with uniform edge weights $\omega(u, v) = 1\%$ and random threshold θ_u for each node.

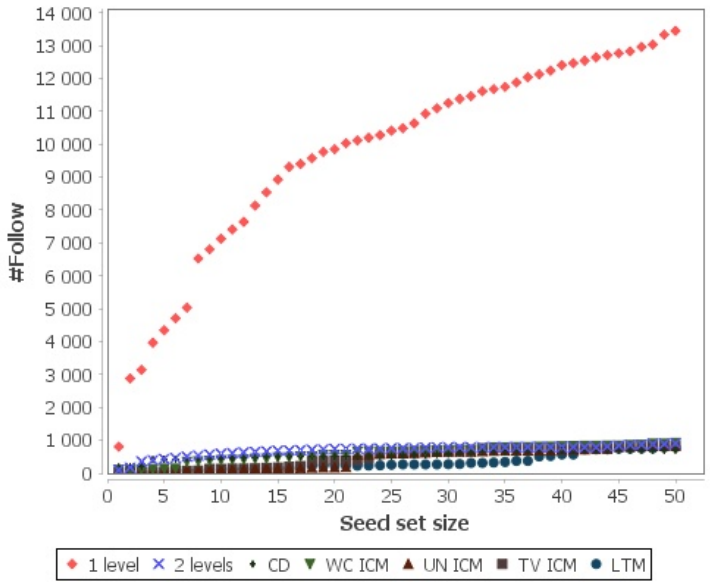
To fix ICM edge probabilities and LTM weights we followed the experiments of previous works [3, 5] and we run the algorithms 10000 times with the Monte-Carlo simulation. Furthermore, to examine the quality of the selected seeds by each method we fixed four comparison criteria which are: the number of followers, *#Follow*, the number of tweets, *#Tweet*, the number of times the user was mentioned and retweeted, *#Mention* and *#Retweet*. In fact, we assume that if a user is an influencer on Twitter he would be necessarily: very active and he has a lot of tweets, he is followed by many users in the network that are interested by his news, he is frequently mentioned in others' tweets and his tweets are retweeted several times.

5.2. Results and discussion

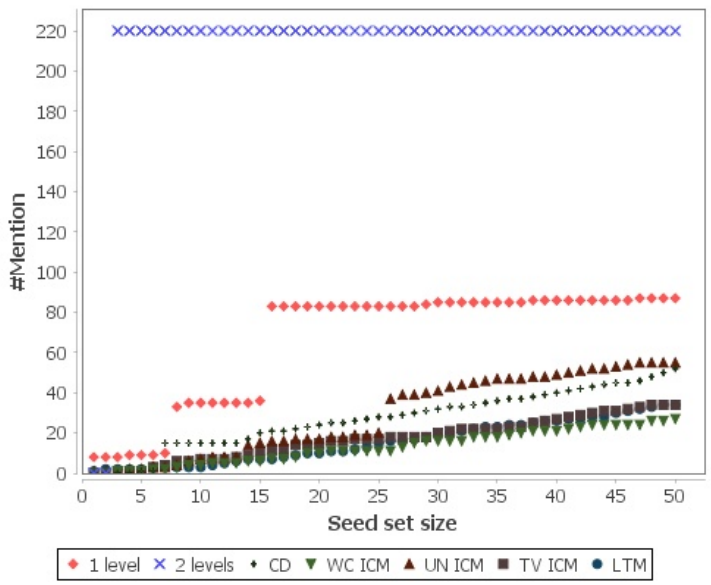
The main goal of our experiments is to show the performance of the proposed approach. We denote by "1 Level" the evidential influence maximization model that uses the formula (24) and by "2 Levels" the evidential model with the formula (25).

In Figure (5), we compare the proposed approach to some existing ones *i.e.* CD, ICM and LTM. As it was very hard to turn the basic models on the whole dataset, this fact was shown by previous works like [5], we used a sampling of 1010 nodes from the original data. In Figure (5a), we observe that "2 Levels", LTM, UN ICM, TV ICM and CD detect weakly connected users at first. However, we observe that the "1 Level" model of the proposed approach detects strongly connected users. Figure (5b) shows that most scatter plots are close to each other except that of "1 Level" and "2 Levels" that detected highly mentioned users. In Figure (5c), we observe that the best results are given by the CD model. Besides, the "2 Levels" has successfully detect highly retweeted users, also, we see that "1 Level", WC ICM, UN ICM and LTM have almost close scatter plots. Finally, Figure (5d) shows that "1 Level" model, "2 Levels" model, WC ICM, UN ICM, TV ICM, LTM and CD detected active users.

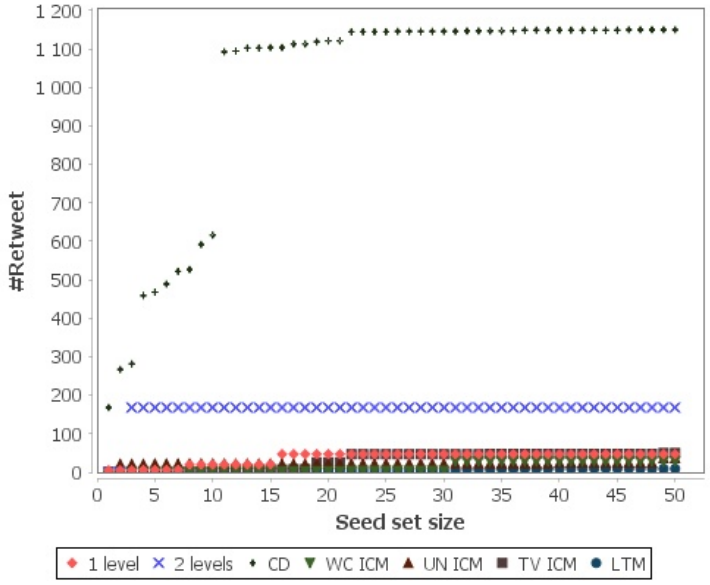
From these observations, we conclude that "1 Level" and "2 Levels" models of the proposed approach detected influencer users that are active and have a good position in the network allowing them to propagate their messages in a short time. Also, we conclude that "1 Level" is the best model in selecting influencer users. In fact, it chooses users that have a good compromise between the four criteria, *i.e.* #Follow, #Mention, #Retweet and #Tweet. In Table (3), we present the running time in milliseconds of methods in the experiment of Figure (5). In fact, all the experimented models are proven to be NP-Hard [3, 5]. As shown in Table (3) the proposed models are faster than existing models. In fact, the "1 Level" model gives its results in 62 milliseconds and the "2 Levels" in 869 milliseconds while LTM and ICM needs many thousands of seconds to give their results.



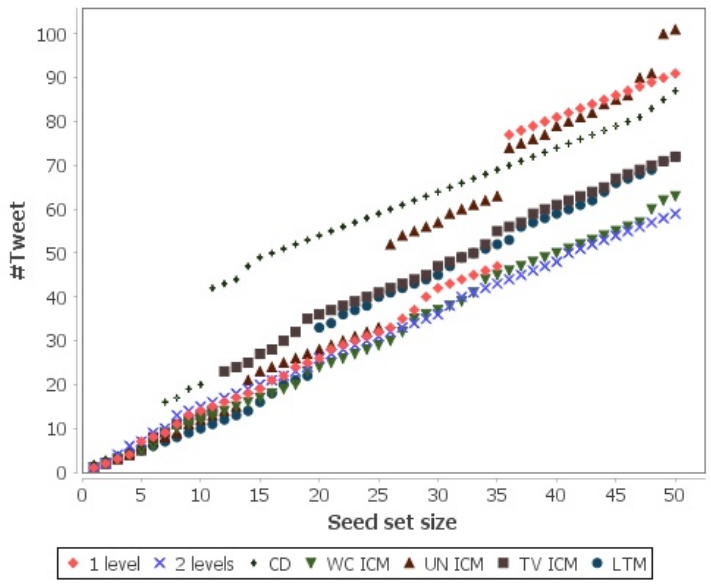
(a) Accumulated follow



(b) Accumulated mention



(c) Accumulated retweet



(d) Accumulated tweet

Figure 5: Comparison between the proposed models ("1 Level" and "2 Levels"), CD model, ICM and LTM

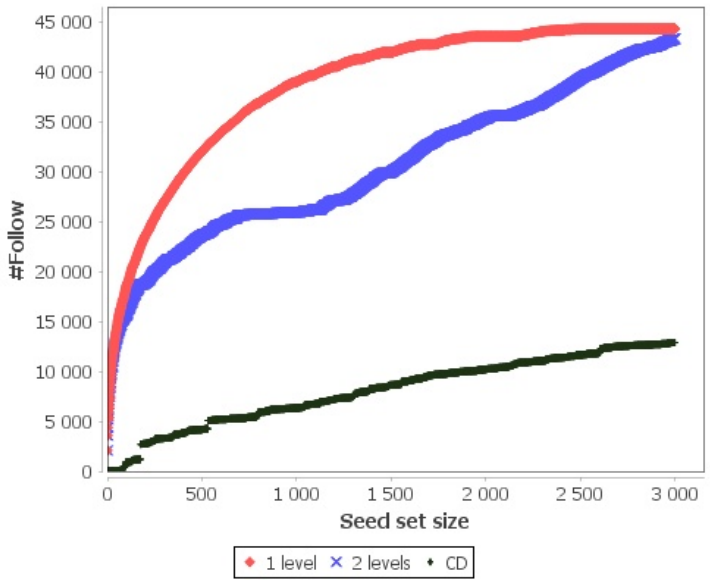
Table 3: Running time in milliseconds

Model	**Time**	**Model**	**Time**
1 Level	**62**	TV ICM	7267904
2 Level	**869**	UN ICM	4844867
CD	4654	WC ICM	4295455
LTM	65963285		

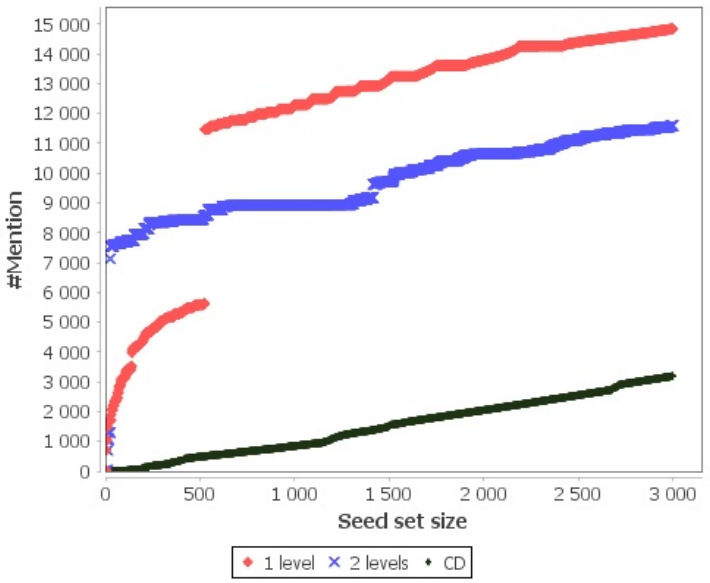
As credit distribution model is the closest in its principle to the proposed models, we use the whole dataset to compare it with "1 Level" and "2 Levels" according to the accumulated #Follow (Figures (6a) and (7a)), the accumulated #Mention (Figures (6b) and (7b)), the accumulated #Retweet (Figures (6c) and (7c)) and the accumulated #Tweet (Figures (6d) and (7d)) of seed set nodes.

Figures (6) and (7) show the performance of the proposed models (1 Level and 2 Levels) against the credit distribution (CD) model. In fact we see that the evidential influence maximization approach detects influencer spreaders that have a good compromise between #Follow, #Mention, #Retweet and #Tweet. We observe that they detected seeds that are followed by many users. Indeed, in Figure (7a) we see that the first 10 seeds are followed by over 6000 users while there are no followers for the first 10 seeds that are detected by CD model. According to Figures (6b) and (7b), detected seeds with the "1 Level" and the "2 Levels" models are mentioned many times whereas the CD model starts to detect mentioned users after over 93 seed nodes detected. In Figure (6c) we see that the CD model has successfully detected users that were retweeted a lot. However, Figure (7c) shows that this model started to detect retweeted users only after about 70 seed nodes detected while the evidential influence maximization models start detecting them from the second seed. Finally, Figures (6d) and (7d) show the accumulated activity size of the detected seeds that is measured by their number of tweets. We see that the CD model has the same behavior as in the retweet scatter plot and it starts to detect active users after about 50 seeds, in the other hand, the proposed approach demonstrates its performance in detecting active users from the second seed detected. From Figures (6) and (7) we conclude that the proposed evidential models are better than the CD model in that the evidential models provide a good compromise between the four influence criteria (#Follow, #Mention, #Retweet and #Tweet) in Twitter. In fact, the selected influencer spreaders are active, have a good position in the network, also, they are highly mentioned in others tweets and their tweets are highly retweeted. However, the CD model starts selecting followed user's after about 40 seeds, mentioned users after about 93 seeds, retweeted users after about 70 seeds and active users after about 50 seeds.

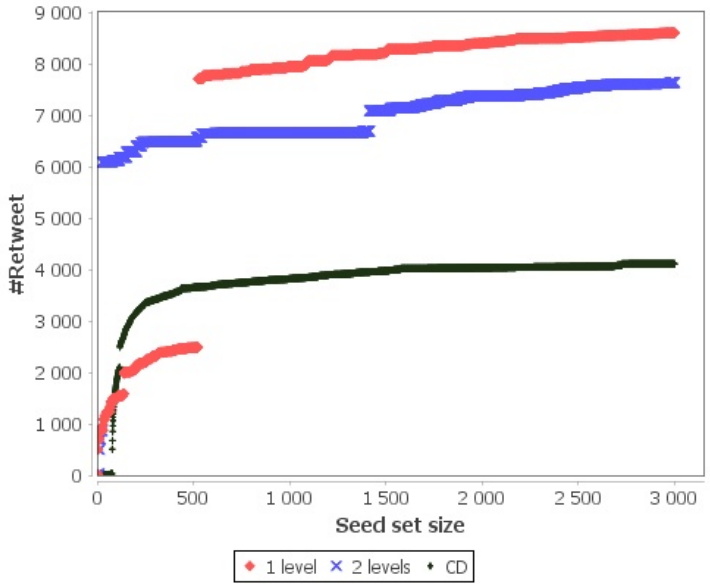
In Figure (8), we examine the number of distinct affected nodes that are connected to the influencers and to their neighbors. We observe that CD model detected about 40 isolated users at first and it started to detect users that are followed by many other users from the seed node 80. In the other hand, we notice a different behavior of scatter plots of "1 Level" and "2 Levels" models. In fact,



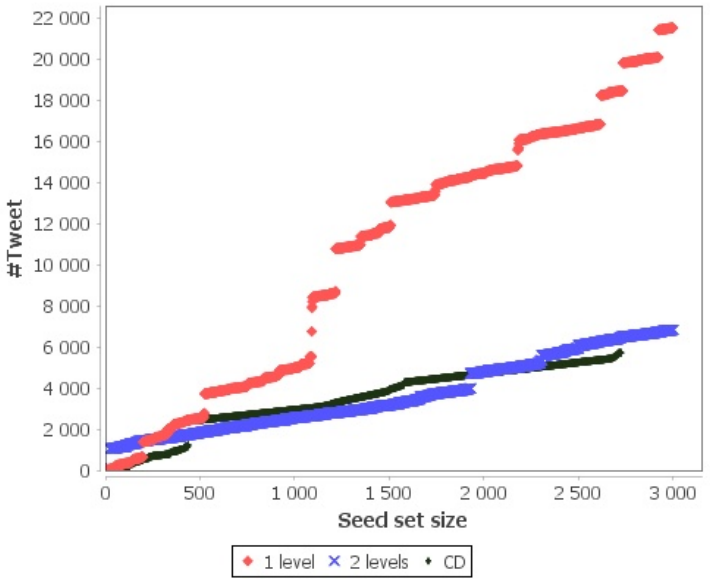
(a) Accumulated follow Comparison



(b) Accumulated mention comparison

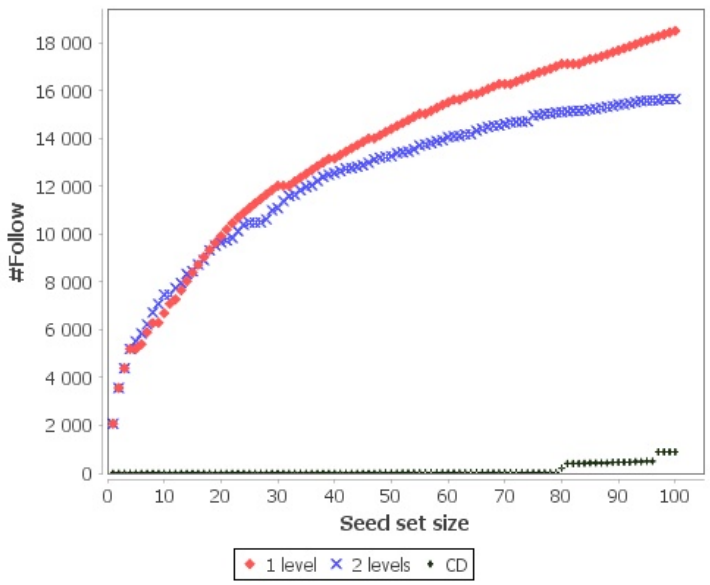


(c) Accumulated retweet comparison

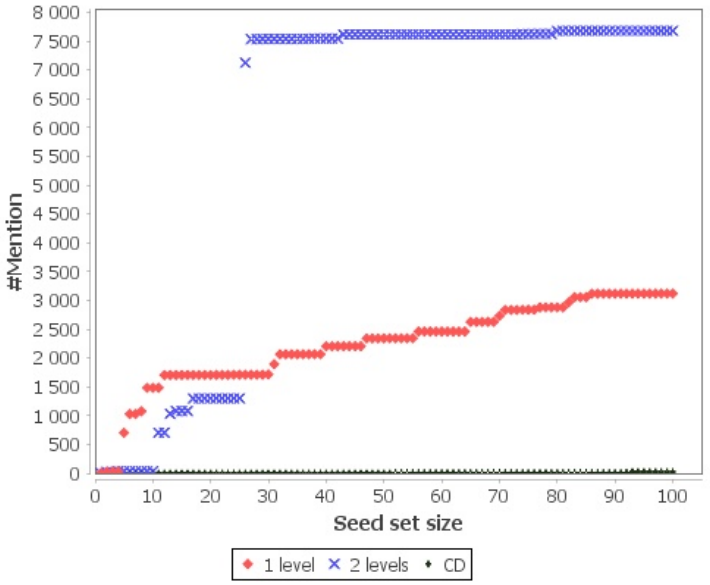


(d) Accumulated tweet comparison

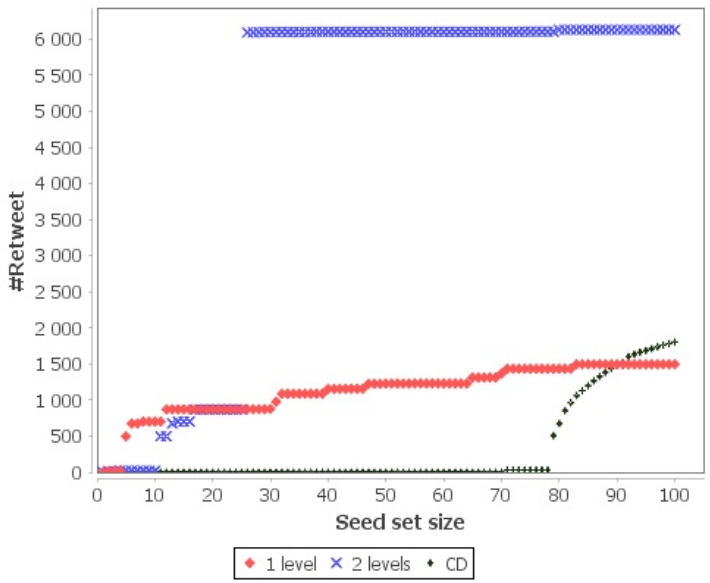
Figure 6: Comparison between the proposed models (1 level and 2 levels) and credit distribution (CD) model with S size = 3000



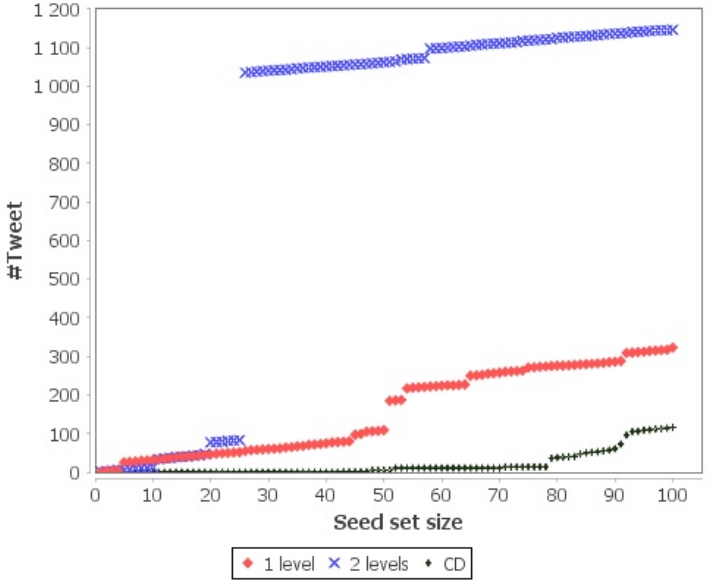
(a) Accumulated follow Comparison



(b) Accumulated mention comparison



(c) Accumulated retweet comparison



(d) Accumulated tweet comparison

Figure 7: Comparison between the proposed models (1 level and 2 levels) and credit distribution (CD) model with S size = 100

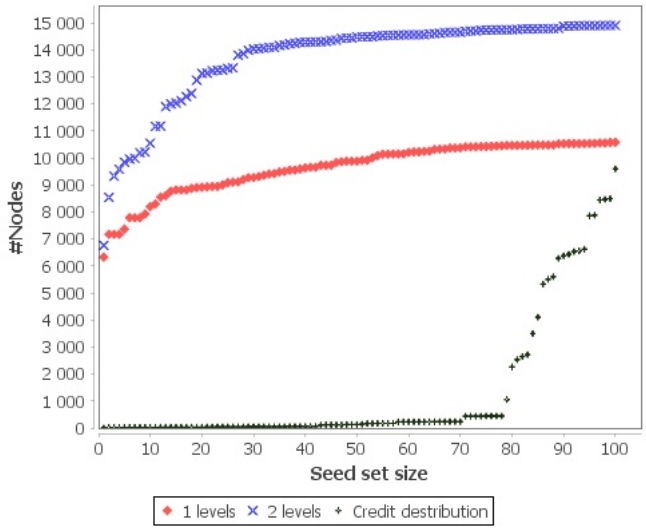


Figure 8: The dependence of the number of affected nodes to the size of S

in Figure (6a) "1 Level" scatter plot is upper than the scatter plot of "2 Levels" model. However, in Figure (8) we observe that "2 Levels" scatter plot is upper than the scatter plot of "1 Level" model. From these observations, we conclude that the "2 Levels" model detects influencer spreader that are connected to highly followed users and the "1 Level" model detects highly followed influencer spreaders. Also, we conclude that our models are better in detecting seeds than CD model. Indeed, "1 Level" and "2 Levels" models detect highly connected seeds at first. However, the CD model selects about 40 isolated seeds before starting to detect some followed seeds.

Our goal in the experiments of Figures (9) and (10) is to show the impact of considering the fact of *"being more influencer if you are connected to influencer users"* on the influence maximization results. This fact is considered in the second step, *i.e.* "Updating step", of our influence estimation process. Then we compare "1 Level" (Figure (9)) and "2 Levels" (Figure (10)) models with and without the updating step. Figure (9) shows that the difference in the "1 Level" is not very significant. However, in Figure (10) we see that the updating step ameliorates the influence maximization results for the "2 Levels" model. Indeed, when we consider the assumption of "being more influencer if you are connected to influencer users", the "2 Levels" model detects better seeds than the model without considering this assumption.

We introduce a last experiment to study and compare the accuracy of the proposed influence maximization models. For this purpose we run the "1 Level" and the "2 Levels" models on the generated data set in which the set of influencers is known. Then, we compute the hit ratio, *i.e.* the percent of correctly detected influencers, in order to compare the set of predicted k influencers with

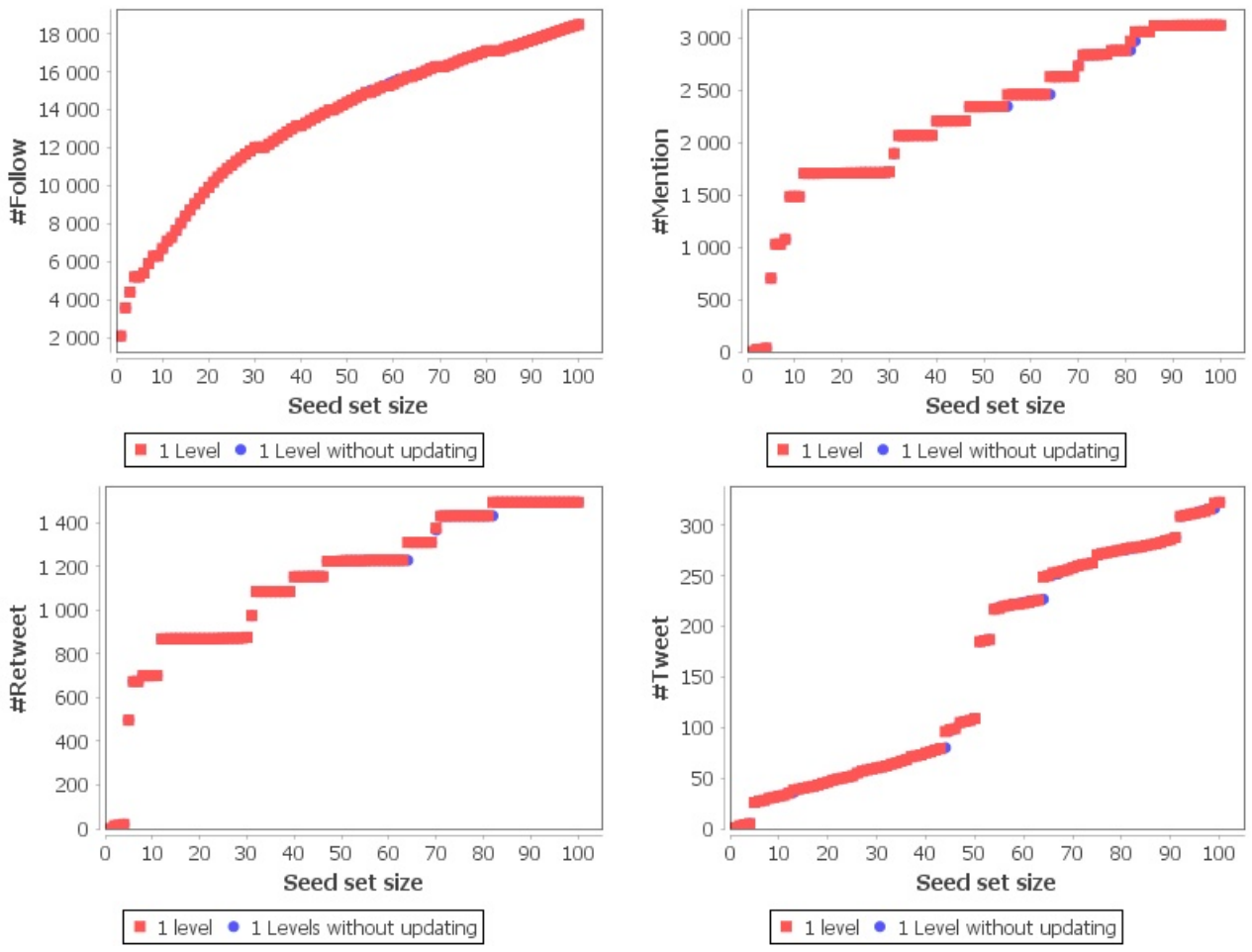


Figure 9: Impact of the weight updating step on influence maximization results: 1 Level

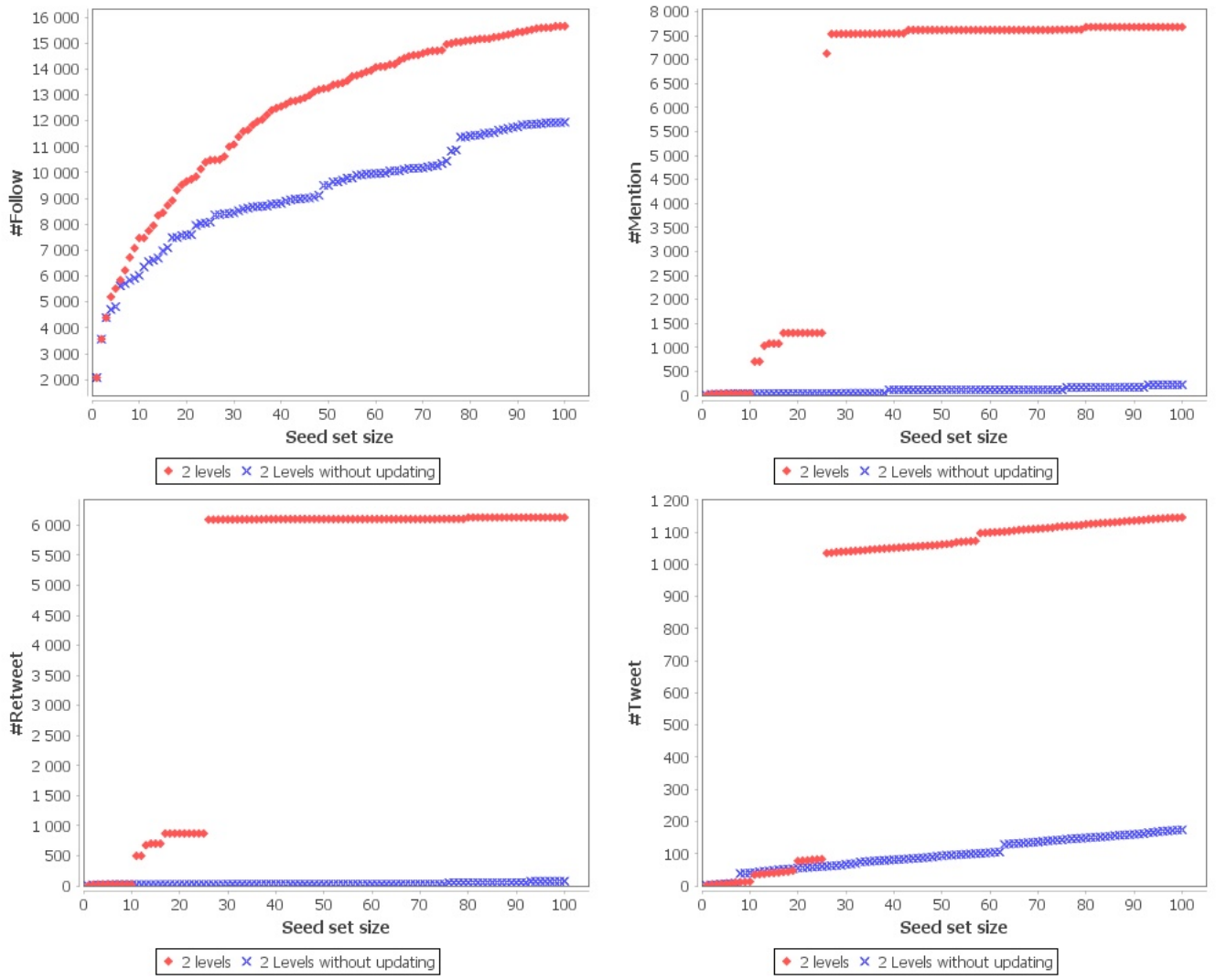


Figure 10: Impact of the weight updating step on influence maximization results: 2 Levels

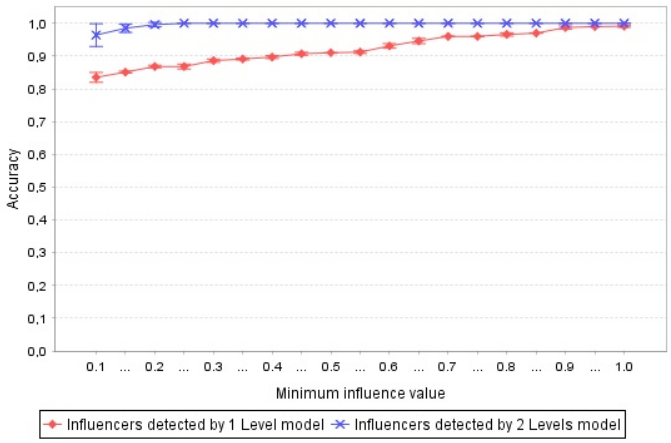


Figure 11: Accuracy of the proposed influence maximization models on generated data

the known set of influencers. As we did this experiment on generated data, then we varied the minimum value of influence given to an influencer. We fixed the size of the seed set k to 50 and we repeated the random process ten times. We obtained the results shown in Figure 11.

According to Figure 11, the proposed models have a good accuracy in detecting influencers. This figure shows the performance of the proposed models. In fact, even with a small influence value, 0.1, the experimented models succeed in detecting influencers with a good accuracy that is no less than $83\% \pm 0.01$. Besides, we notice that the "2 Levels" model have the highest accuracy values.

6. Conclusion

In this work, we introduced two new evidential influence maximization models. The proposed two models are based on a new influence estimation measure for Twitter that considers many influence aspects like the importance of the user in the network structure and the popularity of user's tweets messages. Then, we used the CELF algorithm to solve the influence maximization problem under the proposed evidential approach. To show the performance of the proposed models, we conducted some experiments to compare it with existing models. We proved that the proposed models are better than existing ones in selecting influencer users for the Twitter social network. In fact, the selected seeds performs a good compromise between the basic criteria (#Follow, #Mention, #Retweet and #Tweet) of influence in Twitter. However, we find that the CD model, for example, fails to detect good influencers at first. In fact, it detects about 40 isolated users before starting to detect followed ones. This is not the case of the proposed models. Indeed, the first selected user by our models has about 2000 followers.

In future works, we will search to improve the proposed influence measure by considering the user's profile, the topic of the message and more levels of

influence in the network. Another important objective is to adapt the proposed influence maximization model to other social networks like Facebook and LinkedIn. Finally, we will search to test the evidential influence maximization approach with larger data bases.

7. Acknowledgment

This work was done within the MOBIDOC device launched under the Support Project to the Research and Innovation System (PASRI), funded by the European Union and managed by the National Agency for the Promotion of Scientific Research (ANPR). Also, we thank the "Centre d'Etude et de Recherche des Télécommunications" (CERT) for their support.

8. References

[1] L.-S. Rashotte, *Social Influence*. Oxford: Blackwell Publishing, 10-17 2007, vol. 9, pp. 4426–4429.

[2] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of KDD'01*, 2001, pp. 57–66.

[3] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of KDD'03*, August 2003, pp. 137–146.

[4] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Learning influence probabilities in social networks," in *WSDM'10*, February 2010, pp. 241–250.

[5] ——, "A data-based approach to social influence maximization," in *Proceedings of VLDB Endowment*, August 2012, pp. 73–84.

[6] A. P. Dempster, "Upper and Lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[7] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[8] T. Denœux, S. Sriboonchitta, and O. Kanjanatarakul, "Evidential clustering of large dissimilarity data," *Knowledge-Based Systems*, vol. 106, pp. 179–195, 2016.

[9] Z.-g. Liu, Q. Pan, J. Dezert, and G. Mercier, "Credal c-means clustering method based on belief functions," *Knowledge-Based Systems*, vol. 74, pp. 119–132, 2015.

[10] Z.-g. Liu, Q. Pan, J. Dezert, and A. Martin, "Adaptive imputation of missing values for incomplete pattern classification," *Pattern Recognition*, vol. 52, pp. 85–95, 2016.

[11] D. Wei, X. Deng, X. Zhang, Y. Deng, and S. Mahadeven, "Identifying influential nodes in weighted networks based on evidence theory," *Physica A*, vol. 392, no. 10, pp. 2564–2575, Mai 2013.

[12] C. Gao, D. Wei, Y. Hu, S. Mahadevan, and Y. Deng, "A modified evidential methodology of identifying influential nodes in weighted networks," *Physica A*, vol. 392, no. 21, pp. 5490–5500, November 2013.

[13] Y. A. Kim and M. A. Ahmad, "Trust, distrust and lack of confidence of users in online social media-sharing communities," *Knowledge-Based Systems*, vol. 37, pp. 438–450, 2013.

[14] S. Jendoubi, A. Martin, L. Liétard, and B. B. Yaghlane, "Classification of message spreading in a heterogeneous social network," in *Proceeding of IPMU*, July 2014, pp. 66–75.

[15] S. Jendoubi, A. Martin, L. Liétard, B. Ben Yaghlane, and H. Ben Hadj, "Dynamic time warping distance for message propagation classification in twitter," in *Proceeding of ECSQARU*, July 2015, pp. 419–428.

[16] K. Zhou, A. Martin, Q. Pan, and Z.-g. Liu, "Median evidential c-means algorithm and its application to community detection," *Knowledge-Based Systems*, vol. 74, pp. 69–88, 2015.

[17] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *International Journal of Approximate Reasonning*, vol. 38, pp. 133–147, 2005.

[18] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, pp. 1420–1443, 1978.

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

[20] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of KDD'07*, August 2007, pp. 420–429.

[21] M. Mathioudakis, F. Bonchi, C. Castillo, A. Gionis, and A. Ukkonen, "Sparsification of influence networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, 2011, pp. 529–537.

[22] J. Kim, W. Lee, and H. Yu, "Ct-ic: Continuously activated and time-restricted independent cascade model for viral marketing," *Knowledge-Based Systems*, vol. 62, pp. 57–68, 2014.

[23] C. Aslay, N. Barbieri, F. Bonchi, and R. Baeza-Yates, "Online topic-aware influence maximization queries," in *Proceedings of the 17th International Conference on Extending Database Technology (EDBT)*, March 2014, pp. 24–28.

[24] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, 2012, pp. 81–90.

[25] S. Ahmed and C. I. Ezeife, "Discovering influential nodes from trust network," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, March 2013, pp. 121–128.

[26] R. Mohamadi-Baghmolaei, N. Mozafari, and A. Hamzeh, "Trust based latency aware influence maximization in social networks," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 195–206, March 2015.

[27] B. Liu, G. Cong, D. Xu, and Y. Zeng, "Time constrained influence maximization in social networks," in *12th IEEE International Conference on Data Mining (ICDM)*, December 2012, pp. 439–448.

[28] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, May 2010, pp. 10–17.

[29] P. Brown and J. Feng, "Measuring user influence on twitter using modified k-shell decomposition," in *Proceedings of ICWSM'11 Workshops*, 2011, pp. 18–23.

[30] J. Sung, S. Moon, and J.-G. Lee, "The influence in twitter: Are they really influenced?" in *Behavior and Social Computing*. Springer International Publishing, 2013, pp. 95–105.

[31] L. Ben Jabeur, L. Tamine, and M. Boughanem, "Active microbloggers: Identifying influencers, leaders and discussers in microblogging networks," in *Proceedings of the 19th International Symposium String Processing and Information Retrieval*, October 2012, pp. 111–117.

[32] E. Dubois and D. Gaffney, "The multiple facets of influence: Identifying political influentials and opinion leaders on twitter," *American Behavioral Scientist*, vol. 58, no. 10, pp. 1260–1277, 2014.

[33] A. Rudat and J. Buder, "Making retweeting social: The influence of content and context information on sharing news in twitter," *Computers in Human Behavior*, vol. 46, pp. 75–84, 2015.

[34] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical mechanics and its applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[35] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Procedings of SIAM SDM*, April 2011, pp. 379–390.

[36] M. E. J. Newman, *Networks: An introduction.* Oxford University Press, 2010.