A Reliability-Based Approach for Influence Maximization Using the Evidence Theory

Siwar Jendoubi[1(✉)] and Arnaud Martin[2]

[1] LARODEC, ISG Tunis, University of Tunis, Avenue de la Liberté,
Cité Bouchoucha, 2000 Le Bardo, Tunisia
jendoubi.siwar@yahoo.fr

[2] DRUID, IRISA, University of Rennes 1, Rue E. Branly, 22300 Lannion, France
Arnaud.Martin@univ-rennes1.fr

Abstract. The influence maximization is the problem of finding a set of social network users, called influencers, that can trigger a large cascade of propagation. Influencers are very beneficial to make a marketing campaign goes viral through social networks for example. In this paper, we propose an influence measure that combines many influence indicators. Besides, we consider the reliability of each influence indicator and we present a distance-based process that allows to estimate the reliability of each indicator. The proposed measure is defined under the framework of the theory of belief functions. Furthermore, the reliability-based influence measure is used with an influence maximization model to select a set of users that are able to maximize the influence in the network. Finally, we present a set of experiments on a dataset collected from Twitter. These experiments show the performance of the proposed solution in detecting social influencers with good quality.

Keywords: Influence measure · Influence maximization · Theory of belief functions · Reliability · Social network

1 Introduction

The influence maximization problem has attracted a great attention in these last years. The main purpose of this problem is to find a set of influence users, S, that can trigger a large cascade of propagation. These users are beneficial in many application domains. A well-known application is the viral marketing. Its purpose is to promote a product or a brand through viral propagation through social networks. Then, several research works were introduced in the literature [1,8,14,21] trying to find an optimal set of influence users in a given social network. However, the quality of the detected influence users stills always an issue that must be resolved.

The problem of identifying influencers was first modeled as a learning problem by Domingos and Richardson [6] in 2001. Furthermore, they defined 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" [6].

Moreover, they considered the market to be a social network of customers. Later in 2003, Kempe *et al.* [14] formulated the influence problem as an optimization problem. Indeed, they introduced two influence maximization models: the *Independent Cascade Model* (ICM) and the *Linear Threshold Model* (LTM). These models estimate the expected propagation size, σ_M, of a given node or set of nodes through propagation simulation models. Besides, [14] proved the NP-Hardness of the maximization of σ_M. Then, they proposed the greedy algorithm to approximate the set of nodes that maximizes σ_M. ICM and LTM just need the network structure to select influencers. However, these solutions are shown in [8] to be inefficient to detect good influencers.

When studying the state of the art of the influence maximization problem, we found that most of existing works use only the structure of the network to select seeds. However, the position of the user in the network is not sufficient to confirm his influence. For example, he may be a user that was active in a period of time, then, he collected many connections, and now he is no longer active. Hence, the user's activity is an interesting parameter that must be considered while looking for influencers. Besides to the user's activity in the network, many other important influence indicators are not considered. Among these indicators, we found the sharing and tagging activities of network users. These activities allow the propagation of social messages from one user to another. Also, the tagging activity is a good indicator of the user's importance in the network. In fact, more he is tagged in others' posts, more he is important for them. Therefore, considering such influence behaviors will be very beneficial to improve the quality of selected seeds.

To resolve the influencers quality issue, many influence indicators must be used together to characterize the influence that exerts one user on another [10]. An influence indicator may be the number of neighbors, the frequency of posting in the wall, the frequency of neighbor's likes or shares, *etc.* Furthermore, a refined influence measure can be obtained through the fusion of two or more indicators. A robust framework of information fusion and conflict management that may be used in such a case is the framework of the theory of belief functions [23]. Indeed, this theory provides many information combination tools that are shown to be efficient [4,24] to combine several pieces of information having different and distinct sources. Other advantages of this theory are about uncertainty, imprecision and conflict management.

In this paper, we tackle the problem of influence maximization in a social network. More specifically, our main purpose is to detect social influencers with a good quality. For this goal, we introduce a new influence measure that combines many influence indicators and considers the reliability of each indicator to characterize the user's influence. The proposed measure is defined through the theory of belief functions. Another important contribution in the paper is that we use the proposed influence measure for influence maximization purposes. This solution allows to detect a set of influencers having a good quality and that can maximize the influence in the social network. Finally, a set of experiments is made on real data collected from Twitter to show the performance of the proposed

solution against existing ones and to study the properties of the proposed influence measure.

This paper is organized as follows: related works are reviewed in Sect. 2. Indeed, we present some data-based works and existing evidential influence measures. Section 3 presents some basic concepts about the theory of belief functions. Section 4 is dedicated to explaining the proposed reliability-based influence measure. Section 5 presents a set of experiments showing the efficiency of the proposed influence measure. Finally, the paper is concluded in Sect. 6.

2   Related Works

The influence maximization is a relatively new research problem. Its main purpose is to find a set of k social users that are able to trigger a large cascade of propagation through the word of mouth effect. Since its introduction, many researchers have turned to this problem and several solutions are introduced in the literature [3,8,9,14,15]. In this section, we present some of these works.

2.1   Influence Maximization Models

The work of Kempe *et al.* [14] is the first to define the problem of finding influencers in a social network as a maximization problem. In fact, they defined the influence of a given user or set of users, S, as the expected number of affected nodes, $\sigma_M(S)$, *i.e.* nodes that received the message. Furthermore, they estimated this influence through propagation simulation models which are the *Independent Cascade Model* (ICM) and the *Linear Threshold Model* (LTM). Next, they used a greedy-based solution to approximate the optimal solution. Indeed, they proved the NP-Hardness of the problem.

In the literature, many works were conducted to improve the running time when considering ICM and LTM. Leskovec*et al.* [17] introduced the *Cost Effective Lazy Forward* (CELF) algorithm that is proved to be 700 times faster than the solution of [14]. Kimura and Saito [16] proposed the *Shortest-Path Model (SPM)* which is a special case of the ICM. Bozorgi*et al.* [2] considered the community structure, *i.e.* a community is a set of social network users that are connected more densely to each other than to other users from other communities [22,26], in the influence maximization problem.

The *Credit Distribution model* (CD) [8] is an interesting solution that investigates past propagation to select influence users. Indeed, it uses past propagation to associate to each user in the network an influence credit value. The influence spread function is defined as the total influence credit given to a set of users S from the whole network. The algorithm scans the data (past propagation) to compute the total influence credit of a user v for influencing its neighbor u. In the next step, the CELF algorithm [17] is run to approximate the set of nodes that maximizes the influence spread in the network.

2.2   Influence and Theory of Belief Functions

The theory of belief functions was used to measure the user's influence in social networks. In fact, this theory allows the combination of many influence indicators together. Besides, it is useful to manage uncertainty and imprecision. This section is dedicated to present a brief description of existing works that use the theory of belief functions for measuring or maximizing the influence.

An evidential centrality (EVC) measure was proposed by Wei *et al.* [25] and it was used to estimate the influence in the network. EVC is obtained through the combination of two BBAs defined on the frame {*high*, *low*}. The first BBA defines the evidential degree centrality and the second one defines the evidential strength centrality of a given node. A second interesting, work was also introduced to measure the evidential influence, it is the work of [7]. They proposed a modified EVC measure. It considers the actual node degree instead of following the uniform distribution. Furthermore, they proposed an extended version of the semi-local centrality measure [3] for weighted networks. Their evidential centrality measure is the combined BBA distribution of the modified semi-local centrality measure and the modified EVC. The works of [7,25] are similar in that, they defined their measures on the same frame of discernment, they used the network structure to define the influence.

Two evidential influence maximization models are recently introduced by Jendoubi *et al.* [10]. They used the theory of belief functions to estimate the influence that exerts one user on his neighbor. Indeed, their measure fuses several influence indicators in Twitter like the user's position in the network, the user's activity, *etc.* This paper is based on our previous work [10]. However, the novelty of this paper is that we not only combine many influence indicators to estimate the user's influence, but also we consider the reliability of each influence indicator in characterizing the influence.

3   Theory of Belief Functions

In this section, we present the *theory of belief functions*, also called *evidence theory* or *Dempster-Shafer theory*. It was first introduced by Dempster [4]. Next, the mathematical framework of this theory was detailed by Shafer in his book *"A mathematical theory of evidence"* [23]. This theory is used in many application domains like pattern clustering [5,19] and classification [11,12,18]. Furthermore, this theory is used for analyzing social networks and measuring the user's influence [7,10,25].

Let us, first, define the *frame of discernment* which is the set of all possible decisions:

$$\Omega = \{d_1, d_2, ..., d_n\} \tag{1}$$

The *mass* function, also called *basic belief assignment* (BBA), m^{Ω}, defines the source's belief on Ω as follows:

$$2^{\Omega} \to [0, 1]$$
$$A \mapsto m(A) \tag{2}$$

such that $2^{\Omega} = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1, d_2\}, ..., \{d_1, d_2, ..., d_n\}\}$. The set 2^{Ω} is called *power set*, *i.e.* the set of all subsets of Ω. The value assigned to the subset $A \subseteq \Omega$, $m(A)$, is interpreted as the source's support or belief on A. The BBA distribution, m, must respect the following condition:

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{3}$$

We call A *focal element* of m if we have $m(A) > 0$. The *discounting* procedure allows to consider the *reliability* of the information source. Let $\alpha \in [0, 1]$ be our reliability on the source of the BBA m, then the discounted BBA m^{α} is obtained as follows:

$$\begin{cases} m^{\alpha}(A) &= \alpha.m(A), \forall A \in 2^{\Omega} \setminus \{\Omega\} \\ m^{\alpha}(\Omega) &= 1 - \alpha.(1 - m(\Omega)) \end{cases} \tag{4}$$

The information fusion is important when we want to fuse many influence indicators together in order to obtain a refined influence measure. Then, the theory of belief functions presents several combination rules. The *Dempster's rule* of combination [4] is one of these rules. It allows to combine two distinct BBA distributions. Let m_1 and m_2 be two BBAs defined on Ω, Dempster's rule is defined as follows:

$$m_{1 \oplus 2}(A) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A} m_1(B)m_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} m_1(B)m_2(C)}, & A \subseteq \Omega \setminus \{\emptyset\} \\ 0 & if\ A = \emptyset \end{cases} \tag{5}$$

In the next section, we present some relevant existing influence measures and influence maximization models.

4 Reliability-Based Influence Maximization

In this paper, we propose an influence measure that fuses many influence indicators. Furthermore, we assume that these indicators may do not have the same reliability in characterizing the influence. Then, some indicators may be more reliable than the others. In this section, we present the proposed reliability-based influence measure, the method we use to estimate the reliability of each indicator and the influence maximization model we use to maximize the influence in the network.

4.1 Influence Characterization

Let $G = (V, E)$ be a social network, where V is the set of nodes such that $u, v \in V$ and E is the set of links such that $(u, v) \in E$. To estimate the amount of influence that exerts one user, u, on his neighbor, v, we start first by defining

a set of influence indicators, $I = \{i_1, i_2, \ldots, i_n\}$ characterizing the influence. These indicators may differ from a social network to another. We note that we are considering quantitative indicators. Let us take Twitter as example, we can define the following three indicators: (1) the number of common neighbors between u and v, (2) the number of times v mentions u in a tweet, (3) the number of times v retweets from u.

In the next step, we compute the value of each defined indicator for each link (u, v) in the network. Then, (u, v) will be associated with a vector of values. In a third step, we need to normalize each computed value to the range $[0, 1]$. This step is important as it puts all influence indicators in the same range.

In this stage, we have a vector of values of the selected influence indicators:

$$W_{(u,v)} = \left(i_{(u,v)_1} = w_1, i_{(u,v)_2} = w_2, \ldots, i_{(u,v)_n} = w_n \right) \tag{6}$$

The elements of $W_{(u,v)}$ are in the range $[0, 1]$, i.e. $w_1, w_2, \ldots w_n \in [0, 1]$, and we define a vector $W_{(u,v)}$ for each link (u, v) in the network. Next, we estimate a BBA for each indicator value and for each link. Then, if we have n influence indicators, we will obtain n BBA to model each of these indicators for a given link. Let us first, define $\Omega = \{I, P\}$ to be the frame of discernment, where I models the influence and P models the passivity of a given user. For a given link (u, v) and a given influence indicator $i_{(u,v)_j} = w_j$, we estimate its BBA on the fame Ω as follows:

$$m_{(u,v)_j}(I) = \frac{w_j - \min_{(u,v) \in E} \left(i_{(u,v)_j} \right)}{\max_{(u,v) \in E} \left(i_{(u,v)_j} \right) - \min_{(u,v) \in E} \left(i_{(u,v)_j} \right)} \tag{7}$$

$$m_{(u,v)_j}(P) = \frac{\max_{(u,v) \in E} \left(i_{(u,v)_j} \right) - w_j}{\max_{(u,v) \in E} \left(i_{(u,v)_j} \right) - \min_{(u,v) \in E} \left(i_{(u,v)_j} \right)} \tag{8}$$

After this step, the influence that exerts a user u on his neighbor v is characterized by a set of n influence BBAs. In the next section, we present the method we use to estimate the reliability of each defined BBA.

4.2   Estimating Reliability

The selected influence indicators may do not have the same reliability in characterizing the user's influence. Then, we estimate the reliability, α_j, of each influence indicator. We assume that "*the farthest from the others the indicator is, the less reliable it is*". For that purpose, we follow the approach introduced by Martin *et al.* [20] to estimate reliability. Besides, we note that this operator considers our assumption. In this section we detail the steps of [20] operator we used to estimate the reliability of each influence indicator in this paper.

Let us consider the link (u, v), we have a set of n BBAs to characterize the chosen influence indicators, $\left(m_{(u,v)_1}, m_{(u,v)_2}, \ldots, m_{(u,v)_n} \right)$. Our purpose is to estimate the reliability of each indicator against the others. To estimate the

reliability, α_j, of the BBA $m_{(u,v)_j}$, we start by computing the distance between $m_{(u,v)_j}$ and each BBA from the rest of $n-1$ BBAs that characterizes the influence of u on v, i.e. $\left(m_{(u,v)_1}, m_{(u,v)_2}, \ldots, m_{(u,v)_{j-1}}, m_{(u,v)_{j+1}}, \ldots, m_{(u,v)_n} \right)$:

$$\delta_i^j = \delta(m_j, m_i) \tag{9}$$

To estimate these distances, we can use the *Jousselme* distance [13] as follows:

$$\delta(m_j, m_i) = \sqrt{\frac{1}{2}(m_j - m_i)^T \underline{\underline{D}} (m_j - m_i)} \tag{10}$$

such that $\underline{\underline{D}}$ is an $2^N \times 2^N$ matrix, $N = |\Omega|$ and $D(A, B) = \frac{|A \cap B|}{|A \cup B|}$.

Next, we compute the average of all obtained distance values as follows:

$$C_j = \frac{\delta_j^1 + \delta_j^2 + \ldots + \delta_j^n}{n - 1} \tag{11}$$

such that $(\delta_1, \delta_2, \ldots, \delta_n)$ are the distance values between $m_{(u,v)_j}$ and $\left(m_{(u,v)_1}, m_{(u,v)_2}, \ldots, m_{(u,v)_n} \right)$, $(\delta(m_j, m_j) = 0)$. We use the average distance C_j to estimate the reliability, α_j, of the jth influence indicator in characterizing the influence of u on v as follows:

$$\alpha_j = f(C_j) \tag{12}$$

where f is a decreasing function. The function f can be defined as [20]:

$$\alpha_j = \left(1 - (C_j)^\lambda \right)^{1/\lambda} \tag{13}$$

where $\lambda > 0$.

After applying all these steps, we obtain the estimated value of the BBA reliability, α_j. To consider this reliability, we apply the discounting procedure described in Eq. (4). Then, we apply these steps for all defined BBAs on every link in the network.

4.3   Influence Estimation

After discounting all BBAs of each link in the network, we use them to estimate the influence that exerts one user on his neighbor. For this purpose, let us consider the link (u, v) and its discounted set of BBAs $\left(m_{(u,v)_1}^{\alpha_1}, m_{(u,v)_2}^{\alpha_2}, \ldots, m_{(u,v)_n}^{\alpha_n} \right)$. We define the global influence BBA that exerts u on v to be the BBA that fuses all discounted BBAs defined on (u, v). For this aim, we use the Dempster's rule of combination (see Eq. (5)) to combine all these BBAs as follows:

$$m_{(u,v)} = m_{(u,v)_1}^{\alpha_1} \oplus m_{(u,v)_2}^{\alpha_2} \oplus \ldots \oplus m_{(u,v)_n}^{\alpha_n} \tag{14}$$

The BBA distribution $m_{(u,v)}$ is the result of this combination.

Consequently, we define the influence that exerts u on v to be the amount of belief given to $\{I\}$ as:

$$Inf(u, v) = m_{(u,v)}(I) \tag{15}$$

The novelty of this evidential influence measure is that it considers several influence indicators in a social network and it takes into account the reliability of each defined indicator against the others. Our evidential influence measure can be considered as a generalization of the evidential influence measure introduced in the work of Jendoubi *et al.* [10].

To maximize the influence in the network, we need to define the amount of influence that exerts a set of nodes, S, on the hole network. It is the total influence given to S for influencing all users in the network. Then, we estimate the influence of S on a user v as follows [10]:

$$Inf(S, v) = \begin{cases} 1 & if\ v \in S \\ \sum_{u \in S} \sum_{x \in IN(v) \cup v} Inf(u, x) \, . \, Inf(x, v) & Otherwise \end{cases} \tag{16}$$

where $Inf(v, v) = 1$ and $IN(v)$ is the set of in-neighbors of v, *i.e.* if (u, v) is a link in the network then u is an in-neighbor of v. Next, we define the influence spread function that computes the amount of influence of S on the network as follows:

$$\sigma(S) = \sum_{v \in V} Inf(S, v) \tag{17}$$

To maximize the influence that exerts a set of users S on the network, we need to maximize $\sigma(S)$, *i.e.* $\operatorname*{argmax}_{S} \sigma(S)$. The influence maximization under the evidential model is demonstrated to be NP-Hard. Furthermore, the function, $\sigma(S)$, is monotone and submodular. Proof details can be found in [10]. Consequently, a greedy-based solution can perform a good approximation of the optimal influence users set S. In such cases, the cost effective lazy-forward algorithm (CELF) [17] is an adaptable maximization algorithm. Besides, it needs only two passes of the network nodes and it is about 700 times faster than the greedy algorithm. More details about CELF-based solution used in this paper can be found in [10].

After the definition of the reliability-based evidential influence measure and the influence spread function, we move to the experiments. Indeed, we made a set of experiments on real data to show the performance of our solution.

5   Results and Discussion

This section is dedicated to the experiments. In fact, we crawled Twitter data for the period between the 08-09-2014 and 03-11-2014. Table 1 presents some statistics of the dataset.

To characterize the influence users on Twitter, we choose the following three influence indicators: (1) the number of common neighbors between u and v, (2)

Table 1. Statistics of the data set [10]

Nbr of users	Nbr of tweets	Nbr of follows	Nbr of retweets	Nbr of mentions
36274	251329	71027	9789	20300

the number of times v mentions u in a tweet, (3) the number of times v retweets from u. Next, we apply the process described above in order to estimate the amount of influence that exerts each user u on his neighbor v in the network.

To evaluate the proposed reliability-based solution, we compare the proposed solution to the evidential model of Jendoubi *et al.* [10]. Furthermore, we choose four comparison criteria to compare the quality of the detected influence users by each experimented model. Those criteria are the following: (1) number of accumulated follow, (2) number of accumulated mention, (3) number of accumulated retweet, (4) number of accumulated tweet. Indeed, we assume that an influence user with a good quality is a highly followed user, mentioned and retweeted several times and active in terms of tweets.

In a first experiment, we compare the behavior of the proposed measure with fixed values of indicator reliability. Figure 1 presents the obtained results for two fixed values of $\alpha_j = \alpha$, which are $\alpha_j = \alpha = 0$ and $\alpha_j = \alpha = 0.2$. In Fig. 1 we have a comparison according to the four criteria, *i.e.* #Follow, #Mention, #Retweet and #Tweet shown in the y-axis of the sub-figures. We compare the experimented measure using the set of selected seed for each value of α_j. Besides, we fixed the size of the set S of selected influence users to 50 influencers, *i.e.* shown in the x-axis of each sub-figure.

According to Fig. 1, we notice that when $\alpha = 0$ (red scatter plots), the proposed reliability-based model does not detect good influencers according to the four comparison criteria. In fact, the red scatter plot ($\alpha = 0$) is very near to the x-axis in the case of the four criteria, which means that the detected influencer are neither followed, nor mentioned, nor retweeted. Besides, they are not very active in terms of tweets. However, we see a significant improvement especially when $\alpha = 0.2$ (blue scatter plots). Indeed, the detected influencers are highly followed as they have about 14k accumulated followers in total. Besides, the model detected some highly mentioned and retweeted influencers, especially starting from the 25th detected influencer. Finally, the influence users selected when $\alpha = 0.2$ are more active in terms of tweets than those selected when $\alpha = 0$.

This first experiment shows the importance of the reliability parameter, α, in detecting influencers with good quality. In fact, we see that when we consider that all indicators are totally reliable in characterizing the influence (the case when $\alpha = 0$), we notice that the proposed model detects influencers with very bad quality. However, when we reduce this reliability ($\alpha = 0.2$) we notice some quality improvement.

In a second experiment, we used the process described in Sect. 4.2 to estimate the reliability of each BBA in the network. Then, each BBA in our network is discounted using its own estimated reliability parameter. We note that the
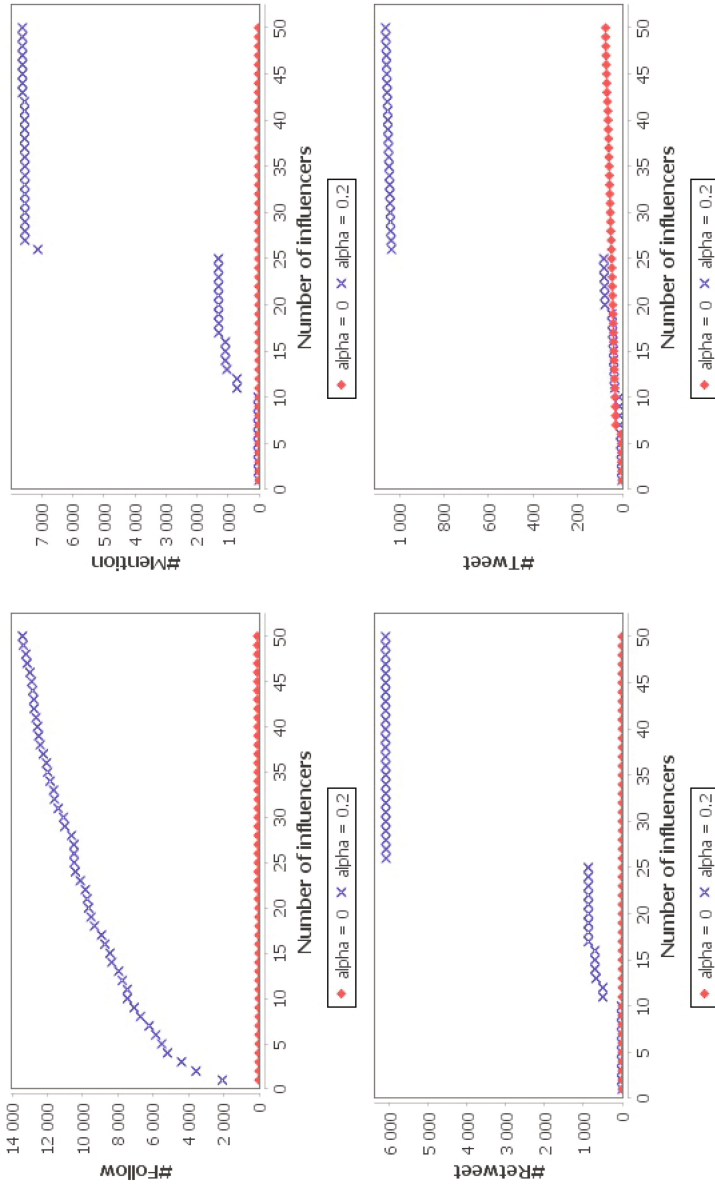
Fig. 1. Comparison of the reliability-based evidential model with three α fixed value (Color figure online)
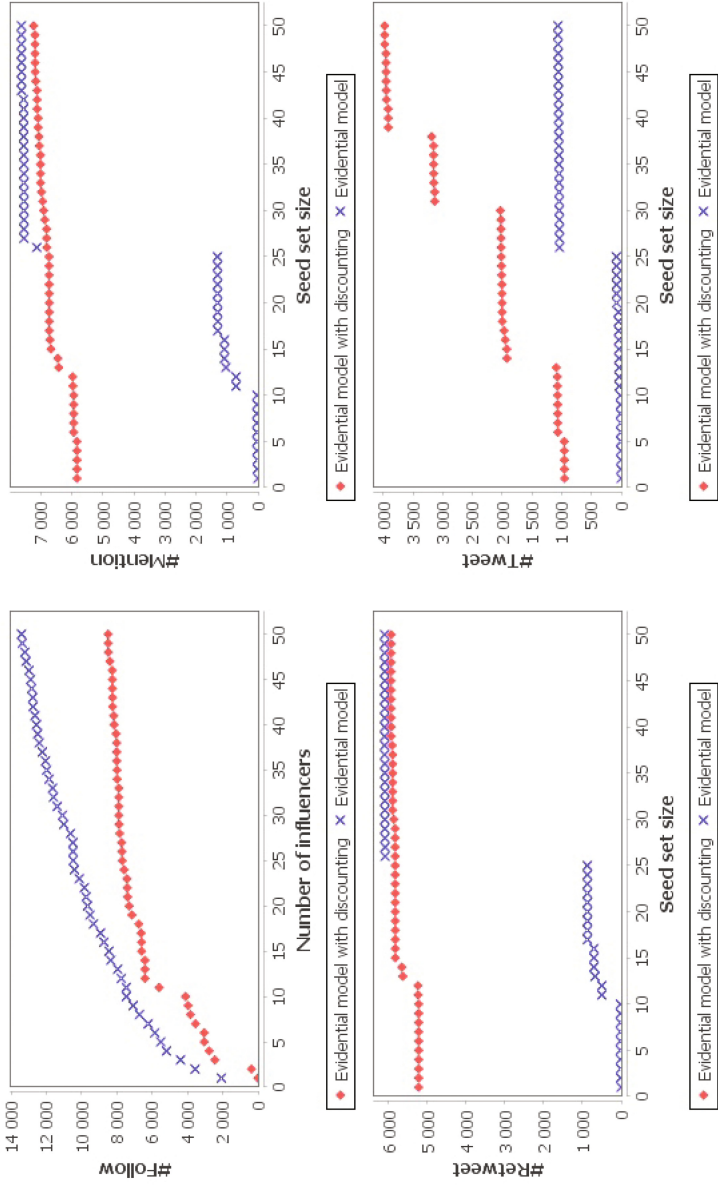
Fig. 2. Comparison between the proposed reliability-based solution and the evidential model of [10]

parameter λ in Eq. (13) was fixed to $\lambda = 5$. To fix this value, we made a set of experiments with different values of λ and the best results are given with $\lambda = 5$. Furthermore, we compare our reliability-based evidential model (also called evidential model with discounting) to the evidential model proposed by [10]. In fact, this last model is the nearest in its principle to the proposed solution in this paper. Besides, we fixed the size of the set S of selected influence users to 50 influencers. Figure 2 presents a comparison between the two experimented models in terms of #Follow, #Mention, #Retweet and #Tweet (shown in the y-axis of the sub-figures).

According to Fig. 2, we note that the two experimented models detect good influencers (shown in the x-axis). However, we see that the best compromise between the four criteria is given by the proposed reliability-based evidential model. In terms of accumulated #Follow, we notice that the most followed influencers are detected by the evidential model, also, our reliability-based model detected highly followed influencers. In terms of #Mention, we see that the evidential model starts detecting some mentioned influencers after detecting about 10 users that are not mentioned. However, the proposed reliability-based model starts detecting highly mentioned users from the first detected influencer. Furthermore, we see a similar behavior in the sub-figure showing the accumulated #Retweet. Indeed, the proposed solution detects highly retweeted influencers from the first user. In the last sub-figure that presents the comparison according to the accumulated #Tweet, the best results are those of the proposed reliability-based model.

This second experiment shows the effectiveness of the proposed reliability-based influence measure against the evidential influence measure of [10]. In fact, the best influence maximization model is always the model that detects the best influencers at first. Indeed, in an influence maximization problem we need generally to minimize the number of selected influencers in order to minimize the cost. For example, this is important in a viral marketing campaign as it helps the marketer to minimize the cost of his campaign and to maximize his benefits. Furthermore, our influence maximization solution gives the best compromise between the four criteria, *i.e.* #Follow, #Mention, #Retweet and #Tweet.

From these experiments we can conclude that the reliability parameter is important if we want to measure the influence in a social network through the consideration of several influence indicators. Indeed, we may have some influence indicators that are more reliable in characterizing the influence of the others. Furthermore, the proposed solution is efficient in detecting influencers with a good compromise between all chosen influence indicators.

6   Conclusion

To conclude, this paper introduces a new reliability-based influence measure. The proposed measure fuses many influence indicators in the social network. Furthermore, it can be adapted for several social networks. Another important contribution of the paper is that we consider the reliability of each chosen

influence indicator to characterize the influence. Indeed, we propose to apply a distance-based operator that estimates the reliability of each indicator against to the others and considers the assumption that "*the farthest from the others the indicator is, the less reliable it is*". Besides, we use the proposed reliability-based measure with an existing influence maximization model. Finally, we present two experiments that show the importance of the reliability parameter and the effectiveness of the reliability-based influence maximization model against the evidential model of Jendoubi *et al.* [10]. Indeed, we obtained a good compromise in the quality of detected influencers and we had good results according to the four influence criteria, *i.e.* #Follow, #Mention, #Retweet and #Tweet.

For future works, we will search to test our influence maximization solution on other social networks. Then, we will collect more data from Facebook and Google Plus and we will prove the performance of the proposed reliability-based influence maximization model. A second important perspective is about the influence maximization within communities. In fact, social networks are generally characterized by a community structure. Then, the main idea is to take profit from this characteristic and to search to select a minimum number of influence users and to minimize the time spent to detect them.

References

1. Aslay, C., Barbieri, N., Bonchi, F., Baeza-Yates, R.: Online topic-aware influence maximization queries. In: Proceedings of the 17th International Conference on Extending Database Technology (EDBT), pp. 24–28, March 2014
2. Bozorgi, A., Haghighi, H., Zahedi, M.S., Rezvani, M.: INCIM: A community-based algorithm for influence maximization problem under the linear threshold model. Inf. Process. Manage. **000**, 1–12 (2016)
3. Chen, D., Lü, L., Shang, M.S., Zhang, Y.C., Zhou, T.: Identifying influential nodes in complex networks. Physica A: Stat. Mech. Appl. **391**(4), 1777–1787 (2012)
4. Dempster, A.P.: Upper and Lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**, 325–339 (1967)
5. Denœux, T., Sriboonchitta, S., Kanjanatarakul, O.: Evidential clustering of large dissimilarity data. Knowl.-Based Syst. **106**, 179–195 (2016)
6. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of KDD 2001, pp. 57–66 (2001)
7. Gao, C., Wei, D., Hu, Y., Mahadevan, S., Deng, Y.: A modified evidential methodology of identifying influential nodes in weighted networks. Physica A **392**(21), 5490–5500 (2013)
8. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: A data-based approach to social influence maximization. In: Proceedings of VLDB Endowment, pp. 73–84, August 2012
9. Jendoubi, S., Martin, A., Liétard, L., Ben Hadj, H., Ben Yaghlane, B.: Maximizing positive opinion influence using an evidential approach. In: Proceedings of the 12th International FLINS Conference, August 2016
10. Jendoubi, S., Martin, A., Liétard, L., Hadj, H.B., Yaghlane, B.B.: Two evidential data based models for influence maximization in twitter. Knowl.-Based Syst. **121**, 58–70 (2017)

11. Jendoubi, S., Martin, A., Liétard, L., Ben Yaghlane, B.: Classification of message spreading in a heterogeneous social network. In: Laurent, A., Strauss, O., Bouchon-Meunier, B., Yager, R.R. (eds.) IPMU 2014. CCIS, vol. 443, pp. 66–75. Springer, Cham (2014). doi:10.1007/978-3-319-08855-6_8

12. Jendoubi, S., Martin, A., Liétard, L., Ben Yaghlane, B., Ben Hadji, H.: Dynamic time warping distance for message propagation classification in twitter. In: Destercke, S., Denoeux, T. (eds.) ECSQARU 2015. LNCS (LNAI), vol. 9161, pp. 419–428. Springer, Cham (2015). doi:10.1007/978-3-319-20807-7_38

13. Jousselme, A.L., Grenier, D., Bossé, E.: A new distance between two bodies of evidence. Inf. Fusion **2**, 91–101 (2001)

14. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of KDD 2003, pp. 137–146, August 2003

15. Kempe, D., Kleinberg, J., Tardos, É.: Influential nodes in a diffusion model for social networks. In: Caires, L., Italiano, G.F., Monteiro, L., Palamidessi, C., Yung, M. (eds.) ICALP 2005. LNCS, vol. 3580, pp. 1127–1138. Springer, Heidelberg (2005). doi:10.1007/11523468_91

16. Kimura, M., Saito, K.: Tractable models for information diffusion in social networks. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 259–271. Springer, Heidelberg (2006). doi:10.1007/11871637_27

17. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of KDD 2007, pp. 420–429, August 2007

18. Liu, Z., Pan, Q., Dezert, J., Martin, A.: Adaptive imputation of missing values for incomplete pattern classification. Pattern Recogn. **52**, 85–95 (2016)

19. Liu, Z., Pan, Q., Dezert, J., Mercier, G.: Credal c-means clustering method based on belief functions. Knowl.-Based Syst. **74**, 119–132 (2015)

20. Martin, A., Jousselme, A.L., Osswald, C.: Conflict measure for the discounting operation on belief functions. In: International Conference on Information Fusion, Cologne, Germany, pp. 1003–1010, juillet 2008

21. Mohamadi-Baghmolaei, R., Mozafari, N., Hamzeh, A.: Trust based latency aware influence maximization in social networks. Eng. Appl. Artif. Intell. **41**, 195–206 (2015)

22. Mumu, T.S., Ezeife, C.I.: Discovering community preference influence network by social network opinion posts mining. In: Bellatreche, L., Mohania, M.K. (eds.) DaWaK 2014. LNCS, vol. 8646, pp. 136–145. Springer, Cham (2014). doi:10.1007/978-3-319-10160-6_13

23. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)

24. Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. **66**, 191–234 (1994)

25. Wei, D., Deng, X., Zhang, X., Deng, Y., Mahadeven, S.: Identifying influential nodes in weighted networks based on evidence theory. Physica A **392**(10), 2564–2575 (2013)

26. Zhou, K., Martin, A., Pan, Q., Liu, Z.: Median evidential c-means algorithm and its application to community detection. Knowl.-Based Syst. **74**, 69–88 (2015)