Evidential Link Prediction Based on Group Information

Sabrine Mallek[1], Imen Boukhris[1], Zied Elouedi[1], and Eric Lefevre[2]

[1] LARODEC, Institut Supérieur de Gestion de Tunis, Université de Tunis, Tunisia
sabrinemallek@yahoo.fr, imen.boukhris@hotmail.com, zied.elouedi@gmx.fr

[2] Univ. Lille Nord de France, UArtois EA 3926 LGI2A, France
eric.lefevre@univ-artois.fr

Abstract. Link prediction has become a common way to infer new associations among actors in social networks. Most existing methods focus on the local and global information neglecting the implication of the actors in social groups. Further, the prediction process is characterized by a high complexity and uncertainty. In order to address these problems, we firstly introduce a new evidential weighted version of the social networks graph-based model that encapsulates the uncertainty at the edges level using the belief function framework. Secondly, we use this graph-based model to provide a novel approach for link prediction that takes into consideration both groups information and uncertainty in social networks. The performance of the method is experimented on a real world social network with group information and shows interesting results.

Keywords: social network analysis, link prediction, uncertain social network, group information, belief function theory

1 Introduction

Social networks are usually conceptualized as a graph representation that provides a mapping of the ties relating the social structures. They are very dynamic and alter quickly over specific time intervals. New connections are established continuously between the network nodes. One of the most popular researches in social network analysis that studies social networks evolving is link prediction. It addresses the problem of predicting the existence of new or missing connections in social networks.

Yet, existing methods for link prediction are devoted to social networks under a certain framework. In fact, most methods assume the links to have binary values, either 1 (exist) or 0 ($\neg$ exist). Still, the structure of social network critically depends on the accurate structure of the data. That is, sparse distortions affect considerably the analysis results. As pointed out in [2, 7], social networks data are frequently noisy and missing, they are also prone to errors of observation (e.g., missing information about the nodes and/or edges from the data). Hence, one would have to deal with two possible problems: take all the nodes and edges into account risking the possibility of considering erroneously false ones into the network or remove all the uncertain nodes and/or edges risking the issue of missing nodes and edges [7]. Furthermore, the unreliability of the tools used for the construction of the social network can lead to distortions [6]. On that point, we propose to incorporate uncertainty into the graph structure of social networks.

Generally, most existing studies consider weighted networks with integer values. Yet, one way for representing an uncertain network is to weight the edges with values in $[0, 1]$ to depict the degrees of uncertainty regarding the links' existence [6]. In fact, several real world social networks are characterized by shifting degrees of uncertainty, more particularly the large scale ones [2]. For this reason, we embrace the theory of belief functions [4, 11] as a general framework for reasoning under uncertainty. We use its assets for the handling of imprecision in data and the modeling of partial and total ignorance to quantify the degrees of uncertainty into the edges of the social network.

Furthermore, we develop a new approach for inferring new links in a social network characterized by uncertain edges based on group information and structural neighborhood measures of nodes. In fact, most of the existing methods are based on the local node neighborhood and global paths measures. Yet, these latter do not take into account a very important aspect in social networks which is its community structure. The participation of actors in social groups can bring important information concerning their characteristics and thus, may enhance the prediction task. To this end, we propose a method that performs exclusively with the belief function tools. The degrees of uncertainty of the similar nodes from the common shared groups are considered. They are revised, transferred and combined as independent sources of information and are afterward employed to get an outlook on the existence of a new link.

The rest of the paper is organized as follows: Section 2 gives a brief survey of the link prediction problem and the existing approaches. Section 3 provides some basic knowledge of the belief function theory notations and definitions. In Section 4, we present our evidential link-based graph model for a social network under an uncertain framework. Section 5 reveals the proposed approach for link prediction under the belief function framework. Section 6 illustrates the proposed method and Section 7 gives the experimental results. Finally, Section 8 concludes the paper.

2 LINK PREDICTION

Due to its great applicability, link prediction constitutes a rich research area and has attracted many researchers from various fields. Namely, in social networks, link prediction is a basic task in social relationships formation. It can be applied to infer the new relations to be formed in the future, expose links which already exist but are not apparent, or even assist users to make new connections.

In most common formulation, the link prediction problem can be defined as follows [8]: given a current state of the social network graph in time t, the aim is to accurately infer the potential edges to be added to the unlinked pairs of nodes given a snapshot of the social network during the time interval $[t, t']$. It may also be considered as the problem of deriving the missing links of the network. In fact, one may construct a social network from a given observable data and try to derive the invisible links that are likely to exist. Most of the state-of-the-art link prediction methods have focused on two groups of network information that can be categorized into local (node neighborhood) and global (path) information. Local information-based approaches use the local similarities of the nodes characteristics in the network. These latter may be the essential attributes, i.e., gender, age, interests, or structural indices which are based solely on the network structure, i.e., common neighbors that two nodes share. Yet, nodes' attributes are generally not available or hidden, thus the majority of local approaches use metrics based on the structural similarities. The global approaches use the proximity of the nodes in the network, they employ metrics based on the ensemble of paths to determine the closest nodes in the network. The intuition is that the more close two nodes are in the network, the more they tend to be linked or to influence each other in the future.

The main advantage of these measures is that they are generic, they can be applied to networks from several fields. While the global methods perform better than the local ones, some path based metrics are time consuming as they inquire for the topological information of the whole network which is in many cases not available. Besides, a relevant aspect characterizing social networks is not considered which is the participation of the actors in social groups (clusters, communities). In fact, in several social networks, users are involved in many social groups at the same time. Thus, hybrid methods that use local and cluster information have been proposed [16, 18, 19]. That is, our proposed method is based on local and group similarity measures. Thus, we recall in this section some state-of-the-art structural measures based on local and group information.

2.1 Local information based measures

Some of the base measures from the literature are "Common Neighbors"[10], "Jaccard's Coefficient"[5] and "Adamic/Adar"[1]. Let $\tau(v_i)$ denote the set of neighbors of the node v_i in the social network. The common neighbors measure denoted by $CN(v_i, v_j)$ characterizes the number of common neighbors between two nodes v_i and v_j. It is defined as:

$$CN(v_i, v_j) = |\tau(v_i) \cap \tau(v_j)| \tag{1}$$

On the other hand, the Jaccard's Coefficient considers all the the neighbors of the pair (v_i, v_j). It is defined as follows:

$$JC(v_i, v_j) = \frac{|\tau(v_i) \cap \tau(v_j)|}{|\tau(v_i) \cup \tau(v_j)|} \tag{2}$$

The Adamic/Adar measure denoted by $AA(v_i, v_j)$ weights the contribution of each common neighbor v_k by the inverse of the logarithm of its degree, it is defined as:

$$AA(v_i, v_j) = \sum_{v_k \in (\tau(v_i) \cap \tau(v_j))} \frac{1}{log|\tau(v_k)|} \tag{3}$$

2.2 Group information based measures

Structural similarity measures based on group information use both local structure of the nodes and group information, they include the Common Neighbors of Groups (CNG) and Common Neighbors Within and Outside of Common Groups (WOCG) [18, 19]. Let $\Lambda^G_{v_i v_j}$ denote the set of common neighbors of the pair (v_i, v_j) belonging to the group G. The CNG depicts the size of the set of common neighbors of (v_i, v_j) that belong to at least one group G to which v_i or v_j is part of. It is defined as:

$$S^{CNG}_{v_i v_j} = |\Lambda^G_{v_i v_j}| \tag{4}$$

Let $\Lambda_{v_i v_j} = \Lambda^{WCG}_{v_i v_j} \cap \Lambda^{OCG}_{v_i v_j}$ be the set of common neighbors of (v_i, v_j) such that $\Lambda^{WCG}_{v_i v_j}$ is the set of common neighbors within common groups (WCG) and $\Lambda^{OCG}_{v_i v_j}$ is the set of common neighbors outside the common groups (OCG). The WOCG measure is defined as:

$$s^{WOCG}_{v_i v_j} = \frac{|\Lambda^{WCG}_{v_i v_j}|}{|\Lambda^{OCG}_{v_i v_j}|} \tag{5}$$

3 BELIEF FUNCTION FRAMEWORK

The belief function theory [4, 11] is a general framework for the representation and management of uncertain evidence. Let Θ be the frame of discernment, an exhaustive and finite set of mutually exclusive events associated to a given problem. 2^{Θ} is the power set of Θ, it includes all the possible subsets and formed unions of events, and the empty set which matches the conflict. A basic belief assignment (*bba*), denoted by m, is the mass assigned to an event given a piece of evidence. It is defined as:

$$\begin{aligned} &m : 2^{\Theta} \rightarrow [0,1] \\ &\sum_{A \subseteq \Theta} m(A) = 1 \end{aligned} \tag{6}$$

A bba with at most one focal element A different from Θ is called a simple support function (ssf). It is defined as [13]:

$$\begin{cases} m(A) & = 1 - \omega, \forall A \subset \Theta \\ m(\Theta) & = \omega, \omega \in [0,1] \end{cases} \tag{7}$$

Beliefs can be fused using combination rules. In particular, the conjunctive rule of combination permits to combine evidence given by two reliable and distinct sources of information characterised by two bba's m_1 and m_2. It is denoted by $\circledcap$ and is defined by [14]:

$$m_1 \circledcap m_2(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B) \cdot m_2(C) \tag{8}$$

While combining evidence on Θ, it is important to take into consideration the reliability of the evidence. For that, a so-called discounting mechanism can be performed [11]:

$$\begin{aligned} &{}^{\alpha}m(A) = (1-\alpha) \cdot m(A), \text{ for } A \subset \Theta \\ &{}^{\alpha}m(\Theta) = \alpha + (1-\alpha) \cdot m(\Theta) \end{aligned} \tag{9}$$

Where $\alpha \in [0,1]$ represents the discount rate (coefficient).

Let Θ and Ω be two disjoint frames of discernment. In order to establish the relation between them, one may use a multi-valued mapping [4]. In fact, a multi-valued mapping function denoted by τ, permits to bring together to different frames of discernment the subsets $B \subseteq \Omega$ that can possibly match under τ to a subset $A \subseteq \Theta$:

$$m_{\tau}(A) = \sum_{\tau(B) = A} m(B) \tag{10}$$

The Transferable Belief Model (TBM), proposed by Smets [15], is one of the well-known interpretations of the belief function theory. In the TBM, decision making is performed at the pignistic level where beliefs are transformed into probabilities using pignistic measures denoted by $BetP$ [12]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \text{for all } A \in \Theta \tag{11}$$

4 EVIDENTIAL LINK-BASED SOCIAL NETWORK

Most social networks graphs include nodes and edges that are assumed to be certain 1 (exist) 0 ($\neg$ exist). The authors in [6] highlighted the importance of incorporating uncertainty when dealing with social networks and proposed to weight the strengths of the links by probabilities. In [3], the authors proposed a belief social network where the nodes, edges and messages are weighted by bba's. The aim is to detect the nature of a message that flows through the network. Yet, the purpose of this work is to treat the uncertainty upon the links. Thus, we introduce our evidential link-based social network graph model where uncertainty is encrypted using the belief function theory. Each edge v_iv_j has assigned a bba defined on $\Theta^{v_iv_j} = \{E_{v_iv_j}, \neg E_{v_iv_j}\}$ denoted by $m^{v_iv_j}$, i.e., $E_{v_iv_j}$ means that v_iv_j exists and $\neg E_{v_iv_j}$ means it is absent. That is, an evidential link-based social network graph is defined as $G(V, E)$ where: $V = \{v_1, \ldots, v_{|V|}\}$ is the set of nodes, and E is the set of edges: An edge $v_iv_j \in E$ has assigned a bba $m^{v_iv_j}$ that depicts the degree of uncertainty regarding its existence.

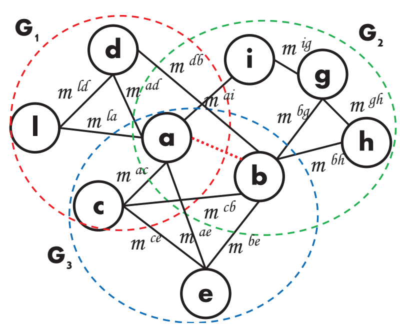

Fig. 1: A social network graph with bba's weighted edges and group belonging nodes

Fig. 1 illustrates an example of such a bba's edge weighted graph structure. In fact, instead of weighting the links by either 1 or 0 to demonstrate whether or not a link is existent, we ascribe a bba with values in $[0, 1]$ to quantify the degree of uncertainty about the link existence. Note that a link v_iv_j is represented if the pignistic probability $BetP^{v_iv_j}(E_{v_iv_j}) > 0.5$ which means that its likelihood to exist is greater than 50%.

5 EVIDENTIAL LINK PREDICTION BASED ON GROUP INFORMATION

Our proposed method for link prediction uses node neighborhood and group information given a snapshot of a graph. In fact, an earlier phase for the partitioning of the network into groups needs to be applied, most works apply algorithms for communities detection with low computational cost. Yet, this makes the prediction quality dependent to the community detection algorithm performance. The authors in [18, 19] proposed to eliminate this dependency by using the natural information of groups, i.e., the information from groups of interests to which users participate to. Thus, each edge $v_i v_j$ has assigned a feature vector that corresponds to the structural similarity measures based on local and group information as explained in Section 2. CN (Equation 1), JC (Equation 2), AA (Equation 3), CNG (Equation 4) and WOCG (Equation 5) are employed as similarity measures since they are simple and have proved their efficiency in several social networks domains [10, 18, 19]. The feature vector is used to compute the similarity between the link to be predicted and its neighbors belonging to the shared groups. The intuition is that in many real world social networks, users with similar experiences or interests are more likely to share a relationship than those that do not share common characteristics. The most similar link is subsequently considered as a source of information. Our formulation of the link prediction problem is as follows:

Given a current state of the graph $G(V, E)$ at time t, predict the existence of an edge $v_i v_j$ between two the unlinked nodes (v_i, v_j) at $t + 1$ by considering the relationships shared in their common groups. To this end, we propose a method for the prediction of the existence of a link between (v_i, v_j) based on the steps outlined below.

5.1 Distance computation

At a first place, the Euclidean distance $D(v_i v_j, v_k v_l)$ between the link $v_i v_j$ and each link $v_k v_l$ included in the common shared clusters is computed. Structural similarity measures based on local and group information are used as features. That is, $D(v_i v_j, v_k v_l)$ is used to evaluate the similarity between $v_i v_j$ and the neighboring links. The most similar link to $v_i v_j$ with the smallest distance is considered in the prediction task. Note that the distance metric is divided by its maximum value in order to get values in $[0, 1]$. It is computed as follows:

$$D(v_i v_j, v_k v_l) = \frac{\sqrt{\sum_{s=1}^{n} (x^s_{v_i v_j} - y^s_{v_k v_l})^2}}{D_{max}} \quad (12)$$

Where s is the index of a structural similarity metric, $x_{v_i v_j}$ and $y_{v_k v_l}$ are respectively its values for $v_i v_j$ and $v_k v_l$ and D_{max} is the maximum value of the Euclidean distance.

5.2 Reliability computation

In order to quantify the degree of reliability of the most similar link, a discounting operation (Equation 9) is applied using the value given by the distance measure as a discount coefficient denoted by $\alpha = D(v_i v_j, v_k v_l)$. In fact, the more similar the two links are, the more reliable the similar link is, i.e., if the two links are totally similar $D(v_i v_j, v_k v_l) = 0$ then $v_k v_l$ is considered as a totally reliable source of evidence i.e., $\alpha = 0$. Thus, $m^{v_k v_l}$ is discounted as follows:

$$\begin{cases} {}^{\alpha}m^{v_k v_l}(\{E_{v_k v_l}\}) = (1-\alpha) \cdot m^{v_k v_l}(\{E_{v_k v_l}\}) \\ {}^{\alpha}m^{v_k v_l}(\{\neg E_{v_k v_l}\}) = (1-\alpha) \cdot m^{v_k v_l}(\{\neg E_{v_k v_l}\}) \\ {}^{\alpha}m^{v_k v_l}(\Theta^{v_k v_l}) = \alpha + (1-\alpha) \cdot m^{v_k v_l}(\Theta^{v_k v_l}) \end{cases} \tag{13}$$

Note that when there is more than one most similar link, i.e., two links with smallest equal distances, the link with the highest mass on the event "exist" is chosen since the degree of certainty of its existence would be higher.

5.3 Information transfer and fusion

To transfer the discounted bba of the most similar link $v_k v_l$ to the frame $\Theta^{v_i v_j}$, a multi-valued operation denoted by τ: $\Theta^{v_k v_l} \rightarrow 2^{\Theta^{v_i v_j}}$ is applied to bring together the elements as follows:

- The discounted mass ${}^{\alpha}m^{v_k v_l}(\{E_{v_k v_l}\})$ is transferred to $m^{v_i v_j}_{v_k v_l}(\{E_{v_i v_j}\})$;
- The discounted mass ${}^{\alpha}m^{v_k v_l}(\{\neg E_{v_k v_l}\})$ is transferred to $m^{v_i v_j}_{v_k v_l}(\{\neg E_{v_i v_j}\})$;
- The discounted mass ${}^{\alpha}m^{\Theta^{v_k v_l}}(\Theta^{v_k v_l})$ is transferred to $m^{v_i v_j}_{v_k v_l}(\Theta^{v_i v_j})$.

Where $\alpha = D(v_i v_j, v_k v_l)$ and $m^{v_i v_j}_{v_k v_l}$ denotes the bba of $v_i v_j$ on $\Theta^{v_i v_j}$ given the most similar link, here $v_k v_l$.

Upon transferring ${}^{\alpha}m^{v_k v_l}$ to $2^{\Theta^{v_i v_j}}$, the bba of $v_i v_j$ is updated given the new evidence obtained from the most similar link. To accomplish this, the initial bba $m^{v_i v_j}$ and $m^{v_i v_j}_{v_k v_l}$ are combined using the conjunctive rule of combination (Equation 8). This step is essential, as it permits to fuse the information provided by the most similar link and treat it as an independent source of evidence.

5.4 Decision making

At last, the pignistic probability $BetP^{v_i v_j}(E_{v_i v_j})$ is computed (Equation 11) to make a decision about the existence of the link $v_i v_j$ on the graph. As a matter of fact, when $BetP^{v_i v_j}(E_{v_i v_j}) > 0.5$ it means that the likelihood that a link exist between v_i and v_j at $t+1$ has probability $> 50\%$, it would be absent otherwise.

6 ILLUSTRATION

To illustrate our link prediction approach, we try to predict the existence of a new link between the pair of nodes (a, b) presented in Fig. 1. To do so, the edge ab is assumed to be present in the graph in order to be able to compare its structural attributes and those of the other links belonging to the shared groups G_2 and G_3. Thus, the neighboring links in the common shared groups are: $ai, ac, ae, bg, bh, hg, be, bc, ce, gi$. We apply the steps presented in Section 5.

Step 1: At first, we compute the Euclidean distance between ab and each neighboring link included in the common groups G_1 and G_2 shared between a and b using Equation 12. The results are reported in Table 1.

Table 1: Distance between ab and the links in the common shared groups of Fig. 1

Distance	ac	ae	bc	be	bg	bh	ce	hg	ai	gi
ab	0.282	0.282	0.283	0.283	0.551	0.599	0.357	0.638	1	1

Hence, the most similar links to ab are ac and ae. That is, we have to use one of them to update m^{ab}. Suppose we have *bba*'s allocated as follows:

$$\begin{cases} m^{ab}(\{E_{ab}\}) = 0.35 \\ m^{ab}(\{\neg E_{ab}\}) = 0.42 \\ m^{ab}(\Theta^{ab}) = 0.23 \end{cases}, \begin{cases} m^{ac}(\{E_{ac}\}) = 0.65 \\ m^{ac}(\{\neg E_{ac}\}) = 0.2 \\ m^{ac}(\Theta^{ac}) = 0.15 \end{cases} \text{and} \begin{cases} m^{ae}(\{E_{ae}\}) = 0.55 \\ m^{ae}(\{\neg E_{ae}\}) = 0.25 \\ m^{ae}(\Theta^{ae}) = 0.2 \end{cases}$$

Thus, ac is chosen as a source of information since $m^{ac}(\{E_{ac}\}) > m^{ae}(\{E_{ae}\})$.

Step 2: The next step is to discount the *bba* m^{ac} using $D(ab, ac)$ to quantify its degree of reliability. We denote $\alpha = D(ab, ac)$ the discount rate. Thus, ${}^{\alpha}m^{ac}$ after the discounting operation is: ${}^{\alpha}m^{ac}(\{E_{ac}\}) = (1-0.282) \cdot 0.65 = 0.4667$, ${}^{\alpha}m^{ac}(\{\neg E_{ac}\}) = (1 - 0.282) \cdot 0.2 = 0.1436$ and ${}^{\alpha}m^{ac}(\Theta^{ac}) = 0.282 + (1 - 0.282) \cdot 0.15 = 0.3897$.

Step 3: When the discounted mass of the most similar link is transferred using the τ function (Equation 10), the mass of ab given ac is: $m^{ab}_{ac}(\{E_{ab}\}) = 0.4667$, $m^{ab}_{ac}(\{\neg E_{ab}\}) = 0.1436$ and $m^{ab}_{ac}(\Theta^{ab}) = 0.3897$. To update the *bba* of the link ab, m^{ab} and m^{ab}_{ac} are fused by applying the conjunctive rule of combination (Equation 8). Thus, we get: $m^{ab} \circledcap m^{ab}_{ac}(\{E_{ab}\}) = 0.407$, $m^{ab} \circledcap m^{ab}_{ac}(\{\neg E_{ab}\}) = 0.257$, $m^{ab} \circledcap m^{ab}_{ac}(\Theta^{ab}) = 0.09$ and $m^{ab} \circledcap m^{ab}_{ac}(\emptyset) = 0.246$.

Step 4: Finally, the pignistic probability $BetP^{ab}$ (Equation 11) is computed to make a decision on the link existence between the nodes a and b. Thus, $BetP(E_{ab}) = 0.575$ and $BetP(\neg E_{ab}) = 0.425$. Hence, there is 57% chance that a link may exist between a and b. That is, a link would be schematized in the graph representation.

7 EXPERIMENTS

In order to test our approach for link prediction, it is necessary to consider an uncertain social network. Yet, uncertain social network data are not available. Thus, we preprocessed a real world social network of 4K nodes and 88K edges of Facebook friendships obtained from [9] in order to transform it into an uncertain social network.

7.1 Network pre-processing

To transform the social network into a belief-link based social network, we follow two major steps: (1) we generate three snapshots of the network from the data (2) then we simulate mass functions on the basis of the three first graphs to get a belief link-based version of the social network.

Graphs generation At first, we create three graphs from the data by removing randomly a portion of the edges. That is, we get three graphs that we call $G(t-2)$, $G(t-1)$ and $G(t)$. Indeed, this technique is widely used in the link prediction literature. In several works, a number of edges is pruned from the graph so that they will be considered in the prediction process [17, 20].

Mass functions simulation In order to generate the belief link-based version of the social network, we weight each link v_iv_j by a simulated *bba* regarding its links existence on the basis of $G(t-2)$, $G(t-1)$ and $G(t)$ as follows:

- If v_iv_j exists in the three graphs $G(t-2)$, $G(t-1)$ and $G(t)$ then a ssf $m^{v_iv_j}$ is assigned such that $m^{v_iv_j}(\{E_{v_iv_j}\}) \in [2/3, 1]$;
- If v_iv_j exists in $G(t-2)$ and $G(t)$ or $G(t-1)$ and $G(t)$ then a mass $m^{v_iv_j}$ is generated such that $m^{v_iv_j}(\{E_{v_iv_j}\}) \in [1/3, 2/3[$, $m^{v_iv_j}(\{\neg E_{v_iv_j}\}) \in]0, 1/3]$;
- If v_iv_j exists only in $G(t)$ then a mass function $m^{v_iv_j}$ is assigned such that $m^{v_iv_j}(\{E_{v_iv_j}\}) \in]0, 1/3]$, $m^{v_iv_j}(\{\neg E_{v_iv_j}\}) \in [1/3, 2/3]$;
- If v_iv_j does not exist in $G(t)$ and exists in $G(t-2)$ and $G(t-1)$ then a ssf $m^{v_iv_j}$ is assigned such that $m^{v_iv_j}(\{\neg E_{v_iv_j}\}) \in]1/3, 2/3]$;
- If v_iv_j exists only in $G(t-2)$ or in $G(t-1)$ then a ssf $m^{v_iv_j}$ is assigned such that $m^{v_iv_j}(\{\neg E_{v_iv_j}\}) \in]0, 1/3]$.

7.2 Results

To test our proposed link prediction method, we produce three different belief link-based versions of the social network that we call G_1,G_2 and G_3. We evaluate the accuracy of our link prediction algorithm using the precision measure. It expresses the number of correctly predicted existent links n_c with respect to the set of analyzed links n. It is defined as follows:

$$precision = \frac{n_c}{n} \quad (14)$$

Table 2: The prediction results measured by the precision

	G_1	G_2	G_3
Precision	0.54	0.56	0.57

Table 2 gives the obtained precision values for the three experiments. As illustrated, the prediction quality measured by the precision gives values higher than 50% reaching a maximum performance of 57% for G_3. Hence, validity and performance of the new approach is empirically confirmed.

8 CONCLUSION

In this paper, we have provided an uncertain graph-based model for social networks whose edges are valued with mass functions given by the belief function theory. Furthermore, we have proposed a novel link prediction approach that takes into consideration both uncertainties in data and group information in social networks. Our method is exclusively based on the belief function framework tools, evidence from the neighbors of the common groups is revised, transferred and combined to successfully predict new connections. As part of future work, extension to the case of both uncertain nodes and edges would be considered. Also, comparison with existing methods is left open for future work.

References

[1] Adamic, L.A., Adar, E.: Friends and neighbors on the web. Social Networks 25(3), 211–230 (2003)

[2] Adar, E., Ré, C.: Managing uncertainty in social networks. Data Engineering Bulletin 30(2), 23–31 (2007)

[3] Ben Dhaou, S., Kharoune, M., Martin, A., Ben Yaghlane, B.: Belief approach for social networks. In: Belief Functions: Theory and Applications. vol. 8764, pp. 115–123 (2014)

[4] Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)

[5] Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37, 547–579 (1901)

[6] Johansson, F., Svenson, P.: Constructing and analyzing uncertain social networks from unstructured textual data. In: Mining Social Networks and Security Informatics. pp. 41–61. Lecture Notes in Social Networks, Springer (2014)

[7] Kossinets, G.: Effects of missing data in social networks. Social Networks 28, 247–268 (2003)

[8] Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. 58(7), 1019–1031 (2007)

[9] McAuley, J.J., Leskovec, J.: Learning to discover social circles in ego networks. In: NIPS. pp. 548–556 (2012)

[10] Newman, M.E.J.: Clustering and preferential attachment in growing networks. Phys. Rev. E (2001)

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

[12] Smets, P.: The transferable belief model for quantified belief representation. In: Handbook of Defeasible Reasoning and Uncertainty Management Systems. vol. 1, pp. 267–301 (1988)

[13] Smets, P.: The canonical decomposition of a weighted belief. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95. vol. 14, pp. 1896–1901 (1995)

[14] Smets, P.: Application of the transferable belief model to diagnostic problems. International Journal of Intelligent Systems 13(2-3), 127–157 (1998)

[15] Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. 66(2), 191–234 (1994)

[16] Soundarajan, S., Hopcroft, J.: Using community information to improve the precision of link prediction methods. In: Proceedings of the 21st International Conference Companion on World Wide Web. pp. 607–608. ACM (2012)

[17] Valverde-Rebaza, J., de Andrade Lopes, A.: Exploiting behaviors of communities of twitter users for link prediction. Social Network Analysis and Mining 3(4), 1063–1074 (2013)

[18] Valverde-Rebaza, J.C., Lopes, A.A.: Link prediction in complex networks based on cluster information. In: SBIA 2012 - 21th Brazilian Symposium on Artificial Intelligence. vol. 7589, pp. 92–101. Springer (2012)

[19] Valverde-Rebaza, J.C., Lopes, A.A.: Link prediction in online social networks using group information. In: ICCSA 2014. vol. 8584, pp. 31–45. Springer (2014)

[20] Zhang, Q.M., Lü, L., Wang, W.Q., Zhu, Y.X., Zhou, T.: Potential Theory for Directed Networks. PLoS ONE 8(2), e55437 (2013)