Evidential Link Prediction in Social Networks based on Structural and Social Information

Sabrine Mallek[a,b,*], Imen Boukhris[a], Zied Elouedi[a], Eric Lefèvre[b]

[a]LARODEC, Institut Supérieur de Gestion de Tunis, Université de Tunis, 41 rue de la liberté, 2000 Le Bardo, Tunisia
[b]Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), F-62400 Béthune, France

Abstract

Social networks are large systems that depict linkage between millions of social entities. The study of their patterning and evolving is one of the major research areas in social network analysis and network mining. It includes the prediction of future associations between unlinked nodes, known as the link prediction problem. Traditional methods are designed to deal with social networks under a certain framework. Yet, data of such networks are usually noisy, missing and prone to observation errors causing distortions and likely inaccurate results. This paper addresses the link prediction problem under the uncertain framework of the belief function theory, an appealing framework for reasoning under uncertainty that permits to represent, quantify and manage imperfect evidence. Firstly, a new graph based model for social networks that handles uncertainties in links' structures is introduced. Secondly, a novel method for the prediction of new links that makes use of the belief functions tools is proposed. It takes advantage of both neighborhood and common groups information in social networks in order to predict new connections. The performance of the new method is validated on real world social networks. Experiments show that our approach performs better than traditional methods based on structural information.

Keywords: Social network analysis; evidential social network; uncertain link prediction; belief function theory; structural information; social circle information

1. Introduction

Social networks are highly dynamic structures that alter quickly over time by the creation of new links. Understanding the processes by which these systems evolve offers a new level of perception of the mechanisms that underlie social networks [1]. Link prediction is a research field in social network analysis that studies the changes of social networks structures. It is an efficient tool to examine the uncertainty and potential connections between unlinked actors.

Link prediction has been drawing the attention of researchers from different domains. The task is to evaluate the potential existence of a link between nonadjacent nodes given a state of the network. Link prediction conducts research in various fields generally related, but not restricted, to computer science, such as recommender systems to

---

*Corresponding author.
Email address: sabrinemallek@yahoo.fr (Sabrine Mallek)

suggest new friends or common interests [2], bioinformatics in the study of protein-protein interaction networks [3] or link analysis and mining [4] for identifying hidden or missing criminals in terrorist networks [5].

It is a well known fact that the structure of social networks highly depends on the precise nature of the data. These latter are often missing and noisy [6, 7] (i.e., missing nodes and/or edges). They are also frequently prone to observation errors which causes distortions. A possible reason is the unreliability of the sources employed for the construction of the network (i.e., open source intelligence, human intelligence) [8]. In many cases, one has to deal with two major problems: consider all the nodes and edges and risk the possibility of adding false ones in the network or delete the uncertain ones and face the problem of missing nodes and edges [9]. On that point, incorporating uncertainty in the structure and analysis of social networks is expected to be crucial.

In recent years, a wide variety of methods have been proposed to address the LP problem. Yet, most of them handle social networks under a certain framework where the links take binary values, either 1 (exist) or 0 (¬ exist). Nevertheless, a possible way to represent an uncertain network is to assign weights with values in [0, 1] to encrypt the uncertainty degrees [8]. The authors in [6] pointed out the shifting degrees of uncertainty characterizing real world large-scale data. They outlined the importance of handling uncertainty in network data which is argued to be more substantial for social networks on account of their large structures. Actually, depiction of uncertainty in social networks has not been addressed in sociology and other fields literature. Traditional analysis focused on interactions in small groups since obtaining large datasets was difficult. Yet, the emergence of online networking and modern database technology allowed to get a great deal of data which stimulated the study of the characteristics of these networks. However, little interest has been dedicated to the examination of their uncertain aspects [10].

The belief function theory also called the Dempster-Shafer theory was first introduced by Dempster [11] in the context of statistical inference then generalized by Shafer [12] as a theory for reasoning under uncertainty. It is an appealing framework considered as a generalization of Bayesian theory. More importantly, it permits to quantify, manage and represent imperfect evidence. Furthermore, it enables combining information from different sources and make decision. We suggest to make use of the belief function theory assets to introduce a new graph-based model for social networks that handles uncertainty at the level of the edges. Additionally, we develop a fruitful approach for LP that operates under uncertainty. Links with and without prior knowledge are considered in the prediction task, information from the common neighbors and shared groups is acquired and whole process of transfer, matching and fusion is applied to get an outlook on the potential new links. Additionally, a new technique for the generation of uncertain social networks is presented.

All in all, the main contributions of this work are as follows:

- The formulation of the problem of link prediction in social networks under the uncertain framework of the belief function theory.

- The development of a graph-based model for social networks that handles uncertainty at the edges level encoded by mass functions.

- The design of a novel framework for link prediction based on our graph model for uncertain social networks that

operates merely with the belief function theory tools.

- The implementation of two algorithms for link prediction according to two strategies: with and without prior knowledge about the links existence.

- Experiments on real life social networks datasets to evaluate the effectiveness of our proposals.

The remaining of the paper is organized as follows: Section 2 summarizes the related work on the LP problem. Section 3 provides some basic knowledge about the belief function theory. Section 4 presents the new framework for LP under uncertainty. Firstly, the new model for uncertain social network is introduced. Then, the proposed algorithms are presented through a detailed description of the individual steps. Section 5 gives the experimental setting and the results. Finally, Section 6 concludes the paper and draws future research directions.

2. Related work on link prediction

Link prediction is becoming increasingly important in many fields such as link analysis, information retrieval and network evolution [13]. In social networks, it is used to predict relationships to be formed, uncover relationships that probably exist but have not been observed, or even to assist individuals in forming new connections [14]. In this section, we formulate the link prediction problem and briefly review some exiting approaches. Then, we give some challenges and present our proposals.

2.1. Problem definition

The most straightforward definition of link prediction is the determination of the likelihood of new associations. In formal terms, given a current snapshot of a social network graph $G(V, E)$ where V is the set of nodes representing the social entities and E is the set of edges encoding social relations. The task is to predict potential future or missing edges according to the considered state of the graph. In fact, when the objective is to predict future links, one should consider a snapshot of the graph at time t and try to determine the likelihood of links' existence at time $t + 1$. In contrast, to predict missing links, one should consider a state of the graph at time t and try to infer the links that may existed at time $t - 1$. They may be unobservable or missing from the data [15]. Yet, to predict future links, one has to consider data with time information regarding the links formulation. However, such data are not usually available. Thus, most methods test LP algorithms on static data where the network is built on the basis of the observable data, then extra links that are not visible but are likely to exist are derived [16].

2.2. Existing methods

Most LP approaches assign weights to links according to the nodes properties. These weights represent computed scores based on the query network. Decision about new links existence is made by ranking the highest L scores, where L is a parameter fixed by the user. Various approaches apply directly this intuition to tackle LP. Meanwhile, another group of methods use the similarity score as features for supervised learning. Other methods use latent models to tackle LP. A brief survey of the latter approaches is presented in the following paragraphs.

Structural features. Structural based approaches can be categorized into two main families: local methods and global methods [17]. The first family compute metrics based on structural similarities between the nodes' characteristics. For example, the number of common neighbors between a pair of nodes. The second family uses nodes' proximity in the network, metrics based on the ensemble of paths are employed to detect the closest nodes. Generally, global methods operate better than the local ones [17]. However, some paths based metrics are time consuming because they use the topological information of the whole network which is not always available. Besides, an important aspect that characterizes social networks is not treated which is the participation of actors in the social groups (clusters, circles). As a matter of fact, in many social networks, users are involved in many social groups at the same time. Thus, methods that use both local and social circles information have been proposed [18, 19, 20, 21].

Supervised learning. Many supervised learning based approaches use similarity measures as features to tackle the LP problem. For instance, LP is formulated as a binary classification task, a set of features based on the nodes' similarities is selected [22]. In [23], the authors used network motifs as a structural feature under supervised learning to predict new links. The authors in [24] addressed LP using classification models to learn the dynamics of social networks using auxiliary networks. In [25], a supervised learning strategy for link prediction is developed using structural metrics as features. In [26], supervised learning is applied to LP using different features such as community information, connectivity, interaction and trust information. Other works extract the essential attributes of nodes and edges and use them as features [27]. However, such information are not generic and choosing the appropriate features can be critical since it depends on the network domain. Besides, node attributes are not usually available in data due to privacy and confidentiality issues. Although, the use of supervised learning allows to benefit of all the advantages of classification, there are three main challenges to be considered: (1) the choice of the appropriate classification algorithm, (2) the suitable combination of features to be considered in learning classifier (3) and handling the class imbalance problem.

Latent models. This group of methods use probabilities to assess the likelihood of links existence [28, 29] where a class is assigned to each node of the network. Latent methods assume a particular arrangement principles of the structure of the network, along with some detailed rules and particular parameters retrieved by maximizing the likelihood of the observed structure. Then, the likelihood of query link is evaluated according to these rules and parameters. They use generally Bayesian probabilistic models, they learn probabilities of the classes by fixing appropriate priors probabilities and adjust them later to estimate the likelihood of co-occurrence of the nodes pairs [28, 29]. For example, in [30], the authors extended matrix factorization and combined latent features with nodes and edges explicit features to predict new links. The authors in [29] addressed link prediction using a local probabilistic model based on Markov Random Field [31], an undirected graphical model. To evaluate the likelihood a new link between connecting u and v, a central neighborhood set based on other nodes that are present in the local neighborhood of u or v is computed. The authors in [32] proposed stochastic relational models based on Gaussian process models for LP. Yet, latent models based methods suffer from high complexity costs du to matrix decomposition, factor matrices and latent features learning which makes such algorithms intractable for large networks.

2.3. Challenges

There are several challenges when dealing with the link prediction problem. The main challenge consists at the rapid growth and evolution of online social networks such as *Facebook*, *Twitter*, *Flickr*, and so on. They are highly dynamic and evolve continuously. That is, effective direct approaches are mandatory to address these constraints [27]. Besides, the cold start problem is a big challenge of LP. It is the problem of not having sufficient initial link structure for training the link prediction task [33]. This problem is especially encountered when the size of network is very small or we do not have any information about the structure of the links. It is also encountered when treating coupled networks, where the task is to predict the links in a network that we have no information except that is it interacting with another known one [34]. Incompletion is another faced challenge while predicting links in online social networks. Actually, almost all social network data contain missing values [35]. This is due to anonymization and privacy preservation concerns. As a result, only part of the data is collected.

2.4. Structural similarity measures

Our proposals draw on structural based models. More specifically, on the intuition of the methods that use local and group metrics. This category of methods is the most basic. Practically, traditional methods use the graph feature measures. One of the most inherent reason is that several similarity score are very easy and simple to compute, especially the local structural metrics [36]. Besides, they are generic since they are related to the graph topology and not to its domain. To this end, we recall, in the following, fundamental definitions of popular state of the art structural local measures and group information scores essential for the understanding of our algorithms.

Local information based measures. Among the well known structural local measures is the "Common Neighbors" [37], denoted by $CN(u, v)$. It gives the number of common neighbors shared between a pair of two nodes u and v in the social network. Let $\tau(u)$ denotes the set of neighbors of a node u. The Common Neighbors measure is defined as:

$$CN_{uv} = |\tau(u) \cap \tau(v)|. \tag{1}$$

Suppose that the graph presented in Figure 1 is a social network. The set of nodes is $\{t, u, v, w, x, y, z\}$. Consider the nodes u and v. The set of neighbors of u is $\tau(u) = \{t, x, y, z\}$ and the set of neighbors of v is $\tau(v) = \{t, w, x, y, z\}$. Hence, the set of common neighbors of the pair (u, v) is $\{t, x, y, z\}$. Thus, $CN_{uv} = 4$.

Another commonly used measure is the Jaccard Coefficient [38] denoted by JC. It accounts all the neighbors of u and v. It is defined as follows:

$$JC_{uv} = \frac{|\tau(u) \cap \tau(v)|}{|\tau(u) \cup \tau(v)|}. \tag{2}$$

For example, the Jaccard Coefficient of the nodes u and v presented in Figure 1 is $JC_{uv} = \frac{4}{5}$.

The Adamic/Adar measure denoted by AA_{uv} weights the intake of each common neighbor x by the inverse of its degree, it is defined as:

$$AA_{uv} = \sum_{x \in (\tau(u) \cap \tau(v))} \frac{1}{log|\tau(x)|}. \tag{3}$$

The Adamic/Adar score of the nodes u and v from Figure 1 is computed as follows: $AA_{uv} = \frac{1}{log|\tau(x)|} + \frac{1}{log|\tau(y)|} + \frac{1}{log|\tau(z)|} + \frac{1}{log|\tau(t)|} = 4.51$.

Preferential Attachment [37], denoted by *PA*, speculates that a new link has probability proportional to $|\tau(u)|$ to connect to u. Thus, the preferential attachment of u and v is proportional to the number of neighbors of u and v. It is computed as follows:

$$PA_{uv} = |\tau(u)| \cdot |\tau(v)| \tag{4}$$

For example PA_{uv} (Figure 1) is computed as follows: $PA_{uv} = |\tau(u)| \cdot |\tau(v)| = 20$.

Resource Allocation [39], denoted by *RA*, is based on the resource allocation process of networks. Common neighbors of two unlinked nodes u and v are considered as transmitters of resources where each one provides a unit so that u sends some resource to v. It is computed as follows:

$$RA(u, v) = \sum_{z \in (\tau(u) \cap \tau(v))} \frac{1}{|\tau(z)|} \tag{5}$$

For example RA_{uv} (Figure 1) is: $RA_{uv} = \sum_{z \in (\tau(u) \cap \tau(v))} \frac{1}{|\tau(z)|} = 1,58$.

The Sørensen index [40] is used generally for ecological community data. It is defined as:

$$Sørensen_{uv} = \frac{2|\tau(u) \cap \tau(v)|}{|\tau(u)| + |\tau(v)|}. \tag{6}$$

The Sørensen score of u and v from Figure 1 is computed as follows: $Sørensen_{uv} = \frac{2|\tau(u) \cap \tau(v)|}{|\tau(u)| + |\tau(v)|} = \frac{8}{9}$.

Salton Index [41] is defined as follows:

$$Salton_{uv} = \frac{|\tau(u) \cap \tau(v)|}{\sqrt{|\tau(u)| \times |\tau(v)|}}. \tag{7}$$

The Salton score of u and v is computed as follows: $Salton_{uv} = \frac{|\tau(u) \cap \tau(v)|}{\sqrt{|\tau(u)| \times |\tau(v)|}} = 0,89$.
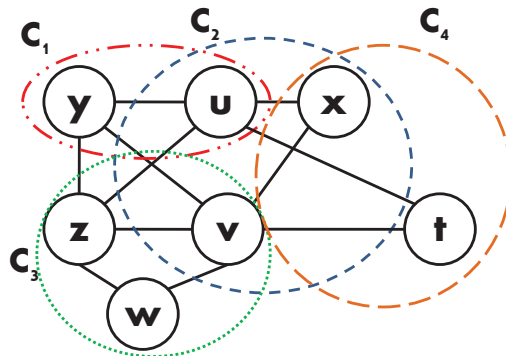

Figure 1: A social network with social circles. The dashed circles cluster nodes belonging to the same social group i.e., y and u belong to the circle C_1 represented by the dashed red circle.

Group information based measures. Group information based measures include the Common Neighbors of Groups (CNG) and Common Neighbors Within and Outside of Common Groups (WOCG) [20]. Let Λ_{uv}^C denotes the set of common neighbors of the pair (u, v) that belong to the social circle C. The CNG describes the size of the set of common neighbors of (u, v) that belong to at least one group C to which u or v is part of. It is defined as:

$$CNG_{uv} = |\Lambda_{uv}^C|. \tag{8}$$

Suppose that the dashed circles C_1, C_2, C_3 and C_4 in the graph presented in Figure 1 describe social groups. The node u belongs to the groups C_1 and C_2 and the node v is member of the groups C_2 and C_3. The set of common neighbors of groups of the pair (u, v) is $\Lambda_{uv} = \{x, y, z\}$. The common neighbor p is not considered since it belongs to the group C_4 which u and v are not part of. Thus, $CNG(u, v) = |\{x, y, z\}| = 3$.

Let $\Lambda_{uv} = \Lambda_{uv}^{WCG} \cup \Lambda_{uv}^{OCG}$ be the set of common neighbors of groups of (u, v) such that Λ_{uv}^{WCG} is the set of common neighbors within common groups (WCG) and Λ_{uv}^{OCG} is the set of common neighbors outside the common groups (OCG). The WOCG measure is defined as:

$$WOCG_{uv} = \frac{|\Lambda_{uv}^{WCG}|}{|\Lambda_{uv}^{OCG}|}. \tag{9}$$

For example, the set of common neighbors within common groups of the pair of nodes (u, v) from Figure 1 is $\Lambda_{uv}^{WCG} = \{x\}$ while the common neighbors outside the common groups is the set $\Lambda_{uv}^{OCG} = \{y, z\}$. Hence, $WOCG_{uv} = \frac{1}{2}$.

Both information of the nodes' neighborhoods and groups are considered in our proposals. We draw on structural local measures since these latter are simple to compute and they have proven their efficiency in many previous works [1, 36, 37].

While the above presented measures (local and group information based) are simple and effective [42], they do not take uncertainty into account. Therefore, in order to handle and manage imperfect social data, the belief function theory is embraced as a framework for reasoning under uncertainty. Mathematical notations and concepts of the belief function theory fundamental for the understanding of our proposals are given in the next section.

3. The uncertain framework of belief functions

The belief function theory [11, 12], is a suitable theory to represent and manage imperfect knowledge. It permits to handle uncertainty and imprecision in data and manage it in a flexible way. Assume the frame of discernment Θ, an exhaustive and finite set of mutually exclusive events associated to a given problem, and let 2^Θ denote the set of all subsets of Θ. Knowledge in the belief function theory is modeled by a basic belief assignment (bba), denoted by m, it is defined as follows:

$$m : 2^\Theta \rightarrow [0, 1]$$
$$\sum_{A \subseteq \Theta} m(A) = 1. \tag{10}$$

A is called a focal element if $m(A) > 0$.

As a special case of bba's, evidence on the frame of discernment is only committed to a particular subset A and Θ. Such a mass function is called a simple support function (ssf). It is defined as [43]:

$$\begin{cases} m(A) & = 1 - \omega, \forall A \subset \Theta \\ m(\Theta) & = \omega. \end{cases} \tag{11}$$

Reliability of the sources of evidence is an important parameter that needs to be considered. When it is quantified, the bba may be weakened before the combination on Θ. For that, a discounting mechanism can be applied [12]:

$$\begin{cases} {}^{\alpha}m(A) & = (1 - \alpha)m(A), \forall A \subset \Theta \\ {}^{\alpha}m(\Theta) & = \alpha + (1 - \alpha)m(\Theta). \end{cases} \tag{12}$$

where $\alpha \in [0, 1]$ is the discount rate.

Evidence given by two reliable and distinct sources of information is combined using the conjunctive rule of combination denoted by ⊙. It is defined as [44]:

$$m_1 \bigodot m_2(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B) \cdot m_2(C). \tag{13}$$

Furthermore, in order to combine two mass functions m_1 and m_2 defined on two disjoint frames Θ and Ω, we need to work under a unified frame. For that, the bba's are extended to the product space $\Theta \times \Omega$. Then the vacuous extension operation, denoted by $\uparrow$, is applied. It is defined by:

$$m^{\Theta \uparrow \Theta \times \Omega}(C) = \begin{cases} m^{\Theta}(A) & \text{if } C = A \times \Omega, A \subseteq \Theta, C \subseteq \Theta \times \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

To define the relation between two different frames of discernment Θ and Ω, the multi-valued mapping may be used [11]. In formal terms, a multi-valued mapping function denoted by φ, joins the subsets $X_i \subseteq \Omega$ that can possibly correspond to $A_i \subseteq \Theta$:

$$m_{\varphi}(A_i) = \sum_{\varphi(X_i) = A_i} m(X_i). \tag{15}$$

To make decisions within the theory of belief functions, one may use the pignistic probability measure denoted by *BetP* [45]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \forall A \in \Theta. \tag{16}$$

4. The evidential link prediction framework

In this section, we start by introducing our evidential graph model for uncertain social networks. Then, we present our algorithms for link prediction under uncertainty which extends state of the art methods and incorporates uncertainty at the same time.

4.1. The model: Evidential social network

A social network is commonly modeled by a graph $G = (V, E)$ where V is the set of nodes and E is the set of links. Yet, such conceptualization does not consider uncertainty that results from imperfect data or unreliability of the tools used for the construction of the network. In this respect, we suggest to encode the uncertainty degrees on the edges level by adopting the belief function theory [46, 47]. Each edge uv has assigned a basic belief assignment denoted by m^{uv} defined on $\Theta^{uv} = \{E_{uv}, \neg E_{uv}\}$, E_{uv} depicts the event that the link between u and v exists and $\neg E_{uv}$ means that it is absent. Thus, this bba describes the degree of uncertainty regarding the existence and absence of uv. Hence, instead of having links with weights that can be either 1 or 0 to show whether or not a link does exist, a mass function with values in [0, 1] that quantifies the degree of uncertainty regarding a link existence is ascribed. Figure 2 gives an example of such a social network graph structure. The dashed circles indicate the social groups to which the users belong. The links are weighted by bba's. For example, m^{ad} is the mass function that encodes the uncertainty degree regarding the existence of an association between nodes a and d. A node may join to multiple circles i.e., the node d participates to the three groups $Group_1$, $Group_2$ and $Group_3$.
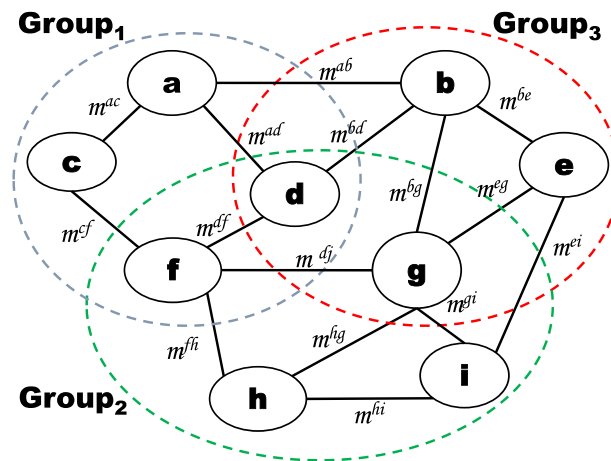

Figure 2: An evidential social network with social groups

4.2. Methodology: Evidential link prediction (ELP) algorithms

Node neighborhood and group information are coupled to forge an overall mechanism for a method that fulfills the process of link prediction under uncertainty. In fact, local metrics are simple and generic, they have proven their efficiency in several social network domains [37, 48]. As such, links connecting similar nodes are more likely to exist. In turn, in many real world social networks, actors who share similar experiences or interests are more likely to share a relationship than those that do not have common characteristics. Therefore, group information add semantic and a better reflection of nodes linkage valuable for LP. The natural arrangement of groups in the network is used e.g., in facebook, the memberships of users to groups of interests.

On the other hand, there are two kinds of evidence regarding the edges existence: the first kind is when we are uncertain about whether the link really exists. In this case, the edge has a bba that quantifies this uncertainty. In the second case, we do not have any knowledge about the link existence. In other terms, the edge has not a mass function. In fact, a preliminary work is done in [47] where we considered the case of not having any knowledge about the edges to be predicted and we inspected the corresponding evidence. However, the effect of group memberships is not taken into account, only structural local information is treated. In [46], we studied the case where the analyzed links already have prior evidence, the task is to update the current knowledge depending on the graph state. In this work, our proposed method takes into consideration these two cases. Besides, it combines local topology and social circles membership to enhance LP.

Accordingly, our proposed problem for link prediction under uncertainty is as follows: Let $\mathcal{G}^e(\mathcal{V}^e, \mathcal{E}^e, C^e)$ be an evidential social network where $\mathcal{V}^e = \{v_1, \dots, v_{|V|}\}$ is the set of nodes, $\mathcal{E}^e = \{(v_i v_j, m^{v_i v_j}) : v_i \neq v_j, m : 2^{\Theta^{v_i v_j}} \rightarrow [0, 1],$ and $\Theta^{v_i v_j} = \{E_{v_i v_j}, \neg E_{v_i v_j}\}\}$ is the set of edges assigned with bba's, and $C^e = \{C_1, \dots, C_{|C|}\}$ is the set of overlapping circles. Let $\mathcal{L}$ be the set of analyzed links. The links in $\mathcal{L}$ might or not have initial bba's. The goal is to update or estimate the bba of query links according to whether or not there is prior information regarding their existence. Prediction is made by considering local and group information.

The steps of the complete method for the prediction of a new link between a pair of nodes (u, v) are presented in the following, they are grouped depending on whether or not there is prior evidence regarding uv.

4.3. Without prior knowledge

Information acquiring. For each common neighbor z, extend vacuously the frames of each link zu and zv to the joint frame $\Theta^{zu} \times \Theta^{zv}$ using Equation 14. This allows to work on a unified referential. The induced bba's from the vacuous extension are then combined using the conjunctive rule of combination to get the masses of the possible pairs on $\Theta^{zu} \times \Theta^{zv}$.

Information transfer and fusion. To successfully transfer the obtained bba's to the frame Θ^{uv} of uv, a multi-valued operation, denoted by φ, is used applied that $\varphi: \Theta^{zu} \times \Theta^{zv} \rightarrow 2^{\Theta^{uv}}$. The φ function (Equation 15) brings together combination sets to the mass of uv considering the common neighbor z denoted by m_z^{uv} as follows:

- The masses of the pairs containing at least an element in $\{E_{zv}, E_{zu}\}$ and not in $\{\neg E_{zv}, \neg E_{zu}\}$ are transferred to $E_{uv} \subseteq \Theta^{uv}$ such that:

$$m_z^{uv}(\{E_{uv}\}) = \sum_{\varphi(S)=E_{uv}} m(S). \tag{17}$$

- The masses of the pairs that contain at least an element in $\{\neg E_{zv}, \neg E_{zu}\}$ and no element in $\{E_{zv}, E_{zu}\}$ are transferred to $\neg E_{uv} \subseteq \Theta^{uv}$ as:

$$m_z^{uv}(\{\neg E_{uv}\}) = \sum_{\varphi(S)=\neg E_{uv}} m(S). \tag{18}$$

- The masses of the pairs including at least an element in $\{E_{zv}, E_{zu}\}$ and an element in $\{\neg E_{zv}, \neg E_{zu}\}$ are transferred to Θ^{uv} such that:

$$m_z^{uv}(\Theta^{uv}) = \sum_{\varphi(S)=\Theta^{uv}} m(S). \tag{19}$$

where $S \subseteq \Theta^{zu} \times \Theta^{zv}$

To get the bba m^{uv}, the masses considering all the m common neighbors $z_i \in \{z_1, \ldots, z_m\}$ are fused using the conjunctive rule of combination (Equation 13) such that:

$$m^{uv} = m_{z_1}^{uv} \bigcirc m_{z_2}^{uv} \bigcirc \ldots \bigcirc m_{z_m}^{uv}. \tag{20}$$

This step permits to fuse the information provided by the neighboring nodes and treat them as independent sources of evidence.

Information reviewing. Once we get a global evidence from of all the neighboring nodes, we review the resulting bba according to whether or not the two nodes uv share a common scial circle. We accomplish this by computing the rate of common neighbors in shared groups $|\Lambda_{uv}^C|$ with the total set of common neighbors CN, that we denote λ_{uv} defined as:

$$\lambda_{uv} = \frac{|\Lambda_{uv}^C|}{|CN_{uv}|}. \tag{21}$$

When all the common neighbors of uv belong to shared groups, λ_{uv} is equal to 1. It is equal to 0 when no common neighbor is part of a mutual circle. We review m^{uv} using a discounting mechanism (Equation 12) such that $\beta = 1 - \lambda_{uv}$ is the discounting rate. The discounted mass denoted by $^\beta m^{uv}$ is computed as follows:

$$\begin{cases} ^\beta m^{uv}(\{E_{uv}\}) & = (1-\beta) \cdot m^{uv}(\{E_{uv}\}) \\ ^\beta m^{uv}(\{\neg E_{uv}\}) & = (1-\beta) \cdot m^{uv}(\{\neg E_{uv}\}) \\ ^\beta m^{uv}(\Theta^{uv}) & = \beta + (1-\beta) \cdot m^{uv}(\Theta^{uv}) \end{cases}$$

4.4. With prior knowledge

Similarity measurement. When we have a prior evidence about uv, we compare it with each link xy included in the common shared circles using the Euclidean distance $D(uv, xy)$. Structural similarity measures based on local and group information are used as features. The most similar link to uv with the smallest distance is considered in the prediction task. The distance metric is divided by its maximum to get values in $[0, 1]$. It is computed as follows:

$$D(uv, xy) = \frac{\sqrt{\sum_{i=1}^n (s_{uv}^i - s_{xy}^i)^2}}{D_{max}} \tag{22}$$

where s_{uv}^i and s_{xy}^i are the values of structural similarity metrics for respectively uv and xy and D_{max} is the maximum value of the Euclidean distance.

Reliability computation. To quantify the degree of reliability of the most similar link, we apply a discounting operation (Equation 12) using the distance value as a discount coefficient denoted by $\alpha = D(uv, xy)$. As such, the more alike the two links are, the more reliable the similar link is. Hence, the discounted mass denoted by $^{\alpha}m^{xy}$ is obtained as follows:

$$\begin{cases} ^{\alpha}m^{xy}(\{E_{xy}\}) & = (1 - \alpha) \cdot m^{xy}(\{E_{xy}\}) \\ ^{\alpha}m^{xy}(\{\neg E_{xy}\}) & = (1 - \alpha) \cdot m^{xy}(\{\neg E_{xy}\}) \\ ^{\alpha}m^{xy}(\Theta^{xy}) & = \alpha + (1 - \alpha) \cdot m^{xy}(\Theta^{xy}) \end{cases}$$

If there is more than one most similar link, we choose the link with the highest mass on "exist" as the degree of certainty of its existence would be higher.

Information transfer and fusion. The discounted bba of the most similar link xy is transferred to the frame Θ^{uv} using a multi-valued mapping $\gamma: \Theta^{xy} \rightarrow 2^{\Theta^{uv}}$ that brings together the elements as follows: $^{\alpha}m^{xy}(\{E_{xy}\})$ is transferred to $m_{xy}^{uv}(\{E_{uv}\})$, $^{\alpha}m^{xy}(\{\neg E_{xy}\})$ is transferred to $m_{xy}^{uv}(\{\neg E_{uv}\})$ and $^{\alpha}m^{\Theta^{xy}}(\Theta^{xy})$ is transferred to $m_{xy}^{uv}(\Theta^{uv})$. Where m_{xy}^{uv} denotes the bba of uv on Θ^{uv} given the most similar link, here xy.

The next step is to update the bba of uv given the new evidence obtained from the most similar link. To do this, the initial bba m^{uv} and m_{xy}^{uv} are aggregated using the conjunctive rule of combination.

4.5. Decision making

After applying one of the two strategies, one has to make a decision about the existence of the link uv on the graph. Decision is made on the basis of the pignistic probability using Equation 16. As a matter of fact, the value of the pignistic probability on the hypothesis "exist" $BetP^{uv}(E_{uv})$ is used as a score to evaluate the likelihood of the existence of new links. Structural methods from literature rank query links on the basis of similarity scores values. In the same manner, we rank the pignistic probabilities value on the hypothesis "exist" in decreasing order and cast the top k predicted links with highest scores.

4.6. Overall summary

Both strategies of the ELP algorithms are schematically given in Figure 3. To sum up, when there is prior knowledge about the link existence, our algorithms collect information from the neighboring links shared with the common neighbors. They are considered as independent sources of information. Then, the collected evidence is pooled to get an overall picture about the link existence. This evidence is boosted through reliability inspection according to groups' information. On the other hand, when there is no prior evidence about the link existence, evidence about the link is gathered from similar links in the shared groups. The similarity is evaluated according to a distance that exploits local and group information structural features. The most similar link represents a source of information, its evidence is reviewed according to its degree of similarity with the query link. The new evidence is pooled with the initial knowledge of the link to be predicted to make decision about its existence. Finally, after passing through one

of the strategies, pignistic probabilities are computed to predict potential links. To sum up, our framework takes an evidential social network as input along with a list of query edges. The analyzed links may already have assigned bba's or not. The ELP algorithms are subsequently applied to provide the potential existing links as the output of the framework.
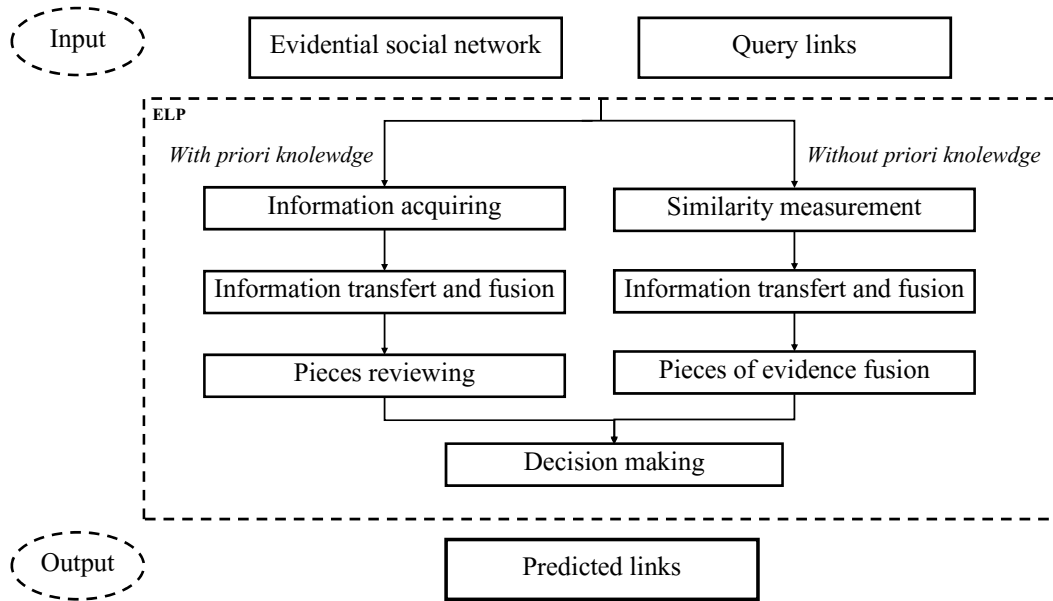

Figure 3: The ELP framework

4.7. Complexity analysis

Analysis of computational time complexity of is essential in social networks. Our novel link prediction framework is based mainly on local information methods, it considers neighboring nodes for prediction. Yet, unlike these methods, it handles uncertainty in social data. Local information link prediction methods are known to have the lowest computational complexity among existing approaches from literature [49]. As our framework is based on local methods, theoretical complexity would be comparable. For the most part, it builds upon the common neighbors approach, thus it follows up the same procedure. Let N denote the number of nodes and let k be the average degree of all nodes in the network. The theoretical complexity of the common neighbors approach (and most local methods) is $O(N.k^2)$ [50]. Additional computational costs due to representing and managing belief masses are very low as we handle frames of discernment with just two elements. Therefore, theoretical complexity of our proposed method is

equal to $O(N.k^2)$. Such complexity is acceptable for link prediction and social network analysis applications.

5. Empirical study

To evaluate the potential benefit of the proposed ELP method, uncertain social networks are needed. Yet, data of such networks are not available. Thus, we preprocessed real-world social networks with social circles of *BlogCatalog* [51], *Flickr* [51] and *Facebook* [52]. Description of the datasets is provided in Table 1.

- BlogCatalog: a social blog directory website. The dataset incorporates a friendship network with group memberships.

- Flickr: an image and video hosting website. It provides web services and online community. The dataset contains the friendship network and the group memberships of the users.

- Facebook: a social media platform enabling users to maintain their network of real-world relationships. The network includes social circles joining people with like-minded interests.

Table 1: Description of the datasets

Dataset	#nodes	#edges	#groups
BlogCatalog	10,312	333,983	39
Flickr	80,513	5,899,882	195
Facebook	4,039	88,234	122

5.1. Network pre-processing

To convert the networks into evidential social networks, we undergo two steps: (1) three graphs of the network are generated from the data (2) then mass distributions are simulated according to the presence of the links in the generated graphs.

5.1.1. Graphs generation

At first, three graphs, that we call G_0, G_1 and G_2, are produced from each dataset by removing randomly a portion of the edges. This procedure is inspired from a widely technique from the literature. In fact, in several works, a portion of the edges is pruned from the graph to be considered in the prediction process [53, 54, 55]. This graph generation technique is applied to all the networks datasets.

5.1.2. Mass functions simulation

To create the evidential version of the social network, each link uv is valued by a bba that is simulated based on G_2, G_1 and G_0 as follows:

- If uv exists in G_2, G_1 and G_0, a bba m^{uv} is created such that $m^{uv}(\{E_{uv}\}) \in [2/3, 1]$ and $m^{uv}(\Theta^{uv}) = 1 - m^{uv}(\{E_{uv}\})$;

- If uv exists in G_2 and G_0 or G_1 and G_0 then a bba m^{uv} is simulated such that $m^{uv}(\{E_{uv}\}) \in [1/3, 2/3[$, $m^{uv}(\{\neg E_{uv}\}) \in]0, 1/3]$ and $m^{uv}(\Theta^{uv}) = 1 - (m^{uv}(\{E_{uv}\}) + m^{uv}(\{\neg E_{uv}\}))$;

- If uv exists only in G_0 then a mass function m^{uv} is ascribed such that $m^{uv}(\{E_{uv}\}) \in]0, 1/3]$, $m^{uv}(\{\neg E_{uv}\}) \in [1/3, 2/3]$ and $m^{uv}(\Theta^{uv}) = 1 - (m^{uv}(\{E_{uv}\}) + m^{uv}(\{\neg E_{uv}\}))$;

- If uv exists in G_2 and G_1 then a ssf m^{uv} is created such that $m^{uv}(\{\neg E_{uv}\}) \in]1/3, 2/3]$ and $m^{uv}(\Theta^{uv}) = 1 - m^{uv}(\{\neg E_{uv}\})$;

- If uv exists only in G_2 or in G_1 then a ssf m^{uv} is assigned such that $m^{uv}(\{\neg E_{uv}\}) \in]0, 1/3]$ and $m^{uv}(\Theta^{uv}) = 1 - m^{uv}(\{\neg E_{uv}\})$.

That is, for each dataset, we get an evidential version from the original graph to which the ELP method is applied to predict the new links under uncertainty.

5.2. Link prediction process

In the experimental phase, the evidential link prediction approach is applied to the evidential version of each graph G_0. The masses of the edges without prior knowledge are determined on the basis of the common neighbors. Edges with prior bba's whose corresponding *BetP*'s claim their absence (i.e., the pignistic probability on the event "not exist" is less than 50%) are compared to the links of the shared groups to update the evidence. Pignistic probabilities are subsequently computed to make decisions about the links existence. The results are finally compared to the initial network. In the similarity measurment step, we use local and group information scores as features to compute distances: CN (Equation 1), JC (Equation 2), AA (Equation 3), PA (Equation 4), RA (Equation 5), Salton (Equation 7), Sørenson (Equation 6), CNG (Equation 8) and *WOCG* (Equation 9).

To quantify the accuracy of prediction algorithms, we use the commonly used metrics: precision and area under the receiver operating characteristic curve (AUC). The AUC evaluates the performance of an algorithm according to predicted links. It depicts the likelihood of the score of an existing randomly picked link being higher than that of A randomly chosen non-existent one. Similarity indices are computed by selecting a random existing link and a non-existent link each time. Then, a missing link and a nonexistent link are randomly picked to compare their scores, if among n independent comparisons, there are n' times the missing link having a higher score and n'' times they have the same score, the AUC value is:

$$AUC = \frac{n' + 0.5n''}{n} \tag{23}$$

Typically, a greater AUC value indicates a better performance quality.

The precision expresses the number of correctly predicted existing links Q_e versus the set of analyzed links Q. It is defined as follows:

$$\text{Precision} = \frac{Q_e}{Q}. \tag{24}$$

5.3. Comparative study with existing methods

A comparative study is conducted with popular approaches local information approaches mostly used in the literature: CN, JC, AA, PA, RA, Salton and Sørensen. Furthermore, we test two other baseline approaches based on group information: *CNG* and *WOCG*. For that, G_0 is considered as a certain graph. In the experimental setup a random fraction of 20% of the links is removed and is considered as the test set along with links that do not exist in the graph. Performance is evualuated by averaging over 5 implementations with independently random divisions of the test set.
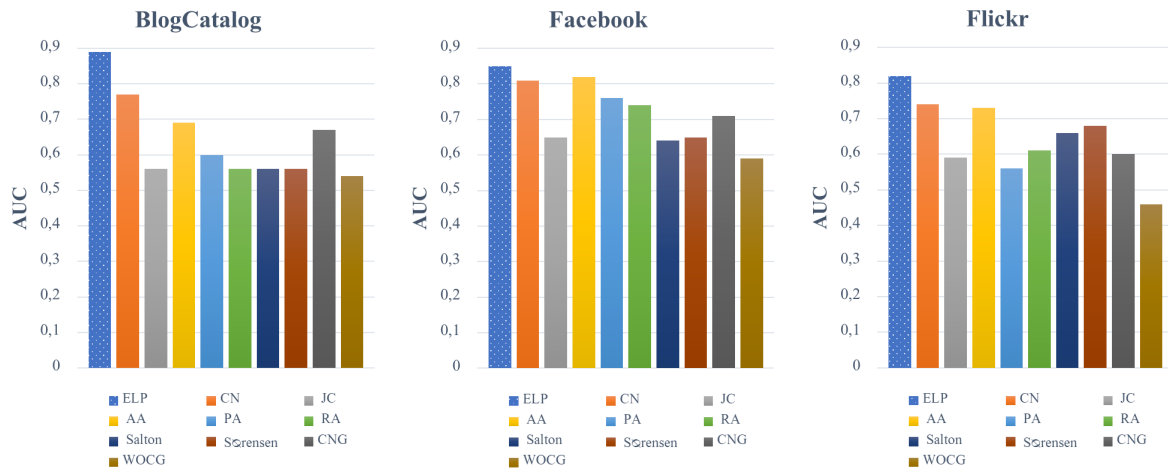

Figure 4: Comparison of link prediction methods on BlogCatalog, Facebook and Flickr networks in terms of AUC

As it can be easily seen from the figure, the best performances are achieved by our novel framework ELP. It gives the highest results in terms of AUC for all the dataset. For BlogCatalog, our method gives by far the best results compared to all other methods, followed by CN, then by AA and CNG, which achieve comparable results. For Facebook, ELP achieves best predictions. Besides AA, CN and PA perform the next best with close accuracies to our best one, while others, such as Salton, Jaccard index, Sørensen and WOCG perform far worse in the cases for Facebook. Same observation for Flickr, ELP gives the highest AUC values. CN, AA perfom the next best accuracies, while WOCG performs the worse. That is, our method clearly outperforms traditional methods. To top it all off, the novel ELP algorithm is significantly better than simple traditional methods. The strength of our approach comes not only from the use of both local and social circles information from the networks, but also from handling uncertainty in the prediction task. Our method operates exclusively using the belief function theory tools. Evidence is adequately collected from neighboring social entities and pooled to get an overall picture thanks to combination rules given by

the belief function theory. Evidence can be even revised by evaluating the degree of trustworthiness of these social entities.
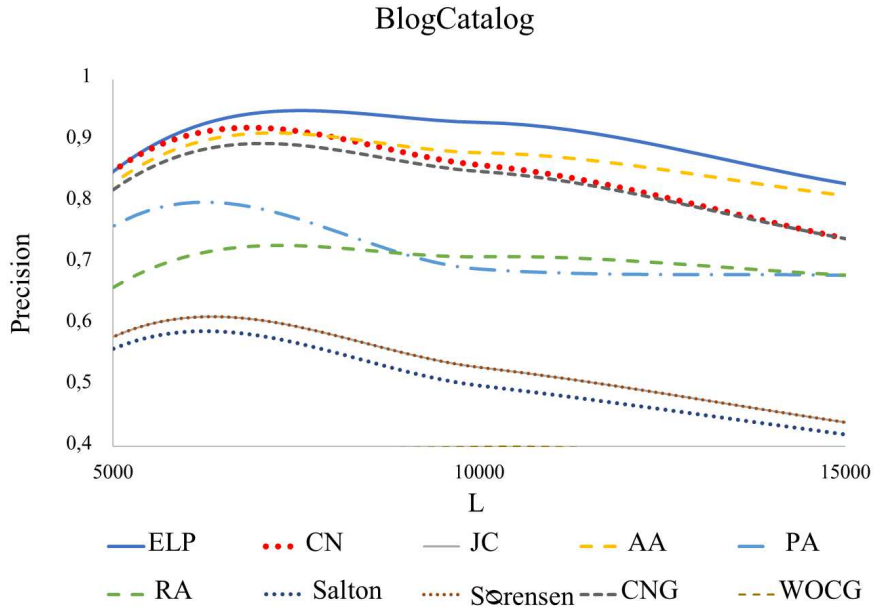
BlogCatalog

Figure 5: Precision of all algorithms for BlogCatalog according to the number of predicted links *L*
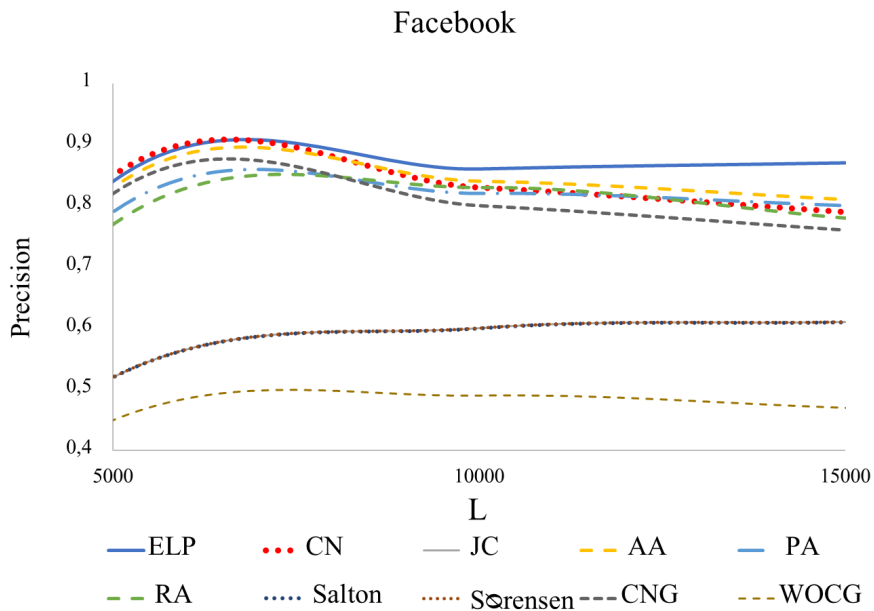
Facebook

Figure 6: Precision of all algorithms for Facebook according to the number of predicted links *L*
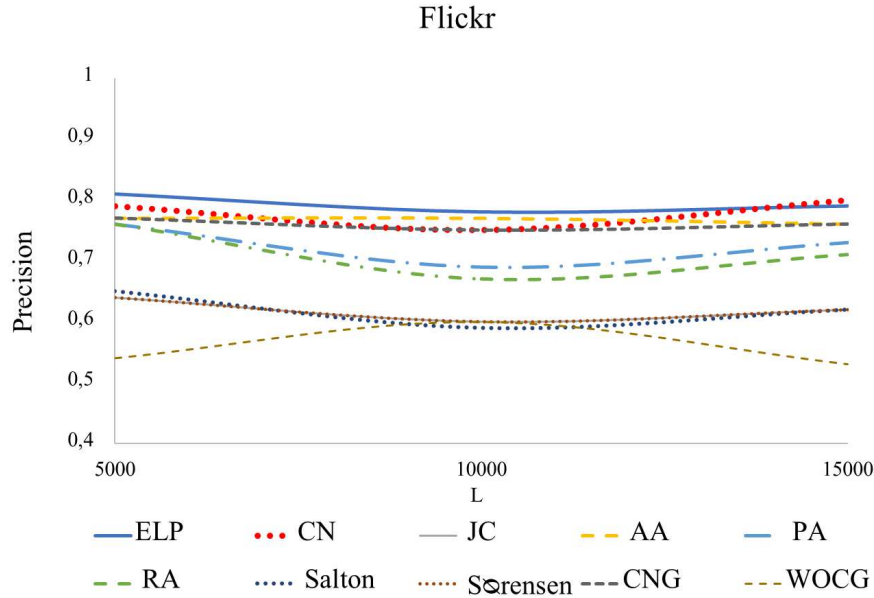
Figure 7: Precision of all algorithms for Flickr according to the number of predicted links *L*

Our proposals are relevant for real world applications since they significantly reduce wrong predictions and boost the right ones.

Figures 5, 6 and 7 report the performance results measured by precision according to the variation of the number of predicted links *L* for all the methods for respectively the BlogCatalog, Facebook and Flickr networks. Precision results are very important to evaluate link prediction algorithms as, in many cases, the main objective is to accurately detect the real existing links. For instance, in Facebook, it is more important not to miss the real actual friends whereas it does not really matter when unknown friends are suggested. As it is observed, the ELP approach has the best precision results compared to state of the art methods for all the datasets. As illustrated, the precision curve of the ELP is above the baseline approaches. That is, the method is able to predict the actual existing associations. It clearly sticks out from these results that handling uncertainty enhances the prediction quality. Moreover, combining measures based on both local information and social circles improves performances. One should also notice that our method is generic as it does not count on the social network domain. That is, it can be applied to social networks from various fields e.g., friendship networks, blog networks, dark networks, collaboration networks, etc.

6. Conclusions and future work

In this paper, we have proposed a new graph model for social networks that handles uncertainty at the edges level. We have developed a new method social links prediction under an uncertain framework. The use of the belief function theory enabled to quantify both the belief regarding the link existence and the uncertainty using mass functions. We

have treated both the cases of having or not evidence about the predicted links. The neighboring nodes are considered as independent sources of information, evidence is transferred and fused to get insight about the edges. It is then, revised and combined to successfully predict new connections. A fruitful technique for the generation of uncertain networks have also been proposed.

A major advantage of our method compared to the existing approaches is that it operates on uncertain social networks. Despite that, using the belief function theory tools, all the information from the neighbors can be combined effectively while remaining cautious about their reliability. Our algorithms can heuristically determine the potential existing links. An empirical study on real social networks clearly demonstrated the effectiveness of our approach.

As part of future work, extension to the case of both uncertain nodes and edges would be considered. It would also be interesting to extend our work to a more complex network structure. For instance, we can treat the case of muti-relationnal networks allowing several types of simultaneous social ties between the actors. Hence, predictions will be made in a more functional manner by inferring both links' existence and types.

References

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

[2] Z. Huang, X. Li, H. Chen, Link prediction approach to collaborative filtering, in: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, JCDL'05, ACM, 2005, pp. 141–142.

[3] C. Lei, J. Ruan, A novel link prediction algorithm for reconstructing proteinprotein interaction networks by topological similarity, Bioinformatics 29 (3) (2013) 355–364.

[4] L. Getoor, C. P. Diehl, Link mining: A survey, SIGKDD Explor. Newsl. 7 (2) (2005) 3–12.

[5] A. Mahesar, Z. Bhatti, A. Waqas, M. Koondhar, M. Rind, S. Nizamani, Efficient link prediction method in dark networks analysis, Sci. Ser. 48 (1) (2016) 81–84.

[6] E. Adar, C. Ré, Managing uncertainty in social networks, Data. Eng. Bull. 30 (2) (2007) 23–31.

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

[8] F. Johansson, P. Svenson, Constructing and analyzing uncertain social networks from unstructured textual data, in: Mining Social Networks and Security Informatics, 2013, pp. 41–61.

[9] J. Dahlin, P. Svenson, A method for community detection in uncertain networks, in: Intelligence and Security Informatics Conference, 2011, pp. 155–162.

[10] P. Svenson, Social network analysis of uncertain networks, in: Proceedings of the 2nd Skövde workshop on information fusion topics, 2008.

[11] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.

[12] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[13] Y. Yang, R. N. Lichtenwalter, N. V. Chawla, Evaluating link prediction methods, Knowl. Inf. Syst. 45 (3) (2014) 751–782.

[14] B. K. Samanthula, W. Jiang, Interest-driven private friend recommendation, Knowl. Inf. Syst. 42 (3) (2013) 663–687.

[15] P. Luo, Y. Li, C. Wu, K. Chen, Detecting the missing links in social networks based on utility analysis, Journal of Computational Science 16 (2016) 51–58.

[16] B. Taskar, M. Wong, P. Abbeel, D. Koller, Link prediction in relational data, in: Proceedings of Neural Information Processing Systems, MIT Press, 2004, pp. 659–666.

[17] L. Lu, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (6) (2011) 1150–1170.

[18] Z. Xie, E. Dong, J. Li, D. Kong, N. Wu, Potential links by neighbor communities, Physica A: Statistical Mechanics and its Applications 406 (2014) 244 – 252.

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

[20] J. C. Valverde-Rebaza, A. A. Lopes, Link prediction in complex networks based on cluster information, in: Proceedings of the 21th Brazilian Symposium on Artificial Intelligence, Vol. 7589, Springer, 2012, pp. 92–101.

[21] J. Wu, G. Zhang, Y. Ren, A balanced modularity maximization link prediction model in social networks, Information Processing & Management 53 (1) (2017) 295–307.

[22] M. A. Hasan, V. Chaoji, S. Salem, M. J. Zaki, Link prediction using supervised learning, in: Proceedings of the 6th SDM workshop on Link Analysis, Counterterrorism and Security, 2006.

[23] F. Aghabozorgi, M. R. Khayyambashi, A new similarity measure for link prediction based on local structures in social networks, Physica A: Statistical Mechanics and its Applications 501 (2018) 12–23.

[24] Z. Lu, B. Savas, W. Tang, I. S. Dhillon, Supervised link prediction using multiple sources, in: Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10, 2010, pp. 923–928.

[25] C. Ahmed, A. ElKorany, R. Bahgat, A supervised learning approach to link prediction in twitter, Social Network Analysis and Mining 6 (1) (2016) 1–11.

[26] H. R. de Sa, R. B. C. Prudêncio, Supervised link prediction in weighted networks, The 2011 International Joint Conference on Neural Networks (2011) 2281–2288.

[27] M. A. Hasan, M. J. Zaki, A Survey of Link Prediction in Social Networks, Springer, 2011, pp. 243–275.

[28] H. Kashima, N. Abe, A parameterized probabilistic model of network evolution for supervised link prediction, in: Proceedings of the Sixth International Conference on Data Mining, ICDM'06, 2006, pp. 340–349.

[29] C. Wang, V. Satuluri, S. Parthasarathy, Local probabilistic models for link prediction, in: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, ICDM'07, 2007, pp. 322–331.

[30] A. K. Menon, C. Elkan, Link prediction via matrix factorization, in: Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD'11, 2011, pp. 437–452.

[31] R. Kinderman, S. Snell, Markov random fields and their applications, American mathematical society, 1980.

[32] K. Yu, W. Chu, Gaussian Process Models for Link Analysis and Transfer Learning, Curran Associates Inc., 2008, pp. 1657–1664.

[33] V. Leroy, B. B. Cambazoglu, F. Bonchi, Cold start link prediction, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'10, ACM, 2010, pp. 393–402.

[34] Y. Dong, J. Zhang, J. Tang, N. V. Chawla, B. Wang, Coupledlp: Link prediction in coupled networks, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'15, ACM, 2015, pp. 199–208.

[35] P. Wang, B. Xu, Y. Wu, X. Zhou, Link prediction in social networks: the state-of-the-art, Sci. China Inf. Sci. 58 (1) (2015) 1–38.

[36] T. Zhou, L. Lü, Y. Zhang, Predicting missing links via local information, The European Physical Journal B-Condensed Matter and Complex Systems 71 (4) (2009) 623–630.

[37] M. E. J. Newman, Clustering and preferential attachment in growing networks, Phys. Rev. E 64 (2) (2001) 025102.

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

[39] T. Zhou, L. Lü, Y.-C. Zhang, Predicting missing links via local information, The European Physical Journal B-Condensed Matter and Complex Systems 71 (4) (2009) 623–630.

[40] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons, Biol. Skr. 5 (1948) 1–34.

[41] G. Salton, M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, Inc., 1986.

[42] S. Gao, L. Zhou, X. Wang, H. Chen, Link prediction via local structural information in complex networks, in: Proceedings of the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2017, pp. 2247–2253.

[43] P. Smets, The canonical decomposition of a weighted belief, in: Proceedings of the Fourteenth International Joint Conference on Artificial

Intelligence, IJCAI'95, Vol. 14, 1995, pp. 1896–1901.

[44] P. Smets, Application of the transferable belief model to diagnostic problems, Int. J Intell. Syst. 13 (2-3) (1998) 127–157.

[45] P. Smets, The Transferable Belief Model for Quantified Belief Representation, Vol. 1, Springer, 1998, pp. 267–301.

[46] S. Mallek, I. Boukhris, Z. Elouedi, E. Lefevre, Evidential link prediction based on group information, in: Proceedings of the 3rd International Conference on Mining Intelligence and Knowledge Exploration, Vol. 9468 of Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 482–492.

[47] S. Mallek, I. Boukhris, Z. Elouedi, E. Lefevre, The link prediction problem under a belief function framework, in: Proceedings of the IEEE 27th International Conference on the Tools with Artificial Intelligence, 2015, pp. 1013–1020.

[48] G. Kossinets, D. Watts, Empirical analysis of an evolving social network, Science 311 (5757) (2006) 88–90.

[49] J. Yang, L. Yang, P. Zhang, A new link prediction algorithm based on local links, in: Proceedings of the 2015 International Workshops on Web-Age Information Management, 2015, pp. 16–28.

[50] F. Gao, K. Musial, C. Cooper, S. Tsoka, Link prediction methods and their accuracy for different social networks and network metrics, Scientific Programming 2015 (2015) 1–13.

[51] L. Tang, H. Liu, Relational learning via latent social dimensions, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09, ACM, 2009, pp. 817–826.

[52] J. J. McAuley, J. Leskovec, Learning to discover social circles in ego networks., in: Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012, 2012, pp. 548–556.

[53] D. Yin, L. Hong, B. D. Davison, Structural link analysis and prediction in microblogs, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, ACM, 2011, pp. 1163–1168.

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

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