An Evidential Method for Multi-Relational Link Prediction in Uncertain Social Networks

Sabrine Mallek[1,2], 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. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), F-62400 Béthune, France eric.lefevre@univ-artois.fr

Abstract. Link prediction is an important problem that permits to analyze networks' evolution. The task is to estimate the likelihood of the existence of future links. Yet, social networks relate individuals via several types of relationships. Thus, it is more interesting to predict the existence and the type of a future link. We focus in this paper on predicting links in multiplex social networks since these latter allow simultaneous relationships with several types. Furthermore, we take into account the uncertainty characterizing the prediction process and social networks noisy and missing data. To this end, we firstly propose an uncertain graph-based model for multiplex social networks that encodes the uncertainty degrees at the edges level using the belief function framework. Furthermore, a novel link prediction approach is subsequently introduced to estimate both the existence and the type of a new link while taking uncertainty into account. Empirical evaluation on a preprocessed real world social network that support our proposals is provided.

Keywords: social network analysis, link prediction, uncertain social network, multiplex social networks, belief function theory

1 INTRODUCTION

During the last years, the World Wide Web has linked tens of thousands to millions of individuals through the Social Web. A great deal of information has become accessible, social network analysis has arisen as a tool to extract and study the patterning of such data. Social networks became the main focus of researchers and analysts from various domains. They are generally conceptualized as graphs where the nodes represent the actors linked by social relationships. One of the major problems handled in network mining and social network analysis is the study of social networks evolving including the prediction of future or hidden links, which is known as the link prediction problem. The task is to evaluate the likelihood of the establishment of a new link between two nodes according to an observed snapshot of the network.

The structure of social networks depends on the quality of the entities and the ties under consideration. They can be homogenous/heterogeneous, uniplex/

multiplex, weighted/unweighted, directed/undirected, etc. For instance, if there is only one type of a relationship between two actors i.e., two friends, two co-workers or two collaborated authors then the tie is called uniplex. On the other hand, if several relationships are shared then it is called multiplex (multi-relational, multi-layered, multi-dimensional) i.e., if two people are friends, co-workers and live in the same building, their association is a multiplex tie (a three-way one). The link prediction has to take into account the topological structure of the social network, especially the characteristics of the links. Most of the traditional methods ignore the relationships labels between the entities, they only treat the existence of the links. However, the likelihood and the type of a connection are frequently interrelated [2, 3].

Yet, most of the state of the art link prediction methods consider links with binary values i.e., 1 (exists) or 0 ($\neg$ exists). Conversely, social networks structure highly rely on the precise nature of the data. Sparse bias alter considerably the analysis results. In contrast, as discussed in [1, 8], social networks data are often exposed to observation errors and are frequently noisy. According to [16], errors about the components of multiplex social networks are expected to be larger as a result of inaccurate experimental settings or technical issues.

In particular, the complexity of multiplex network may affect their properties more adversely. For instance, the extra links can possibly be just duplicates that were generated erroneously by the tools used for the construction of the social network. Besides, real world collections are often missing or have a number of incorrect labels and links. Consequently, one has either to remove a probable valuable information or to take into account all the uncertain information from the data [8]. This imprecision impacts directly the network structure and therefore the outcomes of the analysis. Indeed, we show in previous works [10, 11] the relevance of taking uncertainty into account whether within the structure or through the link prediction process itself. However, we treated the case of uniplex social networks where there are only uni-relational links allowed between the actors.

Accordingly, we embrace the belief function theory [4, 15] to deal with imperfect data and manage uncertain knowledge as it is a general framework for reasoning under uncertainty. We first introduce a new graph model for multiplex social network graph that encodes uncertainty at the edges level. Subsequently, a novel approach for the prediction of new links along with their types in multiplex social networks is proposed. It is inspired from node neighborhood methods and uses exclusively the belief function tools. The common neighbors are considered as independent sources of evidence, information is transferred and combined and is revised afterwords to get a closer picture about the existence of a future link.

This paper is organized as follows. Section 2 presents briefly some related works on link prediction. In section 3, essential belief function notations and concepts are re-called. Section 4 and 5 give our proposals where a new model for uncertain multiplex social network and then a method for the prediction of new connections and their type. Section 6 gives an illustrative example and Section 7 reports the experimental results. Section 8 concludes the paper.

2 RELATED WORK ON LINK PREDICTION

Link prediction has a great applicability in a wide variety of domains as it plays a key role in network analysis. Namely, it is applied to recommend new friends or items in social networks, detect criminals in dark networks, explore missing links in biological networks, etc. The objective is to evaluate the likelihood of a new association between two unlinked nodes at time t' given a snapshot of the network at time $t < t'$ [9].

Most traditional methods consider simple networks allowing only one type of relations. However, a relevant aspect is not treated which is the types of the links. Actually, prediction of link existence and link type are often considered as two independent problems. In the first case, one predicts the future linkage between two nodes, conversely, in the second case, one assumes that the link exists and tries to predict its type. Yet, these two problems are interrelated [2, 3].

As a matter of fact, multiplex social networks highlight the diversity of the links' types and allow simultaneous relationships as it is an aspect of real social life. Formally, a multiplex social network can be defined as a graph $G(V, E_1, \ldots, E_n)$ where V is the set of entities and $E_1, \ldots, E_n$ are the sets of edges each belonging to specific relationship (layer, dimension). In the following, we present traditional methods for link prediction for both the uniplex and the multiplex cases and we introduce the intuition of our proposed approach.

2.1 Link prediction in uniplex networks

There are two groups of link prediction methods depending on the considered information of the network. The first group of methods uses the local structural properties of the nodes i.e., common neighbors, common circles, or the nodes' attributes i.e., age, interests, gender. Yet, the nodes' attributes are usually not available or hidden due to privacy anonymization constraints. Thus, most of the local information based methods use structural similarities metrics. Namely, the most popular ones are "Common Neighbors", "Jaccard's Coefficient" and "Adamic/Adar". For instance, the common neighbors measure, denoted by CN, counts the jointly connected neighbors of a pair of nodes (u, v). It can be defined as $CN_{uv} = |\tau(u) \cap \tau(v)|$ where $\tau(u)$ and $\tau(v)$ are respectively the set of neighbors of u and v. The common neighbors metric predicts links to nodes with many common neighbors. It is based on the intuition that the more two persons share mutual friends, the likely to become friends which has been demonstrated by Kossinets' and Watts analysis [7] made on a large scale social network of student friendships. Furthermore, it is fast and yields to very well results in practice.

The second group of methods uses global information based on global topological properties of the network. Popular approaches include Hitting time, SimRank or the shortest path to reach a node. However, these latter suffer from high complexity since they inquire for global topological properties. Besides, global information is not always available. Also, the additional complexity does not always pay off since local methods can give great performance as well [9].

2.2 Link prediction in multiplex networks

Correspondingly, when dealing with multiplex social networks, two options are available. Treat each layer independently using uniplex graph measures or handle directly the network using multiplex measures [2]. For instance, the neighborhood of a node may be considered in different ways in a multiplex i.e., the union of all the neighbors in all the dimensions or more restrictively, the intersection of the neighbors set across all the layers [6]. Few attempts have been made to address link prediction in multiplex social networks due to manipulation difficulties and the lack of available data of such networks. Some methods tackled it via supervised and unsupervised learning [3, 14]. Other methods applied measures based on local or global information. For instance, the author in [3] proposed a weighted version of the common neighbors method and computed prediction scores for each link type individually.

In this paper, the intuition of the structural local measures is adopted. More precisely, we draw on the method of the common neighbors as it is simple and shows great results in many previous works [7, 13]. Yet, we take both information across local layers and the overall information about the global network into consideration in the prediction task. We adopt the belief function theory to handle uncertainty as it permits to represent and manage imperfect knowledge. It also allows to combine evidence derived from several sources and make decisions.

3 BASICS OF BELIEF FUNCTION THEORY

Mathematical notations and definitions of the belief function theory [4, 15] essential for the understanding of our proposals are given in this section.

Let Θ be the frame of discernment. It is a finite set including exhaustive and mutually exclusive events associated to the problem. Let 2^{Θ} be the power set of Θ. A basic belief assignment (*bba*), denoted by m, is the mass attached to an event given a piece of evidence. It is defined as:

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

When an element $A \subseteq \Theta$ of a mass function m such that $m(A) > 0$, it is called a focal element. Particularly, when a mass function m has only one focal element A such that $m(A) = 1$, it is denoted m_A and it is called *categorical.* The state of total ignorance is represented by the categorical mass function m_{Θ}. In this case, m is called a vacuous *bba*. Perfect knowledge occurs when there is an element $A \in \Theta$ where $m(A) = 1$.

The fusion of two masses m_1 and m_2 derived from two reliable and distinct sources of evidence is ensured using the conjunctive rule of combination [17] denoted by $\textcircled{\cap}$. It is defined by:

$$
m_1 \textcircled{\cap} m_2(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B) \cdot m_2(C). \tag{2}
$$

On the other hand, when at least one of the sources is reliable but we do not know which one it is, the disjunctive rule of combination denoted by $\textcircled{\cup}$ is used [17]. It is defined by:

$$m_1 \textcircled{\cup} m_2(A) = \sum_{B,C \subseteq \Theta : B \cup C = A} m_1(B) \cdot m_2(C). \tag{3}$$

The reliability of the source can be evaluated by a coefficient $\alpha \in [0, 1]$. A discounting mechanism [15] could therefore be performed on m. The discounted mass function is denoted by ${}^{\alpha}m$ and we have:

$$\begin{cases} {}^{\alpha}m(A) = (1 - \alpha) \cdot m(A), \forall A \subset \Theta \\ {}^{\alpha}m(\Theta) = \alpha + (1 - \alpha) \cdot m(\Theta). \end{cases} \tag{4}$$

where α is called the discounting rate. $m(\Theta)$ symbolizes the state of ignorance, it is equal to 1 when $\alpha = 1$. It corresponds to a vacuous mass function. In other terms, the source is fully unreliable. That is, its corresponding evidence is not taken into account in further processing.

In some cases, one may want to revise a mass function m by reinforcing an element A of the frame. This can be done using the reinforcement correction mechanism [12], which is similar to the discounting operation but unlike the later the masses of the focal elements are recovered and redistributed to the element A instead of Θ. Let $\beta \in [0, 1]$ be the reinforcement rate, the reinforcement towards the element A is defined by:

$$\begin{cases} {}^{\beta}m(A) = (1 - \beta)m(A) + \beta \\ {}^{\beta}m(B) = (1 - \beta)m(B), \forall B \subseteq \Theta \text{ and } B \neq A. \end{cases} \tag{5}$$

To set up the relation between two disjoint frames Θ and Ω, we can use a multi-valued mapping operation [4]. Actually, a multi-valued mapping function denoted by τ allows to assign the subsets $B \subseteq \Omega$ that can possibly accord a subset $A \subseteq \Theta$:

$$m_{\tau}(A) = \sum_{\tau(B) = A} m(B). \tag{6}$$

One of the solutions for decision making in the belief function framework is the transformation of the mass functions into pignistic probabilities denoted by $BetP$. It is defined by [18] :

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \forall A \in \Theta. \tag{7}$$

4 EVIDENTIAL MULTIPLEX SOCIAL NETWORK

In previous works [10, 11], we developed an evidential graph-based versions of social networks that handle uncertainty at the edges level. However, the latter model only supports uni-relational homogenous connections between the nodes. Consequently, this paper extends it the multi-relational case.

The proposed evidential multiplex social network graph is defined as $G(V, E_1, \ldots, E_n)$ where $V = \{v_1, \ldots, v_{|V|}\}$ is the set of nodes, and $E_1, \ldots, E_n$ are the sets of edges, each set of edges E_i corresponds to a type i. Each edge $uv \in E_i$ has assigned a mass function defined on the frame of discernment $\Theta_i^{uv} = \{E_{uv}, \neg E_{uv}\}$ denoted by m_i^{uv}. The event E_{uv} means that uv exists and $\neg E_{uv}$ means that it is absent. The *bba* m_i^{uv} quantifies the degree of uncertainty regarding the existence of a link of type i between u and v.

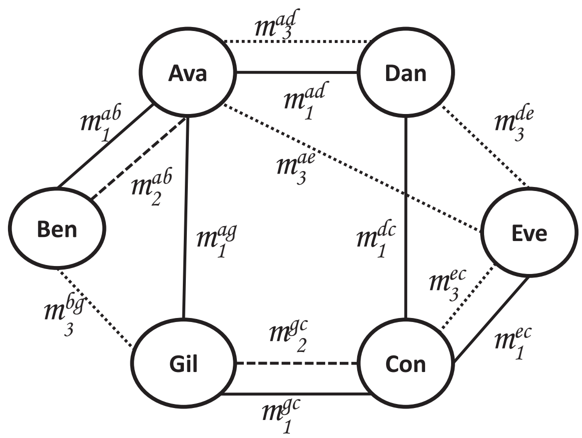

Fig. 1. A multiplex social network graph with *bba*'s weighted edges

Fig. 1 gives an illustration of such a graph structure. The links are weighted with *bba*'s rather than binary values (either 1 or 0) to quantify the uncertainty regarding their existence. The nodes may share three different types of relationships namely m_1, m_2 and m_3. These latter are schematized differently where each type represents a specific association. Hence, there are three layers with the same number of nodes. For the sake of lucidity, a link uv of type i is schematized if its pignistic probability $BetP_i^{uv}(E_{uv}) > 0.5$. In other words, its likelihood to exist is greater than 50%.

5 LINK PREDICTION IN MULTIPLEX EVIDENTIAL SOCIAL NETWORKS

Our proposed approach is based on the intuition of the common neighbors technique. This latter has proved its effectiveness in various real networks and usually got the best performances with respect to other local measures [7, 13]. We have proposed in previous works [10, 11] methods for future links' existence in uncertain social networks. However, we worked on uniplex social network graphs where only a single type of a relationship is allowed. Application of these methods to the multiplex case is not interesting as it involves to carry out the same process for each layer separately. Besides, information regarding the whole structure of the multiplex is not treated. In this work, we take into account the multiplicity of the relations between the nodes. Information about the links' types is treated in order to predict the likelihood of the existence of a link in a specific layer. We draw on the methods based on local structural properties by considering the common neighbors present throughout the global graph.

At first, evidence from the neighboring nodes is gathered from all the layers of the network where each one is considered as an independent source of evidence. Then, the evidence collected from each layer is evaluated according to its reliability. Subsequently, the resulting beliefs are revised according to the distribution of simultaneous links of specific types in the multiplex. Indeed, global information is mandatory for a successful overall link prediction. From this point of view, the steps for the prediction of a future link uv in an evidential multiplex social network $G(V, E_1, \dots, E_n)$, where n is the number of possible types, are as outlined below.

5.1 Information gathering and fusion

Firstly, we extract the subgraph $G_C(V_C, E_C)$ containing the common neighbors of u and v. Since these latter may share various types of relationships with their common neighbors, we decompose G_C into n graphs where each graph $G_i(V_i, E_i)$ includes the links belonging to the type of associations $i \in \{1, \dots, n\}$ such that $V_C = V_i$ and $E_C = \bigcup_{i=1}^{n} E_i$.

Then, for each graph G_i, the masses of each neighboring link xy are transferred to the frame of uv_i using a multivalued mapping operation (Equation 6) $\tau : \Theta_i^{xy} \to 2^{\Theta_i^{uv}}$ in order to get the mass in G_i collected from xy denoted $m_{xy,i}^{uv}$ as follows:

- The mass $m_i^{xy}(\{E_{xy}\})$ is transferred to $m_{xy,i}^{uv}(\{E_{uv}\})$;
- The mass $m_i^{xy}(\{\neg E_{xy}\})$ is transferred to $m_{xy,i}^{uv}(\{\neg E_{uv}\})$;
- The mass $m_i^{xy}(\Theta_i^{xy})$ is transferred to $m_{xy,i}^{uv}(\Theta_i^{uv})$.

Thereafter, we combine the masses that we get from the neighboring links according to the presence of the common neighbors in G_i to get the overall mass of uv_i denoted m_i^{uv}. For instance, if all the common neighbors of the pair (u, v) are in G_i, the masses are fused using the conjunctive rule (Equation 2). That is, all the common neighbors are considered as reliable sources. In contrast, if there are common and uncommon neighbors of (u, v) in G_i, the transferred masses are fused using the disjunctive rule of combination (Equation 3). Hence, we consider at least one of the common neighbors as a reliable source of evidence. G_i is ignored if (u, v) does not share any common neighbor in it and we get a vacuous mass function.

5.2 Reliability evaluation

Next, we focus on the overall reliability of the sub-graphs G_i with respect to the global graph G_C. We compute the distribution of the common neighbors across all the subgraphs G_i defined by $\lambda_i = \frac{|CN_{uv_i}|}{|CN_{uv}|}$. We use $\alpha_i = 1 - \lambda_i$ as a discounting rate to discount m_i^{uv}. Thus, we get a discounted mass function ${}^{\alpha_i}m_i^{uv}$ as follows:

$$\begin{cases} {}^{\alpha_i}m_i^{uv}(\{E_{uv}\}) = (1-\alpha_i) \cdot m_i^{uv}(\{E_{uv}\}) \\ {}^{\alpha_i}m_i^{uv}(\{\neg E_{uv}\}) = (1-\alpha_i) \cdot m_i^{uv}(\{\neg E_{uv}\}) \\ {}^{\alpha_i}m_i^{uv}(\Theta_i^{uv}) = \alpha_i + (1-\alpha_i) \cdot m_i^{uv}(\Theta_i^{uv}) \end{cases} \tag{8}$$

5.3 Evidence reinforcement

The obtained masses are revised according to the distribution of the simultaneous links of more than two types in the multiplex $G(V, E)$. For instance, if we consider a sub-graph G_i and it already exists exactly one link of a type $j \neq i$ between u and v. We compute the distribution of simultaneous 2-relational associations of types i and j denoted by S_{ij}^2 with respect to all the simultaneous relations of exactly two types in G denoted by S_G^2. Generally, when there are $m \leq n-1$ simultaneous links between (u, v), we seek to the distribution S_{*j}^{m+1} where $* = \{1, \ldots, m\}$ are the types of the shared links. If $S_{*j}^{m+1} \neq 0$, we reinforce the mass on the element "exist" using $\beta = S_{*j}^m$ as a reinforcement rate (Equation 5) and obtain the mass ${}^{\beta}m_i^{uv}$.

5.4 Links selection

Most of the methods use the ranking of the similarity scores and consider the L highest ones as the predicted links. In contrast, we first compute the pignistic probability $BetP_i^{uv}$ of the query link uv. Decision about its existence is made later according to the ranking of the pignistic probabilities on the event "exists" of all the analyzed links.

6 ILLUSTRATION

The process of our method is illustrated considering the social network graph in Fig. 1. Suppose we aim to predict the existence of one or multiple relations between *Ava* and *Con*. In the first step of information gathering and fusion (Subsection 5.1), we catch the subgraph containing their common neighbors which are *Dan*, *Eve* and *Gil*. It is subsequently decomposed into three graphs where each one includes the links belonging the a specific type i as shown in Fig. 2.

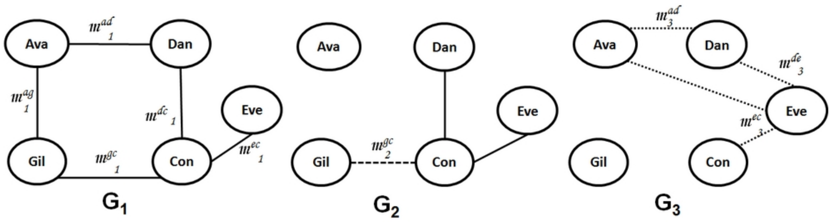

Fig. 2. Decomposition into subgraphs

The masses of the neighboring links in each subgraph G_i are transferred to the mass m^{ac} of the query link ac connecting Ava and Con. Hence, for the each subgraph G_i of type i we get the following *bba*'s:

$$\begin{cases} G_1 : m^{ac}_{ad,1}, m^{ac}_{ag,1}, m^{ac}_{gc,1}, m^{ac}_{dc,1} \text{ and } m^{ac}_{ec,1} \\ G_2 : m^{ac}_{gc,2} \\ G_3 : m^{ac}_{ad,3}, m^{ac}_{de,3} \text{ and } m^{ac}_{ec,3} \end{cases}$$

In order to get the overall mass function for each type, we combine the transferred masses according to the presence of the common neighbors in the subgraphs. For instance, we fuse the obtained masses in G_1 using the disjunctive rule of combination (Equation 3) since there are common and uncommon neighbors i.e., Eve is not a common neighbor anymore. Hence, we get:

$$m^{ac}_1 = m^{ac}_{ad,1} \textcircled{\cup} m^{ac}_{ag,1} \textcircled{\cup} m^{ac}_{gc,1} \textcircled{\cup} m^{ac}_{dc,1} \textcircled{\cup} m^{ac}_{ec,1}$$

The same case is for G_3, the *bba*'s are combined using the disjunctive rule:

$$m^{ac}_3 = m^{ac}_{ad,3} \textcircled{\cup} m^{ac}_{de,3} \textcircled{\cup} m^{ac}_{ec,3}$$

On the other hand, G_2 is discarded since there are no common neighbors on it and we get a final *bba* ($m^{ac}_2(\Theta^{ac}_2) = 1$).

The next step is to evaluate the reliability of the subgraphs according to the distribution of the common neighbors (Subsection 5.2). In fact, Ava and Con share two common neighbors in G_1 which are Gil and Dan. Thus, $\lambda_1 = \frac{2}{3}$. They share one common neighbor in G_3 which is Eve. Hence, $\lambda_3 = \frac{1}{3}$. That is, the reliabilities of G_1 and G_3 are respectively quantified by the reliability coefficients $\alpha_1 = 1 - \frac{1}{3}$ and $\alpha_3 = 1 - \frac{2}{3}$. Therefore, by applying the discounting operation, we get ${}^{\alpha_1}m^{ac}_1$ and ${}^{\alpha_3}m^{ac}_3$.

As shown in Fig. 1, Ava and Con do not share any links. Thus, the obtained *bba*'s after the discounting operation are final. However, if they had a link, for example of type 2. One goes through the reinforcement step (Subsection 5.3) by considering the overall simultaneous 2-relational connections in the global graph. As illustrated in Fig. 1, there are four 2-relational associations i.e., between Ava and Ben, Ava and Dan, Eve and Con, and Gil and Con. Hence, we compute the distribution of simultaneous 2-relational associations of types 1 and 2 i.e., $S^2_{12} = 2$ and the types 3 and 2 i.e., $S^2_{32} = 2$ with respect to all the simultaneous relations of exactly two types in G i.e., $S^2_G = 4$. That is, ${}^{\alpha_3}m_1^{ac}(\{E\})$ and ${}^{\alpha_3}m_3^{ac}(\{E\})$ are reinforced using respectively $\beta_1 = S^2_{12} = 2$ and $\beta_3 = S^2_{32} = 2$ as reinforcement rates. We get the final *bba*'s ${}^{\beta_1}m_1^{ac}$ and ${}^{\beta_3}m_3^{ac}$.

Finally, after the computation of all the bba's of the query links, one decides about the links to be included in the future state of the graph. For that, we compute the pignistic probability of each query link. The highest ones on the event "exists" are selected.

7 EXPERIMENTAL EVALUATION

In order to evaluate our proposals, we test our approach on a real word network of 185 students cooperation linked by 362 edges according to 3 relations: 241 partners links, 23 same computer links and 98 same session links [5]. Since there are no uncertain multiplex social network datasets available, we build it artificially by simulating mass functions on the links regarding their existence. We proposed in previous works [10, 11] methods for the pre-processing of a social network to transform it into an uncertain one where the links are valued with *bba*'s. That is, we simulate mass functions according to the technique proposed in [10]. It is based on a widely applied procedure of graph sampling used in link prediction literature [20]. Actually, one of the traditional solutions to deal with uncertainty is the use of simulation techniques. As pointed out in [19], the use of simulation and sampling techniques is a fruitful way to inspect social networks with uncertain data.

Once evidential multiplex social network is obtained using the proposed approach for the prediction of new links along with their types is applied. We use a popular technique applied in link prediction literature that consists at pruning a number of links from the graph and consider them along with false links that do not already exist to test the performance of the method [20]. Three experiments are conducted using each time a different set of analyzed links. We use precision and recall to evaluate our proposals. The precision depicts the number of relevant links ϵ according to the number of analyzed links δ. It is defined as follows:

$$\text{Precision} = \frac{\epsilon}{\delta}. \tag{9}$$

The recall expresses the correctly predicted existing links ϵ with respect to the number of correctly and falsely predicted existing ones γ. It is defined as follows:

$$\text{Recall} = \frac{\epsilon}{\gamma}. \tag{10}$$

Table 1. Results measured by precision and recall

	E_1	E_2	E_3	Avg
Precision	0.76	**0.80**	0.78	0.78
Recall	0.60	**0.65**	0.62	0.62

Table 1 reports the obtained results in terms of precision and recall with the average values for the three experiments (5th column). As it can be seen, the algorithms give good prediction quality measured by both precision and recall. The precision results are about 78%. That is, our method is able to predict both links' existence and types efficiently. Correspondingly, recall are quite acceptable i.e., a mean value of 62%. Yet, results given by the precision are higher than those of the recall measure. This points out that we are getting more incorrect existing links than incorrect non existing links. Results are close but different in the three experiments due to the variation of the analyzed links. Still, the new approach has proved validity and performance empirically.

To the best of our knowledge, this is the first approach proposed to predict future links' existence and types in multiplex social network while dealing with uncertainty. Consequently, a comparative study is not possible at this stage.

8 CONCLUSION

Multiplex social networks provide a more detailed picture of real life associations. Most traditional methods for link prediction are dedicated to the case of uniplex networks. They only predict future link existence since all the links are assumed to have the same type. They can be adjusted to the multiplex case but this requires the application of the same process several times according to the number of the layers which is time consuming. Besides, global properties of the multiplex will be neglected.

In this paper, we have presented an uncertain multi-relational graph model for multiplex social networks that encapsulates the uncertainty degrees using mass functions encoding the likelihood of the edges' existence. In addition, we have given a new link prediction method that permits to predict both the existence and the type of new associations in multiplex social networks and deals with uncertainty at the same time. Specifically, the proposed method fully operated using the tools of the belief function framework. Evidence from the common neighbors of the global graph is revised and transferred to the query link frame of discernment. It is subsequently updated and fused with the overall information gathered from all the layers to predict both existence and type of the link.

An extension to the case where there are both uncertain edges and nodes is left open for future work. Another direction is to take the actors attributes into consideration. It is clear that some attributes relate directly to the type of the associations between the nodes. Therefore, we intend to investigate the impact of the nodes' attributes on the links' types and consequently the link prediction task.

References

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

[2] Battiston, F., Nicosia, V., Latora, V.: Metrics for the analysis of multiplex networks. CoRR abs/1308.3182 (2013)

[3] Davis, D., Lichtenwalter, R., Chawla, N.V.: Multi-relational link prediction in heterogeneous information networks. In: Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining. pp. 281–288 (2011)

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

[5] Fire, M., Katz, G., Elovici, Y., Shapira, B., Rokach, L.: Predicting student exam's scores by analyzing social network data. In: Proceedings of the 8th International Conference on Active Media Technology. pp. 584–595 (2012)

[6] Kanawati, R.: Multiplex Network Mining: A Brief Survey. IEEE Int. Info. Bull. 16(1), 24–27 (2015)

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

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

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

[10] Mallek, S., Boukhris, I., Elouedi, Z., Lefevre, E.: Evidential link prediction based on group information. In: Proceedings of the 3rd International Conference on Mining Intelligence and Knowledge Exploration. vol. 9468, pp. 482–492 (2015)

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

[12] Mercier, D., Denœux, T., Masson, M.H.: Belief function correction mechanisms. In: Foundations of Reasoning under Uncertainty. vol. 249, pp. 203–222 (2010)

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

[14] Pujari1, M., Kanawati, R.: Link prediction in multiplex bibliographical networks. Int. J. Complex Systems in Science 3(1), 77–82 (2013)

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

[16] Sharma, R., Magnani, M., Montesi, D.: Missing data in multiplex networks: A preliminary study. In: Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on. pp. 401–407 (Nov 2014)

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

[18] Smets, P.: The transferable belief model for quantified belief representation. In: Quantified Representation of Uncertainty and Imprecision. vol. 1, pp. 267–301 (1998)

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

[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)