Evidential Missing 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. Lille Nord de France, UArtois EA 3926 LGI2A, France
eric.lefevre@univ-artois.fr

Abstract. Link prediction is the problem of determining future or missing associations between social entities. Most of the methods have focused on social networks under a certain framework neglecting some of the inherent properties of data from real applications. These latter are usually noisy, missing or partially observed. Therefore, uncertainty is an important feature to be taken into account. In this paper, proposals for handling the problem of missing link prediction while being attentive to uncertainty are presented along with a technique for uncertain social networks generation. Uncertainty is not only handled in the graph model but also in the method itself using the assets of the belief function theory as a general framework for reasoning under uncertainty. The approach combines sampling techniques and information fusion and returns good results in real-life settings.

Keywords: social network analysis, missing link prediction, uncertain social network, belief function theory, information fusion, graph sampling

1 Introduction

Social networks have been witnessing an inconceivable development in recent years, becoming possibly the main actor of the Web 2.0. Social Network Analysis (SNA) has provided a collection of specific models and methods designed for the investigation of social network data and extraction of knowledge from them. The main objective is to determine the conditions under which the patterning of social ties arise and uncover their consequences. From this perspective, link prediction of topology became the focus of many researchers from various domains. It is a task of link mining that aims at predicting new or existent links in the network. In fact, prediction of future links considers the dynamics of the social network. The task is to determine very likely but not yet existing associations based on the previous snapshots of the network. In contrast, prediction of missing links considers its static state rather than its evolution, where the current knowledge is incomplete [15]. In a word, the latter has no temporal aspect, the goal is to predict missing connections to get a more outright picture of the overall structure of the links from the data [27]. Here we are interested at the prediction of missing links under uncertainty in social networks.

As a matter of fact, missing link prediction is of theoretical and practical significance in modern science [26]. In many cases, links might exist at time t but not at t'. A possible reason is a change in privacy settings or when data are partially observed [13], e.g. a facebook user might decide to hide his friends between time t and t', which lead to missing links in the network. This has important ramifications as it may alter estimates of the network statistics [14]. Besides, inferring these missing links raises privacy matters in social networks since several algorithms can be applied to predict new and missing links [9].

On the other hand, data from real world applications are prone to observation errors. They are frequently missing, incomplete and noisy. As pointed out in [2], different degrees of uncertainty characterize several real-world networks especially the large-scale ones. Accordingly, it is an important feature that needs to be taken into account when dealing with social networks from real world data. To handle this uncertainty, the edges might be associated with weights describing their existence in the network. Most of the existing methods use probabilities [3, 13], however, in our case, we propose to use the belief function theory [6, 21] which is considered as a generalization of the probability theory. In fact, one of the practical uses of the belief functions is the representation and management of missing information. It provides convenient ways to handle real life missing data problems [23]. Furthermore, the belief function theory provides tools for combining of evidence induced from several pieces of information. More information about the interest of adopting the belief function theory to handle uncertainty in networks can be found in [5].

Additionally, we design a fruitful approach for missing link prediction that takes into account the uncertain aspects of the social network. It is completely different from methods from link prediction literature as it operates merely with the belief function tools. It uses popular structural measures based on local graph information to compute distances between the links. A fusion procedure is subsequently applied taking into account the reliability of the sources to predict missing links. Besides, a technique based on network sampling is operated for the creation of uncertain social networks to test the validity of our proposals.

This paper is organized as follows: in the next two sections, we examine related literature about link prediction and the belief function theory. In Section 4, we introduce our graph model for uncertain social networks. In Section 5, we design the approach for missing link prediction. In Section 6, we show the experiments we have carried out to test the performance of our approach. Finally, in Section 7, we draw our conclusions and sketch possible future works.

2 Missing link prediction

In recent years, topological link prediction in network evolution has gained the interest of many researchers from various fields. Its applications include exploration of protein-protein interactions, mining food relationships in biological and ecological networks, co-authorship retrieval in collaboration networks, mining frienships, uncovering hidden groups or investigating missing members in social networks.

The most straightforward assumption for link prediction is that two nodes that are similar tend to share a link. To this end, the main concern is how to compute the similarity between nodes accurately. As discussed in several works, methods and measures used in link prediction can be applied for both future and missing link prediction. For a review, see [16].

Typically, social networks are schematized as a graph $G = (V, E)$, where V is the set of social entities and E is the social ties linking them. On the basis of this graph formulation, the link prediction problem can be defined as follows: Let $T_l = (G_l, V_l)$ and $T_k = (G_k, V_k)$ be two states of a social network at times l and k. The link prediction task consists at using T_l to predict the social network structure G_k. We predict new links when $l < k$. In contrast, missing links are predicted when $l > k$ [8]. Most existing methods use the topological information of the networks, including the local or global similarity measures. The local methods consider indices based on neighborhoods in the network while the global methods use the ensemble of paths between the nodes.

2.1 Local information measures

These measures capture node similarity by considering their structural local properties. The most popular property is the set of neighbors $\tau(u)$ of a given node u. The most widely used index is the number of common neighbors [20], denoted by CN. The intuition is that two nodes u and v that share many common neighbors are more likely to form a link. It is defined as follows:

$$CN(u, v) = |\tau(u) \cap \tau(v)| \tag{1}$$

The Jaccard Coefficient (JC) uses all the the neighbors of the pair (u, v) as it considers the number of nodes that are adjacent to at least one of them. It is defined as follows:

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

On the other hand, the AdamicAdar measure [1], denoted by AA, penalizes high degree neighbors since a node with high degree is likely to be in the common neighborhood of other nodes anyway. The AA index is defined as follows:

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

2.2 Global information measures

These measures derive nodes similarity between a pair of nodes (u, v) from paths based on the assumption: the closest two nodes are, the higher the chance for them to be connected. Global information measures include the shortest path, SimRank [12], Hitting time, etc. For instance, the shortest path distance is simply the shortest distance between two nodes. The SimRank index assumes that two nodes tend to be connected if they are linked to similar nodes. The Hitting time consider random walks, it computes the expected number of steps required for a random walker to reach v from u.

2.3 Discussion

Both types of measures are simple and generic, they may be applied to networks from several domains. Yet, they have some shortcomings. The CN measure has proved its efficiency in several real networks and has shown the best performances in many comparisons with others measures based on local information [26, 15]. However, it favors the nodes with large degrees. To solve this problem, variants such as the JC and AA have been proposed to clear up this tendency. On the other hand, path based metrics generally give accurate prediction however they suffer from two major drawbacks. Firstly, they are computationally expensive as they inquire for the global topological information of the network, and are usually impractical on large-scale networks. Secondly, the global topological information is frequently not available [16]. Besides, the additional complexity does not always enhance the prediction, since similar power can be obtained with local methods as well [15]. For that, our approach for missing link prediction uses local information measures.

3 Belief Function Theory

The belief function theory [6, 21], is a suitable theory for the representation and management of imperfect knowledge. It allows to handle uncertainty and imprecision found in data, fuse evidence and make decisions. In fact, belief functions provide convenient solutions to deal with missing information problems, many real life examples are given in [23]. For these reasons, we have adopted this theory to address the missing link prediction problem.

Let Θ be the frame of discernment, 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 represented by a basic belief assignment (*bba*), denoted by m, it is defined as follows:

$$\begin{aligned} m &: 2^{\Theta} \rightarrow [0,1] \\ \sum_{A \subseteq \Theta} m(A) &= 1 \end{aligned} \tag{4}$$

We call A a focal element if $m(A) > 0$.

Evidence induced from two reliable and distinct sources of information may be combined using the conjunctive rule of combination denoted by $\bigcirc\!\!\!\!\cap$. It is defined as [22]:

$$m_1 \bigcirc\!\!\!\!\cap m_2(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B) \cdot m_2(C) \tag{5}$$

On the other hand, to combine two masses m_1 and m_2 defined on two disjoint frames Θ and Ω, the vacuous extension is applied. For that, the *bba*'s have to be extended to the product space $\Theta \times \Omega$. The vacuous extension denoted by $\uparrow$ is defined by:

$$m^{\Theta\uparrow\Theta\times\Omega}(C) = \begin{cases} m^{\Theta}(A) & \text{if } C = A \times \Omega, A \subseteq \Theta, C \subseteq \Theta \times \Omega \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

When combining evidence on Θ, it is important to take reliability of the sources into account. For that, a discounting operation can be applied [21]:

$$\begin{cases} {}^{\alpha}m(A) = (1-\alpha) \cdot m(A), \forall A \subset \Theta \\ {}^{\alpha}m(\Theta) = \alpha + (1-\alpha) \cdot m(\Theta) \end{cases} \tag{7}$$

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

In order to define the relation between two different frames of discernment Θ and Ω, one may use the multi-valued mapping [6]. In fact, a multi-valued mapping operation denoted by τ, joins the subsets $X_i \subseteq \Omega$ that can possibly correspond to $A_i \subseteq \Theta$:

$$m_{\tau}(A_i) = \sum_{\tau(X_i)=A_i} m(X_i) \tag{8}$$

The pignistic probability measure denoted by $BetP$ is usually used to make decisions under the belief function framework [24]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \text{for all } A \in \Theta \tag{9}$$

4 Evidential Social Network

A graph $G = (V, E)$ is the most commonly used representation of social networks where V is the set of nodes representing the actors and E is the set of social links. Yet, binary relationships do not express uncertainty resulting from imperfect data and unreliability of the tools used when constructing the network.

Accordingly, we encapsulate the uncertainty degrees on the edges level using the belief function theory [17, 18]. In fact, each edge uv is weighted by a *bba* denoted by m^{uv} defined on $\Theta^{uv} = \{E_{uv}, \neg E_{uv}\}$, where E_{uv} expresses the event exists and $\neg E_{uv}$ depicts the absence of the link. That is to say, m^{uv} encodes the degree of uncertainty regarding the existence of uv. In other terms, instead of assigning weights that can be either 1 or 0 to describe whether or not a link exists, a mass distribution with values in $[0, 1]$ is ascribed. It is important to notice that links uv with pignistic probability $BetP^{uv}(E_{uv}) < 0.5$ are considered not existing. In other words, the likelihood that uv exists is less than 50%. It is therefore not schematized on the graph.

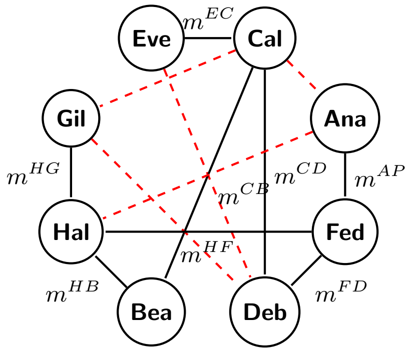

Fig. 1. An evidential social network with missing links and *bba*'s weighted edges

An example of such a graph structure is given in Figure 1, where links are weighted by *bba*'s. The dashed links are the missing ones, they can match to a previous state of the graph or they are unobservable from the data due to noise. Assume that the graph in Figure 1 is a social network of friendships, the nodes represent users and the links describe friend relationships. We can imagine the setting where the user "Gil" decided to hide his list of friends between time t and t', however, one of his friends, here "Hal", is showing his list of friends. Hence, we will have missing links i.e., *Gil Deb* and *Gil Cal* from the social network data at time t'. Since the hidden links do not have assigned masses, our task is to uncover them in order to decide whether the connections actually exist or not.

5 Evidential Missing Link Prediction

To properly deal with uncertainty, it is not enough to just handle links with mass functions attached, we have also to define a proper method for how to take the uncertainties into consideration when applying the link prediction task. As a matter of fact, as discussed in [25], sampling techniques and simulation-based approaches are promising methods to model and analyze social networks with uncertain data. Actually, sampling mechanisms are frequently applied when dealing with missing or partially observed data [4, 13]. One of the reasons is that processing missing data and treatment of sampled data are much alike since what is not sampled can be considered as unobserved. In particular, various effective link tracing and link mining techniques use network sampling [7, 11]. The authors in [10] discussed the connection between sampled and missing data in social networks.

To this end, our proposed method draws on n independent random samples of the social network graph generated from the data. We assume that the links have a priori *bba*'s, the task is to determine the missing ones. For a link to predict, a distance based on local information measures is computed with respect to the links in each graph G_i. The most similar link is considered as the most reliable source of evidence and the information is transferred to the frame of discernment of the analyzed. Finally, evidence gathered from all the graphs is combined to get an overall picture and make decision about the link existence. To this end, we propose a method fulfilling the task of inferring a missing link between a pair of nodes (u, v) based on the five steps presented below.

5.1 Similarity measurement

At first, the Euclidean distance $D(uv, xy)$ between the link uv and each link xy included in each graph G_i is computed. We use structural similarity measures based on local information as features. CN (Equation 1), JC (Equation 2) and AA (Equation 3) are employed since they are simple and they have proved their effectiveness in many scenarios [15, 20, 26]. The most similar link that has the smallest distance is considered. We divide the distance metric by its maximum value to get values in $[0, 1]$. It is computed as follows:

$$D(uv, xy) = \frac{\sqrt{\sum_{s=1}^{n}(sim_{uv}^{s} - sim_{xy}^{s})^2}}{D_{max}} \tag{10}$$

Where s is the index of a structural similarity metric, sim_{uv} and sim_{xy} are respectively its values for uv and xy and D_{max} is the maximum value of the Euclidean distance.

5.2 Reliability computation

Upon determining the most similar link, we quantify its degree of reliability using a discounting operation (Equation 7). The value given by the distance measure is considered as a discount coefficient denoted by $\alpha = D(uv, xy)$. In fact, the more similar the two links are, the more reliable the similar link is, i.e., when the two links are totally similar $D(uv, xy)$ is equal to 0 thus xy is a totally reliable source of evidence ($\alpha = 0$). Hence, m^{xy} is discounted as follows:

$$\begin{cases} {}^{\alpha}m^{xy}(\{E_{xy}\}) = (1-\alpha) \cdot m^{xy}(\{E_{xy}\}) \\ {}^{\alpha}m^{xy}(\{\neg E_{xy}\}) = (1-\alpha) \cdot m^{xy}(\{\neg E_{xy}\}) \\ {}^{\alpha}m^{xy}(\Theta^{xy}) = \alpha + (1-\alpha) \cdot m^{xy}(\Theta^{xy}) \end{cases} \tag{11}$$

One should notice that when there are many similar links, i.e., equal smallest 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 mapping

The discounted *bba* of the most similar link xy defined on the frame of discernment Θ^{xy} has to be transfered to the frame Θ^{uv} of the link to predict. For that, a multi-valued mapping operation (Equation 8) denoted by τ: $\Theta^{xy} \rightarrow 2^{\Theta^{uv}}$ is applied to match the elements as follows:

- The discounted mass ${}^{\alpha}m^{xy}(\{E_{xy}\})$ is transferred to $m^{uv}_{G_i}(\{E_{uv}\})$;
- The discounted mass ${}^{\alpha}m^{xy}(\{\neg E_{xy}\})$ is transferred to $m^{uv}_{G_i}(\{\neg E_{uv}\})$;
- The discounted mass ${}^{\alpha}m^{\Theta^{xy}}(\Theta^{xy})$ is transferred to $m^{uv}_{G_i}(\Theta^{uv})$.

Where $\alpha = D(uv, xy)$ and $m^{uv}_{G_i}$ denotes the *bba* of uv on Θ^{uv} given the most similar link, here xy in the graph G_i.

5.4 Global fusion

Upon gathering information from the n sample graphs, the overall evidence is fused to get the final basic belief assignment denoted by m^{uv}_{f}. The masses $m^{uv}_{G_i}$ obtained from the n graphs are combined using the conjunctive rule of combination such that:

$$m^{uv}_{f} = m^{uv}_{G_1} \textcircled{\cap} m^{uv}_{G_2} \textcircled{\cap} \ldots \textcircled{\cap} m^{uv}_{G_n} \tag{12}$$

At this step, the graphs are treated as independent sources of evidence, the combined information obtained from each most similar link in each graph is fused with the evidence collected from all the graphs.

5.5 Decision process

At the final step, we make decision about whether or not the link is missing (existent). Fot that, we compute the pignistic probability $BetP^{uv}(E_{uv})$ (Equation 9). Actually, if $BetP^{uv}(E_{uv}) > 0.5$ then the likelihood that a link between u and v exists has probability greater than 50%, it would not be considered missing otherwise.

6 Experiments

In our experiments for testing the proposed evidential missing links prediction method, we generated samples of a real social network component of 1500 nodes and 20K edges of facebook friendships obtained from [19]. A simulation phase is subsequently applied in order to transform the samples into evidential graphs. Mass functions are simulated randomly and attached to the edges. The link prediction task is then applied. We compared the predicted missing links with the actual existing ones in the initial graph to test the quality of the results.

6.1 Pre-processing

In the first part of our experiments, we generated 13 samples of the social network graph. A fraction of the existing links is removed and a number of false edges that do not exist either in the sample graph or the original graph is added randomly. Hence, the removed links are the missing ones that we aim to predict when applying the prediction task. Table 1 reports the percentage of false links added to the samples graphs.

Table 1. The percentage of false links

Graphs	G_1,G_2,G$_3$,G_4	G_5,G_6,G_7	G_8,G_9,G_{10}	G_{11},G_{12},G_{13}
False links %	10	15	20	25

At a second step, for all our dataset, we simulate mass functions according to the links' existence in each graph in order to get uncertain versions of the samples. For that, *bba*'s on the links that exist in the original graph corresponding to a pignistic probability that is greater than 0.5 on the event "exist" are randomly created. In contrast, the *bba*'s on the new added links in each sample are generated such that the pignistic probability on the event "not exist" is greater or equals 50%.

6.2 Results

To test our proposals, six experiments E_1, E_2, E_3, E_4, E_5 and E_6 are performed. In E_1, E_2, E_3 and E_4, the missing link prediction approach is applied to respectively three graph samples with the same percentage of false added links, (G_1,G_2,G$_3$), (G_5,G_6,G_7), (G_8,G_9,G_{10}) and (G_{11},G_{12},G_{13}). To analyze the effect of the number of considered graphs, we used in the fifth and sixth experiments respectively two and four graphs samples with the same number of added false edges, (G_1,G_2) and (G_1,G_2,G_3,G_4). The predicted links are subsequently compared with the original graph. The performance is evaluated using two ppopular measures: precision and recall. The precision represents the ratio of the number of relevant predicted existing links n_c to the number of analyzed links n. It is defined as follows:

$$precision = \frac{n_c}{n} \tag{13}$$

The recall catches the correctly predicted existing links n_c versus the correctly and falsely predicted existing ones n_{cf} . It is defined as follows:

$$recall = \frac{n_c}{n_{cf}} \tag{14}$$

In each experiment, 50% of the analyzed links correspond to true missing links that exist in the original graph. The other 50% are false links that do not exist in both the original and sample graphs. The precision and recall results obtained in the experiments are shown in Figure 2 and Figure 3.

As it can be seen in Figure 2 and Figure 3, the prediction quality in terms of precision gives values higher than 60% reaching a maximum performance of 71% in E_6. Besides, the recall measure reaches 61% in E_1 which means that 61% of relevant existing links are predicted by the approach. In other words, the method is able to predict 61% of the actual missing links. It clearly sticks out from these results that our method is applicable on uncertain social networks generated from real world data. That is, validity of our proposals is experimentally showed.

In Figure 2, we observe that prediction accuracy for the four experiments is above 60% for precision and close to 60% for recall. However, the precision

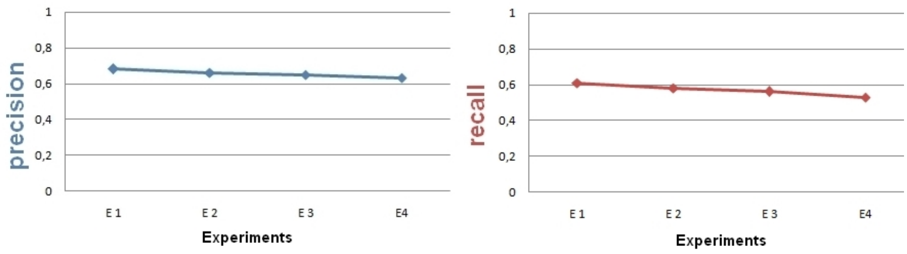

Fig. 2. Precision and recall values obtained in E_1, E_2, E_3 and E_4

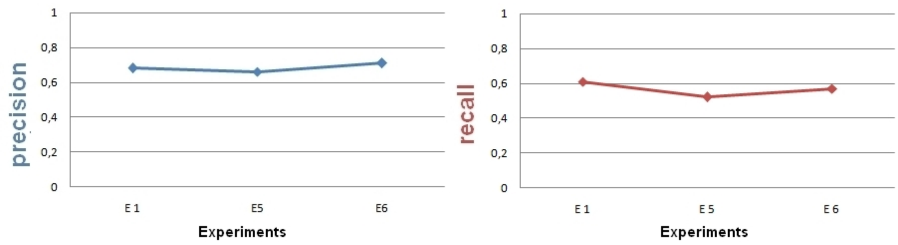

Fig. 3. Precision and recall values obtained in E_1, E_5 and E_6

decreases as the percentage of false edges increases, from 68% in E_1 where the graphs have 10% false edges to 62% in E_4 where the graphs have 25% false added edges. This can be due to the increase of the nodes' degrees. In fact, the proposed method is based on local information measures, these latter are sensitive to nodes neighborhoods. The more the nodes are connected and the more we get similar links when computing distances. The same applies to recall values, it decreases from 61% in E_1 to 53% in E_4. On the other hand, increasing the number of considered graphs enhance the prediction accuracy. As shown in Figure 3, the precision and recall values increase respectively from 64% and 52% in E_5 (two considered graphs) to 71% and 57% in E_6 (four considered graphs). This can be related to the fact that further sources of evidence are considered and new information about the nodes becomes available. Accordingly, more evidence is investigated in the fusion procedure. We also note that in both Figures 2 and 3, precision values are higher than the recall values which points out that the method predicts more incorrect missing links than incorrect non missing links. In other terms, the approach is omitting relevant missing links more than it is predicting false non existing ones. Although the results given by the recall measure are quite satisfactory i.e., 61% in E_1, they are considerably smaller due to the large size of the dataset which challenges the algorithms.

Unfortunately, a comparative analysis cannot be accomplished at this point since, to the best of our knowledge, there is no existing approach that addresses missing link prediction under uncertainty. On the other hand, comparison with the state of the art methods is not engaging since they do not operate on the same graph structures.

7 Conclusion

Missing link prediction is a substantial problem in social networks as it helps analyze and understand social groups. It enables the implementation of efficient tools to discover hidden groups or to investigate missing members, etc. which are very crucial problems in security analysis and criminal investigation.

We have proposed a graph model for social networks that handles uncertainty degrees regarding the links existences using mass functions. A new method for the prediction of missing links have also been investigated. It uses local information of the graph topology to compute distances between the nodes. These information are subsequently transferred and fused using the belief function theory tools to get a global information and make decisions about the links' existence. Our proposals have been evaluated on a real world online social network of facebook friendship. Experimental results given by the precision and recall measures show that our method provides accurate prediction.

Interesting avenues for future research include prediction of jointly missing attributes of the nodes under uncertainty. In fact, several methods use additional information about the nodes and edges to predict future or missing links. Yet, these attributes are frequently missing from the data due to privacy or anonymization issues. Therefore, it would be interesting to study the problem of jointly missing links and attributes under an uncertain framework.

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. Ahmed, N.M., Chen, L.: An efficient algorithm for link prediction in temporal uncertain social networks. Information Sciences 331, 120 – 136 (2016)
4. Clauset, A., Moore, C., Newman, M.E.J.: Hierarchical structure and the prediction of missing links in networks. Nature 453, 98–101 (2008)
5. Dahlin, J., Svenson, P.: A method for community detection in uncertain networks. In: Proceedings of the 2011 European Intelligence and Security Informatics conference. pp. 155–162 (2011)
6. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
7. Gile, K.J., Handcock, M.S.: Respondent-driven sampling: An assessment of current methodology. Sociological Methodology 40(1), 285–327 (2010)
8. Gong, N.Z., Talwalkar, A., Mackey, L., Huang, L., Shin, E.C.R., Stefanov, E., Shi, E.R., Song, D.: Joint link prediction and attribute inference using a social-attribute network. ACM Trans. Intell. Syst. Technol. 5(2), 27:1–27:20 (2014)

9. Gong, N.Z., Talwalkar, A., Mackey, L.W., Huang, L., Shin, E.C.R., Stefanov, E., Shi, E., Song, D.: Predicting links and inferring attributes using a social-attribute network (SAN). CoRR abs/1112.3265 (2011)
10. Handcock, M.S., Gile, K.: Modeling social networks with sampled data. Ann.Appl.Stat 4(1), 5–25 (2010)
11. Heckathorn, D.D.: Comment: Snowball versus respondent-driven sampling. Sociological Methodology 41(1), 355–366 (2011)
12. Jeh, G., Widom, J.: Simrank: A measure of structural-context similarity. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 538–543. KDD '02, ACM (2002)
13. Koskinen, J.H., Robins, G., Wang, P., Pattison, P.: Bayesian analysis for partially observed network data, missing ties, attributes and actors. Social Networks 35(4), 514–527 (2013)
14. Kossinets, G.: Effects of missing data in social networks. Social Networks 28, 247–268 (2003)
15. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. 58(7), 1019–1031 (2007)
16. Lu, L., Zhou, T.: Link prediction in complex networks: A survey. Physica A 390(6), 1150–1170 (2011)
17. 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. Lecture Notes in Computer Science, vol. 9468, pp. 482–492. Springer International Publishing (2015)
18. 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 (ICTAI). pp. 1013–1020 (2015)
19. McAuley, J.J., Leskovec, J.: Learning to discover social circles in ego networks. In: Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012. pp. 548–556 (2012)
20. Newman, M.E.J.: Clustering and preferential attachment in growing networks. Phys. Rev. E 64, 025102 (2001)
21. Shafer, G.R.: A Mathematical Theory of Evidence. Princeton University Press (1976)
22. Smets, P.: Application of the transferable belief model to diagnostic problems. International Journal of Intelligent Systems 13(2-3), 127–157 (1998)
23. Smets, P.: Practical uses of belief functions. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. pp. 612–621. UAI'99 (1999)
24. Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. 66(2), 191–234 (1994)
25. Svenson, P.: Social network analysis of uncertain networks. In: Proceedings of the 2nd Skövde workshop on information fusion topics (2008)
26. Zhou, T., Lü, L., Zhang, Y.: Predicting missing links via local information. The European Physical Journal B-Condensed Matter and Complex Systems 71(4), 623–630 (2009)
27. Zhu, Y.X., L, L., Zhang, Q.M., Zhou, T.: Uncovering missing links with cold ends. Physica A: Statistical Mechanics and its Applications 391(22), 5769 – 5778 (2012)