Physica A 570 (2021) 125795

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

How to predict crime — informatics-inspired approach from link prediction

Nora Assouli [a,1,*], Khelifa Benahmed [a,1], Brahim Gasbaoui [b,1]

[a] *Energarid Laboratory, Tahri Mohamed University, Bechar, Algeria*
[b] *Laboratory of Smart Grids & Renewable Energies (S.G.R.E), Tahri Mohamed University, Bechar, Algeria*

ARTICLE INFO

Article history:
Received 14 February 2020
Received in revised form 31 December 2020
Available online 30 January 2021

Keywords:
Link prediction
Bipartite graph
Internal links
Similarity measures
Criminal activity prediction

ABSTRACT

Many social complex networks are best modeled as a bipartite graph and they evolve during time, thus, predicting links that will appear in them have become highly relevant and critical. Link Prediction is a key direction in social complex network research refers to estimating the possibility of the existence of non-existent links between node pairs. In criminal networks, link prediction can provide a very efficient way in the discovery of missing or hidden links and the detection of the underground groups of criminals. Only few works address the bipartite case, though, despite its high practical interest and the specific challenges it raises. Likewise, most of prior algorithms of link prediction consider a threshold. However, it is difficult to set such a proper threshold in advance for a given dataset. Hence, in this paper, we propose a new method called Latent Link Prediction based on Internal and Local Links (LLPIL) for bipartite networks. LLPIL is based on new proposed topological metric named reliability that can reflect the likelihood of two nodes to be connected. We exploit the proposed model to identifying and preventing future criminal activities. Extensive simulations show that our proposed algorithm has high prediction accuracy and low time complexity.

1. Introduction

Complex network has shown to be highly effective in modeling and analyzing many complex systems, such as social, biological and information systems [1–3]. Such network structure is made up of individuals represented by nodes and their relations expressed by links [1,4,5]. In recent years, networks which are formed via social interactions have been increasingly attracting research attention due to the heterogeneity of their components [6,7]. Dynamic variation and evolution of these links over time make it necessary to predict missing and potential links [5,8–11]. Criminal networks, in particular, exhibit a relatively high propensity to have hidden or missing nodes and links due to the covert and stealthy nature of criminal activities, the incompleteness, incorrectness, and inconsistence of the captured data [12]. Crime can be modeled as a generalized graph consisting of nodes (people, events, etc.) and links (relationships between nodes) based on variables as the group, time, and geographic location… [13]. Link prediction is an important hot topic of complex networks [5] aims to estimate the probability of a link between two nodes [1,3,8,14] using the network structure [8,15,16]. It has applications in different areas [1,3,17] such as bibliographic field, criminal investigations, co-authors scientific

* Corresponding author.
E-mail addresses: mg.assouli@yahoo.fr (N. Assouli), benahmed_khelifa@yahoo.fr (K. Benahmed), gasbaoui_2009@yahoo.com (B. Gasbaoui).
[1] All authors contributed to the final manuscript.

https://doi.org/10.1016/j.physa.2021.125795
0378-4371/

collaboration [18–20], movie actors network [21], sharing habits [22] and recommending systems [14,23,24]. Liben-Nowell [25] et al. proposed one of the earliest link prediction models for social networks. In monitoring networks of criminals, link prediction is used to discover hidden connections between criminals from existing information [5,11,26,27] to detect underground relationships and to prevent crime or terrorist activity [11,28] which it can reduce the effects of an attack in many fields as security threat like life of a victim, stability threat, infrastructure destruction and so on [29]. Recently, many methods have been proposed to predict potential links in bipartite networks [17,24] such as learning and similarity-based ones [30]. A learning-based approach is treating the link prediction problem as a binary classification task [31,32]. Tolan [33] et al. used five classifiers namely: Naïve Bayes (NB), K-Nearest Neighbor (KNN), Tree Induction (C4.5), Iterative Dichotomiser (ID3) and Support Vector Machine (SVM) for terrorism prediction. Experiments were performed on real-life data represented by Global Terrorist Data (GTD) with the help of WEKA as one of open software in data mining written in JAVA [34]. Learning-based methods have the difficulties in feature selection and unbalancing output classes and are suffered from computational cost and limitation of capacity [29,31]. Therefore it is not suitable for large scale and dynamic networks [31]. Similarity-based approaches provide the simplest framework [14,15], they are often very efficient for its low computational complexity [31,35,36]. A similarity-based approach is to compute the similarities on non-connected pairs of nodes by various graph-based similarity metrics like nodes' information, network topology, etc [32] and assumes that the greater the similarity values between nodes are, the higher the likelihood of the existence of links between them [3,5,8,17,24,37,38]. Several methods exist to measure node similarity, such as Adamic–Adar (AA) [39], Resource Allocation (RA) [40] Common Neighbor (CN), Jaccard Coefficient (JC) and preferential attachment (PA) [41–43]. Kumar et al. [44] explored CN, JC, AA and RA for predicting the hidden links or the future ones on GTD. Authors did not differentiate from the predicting link in term of importance (rating the prediction link and its probability of occurrence). CN, JC and PA were used in [45] where Allali et al. address the problem of link prediction in dynamic bipartite graphs by proposing method called Internal Link Prediction (ILP) based on a special introduced kind of links called internal link (IL) that represent the core of their work. The concept of the IL will be further elaborated in Section 2 of this paper. Their main approach consists in transforming bipartite graph into projection graph and assigning weights to the edges in this graph called induced link (IdL). ILP performed very well, it was purely structural. Moreover, it may be extended in several ways. The projection can be performed on either part of the bipartite graph. However, the authors did not explain whether different projections could produce different results or the part on which the projection should be performed to obtain a better result. Furthermore, the considered potential links (ILs) are those with weights greater than a threshold. Therefore, a proper threshold must be predefined to obtain accurate results. However, it is difficult to set a proper threshold in advance for a given dataset. These disadvantages are considered to be study by Gao et al. [46]. In this work, authors proposed the Potential Link Prediction (PLP) method. They defined a new kind of links called Candidate Node Pair (CNP) which was somewhat similar to the IL and the notion of patterns covered by CNP. Patterns were similar to the IdLs in the previous work. Gao et al. defined a new measure called connectivity of the CNP, and they used it as the final score of link prediction. The experimental results showed that PLP could achieve higher speed and superior quality link prediction results in bipartite networks compared with other methods because more topological information is considered in connectivity. In addition, the authors proved that projecting to either part will achieve the same result. Therefore, to reduce the computation time, the graph must be projected only to the part with fewer nodes. The connectivity used as a similarity index is not enough; authors did not specify the connectivity values to be considered in the calculation. In this study, we propose a new method for link prediction in bipartite graph has the advantages of [45] and [46] and overcomes their limitations based on local and adjacency information. Our main contributions are summarized as follows:

- Our proposed method is based on the highly correlated and mutually helpful information contained in nodes connections and it is easy to be implemented with low computation time and higher accuracy. It is based on reliability rather than threshold; the reliability is a new measure that the existing literature on similarity metric for link prediction does not appear to use.
- Our work contributes to predicate the criminal incident to protect the whole world. It will be used as a core framework for early warning system and it can help to develop a crime analysis tool that assists in detecting crime patterns, identify and analyzing common patterns to reduce further occurrences of similar incidence and providing information to formulate effective strategies for crime prevention.
- We evaluated the computation of the proposed method by a detail quantitative analysis. To the best of our knowledge, it is the efficient one for prediction in terms of computation and accuracy overhead. This has been verified by the effective results that have shown for predicting criminal acts using the most comprehensive and world's largest dataset GTD.

The remainder of the paper is organized as follows: Section 2 outlines graph theoretical preliminaries and basic notions, Section 3 gives the framework of our method of link prediction. Experimental results and discussions are presented in Section 4. Finally, Section 5 concludes this paper exploring the future directions.

2. Mathematical preliminaries

In this section, we present a few preliminary definitions that will be used in the rest of this paper.

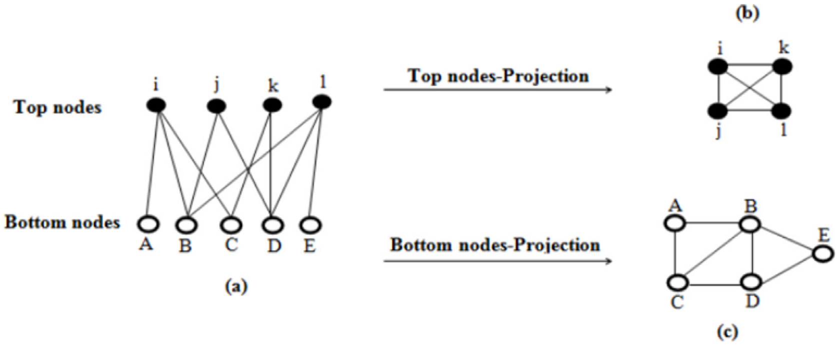


Fig. 1. Bipartite graph (a), and its projected graphs (b) and (c).

Definition 1 (*Bipartite Graph*). $G = (U, V, E)$ is bipartite graph, where U and V are two disjoint sets of vertices and E is the set of edges in which: $\forall e = (u, v) \in E$, $u \in U$ and $v \in V$, $U \cup V = E$ and $U \cap V = \emptyset$[24,45]. Assuming that U has n nodes and V has m nodes, the adjacency matrix A of G takes the following form:

$$A = \begin{vmatrix} 0_{n,n} & B_{n,m} \\ B^T_{m,n} & 0_{m,m} \end{vmatrix} \tag{1}$$

where $0_{n,n}$ and $0_{m,m}$ are n x n and m x m zero matrices, respectively, and B is an n x m nonzero matrix. In this case, the smaller matrix B uniquely represents G, and the remaining parts of A can be discarded as redundant [46]. The neighbors $N(x)$ of a node $x \in U$ in G is the set of all nodes such that there exist at least one edge linking x and this node: $N(x) = \{ y|y \in V, (x, y) \in E \}$. The number of its neighbors represents its degree, namely $D(x)$.

Definition 2 (*projected Graph*). The projection of the bipartite graph $G = (U, V, E)$ is to project it into unipartite graph of one kind of its original nodes. As a consequence, in the projection over the nodes in U-part (G_u), the neighbors of each node in V-part are linked between them in G_u and in the projection over the nodes in V-part (G_v), the neighbors of each node in U-part are linked between them in G_v. The U-projected graph of G is $G_u = (U, E_u)$, where the set of edges [45,46] is:

$$E_u = \{(x, y)|x, y \in U, \exists z \in V, z \in N(x) \cap N(y)\} \tag{2}$$

$N(x)$, $N(y)$ is the set of neighbors of nodes x and y respectively in G.

Similarly, G_v can be defined. Fig. 1 shows example of a bipartite graph and its projections.

Definition 3 (*Internal Links*). (x, y) is an IL in the bipartite graph $G = (U, V, E)$ such that: $x \in U$, $y \in V$ and $(x, y) \notin E$ if and only if it is an IL by U-projection and it is an IL by V-projection.

In other words, an IL in a bipartite graph G is a pair of nodes x and y such that adding the link (x, y) to G does not change its G_{proj}; i.e. $G_u = G'_u$ and $G_v = G'_v$.

In Fig. 1, for example, (B, k) is an IL, because all neighbors of k in G, namely $N(k) = \{C, D\}$ are already linked to B in Fig. 1(c), namely $N_u(B) = \{A, C, D, E\}$, i.e. $N(k) \cap N_u(B) = \{C, D\}$ and the condition is also true in the projection over the top nodes, i.e. all neighbors of B in G, namely $N(B) = \{i, j, l\}$ are already linked to k in (Fig. 1(b)), namely $N_v(k) = \{i, j, l\}$, i.e. $N(B) \cap N_v(k) = \{i, j, l\}$. In either case, the intersection does not equal the empty set.

Definition 4 (*Induced Link*). Given a bipartite graph $G = (U, V, E)$, the set of links induced by any pair of nodes $x \in U$ and $y \in V$ in G_u is:

$$E_u(x, y) = \{x\} * N(y) = \{(x, w), w \in N(y)\} \tag{3}$$

where E_u is the set of links in G_u. Hence, (x, y) is in IdL covered by the link (x, w). Similarly, we can defined the set of IdL in G_v. So, G_u(Likewise G_v) is made entirely of IdLs. IdLs represents the nodes within G that share endpoints. By definition, a link is an internal if and only if all the links it induces are already in G_{proj}. In Fig. 1, for example, let consider U is the bottom nodes and V is the top-nodes, then E_u of the link $(B, k) = \{B\} * N(k) = \{(B, C),(B, D)\} \sqsubseteq G_u$ and E_v of the link $(B, k) = \{k\} * N(B) = \{(k, i),(k, j$ and $(k, l))\} \sqsubseteq G_v$ and therefore (B, k) is an IL.

Table 1
Definition and description of the relevant variables.

Term	Designation
G_u	U-projection of the bipartite graph $G = (U,V,E)$
E_u	Induced links in G_u
$N(A)$	Neighbors of node A
(A,B)	Induced link (A,B)
(A,x)	Internal link (A,x)
$W(A,B)$	Weight of (A,B)
$S(A,x)$	Connectivity of (A,x)
$R_d(A,B)$	Reliability of (A,B)
$R_I(A,x)$	Reliability of (A,x)
I	Set of ILs which induces links satisfying condition in formula (11), these ILs have $S \geq R_I$

3. Algorithm

3.1. Motivation

Our proposed method combines the advantages of the previous works [45] and [46] and it addresses their weaknesses. In our algorithm, the link is unpredictable unless if it is a link in the projection on both sides. This means that the projection will produce the same results, regardless of the projection side. Hence, the projection will be on the subset with the lowest number of nodes to reduce the time of calculation and the identification of the ILs enumerates only the efficient links. Therefore, the time complexity of our proposed algorithm is $O(n)$, where n is the number of the fewer nodes in two subsets. Also, to calculate the weight of IdLs, our algorithm does not depend on a threshold and does not depend on one measurement index that does not have specific values as if there is a use of the threshold indirectly but rather, it depends on two proven metrics which integrates more topological information as measures of similarity and it compares between them to achieve more perfect prediction results.

3.2. Algorithm design

In this section, we describe our LLPIL algorithm applied in criminal act prediction. At a high level, this involves creating a G_{proj} from G, identifying all possible ILs and predicting the potential ones with weights. These phases are detailed below. In order to accurately describe the method, the definitions of the relevant variables are given in Table 1. In the rest of this paper, we utilize uppercase letters, such as A, B to designate the U-part nodes and lowercase letters, such as x, y to designate the V-part nodes. To make the annotation more smooth, we will consider that U-part has the lowest number of nodes and it will be the same way if the opposite is.

We create the projection graph of G over the part with few nodes. Then, we weight the induced links.

3.2.1. How to weight an induced link ?

To weight the IdLs, we use two weighting functions. The weight W defined in [46] and the new proposed metric called the reliability of an IdL, which play a key role in our proposed method.

3.2.1.1. Weight of induced link according to [46]. Suppose that $G_u = (U, E_u)$ is the projected graph of $G = (U, V, E)$. Let (A, B) be an IdL in G_u, then its weight as defined in [46] is as follows

$$W(A,B) = \frac{2}{D(A)+D(B)} \sum_{v\in N(A)\cap N(B)} \frac{1}{D(v)} \tag{4}$$

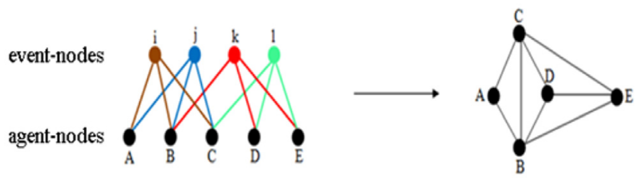
where $D(A)$, $D(B)$ and $D(v)$ are the degrees of nodes A, B and v, respectively in G. $N(A)$ and $N(B)$ are the sets of neighbors of A and B, respectively, in G. This measure is symmetric, it is based on the nearest neighbors and the similarity score between the nodes A and B will be assigned higher if A and B have more common neighbors with lower degrees. If they have no common neighbors, their similarity score equals to zero, which it is somewhat similar and very close to RA defined as [40],

$$RA_{AB} = \sum_{v\in N(A)\cap N(B)} \frac{1}{D(v)} \tag{5}$$

The method of weight [46] has fatal disadvantage. It makes default that the two nodes A and B in the network have the same effect on each other, however, it does not accord with the reality. i.e, each node is dependent on the other node to a different extent. Thus, we propose the reliability of an IdL in which $W'(A, B) \neq W'(B, A)$ and it is defined as follow.

(1)

	i	j	k	l
A	1	1	0	0
B	1	1	1	0
C	1	1	0	1
D	0	0	1	1
E	0	0	1	1



(2)

	A	B	C	D	E
A	0	2/9	2/9	0	0
B	2/6	0	2/9	1/6	1/6
C	2/6	2/9	0	1/6	1/6
D	0	1/9	1/9	0	2/6
E	0	1/9	1/9	2/6	0

Fig. 2. Example of bipartite graph G (1), and its bottom node-projection (2) based on [48].

3.2.1.2. Weight of induced link according to our proposition in this paper. Consider a bipartite network comprised of agents and events. By considering the agents as the holders of resources that flow through the network, we can infer a matrix containing only the agents from these between-mode connections or a matrix containing only the events. If the projection is done over agent nodes, the resources flow from the agents to the events in direct proportion to each agent's degree. Thus, if an agent has links to three events, one third of his resources flow from to each of these events, and so on. It is the same for the projection over the events nodes; if an event has ties to four agents, then a quarter of the resources momentarily parked on the event flow to each of the four agents who share ties to the event [47,48]. Mathematically:

$$W'(A,B) = \sum_{A\neq B,k:m_{Ak};m_{Bk}=1} \frac{1}{D(B) * D(k)} \tag{6}$$

where $W'(A,B)$ represents the strength of the tie that exists between the two nodes A and B as the result of their co-participation in k events in the original bipartite matrix, $m_{Ak} = 1$ means that a link exist between A and k, $m_{Bk} = 1$ means that a link exist between B and k. $D(B)$, $D(k)$ represent the degree of nodes B and k, respectively, in G. Fig. 2 also offers a graphic depiction of this process for a simple bipartite network containing five agents (A, B, C, D and E) and four events (i, j, k and l). Two nodes are structurally equivalent if they participate in the same events do by them as the nodes D and E in Fig. 2, they participate in two events k and l each other which are the same. In a bipartite network, a given node is structurally reliable by another node if the later participates in at least the same events as the former, and if the later node participates in additional events that do not involve the first one, the reliability of the first one by the second one remains unaffected, though the second node would then be only structurally reliable by the first node. for example, the node B participate on the same events i and j as node A, so node B is structurally reliable by node A and node B participates in the additional event k that do not involve the node A, so node A is partially reliable by node B.

Definition 5 (*Reliability of IdL (R_d)*). Reliability (R) is a node-level measure of relationship in bipartite network. It is based on a derivation of the equivalence between nodes. As the measure's name implies, Reliability more accurately scores the strength of the tie and the consistency between two nodes. The projection graph is non-oriented, hence, if we called the IdL (A, B) or (B, A) is the same one, but node B may be structurally reliable by node A, while node A is partially reliable by B, thus:

$W'(A,B) \neq W'(B,A)$. Therefore, the reliability of the IdL (A, B) is the summation of the strength of the ties that exist between (A and B) and the strength of the ties that exist between (B and A)divided by 2. More formally:

Suppose that $G_u = (U, E_u)$ is the projected graph of $G = (U, V, E)$. Let (A, B) be an IdL in G_u, then its reliability is defined as:

$$R_d(A,B) = \frac{W'(A,B) + W'(B,A)}{2} \tag{7}$$

where, $R_d(A,B)$ represents the consistency of agents A and B as the result of co-participation in events.

Once we have created a weighted projection graph, we identify ILs that induce links which already exist within G_u. Recall, this is the definition of an IL in Section 2. Then, we predict them based on the weight of their IdLs satisfying the condition that their weights are greater than or equal their reliabilities (evading the use of the threshold). More formally, we only keep ILs(TG, ta) where $TG \in$ Terrorist group and $ta \in$ Terrorist attack where Induced(TG, ta) is true, in which Induced(TG, ta) returns true if any link induced from the IL(TG, ta) has a weight $W \geq R_d$. Once we have the set of the ILs, we calculate their weights.

3.2.2. How to weight an internal link ?

3.2.2.1. Weight of internal link according to connectivity proposed in [46]. The connectivity (S) of the IL(A, x) in bipartite graph $G = (U, V, E)$ is the accumulation of the weight of each IdL covered by it. $S(A, x)$ is defined as [46]:

$$S(A,x) = \sum_{(A,B)\in\Gamma(A,x)} W(A,B) \tag{8}$$

In which, $W(A, B)$ is the weight of the IdL (A, B), and $\Gamma(A, x)$ is the set of the IdLs covered by the IL(A, x)

$$\Gamma(A,x) = \{(A,B) | B \in N_u(A) \cap N(x)\} \tag{9}$$

where, $N_u(A)$ are the neighbors of node A in G_u and $N(x)$ are the neighbors of node x in G. In [46] authors stated that a CNP (IL) covering higher-weight patterns (IdLs) would have a higher probability of conducting by linker, but they did not identify the values that are considered high as if they had used a threshold equal to zero. So, in this research, we defined the reliability of an IL to give meaning to the values of connectivity and specifying them in one hand and to avoid the use of the threshold [45] in the other hand.

3.2.2.2. Weight of IL according to reliability proposed in this paper.

Definition 6 (*Reliability of IL (R_I)*)**.** We defined the reliability of the IL (B, x) in bipartite graph $G = (U, V, E)$ as:

$$R_I(B,x) = \sum_{(A,B)\in\Gamma(A,x)} R_d(A,B) \tag{10}$$

In which, $R_d(A, B)$ is the reliability of the IdL containing A and B, and $\Gamma(A, x)$ is the set of the IdLs covered by the IL(A, x) according to formula (9) and satisfying condition in (11).

$$\Gamma(A,x) = \{(A,B) \in E_u | B \in N_u(A) \cap N(x), W(A,B) \geq R_d(A,B)\} \tag{11}$$

From formula (10), we can see that the reliability of the IL(A, x) is simply the summation of the reliabilities corresponding to the IdLs that it covered and which have weights superior or equal to their reliabilities. Obviously, the IL which has a higher probability to appear is the one covered a large number of important IdLs. The connectivity of the IL defined by formula (8) and its reliability defined by formula (10) reflect the probability of an IL to be connected. Therefore, they are used as the final score of link prediction. Building off of the ILs structure already established for link prediction, we saw the opportunity to predict ratings for these links as well. We predict an IL if and only if its connectivity is more or equal to its reliability. The framework of the LLPIL algorithm is described below.

LLPIL Algorithm (Latent Link Prediction based on Internal and Local Links)

Input: Bipartite graph $G = (U,V,E)$, such that U-part has the lowest number of nodes;
Output: Set of ILs: I with their weights (connectivity S and Reliability R);
Begin
(1) /* Construct the set of all IdLs according to formula (2) */
    $E_u = \emptyset$; $I = \emptyset$;
    For each node A in U **do**
        For each node x in $N(A)$**do**
            For each node B in $N(x)$**do**
                $E_u = E_u \cup \{(A, B)\}$;
            EndFor
        Endfor
    Endfor
(2) /* Calculate the weight of each IdL */
    For each edge (A,B) in E_u **do**
        Calculate the weight of the IdL(A ,B) according to formula (4);
        Calculate the reliability of the IdL(A,B) according to formula (7);
    Endfor

(3) /* Calculate S of each IL and its R_I and construct the set of ILs that fulfill the condition $S \geq R_I$;
For each node A in projected graph G_u **do**
For each neighbor B of A in projected graph G_u **do**
For each node $x \in N(B)$ in G **do**
if $(A,x) \notin E$ then
if $S(A,B) \geq R_d(A,B)$then
$S(A,x) = S(A,x)+W(A,B)$; $R_I(A,x) = R_I(A,x)+R_d(A,B)$;
if $S(A,x) \geq R_I(A,x)$ then
$I = I \cup \{(A, x)\}$;
Endif
Endif
Endif
Endfor
Endfor
Endfor
(4) Output $\{I\}$.
End

4. Experimental evaluation

In order to prove the performances of our proposed method and to show its adaptability and flexibility, we conduct considerable simulations on 3 benchmark data-sets: GTD, RAND Database of Worldwide Terrorism Incidents (RDWTI) and Southern Women (SW) network through comparisons with ILP and PLP methods in terms of accuracy metric named area under the receiver operating characteristic curve (AUC [48–50]). A machine powered by an Intel Core i5 with 6 GB RAM, running Windows 7 is used to carry out the experiments. The algorithm was coded using Java, and the results are visualized by Excel. To evaluate the accuracy of the results, we use a random partitions of training set (90%) and probe set (10%) for each data-set. Averages are taken to four decimal places; the entries corresponding to the highest AUCs scores among the methods are emphasized in black.

4.1. Matrix representation

We utilize the adjacency matrix M to visualize our graphs. Advantage of this representation with respect to the node link representation is the non-overlapping display of graph edges, and the readability of the graph especially for larger and denser graphs [5,27,33] like our case. In the case of GTD for example, the rows of M correspond to the terrorist groups (TG), and the columns correspond to the terrorist attacks (ta). If there is an edge between TG and ta, the corresponding cell has the value of 1, otherwise, it has the value of 0.

4.2. Experimental design

In this section, we discuss the implementation details of our method and the benchmark ones. For our method, nodal proximity was measured using formula (11), in which we relied on the S and R as weighting functions. We selected two representative methods described in Section 1; ILP [45] and PLP [46] as a benchmark. We benchmarked our method against ILP using two weighting functions, the commonly used in practice CN and we named ILP_CN. For CN, two nodes u and v, are more likely to have a link if they have many common neighbors [13,33]. Using CN, Newman [51] firstly rose up the idea of analyzing the structure of scientific collaboration network , showing a positive correlation between the number of common neighbors and the probability that two scientists will collaborate in the future. CN is defined as:

$$\sigma(u, v) = |N(u) \cap N(v)| \tag{12}$$

and against JC which we called ILP_JC. JC is the earliest local link prediction algorithm, which is proposed by Jaccard in 1901 [49,51]. It is defined as :

$$\gamma(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \tag{13}$$

For benchmarking our method against the PLP method, we use two cases. In the first one, we take all predicted links in account and we named PLP_1, and in the second one, we take only predicted links that achieve a connectivity superior or equal to a threshold in account and we called PLP_2. Table 2 summarizes the benchmark methods.

Table 2
Summary of benchmark methods.

Method	Weight function	Abbreviation
Internal Link Prediction	CN and threshold	ILP_CN
	JC and threshold	ILP_JC
Potential Link Prediction	S (all Values)	PLP_1
	S and threshold	PLP_2
Latent Link Prediction based on Internal and Local links	S and R_l	Our (LLPIL)

4.3. Experimental results and analysis

4.3.1. Experiments on GTD

GTD is a large semi structure dataset includes terrorist attacks over 4 decades, currently the most comprehensive and world's largest unclassified dataset available on terrorism incidents used for the experiment [33]. It contains information on over 140,000 terrorist attacks across the globe such as occurrence of the attack, modus-operandi of terrorist groups, their names, etc. For this article, we extract the two important columns: attacktype1 (terrorist attack) and gname (terrorist group). We generate heterogeneous graph by drawing an edge between these features. Table 3 contains the data selected as a case study, consisting of 50 rows containing the names of 50 terrorist groups and 6 columns containing the 6 types of their attacks. The value in the table is 1, if the group was involved in such attack, otherwise, it has 0.

Table 3
Fifty terrorist groups and their attacks selected from GTD.

Group-Terrorists	A	B	F	AA	HT	H
Tupamaros(Uruguay)	1	1	1	1	1	0
Armed Commandos of Liberation)	1	1	1	0	0	0
Individual	1	1	1	1	1	1
Popular Front for the Liberation of Palestine(PFLP)	1	1	1	1	1	1
Black Nationalists	1	1	1	1	1	0
Left-Wing Militants	1	1	1	1	0	1
White Extremists	1	1	1	1	0	0
Strikers	1	1	1	1	1	0
StudentRadicals	0	1	1	1	0	0
Fuerzas Armadas de Liberacion Nacional (FALN)	0	1	1	1	1	0
$23^r d$ of September Communist League	1	1	0	1	1	0
New Year's Gang	0	1	1	0	0	0
EritreanLiberation Front	0	1	0	1	1	1
$1^s t$ of May Group	0	1	1	0	1	0
Ku Klux Klan	1	1	1	1	1	0
Black Panthers	1	1	0	1	0	1
Japanese Red Army (JRA)	0	1	0	0	1	1
Montoneros (Argentina)	1	1	1	1	1	0
Jewish Defense League(JDL)	1	1	1	1	1	1
Taliban	1	1	1	1	1	1
Al_Qaida in the Arabian Peninsula(AQAP)	1	1	1	1	1	1
Al_Shabaab	1	1	1	1	1	1
Al_Qaida	1	1	0	1	1	1
Anarchists	0	1	1	1	0	0
Popular Revolutionary Vanguard(VPR)	0	1	0	0	1	0
Irish Republican Army(IRA)	1	1	1	1	1	0
Baloch Republican Army (BRA)	1	1	1	1	1	1
United Liberation Front of Assam(ULFA)	1	1	1	1	1	0
Black September	1	1	1	1	1	1
al_Fatah	1	1	1	1	1	1
Basque Fatherland and Freedom (ETA)	1	1	1	1	1	0
Misrata Brigades	0	1	0	1	1	1
Haqqani Network	1	1	0	1	1	0
Huthis	1	1	1	1	1	1
Maoists	1	1	1	1	1	0
Black Liberation Army	1	1	0	1	1	1
Misrata Brigades	0	1	0	1	1	1
Kurdistan Workers'Party (PKK)	1	1	1	1	1	1
Baloch Liberation Front (BLF)	1	1	1	1	1	1
Tehrik_i_Taliban Pakistan (TTP)	1	1	1	1	1	1
New People's Army (NPA)	1	1	1	1	1	0

(*continued on next page*)

Table 3 (*continued*).

Group-Terrorists	A	B	F	AA	HT	H
Boko Haram	1	1	1	1	1	1
Croatian Nationalists	1	1	1	1	1	1
Ulster Volunteer Force (UVF)	1	1	1	1	0	0
Separatists	1	1	1	1	1	0
People's Liberation Army (India)	1	1	0	1	1	0
Jamaat_ul_Ahrar	1	1	0	1	1	0
Paraguayan People's Army (EPP)	1	1	1	1	1	0
Fulani Militants	1	0	1	1	1	1
M_19 (Movement of April 19)	1	1	1	1	1	1

Abbreviation of attacks' names are: **A**: Assassination, **B**: Bombing/Explosion, **F**: Facility Infrastructure Attack, **AA**: Armed Assault, **HT**: Hostage Taking (Kidnapping), **H**: Hijacking.

We create our bipartite graph G = (Terrorist groups, Terrorist attacks), we then create its projection graph over the terrorist attacks. For the account, Table 3 has been represented by a matrix according to the formula (1). Table 4 shows the main topological features of the network in each test.

Table 4
Main topological features of the GTD bipartite network.

Test	Number of terrorist nodes	Number of attack nodes	N_E	N_E^P	N_E^T
1	10	6	46	5	41
2	10	6	46	5	41
3	30	6	94	9	85
4	30	6	94	9	85
5	50	6	239	24	215
6	50	6	239	24	215

Where, N_E means number of links, N_E^P: number of probe links and N_E^T: number of training links.

Accuracies of the algorithms subject to link prediction are represented in Table 5. Each number is obtained by 2 implementations with independently random partitions of training set (90%) and probe set (10%).

Table 5
Comparison of AUCs by our method LLP_*LI* and other methods on the GTD.

Test	ILP_CN $\tau = 3$	ILP_CN $\tau = 5$	ILP_JC $\tau = 1/3$	ILP_JC $\tau = 2$	PLP_1 S	PLP_2 $S \geq 0.5$	OUR
1	0.9286	0.5	0.8571	0.6	**0.9786**	0.9571	**0.9786**
2	0.8214	0.5	0.8429	0.5	0.8857	0.8222	**0.9**
3	0.8833	0.6472	0.8514	0.7111	**0.9125**	0.8333	**0.9125**
4	0.8542	0.7333	**0.8736**	0.7757	0.8653	0.7778	**0.8736**
5	0.7924	0.7097	0.8169	0.6880	**0.8344**	0.7113	**0.8344**
6	0.7886	0.5721	0.7862	0.6760	0.8125	0.6827	**0.8180**
Average	0.8447	0.6104	0.8380	0.6585	0.8815	0.7974	**0.8862**
Average%	84%	61%	84%	66%	88%	80%	**89%**

As shown in Table 5, our method achieves higher accuracy than ILP and PLP and it substantially outperforms each benchmark in almost prediction test. Therefore, we prove the positive contributions of the combined advantages of ILP and PLP algorithms for link prediction. For example, on the 2nd test, OUR algorithm obtains the AUC score 0.9, whereas ILP_CN ($\tau = 3$), ILP_CN ($\tau = 5$), ILP_JC ($\tau = 1/3$), ILP_JC ($\tau = 2$), PLP_1 and PLP_2 achieve 0.8214, 0.5, 0.8429, 0.5, 0.8857, and 0.8222 respectively. This shows that the LLPIL algorithm can obtain higher-quality results than the other methods. Note that high values of UAC obtained by ILP_CN and ILP_JC correspond to small values of the threshold τ for ILs prediction. Similarly, if we take all predicted links in the count, then the PLP has high values of UAC (PLP_1), whereas, if we considered that the high values of connectivity are those superior or equal to 0.5 than PLP has low values of UAC (PLP_2). Our method achieves the highest predictive accuracy (89%) among the benchmark methods, although there is no significant difference with of PLP_1 88%, but LLPIL relies on two weighting functions to predict potential links what distinguishes it and makes it more precise. As a conclusion, our method can obtain results with high quality similar to those by PLP in the same computation time with more efficiency and significance. As a result, links recommended by our method generally have higher utilities than those recommended by benchmark methods.

To further compare the results of our method with PLP_1 [46], we added Fig. 3 and discussed it.

Note: the notation (*ta1*, *TG9*) in Fig. 3 means that the group number nine will do an attack of type 1 and so on.

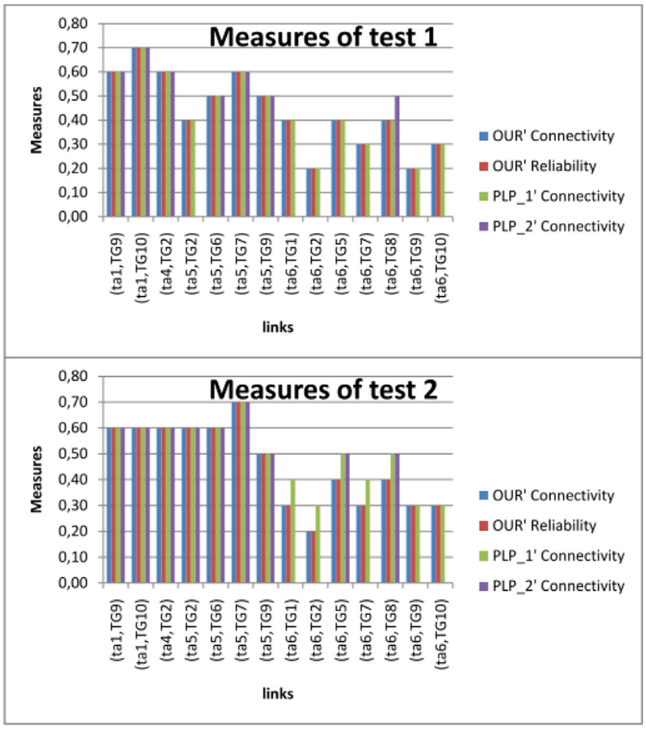


Fig. 3. Comparison the values of the measures obtained by PLP_1, PLP_2 and our method.

As shown in Fig. 3, the connectivity in our method is equal to the reliability which is proposed to replace the use of the threshold just as in ILP [45] and to make the values of connectivity in PLP [46] more significance, we see that the link (*ta6*, *TG2*) in the case of Measures of test 1 is a predicted link by the method PLP_1 although it has low connectivity (0.2), which it is seems as if there is a threshold equal to zero. While in our case it is a predicted link since that its connectivity equal to its reliability what makes our method more expressive, appropriate and acceptable. Also, we can see that some values of connectivity in PLP_1 such as that of the 5 links ((*ta6*, *TG1*), (*ta6*, *TG2*), (*ta6*, *TG5*),(*ta6*, *TG7*) and (*ta6*, *TG8*)) in the case of Measures of test 2 have diminished Compared with their connectivities in our method and this because that some of the IdLs covered by them have connectivities inferior then their reliabilities (the condition we added to identify links that have a high probability of occurring). Thus, the values of connectivity which are taken into account in our case are only those that are equal to or greater than the values of the reliability and not all as in PLP_1 [46] (this explains the low number of the predicted links when we took only the values of the connection which are equal to or superior then 0.5 as in PLP_2 [46] in the case of Measures of test 1, such as the 7 links (*ta5*, *TG2*), (*ta6*, *TG1*),((*ta6*, *TG2*),(*ta6*, *TG5*),(*ta6*, *TG7*),(*ta6*, *TG9*),(and (*ta6*, *TG10*))).

4.3.2. Experiments on RDWTI

As the second example, we test the RDWTI [52,53], a compilation of data from 1968 until 2009, it is managed by the RDWTI Corporation 2016. With over 40,000 incidents of terrorism coded and detailed, RAND provides comprehensive information on international and domestic terrorism. It is a fully searchable and interactive database, it is available in https://www.rand.org/nsrd/projects/terrorism-incidents/download.html. In this paper, the data was selected from 9 February 1968 until 10 December 1974 where the two taken attributes for the analysis are perpetrator and weapon. 5 different labels are used to represent different type of weapon, these labels are Firearms, Explosives, Fire or Firebomb, Remote-detonated explosive and Knives & sharp objects. If the perpetrator was used the specified weapon then a link relies between them in the bipartite graph. The main topological features of the RDWTI selected data is represented in Table 6.

Table 6
Main topological features of the RDWTI bipartite network.

Number of perpetrators nodes	Number of weapons nodes	N_E	N_E^P	N_E^T
100	5	189	19	170

Table 7 presents the AUC scores of 5 tests achieved by the algorithms in the 6 cases; ILP_CN ($\tau = 2$), ILP_CN ($\tau = 4$), ILP_JC ($\tau = 0.1$), PLP_1, PLP_2 (S $\geq 1/3$) and OUR. From the table, we can see that LLPIL obtains the highest AUC scores among the algorithms in the almost tests. LLPIL has lower AUC score than PLP_1 in the 2nd test but the difference was insignificant. LLPIL achieves the highest average (90%), significantly better than PLP_1 (87%) and the other algorithms. This shows that the LLPIL algorithm can achieves higher-quality results than the other algorithms.

Table 7
Comparison of AUCs by our method LLPIL and other methods on the RDWTI dataset.

Test	ILP_CN $\tau = 2$	ILP_CN $\tau = 4$	ILP_JC $\tau = 0.1$	PLP_1 S	PLP_2 $S \geq 1/3$	OUR
1	0.579	0.5	0.1289	**0.8367**	0.7526	**0.8367**
2	0.5526	0.5	0.0367	**0.9101**	0.763	0.8421
3	0.7895	0.5	0.4729	0.8246	0.8246	**0.9298**
4	**0.9474**	0.5	0.8718	**0.9474**	**0.9474**	**0.9474**
5	0.5526	0.5	0.0572	0.8262	0.7156	**0.9269**
Average	0.6842	0.5	0.3135	0.869	8006	**0.8966**
Average%	68%	50%	31%	87%	80%	**90%**

4.3.3. Experiments on southern women network

We have compared our proposed algorithm LLPIL with the PLP one cited in [46], for this reason, we have select a data set used in [46] as the third example to more validate the advantages of our proposed method. The data set called the Southern Women (SW). It was collected by Davis et al. around Mississippi during the 1930s in a wide-ranging study of class and race in the Deep South. Because its community structure is known, this dataset has been widely used as a benchmark for social network analyzers [46,54]. SW depicts the participation of 18 women (bottom_*nodes*) in 14 social events (top_*nodes*). There are 89 edges linking women nodes and event nodes. A visualization of the bipartite graph for the SW Study is shown in Fig. 4. In the network, each edge indicates that the woman attended the corresponding event.

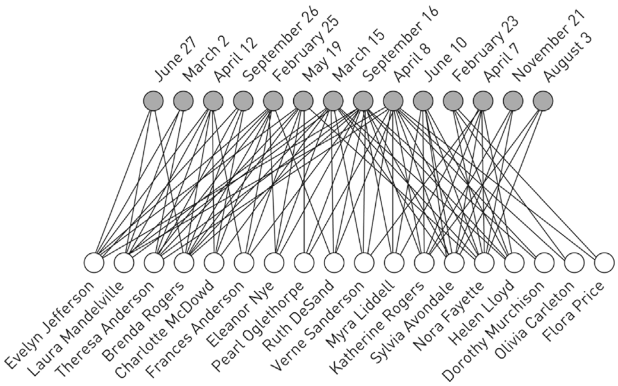


Fig. 4. Southern Women bipartite network.

Table 8 shows the main topological features of SW network.

Table 8
Main topological features of the Southern Women bipartite network.

Number of women nodes	Number of events nodes	N_E	N_E^P	N_E^T
18	14	89	9	80

Table 9
Comparison of AUCs by our method LLPIL and other methods on the Southern Women dataset.

Test	ILP_CN $\tau = 5$	ILP_CN $\tau = 6$	ILP_JC $\tau = 1/3$	ILP_JC $\tau = 0.5$	PLP_1 S	PLP_2 $S \geq 0.5$	OUR
1	0.5556	0.5	0.867	0.7104	**0.9124**	0.6667	**0.9124**
2	0.7417	0.5	0.8116	0.7343	**0.9468**	0.5556	**0.9468**
3	0.6918	0.5	0.8767	0.7350	**0.9464**	0.6889	**0.9464**
4	0.6401	0.5556	0.7669	0.7350	**0.9509**	0.7870	**0.9509**
5	0.7068	0.5	0.8124	0.7728	0.933	0.6667	**0.9425**
Average	0.6672	0.5111	0.8269	0.7375	**0.9379**	0.673	**0.9398**
Average%	67%	51%	83%	67%	**94%**	67%	**94%**

Table 9 presents the AUC scores of 5 tests achieved by different algorithms. From the table, we can see that LLPIL obtains the highest AUC scores among the algorithms in all 5 tests. Although PLP_1 achieved the same height result as our method (94%), our method was better in test 5 with a score of 0.9425 than PLP_1 (0.933), resulting in an average of 0.9398 that exceeds the average achieved by PLP_1, which is 0.9379. This shows that the LLPIL algorithm can achieve higher-quality results than the other methods.

Based on the AUC scores of the results on the three datasets shown in Tables 5, 7 and 9, we conclude that our method is much better than the other methods, as it is not related and it is not limited to any condition, unlike ILP and PLP, which require a small threshold to achieve good results.

4.4. Time efficiency

Fig. 5 shows the experimental results in terms of running time required by the algorithms. As PLP, LLPIL detects the adept links only within the set of ILs and its time complexity algorithm is linear to the lowest number of nodes in the two parts. Thus and according to the results, LLPIL can obtain a better prediction result with low time complexity.

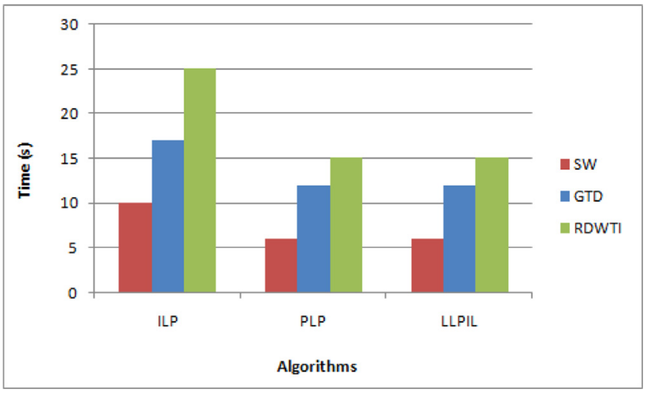


Fig. 5. Computation time of the algorithms.

As a conclusion, our method combines the advantages of both methods PLP and ILP. It is easy in the implementation like ILP and has results with low time complexity as PLP and in a standard situation.

5. Conclusion

In many real world networks, the links and node attributes are often partially observable. In this paper, we have proposed an innovative algorithm in bipartite graph to predict hidden and missing links on the basis of the topological analysis of internal and local links. In order to further increase prediction accuracy, we introduced new weight function called reliability to predict only sufficient and appropriate links. Reliability replaces the threshold such in [45] on the one hand and specifies acceptable values of connectivity in [46] on the other. The proposed algorithm was applied in crime prevention as a case study. Our simulations demonstrated the advantageous performance of our proposed approach over GTD, RDWTI and SW. For GTD, the results show that the proposed algorithm can achieve higher speed and superior quality prediction, it reduce the complexity and it improve prediction accuracy (89%) which is greater than 80,41% [55] and 81,12% [56]. From the various simulations observations, it is evident that the proposed approach is the efficient one for prediction.

Some directions of future research appear attractive: We plan to adapt our approach in order to handle temporal issues, to merge other metrics like the structural consistency and structural perturbation method proposed in [14] which is considered more accurate and robust or to use other metrics like the triadic closure structure.

CRediT authorship contribution statement

Nora Assouli: Conceived of the presented idea, Developed the theory, Designed and performed the computations, Analyzed data, Planned and carried out the simulations, Interpreted the results, Draft and wrote the manuscript with support from Khelifa Benahmed. **Khelifa Benahmed:** Proposing of the idea for the proposed approach; in terms of graph theory, Interpretation of the results, Review experiments, Reviewing the form and content of the paper, Supervised the findings. **Brahim Gasbaoui:** Manuscript review, Provided critical feedback, Helped shape the research and provided guidance and advice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.-H. Liu, Y.-X. Zhu, T. Zhou, Improving personalized link prediction by hybrid diffusion, Physica A 447 (2016) 199–207, http://dx.doi.org/10.1016/j.physa.2015.12.036.
[2] P.D. Meo, E. Ferrara, G. Fiumara, A. Provetti, Enhancing community detection using a network weighting strategy, Inform. Sci. 222 (2013) 648–668, http://dx.doi.org/10.1016/j.ins.2012.08.001.
[3] Z. Wu, Y. Lin, J. Wang, S. Gregory, Link prediction with node clustering coefficient, Physica A 452 (2016) 1–8, http://dx.doi.org/10.1016/j.physa.2016.01.038.
[4] A. Andalib, S.M. Babamir, A class-based link prediction using Distance Dependent Chinese Restaurant Process, Physica A 456 (2016) 204–214, http://dx.doi.org/10.1016/j.physa.2016.03.067.
[5] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (2011) 1150–1170, http://dx.doi.org/10.1016/j.physa.2010.11.027.
[6] A. Wahid-Ul-Ashraf, M. Budka, K. Musial, How to predict social relationships–Physics-inspired approach to link prediction, Physica A 523 (2019) 1110–1129, http://dx.doi.org/10.1016/j.physa.2019.04.246.
[7] J. chao Li, D. ling Zhao, B.-F. Ge, K.-W. Yang, Y.-W. Chen, A link prediction method for heterogeneous networks based on BP neural network, Phyisca A 495 (2018) 1–17, http://dx.doi.org/10.1016/j.physa.2017.12.018.
[8] Z. Wu, Y. Lin, Y. Zhao, H. Yan, Improving local clustering based top-L link prediction methods via asymmetric link clustering information, Physica A 492 (2018) 1859–1874, http://dx.doi.org/10.1016/j.physa.2017.11.103.
[9] Z. Wang, J. Liang, R. Li, Exploiting user-to-user topic inclusion degree for link prediction in social-information networks, Expert Syst. Appl. 108 (2018) 143–158, http://dx.doi.org/10.1016/j.eswa.2018.04.034.
[10] B. Moradabadi, M.R. Meybodi, Link prediction based on temporal similarity metrics using continuous action set learning automata, Physica A 460 (2016) 361–373, http://dx.doi.org/10.1016/j.physa.2016.03.102.
[11] N.M. Ahmed, C. Ling, W. Yulong, L. Bin, L. Yun, L. Wei, Sampling-based algorithm for link prediction in temporal networks, Inform. Sci. 374 (2016) 1–14, http://dx.doi.org/10.1016/j.ins.2016.09.029.
[12] S.D. Moor, C. Vandeviver, T.V. Beken, Assessing the missing data problem in criminal network analysis using forensic DNA data, Social Networks 61 (2020) 99–106, http://dx.doi.org/10.1016/j.socnet.2019.09.003.
[13] A.M. Filho, M. da Silva, G.F. Zebende, Autocorrelation and cross-correlation in time series of homicide and attempted homicide, Physica A 400 (2014) 12–19, http://dx.doi.org/10.1016/j.physa.2014.01.015.
[14] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, H.E. Stanley, Toward link predictability of complex networks, Proc. Natl. Acad. Sci. 112 (8) (2015) 2325–2330, http://dx.doi.org/10.1073/pnas.1424644112.
[15] P. Pei, B. Liu, L. Jiao, Link prediction in complex networks based on an information allocation index, Physica A 470 (2017) 1–11, http://dx.doi.org/10.1016/j.physa.2016.11.069.
[16] S. Daminelli, J.M. Thomas, C. Durán, C.V. Cannistraci, Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks, New J. Phys. 17 (11) (2015) 113037, http://dx.doi.org/10.1088/1367-2630/17/11/113037.
[17] W. Yuan, K. He, D. Guan, L. Zhou, C. Li, Graph kernel based link prediction for signed social networks, Inf. Fusion 46 (2019) 1–10, http://dx.doi.org/10.1016/j.inffus.2018.04.004.
[18] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. 98 (2) (2001) 404–409, http://dx.doi.org/10.1073/pnas.021544898.
[19] M.E.J. Newman, Coauthorship networks and patterns of scientific collaboration, Proc. Natl. Acad. Sci. 101 (1) (2004) 5200–5205, http://dx.doi.org/10.1073/pnas.0307545100.
[20] M.E.J. Newman, Scientific collaboration networks. I. Network construction and fundamental results, Phys. Rev. E 64 (016131) (2001) 1–8, http://dx.doi.org/10.1103/PhysRevE.64.016131.
[21] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442, http://dx.doi.org/10.1038/30918.
[22] R. Lambiotte, M. Ausloos, Uncovering collective listening habits and music genres in bipartite networks, Phys. Rev. E 72 (6) (2005) 066107, http://dx.doi.org/10.1103/PhysRevE.72.066107.
[23] S. Li, X. Song, H. Lu, L. Zeng, M. Shi, F. Liu, Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm, Expert Syst. Appl. 139 (112839) (2020) 1–11, http://dx.doi.org/10.1016/j.eswa.2019.112839.
[24] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender systems, Phys. Rep. 519 (1) (2012) 1–49, http://dx.doi.org/10.1016/j.physrep.2012.02.006.
[25] D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks, J. Am. Soc. Inf. Sci. Technol. 58 (2007) 1019–1031, http://dx.doi.org/10.1002/asi.20591.
[26] L. Lü, C.-H. Jin, T. Zhou, Similarity index based on local paths for link prediction of complex networks, Phys. Rev. E 80 (2009) 1–9, http://dx.doi.org/10.1103/PhysRevE.80.046122.
[27] M. Waniek, K. Zhou, Y. Vorobeychik, E. Moro, T.P. Michalak, T. Rahwan, How to hide one's relationships from link prediction algorithms, Sci. Rep. 9 (12208) (2019) 1–10, http://dx.doi.org/10.1038/s41598-019-48583-6.
[28] V. Martínez, F. Berzal, J.-C. Cubero, A survey of link prediction in complex networks, ACM Comput. Surv. 49 (4) (2016) 1–33, http://dx.doi.org/10.1145/3012704.
[29] S. Saha, A. Kurian, H. Aladi, A. Basu, Future terrorist attack prediction using machine learning techniques, 2017, pp. 1–7, http://dx.doi.org/10.13140/RG.2.2.17157.96488.
[30] S.-H. Yoon, S.-W. Kim, S. Park, C-Rank: A link-based similarity measure for scientific literature databases, Inform. Sci. 326 (2016) 25–40, http://dx.doi.org/10.1016/j.ins.2015.07.036.
[31] L. Yao, L. Wang, L. Pan, K. Yao, Link prediction based on common-neighbors for dynamic social network, Procedia Comput. Sci. 83 (2016) 82–89, http://dx.doi.org/10.1016/j.procs.2016.04.102.
[32] P. Wang, B. Xu, Y. Wu, X. Zhou, Link prediction in social networks: the state-of-the-art, Sci. China Inf. Sci. 58 (2015) 1–38, http://dx.doi.org/10.1007/s11432-014-5237-y.
[33] G.M. Tolan, O.S. Soliman, An experimental study of classification algorithms for terrorism prediction, Int. J. Knowl. Eng. 1 (2) (2015) 107–112, http://dx.doi.org/10.7763/IJKE.2015.V1.18.
[34] M.N. Mohammad, N. Sulaiman, O.A. Muhsin, A novel intrusion detection system by using intelligent data mining in weka environment, Procedia Comput. Sci. 3 (2011) 1237–1242, http://dx.doi.org/10.1016/j.procs.2010.12.198.
[35] C. Stadtfeld, K. Takács, A. Vörös, The emergence and stability of groups in social networks, Social Networks 60 (2020) 129–145, http://dx.doi.org/10.1016/j.socnet.2019.10.008.
[36] Z. Zhang, J. Wen, L. Sun, Q. Deng, S. Su, P. Yao, Efficient incremental dynamic link prediction algorithms in social network, Knowl.-Based Syst. 132 (2017) 226–235, http://dx.doi.org/10.1016/j.knosys.2017.06.035.
[37] R. Pech, D. Hao, Y.-L. Lee, Y. Yuan, T. Zhou, Link prediction via linear optimization, Physica A 528 (121319) (2019) 1–29, http://dx.doi.org/10.1016/j.physa.2019.121319.

[38] J.F. Sánchez-Rada, C.A. Iglesias, Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison, Inf. Fusion 52 (2019) 344–356, http://dx.doi.org/10.1016/j.inffus.2019.05.003.
[39] L.A. Adamic, E. Adar, Friends and neighbors on the Web, Social Networks 25 (2003) 211–230, http://dx.doi.org/10.1016/S0378-8733(03)00009-1.
[40] T. Zhou, L. Lü, Y.-C. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630, http://dx.doi.org/10.1140/epjb/e2009-00335-8.
[41] J. Kunegis, E.W.D. Luca, S. Albayrak, The link prediction problem in bipartite networks, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2010, pp. 380–389, http://dx.doi.org/10.1007/978-3-642-14049-5_39.
[42] Z.N.H. Jeong, A. Barabàsi, Measuring preferential attachment for evolving networks, Europhys. Lett. 61 (4) (2001) 1–4, http://dx.doi.org/10.1103/PhysRevE.64.016131.
[43] M.E.J. Newman, Clustering and preferential attachment in growing networks, Phys. Rev. E 64 (2001) 025102, http://dx.doi.org/10.1103/PhysRevE.64.025102.
[44] A. Anil, D. Kumar, S. Sharma, R. Singha, R. Sarmah, N. Bhattacharya, S.R. Singh, Link prediction using social network analysis over heterogeneous terrorist network, in: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom, SmartCity, 2015, pp. 267–272, http://dx.doi.org/10.1109/SmartCity.2015.82.
[45] O. Allali, C. Magnien, M. Latapy, Link prediction in bipartite graphs using internal links and weighted projection, in: 2011 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS, 2011, pp. 936–941, http://dx.doi.org/10.1109/INFCOMW.2011.5928947.
[46] M. Gao, L. Chen, B. Li, Y. Li, W. Liu, Y. cheng Xu, Projection-based link prediction in a bipartite network, Inform. Sci. 376 (2017) 158–171, http://dx.doi.org/10.1016/j.ins.2016.10.015.
[47] L.M. Gerdes, Dependency centrality from bipartite social networks, Connections 34 (2014) 14–28, http://dx.doi.org/10.17266/34.1.2.
[48] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E 76 (2007) 1–7, http://dx.doi.org/10.1103/PhysRevE.76.046115.
[49] K. ke Shang, T. chen Li, M. Small, D. Burton, Y. Wang, Link prediction for tree-like networks, Chaos 29 (061103) (2019) 1–9, http://dx.doi.org/10.1063/1.5107440.
[50] K. ke Shang, M. Small, W. sheng Yan, Fitness networks for real world systems via modified preferential attachment, Physica A 474 (2017) 49–60, http://dx.doi.org/10.1016/j.physa.2017.01.066.
[51] M.E.J. Newman, Clustering and preferential attachment in growing networks, Phys. Rev. E 64 (025102) (2001) 1–13, http://dx.doi.org/10.1103/PhysRevE.64.025102.
[52] R. Hagan, Filling Terrorism Gaps: VEOs Evaluating Databases and Applying Risk Terrain Modeling to Terrorism, Tech. Rep. PNNL-25727, 2016, http://dx.doi.org/10.2172/1342342, August.
[53] L. Dugan, G. LaFre, K. Cragin, A. Kasupski, Building and Analyzing a Comprehensive Open Source Data Base on Global Terrorist Events, Tech. Rep., RAND Corporation, 2008.
[54] Networks an Introduction, OUP Oxford, 2010.
[55] A. Sachan, D. Roy, TGPM: terrorist group prediction model for counter terrorism, Int. J. Comput. Appl. 44 (10) (2012) 49–52, http://dx.doi.org/10.5120/6303-8516.
[56] A. Sachan, E-TGPS: enhanced terrorist group prediction system for counter terrorism, Int. J. Comput. Appl. 117 (24) (2015) 24–28, http://dx.doi.org/10.5120/20704-3633.