RESEARCH
www.commprac.com
ISSN 1462 2815

LINK PREDICTION IN SOCIAL NETWORKS: A SURVEY

Abraham Domínguez-Melquíades [1], Jorge Pineda-Acosta [2] and Rubén Pérez-Ramos [3]*

[1,2,3] Higher School of Mechanical and Electrical Engineering, National Polytechnic Institute, Av. Juan de Dios Bátiz, esq. Miguel Othón de Mendizábal, Col.Nueva Industrial Vallejo, Gustavo A. Madero, 07700, CDMX, Mexico.
*Corresponding Author Email: alumno.cic.2@gmail.com

DOI: 10.5281/zenodo. 11666524

Abstract

Artificial intelligence has become highly significant in virtually all activities important to human beings. Nowadays, it is common to see that this important branch of science has become a topic of conversation even among groups of people who are not well-versed in science or technology. In this context, social networks and internet applications are increasingly necessary in all types of human interactions. Therefore, the search for solutions to predict possible future links (incomplete or missing) in social networks has become an active branch in scientific research. That is what this proposal is about. Specifically, our research refers to a systematic review of the most recent information available in the state of the art, regarding the way in which it is possible to improve various processes in social networks through link prediction. Considering the importance of being able to know the trends of processes and elements immersed in social networks, it is gratifying to observe that the results of our research promise to positively impact the generation of valuable information in these contexts.

Keywords: Artificial Intelligence, Social Networks, Link Prediction, Machine Learning, Internet.

1. INTRODUCTION

Technology attracts an enormous amount of interest. New technologies emerge every day in the world, causing radical changes in the lifestyles of human beings. A significant amount of material has been written about the ways technology is currently changing our lives, about the possibilities of its development in the future and of the hopes that it will revolutionize the way people learn and exchange information [1].

Technology applied has permeated the world so much that specialists now fiercely argue about its role: some praise emerging technologies as a triumph of modernity which helps society, while others dismiss technology as a misused innovation. Setting aside this debate and considering the enormous influence of emergent technologies applied to various environments, it is pertinent to recognize the speed at which social networks currently influence the world. This influence is clearly manifested in the attitudes and the unprecedented habits acquired by users. Indeed, educational systems have not only kept up in this regard, but it is necessary to appreciate the active role that social networks have taken in engineering educational processes [2].

Let's consider a community of members of some specific and well-identified nature. If the members interact reciprocally with each other, this community can be characterized as a social network. The spontaneous cohesion that these reciprocal interactions produce among the community members, at the same time, prevents access to potential external agents interested in the community. Taking advantage of the benefits and facilities of information and communication technologies, members of a social network can exchange resources, for example: advice, news, general or specific information on a subject of mutual interest, job opportunities, and the corresponding feedback. An advantage of social networks seen as social networks is that it is not required to be geographically close for different people to belong to the network; even more, the simultaneous connection of two or more members is not required [3].

Social networks can be modelled as nodes, consisting of people, linked by edges as connections between those individuals, in other words, a graph. The influence a person has on a social network is determined by his connectivity: the number of people to whom he or she is linked and the number of paths they form throughout the network. Assessing the impact of an individual within the social network has several applications: for example, data mining can extract patterns to determine where cliques are forming, measuring reputations and monitoring social happenings [4].

Taking into account that the importance that social networks have acquired in their specific form of social networks is undeniable, the authors of the present paper consider that searching for answers to the questions regarding usage habits in social networking like those mentioned above is important.

We are convinced that there exists a need for researching the role of social networking. The natural environment is the application of this type of research in a subset of the social networks, which are the social networks. We have selected for our research an application area that consists mainly of published works in computer science topics. Considering the relevance of being able to know the trends of processes and elements immersed in social networks, seen as social networks, it is gratifying to observe that the results of our research promise to positively impact the generation of valuable information in these contexts.

The rest of the paper is organized as follows: some basic concepts about social networks are described in Section 2, while link mining and link prediction in social networks are presented in Sections 3 and 4, respectively. Section 5 addresses the most relevant ideas and events regarding the genesis and evolution of social networks, while Section 6 contains descriptions of some state-of-the-art research work related to link prediction in social networks, which are applied in some areas of interest to human beings. Subsection 6.1 contains methods from several scientific disciplines, while in Subsection 6.2, we emphasize the relevant methods that come from artificial intelligence, machine learning and related fields. Conclusions are presented in Section 7. Finally, CRediT authorship contribution statement, Declaration of competing interest, and bibliographical references are included at the end.

2. SOCIAL NETWORKS: BASIC CONCEPTS

This section is dedicated to describing the social networks. Members of a social network interact with each other. The need for effective communication mechanisms among peers is covered, in some way, by the social networks. The promotion and cultivation of topics about which some cores of social networks are manifested are guaranteed in short times. Contact by affinity in social networks causes the socializing aspect to become a valued capital, whose intrinsic value permeates and positively affects all members of the network [4] [5].

Graphs are an adequate tool to model social networks; nodes represent people and edges are connections between individuals. Personal computers and mobile devices such as smartphones and RFID or Bluetooth tags allow connection to social networks which offer interaction and communication [6]. Fig. 1 represents the concept of a social

ISSN 1462 2815

network. The yellow squares represent the nodes, while the red segments represent the links between nodes.

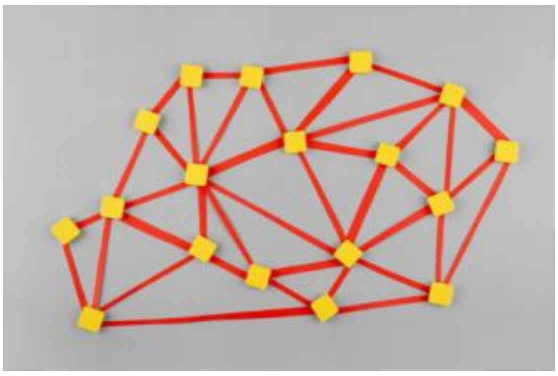

Figure 1: Concept of a social network (www.freepik.com). The nodes are represented by yellow squares, while the links are represented by red segments

Social networks can be characterized by the diameter, distribution and degree of its nodes or by the mean connectivity of the nodes, leading to the identification of various roles in the nodes. An important task in social networks is finding key actors in function of their connectivity, the number of contacts they have and the paths that traverse each node [7].

The influence a person has on a social network is determined by his connectivity: the number of people to whom he or she is linked and the number of paths they form throughout the network. Assessing the impact of an individual within the social network has several applications: for example, data mining can extract patterns to determine where cliques are forming, measuring reputations, and monitoring social happenings [4].

As social networks develop, thanks in great part to the advancement of world wide web-related technologies, groups of researchers worldwide have entered this field, taking measurements, and collecting data. These efforts are geared to applying the scientific method to the study and understanding of the relationships between individuals immersed in social networks [8].

Formally, a social network is defined as a finite set of social actors (nodes). These nodes, which are no more than the network's members, are connected by one or more kinds of relationships (edges), whose characterization is of great help to researchers. Concepts exist which substantiate the measurements and characterizations of networks, such as ties, density, centrality, cliques and other relevant features [3] [9].

2.1. Links

In a network's representation as a graph, links (also known as edges or ties) are connections between two or more of the network's nodes. Links can be directed or undirected. Directed links can also be unidirectional or reciprocal, and they can be modeled as binary relationships or can be weighted so that some ties are stronger, and others are weaker. A clear example of a directed link in the educational context is when one of the members (a student) sends his work to another member (the teacher). An example of a reciprocal directed link is that of two students sharing ideas for solving a problem related to a course that both are taking. An example of weighted links would be the frequency with which two specific students exchange ideas; in some cases, this weight would be almost zero, when those two seldom interact.

2.2. Density

This concept describes the general linkage level in a social network's graph representation. In an informal way we can say density measures how far is a network's graph from being complete; in quantitative terms, density is the cardinality of the link set divided by the cardinality of the set of vertices of a complete graph with the same number of nodes.

2.3. Path, length and distance

Nodes or actors of a social network can be directly connected by edges or ties, or indirectly by a sequence of links. A trail is a sequence of lines in a graph, and a path is a trail in which each point and each line are different. The number of lines that make up a path is defined as its length. Given two nodes in a social network, the distance between them is defined as the length of the shortest path connecting both nodes.

2.4. Centrality

The most active nodes or actors are detected by the measure of the concept called centrality; meaning the nodes which relate to other members with the most intensity and reach. There are three important measurements regarding the concept of centrality: a) Degree centrality: means popularity or intense activity; it is the sum of all the nodes connected to a specific node; b) Between-ness centrality: the capacity of a given node of connecting pairs of other nodes; that is, in a way it is the measure of the node's potential to function as a controller of the resource flow in the network. This measure indicates how powerful is that node in the network as a whole; and c) Closeness centrality: measures efficiency and independence of a specific node. If path lengths from a certain node to others are lesser that two, it is said that that node can reach any other node by itself and does not depend on other nodes to reach any desired node.

2.5. Clique

Given a graph which represents a social network, a clique is a subgraph in which any of its nodes is directly connected to any other node in the subgraph.

In the proposal presented in this survey, these concepts are used while emphasizing the concept of link within the structure of social networks.

3. LINK MINING IN SOCIAL NETWORKS

Links in a social network possess a conceptual wealth such that they exhibit relevant properties for researchers; among these properties we can find the category, importance or rank of the nodes and of the links between them when they are considered as objects of study [7].

It should be noted that in social networks, seen as social networks, phenomena of interest to scientific researchers sometimes occur in topics related to the areas of application of networks. One of the most emblematic cases occurs when it is not possible to observe all the links; from here, it could be interesting to achieve the prediction of the existence of links between specific nodes of the network; and if so, it could be useful to know the weights of those potential links. In other cases of dynamic social networks where links evolve over time, it could be of interest to predict if a specified link might materialize at some future time, near or distant [9].

According to [10], the expression "Link Mining" has been coined to represent all processes, techniques, and data mining models that explicitly consider the links of the nodes in a social network as objects of study. Typically, there are eight tasks in link mining that focus on objects, links, and graphs:

3.1. Link-based object ranking

Based on exploiting the link structure in a graph representing a social network with the goal of assigning priorities to the nodes in the graph. The objective of this task is to arrange the nodes in order of importance according to the measure of their degree centrality.

3.2. Link-based object classification

This link mining task consists of labeling nodes by categories, according to the structure and characteristics of its links. In contrast to traditional pattern classification, this task faces the challenge of the labels being possibly correlated, which adds increased difficulty when designing classification algorithms.

3.3. Object clustering (group detection)

Clustering of nodes has as its purpose detecting groups in the social network that share common properties. The challenge faced with this task is to make contributions in research related to the processes of knowledge discovery.

3.4. Object identification (entity resolution)

The purpose of this task is to determine which references point to the same node within a social network. It is common to consider references between pairs of specific nodes in order to solve a problem known as "entity resolution". In link mining, the aim is to use links to improve entity resolution. For this reason, in addition to specifically considering the attributes of the nodes corresponding to the references intended to be resolved, it is also necessary to take into account the rest of the references that display links pointing to their nodes.

3.5. Link prediction

The task of link prediction is of central importance in the present survey; it involves using the attributes of two specific nodes in a social network as a basis, with the purpose of successfully predicting the existence of a possible link between the two considered nodes. When performing the link prediction task, it is also desirable to take into account other links in the social network. Typically, the researcher observes some links and based on the gathered information predicts links that have not been yet observed; however, the most common way to describe the task of link prediction from the temporal viewpoint is as follows: the state of a set of links is known at the time t and the state of that link set at the time t+1 is predicted.

3.6. Clique discovery

Given a social network, by tending to the task of clique discovery one attempts to discover cliques of interest with common properties, or that conform to a certain specified pattern.

3.7. Graph classification

Differing from the link-based object classification task, which consists in labeling the nodes by category, the task of graph classification aims to assign a category to a graph, from a predetermined set of known categories. That is, this task falls within the supervised learning paradigm.

3.8. Generative models for graphs

This task attempts to discover general principles that dictate the different kinds of networks. Recently, general patterns that lead into generative models for graphs have been observed when studying, among others, the structural properties of biological and social networks.

4. LINK PREDICTION IN SOCIAL NETWORKS

Link prediction stands out among the eight main tasks of link mining due to its reach and relevance in social networking. The authors of the present work have posed the question of what is the way in which we can improve processes in social networks using link prediction [2].

Link prediction is evidently a particular case of link mining. The purpose of this work is to establish the importance of link prediction in social networks [2] [9] [10]. With a social network within the context of several environments, the task of link prediction can be described as a binary classification problem; that is: let n_i and n_j be two different nodes of a social network, which can be potentially linked; the problem to solve is to predict if the link l_{ij} exists or in other words, whether it is equal to zero or one.

Several approaches exist to take on link prediction; among these, we can mention measurements of graph proximity, regression models, probabilistic and statistical models, random Markov fields and relational representations [11].

The task of link prediction is quite relevant in the field of social network analysis, although it has also relevant applications to other domains such as information retrieval, bioinformatics, and e-commerce. By applying such a link prediction algorithm, several tasks are enabled, for instance: identifying spurious interactions, extracting missing information, or evaluating network evolving mechanisms. For example, in social networks currently non-existent but very likely connections may be discovered, and thus recommended as promising friendships, which is a feature that can be helpful for users in finding new friends, improving the user experience, and even strengthening their loyalty to the social network [12]. Another example is the application of similar techniques to elucidate the way or ways in which a particular network (or part of it) grows and evolves over time [13] [14].

In the context of social networks, links between nodes usually represent connections or friendships between users. Thus, the manner in which these links evolve is commonly driven by mutual interests between users, which are intrinsic to the group to which they belong. However, social networks are inherently very fluid and dynamic, with new links and nodes added, as well as removed over time [3].

Understanding the dynamics underlying the evolution of social networks (or even portions of them) is a social problem, due to the large number of variable parameters that must be considered. In this regard, a sensible simplification is to study the connection between two particular nodes. Thus, some questions of interest that may

ISSN 1462 2815

arise are: What factors influence the dynamics of such links? In particular, how is the link between two nodes affected by other nodes? In general, how does the link pattern evolve over time? [10].

According to [9], given a social network, the convention is to denote the sets of nodes and links by N and L, respectively, while $x \in N$, $y \in N$, and $z \in N$ denote specific nodes at time t. Also, $\Gamma(x)$ represents the set of neighboring nodes of $x \in N$. Based on this notation, the task of link prediction may be defined as follows. Let a social network be represented by a graph $G(N, L)$, where N denotes the set of nodes and L the set of links. Thus, a link between nodes $x \in N$ and $y \in N$ at some time t is denoted as $l(x, y)$. If T represents a time after t (i.e. $t < T$), then the clique of G containing the set on links existent between times t and T is represented by $G[t, T]$. Now, let $[t_1, T_1]$ be a training time interval and $[t_2, T_2]$ be a testing time interval, where $T_1 < t_2$. Then, the task of link prediction consists of finding the set of links not present in the clique of $G[t_1, T_1]$, but which will appear in the clique of $G[t_2, T_2]$.

Using this notation, it is possible to state the task of link prediction as a supervised pattern classification problem; for this, the feature space is considered to be the set of all pairs of nodes (x, y) in graph $G(N, L)$. The training set is formed by all pairs or nodes (x_1, y_1) in the training interval $[t_1, T_1]$, which are labeled as follows: the pairs of nodes in $[t_1, T_1]$ such that $l(x_1, y_1) \in L$ are given the label 1, while the pairs of nodes in $[t_1, T_1]$ such that $l(x_1, y_1) \notin L$ are labeled with 0. Now, for the supervised classification phase, let (x_2, y_2) be a pair of nodes in the testing time interval $[t_2, T_2]$; then, the task of link prediction consists of predicting the label of $l(x_2, y_2)$.

Current scientific literature includes several approaches to link prediction, as formulated previously. For instance, there are probabilistic algorithms, such as relational models or those based on the Bayesian decision theory [15].

On the other hand, some methods use a kernel matrix and involve maximum margin classifiers, while there is another group based on maximum likelihood [16] [17]. Also, some algorithms are based on graph evolution models [18], while others make use of a great variety of formulations that have scientific and mathematical foundations [19] [20].

Among this wide range of algorithms, a group based on similarity score calculation between two nodes with the goal of later applying a supervised learning algorithm stands out [21] [22].

In this context, and with the goal of generating the patterns features, there exists a set of effective methods for measuring proximity, taken from ideas in the fields of graph theory, computer science and the social sciences, which will be outlined below.

In this work we describe the most important indices in current literature: the indices based on node neighborhoods and those using similarity scores based on paths.

4.1. Common neighbors

This index reports the number of neighbors that any two nodes, $x \in N$, $y \in N$, share. It is the cardinality of the intersection between the sets containing the neighbors of both nodes:

$$|\boldsymbol{\Gamma}(\mathbf{x}) \cap \boldsymbol{\Gamma}(\mathbf{y})| \quad (1)$$

ISSN 1462 2815

4.2. Jaccard coefficient

This index normalizes the common neighbors index, and is expressed as follows:

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{|\Gamma(x)\cup\Gamma(y)|} \qquad (2)$$

4.3 Salton index

This index, also called Cosine Similarity Index, is defined by the following expression:

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{\sqrt{|\Gamma(x)|\cdot|\Gamma(y)|}} \qquad (3)$$

4.4 Leicht-Holme-Newman index

This index assigns higher similarities to node pairs with several common neighbors, comparing this number not with the maximum possible number of neighbors, but with the expected quantity of neighbors. This index is defined as follows:

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{|\Gamma(x)|\cdot|\Gamma(y)|} \qquad (4)$$

4.5 Sørensen index

This index is defined as:

$$\frac{2|\Gamma(x)\cap\Gamma(y)|}{|\Gamma(x)|+|\Gamma(y)|} \qquad (5)$$

4.6 Hub promoter index

Used for calculating topological features and is defined as follows:

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{\boldsymbol{min}\{|\Gamma(x)|,|\Gamma(y)|\}} \qquad (6)$$

4.7 Hub depressed index

It is a variant of the Hub Promoter Index and is defined as follows:

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{\boldsymbol{max}\{|\Gamma(x)|,|\Gamma(y)|\}} \qquad (7)$$

4.8 Resource allocation index

This index is used for measuring the number of resources that a node receives from another in a social network and is defined as:

$$\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{|\Gamma(z)|} \qquad (8)$$

4.9 Adamic-Adar index

This index assigns higher weights to the least-connected neighbors; it has been used for measuring similitude between two web pages. The Adamic-Adar index is defined as follows:

$$\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{\boldsymbol{log}|\Gamma(z)|} \qquad (9)$$

4.10 Preferential attachment index

The underlying premise of this index's definition is in the evolution of a social network: the probability that a new link has *x* as a final node is proportional to $|\Gamma(x)|$; under this scheme, the probability that a new link connects any two nodes $x \in N$, $y \in N$, is proportional to:

$$|\boldsymbol{\Gamma}(\boldsymbol{x})| \cdot |\boldsymbol{\Gamma}(\boldsymbol{y})| \qquad (10)$$

4.11 Katz index

This index considers the sum of all paths that exist between two nodes $x \in N$, $y \in N$, and is defined as follows:

$$\sum_{t=1}^{\infty} \boldsymbol{\beta}^{t} . \left|\mathbf{paths}_{(x,y)}^{\langle t\rangle}\right| \qquad (11)$$

where β is a positive parameter, *t* represents a path length, and

$$\left|\text{paths}_{(x,y)}^{\langle t\rangle}\right| \qquad (12)$$

is the set of all paths of length *t* that join the nodes *x* and *y*.

4.12 Hitting time

This measure indicates the expected value of the number of steps a random walk would take to traverse the path from *x* to *y*. That is, in a social network, the hitting time $H_{x,y}$ is the expected number of steps in a path starting at node *x* which iteratively moves to some neighbor of the last visited node, which is chosen uniformly at random from the set $\Gamma(x)$ until it arrives to the node *y*.

4.13 SimRank

It is a recursive process in which a parameter called *decay factor*, defined as $\gamma \in [0,1]$, intervenes. For two nodes x, y the SimRank indicator is defined so that $SimRank(x,x) = 1$ and $SimRank(x,y)$ is:

$$\boldsymbol{\gamma} \cdot \frac{\sum_{a \in \Gamma(x)} \sum_{b \in \Gamma(y)} \boldsymbol{SimRank(a,b)}}{|\boldsymbol{\Gamma}(\boldsymbol{x})| \cdot |\boldsymbol{\Gamma}(\boldsymbol{y})|} \qquad (13)$$

The indices described above are not the only ones in the literature, but they are indeed the most important and enjoy the highest usage in scientific works related to link prediction.

5. SOCIAL NETWORKS: GENESIS AND EVOLUTION

The topic of social networks has been present in scientific literature since the mid-fifties [23]. In this pioneering article on the subject, the author explores the dynamics of conjugal roles within urban families and their relationship with social networks. The paper meticulously analyzes the active and passive interactions between couples and their social environments, emphasizing the distribution of responsibilities and power within the household. The intricate connections between individual family units and their broader social contexts are clearly uncovered, highlighting the importance of social networks in shaping family dynamics and conjugal roles in urban settings. These elements allow for an understanding of marital roles and expectations, which are influenced and shaped by the broad social networks that couples are part of.

In the late eighties and early nineties, the overwhelming presence of the first World Wide Web browser arrived. This milestone in technological advances brought impressive consequences for the instantaneous dissemination of data and information, as we have all witnessed who have been fortunate enough to live in this era of rapid scientific and technological development.

It didn't take long for the concept of a social network to crystallize within the realms of the World Wide Web. There is a widespread opinion among experts that the first modern, structured, and organized social network was the so-called Six Degrees Social Engine [24].

The fundamental idea underlying this primordial social network emphatically assures that a chain of interactions can be established to connect any pair of individuals within a maximum of six steps. That is, every pair of human beings in the universe are within six steps (degrees) of separation from each other.

From this point, the advancements in the development of social networks, as we currently know them, accelerated dramatically. Thus, a lustrum after the appearance of Six Degrees, the famous social network LinkedIn made its entry into the business world [25].

LinkedIn is a professional social network, the most famous worldwide. This network is useful for a large number of people around the world. According to the interests and activities of each person, all its benefits can be leveraged to find the dream job, or to connect and strengthen professional relationships. Additionally, it can be an effective guide that allows acquiring the necessary skills to achieve success in professional, technical, or administrative careers.

New types of social networks began to emerge in rapid succession, adapting to the specific needs of individuals. Suddenly, an ultramodern type of entrepreneurs emerged, who began to have (and still have) a significant influence on the habits and activities of millions of people around the world.

In 2004, Mark Zuckerberg and his partners founded Facebook. This social network came to satisfy a deeply felt demand of most human beings in the world: to foster and maintain contact between people. Facebook is an excellent facilitator of very human processes such as sharing information, jokes, memes, news, events, offers, and other content of interest, taking advantage of the benefits of audiovisual media [26].

YouTube is the quintessential video-focused social network. This network has many millions of users worldwide, and memberships grow daily. It's an excellent aid in the educational field, due to the vast amount of courses offered through this social network. Among many other advantages, YouTube allows for the monetization of videos.

Twitter was founded in 2006 as one of the social networks operated by Twitter, Inc. (in addition to Vine and Periscope, which are now obsolete). One of the most attractive and popular advantages of Twitter is that it allows sending messages very quickly and easily, in addition to accompanying them with documents and audiovisual content [27]. Since July 24, 2023, Twitter has been renamed X.

In 2009, the social network WhatsApp was founded. The technology underlying WhatsApp allows for sending messages, making calls, and video calls through the smartphone's internet connection [28].

Instagram emerged in 2010. It is a social network that is also a mobile application. Instagram features facilities that allow people to apply filters to images and videos and then share those contents with friends and family [29].

There is a continuous avalanche of creators generating new social networks to cover various needs in people. Just as many social networks are ephemeral and some of them become obsolete, new social networks are emerging with the aim of satisfying certain specific needs of human beings (educational, emotional, or sexual; for example, Tinder and Onlyfans). That is, there are social networks for all tastes and inclinations: Pinterest: 2011; Snapchat: 2011; TikTok: 2012; Threads: 2023.

According to some publications and internet sites (www.rdstation.com, for instance), new and varied social networks are emerging that, for the moment, are little known: Bumble, Caffeine, Clubhouse, Discord, Facecast, Mastodon, Nextdoor, SteemIt, Supernova, Swarm, and Twitch, among others.

6. LINK PREDICTION: STATE-OF-THE-ART

This section is very relevant in the context of the present survey. Here, the descriptions of the main scientific research works, which are related to the central theme of the survey, are included.

It would be pointless to try to cover exhaustively all the publications that have been recorded on the topic in impact journals. Therefore, we will include a representative sample of what, in our opinion, are the most relevant works. We are confident that this strategy will provide the reader with as complete a perception as possible of the development of the topic over time, as well as a clear view of the state of the art in the central theme of the survey: link prediction in social networks.

The topic gained attention in the scientific community after Liben-Nowell and Kleinberg published, in 2007, their iconic paper on the link prediction problem for social networks [9].

6.1. Methods from several scientific disciplines

One of the first known works is the one published, in 2008, by Murata and Moriyasu, who use the structural properties of online social networks to perform link prediction [30].

The authors consider the general connections of the question-and-answer boards of that time as a social network. They correctly hypothesize that an accurate prediction of the evolution of social networks would improve the communication carried out through networks. The authors base their method on weighted proximity measurement, in order to propose the method presented in the paper that aims to predict links. They also assume that the weights of the links can support obtaining better calculations of the measurements.

The fact that educators [31], networks engineers [14], biologists [32], physicists [33] [34] [35], and medical professionals [36] became interested in the topic of link prediction in social networks cannot be overlooked. They found an appropriate way to frame the problems associated with their areas of study as social networks.

It's interesting to observe how the authors of [31] recognize that among the youth of the recent world, social networks have gained more popularity day by day. The authors are educators, social scientists, and cybersecurity researchers, and they establish that this situation opens up possibilities for educators to use them on their virtual higher education platforms. The article addresses the topic of higher education. Specifically, it refers to knowledge management that has recently been supported by social networks. The authors make a great effort in creating a compendium of open research questions in this context. First, they perform a diagnosis that throws them an unavoidable reality: the fact of honestly recognizing the role that online social networks play in various activities of the actors involved in higher education processes. For example, in teaching and learning, they are invaluable tools that arouse interest both in students and mentors. In addition, the authors propose the possibility of using social networks in activities that have not been adequately addressed by higher education researchers. For example, management, accumulation, and sharing of knowledge.

Engineers dedicated to networks have also managed to adopt and apply techniques similar to link prediction. In [14], the authors elucidate the way in which a particular network grows and evolves over time. They work on network routing through a link discovery algorithm (redirectors, as they call them) that reaches the final route in fewer iterations than those reported in the state of the art. With this faster route convergence, greater energy savings can be achieved, with the corresponding advantages that this entails.

Biologists also timely joined the wave of advances in link prediction in social networks. According to the authors of [32], a large number of biological functions are related to protein-protein interactions. Therefore, link prediction in these types of interaction networks is an important task in biology. In this scientific work, the authors introduce new ways to measure the proximity of nodes, using the eigenvectors of the normalized Laplacian matrix for this purpose. Meanwhile, the spectral clustering model is based on the Bray-Curtis metric.

Shortly after the publication of the paper by Liben-Nowell and Kleinberg on the link prediction problem for social networks, physicists immediately realized the importance of the topic and the opportunities it brought to the forefront of scientific research. Thus, in [33], the authors state that link prediction is an open problem in social networks, currently attracting a lot of interest in the scientific community.

They focus their efforts on one of the emblematic topics of unsupervised learning: clustering. Among their findings, it can be mentioned that as the clustering increases, the accuracy of the prediction methods improves significantly. On the other hand, if the networks are sparse and weakly clustered, the performance of the prediction methods is poor.

In another work published by physicists, the authors of [34] discovered that in biological networks and in other networks taken from real environments, the network links that connect low-degree nodes are those reported as typical missing links. The experimental part of this work included ten indices, calculated with data from four social networks taken from the real world. The authors found that the index with the best performance is the Leicht-Holme-Newman index.

In our opinion, it's worth mentioning one more work published by physicists at the beginning of the decade from 2010 to 2020. Notice how, in a short time, these physics specialists fully understood the importance of the topic that occupies us in the present survey. This is exhibited by the authors of [35], who recognize the importance of link prediction, which "...has been widely used to extract missing information, identify spurious interactions, evaluate network evolving mechanisms, and so on...". Then, they mention what they consider disadvantages of similarity-based algorithms, as they equally consider the contributions of each common neighbor to the connection likelihood of two nodes. Because of this, the authors propose a model for link prediction, which is based on the centrality of nodes of common neighbors.

The topic of link prediction in social networks quickly permeated many disciplines of interest, including those cultivated by health professionals. In [36], the authors combine three topics of great interest: link prediction, the domain of medical research, as well as a recurring theme in the present work that has gained the attention of the scientific community due to its consequences on global scientific production, which are the co-authorship social networks. The authors of the paper start from an incontrovertible fact: researchers are prone to searching for possible collaborators with common interests. In this work, intensive use is made of indices such as Common Neighbors, Adamic/Adar, Preferential Attachment, Jaccard's Coefficient, and Katz. The authors tested their proposals on networks formed by scientists who were co-authors of research works related to coronary artery disease. The data was acquired from the Web of Science (WoS) site.

Researchers in evolutionary algorithms also made significant contributions to the topic. In [37], the authors perform link prediction in dynamic social networks, using a method that includes evolutionary algorithms. The authors introduce the Covariance Matrix Adaptation Evolution Strategy. With this, they manage to predict future links in networks, by optimizing the weights... Then, they successfully applied this method to a large dynamic social network of more than a million nodes: The Twitter reciprocal reply networks.

Scientific research related to the topic of link prediction in social networks that generates methods from several scientific disciplines remains relevant to this day. A plethora of papers in impact journals are published month by month.

Here are some examples of methods proposed from several scientific disciplines: [2]: education; [3]: data engineering; [4][6]: reliability; [5] [18]: graph theory; [7]: data mining; [8]: data and policy; [10]: information; [11]: Markov chains; [12] [13]: psychology, among many others.

6.2. Methods from artificial intelligence, machine learning and related fields

As previously shown, the response from other disciplines to the paper by Liben-Nowell and Kleinberg on the link prediction problem for social networks was very quick (1 year). In contrast, scientific research generating smart algorithms to tackle this problem came a bit later. Thus, scientific researchers in machine learning, artificial intelligence, computational intelligence, pattern recognition, and related areas began producing impactful research works on the topic until 2013.

However, since that year, the scientific production on smart algorithms addressing the link prediction problem for social networks has been constant and of high quality, up to the present day.

Interestingly, one of the first quality papers on the central theme of this survey emerged in the military field. In [38], the authors address the problem of link prediction within terrorist networks. The most relevant aspect of this work is the presence of one of the most used learning paradigms in the state of the art of machine learning and related areas: supervised machine learning. The authors specifically face two main challenges: "(1) how to extract latent link patterns from the (sparse) network topology and combine them with optional explicit features to improve link prediction accuracy; and (2) how to reliably infer links with an incomplete set of only positive examples". In the experimental part, the authors highlight the utility of link prediction "...in a covert human smuggling network...", which is of undeniable relevance in the field of national security.

That same year, 2013, a second paper on the topic of this survey was published in the field of machine learning, artificial intelligence, computational intelligence, pattern recognition, and related areas [39]. The authors venture into social networks containing hundreds of millions of users. They discover that the main bottleneck in link prediction processes consists of the difficulties involved in extracting features, which will form the patterns to classify, to thus determine the prediction of links. Given this, the authors propose the simple calculation of some simple attributes. With those attributes, they form patterns that are fed into machine learning classifiers, which are capable of successfully identifying missing links. The experiments were conducted on ten large social network datasets: Academia.edu, DBLP, Facebook, Flickr, Flixster, Google+, Gowalla, TheMarker, Twitter, and YouTube. In addition to helping users find known contacts, the proposed methods can be useful in uncovering hidden links in social networks.

Four years later, in 2017, a very important paper was published [40]. Its relevance lies in two fundamental reasons: 1) the problem of link prediction in multiplex networks (multiple social networking platforms) is studied; and 2) it uses three of the most well-known and popular classical pattern classifiers in the scientific community: Naive Bayes, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). In the experimental part, the authors consider Twitter and Foursquare. The experimental results show that including information across layers significantly improves prediction performance. The performance of the SVM classifier (average accuracy of 89%) makes it the best performer in the experiments conducted.

The year 2019 brought two papers that deserve to be in this survey. First, the authors of [41] propose a new algorithm that performs link prediction of nodes in a social network. To develop their proposal, the authors combine machine learning and computational intelligence algorithms with hierarchical representation learning algorithms for networks (HARP). A novel aspect of the proposal is the inclusion of the well-known optimizer (widely used in deep learning algorithms) to establish a social network link prediction model based on machine learning and HARP. The experimental results show that the new method is stable and exhibits excellent performance.

The main feature of the second research work is that it uses human trajectory internet data mining [42]. The authors begin their manuscript with an extensive explanation about the long-standing fundamentals of the problem established more than a decade ago by Liben-Nowell and Kleinberg. Then, they hypothesize that the problem to be solved can be modeled as an intelligent prediction problem. Specifically, the prediction of future links between social network users, seen as nodes. The proposed method allows understanding the relationship between nodes, in addition to predicting links in human trajectories. In this work, the proposed method is called IMC-TEM-SUCM, and it is useful for predicting stable links by identifying node-node relationships. Experimentally, the Foursquare and Gowalla networks are examined. An important novelty is that disruptive attributes (reputation and optimism, among others) are considered, in addition to the route length in the link prediction problem. In the execution of this task, pattern classification algorithms based on the similarity of

ISSN 1462 2815

Cosine and Jaccard coefficients play a central role (here, the possibility of forming links is closely related to greater similarity).

Since 2020 and so far this decade, the scientific output of works related to the current survey has significantly multiplied, both in quantity and quality. In recently published articles in high-impact journals, terms like "deep learning" and acronyms like CNN (Convolutional Neural Networks) are now appearing, which denote an update of content towards trending topics. This undoubtedly enriches the research in the link prediction problem for social networks, favoring research in the area.

Given the profusion of recent works in the state of the art, as well as the sophistication of approaches, it is not an easy task to decide on the relevance of papers. Following the extensive documentary research we have conducted, we, the authors of this survey, have noticed several interesting things we wish to share with the kind readers: i) over the months, sophisticated topics appear in the issues of the consulted impact journals; ii) authors are inclined to perform updates and hybrid algorithms of topics already addressed in the development of previous years and decades; iii) it is necessary to note the importance that has been preserved by the classic original paper from 2007, where Liben-Nowell and Kleinberg established in a clear, but rigorous way, the link prediction problem for social networks; iv) few authors have proposed new indices to measure proximity or similarity between nodes; v) the vast majority of authors continue to use, consistently, the indices described in Section 3 of this survey (3. Link Mining in Social Networks); vi) in the most recent papers, the inclusion of trending topics of machine learning, artificial intelligence, computational intelligence, pattern recognition, and related areas is evident (among the most frequent and important ones, "deep learning" can be mentioned); vii) some of the papers are of such high quality that they generate new lines of research; viii) these new lines of research are, potentially, sources of topics for new research, projects, papers, and undergraduate or graduate theses; ix) the theme of this survey (link prediction in social networks) is present, current, and is a "hot" topic; x) it is very encouraging to know that teams from practically the best universities in the world participate in the scientific development on the aforementioned topic.

Considering the sophistication of the scientific ideas underlying recent articles, we, the authors of this survey, have decided not to go into details when describing the scientific works in question. Instead, for each of the ten high-quality works we have selected from high-impact journals, we will briefly mention the concepts on which the research and developments are based. We leave in the hands of the kind reader the unique opportunity to further investigate these scientific concepts and, potentially, to select one of these works as a starting point to initiate their own research on the attractive topic addressed in this survey.

[17] (2020): The problem of Influence Maximization is addressed; to tackle it, an original algorithm based on maximum likelihood and Independent Cascade is proposed.

[43] (2021): The authors apply supervised machine learning algorithms to tackle the problem of link prediction in social social networks. In order to assign weights to the network connections, especially on YouTube where the experiments were conducted, the conceptual idea of the "closed triangle" is used.

[44] (2021): One of the novelties is that the authors revisit and refine an approach used in previous years' work: the unsupervised learning paradigm, specifically the well-

known k-means algorithm. However, there are improvements to the algorithm through shifts. Additionally, differences between global and local attributes are incorporated and taken into account when experimenting on Facebook and Twitter.

[45] (2022): This paper is a prime example of the overwhelming emergence of trending topics such as "deep learning" and "convolution neural networks". Here, the authors apply convolution neural networks and a concept generated by them, that of relational patterns, in link prediction. They also use a resource that is fashionable in the state of the art: heat maps.

[16] (2023): One of the reasons why we have decided to include a brief review of this paper is that the authors address a problem related to the privacy of users on social networks. Specifically, this research work proposes an alternative method to the k-anonymization of user data on social networks, especially on LinkedIn, Facebook, and Twitter. Given that the method is a hybrid between Support Vector Machines (SVM) and Neural Networks (MLP), the authors have named it NeuroSVM. One of the major advantages of this alternative method of k-degree anonymization is that, according to the authors, it reduces data loss in the processes carried out on social networks, including the task of link prediction.

[21] (2023): This research work contains something truly attractive for those of us dedicated to applied scientific research on topics sensitive to human beings: the application of the theoretical results of link prediction in social networks in the detection and prevention of diseases. The authors make use of the entire arsenal possessed by machine learning, data mining, and related areas (especially the k-NN family and Support Vector Machines SVM) to model a Patient Similarity Network. Among the main advantages it exhibits, this novel network models not only patients as nodes but also laboratory data and symptoms. It is worth noting that traditionally, this last concept had been alien to the development of algorithms related to social networks in the domain of medicine.

[46] (2023): The theoretical content of this paper is a clear example of the sophistication achieved since 2007, when Liben-Nowell and Kleinberg published their iconic paper on the link prediction problem for social networks, in the significant advances that have been presented on the central topic of this survey. Firstly, the authors smartly transform the link prediction problem into a signal processing problem. To this end, they introduce an original and innovative technology, which they have called "compressed sensing." Then, surprisingly, they integrate topological concepts with the information present in networks, through the incorporation of a matrix method well-known to physicists: Singular Value Decomposition (SVD), but as they consider a sparse representation K, the new method is named K-Singular Value Decomposition (K-SVD). To make their algorithm even more sophisticated, yet effective, the authors also incorporate the restricted isometry property and the orthogonal matching pursuit algorithm into the matrices. We consider that this content can serve as a starting point for those who wish to develop theoretical work on the topic at hand.

[15] (2024): In our opinion, this paper possesses two characteristics that make it attractive to scientific researchers. On one hand, the authors adapt the link prediction problem for social networks in order to achieve user identification across multiple social networks. On the other hand, they make use of one of the most longstanding, well-known, and appreciated algorithms in the supervised machine learning paradigm: naive Bayes.

[47] (2024): In this research work, the authors take a risk and accept to tackle a task that is social, conceptually deep, and therefore, challenging at the same time: the prediction of links that evolve over time in dynamic social networks. In addition to the usual data in static networks, when working with dynamic networks, temporal changes in the evolutionary information of the social network structure must also be considered. The authors successfully incorporate dynamic information into machine learning algorithms in order to perform link prediction in dynamic social networks.

[48] (2024): If any research group is looking for a cutting-edge research work on link prediction in social networks, the content of this paper is the answer. The article presents several characteristics, which make it an advanced paper. To start with, similar to the previous paper, the authors address the link prediction problem in dynamic social networks. Furthermore, it considers the textual content managed by network users, taking into account temporal changes. To successfully tackle such sociality in the problem, the authors make use of state-of-the-art concepts. For example: temporal Graph Neural Networks (GNNs), pipelines, BERT language, experiments on blockchain, dense preprocessing layers in neural networks, and high-resolution temporal information.

7. CONCLUSIONS

Although the concept of social networks appeared in scientific literature since the mid-1950s (assuming that dynamics of marital roles within urban families behave as social networks), it took half a century for the problem of link prediction in social networks to be established, with particular attention to social networks.

The pioneering article by Liben-Nowell and Kleinberg in formulating the problem (2007) has had such a significant influence on the development of this branch of science that it continues to be cited and considered in the development of state-of-the-art works, which have been published in recent months.

From the outset, the problem of link prediction in social networks has attracted the attention of researchers from a wide range of scientific disciplines. Scientific research related to the topic of link prediction in social networks, which generates methods from several scientific disciplines, remains relevant to this day: education, data engineering, reliability, graph theory, data mining, data and policy, information, Markov chains, psychology, among many others.

In particular, this problem has been addressed using methods from machine learning, artificial intelligence, computational intelligence, pattern recognition, and related areas. In this context, an impressive number of new intelligent methods have been generated to address the aforementioned problem, with such quality that they are published in high-impact journals.

Since the beginning of the current decade, the scientific output of works related to the current survey has significantly multiplied, both in quantity and quality. In recently published articles in high-impact journals, terms like "deep learning" and acronyms like CNN (Convolutional Neural Networks) are now appearing, which denote an update of content towards trending topics. This undoubtedly enriches research in the link prediction problem for social networks, favoring research in the area. It is gratifying to observe that the results of our research promise to positively impact the generation of valuable information for the sake of scientific development.

References

1) Brahma M., Rejula M.A., Srinivasan B., Kumar S.N., Banu W.A., Malarvizhi K., Priya S.S., Kumar A. - Learning impact of recent ICT advances based on virtual reality IoT sensors in a metaverse environment. Measurement: Sensors, 27 (2023) art. no. 100754. doi.org/10.1016/j.measen.2023.100754. doi.org/

2) Lytras M.D., Mathkour H.I., Abdalla H., Yanez-Marquez C., de Pablos P.O. - The social media in academia and education: Research R-evolutions and a paradox: Advanced next generation social learning innovation. Journal of Universal Computer Science, 20 (2014) 1987 - 1994. doi.org/10.3217/jucs-020-15

3) Jin D., Wang X., He D., Dang J., Zhang W. - Robust Detection of Link Communities with Summary Description in Social Networks. IEEE Transactions on Knowledge and Data Engineering, 33 (2021) 2737 - 2749, art. no. 8930274. doi.org/ 10.1109/TKDE.2019.2958806

4) Gupta V.K., Shukla D. - Estimation of Average Degree of Social Network Using Clique, Shortest Path and Cluster Sampling to monitor Network Reliability. Reliability: Theory and Applications, 17 (2022) 326 - 339. doi.org/10.24412/1932-2321-2022-268-326-339

5) Su J., Li J., Chen J. - Probabilistic graph model mining user affinity in social networks. International Journal of Web Services Research, 18 (2021) 22 - 41. doi.org/10.4018/IJWSR.2021070102

6) Anitha M., Venkataraman Y. - Computation of structural parameters of a collaboration graph from social networks. Expert Systems with Applications, 243 (2024) art. no. 122800. doi.org/10.1016/j.eswa.2023.122800

7) Shebaro M., de Moura L.N., Tešić J. - Improving association discovery through multiview analysis of social networks. Social Network Analysis and Mining, 14 (2024) art. no. 42. doi.org/10.1007/s13278-023-01197-3

8) Malaterre C., Lareau F. - Inferring social networks from unstructured text data: A proof of concept detection of hidden communities of interest. Data and Policy, 6 (2024) art. no. e5. doi.org/10.1017/dap.2023.48

9) Liben-Nowell D., Kleinberg J. - The Link Prediction Problem for Social Networks. Journal of the American Society for Information Science and Technology, 58 (2007) 1019 - 1031. doi.org/10.1002/asi.20591

10) Xiu Y., Liu X., Cao K., Chen B., Chan W.K.V. - An extended self-representation model of social networks for link prediction. Information Sciences, 662 (2024) art. no. 120254. doi.org/10.1016/j.ins.2024.120254

11) Dennis L.A., Fu Y., Slavkovik M. - Markov chain model representation of information diffusion in social networks. Journal of Logic and Computation, 32 (2022) 1195 - 1211. doi.org/ 10.1093/logcom/exac018

12) Dávila Espuela N., Reina Paz M.D., Sevilla Sevilla C. - Social networks and customer loyalty: review of loyalty keys and main social networks publications' characteristics. Frontiers in Psychology, 14 (2023) art. no. 1286445. doi.org/10.3389/fpsyg.2023.1286445

13) Jackson J.C., Halberstadt J., Takezawa M., Liew K., Smith K., Apicella C., Gray K. - Generalized Morality Culturally Evolves as an Adaptive Heuristic in Large Social Networks. Journal of Personality and Social Psychology, 125 (2023) 1207 - 1238. doi.org/10.1037/pspa0000358

14) López-Yáñez I., Camacho-Nieto O., Yáñez-Márquez C., Uriarte-Arcia A. - Fast route convergence in dynamic power controlled routing for wireless ad–hoc networks. IEEE Latin America Transactions, 11 (2013) 607-608. doi.org/10.1109/TLA.2013.6502870

15) Ding X., Zhang H., Ma C., Zhang X., Zhong K. - User Identification Across Multiple Social Networks Based on Naive Bayes Model. IEEE Transactions on Neural Networks and Learning Systems, 35 (2024) 4274 - 4285. doi.org/10.1109/TNNLS.2022.3202709

16) Kaur H., Hooda N., Singh H. - k-anonymization of social network data using Neural Network and SVM: K-NeuroSVM. Journal of Information Security and Applications, 72 (2023) art. no. 103382. doi.org/10.1016/j.jisa.2022.103382

17) Liu W., Li Y., Chen X., He J. - Maximum likelihood-based influence maximization in social networks. Applied Intelligence, 50 (2020) 3487-3502. doi.org/10.1007/s10489-020-01747-8

18) Yu D., Zhou Y., Zhang S., Li W., Small M., Shang K.-K. - Information cascade prediction of social networks based on physics-informed graph convolutional network. New Journal of Physics, 26 (2024) art. no. 013031. doi.org/10.1088/1367-2630/ad1b29

19) Acevedo-Mosqueda M. E., Yáñez-Márquez C., Acevedo-Mosqueda M. A. - Bidirectional associative memories: Different approaches. ACM Comp. Surv. (CSUR), 45 (2013) 1-30. doi.org/10.1145/2431211.2431217

20) Yáñez-Márquez C. - Toward the bleaching of the black boxes: minimalist machine learning. IT Profess., 22 (2020) 51-56. doi.org/10.1109/MITP.2020.2994188

21) Dutta S., Singh N., Das S., Chakraborty S. - Proceedings of the International Conference on Computational Intelligence in Pattern Recognition, Springer (2023) 453-464.

22) Villuendas-Rey Y., Rey-Benguría C.F., Ferreira-Santiago Á., Camacho-Nieto O., Yáñez-Márquez C. - The naïve associative classifier (NAC): a novel, simple, transparent, and accurate classification model evaluated on financial data. Neurocomputing, 265 (2017) 105-115. doi.org/10.1016/j.neucom.2017.03.085

23) Bott E. - Urban Families: Conjugal Roles and Social Networks. Human Relations, 8 (1955) 345-384. doi.org/10.1177/001872675500800401

24) Shu W., Chuang Y.-H. - The perceived benefits of six-degree-separation social networks. Internet Research, 21 (2011) 26-45. doi.org/10.1108/10662241111104866

25) Thew D. - LinkedIn - A user's perspective: Using new channels for effective business networking. Business Information Review, 25 (2008) 87-90. doi.org/10.1177/0266382108090810

26) LEADERSHIP MARK ZUCKERBERG Facebook Stares Down Success, https: //fortune.com/2005/11/28/facebook-mark-zuckerberg-ceo-early-stage/ (accessed 14 April 2024).

27) Sotudeh H. - How social are open-access debates: a follow-up study of tweeters' sentiments. Online Information Review, 48 (2024) 159-186. doi.org/10.1108/OIR-09-2022-0502

28) Allagui B. - Writing through WhatsApp: An evaluation of students writing performance. International Journal of Mobile Learning and Organisation, 8 (2014) 216-231. doi.org/10.1504/IJMLO.2014.067022

29) Anderson K.E. - Getting acquainted with social networks and apps: Instagram's instant appeal. Library Hi Tech News, 33 (2016) 11-15. doi.org/10.1108/LHTN-03-2016-0011

30) Murata T., Moriyasu S. - Link prediction based on structural properties of online social networks. New Generation Computing, 26 (2008) 245-257. doi.org/10.1007/s00354-008-0043-y

31) Benson V., Morgan S., Tennakoon H. - A Framework for Knowledge Management in Higher Education Using Social Networking. International Journal of Knowledge Society Research, 3 (2012) 44-54. doi.org/10.4018/jksr.2012040104

32) Symeonidis P., Iakovidou N., Mantas N., Manolopoulos Y. - From biological to social networks: Link prediction based on multi-way spectral clustering. Data & Knowledge Engineering, 87 (2013) 226–242. doi.org/10.1016/j.datak.2013.05.008

33) Feng X., Zhao J.C., Xu K. - Link prediction in complex networks: a clustering perspective. The European Physical Journal B-Condensed Matter and Complex Systems, 85 (2012) 1-9. doi.org/10.1140/epjb/e2011-20207-x

34) Yu-Xiao Z., Lü L., Zhang Q.-M., Zhou T. - Uncovering missing links with cold ends. Physica A: Statistical Mechanics and its Applications, 391 (2012) 5769-5778. doi.org/10.1016/j.physa.2012.06.003

35) Liu H., Hu Z., Haddadi H., Tian H.L. - Hidden link prediction based on node centrality and weak ties. Europhysics Letters, 101 (2013) 18004. doi.org/10.1209/0295-5075/101/18004

36) Yu Q., Long C., Lv C., Shao H., He P., Duan Z. - Predicting Co-Author Relationship in Medical Co-Authorship Networks. PloS ONE, 9 (2014) e101214. doi.org/10.1371/journal.pone.0101214

37) Bliss C.A., Frank M.R., Danforth C.M., Dodds P.S. - An evolutionary algorithm approach to link prediction in dynamic social networks. Journal of Computational Science, 5 (2014) 750–764. doi.org/10.1016/j.jocs.2014.01.003

38) Tran P.V. - Supervised link prediction in social networks with positive and unlabeled examples. Military Operations Research, 18 (2013) 53 - 62. doi.org/10.5711/1082598318353

39) Fire M., Tenenboim-Chekina L., Puzis R., Lesser O., Rokach L., Elovici Y. - Computationally efficient link prediction in a variety of social networks. ACM Transactions on Intelligent Systems and Technology, 5 (2013) art. no. 2542192. doi.org/10.1145/2542182.2542192

40) Jalili M., Orouskhani Y., Asgari M., Alipourfard N., Perc M. - Link prediction in multiplex online social networks. Royal Society Open Science, 4 (2017) art. no. 160863. doi.org/10.1098/rsos.160863

41) Shao H., Wang L., Ji Y. - Link prediction algorithms for social networks based on machine learning and HARP. IEEE Access, 7 (2019) 122722-122729. doi.org/10.1109/ACCESS.2019.2938202

42) Suryakumar B., Ramadevi E. - An enhancement on mobile social network using social link prediction with improved human trajectory internet data mining. International Journal of Advanced Computer Science and Applications, 10 (2019) 146-150. doi.org/10.14569/IJACSA.2019.0100318

43) Sharma U., Khatri S.K., Patnaik L.M. - A supervised learning approach for link prediction in complex social networks. International Journal of Advanced Intelligence Paradigms, 20 (2021) 1 - 15. doi.org/10.1504/IJAIP.2021.117607

44) Bhanodia P.K., Khamparia A., Pandey B. - Supervised shift k-means based machine learning approach for link prediction using inherent structural properties of large online social network. Computational Intelligence, 37 (2021) 660 - 677. doi.org/10.1111/coin.12372

45) Wu J.M.-T., Tsai M.-H., Li T.-W., Pirouz M. - Analysis and construction by convolution neural network of link prediction model on social network. Journal of Intelligent and Fuzzy Systems, 43 (2022) 2167 - 2178. doi.org/10.3233/JIFS-219316

46) Yang J., Wu Y. - A link prediction method based on compressed sensing for social networks. Applied Intelligence, 53 (2023) 29300 - 29318. doi.org/10.1007/s10489-023-05060-y

47) Choudhury N. - Community-Aware Evolution Similarity for Link Prediction in Dynamic Social Networks. Mathematics, 12 (2024) art. no. 285. doi.org/10.3390/math12020285

48) Dileo M., Zignani M., Gaito S. - Temporal graph learning for dynamic link prediction with text in online social networks. Machine Learning, 113 (2024) 2207 - 2226. doi.org/10.1007/s10994-023-06475-x