How to Cite:
Kadam, S., & Bhattacharya, R. (2022). Predicting future link in social network: A review. *International Journal of Health Sciences*, *6*(S2), 8232–8240. https://doi.org/10.53730/ijhs.v6nS2.7087

Predicting future link in social network: A review

Sangeeta Kadam
M. Tech. (Scholar) Department of Computer Science & Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India
Corresponding author email: s.kadam@ssipmt.com

Riju Bhattacharya
Associate Professor, Department of Computer Science & Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India
Email: riju@ssipmt.com

***Abstract*---**A social network can have many different types of links or margins between nodes. Those, for example, could be social contacts, major contacts, or calls. Link Prediction is the problem of predicting edges that may not be present in a given or present time, but that have not yet been discovered and may occur in the near future. We are developing predictive linking methods based on step-by-step analysis of network nodes. Consider a network of collaborative writing among scientists, e.g. The two scientists closest to the network will have similar colleagues, so they may be working together soon. Our goal is to make this accurate idea more accurate and to understand what steps to take to approach the network that lead to the most accurate prediction of a link. Link prediction algorithms can be divided into three categories: Node Neighborhood Mode, Mode-based Mode, and Meta Mode. The node location method is based on network location features, which focuses mainly on the node structure (i.e., based on the number of common friends shared by two users). Local-based measures are: general neighborhood, Jaccard Coefficient, Adamic/Adar, and preferred attachments. Path-based algorithms take into account the interaction of all the pathways between two nodes. Path-based algorithms are: Katz, Sim-Rank, Hit Time and Travel Time, Fixed Page Rank, Prop Flow, and High-Performance Link Prediction. Meta-approaches change data before it is transferred to one of the route-based modes. The algorithms are: low rating, invisible bigrams, and clustering.

***Keywords*---**link prediction, analysis, social network, review.

Introduction

The problem of predicting a link is often defined as the function of predicting how a link can live between a few nodes carelessly. In other words, the prediction of a link is the problem of determining whether a link exists between two things or not. Predicting diversity in a social network is known as a "link guessing problem." The link prediction problem is best defined as both the detection of neglected links in the current network or as a time series problem where the task is to predict which links will be present in the network at t + 1 in terms of network status during t. For example, consider a social networking site for collaborative research among researchers close to a network who may be more likely to participate in the future. Link prediction is a small part of social network analysis, which emphasizes the margins between objects. For this reason, link prediction turns out to be more exciting than conventional data mining sites that focus on objects. Link prediction can be used in many regions as a recommendation and in crime investigation programs. Prediction linking methods are expected to be based on various steps to analyses the proximity of nodes to the network, leading to more accurate link prediction. In the predictive link method, all methods assign a link weight (x, y) to two pairs of nodes, x and y, based on the given approximate estimate and the contribution graph G. A rated list in the point reduction setting (x, y) is generated. This provides predictable new links with a decrease in confidence. The prediction can be tested based on actual observations in the test data sets.

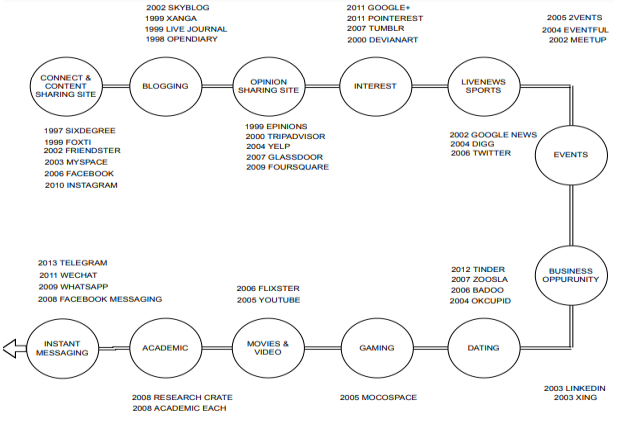

Figure 1.1: Timeline of social networks based on category

Literature Review

Introduced online dating sites, which have grown in popularity in recent years. As a result, new multidisciplinary research guides have emerged as social networking analytics methods are applied to networks that contain hundreds of millions of users. Unfortunately, links between individuals may be lacking due to incomplete detection processes or because they have not yet been displayed on the Internet (i.e., friends in the real world did not build a visible connection). An ever-growing social network that is likely to add daily users with thousands of connections. A key step in link prediction techniques is to extract the structural

features needed to separate links. In this paper, we suggest a set of simple, easy-to-calculate structural features that can be analysed to identify missing links. It shows that a machine learning class trained using the simple features of the proposed structure can effectively identify missing links even when used in the complex problem of distinguishing links between individuals with one partner. The new number of friends we have created is shown as a good predictor of non-existent links, and analytical tests were conducted on five major social media databases: Facebook, Flickr, YouTube, Academia, and The Marker. Our methods can give social network operators the power to help users find known offline contacts and make new friends online. It can also be used to display hidden links in an online social network.

Yugchhaya Dhote et al. (2013) The proposed online social network has captured the attention of users in recent years. Some social networking sites are specifically designed to interact with people, and some of them are app-based in providing content sharing and social networking. Link prediction is a new framework for cross-sectional research in which existing links are analyzed and future links are predicted among millions of social network users. Common predictor prediction methods focus on the use of graph metrics to determine where new links might appear. Little work has been done on analyzing long-term graph styles. This paper performs a survey and analysis of the emergence of a predictive link. It was found that previous graph production models were unreasonable in their predictions and could be fulfilled using short-term metrics, leading to more accurate link predictions. In addition, the paper is completed with a proposed framework that uses a temporary element and similar local features.

Has been studied in recent years, with the emergence of social media, such as Facebook, Twitter, and Sina Microblog, and more and more researchers are paying attention to social media. Link prediction is one of the most intriguing issues in social network analysis, which uses existing network information, such as node features and edges, to predict strong relationships that will be built in the future. This paper summarizes the methods for predicting links on social media, including traditional link prediction methods, link prediction methods for various networks, and temporary link prediction methods.

Gupta et al. (2015)""proposed Research interest in the area of Social Network Analysis, specifically linking scalability, is increasing due to its importance. Previous research activities have revealed a number of problems with link prediction, such as dispersion, accuracy, and efficiency/performance, especially recently applied to related research and emphasizing the importance of finding a solution to the scalability predictive link problem. Fire et al. has also conducted research in which he claims that the rapid growth of social media platforms has provided a number of research guides on the properties and behaviors of major social media platforms. Additionally, you have done some work related to Fire et al., which they wanted but were very clear about as they believed that mutations formed the basis of link prediction algorithms, which included trying to understand the process of dynamic changes and trying to replicate them. In addition, it has identified seven other key elements that represent research in research, which are influence on dissemination, public or group discovery, expert

discovery, recommendation systems, behavioral and emotional analysis, predicting trust and distrust between individuals, and solicitation of ideas.

Ajay Kumar Singh Kushwah et al. (2016)""introduced social network analysis, which is a growing field of research, and the problem of link prediction shows an important role for predicting the structure of a social network. This paper emphasizes existing research on the problem of link prediction. Existing research shows that the link between the problem of predicting complexity and the solutions available to manage effective group communication and social media awareness is strong. The problem of predicting a link across all related networks can be solved by predicting anchor links and link transfers through various related networks. This paper summarizes recent developments regarding link prediction algorithms and evaluates all available link prediction methods.

Among those who have contributed to this work are Kalum P. Udagepola et al. (2016). The proposed rapid production and uncontrolled collection of social network data has raised a real problem now, because the data is large, sound, distributed, uncluttered, and powerful. As this data can be extracted using web mining techniques, social network analysis, and links to predict algorithms, in this article we try to understand the social structure and issues surrounding social network mining data. We will also look at the problems of predicting links to dynamic social networks and key strategies that can be used as a solution effort.

Ahmad Zareie et al. (2020) introduced a social network analysis recently that drew a lot of attention among researchers due to the wide range of interactions in photographic communication. Link prediction, which is related to the possibility of having a link between two offline network nodes, is a major problem in social network analysis. Several ways to solve this problem have been suggested. Between these approaches, similarly-based approaches demonstrate efficiency by considering network structure and using as a primary factor the number of common neighbors between two nodes to establish structural similarity. High structural similarity may suggest that a connection between the two nodes is likely to emerge. However, as shown in the paper, the number of common neighbors may not always be sufficient to provide general information about the similarity of a structure between two nodes. To address this, the local vector is first specified in each location. Then, a novel rating is proposed to determine the similarity of each pair of nodes based on the number of common neighbors and the correlation between local vectors of nodes. link "".

In recent years, the number of proposed social networks has increased significantly. Therefore, the analysis of social media has proved to be very useful and has attracted many researchers in this field of research. In particular, link prediction remains an important problem for social networks. Link Prediction has a wide range of apps where it is used to recommend friends on social media platforms such as Facebook and to recommend products to customers on e-commerce platforms like Amazon. This paper suggests a new algorithm for predicting links by combining the average rating. This algorithm has been used on real-time social networks and works better than the existing basic methods.

Kalpana Prajapati et al. (2020)""learned in recent days that social media is constantly evolving. People connect with each other through social apps. So link prediction is the course of research in the present era. With this link prediction, many people can communicate with each other through communication based on their common characteristics. In this area, a variety of machine learning techniques, such as supervised learning and non-supervised learning, are used to improve performance and similarly based features. Therefore, advanced learning models are used to provide better results as we do a variety of analyses of research papers. The deep neural convolution network provides better location under curve (AUC) results than other strategies. The implementation of such a thing is based on the information obtained from the data collected on various social media platforms. Learning-based algorithms are used to predict future links and provide better performance. There are still some challenges in large networks in predicting future interactions and similarities between nodes. We will therefore develop other statistics related to operators used in embedding algorithms.

Taxonomy of link predictions

Taxonomy is a systematic work associated with the classification. The proposed taxonomy consists of three components that influence link prediction: the prediction approaches, prediction features.

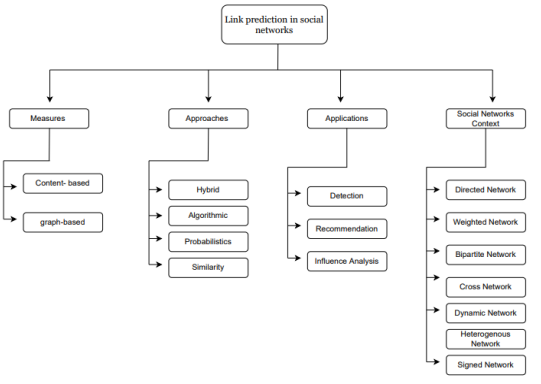

Figure 3.1: Taxonomy of link prediction

New community-based methods are included in the similarity-based approach, following the latest research trends on prediction links that utilize community information as predictive parameters. The similarity-based approaches include the local similarity-based, global similarity-based, quasi-local similarity-based, and community similarity-based methods, as shown in Figure 1.1. In addition, the prediction features are classified into data features, data representation, network features, and network categories. In turn, the measurement methods include the AUC score, accuracy, precision, Area Under the Precision-Recall Curve (AUPRC), recall, and the F1-score.

- Similarity-based approaches: The similarity-based approaches are the simplest ones among the link prediction approaches, as they give a score ranking for each unobserved pair of nodes. Similarity-based approaches can

be applied successfully to some networks, but can also fail for some networks. The proximity of similarities over unconnected node pairs is the basis of this approach. The similarity is calculated based on the determination of potential pairing node candidates that is defined as E (u, v), where u and v of unconnected node pairs are calculated as the index similarity score by using the selected similarity-based methods. The index scores are sorted from the highest to the lowest ones. The highest score corresponds to the pair of nodes with the highest possibility of generating new links or missing links.

- Learning-based approaches: The learning-based approaches are based not only on topological information like node and topology from an observed network but also on other network and data features. Classification methods serve as the basis of the learning-based approaches. Adopting probabilistic, statistical, mathematical models, and machine learning models as predictors

Approaches and Methods

Several link predication approaches have been proposed, including unsupervised approaches such as similarity measures computed on the entity attributes, random walk and matrix factorization based approaches, and supervised approaches based on graphical models and deep learning. Link prediction approaches can be divided into two broad categories based on the type of the underlying network: (1) link prediction approaches for homogeneous networks and (2) link prediction approaches for heterogeneous networks. Based on the type of information used to predict links, approaches can be categorized as topology-based approaches, content-based approaches, or mixed methods.

Topology-based methods

Topology-based methods broadly make the assumption that nodes with similar network structure are more likely to form a link.

- Common neighbors:
  This is a common approach to link prediction that computes the number of common neighbors. Entities with more neighbors in common are more likely to have a link. It is computed as follows:
  $CN(A,B) = |A \cap B|$
  A weakness of this approach is that it does not take into account the relative number of common neighbors.
- Jaccard measure:
  The Jaccard Measure addresses the problem of Common Neighbors by computing the relative number of neighbors in common
  $J(A,B) = |A \cap B| / |AUB|$
- Adamic–Adar measure:
  The Adamic–Adar measure is the sum of the log of the intersection of the neighbors of two nodes. This captures a two-hop similarity, which can yield better results than simple one-hop methods. It is computed as follows:
  $A(x,y) = \sum 1/\log|N(u)|$, $u \epsilon N(x) \cap N(y)$
  Where N(u) is the set of nodes adjacent to u.

- Katz measure:
  Neighbor based methods can be effective when the number of neighbors is large, but this is not the case in sparse graphs. In these situations it is appropriate to use methods that account for longer walks. The Katz Measure is one metric that captures this. It is computed by searching the graph for paths of length t in the graph and adding the counts of each path length weighted by user specified weights.
  Let A be the adjacency matrix of a network under consideration. Elements (a_{ij}) of A are variables that take a value 1 if a node *i* is connected to node *j* and 0 otherwise. The powers of *A* indicate the presence (or absence) of links between two nodes through intermediaries. For instance, in matrix A^3, if element ($a_{2,12}$) = 1, it indicates that node 2 and node 12 are connected through some walk of length 3. If $C_{katz}(i)$ denotes Katz centrality of a node *i*, then mathematically
  $C_{katz}(i) = \sum_{k=1}^{\infty} \sum_{=1}^{\infty} \alpha k(Ak)_{ji}$

Node attribute-based methods

Node-similarity methods predict the existence of a link based on the similarity of the node attributes.

- Euclidean distance:
  The attribute values are represented as normalized vectors, and the distance between the vectors is used to measure similarity. Small distances indicate higher similarity.
- Cosine similarity:
  After normalizing the attribute values, computing the cosine between the two vectors is a good measure of similarity, with higher values indicating higher similarity..

Mixed methods

Mixed methods combine attribute and topology based methods.

- Graph embeddings:
  Graph embeddings also offer a convenient way to predict links. Graph embedding algorithms, such as Node2vec, learn an embedding space in which neighboring nodes are represented by vectors so that vector similarity measures, such as dot product similarity, or Euclidean distance, hold in the embedding space. These similarities are the result of both topological and attribute-based similarity. One can then use other machine learning techniques to predict edges on the basis of vector similarity.
- Probabilistic relationship models:
  A probabilistic relational model (PRM) specifies a template for a probability distribution over a database. The template describes the relational schema for the domain and the probabilistic dependencies between attributes in the domain. A PRM, together with a particular database of entities and unobserved links, defines a probability distribution over the unobserved links.·.
- Probabilistic soft logic (PSL):

Probabilistic soft logic (PSL) is a probabilistic graphical model over a hinge-loss Markov random field (HL-MRF). HL-MRFs are created by a set of first-order logic-like rules that are then grounded over the data. PSL can combine attribute, or local, information with topological, or relational, information. While PSL can incorporate local predictors, such as cosine similarity, it also supports relational rules, such as triangle completion in a network.

- Markov logic networks (MLNs):
  Markov logic networks (MLNs) is a probabilistic graphical model defined over Markov networks. These networks are defined by first-order logic-like rules that are templated and then grounded over training data. For the purpose of link prediction, MLNs are able to incorporate both local and relational rules..
- R-Model (RMLs):
  R-Models (RMLs) is a neural network model created to provide a deep learning approach to the link weight prediction problem. This model uses a node embedding technique that extracts node embeddings (knowledge of nodes) from the known links' weights (relations between nodes) and uses this knowledge to predict the unknown links' weights.

Conclusion

This paper is based on a review of the social network, which predicts the direct correlation between links and estimates the unusual types of estimates for a competent link. Link prediction deals with the problem of predicting the survival of links between vertices in a social network. Predictive communication methods can provide a very effective way to get useful information from existing information. This survey is central to the modification of existing methods to overcome their shortcomings or to use the meta heuristic approach to improve the prediction accuracy of the link in order to easily find relationships between nodes.

References

1. Ajay Kumar Singh Kushwah and Amit Kumar Manjhvar, "A Review on Link Prediction in Social Network", International Journal of Grid and Distributed Computing Vol. 9, No. 2 (2016).
2. Ahmad Zareie&RizosSakellariou, "Similarity-Based Link Prediction in Social Networks using Latent Relationships Between the Users", Scientifc Reports | (2020) 10:20137.
3. Hasti Akbari Deylami, Masoud Asadpour, "Link Prediction in Social Networks Using Hierarchical Community Detection", IKT2015 7th International Conference on Information and Knowledge Technology, 978-1-4673-7485-9/15/$31.00 ©2015 IEEE.
4. Jagadishwari V, Sugato Chakrabarty, "Link Prediction using Influencer nodes of a Social Network", Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020) IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2.
5. Kalum P. Udagepola and Fatima Chiroma, "Review Social Network Analysis and Mining: Link Prediction", International Journal of Information and Electronics Engineering, Vol. 6, No. 4, July 2016.

6. Kalpana Prajapati, Harshal Shah, Rutvik Mehta, “A Survey Of Link Prediction In Social Network Using Deep Learning Approach”, International Journal Of Scientific & Technology, Research Volume 9, Issue 04, April 2020.
7. Michael Fire, Lena Tenenboim, Ofrit Lesser, Rami Puzis, LiorRokach and Yuval Elovici, “Link Prediction in Social Networks using Computationally Efficient Topological Features”, 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 978-0-7695-4578-3/11 $26.00 © 2011 IEEE.
8. S. Gupta, S. Pandey, and K. K Shukla, “Comparison analysis of link prediction algorithms in social network,” International Journal of Computer Applications, vol. 111, no. 16, February 2015.
9. Tingli Wang, Guoqiong Liao, “A Review of Link Prediction in Social Networks”, 2014 International Conference on Management of e-Commerce and e-Government, 978-1-4799-6543-4/14 $31.00 © 2014 IEEE.
10. YugchhayaDhote, Nishchol Mishra, Sanjeev Sharma, “Survey and Analysis of Temporal Link Prediction in Online Social Networks”, 978-1-4673-6217-7/13/$31.00 c 2013 IEEE.