ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.4 Issue.2 February 2018

Survey of Link Prediction Methods using Topology based Features

[1]R.Jayasree, [2]B.Sarojini
[1]PhD Scholar, Department of Computer Science Designation, Avinashilingam University for Women, Coimbatore, India.
[2]Assistant Professor, Department of Computer Science Designation, Avinashilingam University for Women, Coimbatore, India.

Abstract – **In social networks, link prediction calculates missing links in contemporary networks and new or termination links in future networks, is significant for mining and investigating the progress of social networks. In the past period, various works have been done around the link prediction in social networks. The goal of this paper is to carefully review, study and deliberate the state-of-the-art of the topology based link prediction metrics in social networks. The link prediction background and challenges are discussed. Distinctive applications of link prediction are addressed. Finally, comparative study of existing topology feature based methods are discussed.**

Index Terms: Link prediction, Social networks, Comparative study, mining.

I. Introduction

Humans are social mortals and cooperate with one another in a range of manners such that human social networks are universal in nature. Such social networks can range from offline networks based on friendship or kinship ties to online networks in social networking websites like Facebook, LinkedIn, MySpace etc. Social networks are a common way to model the connections among the people in a group or community. They can be visualized as graphs, where a vertex corresponds to a person in some group and an edge represents some form of association among the corresponding persons. The associations are usually driven by mutual interests that are intrinsic to a group [1]. However, social networks are very dynamic, since new edges and vertices are added to the graph over time. Understanding the dynamics that drive the evolution of social network is a complex problem due to a large number of variable parameters. But, a comparatively easier problem is to understand the association among two specific nodes.

Social Networks are most often characterized as graphs or hyper graphs and there is a widespread body of literature on social networks. Various predictive problems have been proposed for social networks, the link prediction problem is the problem of predicting links in a network which may form in the future among the nodes in the network. The link prediction problem was first projected by Liben-Nowell and Kleinberg [1]. The link prediction really consists of a family or sub problem e.g., predicting the existence of the link, type of the link, the strength of the link etc., as shown in fig.1. Most of the research on link prediction deliberate on problem of link existence (whether a [new] link among two nodes in a social network will exist in the future or not). This is since the link existence problem can be easily prolonged to the other two problems of link weight (links have different weights related with them) and link cardinality (more than one link among same pair of nodes in a social network). The fourth problem of link type prediction is a bit dissimilar which gives different roles to relationship among two objects.

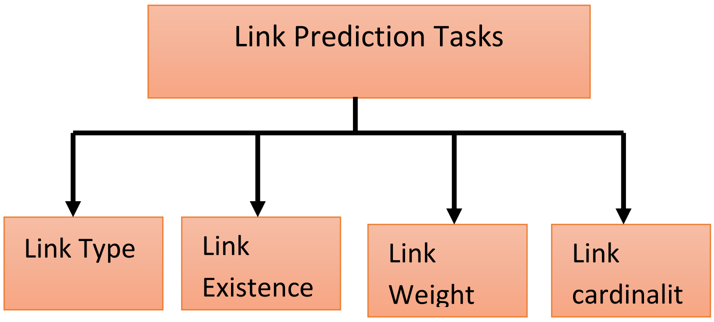

Fig. 1. Link Prediction tasks

When continuous radiation pass through a transparent material, a portion of the radiation may be absorbed. If that occurs, the residual radiation, when it is passed through a prism, yields a spectrum with gaps in it, called an absorption spectrum. The electromagnetic radiation that is absorbed has energy exactly equal to the energy difference between the excited and ground states. Generally, the most probable transition is from the Highest Occupied Molecular Orbital [HOMO] to the Lowest Unoccupied Molecular Orbital [LUMO].

II. BACKGROUND AND CHALLENGES

Liben-Nowell and Kleinberg [1] planned one of the initial link prediction models that works obviously on a social network. Every vertex in the graph represents a person and an edge among two vertices demonstrates the interaction among the persons. Multiplicity of interactions can be modeled explicitly by permitting parallel edges or by accepting an appropriate increment scheme for the edges. The learning paradigm in this system typically excerpts the similarity among a pair of vertices by several graph-based similarity metrics and practices the ranking on the similarity scores to estimate the link among two vertices. They motivated mostly on the performance of several graph-based similarity metrics for the link prediction task. By co-authorship social networks, this seminal work studies topological features for link prediction. The predictors recycled are Adamic/Adar, low-rank inner product, weighted Katz, common neighbors, Katz clustering, rooted PageRank, Jaccard, SimRank, unseen bigrams, and hitting time. The authors found that all predictors nearly always out-performed a baseline random prediction on wholly data sets. Later, Hasan et. al. [2] prolonged this work in two ways. First, they showed that using external data outside the scope of graph topology can expressively improve the prediction result. Second, they used several similarity metric as features in a supervised learning setup where the link prediction problem is demonstrated as a binary classification task. Since then, the supervised classification method has been popular in several other works in link prediction.

The research on social network evolution closely look like the link prediction problem. An advancement model predicts the future edges of a network, taking into account some well-known characteristics of social networks, such as the power law degree distribution and the small world phenomenon. This remains the main difference among development models and the link prediction models. The former focus on the global properties of the network and the latter model the local states of the network to predict the likelihood of the existence of a link among a specific pair of nodes in the network. Nevertheless, the ideas from these models have been instrumental for some research works that straight addressed the task of link prediction.

One of the main contests of link prediction concerns the development of Internet scale social networks like facebook, mySpace, Àickr, and so on. These networks are huge in size and highly dynamic in nature for which earlier algorithms may not scale and adapt well more direct methods are prerequisite to address these limitations. For instance, Tylenda et. al. [3] displays that exploiting the time stamps of past interactions, which explicitly utilize the lineage of interactions, can significantly improve the link prediction performance. Recently, Song et. al. [4] used matrix factorization to estimate similarity among the nodes in a real life social network having nearly 2 millions nodes and 90 millions edges. Any traditional algorithm that purposes to calculate pair-wise similarities among vertices of such a big graph is doomed to fail. Recently, the matrix based factorization works have been prolonged to the richer higher-order models such as tensors [5].

III. Applications Of Link Prediction

Link prediction techniques have found a large number of applications in very diverse fields. Any domain where objects interact in a structured way can possibly benefit from link prediction. Graphs provide a normal abstraction to represent interactions among diverse entities in a network. We can have graphs representing social networks, transport networks, illness networks, email / telephone calls network to list a rare. Link prediction can precisely be applied on these networks to analyze and solve interesting problems like predicting outbreak of a disease, controlling privacy in networks, detecting spam emails, suggesting alternative routes for possible navigation based on the current traffic patterns, etc.

Link prediction techniques are used to increase similar users' selection in recommender systems that follow a collaborative approach, primary to better recommendation results. A similar application is associated to social networks, which have become tremendously popular in modern society. Most social networks are using link prediction techniques to mechanically suggest connections with a high degree of accuracy. Some of the applications are illustrate in the figure 2.

Also the link prediction is appropriate in the area of Internet and web science, it can be used in responsibilities like automatic web hyper-link creation and web site hyper-link prediction. In electronic commerce, one of the most noticeable usage of link prediction is bibliography and library science, also it can be used to find out repetition and record linkage. In security associated applications, it can be used to recognize hidden groups of terrorists and criminals.

ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.4 Issue.2 February 2018

The fig. 2 shows the link prediction applications in various domain like, recommendation system, unsolicited bulk mail detection, privacy setting and control in social network, recognizing missing references in a publication, knowledge expert detection, influence detection, efficient routing in network, finally disease identification are some of the most popular application domains for link prediction [6 – 13].

Fig. 2. Link prediction applications in various domain

IV. TOPOLOGY BASED FEATURES

Even in a simple network without node or edge attributes, there are many metrics available for calculating the similarity of two nodes. Most of the metrics are constructed on the topological information are called as topology-based metrics. Liben-Nowell and Kleinberg have discoursed several metrics based on the graph structural features [1], after their work, many topology-based metrics were projected. Here, we will give a systematical explanation of popular topology-based metrics in link prediction. According to [14] the characteristics of these metrics, they can be separated into neighbor-based metrics, path-based metrics and random-walk-based metrics as shown in fig. 3. The fig. 4 shows the list of metrics which comes under the path based category with explanation. The fig. 5 shows the list of metrics which comes under the category of neighbor based metrics with explanation. Finally fig.6. shows the list of metrics which comes under the category of random walk based metrics with brief explanation.

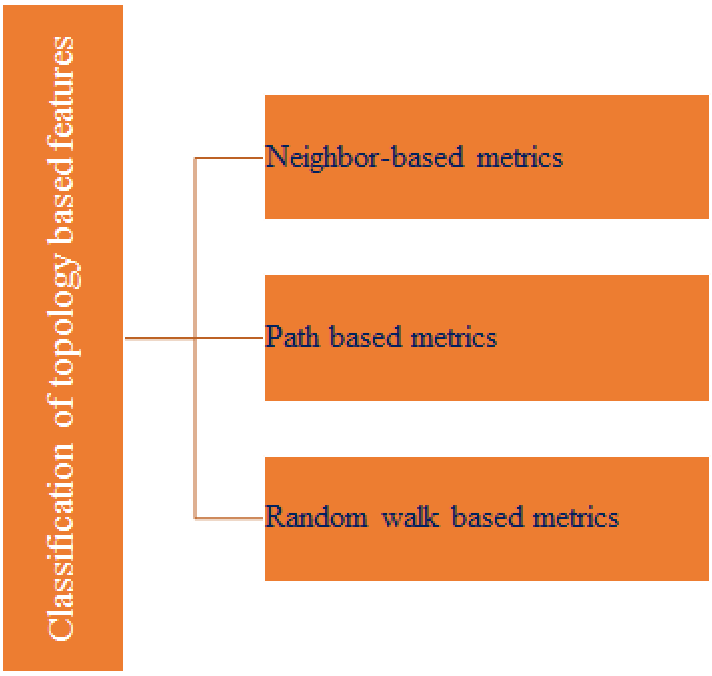

Fig. 3. Classification of topology based features

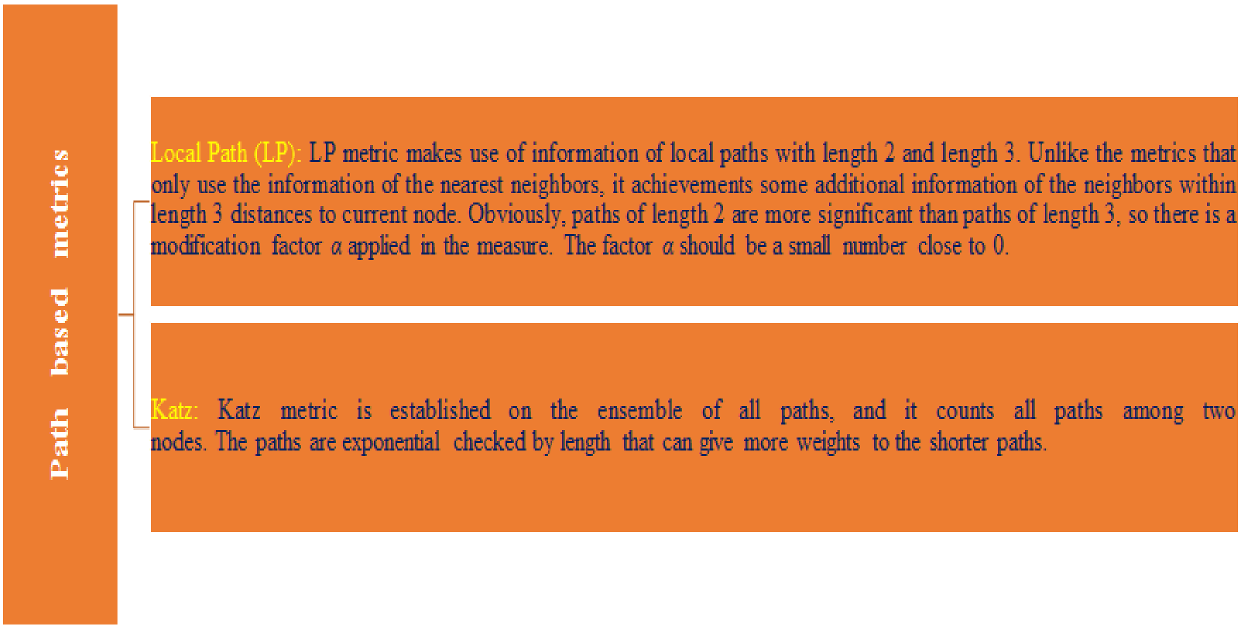

Fig. 4. List of path based metrics with explanation

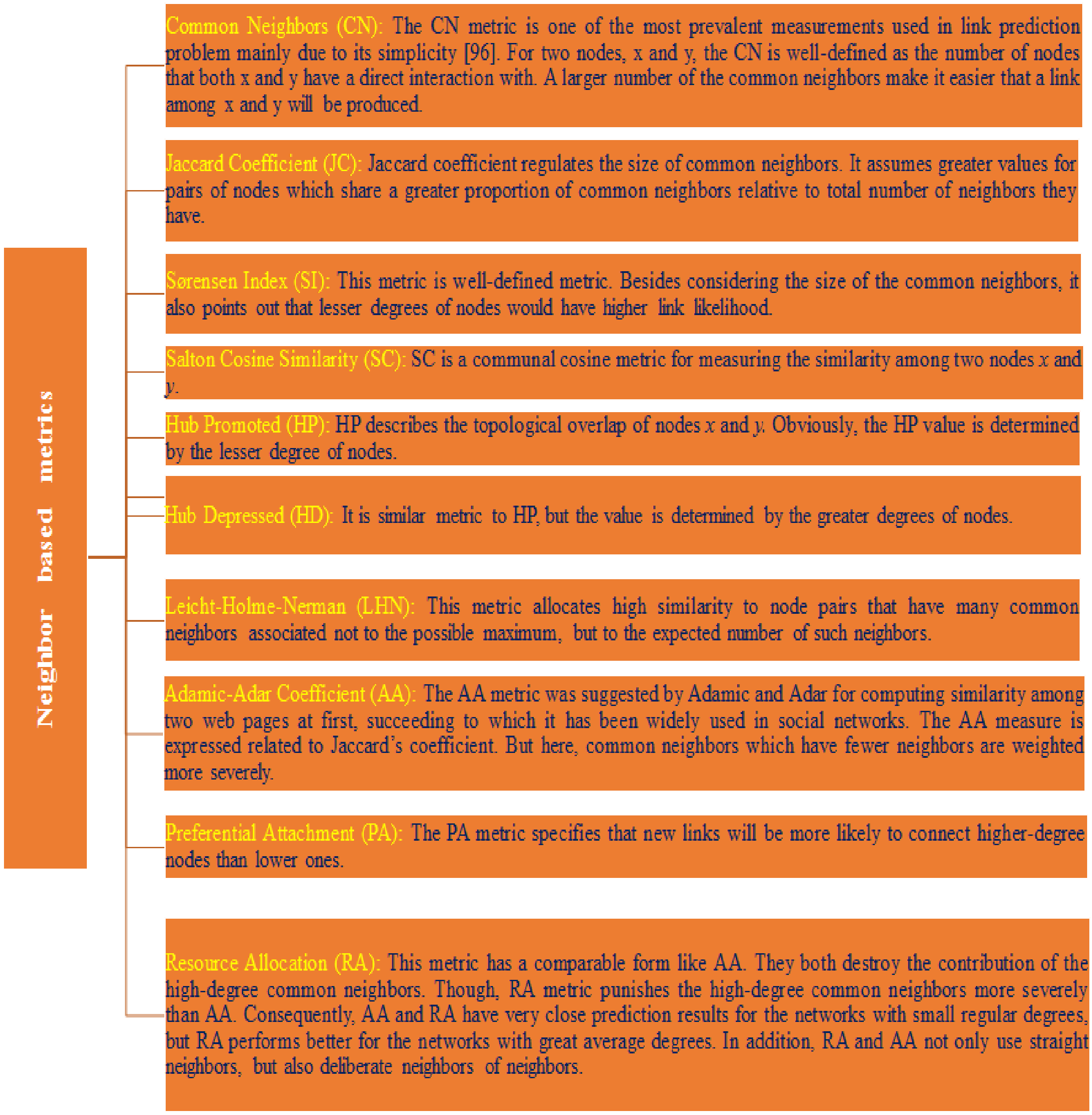

Fig. 5. List of neighbor based metrics with explanation

ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.4 Issue.2 February 2018

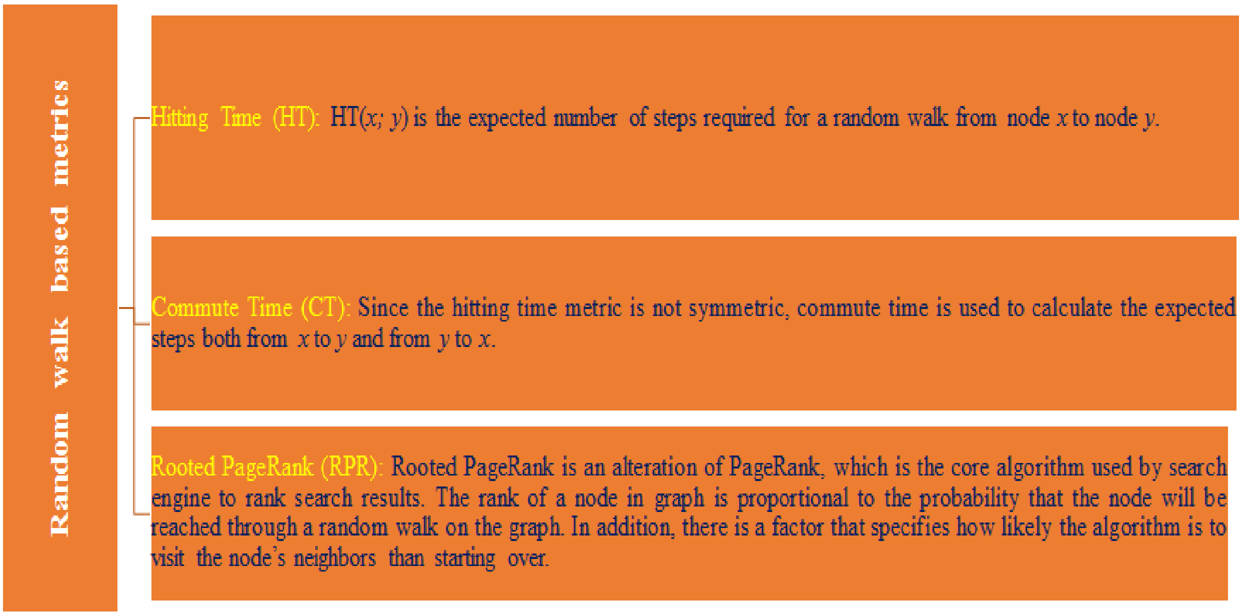

Fig. 6. List of Random-walk based metrics with explanation

V. Comparative Study Of Existing Topology Feature Based Methods

We will look at various topology feature based methods for link prediction in this section. The algorithms will work on the topology based features. The comparison is given in table1.

ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.4 Issue.2 February 2018

TABLE 1. SURVEY OF VARIOUS TOPOLOGY BASED FEATURE METHODS

Author & year	Measures	Method	Dataset
Pavlov et.al (2007) [13]	Shortest path, Common neighbours, Jaccard's coefficient, Adamic / Adar, Katz, etc.	SVM, Decision trees, J48, Decision stump, boosting	IEICE dataset Tokyo.
Wang et.al (2007) [15]	Katz, neighbors, shortest path, Adamic-Adar, Preferential Attachment measures, etc	Logistic regression	DBLP, Genetics, Biochemistry
Scripps et.al (2008) [16]	Path length, Common neighbors	Matrix alignment algorithm	DBLP, TakingItGlobal.org, Webkb
Kahanda et.al (2009) [17]	Acquard coefficient, Adamic – Adar measures	Logistic regression, bagged decision trees, naïve Bayesian classifier	Purdue Facebook dataset
Menon, A. K., & Elkan, C. (2011) [18]	Adamci-Adar, Katz measures	Unilinear and Bilinear Regression	KEGG / PATHWAY dataset, NIPS, PowerGrid, Conflict.
De Sa et. al (2011) [19]	Common neighbors, Jaccard's coefficient, Preferential attachment, Adamic / Adar, path distance,	J48, Naïve Bayesian, IBk, LibSVM, LibLinear	DBLP
Yang, X. et.al (2012) [20]	Common neighbors	Tensor-based Node Similarity (TBNS) link prediction algorithm	Autonomous system, Wiki-edit Esperanto, Facebook Wall posts
Pujari, & Kanawati, (2012) [21]	Common neighbors, Jaccard's coefficient, Adamic Adar, Preferential attachment, shortest path, katz, etc.	Supervised rank aggregation, decision tree, naïve Bayesian, kNN	DBLP
Ahmed, C., & ElKorany, A. (2015) [22]	Common Neighbors, Adamic- Adar, SimRank, FriendTNS	Similarity Algorithm	Twitter dataset
Srilatha, P., & Manjula, R. (2016) [23]	Common neighbors, Salton index, Jaccord coefficient, Sorenson Index, Hub promoted Index, Hub depressed index.	User behavior based features	Facebook
Berlusconi, G et. al (2016) [24]	Common neighbors, Katz, Resource allocation.	Similarity Score	Italian Criminal case dataset

From the table 1. One can infer that the topological based features can predict the link efficiently. The advantages of these approaches include the genericity and ease with which they can incorporate the attributes of the entities in the future model.

VI. CONCLUSION

This survey paper is attempts to analytically summarize the works which are focusing on topology based feature methods. This paper begins with background and challenges in link prediction and followed by applications of link prediction. Next focused on detailed study of topology based link prediction metrics. Finally compared the existing topology feature based methods. Based on the survey the topology features gives efficient results for predicting links in social networks. In future, we use topology features with classification algorithm to predict the links in social networks.

REFERENCES

[1] Liben-Nowell, D. and Kleinberg, J., 2007. The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology*, vol. *58, no.* 7, pp.1019-1031.

[2] Al Hasan, M., Chaoji, V., Salem, S. and Zaki, M., 2006, April. Link prediction using supervised learning. In *SDM06: workshop on link analysis, counter-terrorism and security*.

[3] Tylenda, T., Angelova, R. and Bedathur, S., 2009, June. Towards time-aware link prediction in evolving social networks. In ACM Proceedings *of the 3rd workshop on social network mining and analysis*, p. 9.

[4] Song, H.H., Cho, T.W., Dave, V., Zhang, Y. and Qiu, L., 2009, Scalable proximity estimation and link prediction in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pp. 322-335.

[5] Acar, E., Dunlavy, D.M. and Kolda, T.G., 2009, December. Link prediction on evolving data using matrix and tensor factorizations. *IEEE International Conference on Data Mining Workshops- ICDMW'09*, pp. 262-269.

[6] Gimenes, G.P., Gualdron, H., Raddo, T.R. and Rodrigues, J.F., 2014, March. Supervised-learning link recommendation in the DBLP co-authoring network. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on* (pp. 563-568). IEEE.

[7] Esslimani, I., Brun, A. and Boyer, A., 2011. Densifying a behavioral recommender system by social networks link prediction methods. *Social Network Analysis and Mining*, vol. *1, no.* 3, pp.159-172.

[8] Huang, Z. and Zeng, D.D., 2006, October. A link prediction approach to anomalous email detection. *IEEE International Conference on Systems, Man and Cybernetics - SMC'06.* Vol. 2, pp. 1131-1136.

[9] Al-Oufi, S., Kim, H.N. and El Saddik, A., 2011, August. Controlling privacy with trust-aware link prediction in online social networks. In *Proceedings of the ACM Third International Conference on Internet Multimedia Computing and Service*, pp. 86-89.

[10] Kc, M., Chau, R., Hagenbuchner, M., Tsoi, A.C. and Lee, V., 2009, A machine learning approach to link prediction for interlinked documents. In Springer *International Workshop of the Initiative for the Evaluation of XML Retrieval*, pp. 342-354.

[11] Pavlov, M. and Ichise, R., 2007, Finding experts by link prediction in co-authorship networks. In *Proceedings of the 2nd International Conference on Finding Experts on the Web with Semantics, Vol. 290,* pp. 42-55.

[12] Perez-Cervantes, E., Mena-Chalco, J.P., De Oliveira, M.C.F. and Cesar, R.M., 2013, October. Using link prediction to estimate the collaborative influence of researchers. *IEEE 9th International Conference,* pp. 293-300.

[13] Yadav, A., Singh, Y.N. and Singh, R.R., 2015. Improving routing performance in AODV with link prediction in mobile Adhoc Networks. *Wireless Personal Communications*, vol. *83, no.* 1, pp.603-618.

[14] Wang, P., Xu, B., Wu, Y. and Zhou, X., 2015. Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, *58*(1), pp.1-38.

[15] Wang, C., Satuluri, V. and Parthasarathy, S., 2007, October. Local probabilistic models for link prediction. *Seventh IEEE International Conference on Data Mining,* pp. 322-331.

[16] Scripps, J., Tan, P.N., Chen, F. and Esfahanian, A.H., 2008, December. A matrix alignment approach for link prediction. *19th International Conference on Pattern Recognition, 2008.* pp. 1-4.

[17] Kahanda, I. and Neville, J., 2009. Using Transactional Information to Predict Link Strength in Online Social Networks. *ICWSM*, *9*, pp.74-81.

[18] Menon, A.K. and Elkan, C., 2011, Link prediction via matrix factorization. In *Joint european springer conference on machine learning and knowledge discovery in databases*, pp. 437-452.

[19] De Sá, H.R. and Prudêncio, R.B., 2011, July. Supervised link prediction in weighted networks. In *2011 IEEE International Joint Conference on Neural Networks,* pp. 2281-2288.

[20] Yang, X., Tian, Z., Cui, H. and Zhang, Z., 2012, October. Link prediction on evolving network using tensor-based node similarity. *IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS), 2012,* Vol. 1, pp. 154-158.

[21] Pujari, M. and Kanawati, R., 2012, November. Link prediction in complex networks by supervised rank aggregation. In *IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI), 2012*, Vol. 1, pp. 782-789.

[22] Ahmed, C. and ElKorany, A., 2015, August. Enhancing link prediction in twitter using semantic user attributes. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* pp. 1155-1161.

[23] Srilatha, P. and Manjula, R., 2016, August. User behavior based link prediction in online social networks. In IEEE *International Conference on Inventive Computation Technologies (ICICT),* Vol. 1, pp. 1-3.

[24] Berlusconi, G., Calderoni, F., Parolini, N., Verani, M. and Piccardi, C., 2016. Link prediction in criminal networks: a tool for criminal intelligence analysis. *PloS one*, vol. *11, no.* 4, p.e0154244.