Online-Journals.org

iTDAF IETI Transactions on Data Analysis and Forecasting

iTDAF | eISSN: 2959-0442 | Vol. 1 No. 2 (2023) | OPEN ACCESS

https://doi.org/10.3991/itdaf.v1i2.39675

PAPER

Link Prediction in Human Complex Network Based on Random Walk with Global Topological Features

Syed Shah Hussain[1], Muhammad Arif[2,3](✉), Osama Bin Inayat[1], Haji Gul[1]

[1]Department of Computer Science, City University of Science and Information Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan

[2]Fixed Point Research Laboratory, Fixed Point Theory and Applications Research Group, Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand

[3]Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand

arifmaths667@gmail.com

ABSTRACT

Link Prediction in Human Complex Networks aims to predict the missing, deleted, or future link formations. These complex networks are represented graphically, consisting of nodes and links, also referred to as vertices and edges, respectively. We employ Link Prediction techniques on four different human-related networks to determine the most effective methods in the Human Complex domain. The techniques utilized are similarity-based, primarily focused on determining the similarity score of each network. We select four algorithms that demonstrated superior results in other complex networks and implement them on human-related networks. Our goal is to predict links that have been removed from the network in order to evaluate the prediction accuracy of the applied techniques. To accomplish this, we convert the datasets into adjacency matrices and divide them into training and probe sets. The training session is then conducted, followed by the testing of the data. The selected techniques are implemented to calculate the similarity score, and the accuracy is subsequently measured for each dataset. This approach facilitates a comprehensive comparative analysis of the various predicting techniques to determine the most effective one.

KEYWORDS

human complex network (HCN), link prediction (LP), area under the curve (AUC), preferential attachment (PA), resource allocation (RA)

1 INTRODUCTION

In the context of network theory, a complex network refers to a network graph with intricate practical topological structures that are not typically present in simple networks. These structures are matrices or graphs, are frequently observed in complex networks. Essentially a complex network can be defined as a vast assemblage of interconnected nodes. A node can represent various entities, including individuals, organizations, computers, biological cells, and so forth. The term "connected" signifies the ability of two nodes to be linked together within the network.

Hussain, S.S., Arif, M., Inayat, O.B., Gul, H. (2023). Link Prediction in Human Complex Network Based on Random Walk with Global Topological Features. *IETI Transactions on Data Analysis and Forecasting (iTDAF)*, 1(2), pp. 30–43. https://doi.org/10.3991/itdaf.v1i2.39675

Article submitted 2023-03-18. Resubmitted 2023-05-05. Final acceptance 2023-05-17. Final version published as submitted by the authors.

For instance, in a complex network, two individuals may be connected because they know each other, two groups of companies or organizations may exchange goods, two computers may be linked through a cable, or two neurons' systems may be connected through synaptic connections to transmit signals. These networks are considered complex due to their vast size, making it impossible to comprehend or predict their overall performance by solely examining the performance of individual nodes or links.

The network models can be represented in the form of graphical models [2]. These graphical models can map nodes to social entities or individual persons, while the links represents the relations between these nodes [27]. A complex network is composed of nodes and links, where the nodes can be biological or individual entities, and the links indicate the relationships between these nodes [16][34]. The complex network has numerous applications, spanning across various fields in the real world. It has worked in the legislative drives to influence the maximum number of citizens [7][25]. Furthermore, it is utilized for the development of road networks, aiming to improve transportation routes [31]. The study of these networks is highly dynamic, constantly expanding as new edges and vertices are added over time [23]. The study of these networks has experienced significant growth in recent few years and has found practical applications. Once the networks are modelled, they can provide insights into various questions pertaining to complex networks [10].

Because of the increasing size of these networks, one emerging issue related to complex networks is the problem of missing link prediction (LP). LP revolves around the concept of potential links existing between nodes within complex networks [6]. The objective of LP is to estimate the likelihood of the link existence between pairs of nodes that are not currently connected [5]. Solving of the challenge of predicting links in various complex networks, such as protein-protein interactions in biological networks, but also enable the prediction of future link evolution within these nodes or networks [24][33].

A basic assumption regarding the LP is that two similar nodes are more likely to have a link if they share similarities or similarity. This raises the question of how to calculate the similarity between nodes. The CN method has traditionally been used to address this issue; however, it has several limitations, with one of the most common limitation being its requirement for a larger number of nodes [14]. In order to overcome this limitation, many variants of the CN, such as the Jaccard Index [15] have been employed to mitigate this issue. In addition, other methods such as Katz Index [15] have proven effective in calculating node similarity. When dealing with large real-world datasets, this research strives to deliver the most accurate and efficient LP approaches. This has several advantages, particularly in the context of Human Complex Networks. It allows for comparative analysis of different LP techniques, with the goal of identifying the most effective LP technique in this domain.

2 PROBLEM OVERVIEW

One of the major challenges in LP is that complex networks are often both incomplete and dynamic. This poses a problem as nodes in the network can appear and disappear over time. Unlike static networks, where nodes remain static, real-world networks exhibit dynamic behavior, where links can change dynamically over time. To illustrate, consider the graph shown in Figure 1, representing the nodes $G(V, E)$ with respect to time. V and E represent nodes and links, respectively. The LP problem focuses on predicting the future and missing or deleted links in a network.

For instance, consider the scenario where Sophia is a friend of Maria and Maria is a friend of Adam. At a later time *t*′, it is possible that Maria introduces Sophia to Adam, resulting in a new friendship between them. Therefore, the objective of LP is to predict the emergence of such connections. While static networks pose a different challenge, real-world networks introduce additional complexities that make LP more challenging [32].

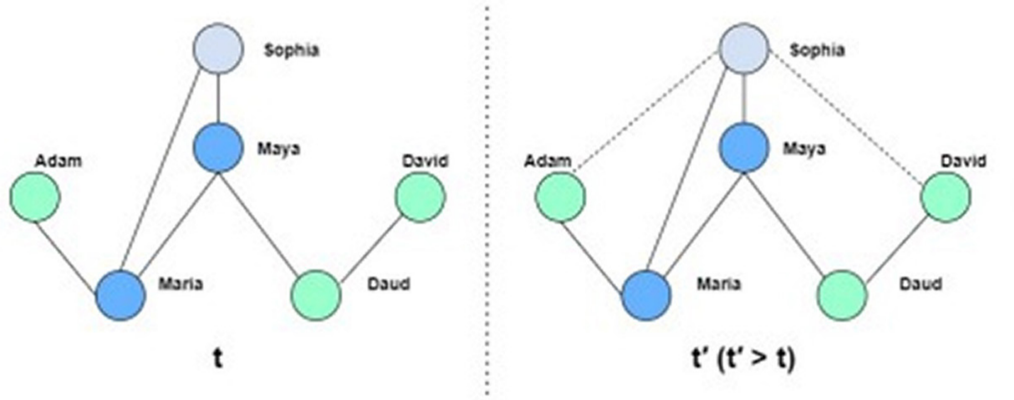

Fig. 1. Physical sketch of common neighbour

3 MOTIVATION

LP has emerged as a prominent issue in network sciences, applicable to various real-world networks such as individual connections, biological elements, transportation systems, and social networks. The primary purpose of LP is to predict missing or future links within complex networks, enabling better decision-making. One notable example is Facebook, one of the most popular social media platforms, where nodes represent users or other social activities, and links represent connections as friendship, likes, and tagging. In this context, LP is utilized to predict when two users are likely to interact. Business revenues are often generated based on the links within a network, where promotional ads are presented to individuals with similar interests. Additionally, friend suggestions can be made based on shared interests or liking in the ads [22]. In the context of Protein-Protein Interaction, a biological network, proteins are represented as nodes, and the links represents interactions between them. LP techniques enable prediction of interactions among pairs of protein [13].

4 APPLICATIONS OF LINK PREDICTION

LP techniques have found extensive applications across various fields. Any field in which entities interact with each other can benefit from LP. Some of the widely used applications of LP are as follows:

4.1 The detection of anomalous E-mail

Anomalous e-mail detection involves analyzing the communication patterns that occur within a network-based structure. These communications can be either one-to-one or one-to-many. The LP approach is used to identify one-to-many communication patterns and detect anomalies. Huang and Zeng proposed an approach that considers e-mail communication as a network, with nodes representing users, and LP techniques are employed to predict future communications [27].

4.2 Predicting co-participation in an event

Social network analysis involves studying persistent relationships and discrete events within a network. The data in social networks is dynamic over time, and participation in events can be predicted based on these relationships and events. For instance, a formalized paraphrase of the prediction question could be, "how likely is it that X will call Y next week?" [17].

4.3 The recommendation of items to the users

A recommendation system is employed to suggest items to users based on a bipartite network. This network consist of user-tags, item-tags, and other interconnected networks. The LP technique is applied to these networks, enabling the recommendation system to make accurate suggestions [12].

4.4 The proteome of yeast

A protein function utilizes LP techniques to improve protein interaction analysis. It uses a role-similarity measure and, following testing, is applied to the proteome of the yeast [34].

5 RELATED WORK

Although LP is a relatively new research topic, there has been notable progress. Linyuan Lu and Tao Zhou conducted a survey in 2011, utilizing LP approaches were on various datasets. They employed local similarity, global similarity, and quasi local similarity as similarity indices, along with maximum likelihood methods and probabilistic models. Precision and accuracy were used as evaluation criteria. The study's assessment demonstrated that the similarity indexes effectively performed in the specified LP scenario, as revealed by the findings [30]. In 2014, LP approaches focusing on the role of networks topology were explored frp, an information theory perspective. Fei Tan, Yongxiang Xia, and Boyao Zhu employed data mining techniques and utilized accuracy as an evaluation criteria. The findings, a led to the implementation of a reciprocal information method [21]. In 2015, the Pearson correlation coefficient approach was introduced and proved successful in analyzing high-order pathways. Data mining techniques were applied to nine empirical networks [28]. In 2016, another survey focused on the computational complexity analysis of various methodologies. Data mining approaches were used, and applied to seven empirical networks representing diverse perspectives and backgrounds. Accuracy and precision were used as evaluation metrics. The survey included a comparison with the old techniques, assessing their effectiveness [25].

In 2017, a rapid technique for solving the LP problem was developed in the field of graph theory. The focus extended beyond predicting emerging network edges, recognizing that LP can offer more. Data mining approaches were employed, and accuracy and precision served as evaluation metrics [20]. In 2018, researchers used a multi-attribute decision making method to identify missing links in a network. This study introduced a novel approach that was tested on ten real-world networks. Accuracy was utilized as a criterion for evaluation, and the Multiple Attribute Decision Making approach was employed to solve the issue [16]. In 2019, a survey was conducted to explore LP techniques, applications, and performance within the field of graph theory. Data mining techniques were applied, and the accuracy of eight different datasets served as an evaluation metric [16]. Additionally, in the same year, both real-world and synthetic networks were analyzed using common neighbor degree penalization. Precision and accuracy were employed as evaluation metrics and the research primarily focused on addressing two key issues: big data and low computing complexity [29].

In 2020, a study on LP focused on local path research within the field of graph theory was conducted. Data mining techniques were applied to 12 separate datasets, using accuracy and precision as evaluation metrics. The results demonstrated improved accuracy compared to previous approaches [4]. In the same year, a method for addressing missing LP utilizing common neighbor and centrality-based parameterized algorithm was introduced. This approach was applied on eight networks. Accuracy and precision served as evaluation metrics, revealing that accuracy outperforms precision when employing the technique [2]. In 2021, several new link predictors were introduced, leveraging the advantages of multiple known algorithms simultaneously [3][9]. Comparative studies have also been published [8][26], although they primarily focus on general comparisons across various complex networks rather than exclusively on Human complex networks.

5.1 Jaccard Index

The Jaccard Index is a local similarity index, similar to CN, but it is more efficient as it normalizes the score. It is defined as the likelihood of pair-wise selection. It is made up of vertices from node neighbors and is written as:

$$S_{ab}^{Jaccard} = \frac{|\Gamma(\mathrm{a}) \cap \Gamma(\mathrm{b})|}{|\Gamma(\mathrm{a}) \text{ U } \Gamma(\mathrm{b})|}$$

where (a) is the set of node a’s neighbors [1].

5.2 Preferential Attachment Index (PAI)

The local similarity index, also known as the Preferential Attachment (*PA*), is determined by the degree of nodes and *b*. It is defined as the likelihood that a link exists at nodes ‘*a*’ and ‘*b*’ being proportional to their respective degrees, k_a and k_b [19]. It is expressed as

$$S_{ab}^{PA} = k_a \times k_b$$

5.3 Index of Resource Allocation (RA)

The Resource Allocation (*RA*) is a local similarity index that utilizes the concept of intermittent node connecting nodes *x* and *y*. It quantifies the number of resources allocated between the two nodes. In this case, node *x* is indirectly connected to node *y* through this intermittent node. This can be expressed as

$$S_{ab}^{RA} = \sum_{z \in \Gamma(a) \cap \Gamma(b)} \frac{1}{|\Gamma(z)|}$$

where k(z) is the degree of node z [18].

5.4 Index Katz

The Katz Index is a global similarity index that is considered a metric form of the shortest path. It operates by aggregating pathways between nodes *x* and *y* and exponentially dampening longer paths. This index combines the utilization of global information and CN index [15], as discussed in this paper.

$$S_{ab}^{Katz} = \sum_{l=1}^{\infty} \beta^{1}.\left|Path^{l}(xy)\right| = \beta A + \beta^{2} A^{2} + \beta^{3} A^{3} + \ldots$$

6 PROPOSED METHODOLOGY

A graph or a network G is defined as an ordered pair, G = (V, E), where V stands for nodes or vertices and E stands for connections or edges. An edge, ei,j represents a link between two nodes i and j. The number of nodes in the graph is represented by |V|, which is also referred to as the network's size. |E| stands for the number of linkages. In the context of LP, undirected graphs are considered, which means that the connections are represented as edges rather than arcs. The edge set is composed of elements from the vertex set, and is written as e = (v, u) E, with (v, u) V. The standard LP formulation involves splitting the edge E into two halves, E T and E P, which represent the Training Set and Probe Set, respectively, with the condition that E T E P = E. The neighbour of x is the collection of nodes connected by the edge, represented by x V, and can be denoted as x. The degree of node x in an undirected graph is denoted as |x|.

The acquisition of datasets serves as the initial step in the comprehensive study technique. In this case, the datasets were numeric and obtained in raw form. To make the data more interpretable, a program, specifically MATLAB, was utilized. The datasets were initially organized in an array within MATLAB. Subsequently, a graphical representation of each dataset was generated, depicting the links between the nodes. This graphical depiction facilitated the creation of an adjacency matrix, which captured the relationships between the nodes.

There were two cases: either the link exists (denoted by 1) or it does not exist (denoted by 0), which is known as the 'Adjacency Matrix' in matrix form. Afterwards, the Adjacency matrix was split into Training and Probe sets. Different division classes were used to split training and probe sets, such as 90–10, 80–20, 70–30, 60–40, and 50–50. The implementation of each division classes provided specific results, which were then compared, and the AUC value closest to '1' was chosen for further processing. After deciding on the train-probe split percentage, each of the strategies was used to forecast the phenomena. The accuracy of each technique on each dataset was determined using post-prediction analysis. Finally, the findings were compared, and the most effective prediction method was determined. Figure 2 depicts the research approach in its entirety.

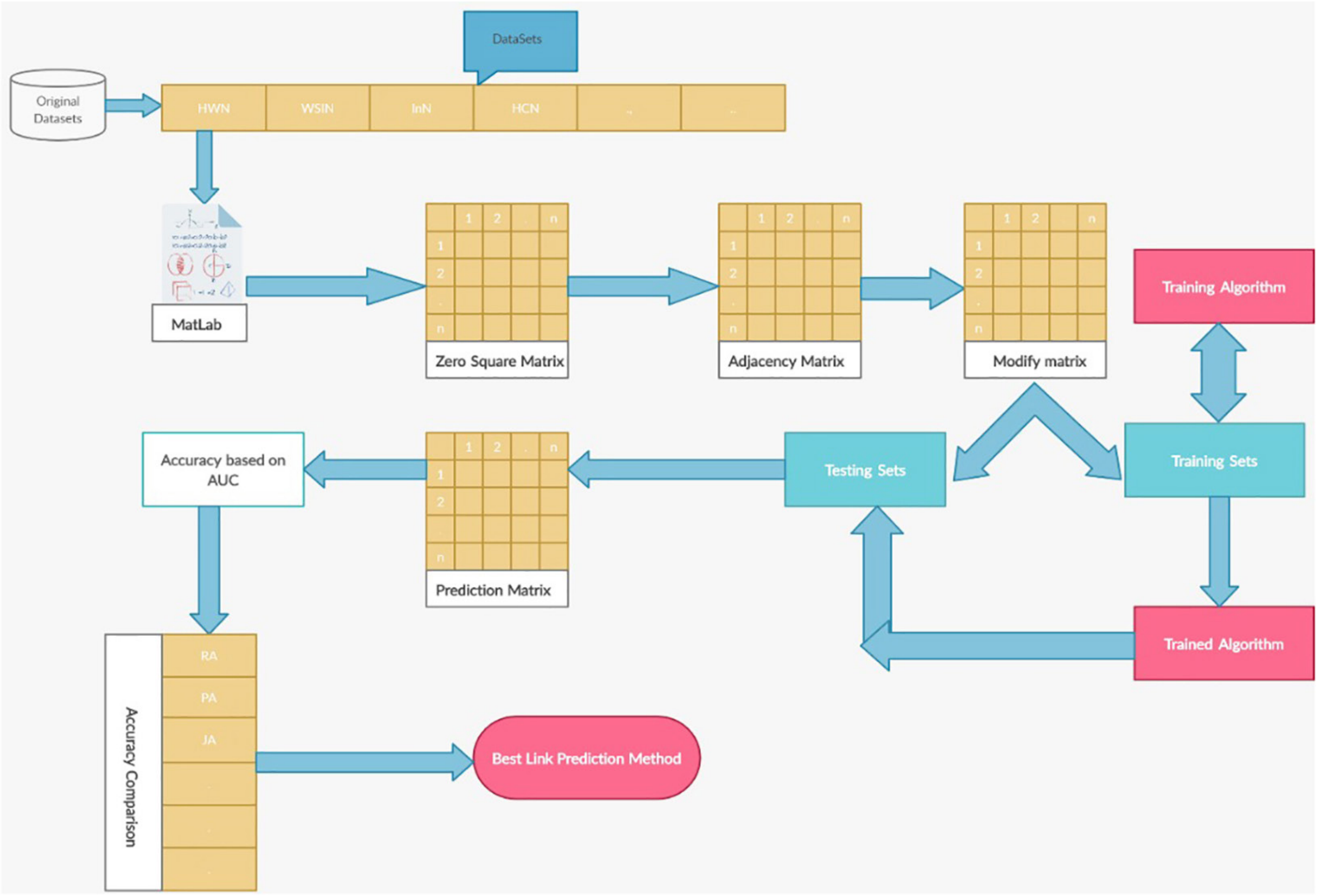

Fig. 2. Methodology

6.1 Datasets

The real-world datasets used for this research are listed as follows. Most of the complex network datasets are taken from, http://konect.uni-koblenz.de/networks/, and https://snap.stanford.edu/data/

a. *Human Contact Network.* A network in which the humans are considered as nodes or vertices and the contact such as calls, messages are considered as the links or edges.
b. *Human Wireless Contact Network.* A network in which the humans are considered as nodes or vertices and the wireless contact among them are considered as links or edges.
c. Women Social Event Interaction Network. A network which is obtained from an event of women where their interaction is noted and the women are considered as nodes or vertices and the interaction among them like talk or meet is considered as links or edges.

d. *Karate Network*. A network obtained from A Karate Club, where the members are considered as nodes or vertices and the interaction among them can be considered as links or edges.

The vertices, edges, average degree of nodes and average density of the above mentioned datasets are given in Table 1.

Table 1. Dataset details

Datasets	$\|V\|$	$\|E\|$	<*k*>	<*d*>
Human Contact Network	43	336	15.628	0.8
Human Wireless Contact network	274	2124	15.5	0.41
Karate Network	34	78	4.588	0.137
Woman Social Event Network	18	63	7	0.592

7 EVALUATION

The area under the curve (AUC) has been widely used as a measure of prediction accuracy [11]. In our study, we use AUC as a measure of accuracy for LP. The performance score Sx.y, is recorded by incorporating information from E T. The prediction score is then compared with n pairs of nodes randomly from E P and E¯. If the score measured from E P is greater than E¯, it is denoted as n0. If E P is equal to E¯, it is denoted as n00. The AUC can be calculated by AUC = (n 0 + 0.5n 00) n Working for AUC. For better understanding, we exemplify the AUC using the LP technique, i.e., the PA Index. From Figure 3, set (a) represents the actual network, set (b) and (c) corresponds to the training and probe sets, respectively, and set (d) includes the non-observed links. We then chose BD from the probe set and a pair of nodes from non-observed links. It is important to note that only the training set is used for node degree computation. For example, we choose an AB pair from non-observed links. Now using the PA Index, the similarity between pairs of nodes BD is ScoreAB = kA × kB = 0 and CD is ScoreAB = kA × kB = 0, the ScoreBD = ScoreAB hence n0 = 0 and n00 = 1, so

$$\mathrm{AUC} = \frac{0 + (0.5 \times 1)}{1} = 0.5$$

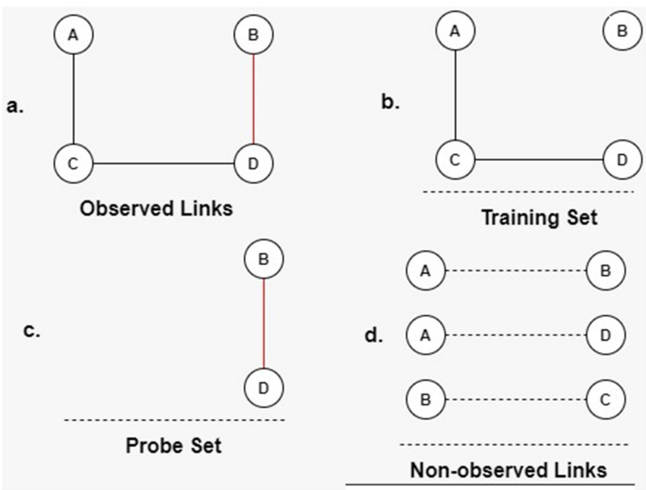

Fig. 3. Flow chart

8 RESULT AND CONCLUSION

The research aims to determine the most effective LP techniques among four options in the context of real-world networks in the Human Complex domain. The experimentation demonstrates that LP techniques can successfully identify missing or future links when applied to HCN. The experiment involved varying percentages of Probe Set (E P) in order to obtain the AUC more precisely. To achieve the objective, the experiment included multiple sets of training and probe sets with varying percentages: 10%, 20%, 30%, 40%, and 50%. The results obtained from each set were then compared to determine the one with the higher AUC value. You can refer to Tables 2–6 for detailed information on the comparative results. Figures 4–8 provide graphical representations of each dataset. Based on the observations of the figures and tables, it is noted that AUC values obtained for 10% probe sets were close to 1. As a result, the 10% probe set was selected for further analysis. Specifically Table 2 representing the 10% probe set, along with Figure 8 as the corresponding graphical representation, demonstrate that the RA Index consistently provides results closer to '1' for each dataset. Therefore, based on these findings, it can be concluded that the RA Index outperforms other LP techniques in the human complex domain in terms of efficiency.

Table 2. Average AUC results of 90–10

Techniques	Human Contact Network	Human Wireless Contact Network	Karate Network	Woman Social Event Network
Jaccard Index	0.8383	0.8892	0.5946	0.7982
Katz Index	0.83	0.9155	0.7209	0.8067
Preferential Attachment Index	0.6974	0.9307	0.7048	0.7005
Resource Allocation Index	0.858	0.9327	0.7323	0.8463

Table 3. Average AUC results of 80–20

Techniques	Human Contact Network	Human Wireless Contact Network	Karate Network	Woman Social Event Network
Jaccard Index	0.8103	0.8886	0.5948	0.7678
Katz Index	0.8237	0.9074	0.6901	0.7968
Preferential Attachment Index	0.7068	0.9344	0.6739	0.6892
Resource Allocation Index	0.841	0.9328	0.6891	0.8215

Table 4. Average AUC results of 70–30

Techniques	Human Contact Network	Human Wireless Contact Network	Karate Network	Woman Social Event Network
Jaccard Index	0.7853	0.8862	0.5896	0.7259
Katz Index	0.8077	0.9042	0.6773	0.7433
Preferential Attachment Index	0.7025	0.9301	0.6775	0.6713
Resource Allocation Index	0.8204	0.9266	0.6633	0.7886

Table 5. Average AUC results of 60–40

Techniques	Human Contact Network	Human Wireless Contact Network	Karate Network	Woman Social Event Network
Jaccard Index	0.7558	0.8822	0.5779	0.6919
Katz Index	0.788	0.8944	0.6309	0.7317
Preferential Attachment Index	0.693	0.9241	0.6564	0.6623
Resource Allocation Index	0.7913	0.9198	0.6233	0.7363

Table 6. Average AUC results of 50–50

Techniques	Human Contact Network	Human Wireless Contact Network	Karate Network	Woman Social Event Network
Jaccard Index	0.7215	0.8732	0.5673	0.6633
Katz Index	0.7566	0.8796	0.611	0.7028
Preferential Attachment Index	0.6827	0.9249	0.6323	0.6559
Resource Allocation Index	0.7524	0.9107	0.596	0.7031

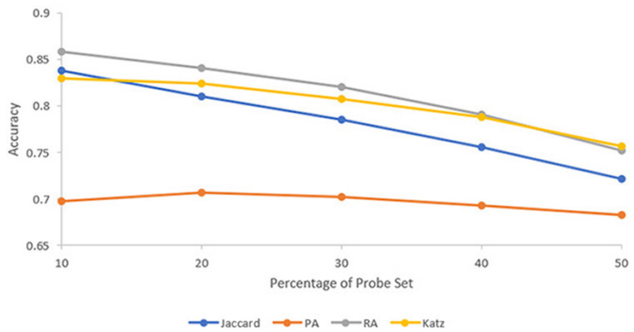

Fig. 4. Human contact network percentage results

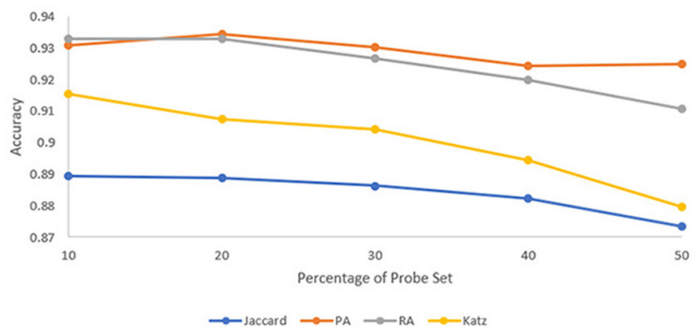

Fig. 5. Human wireless contact network percentage result

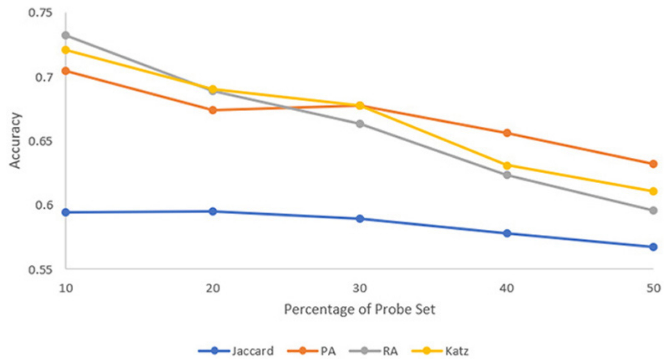

Fig. 6. Karate network percentage results

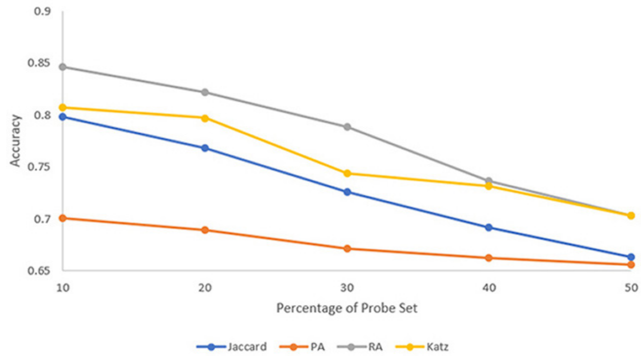

Fig. 7. Woman social event network percentage results

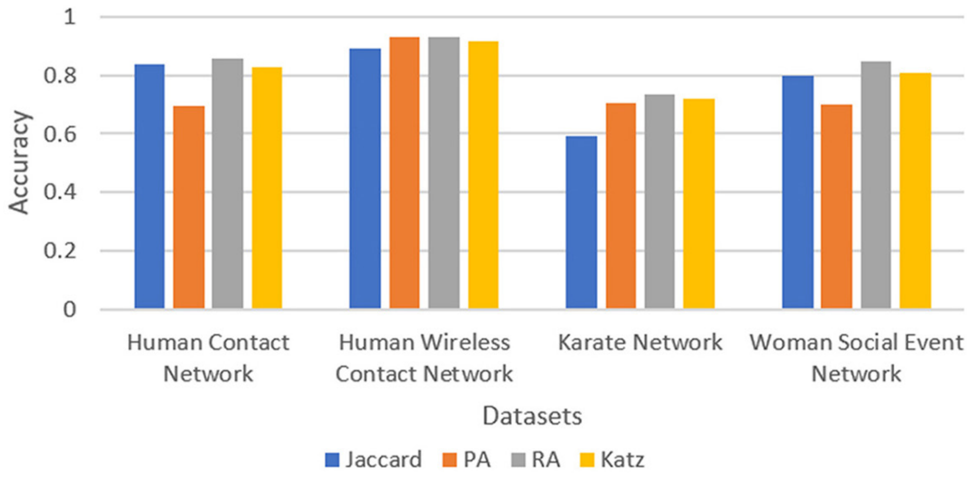

Fig. 8. Average AUC

9 CONCLUSION

Link prediction has emerged as a significant research topic and has garnered considerable attention over the past two decades. Its applications span across various disciplines, benefiting diverse fields. LP in the context of real-world complex networks has proven valuable in analyzing missing links in biological elements such as PPI, Yeast Proteome, Gene Nature, as well as in computer science applications like hyper-links navigation prediction, website predictions, and social elements including Facebook and Twitter. Furthermore, LP has also demonstrated progress in the field of sports by providing future predictions, such as score prediction in cricket. The focus of this paper was specifically on covering real-world networks in the domain of HCNs. Considerable efforts have been made to predict missing, deleted, or future links in various network. However, certain HCNs remain largely unexplored, presenting opportunities for further advancements in LP within this domain. This paper aims to fill this gap by providing a comprehensive investigation into 'Link Prediction in Human Complex Networks.' The research entails a comparative analysis of selected LP techniques, complemented by an extensive literature review to gain insights into prior work in the field of LP. The Complex Networks and LP were explained so as to provide an easier way to understand the challenges and complexities associated with the research.

This paper provides the workflow of the whole research. However, there are some limitations that were left unexplored in this study. One such limitation is the consideration of network size. Predicting links in large network can be computationally intensive and time-consuming, requiring significant computational power. Besides this, the concept of directed graphs was not used, which is another limitation of the study.

10 REFERENCES

[1] L.A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003. https://doi.org/10.1016/S0378-8733(03)00009-1

[2] M.U. Akthar Ahmad, S. Noor and A. Shahnaz, "Missing link prediction using common neighbor and centrality based parameterized algorithm," *Scientific Reports*, vol. 10, no. 364, pp. 1–9, 2020. https://doi.org/10.1038/s41598-019-57304-y

[3] F. Aziz, H. Gul, I. Uddin and G.V. Gkoutos, “Path-based extensions of local link prediction methods for complex networks,” *Scientific Reports,* vol. 10, p. 19848, 2020. https://doi.org/10.1038/s41598-020-76860-2

[4] F. Aziz, H. Gul, I. Muhammad and I. Uddin, “Link prediction using node information on local paths,” *Physica A: Statistical Mechanics and its Applications*, vol. 557, no. 1, p. 124980, 2020. https://doi.org/10.1016/j.physa.2020.124980

[5] L. Backstrom and J. Leskovec, “Supervised random walks: predicting and recommending links in social networks,” In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 635–644, 2011. https://doi.org/10.1145/1935826.1935914

[6] M. Brasoveanu, M. Moodie and R. Agrawal, “Textual Evidence for the Perfunctoriness of Independent Medical Review,” In KDD 2020: Virtual Event / San Diego, CA, USA, vol. 2657, pp. 1–9, 2020.

[7] N.Z. Gong, A. Talwalkar, L. Mackey, L. Huang, E.C.R. Shin, S. Stefanov, E. Shi, and D. Song, “Joint link prediction and attribute inference using a social-attribute network,” *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 2, pp. 1–20, 2014. https://doi.org/10.1145/2594455

[8] H. Gul, A. Amin, A. Adnan and K. Huang, “A systematic analysis of link prediction in complex network,” *IEEE Access,* vol. 9, pp. 20531–20541, 2021. https://doi.org/10.1109/ACCESS.2021.3053995

[9] H. Gul, A. Amin, F. Nasir, S.J. Ahmad and M. Wasim, “Link prediction using double degree equation with mutual and popular nodes,” In *World Conference on Information Systems and Technologies*, vol. 1368, pp. 328–337. Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-72654-6_32

[10] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang and R. Zadeh, “WTF: the who to follow service at Twitter,” In: *WWW ’13: Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 505–514. https://doi.org/10.1145/2488388.2488433

[11] A. Hanley and B.J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve.” *Radiology*, vol. 143, no. 1, pp. 29–36, 1982. https://doi.org/10.1148/radiology.143.1.7063747

[12] P. Holme and M. Huss, “Role-similarity based functional prediction in networked systems: application to the yeast proteome,” *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 327–333, 2005. https://doi.org/10.1098/rsif.2005.0046

[13] Z. Huang and D.D. Zeng, “A link prediction approach to anomalous email detection,” In *2006 IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1131–1136, 2006. https://doi.org/10.1109/ICSMC.2006.384552

[14] P. Jaccard, “Etude de la distribution florale dans une portion des Alpes et du Jura,” *Bull Soc Vaudoise Sci Nat*, vol. 37, pp. 547–579, 1901.

[15] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953. https://doi.org/10.1007/BF02289026

[16] A. Kumar, S.S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Physica A: Statistical Mechanics and its Applications*, vol. 553, no. 1, p. 124289, 2020. https://doi.org/10.1016/j.physa.2020.124289

[17] J. Kunegis, E.W. De Luca and S. Albayrak, “The link prediction problem in bipartite networks,” In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pp. 380–389, Springer, 2010. https://doi.org/10.1007/978-3-642-14049-5_39

[18] L. Lü. L and T. Zhou, “Link prediction in weighted networks: The role of weak ties,” *Europhysics Letters (EPL)*, vol. 89, no. 1, p. 18001, 2010. https://doi.org/10.1209/0295-5075/89/18001

[19] L. Lü, C.-H. Jin and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Physical Review E*, vol. 80, no. 4, p. 046122, 2009. https://doi.org/10.1103/PhysRevE.80.046122

[20] L. Li, S. Bai, L. Leng, L. Wang and X. Chen, "Finding missing links in complex networks: A multiple-attribute decision-making method," *Complexity*, vol. 2018, pp. 1–11, 2018. https://doi.org/10.1155/2018/3579758

[21] H. Liao, Y.-C. Zeng and Y.-C. Zhang, "Predicting missing links via correlation between nodes," *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 216–223, 2015. https://doi.org/10.1016/j.physa.2015.05.009

[22] C. Lin, Y.-r. Cho, W.-C. Hwang, P. Pei and A. Zhang, "Clustering methods in protein-protein interaction network," *Knowledge Discovery in Bioinformatics: techniques, methods and application*, pp. 1–35, 2007.

[23] Z. Liu, Q.-M. Zhang, L. Lü and T. Zhou, "Link prediction in complex networks: A local naïve bayes model," E*urophysics Letters (EPL)*, vol. 96, no. 4, p. 48007, 2011. https://doi.org/10.1209/0295-5075/96/48007

[24] F. Lorrain and H.C. White, "Structural equivalence of individuals in social networks," *The Journal of Mathematical Sociology*, vol. 1, no. 1, pp. 49–80, 1971. https://doi.org/10.1080/0022250X.1971.9989788

[25] V. Martínez, F. Berzal and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, pp. 1–33, 2016. https://doi.org/10.1145/3012704

[26] F. Nasir, H. Gul, M. Bakhsh and A. Salam, "Comparitive analysis of link prediction in complex networks," *International Journal of Technology Diffusion (IJTD)* vol. 12, no. 3, pp. 44–60, 2021. https://doi.org/10.4018/IJTD.2021070103

[27] J. O'Madadhain, J. Hutchins and P. Smyth, "Prediction and ranking algorithms for event-based network data," ACM SIGKDD Explorations Newsletter, vol. 7, no. 2, pp. 23–30, 2005. https://doi.org/10.1145/1117454.1117458

[28] B. Pachev and B. Webb, "Fast link prediction for large networks using spectral embedding," *Journal of Complex Networks*, vol. 6, no. 1, pp. 79–94, 2018. https://doi.org/10.1093/comnet/cnx021

[29] S. Rafiee, C. Salavati and A. Abdollahpouri, "Cndp: Link prediction based on common neighbors degree penalization," *Physica A: Statistical Mechanics and its Applications*, vol. 539, p. 122950, 2020. https://doi.org/10.1016/j.physa.2019.122950

[30] F. Tan, Y. Xia and B. Zhu, "Link prediction in complex networks: a mutual information perspective," *PLoS ONE*, vol. 9, no. 9, p. e107056, 2014. https://doi.org/10.1371/journal.pone.0107056

[31] A. Tumasjan, T. Sprenger, P. Sandner and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, pp. 178–185, 2010. https://doi.org/10.1609/icwsm.v4i1.14009

[32] B. Viswanath, A. Mislove, M. Cha and K.P. Gummadi, "On the evolution of user interaction in Facebook," In *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*, pp. 37–42, 2009. https://doi.org/10.1145/1592665.1592675

[33] D. Wang, D. Pedreschi, C. Song, F. Giannotti and A.-L. Barabasi, "Human mobility, social ties, and link prediction," In *KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1100–1108, 2011. https://doi.org/10.1145/2020408.2020581

[34] T. Zhou, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150–1170, 2011. https://doi.org/10.1016/j.physa.2010.11.027

11 AUTHORS

Syed Shah Hussain is a dedicated and skilled computer scientist who completed his BS in Software Engineering from City University of Science and Information Technology. Throughout his academic career, he has shown a remarkable level of hard work and dedication, which has allowed him to develop a strong expertise in the fields of Graph theory and Complex network analysis.

He has also acquired a wealth of experience working with various tools and programming languages, including Python, MATLAB, and others. He has demonstrated a strong ability to use these tools to develop complex and effective solutions to a wide range of technical problems.

Muhammad Arif received the M.S. degree in applied mathematics from the Department of Mathematics, City University of Science and Information Technology, Peshawar Pakistan. He received M.Sc. Mathematics from the same university. He has published many research papers in different well-reputed high impact factor international journals of the world. His most experience has been working in the academic sector and has expertise in the area of fluid dynamics, two phase flows, MHD flows, nanofluids, couple stress fluid, fractional derivatives, integral transforms, exact solutions, numerical solutions, infectious diseases dynamics, biomathematics, and mathematical modeling. He has more than eight years of academic experience in different reputed institutions. He has attended many international conferences in the world as an invited speaker and presenter. Recently, he is a PhD researcher in the department of Mathematics Faculty of Science King Mongkut's University of Technology Thonburi (KMUTT), Bangkok Thailand. Furthermore, he is invited by the University of York Canada as an International Visiting researcher for six months on a research-based project. In addition to this he is a reviewer for different international journal of the world. (E-mail: arifmaths667@gmail.com)

Osama Bin Inayat is a dedicated and hardworking individual who completed his Bachelor's degree in Computer Science from the City University of Science and Information Technology. With a keen interest in Graph theory and Complex network analysis, he has developed expertise in these areas and has also gained experience working with various tools such as Python and Matlab. Osama's passion for exploring the intricacies of graphs and networks has led him to work on several projects and research papers, where he has applied his knowledge and skills to develop innovative solutions. He is committed to continuously learning and growing as a professional and hopes to make a positive impact in the field of computer science.

Haji Gul is an accomplished computer scientist with a Master of Science degree from IMSciences University. He is currently working as a lecturer at the Computer Science Department of City University of Science and Information Technology, where he imparts his knowledge and expertise in the field to aspiring students. Haji Gul has extensive knowledge and experience in machine learning, graph theory, complex network analysis, and related fields. He has made significant contributions to these areas and has published several papers in high-impact factor journals and international conferences. Throughout his career, Haji Gul has been recognized for his exceptional work and has earned a reputation as a reliable and dedicated professional. He continues to push the boundaries of computer science through his research and teachings, inspiring the next generation of computer scientists to pursue excellence and innovation.