

www.ijprems.com editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062 Impact Factor :

5.725

Vol. 03, Issue 06, June 2023, pp : 91-95

A REVIEW OF TRADITIONAL SIMILARITY BASED LINK PREDICTION METHODS IN COMPLEX NETWORKS

Nirmaljit Singh¹

¹Research Scholar, Sant Baba Bhag Singh University, Jalandhar india

ABSTRACT

Link prediction is a fundamental task in network analysis, aiming to infer missing or future connections between nodes. Numerous link prediction techniques have been proposed, each offering unique insights and methodologies. In this review paper, we comprehensively compare and evaluate five popular link prediction techniques: Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, and Resource Allocation. Through a comprehensive comparison, we discuss the advantages and limitations of each technique, shedding light on their suitability for different network scenarios. We examine their performance in real-world applications, such as social networks, biological networks, and recommendation systems. Moreover, we discuss the potential for future research and improvements in link prediction algorithms. By providing a comprehensive review of these link prediction techniques, we aim to assist researchers, practitioners, and enthusiasts in selecting the most appropriate method for their specific network analysis tasks. Understanding the nuances and trade-offs of these techniques is crucial for advancing link prediction in complex networks.

Keywords: Link prediction, Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, Resource Allocation

1. INTRODUCTION

Link prediction is a fundamental problem in network analysis that involves predicting missing or future connections between nodes based on observed interactions. It plays a crucial role in understanding network structure, uncovering hidden relationships, and making informed predictions about network dynamics. Over the years, numerous link prediction techniques have been proposed, each offering distinct approaches to address this challenging task

2. LITERATURE REVIEW

[1]. In this review paper, we provide a comprehensive analysis and comparison of five popular link prediction techniques: Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, and Resource Allocation. These techniques have gained significant attention in the research community due to their effectiveness and applicability in a wide range of network analysis tasks. The Common Neighbors method [1] focuses on the intuition that nodes with more shared neighbors are more likely to form links. It counts the number of common neighbors between two nodes and assumes that higher overlap indicates a higher probability of link formation. Adamic Adar [2] builds upon this concept by assigning higher weights to shared neighbors with lower degree centrality, considering the significance of connections to less connected nodes. The Jaccard Index[3] calculates the ratio of the number of common neighbors to the total number of distinct neighbors, providing a measure of similarity between nodes. Preferential Attachment

[1], on the other hand, predicts links based on the assumption that popular nodes, i.e., nodes with higher degrees, are more likely to attract new connections. Finally, the Resource Allocation method

[4] incorporates centrality measures to assess the importance of shared connections, redistributing resource scores from neighbors to measure similarity. we aim to compare and evaluate these link prediction techniques comprehensively. We analyze their underlying principles, calculation methods, strengths, limitations, and applications. Through empirical studies, real-world examples, and benchmark datasets, we assess their performance and provide insights into their effectiveness in different network contexts. Furthermore, we discuss the implications and trade-offs associated with each technique, highlighting their relevance and suitability for various network analysis tasks.

The findings of this review paper will be beneficial for researchers, practitioners, and enthusiasts in selecting appropriate link prediction techniques for their specific applications.

By understanding the nuances and comparative performance of these methods, researchers can make informed choices and advance link prediction in diverse domains such as social networks, biological networks, recommendation systems, and more. Overall, this review aims to provide a comprehensive understanding of the strengths and limitations of the Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, and Resource Allocation techniques, facilitating further advancements in link prediction algorithms and fostering their practical applications in complex network analysis.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

2583-1062 Impact Factor : 5.725

e-ISSN:

www.ijprems.com editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 91-95

3. COMPARISON

STRENGTH AND WEAKNESSES OF TRADITIONAL TECHNIQUES

The pros and cons of various traditional and popular link prediction techniques are as follows:

Common Neighbors: The Common Neighbors method [1] has several strengths. Firstly, it is a simple and intuitive method that is easy to understand and implement. Secondly, it has computational efficiency, making it suitable for large-scale networks. Thirdly, it tends to perform well in scenarios with dense networks, where the number of common neighbors between nodes is a good indicator of potential links. However, the Common Neighbors method also has weaknesses. It tends to have limited performance in sparse networks where shared neighbors are less common. Additionally, it ignores the quality of shared neighbors, treating all common neighbors equally. Lastly, the method is susceptible to the "hub" effect, where nodes with high degrees dominate the prediction, potentially leading to inaccurate results.

Adamic Adar:

The Adamic Adar method [2] has strengths that address some of the limitations of the Common Neighbors method. It considers the significance of connections to less connected nodes by assigning higher weights to shared neighbors with lower degree centrality. This approach allows for the identification of rare and valuable connections in the network. Furthermore, the Adamic Adar method is more robust against the "hub" effect compared to Common Neighbors. However, it also has weaknesses. It may be less effective in scenarios where highly connected nodes dominate the network. Additionally, the method requires knowledge of node degrees, which may not always be available. Moreover, there is a possibility that the Adamic Adar method assigns disproportionate importance to low-degree nodes, potentially affecting the accuracy of predictions.

Jaccard Index:

The Jaccard Index[3] provides a similarity-based measure for link prediction. It has strengths such as capturing the local neighborhood structure effectively and handling sparsity better than the Common Neighbors method. By considering the ratio of the number of common neighbors to the total number of distinct neighbors, the Jaccard Index can provide a measure of similarity between nodes. However, it also has limitations. Like the Common Neighbors method, it does not consider the quality of shared neighbors. It can still be affected by the "hub" effect, where nodes with high degrees dominate the prediction. Therefore, it may not be suitable for all network scenarios.

Preferential Attachment:

The Preferential Attachment method [1] has strengths rooted in the concept of popularity. It predicts links based on the assumption that popular nodes, i.e., nodes with higher degrees, are more likely to attract new connections. This approach captures the preferential growth mechanism observed in many real-world networks. However, it has limitations. The method does not consider any structural similarity between nodes and relies solely on the node degree distribution. It may not be suitable for networks where preferential attachment does not play a significant role or when considering additional factors beyond node popularity is crucial for accurate predictions.

Resource Allocation:

The Resource Allocation [4] method incorporates centrality measures into link prediction. By redistributing resource scores from neighbors to measure similarity, it considers the importance of shared connections in the network. This approach allows for the identification of nodes with similar network roles or positions. However, the method also has limitations. It requires the computation of centrality measures for all nodes, which can be computationally expensive for large networks. Additionally, the method assumes that resource allocation accurately reflects the likelihood of link formation, which may not always hold true in real-world scenarios.

Here's a table 1 comparing the five link prediction techniques: Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, and Resource Allocation.

Link Prediction Technique	Strengths	Weaknesses
Common Neighbors	- Simple and intuitive method	- Limited performance in sparse networks
	- Computational efficiency	- Ignores the quality of shared neighbors

Table 1: Attribute comparison of different LP methods



www.ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03 Issue 06 June 2023 pp · 91-95

e-ISSN : 2583-1062

Impact Factor : 5.725

tor@ijprems.c	com		, on 05, 155ac 00, 1	une	2020,		
			- Effective in dense n	etwo	orks	- Susceptible to the "hub" effec	et
Adamic Adar		- Considers connections to less connected nodes		less	- Less effective in highly connected networks		
		- Robust against the "hub" effe		effect	- Requires knowledge of node degrees		
	- E1	mphasiz c	es rare and valuable onnections	- P	otentia	l disproportionate importance giv to low-degree nodes	en
Jaccard Index - Provide		vides a s	imilarity-based measure		- Does not consider the quality of shared neighbors		
	- Handles sparsity Nei		sity better than Common Neighbors		- Can s	- Can still be affected by the "hub" effect	
	Cap	tures lo	cal neighborhood struct	ure			
Preferential Attachment - Captures		otures pi	referential growth mechanism		- Does not consider structural similarity between nodes		
		- Con	siders node popularity		- Lim	ited applicability in networks witho preferential attachment	ut
Resource Allocation - Inco		- Inco	rporates centrality measur	- Compt		omputationally expensive for large networks	
		- Ident	ifies nodes with similar ro	les	- Ass	sumes resource allocation accurately reflects link likelihood	1

It's important to note that this table provides a general overview of the strengths and weaknesses of each technique, but the suitability of a specific technique will depend on the characteristics of the network and the research context

4. OTHER IMPORTANT EVALUATING MEASURES

In addition to the strengths and weaknesses, several other points can be compared when evaluating and comparing link prediction techniques [9, 11, 14]:

- 1. Computational Complexity: Compare the computational requirements of each technique, including time complexity and memory usage. This is especially important when dealing with large-scale networks.
- 2. Data Requirements: Assess the data requirements for each technique, such as the need for node attributes, edge weights, or additional network information. Consider the availability and accessibility of such data.
- 3. Scalability: Evaluate the scalability of each technique, particularly in terms of network size and the ability to handle increasing data volumes. Consider whether the technique can efficiently handle large and dynamic networks.
- 4. Robustness: Assess the robustness of each technique to noise, missing data, or perturbations in the network structure. Determine how well the technique performs in the presence of such challenges.
- 5. Predictive Accuracy: Measure the predictive accuracy of each technique through appropriate evaluation metrics such as precision, recall, F1-score, or area under the receiver operating characteristic curve (AUC-ROC). Compare their performance on benchmark datasets or real-world network examples.
- 6. Sensitivity to Network Characteristics: Evaluate how each technique performs across different types of networks, such as social networks, biological networks, or technological networks. Consider whether the technique is sensitive to network properties like density, degree distribution, or community structure.
- 7. Interpretability: Consider the interpretability of the technique. Some techniques provide easily interpretable measures, while others may require more complex interpretations or have less transparent mechanisms.
- 8. Applicability to Various Domains: Assess the applicability of each technique to different domains or applications, such as social network analysis, recommendation systems, biological network analysis, or information retrieval. Consider the specific requirements and characteristics of the domain.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 91-95

- 9. Extensibility and Adaptability: Consider the potential for extending or adapting the technique to address specific research questions or incorporate additional factors or features.
- 10. Computational Implementation: Evaluate the availability of software packages or libraries that provide implementations of the techniques. Consider the ease of use and availability of code or tools to facilitate practical applications.

By considering these additional points of comparison, researchers can gain a more comprehensive understanding of the strengths, limitations, and suitability of different link prediction techniques for their specific research or application needs.

5. CONCLUSION

Determining the best link prediction technique depends on various factors such as network characteristics, data availability, and the specific research or application context. Each of the link prediction techniques discussed (Common Neighbors, Adamic Adar, Jaccard Index, Preferential Attachment, and Resource Allocation) has its strengths and weaknesses, making them suitable for different scenarios.Common Neighbors can be effective in dense networks where the number of shared neighbors is a good indicator of link formation. Adamic Adar performs well in scenarios where connections to less connected nodes are deemed significant, and it is more robust against the "hub" effect. The Jaccard Index is suitable for capturing similarity-based link formation, particularly in sparser networks. Preferential Attachment is effective when the popularity of nodes plays a significant role in link formation. Resource Allocation incorporates centrality measures and can be valuable for identifying nodes with similar network roles. To determine the best technique, researchers and practitioners should consider the specific characteristics of their network and the research objectives. [19] For example, if the network is dense and common neighbors play a crucial role, Common Neighbors might be a good choice. If the focus is on identifying rare and valuable connections, Adamic Adar can be more appropriate. If similarity-based link prediction is desired, the Jaccard Index might be the preferred choice. In some cases, combining multiple techniques or using hybrid approaches can yield more accurate predictions. Hybrid methods can leverage the strengths of different techniques and provide more robust link predictions. It is important to note that the selection of the best technique is not definitive, and it may vary depending on the specific context and research goals. Therefore, researchers should carefully evaluate the strengths and weaknesses of each technique and consider how they align with the particular characteristics and requirements of their network analysis task.

6. FUTURE SCOPE OF RESEARCH

Despite the strengths of these link prediction techniques, there is still scope for further research and improvement. Some areas that warrant attention include:

- Developing hybrid approaches that combine multiple techniques to leverage their complementary strengths [10].
- Considering temporal aspects and dynamic network characteristics for more accurate predictions in evolving networks.
- Exploring the impact of community structure and homophily in link prediction to account for the presence of tightly knit groups and similarity-driven connections.

Investigating the robustness and generalizability

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my family for their unwavering support and understanding throughout the process of writing this review paper. Their love, encouragement, and patience have been invaluable in enabling me to pursue my research endeavors. I am also deeply grateful to my research guide, for their guidance, expertise, and invaluable insights throughout the preparation of this review paper. Their constant support, valuable feedback, and encouragement have played a pivotal role in shaping my understanding of the subject matter and refining my research work.

7. REFERENCES

- [1] Liben-Nowell, D., & Kleinberg, J. M. (2007). The link prediction problem for social networks. Journal of the American Society for Information Science and Technology, 58(7), 1019-1031.
- [2] Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the Web. Social Networks, 25(3), 211-230.
- [3] Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 547-579.
- [4] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN:

www.ijprems.com editor@ijprems.com

- [5] Zhou, T., Ren, J., Medo, M., & Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. Physical Review E, 76(4), 046115.
- [6] Yang, J., McAuley, J., & Leskovec, J. (2012). Community detection in networks with node attributes. In Proceedings of the 2012 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1158-1166).
- [7] Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1), 39-43.
- [8] Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval. McGraw-Hill.
- [9] Backstrom, L., & Leskovec, J. (2011). 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).
- [10] Dekker, A. H. & Colbert, B. D. The symmetry ratio of a network. In: Proceedings of the 2005 Australasian symposium on Theory of computing Volume 41, CATS, Australian Computer Society, 13–20 (2005).
- [11] NEWMAN, M. E. J. 2001. Clustering and preferential attachment in growing networks. Physical Review E, 64.
- [12] Barrat A, Barthelemy M and Vespignani A, 2008 Dynamical Processes on the Complex Networks (Cambridge: Cambridge University Press)
- [13] L. Lü and T. Zhou, "Link prediction in complex networks: a survey," Physica A: Statistical Mechanics and Its Applications, vol. 390, no. 6, pp. 1150–1170, 2011.
- [14] Bullmore E and Sporns O, Complex brain networks: graph theoretical analysis of structural and functional systems, 2009 Nature Rev.
- [15] 4.Chavez M, Valencia M, Navarro V, Latora V and Martinerie J, Functional modularity of background activities in normal and epileptic brain networks, 2010 Phys. Rev. Lett. 104 118701 [10] Chung F, Lu L and Vu V, The spectra of random graphs with given expected degrees, 2003
- [16] K. Thiel and M. R. Berthold, "Node similarities from spreading activation," in Proceedings of the 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 1085–1090, Sydney, Australia, December 2010.
- [17] P. Symeonidis, E. Tiakas, and Y. Manolopoulos, "Transitive node similarity for link prediction in social networks with positive and negative links," in Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10), pp. 183–190, September 2010.
- [18] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, vol. 73, no. 2, Article ID 026120, 2006.
- [19] Y. Li, P. Luo, and C. Wu, "A new network node similarity measure method and its applications," March 2014.
- [20] W. De Nooy, "A literary playground: Literary criticism and balance theory," Poetics, vol. 26, no. 5-6, pp. 385– 404, 1999.