

A novel similarity measure for missing link prediction in social networks

G. Naga Chandrika¹, E. Srinivasa Reddy²

¹Department of Computer Science and Engineering, ANU College of Engineering and Technology, India

²Department of Computer Science and Engineering, ANU, India

Article Info

Article history:

Received Jan 5, 2020

Revised Mar 9, 2020

Accepted mar 23, 2020

Keywords:

Complex networks

Link prediction

Missing links

Similarity measure

Social networks

ABSTRACT

Social Networks progress over time by the addition of new nodes and links, form associations with one community to the other community. Over a few decades, the fast expansion of Social Networks has attracted many researchers to pay more attention towards complex networks, the collection of social data, understand the social behaviors of complex networks and predict future conflicts. Thus, Link prediction is imperative to do research with social networks and network theory. The objective of this research is to find the hidden patterns and uncovered missing links over complex networks. Here, we developed a new similarity measure to predict missing links over social networks. The new method is computed on common neighbors with node-to-node distance to get better accuracy of missing link prediction. We tested the proposed measure on a variety of real-world linked datasets which are formed from various linked social networks. The proposed approach performance is compared with contemporary link prediction methods. Our measure makes very effective and intuitive in predicting disappeared links in linked social networks.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

G. Naga Chandrika,

Department of Computer Science & Engineering,

ANU College of Engineering & Technology, Guntur, AP, India.

Email: gnchandrika@gmail.com

1. INTRODUCTION

Rapid development of online social networks has grown exponentially by the emergent day-by-day scenario and plays a vital role in creating interest to extend friendships, share common interests and to perform various business [1]. In social networks, nodes represent domain-dependent entities and links represent the relation between entities [2]. The Link prediction problem estimate or find hidden patterns that exist between any two unconnected nodes by observed nodes and links [2]. In general, social networks require link prediction to calculate the uncovered or future links which are useful for people to add friends on networks. First, link prediction is applied to recommender systems and it can be used to understand the entire network built on detected networks. Finally, it can be applied on Bio-informatics, to discover the interaction between proteins [3]. Security field can make use of Social link prediction to recognize secret terrorist activities which in turn identifies the abnormalities of communication in networks [4]. The Social network model is important to understand organizational structures and link behaviors [5]. Link prediction and classification becomes the most challenging issues in social networks, and these are essential due to the vast capacity of data on social networks [6]. The main obstacle in link prediction is the trade-off among the total data which includes nodes, links and features [7]. Evaluating the possible links among nodes that are not observed until now by utilizing on hand link information and feasible node information in social networks is the main role of missing link prediction [8]. However, the prediction of social networks faces two challenging issues that are a) incomplete or missing links, b) dynamic networks. The intension of missing

link prediction is to find disappeared links among the nodes using existing links and connections in the social networks that are recently added or deleted [6, 8]. The appearance of a link among nodes with high scores is significant for the prediction of missing links. Link prediction is the most challenging problem in complex networks [7]. This challenging objective makes interpretations concerned with the surviving links among the nodes of the graph and understands the structure of the network to forecast the links between the entities [7]. In the literature, most of the researchers have investigated link prediction through unsupervised and supervised learning approaches using nodes or paths of social networks. However, the social link prediction prototypes depend on the theory that nodes with high similarity are expected to join networks [9, 10]. The examination of complex network topology and structural features shows that the similarities are calculated based on the similarity metrics like Adamic Adar, common neighbors, Jaccard coefficient, preferential attachment [7]. Thus, these measures are likelihoods of link association in networks. Although the above measures may look outdated, they are far away from being obsolete [7]

This paper address an extension of the work done by Jinxuan Yang and Xiao-Dong Zhang (2016) [11]. Common Neighbors (CN) [7, 12, 13] is a popular approach for link prediction but the weaknesses of CN is that it is not sufficient to disclose the similarities between the nodes of networks with a single common neighbor. The network with part of missing links will not perform well due to uncommon neighbors, moreover it plays a vital role to form dissimilar communities, it in turn, affects network properties like average distance, centrality, and congestion. So, it is significant to develop an algorithm to identify the missing links among nodes without common neighbors. To solve these issues, we developed a novel approach to predict the unobserved links in social networks established on common neighbors and node-to-node distance. In this paper, we developed a new two-stage similarity measure for missing link prediction of networks. Finally, the experimental outcome reveals the dominance of the suggested method over persisting methods. The time complexity of this method is almost similar to the CN method [7, 12, 13] and CN and distance method [11].

2. RELATED WORKS

This section, address a brief review of the contemporary missing link prediction approaches. Recently, missing link prediction has received considerable attention on social networks. In 2004, Ben Taskar et al., [14] presented link prediction by using a probabilistic model in relational datasets using WebKB dataset. It is used to exploit interesting patterns in graph and Relational Markova networks. In 2006, Sevon et al., [15] presented a technique for link discovery from biological databases using three measures with probabilistic interpretation i.e., edge reliability, relevance, and rarity. In 2007, Liben Nowell, and Jon Kleinberg [16] presented a method to identify the closeness of nodes in social networks. Experiments are tested over complex networks of co-authorships which recommend that the information of the existing networks can be used for predicting the upcoming interactions.

In 2008, Clauset et al., [17] presented a measure for predicting missing links in networks and hierarchical structure prediction of networks. In 2009, Zhou et al., [18] offered a novel similarity approach for identifying missing links through local information based on node similarity. This shows that many links in the network are assigned with the same scores when the nearest neighbors are used. Another new measure, influenced by the resource allocation process and it is similar to common neighbors with one step random walk. In 2010, Liu and Linyuan [19] proposed a method for predicting the missing links, built on the local random walk and superposed random walk index and compared with other six well-known methods by using five well-known network datasets. In 2011, Strom and Leskovec [20] described a supervised random walk algorithm for predicting missing links where strengths to the edges are assigned by a function such that the random walker probably looks up the nodes to whom new links are formed in the future. In 2012, Chen et al., [21] used vertex based similarity measures for predicting missing links. It is useful to predict the unseen arrangement of the vertices in the networks.

In 2013, Narang et al., [22] discussed the link prediction task using proximity measures with different network flows which include a random walk, epidemics and applied these heuristics for identifying missing links in social and biological networks. In 2014, Naveen and Anurag Singh [23] proposed a new measure for predicting missing links by using common neighbors and uncommon neighbors which predicts the possibility of a link among two nodes by utilizing local topological information. In 2014, Tan et al., [24] developed a novel method for predictiong missing links in networks by applying mutual information of network topology which is a less time-consuming approach. In 2015, Liao et al., [25] described a new approach for predicting missing links in networks using the Pearson correlation coefficient. It is very proficient for extracting the similarity information in the network.

In 2016, Yang et al., [11] presented an approach for missing link prediction in complex networks using common neighbors and distance. It is accountable for missing links between nodes of the network

without common neighbors. We motivated for this work as an extension to this work of missing link prediction. In 2016, Pan et al., [26] presented an algorithm for predicting missing links to identify unauthentic links in complex networks using likelihood analysis. The limitation of this method is it can be applied only on small networks and time-consuming process. In 2017, Anupam and Bhaskar Biswas [27] have presented a prediction based on the community (CLP) for finding missing links, CLP system with centrality between edges, edge centrality with k-path and centrality with the spanning edge. In 2018, Yasami and Safaei [28] presented a novel approach for link prediction and forecasting of upcoming links in dynamic networks using multilayer model. The implication of this approach is obtained by sampling strategies of Markov Chain Monte Carlo in complex networks. In 2019, Rödderet al., [29] presented a link prediction rested on cross entropy in social networks. This entropy method employed unbiased estimates for weights of missing links. However, the above link prediction methods don't yield effectiveness in common neighbors and accuracy of missing link prediction in social networks.

3. PROPOSED MEASURE

Motivated by the missing link prediction issues and challenges, we address an extension of the work done by Jinxuan Yang and Xiao-Dong Zhang (2016) [11].

Problem Definition: Suppose $G(V, E)$ be a graph of social network at time t , where V represents nodes or vertices with size n , E represents edges or links with the size m respectively. Also, this graph is undirected, unweighted, doesn't allow self-loops and several links. The aim of this measure is the prediction of new links or missing links between nodes of the network.

Definitions: Let E_p be proportion links in graph G , then divide E_p into two subsets that are links among nodes with common neighbors (E_p^C) and without common neighbors (E_p^{NC}).

$$E_p^C = \{i, j \mid \Gamma(i) \cap \Gamma(j) \neq \emptyset \text{ for all } (i, j) \in E_p\} \tag{1}$$

$$E_p^{NC} = \{i, j \mid \Gamma(i) \cap \Gamma(j) = \emptyset, \text{ for all } (i, j) \in E_p\} \tag{2}$$

In (1) & (2) $\Gamma(i)$ represents the set of neighbors of node 'i', and $\Gamma(j)$ denotes the set of neighbors of node 'j'.

$$E_p^C \cup E_p^{NC} = E_p \tag{3}$$

$$|E_p^C| + |E_p^{NC}| = |E_p| = L \tag{4}$$

where L represents the number of missing links.

New Similarity Measure: In literature, complex networks use common neighbor as a popular method for link prediction. A single common neighbor measure does not reveal the similarities in network links accurately. So, it is essential to propose a new measure for identifying missing links among nodes without common neighbors. Keeping these limitations, we suggested a new similarity method for predicting missing links in complex networks. This method computes links if they have more common neighbors with node-to-node distance. Here, node-to-node distance plays a vital task in missing link prediction among nodes without common neighbors. The proposed similarity measure summarized as follows:

$$Sim_{ij} = \begin{cases} \frac{CN_{ij+d_{ij}}}{1+d_{ij}}, & \Gamma(i) \cap \Gamma(j) \neq \emptyset, \\ \frac{1}{d_{ij}}, & \text{otherwise,} \end{cases} \tag{5}$$

where, $CN_{ij} = |\Gamma(i) \cap \Gamma(j)|$ represents the set of neighbours of node 'i' & 'j'. d_{ij} is the distance among nodes.

Using (5) we attain the similarity scores of all the un existing links $U - E_T$, then sort the similarity scores in decreasing form. Now pick top- $|E_p^C|$ links among nodes with common neighbors and select the top- $|E_p^{NC}|$ links among nodes without common neighbors and from $U - E_T$ to compose identified links. In this section, we presented the proposed measure of research and at the same time it gives the thorough examination. Outcomes can be shown in figures, graphs, tables and others that make the reader to interpret effortlessly [2, 5].

4. EXPERIMENTS AND RESULTS

This study uses eight complex social network datasets for testing our algorithm that is (i) Dolphins, (ii) E-mail, (iii) Karate, (iv) Polbook, (v) Power, (vi) Word, (vii) Facebook, (viii) Wiki. The datasets and their properties are presented in Table 1.

Table 1. Description of the datasets and their properties

| Dataset | n | M | c | c_n | $\langle d \rangle$ | $\langle k \rangle$ |
|----------|------|--------|--------|-------|---------------------|---------------------|
| Dolphins | 62 | 159 | 0.26 | 0.76 | 3.35 | 5.12 |
| Email | 1133 | 5451 | 0.22 | 0.77 | 3.06 | 9.62 |
| Karate | 34 | 78 | 0.57 | 0.85 | 2.40 | 4.58 |
| Polbook | 105 | 441 | 0.48 | 0.95 | 3.07 | 8.40 |
| Power | 4941 | 6594 | 0.08 | 0.20 | 18.98 | 2.66 |
| Word | 112 | 425 | 0.17 | 0.72 | 2.53 | 7.58 |
| Facebook | 4039 | 88234 | 0.6055 | 0.687 | 8 | 43.69 |
| Wiki | 7115 | 103689 | 0.1409 | 0.734 | 7 | 14.57 |

where n =nodes, m =links, c = cluster coefficient, c_n =common neighbor coefficient, $\langle d \rangle$ = average distance, $\langle k \rangle$ =average degree.

Dolphins [30]: Data About bottlenose dolphins relationship directed social network. The nodes represent the bottlenose dolphins of the dolphin community living off Doubtful Sound, a fjord in New Zealand during the period 1994-2001.

Email [31]: This dataset contains networks of e-mail transactions between the members of University at Rovira i Virgili (Tarragona)

Karate [32]: This dataset consists of social ties among the associates of the university of Karate club networks composed by W.Zachary 1977.

Polbook [33]: This dataset is a network related to books of US politics printed around 2004 governmental election and available at Amazon.com.

Power [34]: This dataset contains an un weighted network presenting topological information about Power Grid of Western States in the United States.

Word [35]: This dataset was constructed by Newman contains a linkage of noun adjacencies and common adjectives of the book “*David Copperfield*” by Charles Dickens

Facebook [36]: This dataset comprises of friends list from Facebook, includes node features (profiles), circles, and ego networks. Data is collected from a survey through an app.

Wiki [37]: This dataset comprises all the Wikipedia voting related information acquired from the origination of Wikipedia till January 2008.

The computation of link prediction methods: This study compares the proposed method with Adamic-Adar (AA), Common Neighbor (CN), Hub Depressed Index (HDI), Hub Promoted Index (HPI), Jaccard Coefficient, Leicht-Holme-Newman Index (LHN), Salton Index (Sal), Sørensen Index (Sen), and CN+Dist [11] on complex social networks given in Table 1.

Testing metrics: In this experiment, two statistical indices are identified to check the effectiveness of the proposed link-prediction method. The initial index measure is the area under the ROC curve (AUC):

$$AUC = \frac{n' + \frac{1}{2}n''}{n} \quad (6)$$

where n' is the number of periods a disappeared link is allotted a higher probability, n'' is the number of periods they have allotted identical probability, and n denotes the total number of assessments that are the number of disappeared links and the number of unexisting links. This AUC measure quantifies the possibility of assigning a higher score to a missing link than any non-observed link. Another index measure is Precision also known as true positive rate, which is defined as follows:

$$Precision = \frac{L_r}{L_{miss}} \quad (7)$$

This quantifies the percentage of missing links that are properly retrieved. In Table 2, predicted results of various common neighbor index methods under the AUC results on the complex social network datasets are presented. The outcomes are the mean of 40 recognitions for every complex network, and every time the probe set E_p will be removed randomly. The uppermost values for individual networks are highlighted.

Tables 2 and 3, list the predicted results of the various index measures under the AUC metric. The outcomes are the mean of 40 recognitions for every network under 10% and 20% investigation set. Thus, the proposed measure performs better with other measures due to their node-to-node distance computation. Because of less training dataset, the pair of nodes through a common neighbor becoming a minor that raises the complexity of missing link prediction. Similarly, the missing link prediction results of Precision are presented in Table 4 and 5. It quantifies that the percentage of missing links are properly recovered in the linked social networks. The experimental results of state-of-the-art algorithms are well performed for a great ration of E_p than the lower Precision measure.

Table 2. The results of the AUC of various measures under 10% probe in social networks

| | Dolphins | Email | Karate | Polbook | Power | Word | Facebook | Wiki |
|----------|----------|-------|--------|---------|-------|-------|----------|-------|
| AA | 0.776 | 0.849 | 0.720 | 0.899 | 0.585 | 0.677 | 0.867 | 0.846 |
| CN | 0.772 | 0.847 | 0.678 | 0.890 | 0.586 | 0.678 | 0.912 | 0.928 |
| HDI | 0.773 | 0.845 | 0.582 | 0.865 | 0.586 | 0.621 | 0.884 | 0.893 |
| HPI | 0.754 | 0.841 | 0.696 | 0.897 | 0.586 | 0.635 | 0.912 | 0.913 |
| Jaccard | 0.770 | 0.845 | 0.592 | 0.879 | 0.586 | 0.627 | 0.901 | 0.902 |
| LHN | 0.752 | 0.838 | 0.579 | 0.852 | 0.586 | 0.585 | 0.856 | 0.834 |
| Sal | 0.617 | 0.844 | 0.618 | 0.885 | 0.586 | 0.624 | 0.874 | 0.901 |
| Sen | 0.592 | 0.845 | 0.589 | 0.879 | 0.586 | 0.623 | 0.873 | 0.891 |
| CN+Dist | 0.791 | 0.880 | 0.724 | 0.904 | 0.659 | 0.694 | 0.942 | 0.946 |
| Proposed | 0.779 | 0.972 | 0.772 | 0.921 | 0.668 | 0.656 | 0.989 | 0.949 |

Table 3. The results of the AUC of various measures under 20% probe in social networks

| | Dolphins | Email | Karate | Polbook | Power | Word | Facebook | Wiki |
|----------|----------|-------|--------|---------|-------|-------|----------|-------|
| AA | 0.754 | 0.822 | 0.684 | 0.882 | 0.571 | 0.661 | 0.856 | 0.862 |
| CN | 0.741 | 0.822 | 0.662 | 0.858 | 0.571 | 0.858 | 0.938 | 0.904 |
| HDI | 0.751 | 0.820 | 0.591 | 0.847 | 0.571 | 0.609 | 0.823 | 0.884 |
| HPI | 0.739 | 0.816 | 0.662 | 0.868 | 0.571 | 0.623 | 0.927 | 0.902 |
| Jaccard | 0.749 | 0.819 | 0.598 | 0.856 | 0.571 | 0.609 | 0.867 | 0.888 |
| LHN | 0.736 | 0.814 | 0.588 | 0.829 | 0.571 | 0.583 | 0.899 | 0.910 |
| Sal | 0.746 | 0.818 | 0.614 | 0.862 | 0.571 | 0.611 | 0.901 | 0.893 |
| Sen | 0.750 | 0.819 | 0.597 | 0.856 | 0.571 | 0.609 | 0.893 | 0.913 |
| CN+Dist | 0.767 | 0.871 | 0.678 | 0.885 | 0.593 | 0.672 | 0.924 | 0.912 |
| Proposed | 0.795 | 0.973 | 0.715 | 0.881 | 0.655 | 0.680 | 0.988 | 0.946 |

Table 4. The results of precision of various measures under 10% probe

| | Dolphins | Email | Karate | Polbook | Power | Word | Facebook | Wiki |
|----------|----------|-------|--------|---------|-------|-------|----------|-------|
| AA | 0.128 | 0.158 | 0.132 | 0.172 | 0.031 | 0.068 | 0.207 | 0.078 |
| CN | 0.135 | 0.149 | 0.143 | 0.148 | 0.069 | 0.063 | 0.301 | 0.112 |
| HDI | 0.083 | 0.075 | 0.000 | 0.105 | 0.020 | 0.002 | 0.211 | 0.091 |
| HPI | 0.022 | 0.013 | 0.171 | 0.142 | 0.030 | 0.011 | 0.266 | 0.11 |
| Jaccard | 0.087 | 0.074 | 0.004 | 0.122 | 0.016 | 0.002 | 0.193 | 0.088 |
| LHN | 0.017 | 0.004 | 0.007 | 0.077 | 0.009 | 0.001 | 0.192 | 0.032 |
| Sal | 0.075 | 0.056 | 0.000 | 0.120 | 0.015 | 0.000 | 0.278 | 0.118 |
| Sen | 0.087 | 0.074 | 0.004 | 0.122 | 0.016 | 0.002 | 0.192 | 0.121 |
| CN+Dist | 0.227 | 0.225 | 0.221 | 0.188 | 0.079 | 0.193 | 0.324 | 0.125 |
| Proposed | 0.348 | 0.320 | 0.231 | 0.266 | 0.081 | 0.093 | 0.314 | 0.131 |

Table 5. The results of precision of various measures under 20% probe

| | Dolphins | Email | Karate | Polbook | Power | Word | Facebook | Wiki |
|----------|----------|-------|--------|---------|-------|-------|----------|-------|
| AA | 0.161 | 0.188 | 0.155 | 0.278 | 0.023 | 0.102 | 0.194 | 0.194 |
| CN | 0.226 | 0.163 | 0.200 | 0.243 | 0.085 | 0.107 | 0.356 | 0.149 |
| HDI | 0.159 | 0.113 | 0.045 | 0.138 | 0.018 | 0.017 | 0.214 | 0.112 |
| HPI | 0.018 | 0.012 | 0.145 | 0.192 | 0.027 | 0.009 | 0.245 | 0.146 |
| Jaccard | 0.131 | 0.097 | 0.028 | 0.157 | 0.013 | 0.014 | 0.190 | 0.099 |
| LHN | 0.027 | 0.012 | 0.015 | 0.089 | 0.010 | 0.001 | 0.239 | 0.077 |
| Sal | 0.108 | 0.071 | 0.023 | 0.158 | 0.013 | 0.009 | 0.302 | 0.123 |
| Sen | 0.131 | 0.097 | 0.028 | 0.157 | 0.013 | 0.014 | 0.189 | 0.133 |
| CN+Dist | 0.256 | 0.255 | 0.292 | 0.252 | 0.060 | 0.231 | 0.389 | 0.169 |
| Proposed | 0.306 | 0.390 | 0.259 | 0.289 | 0.122 | 0.133 | 0.394 | 0.175 |

5. CONCLUSION

This research presents a novel similarity measure for predicting missing links. The account of missing links among two nodes is due to uncommon neighbors in social networks. The suggested approach is relay on the common neighbors and the node-to-node distance which improves prediction accuracy. Existing measures constructed on common neighbors are unable to identify missing links among nodes, computing the percentage of missing links among nodes without common neighbors in total missing links of the networks. The proposed method makes an efficient way for missing link prediction among nodes without common neighbors. From the investigational results, we conclude that the prediction precision of the suggested approach is higher to other methods used on social networks.

REFERENCES

- [1] M. Fire, et al., "Computationally efficient link prediction in a variety of social networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 1, Dec 2013.
- [2] M. Pujari, "Link Prediction in Large-scale Complex Networks (Application to bibliographical Networks)," PhD diss, Universite Paris Nord, 2015.
- [3] E. M. Airoidi, et al., "Mixed membership stochastic block models for relational data with application to protein-protein interactions," in *Proceedings of International Biometric Society Annual Meetings*, 2006.
- [4] M. Al. Hasan, et al., "Link prediction using supervised learning," in *Proceedings of SDM Workshop on Link Analysis, Counterterrorism and Security*, 2006.
- [5] G. N. Chandrika and E. S. Reddy, "Latent Semantic Analysis and Nearest Neighbor Classification Algorithm for Linked Web Pages," *International Journal of Computer Science and Information Security*, vol. 15, no. 6, pp. 304-308, 2017.
- [6] W. Almansoori, et al., "Link prediction and classification in social networks and its application in healthcare and systems biology," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 1, no. 1-2, pp. 27-36, 2012.
- [7] E. C. Mutlu and T. A. Oghaz, "Review on Graph Feature Learning and Feature Extraction Techniques for Link Prediction," *arXiv preprint arXiv: 1901.03425*, 2019.
- [8] L. L'u 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.
- [9] P. Wang, et al., "Link prediction in social networks: the state-of-the-art," *Science China Information Sciences*, vol. 58, no. 1, pp. 1-38, 2015.
- [10] V. Martínez, et al., "A survey of link prediction in complex networks," *ACM Computing Surveys*, vol. 49, no. 4, p. 69, 2017.
- [11] J. Yang and X. D. Zhang, "Predicting missing links in complex networks based on common neighbors and distance," *Scientific reports*, vol. 6, p. 38208, 2016.
- [12] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, p. 025102, 2001.
- [13] 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.
- [14] B. Taskar, et al., "Link prediction in relational data," *Advances in neural information processing systems*, pp. 659-666, 2004.
- [15] P. Sevon, et al., "Link discovery in graphs derived from biological databases," in *International Workshop on Data Integration in the Life Sciences*, pp. 35-49, 2006.
- [16] D. L. Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for information science and technology*, vol. 58, no. 7, pp. 1019-1031, 2007.
- [17] A. Clauset, et al., "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, pp. 98-101, 2008.
- [18] T. Zhou, et al., "Predicting missing links via local information," *The European Physical Journal B*, vol. 71, no. 4, pp. 623-630, 2009.
- [19] W. Liu and L. Lü, "Link prediction based on local random walk," *EPL (Europhysics Letters)*, vol. 89, no. 5, p. 58007, 2010.
- [20] 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.
- [21] H. H. Chen, et al., "Discovering missing links in networks using vertex similarity measures," in *Proceedings of the 27th annual ACM symposium on applied computing*, pp. 138-143, 2012.
- [22] K. Narang, et al., "Network flows and the link prediction problem," in *Proceedings of the 7th workshop on social network mining and analysis*, pp. 1-8, 2013.
- [23] N. Gupta and A. Singh, "A novel strategy for link prediction in social networks," in *Proceedings of the 2014 CoNEXT on student workshop*, pp. 12-14, 2014.
- [24] F. Tan, et al., "Link prediction in complex networks: a mutual information perspective," *PloS one*, vol. 9, no. 9, p. e107056, 2014.
- [25] H. Liao, et al., "Predicting missing links via correlation between nodes," *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 216-223, 2015.

- [26] L. Pan, et al, "Predicting missing links and identifying spurious links via likelihood analysis," *Scientific Reports*, vol. 6, 2016.
- [27] A. Biswas and B. Biswas, "Community-based link prediction," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18619-18639, 2017.
- [28] Y. Yasami and F. Safaei, "A novel multilayer model for missing link prediction and future link forecasting in dynamic complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 2166-2197, 2018.
- [29] W. Rödder, et al., "Weight Prediction on Missing Links in Social Networks—a Cross-Entropy-Based Approach," *Journal of Applied Logics—IfCoLog Journal of Logics and their Applications*, vol. 6, no. 1, pp. 95-116, 2019.
- [30] D. Lusseau, et al., "The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396-405, 2003.
- [31] R. Guimerà, et al., "Self-similar community structure in a network of human interactions," *Physical Review E*, vol. 68, 2003.
- [32] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452-473, 1977.
- [33] <http://konect.cc/networks/dimacs10-polbooks/>
- [34] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [35] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, p. 036104, 2006.
- [36] <https://snap.stanford.edu/data/ego-Facebook.html>
- [37] <https://snap.stanford.edu/data/wiki-Vote.html>

BIOGRAPHIES OF AUTHORS



G. Naga Chandrika received the B.Tech degree in Computer Science and Information Technology from JNTU, Hyderabad in 2002 and M.Tech degree in Software Engineering from JNTU, Hyderabad, in 2004. She is currently pursuing Ph.D degree in Computer Science and Engineering at ANU, Guntur, AP, India. Since 2011, she has been an Assistant Professor with the Information Technology Department, VNRVJIET, Hyderabad, India. Her research interests include Data Mining, Machine Learning.



Prof E.Srinivasa Reddy received the B.Tech degree in Electronics & Communications from ANU, Guntur in 1988, M.S in Electronics & Control from BITS, Pilani, India in 1997, M. Tech degree in Computer Engineering from VTU, Karnataka, India in 2000 and Ph.D. degree in Computer Science & Engineering from ANU, Guntur in 2008. From 2014, he was working as Principal, ANU College of Engineering, Guntur, India. He is the author of 5 books, more than 100 articles. His research interest includes Image Processing, data Mining.