International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Hybrid Link Prediction Model Combining Node Degree Properties

Mouhoub BELAZZOUG *[1], Mourad CHARIKHI [2], Abdelhamid SAIFI [3], Messaoud BENDIAF [4], Meriem LAIFA [5]

Submitted: 16/03/2024 **Revised**: 26/04/2024 **Accepted**: 02/05/2024

Abstract:

Link prediction is an essential task in the analysis of complex networks. It involves predicting new links based on a prediction algorithm. Local methods are commonly used for link prediction and are effective in many application areas, even for large datasets. However, they are characterized by low precision. To handle this problem, we propose in this paper, a hybrid approach that integrates local link prediction methods and node degree properties to enhance the performances of local methods. To validate our proposal, we tested our approach on several existing similarity methods and performed a series of experiments on twelve datasets for link prediction. Experimental results on almost all datasets indicate the powerful performance of our proposed hybrid approach in terms of AUC scoring.

Keywords: *Link Prediction, Similarity metrics, node degree property, Network Evolution, Feature Engineering.*

1. Introduction

In last decades, the amount of data on the Internet has grown considerably. Data analysis has become principal to extract knowledge as well as managing these websites. Indeed, the information is dispersed between pages which are in reality representing people, factories, magazines or company sites. These sites with their interactions led many researchers to study a very important field which is the analysis of the complex networks. These networks are complex and their structures dynamically developed. Link prediction analyses complex network aims for estimating new links between existing nodes. It is one of the most important fields in network analysis domains. We can find many fields applications of link prediction task such as, recommender systems [1] and in social networks that can help people make new friends [2], [3]. There is some works in academic networks such as co-publishing or co-citation between authors [4] [5]. In e-commerce sites today use link prediction to provide shoppers with exciting new products based on their preferences or purchases. Predict future friendships during social network analyses [3]. In Bioinformatics, we cite for instance the prediction of protein-protein interactions [6]. There are two main approaches to solve link prediction, based node information and graph structure. Local methods are a promising alternative that provide good performances with low complexity. These methods leverage the topological of graph to estimate the new links. Most of them, then don't predict new links for nodes that don't share neighbors. To overcame this drawback we propose the node node degree property as an alternative solution to the current local similarity measure in perform. More specifically, Resources allocation (RA) is used in this study to improve the performance of existing link prediction algorithms. as we know from literature the RA property contribute negatively to estimate the likelihood of predicting link, in other words, more the RA property is low for a given node, the more likely he makes connection in the future. Resource allocation (RA) has been demonstrating a good advance for improving several link prediction methods in numerous applications domain. In [7], [8], [9] and [10] **.** In this paper, the experimental study demonstrates the good performance of our new similarity approach (NP) compared to several methods taken from the state of the art, especially with databases having a low average degree compared to the number of points in the graph, while maintaining the characteristic of low complexity of the local method.

The remainder of the paper is organized as follows. In the Section-2, the related works of the link prediction problem is given, including different classified approaches in this research field. In Section-3, we present the main contribution of our proposed approach, which leverage the Node Property index combined with several existing local based methods. Section-4 we present the concept and the mathematical formulation of existing algorithm used in this study. Section-5 focuses on the experiential study. We give a description of the properties of the twelve (12) datasets used in this study, the evaluation methodology, also the assessment metric. Moreover, the result of experimental tests of algorithms and their improved version with RA index is given in details. Finally, this work is concluded by a conclusion.

2. Related Works

Network science has seen remarkable growth in recent years across many areas of interdisciplinary research. Networks are a powerful tool for representing complex systems, whether

12345Department of Computer Science, Mohamed El Bachir El Ibrahimi University, El Anasser, Algeria
** Corresponding Author Email: Mouhoub.belazoug@univ-bba.dz*

social, biological, or technological [11], [12]. Among the most productive branches of network science is link prediction which aims to evaluate the probabilities of missing links, future links and temporal links by taking advantage of the existing network topology [13]–[15]. Different methodologies are frequently employed in this domain, including similarity-based methods [3], topology-based methods [16], and machine learning approaches [17], among others. We can distinguish three categories of methods based on similarity, namely local, global and quasi-local methods [14] . Local similarity methods mainly use the number of neighbors (or neighbors of neighbors) to estimate future connections in the graph [18]. They are based on the principle that nodes with many shared neighbors are more likely to be interconnected in the future [2]. The main metric of this approach is Common Neighbors [3]. Its variations include, Jaccard Coefficient JC [19], Adamic-Adar index AA [20], preferential attachment PA[21], resource allocation index RA [22], cosine similarity [23], and node clustering coefficient [24].

Recent advancements incorporating eigenvector centrality measures into similarity calculations between nodes [25]. Scalable algorithms based on popularity features extracted from network topology have been proposed, owing to their versatility and domain-independence [26]. Mutual information and high-order clustering structures are proposed to enhance link prediction accuracy [27]. Furthermore, methodologies like the local Naïve Bayes model have been devised to estimate the probability of connection between node pairs by considering the distinct characteristics and roles of common neighbors [28]. In diverse domains, Graph Neural Networks (GNNs) have emerged as a potent tool for learning representations from systems with rich-relational data, finding applications in bioinformatics and beyond [29]. In [30], The DeepLink framework, proposed by Keikha,et al, they employ deep learning techniques to extract features from both content and structural information, thereby advancing link prediction in social networks. Resource allocation (RA) has been demonstrating a good advance for improving several link prediction methods in numerous applications domain. In [7], authors proposed a similarity-network resource allocation index to predict user-item links for improving prediction accuracy and preserving recommendation diversity, and ensuring algorithm scalability by utilizing a limited number of neighbors. [8] Authors propose a new method called Multi-Steps Resource Allocation (MSRA) and they demonstrate its effectiveness in link prediction. Moreover, Liu et al, estimated links in a time-weighted user-item graph in which the latest phasing activities were recorded with the largest weights to predicted users' phasing activities [9]. The work of [10] proposed the neighbor set information allocation index based on a set of neighbors obtained from the process of virtual information allocation to quantify the possible connection of events corresponding to the two nodes by measuring self-information.

Generally, similarity-based methods are popular for link prediction due to their efficiency in capturing network structure. This paper proposes a new link prediction method leveraging node properties, aiming to enhance prediction accuracy. The higher the importance of the nodes, the more likely they are to be connected to each other. This paper utilizes the Python language to conduct experiments using real datasets. Results indicate that the proposed prediction method outperforms existing similarity measure algorithms in link prediction.

3. Proposed Method

Local methods based on neighborhood information, such as Common Neighbor (CN), Adamic Adar (AA), and preferential attachment, are widely used to solve link prediction problems [31]. As their names and concepts suggest, these methods rely solely on the neighborhood property. Consequently, the more neighbors a pair of nodes shares, the more likely they are to be connected in the future. However, the estimated score is zero if the pair of nodes has no common neighbors. This drawback unfortunately persists even in their successive variations. Meanwhile, global methods can predict new links between nodes that do not share a neighborhood. However, these methods are computationally expensive and resource-intensive. To overcome this problem, we propose a new approach in this paper to improve local methods for link prediction. We leverage the node degree property as an alternative measure to the original local method. The degree property of nodes is only taken into account when the candidate pair of nodes does not share any neighbors. Our approach can be integrated with various local methods. In this section, we present an example to demonstrate the enhancement of the common neighbor method with the node degree property. To prove our new local approach for link prediction, we provide the calculations for two different scenarios based on a given graph (see Figure 1). Therefore, we will first present the new formula, followed by the graph used in this demonstration. We will then compare the calculations before and after applying our new measure. This comparison will demonstrate the importance of integrating our new measure for link prediction.

If we consider the CN method as the original metric, the result of our new proposal method is given in Formula 1:

$$\mathbf{CN.NP}(\mathbf{x},\mathbf{y}) = \begin{cases} \mathbf{IF}\ (\mathbf{CN}(\mathbf{x},\mathbf{y}) > \mathbf{0}\ )\ \mathbf{then} \\ \mathbf{CN}(\mathbf{x},\mathbf{y}) \\ \boldsymbol{Else}\ (\mathbf{max}((\mathbf{deg}(\boldsymbol{x})\ , \mathbf{deg}(\boldsymbol{y}))/\boldsymbol{N}) \end{cases} \qquad (1)$$

Where the pair of nodes (x, y) of network is not connected and N is the number of nodes of the graph. To be more consistent and Clair in our paper, we present graphically in figure 1 a scenario that presents obviously this limitation wherein we see clearly that nodes don't share neighbors will be linked in the future while others can't even if they could have new links leveraging other methods or regarding the topological graph structure. In this example, we graphically illustrate two different scenarios to show the difference between them in terms of accuracy for link prediction under the presented network. The first scenario simulates the link prediction task using only the CN method, while in the second scenario that represents our contribution; we combine the CN method and the degree node property. For the scenario1 and from Figure 1, the node pairs (A-D), (B-C), and (C-E) which share neighbors, the CN metric gives a non-null

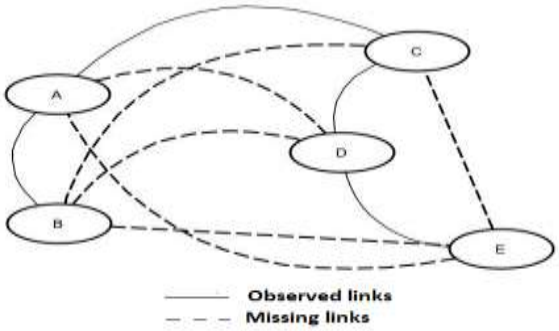

Figure 1 : Example of a graph with 5 nodes for link prediction

score which equal 1.

Regarding, the pairs of unconnected nodes (A-E), (B-D), and (B-E) that don't share neighbors, the estimated score for predicting new links is equal to zero by Common Neighbors (CN) metric. Here it is clear to see the limitation of local method vs this case of situation, therefore they are naïve and they not accurate well. On the other hand, the scenario_2 which represents the combination of CN metric and the node degree property to calculate the score of pair of nodes for link prediction task as it's described in Formula1.

The node pairs (A-D), (B-C), and (C-E) which share neighbors, the score returned for link prediction using formula 1 is equal 1. We notice here, our new formula is not chosen and the score provided is calculated basing only on the basic of CN metric, like as in the first scenario. Meanwhile, in the case of the pairs (A-E), (B-D), and (B-E), the new calculated score here is equal to 2.5, 1.5, and 1.5, respectively and these new recorded scores are calculated by our proposed formula using the node degree property contrarily to the 1st_scenario where the old scores are equal to 0 by CN metric. It is worth to notice here that 3 links, namely (A-E), (B-D), and (B-E) are gained for link prediction task for this current example of network that having 5 nodes. The calculus and the detail of this example are given in table 1. The bold lines are marked to show the different cases that have improvement in score by our new measure, the degree property, for link prediction task.

Table 1. Original vs proposed method for link prediction

Scenario1 Original method : CN measure	**Scenario2** Our approach: CN-NP measure
A-D = 1	A-D = 1
A-E = 0	**A-E = 2/5**
B-D = 0	**B-D = 2/5**
B-C = 1	B-C = 1
B-E = 0	**B-E = 1/5**
C-E = 1	C-E = 1

In this section, we proposed a new method to improve the accuracy of link prediction, particularly for local methods that primarily rely on a node's neighbourhood information. In this section, we have demonstrated the effectiveness of our approach by combining the Common Neighbor (CN) measure with node degree (NP). The network given in Figure 1 and the results presented in Table 1 clearly illustrate the improvement achieved by our approach when incorporating the node degree (NP) measure for link prediction, compared to the original CN measure.

4. Background & Methods

In this section, we show the baseline algorithms that have been improved by our strategy. All these methods will be later, tested and compared against our proposed similarity measures in the experimental section. From existing local methods, we have selected ten methods based on shared neighbor's information, namely: CN, AA, JI, RA, CNC, RACNI, SAM, HPI, IA and LNB-RA [32]. We introduce the definition and mathematical formulation of the three methods that are used in our new similarity measurement, as follows: For the set of similarity measure used, $\Gamma(x)$ represents the neighborhood of *x* which is the set of nodes connected to a node *x* by an edge. The degree of a node *x* is represented by the $|\Gamma x|$ symbol and is defined as the number of edges that bind to the node. Table2 presents the list of ten methods with relative formulation and reference used in this study.

Table 2. List of methods for link prediction

Order	Algorithm name	Formula	Reference
01	***Common Neighbors (CN)***	$S^{CN}(x,y) = \mid \Gamma(x) \cap \Gamma(y) \mid$	[3]
02	***Adamic-Adar Index (AA)***	$S^{AA}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{log\|\Gamma z\|}$	[20]
03	***Resource Allocation index (RA)***	$S^{RA}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\|\Gamma z\|}$	[33], [34]
04	**Jaccard index**	$S^{JC}(x,y) = \frac{\mid \Gamma(x) \cap \Gamma(y) \mid}{\mid \Gamma(x) \cup \Gamma(y) \mid}$	[19]
05	**Hub promoted index**	$S^{HPI}(x,y) = \frac{\mid \Gamma(x) \cap \Gamma(y) \mid}{\min(\mid \Gamma x \mid, \mid \Gamma y \mid)}$	[35]
06	**Sam Similarity**	$S^{SAM}(x,y) = \frac{\frac{\|\Gamma(x) \cap \Gamma(y)\|}{\|\Gamma x\|} + \frac{\|\Gamma(x) \cap \Gamma(y)\|}{\|\Gamma y\|}}{2}$	[36]
07	**Resource Allocation Based on Common Neighbor**	$S_{xy}^{\mathrm{ORA-CNI}}(x,y) = \sum_{Z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\|\Gamma z\|} + \sum_{e_{i,j} \in E, \Gamma(i) < \Gamma(j), i \in \Gamma(x), j \in \Gamma(y)} (\frac{1}{\|\Gamma i\|} - \frac{1}{\|\Gamma j\|}) + \beta \sum_{[x,p,q,y] \in path_{x,y}^{3}} \frac{1}{\|\Gamma p\|\|\Gamma q\|}$	[37]
08	**Common neighbor centrality index**	$S^{CNC}(x,y) = \alpha(\|\Gamma(x) \cap \Gamma(y) \|) + \frac{N(1-\alpha)}{dist(x,y)}$	[38]
09	**The Individual Attraction Index**	$\boldsymbol{S^{IA}} = \sum_{\boldsymbol{Z \in \Gamma(x) \cap \Gamma(y)}} \frac{\boldsymbol{e_z}}{\boldsymbol{\|\Gamma z\|}}$	[39]
10	**Local Naïve Bayes**	$S^{LNB} = \sum_{Z \in \Gamma(x) \cap \Gamma(y)} f(z) \log(oR_z)$	[28]

5. Experimental Results:

In this section ten 10 local based methods for link prediction task are tested as well as improved in the simulation experiments, such as: CN, AA, JI, RA, CNC, and CND among other commonly used indices. Table 3 presents exhaustive list of these algorithms used in this study with references. We have selected 12 networks from different sources and applications domains. These networks were chosen to cover a wide range of properties, including different sizes, average degrees, clustering coefficients, and heterogeneity indices. We categorized tow type of networks: first, social and biological networks, the second type is: contact and communication. The chosen networks are as follows: HPD, YST, and CEG are biological networks. ERD, HTC, are co-authorship networks for different fields of study.

HMT, FBK, and ADV are social networks. EML is a network of individuals who shared emails. PGP is an interaction network. BUP is a network of political blogs. Finally, INF is a network of face-to-face contacts in an exhibition.

Table 4 shows the details of the structural properties of the networks utilized in our experiments. The performance of our suggested similarity method will be evaluated using five-fold cross-validation approach and we have used the AUC (area under the receiver operating characteristic curve) metric that is frequently used to assess how well link prediction algorithms function in complex networks.

To distinguish a data set with a strong average degree from a data set with a low average degree, we have proposed the following hypothesis: if the average number of connections of a node is greater than 1% of the total number of possible connections with all nodes, then it has a strong average degree; otherwise, it has a low average degree. Formally, the mean degree is considered strong relative to the number of nodes if (**100 * <K> /|V| >= 1)**, otherwise it is considered low.

Table 3. List of the algorithms used in this study with references

Algorithm	**Designation**	**References**
CN	Common Neighbors	[3]
AA	Adamic Adar	[40]
JI	Jaccard index	[41]
RA	Resource allocation	[34]
CNC	Common neighbor centrality	[38]
RACNI	Resource Allocation Based on Common Neighbor Interactions	[37]
SAM	Sam Similarity	[36]
HPI	Hub promoted index	[35]
IA	Individual Attraction Index	[39]
LNB-RA	Local Naïve Bayes - Resource Allocation	[28]

Table 4. Description of the structural features of the networks used in our experiments [42]

Name	**|V|**	**|E|**	**<k>**	**C**	**ASPL**	**D**	**H**	**r**
HPD	8756	32331	7,38	0,11	4,19	14	4,5133	-0,051
ERD	6927	11850	3,42	0,12	3,78	4	12,6708	-0,1156
YST	2284	6646	5,82	0,13	4,29	11	2,8479	-0,0991
EML	1133	5451	9,62	0,22	3,61	8	1,9421	0,0782
ADV	5155	39285	15,24	0,25	3,22	9	5,406	-0,0951
PGP	10680	24316	4,55	0,27	7,49	24	4,1465	0,2382
CEG	297	2148	14,46	0,29	2,46	5	18008	-0,1632
INF	410	2765	13,49	0,46	3,63	9	1,3876	0,2258
BUP	105	441	8,4	0,49	3,08	7	1,4207	-0,1279
HTC	7610	15751	4,14	0,49	5,68	19	2,0986	0,2939
HMT	2426	16630	13,71	0,54	3,15	10	31011	0,0474
FBK	4024	87887	43,68	0,59	3,98	13	2,432	0,0707

From left to right: number of nodes, number of links, average degree of each node, average clustering coefficient, average shortest path length, diameter of the network average, heterogeneity and average degree associativity.

5.1. Social and biological networks

The evaluation approach for the experimental results primarily employs the AUC value as the accuracy measure for link prediction simulation experiments, using ADV, FBK, HMT, CEG, HPD, and YST networks. These experiments evaluate the algorithms CN, AA, JI, RA, CNC, and CND, as well as their respective improved versions, see Table 5. From the simulation experiments conducted on these six social networks, it is evident that the gain is significant with data sets whose average degree is low compared to the number of network nodes (100 * |V|/<K> < 1). The gain is zero or negative if the average degree is strong.

Table 5. AUC score values of different algorithms on social and biological networks

Algorithm	Network					
	ADV	FBK	HMT	CEG	HPD	YST
CN	0,8823	0,9897	0,9523	0,8276	0,7141	0,685
CN -NP	0,9349	0,9889	0,9585	0,8354	0,8781	0,8544
AA	0,8856	0,9909	0,9553	0,845	0,7146	0,6855
AA-NP	0,9382	0,9901	0,9615	0,8538	0,8786	0,8548
JI	0,8673	0,9883	0,9492	0,7767	0,7127	0,6837
JI-NP	0,9073	0,977	0,9474	0,6961	0,8762	0,8529
RA	0,8857	0,9921	0,9561	0,8485	0,7145	0,6854
RA-NP	0,9259	0,9821	0,9572	0,8184	0,8784	0,8547
CNC	0,9023	0,9912	0,9612	0,8316	0,8398	0,7904
CNC--NP	0,9349	0,9889	0,9585	0,8354	0,8781	0,8544
sum (Alg)	4,4232	4,9522	4,7741	4,1294	3,6957	3,53
sum(Alg-NP)	4,6412	4,927	4,7831	4,0391	4,3894	4,2712
ecart	0,22	-0,03	0,01	-0,09	0,69	0,74
100 * <K> /|V|	**0,296**	**1,085**	**0,565**	**4,869**	**0,084**	**0,255**

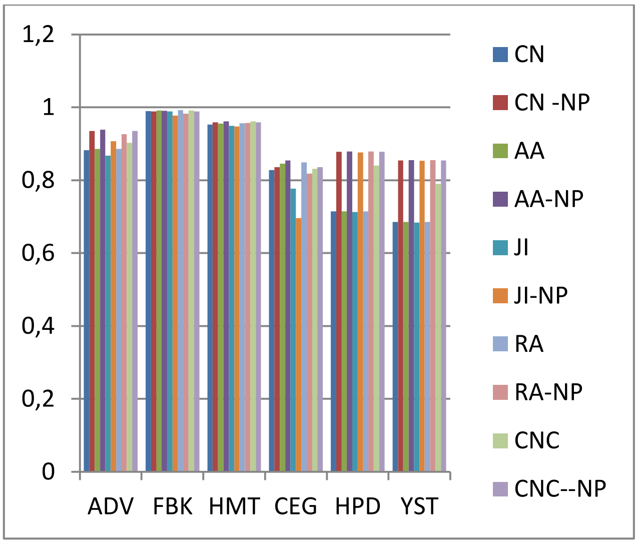

Figure 2: Auc score results presented by algorithm and Network

Most of the improved algorithms show significant enhancements in comparison to the original algorithms. The most important improvement was registered in score is realized in YST and HPD networks with 0.74 and 0.69 of difference respectively. A remarkable gain was contacted in ADV equal to 0.22, while we remark a equivalence in results performance in the other networks which are FBK, HTM and CEG. We notice here that the only exception was registered in the Jaccard Index (JI) where there is degradation in performances of 0.7 with CEG data set whose average degree is very strong (100 * |V|/<K> == 4.86). Therefore, the modified algorithms consistently demonstrate a notable increase in accuracy when compared to the original metrics.

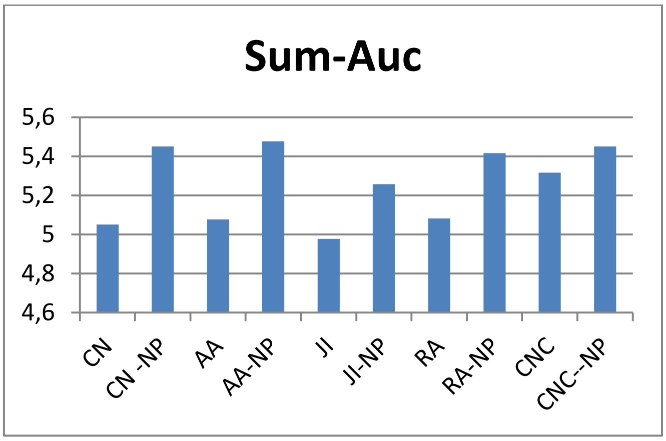

Figure 3 : Cumulative AUC-score across social and biological Networks grouped by algorithm

In Figure 2, we present an illustration of the performance of the original algorithm tested on these networks and its improved version that leverages node properties. The modified algorithm shows significant improvement, particularly with databases having a low average degree in the graph (YST, HPD, and ADV networks), compared to the original algorithms. Figure 3 shows the gain achieved by each algorithm in the cumulative sum score (AUC-measure). Our approach demonstrates significant improvement, although the worst case is observed in the CNC and CNC-NP algorithms, where the improvement is around 0.1 in terms of AUC score. We follow the same process, but this time we test our approach against recent algorithms from the literature: RACN, SAM, HPI, AI, and LBN-RA. Table 6 shows the scores obtained by our approach and the original algorithms. It is always visible that the gain is significant with datasets whose average degree is low compared to the number of network nodes. We observe significant improvements for our approach compared to most algorithms, especially in YST, HPD, and ADV networks. In contrast, for FBK, HMT, and CEG networks, only slight improvements are observed.

Figure 4, illustrates the experimental results of the original and newly improved algorithms: RACN, SAM, HPI, AI, and LBN-RA. The modified algorithm shows significant improvement, particularly in YST, HPD, and ADV networks. Figure 5, complements this by showing the gain achieved by each algorithm in the cumulative AUC-score across all datasets. Here, our approach demonstrates significant improvement, except for the RACN and improved RACN-NP algorithms. In these cases, the cumulative improvement is around 0.1.

Table 6. AUC score values of different algorithms on social and biological networks

	Network					
Algorithm	ADV	FBK	HMT	CEG	HPD	YST
RACN	0,9264	0,9887	0,9652	0,8324	0,8629	0,8203
RACN-NP	0,9366	0,9858	0,9591	0,8402	0,879	0,8556
SAM	0,8667	0,988	0,9496	0,7905	0,7126	0,6835
SAM-NP	0,9193	0,987	0,9556	0,7926	0,8766	0,8529
HPI	0,8658	0,9851	0,9484	0,7933	0,7126	0,6835
HPI-NP	0,9184	0,9842	0,9545	0,7995	0,8766	0,8528
IA	0,8859	0,9911	0,9559	0,8481	0,7146	0,6855
IA-NP	0,935	0,9843	0,961	0,8474	0,8786	0,8548
LNB-RA	0,8857	0,9804	0,9561	0,8451	0,7146	0,6857
LNB-RA A	0,9354	0,9893	0,9618	0,8313	0,8785	0,8551
sum (Alg)	4,4305	4,9333	4,7752	4,1094	3,7173	3,5585
sum(Alg-NP)	4,6447	4,9306	4,792	4,111	4,3893	4,2712
Gain	**0,21**	**0,00**	**0,02**	**0,00**	**0,67**	**0,71**
100 * <K> / |V|	**0,296**	**1,085**	**0,565**	**4,869**	**0,084**	**0,255**

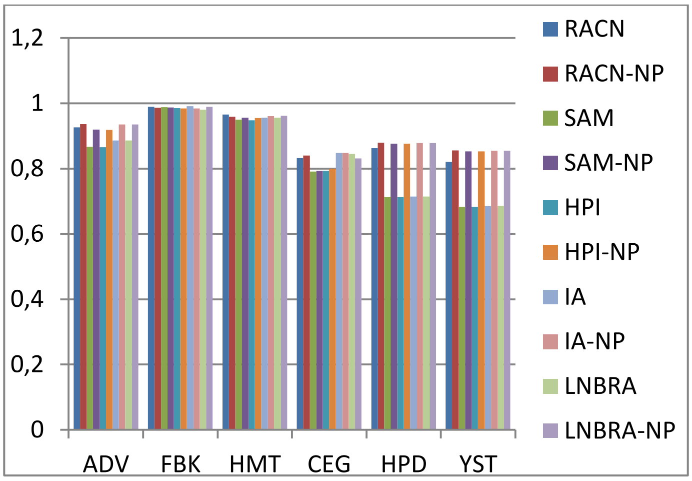

Figure 4: Auc score results presented by algorithm and Network

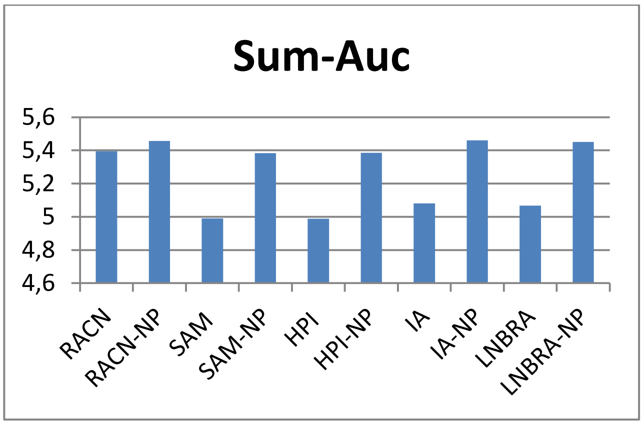

Figure 0 : Cumulative AUC-score across social and biological Networks grouped by algorithm

We continue our experimental study in the same manner, but this time on the second type of networks, namely: BUP, EML, INF, PGP, ERD, and HTC. This allows us to evaluate our enhanced algorithms on a wider range of network categories, specifically communication and contact networks. The following sections will present the evaluations and results obtained for these new networks.

5.2. Communication and contact Networks

We evaluated the performance of several algorithms such as: CN, AA, JI, RA, and CNC and their improved versions using AUC score measure under the communication and contact networks, namely: BUP, EML, INF, PGP, ERD, and HTC. We noted the same remark as previous experiments. All new algorithms work very well with data sets having low average degree. The experiments had shown significant performance improvements, particularly in EML, ERD, and HTC networks, see Table 7. Despite, we observed a slight relative decrease in performance for our approach compared to the original algorithms in the BUP and INF networks whose average degree is very strong (100 * |V|/<K> == 8 and 3,29), this is negligible compared to the significant improvements achieved in other networks. Overall, the modified algorithms consistently demonstrated the substantial advancements achieved compared to the original versions.

Table 7. Comparison of AUC values of different algorithms on communication and contact networks

Algorithm	DataSet					
	BUP	EML	INF	PGP	ERD	HTC
CN	0,8691	0,8234	0,9264	0,8192	0,6792	0,8501
CN A	0,8732	0,8750	0,9241	0,8651	0,9661	0,8645
AA	0,8781	0,8251	0,9300	0,8193	0,6797	0,8502
AA A	0,8822	0,8767	0,9277	0,8653	0,9665	0,8647
JI	0,8559	0,8215	0,9286	0,8191	0,6772	0,8501
JI A	0,7664	0,8705	0,9021	0,8650	0,9631	0,8645
RA	0,8801	0,8248	0,9305	0,8193	0,6796	0,8502
RA A	0,8163	0,8755	0,9188	0,8653	0,9674	0,8647
CNC	0,8895	0,8708	0,9492	0,7712	0,5960	0,8610
CNC A	0,8732	0,8750	0,9241	0,8651	0,9661	0,8645
sum (Alg)	4,3727	4,1656	4,6647	4,0481	3,3117	4,2616
sum(Alg-NP)	4,2113	4,3727	4,5968	4,3258	4,8292	4,3229
Gain	**-0,16**	**0,21**	**-0,07**	**0,28**	**0.15**	**0,06**
100 * <K> / |V|	**8,000**	**0,849**	**3,290**	**0,043**	**0,049**	**0,054**

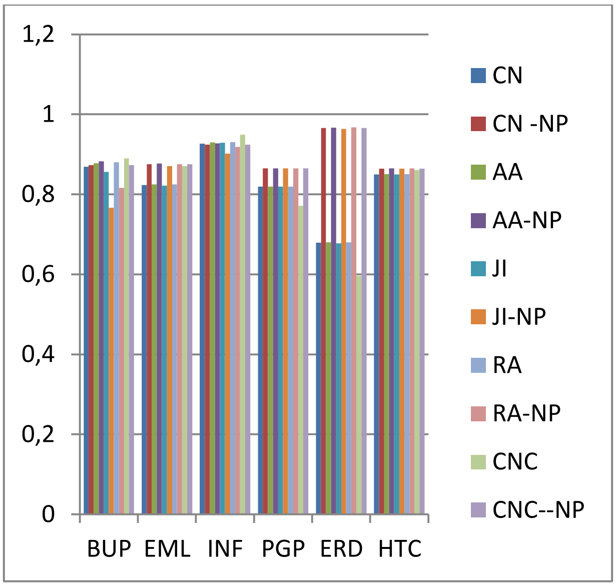

Figure 6: Auc score results presented by algorithm and Network

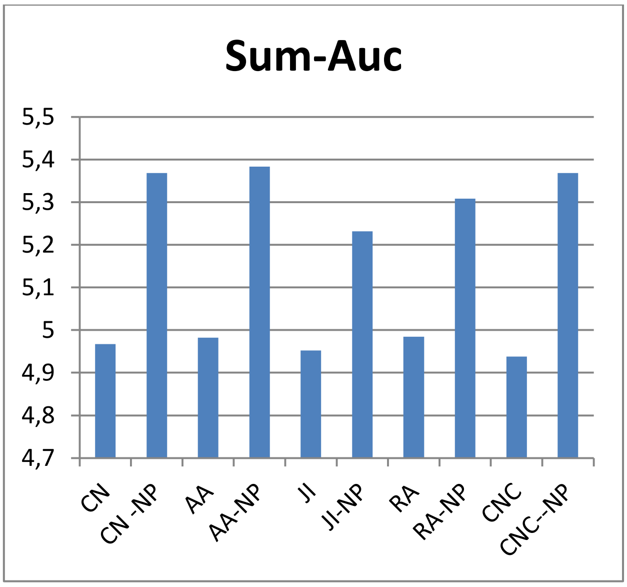

Figure 07: Cumulative AUC-score across communication and contact Networks grouped by algorithm

Figure 6, illustrates the Auc measurement obtained by the original algorithms and our improved algorithms. Obviously, the graph shows the progress of our approach in almost all networks, notably ERD. In Figure 7, which presents the cumulative score of each algorithm registered in all networks, it is clear that there is a big difference between all the original algorithms and their improved versions in all datasets. In other words, our approach achieved perfect performance improvement.

Table 8: Comparison of AUC values of different algorithms on communication and contact networks

Algorithm	Dataset					
	BUP	EML	INF	PGP	ERD	HTC
RACNI	0,8934	0,8889	0,9515	0,8442	0,7323	0,8779
RACNI-NP	0,8766	0,8767	0,9230	0,8651	0,9672	0,8645
SAM	0,8665	0,8197	0,9276	0,8190	0,6774	0,8501
SAM-NP	0,8623	0,8713	0,9243	0,8649	0,9643	0,8645

HPI	0,8682	0,8186	0,9255	0,8189	0,6777	0,8501
HPI-NP	0,8700	0,8702	0,9228	0,8648	0,9645	0,8645
IA1	0,8780	0,8250	0,9302	0,8193	0,6796	0,8502
IA1-NP	0,8664	0,8766	0,9274	0,8653	0,9672	0,8647
LNBRA	0,8772	0,8246	0,9295	0,8193	0,6797	0,8502
LNBRA-NP	0,8438	0,8762	0,9252	0,8659	0,9670	0,8647
sum (Alg)	4,3833	4,1768	4,6643	4,1207	3,4467	4,2785
sum(Alg-NP)	4,3191	4,371	4,6227	4,326	4,8302	4,3229
Gain	**-0,06**	**0,19**	**-0,04**	**0,21**	**0.14**	**0,04**
100 * <K> / |V|	**8,000**	**0,849**	**3,290**	**0,043**	**0,049**	**0,054**

Table 8, shows the evaluation performances of algorithms, such as: RACN, SAM, HPI, AI, and LBN-RA, also their all improved versions using our approach. They are evaluated using Auc score measure under the category communication and contact networks, namely: BUP, EML, INF, PGP, ERD, and HTC. Our proposed hypothesis is confirmed with the fourth experiment. The proposed algorithms still work very well with data sets having low average degree. The experiments had shown significant performance improvements, particularly in EML, PGP, ERD networks, where we notice there is a small advance in HTC networks. We observe there is a slight relative decrease in performance for our approach compared to the original algorithms in the BUP and INF networks, but this is negligible compared to the significant improvements achieved in other networks. Overall, the modified algorithms consistently demonstrated the substantial advancements achieved compared to the original versions.

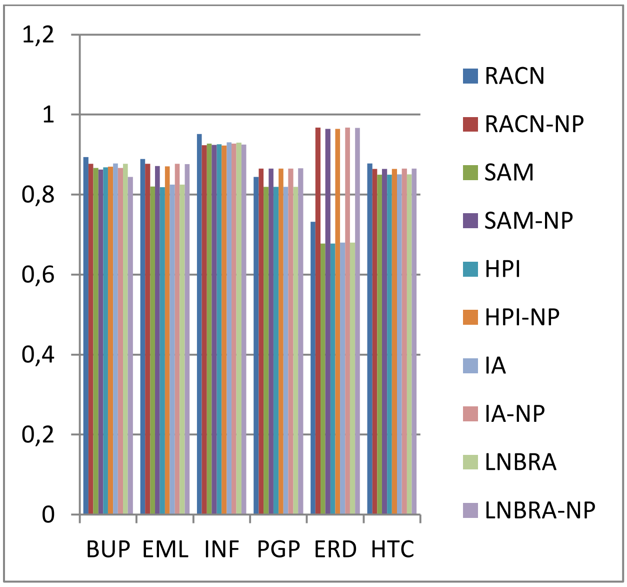

Figure 08 : Auc score results presented by algorithm and Network

Figure 8, illustrates the Auc score measurement obtained by the original algorithms and our improved algorithms. Obviously, the graph shows the progress of our approach in almost all networks, notably ERD. In Figure 9, which presents the cumulative score of each algorithm registered in all networks, it is clear that there is a big difference between all the original algorithms and their improved versions in all datasets. In other words, our approach achieved perfect performance improvement.

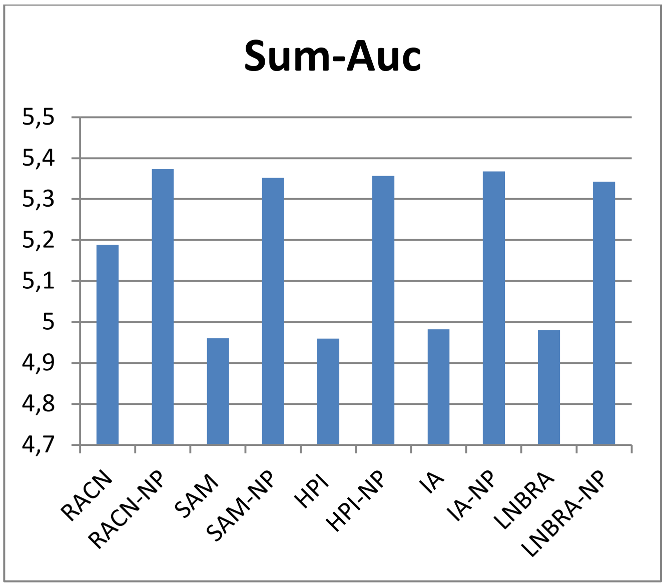

Figure 9: Cumulative AUC-score across communication and contact Networks grouped by algorithm

6. Conclusion

In this paper, we have proposed a novel hybrid link prediction approach leveraging local methods and node degree (NP) properties. This approach successfully addresses the limitations of existing local methods by assigning non-zero scores to unconnected pairs of nodes. From the experimental section, our hybrid method improved the performance of several indices such as CN, RA, AA, JACCARD, HPI, HDI, and PA, as well as other similarity metrics presented in this study. These improvements have been observed across a diverse set of real-world datasets from various categories, including: biological, social, communication and contact networks. Concluding the experimental results, we see a significant improvement in the AUC score obtained with our hybrid approach compared to other existing link prediction algorithms applied in this study. It is particularly effective for datasets characterized by low average degree, such as ERD, EML, HPD and YST, compared to other networks with high average degree.

Future research could explore the applicability of the node degree property to different types of networks and investigate further improvements to this hybridization. It would be useful to examine the performance of our hybrid approach in dynamic networks or oriented networks. Additionally, testing the integration of other network properties, such as community structure or edge weights, could further improve the accuracy of link prediction.

- **Data Availability Statement**

The datasets utilized in our study for link prediction problem are conventionally accessible through the following link: http://noesis.ikor.org/datasets/linkprediction.

- **Conflict-of-Interest Statement**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

7. References

[1] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4321 LNCS, pp. 291–324, 2007.

[2] D. Liben-Nowell, “An algorithmic approach to social networks.” .

[3] D. Liben-Nowell and J. Kleinberg, “The link-prediction problem for social networks,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[4] M. E. J. Newman, “The structure of scientific collaboration networks,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 2, pp. 404–409, Jan. 2001.

[5] D. Mohdeb, A. Boubetra, and M. Charikhi, “Strength-based link prediction in scientific bibliographic networks,” *J. Inf. Technol. Res.*, vol. 10, no. 3, pp. 84–106, Jul. 2017.

[6] D. Bu *et al.*, “Topological structure analysis of the protein–protein interaction network in budding yeast,” *Nucleic Acids Res.*, vol. 31, no. 9, p. 2443, May 2003.

[7] J. Ai, Y. Cai, Z. Su, K. Zhang, D. Peng, and Q. Chen, “Predicting user-item links in recommender systems based on similarity-network resource allocation,” *Chaos, Solitons & Fractals*, vol. 158, p. 112032, 2022.

[8] Z. Wu and Y. Li, “Link Prediction Based on Multi-steps Resource Allocation,” in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 2014, vol. 1, pp. 355–360.

[9] J. Liu and G. Deng, “Link prediction in a user–object network based on time-weighted resource allocation,” *Phys. A Stat. Mech. its Appl.*, vol. 388, pp. 3643–3650, 2009.

[10] P. Pei, B. Liu, and L. Jiao, “Link prediction in complex networks based on an information allocation index,” *Phys. A Stat. Mech. its Appl.*, vol. 470, 2016.

[11] M. Newman, *Networks*. Oxford University Press, 2018.

[12] A.-L. Barabási, “Network science,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 371, no. 1987, p. 20120375, 2013.

[13] P. Wang, B. Xu, Y. Wu, and X. Zhou, “Link prediction in social networks: the state-of-the-art,” vol. 58, pp. 1–38.

[14] V. Martínez, F. Berzal, and J.-C. Cubero, “A Survey of Link Prediction in Complex Networks,” vol. 49, no. 4.

[15] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Phys. A Stat. Mech. its Appl.*, vol. 553, Sep. 2020.

[16] K. K. Shang and M. Small, “Link prediction for long-circle-like networks,” *Phys. Rev. E*, vol. 105, no. 2, Feb. 2022.

[17] M. Hasan, V. Chaoji, S. Salem, and M. Zaki, “Link Prediction Using Supervised Learning,” 2006.

[18] M. Charikhi, “Association of the PageRank algorithm with similarity-based methods for link prediction in complex networks,” *Phys. A Stat. Mech. its Appl.*, vol. 637, p. 129552, Mar. 2024.

[19] P. Jaccard, “Etude comparative de la distribution florale une portion des Alpes et du Jura,” vol. 37, pp. 547–579, 1901.

[20] L. A. Adamic and E. Adar, “Friends and neighbours on web,” vol. 25, pp. 211–230.

[21] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[22] S. Zeng, “Link prediction based on local information considering preferential attachment,” vol. 443, pp. 537–542.

[23] G. Adomavicius and A. Tuzhilin, “Toward the Next Generation of Recommender Systems,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[24] S. Fortunato, “Community detection in graphs,” *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, Feb. 2010.

[25] L. Wang, C. Chen, and H. Li, “Link Prediction of Complex Network Based on Eigenvector Centrality,” *J. Phys. Conf. Ser.*, vol. 2337, p. 12018, 2022.

[26] S. Kerrache, R. Alharbi, and H. Benhidour, “A Scalable Similarity-Popularity Link Prediction Method,” *Sci. Rep.*, vol. 10, no. 1, Dec. 2020.

[27] Y. Yao *et al.*, “Link prediction based on the mutual information with high-order clustering structure of nodes in complex networks,” *Phys. A Stat. Mech. its Appl.*, vol. 610, Jan. 2023.

[28] Z. Liu, Q. M. Zhang, L. Lü, and T. Zhou, “Link prediction in complex networks: A local naïve Bayes model,” *EPL*, vol. 96, no. 4, Nov. 2011.

[29] X. M. Zhang, L. Liang, L. Liu, and M. J. Tang, “Graph Neural Networks and Their Current Applications in Bioinformatics,” *Front. Genet.*, vol. 12, Jul. 2021.

[30] M. M. Keikha, M. Rahgozar, and M. Asadpour, “DeepLink: A novel link prediction framework based on deep learning,” *J. Inf. Sci.*, vol. 47, no. 5, pp. 642–657, Oct. 2021.

[31] L. Lü and T. Zhou, “Link prediction in complex networks: A survey,” vol. 390, pp. 1150–1170.

[32] V. Martínez, F. Berzal, and J.-C. Cubero, “A survey of link prediction in complex networks,” vol. 49, no. 4, pp. 1–33.

[33] J. Scripps, P. N. Tan, and A. H. Esfahanian, “Measuring the effects of preprocessing decisions and network forces in dynamic network analysis,” in *In Proceedings of the 15th {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining*, pp. 747–756.

[34] T. Zhou, L. Lü, and Y. C. Zhang, “Predicting missing links via local information,” *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, Oct. 2009.

[35] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, “Hierarchical organization of modularity in metabolic networks,” vol. 297, no. 5586, pp. 1551–1555, 2002.

[36] A. Samad, M. Qadir, and I. Nawaz, “SAM: A Similarity Measure for Link Prediction in Social Network,” *MACS 2019 - 13th Int. Conf. Math. Actuar. Sci. Comput. Sci. Stat. Proc.*, Dec. 2019.

[37] J. Zhang, Y. Zhang, H. Yang, and J. Yang, “A link prediction algorithm based on socialized semi-local information,” vol. 10, pp. 4459–4466, 2014.

[38] I. Ahmad, M. U. Akhtar, S. Noor, and A. Shahnaz, “Missing Link Prediction using Common Neighbor and Centrality based Parameterized Algorithm,” vol. 10, no. 1, pp. 1–9, 2020.

[39] Y. Dong, Q. Ke, B. Wang, and B. Wu, “Link Prediction Based on Local Information,” in *2011 International Conference on Advances in Social Networks Analysis and Mining*, 2011, pp. 382–386.

[40] L. A. Adamic and E. Adar, “Friends and neighbors on the Web,” *Soc. Networks*, vol. 25, no. 3, pp. 211–230, Jul. 2003.

[41] P. Jaccard, “Étude comparative de la distribution florale dans une portion des Alpes et des Jura,” *Bull. del la Société Vaudoise des Sci. Nat.*, vol. 37, pp. 547–579, 1901.

[42] D. Thesis, F. B. Galiano, J. Carlos, and C. Talavera, “Supervised data mining in networks: link prediction and aplications,” no. May, Oct. 2018.