Comparative Evaluation of Local and Quasi-Local Topological Methods for Link Prediction in Complex Networks

Avelar Rodrigues de Sousa, Maykon Willyam de Sousa Ferreira

Institute of Mathematics and Computer Science (ICMC)
University of São Paulo (USP)
avelars@usp.br, maykonwsf@usp.br

Abstract. Link prediction is a relevant task in complex networks, as it aims to estimate missing or future connections from the observed topology. This research presents an empirical comparison of 19 topological link prediction methods in eight real-world networks from different domains: social, scientific collaboration, communication, and electrical infrastructure. The evaluated methods include local similarity indices, degree-based approaches, enhanced local methods, a local Bayesian method, and quasi-local methods based on short paths and walks. The experimental protocol used five-fold cross-validation, preserving the connectivity of the training graphs and applying balanced negative sampling. The methods were mainly compared using Average Precision (AP), complemented by the area under the receiver operating characteristic curve (ROC-AUC), Precision, Recall, F1-score, and Normalized Discounted Cumulative Gain (NDCG). The results show that performance varies substantially across networks, with higher scores in networks with stronger local closure, such as ego-Facebook and scientific collaboration networks, and lower scores in the power-grid network. Among the evaluated methods, Superposed Random Walk with three steps (SRW-l3) achieved the best overall mean AP, followed by Local Path Index with $\beta = 0.001$ (LPI-beta-0.001) and Resource Allocation (RA) Based on Common Neighbor Interactions (RA-CNI). A statistical analysis (Friedman test followed by Wilcoxon signed-rank tests with Holm correction) indicated that SRW-l3 significantly outperformed PFP-l3 and RA-CNI, but its advantage over LPI-beta-0.001 was not statistically significant. Overall, the findings indicate that short paths and walks provide useful structural information for link prediction, while network properties such as average degree and clustering help explain task difficulty in an exploratory manner.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: complex networks, graph mining, link prediction, topological similarity

1. INTRODUCTION

Complex networks allow the representation of systems composed of interconnected entities, such as social relationships, scientific collaborations, digital communications, and physical infrastructures [Newman 2003; Lü and Zhou 2011]. In this context, link prediction consists of estimating missing or future connections based on the observed structure of the network, being a relevant task for social network analysis, recommender systems, information retrieval, and the study of the evolution of complex systems [Liben-Nowell and Kleinberg 2007; Martínez et al. 2016]. Furthermore, network representations have also been increasingly applied to uncover behavioral patterns and structural organization in legislative and political interactions [Ferreira et al. 2026].

Several topological methods have been proposed for this task. Local methods, such as Common Neighbors (CN), Adamic-Adar (AA), Resource Allocation (RA), and Preferential Attachment (PA), mainly use information from the immediate neighborhood of the vertices [Adamic and Adar 2003;

Copyright©2026 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

Zhou et al. 2009]. Quasi-local methods, in turn, incorporate short paths or walks, seeking to capture structural patterns beyond CN, but without the computational cost of global methods [Lü et al. 2009; Aziz et al. 2020]. However, the performance of these methods may vary according to network properties, such as density, average degree, clustering coefficient, transitivity, and assortativity [Feng et al. 2012].

This research empirically evaluates topological link prediction methods in eight real-world networks from four domains, with two networks per domain: social, scientific collaboration, e-mail communication, and electrical infrastructure. The investigation is guided by the following Research Questions (RQs): **RQ1:** does the performance of the methods vary across networks from different domains? **RQ2:** do quasi-local methods outperform simple local methods? **RQ3:** is there a dominant method among the evaluated networks? **RQ4:** do structural properties help explain the difficulty of the link prediction task?

A total of 19 methods were evaluated using five-fold cross-validation, preserving the connectivity of the training graphs and using balanced sampling of negative pairs. The results indicated that the *Superposed Random Walk* method with three steps (SRW-l3) achieved the best overall average performance, with an *Average Precision* (AP) of 0.9220, followed by the *Local Path Index* (LPI) with $\beta = 0.001$, with 0.9180, and *Resource Allocation Based on Common Neighbor Interactions* (RA-CNI), with 0.9086. A statistical analysis indicated that the advantage of SRW-l3 over LPI-beta-0.001 was not statistically significant, although SRW-l3 obtained a higher mean AP in most networks. In aggregate terms, quasi-local methods achieved the best average performances, while the electrical infrastructure networks were the most difficult. These results suggest that short paths and walks may contribute to greater stability across domains, although the relationship between structural properties and performance should be interpreted as exploratory.

2. RELATED WORK

Link prediction has been widely investigated as a structural inference task. The work of Liben-Nowell and Kleinberg [Liben-Nowell and Kleinberg 2007] is one of the main early formulations of the problem in social networks. Later, Lü and Zhou [Lü and Zhou 2011] and Martínez et al. [Martínez et al. 2016] systematized the field in surveys, organizing methods into categories such as local similarity, paths, random walks, probabilistic models, and supervised approaches, and more recent reviews add feature extraction, representation learning, and machine learning models [Kapoor et al. 2024]. These surveys, however, generally compile results obtained under heterogeneous conditions; the present work instead evaluates a representative subset of local and quasi-local indices under a single, uniform protocol across all networks.

Among local methods, indices based on CN and degree penalization stand out. Adamic and Adar [Adamic and Adar 2003] proposed a measure that assigns greater weight to less connected CN, while Zhou et al. [Zhou et al. 2009] introduced the RA index, also based on the contribution of shared neighbors. These methods are attractive due to their simplicity and low computational cost, but they may be limited when information from the immediate neighborhood is not sufficient to distinguish candidate pairs.

To overcome part of this limitation, quasi-local methods incorporate short paths or walks, exploiting information beyond immediate neighbors without requiring the cost of global methods. The LPI, proposed by Lü et al. [Lü et al. 2009], combines paths of length two and three to improve prediction in complex networks. Other extensions explore enhanced local information, local Bayesian models, and local community patterns [Liu et al. 2011; Cannistraci et al. 2013; Aziz et al. 2020]. In addition, Feng et al. [Feng et al. 2012] analyze link prediction from the perspective of clustering, reinforcing the importance of structural properties in interpreting method performance.

In relation to these works, this research adopts a specific empirical scope: comparing 19 local and quasi-local topological methods in eight real-world networks from distinct domains, while maintaining a uniform experimental protocol across the networks. In addition to the ranking of the methods, statistical significance of the ranking, performance by families, fold-level behavior of the best-performing method, and the exploratory relationship between structural properties and task difficulty are analyzed. Thus, the main contribution does not lie in proposing a new index, but rather in the comparative analysis of the behavior of topological methods on networks with distinct properties.

3. NETWORKS USED AND STRUCTURAL CHARACTERIZATION

The analyzed networks were *ego-Facebook*, *ca-GrQc*, *ca-HepTh*, and *email-Eu-core* (from SNAP); *socfb-Middlebury45*, *email-univ*, and *power-1138-bus* (from the Network Data Repository); and *power-grid* (from Netzschleuder). Two networks were selected for each of the four domains in order to reduce the influence of any single dataset on the domain-level interpretation. Before evaluating the methods, all networks were standardized as simple undirected graphs, with self-loops removed and the largest connected component selected. The originally directed *email-Eu-core* network was converted to an undirected representation. Therefore, the values presented in Table I correspond to the versions effectively used in the experiments; the exact sources are listed in the public repository.

In Table I, $|V|$ and $|E|$ denote the number of vertices and edges. Here, *local closure* refers to the tendency of a vertex's neighbors to be connected to each other, as captured by the average clustering coefficient ("Clust.") and transitivity ("Trans."). These properties describe the structural organization of the networks and support the exploratory interpretation of performance.

Table I. Structural properties of the networks after preprocessing.

Network	**Domain**	$\lvert V\rvert$	$\lvert E\rvert$	**Dens.**	**Avg. Deg.**	**Clust.**	**Trans.**	**Assort.**
ego-Facebook	Social	4,039	88,234	0.010820	43.691	0.606	0.519	0.064
socfb-Middlebury45	Social	3,069	**124,607**	0.026468	**81.204**	0.282	0.211	0.078
ca-GrQc	Scientific collab.	4,158	13,422	0.001553	6.456	0.557	**0.629**	**0.639**
ca-HepTh	Scientific collab.	**8,638**	24,806	0.000665	5.743	0.482	0.281	0.239
email-Eu-core	Communication	986	16,064	**0.033080**	32.584	0.407	0.267	-0.026
email-univ	Communication	1,133	5,451	0.008500	9.622	0.220	0.166	0.078
power-1138-bus	Infrastructure	1,138	1,458	0.002254	2.562	0.087	0.093	-0.080
power-grid	Infrastructure	4,941	6,594	0.000540	2.669	**0.080**	0.103	0.003

The networks exhibit relevant structural differences. Within the social domain, *socfb-Middlebury45* has the highest average degree, while *ego-Facebook* has the highest average clustering coefficient. The infrastructure networks (*power-1138-bus* and *power-grid*) have low average degree, low density, and the lowest clustering coefficients. Among the communication networks, *email-Eu-core* has the highest density and slightly negative assortativity, whereas *email-univ* is considerably sparser.

4. EXPERIMENTAL METHODOLOGY

For each network, a simple undirected graph $G = (V, E)$ was considered. The task consists of removing observed edges, scoring candidate pairs in the training graph, and verifying whether removed edges are ranked above sampled non-edges. All networks were evaluated using five folds. To preserve the connectivity of the training graphs, a spanning tree was kept fixed and only the remaining edges were distributed among the test folds. Thus, spanning-tree edges were always retained in the training graph and were not used as positive test instances. This design avoids disconnected training graphs and allows all methods to be compared under the same structural constraint, but it also means that the evaluation is restricted to the reconstruction of removable edges.

In each fold, the removed edges form the positive set E_i^+, while the negative set E_i^- contains pairs of vertices with no edge in the original network. A balanced ratio was adopted, with one negative pair for each positive pair. This balanced setting provides a controlled comparison among methods rather than reproducing the natural class imbalance of link prediction. Therefore, the protocol evaluates static link reconstruction under sampled negatives, not temporal link formation. The main experimental settings used in the implementation are summarized in Table II.

Table II. Main experimental settings used in the implementation.

Setting	**Value**
Random seed	42
Cross-validation folds	5
Negative sampling ratio	1:1
Fold datasets	40
Candidate pairs	505,084
Positive / negative pairs	252,542 / 252,542
LPI parameter	$\beta = 0.001$
LRW, SRW, and PFP steps	$l = 3$
Reported methods	19 local or quasi-local methods

The training graph in fold i is defined as $G_i^{\text{train}} = (V, E \setminus E_i^+)$ and the candidate set is $C_i = E_i^+ \cup E_i^-$. All methods computed scores exclusively from G_i^{train}. A total of 19 methods were evaluated: CN, AA, RA, PA, Jaccard coefficient (JA), Salton index (SA), Sørensen index (SO), Hub Promoted Index (HPI), Hub Depressed Index (HDI), Local Leicht-Holme-Newman index (LLHN), Individual Attraction 2 (IA2), Local Naive Bayes with Resource Allocation (LNB-RA), CAR-RA, Functional Similarity Weight (FSW), RA-CNI, LPI-beta-0.001, Local Random Walk with three steps (LRW-l3), SRW-l3, and PropFlow with three steps (PFP-l3). All reported methods are local or quasi-local; global and dense global methods were not included in the reported comparison, since one of the motivations for local and quasi-local indices is precisely to avoid the higher computational cost of global approaches. The final execution generated 40 fold datasets and 760 method-fold evaluations, corresponding to 8 networks, 5 folds, and 19 methods.

The hyperparameters of the parametric methods were fixed rather than tuned per network, following common practice and the values reported in the literature: $\beta = 0.001$ for LPI, as suggested by Lü et al. [Lü et al. 2009] for the low-weight regime of longer paths, and $l = 3$ steps for the walk-based methods (LRW, SRW, and PFP), which restricts them to the quasi-local neighborhood. A broader sensitivity analysis for β and for the number of steps is provided in the public repository; using a single fixed configuration keeps the comparison uniform across all networks.

AP was used as the main metric because link prediction is treated as a ranking task. ROC-AUC, Precision, Recall, F1, and NDCG were also computed at $K = |E_i^+|$, that is, using the same number of predictions as the number of removed positive edges. AP was calculated as:

$$AP = \frac{1}{P} \sum_{r=1}^{|C_i|} \text{Prec}(r) \cdot rel(r), \tag{1}$$

where P is the number of positive pairs, $\text{Prec}(r)$ is the precision at rank r, and $rel(r) = 1$ when the item at position r is positive, and $rel(r) = 0$ otherwise. Final values were aggregated using mean and Standard Deviation (SD) over the five folds.

Since the cutoff was defined as $K = |E_i^+|$ and the candidate set was balanced, Precision, Recall, and F1 have the same value at this cutoff. For compactness, the tables report Precision at this cutoff, while Recall and F1 are omitted because they are numerically equivalent to Precision under this setting.

Because several topological indices assign identical scores to multiple candidate pairs, ties were resolved deterministically by the ordering of the previously shuffled candidate table, using the fixed random seed. This makes the evaluation reproducible, although the reported values are not averaged over multiple tie-breaking permutations.

5. RESULTS AND DISCUSSION

5.1 Overall performance of the methods

Table III presents the ten methods with the highest average performance considering all networks. The SRW-l3 method obtained the best overall result, with a mean AP of 0.9220. It is followed by LPI with $\beta = 0.001$ (LPI-beta-0.001), with a mean AP of 0.9180, and RA-CNI, with a mean AP of 0.9086. PFP-l3 presented a result practically equivalent to RA-CNI, with a mean AP of 0.9086.

Table III. Ten methods with the highest average performance across the evaluated networks.

Rank	Method	Family	AP	AP SD	ROC-AUC	Prec.	NDCG
1	SRW-l3	Quasi-local/walks	**0.9220**	0.0835	**0.9223**	**0.8721**	**0.8948**
2	LPI-beta-0.001	Quasi-local/paths	0.9180	0.0818	0.9192	0.8670	0.8901
3	RA-CNI	Enhanced local	0.9086	0.0856	0.9106	0.8555	0.8797
4	PFP-l3	Quasi-local/walks	0.9086	0.0794	0.9139	0.8612	0.8832
5	LRW-l3	Quasi-local/walks	0.8847	0.1185	0.8863	0.8292	0.8567
6	RA	Local similarity	0.8809	0.1231	0.8811	0.8283	0.8574
7	LNB-RA	Local Bayesian	0.8809	0.1230	0.8811	0.8281	0.8573
8	IA2	Enhanced local	0.8800	0.1230	0.8805	0.8278	0.8569
9	AA	Local similarity	0.8796	0.1220	0.8801	0.8267	0.8559
10	FSW	Enhanced local	0.8781	0.1219	0.8787	0.8246	0.8541

Figure 1 shows that top-ranked methods keep high AP in most networks but drop markedly in the infrastructure networks, indicating that task difficulty depends on both the method and the network structure.

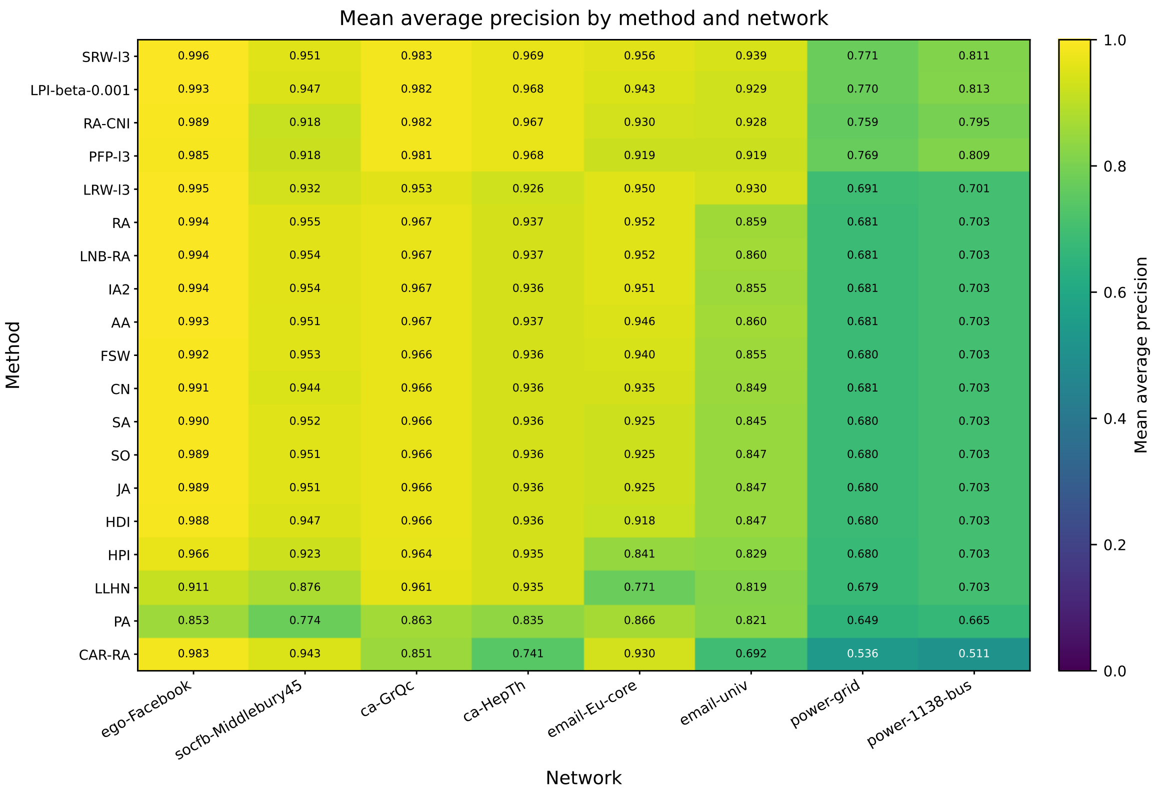

Fig. 1. Mean AP of the link prediction methods in each evaluated network.

The results indicate that quasi-local methods based on short paths or walks presented an advantage over purely local methods, as the top-ranked methods all combine local information with a structural extension (short walks, length-three paths, or interactions among CN). This advantage is not absolute: local methods such as RA, AA, and CN also remained competitive in several networks.

At the fold level, SRW-l3 ranked first in 30 of the 40 evaluated folds; the exceptions occurred mostly in *socfb-Middlebury45* and the infrastructure networks, where local indices were occasionally more competitive.

5.2 Statistical significance of the ranking

Because the differences among the top-ranked methods are small, a statistical analysis was performed to assess whether they are consistent across networks. For each network, the mean AP of every method was computed, and the four best methods (SRW-l3, LPI-beta-0.001, RA-CNI, and PFP-l3) were compared. A Friedman test across the eight networks rejected the null hypothesis of equal performance (statistic = 20.55, $p < 0.001$), indicating that at least one method differs from the others. Pairwise Wilcoxon signed-rank tests with Holm correction were then applied. SRW-l3 was significantly superior to PFP-l3 ($p_{\text{adj}} = 0.047$) and to RA-CNI ($p_{\text{adj}} = 0.047$), and the same held for LPI-beta-0.001 over both methods ($p_{\text{adj}} = 0.047$). However, the difference between SRW-l3 and LPI-beta-0.001 was *not* statistically significant ($p_{\text{adj}} = 0.109$), even though SRW-l3 obtained a higher mean AP in seven of the eight networks. PFP-l3 and RA-CNI were statistically indistinguishable ($p_{\text{adj}} = 0.945$). These results quantitatively support the interpretation that short walks and short paths provide comparable predictive signals, and that no single method dominates the others by a wide margin under the adopted protocol. As the network-level test is based on eight networks, these conclusions should be read as indicative rather than definitive.

5.3 Performance by network

Table IV presents the best method obtained in each network, its AP, AP SD, Precision at $K = |E_i^+|$, the mean AP of all methods in that network, and two relevant structural properties: average clustering coefficient and average degree. SRW-l3 was the best method in six of the eight evaluated networks; in *socfb-Middlebury45* the local index RA was slightly superior, and in *power-1138-bus* LPI-beta-0.001 obtained the highest AP. Performance varied strongly across domains.

Table IV. Best method by network and associated structural properties.

Network	**Best method**	**Best AP**	**AP SD**	**Prec.**	**Mean AP**	**Clust.**	**Avg. Deg.**
ego-Facebook	SRW-l3	**0.9959**	0.0003	**0.9766**	**0.9781**	0.606	43.691
socfb-Middlebury45	RA	0.9549	0.0011	0.8957	0.9313	0.282	**81.204**
ca-GrQc	SRW-l3	0.9830	0.0029	0.9677	0.9569	0.557	6.456
ca-HepTh	SRW-l3	0.9687	0.0019	0.9416	0.9265	0.482	5.743
email-Eu-core	SRW-l3	0.9564	0.0020	0.8942	0.9197	0.407	32.584
email-univ	SRW-l3	0.9387	0.0045	0.8738	0.8596	0.220	9.622
power-1138-bus	LPI-beta-0.001	0.8126	0.0367	0.7540	0.7125	0.087	2.562
power-grid	SRW-l3	0.7707	0.0106	0.6783	0.6899	**0.080**	2.669

The *ego-Facebook* network presented the highest performance, with an AP of 0.9959 for SRW-l3 and a mean AP of 0.9781 considering all methods. This result is consistent with its high average degree and high average clustering coefficient, which favor methods based on neighborhoods, short paths, and triadic closure. In *socfb-Middlebury45*, which has a very high average degree but lower clustering, the simple RA index was marginally the best method.

The scientific collaboration networks also presented high performance. *ca-GrQc* obtained an AP of 0.9830 for the best method, while *ca-HepTh* obtained an AP of 0.9687. Although these networks are sparse, their clustering coefficients are relatively high, suggesting the presence of local coauthorship groups that favor topological methods.

The infrastructure networks presented the lowest AP: even the best method reached only 0.8126 in *power-1138-bus* and 0.7707 in *power-grid*, and Precision at $K = |E_i^+|$ followed the same pattern (SRW-l3 achieved 0.6783 in *power-grid*, versus above 0.89 in the denser networks). With low average degree and clustering, these networks offer fewer CN, triangles, and informative short paths, providing weaker topological signals for local and quasi-local methods.

5.4 Comparison by method families

Table V summarizes the average performance by method family using AP, ROC-AUC, Precision, and NDCG. The column "N" indicates the number of methods evaluated in each family. The path-based quasi-local family obtained the highest mean AP, followed by the walk-based quasi-local family. However, the interpretation of the first position should be made with caution, since this family contains only LPI in the evaluated configuration.

Table V. Average performance by method family.

Family	N	AP	ROC-AUC	Prec.	NDCG
Quasi-local/paths	1	0.9180	0.9192	0.8670	0.8901
Quasi-local/walks	3	0.9051	0.9075	0.8542	0.8782
Local Bayesian	1	0.8809	0.8811	0.8281	0.8573
Local similarity	9	0.8688	0.8749	0.8202	0.8483
Enhanced local	4	0.8600	0.8608	0.8095	0.8368
Degree-based local	1	0.7908	0.7902	0.7174	0.7431

On the other hand, simple local methods still presented relevant performance. The local similarity family achieved a mean AP of 0.8688, showing that the immediate neighborhood contains a strong predictive signal in several networks. The advantage of quasi-local methods lies in exploiting additional information without resorting to global methods with higher computational cost. From a practical standpoint, this suggests that quasi-local walks and paths are most worthwhile in dense networks with strong local closure, whereas in very sparse infrastructure networks the additional structural information yields only limited gains over simple local indices.

Overall, the results provide evidence regarding the proposed research questions. Regarding RQ1, performance varied across networks, with greater difficulty in the infrastructure networks and greater ease in *ego-Facebook*. Regarding RQ2, quasi-local methods presented the best average performances. Regarding RQ3, SRW-l3 was the best method in most networks, although the statistical analysis in Section 5.2 indicated that its advantage over LPI-beta-0.001 was not significant. Finally, regarding RQ4, properties such as clustering, average degree, and local organization help interpret the difficulty of the task, but this relationship should be understood as exploratory, since only eight networks were evaluated.

6. CONCLUSIONS

This study compared 19 local and quasi-local topological methods for link prediction in eight real-world complex networks from different domains. The experimental protocol used five-fold cross-validation, preserved the connectivity of the training graphs, and adopted balanced negative sampling. This setting allowed the methods to be compared under uniform conditions, although it should be interpreted as static link reconstruction rather than temporal link formation.

The results provide evidence that method performance varies across network domains. The highest AP values were observed in *ego-Facebook* and in the scientific collaboration networks, whereas the infrastructure networks were the most difficult. This pattern suggests that structural properties such as average degree, clustering coefficient, and local closure influence the amount of topological evidence available to local and quasi-local predictors.

Among the evaluated methods, SRW-l3 achieved the highest mean AP and was the best method in most networks. However, a statistical analysis (Friedman test with Wilcoxon signed-rank pairwise comparisons and Holm correction) showed that its advantage over LPI-beta-0.001 was not statistically significant, even though it was significant over PFP-l3 and RA-CNI. Therefore, the results do not support broad superiority of a single method, but indicate that short walks and short paths are effective sources of structural information under the adopted evaluation protocol. Simple local methods, such as RA, AA, and CN, also remained competitive, reinforcing the relevance of local neighborhoods as strong baselines.

The main limitations of this study are the still limited number of networks, the static reconstruction protocol, and the use of balanced negative sampling. Future work should include more domains, temporal evaluation, different negative sampling ratios, directed and weighted networks, and comparisons with supervised approaches, graph embeddings, and Graph Neural Networks (GNNs).

Availability of Code and Materials

The source code and experimental materials used in this study are openly available at: `https://github.com/avelando/networks`.

Acknowledgments

This work was developed during the course SME5924 (Dynamical Processes in Complex Networks), taught by Prof. Francisco Aparecido Rodrigues, whom the authors thank for his guidance.

REFERENCES

Adamic, L. and Adar, E. Friends and neighbors on the web. *Social Networks* 25 (3): 211–230, 2003.

Aziz, F., Gul, H., Uddin, I., and Gkoutos, G. Path-based extensions of local link prediction methods for complex networks. *Scientific Reports* 10 (1): 19848, 2020.

Cannistraci, C., Alanis-Lobato, G., and Ravasi, T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Scientific Reports* vol. 3, pp. 1613, 2013.

Feng, X., Zhao, J., and Xu, K. Link prediction in complex networks: A clustering perspective. *The European Physical Journal B* 85 (1): 3, 2012.

Ferreira, M. W., de Sousa, G., and Leal, J. Detecção de comunidades entre os membros da câmara dos deputados do brasil durante o período de 2019 a 2023: Uma abordagem utilizando redes complexas. In *Anais do XLI Simpósio Brasileiro de Bancos de Dados*. SBC, Porto Alegre, RS, Brasil, pp. 701–713, 2026.

Kapoor, P., Kaushal, S., Kumar, H., and Kanwar, K. A survey on feature extraction and learning techniques for link prediction in homogeneous and heterogeneous complex networks. *Artificial Intelligence Review* vol. 57, pp. 348, 2024.

Liben-Nowell, D. and Kleinberg, J. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58 (7): 1019–1031, 2007.

Liu, Z., Zhang, Q.-M., Lü, L., and Zhou, T. Link prediction in complex networks: A local naive bayes model. *EPL* 96 (4): 48007, 2011.

Lü, L., Jin, C.-H., and Zhou, T. Similarity index based on local paths for link prediction of complex networks. *Physical Review E* 80 (4): 046122, 2009.

Lü, L. and Zhou, T. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications* 390 (6): 1150–1170, 2011.

Martínez, V., Berzal, F., and Cubero, J.-C. A survey of link prediction in complex networks. *ACM Computing Surveys* 49 (4), 2016.

Newman, M. The structure and function of complex networks. *SIAM Review* 45 (2): 167–256, 2003.

Zhou, T., Lü, L., and Zhang, Y.-C. Predicting missing links via local information. *The European Physical Journal B* 71 (4): 623–630, 2009.