



HAL
open science

Link Prediction Without Learning

Simon Delarue, Thomas Bonald, Tiphaine Viard

► **To cite this version:**

Simon Delarue, Thomas Bonald, Tiphaine Viard. Link Prediction Without Learning. European Conference on Artificial Intelligence, Oct 2024, Santiago de compostela, Galicia, Spain. hal-04643971

HAL Id: hal-04643971

<https://hal.science/hal-04643971v1>

Submitted on 10 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Link Prediction Without Learning

Simon Delarue^{a,*}, Thomas Bonald^a and Tiphaine Viard^b

^aLTCI, Télécom Paris, Institut Polytechnique de Paris

^bi3, Télécom Paris, Institut Polytechnique de Paris

Abstract. Link prediction is a fundamental task in machine learning for graphs. Recently, Graph Neural Networks (GNNs) have gained in popularity and have become the default approach for solving this type of task. Despite the considerable interest for these methods, simple topological heuristics persistently emerge as competitive alternatives to GNNs. In this study, we show that this phenomenon is not an exception and that GNNs do not consistently establish a performance standard for link prediction on graphs. For this purpose, we identify several limitations in the current GNN evaluation methodology, such as the lack of variety in benchmark dataset characteristics and the limited use of diverse baselines outside of neural methods. In particular, we highlight that integrating feature information into topological heuristics remains a little-explored path. In line with this observation, we propose a simple non-neural model that leverages local structure, node feature, and graph feature information within a weighted combination. Experiments conducted on large variety of networks indicate that the proposed approach outperforms existing state-of-the-art GNNs and increases generalisation ability. Contrasting with GNNs, our approach does not rely on any learning process and therefore achieves superior results without sacrificing efficiency, showcasing a reduction of one to three orders of magnitude in computation time.

1 Introduction

Link prediction aims at estimating the likelihood of a relation between two entities in a network. Considering the ubiquity of relational data, link prediction has numerous real-world applications, such as predicting protein-protein interactions, identifying spurious relations, building knowledge graphs, or recommending items. Recently, Graph Neural Networks (GNNs) emerged as a powerful technique to address downstream machine learning tasks on graphs, such as node classification, graph classification, or link prediction [39, 46]. The expected goal was that, by considering both topological information and explicit node features within a unified neural framework, performance would necessarily rise on several benchmarks.

In practice however, results are more nuanced, as long-established topological heuristics remain robust competitors in link prediction tasks [45, 44, 18, 38, 5]. These simple heuristics rely solely on local node relationships near the predicted edge and do not involve any learning process, contrasting the complexity of neural approaches.

To explain this phenomenon, recent works have highlighted how pitfalls in the evaluation frameworks could lead to an inaccurate assessment of GNN performance. Among these, several studies

have emphasised how GNN models, due to their message passing learning scheme, tend to overfit a restricted number of datasets with highly similar characteristics, namely *Cora*, *Pubmed* and *Citeseer* [12, 27, 48, 29, 25]. Other studies have criticised the lack of unified setting in the evaluation procedures, and have proposed solutions regarding the definition of split size, evaluation metric, and negative sampling strategy [22, 18].

Despite these efforts, both existing studies on GNNs and their critics often overlook the importance of well-designed baseline models in assessing the true progress of recent neural approaches [9]. Consequently, the current literature in the field primarily relies on comparing complex neural approaches, such as specialised GNNs, against simpler neural approaches, such as multi-layer perceptrons or “simple” GNNs.

Non-neural topological heuristics, however, have several advantages that makes them worth pursuing. They are usually significantly less complex and thus more interpretable, they run on simpler architectures, usually at a fraction of the computational cost of neural approaches, and they require significantly less hyperparameter tuning, which makes them more appealing in a variety of contexts.

We argue that to understand and tackle these issues, strong non-neural heuristics and robust evaluation frameworks should be studied jointly. To address these limitations, we shift our focus to non-neural heuristics and propose to solve link prediction tasks by enhancing topological heuristics with explicit feature information. To achieve this, we leverage both the topological information of the graph and features at node-level and graph-level. We define the topological information using traditional local similarity indices, such as Common Neighbours [28], Adamic Adar index [1], or Resource Allocation index [47]. Regarding feature information, at the node-level, we compute node feature similarity between entities near the predicted edge. At the graph-level, we measure the distance in pairwise node feature similarity distributions between positive and negative edges in the network. Similar to existing GNNs, our approach harnesses information from multiple dimensions of the original graph. However, unlike neural approaches, our method does not require any learning mechanism.

Furthermore, we discuss how variations in the evaluation settings make comparisons between models difficult, and propose a robust evaluation framework addressing these limitations. In particular, we rely on various datasets with different structures and distributions, both in terms of topology and of feature homophily. Moreover, we include both simple topological heuristics and our enhanced topological heuristics in the benchmarks.

Through extensive experiments conducted under this evaluation framework, we show that the proposed approach outperforms recent

* Corresponding Author. Email: simon.delarue@telecom-paris.fr.

GNN architectures and exhibits a higher generalisation ability. Additionally, we highlight the scalability performance of the proposed method compared to state-of-the-art GNNs. Specifically, we illustrate that our model enables a reduction in computation time between one to three orders of magnitude compared to its neural competitors. Overall, our experiments suggest that the complexity of GNNs may not be required for link prediction tasks, and that enhanced heuristics can overcome the limitations of neural approaches in terms of generalisation and performance. Our key contributions are summarised as follows:

- We show that GNNs evaluation frameworks (i) contain irregularities, and (ii) lack well-designed non-neural heuristics.
- We introduce a non-neural baseline model for link prediction that combines traditional topological heuristics with node and graph features.
- We design a robust evaluation framework to assess model performance. This setting includes real-world networks with various characteristics, and unifies the selection of data splits, negative samples, and evaluation metrics.
- Under this setting, we show that the proposed non-neural baseline achieves or beats state-of-the-art performance against recent GNNs while running up to 10^3 times faster.

2 Related Work

We begin by presenting a literature review on various link prediction techniques. Then, we focus on the criticisms raised regarding the evaluation procedures within the GNNs field.

2.1 Link prediction

It is a standard procedure to categorise link prediction models into three main categories: topological heuristics, embedding approaches, and GNNs.

Topological heuristics. Early research have investigated prediction metrics based on the topological similarity score between nodes. These metrics heavily rely on the homophily principle, which suggests that entities with similar characteristics tend to connect in a network. Therefore, the likelihood of a connection between two nodes can be estimated based on the number of common relations they share. Various levels of similarity indices exist. Among the simplest ones are local indices, such as Common Neighbours (CN) [28], Adamic Adar (AA) [1], or Resource Allocation (RA) [47], which directly depend on the size of the overlap of the neighbourhoods of two end nodes. In contrast, global indices, such as Katz [13] or PageRank [3], define similarity between nodes by leveraging path-related expressions. All these topological heuristics can be computed directly on the graph and do not require any form of learning process, making them highly efficient approaches. However, they lack general graph structure expressiveness and may therefore struggle to generalise well [20].

Embeddings. Also called latent feature-based methods, these approaches learn low-dimensional representations of nodes and use pairwise embeddings to predict the existence of a link between two nodes. Well-known methods include matrix factorization [16], DeepWalk [31], LINE [36], and node2vec [10]. Limitations of these approaches stem from their naturally transductive setting and their lack of parameter sharing. Furthermore, as for the topological heuristics, they typically do not take additional node properties into account.

GNNs. Neural-based approaches address these challenges. These methods learn shared parameters through a message passing scheme in order to derive embeddings of the elements of the graph. Two main categories of GNNs for link prediction can be distinguished. The first category includes node-based approaches, inspired by auto-encoder architectures such as Graph AutoEncoder (GAE) and Variational Graph AutoEncoder (VGAE) [14]. These models typically combine a Graph Convolutional Network (GCN) [15] architecture with an inner product decoder to predict the existence of links. Consequently, several foundational GNN models, such as GraphSage [11] or GAT [37], can be adapted to the link prediction task. However, researchers have showed that such approaches fail to distinguish non-isomorphic links [45]. They proposed graph-based approaches, such as SEAL, which involves learning the representation of each link using an embedding of its surrounding local subgraph. Despite its great performance, this approach entails the extraction and embedding of a number of subgraphs equals to the number of positive and negative edges, leading to high computational cost. Recent works proposed to address this challenge by learning structural information more efficiently [44, 49, 5, 38]. Nevertheless, in practice, the computational complexity of the GAE model is still a lower-bound for the most recent approaches [44, 38].

Despite the simplicity of topological heuristics, their performance often matches or exceeds that of embedding-based models or GNNs. This has led to numerous studies proposing to boost GNNs by incorporating topological features. For example, SEAL [45] relies on *node labeling* trick to enforce the learning of structural features. NCN/NCNC [38] combines common neighbours index with message passing neural network representations. Neo-GNN [44] learns to generate useful structural features to later enhance GNN representations. Link-MoE [23] uses an ensemble of topological heuristics for weighting GNN outputs. On the other hand, limited work has been done to enhance topological heuristics with additional information. For instance, some early initiative proposed introducing a parameter based on the change of common neighbours' degree [7]. However, this approach did not leverage any node feature information. In this work, we focus on adaptively incorporating explicit feature information into topological heuristics in order to design simple yet robust competitors to GNNs.

2.2 Evaluation frameworks

Several previous works highlighted how the progress achieved by state-of-the-art neural models was actually lower under a rigorous and consistent evaluation framework [34, 19, 32, 17]. The fast moving field of deep learning exposes to several pitfalls for model evaluation. Among them, researchers criticise the low reproducibility for recommender systems [9, 18], underscore the influence of train-validation-test splits on GNN performance [2], evoke the diversity in the chosen evaluation metrics [18], or highlight the bias and overfitting introduced by the limited number of benchmark datasets used in the literature [12, 27, 43, 29]. To address these challenges, some studies propose to include several large-scale graph datasets in benchmarks [12], or to rely on synthetic graph [24]. Others develop a full design space evaluation relying on a large number of model-task combinations [43], or introduce an open-source benchmarking framework for GNN evaluation [8]. Additional work develop a novel evaluation setting based on a personalised and non-trivial negative sampling strategy [18].

However, despite recent calls for designing robust non-neural baselines to correctly assess the performance of deep learning-based

methods [41, 9, 21], the proposals closest to our work [2, 18] lack implementation of simple yet strong methods. We argue that the simultaneous introduction of both a robust evaluation framework and enhanced non-neural baselines is crucial for a comprehensive assessment of model performance.

To address these limitations, we design a robust evaluation framework in which we reassess GNNs performance against both simple and specifically designed non-neural baselines.

3 Problem Definition

We consider the problem of predicting the existence of links in a graph. Let $G = (V, E, X)$ be an attributed graph, where V represents a set of vertices, $E \subseteq V \times V$ is a set of edges, and $X^{|V| \times L}$ is an attribute matrix assigning L numerical attributes to each node in G . The set of nodes connected to u in G , denoted with $\mathcal{N}(u) = \{v : (u, v) \in E\}$ corresponds to the neighbourhood of u . Each node is associated with a class label, and we denote with $Y \in \{0, \dots, C\}^{|V|}$ the vector of node labels given C classes.

The goal of the link prediction task is to infer the existence of missing links in the graph, having knowledge about all nodes and their features, but knowing only a subset of their relations. More formally, we try to learn a mapping $g : E \rightarrow \{0, 1\}$, given a subset of existing links $E' \subseteq E$ as training data.

4 Enhanced Topological Heuristics

In this section, we describe our enhanced non-neural topological heuristics for link prediction. Contrasting with their original form, our enhanced heuristics do not restrict to the structural information of the network to predict the existence of a link, but instead combines three key components: (i) a node-level topological weight, (ii) a node-level feature similarity weight, and (iii) a graph-level feature similarity measure. We illustrate our methodology in Figure 1.

Intuitively, incorporating information from both the structure of the graph and its explicit node features expands the range of information used from the initial data. Similar to the message passing scheme commonly used by GNNs, the aim of this enhancement is to improve the expressiveness of the proposed model. Furthermore, by considering information from both the node-level and the graph-level, we seek to mitigate limitations in the generalisability of topological heuristics.

4.1 Node-level topological and feature similarity weights

In our approach, we compute the node-level topological weight between two nodes u and v , denoted $\omega_{u,v}^A$, using the definition of each selected topological heuristics. Then, applying the homophily principle to the node features of the graph, we propose to incorporate feature information relying on pairwise node feature similarities in the neighbourhood defined by the topological heuristic. Thus, the feature similarity weight between two nodes u and v , denoted $\omega_{u,v}^X$ is computed using any similarity function on node attribute vectors (in this work we use cosine similarity). From these, we define node-level topological and feature similarity weights for three heuristics:

Enhanced Common Neighbours (ECN) [28]:

$$\omega_{u,v}^A = |\mathcal{N}(u) \cap \mathcal{N}(v)| \quad (1)$$

For ECN, we define the feature similarity weight of the link as the feature similarity between its two end nodes:

$$\omega_{u,v}^X = s(X_u, X_v) \quad (2)$$

Enhanced Adamic Adar index (EAA) [1]:

$$\omega_{u,v}^A = \sum_{c \in \mathcal{N}(u) \cap \mathcal{N}(v)} \frac{1}{\log |\mathcal{N}(c)|} \quad (3)$$

Drawing inspiration from the residual connection mechanism which helps reduce oversmoothing in GNN learning [11, 40], we define feature similarity weight for the EAA index by combining the target nodes' feature similarity with the pairwise node feature similarities in their neighbourhood.

$$\omega_{u,v}^X = s(X_u, X_v) + \sum_{c \in \mathcal{N}(u) \cap \mathcal{N}(v)} s(X_u, X_c) s(X_c, X_v) \quad (4)$$

Enhanced Resource Allocation index (ERA) [47]:

$$\omega_{u,v}^A = \sum_{c \in \mathcal{N}(u) \cap \mathcal{N}(v)} \frac{1}{|\mathcal{N}(c)|} \quad (5)$$

Given the closeness of the definitions of the ERA and EAA indexes, we define the feature similarity weight of ERA index using Equation 4.

4.2 Graph-level feature similarity measure

We consider the graph-level feature similarity γ of a network by measuring the distance between distributions of pairwise feature similarity in its positive and negative edges (in practice we use Jensen-Shannon distance and $\gamma \in [0, 1]$). Intuitively, γ measures the relationship between node feature similarity and structural information in the network. Since γ is observed for each network, no learning process is required.

To support our choice for γ , we examine the disparities in pairwise node feature similarity distributions between positive and negative edges in three real-world networks in Figure 2. For both Cora and CS, we observe a strong contrast between pairwise feature similarity among connected nodes and disconnected ones. In these scenarios, node feature information proves crucial in predicting pairwise connections. Conversely, in Wikivitals and Wikivitals-fr, nodes exhibit similar features regardless of their pairwise connectivity.

4.3 Link prediction using enhanced topological heuristics

In the last step of our approach, we combine the node-level topological and feature weights with the graph-level similarity measure γ . In order to predict the existence of an edge between u and v , denoted $e_{u,v}$, each enhanced topological heuristic uses the following:

$$e_{u,v} = (1 - \gamma) \tilde{\omega}_{u,v}^A + \gamma \tilde{\omega}_{u,v}^X \quad (6)$$

where $\tilde{\omega}_{u,v}^A$ and $\tilde{\omega}_{u,v}^X$ are topological and feature similarity weights normalised over all nodes. In contrast to GNN models, the proposed prediction step does not involve semi-supervised learning. Consequently, class label information for the nodes is not required.

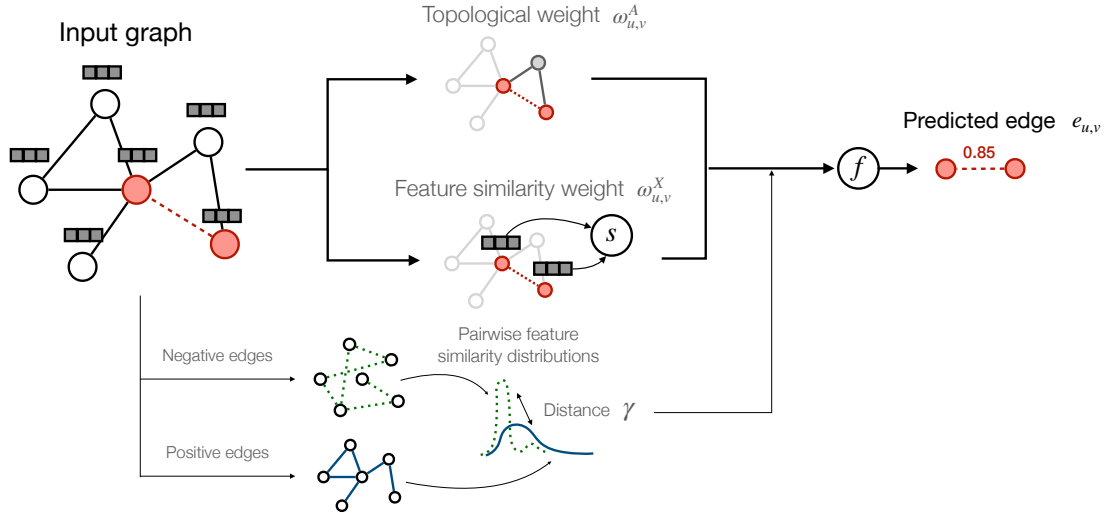


Figure 1. Enhanced topological heuristic for link prediction. For each target link, we compute a node-level topological weight, and a node-level feature similarity weight using a similarity metric s . Additionally, we measure graph-level feature similarity as the distance γ between pairwise feature similarity distributions for positive and negative edges in the graph. Finally, we predict the existence of a link using a linear combination f of both the topological and feature weights and using γ as a controlling parameter.

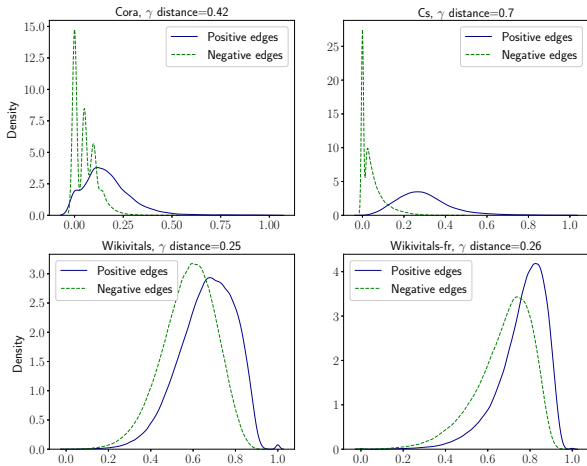


Figure 2. Distributions of pairwise node feature similarity among positive and negative edges. For each network, we report the distance γ between the two distributions.

We expect to observe high values for γ in networks where connected nodes share similar features, and where disconnected nodes exhibit dissimilar features. In such cases, increasing the contribution of node-level feature information should improve the prediction of the existence of an edge. Conversely, we expect low values for γ in networks in which the relation between the graph structure and the node feature similarity is not easily observable. For these networks, our model will adaptively assign greater importance to the topological information over the feature information.

In summary, our topological weight aims to maintain the expressive power of topological heuristics by incorporating node-level structural information near the predicted edges. Integrating the node feature similarity weight expands the range of information beyond the graph topology alone and can be highly discriminatory regarding the existence of an edge, as seen in Figure 2. Finally, to anticipate the variations in relations between node feature similarity and

graph structure across networks, and thus improve generalisability, we adaptively combine both the node-level weights using graph-level information.

5 Limits of Existing Evaluation Frameworks

We discuss two main practices in evaluation procedures that could mislead readers regarding the assessment of GNN performance.

5.1 Diversity of settings

Evaluating link prediction tasks involves several choices in experimental settings, including determining training-validation-test splits, sampling negative edges, choosing evaluation metric, and selecting datasets. Regarding data splits, commonly used datasets like Cora, Pubmed and Citeseer are typically evaluated using random initializations on splits with sizes 85%-5%-10% [14, 49] or 70%-10%-20% [38, 5]. Alternatively, new splitting strategies are proposed [25]. In contrast, OGB [12] datasets are typically used with fixed splits for each network [44, 5, 38]. We argue that the inconsistency in split sizes makes it difficult to assess the progress among different approaches. Furthermore, systematically using fixed splits can hinder generalisation limitations and result in dataset overfitting over time [2]. As for data splits, the size of the set of negative edges used to evaluate models, and the sampling strategy of its elements – whether random [45, 4] or following a specific procedure [35, 12, 49] – can vary from one experiment to the other. Again, such differences in practice can render a fair comparison between approaches difficult. Similar observation applies to the set of existing evaluation metrics; existing works alternates between Area Under the ROC-Curve (AUC) [14, 45, 49, 18], Average Precision (AP) [14, 45, 49], or Mean Reciprocal Rank (MRR) and Hit Rate@ k (Hit@ k) [49, 44, 18, 38, 5]. Considering the choice of network for benchmarks, several works have emphasised how GNNs were evaluated on a subset of datasets with limited diversity [12, 27, 29, 33, 25]. Three of them, namely Cora, Pubmed and Citeseer have been

extensively used in the recent benchmarks. Researchers showed that they all exhibit high homophily. However, there is a strong assumption that this characteristic greatly favours GNNs, since their design is built upon a message passing scheme [48, 24, 29]. Overall, both the lack of unified settings and the systematic use of highly specific elements in the evaluation frameworks makes it challenging to fully assess the progress achieved by the proposed approaches.

5.2 Limited use of simple non-neural baselines

To assess the extent to which simple approaches are considered in current benchmarks, we conduct a literature review on the usage of both simple and enhanced topological heuristics in link prediction with GNNs (we summarise the detailed procedure in the supplementary material [6]). We illustrate the proportion of the literature articles in which different categories of baseline are used in Figure 3 ($n = 88$). We observe that GNNs have become the standard benchmark against other proposed GNNs, appearing in over 95% of the reviewed articles. This trend is unsurprising given the neural nature of the proposed approaches.

However, despite their performance in several recent benchmarks, topological heuristics are considered as baselines against GNNs in less than 15% of the reviewed articles. Additionally, none of the reviewed articles incorporate any form of enhanced topological heuristics. We argue that overlooking simple topological heuristics in GNN benchmarks may result in the assessment of overvalued performance. Moreover, we consider that to ensure fair comparisons, GNNs should be benchmarked against competitive baselines that harness sufficient information.

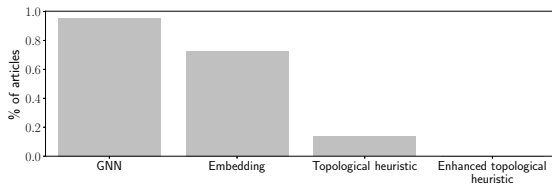


Figure 3. Proportion of reviewed literature articles in which each category of baseline appears. Over 95% of the reviewed articles compare their approach against GNN-based models. This proportion drops down below 15% when considering topological heuristic baselines.

6 Robust Evaluation Framework

We address the limitations of existing evaluation procedures and propose a robust evaluation framework which unifies settings across experiments and relies on a large variety of real-world networks. We provide details about baselines and GNN models in Appendix C in the supplementary material [6].

6.1 Model evaluation

To avoid the aforementioned pitfalls regarding split choice, we randomly divide the edge set of each graph into s training, validation and test splits with size 85%-5%-10%. For each split, we randomly sample a set of negative edges and fix the size of this set to be equal to the number of positive elements in the considered split. For fair comparison, we rely on the same splits and the same set of negative edges across all models (except for GAE and VGAE which involve

sampling new negative edges at each training step). Overall performance is obtained by averaging results over the s different runs (in practice $s = 3$).

Considering the lack of consensus regarding evaluation metrics, we verify the consistency of performance in our benchmarks by relying on the broadest set of metrics. For this purpose, for all networks, we use both AUC and AP, as well as ranking-based metrics MRR and Hit@ k . For all experiments, we use $k \in \{20, 50, 100, 200\}$, which covers the range of values being frequently used in the literature [12, 38, 5, 18].

6.2 Datasets

We ensure to include datasets with a large variety of characteristics considering their size, density and homophily. First, we consider the well-known homophilous networks *Cor*_a, *Pubmed**¹ and *Citeseer* [42]. We also include three other frequently used homophilous graphs: *CS* [2], a co-authorship graph, *Photo* [26], an Amazon co-purchase network, and *Ogbn-arxiv*, a significantly larger citation network [12]. Considering heterophilous datasets, we include *Actor*, an actor-induced subgraph [30], as well as universities web pages graphs *Cornell* and *Wisconsin*. Furthermore, we consider four real-world networks from Wikipedia, each characterised by a substantial number of features. In contrast to traditional datasets that typically exhibit either high node homophily or heterophily, the node homophily distributions of these networks are more uniform.

In summary, in the proposed evaluation framework, we unify training-validation-test splits, negative samples, and evaluation metrics across all evaluated models. To ensure generalisability assessment, we conduct each experiment k times on a several datasets with diverse characteristics regarding size, density, and homophily distribution.

7 Experiments and Results

We evaluate the performance of our enhanced topological heuristics on link prediction tasks. To ensure fair comparison, we include our analysis within the evaluation framework introduced in Section 6. We show that the proposed approach outperforms both simple topological heuristics and GNNs on most datasets, with improved generalisation. Furthermore, we emphasise the performance gains compared to simple topological heuristics and highlight the computational efficiency relative to both foundational and sophisticated GNNs. We detail information about experimental settings and hyperparameters, and present extended performance results for all datasets and evaluation metrics in Appendices D and E in the supplementary material [6]. Source code is made available [6].

7.1 Performance and generalisation ability

We report the average ranking obtained across all datasets and for each of the evaluated performance metric in Table 1. Our experiments indicate that the proposed enhanced non-neural topological heuristics outperform state-of-the-art GNNs as well as traditional topological heuristics, regardless of the evaluation metric. In particular, the enhanced variations of RA and AA indexes systematically rank first or second, ahead of neural approaches. These results highlight the

¹ *Pubmed** denotes the dataset built from sources (more details in Appendix A in the supplementary material [6]).

Table 1. Comparison of the average ranking across all datasets between neural and non-neural models for the link prediction task. We report results for each of the evaluated metric. **Best** and **second best** ranks are highlighted.

Model	AUC	AP	MRR	Hit@20	Hit@50	Hit@100	Hit@200
CN	8.27	8.92	6.04	6.55	7.27	6.82	7.27
AA	6.96	7.46	4.42	4.77	5.91	6.18	6.00
RA	6.81	7.15	4.69	4.14	4.91	5.09	5.45
GCN	6.08	6.77	9.46	9.27	8.05	7.45	7.36
GRAPHSAGE	10.46	10.69	11.81	11.50	10.68	10.82	10.64
GAT	8.62	9.15	11.46	11.09	10.91	10.36	9.73
GAE	6.19	6.15	8.19	7.64	6.91	7.00	6.91
VGAE	5.23	4.77	7.15	6.64	6.09	5.82	5.73
SEAL	6.00	6.08	6.85	6.09	5.91	5.82	6.18
NEO-GNN	11.15	10.85	10.46	11.50	11.50	11.36	11.36
ECN	5.85	5.31	4.31	5.27	5.45	5.95	5.68
EAA	4.59	3.73	3.35	3.73	4.05	4.77	4.59
ERA	5.12	4.19	3.04	3.09	3.64	3.82	4.36

Table 2. Average Precision (with standard deviation) for the link prediction task. The last column indicates the average ranking over all datasets. For each metric, **best** and **second best** scores are highlighted. – is used when a model returns an out-of-memory error or runs for longer than 24 hours.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	71.86 ± 0.70	64.25 ± 0.24	66.46 ± 0.93	59.91 ± 0.21	89.72 ± 0.18	96.40 ± 0.01	57.77 ± 2.27	58.50 ± 2.04
AA	72.24 ± 0.77	64.31 ± 0.23	66.55 ± 0.85	60.01 ± 0.17	89.81 ± 0.17	96.96 ± 0.01	57.94 ± 2.09	58.50 ± 2.04
RA	72.23 ± 0.73	64.31 ± 0.20	66.55 ± 0.85	59.96 ± 0.18	89.80 ± 0.16	97.24 ± 0.02	58.32 ± 1.80	58.50 ± 2.04
GCN	93.01 ± 0.99	92.77 ± 0.76	92.30 ± 1.07	79.29 ± 0.57	93.55 ± 0.20	61.62 ± 20.13	65.50 ± 8.54	62.39 ± 2.85
GRAPHSAGE	53.35 ± 5.85	91.91 ± 0.29	69.97 ± 2.41	81.46 ± 1.19	60.34 ± 4.48	50.00 ± 0.00	60.00 ± 7.32	49.77 ± 2.66
GAT	94.22 ± 0.45	80.24 ± 3.52	93.18 ± 0.55	70.42 ± 2.24	84.18 ± 4.29	50.06 ± 0.04	65.97 ± 8.49	56.21 ± 4.76
GAE	98.38 ± 0.22	97.71 ± 0.13	98.20 ± 0.40	83.87 ± 1.03	96.79 ± 0.06	50.00 ± 0.00	67.42 ± 6.00	62.57 ± 8.58
VGAE	98.89 ± 0.25	95.99 ± 0.27	98.93 ± 0.30	84.08 ± 0.21	98.05 ± 0.05	91.04 ± 4.16	67.59 ± 6.65	67.48 ± 2.44
SEAL	97.66 ± 0.36	98.06 ± 0.20	94.44 ± 1.04	83.95 ± 0.34	94.91 ± 0.03	97.84 ± 0.01	63.87 ± 2.21	66.08 ± 0.87
NEO-GNN	86.03 ± 4.81	70.22 ± 26.83	58.66 ± 21.96	58.25 ± 14.80	49.90 ± 21.55	54.51 ± 5.84	76.28 ± 3.91	57.73 ± 17.55
ECN	89.96 ± 0.33	92.31 ± 0.20	93.35 ± 0.41	65.61 ± 0.53	99.04 ± 0.00	95.66 ± 0.09	75.91 ± 6.40	75.75 ± 4.39
EAA	90.05 ± 0.36	92.37 ± 0.21	93.39 ± 0.37	65.69 ± 0.52	99.26 ± 0.03	96.36 ± 0.07	76.30 ± 6.78	75.75 ± 4.39
ERA	89.72 ± 0.33	92.23 ± 0.20	93.37 ± 0.33	65.64 ± 0.52	99.24 ± 0.02	95.98 ± 0.05	76.30 ± 6.78	75.75 ± 4.39

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	95.24 ± 0.06	94.22 ± 0.07	89.56 ± 0.12	96.70 ± 0.02	75.26 ± 0.02	8.92
AA	95.93 ± 0.06	95.05 ± 0.08	90.70 ± 0.10	96.99 ± 0.02	75.27 ± 0.02	7.46
RA	96.22 ± 0.05	95.30 ± 0.10	90.75 ± 0.12	97.07 ± 0.02	75.26 ± 0.02	7.15
GCN	80.93 ± 3.49	82.00 ± 2.70	87.22 ± 2.48	76.81 ± 23.26	94.11 ± 0.50	6.77
GRAPHSAGE	50.00 ± 0.00	50.00 ± 0.00	51.56 ± 1.34	–	89.48 ± 0.61	10.69
GAT	51.57 ± 2.66	53.95 ± 6.30	54.95 ± 5.25	56.49 ± 9.18	82.04 ± 1.41	9.15
GAE	50.00 ± 0.00	66.01 ± 18.52	88.61 ± 0.50	64.37 ± 24.88	93.73 ± 0.83	6.15
VGAE	75.19 ± 21.82	88.23 ± 1.87	76.17 ± 12.05	94.70 ± 0.51	95.07 ± 0.07	4.77
SEAL	91.26 ± 6.59	91.08 ± 5.53	84.82 ± 4.65	–	–	6.08
NEO-GNN	57.65 ± 4.39	63.27 ± 1.82	51.06 ± 12.49	–	63.08 ± 14.41	10.85
ECN	95.23 ± 0.04	94.69 ± 0.04	93.59 ± 0.16	95.65 ± 0.01	90.31 ± 0.02	5.31
EAA	95.63 ± 0.04	95.23 ± 0.04	94.10 ± 0.14	96.26 ± 0.01	90.31 ± 0.02	3.73
ERA	96.13 ± 0.04	95.66 ± 0.06	93.92 ± 0.11	96.53 ± 0.01	90.30 ± 0.02	4.19

ability for our approach to generalise to several different networks exhibiting various characteristics. Additionally, our results highlight the impact of the evaluation procedure on performance; Findings indicate that traditional topological heuristics are outperformed by neural approaches when using AUC and AP metrics, but systematically outperform several recent GNNs in terms of ranking-based metrics.

To better understand these results, we present the detailed performance for all models across all datasets for the AP metric in Table 2. These results indicate strong performance of recent GNNs on

a limited number of well-known homophilous datasets, among them Cora, Pubmed and Citeseer. This aligns with previous works’ observations [48, 24, 29]; the GNNs’ message passing scheme is particularly efficient on homophilous networks. However, considering heterophilous networks, such as Cornell or Wisconsin, or Wikipedia-based networks which are neither homophilous or heterophilous, the proposed approach consistently outperform GNNs.

VGAE and SEAL are our strongest competitors and achieve the highest performance among the neural approaches. Surprisingly, we

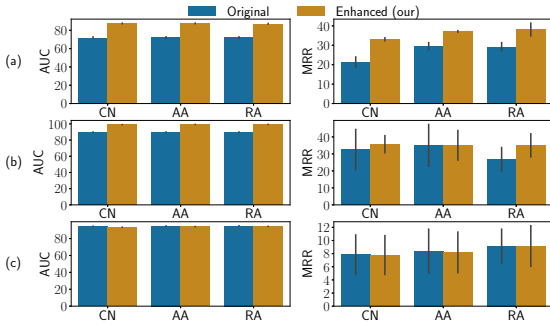


Figure 4. Benefits of weighting topological heuristics with feature information for Cora (a), CS (b) and Wikivitals-fr (c).

find that Neo-GNN is highly unstable and fails to generalise well. Nevertheless, SEAL suffers from high computational cost, rendering it impractical within a reasonable timeframe and resource allocation for networks such as Wikivitals+ and ogbn-arxiv. VGAE is more scalable, but is less generalisable and suffers from poor results on the Wikipedia-based networks.

We show that the proposed methods address both these limitations and achieve robust performance with high scalability and generalisability. Moreover, we emphasise the increase in performance of our methods compared to simple topological heuristics, highlighting the significance of both structural and feature expressivity. Similar conclusions can be drawn regarding the other evaluation metrics (see Appendix E in the supplementary material [6]).

7.2 Performance improvement vs. simple topological heuristics

We evaluate the benefits of incorporating node feature information into topological heuristics. We observed in Figure 2 how different networks exhibit different relations between their topology and their pairwise node feature similarity. Considering the proposed weighting mechanism introduced in Equation 6, we expect our approach to outperform simple topological heuristics for networks in which the distance between node feature similarity distributions among connected and disconnected nodes is relatively large.

To illustrate this, we select three real-world networks and exhibit performance gain of our approach in Figure 4. We show that incorporating explicit feature information into topological heuristics significantly increases performance for two datasets which display large disparity in pairwise feature similarity distributions among positive and negative edges, Cora and CS. This suggests that the proposed mechanism efficiently leverage information from both the structure of the graph and the node attributes. Conversely, in Wikivitals-fr, where pairwise node feature similarity weakly correlates with pairwise connectivity, our approach adaptively balances topological and feature information to maintain competitive performance. We provide additional results for other datasets in Appendix E in the supplementary material [6].

7.3 Scalability

We quantify the computational benefit of the proposed approach by comparing both its average running time and its performance against GNNs. We display in Figure 5 the results obtained for the AP metric on the Wikivitals-fr dataset. This dataset represents the largest network on which all models could be trained within a 24-hour time constraint.

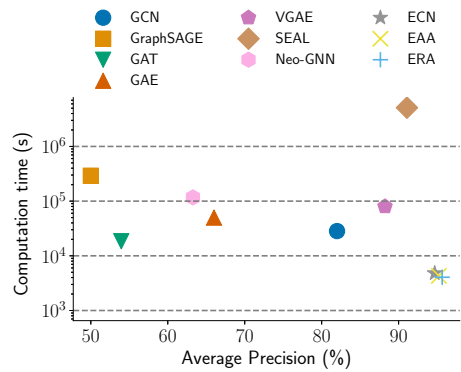


Figure 5. Average Precision and computation time on Wikivitals-fr. Notice the log scale for the y-axis. The proposed approaches do not trade-off performance for computation time, and achieve highest results within a difference of 3 orders of magnitude compared to the best GNN.

We show that SEAL method incurs the highest computational cost. In particular, extracting a number of enclosing subgraphs equals to the number of edges in the graph requires a significant amount of time for datasets of considerable size. Conversely, the proposed enhanced topological heuristics achieve superior performance with a computing time reduction of three orders of magnitude. Additionally, we observe a difference in computation time of one to two orders of magnitude with models such as GAE, VGAE and Neo-GNN, which often position as lower bound reference in attempts to reduce GNN computational cost [44]. These results highlight the capability of simple yet strong non-neural models to consistently outperform GNNs on large datasets, while requiring significantly fewer resources.

In summary, our approach maintains the topological expressivity of simple heuristics while adaptively incorporating node attribute when relevant. Our weighting mechanism helps improving performance as well as generalisation ability. Unlike GNNs, the absence of a learning process makes our approach computationally efficient and highly scalable.

8 Conclusions and Future Work

In this work, we introduced enhanced non-neural heuristics combining topological information with node-level and graph-level features. Our proposed approach adaptively leverages either the structural knowledge from the direct neighbourhood of the predicted edge or their node attribute information. Within a robust evaluation framework, we have showed that our enhanced heuristics achieve state-of-the-art performance against recent GNNs, harnessing their capabilities without relying on any learning mechanism. Experiments on several real-world networks with various characteristics have indicated that adaptively combining topological information with node feature information is a valuable mechanism that allows better generalisation ability than sophisticated GNNs. Additionally, we have highlighted the scalability of our approach, achieving the highest results with up to three-order-of-magnitude difference in computation time compared to the best GNNs.

In future work, we plan to further develop the proposed approach to generalise to various downstream machine learning tasks on graphs, such as node classification and graph classification.

References

- [1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, 2003.
- [2] Aleksandar and S. Günnemann. Pitfalls of Graph Neural Network Evaluation. *arXiv preprint arXiv:1811.05868*, 2018.
- [3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.
- [4] L. Cai, J. Li, J. Wang, and S. Ji. Line graph neural networks for link prediction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(9):5103–5113, 2022.
- [5] B. P. Chamberlain, S. Shirobokov, E. Rossi, F. Frasca, T. Markovich, N. Y. Hammerla, M. M. Bronstein, and M. Hansmire. Graph neural networks for link prediction with subgraph sketching. In *The Eleventh International Conference on Learning Representations*, 2023.
- [6] S. Delarue, T. Bonald, and T. Viard. Code, data, and supplementary material for “Link prediction without learning”, 2024. URL <https://github.com/simondelarue/link-prediction-without-learning>.
- [7] Y. Dong, Q. Ke, J. Rao, and B. Wu. Predicting missing links via local feature of common neighbors. In *Eighth International Conference on Fuzzy Systems and Knowledge Discovery*, pages 1038–1042, 2011.
- [8] V. P. Dwivedi, C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24:43:1–43:48, 2023.
- [9] M. Ferrari Dacrema, P. Cremonesi, and D. Jannach. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 101–109, 2019.
- [10] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD Conference*, pages 855–864, 2016.
- [11] W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems 30*, pages 1024–1034, 2017.
- [12] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, pages 22118–22133, 2020.
- [13] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
- [14] T. N. Kipf and M. Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- [15] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [16] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
- [17] W. Leeney and R. McConville. Uncertainty in gnn learning evaluations: The importance of a consistent benchmark for community detection. In *The International Conference on Complex Networks and their Applications*, 2023.
- [18] J. Li, H. Shomer, H. Mao, S. Zeng, Y. Ma, N. Shah, J. Tang, and D. Yin. Evaluating graph neural networks for link prediction: Current pitfalls and new benchmarking. In *Advances in Neural Information Processing Systems*, 2023.
- [19] Z. C. Lipton and J. Steinhardt. Troubling trends in machine learning scholarship: Some ml papers suffer from flaws that could mislead the public and stymie future research. *Queue*, 17(1):45–77, 2019.
- [20] L. Lü and T. Zhou. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6):1150–1170, 2011.
- [21] M. Ludewig, N. Mauro, S. Latifi, and D. Jannach. Performance comparison of neural and non-neural approaches to session-based recommendation. In *Proceedings of the 13th ACM conference on recommender systems*, pages 462–466, 2019.
- [22] Q. Lv, M. Ding, Q. Liu, Y. Chen, W. Feng, S. He, C. Zhou, J. Jiang, Y. Dong, and J. Tang. Are we really making much progress?: Revisiting, benchmarking and refining heterogeneous graph neural networks. In *Proceedings of the 27th ACM SIGKDD Conference*, pages 1150–1160, 2021.
- [23] L. Ma, H. Han, J. Li, H. Shomer, H. Liu, X. Gao, and J. Tang. Mixture of link predictors. *arXiv preprint arXiv:2402.08583*, 2024.
- [24] S. Maekawa, K. Noda, Y. Sasaki, et al. Beyond real-world benchmark datasets: An empirical study of node classification with gnns. *Advances in Neural Information Processing Systems*, pages 5562–5574, 2022.
- [25] H. Mao, Z. Chen, W. Jin, H. Han, Y. Ma, T. Zhao, N. Shah, and J. Tang. Demystifying structural disparity in graph neural networks: Can one size fit all? In *Advances in Neural Information Processing Systems*, 2023.
- [26] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference*, pages 43–52, 2015.
- [27] P. Mernyei and C. Cangea. Wiki-cs: A wikipedia-based benchmark for graph neural networks. *arXiv preprint arXiv:2007.02901*, 2020.
- [28] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64:025102, Jul 2001.
- [29] J. Palowitch, A. Tsitsulin, B. Mayer, and B. Perozzi. GraphWorld: Fake Graphs Bring Real Insights for GNNs. In *Proceedings of the 28th ACM SIGKDD Conference*, pages 3691–3701, 2022.
- [30] H. Pei, B. Wei, K. C. Chang, Y. Lei, and B. Yang. Geom-gcn: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020.
- [31] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD Conference*, KDD ’14, pages 701–710, 2014.
- [32] J. Pineau, P. Vincent-Lamarre, K. Sinha, V. Larivière, A. Beygelzimer, F. d’Alché Buc, E. Fox, and H. Larochelle. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *Journal of Machine Learning Research*, 22:164:1–164:20, 2021.
- [33] F. Poursafaei, S. Huang, K. Pelrine, and R. Rabbany. Towards better evaluation for dynamic link prediction. In *Advances in Neural Information Processing Systems*, 2022.
- [34] D. Sculley, J. Snoek, A. B. Wiltschko, and A. Rahimi. Winner’s curse? on pace, progress, and empirical rigor. In *International Conference on Learning Representations, Workshop Track Proceedings*, 2018.
- [35] Z. Sun, Z. Deng, J. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations*, 2019.
- [36] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: large-scale information network embedding. In *Proceedings of The ACM Web Conference 2015*, pages 1067–1077, 2015.
- [37] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.
- [38] X. Wang, H. Yang, and M. Zhang. Neural common neighbor with completion for link prediction. *arXiv preprint arXiv:2302.00890*, 2023.
- [39] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.
- [40] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka. Representation learning on graphs with jumping knowledge networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5449–5458. PMLR, 2018.
- [41] W. Yang, K. Lu, P. Yang, and J. Lin. Critically examining the “neural hype”: Weak baselines and the additivity of effectiveness gains from neural ranking models. In *Proceedings of the International ACM SIGIR Conference*, pages 1129–1132, 2019.
- [42] Z. Yang, W. Cohen, and R. Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning*, volume 48, pages 40–48. JMLR.org, 2016.
- [43] J. You, Z. Ying, and J. Leskovec. Design space for graph neural networks. *Advances in Neural Information Processing Systems 33*, 33: 17009–17021, 2020.
- [44] S. Yun, S. Kim, J. Lee, J. Kang, and H. J. Kim. Neo-gnns: Neighborhood overlap-aware graph neural networks for link prediction. In *Advances in Neural Information Processing Systems 34*, pages 13683–13694, 2021.
- [45] M. Zhang and Y. Chen. Link prediction based on graph neural networks. In *Advances in Neural Information Processing Systems 31*, pages 5171–5181, 2018.
- [46] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.
- [47] Zhou, Tao, Lü, Linyuan, and Zhang, Yi-Cheng. Predicting missing links via local information. *Eur. Phys. J. B*, 71(4):623–630, 2009.
- [48] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in Neural Information Processing Systems*, 33:7793–7804, 2020.
- [49] Z. Zhu, Z. Zhang, L. A. C. Xhonneux, and J. Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *Advances in Neural Information Processing Systems*, pages 29476–29490, 2021.

A Dataset Characteristics

A.1 Dataset statistics

We report dataset statistics in Table 3. The density each network is denoted with δ_A . The label homophily of a node u , denoted $h_u = \frac{1}{d_u} \sum_{v \in \mathcal{N}(u)} \mathbb{1}(y_u = y_v)$, corresponds to the proportion of its neighbourhood sharing its label. We denote with \mathcal{H} the average label homophily across all nodes.

We show in Figure 6 the node label homophily distributions for all the considered networks. *Cora*, *Pubmed*, *Citeseer*, *CS*, *Photo* and *Ogbn-arxiv* are highly homophilous networks. On the contrary, *Actor*, *Cornell* and *Wisconsin* are considered heterophilous.

Despite their differences, the aforementioned networks are highly specific. In this work, we also consider real-world networks with smoother label homophily distributions, namely *Wikivitals*, *Wikivitals-fr*, *Wikischools* and *Wikivitals+*.

A.2 Dataset details

Cora, *Pubmed* and *Citeseer* [42] are citation networks, where nodes represent articles and edges represent citation links. We notice differences in the feature matrix between pre-computed online versions of the *Pubmed* graph and the graph we have built from sources, denoted with *Pubmed**. It appears that these differences come from the ordering of the rows of the feature matrix. In this paper, we rely on the *Pubmed** version of the graph, provided in the repository of this project².

Actor dataset is the actor-induced subgraph [30]. In this graph, each node corresponds to an actor and edges are connecting actors whose names co-occur on the same Wikipedia page. The node features are generated from the bag-of-words representation of keywords in these Web pages.

The *Photo* [26] dataset is an Amazon co-purchase network, where a node represents a good and an edge denotes a frequent co-purchase between these items on the platform. Features are constructed from the bag-of-words extracted from product reviews.

The *CS* [2] dataset is a co-authorship graph originating from the KDD Cup 2016 Challenge. Nodes represent authors, and edges denote co-authorship relations between these authors. The features are bag-of-words of the scientific paper keywords.

Cornell and *Wisconsin*³ are universities web pages graphs, where each page is manually classified into a category (for example, student, faculty, etc). Features correspond to bag-of-words from page texts, after removing words with highest Mutual Information with the category variable.

We consider four Wikipedia-based real-world networks⁴. The *Wikivitals* and *Wikivitals+* datasets focus on Wikipedia’s so-called “vital articles”, a community-made selection of Wikipedia pages. They are extracted from respectively levels 4 and 5 from WikiData. *Wikivitals-fr* contains the vital articles written in French, and *Wikischools* contains articles related to material taught in schools. For all these datasets, an edge exist between two articles if they are referencing each other in Wikipedia, and node features correspond to the bag-of-words representations of the articles.

² Remote repository: <https://github.com/simondelaurie/link-prediction-without-learning>.

³ Data related to the World Wide Knowledge Base Project: <https://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wwkb/>

⁴ Wikipedia-based networks: <https://netset.telecom-paris.fr/>.

Finally, *Ogbn-arxiv* [12] is a citation network between Computer Science papers. Each paper comes with a 128-dimensional vector of features built according to the embedding of the words contained in its abstract and title.

B Literature review on GNN for link prediction

We evaluate the extent to which simple heuristics are overviewed in the GNN benchmarks for the link prediction task. Detailed information about all the reviewed literature articles and baseline counting is available in the supplementary material section of the remote repository of the project.

We consider the following four categories of baselines. For each reviewed article, we record the existence of these categories of baselines in the benchmark.

GNN: Includes all GNN-based model (GCN, GAE, SEAL, etc.).

Embedding: Includes all latent feature-based models that do not rely on a message passing framework, such as matrix factorization or random walk methods (DeepWalk, node2vec, TransE, etc.).

Topological heuristic: Includes all indexes relying solely on graph structure (CN, RA, AA, etc.).

Enhanced topological heuristics: Includes all topological heuristics baselines that may have been modified to become valid competitors to GNNs.

We extract GNN-based articles from the Scopus database using the following procedure:

1. We extract all articles mentioning Graph Neural Network and link prediction, but not corresponding to a survey. We restrict our search for articles with a publication year greater than 2016. This year corresponds to the publication of the GCN model [15], which is often considered as a stepping stone in the GNN field evolution. For this purpose, we use the following query `'TITLE-ABS-KEY ("graph neural network" AND "link prediction" AND NOT "survey") AND PUBYEAR > 2016'`, which returns 880 documents.
2. From this corpus, we only consider the top-10% most cited articles that do not contradict our selection criteria; some articles may be discarded if the proposed approach is solely evaluated on a graph classification task or if it does not contain a novel GNN proposal (for instance when the article is actually a book chapter or a tutorial). We end up with 88 articles.
3. For each of the remaining 88 articles, we record which categories of baselines are used in the proposed benchmark.

C Models

To demonstrate the effectiveness of the proposed approach, we compare its performance to three topological heuristics and six GNN models.

Topological heuristics. For topological heuristics, we use Common Neighbours (CN) [28], Adamic-Adar index (AA) [1] and Resource Allocation index (RA) [47]. These approaches do not involve any learning process as they only rely on the topological information of the graph.

Graph Neural Networks. For GNNs, we use GCN [15], GraphSage [11], GAT [37], GAE and VGAE [14]. All these models learn node representations and combine pairwise embeddings through a similarity function to predict the existence of a link. Additionally, we include GNNs specifically boosted to increase structural expressiveness. Drawing inspiration from topological heuristics, SEAL [45]

Table 3. Dataset statistics

Dataset	#nodes	#edges	#features	#labels	δ_A	\mathcal{H}
Cora	2708	10556	1433	7	2.88×10^{-3}	0.825
Pubmed*	19717	88651	500	3	4.56×10^{-4}	0.792
Citeseer	3327	9104	3703	6	1.65×10^{-3}	0.717
Actor	7600	30019	932	5	1.04×10^{-3}	0.224
CS	18333	163788	6805	15	9.75×10^{-4}	0.832
Photo	7650	238162	745	8	8.14×10^{-3}	0.849
Cornell	183	298	1703	5	1.79×10^{-2}	0.203
Wisconsin	251	515	1703	5	1.64×10^{-2}	0.198
Wikivitals	10011	824999	37845	11	8.23×10^{-3}	0.472
Wikivitals-fr	9945	558427	28198	11	5.65×10^{-3}	0.447
Wikischools	4403	112834	20527	16	5.82×10^{-3}	0.411
Wikivitals+	45149	3946850	85512	11	1.93×10^{-3}	0.487
Ogbn-arxiv	169343	1166246	128	40	8.14×10^{-5}	0.632

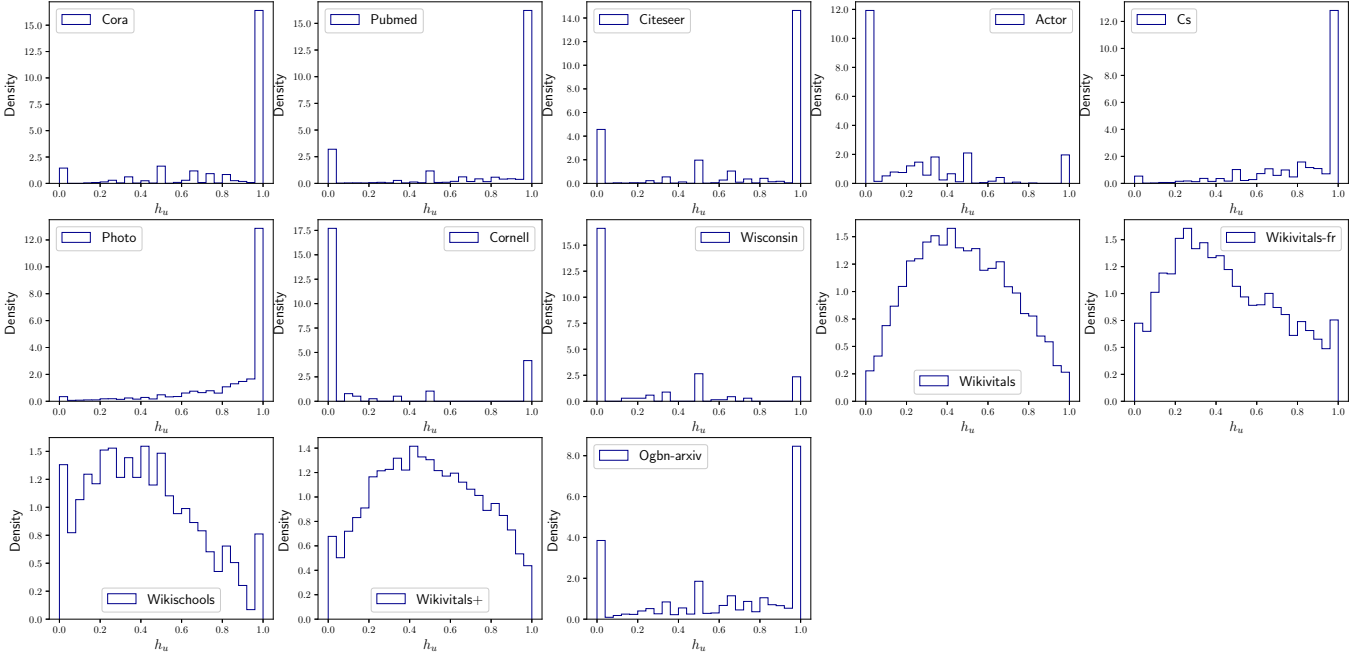


Figure 6. Node label homophily distributions.

and Neo-GNN [44] are designed to incorporate local topological information near the predicted edge. We provide implementation details of each model in the following section.

D Hyperparameters and settings

We detail GNN-based hyperparameters in Table 4. For all models we used the Binary Cross Entropy Loss and Adam optimizer. The following model’s hyperparameters were tuned using grid search, GCN, GRAPH-SAGE, GAT, GAE, VGAE and NEO-GNN. The search space was over hidden and output dimension [16, 32, 64, 128] for all neural models. Additionally, the search space for the number of heads in GAT was over [4, 8]. The search space for the neural link predictor in NEO-GNN was over [16, 32, 64, 128]. All models were tuned on Cora, and the hyperparameters that achieved the best results on validation AUC were used for all other datasets. For SEAL, the best hyperparameters found for Cora induced an intractable learning process for larger datasets. Consequently, we used

the hyperparameters found in the original paper. For all datasets, the number of edges included in training-validation-test sets are 85%-5%-10%. The number of epochs, learning rate and optimizer corresponds to the ones used in the original model implementations.

The original implementation of SEAL uses a 2-hops subgraph extraction in the vicinity of the predicted edge. This setting led to an out-of-memory error for CS, Photo, Wikivitals, Wikivitals-fr, Wikischools, Wikivitals+ and Ogbn-arxiv. Therefore, for these datasets, we reduced the dimension of the extracted subgraph to the 1-hop neighbourhood. Moreover, as proposed in its original implementation [45], we restrict the SEAL approach not to include any explicit feature and rely only on the graph structure features learned during the training.

It is useful to note that both GAE and VGAE models perform negative sampling **at each epoch** during the training procedure, which deviates from our choice for a systematic comparison on the same train-validation and test splits across models. In our experiments, we

Table 4. Final hyperparameters for GNNs in link prediction.

Model	Encoder	Decoder	#epochs	learning rate	Optimizer
GCN	2-layers GCN(64, 64)	Inner product	200	1×10^{-2}	Adam
GRAPHSAGE	2-layers GraphSage(128, 128)	Inner product	100	1×10^{-2}	Adam
GAT	2-layers GAT(16, 64) (heads=4)	Inner product	100	5×10^{-2}	Adam
GAE	2-layers GCN(128, 64)	Inner product	200	1×10^{-2}	Adam
VGAE	2-layers GCN(128, 128)	Inner product	200	1×10^{-2}	Adam
SEAL	2-layers DGCNN(128, 128) + MLP(32)	-	50	1×10^{-4}	Adam
NEO-GNN	3-layers GCN(128, 64) + 3-layers Linear(64, 32)	-	200	1×10^{-3}	Adam

maintained this setting to keep the implementation as close as possible to the original one. However, we noticed that relying on a fixed set of negative edges shared for all epochs drastically lower the overall performance.

Rankings. For all experiments, in cases where settings ran out of memory during training process, we assigned the corresponding models the same rank, which corresponds to the lowest possible rank. We make this decision to avoid introducing bias in favor of computationally expensive methods, aligning with our belief that scalability considerations should be weighted equally with performance in our evaluation.

Considering the small number of edges available in Cornell and Wisconsin, we did not include results from Hit@ k with $k > 20$ in the rankings.

E Detailed results

We provide average results and standard deviations for each of the evaluated performance metric on the link prediction task: AUC (Table 5), Mean Reciprocal Rank (Table 6), Hit@20 (Table 7), Hit@50 (Table 8), Hit@100 (Table 9) and Hit@200 (Table 10). For each metric, **best** and **second best** scores are highlighted. - is used when a model returns an out-of-memory error or runs for longer than 24 hours. For each evaluation metric, the average rank across all datasets is reported in the last column.

We ran all models on Intel(R) Xeon(R) CPU E5-2660 v3 @ 2.60GHz with 362GB memory. For the training of GNN-based models, we chose not to use any GPU. This decision was motivated by our aim to maintain simplicity in the models, ensuring compatibility with various computing environments.

On Wikivitals+ and Ogbn-*arxiv*, SEAL method required longer than 24 hours only for the extraction of the enclosing subgraphs.

E.1 Performance improvement vs. simple topological heuristics

We illustrate the benefits of incorporating node-level and graph-level feature information into topological heuristics in Figure 7.

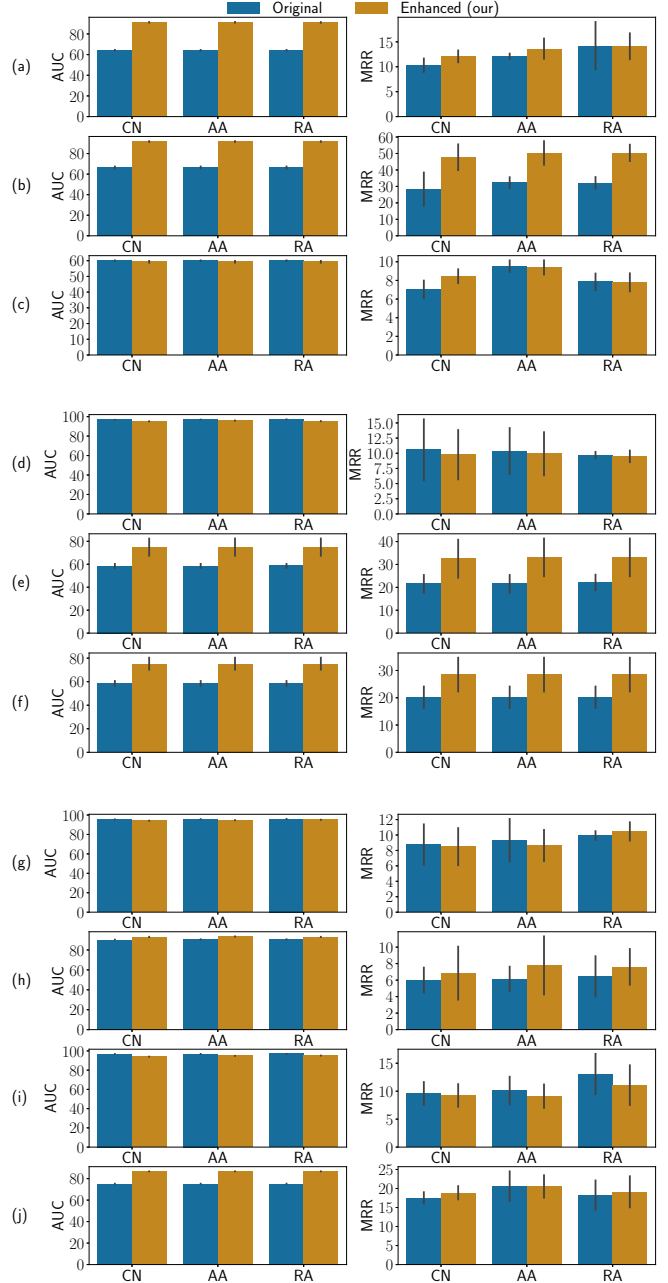


Figure 7. Benefits of weighting topological heuristics with feature information for Pubmed* (a), Citeseer (b), Actor (c), Photo (d), Cornell (e), Wisconsin (f), Wikivitals (g), Wikischools (h), Wikivitals+ (i), and Ogbn-*arxiv* (j).

Table 5. AUC for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	72.06 ±0.71	64.29 ±0.24	66.49 ±0.89	60.00 ±0.18	89.77 ±0.16	96.75 ±0.00	58.54 ±1.73	58.50 ±2.04
AA	72.13 ±0.73	64.29 ±0.23	66.50 ±0.88	60.00 ±0.18	89.80 ±0.16	97.02 ±0.00	58.56 ±1.73	58.50 ±2.04
RA	72.13 ±0.72	64.29 ±0.23	66.50 ±0.88	60.00 ±0.19	89.79 ±0.16	97.24 ±0.01	58.60 ±1.72	58.50 ±2.04
GCN	93.71 ±0.78	92.75 ±0.95	93.46 ±0.63	76.97 ±0.20	94.15 ±0.25	61.98 ±20.75	64.37 ±11.08	63.24 ±3.46
GRAPHSAGE	54.75 ±8.31	90.95 ±0.17	74.25 ±2.59	78.53 ±1.65	63.33 ±4.52	50.00 ±0.00	56.30 ±10.25	48.90 ±5.13
GAT	94.56 ±0.53	81.56 ±3.78	93.35 ±0.43	70.77 ±1.14	87.95 ±3.46	50.11 ±0.07	64.35 ±5.49	57.14 ±2.70
GAE	98.42 ±0.33	97.90 ±0.12	97.96 ±0.26	80.70 ±0.81	97.18 ±0.04	50.00 ±0.00	64.05 ±1.63	58.50 ±9.07
VGAE	98.80 ±0.34	96.25 ±0.23	98.71 ±0.21	80.86 ±0.32	98.32 ±0.04	91.83 ±3.76	63.85 ±6.38	61.86 ±4.08
SEAL	96.75 ±0.39	97.61 ±0.28	93.66 ±1.32	81.09 ±0.33	93.54 ±0.03	97.57 ±0.02	58.68 ±1.71	68.47 ±0.98
NEO-GNN	80.60 ±4.92	66.75 ±31.75	51.28 ±24.87	60.38 ±24.60	34.65 ±36.76	42.30 ±4.27	70.81 ±9.69	48.79 ±20.08
ECN	87.50 ±0.28	91.24 ±0.29	91.67 ±0.37	59.27 ±0.58	99.03 ±0.00	95.19 ±0.10	74.87 ±7.45	75.28 ±5.06
EAA	87.54 ±0.29	91.26 ±0.29	91.69 ±0.35	59.29 ±0.58	99.24 ±0.02	95.93 ±0.08	74.91 ±7.50	75.28 ±5.06
ERA	87.20 ±0.22	91.16 ±0.29	91.67 ±0.32	59.28 ±0.58	99.22 ±0.02	95.37 ±0.06	74.91 ±7.50	75.28 ±5.06

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	95.47 ±0.05	94.59 ±0.08	90.28 ±0.12	96.77 ±0.02	75.26 ±0.02	8.27
AA	95.72 ±0.05	94.89 ±0.09	90.62 ±0.11	96.87 ±0.02	75.26 ±0.02	6.96
RA	95.92 ±0.05	95.04 ±0.10	90.63 ±0.11	96.92 ±0.02	75.26 ±0.02	6.81
GCN	79.83 ±4.26	80.19 ±3.59	86.36 ±2.53	76.46 ±22.95	93.64 ±0.61	6.08
GRAPHSAGE	50.00 ±0.00	50.00 ±0.00	51.76 ±1.50	–	88.47 ±0.78	10.46
GAT	53.67 ±6.25	56.56 ±8.54	58.35 ±7.95	62.23 ±9.49	86.03 ±0.76	8.62
GAE	50.00 ±0.00	68.07 ±17.74	87.44 ±0.65	64.04 ±24.32	93.17 ±1.01	6.19
VGAE	74.92 ±21.59	87.88 ±1.51	80.54 ±8.16	94.08 ±0.51	94.74 ±0.09	5.23
SEAL	91.63 ±3.91	93.21 ±0.09	90.04 ±0.03	–	–	6.00
NEO-GNN	49.49 ±5.92	52.51 ±1.40	36.33 ±17.87	–	53.19 ±10.13	11.15
ECN	94.16 ±0.04	93.40 ±0.07	93.03 ±0.19	94.10 ±0.01	86.74 ±0.03	5.85
EAA	94.69 ±0.04	94.11 ±0.08	93.47 ±0.15	94.90 ±0.01	86.74 ±0.03	4.50
ERA	95.12 ±0.04	94.41 ±0.08	93.11 ±0.10	95.18 ±0.01	86.72 ±0.04	5.12

Table 6. Mean Reciprocal Rank for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	21.33 ±2.64	10.32 ±1.33	28.35 ±10.07	7.04 ±0.94	32.47 ±11.91	10.58 ±5.02	21.53 ±3.86	20.19 ±3.92
AA	29.51 ±1.81	12.13 ±0.52	32.28 ±3.31	9.53 ±0.61	35.05 ±12.30	10.38 ±3.77	21.53 ±3.86	20.19 ±3.92
RA	29.29 ±2.04	14.25 ±4.74	32.15 ±3.53	7.84 ±0.90	26.68 ±7.06	9.69 ±0.54	22.11 ±3.41	20.19 ±3.92
GCN	13.36 ±4.45	6.47 ±0.07	8.37 ±2.54	4.35 ±0.69	3.76 ±0.85	1.26 ±2.17	23.59 ±3.37	13.64 ±0.56
GRAPHSAGE	0.62 ±0.75	5.53 ±1.21	3.88 ±1.19	4.97 ±0.85	0.61 ±0.61	0.01 ±0.00	19.51 ±7.23	5.72 ±2.00
GAT	15.95 ±3.62	1.32 ±0.40	21.57 ±3.22	1.49 ±1.04	0.28 ±0.14	0.01 ±0.00	22.84 ±11.64	9.86 ±3.55
GAE	38.59 ±7.39	9.22 ±0.98	42.51 ±13.42	6.17 ±0.28	5.97 ±1.23	0.01 ±0.00	31.04 ±15.36	16.65 ±5.86
VGAE	47.42 ±11.34	7.58 ±1.69	52.26 ±14.22	5.55 ±0.94	10.07 ±4.50	3.21 ±2.01	24.22 ±3.78	19.26 ±0.31
SEAL	74.76 ±11.23	66.33 ±8.91	66.89 ±16.23	6.14 ±0.06	25.80 ±3.32	13.10 ±3.15	19.14 ±2.78	15.08 ±1.42
NEO-GNN	17.48 ±5.22	2.80 ±2.46	2.83 ±3.14	0.53 ±0.39	2.66 ±4.51	0.05 ±0.05	44.79 ±9.08	18.32 ±12.59
ECN	33.09 ±0.76	12.10 ±1.18	47.72 ±7.81	8.45 ±0.74	35.70 ±5.03	9.77 ±4.06	32.51 ±8.28	28.49 ±6.18
EAA	37.14 ±0.39	13.64 ±2.03	50.33 ±7.14	9.39 ±0.76	35.06 ±8.64	9.92 ±3.55	33.09 ±8.24	28.49 ±6.18
ERA	38.11 ±3.28	14.12 ±2.60	50.38 ±4.97	7.80 ±0.97	35.05 ±6.77	9.49 ±0.95	33.09 ±8.24	28.49 ±6.18

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	8.77 ±2.61	7.84 ±2.99	6.01 ±1.51	9.60 ±2.03	17.54 ±1.47	6.04
AA	9.32 ±2.76	8.39 ±3.32	6.13 ±1.50	10.12 ±2.47	20.64 ±3.86	4.42
RA	9.92 ±0.58	9.16 ±2.56	6.46 ±2.43	13.09 ±3.59	18.28 ±3.83	4.69
GCN	0.46 ±0.12	0.96 ±0.18	2.96 ±0.09	0.17 ±0.15	1.83 ±0.70	9.46
GRAPHSAGE	0.00 ±0.00	0.00 ±0.00	0.12 ±0.09	–	1.14 ±0.27	11.81
GAT	0.00 ±0.00	0.01 ±0.01	0.04 ±0.03	0.00 ±0.00	0.05 ±0.02	11.46
GAE	0.00 ±0.00	0.33 ±0.57	2.82 ±0.56	0.19 ±0.34	1.83 ±0.66	8.19
VGAE	0.29 ±0.28	0.89 ±0.80	0.28 ±0.40	0.26 ±0.09	1.84 ±0.62	7.15
SEAL	2.99 ±1.21	6.30 ±1.76	3.46 ±0.29	–	–	6.85
NEO-GNN	0.01 ±0.00	0.03 ±0.01	0.10 ±0.07	–	0.04 ±0.04	10.46
ECN	8.49 ±2.40	7.78 ±2.95	6.84 ±3.21	9.24 ±2.02	18.88 ±1.74	4.31
EAA	8.63 ±2.02	8.18 ±3.09	7.78 ±3.52	9.11 ±2.09	20.53 ±2.93	3.35
ERA	10.47 ±1.18	9.15 ±3.09	7.61 ±2.16	11.08 ±3.55	19.13 ±4.08	3.04

Table 7. Hit@20 for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	44.80 ±1.54	28.75 ±0.49	33.11 ±1.67	20.30 ±0.27	53.81 ±0.39	26.18 ±0.21	100.00 ±0.00	16.99 ±4.08
AA	44.80 ±1.54	28.67 ±0.36	33.11 ±1.67	20.30 ±0.27	66.76 ±4.43	31.30 ±1.59	100.00 ±0.00	16.99 ±4.08
RA	44.80 ±1.54	28.67 ±0.36	33.11 ±1.67	20.30 ±0.27	65.45 ±5.33	32.38 ±1.56	100.00 ±0.00	16.99 ±4.08
GCN	42.37 ±4.41	16.93 ±1.25	44.40 ±5.61	14.10 ±1.72	12.42 ±1.61	3.36 ±5.83	79.31 ±9.12	60.78 ±7.84
GRAPHSAGE	1.99 ±3.45	22.85 ±2.02	9.89 ±1.92	16.65 ±2.30	1.31 ±1.38	0.00 ±0.00	71.26 ±12.11	3.27 ±4.08
GAT	50.90 ±1.42	6.08 ±1.42	47.22 ±3.61	4.60 ±2.49	0.83 ±0.74	0.00 ±0.00	80.46 ±13.94	43.79 ±7.42
GAE	84.04 ±2.50	29.72 ±0.74	90.04 ±2.04	18.77 ±2.69	18.49 ±1.86	0.00 ±0.00	77.01 ±7.18	42.48 ±9.27
VGAE	92.70 ±1.83	23.84 ±3.80	95.60 ±1.48	20.52 ±1.04	23.65 ±2.53	7.94 ±4.30	77.01 ±5.27	56.21 ±4.08
SEAL	86.73 ±4.00	82.27 ±0.95	80.11 ±2.60	18.84 ±2.07	65.30 ±0.64	29.51 ±0.37	70.11 ±8.68	66.01 ±3.00
NEO-GNN	45.24 ±7.64	8.23 ±7.28	19.49 ±30.76	0.50 ±0.44	4.13 ±6.75	0.00 ±0.00	78.16 ±18.99	39.87 ±20.03
ECN	55.80 ±1.10	28.27 ±1.62	66.74 ±1.65	20.73 ±0.43	61.92 ±1.96	28.30 ±0.82	86.21 ±6.90	77.78 ±7.92
EAA	56.40 ±0.90	31.20 ±1.15	66.74 ±1.65	20.73 ±0.43	71.55 ±2.35	27.53 ±1.14	86.21 ±6.90	77.78 ±7.92
ERA	55.04 ±1.54	29.59 ±1.27	66.41 ±1.71	20.72 ±0.43	71.40 ±2.38	29.04 ±0.73	86.21 ±6.90	77.78 ±7.92

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	18.95 ±0.57	14.87 ±0.40	21.40 ±1.74	22.95 ±0.68	43.50 ±12.20	6.54
AA	20.61 ±1.41	16.89 ±0.43	23.65 ±0.89	26.16 ±0.34	49.01 ±1.54	5.04
RA	26.60 ±1.08	26.13 ±3.52	24.89 ±3.13	34.96 ±3.72	48.61 ±2.09	4.50
GCN	1.20 ±0.14	2.93 ±0.16	9.79 ±2.48	0.54 ±0.47	9.55 ±0.84	8.85
GRAPHSAGE	0.01 ±0.02	0.00 ±0.00	0.06 ±0.05	–	3.49 ±0.78	11.65
GAT	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	0.07 ±0.06	10.46
GAE	0.00 ±0.00	1.18 ±2.04	11.11 ±0.18	0.49 ±0.84	9.77 ±0.42	7.88
VGAE	0.93 ±0.82	2.63 ±2.51	0.31 ±0.54	1.13 ±0.59	10.18 ±0.81	6.88
SEAL	7.84 ±1.47	18.78 ±2.00	12.71 ±2.80	–	–	6.46
NEO-GNN	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	–	0.02 ±0.03	11.12
ECN	18.19 ±0.27	14.12 ±0.34	22.26 ±0.23	22.61 ±1.22	47.56 ±2.33	5.00
EAA	18.12 ±0.44	15.28 ±0.23	23.89 ±0.44	23.23 ±1.29	48.54 ±1.56	3.69
ERA	26.42 ±1.33	23.65 ±3.09	26.49 ±0.36	32.30 ±2.82	48.67 ±1.53	3.15

Table 8. Hit@50 for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	44.80 ±1.54	28.75 ±0.49	33.11 ±1.67	20.30 ±0.27	62.41 ±15.16	35.86 ±1.85	100.00 ±0.00	100.00 ±0.00
AA	44.80 ±1.54	28.75 ±0.49	33.11 ±1.67	20.30 ±0.27	76.12 ±2.73	42.00 ±1.93	100.00 ±0.00	100.00 ±0.00
RA	44.80 ±1.54	28.75 ±0.49	33.11 ±1.67	20.30 ±0.27	76.39 ±2.43	49.45 ±1.12	100.00 ±0.00	100.00 ±0.00
GCN	64.61 ±3.69	28.50 ±1.69	63.70 ±5.78	21.53 ±1.91	19.05 ±0.58	4.54 ±7.87	100.00 ±0.00	99.35 ±1.13
GRAPHSAGE	3.16 ±5.47	34.71 ±1.38	17.07 ±1.43	26.91 ±2.18	2.49 ±2.38	0.00 ±0.00	100.00 ±0.00	94.77 ±5.99
GAT	71.56 ±1.28	10.04 ±1.95	68.28 ±3.37	9.69 ±3.11	2.06 ±1.78	0.00 ±0.00	100.00 ±0.00	98.69 ±1.13
GAE	94.15 ±0.90	48.37 ±1.29	94.76 ±0.73	31.79 ±3.56	29.80 ±1.48	0.00 ±0.00	100.00 ±0.00	97.39 ±4.53
VGAE	97.22 ±1.06	37.02 ±2.78	96.74 ±0.62	31.82 ±1.49	37.44 ±1.08	13.69 ±6.86	100.00 ±0.00	98.04 ±3.40
SEAL	90.52 ±1.52	83.72 ±0.81	83.81 ±1.93	30.31 ±0.57	72.21 ±2.68	42.48 ±0.60	100.00 ±0.00	100.00 ±0.00
NEO-GNN	62.09 ±8.47	13.10 ±11.35	26.70 ±38.75	1.02 ±0.19	6.01 ±9.38	0.00 ±0.00	100.00 ±0.00	84.31 ±22.10
ECN	64.61 ±1.95	38.75 ±0.53	73.88 ±1.82	21.36 ±0.61	70.17 ±1.56	38.73 ±1.32	100.00 ±0.00	100.00 ±0.00
EAA	64.58 ±1.83	40.34 ±0.48	73.88 ±1.82	21.36 ±0.61	79.09 ±1.35	40.99 ±1.41	100.00 ±0.00	100.00 ±0.00
ERA	63.19 ±1.74	38.80 ±0.34	73.81 ±1.76	21.35 ±0.62	78.64 ±1.10	45.08 ±1.77	100.00 ±0.00	100.00 ±0.00

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	25.70 ±2.57	23.27 ±0.63	33.66 ±4.02	30.55 ±0.07	50.54 ±0.05	7.27
AA	28.43 ±2.68	25.74 ±0.19	35.68 ±2.06	33.40 ±0.42	50.54 ±0.05	5.91
RA	36.83 ±1.30	34.23 ±2.10	37.22 ±1.05	42.01 ±2.50	50.54 ±0.05	4.91
GCN	2.85 ±0.78	5.26 ±0.34	16.64 ±4.08	1.15 ±1.07	16.14 ±0.52	8.05
GRAPHSAGE	0.01 ±0.02	0.00 ±0.00	0.06 ±0.05	–	6.30 ±0.68	10.68
GAT	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	0.19 ±0.14	10.91
GAE	0.00 ±0.00	2.61 ±4.52	18.84 ±1.59	1.03 ±1.78	15.74 ±0.50	6.91
VGAE	2.13 ±1.86	5.02 ±4.68	3.97 ±6.88	2.57 ±1.11	16.67 ±0.75	6.09
SEAL	15.18 ±1.86	26.35 ±1.70	25.35 ±1.35	–	–	5.91
NEO-GNN	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	–	0.26 ±0.22	11.50
ECN	24.84 ±2.33	22.80 ±0.79	32.00 ±0.60	29.96 ±0.56	52.56 ±0.17	5.45
EAA	25.76 ±2.42	24.78 ±0.60	35.42 ±1.01	31.19 ±0.33	52.56 ±0.17	4.05
ERA	34.10 ±1.48	32.61 ±0.80	37.34 ±0.28	40.08 ±1.17	52.48 ±0.16	3.64

Table 9. Hit@100 for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	47.24 ± 3.92	100.00 ± 0.00	100.00 ± 0.00
AA	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	53.40 ± 3.65	100.00 ± 0.00	100.00 ± 0.00
RA	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	62.61 ± 2.54	100.00 ± 0.00	100.00 ± 0.00
GCN	81.20 ± 2.84	38.77 ± 0.62	82.05 ± 3.64	29.82 ± 2.62	27.72 ± 0.96	6.34 ± 10.98	100.00 ± 0.00	100.00 ± 0.00
GRAPHSAGE	3.16 ± 5.47	45.92 ± 1.24	19.71 ± 2.89	36.45 ± 0.75	3.11 ± 3.13	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
GAT	84.04 ± 2.37	15.62 ± 2.07	82.82 ± 2.22	15.36 ± 2.57	4.03 ± 3.50	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
GAE	97.38 ± 0.79	65.08 ± 2.77	95.82 ± 0.76	42.01 ± 3.05	40.07 ± 0.52	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
VGAE	98.29 ± 0.78	50.06 ± 2.40	97.36 ± 0.90	42.24 ± 1.06	50.85 ± 0.79	19.24 ± 29.24	100.00 ± 0.00	100.00 ± 0.00
SEAL	92.51 ± 0.83	84.85 ± 0.69	87.00 ± 2.26	42.06 ± 1.32	82.68 ± 0.03	54.45 ± 0.35	100.00 ± 0.00	100.00 ± 0.00
NEO-GNN	71.85 ± 14.86	17.64 ± 15.31	31.87 ± 37.89	2.47 ± 1.39	7.87 ± 12.16	2.54 ± 4.38	100.00 ± 0.00	100.00 ± 0.00
ECN	71.34 ± 1.54	47.29 ± 1.47	79.96 ± 0.71	22.53 ± 0.58	77.72 ± 0.46	48.71 ± 2.88	100.00 ± 0.00	100.00 ± 0.00
EAA	71.28 ± 1.56	48.24 ± 1.14	79.96 ± 0.71	22.53 ± 0.58	83.98 ± 0.26	51.63 ± 1.96	100.00 ± 0.00	100.00 ± 0.00
ERA	70.14 ± 1.31	46.93 ± 1.06	80.04 ± 0.61	22.74 ± 0.51	83.75 ± 0.42	57.01 ± 0.94	100.00 ± 0.00	100.00 ± 0.00

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	32.64 ± 1.64	31.64 ± 1.03	46.54 ± 0.35	37.47 ± 0.51	50.54 ± 0.05	6.82
AA	35.34 ± 2.05	35.22 ± 0.73	45.19 ± 0.81	40.14 ± 0.53	50.54 ± 0.05	6.18
RA	43.33 ± 0.99	42.07 ± 0.92	47.38 ± 1.33	48.21 ± 0.96	50.54 ± 0.05	5.09
GCN	4.77 ± 1.13	8.07 ± 0.87	23.80 ± 5.49	1.93 ± 1.82	20.98 ± 0.54	7.45
GRAPHSAGE	0.01 ± 0.02	0.00 ± 0.00	3.88 ± 3.37	-	11.13 ± 1.22	10.82
GAT	0.00 ± 0.00	0.05 ± 0.08	0.59 ± 1.02	0.00 ± 0.00	0.38 ± 0.23	10.36
GAE	0.00 ± 0.00	3.76 ± 6.51	26.54 ± 2.38	1.82 ± 3.15	20.82 ± 0.92	7.00
VGAE	3.76 ± 3.28	7.86 ± 7.15	7.82 ± 13.55	4.83 ± 1.81	22.48 ± 0.84	5.82
SEAL	22.42 ± 3.28	33.92 ± 1.15	37.18 ± 0.54	-	-	5.82
NEO-GNN	0.00 ± 0.00	0.00 ± 0.00	1.85 ± 2.54	-	0.32 ± 0.28	11.36
ECN	31.60 ± 2.02	31.60 ± 0.44	41.93 ± 1.43	36.36 ± 0.45	53.49 ± 0.29	5.95
EAA	32.83 ± 1.95	33.51 ± 0.32	45.01 ± 1.28	37.44 ± 0.36	53.49 ± 0.29	4.77
ERA	41.45 ± 1.73	40.75 ± 0.50	47.27 ± 0.60	46.44 ± 0.91	53.41 ± 0.28	3.82

Table 10. Hit@200 for link prediction.

Model	Cora	Pubmed*	Citeseer	Actor	CS	Photo	Cornell	Wisconsin
CN	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	62.56 ± 3.31	100.00 ± 0.00	100.00 ± 0.00
AA	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	68.26 ± 1.58	100.00 ± 0.00	100.00 ± 0.00
RA	44.80 ± 1.54	28.75 ± 0.49	33.11 ± 1.67	20.30 ± 0.27	79.77 ± 0.27	75.39 ± 1.15	100.00 ± 0.00	100.00 ± 0.00
GCN	92.58 ± 1.35	52.16 ± 1.21	95.57 ± 0.28	41.25 ± 2.11	39.25 ± 0.89	8.19 ± 14.19	100.00 ± 0.00	100.00 ± 0.00
GRAPHSAGE	3.16 ± 5.47	57.41 ± 1.66	32.16 ± 23.89	47.20 ± 1.87	3.92 ± 4.44	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
GAT	93.14 ± 1.11	22.20 ± 3.00	93.11 ± 1.54	25.24 ± 2.78	10.84 ± 3.71	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
GAE	98.86 ± 0.76	79.21 ± 2.66	97.29 ± 0.67	53.63 ± 3.83	54.09 ± 0.57	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
VGAE	98.89 ± 0.49	65.75 ± 0.99	98.50 ± 0.17	54.82 ± 0.40	67.53 ± 0.84	26.81 ± 11.19	100.00 ± 0.00	100.00 ± 0.00
SEAL	94.60 ± 0.25	86.76 ± 0.63	89.56 ± 2.87	53.04 ± 0.83	83.54 ± 0.03	66.92 ± 0.29	100.00 ± 0.00	100.00 ± 0.00
NEO-GNN	75.55 ± 13.08	28.44 ± 26.39	39.60 ± 33.85	9.29 ± 11.09	9.40 ± 13.67	3.03 ± 5.23	100.00 ± 0.00	100.00 ± 0.00
ECN	79.27 ± 1.56	56.77 ± 0.31	86.52 ± 0.56	24.98 ± 0.49	83.46 ± 0.55	61.62 ± 1.37	100.00 ± 0.00	100.00 ± 0.00
EAA	79.27 ± 1.56	57.10 ± 0.25	86.52 ± 0.56	24.98 ± 0.49	88.21 ± 0.47	64.30 ± 1.63	100.00 ± 0.00	100.00 ± 0.00
ERA	78.58 ± 1.43	55.80 ± 0.36	86.52 ± 0.56	24.97 ± 0.49	87.85 ± 0.52	67.17 ± 0.56	100.00 ± 0.00	100.00 ± 0.00

Model	Wikivitals	Wikivitals-fr	Wikischools	Wikivitals+	Ogbn-arxiv	Rank
CN	39.81 ± 1.77	41.05 ± 0.13	57.41 ± 0.00	43.89 ± 0.68	50.54 ± 0.05	7.27
AA	43.19 ± 1.51	44.65 ± 0.57	58.47 ± 0.68	46.51 ± 0.71	50.54 ± 0.05	6.00
RA	50.91 ± 1.20	50.90 ± 0.70	58.26 ± 0.41	55.02 ± 0.46	50.54 ± 0.05	5.45
GCN	7.61 ± 1.10	11.88 ± 1.27	33.21 ± 7.07	3.13 ± 2.87	27.60 ± 0.83	7.36
GRAPHSAGE	0.01 ± 0.02	0.00 ± 0.00	4.13 ± 3.58	-	17.29 ± 1.20	10.64
GAT	0.00 ± 0.00	0.10 ± 0.17	0.80 ± 1.39	0.00 ± 0.00	0.85 ± 0.42	9.73
GAE	0.00 ± 0.00	5.14 ± 8.90	37.36 ± 1.14	3.03 ± 5.24	27.29 ± 1.90	6.91
VGAE	6.57 ± 5.77	14.83 ± 4.48	11.78 ± 20.40	8.72 ± 1.75	29.36 ± 0.35	5.73
SEAL	30.57 ± 5.30	42.28 ± 0.71	48.16 ± 0.32	-	-	6.18
NEO-GNN	0.00 ± 0.00	0.00 ± 0.00	10.21 ± 11.50	-	0.39 ± 0.34	11.36
ECN	39.32 ± 1.59	41.58 ± 0.19	54.75 ± 0.61	43.08 ± 0.60	54.67 ± 0.13	5.68
EAA	40.95 ± 1.63	43.20 ± 0.38	57.17 ± 0.69	44.38 ± 0.31	54.67 ± 0.13	4.59
ERA	48.75 ± 1.34	49.82 ± 0.84	57.94 ± 0.83	52.88 ± 0.25	54.61 ± 0.12	4.36