Inductive Link Prediction in Static and Temporal Graphs for Isolated Nodes

Ayan Chatterjee
Northeastern University
Boston, MA USA
chatterjee.ay@northeastern.edu

Robin Walters
Northeastern University
Boston, MA USA
r.walters@northeastern.edu

Giulia Menichetti
Harvard Medical School
Boston, MA USA
giulia.menichetti@channing.harvard.edu

Tina Eliassi-Rad
Northeastern University
Boston, MA USA
t.eliassirad@northeastern.edu

Abstract

Link prediction is a vital task in graph machine learning, involving the anticipation of connections between entities within a network. In the realm of drug discovery, link prediction takes the form of forecasting interactions between drugs and target genes. Likewise, in recommender systems, link prediction entails suggesting items to users. In temporal graphs, link prediction ranges from friendship recommendations to introducing new devices in wireless networks and dynamic routing. However, a prevailing challenge in link prediction lies in the reliance on topological neighborhoods and the lack of informative node metadata for making predictions. Consequently, predictions for nodes with low degrees, and especially for newly introduced nodes with no neighborhood data, tend to be inaccurate and misleading. State-of-the-art models frequently fall short when tasked with predicting interactions between a novel drug and an unexplored disease target or suggesting a new product to a recently onboarded user. In temporal graphs, the link prediction models often misplace a newly introduced entity in the evolving network. This paper delves into the issue of observation bias related to the inequity of data availability for different entities in a network, unavailability of informative node metadata, and explores how contemporary models struggle when it comes to making inductive link predictions for low-degree and previously unseen isolated nodes. Additionally, we propose a non-end-to-end training approach harnessing informative node attributes generated by unsupervised pre-training on corpora different from and with significantly more entities than the observed graphs to enhance the overall generalizability of link prediction models.

1 Introduction

Graph datasets are ubiquitous in diverse domains, encompassing social networks (Ball and Newman, 2013), collaboration networks (Wang et al., 2020b), protein-protein interaction (PPI) networks (Qi et al., 2006), drug-target interaction (DTI) networks (Yamanishi et al., 2008), power grids (Pagani and Aiello, 2011), and transportation networks (Lordan and Sallan, 2020). These real-world graphs are often characterized by sparsity and partial observability, making the problem of link prediction for unobserved edges of paramount importance (Liben-Nowell and Kleinberg, 2007). The significance of link prediction extends to numerous applications, ranging from predicting protein interactions (Kovács et al., 2019) to understanding drug responses (Stanfield et al., 2017), recommending products

Temporal Graph Learning Workshop @ NeurIPS 2023, New Orleans.

(Lakshmi and Bhavani, 2021), completing knowledge graphs (Nickel et al., 2016), and suggesting connections in social networks (Adamic and Adar, 2003).

Moreover, Link prediction is a crucial task in temporal networks (Qin and Yeung, 2023). Applications range from friendship and item recommendations (Campana and Delmastro, 2017; Wang et al., 2020a), intrusion detection on the internet (King and Huang, 2023), channel allocation in wireless networks (Gao et al., 2020), traffic detection and dynamic routing in optical networks (Vinchoff et al., 2020; Aibin et al., 2021), and molecular dynamic simulations (Ashby and Bilbrey, 2021). Similar to static graphs, the majority of the temporal link prediction models (Rossi et al., 2020; Xu et al., 2020) combine graph neural networks leveraging neighborhood topology with memory modules for predicting future links in both continuous and discrete time domains.

Link prediction in static graphs has been extensively studied, resulting in various methods including similarity-based indices, probabilistic approaches, and dimensionality reduction techniques (Kumar et al., 2020). For link prediction, latent representations such as Node2Vec (Grover and Leskovec, 2016) are commonly employed, capturing graph topology in low-dimensional feature vectors (Cao et al., 2015; Perozzi et al., 2014; Tang et al., 2015; Wang et al., 2016). Graph neural networks (GNN) (Zhang and Chen, 2018), graph convolutional network (GCN) (Zhang et al., 2022), graph attention network (GAT) (Veličković et al., 2018), and variational graph auto-encoder (VGAE) (Kipf and Welling, 2016) have been recently used in link prediction tasks. Although these topology-based approaches achieve commendable performance in transductive link prediction, where the nodes are shared between train and test datasets, these methods fail in making link predictions for never-before-seen nodes (Chatterjee et al., 2023b,a; Szymborski and Emad, 2022). GraphSAGE (Hamilton et al., 2017) and GraIL (Teru et al., 2019) tackle inductive link prediction in static graphs, although they leverage the neighborhood of a newly arrived node. Thus, these models are unable to handle the nodes lacking neighborhood topology, in other words, the nodes that lack sufficient samples in train and test datasets. DEAL (Hao et al., 2021) makes inductive link prediction for nodes having only node attributes. However, the DEAL alignment mechanism between node attributes and graph structure creates a high overlap between the attributes and neighborhood topology, combined with the lack of informative node metadata in real-world graphs, making the prediction task difficult for unseen nodes lacking sufficiently observed neighborhood topology.

Similar to static graphs, temporal graphs compute the embedding of time-dependent topology via shallow encoders, decomposition approaches, random walk, or autoencoders and train a downstream decoder for making temporal link prediction (Kazemi et al., 2020). Similar to GraphSAGE in static graphs, state-of-the-art (SOTA) temporal inductive link prediction models like DyHATR (Xue et al., 2020), DGCN (Manessi et al., 2020), TGAT (Xu et al., 2020), DyRep (Trivedi et al., 2018), and TGN-attn (Rossi et al., 2020) execute inductive link prediction in temporal graphs where the neighborhood-topology of test nodes is known. Thus, these models are limited to the nodes with plentiful data samples and fail to make predictions for newly arrived single nodes.

Recent research has underscored the importance of inductive link prediction, particularly for nodes lacking neighborhood information, and for improving the interpretability of predictions. For instance, AI-Bind (Chatterjee et al., 2023b) addresses the challenge of predicting the binding between a novel drug and an uncharacterized target, where both entities lack established interaction records in existing databases (Wishart et al., 2017; Liu et al., 2007). The concept of observation bias in PPI networks was first identified by Park et al. (Park and Marcotte, 2012). Link prediction models tend to excel when dealing with proteins that have abundant samples in the training dataset. In scenarios with random train-test splits, the majority of links in both train and test datasets are attributed to the hub nodes (Barabási and Alber, 1999), and overall test performance heavily relies on these highly connected nodes (Chatterjee et al., 2023b). Consequently, during inductive testing, these models struggle to provide meaningful predictions for proteins with fewer or no samples. RAPPPID (Szymborski and Emad, 2022) highlights the necessity of learning from protein molecular structures rather than relying solely on the PPI topology for better prediction on proteins with fewer samples in the databases used in training. Similar to AI-Bind in the context of DTI, RAPPPID proposes inductive tests using node attributes in PPI networks. Analogous situations arise in temporal networks, such as when introducing a new IoT (Internet of Things) device into a wireless network (Gao et al., 2020) or studying the binding of a new molecule via molecular dynamic simulations (Ashby and Bilbrey, 2021).

Contributions: (1) We address the challenge of observation bias due to unequal data availability among entities, leading to inferior model performance when dealing with low-degree and isolated

nodes lacking sufficient examples in both training and testing. **(2)** We propose unsupervised pre-training of node attributes on corpora different from and larger than the observed graph. These attributes combined with a downstream decoder trained in a non-end-to-end fashion enhance the performance of inductive link prediction, particularly for isolated nodes that have never been encountered before.

2 Problem Formulation

2.1 Static Graphs

Consider a graph instance $G = (V, E, X)$, where V represents the set of vertices (or nodes), E represents the set of edges (or links), and the node attributes are captured in the matrix X. $N_w(u)$ and $N'_w(u)$ represent the neighborhoods of node u seen by the link prediction model w during training and testing, respectively. $|N_w(u)|$ and $|N'_w(u)|$ are the number of the immediate neighbors (a.k.a., degree) of node u observed by the link prediction model w in training and testing, respectively. We focus on undirected unipartite graphs, although this formulation can be extended to encompass directed, bipartite, and multilayered graphs as well. For instance, G could represent a PPI network, where nodes correspond to proteins, links represent interactions between proteins and node attributes are molecular structure embeddings obtained using ProtVec (Asgari and Mofrad, 2015) on the amino acid sequences of the proteins.

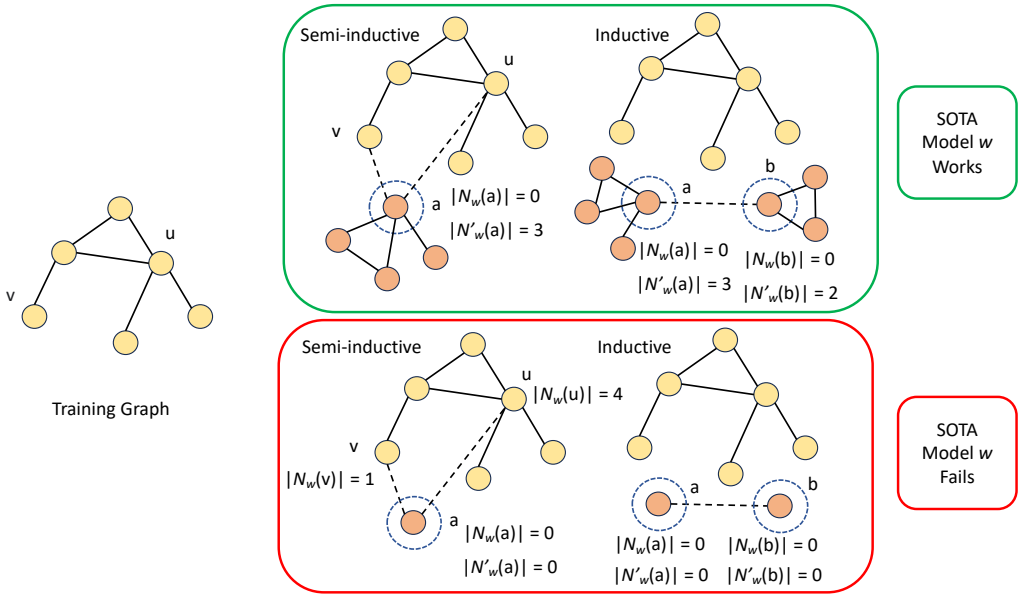

Figure 1: SOTA link prediction models leverage the neighborhood topology of nodes. In the top scenario, since the neighborhood topologies of the inductive test nodes a and b are well-observed, the topology-based encoders can represent these nodes adequately for downstream link prediction. In the bottom scenario, the neighborhoods of the inductive test nodes are unobserved by the model. Thus, in the absence of informative node metadata, the model fails to predict new edges for these newly arrived nodes a and b.

We construct a link prediction model w using supervised learning on the set of links E. The edge set is partitioned into observed and unobserved edges during training as $E = E_o \cup E_u$. Our goal is to learn a function that maps the observed nodes and node attributes (V_o, X_o) to the observed edges E_o, with the hope that it will generalize to the unobserved edges E_u. We also define three types of link prediction scenarios based on the observed and unobserved nodes, denoted as V_o and V_u, respectively:

- Transductive: Predicting $(a, b) \in E_u$, where $a, b \in V_o$,
- Semi-inductive: Predicting $(a, b) \in E_u$, where $a \in V_o$ and $b \in V_u$ or vice-versa,

- Inductive: Predicting $(a, b) \in E_u$, where $a, b \in V_u$.

In this work, we focus on the semi-inductive and inductive link prediction scenarios. The link prediction model takes input $\{V_o, X_o, E_o\}$, and makes predictions on E_u induced by the isolated node(s) u where $|N_w(u)| = 0$ and $|N'_w(u)| = 0 \; \forall u \in V_u$ (see Figure 1).

2.2 Temporal Graphs

We consider discrete-time temporal graphs (Holme and Saramäki, 2012), where the graph instances are captured at intervals of Δt, with time stamps t_0, $t_1 = t_0 + \Delta t$, ..., $t_n = t_0 + n\Delta t$.

The temporal graph instances of the evolving network are denoted as $G_{t_0} = (V_{t_0}, E_{t_0}, X_{t_0})$, $G_{t_1} = (V_{t_1}, E_{t_1}, X_{t_1})$, ..., $G_{t_n} = (V_{t_n}, E_{t_n}, X_{t_n})$, corresponding to time stamps t_0, t_1, ..., t_n, respectively. We define the unobserved node set from time t_j to t_{j+1} as $V_u = V_{t_{j+1}} \setminus V_{t_j}$, and the unobserved edge set as $E_u = E_{t_{j+1}} \setminus E_{t_j}$. In this setting, the link prediction model takes input $G_{t_j} = (V_{t_j}, E_{t_j}, X_{t_j})$ and makes predictions on E_u induced by the isolated node(s) u, where $|N_w(u)| = 0$ and $|N'_w(u)| = 0$ $\forall u \in V_u$.

3 Observation Bias

Power-law degree distributions are common in real-world graphs (Barabási and Alber, 1999; Faloutsos et al., 1999), resulting in a majority of nodes having low degrees while a few are highly connected. Traditional machine learning cross-validation randomly splits edges, leading to both train and test sets being dominated by high-degree nodes (hubs). This allows the SOTA models to effectively learn the neighborhood topology of the hubs, resulting in accurate predictions and overall high performance (Mara et al., 2022, 2020; Chatterjee et al., 2023b,a).

However, this success is not uniform across nodes with varying degrees. In real-world data, a node's degree often correlates with the available information or samples. For example, extensively studied drugs with more samples in databases like DrugBank (Wishart et al., 2017) and BindingDB (Liu et al., 2007) become hubs in the training dataset for DTI prediction. Models like (Huang et al., 2020b,a) heavily rely on these well-explored drugs, leading to accurate hub-related predictions. However, they struggle when predicting interactions involving new drugs with low degrees due to limited data (Chatterjee et al., 2023b). Predicting in data-scarce scenarios is crucial, particularly for rare diseases (Genes, 2022) and developing new drugs for them when information is lacking in established disease databases like DisGeNET (Pinero et al., 2019).

In Figure 2, we delve into this phenomenon within the benchmark drug-drug interaction (DDI) network (Wishart et al., 2017; Guney, 2017) from Open Graph Benchmark (OGB) (Hu et al., 2020). We employ GraphSAGE (Hamilton et al., 2017), an inductive link prediction method designed for static graphs. The drugs in ogbl-ddi do not contain any node metadata. Thus, GraphSAGE only leverages the DDI topology. We illustrate the power-law degree distribution inherent in the DDI graph in Figure 2E, which confirms the presence of low-degree drugs and hubs. Consequently, certain drugs benefit from richer training data. Figure 2F breaks down the performance of GraphSAGE based on observed node neighborhoods during training, revealing its limitations when dealing with nodes with $|N_w(u)| = 0$. Conversely, for drugs with ample training samples and high degrees, GraphSAGE attains significantly higher prediction performance.

4 Experiments

4.1 Methodology

To improve inductive link prediction performance for isolated and low-degree nodes, we model link prediction as a pairwise learning task (Ying and Zhou, 2015) on node attributes. The methodology is visualized in Figure 3. Instead of learning the node representations and training the downstream decoder altogether in a traditional end-to-end setting (Božič et al., 2020), we first learn the node attributes on a large corpus independent of the observed graph and then train the decoder on the training edges. We test our method on three static graphs from OGB (Hu et al., 2020) and one temporal graph from Reddit (Kumar et al., 2018). The methodology for obtaining the pre-trained

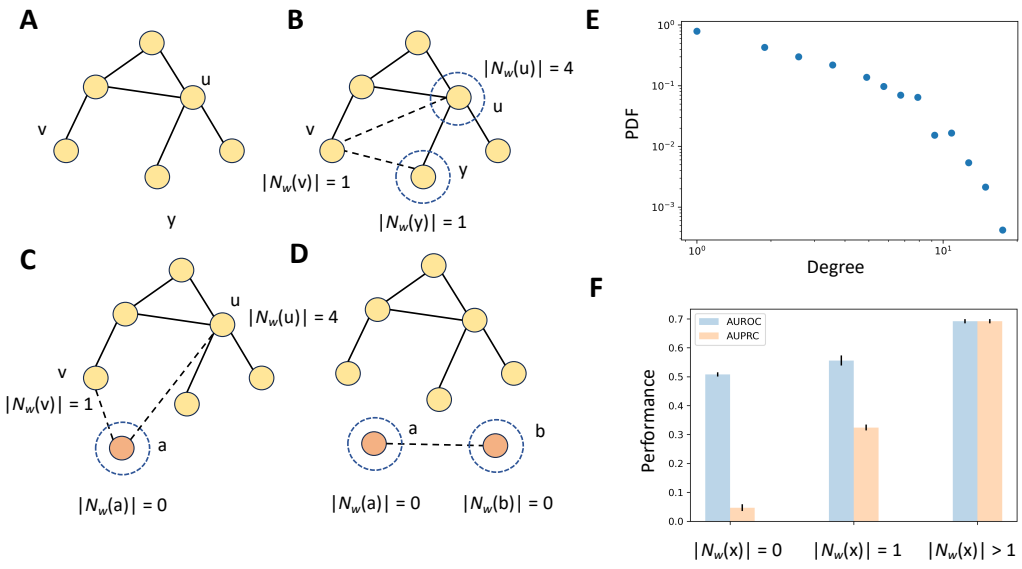

Figure 2: GraphSAGE applied to the ogb-ddi dataset. **(A)** Sample training graph instance. **(B)** Transductive link prediction: Predicting the link between nodes v and y is facilitated by leveraging neighborhood topology. **(C)** Semi-inductive link prediction: Link prediction for a new node a depends on the degree of the existing node in the graph. Preferential attachment (Barabási and Alber, 1999) is a key mechanism in the emergence of power-law degree distributed graphs, and hence the edge u-a is more likely to exist and is easier to predict compared to the edge v-a. **(D)** GraphSAGE struggles with link prediction for nodes a and b whose neighborhoods are unobserved. **(E)** The drug-drug interaction network exhibits a power-law degree distribution. **(F)** GraphSAGE excels in link prediction for high-degree nodes but falters for nodes with limited observed edges or no neighborhood data. Results are reported based on a 5-fold cross-validation.

node attributes is described here: **(a)** In ogbl-ppa (protein-protein interaction network), we employ 100-dimensional ProtVec vectors as node attributes (Asgari and Mofrad, 2015), trained on a larger and more diverse dataset of 1,640,370 amino acid trigram sequences from the Swiss-Prot database (Bairoch, 1996), compared to the ogbl-ppi graph, which contains 576,289 proteins. **(b)** In ogbl-collab (collaboration network), we employ 128-dimensional Word2Vec embeddings trained on a vast corpus of 10 billion Google news articles (Mikolov et al., 2013), which is both disjoint from and considerably larger than the 300 million papers authored by 235,868 researchers in ogbl-collab. The average of the Word2Vec embeddings of the papers for each author is used as node attributes. **(c)** In ogbl-ddi (drug-drug interaction network), we begin by retrieving drug molecular SMILES (Weininger, 1988) from PubChem (Kim et al., 2015). Subsequently, we utilize 300-dimensional Mol2vec embeddings (Jaeger et al., 2018) as node attributes, which were pre-trained on an extensive corpus of 19.9 million chemicals from ZINC (Irwin et al., 2012) and ChEMBL (Gaulton et al., 2011) libraries. This training corpus significantly surpasses the 4,267 drugs found in ogbl-ddi. **(d)** In the Reddit hyperlink network, we create pre-trained node attributes from post content using GloVe word embeddings trained on a massive corpus of 840 billion tokens (Pennington et al., 2014). It's worth noting that the Reddit network (Kumar et al., 2018), consisting of 3.6 billion tokens, involving 118,381 users and 51,278 subreddits, is substantially smaller than the GloVe training dataset.

Training the node attributes on a corpus larger than the observed graph makes the attributes generalizable toward newly arrived nodes since the embeddings are independent of the training graph topology. In the next phase, we train a decoder in a supervised manner on the training graph, with the concatenation of the pair of node embeddings for each training link as input. This decoder is then used for predicting the link probabilities on unseen nodes.

We use random node split inspired by GraIL (Teru et al., 2019) for creating inductive and semi-inductive link prediction scenarios in static graphs. In this train-test-validation split method, we

randomly split the nodes of the original graph V into three groups (V_{train}, $V_{validation}$, and V_{test}) at an 80:10:10 ratio. Then, we obtain the subgraphs G_{train}, $G_{validation}$, and G_{test} induced by V_{train}, $V_{validation}$, and V_{test}, respectively. This approach creates multiple disconnected components in the test data, making the test graph topology significantly different from the train graph.

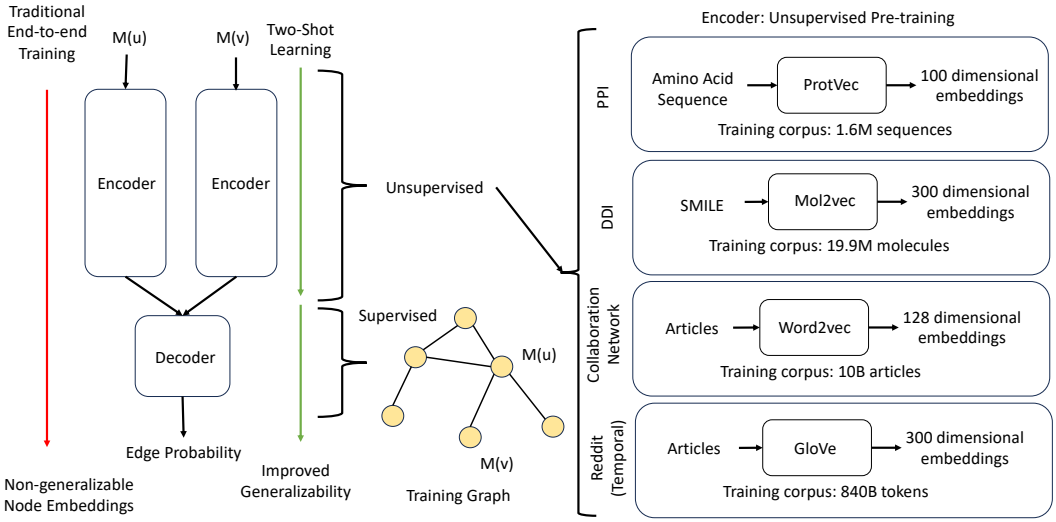

Figure 3: Our methodology pre-trains the node attributes in an unsupervised manner on a corpus significantly different from and larger than the observed training graph. These node attributes are then concatenated and passed through a decoder, which is trained in a supervised fashion using the links in the train graph. In our two-shot learning (unsupervised + supervised) setup, we formulate link prediction as a binary classification task, where the output of the decoder represents the probability of an unobserved edge existing between two unseen nodes. We implemented a multi-layer perception (MLP) as our decoder architecture.

4.2 Inductive Link Prediction on Isolated Nodes in Static Graphs

Our approach does not rely on the neighborhood topology of nodes for making link predictions, but relies solely on the pre-trained node attributes, and is able to make accurate link predictions for unseen isolated nodes. We use as the downstream decoders (see Figure 3) OGB-defined MLPs (see here). These MLPs take concatenated attributes of the two nodes at the end of each edge as input. We compare our approach with the top-performing state-of-the-art model from the OGB leaderboard PLNLP (Wang et al., 2021). PLNLP combines SAGE (Hamilton et al., 2017) neighborhood encoders concatenated with OGB node attributes (Hu et al., 2020) with pairwise learning decoders. ogbl-ddi has no node attributes in the benchmark. ogbl-ppi has 58-dimensional one-hot node attributes that indicate the species that the corresponding protein comes from. ogbl-collab has 128-dimensional features, obtained by averaging the word embeddings of papers that are published by the authors.

PLNLP outperforms multiple well-established static link prediction models like GraphSAGE (Hamilton et al., 2017), GCN (Kipf and Welling, 2017), SEAL (Zhang and Chen, 2017), Node2vec (Grover and Leskovec, 2016), DeepWalk (Perozzi et al., 2014), and CFLP (Zhao et al., 2022). Our findings, summarized in Table 1, demonstrate that the inductive performance of PLNLP is significantly lower compared to its transductive performance. Thus, the majority of the prediction power of PLNLP is derived from leveraging the neighborhood-topology of the nodes. Furthermore, our approach combining pre-trained node attributes with MLP outperforms PLNLP in the inductive test scenario (see Table 2). Thus, learning from neighborhood topology is insufficient in inductive link prediction scenarios involving isolated nodes, and the need to learn from informative node attributes is validated.

4.3 Inductive Link Prediction on Isolated Nodes in Temporal Graphs

We evaluate our approach on temporal networks for link prediction on newly arrived nodes. We utilize the Reddit hyperlink network dataset (Kumar et al., 2018), which spans 3 years (2014-2017) and

Table 1: PLNLP is a top-performing link prediction model from OGB leaderboard leveraging pairwise learning with SAGE neighborhood encoders and OGB-provided node attributes. PLNLP's performance significantly reduces from transductive to inductive link prediction scenarios. Hits@Top K is evaluated with default values: K=100 for ogbl-ppa, K=50 for ogbl-collab, and K=20 for ogbl-ddi, as recommended by OGB benchmark metrics. We conduct a 5-fold cross-validation for each dataset.

Dataset	**Transductive Hits@Top K**	**Inductive Hits@Top K**
ogbl-ppa	32.38 ± 2.58	0.09 ± 0.03
ogbl-collab	70.59 ± 0.29	11.56 ± 0.93
ogbl-ddi	90.88 ± 3.13	0.01 ± 0.02

Table 2: Our non-end-to-end training approach of combining MLP and pre-trained node attributes outperforms PLNLP in inductive tests involving isolated nodes in terms of three performance metrics AUROC, AUPRC, and Hits@Top K. Hits@Top K is evaluated with default values: K=100 for ogbl-ppa, K=50 for ogbl-collab, and K=20 for ogbl-ddi, as recommended by OGB benchmark metrics. AUROC is the Area Under the Receiver Operating Characteristics and AUPRC is the Area Under the Precision Recall Curve (average precision). We conduct a 5-fold cross-validation for each dataset.

Dataset	**PLNLP**		**Pre-trained Node Attributes**	
	AUROC	**AUPRC**	**AUROC**	**AUPRC**
ogbl-ppa	0.51 ± 0.03	0.12 ± 0.04	0.78 ± 0.03	0.35 ± 0.03
ogbl-collab	0.61 ± 0.03	0.23 ± 0.07	0.97 ± 0.02	0.92 ± 0.02
ogbl-ddi	0.50 ± 0.04	0.11 ± 0.07	0.54 ± 0.02	0.21 ± 0.02

Dataset	**PLNLP Hits@TopK(%)**	**Pre-trained Node Attributes Hits@TopK(%)**
ogbl-ppa	0.09 ± 0.03	0.39 ± 0.03
ogbl-collab	11.56 ± 0.93	36.44 ± 3.11
ogbl-ddi	0.01 ± 0.02	0.39 ± 0.02

consists of subreddit communities. We train each year and execute link prediction for newly arrived isolated nodes in the following year. We employ a 3-layer MLP decoder with hidden layer sizes of 100, a learning rate of 0.001, 200 epochs, ReLU activation, and ADAM solver. We compare our approach leveraging pre-trained node attributes with DyHATR (Xue et al., 2020), a state-of-the-art temporal link prediction model, in inductive link prediction. DyHATR uses a hierarchical attention mechanism to learn heterogeneous information from graph topology and incorporates recurrent neural networks with temporal attention to capture evolutionary patterns. DyHATR outperforms multiple SOTA temporal link prediction models such as DHNE (Yin et al., 2019), metapath2vec-GRU and metapath2vec-LSTM (Dong et al., 2017), dyngraph2vec (Goyal et al., 2020), GraphSAGE-LSTM, and DySAT (Sankar et al., 2020). We observe in Table 3 that our approach outperforms DyHATR for newly arrived isolated nodes.

Table 3: We evaluate inductive link prediction performance on newly arrived isolated nodes across various temporal instances of the subreddit network. Non-end-to-end training with pre-trained node attributes yields the best performance compared to DyHATR in inductive tests for isolated nodes.

Model	**2014-2015**		**2016-2017**		**2017-2018**	
	AUROC	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**
Pre-trained Node Attributes	0.70	0.66	0.63	0.60	0.69	0.65
DyHATR	0.45	0.25	0.45	0.48	0.46	0.28

5 Related work

Joachims (1999) highlighted inductive test challenges in machine learning, particularly in text classification. Present link prediction models (Ai et al., 2022) combine graph topology and node attributes. Successful transductive methods like Planetoid (Yang et al., 2016), GraphSAGE (Hamilton et al., 2017), and GraIL (Teru et al., 2019) share effective performance when train and test graphs have similar topologies. However, they struggle with nodes lacking observed neighborhood topologies, especially those with limited data samples. DEAL (Hao et al., 2021) innovatively combines topological data and node attributes for link prediction. SEG (Ai et al., 2022) improves transductive performance using a one-layer graph convolutional network (GCN) for topology encoding and an MLP for node attributes. Yet, the absence of informative node attributes in real-world networks limits these models to graph topology.

Erhan et al. (2010) explained the advantages of unsupervised pre-training, a concept we apply to link prediction. End-to-end training can lead to overfitting to observed nodes, hampering generalization to unseen ones. To address this, we use unsupervised pre-training of node attributes and a downstream decoder trained in a non-end-to-end fashion, coupled with low learning rates and early stopping on validation data. This approach helps the model find a more generalizable minimum, greatly enhancing inductive link prediction for unseen isolated nodes. Erhan et al. (2010) also noted that larger unsupervised pre-training datasets can further improve generalizability.

The inadequacy of SOTA models in inductive link prediction in DTIs and PPIs has recently come under scrutiny in AI-Bind (Chatterjee et al., 2023b) and RAPPPID (Szymborski and Emad, 2022), respectively. AI-Bind's findings reveal a marked decline in the performance of two prominent SOTA DTI prediction models, DeepPurpose (Huang et al., 2020a) and MolTrans (Huang et al., 2020b), when transitioning from transductive tests (DeepPurpose AUROC 0.82 ± 0.003, AUPRC 0.48 ± 0.004; MolTrans AUROC 0.86 ± 0.07, AUPRC 0.80 ± 0.09) to inductive tests (DeepPurpose AUROC 0.60 ± 0.066, AUPRC 0.42 ± 0.063; MolTrans AUROC 0.62 ± 0.02, AUPRC 0.48 ± 0.03). Similar observations apply to two SOTA PPI prediction models, SPRINT (Li and Ilie, 2017) and DeepPPI (Richoux et al., 2019), with their inductive link prediction performance (SPRINT AUROC 0.65; DeepPPI AUROC 0.63) significantly lower than their transductive counterparts (SPRINT AUROC 0.97; DeepPPI AUROC 0.87).

AI-Bind addresses inductive link prediction in DTI bipartite graphs by integrating network-derived negatives and employing a two-shot learning strategy (Chatterjee et al., 2023b). RAPPPID enhances inductive PPI prediction with LSTM-regularized protein embeddings (Hochreiter and Schmidhuber, 1997), but struggles with generalization beyond the training data (Szymborski and Emad, 2022). Our approach combines unsupervised pre-training of node attributes with a two-shot learning-trained downstream decoder, resulting in significant inductive link prediction improvements across diverse unipartite graphs, even with traditional random negative sampling (Yang et al., 2020).

Node attributes play a crucial role in graph stream problems, dynamic network research (Jiang et al., 2015; Li et al., 2018), and cold-start scenarios (Gantner et al., 2010; Li et al., 2021). While pre-training has been used in dynamic network research (Shao et al., 2022) and cold-start problems (Liu et al., 2023; Hao et al., 2023; Wang et al., 2022), it has primarily focused on training GNN model parameters rather than node attributes. Our method introduces pre-training for node attributes, which is novel and has proven to be crucial in developing interpretable and generalizable link prediction models (Chatterjee et al., 2023b; Szymborski and Emad, 2022).

6 Conclusion and Future Work

Our work builds on link prediction models' strengths in transductive tests, emphasizing the need for robust inductive link prediction in both static and temporal graphs. We note a considerable performance drop in state-of-the-art models during inductive tests, especially for isolated nodes. Our two-shot learning method, involving unsupervised pre-training of node attributes on a significantly larger corpus than the observed graph, improves inductive link prediction for new isolated nodes. We will investigate the impact of pre-trained node attributes' training corpus size on inductive link prediction. In essence, our approach addresses data biases in link prediction and bolsters generalizability in both static and temporal graphs.

Our code has been made open-source at: `https://github.com/ChatterjeeAyan/ILP`.

Acknowledgment

R. Walters is supported by NSF grants 2107256 and 2134178.

References

Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the Web. *Social Networks* 25, 3 (July 2003), 211–230. https://doi.org/10.1016/s0378-8733(03)00009-1

Baole Ai, Zhou Qin, Wenting Shen, and Yong Li. 2022. Structure Enhanced Graph Neural Networks for Link Prediction. https://doi.org/10.48550/ARXIV.2201.05293

Michał Aibin, Nathan Chung, Tyler Gordon, Liam Lyford, and Connor Vinchoff. 2021. On Short- and Long-Term Traffic Prediction in Optical Networks Using Machine Learning. In *2021 International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, Gothenburg, Sweden, 1–6. https://doi.org/10.23919/ONDM51796.2021.9492437

Ehsaneddin Asgari and Mohammad R. K. Mofrad. 2015. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLOS ONE* 10, 11 (Nov. 2015), e0141287. https://doi.org/10.1371/journal.pone.0141287

Michael Hunter Ashby and Jenna A. Bilbrey. 2021. Geometric learning of the conformational dynamics of molecules using dynamic graph neural networks. arXiv:2106.13277 [cs.LG]

A Bairoch. 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Research* 24, 1 (Jan. 1996), 21–25. https://doi.org/10.1093/nar/24.1.21

Brian Ball and M.E.J. Newman. 2013. Friendship networks and social status. *Network Science* 1, 1 (apr 2013), 16–30. https://doi.org/10.1017/nws.2012.4

Albert-László Barabási and Réka Alber. 1999. Emergence of Scaling in Random Networks. *Science* 286, 5439 (Oct. 1999), 509–512. https://doi.org/10.1126/science.286.5439.509

Jakob Božič, Domen Tabernik, and Danijel Skočaj. 2020. End-to-end training of a two-stage neural network for defect detection. arXiv:2007.07676 [cs.CV]

Mattia G. Campana and Franca Delmastro. 2017. Recommender Systems for Online and Mobile Social Networks: A survey. *Online Social Networks and Media* 3-4 (oct 2017), 75–97. https://doi.org/10.1016/j.osnem.2017.10.005

Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. GraRep: Learning Graph Representations with Global Structural Information. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (Melbourne, Australia) *(CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 891–900. https://doi.org/10.1145/2806416.2806512

Ayan Chatterjee, Robin Walters, Giulia Menichetti, and Tina Eliassi-Rad. 2023a. Disentangling Node Attributes from Graph Topology for Improved Generalizability in Link Prediction. arXiv:2307.08877 [cs.LG]

Ayan Chatterjee, Robin Walters, Zohair Shafi, Omair Shafi Ahmed, Michael Sebek, Deisy Gysi, Rose Yu, Tina Eliassi-Rad, Albert-László Barabási, and Giulia Menichetti. 2023b. Improving the generalizability of protein-ligand binding predictions with AI-Bind. *Nature Communications* 14, 1 (April 2023), 1–15. https://doi.org/10.1038/s41467-023-37572-z

Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) *(KDD '17)*. Association for Computing Machinery, New York, NY, USA, 135–144. https://doi.org/10.1145/3097983.3098036

Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. 2010. Why Does Unsupervised Pre-training Help Deep Learning? *Journal of Machine Learning Research* 11, 19 (2010), 625–660. http://jmlr.org/papers/v11/erhan10a.html

Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. 1999. On Power-Law Relationships of the Internet Topology. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (Cambridge, Massachusetts, USA) *(SIGCOMM '99)*. Association for Computing Machinery, New York, NY, USA, 251–262. https://doi.org/10.1145/316188.316229

Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Lars Schmidt-Thieme. 2010. Learning Attribute-to-Feature Mappings for Cold-Start Recommendations. In *2010 IEEE International Conference on Data Mining*. IEEE, Sydney, Australia, 176–185. https://doi.org/10.1109/ICDM.2010.129

Weifeng Gao, Zhiwei Zhao, Zhengxin Yu, Geyong Min, Minghang Yang, and Wenjie Huang. 2020. Edge-Computing-Based Channel Allocation for Deadline-Driven IoT Networks. *IEEE Transactions on Industrial Informatics* 16, 10 (2020), 6693–6702. https://doi.org/10.1109/TII.2020.2973754

A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington. 2011. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40, D1 (Sept. 2011), D1100–D1107. https://doi.org/10.1093/nar/gkr777

G. Genes. 2022. Rare diseases, common challenges. *Nature Genetics* 54, 3 (March 2022), 215–215. https://doi.org/10.1038/s41588-022-01037-8

Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. 2020. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems* 187 (Jan. 2020), 104816. https://doi.org/10.1016/j.knosys.2019.06.024

Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 855–864. https://doi.org/10.1145/2939672.2939754

Emre Guney. 2017. Reproducible Drug Repurposing: When Similarity Does Not Suffice. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 22 (2017), 132–143. https://api.semanticscholar.org/CorpusID:32573150

William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. https://doi.org/10.48550/ARXIV.1706.02216

Bowen Hao, Hongzhi Yin, Jing Zhang, Cuiping Li, and Hong Chen. 2023. A Multi-Strategy-Based Pre-Training Method for Cold-Start Recommendation. *ACM Trans. Inf. Syst.* 41, 2, Article 31 (jan 2023), 24 pages. https://doi.org/10.1145/3544107

Yu Hao, Xin Cao, Yixiang Fang, Xike Xie, and Sibo Wang. 2021. Inductive Link Prediction for Nodes Having Only Attribute Information. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI'20)*. ACM, Yokohama, Yokohama, Japan, Article 168, 7 pages.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (nov 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Petter Holme and Jari Saramäki. 2012. Temporal networks. *Physics Reports* 519, 3 (oct 2012), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open Graph Benchmark: Datasets for Machine Learning on Graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS'20)*. Curran Associates Inc., Red Hook, NY, USA, Article 1855, 16 pages.

Kexin Huang, Tianfan Fu, Lucas M Glass, Marinka Zitnik, Cao Xiao, and Jimeng Sun. 2020a. DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics* 36, 22-23 (Dec. 2020), 5545–5547. https://doi.org/10.1093/bioinformatics/btaa1005

Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. 2020b. MolTrans: Molecular Interaction Transformer for drug–target interaction prediction. *Bioinformatics* 37, 6 (Oct. 2020), 830–836. https://doi.org/10.1093/bioinformatics/btaa880

John J. Irwin, Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. 2012. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling* 52, 7 (June 2012), 1757–1768. https://doi.org/10.1021/ci3001277

Sabrina Jaeger, Simone Fulle, and Samo Turk. 2018. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *Journal of Chemical Information and Modeling* 58, 1 (Jan. 2018), 27–35. https://doi.org/10.1021/acs.jcim.7b00616

Maosheng Jiang, Yonxiang Chen, and Ling Chen. 2015. Link Prediction in Networks with Nodes Attributes by Similarity Propagation. arXiv:1502.04380 [cs.SI]

Thorsten Joachims. 1999. Transductive Inference for Text Classification Using Support Vector Machines. In *Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 200–209.

Seyed Mehran Kazemi, Rishab Goel, Kshitij Jain, Ivan Kobyzev, Akshay Sethi, Peter Forsyth, and Pascal Poupart. 2020. Representation Learning for Dynamic Graphs: A Survey. *Journal of Machine Learning Research* 21, 70 (2020), 1–73. http://jmlr.org/papers/v21/19-447.html

Sunghwan Kim, Paul A. Thiessen, Evan E. Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A. Shoemaker, Jiyao Wang, Bo Yu, Jian Zhang, and Stephen H. Bryant. 2015. PubChem Substance and Compound databases. *Nucleic Acids Research* 44, D1 (Sept. 2015), D1202–D1213. https://doi.org/10.1093/nar/gkv951

Isaiah J. King and H. Howie Huang. 2023. Euler: Detecting Network Lateral Movement via Scalable Temporal Link Prediction. *ACM Trans. Priv. Secur.* 26, 3, Article 35 (jun 2023), 36 pages. https://doi.org/10.1145/3588771

Thomas N. Kipf and Max Welling. 2016. Variational Graph Auto-Encoders. arXiv:1611.07308 [stat.ML]

Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 2017 International Conference on Learning Representations (ICLR)*. ACM, Palais des Congrès Neptune, Toulon, France, 1–14.

István A. Kovács, Katja Luck, Kerstin Spirohn, Yang Wang, Carl Pollis, Sadie Schlabach, Wenting Bian, Dae-Kyum Kim, Nishka Kishore, Tong Hao, Michael A. Calderwood, Marc Vidal, and Albert-László Barabási. 2019. Network-based prediction of protein interactions. *Nature Communications* 10, 1 (March 2019), 1240–1248. https://doi.org/10.1038/s41467-019-09177-y

Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, and Bhaskar Biswas. 2020. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications* 553 (Sept. 2020), 124289. https://doi.org/10.1016/j.physa.2020.124289

Srijan Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Community Interaction and Conflict on the Web. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 933–943. https://doi.org/10.1145/3178876.3186141

T. Jaya Lakshmi and S. Durga Bhavani. 2021. Link Prediction Approach to Recommender Systems. https://doi.org/10.48550/ARXIV.2102.09185

Jundong Li, Kewei Cheng, Liang Wu, and Huan Liu. 2018. Streaming Link Prediction on Dynamic Attributed Networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 369–377. https://doi.org/10.1145/3159652.3159674

Shijun Li, Wenqiang Lei, Qingyun Wu, Xiangnan He, Peng Jiang, and Tat-Seng Chua. 2021. Seamlessly Unifying Attributes and Items: Conversational Recommendation for Cold-Start Users. *ACM Trans. Inf. Syst.* 39, 4, Article 40 (aug 2021), 29 pages. https://doi.org/10.1145/3446427

Yiwei Li and Lucian Ilie. 2017. SPRINT: ultrafast protein-protein interaction prediction of the entire human interactome. *BMC Bioinformatics* 18, 1 (Nov. 2017), 1–11. https://doi.org/10.1186/s12859-017-1871-x

David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58, 7 (2007), 1019–1031. https://doi.org/10.1002/asi.20591

T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson. 2007. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research* 35, Database (Jan. 2007), D198–D201. https://doi.org/10.1093/nar/gkl999

Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. 2023. GraphPrompt: Unifying Pre-Training and Downstream Tasks for Graph Neural Networks. arXiv:2302.08043 [cs.LG]

Oriol Lordan and Jose M. Sallan. 2020. Dynamic measures for transportation networks. *PLOS ONE* 15, 12 (Dec. 2020), e0242875. https://doi.org/10.1371/journal.pone.0242875

Franco Manessi, Alessandro Rozza, and Mario Manzo. 2020. Dynamic Graph Convolutional Networks. *Pattern Recogn.* 97, C (jan 2020), 18 pages. https://doi.org/10.1016/j.patcog.2019.107000

Alexandru Cristian Mara, Jefrey Lijffijt, and Tijl De Bie. 2022. An Empirical Evaluation of Network Representation Learning Methods. *Big Data* 3:X (March 2022), 1–20. https://doi.org/10.1089/big.2021.0107

Alexandru Cristian Mara, Jefrey Lijffijt, and Tijl de Bie. 2020. Benchmarking Network Embedding Models for Link Prediction: Are We Making Progress?. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, Sydney, Australia, 1–10. https://doi.org/10.1109/dsaa49011.2020.00026

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. https://doi.org/10.48550/ARXIV.1301.3781

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016. A Review of Relational Machine Learning for Knowledge Graphs. *Proc. IEEE* 104, 1 (jan 2016), 11–33. https://doi.org/10.1109/jproc.2015.2483592

Giuliano Andrea Pagani and Marco Aiello. 2011. The Power Grid as a Complex Network: a Survey. https://doi.org/10.48550/ARXIV.1105.3338

Yungki Park and Edward M Marcotte. 2012. Flaws in evaluation schemes for pair-input computational predictions. *Nature Methods* 9, 12 (Dec. 2012), 1134–1136. https://doi.org/10.1038/nmeth.2259

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA) *(KDD '14)*. Association for Computing Machinery, New York, NY, USA, 701–710. https://doi.org/10.1145/2623330.2623732

Janet Pinero, Juan Manuel Ramirez-Anguita, Josep Sauch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I Furlong. 2019. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 48 (Nov. 2019), 845–855. https://doi.org/10.1093/nar/gkz1021

Yanjun Qi, Ziv Bar-Joseph, and Judith Klein-Seetharaman. 2006. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins: Structure, Function, and Bioinformatics* 63, 3 (Jan. 2006), 490–500. https://doi.org/10.1002/prot.20865

Meng Qin and Dit-Yan Yeung. 2023. Temporal Link Prediction: A Unified Framework, Taxonomy, and Review. arXiv:2210.08765 [cs.SI]

Florian Richoux, Charlène Servantie, Cynthia Borès, and Stéphane Téletchéa. 2019. Comparing two deep learning sequence-based models for protein-protein interaction prediction. https://doi.org/10.48550/ARXIV.1901.06268

Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. 2020. Temporal Graph Networks for Deep Learning on Dynamic Graphs. arXiv:2006.10637 [cs.LG]

Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. 2020. DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (Houston, TX, USA) *(WSDM '20)*. Association for Computing Machinery, New York, NY, USA, 519–527. https://doi.org/10.1145/3336191.3371845

Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. 2022. Pre-Training Enhanced Spatial-Temporal Graph Neural Network for Multivariate Time Series Forecasting. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD '22)*. Association for Computing Machinery, New York, NY, USA, 1567–1577. https://doi.org/10.1145/3534678.3539396

Zachary Stanfield, Mustafa Coşkun, and Mehmet Koyutürk. 2017. Drug Response Prediction as a Link Prediction Problem. *Scientific Reports* 7, 1 (Jan. 2017), 1–12. https://doi.org/10.1038/srep40321

Joseph Szymborski and Amin Emad. 2022. RAPPPID: towards generalizable protein interaction prediction with AWD-LSTM twin networks. *Bioinformatics* 38, 16 (June 2022), 3958–3967. https://doi.org/10.1093/bioinformatics/btac429

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-Scale Information Network Embedding. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1067–1077. https://doi.org/10.1145/2736277.2741093

Komal K. Teru, Etienne Denis, and William L. Hamilton. 2019. Inductive Relation Prediction by Subgraph Reasoning. https://doi.org/10.48550/ARXIV.1911.06962

Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. 2018. Representation Learning over Dynamic Graphs. arXiv:1803.04051 [cs.LG]

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. arXiv:1710.10903 [stat.ML]

C. Vinchoff, N. Chung, T. Gordon, L. Lyford, and M. Aibin. 2020. Traffic Prediction in Optical Networks Using Graph Convolutional Generative Adversarial Networks. In *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. IEEE, Bari, Italy, 1–4. https://doi.org/10.1109/ICTON51198.2020.9203477

Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural Deep Network Embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, San Francisco, CA, USA, 1225–1234. https://doi.org/10.1145/2939672.2939753

Jianling Wang, Kaize Ding, Liangjie Hong, Huan Liu, and James Caverlee. 2020a. Next-Item Recommendation with Sequential Hypergraphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 1101–1110. https://doi.org/10.1145/3397271.3401133

Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. 2020b. Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies* 1, 1 (Feb. 2020), 396–413. https://doi.org/10.1162/qss_a_00021

Yiqi Wang, Chaozhuo Li, Zheng Liu, Mingzheng Li, Jiliang Tang, Xing Xie, Lei Chen, and Philip S. Yu. 2022. An Adaptive Graph Pre-Training Framework for Localized Collaborative Filtering. *ACM Trans. Inf. Syst.* 41, 2, Article 43 (dec 2022), 27 pages. https://doi.org/10.1145/3555372

Zhitao Wang, Yong Zhou, Litao Hong, Yuanhang Zou, Hanjing Su, and Shouzhi Chen. 2021. Pairwise Learning for Neural Link Prediction. https://doi.org/10.48550/ARXIV.2112.02936

David Weininger. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28, 1 (Feb. 1988), 31–36. https://doi.org/10.1021/ci00057a005

David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. 2017. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* 46, D1 (Nov. 2017), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2020. Inductive Representation Learning on Temporal Graphs. arXiv:2002.07962 [cs.LG]

Hansheng Xue, Luwei Yang, Wen Jiang, Yi Wei, Yi Hu, and Yu Lin. 2020. Modeling Dynamic Heterogeneous Network for Link Prediction Using Hierarchical Attention with Temporal RNN. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part I* (Ghent, Belgium). Springer-Verlag, Berlin, Heidelberg, 282–298. https://doi.org/10.1007/978-3-030-67658-2_17

Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda, and Minoru Kanehisa. 2008. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, 13 (July 2008), i232–i240. https://doi.org/10.1093/bioinformatics/btn162

Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. 2016. Revisiting Semi-Supervised Learning with Graph Embeddings. https://doi.org/10.48550/ARXIV.1603.08861

Zhen Yang, Ming Ding, Chang Zhou, Hongxia Yang, Jingren Zhou, and Jie Tang. 2020. Understanding Negative Sampling in Graph Representation Learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. Association for Computing Machinery, New York, NY, USA, 1666–1676. https://doi.org/10.1145/3394486.3403218

Ying Yin, Li-Xin Ji, Jian-Peng Zhang, and Yu-Long Pei. 2019. DHNE: Network Representation Learning Method for Dynamic Heterogeneous Networks. *IEEE Access* 7 (2019), 134782–134792. https://doi.org/10.1109/ACCESS.2019.2942221

Yiming Ying and Ding-Xuan Zhou. 2015. Online Pairwise Learning Algorithms with Kernels. arXiv:1502.07229 [stat.ML]

Haiqi Zhang, Guangquan Lu, Mengmeng Zhan, and Beixian Zhang. 2022. Semi-Supervised Classification of Graph Convolutional Networks with Laplacian Rank Constraints. *Neural Process. Lett.* 54, 4 (aug 2022), 2645–2656. https://doi.org/10.1007/s11063-020-10404-7

Muhan Zhang and Yixin Chen. 2017. Weisfeiler-Lehman Neural Machine for Link Prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) *(KDD '17)*. Association for Computing Machinery, New York, NY, USA, 575–583. https://doi.org/10.1145/3097983.3097996

Muhan Zhang and Yixin Chen. 2018. Link Prediction Based on Graph Neural Networks. arXiv:1802.09691 [cs.LG]

Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, and Meng Jiang. 2022. Learning from Counterfactual Links for Link Prediction. arXiv:2106.02172 [cs.LG]