

A Additional Discussions and Details

A.1 Limitations

One potential limitation of DyGFormer lies in the ignorance of high-order relationships between nodes since it solely learns from the first-hop interactions of nodes. In certain scenarios where nodes' high-order relationships are essential, DyGFormer may be suboptimal compared with baselines that learn the higher-order interactions. However, trivially feeding the multi-hop neighbors of nodes into DyGFormer would incur expensive computational costs. It is promising to design more efficient and effective frameworks to model nodes' high-order relationships for dynamic graph learning.

Another potential limitation is the sensitivity of the neighbor co-occurrence encoding scheme against different negative sampling strategies (discussed in Section 5.7). When the assumption of our neighbor co-occurrence encoding scheme is violated, its performance may drop drastically in some cases. It is an insightful direction to design more robust encoding schemes to tackle this issue.

A.2 Licenses

All the used codes and datasets are publicly available and permit usage for research purposes under either MIT License or Apache License 2.0.

A.3 Overall Procedure of DyGLib

Figure 3 shows the overall procedure of DyGLib.

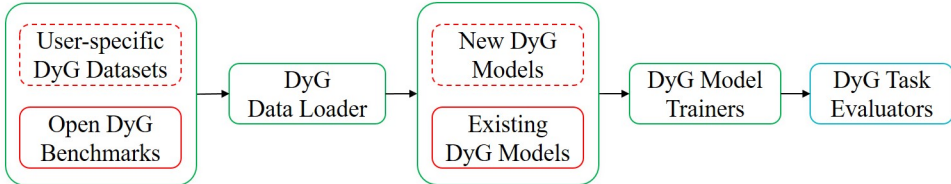


Figure 3: DyGLib is equipped with **standard training pipelines**, **extensible coding interfaces**, and **comprehensive evaluating protocols**. DyG denotes the abbreviation of Dynamic Graph.

B Detailed Experimental Settings

B.1 Descriptions of Datasets

We use thirteen datasets collected by [44] in the experiments, which are publicly available²:

- **Wikipedia** is a bipartite interaction graph that contains the edits on Wikipedia pages over a month. Nodes represent users and pages, and links denote the editing behaviors with timestamps. Each link is associated with a 172-dimensional Linguistic Inquiry and Word Count (LIWC) feature [42]. This dataset additionally contains dynamic labels that indicate whether users are temporarily banned from editing.
- **Reddit** is bipartite and records the posts of users under subreddits during one month. Users and subreddits are nodes, and links are the timestamped posting requests. Each link has a 172-dimensional LIWC feature. This dataset also includes dynamic labels representing whether users are banned from posting.
- **MOOC** is a bipartite interaction network of online sources, where nodes are students and course content units (e.g., videos and problem sets). Each link denotes a student's access behavior to a specific content unit and is assigned with a 4-dimensional feature.
- **LastFM** is bipartite and consists of the information about which songs were listened to by which users over one month. Users and songs are nodes, and links denote the listening behaviors of users.

²<https://zenodo.org/record/7213796#.Y1c06y8r30o>

- **Enron** records the email communications between employees of the ENRON energy corporation over three years.
- **Social Evo.** is a mobile phone proximity network that monitors the daily activities of an entire undergraduate dormitory for a period of eight months, where each link has a 2-dimensional feature.
- **UCI** is an online communication network, where nodes are university students and links are messages posted by students.
- **Flights** is a dynamic flight network that displays the development of air traffic during the COVID-19 pandemic. Airports are represented by nodes and the tracked flights are denoted as links. Each link is associated with a weight, indicating the number of flights between two airports in a day.
- **Can. Parl.** is a dynamic political network that records the interactions between Canadian Members of Parliament (MPs) from 2006 to 2019. Each node represents an MP from an electoral district and a link is created when two MPs both vote "yes" on a bill. The weight of each link refers to the number of times that one MP voted "yes" for another MP in a year.
- **US Legis.** is a senate co-sponsorship network that tracks social interactions between legislators in the US Senate. The weight of each link specifies the number of times two congresspersons have co-sponsored a bill in a given congress.
- **UN Trade** contains the food and agriculture trade between 181 nations for more than 30 years. The weight of each link indicates the total sum of normalized agriculture import or export values between two particular countries.
- **UN Vote** records roll-call votes in the United Nations General Assembly. If two nations both voted "yes" to an item, the weight of the link between them is increased by one.
- **Contact** describes how the physical proximity evolves among about 700 university students over a month. Each student has a unique identifier and links denote that they are within close proximity to each other. Each link is associated with a weight, revealing the physical proximity between students.

We show the statistics of the datasets in Table 6, where #N&L Feat stands for the dimensions of node and link features. We notice a slight difference between the statistics of the Contact dataset reported in [44] (which has 694 nodes and 2,426,280 links) and our own calculations, although both of them are computed based on the released dataset by [44]. We ultimately report our statistics of the Contact dataset in this paper.

Table 6: Statistics of the datasets.

Datasets	Domains	#Nodes	#Links	#N&L Feat	Bipartite	Duration	Unique Steps	Time Granularity
Wikipedia	Social	9,227	157,474	- & 172	True	1 month	152,757	Unix timestamps
Reddit	Social	10,984	672,447	- & 172	True	1 month	669,065	Unix timestamps
MOOC	Interaction	7,144	411,749	- & 4	True	17 months	345,600	Unix timestamps
LastFM	Interaction	1,980	1,293,103	- & -	True	1 month	1,283,614	Unix timestamps
Enron	Social	184	125,235	- & -	False	3 years	22,632	Unix timestamps
Social Evo.	Proximity	74	2,099,519	- & 2	False	8 months	565,932	Unix timestamps
UCI	Social	1,899	59,835	- & -	False	196 days	58,911	Unix timestamps
Flights	Transport	13,169	1,927,145	- & 1	False	4 months	122	days
Can. Parl.	Politics	734	74,478	- & 1	False	14 years	14	years
US Legis.	Politics	225	60,396	- & 1	False	12 congresses	12	congresses
UN Trade	Economics	255	507,497	- & 1	False	32 years	32	years
UN Vote	Politics	201	1,035,742	- & 1	False	72 years	72	years
Contact	Proximity	692	2,426,279	- & 1	False	1 month	8,064	5 minutes

B.2 Descriptions of Baselines

We select the following eight baselines:

- **JODIE** is designed for temporal bipartite networks of user-item interactions. It employs two coupled recurrent neural networks to update the states of users and items. A projection operation is introduced to learn the future representation trajectory of each user/item [28].

- **DyRep** proposes a recurrent architecture to update node states upon each interaction. It also includes a temporal-attentive aggregation module to consider the temporally evolving structural information in dynamic graphs [55].
- **TGAT** computes the node representation by aggregating features from each node’s temporal-topological neighbors based on the self-attention mechanism. It is also equipped with a time encoding function for capturing temporal patterns [64].
- **TGN** maintains an evolving memory for each node and updates this memory when the node is observed in an interaction, which is achieved by the message function, message aggregator, and memory updater. An embedding module is leveraged to generate the temporal representations of nodes [45].
- **CAWN** first extracts multiple causal anonymous walks for each node, which can explore the causality of network dynamics and generate relative node identities. Then, it utilizes recurrent neural networks to encode each walk and aggregates these walks to obtain the final node representation [60].
- **EdgeBank** is a pure memory-based approach without trainable parameters for transductive dynamic link prediction. It stores the observed interactions in the memory unit and updates the memory through various strategies. An interaction will be predicted as positive if it was retained in the memory, and negative otherwise [44]. The publication presents two updating strategies, but their official implementations include two more³. To be specific, the four variants of EdgeBank are: EdgeBank_∞ with unlimited memory to store all the observed edges; EdgeBank_{tw-ts} and EdgeBank_{tw-re}, which only remember edges within a fixed-size time window from the immediate past. The window size of EdgeBank_{tw-ts} is set to the duration of the test split, while EdgeBank_{tw-re} makes it related to the time intervals of repeated edges; EdgeBank_{th} that retains edges with appearing counts higher than a threshold. We experiment with all four variants and report the best performance among them.
- **TCL** first generates each node’s interaction sequence by performing a breadth-first search algorithm on the temporal dependency interaction sub-graph. Then, it presents a graph transformer that considers both graph topology and temporal information to learn node representations. It also incorporates a cross-attention operation for modeling the inter-dependencies between two interaction nodes [57].
- **GraphMixer** shows that a fixed time encoding function performs better than the trainable version. It incorporates the fixed function into a link encoder based on MLP-Mixer [54] to learn from temporal links. A node encoder with neighbor mean-pooing is employed to summarize node features [12].

B.3 Some Problematic Implementations in Baselines

JODIE, DyRep, and TGN models are based on memory networks, and their implementations have designed the raw messages to avoid information leakage. However, they fail to store the raw messages when saving models because the raw messages are maintained in a dictionary and thus cannot be saved as model parameters⁴. Our DyGLib has addressed this issue by additionally saving the correct raw messages when saving models. In the official implementation of CAWN, the encoding of each causal anonymous walk is represented by the embedding at the last position of the walk⁵. However, some walks are padded to support mini-batch training in practice, making the last position of these padded walks meaningless. To get the correct encoding, it is necessary to use the actual length of each walk. Moreover, there are issues with the `nodeedge2idx` dictionary computed by the `get_ts2idx` function⁶ if multiple interactions simultaneously occur at the last timestamp. This would lead to information leakage since the model can potentially access more interactions for a given interaction even if they happen at the same time. Our DyGLib has fixed these problems as well.

³https://github.com/fpour/DGB/blob/main/EdgeBank/link_pred/edge_bank_baseline.py

⁴https://github.com/twitter-research/tgn/blob/2aa295a1f182137a6ad56328b43cb3d8223cae77/train_self_supervised.py#L302

⁵<https://github.com/snap-stanford/CAW/blob/f994ff2b2c29778e6250b6a9928fd9943e0163f7/module.py#L1069>

⁶<https://github.com/snap-stanford/CAW/blob/f994ff2b2c29778e6250b6a9928fd9943e0163f7/graph.py#L79>

B.4 Some Inconsistent Observations with Previous Reports

In the experiments, we find the behaviors of baselines are inconsistent with their previous reports in some cases. We provide detailed illustrations and attribute these phenomena to the following reasons.

Suboptimal Settings of Hyperparameters. Some important hyperparameters in the baselines are not sufficiently fine-tuned, such as the dropout rate, the number of sampled neighbors, the number of random walks, the length of input sequences, and the neighbor sampling strategies. In this paper, we perform the grid search to find the best settings of these hyperparameters and observe that the performance of many baselines can be significantly improved by properly setting certain hyperparameters. Take TGAT for transductive dynamic link prediction with the random negative sampling strategy as an example (see Table 1). Compared with the performance in [44], the results of AP are significantly improved on datasets like MOOC (from 0.61 to 0.86), LastFM (from 0.50 to 0.73), Enron (from 0.59 to 0.71), Social Evo. (from 0.76 to 0.93), Flights (from 0.89 to 0.94), and Contact (from 0.58 to 0.96). Similar improvements can also be found on JODIE, DyRep, and TGN on several datasets.

Usage of Problematic Implementations. Some previous studies utilize problematic implementations in their experiments (illustrated in Section B.3), and the reported results may not be rigorous. Therefore, some methods would obtain worse results after we fix the problems in their implementations. Take CAWN for transductive dynamic link prediction with the random negative sampling strategy as an example (see Table 1). Compared with the results in [44], the performance on the AP metric of CAWN drops sharply on datasets like LastFM (from 0.98 to 0.87), Can. Parl. (from 0.94 to 0.70), US Legis. (from 0.97 to 0.71), UN Trade (from 0.97 to 0.65), and UN Vote (from 0.82 to 0.53). This is because [44] uses the problematic implementations of CAWN to conduct experiments and the results will sometimes become worse after fixing the issues.

Adaptions for Evaluations. There also exist some differences between GraphMixer’s results in the original paper and our work because we modify its implementation to fit our evaluations. The original GraphMixer could only be evaluated for transductive dynamic link prediction. In this work, we remove the one-hot encoding of nodes in GraphMixer to adapt it to the inductive setting, which may lead to decreased performance in certain situations. Take the performance of GraphMixer for transductive dynamic link prediction with the random negative sampling strategy as an example (see Table 1). Compared with the results in [12], the performance of GraphMixer slightly drops on datasets like Wikipedia (from 0.9985 to 0.9725) and Reddit (from 0.9993 to 0.9731).

B.5 Configurations of Different Methods

We first present the official settings of baselines as well as the configurations of DyGFormer. These configurations remain unchanged across all the datasets.

- **JODIE:**
 - Dimension of time encoding: 100
 - Dimension of node memory: 172
 - Dimension of output representation: 172
 - Memory updater: vanilla recurrent neural network
- **DyRep:**
 - Dimension of time encoding: 100
 - Dimension of node memory: 172
 - Dimension of output representation: 172
 - Number of graph attention heads: 2
 - Number of graph convolution layers: 1
 - Memory updater: vanilla recurrent neural network
- **TGAT:**
 - Dimension of time encoding: 100
 - Dimension of output representation: 172
 - Number of graph attention heads: 2

- Number of graph convolution layers: 2
- **TGN:**
 - Dimension of time encoding: 100
 - Dimension of node memory: 172
 - Dimension of output representation: 172
 - Number of graph attention heads: 2
 - Number of graph convolution layers: 1
 - Memory updater: gated recurrent unit [10]
- **CAWN:**
 - Dimension of time encoding: 100
 - Dimension of position encoding: 172
 - Dimension of output representation: 172
 - Number of attention heads for encoding walks: 8
 - Length of each walk (including the target node): 2
 - Time scaling factor α : 1e-6
- **TCL:**
 - Dimension of time encoding: 100
 - Dimension of depth encoding: 172
 - Dimension of output representation: 172
 - Number of attention heads: 2
 - Number of Transformer layers: 2
- **GraphMixer:**
 - Dimension of time encoding: 100
 - Dimension of output representation: 172
 - Number of MLP-Mixer layers: 2
 - Time gap T : 2000
- **DyGFormer:**
 - Dimension of time encoding d_T : 100
 - Dimension of neighbor co-occurrence encoding d_C : 50
 - Dimension of aligned encoding d : 50
 - Dimension of output representation d_{out} : 172
 - Number of attention heads I : 2
 - Number of Transformer layers L : 2

Then, we perform the grid search to find the best settings of some critical hyperparameters, where the searched ranges and related methods are shown in Table 7. It is worth noticing that DyGFormer can directly handle nodes with sequence lengths shorter than the defined length. When the sequence length exceeds the specified length, we select the most recent interactions up to the defined length.

Finally, we show the hyperparameter settings of various methods that are determined by the grid search in Table 8, Table 9, Table 10 and Table 11.

C Detailed Experimental Results

C.1 Additional Results for Transductive Dynamic Link Prediction

We show the AUC-ROC for transductive dynamic link prediction with three negative sampling strategies in Table 12.

C.2 Additional Results for Inductive Dynamic Link Prediction

We present the AP and AUC-ROC for inductive dynamic link prediction with three negative sampling strategies in Table 13 and Table 14 .

Table 7: Searched ranges of hyperparameters and the related methods.

Hyperparameters	Searched Ranges	Related Methods
Dropout Rate [52]	[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6]	JODIE, DyRep, TGAT, TGN, CAWN, TCL, GraphMixer, DyGFormer
Number of Sampled Neighbors	[10, 20, 30]	DyRep, TGAT, TGN, TCL, GraphMixer
Neighbor Sampling Strategies	[uniform, recent]	DyRep, TGAT, TGN, TCL, GraphMixer
Number of Causal Anonymous Walks	[16, 32, 64, 128]	CAWN
Memory Updating Variants	[EdgeBank _∞ , EdgeBank _{tw-ts} , EdgeBank _{tw-re} , EdgeBank _{th}]	EdgeBank
Length of Input Sequences & Patch Size	[32 & 1, 64 & 2, 128 & 4, 256 & 8, 512 & 16, 1024 & 32, 2048 & 64, 4096 & 128]	DyGFormer

Table 8: Configurations of the dropout rate of different methods.

Datasets	JODIE	DyRep	TGAT	TGN	CAWN	TCL	GraphMixer	DyGFormer
Wikipedia	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.1
Reddit	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.2
MOOC	0.2	0.0	0.1	0.2	0.1	0.1	0.4	0.1
LastFM	0.3	0.0	0.1	0.3	0.1	0.1	0.0	0.1
Enron	0.1	0.0	0.2	0.0	0.1	0.1	0.5	0.0
Social Evo.	0.1	0.1	0.1	0.0	0.1	0.0	0.3	0.1
UCI	0.4	0.0	0.1	0.1	0.1	0.0	0.4	0.1
Flights	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1
Can. Parl.	0.0	0.0	0.2	0.3	0.0	0.2	0.2	0.1
US Legis.	0.2	0.0	0.1	0.1	0.1	0.3	0.4	0.0
UN Trade	0.4	0.1	0.1	0.2	0.1	0.0	0.1	0.0
UN Vote	0.1	0.1	0.2	0.1	0.1	0.0	0.0	0.2
Contact	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.0

Table 9: Configurations of the number of sampled neighbors, the number of causal anonymous walks, and the length of input sequences & the patch size of different methods.

Datasets	DyRep	TGAT	TGN	CAWN	TCL	GraphMixer	DyGFormer
Wikipedia	10	20	10	32	20	30	32 & 1
Reddit	10	20	10	32	20	10	64 & 2
MOOC	10	20	10	64	20	20	256 & 8
LastFM	10	20	10	128	20	10	512 & 16
Enron	10	20	10	32	20	20	256 & 8
Social Evo.	10	20	10	64	20	20	32 & 1
UCI	10	20	10	64	20	20	32 & 1
Flights	10	20	10	64	20	20	256 & 8
Can. Parl.	10	20	10	128	20	20	2048 & 64
US Legis.	10	20	10	32	20	20	256 & 8
UN Trade	10	20	10	64	20	20	256 & 8
UN Vote	10	20	10	64	20	20	128 & 4
Contact	10	20	10	64	20	20	32 & 1

Table 10: Configurations of neighbor sampling strategies of different methods.

Datasets	DyRep	TGAT	TGN	TCL	GraphMixer
Wikipedia	recent	recent	recent	recent	recent
Reddit	recent	uniform	recent	uniform	recent
MOOC	recent	recent	recent	recent	recent
LastFM	recent	recent	recent	recent	recent
Enron	recent	recent	recent	recent	recent
Social Evo.	recent	recent	recent	recent	recent
UCI	recent	recent	recent	recent	recent
Flights	recent	recent	recent	recent	recent
Can. Parl.	uniform	uniform	uniform	uniform	uniform
US Legis.	recent	recent	recent	uniform	recent
UN Trade	recent	uniform	recent	uniform	uniform
UN Vote	recent	recent	uniform	uniform	uniform
Contact	recent	recent	recent	recent	recent

Table 11: Configurations of the variants of EdgeBank with three negative sampling strategies.

Datasets	Random	Historical	Inductive
Wikipedia	EdgeBank _∞	EdgeBank _{th}	EdgeBank _{th}
Reddit	EdgeBank _∞	EdgeBank _{th}	EdgeBank _{th}
MOOC	EdgeBank _{tw-ts}	EdgeBank _{tw-re}	EdgeBank _{th}
LastFM	EdgeBank _{tw-ts}	EdgeBank _{tw-re}	EdgeBank _{th}
Enron	EdgeBank _{tw-ts}	EdgeBank _{tw-re}	EdgeBank _{th}
Social Evo.	EdgeBank _{th}	EdgeBank _{th}	EdgeBank _{th}
UCI	EdgeBank _∞	EdgeBank _{tw-ts}	EdgeBank _{tw-re}
Flights	EdgeBank _∞	EdgeBank _{th}	EdgeBank _{th}
Can. Parl.	EdgeBank _{tw-ts}	EdgeBank _{tw-ts}	EdgeBank _{th}
US Legis.	EdgeBank _{tw-ts}	EdgeBank _{tw-ts}	EdgeBank _{tw-ts}
UN Trade	EdgeBank _{tw-re}	EdgeBank _{tw-re}	EdgeBank _{th}
UN Vote	EdgeBank _{tw-re}	EdgeBank _{tw-re}	EdgeBank _{tw-re}
Contact	EdgeBank _{tw-re}	EdgeBank _{tw-re}	EdgeBank _{th}

C.3 Results on Dynamic Node Classification

Table 15 shows the results of various methods for dynamic node classification on Wikipedia and Reddit (the only two datasets with dynamic labels).

C.4 Complete Results of TCL and GraphMixer with Neighbor Co-occurrence Encoding

We show the complete results of TCL and GraphMixer with our neighbor co-occurrence encoding scheme in Table 16.

C.5 Results of Training Time and Memory Usage Comparisons

Table 17 shows the comparisons of running time and memory usage of DyGFormer with and without the patching technique during the training process.

C.6 Results and Discussions of Ablation Study

We validate the effectiveness of several designs in DyGFormer via an ablation study, including the usage of **Neighbor Co-occurrence Encoding (NCoE)**, the usage of **Time Encoding (TE)**, and the **Mixing** of the sequence of **Source** node and **Destination** node (MixSD). We respectively remove these modules and denote the remaining parts as w/o NCoE, w/o TE, and w/o MixSD. We also **Separately** encode the **Neighbor Occurrence** in the source node’s or destination node’s sequence and denote this variant as w/ SepNO. We report the performance of different variants on MOOC, Social Evo., UCI, and UN Trade datasets from four domains in Figure 4.

Table 12: AUC-ROC for transductive dynamic link prediction with random, historical, and inductive negative sampling strategies.

NSS	Datasets	JODIE	DyRep	TGAT	TGN	CAWN	EdgeBank	TCL	GraphMixer	DyGFormer
rnd	Wikipedia	96.33 ± 0.07	94.37 ± 0.09	96.67 ± 0.07	98.37 ± 0.07	98.54 ± 0.04	90.78 ± 0.00	95.84 ± 0.18	96.92 ± 0.03	98.91 ± 0.02
	Reddit	98.31 ± 0.05	98.17 ± 0.05	98.47 ± 0.02	98.60 ± 0.06	99.01 ± 0.01	95.37 ± 0.00	97.42 ± 0.02	97.17 ± 0.02	99.15 ± 0.01
	MOOC	83.81 ± 2.09	85.03 ± 0.58	87.11 ± 0.19	91.21 ± 1.15	80.38 ± 0.26	60.86 ± 0.00	83.12 ± 0.18	84.01 ± 0.17	87.91 ± 0.58
	LastFM	70.49 ± 1.66	71.16 ± 1.89	71.59 ± 0.18	78.47 ± 2.94	85.92 ± 0.10	83.77 ± 0.00	64.06 ± 1.16	73.53 ± 0.12	93.05 ± 0.10
	Enron	87.96 ± 0.52	84.89 ± 3.00	68.89 ± 1.10	88.32 ± 0.99	90.45 ± 0.14	87.05 ± 0.00	75.74 ± 0.72	84.38 ± 0.21	93.33 ± 0.13
	Social Evo.	92.05 ± 0.46	90.76 ± 0.21	94.76 ± 0.16	95.39 ± 0.17	87.34 ± 0.08	81.60 ± 0.00	94.84 ± 0.17	95.23 ± 0.07	96.30 ± 0.01
	UCI	90.44 ± 0.49	68.77 ± 2.34	78.53 ± 0.74	92.03 ± 1.13	93.87 ± 0.08	77.30 ± 0.00	87.82 ± 1.36	91.81 ± 0.67	94.49 ± 0.26
	Flights	96.21 ± 1.42	95.95 ± 0.62	94.13 ± 0.17	98.22 ± 0.13	98.45 ± 0.01	90.23 ± 0.00	91.21 ± 0.02	91.13 ± 0.01	98.93 ± 0.01
	Can. Parl.	78.21 ± 0.23	73.35 ± 3.67	75.69 ± 0.78	76.99 ± 1.80	75.70 ± 3.27	64.14 ± 0.00	72.46 ± 3.23	83.17 ± 0.53	97.76 ± 0.41
	US Legis.	82.85 ± 1.07	82.28 ± 0.32	75.84 ± 1.99	83.34 ± 0.43	77.16 ± 0.39	62.57 ± 0.00	76.27 ± 0.63	76.96 ± 0.79	77.90 ± 0.58
	UN Trade	69.62 ± 0.44	67.44 ± 0.83	64.01 ± 0.12	69.10 ± 1.67	68.54 ± 0.18	66.75 ± 0.00	64.72 ± 0.05	65.52 ± 0.51	70.20 ± 1.44
	UN Vote	68.53 ± 0.95	67.18 ± 1.04	52.83 ± 1.12	69.71 ± 2.65	53.09 ± 0.22	62.97 ± 0.00	51.88 ± 0.36	52.46 ± 0.27	57.12 ± 0.62
	Contact	96.66 ± 0.89	96.48 ± 0.14	96.95 ± 0.08	97.54 ± 0.35	89.99 ± 0.34	94.34 ± 0.00	94.15 ± 0.09	93.94 ± 0.02	98.53 ± 0.01
	Avg. Rank	4.38	5.77	6.00	2.54	4.38	7.31	7.23	5.77	1.62
hist	Wikipedia	80.77 ± 0.73	77.74 ± 0.33	82.87 ± 0.22	82.74 ± 0.32	67.84 ± 0.64	77.27 ± 0.00	85.76 ± 0.46	87.68 ± 0.17	78.80 ± 1.95
	Reddit	80.52 ± 0.32	80.15 ± 0.18	79.33 ± 0.16	81.11 ± 0.19	80.27 ± 0.30	78.58 ± 0.00	76.49 ± 0.16	77.80 ± 0.12	80.54 ± 0.29
	MOOC	82.75 ± 0.83	81.06 ± 0.94	80.81 ± 0.67	88.00 ± 1.80	71.57 ± 1.07	61.90 ± 0.00	72.09 ± 0.56	76.68 ± 1.40	87.04 ± 0.35
	LastFM	75.22 ± 2.36	74.65 ± 1.98	64.27 ± 0.26	77.97 ± 3.04	67.88 ± 0.24	78.09 ± 0.00	47.24 ± 3.13	64.21 ± 0.73	78.78 ± 0.35
	Enron	75.39 ± 2.37	74.69 ± 3.55	61.85 ± 1.43	77.09 ± 2.22	65.10 ± 0.34	79.59 ± 0.00	67.95 ± 0.88	75.27 ± 1.14	76.55 ± 0.52
	Social Evo.	90.06 ± 3.15	93.12 ± 0.34	93.08 ± 0.59	94.71 ± 0.53	87.43 ± 0.15	85.81 ± 0.00	93.44 ± 0.68	94.39 ± 0.31	97.28 ± 0.07
	UCI	78.64 ± 3.50	57.91 ± 3.12	58.89 ± 1.57	77.25 ± 2.68	57.86 ± 0.15	69.56 ± 0.00	72.25 ± 3.46	77.54 ± 2.02	76.97 ± 0.24
	Flights	68.97 ± 1.87	69.43 ± 0.90	72.20 ± 0.16	68.39 ± 0.95	66.11 ± 0.71	74.64 ± 0.00	70.57 ± 0.18	70.37 ± 0.23	68.09 ± 0.43
	Can. Parl.	62.44 ± 1.11	70.16 ± 1.70	70.86 ± 0.94	73.23 ± 3.08	72.06 ± 3.94	63.04 ± 0.00	69.95 ± 3.70	79.03 ± 1.01	97.61 ± 0.40
	US Legis.	67.47 ± 6.40	91.44 ± 1.18	73.47 ± 5.25	83.53 ± 4.53	78.62 ± 7.46	67.41 ± 0.00	83.97 ± 3.71	85.17 ± 0.70	90.77 ± 1.96
	UN Trade	68.92 ± 1.40	64.36 ± 1.40	60.37 ± 0.68	63.93 ± 5.41	63.09 ± 0.74	86.61 ± 0.00	61.43 ± 1.04	63.20 ± 1.54	73.86 ± 1.13
	UN Vote	76.84 ± 1.01	74.72 ± 1.43	53.95 ± 3.15	73.40 ± 5.20	51.27 ± 0.33	89.62 ± 0.00	52.29 ± 2.39	52.61 ± 1.44	64.27 ± 1.78
	Contact	96.35 ± 0.92	96.00 ± 0.23	95.39 ± 0.43	93.76 ± 1.29	83.06 ± 0.32	92.17 ± 0.00	93.34 ± 0.19	93.14 ± 0.34	97.17 ± 0.05
	Avg. Rank	4.38	4.77	5.85	3.46	7.38	5.38	6.08	4.77	2.92
ind	Wikipedia	70.96 ± 0.78	67.36 ± 0.96	81.93 ± 0.22	80.97 ± 0.31	70.95 ± 0.95	81.73 ± 0.00	82.19 ± 0.48	84.28 ± 0.30	75.09 ± 3.70
	Reddit	83.51 ± 0.15	82.90 ± 0.31	87.13 ± 0.20	84.56 ± 0.24	88.04 ± 0.29	85.93 ± 0.00	84.67 ± 0.29	82.21 ± 0.13	86.23 ± 0.51
	MOOC	66.63 ± 2.30	63.26 ± 1.01	73.18 ± 0.33	77.44 ± 2.86	70.32 ± 1.43	48.18 ± 0.00	70.36 ± 0.37	72.45 ± 0.72	80.76 ± 0.76
	LastFM	61.32 ± 3.49	62.15 ± 2.12	63.99 ± 0.21	65.46 ± 4.27	67.92 ± 0.44	77.37 ± 0.00	46.93 ± 2.59	60.22 ± 0.32	69.25 ± 0.36
	Enron	70.92 ± 1.05	68.73 ± 1.34	60.45 ± 2.12	71.34 ± 2.46	75.17 ± 0.50	75.00 ± 0.00	67.64 ± 0.86	71.53 ± 0.85	74.07 ± 0.64
	Social Evo.	90.01 ± 3.19	93.07 ± 0.38	92.94 ± 0.61	95.24 ± 0.56	89.93 ± 0.15	87.88 ± 0.00	93.44 ± 0.72	94.22 ± 0.32	97.51 ± 0.06
	UCI	64.14 ± 1.26	54.25 ± 2.01	60.80 ± 1.01	64.11 ± 1.04	58.06 ± 0.26	58.03 ± 0.00	70.05 ± 1.86	74.59 ± 0.74	65.96 ± 1.18
	Flights	69.99 ± 3.10	71.13 ± 1.55	73.47 ± 0.18	71.63 ± 1.72	69.70 ± 0.75	81.10 ± 0.00	72.54 ± 0.19	72.21 ± 0.21	69.53 ± 1.17
	Can. Parl.	52.88 ± 0.80	63.53 ± 0.65	72.47 ± 1.18	69.57 ± 2.81	72.93 ± 1.78	61.41 ± 0.00	69.47 ± 2.12	70.52 ± 0.94	96.70 ± 0.59
	US Legis.	59.05 ± 5.52	89.44 ± 0.71	71.62 ± 5.42	78.12 ± 4.46	76.45 ± 7.02	68.66 ± 0.00	82.54 ± 3.91	84.22 ± 0.91	87.96 ± 1.80
	UN Trade	66.82 ± 1.27	65.60 ± 1.28	66.13 ± 0.78	66.37 ± 5.39	71.73 ± 0.74	74.20 ± 0.00	67.80 ± 1.21	66.53 ± 1.22	62.56 ± 1.51
	UN Vote	73.73 ± 1.61	72.80 ± 2.16	53.04 ± 2.58	72.69 ± 3.72	52.75 ± 0.90	72.85 ± 0.00	52.02 ± 1.64	51.89 ± 0.74	53.37 ± 1.26
	Contact	94.47 ± 1.08	94.23 ± 0.18	94.10 ± 0.41	91.64 ± 1.72	87.68 ± 0.24	85.87 ± 0.00	91.23 ± 0.19	90.96 ± 0.27	95.01 ± 0.15
	Avg. Rank	5.92	6.15	4.85	4.54	5.15	5.08	5.00	4.77	3.54

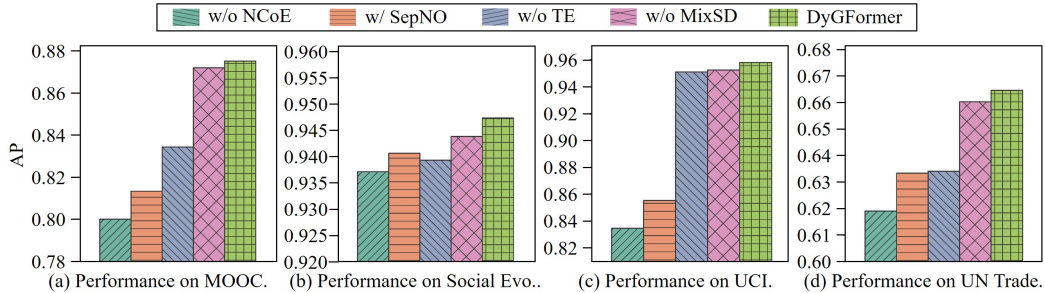


Figure 4: Ablation study of the components in DyGFormer.

Table 13: AP for inductive dynamic link prediction with random, historical, and inductive negative sampling strategies.

NSS	Datasets	JODIE	DyRep	TGAT	TGN	CAWN	TCL	GraphMixer	DyGFormer
rnd	Wikipedia	94.82 ± 0.20	92.43 ± 0.37	96.22 ± 0.07	97.83 ± 0.04	98.24 ± 0.03	96.22 ± 0.17	96.65 ± 0.02	98.59 ± 0.03
	Reddit	96.50 ± 0.13	96.09 ± 0.11	97.09 ± 0.04	97.50 ± 0.07	<u>98.62 ± 0.01</u>	94.09 ± 0.07	95.26 ± 0.02	98.84 ± 0.02
	MOOC	79.63 ± 1.92	81.07 ± 0.44	85.50 ± 0.19	89.04 ± 1.17	81.42 ± 0.24	80.60 ± 0.22	81.41 ± 0.21	86.96 ± 0.43
	LastFM	81.61 ± 3.82	83.02 ± 1.48	78.63 ± 0.31	81.45 ± 4.29	89.42 ± 0.07	73.53 ± 1.66	82.11 ± 0.42	94.23 ± 0.09
	Enron	80.72 ± 1.39	74.55 ± 3.95	67.05 ± 1.51	77.94 ± 1.02	<u>86.35 ± 0.51</u>	76.14 ± 0.79	75.88 ± 0.48	89.76 ± 0.34
	Social Evo.	91.96 ± 0.48	90.04 ± 0.47	91.41 ± 0.16	90.77 ± 0.86	79.94 ± 0.18	91.55 ± 0.09	91.86 ± 0.06	93.14 ± 0.04
	UCI	79.86 ± 1.48	57.48 ± 1.87	79.54 ± 0.48	88.12 ± 2.05	<u>92.73 ± 0.06</u>	87.36 ± 2.03	91.19 ± 0.42	94.54 ± 0.12
	Flights	94.74 ± 0.37	92.88 ± 0.73	88.73 ± 0.33	95.03 ± 0.60	<u>97.06 ± 0.02</u>	83.41 ± 0.07	83.03 ± 0.05	97.79 ± 0.02
	Can. Parl.	53.92 ± 0.94	54.02 ± 0.76	55.18 ± 0.79	54.10 ± 0.93	55.80 ± 0.69	54.30 ± 0.66	55.91 ± 0.82	87.74 ± 0.71
	US Legis.	54.93 ± 2.29	57.28 ± 0.71	51.00 ± 3.11	58.63 ± 0.37	53.17 ± 1.20	52.59 ± 0.97	50.71 ± 0.76	54.28 ± 2.87
	UN Trade	59.65 ± 0.77	57.02 ± 0.69	61.03 ± 0.18	58.31 ± 3.15	65.24 ± 0.21	62.21 ± 0.12	62.17 ± 0.31	<u>64.55 ± 0.62</u>
	UN Vote	56.64 ± 0.96	54.62 ± 2.22	52.24 ± 1.46	58.85 ± 2.51	49.94 ± 0.45	51.60 ± 0.97	50.68 ± 0.44	55.93 ± 0.39
	Contact	94.34 ± 1.45	92.18 ± 0.41	<u>95.87 ± 0.11</u>	93.82 ± 0.99	89.55 ± 0.30	91.11 ± 0.12	90.59 ± 0.05	98.03 ± 0.02
	Avg. Rank	4.69	5.77	5.23	3.77	3.77	5.77	5.23	1.54
hist	Wikipedia	68.69 ± 0.39	62.18 ± 1.27	84.17 ± 0.22	81.76 ± 0.32	67.27 ± 1.63	82.20 ± 2.18	87.60 ± 0.30	71.42 ± 4.43
	Reddit	62.34 ± 0.54	61.60 ± 0.72	63.47 ± 0.36	64.85 ± 0.85	63.67 ± 0.41	60.83 ± 0.25	64.50 ± 0.26	65.37 ± 0.60
	MOOC	63.22 ± 1.55	62.93 ± 1.24	76.73 ± 0.29	<u>77.07 ± 3.41</u>	74.68 ± 0.68	74.27 ± 0.53	74.00 ± 0.97	80.82 ± 0.30
	LastFM	70.39 ± 4.31	71.45 ± 1.76	76.27 ± 0.25	66.65 ± 6.11	71.33 ± 0.47	65.78 ± 0.65	76.42 ± 0.22	76.35 ± 0.52
	Enron	65.86 ± 3.71	62.08 ± 2.27	61.40 ± 1.31	62.91 ± 1.16	60.70 ± 0.36	67.11 ± 0.62	72.37 ± 1.37	67.07 ± 0.62
	Social Evo.	88.51 ± 0.87	88.72 ± 1.10	93.97 ± 0.54	90.66 ± 1.62	79.83 ± 0.38	<u>94.10 ± 0.31</u>	94.01 ± 0.47	96.82 ± 0.16
	UCI	63.11 ± 2.27	52.47 ± 2.06	70.52 ± 0.93	70.78 ± 0.78	64.54 ± 0.47	76.71 ± 1.00	81.66 ± 0.49	72.13 ± 1.87
	Flights	61.01 ± 1.65	62.83 ± 1.31	64.72 ± 0.36	59.31 ± 1.43	56.82 ± 0.57	64.50 ± 0.25	65.28 ± 0.24	57.11 ± 0.21
	Can. Parl.	52.60 ± 0.88	52.28 ± 0.31	<u>56.72 ± 0.47</u>	54.42 ± 0.77	<u>57.14 ± 0.07</u>	55.71 ± 0.74	55.84 ± 0.73	87.40 ± 0.85
	US Legis.	52.94 ± 2.11	62.10 ± 1.41	51.83 ± 3.95	61.18 ± 1.10	55.56 ± 1.71	53.87 ± 1.41	52.03 ± 1.02	56.31 ± 3.46
	UN Trade	55.46 ± 1.19	55.49 ± 0.84	55.28 ± 0.71	52.80 ± 3.19	55.00 ± 0.38	55.76 ± 1.03	54.94 ± 0.97	53.20 ± 1.07
	UN Vote	61.04 ± 1.30	60.22 ± 1.78	53.05 ± 3.10	63.74 ± 3.00	47.98 ± 0.84	54.19 ± 2.17	48.09 ± 0.43	52.63 ± 1.26
	Contact	90.42 ± 2.34	89.22 ± 0.66	94.15 ± 0.45	88.13 ± 1.50	74.20 ± 0.80	90.44 ± 0.17	89.91 ± 0.36	<u>93.56 ± 0.52</u>
	Avg. Rank	5.38	5.46	4.00	4.54	5.92	3.92	3.54	3.23
ind	Wikipedia	68.70 ± 0.39	62.19 ± 1.28	84.17 ± 0.22	81.77 ± 0.32	67.24 ± 1.63	82.20 ± 2.18	87.60 ± 0.29	71.42 ± 4.43
	Reddit	62.32 ± 0.54	61.58 ± 0.72	63.40 ± 0.36	64.84 ± 0.84	63.65 ± 0.41	60.81 ± 0.26	64.49 ± 0.25	65.35 ± 0.60
	MOOC	63.22 ± 1.55	62.92 ± 1.24	76.72 ± 0.30	<u>77.07 ± 3.40</u>	74.69 ± 0.68	74.28 ± 0.53	73.99 ± 0.97	80.82 ± 0.30
	LastFM	70.39 ± 4.31	71.45 ± 1.75	76.28 ± 0.25	69.46 ± 4.65	71.33 ± 0.47	65.78 ± 0.65	76.42 ± 0.22	76.35 ± 0.52
	Enron	65.86 ± 3.71	62.08 ± 2.27	61.40 ± 1.30	62.90 ± 1.16	60.72 ± 0.36	67.11 ± 0.62	72.37 ± 1.38	67.07 ± 0.62
	Social Evo.	88.51 ± 0.87	88.72 ± 1.10	93.97 ± 0.54	90.65 ± 1.62	79.83 ± 0.39	<u>94.10 ± 0.32</u>	94.01 ± 0.47	96.82 ± 0.17
	UCI	63.16 ± 2.27	52.47 ± 2.09	70.49 ± 0.93	70.73 ± 0.79	64.54 ± 0.47	<u>76.65 ± 0.99</u>	81.64 ± 0.49	72.13 ± 1.86
	Flights	61.01 ± 1.66	62.83 ± 1.31	64.72 ± 0.37	59.32 ± 1.45	56.82 ± 0.56	64.50 ± 0.25	65.29 ± 0.24	57.11 ± 0.20
	Can. Parl.	52.58 ± 0.86	52.24 ± 0.28	56.46 ± 0.50	54.18 ± 0.73	57.06 ± 0.08	55.46 ± 0.69	55.76 ± 0.65	87.22 ± 0.82
	US Legis.	52.94 ± 2.11	62.10 ± 1.41	51.83 ± 3.95	61.18 ± 1.10	55.56 ± 1.71	53.87 ± 1.41	52.03 ± 1.02	56.31 ± 3.46
	UN Trade	55.43 ± 1.20	55.42 ± 0.87	55.58 ± 0.68	52.80 ± 3.24	54.97 ± 0.38	55.66 ± 0.98	54.88 ± 1.01	52.56 ± 1.70
	UN Vote	61.17 ± 1.33	60.29 ± 1.79	53.08 ± 3.10	63.71 ± 2.97	48.01 ± 0.82	54.13 ± 2.16	48.10 ± 0.40	52.61 ± 1.25
	Contact	90.43 ± 2.33	89.22 ± 0.65	94.14 ± 0.45	88.12 ± 1.50	74.19 ± 0.81	90.43 ± 0.17	89.91 ± 0.36	<u>93.55 ± 0.52</u>
	Avg. Rank	5.31	5.54	3.85	4.38	5.85	3.92	3.46	3.31

Our findings from Figure 4 indicate that DyGFormer achieves optimal performance when all components are utilized. The removal of any component would lead to worse results. In particular, the neighbor co-occurrence encoding scheme has the most significant impact on the performance as it effectively captures correlations between nodes. Separately encoding neighbor occurrences or encoding the temporal information could also improve performance. Mixing the sequences of the source node and destination node causes relatively minor improvements due to the usage of the neighbor co-occurrence encoding scheme because both of them aim to explore the node correlations. To this end, the necessity of designing these components has been well demonstrated.

C.7 Results on Temporal Graph Benchmark

We also evaluate DyGFormer and baselines on the recently proposed Temporal Graph Benchmark (TGB) [23], which contains a collection of challenging and diverse benchmark datasets. We aim to verify the superiority and scalability of DyGFormer on TGB since it covers small, medium, and large datasets. We only present the analysis here. Please refer to [23] and our other work [68] for details of the datasets, tasks, evaluation metrics, experimental settings, and complete quantitative results.

Superiority of DyGFormer. For the dynamic link property prediction task, DyGFormer ranks first on the TGB leaderboard⁷ on tgbl-wiki-v2. It ranks second/third on tgbl-coin-v2/tgbl-comment and does medium on tgbl-review-v2. We notice that the surprise index (defined as the ratio of test links that are not seen during training) on tgbl-wiki-v2 and tgbl-coin-v2 are 0.108 and 0.120, which indicates nodes tend to interact repeatedly. Such a property may be well handled by the neighbor co-occurrence

⁷https://tgb.complexdatalab.com/docs/leader_linkprop/

Table 14: AUC-ROC for inductive dynamic link prediction with random, historical, and inductive negative sampling strategies.

NSS	Datasets	JODIE	DyRep	TGAT	TGN	CAWN	TCL	GraphMixer	DyGFormer
rnd	Wikipedia	94.33 ± 0.27	91.49 ± 0.45	95.90 ± 0.09	97.72 ± 0.03	98.03 ± 0.04	95.57 ± 0.20	96.30 ± 0.04	98.48 ± 0.03
	Reddit	96.52 ± 0.13	96.05 ± 0.12	96.98 ± 0.04	97.39 ± 0.07	98.42 ± 0.02	93.80 ± 0.07	94.97 ± 0.05	98.71 ± 0.01
	MOOC	83.16 ± 1.30	84.03 ± 0.49	86.84 ± 0.17	91.24 ± 0.99	81.86 ± 0.25	81.43 ± 0.19	82.77 ± 0.24	87.62 ± 0.51
	LastFM	81.13 ± 3.39	82.24 ± 1.51	76.99 ± 0.29	82.61 ± 3.15	87.82 ± 0.12	70.84 ± 0.85	80.37 ± 0.18	94.08 ± 0.08
	Enron	81.96 ± 1.34	76.34 ± 4.20	64.63 ± 1.74	78.83 ± 1.11	87.02 ± 0.50	72.33 ± 0.99	76.51 ± 0.71	90.69 ± 0.26
	Social Evo.	93.70 ± 0.29	91.18 ± 0.49	93.41 ± 0.19	93.43 ± 0.59	84.73 ± 0.27	93.71 ± 0.18	94.09 ± 0.07	95.29 ± 0.03
	UCI	78.80 ± 0.94	58.08 ± 1.81	77.64 ± 0.38	86.68 ± 2.29	90.40 ± 0.11	84.49 ± 1.82	89.30 ± 0.57	92.63 ± 0.13
	Flights	95.21 ± 0.32	93.56 ± 0.70	88.64 ± 0.35	95.92 ± 0.43	96.86 ± 0.02	82.48 ± 0.01	82.27 ± 0.06	97.80 ± 0.02
	Can. Parl.	53.81 ± 1.14	55.27 ± 0.49	56.51 ± 0.75	55.86 ± 0.75	58.83 ± 1.13	55.83 ± 1.07	58.32 ± 1.08	89.33 ± 0.48
	US Legis.	58.12 ± 2.35	61.07 ± 0.56	48.27 ± 3.50	62.38 ± 0.48	51.49 ± 1.13	50.43 ± 1.48	47.20 ± 0.89	53.21 ± 3.04
	UN Trade	62.28 ± 0.50	58.82 ± 0.98	62.72 ± 0.12	59.99 ± 3.50	67.05 ± 0.21	63.76 ± 0.07	63.48 ± 0.37	67.25 ± 1.05
	UN Vote	58.13 ± 1.43	55.13 ± 3.46	51.83 ± 1.35	61.23 ± 2.71	48.34 ± 0.76	50.51 ± 1.05	50.04 ± 0.86	56.73 ± 0.69
	Contact	95.37 ± 0.92	91.89 ± 0.38	96.53 ± 0.10	94.84 ± 0.75	89.07 ± 0.34	93.05 ± 0.09	92.83 ± 0.05	98.30 ± 0.02
	Avg. Rank	4.69	5.85	5.31	3.38	4.00	6.00	5.31	1.46
hist	Wikipedia	61.86 ± 0.53	57.54 ± 1.09	78.38 ± 0.20	75.75 ± 0.29	62.04 ± 0.65	79.79 ± 0.96	82.87 ± 0.21	68.33 ± 2.82
	Reddit	61.69 ± 0.39	60.45 ± 0.37	64.43 ± 0.27	64.55 ± 0.50	64.94 ± 0.21	61.43 ± 0.26	64.27 ± 0.13	64.81 ± 0.25
	MOOC	64.48 ± 1.64	64.23 ± 1.29	74.08 ± 0.27	77.69 ± 3.55	71.68 ± 0.94	69.82 ± 0.32	72.53 ± 0.84	80.77 ± 0.63
	LastFM	68.44 ± 3.26	68.79 ± 1.08	69.89 ± 0.28	66.99 ± 5.62	67.69 ± 0.24	55.88 ± 1.85	70.07 ± 0.20	70.73 ± 0.37
	Enron	65.32 ± 3.57	61.50 ± 2.50	57.84 ± 2.18	62.68 ± 1.09	62.25 ± 0.40	64.06 ± 1.02	68.20 ± 1.62	65.78 ± 0.42
	Social Evo.	88.53 ± 0.55	87.93 ± 1.05	91.87 ± 0.72	92.10 ± 1.22	83.54 ± 0.24	93.28 ± 0.60	93.62 ± 0.35	96.91 ± 0.09
	UCI	60.24 ± 1.94	51.25 ± 2.37	62.32 ± 1.18	62.69 ± 0.90	56.39 ± 0.10	70.46 ± 1.94	75.98 ± 0.84	65.55 ± 1.01
	Flights	60.72 ± 1.29	61.99 ± 1.39	63.38 ± 0.26	59.66 ± 1.04	56.58 ± 0.44	63.48 ± 0.23	63.30 ± 0.19	56.05 ± 0.21
	Can. Parl.	51.62 ± 1.00	52.38 ± 0.46	58.30 ± 0.61	55.64 ± 0.54	60.11 ± 0.48	57.30 ± 1.03	56.68 ± 1.20	88.68 ± 0.74
	US Legis.	58.12 ± 2.94	67.94 ± 0.98	49.99 ± 4.88	64.87 ± 1.65	54.41 ± 1.31	52.12 ± 2.13	49.28 ± 0.86	56.57 ± 3.22
	UN Trade	58.73 ± 1.19	57.90 ± 1.33	59.74 ± 0.59	55.61 ± 3.54	60.95 ± 0.80	61.12 ± 0.97	59.88 ± 1.17	58.46 ± 1.65
	UN Vote	65.16 ± 1.28	63.98 ± 2.12	51.73 ± 4.12	68.59 ± 3.11	48.01 ± 1.77	54.66 ± 2.11	45.49 ± 0.42	53.85 ± 2.02
	Contact	90.80 ± 1.18	88.88 ± 0.68	93.76 ± 0.41	88.84 ± 1.39	74.79 ± 0.37	90.37 ± 0.16	90.04 ± 0.29	94.14 ± 0.26
	Avg. Rank	5.08	6.00	4.23	4.54	5.38	4.00	3.69	3.08
ind	Wikipedia	61.87 ± 0.53	57.54 ± 1.09	78.38 ± 0.20	75.76 ± 0.29	62.02 ± 0.65	79.79 ± 0.96	82.88 ± 0.21	68.33 ± 2.82
	Reddit	61.69 ± 0.39	60.44 ± 0.37	64.39 ± 0.27	64.55 ± 0.50	64.91 ± 0.21	61.36 ± 0.26	64.27 ± 0.13	64.80 ± 0.25
	MOOC	64.48 ± 1.64	64.22 ± 1.29	74.07 ± 0.27	77.68 ± 3.55	71.69 ± 0.94	69.83 ± 0.32	72.52 ± 0.84	80.77 ± 0.63
	LastFM	68.44 ± 3.26	68.79 ± 1.08	69.89 ± 0.28	66.99 ± 5.61	67.68 ± 0.24	55.88 ± 1.85	70.07 ± 0.20	70.73 ± 0.37
	Enron	65.32 ± 3.57	61.50 ± 2.50	57.83 ± 2.18	62.68 ± 1.09	62.27 ± 0.40	64.05 ± 1.02	68.19 ± 1.63	65.79 ± 0.42
	Social Evo.	88.53 ± 0.55	87.93 ± 1.05	91.88 ± 0.72	92.10 ± 1.22	83.54 ± 0.24	93.28 ± 0.60	93.62 ± 0.35	96.91 ± 0.09
	UCI	60.27 ± 1.94	51.26 ± 2.40	62.29 ± 1.17	62.66 ± 0.91	56.39 ± 0.11	70.42 ± 1.93	75.97 ± 0.85	65.58 ± 1.00
	Flights	60.72 ± 1.29	61.99 ± 1.39	63.40 ± 0.26	59.66 ± 1.05	56.58 ± 0.44	63.49 ± 0.23	63.32 ± 0.19	56.05 ± 0.22
	Can. Parl.	51.61 ± 0.98	52.35 ± 0.52	58.15 ± 0.62	55.43 ± 0.42	60.01 ± 0.47	56.88 ± 0.93	56.63 ± 1.09	88.51 ± 0.73
	US Legis.	58.12 ± 2.94	67.94 ± 0.98	49.99 ± 4.88	64.87 ± 1.65	54.41 ± 1.31	52.12 ± 2.13	49.28 ± 0.86	56.57 ± 3.22
	UN Trade	58.71 ± 1.20	57.87 ± 1.36	59.98 ± 0.59	55.62 ± 3.59	60.88 ± 0.79	61.01 ± 0.93	59.71 ± 1.17	57.28 ± 3.06
	UN Vote	65.29 ± 1.30	64.10 ± 2.10	51.78 ± 4.14	68.58 ± 3.08	48.04 ± 1.76	54.65 ± 2.20	45.57 ± 0.41	53.87 ± 2.01
	Contact	90.80 ± 1.18	88.87 ± 0.67	93.76 ± 0.40	88.85 ± 1.39	74.79 ± 0.38	90.37 ± 0.16	90.04 ± 0.29	94.14 ± 0.26
	Avg. Rank	5.08	5.92	4.15	4.54	5.38	4.00	3.77	3.15

Table 15: AUC-ROC for dynamic node classification.

Methods	Wikipedia	Reddit	Avg. Rank
JODIE	88.99 ± 1.05	60.37 ± 2.58	4.50
DyRep	86.39 ± 0.98	63.72 ± 1.32	5.00
TGAT	84.09 ± 1.27	70.04 ± 1.09	4.00
TGN	86.38 ± 2.34	63.27 ± 0.90	6.00
CAWN	84.88 ± 1.33	66.34 ± 1.78	5.00
TCL	77.83 ± 2.13	68.87 ± 2.15	5.00
GraphMixer	86.80 ± 0.79	64.22 ± 3.32	4.00
DyGFormer	87.44 ± 1.08	68.00 ± 1.74	2.50

Table 16: AP for different methods when equipped with the neighbor co-occurrence encoding.

Datasets	TCL			GraphMixer			DyGFormer
	Original	w/ NCoE	Improv.	Original	w/ NCoE	Improv.	
Wikipedia	96.47	99.09	2.72%	97.25	97.90	0.67%	<u>99.03</u>
Reddit	97.53	<u>99.04</u>	1.55%	97.31	97.63	0.33%	99.22
MOOC	82.38	<u>86.92</u>	5.51%	82.78	83.58	0.97%	87.52
LastFM	67.27	<u>84.02</u>	24.90%	75.61	76.48	1.15%	93.00
Enron	79.70	<u>90.18</u>	13.15%	82.25	88.83	8.00%	92.47
Social Evo.	93.13	<u>94.06</u>	1.00%	93.37	<u>94.37</u>	1.07%	94.73
UCI	89.57	<u>94.69</u>	5.72%	93.25	<u>93.48</u>	0.25%	95.79
Flights	91.23	<u>97.71</u>	7.10%	90.99	96.90	6.50%	98.91
Can. Parl.	68.67	69.34	0.98%	<u>77.04</u>	76.38	-0.86%	97.36
US Legis.	69.59	69.47	-0.17%	<u>70.74</u>	70.26	-0.68%	71.11
UN Trade	62.21	<u>63.46</u>	2.01%	62.61	62.77	0.26%	66.46
UN Vote	51.90	51.52	-0.73%	52.11	<u>52.13</u>	0.04%	55.55
Contact	92.44	<u>97.98</u>	5.99%	91.92	97.94	6.55%	98.29
Avg. Rank	4.62	<u>2.62</u>	—	3.85	2.85	—	1.08

Table 17: Comparisons of training time and memory usage of DyGFormer with and without the patching technique, where OOM stands for Out-Of-Memory.

Datasets	Input Lengths	Metrics	DyGFormer w/ patching	DyGFormer w/o patching	Reduced Ratios
LastFM	64	running time	13min 58s	21min 01s	1.50
		memory usage	3,945 MB	7,953 MB	2.02
	128	running time	16min 54s	45min 29s	2.69
		memory usage	4,677 MB	7,585 MB	1.62
256	running time	24min 41s	2h 5min 50s	5.10	
	memory usage	4,635 MB	14,583 MB	3.15	
512	running time	37min 04s	—	—	
	memory usage	7,547 MB	OOM	—	
Can. Parl.	64	running time	58s	1min 14s	1.28
		memory usage	2,121 MB	3,263 MB	1.54
	128	running time	1min 02s	2min 39s	2.56
		memory usage	2,369 MB	6,417 MB	2.71
256	running time	1min 26s	6min 37s	4.62	
	memory usage	2,855 MB	14,923 MB	5.23	
512	running time	1min 57s	—	—	
	memory usage	4,511 MB	OOM	—	

encoding scheme in DyGFormer, leading to excellent performance. Correspondingly, the surprise index on tgbl-review-v2 and tgbl-comment are 0.987 and 0.823, showing that these two datasets contain many new links. This characteristic may violate the motivation of our neighbor co-occurrence encoding scheme and bring suboptimal performance. For the dynamic node property prediction task, DyGFormer also performs better than other methods with trainable parameters, but its performance is still worse than the simple non-trainable Persistent Forecast and Moving Average methods on the TGB leaderboard⁸. This observation demonstrates the necessity of proposing customized models for the dynamic node property prediction task.

Scalability of DyGFormer. DyGFormer can be scaled up to larger datasets that contain hundreds of thousands of nodes and tens of millions of links (e.g., tgbl-coin-v2, tgbl-comment, tgbl-genre, and tgbl-reddit) while many baselines (e.g., JODIE, DyRep, TGN, and CAWN) fail. This demonstrates the good scalability of DyGFormer.

⁸https://tgb.complexdatalab.com/docs/leader_nodeprop/