
BENCHTEMP: A General Benchmark for Evaluating Temporal Graph Neural Networks

Authors' Response to Reviewer J4Uw - Second Round

Opportunities For Improvement:

Thank you. I have checked out the response.

W1. Two added datasets, eBay-Small and eBay-Large, may not be relevant to be included in the paper if they can only be shared after personally contacting the authors. The CFP states the following:

- A key criterion is accessibility: datasets should be available and accessible, i.e. the data can be found and obtained without a personal request to the PI.

W2. Model efficiency results are not provided for the newly added datasets.

General Response:

Dear Reviewer J4Uw:

We sincerely appreciate your feedback and valuable suggestions!

Indeed, the CFP states the accessibility of the datasets. After a discussion with our industrial partner eBay, we are working on sharing the **eBay-Small** and **eBay-Large** datasets in a way that ensures availability and justifies the research purpose:

1. We will build a website that describes the eBay datasets and provides an application form.
2. The applicants input their email and affiliation in the form, and agree to the access terms (similar to ImageNet).
3. The backend will check the applicant's information and send a download link to the corresponding email.

Note that, many large-scale datasets also adopt this routine, e.g., YFCC100M from Yahoo (<http://www.multimediacommons.org/>) and ImageNet (<https://www.image-net.org/download.php>). We hope this solution can address the reviewer's concern. We will optimize this procedure according to the reviewer's further suggestions.

In the meantime, eBay provide a Google form for the applicants to obtain the eBay datasets: <https://forms.gle/bP1RmyVJ1C6pgyS66> (**the applicants can remain anonymous**).

We have added model efficiency results for the newly added datasets. Please refer to Comment 2 in this response file for details.

Thank you and best regards!

Yours sincerely,

Qiang Huang, Jiawei Jiang, Xi Susie Rao, Ce Zhang, Zhichao Han, Zitao Zhang, Xin Wang, Quanqing Xu, Yang Zhao, Chuang Hu, Shuo Shang, Yongjun He, Bo Du

Submitted to the 37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks. Do not distribute.

26 **We provide our response to each individual comment below:**

Comment 1

W1. I thank the authors for addressing my concerns in detail. In terms of W2 (novel datasets), the authors have added four novel large-scale datasets (with subsets forming the smaller version of these datasets) along with experiments on them. These datasets can be valuable to the temporal graph learning community. I appreciate the efforts in contributing these datasets.

27

28 **Response:**

29 We thank the reviewer for the suggestions!

30 After a discussion with our industrial partner eBay, we are working on sharing the **eBay-Small** and
31 **eBay-Large** datasets in a way that ensures availability and justifies the research purpose:

- 32 1. We will build a website that describes the eBay datasets and provides an application form.
- 33 2. The applicants input their email and affiliation in the form, and agree to the access terms (similar
34 to ImageNet).
- 35 3. The backend will check the applicant's information and send a download link to the corresponding
36 email.

37 Note that, many large-scale datasets also adopt this routine, e.g., YFCC100M from Yahoo (<http://www.multimediacommons.org/>) and ImageNet (<https://www.image-net.org/download.php>). We hope this solution can address the reviewer's concern. We will optimize this procedure
40 according to the reviewer's further suggestions.

41 In the meantime, eBay provide a Google form for the applicants to obtain the eBay datasets:
42 <https://forms.gle/bP1RmyVJ1C6pgyS66> (**the applicants can remain anonymous**).

43

44

Comment 2

W2. Model efficiency results are not provided for the newly added datasets.

45

46 **Response:**

47 We appreciate your valuable suggestion! We have added model efficiency results for the newly added
48 datasets as follows. We will add all these results to the Appendix (https://openreview.net/attachment?id=rnZm2vQq31&name=supplementary_material).

50 Since many real-world graphs are extremely large, we believe efficiency is a vital issue for TGNNs in
51 practice. We thereby compare the efficiency of the evaluated models on the newly added datasets
52 (eBay-Small, eBay-Large, Taobao-Large, DGraphFin, YouTubeReddit-Small, YouTubeReddit-Large),
53 and present the results for dynamic link prediction task in Table 1, while dynamic node classification
54 task Table 2.

55 The Runtime in Table 1 and Table 2 shows that NAT is always trained much faster than the others
56 and need a low RAM and GPU Memory. TGAT obtains the second-best efficiency performance on
57 the newly added datasets. JODIE, DyRep, TGN achieve similar efficiency performance. We observe
58 similar results as the main paper, NeurTW performs poorly on model efficiency.

Table 1: Model efficiency for the newly added datasets on *the link prediction task*. We report seconds per epoch as **Runtime**, the maximum RAM usage as **RAM**, and the maximum GPU memory usage as **GPU Memory**, respectively. The best and second-best results are highlighted as **bold red** and underlined blue.

Model Dataset	Runtime (second)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	749.80	801.58	905.19	<u>61.05</u>	1,385.54	1,556.32	25.12
YouTubeReddit-Small	213.92	227.99	214.17	<u>85.59</u>	378.94	7,459.92	29.51
eBay-Large	28,203.53	30,151.18	30,286.88	<u>791.86</u>	52,116.62	58,540.48	117.38
DGraphFin	4,579.52	4,210.48	4,397.32	<u>1,708.71</u>	30,144.25	81,653.89	904.38
Youtube-Reddit-Large	4,630.49	4,935.05	4,635.91	<u>1,852.67</u>	8,202.50	161,476.80	638.77
Taobao-Large	3,108.45	2,931.87	2,860.83	<u>2,658.34</u>	12,143.02	148,922.55	6654.56
Model Dataset	RAM (GB)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	7.8	<u>6.2</u>	6.8	4.3	9.1	7.8	4.3
YouTubeReddit-Small	6.8	7.2	6.6	<u>5.3</u>	13.1	8.1	4.5
eBay-Large	20.2	18.3	19.1	<u>5.2</u>	17.1	10.1	5.5
DGraphFin	17.5	15.3	17.5	<u>8.3</u>	23.2	24.3	6.9
Youtube-Reddit-Large	26.3	16.6	18.9	<u>7.9</u>	18.5	21.3	6.3
Taobao-Large	14.3	12.1	13.4	<u>7.5</u>	18.1	20.7	6.2
Model Dataset	GPU Memory (GB)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	2.0	1.9	2.0	1.9	<u>1.8</u>	1.6	2.2
YouTubeReddit-Small	<u>1.3</u>	1.4	2.1	<u>1.3</u>	1.8	1.1	1.1
eBay-Large	29.7	24.6	30.9	5.8	<u>5.7</u>	3.0	5.9
DGraphFin	19.3	18.5	16.1	6.3	6.9	<u>6.1</u>	6.0
Youtube-Reddit-Large	22.1	23.0	23.4	7.8	6.3	7.2	<u>7.1</u>
Taobao-Large	20.3	21.8	19.6	7.7	7.3	<u>6.8</u>	5.6

Table 2: Model efficiency for the newly added datasets on *the node classification task*. We report seconds per epoch as **Runtime**, the maximum RAM usage as **RAM**, and the maximum GPU memory usage as **GPU Memory**, respectively. The best and second-best results are highlighted as **bold red** and underlined blue.

Model Dataset	Runtime (second)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	765.05	794.03	718.56	<u>55.05</u>	226.56	583.08	13.05
eBay-Large	29,153.28	29,867.17	27,028.53	<u>629.52</u>	8,522.04	25,693.71	97.54
Model Dataset	RAM (GB)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	6.5	6.8	6.7	<u>4.2</u>	6.9	7.2	4.1
eBay-Large	41.8	39.2	20.5	5.2	15.1	7.4	<u>5.8</u>
Model Dataset	GPU Memory (GB)						
	JODIE	DyRep	TGN	TGAT	CAWN	NeurTW	NAT
eBay-Small	1.8	1.2	<u>1.5</u>	1.8	1.9	1.8	2.3
eBay-Large	31.7	31	31.4	<u>5.8</u>	<u>5.8</u>	2.9	5.9