KDD 2024 Rebuttal Reviewer 8Aps

April 12, 2024

1 Reviewer 8Aps

Dear reviewer, thank you so much for your valuable comments. We address all of your concerns as follows

1.1 Q1

The performance gain is very limited (significant improvement only in "ogbn-products"). To make a practical impact, much larger improvements are required

1.2 A1

We first emphasize that our proposed methods aim to reduce staleness, which is typically significant when the batch size is small or the datasets are very large. To showcase the effectiveness of REST, we have added Tables 2, 3, and 4 in our submission to demonstrate the advantages of our model when the batch size is small. For instance, REST can improve performance from 67.8% to 71.9% when using a small batch size on ogbn-arxiv, illustrating the advantage of our proposed method.

Furthermore, we carry out experiments using ogbn-papers100m, which is significantly larger in scale compared to other datasets. We report the performance and efficiency in Table 1.

The larger size of the ogbn-papers100M dataset exacerbates the staleness issue for GAS, leading to decreased accuracy and slower convergence, as anticipated. In contrast, REST showcases both high accuracy and efficiency, in line with our primary claim in the submission.

1.3 Q2

Experiments need to be improved: (1) In Table 1, the results of some cases are missing. In Table 4, ogbn datasets are missing. For some difficult to reproduce, please specify why. (2) The effectiveness of REST may vary depending on the specific graph and embedding method used. Some large datasets are well-known for GNN training, e.g., Papers-100M, MAG240, and IGB. Please consider using more large datasets for the experiment.

Table 1: Memory usage (MB) and running time (seconds) ogbn-papers100m.

Models	**Accuracy**	MEMORY(MB)	TIME(S)
GAS	64.9	15705	8840
REST	67.3	16808	4100

1.4 A2

(1) In our experiments, we strictly adhere to the performance reported in the original papers or other existing related works. Those missing results are due to the fact that some codes are stale and not easily runnable on recent large-scale datasets. It's worth noting that this challenge of missing results is not unique to our study; all other papers also lack certain results due to similar limitations. For instance, the MVS-GCN model was published earlier than OGB, and thus, certain experiments may not have been conducted on recent large-scale datasets. However, we plan to address these gaps and include additional results after the rebuttal period, taking into consideration the time constraints.

(2) Please refer last answer for the results and analysis on ogbn-papers100M.

1.5 Q3

Tables need to be improved as well. Please color the important values in the tables, as the tables in this paper are many and varied, making them hard to analyze.

1.6 A3

Given the time constrain, we will make revisions after the rebuttal period.

1.7 Q4

In line 373, "embeddings These" to "embeddings. These"

1.8 A4

Given the time constrain, we will make revisions after the rebuttal period.