Accurate and Efficient Graph Neural Network Training on Very Large Knowledge Graphs via Meta-sampling

Hussein Abdallah*, Yasaman Sabbagh Ziarani*, Panos Kalnis§, Essam Mansour*
*Concordia University, Canada §KAUST, Saudi Arabia
{first}.{last}@concordia.ca panos.kalnis@kaust.edu.sa

PVLDB Reference Format:
Hussein Abdallah*, Yasaman Sabbagh Ziarani*, Panos Kalnis§, Essam Mansour*. Accurate and Efficient Graph Neural Network Training on Very Large Knowledge Graphs via Meta-sampling. PVLDB, 14(1): XXX-XXX, 2020.
doi:XX.XX/XXX.XX

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at URL_TO_YOUR_ARTIFACTS.

1 EXPERIMENTAL STUDY

1.1 Open KG benchmark datasets

Existing GNN datasets are not suitable for evaluating our problem as they are either homogeneous graphs datasets as in [4] or heterogeneous datasets extracted from real KGs as in [4, 5] with small size and a few numbers of node/edge types [5].

We built the open KG benchmark that contains 4 heterogenous directed KG datasets and used them for the evaluation, as shown in Table 1. These datasets were used to train 6 GNN node classification tasks and 2 GNN link prediction tasks as in table 2. DBLP26K, IMDB12K, and OGBN-MAG are existing datasets in [8, 9] with small to medium size that we use to evaluate the performance and efficiency of RGCN-based GNN and HGNN methods to decide which methods are best to integrate with our meta-sampler with.

For each node classification task, the model predicts the target class O for the subject vertex S based on ground truth data linked with the predicate P. For each link prediction, the model predicts the target link (predicate) P is exist between the subject vertex S and object vertex O based on ground truth edges inside the dataset. The dataset nodes and edges are in tens to hundred million with node/edge types in tens to hundreds which is 10x larger than the existing datasets. We followed the OGB benchmarks guidelines while building these datasets and used KG-TOSA Hetero-GNN dataset transformer component to generate the GNN dataset. more details about the datasets are in [1] and github is available at[1].

1.2 RGCN-Based GNNs V.s. HGNNs

To determine the appropriate GNN methods to integrate our meta-sampler with, we compare RGCN-based with HGNNs to choose a representative method for each class where the used methods are the state-of-the-art methods in each category.

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX
[1]https://gitfront.io/r/CODS/4JLJKBtaniud/OpenKG-Benchmark/

Table 1: Statistics of KG Datasets (n-type: node type, e-type: edge type. The number of nodes and edges is in millions. The number of node/edge types in tens to hundred.

KG-Dataset	#nodes	#edges	#n-type	#e-type
MAG-42M	42.4M	166M	58	62
DBLP-15M	15.6M	252M	42	48
YAGO-30M	30.7M	400M	104	98
YAGO3-10	123K	1.1M	-	37
OBGN-MAG	1.9MK	62M	4	4

Table 2: The GNN Tasks. The GNN task type (TT) is either node classification (NC) or link prediction (LP). Micro-F1 is the metric to evaluate the NC task and MRR@10 is the metric to evaluate the LP task. More details are in [1].

TT	KG-Dataset	S/P/O
NC	MAG-42M	**Paper**/publishedIn/Venue
NC	MAG-42M	**Paper**/Discipline/FieldOfStudy
NC	BDLP-15M	**Author**/Affiliation/Country
NC	DBLP-15M	**Publication**/publishedIn/Venue
NC	YAGO-30M	**CreativeWork**/Type/Geners
NC	YAGO-30M	**Place**/Locatedin/Countries
LP	YAGO3-10	Airport/**ConnectTo**/Airport
LP	DBLP-15M	Author/**AffiliatedWith**/Affaliation
NC	OBGN-MAG	**Paper**/PublishedIn/Venue

In previous studies, the meta-path-based HGGN methods have been evaluated using small datasets with 10K nodes and 100K edges, having 3-4 edge/node types as reported in [9, 10] which keep them not well evaluated with large heterogeneous graphs. The IMDB-12k and DBLP-26K datasets are used to compare the performance of these methods in small-size datasets. Figure 1 demonstrate that the RGCN-Based GNNs achieve close to the best accuracy while reducing the training memory significantly in both datasets.

Moving a step forward, we compare these methods using medium/large-size datasets extracted from MAG KG. Figures **??** and **??**

The ShadowSAINT, SeHGNN and GraphSAINT methods in figure 2 achieve the best accuracy for OBGN-MAG and MAG42M datasets with savings in training time and memory. To ensure fairness, we only compare KG-TOSA with the 3 best-performing methods on the MAG42M dataset.

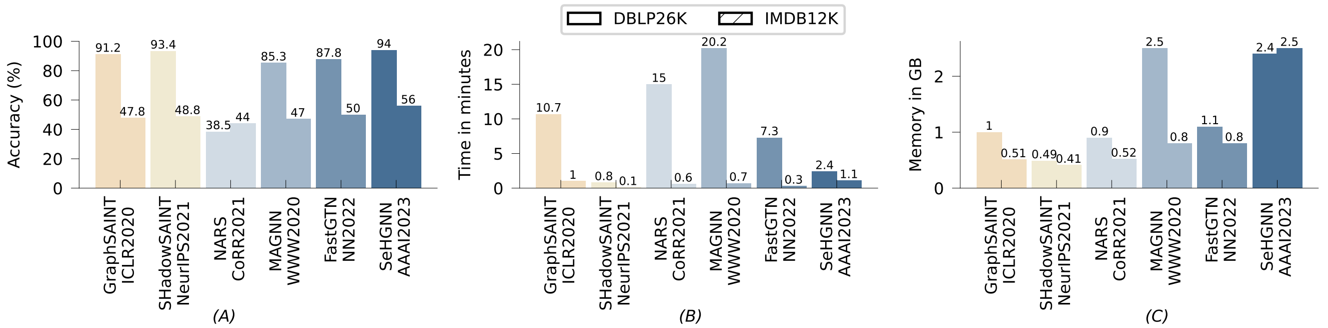

Figure 1: (A) Accuracy (higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better). GNN Node classification task using state-of-the-art RGCN-based GNNs (Graph-SAINT[12], Shadow-SAINT[11]) and Heterogeneous GNNs (HGSL[13], NARS[8], MAGNN[3], and FastGTN[10]) on DBLP26K and IMDB12K dataset. The RGCN-based GNNs are able to achieve comparable accuracy with significant savings in training time with this small-size dataset.

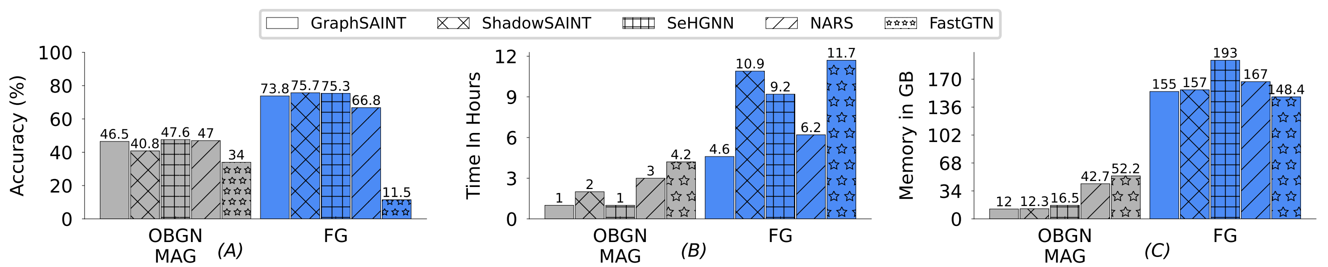

Figure 2: (A) Accuracy (higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better).GNN Node classification task using Graph-SAINT[12] (Homogenous GNN) and NARS[8], FastGTN[10] (Heterogeneous GNNs) on MAG dataset versions (OBGN-MAG, and FG). Graph-SAINT is able to achieve the best accuracy on all datasets with significant savings in training time and memory.

1.3 Node Classification Tasks

We evaluate the efficiency and accuracy of KG-TOSA meta-sampling on classification tasks. The figures 3,4,5 show the three metrics Classification accuracy(Micro F1), Training time, and Training Memory for node classification tasks on MAG, DBLP, and YAGO KG datasets respectively.

Overall, KG-TOSA enables GNN methods to achieve at least comparable accuracy or better accuracy than training with FG which significantly improves the training time and memory in all experiments. KG-TOSA allows accurate and efficient GNN training on various node classification tasks on different KGs with different sizes from different domains.

1.4 Link Prediction Tasks

We evaluate the efficiency and accuracy of KG-TOSA meta-sampling on link prediction tasks. Figure 6 shows the three metrics (link prediction score (MRR, Hits @ K), Training time, and Training Memory) to evaluate the performance of KG-TOSA on the four datasets. These datasets are varying in source KG, size, domain, node statistics, and the link to predict.

The figures compare the results of training The GNN methods with two varieties of datasets namely the **FG**, and **KG-TOSA** a meta-sampled version of the FG with PQ variation for the same task. The GNN methods used for evaluation are RGCN as a full batch method, ComplEx as a KGE method, and MorsE as A GNN sampling-based method.

REFERENCES

[1] Hussein Abdallah, Yasaman Sabbagh, Panos Kalnis, and Essam Mansour. 2023. Open Knowledge Graph Benchmark (OKGB): GNN methods on large-scale heterogeneous datasets. *CoRR* (2023). https://www.overleaf.com/read/nsmktybzpjvf

[2] Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, and Huajun Chen. 2022. Meta-Knowledge Transfer for Inductive Knowledge Graph Embedding. In *Proceedings of the 45th International ACM SIGIR Conference (SIGIR '22)*. Association for Computing Machinery, 927–937. https://doi.org/10.1145/3477495.3531757

[3] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. 2020. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*. ACM / IW3C2, 2331–2341. https://doi.org/10.1145/3366423.3380297

[4] 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 *Advances in Neural Information Processing Systems 33:NeurIPS*.

[5] Qingsong Lv, Ming Ding, and et. al. 2021. Are we really making much progress?: Revisiting, benchmarking and refining heterogeneous graph neural networks. In *KDD 21*. ACM, 1150–1160. https://doi.org/10.1145/3447548.3467350

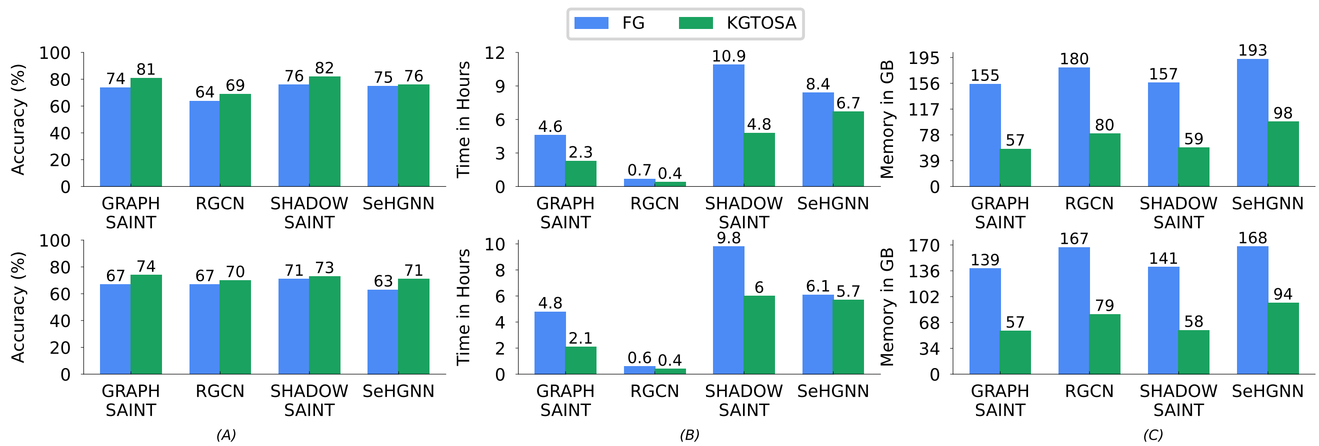

Figure 3: (A) Accuracy (higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better). Analyzing MAG-42M (FG) dataset performance on two node classification tasks using Graph-SAINT, Shadow-SAINT, RGCN, and SeHGNN. The figures on top show the results for the paper-venue classification task. The figures at the bottom show the results for the paper-discipline classification task. The KG-TOSA achieves the highest accuracy with minimal training cost.

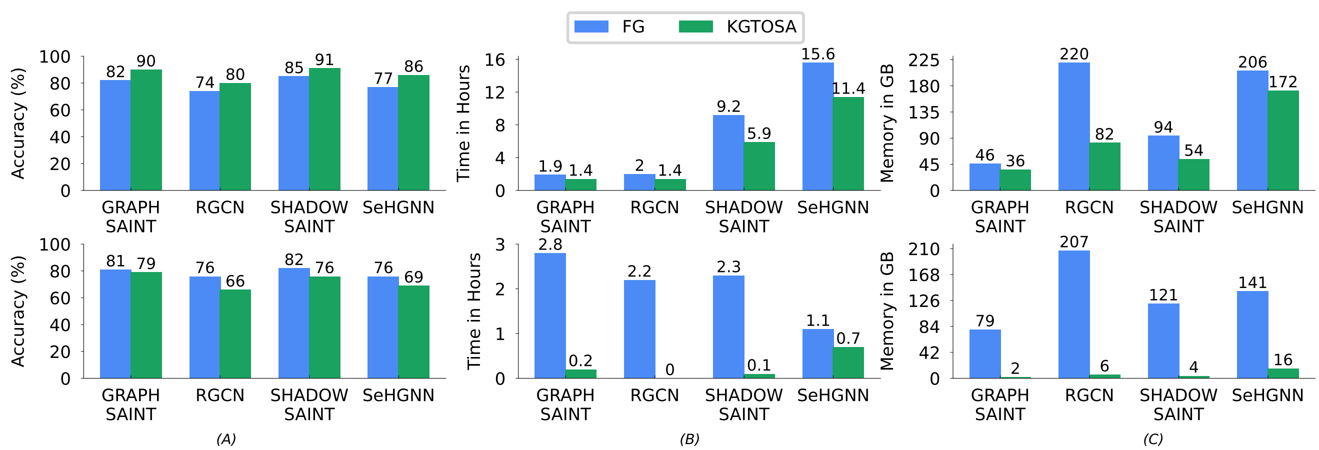

Figure 4: (A) Accuracy (higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better). Analyzing DBLP-15M (FG) dataset performance on Two node classification tasks using Graph-SAINT, Shadow-SAINT, RGCN, and SeHGNN. The figures on top show the results for the paper-venue classification task. The figures at the bottom show the results for the author-country classification task. The KG-TOSA achieves the highest accuracy with minimal training cost for the paper-venue classification task. For the author-country task, the tasked-based sampled (BGP Star-query) sub-graph achieves comparable accuracy to the trained full graph with significantly lower memory and time consumption.

[6] Michael Sejr Schlichtkrull, Thomas N. Kipf, and et al. Peter Bloem. 2018. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, Vol. 10843. Springer, 593–607. https://doi.org/10.1007/978-3-319-93417-4_38

[7] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016 (JMLR Workshop and Conference Proceedings)*, Maria-Florina Balcan and Kilian Q. Weinberger (Eds.), Vol. 48. JMLR.org, 2071–2080. http://proceedings.mlr.press/v48/trouillon16.html

[8] Lingfan Yu, Jiajun Shen, Jinyang Li, and Adam Lerer. 2020. Scalable Graph Neural Networks for Heterogeneous Graphs. https://doi.org/10.48550/ARXIV.2011.09679

[9] Pengyang Yu, Chaofan Fu, Yanwei Yu, Chao Huang, Zhongying Zhao, and Junyu Dong. 2022. Multiplex Heterogeneous Graph Convolutional Network. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*. ACM, 2377–2387. https://doi.org/10.1145/3534678.3539482

[10] Seongjun Yun, Minbyul Jeong, Sungdong Yoo, Seunghun Lee, Sean S. Yi, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. 2022. Graph Transformer Networks: Learning meta-path graphs to improve GNNs. *Neural Networks* 153 (2022), 104–119. https://doi.org/10.1016/j.neunet.2022.05.026

[11] Hanqing Zeng, Muhan Zhang, Yinglong Xia, and et.al. 2022. Decoupling the Depth and Scope of Graph Neural Networks. *CoRR* abs/2201.07858 (2022). https://arxiv.org/abs/2201.07858

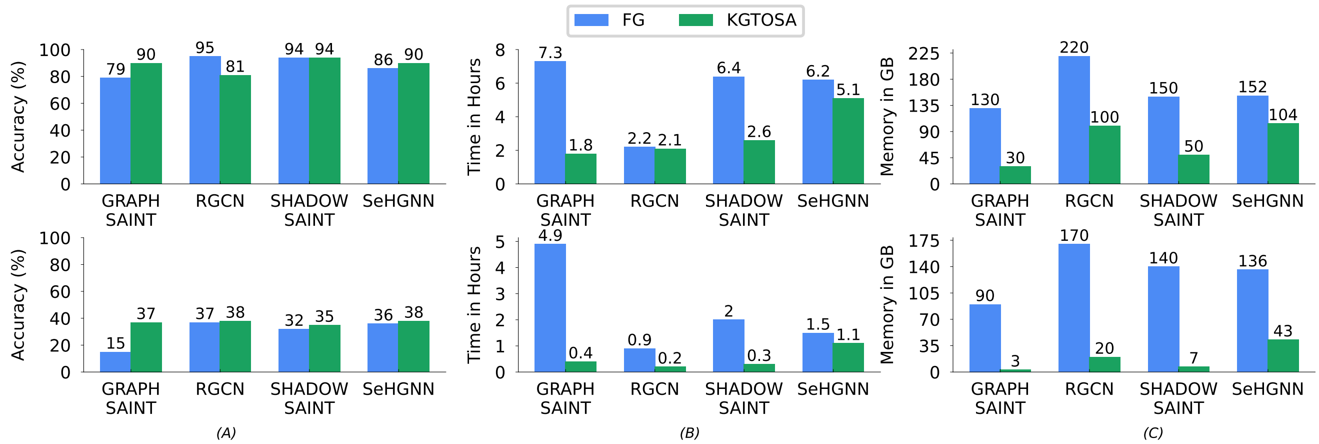

Figure 5: (A) Accuracy (higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better). Analyzing YAGO-30M (FG) dataset performance on Two node classification tasks using Graph-SAINT, Shadow-SAINT, RGCN, and SeHGNN. The figures on top show the results for the place-country classification task. The figures at the bottom show the results for the creative work-genre classification task. The KG-TOSA achieves the highest accuracy with minimal training cost for the creative work-genre classification task. For the place-country classification task, KG-TOSA achieves comparable accuracy to FG with significantly lower memory and time.

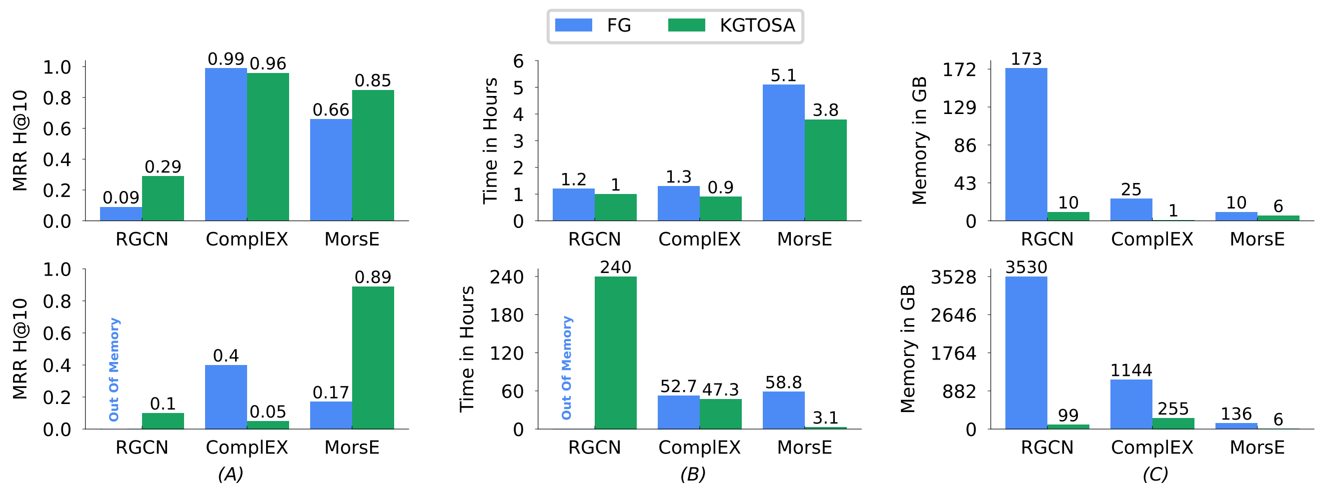

Figure 6: (A) Mean reciprocal rank (MRR Hits@10 higher is better), (B) Training-Time (lower is better), (C) Training-Memory (lower is better). Link prediction tasks on 2 KGs using RGCN[6], ComplEx[7], and MorsE[2] methods. The top figure uses the Yago3-10 (FG) dataset. The figure at the bottom uses DBLP-15M (FG) dataset. OOM is out of memory in DBLP-15M FG using RGCN. The KG-TOSA sub-graphs enable comparable scores to FG, saving training time and memory.

[12] Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor K. Prasanna. 2020. GraphSAINT: Graph Sampling Based Inductive Learning Method. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020.* https://openreview.net/forum?id=BJe8pkHFwS

[13] Jianan Zhao, Xiao Wang, Chuan Shi, Binbin Hu, Guojie Song, and Yanfang Ye. 2021. Heterogeneous Graph Structure Learning for Graph Neural Networks. In *AAAI*. AAAI Press, 4697–4705. https://ojs.aaai.org/index.php/AAAI/article/view/16600