A Appendix

A.1 More dataset details

On MAG and ACM, Paper (P) objects are associated with 128-dimensional real features. On IMDB, Movie (M) objects are associated with 14-dimensional real features. On DBLP, all the objects have no real object features. For the objects that are not associated with real features, following our previous work [8], we generate random features for them by the Xavier uniform distribution [2].

For some baseline methods [6, 1, 4, 7], they need users to input several useful meta-paths, which are specified in Table 1. Our SHGP does not require meta-paths.

Table 1: Meta-paths used by baselines.

Datasets	Meta-paths
MAG	PP, PFP, PAP, $PAIAP$
ACM	PAP, PSP
DBLP	APA, $APCPA$, $APTPA$
IMDB	MAM, MUM, MDM

Note that, although the datasets used in the papers of some baselines may have the same name as ours, they may not be identical. Please refer to their dataset statistics. For example, DBLP in [4] and DBLP in [7] are quite different, and IMDB in [4] and IMDB in [5] are also different. We reproduce the experimental results of all the baselines on our datasets.

A.2 Efficiency of SHGP

We follow GCN [3] to show the efficiency of our SHGP. Specifically, we construct eight HINs with different number of links, and use SHGP to pre-train each of them for 100 epochs on our GPU. The actual total time cost is reported in Figure 1. As shown, overall, the time cost increases roughly linearly with respect to the number of links, which demonstrates the high efficiency of our SHGP.

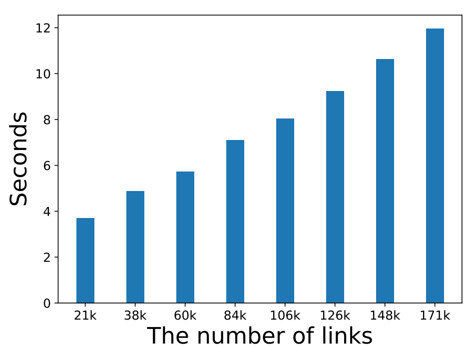
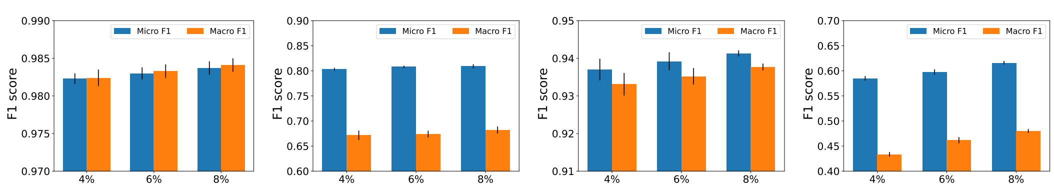


Figure 1: The total time cost (seconds) of 100 pre-training epochs of SHGP.



(a) MAG (b) ACM (c) DBLP (d) IMDB

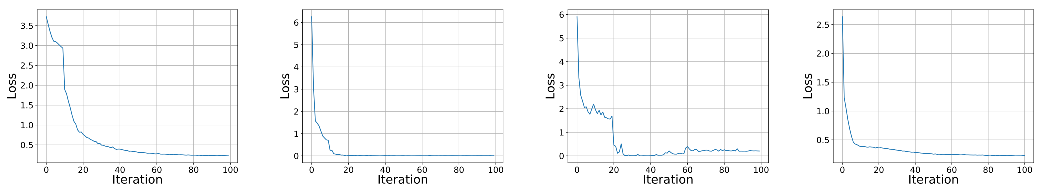
Figure 2: Error bar of SHGP in term of object classification.

A.3 Impact of training ratio

To more intuitively observe the impact of training ratio on object classification performance. In Figure 2, we show the error bar of SHGP with respect to ten random repeats of object classification. We can see, overall, the classification performance improves as the training ratio increases, slightly.

A.4 Convergence of SHGP

In Figure 3, we plot the pre-training loss curves of SHGP on the four benchmark datasets. As we can see, our SHGP converges fast. Specifically, it converges in about 20 iterations on ACM, 30 iterations on DBLP, and 80 iterations on MAG and IMDB.



(a) MAG (b) ACM (c) DBLP (d) IMDB

Figure 3: The pre-training loss curves of SHGP on four datasets.

References

[1] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*, pages 135–144. ACM, 2017.

[2] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, pages 249–256, 2010.

[3] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv:1609.02907*, 2016.

[4] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *AAAI*, pages 5371–5378, 2020.

[5] Yuxiang Ren, Bo Liu, Chao Huang, Peng Dai, Liefeng Bo, and Jiawei Zhang. Heterogeneous deep graph infomax. *arXiv:1911.08538*, 2019.

[6] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *WWW*, pages 2022–2032, 2019.

[7] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In *SIGKDD*, page 1726–1736, 2021.

[8] Yaming Yang, Ziyu Guan, Jianxin Li, Wei Zhao, Jiangtao Cui, and Quan Wang. Interpretable and efficient heterogeneous graph convolutional network. *TKDE*, 2021.