A Appendix

A.1 More details about datasets

In this study, the datasets from the following references were used, and the datasets can be downloaded from the provided links.

- IMDB Yang et al. (2022)
  IMDB consists of Movie (M), Actor (A), User (U), and Director (D). They are connected by six relationships(M↔A, M↔D, M↔U). Target node is Movie(M) node and has 14-dimensional feature. https://github.com/kepsail/SHGP
- DBLP Fu et al. (2020)
  DBLP consists of 4 distinct node types: Paper (P), Author (A), Conference (C), and Term(T). They are connected with 3 types of relationships(A↔P, P↔C, P↔T). Target node is Author(A) node, and has 334-dimensional feature. https://github.com/cynricfu/MAGNN
- ACM Zhao et al. (2020)
  ACM consists of 3 distinct node types: Paper (P), Author (A), and subject(S). They are connected with 2 types of relationships(P↔A, P↔S). Target node is Paper(P) node, and has 1902-dimensional features. https://github.com/AndyJZhao/NSHE
- MAG Yang et al. (2022)
  MAG consists of 4 distinct node types: Paper (P), Author (A), Institution (I), and Field (F). They are connected with 4 types of relationships(P↔P, P↔F, P↔A, A↔I). Target node is Paper(P) node, and has 128-dimensional features. https://github.com/kepsail/SHGP

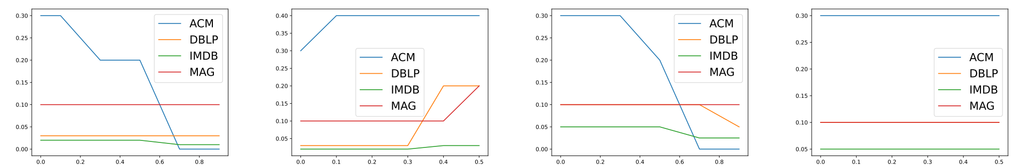
A.2 Details about the baselines

Baselines GraphSAGE (abbreviated as SAGE) Hamilton et al. (2018) is an unsupervised homogeneous graph model that minimizes the representation difference between nodes. Mp2vec Dong et al. (2017) (abbreviated as M2V) Jiang et al. (2017) proposes a metapath-guided random walk strategy to determine the correlation among different types of nodes and edges. DMGI Park et al. (2020) is an extended version of DGI Veličković et al. (2018) for the SSL in HINs. CKD Wang et al. (2022) utilized knowledge distillation to model the correlation between metapaths. HeCo Wang et al. (2021) proposed a co-contrastive learning framework between a metapath-based view and a network schema-based view. SHGP Yang et al. (2022) proposed SSL model without positive pairs by utilizing pseudo labels from structural clustering. HDMI Jing et al. (2021) first utilized both extrinsic and intrinsic mutual information.

Implementation details For mp2vec Dong et al. (2017), we selected walk length to 100 and the number of walks per node was 50. For all baseline models, we tested five times and reported the average results. For HeCo Wang et al. (2021), this method was designed under the assumption that all target nodes maintain the same network schema. Therefore, it cannot be applied to cases like IMDB where there are missing links. For such instances, we modified the model to receive a zero vector from neighboring nodes.

A.3 Hyperparameter analysis

In this section, we aim to explore the behavior of the balance hyperparameter α and the alignment hyperparameter β in response to changes in the balance between features and topology. To illustrate this, Figure 1 depicts how the optimal values of α and β are affected by varying levels of feature mask noise and edge drop noise. Upon examining Figure 1 (a) and (b), it is observed that the optimal alpha values for ACM and IMDB decrease with the increase in feature mask noise and those for four datasets increase with the rise in edge drop noise. This shows that the model is responding to situations where the importance of topology increases due to feature mask noise, or the importance of features increases due to edge drop noise. These results support our claim that our model can balance between the topology and features according to the relative importance of features and topology. Since MAG has an extremely informative topology compared to the node features (Table 1), the balance hyperparameter α did not change a lot. Moreover, in Figure 1 (c), we can see the alignment hyperparameter β decreases as the feature mask noises increase. This also demonstrates that the model adjusts its learning because the representation from the node features becomes less informative. This result reinforces our statement in section 3.4 that if the representation from one modality significantly lags behind the other, aligning two representations will result in the loss of crucial information. Lastly, in Figure 1 (d), it can be observed that beta remains consistent, which indicates that the influence of the alignment hyperparameter is low when correcting the balance affected by edge drop noise. Additionally, in Figure 1 (d), MAG and DBLP are shown overlapping as they have exactly the same values.



(a) α & feature mask noise (b) α & edge drop noise (c) β & feature mask noise (d) β & edge drop noise

Figure 1: Change in the optimal α and β according to the noise

References

Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. 2017.

Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of The Web Conference 2020*. ACM, apr 2020. doi: 10.1145/3366423.3380297. URL `https://doi.org/10.1145%2F3366423.3380297`.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018.

He Jiang, Yangqiu Song, Chenguang Wang, Ming Zhang, and Yizhou Sun. Semi-supervised learning over heterogeneous information networks by ensemble of meta-graph guided random walks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 1944–1950, 2017. doi: 10.24963/ijcai.2017/270. URL `https://doi.org/10.24963/ijcai.2017/270`.

Baoyu Jing, Chanyoung Park, and Hanghang Tong. HDMI: High-order deep multiplex infomax. In *Proceedings of the Web Conference 2021*. ACM, apr 2021. doi: 10.1145/3442381.3449971. URL `https://doi.org/10.1145%2F3442381.3449971`.

Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding, 2020.

Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax, 2018.

Can Wang, Sheng Zhou, Kang Yu, Defang Chen, Bolang Li, Yan Feng, and Chun Chen. Collaborative knowledge distillation for heterogeneous information network embedding. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pp. 1631–1639, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450390965. doi: 10.1145/3485447.3512209. URL `https://doi.org/10.1145/3485447.3512209`.

Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning, 2021.

Yaming Yang, Ziyu Guan, Zhe Wang, Wei Zhao, Cai Xu, Weigang Lu, and Jianbin Huang. Self-supervised heterogeneous graph pre-training based on structural clustering. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL `https://openreview.net/forum?id=fBU4qsM6Fkf`.

Jianan Zhao, Xiao Wang, Chuan Shi, Zekuan Liu, and Yanfang Ye. Network schema preserving heterogeneous information network embedding. In Christian Bessiere (ed.), *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pp. 1366–1372. International Joint Conferences on Artificial Intelligence Organization, 7 2020. doi: 10.24963/ijcai.2020/190. URL `https://doi.org/10.24963/ijcai.2020/190`. Main track.