Technical Report

1 PROOF OF LEMMA 4.1

Lemma 4.1. Optimizing the loss function in Equation 3 is equivalent to maximizing $I(f(\mathcal{G}[S_1]); f(\mathcal{G}[S_2]))$, leading to the maximization of $I(f(\mathcal{G}[S_1]); \mathcal{G}[S_2])$, where I represents mutual information.

Proof. Since the objective of the encoder is to maximize the embedding similarity between positive subgraph pairs and minimizing the embedding similarity between negative pairs, the loss function for a pair of subgraphs $G[S_1]$ and $G[S_2]$ with embeddings $e(G[S_1])$ and $e(G[S_2])$ is equivalent to:

$$\mathcal{L}_{1,2} = -\log \frac{\exp(-D(G[S_1], G[S_2])}{\sum_{i,j} \exp(-D(G[S_i], G[S_j]))} . \tag{1}$$

And minimizing Equation 1 is equivalent to maximizing a lower bound of the mutual information between the learned embeddings of two subgraphs as proven in [4], which is equivalent to maximizing the mutual information between their embeddings $I(f(\mathcal{G}[S_1]); f(\mathcal{G}[S_2]))$. According to Lemma A.5 in [3], it leads to the maximization of $I(f(\mathcal{G}[S_1]); \mathcal{G}[S_2])$. □

2 PROOF OF THEOREM 4.2

Theorem 4.2. Let f_1 denote our proposed abnormal subgraph encoder with financial distribution considered, and let f_2 denote the encoder without financial distribution modeling as in [1, 2]. After sufficient training of f_1 and f_2, $I(f_1(\mathcal{G}[S_1]); y) > I(f_2(\mathcal{G}[S_1]); y)$, where y represents the graph label.

Proof. The encoder f_1 can differentiate financial subgraphs with different financial distributions that f_2 cannot. It means that f_1 can capture more information of subgraphs $\mathcal{G}[S_1]$ than f_2:

$$H(\mathcal{G}[S_1]) \geq H(f_1(\mathcal{G}[S_1])) > H(f_2(\mathcal{G}[S_1])) , \tag{2}$$

where $H(\cdot)$ represents information entropy. Since $f_1(\mathcal{G}[S_1])$ and $f_2(\mathcal{G}[S_1])$ are functions of $\mathcal{G}[S_1]$, we have

$$I(f_1(\mathcal{G}[S_1]); \mathcal{G}[S_1]) > I(f_2(\mathcal{G}[S_1]); \mathcal{G}[S_1]) . \tag{3}$$

According to Theorem A.6 in [3] and Lemma 4.1, it holds that $I(f_1(\mathcal{G}[S_1]); y) > I(f_2(\mathcal{G}[S_1]); y)$. □

REFERENCES

[1] Zhaoyu Lou, Jiaxuan You, Chengtao Wen, Arquimedes Canedo, Jure Leskovec, et al. 2020. Neural subgraph matching. *arXiv preprint arXiv:2007.03092* (2020).

[2] Xixi Wu, Yun Xiong, Yao Zhang, Yizhu Jiao, Caihua Shan, Yiheng Sun, Yangyong Zhu, and Philip S Yu. 2022. Clare: A semi-supervised community detection algorithm. In *SIGKDD*. 2059–2069.

[3] Yucheng Wu, Leye Wang, Xiao Han, and Han-Jia Ye. 2024. Graph Contrastive Learning with Cohesive Subgraph Awareness. *arXiv preprint arXiv:2401.17580* (2024).

[4] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in neural information processing systems* 33 (2020), 5812–5823.