SCIENCE CHINA
Information Sciences

• Supplementary File •

# Multi-topology Contrastive Graph Representation Learning

Yu XIE[1], Jie JIA[1], Chao WEN[1], Deyu LI[1] & Ming LI[2*]

[1]*Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China*
[2]*Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua 321004, China*

## Appendix A Notation definition

**Table A1** Notations for MCGRL.

| Notations | Descriptions |
|---|---|
| $G$ | An undirected graph $G = \{V, E, A, X\}$ |
| $\mathcal{G}$ | Multi-topology graphs $\mathcal{G} = \{G_1, G_2, ..., G_K\}$ |
| $K$ | The number of topologies |
| $V$ | The set of nodes in $G$ |
| $v_i$ | A node $v_i \in V$ |
| $E$ | The set of edges in $G$ |
| $e_{ij}$ | An edge $e_{ij} \in E$ between $v_i$ and $v_j$ |
| $A \in \mathbb{R}^{N\times N}$ | The adjacency matrix of a graph |
| $X_i$ | Original feature vector of node $v_i$ |
| $D$ | The degree matrix of $A$ |
| $I$ | The identity matrix of $A$ |
| $H^k \in \mathbb{R}^{N\times d}$ | The global node representations for the $k$-th topology |
| $\tilde{H}^k \in \mathbb{R}^{N\times d}$ | The local node representations for the $k$-th topology |
| $Z^k \in \mathbb{R}^{N\times d}$ | The multi-granularity node representations for the $k$-th topology |
| $z^k_{v_i} \in \mathbb{R}^{d\times 1}$ | The multi-granularity node representations for node $v_i$ in the $k$-th topology |
| $s_{ij}$ | The similarity score between nodes $v_i$ and $v_j$ |
| $S$ | The node similarity matrix |
| $C$ | The number of clusters for each topology |
| $cluster^k_i$ | The $i$-th cluster for the $k$-th topology |

## Appendix B Theoretical analysis

**Theorem 1.** Let $f_{MCGRL}$ represent our proposed method with multi-topology interaction and semantic consistency among topolpgies. Let $f_{GCN}$ denote graph contrastive learning methods based on GCN. From the information theory perspective, $f_{MCGRL}$ can capture more information of $G$ than $f_{GCN}$, leading to $H(f_{MCGRL}(G)) > H(f_{GCN}(G))$, where $H(\cdot)$ denotes the entropy function.

Proof. Information theory provides a natural avenue to state the assumption for multi-topology contrastive learning. In particular, for discrete random variables $\mathcal{V}_A$, $\mathcal{V}_B$, and $\mathcal{V}_C$, the conditional mutual information $I(\mathcal{V}_A; \mathcal{V}_B|\mathcal{V}_C)$ measures the degree of shared information between $\mathcal{V}_A$ and $\mathcal{V}_B$ given $\mathcal{V}_C$. In multi-topology graph representation learning, considering a set of topological graphs $\mathcal{G}$, we generate a new topology $G$ from the latent complete space $\mathbb{X}$ and incorporate it into $\mathcal{G}$. For instance, in a scenario where $\mathcal{G}$ already contains the topologies $G_{k_1}$ and $G_{k_2}$, the newly generated topology $G_{k_3}$ introduces additional complementary information and the magnitude of its informative value can be measured as follows

$$I(\mathbb{X}; G_{k_3} \left| G_{k_2}, G_{k_1} \right.) \geqslant \varepsilon_{\text{info}}, \tag{B1}$$

* Corresponding author (email: mingli@zjnu.edu.cn)

where $\varepsilon_{info}$ is a variable larger than zero. Moreover, considering the generation of $K$ topologies $G_1, ..., G_K$ from the latent space $\mathbb{X}$, the information obtained from $\mathbb{X}$ can be expressed as follows

$$I(\mathbb{X}; G_1, ..., G_K) = \sum_{k=1}^{K} I(\mathbb{X}; G_k|G_{k-1}, ..., G_1). \tag{B2}$$

Similarly, we can have

$$I(\mathbb{X}; cluster_1^{k_1}, cluster_2^{k_1}, ..., cluster_{|C|}^{k_K}) = \sum_{k=1}^{K} I(\mathbb{X}; cluster_{|C|}^{k_K}|cluster_{|C-1|}^{k_K}, ..., cluster_1^{k_1}). \tag{B3}$$

While a single topology may inadequately depict $\mathbb{X}$, exploiting the complementary among topologies enables the acquisition of enriched information to effectively characterize the latent complete space $\mathbb{X}$. MCGRL not only considers multiple topologies generated from different perspectives but also utilizes the information by clustering to generate new topologies for each topology, which enables the capture of richer and different-granularity information. Hence, $f_{MCGRL}$ can capture more information about $\mathbb{X}$ than $f_{GCN}$, so we can obtain that

$$H(G) > H(f_{MCGRL}(G)) > H(f_{GCN}(G)). \tag{B4}$$

**Theorem 2.** After sufficient training, $f_{MCGRL}$ can acquire node representations that preserve more information associated with $Y$ through multi-topology contrastive learning, i.e. $I(f_{MCGRL}(G); Y) > I(f_{GCN}(G); Y)$, where $Y \in \mathbb{R}^{n\times 1}$ is the label matrix of nodes in $G$.

Proof. According to Theorem 1, we can obtain $H(f_{MCGRL}(G)) > H(f_{GCN}(G))$, so

$$I(f_{MCGRL}(G); G) > I(f_{GCN}(G); G), \tag{B5}$$

$$I(f_{MCGRL}(G); G) = I(f_{MCGRL}(G); G_{k_1}) + I(f_{MCGRL}(G); G|G_{k_1}). \tag{B6}$$

The objective function $\mathcal{L}$ incorporates both the multi-topology contrastive loss and the semantic consistency loss, in which the multi-topology contrastive loss attempts to maximize the consistency for positive pairs of node-level representations between any two topologies, and the semantic consistency loss aims at maximizing that of subgraph-level representations between any two topologies. Additionally, GraphCL [1] theoretically demonstrates that minimizing the InfoNCE based graph contrastive loss is equivalent to maximizing the lower bound of the mutual information between latent representations of two topologies, which can be regarded as a form of maximizing mutual information between latent representations. Therefore, optimizing the objective function $\mathcal{L}$ is equivalent to respectively maximizing the mutual information $I(f(G_{k_1}); f(G_{k_2}))$ for the node-level representations and subgraph-level representations between any two topologies.

$$I(f(G_{k_1}); G_{k_2}) = I(f(G_{k_1}); f(G_{k_2})) + I(f(G_{k_1}); G_{k_2}|f(G_{k_2})). \tag{B7}$$

When the objective function $\mathcal{L}$ is optimized, $I(f(G_{k_1}); G_{k_2}|f(G_{k_2}))$ approaches its minimum value 0 and $I(f(G_{k_1}); f(G_{k_2}))$ also reaches its maximum value i.e. $I(f(G_{k_1}); G_{k_2})$. This establishes a direct consequence that the maximization of the objective function $\mathcal{L}$ inherently also results in the maximization of $I(f(G_{k_1}); G_{k_2})$. This give, $I(f_{MCGRL}(G); G|G_{k_1})$ is infinitely close to its minimum value 0, we have that

$$I(f_{MCGRL}(G); G) \approx I(f_{MCGRL}(G); G_{k_1}), \quad I(f_{GCN}(G); G) \approx I(f_{GCN}(G); G_{k_1}). \tag{B8}$$

By Eq. (B5) and (B8), it is easy to verify that

$$I(f_{MCGRL}(G); G_{k_1}) > I(f_{GCN}(G); G_{k_1}), \tag{B9}$$

$$I(f_{MCGRL}(G); G_{k_1}; ...; G_K) > I(f_{GCN}(G); G_{k_1}; ...; G_K), \tag{B10}$$

$$\begin{aligned} I(f_{MCGRL}(G); G_{k_1}) &= I(f_{MCGRL}(G); G_{k_1}; Y) + I(f_{MCGRL}(G); G_{k_1}|Y) \\ &= I(f_{MCGRL}(G); Y) - I(f_{MCGRL}(G); Y|G_{k_1}) + I(f_{MCGRL}(G); G_{k_1}|Y). \end{aligned} \tag{B11}$$

Similarly,

$$\begin{aligned} I(f_{MCGRL}(G); G_{k_1}|Y) &= I(f_{MCGRL}(G); G_{k_1}; G|Y) \\ &\leqslant I(f_{MCGRL}(G); G_{k_1}; G|Y) + I(G_{k_1}; G|Y, f_{MCGRL}(G)) \\ &= I(G_{k_1}; G|Y), \end{aligned} \tag{B12}$$

$$I(f_{MCGRL}(G); G_{k_1}; ...; G_K|Y) \leqslant I(G_{k_1}; ...; G_K; G|Y), \tag{B13}$$

since based on the non-negativity of mutual information, we can have

$$I(G_{k_1}; G|Y, f_{MCGRL}(G)) \geqslant 0, \tag{B14}$$

$$I(G_{k_1};...;G_K;G|Y,f_{MCGRL}(G)) \geqslant 0. \tag{B15}$$

By Eq. (B11), (B12),(B13),(B14), and (B15) we deduce

$$I(f_{MCGRL}(G);G_{k_1}) \approx I(f_{MCGRL}(G);Y) + I(G_{k_1};G|Y), \tag{B16}$$

$$I(f_{MCGRL}(G);G_{k_1};...;G_K) \approx I(f_{MCGRL}(G);Y) + I(G_{k_1};...;G_K;G|Y). \tag{B17}$$

In a similar vein, it can be argued that

$$I(f_{GCN}(G);G_{k_1}) \approx I(f_{GCN}(G);Y) + I(G_{k_1};G|Y), \tag{B18}$$

$$I(f_{GCN}(G);G_{k_1};...;G_K) \approx I(f_{GCN}(G);Y) + I(G_{k_1};...;G_K;G|Y). \tag{B19}$$

Based on Eqs. (B9), (B10), (B16), (B17), (B18), and (B19), we can verify that

$$I(f_{MCGRL}(G);Y) > I(f_{GCN}(G);Y). \tag{B20}$$

We provide a rigorous mathematical analysis of the expressive power of MCGRL by the information theory, which proves that our method can capture more information. Additionally, we investigate the reasons behind the superior performance of our proposed method compared to existing graph contrastive learning methods based on GCN.

## Appendix C Visualization

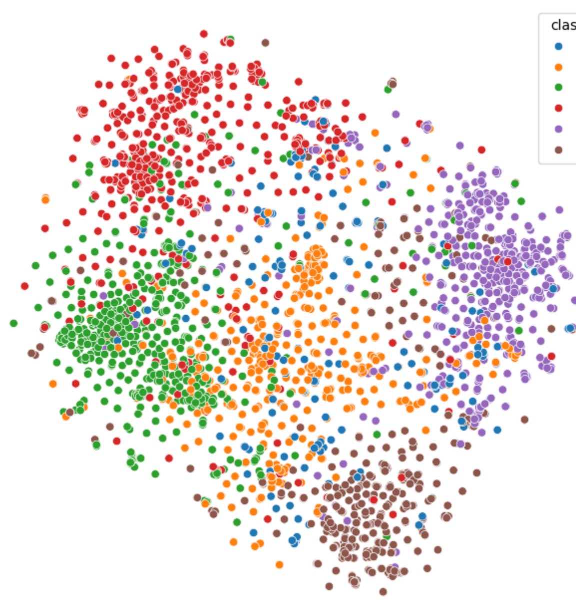


**Figure C1** t-SNE visualization on Citeseer dataset.

To further validate the performance of MCGRL, we utilize the t-SNE [2] visualization tool to present its node representation results in 2D layouts. As shown in Figure C1, different colors correspond to distinct class labels within the Citeseer dataset. It exhibits a clearer classification in the visualization with small intra-class distances and large inter-class distances, which is attributed to the similarity constraint that brings nodes of the same class closer in different topologies and nodes of the different classes in the same topology farther. These results further emphasize the advantages of MCGRL in node classification.

**References**

1 You Y, Chen T, Sui Y, et al. Graph contrastive learning with augmentations. In: Proceedings of Advances in Neural Information Processing Systems, 2020, 33: 5812-5823

2 Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res, 2008, 9(11)