图约束汇总

Leibniz-Institut für Sozialwissenschaften

Reported by Xinsheng Wang

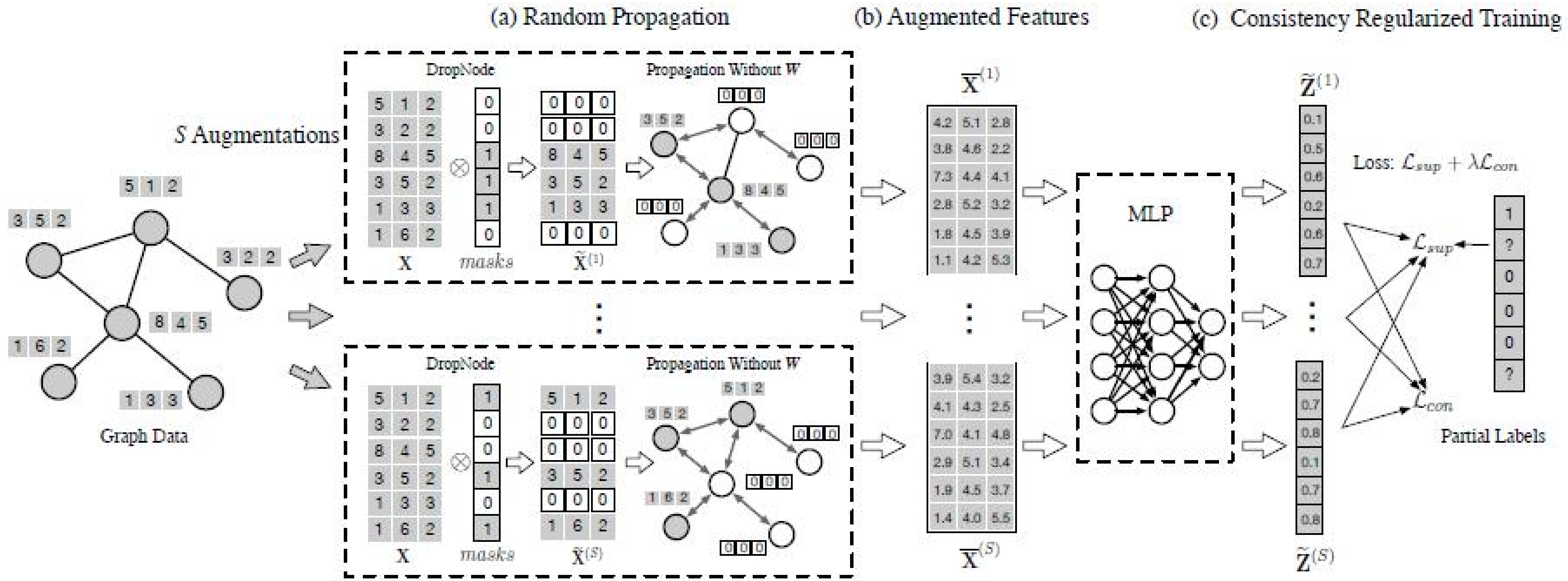


Figure 1: Illustration of GRAND with DropNode as the perturbation method. GRAND designs random propagation (a) to generate multiple graph data augmentations (b), which are further used as consistency regularization (c) for semi-supervised learning.

$$\mathcal{L}_{con} = \frac{1}{S} \sum_{s=1}^{S} \sum_{i=0}^{n-1} \|\overline{\mathbf{Z}}_i' - \widetilde{\mathbf{Z}}_i^{(s)}\|_2^2. \quad (3)$$

NIPS_2020_Graph Random Neural Networks for Semi-Supervised Learning on Graphs

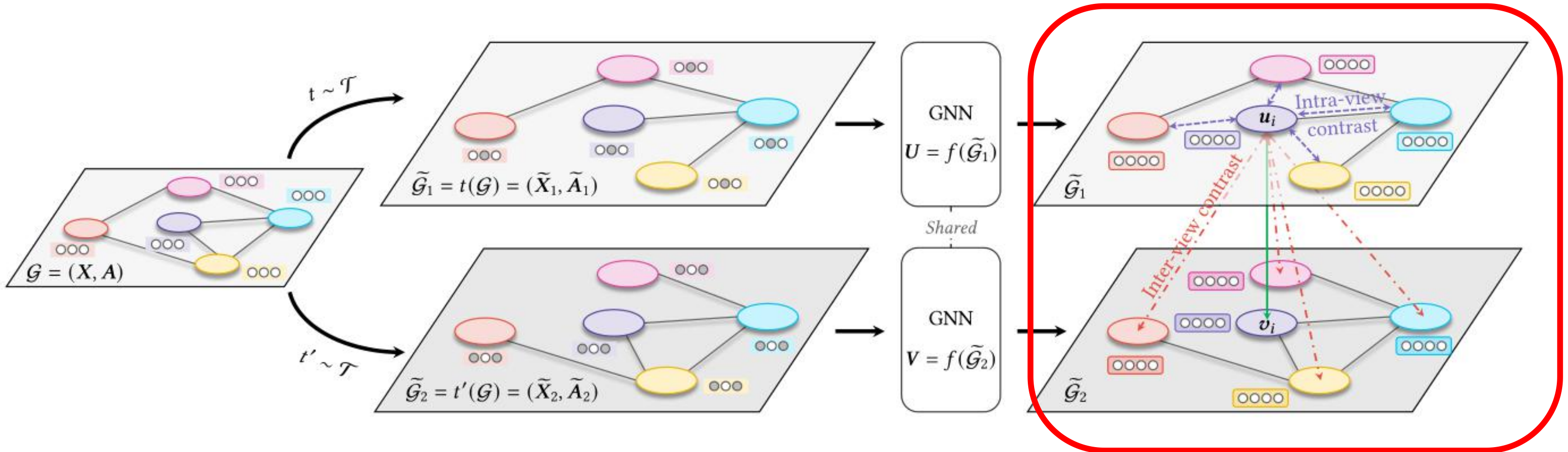


Figure 1: Our proposed deep Graph Contrastive representation learning with Adaptive augmentation (GCA) model. We first generate two graph views via stochastic augmentation that is adaptive to the graph structure and attributes. Then, the two graphs are fed into a shared Graph Neural Network (GNN) to learn representations. We train the model with a contrastive objective, which pulls representations of one node together while pushing node representations away from other node representations in the two views. N.B., we define the negative samples as all other nodes in the two views. Therefore, negative samples are from two sources, intra-view (in purple) and inter-view nodes (in red).

$$\ell(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \frac{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}{\underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_k)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k \neq i} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_k)/\tau}}_{\text{intra-view negative pairs}}}, \quad (1)$$

$$\mathcal{J} = \frac{1}{2N} \sum_{i=1}^{N} \left[\ell(\boldsymbol{u}_i, \boldsymbol{v}_i) + \ell(\boldsymbol{v}_i, \boldsymbol{u}_i) \right]. \quad (2)$$

WWW_2021_Graph Contrastive Learning with Adaptive Augmentation

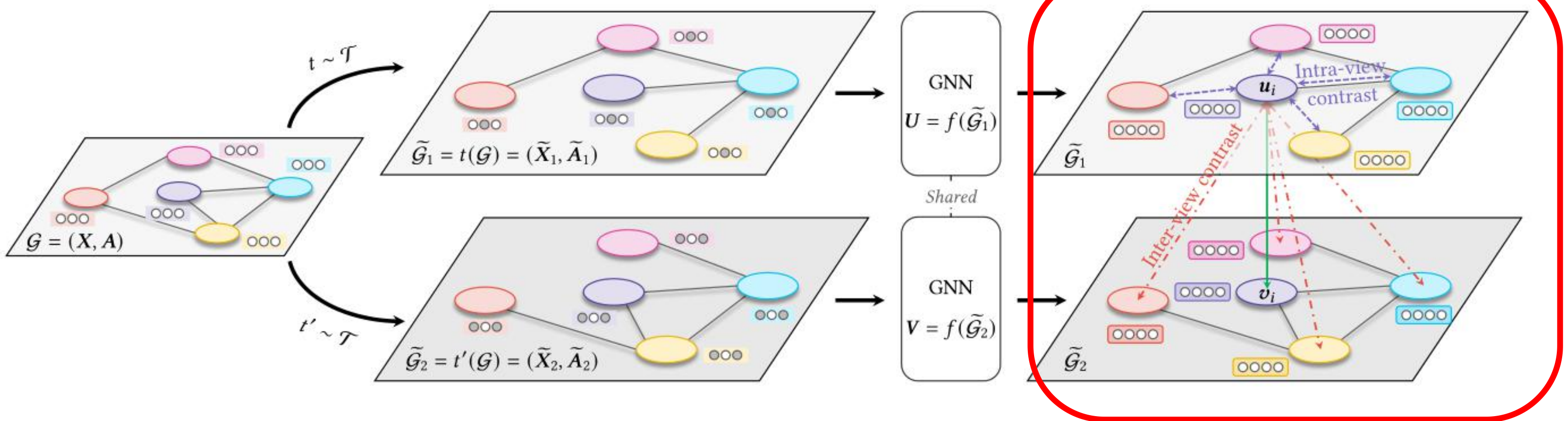


Figure 1: Our proposed deep Graph Contrastive representation learning with Adaptive augmentation (GCA) model. We first generate two graph views via stochastic augmentation that is adaptive to the graph structure and attributes. Then, the two graphs are fed into a shared Graph Neural Network (GNN) to learn representations. We train the model with a contrastive objective, which pulls representations of one node together while pushing node representations away from other node representations in the two views. N.B., we define the negative samples as all other nodes in the two views. Therefore, negative samples are from two sources, intra-view (in purple) and inter-view nodes (in red).

思想

graph. Specifically on the topology level, we design augmentation schemes based on node centrality measures to highlight important connective structures. On the node attribute level, we corrupt node features by adding more noise to unimportant node features, to enforce the model to recognize underlying semantic information. We

WWW_2021_Graph Contrastive Learning with Adaptive Augmentation

节点对，可有边，可无边

Method

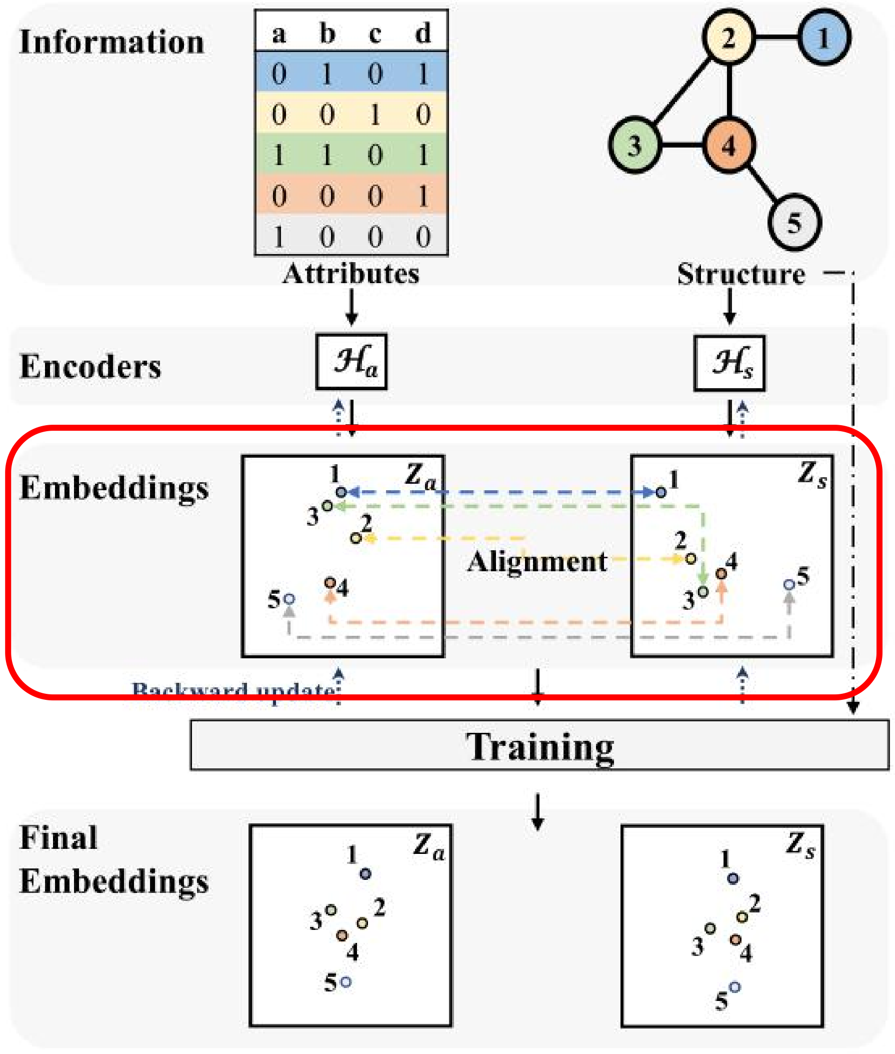


Figure 2: An illustration of the proposed model DEAL.

1. Tight Alignment (*T-align*)

$$\mathcal{L}_{T\text{-}align}(\mathcal{Z}_s, \mathcal{Z}_a) = -\frac{1}{|V|}\sum_{v_i \in V} s(\mathbf{z}_s^i, \mathbf{z}_a^i)$$

2. Loose Alignment (*L-align*)

$$\mathcal{L}_{L\text{-}align}(\mathcal{Z}_s, \mathcal{Z}_a) = \frac{1}{|B|}\sum_{(v_{p_i}, v_{q_i}) \in B} \text{有边} \; [\, y_i \phi_2(s(\mathbf{z}_s^{p_i}, \mathbf{z}_a^{q_i}))$$

$$+ (1 - y_i)\alpha(v_{p_i}, v_{q_i})\phi_1(-s(\mathbf{z}_s^{p_i}, \mathbf{z}_a^{q_i}))\,] \; \text{无边}$$

above observations, we propose the following loss to be optimized for a given mini-batch of node pair samples $B = \{(v_{p_1}, v_{q_1}), \ldots, (v_{p_k}, v_{q_k})\}$ where $p_i \neq q_i$ with $i \in [1, k]$ (obtained by sampling k pairs of nodes from the graph):

where $s(\cdot,\cdot)$ is the function to measure the similarity between node embeddings ($\mathbf{z}^{p_i}$ and $\mathbf{z}^{q_i}$), y_i is the link relation label with $y_i = 1$ if v_{p_i} and v_{q_i} are connected and $y_i = 0$ otherwise, $\alpha(\cdot,\cdot)$ is a weight function, and $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are derived from the function $\phi(\cdot)$ with different hyper-parameters.

$$\phi(x) = \frac{1}{\gamma}\log(1 + e^{-\gamma x + b}),$$

$$\alpha(v_p, v_q) = \exp\left(\frac{\beta}{d_{sp}(v_p, v_q)}\right)$$

where $\beta > 0$ is a hyper-parameter, and $d_{sp}(\cdot,\cdot)$ denotes the shortest path distance between a node pair. If node v_p cannot reach node v_q, $d_{sp}(v_p, v_q) = \infty$. This weight aims to help

$$\mathcal{L}_B(\mathcal{Z}) = \frac{1}{|B|}\sum_{(v_{p_i}, v_{q_i}) \in B} [\,(1 - y_i)\alpha(v_{p_i}, v_{q_i})\phi_1(-s(\mathbf{z}^{p_i}, \mathbf{z}^{q_i})) + y_i\phi_2(s(\mathbf{z}^{p_i}, \mathbf{z}^{q_i}))\,], \quad (4)$$

$$\mathcal{L} = \theta_1\mathcal{L}_B(\mathcal{Z}_s) + \theta_2\mathcal{L}_B(\mathcal{Z}_a) + \theta_3\mathcal{L}_{align}(\mathcal{Z}_s, \mathcal{Z}_a), \quad (9)$$

IJCAI_2021_Inductive Link Prediction for Nodes Having Only Attribute Information

Method

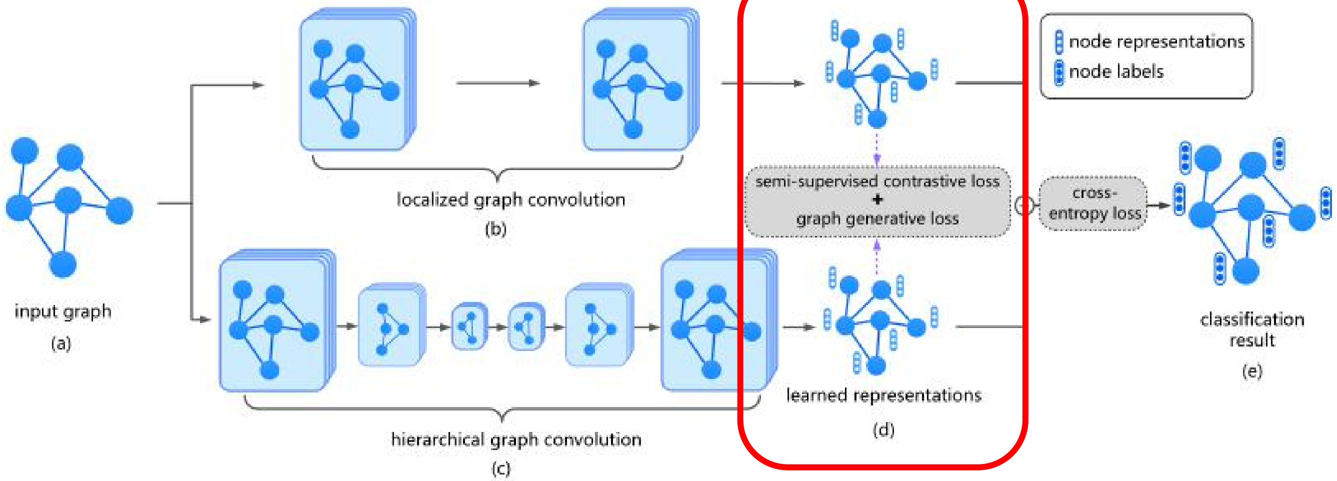


Figure 1: The framework of our approach. In (a), the original graph is adopted as the input of (b) the localized GCNs and (c) the hierarchical GCNs, respectively, where (c) is utilized to capture the global information and serves as the augmented view of (b). In (d), the node representations are generated from (b) and (c), and then constitute the contrastive loss and graph generative loss collaboratively, in order to provide additional supervision signals to improve the representation learning process. In (e), the classification result is acquired via integrating the outputs of (b) and (c), where the cross-entropy loss is used to penalize the difference between the model prediction and the given labels of the initially labeled nodes.

无监督

$$\mathcal{L}_{uc} = \frac{1}{2n}\sum_{i=1}^{n}\left(\mathcal{L}_{uc}^{\phi_1}(\mathbf{x}_i) + \mathcal{L}_{uc}^{\phi_2}(\mathbf{x}_i)\right), \quad (3)$$

$$\mathcal{L}_{uc}^{\phi_1}(\mathbf{x}_i) = -\log\frac{\exp(\langle \mathbf{h}_i^{\phi_1}, \mathbf{h}_i^{\phi_2}\rangle)}{\sum_{j=1}^{n}\exp(\langle \mathbf{h}_i^{\phi_1}, \mathbf{h}_j^{\phi_2}\rangle)}, \quad (4)$$

$$\mathcal{L}_{uc}^{\phi_2}(\mathbf{x}_i) = -\log\frac{\exp(\langle \mathbf{h}_i^{\phi_2}, \mathbf{h}_i^{\phi_1}\rangle)}{\sum_{j=1}^{n}\exp(\langle \mathbf{h}_i^{\phi_2}, \mathbf{h}_j^{\phi_1}\rangle)}. \quad (5)$$

有监督

$$\mathcal{L}_{sc} = \frac{1}{2l}\sum_{i=1}^{l}\left(\mathcal{L}_{sc}^{\phi_1}(\mathbf{x}_i) + \mathcal{L}_{sc}^{\phi_2}(\mathbf{x}_i)\right). \quad (6)$$

$$\mathcal{L}_{sc}^{\phi_1}(\mathbf{x}_i) = -\log\frac{\sum_{k=1}^{l}\mathbb{1}_{[y_i=y_k]}\exp(\langle \mathbf{h}_i^{\phi_1}, \mathbf{h}_k^{\phi_2}\rangle)}{\sum_{j=1}^{l}\exp(\langle \mathbf{h}_i^{\phi_1}, \mathbf{h}_j^{\phi_2}\rangle)}, \quad (7)$$

$$\mathcal{L}_{sc}^{\phi_2}(\mathbf{x}_i) = -\log\frac{\sum_{k=1}^{l}\mathbb{1}_{[y_i=y_k]}\exp(\langle \mathbf{h}_i^{\phi_2}, \mathbf{h}_k^{\phi_1}\rangle)}{\sum_{j=1}^{l}\exp(\langle \mathbf{h}_i^{\phi_2}, \mathbf{h}_j^{\phi_1}\rangle)}, \quad (8)$$

$$\mathcal{L}_{ssc} = \mathcal{L}_{uc} + \mathcal{L}_{sc}. \quad (9)$$

$$p(\mathcal{G}|\mathbf{H}^{\phi_1}, \mathbf{H}^{\phi_2}) = \prod_{i,j} p(e_{ij}|\mathbf{h}_i^{\phi_1}, \mathbf{h}_j^{\phi_2}) = \prod_{i,j}\delta([\mathbf{h}_i^{\phi_1}, \mathbf{h}_j^{\phi_2}]\mathbf{w})$$

$$\mathcal{L}_{g^2} = -p(\mathcal{G}|\mathbf{H}^{\phi_1}, \mathbf{H}^{\phi_2})$$

$$\mathbf{O} = \lambda^{\phi_1}\mathbf{H}^{\phi_1} + (1-\lambda^{\phi_1})\mathbf{H}^{\phi_2}, \quad (12)$$

$$\mathcal{L}_{ce} = -\sum_{i=1}^{l}\sum_{j=1}^{c}\mathbf{Y}_{ij}\ln\mathbf{O}_{ij}. \quad (13)$$

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda_{ssc}\mathcal{L}_{ssc} + \lambda_{g^2}\mathcal{L}_{g^2}, \quad (14)$$

AAAI_2021_Contrastive and Generative Graph Convolutional Networks for Graph-based Semi-Supervised Learning

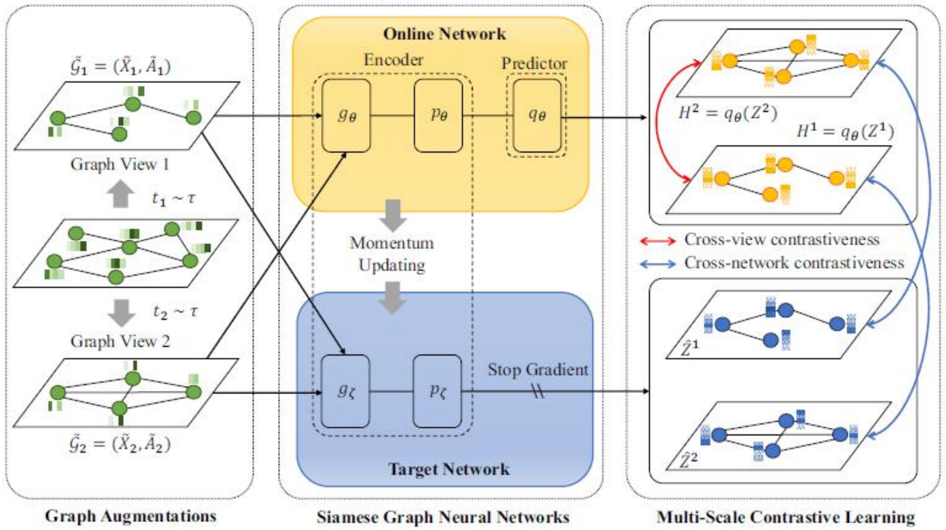


Figure 1: The overall framework of MERIT. Through graph augmentations, we construct two graph views, based on which an online network and a target network are employed to generate node representations for each view. A multi-scale contrastive learning scheme, which utilizes both cross-network and cross-view contrastive modules, is deployed to learn effective node embeddings. g_θ and g_ζ denotes a GNN-based graph encoder. p_θ, p_ζ, and q_θ are two-layer MLP with the batch normalization. $t_1 \sim \tau$ and $t_2 \sim \tau$ are two different graph augmentations

Cross-Network Contrastive Learning

$$\mathcal{L}_{cn}^{1}(v_i) = -\log \frac{\exp(\text{sim}(h_{v_i}^1, \hat{z}_{v_i}^2))}{\sum_{j=1}^{N} \exp(\text{sim}(h_{v_i}^1, \hat{z}_{v_j}^2))},$$

$$\mathcal{L}_{cn}^{2}(v_i) = -\log \frac{\exp(\text{sim}(h_{v_i}^2, \hat{z}_{v_i}^1))}{\sum_{j=1}^{N} \exp(\text{sim}(h_{v_i}^2, \hat{z}_{v_j}^1))}.$$

$$\mathcal{L}_{cn} = \frac{1}{2N} \sum_{i=1}^{N} \left(\mathcal{L}_{cn}^{1}(v_i) + \mathcal{L}_{cn}^{2}(v_i)\right).$$

Cross-View Contrastive Learning

$$\mathcal{L}_{inter}^{1}(v_i) = -\log \frac{\exp(\text{sim}(h_{v_i}^1, h_{v_i}^2))}{\sum_{j=1}^{N} \exp(\text{sim}(h_{v_i}^1, h_{v_j}^2))}.$$

$$\mathcal{L}_{intra}^{1}(v_i) = -\log \frac{\exp(\text{sim}(h_{v_i}^1, h_{v_i}^2))}{\exp(\text{sim}(h_{v_i}^1, h_{v_i}^2)) + \Phi},$$

$$\Phi = \sum_{j=1}^{N} \mathbb{1}_{i \neq j} \exp(\text{sim}(h_{v_i}^1, h_{v_j}^1)),$$

$$\mathcal{L}_{cv}^{k}(v_i) = \mathcal{L}_{intra}^{k}(v_i) + \mathcal{L}_{inter}^{k}(v_i), \quad k \in \{1, 2\}.$$

$$\mathcal{L}_{cv} = \frac{1}{2N} \sum_{i=1}^{N} \left(\mathcal{L}_{cv}^{1}(v_i) + \mathcal{L}_{cv}^{2}(v_i)\right),$$

IJCAI_2021_Multi-Scale Contrastive Siamese Networks for Self-Supervised Graph Representation Learning

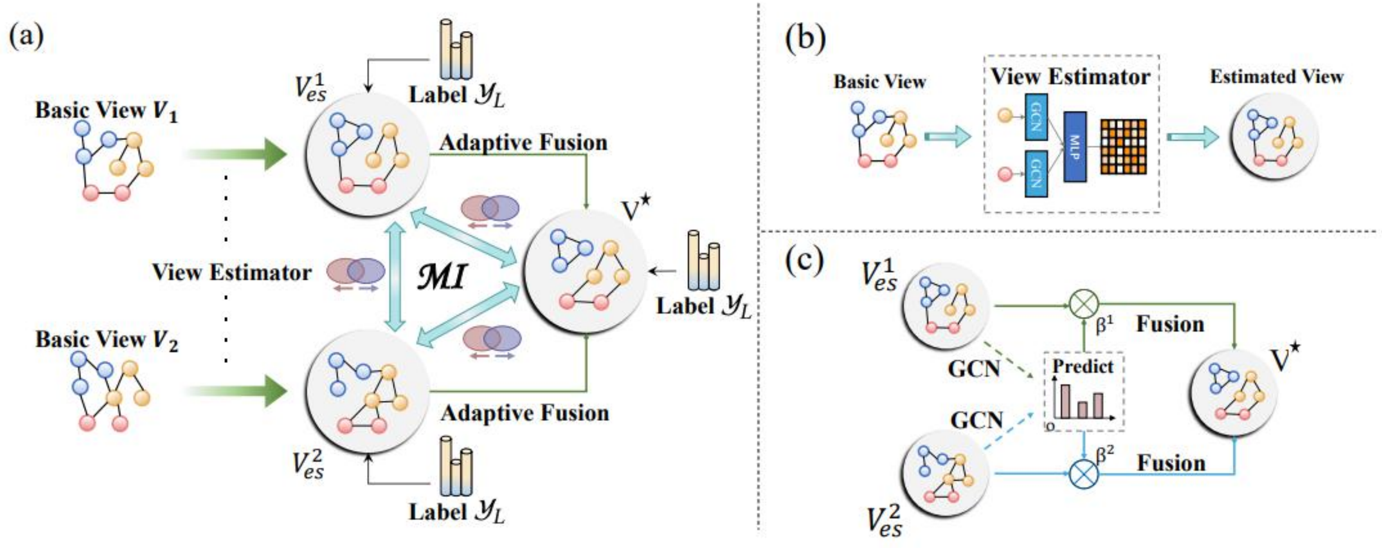


Figure 1: The overview of our proposed CoGSL. (a) Model framework. (b) View estimator. (c) Adaptive fusion.

$$L(V^\star, V_{es}^1) = -\frac{1}{2|B|}\sum_{i=1}^{|B|}\left[\log \frac{e^{sim(h_{p_i}^\star, h_{p_i}^1)/\tau}}{e^{sim(h_{p_i}^\star, h_{p_i}^1)/\tau} + \sum_{k\neq i} e^{sim(h_{p_i}^\star, h_{p_k}^1)/\tau}} + \log \frac{e^{sim(h_{p_i}^1, h_{p_i}^\star)/\tau}}{e^{sim(h_{p_i}^1, h_{p_i}^\star)/\tau} + \sum_{j\neq i} e^{sim(h_{p_i}^1, h_{p_j}^\star)/\tau}}\right], \quad (17)$$

$$\mathcal{L}_{MI} = L(V^\star, V_{es}^1) + L(V^\star, V_{es}^2) + L(V_{es}^1, V_{es}^2). \quad (18)$$

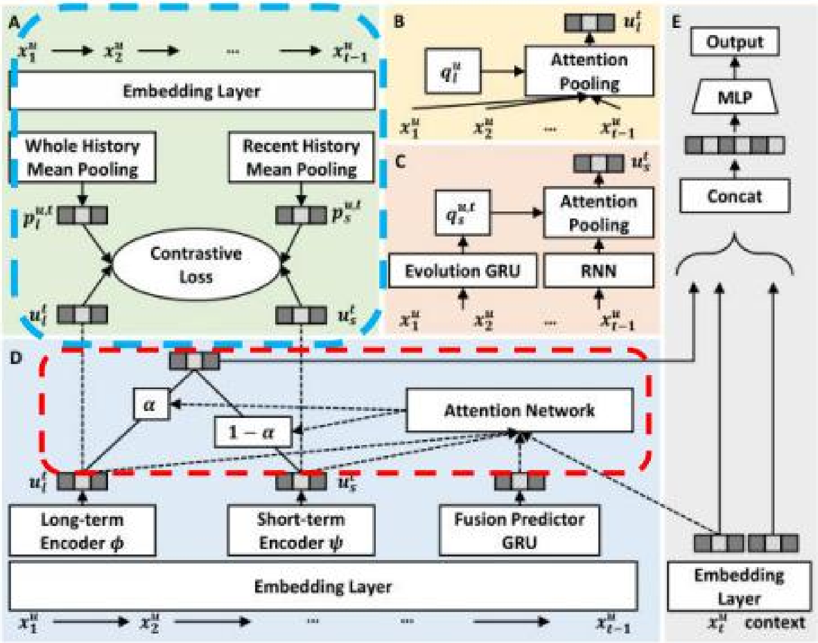


Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

denotes a positive margin value. Both $\mathcal{L}_{bpr}$ and $\mathcal{L}_{tri}$ are designed for making the anchor a more similar to the positive sample p than the negative sample q. Thus the contrastive loss for self-supervised

$$p_l^{u,t} = \text{MEAN}(\{x_1^u, \cdots, x_t^u\}) = \frac{1}{t}\sum_{j=1}^{t} E(x_j^u), \quad (17)$$

$$p_s^{u,t} = \text{MEAN}(\{x_{t-k+1}^u, \cdots, x_t^u\}) = \frac{1}{k}\sum_{j=1}^{k} E(x_{t-j+1}^u), \quad (18)$$

$$sim(\boldsymbol{u}_l^t, \boldsymbol{p}_l^{u,t}) > sim(\boldsymbol{u}_l^t, \boldsymbol{p}_s^{u,t}), \quad (19)$$

$$sim(\boldsymbol{p}_l^{u,t}, \boldsymbol{u}_l^t) > sim(\boldsymbol{p}_l^{u,t}, \boldsymbol{u}_s^t), \quad (20)$$

$$sim(\boldsymbol{u}_s^t, \boldsymbol{p}_s^{u,t}) > sim(\boldsymbol{u}_s^t, \boldsymbol{p}_l^{u,t}), \quad (21)$$

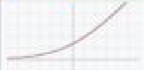
$$sim(\boldsymbol{p}_s^{u,t}, \boldsymbol{u}_s^t) > sim(\boldsymbol{p}_s^{u,t}, \boldsymbol{u}_l^t), \quad (22)$$

$$\mathcal{L}_{\text{bpr}}(a, p, q) = \sigma(\langle a, q\rangle - \langle a, p\rangle), \quad (23)$$

$$\mathcal{L}_{\text{tri}}(a, p, q) = \max\{d(a, p) - d(a, q) + m, 0\}, \quad (24)$$

$$\mathcal{L}_{\text{con}}^{u,t} = f(\boldsymbol{u}_l, \boldsymbol{p}_l, \boldsymbol{p}_s) + f(\boldsymbol{p}_l, \boldsymbol{u}_l, \boldsymbol{u}_s) + f(\boldsymbol{u}_s, \boldsymbol{p}_s, \boldsymbol{p}_l) + f(\boldsymbol{p}_s, \boldsymbol{u}_s, \boldsymbol{u}_l) \quad (25)$$

SoftPlus[20]



$f(x) = \ln(1 + e^x)$

$$\boldsymbol{h}_t^u = \text{GRU}(\{\boldsymbol{E}(x_1^u), \ldots, \boldsymbol{E}(x_t^u)\}), \quad (26)$$

$$\alpha = \sigma(\tau_f(\boldsymbol{h}_t^u \| \boldsymbol{E}(x_{t+1}^u) \| \boldsymbol{u}_l^t \| \boldsymbol{u}_s^t), \quad (27)$$

$$\boldsymbol{u}^t = \alpha \cdot \boldsymbol{u}_l^t + (1 - \alpha) \cdot \boldsymbol{u}_s^t, \quad (28)$$

WWW_2022_Disentangling Long and Short-Term Interests for Recommendation

Datasets	Nodes	Edges	Attributes
CS ([Shchur *et al.*, 2018])	18,333	81,894	6,805
PPI ([Zitnik and Leskovec, 2017])	1,767	16,159	50
Cora ([McCallum *et al.*, 2000])	2,708	5,278	1,433
CiteSeer ([Sen *et al.*, 2008])	3,327	4,552	3,703
PubMed ([Namata *et al.*, 2012])	19,717	44,324	500
Computers ([McAuley *et al.*, 2015])	13,752	245,861	767
Photo ([McAuley *et al.*, 2015])	7,650	119,081	745

IJCAI_2021_Inductive Link Prediction for Nodes Having Only Attribute Information

Dataset	#Nodes	#Edges	#Features	#Classes
Wiki-CS[1]	11,701	216,123	300	10
Amazon-Computers[2]	13,752	245,861	767	10
Amazon-Photo[3]	7,650	119,081	745	8
Coauthor-CS[4]	18,333	81,894	6,805	15
Coauthor-Physics[5]	34,493	247,962	8,415	5

[1] https://github.com/pmernyei/wiki-cs-dataset/raw/master/dataset
[2] https://github.com/shchur/gnn-benchmark/raw/master/data/npz/amazon_electronics_computers.npz
[3] https://github.com/shchur/gnn-benchmark/raw/master/data/npz/amazon_electronics_photo.npz
[4] https://github.com/shchur/gnn-benchmark/raw/master/data/npz/ms_academic_cs.npz
[5] https://github.com/shchur/gnn-benchmark/raw/master/data/npz/ms_academic_phy.npz

WWW_2021_Graph Contrastive Learning with Adaptive Augmentation

Datasets	Nodes	Edges	Features	Classes
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
PubMed	19,717	44,338	500	3
Amazon Computers	13,752	245,861	767	10
Amazon Photo	7,650	119,081	745	8
Coauthor CS	18,333	81,894	6,805	15

AAAI_2021_Contrastive and Generative Graph Convolutional Networks for Graph-based Semi-Supervised Learning

不同图空间的约束总结(4篇)

WWW_2021_Graph Contrastive Learning with Adaptive Augmentation

IJCAI_2021_Inductive Link Prediction for Nodes Having Only Attribute Information

AAAI_2021_Contrastive and Generative Graph Convolutional Networks for Graph-based Semi-Supervised Learning

IJCAI_2021_Multi-Scale Contrastive Siamese Networks for Self-Supervised Graph Representation Learning

知识蒸馏和孪生网络对比总结(2篇)

IJCAI_2021_Multi-Scale Contrastive Siamese Networks for Self-Supervised Graph Representation Learning

A Dual-Channel Knowledge Distillation Framework for Node Classification(ours)