MULTI-SCALE CONTRASTIVE SIAMESE NETWORKS FOR SELF-SUPERVISED GRAPH REPRESENTATION LEARNING

IJCAI 2021 MONTREAL

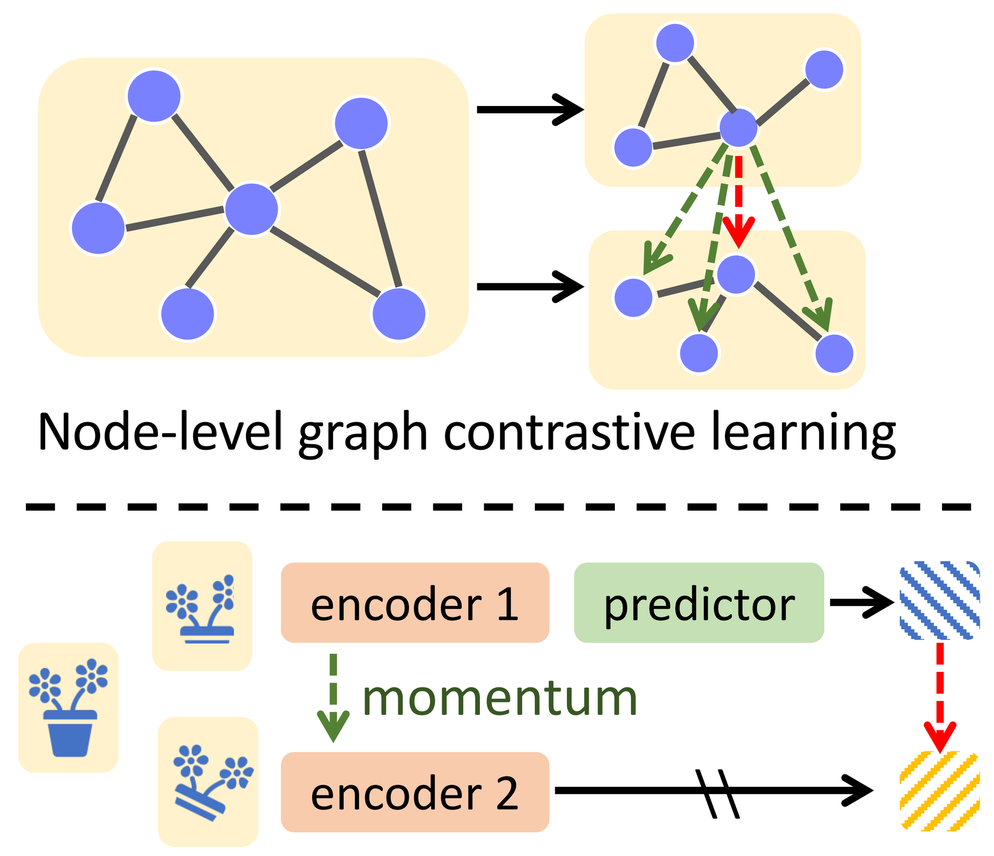
MONASH University

MING JIN, YIZHEN ZHENG, YUAN-FANG LI, CHEN GONG, CHUAN ZHOU, SHIRUI PAN

ABSTRACT

Prior arts on graph representation learning heavily rely on labeling information. To overcome this problem, we propose a novel self-supervised approach to learn node representations by enhancing Siamese self-distillation with multi-scale contrastive learning. Specifically, we first generate two augmented views from the input graph based on local and global perspectives. Then, we employ two objectives called cross- view and cross-network contrastiveness to maximize the agreement between node representations across different views and networks. To demonstrate the effectiveness of our approach, we perform empirical experiments on five real-world datasets.

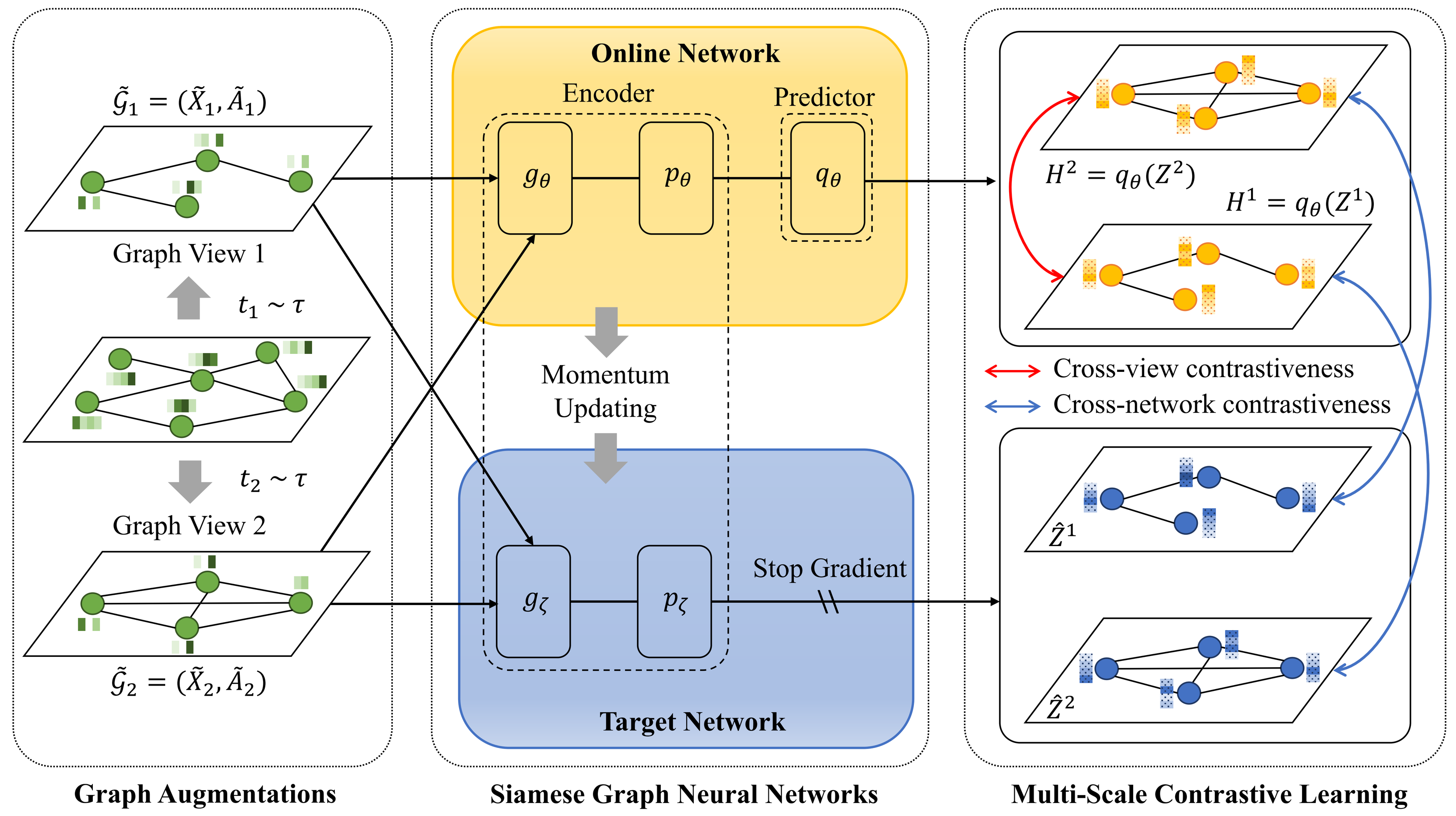
MOTIVATION



Node-level graph contrastive learning

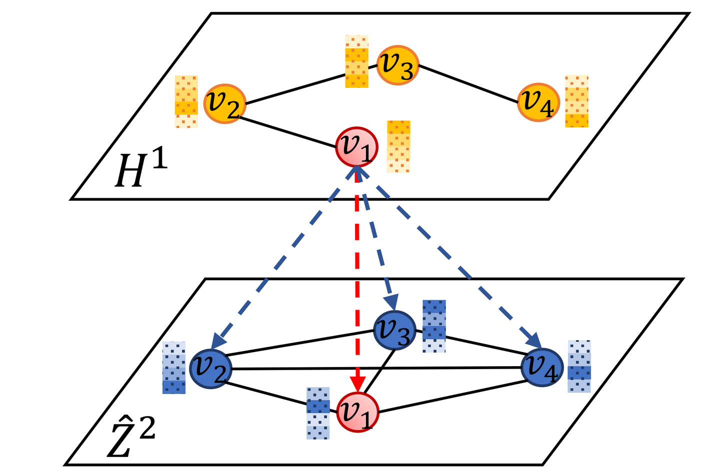
Bootstrapping representation learning

FRAMEWORK



- Two graph views are first generated via augmentations. Then, online and target networks are employed to generate node representations for each view.
- A multi-scale contrastive schema with the self-knowledge distillation is proposed to train the online graph encoder.
- g_θ and g_ζ are two graph encoders. p_θ, p_ζ and q_θ are two-layer MLPs with the batch normalization.

METHODOLOGY



Cross-network contrastiveness aims to distill the knowledge from historical observations and stabilize online graph encoder training.

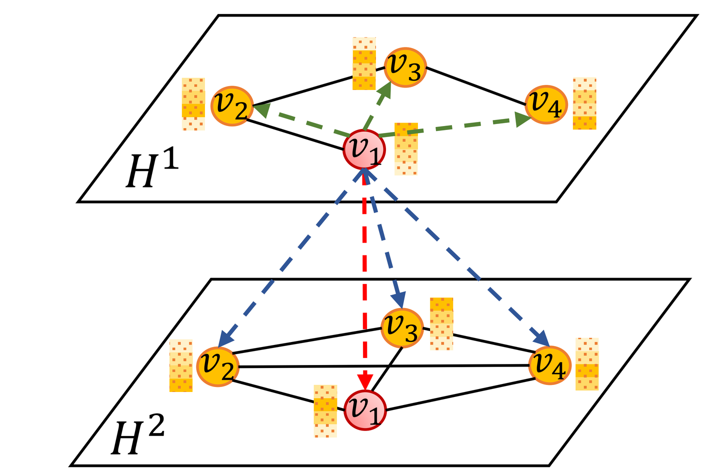
$$\mathcal{L}_{cn} = \frac{1}{2N}\sum_{i=1}^{N}\left(\mathcal{L}^1_{cn}(v_i) + \mathcal{L}^2_{cn}(v_i)\right),$$

where we have:

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

More importantly, a momentum parameter updating mechanism is applied to facilitate the knowledge distillation:

$$\zeta^t = m \cdot \zeta^{t-1} + (1-m)\cdot\theta^t.$$



Cross-view contrastive learning regularizes our bootstrapping objective by contrasting between online representations of two views.

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

where the inter-view contrasting is defined as:

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

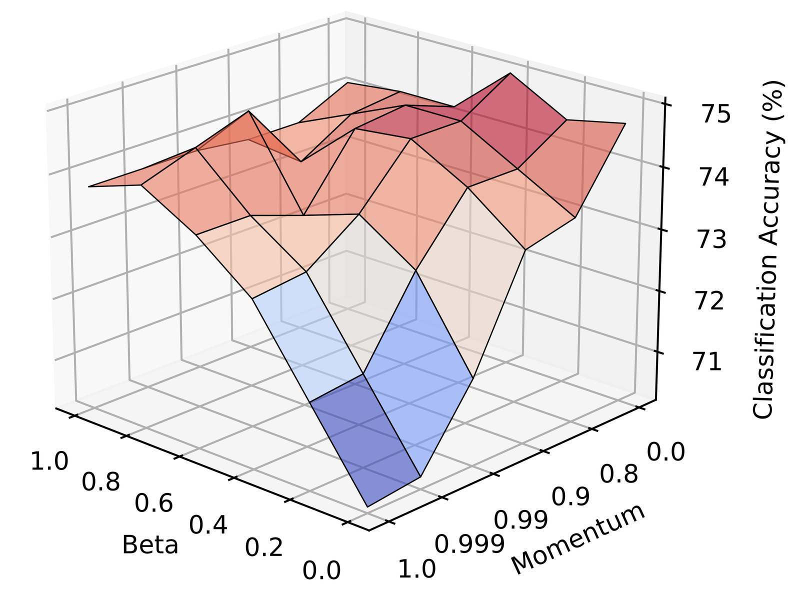
The intra-view contrasting is formulated as:

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

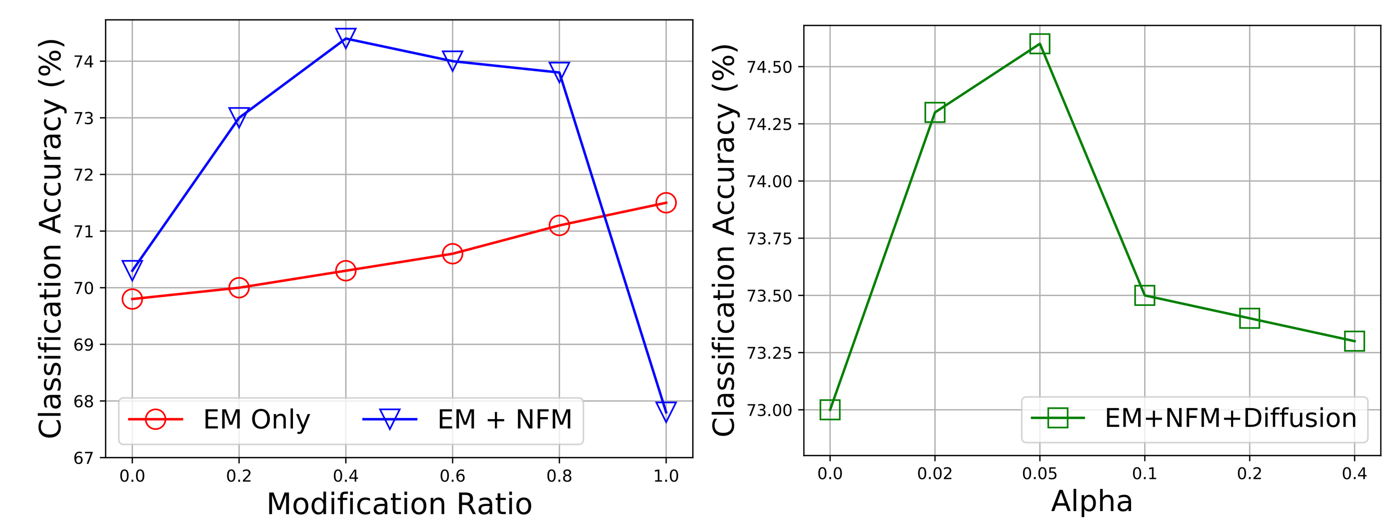
where $\Phi = \sum_{j=1}^{N}\mathbb{1}_{i\neq j}\exp(\text{sim}(h^{1\text{ or }2}_{v_i}, h^{1\text{ or }2}_{v_j}))$.

EXPERIMENTS

Information Used	Method	Cora	CiteSeer	PubMed	Amazon Photo	Coauthor CS
A, Y	LP	68.0	45.3	63.0	67.8±0.0	74.3 ±0.0
X, A, Y	Chebyshev	81.2	69.8	74.4	74.3±0.0	91.5 ±0.0
X, A, Y	GCN	81.5	70.3	79.0	87.3±1.0	91.8 ±0.1
X, A, Y	GAT	83.0 ±0.7	72.5 ±0.7	79.0 ±0.3	86.2 ±1.5	90.5 ±0.7
X, A, Y	SGC	81.0 ±0.0	71.9 ±0.1	78.9 ±0.0	86.4 ±0.0	91.0 ±0.0
X, A	DGI	81.7 ±0.6	71.5 ±0.7	77.3 ±0.6	83.1 ±0.5	90.0 ±0.3
X, A	GMI	82.7 ±0.2	73.0 ±0.3	**80.1 ±0.2**	85.1 ±0.1	91.0 ±0.0
X, A	MVGRL	82.9 ±0.7	72.6 ±0.7	79.4 ±0.3	87.3 ±0.3	91.3 ±0.1
X, A	GRACE	80.0 ±0.4	71.7 ±0.6	79.5 ±1.1	81.8 ±1.0	90.1 ±0.8
X, A	MERIT	**83.1 ±0.6**	**74.0 ±0.7**	**80.1 ±0.4**	**87.4 ±0.2**	**92.4 ±0.4**



Classification accuracies on CiteSeer with different β and m. A warmer color denotes a higher accuracy.



Classification accuracies on CiteSeer versus graph augmentation in varying types and degrees.

KEY REFERENCES

[1] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Pires, Zhaohan Guo, Mohammad Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. In *NIPS*, 2020.

[2] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep Graph Contrastive Representation Learning. In *ICML Workshop*, 2020.

CONTACT INFORMATION

Email Us

GitHub Site