ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888092

Graph Contrastive Learning with Decoupled Augmentation

Shihao Gao[1,5], Caoshuo Li[1], Cunli Mao[2,3], Xulong Zhang[5], Xiaoyang Qu[5], Taisong Jin[1,4*], Jianzong Wang[5]
[1]Key Laboratory of Multimedia Trusted Perception and Efficient Computing,
Ministry of Education of China, School of Informatics, Xiamen University, China
[2]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, China
[3]Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, China
[4]Key Laboratory of Oracle Bone Inscriptions Information Processing,
Ministry of Education of China, Anyang Normal University, China
[5]Ping An Technology Co., Ltd., Shenzhen, China

***Abstract*—Graph contrastive learning based on augmentation strategies has recently demonstrated remarkable performance. Existing methods typically jointly leverage attribute and structural augmentations to generate graph views, learning data invariance information through contrasting sample pairs. However, this joint approach may deviate from the expectation of semantically similar before and after augmentation. The propagation of attribute information in graphs usually occurs through their structure, meaning that structural and attribute augmentations can interfere with each other and potentially distort the graph's semantics. To address this, we propose a decoupled augmentation framework for graph contrastive learning, which eliminates the mutual interference between the two levels of augmentation while fully exploring graph information. Specifically, our framework employs separate encoders to learn data invariance under different augmentation levels, and it considers the positive gains generated between these levels. Experimental results on five public datasets show that the proposed method is more competitive than state-of-the-art approaches.**

***Index Terms*—Self-supervised learning, Graph representation, Contrastive learning**

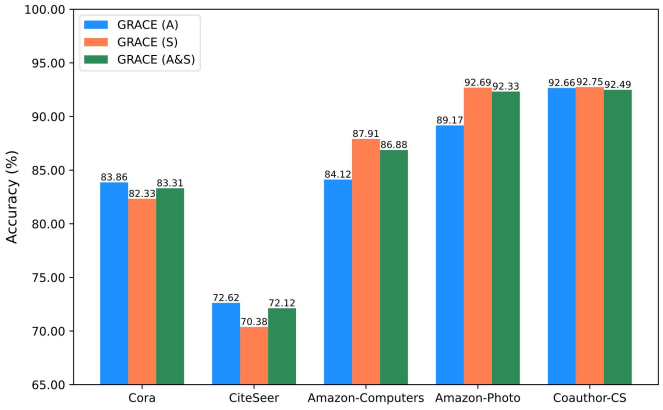


Fig. 1. Performance of different variants of GRACE for node classification on different datasets. GRACE is one of the most popular frameworks in GCL. Where (A), (S), and (A&S) represent using attribute augmentation only, using structural augmentation only, and jointly using two levels of augmentation, respectively.

I. Introduction

Graph neural networks (GNNs) are a powerful tool for analyzing graph-structured data, aiming to extract low-dimensional representations of nodes from structures and attributes. GNNs have demonstrated significant potential across various application areas [1, 2, 3, 4]. However, traditional GNNs [5, 6] are built in a supervised manner, requiring a large amount of costly labeled data for training. To reduce this dependency on labels, graph contrastive learning (GCL), a prominent form of self-supervised learning, has been proposed and has shown promising results in numerous downstream tasks [7, 8, 9].

A typical GCL method first applies joint augmentation of attributes and structures to generate two new views, then learns node representations by contrasting sample pairs from these views. The performance of graph contrastive models relies heavily on effective augmentation strategies, which have been a focal point for researchers. For instance, GRACE [10] constructs graph views by randomly removing edges and masking node attributes. GCA [11] refines GRACE by using node centrality as guiding information to preserve important links and attributes. CSGCL [12] introduces perturbations to edges and attributes, guided by community strength, to preserve differences among communities in the generated views.

Although these approaches have been successful, they all employ a strategy of jointly leveraging structural-level and attribute-level augmentations. Current GCL models [13, 14] typically use GNNs [5] as the base encoder, and attribute information to propagate based on the structure. The new views generated from this two-level joint augmentation may lead to significant semantic shifts compared to the input graph, resulting in suboptimal performance. Empirically, we observed that using a single-level augmentation could yield

*The corresponding author. jintaisong@xmu.edu.cn Supported by the National Natural Science Foundation of China (No. 62072386), Yunnan Provincial Major S&T Special Plan Project (No. 202402AD080001), Henan Key R&D Project (No. 231111212000), Open Foundation of Henan Key Lab of General Aviation Technology (No. ZHKF-230212), Key Lab of Oracle Information Processing of MOE (No. OIP2024E002), and Guangdong Key R&D Program (No. 2021B0101400003).

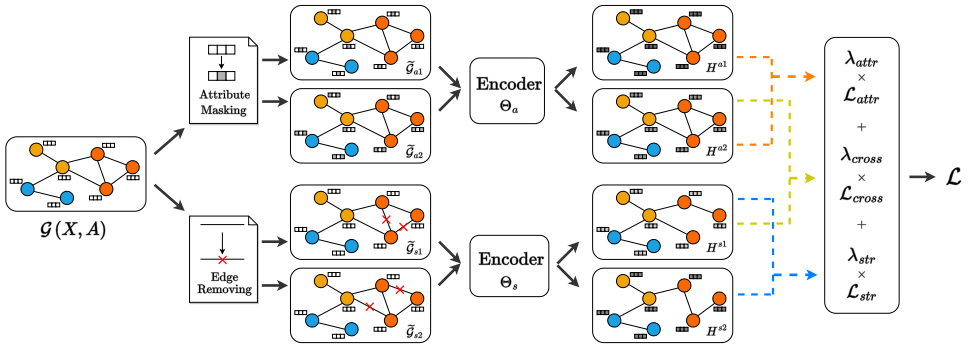


Fig. 2. The overview of our proposed framework: GDA. We first apply attribute augmentation and structural augmentation separately to the input graph to obtain new views at different levels. Then these graph views are fed into the encoder at the corresponding level to obtain node embeddings. GDA learns efficient node representations by minimizing attribute level contrastive loss $\mathcal{L}_{attr}$, structure level contrastive loss $\mathcal{L}_{str}$, and cross-level contrastive loss $\mathcal{L}_{cross}$ between H^{a2} and H^{s1}.

better performance for GCL models, as shown in Fig. 1. A natural question arises thereby: Is there an effective way to better leverage the rich invariant information provided by augmentations at both the attribute and structural levels?

To answer this question, we propose a novel Graph contrastive learning with Decoupled Augmentation framework, termed GDA. It mines graph information under different augmentation by decoupling attribute and structure augmentation. We first apply attribute and structural augmentations separately to the input graph and use two distinct encoders to capture invariant information at various augmentation levels. In addition, to align the representation between different levels and to mine the beneficial information between different levels, a novel cross-level contrastive loss is proposed. By jointly optimizing attribute level loss, structure level loss, and cross-level loss, our framework could learn effective node representations without manual labels. The contributions of this work can be summarized as follows:

- We empirically show that joint augmentation of both attribute and structure levels may be a suboptimal choice for GCL.
- We propose a novel GCL framework, GDA, which more effectively captures distinct aspects of invariance information provided by the two-level augmentation of attributes and structures.
- Experimental results demonstrate that our approach is more competitive than state-of-the-art approaches.

II. Methodology

In this section, we present the proposed GDA framework in detail. Fig. 2 shows the overall architecture of GDA.

A. Preliminaries and Notations

let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a graph, where $\mathcal{V} = \{v_1, v_2, \cdots, v_N\}$ represents the set of nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represent the set of edges. The attribute matrix is denoted as $X \in \mathbb{R}^{n \times d}$, where $x_i \in \mathbb{R}^d$ is the attribute vector of node v_i. The adjacency matrix is denoted as $A \in \{0,1\}^{n \times n}$, where $A_{ij} = 1$ iff $(v_i, v_j) \in \mathcal{E}$. Besides, in the self-supervised training setting, the class information Y of nodes is unknown.

Given a graph $\mathcal{G}$ with no labels, the goal of GCL is to learn one or more encoders and generate node embeddings that can be used in downstream tasks. At the start of training, GCL generates multiple graph views $(p_1, \ldots, p_K)$ by applying random perturbations to the input graph. Graph perturbations include attribute masking, edge removing, subgraph sampling, etc. Then views are fed into corresponding encoders $(f_1, \ldots, f_M)$ to obtain node embeddings $\left(h^1, \ldots, h^K\right)$ in the different views. These embeddings are used to train the encoders by optimizing contrastive objective that aims to maximize the mutual information $\mathcal{MI}\left(h^i, h^j\right)$ between the same instance in different views:

$$\max_{\{f_k\}_{k=1}^K} \frac{1}{\sum_{i \neq j} \sigma_{i,j}} \left[\sum_{i \neq j} \sigma_{i,j} \mathcal{MI}\left(h_i^{f_k}, h_j^{f_k}\right) \right]. \tag{1}$$

where $\sigma_{i,j} \in \{0,1\}$, and $\sigma_{i,j} = 1$ if the mutual information is computed between $h_i^{f_k}$ and $h_j^{f_k}$, and $\sigma_{i,j} = 0$ otherwise. The most commonly used optimization objective in GCL is InfoNCE [15], which is an estimator of the lower bound of mutual information.

B. Attribute-level Contrastive Learning

GCL typically uses GNNs as the base encoder, which propagates attribute information through the graph structure. When both attributes and structure are perturbed simultaneously, the semantics of the graph data may undergo significant changes. Moreover, specific attribute-level invariant information may be disrupted or even overshadowed by structural changes. Hence, we propose using a separate module to learn invariant information under attribute augmentation.

We adopt the widely used attribute masking technique for attribute-level augmentation. A two-layer GCN serves as the base encoder, which can be formally defined as follows:

$$f(X, A) = \sigma\left(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} X W^l\right), \tag{2}$$

where $\hat{A} = A + I$ is the adjacency matrix with self-loops, $\hat{D} = \sum_i \hat{A}_i$ is the degree matrix, $\sigma(\cdot)$ is a nonlinear activation function, and W^l is the trainable weight matrix.

The new views of the attribute level are fed into the encoder to obtain the node embeddings H^{a1}, H^{a2}. For node v_i, node embeddings h_i^{a1} and h_i^{a2} form the positive samples, and node embeddings of other nodes are regarded as negative samples. Therefore, the loss function of attribute-level contrastive learning can be defined by the InfoNCE loss as

$$\mathcal{L}_{attr}\left(v_i\right) = \log \frac{e^{\theta\left(h_i^{a1}, h_i^{a2}\right)/\tau}}{e^{\theta\left(h_i^{a1}, h_i^{a2}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{a1}, h_j^{a2}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{a1}, h_j^{a1}\right)/\tau}}. \tag{3}$$

where τ is the temperature parameter and $\theta\left(\cdot,\cdot\right)$ is the cosin similarity function.

C. Structure-level Contrastive Learning

The specific invariant information at the structural level may also be affected by changes in attributes. We similarly use a separate module to learn the specific invariant information under structural augmentation. We choose the widely used edge removing as the augmentation method for the structural level. Similar to attribute-level contrastive learning, a two-layer GCN is used as the encoder, but with different weights.

The new views of the structural level are fed into the corresponding encoder to obtain the node embeddings H^{s1}, H^{s2}. Given node embeddings h_i^{s1} and h_i^{s2} of node v_i, the loss function of structure-level contrastive learning can be defined as

$$\mathcal{L}_{str}\left(v_i\right) = \log \frac{e^{\theta\left(h_i^{s1}, h_i^{s2}\right)/\tau}}{e^{\theta\left(h_i^{s1}, h_i^{s2}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{s1}, h_j^{s2}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{s1}, h_j^{s1}\right)/\tau}}. \tag{4}$$

D. Cross-level Loss

To align the representations of different layers and leverage the positive gains between augmentations at different layers, we introduced a cross-level contrastive loss. The cross-level loss aims to align the representations of H^{a2} and H^{s1}, and we have observed that contrasting views from different levels can enhance the performance of the proposed framework.

$$\mathcal{L}_{cross}\left(v_i\right) = \log \frac{e^{\theta\left(h_i^{a2}, h_i^{s1}\right)/\tau}}{e^{\theta\left(h_i^{a2}, h_i^{s1}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{a2}, h_j^{s1}\right)/\tau} + \sum_{j \neq i} e^{\theta\left(h_i^{a2}, h_j^{a2}\right)/\tau}}. \tag{5}$$

E. Overall Objective Loss.

The overall objective loss of GDA is defined to be a sum of attribute-level loss, structure-level loss, and cross-level loss:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left[\lambda_1 \mathcal{L}_{attr}\left(v_i\right) + \lambda_2 \mathcal{L}_{str}\left(v_i\right) + \lambda_3 \mathcal{L}_{cross}\left(v_i\right)\right].$$

TABLE I
STATISTICS OF DATASETS USED IN OUR EXPERIMENTS.

Dataset	#Nodes	#Edges	#Features	#Classes
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
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

where λ_1, λ_2, and λ_3 are tuning hyperparameters to weight different loss terms, and different datasets usually correspond to different values.

III. EXPERIMENTS

A. Experiment Setting

Datasets. To enable a comprehensive comparison, we utilized five open graph datasets, including Cora [5], CiteSeer [5], Amazon-Computers [16], Amazon-Photo [16], and Coauthor-CS [16], to evaluate node classification performance. The statistics of the datasets are provided in Table I.

Baselines. We compared with classic unsupervised models including Logistic regression (LogReg) [17], and node2vec [18]. Additionally, we compared it with the recently proposed excellent self-supervised models including DGI [19], MVGRL [13], GRACE [10], GCA [14], BGRL [20], COSTA [7], and CSGL[12]. To better demonstrate the effectiveness of our proposed approach, we also compared it with supervised methods GCN [5] and GAT [6]. The official implementations were used to report the performance of all baselines.

Evaluation protocols. We follow the linear evaluation scheme proposed in [19], where each model is first trained in an unsupervised manner, and then the node representations outputted by the encoder are fed into a simple logistic regression classifier. All datasets are randomly split the remaining four datasets into 10%, 10%, and 80% for training, validation, and testing, respectively.

B. Node Classification

The experimental results are shown in Table II, where the best performance is highlighted in bold and the second best performance is underlined. It is observed that, apart from the Amazon-Computers dataset, GDA outperforms all self-supervised baselines on the other four datasets. We attribute this to its ability to effectively learn beneficial information brought by different levels of augmentation. Furthermore, we also note that the performance of most contrastive learning models surpasses traditional supervised methods, highlighting the superiority of contrastive learning.

Additionally, we observe that GDA achieved a significant performance improvement compared to GRACE. Both use the same perturbation augmentations at the attribute and structural levels but differ in that one uses joint augmentation while the other uses decoupled augmentation. This phenomenon may offer a new research direction for graph contrastive learning.

TABLE II
PERFORMANCE ON NODE CLASSIFICATION IN TERMS OF ACCURACY (%).

Method	Input	Cora	CiteSeer	Amazon-Computers	Amazon-Photo	Coauthor-CS
Sup.GCN	X,A,Y	81.50 ± 0.10	71.62 ± 0.26	86.51 ± 0.54	92.42 ± 0.22	93.03 ± 0.31
Sup.GAT	X,A,Y	82.34 ± 0.08	72.24 ± 0.15	86.93 ± 0.29	92.56 ± 0.35	92.32 ± 0.24
Raw (LogReg)	X	64.61 ± 0.22	65.77 ± 0.15	71.98 ± 0.00	78.53 ± 0.00	90.37 ± 0.00
node2vec	A	74.32 ± 0.20	52.28 ± 0.07	71.79 ± 0.05	85.08 ± 0.03	85.08 ± 0.03
DGI	X,A	82.63 ± 0.45	68.83 ± 0.71	83.95 ± 0.47	91.61 ± 0.22	92.15 ± 0.63
MVGRL	X,A	83.11 ± 0.12	73.33 ± 0.03	87.52 ± 0.11	91.74 ± 0.07	92.11 ± 0.12
GRACE	X,A	83.31 ± 0.25	72.12 ± 0.28	86.88 ± 0.25	92.33 ± 0.25	92.49 ± 0.03
GCA	X,A	82.89 ± 0.21	72.49 ± 0.31	87.85 ± 0.22	92.49 ± 0.18	93.06 ± 0.01
BGRL	X,A	83.80 ± 0.34	72.76 ± 0.16	87.37 ± 0.40	91.57 ± 0.44	92.59 ± 0.14
COSTA	X,A	84.32 ± 0.22	72.92 ± 0.31	88.32 ± 0.03	92.56 ± 0.45	92.95 ± 0.12
CSGL	X,A	84.27 ± 0.32	72.56 ± 0.20	**89.98 ± 0.26**	92.87 ± 0.33	93.17 ± 0.12
GDA	X,A	**85.42 ± 0.23**	**73.61 ± 0.22**	89.48 ± 0.29	**93.41 ± 0.32**	**93.29 ± 0.08**

C. Ablation Study

To evaluate the impact of main components on model performance, we conducted a series of ablation experiments. GDA-A, GDA-S, and GDA-C represent the removal of attribute-level contrastive learning, structure-level contrastive learning, and cross-level loss, respectively. The results shown in Table III, indicate that each component contributes positively to the model performance. Additionally, when the cross-level loss is removed, performance decreases across all three datasets, suggesting a positive gain between the augmentations at the two different levels.

TABLE III
ABLATION STUDY ON NODE CLASSIFICATION.

Variant	Cora	Amazon-Computers	Coauthor-CS
GDA-A	83.25 ± 0.10	89.24 ± 0.24	90.13 ± 0.04
GDA-S	82.97 ± 0.22	80.83 ± 0.21	91.25 ± 0.13
GDA-C	83.44 ± 0.24	88.94 ± 0.13	92.24 ± 0.17
GDA	**85.42 ± 0.23**	**89.48 ± 0.29**	**93.29 ± 0.08**

D. Case Study

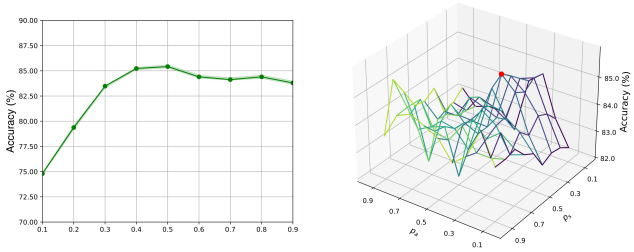
In principle, our framework allows the integration of various existing attribute and structure-level augmentation strategies. GCA is one of the most popular GCL frameworks and designs augmentations at two levels by leveraging node centrality. We replaced the augmentations in GDA with those from GCA and evaluated the performance of GDA (GCA) on three datasets. As shown in Table IV, the results demonstrate consistent improvements of GDA over the base GCA, further validating the potential of the proposed architecture.

E. Hyperparameter Analysis

In Fig. 3, we conduct a hyperparameter analysis of the perturbation rates for view generation (attribute masking rate and edge removing rate) and the temperature coefficient τ. Both hyperparameters are varied from 0.1 to 0.9. Fig. 3 (a) shows GDA's performance on the Cora dataset under different τ. The model performs best when τ is set to 0.5, with both larger and smaller values leading to performance degradation. Fig. 3 (b) illustrates GDA's performance on the Cora dataset across different perturbation rates. p_a and p_s represent the perturbation rates of the two levels respectively. The results indicate that our model is relatively insensitive to the perturbation hyperparameter, due to its ability to capture invariant information from various aspects of the data.

TABLE IV
THE PERFORMANCE OF THE GDA MODEL USING THE GCA AUGMENTATION STRATEGY IN TERMS OF ACCURACY.

Method	CiteSeer	Amazon-Computers	Amazon-Photo
GCA	72.49 ± 0.31	87.85 ± 0.22	92.49 ± 0.18
GDA (GCA)	**73.15 ± 0.19**	**88.75 ± 0.28**	**92.87 ± 0.21**



(a) temperature coefficient τ (b) perturbation rates

Fig. 3. Hyperparameter analysis on Cora.

IV. CONCLUSION

In this paper, we propose a novel graph contrastive representation framework with decoupled augmentation. Our model eliminates potential interference from the joint effects of attributes and structural levels by decoupling their augmentations. By effectively utilizing the rich information provided by different-level augmentations, the encoders can learn expressive node representations. Experimental results on five public datasets validate the effectiveness of the proposed method.

REFERENCES

[1] Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi, "A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning," *The VLDB Journal*, vol. 31, no. 5, pp. 1143–1168, 2022.

[2] Yangqin Jiang, Chao Huang, and Lianghao Huang, "Adaptive graph contrastive learning for recommendation," in *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, 2023, pp. 4252–4261.

[3] Jiahua Rao, Shuangjia Zheng, Sijie Mai, and Yuedong Yang, "Communicative subgraph representation learning for multi-relational inductive drug-gene interaction prediction," in *IJCAI*, 2022.

[4] Zhenxing Wu, Jike Wang, Hongyan Du, Dejun Jiang, Yu Kang, Dan Li, Peichen Pan, Yafeng Deng, Dongsheng Cao, Chang-Yu Hsieh, et al., "Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking," *Nature Communications*, vol. 14, no. 1, pp. 2585, 2023.

[5] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[6] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio, "Graph attention networks," in *ICML*, 2018.

[7] Yifei Zhang, Hao Zhu, Zixing Song, Piotr Koniusz, and Irwin King, "Costa: covariance-preserving feature augmentation for graph contrastive learning," in *KDD*, 2022, pp. 2524–2534.

[8] Yifei Zhang, Hao Zhu, Zixing Song, Piotr Koniusz, and Irwin King, "Spectral feature augmentation for graph contrastive learning and beyond," in *AAAI*, 2023, pp. 11289–11297.

[9] Wen-Zhi Li, Chang-Dong Wang, Hui Xiong, and Jian-Huang Lai, "Homogcl: Rethinking homophily in graph contrastive learning," in *KDD*, 2023, pp. 1341–1352.

[10] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang, "Deep graph contrastive representation learning," in *ICML*, 2020.

[11] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen, "Graph contrastive learning with augmentations," in *NeurIPS*, 2020, pp. 5812–5823.

[12] Han Chen, Ziwen Zhao, Yuhua Li, Yixiong Zou, Ruixuan Li, and Rui Zhang, "Csgcl: community-strength-enhanced graph contrastive learning," in *IJCAI*, 2023, pp. 2059–2067.

[13] Kaveh Hassani and Amir Hosein Khasahmadi, "Contrastive multi-view representation learning on graphs," in *PMLR*, 2020, pp. 4116–4126.

[14] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang, "Graph contrastive learning with adaptive augmentation," in *WWW*, 2021, pp. 2069–2080.

[15] Aaron van den Oord, Yazhe Li, and Oriol Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[16] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann, "Pitfalls of graph neural network evaluation," *arXiv preprint arXiv:1811.05868*, 2018.

[17] James MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.

[18] Aditya Grover and Jure Leskovec, "node2vec: Scalable feature learning for networks," in *KDD*, 2016, pp. 855–864.

[19] Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm, "Deep graph infomax," in *ICLR*, 2019.

[20] Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Rémi Munos, Petar Veličković, and Michal Valko, "Bootstrapped representation learning on graphs," in *ICLR*, 2021.