Domain-aware Node Representation Learning for Graph Out-of-Distribution Generalization

Yi Qiao*†, Yang Liu*†§, Qing He*†, Xiang Ao*†‡§
* Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China
Email: {liuyang2023,heqing,aoxiang}@ict.ac.cn
†University of Chinese Academy of Sciencs, Beijing 100049, China
Email: qiaoyi19@mails.ucas.ac.cn
‡Institute of Intelligent Computing Technology, Suzhou, CAS

***Abstract*—Graph Neural Networks (GNNs) have demonstrated impressive success across diverse fields when data satisfies in-distribution (ID) assumption. Nevertheless, GNN performance significantly declines in cases of distribution shifts between training and testing graph data. This degradation primarily stems from spurious correlations between irrelevant domain information and target labels in out-of-distribution (OOD) scenarios. Thus, maximizing the utilization of domain information becomes imperative. In light of this, we propose a novel approach named Domain-aware Node Representation Learning (DNRL), comprehensively incorporates domain information to bolster generalization capability. Specifically, DNRL selectively interpolates nodes with the same label but different domains, extending training data into unseen domains and alleviating the effects caused by domain-related spurious correlations. Futhermore, by introducing a domain-aware contrastive learning strategy, our method implicitly decouples domain information from node information to learn domain-independent node representations. Extensive experiments on graph out-of-distribution benchmarks demonstrate that DNRL can achieve effective OOD generalization performance across diverse domains.**

***Index Terms*—graph neural network, out-of-distribution, domain augmentation**

I. Introduction

Graph Neural Networks (GNNs) have achieved remarkable success on multiple graph learning tasks [1]–[6]. Most GNNs assume in-distribution (ID) data, yet in real-world scenarios, the testing distribution may deviate from the training distribution due to unobserved or uncontrolled shifts. This discrepancy can adversely affect GNN performance. Hence, it is imperative to learn GNNs with robust out-of-distribution (OOD) generalization capabilities.

Several efforts have been dedicated to enhance the generalization capability of GNNs. Invariant learning [7], [8] aims to capture invariant relationships between features and labels across different domains, with its efficacy contingent upon the domain diversity. Moreover, graph data augmentation [9] boost data diversity and quality by generating new training instances, thereby improving the model's generalization ability.

Taking a toy graph in Fig. 1 as an example, where node colors represent domain information, with shape labels y_1 and y_2 in the training set spuriously correlated with the colors blue and red, respectively. Then in the testing set, the model can easily be affected to misclassify node 1 as y_2 and node 2 as y_1 by the color. In such an OOD case, graph invariant learning methods [10] can capture the invariant relationship between feature and label but cannot avoid spurious correlation when the number of domains is limited in the training set. Traditional graph augmentation methods [9] do not use the domain information thus cannot avoid spurious correlation either. Our solution is to interpolate the nodes with the same shape but different colors. Consequently, we obtain node samples with diverse colors and the spurious correlation between color and labels would be eliminated for better OOD generalization.

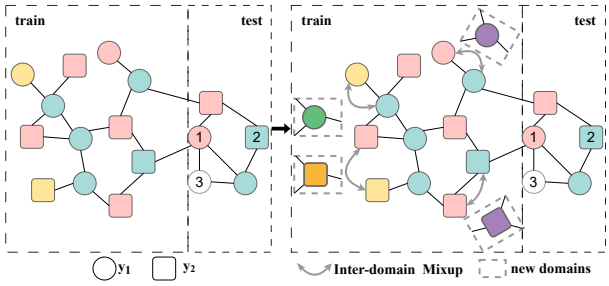


Fig. 1. A toy example of OOD node classification. *Left*: The label (shape of the nodes) is spuriously associated with domain (color). *Right*: Domain-aware Augmentation.

As the mainstream solution for graph OOD problem, existing graph invariant learning methods do not fully utilize domain information [9], [10]. While these methods incorporates domain information within the loss function, yet this approach risks constraining the model's capacity to learn optimal feature representations. Consequently, the model may overly emphasize domain disparities at the expense of fundamental feature characteristics. Moreover, the efficacy of invariant learning techniques can be influenced by the number of distinct domain.

In this study, we propose a domain-aware node representation learning method (DNRL) that utilizes domain information to eliminate the effect of spurious correlations. Specifically, our method exploits domain information at both the data level and the representation level to improve the generalization ability of GNNs. As depicted in Fig. 1(*Right*),

§ Corresponding authors

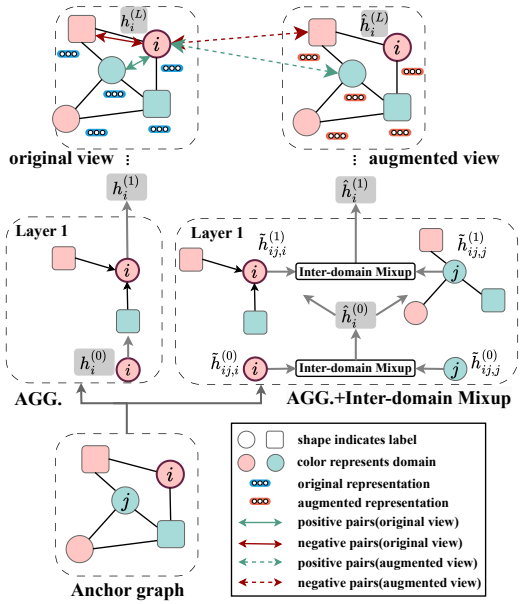


Fig. 2. The framework of DNRL on an example graph: Node shape indicates label and node color represents domain.

DNRL first performs inter-domain interpolation at the data level, facilitating the generation of new domains within the same class. This process enables the extension of training data into new domains and effectively eradicates domain-related spurious correlations. Furthermore, at the representation level, we introduce a domain-aware contrastive loss targeting domain correlations. DNRL implicitly disentangles irrelevant domain information from node representations, thereby enhancing OOD robustness. Extensive experimental results demonstrate that DNRL substantially boosts graph OOD performance while concurrently enhancing the model robustness.

II. Preliminary

In this work, let $\mathcal{X}$ be the input space, $\mathcal{Y}$ be the target label space, and $\mathcal{D}$ be the domains. Let $(x, y, d) \sim P(x, y, d)$ be a sample, where x denotes the input feature, y denotes the class label, and d denotes the domain.

Problem Statement. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a graph, where $\mathcal{V} = \{v_1, \ldots, v_N\}$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represent the node set and edge set respectively. Each node v_i is associated with a feature vector $x_i \in \mathbf{X}$ and a label $y_i \in \mathbf{Y}$. The adjacency matrix $\mathbf{A} \in \{0, 1\}^{N \times N}$ is defined such that $\mathbf{A}_{ij} = 1$ if and only if $(v_i, v_j) \in \mathcal{E}$. Consider the setting where one predicts the label $y \in \mathbf{Y}$ based on the input feature $x \in \mathbf{X}$. This paper focuses on the out-of-distribution node classification and learns a multi-classification function based on $\mathbf{D}_{\text{train}} = \{v_i | (x_i, y_i, d_i) \sim P_{\text{train}}\}$ to predict the unlabeled nodes in $\mathbf{D}_{\text{test}} = \{v_i | (x_i, y_i, d_i) \sim P_{\text{test}}\}$. $\mathbf{D}_{\text{train}}$ and $\mathbf{D}_{\text{test}}$ follow different distributions. Formally,

$$f_\theta : (\mathbf{X}, \mathbf{A}) \to \mathbf{Y}. \tag{1}$$

where θ represents the parameters of the model.

III. Methodology

This section gives a detailed presentation of DNRL, a domain-aware contrastive graph augmentation method. As shown in Fig. 2, we selectively interpolate pairs of nodes with the same label but different domains to generate samples spanning new domains. DNRL obtains node embeddings from two distinct perspectives: the original view and the interpolation-augmented view. Subsequently, both the original and augmented representations of adversarial samples are fed into the contrastive loss to learn domain-independent information. Afterwards, the model is optimized by minimizing the DNRL loss function, comprising the empirical risk minimization loss integrated with mixup and contrastive loss.

A. Inter-domain Mixup

We enhance domain generalization robustness through inter-domain mixup, selectively interpolating node pairs with the same label but different domains. We incorporate both attribute features and topological structure in this process. We randomly sample the Mixup weight λ from a Beta distribution [11] with a hyperparameter α.

Specifically, for v_i, we select v_j that has the same label but belongs to a different domain. Then we mix the node attributes of two nodes before the input layer:

$$\tilde{x}_{ij} = \lambda x_i + (1 - \lambda) x_j, if\ y_i = y_j\ and\ d_i \neq d_j. \tag{2}$$

Subsequently, we undertake topological-level interpolation [9]. In each layer of GNNs, message passing and aggregation are performed for v_i and v_j to derive the representation:

$$\begin{aligned} \tilde{h}_{ij,i}^{(l)} &= \text{AGG}\left(z_{ij}^{(l-1)}, \left\{h_k^{(l-1)} | k \in \mathcal{N}(i)\right\} | W^{(l)}\right), \\ \tilde{h}_{ij,j}^{(l)} &= \text{AGG}\left(z_{ij}^{(l-1)}, \left\{h_k^{(l-1)} | k \in \mathcal{N}(j)\right\} | W^{(l)}\right), \\ &\qquad if\ y_i = y_j\ and\ d_i \neq d_j, \end{aligned} \tag{3}$$

where $\mathcal{N}(i)$ denotes the set of neighbors of the node i and $W^{(l)}$ is the layer-specific trainable weight matrix. Here, h denotes the intermediate node representations at each layer, while z represents the final node representations obtained after layer-wise interpolation. And we interpolate the representations obtained from the two topologies before the next layer as shown in Eq.(4), where $z_{ij}^{(0)} = \tilde{x}_{ij}$.

$$z_i^{(l)} = z_j^{(l)} = z_{ij}^{(l)} = \lambda \tilde{h}_{ij,i}^{(l)} + (1 - \lambda) \tilde{h}_{ij,j}^{(l)}, \tag{4}$$

Nevertheless, under practical scenarios, it is not always feasible for each training node sample v_i to identify node v_j with the same label but different domain. For such nodes, we instead encode them by aggregating the interpolated representation z of their neighbors:

$$z_i^{(l)} = \text{AGG}\left(z_i^{(l-1)}, \left\{z_k^{(l-1)} | k \in \mathcal{N}(i)\right\} | W^{(l)}\right). \tag{5}$$

where $z_i^{(0)} = x_i$ holds. In this way, the interpolated representation with richer domain information is passed to node (x_i, y_i, d_i) through its neighborhood $\mathcal{N}(i)$.

Consequently, the loss of interpolated nodes is:

$$\mathcal{L}_m = \mathbb{E}_{(x_i,y_i,d_i)\sim P_{\text{train}}}[\ell_m(z_i, \tilde{y}_i)]. \tag{6}$$

where $\tilde{y}_i = \lambda y_i + (1-\lambda)y_j = y_i$ and ℓ_m is the cross-entropy loss. Through inter-domain mixup, we shift the domain distribution of the newly generated data away from the original domain distribution, extending it to unseen domains. This effectively mitigates domain-related spurious correlations.

B. Domain-aware Contrastive Strategy

Current contrastive methods exhibit limited generalization capabilities under distribution shifts since they do not consider the domain information. To mitigate this issue, we incorporate the domain information into the contrastive learning process, which selectively identifies positive and negative samples and generates corresponding augmented sample representations for graph OOD node classification.

Positive Sample Selection For v_i in $\mathbf{D}_{\text{train}}$, we identify positive samples from $\mathbf{D}_{\text{train}}$ by selecting those nodes with the same label but different domains, and add them into the set $\mathcal{P}_i = \{v_j \mid v_j \in \mathbf{D}_{\text{train}}, y_i = y_j, d_i \neq d_j\}$.

In this way, data points with different domain information within a shared label space are brought closer together, so that the model ignores domain-specific information unrelated to labels and better utilizes label-related feature information.

Negative Sample Selection For v_i in $\mathbf{D}_{\text{train}}$, we identify negative samples from $\mathbf{D}_{\text{train}}$ by selecting those nodes with the same domain but different labels, and add them into the set $\mathcal{N}_i = \{v_j \mid v_j \in \mathbf{D}_{\text{train}}, y_i \neq y_j, d_i = d_j\}$.

This not only induces greater separation in the embedding space for samples with different labels but also leads the model to disregard spurious domain correlations.

In particular, we acquire node embeddings $\mathbf{H}_{\mathcal{P}_i}$ and $\mathbf{H}_{\mathcal{N}_i}$ for both positive samples $\mathcal{P}_i$ and negative samples $\mathcal{N}_i$ via a GNN encoder, serving as the original view. Additionally, due to the design of Inter-domain Mixup in section III-A, the interpolated node representation $\mathbf{Z}_{\mathcal{P}_i}$ and $\mathbf{Z}_{\mathcal{N}_i}$ can effectively serve as the interpolation-augmented view. Inter-domain mixup selects pairs with the same label but different domains for interpolation, which ensures the node categories for both positive and negative samples remain unaltered after interpolation. Formally, we denote $\mu(h_i, h_j) = s(g(h_i), g(h_j))$, where s is the cosine similiarity and g is a two-layer multilayer perceptron (MLP) to enhance the expression power of embedding [12]. We define the sum of similarities between node v_i and set $\mathbf{H}$ as:

$$S_i(\mathbf{H}) = \sum_{h_j \in \mathbf{H}} e^{\mu(h_i,h_j)/\tau}, \tag{7}$$

where τ is a temperature parameter. Then the contrastive loss for node v_i is shown in Eq.(8).

$$\ell_c(v_i) = -\log \frac{S_i(\mathbf{H}_{\mathcal{P}_\mathbf{i}}) + S_i(\mathbf{Z}_{\mathcal{P}_\mathbf{i}})}{S_i(\mathbf{H}_{\mathcal{P}_\mathbf{i}}) + S_i(\mathbf{Z}_{\mathcal{P}_\mathbf{i}}) + S_i(\mathbf{H}_{\mathcal{N}_\mathbf{i}}) + S_i(\mathbf{Z}_{\mathcal{N}_\mathbf{i}})}, \tag{8}$$

The contrastive loss in DNRL is defined as the average loss over all training nodes as shown in Eq.(9).

$$\mathcal{L}_c = \mathbb{E}_{(x_i,y_i,d_i)\sim P_{\text{train}}}[\ell_c(v_i)]. \tag{9}$$

By adding the contrastive loss L_c to the classification loss L_m, we optimize the parameter θ of f_θ to minimize the overall loss as shown in Eq.(10).

$$\mathcal{L} = \mathcal{L}_m + \eta\mathcal{L}_c. \tag{10}$$

IV. Experiments

A. Experimental Details

1) Datasets: We assessed the generalization capability of DNRL on the graph OOD benchmark GOOD [13], which comprises three datasets—GOOD-Cora, GOOD-WebKB, and GOOD-CBAS—each involving two types of distributional shifts: covariate shift and concept shift. The first one is citation netorks, where nodes are documents and edges are citation links. In GOOD-WebKB, a node represents a webpage and edges are hyperlinks between webpages. GOOD-CBAS is a synthetic dataset modified from BA-Shapes [14].

2) Baselines: Using the empirical risk minimization (ERM) as the initial baseline, we compare our method DNRL with multiple invariant learning methods, including IRM [7], VREx [15], GroupDRO [16], SRGNN [17] and EERM [10]. Besides, two typical domain adaption methods DANN [18] and DeepCoral [19] are used for comparison. We further compare against the graph data augmentation method Mixup [9].

3) Experimental setup: The code is implemented based on PyTorch Geometric. To ensure fair comparisons, we use a three-layer GCN with ReLU activation and a 300-dimensional hidden layer as the backbone for all methods. The ReLU activation function is employed, and the hidden layer dimension is set to 300. The training process consists of 100 epochs for three real-world datasets, while the synthetic dataset GOOD-CBAS runs for 200 epochs. The batch size is selected from the set $\{1000, 4096\}$, and the learning rate is chosen from $\{1e-3, 5e-3\}$. Performance is assessed using the **Accuracy** metric, with higher scores indicating better performance.

B. Overall Performance

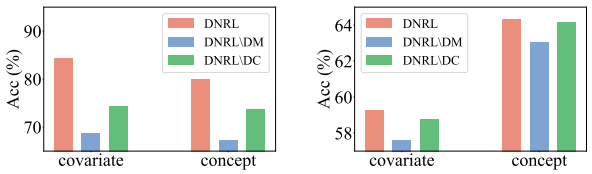
In TABLE I, we report the accuracy results, revealing following findings: (1) DNRL consistently achieves the highest average accuracy over three datasets, outperforming baseline methods. This suggests that our method exhibits robust domain generalization capabilities, effectively handling various out-of-distribution (OOD) scenarios and shift types. (2) Invariant learning methods, which aim to minimize risk across domains during the loss function stage, exhibit inconsistent performance across different datasets. For instance, while the IRM method achieves high accuracy on GOOD-WebKB, it fails on GOOD-Cora. This observation highlights the challenge of fully and effectively leveraging domain information. (3) Despite suboptimal performance in a minority of cases, our methods yield comparable results and consistently outperform the data-augmented method Mixup. These findings underscore the robust generalization capabilities of DNRL.

TABLE I
OOD PERFORMANCE COMPARISONS UNDER COVARIATE (COV.) AND CONCEPT (CON.) SHIFT

Dataset	GOOD-Cora						GOOD-WebKB			GOOD-CBAS		
Domain	word			degree			university			color		
Shift	cov.	con.	avg.	cov.	con.	avg.	cov.	con.	avg.	cov.	con.	avg.
ERM	65.06	63.93	64.50	55.12	60.68	57.90	15.08	24.77	19.93	78.57	81.43	80.00
IRM	64.72	63.93	64.33	56.34	61.50	58.92	17.46	28.44	22.95	78.57	80.71	79.64
VREx	63.77	64.44	64.11	54.95	60.55	57.75	15.87	27.52	21.70	77.14	**83.57**	80.36
GroupDRO	64.53	64.10	64.32	54.95	60.80	57.88	15.87	27.52	21.70	80.00	82.86	81.43
SRGNN	64.53	64.93	64.73	56.11	61.00	58.56	11.11	28.44	19.78	71.43	82.86	77.15
EERM	65.06	62.84	63.95	56.34	58.25	57.30	17.06	27.52	22.29	72.86	65.71	69.29
DANN	65.06	63.96	64.51	55.15	60.55	57.85	14.29	25.69	19.99	78.57	**83.57**	81.07
DeepCoral	65.04	63.98	64.51	55.82	60.55	58.19	15.87	27.52	21.70	78.57	81.43	80.00
Mixup	63.74	64.60	64.17	56.31	63.23	59.77	**19.84**	31.19	25.52	77.14	64.29	70.72
DNRL	**66.17**	**65.41**	**65.79**	**59.24**	**64.34**	**61.79**	**19.84**	**32.11**	**25.98**	**84.29**	80.00	**82.15**

C. Ablation Study

We conduct two ablation tests to evaluate the validity of each module. Initially, we evaluate the efficacy of inter-domain node mixup by switching it to random interpolation, denoted as "DNRL\DM". As illustrated in Fig. 3(a) and Fig. 3(b), DNRL is more effective in all cases, indicating that Inter-domain succeeds in extending the training data to invisible domains and improves the generalization of the model. The removal of the contrast strategy, denoted by "DNRL\DC", results in a consistent decrease in the model's performance across both datasets. This suggests that the model encounters challenges in disentangling irrelevant domain information.

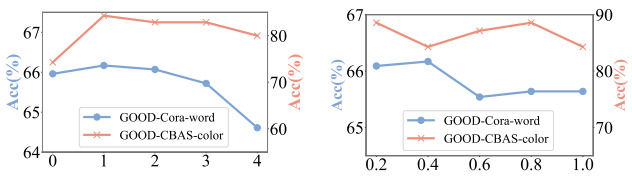


(a) GOOD-CBAS-color (b) GOOD-Cora-degree

Fig. 3. The validity of Inter-domain Mixup and Contrastive strategy with domain information.

D. Sensitivity Analysis

Fig. 4(a) illustrates the impact of the loss control hyper-parameter η. DNRL exhibits a consistent trend across both datasets, highlighting its stability. Additionally, in Fig. 4(b), we assess the sensitivity of the hyperparameter α, which regulates the interpolation ratio. Our results indicate that DNRL exhibits robustness across various interpolation ratios.



(a) Loss balance η (b) Interp. ratio α

Fig. 4. Sensitivity of the balance and interpolation ratio.

V. RELATED WORK

To improve the generalization ability of GNNs [20], [21] under OOD scenarios, prior works have proposed various strategies to learn node representations. Invariant learning [7], [8], [17], [22], [23], positioned as the mainstream method for addressing OOD challenges, aims to exploit the invariant relationships between features and labels across different distributions while disregarding the spurious correlations. Its generalization ability relies on the diversity of training environments. Data augmentation techniques [9], [24] contribute to enriching the distribution of training data. The OOD generalization ability of graph models rely on the diversity and quality of training data [25], thus, data augmentation methods are effective in solving the OOD generalization problem. Other methods [26]–[28] construct new graph network structures based on the causal inference theory, facilitating the generation of node and graph representations. Furthermore, recent research [29] has explored the application of unsupervised learning for OOD generalization in graph models, specifically through the framework of information bottlenecks.

VI. CONCLUSION

This paper focuses on the problem of out-of-distribution generalization for node classification on graph. We introduce DNRL, a novel model that leverages domain information in both data and representation spaces. Employing inter-domain interpolation, DNRL extends data to new domains, while a domain-aware contrastive strategy enables GNNs to learn domain-independent node representations. This approach mitigates spurious correlations, enhancing the model's generalization capacity.

ACKNOWLEDGMENT

The research work is supported by National Key R&D Plan No. 2022YFC3303302, the National Natural Science Foundation of China under Grant No. 62476263, No. 62406307.

Xiang Ao is also supported by Beijing Nova Program 20230484430 and the Innovation Funding of ICT, CAS under Grant No. E461060.

Yang Liu is also supported by the China Postdoctoral Science Foundation under Grant No. 2023M743567.

REFERENCES

[1] Y. Liu, X. Ao, Z. Qin, J. Chi, J. Feng, H. Yang, and Q. He, "Pick and choose: a gnn-based imbalanced learning approach for fraud detection," in *Proceedings of the web conference 2021*, 2021, pp. 3168–3177.

[2] J. Zhang, R. Xue, W. Fan, X. Xu, Q. Li, J. Pei, and X. Liu, "Linear-time graph neural networks for scalable recommendations," in *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 3533–3544.

[3] Y. Liu, X. Ao, F. Feng, and Q. He, "Ud-gnn: Uncertainty-aware debiased training on semi-homophilous graphs," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 1131–1140.

[4] K. Li, Y. Liu, X. Ao, and Q. He, "Revisiting graph adversarial attack and defense from a data distribution perspective," in *The Eleventh International Conference on Learning Representations*, 2023.

[5] Z. Qin, Y. Liu, Q. He, and X. Ao, "Explainable graph-based fraud detection via neural meta-graph search," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 4414–4418.

[6] Z. Gong, G. Wang, Y. Sun, Q. Liu, Y. Ning, H. Xiong, and J. Peng, "Beyond homophily: Robust graph anomaly detection via neural sparsification." in *IJCAI*, 2023, pp. 2104–2113.

[7] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[8] H. Li, Z. Zhang, X. Wang, and W. Zhu, "Learning invariant graph representations for out-of-distribution generalization," *Advances in Neural Information Processing Systems*, vol. 35, pp. 11 828–11 841, 2022.

[9] Y. Wang, W. Wang, Y. Liang, Y. Cai, and B. Hooi, "Mixup for node and graph classification," in *Proceedings of the Web Conference 2021*, 2021, pp. 3663–3674.

[10] Q. Wu, H. Zhang, J. Yan, and D. Wipf, "Handling distribution shifts on graphs: An invariance perspective," *arXiv preprint arXiv:2202.02466*, 2022.

[11] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC press, 2004.

[12] M. Tschannen, J. Djolonga, P. K. Rubenstein, S. Gelly, and M. Lucic, "On mutual information maximization for representation learning," *arXiv preprint arXiv:1907.13625*, 2019.

[13] S. Gui, X. Li, L. Wang, and S. Ji, "Good: A graph out-of-distribution benchmark," *Advances in Neural Information Processing Systems*, vol. 35, pp. 2059–2073, 2022.

[14] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," *Advances in neural information processing systems*, vol. 32, 2019.

[15] D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, and A. Courville, "Out-of-distribution generalization via risk extrapolation (rex)," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5815–5826.

[16] S. Sagawa, P. W. Koh, T. B. Hashimoto, and P. Liang, "Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization," *arXiv preprint arXiv:1911.08731*, 2019.

[17] Q. Zhu, N. Ponomareva, J. Han, and B. Perozzi, "Shift-robust gnns: Overcoming the limitations of localized graph training data," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 965–27 977, 2021.

[18] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[19] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 443–450.

[20] K. Li, Y. Chen, Y. Liu, J. Wang, Q. He, M. Cheng, and X. Ao, "Boosting the adversarial robustness of graph neural networks: An ood perspective," in *The Twelfth International Conference on Learning Representations*.

[21] Z. Gong and Y. Sun, "An energy-centric framework for category-free out-of-distribution node detection in graphs," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 908–919.

[22] S. Chang, Y. Zhang, M. Yu, and T. Jaakkola, "Invariant rationalization," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1448–1458.

[23] Y. Liu, X. Ao, F. Feng, Y. Ma, K. Li, T.-S. Chua, and Q. He, "Flood: A flexible invariant learning framework for out-of-distribution generalization on graphs," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 1548–1558.

[24] T. Zhao, Y. Liu, L. Neves, O. Woodford, M. Jiang, and N. Shah, "Data augmentation for graph neural networks," in *Proceedings of the aaai conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 015–11 023.

[25] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. Yu, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[26] H. Li, X. Wang, Z. Zhang, and W. Zhu, "Ood-gnn: Out-of-distribution generalized graph neural network," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[27] S. Fan, X. Wang, C. Shi, P. Cui, and B. Wang, "Generalizing graph neural networks on out-of-distribution graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[28] Q. Wu, F. Nie, C. Yang, T. Bao, and J. Yan, "Graph out-of-distribution generalization via causal intervention," in *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 850–860.

[29] Y. Zhu, H. Shi, Z. Zhang, and S. Tang, "Mario: Model agnostic recipe for improving ood generalization of graph contrastive learning," in *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 300–311.