

A Sample-driven Selection Framework: Towards Graph Contrastive Networks with Reinforcement Learning

Anonymous Authors

1 A. OVERVIEW OF DATA AUGMENTATIONS FOR GRAPHS

Table 1 shows the overview of four data augmentations for graphs. For different domains of graph datasets, we adopt [15] to strategically select data augmentations. Specifically, for Molecules and Bioinformatics, we employ two strategies as: node dropping and subgraph. For Social networks, we use all data augmentation strategies.

2 B. ALGORITHM

The entire learning process is described in **Algorithm 1**.

Algorithm 1: SDG Algorithm

Input: Graph dataset $\mathcal{D} = \{B_1, B_2, \dots, B_N\}; \gamma; \mathbf{W}(\text{SDG});$
Graph representation Φ_{B_j} and $\Phi_{A_j}; d(\cdot, \cdot);$ epochs L .

Output: Updated \mathbf{W} .

```
1 Initialize  $\mathbf{W}$  with standard Gaussian distribution;
2 while not converging do
3    $\Sigma = 0;$ 
4   while  $k < L$  do
5      $\Sigma_r = 0;$ 
6      $\mathcal{D} = \{B_1, B_2, \dots, B_N\};$ 
7     for  $B_j \in \mathcal{D}$  do
8        $\Phi_{B_j}^{s_j} \leftarrow \Theta_{j-1}(B_j); \Phi_{A_j}^{s_j} \leftarrow \Theta_{j-1}(A_j);$ 
9       On current bag state  $s_j$ ,
10       $D_j \leftarrow \mathbf{W}(\Phi_{A_j}^{s_j});$ 
11      select  $\Phi_{\hat{B}_j}^{s_j}$  from  $\Phi_{B_j}^{s_j}$  via  $D_j$  (take action  $a_j$ );
12       $r(s_{j-1}, a, s_j) \leftarrow d(\Phi_{\hat{B}_{j-1}}^{s_{j-1}}, \Phi_{\hat{A}_{j-1}}^{s_{j-1}}) - \gamma d(\Phi_{\hat{B}_j}^{s_j}, \Phi_{\hat{A}_j}^{s_j});$ 
13       $\Sigma_r \leftarrow \Sigma_r + \gamma^{j-1} r(s_{j-1}, a_j, s_j)$ 
14    end
15     $\Sigma \leftarrow \Sigma + \Sigma_r^N \nabla_{\omega} \log \pi_{\mathbf{W}}(a_j^k | s_j^k) \Sigma_r;$ 
16  end
17   $\nabla_{\omega} \tilde{J}(\mathbf{W}) \leftarrow \frac{1}{L} \Sigma;$ 
18   $\omega \leftarrow \omega + \tau \nabla_{\omega} \tilde{J}(\omega);$ 
19 end
```

3 C. DATASETS

To evaluate our model, we conduct experiments on eight widely used benchmark datasets¹ in three fields, including three molecules datasets: MUTAG, PTC, NCI-1, three social network datasets: REDDIT-BIN, IMDB-BINARY (IMDB-B), IMDB-MULTI (IMDB-M), and two

bioinformatics datasets: PROTEINS, D&D. More details about the data statistics and properties can be seen in **Table 2**.

4 D. BASELINE METHODS

The following models, including the state-of-art and closely related works, are used as representative baselines to evaluate the performance of the proposed model.

- **HGCL** [6]: HGCL investigates the hierarchical structural semantics of a graph at both node and graph levels.
- **AD-GCL** [9]: AD-GCL enables GNNs to avoid capturing redundant information during the training by optimizing adversarial graph augmentation strategies used in GCL.
- **SimGRACE** [11]: SimGRACE does not require data augmentations, and is inspired by the observation that graph data can preserve their semantics well during encoder perturbations.
- **GraphMAE** [5]: GraphMAE that mitigates these issues for generative self-supervised graph pretraining, and simple graph autoencoder with careful designs-can consistently generate outperformance over both contrastive and generative state-of-the-art baselines.
- **LaGraph** [12]: LaGraph is a theoretically grounded predictive SSL framework based on latent graph prediction, and provides explanations for recent successes of predictive models that include invariance-based objectives.
- **CuCo** [1]: CuCo propose a novel curriculum contrastive learning based graph representation learning model, which effectively combines curriculum learning and contrastive learning.
- **GraphCL** [15]: GraphCL designs four types of graph data augmentations, each of which imposes certain prior over graph data and parameterized for the extent and pattern.
- **InfoGraph** [8]: InfoGraph maximizes the mutual information between the graph-level representation and the representations of substructures of different scales (e.g., nodes, edges, triangles).
- **JOAO** [14]: JOAO is to automate the augmentation selection when performing contrastive learning on specific graph data.
- **LG2AR** [4]: The work is an end-to-end automatic graph augmentation framework that helps encoders learn generalizable representations on both node and graph levels.
- **GCN** [7]: GCN based on an efficient variant of convolutional neural networks which operate directly on graphs via a localized first-order approximation of spectral graph convolutions.
- **GraphSAGE** [3]: GraphSAGE, a general inductive framework that leverages node feature information (e.g., text attributes) to efficiently generate node embeddings for previously unseen data.
- **GIN** [13]: GIN show that its discriminative/representational power is equal to the power of the WL test.
- **GAT** [10]: GAT is a novel neural network architecture that operate on graph-structured data, leveraging masked self-attentional

¹Eight widely used datasets are publicly available at <https://ls11-www.cs.tudortmund.de/staff/morris/graphkerneldatasets>.

Table 1: Overview of data augmentations for graphs

Data augmentation	Type	Detailed Description
Node dropping	Nodes, edges	randomly discarding a certain portion of nodes along with their connections
Edge perturbation	Edges	randomly adding or dropping certain ratio of edges
Attribute masking	Nodes	recovering masked node attributes using their context information
Subgraph	Nodes, edges	samples a subgraph from \mathcal{G} using random walk

Table 2: Properties of the Biological and Social Network Datasets.

Dataset	Classes	Size	Avg.Nodes	Label	Category
MUTAG	2	188	17.9	7	Small molecules
PTC	2	344	25.5	19	Small molecules
NCI-1	2	4110	29.8	37	Small molecules
PROTEINS	2	1113	39.1	3	Bioinformatics
D&D	2	1178	284.31	82	Bioinformatics
IMDB-BINARY	2	1000	19.77	-	Social
IMDB-MULTI	3	1500	13	-	Social
REDDIT-BINARY	2	2000	429.61	-	Social

layers to address the shortcomings of prior methods based on graph convolutions or their approximations.

5 E. OTHER EXPERIMENTS

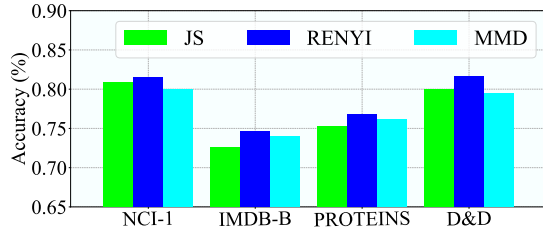


Figure 1: The results of different DDM.

5.0.1 Analysis on DDM. We summarize the results for three different distribution discrepancy measurements (DDM) calculated as a reinforcement reward to automatically guide the model to select appropriate samples. To evaluate the effect of DDM, we choose datasets: NCI-1, IMDB-B, PROTEINS, D&D as examples and conduct experiments. As shown in **Figure 1**, it is evident that three DDM achieve better results across four datasets, especially the RENYI that yields superior performance compared to the other two methods. Meanwhile, JS and MMD have similar performance. Finally, we conclude that three DDM methods are all suitable as reinforcement rewards to guide our model so as to achieve better performance.

5.0.2 Analysis on the Position of Selection. We conduct this experiment to explore the impact of different positions on negative sample selection, as shown in **Figure 2**. There are two lines: **A line**

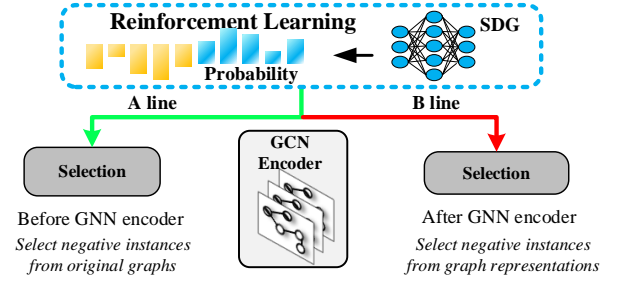


Figure 2: Two positions of selection.

Table 3: Analysis on the position of selection

Ours	NCI-1	IMDB-B	PROTEINS
GraphSaSe-A	78.8%	73.1%	75.9%
GraphSaSe-B	81.5%	75.8%	77.1%

represents that we select negative samples from original graphs in a batch before conducting GNN embedding. **B line** denotes that we select negative samples from graph representation after conducting GNN embedding. The results in **Table 3** show that the B line achieves better performance compared to A line. We conjecture that suppose positive pairs may have a considerable distance initially, but they become close to each other in the latent space after the GNN encoder. Thus, we believe that employing selection for the negative samples after the GNN encoder could further improve the performance.

5.0.3 Hyperparameters Analysis. In this section, we present the sensitivity analysis of two critical hyper-parameters employed by the GraphSaSe model, namely, the hyper-parameter λ to control the magnitude of RL task, and the number of layers N of GNNs. For brevity, we conducted the experiments on three datasets: NCI-1, IMDB-B, and PROTEINS. Additionally, the default parameter settings of the analysis are $\lambda=0.05$, and $N=4$. To be strict, when we tested the hyper-parameters specific to GraphSaSe, the other parameters were set to the default settings.

- **The impact of the hyper-parameter λ :** We analyze the sensitivity to the graph embedding dimension size $\lambda \in \{0.01, 0.05, 0.1, 0.5\}$. **Figure 3** demonstrates that GraphSaSe with different λ all achieves good results in general, which verifies that our GraphSaSe is not

very sensitive to hyper-parameter λ , stable over hyper-parameters in most cases. To get better performance, we select $\lambda=0.05$.

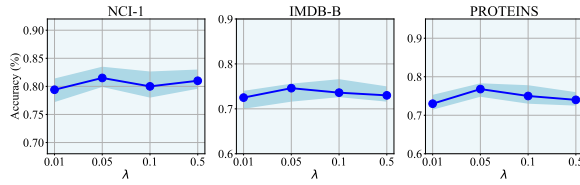


Figure 3: The hyper-parameter λ .

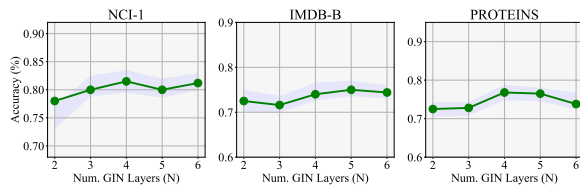


Figure 4: The number layers N of GNNs.

- **The impact of the hyper-parameter N :** Next, we examine the effects of the number layers N of GNNs, which is a hyper-parameter for the graph encoder. We select different number of layers $N \in \{2, 3, 4, 5, 6\}$. The results are shown in Figure 4. We observe that $N = 3$ generally gives the best performance, and the performance is rather stable over various datasets. Although $N = 4$ brings limited performance improvement on IMDB-B than $N = 3$, the computational complexity is more in calculating. So we choose $N = 3$ to reduce the computational complexity and keep the rather better performance.

REFERENCES

- [1] Guanyi Chu, Xiao Wang, Chuan Shi, and Xunqiang Jiang. 2021. CuCo: Graph Representation with Curriculum Contrastive Learning. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, Zhi-Hua Zhou (Ed.). ijcai.org, 2300–2306. <https://doi.org/10.24963/ijcai.2021/317>
- [2] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Ávila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. 2020. Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). <https://proceedings.neurips.cc/paper/2020/hash/f3ada80d5c4ee70142b17b8192b2958e-Abstract.html>
- [3] William L. Hamilton, Zitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 1024–1034. <https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7e9ea9-Abstract.html>
- [4] Kaveh Hassani and Amir Hosein Khas Ahmadi. 2022. Learning Graph Augmentations to Learn Graph Representations. *CoRR* abs/2201.09830 (2022). arXiv:2201.09830 <https://arxiv.org/abs/2201.09830>
- [5] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. 2022. GraphMAE: Self-Supervised Masked Graph Autoencoders. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, Aidong Zhang and Huzefa Rangwala (Eds.). ACM, 594–604. <https://doi.org/10.1145/3534678.3539321>
- [6] Wei Ju, Yiyang Gu, Xiao Luo, Yi-Fan Wang, Haochen Yuan, Huasong Zhong, and Ming Zhang. 2023. Unsupervised graph-level representation learning with hierarchical contrasts. *Neural Networks* 158 (2023), 359–368. <https://doi.org/10.1016/j.neunet.2022.11.019>
- [7] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. <https://openreview.net/forum?id=SJU4ayYgl>
- [8] Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. 2020. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. <https://openreview.net/forum?id=r1lF2NYvH>
- [9] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. 2021. Adversarial Graph Augmentation to Improve Graph Contrastive Learning. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.). 15920–15933. <https://proceedings.neurips.cc/paper/2021/hash/854f1fb6f65734d9e49f708d6cd84ad6-Abstract.html>
- [10] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph Attention Networks. *CoRR* abs/1710.10903 (2017). arXiv:1710.10903 <http://arxiv.org/abs/1710.10903>
- [11] Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z. Li. 2022. SimGRACE: A Simple Framework for Graph Contrastive Learning without Data Augmentation. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini (Eds.). ACM, 1070–1079. <https://doi.org/10.1145/3485447.3512156>
- [12] Yaochen Xie, Zhao Xu, and Shuiwang Ji. 2022. Self-Supervised Representation Learning via Latent Graph Prediction. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA (Proceedings of Machine Learning Research, Vol. 162)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (Eds.). PMLR, 24460–24477. <https://proceedings.mlr.press/v162/xie22e.html>
- [13] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. <https://openreview.net/forum?id=ryGsg6iA5Km>
- [14] Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. 2021. Graph Contrastive Learning Automated. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 12121–12132. <http://proceedings.mlr.press/v139/you21a.html>
- [15] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph Contrastive Learning with Augmentations. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). <https://proceedings.neurips.cc/paper/2020/hash/3fe230348e9a12c13120749e3f9f4cd-Abstract.html>