Graph Self Supervised Learning: the BT, the HSIC, and the VICReg.

Sayan Nag*
University of Toronto, Ontario, Canada
nagsayan112358@gmail.com

1. Introduction

Supervised Learning calls for a large number of labels thereby rendering it practically inapplicable in many applications because of the unavailability of such large annotated datasets. Dataset labeling comes with huge overhead costs and can also be biased. Therefore, self-supervised learning (SSL) strategy was adopted which enables the training of models on unlabeled data. In the presence of a handful of labeled data, using SSL strategy representations from the remaining unlabeled data can be learnt (pre-training) after which labeled data are used to fine-tune the pre-trained models. These fine-tuned models can either be used for downstream tasks, or as an auxiliary training task that enhances the performance of main tasks [1]. In our preliminary study, we have used four different loss functions namely Barlow Twins [1], HSIC [2], VICReg [3] and proposed hybrid VICRegHSIC in the Graph based SSL paradigm and have compared their perfomances on node classification tasks abreast performing ablation studies. Code is available at: (**https://github.com/sayannag/GraphSSL_HSICVICReg**)

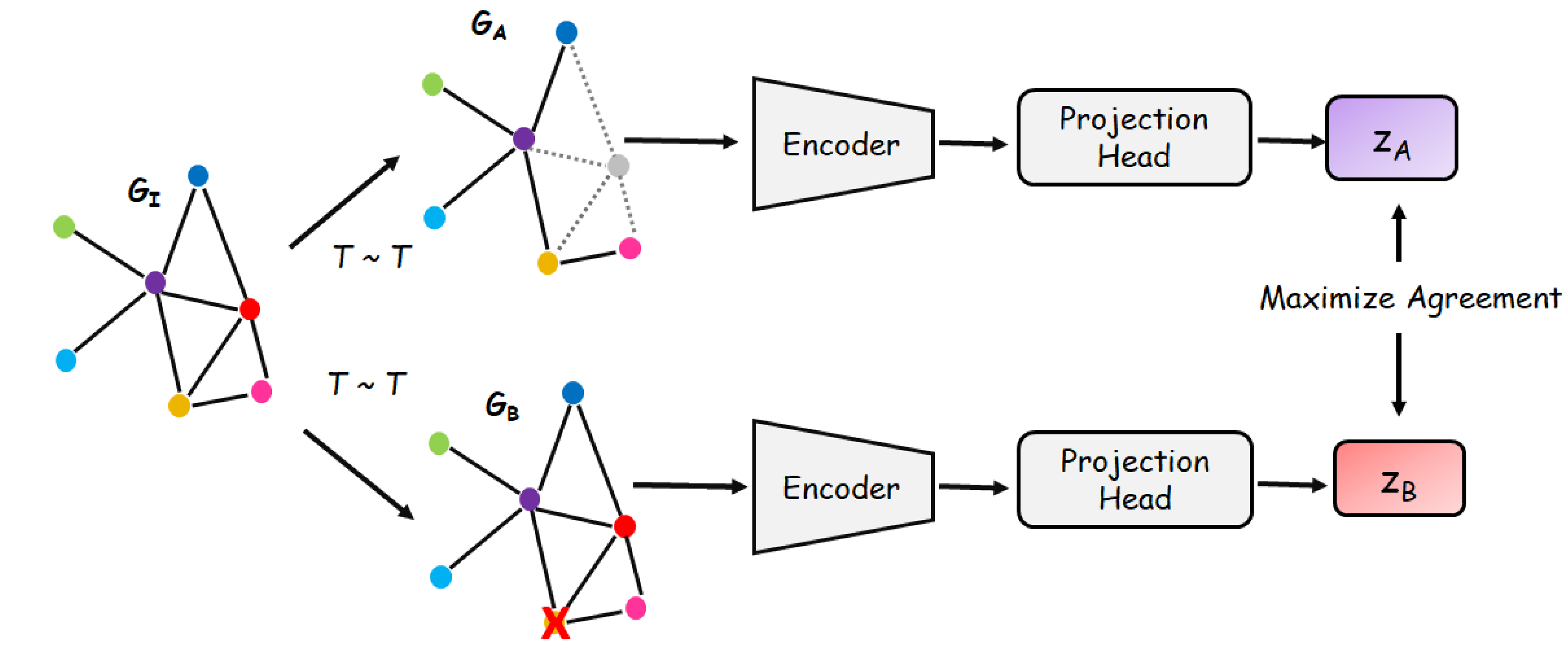
2. Data Augmentations

For fair comparisons, in our work we have followed the graph augmentations used in [4] which are as follows:

1. **Node Dropping (ND)** → Does not alter the semantics of the graph.
2. **Subgraph (SG)** → The local structure is able to hint the full semantics.
3. **Edge Perturbation (EP)** → Underlying prior is that the graph has a certain degree of semantic robustness to the connectivity variations.
4. **Attribute Masking (AM)** → There exists a certain extent of semantic robustness against losing partial attributes.

3. Framework

A standard paradigm of graph based self-supervised learning is shown below. There are 4 basic steps in order:
→ Graph Augmentations on input graph G_I resulting in two transformed graphs G_A (via ND augmentation) and G_B (via AM augmentation) following each transformation T.
→ Encode representations via Encoder.
→ Project the representations to a different dimension via Projection Head
→ Compute a self-supervised loss between the encoded representations z_A and z_B in order to maximize the agreement.



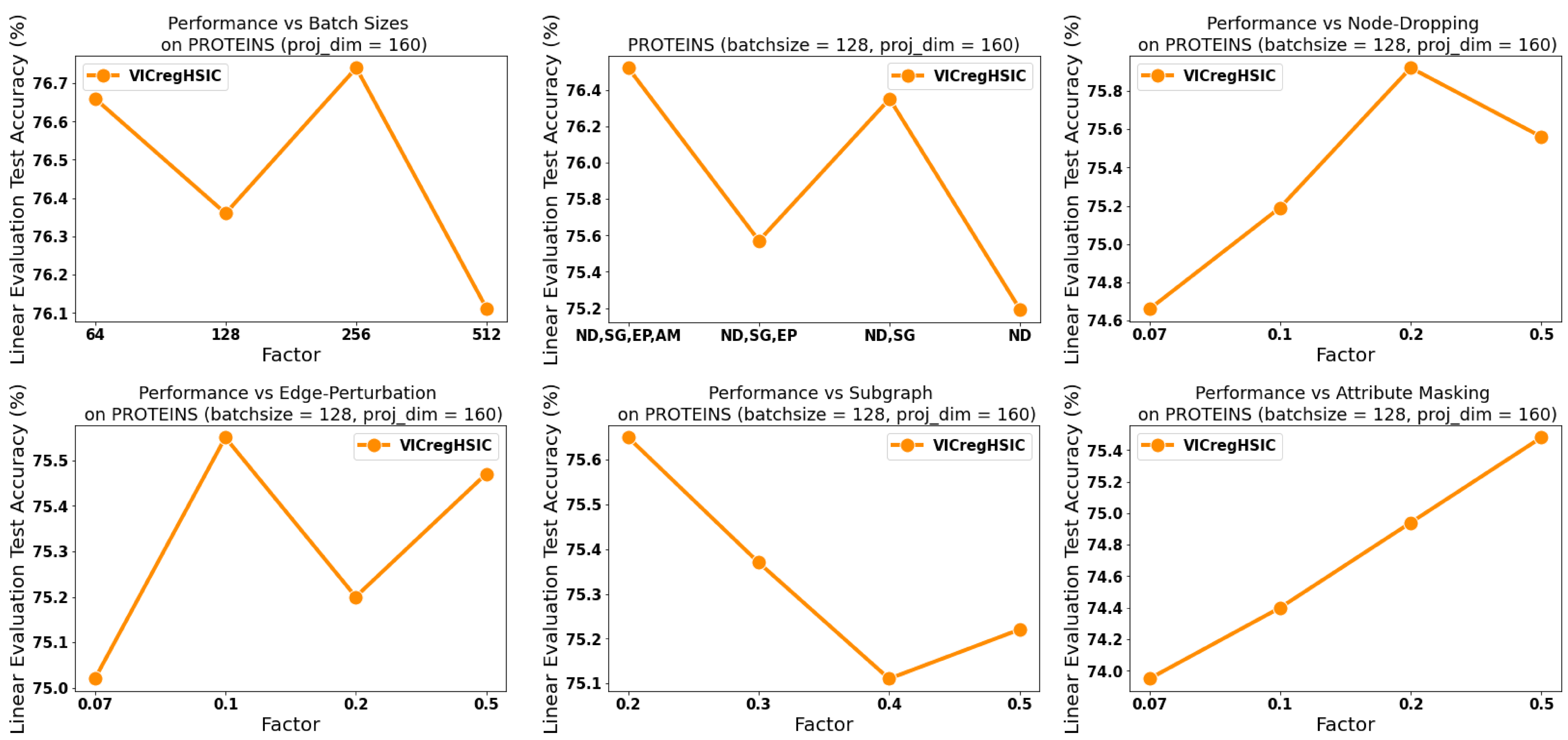
4. Benchmarks

Linear Evaluation Test Accuracy scores between different state-of-the-art methods show that the proposed VICRegHSIC loss function gives the best accuracy score in 2 out of 3 cases.

Methods	MUTAG	PROTEINS	IMDB-B
InfoGraph [Sun *et al.*, 2019]	89.13	74.48	73.05
GCL [You *et al.*, 2020]	86.98	74.43	72.21
GBT	89.41	76.12	72.50
GHSIC	88.68	76.08	71.93
GReg	88.83	**76.37**	72.9
GVICRegHSIC	**90.05**	76.35	**73.2**

5. Ablation Studies

- Removing the data augmentations reduces the performance.
- EP leads to a decrease in performance overall for both PROTEINS and MUTAG because for biomolecules datasets, the semantics are less robust against connectivity variations [4].
- ND augmentation have been seen to be useful.
- SG augmentation has reduced the performance of the algorithm drasctially.
- Increasing the masking factor for AM augmentation leads to an increase in performance



6. Conclusions

1. We have proposed a hybrid self-supervised loss function called VICRegHSIC.
2. We compared VICRegHSIC with BT, HSIC and VICReg on PROTEINS, MUTAG and IMDB.
3. We did ablation studies to show the impact of batch sizes, projector dimensions and different graph level augmentations on the performance of the algorithm.

7. References

[1] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. *arXiv preprint arXiv:2103.03230*, 2021.

[2] Yao-Hung Hubert Tsai, Shaojie Bai, Louis-Philippe Morency, and Ruslan Salakhutdinov. A note on connecting barlow twins with negative-sample-free contrastive learning. *arXiv preprint arXiv:2104.13712*, 2021.

[3] Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.

[4] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33, 2020.