Self-supervised Learning for Graph Classification

Selim Yagci [1]

Abstract

This study analyzes the performance and problems of self-supervised approach on the challenges of graph-based deep learning, focusing on graph classification task and generalization. Two proposed approaches, namely Graph Adversarial Pseudo Group Contrast (GAPGC) and a self-supervised domain adaptation method for size generalization, are briefly introduced. GAPGC enhances the relevance between pretext and downstream tasks, while the self-supervised domain adaptation approach exploits the d-pattern characterization of local structures. Both have proven superior performance over state-of-the-art methods. However, problems of limitations include assumptions on node and edge features, the need for domain knowledge, and access to the test distribution. Additionally, we present possible direction for exploring automated pretext task selection for different domains.

1. Introduction

Graph-structured data is widely used in variety of domains, including social networks (Fan et al., 2019), molecular sciences (Wang et al., 2023), transportation (Peregrino et al., 2021) and e-commerce (Liu et al., 2021). In recent years, deep learning models that operate over graphs have become increasingly popular. For example, in social networks, a graph-based approach can learn the interactions between users and contents to make highly accurate recommendations (Wu et al., 2019). The complexity of graph data has shown challenging problems on state-of-the-art machine learning methods that are highly successful on image and text data. Despite the challenges, supervised scenarios on specific downstream tasks such as graph classification with annotated labels has been performing well.

However, being heavily reliant on labeled data and poor generalization are two significant shortcomings of supervised learning for graph classification. To address this problem, self-supervised learning (SSL) frameworks have started gaining popularity also in graph-based deep learning research after their success in natural language processing and computer vision (Gui et al., 2023).

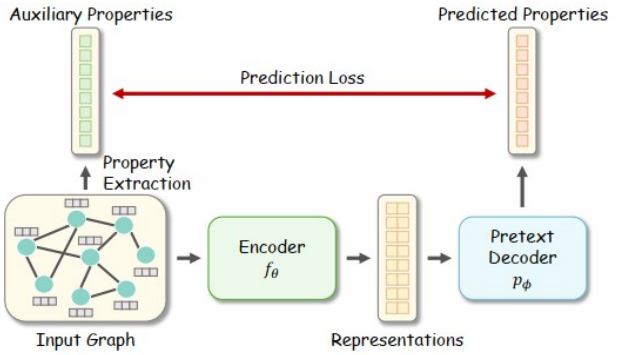

Figure 1. The figure shows Auxiliary property-based graph SSL method as an example for SSL on graph classification (Liu et al., 2022)

SSL enables model to learn by pretext tasks (Liu et al., 2022), acquiring its own supervision signals. In SSL, pretext task is defined as the tasks designed for acquiring a promising pretrained model. Rotation prediction and instance discrimination are some common pretext tasks (Gui et al., 2023). Depending on the design of pretext task and its objective function SSL in graphs can be divided into four concepts, including generation-based, auxiliary property-based, contrast-based and hybrid methods. All concepts have their disadvantages. Generation-based methods are prone to memory consumption on large-scale graph, auxiliary property-based methods needs domain knowledge to select helpful auxiliary properties and contrast-based methods may need longer time in empirical experiments (Liu et al., 2022).

Not only designing good pretext tasks, but the relationship between downstream task and pretext task is also crucial in SSL frameworks. Graph encoders can be pretrained on pretext task and finetuned on downstream task, jointly trained on both tasks in a multitask learning manner or can leverage

[1]Department of Computer Science, Technical University Munich, Munich, Germany. Correspondence to: Selim Yagci <selim.yagci@tum.de>.

fixed representations from unsupervised first step to learn downstream decoder in second step (Xie et al., 2022).

Choosing the well-designed pretext task and training strategies, SSL showed great results in image-based and text-based learning. For example SimCLR framework has performed well in augmenting image data by various means and training Convolutional Neural Networks to capture dependencies between different augmentations. And, BERT has achieved higher performance than conventional methods in several downstream tasks by pretraining its model with pretexts like masked language model and next sentence prediction (Wu et al., 2021).

Inspired by that success, similar approaches have been adapted to graph-data domain, especially from image-based methods. However the complex structure of the graph data constitutes a limit on these methods. For example, augmentation-based methods referred above for SimCLR cannot be extended trivially to graphs due to compound augmentations of graphs (Chen et al., 2022). Thus, recent advances in SSL on graphs seeks characteristic methods for graphs and the problems rises from its structure.

One major problem have been attempted to be addressed by graph specific SSL methods is out-of-distribution generalization. OOD generalization arises from shifts in test and train data distributions (Li et al., 2022). While most of the methods in literature assumes in-distribution test and training graph data, in real life this hypothesis cannot be satisfied. And since the performance decreases, generalization of the model becomes problematic. These distribution shifts can occur on node features, graph size or other structural properties. The graph size distribution shift is essential principally for graph distributions where local structure depends on graph size (Yehudai et al., 2021).

This paper aims at providing brief introductions of two different SSL approaches on this OOD generalization problem in learning with graph neural networks (GNN). More precisely, in section 2, we present a test time adaptation (TTA) method and how it is used for graph-base data briefly. In section 3, we provide background on a specific OOD generalization problem, namely size generalization of GNNs and the idea behind d-pattern characterization. The performance review and possible problems are then discussed in section 4. Finally, in section 5, we conclude the paper with presenting a possible future research direction.

2. Out-Of-Distribution Generalization

Recently TTA methods proposed to solve OOD generalization problem in visual domain (Wang et al., 2021). Main objective of TTA methods is to adapt pretrained model

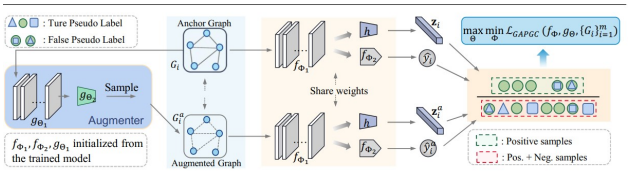

Figure 2. The figure shows overall framework of GAPGC for graph test time adaptation (Chen et al., 2022)

to unlabeled data during testing before inference. For instance, a test-time training (TTT) method has shown great performance with introducing a rotation task (as a self-supervised pretext task) to be jointly trained with the downstream task and fine tuning trained model during the test time (Sun et al., 2020). However image-based methods are limited in graph domain and there are several reasons. Graphs are structured irregularly, i.e. its non-Euclidian nature (Li et al., 2022). Thus, many image augmentation-based methods cannot be extended to graphs. Moreover, data-agnostic methods cause confidence bias and models may learn redundant information causing sub-optimal performance in downstream task (Chen et al., 2022).

Tackling these problems, a novel TTA method, Graph Adversarial Pseudo Group Contrast (GAPGC) has been proposed. GAPGC makes pseudo-positive group of graph augmentations, where the class information comes from the pseudo-labels output in test time. These labels are initialized with the offline trained model. The contrastive loss is enabled to exploit pseudo-labels in training, which gets the discriminative information from the pre-trained model. Therefore, the relevance between self-supervised pretext task and the downstream task gets improved. Moreover, GAPGC adopts adversarial learning. Optimizing the encoder to maximize correspondence between augmented disturbed graph pairs by minimizing the contrastive loss, the method reduces redundant information for downstream task. Results on molecular OOD datasets showed success in generalization of GAPGC (Chen et al., 2022).

3. Size Generalization

For graph classification in many domains, it is hard to collect labels for large graphs. Therefore, finding methods that can generalize from small graphs to larger ones is significant part of GNN research recently. Specifically, graphs with the distribution of its local structures depends on size is challenging in terms of OOD generalization problem. For example, in preferential attachment model in biological networks, maximal node degree grows with size of the graph. Moreover, size generalization is important because it is very common to have different size graphs in the same domain (Buffelli et al., 2022). For example graph sizes are

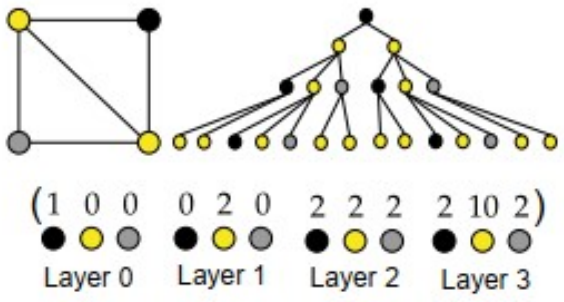

Figure 3. Top left: a graph with node features represented by colors. Top right: A tree that represents the d-patterns for the black node. Bottom: The tree descriptor is a vector with each coordinate containing the number of nodes from each class in each layer of the tree. (Yehudai et al., 2021).

ranging from dozens of nodes to billions of nodes in social networks.

First step of identifying this generalization problem is to get a representation of local structures as capturing the values of a node and its neighbors, which is also called d-patterns. With this characterization we can see that even a small discrepancy in distribution of d-patterns between test and train data may cause the global minima fail on generalizing for larger sizes because of bad weight assignments. And as d-pattern discrepancy grows, the generalization of GNNs degrades more.

Formulating learning problem as a domain adaptation problem where source domain have small graph size distribution and target consists of larger sized ones a solution is proposed. In the proposed method, the pretext task of learning useful d-pattern representation from both small and large graphs exploits the relation between d-pattern discrepancy and generalization. More precisely, model is pretrained for predicting the node labels holding important information about the node's d-pattern (see Figure 3). This pretext task helps aligning source and target domains representations. Results showed that, classification accuracy increases on large graphs by 0.04 on average, which is noticeable (Yehudai et al., 2021).

4. Performance and Problems of Proposed Approaches

Applying self-supervised approach to solve two significant issues related to generalization of graph-base deep learning, there are still limitation on proposed methods.

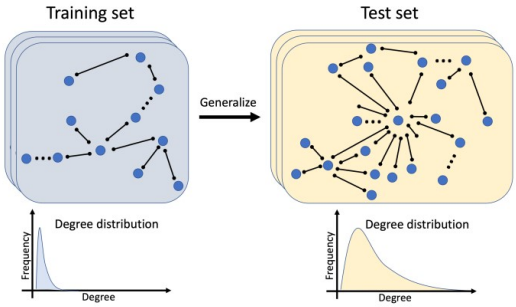

Figure 4. The figure shows two graph distributions that differ in size and degree distribution (Yehudai et al., 2021).

For the size generalization problem, Yehudai et al. (2021) approach assumed strict use cases like categorical node features and edges without features. The expansion of the approach does not seem to happen also due to the limit of d-pattern characterization over different domains. The approach also focuses on representing the local structure of each node. Thus, for other downstream tasks, requiring a representation that takes into account the whole graph, this approach does not really apply.

In test time adaptation, although Chen et al. (2022) approach do not require ad-hoc use cases or models, we may need to know about graphs' domain, which might be hard to obtain in advance. Moreover, we do not always have access to the test distribution in real datasets (Buffelli et al., 2022). And lastly the augmenter in the model parameterized only on specific graph generators.

Furthermore, a general problem of proposed methods is their explainablity. The methods are not interpretable with respect to what exactly they learn from self-supervised pretext tasks except the empirical correlations that are given. The question of which feature patterns or structures are learned by SSL is crucial especially in security or privacy related downstream tasks (Wu et al., 2021).

5. Conclusion and Future Work

Different from data instances in the image and text domains, graphs have more complex and irregular structure. Main challenges of graph classification are heavy label reliance, poor generalization, and consequently weak robustness. SSL approach may be leveraged to address these issues by extract knowledge through specific pretext tasks without having manual labels, resulting in better generalization. However two proposed approaches has shown either some level of reliance on data and model specific assumptions or requirement of accessing test distribution. For the future work, it would worth exploring strategies that

automatically finds best suitable pretext task and objective functions for different domains, while preserving the robustness to size shifts (Jin et al., 2022).

References

Buffelli, D., Lió, P., and Vandin, F. Sizeshiftreg: a regularization method for improving size-generalization in graph neural networks. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 31871–31885. Curran Associates, Inc., 2022.

Chen, G., Zhang, J., Xiao, X., and Li, Y. Graphtta: Test time adaptation on graph neural networks, 2022.

Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., and Yin, D. Graph neural networks for social recommendation, 2019.

Gui, J., Chen, T., Cao, Q., Sun, Z., Luo, H., and Tao, D. A survey of self-supervised learning from multiple perspectives: Algorithms, theory, applications and future trends, 2023.

Jin, W., Liu, X., Zhao, X., Ma, Y., Shah, N., and Tang, J. Automated self-supervised learning for graphs, 2022.

Li, H., Wang, X., Zhang, Z., and Zhu, W. Out-of-distribution generalization on graphs: A survey, 2022.

Liu, W., Zhang, Y., Wang, J., He, Y., Caverlee, J., Chan, P., Yeung, D., and Heng, P.-A. Item relationship graph neural networks for e-commerce. *IEEE Transactions on Neural Networks and Learning Systems*, PP:1–15, 03 2021. doi: 10.1109/TNNLS.2021.3060872.

Liu, Y., Jin, M., Pan, S., Zhou, C., Zheng, Y., Xia, F., and Yu, P. Graph self-supervised learning: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2022. doi: 10.1109/tkde.2022.3172903.

Peregrino, A. A., Pradhan, S., Liu, Z., Ferreira, N., and Miranda, F. Transportation scenario planning with graph neural networks, 2021.

Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A. A., and Hardt, M. Test-time training with self-supervision for generalization under distribution shifts, 2020.

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., and Darrell, T. Tent: Fully test-time adaptation by entropy minimization, 2021.

Wang, Y., Li, Z., and Farimani, A. B. Graph neural networks for molecules, 2023.

Wu, L., Lin, H., Gao, Z., Tan, C., and Li, S. Z. Self-supervised learning on graphs: Contrastive, generative,or predictive, 2021.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. A comprehensive survey on graph neural networks. *CoRR*, abs/1901.00596, 2019. URL http://arxiv.org/abs/1901.00596.

Xie, Y., Xu, Z., Zhang, J., Wang, Z., and Ji, S. Self-supervised learning of graph neural networks: A unified review, 2022.

Yehudai, G., Fetaya, E., Meirom, E., Chechik, G., and Maron, H. From local structures to size generalization in graph neural networks, 2021.