

# MOTIF-Driven Contrastive Learning of Graph Representations

Arjun Subramonian\*

University of California, Los Angeles  
405 Hilgard Avenue  
Los Angeles, California 90095  
arjunsub@ucla.edu

## Abstract

We propose a MOTIF-driven contrastive framework to pre-train a graph neural network in a self-supervised manner so that it can automatically mine motifs from large graph datasets. Our framework achieves state-of-the-art results on various graph-level downstream tasks with few labels, like molecular property prediction.

## Introduction

Graph representation learning aids numerous applications, like predicting properties of molecules (Hu et al. 2020b). These applications are often low-resource because labels are expensive to acquire, which makes them amenable to self-supervised learning. Self-supervised learning is a form of unsupervised learning in which pseudo-labels intrinsic to the data provide supervision. There exist self-supervised methods for learning graph-level representations, including Info-Graph (Sun et al. 2019), GPT-GNN (Hu et al. 2020c), and Context Prediction (Hu et al. 2020b).

Furthermore, we initially experimented with a contrastive self-supervised learning method for graph-level representations in which the representations of positive subgraph pairs (i.e. subgraphs from the same graph) are learned to be more similar while the representations of negative subgraph pairs (i.e. subgraphs from different graphs) are learned to be more dissimilar. However, this method was empirically ineffective because it inherently suffers from false negatives; that is, the representations of semantically-similar subgraphs from different graphs are erroneously learned to be more dissimilar.

## MOTIF Overview

The aforementioned methods fail to leverage the commonality of significant subgraphs across graphs in a dataset, known as motifs, which often indicate graph-level semantic properties (Milo et al. 2002). We propose a MOTIF-driven contrastive self-supervised framework (MOTIF) that simultaneously trains a graph neural network (GNN) to learn subgraph representations and clusters these representations based on

their semantics to mine motifs, thereby pretraining the GNN for various graph-level downstream tasks.

On each iteration of MOTIF, we obtain subgraph representations by sampling subgraphs from a batch of graphs, passing these subgraphs through the GNN, and pooling node representations. We subsequently compute a soft assignment of each subgraph representation to  $K$  clusters, each defined by a center vector. The  $K$  center vectors represent motifs and are interpreted as the semantic meaning shared by similar subgraphs. We represent the soft cluster assignments as a matrix  $Q$ , where the  $i$ th column is a  $K$ -dimensional probability vector representing how likely the  $i$ th subgraph belongs to each of the  $K$  clusters. Our goal is to optimize  $Q$  by adjusting the cluster centers to maximize the similarity between subgraph representations and the centers. To prevent all the subgraphs from being assigned to a single cluster, we introduce an entropy term and a constraint on  $Q$  to force all the clusters to have similar size. Thereby, computing  $Q^*$  reduces to an optimal transportation problem with a closed-form solution (Asano, Rupprecht, and Vedaldi 2020).

The elegance of our solution is in that  $Q^*$  provides us with pseudo-labels with which we can contrastively train the GNN in a self-supervised manner. We predict the probabilities of each subgraph belonging to the optimal clusters via representation-center similarity and learn subgraph representations to be more similar or dissimilar to the appropriate cluster centers by minimizing the cross-entropy loss.

In our experiments, MOTIF was ineffective when employing random walk or k-hop neighbor subgraph sampling, as these methods solely rely on topology without considering node representations learned by the GNN. To solve this issue, I experimented with many dynamic subgraph samplers that exploit learned node representations. I implemented the samplers with PyTorch Geometric (Fey and Lenssen 2019), writing subgraph visualization notebooks and utilizing homogeneity, completeness, and intersection over union to evaluate the quality of subgraphs. Ultimately, to sample subgraphs from a graph, I computed the similarities between the representations of adjacent nodes and normalized them to produce an affinity matrix  $A$ , where  $A_{i,j}$  is the affinity score between adjacent nodes  $i$  and  $j$ . I then applied spectral clustering with  $A$  to recursively bisect the graph into subgraphs of multiple sizes (1). Intuitively, as we train the GNN using the clustering scheme, the representations of nodes belong-

\*In collaboration with Shichang Zhang, Ziniu Hu, and Yizhou Sun, all from the same institution.  
Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

	bace	bbbp	clintox	hiv	sider	tox21	toxcast	Average
Non-Pretrain	72.80 $\pm$ 2.12	82.13 $\pm$ 1.69	74.98 $\pm$ 3.59	73.38 $\pm$ 0.92	55.65 $\pm$ 1.35	76.10 $\pm$ 0.58	63.34 $\pm$ 0.75	71.19
ContextPred	73.02 $\pm$ 2.59	80.94 $\pm$ 2.55	74.57 $\pm$ 3.05	73.85 $\pm$ 1.38	54.15 $\pm$ 1.54	74.85 $\pm$ 1.28	63.19 $\pm$ 0.94	70.65 (-0.7%)
InfoGraph	76.09 $\pm$ 1.63	80.38 $\pm$ 1.19	<b>78.36 <math>\pm</math> 4.04</b>	72.59 $\pm$ 0.97	56.88 $\pm$ 1.80	76.12 $\pm$ 1.11	64.40 $\pm$ 0.84	72.11 (+1.3%)
GPT-GNN	75.56 $\pm$ 2.49	83.35 $\pm$ 1.70	74.84 $\pm$ 3.45	74.82 $\pm$ 0.99	55.59 $\pm$ 1.58	76.34 $\pm$ 0.68	64.76 $\pm$ 0.62	72.18 (+1.4%)
MOTIF	<b>76.16 <math>\pm</math> 2.51</b>	<b>83.78 <math>\pm</math> 1.77</b>	77.50 $\pm$ 3.35	<b>75.51 <math>\pm</math> 0.67</b>	<b>57.28 <math>\pm</math> 1.09</b>	<b>76.68 <math>\pm</math> 0.36</b>	<b>65.42 <math>\pm</math> 0.62</b>	<b>73.19 (+2.8%)</b>

Table 1: We evaluated transfer learning performance using ROC-AUC. For multi-task prediction, we averaged the ROC-AUC across all tasks. We report the test result of the best model on the validation set across 10 runs.

ing to the same motif will become more similar and cause motif-like subgraphs to be sampled. Hence, the sampler will reinforce clustering and vice versa, thereby boosting MOTIF’s performance. To reduce the runtime of my sampling implementation, I introduced multiprocessing and increased batch, vectorized, and scatter operations.

## Results and Discussion

We evaluated MOTIF on molecular property prediction datasets from the Open Graph Benchmark (OGB) (Hu et al. 2020a). We pretrained GNNs using MOTIF and state-of-the-art baselines on the large ogbg-molhiv dataset. Then, we emulated the scarcity of labels in the real world by finetuning the pretrained model on all the ogbg molecule datasets with only 10% of labels (Hu et al. 2020a). We used the Deeper Graph Convolutional Network (DeeperGCN) as our GNN with the recommended model hyperparameters (Li et al. 2020). We used PyTorch and PyTorch Geometric for all implementations (Paszke et al. 2017; Fey and Lenssen 2019).

I implemented and ran pretraining and finetuning for all baselines. The transfer learning results of the non-pretrained model, MOTIF-pretrained model, and models pretrained using the baselines on the ogbg molecule datasets are reported in Table 1. The MOTIF-pretrained model consistently performed better than or on par with models pretrained with the baselines when transferring to a supervised finetune task with few labels. Furthermore, the MOTIF-pretrained model outperformed the non-pretrained model by 2.8% on average.

These results suggest that valuable graph-level information can be extracted by automatically mining motifs from a

large graph dataset and leveraging the motifs to contrastively learn graph representations. Compared to the baselines, MOTIF further enhances molecular property prediction with few labels, which can advance drug discovery and quantum chemistry. Furthermore, MOTIF is more interpretable, as one can inspect the motifs mined to intuit the graph representations learned. In the future, we will explore introducing a sampling loss term and end-to-end pooling for subgraph representations.

## References

- Asano, Y. M.; Rupprecht, C.; and Vedaldi, A. 2020. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations (ICLR)*.
- Fey, M.; and Lenssen, J. E. 2019. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428*.
- Hu, W.; Fey, M.; Zitnik, M.; Dong, Y.; Ren, H.; Liu, B.; Catasta, M.; and Leskovec, J. 2020a. Open Graph Benchmark: Datasets for Machine Learning on Graphs. *arXiv preprint arXiv:2005.00687*.
- Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; and Leskovec, J. 2020b. Strategies for Pre-training Graph Neural Networks. In *International Conference on Learning Representations*.
- Hu, Z.; Dong, Y.; Wang, K.; Chang, K.-W.; and Sun, Y. 2020c. GPT-GNN: Generative Pre-Training of Graph Neural Networks. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Li, G.; Xiong, C.; Thabet, A.; and Ghanem, B. 2020. DeeperGCN: All You Need to Train Deeper GCNs. *arXiv preprint arXiv:2006.07739*.
- Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; and Alon, U. 2002. Network motifs: simple building blocks of complex networks. *Science* 298(5594): 824–827.
- Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch. *NIPS 2017 Autodiff Workshop*.
- Sun, F.-Y.; Hoffmann, J.; Verma, V.; and Tang, J. 2019. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. *arXiv preprint arXiv:1908.01000*.

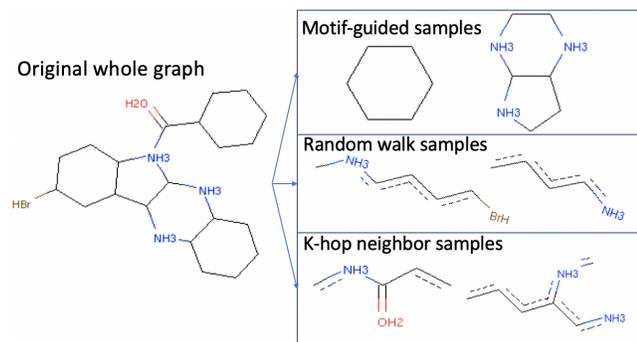


Figure 1: Motif-like subgraphs sampled by the dynamic sampler vs. subgraphs sampled by random walk and k-hop neighbor samplers. Figure created in collaboration with Shichang Zhang.