On Graph Neural Networks versus Graph-Augmented MLPs

Lei Chen*, Zhengdao Chen*, Joan Bruna

New York University

ICLR 2021

Summary: We show advantages of GNNs in expressive power and learning compared to Graph-Augmented MLPs

- GNNs based on message passing:

$$M_i^{(k)} = \text{AGGREGATE}^{(k)}(\{H_j^{(k-1)} : j \in \mathcal{N}(i)\})$$
$$H_i^{(k)} = \text{COMBINE}^{(k)}(H_i^{(k-1)}, M_i^{(k)})$$

- **Graph-Augmented MLPs**: Simplifying GNNs by replacing depth with multi-hop graph operators

Summary: We show advantages of GNNs in expressive power and learning compared to Graph-Augmented MLPs

- GNNs based on message passing:

$$M_i^{(k)} = \text{AGGREGATE}^{(k)}(\{H_j^{(k-1)} : j \in \mathcal{N}(i)\})$$
$$H_i^{(k)} = \text{COMBINE}^{(k)}(H_i^{(k-1)}, M_i^{(k)})$$

- **Graph-Augmented MLPs**: Simplifying GNNs by replacing depth with multi-hop graph operators

Summary: We show advantages of GNNs in expressive power and learning compared to Graph-Augmented MLPs

- GNNs based on message passing:

$$M_i^{(k)} = \text{AGGREGATE}^{(k)}(\{H_j^{(k-1)} : j \in \mathcal{N}(i)\})$$
$$H_i^{(k)} = \text{COMBINE}^{(k)}(H_i^{(k-1)}, M_i^{(k)})$$

- **Graph-Augmented MLPs**: Simplifying GNNs by replacing depth with multi-hop graph operators

Graph-Augmented Multi-layer Perceptrons (GA-MLPs)

- Node features as input: $X \in \mathbb{R}^{n \times d}$
  Node embeddings as output: $Z \in \mathbb{R}^{n \times d'}$
- Ingredients:
  1. A family of **graph operators**: $\Omega = \{\omega_1(A), ..., \omega_K(A)\}$
  2. $\varphi : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$ and $\rho : \mathbb{R}^{K\tilde{d}} \to \mathbb{R}^{d'}$ being learnable **node-wise** functions, e.g. MLPs
- Steps:
  1. For each k, compute $\tilde{X}_k = \omega_k(A) \cdot \varphi(X) \in \mathbb{R}^{n \times \tilde{d}}$
  2. Concatenate $\tilde{X} = [\tilde{X}_1, ..., \tilde{X}_K] \in \mathbb{R}^{n \times (K\tilde{d})}$
  3. Compute $Z = \rho(\tilde{X}) \in \mathbb{R}^{n \times d'}$
- Examples: SGC Wu et al. (2019), GFN Chen et al. (2019), SIGN Rossi et al. (2020)

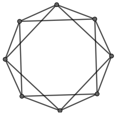
Expressive power of GNNs

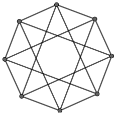
From the viewpoint of **graph isomorphism tests** -

Question: Can GNNs distinguish all pairs of non-isomorphic graphs?

Answer Xu et al. (2019); Morris et al. (2019)

No, for GNNs based on message passing. For example,





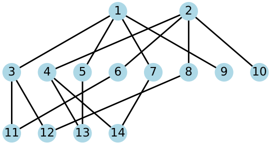
Expressive power of GA-MLPs

Question: Are GNNs more powerful than GA-MLPs in distinguishing non-isomorphic graphs?

Answer:
For certain choices of the operator family, yes.

Proposition (1)

If $\Omega \subseteq \{\tilde{A}^k : k \in \mathbb{N}\}$, with either $\tilde{A} = A$ or $\tilde{A} = D^{-\alpha} A D^{-(1-\alpha)}$ for some $\alpha \in [0, 1]$, there exists a pair of graphs which can be distinguished by GNNs but not this GA-MLP.



Expressive power of GA-MLPs

Question: Are GNNs more powerful than GA-MLPs in distinguishing non-isomorphic graphs?

Answer:
For certain popular choices of the operator family, yes.
However, the fraction of such examples is small.

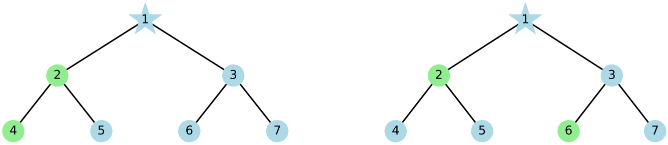
Proposition (2)

For all $n \in \mathbb{N}_+$, $\exists \alpha_n > 0$ such that any GA-MLP that has $\{D, AD^{-\alpha_n}\} \subseteq \Omega$ can distinguish almost all pairs of non-isomorphic graphs of at most n nodes, in the sense that the fraction of graphs on which such a GA-MLP fails to test isomorphism is $1 - o(1)$ as $n \to \infty$.

Towards a node-wise perspective on expressive power

Both GNNs and GA-MLPs can be viewed as functions on rooted graphs (i.e., egonets or neighborhooods).

They partition rooted graphs into equivalence classes.



More powerful models induce finer equivalence classes.

GNNs induce finer equivalence classes of rooted graphs than GA-MLPs

Proposition (3)

Assume that $|\mathcal{X}| \geq 2$ and $m \geq 3$. The total number of equivalence classes of rooted graphs induced by GNNs of depth K grows at least doubly-exponentially in K.

Proposition (4)

Fix $\Omega = \{I, \tilde{A}, \tilde{A}^2, ..., \tilde{A}^K\}$, where $\tilde{A} = D^{-\alpha} A D^{-\beta}$ for some $\alpha, \beta \in \mathbb{R}$. Then the total number of equivalence classes in $\mathcal{E}$ induced by such GA-MLPs is poly-exponential in K.

Corollary

The VC dimension of all GNNs of K layers as functions on rooted graphs grows at least doubly-exponentially in K; Fixing $\alpha, \beta \in \mathbb{R}$, the VC dimension of all GA-MLPs with $\Omega = \{I, \tilde{A}, \tilde{A}^2, ..., \tilde{A}^K\}$ as functions on rooted graphs is at most poly-exponential in K.

GNNs induce finer equivalence classes of rooted graphs than GA-MLPs

Proposition (5)

If Ω is any family of equivariant linear operators on the graph that only depend on the graph topology of at most K hops, then there exist exponentially-in-K many equivalence classes in $\mathcal{E}$ induced by the GA-MLPs with Ω, each of which intersects with doubly-exponentially-in-K many equivalence classes in $\mathcal{E}$ induced by depth-K GNNs, assuming that $|\mathcal{X}| \geq 2$ and $m \geq 3$. Conversely, in constrast, if $\Omega = \{I, \tilde{A}, \tilde{A}^2, ..., \tilde{A}^K\}$, in which $\tilde{A} = D^{-\alpha} A D^{-\beta}$ with any $\alpha, \beta \in \mathbb{R}$, then each equivalence class in $\mathcal{E}$ induced by depth-$(K+1)$ GNNs is contained in one equivalence class induced by the GA-MLPs with Ω.

GA-MLPs cannot count attributed walks

Proposition (6)

For any sequence of node features $\{x_k\}_{k \in \mathbb{N}_+} \subseteq \mathcal{X}$, consider the sequence of functions $f_k(G^{[i]}) := |\mathcal{W}_k(G^{[i]}; (x_1, ..., x_k))|$ on $\mathcal{E}$. For all $k \in \mathbb{N}_+$, the image under f_k of every equivalence class in $\mathcal{E}$ induced by depth-k GNNs contains a single value, while for any GA-MLP using equivariant linear operators that only depend on the graph topology, there exist exponentially-in-k many equivalence classes in $\mathcal{E}$ induced by this GA-MLP whose image under f_k contains exponentially-in-k many values.

GA-MLPs cannot count attributed walks

Model	Cora Train	Cora Test	RRG Train	RRG Test
GIN	**3.98E-6**	**9.72E-7**	**3.39E-5**	**2.61E-4**
GA-MLP-A	1.23E-1	1.56E-1	1.75E-2	2.13E-2
GA-MLP-$A+$	1.87E-2	6.44E-2	1.69E-2	2.13E-2
GA-MLP-$\tilde{A}_{(1)}$	4.22E-1	5.79E-1	1.02E-1	1.58E-1
GA-MLP-$\tilde{A}_{(1)}+$	4.00E-1	5.79E-1	1.12E-1	1.52E-1

Table 1: MSE loss divided by label variance for counting attributed walks on the Cora graph and RRG. The models denoted as "+" contain twice as many powers of the operator.

Another advantage of GNNs over GA-MLPs: learning the operators

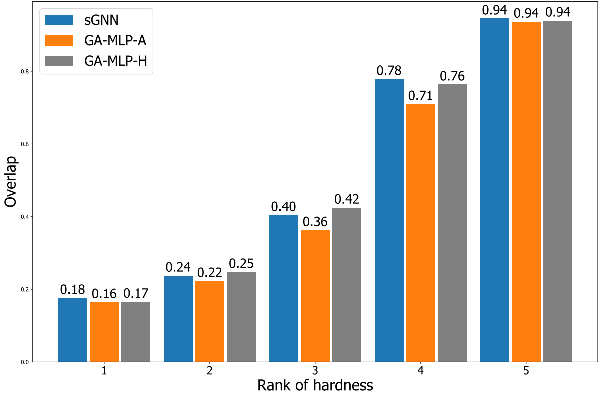


Figure 1: Community detection on binary SBM with 5 choices of in- and out-group connectivities, each yielding to a different SNR. Higher overlap means better performance.

Thanks!

References

Chen, T., Bian, S., and Sun, Y. (2019). Are powerful graph neural nets necessary? a dissection on graph classification. arXiv preprint arXiv:1905.04579.

Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., and Grohe, M. (2019). Weisfeiler and leman go neural: Higher-order graph neural networks. Association for the Advancement of Artificial Intelligence.

Rossi, E., Frasca, F., Chamberlain, B., Eynard, D., Bronstein, M., and Monti, F. (2020). Sign: Scalable inception graph neural networks. arXiv preprint arXiv:2004.11198.

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., and Weinberger, K. (2019). Simplifying graph convolutional networks. volume 97 of Proceedings of Machine Learning Research, pages 6861–6871, Long Beach, California, USA. PMLR.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019). How powerful are graph neural networks? In International Conference on Learning Representations.