Towards Bridging Generalization and Expressivity of Graph Neural Networks

Shouheng Li[1,2], Floris Geerts[3], Dongwoo Kim[4], Qing Wang[2]

[1]Data61 CSIRO [2]Australian National University
[3]University of Antwerp [4]POSTECH

This research was supported partially by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP210102273) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(RS-2024-00337955 and RS-202300217286).

- There is a lot of recent research on GNN expressivity, but much less is known about GNN generalization.
- Common belief: More expressivity $\rightarrow$ risk of **overfitting**.
- **Contradiction**: Empirically some **powerful GNNs generalize well**.
- **Challenge**: Understanding how generalization is influenced by graph structure.

Goal: Explain **how** and **when** added *expressivity* can lead to **improved generalization**.

Key Contributions

- Propose a **novel generalization bound** for GNNs that incorporates graph structure variance.
- Identify conditions for GNNs to **generalize well** from their **bounded expressivity**:
  - Embeddings within a class are **well-clustered**
  - Classes are **separable** in the embedding space
- Show how expressivity influences generalization by case studies
- Validate the theoretical insights by showing the alignment with empirical findings

A graph encoder maps **a graph** $G \in \mathcal{G}$ to **a vector** in some space $\mathcal{Z}$.

- GNNs are graph encoders: MPNN/$\mathcal{F}$-MPNN/k-GNN/...
- Weisfeiler-Lehman are also graph encoders: 1-WL/$\mathcal{F}$-WL/k-WL/...
- Random walk encoder, subtree encoder, ...

For a graph encoder ϕ, bound its **generalization error**

$$R(\phi) - \hat{R}(\phi)$$

- $R(\phi)$: true/population error
- $\hat{R}(\phi)$: empirical error

A Tale of Two Graph Encoders

- Weaker encoder: $\phi : \mathcal{G} \to \mathcal{Z}_\phi$
- Stronger encoder: $\lambda : \mathcal{G} \to \mathcal{Z}_\lambda$
- λ bounds ϕ in expressivity
  - if λ can distinguish all graphs that ϕ can distinguish
  - there is $f : \mathcal{Z}_\lambda \to \mathcal{Z}_\phi$ such that $\phi = f \circ \lambda$

A Tale of Two Graph Encoders

- Weaker encoder: $\phi : \mathcal{G} \rightarrow \mathcal{Z}_\phi$
- Stronger encoder: $\lambda : \mathcal{G} \rightarrow \mathcal{Z}_\lambda$
- λ bounds ϕ in expressivity
  - if λ can distinguish all graphs that ϕ can distinguish
  - there is $f : \mathcal{Z}_\lambda \rightarrow \mathcal{Z}_\phi$ such that $\phi = f \circ \lambda$

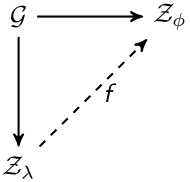
ϕ: MPNN/$\mathcal{F}$-MPNN/k-GNN/...

is **bounded by**

λ: 1-WL/$\mathcal{F}$-WL/k-WL/...

Margin-Based Generalization Bound (Informal)

Let encoder λ bounds encoder ϕ:



Lipschitz constant of f

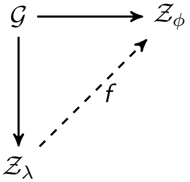
$$R(\phi) - \hat{R}(\phi) \leq \mathcal{O}\left(\text{Lip}(f) \cdot C(\lambda) \right)$$

Intra-class embedding concentration[1] of λ

Main Theoretical Result (2/3)

Margin-Based Generalization Bound (Informal)

Let encoder λ bounds encoder ϕ:



Lipschitz constant of f

$$R(\phi) - \hat{R}(\phi) \leq \mathcal{O}\left(\mathrm{Lip}(f) \cdot C(\lambda) \right)$$

Intra-class embedding concentration[1] of λ

Takeaway:

The *stronger encoder* λ (e.g. 1-WL/$\mathcal{F}$-WL/k-WL etc.)
hints at the generalization of
the *weaker encoder* ϕ (e.g. MPNN/$\mathcal{F}$-MPNN/k-GNN)

[1] Measured by Wasserstein distance

The above generalization bound is lower bounded by (Informal)

Intra-class embedding concentration[1] of λ

$$\mathcal{O}\left(\frac{\mathrm{Lip}(f)\cdot C(\lambda)}{B(\lambda)}\right).$$

Inter-class embedding separation[1] of λ

Main Theoretical Result (3/3)

The above generalization bound is lower bounded by (Informal)

Intra-class embedding concentration[1] of λ

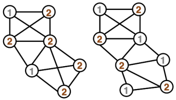
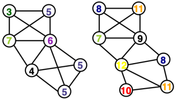
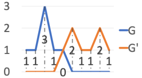
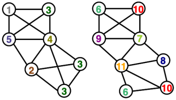
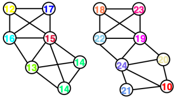
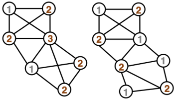
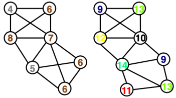
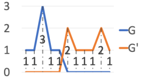
$$\mathcal{O}\left(\frac{\mathrm{Lip}(f)\cdot C(\lambda)}{B(\lambda)}\right).$$

Inter-class embedding separation[1] of λ

Key Insight: *Low intra-class variance* + *high inter-class separation* $\rightarrow$ better generalization.

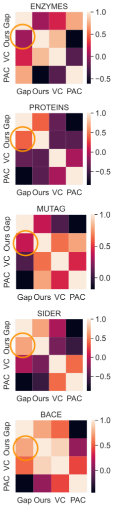
Also, we can predict GNNs' generalization **before training them!**

[1] Measured by Wasserstein distance

	Initial Vertex Colors G G'	After One Iteration G G'	Graph Embeddings (Difference)	Wasserstein Distance
(a) 1-WL				4.796
(b) C_4-WL				4.123
(c) K_4-WL				5.000

Takeaway: The choice of patterns leads to different generalization.

# Layers		Dataset ENZYMES	PROTEINS	MUTAG	SIDER	BACE
1	Loss gap	0.248±0.040	0.029±0.015	-0.070±0.017	0.037±0.003	0.018±0.017
	Our Bound	7.926±1.279	2.193±0.702	1.216±0.169	0.511±0.286	1.479±0.301
	VC dimension	586	929	51	960	621
	VC bound	1.302±0.000	1.292±0.001	1.100±0.004	1.302±0.000	1.301±0.000
	PAC bound	3.48	5.04	3.06	52.39	21.525
1	Loss gap	0.242±0.026	0.032±0.010	-0.074±0.007	0.038±0.003	0.037±0.019
	Our bound	7.425±0.982	1.404±0.144	1.247±0.155	0.620±0.463	1.729±0.251
	VC dimension	595	996	121	1300	1060
	VC bound	1.302±0.000	1.292±0.000	1.281±0.003	1.302±0.000	1.302±0.000
	PAC bound	12.75	31.94	8.17	132.79±8.12	51.573
2	Loss gap	0.237±0.035	0.025±0.009	-0.058±0.012	0.038±0.002	0.032±0.011
	Our Bound	6.513±0.951	1.421±0.220	1.649±0.158	0.409±0.253	1.789±0.226
	VC dimension	595	996	135	1309	1089
	VC bound	1.302±0.000	1.293±0.000	1.286±0.002	1.302±0.000	1.302±0.000
	PAC bound	56.98	276.78	21.96±0.00	341.04	124.605
4	Loss gap	0.235±0.038	0.027±0.005	-0.073±0.009	0.036±0.001	0.022±0.030
	Our Bound	6.825±0.796	1.434±0.297	1.535±0.115	0.298±0.080	1.686±0.377
	VC dimension	595	996	139	1309	1093
	VC bound	1.302±0.000	1.292±0.001	1.291±0.002	1.302±0.000	1.302±0.000
	PAC bound	308.43	2331.63	57.69	845.62	310.732
5	Loss gap	0.256±0.037	0.020±0.007	-0.071±0.021	0.035±0.001	0.020±0.020
	Our Bound	6.384±0.813	1.308±0.165	1.773±0.194	0.369±0.172	1.662±0.120
	VC dimension	595	996	139	1309	1093
	VC bound	1.302±0.000	1.292±0.001	1.292±0.002	1.302±0.000	1.302±0.000
	PAC bound	1615.10	17992.81	155.74	2179.21	744.08
6	Loss gap	0.264±0.025	0.030±0.008	-0.078±0.019	0.034±0.002	0.022±0.016
	Our Bound	6.151±0.798	1.340±0.316	1.627±0.038	0.353±0.156	1.785±0.237
	VC dimension	595	996	139	1309	1093
	VC bound	1.302±0.000	1.292±0.001	1.291±0.002	1.302±0.000	1.302±5.870
	PAC bound	8931.00	135762.52	410.31	5254.88	1860.94



Takeaway: The proposed bound reflects the empirical gaps well.

Conclusion

- Develop a **theoretical framework** linking GNN expressivity and generalization.
- Demonstrate how **well-clustered embeddings** and **separable classes** lead to improved generalization.
- Provide guidance for designing **robust, expressive, and generalizable GNNs**.

Future Work: Extend to understand how the training dynamics changes f and further influences generalisation.