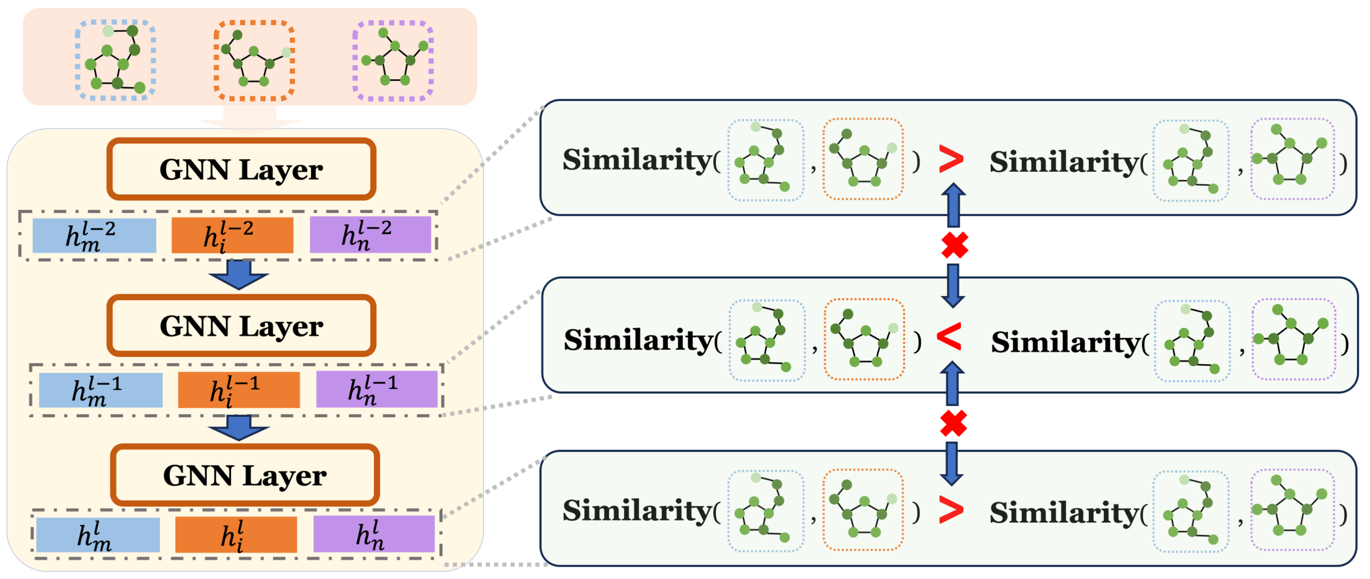
Exploring Consistency in Graph Representations: from Graph Kernels to Graph Neural Networks

Xuyuan Liu, Yinghao Cai, Qihui Yang, Yujun Yan
Dartmouth College

Motivation

GNNs fail to capture the similarity structure!



Observation: Inconsistency in similarity order across layers.

Limitations

Graph Kernels:

+ Effective at capturing relative graph similarities
- Depends on predefined kernels and lacks adequate non-linearities

GNNs:

+ Good at capturing non-linearities
- Ineffective at capturing relative graph similarity

Analogy between GNNs and IGS

Iterative Graph Kernels (IGK): Graph kernels obtained from an iterative coloring process

IGKs

Given: A graph G with a set of nodes V.

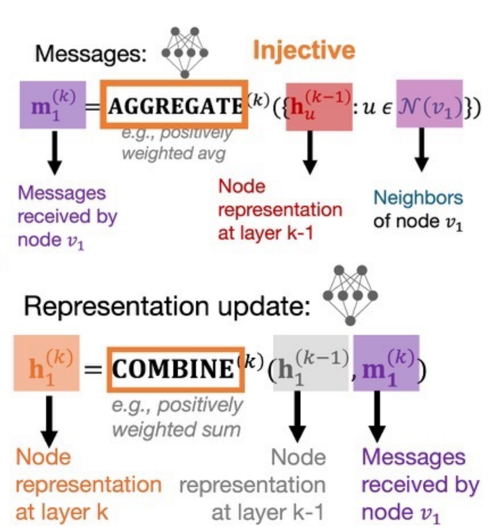
- Assign an initial color $c^{(0)}(v)$ to each node v.
- Iteratively refine node colors by

$$c^{(k+1)}(v) = \mathrm{HASH}\left(\left\{c^{(k)}(v), \{c^{(k)}(u)\}_{u \in N(v)}\right\}\right)$$

maps different inputs to different colors

- After k steps of color refinement, $c^{(k)}(v)$ summarizes the structure of k-hop neighborhood

GNNs



Given the analogy between GNNs and IGKs, can we bridge the two worlds?

Principles

[Monotonically Decreasing] The normalized iterative graph kernels $\mathbb{K}_{\mathcal{F}_c,\phi}(x, y, i)$ are said to be monotonically decreasing if and only if:

$$\tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, y, i) \geq \tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, y, i+1) \quad \forall x, y \in \chi$$

[Order Consistency] The normalized iterative graph kernels $\tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, y, i)$ are said to preserve order consistency if the similarity ranking remains consistent across different iterations for any pair of graphs:

$$\tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, y, i) > \tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, z, i) \Rightarrow \tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, y, i+1) \geq \tilde{\mathbb{K}}_{\mathcal{F}_c,\phi}(x, z, i+1) \quad \forall x, y, z \in \chi$$

Theorem: Two principles guarantee provable generalizability in graph classification tasks.

Verification:

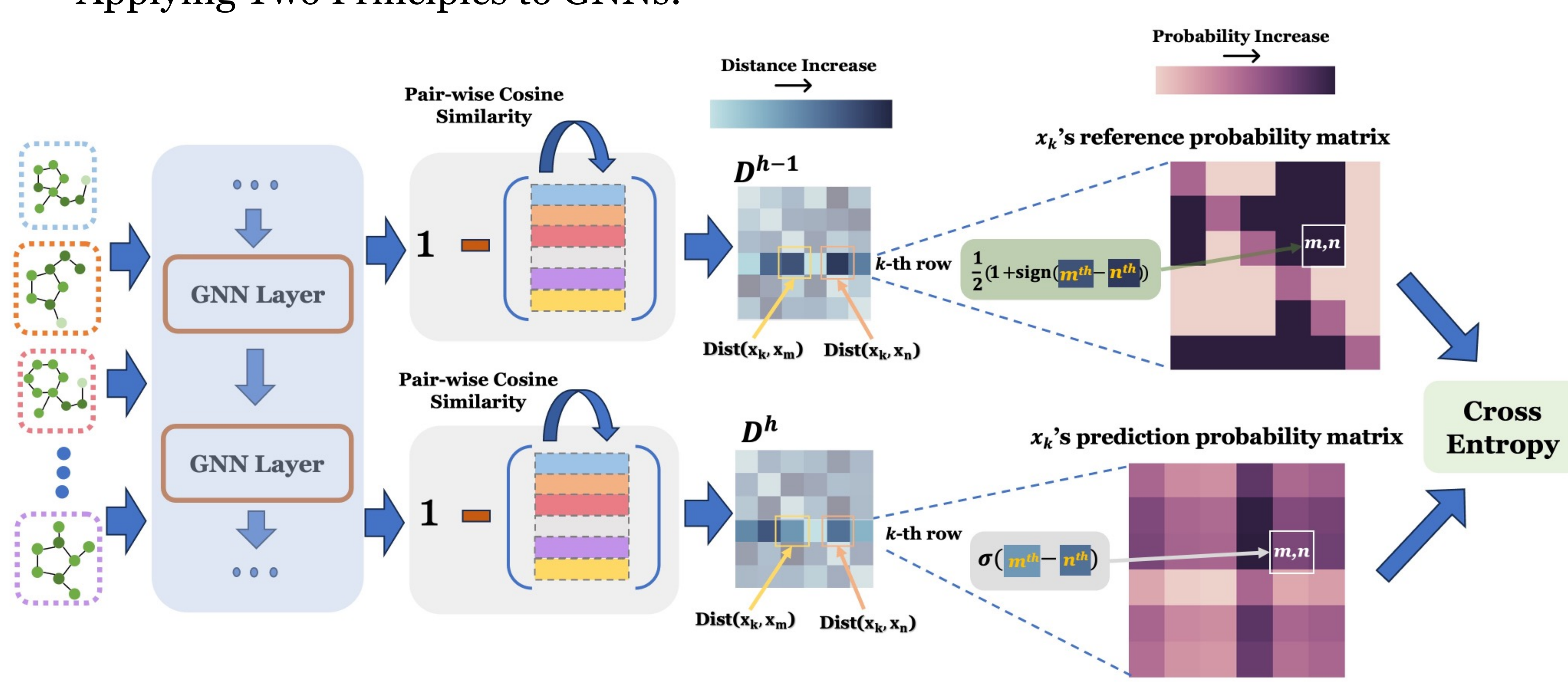
- WL-subtree Kernel[1]:
  - Does not follow either principle.
- WLOA Kernel[2]:
  - Monotonically decreases.
  - Preserves the order consistency asymptotically.

WLOA outperform WL-subtree in various benchmark.

Question: Can we apply the two principles to enhance GNN performance?

Consistency Loss

Applying Two Principles to GNNs:



Learn similarity order using signals from the previous layer.

Experiments

Performance

		NCI1	NCI109	PROTEINS	D&D	IMDB-B	IMDB-M	COLLAB	COIL-RAG	OGB-HIV	REDDIT-T
	#Graphs	4110	4127	1113	1178	1000	1500	5000	3900	41127	203088
	Avg. #nodes	29.87	29.68	39.06	284.32	19.77	13.00	74.49	3.01	25.50	23.93
Kipf et al. ICLR' 17	GCN	73.96±2.37	74.04±3.09	73.24±6.93	74.92±2.66	75.40±2.97	55.07±1.24	81.72±0.84	91.72±1.65	72.86±1.90	76.00±0.44
	$+\mathcal{L}_{\text{consistency}}$	75.12±1.19	73.25±1.25	75.07±5.05	78.56±3.32	75.85±1.82	56.27±1.00	83.44±0.45	93.38±1.64	73.75±0.89	77.12±0.12
Xu et al. ICLR' 19	GIN	78.13±2.11	76.75±2.91	72.97±4.59	71.10±4.63	70.80±4.07	52.13±1.42	79.84±1.05	93.33±1.48	71.60 ±2.36	77.50±0.16
	$+\mathcal{L}_{\text{consistency}}$	79.45±1.09	77.46±1.96	74.98±4.57	75.51±2.63	74.50±3.06	53.46±2.44	84.16±0.81	94.03±1.33	74.57±1.61	77.64±0.05
Xu et al. NeurIPS' 17	GraphSAGE	74.40±1.83	73.17±0.47	74.96±3.14	76.44±4.16	73.90±2.17	51.33±2.95	78.92±1.20	89.56±2.37	77.03±1.65	76.67±0.11
	$+\mathcal{L}_{\text{consistency}}$	78.26±1.08	74.10±2.10	76.40±3.12	77.50±3.38	74.75±3.06	54.27±1.24	82.12±0.78	92.31±1.32	78.60±1.44	77.57±0.05
Shi et al. IJCAI' 21	GTransformer	75.72±2.69	74.79±1.82	73.33±4.80	75.42±3.22	72.20±3.49	53.33±1.12	80.36±0.56	83.74±3.17	76.81±1.34	76.75±0.12
	$+\mathcal{L}_{\text{consistency}}$	76.83±1.36	75.82±1.53	77.03±3.79	76.57±2.54	73.75±2.56	56.53±1.54	80.48±0.47	91.67±1.88	76.90±3.25	77.14±0.06
Baek et al. ICLR' 21	GMT	75.04±1.43	73.90±2.29	72.70±4.21	72.80±2.19	79.80±1.08	54.13±2.90	80.36±1.15	90.85±1.91	74.86±2.26	72.06±10.15
	$+\mathcal{L}_{\text{consistency}}$	75.52±1.07	75.20±0.95	74.86±2.03	73.14±2.28	79.60±1.91	54.80±1.42	82.80±0.61	92.00±1.43	76.00±1.99	77.19±0.14

Table 1: Performance on the graph classification tasks, with and without consistency loss. The highlighted cells indicate instances where GNNs with our proposed consistency loss outperform the base GNNs.

Consistency

	NCI1	NCI109	PROTEINS	D&D	IMDB-B
GCN	0.753	0.920	0.584	0.709	0.846
$+\mathcal{L}_{\text{consistency}}$	0.859	0.958	0.946	0.896	0.907
GIN	0.666	0.674	0.741	0.721	0.598
$+\mathcal{L}_{\text{consistency}}$	0.877	0.821	0.904	0.847	0.816
GraphSAGE	0.903	0.504	0.845	0.741	0.806
$+\mathcal{L}_{\text{consistency}}$	0.911	0.709	0.916	0.872	0.933
GTransformer	0.829	0.817	0.867	0.865	0.884
$+\mathcal{L}_{\text{consistency}}$	0.863	0.883	0.915	0.880	0.917
GMT	0.872	0.887	0.980	0.826	0.893
$+\mathcal{L}_{\text{consistency}}$	0.906	0.908	0.983	0.856	0.908

Table 2: Spearman Correlation in Consecutive Layer Graph Representations. The representation space becomes more aligned with the proposed consistency loss.

Efficiency

	GMT	GTransformer	GIN	GCN	GraphSAGE
GCN	8.380	4.937	4.318	4.221	3.952
GCN+$\mathcal{L}_{\text{consistency}}$	8.861	6.358	5.529	5.382	5.252

Table 3: Training Time per Epoch (Seconds, OGBG-MOLHIV): Impact of Consistency Loss is Minimal

References

[1]Nino Shervashidze et al. (2010). "Weisfeiler-lehman graph kernels." In: Journal of Machine Learning Research.
[2]Nils M. Kriege, et al. (2016). "On valid optimal assignment kernels and applications to graph classification" In: In Advances in Neural Information Processing Systems.
[3]Keyulu Xu et al. (2019). "How Powerful are Graph Neural Networks?" In: International Conference on Learning Representations.

Code

Paper