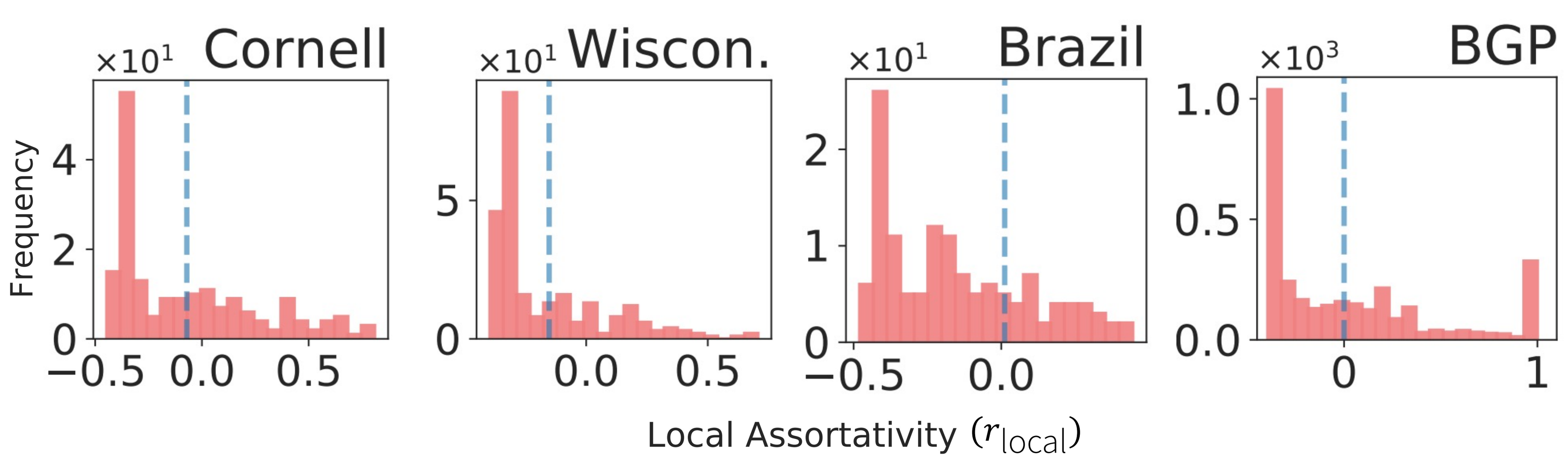
Breaking the Limit of Graph Neural Networks by Improving the Assortativity of Graphs with Local Mixing Patterns

Susheel Suresh[1], Vinith Budde[2], Jennifer Neville[1], Pan Li[1] and Jianzhu Ma[1] | [1]Dept. of Computer Science, Purdue University, [2]Alexa AI, Amazon

Introduction

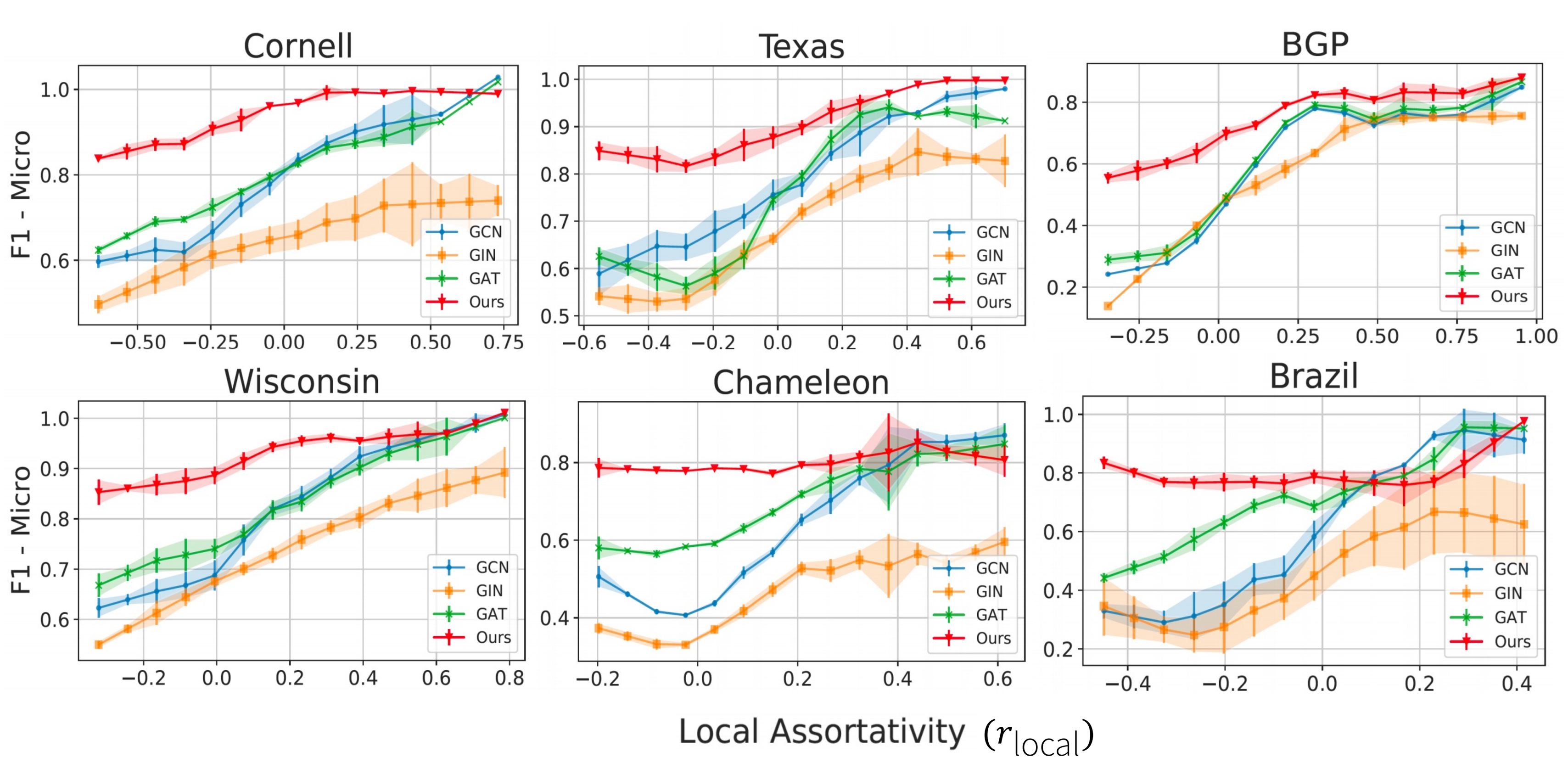
- GNNs are the de facto standard for graph representation learning. They perform neighborhood smoothing guided by the principle: proximity information from the surroundings of a node is a useful descriptor for predicting its labels.
- Mixing patterns in graphs characterize how nodes mix/connect based on their attributes. We show that the prediction performance of a wide range of GNNs are highly correlated with the notion of local assortativity in graphs.
- Motivated by this limit, we propose a graph transformation technique to boost GNN performance.

Heterogeneous mixing observed in real world networks



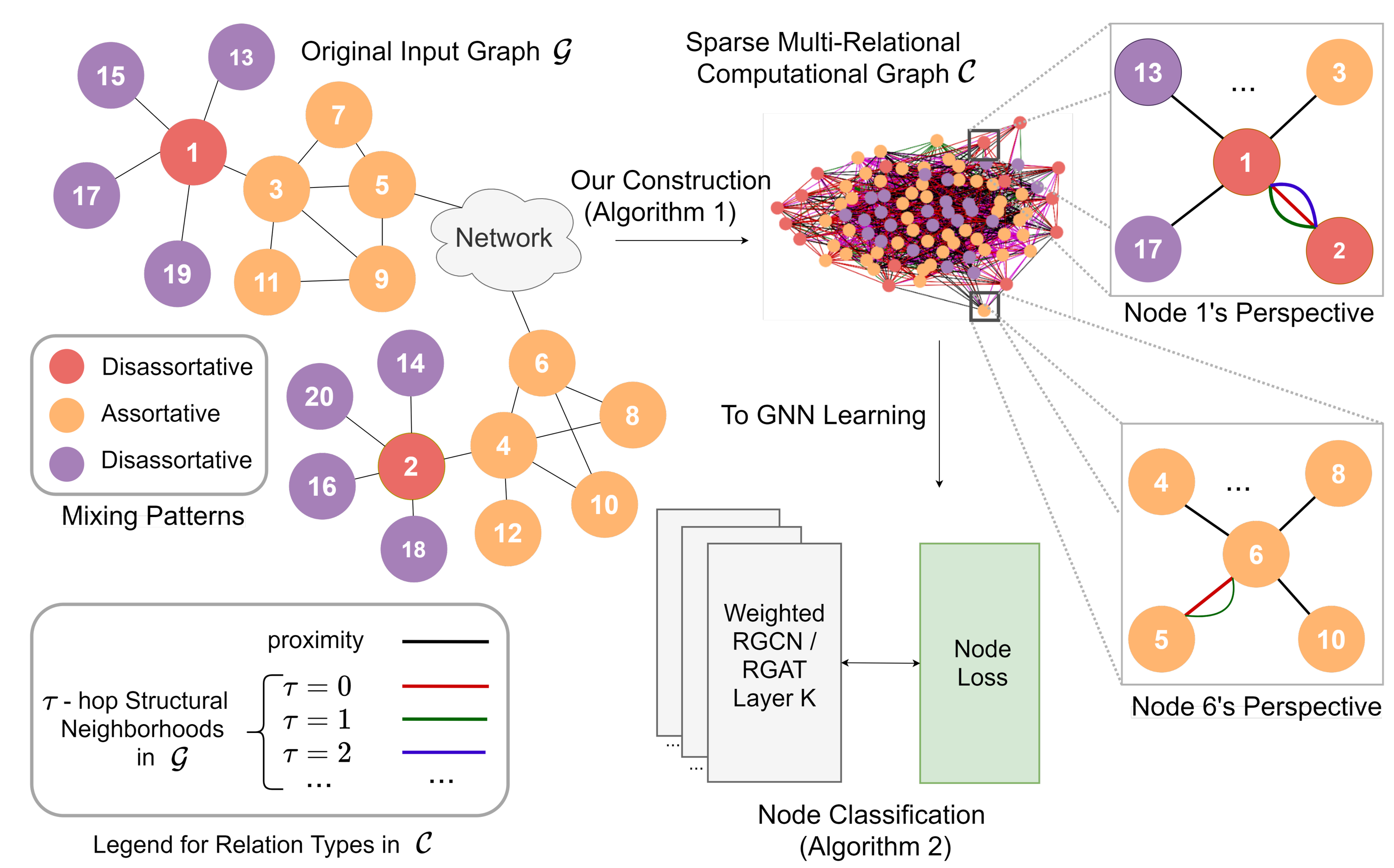
- Skewed and multimodal distributions observed.
- r_{global} (blue dotted line) measures average mixing pattern for the whole network. Fails to capture heterogeneous mixing patterns.

GNNs and local mixing



- GNNs behave poorly in disassortative regime. However, in assortative regime, they show strong performance.

Improving GNNs in disassortative regime



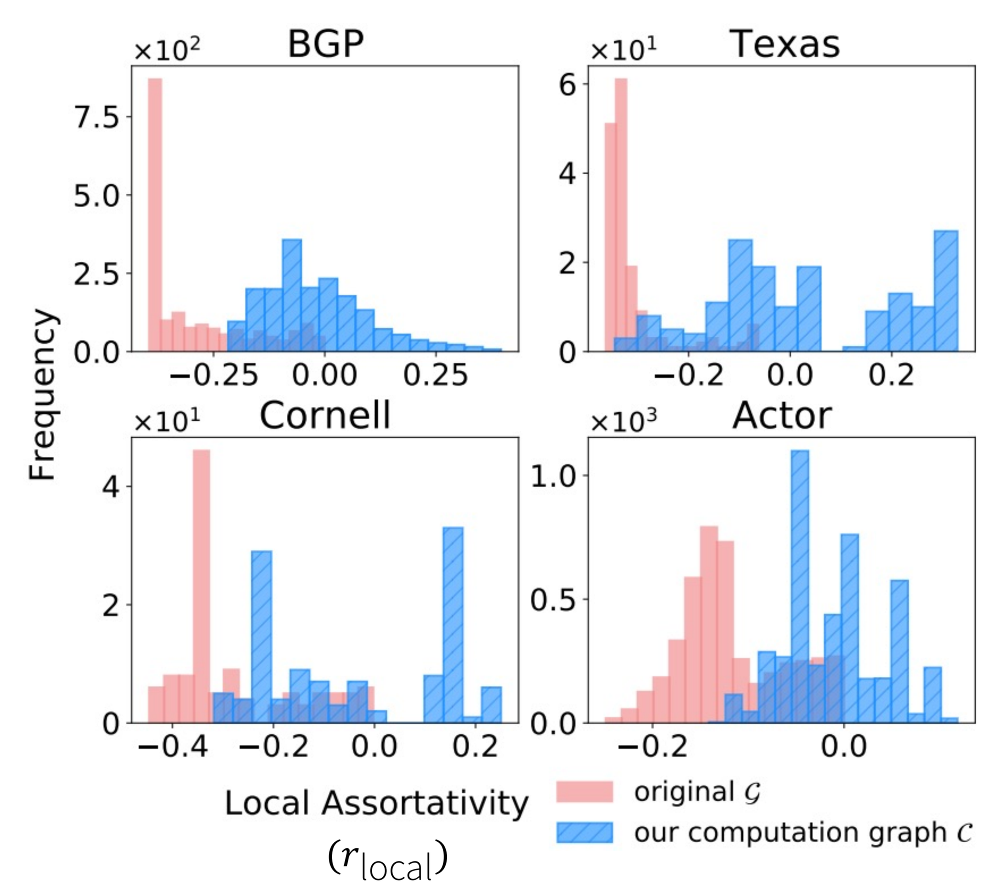
Our idea:

- Transform the original graph using both structural and proximity information to form a computation graph.
- Run GNNs on the computation graph. Adaptively choose between structure and proximity using attention.

Key ingredients:

- Pairwise node structural similarity measure by comparing ordered sequences at various neighborhoods.
- Pairwise node proximal similarity measure using original edges.
- Relations in computation graph encode the above notions.

Our computation graph is more assortative



- The shift in local assortativity shows computation graph is inherently more assortative.
- Explicitly captures structural similarity of far away disassortative nodes.

Experiments and analysis

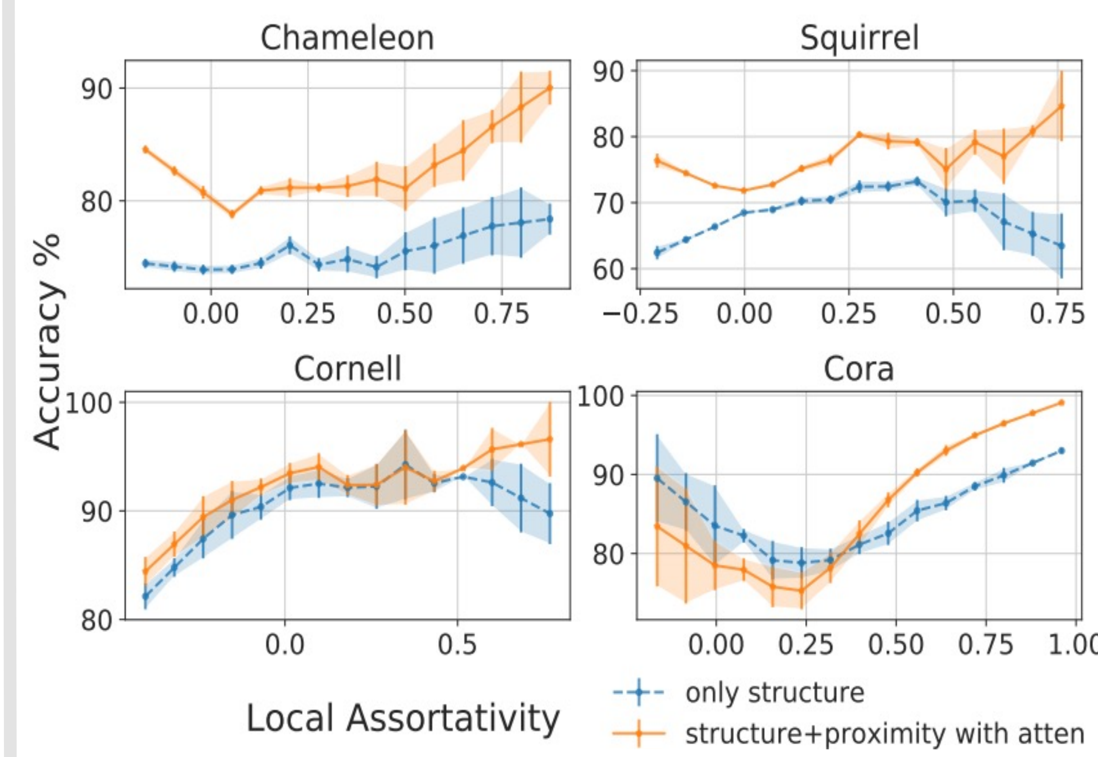
Table 1: Semi-supervised node classification showing mean test accuracy ± std. over 10 runs. Club Suit [♣] denotes result obtained from the best model variant of respective papers.

	Chameleon	Squirrel	Actor	Cornell	Texas	Wisconsin	Cora	Citeseer	Pubmed
GCN	59.82 ± 2.58	36.89±1.34	30.26±0.79	57.03±4.67	59.46±5.25	59.80±6.99	87.28±1.26	76.68±1.64	87.38±0.66
GraphSage	58.73±1.68	41.61±0.74	34.23±0.99	75.95±5.01	82.43±6.14	81.18±5.56	86.90±1.04	76.04±1.30	88.45±0.50
GAT	54.69±1.95	30.62±2.11	26.28±1.73	58.92±3.32	58.38±4.45	55.29±8.71	86.37±1.69	75.46±1.72	87.62±0.42
GCN-Cheby	55.24±2.76	43.86±1.64	34.11±1.09	74.32±7.46	77.30±4.07	79.41±4.46	86.86±0.96	76.25±1.76	88.08±0.52
MixHop	60.50±2.53	43.80±1.48	32.22±2.34	73.51±6.34	77.84±7.73	75.88±4.90	83.10±2.03	70.75±2.95	80.75±2.29
Geom-GCN ♣	60.90	38.14	31.63	60.81	67.57	64.12	85.27	77.99	90.05
H_2GCN ♣	59.39±1.98	37.90±2.02	35.86±1.03	82.16±4.80	84.86±6.77	86.67±4.69	87.67±1.42	76.72±1.50	88.50±0.64
Ours (WRGAT)	65.24±0.87	48.85±0.78	36.53±0.77	81.62±3.90	83.62±5.50	86.98±3.78	88.20±2.26	76.81±1.89	88.52±0.92

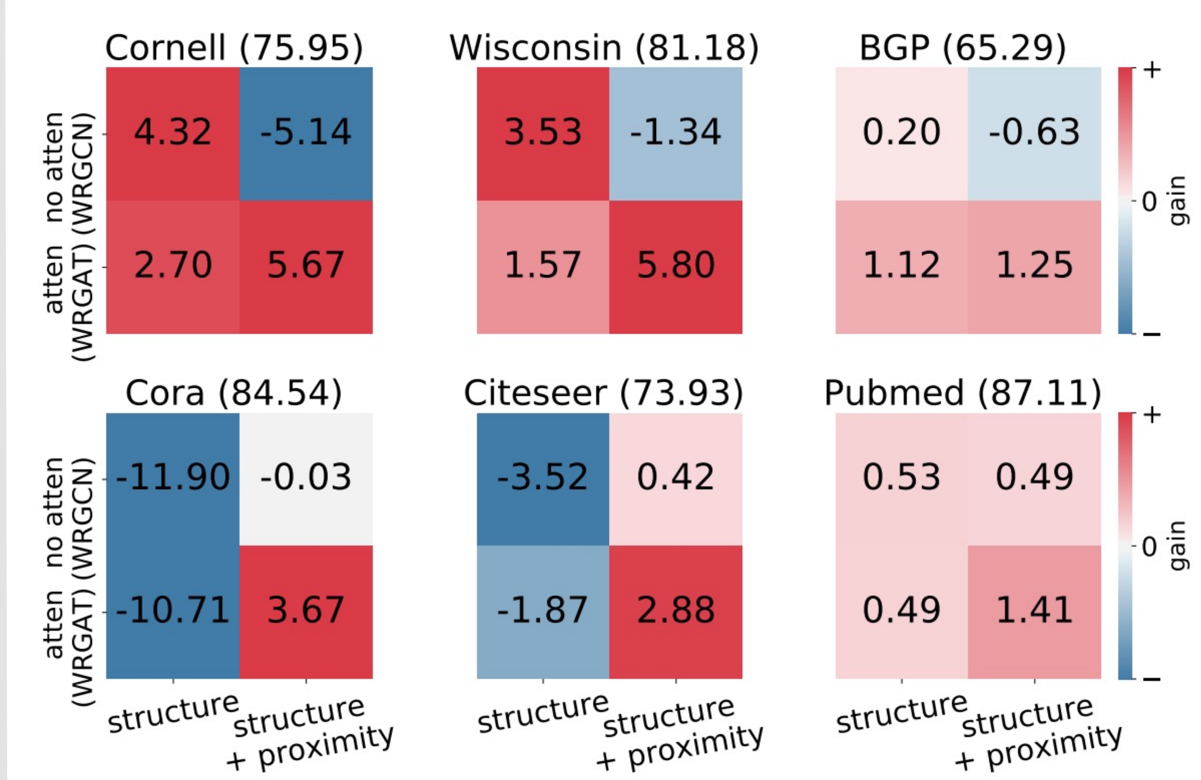
Table 2: Node classification on Air Traffic Networks and BGP Network. Mean test acccuracy ± std. is shown over 20 runs.

	Brazil	Europe	USA	BGP
GCN	64.55±4.18	54.83±2.69	56.58±1.11	53.33±0.18
GraphSage	70.65±5.33	56.29±3.21	50.85±2.83	65.19±0.28
GIN	71.89±3.60	57.05±4.08	58.87±2.12	49.51±1.52
Struc2vec	70.88±4.26	57.94±4.01	61.92±2.61	48.40±1.39
Ours	76.92±5.45	57.12±2.81	63.02±1.87	66.54±0.48

- Stacking multiple GNN layers known to cause over-smoothing issues. Exacerbated when higher order neighborhoods are utilized (MixHop, H2GCN)
- Our WRGAT method achieves best overall rank among 12 benchmark datasets.



- In high assortative regime, structure is not so important.
- Attention mechanism helps.



- Significant gain over proximity only baseline when structure + proximity information is adaptively chosen.

Contact and code

suresh43@purdue.edu
@susheel_suresh

github.com/susheels/gnns-and-local-assortativity