SAND2020-1608C

How Robust Are Graph Neural Networks to Structural Noise?

James Fox, Sivasankaran Rajamanickam

Georgia Institute of Technology, Sandia National Laboratories

Georgia Tech College of Computing — Computational Science and Engineering

Sandia National Laboratories

Introduction

GNNs are popular models for learning on graph-structure data, but their robustness is not well-understood. We study robustness in context of node structural identity predictions and explore augmented training for improving robustness.

Key points:

- GNNs can perfectly distinguish structural identity (without noise)
- GNN accuracy sharply declines with structural noise (random edge additions)
- Augmented training with generated noisy samples can improve GNN robustness

Generated Graphs

Graphs are generated from structural motifs according to [1]. Node labels are from well-defined structural role.

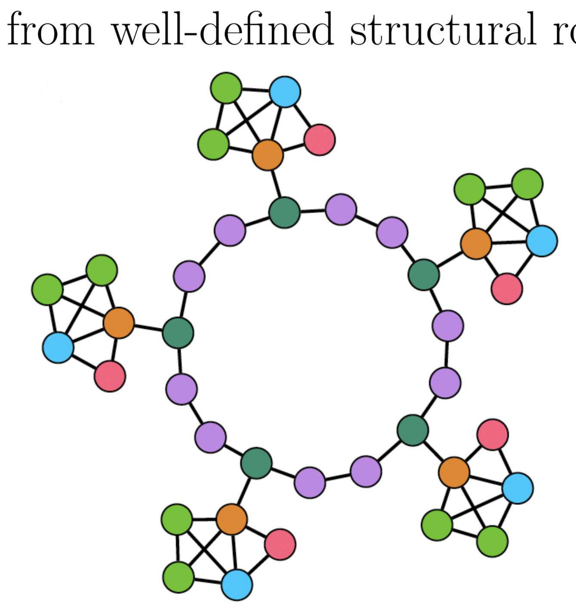


Figure: 'Ring of houses' graph. Image from [1].

Size of base graphs used. G' is downsized version of G. G is ring-of-houses in all experiments.

Name	Nodes	Edges	Classes
Ring of houses G	2664	3996	6
Ring of houses G'	264	396	6

Structural Noise

We introduce structural noise in the form of random edge additions, but keep original node labels. Labels no longer neatly match structural identity.

Noise model parameters

- Noise ratio p: how many edge additions as ratio of original edges
- Distance k: can only form new edges from nodes within k hops.

New edge pairs are sampled uniformly at random, under k-hop constraint.

Augmented Robustness Training

Graph training samples is often limited. Can we augment training with generated noisy samples to improve robustness?

Noisy Augmentation Method

- Generating from same distribution: $G_p^{(j)}$ is j-th noisy graph generated from G (of same size). We use 1 in practice to augment training.
- Generating from similar distribution: $G_p'^{(j)}$ is j-th noisy graph generated from G' (smaller version of G). We use 10 such graphs to augment.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525

Model

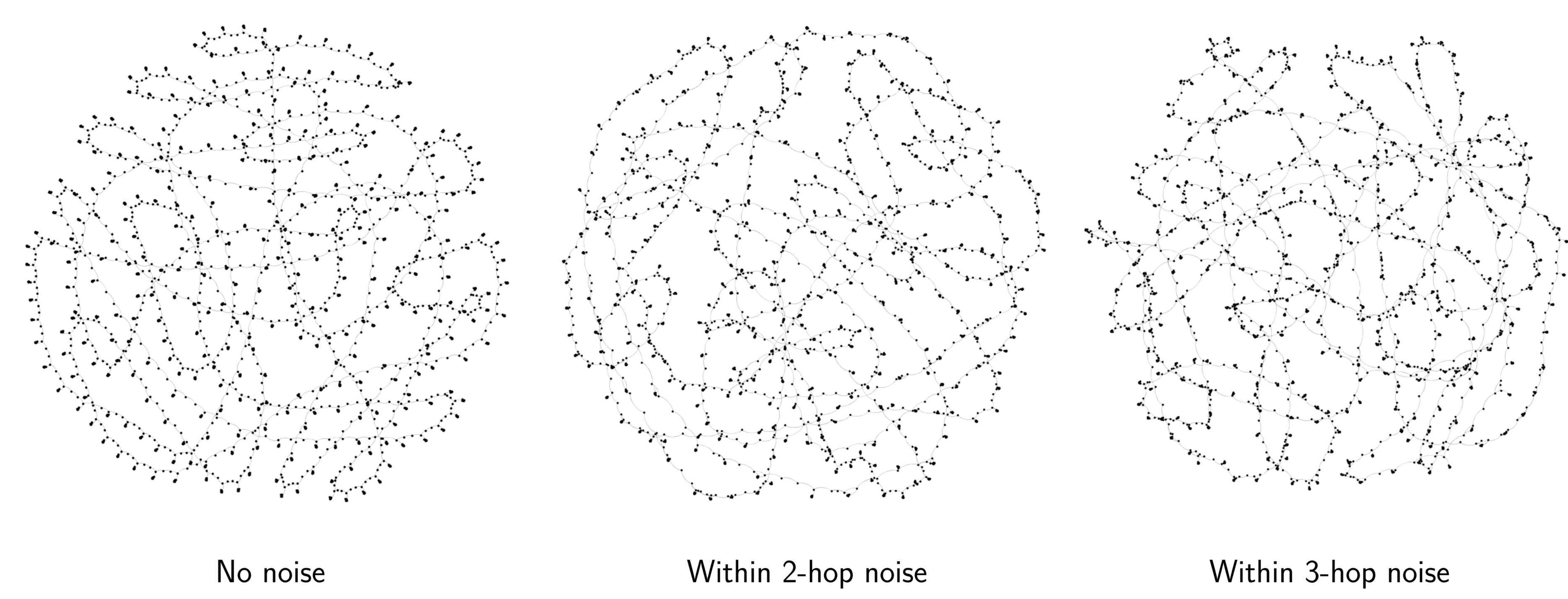
We use Graph Ismorphism Network (GIN) [3] as the GNN. Implemented using PyTorch Geometric [2].

- Architecture: 3 GIN layers, followed by two fully-connected (FC) layers
- Each GIN layer is also composed of two FC layers. Batchnorm applied follows each GIN layer. ReLU activation after linear transformations.
- Only feature is node degree (normalized)

Experiment Setup

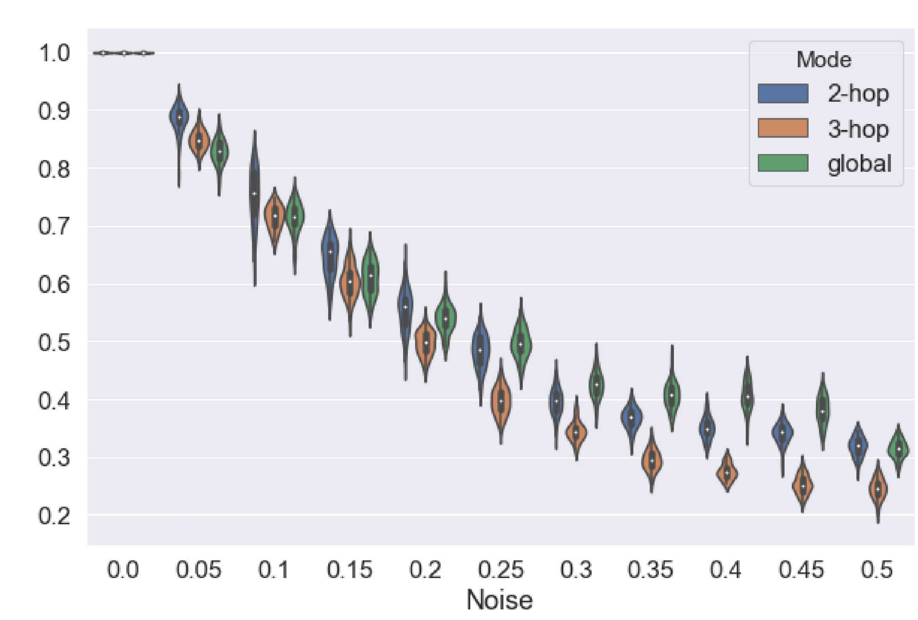
- Results from 50 independent trials
- Training set: 20 node labels per class from G_p
- Validation set: 200 node labels
- Test set: 1000 nodes
- Test score (F_1-macro) is from evaluating model achieving best validation score after training for 200 epochs

Experiments and Results



GNN performance vs. structural noise

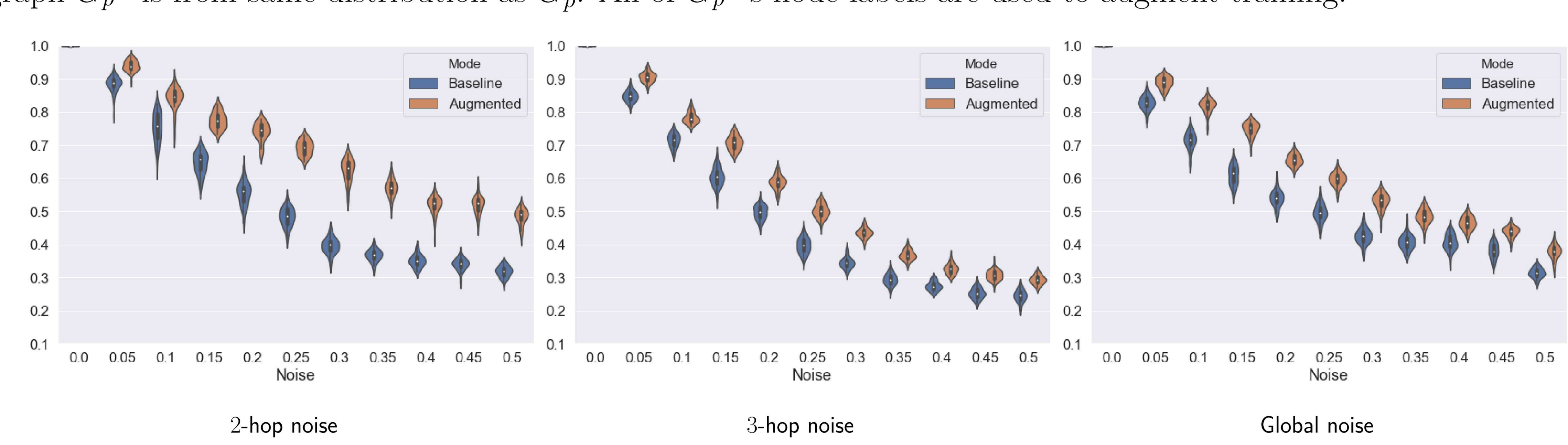
We vary the ratio p of noisy edges added to G in increments of 0.05, and evaluate performance trained on each version of G_p. We evaluate for 3 different modes of noise: 2-hop, 3-hop, or unconstrained (global).



Findings: With no noise ($p = 0$), the GIN learns to classify nodes near perfectly. $F1$-score declines sharply with increasing p (randomly added edges)–median performance is below 50% at 25% noise, and below 35% at 50% noise, across all modes.

Augmented training from same distribution

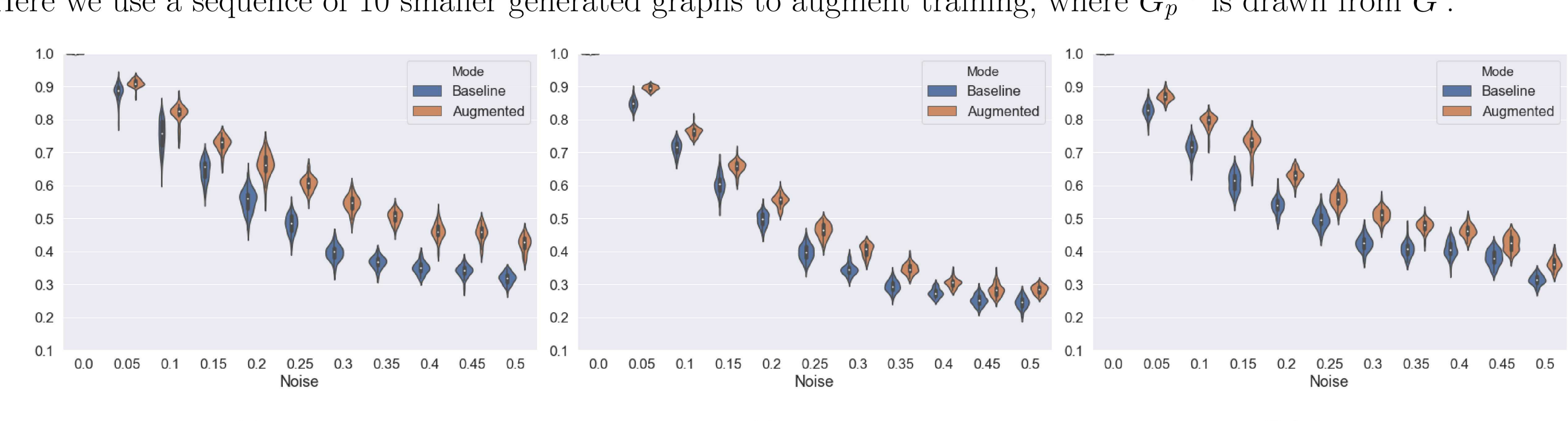
We compare performance from augmented training vs. non-augmented training (baseline) on G_p. The augmenting graph $G_p^{(1)}$ is from same distribution as G_p. All of $G_p^{(1)}$'s node labels are used to augment training.



Findings: Training augmentation with graph drawn from the same noise distribution gives relative improvement of median F1 score up to 59%, 26%, and 26% for 2-hop, 3-hop, and global noise modes.

Augmented training with smaller graphs

Here we use a sequence of 10 smaller generated graphs to augment training, where $G_p'^{(j)}$ is drawn from G'.



Findings: Augmented training is beneficial even when the smaller graphs are not of exactly same distribution as G_p. Relative improvements (of median) are 38%, 18%, and 20% for 2-hop, 3-hop, and global noise modes.

References

[1] C. Donnat, M. Zitnik, D. Hallac, and J. Leskovec. Learning structural node embeddings via diffusion wavelets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1320–1329. ACM, 2018.

[2] M. Fey and J. E. Lenssen. Fast graph representation learning with pytorch geometric, 2019.

[3] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.