Learning GNN in Noisy Settings

JFLI Tokyo Tech Workshop 2019

Hoang NT

Murata Laboratory
Tokyo Tech

2019/05/16

Overview

GCN Introduction

There are two main problems:

Popular Problems on Graph-Structured Data

There are two main problems:

- Answering questions on vertices.

There are two main problems:

- Answering questions on vertices.
- Answering questions about the graph.

There are two main problems:

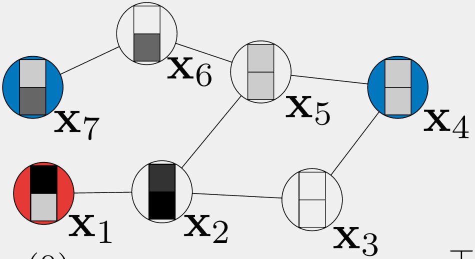
- Answering questions on vertices.
- Answering questions about the graph.

Semi-supervised vertex classification

Given a graph structured data $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}})$, we want to find: $\mathcal{C}_i \ \forall i \in V(A) - J_{\text{train}}$.

Semi-supervised graph classification

Given a collection of graphs $\mathcal{S} = (\{A_i\}, \mathcal{X}_i, \mathcal{C}_{\mathcal{G}}, J_{\text{train}})$, we want to find: $\mathcal{C}(\mathcal{G}_i) \ \forall i \in \{A_j\} - J_{\text{train}}$.



$$\mathbf{H}^{(0)} = \mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$$

Figure: Vertex Classification [10]

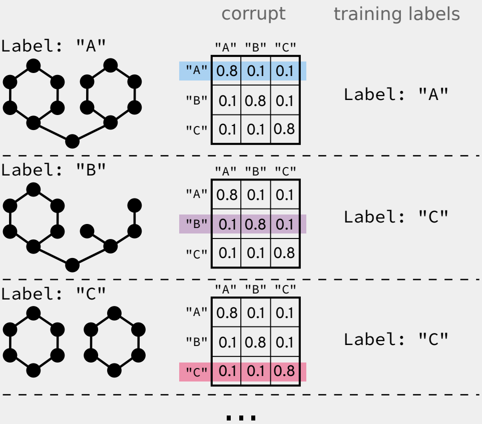


Figure: Graph Classification [4]

Graph representation learning:

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]
- Weifesler-Lehman isomorphism test

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]
- Weifesler-Lehman isomorphism test

Common theme: Graph

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]
- Weifesler-Lehman isomorphism test

Common theme: Graph $\rightarrow$ Surrogate Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]
- Weifesler-Lehman isomorphism test

Common theme: Graph $\rightarrow$ Surrogate Structure $\rightarrow$ Manifold Learning

Approach: Only Use Graph Structure

Graph representation learning:

- Spectral Clustering [7], Modularity [6]
- Label Propagation [13]
- Deepwalk [5], node2vec [1]
- Weifesler-Lehman isomorphism test

Common theme: Graph $\rightarrow$ Surrogate Structure $\rightarrow$ Manifold Learning $\rightarrow$ Vertex/Graph Representations.

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

Approach: Graph Structure + Other Information

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]

Approach: Graph Structure + Other Information

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]

Approach: Graph Structure + Other Information

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]
- GIN [11]

Approach: Graph Structure + Other Information

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]
- GIN [11]

Common theme: Graph + Features

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]
- GIN [11]

Common theme: Graph + Features $\rightarrow$ Average Features

Approach: Graph Structure + Other Information

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]
- GIN [11]

Common theme: Graph + Features $\rightarrow$ Average Features $\rightarrow$ Loss Backpropagation

Using $\mathcal{G} = (A, \mathcal{X}, \mathcal{C}, J_{\text{train}}, \mathcal{C})$

- SemiEmb [9], Planetoid [12]
- GCN [3], SGC [10]
- GraphSAGE [2], DGI [8]
- GIN [11]

Common theme: Graph + Features $\rightarrow$ Average Features $\rightarrow$ Loss Backpropagation $\rightarrow$ Vertex/Graph Representations.

Visually...

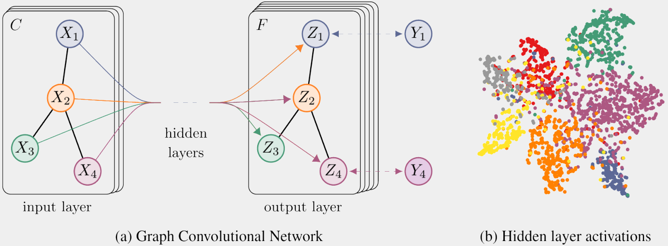


Figure: A Doodle of GCN[3]

Label Noise

Wrong labels are given to training data

We addresses the error of GCN models under “wrong” labels.

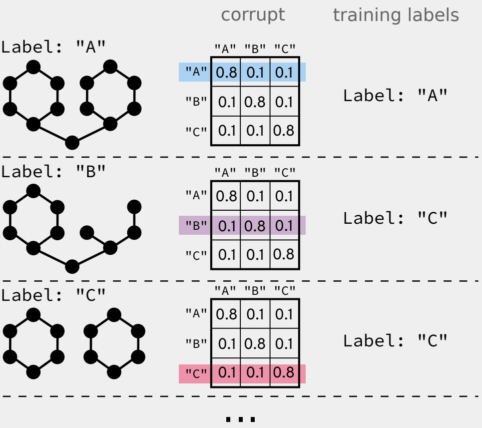


Figure: Noisy training

Observation

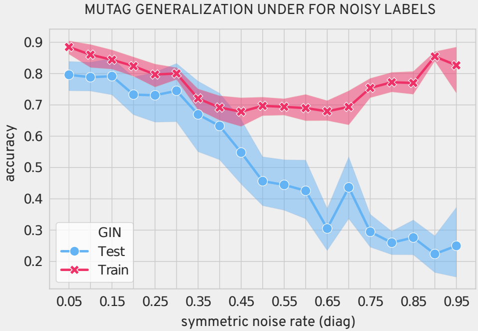


Figure: MUTAG (binary clf) under symmetric label noise

1. Accumulate neighborhood information to each vertex $\boldsymbol{v}$ and neural network layer (k):

$$\mathbf{a}_v^{(k)} = \texttt{AGGREGATE}^{(k)}(\{\mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v)\}),$$
$$\mathbf{h}_v^{(k)} = \texttt{COMBINE}^{(k)}(\mathbf{h}_v^{(k-1)}, \mathbf{a}_v^{(k)})$$

2. Employ a function to learn the overall representation of the graph, then the objective ℓ is optimized with standard backpropagation.

$$\mathbf{h}_{\mathcal{G}} = \texttt{READOUT}(\{\mathbf{h}_v^{(K)} : v \in \mathcal{G}\}),$$
$$\ell(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}}) = \texttt{XENTROPY}(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}})$$

Visually...

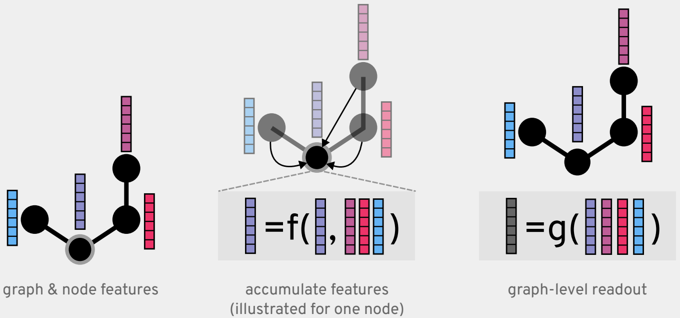


Figure: Simple GNN

Simple Loss Correction

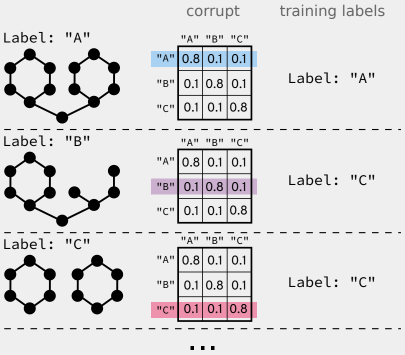


Figure: Graph Classification [4]

$$\ell^{\leftarrow}(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}}) = \boldsymbol{C}^{-1} \cdot \texttt{CROSS\_ENTROPY}(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}})$$

	IMDB-M	COLLAB	IMDB-B	PROTEINS	PTC	NCI1
GIN	.4476	.6544	.6573	.6257	.4824	.6472
GraphSAGE	.4373	-	.6410	.6583	.4892	.6053
D-GNN-C	**.4747**	.5979	**.6940**	**.6693**	**.5557**	.6170
D-GNN-A	**.4505**	**.6917**	**.7088**	**.6769**	**.5001**	.6405
D-GNN-E	**.4633**	**.6960**	**.7190**	**.6917**	**.5235**	**.6638**

The Bad

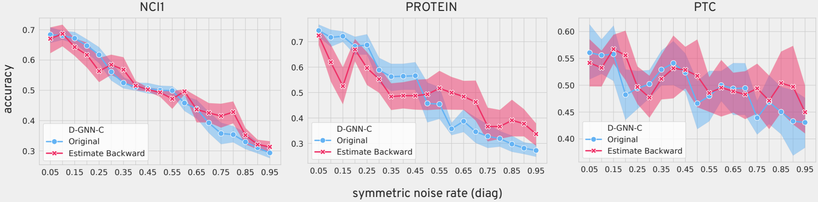


Figure: Performance is not particularly good

Another Perspective

GNN is just denoising

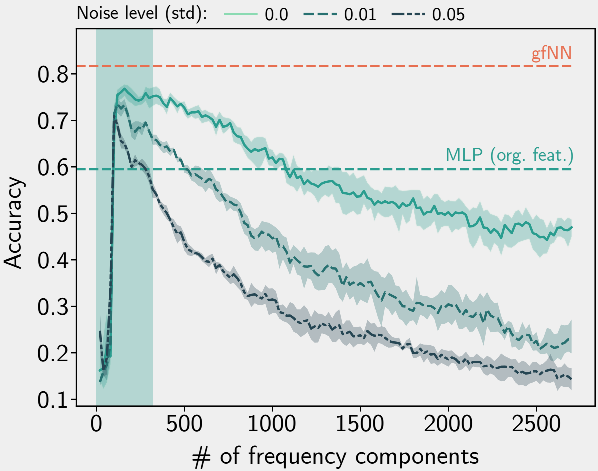


Figure: Accuracy of a MLP on Cora dataset.

Thanks for listening!

Aditya Grover and Jure Leskovec.
node2vec: Scalable feature learning for networks.
In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.

Will Hamilton, Zhitao Ying, and Jure Leskovec.
Inductive representation learning on large graphs.
In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

Thomas N Kipf and Max Welling.
Semi-supervised classification with graph convolutional networks.
arXiv preprint arXiv:1609.02907, 2016.

Hoang NT, Jun Jin Choong, and Tsuyoshi Murata.
***LLD Workshop, ICLR*, 2019.**

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena.
Deepwalk: Online learning of social representations.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

L. Tang and H. Liu.
Efficient graphlet kernels for large graph comparison.
pages 817–826, 2009.

L. Tang and H. Liu.
Reconstructing markov processes from independent and anonymous experiments.
pages 447–478, 2011.

Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm.
Deep graph infomax.
arXiv preprint arXiv:1809.10341, 2018.

Jason Weston, Frédéric Ratle, Hossein Mobahi, and Ronan Collobert.
Deep learning via semi-supervised embedding.
In *Neural Networks: Tricks of the Trade*, pages 639–655. Springer, 2012.

Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr, Christopher Fifty, Tao Yu, and Kilian Q Weinberger.
Simplifying graph convolutional networks.
arXiv preprint arXiv:1902.07153, 2019.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka.
How powerful are graph neural networks?
International Conference on Learning Representations (ICLR), 2019.

Zhilin Yang, William W Cohen, and Ruslan Salakhutdinov.
Revisiting semi-supervised learning with graph embeddings.
arXiv preprint arXiv:1603.08861, 2016.

Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty.
Semi-supervised learning using gaussian fields and harmonic functions.
pages 912–919, 2003.