Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning

Qimai Li, Zhichao Han, Xiao-Ming Wu
Department of Computing
The Hong Kong Polytechnic University

Supervised Learning

Tons of labeled data ➡ A good model

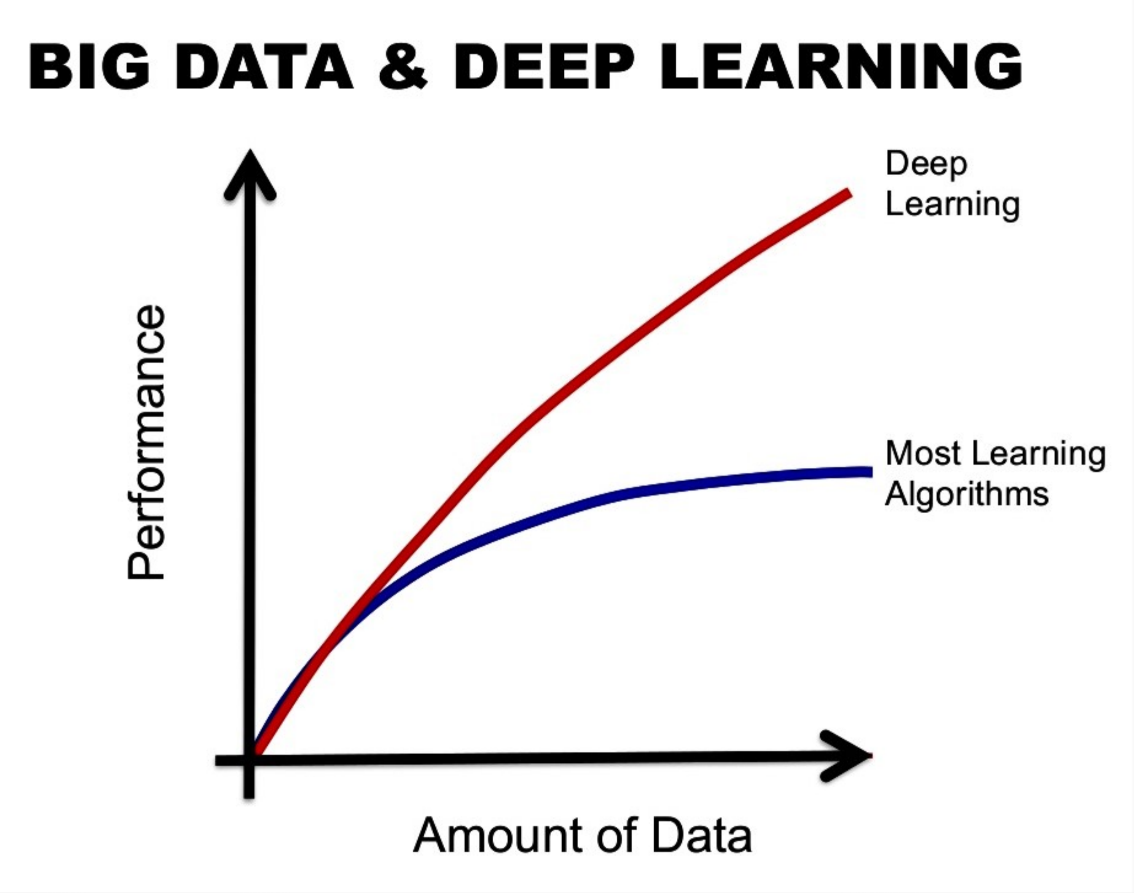
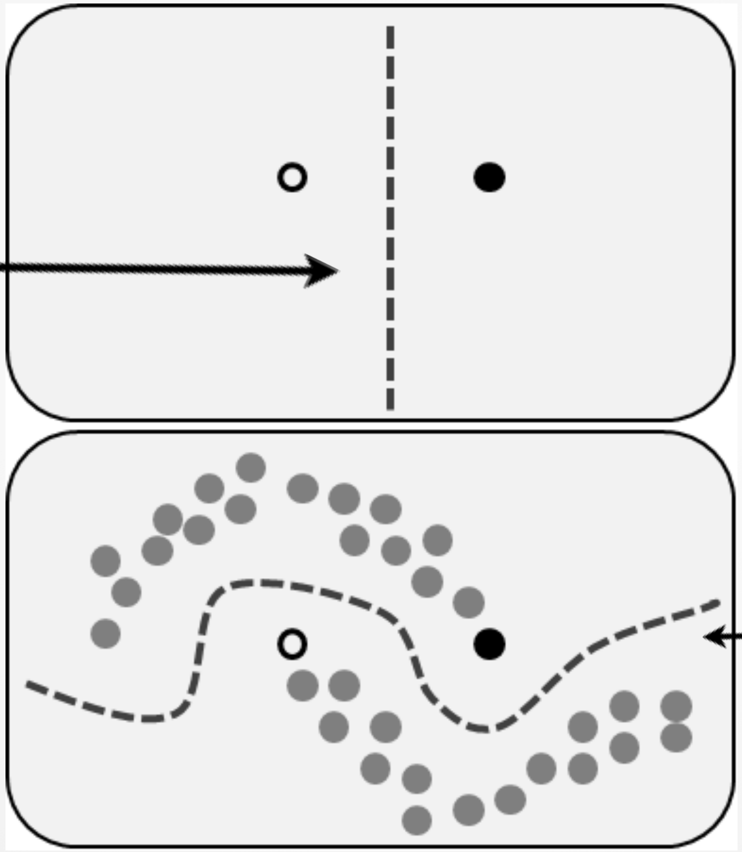


Image via “https://www.linkedin.com/pulse/deep-learning-aviation-francis-j-duque”

Semi-Supervised Learning (SSL)

How unlabeled data helps?

Decision
Boundary



Better
Decision
Boundary

Image via “https://en.wikipedia.org/wiki/Semi-supervised_learning”

How to Leverage Unlabeled Data

Node -> Document   Edge -> Citation Link

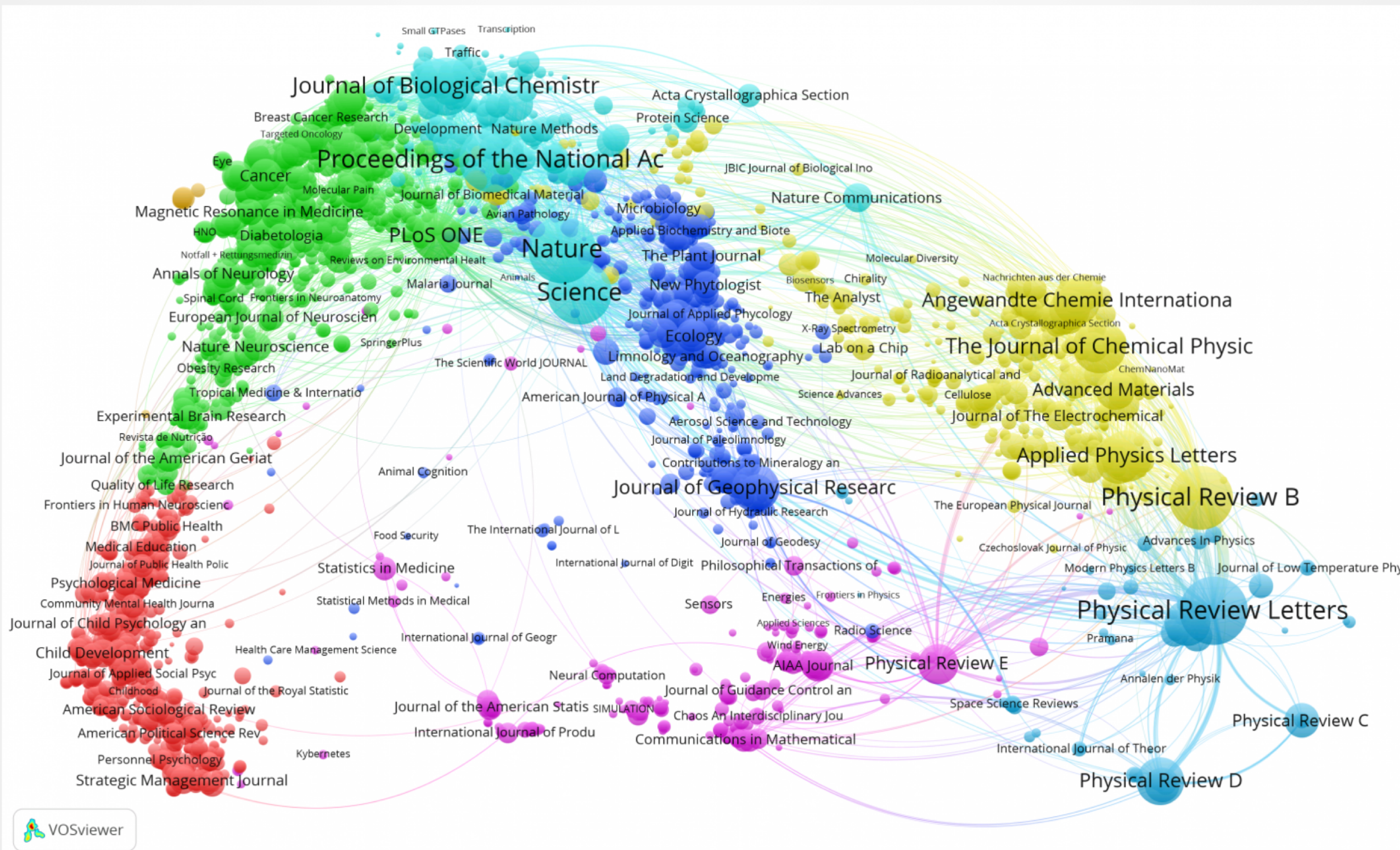


Image via https://www.cwts.nl/media/images/content/b515d3b727bc41fe7e858df0ffd062bf_large.png

Graph Convolutional Networks

(Kipf & Welling, ICLR, 2017)

Layer-wise propagation rule:

$$H^{(l+1)} = \sigma\left(AH^{(l)}\Theta^{(l)}\right)$$

Convolution layer: preprocessed adjacency matrix

Projection layer: fully connected networks

GCNs for semi-supervised classification:

$$Z = \mathrm{softmax}\left(A\sigma\left(AX\Theta^{(0)}\right)\Theta^{(1)}\right)$$

$$\mathcal{L} := -\sum_{i\in\mathcal{V}_l}\sum_{f=1}^{F} Y_{if}\ln Z_{if}$$

Why GCNs Work

$$H^{(1)} = \sigma\left(AX\Theta^{(0)} \right)$$

Convolution layer: preprocessed adjacency matrix

Projection layer: fully connected networks

Table 1: GCNs vs. Fully-connected networks

One-layer FCN	Two-layer FCN	One-layer GCN	Two-layer GCN
0.530860	0.559260	**0.707940**	**0.798361**

Laplacian Smoothing

$$H^{(l+1)} = \sigma\left(\boxed{AH^{(l)}} \Theta^{(l)} \right)$$

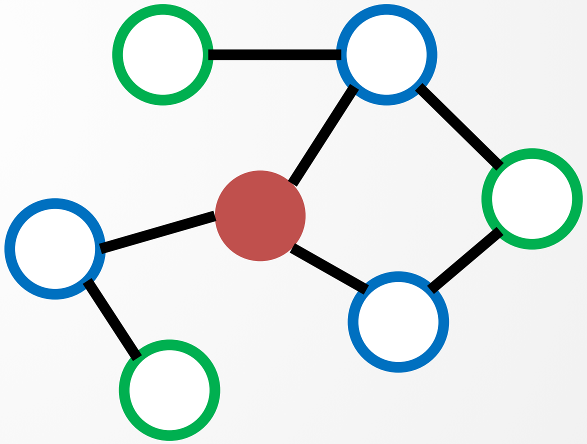
Smoothing

Limitations of GCNs (1)

$$H^{(l+1)} = \sigma\left(\boxed{A} H^{(l)} \Theta^{(l)}\right)$$

Localized filter

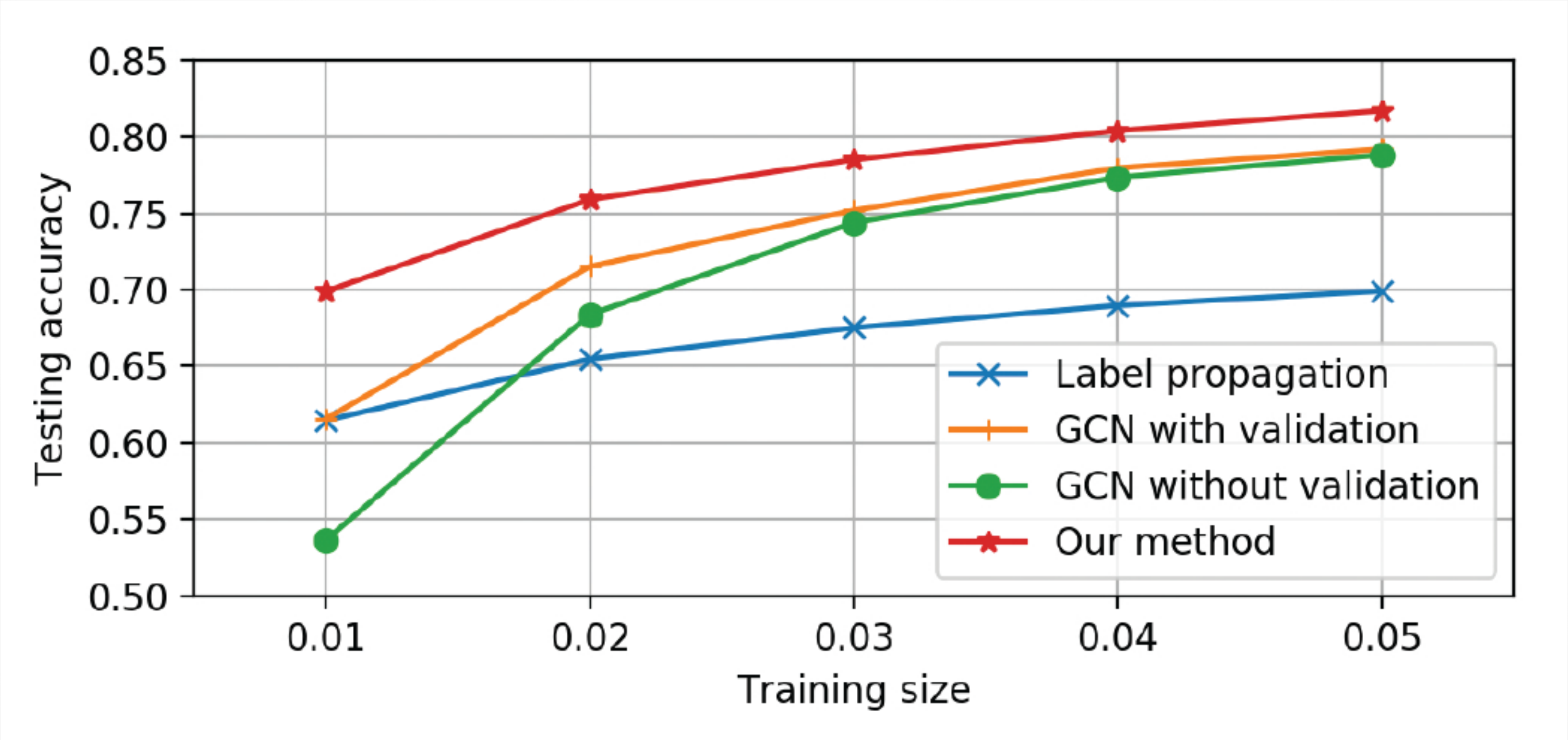
- Labeled instance
- Unlabeled instance
- Instance adjacent to a labeled instance



Need to stack many layers to explore global graph topology when labeled data is few - overfitting

Limitations of GCNs (2)

Need additional labeled data for model selection



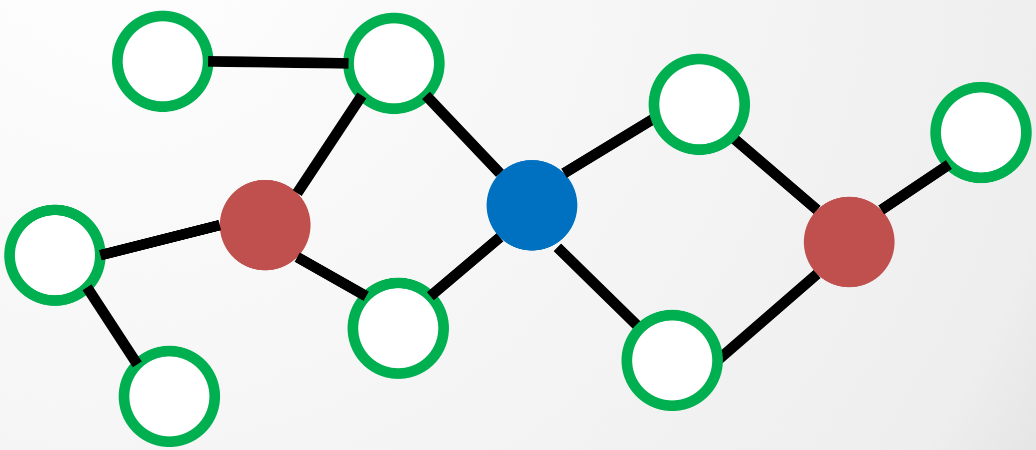
Our Solutions (1)

- Co-train a GCN with a random walk model
  - Use random walks to explore global topology
  - Extend the labeled set with high-confidence predictions by the random walk model

Labeled instance

Unlabeled instance

Pseudo labeled instance



Our Solutions (2)

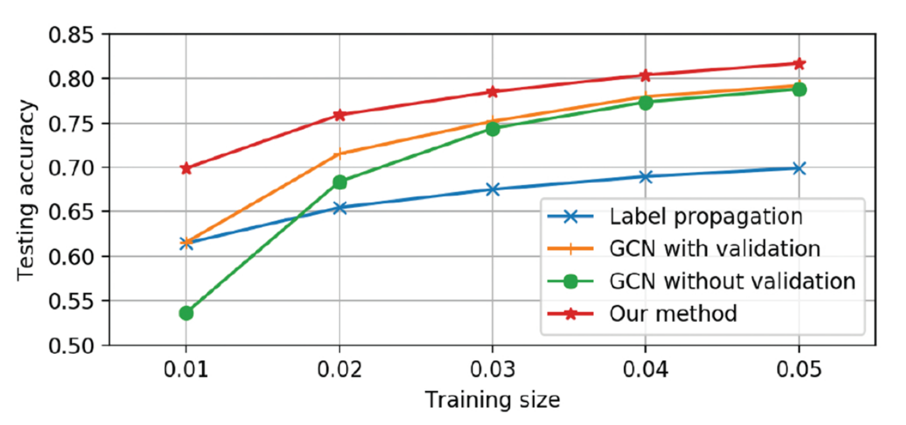
- Self-training
  - Extend the labeled set with high-confidence predictions by a pre-trained GCN
- Union of Self-training and Co-training
  - Add to the diversity of pseudo labels
- Intersection of Self-training and Co-training
  - Get more accurate pseudo labels

Experimental Results

Significant improvements on 3 citation networks

Cora

Label Rate	0.5%	1%	2%	3%	4%	5%
LP	56.4	62.3	65.4	67.5	69.0	70.2
Cheby	38.0	52.0	62.4	70.8	74.1	77.6
GCN-V	42.6	56.9	67.8	74.9	77.6	79.3
GCN+V	50.9	62.3	72.2	76.5	78.4	79.7
Co-training	56.6	66.4	73.5	75.9	78.9	80.8
Self-training	53.7	66.1	73.8	77.2	79.4	80.0
Union	**58.5**	**69.9**	**75.9**	**78.5**	**80.4**	**81.7**
Intersection	49.7	65.0	72.9	77.1	79.4	80.2



CiteSeer

Label Rate	0.5%	1%	2%	3%	4%	5%
LP	34.8	40.2	43.6	45.3	46.4	47.3
Cheby	31.7	42.8	59.9	66.2	68.3	69.3
GCN-V	33.4	46.5	62.6	66.9	68.4	69.5
GCN+V	43.6	55.3	64.9	67.5	68.7	69.6
Co-training	**47.3**	55.7	62.1	62.5	64.5	65.5
Self-training	43.3	58.1	68.2	69.8	70.4	71.0
Union	46.3	**59.1**	66.7	66.7	67.6	68.2
Intersection	42.9	**59.1**	**68.6**	**70.1**	**70.8**	**71.2**

PubMed

Label Rate	0.03%	0.05%	0.1%	0.3%
LP	61.4	66.4	65.4	66.8
Cheby	40.4	47.3	51.2	72.8
GCN-V	46.4	49.7	56.3	76.6
GCN+V	60.5	57.5	65.9	77.8
Co-training	**62.2**	**68.3**	**72.7**	78.2
Self-training	51.9	58.7	66.8	77.0
Union	58.4	64.0	70.7	**79.2**
Intersection	52.0	59.3	69.4	77.6

Summary

- Contributions
  - Principled understanding of the working mechanisms and limitations of GCNs for SSL
  - Solutions to improve GCNs

- Future directions
  - Designing more powerful convolution filters
  - Techniques for training GCNs