An Anatomy of Graph Neural Networks Going Deep via the Lens of Mutual Information:

Exponential Decay vs. Full Preservation

DS3Lab@
ETH ZURICH

ETH
Systems@ETH Zürich

Summary

> *Graph Convolutional Networks (GCN) cannot benefit from using a deep architecture!*

Li et al. 2018/2019, Oono et al. 2019, identified the problem as oversmoothing.

Can we make GCN as _powerful_ as Graph CNN?

Pillar 1. Empirical Study

- How fundamental is the decomposition strategy in GraphCNN?
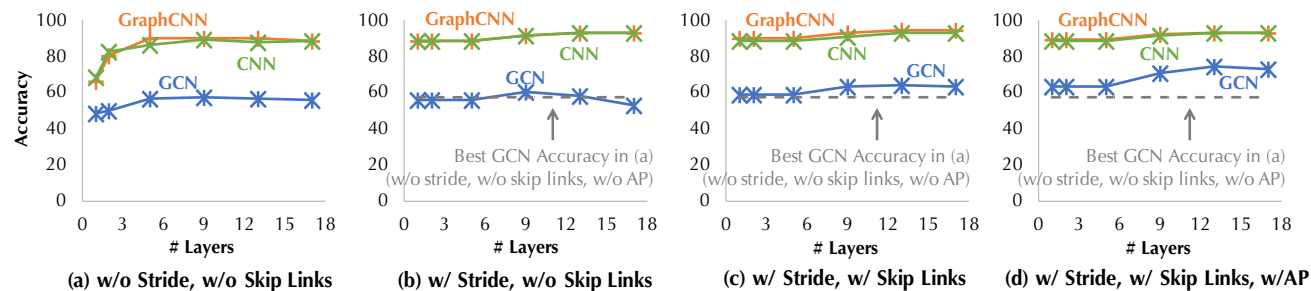- Can GCN match the accuracy of ResNet empirically?

(a) w/o Stride, w/o Skip Links (b) w/ Stride, w/o Skip Links (c) w/ Stride, w/ Skip Links (d) w/ Stride, w/ Skip Links, w/AP

Best GCN Accuracy in (a) (w/o stride, w/o skip links, w/o AP)

Figure: Test Accuracy on CIFAR-10

(a) GCN Layer

$$AXW =$$

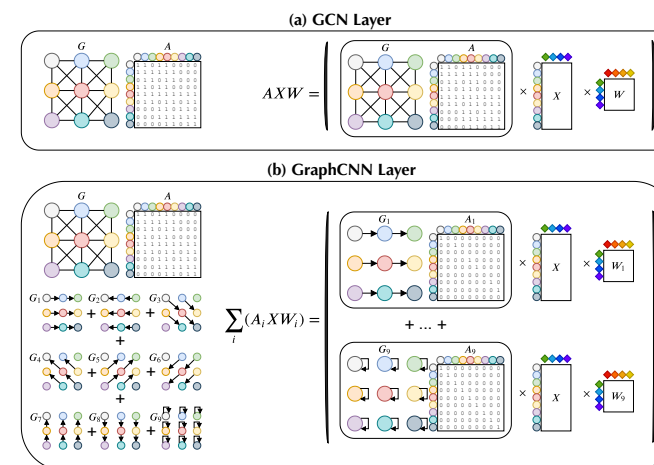(b) GraphCNN Layer

$$\sum_i (A_i X W_i) =$$

Figure: GCN vs. GraphCNN

Pillar 2. Theoretical Study

- What are the benefits introduced by graph decomposition in GraphCNN, compared with GCN?
  - *Mutual information exponentially decays with number of layers in GCN!*
  - *The regime where GCN loses information is much larger compared to that of GraphCNN!*