



SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS

CS 598 Final Project

**Jiahao Zhang,
Sixian Jia**

About the paper

This paper presents a graph-based semi-supervised learning approach for classifying nodes in a graph, where labels are only available for a small subset of nodes.

- Introduce a simple and well-behaved layer-wise propagation rule for neural network models that operate directly on graphs and demonstrate how this form of a graph-based neural network model can be used for fast and scalable semi-supervised classification of nodes in a graph.

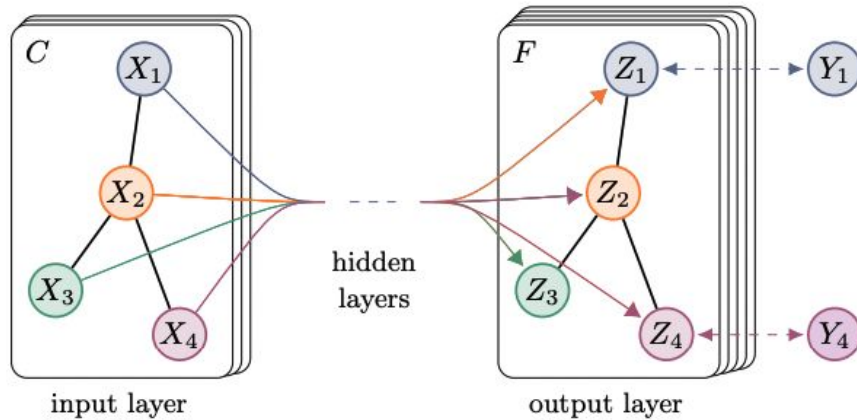
Our Goal

Claim 1: The paper's method is reproducible and has the potential to achieve higher accuracy than previous methods (such as ManiReg, SemiEmb, DeepWalk, ICA, and Planetoid) on the Cora dataset (the dataset used in the paper) when the amount of labeled data is limited.

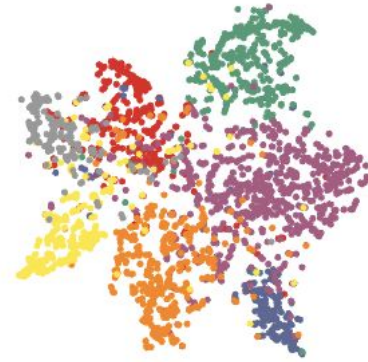
Claim 2: The performance of the GCN model can be influenced by the choice of hyperparameters selected for the specific task.

Claim 3: The model presented in the paper has demonstrated excellent performance on graph-structured datasets not specifically utilized in the paper.

Graph Convolutional Network

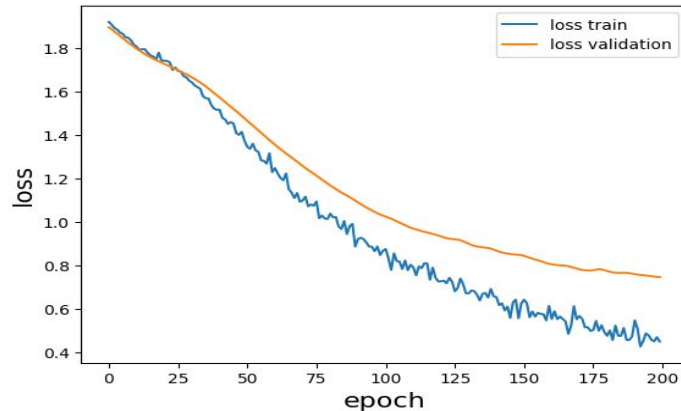
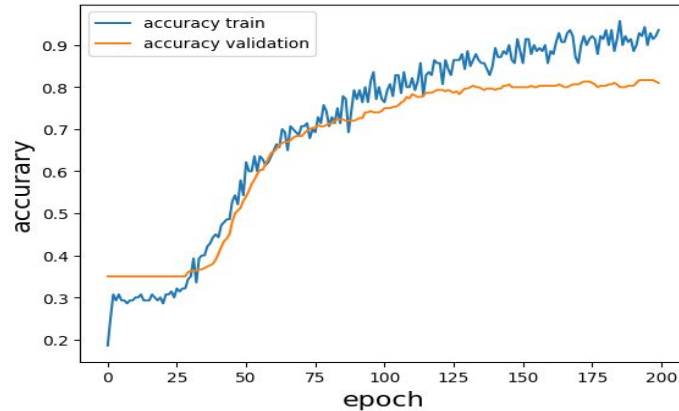


Hidden Layer Activations



Two - Layer GCN:
$$Z = f(X, A) = \text{softmax}\left(\hat{A} \text{ReLU}\left(\hat{A}XW^{(0)}\right)W^{(1)}\right)$$

Reproduction - Model Performance



Cora Dataset Statistics

Type	Citation Network	Nodes	2708
Edges	5429	Classes	7
Features	1433	Label Rate	0.052

Method	Classification Accuracy
ICA	75.1
Deep Walk	67.2
GCN (paper)	81.5
GCN (our results)	83.6

Reproduction - Hyperparameters



There are **five** hyperparameters for the GCN model: learning rate, number of epochs, weight decay, number of hidden units, and dropout rate

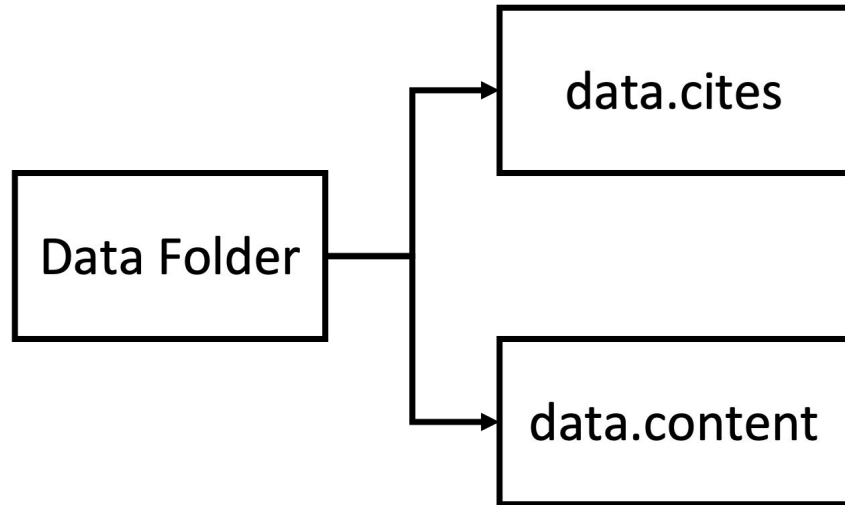
Learning Rate	Classification Accuracy
0.00001	0.155
0.0001	0.138
0.001	0.340
0.01	0.824
0.1	0.838

Dropout Rate	Classification Accuracy
0.1	0.831
0.2	0.835
0.3	0.838
0.4	0.833
0.5	0.824
0.6	0.836
0.7	0.805
0.8	0.798
0.9	0.701

Reproduction - WebKB dataset



The architecture of WebKB dataset



WebKB has **877 webpages** classified into one of the **five categories**: course, faculty, student, project, and staff. The webpages gathered from **four different universities**.

Dataset	Classification Accuracy
Cornell	0.829
Texas	0.834
Washington	0.836
Wisconsin	0.823