Follow-up questions on assignment 3

1. Confusion about the result

In this assignment, we wanted to develop a classifier which can classify nodes in a network into appropriate category. I see there are two goals model aims to achieve, first to minimize the cross-entropy loss of the classification task. Second, create node embeddings that captures not only the available features but also the properties of the given graph. The second goal then helps to achieve better results for the first task.

I observed that when depth of the GraphSage model is increased the classification performance reduces. Although, the embeddings with deeper model are better than the shallow model (atleast visually). Initially when I saw the lower performance of classification for deeper model, I assumed that the embeddings created by the model might not be much better than the input features. But this was not the case, node embeddings get better with deeper model. The figures below show this observation,

Layer wise embeddings:

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Classification performance:

*(the best results are observed with depth = 1, around 75% accuracy on validation set)*

Chart

Description automatically generated

Graphical user interface, chart

Description automatically generated

1. The noisy results on validation dataset:

With the increase in depth of the GraphSage model, I saw more variations in the model’s performance on validation dataset. The following figures show results with depth equal to one and three,

I am not able understand why we might be seeing such differences in the results.

Depth = 1

Chart, line chart

Description automatically generated

Depth = 3

Chart

Description automatically generated

1. Would stochastic approximation method work for updating embeddings?

In the GraphSage algorithm we update the hidden feature vectors (node embeddings) as follows,

In the developed model,

is BERT representation of the title of node ‘u’

is neighborhood of node ‘v’

is neural network for layer ‘k’

Now, we can consider (or can we consider?),

as our estimate of the embeddings that we want to update with the help of some sampled information i.e.

Therefore, can we update the embeddings as follows,

According to Robbins-Monro stochastic approximation,

Where, = 0.9 (say)

I am doing some experiments to compare results but I do not have results yet.