Experiment on the Model

We test our construction model in with the first layer, classifying all nodes into three clusters of Communication, Information Theory and signal theory.

We first construct the node with the given data set, add reverse sides and insert attributes as the origin attribute plus “by”. Then we get a graph with 812 nodes, 1834 sides and 40 side attributes, where 28(3.4% of all) nodes are artificially labeled. Then we test the SCDML model in the data set. With node embedding algorithm, we limit the embedding dimension from the origin 812-dimension One-Hot vector to embedding code as 2-dimension vector for visualization. It loses much information as actually we only reduce the dimension to about 50- dimension for the next step, and here, we mean to show a 2-D results.

The results have been shown in the paper above in Model part. From which we could see that the nodes with deep connection are mostly close to each other. The l-2 norm of the embedding vector could be a great metric for structure similarity.

Then, we first test weighted kernel KNN clustering(here we didn’t introduce any kernel function) without deep metric learning. In this situation, the model is a supervised learning pattern, but clustering standard is mainly based on structure feature. Then we test the same method but with deep metric learning. This is a semi-supervised pattern and the F function makes sure that the clustering standard is based on semantic feature of all nodes learning from the 28 labeled nodes. The results are completely different as we shown in this paper above in Model part. KNN without deep metric learning gets a confusing clusters, and it can’t be generalized with newly input information. But its counterpart demonstrates a great generalization ability. The F function mapping the structure feature to semantic feature and clusters all nodes clearly in three part.

We believe that in real scenes, the initial labels will be more accurate by experts, and the labeled nodes will take a greater proportion and cover a more complete classification. Thus, our model will generate a more accurate knowledge graph system.

Finally, we generally test our GCN model in this graph. But the graph is in a very small scale, our model is inaccurate. As this model is mainly designed to be applied in real scenes with great amount of input data, we can’t show the results here in this small data set.

We believe that in real scenes, with much greater amount of data input and more complex graph structure, our model will shown a greater ability.