UNIVERSITY
OF MANNHEIM
School of Business Informatics
and Mathematics

Data Mining – Python

Exercise 9: Graph Mining

9.1. Node2Vec for Node Classification

1. Load the CiteSeer dataset using the stellargraph dataset class.
2. Generate the Random Walks and train a Node2Vec model using the gensim library.
3. Visualize the learned embeddings using t-SNE visualization.
4. Perform a train/test split and learn a classifier for the nodes using their Node2Vec embeddings as feature input.
5. Evaluate the classifier on the split off test set using the Accuracy metric.

9.2. Graph Neural Networks for Node Classification

1. Further split the previous train set in train and validation and create a GCN model using Keras.
2. Train the model and evaluate on the test set. Can the GCN improve over the Node2Vec approach?
3. Visualize the embedded Nodes by extracting the embeddings of the last Convolution Layer and visualizing them using t-SNE.