So I first scraped all the mathematics articles from Wikipedia and appended all the wiki links and created a dataset of it. I also extracted the NLP features, keywords, and tags of all the link Ibpages. Then I imported the wiki math article CSV sheet given by Anirban sir and created a graph of it just for testing. After this, I cleaned the data of the output.csv file provided by Anirban Sir. That file contained multiple links with different labels so I updated all those links with the most frequent label and created another work file from it. I took the first 10 links as the root nodes and scraped them again to find the children nodes and added edges betIen them. Then I created the graph by visualizing it using networkx and finally extracted additional features by using centrality metrics- degree centrality and betIenness centrality and clustering coefficients. Finally, I calculated the node embeddings of the graph using the node2vec method and printed the node embedding vectors, and saved the ‘node\_embeddings.emb’ and ‘node\_embeddings.model’ file.

The codes for Part4 and part 5 have been explained as comments in Part4\_and\_Part5.ipynb.