MSc Project - Reflective Essay

Project Title:	**Investigating Ecological Networks based on Structural Identity**
Student Name:	**Paras Vasant Mistry**
Student Number:	**210877245**
Supervisor Name:	**Athen Ma**
Programme of Study:	**Artificial Intelligence**

1 Analysis of strengths/weaknesses

Strengths

- **Get to know the conversion of the ".csv" file into graphical data using networkx**
- **Learned different graph embeddings algorithms along with their usability on graphs**
- **Word2vec model in comparison with node2vec and deep walk model**
- **Using seaborn on the dataset to visualise it using t-SNE was an achieving task**
- **Getting the graph embeddings using different models and methods and make it work for further evaluation**

Weakness

- **The dataset was not properly organized according to the data type**
- **The node indexing was giving an error which was annoying at some point even when the data was cleaned and reused**
- **A larger file size can lead to kernel restart**
- **Data type conversion was not on point**
- **Increasing the number of random walks can lead to the time-consuming task**

2 Presentation of possibilities for further work

We can work on the model further to implement the dataset more accurately and can also use different machine learning algorithms to visualize the data in a more proper manner rather than just getting inaccurate outputs

- **We can split the data into proper sets and perform pre-processing**
- **Getting the network graph in a more resolution state and comparing it with different graphs**
- **Using a higher GPU can make the model run faster**

3 Critical analysis of the relationship between theory and practical work produced

When we start working on the theory side of the project, we get to know the real struggle of finding the data and connecting each and everything in place to make the project work and provide the desired outcome. Learning new packages and techniques from scratch is very time-consuming because you don't know where

you go wrong. We get to learn different methods and ideas to implement the same problem in a more better and accurate manner. Going through multiple research papers of different authors and closely understanding the goal behind the execution took a lot of time but the final outcome gives an overall satisfaction. However, on the practical side, I got an error while installing packages because of the version compatibility issues depending on different environments. Some packages were working properly on colab whereas the same was raising an error on the Jupyter notebook. Finding the way to get node embeddings was the main task for me. I was not able to get the embeddings for the graph because of uneven data and different data type conversion errors. In the end, I figure out the problem and manage to get the embeddings for the graph then again new error of int to float was raised which made the process slow. Overall get to know in-depth about graphs and how to visualise them using different graph embeddings models by using different algorithms.