Shahar Berenson 208608414  
Shlomi Fridman 318187002  
Omer Goldstein 205906258

Graph-Embedding Summary Report

Process of the project study:

We received the project subject “Graph-Embedding” with a link to a GitHub Repository related to it. There were four scripts in the repository and all of them didn’t work. Two that were beyond saving. And two that needed updates to the deprecated packages and commands.  
There were some challenges understanding the code as there was little documentation. All of it in Chinese.

After we were able to compile two of the scripts. We began to research the project by reading papers about Graph-Embedding, the motivation and the known algorithms, Especially DeepWalk and Node2Vec which were used in the scripts that worked.

Afterward, We made a new Google Colab notebook, loaded the data from GitHub project and compiled the model in the cloud.

Then we analyzed the dataset files, and the flow of the scripts. We summarized the script process and printed the obtained results. Along with statistical data (such as F1 scores, AUC, etc..) and a confusion matrix.

Project flow:

We met once a week to make progress on the presentation and share research insights between us. During the week, we discussed the subject, the presentation and the general schema of the project.

We met with the lecturer (Renata) to discuss the direction of the project, and get clarifications about the presentation.

Obtained results:

The average F1 score moves in range of 65-70% across multiple runs of the scripts. And the average AUC score is in range of 90-95%.

We noticed that the Node2Vec script was slightly more accurate than the DeepWalk script, average improvement of 1% to 2%.

Conclusion:

Compared to other algorithms and implementations, during our research phase, we concluded that the scripts given were not giving very accurate results. The cause of which might be derived from the implementation and the dataset used by the author of the scripts.

We also conclude that the Node2Vec is slightly more accurate than DeepWalk as implemented in the given scripts, using the Cora dataset.   
Based on multiple runs of the scripts, and the difference between the two scripts, that Node2Vec random walks is more likely to be going-backwards during its random walk, due to Node2Vec p=5 as opposed to DeepWalk that has p=q=1.

External links:

[Github Repository](https://github.com/shachar700/Graph-Embedding/)

[Google Colab script](https://colab.research.google.com/drive/11eVIbhSCO4OnrZEPRfTYLs7bEZRxLuLQ?usp=sharing)