1. Introduction

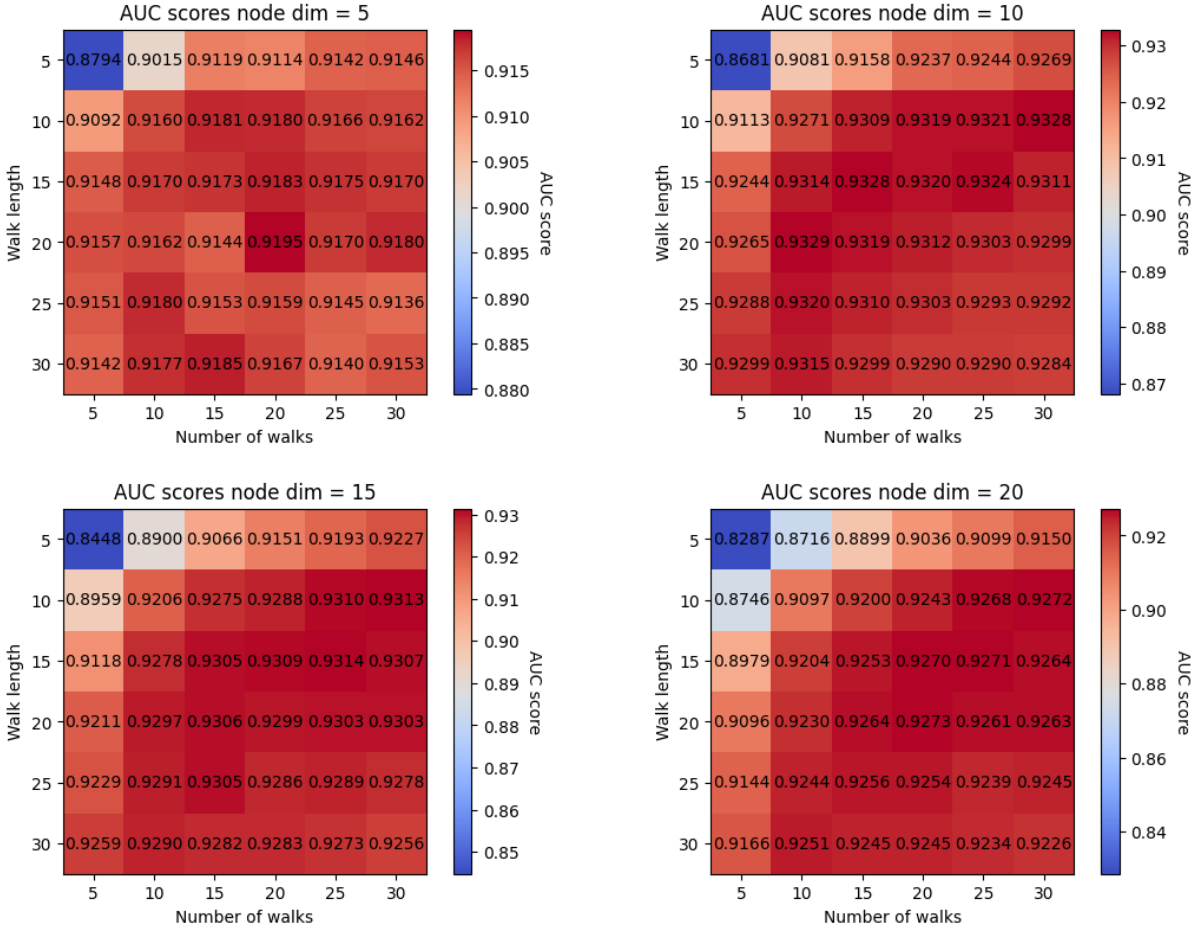
DeepWalk and Node2Vec were the embedding algorithm we used. After several trial and error, we got the best hyperparameter values and assessed the performance of the model by using the validation set. The best AUC score of 0.9329 were obtained with DeepWalk model.

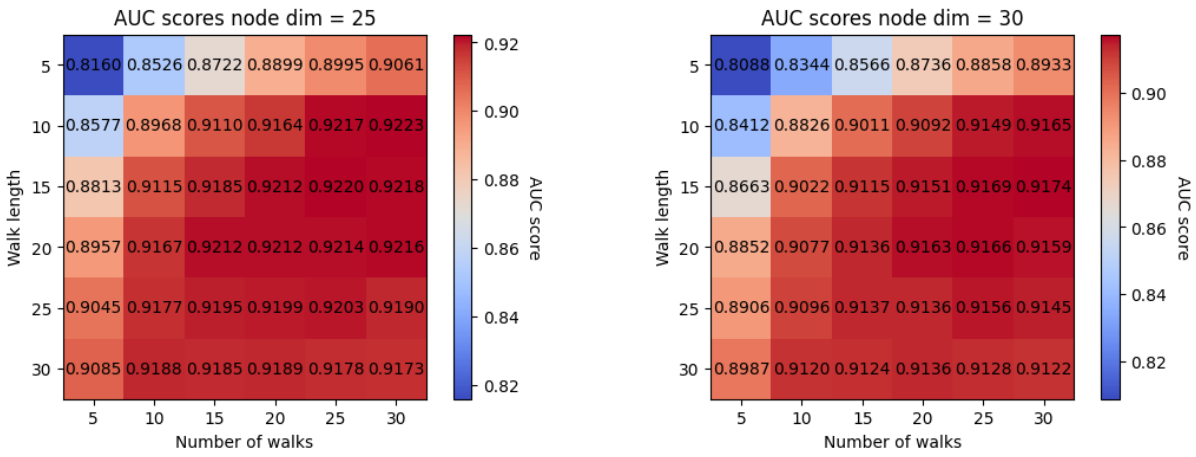
2. Hyperparameters

We utilized ‘for loops’ to test parameter pairs in order to conduct hyperparameter tuning for both DeepWalk and Node2Vec. For DeepWalk model, the hyperparameters tested included the number of walks, walk length, and node dimension, which were set to values ranging from 5 to 30 with increment of 5. For Node2Vec model, we first tested p and q values ragning from 0.25 to 0.75 with increment of 0.25. The number of walks tested were 5 ,10 ,20, 40 and the walk length tested were 10, 20, 40. Then, we found that when the number of walks is 10 and the walk length is 20, the average AUC score was the best. Thus, we then tested the p and q values ranging from 0.05 to 0.95 with increment of 0.1. For the Node2Vec model, the best AUC score is 0.9317 which is worse than the best AUC score of DeepWalk model which is 0.9329. We found that the higher num_walks and walk_length is resulted in the longer training time. After conducting the hyperparameter tuning, we observed that optimal DeepWalk model had higher AUC-ROC scores than the optimal Node2Vec.

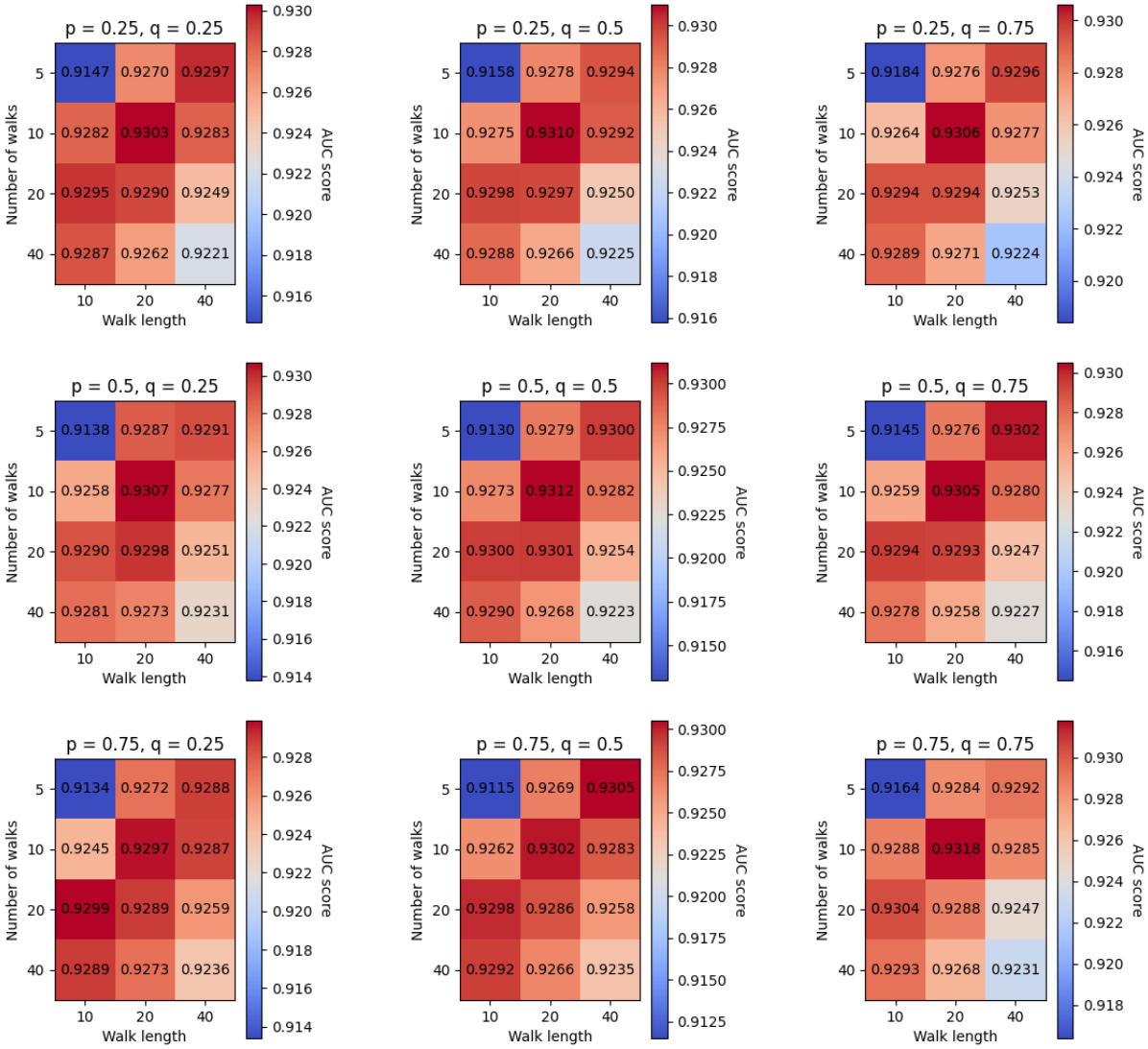
3. Data visualization

We visualized the AUC score on different parameter combinations using heatmaps for the DeepWalk model and Node2Vec model. For DeepWalk model, we generated different heatmaps based on node dim and plot walk length against number of walks.

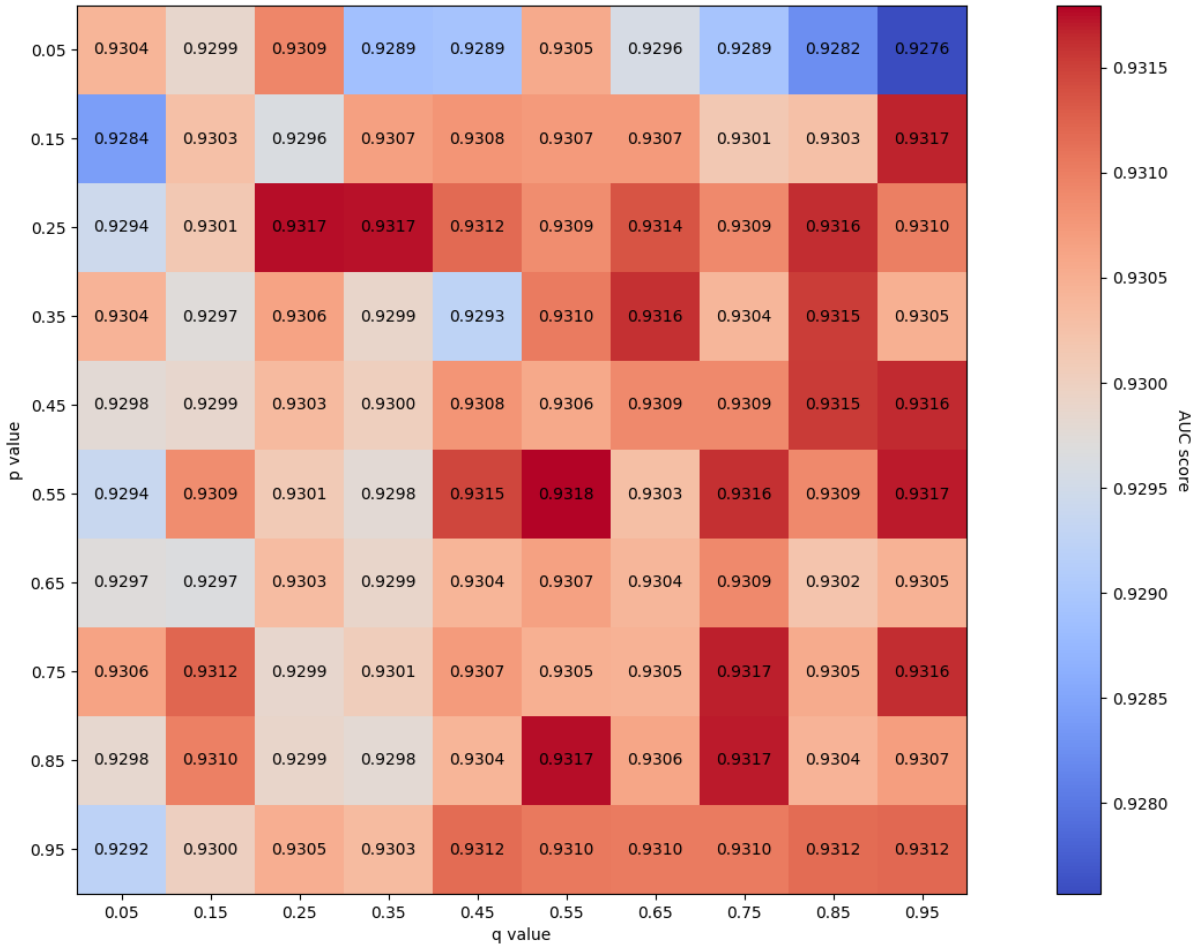




For Node2Vec model, we first generated different heatmaps based on p and q values and plot walk length against number of walks.



Then, we found that the AUC score was good in average at node_dim=10, num_walks=10, and walk_length=20. WIth these values, we tested and generated heatmap which has q as x-axis and p as y-axis.



In conclusion, after conducting hyperparameter tuning on the training dataset and comparing the AUC scores generated by different hyperparameter combinations, we determined that the optimal hyperparameter values for DeepWalk model were node_dim=10, num_walks=10, and walk_length=20 with the AUC score of 0.9329 and the the optimal hyperparameter values for Node2Vec model were node_dim=10, num_walks=10, walk_length=20, p=0.55, and q=0.55 with the AUC score of 0.9318. The AUC score obtained from DeepWalk model with optimal hyperparameter was higher than the AUC score obtained from Node2Vec model with optimal hyperparameter.