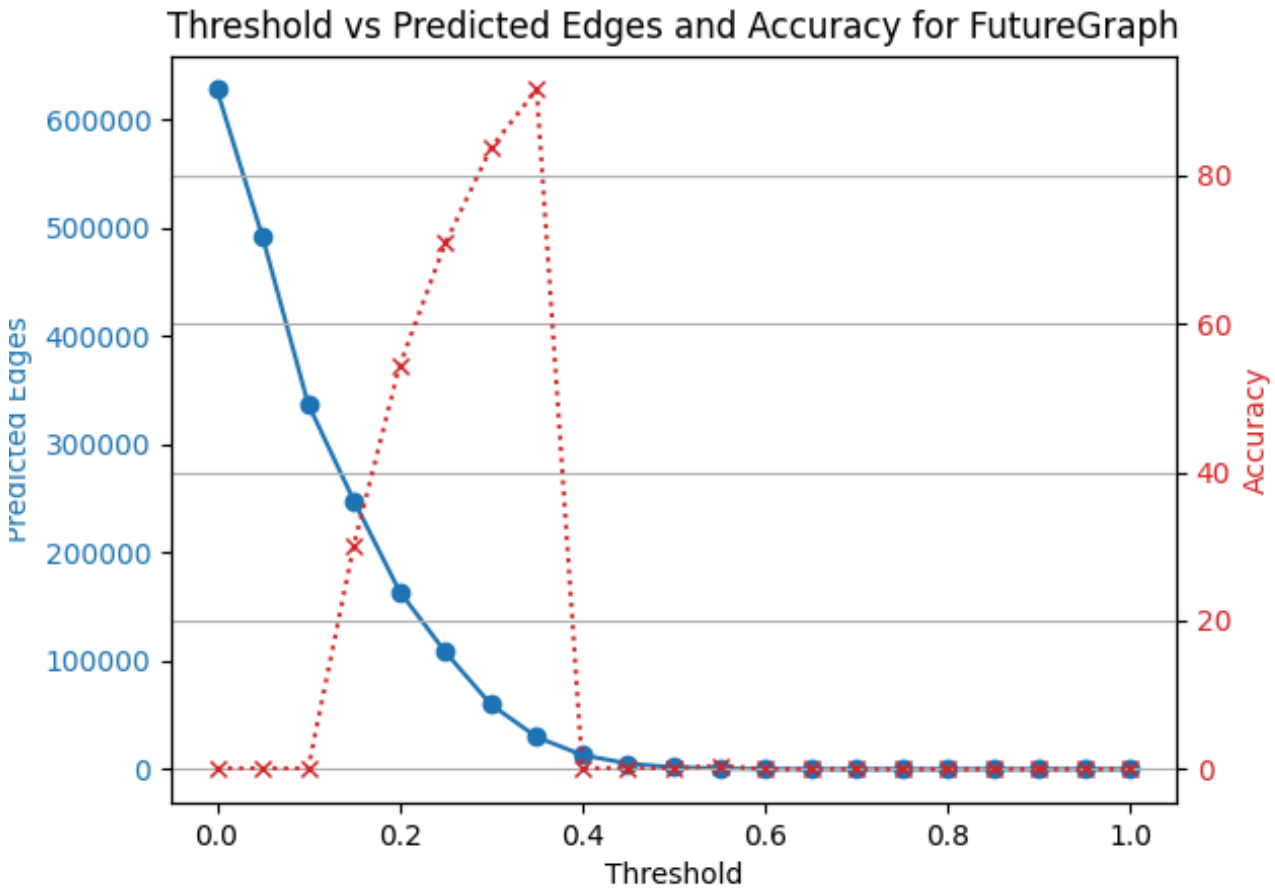
Bonus Task:

Link Prediction is a task in graph and network analysis where the goal is to predict missing or future connections between nodes in a network. As before, multiple algorithms exist for this task. However, you are required to implement both a graph neural network, and a classic algorithm like DeepWalk or Node2Vec. For training, you can use the citation network at any time interval [0-T] and for testing/validation, use nodes that appear after time T. ,. Compare the results of the two approaches, and analyze whether the GNN performs better, and if so, why. You will be evaluated on how well the model can predict these edges, as well as your understanding of the link prediction task in graphs

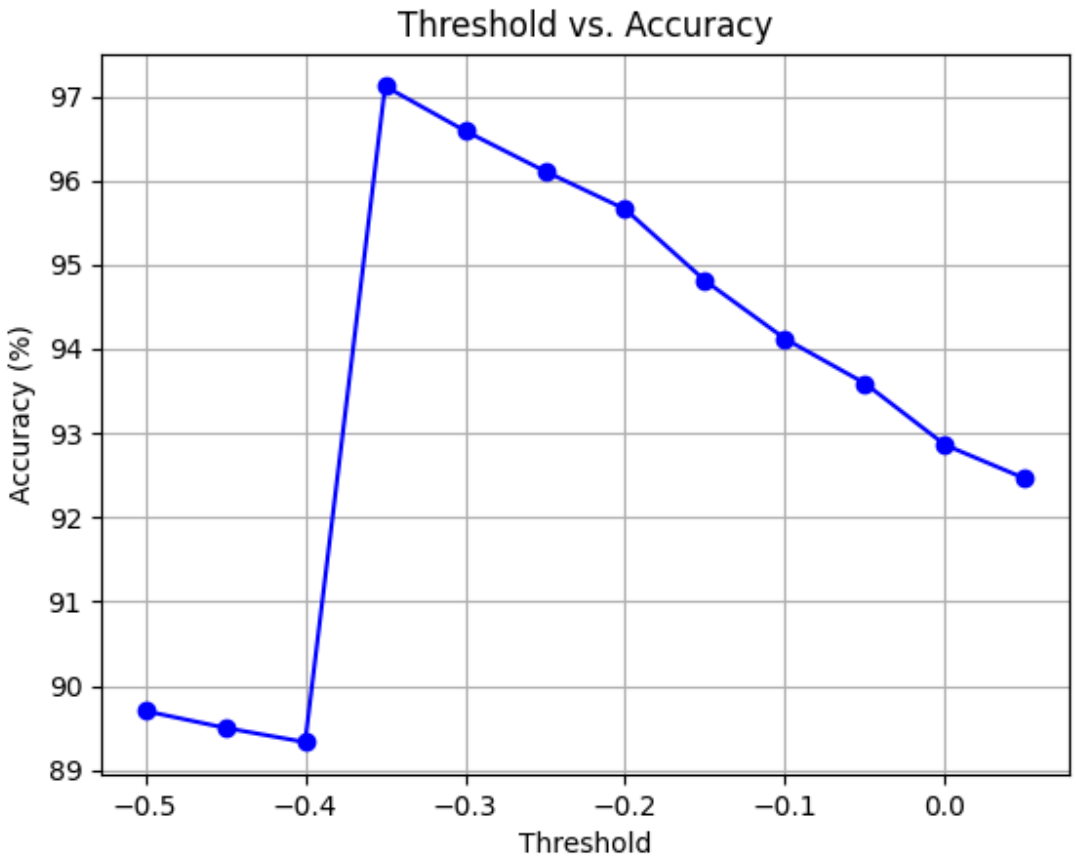
For link prediction graph at time 1993-12-01 was taken and the two models NODE2VEC and GNN were trained and testing was done on the graph at time 1994-03-01

A threshold was set for Node2Vec , and accuracy was compared for different values of threshold ….



This graph accuracy decreases with increase in no of nodes and suddenly drops to 0 when predicted nodes goes to minimal constant set.
Therefore actual accuracy is
87.4 %

The GNN model:
Similar graph- linkage was predicted using neural networks and the plot was plotted



The GNN provides the predicted values from -1 to 1 and therefore such values of threshold. Also note with increase in threshold no of predicted edges decreases , and no decreases to such minimal value that after 0.05 , accuracy either shoots up or shoots down randomly every time you run the script.

The Accuracy being around 97 %

Clearly the accuracy of GNN is better.

Possible Reasons:

- Node2Vec is effective for capturing local neighborhood structures and can generate embeddings that preserve the structural similarity between nodes. Though, Node2Vec may not explicitly consider global graph structure, potentially missing important higher-level relationships.
- GNNs are designed to capture both local and global graph structures, making them potentially more powerful for capturing complex relationships.
  They can adapt to different types of graph-structured data and learn hierarchical representations.
- Since the citation network has a global complex structure and the new edge don't just depend on the local structure but the global topography of the graph GNN performs better.
- GNNs are more expressive in terms of modeling intricate relationships within the graph, potentially making them more suitable for tasks requiring a deeper understanding of the overall structure.