UE19CS345 – Network Analysis and Mining
Course Project

Project Title : Predicting drug-drug interaction using Graph Neural Network

PES1UG19CS012 : Abhijit Sethi
PES1UG19CS015 : Abhijnya Bhat
PES1UG19CS427 : Sammith S Bharadwaj
PES1UG19CS444 : Shamanth K M

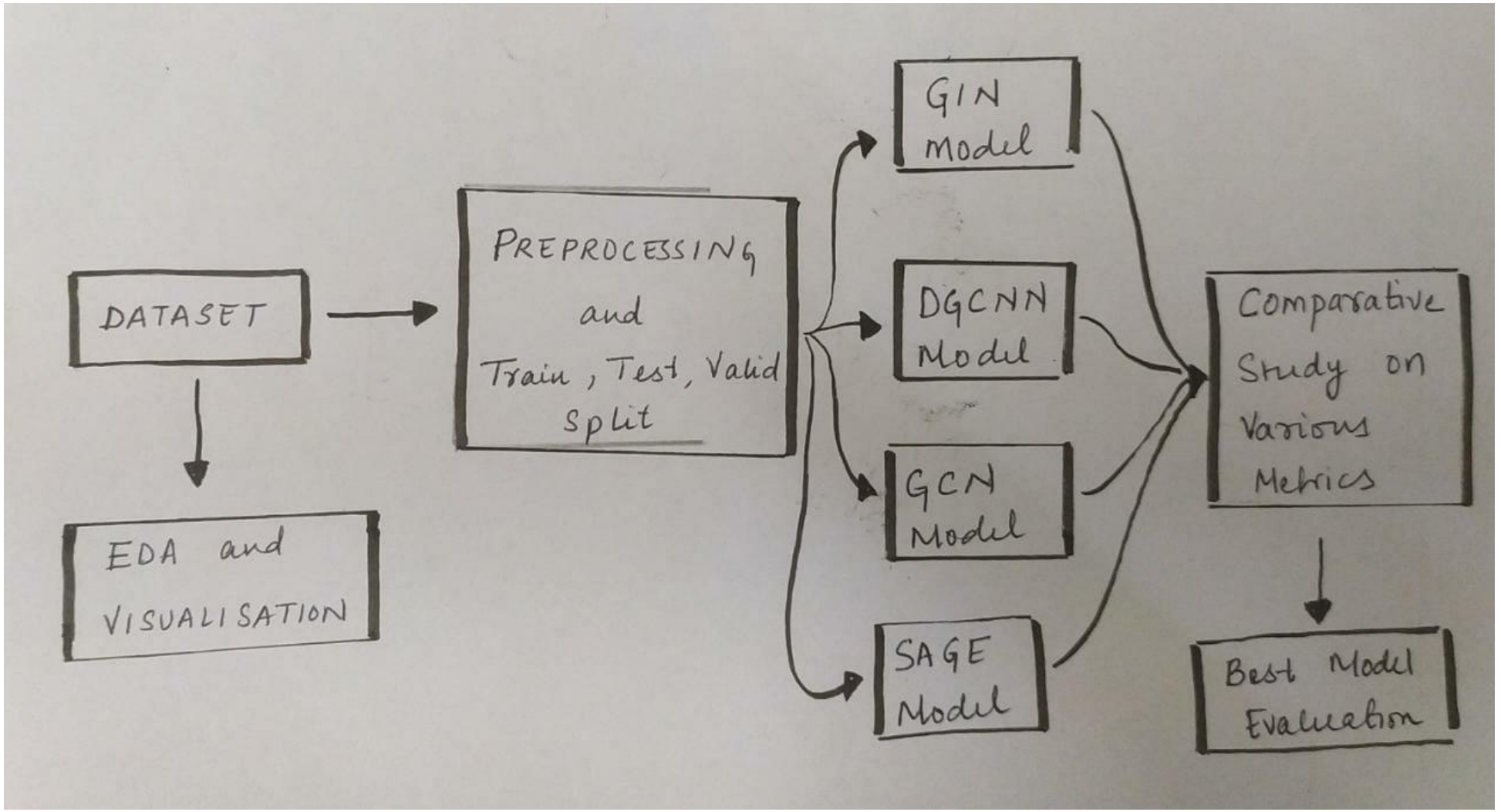
Topic and its uniqueness

- Drug-Drug interaction is one of the most sought out areas due to its immense relevance.
- It is defined by the reaction between two or more drugs when in contact with one another. These interactions can either be supportive, or might lead to hazardous consequences.
- The use of graph neural networks instead of traditional manual techniques not only makes the process more efficient, but also helps in enhancing the accuracy by including connections that would not have been seen by manual approaches.
- We aim to build graph based link predictions and generate a comparative study of the same.

Dataset

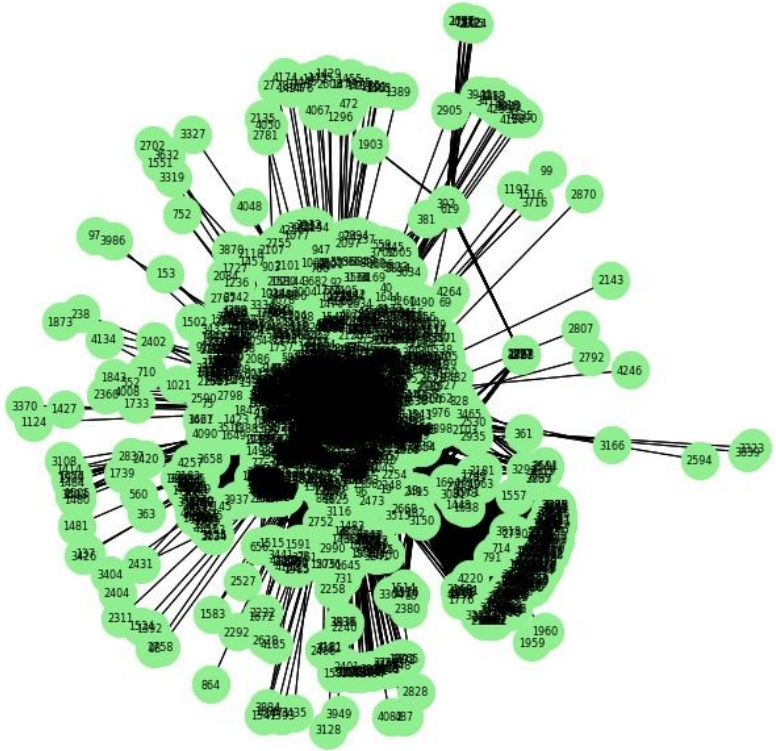
- Ogbl-ddi from Open Graph Benchmark
  Nodes - 4267
  Edges - 13,34,889
- It is a homogenous, undirected, unweighted graph representing drug-drug interactions.
- Source- link

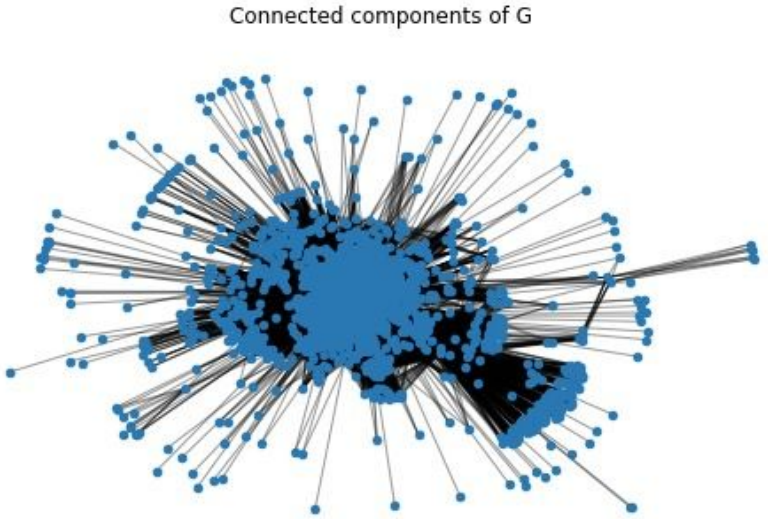
Overall design or approach in a free hand diagram

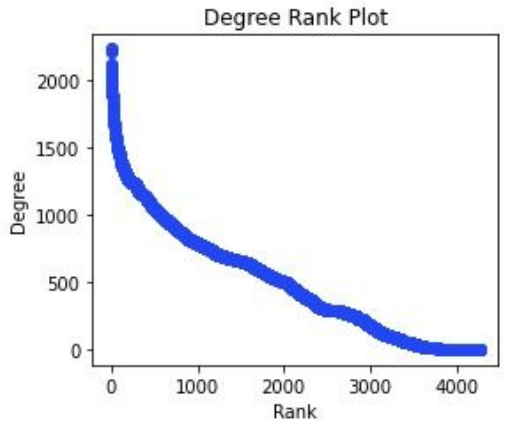


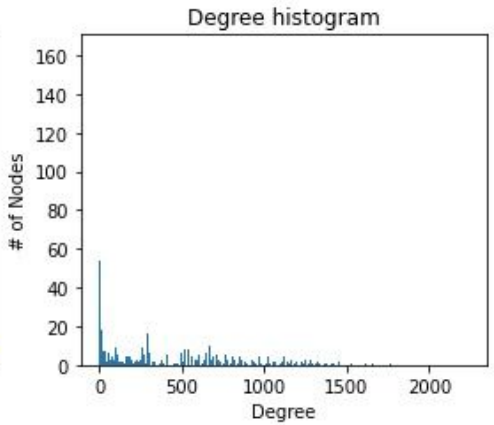
Visualization:

Final results









Final results

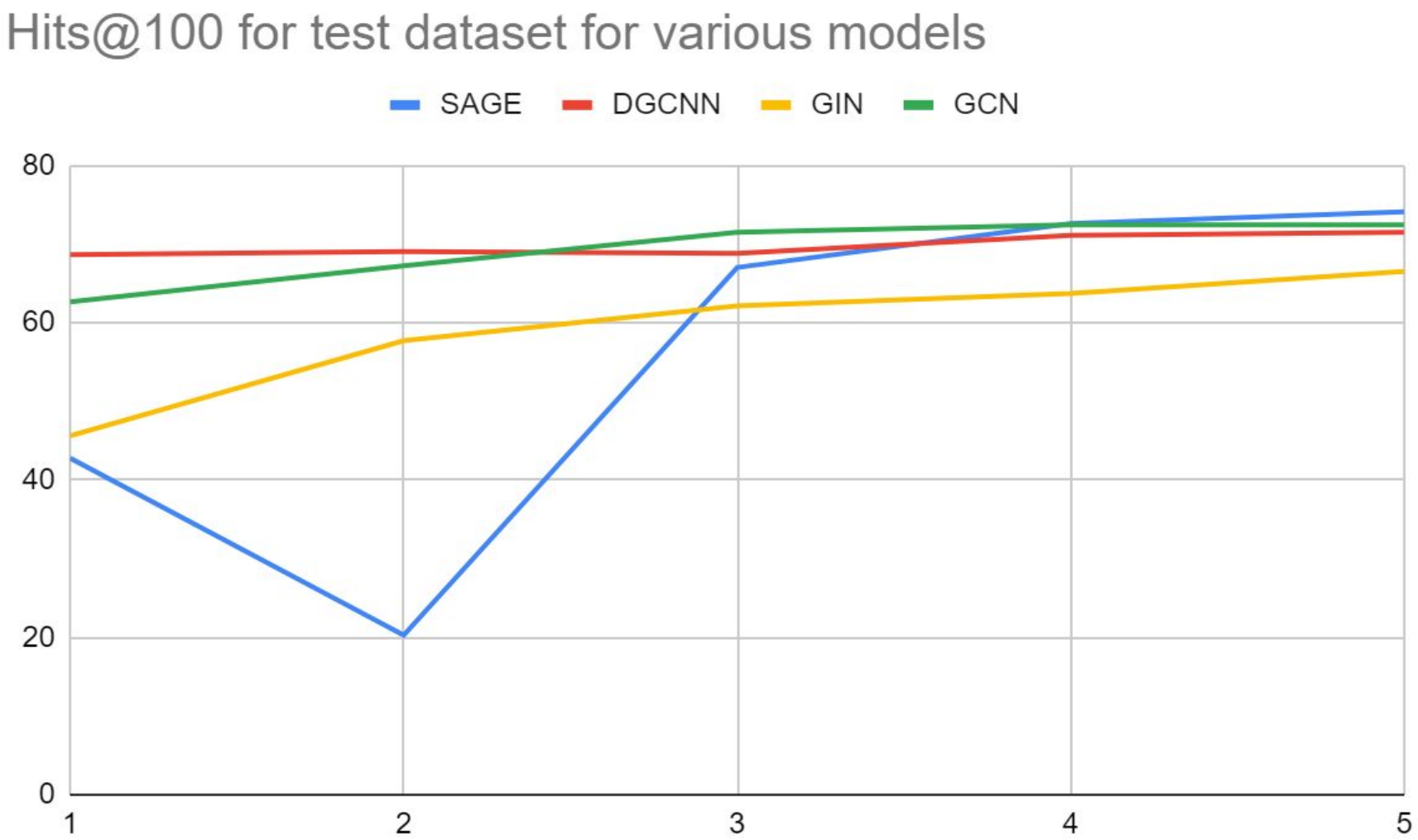
Metric :

- Loss - BCEWithLogitsLoss
- Accuracy
- Hits : On validation set

Models	Hits@20	Hits@50	Hits@100
DGCNN	53.93	63.34	71.52
SAGE	58.23	73.18	73.18
GIN	47.89	58.54	66.54
GCN	0.4819	67.47	67.47

Visualization:

Final results



Visualization:

Final results

- Based on the test metrics, we see that SAGE performs the best on drug-drug interaction, closely followed by DGCNN

What are the remaining portions in this project ?

Currently, we have not used any node embeddings initially. Exploring Deepwalk and Node2Vec for node embeddings and comparing results is yet to be done.

Quantity and quality of work

No	Code functionality	% Complete	Runs without problem (Y/N)	If there are minor issues, indicate
1	Load Dataset - Pyg	100	Y	
2	DGCNN Model	100	Y	Google Colab resource limit disallowed running on entire train, test or val split. Epochs limited to 1 for submission purpose due to same reason.
3	GCN Model	100	Y	
4	SAGE Model	100	Y	
5	GIN Model	100	Y	
6	Evaluation - Hits, MRR, AUC	100	Y	Demo purpose, only HITS shown.

Top few learning

Serial No	Top learning in this project
1	Given the different techniques and implementations of Link Prediction - GraphSAGE, GCN , DGCNN, GIN – Putting them all together into one framework so that we don’t have to re-code train, test and evaluation based on different formats of implementation.
2	Fascinating to model and learn the inner workings of drug-drug interactions via Graph Neural Networks.
3	Key choice of evaluation metrics, overcoming hurdles of compute resource shortage.

Top unresolved challenges

Serial No	Brief description of unresolved challenges	Type of challenge (scope/data/design/implementation / others)
1	Running model on entire dataset with multiple epochs.	Lack of GPU

Reference papers, if any

No	Paper Title	Authors
1	Link Prediction Based on Graph Neural Network	Muhan Zhang, Yixin Chen
2	DGCNN: Disordered Graph Convolutional Neural Network Based on the Gaussian Mixture Model	Bo Wu, Yang Liu, Bo Lang, Lei Huang
3	Semi-Supervised Classification with Graph Convolutional Networks	Semi-Supervised Classification with Graph Convolutional Networks