

Modeling Data with Explainable Graph Neural Network

Shubham Ganesh Modi (A20492276)
Oleksandr Shashkov (A20229995)

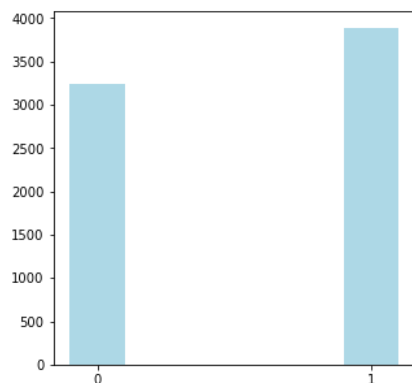
Problem Statement

- The field on Artificial Intelligence has witnessed exponential advancements since the introduction of Deep Neural Networks.
- The working of Deep Neural Models and their inner structure however lacks transparency and creates a black box which provides little understanding of how the predictions are done and which part of the information was critical for the decision
- This drawback has led to research on many explainable Machine learning models and one such branch is Graph Neural Network (GNN).
- In this project we reproduce a published GNN explanation model which allows us to understand and visualize the logic behind predictions.

Project Datasets

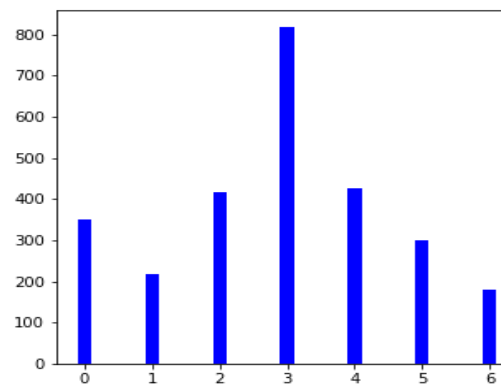
- Twitch Dataset Explanation

- Twitch dataset is a graph dataset consisting of 7126 nodes, where each node consists of 128 node features and the dataset has 2 classes.
- Each node represents a Twitch Streamer and edges are followership between them.
- The node features represent embeddings of games played by the Twitch users.
- The task is to predict whether a user stream make use of explicit language in their content.



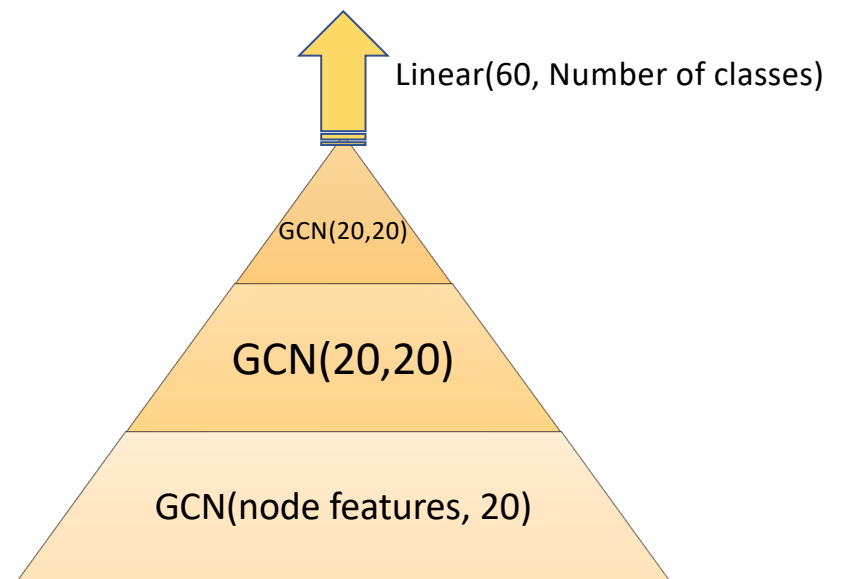
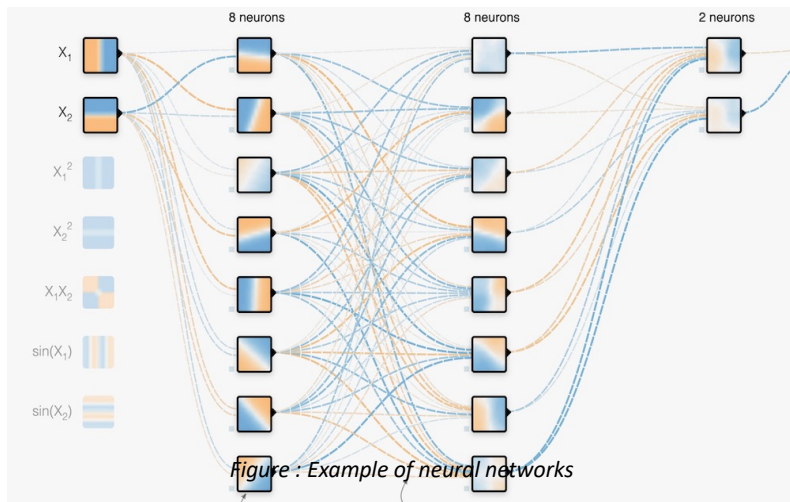
- Cora Dataset Explanation

- Cora dataset is a graph dataset consisting of 2,708 nodes, where each node consists of 1433 node features and the dataset has 7 classes.
- Each node represents a document and edges represent citation links.
- The node features represent embeddings of words in a document.
- The task is to classify the documents in one of the 7 available categories such as Neural Networks, Reinforcement Learning, Probabilistic Methods, Theory, Rule Learning, Genetic Algorithms, Case Based.



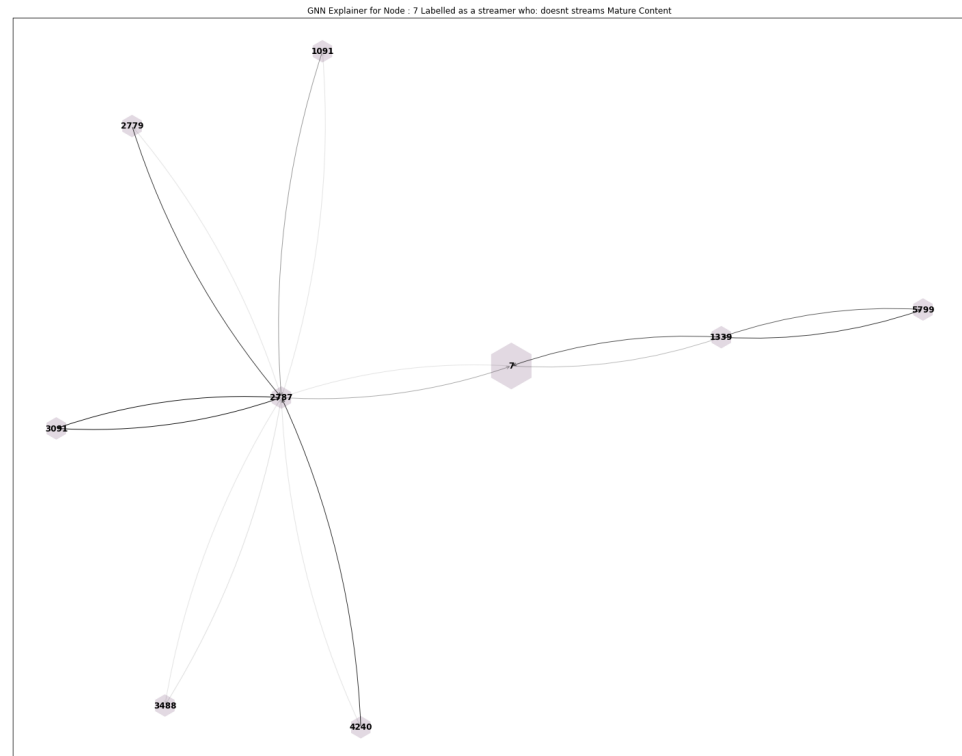
Graph Neural Network Model

- The GNN model makes use of 3 GCN Convolution layers and a Linear layer.
- The model uses ReLU activation function.
- The model also makes use of L2 normalization for edge weights.



Graph Neural Network Explainer

- After training the GNN model and initializing the GNN explainer using the best model we can select a node and explain it.
- The model aggregates information obtained from the target node's neighbors and understands which nodes contain important information which is passed on to the target node.
- This piece of information is important in determining the explanations for the target node to be classified with its correct label.



GNN Explanation for Node in Cora dataset

- Edges

- The subgraph is an explanation for why publication with index 1000 was classified as topic 3
- The edges highlighted in bold are the most influential connections explaining the prediction

