USC Viterbi School of Engineering

Stock Market Community Detection

Discover lucrative stock market trends by analyzing social media sentiment and pinpointing profitable communities through advanced Graph Neural Networks (GNNs).

Anneketh Vij
Eric Liu
Haoxuan Lu
Rishabh Ghosh
Sheffali Rawat

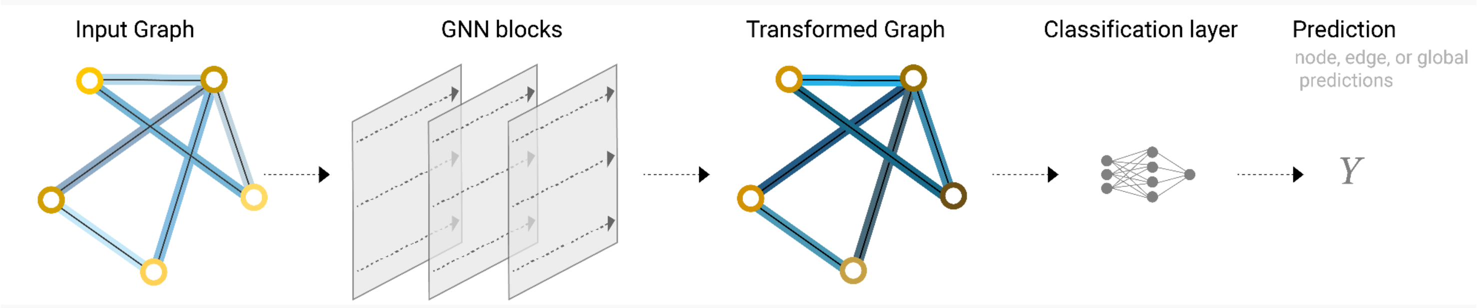
Introduction & Background

- This project's goal is to leverage SOTA techniques in sentiment analysis and graph analytics to enhance the predictive capabilities related to stock market trends, ultimately providing valuable insights for decision-makers in the financial domain.
- While previous research focuses on financial sentiment analysis, with models like FinBERT, often fell short in quantifying the overall net profit or loss. The emergence of graph neural networks (GNN's) offers potential solutions by combining sentiment analysis with structural data analysis.
- The key challenge lies in improving the accuracy of predicting financial information from social media. Traditional models, though successful in sentiment analysis, often struggle to capture the intricate relationships and contextual dependencies.

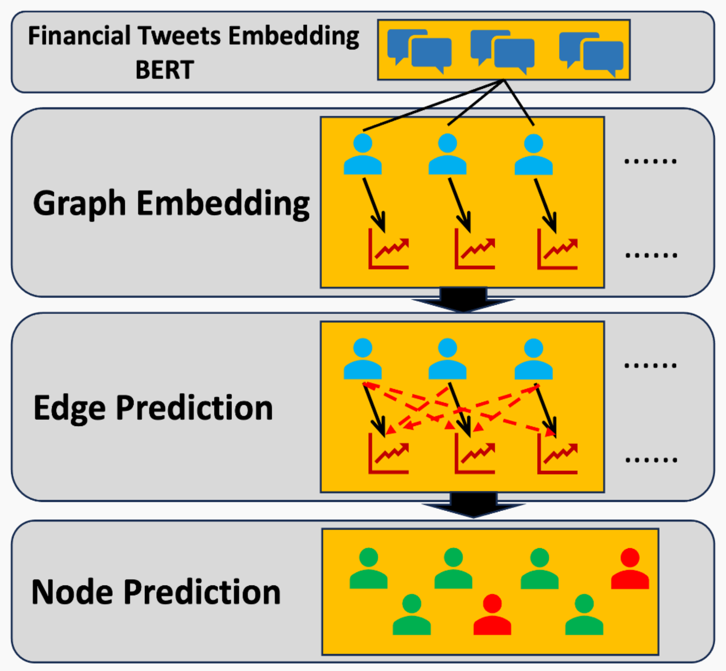
Basic Overview

- By combining sentiment analysis with graph-based contextual understanding, outcomes are more adaptable and informed decision-making tool for navigating dynamic market conditions.
- The study employs financial market-related tweets for analysis. These datasets offer valuable sentiment and contextual information. Data preprocessing, including tweet cleaning and structuring, is essential for effective application of BERT and the GCN model.

(Below figure shows simple working of GNN)



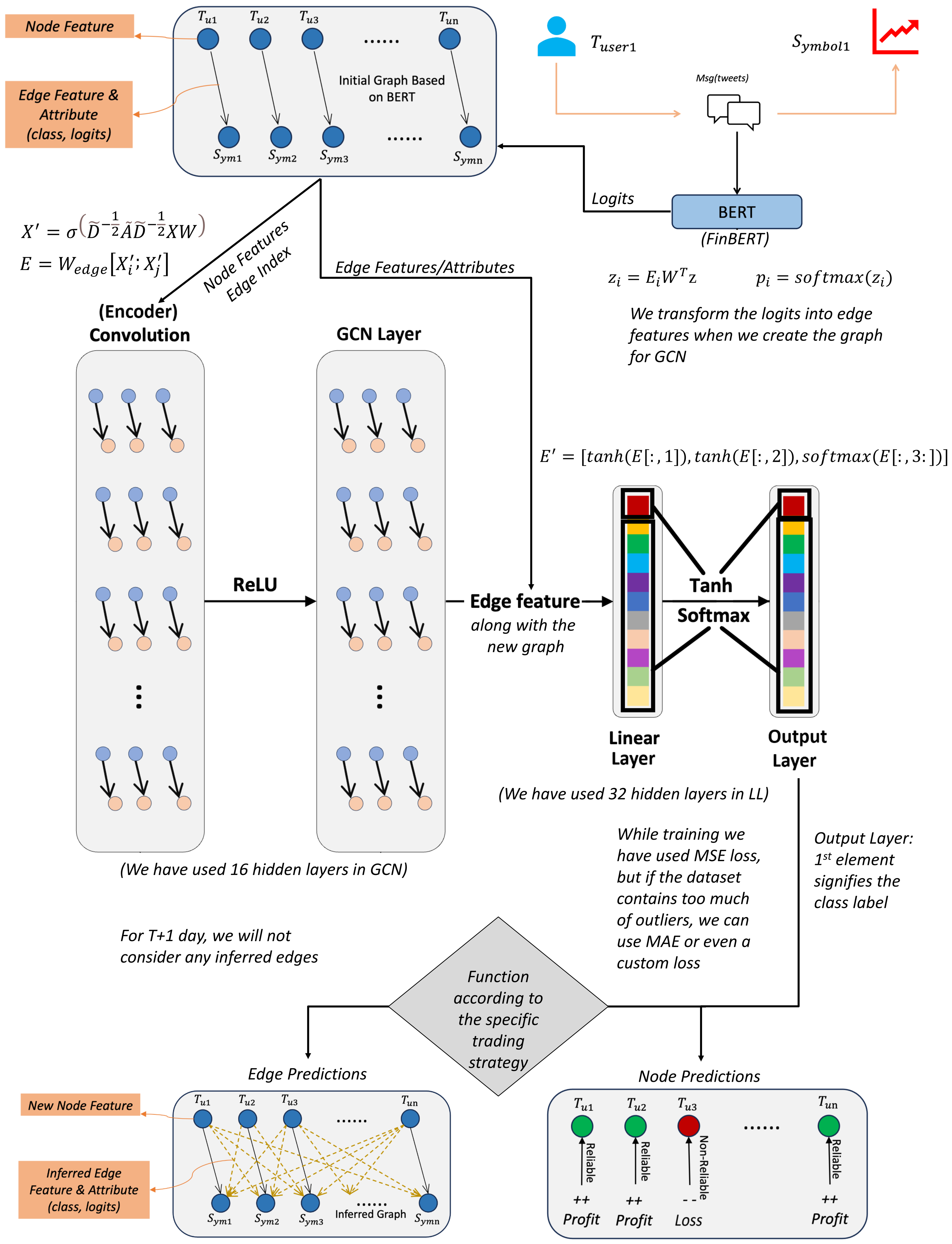
- We use the FinBERT transformer to obtain embeddings for financial tweets and then transform tweet embeddings into graph embeddings for graph-based analysis. These are added as edge features in the graph.



- We then apply a Graph Convolutional Network (GCN) for node and edge predictions. GCN acts as an encoder for node features which captures the structural relationships, which is then passed to a linear layer along with the edge features.

- We then apply a Graph Convolutional Network (GCN) for node and edge predictions. GCN acts as an encoder for nodes, capturing structural relationships, and a linear layer decodes edge features.

Model Architecture

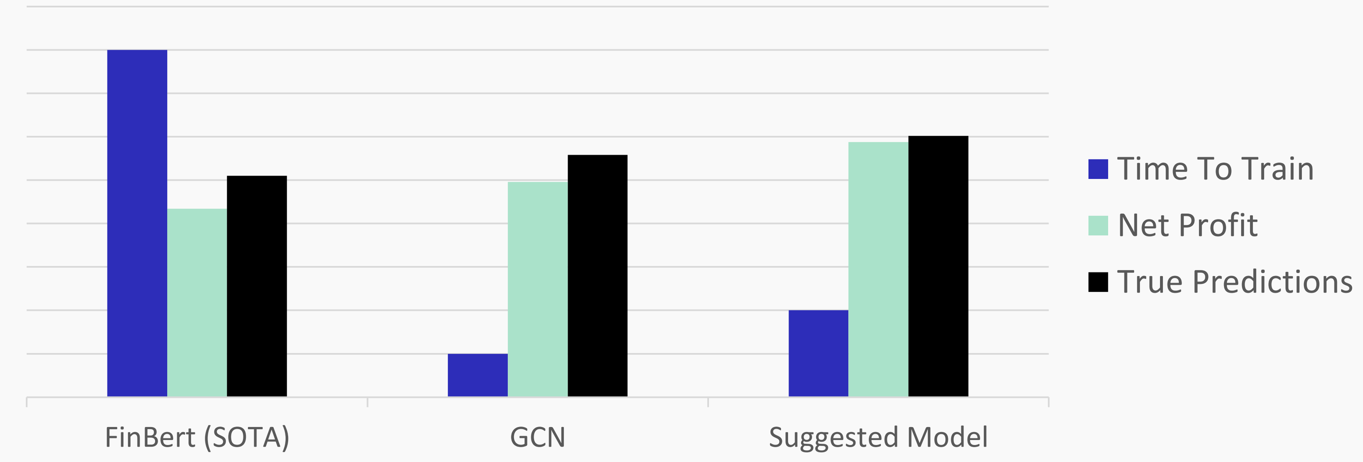


~2 Million new inferred edges were predicted within a minute. We had only ~20k initial edges. Now during trading, traders can cross verify new message.

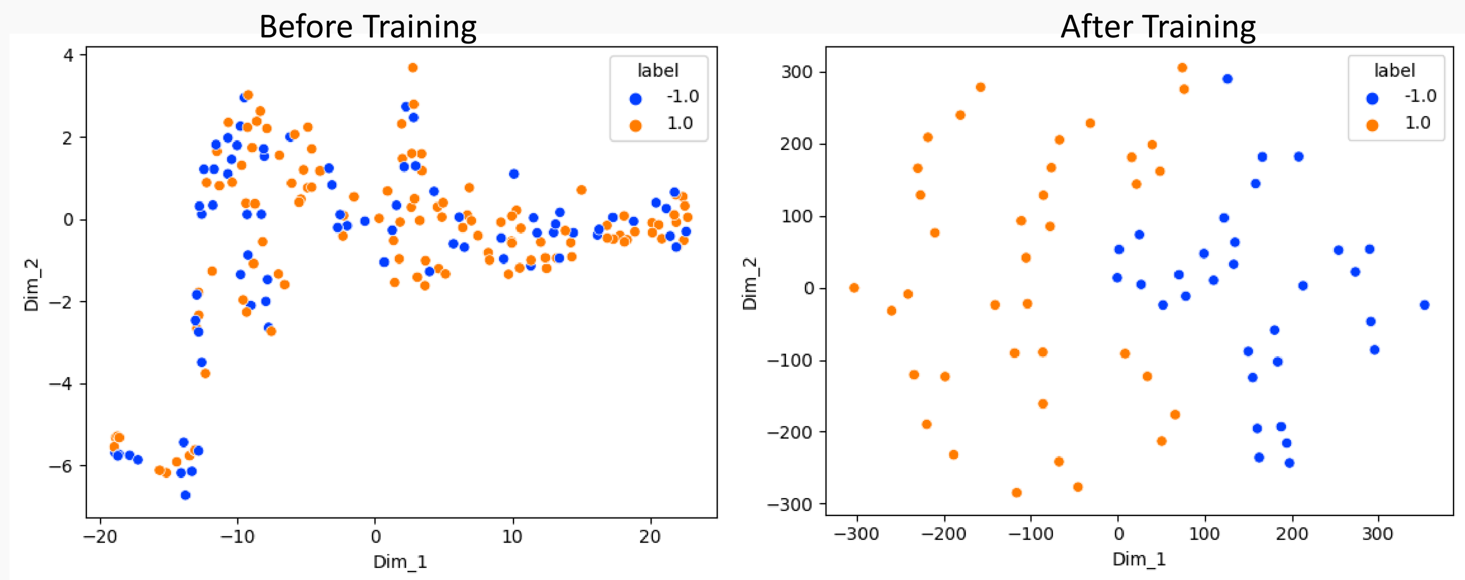
~5k initial nodes were classified into reliable and non-reliable nodes. Reliable nodes CAN have more cumulative more profit or more positive predictions.

Performance & Visualization

Performance of various models during trading hours



t-SNE Visualization



Net Profit Performance of the Model in Various Strategies

Total Positive Edge	True EOD Outcome	Predicted Reliable Nodes	Predicted Profitable edges	Total Profit %
High	-	3173	3882	0.85
-	High	3095	3743	0.86
High	Normal	3022	3738	0.88
Normal	High	3075	3535	0.89
Normal	**Normal**	**1079**	**2143**	**0.94**

Note: Baseline Model (FinBERT) has a Total Profit of 0.87 %

Conclusion

- Combining FinBERT with a GCN model offers a promising avenue for analyzing financial information. Subsequent research could investigate the scalability of the model, its applicability to diverse financial data, and improvements in handling real-time data streams for more timely predictions.
- Furthermore, a BERT will always lack the capability to quantify profit and loss, whereas the combined model addresses this limitation. This ability is crucial in the stock market, providing the ultimate advantage for profit by executing accurate trades.
- The integrated model can offer a more holistic analysis of financial data. This includes considering not only sentiment but also the relationships and dependencies between entities which is needed in most of the trading strategy.