

Networks of Presidential Candidates and the US Politics

Student: Zhang Ningyi, Advisor: Nishant Malik

Introduction

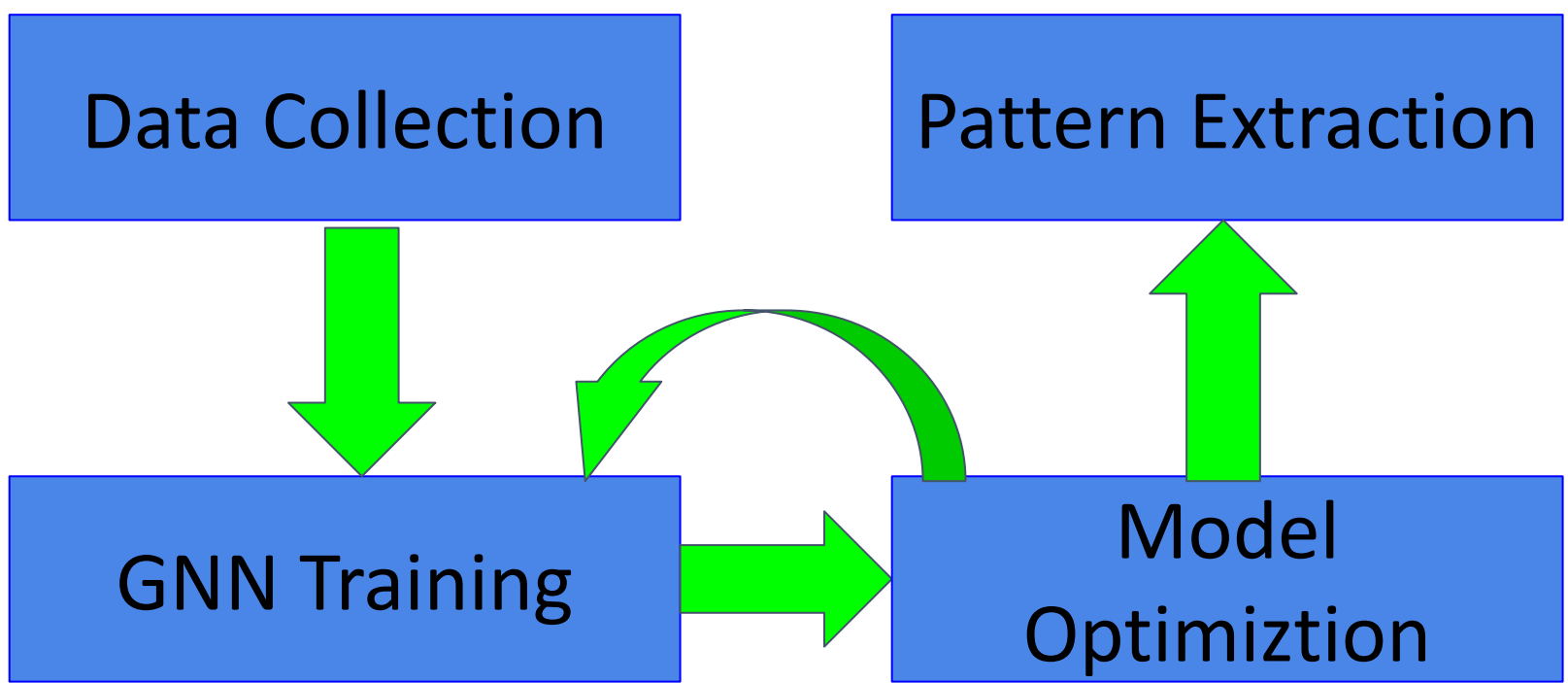
- The relationships of presidential candidates with donors and other influential individuals in the American politics can be represented by networks

Project Design

- Try to implement a reasonable model trained to predict election outcomes based on relationship dataset
- Add some ML techniques and complex system techniques to make the model better on performance and robustness
- Extract patterns that may play important role in election victory.

Methodology

- Data Collection:** Download programmatically from LittleSis.org, an online watchdog which has compiled a large data on influential individuals and organizations and relationships among them.
- Training:** Implement GNN(Graph Neural Network) and train using the downloaded and pre-processed dataset
- Model Optimization:** Involve GAT(Graph Attention Network) and Resilience Optimization to improve model
- Pattern Extraction:** Using GAT score and Leiden Community Detection to find weight difference between relationship clusters

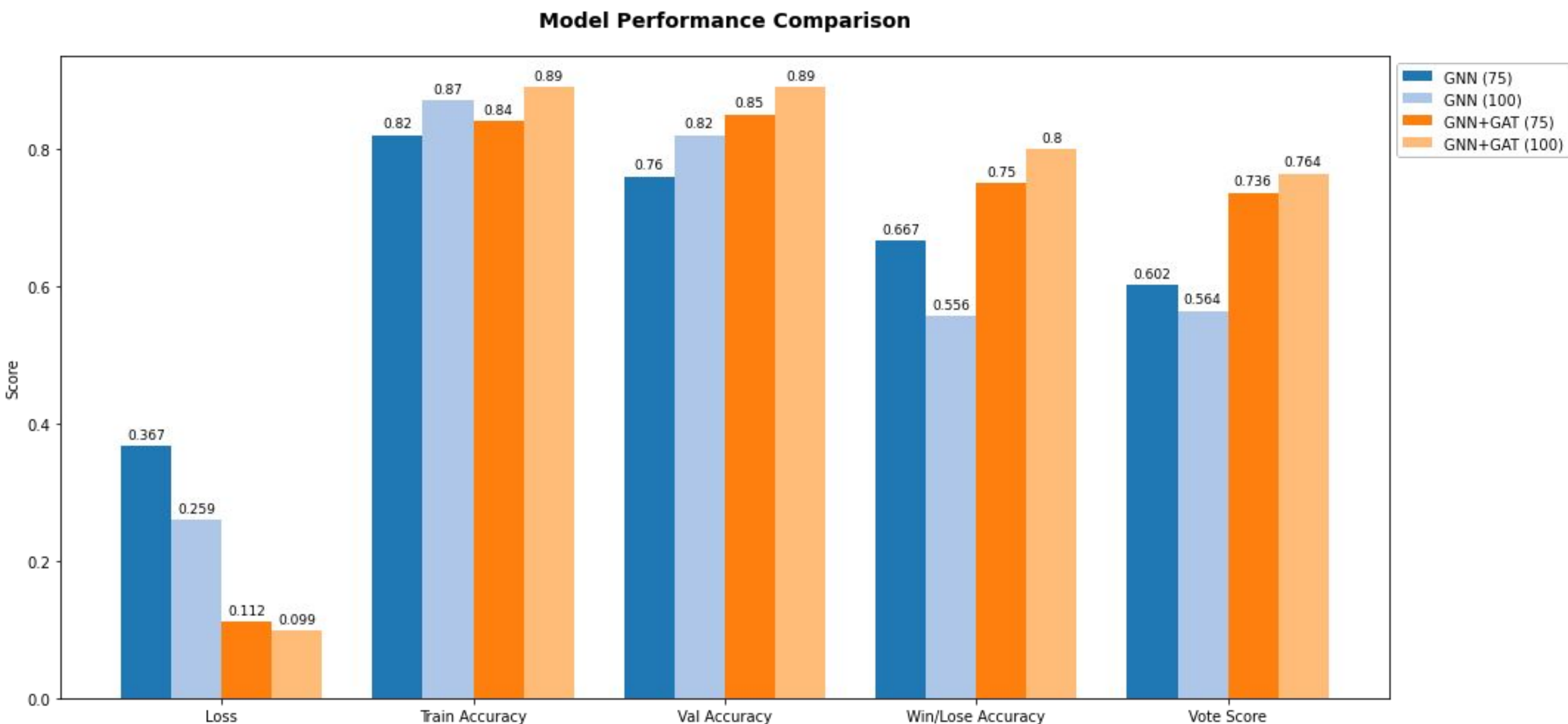


Data Processing

- Data Selection:** Pick presidential candidates of US election from 1988-2024, whose candidates have a moderate amount of clear relationship data.
- Data Fetching:** Use requests in python script to download relationship dataset from LittleSis.org to local repository. After filtering and pre-processing, there's over 100,000 relationship records stored in graph for training.
- Pre-processing:** Note that same candidate in different elections should be regarded as different candidates. Only small portion of useful attributes of relationship data is picked for training, such as start_date, donation amount, donors' type .etc

Expected Results

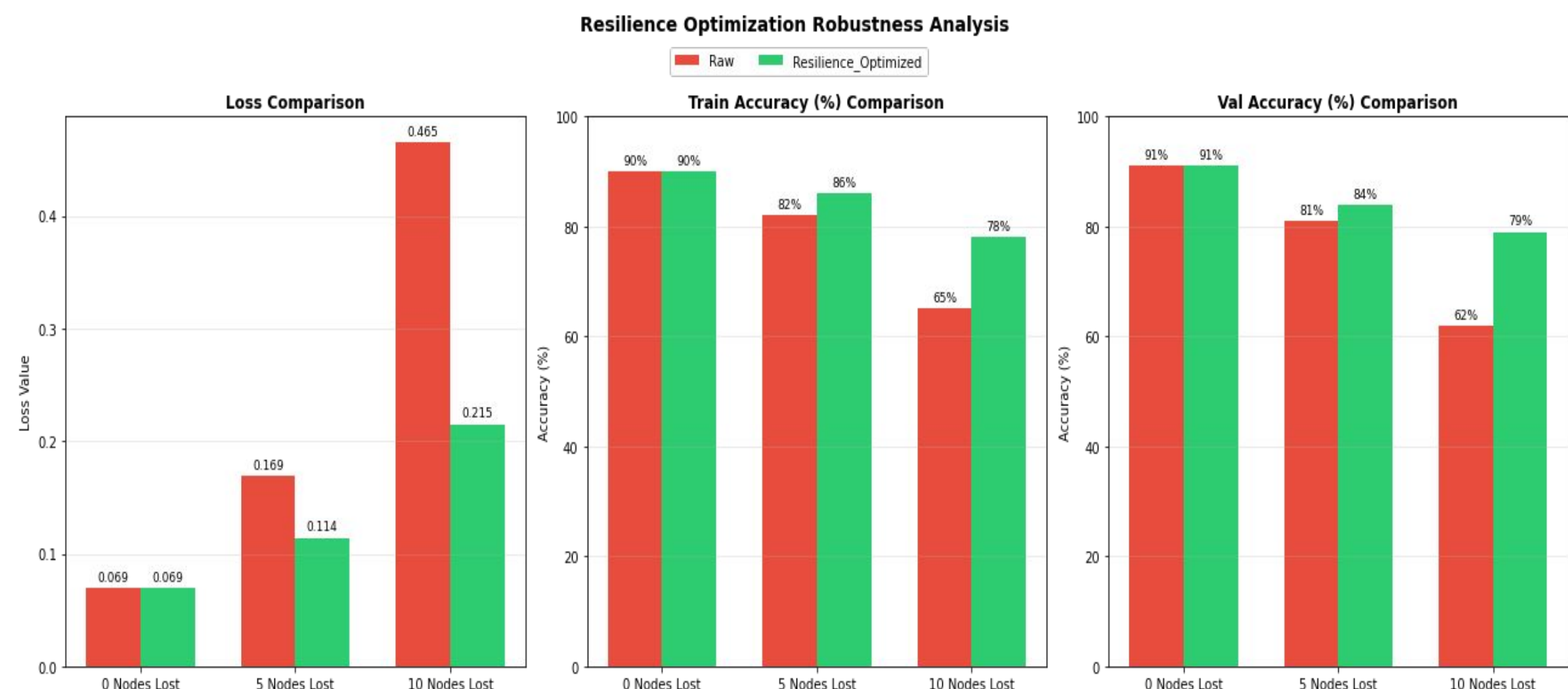
Model Performance



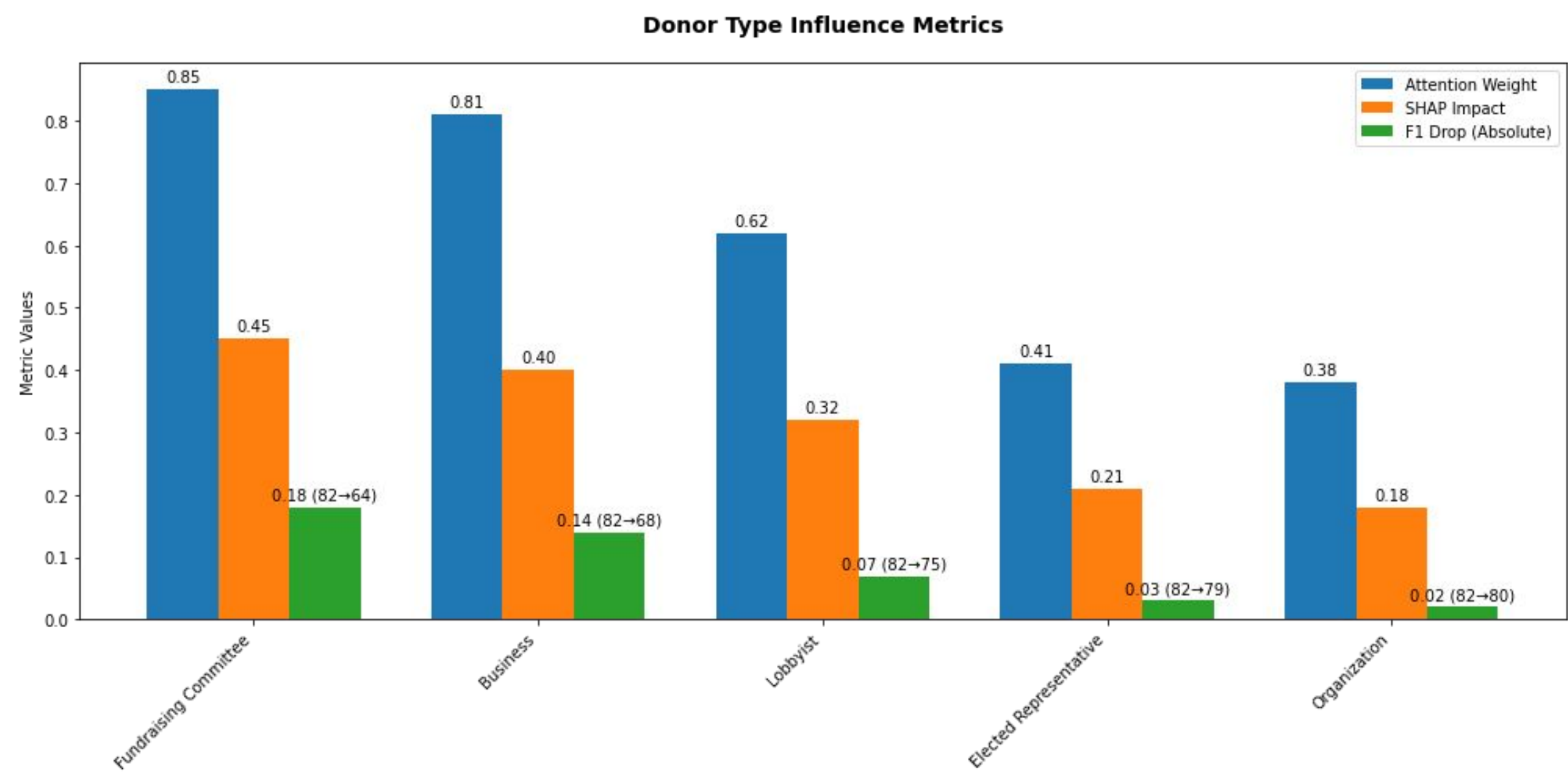
This picture shows the performance improvement adding GAT that involves self-attention mechanism.

Model Robustness

This diagram shows the robustness change of the model after applying Resilience Optimization, which is practical when facing data lost, data contamination or attacks.



Pattern Extraction



Using GAT score, we got the 5 most heavy-weighted type of the donor/endorser

Conclusions

- GNN can be trained to predict US election outcomes under un-exact standard.
- Resilience Optimization can be used to deal with dataset that is not ideal
- Leiden Community detection and more complex-system related techniques can be added to improve the model analysis, far from exhausted explore in this project
- Interactive relationships between “neighbor” of one node is significant for this kind of election network analysis, which has much larger memory and time cost, cannot be achieved in this project due to data process efficient and techniques