Industry, Committee, and Lobbying - Uncovering Congressional Stock Trading using Graph Data

Suyeol Yun

May 8, 2023

Literature Review: Approaches & Puzzles

- Eggers & Hainmueller (2014) examine congressional portfolios to determine disproportionate investments in firms with PAC contributions (+), district connections (+), or lobbying activities related to the congressperson's committee (-).
- *"... We find no evidence that members disproportionately invest in companies to which they are connected through their committee assignments."*
- This presents a new puzzle:
  - Extensive research exists on committee specialization (King, 1994; Asher et al., 1974; Myers, 2007)
  - Why is there an absence of a relationship between lobbying, committee assignments, and congressional investments?

Limitation of Eggers & Hainmueller (2014)

$$w_{ij} = \alpha + \beta_1 \text{ District }_{ij} + \beta_2 \text{ Contributions }_{ij} + \beta_3 \text{ Lobbying \& CA }_{ij} + \theta_i + \theta_j + \varepsilon_{ij}$$

- Firm-level analysis: w_{ij} is portfolio weight of a firm j for congressperson i
- Congresspeople invest at industry level
- Approximately 60% of reported Senators' trades are ETFs or mutual funds, reflecting industry-level trading
  - e.g. Sheldon Whithouse traded US Medical Devices ETF (IHI)
- Lobbying involves complex industry-level interactions that a binary firm-level indicator and linear model cannot capture

Resolution: Using Graph Structured Data

- Graphs are network data with nodes and edges
- They capture intricate interactions among firms, bills, committees, and congresspersons
- Graph data compiled from Lobbyview, Senate/House Financial Disclosure, Congress, and naics.com
- url-based entity disambiguation; No more similarity based matching; $O(nm) \rightarrow O(n+m)$
- The graph includes 55,700 nodes and 264,000 edges
- Data spans the 110th-117th Congresses (2007-2021)

Graph Data Specification

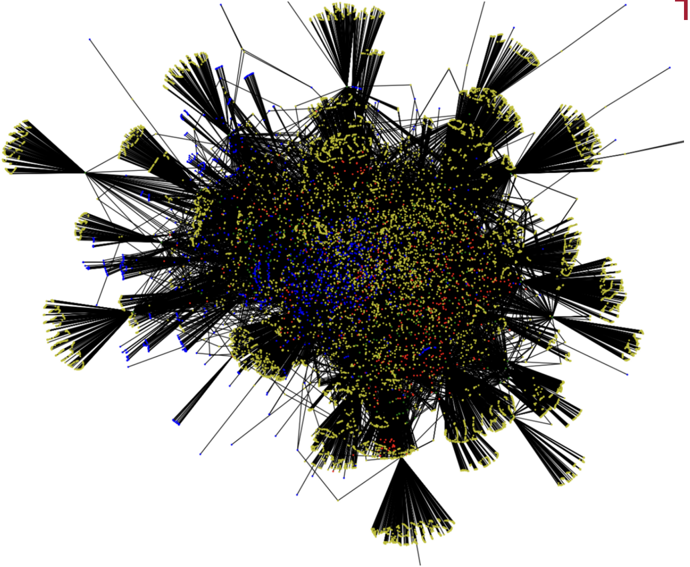
Table 2: Heterograph (Nodes)

Node Type	N	Period	Source
Firm (Ticker)	4,202	-	Lobbyview & Finance Disclosure
Bills	47,767	110-117th Congress	Lobbyview
Congressperson	2,431	113-118th Congress	Lobbyview & Finance Disclosure
Committee	556	-	Lobbyview
NAICS code	744	-	naics.com
Total	55,700	-	-

Table 3: Heterograph (Edges)

Edge Types	N	Period	Source
Congressperson- Buy/Sell- Firm (Ticker)	24,675	[2013-01-24, 2023-03-08]	Finance Disclosure
Firm (Ticker) - Lobby On - Bill	148,487	[2016-01-02, 2022-02-24]	Lobbyview
Ticker- Classified as - NAICS Codes	4,147	-	Finance Disclosure & naics.com
Bill- Referred to - Committee	75,626	[2016-01-05, 2021-12-17]	Lobbyview
Congressperson- Assigend to - Committee	11,698	115-117th Congress	Finance Disclosure & Lobbyview
Total	264,633	-	-

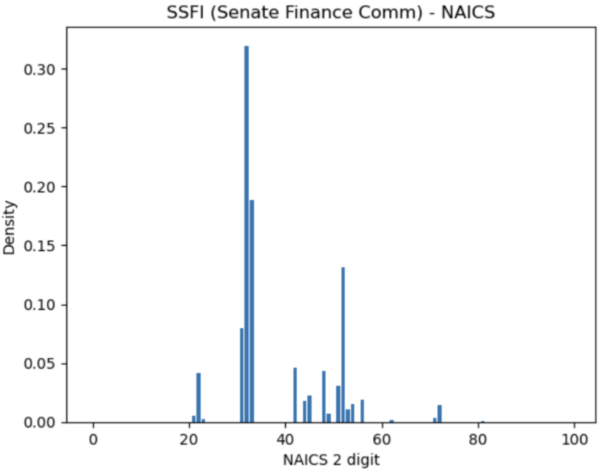
Entire Graph



very dense, hard to interpret

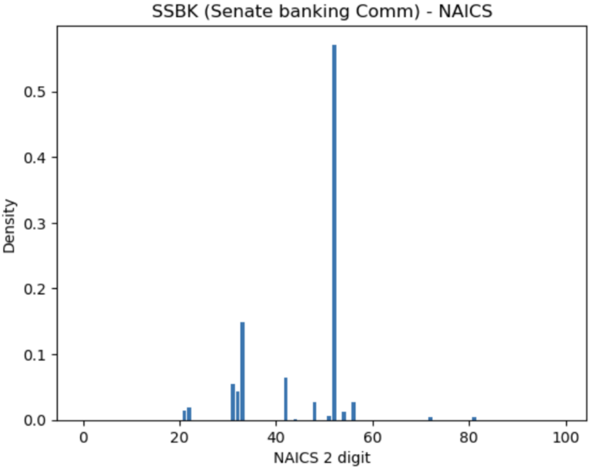
Industry-Level Specialization

- Committee specialization can be quantified by aggregating the NAICS codes of firms lobbying on bills referred to the committee.
- NAICS PMF of Senate Finance Committee



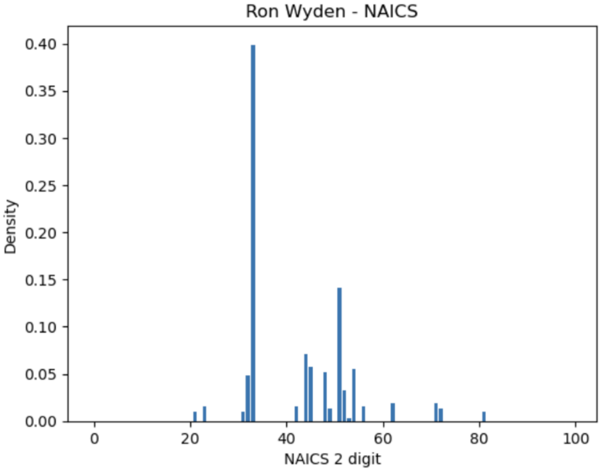
Industry-Level Specialization

- NAICS PMF of Senate Banking Committee



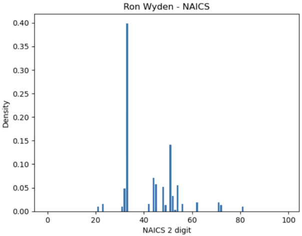
Industry-Level Specialization

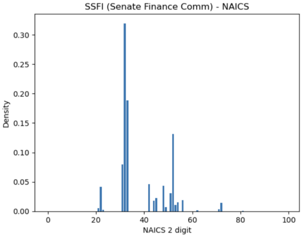
- Similarly, one can quantify a congressperson's industry-level specialization in their stock portfolio by aggregating the NAICS codes of firms they transacted with.
- NAICS PMF of Senator Ron Wyden

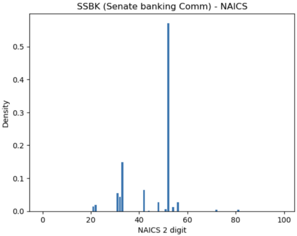


Industry-Level Specialization

- We can directly measure the similarity between NAICS PMF of congressperson and committee.
- Using Cross-Entropy: $H(P, Q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x)$
- Lower, the similar

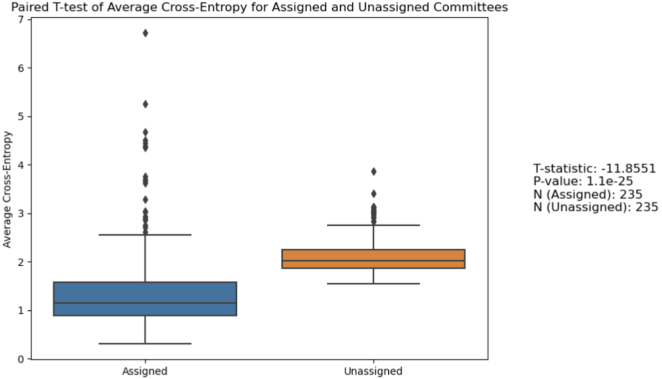






- CE(Ron Wyden, SSFI) $= 0.7 <$ CE(Ron Wyden, SSBK) $= 3.3$
- Sen. Ron Wyden is a member of the Senate Committee on Finance
- His portfolio resembles the committee's industry specialization

Average Cross Entropy: Assinged vs Un-Assigned



- Congresspeople's stock portfolio resembles their committee's industry specialization
- Contrast to Eggers & Hainmueller (2014) *"we find no evidence that members disproportionately invest in companies to which they are connected through their committee assignments."*

Predictive Modeling: Graph $\rightarrow$ Transaction

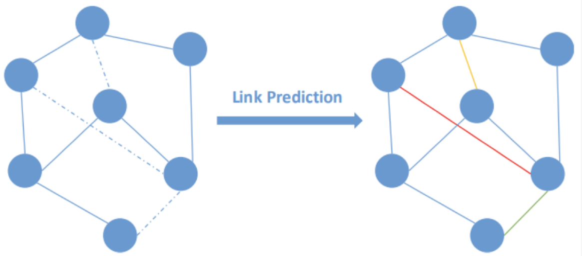
Egger & Hainmueller (2014) model:

$$w_{ij} = \alpha + \beta_1 \text{ District }_{ij} + \beta_2 \text{ Contributions }_{ij} + \beta_3 \text{ Lobbying }_{ij} + \theta_i + \theta_j + \varepsilon_{ij}$$

- w_{ij} is portfolio weight of a firm j for congressperson i
- How predictive is the graph for stock transactions?

Link Prediction using Graph Neural Network

- Link prediction is a task of predicting the existence of a link between two nodes in a graph
- Given two nodes u and v, we want to predict whether there is an edge between them
- u is a congressperson and v is a firm (ticker)



Link Prediction using Graph Neural Network

Idea is similar to logistic regression:

$$\pi_i = \text{sigmoid}\left(X_i^\top \beta\right) \equiv \frac{\exp\left(X_i^\top \beta\right)}{1 + \exp\left(X_i^\top \beta\right)}$$

where $\pi_i \in [0, 1]$

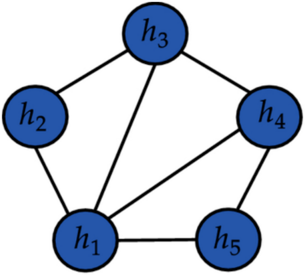
- Replace $X_i^\top \beta$ (logit) with $PredHead(h_{congressperson}, h_{ticker})$
- $PredHead$ is normally a dot product or a neural network
- $h_{congressperson}, h_{ticker}$ are multi-dimensional vectors that represents each node (similar idea like D(W)-Nominate score)
- How to learn $h_{congressperson}, h_{ticker}$ to encode the information embedded in the graph structured data?
- Using Graph Neural Network (GNN)
- GNN should map $h_{congressperson}, h_{ticker}$ into a close distance if they are connected in the graph

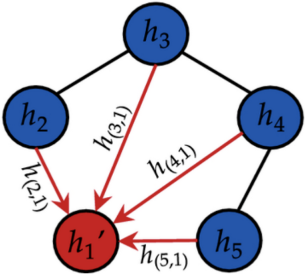
GNN Architecture

I used Edge-Conditioned Convolution (Simonovsky, 2017)

$$\mathbf{h}'_i = \mathbf{\Theta}\mathbf{h}_i + \sum_{j \in \mathcal{N}(i)} \mathbf{MLP}\left(\mathbf{e}_{i,j}\right) \cdot \mathbf{h}_j$$

- GNN is computation graph of iteratively updating node representations
- Message Passing, Aggreagation, and Update

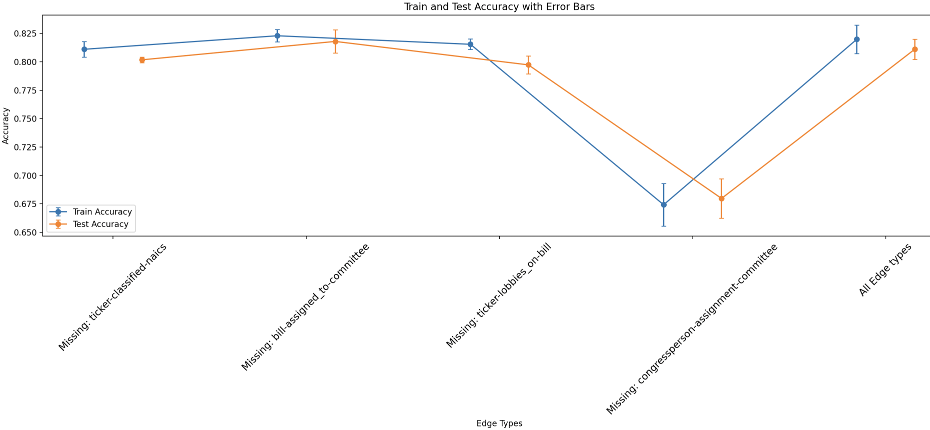




Training and Evaluation

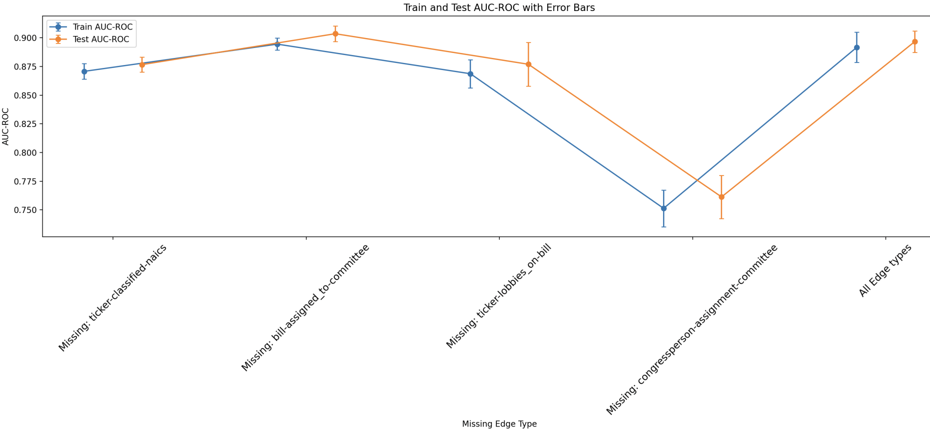
- Total 24,675 edges for edge-type (congressperson, buy/sell, ticker)
- 80% of the edges are used for training, 20% for testing
- Same number of negative edges are sampled for training (Negative Sampling; For balanced training)
- 5-fold cross validation for uncertainty statistics
- Ablation study for feature importance - removing each edge-types and see how the performance changes

Performance: Accuracy



Use All Edge Types: 82% accuracy
Remove Committee Assignments of MC: 65% accuracy

Performance: AUC-ROC



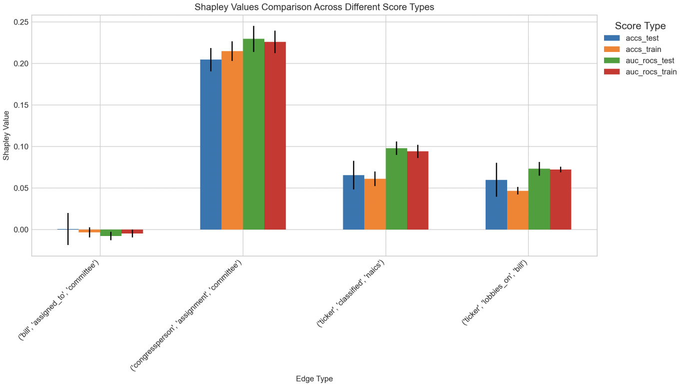
- The performance drops the most when we remove the committee assignments for congresspersons.
- Committee assignments are the most important feature for predicting stock trading by congresspersons.

Shapley Value

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!}(v(S \cup \{i\}) - v(S))$$

- Shapley value represents the fair contribution or importance of each player
- Shapley value can be applied to compute the significance of different edge types in link prediction task
- Measure: How much each edge type contributes to the prediction?
- Can be computed by $16(= 2^4)$ different combinations of 4 edge types

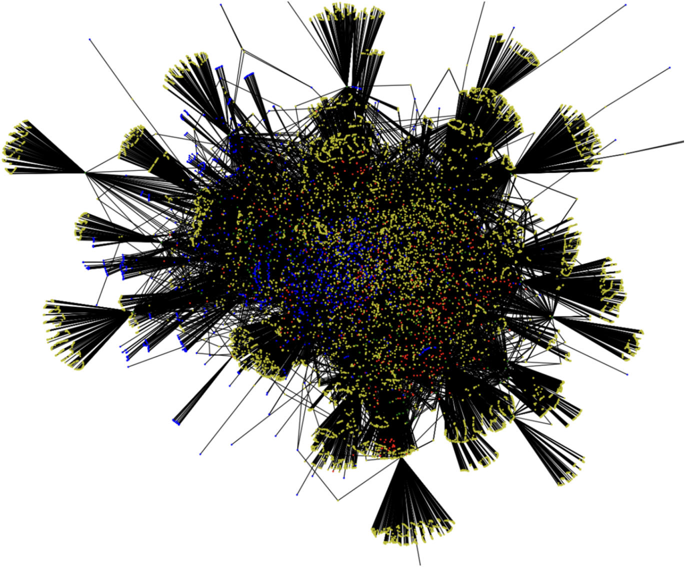
Shapley Value



- Committee assignment is key for predicting congresspersons' stock trading.
- Firm lobbying on bills and NAICS code also matter.
- Bill referral to committees isn't helpful—it hurts performance.
- Incomplete Lobbyview data could be a factor - Parsing bills from lobbying reports is hard.

GNNExplainer

Which nodes and edges are important for the prediction?



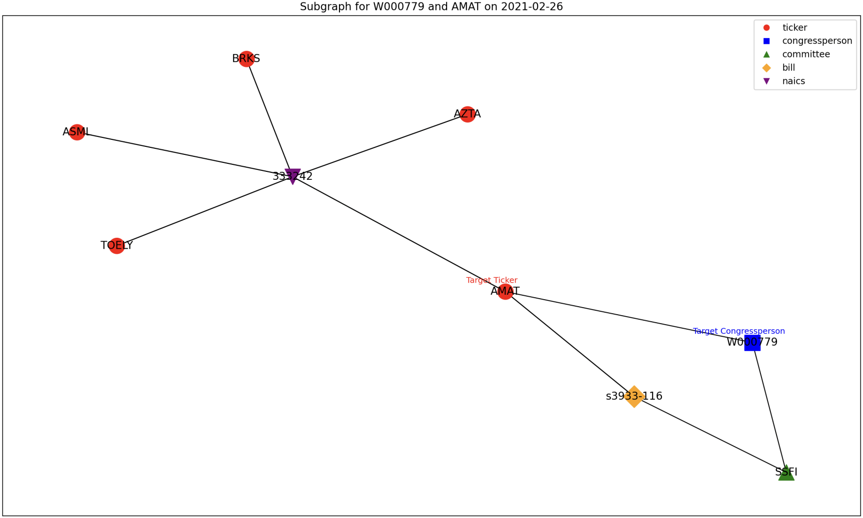
GNNExplainer

GNNExplainer: Generating Explanations for Graph Neural Networks (Ying et al., 2019)

$$\min_M - \sum_{c=1}^{C} \mathbf{1}[y=c] \log P_\Phi \left(Y = y \mid G = A_c \odot \sigma(M), X = X_c \right)$$

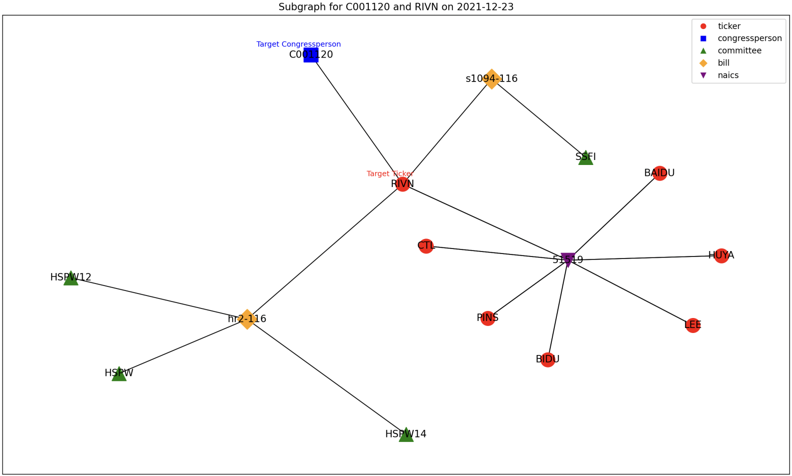
- If model predicts Ron Wyden traded Applied Materials Inc. (AMAT), then which nodes and edges are important for the prediction?
- Optimize node and edge masks M that minimize the difference between prediction on the original graph and the masked graph.
- Can add L1 regularization to control the sparsity - how sparse (simple) explanation do you want?
- Aim to recover most simple but powerful explanation in form of subgraph of the original graph.

Ron Wyden - Applied Materials Inc. (AMAT)



- S.3933-116: CHIPS ACT

Dan Crenshaw - Rivian Automotive Inc (RIVN)



- S.1094-116: Driving America Forward Act

Conclusion

- Committee Assignments and Firm-level Lobbying on Bills are the most significant predictors of congresspersons' stock trading behavior, contrary to the findings of Egger and Hainmueller (2014).
- Congressional activities are informative for predicting congresspersons' stock trading.
- Future Research: Predict "excess returns" using the same graph data to assess the extent to which congressional activities can explain variations in excess returns.