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

Suyeol Yun *

May 10, 2023

Abstract

In this paper, I revisit a claim from a previous study asserting that a U.S. Congressperson's stock trading behavior lacks significant associations with their committee assignments and firm-level lobbying on bills. I challenge this view by presenting evidence suggesting that Congresspersons tend to exhibit industry specialization in their stock trading, aligning with the industries pertinent to their committee assignments. To explore this relationship, I employ a graph-structured dataset that comprehensively encapsulates a variety of legislative activities. Leveraging the Graph Neural Network (GNN), I elucidate the complex interplay between congressional activities and stock trading behavior. Through an ablation study on the GNN's predictions, I identify the committee assignment of a congressperson and firm-level lobbying on specific bills as crucial factors in predicting a congressperson's specific stock choices. This research thus offers a fresh perspective that aligns more closely with traditional literature on committee specialization, demonstrating that such legislative roles also significantly influence a congressperson's stock trading behavior. This finding underscores the relevance of committee-level specialization in understanding the financial decisions of Congresspersons.

*PhD Student, Department of Political Science, MIT. Email: syyun@mit.edu

1 Introduction

In democratic societies, electoral accountability is a central mechanism through which politicians, vested with the power to make significant public decisions, remain answerable for their actions (Besley, 2006; Fearon, 1999; Ferejohn, 1986). However, the intermingling of public service and personal financial interests can blur the lines of accountability. A key question arises: how closely intertwined are a politician's personal financial gains and their public service role? Are these two realms neatly partitioned, or do they bleed into each other?

The moral hazard theory raises concerns that when individuals are shielded from the repercussions of their actions, they might behave differently than if they bore the full consequences (Holmström, 1979). This theory is especially pertinent in the political sphere, where politicians could leverage their positions for personal financial gains, leading to potential conflicts of interest. The theory underscores the need for a deeper theoretical and empirical examination of politicians' behavior in this regard.

Public choice theory, in addition, posits that it is unrealistic to presume a clear-cut divide between the public and private domains of a politician's life (Buchanan and Tollison, 1984). The theory suggests that individuals aim to maximize their personal utility in both their public and private roles. This notion aligns with Fenno's home style approach, which suggests that personal interests can often shape political actions (Fenno, 1977).

Given these theoretical considerations, the financial behavior of politicians, particularly their stock trading activities, warrants closer examination. The trading activities of U.S. Congresspersons offer a unique opportunity to investigate the relationship between public service and personal financial interests. By analyzing the stock trading patterns of Congresspersons and their legislative activities, we can gain valuable insights into the potential conflicts of interest and the implications for electoral accountability.

This issue of financial behavior among politicians has been explored from various angles. Legal scholars have examined the use of political intelligence for profit and the potential for insider trading within the political sphere (Jerke, 2010; Bainbridge, 2010). These studies highlight the challenges of designing legal and enforcement structures to prevent and deter such behavior. Additionally, research by Lenz and Lim (2009) provides an indirect test of corruption in Congress by comparing wealth accumulation among U.S. House members and the general public. Their findings suggest that representatives accumulate wealth about 50 percent faster than expected, raising questions about the sources of this wealth accumulation and its implications for political integrity.

While these studies provide valuable insights, there has also been attention given to anecdotal approaches that explore specific instances of insider trading by politicians. For example, Schweizer (2011) presents a series of case studies that highlight how politicians and their associates have profited from insider stock tips, land deals, and cronyism. These accounts, while informative, underscore the need for a more systematic and empirical investigation of stock trading behavior among politicians, moving beyond individual anecdotes to a broader analysis of patterns and trends.

Recognizing this need to move beyond individual anecdotes to a broader analysis of patterns and trends, researchers have turned to the excess return approach, which has been used in the broader literature on insider trading, to assess whether politicians achieve abnormal returns on their investments. The excess return approach provides a foundation for understanding how informational advantages may lead to excess returns in financial markets. For example, Jeng et al. (2003) estimated the returns to insider trading from a performance-evaluation perspective, while Ivković and Weisbenner (2005) examined the information content of the geography of individual investors' common stock investments. Seasholes and Zhu (2009) similarly tested whether individual investors excel in investing in local non-S&P 500 companies. Building on this approach, several studies have specifically examined the stock trading activities of U.S. Congresspersons. Ziobrowski et al. (2004) and Ziobrowski et al. (2011) conducted pioneering research in this area, analyzing the common stock investments of members of the U.S. Senate and House of Representatives, respectively. Their findings suggest that members of Congress achieve abnormal returns on their stock investments, raising questions about the potential use of privileged information in their trading decisions.

However, the notion that members of Congress achieve excess returns on their investments has been challenged by other researchers. Eggers and Hainmueller (2013) call into question the consensus that members of Congress trade with an information advantage. They reinterpret existing studies of congressional stock trading between 1985 and 2001 and conduct their own analysis of trades in the 2004–2008 period. Their findings challenge the notion of widespread "insider trading" in Congress, concluding that in neither period do members of Congress trade with an information advantage. Furthermore, they conduct the first analysis of members' portfolio holdings, showing that between 2004 and 2008, the average member of Congress would have earned higher returns in a passive index fund.

The studies that employ the "excess return" approach to analyze congressional stock trading in the U.S. typically focus on the "average" performance of Congresspersons as a whole. This approach aggregates the stock trading activities and returns of all members of Congress, calculating an average excess return for the entire group. While this methodology provides insights into the overall trading performance of Congress as a collective entity, it may not fully capture the individual-level behavior of specific politicians. By aggregating the data and calculating average returns, the approach may mask variations and nuances in the trading behavior of individual members of Congress. As a result, instances of potentially unethical trading behavior by specific politicians, who may leverage privileged information for personal financial gain, could be obscured within the average calculations. This limitation is particularly relevant in light of anecdotal findings, such as those presented by Schweizer (2011), which highlight case studies of individual politicians who have allegedly profited from insider stock tips, land deals, and cronyism. These accounts underscore the need for a more granular examination of stock trading behavior at the individual level, moving beyond the analysis of average excess returns for the entire group of Congresspersons.

Additionally, the excess return approach on congressional stock trading mostly adopts a calendar time-based synthetic portfolio approach (Shanken, 1996; Annaert et al., 2009), which involves buying and selling stocks according to the Congressperson's decisions and then selling them a year later. This methodology obscures the importance of the "timing" aspect of such Congresspersons' stock trading, an aspect already well-highlighted by researches like Schweizer (2011) and Tahoun (2014). By focusing on calendar time and ignoring the timing of trades, these studies may miss critical insights into the potential exploitation of non-public information by individual politicians. This underscores the need for further research and alternative methodologies that take into account both the individual-level trading behavior and the timing of trades to provide a more comprehensive understanding of congressional stock trading practices.

In this regard, my research aims to delve deeper into the stock trading activities of individual Congresspersons, estimating the excess return of each Congressperson-ticker level transaction for each specific cycle of such transactions. These transaction cycles begin with a series of iterative purchases up to a certain point, after which they transition into a series of sales. This granular analysis allows for a more precise examination of individual trading behavior, shedding light on potentially questionable activities that might otherwise be concealed within broader, aggregate assessments.

Throughout this estimation, the findings of this study lend further support to the conclusions drawn by Eggers and Hainmueller (2013). Their research found no widespread evidence of excess returns among Congresspersons, suggesting that insider trading is not a pervasive issue within the U.S. Congress. However, they noted that certain anomalies did exist, implying that a few individual Congresspersons could potentially be generating excess returns on specific tickers.

In line with these observations, my estimation shows a positively skewed distribution of excess returns. While negative excess returns are limited, positive excess returns span a wider range and include some notable outliers. This pattern suggests that while most Congresspersons do not achieve significant excess returns, a small number of individuals may be generating substantial positive returns on specific tickers. This lends further credence to the idea that a few individual Congresspersons could potentially be leveraging their informational advantages for financial gain.

In addition, the breadth and depth of political connections, and their financial implications for both firms and politicians, have been extensively examined across a wide range of literature. Connections have been shown to significantly impact firm value (Roberts, 1990; Khwaja and Mian, 2005; Goldman et al., 2009), and politicians have been found to derive financial benefits from them (Diermeier et al., 2005; Lenz and Lim, 2009; Querubin and Snyder, 2013). Boller (1995) even demonstrated that members of the United States Congress regularly purchase common stock in companies that they regulate through legislation, highlighting the direct intersection between political activities and financial decision-making. Following this line of inquiry, Eggers and Hainmueller (2014) delved further into the topic, attempting to understand why congressional stock transactions generally appeared to be low in profitability. Their research suggested that transactions with a "political connection", such as local ties or campaign contributions, performed better than transactions without such connections. This revelation implies that political connections could facilitate a certain kind of information arbitrage where politicians can leverage their privileged knowledge obtained from their positions and legislative activities in their stock transactions.

However, Eggers and Hainmueller (2014) also found that "we find no evidence that members disproportionately invest in companies to which they are connected through their committee assignments", despite finding such behaivors in local-connection or campaign finance connections. This introduces a new puzzle, which seems to stand in contrast with the well-established findings regarding committees' specialization in various topics within their authority conferred by statutory jurisdiction (Myers, 2009) or by their bill referral (King, 1994), as well as members' specialization in such topics (Asher, 1974).

The nature of committee assignments and the specialized knowledge they confer have been a central focus of legislative studies, with scholars recognizing their critical role in shaping legislative outcomes and individual member's behavior (King, 1994; Patterson, 1970). Despite this, the connection between these committee assignments and stock trading behavior remains underexplored, presenting an interesting avenue for this research.

In this study, I re-examine the relationship between legislative activities and their stock transactions, using graph-structured data. This approach is motivated by the nature of the data we collected, which shows that 60% of the Senators' stock transactions are mostly ETFs or mutual funds, which are typically targeted at specific industries, not individual firms.

Previous research, such as the study conducted by Eggers and Hainmueller (2014), focused on estimating the impact of firm-level lobbying or committee-assignment of bills lobbied by specific firms on the increase of that specific firm's weight in a congressperson's portfolio. However, this approach does not account for a congressperson's specialization in industry-level knowledge, gained through their committee assignments, and the potential utilization of such industry-level knowledge in shaping their personal investment portfolio.

To address this gap, I collected data that captures diverse aspects of legislative activities - such as firms' lobbying on specific bills, bills assignment to specific committees, and committee membership of congresspersons - alongside with congresspersons' stock trading data. I sourced this data from various relevant platforms, such as Lobbyview (Kim, 2018), Senate & House's financial disclosures, and Congress.

I represented the collected information in a graph-structured data format. This allowed me to conduct an analysis that could reveal a clear resemblance between a congressperson's portfolio and the industry specialization of the committees they are assigned to. To further investigate this relationship, I conducted a paired t-test (Hsu and Lachenbruch, 2014) to determine if there is a significant difference between two sets of average cross-entropy values, each measuring the similarity between the industry distribution of a congressperson's stock transactions and the industry distribution of committees. In the paired t-test, each pair consisted of the average cross-entropy value for a congressperson's assigned committees and the average cross-entropy value for their unassigned committees.

The result of the paired t-test shows that a congressperson's stock trading pattern significantly resembles the industry distribution of their assigned committees more than that of non-assigned committees. This finding is in stark contrast to the conclusions drawn by Eggers and Hainmueller (2014), highlighting one of the novel contributions of this research in understanding the relationship between committee assignments and stock trading behavior among Congress members.

While this research arrives at a conclusion that is diametrically opposed to that of Eggers and Hainmueller (2014) regarding the importance of committee assignments in shaping a congressperson's investment portfolio, it aligns with the broader trend in the study of congressional stock trading. The work of Eggers and Hainmueller (2014) is valuable in that it moves beyond the traditional excess return approach and instead seeks to identify meaningful associations between various factors that could influence a congressperson's choice of stocks, such as PAC donations, district-level connections, or committee assignments.

In a similar vein, this study emphasizes the need to understand how legislative activities in general, which encompass a vast and complex network of information flows and interactions, can impact a congressperson's choice of specific stocks. These activities include firms lobbying on bills of particular interest to them, the referral of these bills to specific committees based on statutory and historical jurisdiction, and the assignment of congresspersons to these committees based on their expertise.

This approach is further justified by public choice theory, which posits that Congresspeople possess hybrid identities as both public and private entities (Buchanan and Tollison, 1984). Given that individual investors often experience failures (Barber and Odean, 2000; Barberis and Thaler, 2003), it is reasonable to assume that Congresspeople may also face similar failures when investing as individuals, rather than utilizing privileged information as public figures. This underscores the importance of studying the direct relationship between congressional activities and stock trading patterns, rather than relying solely on tests of excess returns.

Considering the importance of timing in trading activities (Tahoun, 2014), this research proposes a more fundamental approach to study these behaviors, directly tackling the problem from an information perspective. The aim is to test whether congressional activities, which are often centered around legislative activities, provide information to predict each Congressperson's specific ticker transactions. To achieve this, a predictive model is designed that takes into account a variety of factors tied to each Congressperson's legislative activities. This will include elements such as the committees they are assigned to, the bills being legislated through those committees, and the potential interests of various firms or industries related to those bills. The goal is to ascertain whether these factors can reliably predict a Congressperson's transaction with a specific ticker at a specific time. For example, if a Congressperson, as posited by Public Choice Theory (Buchanan and Tollison, 1984), seeks to maximize personal utility, they may be motivated to transact with tickers of specific firms they are legislating about.

In order to effectively capture the complex nature of congressional activities, I have collected and organized data using a heterograph, a type of graph structure that incorporates different types of nodes and edges. This graph-structured data is particularly useful for representing various entities and their relationships, which are inherent in the legislative process. These relationships include firms lobbying on bills of particular interest to them, the referral of these bills to specific committees, and the assignment of congresspersons to these committees. Each of these elements can be represented as nodes in the heterograph, with edges indicating the relationships between them. For example, an edge can represent a firm lobbying on a bill, a bill being referred to a committee, or a congressperson being assigned to a committee. Therefore, the heterograph structure allows for the inclusion of multiple types of relationships, or edges, between the same pair of nodes. This is crucial in our context, where a congressperson can have multiple types of connections to a committee or a bill.

Pursuing the approach of Eggers and Hainmueller (2014), this research endeavors to predict the specific stock transactions of given congresspersons. If these transactions can be forecast using the heterograph, it implies that congressional activities contain vital information that can explain their stock trading patterns. Such a finding is significant as it highlights a potentially strong association between a congressperson's legislative activities and their stock transactions.

The predictive task, in this case, is modeled as a link prediction task. Essentially, this task aims to ascertain the presence or absence of links (edges) between a congressperson (node) and a specific stock ticker (node). Successfully predicting the presence of an edge would mean that a relationship exists between a congressperson's legislative activities and their likelihood to transact with a specific stock ticker.

In conducting this prediction task, the research successfully trained a Graph Neural Network (GNN) model (Zhou et al., 2020; Wu et al., 2020; Scarselli et al., 2008; Zhang et al., 2019), achieving an accuracy of 0.81 and an AUC-ROC score of 0.89. These results indicate that congressional activities provide substantial information to explain the stock choices of congresspersons, further supporting the hypothesis of a significant correlation between legislative activities and stock transactions.

To understand the varying contribution of different types of edges in the heterograph to the prediction task, an ablation study was conducted. It systematically removed particular edge types from the training and calculated the Shapley values (Winter, 2002; Hart, 1989; Littlechild and Owen, 1973), which measure the contribution of each edge type to the prediction. This study found that a congressperson's committee assignments and firm-level lobbying on bills were the most important factors contributing to the model's predictive power. Interestingly, this finding contradicts Eggers and Haimuller's (2014) conclusion, which argued that there's no significant evidence that firm-level lobbying on bills and a congressperson's committee assignments explain the weight of specific stocks in their portfolio.

Furthermore, to address the black-box nature of GNN predictions (Dayhoff and DeLeo, 2001; Buhrmester et al., 2021; Olden and Jackson, 2002), this research incorporates the use of GNNExplainer (Ying et al., 2019). This tool allows for the identification of the most critical nodes and edges for each prediction, which in turn enhances the interpretability of the model. It provides a clearer understanding of the reasons behind the model's prediction regarding the existence of an edge (transaction) between the congressperson and the ticker nodes. I will provide examples illustrating this interpretive capability, demonstrating how it brings greater transparency to the decision-making process of the GNN model. This step is crucial in reinforcing the trustworthiness and understandability of the model, and in promoting its practical applicability in real-world scenarios.

To summarize, this research contributes to the field in several significant ways. First, it corroborates the findings of Eggers and Hainmueller (2013) by affirming the absence of widespread excess returns in congressional stock trading. Second, it reveals the presence of asymmetrically high excess returns when congresspersons earn, as opposed to when they incur losses. This suggests a more nuanced understanding of the financial outcomes of congressional stock trading. Third, the research proposes a direct measurement of the resemblance between the industry-level specialization of committees and the industry-level distribution of a congressperson's stock portfolio. This novel approach enhances our understanding of how legislative roles may influence financial decisions. Fourth, this study advocates the use of graph-structured data and designs a predictive analysis using a Graph Neural Network (GNN). This approach allows for a more comprehensive understanding of the relationship between congressional activities and stock transactions. Finally, the research underscores the significance of a congressperson's committee assignments and firm-level lobbying in explaining the choice of stock transactions. This conclusion contradicts some previous studies, but aligns more closely with traditional literature on the influence of committee specialization and firm-level lobbying. Thus, it provides a fresh perspective on the intricate dynamics of legislative activities and their potential financial implications.

2 Estimating Excess Returns of Congressional Stock Trading

[1]The broader literature on insider trading has long explored the information advantages that certain individuals, such as corporate insiders or well-connected investors, may possess when trading in the stock market. For example, Jeng et al. (2003) estimated returns to insider trading from a performance-evaluation perspective, while Ivković and Weisbenner (2005) studied the information content of the geography of individual investors' common stock investments. These studies highlight the importance of understanding the impact of information asymmetry and potential insider trading in financial markets.

Despite the extensive research on insider trading in general, the application of this approach to congressional stock trading has been limited. In the context of congressional stock trading, previous studies, such as those conducted by Ziobrowski et al. (2004), Ziobrowski et al. (2011) and Eggers and Hainmueller (2013), have predominantly used calendar-time based portfolio approaches (Hoechle and Zimmermann, 2007) to estimate excess returns. This involves creating synthetic buy and sell portfolios (Shanken, 1996; Annaert et al., 2009) that mimic congresspersons' stock purchases and sales but sell or buy such stocks after a year.

This approach, however, neglects the importance of transaction timing, which is a crucial aspect of insider trading (Tahoun, 2014; Schweizer, 2011). It does not account for the short-term fluctuations in stock prices that congresspersons might anticipate based on their access to privileged information. Congresspersons could potentially profit from these expected fluctuations by strategically timing their transactions using their privileged knowledge. This presents a severe limitation in understanding the true extent of potential insider trading among U.S. members of Congress.

In addition, averaging excess returns across congresspersons may not capture the full extent of insider trading within the inner circle of Washington D.C. politics. Schweizer (2011) provided anecdotal evidence of politicians and their friends profiting from insider stock tips, while Lenz and Lim (2009) studied corruption and wealth accumulation in Congress, and Jerke (2010) and Bainbridge (2010) examined the use of political intelligence for insider trading with several anecdotes. These case studies suggest that certain congresspersons might engage in insider trading with specific firms or industries, which would be overlooked in an aggregate analysis.

In this section, therefore, I aim to address these limitations by estimating the excess returns at

[1]Reproducible code for this section is available at `https://github.com/syyunn/efd/blob/main/anlys/cashout/fifo-rd-fed-pppsss-include-etf.py`

the congressperson-ticker level, with a focus on the life cycle of each buy/sell chain of specific tickers consecutively transacted by a congressperson. This approach offers a more granular analysis of potential insider trading among U.S. members of Congress, allowing us to better evaluate whether widespread insider trading exists at the congressperson-ticker level. By doing so, we build upon both the general insider trading literature and the existing research on congressional stock trading, contributing to a more comprehensive understanding of the potential information advantages leveraged by politicians in the stock market and the importance of transaction timing.

2.1 Data

To estimate excess returns at the congressperson-ticker level, I first needed to compile comprehensive data on the stock transactions of U.S. members of Congress. I obtained this data by scraping the Senate Financial Disclosure website[2], which provides detailed information about the stock transactions made by congresspersons, including the date, ticker symbol, and the amount of each transaction.

The resulting dataset consists of 25,023 transactions, spanning a period from January 1, 2014, to August 5, 2022. These transactions involve 74 distinct Senators and 2,114 distinct tickers. Among these tickers, around 40% (832) are individual company-level tickers, such as AAPL for Apple Inc. and AMAT for Applied Materials Inc., while the remaining 60% (1,282) are ETFs or mutual funds, like QQQ for Nasdaq-100 index funds or IHI for U.S. Medical Devices ETF.

For each transaction, I added the Volume Weighted Average Price (VWAP) in USD acquired from a commercial stock data API [3]. VWAP is a widely used trading benchmark (Madhavan, 2002; Białkowski et al., 2008; Duffie and Dworczak, 2021) that represents the average price at which a security is traded throughout the day, weighted by the volume of each trade. By using VWAP, I obtained a more representative price for each stock transaction, taking into account the varying trading volumes and prices during the entire trading day. However, the Senate Financial Disclosure data is range-censored in terms of the "amount", which represents the value of the stock transaction for that date. The amount is reported as one of the following ranges in Table 1.

It is important to note that not all of the transactions have a clear ticker because some assets are not publicly traded on an exchange. Additionally, not all transactions have VWAP values, as not all tickers or asset names have available stock price data from the data provider side. This may lead to some

[2] https://efdsearch.senate.gov/search/home/
[3] https://polygon.io/stocks

Amount Range (USD)
1,001 - 15,000
15,001 - 50,000
50,001 - 100,000
100,001 - 250,000
250,001 - 500,000
500,001 - 1,000,000
1,000,001 - 5,000,000
5,000,001 - 25,000,000

Table 1: Range of the min-max amount of each stock transaction.

first_name	last_name	ticker	asset_name	trans_date	amount_min	amount_max	vwap
John	Hoeven	QCOM	QUALCOMM Incorporated	2017-03-02	100,001	250,000	[NULL]
David A	Perdue , Jr	[NULL]	Alliant Energy Corp CMN	2015-10-21	15,001	50,000	[NULL]
Benjamin L	Cardin	VO	Vanguard Mid-Cap ETF	2021-07-23	1,001	15,000	238.6192
Pat	Roberts	BAC	Bank of America Corporation	2018-07-05	1,001	15,000	27.9392
Patrick J	Toomey	IWF	iShares Russell 1000 Growth ETF	2021-01-14	1,001	15,000	241.4834
Timothy M	Kaine	ODVYX	Oppenheimer Developing Markets Y	2015-07-13	1,001	15,000	[NULL]
Kamala D	Harris	BSV	Vanguard Short-Term Bond ETF	2017-02-28	1,001	15,000	[NULL]
Steve	Daines	[NULL]	AMERICAN TAX EXEMPT BOND FUND	2014-10-22	1,001	15,000	[NULL]
Sheldon	Whitehouse	PANW	Palo Alto Networks, Inc.	2017-01-11	1,001	15,000	[NULL]
Mark R	Warner	ANGIX	Angel Oak Multi-Strategy Income Instl	2017-05-01	15,001	50,000	[NULL]
A. Mitchell	McConnell, Jr.	VFIAX	Vanguard 500 Index Fund Admiral Shares	2018-03-23	15,001	50,000	[NULL]
Sheldon	Whitehouse	TRBCX	T. Rowe Price Blue Chip Growth Fund	2018-05-07	15,001	50,000	[NULL]
Mark R	Warner	DBLTX	DoubleLine Total Return Bond Fund Class I	2018-06-01	15,001	50,000	[NULL]
Ron L	Wyden	BLL	Ball Corporation	2020-05-07	50,001	100,000	65.1487
John W	Hickenlooper	QRTEA	Qurate Retail, Inc. - Series A Common Stock	2021-05-10	250,001	500,000	14.2215
Mark R	Warner	AGG	iShares Core U.S. Aggregate Bond ETF	2021-03-05	1,001	15,000	114.2362
Christopher A	Coons	MSAIX	Invesco American Value Fund Class Y	2020-09-10	15,001	50,000	[NULL]
Robert J	Portman	WCMIX	WCM Focused International Growth Fund Institutiona	2019-07-11	1,001	15,000	[NULL]
Christopher	Murphy	[NULL]	Aggressive Managed Allocation Age 4-7	2015-12-21	1,001	15,000	[NULL]
A. Mitchell	McConnell, Jr.	PRSCX	T. Rowe Price Science & Tech	2017-12-15	1,001	15,000	[NULL]
Rick	Scott	SHV	iShares Short Treasury Bond ETF	2019-03-21	250,001	500,000	110.4938

Figure 1: **Senator's Stock Transactions Data (Compiled)** The table shows the compiled stock transactions data which includes the name, ticker, date, amount min/max, and VWAP for each transaction.

limitations in the analysis, but the dataset still provides a rich source of information for understanding potential insider trading among U.S. members of Congress. An excerpt of a few rows of such compiled transaction data is provided in Fig 1.

2.2 Uncanny Timing of Congressional Stock Trading

Firstly, I gained insights into the mechanisms behind their trading decisions by reviewing news articles. For example, there were several media reports[4,5] regarding Ron Wyden's semiconductor stocks trading. I searched for Senator Ron Wyden's stock transactions with a NAICS code beginning with 334, which indicates computer and electronic product manufacturing. I found that three different tickers (AMAT, AVGO, KLAC) of the transactions that met this condition have a commonality in that they all started on the same date, April 6th, 2020, and ended on either April 6th or April 16th, 2021. Furthermore, all

[4] Theo Wayt, "US Sen. Ron Wyden boosts chipmakers while his wife buys their shares", New York Post, May 20, 2021, https://nypost.com/2021/05/20/us-sen-ron-wyden-boosts-chipmakers-while-his-wife-buys-their-shares/

[5] Alicia Parlapiano, Adam Playford, and Kate Kelly, "These 97 Members of Congress Reported Trades in Companies Influenced by Their Committees", The New York Times, Sept. 13, 2022, https://www.nytimes.com/interactive/2022/09/13/us/politics/congress-members-stock-trading-list.html

of them follow a similar pattern of multiple purchases followed by sales after certain critical points, such as $Purchase - Purchase - \ldots - Purchase \mid Sales - Sales - \ldots - Sales$ as shown in Fig 2.

	fi	la	ticker	trans_type	trans_date	trans_type	amount_min	amount_max
1	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
2	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
3	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
4	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
5	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
6	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
7	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
8	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
9	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
10	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
11	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
12	Ron L	Wyden	AVGO	Purchase	2020-06-04	Purchase	15,001	50,000
13	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
14	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
15	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
16	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
17	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
18	Ron L	Wyden	AVGO	Purchase	2020-06-23	Purchase	15,001	50,000
19	Ron L	Wyden	AVGO	Purchase	2021-03-04	Purchase	1,001	15,000
20	Ron L	Wyden	AVGO	Purchase	2021-03-04	Purchase	1,001	15,000
21	Ron L	Wyden	AVGO	Purchase	2021-03-04	Purchase	1,001	15,000
22	Ron L	Wyden	AVGO	Purchase	2021-03-04	Purchase	1,001	15,000
23	Ron L	Wyden	AVGO	Purchase	2021-03-04	Purchase	1,001	15,000
24	Ron L	Wyden	AVGO	Sale (Partial)	2021-03-30	Sale (Partial)	1,001	15,000
25	Ron L	Wyden	AVGO	Sale (Partial)	2021-03-30	Sale (Partial)	1,001	15,000
26	Ron L	Wyden	AVGO	Sale (Partial)	2021-03-30	Sale (Partial)	1,001	15,000
27	Ron L	Wyden	AVGO	Sale (Partial)	2021-03-30	Sale (Partial)	1,001	15,000
28	Ron L	Wyden	AVGO	Sale (Partial)	2021-03-30	Sale (Partial)	1,001	15,000
29	Ron L	Wyden	AVGO	Sale (Full)	2021-04-06	Sale (Full)	100,001	250,000
30	Ron L	Wyden	AVGO	Sale (Full)	2021-04-06	Sale (Full)	100,001	250,000
31	Ron L	Wyden	AVGO	Sale (Full)	2021-04-06	Sale (Full)	100,001	250,000
32	Ron L	Wyden	AVGO	Sale (Full)	2021-04-06	Sale (Full)	100,001	250,000
33	Ron L	Wyden	AVGO	Sale (Full)	2021-04-06	Sale (Full)	100,001	250,000

Figure 2: **Senator Ron Wyden's stock transactions for Broadcom Inc. (ticker: AVGO)** The transactions exhibits a pattern of multiple purchases followed by sales after certain critical points, spanning from April 6th, 2020 to April 6th, 2021.

On April 1st, 2021, President Biden announced a plan to invest $50 billion to boost the U.S. chip industry[6]. After this announcement, Senator Ron Wyden sold all of his semiconductor stocks. This suggests that members of Congress may have access to not only legislative information but also the publicization of such information that can potentially move the stock market beforehand. This enables them to not just design their portfolio, but also determine when to buy and when to sell, with some anticipation of specific events and their impact on the market.

[6] Alex Leary and Paul Ziobro, "Biden Calls for $50 Billion to Boost U.S. Chip Industry", The Wall Street Journal, March 31, 2021, `https://www.wsj.com/articles/biden-urges-50-billion-to-boost-chip-manufacturing-in-u-s-11617211570`

2.3 Sub-sequences of Congressional Stock Transactions

Based on the observation introduced in Section 2.2, I partitioned each transaction sequence into sub-sequences, where each sub-sequence consists of consecutive purchase transactions followed by consecutive sale transactions, all arranged in chronological order as illustratively shown in Fig 3.

This kind of sub-sequence partitioning is based on the assumption that if congressional investments involve insider trading—using privileged knowledge—there should be a timing of both the beginning and end of the investment that is driven by a certain event (Cziraki et al., 2021; Sivakumar and Waymire, 1994). Specifically, the event of interest would be one that, upon being publicized, moves the stock market into a different phase. In the case of insider trading, we would expect to see a pattern where a congressperson accumulates a long position in a stock ahead of a positive event and subsequently monetizes that position by selling the stock after the event becomes public and positively impacts the stock price. Conversely, a congressperson may sell a stock ahead of a negative event and avoid losses when the event becomes public and negatively impacts the stock price.

It is important to note that in this analysis, we are only considering the case of congresspersons taking long positions and subsequently selling those positions, as this is the type of transaction that is reported in Financial Disclosure reports. In these reports, there are no "stock-shorting" transactions, which involve betting against a stock and profiting from its decline. As such, our partitioning approach focuses on identifying sub-sequences of consecutive purchases followed by consecutive sales, which may reflect the use of privileged knowledge to take advantage of market-moving events and realize profits from long positions.

Through the partitioning process, I obtained a total of 435 sub-sequences spanning across 358 unique combinations of Senator-Ticker pairs. Each sub-sequence represents a long position taken by a senator in a specific stock and is characterized by a start date and an end date. The start date corresponds to the date of the first purchase transaction in the sub-sequence, and the end date corresponds to the date of the last sale transaction in the sub-sequence. The duration of each sub-sequence, measured in days, represents the length of time the senator held the long position. The frequencies of durations for all subchains are illustrated in Figure 4.

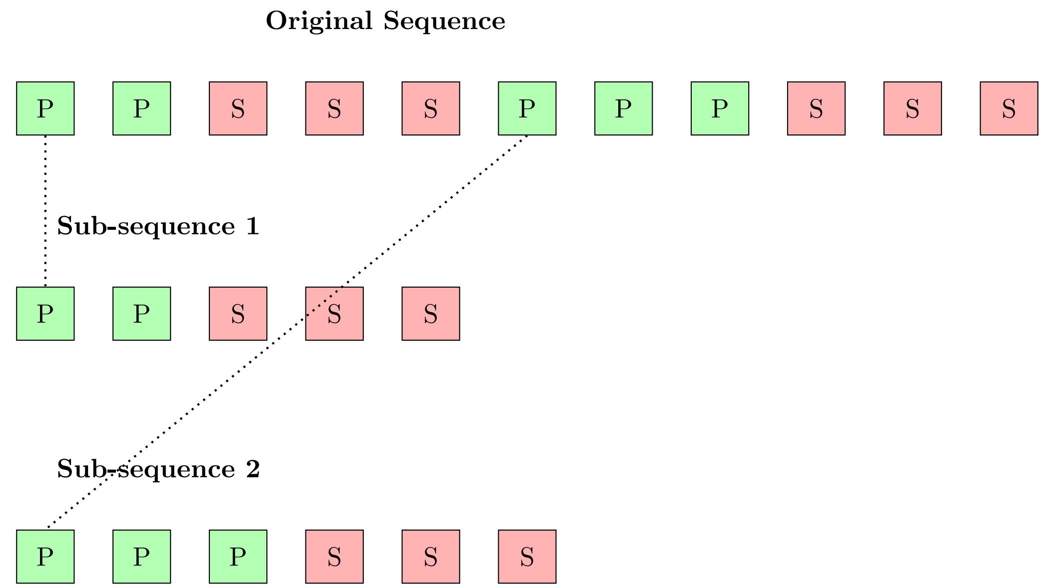


Figure 3: Partitioning of transaction sequences into sub-sequences based on consecutive purchase and sale transactions.

2.4 Estimating Excess Returns of Sub-sequences

Estimating the excess return of the 435 sub-sequences, which were acquired following the procedure described in Section 2.3, presented a methodological challenge due to the nature of the Finance Disclosure data. The data provides only the minimum and maximum range of amounts spent on purchasing or selling each ticker on a specific day (as shown in Table 1), rather than exact transaction amounts. To address this challenge and estimate the excess returns for each Purchase-Sale sub-sequence, the following approach was taken:

1. **Random Sampling of Transaction Amounts**: For each transaction (purchase or sale) within a sub-sequence, an amount was randomly sampled from a uniform distribution with support equal to the minimum and maximum range of the transaction amount provided in the data.

2. **Estimation of Shares Bought or Sold**: The sampled amount was divided by the volume-weighted average price (VWAP) of the stock on the corresponding transaction date to estimate the number of shares bought or sold by the congressperson.

3. **Creating Settled Pairs**: Within each sub-sequence, settled pairs of buy-sell transactions are identified. A settled pair consists of one unit of a buy transaction matched with one unit of a subsequent sell transaction. The pairing process is based on a first-in/first-out principle, meaning that stocks purchased earlier are matched first to sales, before those bought later. This ensures that the sale always occurs after the purchase. Multiple settled pairs can be created within a single sub-sequence.

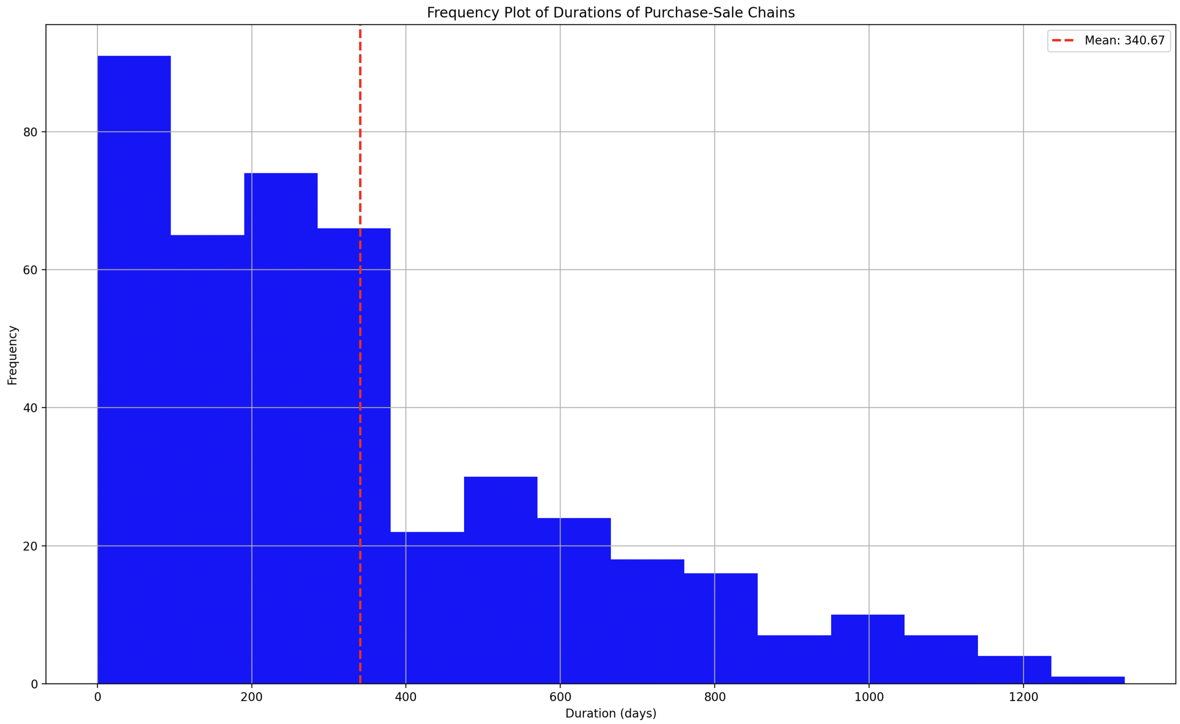


Figure 4: **Frequencies of durations of long positions held by congressmen at the Senator-Ticker level** The durations are measured in days and represent the length of time between the start and end dates of each subchain. The mean duration of holding such long positions is approximately 340 days. Notably, around 65% of these long positions are held for less than a year

4. **Computing Profit Return Rate for Each Settled Pair**: For each settled pair, the profit return rate is calculated as the relative profit or loss from the buy-sell transaction. The profit return rate is computed using the formula:

$$\text{Profit Return Rate} = \frac{\text{Sale Price} - \text{Purchase Price}}{\text{Buy Price}} * 100$$

where "Sale Price" is the price at which the stock was sold, and "Purchase Price" is the price at which the stock was purchased.

5. **Penalizing the Profit Return Rate**: The profit return rate for each settled pair is then penalized by the average Federal Reserve Rate during the holding period of that specific pair. The holding period is defined as the time interval between the purchase date and the sale date of the settled pair. The penalized return, or "excess return" for each pair is calculated as:

$$\text{Excess Return } = \text{Profit Return Rate} - \text{Average Federal Reserve Rate}$$

The Average Federal Reserve Rate represents the risk-free rate of return that could have been earned from a risk-free investment during the same holding period.

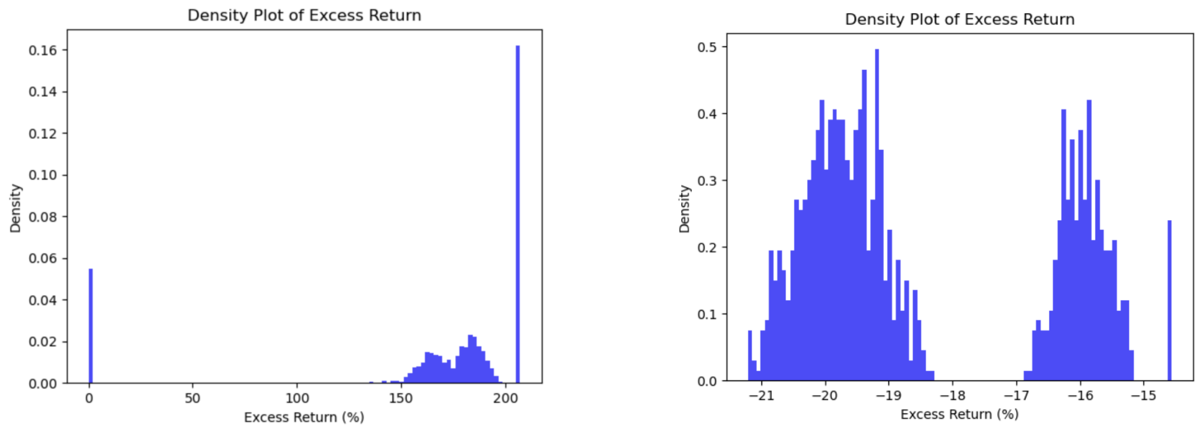
6. **Averaging Excess Returns**: The final excess return for the entire sub-sequence is computed by averaging the excess returns of all the individual pairs of settled buy-sell transactions within the sub-sequence. This approach provides a comprehensive measure of the excess return for the sub-sequence, accounting for profit ratio above the risk-free rate of return for each holding period.

By following this approach, the excess return of each sub-sequence was estimated while accounting for the limitations of the available data. Using the Federal Reserve Rate as the baseline for comparison is a more conservative approach because it represents a risk-free rate of return (Bauer and Rudebusch, 2014; Sarno and Thornton, 2003) that is typically higher than the interest rates offered by most savings accounts. As a result, the excess return is measured against a higher baseline, potentially lowering the final estimation result and providing a more cautious assessment of the excess return earned by the congressperson.

To assist in understanding the concepts explained previously, I am presenting the estimated excess return distributions for Senator Ron Wyden's sub-sequences involving two different companies: Applied Materials Inc. (AMAT), which provides manufacturing equipment, services, and software to the semiconductor industry, and Marriott International Inc. (MAR), a global hotel brand. These distributions were computed using the random sampling method explained earlier, where the randomness is inherited from the uniform random sampling of transaction amounts from the provided minimum and maximum ranges. It is important to note that each sub-sequence is uniquely identified not only by the congressperson-ticker level but also by the start and end dates of the sub-sequence.

In Figure 5, the mean of the excess return distributions for Senator Ron Wyden's sub-sequences involving AMAT and MAR are 166.30% and -18.43%, respectively. As we can see, even the same senator sometimes achieves great excess returns while also experiencing failures. In a similar vein, I collected the mean values of the excess return distributions for all 435 subsequences, which are presented in Figure 6.

In Figure 6, among the mean estimated excess returns, Ron Wyden's semiconductor-related stocks like Applied Materials Inc (AMAT), KLA Corp. (KLAC), and Broadcom Inc. (AVGO) are highly ranked, scoring from 80 to 166% of excess returns. These transactions have already been spotlighted by the media, as introduced in Section 2.2. This suggests that this method can reveal such dubious transactions spotlighted by the media, ranking them as acquiring high-performing excess returns.



(a) Ron Wyden's excess returns from transactions involving Applied Materials Inc. (AMAT) from April 2020 to April 2021.

(b) Ron Wyden's excess returns from transactions involving Marriott International Inc. (MAR) from May to August 2020.

Figure 5: Estimated Excess return distributions of Senator Ron Wyden's transactions for AMAT and MAR

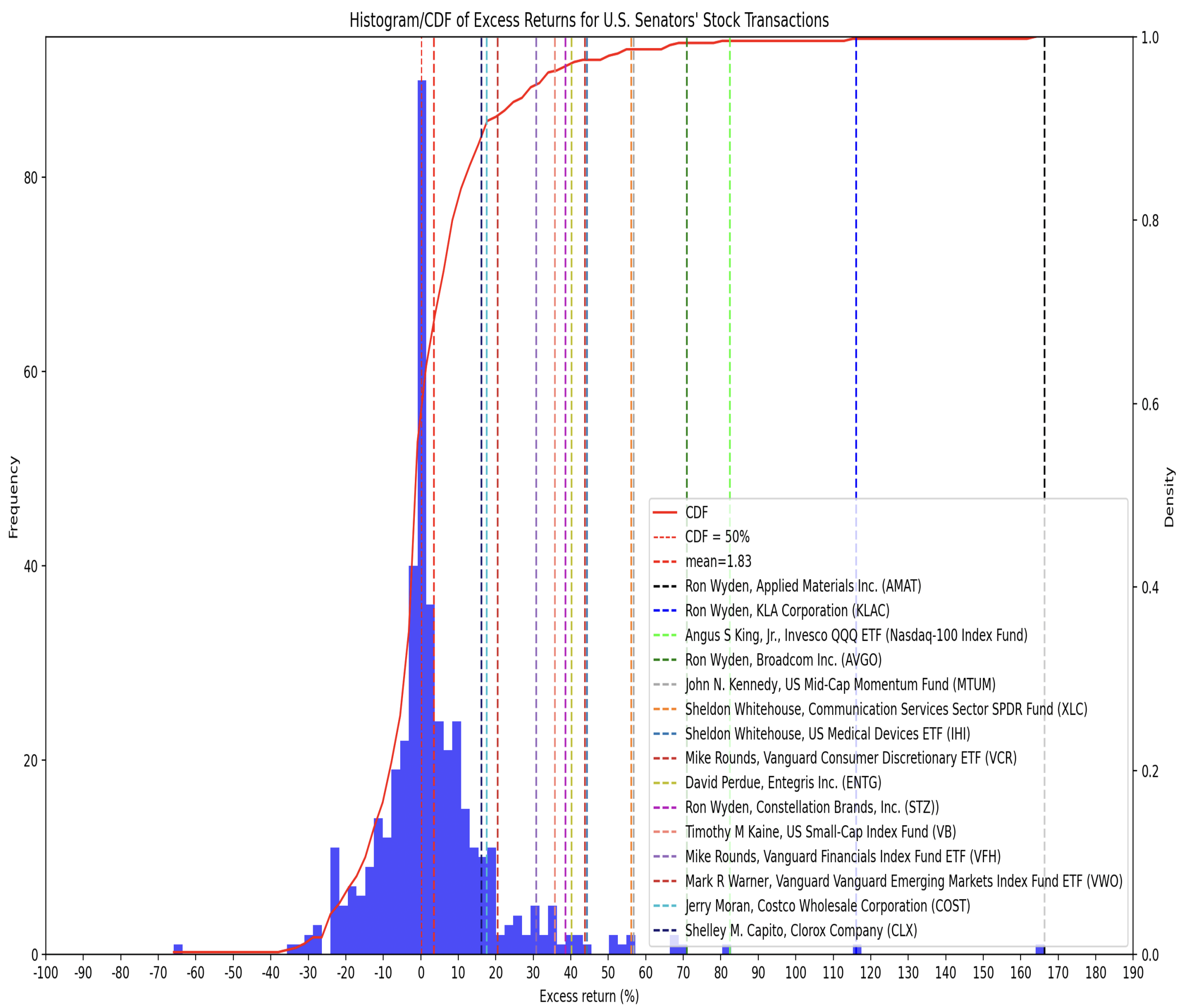


Figure 6: Distribution of Senators' Mean of Estimated Excess Returns Distribution.

What is noticeable is that the cumulative density of the excess return distribution reaches 0.5 when the excess return is 0. It means that not all transactions of Senators are always successful, but it's more like random whether they actually acquire good excess profit. This finding aligns more with the results from Eggers and Hainmueller (2013) than those of Ziobrowski et al. (2004) and Ziobrowski et al. (2011), suggesting that Senators are more like mediocre investors who align with the literature about failing individuals as investors, as explained in broad finance literature such as Barberis and Thaler (2003), and (Barberis and Thaler, 2003).

However, one thing that is important to notice is that the distribution in Figure 6 scores a skewness of 2.804, which means the tail on the right side of the distribution is longer. This indicates that compared to the case of Senators losing money, at least in situations where they are acquiring excess return from it, they may be more related to some privileged knowledge that can back up the performance of such stock transactions. The presence of the long tail suggests that, in some cases, Senators might be involved in transactions that benefit from privileged knowledge, as Eggers and Hainmueller (2014) suggest, which could be originating from their political connections, including connections established through PAC (Political Action Committee) donations or district-level affiliations with the firms in question.

While the current findings provide valuable insights into the nature of congressional stock trading, it is essential to delve deeper into the fundamental relationship between congressional activities and stock transactions. In order to gain a better understanding of the underlying factors that drive congressional stock trading behavior and the potential impact of political connections and privileged information on investment decisions among lawmakers, we need to explore an approach that quantifies the predictive power embedded in the congressional activities and assesses the extent to which these activities are associated with stock transactions.

It is worth noting that this approach is not entirely novel, as Eggers and Hainmueller (2014) have already studied the connections between congressional-related activities and the specific firm's stock transactions of a congressperson. Their research examined factors such as PAC contributions, lobbying, geographical connections based on district, and congressperson's committee membership and firm-level lobbying, to determine whether these factors could predict a congressperson's stock transactions.

Given that the innate difficulty in identifying the intention behind such transactions (Ziobrowski et al., 2011, 2004; Eggers and Hainmueller, 2013, 2014), due to the limited information available or the more fundamental challenge of distinguishing between a congressperson's private and public life (Buchanan and Tollison, 1984), it is essential to examine how dynamically these transactions are connected with congressional activities in terms of information.

One possible avenue for this exploration is to represent congressional activities as a graph, which can effectively capture the complex relationships between various actors and actions (Henaff et al., 2015; Kaushik et al., 2002) within the legislative process. Graph-based representations are well-suited for modeling the interconnected nature of congressional activities, taking into account not only the individual actions of Congress members but also the broader context of committee memberships, lobbying efforts by firms on specific bills of their interest, the referral of bills to specific committees, and the assignment of Congresspersons to certain committees.

Eggers and Hainmueller (2014) studied the impact of such factors, particularly committee membership and firm-level lobbying on bills, on stock transactions at a binary level, considering whether or not this information existed for each congressperson-stock pair. However, congressional activities are more complex and interconnected, with various entities involved in these relationships simultaneously rather than in isolation. For example, multiple firms in the semiconductor industry, such as Intel, Qualcomm, Broadcom, Apple, and IBM, participate in lobbying efforts for bills related to their sector, like the CHIPS Act or FABs Bills. These activities are governed by specific congresspeople within particular committees, and all this information collectively can forms the detailed context in which a congressperson transacts stock.

In addition, as explained in Section 2.1, a congressperson's securities transactions are not limited to individual firm levels. In fact, 60% of these transactions involve exchange-traded funds (ETF) or mutual funds that target a wide range of specific industries such as wireless communication, medical devices, or mid-cap or small-cap companies. Therefore, the full context of a congressperson's stock transactions extends beyond individual companies to encompass broader industry trends and movements.

In light of this, the next section will introduce a newly compiled dataset that captures congressional activities as a whole, in the form of graph-structured data. I will then demonstrate how this graph-structured data can be useful, for example, by directly computing the similarity between a committee's industry-level specialization and the industry-level distribution of a congressperson assigned to that committee in Section 4. Additionally, I will present a method for modeling the predictive task, which can directly take graph-structured data as input using Graph Neural Networks in Section 5. An array of analyses using graph-structured data will enhance our comprehension of the intricate relationships between congressional activities and stock transactions. This approach will offer more profound insights into the potential influence of political connections and privileged information on lawmakers' investment decisions.

3 Graph-Structured Data for Representing Congressional Activities

[7] The data utilized in the following sections is a large graph-structured dataset that captures both congressional activities. The graph-structured data encompasses information on congressional activities, such as committee assignments, bills being lobbied by firms, bill assignments to committees, and firms classified under specific NAICS codes. The detailed specifications of the node types can be found in Table 2, while the edge types are described in Table 3. Different types of nodes and their relationships, captured by different types of edges, are provided in Figure 7.

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	-	-

To provide a more concrete understanding of the data, Figure 8 displays a subgraph related to Senator Ron Wyden's transaction in Trip Advisor stock (Ticker: TRIP). This subgraph illustrates the relationships between Senator Ron Wyden's congressional activities, including his membership in the Senate Finance Committee, his involvement with a specific bill related to airport improvements, and the

[7]Reproducible code for this section is available at `https://github.com/syyunn/gnnex/blob/main/data/graph.ipynb`

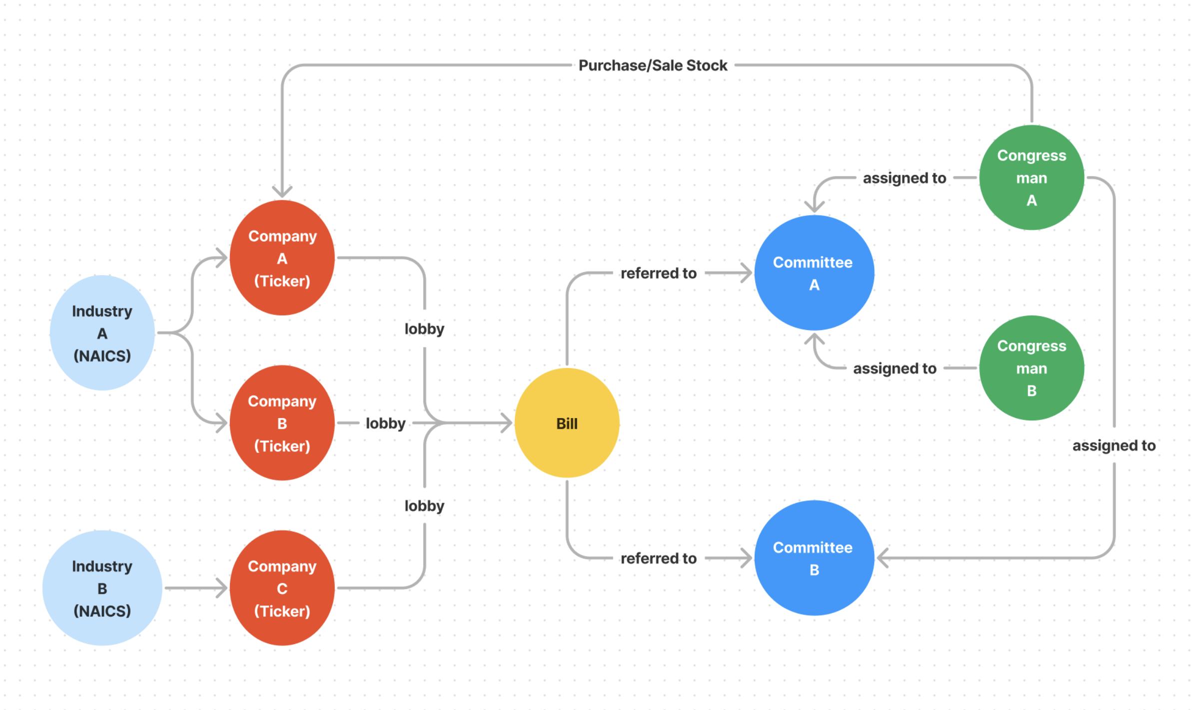


Figure 7: The network data includes various types of nodes and edges that represent different entities and interactions within the congressional activities and investment behavior of Congresspersons.

economic sectors represented by NAICS codes, thereby providing insights into how these activities could potentially influence or be influenced by his stock transactions.

In addition to the subgraph related to Senator Wyden's transactions, I also provide two more subgraphs in Figure 9 and Figure 10 to further aid understanding. These figures capture the industry-level congressional activities, illustrating how firms project their interests through lobbying to specific bills, and how committees oversee such bills. These subgraphs collectively demonstrate how firm/industry-level specific interests are funneled through lobbying to bills and aggregated to specific committees, ultimately conveying this information to congresspersons assigned to those committees. By examining these interconnections, we can gain a deeper understanding of the potential influences on stock transactions made by members of Congress.

3.1 Data Merging and Entity Disambiguation

One of the key challenges to make this graph-structured data is the effective disambiguation of entities, as the data is collected from multiple sources, including LobbyView, Senate/House Financial Disclosures, and naics.com. In this graph-structured dataset, entities such as Congresspersons and firms may appear under different names or expressions. For example, "Ron Wyden" may also be referred to as "Ron L.

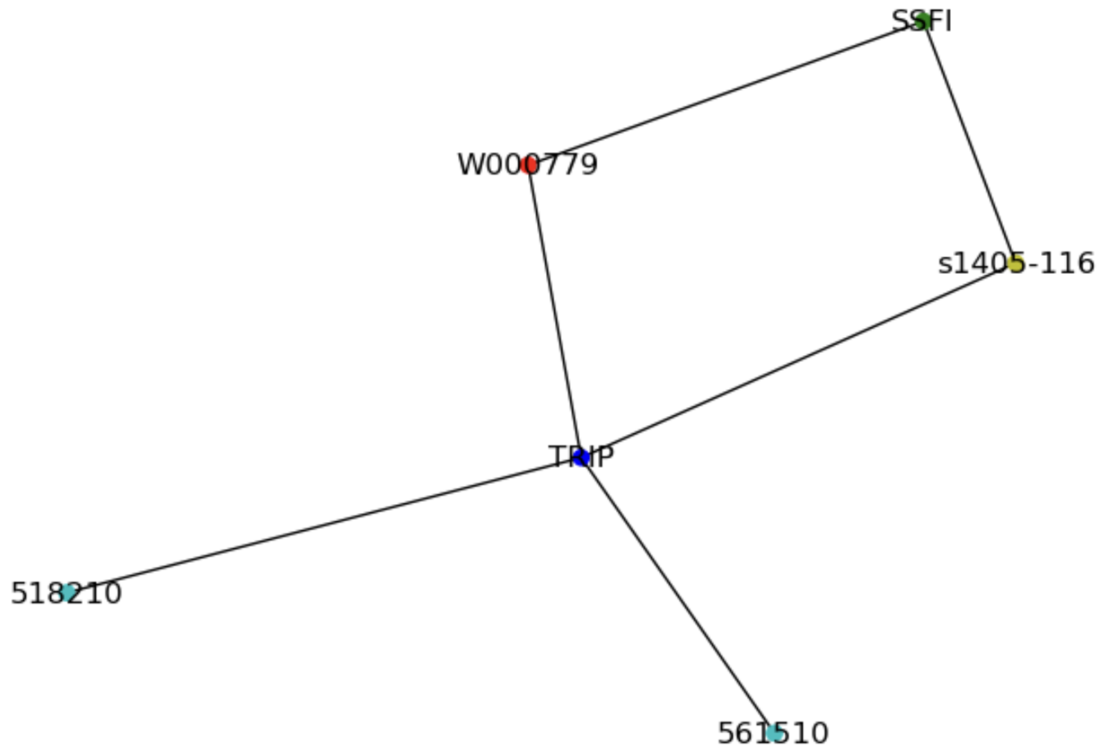


Figure 8: **A subgraph illustrating the congressional network related to the transaction of Senator Ron Wyden's Trip Advisor stock.** The node labeled W000779 corresponds to Ron Wyden's bioguide-id, which is a unique identifier provided by Congress for each senator. SSFI represents the Senate Finance Committee, of which Ron Wyden is a member. S1405-116 is a bill in the 116th Congress that revises requirements for the airport improvement program and pilot program for passenger facility charges at nonhub airports. The node labeled 518210 represents the NAICS Code for Data Processing, Hosting, and Related Services, while 561510 represents Travel Agencies.

Wyden", and "Apple" may appear as "Apple Inc.". To accurately disambiguate these differing text representations of entities, it is essential to establish a unique identifier for each entity, regardless of the variations in their names.

Theoretically, matching entities based on text similarity between two datasets with n and m rows has a computational complexity of $O(nm)$ (Román et al., 2012). Therefore, as the datasets grow larger, this complexity becomes prohibitively expensive. For instance, matching 70,000 firm names from LobbyView to 4,000 firm names appearing in the ticker table would require 280,000,000 times of computations for text similarity. To address this challenge, I developed a novel approach that leverages URLs as unique identifiers for entities.

The approach involves acquiring the corresponding URL for each entity through Google searches, such as `https://en.wikipedia.org/wiki/Ron_Wyden` for Ron Wyden and `https://www.apple.com/`

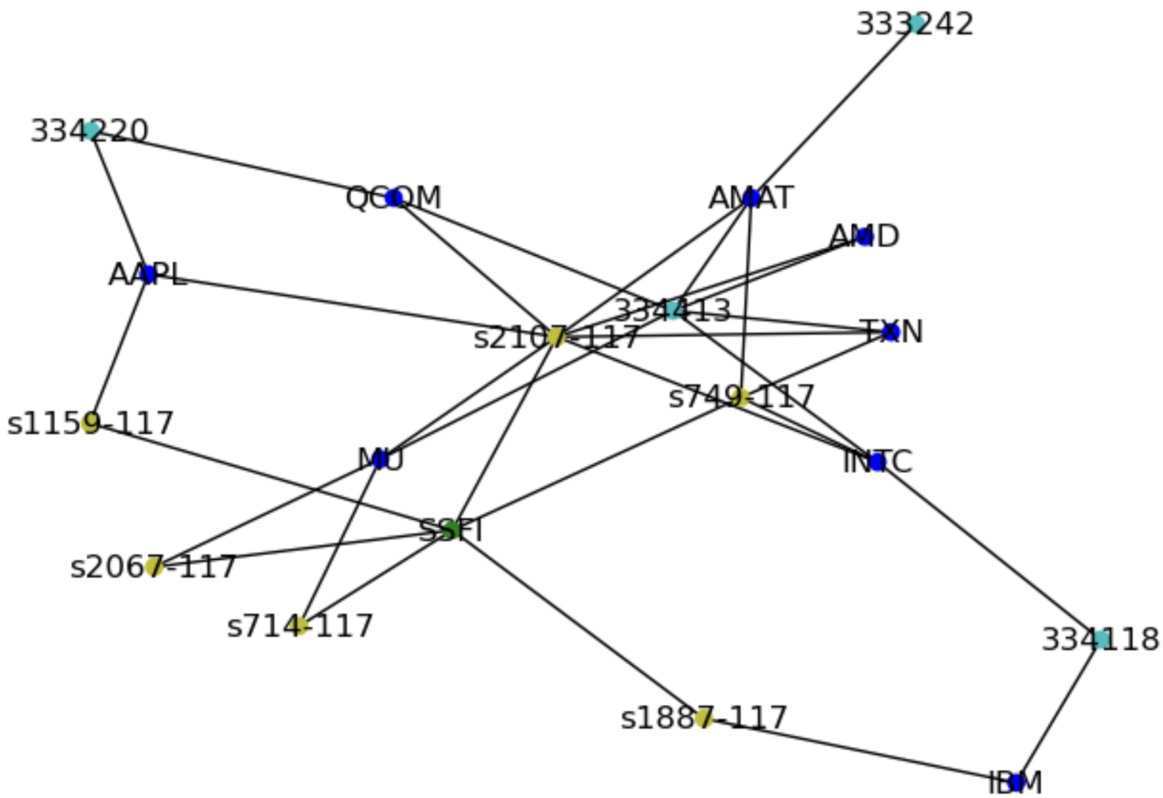


Figure 9: **Subgraph capturing how the semiconductor industry's interests were funneled through the Senate Finance Committee during the 117th Congress** The NAICS code 334413 indicates Semiconductor and Related Device Manufacturing, which involves companies such as Qualcomm (QCOM), Intel (INTC), and Advanced Micro Devices (AMD), lobbying for bills such as S.2107 FABS Act and S.749 American Innovation and Jobs Act that are closely related to the subsidization of semiconductor manufacturing facilities. Relevant companies such as Apple Inc. and IBM, with NAICS codes of 334220 Wireless Communications Equipment Manufacturing and 334118 Computer Equipment Manufacturing, respectively, are direct customers of these semiconductor chips for manufacturing smartphone and computer hardware. The bills in which these companies have an interest are assigned to the Senate Finance Committee as well.

for Apple, Inc. A key advantage of using URLs as unique identifiers is that they facilitate effective entity disambiguation. For example, if two different expressions, "Ron Wyden" and "Ron L. Wyden" are both assigned the same URL `https://en.wikipedia.org/wiki/Ron_Wyden`, we can confidently recognize that these two expressions refer to the same entity. This approach allows us to accurately consolidate information about entities that may be represented in various ways across different data sources. Additionally, this method reduces the computational complexity to $O(n+m)$, as only one query is required for each row of data. To further scale up this process, I parallelized the URL acquisition process by batching queries and distributing them across multiple servers available through commercial cloud services like AWS.

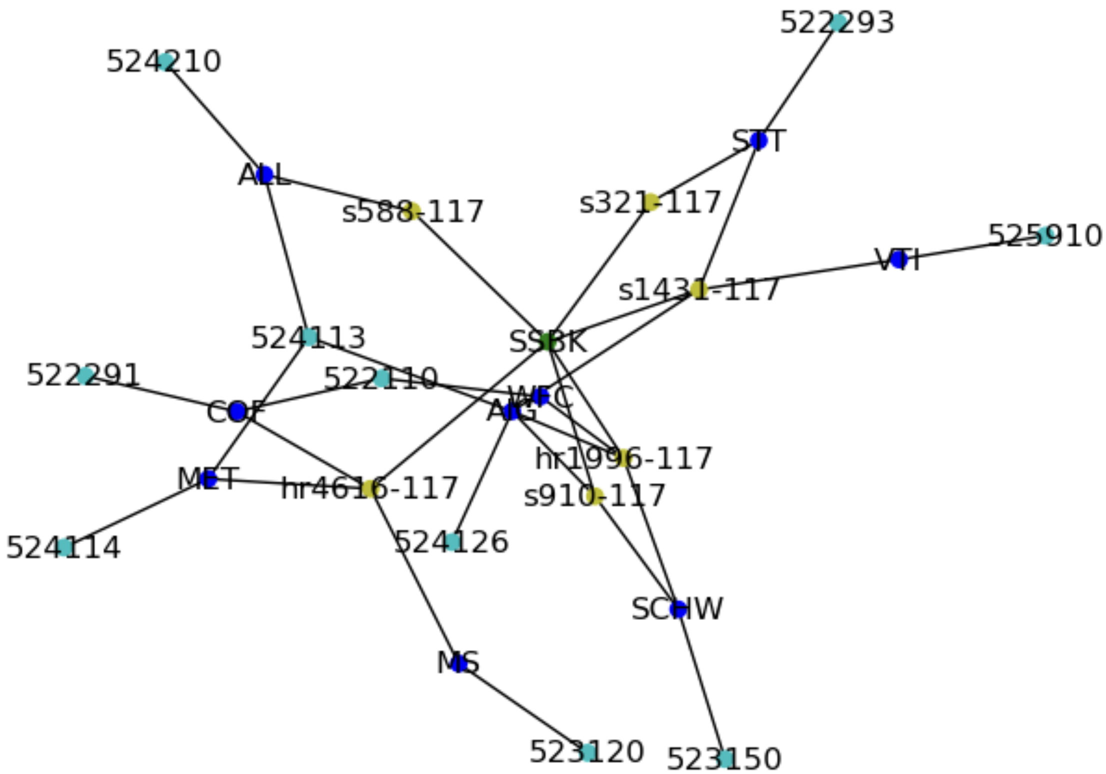


Figure 10: **Subgraph capturing how the financial industry's interests were funneled through the Senate Banking Committee during the 117th Congress** the Senate Banking Committee (SSBK) serves as a channel for different financial companies to project their lobbying interests over bills. NAICS codes starting from 52 generally relate to the financial industry in Figure 10. For instance, Wells Fargo (WFC) and AIG are lobbying for H.R.1996 - SAFE Banking Act of 2021, which prohibits a federal banking regulator from penalizing a depository institution for providing banking services to a legitimate cannabis-related business. Capital One Inc. (COF) and MetLife Inc. (MET) are lobbying for H.R.4616, which allows for the transition of certain financial contracts away from the London Interbank Offered Rate (LIBOR). These bills are all funneled through the Senate Banking Committee, thus a committee member of SSBK is more likely to be equipped with in-depth knowledge of the financial industry.

In summary, this innovative approach to entity disambiguation through URL acquisition and parallelization enables efficient data merging from diverse sources, ensuring the accuracy and scalability of the analysis.

3.2 Effective Parsing Technique for Financial Disclosures

Financial Disclosures from the House are provided as encrypted PDF files. While text can be extracted from these files, the encryption results in irregular patterns, particularly in the tables that contain information about Congresspersons' stock buying and selling activities. These irregular patterns make it challenging to parse the data using manually coded patterns, as the deviations are difficult to anticipate and account for. To address this challenge, I utilized OpenAI's APIs, specifically the GPT-3.5 Turbo language model, to parse the PDFs into a CSV format that includes information such as when and who bought or sold which ticker, and how much.

The process involves querying the Large Language Model (LLM) with the extracted text from the PDFs and instructing the model to convert the irregularly formatted tables into structured CSV data which includes columns such as the date of the transaction, the name of the Congressperson, the ticker symbol of the stock, the type of transaction (buy or sell), and the amount of the transaction.

By leveraging the capabilities of the GPT-3.5 Turbo language model, I was able to effectively parse information contained in PDF files that would normally require manual human labor. This approach significantly streamlines the data extraction process and ensures the accuracy and consistency of the parsed data.

4 Industry-level Similarities between Congresspersons and Committees

In this section[8], I aim to examine the relationship between committee assignments and stock trading behavior among congress members, addressing the puzzle arising from Eggers and Hainmueller (2014)'s findings. While they discovered that political connections played a role in profitable transactions, they found no evidence that committee membership influenced investment decisions. This observation contrasts with previous studies highlighting the importance of committee assignments and the specialized knowledge they confer, such as Patterson (1970) who emphasized the role of committee assignments in shaping legislative outcomes, King (1994) who focused on the impact of bill referral on committee specialization, and Asher (1974) who studied members' specialization in topics realted to committees' jurisdiction.

Committee assignments provide congresspersons with unique access to information and resources related to specific industries and policy areas. As members of these committees, they are privy to the latest policy developments (Price, 1978), market trends, and regulatory changes (Weiss, 1989) that could potentially affect the performance of companies in the industries they oversee. The specialized knowledge and insights gained from participating in committee activities could inform their investment decisions and

[8]Reproducible code for this section is available at `https://github.com/syyunn/efd/blob/main/anlys/cycle/main9-transactions-desc-house-included.ipynb`

potentially influence their stock trading behavior. For example, they might be more inclined to invest in companies within their committee's jurisdiction due to their deeper understanding of the industry dynamics and future prospects. Furthermore, their committee roles could enable them to establish connections with industry stakeholders and gain access to non-public information that could potentially offer them an edge in making investment decisions.

This section aims to provide more robust statistical evidence on the relationship between committee specialization and its potential implications on the stock trading behavior of Congress members by investigating the extent to which committee specialization resembles congresspeople's stock portfolio. I analyze the industry distributions of their stock transactions and committee assignments by computing the similarity between a committee's NAICS code distribution and that of the stocks transacted by the congressperson.

4.1 Measuring Industry-level Specialization

In this subsection, I will discuss the measurement of committee specialization in terms of industry-level specialization. By aggregating NAICS codes for the bills lobbied by various industries and those bills being referred to specific committees, I aim to quantify the industry-level specialization of each committee.

As an example, we can expect the Senate Banking Committee to have a higher degree of jurisdiction over banking-related issues, such as regulations tied to LIBOR (London Interbank Offered Rate) rates. This, in turn, would influence firms associated with NAICS codes related to banking (e.g., 52) to lobby the bills assigned to this committee more actively. This industry-level specialization can be effectively captured using the graph-structured data discussed in Section 3. Also, Figure 9 and Figure 10 from the previous section well illustrate industry-level specializations of Senate Fiinance and Banking committee for the semiconductor and banking industries, respectively.

To represent the committee-level specialization, I create a discrete probability distribution for each committee by aggregating the NAICS code distributions of firms that lobby bills assigned to those committees. Specifically, I count the occurrences of each NAICS code associated with firms that lobby bills referred to a particular committee. This method highlights the committee's industry focus, as it captures the concentration of lobbying efforts by firms within specific industries. Once we have the count-based frequency plot for each committee, we can easily convert it into a probability distribution function (PDF) by normalizing the counts with the total occurrences of all NAICS codes.

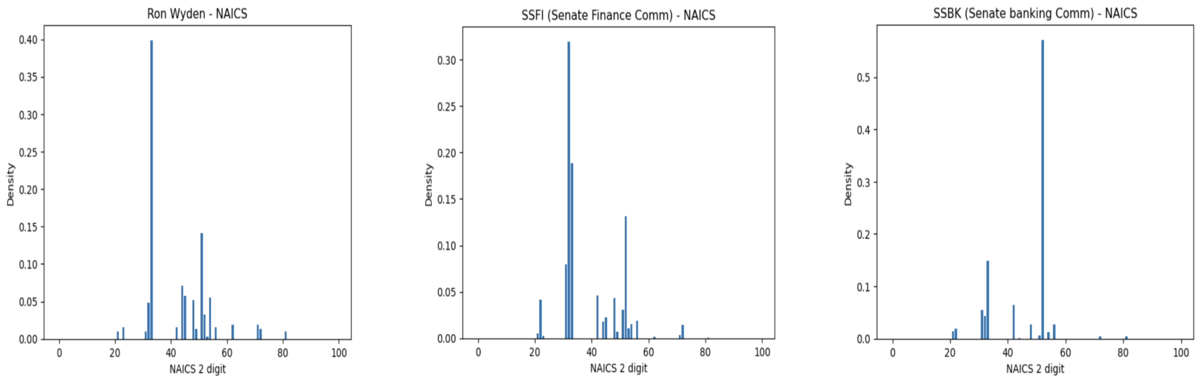
Similarly, to aggregate the NAICS code distribution of firms involved in each congressperson's stock transactions, we can simply count the occurrences of NAICS codes associated with all stock transactions executed by the congressperson (both purchases and sales). This provides a clear picture of the industries in which they invest, as the NAICS codes are derived from the firms whose stocks are being bought or sold by the congressperson in their stock transactions. We then normalize the count-based frequency plot with the total occurrences of all NAICS codes to obtain a PDF representing the congressperson's industry preferences.

As we've seen in Section 2.4, the data used for estimating Senator's excess return shows that around 60% of tickers are Exchange-Traded Funds (ETF) or mutual funds, which are not single firms but representing certain industries or the stock market in general. For example, in Figure 6, one of the highest excess returns, like the 45% profit shown by Sen. Sheldon Whitehouse, is the ticker IHI, which is an ETF that invests in the US medical device market. Sen. Sheldon Whitehouse supported the Biden plan to fully utilize the Defense Production Act, increase the supply of necessary medical equipment and supplies[9]. As shown, Senators do not always reflect their knowledge gained from congressional activities at the firm level but at the industry level. In either case, this measurement can effectively capture their level of specialization in certain industries in terms of their transaction patterns. By comparing the similarity between the two distributions, one from the committee and one from the congressperson, we can statistically test how similar they are to each other.

Figure 11 provides an example of this measurement for Sen. Ron Wyden's case. It displays the NAICS code distributions for Sen. Wyden's stock transactions, the Senate Finance Committee (SSFI), and the Senate Banking Committee (SSBK). The figure illustrates that the distribution of Sen. Wyden's stock transactions resembles the distribution of the Senate Finance Committee more closely than that of the Senate Banking Committee, highlighting the connection between his transactions and his committee assignments.

To measure the similarity between the two discrete probability distributions, we can use cross-entropy. Cross-entropy is a useful statistical tool for determining the similarity between two distributions (Wu et al., 2018; Mao et al., 2013), making it an ideal choice for comparing the distributions of NAICS codes for committee assignments and stock transactions. In Figure 11, we can calculate the cross-entropy between

[9] See `https://www.whitehouse.senate.gov/news/release/whitehouse-supports-biden-plan-to-fully-utilize-defense-production-act-increase-supply-of-necessary-medical-equipment-and-supplies`



(a) Sen. Ron Wyden's NAICS Distribution

(b) Senate Finance Committee (SSFI) NAICS Distribution

(c) Senate Banking Committee (SSBK) NAICS Distribution

Figure 11: Comparison of NAICS code distributions for Sen. Ron Wyden, Senate Finance Committee (SSFI), and Senate Banking Committee (SSBK). The figure illustrates how the distribution of Sen. Wyden's stock transactions resembles the Senate Finance Committee's distribution more closely than that of the Senate Banking Committee.

Sen. Wyden's stock transaction distribution and the distributions of Senate Finance Committee (SSFI) and Senate Banking Committee (SSBK). The results are as follows:

$$\text{Cross entropy (Ron Wyden, SSFI)} = 0.717$$

$$\text{Cross entropy (Ron Wyden, SSBK)} = 3.311$$

These values indicate that Sen. Wyden's investment portfolio is more similar to Senate Finance Committee in terms of NAICS code distribution, as a lower cross-entropy value represents a closer resemblance between the distributions. This reflects that Sen. Wyden's stock portfolio more closely resembles the industry distribution of his own committee, the Senate Finance Committee (SSFI), compared to the Senate Banking Committee (SSBK), to which he does not belong. Also, this suggests that the cross-entropy measure effectively captures the similarity between the industry-level specialization of a committee and the preferences reflected in a congressperson's stock portfolio.

4.2 Paired T-Test: Comparing Assigned and Unassigned Committees

In this subsection, I investigate whether there is a significant difference in the similarity between the industry distributions of Congress members' stock transactions and the industry distributions of their assigned and unassigned committees. To conduct this analysis, I computed the cross-entropy between the NAICS code distribution of stock transactions and committees for the 115th, 116th, and 117th

Congresses. I restricted the stock transaction dates to match each congressional term (e.g., for the 115th Congress, from January 2017 to January 2019) to ensure that only transactions during these periods were considered.

Senators and House Representatives are typically assigned to multiple committees for each congressional term. For each Congress member, I computed the average cross-entropy between their stock transactions and both their assigned and unassigned committees. I then tested the significance of the difference in mean cross-entropy values between the two groups using a paired t-test.

The paired t-test (Hsu and Lachenbruch, 2014) is appropriate in this context because the cross-entropy values within each pair (assigned and unassigned committees) are expected to be correlated. This correlation arises because committee assignments are based on a member's relative specialization in various topics and industries. However, the assumption of independence between pairs may not be fully justified due to the dense connections between Congress members and the potential for correlated behaviors.

The results of the paired t-test are presented in Figure 12, with a sample size of 235 unique pairs of (Congressperson, Congressional year). For example, one of the pairs includes Ron Wyden's average cross-entropy between the NAICS code distribution of his stock trading and assigned committees (Senate Finance, Budget, Intelligence, and Energy and Nuclear Resources Committee) and the remaining committees not assigned to him.

Despite the limitations of using a paired t-test, the results indicate that the average cross-entropy for the assigned group is significantly lower than that of the unassigned group. This suggests that the stock trading patterns of Congress members more closely resemble the industry distribution of their assigned committees compared to unassigned committees.

This result is directly opposite to the findings of Eggers and Hainmueller (2014), which conclude that "*...In contrast, we find no evidence that members disproportionately invest in companies to which they are connected through their committee assignments...*(p.4)". This result, on the other hand, suggests that committee assignments affect the congresspersons' decisions to invest disproportionately in industries connected to their committee assignments.

There could be several reasons for this discrepancy. For example, the data range is different: Eggers and Hainmueller (2014) consider the period from January 1, 2004, to December 31, 2008, while in this case, we consider the 115th, 116th, and 117th Congresses, which correspond to the years 2017 to 2022.

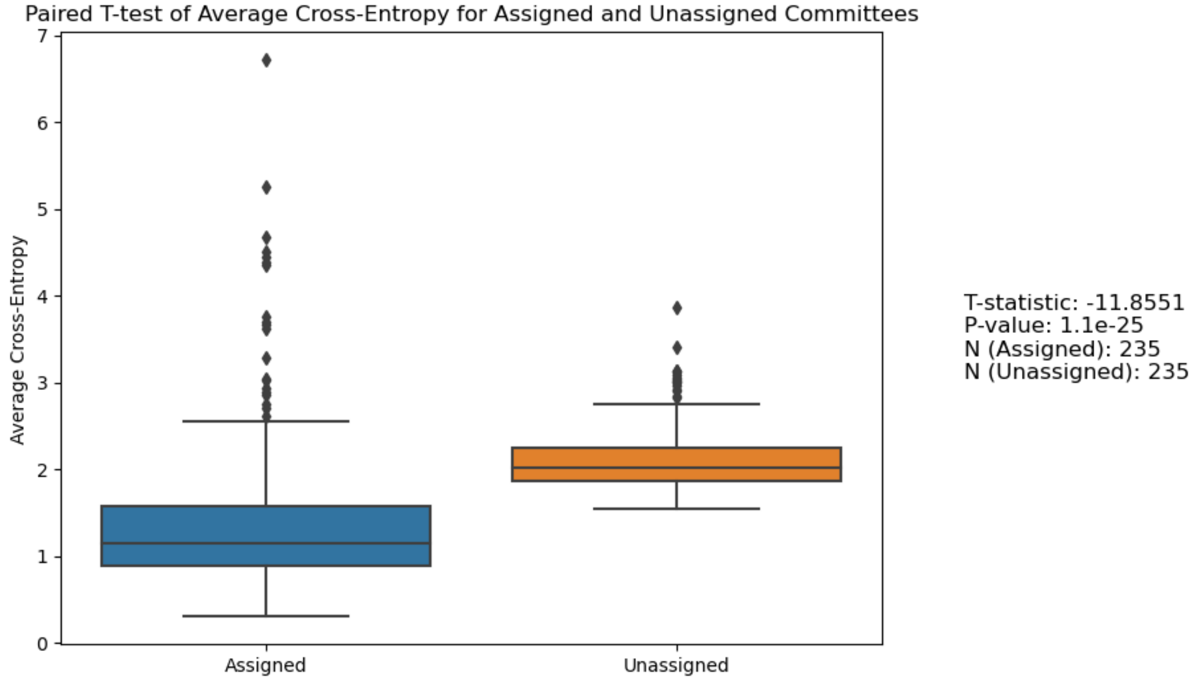


Figure 12: **Paired t-test results for the cross-entropy of assigned and unassigned committees.** The figure shows the comparison between the average cross-entropy of assigned and unassigned committees, with a sample size of 235 unique pairs of (Congressperson, Congressional year). The significantly lower average cross-entropy for the assigned group suggests that the stock trading patterns of Congress members more closely resemble the industry distribution of their assigned committees compared to unassigned committees.

However, more fundamentally, the difference may reside in the measurement approach.

Eggers and Hainmueller (2014) use a binary encoding to indicate whether a specific firm has engaged in lobbying behavior on a certain bill that is referred to a particular congressional committee. They then design a linear regression model to predict the weight of that specific firm's stock in a congressperson's entire investment portfolio. In this way, they investigate whether there is a relationship between a firm's lobbying activities on bills assigned to a committee and the investment decisions of congresspersons who are members of that committee. However, this approach only captures a specific company's lobbying behavior rather than industry-level information in its entirety. For example, multiple firms can lobby on the same bill assigned to a certain committee, then a congressperson on that committee can evaluate more broadly how this could impact the industry as a whole and selectively redesign their own portfolio. This means that there is no reason to assume that a congressperson would buy a specific stock that is being lobbied for — instead, it is more plausible to understand that such lobbying provides more detailed context about a specific industry, which the congressperson can utilize in their personal financial investment decisions.

Therefore, the cross-entropy approach, which directly measures the industry-level similarity between a committee's interests and a congressperson's portfolio, is a more intuitive and appropriate approach. This approach allows us to better understand the potential influence of committee assignments on the investment behavior of members of Congress, and it provides evidence that is complementary to the findings of previous studies.

However, another potential limitation of this analysis is the difficulty in collecting NAICS codes for ETFs and Mutual Funds, which comprise approximately 60% of the stock transactions in our dataset. Although each ETF or Mutual Fund's website provides information about their holdings, it was challenging to establish a generalizable pattern across different providers to obtain the composition of stocks held by these funds and their corresponding NAICS codes. Consequently, this section focuses on individual stocks and their corresponding NAICS codes, which might not fully capture the reality of Congress members' investments. However, the findings still indicate that the individual stock trading behaviors of Congress members are significantly influenced by their assigned committees.

5 Predicting Congressional Stock Transactions using Graph Neural Networks

In the previous section, I discussed the limitations of Eggers and Hainmueller (2014) linear prediction model that used a binary encoding of lobbying and committee assignments to predict the weight of a specific firm's stock in a congressperson's portfolio. Although I argued that this model was inappropriate because it failed to capture the complex interactions between different entities involved in congressional activities, I acknowledge that the model attempted to explain congressional stock transactions using possible explainable components such as district, PAC, lobbying, and committee assignments.

In this section[10], I propose to use a graph neural network (GNN) (Zhou et al., 2020; Wu et al., 2020; Scarselli et al., 2008; Zhang et al., 2019) to predict congressional stock transactions using the information embedded in the congressional activities captured in the data explained in Section 3. The GNN approach is well-suited to this task because it can capture the different types of relationships between different entities involved in the graph structured data (Zhao et al., 2021; Hong et al., 2020; Jin et al., 2021).

By leveraging a graph representation of the relationships between firms, bills, committees, and con-

[10]Reproducible code for this section is available at `https://github.com/syyunn/gnnex/blob/main/hetero/train_kfold_auto.py`

gresspersons, we can train a GNN to predict whether a congressperson is likely to buy a particular stock. This will enable us to investigate whether there is an associational relationship between congressional activities and stock transactions, and to identify the most relevant types of interactions between entities that contribute to the predictions.

While (Eggers and Hainmueller, 2014)'s linear prediction model attempted to explain congressional stock transactions using a limited set of explainable components, the GNN approach will enable us to capture the complexity of the relationships between different entities involved in congressional activities. This will provide new insights into the potential influence of congressional activities on stock transactions and complement the findings of previous studies.

5.1 Designing a Binary Classifier with Graph Neural Networks

To predict congressional stock transactions using a graph neural network (GNN) approach, I design a binary classifier that takes as input a graph G, a congressperson and a ticker (stock symbol). The classifier, denoted as $f(G, \text{congressperson}, \text{ticker})$, will output a binary prediction of either 0 or 1, indicating whether an edge (a buy or sell relationship) exists between the given congressperson and the ticker.

The hidden representations (Rauber et al., 2016; Das et al., 2020)o f the congressperson and the ticker, denoted as $h_{\text{congressperson}}$ and h_{ticker} respectively, are obtained as outputs of the GNN model. The main task in this approach is to train the GNN model to learn a computational graph that generates "good" representation of the congressperson and the ticker, $h_{congressperson}$ and h_{ticker}, which involves how to effectively encode the information embedded in the network to perform the downstream task of binary classification (Féraud and Clérot, 2002).

To design the classifier, a probabilistic modeling approach is used that comprises of a sigmoid function applied to the logit, which is the output of the model. The logit of the model is obtained by passing the representation learned by the GNN, $h_{congressperson}$ and h_{ticker}, to an MLP (Multi-layer perceptron) (Gardner and Dorling, 1998; Tang et al., 2016) that maps the representations of the congressperson and the ticker to a single logit. In other words, the MLP takes as input the representations of the congressperson and the ticker learned by the GNN, and outputs a logit that will be used to compute the probability of the existence of edge between them. MLP is simply an affine transformation over the concatenation of two representations, $h_{congressperson}$ and h_{ticker}, follwed by a non-linear activation function (Lu and Lu, 2020), which is ReLU (Agarap, 2018) in this case.

Formally, the logit of the classifier is defined as:

$$\begin{aligned}
\text{logit} &= \text{MLP}\left(h_{\text{congressperson}}, h_{\text{ticker}}\right) \\
\text{where } \text{MLP}(x) &= \text{ReLU}(Ax + b) \\
x &= \text{concat}\left(h_{\text{congressperson}}, h_{\text{ticker}}\right) \\
A &\in \mathbb{R}^{(d+d)\times 1} \\
b &\in \mathbb{R}^{1}
\end{aligned}$$

The sigmoid function is then applied to the logit to obtain a probability value:

$$\text{prob} = \sigma(\text{ logit })$$

where $\sigma(x)$ is the sigmoid function. The probability value indicates the likelihood of an edge existing between the given congressperson and the ticker. If the probability value is above a certain threshold, we predict that an edge exists between them, otherwise we predict that there is no edge.

Then remaining task is how to design a GNN model that can effectively learn the representations of the congressperson and the ticker, $h_{\text{congressperson}}$ and h_{ticker}, respectively, which can be used to train the classifier. In the following section, I will discuss the design of the GNN model.

5.2 Design of the Graph Neural Network

To obtain the representations $h_{\text{congressperson}}$ and h_{ticker}, I use a GNN approach that is designed to handle the complexity and dynamics of the congressional graph. The GNN approach is based on the idea of message passing and updating (Zhou et al., 2020; Wu et al., 2020), which is a process of aggregating information from the neighbors and updating the representation of each node accordingly.

In the case of the congressional graph, I use an edge-conditioned convolution GNN model (Gilmer et al., 2017; Simonovsky and Komodakis, 2017), which takes into account the edge attributes, such as the date, to better capture the complex relationships in the graph. The message passing, aggragation and updating in this model is defined as:

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

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the representations of nodes i and j, respectively, $\mathbf{e}_{i,j}$ is the edge attribute between nodes i and j, $\mathcal{N}(i)$ is the set of neighbors of node i, $\mathbf{\Theta}$ is a learnable matrix of size $d \times d$, where d is the dimension of the representation space, and $\mathbf{MLP}$ takes the edge attribute $\mathbf{e}_{i,j}$ as input and outputs a weight matrix of size $d \times d$. This weight matrix is then multiplied with the representation $\mathbf{h}_j$ of the neighbor node j to obtain a message $\mathbf{m}_{i,j} = \mathbf{MLP}\left(\mathbf{e}_{i,j}\right) \cdot \mathbf{h}_j$. In the updating step, the message from each neighbor node is aggregated by summing them up, and the resulting sum is added to the current representation $\mathbf{h}_i$ of node i multiplied by the learnable parameter matrix $\mathbf{\Theta}$ to obtain the updated representation $\mathbf{h}_i'$.

In the case of the congressional graph, the edge attribute $\mathbf{e}_{i,j}$ represents the relationship between nodes i and j at a specific date, which is represented as the elapsed time from a reference date (in this case, January 1, 2016). However, in our case, we have different types of edges, which means that $\mathbf{MLP}\left(\mathbf{e}_{i,j}\right)$ should be differently defined for different types of edges. This is because parsing the information of start and end dates should be considered differently across different edge types. For example, committee assignments of a congressperson that occurred over a specific congressional year should be considered differently from the date information that a certain firm lobbied on a certain bill. To account for this, I used the expanded version of the above formula:

$$\mathbf{h}_i^{(l+1)} = \mathbf{\Theta}^{(l)}\mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \mathrm{MLP}_k^{(l)}\left(\mathbf{e}_{i,j}^{(k)}\right) \cdot \mathbf{h}_j^{(l)}$$

Here, l represents a layer, and we can expand the expressivity of such message passing and updating process by stacking up the repeated layers of this operation. This allows the model to learn a more complex representation of each node, which is essential for capturing the intricate relationships in the congressional graph. Experimentally, I found that using 2 layers of message passing and updating was sufficient to learn the best representation of each node and used this configuration for the GNN model.

In conclusion, our GNN aims to learn the optimal parameter set that defines $\mathbf{\Theta}^{(l)}$ and $\mathrm{MLP}_k^{(l)}$ to output the best representations $\mathbf{h}_i^{(l)}$ and $\mathbf{h}_j^{(l)}$, which helps to perform the downstream task successfully. In this case, the downstream task is generating the best logit in the prediction head, $\mathrm{MLP}(\mathbf{h}_{\text{congressperson}}, \mathbf{h}_{\text{ticker}})$. It is important to note that the representations $\mathbf{h}_i^{(l)}$ and $\mathbf{h}_j^{(l)}$ are initialized randomly before they are provided into the first layer of the message passing and updating process. Through multiple rounds of message passing and updating, the GNN is tuned to output the best representation of each node that scores the best performance as possible in binary classification of edge existence.

5.3 Training & Evaluation of the GNN

5.3.1 Dataset Preparation

In the context of our GNN architecture, the goal is to predict the existence of edges between two nodes, a task commonly known as link prediction. To train the GNN for this task, the dataset must be prepared for training and evaluation (test). The dataset consists of a total of 24,675 edges, which represent the relationship (congressperson, buy-sell, ticker).

To create a balanced dataset for the link prediction task, the dataset is divided into a train and test set with an 8:2 ratio, resulting in 19,740 instances for training and 4,935 instances for testing. The network is then trained using the 19,740 instances and its performance is evaluated on the 4,935 test instances.

In addition, to ensure a balanced dataset, the same number of randomly sampled negative edges (Yang et al., 2020) is prepared. These negative edges are created by randomly selecting pairs of nodes (congressperson and ticker) that do not have a connection in the original dataset. This results in a total of 39,480 edges for training and 9,870 edges for testing. Including both positive and negative examples in the training process helps the model to better differentiate between true and false existence of edges between congressperson and tidcker nodes, improving its ability to predict links in the graph.

5.3.2 Training of the GNN

For the training of the GNN, a two-layer GNN architecture, as $l = 2$, is employed. Additionally, node embeddings h_i are represented as vectors in a 64-dimensional space ($h \in \mathbf{R}^{64}$). This hyperparameter is also set experimentally.

In order to measure the performance of the model during the training process, binary cross-entropy loss is used as the loss function. Binary cross-entropy loss is particularly suitable for binary classification problems (Ruby and Yendapalli, 2020), such as link prediction (Zhang and Chen, 2018), where the goal is to differentiate between the presence and absence of a connection between two nodes. This loss function quantifies the difference between the predicted probabilities and the true labels, and penalizes the model for incorrect predictions. Formally, the binary cross-entropy loss for a set of samples is defined as:

$$L = -\sum_{i=1}^{N} \left(y_i \cdot \log\left(p_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - p_i\right)\right)$$

where L represents the total binary cross-entropy loss. N is the total number of samples. y_i is the true label for the ith sample (1 for the presence of a connection, and 0 for the absence of a connection). p_i is the predicted probability of a connection existing between two nodes for the ith sample.

By minimizing the binary cross-entropy loss, the GNN learns to accurately predict the existence or non-existence of links in the network, ultimately improving its performance on the link prediction task. For the minimization, the Adam optimizer (Kingma and Ba, 2014) with stochastic gradient descent (SGD) is utilized. SGD is an iterative optimization algorithm that updates the model's parameters based on a random sample (or minibatch) of training data in each iteration (Amari, 1993). This approach helps in converging faster and reduces the impact of noisy gradients, thus improving the optimization process. Adam is an adaptive learning rate optimization algorithm, combining the advantages of two other popular optimization methods, AdaGrad and RMSProp (Kingma and Ba, 2014). This optimizer is well-suited for large-scale problems and is known for its ability to efficiently handle noisy and sparse gradients, making it a suitable choice for training GNNs.

To obtain a more robust estimation of the model's performance and uncertainty, a 5-fold cross-validation (Hastie et al., 2001) is performed. In this approach, the entire dataset is randomly split into five equal-sized chunks. For each fold, one chunk is used as the test set, while the remaining chunks are combined to form the training set. This process is repeated five times, with each chunk being used once as the test set. This technique allows for a better understanding of the model's performance across different subsets of the dataset and provides uncertainty statistics of overall prediction performance.

5.3.3 Evaluation & Ablation Study

In this study, we conducted a link-prediction (Zhang and Chen, 2018) task to predict the existence of an edge between a congressperson and a ticker, symbolizing the trade relationship - whether the given congressperson would sell or buy a particular stock. This task was performed using a variety of edge types, and the performance was evaluated using two metrics: accuracy and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

The results of this evaluation are depicted in Figures 13 and 14. With all edge types included, the model achieved an accuracy of approximately 81% and an AUC-ROC of 0.89. These results indicate that the model was generally effective at predicting the stock transactions of congresspersons.

To further understand the importance of each edge type, I conducted an ablation study, where I systematically removed each edge type from the training and testing data and observed the resulting performance drop. The most significant drop in performance was observed when the edge type ('congressperson', 'assignment', 'committee') was removed. This resulted in a decrease in accuracy from 81% to 67%, and a decrease in AUC-ROC from 0.89 to 0.76. This suggests that the ('congressperson', 'assignment', 'committee') edge type carries significant information for predicting a congressperson's stock transactions.

In comparison, the removal of other edge types, such as ('bill', 'assigned_to', 'committee'), or ('ticker', 'lobbies_on', 'bill'), resulted in less dramatic performance drops. This further underscores the relative importance of the ('congressperson', 'assignment', 'committee') edge type in this prediction task.

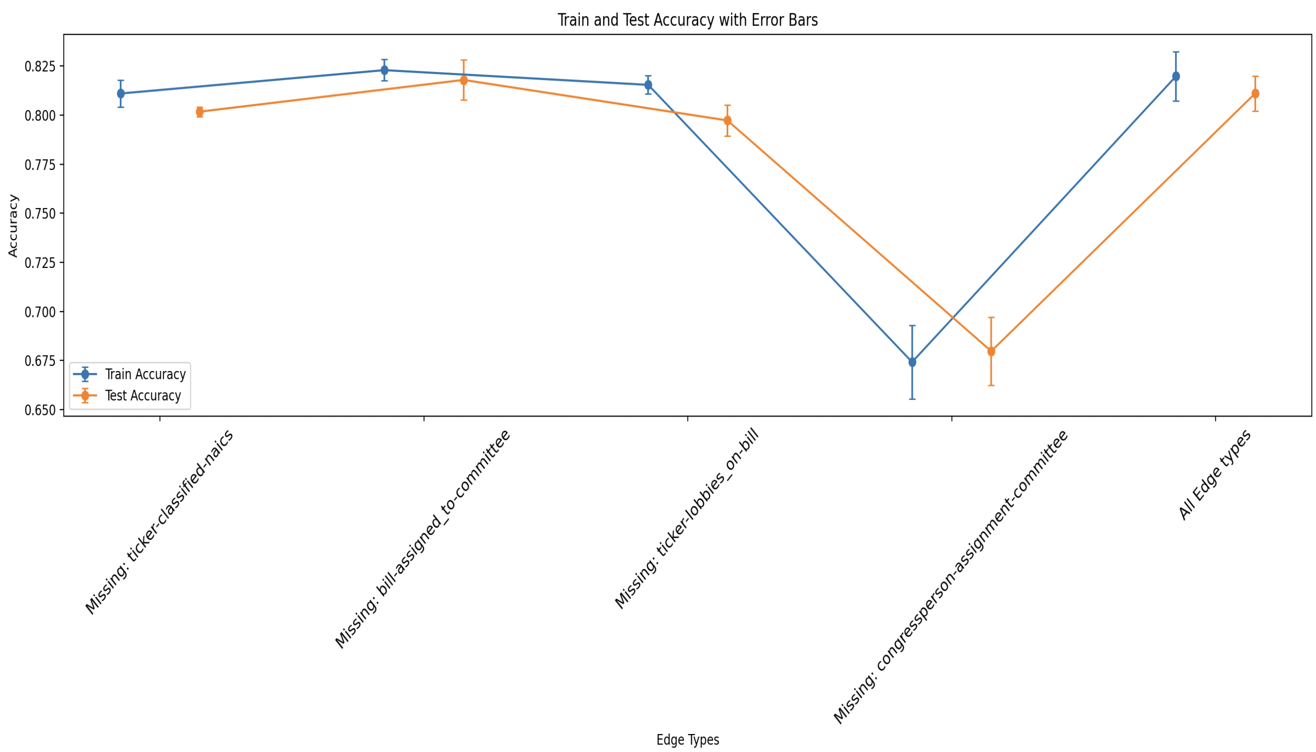


Figure 13: **Accuracy drop for different edge types.** The figure shows the accuracy of the model with all edge types included and with each edge type removed one at a time. With all edge types included, the model achieved an accuracy of approximately 81%. The most significant drop in accuracy, to 67%, was observed when the edge type ('congressperson', 'assignment', 'committee') was removed. This suggests that the ('congressperson', 'assignment', 'committee') edge type carries significant information for predicting a congressperson's stock transactions.

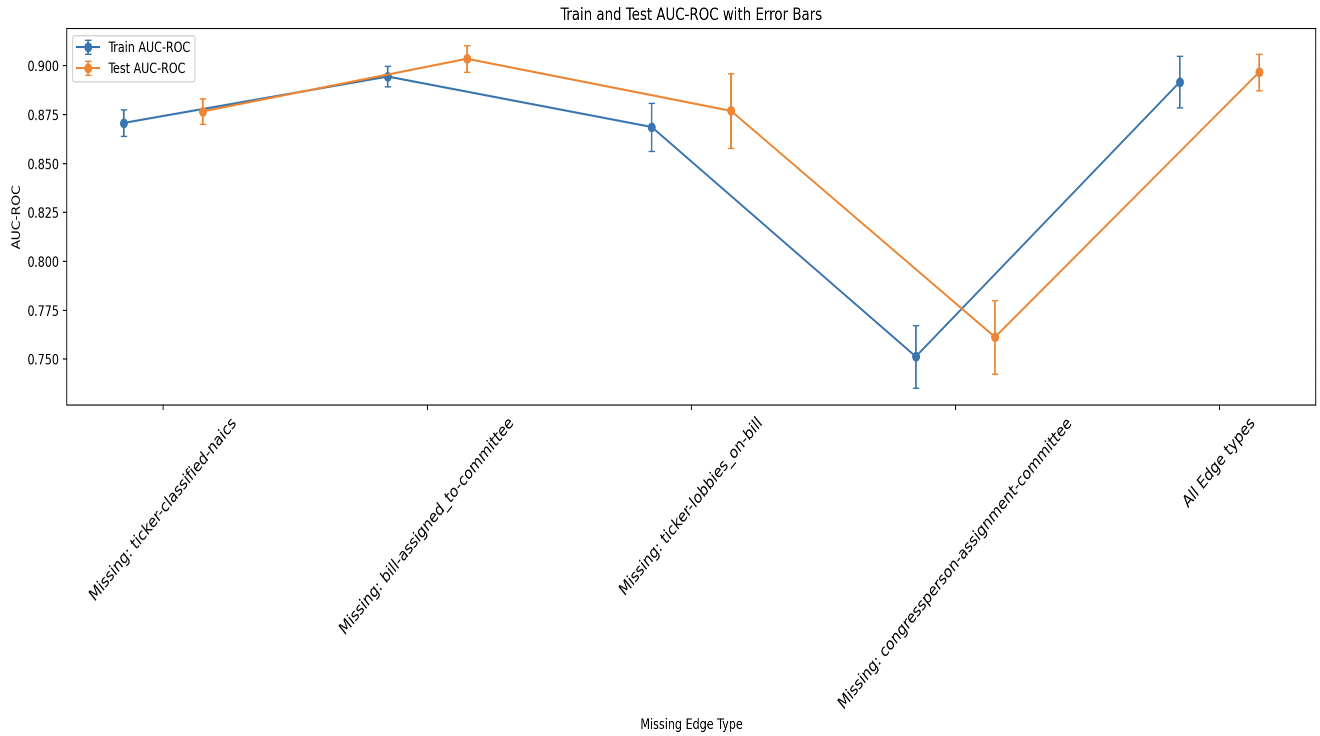


Figure 14: **AUC-ROC drop for different edge types.** The figure shows the AUC-ROC of the model with all edge types included and with each edge type removed one at a time. With all edge types included, the model achieved an AUC-ROC of approximately 0.89. The most significant drop in AUC-ROC, to 0.76, was observed when the edge type ('congressperson','assignment', 'committee') was removed. This suggests that the ('congressperson', 'assignment', 'committee') edge type carries significant information for predicting a congressperson's stock transactions.

To further quantify the importance of each edge type, I employed the concept of Shapley values (Winter, 2002; Hart, 1989; Littlechild and Owen, 1973), a concept borrowed from cooperative game theory. In this context, each edge type can be considered as a player in a cooperative game, where the "payout" is the performance of the model.

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

Here, $\varphi_i(v)$ is the Shapley value for edge type i, representing the average marginal contribution of edge type i to the performance of the model, considering all possible combinations of edge types. N is the set of all edge types, S is a subset of N that does not include edge type i, $|S|$ is the number of edge types in subset S, and n is the total number of edge types. $v(S \cup i)$ and $v(S)$ represent the performance of the model when edge type i is added to and excluded from the subset S of edge types, respectively.

I computed Shapley values over all $16 (= 2^4)$ possible combinations of the four different edge types. The results of this analysis are shown in Figure 15. Consistent with our ablation study, the Shapley value analysis indicated that the most important feature was (‘congressperson’, ‘assignment’, ‘committee’), followed by (‘ticker’, ‘classified_as’, ‘naics’) and (‘ticker’, ‘lobbies_on’, ’bill’).

This further reinforces the conclusion that the (‘congressperson’, ‘assignment’, ‘committee’) edge type plays a crucial role in predicting congressperson’s stock transactions.

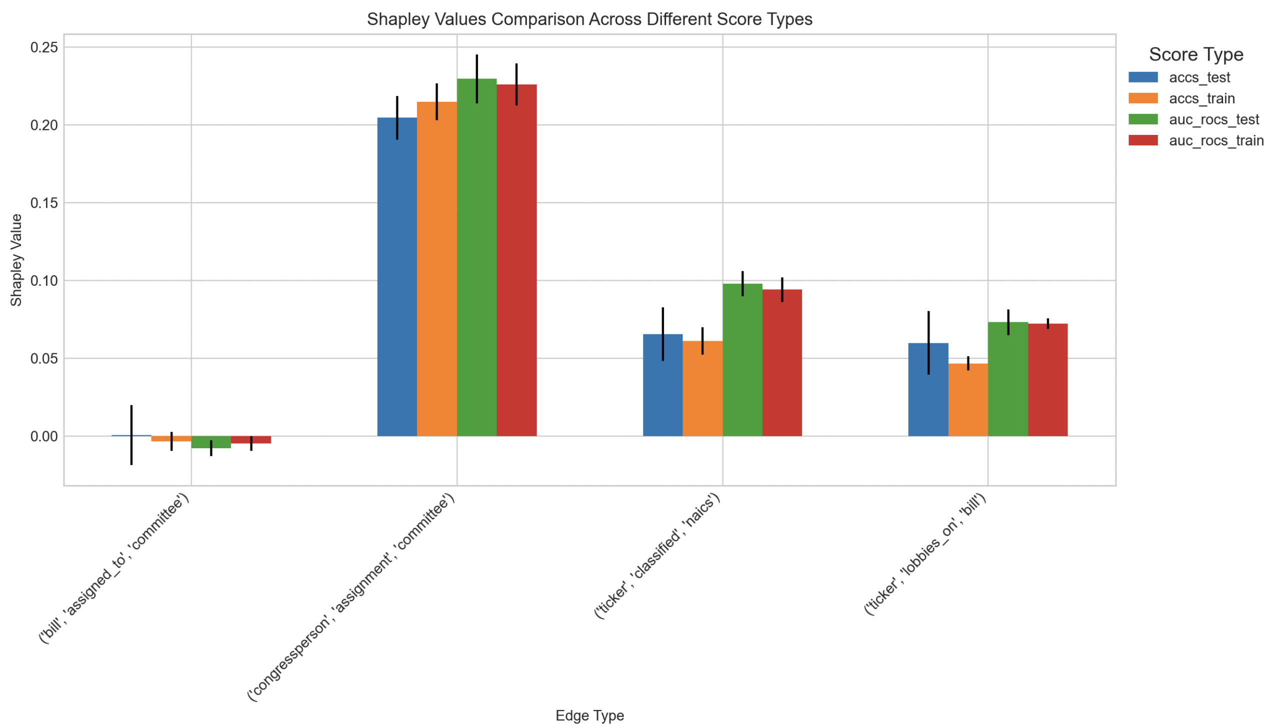


Figure 15: **Shapley values for different edge types.** The figure shows the Shapley values for each edge type, computed over all $16(2^4)$ possible combinations of the four different edge types. The Shapley value for an edge type represents its average marginal contribution to the performance of the model, considering all possible combinations of edge types. The most important feature, according to the Shapley value analysis, was (‘congressperson’, ‘assignment’, ‘committee’), followed by (‘ticker’, ‘classified_as’, ‘naics’) and (‘ticker’, ‘lobbies_on’, ‘bill’). This further reinforces the conclusion that the (‘congressperson’, ‘assignment’, ‘committee’) edge type plays a crucial role in predicting congressperson’s stock transactions.

In the Shapley value analysis, I observed that the edge type (‘bill’, ‘assigned_to’, ‘committee’) had a Shapley value of zero or even negative. This suggests that this type of edge does not contribute to increasing the performance of the model. In fact, it appears to harm the performance when included.

The reason for the zero or negative contribution of the edge type (‘bill’, ‘assigned_to’, ‘committee’) is not immediately clear and warrants further investigation. One possible explanation could be that bills can be assigned to different committees, making this information more complex and potentially harder for the model to utilize effectively. In contrast, the firm-level lobbying information and industry-level classification of firms provided by the edge types (‘ticker’, ‘classified_as’, ‘naics’) and (‘ticker’, ‘lobbies_on’, 'bill') are more straightforward. These edge types may allow the model to more easily discern patterns in company behavior and use this information to make accurate predictions.

5.4 Interpreting Predictions with GNNExplainer

[11]To further explain which nodes and edges the trained model focuses on to output such predictions, I used GNNExplainer (Ying et al., 2019), which trains soft node and edge masks that can be applied to the original graph to extract the subgraph most relevant to the prediction.

The detailed implementation of GNNExplainer involves modifying the update rule for the node representations in the GNN. The original update rule is:

$$\mathbf{h}_i^{(l+1)} = \mathbf{\Theta}^{(l)}\mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \mathrm{MLP}_k^{(l)}\left(\mathbf{e}_{i,j}^{(k)}\right) \cdot \mathbf{h}_j^{(l)}$$

In the modified update rule, we introduce soft node and edge masks, denoted by m_i and $m_{i,j}$ respectively, which are element-wise multiplied with the node and edge representations:

$$\mathbf{h}_i^{(l+1)} = m_i \cdot \mathbf{\Theta}^{(l)}\mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} m_{i,j} \cdot \mathrm{MLP}_k^{(l)}\left(\mathbf{e}_{i,j}^{(k)}\right) \cdot \mathbf{h}_j^{(l)}$$

The soft masks are continuous values between 0 and 1, as opposed to hard masks which are either 0 or 1. This allows us to optimize the masks using stochastic gradient descent (SGD).

The objective of the optimization is to minimize the L2 loss between the predictions of the original graph and the masked graph, denoted by y_{original} and y_{masked} respectively:

$$\mathcal{L} = \|y_{\text{original}} - y_{\text{masked}}\|^2 + \lambda \cdot \left(\|m_i\|_1 + \|m_{i,j}\|_1\right)$$

Here, λ is a regularization parameter that controls the complexity of the subgraph by encouraging sparsity in the masks. For this study, I used a value of 0.01 for λ.

The masks are trained separately for each prediction, which makes the method less scalable but provides insights into which nodes and edges are important for mimicking the original model's prediction. After training the node and edge masks, I can generate a subgraph by applying the masks to the original graph. The complexity of the subgraph can be controlled by setting a cutoff level for the mask values,

[11]Reproducible code for this subsection is available at `https://github.com/syyunn/gnnex/blob/main/hetero/explain_edge.py`

or by adjusting the regularization parameter λ. Figures 16 and 17 provide examples of the output from GNNExplainer for specific stock transactions.

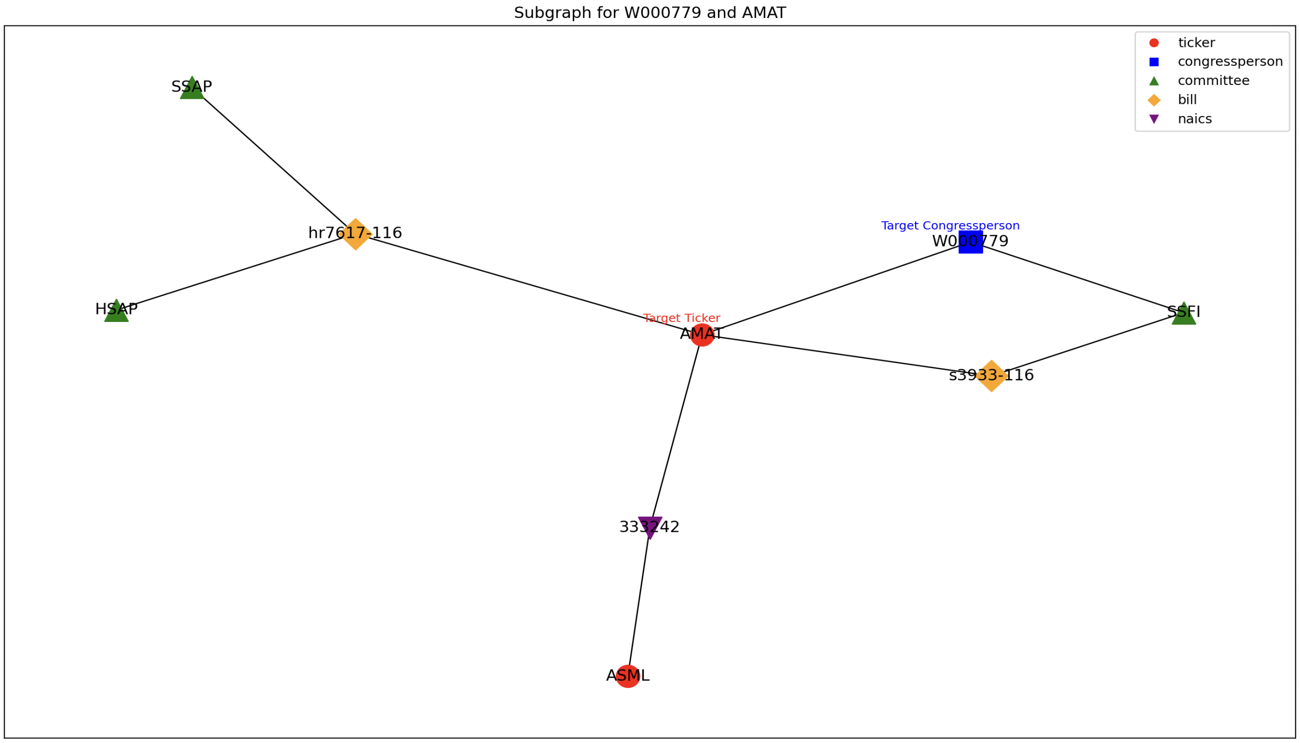


Figure 16: **GNNExplainer output for Senator Ron Wyden's transaction of AMAT's stock.** The figure shows the subgraph extracted from the entire graph using the node and edge masks trained by the GNNExplainer. The GNNExplainer identified S3933-116 (the CHIPS Act) and HR7617-116th among the bills that AMAT lobbied on, NAICS code 333242 (Semiconductor Machinery Manufacturing) among the NAICS code classifications of AMAT, and ASML, among the firms classified as 3333242 as the most influential factors for this transaction.

Figure 16 focuses on Senator Ron Wyden's transaction of Applied Materials Inc. (AMAT)'s stock. To generate the subgraph from the entire graph, I applied the node and edge masks trained by the GNNExplainer. This process involved selecting the nodes and edges with the highest scores from the masks. For instance, among all 56 bills that AMAT lobbied on, I selected the two bills that had the highest scores in the edge mask. Similarly, among the two NAICS code classifications of AMAT, I selected the NAICS code 333242 (Semiconductor Machinery Manufacturing), which had the highest score in the edge mask.

The GNNExplainer successfully identified the most relevant bill for this transaction, S.3933-116, the CHIPS Act, which subsidizes the US semiconductor industry. Interestingly, the GNNExplainer also highlighted H.R.7617-116, a more general appropriations act, titled "Defense, Commerce, Justice, Science, Energy and Water Development, Financial Services and General Government, Labor, Health and Human Services, Education, Transportation, Housing, and Urban Development Appropriations Act, 2021 ". While this bill may not be directly related to subsidization of the semiconductor industry, it is indicative of the broader legislative environment. It's worth noting that the National Defense Authorization Act (NDAA) for Fiscal Year 2021, which is often associated with appropriations for the semiconductor industry, was also part of the data. However, the GNNExplainer did not identify it as a highest score node for this particular transaction. This could be due to a variety of reasons, such as the complexity of the appropriations process or the indirect relationship between the NDAA and H.R.7617-116.

In addition to identifying relevant bills, the GNNExplainer also provided insights into the industry context of Senator Wyden's transaction of AMAT's stock. The NAICS code 333242, which corresponds to Semiconductor Machinery Manufacturing, includes four different companies: Applied Materials, ASML LLC (ASML), Azenta Inc (AZTA), and Tokyo Electron America Inc (TOELY). The GNNExplainer ranked ASML as the most relevant node for this transaction. This makes sense given the industry dynamics. While Azenta is a semiconductor company, it primarily focuses on bio-related semiconductor products, which may not be as directly relevant to AMAT's business. Tokyo Electron, on the other hand, is a much smaller firm compared to ASML or AMAT. Most importantly, ASML and AMAT are known to have a competitive relationship, with AMAT consistently striving to capture market share from ASML. Therefore, the GNNExplainer's identification of ASML as the most relevant node for this transaction is consistent with the industry context.

As the chair of the Senate Finance Committee (SSFI), Senator Wyden has been a key player in initiatives related to the semiconductor industry. Despite the unexpected selection of H.R.7617-116 by the GNNExplainer, the overall results still highlight the relevance of Senator Wyden's legislative activities and industry context to his stock transactions.

Figure 17 presents another example, focusing on Senator Pat Roberts's transaction of Amazon (AMZN)'s stock. Among the 119 bills that Amazon lobbied on, the GNNExplainer identified H.R.2930-115 (Drone Innovation Act of 2017) and S.1410 (Safe DRONE Act of 2017) as the most relevant bills. This is particularly interesting because Senator Roberts co-sponsored the bill S.2730-116, which establishes a Drone Advisory Committee. Furthermore, 2017 was the year when Amazon started to publicize their plans for drone delivery[12]. Therefore, we can interpret that Senator Roberts's transaction of Amazon's

[12] https://www.businessinsider.com/amazon-takes-critical-step-toward-drone-delivery-2017-5

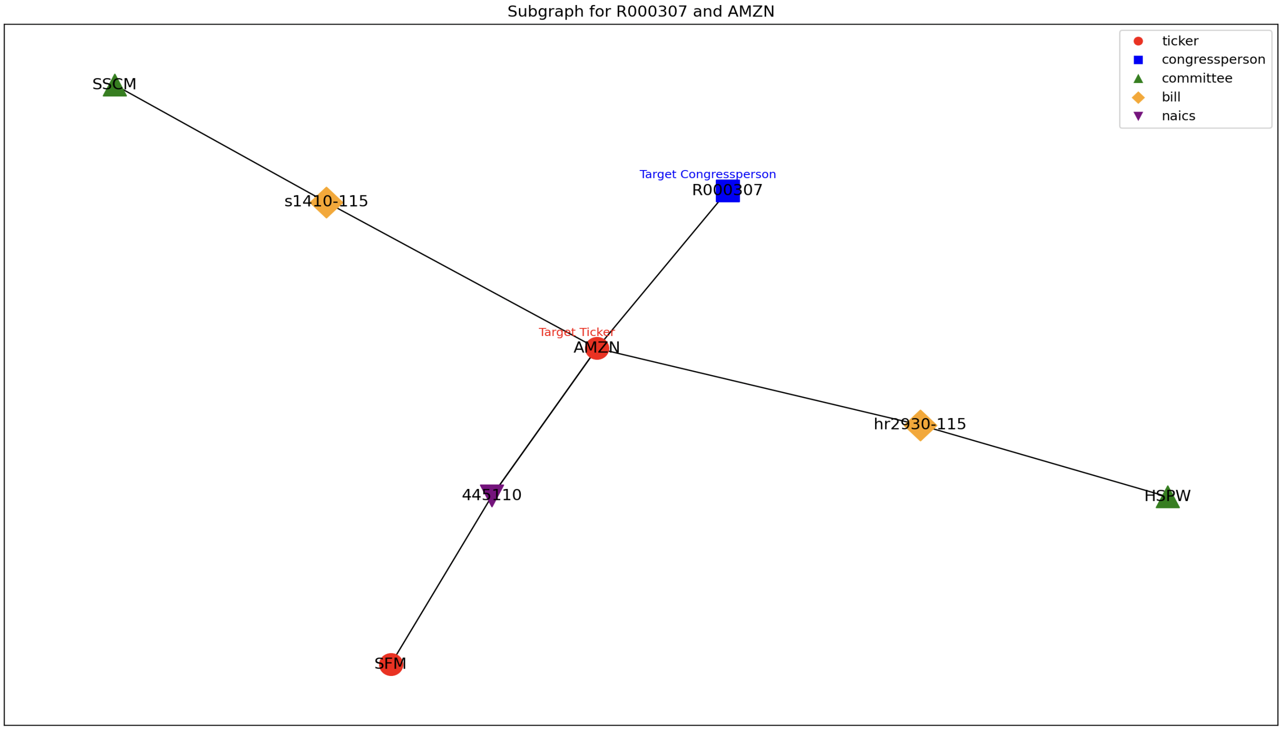


Figure 17: **GNNExplainer output for Senator Pat Roberts's transaction of Amazon's stock.** The figure shows the subgraph extracted from the entire graph using the node and edge masks trained by the GNNExplainer. The nodes and edges in the subgraph were selected based on their high scores in the masks, indicating their relevance to the model's prediction for this specific transaction. The GNNExplainer identified H.R.2930-115 (Drone Innovation Act of 2017) and S.1410 (Safe DRONE Act of 2017) among the bills that Amazon lobbied on as the most influential factors for this transaction. This is particularly interesting given Senator Roberts's known legislative activities related to drone technology.

stock was likely influenced by his legislative activities related to drone technology.

These examples demonstrate the ability of the GNNExplainer to highlight relevant legislative and industry context for specific stock transactions. It's noteworthy that the GNNExplainer was able to identify these relationships even without explicit information about bill titles, text, or sponsorship relationships. This suggests that the GNNExplainer is effectively capturing the underlying patterns in the data that are relevant for the prediction task.

In conclusion, the GNNExplainer provides valuable insights into the specific nodes and edges that the model focuses on for its predictions. This allows us to pinpoint the specific legislative and industry factors that a congressperson is likely to consider when making stock transactions.

6 Conclusion

In this study, I delved into the dynamics of congressional stock investment, exploring what exactly influences these investment choices. My analysis aligned with the traditional financial literature's approach of excess return, which provided a direct estimation of possible excess return. This estimation addressed the range-censored limitation of the financial disclosure at the specific congress-ticker level, considering the life-cycle of transactions - from consecutive purchases to consecutive sales. This reconfirmed the findings of Eggers and Hainmueller (2013), which argued there was no widespread excess return among congressional investments. However, my findings indicate that such excess returns do exist at least abnormally and asymmetrically, more pronounced in the positive skewness compared to the negative returns. This suggests that some privileged information may drive such asymmetry in their excess return overall.

Secondly, I addressed a puzzle originating from the conclusions drawn by Eggers and Hainmueller (2014), who found no clear evidence that congresspersons disproportionately invested in stocks linked to their lobbying and committee assignments. This finding seemed somewhat counterintuitive and diverged from the extensive research on committee assignment and congresspersons' specialization in specific topics governed by committees. Stemming from this, I compiled a novel graph-structured dataset that more comprehensively captures congressional activities. This data utilizes a hetero-graph type, representing the interactions between different types of entities. Leveraging this dataset, I proposed a novel measure using cross-entropy, demonstrating that a congressperson's stock portfolio significantly resembles the stocks related to their assigned committees, compared to unassigned ones.

Furthermore, I expanded on the work of Eggers and Hainmueller (2014) by using the graph neural network to determine how possibly relevant factors, such as congressional activities captured in the graph data, predict congresspersons' stock transactions. The results showed that, contrary to Eggers and Hainmueller (2014), the committee assignment of congresspersons and lobbying activities of firms are the most important features predicting their stock selections. In addition, to address the black-box nature of predictions leveraging neural networks, I proposed using GNNExplainer (**?**). This type of explainability method complements the evaluation metric to interpret case-by-case predictions and provides a more interpretable and semantically rich explanation of why a particular congressperson may choose certain stocks.

In conclusion, this research contributes to existing literature in three significant ways. First, it confirms the non-widespread excess return among congresspersons but suggests that they systematically acquire positive excess return compared to their losses. Secondly, this research highlights the importance of committee assignments in explaining congressional stock trading, resolving the potential puzzle instigated by Eggers and Hainmueller (2014). Thirdly, it illustrates the utility of graph-structured data and graph neural networks in understanding which features carry the most critical information to explain congressional stock trading overall.

References

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *CoRR*, abs/1803.08375.

Amari, S. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4):185–196.

Annaert, J., Van Osselaer, S., and Verstraete, B. (2009). Performance evaluation of portfolio insurance strategies using stochastic dominance criteria. *Journal of Banking & Finance*, 33(2):272–280.

Asher, H. B. (1974). Committees and the norm of specialization. *The Annals of the American Academy of Political and Social Science*, 411:63–74.

Bainbridge, S. (2010). Insider trading inside the beltway. Law-Econ Research Paper 10-08, UCLA School of Law.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2):773–806.

Barberis, N. and Thaler, R. (2003). A survey of behavioral finance. In *Handbook of the Economics of Finance*, volume 1, pages 1053–1128. Elsevier.

Bauer, M. and Rudebusch, G. D. (2014). The signaling channel for federal reserve bond purchases. *International Journal of Central Banking*.

Besley, T. (2006). *Principled Agents?: The Political Economy of Good Government*. Oxford University Press, Oxford, UK.

Białkowski, J., Darolles, S., and Le Fol, G. (2008). Improving vwap strategies: A dynamic volume approach. *Journal of Banking & Finance*, 32(9):1709–1722.

Boller, G. (1995). Taking stock in congress. *Mother Jones*.

Buchanan, J. M. and Tollison, R. D., editors (1984). *The Theory of Public Choice - II*. University of Michigan.

Buhrmester, V., Münch, D., and Arens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, 3(4):966–989.

Cziraki, P., Lyandres, E., and Michaely, R. (2021). What do insiders know? evidence from insider trading around share repurchases and seos. *Journal of Corporate Finance*, 66:101544.

Das, L., Sivaram, A., and Venkatasubramanian, V. (2020). Hidden representations in deep neural networks: Part 2. regression problems. *Computers & Chemical Engineering*, 139:106895.

Dayhoff, J. E. and DeLeo, J. M. (2001). Artificial neural networks: opening the black box. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 91(S8):1615–1635.

Diermeier, D., Keane, M., and Merlo, A. (2005). A political economy model of congressional careers. *The American Economic Review*, 95(1):347–373.

Duffie, D. and Dworczak, P. (2021). Robust benchmark design. *Journal of Financial Economics*, 142(2):775–802.

Eggers, A. and Hainmueller, J. (2013). Capitol losses: The mediocre performance of congressional stock portfolios, 2004-2008. *Journal of Politics*, 75.

Eggers, A. C. and Hainmueller, J. (2014). Political capital: Corporate connections and stock investments in the u.s. congress, 2004-2008. *ERN: Models of Political Processes: Rent-Seeking*.

Fearon, J. D. (1999). Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance. In Przeworski, A., Stokes, S. C., and Manin, B., editors, *Democracy, Accountability, and Representation*, pages 55–97. Cambridge University Press, Cambridge.

Fenno, Richard F., J. (1977). U.s. house members in their constituencies: An exploration. *The American Political Science Review*, 71(3):883–917.

Féraud, R. and Clérot, F. (2002). A methodology to explain neural network classification. *Neural networks*, 15(2):237–246.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1/3):5–25.

Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32:2627–2636.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). Neural message passing for quantum chemistry. *CoRR*, abs/1704.01212.

Goldman, E., Rocholl, J., and So, J. (2009). Do politically connected boards affect firm value? *Review of Financial Studies*, 22(6):2331–2360.

Hart, S. (1989). *Shapley value*. Springer.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Henaff, M., Bruna, J., and LeCun, Y. (2015). Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*.

Hoechle, D. and Zimmermann, H. (2007). A generalization of the calendar time portfolio approach and the performance of private investors. *Faculty of Business and Economics - University of Basel, Working papers*.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91.

Hong, H., Guo, H., Lin, Y., Yang, X., Li, Z., and Ye, J. (2020). An attention-based graph neural network for heterogeneous structural learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 4132–4139.

Hsu, H. and Lachenbruch, P. A. (2014). Paired t test. *Wiley StatsRef: statistics reference online*.

Ivković, Z. and Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60(1):267–306.

Jeng, L. A., Metrick, A., and Zeckhauser, R. (2003). Estimating the returns to insider trading: A performance-evaluation perspective. *Review of Economics and Statistics*, 85(2):453–471.

Jerke, B. W. (2010). Cashing in on capitol hill: Insider trading and the use of political intelligence for profit. *University of Pennsylvania Law Review*, 158:1451–1523.

Jin, D., Huo, C., Liang, C., and Yang, L. (2021). Heterogeneous graph neural network via attribute completion. In *Proceedings of the Web Conference 2021*, pages 391–400.

Kaushik, R., Shenoy, P., Bohannon, P., and Gudes, E. (2002). Exploiting local similarity for indexing paths in graph-structured data. In *Proceedings 18th International Conference on Data Engineering*, pages 129–140. IEEE.

Khwaja, A. I. and Mian, A. (2005). Do lenders favor politically connected firms? rent provision in an emerging financial market. *The Quarterly Journal of Economics*, 120(4):1371–1411.

Kim, I. S. (2018). Lobbyview: Firm-level lobbying and congressional bills database. Working Paper.

King, D. C. (1994). The nature of congressional committee jurisdictions. *The American Political Science Review*, 88(1):48–62.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lenz, G. and Lim, K. (2009). Getting rich(er) in office? corruption and wealth accumulation in congress.

Littlechild, S. C. and Owen, G. (1973). A simple expression for the shapley value in a special case. *Management Science*, 20(3):370–372.

Lu, Y. and Lu, J. (2020). A universal approximation theorem of deep neural networks for expressing distributions. *CoRR*, abs/2004.08867.

Madhavan, A. N. (2002). Vwap strategies. *Trading*, 2002(1):32–39.

Mao, J., Yao, D., and Wang, C. (2013). A novel cross-entropy and entropy measures of ifss and their applications. *Knowledge-Based Systems*, 48:37–45.

Myers, B. W. (2009). Firms, politicians, and capital structure. Working Paper.

Olden, J. D. and Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*, 154(1-2):135–150.

Patterson, S. C. (1970). The professional staffs of congressional committees. *Administrative Science Quarterly*, 15(1):22–37.

Price, D. E. (1978). Policy making in congressional committees: The impact of “environmental” factors. *American Political Science Review*, 72(2):548–574.

Querubin, P. and Snyder, James M., J. (2013). The control of politicians in normal times and times of crisis: Wealth accumulation by u.s. congressmen, 1850–1880. *Quarterly Journal of Political Science*, 8(4):409–450.

Rauber, P. E., Fadel, S. G., Falcao, A. X., and Telea, A. C. (2016). Visualizing the hidden activity of artificial neural networks. *IEEE transactions on visualization and computer graphics*, 23(1):101–110.

Roberts, B. E. (1990). A dead senator tells no lies: Seniority and the distribution of federal benefits. *American Journal of Political Science*, 34(1):31–58.

Román, J. H., Hulin, K. J., Collins, L. M., and Powell, J. E. (2012). Entity disambiguation using semantic networks. *Journal of the American Society for Information Science and Technology*, 63(10):2087–2099.

Ruby, U. and Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, 9.

Sarno, L. and Thornton, D. L. (2003). The dynamic relationship between the federal funds rate and the treasury bill rate: An empirical investigation. *Journal of Banking & Finance*, 27(6):1079–1110.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80.

Schweizer, P. (2011). *Throw Them All Out: How Politicians and Their Friends Get Rich Off Insider Stock Tips, Land Deals, and Cronyism That Would Send the Rest of Us to Prison.* Houghton Mifflin Harcourt (HMH).

Seasholes, M. S. and Zhu, N. (2009). Individual investors and local bias. *Journal of Finance*. Forthcoming.

Shanken, J. (1996). 23 statistical methods in tests of portfolio efficiency: A synthesis. *Handbook of statistics*, 14:693–711.

Simonovsky, M. and Komodakis, N. (2017). Dynamic edge-conditioned filters in convolutional neural networks on graphs. *CoRR*, abs/1704.02901.

Sivakumar, K. and Waymire, G. (1994). Insider trading following material news events: Evidence from earnings. *Financial Management*, pages 23–32.

Tahoun, A. (2014). The role of stock ownership by us members of congress on the market for political favors. *Journal of Financial Economics*, 111(1).

Tang, J., Deng, C., and Huang, G.-B. (2016). Extreme learning machine for multilayer perceptron. *IEEE Transactions on Neural Networks and Learning Systems*, 27(4):809–821.

Weiss, C. H. (1989). Congressional committees as users of analysis. *Journal of Policy Analysis and Management*, 8(3):411–431.

Winter, E. (2002). The shapley value. *Handbook of game theory with economic applications*, 3:2025–2054.

Wu, H., Yuan, Y., Wei, L., and Pei, L. (2018). On entropy, similarity measure and cross-entropy of single-valued neutrosophic sets and their application in multi-attribute decision making. *Soft Computing*, 22:7367–7376.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24.

Yang, Z., Ding, M., Zhou, C., Yang, H., Zhou, J., and Tang, J. (2020). Understanding negative sampling in graph representation learning. *CoRR*, abs/2005.09863.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32.

Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019). Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 793–803.

Zhang, M. and Chen, Y. (2018). Link prediction based on graph neural networks. *CoRR*, abs/1802.09691.

Zhao, J., Wang, X., Shi, C., Hu, B., Song, G., and Ye, Y. (2021). Heterogeneous graph structure learning for graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4697–4705.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, 1:57–81.

Ziobrowski, A. J., Boyd, J. W., Cheng, P., and Ziobrowski, B. J. (2011). Abnormal returns from the common stock investments of members of the u.s. house of representatives. *Business and Politics*, 13(1):1–22.

Ziobrowski, A. J., Cheng, P., Boyd, J. W., and Ziobrowski, B. J. (2004). Abnormal returns from the common stock investments of the u.s. senate. *The Journal of Financial and Quantitative Analysis*, 39(4):661–676.