Unveiling the Connection: Congressional Knowledge, Committee Assignments, and Stock Trading Patterns Using Graph Neural Networks

Suyeol Yun *

March 16, 2023

Abstract

This paper explores the connection between congressional knowledge and the stock trading behaviors of Congress members. While earlier research has focused on excess returns as proof of insider trading, this study aims to provide direct evidence of whether Congress members utilize their privileged information to guide their trading decisions. To do this, the paper creates a network graph that captures congressional knowledge, including Congress members, committees, bills, firms, and their interactions. However, transforming this graph-structured data into a numerical representation suitable for estimating causal effects presents a methodological challenge. To address this, the paper employs a Graph Neural Network (GNN) to generate a numerical representation of the graph-structured data. This representation is then used in conjunction with meta-learners to effectively estimate the conditional average treatment effect of committee assignments on Congress members' stock portfolios, making them more similar to industry-level information in the committee. In conclusion, the study offers two significant contributions by revealing the impact of committee assignments on members' trading patterns and demonstrating a new method for estimating CATE using graph-structured data.

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

1 Introduction

The investment behavior of congressmen has always been a topic of interest for researchers as it raises concerns about conflicts of interest (Tahoun, 2014) and insider trading (Kim, 2012). Members of Congress have access to privileged and confidential information, such as pending legislation, government contracts, and upcoming regulations. If members of Congress use this information to make investments, it can be seen as unfair and unethical as they have an advantage over other market participants who do not have access to this information. Furthermore, If members of Congress use their positions to gain financial benefits, it can be seen as a breach of public trust and may undermine the integrity of the democratic system.

Numerous studies have analyzed the investment behavior of Congress members, primarily focusing on measuring their excess returns. Research by Ziobrowski et al. (2011) and Ziobrowski et al. (2004) discovered that Senators and House representatives earn more than the market and enjoy excess returns, while a study by Eggers and Hainmueller (2012) found that they do not.

These studies attempt to prove the existence of insider trading by measuring Congress members' excess returns, which involves calculating their average yearly return on investment. The approach assumes that, all else being equal, Congress members' returns should not exceed the market's average return, represented by index funds like the S&P 500.

However, this method has been criticized because the main interest lies in whether Congress members use their knowledge gained from congressional activities for personal investment, rather than whether they enjoy excess returns or not. As a result, excess return is merely considered indirect evidence. In the preliminary section, the paper will present real-world cases in which Senators appear to have relevant knowledge but fail to enjoy excess returns.

Moreover, previous literature has primarily focused on measuring the average performance of the Senate (Ziobrowski et al., 2011) or House (Ziobrowski et al., 2004) at the chamber level. However, it is reasonable to expect significant variations in behavior between individuals due to the different sources of congressional knowledge they acquire during their activities. Therefore, unlike prior research that solely relies on the securities transaction data of Congress members without specifying the source of congressional knowledge that could affect their investment behavior, this paper incorporates the source of congressional knowledge as a covariate in the analysis.

For instance, interest groups are known to influence the legislative process by revealing their policy preferences and lobbying lawmakers (Smith, 1995). They also provide policy information, political intelligence, and legislative assistance to policymakers (Hall and Deardorff, 2006). Additionally, Congress members are often assigned to multiple committees, where they oversee specific topics and organized interests. Consequently, members of Congress have access to detailed industry and firm-level information that can help them understand policy preferences and the potential impact of certain legislation or political events on specific firms or industries. All this information can be considered congressional knowledge that may affect their financial investments.

To capture this congressional knowledge, the paper employs graph-structured data[1][2] that encompasses the interactions between Congress members, committees, firms, and bills in terms of their legislative and lobbying activities.

To examine whether Congress members utilize the knowledge acquired through their positions for personal stock trading, this study estimates the causal effect of committee assignments on their stock transaction patterns. Specifically, the study investigates how being assigned to a particular committee influences the similarity between the NAICS code distribution of a Congress member's portfolio and that of the assigned committee. This is because committee assignments equip Congress members with essential expertise in specific issue areas or industries (Asher, 1974).

If the transaction patterns of Congress members become more similar to the industry-level information within a committee after being assigned to that committee, it would suggest that they are using their congressional knowledge for personal investment purposes.

The primary focus of this study is to estimate the conditional average treatment effect (CATE) of how a Senator's committee assignment affects the similarity between the Senator's and the assigned committee's NAICS code distribution. To estimate this, it is crucial to control for confounding variables that can influence both committee assignment and a Senator's investment behavior. The study assumes that the graph-structured data, which includes information on which firms lobby on which bills, which bills are assigned to which committees, and which Senators are members of which committees, provides all the necessary information to explain a Senator's assignment to a particular committee and their securities transactions. To support this assumption with empirical evidence, the paper will present a

[1] In network science, the terms "graph" and "network" are often used interchangeably to refer to a collection of nodes and edges, where the edges represent some sort of relationship or interaction between the nodes.

[2] Unlike the usual tabular data used in most political science research, graph-structured data has the advantage of including the relationship between covariates in the form of connected edges between two nodes.

preliminary analysis.

However, unlike widely used tabular data, graph-structured data cannot be directly employed to estimate a quantity of interest through a model. Since the estimation involves mathematical operations over continuous variables, graph data, being a collection of discrete objects, does not have defined operations. Nonetheless, the research emphasizes the importance of using graph-structured data, as it can encompass additional information that explains the relationships between different entities in the network during interactions in congressional activities. To overcome this computational challenge, this paper utilizes Graph Neural Networks (GNN) to learn the appropriate numerical representation specific to each Senator, who resides as a node in the graph. This representation is a numeric vector and is expected to include all the information available from the graph-structured data to explain both a Senator's stock trading behavior and their committee assignment.

There are a total of 62 different committees, and CATE is estimated for each committee separately. By evaluating these CATE values, the paper aims to quantitatively evaluate how committee assignment heterogeneously affects Senators' behavior across committees and among Senators themselves.

The paper aims to make two main contributions. Firstly, it substantively proves the existence of insider trading by explicitly showing that committee membership changes affect Congressmen's investment behavior to become similar to the information available from the committee. Secondly, it demonstrates a novel methodological approach to estimate causal quantities using graph-structured data by leveraging representation learning via Graph Neural Network. Therefore, the paper proposes a model and learning scheme that combines a Graph Neural Network and meta-learners to learn CATE, conditioning on the output of the Graph Neural Network.

2 Data

The data utilized in this study is a congressional network that records the legislative activities of Senators and their investment behavior. The network is created by merging data from LobbyView with Senate financial disclosures. It contains details about the Senators' buying/selling of company stocks, the dates of these transactions, their committee assignments during specific Congresses, the bills that are being lobbied by companies, and which bills are assigned to which committee. The firms are classified according to NAICS codes, as shown in Figure 1.

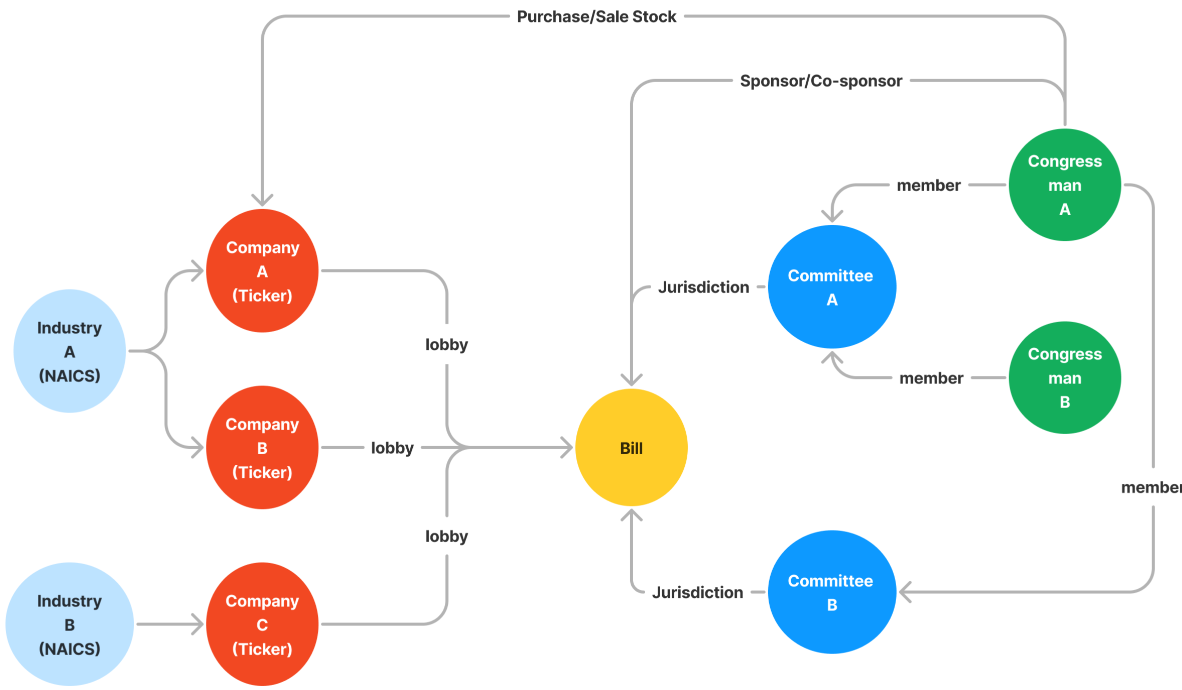


Figure 1: The network data includes various types of nodes and edges that represent different entities and interactions within the congressional activities.

In addition, descriptive statistics for the nodes and edges included in the congressional network are presented in Table 1 and Table 2, respectively. Table 1 provides information about the number and type of nodes in the network, including firms, bills, senators, committees, and NAICS codes. Table 2 presents data on the number and type of edges in the network, including the number of links between firms and bills, senators and bills, and committees and bills, as well as the number of bills assigned to each committee.

Table 1: Data for Congressional Network (Nodes)

Node Type	N	Period	Source
Firms (Ticker)	3,552	2014-2021	Lobbyview & Senate Finance Disclosure
Bills	64,031	113-117th Congress	Lobbyview
Senators	95	114-117th Congress	Lobbyview & Senate Finance Disclosure
Committee	62	-	Lobbyview
NAICS codes	414	-	Additional scraping (from naics.com)

To provide a more concrete understanding of the data, Figure 2 displays a subgraph related to Senator

Table 2: Data for Congressional Network (Edges)

Edge Types	N	Period	Source
Purchase/Sale of Stock	29,475	2014-2021	Senate Finance Disclosure
Firms' Lobbying on Bills	1,534,097	2012-2021	Lobbyview
Ticker-NAICS Codes	414	2014-2021	Senate Finance Disclosure and naics.com
Bill-Committee Assignments	109,307	113-117th Congress	Lobbyview
Senator-Committee Assignments	6,708	114-117th Congress	Senate Finance Disclosure and Lobbyview

Ron Wyden's transaction in Trip Advisor stock (Ticker: TRIP). 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.

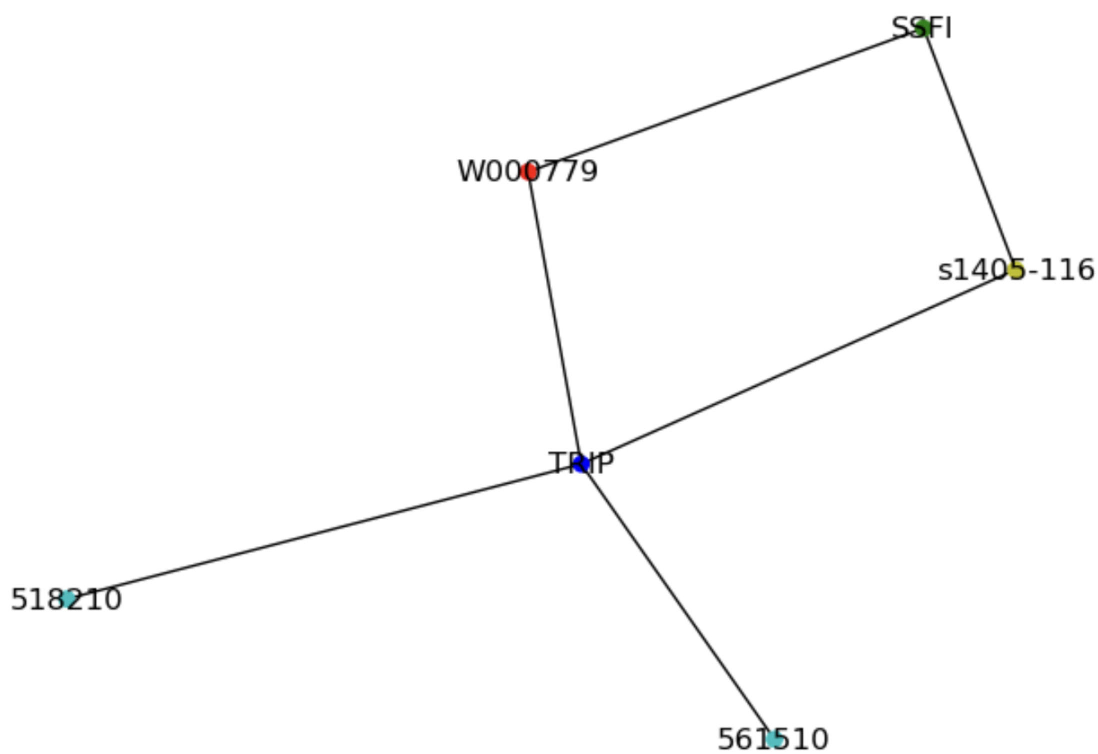


Figure 2: A subgraph illustrating the congressional network related to the transaction of Senator Ron Wyden's Trip Advisor stock.

3 Empirical Strategy

The mission of this paper is to provide evidence that Congressmen's investment patterns are affected by the information they acquire through their congressional activities. To establish a causal relationship, it is necessary to identify a proper treatment and its associated dependent variable that can demonstrate such behavioral changes.

To accomplish this, I conducted preliminary research to identify those who are obtaining a high level of excess return, which is indicated by abnormally high returns from their stock transactions. As previously emphasized in the introduction, excess return alone does not constitute direct evidence of insider trading. However, it can provide a hint about the information that Congressmen are primarily using for their investment decisions, and aid in determining a viable treatment and dependent variable to establish the impact of congressional activities on investment patterns.

3.1 Preliminary Analysis: Excess Returns on Investments of Congressmen

Firstly, I gained insights into the mechanisms behind their trading decisions by reviewing news articles. For example, there were several media reports[3] in May 2021 regarding Ron Wyden's semiconductor stocks trading. I searched for Senator Ron Wyden's stock transactions that occurred prior to May 2021, but within a year, 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 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 3.

On April 1st, 2021, President Biden announced a plan to invest \$50 billion to boost the U.S. chip industry[4]. After this announcement, Senator Ron Wyden sold all of his semiconductor stocks. This suggests that members of Congress may have access to legislative information that can impact certain industries and enable them to design their own portfolio based on this information, potentially resulting in profitable stock trades.

Based on this observation, I developed an algorithm to calculate the excess returns for each (Senator,

[3] https://nypost.com/2021/05/20/us-sen-ron-wyden-boosts-chipmakers-while-his-wife-buys-their-shares/
[4] https://www.wsj.com/articles/biden-urges-50-billion-to-boost-chip-manufacturing-in-u-s-11617211570

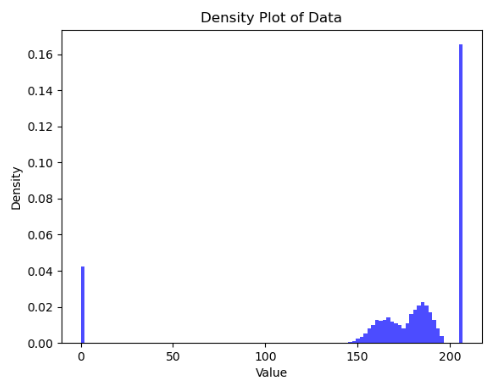
	fi	la	ticker	trans_type	trans_date	trans_type	amount_min	amount_max
1	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
2	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
3	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	15,001	50,000
4	Ron L	Wyden	AVGO	Purchase	2020-04-06	Purchase	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 3: Senator Ron Wyden's stock transactions for Broadcom Inc. (ticker AVGO) exhibited a pattern of multiple purchases followed by sales after certain critical points, spanning from April 6th, 2020 to April 6th, 2021.

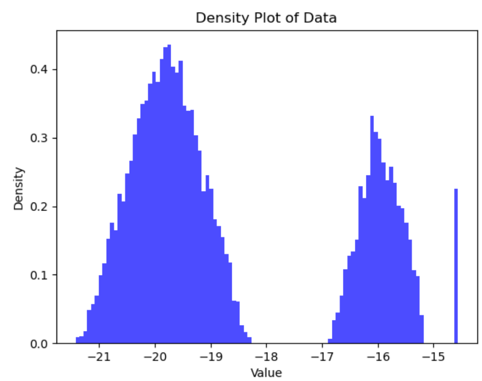
Ticker) pair. Unlike previous literature (Ziobrowski et al., 2011, 2004; Eggers and Hainmueller, 2012) that computes excess returns at the year-chamber level, such as the excess return of the Senate or House over a specific year, I hypothesized that insider trading occurs in a time-specific manner. Therefore, it should be estimated accordingly, rather than a year-based approach. As a result, I selected all transaction chains that start with multiple purchases followed by sales within six months of the transition from purchase to sale.

There is a methodological challenge with using the Senate Finance Disclosure data, as it only provides the minimum and maximum range of amounts spent on purchasing or selling each ticker on a specific day (See Figure 3). To estimate the distribution of excess returns for each particular Purchase-Sale chain at the (Senate, Ticker) level, I randomly sampled the amount from a uniform distribution with support equal to the min/max range of the amount in the data. To obtain a more conservative estimate of the excess return, I penalized the monetization of a unit stock by the average Federal Reserve Rate during the holding period, even though the savings account rate is typically lower.

I am presenting the excess return distributions for Senator Ron Wyden's transactions involving two 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. The X-axis of Figure 4 represents the percentage of return for a unit investment, with a value of 120 indicating a gain of \$1.20 for every \$1 invested.



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

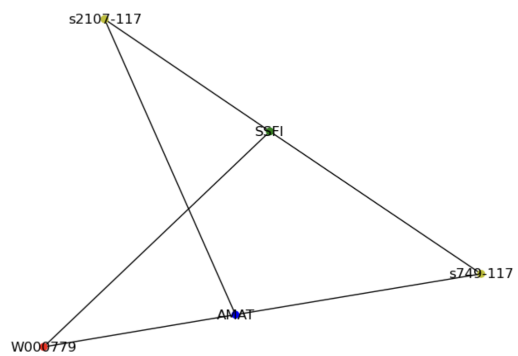


(b) Ron Wyden's excess returns from transactions involving Marriott International Inc. (MAR) from May to August 2020.

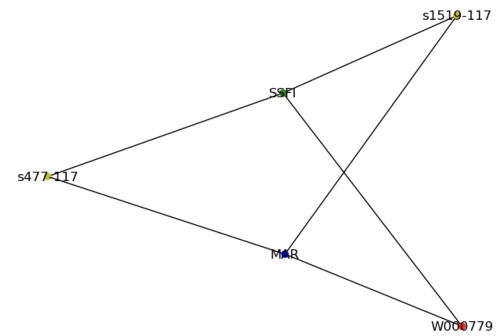
Figure 4: Excess return distributions of Senator Ron Wyden's transactions for AMAT and MAR

As shown in Figure 5, it appears that Senator Ron Wyden may have access to legislative information regarding bills that could impact both the semiconductor and hospitality industries. During the 117th Congress, which coincides with Senator Ron Wyden's transaction involving AMAT, the company lobbied for two bills: S.2107 FABS Act and S.749 American Innovation and Jobs Act. Both bills include provisions that provide subsidies to the semiconductor industry. Additionally, during the 117th Congress, which overlaps with Senator Ron Wyden's transaction involving MAR, the company lobbied for two bills: S.477 Hospitality and Commerce Job Recovery Act of 2021, which extends tax credits to assist the hospitality and restaurant industry, and S.1519, which grants awards to hotel owners. It is worth noting that this legislative information likely flowed from the firms' lobbying efforts directed towards the Senate Finance Committee, which oversees the bills listed.

Based on both Figure 4 and Figure 5, we can conclude that although Senator may have access to legislative information about certain industries, it is not always possible to profit from it due to external factors that the legislative information from congressional activity cannot explain. This example provides suggestive evidence as to why a simple excess return approach cannot provide a persuasive argument



(a) Subgraph extracted from the congressional network that is in close proximity to Senator Ron Wyden's purchase or sale transaction involving AMAT.



(b) Subgraph extracted from the congressional network that is in close proximity to Senator Ron Wyden's purchase or sale transaction involving MAR.

Figure 5: Potential Access to Legislative Information by Senator Ron Wyden Regarding Bills Impacting Semiconductor and Hospitality Industries during the 117th Congress

that proves or disproves the existence of insider trading.

Finally, I provide a distribution of the mean estimated excess return for each (Senator, Ticker) pair in Figure 6. The plot shows that only a few transactions are acquiring abnormal positive profits over 50%, and almost half of them record negative returns. This statistic is consistent with the findings reported by Eggers and Hainmueller (2012), which suggests that there is little evidence of systematic abnormal returns for members of Congress. However, thanks to the current approach that delves more into the Senator and Ticker level, we can conclude that it is hard to deny that a few transactions are acquiring abnormally high returns connected to the relevant legislative information that is accessible during congressional activity.

3.2 Committees: A Channel for Legislators to Acquire Industry-Specific Information

Based on multiple anecdotes and observations, I hypothesize that legislators acquire industry-specific information from their committee assignments that they may use for personal investments. For example, Senator Ron Wyden's semiconductor stock transactions involving multiple tickers in the same industry, such as AMAT, AVGO, and KLAC, can be explained by this hypothesis. These companies lobbied for bills relevant to the semiconductor industry, and the industry-level preferences and possible impacts of legislation on the semiconductor industry were aggregated in the Senate Finance Committee, which oversees such bills. As a member of the Senate Finance Committee, Ron Wyden acquired a more

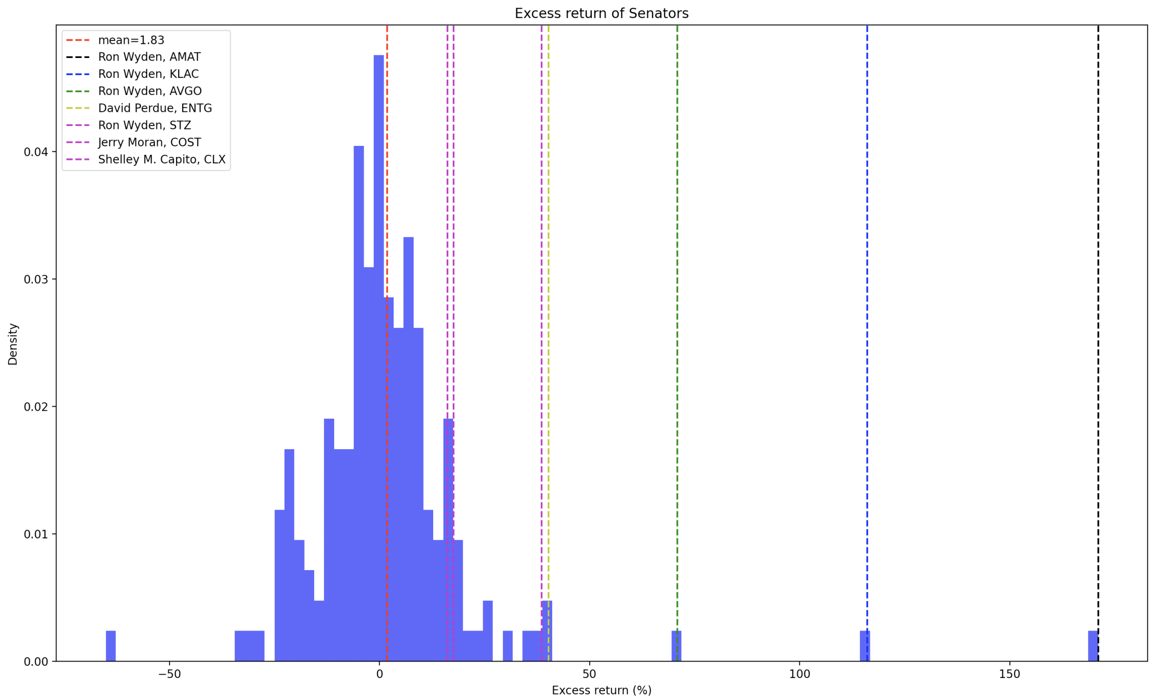


Figure 6: Distribution of congressmen's excess returns at the Senate-ticker level.

detailed understanding of how vital these bills were to the industry, and may have timed his transactions accordingly.

I also provide an additional illustration that shows how the preferences of the semiconductor industry were focused on the Senate Finance Committee during the 117th Congress in Figure 7. 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 the CHIPS Act and FABS 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.

As another example, I provide how 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 8. 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

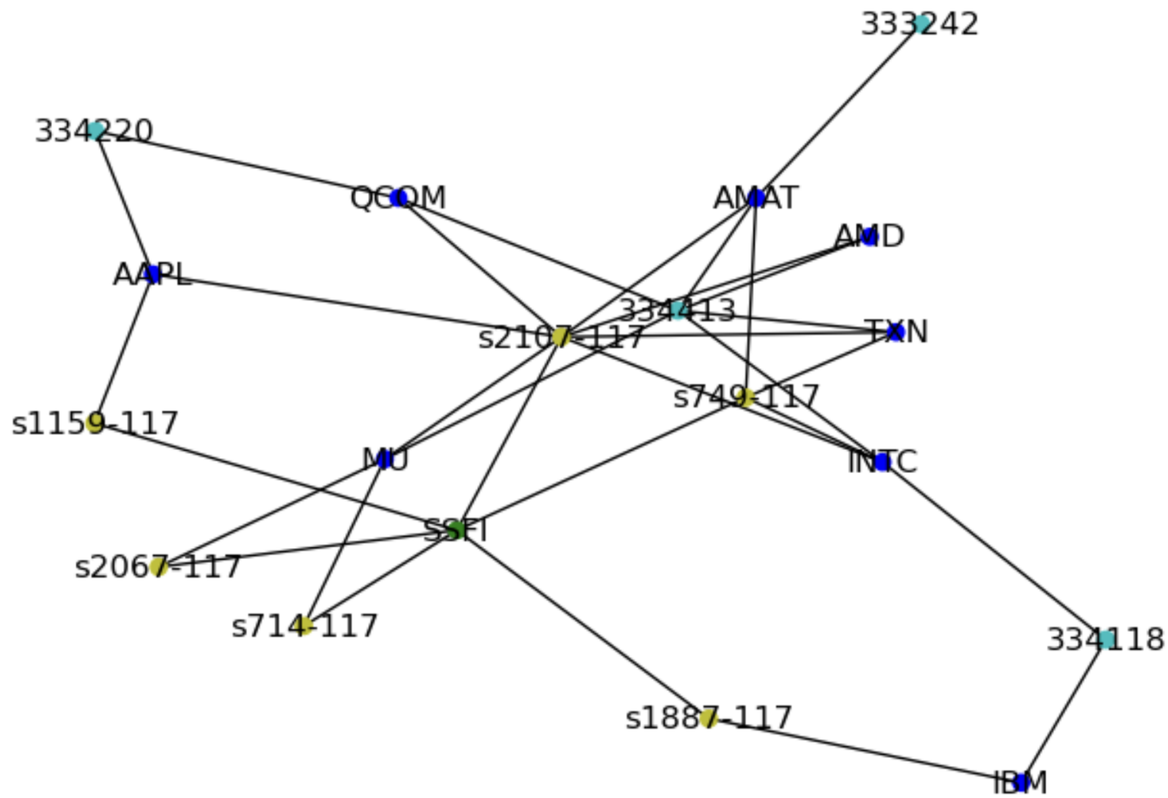


Figure 7: The illustration demonstrates how the semiconductor industry's interests were funneled through the Senate Finance Committee during the 117th Congress, with relevant bills assigned to the committee.

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.

3.3 Causal Quantity of Interest: Effect of Committee Membership on Trading Behavior

In Section 3.2, illustrative examples were provided to suggest that committee assignments can influence a congressman's financial behavior by utilizing industry-level specialized information funneled through the committee. Building on this insight, I propose a Directed Acyclic Graph (DAG) in Figure 9 that captures the hypothesized causal relationships between committee assignments, congressional knowledge, and financial behavior. Specifically, I expect that Congressional Knowledge before joining a committee will act as a confounder, influencing both the committee assignment and financial behavior. After joining the committee, Congressional Knowledge will mediate the effect of the committee assignment on financial

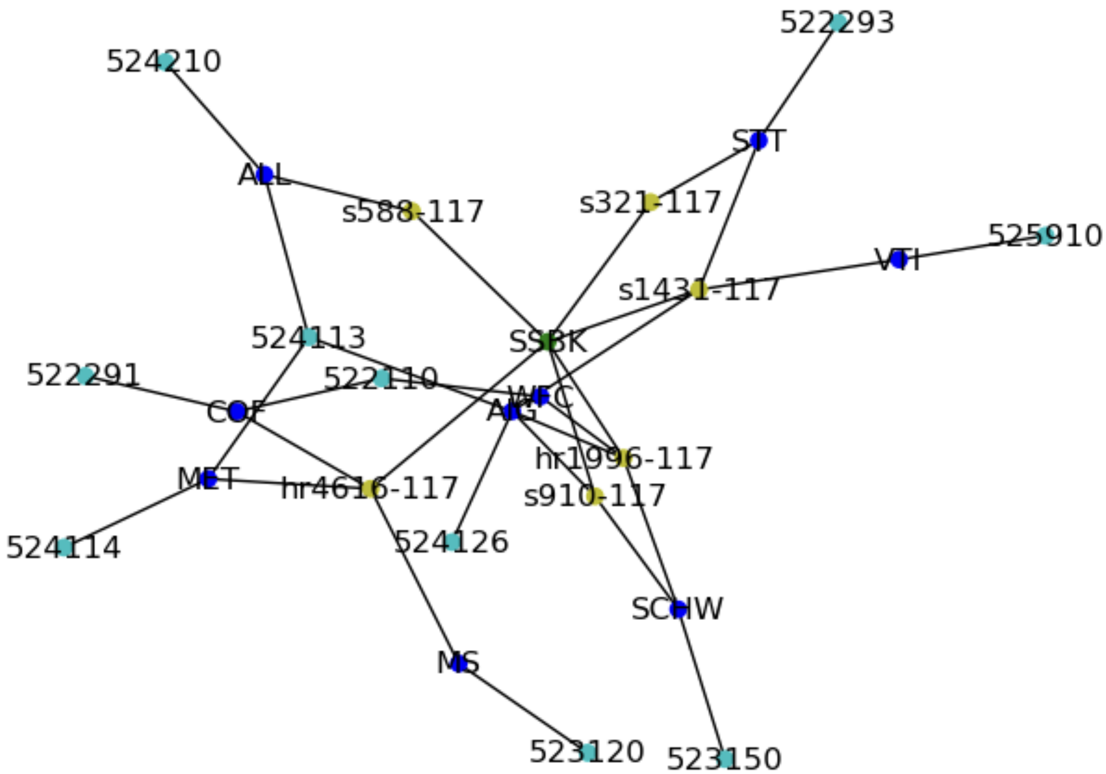


Figure 8: The illustration demonstrates how the financial industry's interests were funneled through the Senate Banking Committee during the 117th Congress, with relevant bills assigned to the committee.

behavior. At the outcome level, I expect that the industry distribution of a senator's portfolio after joining the committee will become more similar to that of the committee. The industry distribution of the committee will be evaluated by examining the firms' industry distribution lobbying for bills assigned to that committee.

To justify the measurement of the outcome variable as a similarity between the industry-level distribution of a senator's stock trading portfolio and that of the committee, I analyzed the stock trading portfolio and its NAICS code distribution for Senator Ron Wyden, as well as those of the Senate Finance Committee (SSFI) and Senate Banking Committee (SSBK) over the 117th Congress, as shown in Figure 8. I calculated the discrete probability distribution of the collection of 2-digit level NAICS codes for each entity, and used cross-entropy to measure the similarity between the Senator and each committee's NAICS code distribution. The results indicate that the cross-entropy between Senator Ron Wyden's NAICS code distribution and that of SSFI is 0.71, while the cross-entropy between Senator Ron Wyden's NAICS code distribution and that of SSBK is 3.31. A smaller cross-entropy value signifies

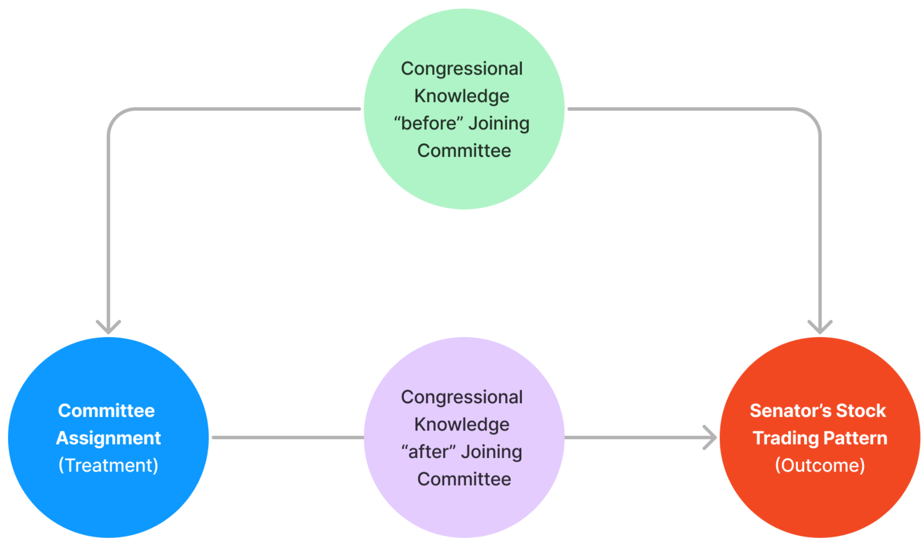
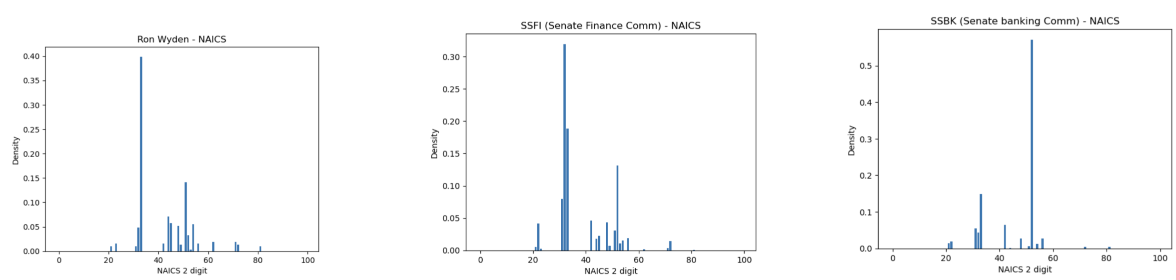


Figure 9: Causal diagram that demonstrates how committee assignments affect the congressmen’s transaction behavior

a more similar distribution, indicating that Senator Ron Wyden’s NAICS code distribution is closer to that of SSFI than SSBK. This supports the hypothesis that committee assignments influence a senator’s investment choices, as Ron Wyden is a member of SSFI but not a member of SSBK.



(a) NAICS code distribution of Senator Ron Wyden’s stock portfolio over the 117th Congress.

(b) NAICS code distribution of firms lobbying for bills assigned to Senate Finance Committee in the 117th Congress.

(c) NAICS code distribution of firms lobbying for bills assigned to Senate Banking Committee in the 117th Congress.

Figure 10: NAICS code distribution of Senator and Committees.

Therefore, I conclude that by estimating the causal effect of committee assignment on the cross-entropy between committees and senators, we can determine whether the information funneled through committee in congressional activity, after joining such committee, affects congressmen’s trading behavior or not, provided that we properly control for the level of congressional knowledge before joining the committee.

Then there exists a remaining question: how to measure congressional knowledge "before" joining a committee? Before proceeding, I assume that the congressional network G described in Section 2 is a comprehensive source of relevant information, meaning that it is sufficient to satisfy the backdoor-adjustment criterion for correctly estimating the causal effect of committee assignment on the outcome. However, although G includes this sufficient information, we need to separate the congressional knowledge "before" and "after" joining the committee. Additionally, there is another challenge: G is a discrete object that cannot be directly used in any model to estimate the quantity of interest in its raw format. To overcome these challenges, I suggest using Graph Neural Network (GNN) to learn the corresponding numerical representations that correspond to the congressional knowledge "before" joining the committee.

Figure 11 briefly illustrates how Graph Neural Networks (GNN) work. Given an original graph whose nodes are represented as x_i, we learn the corresponding numeric representation of x_i, which is $h_i \in \mathbb{R}^k$, by aggregating information from the neighboring nodes. In the case of x_2, for example, we aggregate information from its neighbors (x_1, x_5, x_6) to generate h_2. Over the layers of the neural network, this process is repeated until we learn the representation that achieves the best performance in the downstream task, in this case estimating the potential outcomes.

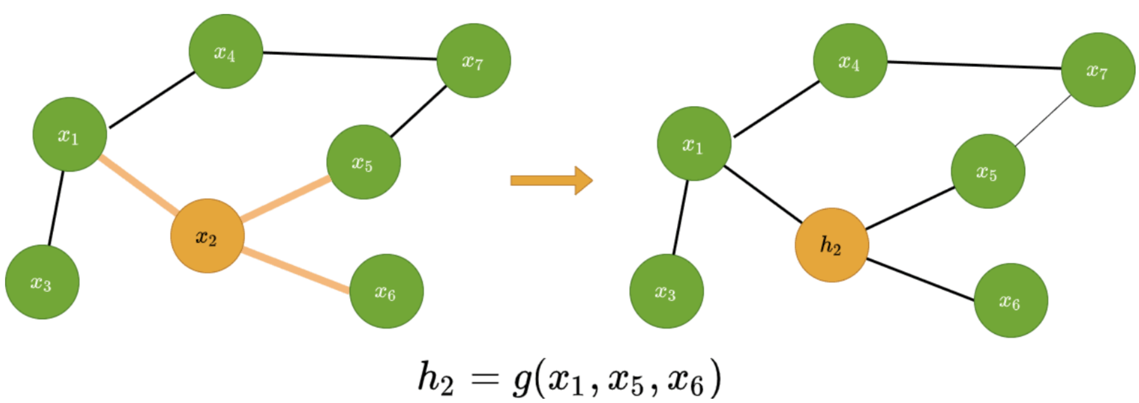


Figure 11: A brief illustration of how Graph Neural Networks (GNN) work is presented below. The diagram is borrowed from `https://theaisummer.com/gnn-architectures/`

Then general idea to learn the representation of congressional knowledge before joining a committee is to prepare additional nodes that encode the information of each congress, and learn the joint representation that aggregates a senator's information and those of congresses before joining such committees, as illustrated in Figure 12.

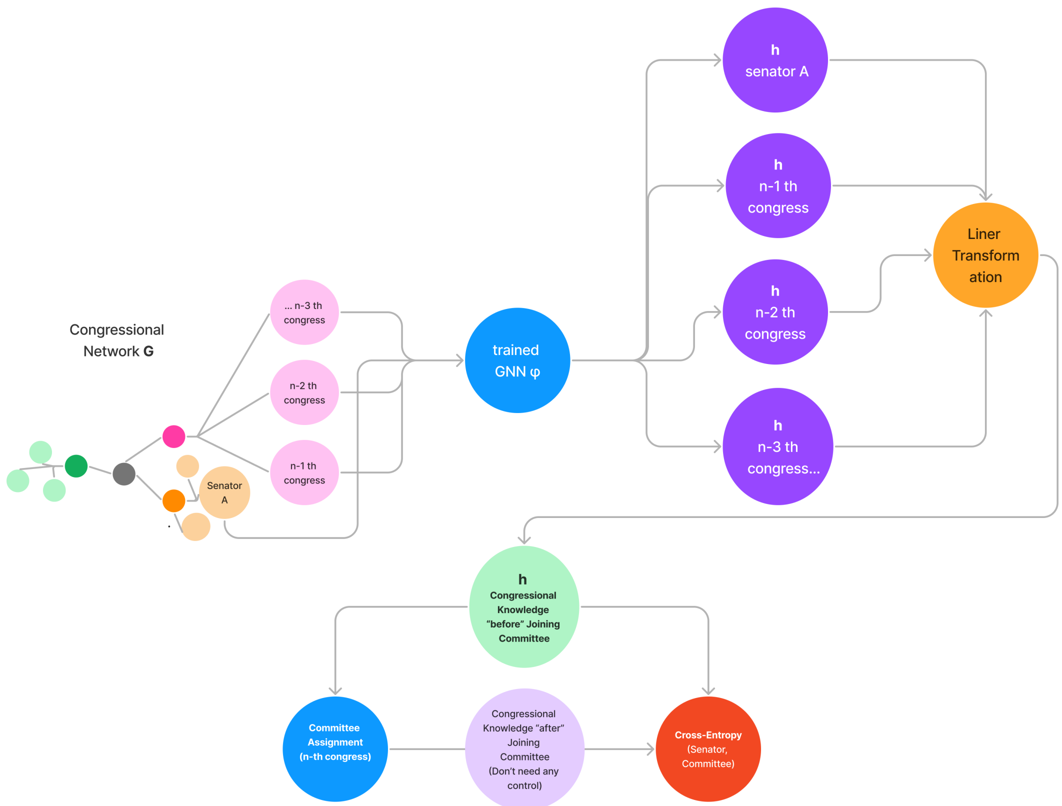


Figure 12: An illustration of how to utilize GNN to learn the numeric representation of congressional knowledge before joining a committee

3.4 Estimation of CATE

Finally, by denoting the learned representation of Senator i's congressional knowledge before joining the committee at time t as $h_i = \phi_{GNN}(i, G; t-1)$, the study can estimate the conditional average treatment effect (CATE) by learning estimators for both $\mu_0(h_i) \equiv \mathbb{E}[Y_i \mid H = h_i, T_i = 0]$ and $\mu_1(h_i) \equiv \mathbb{E}[Y_i \mid H = h_i, T_i = 1]$, and using the formula $\text{CATE}(h_i) = \mu_1(h_i) - \mu_0(h_i)$. This allows the study to estimate the effect of committee assignment on the similarity between a Senator's and the assigned committee's NAICS code distribution, while controlling for potential confounders through the learned representation $h_i = \phi_{GNN}(i, G)$. For the estimator $\mu_0(h_i)$ and $\mu_1(h_i)$, the study will use the Treatment Agnostic Regression Network (TARNet) (Shalit et al., 2016), which extends T-learners (Johansson et al., 2020) with additional layers of neural networks that learn a shared representation of $h_i = \phi_{GNN}(i, G; t-1)$

for the treatment and control groups in a balanced way. By doing so, the neural network can effectively perform its regression task to estimate CATE, even though h_i is asymmetrically distributed between the treatment and control groups.

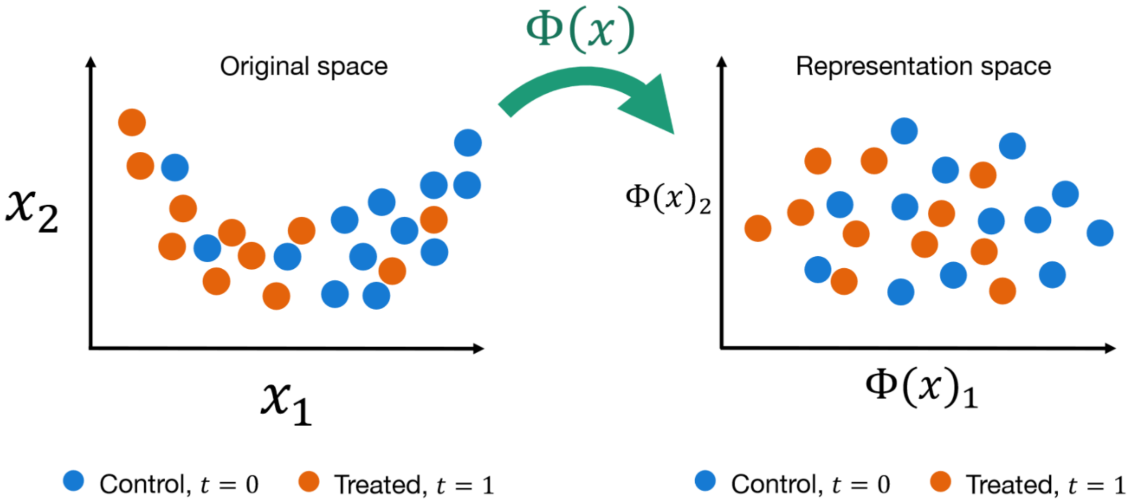


Figure 13: An illustration of how balancing works using TARNet's shared representation. The diagram is borrowed from (Koch et al., 2021).

The necessity of adding an additional balancing layer is quite straightforward. For a given committee, the number of senators assigned to that committee is relatively small compared to those who are not assigned to it. By incorporating additional layers with appropriate loss functions, it is expected that h_i will be mapped into a more balanced space, where $\phi_{\text{balancing}}(h_i; T = 0)$ and $\phi_{\text{balancing}}(h_i; T = 1)$ are evenly distributed as illustrated in Figure 13.

3.5 Evaluation of CATE

There are a total of 62 different committees, and CATE is estimated for each committee separately using the transaction information of 95 Senators across various congresses, spanning from the 114th to the 117th Congress. By evaluating these CATE values, the paper aims to quantitatively assess how committee assignments heterogeneously affect Senators' behavior across committees and among the Senators themselves.

4 Conclusion and Contributions

The paper aims to make two main contributions. Firstly, it substantively proves the existence of insider trading by explicitly showing that committee membership changes affect Congressmen's investment behavior to become similar to the information available from the committee. Secondly, it demonstrates a novel methodological approach to estimate causal quantities using graph-structured data by leveraging representation learning via Graph Neural Network. Therefore, the paper proposes a model and learning scheme that combines a Graph Neural Network and meta-learners to learn CATE, conditioning on the output of the Graph Neural Network.

References

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

Eggers, A. and Hainmueller, J. (2012). Capitol losses: The mediocre performance of congressional stock portfolios, 2004-2008. *Journal of Politics*, 75.

Hall, R. L. and Deardorff, A. V. (2006). Lobbying as legislative subsidy. *American Political Science Review*, 100(1):69–84.

Johansson, F. D., Shalit, U., Kallus, N., and Sontag, D. A. (2020). Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *CoRR*, abs/2001.07426.

Kim, S. H. (2012). The last temptation of congress: Legislator insider trading and the fiduciary norm against corruption. *Microeconomics: Welfare Economics & Collective Decision-Making eJournal*.

Koch, B., Sainburg, T., Geraldo, P., Jiang, S., Sun, Y., and Foster, J. G. (2021). Deep learning of potential outcomes. *CoRR*, abs/2110.04442.

Shalit, U., Johansson, F. D., and Sontag, D. (2016). Estimating individual treatment effect: generalization bounds and algorithms.

Smith, R. A. (1995). Interest group influence in the u. s. congress. *Legislative Studies Quarterly*, 20(1):89–139.

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):86–110.

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.