Donor Network Analysis: A Novel Approach to Measure Political Ideology

Viktor Due Pedersen
vipe@itu.dk

15 December 2023, Copenhagen, Denmark

1 Abstract

In an era of declining public trust in political institutions, this study probes the relationship between political donations and the ideological stances of politicians. Utilizing data from LittleSis and Voteview, it examines the 117th Congress to explore whether the topology of donation networks can reveal insights into a politician's ideological leanings. The analysis employs both standard and Laplacian Euclidean distances to assess the ideological positioning of politicians based on their donation histories and the underlying network structures. The study finds that donation patterns excluding network topology to a high degree align with Voteview's ideological scores, Including the topology can effectively distinguish between Republicans and Democrats, but do not consistently align with Voteview's ideological scores. Instead, the integration of network topology in the analysis uncovers additional dimensions of political ideology, suggesting that a politician's ideological position is shaped by a complex interplay of factors beyond financial contributions. This research contributes to the understanding of political behavior, highlighting the nuanced relationship between political funding, party affiliation, and ideology.

2 Introduction

Since the National Election Study began surveying public trust in politicians, it has been falling. As of November 6th, 2023, only 16% of the American people say they *trust the government to do what is right just about always/most of the time* [1]. This is an all-time low. Several factors contribute to this dwindling trust, one of which is the perception that politicians are influenced by external entities such as corporations or powerful individuals.

Voteview [2] offers an interesting lens to examine political behavior, providing a unique score for each politician on a scale from -1 to 1, the so-called nominate score. This score reflects their voting patterns in the House of Representatives or the Senate. However, the question arises: can a politician's ideological score be accurately inferred from their donation history alone?

This paper investigates the possibility of estimating a politician's ideology score based solely on their received donations and the topology of the network of nodes in which each politician exists. The primary data source for this analysis is LittleSis [3], described as "*an involuntary Facebook of the 1%*," which maps the network of donations from various companies, organizations, and individuals. Focusing specifically on the 117th Congress, this study aims to determine whether the Voteview score can be replicated using donation data alone, thereby providing insight into the influence of financial contributions on political ideology and actions.

I show that the collection of donations without network topology is adequate to clearly divide Republicans and Democrats and that the estimates are similar to the Voteview score. I demonstrate that including the topology of the network does not approximate the Voteview score but potentially approximates a different dimension of political ideology.

3 Related Work

Voteview [2], a project by the Department of Political Science at UCLA, endeavors to quantify political ideology on a scale ranging from conservatism (1) to liberalism (-1). The metric *nominate_dim1* specifically interprets this scale as a measure of economic conservatism versus liberalism. In the US, where the political landscape is predominantly divided into Democrats and Republicans, it's commonly perceived that Democrats lean towards liberal economic ideologies, while Republicans incline towards conservatism. Figure 1 illustrates the distribution of politicians along this economic ideology spectrum, as calculated by Voteview. It's important to note, however, that Voteview's methodology assumes a static ideological position for politicians, an assumption not universally accepted. Given that my analysis does not confine donations to a specific time period, adopting an 'average ideology' as a static metric seems most appropriate. Voteview's nominate scores are derived from roll-call voting patterns, a method pioneered by Poole in the 1980s [4–6], essentially representing the principal component of a politician's voting record.

The LittleSis database, curated by the Public Accountability Initiative [7], has been the focus of various studies ranging from the philanthropy of the ultra-rich [8] to the analysis of donations to art museums [9]. One notable study [10] examines evolving patterns in U.S. political donations, highlighting a significant increase in donation amounts, polarization, and concentration. This research underscores a shift over the past 35 years towards a landscape increasingly dominated by a few 'mega-donors'.

In the field of complex networks, a typical approach of describing a node involves analyzing a node's neighborhood. This includes employing metrics like degree centrality, which deems a node significant if it has numerous connections. Additionally, algorithms such as PageRank extend this notion by not only considering the quantity but also the quality of these connections. Path-based methods, such as

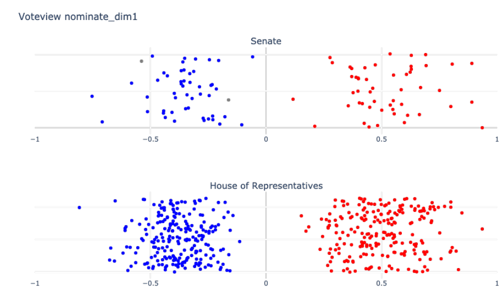

Figure 1: The Voteview *nominate_dim1* variable for Republicans and Democrats, representing the first dimension (often interpreted as economic liberalism-conservatism) of politicians as estimated by NOMINATE. Note that the y-coordinate in this 1D plot merely represents point density and does not encode an additional variable.

closeness centrality (the inverse of the average shortest distance to all other nodes) and betweenness centrality (measuring the frequency of a node appearing on shortest paths), are also utilized. Both neighbor- and path-based methods are sensitive to the overall network structure.

Recent research has explored methods to measure distances between nodes represented as vectors [11, 12] providing another approach to describe a node. These methods treat nodes as vectors and employ distance metrics, like the Euclidean distance, to measure similarity between nodes. They also propose incorporating network topology into distance calculations for enhanced accuracy. My project aims to apply similar methods, integrating network topology into the analysis of political donation data.

4 Data Description

The network provided by LittleSis [3], is a weighted un-directed network. Each of the nodes represents a politician, company, individual or organization. There exist 12 different kinds of relationship types connecting nodes in many different ways ranging from *familiy*, *ownership* and *donations*. All 12 relationship types are included in the network, but when I collect the list of donors I only inspect two of the types:

5. Donation/Grant: A gift transfer of money, goods, or services with nothing due in return. Examples - political funding, contributions to charities, government grants, prizes [13].
7. Lobbying: When an organization directly lobbies a government agency or official. Examples - organization that employs lobbyists in-house, lobbying firm hired by an organization [13].

The network consists of multiple connected components, but all politicians from the 117th congress, as well as all their donors are included in the biggest connected component. The entire analysis is therefore performed on either that component or subgrahs of it.

The initial graph has $413,057$ nodes and $1,685,094$ edges. It consists of $51,414$ connected components where the biggest contain 85.83% of the nodes. When I exclude all nodes except the biggest connected component the graph consist of $354,538$ nodes and $1,677,720$ edges.

5 Methods

This section is a collection of the methods I use throughout the paper. Section 5.1 provides the overview of the entire pipeline of methods that are generally used. Section 5.2 explain the methodology behind the distance metric of vectors and 5.3 then explain the projection methodology. Finally section 5.4 describe the node ranking strategy used for the discussion.

5.1 Data transformation from network to ideology estimation

All three experiments are done on the vector representation of a politician. Each politician is therefore represented by the donations they receive illustrated with a toy example in figure 2. All donors to a politician is given a unique index in the vector and the amount is inserted. Since some politicians have received 0 dollars in donation and others received 96 million dollars in donations, the values are increased by 1 and log-transformed such that 96 million become 7.98 and 0 dollars stay 0. This process is illustrated in figure 2.

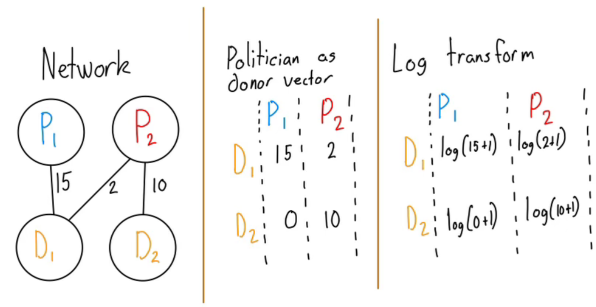

Figure 2: The process of converting a politician to a donor vector. In the figure P_1 is the first politician, P_2 the second etc. and D_1 and D_2 donor 1 and 2.

The distance between each pair of politician vectors are now calculated using the method described in section 5.2 and projected to one and two dimensions with T-SNE described in section 5.3.

Then the projections are pushed to a range of -1 to 1 by first normalizing to a range of 0 to 1 $X'_{ij} = \frac{X_{ij}-\min(X)}{\max(X)-\min(X)}$ and thereafter -1 to 1: $X''_{ij} = 2X'_{ij} - 1$. For an example of these calculations see section 9.1 in the appendix.

Since t-SNE, when projecting the data, do not have any knowledge about the structure of Voteview I added a minor change. Republicans and Democrates was *flipped* when projected so I multiplied all values with -1 if the values was flipped.

5.2 Node Vector Distance

This paper employs methods from [11] for computing distances between nodes, using both standard and Laplacian euclidean distances. These distances are based on the generalized euclidean form:

$$\delta_{A_{t1},A_{t2}} = \sqrt{(A_{t1} - A_{t2})^T Q (A_{t1} - A_{t2})} \quad (1)$$

where Q can be substituted depending on the method.

In the case of the *standard* euclidean, Q is simply the identity matrix I, since A_{t1} and A_{t2} would live in an undistorted space, with no weighting of any sort. A substitution can be written, perhaps more familiarly, as: $\delta_{A_{t1},A_{t2}} = ||A_{t1} - A_{t2}||^2$

In the case of Q being the Laplacian matrix L, the interpretation is significantly different. The Laplacian is defined as $L = D - A$ where D is a diagonal degree matrix and A is the adjacency matrix. The Laplacian is an important matrix in network science for many reasons outside of the scope of the current project, but is used to incorporate the network topology as a weight for each of the dimensions, when representing the donors as a separate dimension in the vector.

To be precise, it is not the Laplacian L, but rather the Moore-Penrose pseudoinverse L^+, I need. The Laplacian matrix can represent a flow of information [14], and [12] argues that the inverse L^{-1} represents distances. The issue is that we can not simply take the inverse of L as L is singlular and therefore can not be inversed. The Moore-Penrose pseudoinverse L^+ can then be used, but again a problem arise. This pseudoinverse comes with a computational cost I can not accept.

This naturally leads me to use an approximation of L^{-1} rather that L^+. In order to approximate the Laplacian, I use the approxchol_lap function from the `Laplacians` julia package. The specifics of their method of approximating the Laplacian can be studied in [15]. The method only works for connected components and the full graph is therefore never used. In general the Laplacian can be intractable to calculate and approximations are therefore common for large networks. While Laplacian solvers do not provide exact solutions, they have been found to provide approximations that are negligible in their impact for real-world networks. [16].

When using the Laplacian in our context, the calculated euclidean distance reflects not only the donation disparities between politicians (as represented by A_{t1} and A_{t2}), but also the networks topology. Thus, this approach integrates quantitative donation data with the qualitative structure of the political network, offering a comprehensive view of donation dynamics influenced by network connectivity.

5.3 T-SNE

In order to project the node vectors, with all the donation information, to a single dimension I used the *t-Distributed Stochastic Neighbor Embedding* or simply T-SNE algorithm. T-SNE was first introduced in [17], and has gained a lot of popularity.

Even though t-SNE are difficult to initialize and tend to prioritize global structure rather than the local structure [18], I decided to use it based on its ability to take a precomputed distance matrix as input. When t-SNE tries to project a dataset to lower dimensions it require a pairwise comparison. By providing this myself, I can decide what kind of similarity the t-SNE should consider. As described in section 5.2 the *similarities* are euclidean distance and Laplacian euclidean distance.

One thing that is important to note is that two very different datasets can in theory have exactly the same distance matrix. Namely for this reason it becomes increasingly important to embed network topology in the distance matrix as done with the Laplacian euclidean distance calculations.

5.4 Node Ranking

In general when inspecting importance of nodes I use the PageRank of the given node. PageRank is an algorithm, utilizing random walks to estimate the importance of a given node. It was developed for directed networks, but can be applied to undirected aswell. In that case an edge is converted to a birectional edge. PageRank is a global measure, but has been shown to be approximated well, by inspecting the local in-degree (or simply degree for un-directed networks) of a node [19]. This metric is used in section 7 for discussing individual node importance.

6 Results

6.1 Politicians as donor vectors

The first attempt to estimate the ideology was the simple euclidean distance where each vector is a politician and the values are the log transformed donation, where each index is a donor as described in section 5.1. It therefore include no other information that the donations and no understanding that the data is modelled as a network.

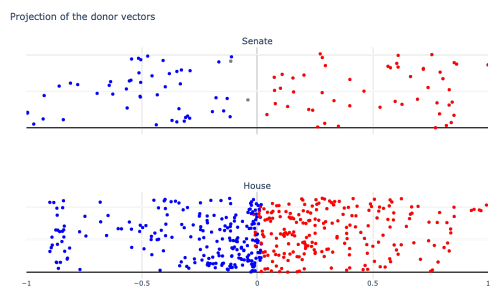

Figure 3: Projection of the euclidean distance matrix between each of the politicians for Democrates and Republicans. Note that the y-coordinate in the 1d plot merely represent point density, and does not encode a variable.

As seen in figure 3 the parties are indeed clearly divided, indicating that there are in-fact a difference between who donates to what parties. Dividing the parties is one thing, but has the t-SNE embedded an ideological meaning in the projection? And if so, is it the same as Voteview approximated?

The absolute distance between the Voteview score and my estimation (figure 4) are pretty close. With a median under 0.24 for all except Democrates in the House of Representatives, I would argue that the scores are sufficiently close. This indicates that even though no topology of the network is included in the distance matrix, the nominate score from Voteview is possible to approximate.

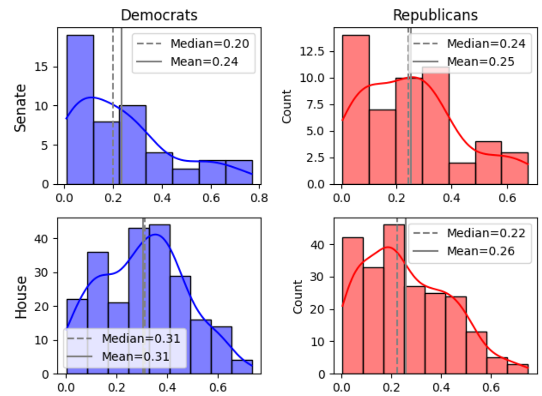

Figure 4: Absolute distances between Voteview nominate_dim1 and the 1 dimensional projection of the euclidean distances of donation vectors.

6.2 Politicians as a network vector

But LittleSis allow us to extend the estimations for more advanced methods. Since LittleSis provide a network of politicians and its donors, the distance metrics can include the Laplacian as explained in 5.2.

What I define as the network is not the entire network, but rather the largest connected component that include all politicians from the 117th congress. Since the vector represents the entire network, but every node in the network haven't donated, a lot of noise will be added as well. The hypothesis before projecting the distance metric would therefore be that the estimations are worse. This is due to the vast amount of noise that have suddenly been added, even though the Laplacian should help weighing the values.

Figure 5 clearly shows that the vectors contain to much noise to clearly divide the parties. The 1 dimensional projection for the Senate are able to pull many of the republicans towards the right, but they end up in the middle with all the Democrats completely towards the left. Projecting the distances of 557 politicians to a single number is not an easy task for the t-SNE algorithm with a vector being as big as the number of nodes. The scores in figure 6 are also so scattered that arguing for any comparison with Voteview would be ambigious.

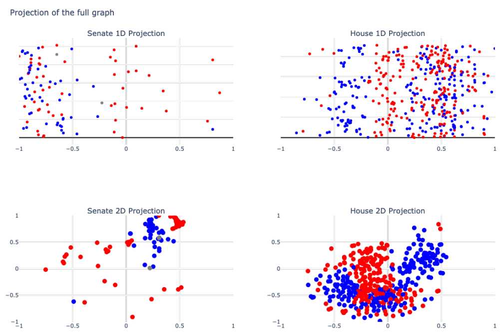

Figure 5: Projection of the Laplacian weighted euclidean distance matrix between each of politicians for Democrates and Republicans. Note that the y-coordinate in the 1d plot merely represent point density, and does not encode a variable.

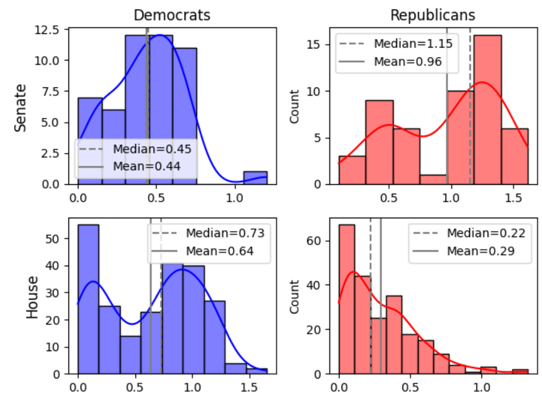

Figure 6: Absolute distances between Voteview nominate_dim1 and the 1 dimensional projection of the Laplacian weighted euclidean distances of full graph donation vectors.

For this reason I tried to project the distances to a two dimensional space to see the capabilities in a higher dimension. In this we see that many republicans are grouped in the House of Representatives, with some Democrats clustered as well. The division is in no way clear, but indicate that a projecting to two dimensions are more meaning full. Comparing this to the Voteview nominate value is not possible but measuring ideology can be done in many dimensions. Perhaps the full network does not tell a story about the nominate, but another metric, which is shown in the two dimensions.

6.3 Politicians as a reduced network vector

With the inability to estimate the nominate value from the full graph structure using the Laplacian matrix, perhaps it is possible using a network of only the politicians and the donors. I therefore excluded all nodes, not being a politician or a donor and conducted the same analysis as with the full graph. Note that the analysis can only be done on a connected component. I therefore excluded the seven politicians with no data about donations. The network now contains $41,439$ nodes and $712,705$ edges.

Projecting the distance matrix to one dimensional in figure 7 visually looks equally random. But as we see in figure 8, the absolute distance increase for both parties in the House of Representatives. For the Senate we see that the distance drops significantly for the republicans while staying almost the same for the democrats. These results again indicate that the network structure does not tell a story about the nominate variable. Again it seems more probable that the topology of network is simply another than the nominate estimation.

When inspecting the two dimensional plot in figure 7, t-SNE are now able to clearly divide the network. We can therefore conclude that the topology of the network do in fact represent a difference between the parties - just not the nominate difference.

7 Discussion

When dealing with such a complex network with so many nodes and processing steps the error analysis and discussion of the finer details of the project is important.

I do see the best performance in section 6.1 and 6.3, concluding that the performance do indeed depend on the structure of the network and the representation. How I choose to build the network, can

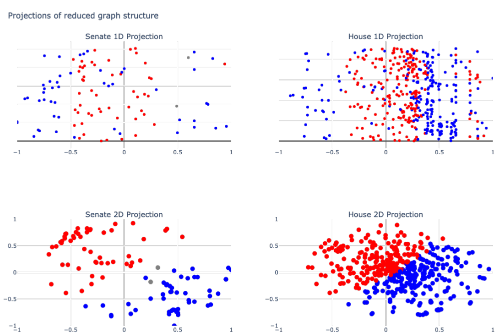

Figure 7: Projection of the Laplacian weighted euclidean distance matrix for the reduced network of only politicians and donors. The color represents Democrats and Republicans. Note that the y-coordinate in the 1d plot merely represent point density, and does not encode a variable.

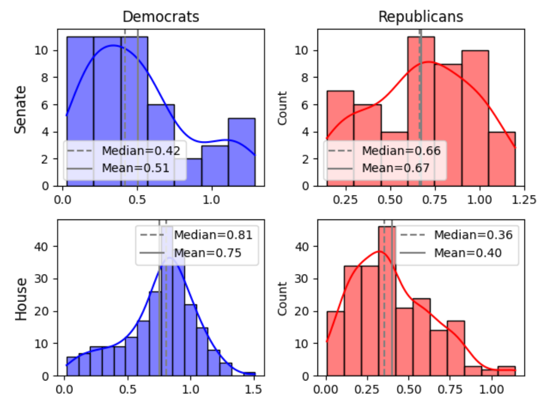

Figure 8: Absolute distances between Voteview nominate_dim1 and the 1 dimensional projection of the Laplacian weighted euclidean distances of donation vectors.

Chamber	Party	Total
House	Democrats	\$141.76M
	Republicans	\$100.39M
Senate	Democrats	\$185.07M
	Republicans	\$294.17M

Table 1: The total amount donated to each party for each chamber.

therefore make a big difference.

Throughout the paper I visualize the Senate and the House of Representatives as separate entities. I should have done that as well in the methodology. Separating the networks, vector distance calculations and projections, would potentially allow the methodology to actually catch the differences that could potentially exist between the two institutions. The institutions are similar in the sense that they are both elected by the public and both have to agree when passing legislation. The Congress consists of 100 senators (two from each state) and the House of Representatives consist of 435 voting members. The difference in size already propose an important structural difference.

And then the discrepancies in the amount of donation each party and chamber have received. Table 1 clearly shows that, the amount of money donated to each chamber is uneven both in regards to chamber as well as party. This indicate that the intrinsic nature of donating are different for the Senate and the House [10].

The fact that the Senate score decrease so drastically for the republicans from figure 6 to 8, could be an indicator that the network is stronger towards the senate and especially the Republicans. When inspecting the top 2% of the politicians who have received the highest total donations in table 2, it become clear that the Senate receives a greater amount of donations.

The donation amount of the senate are larger to such a degree, that even when log transformed the values could potentially weigh the t-SNE such that the Senate have more *weight* in the projections.

One thing that become clear as well, is that donation activities are not only bound by either the Senate or the House but are historical as well. Consider Mitt Romney, receiving \$$50M$, more than Kelly Loeffler. He ran for president in 2012 against Barack Obama. A large proportion of the donations are properly from that period and not since 2019 where he has served as a senator. Since the Voteview scores are based on voting behavior, the donations could potentially

Name	Chamber	Party	Total
Mitt Romney	Senate	Republican	$96.82M
Kelly Loeffler	Senate	Republican	$47.29M
Chuck Schumer	Senate	Democrat	$18.77M
Nancy Pelosi	House	Democrat	$17.83M
Mitch McConnell	Senate	Republican	$13.50M
Rob Portman	Senate	Republican	$10.59M
Kirsten Gillibrand	Senate	Democrat	$10.53M
Daniel Scott Sullivan	Senate	Republican	$10.16M
Kevin McCarthy	House	Republican	$10.08M
Ted Cruz	Senate	Republican	$8.93M
Marco Rubio	Senate	Republican	$8.41M
John Cornyn	Senate	Republican	$8.36M

Table 2: The politicians that have received top 2% of the donations.

have been limited to the time served in either of the chambers. Modelling ideology should probably be modelled as a temporal problem rather than a static.

Measuring absolute distance between donation ideology and the Voteview estimation relies on the assumption that Voteview has the *true* estimation. Ideology can be expressed in many ways. It could be that the donor projections are not to be compared with nominate score. Due to the political landscape of the U.S. some states are considered either republican, democratic or swing-states, i.e. states where the republicans are likely to win, vice versa or both have a reasonable chance. A Democrat elected in a republican state are likely to have an ideology closer to republicans while still being a democrat. Further analysis could be made to inspect the estimation of such specific individuals for a more nuanced view on the one dimensional scores. Representing a Democrat with a slight republican-ideology are likely to be easier in two dimension rather that one.

When inspecting the ten highest pagerank scores from the biggest connected component, Mitt Romney stands out, with a significantly higher score than the rest, again consolidating the weight he carries in the network. We clearly see that the network is not solely a donation network as Joe Biden have the third highest degree but are one in the top 2% receivers of donations. There is a large overlap of politicians between table 2 and 3, that does indicate that donations are important for the network.

Party	Politician	PageRank
Republican	Mitt Romney	0.00214
Democrat	Chuck Schumer	0.00089
Democrat	Joe Biden	0.00047
Republican	Mitch McConnell	0.00047
Democrat	Kirsten Gillibrand	0.00043
Republican	Marco Rubio	0.00039
Democrat	Dianne Feinstein	0.00036
Democrat	Maria Cantwell	0.00035
Republican	Rob Portman	0.00035
Democrat	Elizabeth Warren	0.00034

Table 3: Politicians with the ten highest pagerank value

8 Conclusion

This study embarked on an exploration to determine if a politician's ideology could be quantified based on their donation history. The analysis was twofold: it considered the donation data in isolation and in conjunction with the topology of the network from which this data was derived. The findings revealed that while donation patterns alone offer a partial approximation of Voteview scores, the inclusion of network topology uncovers additional dimensions of political ideology that Voteview does not capture.

It is important to note, however, that this study does not conclusively establish a direct impact of political donations on a politician's decisions or actions. Instead, the donation patterns primarily serve as indicators of party affiliation.

Link to code: `https://github.com/DueViktor/DonorIdeo`

References

1. Pew Research Center for the People and the Press. *Public Trust in Government: 1958-2023* Accessed: 2023-11-27. Sept. 2023. https://www.pewresearch.org/politics/2023/09/19/public-trust-in-government-1958-2023/.
2. Lewis, J. B. *et al. Voteview: Congressional Roll-Call Votes Database* Accessed: 2023-09-01. 2023. https://voteview.com/.
3. LittleSis.org. *LittleSis is a free database detailing the connections between powerful people and organizations* Online resource. Sept. 2023. https://littlesis.org.
4. Poole, K. T. *Spatial models of parliamentary voting* (Cambridge University Press, 2005).
5. Poole, K. T. Recovering a basic space from a set of issue scales. *American Journal of Political Science,* 954–993 (1998).
6. Poole, K. T. & Rosenthal, H. A spatial model for legislative roll call analysis. *American journal of political science,* 357–384 (1985).
7. Public Accountability Initiative. *Public Accountability Initiative* [Online; accessed 12-December-2023]. 2023. https://public-accountability.org/.
8. Coupe, T. & Monteiro, C. The charity of the extremely wealthy. *Economic Inquiry* **54,** 751–761 (2016).
9. Barabási, A.-L. & Shekhtman, L. Who Supports American Art Museums? Introducing a New Dataset and Data Sources about Museum Funding.
10. Zakir, N. *The economics of political donations* PhD thesis (University of Leicester, 2019).
11. Coscia, M., Gomez-Lievano, A., Mcnerney, J. & Neffke, F. The node vector distance problem in complex networks. *ACM Computing Surveys (CSUR)* **53,** 1–27 (2020).
12. Coscia, M. *Generalized Euclidean measure to estimate network distances* in *Proceedings of the International AAAI Conference on Web and Social Media* **14** (2020), 119–129.
13. Public Accountability Initiative. *Navigating the Relationship Page* https://littlesis.org/help/relationships. Accessed: 1 Oct 2023. 2023.
14. Coifman, R. R. & Lafon, S. Diffusion maps. *Applied and computational harmonic analysis* **21,** 5–30 (2006).
15. Gao, Y., Kyng, R. & Spielman, D. A. Robust and Practical Solution of Laplacian Equations by Approximate Elimination. *arXiv preprint arXiv:2303.00709* (2023).
16. Coscia, M. & Devriendt, K. Fast Node Vector Distance Computations using Laplacian Solvers. *arXiv preprint arXiv:2310.11222* (2023).
17. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9** (2008).
18. Wattenberg, M., Viégas, F. & Johnson, I. How to use t-SNE effectively. *Distill* **1,** e2 (2016).
19. Fortunato, S., Boguñá, M., Flammini, A. & Menczer, F. *Approximating PageRank from indegree* in *Algorithms and Models for the Web-Graph: Fourth International Workshop, WAW 2006, Banff, Canada, November 30-December 1, 2006. Revised Papers 4* (2008), 59–71.

9 Appendix

9.1 Example of transformation from 5.1

$$X = \begin{pmatrix} 0.5 & 0.8 \\ 0.3 & 0.7 \end{pmatrix}$$

$$\min(X) = 0.3, \quad \max(X) = 0.8$$

$$
\begin{aligned}
X' &= \frac{X - \min(X)}{\max(X) - \min(X)} \\
&= \frac{\begin{pmatrix} 0.5 & 0.8 \\ 0.3 & 0.7 \end{pmatrix} - 0.3}{0.8 - 0.3} \\
&= \begin{pmatrix} \frac{0.5-0.3}{0.5} & \frac{0.8-0.3}{0.5} \\ \frac{0.3-0.3}{0.5} & \frac{0.7-0.3}{0.5} \end{pmatrix} \\
&= \begin{pmatrix} 0.4 & 1.0 \\ 0.0 & 0.8 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
X'' &= 2X' - 1 \\
&= 2 \cdot \begin{pmatrix} 0.4 & 1.0 \\ 0.0 & 0.8 \end{pmatrix} - 1 \\
&= \begin{pmatrix} -0.2 & 1.0 \\ -1.0 & 0.6 \end{pmatrix}
\end{aligned}
$$