Applied Network Science

Coalitions and coordination in Washington think tanks: board interlock among Washington D.C.-based policy research and planning organizations

Alexander C. Furnas*

*Correspondence:
zfurnas@gmail.com
Center for Science of Science
and Innovation, Kellogg
School of Management,
Northwestern University,
Chambers Hall, 600 Foster St,
Evanston, IL 60208, USA

Abstract

Think tanks have become central players in the political and policy ecosystem of the United States, yet the communication and coordination strategies and connections between them remain relatively unexamined. This paper begins to remedy that by using IRS 990 data to construct and analyze the largest board interlock network among policy planning organizations to date. An analysis of 277 policy research organizations located in Washington D.C. reveals high levels of resource seeking, issue-area coordination and weak evidence of ideological coordination among these substantively important organizations in American politics.

Keywords: Networks, Interlock, Interest groups, Think tanks, ERGM

Introduction

The last 40 years have seen a precipitous rise in both the number and influence of policy-planning organizations, often called "think tanks" (Medvetz 2012). Recent research in American politics has begun to reckon with the role that policy-planning organizations play in substantive policy-making and party politics. Indeed, incoming Presidents routinely staff their administrations with high-level think tankers, and think tanks have been central actors in the rise of policies like Welfare Reform, and the Affordable Care Act (Medvetz 2012). Think tanks' allocation of attention across issues tends not only to reflect, but also to shape partisan issue ownership (Fagan 2019). Both the ideological perspective and proximity to power of a think tank shape how their work is used by Congress (Lerner 2018). Moreover, congressional staffers both disproportionately trust policy evidence from partisan-aligned think tanks (Furnas 2019), and are more likely to favorably evaluate petitioners presenting evidence from aligned think tanks (Furnas et al. Unpublished).

Despite this recent growth in attention to think tanks, the full ecosystem of these policy-planning organizations remains under-mapped. Following work in interest organization population ecology, and population demography, (Gray and Lowery 2000; Carroll

1984), this study catalogues policy research and analysis organizations active in Washington D.C. In particular, I focus on the connections between these organizations, examining how they coordinate and share information, as well as mitigate systemic risk

I take an inductive approach at how these central—yet understudied—actors in the American political landscape coordinate among themselves. I conduct the largest mapping of the Washington D.C. think tank ecosystem to date (Burris 2008). Following an approach common in organizational sociology, I leverage interlocking directorates of organizations to examine patterns of organizational coordination (Mizruchi 1996). Two organizations are said to have an "interlock" between their directorates when one (or more) people sit on the boards of both organizations. This signifies a strong organizational tie between these two groups. I construct and analyze the board interlock network for 277 Washington D.C.-based organizations, using IRS 990 disclosure data.

Prior interlock scholarship has tended to focus largely on connections between corporate boards (Mizruchi 1996). However, even in these contexts scholars have found that interlock has been central in enabling political activity (Mizruchi 1982), with highly interlocked directors being more active in policy associations (Useem 1979). Research on bank control over corporate boards has indicated that centrality in the interlock network can be understood as power or importance in the community (Mariolis 1975), with some arguing that these ties are particularly important as they facilitate social capital and access to information flowing through the network of organizations (Davis 1991; Mizruchi 1996).

A more thorough accounting of policy-planning organizations and the manner in which they coordinate is of central importance in understanding contemporary American political institutions. As legislative staff capacity has declined in members' offices (Crosson et al. 2019), policy-making in Congress has become centralized in party leadership (Curry 2015). We can understand this partisan legislating in the context of a recently prominent theory of American politics, which conceptualizes political parties as extended networks of policy demanding constituencies, organizations and interest groups (Bawn et al. 2012). This theory is usually applied to the role these networks play in setting party agendas or nominating and supporting candidates for elected office. To be sure, the nomination and election process serves a fundamental role in identifying and prioritizing issues for government attention. However, I argue that think tanks serve as the policy-apparatus of these extended party networks of policy-demanders. As such, we should expect them to coordinate along ideological dimensions in their attempts to support and influence partisan and ideological lawmakers (Noel 2014).

Prior interlock research among policy planning organizations has focused on small subsets of the full network, with particular attention given to the largest actors. Burris (2008) evaluated interlock among twelve prominent think tanks, and Salas-Porras (2018) focused on a more inclusive set of connections—beyond board interlock—among 33 economic policy think tanks during the financial crisis. These studies have found strong evidence of ideological coordination among these elite policy planning actors. In a study of policy networks around estuaries, Berardo and Scholz (2010) demonstrate that organizations connect both to popular and well-resourced organizations. These connections enable both information flow in the network, and the establishment of bonding structures that facilitate higher-stakes coordination.

This paper differs from this prior work on think-tank interlock in two important ways. First, I analyse a far larger set of organizations, as policy information, especially on niche issues, may come from players beyond the few most prominent. Current work gives us no sense of whether or how these organizations are integrated into the network. Second, I use exponential random graph modeling to more systematically evaluate the interlock network. Beyond simply describing the observed characteristics of the network (Burris 2008; Salas-Porras 2018), this enables me to conduct a statistical assessment of what features of individual organizations are directly associated with greater connectivity in this network.

Because of the inductive nature of this investigation of coordination and coalition formation among policy-planning organizations, I do not detail explicit theory-driven hypotheses to test. However, it is useful to lay out some general expectations about what evidence of different network structures and interlock patterns might indicate. Following the findings in Burris (2008) and Salas-Porras (2018), I expect that organizations will be more likely to foster ties with other organizations that are ideologically aligned with them. However, it remains to be seen if this effect extends to this broader set of organizations beyond the top, often quite ideological, players. I also expect organizations with access to greater financial resources to be more heavily interlocked, as access to resources is a substantial motivator of interlock in other contexts (Mizruchi 1996).

I do not have strong prior beliefs about the direction or magnitude of the relationships between other dyadic and organizational factors and the probability of connection. Common explanations for strategic interlock in the corporate context include preferences for diverse connections to increase information gathering breadth, and preferences for connections to other similar organizations to monitor competition, or engage in collusion (Mizruchi 1996). Similar explanations map to this case: Organizations may seek to build cross-issue coalitions increasing the scope of their information networks (issue heterophily), or they may choose ties within their local issue space to increase efficiency and avoid duplicated effort (issue homophily). Similarly, we might imagine justifications for either homo- or heterophilic preferences in other dimensions of organization type, like whether the organization engages in lobbying, has dues paying members, or hires contractors.

Following (Berardo and Scholz 2010), we can interpret the presence of centrally located bridging ties as indicative of the network's ability to facilitate information transfer and mitigate low-level risk, while a tendency for transitivity and triadic closure can be understood as offering the potential for greater collaboration within communities.

In this paper, I have three main goals. First, I set out to describe the scope and properties of the U.S. federal policy-planning network, at a much larger scale than has previous work. Second, I model the organizational factors which are associated with a greater embeddedness in this community. Finally, I detail the communication and coordination strategies that these interlock trends imply.

Describing the U.S. Federal Policy-planning network

In this project I construct and analyze the board interlock network of policy planning organizations headquartered in Washington D.C. I analyze all organizations active between 2008 and 2015 with average annual budgets over $100,000, which are classified

Fig. 1 Breakdown of research institutes & public policy analysis by NTEE-CC major codes
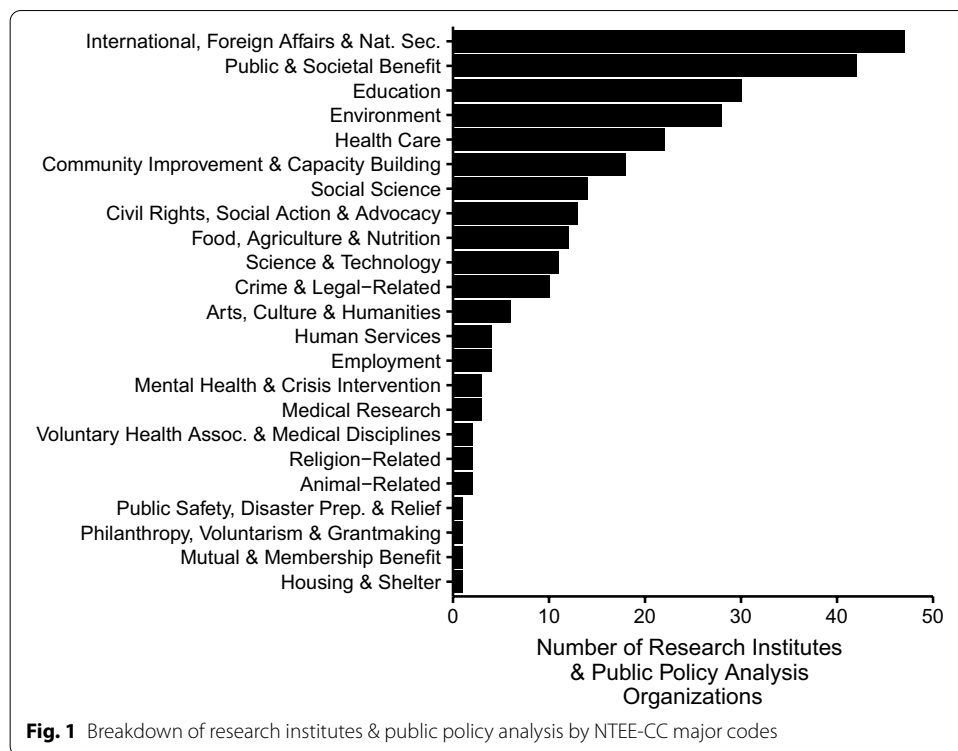
as subtype "Research Institutes & Public Policy Analysis" according to the IRS' National Taxonomy of Exempt Entities Core Codes (NTEE-CC). As part of their determination of tax exemption status, "Determination specialists" at the IRS classify organizations by type using the NTEE-CC. The first digit of the NTEE-CC is a letter A-Z, which represents the "Major Group," a broad sectoral categorization (e.g. Education, Environment, Crime & Legal-Related etc.). The second and third digit of the NTEE-CC subdivide organanizations by specific areas, organization types, or activities. The NTEE includes a set of, so-called, "common codes," that code for particular types of activity that are common across all major groups, such as advocacy, technical assistance, fundraising, or research. In this analysis, I include all organizations classified with the "05" common code for "Research Institutes & Public Policy Analysis," regardless of their major group.

Think tank data

Figure 1, shows the breakdown of the 277 research institutes and public policy analysis organizations included in this analysis by their major group. Notably, the most common type in the dataset are International, Foreign Affairs & National Security organizations, followed by Public & Societal Benefit, Education, and Environment.

Data about these organizations, including their classifications, complete listings of their boards of directors, annual revenues, membership dues, lobbying expenditures, and contracting comes from their Internal Revenue Service (IRS) form 990 mandatory annual disclosures, as collected and digitized by GuideStar. In addition to IRS 990 data, I gathered campaign contributions made by employees of these organizations, summed by party using data from the Center for Responsive Politics, and organizations' ideal points, IGScores (Crosson et al. 2020). IGScores are ideal points on a unidimensional,

Table 1 Summary statistics

Variable	Percent true	Percent false
Has IG score	28.9	71.1
Has members	14.8	85.2
Hires contractors	55.6	44.4
Engages in lobbying	90.6	9.4
Variable	**Mean**	**SD**
Total revenue	$4,085,478.76	$10,572,496.46
Membership dues	$61,838.29	$332,845.65
Lobbying spending	$12,712.53	$151,517.58
Number of contractors	1.39	3.80
Contributions to democrats	$21,936.59	90,503.17
Contributions to republics	$9934.49	60,659.49
IGscore (all)	− 0.07	0.88
IGscore (real)	0.06	1.09
IGscore (imputed)	− 0.12	0.77

left-right scale based on organizations' public position-taking activity between 2005 and 2016, which are particularly well-suited to this application (Crosson et al. 2020). Edges have been dichotomized to code for the existence of an interlock tie between organizations. Table 1, below, shows descriptive statistics for the variables which are used in the subsequent analysis.

As Table 1 shows, IGScores were only available for 28.9% of the organizations included in the analysis. Because of this missingness I employed multiple imputation using `Amelia` (Honaker et al. 2011), to generate imputed values for IGScores for think-tanks without scores. In service of a more accurate imputation model, I then collected two campaign finance variables for each organization using data from the Center for Responsive Politics: (1) total contributions to democratic candidates from the organizations' employees, and (2) total contributions to republican candidates from the organizations' employees. I found contributions to democratic candidates from 61.0% of organizations and contributions to republican candidates for 54.1% of organizations. Organizations with no contributions found from employees to either republicans or democrats were treated as missing. I also included variables for the average IGscore of the three and five organizations with the most semantically similar names in a 300 dimensional vector space constructed using latent semantic analysis (Furnas et al. 1988; Deerwester et al. 1990; Rehurek and Sojka 2011). Both the contribution variables and the proximate IGScore variables were used in the imputation model, along with average revenue, average membership dues, average number of contractors, average total revenue, NTEE code, and average end of year assets during the period 2008–2015. Contributions to democrats, contributions to republicans, revenue, membership dues, total revenue and end of year assets were logged, while NTEE was treated as a nominal variable. Bounds for the IGScores to be imputed as well as the campaign finance variables were set at the empirically observed minimum and maximum values in the dataset. I generate 100 imputed datasets and conduct my subsequent analysis in parallel across datasets, combining results.

Fig. 2 Real versus imputed IGScores for organizations in the interlock network
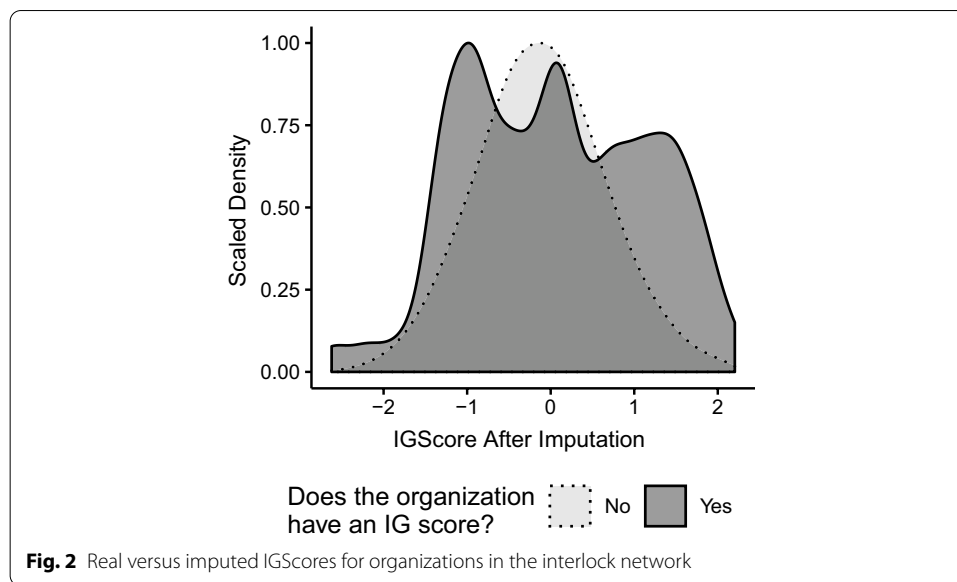
The distribution of real versus imputed IGScore values is shown in Fig. 2, below. The imputed values for missing IGScores is substantially more unimodal and centrally aligned than the distribution of actual IGScores. Additional imputation diagnostics are shown in Additional file 1: Electronic Supplementary Information "Appendix A, in Figures 7 and 8". It is important to note that because I use these scores to measure ideological distance between organizations, the moderate, unimodality of the imputed scores will tend to yield conservative estimates of ideological distance, and bias against finding substantive ideological results.

In expectation, overly moderate IGScores for these organizations will artificially shrink the distance between organizations—the key dyadic measure of ideological alignment. This should attenuate estimated effects, and as a result the test of dyadic alignment presented here is a conservative one.

Detecting think tank interlock

To facilitate the detection of interlocks, I standardized directors names using a three-stage process. First, I used key collision clustering with a two letter fingerprint with manual verification to correct for typos. Next, I used key collision clustering based on word-tokens, with manual verification to identify names that may have been entered in different formats (e.g. FirstName LastName vs. LastName, FirstName). This clustering and merging was done in OpenRefine (Verborgh and De Wilde 2013; Kusumasari et al. 2016). Finally, I removed titles, honorifics, and post-nominal letters or initials such as Esq., PhD or Jr., as they are applied extremely inconsistently throughout the data.[1] Following this name standardization procedure, the 277 think tanks included in this analysis had, in total, 9469 unique directors between 2008 and 2015. Two organizations were

---

[1] While it is certainly true that titles and post-nominal letters provide additional information which may distinguish between otherwise identically-named individuals, they are applied so inconsistently (even when the same person's name is being entered by the same organization in successive years) that matching on names including titles and post-nominals would induce substantial false negatives.
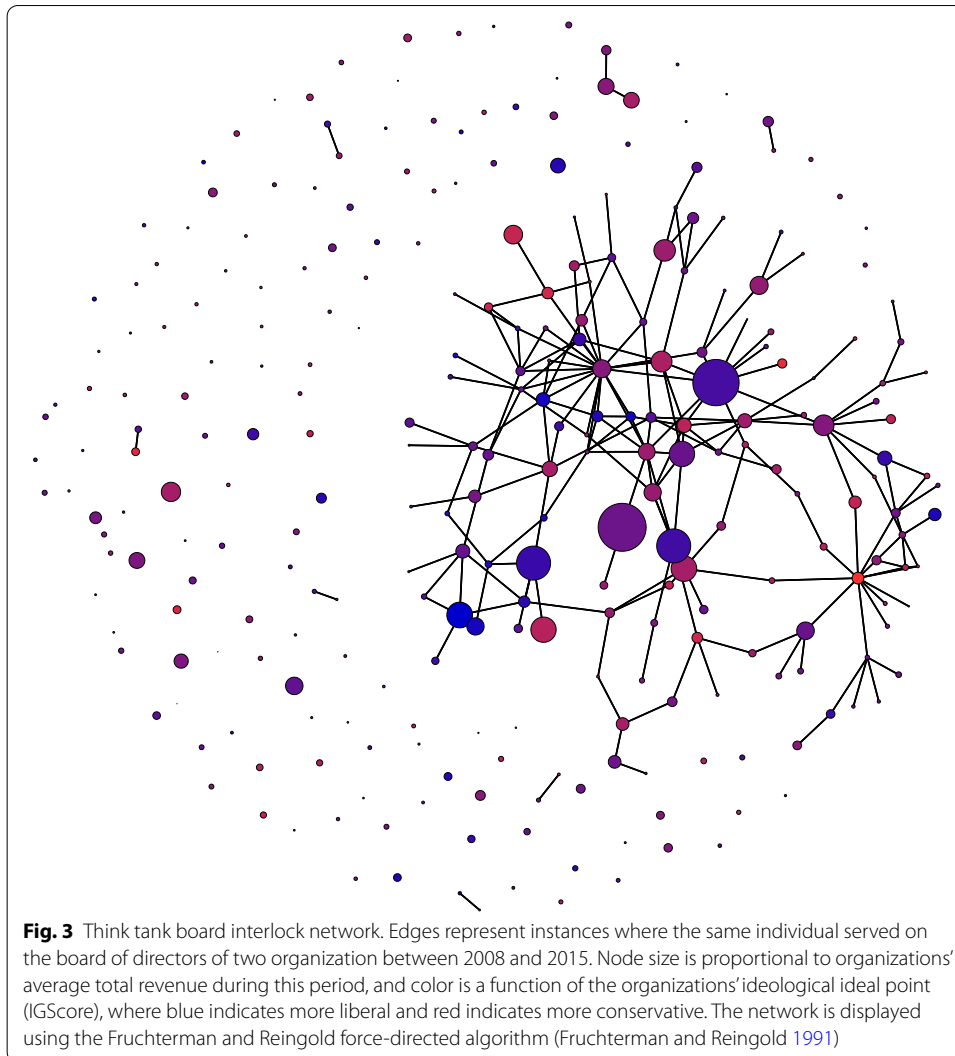
coded as having an interlock tie between them if they had directors with identical post-standardization names at any point during that eight year window. Using this method, 234 of these directors served on the boards of more than one organization in the dataset. For example, Ambassador C. Boyden Gray, founding partner of the DC-based law firm, Boyden Gray & Associates LLP, served on the board of 5 organizations in the dataset during this period, the most of any director,—American Action Forum, Atlantic Council of the US, The European Institute, Center for Global Development, and Freedomworks Foundation.

Because having a common board member within this time window is coded as an interlock tie, organizations can be coded as being interlocked without actually having a board member simultaneously serve on both boards. For example, if John Smith serves on the board of organization A from 2008 to 2011 and then organization B from 2013 to 2015, my procedure would code organization's A and B as being interlocked although they were not, in the strictest sense, actually interlocked. In this analysis I attempt to analyze the structure of the policy-planning network during the Obama Administration, looking at the whole time period. Because I use interlock ties as an indicator of coordination and communication, I count asynchronous ties like the hypothetical one between organisations A and B. As a member of the board of directors for organization B, John Smith would retain social-network connections to the board members he formerly served with on the board of organization A. Because these existing social ties would still serve as effective means of coordination and communication, I consider these asynchronous interlocks as valid for my analytic purposes.

Properties of the think tank interlock network

The interlock network of these D.C.-based research institutes and public policy analysis organizations is shown in Fig. 3, below. The degree distribution of this network is shown in Fig. 4.

Following (Gerber et al. 2013) who conduct a similar network analysis on the regional planning network in California, I report a series of network statistics on the observed interlock network. These are shown in Table 2. The 277 node undirected interlock network has a density of 0.005, which means that only $\frac{1}{200}$ of all possible dyads are connected by an interlock. The average organization in the network has a degree of 1.4 other organizations, while the most connected organization, the Atlantic Council of the US, is connected to 20 other organizations. There are 127 organizations in the largest connected component of the network, which comprises $\sim 46\%$ of the network. Path length, or the degree of separation between nodes, refers to the minimum number of "hops" that are required to travel from one node to the other. Mean path length is equal to $\frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j)$ where n is the number of vertices in the network, and $d(v_i, v_j)$ is the length of the shortest path between vertex i (v_i) and vertex j (v_j). Networks with longer mean path lengths are more disparately connected, while those with shorter mean path lengths are more closely connected. The interlock network has a mean path length of 4.825, substantially lower than the expected average path length of 9.29 calculated from 100 randomly generated Erdős–Rényi random graphs (Erdős and Rényi 1960; Newman et al. 2001), in which the probability of ties between any two of the 277 nodes was set equal to the observed density in the interlock network. The global clustering

Fig. 3 Think tank board interlock network. Edges represent instances where the same individual served on the board of directors of two organization between 2008 and 2015. Node size is proportional to organizations' average total revenue during this period, and color is a function of the organizations' ideological ideal point (IGScore), where blue indicates more liberal and red indicates more conservative. The network is displayed using the Fruchterman and Reingold force-directed algorithm (Fruchterman and Reingold 1991)

coefficient is the share of all triplets (sets of three nodes) which are closed (i.e. all connected to each other). Networks with higher global clustering coefficients will tend to have more tightly-knit clusters of nodes. The interlock network has a global clustering coefficient of 0.158 which far exceeds the expected global clustering coefficient estimated from Erdős–Rényi random graphs.

This pattern of comparatively small mean path length and a substantially higher clustering coefficient than we would expect by random chance suggests that the interlock network is a "small-world network" (Watts 1999), a type of network common in social and other real-world phenomenon that tend to have more highly connected subgraphs with high-degree hubs and relatively short paths between nodes.

Scholars have long noted the importance of an actor's position in a network for a variety of salient outcomes such as access to information (Granovetter 1985; Carpenter et al. 2004) ability to engage in brokerage (Burt 1992; Heaney 2006), and status (Podolny 2010; Heaney and Lorenz 2013). Following (Heaney and Lorenz 2013), I focus on betweenness centrality to identify potentially influential organizations in the interlock network.
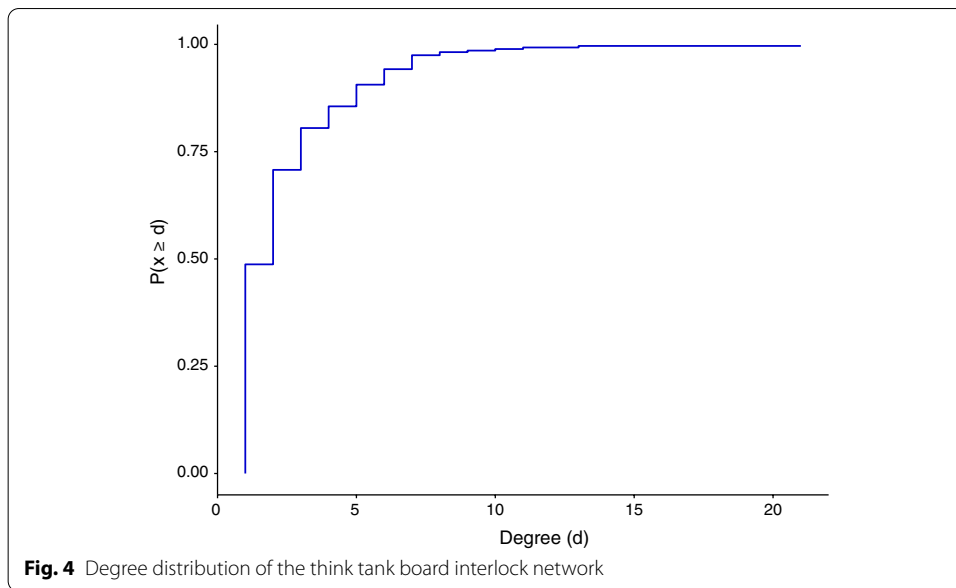
Fig. 4 Degree distribution of the think tank board interlock network

Table 2 Think tank interlock network statistics

Statistic	Value
Size	277
Density	0.005
Average degree	1.408
Maximum degree	20
Size of largest component	127
% in largest component	0.458
Global clustering	0.158
Mean path length	4.825
Expected global clustering	0.005
Expected path length	9.29

Betweenness centrality measures how frequently a given organization lies on the shortest path between other organizations in the network (Freeman 1978). For descriptive purposes I present the top twenty most central nodes according to betweenness centrality in Table 3.

All but one of the twenty most central organizations, the Albert Shanker Institute, have average yearly revenues above the median revenue in the network of $1,189,176. This suggests that revenue is strongly related to connectedness in the network, a proposition which I test more rigorously below.

Modeling think tank interlock connections

In this section I model the individual organizational, dyadic, and endogenous network factors which shape the patterns we observe in the interlock network I have described above.

Table 3 **Most central policy planning organizations**

Organization	Betweenness centrality	Revenue
Atlantic Council of the U.S.	2932.31	$14,830,739.38
National Quality Forum	2129.40	$19,507,711.62
Economic Policy Institute	1811.19	$6,308,457.50
Brookings Institution	1799.28	$98,120,399.86
Center for Global Development	1150.49	$12,671,746.12
World Resources Institute	954.05	$52,983,077.14
American Council for Voluntary International Action	847.39	$9,080,699.25
Families USA Foundation	847.33	$6,925,062.38
Washington Institute for Near East Policy	785.95	$10,627,693.75
Committee for Economic Development	768.53	$4,508,123.43
Center on Budget and Policy Priorities	754.52	$29,939,195.12
Center for us Global Leadership	694.39	$4,486,452.25
AAA Foundation for Traffic Safety	686.79	$4,241,489.75
Environmental Law Institute	675.64	$5,278,672.88
New America Foundation	618.39	$19,859,039.00
Albert Shanker Institute	614.00	$907,843.57
Competitive Enterprise Institute	570.56	$5,887,372.14
American Enterprise Institute	568.04	$53,042,058.86
Institute on Taxation and Economic Policy	510.22	$1,452,101.25
Results for Development Institute	491.23	$13,642,474.25

Methods

I use an exponential-family random graph model (ERGM) to interrogate the interlock network described above. Interdependence between ties presents a critical challenge to inference when working with network data, as the higher-order relational structure of the data violates the usual assumptions of simple linear regression models. ERGMs allow an analyst to explicitly model these higher-order network structures in addition to node-level and dyad-level covariates of interest (Robins et al. 2007, 2007).

Exponential random graph modeling treats the full observed network as the dependent variable, and node variables, dyadic variables, and endogenous network structures act as independent variables which influence the probability of any given tie to form within that network. The observed network is understood as a random draw from a probability distribution of possible networks given those independent variables, maximized via Monte Carlo maximum likelihood (MCMLE).

The ERGMs in this paper are fit using MCMLE (Handcock et al. 2008), using a single chain of 33,562,620 iterations and a thinning interval of 8196 for an effective sample size of 4090.

The full model of board interlock estimated here includes the following "exogenous" terms:

> *Ideological Distance* (`absdiff`)—dyadic covariate for each (i, j), defined as the absolute value of the difference between the IGScores of nodes i and j.
> *Match NTEE* (`nodematch`)— a statistic for "uniform homophily." Accounts for the propensity for groups to connect with others of the same NTEE-CC type (assumes this probability is *uniform* across types).

Match Dues Collecting (`nodematch`)—a statistic for "uniform homophily." Accounts for the propensity for groups to connect with others that also (do not) have dues paying members.

Match Lobbying (`nodematch`)—a statistic for "uniform homophily." Accounts for the propensity for groups to connect with others that also (do not) spend money on lobbying.

Match Contracting (`anodematch`)—a statistic for "uniform homophily." Accounts for the propensity for groups to connect with others that also (do not) hire contractors.

Log(Revenue) (`nodecov`)—a node covariate which captures the propensity for organizations with higher revenues to have more connections. Logged to aid convergence, as the distribution of revenue is extremely left-skewed.

Membership Dues (`nodecov`)—a node covariate which captures the propensity for organizations to connect to others with larger revenues from membership dues specifically.

Lobbying Fees (`nodecov`)—a node covariate which captures the propensity for organizations to connect to others that spend more money on lobbying.

When I take endogenous network structure into effect using the ERGM, I include the following "structural" terms:

Edges (`edges`)—a count of the number of edges in the graph. Can be thought of as analogous to an intercept in a linear model.

Concurrent (`concurrent`)—a network statistic equal to the number of nodes with degree two or higher.

Isolates (`isolates`)—a network statistic equal to the number of nodes with degree zero.

GWESP (`gwesp.fixed`)—a statistic equal to the *Geometrically weighted edgewise shared partner distribution* with a decay parameter of $\theta = 0.8$. See (Hunter 2007) for details. This statistic captures the tendency of the network towards transitivity, but is less sensitive to degeneracy than a simple triangles statistic.

The *Edges* term is included to capture the base-rate of connections within the network, and the *Concurrent*, *Isolates*, and *GWESP* terms are included to account for endogenous network dependency in the interlock network. These terms capture the tendency of the interlock network to form clusters while having a relatively large percentage of isolates.

Following best practices, I report several intermediary models prior to the full ERGM specification. Cranmer and Desmarais (2011) and Schrodt (2014) I present models which include only the primary variables of interest (Ideological distance and NTEE typo homophily, as well as those without endogenous network controls.

The models are run using identical specifications across 100 imputed datasets. Each of these datasets is identical in all regards other than that they contain different values of the imputed IGScores used to calculate the ideological distance between nodes. Results from these 100 separate estimates are combined using Rubin's rules (Rubin 1987; Honaker et al. 2011).

Table 4 Results from ERGMs run on 100 datasets with imputed IGScores

	Model 1	Model 2	Model 3	Model 4
Ideological distance	0.014	0.037	− 0.100	− 0.066
	(0.141)	(0.142)	(0.121)	(0.053)
Match NTEE		1.447	1.355	1.066
		(0.158)	(0.160)	(0.022)
Match dues collecting			− 0.628	− 0.410
			(0.166)	(0.017)
Match is lobbying			0.127	0.103
			(0.189)	(0.017)
Match is contracting			0.193	0.108
			(0.167)	(0.014)
Log(Revenue)			0.522	0.253
			(0.037)	(0.013)
Membership dues			0.000	0.000
			(0.000)	(0.000)
Lobbying fees			0.000	0.000
			(0.000)	(0.000)
Edges	− 5.290	− 5.571	− 20.496	− 12.073
	(0.158)	(0.168)	(1.122)	(0.469)
Concurrent				− 0.527
				0.135
Isolates				0.960
				0.074
GWESP($\theta = 0.8$)				0.761
				0.072
Model type	Dyadic-independent	Dyadic-independent	Dyadic-independent	Network-dependent
Replicates	100	100	100	100
Median AIC	2450.750	2386.328	2128.629	2008.582
Median BIC	2467.852	2411.982	2205.590	2111.198

Results

The results from the ERGM analysis is presented in Table 4 ERGM coefficients can be interpreted similarly to those of a logistic regression. In fact because models 1–3 contain no model terms for endogenous network effects, the ERGM estimation procedure defaults to logistic regression on the network dyads. Coefficients are the log-odds of a particular tie given the rest of the interlock network (Leifeld and Schneider 2012). Model 4 is the full ERGM model including endogenous network statistics, estimated using MCMC MLE on the 100 imputed datasets. If the p value from the multivariate geweke test was greater than or equal to 0.95, indicating poor convergence of the MCMC model, the model was re-estimated for that dataset. ERGM convergence and goodness-of-fit diagnostics are presented in the Additional file 1: Supplementary Information.

Compared to the simpler specifications, Model 4, which includes all substantive covariates as well as endogenous network effects, offers the best fit as indicated by the lowest AIC and BIC. It is worth noting, however, that across all specifications there is strong evidence of homophily by issue area of the think tank as indicated by the Match NTEE parameter estimates.

Fig. 5 Marginal effect of ideological distance in each imputed dataset
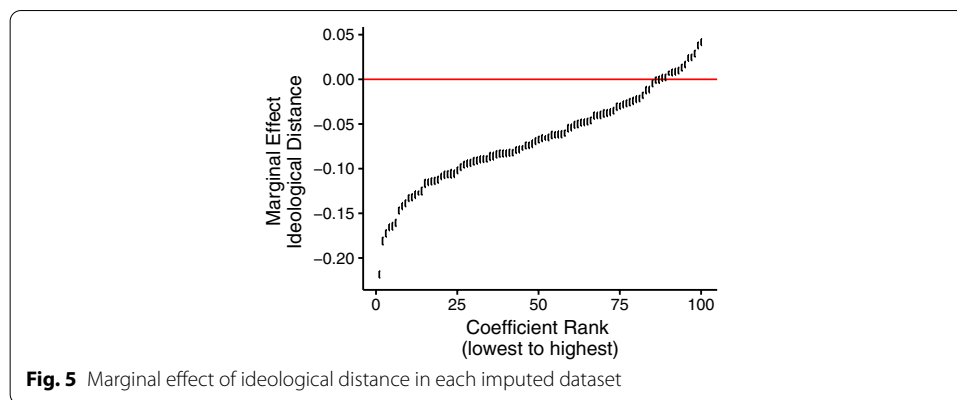
However, these results in Table 4 offer mixed evidence as to whether ideological distance is associated with a decreased probability of connection between organizations. In the complete model, think tanks are less likely to be connected to those that are ideologically more distant from them. However, the magnitude of this effect is quite small, and after propagating uncertainty from the imputation using 100 imputed datasets the effect is not significant at conventional levels. In order to assess whether this uncertainty is the result of missingness and imputation, I plot the separate estimates for each imputed dataset in Fig. 5, below. This plot shows marginal effect of an increase of 1 unit in distance between organizations i and j on the probability of a tie between those two organizations across each of the ERGMs estimated. In 85 out of 100 replicates (85%), ideological distance was associated with a significantly lower likelihood of a tie between two organizations. This suggests that these results are substantially attenuated by measurement error induced by missingness in the ideology measure, and variance introduced by imputation.

The most persistent and substantively significant finding across models 2–4, is that organizations are significantly more likely to be connected to each other if they are both of the same NTEE-CC classification. With log-odds of ~ 1.07, a given think tank is $\sim 190\%$ more likely to be connected to another think tank that works in the same issue area (according to NTEE classification), than it is to be connected to a think tank that works in a different issue area. This is strong evidence of issue-based homophily in the interlock network. This is consistent with several existing explanations for interlock behavior. Because the information transfer and coordination enabled by interlock can mitigate risk (Mizruchi 1996), scholars have argued that organizations tend towards homophilic interlock when they are dependent on other organizations in their sector. For example, organizations active within the same policy area may need to rely on the same sets of funders, and knowledge of each others' activities may allow them to avoid challenging each others' funding streams.

While not as strong as the propensity for organizations to connect with those working in the same area, organizations are also $\sim 10\%$ more likely to be connected to those that lobby if it also hires lobbyists. Similarly, organizations are $\sim 11\%$ more likely to interlock if they both hire contractors. One possible explanation for this is that contractors and lobbyist may serve as potential vectors of informal relationships between organizations, which ultimately facilitates later institutionalization via interlock.

Table 5 Subgraph density by organization NTEE-CC code

NTEE-CC code	Organizations	Edges	Density
International, foreign affairs & nat. sec.	47	26	0.0241
Public & societal benefit	42	11	0.0128
Education	30	5	0.0115
Environment	28	4	0.0106
Health care	22	6	0.0260
Community improvement & capacity building	18	1	0.0065
Social science	14	3	0.0330
Civil rights, social action & advocacy	13	0	0.0000
Food, agriculture & nutrition	12	1	0.0152
Science & technology	11	0	0.0000
Crime & legal-related	10	0	0.0000
Arts, culture & humanities	6	0	0.0000
Employment	4	0	0.0000
Human services	4	1	0.1667
Medical research	3	0	0.0000
Mental health & crisis intervention	3	0	0.0000
Animal-related	2	0	0.0000
Religion-related	2	0	0.0000
Voluntary health assoc. & medical disciplines	2	0	0.0000
Mutual & membership benefit	1	0	N/A
Public safety, disaster prep. & relief	1	0	N/A
Philanthropy, voluntarism & grantmaking	1	0	N/A
Housing & shelter	1	0	N/A
Full network	277	195	0.0051

In contrast, membership organizations that collect dues are significantly less likely to be connected to others that collect membership dues. Organizations connect to those that are dissimilar along this dimension, suggesting that they find benefit in the differential expertise and resources that the other can provide. However, the amount of lobbying an organization does, or amount of membership dues it collects are not associated with interlock. Non-member organizations and member organizations appear to seek each other out for interlock.

As expected, think tanks are substantially more likely to have interlocking boards as the revenue of the organizations increase; particularly successful fundraisers appear to be more attractive targets for interlock. Because the dependent variable in an ERGM is represented as log odds, the estimated effect of revenue can be interpreted as a log-log model, in which the proportional change in the probability of tie associated with a p percent increase in $Revenue = exp(a\beta_R)$, where $a = log([100 + p]/100)$ and β_R is the coefficient estimated for $Log(Revenue)$ (Benoit 2011). Using this, an increase in revenue of 100% is associated with a 19.1% increase in the likelihood of a tie. An organization with the revenue at 75th percentile ($3,433,207) has a 65.0% higher probability of a given tie than an organization with revenues at the 25th percentile ($473,935).

In order to better understand the within-issue area homophily that the ERGM results indicate, I present a breakdown of the subgraph density of the interlock network by NTEE-CC code of the organization. These results are shown in Table 5.

Density within the subgraph composed of the 47 organizations in the International, Foreign Affairs & National Security sector is almost five-times higher (0.024) than in the full interlock network (0.005). In fact, the five most common categories all have higher subgraph density than the density in the full graph. This suggests that organizations form interlock communities with each other based on their activity in the same policy sector. This effect is particularly strong among the most common types of organizations in the network.

Discussion

The results presented here represent the most extensive look at the network structure of interlock among active policy-planning organizations in Washington D.C. The network appears to exhibit small-world properties, suggesting it is relatively efficient for transmitting information within the connected component. The tendency for triadic closure represented by the positive effect of the GWESP parameter in the ERGM is consistent with organizations forming dense ties within their local communities, a form of network structure often thought to facilitate collaboration.

In contrast to prior work on small subsets of the policy network, I find mixed and inconclusive evidence of ideological alignment as a substantial driver of interlock. This may in part be related to ideology indeed playing a substantially less significant role among niche organizations, for whom issue area is a more salient part of their organizational identity than ideology. However, substantial missingness in my measure of ideology—and a conservative imputation process likely—may have attenuated these results as well.

Organizations tend to interlock with those with the most resources, and those active in their own issue areas. In the case of membership organizations, the network structure indicates that organizations tend to connect with organizations unlike themselves; membership organizations are more likely to connect with non-membership organizations and vice versa. This is particularly interesting because, by virtue of their organizational structure, membership organizations may tend to have better access to diffuse information from their membership,or a greater ability to mobilize an outside lobbying campaign (Kollman 1998).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1007/s41109-020-00318-7.

Additional file 1: Supplemental Appendices.

Abbreviations
ERGM: Exponential random graph model; NTEE-CC: National Taxonomy of Exempt Entities Core Codes; IRS: Internal Revenue Service; D.C.: District of Columbia; MCMLE: Monte Carlo Maximum Likelihood Estimation.

References
Bawn K, Cohen M, Karol D, Masket S, Noel H, Zaller J (2012) A theory of political parties: groups, policy demands and nominations in American politics. Perspect Polit 10(3):571–597
Benoit K (2011) Linear regression models with logarithmic transformations. Lond Sch Econ 22(1):23–36
Berardo R, Scholz JT (2010) Self-organizing policy networks: risk, partner selection, and cooperation in estuaries. Am J Polit Sci 54(3):632–649
Burris V (2008) The interlock structure of the policy-planning network and the right turn in US state policy. Res Polit Sociol 17(1):3–43
Burt RS (1992) Structural holes: the social structure of competition. Harvard University Press, Cambridge
Carpenter DP, Esterling KM, Lazer DM (2004) Friends, brokers, and transitivity: Who informs whom in Washington politics? J Polit 66(1):224–246
Carroll GR (1984) Organizational ecology. Ann Rev Sociol 10(1):71–93
Cranmer SJ, Desmarais BA (2011) Inferential network analysis with exponential random graph models. Polit Anal 19(1):66–86
Crosson JM, Furnas AC, Lapira T, Burgat C (2019) Partisan competition and the decline in legislative capacity among congressional offices. Legisl Stud Q
Crosson JM, Furnas AC, Lorenz GM (2020) Polarized pluralism: organizational preferences and biases in the American pressure system. Am Polit Sci Rev 114(4):1117–1137. https://doi.org/10.1017/S0003055420000350
Curry JM (2015) Legislating in the dark: information and power in the house of representatives. University of Chicago Press, Chicago
Davis GF (1991) Agents without principles? The spread of the poison pill through the intercorporate network. Admin Sci Q 36:583–613
Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci 41(6):391–407
Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5(1):17–60
Fagan E (2019) Issue ownership and the priorities of party elites in the United States, 2004–2016. Party Politics 1354068819839212
Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1(3):215–239
Fruchterman TM, Reingold EM (1991) Graph drawing by force-directed placement. Softw Pract Exp 21(11):1129–1164
Furnas AC (2019) Biasing their bosses: staff ideology, motivated reasoning, and the distortion of information in congress. Unpublished Working Paper. https://bit.ly/2ksg8i5
Furnas AC, LaPira T, Hertel-Fernandez A, Drutman L, Kosar K Moneyed interests, information, and action in congress: a survey experiment. Unpublished Working Paper. https://bit.ly/2lwUYzE
Furnas G, Deerwester S, Dumais S, Landauer T, Harshman R, Streeter L, Lochbaum K (1988) Information retrieval using a singular value decomposition model of latent semantic structure. In: Proceedings of the 11th annual international ACM SIGIR conference on research and development in information retrieval, pp 465–480
Gerber ER, Henry AD, Lubell M (2013) Political homophily and collaboration in regional planning networks. Am J Polit Sci 57(3):598–610
Granovetter M (1985) Economic action and social structure: the problem of embeddedness. Am J Sociol 91(3):481–510
Gray V, Lowery D (2000) The population ecology of interest representation: lobbying communities in the American States. University of Michigan Press, Ann Arbor
Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008) statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24(1):1548
Heaney MT (2006) Brokering health policy: coalitions, parties, and interest group influence. J Health Polit Policy Law 31(5):887–944
Heaney MT, Lorenz GM (2013) Coalition portfolios and interest group influence over the policy process. Interest Groups Advocacy 2(3):251–277
Honaker J, King G, Blackwell M et al (2011) Amelia II: a program for missing data. J Stat Softw 45(7):1–47
Hunter DR (2007) Curved exponential family models for social networks. Soc Netw 29(2):216–230
Kollman K (1998) Outside lobbying: public opinion and interest group strategies. Princeton University Press, Princeton
Kusumasari TF et al (2016) Data profiling for data quality improvement with openrefine. In: 2016 international conference on information technology systems and innovation (ICITSI), pp 1–6. IEEE
Leifeld P, Schneider V (2012) Information exchange in policy networks. Am J Polit Sci 56(3):731–744
Lerner JY (2018) Getting the message across: evaluating think tank influence in congress. Public Choice 175(3–4):347–366
Mariolis P (1975) Interlocking directorates and control of corporations: the theory of bank control. Soc Sci Q 56:425–439
Medvetz T (2012) Think tanks in America. University of Chicago Press, Chicago
Mizruchi MS (1982) The American Corporate Network, 1904–1974, vol 138. Sage Publications Inc, New York

Mizruchi MS (1996) What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates. Ann Rev Sociol 22(1):271–298

Newman ME, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64(2):026118

Noel H (2014) Political ideologies and political parties in America. Cambridge University Press, Cambridge

Podolny JM (2010) Status signals: a sociological study of market competition. Princeton University Press, Princeton

Rehurek R, Sojka P (2011) Gensim—statistical semantics in python. Retrieved from https://is.muni.cz/publication/884893/en

Robins G, Pattison P, Kalish Y, Lusher D (2007a) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29(2):173–191

Robins G, Snijders T, Wang P, Handcock M, Pattison P (2007b) Recent developments in exponential random graph (p*) models for social networks. Soc Netw 29(2):192–215

Rubin DB (1987) Multiple imputation for survey nonresponse. Wiley, New York

Salas-Porras A (2018) American think tank networks and expert debates around the global financial crisis: Keynesian insurgents against austerity defenders. Policy Soc 37(2):243–259

Schrodt PA (2014) Seven deadly sins of contemporary quantitative political analysis. J Peace Res 51(2):287–300

Useem M (1979) The social organization of the American business elite and participation of corporation directors in the governance of American institutions. Am Sociol Rev 44:553–572

Verborgh R, De Wilde M (2013) Using OpenRefine. Packt Publishing Ltd, Birmingham

Watts DJ (1999) Networks, dynamics, and the small-world phenomenon. Am J Sociol 105(2):493–527

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.