

# Measuring the dynamical evolution of the United States lobbying network

Karol A. Bacik,<sup>1</sup> Jan Ondras,<sup>1</sup> Aaron Rudkin,<sup>2</sup> Jörn Dunkel,<sup>1</sup> and In Song Kim<sup>2</sup>

<sup>1</sup>*Department of Mathematics, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02142, USA*

<sup>2</sup>*Department of Political Science, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02142, USA*

(Dated: March 18, 2025)

**Lobbying systems are complex political networks that influence governmental decisions, with often profound socio-economic consequences on national and global scales [1, 2].** For most political systems, a comprehensive understanding of lobbying strategies and dynamics is likely to remain elusive as time-resolved system-spanning data and analysis are lacking [3, 4]. A notable exception is the United States (U.S.), where the Lobbying Disclosure Act (LDA) of 1995 mandates that all federal lobbying activities be disclosed in detailed quarterly reports [5]. Here, we introduce our recently completed relational *LobbyView* database [6] that accounts for every reported lobbying instance since the implementation of the LDA. We demonstrate how *LobbyView* can be used as a resource to quantify the salient aspects of the U.S. lobbying, such as dynamical evolution, economic impacts, or political polarization. By analyzing the dynamic evolution of the lobbying network, we identify fundamental self-organization principles, such as the self-accelerating accumulation of influence within a small group of powerful lobbying firms. We further show how *LobbyView* data can be used to accurately measure the synchronization of lobbying activities with election cycles. Finally, as a guide to future research, we illustrate how this data resource can enable quantitative time-resolved analysis to investigate a wide range of critical issues, including concentration of resources, demographic representation, and polarization dynamics in the U.S. lobbying system. We envisage our database [6] to be not only a potent resource for political and social scientists but also a starting point for quantitative interdisciplinary research, by leveraging insights and methods from statistical physics, systems biology, and machine learning.

Political networks [7–10] are complex dynamical systems that evolve and adapt in response to national and global events [11, 12]. A prime example is the U.S. lobbying system [1, 2], a self-organized multilayered network (Fig. 1A) that ingests billions of dollars annually [1] and has orchestrated a significant part of both the legislative and electoral processes in the U.S. [13–15]. The dynamically evolving interactions and monetary flows within the U.S. lobbying network have profound consequences

for national [16] and international [17] issues, from global climate [18, 19] and health [20] crises to economic [21] and armed [22] conflicts. The complex interactions between special interest groups, lobbyists, government agencies, and politicians pose central challenges in the social and political sciences: Which political entities engage in lobbying and which strategies do they pursue [23]? What determines the scale of the monetary flows [24] and how do these affect political decisions [15]? How does the lobbying network re-organize in response to disruptive events such as the 2007–2008 financial crisis or the COVID-19 pandemic?

Despite the critical need for a systemic understanding of these and other critical issues, and notwithstanding important methodological advances [3, 4, 25], a quantitative end-to-end analysis of the network of actors in lobbying and their interactions remains elusive. This gap primarily persists due to the absence of comprehensive, high-fidelity data that captures both the monetary and informational flows, as well as the intricate dynamics among the various components of the U.S. lobbying system. The availability of such data holds the key to substantial future progress as it opens the possibility of utilizing recently developed methods from network theory [26–30], applied mathematics [31, 32] and statistical physics [33, 34] to achieve a predictive understanding of political decision processes. In particular, such data will offer a chance to compare the evolution and adaptation within an influential human interaction network with the emergent dynamical behaviors and scaling laws found in other complex social [35–37], biological [38–40], physical [41, 42] and information-processing [43, 44] network systems.

Here, we introduce a new research resource that can be used for quantitative empirical study of all the federal lobbying activities in the U.S. from 1999 to 2023: our comprehensive *LobbyView* database [6], which we have developed over the past decade and made available to the general public. As of 2024, *LobbyView* includes 1.6 million public records that encompass the complete history of federal lobbying since 1999, accounting for more than 87 billion USD in lobbying expenditures. These records provide granular details on each reported issue-specific interaction among clients, lobbying firms, lobbyists, government entities, and politicians, with half-yearly (and more recently quarter-yearly) time resolution.

To showcase the potential of *LobbyView* as a resource for interdisciplinary quantitative research, we use our

data to characterize the complex dynamics of the lobbying network at different time scales. We find that the long-term evolution of professional connections is governed by universal principles such as preferential attachment, the mid-term adaptations are synchronized with the election cycle, and the short-term anomalies can be understood as a response to a perturbation caused by a critical event, such as the COVID-19 pandemic. Moreover, we show how the LDA data can be used to measure concentration of resources, demographic representation and political polarization in lobbying.

## THE LOBBYVIEW DATABASE

*Lobbying reports.* The Lobbying Disclosure Act (LDA) of 1995, amended by the Honest Leadership and Open Government Act of 2007, mandates that federally registered lobbyists and lobbying firms (registrants) file quarterly reports [5]. These reports, known as LD-2 disclosure forms, detail their lobbying activities on behalf of their clients. The disclosures include the total lobbying expenditures and payments by the clients during the relevant period, along with a breakdown across 69 policy issues of concern, including taxation and international trade. For each issue area, registrants must provide detailed information about the specific subjects of their lobbying, including any lobbied congressional bills or resolutions, presented as free text entries. Additionally, the reports are required to identify the federal agencies, departments, and chambers of Congress that were targeted. The names of individual lobbyists involved and any past working relationships between these lobbyists and government officials or legislators are also disclosed. We have collected and curated this data into the *LobbyView* database [6, 45], which is continuously updated with new records as they become available.

*Technical aspects.* An easily underappreciated technical challenge is that the same lobbying actor (e.g., a lobbyist) often appears in the textual LDA reports under different names (Joseph Smith, Joe Smith, J. Smith); likewise, firms represented in different fashions, and when firms change their names or are merged or acquired, no identifier in the data is sufficient to uniquely identify them. We introduce novel algorithms (see [45]) to resolve these text disambiguation challenges and produce unique identifiers for every agent and firm in our database. This allows us to accurately infer the connectivity patterns of the lobbying network for the first time. It allows use to interface unstructured LDA reports to major external databases, such as WRDS Compustat, BoardEx, Moody’s Orbis, NOMINATE congressional ideological scores, interest group scores of legislators, and personal political donation records. As a service to researchers we provide the necessary identifiers to complete these cross-walks.

## RECONSTRUCTING THE LOBBYING NETWORK

We will now show how the *LobbyView* data can be used to discern systemic patterns within the complex lobbying network. Our focus is on elucidating the network dynamics and statistical properties of political interactions among various entities, while deliberately omitting personal and corporate specifics, as well as detailed descriptions of the lobbying activities themselves. As an initial step towards creating a data-driven, reduced-order representation of U.S. lobbying, we limit our analysis to five primary categories of lobbying actors: *clients* (firms or interest groups initiating lobbying efforts), *registrants* (lobbying firms responsible for filing reports), *lobbyists* (individuals directly involved in lobbying activities), *government entities* (House(s) of Congress, and federal agencies and departments targeted by lobbyists), and *legislators* (members of the U.S. Congress).

Due to a clear direction from the lobbying initiative to target stakeholders, we can visualize the lobbying network as a multipartite layered graph (Fig. 1A) [46]. The first (most upstream) layer of the network corresponds to clients (green), contracting registrants in the second layer (brown), who employ lobbyists in the third layer (gray). Importantly, the registrants can be either *In-House* departments of the client firm, or external lobbying firms (often described with a metonym *K-Street*). We construct the connections between the first three layers by adding edges  $c \rightarrow r$  and  $r \rightarrow l$  if lobbyist  $l$  lobbied for client  $c$  through registrant  $r$ . It is worth noting that this representation obscures the intrinsically polyadic character of lobbying interactions [32, 47], but as we shall see it is a useful first-order approximation (see Sec. VI of [45]).

The two downstream layers represent the target entities of the lobbying efforts: government entities (turquoise) and legislators (Congress members) (red/blue). Lobbyists are connected to government entities and legislators through past employment (internships, clerkships, etc.), which the LDA mandates lobbyists report and which we make available through *LobbyView*. We observe a steady increase in documented past connections to government entities (in 2023, 20% of lobbyists had a known connection to the government) (Sec. III of [45]). Moreover, in the recent years for 80% of Senators and 60% of Representatives, we can find at least one active lobbyist related to them through past employment (Fig. S9 of [45]). Obviously, the lobbyists’ contacts are not limited to their past employers [48], but the professional network (and the so-called *revolving-door effect*) is believed to play an important role in the lobbying process [49, 50].

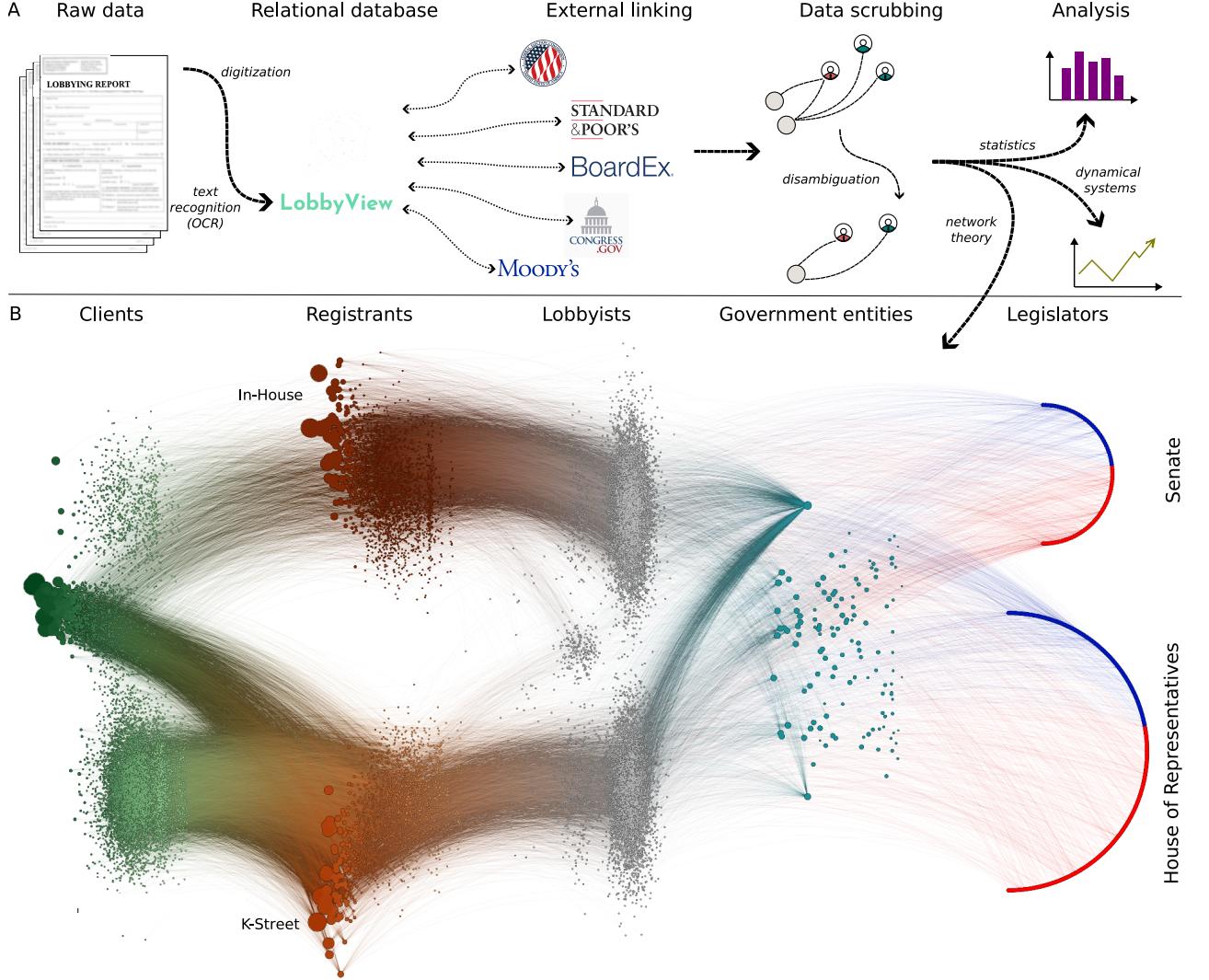
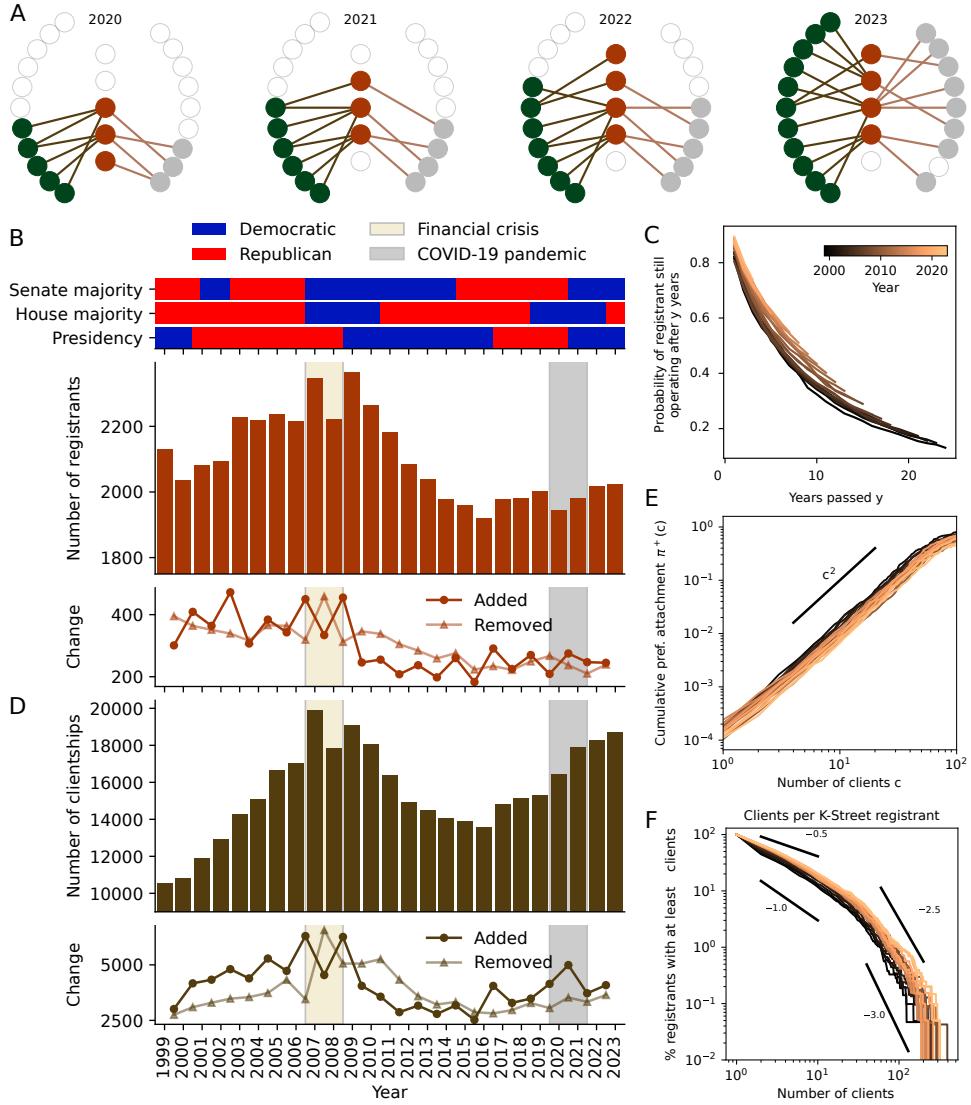


FIG. 1. Our *LobbyView* database comprises over 1.6 million public records, covering all federal U.S. lobbying since 1999 and enabling a time-resolved reconstruction of the multilayered architecture and scaling behaviors of the lobbying network. (A) The data processing from the Lobbying Disclosure Act (LDA) reports to lobbying network starts with the digitized text of publicly available disclosures. The atomistic data is then compiled into a relational database, which can be interfaced with other political datasets. To correctly infer the statistics of lobbying activities and infer the lobbying network structure, we clean the data and disambiguate the lobbying actors. (B) Network representation of all U.S. federal lobbying activities during the year 2017. The first layer comprises 10,694 clients initiating lobbying (green), and the second layer comprises 4,405 registrants (brown). The node size in these layers is proportional to the cumulative USD amount spent/received by each individual client/registrant. The registrants are connected to 11,543 lobbyists they employ (third layer, gray). We also reconstructed historical associations of lobbyists with 128 specific government agencies (turquoise) or current members of Congress (374 legislators; Democrats in blue and Republicans in red). The details of the visualization can be found in [45], Sec. VII.

#### LONG-TERM DYNAMICAL EVOLUTION DRIVEN BY ACCUMULATED ADVANTAGE

*Ever-changing actors.* The lobbying network is dynamic, undergoing continual evolution and adaptation, and our extensive database encompassing lobbying reports spanning from 1999 to 2023, offers a unique opportunity

for longitudinal analysis of such processes [38, 51–53] at the level of individual lobbying actors. Focusing on the K-Street registrants, we observe an annual influx of 200–500 new registrants into the network, alongside a comparable number exiting (Fig. 2B). As a result of this attachment-detachment process, the set of registrants regenerates and rejuvenates continuously, and if



**FIG. 2. The lobbying network is evolving and adapting in response to political landscape and crisis events.** (A) Attachment and detachment processes in a sample client-registrant-lobbyist (green-brown-gray) network component during 2020–2023 (for selection methodology see Sec. VIII.D of [45]). (B) Number of active K-Street registrants per year (bars) and their yearly increments (circles) and decrements (triangles). The top horizontal bars show the party (Democratic in blue and Republican in red) of Senate majority, House majority, and President by year. The financial crisis (2007–2008) and COVID-19 pandemic (2019–2021) are shown by gray overlays. (C) Number of active clientships (client-registrant connections) per year (bars) and their yearly increments (circles) and decrements (triangles). (D) K-Street registrant survival probability. (E) Cumulative pref. attachment function  $\pi^+(c)$  in terms of K-Street registrant in-degree. The quadratic functional form implies that the probability that a new link ‘selects’ a registrant with  $c$  clients is proportional to  $c$ . (F) Complementary cumulative distribution (CCDF) of the number of clients per K-Street registrant (K-Street registrant in-degree). Quantities in (D–F) are computed separately for each year in the range 1999–2023 and their precise mathematical definitions can be found in [45], Sec. VIII.

we compare the set of registrants in 2000 and 2020, we find that only 20% of the registrants from 2000 are still active (Fig. 2C). In other words, there exists fairly significant churn in the membership of core political actors involved in lobbying, while the fundamental structure of the lobbying network remains unchanged.

*Preferential attachment and detachment.* Even if a registrant remains in operation, the set of its clients may change (Fig. 2C). Every year some clients leave the lobbying network, new clients join, and the remaining clients establish relations with new registrants. The scale of this rewiring process is quite substantial, as it involves up to

a third of the clientship connections each year. Critically, the rewiring is not entirely random. Instead, it appears to be driven by the principle of preferential attachment [54]. We find strong evidence that the probability of attracting a new client is proportional to the number of existing clients (linear preferential attachment) (Fig. 2E) [38, 52, 53]. Preferential attachment is one type of accumulated advantage (the Matthew Principle, wherein large or successful incumbents accrue further size or success), which is a well-known mechanism for generating heavy-tailed degree distributions [54–57]. Hierarchical advantage is commonly observed in economic networks [54, 58, 59], and because K-Street registrants operate in a free market, it would not be entirely surprising if their organization was hierarchical, too.

Fig. 2F shows that the distribution of clients per K-Street registrant is indeed hierarchical, locally approximated by power laws, with a small group of elite registrants. Notably, 80% of K-Street income is generated by only 20% of registrants and 60% of the lobbyist-registrant contracts are signed by only 20% of registrants as well (Sec. V of [45]). Interestingly, this concentration of distribution may have profound consequences for the lobbying industry, as discussed in detail below in Fig. 4.

In summary, by analyzing the dynamical evolution of the client-registrant ties, we found strong evidence of preferential attachment, which can rationalize the hierarchical structure of K-Street. Nevertheless, we must note that preferential attachment is not the only principle governing the evolution of the lobbying network. For example, network edges are not merely added as the network evolves; they are removed as well. The detachment probability is also proportional to the number of existing clients (Fig. S24 of [45]), so presently the preferential attachment and preferential detachment act as largely countervailing forces. A more detailed mathematical study, beyond the scope of this article, could take into account these processes in order to explain the exact shape of the degree distributions and forecast the long-term evolution of the lobbying network [53, 60]. We need to emphasize, however, that idealized models will never be able to predict the future of the lobbying network with absolute certainty, due to the shifting political landscape.

## SHORT-TERM ADAPTATION DRIVEN BY THE POLITICAL LANDSCAPE

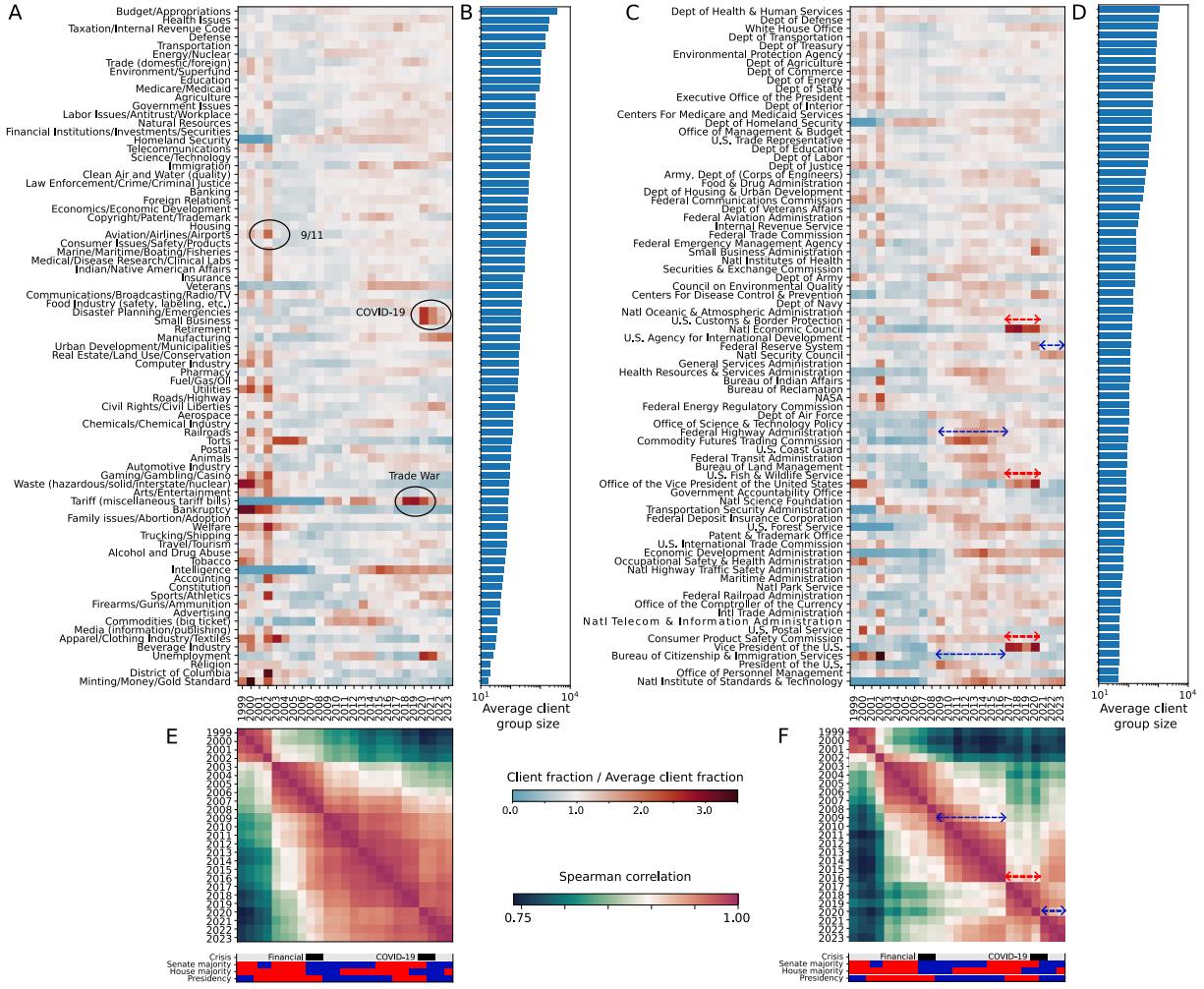
The lobbying network is not simply a self-sustaining, autonomous ecosystem. It develops in the context of and in reaction to current affairs. For example, Fig. 2B clearly shows the impact of the financial crisis in 2007–2008, which led to a downturn in the lobbying industry persisting throughout the subsequent decade. We also anticipate that even if the lobbying firms remain in operation, they adapt their lobbying instruments in response to events. What are the general characteristics of such adaptations?

*Lobbying subnetworks: robust vs. volatile interests.* To address this question, we can decompose the lobbying network into subnetworks associated with various general issue areas [5] and we present the size of the associated client group in Fig. 3A-B. Large interest groups are associated with ‘popular’ issues, such as taxation or health-care, consistently taking up a significant portion of the lobbying budget annually. These issues represent fundamental government activities that hold enduring importance and are consistently targeted by lobbyists. Conversely, the popularity of less prominent issues may fluctuate depending on current controversies or policy initiatives. For instance, we observe substantial spikes in lobbying activity concerning disaster planning and small business during the COVID-19 pandemic in 2020–2021.

The lobbying network can also be split into subnetworks comprising lobbying groups that target various executive government entities (Fig. 3C-D). At the top of the ranking of targeted government entities are federal executive departments, with the Department of Health and Human Services taking the lead. Such institutions are lobbied every year by a large number of clients, regardless of the political situation. Significant year-to-year fluctuations can be observed in the case of less frequently targeted government institutions. Particularly notable adaptations of the lobbying network, synchronized with the presidential election cycle, can be observed in the case of White House bodies, such as the National Economic Council.

To quantify the year-to-year differences in lobbying activities, we compute the *lobbying issue portfolio* and *lobbying target portfolio*. Mathematically, the annual lobbying issue portfolio and the annual lobbying target portfolio can be understood as feature vectors: columns of the matrix in Fig. 3A and Fig. 3C, respectively. To systematically compare annual lobbying portfolios, we compute the year-to-year Spearman correlation matrices. This analysis reveals that the lobbying issue portfolio is responsive to crisis events, e.g., it underwent a major rearrangement, following the 9/11 terrorist attacks and during the COVID-19 pandemic (Fig. 3E). We find that the lobbying target correlation matrix exhibits a distinct block-diagonal structure, manifestly synchronized with the election cycle (Fig. 3F). For example, it appears that lobbyists tend to approach the same government entities with consistent frequency during a presidency, but adapt their lobbying efforts swiftly after a change in leadership. This adaptation underscores the lobbying network’s sensitivity to the differing views of the two major political parties in the U.S regarding the scope and responsibilities of government departments and agencies.

More broadly, these results indicate that information flows can follow different paths through the lobbying network, depending on the nature of the issue and the political situation. To illustrate the level of detail accessible within our database, we present in Sec. XIV of the Supplementary Information [45] a detailed case study of the information flow rearrangements for Small Business

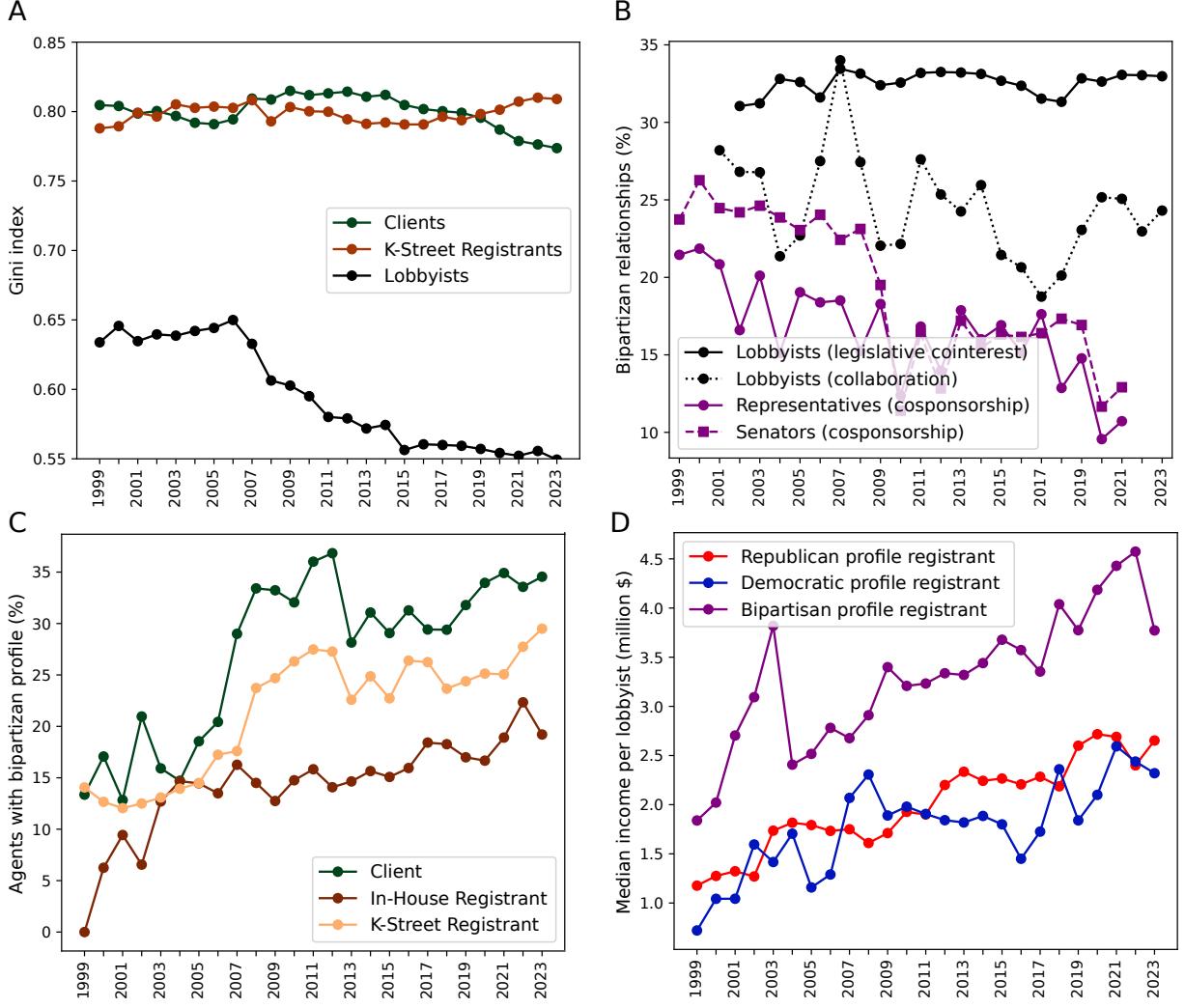


**FIG. 3. Our data enables fine-grained analysis of the lobbying network adaptations.** (A) Fraction of clients lobbying on a given issue, scaled by its average value in the period 1999–2023. Some of the spikes of interest can be easily associated with crisis events, such as the terrorist attacks of 9/11, or the COVID-19 pandemic. (B) Average number of clients lobbying on a given issue. (C) Fraction of clients approaching a given government entity in a given year, scaled by the average number of targeting clients. In this Figure, we present only the top approached government entities; a more extensive list can be found in [45], Sec. XI. (D) Average number of clients approaching different government entities. (E) Year-to-year correlation of the *issue portfolio vectors* (columns of the matrix in panel (A)). (F) Year-to-year correlation of the *government approach portfolios* presented as columns in panel (C). The correlation matrix has a manifest block-diagonal structure, synchronous with the presidential terms.

lobbying during the COVID-19 pandemic. Such a fine-grained analysis uncovers the specific pathways of ‘venue shopping’, demonstrating how lobbyists select their target government institutions and legislators in response to novel legislative initiatives.

## DETECTING AND QUANTIFYING STRUCTURAL CHANGES IN THE U.S. LOBBYING DYNAMICS

*LobbyView* currently spans more than two decades and will be continuously updated with forthcoming lobbying reports. It thus provides an empirical foundation to quantify historical trends and detect structural transitions in the U.S. lobbying network. A comprehensive in-depth analysis of the longitudinal dynamical trends, exploring the fine-grained *LobbyView* data, will require concerted interdisciplinary efforts beyond the scope of this



**FIG. 4. The network connectivity allows us to measure the degree of resource concentration and polarization in U.S. lobbying.** (A) Evolution of global Gini indices based on lobbying expenditure (clients) or income (K-Street registrants, lobbyists). High Gini index for client and registrants points to a substantial disparity in the distribution of lobbying influence. A drop in the lobbyist Gini index indicates a lasting economic transformation of the lobbying industry after the financial crisis of 2007–08. (B) The share of bipartisan professional relationships can be used to evaluate the degree of polarization. Our analysis suggests that, unlike the polarization of the lawmakers, the polarization of the lobbying industry is not increasing (see Sec. XIII of [45] for the methodology). (C) Our data indicates that the share of lobbying agents (clients and registrants) with bipartisan profiles increases. (D) Registrants employing lobbyists with known historical connections to both of the major political parties on average generate higher income per lobbyist.

article. To illustrate the future potential for such analysis, we present exploratory diagnostic measurements of three vital aspects of the U.S. lobbying: resource concentration and partisan polarization (Fig. 4).

*Concentration of resources.* Market dominance of a few large lobbying firms can increase the cost of lobbying and thus decrease access of smaller groups or individuals to political decision makers [61, 62]. *LobbyView* makes it possible to quantify the concentration dynamics of lobbying activities within and across the various components of the lobbying network. To demonstrate this, we calcu-

late the time-resolved Gini indices for lobbying budget distributions. The Gini index, ranging from 0 (egalitarian) to 1 (monopolistic), measures distribution dispersion and has been used to compare income and other distributions [63, 64]. Here, we adopt this methodology to quantify the distribution of lobbying resources within the client, registrant, and lobbyist layers (see Sec. V of [45] for details). We find that, over the last two decades, the cumulative Gini indices (based on lobbying activities across all issues and agencies) for both clients and K-street lobbying firms have remained approximately con-

stant at a remarkably high value  $\sim 0.8$ , indicating that lobbying activities have been persistently dominated by affluent clients and a few large external lobbying firms (Fig. 4A).

The Gini index for lobbyists income is relatively lower ( $\sim 0.6$ ) but still markedly higher than the Gini index for U.S. income distribution ( $\sim 0.4$  according to the World Bank estimate). In other words, the lobbying industry appears to have a higher degree of wealth concentration than the U.S. society as a whole. A more detailed analysis of revenue distribution among different lobbyists of a given registrant is presented in Sec. VI of [45]. Nevertheless, we also observe a fall in the Gini index for lobbyists immediately after 2007, suggesting that the financial crisis of 2008 contributed to a decentralization of the lobbying profession and increased opportunities for new entry (Fig. 4A). We next outline how the *LobbyView* data could be used in the future to investigate potential causes and consequences of such changes, for instance, by studying partisan polarization.

*Estimating partisan polarization.* Polarization, the tendency of ideological or demographic groups to sort into partisan groups, can impede cooperation and consensus while promoting more extreme voices and policy outcomes [65]. *LobbyView* provides detailed information about professional connections to legislators that can be used to infer the political leanings (democratic vs. republican) of clients, registrants and lobbyists. By analyzing the connectivity patterns in the lobbying network, one can estimate the degree of partisan polarization in lobbying (Sec. XIII of [45]). Perhaps unsurprisingly, our data analysis reveals strong evidence for an increasing polarization among the legislators, highlighted by the fact that the rate of bipartisan co-sponsorship of bills in the U.S. Senate has dropped by about 40% over the last 25 years (Fig. 4B). This fact raises the question whether a similar trend towards polarization has occurred within the lobbying industry. Unexpectedly, our data suggest that this has not been the case: The number of instances when the lobbyists of opposite political leaning work together or lobby on the same bill has remained relatively stable (Fig. 4B). Moreover, the proportion of clients and registrants that employ a bipartisan team of lobbyists is weakly increasing (Fig. 4C), suggesting that both clients and registrants recognize a need to build connections with both parties to maximize impact. The incentive on the registrant side is clear: bipartisan teams of lobbyists generate more income (Fig. 4D). Thus, our data support the hypothesis that lobbying is a strategic, rather than

ideological activity.

## DISCUSSION

In summary, we introduced high-dimensional dynamical data made available through *LobbyView* [6], which enables a quantitative end-to-end analysis of the U.S. lobbying system. To understand the interactions between the various political stakeholders, we represented the lobbying system as a layered network. We illustrated the potential of *LobbyView* as a research resource by analyzing the fundamental empirical characteristics of this complex dynamical network: agent statistics, distribution of connections, dynamic evolution principles, and event-induced adaptations. Our analysis underscores the importance of preferential attachment for the long-term evolution of connections, the dominant role of the election cycle for the selection of lobbying instruments, and the adaptability of information flow in response to critical events. Moreover, we showcased how *LobbyView* can be used to quantify salient political phenomena such as polarization. Owing to the complexity of the underlying dataset, many important facets of the lobbying dynamics, such as the role of higher-order and multilayer interactions [32, 46, 47], or the multiscale community structure of the lobbying network [66, 67], remain to be explored. More generally, we hope that this work and the continuously updated *LobbyView* [6] database will initiate a comprehensive evidence-based reflection on interest politics that can increase public awareness and help improve the democratic process. Collective interdisciplinary effort across and beyond the academic community will be essential for achieving this goal.

## ACKNOWLEDGMENTS

This research received support through Schmidt Sciences, LLC (to J.D.), the MathWorks Professorship Fund (to J.D.), the National Science Foundation (SES-1725235, SES-2017315) (to I.K.), and the Russell Sage Foundation (# 1908-17912) (to I.K.). I.K. and J.D. are grateful for seed grant support from the Social and Ethical Responsibilities of Computing (SERC) Initiative of the MIT Schwarzman College of Computing, and from the ICSR Seed Fund of the MIT Institute for Data, Systems, and Society.

- 
- [1] M. Bombardini and F. Trebbi, Empirical models of lobbying, *Annual Review of Economics* **12**, 391 (2020).
  - [2] J. M. D. Figueiredo and B. K. Richter, Advancing the empirical research on lobbying, *Annual Review of Political Science* **17**, 163 (2014).
  - [3] D. M. J. Lazer, A. Pentland, D. J. Watts, S. Aral, S. Athey, N. Contractor, D. Freelon, S. Gonzalez-Bailon, G. King, H. Margetts, A. Nelson, M. J. Salganik, M. Strohmaier, A. Vesplignani, and C. Wagner, Computational social science: Obstacles and opportunities, *Science* **369**, 1060 (2020).

- [4] H. Y. You, Ex post lobbying, *The Journal of Politics* **79**, 1162 (2017).
- [5] U.S. Senate Office of Public Records, *Lobbying Disclosure Act Guidance* (2021), <https://lobbyingdisclosure.house.gov/ldaguidance.pdf>.
- [6] LobbyView Database, <https://www.lobbyview.org/> (2024).
- [7] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, A network analysis of committees in the U.S. House of Representatives, *Proceedings of the National Academy of Sciences U.S.A.* **102**, 7057 (2005).
- [8] J. H. Fowler, Connecting the Congress: A study of cosponsorship networks, *Political Analysis* **14**, 456 (2006).
- [9] D. M. J. Lazer, Networks in political science: Back to the future, *PS: Political Science and Politics* **44**, 61 (2011).
- [10] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, A 61-million-person experiment in social influence and political mobilization, *Nature* **489**, 295 (2012).
- [11] D. Igan, P. Mishra, and T. Tressel, A fistful of dollars: Lobbying and the financial crisis, *NBER Macroeconomics Annual* **26**, 195 (2012).
- [12] J. J. V. Bavel *et al.*, Using social and behavioural science to support COVID-19 pandemic response, *Nature Human Behaviour* **4**, 460 (2020).
- [13] E. E. Schattschneider *et al.*, *Politics, Pressures and the Tariff* (Prentice Hall, 1935).
- [14] M. Tripathi, S. Ansolabehere, and J. M. Snyder, Are pac contributions and lobbying linked? new evidence from the 1995 lobby disclosure act, *Business and Politics* **4**, 131 (2002).
- [15] R. L. Hall and A. V. Deardorff, Lobbying as legislative subsidy, *American Political Science Review* **100**, 69 (2006).
- [16] K. Kang, Policy influence and private returns from lobbying in the energy sector, *The Review of Economic Studies* **83**, 269 (2016).
- [17] I. S. Kim and H. V. Milner, Global Goliaths: Multinational corporations in the 21st century economy, in *Multinational Corporations in a Changing Global Economy*, edited by D. Wessel, J. Hines, and C. F. Foley (Brookings Institution, 2020).
- [18] R. J. Brulle, The climate lobby: a sectoral analysis of lobbying spending on climate change in the usa, 2000 to 2016, *Climatic Change* **149**, 289 (2018).
- [19] A. Kennard, The enemy of my enemy: When firms support climate change regulation, *International Organization* **74**, 187 (2020).
- [20] R. Steinbrook, Lobbying, campaign contributions, and health care reform, *New England Journal of Medicine* **361**, e52 (2009).
- [21] I. S. Kim, Political cleavages within industry: Firm level lobbying for trade liberalization, *American Political Science Review* **111**, 1 (2016).
- [22] H. V. Milner and D. Tingley, *Sailing the water's edge: The domestic politics of American foreign policy* (Princeton University Press, 2015).
- [23] M. Bertrand, M. Bombardini, and F. Trebbi, Is it whom you know or what you know? An empirical assessment of the lobbying process, *American Economic Review* **104**, 3885 (2014).
- [24] S. Ansolabehere, J. M. De Figueiredo, and J. M. Snyder Jr, Why is there so little money in us politics?, *Journal of Economic Perspectives* **17**, 105 (2003).
- [25] D. M. J. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, Computational social science, *Science* **323**, 721 (2009).
- [26] A.-L. Barabási, *Network Science* (Cambridge University Press, 2016).
- [27] M. Newman, *Networks* (Oxford University Press, 2018).
- [28] S. Goyal, *Networks, an economic approach* (MIT Press, 2023).
- [29] D. J. Watts, The “new” science of networks, *Annual Review of Sociology* **30**, 243 (2004).
- [30] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Complex networks: Structure and dynamics, *Physics Reports* **424**, 175 (2006).
- [31] S. H. Strogatz, Exploring complex networks, *Nature* **410**, 268 (2001).
- [32] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub, What are higher-order networks?, *SIAM Review* **65**, 686 (2023).
- [33] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics* **74**, 47 (2002).
- [34] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli, The physics of financial networks, *Nature Reviews Physics* **3**, 10.1038/s42254-021-00322-5 (2021).
- [35] J. H. Fowler and N. A. Christakis, Cooperative behavior cascades in human social networks, *Proceedings of the National Academy of Sciences U.S.A.* **107**, 5334 (2010).
- [36] M. Szell, R. Lambiotte, and S. Thurner, Multirelational organization of large-scale social networks in an online world, *Proceedings of the National Academy of Sciences U.S.A.* **107**, 13636 (2010).
- [37] A. L. Traud, P. J. Mucha, and M. A. Porter, Social structure of Facebook networks, *Physica A: Statistical Mechanics and its Applications* **391**, 4165 (2012).
- [38] H. Jeong, Z. Néda, and A.-L. Barabási, Measuring preferential attachment in evolving networks, *Europhysics Letters* **61**, 567 (2003).
- [39] S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, and I. D. Couzin, Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion, *Proceedings of the National Academy of Sciences U.S.A.* **112**, 4690 (2015).
- [40] D. S. Bassett and O. Sporns, Network neuroscience, *Nature Neuroscience* **20**, 353 (2017).
- [41] L. Papadopoulos, M. A. Porter, K. E. Daniels, and D. S. Bassett, Network analysis of particles and grains, *Journal of Complex Networks* **6**, 485 (2018).
- [42] M. Pósfai, B. Szegedy, I. Baćić, L. Blagojević, M. Abért, J. Kertész, L. Lovász, and A.-L. Barabási, Impact of physicality on network structure, *Nature Physics* **20**, 142 (2024).
- [43] H. Albert, R. and Jeong and A.-L. Barabási, Diameter of the world-wide web, *Nature* **401**, 130 (1999).
- [44] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Graph structure in the web, *Computer Networks* **33**, 309 (2000).
- [45] *Supplementary Information*.

- [46] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Multilayer networks, *Journal of Complex Networks* **2**, 203 (2014).
- [47] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J. G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, *Physics Reports* **874**, 1 (2020).
- [48] D. P. Carpenter, K. M. Esterling, and D. M. J. Lazer, The strength of weak ties in lobbying networks: Evidence from health-care politics in the united states, *Journal of Theoretical Politics* **10**, 417 (1998).
- [49] J. Blanes i Vidal, M. Draca, and C. Fons-Rosen, Revolving door lobbyists, *American Economic Review* **102**, 3731 (2012).
- [50] M. E. Shepherd and H. Y. You, Exit strategy: career concerns and revolving doors in congress, *American Political Science Review* **114**, 270 (2020).
- [51] G. Kossinets and D. J. Watts, Empirical analysis of an evolving social network, *Science* **311**, 88 (2006).
- [52] M. E. J. Newman, Clustering and preferential attachment in growing networks, *Physical Review E* **64**, 025102 (2001).
- [53] J. Wang, Y. J. Zhang, C. Xu, J. Li, J. Sun, J. Xie, L. Feng, T. Zhou, and Y. Hu, Reconstructing the evolution history of networked complex systems, *Nature Communications* **15**, 2849 (2024).
- [54] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- [55] G. Bianconi and A.-L. Barabási, Competition and multi-scaling in evolving networks, *Europhysics Letters* **54**, 436 (2001).
- [56] O. T. Courtney and G. Bianconi, Dense power-law networks and simplicial complexes, *Physical Review E* **97**, 052303 (2018).
- [57] B. Bassetti, M. Zarei, M. C. Lagomarsino, and G. Bianconi, Statistical mechanics of the “Chinese restaurant process”: Lack of self-averaging, anomalous finite-size effects, and condensation, *Physical Review E* **80**, 066118 (2009).
- [58] M. P. H. Stumpf and M. A. Porter, Critical truths about power laws, *Science* **335**, 665 (2012).
- [59] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, Economic networks: The new challenges, *Science* **325**, 422 (2009).
- [60] S. Saavedra, F. Reed-Tsochas, and B. Uzzi, Asymmetric disassembly and robustness in declining networks, *Proceedings of the National Academy of Sciences U.S.A.* **105**, 16466 (2008).
- [61] L. M. Bartels, *Unequal democracy: The political economy of the new gilded age* (Princeton University Press, 2016).
- [62] M. Gilens, *Affluence and influence: Economic inequality and political power in America* (Princeton University Press, 2012).
- [63] N. C. Kakwani, Applications of Lorenz curves in economic analysis, *Econometrica: Journal of the Econometric Society* , 719 (1977).
- [64] L. Wittebolle, M. Marzorati, L. Clement, A. Balloi, D. Daffonchio, K. Heylen, P. De Vos, W. Verstraete, and N. Boon, Initial community evenness favours functionality under selective stress, *Nature* **458**, 623 (2009).
- [65] M. P. Fiorina, *Unstable majorities: Polarization, party sorting, and political stalemate* (Hoover Press, 2017).
- [66] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Stability of graph communities across time scales, *Proceedings of the National Academy of Sciences U.S.A.* **107**, 12755 (2010).
- [67] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, *Science* **328**, 876 (2010).

**Supplementary Information for  
‘Measuring the dynamical evolution of the United States lobbying network’**

Karol A. Bacik, Jan Ondras, Aaron Rudkin, Jörn Dunkel, and In Song Kim

**Contents**

I.	Data sources and processing	1
A.	Lobbying report structure	1
B.	<i>LobbyView</i> database	1
C.	Lobbying data	2
D.	Lobbyist association data	7
E.	Political data	7
II.	Lobbying network construction	8
A.	Nodes	8
B.	Edges	9
C.	Weights	9
III.	Order and size of the lobbying network	10
IV.	Degree distributions	10
A.	Average degree	10
B.	Distribution inference	10
C.	Weighted degree	16
D.	Degree correlation	16
V.	Concentration analysis	16
A.	Pareto charts	16
B.	Gini index	17
VI.	Higher order interactions	22
VII.	Lobbying network visualization	22
A.	Node placement	22
B.	Node size and color	23
C.	Final assembly	24
VIII.	Network evolution	24
A.	Rejuvenation	25
B.	Preferential attachment/detachment inference	26
C.	Registrant out-degree dynamics	28
D.	Small component example	29
IX.	Further analysis of bipartite subgraphs	31
A.	Clustering coefficient	31
B.	Largest component	32
X.	Registrant centrality	32
XI.	Portfolio analysis	33
A.	Portfolio matrices	33
B.	Budget-based analysis	36
C.	Principal component analysis	36

XII. Polarization analysis	40
A. Relation-based measures	40
B. Reach-based measures	41
XIII. Lobbying pathways	42
A. Methodology	42
1. Coarse-graining	42
2. Flow construction	43
B. Findings	44
XIV. Data and code availability	45
References	45

## I. Data sources and processing

In this Section, we describe the data processing procedure, from a lobbying disclosure report, to the lobbying network. We start by describing the nature of the data and the general features of the *LobbyView* database, which contains plenty of additional data that we do not analyze here. We then specialize to the subset of data that we use in this article and describe its structure more formally.

### A. Lobbying report structure

The origin of all data provided in *LobbyView* [1] and referred to in this article are disclosures made by registrants, and legally required under the Lobbying Disclosure Act of 1995 [2]. Under current federal regulations, registrants file three standard disclosure forms to maintain compliance with the act, labeled LD-1, LD-2, and LD-203 [3]. Form LD-1 contains the initial registration of a registrant or the initial registration of a new client for the registrant. Form LD-2 is the standard periodic report that forms the basis for most of our analysis. Form LD-203 discloses direct financial contributions from registered lobbyists. LD-2 filings contain a variety of data including the details of the registrant (including unique identification numbers for the U.S. House and U.S. Senate), the client's name, general report metadata, and specific lobbying income or expenses. For each of the 69 "issue codes", registrants report any work completed lobbying these issues during the filing period. They are asked to provide a free-form textual description of what lobbying occurred on this issue. These vary in format but often include a general description of the policy being lobbied (e.g., "Corporate Tax Reform", "Accessibility", "Open Internet"), and/or a list of specific bills (e.g., "S. 698 - Marketplace Fairness Act", "H.R. 2315 Mobile Workforce State Income Tax Simplification Act"). Filings prior to 2007 are typically digitized as scanned versions of text filings, while filings post-2007 are digitally native. In addition to work done by the U.S. Congress to provide filing metadata, we make use of text extraction methods, including regular expressions, which we describe further below. Registrants must also specify which of their lobbyists worked on this particular issue in the period. Most importantly, registrants specify which specific "Contacts" were made in the course of the lobbying. These contacts are either Federal agencies, broadly defined, or the chambers of Congress: Senate and House of Representatives.

Below, in Fig. S1 we provide examples of individual pages taken from two real-world LDA filings, by way of introducing a number of the considerations we describe in more detail below. The leftmost page is a filing from Shell Oil for the first half of 2004 (supplied as a digitized paper copy). This depicts the basic registrant, client, and filing data. Here Shell is acting as an in-house registrant or self-filer (as opposed to a registrant representing an external client). The rightmost page is a filing from Boeing for the third quarter of 2022 (as we describe below, the filing requirements changed from being biannual to quarterly as the LDA was ultimately amended). This page depicts an issue disclosure: Boeing has lobbied a number of bills and general issues under the AER (Aerospace) issue code. The variation in textual representation and the presence of incomplete metadata (e.g., describing the annual NASA Authorization Act as 'H.R. XXXX/S. XXXX') depicts typical difficulties associated with processing data. Boeing reports the names of many lobbyists who worked on this issue, none of whom have new covered position data to report.

### B. *LobbyView* database

The *LobbyView* database is a large, frequently updated, relational database that consists of three major components: (1) all available lobbying reports, registrants, lobbyists (including lobbyist donation activity), and clients. For each of these groups, we retain raw data but also produce cleaned, de-duplicated, standardized versions of the data which add unique identifiers to lobbyist and client entities, as well as providing commonly used firm identifiers necessary for researchers to attach *LobbyView* data to external business datasets including S&P CompuStat [4] and Moody's Orbis [5]; (2) all political donations made by members of the public, which can be linked to firms via firm-level identifiers and donation 'employer' fields (we omit a more detailed treatment of donation data because it is not discussed in this paper); and (3) a complete database of all U.S. Congress legislative activity, legislators, and committee assignments which can be linked to lobbying activity through bill identifiers and bill-committee assignments.

In this section, we describe the general approach to gathering, ingesting, and storing this data. We provide more details about the ingestion and cleaning of specific portions of the data as they are discussed below in the context of the data used in this paper.

*LobbyView* begins by gathering data from public and private sources. Lobbying data is gathered from the U.S. House of Representatives' Lobbying Disclosure website ([disclosurespreview.house.gov](http://disclosurespreview.house.gov)) and the U.S. Senate Lobbying Disclosure website ([lda.senate.gov](http://lda.senate.gov)). The U.S. House of Representatives makes available a bulk data export, which we

Clerk of the House of Representatives      Secretary of the Senate Legislative Resource Center      Office of Public Records B-106 Cannon Building      232 Hart Building Washington, DC 20515		SECRETARY OF THE SENATE 04 AUG 12 AI		
<b>LOBBYING REPORT</b>				
Lobbying Disclosure Act of 1995 (Section 5) – All Filers Are Required to Complete This Page				
<p>1. Registrant Name      Shell Oil Company</p> <p>2. Address      <input type="checkbox"/> Check if different than previously reported 1401 Eye Street, N.W., Suite 1030      Washington, D.C.</p> <p>3. Principal Place of Business (if different from line 2) City: Houston      State/Zip (or Country): Texas, 77002</p> <p>4. Contact Name      Brian P. Malnak      Telephone: 202/466-1410      E-mail (optional):</p> <p>5. Senate ID#      35092-12</p> <p>6. House ID#      31689000</p> <p>7. Client Name      <input checked="" type="checkbox"/> Self</p>				
<p><b>TYPE OF REPORT</b> 8. Year <input checked="" type="checkbox"/> 2004 Midyear (January 1-June 30)      <input type="checkbox"/> Year End (July 1- December 31)</p> <p>9. Check if this filing amends a previously filed version of this report</p> <p>10. Check if this is a Termination Report      <input type="checkbox"/> Termination Date _____</p> <p>11. No Lobbying Activity      <input type="checkbox"/></p>				
<p><b>INCOME OR EXPENSES</b> – Complete Either Line 12 OR Line 13</p> <table border="1"> <tr> <td>12. Lobbying Firms INCOME relating to lobbying activities for this reporting period was: Less than \$10,000      <input type="checkbox"/> \$10,000 or more      <input type="checkbox"/> Income (nearest \$10,000)</td> <td>13. Organizations EXPENSES relating to lobbying activities for this reporting period were: Less than \$10,000      <input type="checkbox"/> \$10,000 or more      <input checked="" type="checkbox"/> Expenses (nearest \$10,000)</td> </tr> </table> <p>14. REPORTING METHOD. Check box to indicate accounting method. See instructions for description of various methods.</p> <p><input type="checkbox"/> Method A. Reporting amounts using LDA data</p> <p><input type="checkbox"/> Method B. Reporting amounts under section 6106 of the Internal Revenue Code</p> <p><input checked="" type="checkbox"/> Method C. Reporting amounts under section 6106 of the Internal Revenue Code</p>			12. Lobbying Firms INCOME relating to lobbying activities for this reporting period was: Less than \$10,000 <input type="checkbox"/> \$10,000 or more <input type="checkbox"/> Income (nearest \$10,000)	13. Organizations EXPENSES relating to lobbying activities for this reporting period were: Less than \$10,000 <input type="checkbox"/> \$10,000 or more <input checked="" type="checkbox"/> Expenses (nearest \$10,000)
12. Lobbying Firms INCOME relating to lobbying activities for this reporting period was: Less than \$10,000 <input type="checkbox"/> \$10,000 or more <input type="checkbox"/> Income (nearest \$10,000)	13. Organizations EXPENSES relating to lobbying activities for this reporting period were: Less than \$10,000 <input type="checkbox"/> \$10,000 or more <input checked="" type="checkbox"/> Expenses (nearest \$10,000)			
<p>Signature _____</p> <p>Filing #5c7366ee-1ce1-4e51-a50f-e0b117eacfca - Page 1 of 26</p>				

Shell H1 2004

Boeing Q3 2022

Figure S1. Sample pages from LDA filings. (**Left**) Filing from Shell Oil for the H1 2004 period. It is a digitized paper copy and depicts the registrant and client information as well as the total monetary expenditure for the period. Shell is filing as a self-filer, reporting lobbying activity done by lobbyists directly employed by Shell on its behalf. (**Right**) The right page is a filing from Boeing for the Q3 2022 period. This page depicts disclosure for a particular issue: Boeing is lobbying Aerospace issues. This page also reports the names of Lobbyists employed by Boeing to lobby for Aerospace issues.

download, extract, and ingest. The U.S. Senate makes available a data API, which we query exhaustively before ingesting. Both sources provide separate unique identifiers for registrants, which we use internally, and for registrant-client relationships. Clients and lobbyists are identified only by a textual representation. In the relevant sections below we describe our efforts to disambiguate clients and lobbyists. We extract issue text descriptions from these filings using regular expressions in order to identify specific legislative bills. We match registrants to associated private sector firms (using CompuStat and Orbis data). We then download relevant congressional data from the United States digital service [6], including a complete record of legislators who have served in the U.S. Congress, their committee assignments, bill sponsorship and cosponsorship data, and bill histories. Thus, we are able to link a corporate firm to a lobby registrant to an LDA report to a bill debated by Congress to the committee that debated it to the legislators on that committee at the time, allowing for some of the rich data flows illustrated in this paper. We use Python to automate downloading the data, which are ingested into a PostgreSQL relational database and indexed appropriately.

The data in *LobbyView* are made available to researchers in a variety of forms – the access options are described in Section XIV.

### C. Lobbying data

We will now describe more precisely, the subset of the *LobbyView* data that we analyze in this study. Here, we use only the LD-2 reports (see Sec. IA) and we exclude the reports that indicate 'no lobbying activity' in line 11, as well

as the reports that are superseded by updated filings, based on line 9 of LD-2 filings.<sup>1</sup> Each lobbying report that we analyze is assigned a unique filing ID  $f$ , as well as nine other features of interest:

1. Filing year  $y(f) \in \{1999, 2000, \dots, 2023\}$ ,
2. Self-filing flag  $s(f) \in \{\text{True}, \text{False}\}$ ,
3. Monetary amount  $m(f) \in \mathbb{R}^+$ ,
4. Client ID  $c(f) \in \{c_1, c_2, \dots\}$ ,
5. Registrant ID  $r(f) \in \{r_1, r_2, \dots\}$ ,
6. Set of lobbyist IDs  $L(f) \subseteq \{l_1, l_2, \dots\}$ ,
7. Set of approached government entities (House(s) of Congress and Federal agencies)  $G(f) \subseteq \{g_1, g_2, \dots\}$
8. Set of general issue areas  $A(f) \subseteq \{a_1, a_2, \dots, a_{69}\}$  (c.f. Fig. 4 of the main text).
9. Set of bills  $B(f)$

Combined into a tuple,

$$\mathcal{LD}(f) = (f, y(f), s(f), m(f), c(f), r(f), L(f), G(f), A(f)), \quad (\text{S.1})$$

they form a *lobbying datum*, which is the smallest atom of our *lobbying data*. We will now describe the basic characteristics and technical challenges associated with each one of the eight features. It is worth emphasizing that for the purposes of this study, the filing IDs, client IDs, registrant IDs, and lobbyist IDs have been anonymized.

*Filing year*

Reports analyzed in this study cover the period 1999–2023. Although the LDA came into force in 1996, reports from 1996–1998 were not centrally archived and are not widely available.

Figure S2 presents the number of reports considered in our analysis for each year  $y \in \{1999, 2000, \dots, 2023\}$ . We immediately notice an abrupt increase in the number of reports in 2008. This is because in 2007 Congress passed the Honest Leadership and Open Government Act [7] which, among numerous other provisions, altered the mandatory reporting frequency provisions of the LDA 1995. Prior to 2008, reports were submitted biannually (twice a year). Since 2008, reports have been required each quarter. This is reflected in our data, where the number of reports doubled beginning in 2008. Filing and clerical errors during the transition period contribute to a small number of reports that have invalid filing period codes (e.g., reporting for Q2 in 2006 or for H2 in 2009). In order to unify our data across this policy transition, we aggregate the periodic data into annualized data, thus discarding the exact filing code and considering only the reporting year  $y$  (line 8 of LD-2 form).

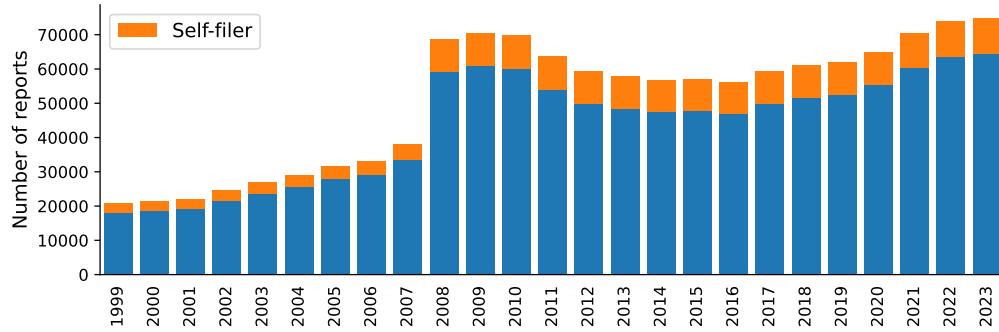


Figure S2. Number of lobbying reports used in our analysis. Most of the reports are filed by lobbying firms lobbying on behalf of a client (as opposed to self-filers). Note that the quantity of filings before 2008 is not directly comparable to those since 2008 due to a legislative change in reporting frequency. In total, we include in our analysis 1,277,411 filings.

---

<sup>1</sup> When we refer to specific lines of the LD-2 form, the reader is welcome to consult Fig. S1.

Table S1. Lobbying activity minimum reporting amount in terms of the year and registrant type. Extracted from [2, 7–9].

Reporting year $y$	Not a self-filer	Self-filer
< 2008	\$5,000.00	\$10,000.00
$\geq 2008$	\$3,000.00	\$5,000.00

### Self-filing

Line 7 of LD-2 provides information on whether the filing pertains to an organization lobbying on its own behalf (self-filer flag  $s(f) = \text{True}$ ) or a lobbying firm lobbying on behalf of a different client (self-filer flag  $s(f) = \text{False}$ ). Figure S2 shows that most of the reports are filed on behalf of someone else.

### Monetary amount

If the lobbying costs exceed a certain threshold (Table S1), the lobbying report must report a dollar value (lines 12 and 13 of LD-2 filings). This field differs depending on the identity of the filer. Registrants who are self-filers (for whom  $s(f) = \text{True}$ ) report the *cost* of their lobbying activities (line 13 of LD-2 filings), while registrants lobbying on behalf of a client ( $s(f) = \text{False}$ ) report the *income* they received for lobbying (line 12 of LD-2 filings). We expect that firms book income that meets or exceeds the actual costs of their activities. Additionally, self-filers are subject to a higher numeric threshold for reporting lobbying costs than non-self-filers are for reporting income (Tab. S1).

To estimate the money flow in lobbying, we neglect some of these subtleties and we attach a single monetary amount  $m(f)$  to each filing. For most filings,  $m(f)$  is simply the amount reported in either line 12 or line 13. If the amount is below the reporting threshold (listed in Table S1), we set  $m(f)$  to this threshold value (for low-cost lobbying, this is an upper bound of the actual unknown value). Figure ?? compares the amount of this low-cost lobbying,

$$\sum_f \mathbb{1}[m(f) \text{ bounded heuristically} \wedge y(f) = y] m(f)$$

to the total amount and shows that this lobbying accounts for a very small percentage of our data (in part because lobbying firms would prefer to only report when they are legally compelled to do so, and in part because lobbying is expensive). We do not believe that this heuristic adjustment has any material impact on our results.

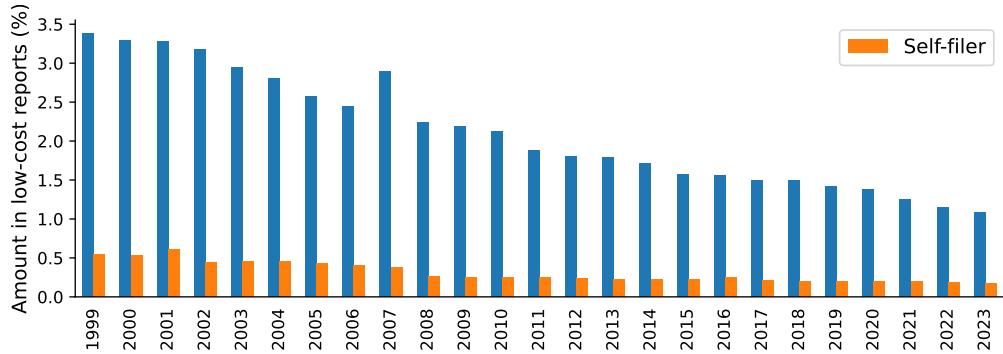


Figure S3. The upper bound of the total monetary amount associated with the ‘low-cost reports’, i.e., the reports where the lobbying income/expenditure falls below the threshold value. Note a steady decrease in the proportion of the lobbying market taken by the low-cost activities.

In Fig. S4, we show the total annual amount associated with self-lobbying

$$\sum_f \mathbb{1}[s(f) = \text{True} \wedge y(f) = y] m(f)$$

and the total annual amount associated with third-party lobbying

$$\sum_f \mathbb{1}[s(f) = \text{False} \wedge y(f) = y] m(f).$$

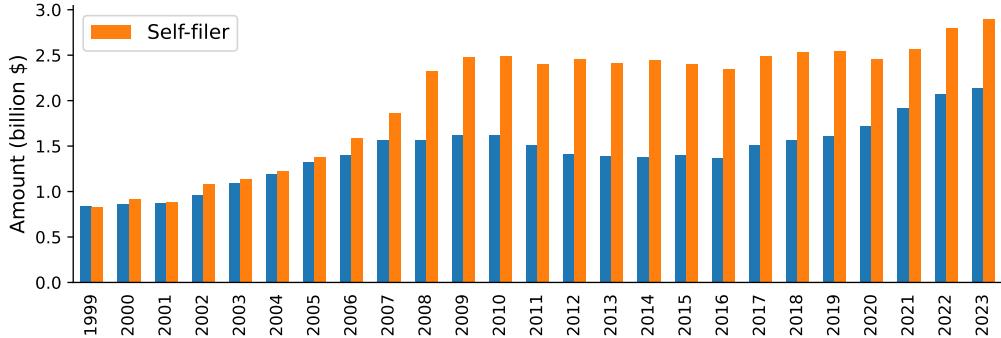


Figure S4. Total lobbying amount for all the reports in a given year. Due to a different definition of the lobbying amount, the figures for self-filers and not-self-filers are not exactly comparable. The total amount reported in these reports is \$87.5 billion.

We note that the latter amount is often larger, even though third-party lobbying is responsible for a larger number of reports (Fig. S2). This difference might be partly due to the differences in filing requirements, but also due to structural differences between externally lobbying firms and in-house lobbying departments.

#### *Client*

Each LD-2 filing discloses the name of the associated client in line 7. Clients include private firms, NGOs and interest groups, municipalities, and more. Because this is a free-form text block, it is non-trivial to isolate the exact real-world entity associated with a client name. The same firm might be a lobbying client on many filings under a variety of different representations. These issues range from minor ('APPLE, INC.' and 'APPLE INC.' are the same firm) to more complex ('APPLE COMPUTER' and 'APPLE INC.' are the same firm). Clients are not assigned unique identifiers under the LDA. Researchers who fail to bridge the gap between representations and underlying entities run the risk of dramatically overestimating the number of unique clients, as well as failing to accurately assess each client's expenditure or level of activity. To solve this problem, we use a search engine-based disambiguation process. In summary, using an online search engine, we perform a search for the representation of a client name. From the results, we identify the website the search engine associates with the client name. When two representations are attached to the same website, we consider them the same entity. This approach empirically performs very well for current entities, and perfectly for S&P 500 companies; effectively, search engines are designed to perform disambiguation of ambiguous search queries, so this should not be a surprise. Although this approach has some limitations,<sup>2</sup> it improves on rules-based approaches (e.g., string distance, cosine similarity, etc.) typically used in these settings.

The total number of distinct clients that we identify in any given year will be presented in Fig. S6A, alongside other entities.

#### *Registrant*

The lobbying report is submitted by the registrant  $r$ , which could be an organization, lobbying firm, or a self-employed individual. Registrants are identified by unique U.S. Senate and U.S. House IDs, which must be disclosed on lines 5 and 6 of the LD-2 filing, and so no further disambiguation is required.<sup>3</sup> Some registrants file only as self-filers. We call these registrants *In-House* registrants – as one example, consider industry-based interest groups such as the Chamber of Commerce of the United States of America, which has reported \$1.73 billion in lobbying expenditures over the life of our dataset, all self-filed; or larger corporate entities such as General Electric, which has reported \$364 million in in-house lobbying expenditures over the life of our dataset. Other registrants file also on behalf of other (as non-self-filers), in which case we call them *K-Street* registrants. We classify each registrant as

<sup>2</sup> Key limitations include the fact that long-bankrupt firms typically have very poor internet presence and that certain classes of non-public-facing firms are more likely to resolve to third-party directories and resources than to have their own internet presence. We take measures to mitigate these limitations: we attempt to use historical data taken from Wikipedia to identify cases where firms are defunct in order to identify the risk of a false positive, and we use a blocklist to remove a wide swath of online firm directories.

<sup>3</sup> As with clients, we are also able to connect registrants to external corporate identifiers by the same search engine-driven disambiguation method described above; although this is not directly relevant to this paper, we make linking identifiers available in our broader dataset.

*In-House* or *K-Street*, reviewing all the lobbying reports in the database, such that

$$\begin{aligned} r \in \text{In-House} &\Leftrightarrow \forall_{f: r(f)=r} s(f) = \text{True}, \\ r \in \text{K-Street} &\Leftrightarrow \exists_{f: r(f)=r} s(f) = \text{False}. \end{aligned} \quad (\text{S.2})$$

The evolution of the total number of registrants is presented in Fig. S6B, and an in-depth analysis of the dynamics of K-Street registrants is shown in Fig. 2B of the main article.

K-Street registrants (in the sense of our definition) sometimes also submit reports as self-filers, but Fig. S5 shows that such cases constitute only a modicum of all the lobbying data.

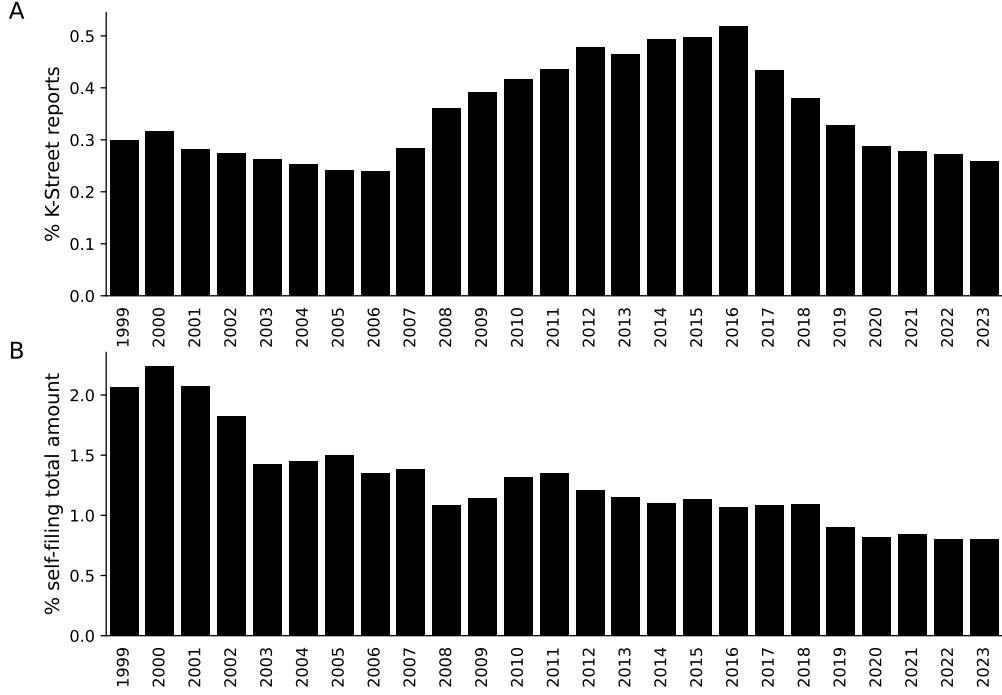


Figure S5. (A) Percentage of lobbying reports submitted by K-Street registrants (according to definition (S.2)) on their own behalf (as self-filers). (B) Percentage of the total self-filing amount generated by self-reports of K-Street registrants decreases in time.

### Lobbyists

In addition to client  $c$  and registrant  $r$ , the lobbying report also provides information about lobbyists  $L(f)$  (line 18 of LD-2 filings). Note that each report contains exactly one client and exactly one registrant, but there is no limit to the number of lobbyists, and some reports mention no lobbyists at all.

Because lobbyists are identified uniquely by names, and not any other identifier, they also require disambiguation. If a lobbyist named Maria Gutierrez appears on multiple filings across multiple registrants, it may be the case that she is a single lobbyist who worked for multiple registrants. Or it may be the case that there are multiple lobbyists with the same name. Other cases are more complex: Are Jon Smith, John Smith, and Jonathan Smith the same lobbyist? Disambiguating individual names using our firm-name approach is currently infeasible, so we use a rules-based approach.

Figure S6C shows how the total number of active lobbyists has been changing over time.

### Government entities

A lobbying report provides information about the approached government entities (House(s) of Congress and Federal agencies)  $G(f)$  (line 17). Similarly to the set of lobbyists  $L(f)$ , the set of approached government entities can be

empty or arbitrarily large. To be precise, the report can mention one government entity multiple times, in relation to different issues, but we disregard this multiplicity. Because entities are reported in an open text format, work is needed to disambiguate entities (e.g., the U.S. Department of Defense might be represented as ‘Defense’, ‘Dept of Defense’, ‘Defense - Dept of (DOD)’, or other representations). Because U.S. agencies are a closed set, the matching task is simpler than the previously described fields: we perform regular expression matching to extract and clean these entities. The full set of government entities considered will be discussed in Sec. XI.

#### *General issue areas*

The set of general issue areas  $A(f)$  is derived from their abbreviations/codes on line 15 of the Lobbying Disclosure Form. The issue areas come from a standardized list of 79 items. All of them can be found on the y-axis of Fig. 4A of the main article.

#### *Bills*

This set includes all the bills that the lobbyists reportedly lobbied on. The bills are referenced with their number, which allows us to relate the lobbying activity to the Congressional data. In this study, we only use the bill data in the section dedicated to polarization (c.f. Sec. XII).

#### **D. Lobbyist association data**

From mandatory disclosures made in LD-2 filings (line 18), we also gather *association data*, which describes the past professional relationships of registered lobbyists.

When registering, a lobbyist discloses if they have a past history as a *covered official*, which is a legal term under the LDA that includes a variety of positions and offices. In brief, covered officials include all elected officials in the federal government, their assistants and Chiefs of Staff, civil servants who are political appointees, executives of federal agencies or departments (this requirement is broad enough to cover U.S. Parole Commission commissioners, National Parks Service directors, and federally-operated regional electrical agencies), high-ranking military officers, and more. The LDA requires that officials disclose such past positions as part of the public interest in understanding “revolving-door” lobbying. Once disclosed in a filing, the positions need not be disclosed again. Thus, an individual lobbyist will typically disclose their past positions when they begin lobbying and never again. Should the lobbyist return to public service and serve in other covered positions, they are obligated to disclose if they subsequently lobby. The lobbyists do not need to provide detailed timeline information for covered positions, and so it is impossible to identify from the filings whether a lobbyist’s employment was recent or in the distant past. We apply text processing methods including regular expressions to identify the exact position from the free-form text and match the official to agencies, departments, or individual legislators.

We compile our findings in a *government association function*  $g(l, y)$  and the *political association function*  $p(l, y)$  which take as an input lobbyist ID  $l$  and year  $y$ .

The government association function outputs the set of government entities that the lobbyist worked for that have been mentioned in the LDA reports in year  $y$  or earlier. Controlling for the year, i.e., using  $g(l, y)$ , instead of  $g(l)$ , allows us to avoid using associations reported in the future to make conclusions about the past. By construction,

$$y_1 < y_2 \Rightarrow g(l, y_1) \subseteq g(l, y_2).$$

The political association function  $p(l, y)$  is an analogous set of all the present legislators (members of the Congress in year  $y$ ) that lobbyist  $l$  reported working for (in year  $y$  or earlier). Note that this time

$$y_1 < y_2 \not\Rightarrow p(l, y_1) \subseteq p(l, y_2),$$

as the legislator might not be in office anymore.

#### **E. Political data**

Our data on legislators is sourced directly from the United States digital service congressional legislator database [6]. We retrieve and ingest this *political data* for each legislator (politician)  $p$ , and we construct the *affiliation functions*

$\text{Party}(p, y)$  and  $\text{Chamber}(p, y)$  providing information about the partisan affiliation and Congress chamber of the legislator  $p$  in year  $y$ .

Party affiliations of the President, House majority, and Senate majority (shown in Fig. 2–4 of the main text) are taken from <https://history.house.gov/Institution/Presidents-Coinciding/Party-Government/>.

## II. Lobbying network construction

After the initial processing described in Sec. I, we use the lobbying data (Sec. IC), affiliation data (Sec. ID), and the political data (Sec. IE) to construct year-indexed multi-layer directed graphs  $\mathcal{G}_{1999}, \mathcal{G}_{2000}, \dots, \mathcal{G}_{2023}$ , that we call *lobbying networks*. The nodes and edges of  $\mathcal{G}_y$  are denoted by  $V(\mathcal{G}_y)$  and  $E(\mathcal{G}_y)$  respectively. The lobbying networks are constructed for each year separately. In this Section, we describe the details of their construction.

### A. Nodes

The lobbying network  $\mathcal{G}_y$  has 5 different sets of nodes, organized in layers:

1. The first (most upstream) layer of  $\mathcal{G}_y$  corresponds to *client nodes*

$$\text{Clients}(y) = \bigcup_{f: y(f)=y} c(f). \quad (\text{S.3})$$

2. The second layer of  $\mathcal{G}_y$  comprises *registrant nodes*

$$\text{Registrants}(y) = \bigcup_{f: y(f)=y} r(f) \quad (\text{S.4})$$

which are further divided based on the registrant type (eq. (S.2)) into two disjoint subsets

$$\text{In-House Registrants}(y) = \{r \in \text{Registrants}(y) : r \in \text{In-House}\}, \quad (\text{S.5})$$

and

$$\text{K-Street Registrants}(y) = \{r \in \text{Registrants}(y) : r \in \text{K-Street}\}. \quad (\text{S.6})$$

3. The third layer of  $\mathcal{G}_y$  consists of *lobbyist nodes*

$$\text{Lobbyists}(y) = \bigcup_{f: y(f)=y} L(f). \quad (\text{S.7})$$

These first three layers use only the lobbying data  $\mathcal{LD}$  defined in eq. (S.1).

4. To construct the fourth layer of  $\mathcal{G}_y$  representing *government entity nodes*, we also use the government association function  $g(l, y)$  defined in Sec. ID. Thus,

$$\text{Government entities}(y) = \bigcup_{f: y(f)=f} \bigcup_{l \in L(f)} g(l, y), \quad (\text{S.8})$$

is the set of all the government entities with a historical employment link to at least one of the lobbyists, with the restriction that the connection must have been mentioned in year  $y$  or earlier.

5. The fifth layer of  $\mathcal{G}_y$  corresponding to *legislator nodes*, is constructed similarly by using the political association function  $p(l, y)$  defined in Sec. ID. Thus,

$$\text{Legislators}(y) = \bigcup_{f: y(f)=f} \bigcup_{l \in L(f)} p(l, y) \quad (\text{S.9})$$

is the set of all active legislators (present in Congress in year  $y$ ) with a historical employment link to at least one of the lobbyists, with the restriction that the connection must have been mentioned in year  $y$  or earlier.

## B. Edges

The lobbying network  $\mathcal{G}_y$  has 5 different sets of edges:

1. Clientships( $y$ ): connections between clients and registrants (first to second layer), such that a directed edge  $(c, r)$  linking client  $c$  and registrant  $r$  exists if and only if the registrant lobbied for the client, i.e.,

$$(c, r) \in \text{Clientships}(y) \Leftrightarrow \exists f : y = y(f) \wedge c = c(f) \wedge r = r(f). \quad (\text{S.10})$$

which are further divided based on the registrant type (eq. (S.2)) into two disjoint subsets

$$\text{In-House Clientships}(y) = \{(c, r) \in \text{Clientships}(y) : r \in \text{In-House}\}, \quad (\text{S.11})$$

and

$$\text{K-Street Clientships}(y) = \{(c, r) \in \text{Clientships}(y) : r \in \text{K-Street}\}. \quad (\text{S.12})$$

2. Lobbyist contracts( $y$ ): connections between registrants and lobbyists (second to third layer), such that a directed edge  $(r, l)$  linking registrant  $r$  and lobbyist  $l$  exists if and only if the lobbyist was mentioned in a report filed by the registrant, i.e.,

$$(r, l) \in \text{Lobbyist contracts}(y) \Leftrightarrow \exists f : y = y(f) \wedge r = r(f) \wedge l \in L(f). \quad (\text{S.13})$$

which are further divided based on the registrant type (eq. (S.2)) into two disjoint subsets

$$\text{In-House Lobbyist contracts}(y) = \{(r, l) \in \text{Lobbyist contracts}(y) : r \in \text{In-House}\}, \quad (\text{S.14})$$

and

$$\text{K-Street Lobbyist contracts}(y) = \{(r, l) \in \text{Lobbyist contracts}(y) : r \in \text{K-Street}\}. \quad (\text{S.15})$$

3. Government associations( $y$ ): connections between lobbyists and government entities (third to fourth layer), such that a directed edge  $(l, g)$  linking lobbyist  $l$  and government entity  $g$  exists if and only if an employment connection between the two has been documented in year  $y$  or earlier, i.e.,

$$(l, g) \in \text{Government associations}(y) \Leftrightarrow l \in \text{Lobbyists}(y) \wedge g \in g(l, y), \quad (\text{S.16})$$

where  $g(l, y)$  is the government association function defined in Sec. ID.

4. Legislator associations( $y$ ): connections between lobbyists and legislators (third to fifth layer), such that a directed edge  $(l, p)$  linking lobbyist  $l$  and legislator  $p$  exists if and only if an employment connection between the two has been documented in year  $y$  or earlier, and the legislator is a Congress member in year  $y$ , i.e.,

$$(l, p) \in \text{Legislator associations}(y) \Leftrightarrow l \in \text{Lobbyists}(y) \wedge p \in p(l, y), \quad (\text{S.17})$$

where  $p(l, y)$  is the political association function defined in Sec. ID.

All the edges are directed and unweighted (c.f. Sec. II C) and we disregard any multi-edges.

## C. Weights

Although by default our lobbying network is unweighted, we could attach weights  $w$  to the client-registrant connections by adding all the monetary values in corresponding filings:

$$w[(c, r) \in E(\mathcal{G}_y)] = \sum_f \mathbb{1}[y(f) = y \wedge c(f) = c \wedge r(f) = r] m(f). \quad (\text{S.18})$$

### III. Order and size of the lobbying network

In this Section, we discuss the number of nodes (order) and the number of edges (size) of the lobbying network. The year-to-year evolutions of order and size are presented in Fig. S6 and Fig. S7, respectively.

The number of clients (c.f. eq. (S.3)) steadily increases until the financial crisis (2007–2008) after which it declines for about the next 7 years before the steady increase is recovered. A similar pattern can be observed for the total number of registrants and lobbyists (c.f. eq. (S.4) and (S.7)), as well as the number of connections between the three upstream layers (clientships and lobbyist contracts, c.f. eq. (S.10) and (S.13)).

Note that the K-Street registrant count (c.f. eq. (S.6)) dominates in the early years while the In-House registrant count (c.f. eq. (S.5)) dominates in the more recent years. Moreover, the clientships involving K-Street registrants outnumber the clientships involving In-House registrants. Nevertheless, lobbyist contracts with In-House registrants slightly outnumber those with K-Street registrants, especially in the more recent years.

The number of government entities (c.f. eq. (S.8)) included in the lobbying network, and the corresponding number of lobbyist-government associations (c.f. eq. (S.16)) grows steadily over time. In Fig. S8, we show which government entities have best known ties to the lobbyists (highest in-degree) and how this quantity varies in time.

The number of legislators (c.f. eq. (S.9)) and the lobbyist-legislator associations (c.f. eq. (S.17)) also steadily grows in time. In Fig. S9 we show that since 2008 around 80% of Senators and 50% of Representatives, from both the Democratic and Republican parties, are included in our lobbying network every year. By construction, the legislator is included if there is at least one active lobbyist who, according to our database, worked for this legislator in the past.

### IV. Degree distributions

In a given year  $y$ , let the out-degree (the number of outgoing edges) of a node type  $i$  and in-degree (the number of incoming edges) of a node type  $j$  be

$$\text{out-deg}(i, y) = |\{(i, j) \in E(\mathcal{G}_y)\}|, \quad \text{and} \quad \text{in-deg}(j, y) = |\{(j, i) \in E(\mathcal{G}_y)\}|. \quad (\text{S.19})$$

For client nodes, we can further distinguish their out-degree to K-Street and In-House registrants as

$$\text{out-deg}_K(c, y) = |\{(c, r) \in E(\mathcal{G}_y) : r \in \text{K-Street}\}|, \quad \text{and} \quad \text{out-deg}_I(c, y) = |\{(c, r) \in E(\mathcal{G}_y) : r \in \text{In-House}\}|. \quad (\text{S.20})$$

#### A. Average degree

First insight into the connectivity of the lobbying network and its temporal evolution can be gained by simply analyzing the average in- and out-degree of each network layer. These data are plotted in Fig. S10. The most dynamic is the average in-degree of government entities and registrants indicating that, at least as far as our database is concerned, government entities become connected to increasingly more lobbyists and registrants to increasingly more clients. The average degrees for the remaining layer-to-layer connections seem to be close to a constant.

#### B. Distribution inference

Following the methodology of Ref. [10], in this study, we characterize the degree distribution by computing the complementary cumulative distribution function (CCDF). For example, in Fig. 2F of the main article, we plot the CCDFs of  $\text{in-deg}(r, y)$  for  $r \in \text{K-Street}$ . To be precise, this CCDF  $\rho$  for this particular distribution is defined as

$$\rho(d, y) = \frac{\sum_{r \in \text{K-Street Registrants}(y)} \mathbb{1}[\text{in-deg}(r) \geq d]}{|\text{K-Street Registrants}(y)|}. \quad (\text{S.21})$$

In a similar way, in Fig. S11 we plot the CCDFs of other degree distributions.

Some of the distributions, e.g. the out-degree of clients (Fig. S11A) or registrants (Fig. S11B) can be classified as heavy-tail, locally approximated by power laws. Note that the difference  $\rho'(d) - \rho(d+1)$  is the probability mass function (PMF), so when CCDF can be locally approximated as  $\rho(d) \sim d^{-\gamma+1}$ , then the PMF can be locally approximated with a power law  $\sim d^{-\gamma}$ . While exact exponent estimation is beyond the scope of our study, we can still make broad-stroke comparisons between the degree distributions in the lobbying network and in other complex networks [10].

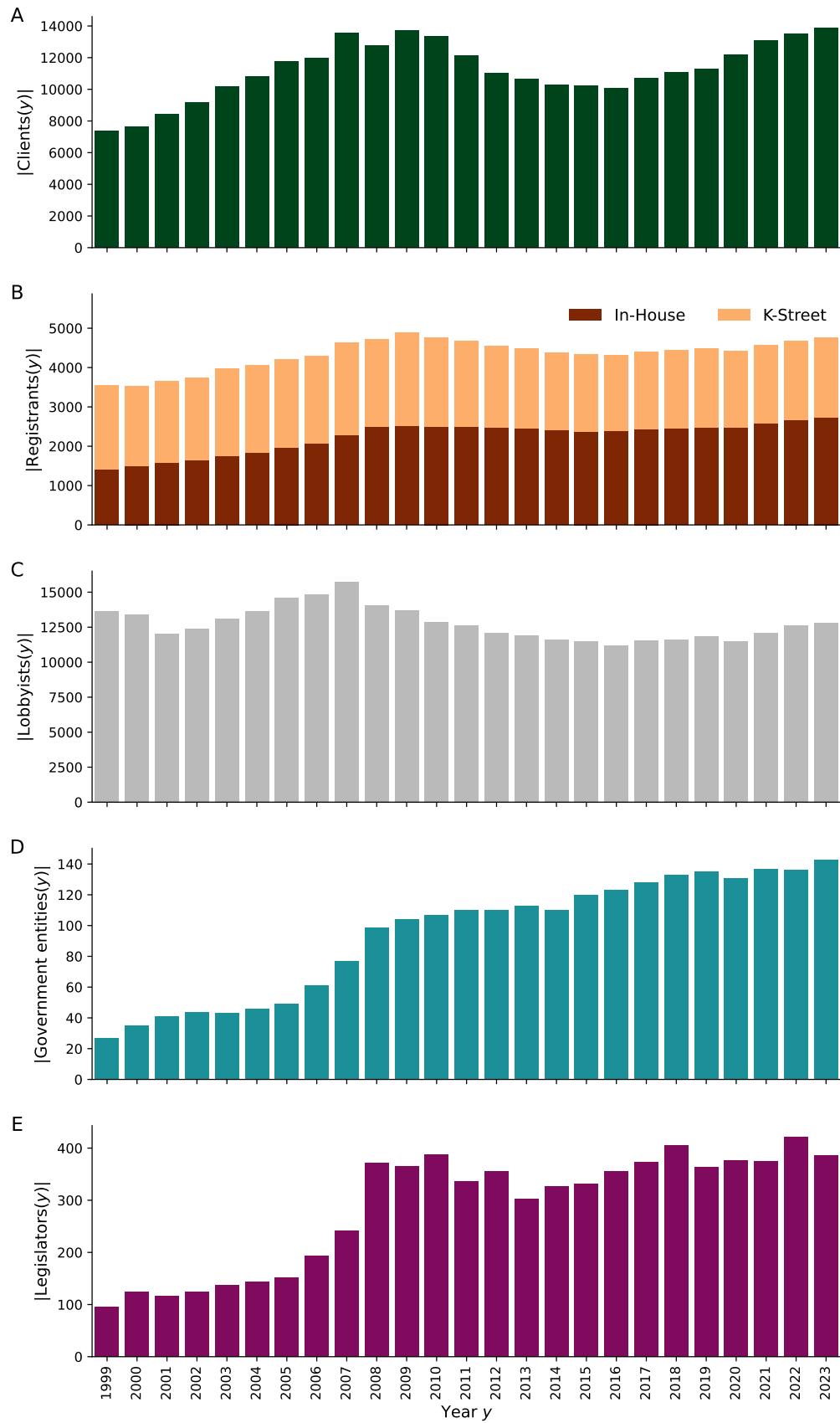


Figure S6. Order of the lobbying network each year. (A) The number of distinct clients in a given year. (B) The number of distinct K-Street and In-House registrants in a given year. (C) The number of distinct lobbyists in a given year. (D) The number of distinct government entities in a given year. (E) The number of distinct legislators in a given year. Note that the set of registrants is significantly smaller than the sets of clients and lobbyists, which are of comparable size.

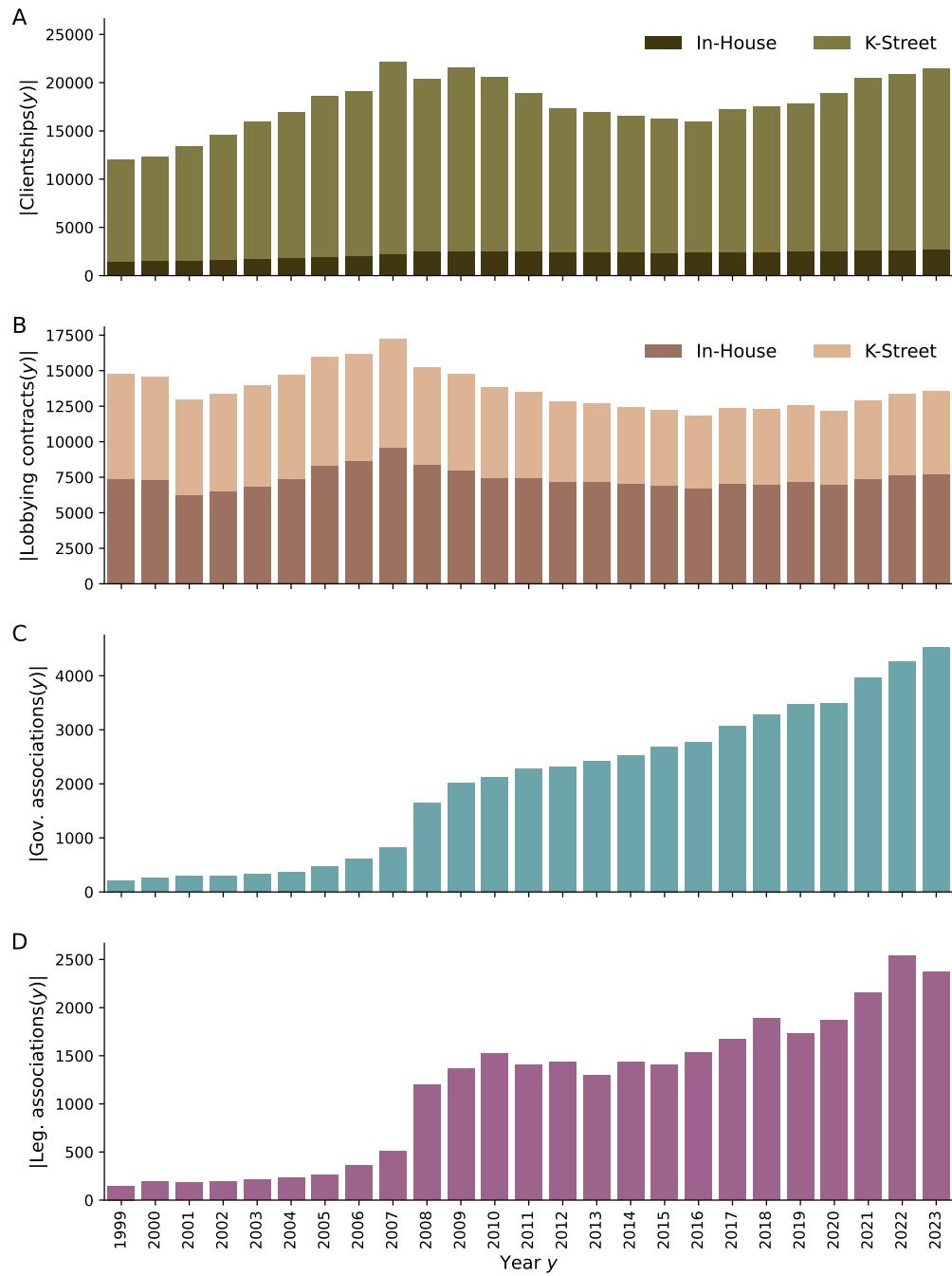


Figure S7. Size of the lobbying network each year. (A) The number of distinct clientships (client-registrant edges) in a given year. (B) The number of distinct lobbying contracts (registrant-lobbyist edges) in a given year. (C) The number of distinct government associations (lobbyist-government entity edges) in a given year. (D) The number of distinct legislator associations (lobbyist-legislator edges) in a given year.

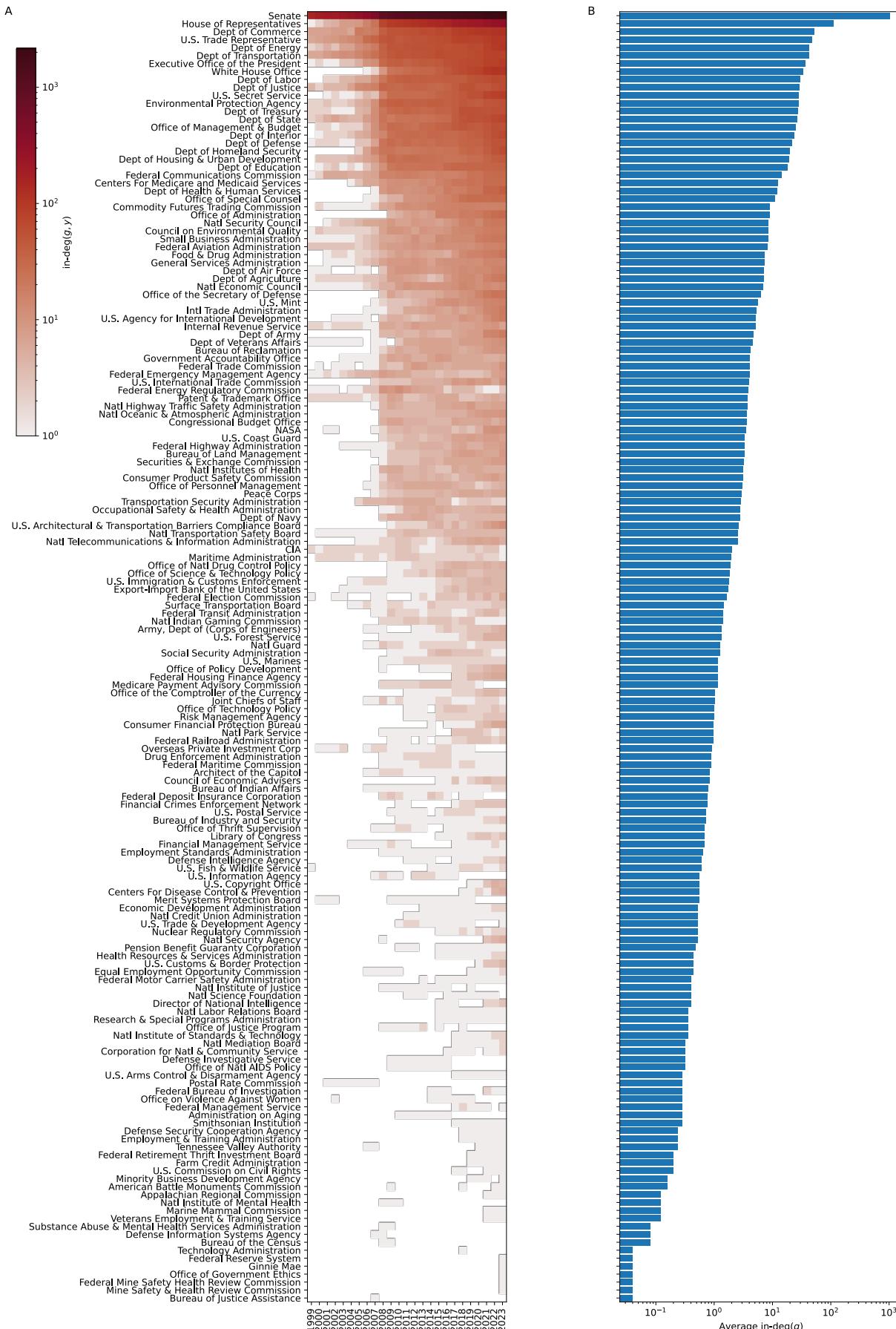


Figure S8. In-degree (number of identified lobbyist associations) of specific government entities. (A) By year. (B) Time-averaged. The entities are sorted by the average government entity in-degree.

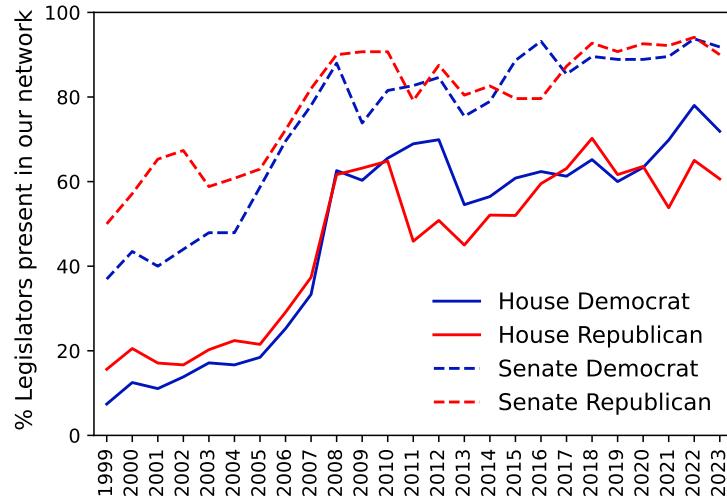


Figure S9. Percentage of legislators included in the lobbying network, in terms of the legislator's chamber and party affiliations.

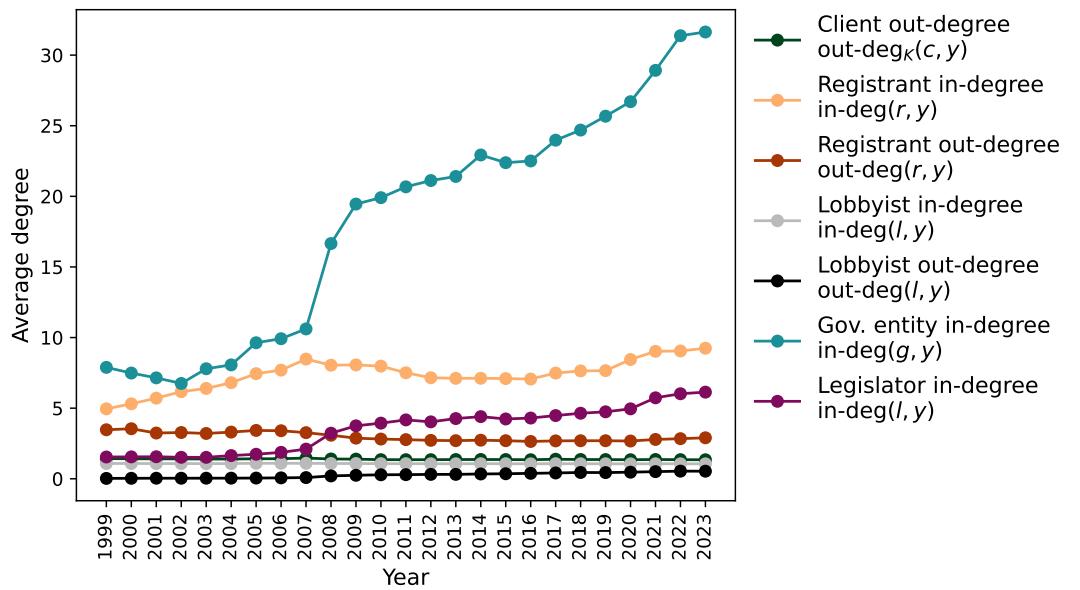


Figure S10. Average in/out-degree of each lobbying network layer in years 1999–2023. The client out-degree, registrant in-degree, and registrant out-degree are all computed for K-Street registrants.

For example, the client out-degree distribution (Fig. S11A) locally follows a power law with the exponent  $\approx 3$  for  $\text{out-deg}_K(c, y) \lesssim 30$ . This value is on a par with  $\gamma = 3$  found for the Barabasi-Albert model [11], and in the range of real-world networks [10]. We need to remember, however, that all the degree distributions we compute appear in the context of a multipartite graph, so the Barabasi-Albert model is not exactly applicable. The generative process of bipartite network connectivity might be more akin to the Pitman-Yor process [12, 13], which can yield exponents  $\gamma \in (1, 2]$ . Nevertheless, fitting a network evolution model to our lobbying network is beyond the scope of this work.

The heavy-tail degree distributions of clients and registrants is in stark contrast with the exponential distribution of contracts and associations among lobbyists. Most lobbyists work for only one registrant and the number of lobbyists with multiple contracts decays exponentially with the number of contracts (Fig. S11C). In Sec. VI of [14], we show that the distribution of reports (jobs) among individual lobbyists in large lobbying firms is also exponential (except for a small set of ‘super-lobbyists’ who work with more than 200 clients). Furthermore, most lobbyists (ca. 90%) do not have any documented historical associations to government entities and legislators, and the number of lobbyists with  $a > 0$  associations decays exponentially with  $a$  (Fig. S11D). Our results indicate, therefore, that the distribution of connections among individual lobbyists is more egalitarian than the distribution of influence among the lobbying firms.

The structural differences between the registrant layer and the lobbyist layer are likely a consequence of distinct evolutionary principles, which govern the development of the lobbying network. Hierarchical degree distributions can spontaneously arise as a consequence of accumulated advantage (Matthew Principle) [13, 15, 16], and exponential distribution is reminiscent of random Erdős-Renyi graphs, where each new link is added independently [17].

For completeness, in (Fig. S11E-F) we show the in-degree distribution of government entities and legislators. The graph of the former has a characteristic kink indicating the possibility of two qualitatively different sets of government entities (large and small). The latter distribution can be well approximated by an exponential function in the entire range.

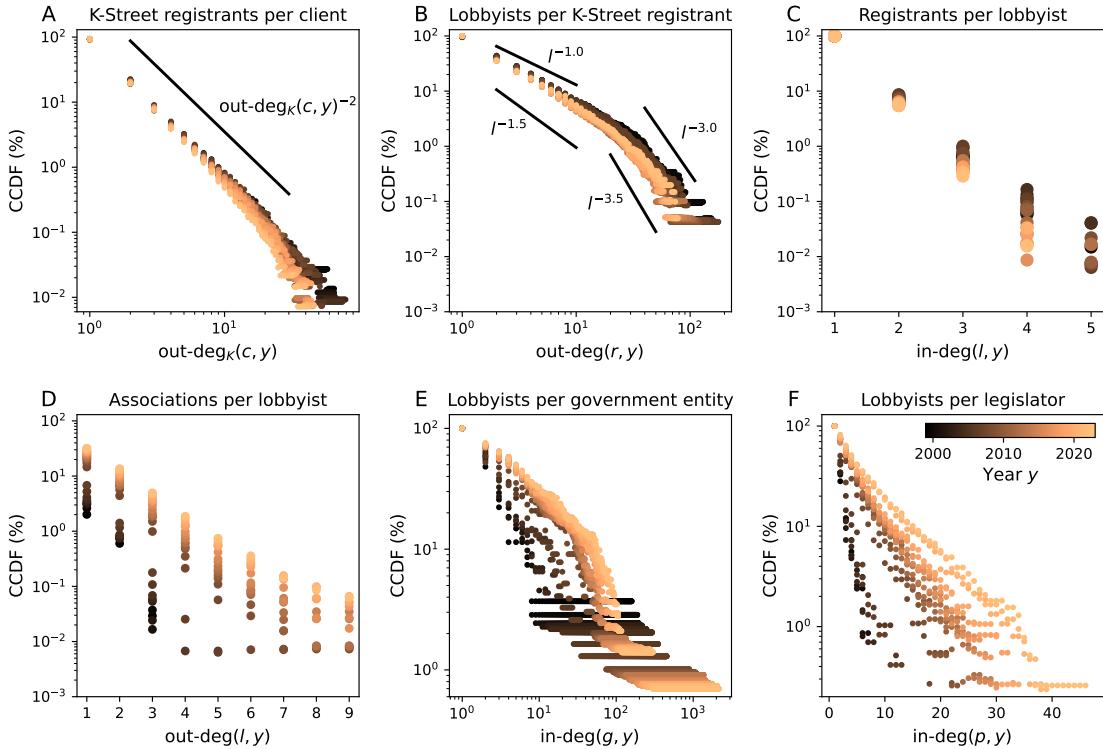


Figure S11. Complementary cumulative distributions (CCDFs) of node in/out-degrees. (A) CCDF of the client out-degree  $\text{out-deg}_K(c, y)$ : % of clients with  $\text{out-deg}_K(c, y)$  or more K-Street registrants. (B) CCDF of K-Street registrant out-degree  $\text{out-deg}(r, y)$ : % of K-Street registrants with  $\text{out-deg}(r, y)$  or more lobbyists. (C) CCDF of lobbyist in-degree  $\text{in-deg}(l, y)$ : % of lobbyists employed by  $\text{in-deg}(l, y)$  or more registrants. (D) CCDF of lobbyist out-degree  $\text{out-deg}(l, y)$ : % of lobbyists with  $\text{out-deg}(l, y)$  or more associations (government entities and legislators). (E) CCDF of the government entity in-degree  $\text{in-deg}(g, y)$ : % of government entities with  $\text{in-deg}(g, y)$  or more lobbyists. (F) CCDF of the legislator in-degree  $\text{in-deg}(p, y)$ : % of legislators with  $\text{in-deg}(p, y)$  or more lobbyists. All CCDFs are computed separately for each year in the range 1999–2023.

### C. Weighted degree

As explained in Sec. II C, we can attach weights (monetary values) to the client-registrant edges  $(c, r)$ . We thus define the weighted out-degree of a client  $c$  to K-Street and In-House registrants as

$$\text{w-out-deg}_K(c, y) = \sum_{\substack{(c,r) \in E(\mathcal{G}_y), \\ r \in \text{K-Street}}} w[(c, r)]. \quad \text{and} \quad \text{w-out-deg}_I(c, y) = \sum_{\substack{(c,r) \in E(\mathcal{G}_y), \\ r \in \text{In-House}}} w[(c, r)]. \quad (\text{S.22})$$

which represent the total *lobbying expenses* of the client  $c$  via K-Street and In-House registrants in year  $y$ . The overall weighted out-degree of a client  $c$  is then

$$\text{w-out-deg}(c, y) = \text{w-out-deg}_K(c, y) + \text{w-out-deg}_I(c, y). \quad (\text{S.23})$$

Similarly, we let the weighted in-degree of a K-Street or In-House registrant  $r$  be

$$\text{w-in-deg}(r, y) = \sum_{(c,r) \in E(\mathcal{G}_y)} w[(c, r)], \quad (\text{S.24})$$

which represents the total *monetary flow* of the K-Street or In-House registrant  $r$  in year  $y$ .

The CCDFs of the weighted degrees  $\text{w-out-deg}_K(c, y)$ ,  $\text{w-out-deg}_I(c, y)$ , and  $\text{w-in-deg}(r, y)$  are shown in Fig. S12. The weighted in/out-degree distributions also follow heavy-tailed distributions, thereby corroborating the existence of 'hyper-influential' clients and registrants (not only in terms of the number of connections but also in terms of money flow). Comparing the client lobbying expenses via K-Street vs. In-House registrants (Fig. S12A–B), it is evident that lobbying expenses via In-House registrants have higher monetary amounts while there are almost 10 times more clients who lobby via K-Street registrants. We also observe interesting temporal trends, e.g., clients with the highest lobbying expenses via K-Street registrants have reduced their spending.

### D. Degree correlation

For the lobbying network's inner layers (K-Street registrants and lobbyists), we investigate the correlation between the in- and out-degrees of nodes of the same layer. We compute 2D histograms of K-Street registrant/lobbyist counts such that one datapoint corresponds to one registrant/lobbyist per year. Figure S13A shows that K-Street registrants with more clients tend to be connected to more lobbyists, supporting the notion of well-connected influential K-Street registrants. In contrast, Fig. S13B shows that the lobbyists with a very large number of government and legislator associations (ca. more than 10) work for only one registrant.

## V. Concentration analysis

One corollary of the hierarchical organization of a heavy-tail degree distribution is the Pareto  $X$ - $Y$  (e.g., 80-20) rule. We can illustrate this property with a so-called *Pareto chart*, or quantify it with a *Gini index*.

### A. Pareto charts

We will now introduce the Pareto chart (Lorentz curve) construction by using the distribution of reports among clients in a given year as an example. First, we sort all the clients  $c$  by the number of filed reports  $n(c)$  (from largest to smallest):  $n_1, \dots, n_N$ . The Pareto chart is then obtained by plotting the normalized ranking score  $R(i) = i/N$  on the x-axis, against the cumulative number of reports

$$C(i) = \frac{\sum_{j=1}^i n_j}{\sum_{j=1}^N n_j} \quad (\text{S.25})$$

on the y-axis. This specific plot is shown in Fig. S14A.

More relevant to the problem of disparity is Figure S14B, shows the Pareto chart based on the total lobbying budget (monetary value of the associated reports). It confirms the hierarchical organization of the first layer of the lobbying network, and displays almost exact the archetypal 80-20 rule, where the 80% of lobbying expenses are due to 20% of

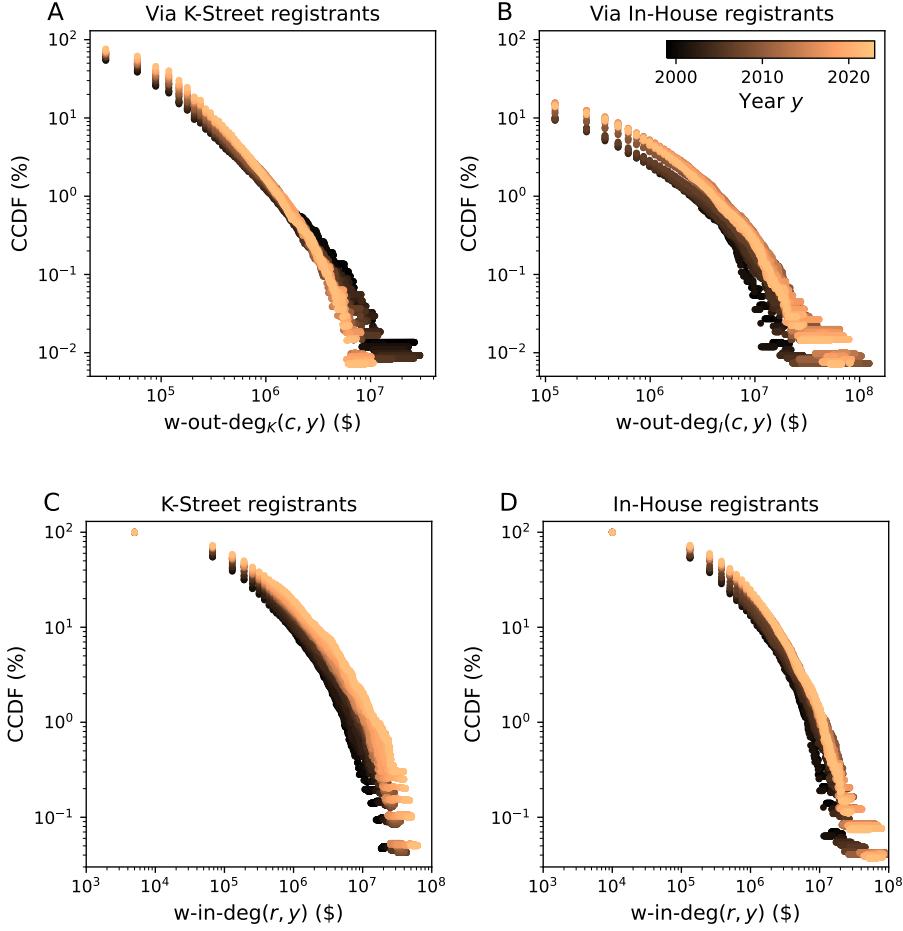


Figure S12. Complementary cumulative distributions (CCDFs) of weighted node in/out-degrees. (A) CCDF of the weighted client out-degree  $w\text{-out-deg}_K(c, y)$ : % of clients with lobbying expenses via K-Street registrants of  $w\text{-out-deg}_K(c, y)$  or more. (B) CCDF of the weighted client out-degree  $w\text{-out-deg}_I(c, y)$ : % of clients with lobbying expenses via In-House registrants of  $w\text{-out-deg}_I(c, y)$  or more. (C) CCDF of the weighted K-Street registrant in-degree  $w\text{-in-deg}(r, y)$ : % of K-Street registrants with monetary flow of  $w\text{-in-deg}(r, y)$  or more. (D) CCDF of the weighted In-House registrant in-degree  $w\text{-in-deg}(r, y)$ : % of In-House registrants with monetary flow of  $w\text{-in-deg}(r, y)$  or more. All CCDFs are computed separately for each year in the range 1999–2023.

clients. In Fig. S15, we perform an analogous analysis for registrants, based on the number of reports they file, their monetary flow, and their number of clients. In all cases, we find a variant of the Pareto rule.

In Fig. S16, we contrast the hierarchical organization of registrants, with a random organization of lobbyists. While 20% of K-Street registrants are responsible for more than 70% of the lobbyist contracts (registrant-lobbyist connection in the network), the connections are distributed among the lobbyists in an egalitarian fashion.

## B. Gini index

The shape of the Pareto chart is often characterized with a *Gini index*, which can be computed by looking at the area under the Pareto curve. For the example of Fig. S16A introduced in the previous section,

$$\text{Gini} = \frac{2}{N} \sum_{i=1}^N C(i) - R(i). \quad (\text{S.26})$$

For a uniform distribution of reports per clients,  $C(i) = R(i)$  so the Gini index would be zero. As the disparity between large and small clients increases, the Gini index approaches 1.



Figure S13. Correlation between in- and out-degrees of same-layer nodes for the lobbying network's inner layers. (A) Number of K-Street registrants in terms of their in- and out-degrees. (B) Number of lobbyists in terms of their in- and out-degrees. The 2D histograms are computed such that one datapoint represents one K-Street registrant/lobbyist per year, spanning the years 1999–2023.

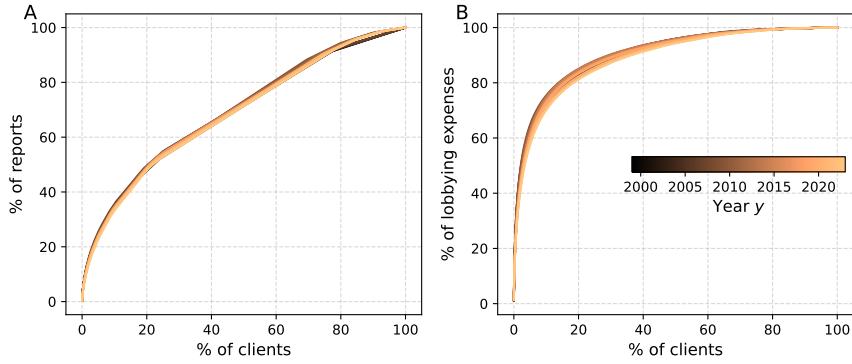


Figure S14. Pareto charts for clients in 1999–2023. (A) The number of reports filed by clients. The top quartile of clients files around half of the LDA reports. (B) The distribution of the lobbying expenses among clients follows the 80-20 rule.

In Fig. 4A of the main text, we focus on the monetary Gini indices quantifying the distribution of the lobbying budget. In Fig. S17, we show all the Pareto curves that were used to calculate the Gini indices in Fig. 4A. One important technical remark is that the lobbyist income is computed by dividing the monetary value of a given filing equally among all the lobbyists mentioned therein.

It is also interesting to compute issue-specific Gini indices (for clients and K-Street registrants) are computed by using only the filings  $f$  mentioning given general issue area  $a$  ( $a \in A(f)$ ). Fig. S20 shows K-Street registrant Pareto charts for three selected issue areas with different client group size. Generally, we find that the larger the issue, the steeper the Pareto chart, and so the larger the registrant Gini index (Fig. S18). The size of the issue can be measured either by the client group size or the total budget, which are correlated (Fig. S19).

The Gini Index and the lobbying network size are correlated also for randomly generated subnetworks (dashed line and shading in Fig. S18). For the Registrant Gini Index (Fig. S18B) we choose a random year and a random set of  $n$  clients connected to K-Street registrants. We then look at all their filings through K-Street registrants and we use them to construct a partial budget distribution among K-Street registrants. This procedure is repeated 100 times for each  $n \in \{10, 20, 50, 100, 200, 300, 400, 500, 600, 1000, 1200, 2000, 3000, 6000\}$ , and each time we use the partial distribution to compute the K-Street Registrant Gini Index. The dashed line and the shade in Fig. S18B reports the mean and standard deviation of their values, respectively. In Fig. S18A, we use an analogous procedure to compute the Client Gini Index in a randomized sample. This time we do not restrict our sample to K-Street Clients, but we use the same values of  $n$ . The relationship between the size of the subnetwork and the Gini Index underscores the importance of working with a complete dataset. Indeed, subsampling could introduce systematic biases.

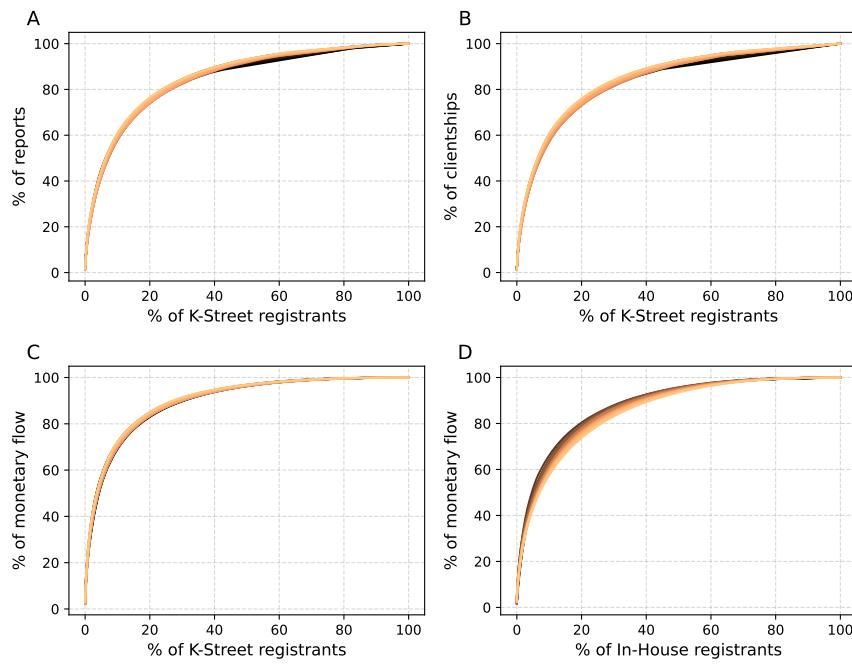


Figure S15. Pareto charts for registrants in 1999–2023. (A) The number of reports per K-Street registrant. (B) The number of clientships per K-Street registrant. (C) Monetary flow per K-Street registrant. (D) Monetary flow per In-House registrant.

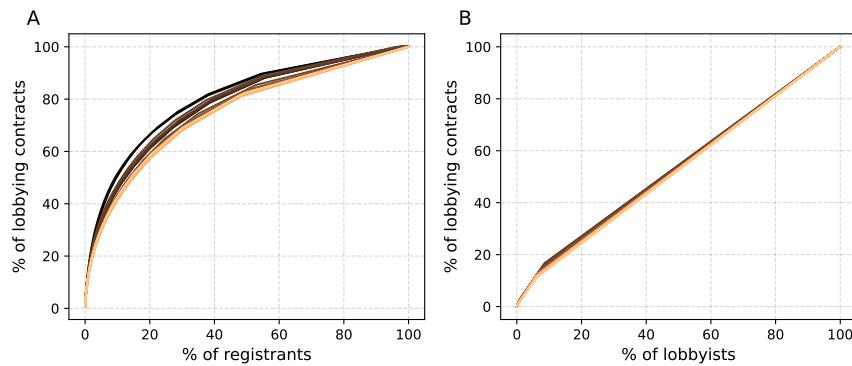


Figure S16. Pareto charts for registrant-lobbyist connections (lobbying contracts) in 1999–2023. (A) The distribution of lobbying contracts among registrants follows the Pareto rule (hierarchy). (B) The number of lobbying contracts among the lobbyists does not follow the Pareto rule (egalitarianism).

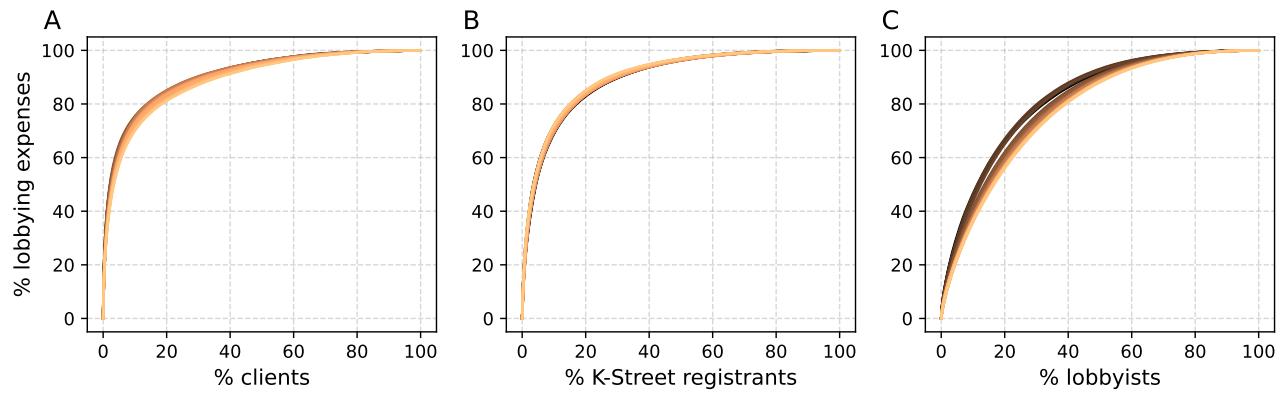


Figure S17. Pareto charts for the budget distribution in 1999–2023. (A) The Pareto chart of the distribution of the total lobbying budget (monetary value of the filings) among the clients. (B) The Pareto chart of the distribution of the total lobbying budget among K-Street registrants. (C) The Pareto chart of the distribution of the total lobbying budget among the lobbyists.

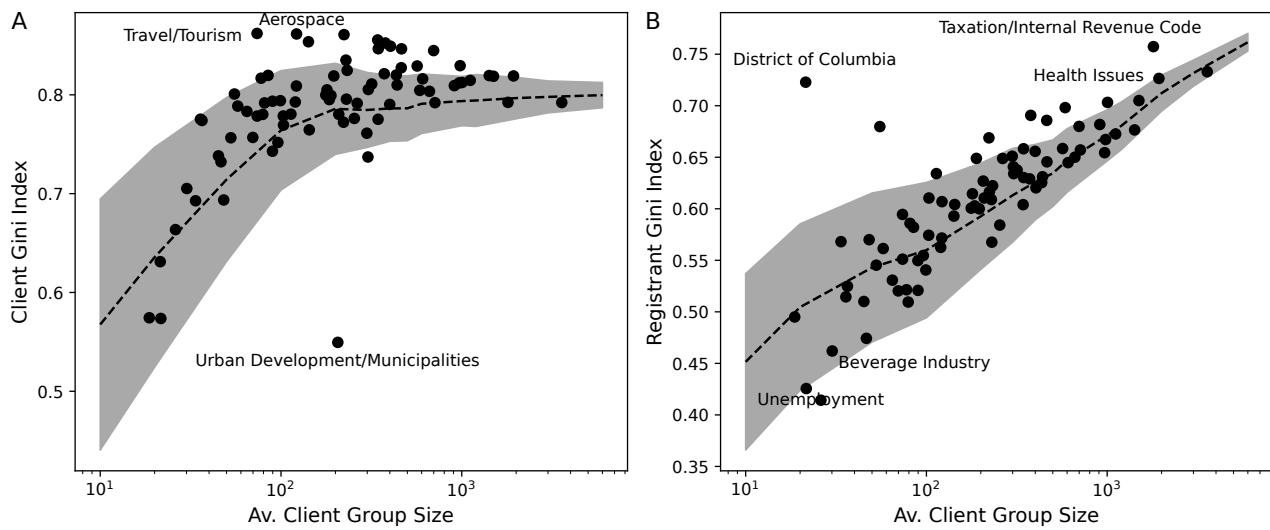


Figure S18. Concentration in area-specific subnetworks. (A) Client Gini Index. (B) Registrant Gini Index. The dashed line and the shaded region corresponds to the randomized network with a given number of clients.

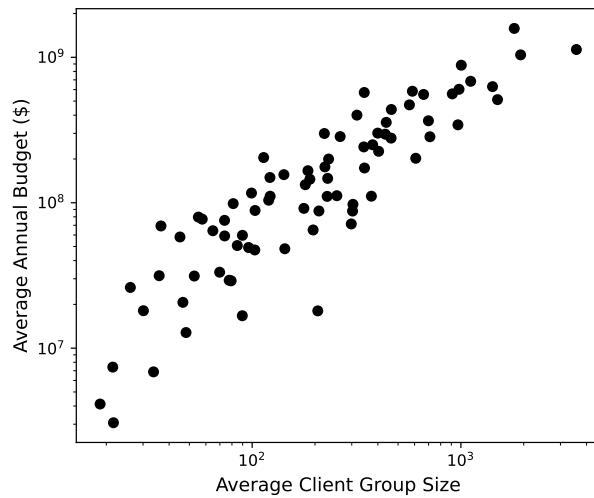


Figure S19. Average value of the client group size interested in one of the standard issue areas in years 1999–2023 plotted against the average annual budget in this area.

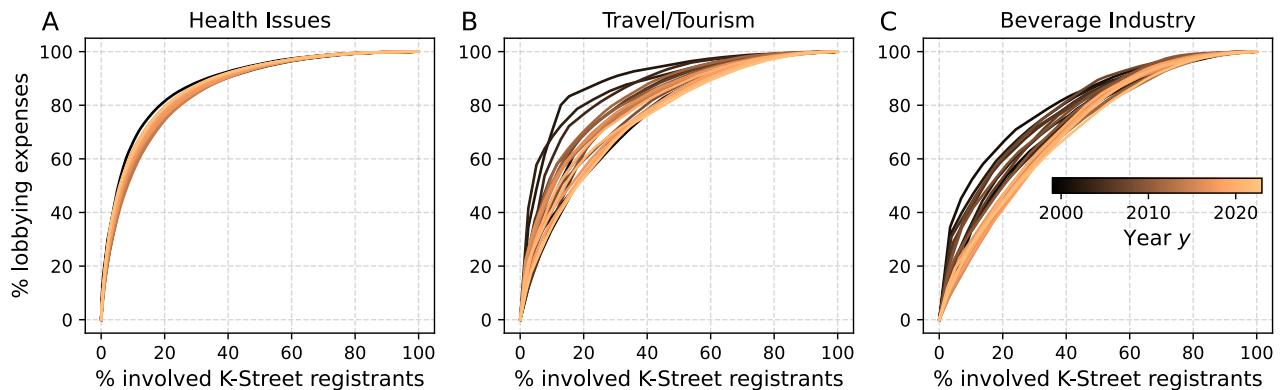


Figure S20. Example Pareto charts for the area-specific budget distribution among K-Street registrants in 1999–2023. (A) Health is an example of a popular (high budget) issue area. The fact that the Pareto curve is steep indicates high degree of concentration. (B) Travel/Tourism is an example of an average-sized issue area. Note that in response to the 9/11 crisis, the Pareto curve got steeper. (C) Beverage Industry is an example of a small issue area with a handful of interested clients. In such categories, the budget distributions tend to be most egalitarian.

## VI. Higher order interactions

So far we have analyzed the ties between clients and registrants, as well as the ties between registrants and lobbyists. We can also more directly ask about the connections between clients and lobbyists by introducing the notion of *client-registrant-lobbyist triads*

$$\text{Lobbying triads}(y) = \{(c, r, l) : \exists f : y = y(f) \wedge c = c(f) \wedge r = r(f) \wedge l \in L(f)\}. \quad (\text{S.27})$$

We will now use the notion of lobbying triads to answer questions about the distribution of jobs in a given lobbying firm. Specifically, we can analyze the number  $n(r, l)$  of clients of registrant  $r$  that lobbyist  $l$  worked for, i.e., all the lobbying triads of the form  $(*, l, r)$ . By construction,  $n(r, l)$  cannot be larger than the number of clients of  $r$  (i.e.,  $\text{in-deg}(r)$ ). In Fig. S21A, we show that the number of clients a given lobbyist works for almost never exceeds 100, even if the number of clients of a given registrant is much larger. For most registrant-lobbyist pairs, the number of clients per lobbyist  $n(r, l)$  decays exponentially (Fig. S21B). Nevertheless, we also find examples of lobbyists whose names appear on virtually all the reports of a given registrant, up to 300 clients per lobbyist (Fig. S21A). On the one hand, this finding corroborates the general picture that the distribution of lobbying cases per individual lobbyist is more egalitarian than the distribution of lobbying cases per lobbying firms that have the capacity to accumulate advantage. On the other hand, we also find a small number of *super-lobbyists*, who appear to be an exception to this rule. Further analysis of the profiles of individual lobbyists (generalists vs. specialists, etc.) can be the subject of an interesting future study.

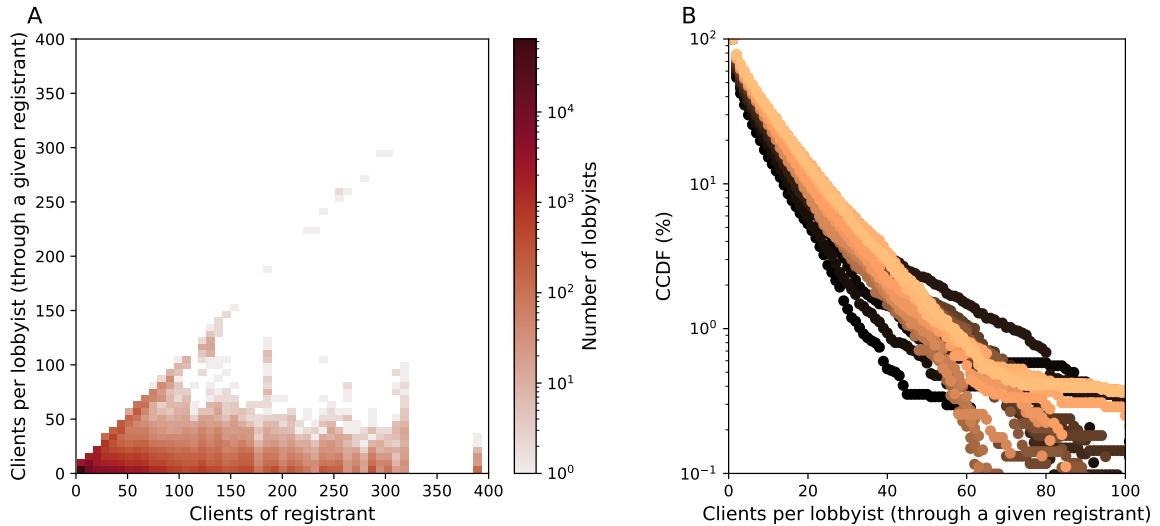


Figure S21. Assignment of clients of a given registrant to their lobbyists. (A) The distribution of the number of clients of registrant  $r$  that lobbyist  $l$  worked for, as a function of the total number clients of a given registrant. (B) The marginal distribution of the number of clients of registrant  $r$  that lobbyist  $l$  worked for, broken by year. Data from years 1999–2023.

## VII. Lobbying network visualization

To visualize the entire network graph for one year (as shown in Fig. 1A for 2017), we use a customized multilayer layout, with a particular embedding strategy for each set of nodes. We will now describe the placement of nodes in each layer separately, and then the way they are assembled together. The final visualization is facilitated with the *PyGraphistry* library and its cloud-based visualization tool [18].

### A. Node placement

The  $x$ -coordinate of a client node  $c$  represents its lobbying expenses in year  $y$  and is proportional to  $-\log [\text{w-out-deg}(c, y)]$ , where  $\text{w-out-deg}(c, y)$  has been defined in eq. (S.23). The  $y$ -coordinate of node  $c$  is proportional to the client's incli-

nation  $i_C(c, y)$  to lobby through In-House registrants such that

$$i_C(c, y) = \frac{\text{w-out-deg}_I(c, y) - \text{w-out-deg}_K(c, y)}{\text{w-out-deg}(c, y)}.$$

The  $x$ -coordinate of a registrant node  $r$  represents its monetary flow in year  $y$  and is proportional to  $-\log[\text{w-in-deg}(r, y)]$ . The  $y$ -coordinate of node  $r$  represents its out-degree and is proportional to  $\chi(r) \log[\text{out-deg}(r, y)]$  where

$$\chi(r) = \begin{cases} 1 & \text{if } r \in \text{In-House} \\ -1 & \text{if } r \in \text{K-Street.} \end{cases}$$

The  $x$ -coordinate of a lobbyist node  $l$  represents its in-degree in year  $y$  and is proportional to  $-\text{in-deg}(l, y)$ . The  $y$ -coordinate of node  $l$  is proportional to the lobbyist's inclination  $i_L(l, y)$  to lobby through In-House registrants such that

$$i_L(l, y) = \frac{\sum_{(r, l) \in E(\mathcal{G}_y)} \chi(r)}{\text{in-deg}(l, y)}.$$

The  $x$ -coordinate of a government entity node  $g$  represents its monetary flow  $\mu(g, y)$  in year  $y$ . Specifically, it is a piece-wise linear, order-preserving transformation of  $-\log[\mu(g, y)]$ , where

$$\mu(g, y) = \sum_f \mathbb{1}[y(f) = y \wedge g \in G(f)] m(f).$$

The  $y$ -coordinate of node  $g$  reflects its similarity (measured by co-mention count in filings) to other government entity nodes. In particular, we construct virtual weighted edges between government entity nodes such that the edge weight is set to the normalized government entity co-mention count in filings and the edges with a small co-mention count are removed. We then run the Fruchterman-Reingold force-directed algorithm [19] on the subgraph consisting of government entity nodes and their internal edges and use the obtained close-to-equilibrium node  $x$ -coordinates to set the  $y$ -coordinates of the government entity nodes. The two largest government entity nodes (Senate and House of Representatives) are positioned manually to align them with the corresponding division in the last (legislators) layer.

The legislator nodes are arranged on semicircles with the radii highlighting the difference in size between the Senate and House of Representatives. On each semicircle, the legislator nodes are spaced/distributed evenly, grouped by party affiliation (such that Democrats have higher  $y$ -coordinates than Republicans), and ordered randomly within each group. In 2017, our data contained also one Independent Senator, but we neglected the corresponding node in the visualization.

Except for legislator nodes, we perturb all node positions with white noise for better node separation in graph visualization. The differences between node placements without and with the white noise perturbation for all affected types of nodes are shown in Fig. S22–S23. As we can see, the white noise perturbation reduces the overlapping of nodes, while minimally disturbing/violating the above-described node placement based on network graph measures.

## B. Node size and color

The client node size is proportional to the client's lobbying expenses  $\text{w-out-deg}(c, y)$ . The registrant node size is proportional to the registrant's monetary flow  $\text{w-in-deg}(r, y)$ . The size of the government entity node  $g$  represents its total mention count  $n(g, y)$  in filings in year  $y$  and is proportional to  $\log[n(g, y)]$ , where

$$n(g, y) = \sum_f \mathbb{1}[y(f) = y \wedge g \in G(f)].$$

The lobbyist and legislator node sizes are fixed. Note that the above-described node sizes are passed to *PyGraphistry*, where a further order-preserving transformation is applied [18].

The color-coding of nodes is summarized in Tab. S2. The color of client and K-Street registrant nodes ranges between the specified minimum and maximum values in proportion to the corresponding scaling law. Specifically, the client node color represents the client's lobbying expenses and the K-Street registrant node color reflects the registrant's monetary flow. The colors of other node types are all fixed.

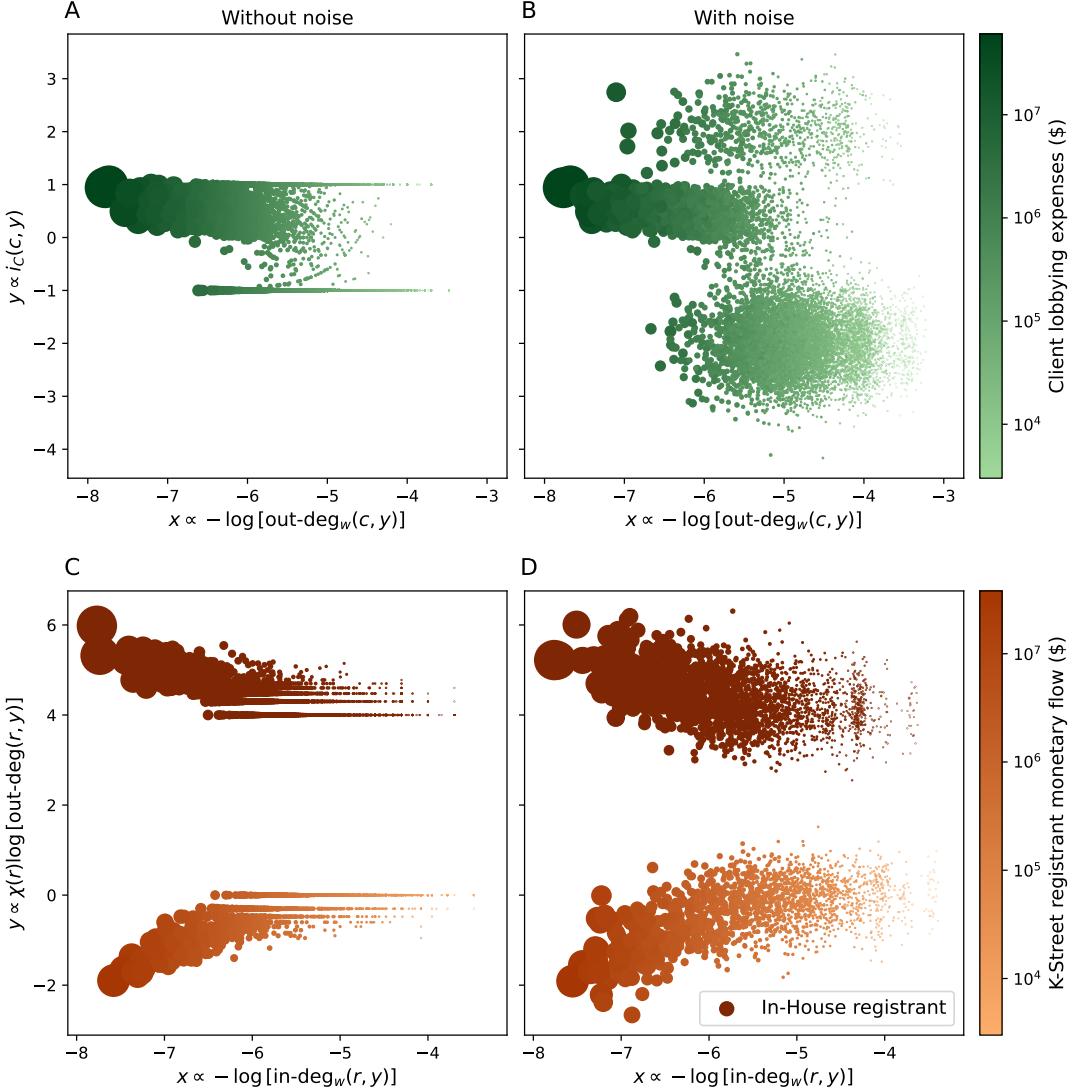


Figure S22. Node placement based on network graph measures without (left) and with (right) the white noise perturbation. Used to construct the whole lobbying network graph for the year 2017 in Fig. 1A. (A, B) Clients. (C, D) Registrants.

### C. Final assembly

We embed the above-described sets of nodes in a common 2D space using translation and scaling operations as needed for a reasonable layered network arrangement. The resulting  $(x, y)$ -coordinates, along with the color and size of all nodes constitute the final *node data*. We then construct unweighted edges as described in Sec. II B such that the obtained list of (source node, destination node) pairs forms *edge data*. We also specify the following *global properties*: node opacity = 100% (default), edge opacity = 5%, edge color interpolates the colors of the connecting nodes along the edge (default), edge size/width = 50 (default), and edge curvature = 20% (default).

The *node data*, *edge data*, and *global properties* are fed to the *PyGraphistry* library to render the lobbying network visualization in the browser using the cloud-based visualization tool [18].

## VIII. Network evolution

In this Section, we describe the technical aspects of our network evolution analysis.

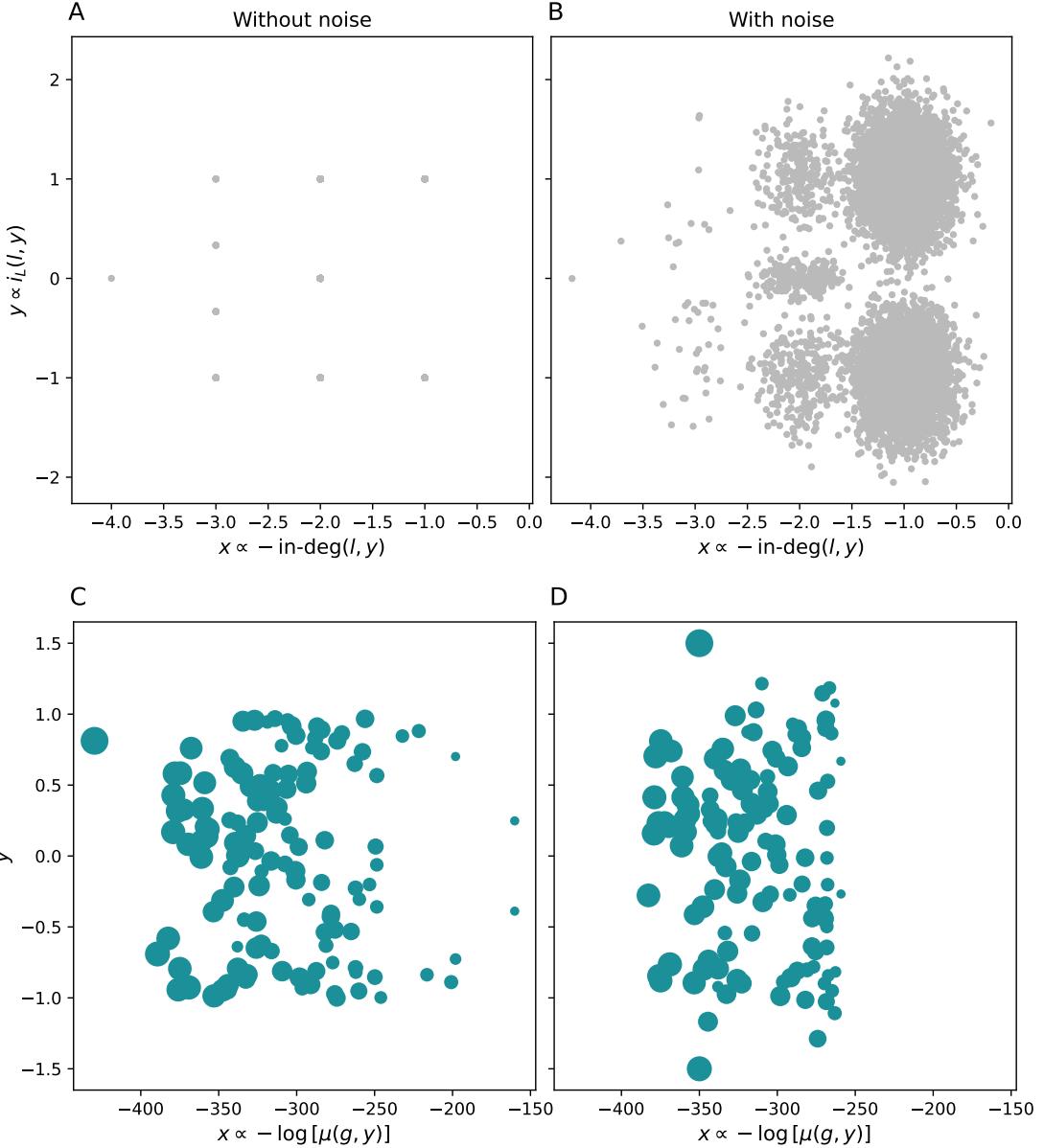


Figure S23. Node placement based on network graph measures without (left) and with (right) the white noise perturbation. Used to construct the whole lobbying network graph for the year 2017 in Fig. 1A. (A, B) Lobbyists. (C, D) Government entities: Panel (C) does not include the piece-wise linear, order-preserving transformation of  $x$ -coordinates nor the manual re-positioning of the two largest nodes (Senate and House of Representatives). For  $y$ -coordinate computation see Sec. VII A.

### A. Rejuvenation

By comparing the nodes between the lobbying networks  $\mathcal{G}_{y_1}$  and  $\mathcal{G}_{y_2}$  in two different years  $y_1$  and  $y_2$ , we can quantify the degree to which the lobbying network refreshes and rejuvenates.

In Fig. S24A-C, we estimate the empirical (frequentist) probability of a node remaining in the network after  $y$  years. For clients (Fig. S24A), this quantity is defined as

$$\mathbb{P} [\text{Client lobbying in year } y_0 \text{ remains in the network after } y \text{ years}] = \frac{|\text{Clients}(y_0) \cap \text{Clients}(y_0 + y)|}{|\text{Clients}(y_0)|}, \quad (\text{S.28})$$

and the definition is analogous for K-Street registrants (Fig. S24B, and Fig. 2 of the main text) and lobbyists (Fig. S24C). The different curves in this figure correspond to different starting years  $y_0 \in \{1999, 2000, \dots, 2022\}$ . For

Table S2. Lobbying network node colors. Used in Fig. 1A and Fig. 2B of the main text.

Node type	HEX color	Min	Max	Scaling law
Client	#a1d99b	#00441b		$\log [w\text{-out-deg}(c, y)]$
K-Street registrant	#fd9e6b	#a63603		$\log [w\text{-in-deg}(r, y)]$
In-House registrant	#7f2704			—
Lobbyist	#bababa			—
Government entity	#1c9099			—
Democratic legislator	#0015bc			—
Republican legislator	#ff0000			—

clients, these curves do not collapse, most likely due to economical interruptions. For registrants and lobbyists, the decay of operation probability to first approximation follows a universal law, but we also observe a steady increase in longevity in recent years.

The probability of leaving the lobbying network is not the same for all the clients/registrants/lobbyists, but it depends on individual characteristics. One particularly interesting predictor of operation ceasing is age, defined as the maximum number of consecutive years for which node  $x$  was present in the network prior to year  $y$ ,

$$\text{Age}(x, y) = \sum_{y' < y} \mathbb{1} [\forall_{y' \leq y'' \leq y} : x \in V(\mathcal{G}_{y''})]. \quad (\text{S.29})$$

Figure S24D shows that the conditional probability that a client  $c$  of a given age ceases lobbying in year  $y_0$  (i.e.,  $y_0$  is the last year the client is observed in the lobbying network) conditioned on its age  $y$

$$\mathbb{P}[c \text{ ceases lobbying in year } y_0 | \text{Age}(c, y_0) = y] = \frac{\sum_{c \in V(\mathcal{G}_{y_0})} \mathbb{1} [\text{Age}(c, y_0) = y \wedge c \notin V(\mathcal{G}_{y_0+1})]}{\sum_{c \in V(\mathcal{G}_{y_0})} \mathbb{1} [\text{Age}(c, y_0) = y]} \quad (\text{S.30})$$

is a decaying function of age. The same is true for an analogous quantity for K-Street registrants (Fig. S24E), and lobbyists (Fig. S24F). In other words, long-operating lobbying actors are less likely to leave the lobbying network. This trend over time may lead to ‘aging’ and accumulation of expertise. Interestingly, the *aging curve* has a similar shape for clients, K-street registrants, and lobbyists, with a characteristic timescale of approximately 10 years.

## B. Preferential attachment/detachment inference

Following the works of Barabasi et al. [15], we analyze the dynamics of registrant connections through the lens of preferential attachment. In the main text, we present the results of this analysis for clientships (client-registrant connections) in Fig. 2E-F. The specific quantity that we look at is defined as follows. First, we identify K-Street registrants with degree  $d$  in year  $y - 1$ :

$$\text{K-Street Registrants}_{\text{in}=d}(y - 1) = \{r \in \text{K-Street Registrants}(y - 1) : \text{in-deg}(r, y - 1) = d\}. \quad (\text{S.31})$$

Then, we find all the new clientships that involve an existing registrant of in-degree  $d$

$$\text{Attached clientships}(y; d) = \{(c, r) \in E(\mathcal{G}_y) \setminus E(\mathcal{G}_{y-1}) : r \in \text{K-Street} \wedge r \in V(\mathcal{G}_{y-1}) \cap V(\mathcal{G}_y) \wedge \text{in-deg}(r, y - 1) = d\} \quad (\text{S.32})$$

and removed clientships

$$\text{Detached clientships}(y; d) = \{(c, r) \in E(\mathcal{G}_{y-1}) \setminus E(\mathcal{G}_y) : r \in \text{K-Street} \wedge r \in V(\mathcal{G}_{y-1}) \cap V(\mathcal{G}_y) \wedge \text{in-deg}(r, y - 1) = d\} \quad (\text{S.33})$$

The absolute probability that one of the attached clientships involves an existing registrant of in-degree  $d$  is simply the ratio

$$p^+(d, y) = \frac{|\text{Attached clientships}(y; d)|}{\sum_{d'} |\text{Attached clientships}(y; d')|}. \quad (\text{S.34})$$

Analogously, the absolute probability that one of the detached clientships involves an existing registrant of in-degree  $d$  is

$$p^-(d, y) = \frac{|\text{Detached clientships}(y; d)|}{\sum_{d'} |\text{Detached clientships}(y; d')|}. \quad (\text{S.35})$$

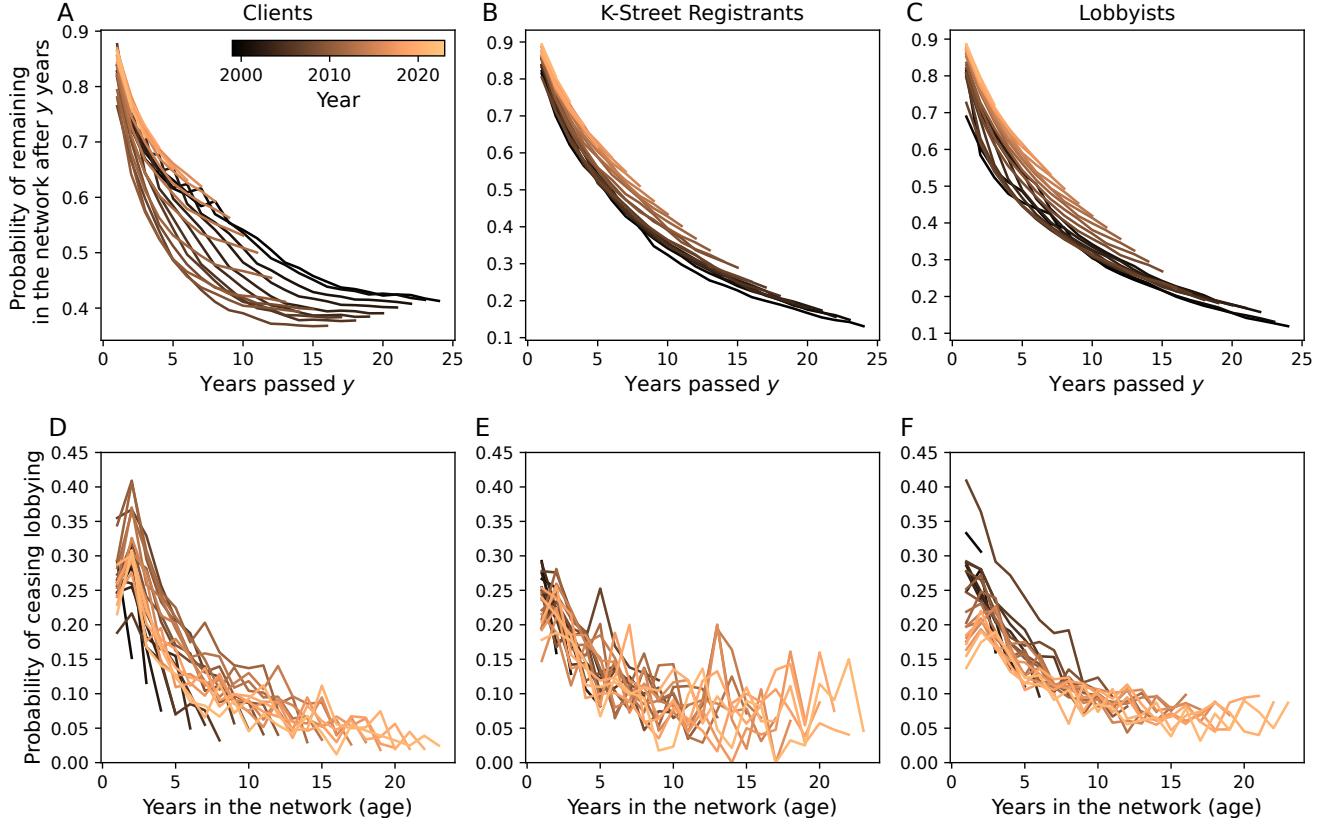


Figure S24. Node removal probability. (A) Client retention probability. (B) K-Street registrant retention probability. (C) Lobbyist retention probability. (D) Probability of client ceasing lobbying conditioned on its age. (E) Probability of K-Street registrant ceasing lobbying conditioned on its age. (F) Probability of lobbyist ceasing lobbying conditioned on its age.

Of more interest is the *attachment function*

$$\Pi^+(d, y) = \left( \frac{|\text{Attached clientships}(y; d)|}{|\text{K-Street Registrants}_{\text{in}=d}(y-1)|} \right) \left( \sum_{d'} \frac{|\text{Attached clientships}(y; d')|}{|\text{K-Street Registrants}_{\text{in}=d'}(y-1)|} \right)^{-1}, \quad (\text{S.36})$$

and the *detachment function*

$$\Pi^-(d, y) = \left( \frac{|\text{Detached clientships}(y; d)|}{|\text{K-Street Registrants}_{\text{in}=d}(y-1)|} \right) \left( \sum_{d'} \frac{|\text{Detached clientships}(y; d')|}{|\text{K-Street Registrants}_{\text{in}=d'}(y-1)|} \right)^{-1}, \quad (\text{S.37})$$

which capture the preferential tendencies in network development. Figure S25A-B shows a direct evaluation of them for our lobbying network. As the attachment function may be noisy, in Fig. S25C-D we show the equivalent, but more robust to estimate, *cumulative attachment/detachment functions* [10]

$$\pi^\pm(d, y) = \sum_{k=1}^d \Pi^\pm(k, y). \quad (\text{S.38})$$

Critically, functions  $\Pi^+(d, y)$  are independent of  $y$  and linear in  $d$  (equivalently,  $\pi^+(d, y) \sim d^2$ ), i.e., the clientship dynamics satisfies linear preferential attachment [15]. Interestingly, the detachment function  $\Pi^-(d, y)$  is also linear in  $d$ , so we also find linear preferential detachment.

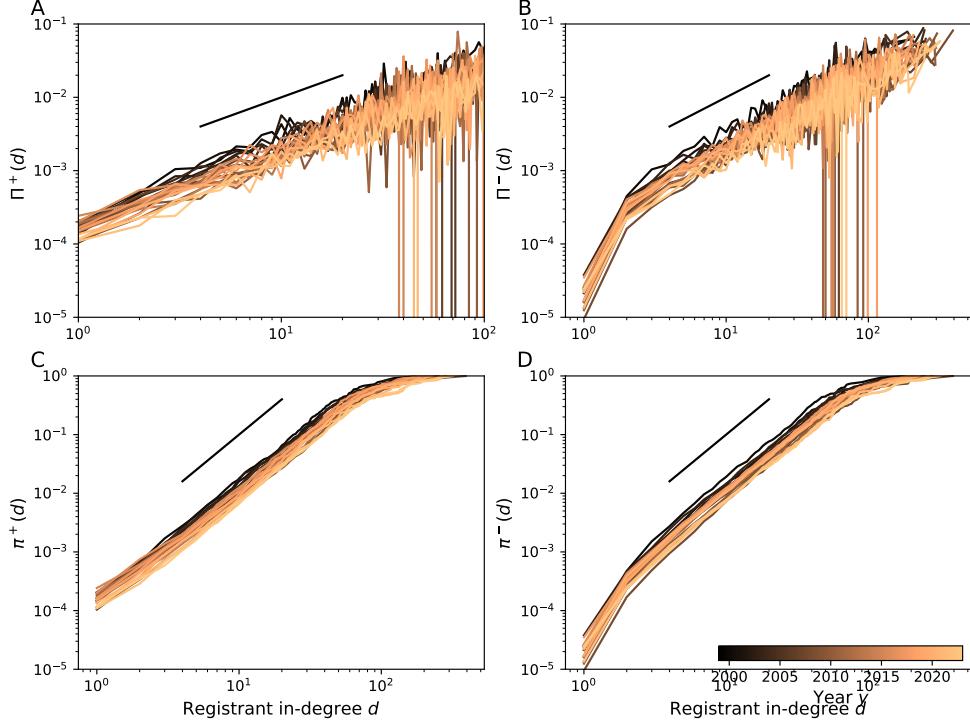


Figure S25. Clientship dynamics in terms of K-Street registrant in-degree  $\text{in-deg}(r, y)$ . (A) Preferential attachment function  $\Pi^+(d, y)$ . (B) Preferential detachment function  $\Pi^-(d, y)$ . (C) Cumulative preferential attachment function  $\pi^+(d, y)$ . (D) Cumulative preferential attachment function  $\pi^-(d, y)$ . For definitions of attachment functions see eq. (S.36)–(S.38). All attachment functions are computed separately for each year in the range 1999–2023.

### C. Registrant out-degree dynamics

It is interesting to ask whether the preferential attachment is also present in the lobbying contract dynamics of registrant-lobbyist connections. In Figures S26 and S27, which are analogous to Fig. 2 of the main text, we analyze the evolution dynamics of the lobbying contracts involving K-Street and In-House registrants, respectively.

Panels A and C of Fig. S26 are identical to the respective panels of Fig. 2. Figure S26C, however, instead of focusing on registrants' in-coming edges (clientships), reports the number of their out-going edges (lobbying contracts). Figure S26D,E presents the preferential attachment/detachment analysis, analogous to the one outlined in Sec. VIIIB. Specifically, we define

$$\text{K-Street Registrants}_{\text{out}=d}(y-1) = \{r \in \text{K-Street Registrants}(y-1) : \text{out-deg}(r, y-1) = d\}, \quad (\text{S.39})$$

$$\text{Attached lobbying contracts}(y; d) = \{(r, l) \in E(\mathcal{G}_y) \setminus E(\mathcal{G}_{y-1}) : r \in \text{K-Street} \cap V(\mathcal{G}_{y-1}) \cap V(\mathcal{G}_y) \wedge \text{out-deg}(r, y-1) = d\}, \quad (\text{S.40})$$

$$\text{Detached lobbying contracts}(y; d) = \{(r, l) \in E(\mathcal{G}_{y-1}) \setminus E(\mathcal{G}_y) : r \in \text{K-Street} \cap r \in V(\mathcal{G}_{y-1}) \cap V(\mathcal{G}_y) \wedge \text{out-deg}(r, y-1) = d\}, \quad (\text{S.41})$$

and compute the cumulative preferential attachment/detachment function analogous to eq. (S.38), except this time

$$\Pi^+(d, y) = \left( \frac{|\text{Attached lobbying contracts}(y; d)|}{|\text{K-Street Registrants}_{\text{out}=d}(y-1)|} \right) \left( \sum_{d'} \frac{|\text{Attached lobbying contracts}(y; d')|}{|\text{K-Street Registrants}_{\text{out}=d'}(y-1)|} \right)^{-1}, \quad (\text{S.42})$$

and

$$\Pi^-(d, y) = \left( \frac{|\text{Detached lobbying contracts}(y; d)|}{|\text{K-Street Registrants}_{\text{out}=d}(y-1)|} \right) \left( \sum_{d'} \frac{|\text{Detached lobbying contracts}(y; d')|}{|\text{K-Street Registrants}_{\text{out}=d'}(y-1)|} \right)^{-1}. \quad (\text{S.43})$$

As  $\pi^-(l) \sim l^2$ , we conclude that similarly to the result of Sec. VIII B, the registrant-lobbyist connections are removed according to linear preferential detachment. Nevertheless, this time the preferential attachment is sublinear, and the preference is more pronounced for larger (higher out-degree) registrants.

Figure S27 presents analogous evolution analysis for the registrant-lobbyist connections, where the registrant is In-House. The preferential attachment/detachment (approximately linear) is still at play, but we find a different trend in the lobbying probability as a function of time (Fig. S27C). It appears that prior to the financial crisis of 2007–08, In-House registrants were much more stable, and nowadays they follow a similar drop-out dynamics to K-Street registrants.

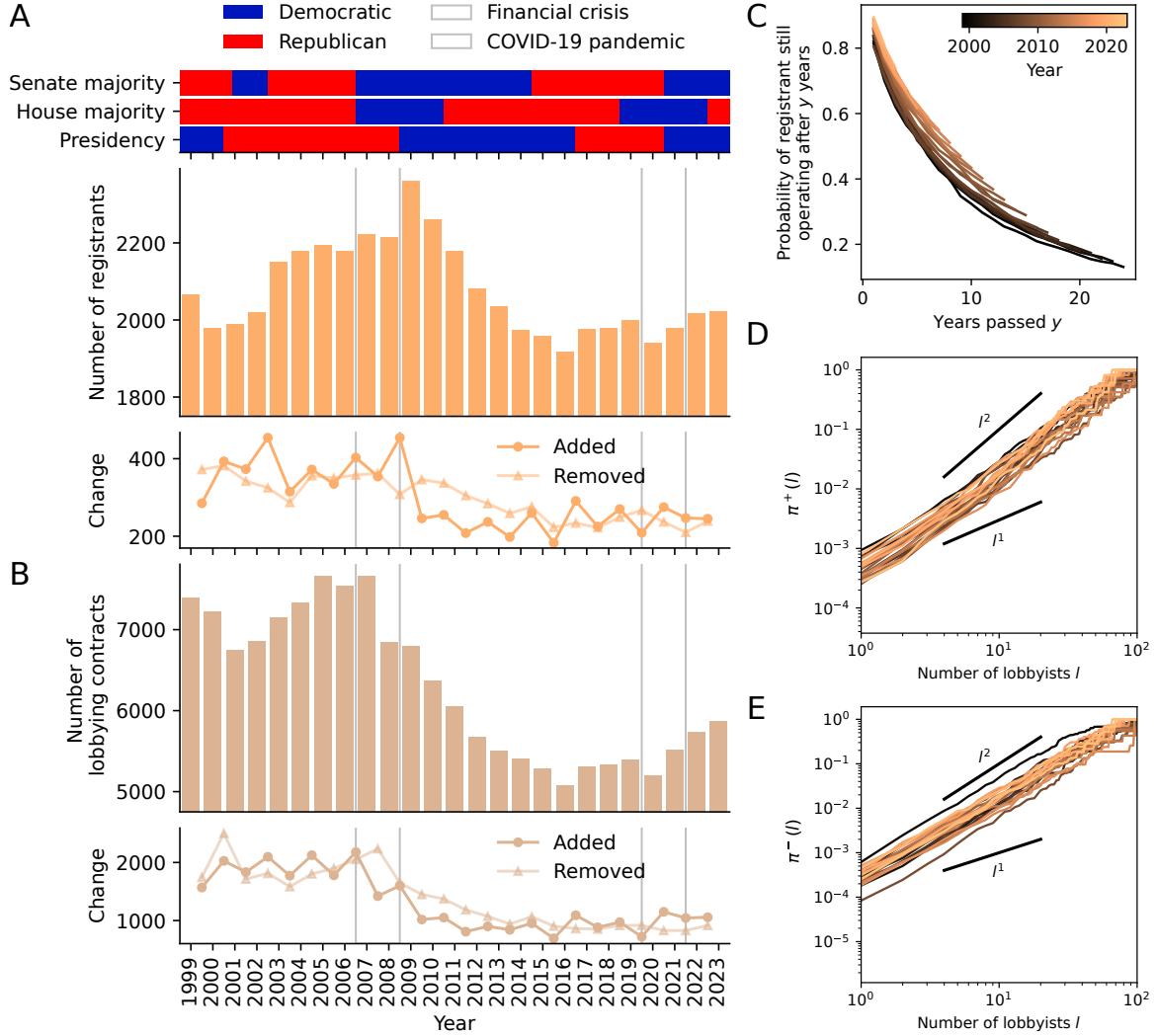


Figure S26. Lobbying contract dynamics for K-Street registrants. (A) Number of active K-Street registrants per year (bars) and their yearly increments (circles) and decrements (triangles). The top horizontal bars show the party (Democratic in blue and Republican in red) of Senate majority, House majority, and President by year. The financial crisis (2007–2008) and COVID-19 pandemic (2019–2021) are shown by gray overlays. (B) Number of active lobbying contracts (register-lobbyist connections) per year (bars) and their yearly increments (circles) and decrements (triangles). (C) K-Street registrant ‘survival probability’. (D) Cumulative preferential attachment function  $\pi^+(l)$  in terms of K-Street registrant out-degree. (E) Cumulative preferential detachment function  $\pi^-(l)$ . Quantities in (C–E) are computed separately for each year in the range 1999–2023.

#### D. Small component example

In Fig. 2A of the main text, we show the attachment and detachment processes in a sample client-registrant-lobbyist network component. We have selected this component through the following principled search algorithm.

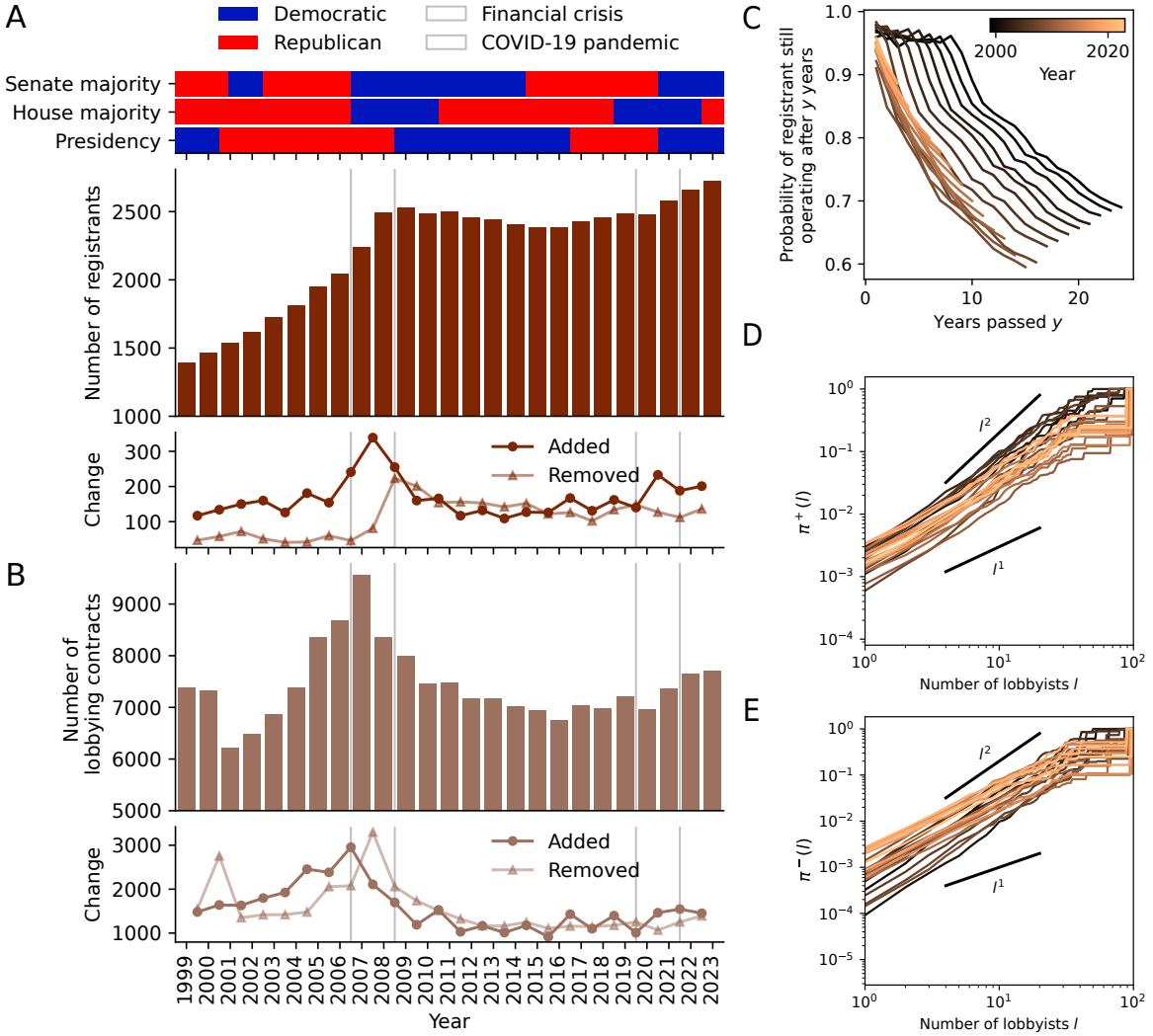


Figure S27. Lobbying contract dynamics for In-House registrants. (A) Number of active In-House registrants per year (bars) and their yearly increments (circles) and decrements (triangles). The top horizontal bars show the party (Democratic in blue and Republican in red) of Senate majority, House majority, and President by year. The financial crisis (2007–2008) and COVID-19 pandemic (2019–2021) are shown by gray overlays. (B) Number of active lobbying contracts (registrant-lobbyist connections) per year (bars) and their yearly increments (circles) and decrements (triangles). (C) In-House registrant ‘survival probability’ is generally larger than the ‘survival probability’ of K-Street registrants (Fig. S27C). We also see a qualitative shift around the financial crisis of 2007–08 to less stable In-House registrants (higher drop-out probability). (D) Cumulative preferential attachment function  $\pi^+(l)$  in terms of In-House registrant out-degree. (E) Cumulative preferential detachment function  $\pi^-(l)$ . Quantities in (C–E) are computed separately for each year in the range 1999–2023.

Given the graphs  $\mathcal{G}_y$  for  $y \in \{1999, 2000, \dots, 2023\}$  constructed according to Sec. II, we first remove all In-House registrant, government entity, and legislator nodes along with their edges, obtaining the subgraphs

$$\begin{aligned} \mathcal{G}_y^* = & (V(\mathcal{G}_y) \setminus \text{In-House Registrants}(y) \setminus \text{Government entities}(y) \setminus \text{Legislators}(y), \\ & E(\mathcal{G}_y) \setminus \text{In-House Clientships}(y) \setminus \text{In-House Lobbyist contracts}(y) \\ & \setminus \text{Government associations}(y) \setminus \text{Legislator associations}(y)). \end{aligned}$$

For technical reasons, we shall treat  $\mathcal{G}_y^*$  as undirected. For each year, we compute the set  $\mathcal{W}_y$  of connected components of  $\mathcal{G}_y^*$ . We denote the number of registrants in a component  $C \in \mathcal{W}_y$  by  $\nu(C, y) = |\{n \in V(C) : n \in \text{Registrants}(y)\}|$  and select only the components that have  $\nu(C, y) = 3$  registrant nodes, creating a set of *candidate components*  $\mathcal{C}_y = \{C \in \mathcal{W}_y : \nu(C, y) = 3\}$ . After sorting the components for a given year by the number of registrants  $\nu(C, y)$  in descending order, we find that for all years the largest  $\nu(C, y)$  is of the order of  $10^3$  and the second-largest  $\nu(C, y) \leq 4$  (c.f. Sec. IX B). We choose candidate components with  $\nu(C, y) = 3$  as they illustrate the attachment and detachment

processes better than those with  $\nu(C, y) = 4$  and still have a reasonable (not too small) size.

Next, we track the evolution of each component  $C \in \mathcal{C}_y$  two years forward and two years backward (for years  $\max(y - 2, 1999) \leq y' \leq \min(y + 2, 2023)$ ). In particular, first, new client and lobbyist nodes from  $\mathcal{G}_{y\pm 1}^*$  are added to the component via existing registrant nodes, and new registrant nodes from  $\mathcal{G}_{y\pm 1}^*$  are added to the component via existing client and lobbyist nodes. Crucially, the obtained set of nodes is then connected with the edges present  $\mathcal{G}_{y\pm 1}^*$  to remove any nodes no longer present in year  $y \pm 1$ .

This operation can be formalized by defining a *one-year forward/backward evolution map* for vertices

$$\epsilon^\pm(V(C), y) = [V(\mathcal{G}_{y\pm 1}^*) \cap V(C)] \cup N(V(\mathcal{G}_{y\pm 1}^*) \cap V(C))$$

where  $N(X, \mathcal{G})$  is the set of all the neighbors of nodes  $n \in X$  in graph  $\mathcal{G}$ , and the evolution map for edges

$$\epsilon^\pm(E(C), y) = E(\mathcal{G}_{y\pm 1}^*) \cap [(n_1, n_2) : n_1, n_2 \in \epsilon^\pm(V(C), y)].$$

Thus, for each  $C \in \mathcal{C}_y$ , we then compute a sequence of 5 graphs

$$\mathcal{E}(C, y) = \left( \epsilon^-(\epsilon^-(C, y), y-1), \epsilon^-(C, y), C, \epsilon^+(\epsilon^+(C, y), y+1) \right)$$

and we select one of them to illustrate the graph evolution with a real example.

## IX. Further analysis of bipartite subgraphs

In this section, we present further analysis of the lobbying network topology. Specifically, we look at various bipartite subgraphs of the network and compute their average clustering coefficient and the size of the largest connected component. The bipartite graphs that we consider are naturally constructed by restricting attention to two adjacent layers of the lobbying network  $\mathcal{G}_y$  and the edges between them (c.f. Sec. II). All of the bipartite subgraphs are listed in Table S3, where we make a distinction between the upstream set and the downstream set. As the lobbying network is directed, the induced subgroups are also directed. Nevertheless, for notational simplicity, in this section, we will treat the bipartite graphs as undirected. We also exclude In-House Registrants, as by definition their in-degree is fixed to 1.

Table S3. Bipartite subgraphs of our lobbying network, excluding In-House registrants.

Upstream Set	Downstream Set
Clients	K-Street Registrants
K-Street Registrants	Lobbyists
Lobbyists	Government entities
Lobbyists	Legislators

### A. Clustering coefficient

Let  $\mathcal{B}$  be a bipartite graph with the upstream node-set  $U$  and the downstream node-set  $D$ . Without loss of generality, we will now consider a node in the upstream set  $u \in U$ . We define the set of its first neighbors  $N(u) \subseteq D$  and the set of its second neighbors  $N(N(u)) \subseteq U$ . The *clustering coefficient* of node  $u$  is defined as

$$C(u) = \frac{1}{|N(N(u))|} \sum_{v \in N(N(u))} \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}, \quad (\text{S.44})$$

and by averaging

$$C(U) = \frac{1}{|U|} \sum_{u \in U} C(u), \quad (\text{S.45})$$

we can compute the *average clustering coefficient* for the upstream set  $U$  in graph  $\mathcal{B}$ . The clustering coefficient  $C(D)$  for the downstream set  $D$  is defined analogously.

In Fig. S28 we report the average clustering coefficient for both the upstream set and the downstream set of  $\mathcal{B}$ , where  $\mathcal{B}$  is one of the subgraphs (c.f. Tab. S3) of the lobbying network  $\mathcal{G}_y$  in year  $y$ . Two coefficients stand out due to their consistently high value:  $C(U)$  for the Client–K-Street Registrant graph, and  $C(D)$  for the K-Street Registrant–Lobbyist graph (noting that by construction the clustering coefficient is bounded between 0 and 1). Both of them reflect the focusing characteristic of the registrant layer. Indeed, a typical registrant will have a cluster of clients and a cluster of lobbyists. We also note some slight temporal trends in the clustering coefficient (increasing for the upstream set), which might be related to the preferential attachment dynamics.

The dashed lines in Fig. S28 correspond to the expected value of the corresponding coefficient in a bipartite configuration model, i.e., a randomized bipartite network with the same partition and degree sequence [20]. We generally find a close match between this expected value and the empirically computed value.

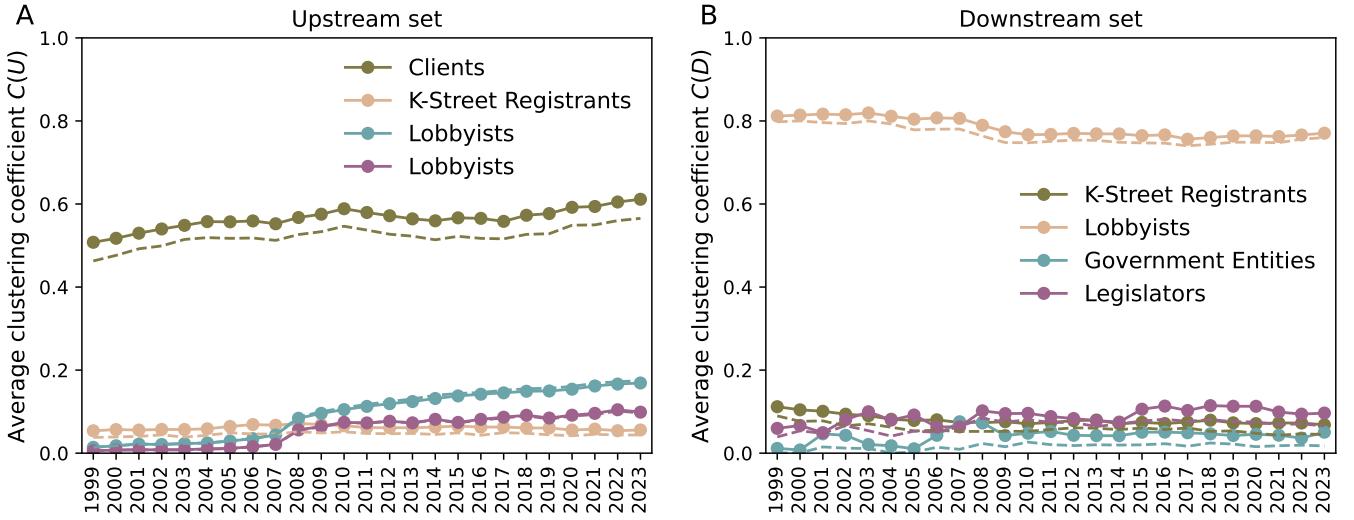


Figure S28. Clustering coefficient for bipartite subgraphs averaged over (A) the upstream set  $U$  and (B) the downstream set  $D$ . The solid line with circles denotes empirical measurements and the dashed line denotes the expected value of the configuration model.

## B. Largest component

Another question we can ask about a bipartite subgraph is the size of its largest connected component. Recall that we treat the subgraph as undirected; alternatively, in a directed graph, we would consider a weakly connected component.

In Fig. S29 we analyze what fraction of each set (upstream and downstream), the largest connected component captures. For all subgraphs, the results are consistent with the prediction of the bipartite configuration model, but we find that only the Client–K-Street Registrant subgraph possesses a large (giant) connected component at all times. For the Lobbyist–Government Entity and Lobbyist–Legislator subgraphs, we observe a transition with a large component starting to appear as the number of associations (edges) increases (c.f. Fig. S7C,D). The reader might notice these components contain almost all nodes in the downstream set (government entities or legislators), but only a minority of the nodes in the upstream set (lobbyists). This is because, even in recent years, only a minority of lobbyists have known governmental or lobbyist associations (c.f. Fig. S10).

## X. Registrant centrality

In this section, we explain the methodology of generating Fig. 3D of the main text, which relates the total income of a K-Street registrant  $r$  ( $w\text{-in-deg}(r, y)$ , as defined in eq. (S.24)) to the number of *reachable* associations (government entities and legislators). This number of associations  $\rho(r, y)$  can be thought of as a centrality score of a registrant node  $r$  and is simply the total number of government entities and legislators that can be reached from  $r$  by a path in  $\mathcal{G}_y$ . In Fig. 3D of the main text, we present a 2D histogram of  $\rho(r, y)$  (controlled for the weighted in-degree of the registrant  $w\text{-in-deg}(r, y)$ ) combined for all years  $y$  and all K-Street Registrants( $y$ ) (c.f. eq. (S.6)). This plot

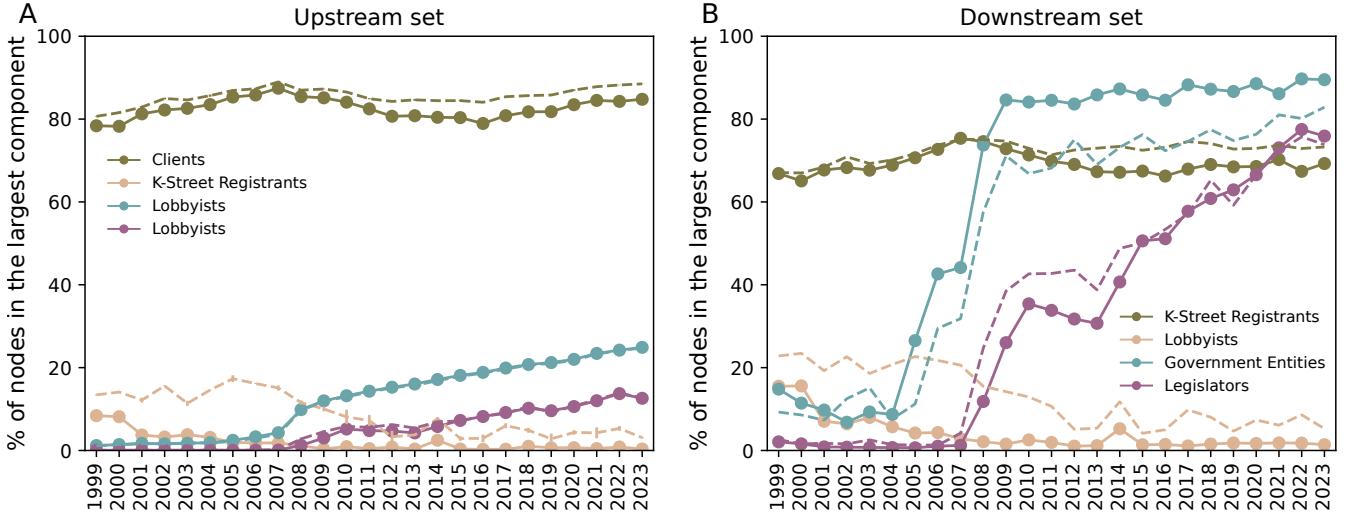


Figure S29. Largest component analysis of bipartite subgraphs. (A) The fraction of the upstream set contained in the largest component. (B) The fraction of the downstream set contained in the largest component. The dashed line denotes the expected value in the bipartite configuration model.

is reproduced in Fig. S30C. In Fig. S30A we present an analogous plot but only count the number of government entity associations, and in Fig. S30B we consider only the number of legislator associations. The combined number of associations (Fig. S30C) is the sum of the two scores. Nevertheless, we find that each of these scores in isolation (number of government associations and number of legislator associations) is a sufficient condition for high registrant income in and of itself.

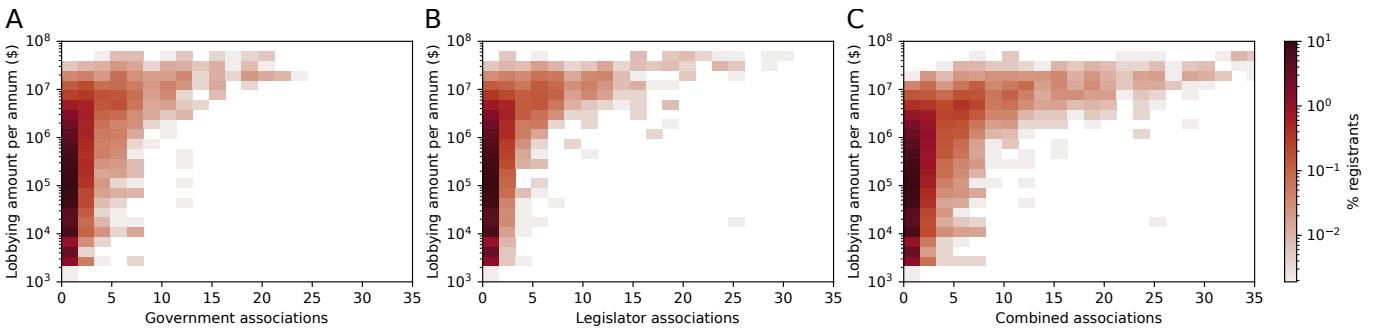


Figure S30. Total annual income of K-Street registrants in relation to their (A) number of government associations, (B) number of legislator associations, and (C) the total number of associations. The 2D histograms comprise all the data from 1999–2023, where one active registrant in one year contributes one data point.

## XI. Portfolio analysis

In this section, we describe the methodology of *lobbying issue portfolio* and *lobbying target portfolio* analysis presented in Fig. 3 of the main text respectively.

### A. Portfolio matrices

We start the lobbying issue portfolio analysis by counting the number of clients who in year  $y$  lobbied on issue area  $a$  and combine these scores into a matrix  $M$ , such that

$$M_{ay} = |\{c \in \text{Clients}(y) : \exists_f c(f) = c \wedge a \in A(f) \wedge y(f) = y\}|. \quad (\text{S.46})$$

This data is presented in Fig. S31A. The number of clients lobbying on a given issue can also reflect the total number of clients lobbying in a given year (c.f. Fig. S6A). To discount potential spurious temporal trend, instead of looking at the number of clients, we compute the fraction of clients (Fig. S31B)

$$\overline{M}_{gy} = \frac{M_{gy}}{|\text{Clients}(y)|}. \quad (\text{S.47})$$

Finally, to account for disparities in ‘popularity’ of different areas, we compute the issue portfolio matrix  $\overline{\overline{M}}$  by normalizing the matrix  $\overline{M}$  row by row such that

$$\overline{\overline{M}}_{gy} = \frac{\overline{M}_{gy}}{\frac{1}{N_y} \sum_y \overline{M}_{gy}}, \quad (\text{S.48})$$

where  $N_y = 25$  is the number of years considered. If  $\overline{\overline{M}}_{gy} > 1$ , it means that the fraction of reports mentioning entity  $g$  in year  $y$  is above the average for years 1999–2023.

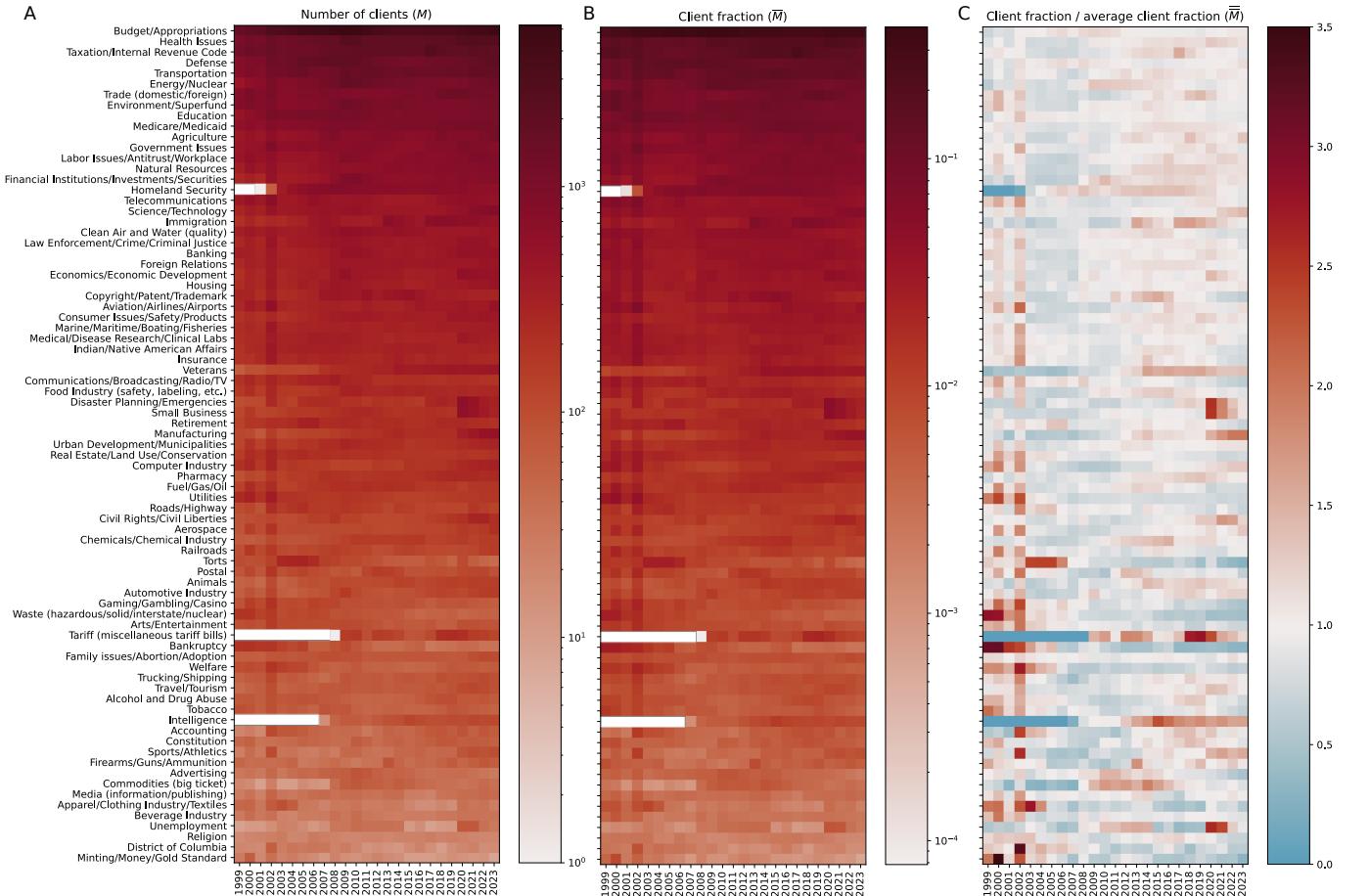


Figure S31. Issue portfolio analysis. (A) The number of clients lobbying on issue area  $a$  in year  $y$  ( $M_{ay}$ ). (B) The fraction of clients lobbying on issue area  $a$  in year  $y$  ( $\overline{M}_{gy}$ ). (C) Row-normalized fraction of clients lobbying on issue area  $a$  in year  $y$  ( $\overline{\overline{M}}_{gy}$ ).

The target portfolio analysis follows a very similar methodology, with the target portfolio matrix defined as

$$\overline{\overline{M}}_{ay}^I = \frac{\overline{M}_{gy}}{\frac{1}{N_y} \sum_y \overline{M}_{gy}}, \quad (\text{S.49})$$

where

$$\overline{M}_{ay}^I = \frac{|\{c \in \text{Clients}(y) : \exists_f c(f) = c \wedge g \in G(f) \wedge y(f) = y\}|}{|\text{Clients}(y)|}. \quad (\text{S.50})$$

Figure 3 of the main text presents the target portfolios for the top 79 government entities, i.e., the entities with the largest average number of clients. In Figures S32, we show such scores for the next 79 entities in this ranking, as well as the corresponding year-to-year Spearman correlation matrix, which is somewhat different to the correlation matrix for the top 79 issues. Nevertheless, we note that many government entities in this range are approached by only a handful of clients, so the ranking can be sensitive to small variations. For this reason, we do not pursue the target portfolio analysis for all 250 government entities mentioned in the reports, but we focus on the most popular ones.

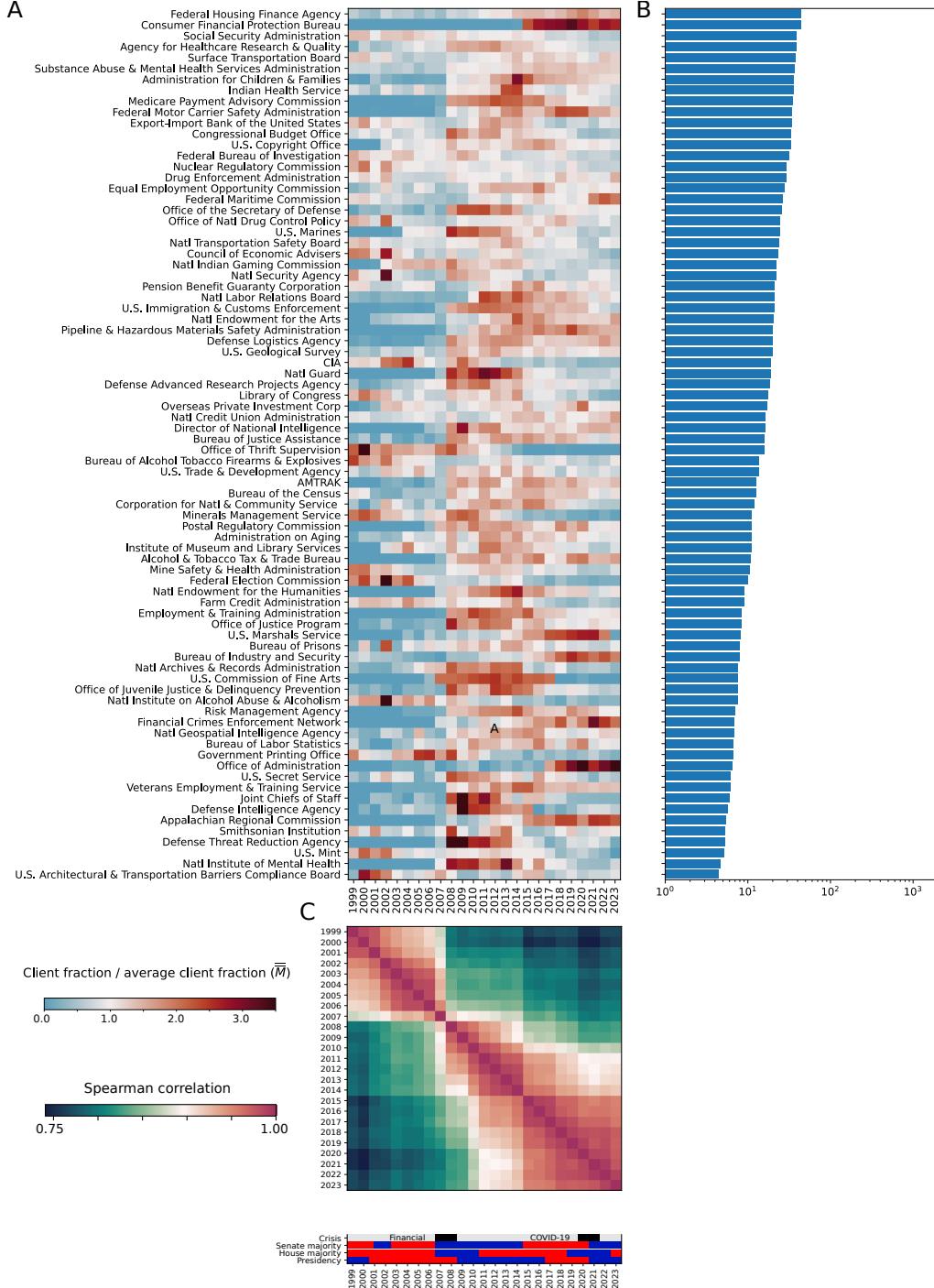


Figure S32. Target portfolio for less often mentioned government entities (80 – 159 in the ranking). (A) Portfolio matrix ( $\bar{M}$ ). (B) Number of clients lobbying with different government entities ( $M$ ). (C) Spearman correlation matrix for these government entities.

## B. Budget-based analysis

Instead of client group size, the lobbying portfolios can also be defined base on the lobbying budget, e.g. w could choose

$$\overline{M}_{ay} = \frac{\sum_f \mathbb{1}[a \in A(f) \wedge y(f) = y \wedge r(f)]}{\sum_f \mathbb{1}[y(f) = y]}. \quad (\text{S.51})$$

This idea is pursued in Fig. S33, which can be directly compared with Fig. 3 of the main text. We find that the correlation structure of the budget-based portfolios is less clear than for the client-group-based portfolios, e.g. the block-diagonal structure of Fig. S33F appears to be less apparent than the block-diagonal structure of Fig. 3F of the main text.

## C. Principal component analysis

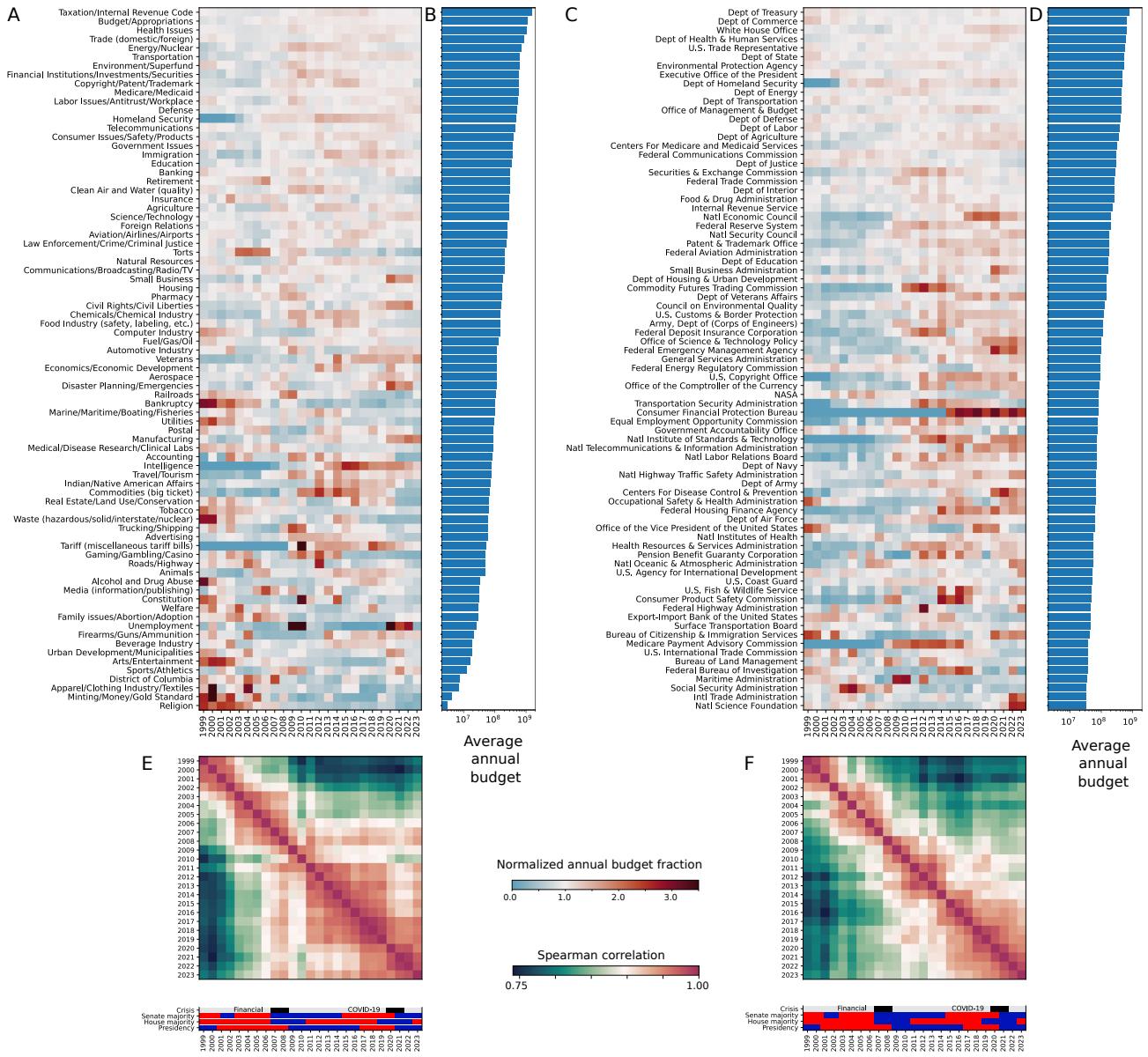
We can further analyze the similarity of annual portfolios by performing the principal component analysis (PCA). For the issue portfolio, we normalize  $\overline{M}_{ay}$  defined in eq. (S.48) one more time by column-wise subtraction of mean and division by standard deviation. Thus, we obtain the matrix  $\overline{\overline{M}}_{ay}$  that we decompose as

$$\overline{\overline{M}}_{ay} = \sum_i \sigma_i \mathbf{e}_i \mathbf{u}_i, \quad (\text{S.52})$$

where  $\sigma_1 \geq \sigma_2 \geq \dots$  are singular values,  $\mathbf{e}_i \in \mathbb{R}^{N_g \times 1}$  are principal components,  $N_a = 79$  is the number of issue areas,  $\mathbf{u}_i \in \mathbb{R}^{1 \times N_y}$  are score vectors, and  $N_y = 25$  is the number of analyzed years. Each entry of a principal component corresponds to one government entity while each entry of a score vector corresponds to one year.

Figure S34 shows the weights of the two highest-variance principal components  $\mathbf{e}_1$  and  $\mathbf{e}_2$  that explain more than 48% and 13% of variance, respectively. The projection of the target portfolio onto the 2D space  $(\mathbf{e}_1, \mathbf{e}_2)$  is shown in Fig. S35A. We can observe a steady evolution along the circle-like path from early years (top right) to recent years (top left). Also, note that the pairs of consecutive years (2007, 2008) and (2008, 2009) are considerably more dissimilar than others suggesting significant changes in the target portfolio between the two years. These changes might be caused by the financial crisis.

Analogously, we normalize the target portfolio matrices  $\overline{M}_{gy}$  and apply PCA to the obtained normalized matrices  $\overline{\overline{M}}_{gy}$ . Figure S35B shows that similarly to the issue portfolio,  $\mathbf{e}_1$  also co-evolves with time. For completeness, the weights of the first two principal components  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are shown in Fig. S36, but the interpretation of this result is beyond the scope of this study.



**Figure S33. Lobbying portfolio analysis based on budget.** (A) Fraction of the lobbying budget dedicated to a given issue, scaled by its average value in the period 1999–2023. (B) Average annual lobbying budget associated with a given issue. (C) Fraction of the lobbying budget that can be associated with lobbying a given government entity in a given year, scaled by the average value in the period 1999–2023. The government entities are listed according to the average annual budget. In this Figure, we present only the top approached government entities; note that the ranking may be different than in Fig. 3. (D) Average annual lobbying budge dedicated to lobbying different government entities. (E) Year-to-year correlation of the *issue portfolio vectors* (columns of the matrix in panel (A)). (F) Year-to-year correlation of the *government approach portfolios* presented as columns in panel (C).

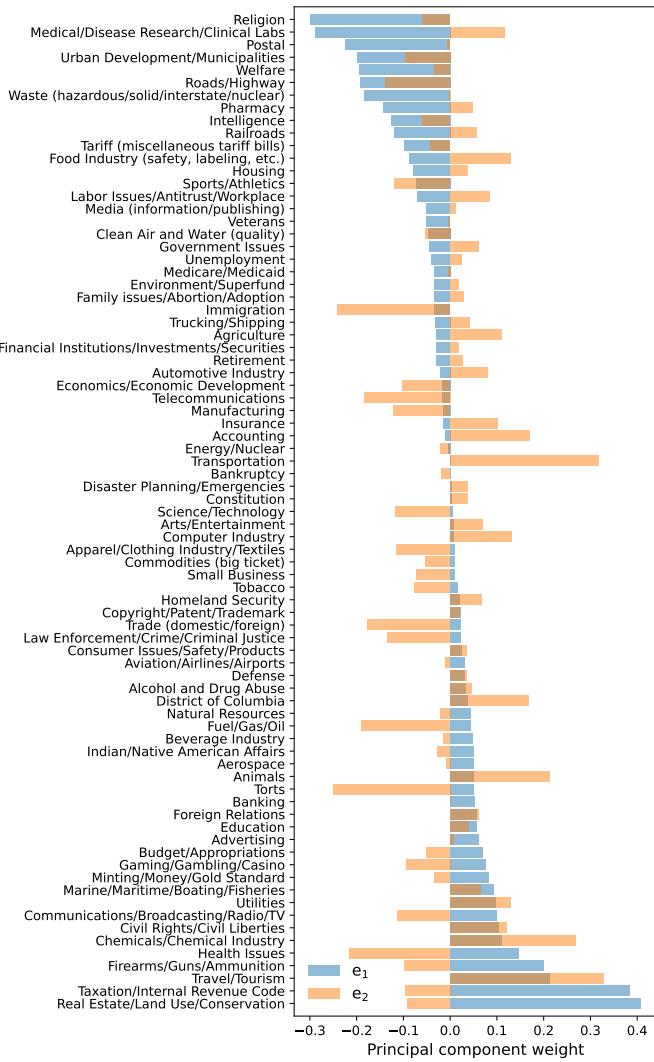


Figure S34. Weights of the first two principal components  $e_1$  and  $e_2$  of the issue portfolio matrix  $\bar{\bar{M}}_{ay}$ . Components  $e_1$  and  $e_2$  explain more than 38% and 13% of variance respectively. Issue areas are sorted by increasing  $e_1$  weight.

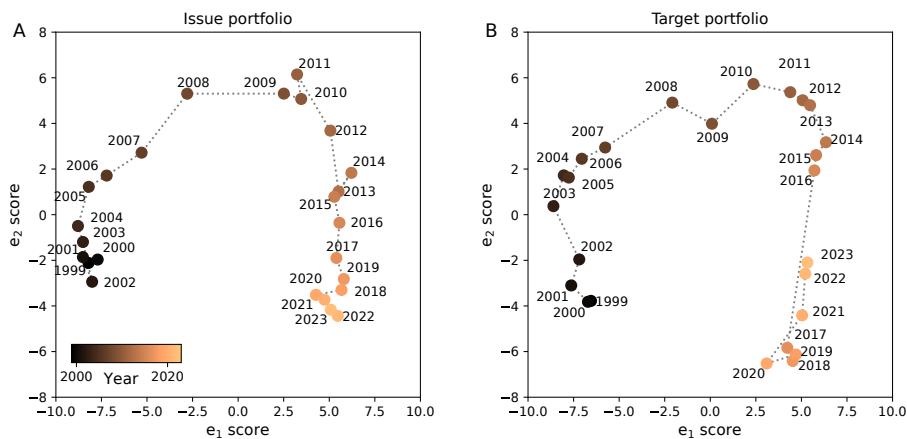


Figure S35. Projection of the annual portfolios onto 2D space spanned by the principal components  $e_1$  and  $e_2$ . (A) Issue portfolios. (B) Target portfolios (top 79 entities).

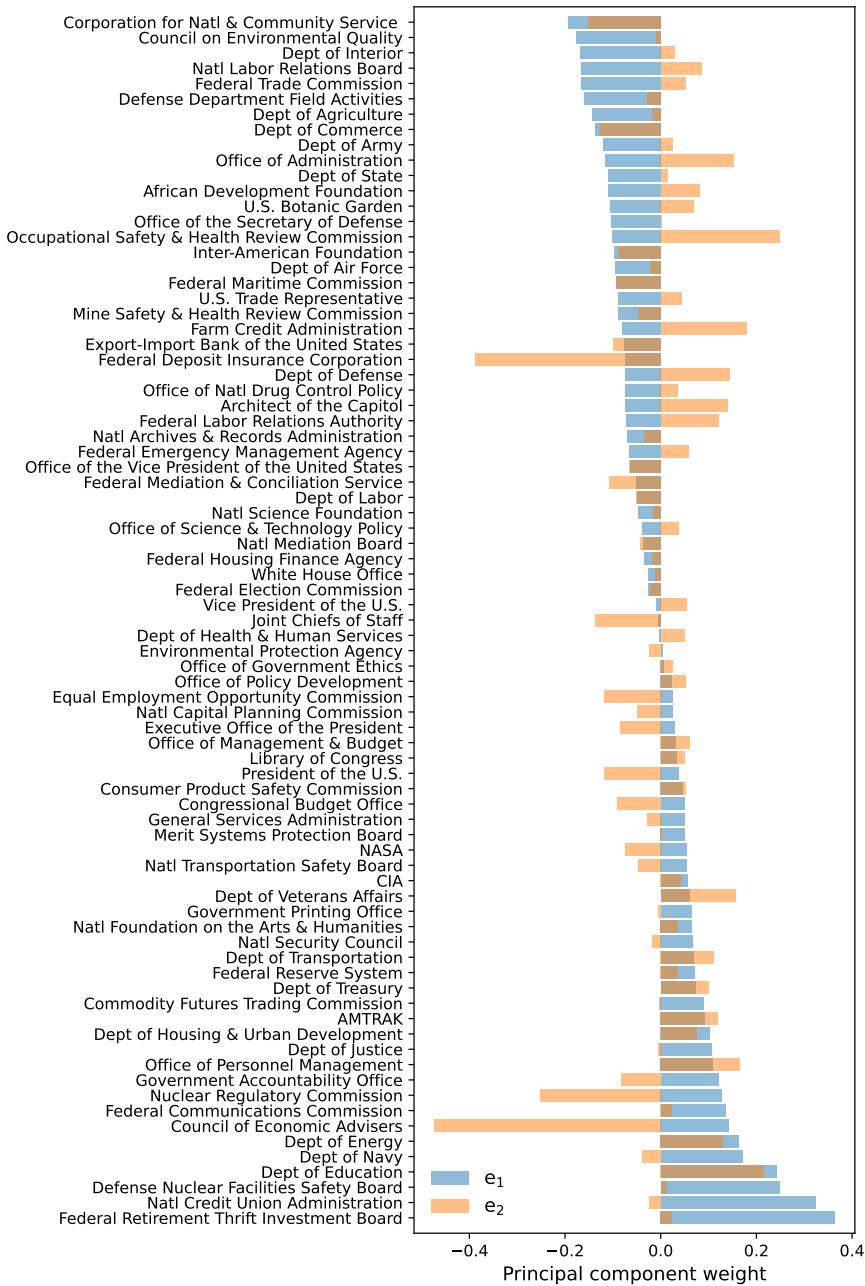


Figure S36. Weights of the first two principal components  $e_1$  and  $e_2$  of the target portfolio matrix  $\bar{M}_{gy}$ . Components  $e_1$  and  $e_2$  explain more than 42% and 21% of variance respectively. Government entities are sorted by increasing  $e_1$  weight.

## XII. Polarization analysis

In the last paragraphs of the article, we use our data to quantify political polarization. Our analysis is exploratory in nature. Here, we explain its methodology.

### A. Relation-based measures

In order to compare the polarization of lobbyists (trend unknown) to the polarization of lawmakers (polarization known to progress), we need a methodology that could be applied to both. In Fig. 4D we used a simple network-based measure that can be applied to any set of entities  $\mathcal{S}$  subdivided into two subsets  $\mathcal{S}_\infty, \mathcal{S}_\epsilon$ , which are joined by some kind of relational ties.

In case of legislators,  $\mathcal{S}_1, \mathcal{S}_2$  correspond to representatives/senators who belong to one of the two main political parties, e.g. we can take

$$\mathcal{S}_1(y) := \text{Democratic Senators}(y) = \{p : \text{Party}(p, y) = \text{Democratic}, \text{Chamber}(p, y) = \text{Senate}\},$$

$$\mathcal{S}_2(y) := \text{Republican Senators}(y) = \{p : \text{Party}(p, y) = \text{Republican}, \text{Chamber}(p, y) = \text{Senate}\},$$

and the superset is simply

$$\mathcal{S} = \mathcal{S}_1(y) \cup \mathcal{S}_2(y).$$

Thus, we neglect here the independent legislators, but their number is small and it is not always straightforward to define how they might contribute to polarization.

The relations between legislators can be defined in all sorts of ways. Here, we focus on bill cosponsoring. We can envisage the legislators as nodes in the multigraph, and we add one edge  $(p_1, p_2)$  between any pair of politicians for each bill they co-sponsored together. We treat each chamber separately, neglect resolutions, and treat sponsors and cosponsors in the same way.

The edges can connect members of the same party or members of two parties. To evaluate the degree of polarization, we look at the number of bipartisan cosponsorships. Specifically, we compute

$$\text{Legislator Bipartisan Index}(y) = \frac{|(p_1, p_2) : p_1 \in \mathcal{S}_1, p_2 \in \mathcal{S}_2|}{|(p_1, p_2) : p_1, p_2 \in \mathcal{S}|}. \quad (\text{S.53})$$

We compute the bipartisan index for House and Senate separately, and we report it in Fig. 4D of the main paper. The fact that it has been consistently decreasing throughout the 21st century can be interpreted as a symptom of Congress polarization.

Crucially, we can use a very similar measure to assess polarization among the lobbyists. Now, we define

$$\mathcal{S}_1(y) := \{l \in \text{Lobbyists}(y) : p(l, y) \cap \text{Democratic Legislators}(y) \neq \emptyset, p(l, y) \cap \text{Republican Legislators}(y) = \emptyset\},$$

$$\mathcal{S}_2(y) := \{l \in \text{Lobbyists}(y) : p(l, y) \cap \text{Republican Legislators}(y) \neq \emptyset, p(l, y) \cap \text{Democratic Legislators}(y) = \emptyset\},$$

where

$$\text{Democratic Legislators}(y) = \{p : \text{Party}(p, y) = \text{Democratic}\},$$

$$\text{Republican Legislators}(y) = \{p : \text{Party}(p, y) = \text{Republican}\}.$$

Again, we neglect lobbyists associated with independent legislators, and lobbyists associated with legislators from both of the major parties, but such cases are rare (c.f. Sec. I.D).

As before,  $\mathcal{S} = \mathcal{S}_1(y) \cup \mathcal{S}_2(y)$  and we want to treat the lobbyists as nodes of a graph, with links representing some kind of ties. In our analysis we considered two different kinds of links. First, we constructed the graph based on professional filings, i.e. we add a link  $(l_1, l_2)$  for every filing where the two lobbyists featured together. This led to the collaboration-based bipartisan index. As an alternative, we consider a graph where an edge is added for every bill

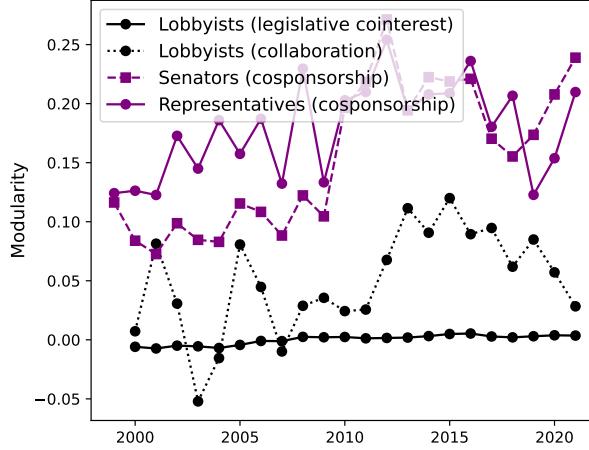


Figure S37. Modularity of the cross-partisan collaboration graphs - an analogue o Fig. 4D of the main text with a different measure.

that the two lobbyists lobbied on together (potentially for different clients). This lead to the bipartisan index based on legislative interest. Both of these indices are computed with the same formula

$$\text{Lobbyist Bipartisan Index}(y) = \frac{|(l_1, l_2) : l_1 \in \mathcal{S}_1, l_2 \in \mathcal{S}_2|}{|(l_1, l_2) : l_1, l_2 \in \mathcal{S}|}, \quad (\text{S.54})$$

which is exactly analogous to (S.53), except the edges are constructed based on a different notion of a relationship. The lobbyists bipartisan indices do not seem to decrease in the same way they do for legislators. Thus, as far as this simple measure is concerned, we do not find symptoms of increasing polarization in the lobbying industry.

### Modularity

As an alternative to the bipartizan index, we could also take the the *modularity* of the relationship graph [20], partitioned into  $\mathcal{S}_1(y)$  and  $\mathcal{S}_2(y)$ . Let  $A_{ij}$  be the number of bills co-sponsored by legislators  $p_i$  and  $p_j$  (edge weight),  $k_i$  be the total number of co-sponsorship ties of legislator  $i$  (node degree), and  $m$  be the total number of edges. The modularity is defined as

$$Q(y) = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \mathbb{1}[\text{Party}(p_1, y) = \text{Party}(p_2, y)],$$

and analogously for the lobbyists.

The results are shown in Fig. ???. For lobbyists, the modularity is close to zero, which indicates a lack of polarization. For legislators, we find consistently higher values of modularity, and in the case of senators we observe a marked increase of modularity over time.

## B. Reach-based measures

In Fig. 4E of the main paper, we present a complementary measure of polarization in lobbying, focused on registrants and clients. We employ here the notion of *reach* introduced in Sec. X. For example, we say that an agent (registrant or client) is able to ‘reach’ the Democratic Party, if it employs a lobbyist with a historical association to a Democratic legislators. Based on this idea, we divide clients and registrants into four categories: those who are not connected to any political, agents with documented connections to the Democratic Party only (democratic profile), agents with documented connections to the Republican Party only (republican profile), and agents connected to both Democrats and Republicans (bipartisan profile). We neglect the first group, ant the sizes of the other three groups across the years are shown in Fig S38.

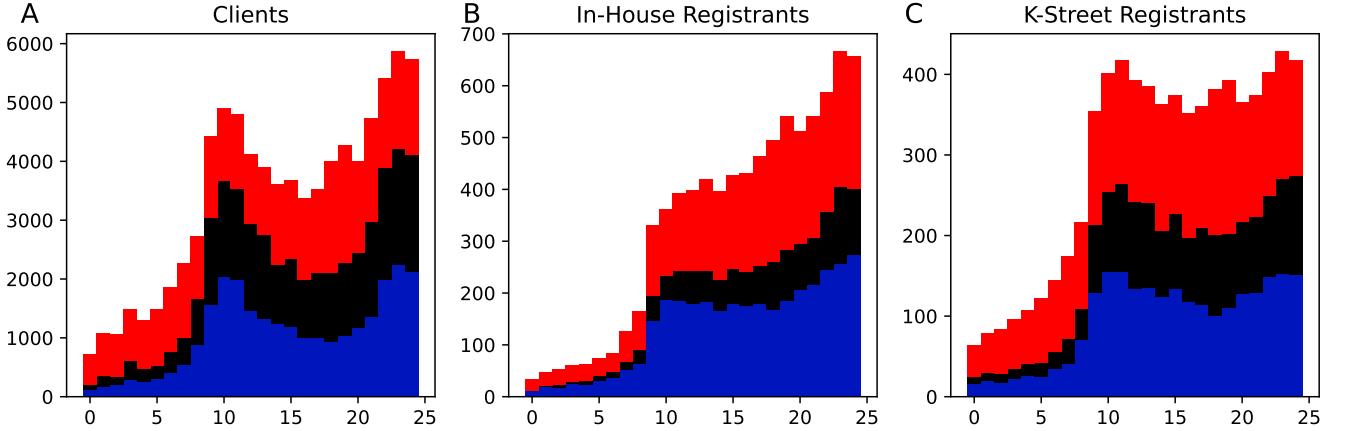


Figure S38. Reach-based polarization analysis. (A) The number of clients with democratic profile (blue), bipartisan profile (black), and republican profile (red). (B) Analogous plot for In-House registrants. (C) Analogous plot for K-Street registrants. In all cases, the relative number of the bipartisan profile agents increases in time, which could be indicative of depolarization.

In Fig. 4B, we use the relative size of the bipartisan set as a measure of polarization, e.g. for clients we compute

$$\frac{|\text{Bipartisan profile clients}|}{|\text{Bipartisan profile clients}| + |\text{Democratic profile clients}| + |\text{Republican profile clients}|}.$$

We note that these measures increase for all the agent types, which indicates that polarization in lobbying might even be decreasing.

### XIII. Lobbying pathways

In this section, to illustrate in more detail the complex dynamics of the lobbying network, we perform a fine-grained analysis of the information flow from a client (lobbying initiative) to a legislator (lobbying target) in response to the COVID-19 pandemic. One of the most prominent spikes in lobbying activity during that period is related to Small Business issues (Fig. 3A of the main body). We are specifically interested in which lobbying pathways that contributed most significantly to Small Business lobbying before (up to and including year 2019) and during (years 2020 and 2021) the pandemic. To this end, we will construct information flow diagrams, constructed from the fine-grained network data (c.f. Fig. 1B) by clustering nodes of similar type (clients based on the lobbying budget, registrants based on their profile, and legislators based on chamber and partisan affiliation). We first describe the methodology of constructing the alluvial diagrams, and then briefly describe the findings.

#### A. Methodology

Our construction can be divided into two steps: network coarse-graining and flow inference. In our example, we focus on lobbying related to small business in years 2019 and 2020, but this methodology can be generalized to any issue area  $a$  in any year  $y$ .

##### 1. Coarse-graining

We coarse-grain the lobbying graph  $\mathcal{G}_y$  by clustering the client nodes into two disjoint sets  $C_<$  and  $C_{\geq}$  based on their lobbying expenditure such that

$$C_< = \{c \in \text{Clients}(y) : c \in V(\mathcal{G}_y) \wedge \text{w-out-deg}(c, y) < \theta_c\},$$

and

$$C_{\geq} = \{c \in \text{Clients}(y) : c \in V(\mathcal{G}_y) \wedge \text{w-out-deg}(c, y) \geq \theta_c\},$$

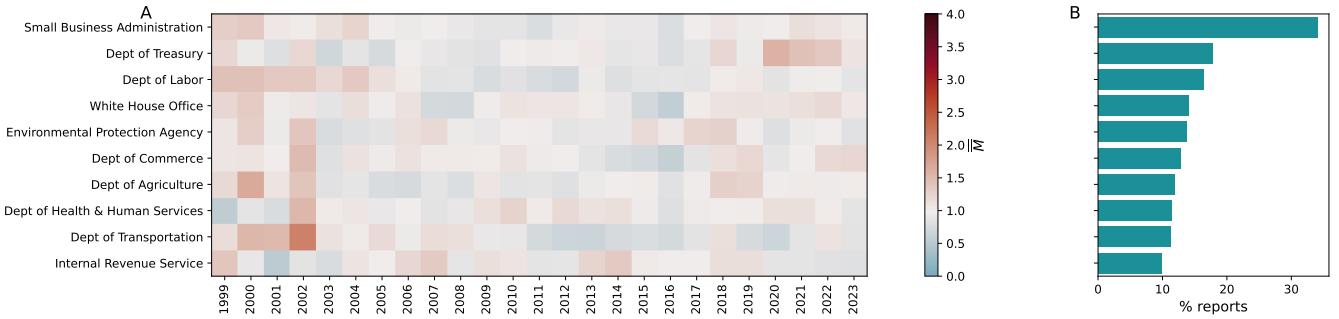


Figure S39. (A) Target portfolio  $\overline{M}_{gya}$  for small business lobbying for the 10 most approached government entities (excluding Senate and House of Representatives). (B) Average fraction of reports that mention the small business issue area and a given government entity.

where we set the threshold  $\theta_c = \$1,000,000$ .

The registrant nodes are divided into K-Street ( $R_K$ ) and In-House ( $R_I$ ) according to definition (S.2). The lobbyist nodes are not explicitly represented in our coarse-grained graph.

In the government entity layer, we retain only four nodes  $G_1, G_2, G_3$  and  $G_4$  corresponding to entities with the largest numbers of mentions in the specified issue area  $\frac{1}{N_y} \sum_y \overline{M}_{gya}$  excluding Congress, where

$$\overline{M}_{gya} = \frac{\sum_f \mathbb{1}[g \in G(f) \wedge y(f) = y \wedge a(f) = a]}{\sum_f \mathbb{1}[y(f) = y \wedge a(f) = a]}. \quad (\text{S.55})$$

For the Small Business issue area, we find  $G_1 = \text{Small Business Administration}$ ,  $G_2 = \text{Department of Treasury}$ ,  $G_3 = \text{Department of Labor}$ , and  $G_4 = \text{White House Office}$ . The ranking of the top 10 most approached government entities is presented in Fig. S39, where we also show the associated target portfolio  $\overline{M}_{gya}$ .

The legislator nodes are classified into four disjoint sets based on their chamber and party affiliation: Democratic Senators ( $P_{DS}$ ), Republican Senators ( $P_{RS}$ ), Democratic Representatives ( $P_{DH}$ ), and Republican Representatives ( $P_{RH}$ ). Legislators without a party affiliation are not considered in this analysis.

## 2. Flow construction

The key mathematical object that represents our year-specific, issue area-specific, alluvial diagram is the flow graph  $\mathcal{F}_{y,a}$ . The vertices of  $\mathcal{F}_{y,a}$  are the 12 sets introduced in the previous section, i.e.,

$$V(\mathcal{F}_{y,a}) = \{C_<, C_>, R_K, R_I, G_1, G_2, G_3, G_4, P_{DS}, P_{RS}, P_{DH}, P_{RH}\}.$$

The edges of  $\mathcal{F}_{y,a}$  are weighted and constructed in the following manner. First, we set all the edge weights to zero ( $\Phi(N_1, N_2) = 0$  for all pairs of nodes  $N_1$  and  $N_2$ ). We then sequentially update the weights by scanning all the filings  $f$  filed in year  $y$  that mention the issue area  $a$ , i.e., the filings where

$$y(f) = y \quad \wedge \quad a \in A(f). \quad (\text{S.56})$$

It is worth noting that the reports that actually contribute to the flow graph are the ones that satisfy

$$y(f) = y \quad \wedge \quad a \in A(f) \quad \wedge \quad \{G_1, G_2, G_3, G_4\} \cap G(f) \neq \emptyset \quad \wedge \quad \bigcup_{l \in L(f)} p(l, y) \neq \emptyset. \quad (\text{S.57})$$

This stricter condition ensures that we are able to reconstruct the lobbying path from client to legislator. Necessarily, the filings that satisfy condition (S.57) are only a subset of all the filings that satisfy condition (S.56). For the Small Business issue area in 2019 and 2020, the set of filings satisfying (S.57) comprises around 16% of the filings satisfying (S.56), namely, 113 out of 686, and 351 out of 2,046 filings respectively. The details of the flow graph construction algorithm is presented below. The resulting flow graph  $\mathcal{F}_{y,a}$  is then passed to the MATLAB *Sankey plot* package [21] to generate the alluvial diagram.

```

1: for  $f$  satisfying (S.57) do
2:
3:   if w-out-deg( $c(f), y) < \theta_c$  then                                 $\triangleright$  Determine client type
4:      $C := C_<$ 
5:   else
6:      $C := C_{\geq}$ 
7:   end if
8:
9:   if  $r(f) \in \text{K-Street}$  then                                          $\triangleright$  Determine registrant type
10:     $R := R_K$ 
11:   else
12:     $R := R_I$ 
13:   end if
14:
15:    $\mathcal{GOV} := \{G_1, G_2, G_3, G_4\} \cap G(f)$                                 $\triangleright$  Determine involved government entities
16:
17:    $\mathcal{LEG} := \bigcup_{l \in L(f)} p(l, y)$                                           $\triangleright$  Determine associated legislators
18:
19:    $\Phi(C, R) := \Phi(C, R) + 1$                                                $\triangleright$  Update client-registrant flow
20:
21:   for  $G \in \mathcal{GOV}$  do
22:      $\Phi(R, G) := \Phi(R, G) + \frac{1}{|\mathcal{GOV}|}$                                  $\triangleright$  Update registrant-government flow (normalize for flow conservation)
23:     for  $p \in \mathcal{LEG}$  do
24:       if Party( $p, y$ ) = Democratic  $\wedge$  Chamber( $p, y$ ) = Senate then
25:          $P := P_{DS}$ 
26:       else if Party( $p, y$ ) = Republican  $\wedge$  Chamber( $p, y$ ) = Senate then
27:          $P := P_{RS}$ 
28:       else if Party( $p, y$ ) = Democratic  $\wedge$  Chamber( $p, y$ ) = House then
29:          $P := P_{DH}$ 
30:       else if Party( $p, y$ ) = Republican  $\wedge$  Chamber( $p, y$ ) = House then
31:          $P := P_{RH}$ 
32:       end if
33:        $\Phi(G, P) := \Phi(G, P) + \frac{1}{|\mathcal{GOV}||\mathcal{LEG}|}$   $\triangleright$  Update government-legislator flow (normalize for flow conservation)
34:     end for
35:   end for
36:
37: end for

```

## B. Findings

In Fig. S40 we present the resulting alluvial diagrams for a range of years (2018–2023) in the neighborhood of the COVID-19 pandemic onset. First, we note that the diagram for 2018 is very similar to 2019, i.e., in the absence of perturbations (in this case the COVID-19 pandemic), the lobbying flow is relatively stable. Following the pandemic we observe that large clients started lobbying on Small Business issues through K-street firms. Moreover, we note a marked increase of lobbying approaches to the Department of Treasury, reflecting its pivotal role in the implementation of the Coronavirus Aid, Relief, and Economic Security (CARES) Act. Furthermore, the information flow diagrams traces of ‘venue shopping’, with lobbyists in 2020 strategically Republican-controlled U.S. Senate. Some of the pandemic-induced redistribution of the lobbying flow persists (e.g., the larger participation of the K-Street firms), while some revert back to mean pre-pandemic distributions (e.g., the increased involvement of the Department of Treasury).

This case study is only one example of how the *LobbyView* database can serve as a resource for quantitative future analysis and model development to understand and predict the issue-specific adaptive response behaviors of the lobbying network.

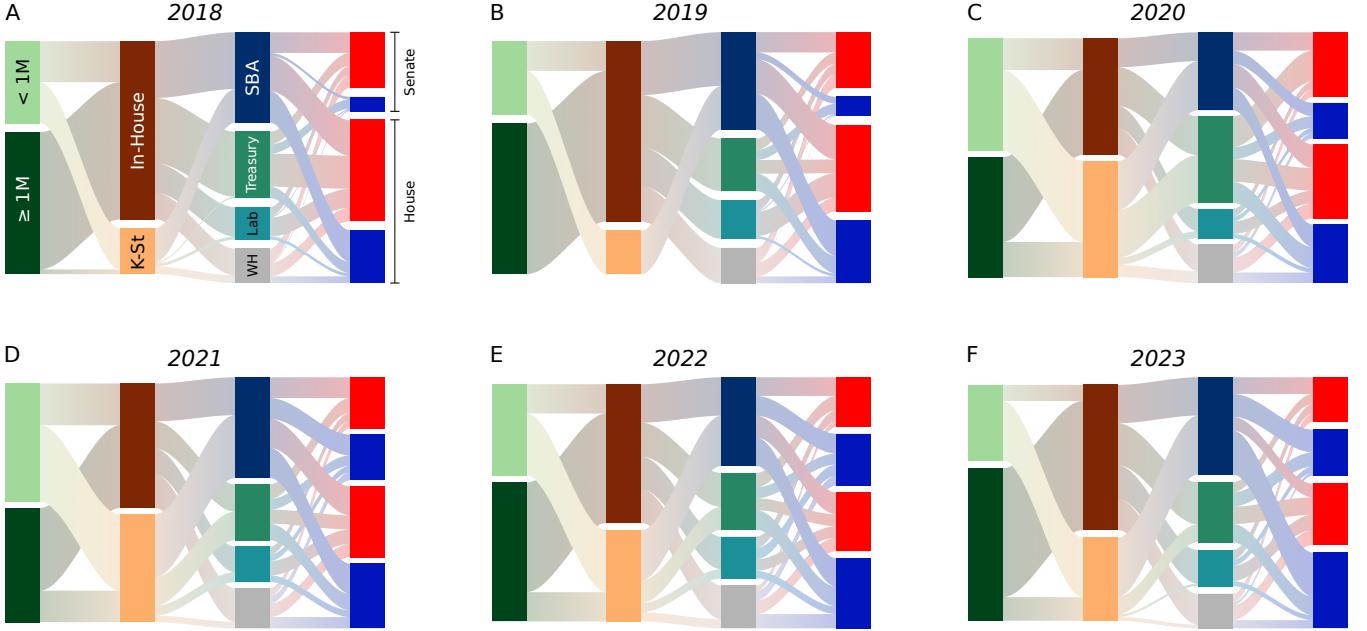


Figure S40. End-to-end (client-to-legislator) lobbying flow for the small business issue in 2018–2023 in the neighborhood of the COVID-19 pandemic onset. Four layers (left to right) represent clients (categorized based on their lobbying expenditure), registrants (K-Street, In-House), government entities (White House Office (WH), Department of Labor (Lab), Department of Treasury (Treasury), Small Business Administration (SBA)), and legislators (Democrats in blue, Republicans in red, either in Senate or House of Representatives).

#### XIV. Data and code availability

Anonymized relational data, sufficient to reproduce the results of this paper, is publicly available at (*link released upon acceptance*). The code used to analyze the data and produce the figures can be accessed at a GitHub repository (*link released upon acceptance*). Researchers who are interested in interfacing with the broader *LobbyView* database can do so by visiting our website, [lobbyview.org](http://lobbyview.org). Registration is free and instant, and allows access to the CSV data downloads and query API for report, client, bill, issue, and lobbyist-level data as well as issue text and client-politician connection dyad data. Researchers who require access to more detailed or granular data or fields not exposed in the CSVs or API are invited to contact the authors for access as necessary.

- 
- [1] LobbyView Database, <https://www.lobbyview.org/> (2024).
  - [2] U.S. Congress, Lobbying Disclosure Act of 1995 (1995). Public Law No: 104-65, 109 Stat. 691.
  - [3] U.S. Senate Office of Public Records, Lobbying Disclosure Act Guidance (2021).  
<https://lobbyingdisclosure.house.gov/ldaguidance.pdf>.
  - [4] S. G. M. Intelligence, Compustat, <https://www.computstat.com>.
  - [5] Moody's, Orbis, <https://www.moody's.com/web/en/us/capabilities/company-reference-data/orbis.html>.
  - [6] United States Digital Service, U.S. Digital Service GitHub Repository: Congress (2024).  
<https://github.com/unitedstates/congress/>.
  - [7] U.S. Congress, Honest Leadership and Open Government Act of 2007 (2007).  
<https://www.govinfo.gov/content/pkg/PLAW-110publ181/pdf/PLAW-110publ181.pdf>.
  - [8] U.S. Congress, Lobbying Disclosure Technical Amendments Act (1998).  
[https://lobbyingdisclosure.house.gov/lda\\_technical.html](https://lobbyingdisclosure.house.gov/lda_technical.html).
  - [9] U.S. Congress, Lobbying Disclosure Act Guidance (2008).  
[https://lobbyingdisclosure.house.gov/amended\\_lda\\_guide.html](https://lobbyingdisclosure.house.gov/amended_lda_guide.html).
  - [10] A.-L. Barabási, *Network Science* (Cambridge University Press, 2016).
  - [11] A.-L. Barabási, R. Albert, H. Jeong, *Physica A: statistical mechanics and its applications* **281**, 69 (2000).
  - [12] O. T. Courtney, G. Bianconi, *Physical Review E* **97**, 052303 (2018).
  - [13] B. Bassetti, M. Zarei, M. C. Lagomarsino, G. Bianconi, *Physical Review E* **80**, 066118 (2009).
  - [14] *Supplementary Information*.

- [15] A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
- [16] G. Bianconi, A.-L. Barabási, *Europhysics Letters* **54**, 436 (2001).
- [17] P. Erdős, A. Rényi, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17 (1960).
- [18] Graphistry, Inc., PyGraphistry: Dataframe-native Python visual graph AI library (Version 0.31.1). (2024).  
<https://github.com/graphistry/pygraphistry>.
- [19] T. M. J. Fruchterman, E. M. Reingold, *Software: Practice and experience* **21**, 1129 (1991).
- [20] M. Newman, *Networks* (Oxford University Press, 2018).
- [21] Z. Liu, MATLAB Sankey plot package (Version 2.0.1). (2024).  
<https://www.mathworks.com/matlabcentral/fileexchange/128679-sankey-plot>.