Governments in modern societies undertake an array of complex functions that shape politics and economics, individual and group behavior, and the natural, social, and built environment. How are governments structured to execute these diverse responsibilities? How do those structures vary, and what explains the differences? To examine these longstanding questions, we develop a technique for mapping Internet “footprint” of government with network science methods. We use this approach to describe and analyze the diversity in functional scale and structure among the 50 US state governments reflected in the webpages and links they have created online: 32.5 million webpages and 110 million hyperlinks among 47,631 agencies.

We first verify that this extensive online footprint systematically reflects known characteristics: 50 hierarchically organized networks of state agencies that scale with population and are specialized around easily identifiable functions in accordance with legal mandates. We also find that the footprint reflects extensive diversity among these state functional hierarchies. We hypothesize that this variation should reflect, among other factors, state size, economic structure, ideology, and location. We find that government structures are most strongly associated with state economic structures, with location and income playing more limited roles. Voters’ recent ideological preferences about the proper roles and extent of government are not significantly associated with the scale and structure of their state governments as reflected online. We conclude that the online footprint of governments offers a broad and comprehensive window on how they are structured that can help deepen understanding of those structures.

Functional structures of US state governments

Stephen Kosacka,b,c,1, Michele Coscia,d,1, Evann Smitha, Kim Albrecht,e,f,g, Albert-László Barabási,f,g,h,1,2, and Ricardo Hausmanni,j,k

To whom correspondence may be addressed. Email: ricardo_hausmann@hks.harvard.edu or albi@northeastern.edu.

1Evans School of Public Policy and Governance, University of Washington, Seattle, WA 98195; 2Ash Center for Democratic Governance and Innovation, Kennedy School of Government, Harvard University, Cambridge, MA 02139; 3Center for International Development, Harvard University, Cambridge, MA 02139; 4Department of Computer Science, IT University of Copenhagen, 2300 Copenhagen, Denmark; 5metaLAB (at) Harvard, Berkman Klein Center for Internet & Society at Harvard University, Cambridge, MA 02138; 6Network Science Institute, Northeastern University, Boston, MA 02115; 7Physics Department, Northeastern University, Boston, MA 02115; 8Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115; 9Center for Network Science, Central European University, Budapest 1051, Hungary; 10Kennedy School of Government, Harvard University, Cambridge, MA 02139; and 11Santa Fe Institute, Santa Fe, NM 87501

*Evans School of Public Policy and Governance, University of Washington, Seattle, WA 98195; Ash Center for Democratic Governance and Innovation, Kennedy School of Government, Harvard University, Cambridge, MA 02139; Center for International Development, Harvard University, Cambridge, MA 02139; Department of Computer Science, IT University of Copenhagen, 2300 Copenhagen, Denmark; metaLAB (at) Harvard, Berkman Klein Center for Internet & Society at Harvard University, Cambridge, MA 02138; Network Science Institute, Northeastern University, Boston, MA 02115; Physics Department, Northeastern University, Boston, MA 02115; Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115; Center for Network Science, Central European University, Budapest 1051, Hungary; Kennedy School of Government, Harvard University, Cambridge, MA 02139; and Santa Fe Institute, Santa Fe, NM 87501

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and approved October 3, 2018 (received for review March 21, 2018)

Significance

Understanding of modern government is limited by a lack of comprehensive, reliable, comparable data on what governments do and how they are organized to execute their diverse responsibilities. We demonstrate that such data can be collected from the extensive footprint that governments leave on the Internet, opening a range of unresolved puzzles and questions about modern government to closer empirical inquiry. The online footprint of the 50 US state governments reflects their close embeddedness with state economies and suggests that other factors widely hypothesized to influence government play more limited roles, including location and income. It also casts doubt on the degree to which state government functional structures systematically reflect voters’ recent ideological preferences.

Author contributions: S.K., M.C., and R.H. conceived the study; M.C. designed the data-collection methodology; S.K., M.C., E.S., A.-L.B., and R.H. designed the analyses; E.S. developed and implemented the functional classification; K.A. created the figures; and S.K. wrote the paper with assistance from A.-L.B. and R.H.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Replication scripts and all data used in the paper have been deposited in the Harvard Dataverse (https://dataverse.harvard.edu/dataverse/govmap) and are also available at govmaps.cit.hks.harvard.edu.

1S.K. and M.C. contributed equally to this work.
2To whom correspondence may be addressed. Email: ricardo_hausmann@hks.harvard.edu or albi@northeastern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1803228115/-/DCSupplemental.

Published online October 29, 2018.

www.pnas.org/cgi/doi/10.1073/pnas.1803228115

11748-11753 | PNAS | November 13, 2018 | vol. 115 | no. 46
online presence of governments has opened an analytically useful window on modern government, and end with several additional puzzles and questions about modern governments that their Internet footprints may open to closer empirical inquiry.

**The Internet Footprint of US State Governments**

We crawled the web to collect publicly available digital traces of state government agencies in 2014, including the content of agency webpages and the hyperlinks among them (Fig. 1).

We interpret agency websites as representations of public agencies implementing public responsibilities; the title of each agency’s page as representing the agency’s function; and the number of pages as an approximate reflection of the scope or density of its activities in implementing that function; and hyperlinks between agency webpages as one agency reporting the relevance of another to its activities. Although some government functions lack an open Internet presence, our crawl recovered an extensive online footprint: the websites of 47,631 state agencies, with a total of 32.5 million webpages and 110 million hyperlinks among them. For example, among the 2,453 public agencies we identify in the State of New York are the State Assembly, which has the majority of its connections to the Senate; the Lottery Division, which is strongly connected to the Gaming Commission, which oversees its operations, and more weakly connected to the Alcoholism and Substance Abuse Office; and the Joint Commission on Public Ethics, which is almost completely isolated aside from links from several universities. The number of these websites grew rapidly through the 1990s and early 2000s, averaging 79% annually from 1996 to 2002, and in recent years has slowed to less than 2% annually, suggesting that most agencies that are going to represent themselves online have already done so. (See SI Appendix, sections 1 and 2 on data gathering and limitations.)

Altogether, these digital traces offer a detailed reflection of the bureaucratic structures by which state governments implement their diverse responsibilities.

**Functional Specialization in the Footprint**

Based on their websites, the functional range of these 47,631 agencies exceeds the most comprehensive current classification of government functions, the US Census Bureau’s Census of Governments. For example, the Census does not include international trade; financing authorities; ethics, lobbying, and campaign finance commissions; utilities regulation; civil and human rights; support for children and families; and public and independent defense—all activities that the websites of government agencies regularly describe as their primary functions.

To develop a classification system that includes the full functional range of state governments, we used Latent Derelict Analysis to generate keywords representing each function as described in agency website titles, matched each to the corresponding functional category in the Census, and appended new categories for functions not among the Census categories. The result is a classification system that expands the 51 Census categories into 166 specific functions (e.g., Consumer Protection and Education; Fire Protection) nested within 28 general categories (e.g., Commerce and Economic Development; Public Safety). Table 1 lists the 28 general categories; see SI Appendix, section 3 for specific functions. Agencies whose website titles contained keywords pertaining to only one functional category were assigned to that category; we manually reviewed and assigned all agencies with keywords pertaining to multiple categories.

To validate our classifications, we asked independent assessors on Mechanical Turk with no expertise or training in government to classify the functions of 8,574 agencies with our system: a random sample of agencies engaged in most common functions (schools, libraries, and municipal and county administrations) and every other agency. Agreement among individual assessors was high (κ = 0.89), and agreement with our classifications was 95%.

In short, the Internet footprint of US state governments provides a detailed and comprehensive picture of 47,631 agencies specialized to perform a wide array of functions that are easily identifiable to independent, untrained assessors, and the range of which exceeds the most comprehensive current national classification of government functions.

**Formal Mandates and Hierarchy in the Footprint**

To validate the aggregate reflection the websites and hyperlinks of these 47,631 public agencies offer of the bureaucratic structures of state governments, we first establish that it reflects two known structural characteristics of modern democratic government: formal legal mandates and hierarchical organization around principal–agent relationships.

To examine the degree to which state agency websites and hyperlinks reflect the structures mandated by law, we manually generated the functional structure mandated by the laws of one state, Massachusetts, and compared it to the structure reflected on the Internet in 2014. Both agency websites and the relationships among them clearly reflect the mandates of the Massachusetts General Laws (Fig. 1B). Of the 552 agencies mandated by law, 392 had websites; those that did not were largely small advisory groups and committees (for example, the state’s American and Canadian...
Similar to the connections mandated by law than the connections an agency connects to and from were randomly determined. Connections number of agencies and connections to and from each agency as doubt: We randomly generated 10,000 networks with the same number of agencies and connections to and from each agency. Thirty-six percent of these mandated connections are reflected in the mandated network have an average of four hyperlinks; and agencies connected directly in the mandated network have an average of 105 hyperlinks between them; those with two degrees of separation in the mandated network have an average of four hyperlinks; and those with three degrees of separation have fewer than one hyperlink on average between them (SI Appendix, Fig. S6.1). This overlap is the first indication that the online structure of government reflects its real-world functional structure systematically, if imperfectly. See SI Appendix, section 6 for detailed measures and tests with high-level organizational structures for three additional states, each showing meaningful but nonmandated connections among mandated agencies. Agencies engaged in other functions: international trade, homeland security, and ethics, lobbying, and campaign finance commissions. (See SI Appendix, section 5 and Table S5.1; ethics is also the least reciprocated function: just 6% of links from ethics agencies are reciprocated.) Fig. 2C displays online connections between functions, aggregated across all states, and Fig. 2D state-specific networks for Massachusetts, Georgia, and South Carolina. Fig. 2E depicts the hierarchy of the aggregate state government by organizing functions by betweenness centrality, with the most centrally connected function at the top and center. (For comparison, SI Appendix, Fig. S5.1B displays a random network lacking hierarchical organization.)

Altogether, we find that the online footprint of state governments, in addition to offering a substantially valid and unusually comprehensive reflection of the functions state governments implement, also reflects several known characteristics of the bureaucratic structures they form to do so, including accordance with legal mandates and hierarchical organization.

**How Much Do State Functional Structures Vary?**

The United States is a decentralized federation in which states have wide latitude to govern themselves differently. Their online footprints reflect considerable variation in their bureaucratic structures. Agencies engaged in research and science, international trade, homeland security, and ethics, lobbying, and campaign finance commissions.

First, the footprint suggests variation in what state governments do. Some functions, like research and science; international trade; and ethics, lobbying, and campaign finance commissions.

Second, the footprint suggests variation in what state governments do. Some functions, like research and science; international trade; and ethics, lobbying, and campaign finance commissions, including the governor, legislature, and county and municipal administrations. Agencies engaged in these functions link to and from clusters of agencies and departments responsible for other functions; these in turn link to and from the websites of agencies engaged in more specific functions, such as schools or hospitals. Several network science measures indicate that these interconnections strongly reflect a flow hierarchy, in which higher-order structures contain and connect to structures directly beneath, as when information and direction flows from managers to subordinates. For instance, 55% of interagency connections lead in only one direction, offering no path back to that agency or to other agencies engaged in the same function (22). (See SI Appendix, section 5 for additional tests.) We generated 10,000 networks with the same number of agencies and connections to and from each agency as the online footprint, but in which all connections are randomly distributed; on average only 5% of connections in these random networks were hierarchical (±2%). In addition, reciprocity—one agency connecting back to an agency that connects to it—is more common within than between most functions, suggesting clustering in functional communities. Only three functions have reciprocal connections more commonly with agencies engaged in other functions: international trade, homeland security, and ethics, lobbying, and campaign finance commissions. (See SI Appendix, section 5 and Table S5.1; ethics is also the least reciprocated function: just 6% of links from ethics agencies are reciprocated.) Fig. 2C displays online connections between functions, aggregated across all states, and Fig. 2D state-specific networks for Massachusetts, Georgia, and South Carolina. Fig. 2E depicts the hierarchy of the aggregate state government by organizing functions by betweenness centrality, with the most centrally connected function at the top and center. (For comparison, SI Appendix, Fig. S5.1B displays a random network lacking hierarchical organization.)

States’ online footprints also suggest wide variation in the bureaucratic structures they have formed to implement most functions. For instance, the diversity of implementation across states on three dimensions observable in the footprint: (i) the number of agencies engaged in each function, representing “scale” of implementation; (ii) the average number of pages per agency website, representing “density” of the activities around that function; (iii) the proportion of agencies engaged in the function that are connected by hyperlinks, directly or through other agencies, representing the function’s degree of “connectedness”; and (iv) overall diversity on all three measures. (See SI Appendix, section 7 for details, alternative measures, and visualizations of six functions of varying diversity.)

As government scales, it is thought to converge around common forms, as agencies learn from each other and develop common understandings of how to organize themselves effectively and legitimately (23, 24). The online footprint of the most common state functions, education and local administration, reflects this tendency: They are relatively more organizationally similar across states. Several of the rarest state functions are also more structurally similar.

### Table 1. The functions of state governments

| Administrative Law (4) | Historic Preservation (3) |
|------------------------|--------------------------|
| Audits, Accountability, and Inspectors General Licensing (22) | Boards and Professional Licensing (22) |
| Homeland Security International Trade | International Trade |
| Business Regulation (8) Judicial and Legal (6) | Commerce and Economic Development (8) Executive and General Administration (10) |
| Criminal Justice (6) Legislative | Data and Information (4) Military |
| Education (9) Public Safety (6) | Environment, Energy, Agriculture, and Natural Resources Campaign Finance |
| Research and Science (3) Social Services (10) | Labor and Human Resources (5) Tourism and Travel (2) |
| Financial Administration (8) Transportation (6) | Financing Authorities (5) Treasury and Revenue (2) |
| Health and Human Services (14) Utilities (4) | |
including international trade and ethics, lobbying, and campaign finance. However, most functions exhibit considerable organizational diversity across states, and even the most similar exhibit important relative differences. For example, the scale of state education systems ranges from 23 schools and other education agencies per million inhabitants in Texas to 573 per million inhabitants in Vermont; their average density from 230 pages per agency website in Hawaii to 2,361 in Wyoming; and the proportion of their connections that are with other education agencies from 89% in Kentucky to 49% in Hawaii.

An additional known characteristic of modern government is that it scales with population (17, 25). Although some state functions, such as the governor or the legislature, are typically performed by the same number of bodies in states large and small, many others involve localized implementation and face-to-face contact with citizens (18). In general, both online scale and density of state governments is greater in more populous states. But only two functions account for most of this scaling, schools and local administrations; their websites are more numerous, denser, and more connected in states with larger populations. For the other 26 functional categories, state population explains only 4% of cross-state variance in scale, 12% of density, and 4% of connectedness (SI Appendix, section 7 and Fig. S7.1). Even scaling in schools and local administrations is less than linear, suggesting economies of scale similar to those underlying growth in most complex social and biological systems (26, 27).

In short, we find substantial structural diversity among state governments. Although the most common functions tend toward structural similarity across states, most functions exhibit wide diversity in scale, density, and connectedness, variation that only partly reflects scaling with state population.

Why Else Do State Governments Differ?

As governments shape and respond to the people and places they govern, they come to reflect their environments. However, there is substantial debate about the factors that modern governments reflect. We focus here on four factors widely hypothesized to be associated with what governments do and how they work:

i) Ideology. Citizens differ in their preferences about the roles and extent of government, differences in part reflected in the liberal-conservative ideological continuum. Thus, ideologically similar states may have more similar governments (1–3).

ii) Economic Structure. Scholars have long noted a strong and likely strengthening (28, 29) relationship between government and the economy, as economic agents demand enabling goods...
and services from government for their industry or occupation (4–8) and government rules, regulations, goods, and services influence economic activity (9–12). Thus, economically similar states may have more similar governments.

iii) Income. Governments draw much of their revenue from their citizens, and the elasticity of their demand for government may vary across functions, such that state governments may be more similar when their citizens have similar incomes (13, 14).

iv) Location. The United States spans a large and varied natural environment, differences in which may have implications for what government does and how (15, 16). Location may also facilitate organizational learning and convergence around common forms, in general as well as for particular functions (23, 24).

Thus, similarly located states may have similar governments.

To investigate the relative role of these four factors, for every pair of states we calculate similarity in the (i) scale, (ii) density, and (iii) connectedness of state agencies implementing each of the 28 functional categories, as well as (iv) similarity across all three dimensions (n = 1225). Table 2 reports whether the structures of two state governments across the vector of all functions, normalized across states, are significantly more similar on these four measures when they are similar when their citizens have similar incomes (13, 14).

Fig. 4. The functional structures of state governments vary widely. A–C display the SD (σ) across states in the scale, density, and connectedness of each function. (D) Structural diversity among states for each function, normalized across the three dimensions, from bottom to top in ascending order of diversity. The SD (shaded) indicates the diversity of the function across states (see SI Appendix, section 7 for details and additional visualizations).

The explanatory power of economic structure far exceeds the others. How citizens in a state earn their living—measured by the proportion employed in grocery stores, doctors offices, coal mining, scientific research and development, or the other 311 industries measured by the Census (SI Appendix, section 8)—is strongly related to similarities in state governments on all four measures. Location and income play more limited roles. Location predicts similarities in the density and connectedness of state functional structures but not similarity in their scale. Income is weakly associated with similarity in scale and unrelated to similarity in either density or connectedness. Similarity in ideological preferences is unassociated with similarities in state functional structures on any of the four measures. (See SI Appendix, section 8 for detailed results and robustness checks.)

The relative strength of the association between economic and government structure may reflect state governments responding to their economies by providing complementary public goods and services (4, 6); shaping their economies, as capital and labor adjust to the enabling goods and services that the state’s government provides (10, 12); or both, as state governments and economies coevolve (7, 11). The cross-sectional correlations in our data cannot disentangle these dynamics. However, they strongly support the relationship itself. A few examples illustrate the overriding relative importance of state economic structures to structural similarities between state governments. Georgia and Massachusetts share similar governments and economic structures (top 4% and 19%, respectively), despite their physical separation and dissimilarity in ideology (bottom 1%) and income (bottom 13%); indeed, based on their online footprints, Georgia’s government is similar to Massachusetts’ than to its northern neighbor South Carolina’s (Fig. 2D). California and Florida have among the most similar economies of any two states (top 2%) as well as similar government structures (top 11%) despite differences in location, ideology (bottom 32%), and income (bottom 26%). Illinois and Indiana differ in average ideology and income (bottom 4% and 3% of other pairs of states, respectively) but share similar economic structures (top 1% of state pairs) and a border, and their governments appear online to have similar structures (top 2% of state pairs). Georgia and Virginia are nearby and have similar economies (top 12%), and their governments are among the top 2% in online structural similarity, despite being among the bottom 15% in ideological overlap and bottom 23% in income similarity.

In short, the strongest predictor of whether two state governments have similar functional structures is similarity in their economic structures, outweighing similarities in location, income, and the ideological preferences of voters.
Table 2. What explains similarities in the functional structures of state governments

| Factors          | Scale | Density | Connect. | Overall |
|------------------|-------|---------|----------|---------|
| Ideology         | 0.001 | −0.002  | −0.002   | −0.002  |
| Income           | 0.003*| −0.003  | −0.004   | −0.002  |
| Economic structure| −0.016**| −0.012**| −0.011* | −0.012**|
| Location         | −0.002| −0.007**| −0.009**| −0.006**|
| FE               | Y     | Y       | Y        |         |
| Obs.             | 1,225 | 1,225    | 1,225    | 1,225   |
| R²               | 0.91  | 0.54     | 0.44     | 0.53    |
| Adj R²           | 0.9   | 0.5      | 0.39     | 0.49    |

**P < 0.01; *P < 0.05. All models also include a control for network position. See SI Appendix, section 8 for bivariate results, variable definitions, and robustness checks. Connect., connectedness; Obs., observations.

Concluding Remarks

Like the states they govern, US state governments are enormously complex and multifaceted, limiting understanding of how they work. We develop a technique for mapping their functional structures using the extensive footprint that they leave online. Online reflections are always imperfect approximations of real-world phenomena (e.g., social networks on Facebook; ref. 30). However, we find strong evidence that in aggregate, these digital traces offer a comprehensive and valid reflection of the bureaucratic structures states have formed to implement their varied responsibilities, a reflection that can advance our understanding of those structures and the factors that explain them. The representation that emerges is recognizable in important ways, reflecting state-centric hierarchies that are functionally specialized, conform to legal mandates, and scale with population. We also find that interstate similarities in these structures vary significantly according to the industries in which citizens work and to a lesser degree with income and location, but not with voters’ recent ideological preferences about the proper roles and extent of government.

These patterns raise several questions. Does the relative similarity of education and local administration reflect convergence around an optimal form (10, 31)? Are economic and geographic conditions more important in determining government structure than citizens’ ideological preferences, or does ideology influence government structure more slowly, such that current structures reflect the preferences of voters from decades or even a century past (32–35)? Why is location not associated with the scale of state government as well as its density and connectedness? Why is income associated with scale but not density or connectedness? More broadly, what else has shaped the functional variation in state governments observable online? Our full model, including all four factors, explains 91% of similarity in the scope of each pair of state governments, but only 50% of similarity in density and 39% of similarity in connectedness. Inasmuch as the remaining variation reflects real differences in state governments rather than artifacts of website designs (SI Appendix, section 2), what explains it? Is the key factor forms of economic or organizational competition (9, 24, 31, 36) or organizational learning and emulation (23, 24, 37, 38) not captured in our measures, economic and industrial policies (12, 39), the nature or timing of industrial transformation (4, 20, 40), the accompanying transformation of social groups and relationships (41–45), vested interests (8, 46–48), inequality (29, 49, 49), shocks or critical junctures in the political, economic, or social environment (33, 34), all or some combination of the above, or factors yet to be considered? Finally, as agencies increasingly interact with businesses, interest groups, and other organizations online or in ways reflected online, these reflections may allow more systematic examination of agency relations with the societies they govern (7, 8, 11, 23, 29, 44).

These are the sorts of questions and puzzles we hope this window on government enables scholars to more closely explore. In addition, we expect the picture itself to improve over time, as more government activity and interaction with citizens is reflected online, and as future crawls permit a dynamic view on the structural evolution of government.

ACKNOWLEDGMENTS. This work was supported by the National Science Foundation Grant ICE-1216028. A.-L.B. was supported by Templeton Foundation Grant 61066.