Towards a network theory of regulatory burden

Jeremy Straughter* and Kathleen Carley

Abstract
Government overlap creates problems for the public and private sector across several government systems. Research suggests that overlap burdens the political and economic system by increasing costs in terms of money, time, and complexity, decreasing innovation and investment, and dampening economic growth. Previous research has identified frameworks for categorizing regulatory overlap. However, the literature has not considered factors that contribute to regulatory overlap nor effective methods for its reduction. This research introduces a proxy measure for regulatory overlap which measures the regulatory burden that a federal agency imposes within the regulatory system. This research proposes a network theory of regulatory burden which finds that burden is a function of network properties characterizing a federal agency’s position within the regulatory system. This research suggests that the federal government may reduce occurrences of regulatory overlap by managing its regulatory network and fostering collaboration among federal agencies within the regulatory system.

Keywords: Network science, Government regulation, Public policy, Public administration

Introduction
Several countries have expressed concern about the burdens imposed by excessive, fragmented, and duplicative regulations—a phenomenon which the literature has coined regulatory overlap. Public and private sector leaders have provided anecdotal evidence to suggest that overlapping regulations have a detrimental effect on the economy. Regulations carry administrative costs in terms of money, time, and complexity for which government and industry leaders must account for compliance purposes. Business Roundtable (2019) lists several inefficiencies that result from overlapping regulations which inhibit business investment and innovation, including conflicting policy guidance and the requirement to deconflict with multiple oversight bodies (6). Several governments have commissioned studies to better understand the burdens that regulatory overlap impose and recommend practical solutions to the problem (Government Accountability Office 2015; European Union 2014; Li 2015; Commonwealth of Australia 2014).

Research around government regulation spans several academic disciplines, including law, political science, public administration, and public policy. A major focus of these
disciplines is developing frameworks that measure or reduce legal complexity (Tullock 1995; Kaplow 1995; Epstein 2004). For example, Daniel Katz and Michael Bommarito conducted extensive research using network science to measure complexity in the United States (U.S.) Code—a compilation statutes enacted by the U.S. Congress (Katz and Bommarito 2014; Bommarito and Katz 2010). Likewise, researchers from the Mercatus Center at George Mason University have developed frameworks for measuring complexity in federal, state and local regulations; specifically, they created a regulations database (called RegData 3.0) which they used, in turn, to derive the Federal Regulations and State Enterprise (FRASE) Index, a ranking of the 50 states and District of Columbia according to the impact of federal regulation on private-sector industries in each state’s economy (Al-Ubaydli and McLaughlin 2017).

While the academic literature on legal complexity is abundant, there is far less research into regulatory overlap. This research considers overlap in U.S. federal regulations. The U.S. Government Accountability Office (GAO) identifies three categories of government overlap—fragmentation, duplication, and overlap (GAO 2015). Fragmentation refers to instances where more than one federal agency (or more than one organization within a federal agency) is involved in the same broad area of national need. Duplication occurs when the two or more agencies or programs are engaged in the same activities or provide the same services to the same beneficiaries. Overlap occurs when multiple agencies or programs have similar goals, engage in similar activities, or target similar beneficiaries. Previous research identified two categories of regulatory overlap—horizontal and vertical (Brown 1994). Horizontal overlap exists when two or more government agencies who operate on the same level, such as two or more federal agencies, are involved in the same regulatory activities. Vertical overlap exists when two or more government agencies who operate on different levels, such as two or more federal, state, and local agencies, are involved in the same regulatory activities.

The existing research into regulatory overlap has several limitations. The extant literature is entirely qualitative, most being case studies that involve a small number of government agencies (Brown 1994). Previous research established a typology for regulatory overlap but there is still much to learn about the factors that contribute to its existence. Further, there is insufficient research into effective strategies for reducing the occurrence of regulatory overlap. Our paper focuses on U.S. federal agencies and their respective federal regulations and contributes to research on horizontal overlap using network analysis of the U.S. federal regulatory environment. Our study models this environment as a complex network of shared interactions between federal agencies based on their co-sponsorship of federal regulations and shared keywords in their federal regulations. Our research proposes a proxy for regulatory overlap that measures the regulatory burden imposed by federal agencies within the regulatory network. Finally, we present a network-based theory of regulatory overlap that models the regulatory burden as a function of network measures that describe federal agencies’ position within the shared regulations and shared keywords networks. In the sections that follow, we ground our research on federal regulations data collected from the U.S. Federal Register database, define key terms and features used to conceptualize our network-based model, explain our research methods and results, and discuss the implications of our research and provide direction for future research on this topic.
Methods
Our research uses data collected from the U.S. Federal Register. The Federal Register is the central repository of daily federal government activity, most important of which are the final regulations that we use as the primary data source. Final regulations are enacted by federal government agencies following a strict process that is codified in the law and has U.S. congressional oversight; as a result, final regulations have the same authority and force of law as any other congressional statute (Davidson et al. 2018). Beginning in 1993, the U.S. federal government began to catalog this daily activity in an online database. We began this research with three primary data sources. First, we used metadata about final regulations published from January 1, 1993, until December 31, 2019. The metadata consisted of twenty-six files which compiled the regulatory activity for each year formatted in java script object notation (JSON). Second, we used a data set that provided additional information about each U.S. federal agency within the register, notably its acronym, official name, the name of its parent agency, and the name of its subordinate agencies. The federal agencies in our data have one of three primary classifications—department, subordinate, or independent. Executive departments are generally the largest federal agencies, managed by politically appointed public officials (called secretaries) who serve on the president’s cabinet; there are fifteen executive departments. Subordinate agencies are those federal agencies who are situated under some other agency in the federal hierarchy. Independent agencies are those federal agencies who are not situated under any other agency in the federal hierarchy but are not executive departments. The leaders of independent agencies (except for the attorney general) are not cabinet members, which distinguishes them from executive departments. Finally, we used the raw text of each federal regulation published in the register from January 1, 1993, until December 31, 2019.

We used a social network analysis and visualization software developed at Carnegie Mellon University called Organizational Risk Analyzer (ORA) to formulate semi-structured networks from the unstructured primary data sources (Carley et al. 2013). The resulting networks included adjacency matrices which mapped the organizational structure of the U.S. federal executive branch, mapped each U.S. federal agency to its respective regulation, and mapped each federal regulation to its respective keywords. We used a Python natural language processing library to pre-process, tokenize, and filter keywords which were classified as nouns or proper nouns within each regulation to derive the network mapping regulations to keywords. There were three node sets in our network: 453 U.S. federal agencies; 99,014 federal regulations; and 266,666 keywords.

Figure 1 shows the distributions of total regulations for each federal agency in the network. The minimum value was one regulation, the maximum value was 24,744 regulations, the median value was twenty-two regulations, and the mode was one regulation. The average was 522 regulations, but since this value is heavily skewed towards the maximum value, the median is a more accurate measure for this network. Figure 2 shows the distributions of federal agencies for each federal regulation in the network. Nearly 98% of the regulations involved two agencies or fewer. The minimum value was one federal agency, the maximum value was forty-three federal agencies; the mean, median, and mode were each two federal agencies. Of the federal regulations that involved two or
fewer agencies, roughly forty percent involved a single agency and the remaining sixty percent involved two agencies.

Our network approach is advantageous because it accounts for the complexity that arises from interdependence among federal agencies when enacting federal regulations. This is a major advantage over the qualitative approaches which are prevalent in the existing research. Our approach captures the macro-level behavior of the entire executive branch, which generalizes our theory to the entire federal system. We projected the bipartite network mapping federal agencies to their respective regulations into a one-mode network that captured the shared regulations (co-sponsorship) relationship between federal agencies. Bipartite network projection is a frequently used method in network science for measuring the level of interdependence that exists among nodes in a network (Neal 2013). This method has been successfully used to theorize about legislative networks and congressional statutes (Kirkland and Gross 2014); however, to the best of our knowledge, our research is the first to use this method for theorizing about federal agencies and federal regulations.

Table 1 provides summary statistics for the shared regulations network. The shared regulations network was sparse, having a density of 0.014, suggesting that U.S. federal
agencies do not collaborate when enacting regulations, as only 1% of potential links were present. There were 222 isolates in the network, meaning that roughly 51% of federal agencies had not collaborated with any other agencies to enact a regulation. There were two dyads and one triad, which means that two pair of federal agencies only collaborated among themselves and one group of three federal agencies only collaborated among themselves. Finally, there was one large component of federal agencies which constituted the preponderance of regulatory activity in the network and required further analysis.

We calculated several network measures to better understand the activity taking place in the large component of the network. Table 2 includes the name and definition for each of the network measures considered in this paper. Figure 3 shows the distribution of total degree centrality based on the binarized shared regulations network, which reveals the total number of other agencies within the federal system with whom each agency collaborated. Most federal agencies in the network collaborated with between zero and three other agencies. The distribution of total degree centralities approximates a power-law that reflects a scale-free topology, supporting prior research on the topology of real-world networks (Barabasi 2009), where a small number of agencies account
for most of the regulatory activity (so-called hubs) and lower-degree agencies connect with them by preferential attachment. The weighted values would then show the number of collaborations that occurred between agencies, which would serve as a proxy for the strength of the relationship that exists between two federal agencies. These findings were meaningful, potentially having a significant effect on regulatory overlap, requiring further investigation.

Using the metadata on each U.S. federal agency, the network mapping of the U.S. federal government executive branch, and bipartite network projection, we created four additional networks to investigate the effect of shared organizational type and the parent–child relationship on the regulatory activity that we observed in the network. These additional adjacency matrices mapped each federal agency to other agencies with whom there was a shared organizational type (i.e., department, subordinate, independent) or a parent–child relationship). Since the shared regulations, shared organizational type, and parent–child relationship networks are dependent in nature and violate the assumptions of traditional regression models, we used the multiple regression quadratic assignment procedure (MRQAP), a network science method for measuring the effects of shared relationships which are inherent in network data (Krackhardt 1988).

The MRQAP algorithm correlates adjacency matrices by reshaping them and calculating ordinary tests of statistical association on the re-shaped matrices (this is the so-called observed correlations). MRQAP addresses dependence in network data by permuting one of the adjacency matrices, meaning that it randomly rearranges rows and columns of the matrix, which results in a matrix that is independent of the original matrix but maintains its properties. The algorithm determines the significance of the observed correlation by comparing it to a reference set of correlations based on the permuted matrices and assigns a p value by counting the proportion of correlations among permuted matrices which were as large as the observed correlation. A p value of less than 5% constitutes a significant relationship, and several permutations is used to stabilize the p value and reduce its variability (we used 20,000 permutations for this experiment). Table 3 lists the MRQAP results. The shared department relationship was significant with a regression coefficient of 9.618 and a p value of 0.012. The parent–child relationship was also significant with a regression coefficient of 253.492 and a p value of zero. These results
suggest that interdepartmental collaboration accounts for part of the regulatory activity we observe in the shared regulations network; however, the parent–child relationship is an order of magnitude larger and suggests that most shared regulations occur between a parent agency and its subordinate agency.

A preferred method of detecting overlap between federal regulations requires projecting our bipartite networks into one-mode networks that capture their shared agency and shared keyword relationships. The disadvantage of bipartite network projection is that it does not scale well for large and dense networks, which was the case for our network mapping regulations to keywords. Instead, we considered ways to reduce the computational requirements associated with the federal regulations network. To accomplish this, we multiplied the networks that mapped federal agencies to federal regulations and federal regulations to keywords to create an additional network that mapped federal agencies to the keywords found in their respective federal regulations. The resulting network offered two advantages—reduced computational complexity and it enabled us to consider a proxy measure for regulatory overlap that models the regulatory burden imposed by each federal agency. We based our measure on the two networks which mapped federal agencies to federal regulations and federal agencies to keywords. From these networks, we calculated the total regulations, exclusive regulations, shared regulations, total keywords, exclusive keywords, and shared keywords for each federal agency within the network (Fig. 4).

We modeled the regulatory burden imposed by each federal agency as its number of shared keywords weighted by its proportion of exclusive regulations to total regulations (we also rounded this result and classified it as an integer for convenience—this change had no effect on the results of our statistical model). A federal agency’s regulatory burden may range from zero to the total number of keywords in the data set and the burden is largest when both the number of shared keywords and exclusive regulations are large. To capture the number of shared keywords for each agency, we measured their out-degree centrality and row exclusivity in the network mapping federal agencies to keywords. In this network, a federal agency’s out-degree centrality represents the total

| Variables | Coef  | Std. coef | Std. errors | Robust std. errors | Bootstrapped std. errors | Sig. Y-perm |
|-----------|-------|-----------|-------------|--------------------|--------------------------|-------------|
| Intercept | 0.178 | 0         | 0.288       | 0.204              | 1.000                    |             |
| X1        | 270.395 | 0.181    | 4.600       | 4.350              |                          |             |
| X2        | 11.879 | 0.006    | 6.562       | 6.966              |                          |             |
| X3        | −0.097 | −6.127e−04 | 0.514     | 0.561              |                          | 0.484       |
| X4        | −0.201 | −0.001   | 0.543       | 0.584              |                          | 0.547       |
| X5        | 0.911  | 0.002    | 1.667       | 1.703              |                          | 0.086       |
number of keywords that an agency has used in its regulatory corpus, while its row exclusivity represents the number of exclusive keywords—the number of keywords to which the federal agency is connected that have a total degree centrality of one, meaning they are not connected with any other federal agencies. The number of shared keywords is the difference between the total keywords and the exclusive keywords. Likewise, in the network mapping federal agencies to federal regulations, the out-degree centrality and row exclusivity represent the total regulations and exclusive regulations, respectively.

Our analysis of the U.S. federal regulations network provides the foundation for a network theory of regulatory burden. The primary assumption of our theory follows directly from prior research, namely that there exist duplicative, fragmented, and overlapping regulations which burden public and private stakeholders. A second assumption of our theory is that the U.S. federal government can reduce its regulatory burden by managing regulatory overlap, specifically horizontal overlap. Given these assumptions, we propose the following theory—that regulatory burden is a function of a federal agency’s position within the regulatory network as modeled by network features derived from that system. Figure 5 shows the distribution of each federal agency’s contribution to the regulatory burden. Our measure of regulatory burden is based on each federal agency’s count of shared keywords, weighted by its proportion of exclusive regulations, and best represented by a Poisson probability distribution (Fig. 6).

The Poisson distribution includes the set of all non-negative integers (e.g., count data), although our data set is technically bounded by the total number of keywords. The Poisson regression model is a generalized linear model that is used for modelling count data. The model assumes that response variables follow the Poisson distribution, have a positive mean (uses the exponential to enforce this positivity requirement), and have equal mean and variance. We used R to generate a quasi-Poisson regression model, which relaxes the requirement for equal mean and variance, yet still produces accurate estimates for the model’s coefficients (Dunn and Smyth 2018).

Results
Figure 7 depicts our Poisson regression model. The intercept estimate in our model establishes a baseline measure of regulatory burden which includes that of executive departments. The remaining values in our model are independent variables, whose

Let Matrix A represent the network mapping federal agencies to federal regulations.
Let Matrix B represent the binarized network mapping federal regulations to keywords.
Let d = column sum (Matrix B) be the degree of each column node (e.g., keywords).
Total Regulations = row sum (Matrix A)
Exclusive Regulations = row sum (Matrix A where (d = 1))
Shared Regulations = Total Regulations – Exclusive Regulations
Total Keywords = row sum (Matrix B)
Exclusive Keywords = row sum (Matrix B where (d = 1))
Shared Keywords = Total Keywords – Exclusive Keywords

Regulatory Burden = (Exclusive Regulations / Total Regulations) * Shared Keywords

Fig. 4 Regulatory burden
estimates measure the change in burden for a one unit increase in the variable, while leaving the remaining variables constant; we exponentiate the regression estimates to determine their effect on the response variable. The burden imposed by independent agencies were 120% higher than baseline, on average, while that imposed by subordinate agencies were 92% lower. Increases in effective network size are associated with a 5% decrease in burden from the baseline, while total degree centrality also reduces burden by a small amount. Increases in clique count and triad count were associated with minor increases to burden within our model. The ego-betweenness and
Discussion

We have premised our network theory of regulatory burden on our statistical model, which uses regulatory burden (a proxy measure for regulatory overlap) as its response variable, and models regulatory burden as a function of each federal agency's organizational type, clique count, triad count, effective network size, betweenness, eigenvector, and total degree centralities. U.S. federal agencies may be categorized into one of three primary organizational types—executive departments, subordinate agencies, or independent agencies. In general, executive departments are the largest of the three organizational types and answer directly to the executive office the president. Most subordinate agencies are aligned under departments, but a small number are aligned under independent agencies. Independent agencies vary in size, and, unlike executive departments, independent agencies do not answer directly to the executive office of the president.

Our research found that the majority of U.S. federal regulatory activity occurs between executive departments and their respective subordinate agencies. Each federal agency manages a particular functional area, including regulations within that functional area. However, there are certain policy issues which require collaboration across functional areas. Our research found substantial interdepartmental collaboration, albeit to a lesser degree than intradepartmental collaboration. There was not a significant amount of collaboration among subordinate agencies (neither intradepartmental nor interdepartmental) nor among independent agencies. Moreover, we found that organizational type had a significant impact on the level of burden in the regulatory system. By our measure, independent agencies impose the highest burden on the system (significantly higher than departments) while subordinate agencies impose the lowest burden on the system. We discovered that higher connectivity (e.g., effective network size and total degree centrality) was associated with reduced regulatory burden within our model. We also observed that being connected to highly connected agencies (e.g., clique count, triad count, ego-betweenness, and eigenvector centrality) is associated with regulatory burden within our model.

Our findings suggest that the federal government can reduce the regulatory burden by attending to the network properties of federal agencies within the regulatory system. Several recommendations follow from this analysis. First, the federal government should make greater use of working groups for enacting regulations. Based on the large percentage of regulations enacted with two or fewer federal agencies, working groups will increase the network connectivity as well as the network properties of agencies within the system. The government should identify which policy areas require collaboration and which federal agencies are stakeholders in those areas. Second, the government should consider consolidating federal agencies with large numbers of exclusive regulations and who significantly burden the regulatory system. For instance, consolidating certain independent agencies under departmental leadership would constrain their actions and could reduce the overall burden on the system. Based on our data, departments are inherently brokers, so by consolidating certain independent agencies under the direction of a department, this would
reduce their number of exclusive regulations. Combining this action with greater use of working groups would ensure that certain policy areas remain a priority.

An advantage of this theory is that by focusing on the network properties of entities in the regulatory process, we lay the foundation for future research to generalize this theory to multiple levels of government or political systems. A limitation of this theory is that it does not account for regulatory overlap directly and incorporating the total number of overlapping regulations may allow us to strengthen this model. Also, we considered a limited number of network measures as independent variables in our model, but additional network features could be included to learn more about the behavior of this system. In future research, we will consider computationally efficient methods for dealing directly with large-scale regulations data and discovering the community structure within. In this way, we hope to examine horizontal and vertical overlap in greater detail and better understand this phenomenon.

**Conclusion**

This research approached the issue of regulatory overlap using network science to model the regulatory burden imposed by U.S. federal agencies on the regulatory system. We propose a network theory of regulatory burden, which posits that regulatory overlap is a function of certain network properties of federal agencies within the system. The theory is bounded on U.S. federal agencies and grounded on U.S. federal regulations data taken from the U.S. federal register database. This research has established the foundation for a network theory that may generalize to multiple levels of government and political systems after consideration of additional data. Our research constitutes an improvement upon existing literature in the field, which up until now, has only proposed frameworks and considered single policy areas using qualitative methods.

**Abbreviations**

JSON: Java script object notation; MRQAP: Multiple regression quadratic assignment procedure; U.S.: United States; ORA: Organizational risk analyzer.

**Acknowledgements**

Not applicable.

**Authors’ contributions**

We attest that all authors contributed significantly to the creation of this manuscript. All named authors had substantial contributions to the conception of the work, the acquisition, analysis, and interpretation of data for the work. All authors wrote, read and approved the final manuscript.

**Funding**

Not applicable.

**Availability of data and materials**

The data sets used during the current study are available from the corresponding author on reasonable request.

**Declarations**

**Competing interests**

The authors declare that they have no competing interests.

Received: 26 June 2021   Accepted: 27 August 2021
Published online: 22 September 2021
References
Al-Ubaydli O, McLaughlin PA (2017) RegData: a numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012. Regul Gov 11(1):109–123
Barabasi AL (2009) Scale-free networks: a decade and beyond. Science 325:412–413
Bommarito MJ, Katz DM (2010) Properties of the United States code citation network
Brown G (1994) Canadian federal-provincial overlap and presumed government inefficiency. Publius J Fed 24(1):21–38
Business Roundtable (2019) Reducing regulatory overlap in the 21st century
Carley et al (2013) Users guide 2013, Carnegie Mellon University, School of Computer Science, Institute for Software Research. Technical report, CMU-ISR-13–108
Commonwealth of Australia (2014) The Australian government guide to regulation
Davidson RH et al (2018) Congress and its members. CQ Press, Washington
Dunn PK, Smyth GK (2018) Generalized linear models. Springer, Berlin
Epstein R (2004) The optimal complexity of legal rules. University of Chicago Law & Economics, Olin Working Paper No. 210
European Union (2014) Gaps, overlaps and challenges: a landscape review of EU accountability and public audit arrangements
Government Accountability Office (2015) Evaluating and managing fragmentation, overlap and duplication
Kaplow L (1995) A model of the optimal complexity of legal rules. J Econ Organ 11:150
Katz DM, Bommarito MJ (2014) Measuring the complexity of the law: the United States code. Artif Intell Law 22:337–374
Kirkland JH, Gross JH (2014) Measurement and theory in legislative networks: the evolving topology of congressional collaboration. Soc Netw 36:97–109
Krackhardt D (1988) Predicting with networks: non-parametric multiple regression analysis of dyadic data. Soc Netw 10:359–381
Li X (2015) An economic analysis of regulatory overlap and regulatory competition: the experience of interagency regulatory competition in China’s regulation of inbound foreign investment. Adm Law Rev 67(4):685–750
Neal Z (2013) Identifying statistically significant edges in one-mode projections. Soc Netw Anal Min 3:915–924
Tullock G (1995) On the desirable degree of detail in the law. Eur J Law Econ 2(3):199–209

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