Exploiting Network Structure to Improve Government Performance

Jeremy Straughter

CMU-ISR-22-105
April 2022

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair
L. Rick Carley
Michael I. Shamos
Douglas C. Sicker (University of Colorado)
Iain J. Cruickshank (U.S. Army)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Societal Computing

Copyright © 2022 **Jeremy Straughter**

Keywords: Public policy, government regulation, network science, public interinstitutional collaboration, performance measurement, public affairs, social media

Abstract

This thesis examines the public interinstitutional collaboration among U.S. federal agencies and associated performance measures. Public interinstitutional collaboration refers to formal or informal interaction between formally autonomous institutions of equal status to reach one or several policy goals. Research suggests that this form of collaboration is a viable alternative for addressing complex, unpredictable, and intractable public policy issues whose solutions are beyond the capacity of any single public organization. The fact that public organizations operate with such high levels of autonomy constrains public interinstitutional collaboration and performance.

Previous research found that frameworks for public interinstitutional collaboration performance measurement were lacking, and recommended performance measurement approaches tailored to specific collaborative contexts, given the fact that each policy issue has specific desired outcomes, boundaries, and stakeholders. I propose that public organizations can improve their performance by exploiting their network structure. Using social network analysis to examine public interinstitutional collaboration within the context of government regulation and strategic communication, I will demonstrate how public organizations can use network methodology to establish performance measures and subsequently assess how network properties affect performance. Specifically, this thesis comprises five

studies that investigate how governments may reduce the regulatory burden, reduce regulatory overlap, minimize regulatory complexity, combat misinformation, and improve customer satisfaction.

My research will contribute to the extant literature by creating performance measures for public interinstitutional collaboration; creating context-specific frameworks which directly incorporate network methodology into the public policy process; identifying latent and dynamic networks that exist within regulatory and strategic communication environments and examining their effect on the desired outcomes; solving complex policy issues which affect several public organizations but are unsolvable by any single organization; creating novel data sets for future research in the disciplines of law, public policy, public relations, and social cybersecurity.

Acknowledgements

I would like to thank Dr. Kathleen Carley for patiently guiding me through the doctoral process. I am grateful to have had her advice and support while my attending the university. Additionally, I would like to thank Drs. Rick Carley, Michael Shamos, Doug Sicker, and Iain Cruickshank for joining my thesis committee and allowing me to complete this milestone.

I have been fortunate to learn from the brightest minds in the computer science discipline. Thanks to all the professors who took the time to share their thoughts and experience with me, of which there are too many to name here. Thanks to the administrators within the Institute for Software Research; and a special thanks to Sienna Watkins, Connie Herold, and my CASOS research group.

The research reported in this thesis has been supported in part by the Knight Foundation, the Center for Computational Analysis of Social and Organization Systems (CASOS) and the Center for Informed Democracy & Social - cybersecurity (IDeaS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the funding agencies.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Background

“A roadmap to the future” is a study that was published in February 2020 by The Partnership for Public Service, which is a nonprofit organization that teams up with federal agencies and other stakeholders to make our government more effective and efficient. Their key finding was that federal agencies must have better collaboration internally and with external stakeholders, as well as do a better job with public engagement. The academic literature coins the phrase public interinstitutional collaboration to refer to formal and informal interactions between government agencies. Research suggests that this form of collaboration is key to addressing wicked policy problems, which are interdisciplinary and unsolvable by a single organization alone.

Public interinstitutional collaboration refers to the interaction between formally autonomous institutions of equal status or those at different levels of government that collaborate to reach one or several policy goals (Costumato, 2021). This form of collaboration has received increasing attention from academia as a viable solution for addressing wicked policy problems. Wicked problems are complex, unpredictable, and intractable problems whose solutions are beyond the capacity of any single public organization (Head and Alford, 2015). Research suggests that a collaboration enhances our understanding of wicked problems and their potential solutions (Huxham and Vangen, 2005).

Given that public organizations operate with such high levels of autonomy, interinstitutional collaboration and performance management is complicated (Provan et al., 2007). The performance management of public organizations centers on the traditional, vertical hierarchies of government departments where the standards are developed within each individual organization (Ryan and Walsh, 2004). Public organizations at the same hierarchical level often consider different goals and ignore their interdepence with others (Borgonovi et al., 2018).

Public interinstitutional collaboration may occur at different stages of the public policy process, ranging from strategic to operational issues, through formal collaborative mechanisms or informal networks. It may occur horizontally, meaning different institutions of equal status collaborate to achieve common goals, or vertically, meaning institutions at different, often hierarchically ordered, levels of government collaborate. There are two main literature streams associated with public interinstitutional collaboration and related performance management – collaborative governance and networked administration; these streams are guided by different theoretical frameworks.

Theoretical Foundation

Collaborative governance refers to the process and structures of public policy decision-making and management that engage people across the boundaries of public agencies and levels of government, and public, private, or civic spheres to carry out a public purpose that could not otherwise be accomplished (Emerson et al., 2012). It provides useful insights when interdependence involves few organizations who collaborate to reach a shared goal; particularly, the managerial practices that facilitate a better relationship among organizations. However, the theory fails to identify and analyze drivers that foster performance, and whether, when or how collaboration improves outcomes.

Networked administration denotes structures of interdependence involving multiple organizations where one unit is not merely the formal subordinate of the others in some larger hierarchical arrangement (O'Toole, 1997). The network is central to networked administration theory, but only partially relates to collaborative governance theory (Favoreau et al., 2016). The networked administration literature has analyzed the main drivers of network performance in different policy areas and contexts, including network structure, level of trust, management behavior, management strategy, and resource distribution. The literature has also considered network performance from multiple perspectives: single organization perspective, entire network perspective, and beneficiary perspective. Networked administration theory assumes a formalized integration between public organizations as an alternative to failures in hierarchical and market types of collaboration; however, this precondition rarely corresponds to the type of collaboration that occurs between autonomous institutions (Costumato, 2021). The theory fails to analyze which factors may also be valid in measuring public interinstitutional performance.

Thesis Proposal

Several gaps exist within the literature on public collaboration. Previous research identified that only a small number of papers dealing with public interinstitutional collaboration also analyzed how this form of collaboration can be effective. The literature does not provide comprehensive frameworks or tools to measure output and policy outcomes deriving from public interinstitutional collaboration. It recognizes that defining a unique performance framework is an arduous task, and therefore recommends performance measurement approaches tailored to specific collaborative contexts, given the fact that each policy has specific needs, boundaries, and various actors.

This research examines public interinstitutional collaboration and performance measurement in two collaborative contexts which the government considers wicked yet have not been previously addressed in the literature - government regulation and strategic communications. Further, it proposes that a government may improve its performance by exploiting its network structure. The United

States (U.S.) federal agencies constitute a boundary for the thesis; as a result, it will consider horizontal public interinstitutional collaboration (also identified as a gap in the literature). The thesis will not consider vertical collaboration (state and local government), international collaboration, nor non-governmental entities.

Data Sources

Figure 1 Proposed data by chapter

Data	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
Federal Regulations	✓	✓	✓		✓
Federal Regulations Metadata	✓	✓	✓		✓
Federal Agency Metadata	✓	✓		✓	✓
Federal Agency Twitter Data				✓	✓
Customer Satisfaction Surveys					✓

Figure 1 shows how data sources will align with each chapter in the thesis. It will be grounded upon primary and derived data associated with federal agencies, including data from federal regulations, social media, and customer satisfaction surveys. The primary data used for this thesis include:

- Federal regulations, 1993 – 2019
- Federal regulations metadata, 1993 – 2019
- Federal agency metadata
- Federal agency Twitter data
- Government customer satisfaction survey, 2019

Research Questions

The thesis will answer the following research questions:

- Chapter 2. How can we measure, assess, and benchmark the regulatory activity of governmental entities?
- Chapter 3. How can we detect regulation overlap within the regulatory system?
- Chapter 4. How can we measure the level of burden that governmental entities impose within the regulatory system?
- Chapter 5. How can we measure, assess, and benchmark the policy dialogue of governmental entities on social networking platforms?

Chapter 2: Structural Analysis of the U.S. Federal Regulations Network

Introduction

This chapter considers the rulemaking process of the United States' federal government from a network perspective and uses social network analysis to derive insights about rulemaking that may not be readily apparent when applying traditional tools of public policy analysis. There is no well-defined definition of public policy nor is there a well-established process for its development and implementation. The approaches to policy management vary based on several factors, including the country and its form of government and level the government being considered – federal, state, or local government. There is an entire field of public policy – comparative policy analysis – that is devoted to understanding these differences and is beyond the scope of this paper. This chapter broadly defines the policy environment as a system of laws, regulations, budgets, and court decisions (which we explain in greater detail in a subsequent section).

The policy making process may occur in four distinct stages – agenda setting, policy formulation, policy adoption, and policy implementation (Davidson et al., 2014). The process begins with Congress setting the legislative agenda, then formulating policies within its committees and subcommittees, and finally adopting policy by majority vote. Federal agencies implement legislative decisions using a subsequent rulemaking process for which we provide background in the section that follows. The rulemaking process does not occur within a vacuum; rather each agency serves a purpose to ensure that the federal government functions effectively.

The rulemaking process requires interdepartmental coordination within a particular functional area as well as interagency coordination when policies impact multiple interests. Congress may draft statutes that require agencies to collaborate on particular policy issues, but in many cases, this collaboration naturally occurs during the course of operations. This paper considers the so-called regulatory networks which might emerge within the federal government as a result of interagency coordination. To the best of our knowledge, this is the first work to consider the network structure that exists within the executive branch of the federal government.

The purpose of this research is to capture an underlying network of federal agencies based on their co-occurrences (or collaborations) on federal regulations within the federal register database. This research identifies the federal agencies who enacted the most regulations in the regulatory network, the federal agencies with the greatest centrality based on total degree and betweenness, and the variables which have the greatest contribution to the regulatory network structure. Lastly, this research seeks to understand how collaborative federal agencies are when enacting regulations and which policy areas required the greatest amount of collaboration, which may be deemed as having the most important (or most complex) policy issues.

Background

The Policy Environment

The policy environment is a system of laws, regulations, budgets and court decisions that span multiple levels of government and involve both internal and external stakeholders. The iron triangle provides a visual representation of the policy environment and describes the relationships that exist between legislative committees, subcommittees, or commissions having legislative authority; government bureaucrats holding senior positions in executive agencies; and interest groups representing individuals, corporations, and non-profit organizations. The political science literature also refers to them as subgovernments based on the reality that small groups of actors dominate certain sectors of the political system by organizing into mutually reinforcing relationships (Davidson et al., 2014; Hayden, 2002; Heclo, 1978).

The legislative branch represents the will of its constituents throughout the policy making process. Congress is responsible for setting a legislative agenda, formulating policy through rigorous debate, and voting on whether proposed policies are adopted into law. Congress is bicameral, consisting of two policy making bodies, the House of Representatives and the Senate. The House and Senate are further divided into committees and subcommittees having legislative and oversight authority over specific policy areas.

The executive branch enforces the law. It is comprised of various components, including the executive office of the president, executive departments and their subordinate agencies, independent agencies, commissions, boards, and committees. Federal agencies are aligned with congressional committees and subcommittees for their statutory authority, budget and oversight. Interest groups assert influence on both branches by keeping a watchful eye over legislative activities and federal agencies. They consider the effects of legislation on their constituents and the performance of federal programs. Finally, interest groups seek to influence statutes by making recommendations for legislative language, a process commonly referred to as lobbying.

The Rulemaking Process

Congress establishes federal agencies, defines their legal mandates, funds and provides oversight of their activities. Federal agencies have legislative authority to draft the rules and regulations (these terms are interchangeable) necessary to fulfill their lawful mandates. Congress is often unable to specify the details needed to effectively implement programs and policies; therefore, they delegate this task to the executive branch. Final regulations have the full force of legislative statutes;

using their regulatory authority, federal agencies enact more laws each year than the Congress (Davidson et al., 2014).

Federal agencies follow standards established in the Administrative Procedure Act (APA) when drafting regulations. The APA requires federal agencies to publish a notice of proposed regulations in the federal register. The federal register is a daily publication of regulatory activity. Notices explain the terms whereby interested parties may comment on proposed regulations; the public comment period typically runs for at least thirty days and allows stakeholders to comment on the merits or demerits of a proposed regulation (5 U.S.C. §551).

The Congressional Review Act (CRA) specifies expedited procedures whereby Congress may nullify any regulation proposed by a federal agency. Under the CRA, federal agencies must submit each proposed regulation considered major – having an economic impact of $100 million in a single year – to the House and Senate for review. Congress has sixty days from the time that a proposed regulation is posted in the federal register to enact a joint resolution of disapproval. The president may veto the disapproval resolution and Congress will need a two-thirds majority to override the president's veto. Once a regulation has posted in the federal register as final, the executive branch may not repeal it without following the guidelines specified in the APA. Federal courts ensure that agencies do not repeal federal regulations by fiat (5 U.S.C. §802).

Data

The data for this research comes from the Federal Register database. The Federal Register is the federal government's central regulatory publishing system and is managed by the Office of the Federal Register. Enacted by the Federal Register Act in 1935, the Federal Register provides an efficient way for citizens to stay abreast of the regulations that affect them (44 U.S.C. §15). In addition to regulations, the Federal Register publishes presidential proclamations, executive orders, notices, and other documents that are required to be published by statute (Carey, 2013).

Since 1993, the federal government has maintained the Federal Register electronically. We collected metadata on each regulation published as a final rule during the twenty-six-year period from January 1, 1993 to December 31, 2019. Our initial data included the title of the regulation and the federal agencies who collaborated to enact the regulation. We uploaded the data into a powerful network analysis application called ORA-Pro, which is used for network analysis, visualization and forecasting. This resulted in a two-mode affiliation matrix that connected each federal agency to the regulations it sponsored as well as its associated bipartite graph structure consisting of 424 federal agencies and 63,065 federal regulations. The total regulatory activity and bipartite network statistics is depicted in figure 1 and table 1, respectively.

Figure 2 Federal regulatory activity from 1995 - 2019

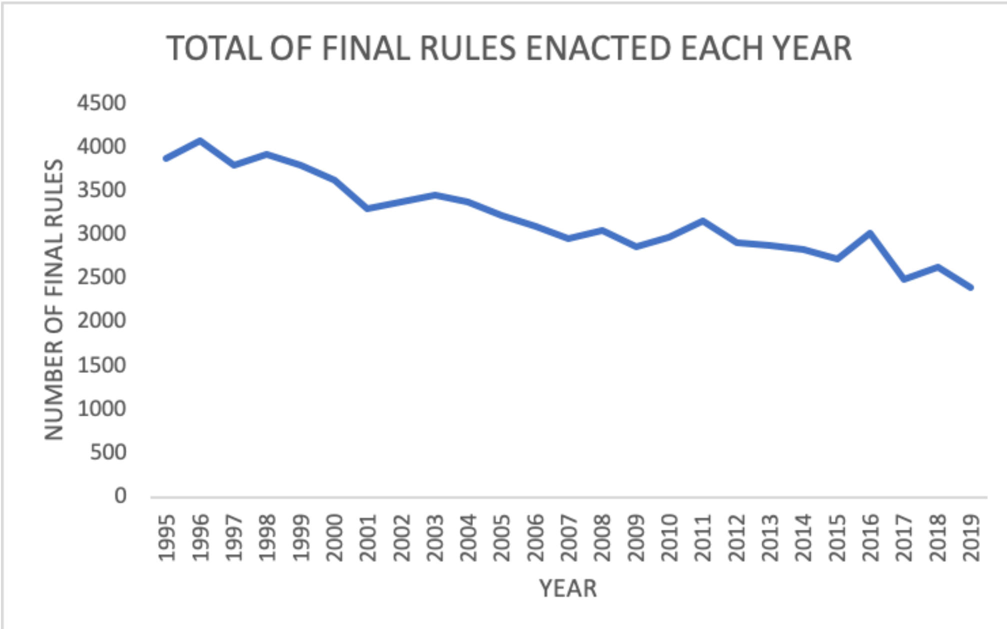


Table 1 Agency by regulation network statistics

General Statistics	
Total Federal Agencies:	453
Total Regulations:	99,014
Density:	0.004
Link Statistics	
Total Links:	162,632
Link Values:	Min: 0; Max: 4; Mean: 1; Std Dev: .05; Sum: 162,955
Component Statistics	
Isolates:	265
Dyads:	16
Triads:	14
Larger:	55
Larger Sizes:	Min: 4; Max: 96,304; Mean: 1,802.33; Std Dev: 12,860.53

Methods

Social Network Analysis

Social network analysis facilitates our understanding of complex social systems by focusing on the relationships that exist among the various entities within them. These systems are referred to as networks in the literature and we may use graphs to represent networks mathematically. Graphs consist of sets of vertices and edges – vertices are commonly referred to as nodes and edges are commonly referred to as links or ties. Links connect node pairs in the network – when two nodes are joined by a link, we say that the nodes are adjacent. Each node has attributes that distinguish it from other nodes within the system. The attributes that characterize nodes and the links that characterize relationships between nodes may take either an ordinal or continuous value (Wasserman et al., 1994).

Graphs may be directed or undirected. Directed graphs are used to represent relationships that have a logical sense of direction – for example, node A gives advice to node B. Undirected graphs are used to represent relationships where no logical direction exists or where the relationship is reciprocated – for example, node A is kin to node B. When a link connects node A and node B, and another link connects node A and node C, we say that the two links are incident upon node A. The number of links incident on a node is referred to as its degree. Nodes that have zero links in a system are called isolates. Nodes having only one link in a system are called pendants (Wasserman et al., 1994). In social network analysis, it is conventional to organize graphs so that each link represents the same social relationship. In general, we expect each social relationship to exhibit a different network structure and have different implications for the nodes involved (Borgatti et al., 2018; Wasserman et al., 1994).

The adjacency matrix is another tool for conceptualizing networks mathematically. The rows and columns of the adjacency matrix represent nodes, where an entry in row i and column j represents a link from i to j. For the adjacency matrix A of a non-valued graph a_{ij} = 1 if a link exists from i to j and a_{ij} = 0 otherwise. The direction of the relationship matters in network analysis; by convention, the direct goes from the rows to the columns. Additionally, for a graph with valued edges, we simply use values in the adjacency matrix instead of ones or zeros. When the graph is reflexive, meaning that a node has a link to itself (i.e. self-loops), there will be values down the main diagonal of the adjacency matrix. In this and many other research cases, self-loops are not allowed, so the main diagonal will hold all zeros (Borgatti et al., 2018; Wasserman et al., 1994).

The adjacency matrix will be symmetric in the case of an undirected graph, meaning that the top half of the matrix (the area above the main diagonal) will mirror the bottom half so that x_{ij} will always equal x_{ji}. This equality may not apply to the adjacency matrix of a directed graph. The adjacency matrix for a graph is always square (i.e., has the same number of rows as columns). It is also one-mode matrix, meaning that its rows and columns both refer to the same set of nodes. In contrast, the rows and columns of a two-mode matrix represent two different sets of nodes. The two-mode matrix is also called an affiliation matrix and it has been applied in several different contexts throughout the network literature (Davis et al., 1941; Mizruchi, 1996; Fowler, 2006). In the same manner, it is foundational to our analysis in subsequent sections (Borgatti et al., 2018; Neal, 2014; Wasserman et al., 1994).

The density and average degree are quantitative measures that characterize the cohesiveness of the entire network structure. In other words, they measure how connected the nodes in the graph are. The density is the total number of links in the network graph expressed as a proportion of the total number of possible links. For undirected non-reflexive graphs, where n is the number of nodes in the network, the total possible links is $\frac{n(n-1)}{2}$ (Wasserman et al., 1994). We may interpret density as the probability that a tie exists between any pair of randomly selected nodes. Density is best used to compare changes within or across networks. In general, density is lower in large networks than in small networks which may pose an issue for network comparisons. For this reason, some researchers prefer the average degree measure. We calculate average degree by averaging the row sums of the adjacency matrix (i.e., the number of links for each node in the network). The average degree is simpler to compute and easier to interpret than the density (Borgatti et al., 2018).

The principal objective of social network analysis is to identify which nodes are structurally important (i.e., measure the contribution that each node makes to the structure of the network). This structural importance is a property referred to as centrality. The degree and betweenness centralities are among the most widely used measures in the network literature. The degree centrality measures the number of links a given node has within a network indicated by the row or column sum of the adjacency matrix for an undirected network where x_{ij} is the (i, j) entry of the matrix and is calculated as $\sum_j x_{ij}$ (Borgatti et al., 2018). The betweenness centrality measures how often a given node falls along the shortest path between two other nodes where g_{ijk} is the number of shortest paths connecting i and k through j, and g_{ik} is the number of shortest paths connecting i and k and whose formula is given by $\sum_{i<k} \frac{g_{ijk}}{g_{ik}}$ (Freeman, 1979; Borgatti, 2005).

The centrality measures are key to network analysis because they assert that opportunity or advantage is afforded to a node based on its position within the network. This assertion is based on the idea that information and resources flow through networks and that the most central nodes are positioned to receive network flows sooner than other nodes, which allows them to act on information earlier or control the distribution of resources. These central nodes may take a gatekeeping or toll-taking role within the network (Brass, 1984). For instance, the nodes who are highest in betweenness centrality are better positioned to disrupt network operations, filter information, and color or distort information as it passes along (Brass, 1984; Pitts, 1979; Borgatti, 2005).

Bipartite Network Projection and Backbone Extraction

The bipartite network consists of two mutually exclusive sets of nodes where links may exist between nodes in different sets but not between nodes in the same set. The bipartite network may also be represented as an m-by-n two-mode affiliation A in which A_{ik} = 1 if there is a link between i and k and zero otherwise. Bipartite networks may be projected (also referred to as folding) onto a one-mode m-by-m network B by multiplying the bipartite matrix by its transpose AA' (Breiger, 1974; Neal, 2014). Using this approach on our data from the federal register, we form a one-mode network of federal agencies linked to one another by their co-sponsorship of federal regulations – a so-called regulatory network. The weight of an edge in the regulatory network B_{ij} reflects the number of regulations that agencies i and j both sponsored.

Figure 3 Regulatory network with nodes sized by total degree centrality

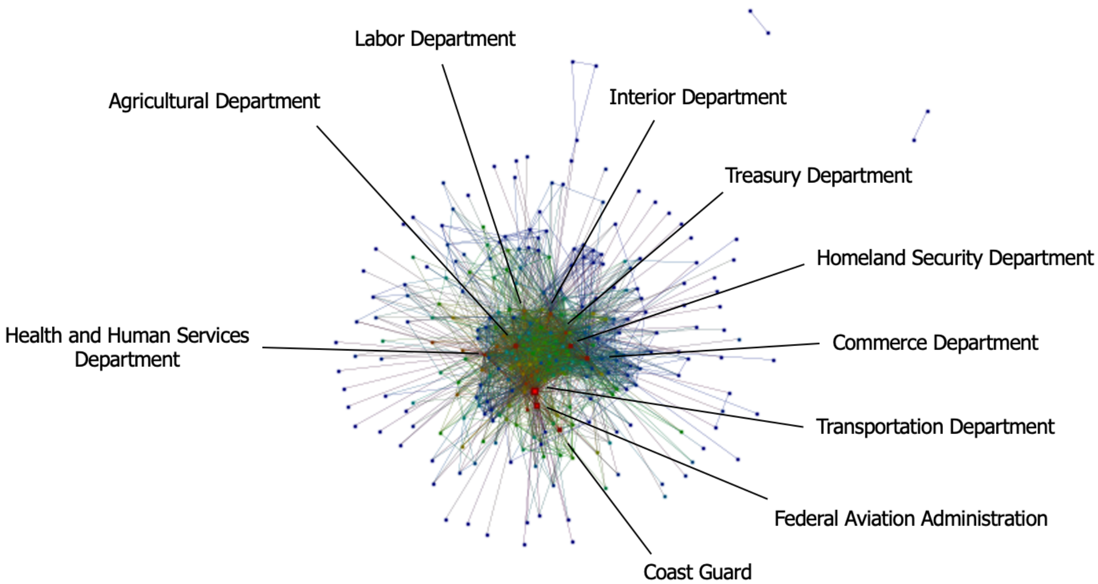
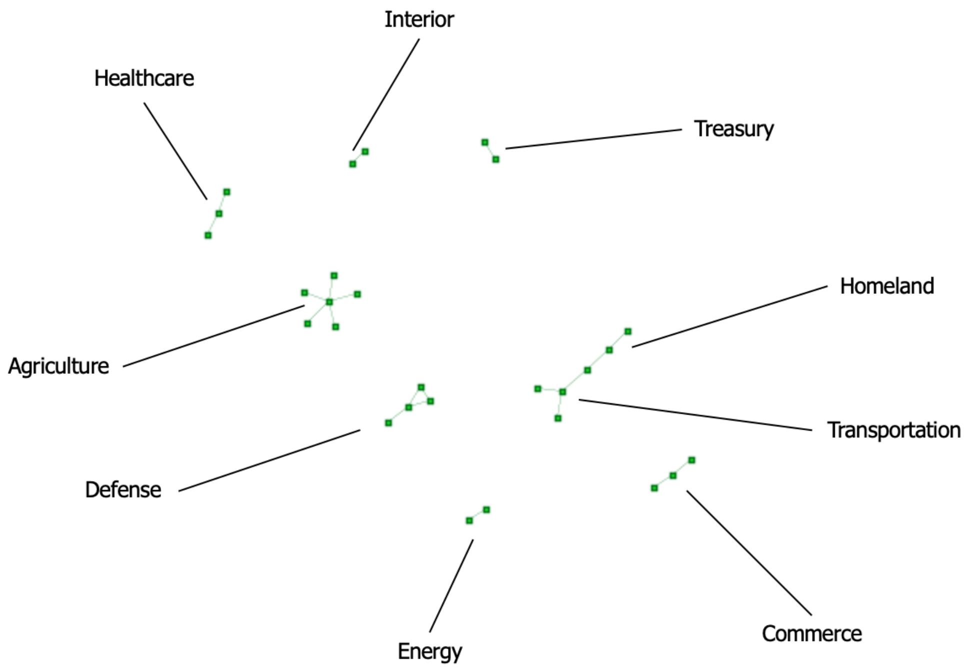


Table 2 Regulatory network statistics

General Statistics	
Total Federal Agencies:	453
Density:	0.014
Link Statistics	
Total Links:	2906
Link Values:	Min: 0, Max: 18506, Mean: 50, Std Dev: 571, Sum: 143798
Component Statistics	
Isolates:	222
Dyads:	2
Triads:	1
Larger:	1
Larger Sizes:	Min: 224; Max: 224; Mean: 224; Std Dev: 0

The backbone extraction of a bipartite projection involves the use of a threshold, where links whose weight exceed the threshold value are retained in the backbone and weights that fall below the threshold are omitted (Neal, 2014). We applied a commonly used threshold of standard deviation plus mean to derive the backbone of our regulatory network (see figure 3). This method is called an unconditional threshold approach; it is the simplest and most widespread approach to extracting the backbone of a bipartite projection and various thresholds have been used throughout the network literature (Derudder and Taylor, 2005; Latapy et al., 2008; Neal, 2013). This major shortcoming of this approach is arbitrariness, as the structure of the backbone network largely depends on the specific threshold value chosen (Neal, 2014). The threshold that we applied to our regulatory network revealed a clear pattern of collaboration between certain federal agencies for which we provide additional analysis in the results section of this article.

Figure 4 Backbone of the regulatory network



Multiple Regression Quadratic Assignment Procedure

The multiple regression quadratic assignment procedure (QAP) is a non-parametric bootstrapping approach designed to test correlations between different networks on the same set of nodes and models the values of a dependent variable network using multiple independent variable networks in matrix form (Krackhardt, 1988). The QAP method correlates two matrices by reshaping them into two long columns and calculates an ordinary measure of statistical association (e.g. Pearson's r) called the observed correlation. Next, it uses a permutation test to calculate the significance of the observed correlation by comparing the correlation to a reference set of thousands of pairs of matrices that are similar to the data matrices but independent of one another.

The permutation test randomly rearranges the rows and matching columns of the data matrix resulting in a new matrix with the same properties but independent of the original matrix. It derives a p-value by counting the proportion of correlations among the independent matrices that were as large as the observed correlation and considers p-values less than five percent as significant (Borgatti et al., 2018; Krackhardt, 1988). We used our regulatory network in figure 2 as the dependent variable and a total of 18 attribute features of the federal agencies in our dataset as independent variables (see table 3). We included the independent variables in their original column-repeated form as well as projected them into one-mode shared attribute networks to determine which, if any, were statistically significant and best contributed to interagency collaboration on federal regulations.

Table 3 Description of independent variables

Independent Variable	Description
Agency	Categorical variable that identifies an entity as an agency in federal register database.
Board	Categorical variable that identifies an entity as a board in the federal register database.
Bureau	Categorical variable that identifies an entity as a bureau in the federal register database.
Child	Categorical variable that identifies an entity as a child of another entity within the federal register database.
Commission	Categorical variable that identifies an entity as a commission in the federal register database.
Committee	Categorical variable that identifies an entity as a committee in the federal register database.
Corporation	Categorical variable that identifies an entity as a corporation in the federal register database.
Department	Categorical variable that identifies an entity as a department in the federal register database.
Department Identification	Categorical variable that associates an entity with a specific department within the federal register database.
Independent	Categorical variable that identifies an entity as independent of other entities within the federal register database.
Judicial Agency	Categorical variable that identifies an entity as a judicial agency within the federal register database.
Legislative Agency	Categorical variable that identifies an entity as a legislative agency within the federal register database.
Parent	Categorical variable that identifies an entity as a parent within the federal register database.
Parent Name	Categorical variable that associates an entity with a specific parent within the federal register database.
Quasi-official	Categorical variable that identifies an entity as quasi-official within the federal register database.
Total Children	Quantitative variable that identifies the total number of children associated with an entity has in the federal register database.

Results

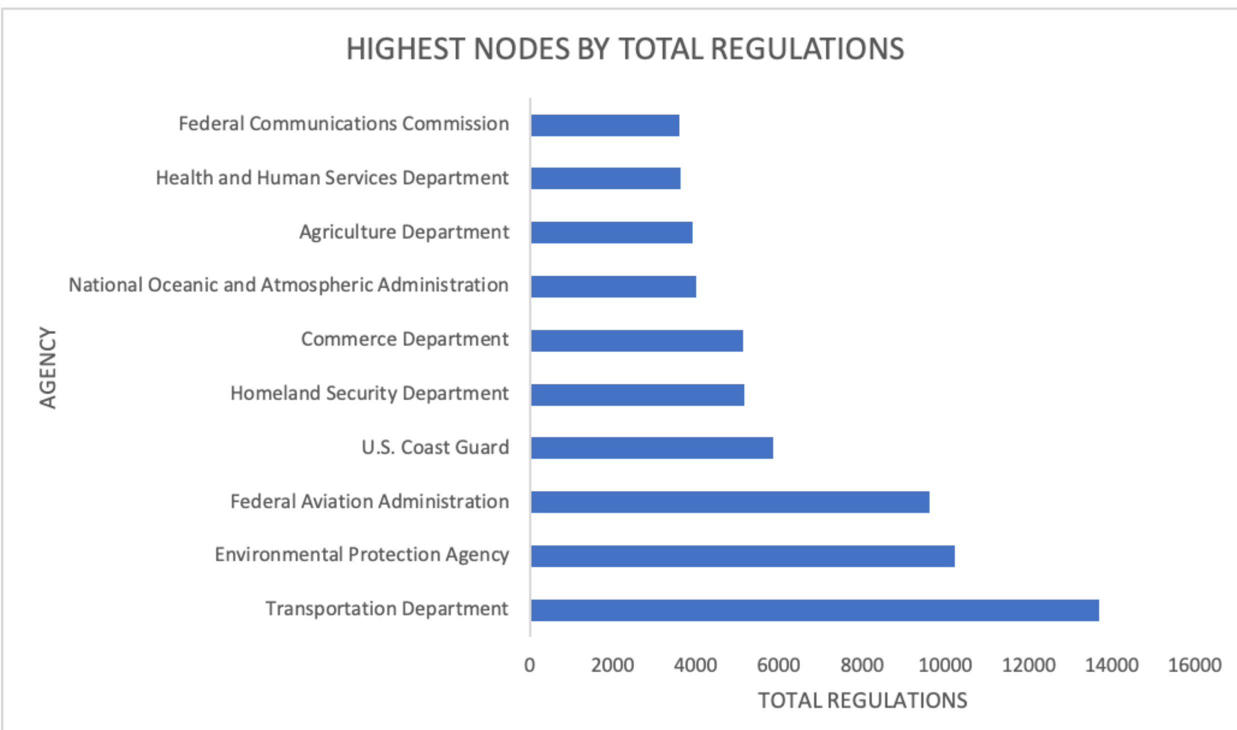
The following table identifies the top federal agencies based on total regulations enacted, total degree centrality, and betweenness centrality. The agencies having high regulatory activity could be considered as burdensome on businesses, individuals, and as a result, the economy. Of course, there are several additional factors that we must consider before attempting to measure the regulatory burden of a federal agency; in the case of this article, we consider the total degree and betweenness centrality. The total degree centrality measures the number of co-sponsor relationships that an agency has in the regulatory network. A high total degree centrality relative to other agencies in the network might suggest that an agency coordinates its efforts with other agencies, which in turn reduces their burden on the system. The betweenness centrality measures the number of instances where an agency sits on the shortest path between two other agencies in the regulatory network. A high betweenness centrality could suggest that an agency may not the source of regulatory activity but influential because enacting a regulation requires their coordination.

Table 4 Top federal agencies by network measure

Total Regulations	Total Degree	Betweenness
Transportation Department	Transportation Department	Treasury Department
Environmental Protection Agency	Federal Aviation Administration	Interior Department
Federal Aviation Administration	Coast Guard	Transportation Department
Coast Guard	Homeland Security Department	Commerce Department
Homeland Security	Commerce Department	Health and Human Services Department
Commerce Department	Agriculture Department	Justice Department
National Oceanic and Atmospheric Admin.	National Oceanic and Atmospheric Administration	Agriculture Department
Agriculture Dept.	Health and Human Services Department	Housing and Urban Development
Health and Human Services Department.	Treasury Department	Postal Regulatory Commission
Federal Comm. Commission	Defense Department	Defense Department

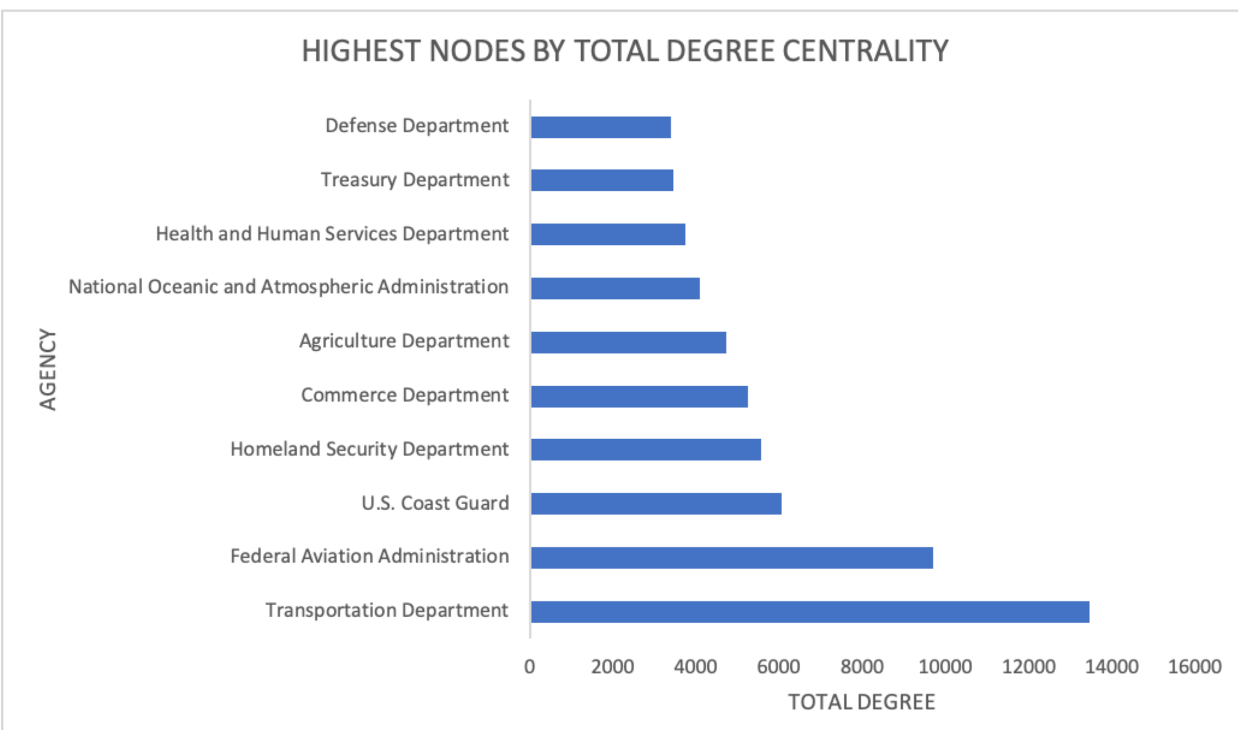
The following chart identifies the federal agencies who were highest in federal regulations enacted during the twenty-six years from 1993 – 2019. The Department of Transportation and the Environmental Protection Agency enacted the greatest number of federal regulations during that period with 13,683 and 10,222 federal regulations, respectively – and accounted for nearly 40 percent of the regulations in our dataset. This constitutes a massive regulatory burden on the federal system and seems to justify many of the complaints that businesses have made about burdensome and often redundant regulations (E.O. 13771, 2017; Business Roundtable, 2019). On the surface, we may use the total number of regulations enacted as a measure of relative importance within the federal agency ecosystem, but this research argues that there are other factors that we must consider; network measures (e.g., degree and betweenness centrality) may give us additional insight or lead us to ask additional questions.

Figure 5 Top federal agencies by total regulation



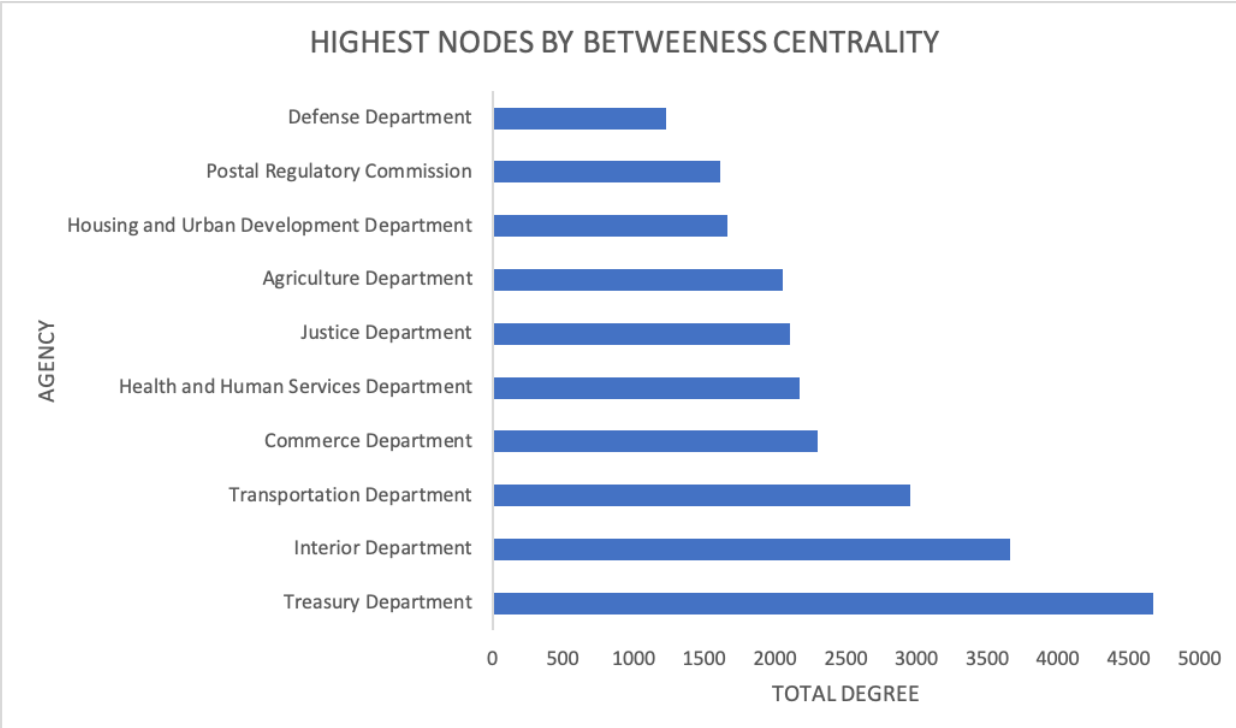
The agencies who are highest in total degree centrality are of particular interest. The following chart depicts the centrality measures for the top ten federal agencies based on total degree. The Department of Transportation leads in total degree centrality followed by the Federal Aviation Administration (which is a sub-agency of the Department of Transportation) and suggests that aviation issues account for a majority of the regulatory activity in the network. The Coast Guard and the Department of Homeland Security follow (the Coast Guard is a sub-agency of the Department of Homeland Security) and may suggest that maritime issues are of major importance in addition to aviation issues. Notably, the Environmental Protection Agency did not count among the agencies with highest degree centrality, which is surprising based on the total number of regulations that it enacts each year; this might also suggest that the Environmental Protection Agency does not need to coordinate its efforts with other agencies in the network.

Figure 6 Top federal agencies by total degree centrality



The agency hierarchies based on total enacted regulations and total degree centrality are similar, with the exception of a few agencies; however, the agency hierarchy based on betweenness centrality is much different. The Treasury Department (responsible for stewarding the U.S. economic and financial systems) and the Interior Department are preeminent, in addition to several other smaller agencies and commissions. This suggests that an entire class of agencies exist that exert subtle influence on the system but without the regulatory drag and that other federal agencies in the network may be unable to enact their desired changes without the coordination and approval of the entities high in betweenness centrality.

Figure 7 Top federal agencies by betweenness centrality



The backbone network that we extracted from the regulatory network has a structure that lends itself to nine central policy areas which may represent the most complex policy issues in government. We found that the preponderance of interagency collaboration occurred between the Departments of Homeland Security and Transportation, a network that is bridged by the Coast Guard. The Department of Defense collaborates frequently with the National Aeronautics and Space Administration (assuming on issues of defense and space) and the General Services Administration and Defense Acquisitions Regulations System (presumably to manage its massive real estate portfolio and to ensure it is able to acquire the cutting-edge equipment quickly and efficiently). The other major policy areas are largely self-contained and managed intra-departmentally.

Table 5 Backbone of the regulatory network

Policy Area	Agency
Agriculture	Department of Agriculture, Rural Utilities Service, Farm Service Agency, Commodity Credit Corporation, Agricultural Marketing Service, Animal and Plant Health Inspection Service
Commerce	Department of Commerce, Industry and Security Bureau, National Oceanic and Atmospheric Administration
Defense	Department of Defense, Defense Acquisition Regulations System, General Services Administration, National Aeronautics and Space Administration
Energy	Department of Energy, Federal Energy Regulatory Commission
Healthcare	Department of Health and Human Services, Centers for Medicare and Medicaid Services, Food and Drug Administration
Homeland	Department of Homeland Security, Coast Guard, Customs and Border Protection
Interior	Department of Interior, Fish and Wildlife Service
Transportation	Department of Transportation, National Highway Traffic Safety Administration, Federal Aviation Administration, Coast Guard
Treasury	Department of Treasury, Internal Revenue Service

The six statistically significant variables that we found through MRQAP help provide some explanation for the regulatory network structure. Whether two agencies fall under the same department explains the structure of the network backbone and certain major policy areas being largely self-contained. Whether two linked agencies share a parental relationship explains several (though not all) of the departmental agencies are high in centrality. The total number of children managed by a parent agency in an additional significant variable that supports the high centrality to departmental agencies. While these variables are significant, it represents only a small fraction of the variance that we find the regulatory network, and for that reason, there is still much work to be done in explaining the nature of the regulatory network that exists between federal agencies.

Table 6 MRQAP Results

Types of Collaboration	Coef.	P-value
Intercept	5.745	**0.04**
Shared Organization	16.503	**0.00**
Department and Department	9.618	**0.01**
Independent and Independent	-0.294	0.45
Subordinate and Subordinate (External)	-0.558	0.58
Subordinate and Subordinate (Internal)	-15.020	1.00
Parent and Child	253.492	**0.00**
Department and Independent	-3.578	0.90
Department and Subordinate (External)	-1.887	0.71
Independent and Subordinate (External)	-4.889	0.97

Discussion

The findings of this research are important because they give us insight into the federal government which might not be readily apparent when using the traditional statistical methods and processes. Social network analysis allows us to view the executive branch, for instance, as a complex network of federal agencies who connect through the regulations that they enact within and across different policy issues. In fact, considering the other branches of government – legislative and judicial – and other levels of government – state and local – renders a complex ecosystem of entities who establish ties at different areas of policy at various points in time – what the literature calls a dynamic meta network (Carley, 2006).

Public policy is the language of the federal government and federal agencies create federal regulations in its daily operations. Those regulations become artifacts that connect federal agencies to particular policy issues at particular points in time; further, it connects federal agencies to any other agency required to coordinate or approve the regulation. Over time, networks transpire between federal agencies around those policy issues which we have coined regulatory networks. These so-called regulatory networks may provide key insights into the inner workings of our government based on the nature of the data available to us and the nature of the questions that we ask of it.

The principal objective of this research was to determine whether a network existed that we could capture from the data available to us in the federal register database to apply network analytical tools to better understand the network structure were a network to exist. Based on the structure of the data in the federal register database, we were able to construct an undirected network graph that connected each agency in the database to its associated federal regulations. Subsequently, we folded that undirected graph into an agency-by-agency shared regulations network where a regulation co-occurrence was the tie and the number of co-occurrences was the tie strength. A major shortcoming of how the federal register database structured the data was that it lacked a data field to identify the sponsor agency. It only provided a list of agencies who co-occurred on the regulation. Of course, we could have made an assumption that the first agency in the list was the regulations sponsor. Doing so would have allowed us to create a directed graph, which would have allowed us to use additional network measures and methods; in turn, we would have been able to uncover deeper insights. However, a failed assumption would threaten the validity of our results. The undirected graph served the purposes of this research, and we leave additional data manipulation efforts open to future research.

More than half of the regulations in our dataset involved the Department of Transportation (or its sub-agencies) or the Environmental Protection Agency. If we were to use the amount of regulatory activity as a measure of performance, these agencies would appear to be involved in more issues and work harder than any other federal agency. The Federal Communications Commission was another agency that was high in regulatory activity in our data set. We emphasize the Environmental Protection Agency and the Federal Communications Commission in particular because while they were high in regulatory activity, they were not present in the hierarchies based on the centrality measures. This is important because it indicates that regulatory activity is not a proxy for importance, nor should it be used as proxy for performance. In contrast, the Department of Defense and the Treasury Department were not high on regulatory activity, but they were high in degree centrality. They were connected to more agencies on more regulations which suggests that they are more important within the executive branch because they may be involved in more complex policy issues or that more agencies in the network solicit their advice or support on policy issues. Similar observations exist for the agencies high in betweenness centrality. The Treasury Department led in betweenness centrality, which supports our assertion that treasury may be involved in more complex policy issues or more solicited for advice and/or support by other agencies in the network. Notably, the Justice Department, Housing and Urban Development, and the Postal Service weren't on any of our earlier lists, but they may lead in betweenness centrality because of the support functions they play within the network.

This research finds that the federal government is not as collaborative as one might expect. The backbone of our regulatory network rendered nine components that we categorized as major policy areas (largely aligned by cabinet departments but not in each instance, which makes this finding not so obvious). We have determined that these policy areas of greatest significance (or complexity) based on the total number of collaborations (ties) surrounding policy issues within the area, as represented by the regulations enacted and the agencies who enacted them. There are certain policy areas (e.g., agriculture and energy) which are largely self-contained, but we do identify the central agencies and sub-agencies within those particular policy areas. There are other policy areas which require massive amounts of collaboration (e.g., homeland security, defense, and transportation). Then there were policy areas who passed vast amounts of regulations in isolation (e.g., environment and communications) which have also been the source of friction and considered burdensome on the federal system.

We leave several areas for future research. Future research may attempt to structure a directed graph from the available data and calculate additional network measures (e.g., in-degree and out-degree centrality) or apply more advanced network methods (e.g., community detection). While the focus of this research centered on regulatory activity, several questions may also be posed about the regulatory inactivity of many agencies in the data. For instance, there were more than one hundred isolates in each network we considered, meaning there were federal agencies with zero regulations or collaborations in the network – future research might explore the nature of this inactivity and the consequent low centralities of those federal agencies. There is a clear need within the field of public policy for finding ways to reduce the regulatory burden on businesses and individuals; future research might consider ways to accomplish this by restructuring certain agencies or commissions (e.g., Environmental Protection Agency or Federal Communications Commission) within the executive branch. We identified a number of significant policy areas within the federal government, but it raises a concern about other policy areas of consequence that weren't included in our data set, but which current events have proven are of critical importance – for example, social cybersecurity and disinformation (Carley et al., 2018).

This research contributes to the literature of public policy and administration by studying the regulatory activity of federal government agencies using social network analysis. Regulatory networks provide insight on federal agencies and their activities, and the relationships that exist (and maybe shouldn't) or don't exist (and maybe should) between federal agencies in the network. When measured over time, regulatory networks provide a useful measure of federal agency performance, in terms of the policy issues that federal agencies work in tandem to address. Ultimately, regulatory networks may increase federal accountability and render a government that is less burdensome and addresses the needs of all citizens.

Chapter 3: A Network-Based Framework for Detecting Regulation Overlap

Background

Our research examines regulation overlap using network science as its analytical approach. The regulatory overlap concept originated with public policy practitioners and stakeholders who identified instances where multiple policies existed to guide a single policy area which were often sponsored by different government agencies. In the best-case scenario, the multiple policies were only superfluous; in a worst case, the multiple policies offered competing guidance. In either case, this multiple policy guidance imposed significant costs for both the regulator and the regulated, alike (Business Roundtable, 2019).

The Government Accountability Office (GAO) is a United States (U.S.) government agency who formally defined regulation overlap using three distinct categories – fragmentation, duplication, and overlap (GAO, 2015). Fragmentation occurs when multiple agencies are involved in the same policy. Duplication occurs when two or more agencies are engaged in the same activities or provide the same services. Overlap occurs when multiple agencies have similar goals, engage in similar activities, or target similar beneficiaries. While our research makes frequent use of articles and anecdotes centered on U.S. government agencies and their regulations and stakeholders, the challenges posed by regulation overlap are not confined to the U.S. alone. Policy practitioners from several different countries, representing various forms of government, have written about the perniciousness of this phenomenon (European Union, 2014; Li, 2015; Commonwealth of Australia, 2014).

Generally, regulations research intersects several disciplines, namely political science, public policy and administration, and law. A major goal of regulations research is to measure and reduce complexity. The Mercatus Center is a research institute at the George Mason University which has developed frameworks for measuring complexity in federal, state, and local regulations. Its RegData and Federal Regulations and State Enterprise (FRASE) Index compiles and ranks state regulations according to their impact on private-sector industries in each state's economy, respectively (Al-Ubaydli and McLaughlin, 2017). Additionally, Daniel Katz and Michael Bommarito have applied network science to measure complexity in the U.S. Code, which is a compilation of the statutes enacted by the U.S. Congress (Bommarito and Katz, 2010).

While policy practitioners and stakeholders have written extensively about regulation overlap, much less research on the subject exists within the academic literature. Gordon Brown (1994) proposed two categories of regulation overlap – horizontal and vertical overlap. Horizontal overlap occurs between two or more government agencies on the same level (e.g., overlap between two federal agencies). Vertical overlap occurs between two or more government agencies on separate levels (e.g., overlap between a federal agency and a state agency). Robyn Hollander (2010) found that efforts within the Australian government to eliminate duplication and overlap in environmental policy imposed artificial divisions on a complex policy domain; by limiting opportunities for political engagement, surrendered some of the strength of a federal system.

Straughter and Carley (2021) modeled the U.S. federal regulatory environment as a complex network of federal agencies, federal regulations, and keywords, and used network science as an analytical approach to studying the problem of regulatory overlap. In efforts to identify a framework for measuring regulatory overlap directly, they proposed a regulatory burden concept which measures the burden imposed by each federal agency within the regulatory system based on their usage of shared keywords and regulations, which addresses the issue of regulatory overlap indirectly. Our research extends that of Straughter and Carley (2021) with additional modeling of the U.S. federal regulatory system and proposing a framework for the direct measurement of regulatory overlap. We analyze two additional views of the federal regulations network which dissects the regulation overlap problem down into two distinct yet related concerns which we conceptualize as agency and keyword overlap. These additional considerations allow us to identify the sources of regulation overlap, make further recommendations that optimize the government's organizational structure, and identify latent collaborative opportunities.

Methods

Our data set includes 99014 federal regulations enacted between 1993 and 2019 along with their associated metadata (i.e., the enactment date, the list of federal agencies who co-sponsored the regulation, and additional information about the federal agencies). Each federal regulation is a text file extracted from the federal register database. The federal register is the central repository for all federal regulatory activity. Federal agencies may enact regulations unilaterally or in conjunction with other agencies, following strict procedures which are codified in law to ensure congressional oversight and allow for public notification and response.

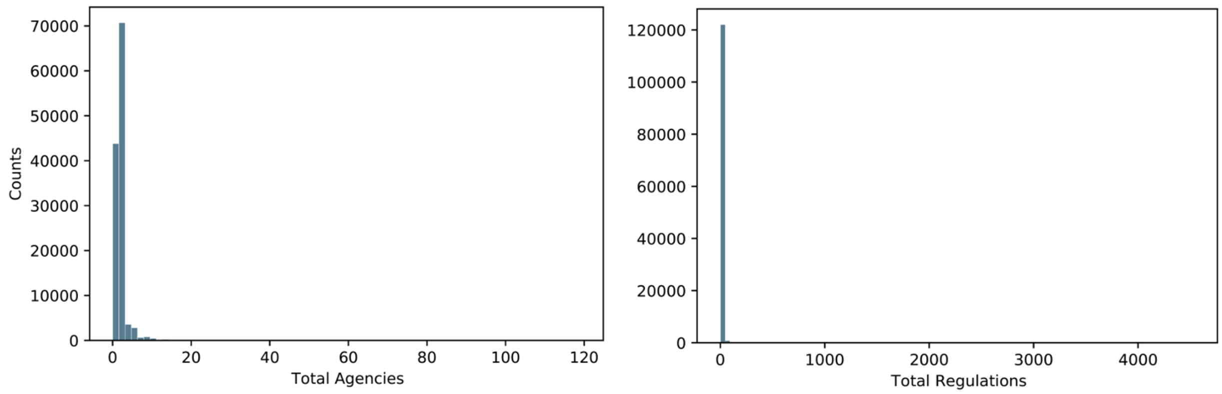
We used a social network analysis and visualization software developed at Carnegie Mellon University called Organizational Risk Analyzer (ORA) to formulate our networks (Carley et al., 2013). Straughter and Carley (2021) formulated a regulations network which mapped each regulation to its respective keywords, which they defined as the nouns within each regulation. Their process resulted in a network of more than 200,000 keywords whose dimensionality led to downstream issues which prevented them from a direct identification of overlap in regulations. In this research, we overcome this dilemma by calculating the term-frequency inverse document-frequency (TFIDF) for each word in the regulation corpus, where:

$$w_{i,j} = tf_{i,j}\ x \log\left(\frac{N}{df_i}\right)$$
$$tf_{i,j} = frequency\ of\ i\ in\ j$$
$$df_i = number\ of\ documents\ containing\ i$$
$$N = total\ number\ of\ documents$$

The TFIDF calculates values for each word in a document through an inverse proportion of the frequency of the word in a particular document to the percentage of documents the word appears in, where words with high TFIDF values imply a strong relationship with the document they appear in (Ramos, 2003).

Our definition of keyword extends the previous definition to include bigrams but considers only those keywords that appear in two or more regulations (we assume that regulation overlap results from a keyword appearing in multiple regulations). Using TFIDF, we extract the top five words from each regulation as keywords, which results in a network that is distinct from the previous research. Akin to previous research, the network includes 312 agencies and 99,014 regulations; however, unlike previous research, the network has 123,209 keywords, which constitutes a greater than fifty percent reduction in dimensionality from the previous research.

Figure 8 Keyword Distribution



We also noticed that many keywords were associated with a single regulation or agency and wouldn't contribute to overlap, so we removed them from the network (Figure 1). Further, we removed subordinate agencies from the network, based on previous research which found that most regulatory activity in the network occurs between subordinate agencies and their respective executive departments – meaning subordinate agencies do not enact regulations unilaterally apart from their executive departments – and maintaining them in the network would bias our results by showing that more overlap is present in the network than what actually exists (Straughter & Carley, 2021). The trimmed network consists of 153 agencies, 99,014 regulations, and 15,599 keywords. Next, we map each of these node sets to one another, yielding three additional network views. This approach allows us to conceptualize three distinct yet related forms of government overlap – regulation, keyword, and agency overlap.

Figure 9 Regulation Overlap

$$
\begin{aligned}
&\textit{Let Matrix A represent the regulation x keyword network.}\\
&\textit{Let Matrix B represent the regulation x agency network.}\\
&Shared\ Keyword\ =\ Binarize\ (A^T A)\\
&Shared\ Agency =\ Binarize\ (B^T B)\\
&Regulation\ Overlap = Shared\ Keyword - Shared\ Agency\\
&\textit{Recode link values less than zero to zero.}
\end{aligned}
$$

Regulation Overlap. Regulation overlap occurs when two or more regulations share one or more keywords without sharing a regulation. Measuring regulation overlap requires the projection and binarization of the regulation-agency and regulation-keyword bipartite networks. Bipartite network projection is used frequently in applied research to transform a two-mode network into a one-mode network of shared attributes between nodes (Neal, 2013). Projection yields a network detailing the number of agencies and keywords that any two regulations share, while binarization indicates whether any two regulations share a keyword or agency. Regulation overlap is the difference between the *regulations (shared keyword)* and *regulations (shared agency)* networks (Figure 2). The regulation overlap network is an unweighted network where a link value of one constitutes overlap and zero otherwise. Negative link values (indicating that regulations shared an agency but not a keyword) were recoded to zero.

Figure 10 Keyword Overlap

$$\begin{aligned}
&\textit{Let Matrix A represent the keyword x agency network.}\\
&\textit{Let Matrix B represent the keyword x regulation network.}\\
&\textit{Shared Agency} = \textit{Binarize}\ (A^T A)\\
&\textit{Shared Regulation} = \textit{Binarize}\ (B^T B)\\
&\textit{Keyword Overlap} = \textit{Shared Agency} - \textit{Shared Regulation}
\end{aligned}$$

Keyword Overlap. Keyword overlap occurs when two or more keywords share one or more agencies without sharing a regulation. Measuring keyword overlap requires the projection and binarization of the keyword-regulation and keyword-agency bipartite networks. Projection yields a network detailing the number of regulations and agencies that any two keywords share, while binarization indicates whether any two keywords share a regulation or agency. Keyword overlap is the difference between the *keywords (shared agency)* and *keywords (shared regulation)* networks (Figure 3). The keyword overlap network is an unweighted network where a link value of one constitutes overlap and zero otherwise. This keyword view supplements the regulation overlap network by indicating the keywords upon which regulation overlap is centered.

Figure 11 Agency Overlap

$$\begin{aligned}
&\textit{Let Matrix A represent the agency x keyword network.}\\
&\textit{Let Matrix B represent the agency x regulation network.}\\
&\textit{Shared Keyword} = \textit{Binarize}\ (A^T A)\\
&\textit{Shared Regulation} = \textit{Binarize}\ (B^T B)\\
&\textit{Agency Overlap} = \textit{Shared Keyword} - \textit{Shared Regulation}
\end{aligned}$$

Agency Overlap. Agency overlap occurs when two or more agencies share one or more keywords without sharing a regulation. Measuring agency overlap requires the projection and binarization of the agency-keyword and agency-regulation bipartite networks. Projection yields a network detailing the number of keywords and regulations that any two agencies share, while binarization indicates whether any two agencies share a keyword or regulation. Regulation overlap is the difference between the *agency (shared keyword)* and *agency (shared regulation)* networks (Figure 4). The regulation network is an unweighted network where a link value of one constitutes overlap and zero otherwise. This agency view supplements the regulation overlap network by indicating which agencies are the sources of overlap.

Figure 12 Overlap Potential

$$Overlap\ Potential = \sum_{i} \frac{\#\ Shared\ Keywords}{\#\ Shared\ Regulations + \epsilon}$$

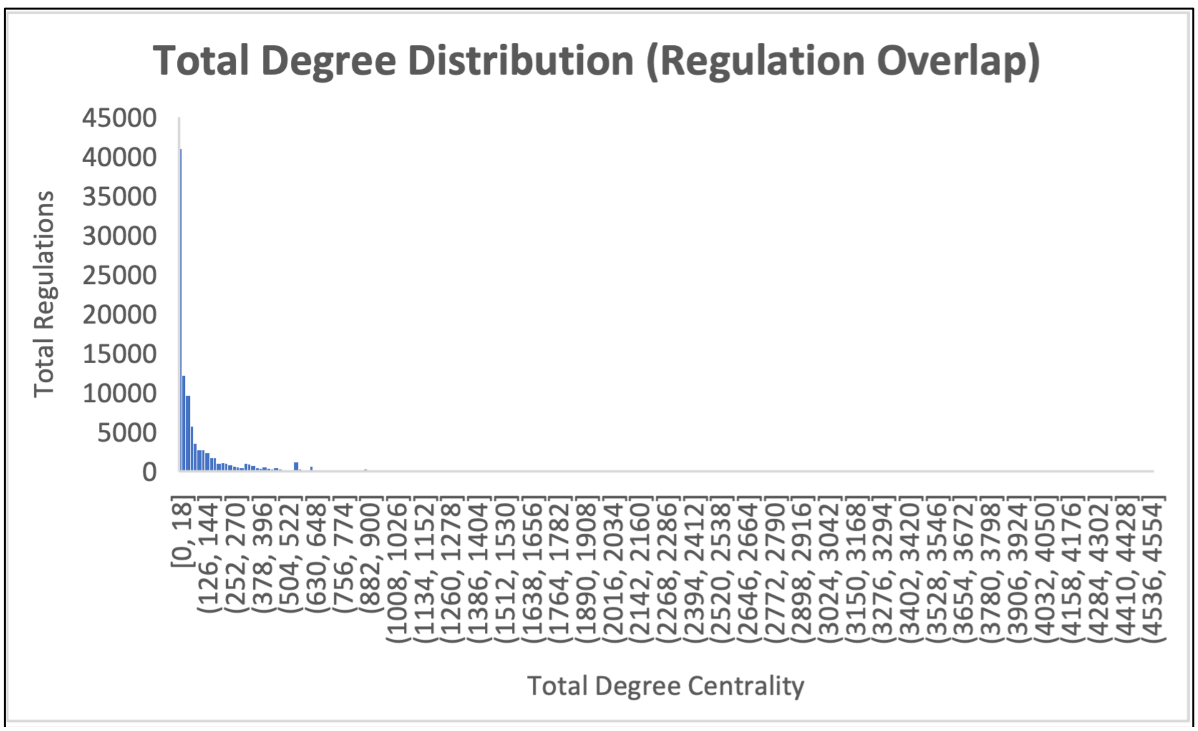
An alternative measure that we propose is the overlap potential. The overlap potential is an agency attribute which measures how likely an agency is to overlap with each of its neighbors (Figure 5). Essentially, this measure creates a fourth view of the network; however, since it is an indirect measure on overlap, we will use this measure to supplement the results of our agency overlap framework. The overlap potential of an agency is determined by taking the ratio of its shared keywords to shared regulations between its respective neighbors, where its total degree would represent an agency's total overlap potential across the network. In this way, we can gauge the risk of overlap across the entire system, and better understand which agencies pose the greatest overlap risks.

Results

Table 7 Regulation Overlap - Network Statistics

General Statistics	
Total Federal Agencies:	99014
Density:	0.0011
Link Statistics	
Total Links:	10926504
Component Statistics	
Isolates:	9507
Dyads:	31
Triads:	2
Larger:	5
Larger Sizes:	Min: 4, Max: 89419, Mean: 17888, Std Dev: 35766

Figure 13 Total Degree Distribution - Regulation Overlap Network

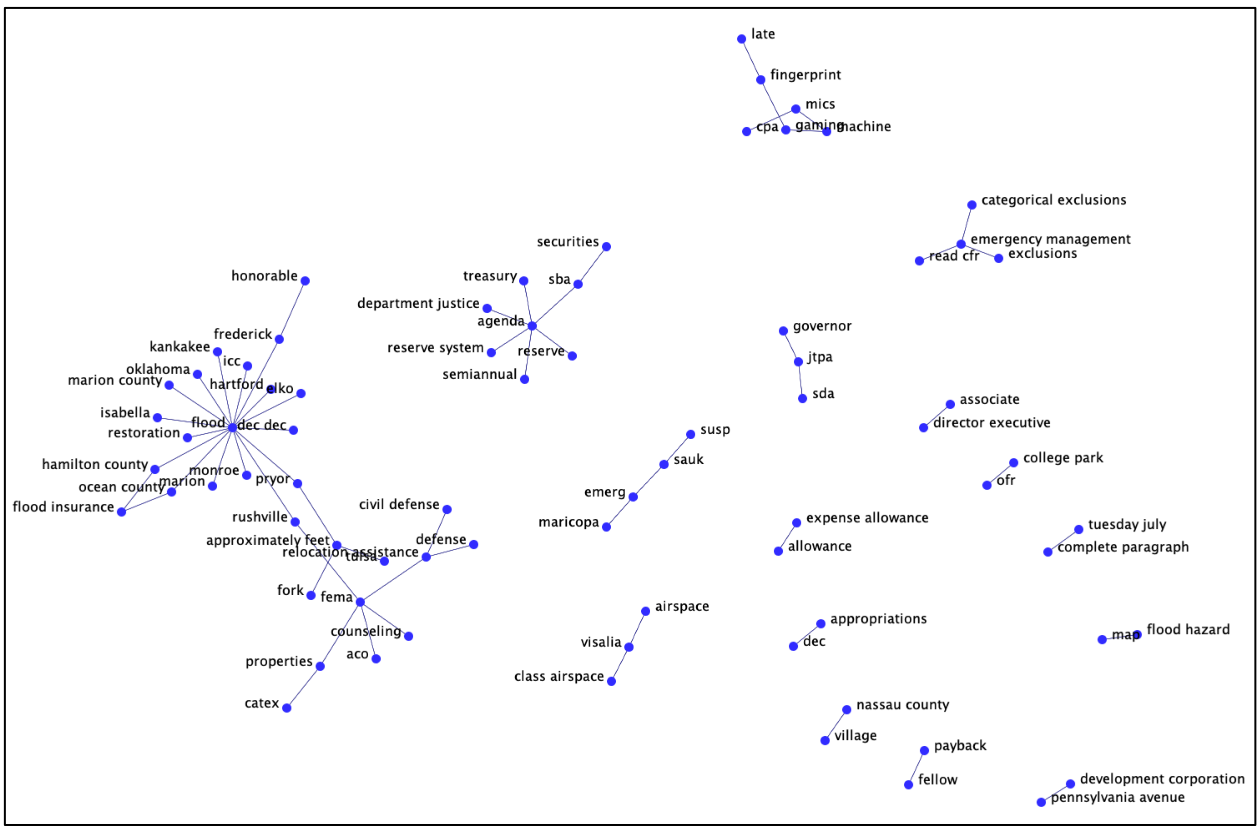


This research conceptualizes three forms of government overlap – regulation, keyword, and agency overlap. Using this framework, our research found a significant level of federal government overlap across all three typologies. First, we determined that 90 percent of the regulations in our data overlapped with at least one other regulation based on shared keywords; only 9,507 regulations were non-overlapping (Figure 5). The total degree centrality measure is a great indicator of the level of regulation overlap existing within the network. The total degree centrality ranges from the minimum value of zero regulations to a maximum value of 4,550 regulations, with a median value of 30 regulations and an average value of 110 regulations (Figure 6). Most of the regulations overlap with twenty or fewer other regulations (<< 1% of total regulations). Keeping in mind that our data spans a twenty-five-year time horizon, one could argue that these results support assertions found in the previous research that the overlap that exists is "well-managed" (Brown, 1994).

Table 8 Keyword Overlap - Network Statistics

General Statistics	
Total Federal Agencies:	15599
Density:	0.0000
Link Statistics	
Total Links:	122
Component Statistics	
Isolates:	15523
Dyads:	9
Triads:	2
Larger:	5
Larger Sizes:	Min: 4, Max: 30, Mean: 10; Std Dev: 10

Figure 14 Keyword Overlap - Network Visualization



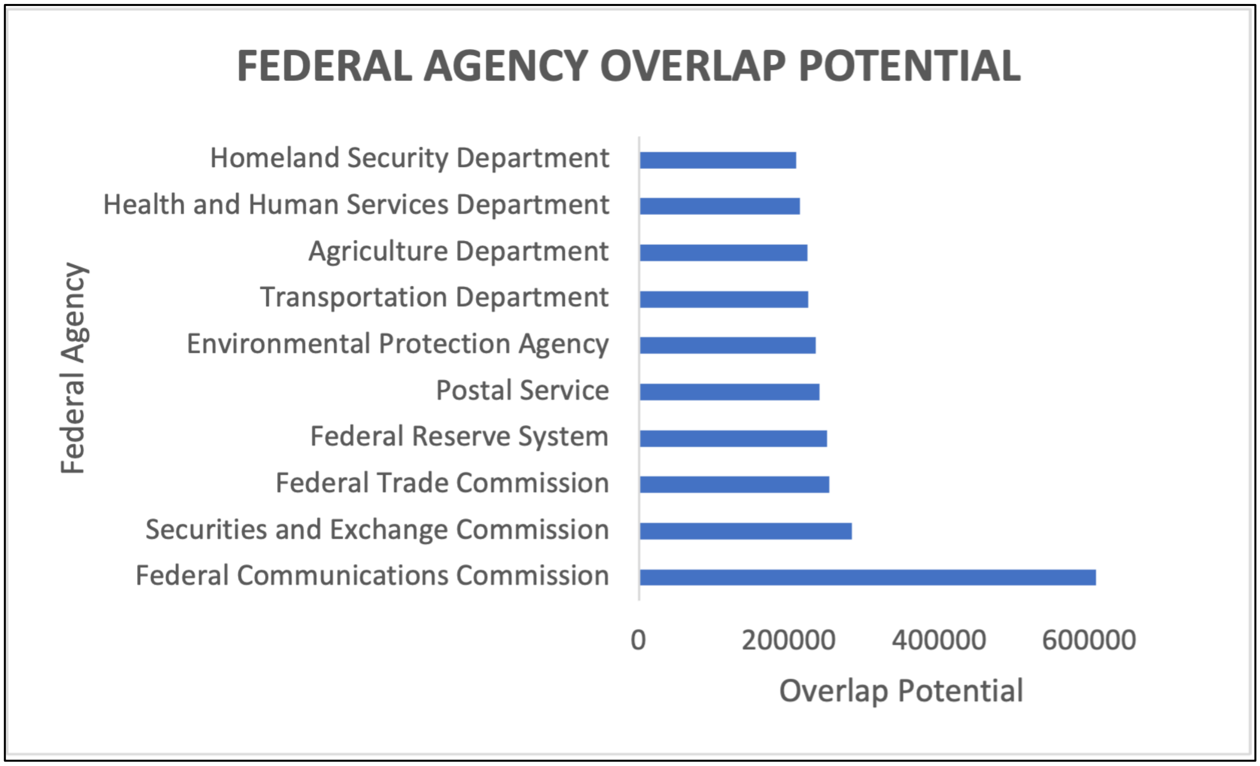
The regulation overlap network alone is not sufficient to fully understand the nature and scope of government overlap that exists. The keyword overlap network provides an additional view of government overlap that is related to regulation overlap yet distinct in the information that it provides (Figure 7). We propose that keyword overlap exists when two or more keywords share an agency without sharing a regulation. Using this definition, we determined that regulation overlap, in large part, results from 79 keywords – much less than one percent of the keywords in the dataset. In the same manner that we used total degree centrality to measure regulation overlap, we can also use the measure to assess keyword overlap. The maximum total degree centrality of the 79 overlapping keywords is 16 keywords and the mode is one keyword. Since the keyword overlap network is small, we can visualize this network to see the relationship between keywords (Figure 8). Our research determined that keyword overlap, and hence regulation overlap, centers on a few key policy areas – namely, emergency management and airspace management.

Table 9 Agency Overlap - Network Statistics

General Statistics	
Total Federal Agencies:	153
Density:	1.0000
Link Statistics	
Total Links:	23256
Component Statistics	
Isolates:	0
Dyads:	0
Triads:	0
Larger:	1
Larger Sizes:	Min: 153, Max: 153, Mean: 153, Std Dev: 0

Having determined the scope and nature of regulation and keyword overlap, the final aspect of our framework considers agency overlap. We propose that agency overlap exists when multiple federal agencies share one or more keywords without sharing a regulation. Using this definition resulted in a network that was fully connected; suggests that every agency in the network overlaps with every other agency. While this result is helpful, and unexpected, it requires us to perform additional analysis. Hence, we conceptualize a fourth term called the ***overlap potential***. Akin to the regulatory burden measure proposed by Straughter & Carley (2021), overlap potential measures each agency's overlap relative to its neighbors in the network. Overlap potential is the ratio of shared keywords to shared regulations between any two agencies and used as the link value in the refined agency overlap network (Figure 9).

Figure 15 Federal Agency Overlap Potential



The overlap potential allows us to measure the level of overlap that exists between any two government agencies. Zero valued links in the shared keyword and shared regulation networks were given a small amount of weight to avoid zero-division errors. This ratio results in large weights when the shared keywords are high and shared regulations are low, and small weights when the shared keywords are low and shared regulations are high. The overlap potential ranges from a minimum value of zero to a maximum value of 75,300 for this dataset. We could scale this measure to avoid dealing with large numbers, but we avoid this action because the main driver of this measure is the magnitude of the potential relative to other agencies in the network. The mean value is approximately 422 with a standard deviation of 1,624, which suggests that there are several agencies with a disproportionate amount of overlap potential relative to its neighbors (Figure 10). Further analysis determines that 60 percent of the top ten agencies in overlap potential are independent agencies. Additionally, the four executive departments highest in overlap potential are also responsible for emergency management, further supporting our analysis of the keyword overlap network.

Discussion

In this research paper, our goal has been to provide a framework by which government overlap may be measured and analyzed; hopefully, this framework will facilitate its reduction. Both public policy practitioners and industry officials have provided anecdotal evidence to suggest that government overlap is inefficient and impedes economic growth. Up to this point, prior research has been largely qualitative; however, there is an academic literature forming at the intersection of government regulation and network science. Brown (1994) proposed a framework to classify government overlap as horizontal and vertical. Using that definition, our research considers horizontal overlap. Of course, this research uses data from the U.S. federal register; based on the nature of our results, one might argue that these results are not generalizable to the vertical case or even international governments. We believe that this framework is general enough that when tested in the vertical case (e.g., federal and state or local regulations, or federal and international regulations) the framework will perform equally well.

Our framework uses the TFIDF to extract the top five words from each regulation and uses them as keywords to represent the document. TFIDF allows us to drastically reduce the dimensionality of the dataset relative to previous research (Straughter & Carley, 2021). The benefit of this method is that it allows us to reduce the dimensionality in a manner that facilitates further projection and analysis of the network. The limitation here is that there is a significant loss of information. Using this method, we identified significant overlap in all three network views; however, that doesn't mean there are no other points of overlap that this framework doesn't consider. An alternative approach would be to incorporate additional keywords by TFIDF (e.g., the top ten words) to represent each document; while this might be more comprehensive, it may also require additional data cleaning downstream or generate additional links in the network that are of no analytic value. We noticed in our research that certain words (e.g., agenda) were a source of overlap, but based on our domain knowledge and review of the data, we understand that this relationship isn't indicative of overlap. Policy practitioners and academics who apply this framework must also use the same judgment in their analysis.

Another limitation of this framework is that it requires specific information be extracted from the data, namely identifier for the regulation node (e.g., a document number or regulation name), the agencies who co-sponsored the regulation, and then the regulation raw text. However, this framework may be adapted if not all the information is present. For example, a unique identifier may be assigned to regulations; if co-sponsor information is not present, regular expressions may be used to extract this information from within the document, or the user may opt to consider two of the three views (regulation and keyword) presented in this research, as opposed to an agency view. Also, if the keywords of a regulation a priori knowledge, then a user may forego the TFIDF calculation altogether. This framework provides flexibility keeping the policy practitioner in mind.

In the case of federal regulations, we found that there was a significant amount of overlap over the past twenty-five years. Most federal regulations overlap with at least one other regulation. This framework allows us to pinpoint exactly which regulations overlap with which others; however, there are caveats to our analysis that practitioners must be aware. First, domain knowledge is required to determine the severity of the overlap. Unfortunately, this framework doesn't allow us to completely automate this process. A practitioner must still review the two documents to determine the severity the overlap. While two regulations may use the same keyword, they may regulate different aspects of a particular topic. The

GAO identified this as fragmentation, when two or more agencies are responsible for a certain portion of a policy area. However, a benefit of this framework is that it points to potential areas of fragmentation, which may be consolidated by forcing agencies to collaborate on regulations within the same policy area, which is a net benefit from the beneficiary perspective.

Further, we identified that most overlap within the federal regulatory system centered on emergency management and airspace management policies. This was indicated by the clusters in the keyword overlap network which were connected by such concepts as ‘flood’, ‘flood management’, ‘emergency management’, and ‘airspace’. We also identified concepts such as ‘securities’, ‘sba’, ‘homeland’, and ‘fcc’ which are indicative of both keyword and agency overlap, as these concepts appeared in regulations which were not offset by a shared regulation. This also points to the fact that there are regulations in which agencies may be mentioning other agencies, but not doing the work of collaborating with said agency on the regulation. In effect, this passes the responsibility to the beneficiary to determine how best to address disparate regulations and hearkens to previous research on the regulatory burden imposed by government agencies onto their beneficiaries (Straughter & Carley, 2021).

Our research finds that there is some baseline level of overlap that exists between all federal agencies in the network. As our framework makes inferences using an unweighted network, a fully connected network such as the agency overlap network is of no analytic value. However, we have not completely removed the notion of agency overlap from the framework because while it isn't predictive in this instance, we believe that over a shorter time horizon or when considering the case of vertical overlap, the agency overlap network may show relationships that are of importance to the policy practitioner. We propose an alternative measure, which we coin as the overlap potential of a government agency. The overlap potential is measured upon an agency relative to its neighbors in the agency overlap network. Essentially, it is a separate view on the network which may be used in lieu of or as a supplement to the agency overlap concept presented as part of the framework.

Using overlap potential, we can assign a risk level for each agency in the network. This measure considers an agency's past practices relative to other agencies present in the network at the same time. Based on our assumption that an agency will continue in those practices into the future, unless some other authority steps in to change said behavior, we also posit that this measure is also forward-looking. The overlap potential has implications for the organizational structure of a government. An ego network analysis of agencies with high overlap potential would lead to recommendations on potential reorganization, consolidation, and/or working group formations. We leave this task to policy practitioners and to future academic research. We also believe that future research in this area should test this framework under different policy environments (e.g., in support of previous qualitative research into environmental or emergency management policies), across different policy levels (e.g., vertical overlap between federal and state regulations), and using different systems of government (e.g., international regulations).

This research paper presents a framework for measuring and analyzing government overlap using an applied networks approach. This framework conceptualizes government overlap as three distinct yet related network views – regulation, keyword and agency overlap. Regulation overlap exists when two or more regulations share one or more keywords but not an agency; keyword overlap exists when two or more keywords share one or more agencies but not a regulation; agency overlap exists when two or more agencies share one or more keywords but not a regulation. This framework extends previous government overlap research by approaching the subject empirically. It provides a mechanism whereby policy scholars may advance future network-centric policy research and practitioners may derive actual insights from their analysis.

Chapter 4: Towards a Network Theory of Regulatory Burden

Background

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 2020).

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 & Bommarito 2013; Bommarito & 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 & 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 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 16 Distribution of Total Regulations

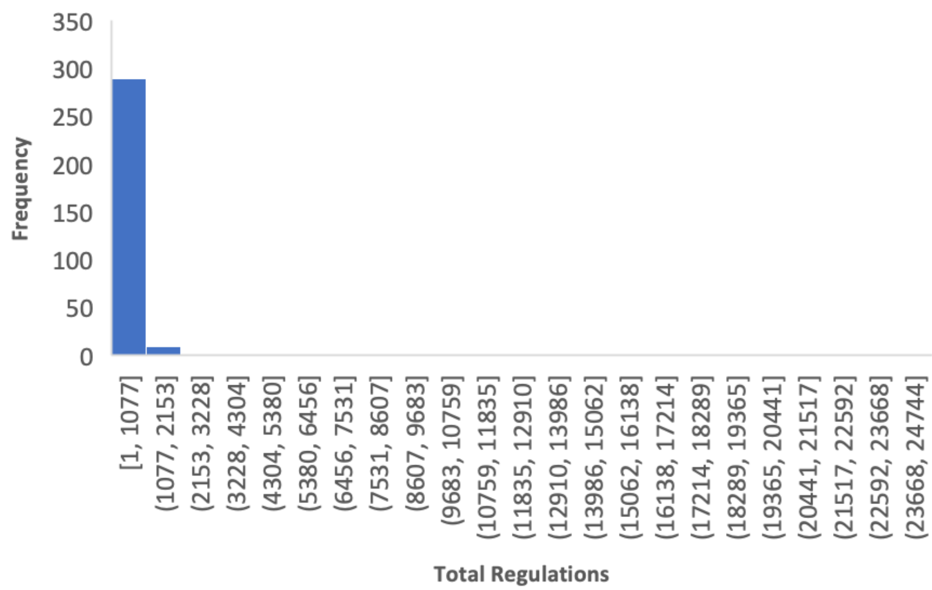


Figure 17 Distribution of Federal Agencies

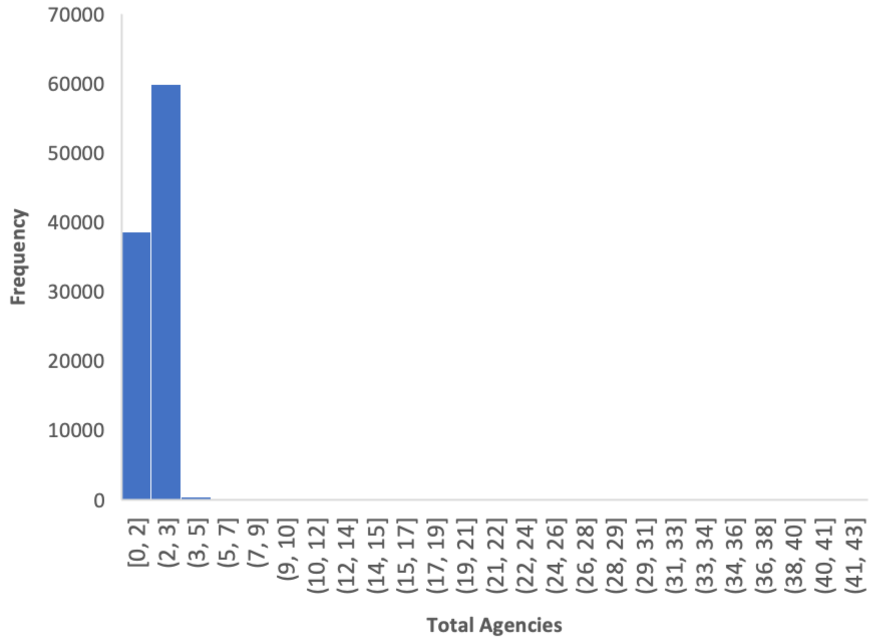


Figure 1 shows the distributions of total regulations for each federal agency in the network. The minimum value was one regulation, the maximum value was 24,744 regulations, the median value was twenty-two regulations, and the mode was one regulation. The average was 522 regulations, but since this value 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 ninety-eight percent 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 10 Shared Regulations Network

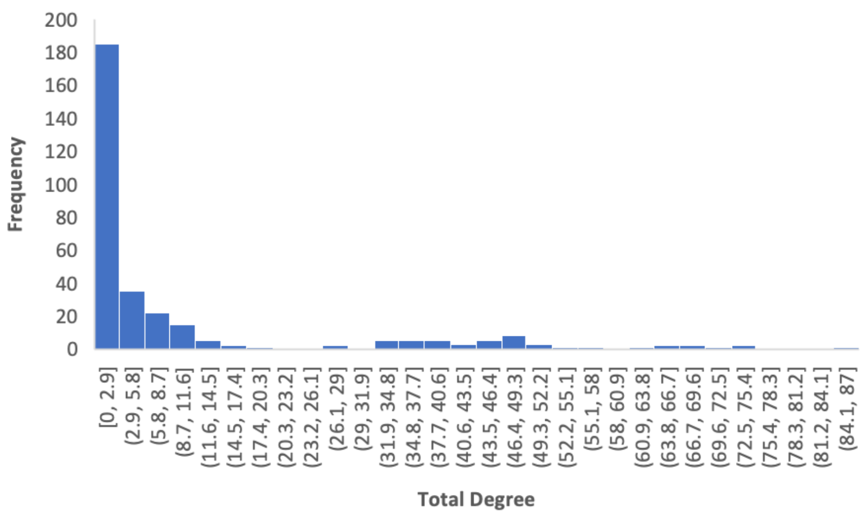
General Statistics	
Total Federal Agencies:	453
Density:	0.014
Link Statistics	
Total Links:	2,906
Link Values:	Min: 0; Max: 18,506; Mean: 49.48; Std Dev: 571.17 Sum: 143,798
Component Statistics	
Isolates:	222
Dyads:	2
Triads:	1
Larger:	1
Larger Sizes:	Min: 224; Max: 224; Mean: 224; Std Dev: 0

Table 1 provides summary statistics for the shared regulations network. The shared regulations network was sparse, having a density of .014, suggesting that U.S. federal agencies do not collaborate when enacting regulations, as only one percent of potential links were present. There were 222 isolates in the network, meaning that roughly fifty-one percent 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.

Table 11 Network Measures

Network Measure	Definition
Total Degree Centrality	The sum of a node's out-degree centrality (row sum) and in-degree centrality (column sum).
Ego Betweenness Centrality	The betweenness score of a node within its own ego network (includes the node itself, its immediate neighbors, and all links between them).
Eigenvector Centrality	Calculates the central eigenvector of the adjacency matrix; a node is central to the extent that its neighbors are central.
Effective Network Size	The effective size of a node's ego network based on redundancy of ties.
Triad Count	The number of distinct triads associated with a node.
Clique Count	The number of distinct cliques associated with a node.

Figure 18 Degree Distribution



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).

Table 12 MRQAP Results

Types of Collaboration	Coef.	P-value
Intercept	5.745	**0.04**
Shared Organization	16.503	**0.00**
Department and Department	9.618	**0.01**
Independent and Independent	-0.294	0.45
Subordinate and Subordinate (External)	-0.558	0.58
Subordinate and Subordinate (Internal)	-15.020	1.00
Parent and Child	253.492	**0.00**
Department and Independent	-3.578	0.90
Department and Subordinate (External)	-1.887	0.71
Independent and Subordinate (External)	-4.889	0.97

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 five percent 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.

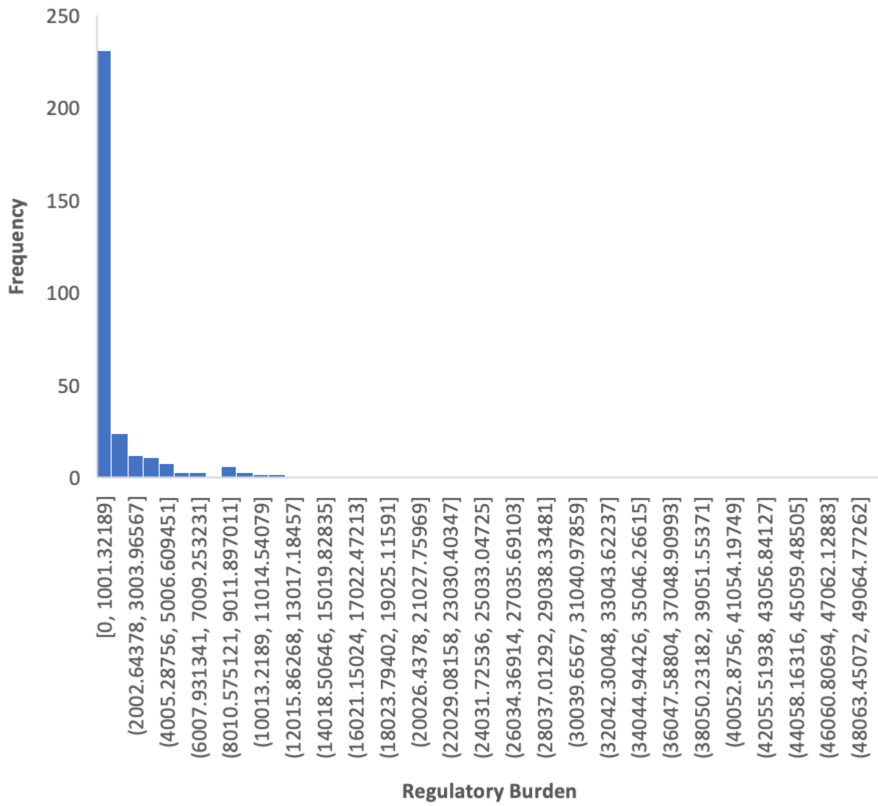
A preferred method of detecting overlap between federal regulations requires projecting our bipartite networks into one-mode networks that capture their shared agency and shared keyword relationships. The disadvantage of bipartite network projection is that it does not scale well for large and dense 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.

Figure 19 Regulatory Burden

$$\boldsymbol{A} = Agency\ x\ Regulation\ Network$$
$$\boldsymbol{B} = Agency\ x\ Keyword\ Network\ (Binarized)$$
$$\boldsymbol{Total\ Regulations} = Row\ Sum\ (A)$$
$$\boldsymbol{Exclusive\ Regulations} = Row\ Sum\ (A[degree == 1])$$
$$\boldsymbol{Total\ Keywords} = Row\ Sum\ (B)$$
$$\boldsymbol{Exclusive\ Keywords} = Row\ Sum\ (B[degree == 1])$$
$$\boldsymbol{Shared\ Keywords} = Total\ Keywords - Exclusive\ Keywords$$
$$\boldsymbol{Regulatory\ Burden} = \frac{Exclusive\ Regulations}{Total\ Regulations}\ x\ Shared\ Keywords$$

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 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.

Figure 20 Distribution of Regulatory Burden



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.

Figure 21 Summary of Poisson Generalized Linear Model

Probability Density Function

$$P(y;\mu) = \frac{\exp(-\mu)\mu^{y}}{y!}$$

Systematic Component:

$$y \sim Pois(\mu)$$
$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)$$

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 & Smyth 2018).

Results

Figure 22 Poisson Regression Model

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       6.5838907  0.3652295  18.027  < 2e-16 ***
independent       0.6549157  0.3222765   2.032 0.043010 *
subordinate      -2.8412030  0.5083749  -5.589 5.10e-08 ***
clique_count      0.1525042  0.0420095   3.630 0.000332 ***
effective_network -0.0723904  0.0248102  -2.918 0.003789 **
ego_betweenness   1.9373259  0.2891533   6.700 1.02e-10 ***
eigenvector       8.6949488  4.2458379   2.048 0.041434 *
total_degree     -0.0003452  0.0001051  -3.286 0.001136 **
triad_count       0.0041638  0.0008096   5.143 4.86e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

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 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 percent higher than baseline, on average, while that imposed by subordinate agencies were ninety-two percent lower. Increases in effective network size are associated with a five percent 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 eigenvector centrality values range from zero to one and increases in these measures result in the most significant increases in burden within our model.

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.

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.

Chapter 5: Measuring the Reach & Engagement of U.S. Federal Agencies on Twitter

Introduction

In 2009, the Obama administration directed federal agencies to harness technology to place information about government operations and decision-making online and readily available to the public through the Open Government Initiative (Obama, 2009). Under the initiative, federal agencies established a presence on major social networking sites, including Facebook, Twitter, and YouTube. This and other efforts to make government more transparent have created research opportunities for analyzing and measuring the effectiveness of government social media practices across the myriad platforms in existence. Previous research identified that government use of social media has been directed largely at service provision and informing citizens about government services, while neglecting citizen engagement (Brainard & McNutt, 2010).

Efforts to increase citizen engagement are a recent trend within public administration, and research towards this end has increasing prevalence within the public administration literature. Government use of social networking sites falls under the umbrella term “E-government” whose goal is extending government services and reaching a greater audience to offer government information and engage the public regarding government efforts (Bertot et al., 2010). The federal government has faced scrutiny for its slow adoption of social media and its reluctance to use the full functionality of social networking applications (Mergel, 2012). Current practice across the federal government also fails to engage citizens directly (Mergel, 2013).

Mergel (2013) was among the first to study government use of social media to engage stakeholders focused specifically on the executive branch. Our research is a direct extension of her prior work which considered the formal and informal information sharing mechanisms used by social media directors to observe and share best practices. Mergel interviewed the social media directors of all fifteen executive departments and found that social media directors' target audiences were those using social media to receive information on their newsfeeds in lieu of visiting their official website; therefore, social media platforms were used to disseminate policy statements or major press releases. Further, social media directors observed one another's online behavior and emulated practices that fit into their strategies, but there were very few role models within government to mirror an interactive engagement approach. Finally, social media directors held that reciprocated feedback and interaction was a desirable goal for their social media use.

Implementing social media in government requires more than simply creating a social media account with an officer to upload content (Goncalves et al., 2015).

Mergel suggested that additional research is necessary to understand the implications of social media strategies, including whether they lead to increased transparency, accountability, participation, and collaboration in government and if social media interactions are designed effectively enough to reach the right audiences of a government agency. Mergel also identified the need for appropriate metrics that will enable the federal government to understand the impact of its social media as well as its use of social media platforms.

Having identified these gaps in knowledge, this study will make an empirical analysis of the federal government's social network on the Twitter application. This paper focuses on the Twitter application because it is listed among the top websites and is the leading microblogging site globally (Alexa, 2022). Twitter has approximately 290 million monthly active users worldwide, with 37 million users living in the United States and representing the target audience of this research. In addition to analyzing the government's use of Twitter, this research will also expand the sample from executive departments to the entire population of federal agencies using the Twitter platform. In this manner, we make an extension of the previously cited research in this area. The objective of this paper is to establish a baseline measure of performance for the federal government's use of Twitter by answering the following research questions:

RQ1. How do we characterize the federal government's aggregate use of the Twitter platform?

RQ2. How do federal agencies use Twitter for communication and engaging the public?

RQ3. How does the public engage with federal agencies on Twitter based on the information that federal agencies post on the platform?

Our findings will inform federal government social media managers about their social media presence relative to other federal agencies within the executive branch. Further, this study will establish a baseline performance and benchmark by which federal agencies may measure future changes in strategy on the platform. Finally, this study will consider which of Twitter's available mechanisms are most amenable to engaging with federal agencies' target audience, given that agencies make full use of the engagement mechanisms available on Twitter.

Background

There is scarce literature on public organizations' use of Twitter for public relations. There is a vast amount of research that considers political discourse around emergent topics, including elections (Tumasjan et al., 2011; McKinnon et al., 2016), pandemics (Kim et al., 2021; Osakwe et al., 2021), and other emergencies (Panagiotopoulos et al., 2016; Liu & Xu, 2019) to name a few. Several studies consider the activity of prominent individuals on Twitter, namely politicians or other public officials (Kim et al., 2021; Brookey & Ott, 2019). A small number of studies consider the strategies employed by organizations on the platforms (Bonson et al., 2013), and even fewer cover public organizations (Bonson et al., 2017).

Linders (2012) and Mergel (2013) have conducted seminal research within this subject area, and whose research offers a theoretical foundation for citizen engagement using social media platforms. Linders proposed a topology of citizen e-participation based around three distinct models: citizen sourcing which reflects a citizen to government model where citizens are enabled to dialog and share opinions with public officials on social media; government as a platform which reflects a government to citizen model which enables citizens through make informed decision through data-sharing and arming them with information; and do-it-yourself government which reflects a citizen to citizen model where the public self-organizes around issues of importance to them at a particular point in time. For her part, Mergel identified that public organizations employ one of three social media strategies: a push strategy that uses social media as an additional channel of communication to post information for constituents; a pull strategy that seeks to solicit feedback from constituents; and a networking strategy that mixes both push and pull strategies to create an environment of interaction and bidirectional responsiveness and generate reciprocal feedback cycles.

Other research has suggested that social media promotes transparency and improvement in the government provision of services (Bertot et al., 2010) as interactions on social media are bi-directional and afford frequent communication and feedback between government officials and the public (Bonson et al., 2017). Further, social media enables citizen-created content that enriches socio-political debate and offers diversity of thought on a vast number of subjects (Bonson et al., 2012), because of coproduction and crowdsourcing of solutions and real-time information updates (Mergel & Bretschneider, 2013; Bonson et al., 2017). Although government entities have increasingly embraced social media as a tool for communicating and engaging with citizens about public policy, they are doing so in large part through an antiquated policy structure that fails to reach a substantial portion of their constituents (Goncalves et al., 2013; Goncalves et al., 2014). Government agencies have discovered that implementing social media requires more than creating an account and uploading content; investment and sound policy may have just a marginal influence on the success of social media use in government (Goncalves et al., 2015; Hosio et al., 2014). There is a clear gap in knowledge and practice regarding the dialog between government and their constituents on social media.

Previous research has considered what makes a social media post effective. For instance, user reactions have large influence on the success of a post (Liu et al., 2014; Ma, 2013). As users become inundated with a vast number of posts, they are likely to sacrifice paying attention to certain posts in favor of others. Users may pay closer attention to the posts in which other users in their social network are responding (Salganik et al., 2006). Therefore, the successful post is one which gets exposure but also solicits a response (Romero & Galuba, 2011; Venkatanathan et al., 2012; Wen & Lin, 2010). Research has also determined that media richness produces greater engagement on social media (Simon & Peppas, 2010; Goncalves et al., 2015; Ma, 2013). Users initiate more positive attitudes and higher level of satisfaction towards sites providing richer media. Plain text is no longer the best type of medium to articulate information and that multimedia content is the most powerful tool to enhance the potency of a given message. Regarding the Twitter application, studies have found that retweets were the most frequent way for constituents to interact with the government entity, that photos and videos generated more engagement than other media types, and that sports and environmental issues were the most engaged topics (Davison, 2007; Cho et al., 2009; Bonson et al., 2019).

Methods

This research began with a complete list of federal agencies within the executive branch. This list was derived from the U.S. Federal Register Database, which is the central repository for all federal activity. In addition to maintaining metadata on all federal agencies, the register also maintains a record of all government correspondence, including regulations and public notices. It is one of the many e-government mechanisms that exist for constituents to remain apprised of the federal government's operations. Next, I mapped each federal agency within the database to the list of active Twitter accounts. There was a total of 241 federal agencies with active Twitter accounts. This represents 53 percent of the federal agencies who had also enacted a regulation. Also, 84 percent of the Twitter accounts were verified accounts on the platform.

Having the 241 twitter handles, I next used the Twitter API to map each handle to its unique identifier, which allowed me to collect additional metadata on each account, as well as granted me access to the previous thirty days of activity on each account's timeline. The metadata that I used for this chapter included the time that the account joined Twitter (referred to in this chapter as the tenure), the total followers, and the total following. I also collected the previous four weeks of Twitter activity (from Feb 6 – Mar 6, 2022) – a total of 8,556 tweets across the accounts.

Table 13 Tweet Statistics

Descriptive Statistics (Tweets) Total: 8,556 Tweets					
	Average	Median	Minimum	Maxixum	Standard Deviation
Tweets	36.41	34.00	0.00	99.00	28.93
Responses	4.78	0.00	0.00	62.00	10.16
Likes	2285.87	146.00	0.00	155414.00	11935.19
Replies	258.57	10.00	0.00	10256.00	1090.98
Retweets	1779.48	178.00	0.00	77454.00	6746.09
Quotes	98.22	10.00	0.00	6634.00	475.80
Has URL	24.14	19.00	0.00	93.00	21.68
Has Hashtag	17.86	11.00	0.00	95.00	18.96
Has Mentions	18.92	12.00	0.00	88.00	19.52

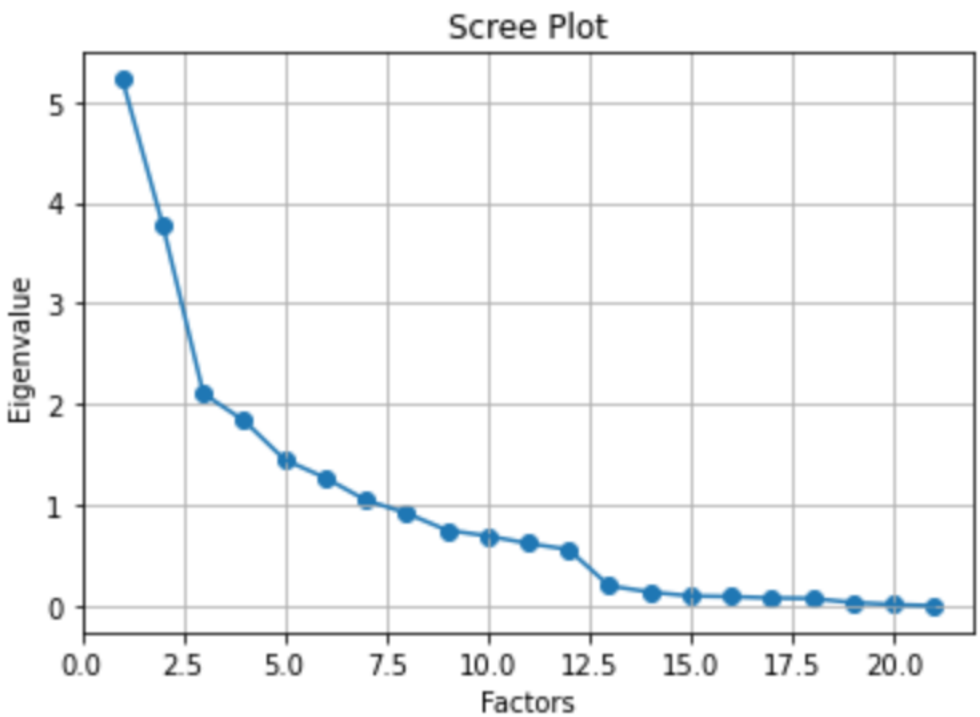
The tweet object returned by the API allows for the introduction of novel metrics which, to the best of my knowledge, have not been proposed in the literature. This chapter researched a total of 21 performance metrics that considered both the reach and engagement of federal agencies. Table 13 provides the summary statistics for federal agency Twitter activity. On average, government agencies generated 36 tweets during the month. I find that of those tweets only a small percentage of them are responses, which are replies made by an agency to a tweet from another account or in response to a reply on one of its own. The likes, replies, retweets, and quotes are right skewed based on the averages being far larger than their medians. On median, agencies receive far more likes and retweets than they do replies and quotes. Additionally, based on the average number of tweets, and the averages for whether a tweet has a URL, hashtag, or mention, that most tweets incorporate a URL. There is far less use of the hashtag and mention.

Table 14 Engagement Metrics

Engagement Metrics	
Metric	Formula
URL Usage	Has URL / Tweets
Hashtag Usage	Has Hashtag / Tweets
Mentions Usage	Has Mention / Tweets
Response Ratio	Responses / Tweets
Responsiveness Ratio	Responses / Replies
Like Ratio	Likes / Tweets
Reply Ratio	Replies / Tweets
Retweet Ratio	Retweets / Tweets
Quote Ratio	Quotes / Tweets

Nine of the 21 performance metrics presented in this chapter are engagement metrics. Table 14 lists the engagement metrics and their formulae. The usage ratios for URL, hashtag, and mentions measure the likelihood of an agency engaging in a particular form of media based on its previous usage in the sample. In the same manner, the response ratio measures the likelihood of an agency tweets being a response, and the responsiveness ratio measures the likelihood of an agency responding should some reply to one of its tweets. The usage, response, and responsiveness ratios are measures on each agency. On the other hand, the like, reply, retweet, and quote ratios are measures on the population who engages with the federal agency (i.e., the likelihood of an agency receiving a like, reply, retweet, or quote).

Figure 23 Factor Analysis Scree Plot



This chapter uses factor analysis as a method for reducing the dimensionality of the performance measures. Factor analysis allows for the consolidation of these performance metrics into a few factors representing latent behaviors which may account for the activity of federal agencies and stakeholders on the Twitter platform. Factor analysis extracts the maximum common variance from all variables and puts them into a common score. I used the Bartlett Sphericity and KMO statistical tests to validate that this set of data is a good candidate for factor analysis. The scree plot in Figure 23 shows that this dataset may be represented as six distinct factors, based upon their associated eigenvalues.

Figure 24 Factor Analysis Loadings

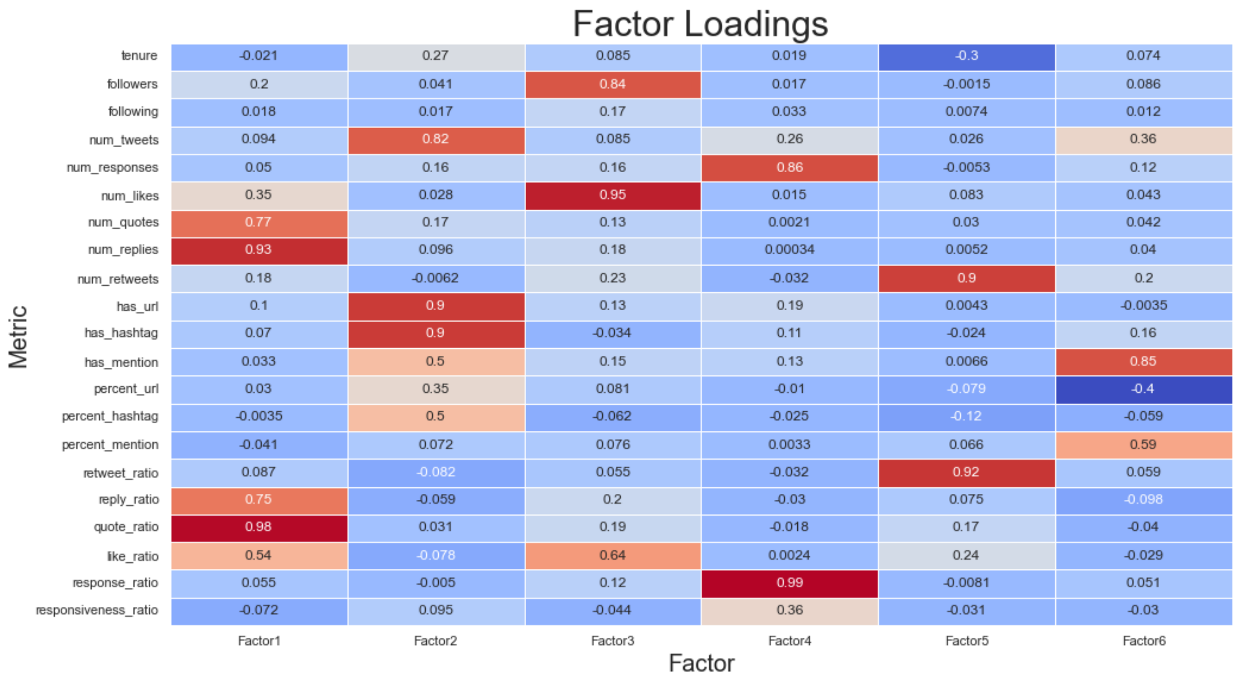
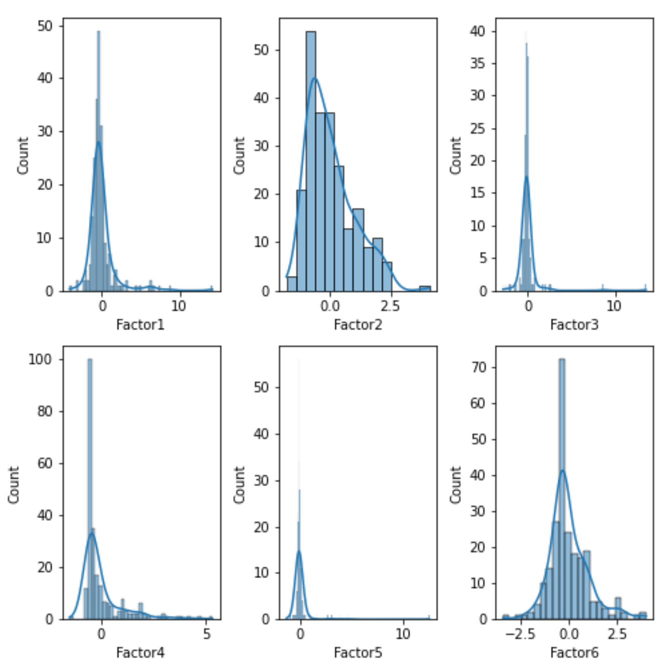


Figure 24 depicts the factor loadings which assigns a correlation coefficient ranging from -1 to 1 to each performance metric. The metrics with the highest correlation explains the latent behavior of the factor. I represent these factors as a heat map to provide a better graphical depiction of these relationships. Factor #1 represents “active agency engagement” based on the high correlation with the like, reply and quote metrics. The term active agency engagement reflects the fact that citizens use the capabilities to directly engage federal agencies on the Twitter platform. Factor #2 represents “passive citizen engagement” based on the high correlation with the tweet, URL, and hashtag metrics. The term passive citizen engagement reflects the fact that agencies use the capabilities to passively engage their stakeholders on the Twitter platform. Factor #3 represents “popularity” of an agency, based on the high correlation with the number of followers as well as the number of likes that its tweets receive. Factor #4 represents “active citizen engagement” based on the high correlation with responses and suggests how an agency actively dialogues with its stakeholders about policy on the platform. Factor #5 represents “passive agency engagement” based on the high correlation with the retweet metric and suggests how stakeholders might passively interact with an agency on the platform. Factor #6 represents “collaboration” based on the high correlation with mentions and suggests how well agencies may include other agencies into policy discussions.

Figure 25 Factor Distribution



Finally, I use the factor loadings to assign a unique value for each factor to each federal agency. Figure 25 shows the distribution for each factor. The values for each factor range from negative to positive number; for the sake of interpretation, we assume ordinality of the measures (e.g., larger number reflect better scores relative to other agencies within the data). While some factors may appear normally distributed, I used the Shapiro test of normality to determine that none of the factors are normally distributed. Finally, I performed statistical tests to determine whether there were significant differences in platform use between organizational and account types.

Results

Table 15 Engagement metrics by organizational type

	Department	Independent	Subordinate	P-value
Active Citizen Engagement	-0.12	0.22	-0.18	**0.04**
Passive Citizen Engagement	0.34	-0.26	0.19	**0.00**
Active Agency Engagement	1.19	-0.18	0.01	0.13
Passive Agency Engagement	-0.13	0.08	-0.06	0.82
Popularity	-0.46	0.09	-0.02	**0.01**
Collaboration	0.60	-0.08	-0.01	0.11

Table 15 shows the results of a Kruskal-Wallis test of the engagement metrics against the organizational types. This chart shows that there are significant differences in how departments, independent, and subordinate agencies use the Twitter platform, based on three of the five metrics under consideration. On average, independent agencies show higher levels of active citizen engagement relative to departments and subordinate agencies. Independent agencies are also more popular on Twitter. Departments and subordinate agencies show higher levels of passive citizen engagement relative to independent agencies. While the differences in agency engagement are not significant, the measures do reflect that the public seems to have greater interest in actively responding to departments and more passively respond to independent agencies; also, departments seem to be the most collaborative, followed by subordinate agencies.

Table 16 Engagement metrics by account type

	Verified	Unverified	P-value
Active Citizen Engagement	0.05	-0.29	0.22
Passive Citizen Engagement	0.10	-0.55	**0.00**
Active Agency Engagement	0.09	-0.50	0.24
Passive Agency Engagement	0.01	-0.08	0.16
Popularity	0.01	-0.08	0.90
Collaboration	0.03	-0.18	0.46

Table 16 shows the results of a Kruskal-Wallis test of the engagement metrics by account types. There is a significant difference between verified and unverified accounts in terms of passive citizen engagement. It is also worth noting that verified accounts seemed to outperform unverified accounts in all measures under consideration.

Discussion

This chapter considers the reach and engagement of federal agencies on the Twitter platform. The objective of this chapter is to find suitable measures of how well agencies communicate policy to their stakeholders. The literature refers to it as civic engagement. Dr. Ines Mergel connected the civic engagement research to federal agencies when she interviewed the social media mangers of the fifteen executive departments to determine the strategies and any latent behaviors associated with their use of the platform. This study occurred in 2013 and she found that most federal agencies utilize Twitter as an extension of their official websites, and that their target audiences were individuals seeking to use their timelines on the platform as a supplementary form of news. In short, departments were not using the platform to engage and dialogue with their stakeholders, only to notify them and keep them informed.

As government seeks to become more data-driven, the goal shifts from notification and information to active dialogue. The research points at a need for metrics that measure how well federal agencies utilize social networking sites. Dr. Mergel found that across the departments, they all hoped to improve how well they leveraged the platform. Unfortunately, my research has found that their strategies have not changed much over the last decade. Departments still use the platform primarily for notification and information purposes.

I began this research by looking at the timelines of federal agencies with active Twitter accounts. There were over 8,000 tweets, and the average agency seems to post at least once per day. I have presented novel metrics which have not been used before in the literature to analyze government strategic communications on a social networking platform. The metrics consider both the usage and engagement. Usage refers to how well federal agencies leverage the functionality that Twitter makes available to users of the platform, e.g., tweets, hashtags, URLs, and mentions. Engagement refers to how well federal agencies leverage the relationship aspects of the platform, e.g., replies, retweets, quotes, and likes. In terms of usage, federal agencies use the tweet and the URL most often, while citizens respond by liking and retweeting most frequently. In this manner, policy communications propagate throughout the network. If we were to consider areas in which federal agencies could improve their use of the platform, the most improvement may be found by increasing their use of replies, hashtags, and mentions. This allows federal agencies to move from passive to active use of the platform.

I used a factor analysis to reduce the dimensionality of the metrics to more meaningful measures. Factor analysis looks for latent variables that could represent the observed behaviors in the data. In our case, six factors represent the preponderance of the combined variable found in the data. The heatmap in Figure 24 shows how each metric maps to a particular factor as well as how strongly correlated that metric is with that factor. One caveat about factor analysis is that its results are open to interpretation. I have tried to present a meaningful interpretation for these factors based on my experience collecting and working with this data over previous chapters, but a different researcher may look at this data and interpret it in a different way.

The measurements that I determined (and which I think are suitable to this data) are active and passive citizen engagement, active and passive agency engagement, popularity, and collaboration. The citizen engagement refers to how agencies engage their stakeholders, the agency engagement refers to how stakeholders engage agencies, the popularity refers to how the public views agencies, and collaboration refer to how well agencies include others in the conversation. The fact that departments are highest in passive collaboration supports the findings of previous research, namely that department strategy leverages the platform simply as an extension of other forms of information and notification. Independent agencies are the most popular and the most responsive, but I can’t say for certain whether popularity leads to responsiveness or responsiveness leads to popularity. This would require additional testing, which is left for future research.

There also seems to be a difference in how the public responds to agencies. We see that the government seems to respond more actively to departments. I have not taken the additional steps to consider why this might occur. Future research might consider the sentiment of the public response, e.g., are they happy with the policy choices that departments notify them of online. Also, like our findings in previous chapters, agencies are also not collaborative online. Agencies made little use of the mentions feature on Twitter, nor did they make use of the hashtag. One limitation of the research presented in this chapter is that expert judgment is required at several stages of the methodology to establish a benchmark for the measure. For example, is one tweet per day good or bad? Is there a reason why certain agencies don’t use hashtags or mentions, while others do? What constitutes a suitable level of engagement? Do agencies have the resources to enact more aggressive social media strategies? These are questions which result from my findings, and which future research should consider.

Chapter 6: Implications & Future Research

This thesis proposes that the government can improve its performance by exploiting its network structure. To support this assertion has required that we consider the interactions between government entities as occurring within a complex system. I use model this system of interactions as a network and the network methodology offers us several ways of analyzing and measuring the behaviors which take place within this system. This methodology offers a different perspective than that of traditional policy analysis. This thesis considers the interactions of U.S. federal agencies within a regulatory and strategic communications context.

An open research problem in the field of public administration and policy is finding ways to reduce government overlap. Overlap may occur in the government provision of goods and services, as well as in government regulation. The existing literature has written about overlap in each of these areas. The US Government Accountability Office (GAO) is a legislative agency who acts as a watchdog over the federal government operations on behalf of the legislative branch. The GAO breaks down three different types of overlap – fragmentation, duplication, and overlap. Fragmentation involves multiple agencies being responsible for an aspect of a particular policy. Duplication involves multiple agencies providing the same product or service. Overlap involves multiple agencies whose responsibilities conflict with one another. This is a problem that the US federal government, and indeed other governments internationally, are seeking a manageable solution. My research has considered how we detect overlap in government regulations, without attempting to categorize the overlap. This may be considered a limitation of my research; that while I propose what I consider an efficient method for detecting overlap, I leave the categorization of overlapping regulations to future research.

One way to limit overlap is to increase the level of collaboration on government regulations. This is certainly a challenging practice to implement, given the autonomy with which government agencies operate. There is also the question of whether the lack of collaboration that we observed a bad thing. The legislature grants each agency with a mandate for regulating within its functional area. Whose responsibility is it to know when one federal agency has an equity in the policy decisions of another. This is a critical point that has been brought up, but which my research does not consider. My research has presented performance measures to gauge both regulatory activity and collaboration within that context. I have found that there are significant differences between organizational types and their regulatory activity. For example, departments enact the greatest number of regulations, but how much is too much? This would require someone in government to establish some benchmark or threshold value above which there is increased scrutiny upon the agency. We recognize that there is some baseline level of regulatory activity to be expected, but we don't know what's reasonable? I believe that my research puts us on a path to begin answering those questions.

Additionally, we know that independent agencies enact the fewest regulations, on average. What actions should the government take about independent agencies who enact regulations at the level of a department? Should those agencies in fact operate with such autonomy? Should they be a department and report to the president, or based on their keyword usage, should they be rolled under some other department or subordinate agency. There certainly is precedent for this. I believe that my research opens the door to answer this question. I conceptualize the notion of regulatory burden to begin moving in that direction. The regulatory burden is a composite measure which considers both agency collaboration as well as legal complexity. I found that the most burdensome agencies within the system are independent agencies, many enacting regulations in isolation and in greater number than executive departments. In my opinion, this is problematic. And this is only at the federal level. I expect that when future research considers the case of vertical collaboration and overlap, this burden will be worse.

This research found that the lack of collaboration also occurs in the strategic communication context. We witnessed the implications of a lack of collaboration in the government response to the pandemic. There was not a unified response and agencies appeared to offer competing guidance. I considered strategic communication on the Twitter platform. My research finds that the government social media strategy has changed little since 2013. My research presents several metrics to analyze the reach and engagement of federal agencies on the platform. The use of these metrics requires a bit of judgment on the part of the researcher, which is one limitation of applying any metric; someone must determine the benchmark or threshold value. In this case, these metrics are best considered relative to other government agencies in the system. Maybe there is some baseline level of behavior expected of a government agency as well as its stakeholders who use the platform? There may not be the expectation that agencies and their stakeholders use the platform in the same manner as a social media influencer, per se. However, there are improvements that the government can make to its social media strategy. This is evidenced by the low number of hashtags that they use, they low number of mentions, and the low response and responsiveness. The federal government has a high level of passive engagement, which might suggest that posting on the platform is simply a block check to suggest that the government is being open. These are the challenges that future research should attempt to address, and I believe that my research has opened the door for this.

My research has made several important contributions to the existing research. I have created performance measures for public collaboration that may improve how governments regulate and communicate those regulations to stakeholders, thereby reducing the occurrence of government overlap and online misinformation. Moreover, I have shown how a more advanced use of network methodology might be used to further policy studies and how networks might integrate within the public policy process. I have shown that there are latent and dynamic networks existing within the regulatory system; understanding these networks should lead to more informed decisions about policy collaboration and communication, as well as organizational design. I have shown how the network methodology may be used to address wicked policy problems – in this case, regulatory overlap and strategic communications, and presenting novel concepts related to those problems (e.g., regulatory network, regulatory overlap, regulatory burden, and associated performance measures). Finally, I have created novel data sets for future research spanning the disciplines of law, public policy, public relations, and social cybersecurity.

References

Administrative Procedure Act (APA), 5 U.S.C. § 551 – 559 (1946).

Al-Ubaydli, Omar and McLaughlin, Patrick A. 2017. RegData: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997 – 2012. *Regulation & Governance, 11(1): 109 – 123.*

Alexa, 2022. "Alexa Top 500 Global Sites." http://www.alexa.com/topsites.

Barabasi, Albert L. 2009. Scale-free networks: A decade and beyond. *Science, 325, 412-413*.

Bertot et al. 2010. Social Media Technology and Government Transparency. *Computer* 43(11): 53 – 9.

Bommarito, Michael J., and Katz, Daniel M. 2010. *Properties of the United States Code Citation Network.*

Bonson et al. 2012. Local e-government 2.0: social media and corporate transparency in municipalities. *Government Information Quarterly*, 29, 123 – 132.

Bonson et al. 2013. A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review* 37(5): 787 – 803.

Bonson et al. 2017. Facebook Practices in Western European Municipalities: Empirical Analysis of Activity and Citizens' Engagement. *Administration & Society* 49(3): 320 – 347.

Bonson et al. 2019. Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities. *Government Information Quarterly*, 36, 480 – 489.

Borgatti, Stephen P. 2005. "Centrality and Network Flow." *Social Networks 27: 55-71*.

Borgatti, Stephen P., Martin G. Everett., and Jeffrey C. Johnson. 2018. *Analyzing Social Networks*. London: SAGE Publications Ltd.

Brainard, L. & McNutt, G. 2010. Virtual government-citizen relations: Informational, transactional, or collaborative? *Administration & Society, 42*, 836 – 858.

Brass, D.J. 1984. "Being in the right place: A structural analysis of individual influence in an organization." *Administrative Science Quarterly 29: 518-539.*

Breiger, R.L. 1974. "The duality of persons and groups." *Social Forces 53(2): 181-190.*

Brookey, R. & Ott, B. 2019. Trump and Twitter 2.0. *Critical studies in media communication*, 36(1), 92 – 96.

Brown, Gordon. 1994. Canadian Federal-Provincial Overlap and Presumed Government Inefficiency. *Publius: The Journal of Federalism, 24(1), 21 – 38.*

Burt, Ronald. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press.

Business Roundtable. 2019. *Reducing regulatory overlap in the 21st century*.

Carey, Maeve P. 2013. "The Federal Rulemaking Process: An Overview." Congressional Research Service Report for Congress, June 17, 2013.

Carey, Maeve P. 2019. "Counting Regulations: An Overview of Rulemaking, Types of Federal Regulations, and Pages in the Federal Register." Congressional Research Service Report for Congress, September 3, 2019.

Carley, Kathleen. 2006. "A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action." Carnegie Mellon: Institute for Software Research.

Carley, K., Cervone, G., Agarwal, N., Liu, H. 2018. "Social Cyber-Security." SBP-BRiMS.

Carley et al. 2013. Users Guide 2013, Carnegie Mellon University, School of Computer Science, Institute for Software Research Technical Report, CMU-ISR-13-108.

Cho et al. 2009. Media richness, user trust, and perceptions of corporate social responsibility: An experimental investigation of visual web site disclosures. *Auditing & Accountability Journal*, 22(6).

Commonwealth of Australia. 2014. *The Australian Government Guide to Regulation*.

Congressional Review Act (CRA), 5 U.S.C. § 802 (1966).

Davis, A., Gardner, B.G., Gardner, M.R. 1941. *Deep South: A Social Anthropological Study of Caste and Class*. Chicago: University of Chicago Press.

Davidson, Roger H., Walter J. Oleszek, Francis E. Lee, and Eric Schickler. 2014. *Congress and its Members*. Thousand Oaks, CA: CQ Press.

Davison, J. 2007. Photographs and accountability: Cracking the codes of an NGO. *Accounting, Auditing & Accountability Journal*, 20(1), 133 – 158.

Derudder, B., Taylor, P. 2005. "The cliquishness of world cities." *Global Networks 5: 71-91*.

Dunn, Pete K., and Smyth, Gordon K. 2018. *Generalized Linear Models*. Springer.

Epstein, Richard. 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*.

Executive Order 13771, Vol. 82, No. 22 (February 3, 2017).

Federal Register Act (FRA), 44 U.S.C. §15 (1935).

Fowler, J.H. 2006. "Legislative co-sponsorship networks in the U.S. House and Senate." *Social Networks 28: 454-465*.

Freeman, L.C. 1979. “Centrality in social networks: Conceptual clarification.” *Social Networks 1: 215-239*.

Goncalves et al. 2013. Narrowcasting in Social Media: Effects and Perceptions. In *Proceedings of the 2013 ACM International Conference on Advances in Social Network Analysis and Mining*. New York: ACM, 502 – 9.

Goncalves et al. 2014. Eliciting Situated Feedback: A Comparison of Paper, Web Forms and Public Displays. *Displays* 35 (1): 27 – 37.

Goncalves et al. 2015. Increasing the Reach of Government Social Media: A Case Study in Modeling Government-Citizen Interaction on Facebook. *Policy and Internet*, 80 – 102.

Government Accountability Office. 2015. *Evaluating and managing fragmentation, overlap, and duplication*.

Hayden, Gregory. 2002. “Policymaking Network of the Iron-Triangle Subgovernment for Licensing Hazardous Waste Facilities.” University of Nebraska – Lincoln Faculty Publications.

Heclo, Hugh. 1978. “Issue Networks in the Executive Establishment.” *The New American Political System: 87-124*.

Hollander, Robyn. 2010. Rethinking Overlap and Duplication: Federalism and Environmental Assessment in Australia. *Publius: The Journal of Federalism, 40(1), 136 – 170*.

Hosio et al. 2014. Exploring Civic Engagement on Public Displays. In *User-Centric Technology Design for Non-profit and Civic Engagements, Public Administration and Information Technology* 6, ed. S. Saaed. Switzerland: Springer, 91 – 111.

Kaplow, Louis. 1995. A Model of the Optimal Complexity of Legal Rules. *Journal of Economics and Organization, 150.*

Katz, Daniel M., and Bommarito, Michael J. 2014. Measuring the Complexity of the Law: The United States Code. *Artificial Intelligence and Law, 22, 337 – 374*.

Kim et al. 2021. Attention to the COVID-19 pandemic on Twitter: Partisan differences among U.S. state legislators. *Legislative studies quarterly*.

Kirkland, Justin H., and Gross, Justin H. 2014. Measurement and theory in legislative networks: The evolving topology of Congressional collaboration. *Social Networks, 36, 97-109.*

Krackhardt, David. 1988. Predicting with networks: Non-parametric multiple regression analysis of dyadic data. *Social Networks, 10, 359-381*.

Latapy, M., Magnien, C., del Vecchio, N. 2008. “Basic notions for the analysis of large two-mode networks.” *Social Networks 30: 31-48*.

Li, Xingxing. 2015. An economic analysis of regulatory overlap and regulatory competition: The experience of interagency regulatory competition in China’s regulation of inbound foreign investment. *Administrative Law Review, 67(4), 685 – 750.*

Linders, D. 2012. From e-government to we-government: Defining a typology for citizen e-participation in the age of social media. *Government Information Quarterly*, 29, 446 – 454.

Liu et al. 2014. Modelling what friendship patterns reveal about personality and social capital. *ACM Transactions on Computer Human Interaction* 21(3).

Liu, W. & Xu, W. Tweeting to engage: How government agencies target stakeholders on Twitter during Hurricane Harvey. *International Journal of Communication*, 4917.

Long et al. 2013. Bridges, brokers, and boundary spanners in collaborative networks: a systematic review. *BMC Health Services Research, 13(1): 158*.

Ma, L. 2013. The diffusion of government microblogging. *Public Management Review*, 15(2), 288 – 309.

McKinnon et al. 2016. Science, Twitter, and election campaigns: Tracking #auspol in the Australian federal elections. *Journal of science communication*, 15(6), 1.

Mergel, I. 2012. Connecting to Congress: The Use of Twitter by Members of Congress. *Journal of Policy Advice and Political Consulting* 3: 108 – 114.

Mergel, I. 2013. Social Media Adoption and Resulting Tactics in the U.S. Federal Government. *Government Information Quarterly* 30 (2): 123 – 130.

Mergel, I., & Bretschneider, S. 2013. A three-stage adoption process for social media use in government. *Public Administration Review*, 73, 390 – 400.

Mizruchi, M.S. 1996. "What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates." *Annual Review of Sociology 22: 271-298*.

Neal, Zachary. 2013. Identifying statistically significant edges in one-mode projections. Social Network Analysis, http://dx.doi.org/10.1007/s13278-013-0107-y.

Neal, Zachary. 2014. "The backbone of bi-partite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors." *Social Networks 39:84-97*.

Newman, Mark. 2018. *Social Networks*. Oxford University Press.

Open Government Initiative. (2009). Retrieved from http://www.whitehouse.gov/open.

Osakwe et al. 2021. Identifying public concerns and reactions during the COVID-19 pandemic on Twitter: A text-mining analysis. *Public health nursing*, 38(2), 145 – 151.

Panagiotopoulos et al. 2016. Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technological forecasting & social change*, 10(111), 86 – 96.

Pitts, F. 1979. "The medieval trade network revisited." *Social Networks 1: 285-292*.

Ramos, J.E. 2003. Using TF-IDF to determine word relevance in document queries.

Romero, D. & Galuba, W. 2011. Influence and Passivity in Social Media. In *Proceedings of the 20th International World Wide Web Conference*. New York: Springer, 18 – 33.

Salganik et al. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311 (5762): 854 – 856.

Simon, S., & Peppas, C. 2004. An examination of media richness theory in product web site design: an empirical study. *The Journal of Policy* 6(4): 270 – 281.

Straughter, Jeremy T., and Carley, Kathleen. 2021. Towards a network theory of regulatory burden. *Applied Network Science 6(70).*

Tullock, Gordon. 1995. *On the Desirable Degree of Detail in the Law*.

Tumasjan et al. 2011. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social science computer review*, 29(4), 402 – 418.

Venkatanathan et al. 2012. Network, personality, and social capital. In *Proceedings of the 4th Annual ACM Web Science Conference*. New York: ACM, 326 – 329.

Wasserman, Stanley and Faust, Katherine. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.