Modelling Cabinet Networks in Parliamentary Democracies

Wojciech Gryc

Trinity College

University of Oxford

A thesis submitted for the degree of

M.Sc. in Mathematical Modelling and Scientific Computing

This dissertation is dedicated to my parents, Jolanta and Zbigniew Gryc.

Acknowledgements

I would like to thank my supervisor, Mason Porter, for his advice and guidance. I would also like to thank Brian Uzzi and Daniel Diermeier for their advice and for the data sets being analyzed in this dissertation. Furthermore, I'd like to thank Peter Mucha for his advice, as well as Kathryn Gillow and all those involved in running the M.Sc. in Mathematical Modelling and Scientific Computing. Finally, I would like to thank the Rhodes Trust for their support.

Abstract

Parliamentary democracies represent a common type of governance structure in numerous countries. While details vary from one country to another, the structure of a parliamentary democracy entails having cabinet ministers who each have a specific portfolio of policy interests, such as healthcare, industry, or education. A set of ministers forms a government, and such governments can change due to elections, political scandals, or health problems. A great deal of work has already explored specific types of ministers, such as prime ministers and foreign ministers, including their survival rates and levels of experience. What is absent from this research, however, is analysis of the network structures surrounding cabinets, and how these structures vary over time. This research project focuses on building mathematical models that incorporate the network structure of ministerial cabinets in twelve parliamentary democracies, analyzing whether such network information can be used to build a comparative analysis of parliamentary democracies or help predict survival rates of cabinet ministers. Governmental and ministerial models are built to analyze cabinet evolution at the country and individual levels. The conclusions reached by both types of models show that network information is extremely useful in predicting both the country- and individual-level qualities of such democracies, including rates of incumbents in governments, as well as political survival rates of individual ministers.

Contents

List of Figures

Chapter 1

Introduction

Social network analysis is a burgeoning academic field helping researchers analyze how different types of actors relate to each other through information on their relationships, transactions, attendance of events, and so on. While social network research has roots that extend back to the 1930s [44], and almost three hundred years further if one considers the origins of graph theory, the last two decades have seen a great deal of growth in the literature on social networks. The premise of such research is that actors or other entities can be represented as nodes in a network, with edges (the connections between the nodes) representing different types of relationships. Such an approach to analyzing and understanding data is extremely useful, and has been explored in numerous different applications [13, 71], from biological food webs, to computer systems, and even actors in the global financial system. An analytical approach that uses networks begins by defining a set of relationships between different actors, and using graph theory, then attempts to reach conclusions about some of the properties of the system itself. This is relevant to political science, where relationships between politicians, departments, and other actors and institutions are key to understanding decision making processes, public policy, and other governmental issues.

With a basic understanding of networks, explored further in Chapter 2, we approach parliamentary democracies and the people who help administrate and run them. Specifically, we are interested in cabinets and their ministers. While the way in which parliamentary democracies operate can vary significantly between different countries, ministers are always responsible for a specific portfolio, such as healthcare, industry, or education, and make important policy decisions about how the government and

the political party in power decides policies surrounding those portfolios. Within a parliamentary democracy, a "government" represents a specific set of ministers in power. Ministers serve this government, and can also be called "cabinet members".

A formal overview of existing political network research is given in Section 2.3. A great deal of research exists on the qualitative characteristics of leaders in ministries [20], as well as the life spans of individual governments [34]. To our knowledge, however, there is a lack of a formal exploration of how the actors within individual cabinets affect each other over time, and whether the connections between them serve any significant roles in ensuring long-term political survival.

With this in mind, our goal is to determine whether network-specific variables provide information on how parliamentary democracies function over time, and whether such variables play a role in cabinet members' survival in office. In the former case, we explore the role of incumbents in governments over time, while in the latter case, we examine whether specific cabinet members return as members in future governments.

1.1 The Mathematics of Networks

Social network data comes in numerous forms, with the two relevant ones to this dissertation being bipartite and weighted, undirected networks. A "graph" or "network" is an object, G, composed of two sets. The first of these are vertices, V, or nodes of the network. The second set, E, is the edge list, which contains objects that signify relationships between two nodes. Edges can be directed or undirected, as well as have weights associated with them. A graph's size n is defined as the size of V.

For example, if we have an undirected graph of size 3 where $V = \{A, B, C\}$ and $E = \{(A, B), (B, C)\}$, then this represents a graph with three nodes, A, B, and C, with a "path" going from A to B to C. Note that in an undirected graph, $(B, C) = (C, B)$. This would not be true in a directed case.

For the purpose of this dissertation, "self-edges", where edges start and end at the same node, are presumed not to exist and are ignored. Specifically, self-edges would represent ministers who work with themselves in a cabinet.

The data being analyzed in this dissertation focuses on cabinet members and the

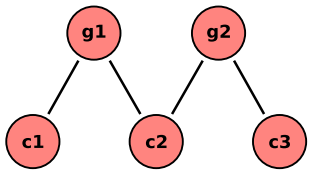


Figure 1.1: A bipartite network with two governments, g_1 and g_2, and three cabinet members, c_1, c_2, and c_3. Observe that c_2 is an incumbent in g_2.

governments they served[1]. This can be represented as a network with two types of nodes, g_i (governments) and c_j (cabinet members). Note that a government, g_i, represents the i^{th} *administration*, rather than the institutions and laws of the country. If a cabinet member c_j served in government g_i, then an undirected edge (c_j, g_i) is present in the network's set of edges. Note that edges cannot connect between two governments or two cabinet members. This type of network is called an "affiliation network", and is a "bipartite graph" [70], where the set of vertices are composed of two subsets, with no edges between members of the same subset.

As indicated, the cabinet data that we analyze in this dissertation is a collection of bipartite graphs. For an individual country, our data consists of a set of cabinet members and the cabinets on which they served. For example, consider a country with two governments, g_1, and g_2, and a total of three cabinet members, c_1, c_2, and c_3. If c_1 and c_3 only served in g_1 and g_2, respectively, and c_2 served in both governments, we would have the bipartite graph shown in Figure 1.1. This yields a graph G with vertex set $V = \{c_1, c_2, c_3, g_1, g_2\}$ and edge set $E = \{(c_1, g_1), (c_2, g_1), (c_2, g_2), (c_3, g_2)\}$.

Bipartite graphs, however, can be projected onto unipartite graphs. If we define the relationship or edge between two cabinet members as existing if and only if they both served in the same government, then we can collapse the above edge set E into a new set $\hat{E}$ such that

$$\begin{aligned} (c_1, g_1), (c_2, g_1) \in E \quad &\Rightarrow \quad (c_1, c_2) \in \hat{E}, \\ (c_2, g_2), (c_3, g_2) \in E \quad &\Rightarrow \quad (c_2, c_3) \in \hat{E}. \end{aligned}$$

Thus, in the collapsed network, we would have $\hat{V} = \{c_1, c_2, c_3\}$ and $\hat{E} = \{(c_1, c_2), (c_2, c_3)\}$. This network is shown in Figure 1.2.

[1] In terms of notation in this report, we use v_i to denote vertices when discussing graphs abstractly, and c_i to denote cabinet members when discussing nodes in the context of a cabinet network.

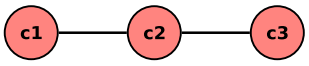


Figure 1.2: A network of cabinet members obtained from the bipartite network shown in Figure 1.1.

In either case, networks with an edge set and node set can also be represented in the form of a matrix, called an "adjacency matrix". In such a matrix, $A = \{a_{ij}\}$, one can arbitrarily order the nodes in V and set the i^{th} row and column to represent the i^{th} node in V. One then sets

$$a_{ij} = \begin{cases} 1 & \text{if } (c_i, c_j) \in E \\ 0 & \text{if } (c_i, c_j) \notin E \end{cases}.$$

A final important property of the networks studied in this dissertation is that when bipartite cabinet networks are collapsed, one can obtain "weighted" networks. This occurs because if we have $(c_1, g_1), (c_2, g_1) \in E$ and $(c_1, g_2), (c_2, g_2) \in E$ — when two politicians serve together in two separate governments — we lose a great deal of information by setting $a_{1,2} = 1$. Instead, we project the network into a weighted one, where the actual value of a_{ij} represents the number of cabinets both nodes served together. Thus, in the aforementioned example, the edge (c_1, c_2) has a value of 2, and all other edges have non-negative integer values, rather than just 0 or 1.

It is useful to note that the above projection of a bipartite to unipartite network is just one approach to this problem. One can set edge weights in other ways, allowing for non-integer values. For example, by normalizing edge weights or taking logarithms of the above values.

1.2 Network- and Node-Level Paradigms

We use two approaches to model cabinet networks:

- **Governmental Level:** In this dissertation, networks represent team collaborations, where each team member is a cabinet minister. Analyzing networks and modelling network-level statistics and properties helps us understand how governments in different countries work and how they differ. By exploring node-level properties and seeing how they aggregate at the "global" (i.e. network) level, one can better understand some of the mechanics of the nodes' decision

making processes, and various limits on how nodes survive, thrive, and make decisions. A key motivation for such work is based on collaboration networks of actors in Broadway musicals [27].

- **Ministerial Level:** Individual nodes in the network represent cabinet members and over time show the evolution of the nodes' political careers. This allows us to make predictions on how long nodes will survive, if they will be promoted, or which will return as cabinet members.

The main goal of this dissertation is to investigate whether structural (i.e. network) properties of governments and individual cabinet members help explain why politicians survive, and how governments work. While other variables can be included in such an analysis, our goal is to explore the effects of networks. Politics is competitive, and being able to navigate various groups (even within cabinets themselves, especially in coalition governments) can make or break a political career. It is intuitive to think that incorporating network measures, and exploring how individual politicians relate to larger groups, can help improve pre-existing models of cabinet formation and political longevity.

Chapter 2

Literature Review

To the best of our knowledge, there has been little work on modelling the social and affiliation networks that pervade cabinets in parliamentary democracies, or governments in general. A key reason for the lack of work in this area is that data surrounding governments and the actors therein is difficult to obtain in a usable format, especially if one tries to do historical research. While parliamentary democracies aim to be transparent and provide information on politicians, historical data is often not digitized, or is stored in formats that are difficult to combine with other data sets.

2.1 Power and Social Networks

Politics and the political survival of cabinet members requires that those members have an acute understanding of the social dynamics of their system. Such dynamics can include how information flow, access to various resources, and understanding the costs of cooperating or competing with other members. A key reason for employing network analysis in the study of social systems is the connection between power, the ability to mobilize resources, and a node's location in a network.

While such network-focused research is relatively new in political science, network analysis has been used in other social systems as early as the 1930s [44, 71]. For example, individual studies have shown that people weakly familiar with or connected to large sets of disparate groups are better at finding jobs [25]; network position helps non-profit organizations succeed [26]; and well-connected managers earn more money

and make decisions that are better implemented in their corporations [12]. Research has delved into the details of the networks pervading these various systems, analyzing characteristics of individual nodes, properties of nodes in relation to other nodes, and "global" characteristics and measures of the entire networks themselves (see, for example, work on the strategic networks of firms [28]).

Some sociologists have gone as far as to promote social network analysis as a way to quantify people's social capital and ability to gain power and status over others [11, 38, 39]. Some common measures related to status in a network are described in Chapter 4.

2.2 Modelling Networks of Teams and Collaborators

To understand a social system, one should go beyond applying descriptive measures on nodes within a network and attempt to model the dynamics of the network itself. We explore models on two levels. First we look at the "governmental" level, where properties of the entire network are modelled. One can analyze the degree (or strength, in the case of weighted networks) distribution of a network, rank nodes, and analyze the properties of such a distribution. For example, Guimerà et al. [27] built models to analyze how Broadway actors work together based on new actors working with established ones, and analyzed the success of the models using degree distributions of actors. Similar strategies have been used to analyze cooperative networks in nature [61].

A second approach to modelling networks is through the analysis of individual nodes. If dealing with temporal data, one can build "longitudinal networks", where edges and nodes vary over time. With such data sets, one can attempt to predict survival rates of specific types of nodes.

2.3 Political Theory and Political Networks

Asking whether the network-based context of a cabinet minister – in relation to other nodes of a specific government, the specific government, and historical properties of

the government – has any effect on the ability of the cabinet minister to succeed (i.e. serve in future governments) is an important question.

Mathematical modelling of governments has received some academic attention, both at the macro level by exploring grassroots support for different types of governments, such as dictatorships and democracies [1], as well as the micro level, by exploring characteristics of individual politicians. One example of the latter approach is the creation of survival models for individual cabinet ministers in the United Kingdom [7], looking at properties of governments, ministerial experience and personal histories, as well as external events to predict how long ministers will serve. Furthermore, analysis of entire governments, such as Congresses in the United States [79], political parties [74], legislative committees and subcommittees [14, 54, 55], as well as supreme court precedents [21], has also been performed. Some of these studies employ network measures like modularity (discussed in Section 4.3).

In the context of cabinet memberships and survival rates of individual ministers, a relatively small body of research exists for modelling coalition governments and the longevity thereof [18]. Indeed, numerous factors and variables affect how governments form following an election, and how parties select their ministers. Institutional constraints, such as internal party politics, constitutional guidelines, and electoral outcomes all play a significant role in whether a politician will become, or continue to be, a minister in a future government [67]. In general, it appears such research is split into two fronts [34]:

1. Cabinets terminate or change due to various types of events, all of which can be treated as random events. As such, the models themselves are stochastic.

2. Cabinet duration is based on the individual properties of the cabinet and its members. Properties could include election results and the institutional constraints mentioned earlier.

Mathematically modelling governments, cabinets, and political systems is a nascent and increasingly active research effort. At this stage, it appears that cabinet members' affiliation networks and prior cabinet experience has not been studied. This is rather surprising because evidence shows that the relationships between ministers has an impact on their survival rates [20], as does their membership in coalition governments [68]. While studies of individual types of ministers, such as foreign

ministers [20], have been performed, studying relationships been all the ministers of specific cabinets seems to be absent [59, 69].

With such a stark absence of relational, network-focused analysis of cabinet members and their long term involvement in cabinets, a pertinent research question is whether such information is useful and helps build fruitful models. This is a key question within this research project, and we try to explore network effects at both the individual node level, as well as the global level mentioned in Section 2.2. With this in mind, we aim to isolate and analyze network variables individually, and investigate if they allow us to build models that outperform baseline results.

Chapter 3

Data

The data being analyzed and modelled in this dissertation is based on cabinet data of twelve parliamentary democracies from 1945 to 1990 [77]. This data was provided to us in a digital format by Dr. Brian Uzzi and Dr. Daniel Diermeier. Parliamentary democracies differ greatly, and may change "governments" – the portfolios of cabinet members – for various reasons, including elections, corruption scandals, or deaths. Hence, the number of governments that were formed during the 45 year period being analyzed differs greatly between countries. Table 3.1 shows the countries being analyzed, the number of governments the period encompasses, and other general information.

One can create graphs by aggregating information over multiple years. Specifically, one can see which ministers worked together in the past and weight edges accordingly. In the case of this dissertation, edge weights represent the number of times the two cabinet members in the edge have worked together up to and including the current government (as discussed in Section 1.1). In previous studies on different networks, researchers [27] defined edges as ones between different types of groups, specifically: `newcomer-newcomer`, `newcomer-incumbent`, `incumbent-incumbent`. Another possibility is to explore inter- versus intra-party edges, as countries that permit coalition governments or have proportional representation voting systems often have governments composed of members from multiple parties.

The biggest concern about the data is its completeness and applicability to relevant modelling questions. The data is composed of membership lists of governmental cabinets in various countries, along with party information and, sometimes, the ministers' portfolios. However, a minister's ability to stay in power, a coalition government's

Country	Govts	Positions	CMs
Canada	18	406	197
Denmark	27	366	156
France	28	532	146
Germany	23	510	224
Israel	35	317	65
Japan	35	508	318
Luxembourg	15	112	55
Netherlands	21	296	159
New Zealand	19	310	139
Norway	23	360	186
Sweden	21	374	127
UK	18	367	206

Table 3.1: Overview of cabinet data. "Govts" represents the total number of governments between 1945 and 1990, "Positions" is the number of cabinet positions available during the period, and "CMs" is the number of unique cabinet members (nodes) that filled those positions.

ministerial composition, and other related variables are often dependent on external factors, such as election results and other aspects of the historical context of the government. Because this data was not available to us, our models are inherently incomplete. If our aim were to build a complete model of governments in parliamentary democracies, we would be required to combine the current data set with other information.

Nevertheless, the data that we have allows us to attempt to answer important questions. Due to the dearth of academic research on network-based or structural variables and their correlation with cabinet member survival rates, our project aims to see if such variables provide *any* useful information when modelling and exploring cabinet memberships. With positive results, it has the potential to improve pre-existing models of politician longevity and cabinet formation theories discussed in Chapter 2.

A discussion on the code written to analyze the data is provided in Appendix B.

Chapter 4

Network Measures

Summary statistics provide a general overview of important network measures. Depending on the mathematical underpinnings of the measures, one can compare different networks to each other, or simply use the information to see how a network evolves over time. A further discussion on some of the measures below is included in Appendix D.

4.1 Centrality and Centralization

The idea behind a "centrality" measure is describing the role that a node plays within a network [71]. Specifically a more "central" node is considered more important, potentially because it is connected to more nodes, controls relationships between disparate groups, or for other reasons. The measures below strive to define a node's centrality in a network in different ways. While this is not a complete list of centrality measures, the choices below all focus on different aspects of a network and show how certain nodal properties can be used to judge the importance of a node to a network.

Once node centralities are calculated for all the individual nodes in a network, one can calculate a graph "centralization", which examines how centralities are dispersed in a network. The goal of a graph centralization is to be able to compare diverse networks to each other.

In general, any measure of centrality can be generalized into a graph centralization

measure using Freeman's theoretical maximum [23].

$$\mathcal{C}_X = \frac{\sum_{i=1}^{n} [C_X(v^*) - C_X(v_i)]}{\max \sum_{i=1}^{n} [C_X(v^*) - C_X(v_i)]} \qquad (4.1)$$
$$v^* = \max_i C_X(v_i)$$

This normalization guarantees that $0 \leq \mathcal{C}_X \leq 1$, because the denominator is the theoretical maximum for a network with n nodes.

4.1.1 Degree Centrality

The simplest centrality measure is degree centrality [71], which represents the number of edges associated with a specific node, also called the node's "degree". That is,

$$C_D(v_i) = \sum_{j=1}^{n} a_{ij}. \qquad (4.2)$$

Note that one can set $C'_D(v_i) = C_D(v_i)/(n-1)$ to normalize the value for a network with n nodes.

In a weighted network one has "strength" rather than degree, where

$$k_i = \sum_{j=1}^{n} a_{ij}. \qquad (4.3)$$

Note, however, that since $a_{ij} \geq 0$ rather than $a_{ij} \in \{0, 1\}$ in the unweighted case, one would have to normalize by the maximum value of k_j for $j = 1, \ldots, n$.

The graph that attains the maximum theoretical value for unweighted degree centrality is a star graph. In an n-star graph, one has n nodes, with one node in the center of the graph and the $n-1$ other nodes only connected to this initial node. The degree of the central node is $n-1$ and all other nodes have degree 1. This yields

$$\max \sum_{i=1}^{n} (C_D(v^*) - C_D(v_i)) = \sum_{i=1}^{n-1} n - 2 = (n-1)(n-2).$$

With this, we can define degree centralization as

$$\mathcal{C}_D = \frac{\sum_{i=1}^{n} [C_D(v^*) - C_D(v_i)]}{(n-1)(n-2)}. \qquad (4.4)$$

4.1.2 Closeness Centrality

A second centrality measure is based on distances, with the idea that a vertex with the smallest average distance to all other vertices is the most central and thus the most powerful (e.g., can spread information most quickly). If we define the distance between two vertices as $d(v_i, v_j)$, then the closeness centrality [23] of a vertex is the inverse of the averages of the distances:

$$C_C(v_i) = \frac{n-1}{\sum_{j=1}^{n} d(v_i, v_j)}. \tag{4.5}$$

The reason we take the inverse is that in an unconnected graph, the distance between nodes without a path between them is infinite. By taking the inverse, one simply defines the inverse of the infinite distance as zero.

For the centralization measure, a star graph provides the maximum theoretical value of closeness centrality. This is because the value of $C_C(v^*)$, where v^* is the central node in the graph, is equal to 1, as the distance from the central vertex to all others is 1. In the case of other vertices, each vertex has a distance of 2 to $n-2$ vertices, and a distance of 1 to the central vertex. Thus,

$$\max \sum_{i=1}^{n} \left(C_C(v^*) - C_C(v_i)\right) = \sum_{i=1}^{n-1} \left(1 - \frac{n-1}{2(n-2)+1}\right) = \frac{(n-2)(n-1)}{2n-3}.$$

With this, we then define closeness centralization as

$$\mathcal{C}_C = \frac{(2n-3)\sum_{i=1}^{n} \left[C_C(v^*) - C_C(v_i)\right]}{(n-1)(n-2)}. \tag{4.6}$$

4.1.3 Betweenness Centrality

Betweenness centrality is defined as the proportion of shortest paths (called "geodesics") that a specific node is in, within the entire network. Define $q_{ij}(v_k)$ as the number of geodesics that the node v_k is on for the nodes v_i and v_j, with q_{ij} being the total number of such paths. Then set $b_{ij}(v_k) = q_{ij}(v_k)/q_{ij}$. One can then define betweenness centrality [22] as

$$C_B(v_k) = \sum_{j=1}^{n} \sum_{i=1}^{j-1} b_{ij}(v_k), \tag{4.7}$$

where $k \neq i, j$.

The maximal value for v_k in such a network occurs with a star graph [22]. This is shown by starting with a completely disconnected graph G with n nodes and $m = 0$ edges. Adding one edge (i.e., letting $m = 1$) does not affect betweenness. If in the previous example one adds the edge (v_k, v_i) for some node v_i, one can now add another edge (v_k, v_j) for some other node v_j. This now maximizes the betweenness centrality of v_k for the graph G at $m = 2$. One can do this $n-1$ times to maximize the value of the betweenness centrality for v_k, at which point adding any new edges will decrease this value.

In a star graph, there are $(n-1)(n-2)/2$ possible paths between the $n-1$ nodes that are not the central node. Because each v_i, v_j combination has only one geodesic with the only intermediary being the central node, the central node has a betweenness centrality of $(n-1)(n-2)/2$ and all others have 0.

Freeman [23] uses this information to normalize the measure above to obtain

$$C'_B(v_k) = \frac{2C_B(v_k)}{(n-1)(n-2)}. \tag{4.8}$$

A star graph still has the maximal value for the normalized betweenness centrality in (4.8), as one is simply dividing by a constant. This yields the theoretical maximum,

$$\begin{aligned}
\max\left(\sum_{i=1}^{n} C'_B(v^*) - C'_B(v_i)\right) &= \sum_{i=1}^{n} C'_B(v^*) - \sum_{i=1}^{n} C'_B(v_i) \\
&= \frac{2}{(n-1)(n-2)}\left[\sum_{i=1}^{n} C_B(v^*) - \sum_{i=1}^{n} C_B(v_i)\right] \\
&= \frac{2}{(n-1)(n-2)}\left[(n-1)-1\right] \\
&= \frac{2}{n-1}.
\end{aligned}$$

Thus, one can define betweenness centralization as

$$\mathcal{C}_B = \frac{2\left(\sum_{i=1}^{n}\left(C'_B(v^*) - C'_B(v_i)\right)\right)}{2(n-1)} = \frac{\sum_{i=1}^{n}\left(C'_B(v^*) - C'_B(v_i)\right)}{(n-1)}. \tag{4.9}$$

4.1.4 Eigenvector Centrality

The idea behind eigenvector centrality [10, 33] is that a node is important based on how important its neighbours are. For example, a node with a high degree need not be the most important one, if its neighbours have very low degrees. Indeed, its underlying theory is similar to that of Google's PageRank [52]. As such, one can define the centrality measure as being proportional to the average value of one's neighbours. For a node v_i, its centrality, x_i can be defined as

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} x_j \tag{4.10}$$

where $A = \{a_{ij}\}$ is the adjacency matrix of the network. Multiplying by λ, one can easily see that this is simply the calculation of the eigenvector. In other words, we have

$$\lambda \mathbf{x} = A\mathbf{x}. \tag{4.11}$$

The eigenvector centrality of a node v_i is the i^{th} component of the eigenvector associated with the highest eigenvalue of A, and all entries in this eigenvector are non-negative [48]. This non-negativity is discussed and proven in Appendix E.

4.2 Clustering Coefficients

The clustering coefficient is a measure of transitivity of relationships; it seeks to measure how often two neighbours of a node are likely to be connected with each other. This can be defined as the proportion of complete triangles to paths of length two in the network [47]:

$$T = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}. \tag{4.12}$$

A second definition exists as well, and provides for a localized interpretation of the clustering coefficient. In this case, the clustering coefficient [73] for a node v_i is

$$T_i = \frac{2 \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} a_{ik} a_{jk}}{k_i (k_i - 1)}. \tag{4.13}$$

The advantage of (4.13) is that it provides information about individual nodes in the network. This is useful if we want to judge whether certain nodes tend to have

more transitive social ties than others. The network-level clustering coefficient is then taken as the average of the values T_i for $i = 1, \ldots, n$.

There is no commonly accepted weighted definition of clustering coefficient, and numerous versions exist. This provides a useful opportunity for using this data set to help understand the differences between coefficient definitions while also using the coefficients to model cabinet networks. Papers are still being written to compare definitions and individual properties [62]. Some examples of clustering coefficients are given below.

4.2.1 Subgraph Intensities

Onnela et al. [51] describe a definition for the weighted clustering coefficient using a concept of "subgraph intensities". Given a subgraph $\hat{G}$ of G, where $\hat{G}$ is of size $\hat{n}$, one can define the subgraph intensity $I(\hat{G})$ as

$$I(\hat{G}) = \left(\prod_{i=1}^{\hat{n}} \prod_{(j=1,\ j \neq i)}^{\hat{n}} a_{ij} \right)^{1/[\hat{n}(\hat{n}-1)]} . \tag{4.14}$$

Onnela et al. justify such a definition for a subgraph intensity because low values of $I(\hat{G})$ imply that the subgraph is of lower importance that one with a high value. One can then define a local clustering coefficient by considering all of the subgraphs containing a specific node v_i and every permutation of two neighbours. Normalizing the weights of the edges by the maximal edge in the graph will always give a value between 0 and 1. Thus, the weighted clustering coefficient is

$$T_{O,i} = \frac{2}{\kappa_i(\kappa_i - 1)} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\hat{a}_{ij}\hat{a}_{jk}\hat{a}_{ki}\right)^{1/3}, \tag{4.15}$$

$$\hat{a}_{ij} = \frac{a_{ij}}{\arg\max_{i,j} a_{ij}}, \tag{4.16}$$

where κ_i is the number of edges connected to v_i.

4.2.2 Upper Bounds

Zhang and Horvath [78] define an unweighted clustering coefficient as $C_i = w_i/\pi_i$, where

$$\pi_i = \frac{\kappa_i(\kappa_i - 1)}{2}, \qquad w_i = \frac{1}{2}\sum_{u \neq i} \sum_{\{v \mid v \neq i,\ v \neq u\}} a_{iu}a_{uv}a_{vi}. \tag{4.17}$$

For a weighted version where $0 \leq a_{ij} \leq 1$ for all $i,\ j$, Zhang and Horvath argue that π_i can be preserved, and one must then find an upper bound for w_i. Note that if edge weights are larger than 1, then they are divided by the largest weight in the network.

An upper bound for w_i can be calculated using the inequality $a_{ij} \leq 1 - \delta_{ij}$ where δ_{ij} is the Kronecker delta. We then have,

$$w_i = \frac{1}{2}\sum_{u \neq i} \sum_{\{v \mid v \neq i,\ v \neq u\}} a_{iu}a_{uv}a_{vi} \leq \frac{1}{2}\left(\left[\sum_{u \neq i} a_{iu}\right]^2 - \sum_{u \neq i} a_{iu}^2\right). \tag{4.18}$$

Steps showing how to obtain the above are discussed in Appendix D.1.

Thus, one can set π_i as in Equation (4.17) and define the weighted version of the clustering coefficient as $C_i = w_i/\pi_i$. Note that this is equivalent to setting [32],

$$T_{Z,i} = \frac{\sum_{j=1}^{N}\sum_{k=1}^{N}\hat{a}_{ij}\hat{a}_{jk}\hat{a}_{ik}}{\sum_{j \neq k}\hat{a}_{ij}\hat{a}_{ik}}. \tag{4.19}$$

4.3 Homophily

A common question in social network analysis is whether nodes tend to connect to similar or dissimilar nodes, and this feature is called "homophily". Homophily comes in numerous forms [41], including nodes connecting with other nodes of similar strength or degree, similar social status, gender, age, or numerous other characteristics. While such factors play a crucial role in sociological research, a key issue is measuring homophily within a network in a quantitative fashion.

4.3.1 Heterogeneity

A simple approach to defining homophily is by counting categorical node groups within a network. Rather than exploring the actual edges in a network, one simply

delimits the different categorical groups that nodes belong to. One then calculates what the probability is that, if one picks two nodes uniformly at random in a network, the two nodes will come from different groups. This measure, called "heterogeneity" [8] is sometimes defined by

$$Q_H = 1 - \sum_{i=1}^{g} p_i^2 \tag{4.20}$$

where p_i is the proportion of the nodes in the network that belong in the i^{th} group.

A problem with the above is that if the size of the network is small, one must sample without replacement. Thus, an improved version of Equation (4.20) is [64]

$$Q_S = 1 - \sum_{i=1}^{g} \frac{s_i}{n} \frac{s_i - 1}{n - 1}, \tag{4.21}$$

where n is the number of nodes in the network, and s_i is the number of nodes in the i^{th} category.

4.3.2 Modularity and Assortativity

If nodes in a network can be categorized into discrete categories, one can build a matrix representing the number or strength of connections between and within categories. Define a matrix $M = \{m_{ij}\}$ where i represents the i^{th} category, and j the j^{th} category. Then define m_{ij} as the sum of strengths of edges from nodes in category i to category j, divided by the sum of all strengths in the network. The matrix M can be called a "mixing matrix". We can then define the network's "modularity" [49],

$$Q_M = \sum_i m_{ii} - \sum_{ijk} m_{ij} m_{ki}. \tag{4.22}$$

Modularity is zero when there is no homophily or heterophily (nodes from different categories being attracted to each other), negative when heterophily is prevalent, and positive when the network is homophilous.

If comparing multiple networks, one needs to normalize the modularity. For example, one can define the assortativity coefficient [46]

$$Q_A = \frac{\sum_{i=1}^{n} m_{ii} - \sum_{ijk} m_{ij} m_{ki}}{1 - \sum_{ijk} m_{ij} m_{ki}} = \frac{Q_M}{1 - \sum_{ijk} m_{ij} m_{ki}}. \tag{4.23}$$

Chapter 5

Descriptive Overview of the Cabinet Data

With the overview of network measures provided in Chapter 4, it is possible to explore the data and try to understand the various countries and political systems in the data set.

5.1 Individual Overview

At the individual level, we observed few universal patterns in the parliamentary democracies in this study. The correlations between clustering coefficients and maximum edges in governments appear significant and negative for all countries, ranging from -0.9203 for Canada to -0.6568 for Israel. A relationship between clustering coefficient and maximal edge could be a byproduct of mathematical definitions, as discussed in Section D.3.

Another interesting observation is recorded in Table 5.1, showing the correlation between the number of parties involved in the current government and the clustering coefficient of the government itself. While we do not have an intuitive explanation for this, it might be an artifact of the data gathering and interpretation process. Specifically, the correlation between clustering coefficient and number of parties for Luxembourg is -0.06492 rather than 0.5386 if the first government (where clustering coefficient is 1 for all nodes) is ignored.

By observing clustering coefficients over time, as shown in Figure 5.1 for Canada and

Country	Cor(# parties, T_O)
Canada	.
Denmark	0.1366
France	−0.1386
Germany	−0.2304
Israel	−0.1369
Japan	0.3982
Luxembourg	0.5386
Netherlands	−0.1715
New Zealand	.
Norway	0.2656
Sweden	0.5978
UK	.

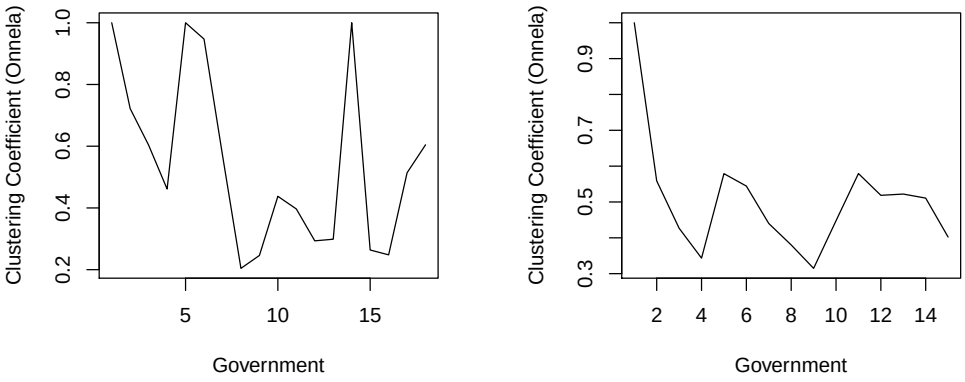
Table 5.1: Correlations between the number of parties and T_O. Dots represent an undefined correlation because the number of parties stayed constant at 1.

Luxembourg, ones sees that the initial measurements for clustering coefficient could be misleading. In the case of Canada, in Figure 5.1(a), the clustering coefficient shoots to 1 every few years. This occurs whenever an entirely new government (one without any previous connections between members) is instituted. This is less prominent in the case of Luxembourg, in Figure 5.1(b), but one can still see that clustering coefficient increases noticeably in a similar fashion. Many network measures, including most centralizations, appear to be volatile like this — whenever a country experiences a large change in government, the network measures change drastically as well.

At the individual level, what is perhaps more telling about this research is how countries differ due to their political histories. Figure 5.2 shows heterogeneity (Q_H) for Japan. Notice how the measure changes for every government, and eventually drops off to zero. This signals the transition of power to the Liberal Democratic Party, which was formed in 1955 and maintained a continuous majority in government past 1990 [37].

As such, while individual network measures provide interesting insights on a case-by-case basis, it appears that little can be gleaned from a descriptive point of view, especially without incorporating political contexts and histories.

When patterns do appear to emerge, they are often remnants of the mathematical relationships between various network measures and simple properties of the countries themselves. For example, there is a strong correlation between degree centralization and closeness centralization in the context of the entire historical cabinet network



(a) Canada.

(b) Luxembourg.

Figure 5.1: Clustering coefficients over time.

(and evolution thereof) for each country, except if the countries have networks with more than one connected subgraph (i.e. "component") [9]. If a country has a fully connected ministerial network (i.e. in the case of countries with proportional representation and coalition governments), as degree centrality decreases, there is less inequality between nodes, in terms of degree. Thus, nodes likely have smaller distances between each other, which brings down closeness centrality measures as well. Such a phenomenon is not observed when multiple components are present because closeness centrality is always extremely low in this case. Low closeness centralities occur in graphs with multiple components because there are infinite distances between nodes in distinct components.

Nevertheless, this provides a useful caveat for any models we develop. We must ensure that our network measures are not simply proxies for simpler or qualitative properties of governments or nodes but actually depend on the network structure of individual cabinet members.

5.2 Comparative Overview

Table F.1 (in Appendix F) lists the countries in this study and some of the mean measures calculated for every government in the data set. A number of very interesting properties seem to emerge. While not listed in the table, the number of components in cabinet networks aggregated over the entire historical period tends

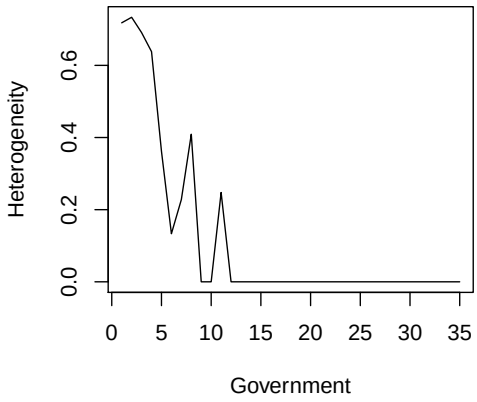


Figure 5.2: Blau's Heterogeneity measure over time in Japan.

to be 1. Indeed, even in countries that never see coalition governments, the entire historical cabinet network tends to be one component due to political mechanisms like ministers switching parties during their careers. One interesting case is that of New Zealand, which has three components (i.e. three sets of ministers have never worked together). Furthermore, a majority (seven of twelve) of countries have mean cabinet sizes between 15 and 25 people, while the rest tend to be smaller.

In some cases, it is the lack of a relationship that proves interesting. A useful property is the maximal edge weight of a cabinet network, which represents the largest number of cabinets two ministers have in common with each other in a given government. The number of governments and mean maximal edge weight of a country do not seem to be highly related. The correlation between the two has a value of 0.4136, with a p value of 0.1814, meaning this relationship is not statistically significant. Specifically, this implies that having a large number of governments within the 45 year period we analyze, such as the case with Israel or Japan, does not necessarily translate to cabinet comemberships lasting in more governments. Furthermore, there is a significant relationship (p value = 0.002427, correlation 0.7862) between the mean cabinet size and the mean number of newcomers to a government, but this relationship disappears when one examines the ratio of newcomers to cabinet size rather than absolute numbers.

Another way to analyze the cabinet networks is by visualizing them, and this is discussed in Appendix C.

Chapter 6

Governmental Models

We use "governmental" models to try to explain and predict network-wide patterns within countries and their governments. While nodes still represent cabinet members, one does not try to follow the progression of a specific node but rather tries to determine whether specific types of nodes tend to survive and progress through multiple cabinets.

How such global parameters cause networks from different countries to vary is useful from a comparative political perspective, as they allow us to see how countries differ. For example, consider a model that accepts a parameter that uses the number of years a node has been in power to calculate the probability the node will survive into a future cabinet position. Upon fitting the model to the various countries in the data set, we might be able to see that some governments seem to treat experienced nodes favourably, whereas other governments do not. Such a conclusion, especially if combined with other data about the individual countries, could be politically significant. For instance, comparing parameter values between countries might allow us to see how different constitutional or institutional properties affect which types of ministers actually gain and keep power.

6.1 Basic Model

A simple approach for modelling cabinets within parliamentary democracies assumes that nodes and their structural properties have no effect on whether a cabinet member will be included in a future cabinet. This model thus assumes that every member of

a cabinet at time t has an equal probability of being in a cabinet at $t+1$.

This model, termed the "basic model", assumes that countries have a fixed cabinet portfolio of n members[1], and each member has a probability p of being re-elected and included in the next cabinet. If a member drops out of the cabinet, he or she will never return. Because the model is based on probabilities, the survival rates of cabinet members and percentages of incumbents in specific governments become random variables as well.

We assume that the size of government and probability of being selected into the following cabinet are both fixed. Because the probability is the same for every member of cabinet, the number of ministers who move on to the next cabinet follows a binomial distribution, $\texttt{bin}(n, p)$, where n is the size of government, and p is the probability of moving on. Specifically, the probability that k cabinet members will be present in governments at times t and $t+1$ is [60]

$$\Pr(K = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}. \tag{6.1}$$

Thus, using the mean and variance of a binomial distribution, we know that the number of incumbents in a cabinet is np with a variance of $np(1-p)$.

Because incumbent cabinet members are selected from the prior cabinet only, a cabinet member will stay in office following a geometric distribution, $\texttt{geo}(p)$ [60]:

$$\Pr(K = k) = p^{k-1}(1-p) \tag{6.2}$$

This implies that the expected number of governments that a cabinet member will serve in is $1/(1-p)$, with a variance of $p(1-p)^{-2}$.

6.1.1 Membership Predictions

Due to the simplicity of the basic model, it is possible to analyze the expected rates of incumbents and survival rates of politicians in detail. Specifically, we are interested not only in the number of incumbents in a cabinet, but also the number of cabinets each member has already served. Thus, given a government g_t, let I_1 represent

[1] We use n to represent the size of the government because the government exists in a network context.

the number of incumbents from the previous government (i.e. from g_{t-1}). Thus, $I_1 \sim \texttt{bin}(n, p)$, so that I_1 is a random variable taken from a binomial distribution.

To see how many incumbents have been in office since $g_{\hat{t}}$, where $\hat{t} < t$, we use an inductive approach. For every government between $\hat{t}$ and t, the incumbent had to be reselected in our model. Because each selection takes place with a probability of p and we make $t - \hat{t}$ selections, there is a probability $p^{t-\hat{t}}$ that the specific node in $g_{\hat{t}}$ will still remain in g_t. As such, $I_{\hat{t}} \sim \texttt{bin}(n, p^{t-\hat{t}})$, yields an expectation of $np^{t-\hat{t}}$.

Another way of looking at the above is that at g_{t-1} the expected number of nodes from $g_{\hat{t}}$ is equal to $np^{t-1-\hat{t}}$. Because each of these nodes continues to serve in g_t with a probability of p, it follows that $I_{\hat{t}} \sim \texttt{bin}(np^{t-1-\hat{t}}, p)$.

Let $J_{t,u}$ be the set of ministers who entered the model at time u and are still present at time t. Also, let $\mathcal{G}_t$ represent the set of ministers that served in government g_t. Thus, we have a government g_t composed of the set

$$\begin{aligned} \mathcal{G}_t &= \cup_{i=1}^{t} J_{t,i} \\ &= J_{t,t} + \cup_{i=1}^{t-1} J_{t,i} \\ &= \{\text{newcomers}\} \cup \left(\cup_{i=1}^{t-1} J_{t,i}\right). \end{aligned}$$

It is possible to get the expected numbers of cabinet ministers in g_t that began serving as ministers from times prior to t. Specifically, we have $\mathcal{G}_2 = J_2 \cup J_1 = \{\text{newcomers}\} \cup \{\text{incumbents from } g_1\}$. Because the distribution for each step is binomial, and newcomers and incumbents are disjoint sets, we obtain: $n = \mathbb{E}(|\mathcal{G}_2|) = \mathbb{E}(|J_2|) + \mathbb{E}(|J_1|) = (n - np) + np$.

Similarly, we have that $n = \mathbb{E}(|\mathcal{G}_3|) = \sum_{i=1}^{3} \mathbb{E}(|J_i|)$. Because every binomial step (i.e. government change) is considered to be independent of the last one, we thus have $n = \mathbb{E}(|J_3|) + \mathbb{E}(|J_2 \cup J_1|) = (n - np) + np = (n - np) + (n - np + np)p$. This leads to

$$n = \underbrace{(n - np)}_{\text{newcomers}} + \underbrace{(np - np^2)}_{\text{incumbents from } \mathcal{G}_2} + \underbrace{np^2}_{\text{incumbents from } \mathcal{G}_1}. \tag{6.3}$$

Using induction, we then get that for $\mathcal{G}_t$, the expected number of incumbents from

different governments is:

$$\begin{aligned}
\mathbb{E}(|J_{t,t}|) &= np^0 - np^1 = n - np \qquad \text{(newcomers)}, \\
\mathbb{E}(|J_{t,t-1}|) &= np^1 - np^2, \\
&\vdots \\
\mathbb{E}(|J_{t,k}|) &= np^{t-k} - np^{t-k+1}, \\
&\vdots \\
\mathbb{E}(|J_{t,2}|) &= np^{t-2} - np^{t-1}, \\
\mathbb{E}(|J_{t,1}|) &= np^{t-1}.
\end{aligned} \tag{6.4}$$

In other words, for g_t, one has an expected number incumbents that entered the cabinet at time k equal to $np^{t-k}(1-p)$.

Using this information, one can create an "experience matrix" that shows the number of newcomers and incumbents from specific governments:

$$\Omega = \{\omega_{ij}\} = \begin{bmatrix} n & 0 & 0 \\ n - np & np & 0 \\ n - np & np - np^2 & np^2 \end{bmatrix} \tag{6.5}$$

Equation (6.5) shows the expected number of cabinet members for a country with three successive governments of size n with probability p of a cabinet member being reselected for a post. Matrix rows represent successive governments, and columns represent the government in which the cabinet member started his or her ministerial career. Thus, ω_{ij} represents the nodes in government i that joined $j-1$ governments ago.

One can calculate the variance of the values in the experience matrix similarly. Specifically, observe the third row of the experience matrix. In this case, one has an expectation of $n - np$ for the number of newcomers, $np - np^2$ for incumbents from the second government, and np^2 for incumbents from the first government. Observe, however, that in each case we look at expected numbers of incumbents, we are actually observing various binomial distributions:

$$\begin{aligned}
(np - np^2) &= \mathbb{E}[(\texttt{bin}(n, p - p^2)], \\
(np^2) &= \mathbb{E}[\texttt{bin}(n, p^2)].
\end{aligned}$$

Thus, using the expectation in Equation (6.4), one can see that for the k^{th} row and m^{th} column of the experience matrix, where $1 < m < k$, each entry is represented by the random distribution $\texttt{bin}(n, p^{m-1} - p^m)$. This implies the variance is $n(p^{m-1} - p^m)(1 - p^{m-1} + p^m)$.

6.2 Single-Measure Models

As discussed in Chapter 2, a key question within cabinet and coalition theory is understanding whether the network structure of the cabinet plays any significant role in ensuring that cabinet members remain in power. Furthermore, the position a node holds within a social network can significantly impact its ability to gain access to resources, information, and other forms of social capital. A politician's position in a cabinet network could similarly impact his or her access to forms of capital, and it is worth testing whether the structural properties of a network do indeed impact nodes in such a way.

With this in mind, we build a probabilistic model that works in a similar fashion as the basic model, but with probabilities that depend on politicians' specific structural properties in the cabinet networks. As in the case of the basic model, we assume that the cabinet members are only selected from the preceding government, and that their probability of doing so is proportional to some node-based measure, $f(c_i)$, where f is a function that outputs the relevant measure for the node in question. Thus,

$$\begin{aligned} P(c_i) &\propto f(c_i) \\ P(c_i) &= \alpha f(c_i) \end{aligned}$$

where $P(c_i)$ is the probability that node c_i will have a portfolio in the next cabinet and α is a constant.

We can thereby fit the model to any node-level network metric. Strength, centralities, clustering coefficients, assortativity, and other measures are all relevant and useful candidates for such a model. Furthermore, if such a model does better in reducing error (discussed in Chapter 8) compared to the basic model in Section 6.1, we can conclude that network measures play a role in ensuring the survival (or death) of politicians' careers as cabinet members.

A further advantage of using single-parameter models is that one can use simple and fast optimization techniques to analyze them. One such technique is the Golden Section Search described in Section 7.1.

6.2.1 Strength Model

As an example, consider a model based on the strength of individual nodes. In this case, the model assumes that the probability a cabinet member will be a member again is proportional to his or her strength. From a political science perspective, strength represents how long a certain party or group of politicians have been in office, and whether or not politicians are more likely to survive in cabinets if they work with other survivors. This yields

$$P(c_i) \propto \sum_{j=1,\ i \neq j}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} \tag{6.6}$$

where we recall that $a_{ii} = 0$ for all i.

Suppose we denote the expected number of incumbents from the previous year by β. Then for some constant b we have,

$$\begin{aligned} \beta \equiv \mathbb{E}[|\{\text{incumbents}\}|] &= \sum_{i=1}^{n} P(c_i) \\ &= \sum_{i=1}^{n} \left(b \sum_{j=1}^{n} a_{ij} \right) \\ &= b \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}. \end{aligned} \tag{6.7}$$

At the start of a simulation, we have a matrix $A = \{a_{ij}\}$ where $a_{ij} = 1$ for $i \neq j$ and $a_{ii} = 0$. Thus, the sum in Equation (6.7) is $n(n-1)$. This means that for the first iteration of the model, if we want to expect β incumbents, we must have

$$b = \frac{\beta}{n(n-1)}. \tag{6.8}$$

6.2.2 The Strength Model as a Bipartite Network Model

One reason that we consider the example of the strength model is that it is equivalent to a bipartite network model. Based on our current approach to modelling

governments, our networks have very specific properties:

1. Every cabinet is represented by a clique of a fixed size n.
2. Edge weights represent the number of times two politicians have served in the same cabinet.

Using these points, there is a direct correspondence between a minister's strength and the number of times he or she served in a cabinet position. Specifically, the strength k_i of node c_i, is a function of the number of governments ϕ_i in which a node has served:

$$k_i = \phi_i(n-1). \tag{6.9}$$

The equivalent bipartite model has two types of nodes: governments, g_t for some discrete time value t, and the cabinet members themselves, c_i. We can set up a system where members of g_t have a probability of remaining in g_{t+1} proportional to their experience:

$$P(c_i \text{ in } g_t \text{ and in } g_{t+1}) \propto \deg(c_i), \tag{6.10}$$

where $\deg(c_i)$ represents the degree of c_i in the bipartite network. Note that in this case, degree is equal to the number of governments served by c_i, because nodes have edges connecting them to the governments they served. Thus, we have

$$k_i = (n-1)\phi_i = (n-1)\deg(c_i). \tag{6.11}$$

Inserting this into the strength model discussed in Section 6.2.1 gives the following probability:

$$P(c_i) = bk_i = b(n-1)\ \deg(c_i) = \hat{b}\ \deg(c_i),$$

where $\hat{b}$ is a new constant. Thus, the strength model with parameter b is equivalent to a bipartite network model with a parameter $\hat{b} = b(n-1)$.

6.2.3 Other Properties

The start of every model that we consider involves a network of n nodes connected to all other nodes in the network with edge weights of 1, so one can make basic

Measure Type	ψ value
Strength	$n-1$
Clustering Coefficient (*any*)	1
Eigenvector Centrality	any

Table 6.1: Various network measures and their ψ values. Note that the eigenvector centrality can have any ψ value, as the eigenvector for a clique is any vector with equal entries. This is discussed in Appendix G.

predictions on the expected number of incumbents in the second government. Because every node in the network is structurally equivalent – that is, it has an equivalent network position when one considers neighbours and descriptive network measures – we have

$$\begin{aligned} \mathbb{E}[|\{\text{incumbents}\}|] &= \sum_{i=1}^{n} P(c_i) \\ &= \sum_{i=1}^{n} bf(c_i) \\ &= nb\psi, \end{aligned}$$

where ψ is the measure's value for a clique with edge weights of 1. Table 6.1 shows ψ values for typical network measures.

6.2.4 Perturbing the Optimal Probability

Another approach we use to build models based solely on one network measure is having the probability that nodes remain in a cabinet position be based on a probability that is then slightly modified based on a network measure. In this case, we have

$$P(c_i) = P_0 + \alpha f(c_i) \tag{6.12}$$

for some constant α and an initial probability of reselection P_0. The advantage of this model is that one can obtain the optimal probability for the basic model and use this as the value for P_0, with $\alpha = 0$. One can then use simulated annealing or another optimization algorithm to see if slightly changing P_0 and α will yield lower error rates.

6.3 Preferential Attachment

The preferential attachment model is based on the idea that those who already have a large degree or strength are more likely to receive connections from new nodes entering a network. In social science, this is sometimes called the "Matthew Effect" [42], based on the Bible verse that discusses how rich people tend to get richer. The idea of the Matthew Effect applies to network modelling, as it has been observed that different types of social networks exhibit properties where nodes with large strengths or degrees tend to gain more edges disproportionately faster than other nodes [57, 63].

The "preferential attachment" mechanism [5] is an application of the Matthew Effect idea to network analysis. It strives to explain scale-free distributions of degrees within networks. Specifically, a network is considered "scale-free" when, given a constant γ, the distribution of nodes with degree k is the power law,

$$P(k) \sim k^{-\gamma}. \tag{6.13}$$

The reason the above type of network is called "scale-free" deals solely with power law form of the degree distribution [47]. Specifically, a scale free function satisfies $P(ak) = bP(k)$ for some a and b, and all k.

The preferential attachment model assumes that a network increases in size by the addition of new nodes. The degree of new nodes is bounded by a predetermined value, and the probability that the newly-added edge connects to a specific pre-existing node depends on the latter node's degree. The probability that a node v_i will be connected to the new node is

$$P(k_i) = \frac{k_i}{\sum_{j=1}^{N} k_j}, \tag{6.14}$$

where k_i is the degree of node v_i and $P(k_i)$ is the probability that the new node will have an edge connected to v_i.

The models described earlier all focus on selecting cabinet members that have served only the previous government, while the present preferential attachment model selects from a pool of all ministers in the entire history of the country. A second variable, representing the aging bias is included as well. Ministers cannot serve forever, and depending on how old a minister is, he or she may be forced into retirement. The

preferential attachment model is thus

$$P_c(k_i, z_i) = \frac{\alpha_1 k_i + \alpha_2 z_i}{\sum_j (\alpha_1 k_j + \alpha_2 z_j)}, \tag{6.15}$$

where z_i represents the number of governments (i.e. "experience") that the cabinet minister has served in the past. Weights associated with the strength and experience variables are represented by α_1 and α_2, respectively.

We built another model where the probability is based on $\alpha_1 k_i + \alpha_2 z_i + \alpha_3 z_i^2$. The reason for this is that experience in multiple governments may be beneficial in the short-term, but due to the limited biological capacity of politicians, they eventually have to remove themselves form political careers. Thus, one would expect to have $\alpha_2 > 0$ and $\alpha_3 < 0$.

A second difference between Barabási and Albert's [5] preferential attachment mechanism and ours is that in our case, a clique of new cabinet ministers is added at every time step, rather than just an individual node. This clique is of a predetermined size $\hat{n}$, and $n - \hat{n}$ incumbents are then selected to join the newcomers. Each cabinet member is chosen with a probability equal to P_c, and incumbents are sampled without replacement.

While not implemented in this dissertation, a second approach to dealing with age and other factors that might affect the attractiveness of a node to an incoming node is to multiply node v_i's strength by a weight [2],

$$P(k_i) = \frac{\eta_i k_i}{\sum_j \eta_j k_j}, \tag{6.16}$$

where η_i is some fitness parameter. In the case of the cabinet networks and our function P_c, this function could be dependent on the age of the politician.

Chapter 7

Model Optimization

We implemented a number of optimization algorithms for the modelling task outlined in this report.

7.1 Golden Section Search

The Golden Section Search algorithm [56] is based on the idea that if we have three points, $x_1 < x_2 < x_3$ on an interval with $f(x_2) < f(x_1)$ and $f(x_2) < f(x_3)$ then, if the function is unimodal and continuous, there is a $x^* \in (x_1, x_3)$ such that $f(x^*)$ is a local minimum.

As a consequence of the previous text, the algorithm begins with points a and c on an interval[1] $[a, c]$, and looks for b and x as illustrated in Figure 7.1. We then define w to satisfy

$$\frac{b-a}{c-a} = w, \qquad \frac{c-b}{c-a} = 1 - w.$$

Thus, w is the ratio of the distance between b and a versus the length of the entire interval. If we assume that the next point in the interval will be x, we can then define the distance z as the distance between x and b over the entire interval length:

$$\frac{x-b}{c-a} = z$$

[1] While in the initial example the interval for the minimum point was open, we treat it as closed in the actual algorithm because aside from the first iteration, we do not know if our minimum point will appear on a boundary.

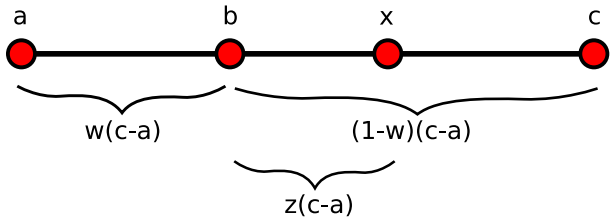


Figure 7.1: Illustration of the Golden Section Search algorithm, where $w = (b - a)/(c - a)$ *and* $z = (x - b)/(c - a)$.

We then choose either x to form a new interval $[a, x]$ or b to form the interval $[b, c]$. To minimize the worst-case outcome, we require that $1 - w = w + z$, which allows us to obtain $z = 1 - 2w$.

Furthermore, by this approach, we thus require $|b - a| = |x - c|$.

We now have an optimal way of choosing z based on w. Because this is an iterative approach, we can assume that w was picked in a similar way. Thus, we require that the ratio between z and $1 - w$ be equal to w. This yields

$$\begin{aligned} &\frac{z}{1-w} = w \\ &\Rightarrow w^2 - 3w + 1 = 0 \\ &\Rightarrow w = \frac{3 \pm \sqrt{5}}{2}. \end{aligned}$$

Because we require $0 < w < 1$, we take the negative root above, so $w \approx 0.38197$.

With this in mind, we can then choose points x and b based on the value of w, above. If we see that $f(x) < f(b)$, then our new interval in the next iteration will be $[b, c]$. If $f(x) > f(b)$, the interval will be $[a, x]$.

7.2 Nelder-Mead Algorithm

The Nelder-Mead algorithm is a nonlinear optimization technique that uses an $(n+1)$-dimensional simplex to find maxima or minima of a given n-dimensional function [45]. This is a popular algorithm for non-linear optimization and is implemented in MATLAB in the `fminsearch` function. Because it is difficult to conclude whether the error functions associated with the parameter space of the models described in Chapter 6 are linear or follow any strict patterns, implementing a nonlinear search technique

can help determine optimal parameter values. Note, however, that we implement our own version of the algorithm.

We begin an implementation of the Nelder-Mead algorithm by being given $n+1$ points $\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}$ in a parameter space $\mathbb{R}^n$, ordered in such a way that $f(\mathbf{x}_1) \leq \ldots.. \leq f(\mathbf{x}_{n+1})$. In our case, the goal of every iteration of the algorithm is to replace $\mathbf{x}_{n+1}$ with a vector that decreases the error function (i.e. our objective function). The actual implementation of the algorithm follows MATLAB's own [35], and the steps of an iteration are:

1. Order the points as described above.
2. Define $\bar{x} = \sum_{i=1}^{n} \mathbf{x}_i/n$. The first candidate is $\mathbf{x}_r = (1+\rho)\bar{\mathbf{x}} - \rho\mathbf{x}_{n+1}$. If $f(\mathbf{x}_1) \leq f(\mathbf{x}_r) < f(\mathbf{x}_n)$, then replace $\mathbf{x}_{n+1}$ and quit the iteration.
3. If $f(\mathbf{x}_r) < f(\mathbf{x}_1)$, set $\mathbf{x}_e = (1+\rho\chi)\bar{\mathbf{x}} - \rho\chi\mathbf{x}_{n+1}$ and replace $\mathbf{x}_{n+1}$ with $\mathbf{x}_e$ if $f(\mathbf{x}_e) < f(\mathbf{x}_r)$. Otherwise, replace it with $\mathbf{x}_r$.
4. If $f(\mathbf{x}_r) \geq f(\mathbf{x}_n)$, then perform one of two steps:
   (a) If $f(\mathbf{x}_n) < f(\mathbf{x}_r) < f(\mathbf{x}_{n+1})$, then set $\mathbf{x}_c = (1+\rho\gamma)\bar{\mathbf{x}} - \rho\gamma\mathbf{x}_{n+1}$, and if $f(\mathbf{x}_c) < f(\mathbf{x}_r)$ then replace. Otherwise, exit this conditional step.
   (b) If $f(\mathbf{x}_r) > f(\mathbf{x}_{n+1})$, set $\mathbf{x}_{cc} = (1-\gamma)\bar{\mathbf{x}} + \gamma\mathbf{x}_{n+1}$. If $f(\mathbf{x}_{cc}) < f(\mathbf{x}_{n+1})$ then replace with $\mathbf{x}_{cc}$.
5. If all of the above fails, reset the entire set of points by setting $\mathbf{y}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$ for $i = 2, \ldots, n+1$.

Following the MATLAB implementation [35], we set $\rho = 1$, $\chi = 2$, $\gamma = 1/2$, and $\sigma = 1/2$.

In our implementation of the algorithm, the optimizer accepts one initial guess for the optimal points, $\mathbf{x}^* = [a_1, \ldots, a_n]$ and sets it as $\mathbf{x}_{n+1}$. Then it sets $\mathbf{x}_i$ as all zeroes except for the i^{th} position, where we have a_i.

7.3 Simulated Annealing

Simulated annealing is an optimization method based on energy states in statistical mechanics [17]. Specifically, one is looking for low-energy states for a set of particles.

Given a starting point of variables, one then attempts to minimize the value of an error function in the system.

If one simply performs a localized search for a minimum, it is likely that with complex objective functions, the searcher will get stuck in a local minimum. It is thus important that we consider both nearby and distant parameter values, and that sometimes values that worsen (i.e. increase) the objective function are used as steps in the algorithm. Of course, this still does not guarantee that a global optimum will be found.

Below is a basic overview of how simulated annealing works [4]. First, we choose an initial starting point or guess, $\mathbf{x}_0$, which is evaluated using an objective function, f. Given a new set of parameters, $\mathbf{x}'$, one can move from $\mathbf{x}$ to $\mathbf{x}'$ on two occasions:

1. When $f(\mathbf{x}') < f(\mathbf{x})$.
2. If $f(\mathbf{x}') \geq f(\mathbf{x})$, then one moves to $\mathbf{x}'$ uniformly at random, using a probability p (see below). Alternatively, we pick a new point $\mathbf{x}''$ based on the current $\mathbf{x}$.

Candidates for the next iteration of the algorithm can be selected in various ways. In this dissertation, one has a set $\mathcal{X}$ of candidates such that, given the current state of $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathcal{X} = \{\mathbf{y} \in \mathbb{R}^n \mid y_j = x_j \pm \epsilon \text{ for } j \text{ in } 1, 2, \ldots, n \text{ and } y_i = x_i \text{ elsewhere.}\}. \tag{7.1}$$

Given the set of candidates $\mathcal{X}$, a member $\mathbf{x}' \in \mathcal{X}$ is chosen uniformly at random and the function f is evaluated. If $f(\mathbf{x}') < f(\mathbf{x})$, then $\mathbf{x}'$ is set as the new value, and a new set $\mathcal{X}$ is generated at the next iteration. Otherwise, $\mathbf{x}'$ still replaces $\mathbf{x}$ with a probability p, defined as:

$$p = e^{\frac{f(\mathbf{x}') - f(\mathbf{x})}{\tau}}, \tag{7.2}$$

where $\tau = r^{i-1}T$, and where $0 < r < 1$ is a cooling rate, T is the initial temperature, and we are on the i^{th} iteration. We then define the perturbation at every step as $\epsilon = \tau\epsilon_0$, where ϵ_0 is the initial value for the perturbation and is provided by the user.

The algorithm runs a predefined set of iterations, or exits once it reaches a certain tolerance for the error (i.e. objective function).

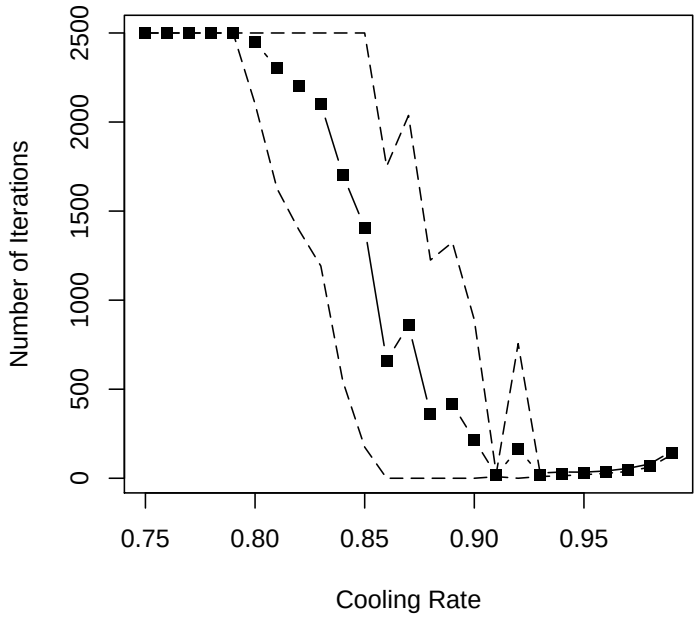


Figure 7.2: Number of iterations necessary to either quit the simulated annealing algorithm (which we have chosen to be 2500 *iterations) or reach convergence with error tolerance* 0.1*. The dashed lines represent one standard deviation away from the mean number of iterations (or* 0 *or* 2500 *where relevant) when running the algorithm* 50 *times for* $T = 1$*, and an initial perturbation of* 2*. The function being minimized is* $f(x, y, z) = x^2 + y^2 + z^2$*.*

The biggest problem with regard to using simulated annealing is the fact that, depending on how one sets the cooling rate, temperature, and perturbations, the algorithm may converge to a local (or, ideally, a global) optimum extremely slowly. Figure 7.2 shows how for a function $f(x, y, z) = x^2 + y^2 + z^2$, the number of iterations necessary to find the minimal point ($x = y = z = 0$) can vary significantly. Based on the figure, it is only within a relatively small subset of values in $(0, 1)$, namely between about 0.8 and 1.0 that the algorithm actually converges without reaching the maximum number of iterations allowed.

Chapter 8

Optimized Models

The models described in Chapter 6 incorporate important aspects of dynamics within cabinet networks. However, to actually judge whether or not such dynamics are relevant in any of the parliamentary democracies being explored, it is important to optimize the parameters of models and see how well they do compared to the baseline model.

8.1 Error Definition

The errors being analyzed are based on the observed experience matrices of the parliamentary democracies. Specifically, if $\hat{O} = \{\hat{o}_{ij}\}$ represents the simulation means and $O = \{o_{ij}\}$ is the observed set of values, then one way to define error is using the 2-norm of the difference. That is

$$e = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (\hat{o}_{ij} - o_{ij})^2 \right)^{1/2} = \left(\sum_{i=1}^{n} \sum_{j=1}^{i} (\hat{o}_{ij} - o_{ij})^2 \right)^{1/2}. \tag{8.1}$$

Note that the second equality exists because the experience matrix is lower-triangular.

8.2 Model Fits

We calculated results for the basic model, single-measure models, and preferential attachment model, and a complete set of results is included in Appendix H. The

basic model results are reported in Table H.1, with most of the results showing the lowest error rates at around $P = 0.5$. Because calculating the variance of a matrix with random entries is non-trivial, we also present the results of the basic model's simulations, rather than analytical values.

Results for single measure models are reported in Table H.2, with models for perturbed probabilities in Table H.3. Finally, results for the preferential attachment model are presented in Table H.4.

Most of the models do not have lower error rates than the basic model (i.e. the baseline). What is promising, however, is that five countries outperform the basic model in one network-centric model each. New Zealand and Luxembourg both outperform the baseline results through the perturbed probability model, where the initial probability is shifted using nodes' strengths. Japan outperforms the baseline using the strength model, and Canada does so using the single-measure model based on T_O (discussed in Section 4.2.1). Finally, France improves upon the basic model through the preferential attachment model. While the differences are not extremely large, they are statistically significant. That is, all of these examples have lower mean error values at a statistically significant difference compared to the simulations of the basic model.

These results imply a number of things. In the Japanese case, if one treats the strength model as a bipartite one (as discussed in Section 6.2.2), this implies that the more governments a cabinet member serves, the more likely that he or she will be reselected for a future portfolio. One can argue a similar result for Luxembourg and New Zealand, though since here we are perturbing initial probabilities, it implies the effect of serving prior governments is smaller.

France also has a similar interpretation. With the preferential attachment model, it appears that ministers who serve many governments and work with others who do the same are more likely to be selected for future posts.

The outlier in this case is Canada, which has clustering coefficients play an important role. This is particularly interesting, because it implies that cabinet members succeed based not only on who they work with, but if those ministers also work with each other.

While models with error rates below the baseline were developed for only five of the twelve democracies, this shows that models based solely on our data set may yield

useful and promising results, and that significant network effects are present in at least some of the parliamentary democracies analyzed. Furthermore, because four different model types had significant results, this implies that the underlying mechanics of ministerial survival may be different depending on the country. It is possible that other types of measures may be significant for the seven other parliamentary democracies being analyzed, or that more complex network mechanics would be necessary to improve beyond the baseline models.

Chapter 9

Ministerial Models

Thus far, we have explored global models of networks. That is, the earlier models were optimized for global variables and properties of the network, rather than seeing how well we could do in predicting the evolution and development of individual nodes. In this chapter, we set out to explore whether we can make predictions about individual nodes. Specifically, we hope to predict whether or not members of a specific cabinet will be selected for another cabinet position in the future.

The problem definition in this case is as follows: given as much information as possible about a specific node, such as political party, assortativity coefficient of the government, number of governments the node has been involved in, eigenvector centrality of the node, and so on, can we predict if the node will be in power in the future? As discussed in Section 2.3, an important question is how cabinet ministers evolve over time and whether they stay in power. A key question to ask, then, is given network-focused information, can we predict with any sort of accuracy whether or not a cabinet member will serve in a future government?

9.1 Supervised Learning

The approach we use for this problem can be termed as a "supervised" predictive modelling approach. The term "supervised" is based on the fact that we train the model on results that we have already observed — in this case, it is knowing whether or not a cabinet member returns to power in a later cabinet. Every algorithm we use seeks to map a data vector with n variables, $\mathbf{x} \in \mathbb{R}^n$, to a binary result, $\{0, 1\}$. Thus,

we seek to fit a function $\ell : \mathbf{x}^n \to \{0, 1\}$ to minimize another objective function.

The objective function, as before, is based on a measure of error. Because our function returns one of two values, 0 for ministers that never return to cabinet, and 1 for ministers that do, we can define the error function as the accuracy rate of our model. Specifically,

$$\text{accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \tag{9.1}$$

Thus, accuracy takes a value in the interval $[0, 1]$, and we seek to maximize the accuracy.

9.2 Preparing the Data

Because we aim to develop a model that predicts whether or not specific cabinet members return to a cabinet position in the future, we can input various variables into the model. Since our parameter for ℓ is a vector $\mathbf{x}$, we can input any information into $\mathbf{x}$ that may help our function discriminate between ministers that serve in another cabinet and those who do not. For example, x_1 may contain the T_O clustering coefficient, while x_4 may contain the node's strength. We hope that our algorithm will be able to take these element values and learn to discriminate between the two classes, $\{0, 1\}$, based on the entries in $\mathbf{x}$. In our case, $\mathbf{x}$ is composed of the following entires:

- **Global Network Properties:** average clustering coefficient (T_O, T_H, T_Z)
- **Node-Level Network Properties:** strength, clustering coefficient ($T_{O,i}$, $T_{H,i}$, $T_{Z,i}$), eigenvector centrality, betweenness centrality
- **Party-Based Properties:** assortativity coefficient, heterogeneity
- **Historical Properties:** whether the node is a newcomer or incumbent, number of governments the node has served in

Within the realm of machine learning, one must decide what data one should train a model on, and what data one should test on. A key problem is ensuring that we do not overtrain a model [40]. This occurs when one inputs biases into the training process by selecting biased training sets. For example, if one were to simply randomize all

the data vectors (variables listed above) from all governments and pick a random set of training instances and random set of test instances, we would likely overtrain the models. This would occur because the global network-level properties stay the same for all nodes in a specific government, and some of these nodes would be randomly placed in the training set, while others in the test set. Thus, we would build and evaluate our model based on data that we would never be able to access in practice. Implicitly, we would be building models with information about future governments, and so biasing our results.

A second issue is understanding how much historical information is necessary for accurate predictions. For example, Canada had 18 governments between 1945 and 1990. If we limit ourselves to forward-looking models, we can build models for 16 governments. The first one would be dropped due to the absence of training data, and the last one cannot be validated. However, having a training set with only one government would likely yield extremely poor results. Similarly, having a test set that only incorporates the 17^{th} government could yield poor results if the transition to the 18^{th} government was a particularly tumultuous one.

With this in mind, we built $N - 1$ models for each country, where N is the number of governments in the data set for the country in question. Thus, $N - 1$ splits were used to create training and test data sets. If we take the Canada example, with 18 governments, suppose we have a split at government seven. This means we take all the nodes in the first seven governments and their associated network properties, and see if they return to power at any point in the future. We use this to train the model, and then validate it on the 11 governments we did not include in the training set.

Note that we dropped Israel and Germany from this analysis because we lacked complete data on political party membership.

9.3 Voting by Majority Class Membership

The simplest type of model, and the one used as a baseline for this research project, deals with simply applying the class with the highest membership rate to all future instances. For example, given 100 training instances, if 60% are in class A and 40% in class B, then the model will categorize all future instances as members of class A.

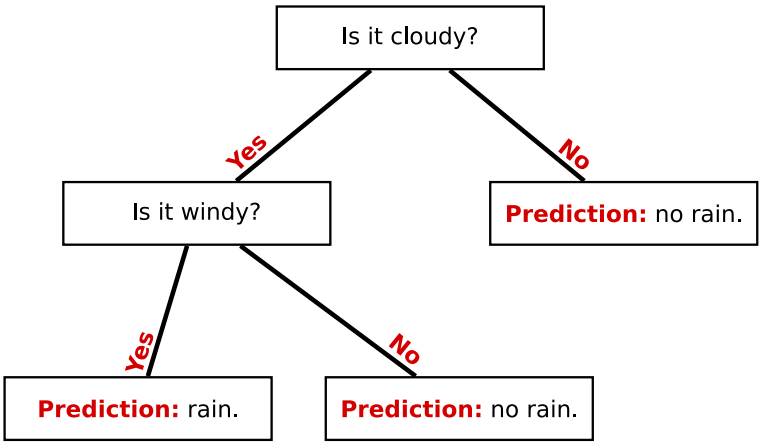


Figure 9.1: An example of a binary tree decision system. One variable (cloudiness) is evaluated first, either leading to a decision or another variable evaluation.

This is a crude approach, as it assumes that class membership rates do not vary, and tries to maximize accuracy rates without any learning.

9.4 Binary Decision Trees

This is a very basic model, but one that worked better than any of the other model types initially attempted, including logistic regressions [36], support vector machines [29], or nearest neighbour comparisons [15]. All of these algorithms were explored in the WEKA software package [76]. Due to the success of binary trees, only they were explored in detail in this project.

A binary tree is a classifier that tries to build decision-making rules for predicting the specific class of an input instance (i.e. vector) [76]. For example, if we were to build a classifier that answers the question, "Will it be raining in an hour?", the tree in Figure 9.1 seems to be a reasonable approach. The first question is "Is it cloudy?", and this splits into two options. If "yes" or "no", then the second question could be "Is it windy?" One hopes that by breaking down the tree into various categories, one can place all the possible events in such a tree and be able to classify any given condition.

A random tree is built by selecting x number of attributes uniformly at random and choosing the attribute that best discriminates between classes. In our case, 4 random attributes were chosen and the best one was used to split the tree. At the next level, 4

new attributes were randomly selected again. This process continues until all training instances are categorized in the tree.

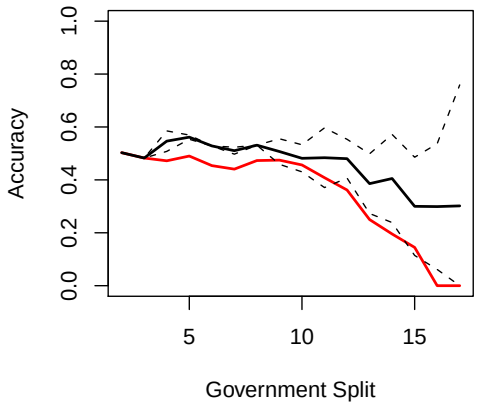
9.5 Model Evaluation

Like the governmental models discussed in Chapter 6, the models in this section vary substantially between countries. Even though the underlying mathematical and computational logic may be the same, the accuracies between countries vary a great deal, implying that the underlying political dynamics may be different. For example, Figures 9.2(a) and 9.2(b) show a typical situation for the data sets, where as we have more training instances, the random tree algorithm tends to do better than the one based on majority classes. However, Figure 9.2(c) shows that Japan provides a more challenging data set for the algorithm, and it rarely has a better accuracy than the baseline.

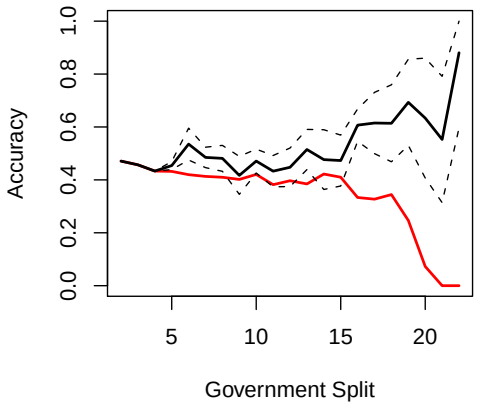
Table 9.1 provides the results of building random trees (20 different times) for each country, compared to the baseline model. Observe that aside from Japan, New Zealand, and Sweden, the models outperform the baseline case.

Country	**Baseline Acc.**	**Mean Acc.**	**Standard Dev., Acc.**
Canada	0.2500	0.3858	0.1128
Denmark	0.4103	0.4891	0.0865
France	0.3750	0.4688	0.0869
Japan	0.8816	0.7270	0.0294
Luxembourg	0.2500	0.4962	0.1365
Netherlands	0.5000	0.5153	0.0619
New Zealand	0.4333	0.4467	0.1347
Norway	0.3444	0.6139	0.1455
Sweden	0.5000	0.4321	0.1060
UK	0.2574	0.5812	0.0613

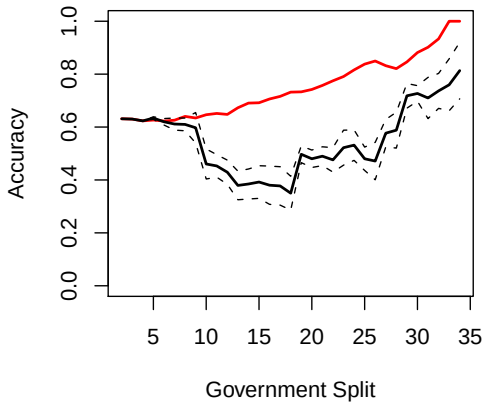
Table 9.1: Results of fitting the binary tree classifier twenty different times. "Acc." represents accuracy. Note that the baseline model does not have a standard deviation as it always builds the same model.



(a) Canada.



(b) Norway.



(c) Japan.

Figure 9.2: Accuracies for the random tree algorithm (dark line) and majority vote (red line). The dashed lines represent one standard deviation away from the mean of the random tree algorithm. The split represents at which government we begin to built the test (i.e. validation) data set, also known as the "holdout" set.

Chapter 10

Discussion and Conclusion

The goal of this research was to explore whether network-centric variables can be used to help explain some of the dynamics of cabinet members between 1945 to 1990. We applied this to twelve parliamentary democracies, with models focusing both on global properties of the networks themselves, as well as predicting specific results for individual nodes. Our considerations did not produce a model that adequately helps explain cabinet member longevity or survival in a majority of the twelve countries. This should not come as a surprise: "parliamentary democracy" is a broad political term, and the twelve countries in this study differ in electoral systems, political views of constituents, constitutions, and other political variables. As such, building a mathematical model that explains all twelve countries might be impossible – at the very least, it would require more than just network information.

Importantly, some of the network-centric governmental models did outperform the baseline "basic model". This was the case with France, Luxembourg, New Zealand, Japan, and Canada. While these models fare better than the baseline model in a statistically significant way, the actual improvement in error rates is quite low. Exploring whether more complex models improve results further would be very useful. Furthermore, there might be interaction effects between network variables and standard political science variables, such as experience of cabinet members or occupations prior to becoming politicians [6].

We were more successful in predicting individual nodes' performance over time, and whether they will return as cabinet members in the future. Again, more algorithms should be implemented in this case, but it is encouraging to see that some of the variables seem to help improve predictions as to whether cabinet members will return

to office in the future. Indeed, only three of the ten countries analyzed in this research fare worse than their baseline models when the holdout set for model validation includes five governments. Of course, success rates vary depend on how large the holdout set is, and some countries (e.g. Japan) present a formidable challenge to the algorithms used. Exploring more algorithms and optimizing them for this specific problem is a useful extension to this research.

Japan is a very unique example in this analysis. Whereas most ministerial models outperform the baseline results and most governmental models do worse than the baseline, the Japanese strength model (at the governmental level) is better than the basic model, and the binary tree (at the ministerial level), is worse. As such, Japan presents a stark reminder that having an understanding, however slight, of cabinet networks at the nodal level does not mean we can extrapolate our knowledge to the whole network.

10.1 Future Work

A number of interesting extensions are possible. First, the models and algorithms implemented and developed above are by no means complete. Other models could be built that incorporate more information about the network structure of the cabinets, or look at the dynamics of cabinets in a different way. Similarly, other algorithms can be tested on the data sets in question to predict whether nodes will return as cabinet members in the future. Indeed, numerous network-based measures exist, and it is possible that some of the ones that were not explored in this work may be fruitful.

A very interesting opportunity for further work lies in implementing exponential random graph models (ERGMs) [31, 72] to help understand and evaluate the underlying structure of the cabinet data in this report. Specifically, ERGMs are a type of logit model [3] that estimates the probability of a given network in a specific space of networks. Thus, given a fixed parameter space, one can model the cabinet networks and whether they come from different types of random distributions of networks. ERGMs have been applied to bipartite graphs [65], as well as "neighbourhoods" (e.g. groups of nodes) in networks [53].

Second, it would be beneficial to combine our data with other data. Corruption indices, voting records, election results, legislative debates, committee memberships,

and numerous other forms of data are available about politicians in many of the countries discussed in this report, given the requisite time to find and extract them. Combining our data with other pieces of information can help build better models – ones that explore the qualitative, quantitative, and structural aspects of government and politicians' careers, and how they affect each other in the long run.

The question underlying much of this work deals with seeing whether it is possible to build models of different types of governance systems. Many aspects of government can be modelled and explored. Electoral systems, legislatures, and even grassroots revolutions are all within reach [1]. It might ultimately be possible to build more detailed models that take election results, cabinet membership decisions, and even policy outcomes into account.

Returning to a more "micro" level, every country is unique in terms of history, governing institutions, and laws. Rather than building a comparative model that tries to look at a group of parliamentary democracies, it may be more useful to build models for individual countries.

10.2 Conclusion

Based on the discussion above, network variables seem to play a role in the careers of cabinet ministers. While it is difficult to say whether their role is causal or simply correlated with political success, such variables nonetheless allow one to predict future success of individual cabinet members. Indeed, combining them with variables already explored by political scientists may yield even better results and, ideally, help drive development of political theory. Similarly, while the results for governmental models are less successful compared to the ministerial ones, they also provide some promising results, especially in the cases of France, Canada, New Zealand, Japan, and Luxembourg. Because the models focusing on these five countries did better than the baseline ones, there is support for the idea that network-centric variables play a role in understanding governments and cabinet-level politics in parliamentary democracies. With this in mind, a great deal of work lies ahead for those interested in investigating cabinet memberships using mathematical and statistical frameworks.

Appendix A

Symbols

Table A.1 lists some of the symbols used in this paper, and what they mean. Note that due to the size of this dissertation and the number of different symbols used, some of the conventional symbols (e.g. for centralities and clustering coefficients, which both use C) were substituted with more convenient ones.

Symbol	Description
G	A graph.
V	The vertex set of a graph.
v	A member of V.
v_i	The i^{th} member of V.
E	The edge set of a graph.
t	A unit of discrete time.
g_t	The t^{th} government.
c_i	The i^{th} cabinet member.
$\mathcal{G}_t$	The set of ministers present in government g_t.
$I_{\hat{t}}$	The set of incumbents from government $g_{\hat{t}}$.
$J_{t,\hat{t}}$	The set of incumbents in G_t that entered as newcomers in $G_{\hat{t}}$.
$\Omega = \{\omega_{ij}\}$	An experience matrix.
$\mathcal{C}_D$, $\mathcal{C}_B$, $\mathcal{C}_C$	Graph centralizations — degree, betweenness, and closeness, respectively.
$C_{D,i}$	The graph centrality score (in this case, degree) of the i^{th} node.
$d(c_i, c_j)$	The distance between c_i and c_j.
k_i	The degree or strength of the node v_i, in the case of unweighted and weighted networks, respectively.
T	The clustering coefficient of an unweighted network.
T_i	The clustering coefficient of the i^{th} node in an unweighted network.
T_O, T_Z, T_H	The clustering coefficient for a weighted network, under the definitions of Onnela et al., Zhang et al., and Holme et al., respectively.
$T_{O,i}$	The clustering coefficient of the i^{th} node in a network.
Q_H	The heterogeneity of a network.
Q_M	The modularity of a network.
Q_A	The assortativity of a network.
α	A coefficient in a model.
α_i	The i^{th} coefficient in a model.
z_i	The number of governments that c_i has served.

Table A.1: Symbols used in the report.

Appendix B

Code Overview

Due to the large amount of information stored in our data sets and the relative novelty of the work being undertaken, much of the software for the analysis and modelling of the data had to be built specifically for this project. The programming language used in the software development effort associated with this dissertation was Python. The software developed for this dissertation has a number of important parts, which we outline in Figure B.1.

The first part of the software is the "network data", which is the way that network information is stored at run time. There are several strategies for storing such information. Because a social network can be represented as an adjacency matrix, one can simply store a two-dimensional array. Another strategy is to store lists of neighbours for each node. In the case of this software, network data is stored in the form of edge lists: this is a list of values that create a function where, given a source and target node, an edge weight is returned, or 0 is returned if such a connection does not exist.

Network data is most often used by "network functions", which take the aforementioned edge list and calculate properties like clustering coefficients or strengths. Node-level functions return an array of values, usually one for every node in the network, while global functions return one value for the entire network. One can use network functions on their own, or in actual "models", which use the functions to make decisions on how the network evolves within multiple time steps. Because most of the models in this dissertation are probabilistic, they depend on running a large number of simulations to understand the range of possible values attained by the networks. With this in mind, "simulators" control how many times a model is run and how data on the multiple simulations is stored.

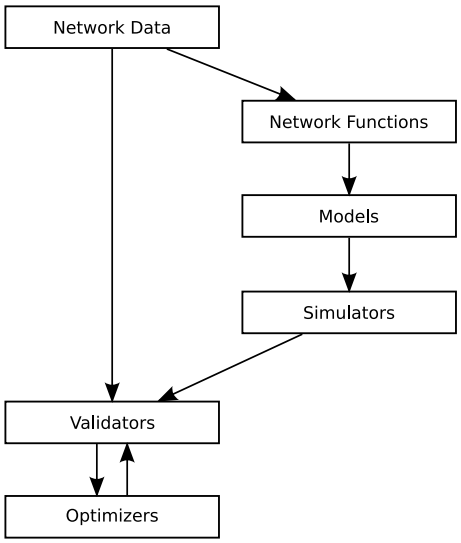


Figure B.1: The flow of information within the code developed to analyze the cabinet network data. Text surrounded by squares represents "entities" such as classes, and arrows represent the flow of information and data.

Simulators, network data, and network functions are used by "validators" to check how a specific model with specific parameter values compares to real-world data. At their simplest, validators calculate errors between the simulated networks and real-world, country-specific data. This is then used by "optimizers" to find optimal parameter values for the models being analyzed.

Unless otherwise stated, the measures, models, and optimization techniques discussed in this dissertation have all been coded in Python by the author.

Appendix C

Graph Visualization

C.1 Algorithms

Graph visualization is a complex and rich field of study, and while it is beyond the scope of this dissertation, it is useful to briefly review the methods we use to visualize the graphs in the cabinet network data set. The algorithm used to display graphs is the Fruchterman-Reingold [24] algorithm, implemented in the "igraph" [16] library. This algorithm is part of a family of models that treat a network as a set of masses (i.e. nodes) connected to each other by springs (i.e. edges) [19]. While these algorithms eschew real-world physics, their approach to laying out graphs has two important properties:

1. All nodes within a certain radius repel each other.
2. Neighbours attract each other and, at a certain radius, balance out the repulsive forces between each other.

Thus, laying out a graph becomes a computational exercise in finding an equilibrium between repulsive and attractive forces. The Fruchterman-Reingold algorithm calculates repulsive and attractive forces for a specific node $v \in V$ with the following equations:

$$\begin{aligned} f_{\text{attractive}}(d) &= d^2/k \\ f_{\text{repulsive}}(d) &= -k^2/d \end{aligned} \tag{C.1}$$

where d is the distance from the node in question to another node, and k is a constant. The borders of the layout also repel a node, ensuring that all nodes remain in the

boundaries of the image. Finally, to ensure that disconnected graphs do not repel each other indefinitely, two other points are worth mentioning:

1. The repulsive force only acts up to a certain radius.
2. If a graph contains N components, the layout itself is first split into N partitions, each designated for a specific component. This prevents any individual components from centering themselves in the layout and repelling all others to the boundaries.

A graph is thereby laid out by calculating forces on individual nodes, moving them accordingly, and iterating as long as needed, or until an equilibrium is reached. The resulting forces are one-dimensional – rather than moving nodes according to the sum of forces in a two-dimensional plane, simulated annealing or other optimization techniques are used to find an optimal layout.

Because visualization algorithms were beyond the broader scope of this dissertation, the "igraph" [16] library was used to visualize the networks in the cabinet data set. Within this library, 500 annealing iterations are used to lay out the graph, with $k = |V|$.

In the case of weighted graphs, the edge weight acts as a multiplier for the attractive force outlined in Equation (C.1). For example, an edge weight of 2 causes twice as much attractive force between neighbours as an edge weight of 1.

An important recent result [50] shows that using a force-directed layout like the Fruchterman-Reingold algorithm is related to grouping nodes in a way that maximizes modularity (discussed in Section 4.3.2) between densely connected groups. This implies that optimal layouts – that is, ones with low energy states – group nodes in densely connected clusters near each other.

C.2 Visualizations

The visualizations in this section show the diversity of the data being analyzed. Figure C.1 shows how the algorithm lays out cliques — where all nodes are connected to all other nodes – and how newcomers affect this layout in C.1(b). Observe that the optimal layout for cliques is by setting nodes in an equidistant fashion. In Figure C.1(b),

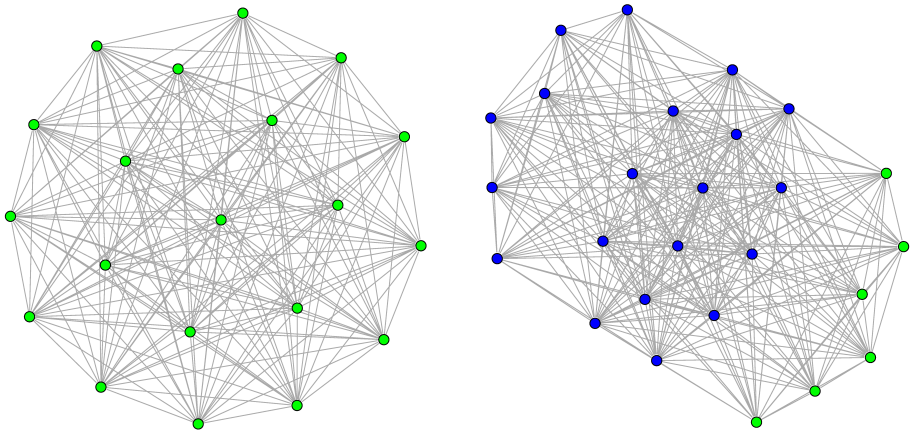
newcomers are shown in green, and the most connected nodes – the incumbents – are placed at the centre of the visualization.

Similarly, Figure C.2 shows the first 21 cabinets in Sweden. Tightly knit clusters are present here as well. Green nodes represent newcomers joining a group in the middle right of the graph. The central node in the graph is also one connected to many different politicians, indicating that the node likely participated in a relatively large number of cabinets.

If one chooses to analyze an individual cabinet, one can also see a wealth of diverse information. Figure C.3 shows Israel's 16^{th} cabinet. Edge weights are illustrated by the thickness of the line between nodes, and square nodes are newcomers. Edges representing the current government are ignored, and node colour represents political party. Observe how there are four parties participating in the cabinet itself, with some of the nodes connected to thicker edges than others. This suggests that some underlying dynamics – whether it be experience of politicians, party affiliations, or other nodal or network-level characteristics – help decide who will remain in power over an extended period of time.

We show a final example, Denmark's 27^{th} government, in Figure C.4. The largest component in the graph is composed of two separate parties, and it is interesting to see that not all incumbents have worked with each other in the past. This implies that being in the same political party does not serve as a predictor for prior cabinet co-membership.

Although visualizations help illustrate the wealth of data and information available in cabinet networks, they also underscore the need for quantitative analysis and modelling. There are numerous interesting features that one can try to better understand through modelling, and some of these are explored in sections of this dissertation. Importantly, the visualizations themselves are not meant to provide quantifiable information about the cabinets and their networks. Rather they are meant to aid in data exploration, and help us create hypotheses to subsequently investigate through quantitative analysis.



(a) Canada's first government. Notice that it is a simple clique with all nodes connected to all other nodes.

(b) Canada's second and first governments. Blue nodes represent cabinet members who were present in the first government, while green nodes are newcomers.

Figure C.1: Canada's first and second cabinets. Note that the edges in these visualizations are unweighted.

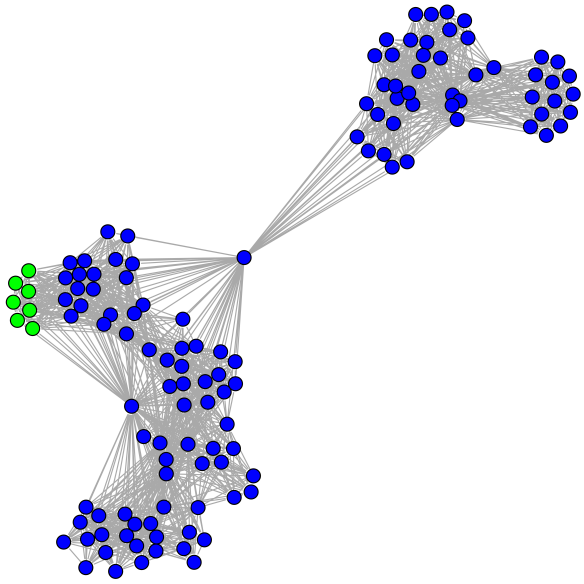


Figure C.2: Sweden's first 21 *cabinets. Notice that nodes organize into clusters, with a few nodes connected to many tightly-knit groups. Edges in this visualization are unweighted.*

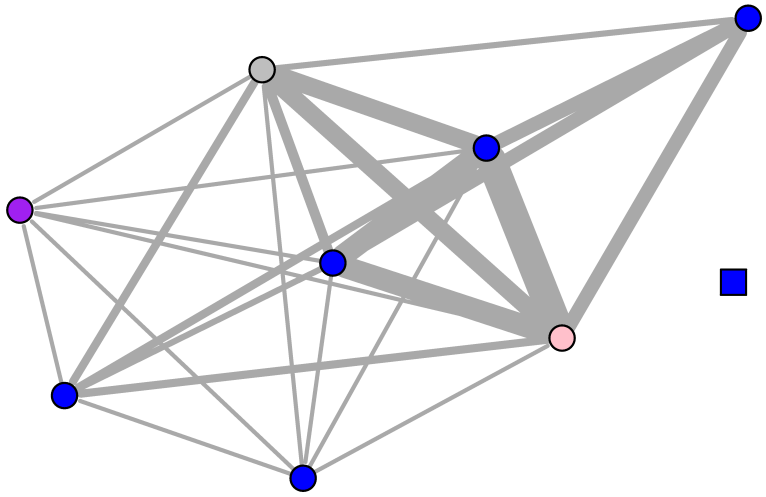


Figure C.3: Israel's 16^{th} cabinet. Political parties are represented by colour, and edges are thicker if their weights are higher. Notice the experienced group composed of blue nodes.

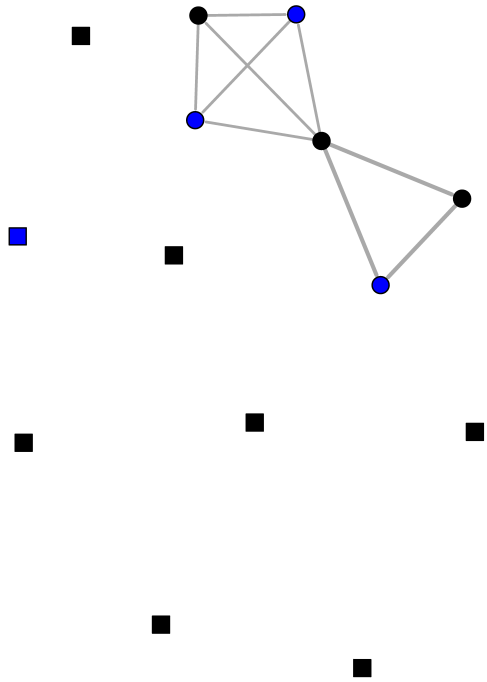


Figure C.4: Denmark's 27^{th} cabinet. Notice the presence of two parties (node colours), and the newcomers (squares). Also, notice that different nodes have only previously worked with a subset of incumbents.

Appendix D

Clustering Coefficients

The appendix contains three important notes on clustering coefficients which were not included in the main text of the dissertation due to space limitations.

D.1 Upper Bounds

An upper bound for w_i can be calculated using the inequality $a_{ij} \leq 1 - \delta_{ij}$ where δ_{ij} is the Kronecker delta. The specific steps for this are,

$$\begin{aligned}
w_i &= \frac{1}{2} \sum_{u \neq i} \sum_{\{v \mid v \neq i,\ v \neq u\}} a_{iu} a_{uv} a_{vi} \\
&\leq \frac{1}{2} \sum_{u \neq i} \left[a_{iu} \left(\sum_{v \neq i} (1 - \delta_{uv}) a_{vi} \right) \right] \\
&= \frac{1}{2} \sum_{u \neq i} a_{iu} \left[\sum_{v \neq i} (a_{vi} - \delta_{uv} a_{vi}) \right] \\
&= \frac{1}{2} \sum_{u \neq i} \left(a_{iu} \left[\sum_{v \neq i} a_{vi} - \sum_{v \neq i} \delta_{uv} a_{vi} \right] \right) \\
&= \frac{1}{2} \sum_{u \neq i} \left(a_{iu} \left[\left(\sum_{v \neq i} a_{vi} \right) - a_{ui} \right] \right) \\
&= \frac{1}{2} \left(\left[\sum_{u \neq i} a_{iu} \right]^2 - \sum_{u \neq i} a_{iu}^2 \right). \qquad \text{(D.1)}
\end{aligned}$$

D.2 Matrix-based Definition

Another way to view the Watts and Strogatz definition is through matrix multiplication [30]. Specifically,

$$T_i = \frac{\sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} a_{jk} a_{ki}}{\sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} a_{ki}}. \tag{D.2}$$

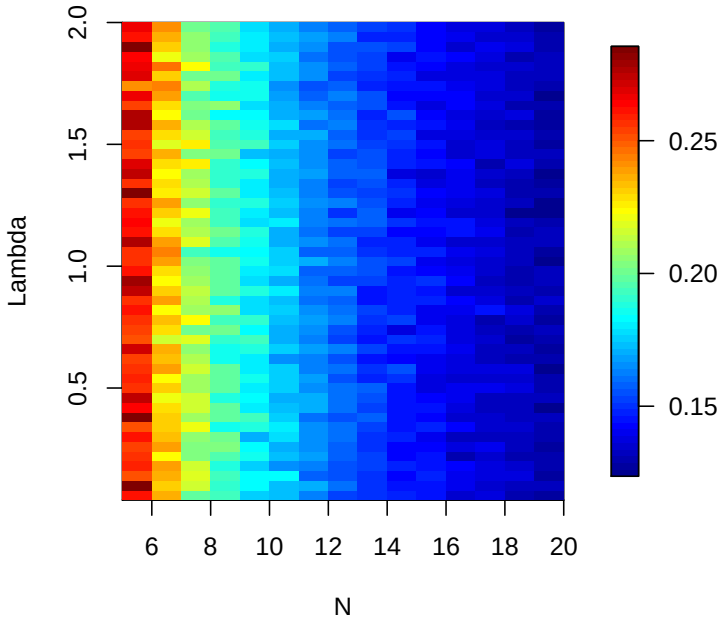
One can expand on this idea by defining a weighted clustering coefficient as

$$T_{H,i} = \frac{\sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij} a_{jk} a_{ki}}{\max(a) \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij} a_{ki}}. \tag{D.3}$$

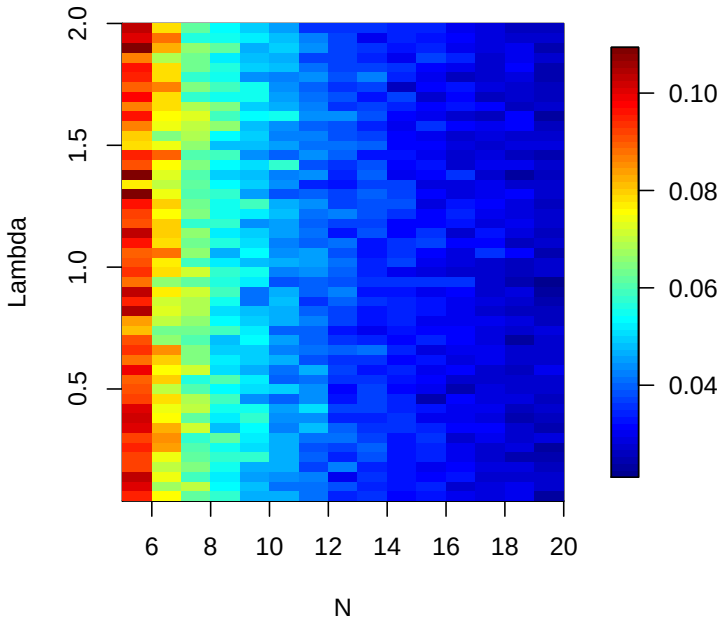
D.3 Discussion and Critique

Dealing with weighted clustering coefficients is challenging when one has different distributions of edge weights. In Equations (4.15) and (4.19), edges are resized to fit within $0 \leq a_{ij} \leq 1$ by dividing all edge values by the maximal edge value in the entire network. Hence, depending on the distribution of edge weights among edges (specifically, the value of the maximal edge compared to the mean values of edges), the clustering coefficient values vary. To show this, we ran simulations on networks of size 3 to 20 with edge weights being taken from an exponential distribution, $\texttt{exp}(\lambda)$, with λ varying between 0.05 and 2.0. The results are shown in Figure D.1(a) and Figure D.1(b), and show that as N, the number of nodes in a network, grows and thus increases the value of $\max_{ij} a_{ij}$ in the network, the mean clustering coefficient generally attains a lower value.

This is important to note, as it implies that one cannot necessarily take values of clustering coefficients from different networks and compare them. One first needs to ensure, for example, that edge weights are distributed in a similar manner.



(a) Mean T_O values for 100 simulations of graphs of various sizes (N) and edge distribution ($\texttt{exp}(\lambda)$).



(b) Standard deviations for Figure D.1(a).

Figure D.1: Various results for simulations for T_O in random graphs whose edge weights are values from a random variable following an exponential distribution.

Appendix E

Perron-Frobenius Theorems

The Perron-Frobenius theorems guarantee that given an adjacency matrix, the largest eigenvalue and eigenvector pair will be non-negative. This is crucial in the definition of eigenvector centrality (which is described in Section 4.1.4). Please note that the proofs below are taken from Meyer [43].

In the context of this dissertation, adjacency matrices are non-negative matrices, as we have $a_{ii} = 0$ for all i, and $a_{ij} \geq 0$ when $i \neq j$. Before seeing that such conditions allow for eigenvector centrality to exist, we first prove it for a simpler case, where all values $a_{ij} > 0$ for all i and j.

Following the notation of Meyer [43], $\rho(A)$ is the spectral radius of A, and given a vector $\mathbf{x}$, we define $|\mathbf{x}|$ as a vector with all absolute values of elements in $\mathbf{x}$. Given two vectors, $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} > \mathbf{b}$ means that for all entries in $\mathbf{a}$ and $\mathbf{b}$, we have $a_i > b_i$. Also, note that $||\mathbf{a}||_1 = \sum_i |a_i|$.

Similarly, given two matrices A and B, $|A|$ implies we take A with all absolute values for entries, and $A > B$ means the inequality holds true for all corresponding elements. Finally, $||A||_\infty = \max_{ij} |a_{ij}|$.

With Theorem 1, we prove that a matrix with positive entries has a positive eigenvector and eigenvalue pair (called an "eigenpair").

Theorem 1. *Given a positive matrix A, there exists an eigenvector with all non-negative entries.*

Proof:

Without loss of generality, assume $\rho(A) = 1$. *Suppose* $(\lambda, \mathbf{x})$ *is an eigenpair for* A *such that* $|\lambda| = 1$. *We then have:*

$$|\mathbf{x}| = |\lambda| \ |\mathbf{x}| = |\lambda \mathbf{x}| = |A\mathbf{x}| \leq |A| \ |\mathbf{x}|$$

This implies $|\mathbf{x}| \leq A \ |\mathbf{x}|$.

Now, let $\mathbf{z} = A \ |\mathbf{x}|$ *and let* $\mathbf{y} = \mathbf{z} - |\mathbf{x}|$. *Thus,* $\mathbf{y} \geq \mathbf{0}$.

Suppose that $\mathbf{y} \neq \mathbf{0}$, *so that we have* $A \ \mathbf{y} > \mathbf{0}$. *Furthermore, since* $\mathbf{y} = \mathbf{z} - |\mathbf{x}| > \mathbf{0}$, *we also have* $\mathbf{z} > |\mathbf{x}|$, *and thus and* $\mathbf{z} > \mathbf{0}$.

Thus, we know there is a number ϵ *such that* $A\mathbf{y} > \epsilon\mathbf{z}$. *Thus*

$$\begin{aligned} A\mathbf{y} &> \epsilon\mathbf{z}, \\ A(\mathbf{z} - |\mathbf{x}|) &> \epsilon\mathbf{z}, \\ A\mathbf{z} - \mathbf{z} &> \epsilon\mathbf{z}, \\ \frac{A}{1+\epsilon}\mathbf{z} &> \mathbf{z}. \end{aligned}$$

We then set $B = A/(1+\epsilon)$, *and we have that* $B^2\mathbf{z} > B\mathbf{z} > \mathbf{z}$, $B^3\mathbf{z} > B^2\mathbf{z} > \mathbf{z}$, *and so on. Thus, we have that* $B^k\mathbf{z} > \mathbf{z}$ *for* $k = 1, 2, \ldots$. *However, because we have that* $\rho(A) = 1$ *and* $\rho(B) = \rho(A/(1+\epsilon)) = 1/(1+\epsilon) < 1$, *we have that* $B^k \to 0$ *as* $k \to \infty$.

This leads to the conclusion that $\mathbf{0} > \mathbf{z}$, *which is a contradiction.*

Because our initial assumption was that $\mathbf{y} > \mathbf{0}$ *and this led to a contradiction, we have that* $\mathbf{y} = \mathbf{0}$. *By definition, this results in* $\mathbf{0} = \mathbf{z} - |\mathbf{x}| = A|\mathbf{x}| - |\mathbf{x}|$.

Thus, $A|\mathbf{x}| = |\mathbf{x}|$, *which means that* $|\mathbf{x}|$ *is an eigenvector for* 1.

□

Before we move on to non-negative matrices, we need to prove a lemma that allows us to transfer a form of the triangle inequality to matrices. Specifically, if we have $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

Lemma 1. $|A| \leq B$, *then* $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

Proof:

With the triangle inequality, we have $|A^k| \leq |A|^k$ *for every positive integer* k*. Also,* $|A| < B$ *implies* $|A|^k < B^k$*. Thus:*

$$\begin{aligned} & ||A^k||_\infty = || \ |A^k| \ ||_\infty \leq || \ |A|^k \ ||_\infty \leq ||B^k||_\infty \\ \Rightarrow \quad & \lim_{k\to\infty} ||A^k||_\infty^{1/k} \leq \lim_{k\to\infty} || \ |A|^k \ ||_\infty^{1/k} \leq \lim_{k\to\infty} ||B^k||_\infty^{1/k} \\ \Rightarrow \quad & \rho(A) \leq \rho(|A|) \leq \rho(B) \end{aligned}$$

□

Before moving on to proving Theorem 1 for non-negative matrices, it is useful to note the Bolzano-Weierstrass Theorem [66]. The theorem states, "Every bounded sequence has a convergent subsequence." This is used in the proof below.

Theorem 2. *If* A *is a non-negative matrix, then there exists an eigenvalue* $r = \rho(A)$ *such that* $A\mathbf{x} = r\mathbf{x}$ *with* $\mathbf{x} \geq 0$*.*

Proof:

For $k = 1, 2, \ldots,$ *define* $A_k = A + (1/k)E > 0$ *where* E *is a matrix with* 1 *in all entries. Thus, Theorem 1 applies to all* A_k*.*

Now, define the set $\mathcal{P} = \{\mathbf{p}_k\}_{k=1}^\infty$ *where* $\mathbf{p}_k$ *is the vector associated with* $r_k = \rho(A_k)$*. From Theorem 1,* r_k *is an eigenvalue.*

Furthermore, $\mathcal{P}$ *is a bounded set because it is contained in the unit-*1 *sphere. This is a result of Perron-Frobenius theory, but for the sake of brevity, is not included here. Furthermore, the theory states that* $||\mathbf{p}_k||_1 = 1$*.*

Using the Bolzano-Weierstrauss Theorem, there is a convergent subsequence $\{\mathbf{p}_{k_i}\} \to \mathbf{p}^\star$ *where* $\mathbf{p}^\star \geq \mathbf{0}$*.*

$\mathbf{p}^\star \neq \mathbf{0}$ *because* $\mathbf{p}_{k_i} > \mathbf{0}$ *and* $||\mathbf{p}_{k_i}||_1 = 1$*.*

Now, because $A_1 > A_2 > \cdots > A$*, Lemma 1 above guarantees that* $r_1 \geq r_2 \geq \cdots \geq r$*, so* $\{r_k\}_{k=1}^\infty$ *is a monotonic sequence of positive numbers that is bounded below by* r*. Thus,*

$$\lim_{k\to\infty} r_k = r^\star \qquad r^\star \geq r \qquad \lim_{i\to\infty} r_{k_i} = r^\star \geq r \tag{E.1}$$

Because $A_k \to A$ as $k \to \infty$, we also have $A_{k_i} \to A$ as $i \to \infty$. Thus, we can also establish:

$$A\mathbf{p}^\star = \lim_{i\to\infty} A_{k_i}\mathbf{p}_{k_i} = \lim_{i\to\infty} r_{k_i}\mathbf{p}_{k_i} = r^\star\mathbf{p}^\star \tag{E.2}$$

This shows that $r^\star$ is an eigenvalue for A, and since r is the maximal eigenvalue, we then have $r^\star \leq r$.

Therefore, $r^\star = r$ and $A\mathbf{p}^\star = r\mathbf{p}^\star$ with $\mathbf{p}^\star \geq \mathbf{0}$ and $\mathbf{p}^\star \neq \mathbf{0}$.

□

With Theorem 2, we know that given a matrix with non-negative entries, the eigenvector corresponding to the largest eigenvalue will be have non-negative entries as well. In a connected graph, this allows us to define and use eigenvector centrality.

Appendix F

Data Tables

Table F.1 shows a number of different measures and summary statistics for the twelve parliamentary democracies in this study.

Country	**Govts**	**Mean(# parties)**	**Mean (Cab.Size)**	**Mean (Newcomers)**	**Mean (Max Edge)**	**Min(Q_A)**	**Max(Q_A)**
Canada	18	1	22.5556	10.9444	3.2778	1	1
Denmark	27	2.1481	13.5556	5.7778	3.8889	−0.15294118	1
France	28	5.0714	19	5.2143	7.3929	−0.11807732	1
Germany	23	.	22.1739	9.7391	5.6957	.	.
Israel	35	.	9.0571	1.8571	7.7714	.	.
Japan	35	1.5429	14.5143	9.0857	2.2	−0.10786845	1
Luxembourg	15	2.4667	7.4667	3.6667	2.8	−0.21518987	0
Netherlands	21	3.5714	14.0952	7.5714	3.1905	−0.1343361	0.01320423
New Zealand	19	1	16.3158	7.3158	4	1	1
Norway	23	1.7826	15.6522	8.087	2.8696	−0.08454065	1
Sweden	21	2.0952	17.8095	6.0476	5.5714	−0.09939585	1
UK	18	1	20.3889	11.4444	2.9444	1	1

Table F.1: Some descriptive statistics by country. "Govts" represents the number of governments, which is followed by the mean number of parties in each government, the mean cabinet size, mean number of newcomers, and the mean maximal edge of each government. "Min(Q_A)" and "Max(Q_A)" give, respectively. the maximum and minimum of the assortativity coefficient.

Appendix G

ψ-Value Proofs

The proofs below show that the ψ-value for eigenvector centrality is any real number, and that all values for all nodes are equal. We do this using LU decomposition [58].

The problem statement itself is proving that the largest eigenvalue of an adjacency matrix, A, representing a clique of size n is $n-1$, and the eigenvector is any real-valued vector with equal entries. By definition, we have

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x}, \\ (A-\lambda I)\mathbf{x} &= \mathbf{0}, \\ \hat{A}\mathbf{x} &= \mathbf{0}, \\ \hat{A} &= \{\hat{a}_{ij}\}, \end{aligned}$$

where

$$\hat{a}_{ij} = \begin{cases} -\lambda & i = j \\ 1 & i \neq j \end{cases}.$$

We can decompose $\hat{A}$ into upper and lower triangular matrices, U and L respectively, to solve for the general $n \times n$ case. The theorem below helps us do this.

Theorem 3. *For the LU decomposition of $\hat{A}$, we have $L = \{l_{ij}\}$ be such that $l_{ii} = 1$, and entries below the diagonal are $l_{ij} = (\lambda - j + 1)^{-1}$. Also, $U = \{u_{ij}\}$ has elements*

$$u_{ij} = \begin{cases} \frac{(-1)(\lambda-j+1)(\lambda+1)}{\lambda-j+2} & \text{for } i = j \\ \frac{\lambda+1}{\lambda-j+2} & \text{for } i \neq j \end{cases}. \tag{G.1}$$

Proof:

This can be proven using the LU decomposition algorithm. The proof is inductive, with the initial result, for the first iteration (at $j = 1$) being proven by simply running through the algorithm.

At the k^{th} step, we have:

$$\begin{aligned} l_{ik} &= \frac{-\frac{\lambda+1}{\lambda-k+2}}{-\frac{(\lambda-k+1)(\lambda+1)}{(\lambda-k+2)}} \\ &= \frac{(-1)(\lambda+1)}{(-1)(\lambda-k+1)} \\ &= \frac{1}{\lambda-k+1} \end{aligned}$$

At this step, we also have

$$\begin{aligned} u_{k+1,k+1} &= \left(\frac{(-1)(\lambda-k+1)(\lambda+1)}{\lambda-k+2}\right) + \left(\frac{1}{\lambda-k+1}\right)\left(\frac{\lambda+1}{\lambda-k+2}\right) \\ &= \frac{(-1)(\lambda+1)(\lambda+k)}{\lambda-k+1} \\ \Rightarrow u_{\hat{k}\hat{k}} &= \frac{(-1)(\lambda+1)(\lambda+\hat{k}-1)}{\lambda-\hat{k}+2} \qquad \text{where } \hat{k} = k+1 \end{aligned}$$

Finally,

$$\begin{aligned} u_{k+1,j} &= \left(\frac{\lambda+1}{\lambda-k+2}\right) + \left(\frac{1}{\lambda-k+2}\right)\left(\frac{\lambda+1}{\lambda-k+2}\right) \\ &= \frac{\lambda+1}{\lambda-k+1} \end{aligned}$$

This implies,

$$u_{\hat{k},j} = \frac{\lambda+1}{\lambda-\hat{k}+2} \qquad \text{where } \hat{k} = k+1$$

□

To obtain the eigenvector centrality, we begin with the characteristic polynomial. Since $\det(\hat{A}) = \det(L)\det(U) = \det(U)$, it follows that:

$$\det(U) = \mathrm{Tr}(U) = (\lambda+1)^{n-1}(\lambda-n+1)$$

Thus, the eigenvalues for $\hat{A}$ are -1 and $n-1$, with the latter being positive and the largest value. Thus, this is the eigenvalue we use to find eigenvector centrality.

To find the associated eigenvector, we have that $A\mathbf{x} = \lambda\mathbf{x}$, which is equivalent to $\hat{A}\mathbf{x} = \mathbf{0} = LU\mathbf{x}$. Because the diagonal entries of L are all 1s and $\mathbf{x} \neq \mathbf{0}$, we can conclude that $L\mathbf{y} = \mathbf{0}$. This implies $\mathbf{y} = \mathbf{0}$, so our eigenvector search leads to $U\mathbf{x} = \mathbf{0}$.

Here, we can use back-substitution to solve for $\mathbf{x}$. By Theorem 3, we have $u_{nn} = \frac{(-1)(n-n)(n)}{n-1} = 0$, so x_n can be any real number. For $x_{n-1,n}$ we have,

$$\begin{aligned} u_{n-1,n-1} &= \frac{(-1)(n-n+1)(n)}{n-1-(n-1)+2} \\ &= \frac{-n}{2} \\ u_{n-1,n} &= \frac{n}{2} \\ \Rightarrow x_{n+1} &= x_n \end{aligned}$$

Using induction, we then get the $(n-k)^{\text{th}}$ row:

$$\begin{aligned} u_{n-k,n-k} &= \frac{(-1)(n-1-n+k+1)(n)}{n-1-n+k+2} \\ &= \frac{-kn}{k+1} \\ u_{n-k,n-k+i} &= \frac{n-1+1}{n-1-n+k+2} = \frac{n}{k+1} \qquad \text{for } i = 1, \ldots, k \end{aligned}$$

This implies,

$$\begin{aligned} u_{n-k,n-k}x_{n-k} + \sum_{i=1}^{k} u_{n-k,n-k+i}x_{n-k+i} &= 0 \\ u_{n-k,n-k}x_{n-k} &= \frac{-kn}{k+1}x_n \end{aligned}$$

Thus,

$$x_{n-k} = x_n.$$

Thus, the eigenvector for $\lambda = n-1$ is any vector where all entries are equal. This implies the ψ value for eigenvector centrality is any real number as well.

Appendix H

Model Optimization Results

The tables below list results for the model parameters that return, on average, the lowest error values for specific models. Table H.1 shows results for the basic model. Tables H.2 and H.3 report on the single-measure model and the variant which perturbed an initial probability. Finally, Table H.4 shows the results of simulations on the preferential attachment model.

Country	P	$\bar{e}$	P_{sim}	$\bar{e}_{\text{sim}}$	SD_e
Canada	0.5600	1.6331	0.5500	1.6466	0.0409
Denmark	0.5676	2.3052	0.5500	2.3132	0.0410
France	0.7400	1.4686	0.7500	1.4978	0.0428
Japan	0.3800	1.6470	0.3691	1.6717	0.0382
Luxembourg	0.5676	1.0204	0.5691	1.0390	0.0335
Netherlands	0.5000	1.7000	0.5000	1.7155	0.0429
New Zealand	0.5276	1.9871	0.5500	1.9997	0.0431
Norway	0.5276	1.7781	0.5000	1.7974	0.0396
Sweden	0.7124	1.6668	0.7000	1.6836	0.0411
UK	0.4800	1.4861	0.5000	1.4993	0.0404
Israel	0.8200	2.1396	0.8191	2.1704	0.0408
Germany	0.5876	1.4148	0.6000	1.4400	0.0461

Table H.1: Probability and error score for the basic model. Models were fitted using the Golden Section Search. The symbols P and $\bar{e}$ represent results from the analytical values, while P_{sim}, $\bar{e}_{sim}$, and SD_e represent the probability, mean error value, and standard deviation of the error value, respectively, based on simulations of the model.

Country	α_{str}	$\bar{e}_{\text{str}}$	$\hat{\sigma}_{\text{str}}$	α_{onn}	$\bar{e}_{\text{onn}}$	$\hat{\sigma}_{\text{onn}}$	α_{ev}	$\bar{e}_{\text{ev}}$	$\hat{\sigma}_{\text{ev}}$
Canada	0.0180	1.8202	0.0294	0.8000	1.5836	0.0902	1.7000	1.6752	0.0419
Denmark	0.0180	2.4513	0.0331	0.7000	2.7047	0.0612	1.5691	2.4546	0.0423
France	0.0190	1.8906	0.0380	0.9382	2.3256	0.0859	1.7191	2.0582	0.0430
Japan	0.0160	1.6362	0.0269	0.5382	1.8272	0.0559	1.2500	1.6697	0.0408
Luxembourg	0.0184	1.2267	0.0272	0.7382	1.3630	0.0445	1.7500	1.0882	0.0285
Netherlands	0.0180	1.7706	0.0354	0.7000	1.8861	0.0777	1.5000	1.7402	0.0380
New Zealand	0.0180	2.0514	0.0260	0.7000	2.3666	0.0552	1.5691	2.0839	0.0481
Norway	0.0174	1.9642	0.0283	0.8000	1.8212	0.0727	1.6500	1.8113	0.0366
Sweden	0.0190	2.1300	0.0420	0.9382	1.9900	0.0877	1.8000	1.9539	0.0361
UK	0.0174	1.5871	0.0277	0.6382	1.6078	0.0609	1.4500	1.5462	0.0427
Israel	0.0190	2.8385	0.0528	0.9382	2.9417	0.0776	1.6500	3.0385	0.0353
Germany	0.0180	1.6012	0.0258	0.8000	1.7534	0.0656	1.7000	1.4977	0.0383

Table H.2: Single measure model fits and errors. 3600 *simulations per parameter set were run for the strength, T_O models (denoted by "onn"), and eigenvector centrality models. Models were fitted using the Golden Section Search algorithm. α_X, $\bar{e}_X$, and $\hat{\sigma}_X$ represent the coefficient, mean error, and standard deviation of the error for each model, respectively, with "str", "onn", and "ev" referring to strength, T_O, and eigenvector centrality models. Red text means that the model's error rate is lower than the baseline model at a statistically significant level, based on t-tests [75].*

Country	$P_{0,str}$	α_{str}	$\bar{e}_{\text{str}}$	$\hat{\sigma}_{\text{str}}$	$P_{0,onn}$	α_{onn}	$\bar{e}_{\text{onn}}$	$\hat{\sigma}_{\text{onn}}$	$P_{0,ev}$	α_{ev}	$\bar{e}_{\text{ev}}$	$\hat{\sigma}_{\text{ev}}$
Canada	0.5367	0.0000	1.6364	0.0319	0.5498	0.0613	1.6494	0.0699	0.5606	0.0500	1.6526	0.0528
Denmark	0.5488	0.0013	2.3080	0.0308	0.5676	0.0000	2.3440	0.0674	0.5676	-0.0619	2.3257	0.0489
France	0.7355	0.0000	1.4798	0.0281	0.7697	0.0000	1.5459	0.0728	0.7400	0.0461	1.5101	0.0490
Japan	0.3741	0.0001	1.6545	0.0261	0.4145	-0.0248	1.6959	0.0600	0.3757	-0.0464	1.6707	0.0408
Luxembourg	0.5461	0.0006	1.0263	0.0241	0.5676	0.0000	1.0650	0.0513	0.5676	0.0500	1.0464	0.0386
Netherlands	0.4774	0.0000	1.7043	0.0314	0.5287	-0.0489	1.7263	0.0640	0.5000	0.0000	1.7143	0.0488
New Zealand	0.5119	0.0004	1.9868	0.0308	0.5276	0.0000	2.0137	0.0729	0.5482	-0.0506	1.9915	0.0537
Norway	0.5498	-0.0006	1.7836	0.0279	0.5223	-0.0155	1.8136	0.0620	0.5135	0.0757	1.7909	0.0455
Sweden	0.7197	0.0000	1.6735	0.0313	0.7238	0.0170	1.7005	0.0573	0.7403	-0.0757	1.6863	0.0508
UK	0.4861	0.0000	1.4882	0.0290	0.4800	0.0000	1.5227	0.0711	0.4818	-0.0409	1.5004	0.0525
Israel	0.8087	0.0000	2.1497	0.0269	0.8506	-0.0511	2.1846	0.0685	0.7968	0.0607	2.1679	0.0507
Germany	0.5876	0.0000	1.4238	0.0312	0.5876	0.0000	1.4628	0.0654	0.5678	-0.0505	1.4460	0.0485

Table H.3: Single measure models with perturbed probabilities were fitted and errors were calculated. This represents 3600 *simulations per perturbation, with* 50 *iterations of simulated annealing used. The symbols* α_X*,* $\bar{e}_X$*, and* $\hat{sigma}_X$ *represent the coefficient, mean error, and standard deviation of the error for each model, respectively. "str", "onn", and "ev" refer to the strength, clustering coefficient (*T_O*), and eigenvector centrality models. Red text means that the model's error rate is lower than the baseline model at a statistically significant level, based on t-tests [75].*

Country	α_1	α_2	P_{new}	$\bar{e}$	$\hat{\sigma}$
Canada	0.5000	0.5000	0.5000	1.6975	0.0179
Denmark	0.4993	0.4993	0.4997	2.4021	0.0170
France	0.1667	0.1667	0.3333	1.4554	0.0200
Japan	0.5000	0.4688	0.4688	1.8660	0.0169
Luxembourg	0.4995	0.4995	0.5000	1.0843	0.0166
Netherlands	0.5000	0.5000	0.5000	1.7497	0.0160
New Zealand	0.5000	0.5000	0.5000	2.0284	0.0212
Norway	0.5000	0.5000	0.5000	1.8352	0.0151
Sweden	0.5417	0.5417	0.3333	1.7860	0.0211
UK	0.5000	0.5000	0.5000	1.5549	0.0181
Israel	0.5833	0.5833	0.1667	2.2964	0.0233
Germany	0.5208	0.5208	0.4167	1.4605	0.0199

Table H.4: Preferential attachment model fits, fitted using the Nelder-Mead algorithm, with 50 *iterations, and* 3600 *simulations per parameter set. Red text means that the model's error rate is lower than the baseline model at a statistically significant level, based on t-tests [75].*

References

[1] D. Acemoglu and J.A. Robinson. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press, 2006.

[2] L.A. Adamic, B.A. Huberman, A.L. Barabási, R. Albert, H. Jeong, and G. Bianconi. Power-Law Distribution of the World Wide Web. *Science*, 287(5461):2115, 2000.

[3] A. Agresti. *Categorical Data Analysis*. Wiley-Interscience, 2002.

[4] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer Verlag, 1999.

[5] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509, 1999.

[6] L. Beckman. The Competent Cabinet? Ministers in Sweden and the Problem of Competence and Democracy. *Scandinavian Political Studies*, 29(2):111–129, 2006.

[7] S. Berlinski, T. Dewan, and K. Dowding. The Length of Ministerial Tenure in the United Kingdom, 1945–97. *British Journal of Political Science*, 37(02):245–262, 2007.

[8] P.M. Blau. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. Free Press New York, 1977.

[9] B. Bollobás. *Modern Graph Theory*. Springer, 1998.

[10] P. Bonacich. Factoring and Weighting Approaches to Status Scores and Clique Identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.

[11] R.S. Burt. *Structural Holes: The Social Structure of Competition*. Belknap Press, 1995.

[12] R.S. Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004.

[13] C.T. Butts. Revisiting the Foundations of Network Analysis. *Science*, 325(5939):414, 2009.

[14] W.K.T. Cho and J.H. Fowler. Legislative Success in a Small World: Social Network Analysis and the Dynamics of Congressional Legislation. *Journal of Politics*, forthcoming.

[15] J.G. Cleary and L.E. Trigg. K*: An instance-based learner using an entropic distance measure. In *12th International Conference on Machine Learning*, pages 108–114, 1995.

[16] G. Csardi and T. Nepusz. The igraph Software Package for Complex Network Research. *InterJournal Complex Systems*, 1695, 2006.

[17] L. Davis. *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1987.

[18] D. Diermeier and A. Merlo. Government Turnover in Parliamentary Democracies. *Journal of Economic Theory*, 94(1):46–79, 2000.

[19] P. Eades. A Heuristic for Graph Drawing. *Congressus Numerantium*, 42:194–202, 1984.

[20] A.Q. Flores. The Political Survival of Foreign Ministers. *Foreign Policy Analysis*, 5:117–133, 2009.

[21] J.H. Fowler, T.R. Johnson, J.F. Spriggs II, S. Jeon, and P.J. Wahlbeck. Network Analysis and the Law: Measuring the Legal Importance of Supreme Court Precedents. *Political Analysis*, 15(3):324–346, 2007.

[22] L.C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, 1977.

[23] L.C. Freeman. Centrality in Social Networks: Conceptual Clarification. *Social networks*, 1(3):215–239, 1979.

[24] T.M.J. Fruchterman and E.M. Reingold. Graph Drawing by Force-Directed Placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.

[25] M.S. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360, 1973.

[26] W. Gryc. The Social and Communication Networks of a Grassroots Organization in Kibera, Kenya. *Proceedings of the 2009 International Workshop on Intercultural Collaboration*, 2009.

[27] R. Guimerà, B. Uzzi, J. Spiro, and L.A.N. Amaral. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308(5722):697–702, 2005.

[28] R. Gulati, N. Nohria, and A. Zaheer. Strategic Networks. *Strategic Management Journal*, 21(3):203–215, 2000.

[29] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[30] P. Holme, S. Min Park, B.J. Kim, and C.R. Edling. Korean University Life in a Network Perspective: Dynamics of a Large Affiliation Network. *Physica A*, 373:821–830, 2007.

[31] D.R. Hunter, M.S. Handcock, C.T. Butts, S.M. Goodreau, and M. Morris. ERGM: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, 24(3):1–29, 2008.

[32] G. Kalna and D.J. Higham. A Clustering Coefficient for Weighted Networks, With Application to Gene Expression Data. *AI Communications*, 20(4):263–271, 2007.

[33] L. Katz. A New Status Index Derived from Sociometric Analysis. *Psychometrika*, 18(1):39–43, 1953.

[34] G. King, J.E. Alt, N.E. Burns, and M. Laver. A Unified Model of Cabinet Dissolution in Parliamentary Democracies. *American Journal of Political Science*, pages 846–871, 1990.

[35] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1999.

[36] S. Le Cessie and J.C. Van Houwelingen. Ridge Estimators in Logistic Regression. *Applied Statistics*, pages 191–201, 1992.

[37] M. Leiserson. Factions and Coalitions in One-Party Japan: An Interpretation Based on the Theory of Games. *The American Political Science Review*, pages 770–787, 1968.

[38] N. Lin. Social Networks and Status Attainment. *Annual Review of Sociology*, 25(1):467–487, 1999.

[39] N. Lin. Building a Network Theory of Social Capital. *Social Capital: Theory and Research*, pages 3–29, 2001.

[40] C.D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press New York, NY, USA, 2008.

[41] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[42] R.K. Merton. The Matthew Effect in Science: The Reward and Communication Systems of Science are Considered. *Science*, 159(3810):56–63, 1968.

[43] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[44] J.L. Moreno and H.H. Jennings. Statistics of Social Configurations. *Sociometry*, 1(3/4):342–374, 1938.

[45] J.A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308, 1965.

[46] M.E.J. Newman. Mixing Patterns in Networks. *Physical Review E*, 67(2):026126, 2003.

[47] M.E.J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, 2003.

[48] M.E.J. Newman. The Mathematics of Networks. *The New Palgrave Encyclopedia of Economics*, 2007.

[49] M.E.J. Newman and M. Girvan. Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69(2):26113, 2004.

[50] A. Noack. Modularity Clustering is Force-Directed Layout. *Physical Review E*, 79(2):26102, 2009.

[51] J.P. Onnela, J. Saramäki, J. Kertész, and K. Kaski. Intensity and Coherence of Motifs in Weighted Complex Networks. *Physical Review E*, 71(6):65103, 2005.

[52] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. 1998.

[53] P. Pattison and G. Robins. Neighborhood-Based Models for Social Networks. *Sociological Methodology*, pages 301–337, 2002.

[54] M.A. Porter, P.J. Mucha, MEJ Newman, and AJ Friend. Community Structure in the United States House of Representatives. *Physica A: Statistical Mechanics and its Applications*, 386(1):414–438, 2007.

[55] M.A. Porter, P.J. Mucha, MEJ Newman, and C.M. Warmbrand. A Network Analysis of Committees in the US House of Representatives. *Proceedings of the National Academy of Sciences*, 102(20):7057–7062, 2005.

[56] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

[57] D.S. Price. A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science*, 27(5), 1976.

[58] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer, 2006.

[59] R. Rose. The Making of Cabinet Ministers. *British Journal of Political Science*, pages 393–414, 1971.

[60] S.M. Ross. *Introduction to Probability Models*. Elsevier, 2007.

[61] S. Saavedra, F. Reed-Tsochas, and B. Uzzi. A Simple Model of Bipartite Cooperation for Ecological and Organizational Networks. *Nature*, 457(7228):463–466, 2008.

[62] J. Saramäki, M. Kivelä, J.P. Onnela, K. Kaski, and J. Kertész. Generalizations of the Clustering Coefficient to Weighted Complex Networks. *Physical Review E*, 75(2):27105, 2007.

[63] H.A. Simon. On a Class of Skew Distribution Functions. *Biometrika*, 42(3-4):425–440, 1955.

[64] J. Skvoretz. Salience, Heterogeneity and Consolidation of Parameters: Civilizing Blau's Primitive Theory. *American Sociological Review*, pages 360–375, 1983.

[65] J. Skvoretz and K. Faust. Logit Models for Affiliation Networks. *Sociological Methodology*, pages 253–280, 1999.

[66] M. Spivak. *Calculus*. Addison Wesley, 1967.

[67] K. Strom, I. Budge, and M.J. Laver. Constraints on Cabinet Formation in Parliamentary Democracies. *American Journal of Political Science*, pages 303–335, 1994.

[68] M. Tavits. The Role of Parties' Past Behavior in Coalition Formation. *American Political Science Review*, 102(04):495–507, 2008.

[69] B. Uzzi, D. Diermeier, P.J. Mucha, and M.A. Porter. Personal communication, July 2009.

[70] W.D. Wallis. *A Beginner's Guide to Graph Theory*. Birkäuser, 2000.

[71] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 2005.

[72] S. Wasserman and P. Pattison. Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^*. *Psychometrika*, 61(3):401–425, 1996.

[73] D.J. Watts and S.H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393(6684):440, 1998.

[74] A.S. Waugh, L. Pei, J.H. Fowler, P.J. Mucha, and M.A. Porter. Party Polarization in Congress: A Social Networks Approach. *arXiv:0907.3509v1*, 2009.

[75] B.L. Welch. The Generalization of "Student's" Problem When Several Different Population Variances are Involved. *Biometrika*, 34(1-2):28–35, 1947.

[76] I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, second edition, 2005.

[77] J. Woldendorp, H. Keman, and I. Budge. *Handbook of Democratic Government: Party Government in 20 Democracies (1945-1990)*. Kluwer Academic Publishers, Dordrecht; Boston, 1993.

[78] B. Zhang and S. Horvath. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1):1128, 2005.

[79] Y. Zhang, A.J. Friend, A.L. Traud, M.A. Porter, J.H. Fowler, and P.J. Mucha. Community Structure in Congressional Cosponsorship Networks. *Physica A*, 387:1705–1712, 2008.