Research Objectives

- State a precise definition of network compartmentalization.
- Develop a generative model for compartmentalized networks.
- Introduce an analytic measure of network compartmentalization.
- Apply to the measurement of political polarization in congress.

Definition

Let the degree to which a graph is characterized by separation on group membership as a result of a preference for within-group edge formation be the compartmentalization of that graph.

Generative Model

A generative model for compartmentalized networks should capture a preference for in-group edge formation that is mediated by the relative number of available in-group edges remaining. The equation below describes the probability of selecting an in-group edge for a given network $G$, group memberships $(M)$ and preference for in-group edge formation $(\rho)$:

$$
\gamma = \frac{(D_{ij} - D_{ii}) \rho}{(D_{ij} - D_{ii}) \rho + ((1 - D_{ij}) - D_{ii}) (1 - \rho)}
$$

(1)

To generate a network using this probability, we simply repeat the process until the desired number of edges is achieved:

1. For $k \in [1, K]$
2. Sample Edge Within Community $\sim \gamma(T, \rho, M)$
3. if Edge Within Community then Sample S, R from Shared Community else Sample S, R from Different Community
4. end if
5. end for

This generative process has a number of desirable properties including respecting a perfect preference for in (out) group edges as long as they exist and producing a constant proportion of in-group edges when $\rho = 0.5$ (no preference for in or out group edges).

Modularity

Modularity is the most used measure of the degree to which groups are disconnected given a particular network structure. However, it is not invariant to the number of groups or network size.

Following Newman [1, 2], for a division of the graph into $L$ distinct communities, define an $L \times L$ matrix $Q$ whose $ij$ component is the proportion of edges in the original graph that connect nodes in group $i$ to those in group $j$.

The modularity of the graph is then defined by:

$$
Q = \sum_{ij} \left( e_{ij} - Q_{ij} \right)
$$

(2)

The first term $Q_{ij}$ is the fraction of edges that lie within communities while $e_{ij}$ is the expected proportion of edges that lie within communities in a graph in which the nodes have the same degrees but edges are placed at random without regard for the communities.

Compartmentalization

Due to the limitations posed by existing measures, I introduce a new measure of graph compartmentalization $T$. To be a valid measure of the intuitive definition of compartmentalization stated previously, $T$ must satisfy three properties:

1. $T$ must be invariant in $N$ and the number and relative size of communities for a constant $D_{ij}$.
2. $T$ must be bounded above and below to give a consistent measure of compartmentalization or anti-compartmentalization.
3. $T$ must only attain its global maximum (minimum) value when $D_{ij} = D_{ii}$ and ties are only present within (between) community.

Let $A$ be the graph adjacency matrix (with $|\mathcal{A}|$ the sum over the adjacency matrix). Then we can define $F$, the fraction of observed edges that occur within-groups as follows:

$$
F = \frac{\gamma(M, A)}{|\mathcal{A}|}
$$

(3)

For a given $F$ and $D_{ii}$, we can then define a measure of the compartmentalization of a graph $T$ as:

$$
T = \left[ F - D_{ii} \right] \times \begin{cases} 
\frac{[1 - (1 - 0.5) \gamma]}{[1 - (1 - 0.5) \gamma]} & \text{if } F \geq D_{ii} \\
\frac{[1 - (0.5) \gamma]}{[1 - (1 - 0.5) \gamma]} & \text{if } F < D_{ii}
\end{cases}
$$

(4)

The first term $F - D_{ii}$ bears a strong analogy to the measure of modularity $Q$ as it is just the proportion of in-community edges minus the expected proportion of in-community edges if $G$ were generated from the generative process described previously with $\rho = 0.5$, indicating no preference for within-group edge formation (see Figure 2). Panel b). $T$ is increasing in $F$ and decreasing in $D_{ii}$ as we can see by taking partial derivatives of $T$ with respect to $F$ and $D_{ii}$.

$$
\frac{\partial T}{\partial F} = \begin{cases} 
1 & \text{if } F \geq D_{ii} \\
\frac{1}{[1 - (1 - 0.5) \gamma]} & \text{if } F < D_{ii}
\end{cases}
$$

(5)

$$
\frac{\partial T}{\partial D_{ii}} = \begin{cases} 
\frac{[1 - (1 - 0.5) \gamma]}{[1 - (0.5) \gamma]} & \text{if } F \geq D_{ii} \\
\frac{[1 - (1 - 0.5) \gamma]}{[0.5 \gamma]} & \text{if } F < D_{ii}
\end{cases}
$$

(6)

This captures the intuition that more compartmentalized graphs have a higher proportion of within-group edges and that dense graphs are generally less partitioned, respectively.

Figure: Plots of average simulated proportion of edges within community versus graph density with varying $\rho$ and max within community density. All proportions averaged over 20,000 simulations.

Figure: Componentation coefficient $T$ values across different maximal within-group density $- $ density combinations. Graphs were simulated from generative process and procedures averaged over 20,000 simulations. The level plots display compartmentalization coefficients recovered from graphs generated with $(a) \; \rho = 0$, $(b) \; \rho = 0.5$, $(c) \; \rho = 0.9$, $(d) \; \rho = 1$.

Figure: Plot of political party modularity and compartmentalization in the Senate directed cosponsorship network.

Figure: Plot of political party modularity and compartmentalization in the Senate influence network.

Figure: Plot of political party modularity and compartmentalization in the Senate co-bill-cosponsorship network (left scale) and difference in party mean NOMINATE scores, used as a ground-truth measure of ideological polarization (right scale) from the 96th term of Congress (1979-1980) to the 109th term (2003-2004).

Measure Null Distribution

Table: Directed Cosponsorship

<table>
<thead>
<tr>
<th>Measure Value</th>
<th>P</th>
<th>Polarity</th>
<th>Measure Null Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization</td>
<td>0.764</td>
<td>0.055</td>
<td>1</td>
</tr>
<tr>
<td>Modularity</td>
<td>0.152</td>
<td>0.050</td>
<td>1</td>
</tr>
</tbody>
</table>

Table: Influence

<table>
<thead>
<tr>
<th>Measure Value</th>
<th>P</th>
<th>Polarity</th>
<th>Measure Null Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization</td>
<td>0.764</td>
<td>0.055</td>
<td>1</td>
</tr>
<tr>
<td>Modularity</td>
<td>0.152</td>
<td>0.050</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure: Plot of $T$ calculated for 99 million simulated networks with $\rho = 0.5$ across all $D_{ij} - D_{ii}$ combinations. The red density is for $D_{ij} = 0.1$, the orange density is for $D_{ij} = 0.3$, and the blue is averaged across all $D_{ij}$. The 95% confidence interval for the null distribution is $[0.4234, 0.4234]$.

Table: Permutated Regressions of Modularity and Compartmentalization on Polarization

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Polarity</th>
<th>Polarity</th>
<th>Modularity</th>
<th>Companynentation</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization</td>
<td>0.093</td>
<td>0.096</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>Modularity</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Correlation coefficients for two measures against party-mean nominee differences for the 96th-109th Congresses in the Senate co-bill-cosponsorship network.
