An Analytic Measure of Graph Compartmentalization

Matthew J. Denny

University of Massachusetts, Amherst | mdenny@polsci.umass.edu

Research Objectives

- State a rigorous 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 tie formation (ρ):

$$\gamma = \frac{(D_M - D_{in})\,\rho}{(D_M - D_{in})\,\rho + ((1 - D_M) - D_{out})\,(1 - \rho)} \qquad (1)$$

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

for $k \in K$ **do**
  Sample Edge Within Community $\sim \gamma(T, \rho, M)$
  if Edge Within Community **then**
    Sample S, R from Shared Community
  else
    Sample S, R from Different Community
  end if
end for

This generative process has several desirable properties including respecting a perfect preference for in (out) group edges so 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).

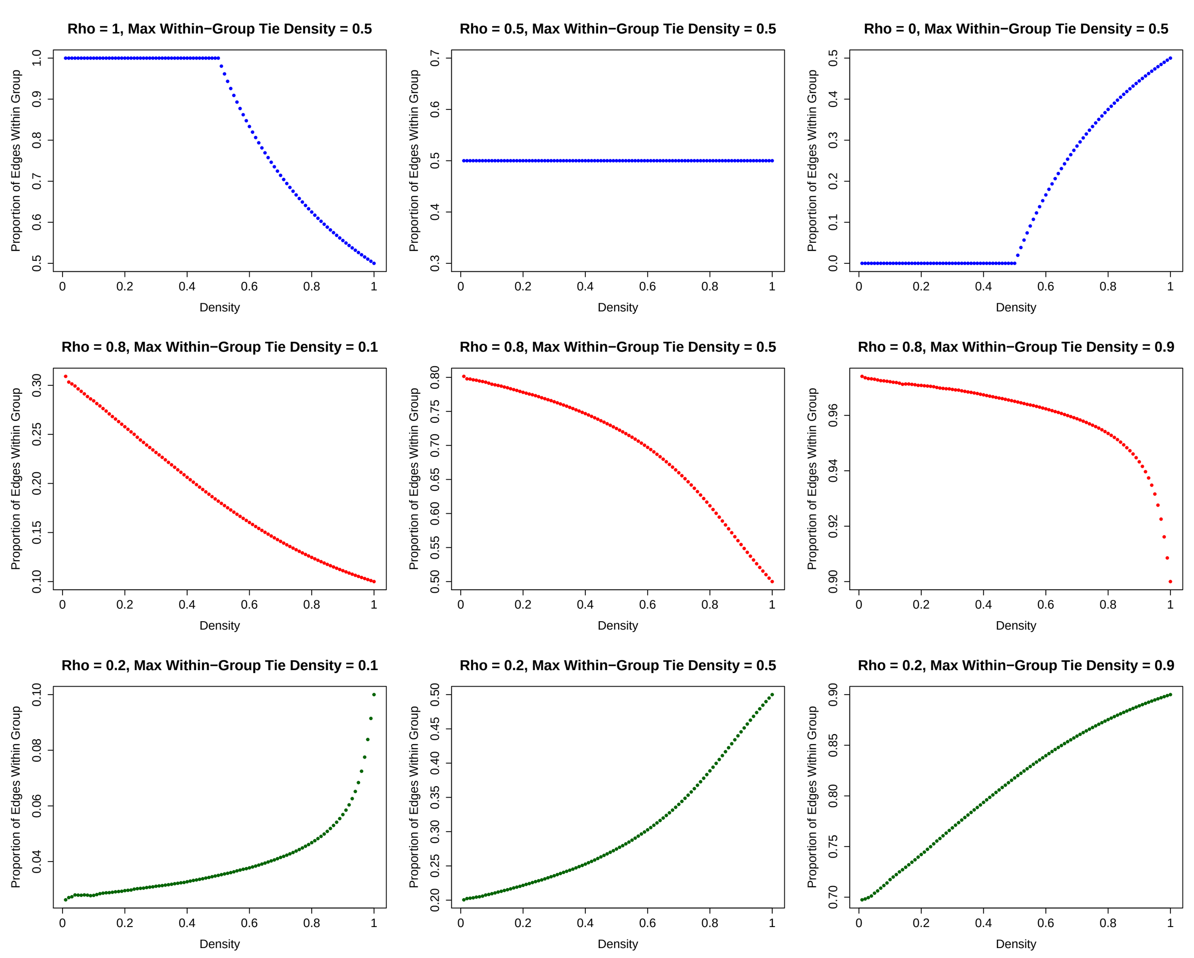


Figure: Plots of average simulated proportion of edges within community versus graph density with varying ρ and max within community density. All proportions averaged over 20,000 simulations.

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 e whose e_{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 to be:

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij} e_{ki} = \text{Tr } \mathbf{e} - ||\mathbf{e}^2|| \qquad (2)$$

The first term (Tr $\mathbf{e}$) is the fraction of edges that lie within communities, while $||\mathbf{e}^2||$ 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 Υ. To be a valid measure of the intuitive definition of compartmentalization stated previously, Υ must satisfy three properties:

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

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

$$F = \frac{\Sigma_i \Sigma_j M_{i,j} A_{i,j}}{||A||} \qquad (3)$$

For a given F and D_M, we can then define a measure of the compartmentalization of a graph Υ as:

$$\Upsilon = [F - D_M] \times \begin{cases} \text{if } F \geq D_M : \dfrac{\left[1-(D-D_M)^2\right]}{1-D_M} \\ \text{if } F < D_M : \dfrac{\left[1-(D-(1-D_M))^2\right]}{D_M} \end{cases} \qquad (4)$$

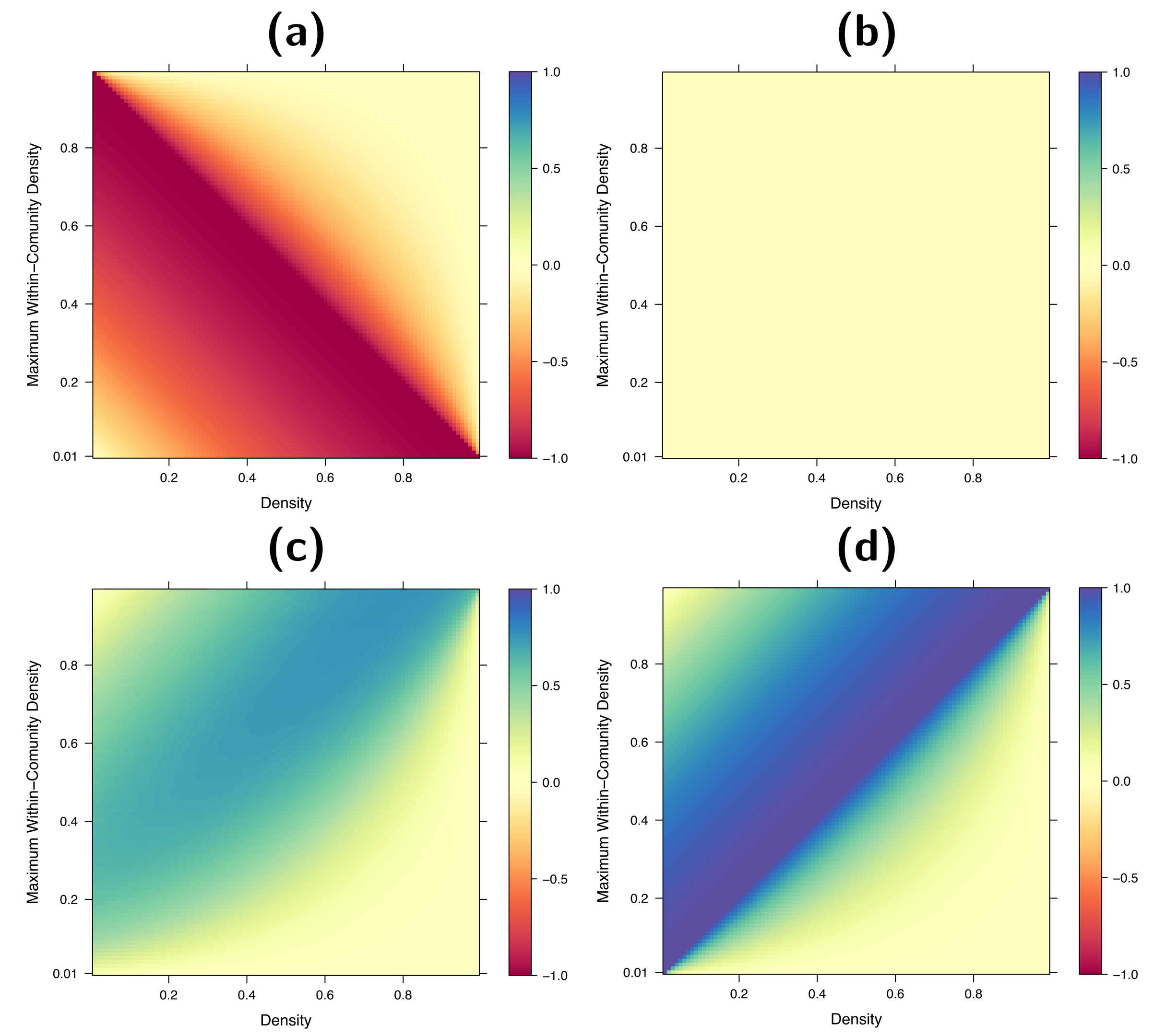
The first term, $[F - D_M]$ bears a strong analogy to the measure of modularity Q, as it is just the proportion of in-community edges minus the expected proportion of in-community edges if G were generated from the generative process described previously with $\rho = 0.5$, indicating no preference for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

$$\frac{\partial \Upsilon}{\partial F} = \begin{cases} \text{if } F \geq D_M : \dfrac{\left[1-(D-D_M)^2\right]}{1-D_M} \\ \text{if } F < D_M : \dfrac{\left[1-(D-(1-D_M))^2\right]}{D_M} \end{cases} \geq 0 \qquad (5)$$

$$\frac{\partial \Upsilon}{\partial D} = \begin{cases} \text{if } F \geq D_M : \dfrac{-2[(F+D_M)(D+D_M)]}{1-D_M} \\ \text{if } F < D_M : \dfrac{-2[F(1+D-D_M)+D_M(1+D)]}{D_M} \end{cases} \leq 0 \qquad (6)$$

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

Figure: Compartmentalization coefficient Υ values across different maximal within-group density – density combinations. Graphs were simulated from generative process and proportions 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)** : $\rho = 1$.



Polarization in Congress

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 108th term (2003-2004)

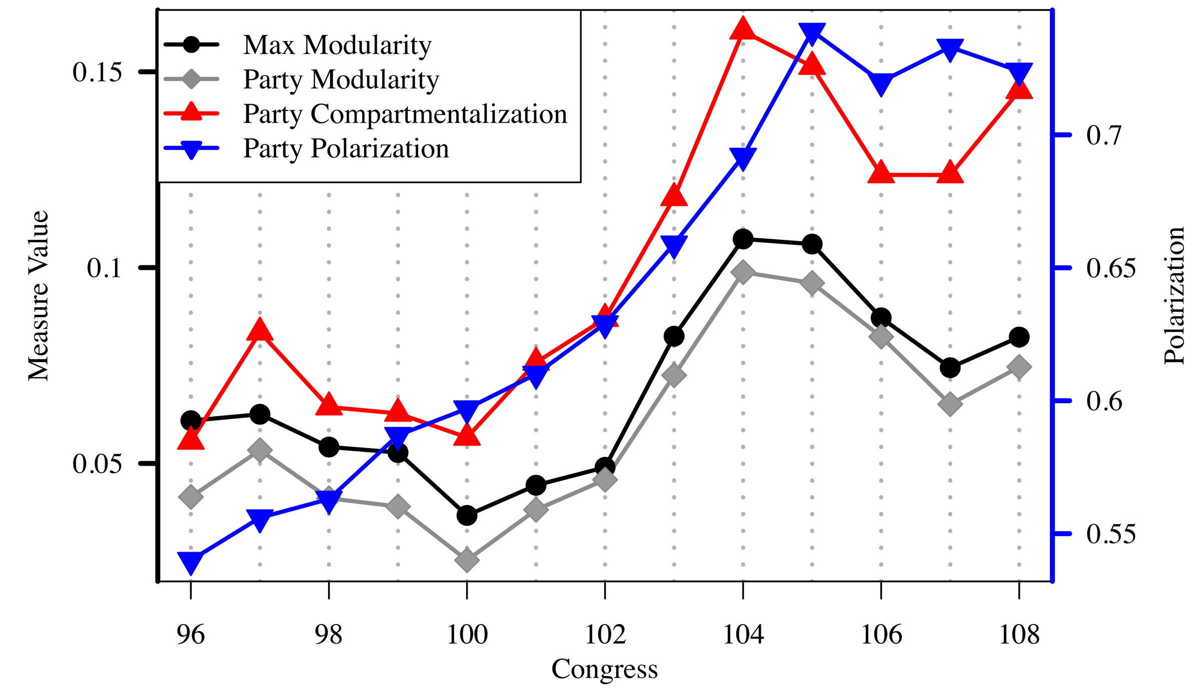


Table: Permuted Regressions of Modularity and Compartmentalization on Polarization

	Dependent variable:		
	Polarization	Polarization	p value
Modularity	2.455		0.0016
Compartmentalization		1.719	0.0002
Observations	13	13	
Adjusted R^2	0.6091	0.785	
F Statistic	19.7***	44.82***	
Iterations	62549	668122	

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

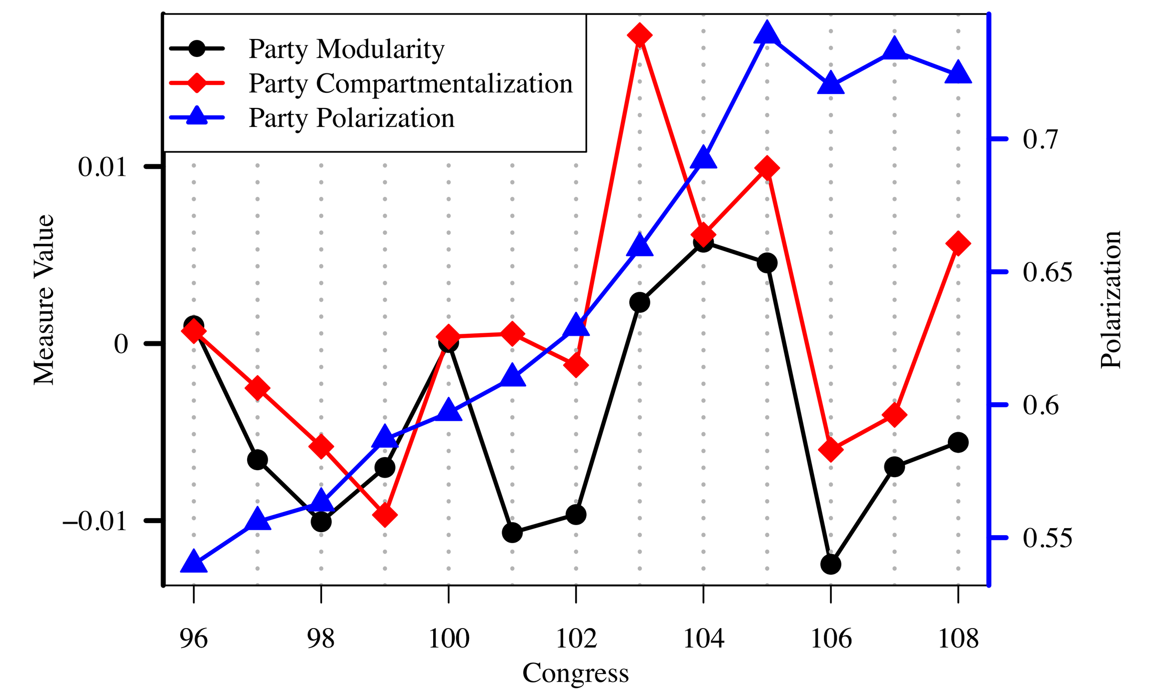


Figure: Plot of political party modularity and compartmentalization in the Senate influence network.

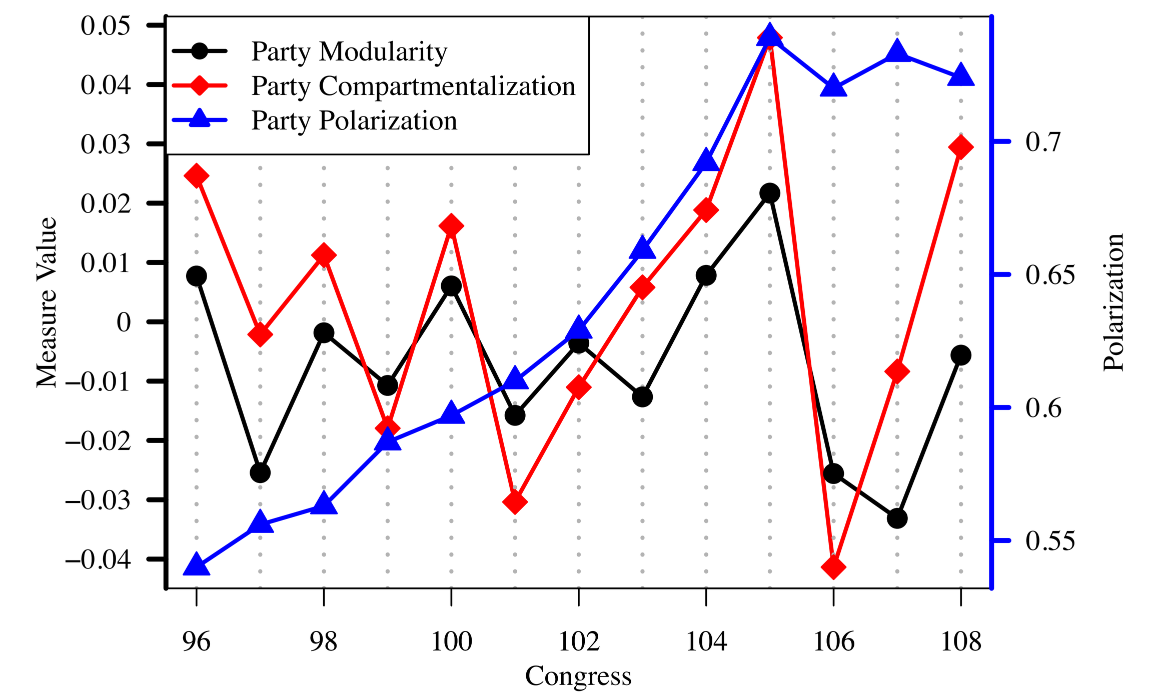


Table: Directed Cosponsorship

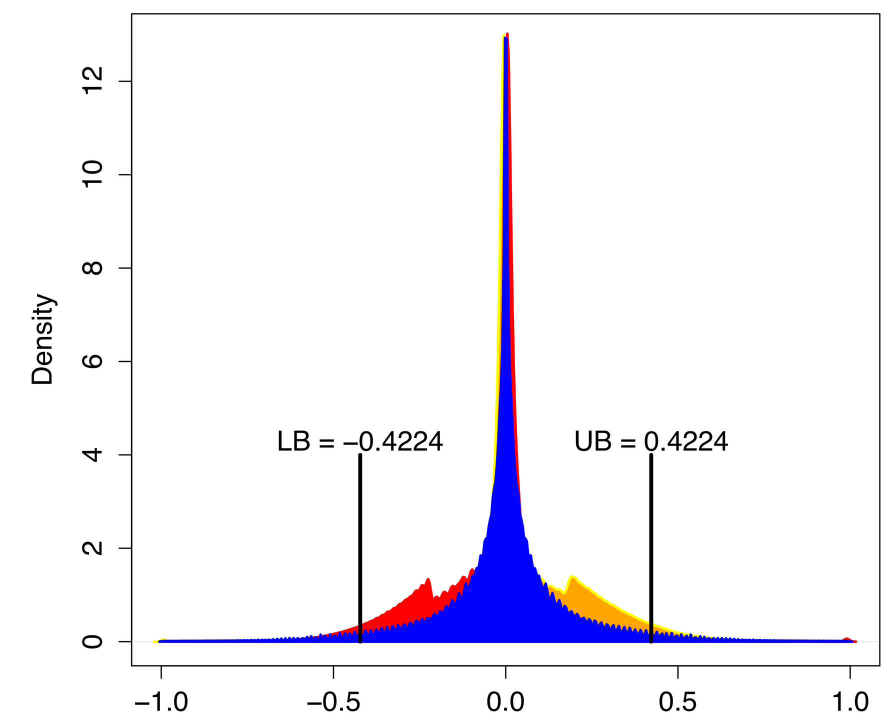
	Q	Υ	P
Υ	0.714	1	
P	0.152	0.359	1

Table: Influence

	Q	Υ	P
Υ	0.764	1	
P	-0.055	0.103	1

Measure Null Distribution

Figure: Plot of Υ calculated for 99 million simulated networks with $\rho = 0.5$ across all D - D_M combinations. The red density is for $D_M = 0.1$, the orange density is for $D_M = 0.9$, and the blue is averaged across all D_M. The 95% confidence interval for the null distribution is [-0.4224,0.4224].



[1] M. E. J. Newman. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter*, 38(2):321–330, March 2004.

[2] Mark E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–82, June 2006.

Correlation coefficients for two measures against party-mean nominate differences for the 96th-108th Congresses in the Senate co-bill-cosponsorship network.

	Q	Υ	P
Q	1		
Υ	0.956	1	
P	0.801	0.896	1

UMassAmherst

INSTITUTE FOR SOCIAL SCIENCE RESEARCH
UNIVERSITY OF MASSACHUSETTS AMHERST