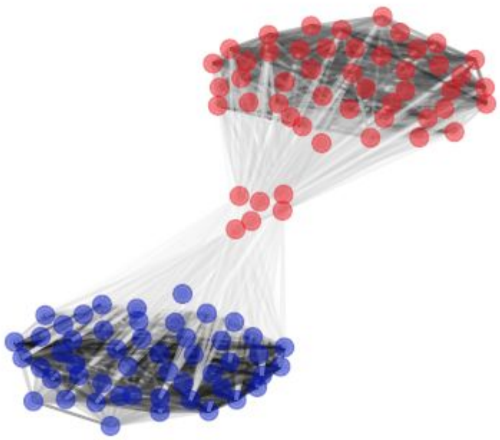
Centrality In Congressional Voting Networks

by Himanshu Rathi, Kevin Nguyen
supervised by Jeremy Kazimer

Motivation for Project

- In the news, we can see many headlines of Congress and its members being at a standoff.
  - “... Congress is so polarized” (CNN)
  - “Polarization defines the midterm election” (Fox News)
  - “Measuring America’s Divide: ‘It’s Gotten Worse’” (NYT)
- There are many ways to quantify this, so we want to study one of them to see if this is really the case.
- One of the ways which would help not only quantify, but also visualize, polarization is **network science**, which is the study and modeling of interactions between objects.
  - A very common network we encounter is a social network, or even a group of friends
- The U.S. Senate can be expressed as a network.



Overview

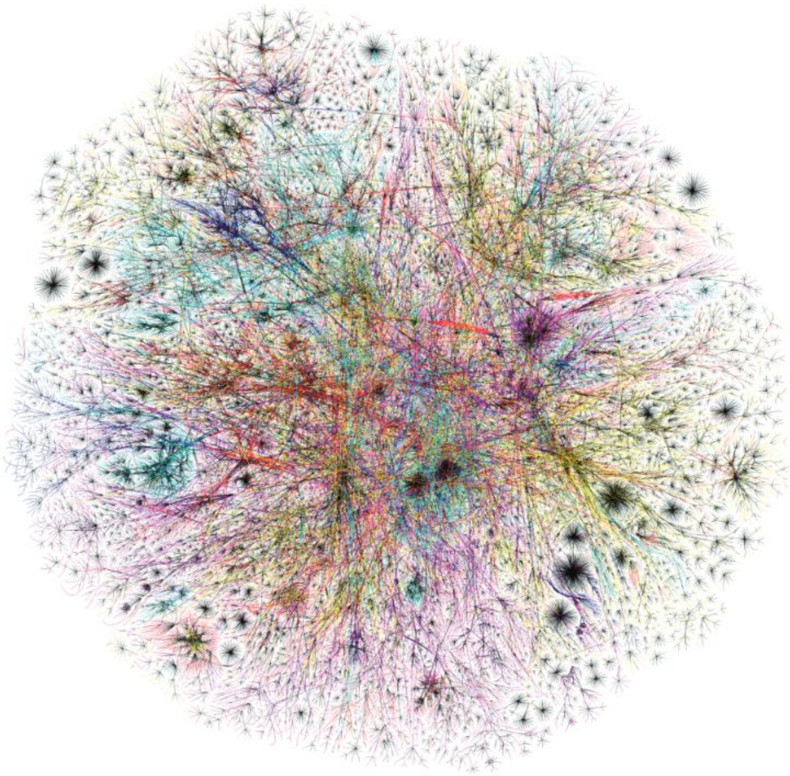
Part 1: Network Science Tools

Part 2: Constructing Congress as a Network

Part 3: Analysis of the Network

Part 4: Conclusions & Future Exploration

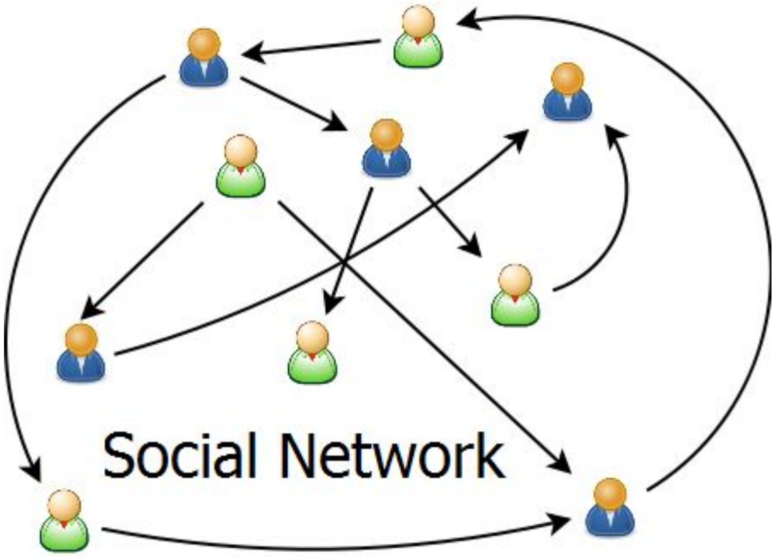
PART 1: Network Science Tools



Newman, Networks 2ed.

Motivations for network science

- Networks were developed as *sociograms* to study human interaction.
- **Network**: A mathematical object made up of **nodes** and **edges**, contained in some latent space.
- **Nodes** represent objects.
  - Examples include people, animals, and websites.
- **Edges** represent some relationship between the nodes.
  - Edges could represent friendship (people), predator/prey (animals), or links between pages (websites).



Josić, *Friendship Paradox*

Fundamental properties of a network

- The structural representation of a network in mathematics is an **adjacency matrix.**

$$A_{ij} = \begin{cases} 1 & (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

- The **degree** is a term used to describe the number of edges the are connected to the node.

$$d_i = \sum_{j=1}^{N} A_{ij}$$

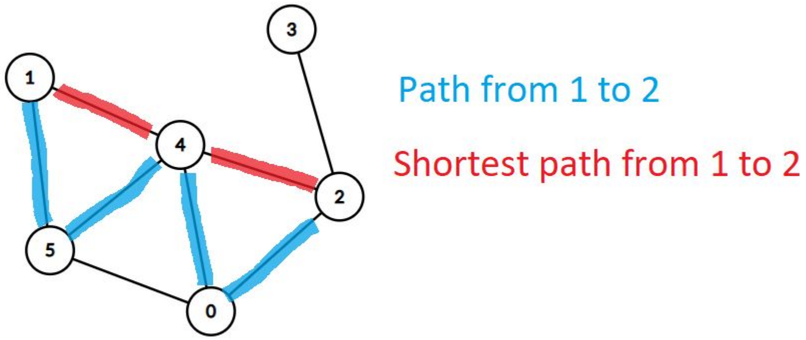
- The **number of edges** in a network is half the total degree of that network.

$$M = \frac{1}{2} \sum_{i=1}^{N} d_i$$

- The **density (ϱ)** is a measure of how connected a graph is, defined as the ratio of the number of edges in the graph to the maximum number of edges possible.
    - High density indicates that the network is well connected.

Paths and connected components

- A **path** is a sequence of nodes where each consecutive pair of unique nodes is connected by an edge.
- A **shortest path** between two nodes is a path with the least number of nodes.
- A **connected component** is a set of nodes that are connected to at least one other node by an edge.

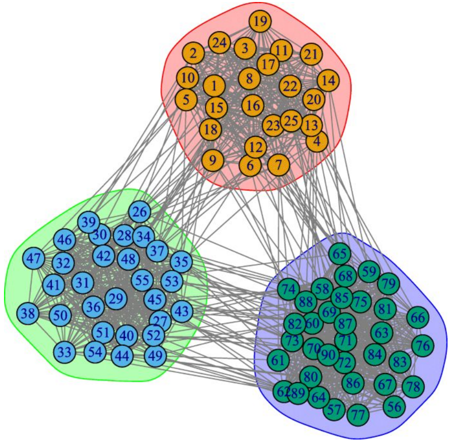


Community structure in networks

- **Community structure** is a division of set of nodes in a network into smaller subgroups or communities.
  - These nodes are more densely related to each other.
  - Important because they show the functions of the network.
- **Modularity** shows us the extent to which a network can be divided into distinct, relatively independent modules or clusters.
  - Modules are defined as groups of nodes that have a higher density of connections among themselves.

$$Q = \frac{1}{4M}\mathrm{Tr}\left(S^T\left[A - \frac{\mathbf{k}\mathbf{k}^T}{2M}\right]S\right)$$

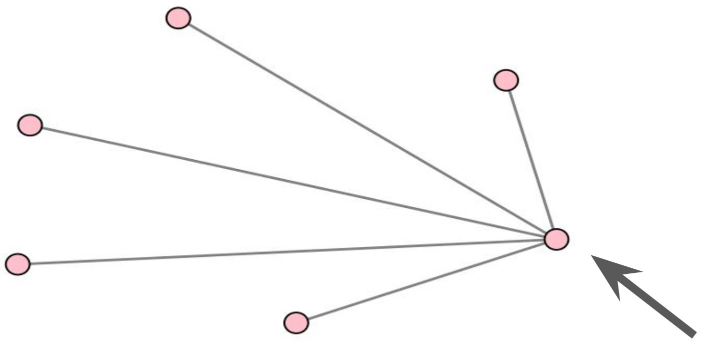
- We use **modularity maximization** to get a clear cut partition in the network.
  - Using statistical inference or iterative methods, we find the permutation of class identification that gives the largest modularity.



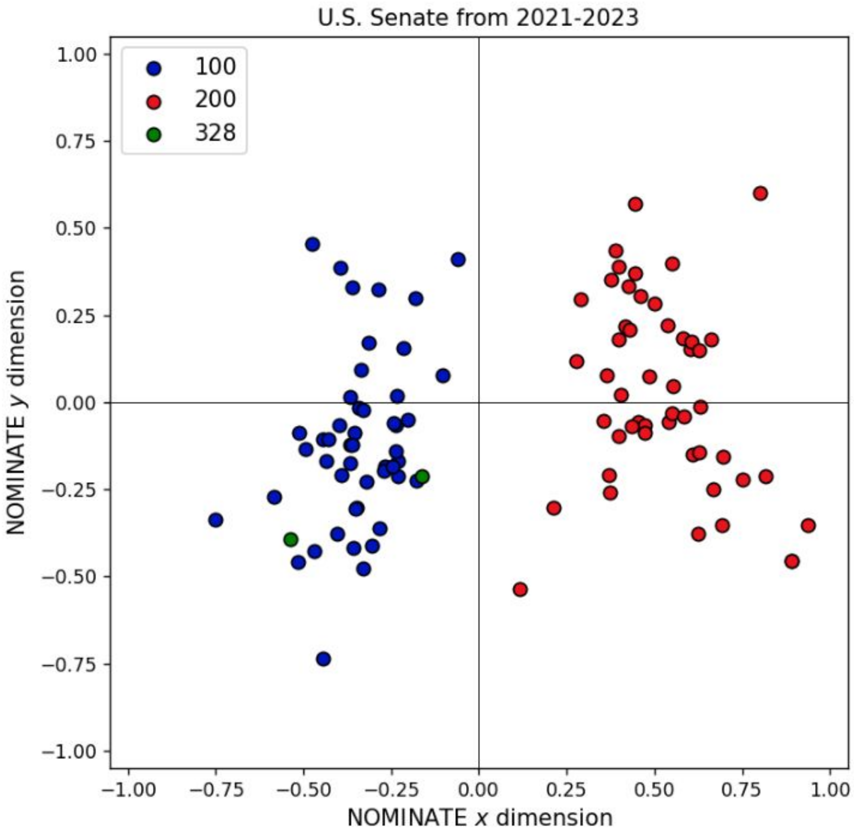
Lee, *A review of stochastic block models and extensions for graph clustering*

Centralities

- **Centrality** in a network determines the most important nodes, edges, or subgraphs.
- There are varying definitions of “important,” so there are many different measures of centrality:
  - **Degree**: a node’s centrality score is its degree.
  - **Eigenvector**: a node’s centrality score is proportional to the centrality of its neighbors.
    - This is the largest eigenvalue of the adjacency matrix and its corresponding eigenvector.
  - **Betweenness**: a node’s centrality score is the number of shortest paths from all nodes that include the node of interest, divided by all possible shortest paths.
    - This is very useful for determining connected components, community structure, etc.



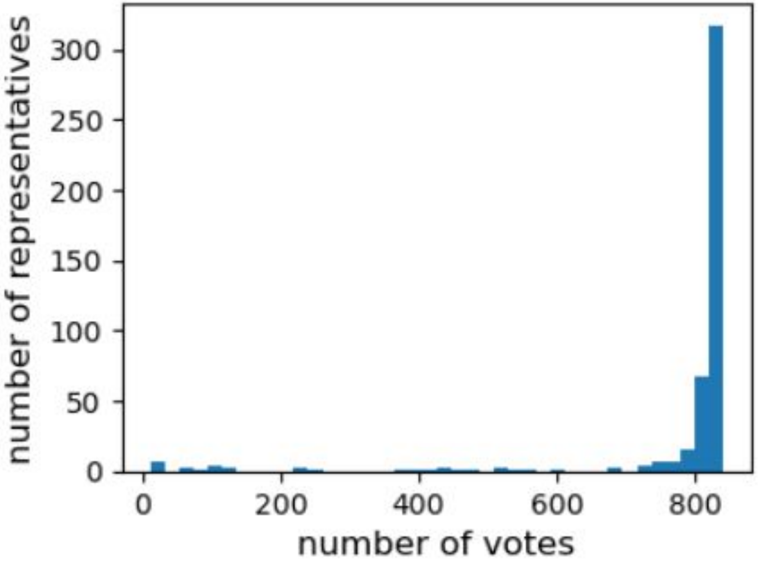
PART 2: Constructing Congress as a Network



VoteView

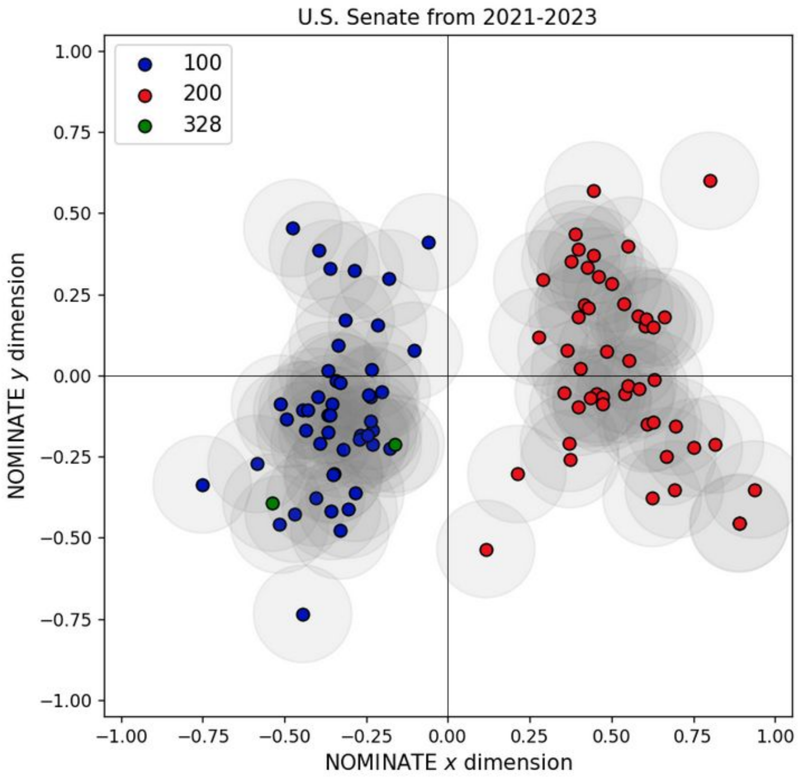
- VoteView is an online database that stores voting records and party ideologies for Congress.
    - We used the two NOMINATE dimensions as the criteria for our project.
        - NOMINATE represents the ideological similarity between congresspeople.
        - Both are represented as scores from a range of -1 to 1, which is an embedding.

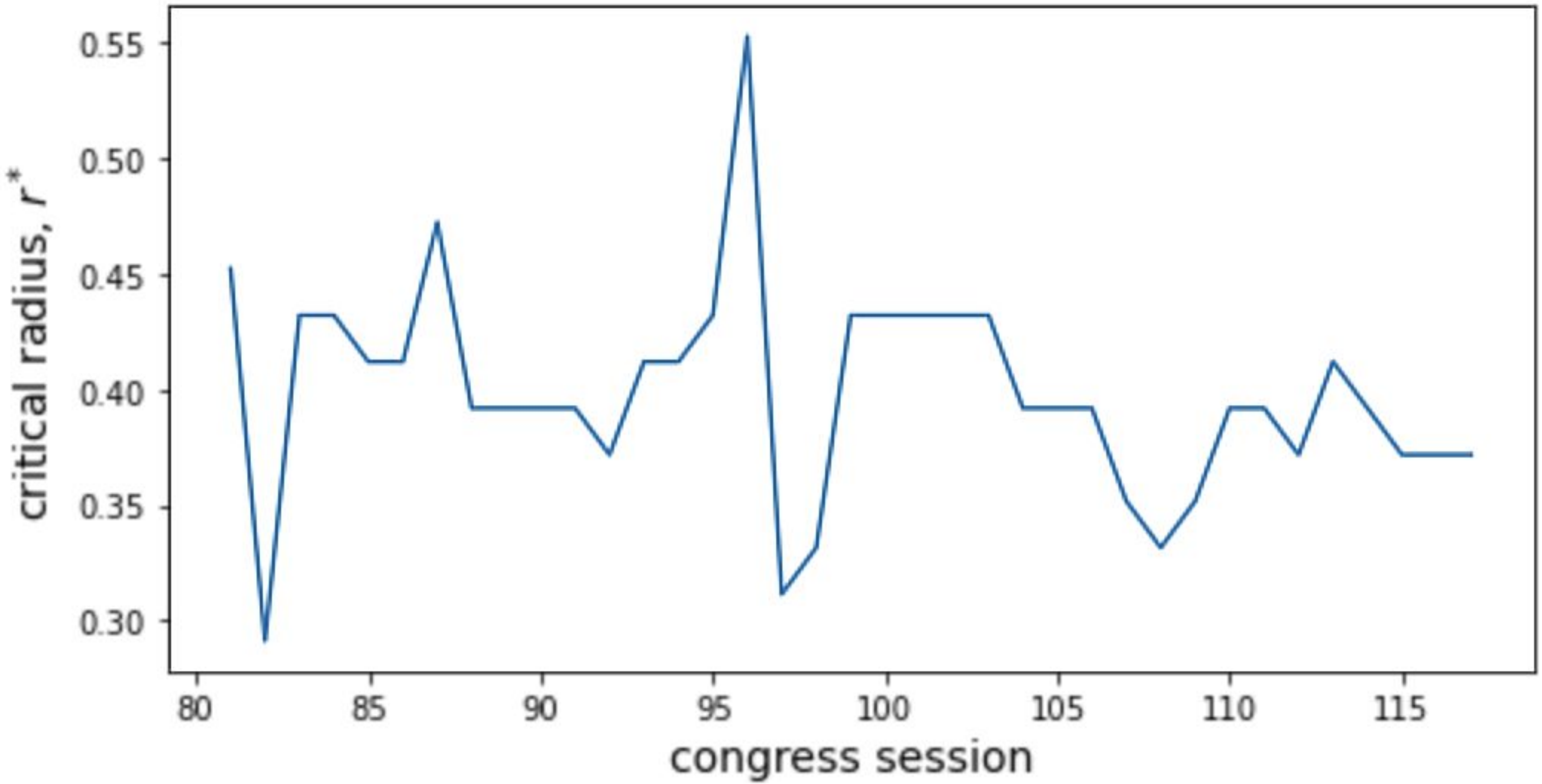
- To make the data more reliable, we removed the Vice President and the senators that voted consistently less that the other senators.
    - We assumed that the data was normally distributed and removed statistically insignificant senators.



Forming the Network

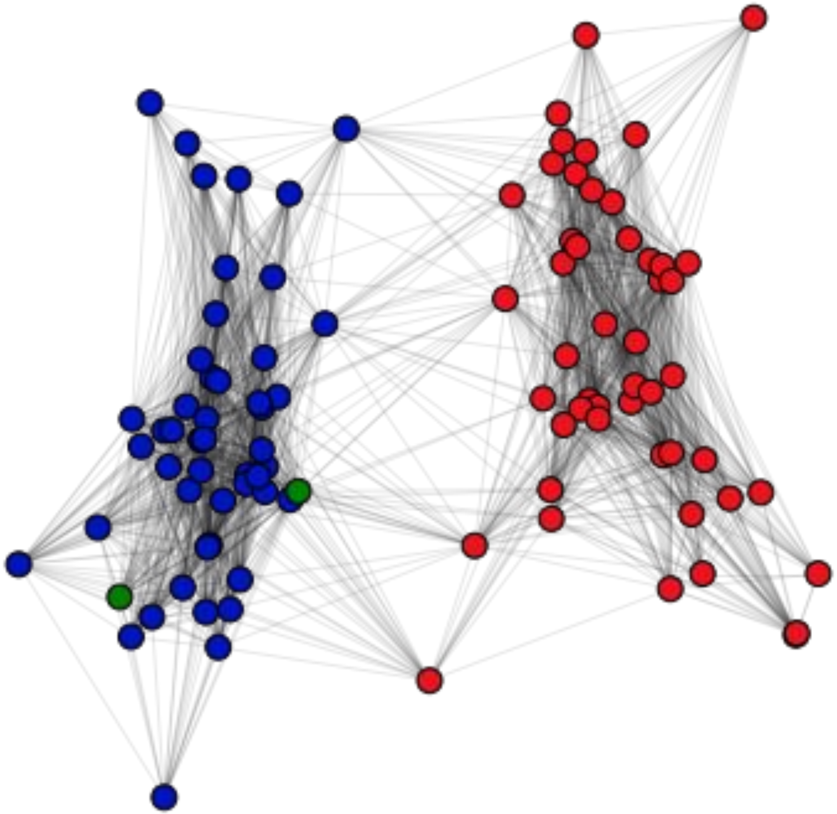
- We considered Senates from the 81st onwards because this is the start of the most modern political landscape.
- Using Euclidean distance, we computed each senator's ε-ball.
  - Using modularity maximization under the constraint of there being only one connected component, we find ε.
  - We took the maximum ε of all the values for each session to be the ε to use for every session.
- To form the adjacency matrix, two nodes (NOMINATE points) are connected if they lie in each other's ε-ball.





critical radius, r^*
0.55
0.50
0.45
0.40
0.35
0.30
80
85
90
95
100
105
110
115
congress session

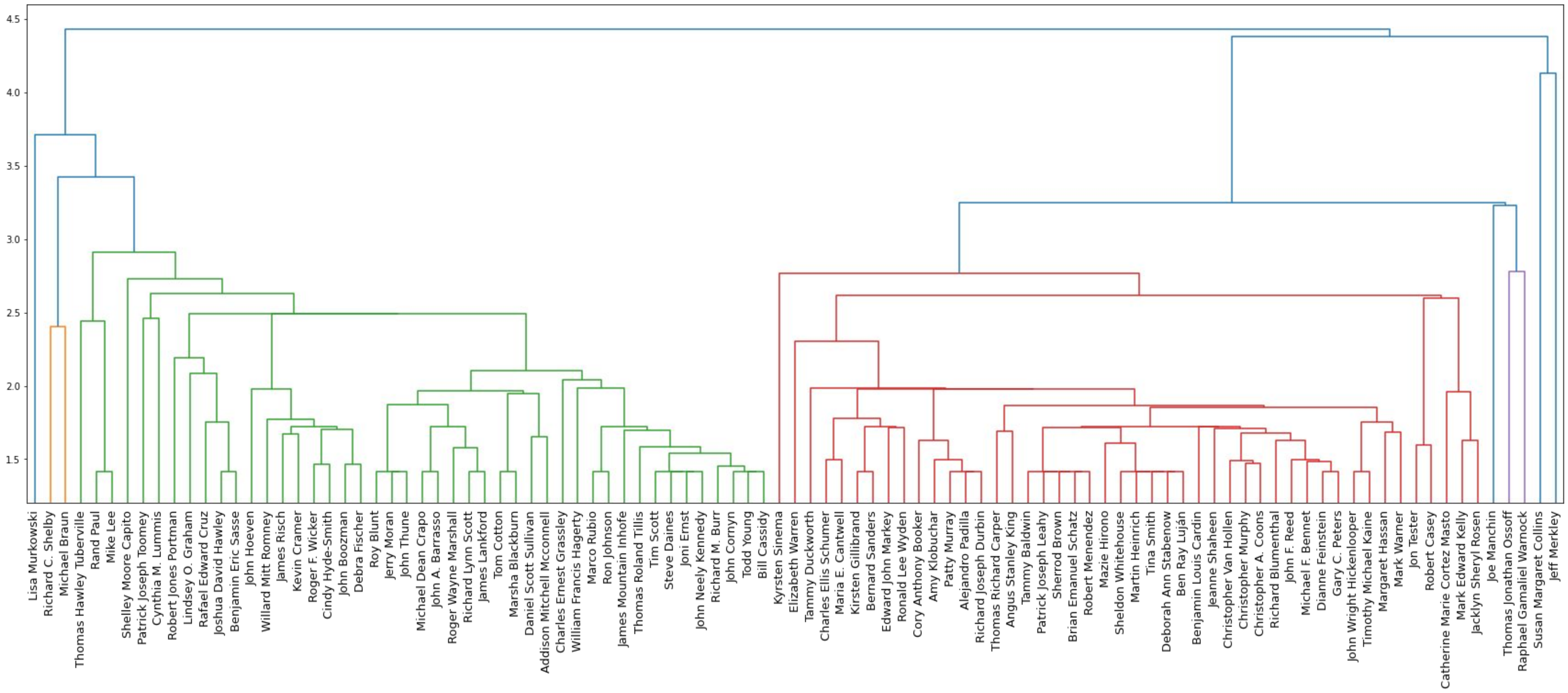
PART 3: Analysis of the Network



Dendrogram Algorithm

1. Begin by placing every senator at height 0.
2. Compute inverse edge betweenness of the different senators.
   a. Edge betweenness is like a similarity measure.
3. Find the closest senators by inverse edge betweenness.
   a. There can be multiple groups if (x, y, z) have distance d and (u, v, w) have distance d, but the distance between x and y is greater than d.
4. Connect the closest senators at their inverse betweenness value.
5. Combine the groups of senator into one pseudo-senator.
6. Take the new distance between this pseudo-senator as the minimum distance between the senators in this group and the other senators.
7. Repeat 2-6 until all senators are in one cluster.

- This is a “dendogram;” at whatever height you cut the dendrogram forms the communities.



4.5
4.0
3.5
3.0
2.5
2.0
1.5
Lisa Murkowski
Richard C. Shelby
Michael Braun
Thomas Hawley Tuberville
Rand Paul
Mike Lee
Shelley Moore Capito
Patrick Joseph Toomey
Cynthia M. Lummis
Robert Jones Portman
Lindsey O. Graham
Rafael Edward Cruz
Joshua David Hawley
Benjamin Eric Sasse
John Hoeven
Willard Mitt Romney
James Risch
Kevin Cramer
Roger F. Wicker
Cindy Hyde-Smith
John Boozman
Debra Fischer
Roy Blunt
Jerry Moran
John Thune
Michael Dean Crapo
John A. Barrasso
Roger Wayne Marshall
Richard Lynn Scott
James Lankford
Tom Cotton
Marsha Blackburn
Daniel Scott Sullivan
Addison Mitchell Mcconnell
Charles Ernest Grassley
William Francis Hagerty
Marco Rubio
Ron Johnson
James Mountain Inhofe
Thomas Roland Tillis
Tim Scott
Steve Daines
Joni Ernst
John Neely Kennedy
Richard M. Burr
John Cornyn
Todd Young
Bill Cassidy
Kyrsten Sinema
Elizabeth Warren
Tammy Duckworth
Charles Ellis Schumer
Maria E. Cantwell
Kirsten Gillibrand
Bernard Sanders
Edward John Markey
Ronald Lee Wyden
Cory Anthony Booker
Amy Klobuchar
Patty Murray
Alejandro Padilla
Richard Joseph Durbin
Thomas Richard Carper
Angus Stanley King
Tammy Baldwin
Patrick Joseph Leahy
Sherrod Brown
Brian Emanuel Schatz
Robert Menendez
Mazie Hirono
Sheldon Whitehouse
Martin Heinrich
Tina Smith
Deborah Ann Stabenow
Ben Ray Luján
Benjamin Louis Cardin
Jeanne Shaheen
Christopher Van Hollen
Christopher Murphy
Christopher A. Coons
Richard Blumenthal
John F. Reed
Michael F. Bennet
Dianne Feinstein
Gary C. Peters
John Wright Hickenlooper
Timothy Michael Kaine
Margaret Hassan
Mark Warner
Jon Tester
Robert Casey
Catherine Marie Cortez Masto
Mark Edward Kelly
Jacklyn Sheryl Rosen
Joe Manchin
Thomas Jonathan Ossoff
Raphael Gamaliel Warnock
Susan Margaret Collins
Jeff Merkley

Node betweenness

- We transformed the data linearly for more interpretability.
  - Betweenness is already within [0, 1], but we reversed order and shifted it so that 0 is most between and 1 is least between.

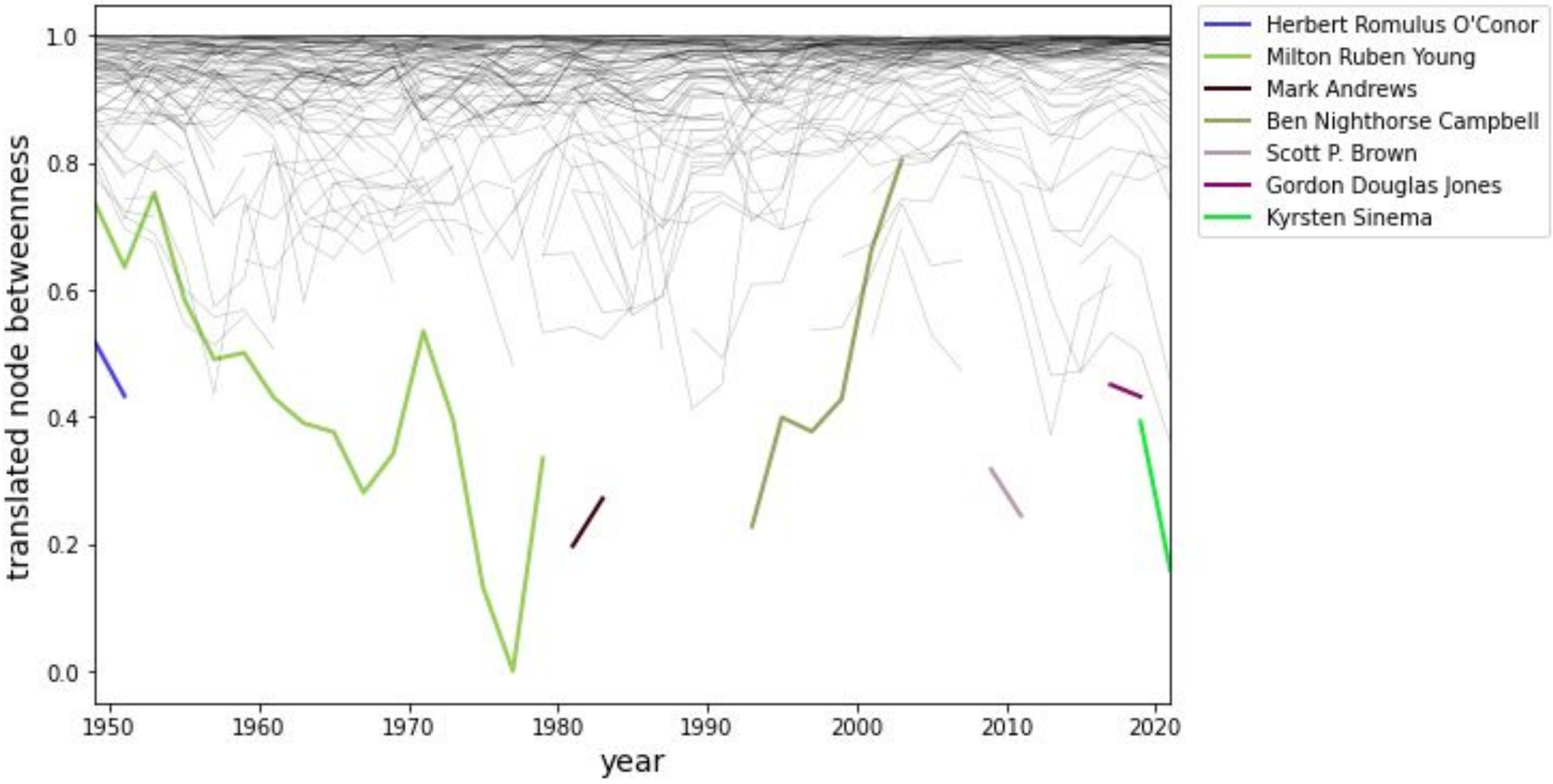
$$\vec{y} = 1 - \frac{\vec{x} - \min(\vec{x})}{\max(\vec{x}) - \min(\vec{x})}$$

Three most between senators

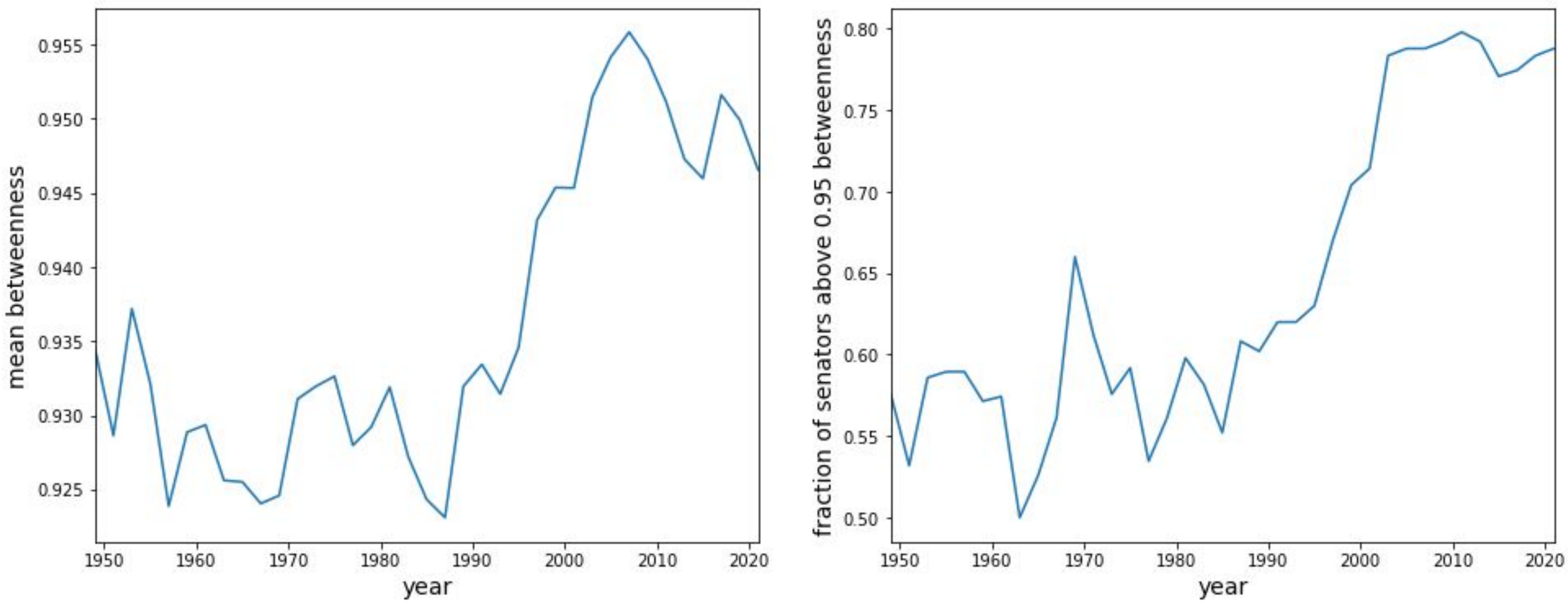
- Kyrsten Sinema (D, AZ): 0.00
- Lisa Murkowski (R, AK): 0.24
- Shelley Moore Capito (R, WV): 0.35

Three least between senators

- Thomas Jonathan Ossoff (D, GA): 1.00
- Jeff Merkley (D, OR): 1.00
- Michael Braun (R, IN): 1.00



translated node betweenness
year
0.0
0.2
0.4
0.6
0.8
1.0
1950
1960
1970
1980
1990
2000
2010
2020
Herbert Romulus O'Conor
Milton Ruben Young
Mark Andrews
Ben Nighthorse Campbell
Scott P. Brown
Gordon Douglas Jones
Kyrsten Sinema


mean betweenness
0.955
0.950
0.945
0.940
0.935
0.930
0.925
1950
1960
1970
1980
1990
2000
2010
2020
year
fraction of senators above 0.95 betweenness
0.80
0.75
0.70
0.65
0.60
0.55
0.50
1950
1960
1970
1980
1990
2000
2010
2020
year

PART 4: Conclusion & Future Explorations

Conclusion

- Congress is a polarized body.
    - Network science is one such tool which can represent that.
- Networks are representations of graphs, which encode the relationships (edges) between nodes (objects).
    - Centralities can be used to measure the importance of nodes or edges.
- Some network centralities are typically associated with community structure and hence can be used as a proxy for polarization.
    - Node and edge betweenness are often used in parallel with community detection.
- We used the NOMINATE dimensions from the VoteView database to embed senators.
    - The NOMINATE dimensions convert high-dimensional voting patterns into 2D coordinates.
- We examined node betweenness for different sessions of the Senate.
    - We found that the average value is increasing over time.
    - We found that the fraction of people with high betweenness increases over time.

Future aspects

- We can repeat these analyses for the House of Representatives.
  - There are 400+ representatives, so the data is more rounded.
  - Polarization should be occurring in both chambers.
- Creating a different embedding to represent the network.
  - PCA, diffusion maps, eigenmaps for different choices of data (like voting patterns, cosponsorship, etc.)
- Using different types of centralities for a different analysis.
  - PageRank, eigenvector, degree, closeness, etc.
- Implementing machine learning on the data to predict important metrics.
  - Modularity, party classification, etc.

Questions?