Introduction to Network Analysis

Soda 496

Why Networks?

- Network are one of the only ways we have to measure relationships.
  - People
  - Groups
  - Geographies
  - Things
  - Concepts

- Relationships are a critical part of almost all social behavior.
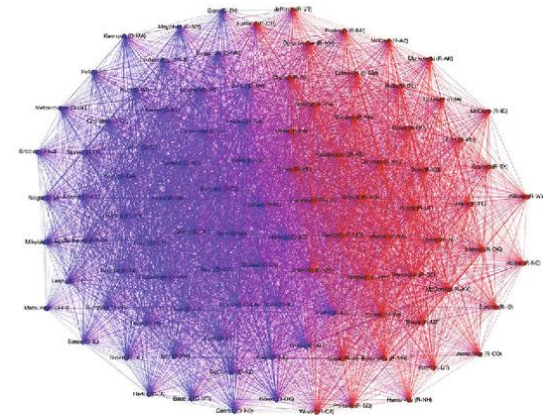
Why Networks?

- Network are one of the only ways we have to measure relationships.
  - People
  - Groups
  - Geographies
  - Things
  - Concepts

- Relationships are a critical part of almost all social behavior.
  - Politics

Senators casting the same votes
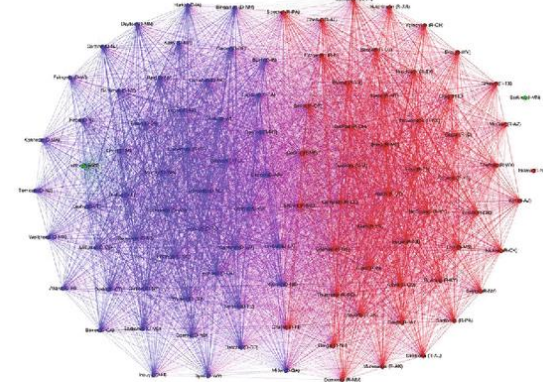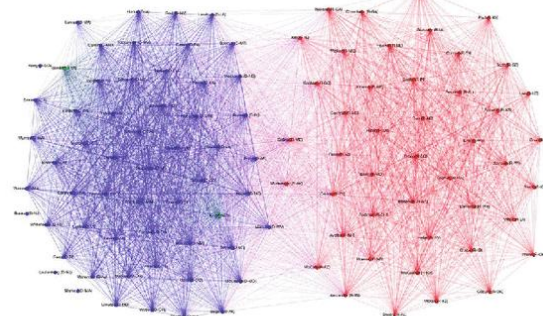
Democrat   Republican

101st Congress, 1989 session

107th Congress, 2002 session

113th Congress, 2013 session

Sources: GovTrack.us, Renzo Lucioni   ■ Independent
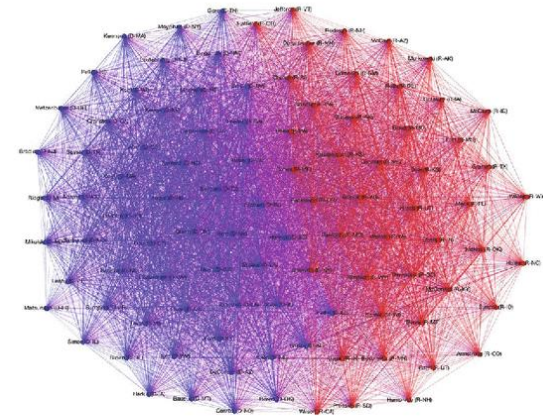
Why Networks?

- Network are one of the only ways we have to measure relationships.
  - People
  - Groups
  - Geographies
  - Things
  - Concepts

- Relationships are a critical part of almost all social behavior.
  - Politics
  - Sports
  - Infectious disease
  - Business and Trade
  - War
  - Information and beliefs
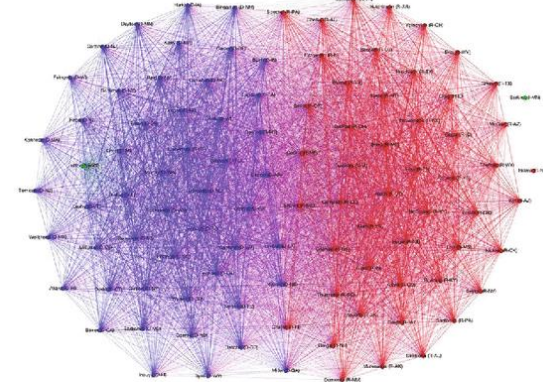

Senators casting the same votes
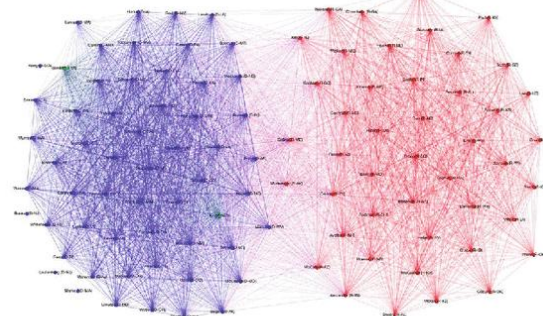
Democrat   Republican

101st Congress, 1989 session

107th Congress, 2002 session
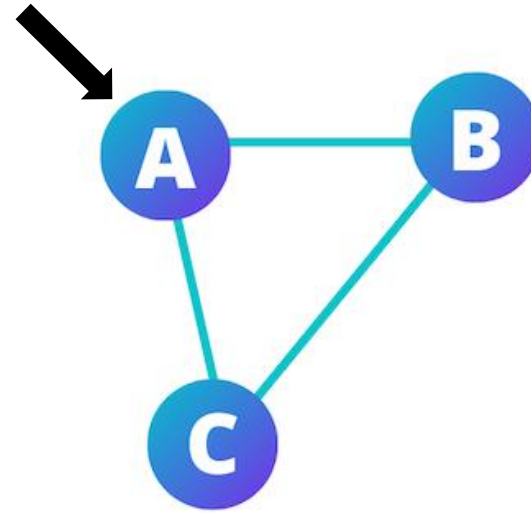
113th Congress, 2013 session

Sources: GovTrack.us, Renzo Lucioni   ■ Independent
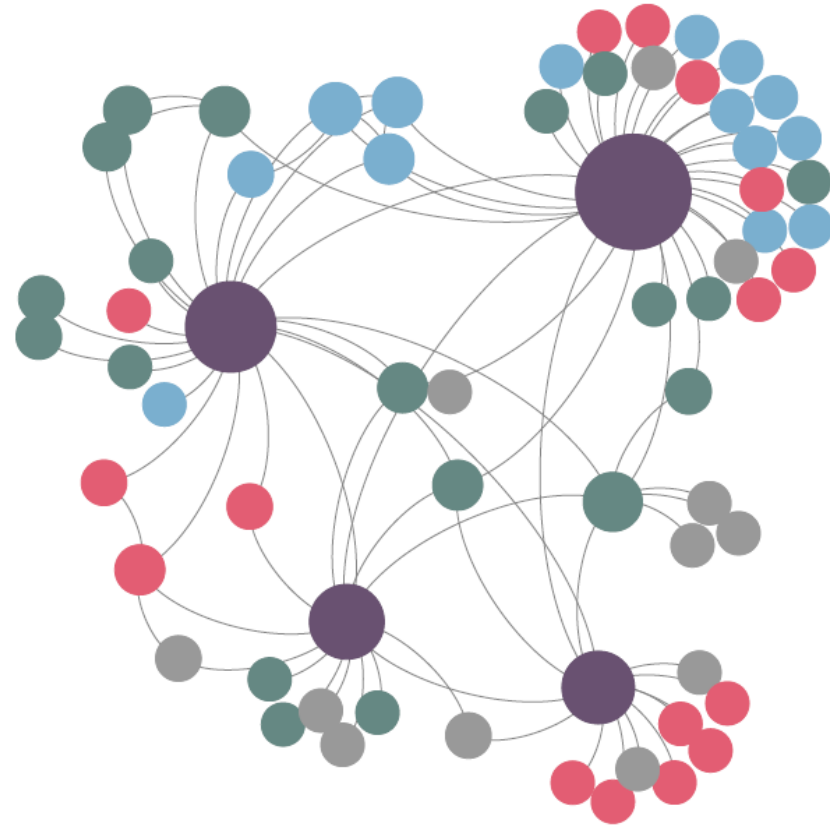
Basic Concepts

Nodes (vertices)

- Represent an entity in the network

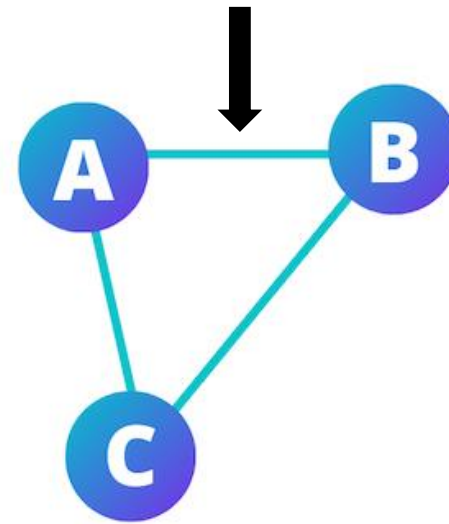- Can contain entity level features or meta data

Nodes (vertices)

- Represent an entity in the network

- Can contain entity level features or meta data

Edges (link, relation)

- Represent a relationship between nodes

- Characteristics:
  - Weight
  - Direction
  - Qualitative attributes

- One of the most common distinctions is between directed and undirected graphs

Edges (link, relation)

- Represent a relationship between nodes

- Characteristics:
  - Weight
  - Direction
  - Qualitative attributes

- One of the most common distinctions is between directed and undirected graphs
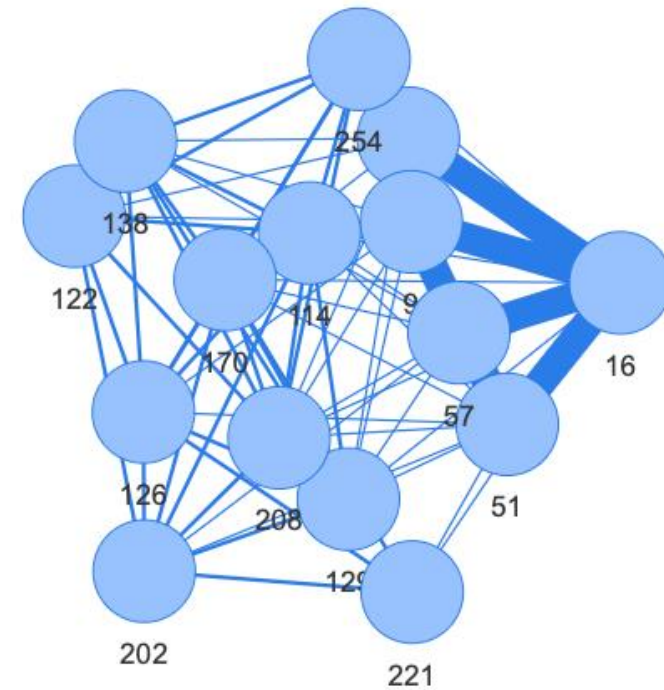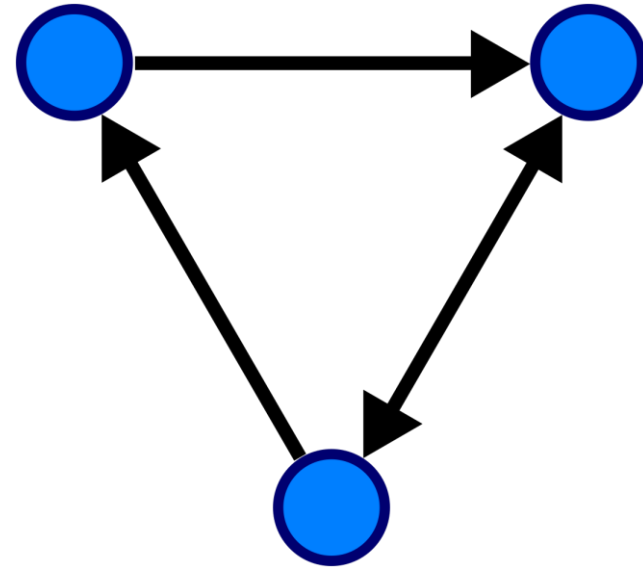
Edges (link, relation)

- Represent a relationship between nodes

- Characteristics:
  - Weight
  - Direction
  - Qualitative attributes

- One of the most common distinctions is between directed and undirected graphs

Data Management

- Networks are usually stored in one of three ways:
  - Edge list

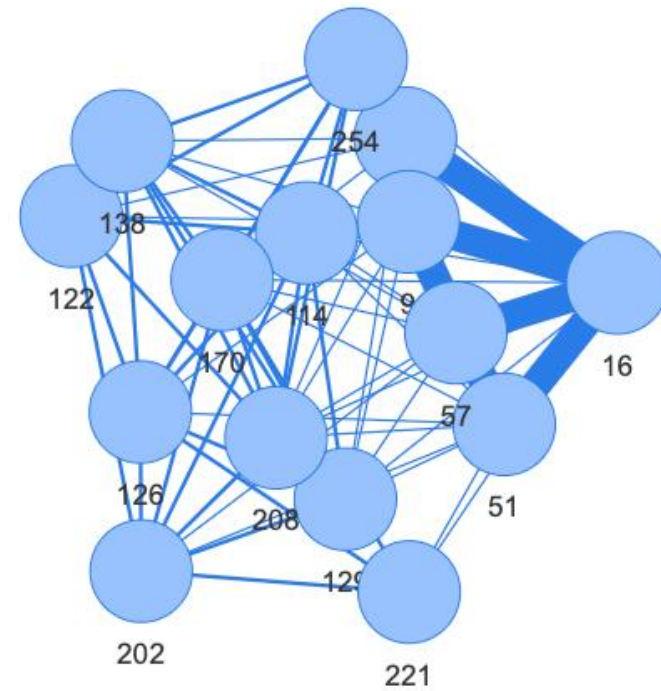| source	target
A	B
A	B
A	C
A	D
A	F
F	A
B	E

Data Management

- Networks are usually stored in one of three ways:
  - Edge list
  - Adjacency matrix

|	A	B	C	D	E
A	1	0	1	1	2
B	0	0	1	0	1
C	1	0	0	0	2
D	2	1	1	1	1
E	0	1	1	2	1

Data Management

- Networks are usually stored in one of three ways:
  - Edge list
  - Adjacency matrix
  - As a network object

Data Management

- Networks are usually stored in one of three ways:
  - Edge list
  - Adjacency matrix
  - As a network object

- Edge lists and adjacency matrices are most common and good for data sharing.
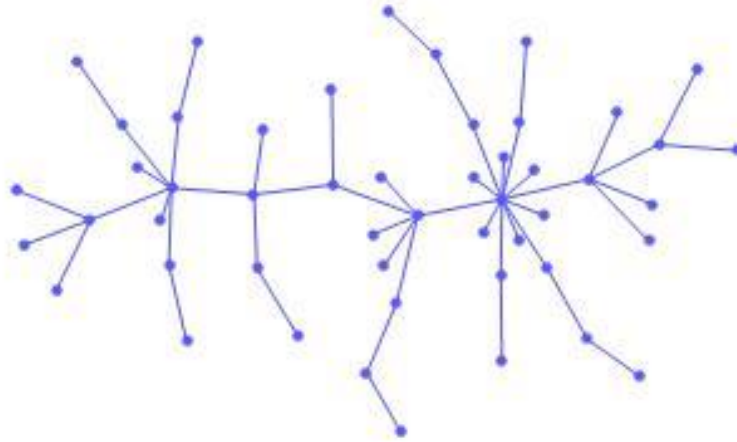 - Often accompanied by a second data set of node attributes.
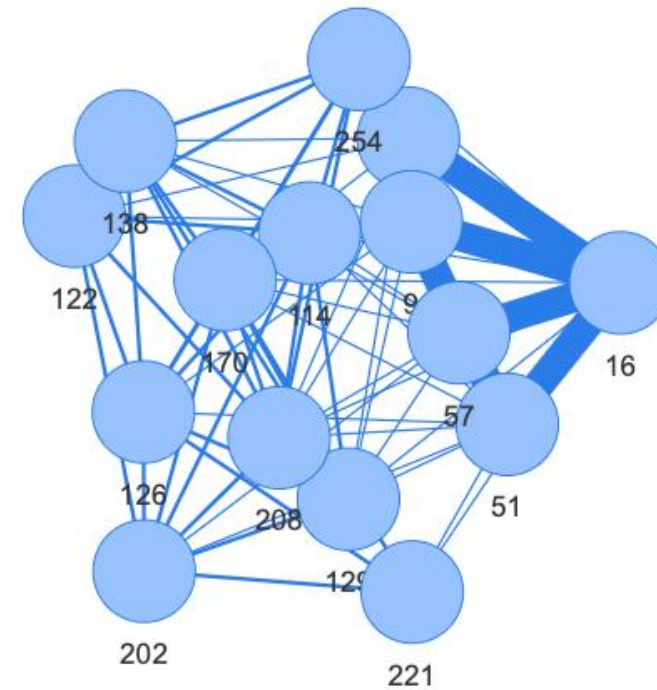
Description

Subgraphs

- Dyad: Pair of nodes

- Triad: Triple of nodes

- Subgroup: a subnetwork of any size

Popularity

- Degree: How many edges are attached to a node.

- Popularity is one measure of importance in a network.

- Preferential attachment: The more connected a node is the more likely it is to receive new links.
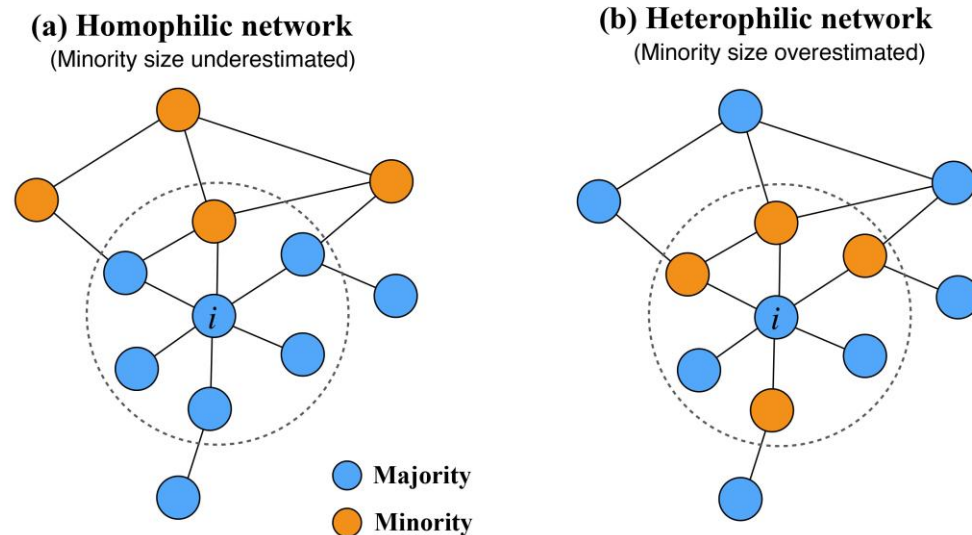
Centrality

- Measure of importance, distance, and time to spread

- Two types:
  - Radial: Walks that originate or terminate at a node.
  - Medial: Walk that pass through a node.

- Ways to measure:
  - Eigenvector Centrality: centrality is proportional to the centrality of its neighbors.
  - Betweenness centrality: Nodes that have a high number of "shortest paths" that pass through them are more central.
  - Closeness centrality: Sum of the distance between nodes.
 - Degree is also considered a measure of centrality
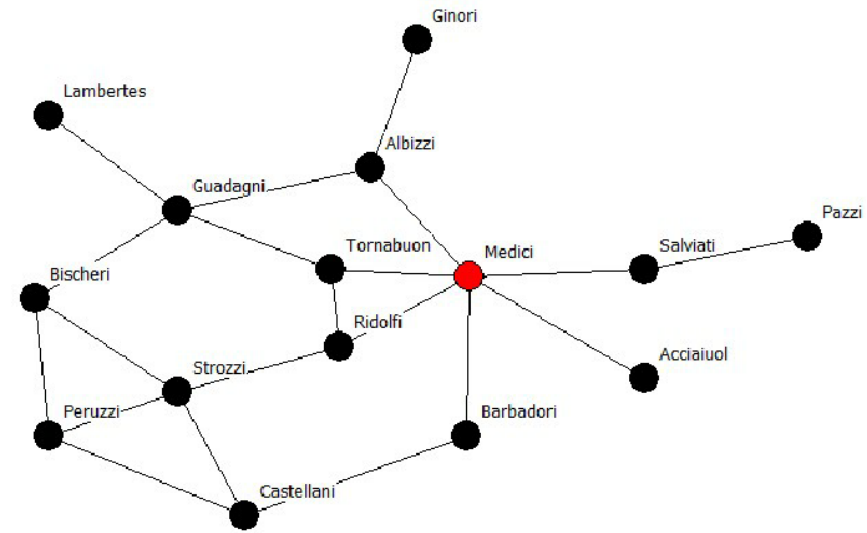
Mixing

- Homophily: The tendency for similar nodes to have connections.

- Assortative mixing: nodes associate with other like them

- Disassortative mixing: nodes associate with those who are different

- Almost all social networks are assortative. Biological networks tend to be disassortative.

(a) **Homophilic network**
(Minority size underestimated)

(b) **Heterophilic network**
(Minority size overestimated)

Majority
Minority

Transitivity

- AKA clustering coefficient. A measure of the tendency for nodes to group together.

- Density: the degree of connectedness between nodes. And "everywhere dense" network is one in which all nodes connect to each other.

- Clustering coefficient: Proportion of a nodes "neighbors" that are tied.

- Triads are and import part of measuring transitivity.
  - E.g. compare the number of triads to a null distribution of interest.
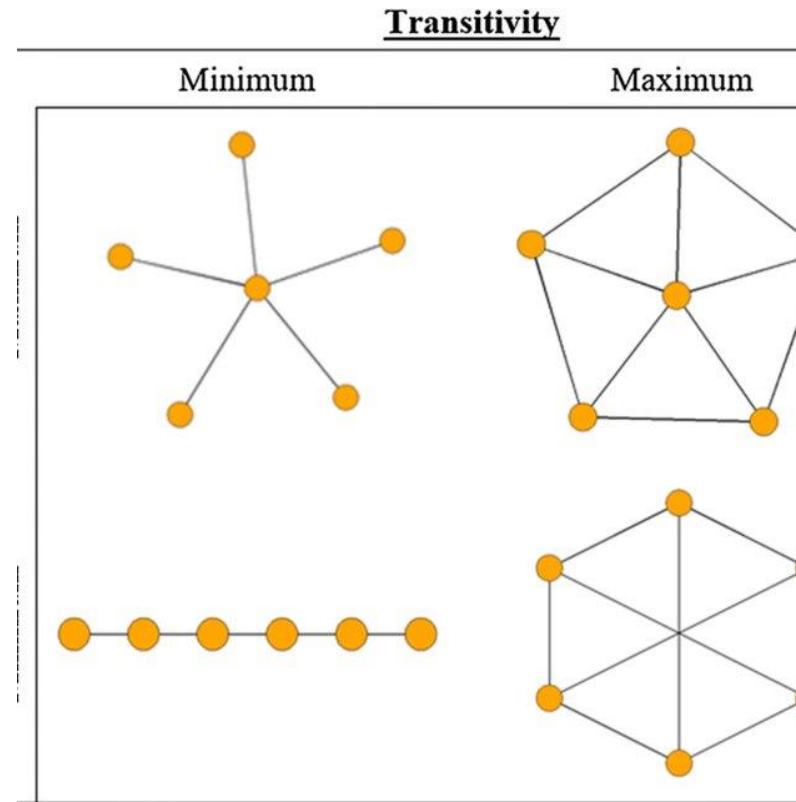
Reciprocity

- Number of bidirectional links/total number of links

- Can measure:
  - Hierarchy
  - Cohesion
  - Commitment
  - Retaliation