# Fundamental Statistics of Graphs: Centrality and Centralization

Mark S. Handcock

handcock@ucla.edu

Statistical Analysis of Networks
October 8, 2024

# Node-Level Indices (NLIs)

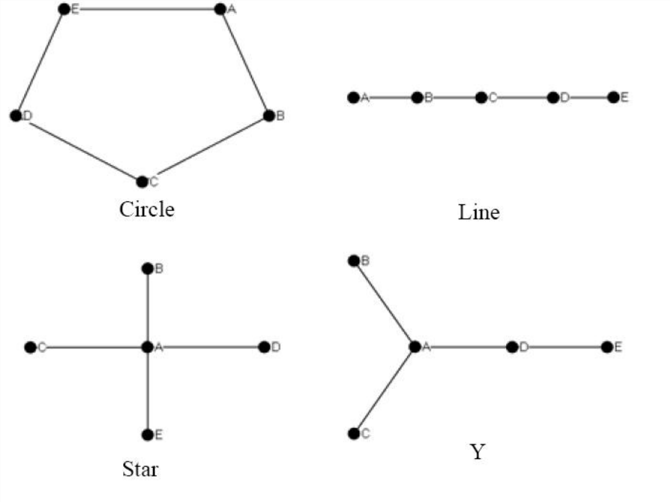
- Primary uses:
    - Quantify properties of individual positions
    - Describe local neighborhood
- Several common families:
    - Centrality
    - Ego-net structure
    - Alter covariate indices
- Centrality is the most prominent

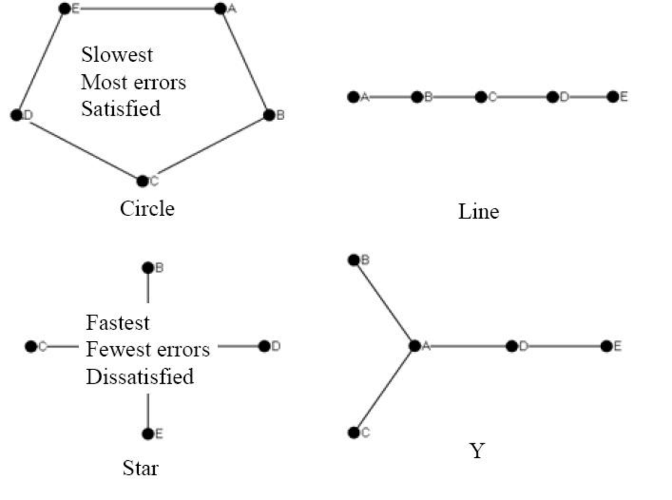
# Centrality

- Core question: how do individual positions vary?
- One manner in which positions vary is the extent to which they are "central" in the network
  - Important concern of social scientists (and junior high school students)
- Many distinct concepts
  - No one way to be central in a network – many different kinds of centrality
  - Different types of centrality aid/hinder different kinds of actions
  - Being highly central in one respect doesn't always mean being central in other respects (although the measures generally correlate)

# Early experimental studies of problem solving

- People in positions passed messages to one another to solve a problem
- Effect of structure on:
    - Efficiency
    - Leadership
    - Satisfaction
- Bavelas, Levitt, Guetskow, Freeman


E
A
D
B
C
Circle
A
B
C
D
E
Line
B
C
A
D
E
Star
B
A
D
E
C
Y


E
A
Slowest
Most errors
Satisfied
D
B
C
Circle
A
B
C
D
E
Line
B
Fastest
C
Fewest errors
D
Dissatisfied
E
Star
B
A
D
E
C
Y

# Types of Centrality Measures

One attempted classification by Koschutzki et al. (2005):

- Reach: Centrality based on ability of ego to reach other vertices
  - Ex: Degree, closeness
- Flow Mediation: Centrality based on quantity/weight of walks passing through ego
  - Ex: Stress, betweenness
- Vitality: Centrality based on effect of removing ego from the network
  - Ex: Flow betweenness (oddly), cutpoint status
- Feedback: Centrality of ego defined as a recursive function of alter centralities
  - Ex: Eigenvector centrality, Bonacich Power

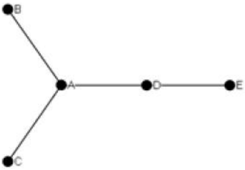
# Centrality and Centralization Measures

- Actor Centrality
  - Degree
  - Betweenness
  - Closeness
- Standardized measures compare to theoretical maximum – so values range from 0 to 1; allows for comparison across networks
- Network Centralization

# Degree

- Degree: number of direct ties
  - Overall activity or extent of involvement in relation
  - High degree positions are influential, but also may be subject to a great deal of influence or stress from others
- Formulas:
  - Degree (undirected): $C_D(n_i) = \sum_{j=1}^{g} x_{ij}$
  - Indegree: $C_I(n_i) = \sum_{j=1}^{g} x_{ji}$
  - Outdegree: $C_O(n_i) = \sum_{j=1}^{g} x_{ij}$

## Degree Centralities



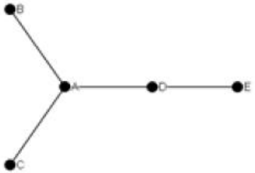

Y

### Adjacency Matrix

| | A | B | C | D | E | |
|---|---|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 | 0 | 3 |
| B | 1 | 0 | 0 | 0 | 0 | 1 |
| C | 1 | 0 | 0 | 0 | 0 | 1 |
| D | 1 | 0 | 0 | 0 | 1 | 2 |
| E | 0 | 0 | 0 | 1 | 0 | 1 |
| | 3 | 1 | 1 | 2 | 1 | |

# Standardized Degree Centrality

Divide by the maximum possible: $(g-1)$



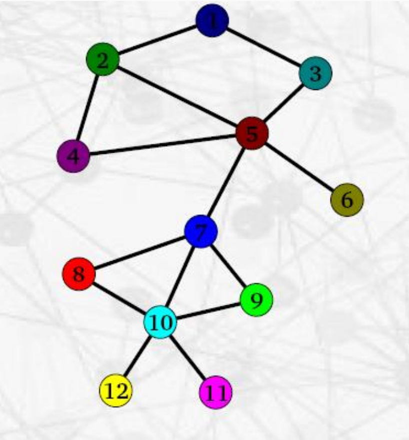

Y

| | A | B | C | D | E | |
|---|---|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 | 0 | ¾=.75 |
| B | 1 | 0 | 0 | 0 | 0 | ¼=.25 |
| C | 1 | 0 | 0 | 0 | 0 | ¼=.25 |
| D | 1 | 0 | 0 | 0 | 1 | 2/4=.5 |
| E | 0 | 0 | 0 | 1 | 0 | ¼=.25 |
| | ¾=.75 | ¼=.25 | ¼=.25 | 2/4=.5 | ¼=.25 | |

$$C_D'(n_i) = \frac{\sum_{j=1}^{g} x_{ij}}{(g-1)}$$

# Reminder: Shortest paths

- A shortest path from $i$ to $j$ is called an $i, j$ *geodesic*
    - Can have more than one (but all same length, obviously)
    - The length of an $i, j$ geodesic is called the geodesic distance from $i$ to $j$

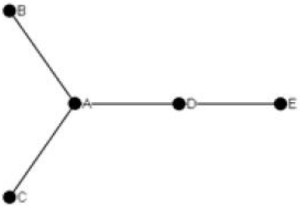



$g(1,2) = 1, g(2,3) = 2,$
$g(4,12) = 1$
$D(1,2) = 1, D(2,3) = 2,$
$D(4,12) = 4$

# Betweenness

- Betweenness: tendency of ego to reside on shortest paths between third parties
  - Quantifies extent to which position serves as a bridge
  - High betweenness positions are associated with “broker”or “gatekeeper” roles; may be able to “firewall” information flow
- Formula: $C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$
  - where $g_{jk}$ is the number of $j, k$ geodesics, $g_{jk}(n_i)$ is the number of $j, k$ geodesics including $i$

# Betweenness




Betweenness

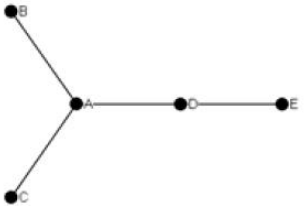
------------

| | |
|---|---|
| A | 5.000 |
| B | 0.000 |
| C | 0.000 |
| D | 3.000 |
| E | 0.000 |

| | | Geodesic |
|---|---|---|
| A | B | A-B |
| A | C | A-C |
| A | D | A-D |
| A | E | A-D-E |
| B | C | B-A-C |
| B | D | B-A-D |
| B | E | B-A-D-E |
| C | D | C-A-D |
| C | E | C-A-D-E |

# Standardized Betweenness Centrality

Divide by the maximum possible: $\frac{(g-1)(g-2)}{2}$ - the number of all pairs of nodes not including $n_i$




Y

| | | Geodesic |
|---|---|---|
| A | B | A-B |
| A | C | A-C |
| A | D | A-D |
| A | E | A-D-E |
| B | C | B-A-C |
| B | D | B-A-D |
| B | E | B-A-D-E |
| C | D | C-A-D |
| C | E | C-A-D-E |

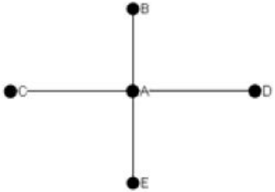
| | Betweenness | Std. Betweenness |
|---|---|---|
| | ------------ | ------------ |
| 1 | 5.000 | 83.333 |
| 2 | 0.000 | 0.000 |
| 3 | 0.000 | 0.000 |
| 4 | 3.000 | 50.000 |
| 5 | 0.000 | 0.000 |

$$\frac{5}{\frac{(g-1)(g-2)}{2}} = \frac{5}{6} = .833$$

# Closeness

- Closeness: how close is a node to other nodes, measured by geodesic distances
    - Extent to which position has short paths to other positions
    - High closeness positions can quickly distribute information (or spread disease), but may have limited direct influence
    - Limitation: not useful on disconnected graphs
- Formula: $C_C(n_i) = \sum_{j=1}^{g} \frac{1}{d(n_i, n_j)}$
    - where $d(n_i, n_j)$ is the geodesic distance from $i$ to $j$

# Closeness




$$C_C(n_i) = \frac{1}{\sum_{j=1}^{g} d(n_i, n_j)}$$

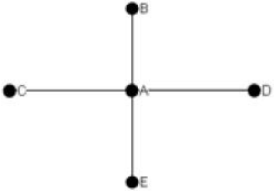
Geodesic Distances

reciprocal

| | A | B | C | D | E | sum | |
|---|---|---|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 | 1 | 4 | 1/4 |
| B | 1 | 0 | 2 | 2 | 2 | 7 | 1/7 |
| C | 1 | 2 | 0 | 2 | 2 | 7 | 1/7 |
| D | 1 | 2 | 2 | 0 | 2 | 7 | 1/7 |
| E | 1 | 2 | 2 | 2 | 0 | 7 | 1/7 |

# Standardized Closeness Centrality

Divide by the maximum possible: $\frac{1}{(g-1)}$ - the number of all pairs of nodes not including $n_i$

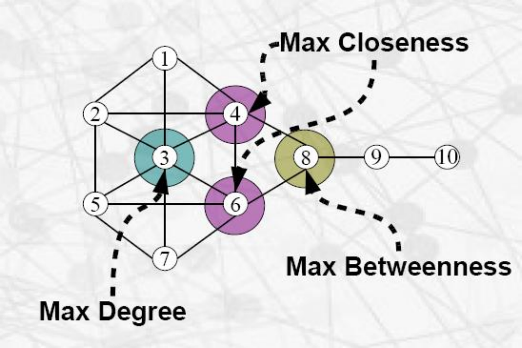



$$C'_C(n_i) = \frac{(g-1)}{\sum_{j=1}^{g} d(n_i, n_j)}$$

Geodesic Distances reciprocal standardized

| | A | B | C | D | E | sum | | |
|---|---|---|---|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 | 1 | 4 | 1/4 | 4/4 = 1 |
| B | 1 | 0 | 2 | 2 | 2 | 7 | 1/7 | 4/7 =.57 |
| C | 1 | 2 | 0 | 2 | 2 | 7 | 1/7 | 4/7 =.57 |
| D | 1 | 2 | 2 | 0 | 2 | 7 | 1/7 | 4/7 =.57 |
| E | 1 | 2 | 2 | 2 | 0 | 7 | 1/7 | 4/7 =.57 |

# Classic Centrality Measures Compared




Top 3 by Degree

1. Node 3
2. Nodes 4 and 6
3. Nodes 2 and 5

Top 3 by Closeness

1. Nodes 4 and 6
2. Nodes 3 and 8
3. Nodes 2 and 5

Top 3 by Betweenness

1. Node 8
2. Nodes 4 and 6
3. Node 9

# Recursive Centrality

- Eigenvector centrality (Bonacich, 1972): the centrality of each vertex is equal to the sum of the centralities of its neighbors, attenuated by a scaling constant ($\lambda$)
  - Formula: $C_E(n_i) = \frac{1}{\lambda} \sum_{j=1}^{g} x_{ij} C_E(n_j)$
  - Central vertices are those with many central neighbors
  - A variant of eigenvector centrality is employed by Google to rank Web pages

# Google Describing PageRank

*PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."*

# Social process interpretation of Eigenvector centrality

- Let $x^l = x^{l-1}x, \quad l = 2, \ldots$ Then $x^l_{ij}$ can be interpreted as the number of walks of length $l$ from vertex $i$ to $j$.
- Formula: $C_E(n_i) = \sum_{l=1}^{\infty} \frac{1}{\lambda^l} \sum_{j=1}^{g} x^l_{ij}$
- Walks are weighted inversely by their length.
- Also a core-periphery measure: high core have little interaction with the periphery.
- Also “best” one-dimensional approximation to structure of the sociomatrix.

# Recursive Centrality

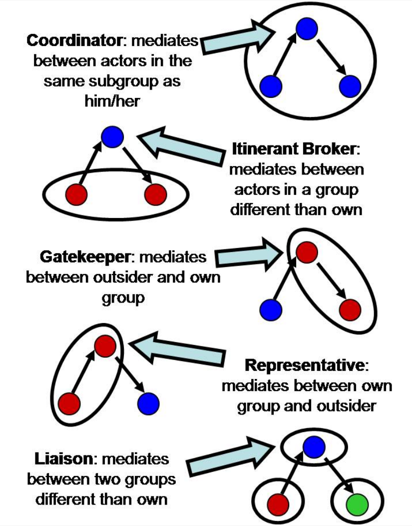
- Bonacich power centrality (beta-centrality) (Bonacich, 1987): the power of a vertex depends on the sum of the power of its alters
    - Formula: $C_{BP}(\alpha, \beta) = \sum_{j=1}^{g} (\alpha + \beta c_j) x_{ij}$
        - $\beta > 0$ implies that vertices become more powerful as their alters become more powerful (as occurs in cooperative relations)
        - $\beta < 0$ implies that vertices become more powerful only as their alters become weaker (as occurs in competitive or antagonistic relations)

# Recursive Centrality

- Flow betweenness (Freeman et al, 1991): a variant of betweenness that takes into account all paths between 2 nodes, not just the geodesics
  - Can be calculated both for valued graphs, and for simple graphs
  - Formula for simple graphs: $C_F(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$
    - where $g_{jk}$ is the number of $j, k$ paths, $g_{jk}(n_i)$ is the number of $j, k$ paths including $i$
  - Formula for valued graphs: $C_F(n_i) = \sum_{j<k} \frac{m_{jk}(n_i)}{m_{jk}}$
    - where $m_{jk}$ is the maximum flow between nodes $j, k$ and $m_{jk}(n_i)$ is the maximum flow between $j, k$ along paths including $i$

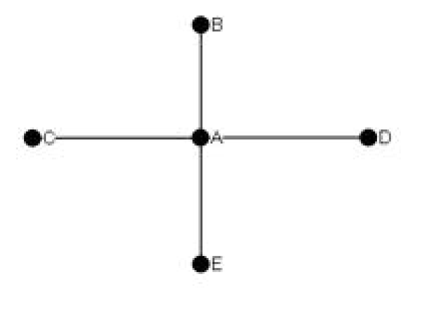
# Brokerage

- Brokers are actors in a network who mediate the flow of information or resources between third parties who are not directly tied
- Actors may belong to distinct subgroups, leading to 5 possible brokerage roles (4 for undirected networks, where the gatekeeper and representative roles are the same)



# Network Centralization

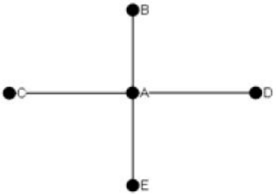
- Network level measure
- A single actor or a few actors are more central than others
- Related to variability in actor centralities
- The star is always the most centralized graph, and thus the reference point



# Differences in Centralities in Observed Network

| | |
|---|---|
| $C_A(n_i)$ | Actor Centrality |
| $C_A(n_i^*) = \max_i C_A(n_i)$ | Maximum Actor Centrality Observed in the network |
| $\sum_{j=1}^{g} [C_A(n_i^*) - C_A(n_j)]$ | Total Differences between the maximum actor centrality and other actor centralities observed in the network |

# Differences in Centralities in Star Network (Theoretical Maximum)




$$\max \sum_{j=1}^{g} [C_A(n_i^*) - C_A(n_j)]$$

Total Differences between most central actor in the star and the other actor centralities

# Network Centralization Measure

$$C_A = \frac{\sum_{j=1}^{g} [C_A(n_i^*) - C_A(n_j)]}{\max[\sum_{j=1}^{g} C_A(n_i^*) - C_A(n_j)]}$$

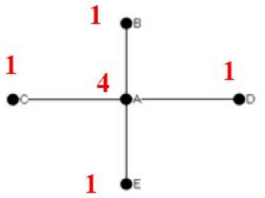
Total differences between actor centralities and the maximum actor centrality <u>observed</u> in the network

Theoretical maximum possible

# Degree Centralization Measure

- Denominator: Find theoretical maximum possible differences in centralities
- Numerator: Find differences in centralities in the observed network

$$C_D(n_i^*) = (g-1)$$




$$\max \sum_{j=1}^{g} [C_D(n_i^*) - C_D(n_j)] = (g-1) \times [(g-1)-1]$$
$$= (g-1)(g-2)$$

# Network Degree Centralization Measure

$$C_D = \frac{\sum_{j=1}^{g}[C_D(n_i^*) - C_D(n_j)]}{(g-1)(g-2)}$$

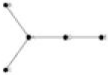
# Network Degree Centralization Measure

Circle

$$\frac{0}{12} = 0$$

| 1 Degree | 2 NrmDegree |
|---|---|
| 2 | 50 |
| 2 | 50 |
| 2 | 50 |
| 2 | 50 |
| 2 | 50 |

Y



$$\frac{7}{12} = .583$$

| 1 Degree | 2 NrmDegree |
|---|---|
| 3 | 75 |
| 1 | 25 |
| 1 | 25 |
| 2 | 50 |
| 1 | 25 |

Line

$$\frac{2}{12} = .167$$

| 1 Degree | 2 NrmDegree |
|---|---|
| 1 | 25 |
| 2 | 50 |
| 2 | 50 |
| 2 | 50 |
| 1 | 25 |

Star

$$\frac{12}{12} = 1.0$$

| 1 Degree | 2 NrmDegree |
|---|---|
| 4 | 100 |
| 1 | 25 |
| 1 | 25 |
| 1 | 25 |
| 1 | 25 |