Working with Non-Symmetric Relations

Steve Borgatti

Unexpected Asymmetry

- Monica claims to have "relations" with Bill, but Bill does not claim to have relations with Monica
  - The relation is logically symmetric, but empirically asymmetric
  - errors of recall; strategic response
- Can measure (and model) the degree of asymmetry
  - Reciprocity and symmetry indices
- Logically symmetric data may be symmetrized
  - if either A or B mentions the other, it's a tie
  - Only if both mention the other is it a tie

Measuring symmetry

- Index
  - How often the value of x_{ij} is the same as x_{ji}
  - T = number of unordered pairs (i,j) in which $x_{ij} = x_{ji}$
  - P = number of unordered pairs = n(n-1)/2
  - Symmetry = T/P
- Equivalently, we are asking whether X = X'
  - Test this via QAP correlation

Reciprocity

- How often a tie is reciprocated
- Measure: $$\frac{|iRj \text{ AND } jRi|}{|iRj \text{ OR } jRi|}$$ |X| indicates a count of the number of times X occurs, across all pairs i,j
  - How often i and j nominate each other as a proportion of the number of times at least one nominates the other
- Can be calculated separately for each node – what proportion of node's outgoing ties are reciprocated?

Degree Centrality

- Concept
  - Number of ties a node has
- Directed case
  - Indegree: colums sums of adjacency matrix
  - Outdegree: row sums
- Scatter plot:

Indegree ↑		
	Authority	High involvement
	Low involvement	Apprentice
	Outdegree →	

	Mary	Bill	John	Larry	Out
Mary	0	1	1	1	3
Bill	1	0	1	0	2
John	0	0	0	1	1
Larry	0	0	0	0	0
In degree	1	1	2	2	6

Closeness Centrality

- Concept
  – Distance from/to all other nodes
- Directed
  – Row and column sums of the distance matrix
- Problems
  – Directed graphs usually not connected. Many distances undefined
- Alternative
  – Sum reciprocals the distance matrix instead. Substitute zeros whenever a distance is undefined
  – Or count number of nodes reached

Betweenness

- Concept
  - How often a node lies along a geodesic path between two others
- Directed graphs
  - No adjustment needed

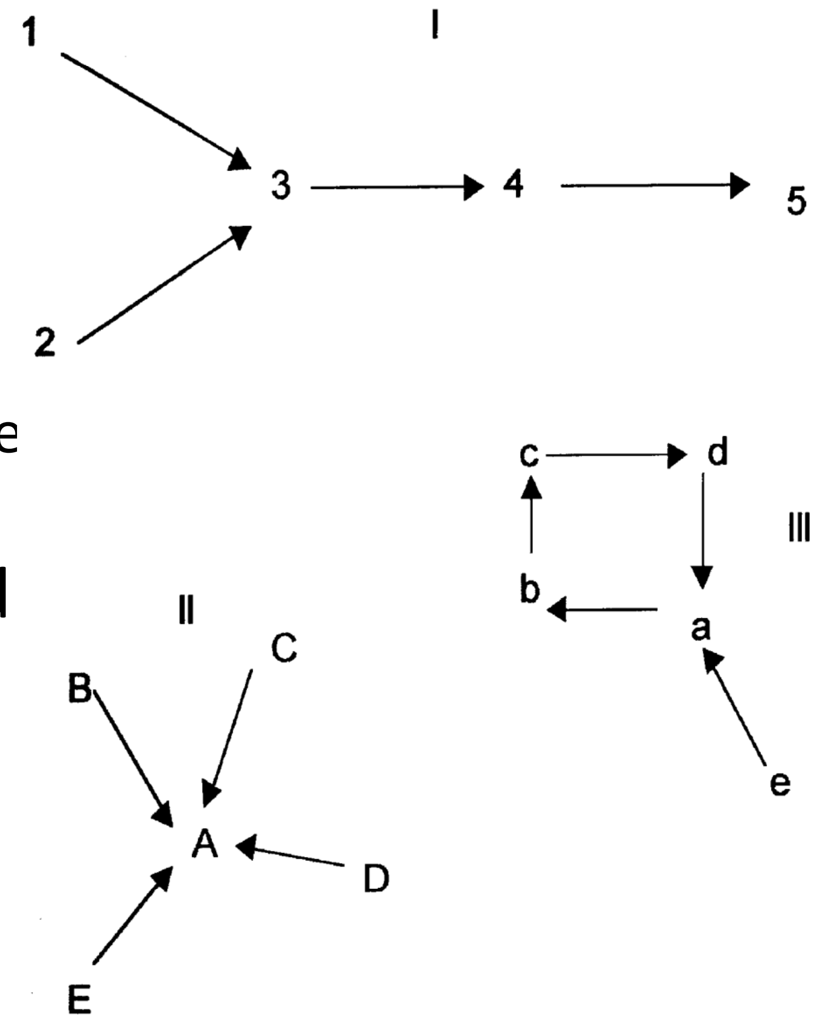
$$b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}}$$

Eigenvector

- Concept
  - A person is central to the extent they are connected to many people who are well connected (to people who are well … etc)
- Directed graphs
  - (columns) A person has high status to the extent that they are nominated by many people who are themselves frequently nominated
    - Left eigenvector $\mathbf{x}'A = \lambda\mathbf{x}$ or $A'\mathbf{x} = \lambda\mathbf{x}$
  - (rows) A person has influence to the extent they influence many who themselves influence many
    - Right eigenvector $A\mathbf{x} = \lambda\mathbf{x}$

Eigenvector for Directed graphs

- Often not calculable
- Can give useless answers
  - Nets I and II give all zeros for all nodes
    - Nodes with 0 indegree have no status to pass along …
  - In net III, nodes a, b, c and d have same score, even though a has greater indegree



Figures from Bonacich and LLoyd

Alpha Centrality

- Same as eigenvector when applied to symmetric matrices, but better results when applied to non-symmetric matrices
- Basically same as measures by Katz and Hubbell
  - Right alpha centrality: $\boldsymbol{x} = \alpha A \boldsymbol{x} + \boldsymbol{e} = (I - \alpha A)^{-1} \boldsymbol{e}$
    - Assume **e** is vector of 1s
  - left alpha centrality: $\boldsymbol{x} = \alpha A^T \boldsymbol{x} + \boldsymbol{e} = (I - \alpha A^T)^{-1} \boldsymbol{e}$
- In left (right) alpha centrality …
  - If α is positive then a person gets a high score for receiving ties from (sending ties to) people with high scores
  - If α is negative, then a person gets a high score for receiving ties from (sending ties to) people with low scores

Katz Influence

- If i does not have a tie to j, i can still influence j by influencing someone who influences someone … who influences j.
  - more chains from I to j, the more certain the influence,
  - but also the longer the chains the weaker the influence
- Given adjacency matrix R, the number of chains of length k is given by R^k , so we need a sum like this: $\mathbf{R}^1 + \mathbf{R}^2 + \mathbf{R}^3 + ...$ except we want to weight the longer chains less
- A parameter α^k (smaller than 1) can be introduced which goes to zero as k approaches infinity
  - $\mathbf{Q} = \alpha^1\mathbf{R}^1 + \alpha^2\mathbf{R}^2 + \alpha^3\mathbf{R}^3 + ... \alpha^\infty\mathbf{R}^\infty$
  - The row sums of Q give the total influence of a node on the network

– It turns out that when $\alpha < 1/\lambda_1$ where λ_1 is the largest eigenvalue of R, this series converges to $\mathbf{Q} = (\mathbf{I}-\alpha\mathbf{R})^{-1} - 1$, which leads to a row sum that is just 1 less than alpha centrality

Singular Value Decomposition (SVD)

- Every matrix A can be decomposed as follows:

$$A_{n\times m} = U_{n\times m} D_{m\times m} V^T_{m\times m}$$

D is a diagonal matrix of singular values

- We can approximate A with lower dimensionality k << m

$$A_{n\times m} = U_{n\times k} D_{k\times k} V^T_{m\times k}$$

- A 1-dimensional solution:

$$A = u\lambda^{1/2}v'$$

- The u-scores and column scores can be written in terms of each other

$$u_i = \lambda^{-1/2}\sum_j a_{ij}v_j$$

$$v_j = \lambda^{-1/2}\sum_i a_{ij}u_i$$

Hubs and Authorities

- Run an SVD on an adjacency matrix A, and retain only the first dimension $A = u\lambda^{1/2}v'$
- The u and v scores measure the extent to which a node is playing the role of a hub or authority respectively
  - The u-score (hub) measures the extent to which the node sends ties to nodes that have high v-scores (are authorities)
  - The v-score (authority) measures the extent to which the node receives ties from nodes with high u-scores (are hubs)