Social Network Analysis

Master Sociology 2014-15
University of Amsterdam

Programme for today

Session 1: Monday, Jan 12, REC-H 2.01 (computer lab)

- Introduction to graph theory
- Cohesion

Session 2: Wednesday, Jan 14, REC-H 2.01 (computer lab)

- Cohesion
- Centrality

Session 3: Thursday, Jan 15, REC-G S.09

- QAP correlations and regression

Session 4: Friday, Jan 16, 13-15, REC-G S.14

- Test 2

What is cohesion?

- Generally: How strongly connected a network is
- Why this that relevant? What happens in cohesive clusters?
- Cohesive clusters
  - Enforce behavioral norms
  - Share and make sense of information
  - Develop an identity, ingroup-outgroup
  - Can coordinate for collective action
  - Share a common fate

Why care about cohesion?

- Why do some communities, organizations, families perform certain functions better than others?
- How does a community's, family's, organization's internal structure affect its functions?
- Examples:
  - Charlestown resisting the urban renewal plan
  - Failure to establish a work council
  - Performance of R & D teams

Investigating cohesion: A family of measures

- Distance
- Density
- Reciprocity
- Transitivity
- Clustering
- (Non-)Fragmentation

Reciprocity

Reciprocity

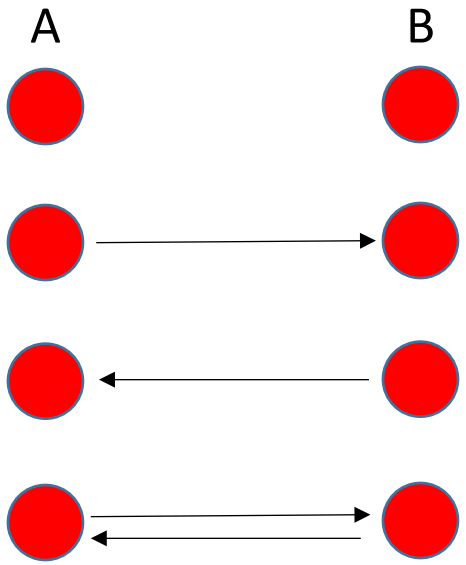
- Why is reciprocity sociologically interesting?

Reciprocity

- Why is reciprocity sociologically interesting?
  - One of the most universal social norms (Gouldner)
  - Indicator for stability (less reciprocity, less stability)
  - Indicator for hierarchy and equality

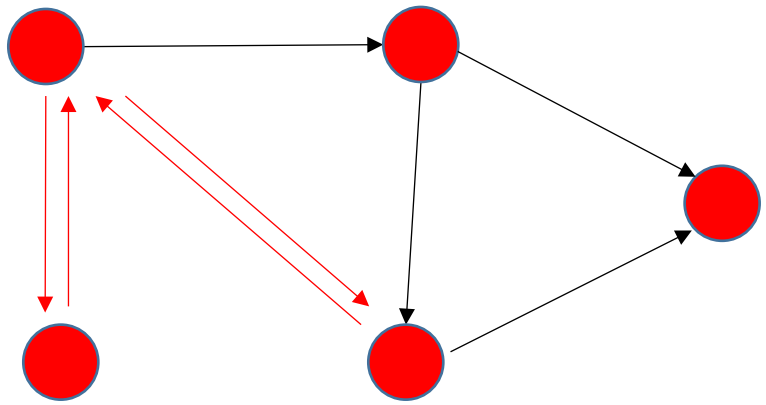
Reciprocity

- Makes sense only in directed graphs
- There are four dyadic possibilities:
  No tie, A to B, B to A, Both



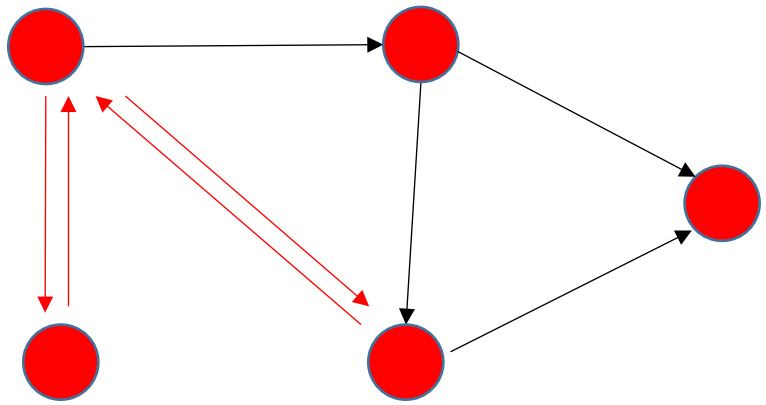
Reciprocity

- Dyad method (focus on pairs of nodes): The number of pairs with reciprocated ties divided by the number of pairs with any tie
- Arc method (focus on ties): Number of ties involved in a reciprocal relation divided by the number of ties



Reciprocity

- Dyad method (focus on pairs of nodes): The number of pairs with reciprocated ties divided by the number of pairs with any tie
- Arc method (focus on ties): Number of ties involved in a reciprocal relation divided by the number of ties



Dyad method: 2/6 = 0.33
Arc method: 4/8 = 0.50

Reciprocity in UCInet

Transitivity

Transitivity

- Remember why transitivity is sociologically interesting?

Transitivity

- Remember why transitivity is sociologically interesting?
  - Principle that leads to tie formation and clustering
  - Triad is the smallest “society”
  - Triads can show the emergence of consistency and balance – institutionalization of social structures
  - Triads may also show the emergence of a hierarchy

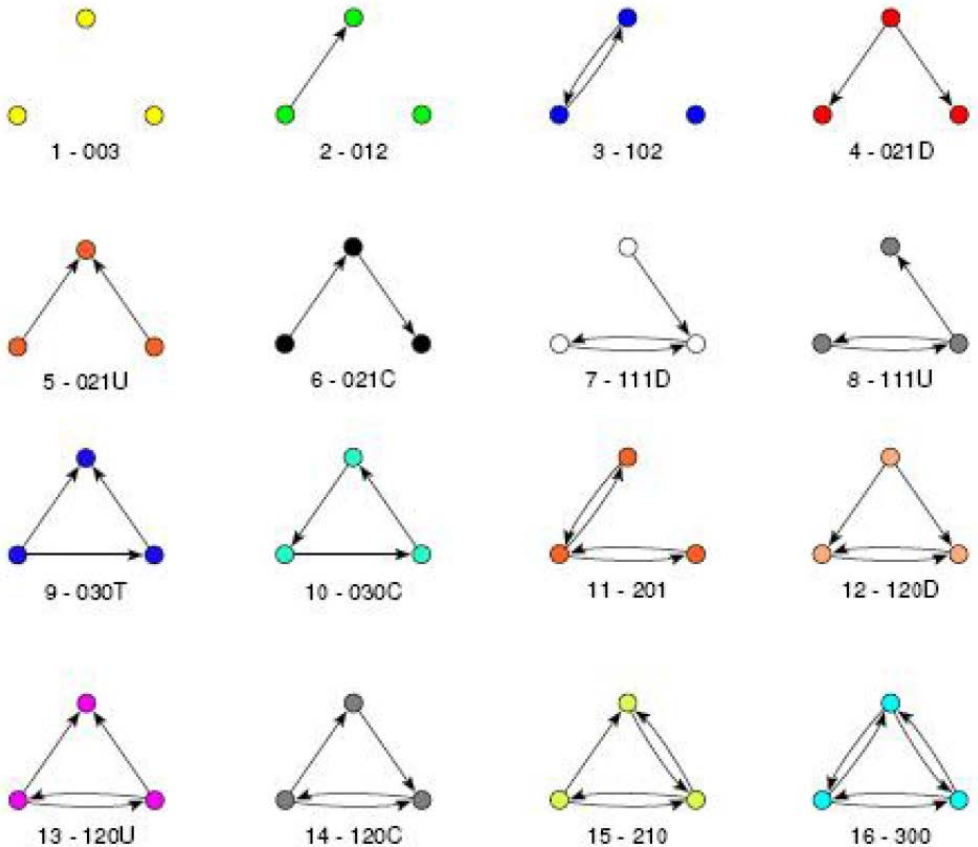
Transitivity

- High transitivity means that networks are clumpy
- Most ties are within clumps, fewer between clumps
- This means that the network is locally cohesive but fragmented on a global scale

Transitivity

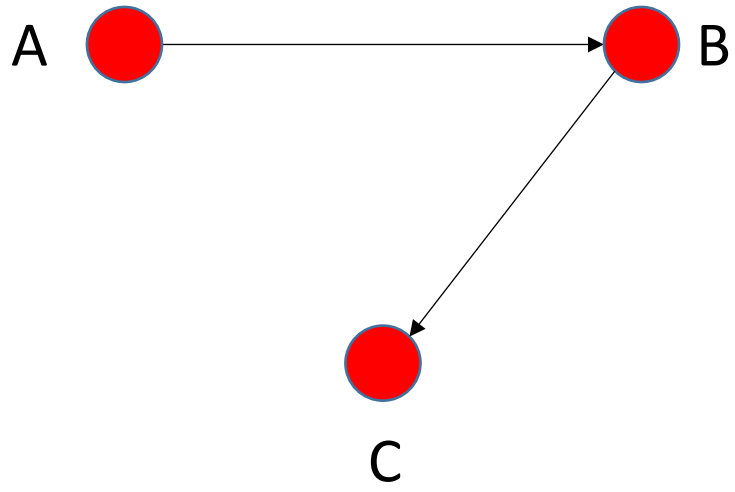
- Triad census: With directed data, there are 16 possible types of relations among 3 actors
  - Including relationships that show hierarchy, equality, exclusion
  - That is, a variety of forms of all social relationships

Transitivity



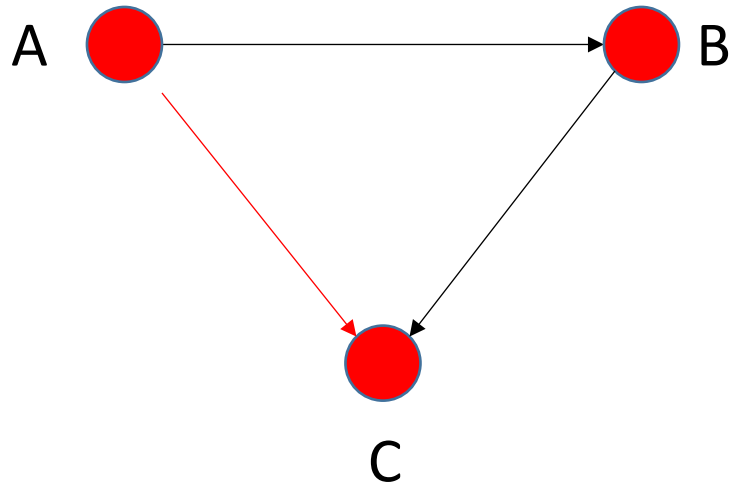
Transitivity

- Triplet transitivity: Chance that if we observe A → B and B → C



Transitivity

- Triplet transitivity: Chance that if we observe A → B and B → C that we will also observe A → C



Transitivity in UCInet

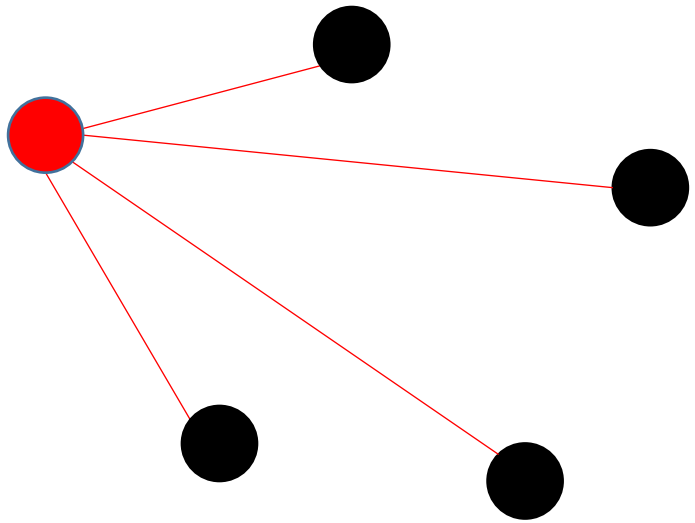
Clustering coefficient

- Remember the small world?
  - High density locally, low density globally
    (but some cross-cutting ties that make the network a small world; so geodesic distances are small on average)
  - How can we measure clustering in local neighborhoods?

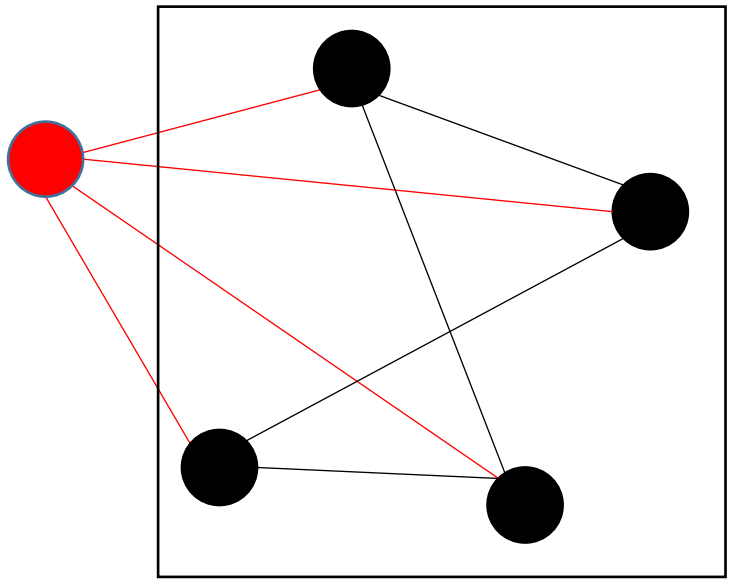
Clustering coefficient

- Remember the small world?
  - High density locally, low density globally
    (but some cross-cutting ties that make the network a small world; so geodesic distances are small on average)
  - How can we measure clustering in local neighborhoods?
    - Take every actor (ego), examine his/her local neighborhood (i.e., all nodes to which he/she is directly tied), and calculate the density in this neighborhood (excluding ego)
    - Do this for all actors and take the average

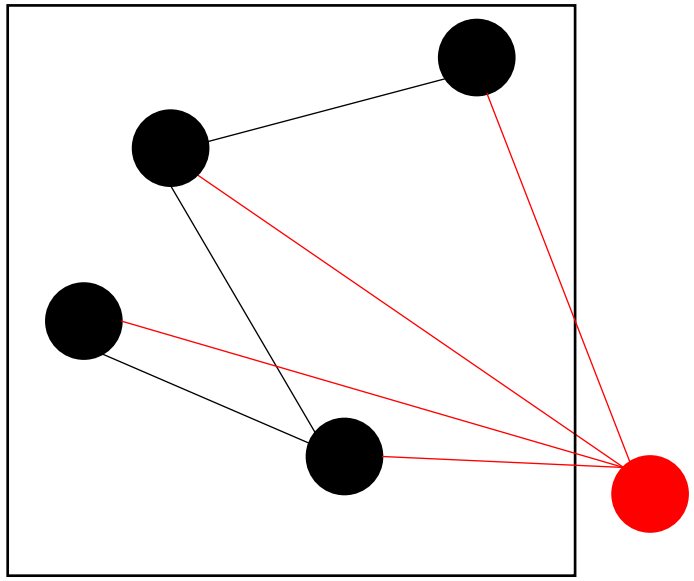
Clustering coefficient



Clustering coefficient



Clustering coefficient

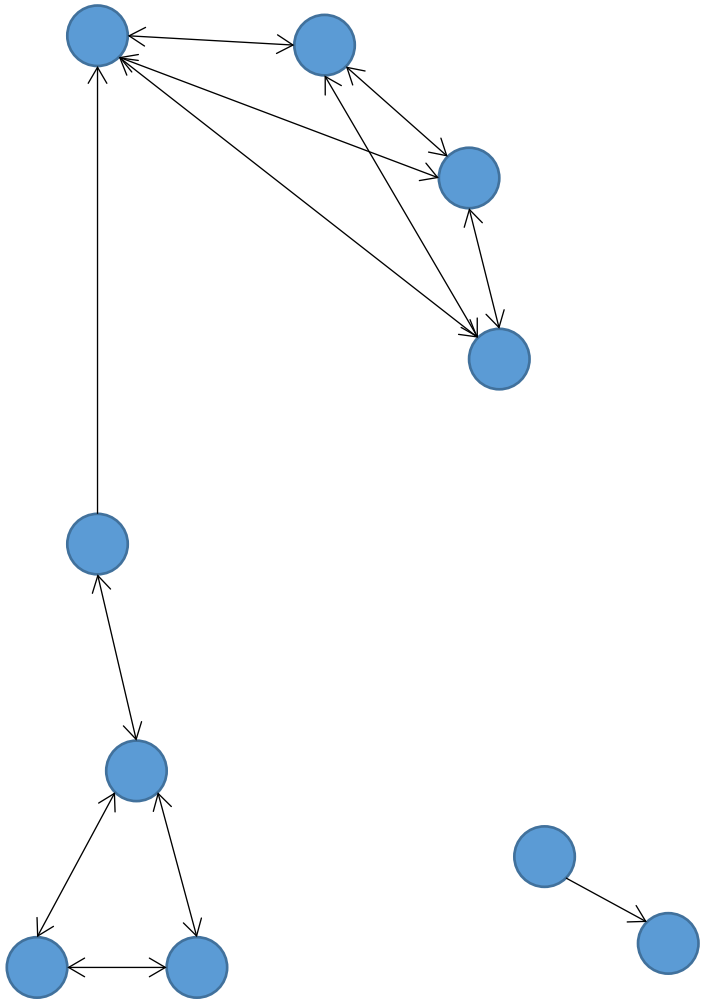


Clustering coefficient

- Overall clustering coefficient
  - Simply the average of the densities of the neighborhoods of all actors
- Weighted clustering coefficient
  - Gives weight proportional to neighborhood size
    → larger ones get more weight
  - Because larger subgraphs also tend to be less dense, the weighted version is usually smaller than the overall version
- Important: Compare to overall density to see whether local clustering really exists

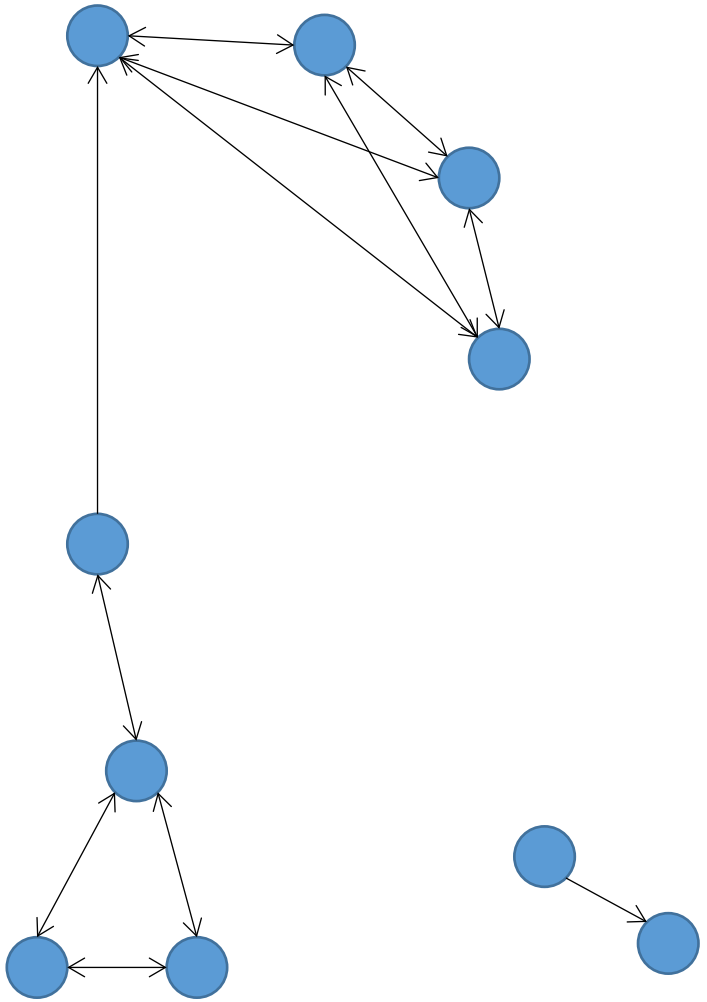
Clustering coefficient in UCInet

Fragmentation



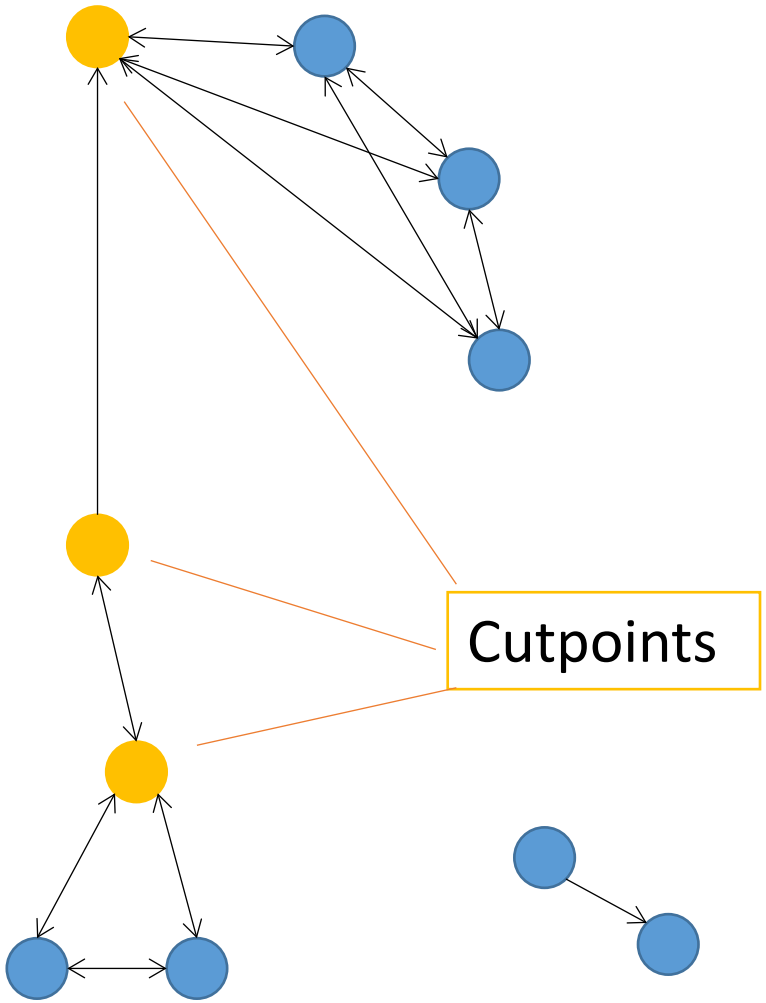
Fragmentation

- Measure of fragmentation:
  Proportion of pairs of nodes that cannot be reached from each other
- Cutpoint:
  A node that would increase the number of components if removed



Fragmentation

- Measure of fragmentation:
  Proportion of pairs of nodes that cannot be reached from each other
- Cutpoint:
  A node that would increase the number of components if removed



Fragmentation in UCInet

Centrality

What is centrality?

- A node-level characteristic
- Measuring a node’s position in the network
- What does a central node have that other nodes don’t?

What is centrality?

- A node-level characteristic
- Measuring a node’s position in the network
- What does a central node have that other nodes don’t?
  - Social capital
  - Risk of contagion (active and passive)
  - Power / influence / control / independence (?)

What is centrality?

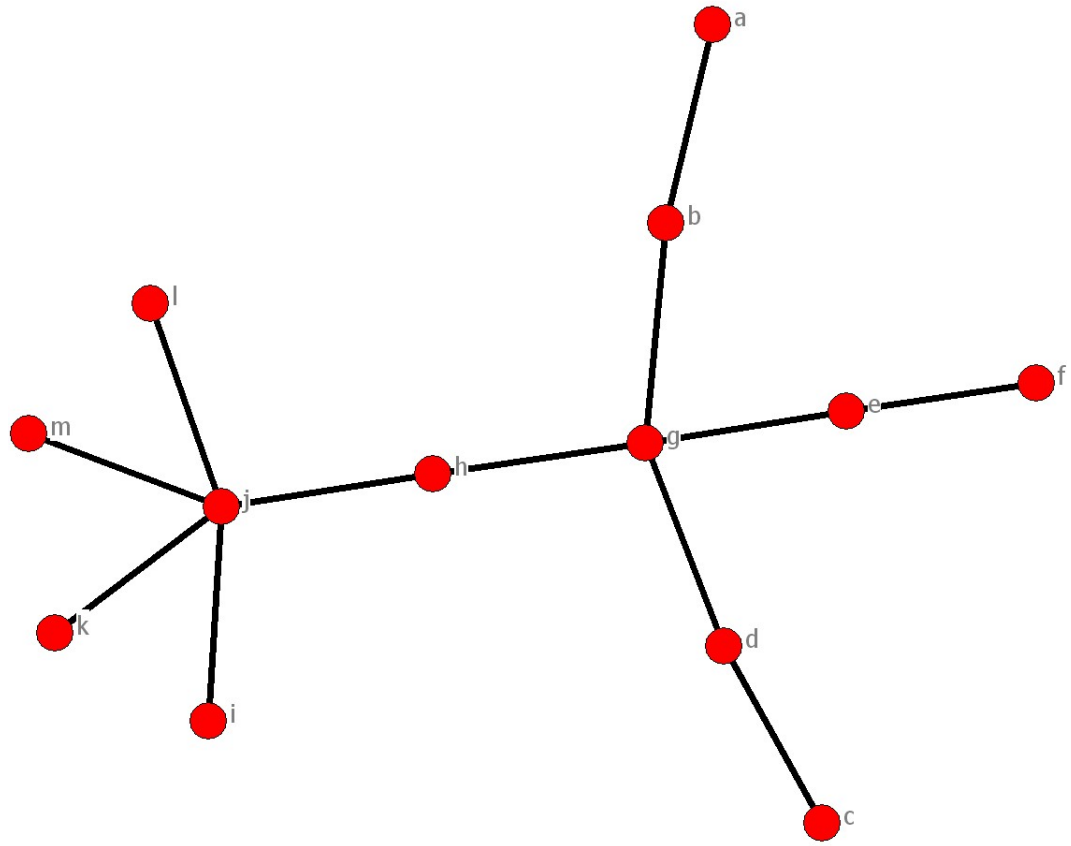
- A family of measures
  - There are different centrality concepts for different aspects of network position
  - The choice of a concept is primarily motivated by theory
  - This theory is typically about things that flow: What flows, and how does it flow?
  - However, the measures are not about the actual flows but about the architecture through which they flow

What is centrality?

- We consider four centrality measures (Freeman 1979)
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvector centrality

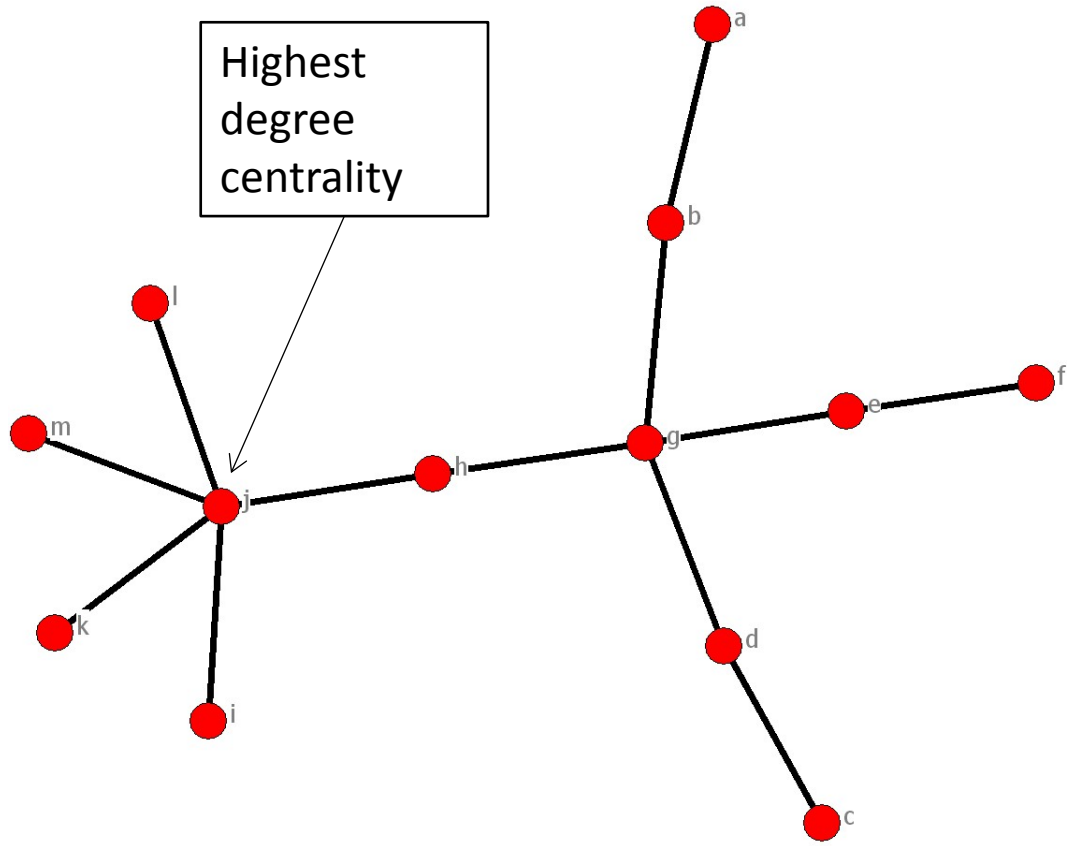
Degree centrality

- The simplest measure: A node's number of ties
- Row sum of the adjacency matrix
- Which node has the highest degree centrality in this graph?



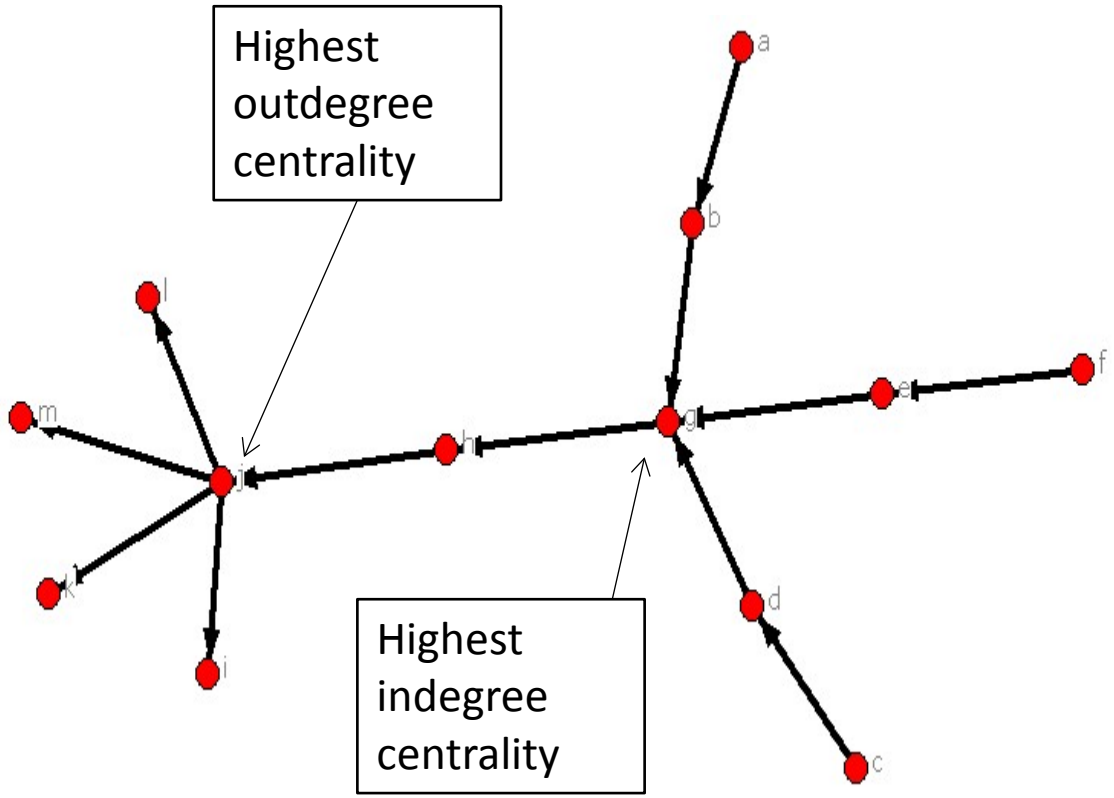
Degree centrality

- The simplest measure: A node's number of ties
- Row sum of the adjacency matrix
- Which node has the highest degree centrality in this graph?



Degree centrality

- In directed graphs: Indegree and outdegree
- What does that mean socially?

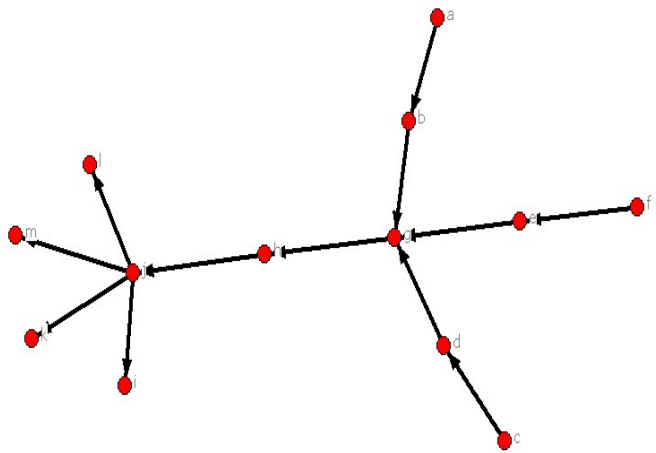


Degree centrality

- High indegree
  - Good for … ?
  - Bad for … ?

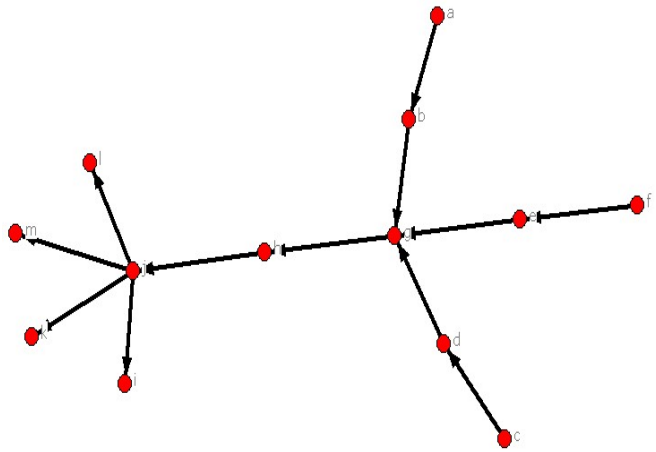
- High outdegree
  - Good for … ?

  - Bad for … ?



Degree centrality

- High indegree
  - Good for obtaining important information
  - Bad for catching the flu
  - Higher chance of being influenced
- High outdegree
  - Good for influencing or activating others
  - Bad for spreading harmful flows



Degree centrality

- Calculated as the row sum and col sum of the adjacency matrix
- In valued graphs: Same procedure

	HOLLY	BRAZEY	CAROL	PAM	PAT	JENNIE	PAULINE	ANN	MICHAEL	BILL	LEE	DON	JOHN	HARRY	GERY	STEVE	BERT	RUSS	OUTDEG
HOLLY	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	3
BRAZEY	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	3
CAROL	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	3
PAM	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	3
PAT	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3
JENNIE	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	3
PAULINE	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3
ANN	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	3
MICHAEL	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	3
BILL	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	3
LEE	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	3
DON	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	3
JOHN	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	3
HARRY	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	3
GERY	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	3
STEVE	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	3
BERT	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	3
RUSS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	3
INDEG	4	1	2	5	4	3	4	2	4	0	3	4	0	3	2	5	4	4	54

Degree centrality

- Most important limitation?

Degree centrality

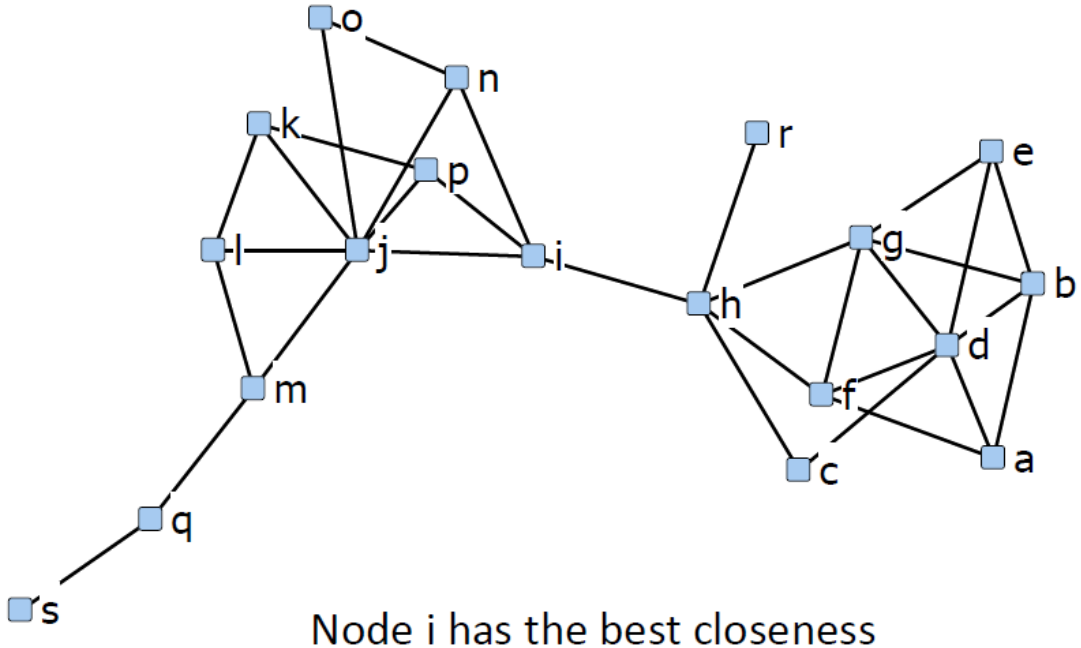
- Most important limitation?
  - Does not account for the wider structure of the network

Closeness centrality

- How far is a node, on average, from all other nodes in the network?
- Nodes that are closer to all other nodes are viewed as more central
- Calculation based on the geodesic distance matrix
  - Number of links in the shortest paths to the other nodes
  - Large numbers mean “less central”

Closeness centrality

ID	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	sum
a	0	1	2	1	2	1	2	2	3	4	5	5	5	4	5	4	6	3	7	62
b	1	0	2	1	1	2	1	2	3	4	5	5	5	4	5	4	6	3	7	61
c	2	2	0	1	2	2	2	1	2	3	4	4	4	3	4	3	5	2	6	52
d	1	1	1	0	1	1	1	2	3	4	5	5	5	4	5	4	6	3	7	59
e	2	1	2	1	0	2	1	2	3	4	5	5	5	4	5	4	6	3	7	62
f	1	2	2	1	2	0	1	1	2	3	4	4	4	3	4	3	5	2	6	50
g	2	1	2	1	1	1	0	1	2	3	4	4	4	3	4	3	5	2	6	49
h	2	2	1	2	2	1	1	0	1	2	3	3	3	2	3	2	4	1	5	40
i	3	3	2	3	3	2	2	1	0	1	2	2	2	1	2	1	3	2	4	39
j	4	4	3	4	4	3	3	2	1	0	1	1	1	1	1	1	2	3	3	42
k	5	5	4	5	5	4	4	3	2	1	0	1	2	2	2	1	3	4	4	57
l	5	5	4	5	5	4	4	3	2	1	1	0	1	2	2	2	2	4	3	55
m	5	5	4	5	5	4	4	3	2	1	2	1	0	2	2	2	1	4	2	54
n	4	4	3	4	4	3	3	2	1	1	2	2	2	0	1	2	3	3	4	48
o	5	5	4	5	5	4	4	3	2	1	2	2	2	1	0	2	3	4	4	58
p	4	4	3	4	4	3	3	2	1	1	1	2	2	2	2	0	3	3	4	48
q	6	6	5	6	6	5	5	4	3	2	3	2	1	3	3	3	0	5	1	69
r	3	3	2	3	3	2	2	1	2	3	4	4	4	3	4	3	5	0	6	57
s	7	7	6	7	7	6	6	5	4	3	4	3	2	4	4	4	1	6	0	86
sum	62	61	52	59	62	50	49	40	39	42	57	55	54	48	58	48	69	57	86	1048



Node i has the best closeness

Closeness centrality

- What does high closeness centrality mean socially?

Closeness centrality

- What does high closeness centrality mean socially?
  - Typical interpretation: Time-until-arrival of a flow
  - For example, central nodes hear gossip early
  - Simulations show that when things travel along the geodesics, closeness gives an exact measure of the expected time until arrival

Closeness centrality

- Closeness in directed graphs
  - In-closeness
  - Out-closeness
- Closeness in valued graphs
  - Draw on adjustment algorithms (depends on what kind of valued tie you study)

Closeness centrality

- Most important limitation
  - If the graph is disconnected, not all distances are defined
  - Time until arrival is infinite when there is no path for delivery
  - Possible solutions
    - Calculate only within components
    - Replace undefined distances with a large value

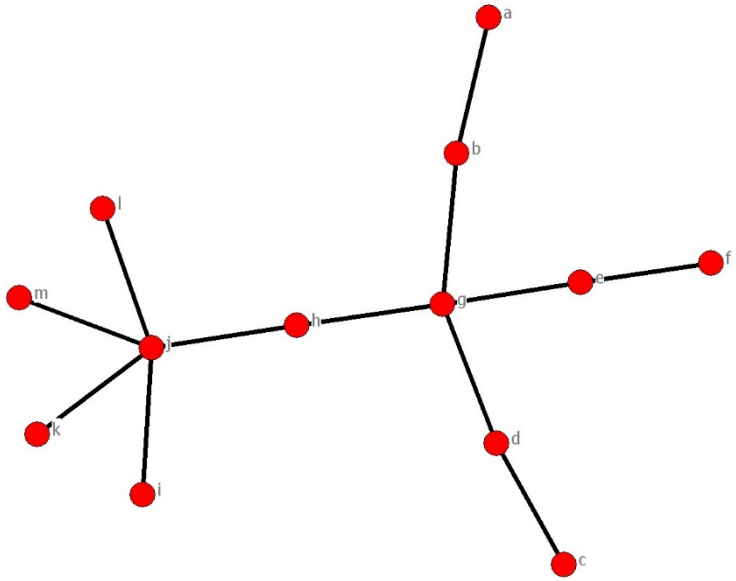
Betweenness centrality

- Nodes that are frequently on the shortest paths between other nodes are viewed as central
- Share of geodesics between all pairs of nodes that pass through a node

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

gij is number of geodesic paths from i to j
gikj is number of geodesics from i to j that pass through k

- Which node has the highest betweenness centrality?



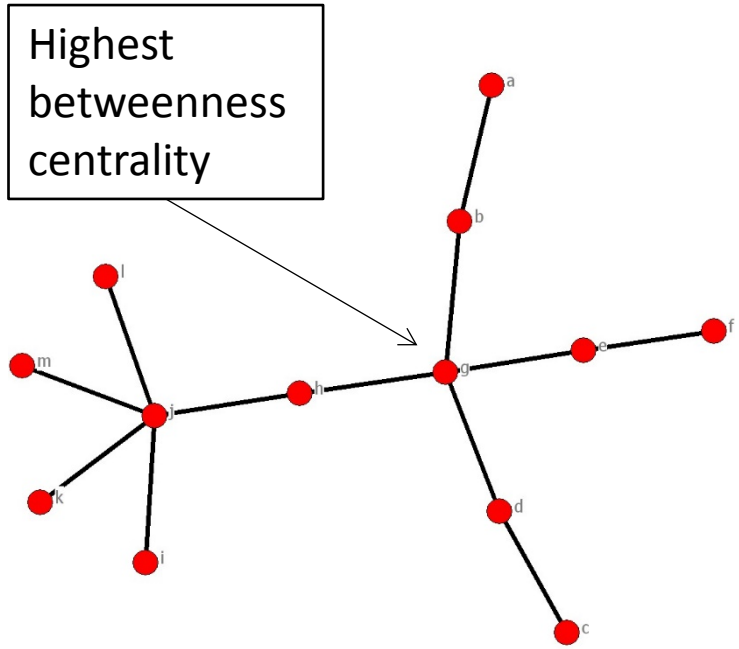
Betweenness centrality

- Nodes that are frequently on the shortest paths between other nodes are viewed as central
- Share of geodesics between all pairs of nodes that pass through a node

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

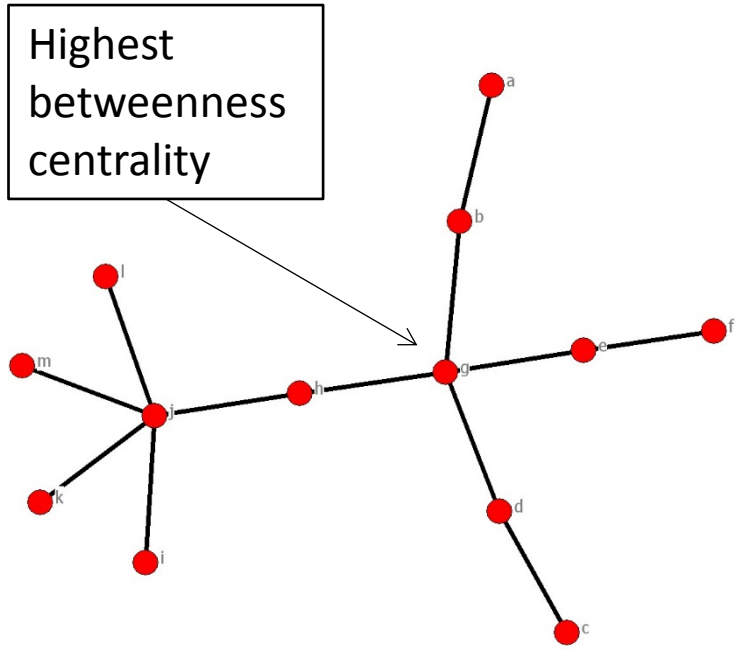
gij is number of geodesic paths from i to j
gikj is number of geodesics from i to j that pass through k

- Which node has the highest betweenness centrality?



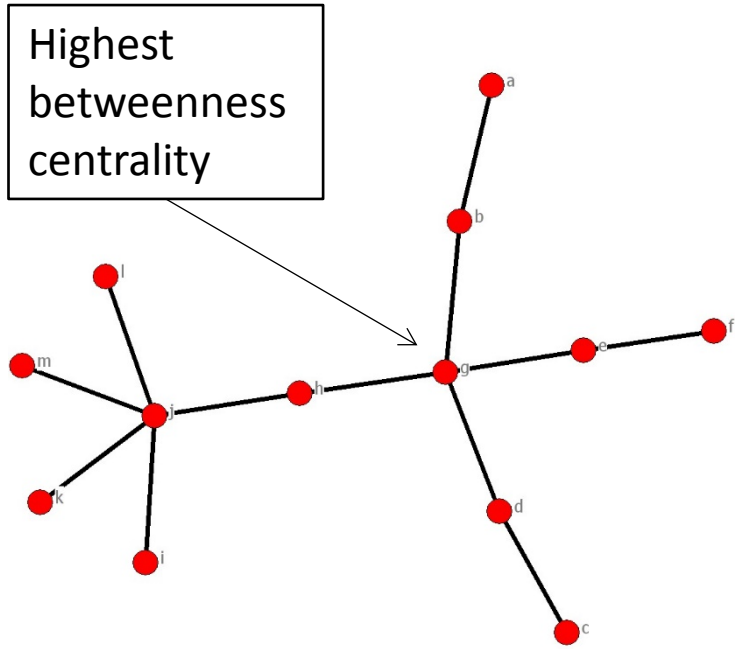
Betweenness centrality

- What does betweenness centrality mean socially?



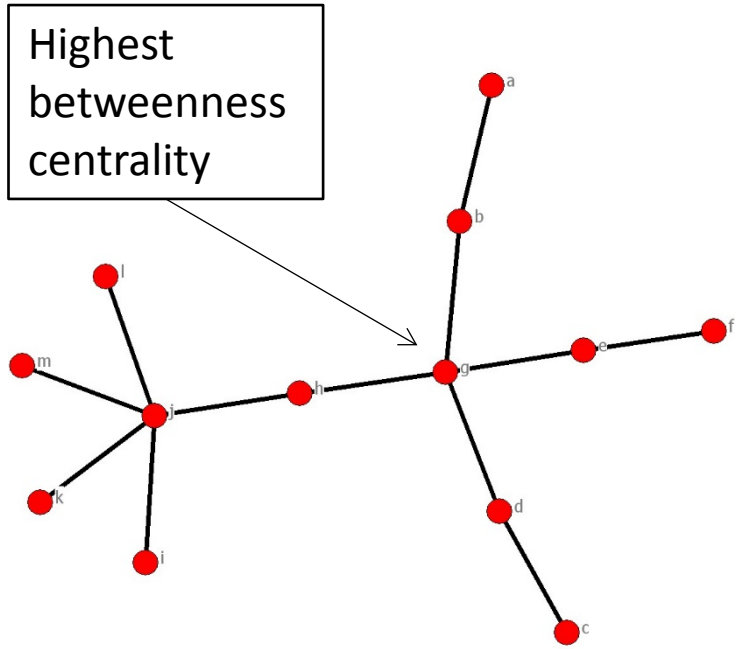
Betweenness centrality

- What does betweenness centrality mean socially?
  - Potential of filtering or modifying the things that flow through the network (gatekeeper)
  - Frequency of exposure:
    Expected number of times a flow would pass through the node
  - Important for global cohesion
  - This power is often invisible
    (in contrast to degree)



Betweenness centrality

- Betweenness in directed graphs
  - No adjustment needed
- Betweenness in valued graphs
  - Some adjustment (depends)

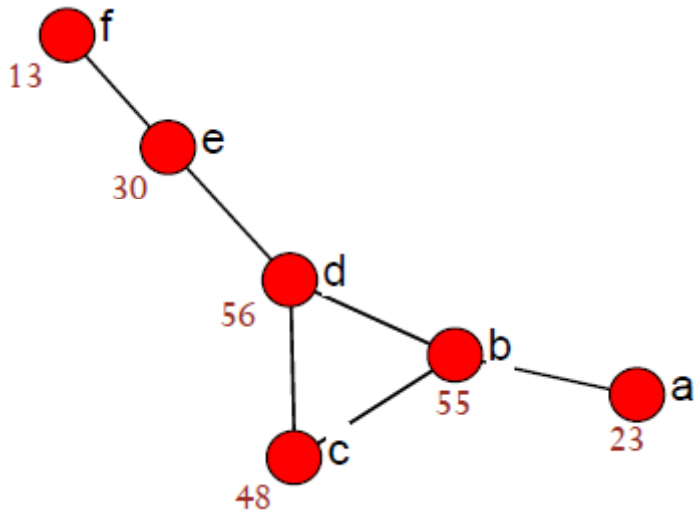


Eigenvector centrality

- A high eigenvector score means that a node is connected to many nodes who themselves have high scores
- “Turbo-charged” degree centrality

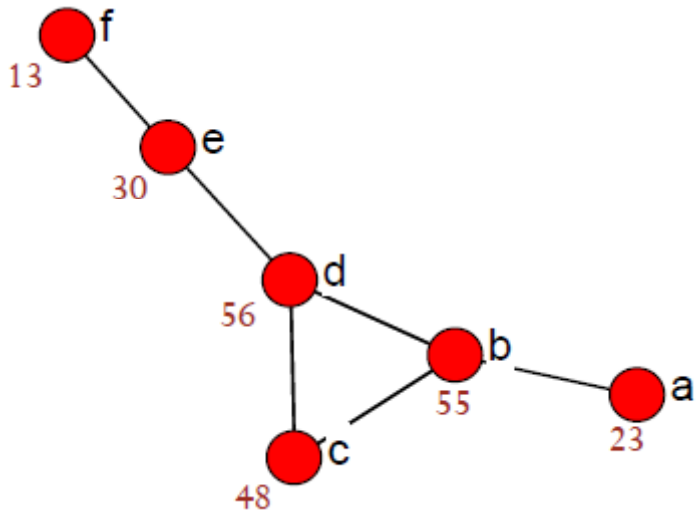
Eigenvector centrality

- A high eigenvector score means that a node is connected to many nodes who themselves have high scores
- “Turbo-charged” degree centrality
- Why is node **a** more central than node **f** (each having degree 1)?



Eigenvector centrality

- A high eigenvector score means that a node is connected to many nodes who themselves have high scores
- “Turbo-charged” degree centrality
- Why is node **a** more central than node **f** (each having degree 1)?
  - Because **b** has more friends than **e**



Eigenvector centrality

- What does eigenvector centrality mean socially?

Eigenvector centrality

- What does eigenvector centrality mean socially?
  - Popularity / status: Have ties not just to many others but to many well-connected others
  - Influence

Eigenvector centrality

- In directed graphs
  - Left eigenvector (columns) is popularity: a node scores high if chosen by nodes who are often chosen
  - Right eigenvector (rows) is influence: node scores high if choosing nodes who choose many others
  - Can not always be calculated (beta centrality is a substitute)
- In valued graphs
  - No adjustment needed

Eigenvector centrality

- Limitations
  - Problematic in disconnected graphs
    - Solution: study each component separately or use beta centrality
  - Problematic in clumpy networks (favors nodes in larger cliques)
  - Not good as a measure of risk/exposure
  - Sometimes odd or non-calculable in directed graphs

Exercises

- Use UCInet to unpack **STUDENTS**
  - Open STUDENTS-GOSSIP in Netdraw to familiarize yourself with the data.
  - Use Data>Filter/Extract to extract the main component of STUDENTS-GOSSIP
  - Then use UCInet to calculate the four centrality measures for the main component: Degree, Betweenness, Closeness, and Eigenvector (Network>Centrality and Power)
  - Which student can be expected to hear earliest of a new gossip?
  - Which student is in the best position to modify the gossip, slow it down or stop its spread?
- Open PRISON in Netdraw. These are directed friendship relations among prisoners
  - Who would you say is the most popular prisoner? Calculate the appropriate measure in UCInet and explain your rationale.
  - Use Netdraw to illustrate prisoners' popularity using to the measures of Indegree and InEigenvector
  - Compare prisoners 22 and 56 on these measures. What do you conclude?