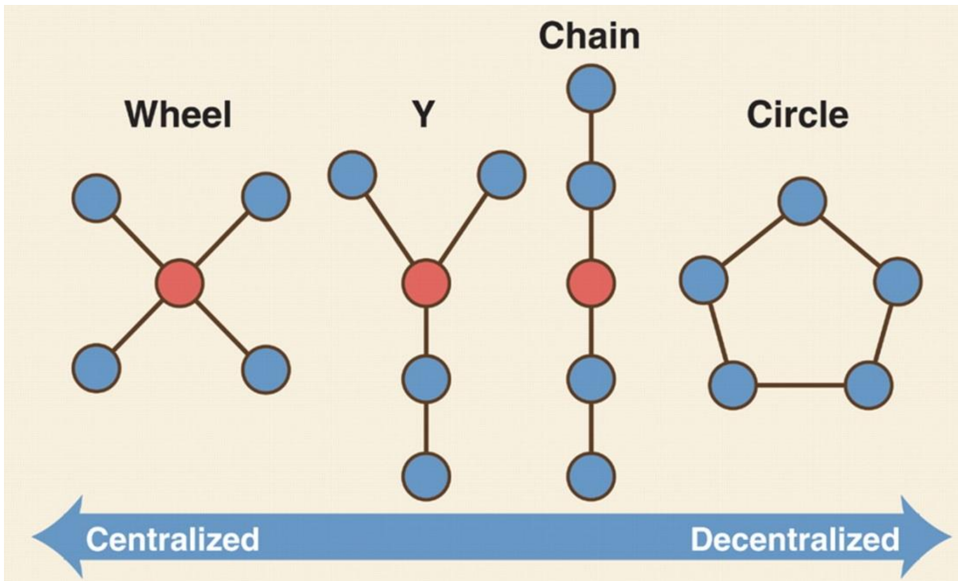
SNA

Centrality

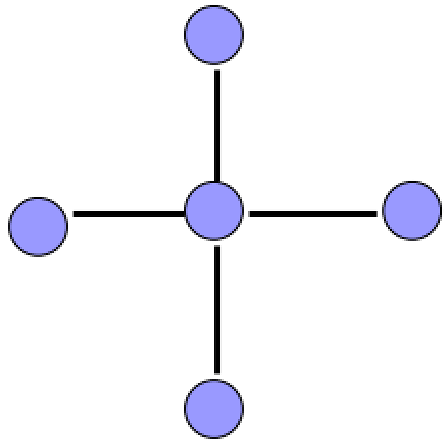
The Bavelas-Leavitt Experiment



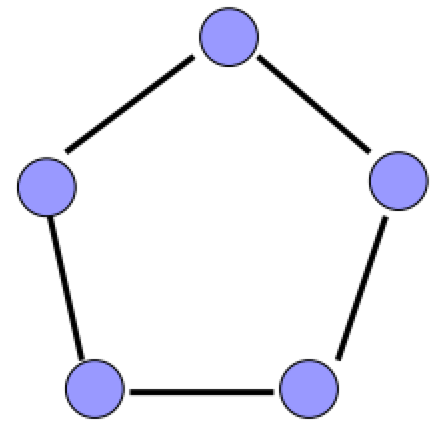
Variables	Simple Task	Complex task
Fewest messages	Centralized	Centralized
Least errors	Centralized	Decentralized
Least time	Centralized	Decentralized
Satisfaction	Decentralized	Decentralized

Network centralization

- One or few actors are quite central, other actors are not central
- Intuition



Centralized



Not centralized

Centrality

- Actor centrality (Micro)
  - Identify important or prominent actors
- Network centralization (Macro)
  - Characterize the structure of the network
    - Equality or evenness
  - If a system is very loosely coupled (sparse linkage) not much power can be exerted; in high density systems there is the potential for greater power.

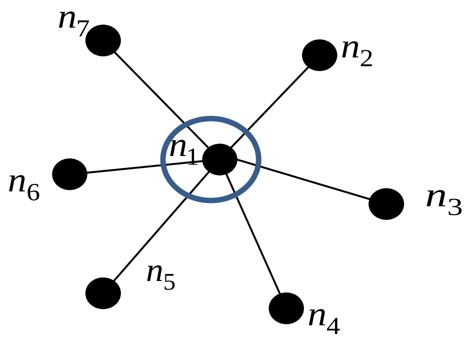
Prominent individuals in a network

- Not all nodes are created equal
  - Centrality
    - Prominent actors have higher involvement in many relations
  - Prestige (in a directional network)
    - Where prominent actors initiates few relations but receives many directed ties.
    - E.g. PageRank

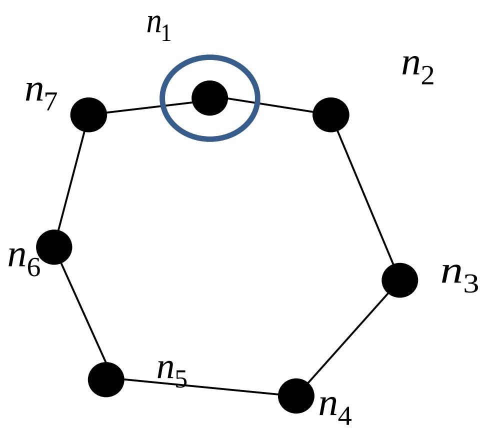
Centrality

- Network view of "power"
  - Manifest itself in social relationship
  - An ego's power is alter's dependence
- Centrality analysis identifies individuals having a favored position, having more opportunities and fewer constraints, and presumably possess more power

Micro and Macro view of centrality

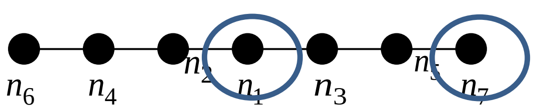


(a) Star Graph



(b) Circle Graph

(c) Line Graph



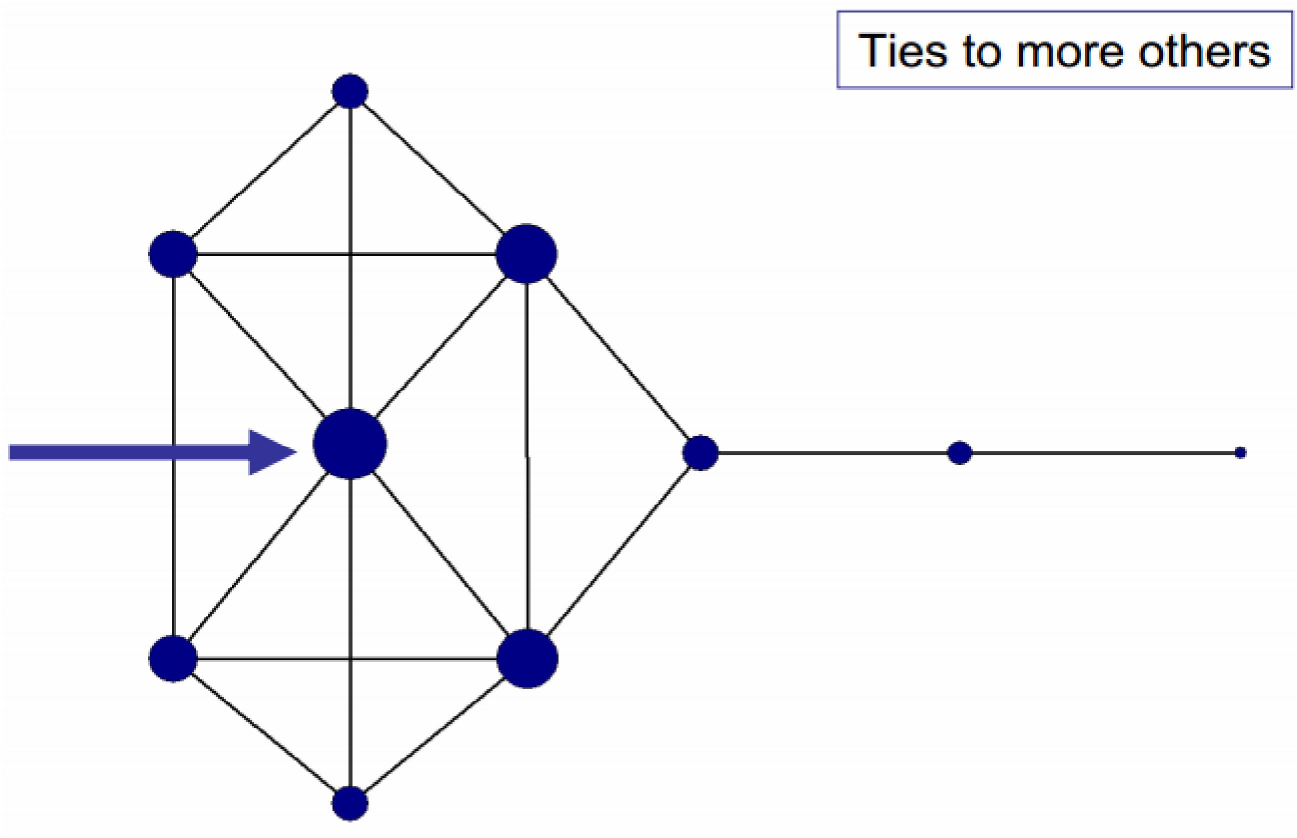
Micro: which node is the most central in each graph?

Macro: which graph has the most unevenly distributed degree ??

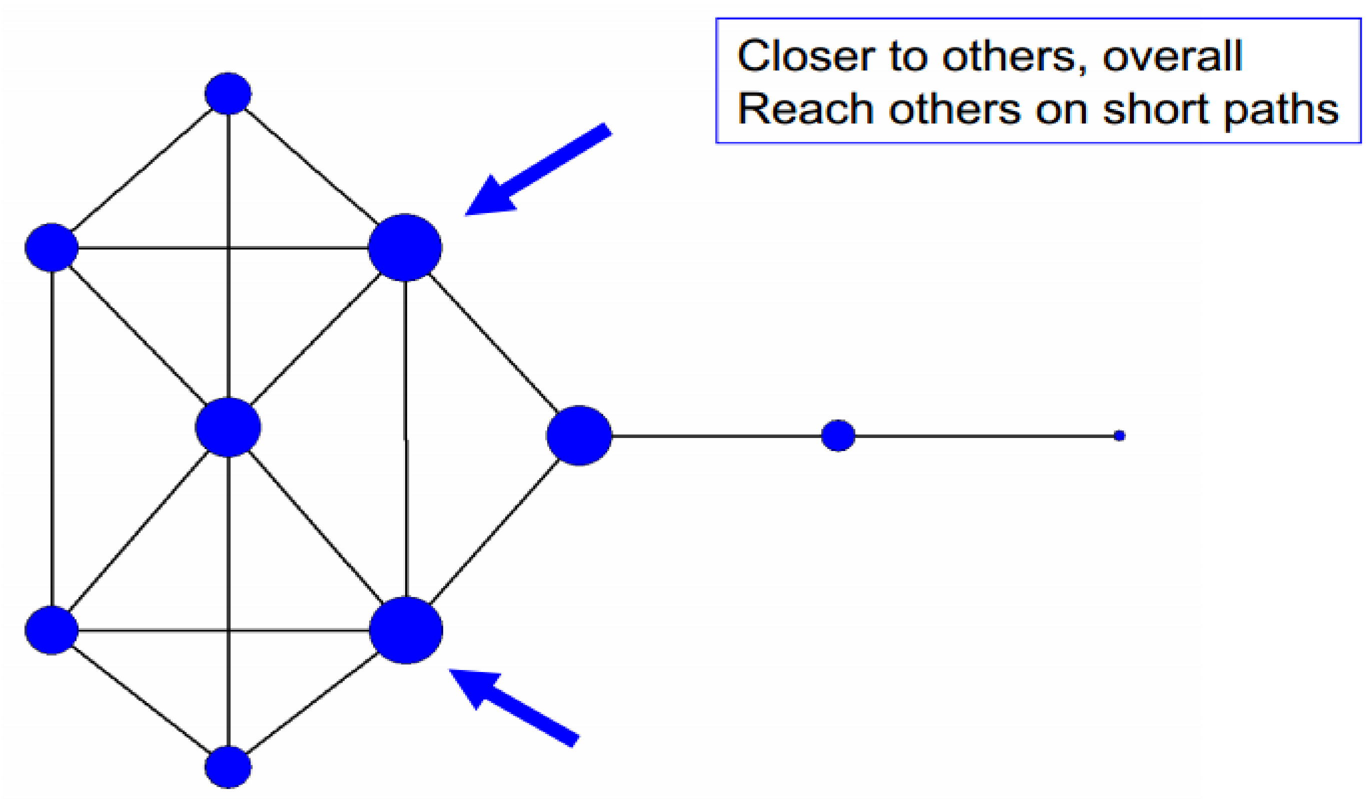
Centrality

- Degree centrality
  - Counting the number of paths of length 1 emanating from a node
- Closeness centrality
  - The total geodesic distance from a given node to all other nodes
- Betweenness centrality
  - The extent that that node falls on the geodesic paths between other pairs of nodes
- Eigenvector centrality
  - assigns relative scores to all nodes in the network based on the principle that **connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes**.

Actor degree centrality

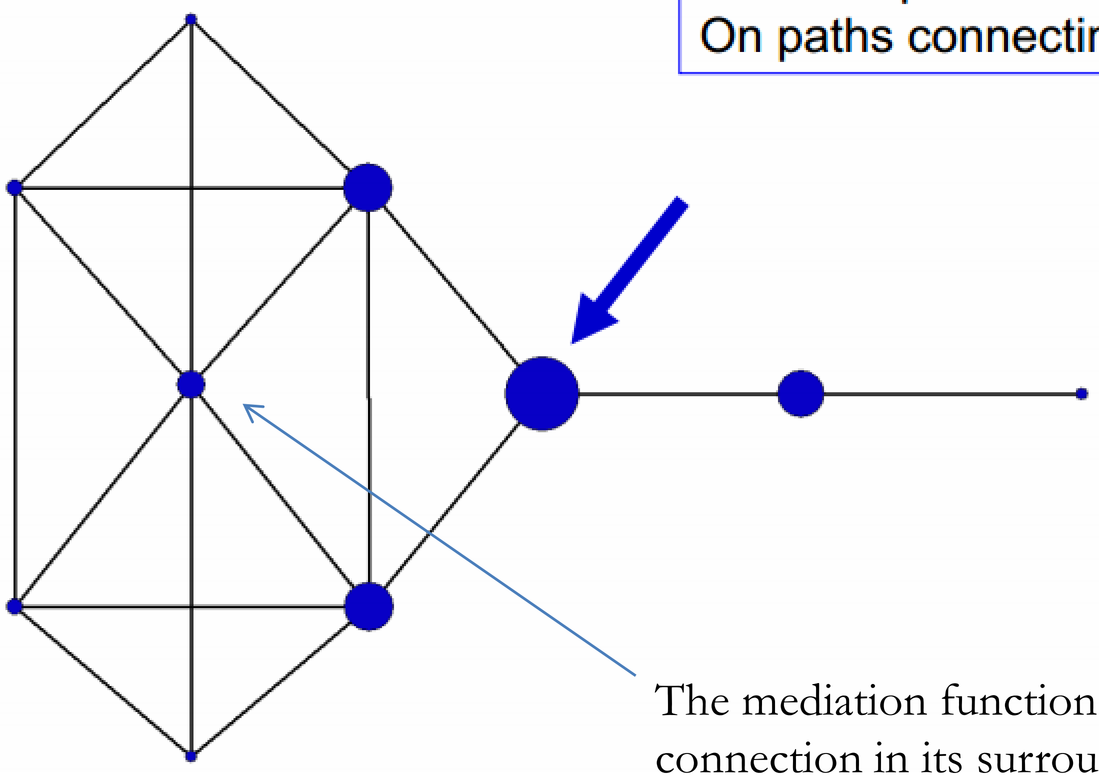


Actor closeness centrality



Actor betweenness centrality

Between pairs of actors
On paths connecting other actors

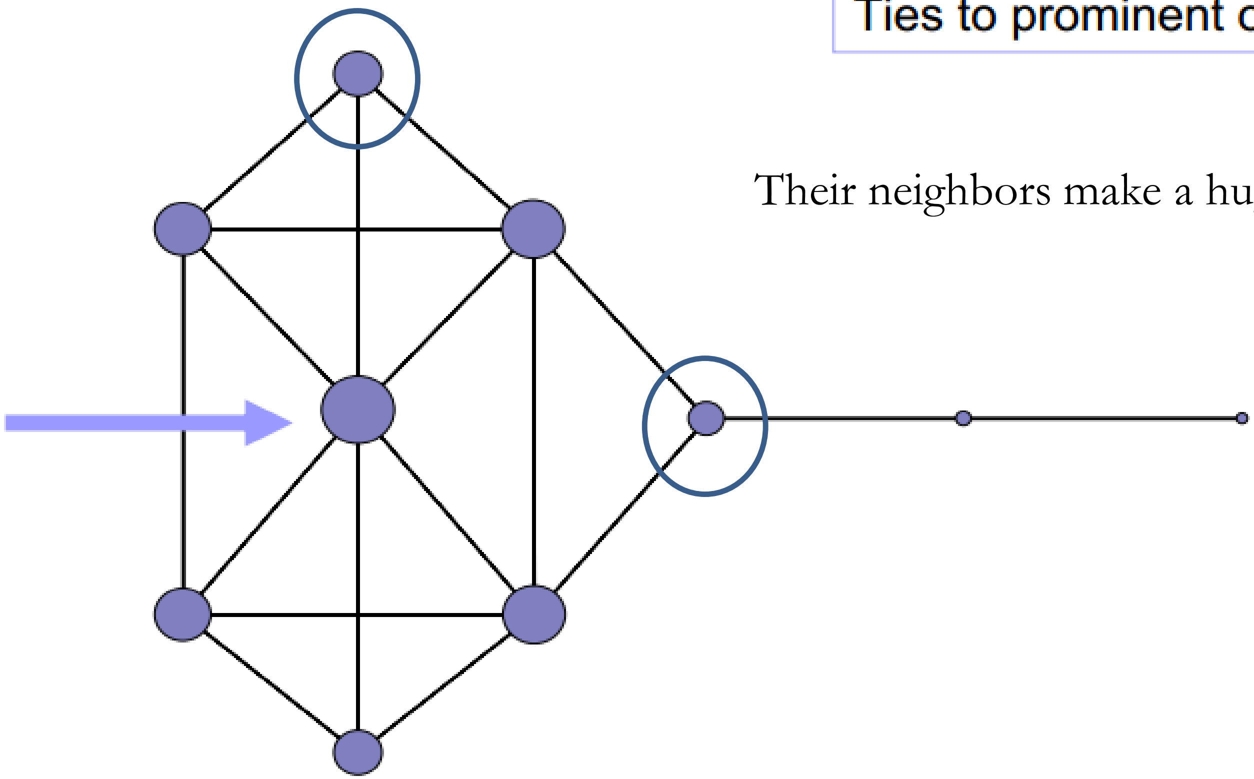


The mediation function is hurt by the dense connection in its surrounding

Actor eigenvector centrality

Ties to prominent others

Their neighbors make a huge difference



Structural intuitions for actor centrality

- Intuition
  - Activity, visibility
  - Brokerage, control
  - Efficiency
  - Status, rank
- Measure
  - Degree
  - Betweenness
  - Closeness
  - Eigenvector

Actor Degree Centrality

- Measures the extent to which a node connect to all other nodes

$$C_D(Ni) = \sum_{j=1}^{g} X_{ij} \; (i \neq j)$$

For a network of size g,
Counts the number of direct ties that node i has to the g-1 other J nodes

General form of network attributes

1. Raw Measure—raw measures of network characteristics reflect a given characteristic that is not sensitive to the size or boundaries of a network or any of its websites
2. Normalized Measures—take on the general form of:

$$S_x = \frac{raw\ measure_x}{maximum\ value\ of\ measure\ given\ network\ characteristics}$$

$\rightarrow \mathbf{0 \leq S_x \leq 1}$

Normalized/standardized Actor Degree Centrality

- To eliminate the effect of variation in network size
  - Allow meaningful comparison of actors across different networks
- Vary from 0, indicating no connections with any actors (i.e. an isolate) to 1, reflecting direct ties to everyone

$$C'_D(Ni) = \frac{C_D(Ni)}{g-1}$$

Freeman normalize group centralization measure

- Macro-view: measure the extent to which the actors in a social network differ from one another in their individual degree of centrality
- Express the degree of variability in the degrees of all actors in the network
- Normalized by the theoretical maximums (i.e. of a perfect star network of the same size)

Normalized Centrality and Centralization

- **Actor** centrality measure
  - Raw
  - Normalized
- Normalize **group** centralization
  - **Variation measure**

The largest actor degree centrality observed

$$C_A = \frac{\sum_{i=1}^{g}[C_A(N^*) - C_A(Ni)]}{\max \sum_{i=1}^{g} C_A(N^*) - C_A(Ni)]}$$

Theoretical maximum hint: star network

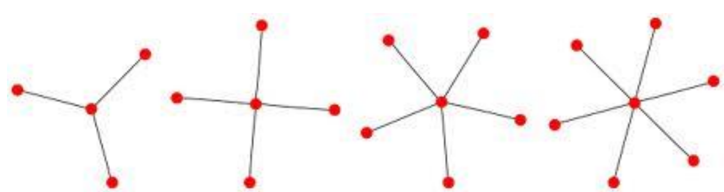
Group Degree Centralization (Macro)

- Measure the extent to which the actors in a social network differ from one another in their individual degree centralities
- A variance measure
- How does it translate into social world?

The largest actor degree centrality observed

$$C_A = \frac{\sum_{i=1}^{g}[C_A(N^*) - C_A(Ni)]}{\max \sum_{i=1}^{g} C_A(N^*) - C_A(Ni)]}$$

a perfect star network of the same size.



Network>Centrality>Degree KNOKBUR

	1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	4.000	5.000	44.444	55.556
2	7.000	8.000	77.778	88.889
3	6.000	4.000	66.667	44.444
4	4.000	5.000	44.444	55.556
5	8.000	8.000	88.889	88.889
6	3.000	1.000	33.333	11.111
7	3.000	9.000	33.333	100.000
8	6.000	2.000	66.667	22.222
9	3.000	5.000	33.333	55.556
10	5.000	2.000	55.556	22.222

DESCRIPTIVE STATISTICS

		1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	Mean	4.900	4.900	54.444	54.444
2	Std Dev	1.700	2.625	18.889	29.165
3	Sum	49.000	49.000	544.444	544.444
4	Variance	2.890	6.890	356.790	850.617
5	SSQ	269.000	309.000	33209.875	38148.148
6	MCSSQ	28.900	68.900	3567.901	8506.173
7	Euc Norm	16.401	17.578	182.236	195.316
8	Minimum	3.000	1.000	33.333	11.111
9	Maximum	8.000	9.000	88.889	100.000

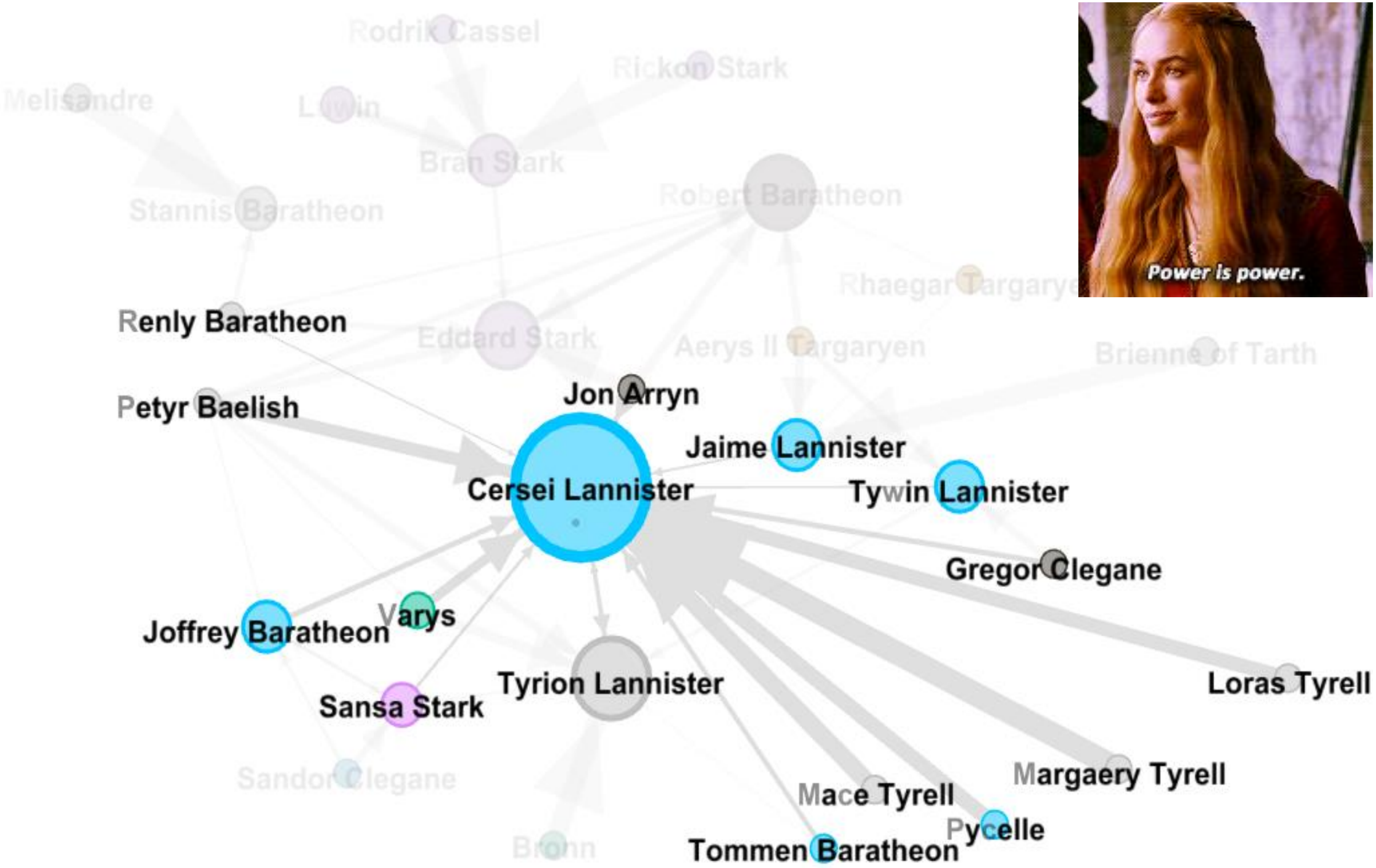
Network Centralization (Outdegree) = 38.272%
Network Centralization (Indegree) = 50.617%

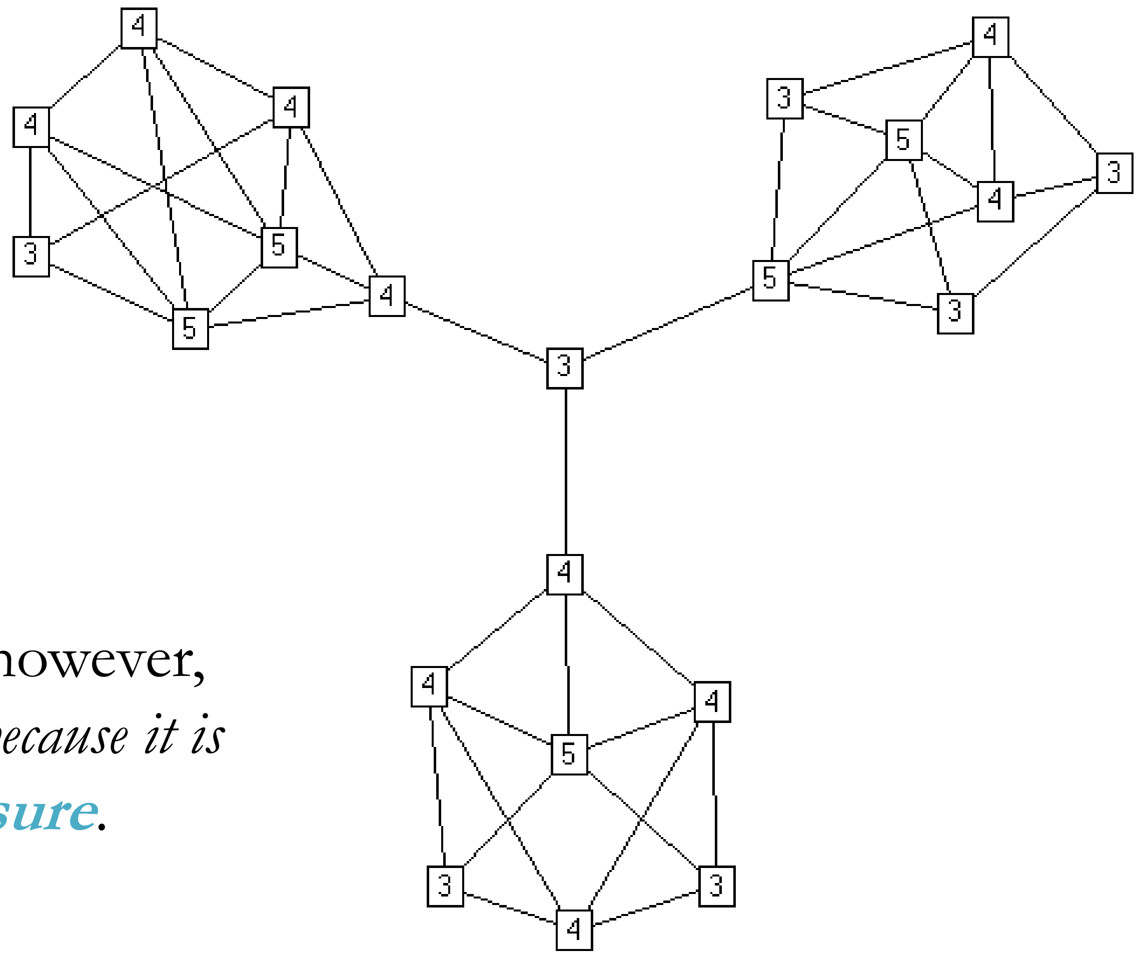
Drawbacks of degree measure

- An actor could be not quite central even with lots of connection
  - Overlook indirect links
  - One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. (occur more likely in a large network)
  - A "local" measure
- Or, a person with few ties yet exert great influence

Cersei is the main character?!

We use the size of node represents the In-degree





Degree centrality, however, can be deceiving, *because it is a purely* ***local measure***.

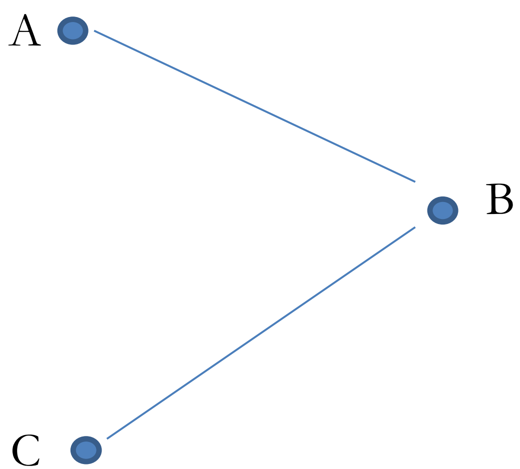
Actor Closeness Centrality

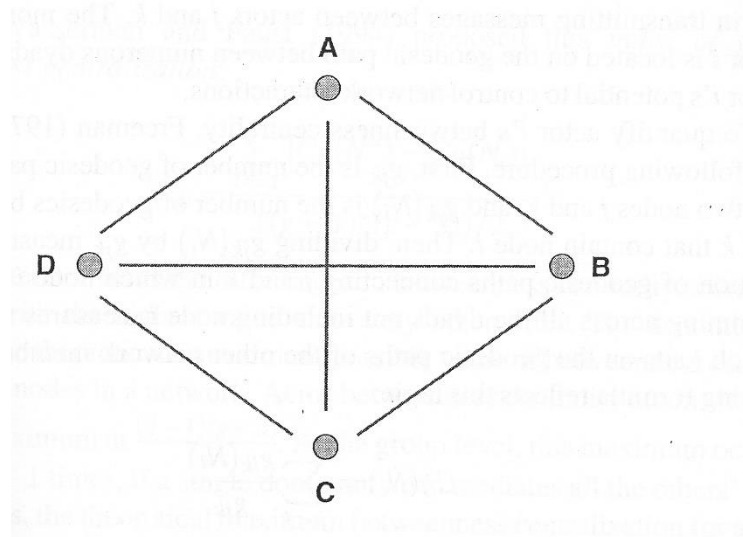
- Emphasizes the distance of an actor to all (not just the neighbors) others
- Reflect how far/near a node is to the other nodes in a social network
- How quickly an actor can interact with others

$$C_C(Ni) = \frac{1}{\left[\sum_{j=1}^{g} d(N_i, N_j)\right]} (i \neq j)$$

Can’t be computed for isolated nodes, whose distances from i are undefined

Closeness centrality normalization





A Four-Node Completely Connected Network

The closeness centrality of A is
1/(1+2) = 1/3 (i.e. ab, ac)

The normalized closeness centrality is?
2/3

The closeness centrality of A is ?
AB, AC, AD **1/(1+1+1)**

The normalized closeness centrality is? 1

Need for normalization (n-1)

Normalized Actor Closeness Centrality

Computed as the inverse of the sum of the geodesic distances between actor i and the g – 1 other actors

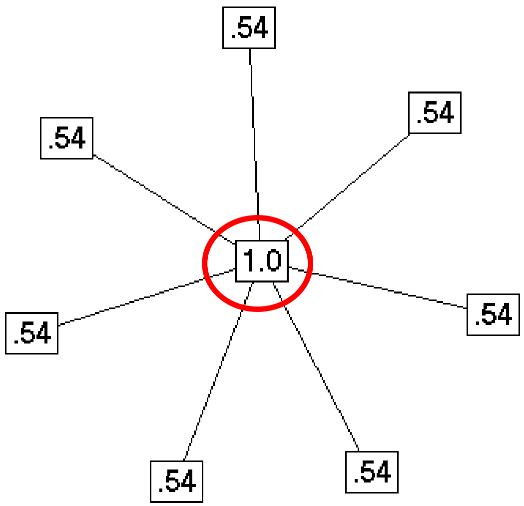
$$C_C(Ni) = \frac{1}{[\sum_{j=1}^{g} d(N_i, N_j)]}(i \neq j)$$

which varies with network size,

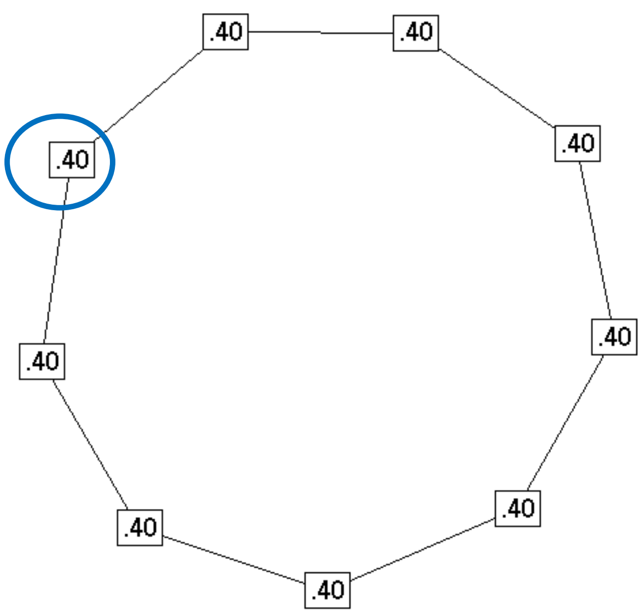
to control for the size of the network

$$C'_C(N_i) = (g-1)(C_C(N_i))$$

Closeness Centrality of two graphs



Distance	closeness	normalized
0 1 1 1 1 1 1 1	.143	1.00
1 0 2 2 2 2 2 2	.077	.538
1 2 0 2 2 2 2 2	.077	.538
1 2 2 0 2 2 2 2	.077	.538
1 2 2 2 0 2 2 2	.077	.538
1 2 2 2 2 0 2 2	.077	.538
1 2 2 2 2 2 0 2	.077	.538
1 2 2 2 2 2 2 0	.077	.538



Distance	closeness	normalized
2 3 4 4 3 2 1	.050	.400
1 2 3 4 4 3 2	.050	.400
0 1 2 3 4 4 3	.050	.400
1 0 1 2 3 4 4	.050	.400
2 1 0 1 2 3 4	.050	.400
3 2 1 0 1 2 3	.050	.400
4 3 2 1 0 1 2	.050	.400
4 4 3 2 1 0 1	.050	.400
3 4 4 3 2 1 0	.050	.400

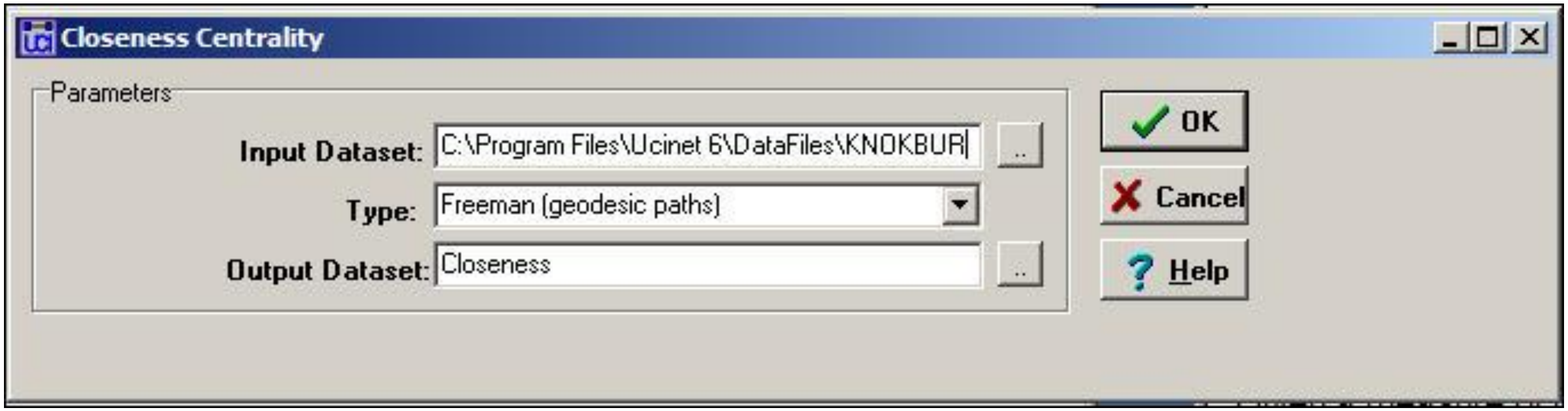
Index of Group Closeness Centralization (Macro)

- The extent to which actors in a given network differ in their closeness centrality
- A dispersion measure indicating the hierarchy of closeness centralities within a network (p. 67)
- It's sociological implications?

The largest actor closeness centrality observed

$$C_C = \frac{\sum_{i=1}^{g}[C'_C(N^*) - C'_C(N_i)]}{[(g-2)(g-1)]/(2g-3)}$$

Path distances

- Network>Centrality>Closeness
- 

Closeness : the reciprocal of “farness”
(i.e. the sum of the lengths of the shortest paths from an ego to all other nodes

Closeness Centrality Measures

	1 inFarness	2 outFarness	3 inCloseness	4 outCloseness
7	9.000	16.000	100.000	56.250
5	10.000	10.000	90.000	90.000
2	10.000	11.000	90.000	81.818
4	13.000	15.000	69.231	60.000
9	13.000	16.000	69.231	56.250
1	14.000	15.000	64.286	60.000
3	14.000	12.000	64.286	75.000
10	16.000	13.000	56.250	69.231
8	17.000	13.000	52.941	69.231
6	22.000	17.000	40.909	52.941

Statistics

		1 inFarness	2 outFarness	3 inCloseness	4 outCloseness
1	Mean	13.800	13.800	69.713	67.072
2	Std Dev	3.682	2.227	17.584	11.616
3	Sum	138.000	138.000	697.133	670.721
4	Variance	13.560	4.960	309.201	134.925
5	SSQ	2040.000	1954.000	51691.488	46335.906
6	MCSSQ	135.600	49.600	3092.015	1349.255
7	Euc Norm	45.166	44.204	227.358	215.258
8	Minimum	9.000	10.000	40.909	52.941
9	Maximum	22.000	17.000	100.000	90.000

Network in-Centralization = 71.51%
Network out-Centralization = 54.14%

Measure of inequality in the distribution of distances across the actors

Actor 6 has the largest sum of geodesic distances from other actors (inFarness of 22) and to other actors (outFarness of 17).

The farness figure can then be re-expressed as nearness and normed relative to the greatest nearness observed in the graph (the inCloseness of actor 6)

Normalized Freeman closeness using new closeness procedure

Matrix: KNOKI

		1 OutFreeClo	2 InFreeClo	3 OutValClo	4 InValClo	5 OutRecipClo	6 InRecipClo
1	COUN	0.600	0.643	0.778	0.815	0.704	0.759
2	COMM	0.818	0.900	0.926	0.963	0.889	0.944
3	EDUC	0.750	0.643	0.889	0.815	0.833	0.722
4	INDU	0.600	0.692	0.778	0.852	0.704	0.778
5	MAYR	0.900	0.900	0.963	0.963	0.944	0.944
6	WRO	0.529	0.409	0.704	0.519	0.630	0.463
7	NEWS	0.563	1.000	0.741	1.000	0.648	1.000
8	UWAY	0.692	0.529	0.852	0.704	0.815	0.593
9	WELF	0.563	0.692	0.741	0.852	0.648	0.778
10	WEST	0.692	0.563	0.852	0.741	0.778	0.611

Different operationalization of closeness centrality

Sum of geodesic distances(Freeman) distances are lengths of shortest paths (geodesic), the standard Freeman measure.

Sum of reciprocal distances As an alternative to taking the reciprocal after the summation, the reciprocals can be taken before. In this case the closeness is the sum of the reciprocated distances (i.e. undefined) so that infinite distances contribute a value of zero. (Harmonic closeness centrality), **used when graph is NOT strongly connected THIS IS CALLED "HARMONIC CLOSENESS CENTRALITY" IN GEPHI**

Avg of reversed distances (Valente-Foreman) the reversed distance is the diameter minus the geodesic distance.

*When a graph is not strongly connected, a widespread idea is that of using the sum of reciprocal of distances, instead of the reciprocal of the sum of distances, with the convention :why?

Betweenness Centrality

- Reflects how other actors control or mediate the relations between dyads that are not directly connected
- Views an actor as being in a favored position to the extent that the actor falls on the geodesic paths between other pairs of actors in the network. That is, the more people depend on me to make connections with other people, the more power I have.

Betweenness centrality

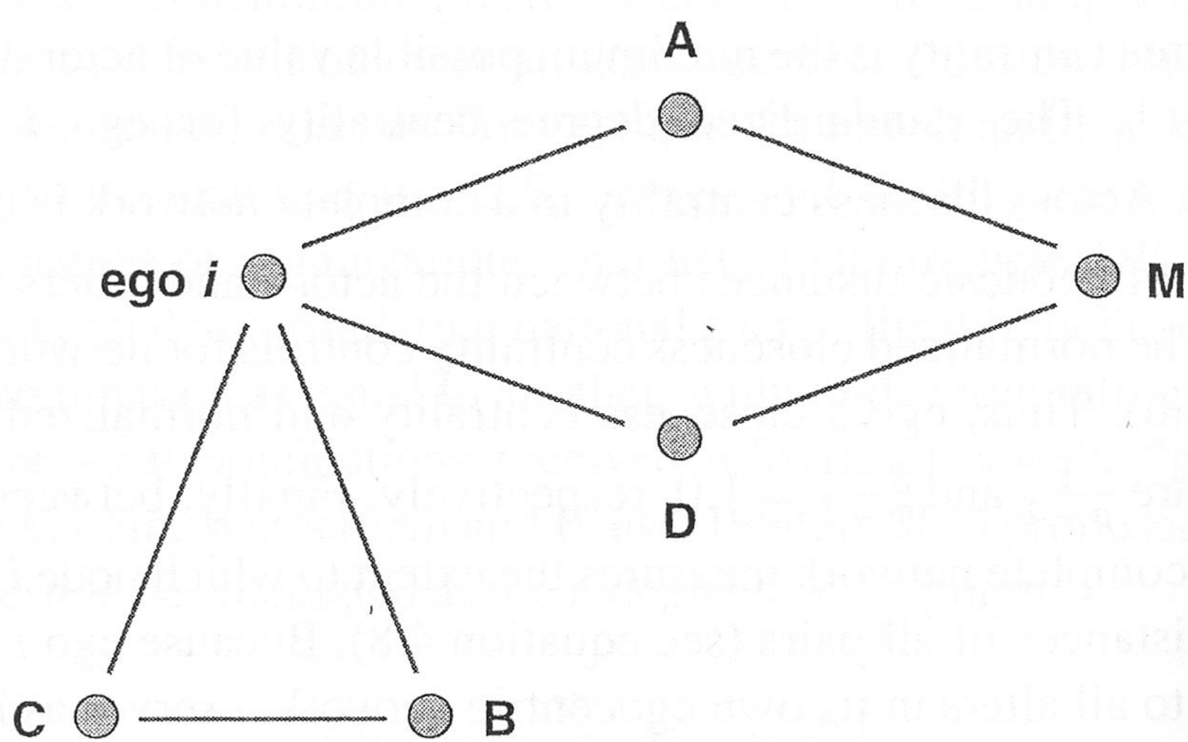


Figure 4.9 Betweenness Centrality in an Egocentric Network

Actor Betweenness Centrality

- Measures the extent to which other actors lie on the geodesic path (shortest distance) between pairs of actors in the network.

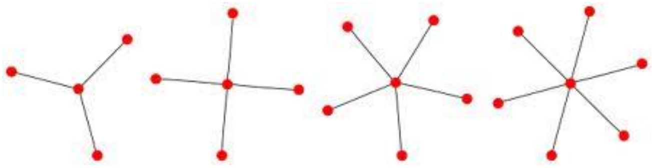
$$C_B = (N_i) = \sum_{j<k} \frac{g_{jk}(N_i)}{g_{jk}}$$

Sum the number of geodesic path between J and K that Contain node i

The number of geodesic path between J and K

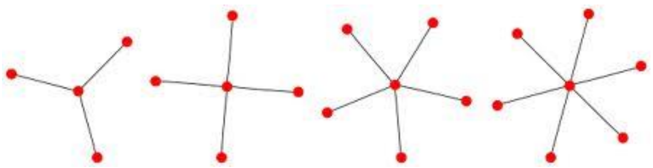
Normalized Actor Betweenness Centrality

- Normed by the maximum theoretical value of actor betweenness
- When node i falls on every geodesic path for all dyads, assuming each pair has only one geodesic path



Normalized Actor Betweenness Centrality

The normalized value varies from 0 to 1



$$C'_B(N_i) = \frac{C_B(N_i)X\boxed{2}}{\boxed{(g-1)(g-2)}}$$

Excluding node i, the total number of geodesic paths among the (g-1) nodes will be?
(g-1)(g-2)/2

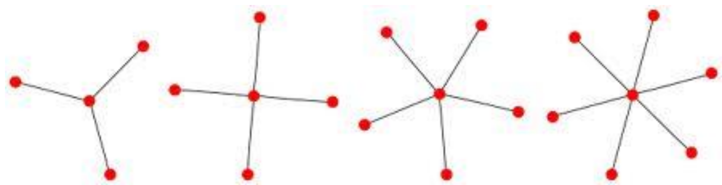
Index of Group Betweenness Centralization

- Reach 1 when a single dominant actor sits on all geodesic paths, reaches 0 when every node has the same betweenness centrality

Sum the differences in betweenness centrality for the actor with the highest value and every other actor

$$C_B = \frac{\sum_{i=1}^{g}[C_B(N^*) - C_B(N_i)]}{((g-1)^2(g-2))/2}$$

Maximum possible value of betweenness centralities for all nods in a network



Network>Centrality>Betweenness>Nodes

	1 Betweenness	2 nBetweenness
5	17.833	24.769
2	12.333	17.130
3	11.694	16.242
7	2.750	3.819
9	1.222	1.698
4	0.806	1.119
1	0.667	0.926
10	0.361	0.502
6	0.333	0.463
8	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 Betweenness	2 nBetweenness
1	Mean	4.800	6.667
2	Std Dev	6.220	8.639
3	Sum	48.000	66.667
4	Variance	38.689	74.632
5	SSQ	617.290	1190.760
6	MCSSQ	386.890	746.316
7	Euc Norm	24.845	34.507
8	Minimum	0.000	0.000
9	Maximum	17.833	24.769

Network Centralization Index = 20.11%

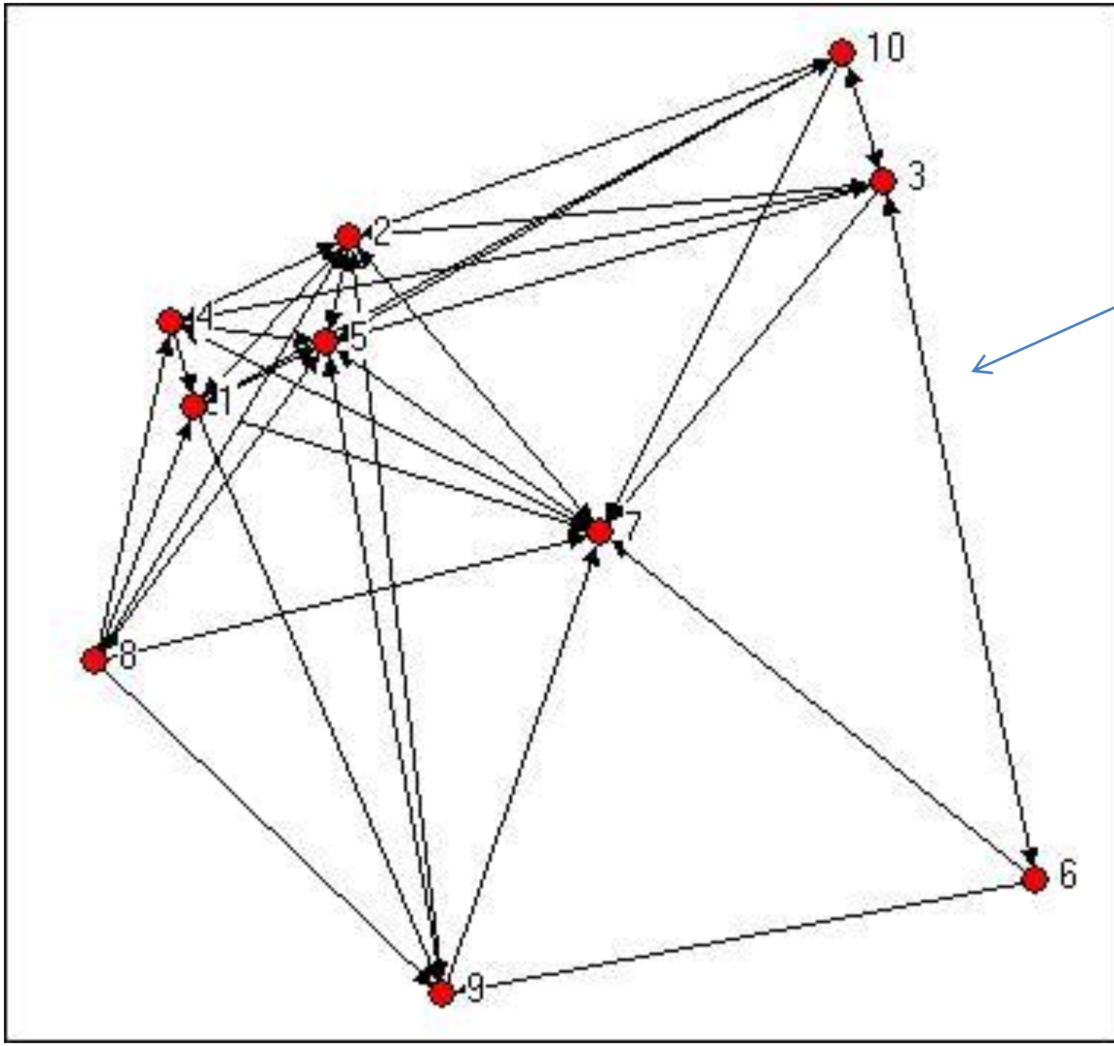
Kind of low, despite wide
Variation in betweenness
From 17.833 to 0
why?

Network>Centrality>Betweenness>Lines (edges)

Another way to think about betweenness is to ask which **relations** are most central, rather than which **actors**. Freeman's definition can be easily applied: a relation is between to the extent that it is part of the geodesic between pairs of actors.

Edge Betweenness

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	0.000	2.833	0.000	0.000	3.833	0.000	1.333	0.000	1.667	0.000
2	2.417	0.000	7.500	1.917	1.500	0.000	1.000	4.500	2.500	0.000
3	0.000	2.694	0.000	2.111	2.694	9.333	1.000	0.000	0.000	2.861
4	2.139	2.833	0.000	0.000	3.833	0.000	1.000	0.000	0.000	0.000
5	2.417	1.000	7.000	1.917	0.000	0.000	1.000	4.500	2.500	6.500
6	0.000	0.000	3.944	0.000	0.000	0.000	2.833	0.000	2.556	0.000
7	0.000	3.944	0.000	2.861	4.944	0.000	0.000	0.000	0.000	0.000
8	1.000	2.000	0.000	1.000	3.000	0.000	1.000	0.000	1.000	0.000
9	0.000	3.944	0.000	0.000	4.944	0.000	1.333	0.000	0.000	0.000
10	1.694	2.083	2.250	0.000	2.083	0.000	1.250	0.000	0.000	0.000



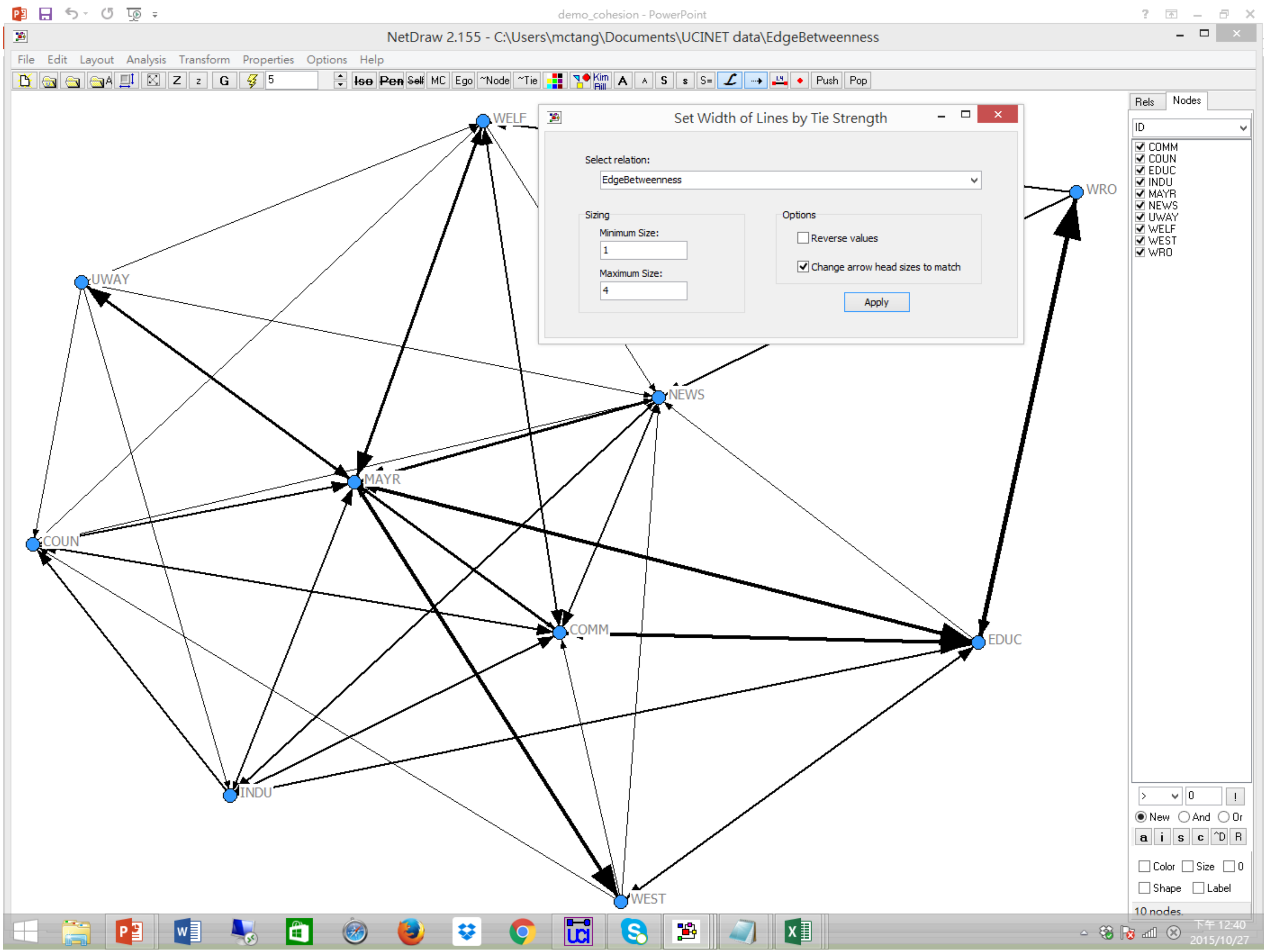
Betweenness between 3 and 6, 3.944
Relative high

This particular high value arises because without the tie to actor 3, actor 6 would be largely isolated.

Visualizing centrality

- Netdraw
  - Analysis>Centrality Measures
    - High tech (advise, friend, report-to)

Visualize line-betweenness



Betweenness: flow centrality

- Assume that actors will use all pathways (instead of the most efficient one only; blocked by a reluctant broker)
- Adds up how involved that actor is in all of the flows between all other pairs of actors (i.e. independent paths)
- Normalized by network size and density

Network>Centrality>Betweenness>Flow Betweenness

	1 FlowBet	2 nFlowBet
1	3.854	5.352
2	20.783	28.866
3	16.954	23.547
4	4.220	5.861
5	25.876	35.939
6	1.500	2.083
7	8.401	11.668
8	2.954	4.102
9	4.054	5.630
10	4.092	5.683

Network Centralization Index = 25.629%

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 FlowBet	2 nFlowBet
1	Mean	9.269	12.873
2	Std Dev	8.230	11.430
3	Sum	92.687	128.732
4	Variance	67.725	130.642
5	SSQ	1536.335	2963.609
6	MCSSQ	677.249	1306.421
7	Euc Norm	39.196	54.439
8	Minimum	1.500	2.083
9	Maximum	25.876	35.939

	1 Betweenness	2 nBetweenness
5	17.833	24.769
2	12.333	17.130
3	11.694	16.242
7	2.750	3.819
9	1.222	1.698
4	0.806	1.119
1	0.667	0.926
10	0.361	0.502
6	0.333	0.463
8	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 Betweenness	2 nBetweenness
1	Mean	4.800	6.667
2	Std Dev	6.220	8.639
3	Sum	48.000	66.667
4	Variance	38.689	74.632
5	SSQ	617.290	1190.760
6	MCSSQ	386.890	746.316
7	Euc Norm	24.845	34.507
8	Minimum	0.000	0.000
9	Maximum	17.833	24.769

Network Centralization Index = 20.11%

Actor eigenvector centrality

- Ties to prominent others gives you more power
- *Network>Centrality>Eigenvector* routine calculates individual actor centrality, and graph centralization using weights on the first eigenvector.
- The assumption is that each node's centrality is the sum of the centrality values of the nodes that it is connected to.

Review of centrality in undirected networks

Comparison

Comparing across these 3 centrality values

- Generally, the 3 centrality types will be positively correlated
- When they are not (low) correlated, it probably tells you something interesting about the network.

	Low Degree	Low Closeness	Low Betweenness
High Degree		Embedded in cluster that is far from the rest of the network	Ego's connections are redundant - communication bypasses him/her
High Closeness	Key player tied to important important/active alters		Probably multiple paths in the network, ego is near many people, but so are many others
High Betweenness	Ego's few ties are crucial for network flow	Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others. As in a hierarchy? (unidirectional)	

slide: Jim Moody