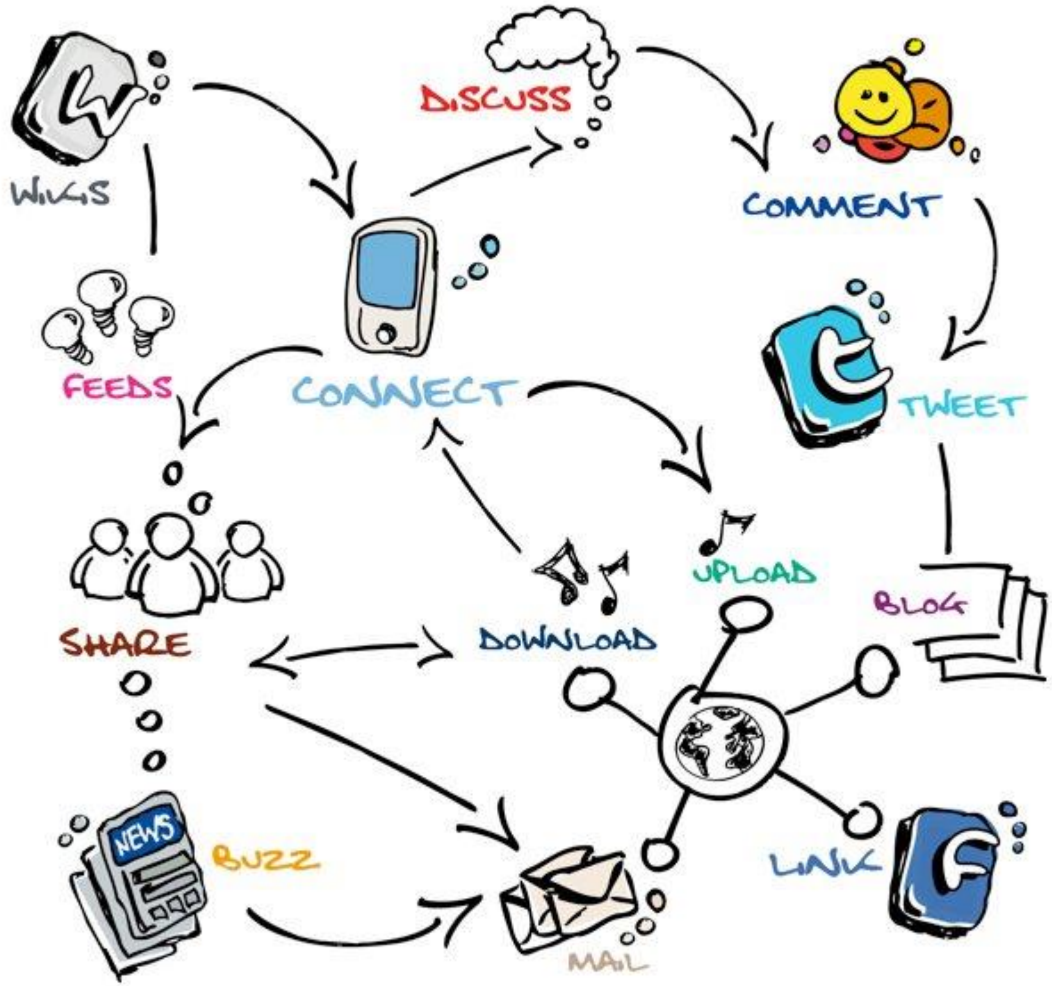
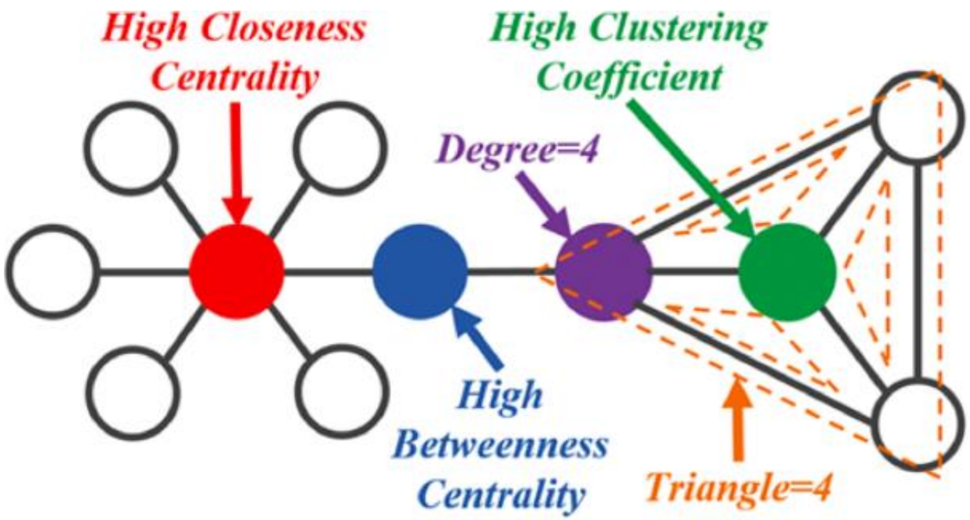


Centrality Measures

Ahmet Onur Durahim

Centrality

- ***Centrality*** is used often for detecting:
  - How ***influential*** a person is in a social network?
  - How ***well used*** a road is in a transportation network?
  - How ***important*** a web page is?

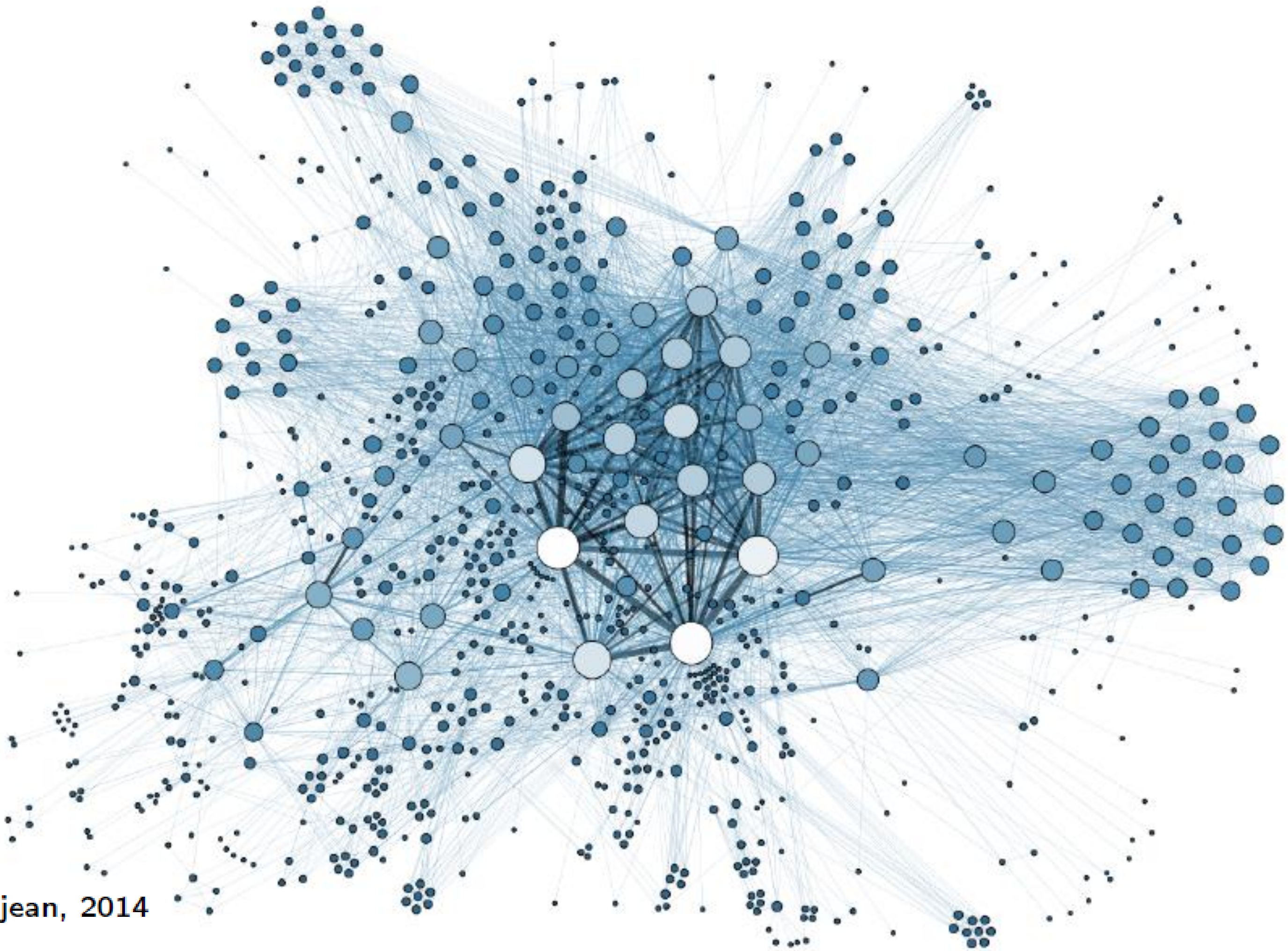


Centrality

- Finding out the ***most important or central node*** in the network is important:
  - It could help ***disseminating information*** in the network faster
  - It could help ***stopping epidemics***
  - It could help ***protecting the network from breaking***

Graph Theoretic Measures

Which vertices are important?

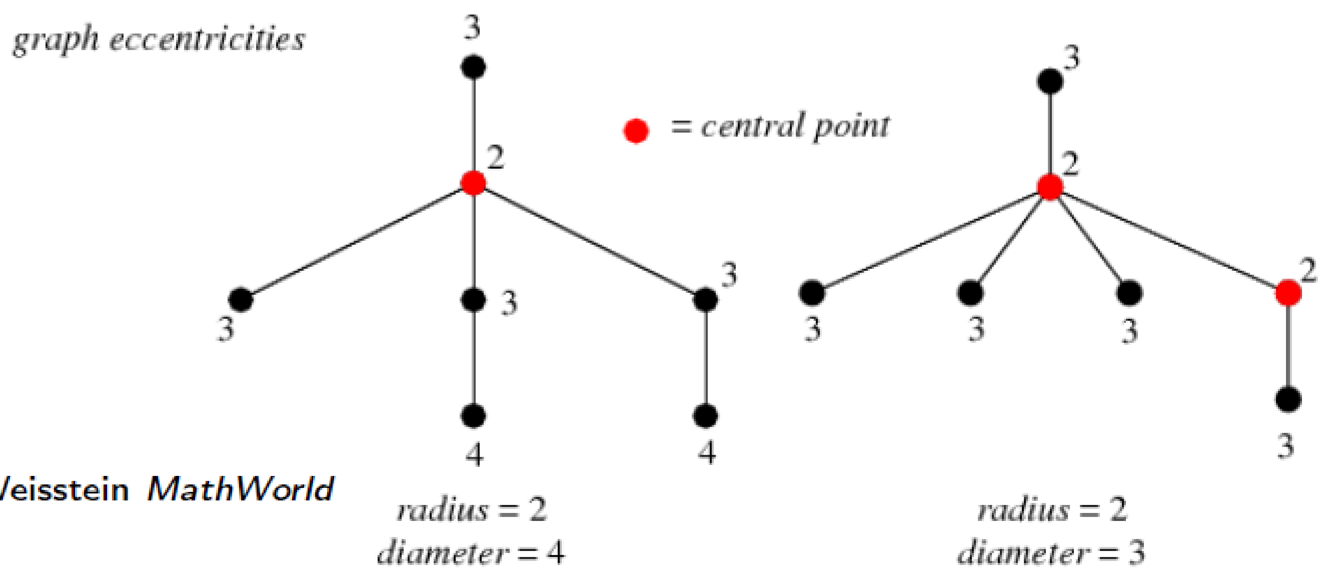


M.Grandjean, 2014

Graph-theoretic Measures

- The ***eccentricity*** ϵ(v) of a vertex v is the *maximum distance* between v and any other vertex u of the graph $\epsilon(v) = \max_{u \in V} d(u, v)$
  - Graph ***diameter*** is the *maximum eccentricity* $d = \max_{v \in V} \epsilon(v)$
  - Graph ***radius*** is the *minimum eccentricity* $r = \min_{v \in V} \epsilon(v)$
  - A point *v* is a ***central point/vertex*** of a graph if the eccentricity of the point equals the graph radius $\epsilon(v) = r$

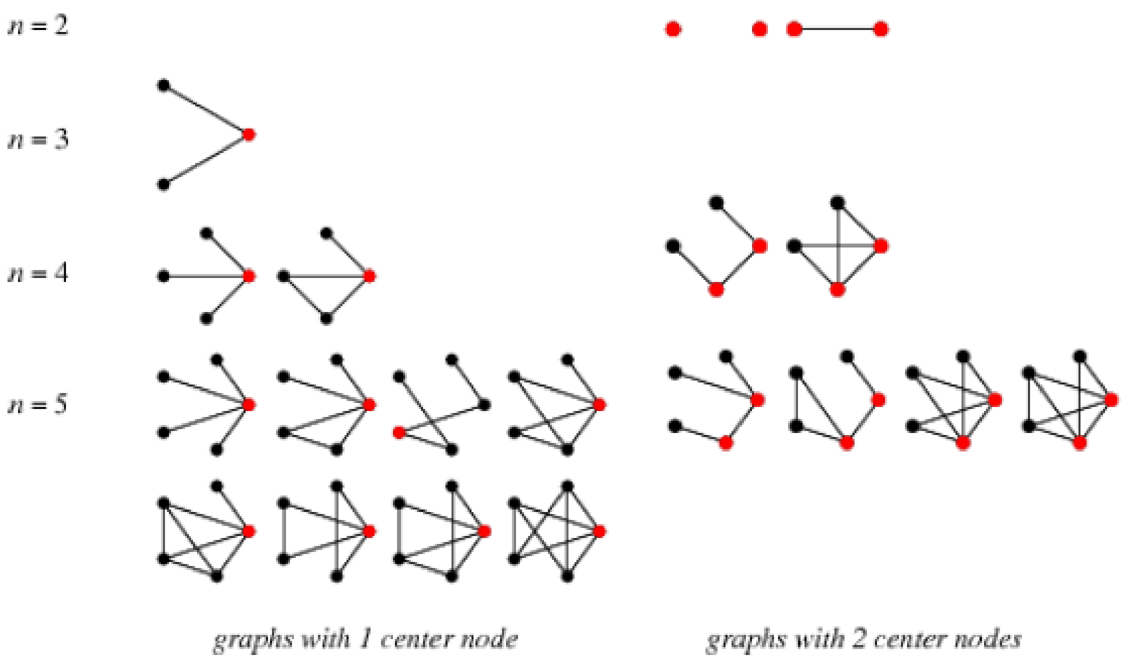
*A **peripheral vertex** of a graph has the eccentricity value equal to the **graph diameter***



from Eric Weisstein *MathWorld*

Graph-theoretic Measures

- Graph ***center*** is a set of of vertices with graph eccentricity equal to the graph radius *ϵ(v) = r*
  - set of central points
- Graph ***periphery*** is a set of vertices that have graph eccentricities equal to the graph diameter *ϵ(v) = d*



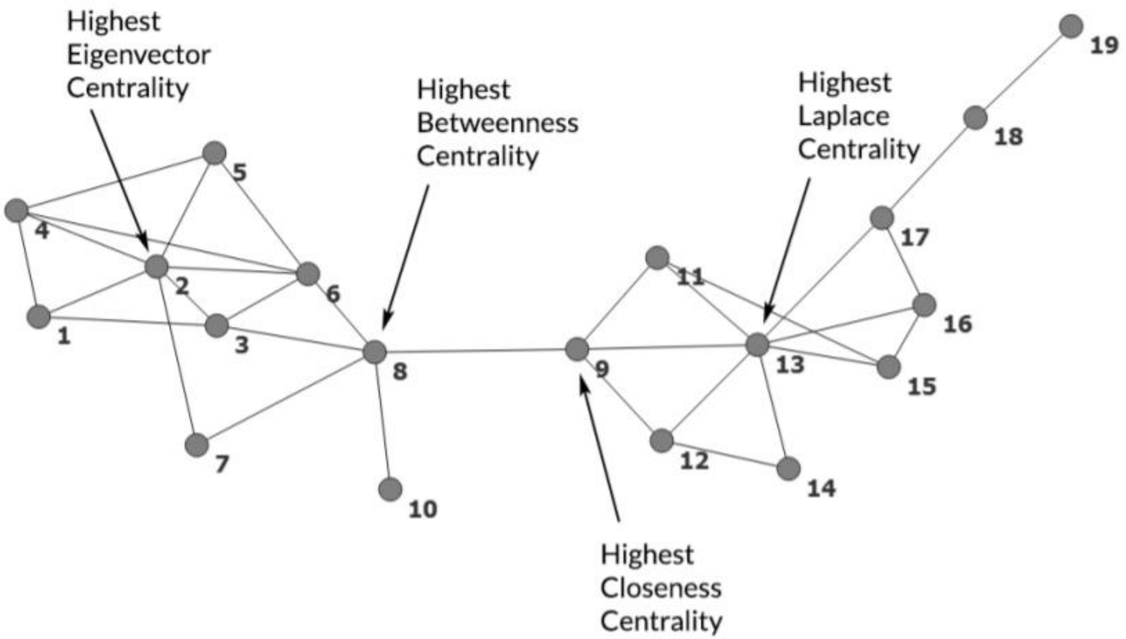
from Eric Weisstein *MathWorld*

Problems with Eccentricity related Metrics

- ***Computational cost***
  - need to calculate all the shortest paths from all the nodes to other nodes
- Diameter/radius and other *metrics* are very ***sensitive to small changes*** in the network
  - Adding 1 more node will change everything
  - use them with caution because any small changes in the graph changes these metrics dramatically

Centrality Measures

- ***Geometric Measures***
  - Importance is a **function of distances** to other nodes
- ***Spectral Measures***
  - Based on the **eigen-structure** of some graph-related matrix
- ***Path-based Measures***
  - Take into account **all (shortest) paths** coming into a node



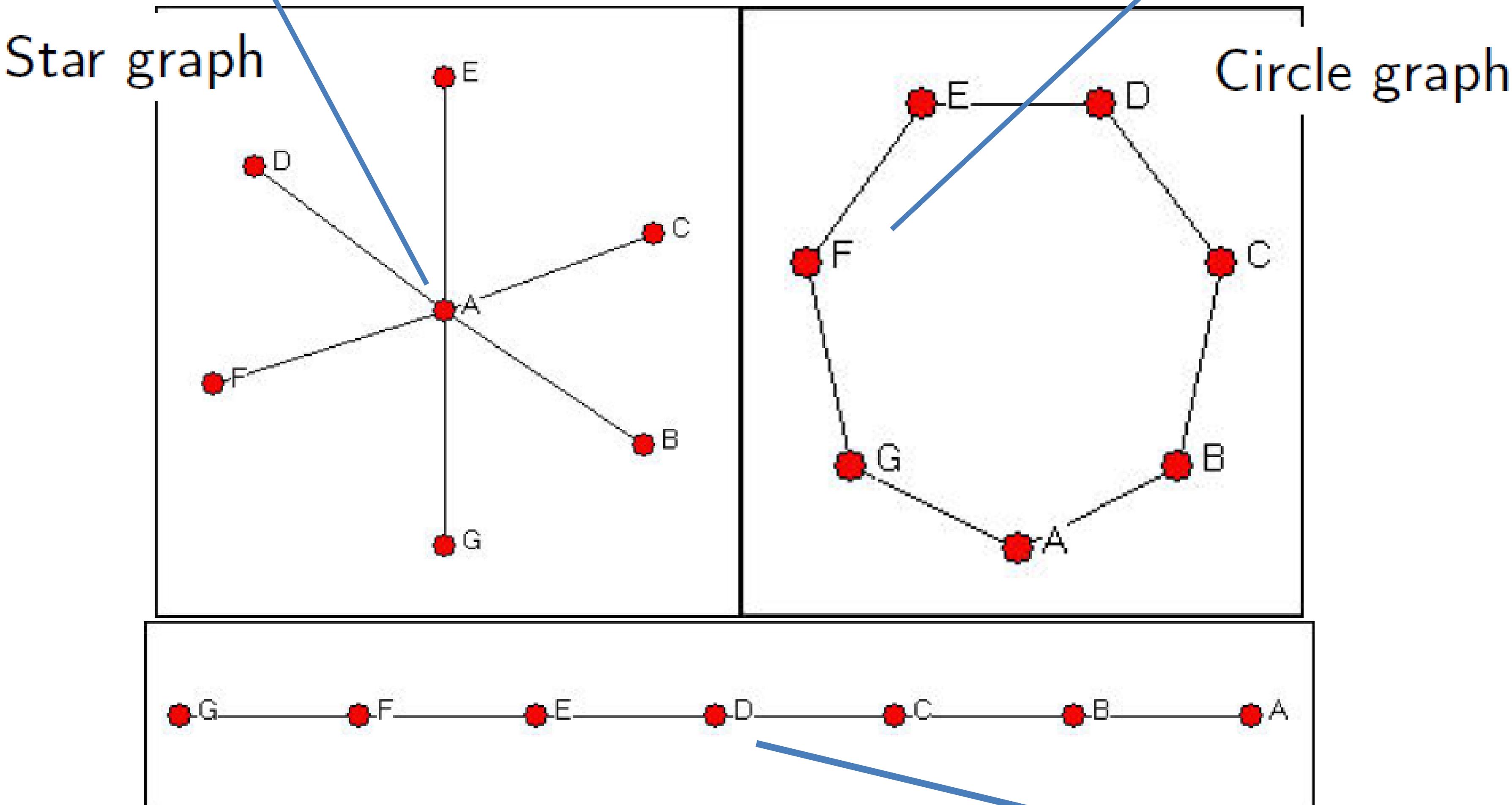
Centrality Measures
Sociology Notion of Node Importance

- Most ***"important"*** actors: actor location in the social network (*Linton Freeman, 1979*)
  - importance in terms of ***propagating information***, ***influencing*** other people, etc.
  - ***Actor centrality*** - involvement with other actors, many ties, source or recipient
    - *Centrality* refers to ***Undirected networks***
  - ***Actor prestige*** - recipient (object) of many ties, ties directed to an actor (in-links)
    - *Prestige* in ***Directed networks***
    - How many citations you get ⇔ prestige of your paper
- Three graphs:
  - *Star* graph : *Circle* graph : *Line* graph

Three Graphs

Node A is the ***central node***
Others are the ***peripheries***

All nodes are ***equivalent***
Graph is ***symmetric***



Line Graph

nodes ***closer to the center*** (node D then nodes C&E) are ***more central nodes***

Degree Centrality

- ***Degree centrality***: number of nearest neighbours

$$C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$$

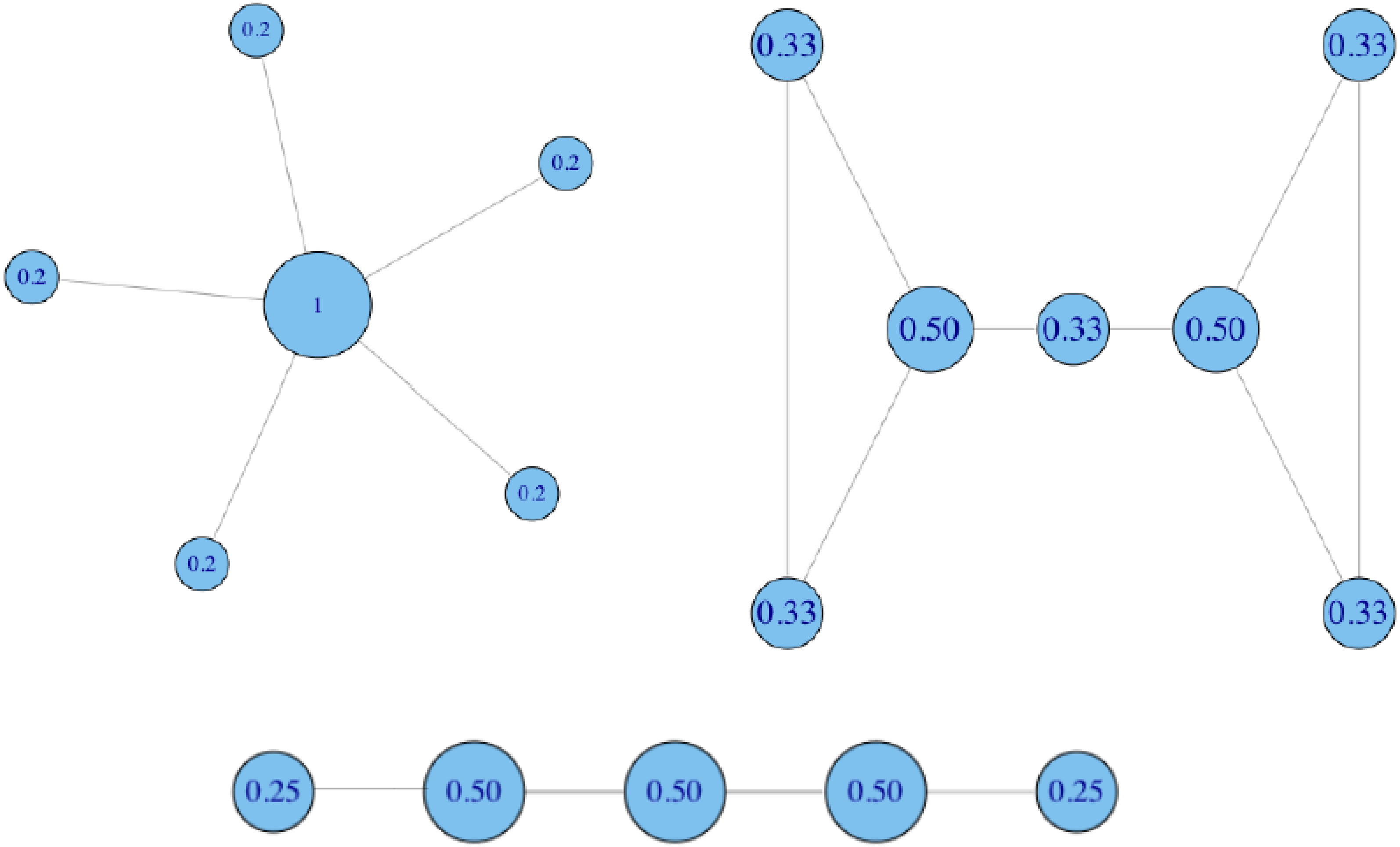
- ***Normalized degree centrality***

$$C_D^*(i) = \frac{1}{n-1} C_D(i)$$

If we want to compare the centralities of two nodes in two different networks we need a comparable metric
=> the ***normalized values*** of different centrality metrics

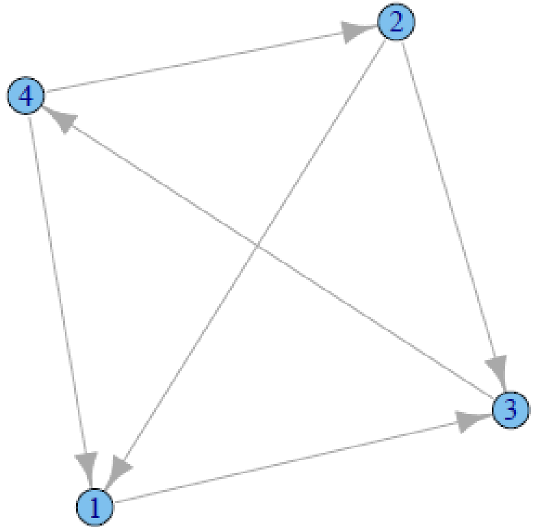
- ***High*** centrality degree
  - direct contact with many other actors
- ***Low*** degree
  - not active, peripheral actor

Normalized Degree Centrality



Outdegree centrality and indegree prestige for digraph

- The nodes with ***higher outdegree*** is ***more central*** (choices made)
- The nodes with ***higher indegree*** is ***more prestigious*** (choices received)



Node	1	2	3	4
Outdegree	1	2	1	2
Indegree	2	1	2	1

Closeness Centrality

- ***Closeness centrality***: how close an actor to all the other actors in network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

- ***Normalized closeness centrality***

$$C_C^*(i) = (n-1)C_C(i)$$

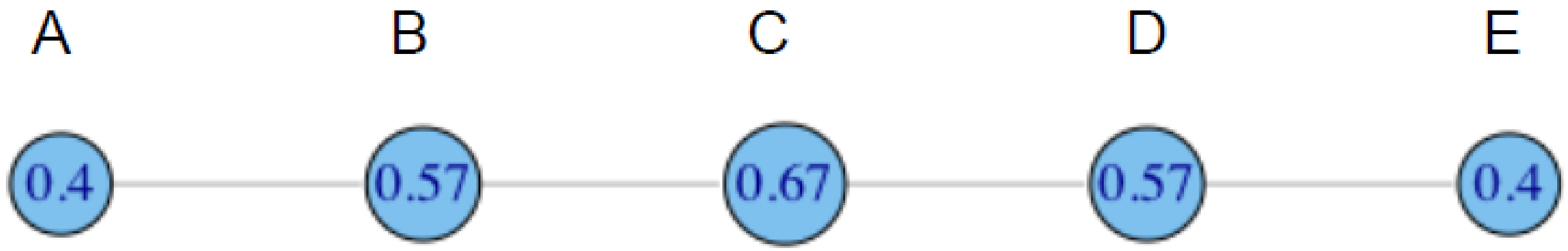
Problem: What happens if the graph is ***not connected*** ?

- Actor in the center can ***quickly interact with all others***, short communication path to others, minimal number of steps to reach others

[*** Harmonic centrality $= \sum_j \frac{1}{d(i,j)}$ ***] If any distance is infinite then it does not have an impact on Harmonic cen.

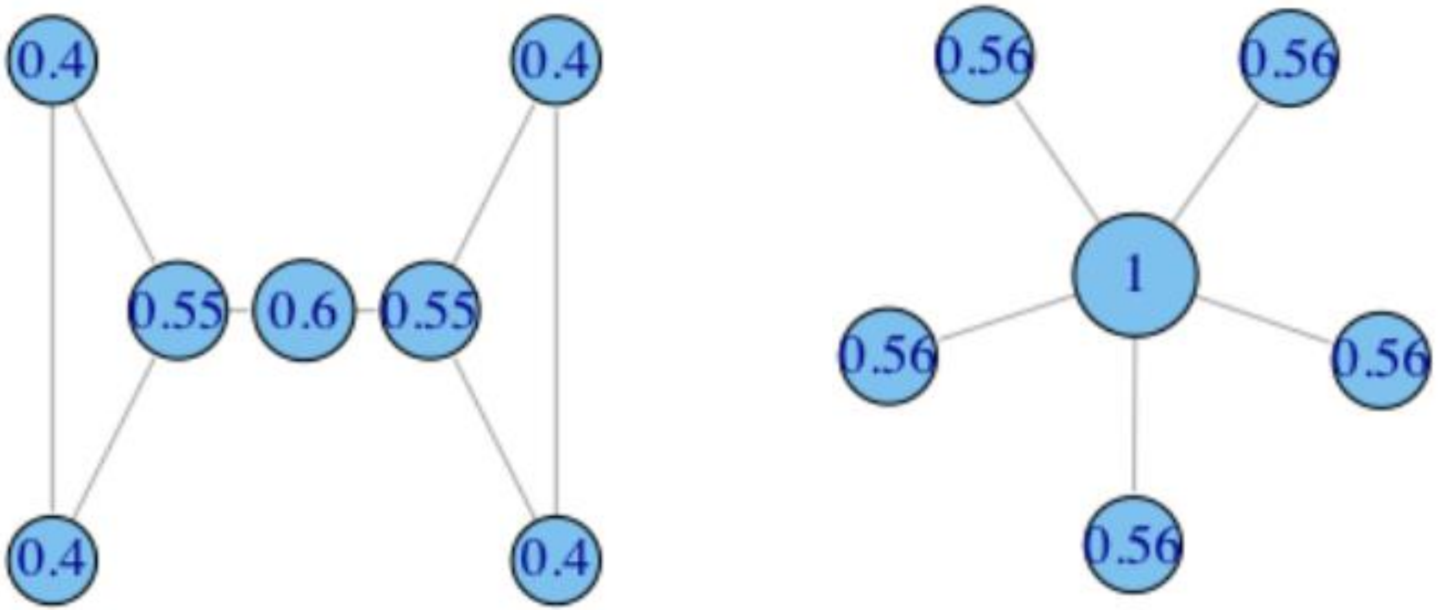
Alex Bavelas, 1948

Closeness Centrality



$$C_c'(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$

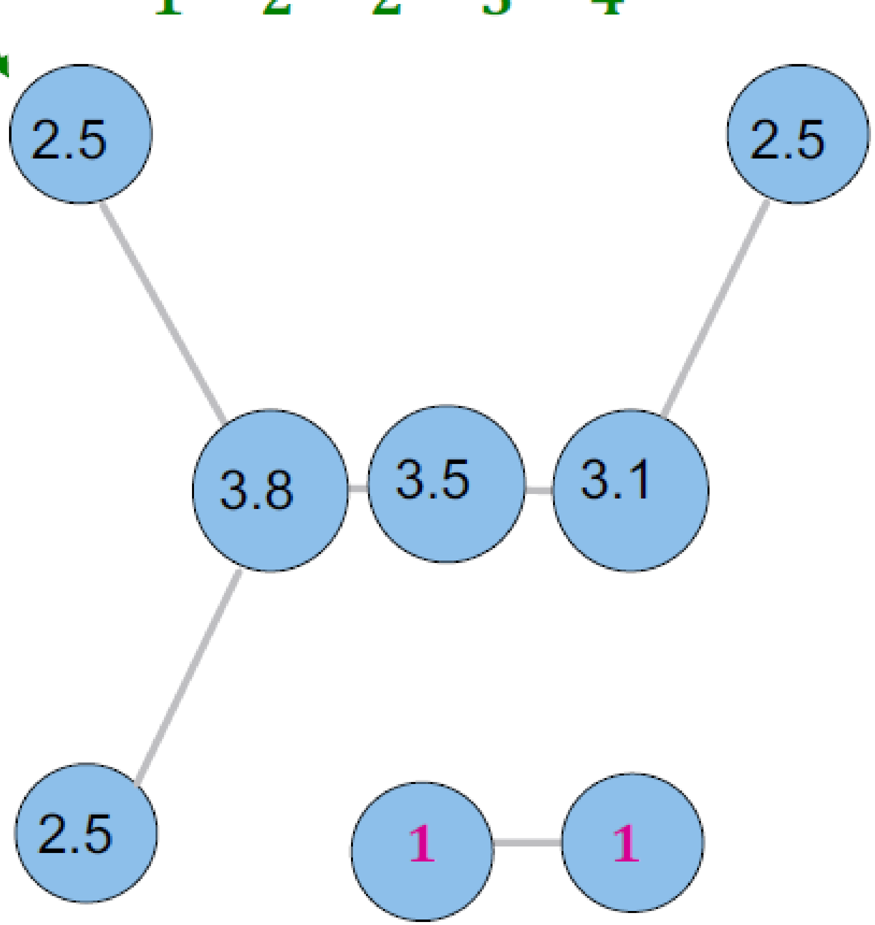
Closeness Centrality



(1+1+2+3+4+4/6)^-1=6/15=0.4

Harmonic Centrality

$$c_{harm} = \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = 2.5$$



Betweenness Centrality

- Betweenness centrality ***quantifies the number of times a node acts as a bridge along the shortest path*** between two other nodes
- It was introduced as a measure for quantifying the *control of a human on the communication between other humans* in a social network by Linton Freeman
- Nodes that have a ***high probability to occur*** on a randomly chosen shortest path between two randomly chosen nodes have a ***high betweenness***
- The ***betweenness centrality*** for node v is
  - the ***probability*** that a shortest path passes through v

Betweenness Centrality

- The ***betweenness of a vertex v*** in a graph G = (V, E) is computed as:
  - For each pair of vertices (s, t)
    - compute the shortest paths between them
  - For each pair of vertices (s, t)
    - determine the fraction of shortest paths that pass through the vertex in question (v)
  - Sum this fraction over all pairs of vertices (s, t)

Betweenness Centrality

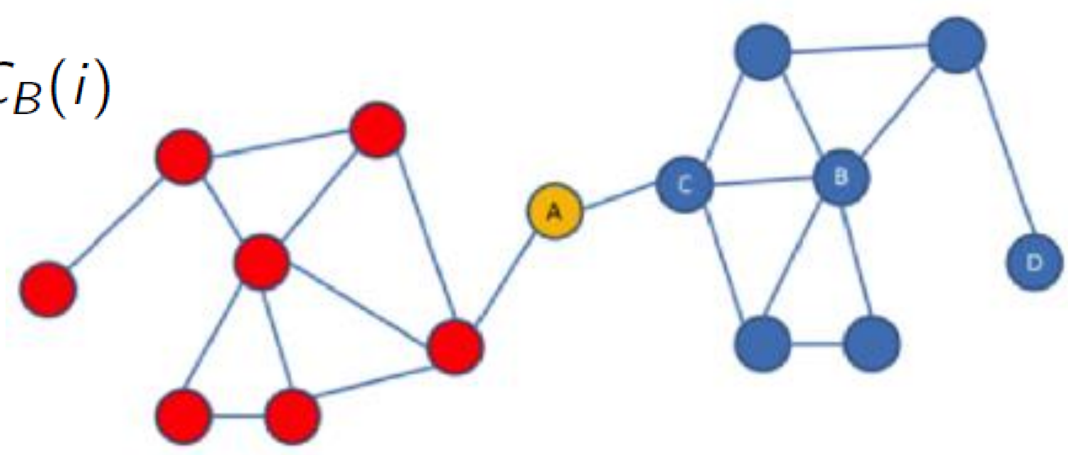
- ***Betweenness centrality***: number of shortest paths going through the actor $\sigma_{st}(i)$

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

- ***Normalized betweenness centrality***

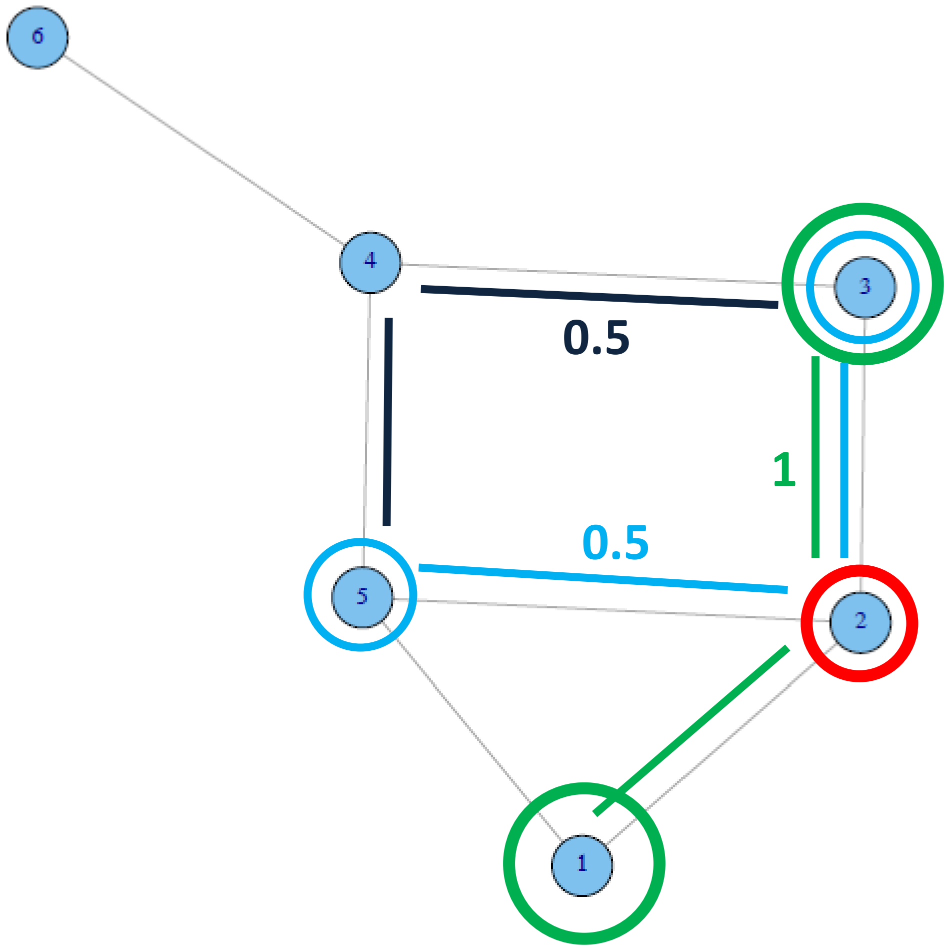
$$C_B^*(i) = \boxed{\frac{2}{(n-1)(n-2)}} C_B(i)$$

- n-1 ways to select each node other than the central node
- n-2 shortest paths going through the central node
  → (n-1)*(n-2)
- don't double count (on undirected networks)
  → (n-1)*(n-2) / 2



- Probability that a communication from s to t will go through i (geodesics) **Linton Freeman, 1977**

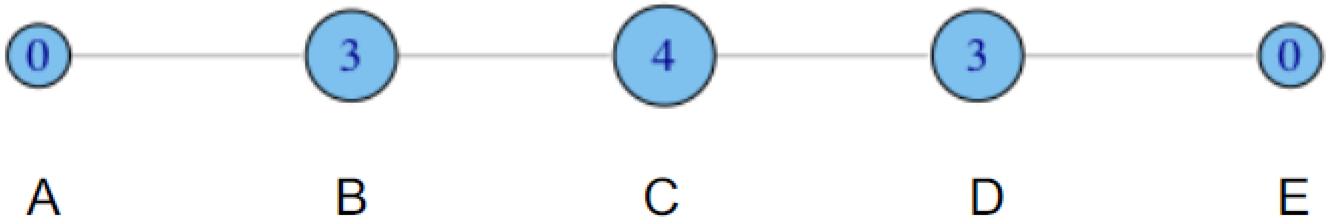
Betweenness Centrality



Node	(Unnormalized) Betweenness Centrality
1	0
2	1.5
3	1
4	4.5
5	3
6	0

Betweenness Centrality Example

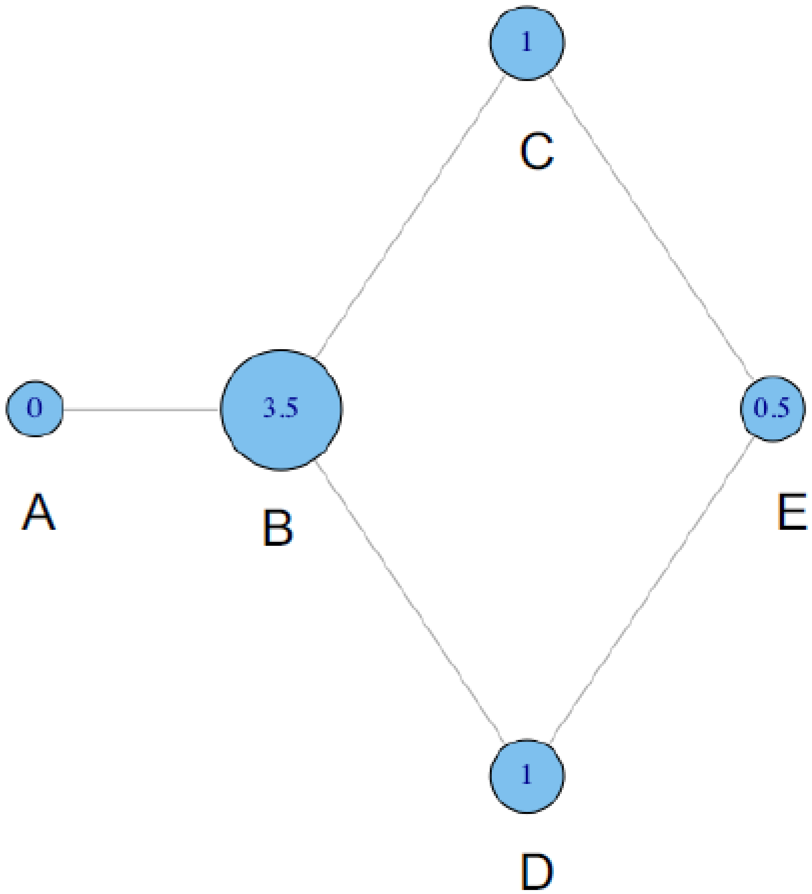
- Removing nodes in betweenness order causes a very quick disruption of the network



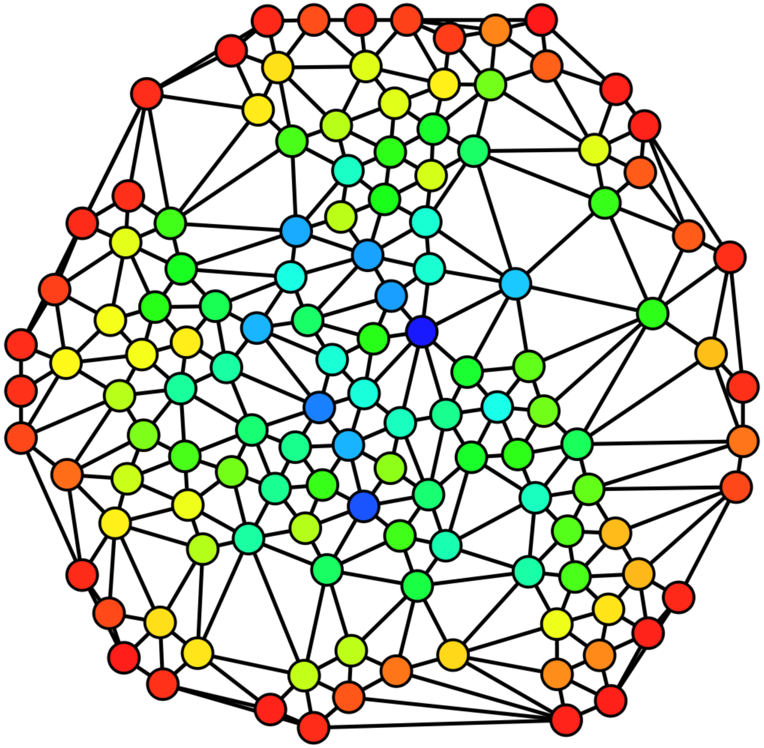
- A & E lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices
  - (A,D) - (A,E) - (B,D) - (B,E)
  - there are no alternate paths for these pairs to take, so C gets full credit

Betweenness Centrality Example

- Why do C and D each have betweenness 1?
  - They are both on shortest paths for pairs (A,E), and (B,E), and so must share credit:
    - ½+½ = 1
- Can you figure out why B has betweenness 3.5 while E has betweenness 0.5?



Betweenness Centrality

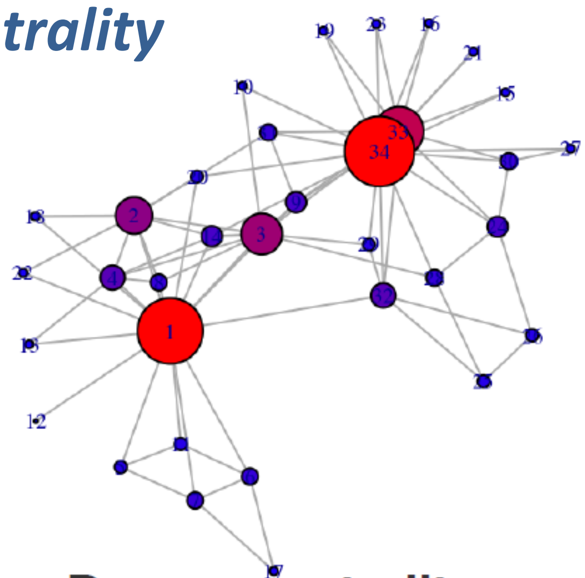


Hue (from red = 0 to blue = max) shows the node betweenness.

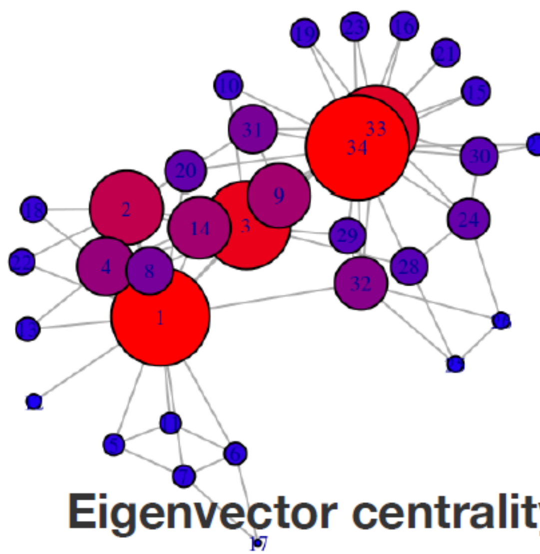
https://en.wikipedia.org/wiki/Centrality

Spectral Measures of Centrality

- Compute the left dominant ***eigenvector*** of some matrix derived from the graph
- ***Idea:*** A node's centrality is a function of the centrality of its neighbors
  - Nodes connected to central nodes has a larger centrality score than those connected to non-central nodes
  - ***Eigenvector Centrality***
  - ***Katz's Index***
  - ***Page Rank***
  - ***Hits***



Degree centrality



Eigenvector centrality

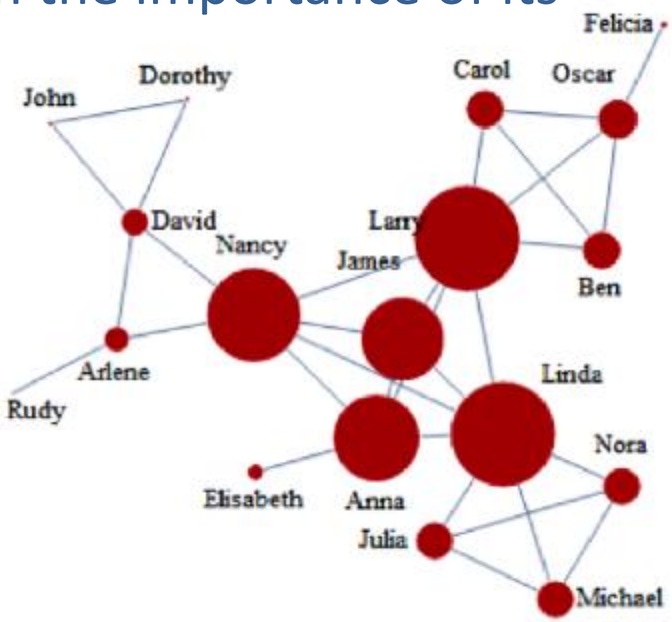
Eigenvector Centrality

- Importance of a node depends on the importance of its neighbors (recursive definition)

$$v_i \leftarrow \sum_j A_{ij} v_j$$

$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$



- Select an eigenvector associated with largest eigenvalue $\lambda = \lambda_1$, $v = v_1$
- The eigenvector centrality defined in this way depends both on the ***number of neighbors |N(i)|*** and the ***quality of its connections***

Phillip Bonacich, 1972

Problem: Graph should be *strongly connected*

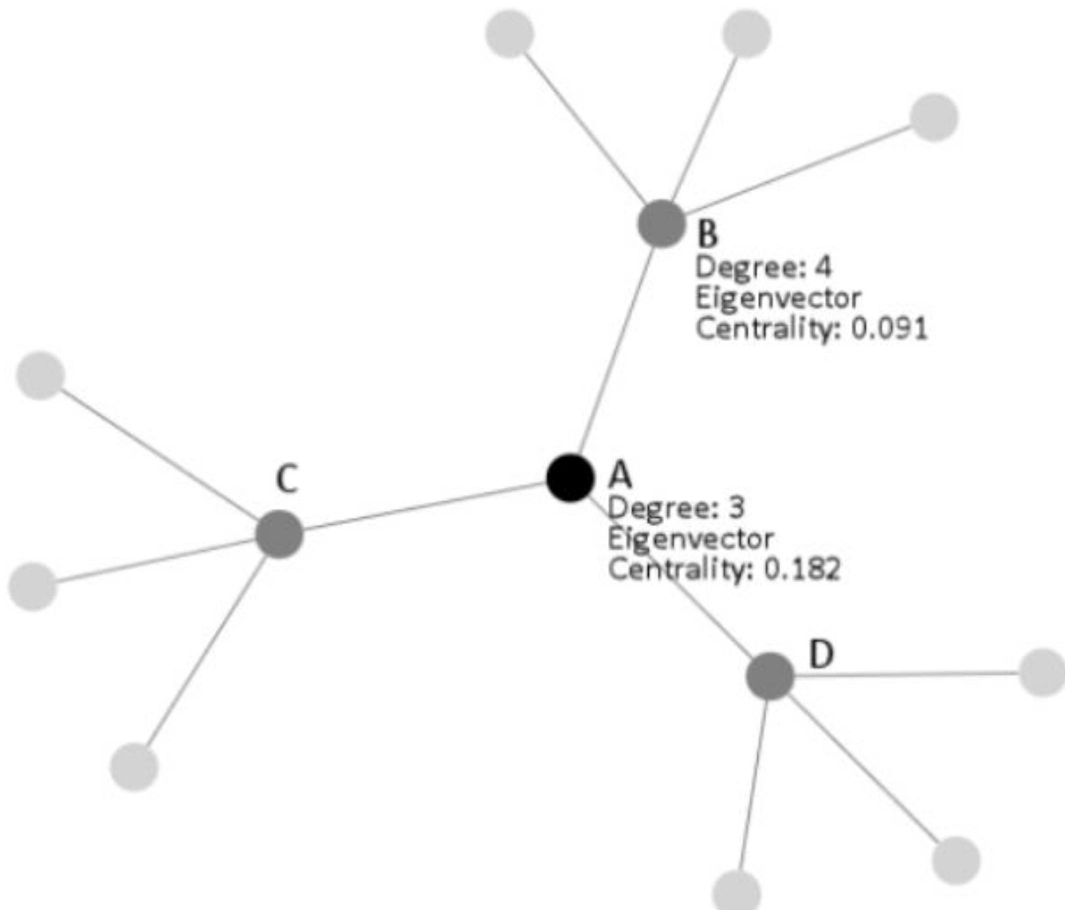
Eigenvector Centrality – Algorithm

1) Start by assigning some centrality score to all nodes
   - e.g., $v_i = 1$ for all i in the network
2) Recompute scores of each node as weighted sum of centralities of all nodes in a node's neighborhood:
   - $v_i = \sum_{j \in N} a_{ij} * v_j$
3) Normalize v by dividing each value by the largest value
4) Repeat steps 2 and 3 until values of v stops changing

Intuitively:

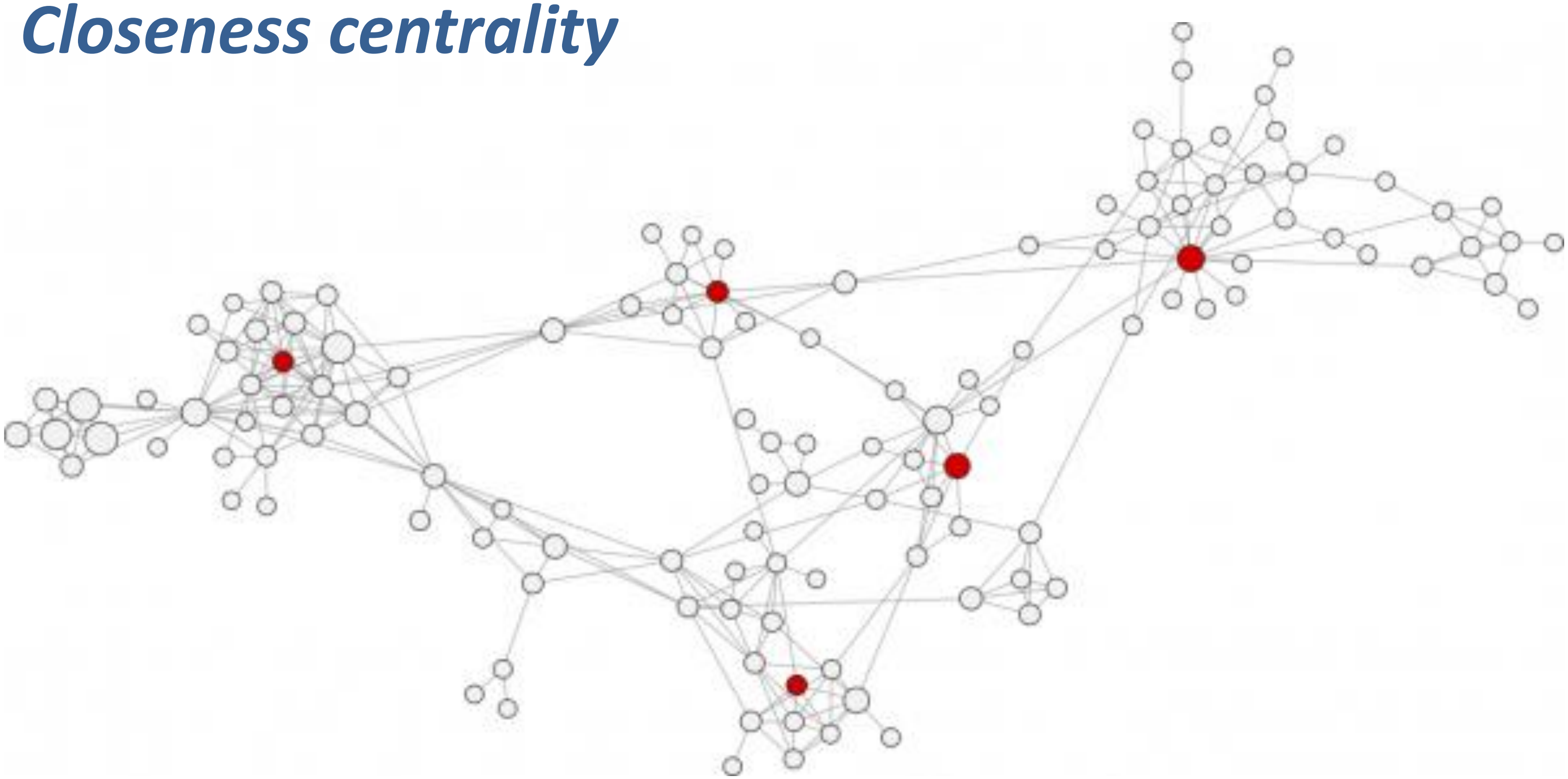
- ***Degree*** counts walks of length one
- ***Eigenvalue Centrality*** counts walks of length infinity

Eigenvector Centrality



Centrality Examples

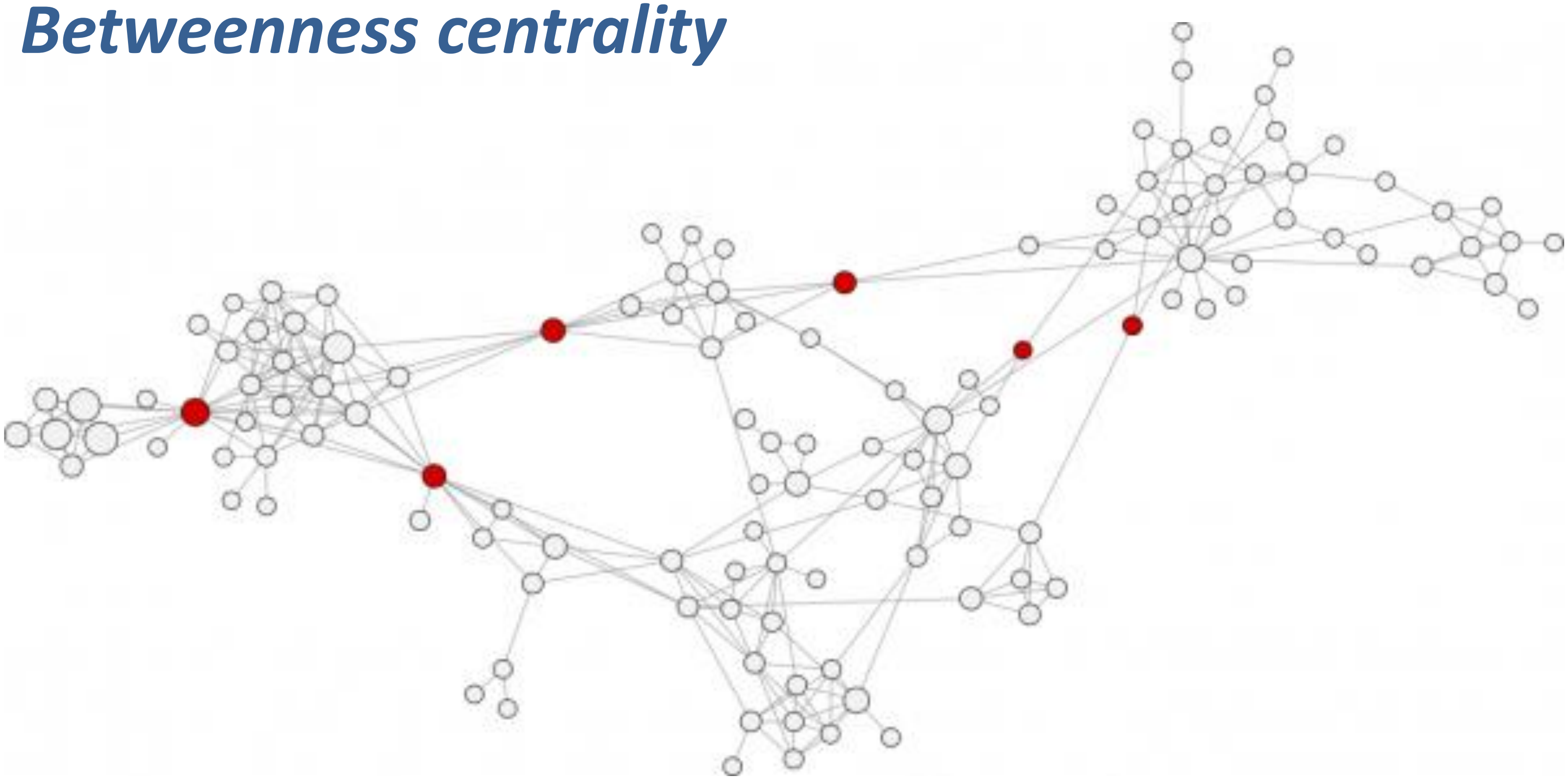
Closeness centrality



These are the nodes that are very well connected with their neighbors

Centrality Examples

Betweenness centrality



Nodes that act as ***bridges***

If you consider the ***information transfer*** over/in the network, these nodes are the key players information goes through them

Centrality Examples

Eigenvector centrality

If a node has high EVC then closest neighbors also has high EVC

Katz status index

- Measures influence by considering the total number of walks between a pair of nodes
- Weighted count of all paths coming to the node
  - the weight of path of length n is counted with ***attenuation factor*** $\beta^n, \beta < \frac{1}{\lambda_1}$

$$k_i = \beta \sum_j A_{ij} + \beta^2 \sum_j A_{ij}^2 + \beta^3 \sum_j A_{ij}^3 + \ldots$$

$$\mathbf{k} = (\beta \mathbf{A} + \beta^2 \mathbf{A}^2 + \beta^3 \mathbf{A}^3 + \ldots)\mathbf{e} = \sum_{n=1}^{\infty} (\beta^n \mathbf{A}^n)\mathbf{e} = (\sum_{n=0}^{\infty} (\beta \mathbf{A})^n - \mathbf{I})\mathbf{e}$$

$$\sum_{n=0}^{\infty} (\beta \mathbf{A})^n = (\mathbf{I} - \beta \mathbf{A})^{-1}$$

$$c_{\text{katz}}(x) = \boldsymbol{\beta} \sum_{k=0}^{\infty} \sum_{x \to y} \alpha^k \left(A^k\right)_{xy}$$

Total number of walks of length k between nodes x and y

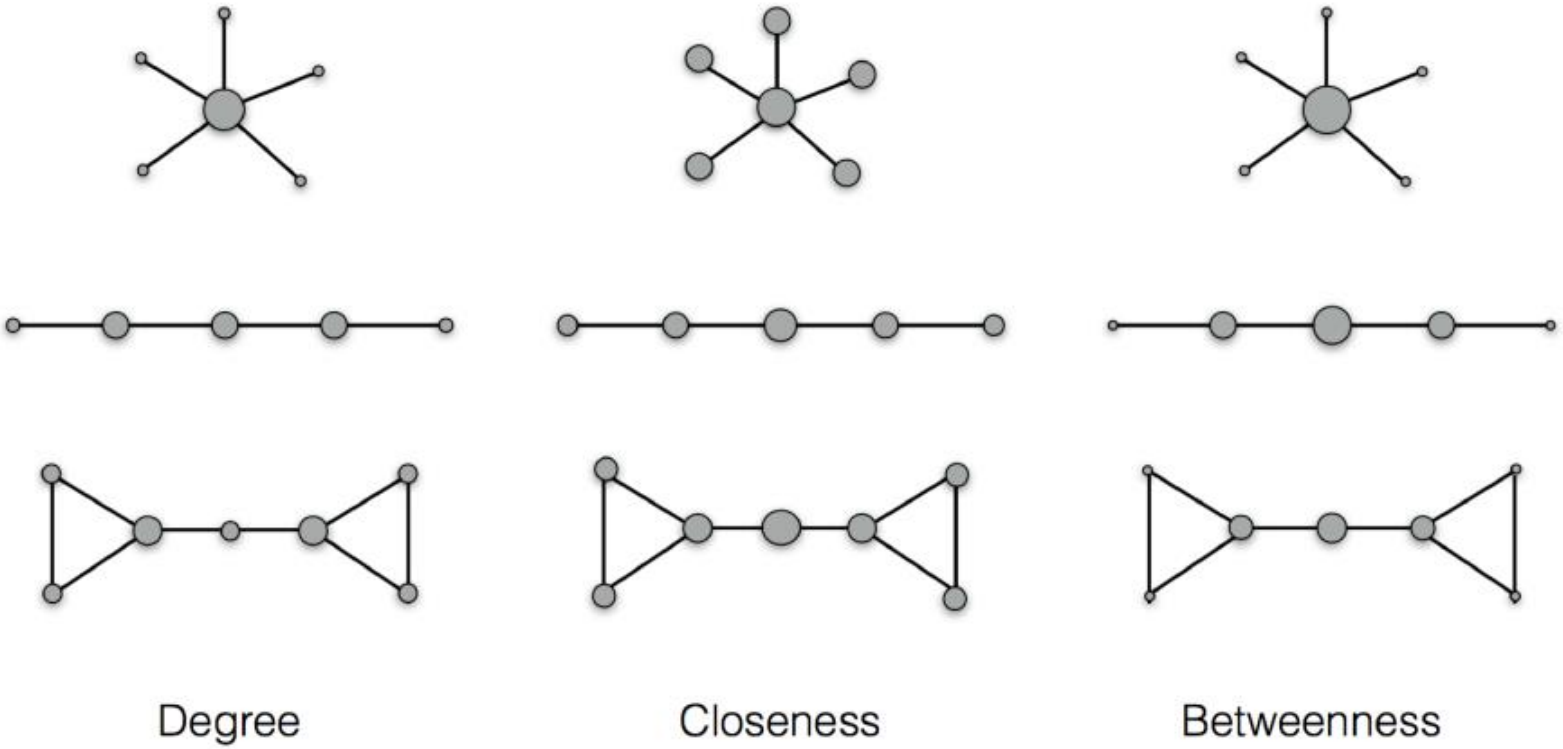
$$\mathbf{k} = ((\mathbf{I} - \beta \mathbf{A})^{-1} - \mathbf{I})\mathbf{e}$$

$$(\mathbf{I} - \beta \mathbf{A})\mathbf{k} = \beta \mathbf{A} \mathbf{e}$$

Leo Katz, 1953

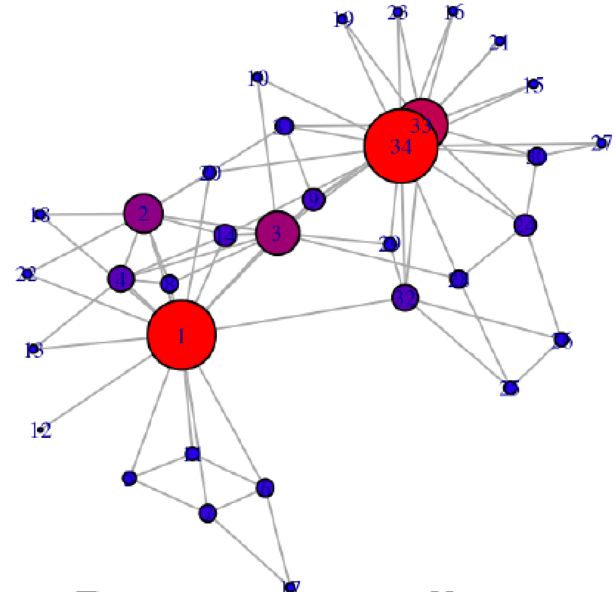
$$\mathbf{k} = \beta (\mathbf{I} - \beta \mathbf{A})^{-1} \mathbf{A} \mathbf{e}$$

Comparing Centrality Measures

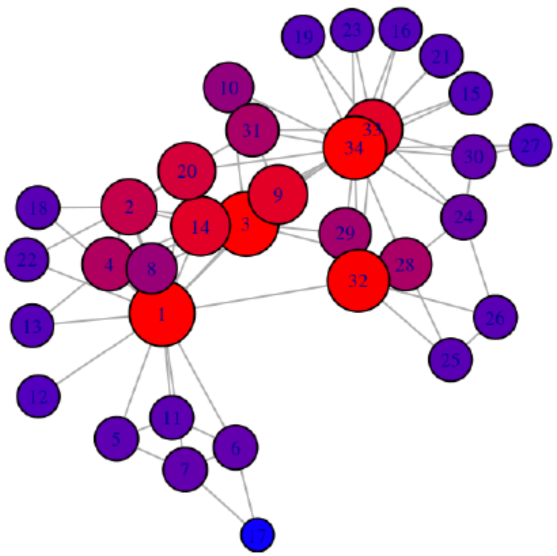


They are clearly related, but they each get at something slightly different

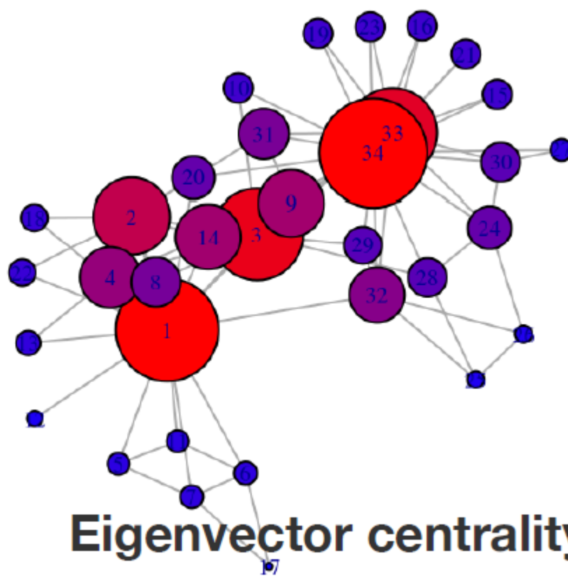
Comparing Centrality Measures



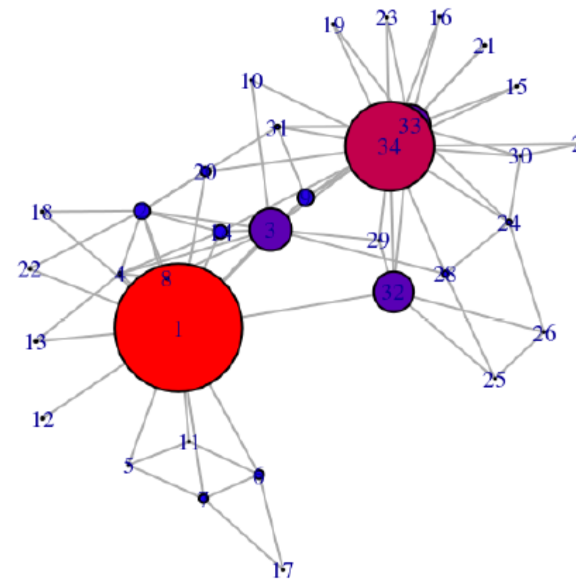
Degree centrality



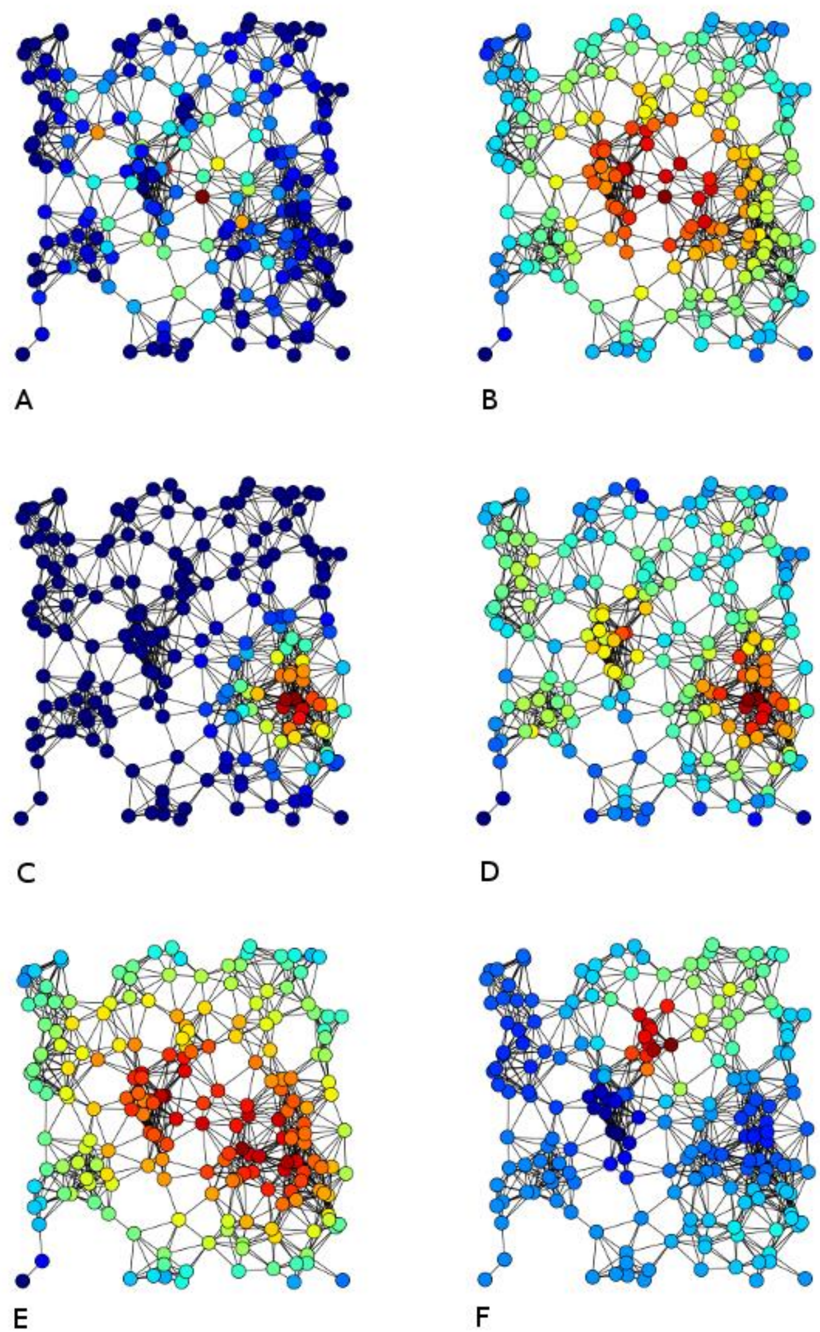
Closeness centrality



Eigenvector centrality



Betwenness centrality



Examples of

A) Betweenness centrality

B) Closeness centrality

C) Eigenvector centrality

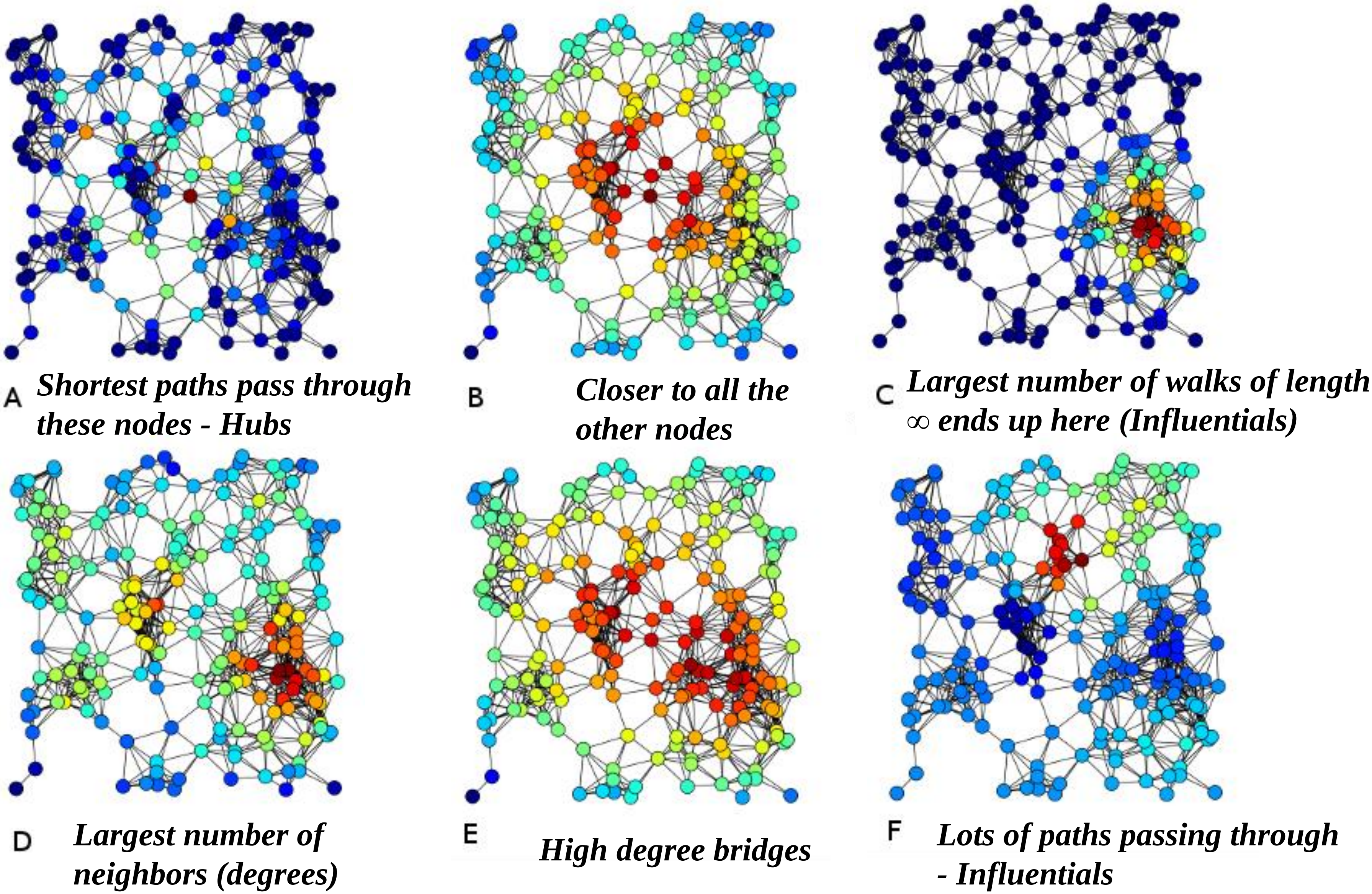
D) Degree centrality

E) Harmonic centrality

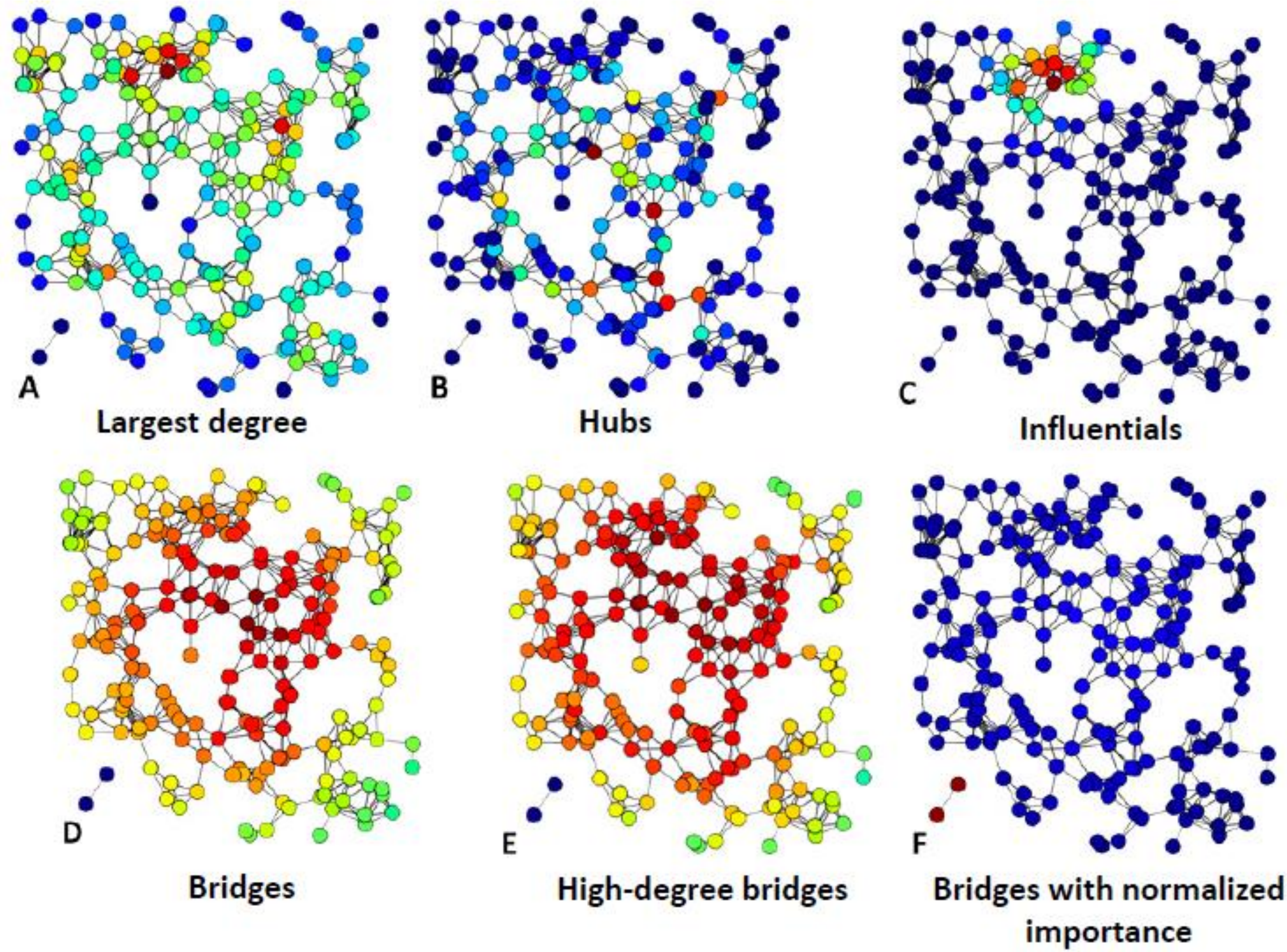
F) Katz centrality

of the same graph

- **Bridges**
- **Geometric Centers**
- **Self-Reinforced Centralities**



Examples of A) Betweenness centrality, B) Closeness centrality, C) Eigenvector centrality, D) Degree centrality, E) Harmonic centrality, F) Katz centrality **of the same graph**



A) Degree, B) Betweenness, C) Eigenvector,

D) Closeness (not normalized) E) Harmonic Centrality, F) Closeness (normalized)

Comparison among centrality measures for the Padgett Florentine families

- Business ties network of the Padgett Florentine families
- The top three ranks by different methods are summarized as follows:

Rank	Degree	Closeness	Betweenness	Eigenvector	PageRank
1	MEDICI	MEDICI	MEDICI	MEDICI	MEDICI
2	GUADAGNI	RIDOLFI	GUADAGNI	STROZZI	GUADAGNI
3	STROZZI	ALBIZZI	ALBIZZI	RIDOLFI	STROZZI

- Deciding which are most appropriate for a given application clearly requires consideration of the context

Comparison of Betw/Cls/Degree Cent.

- 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/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. ego monopolizes the ties from a small number of people to many others.	

Wasserman & Faust, Social Network Analysis Cambridge Univ. Press, 1994: pp730

- "..., we do not expect that the most fruitful development in descriptive techniques will be the continued addition of yet another definition of centrality measure or yet another subgroup definition or yet another definition of equivalence
- Rather, we expect that careful assessment of the usefulness of current methods in substantive and theoretical applications will be helpful in determining when, and under what conditions, each method is useful
  - perhaps in conjunction with statistical assumptions
- Considerable work also needs to be done on measurement properties (such as sampling variability) of the current measures."

Centralization (Network Centrality)

- ***Centralization*** (network measure)
  - How central the most central node in the network in relation to all other nodes?
  - To check how much variation there is among the nodes (heterogeneity?)

$$C_x = \frac{\sum_i^N [C_x(p_*) - C_x(p_i)]}{\max \sum_i^N [C_x(p_*) - C_x(p_i)]}$$

C_x - one of the centrality measures

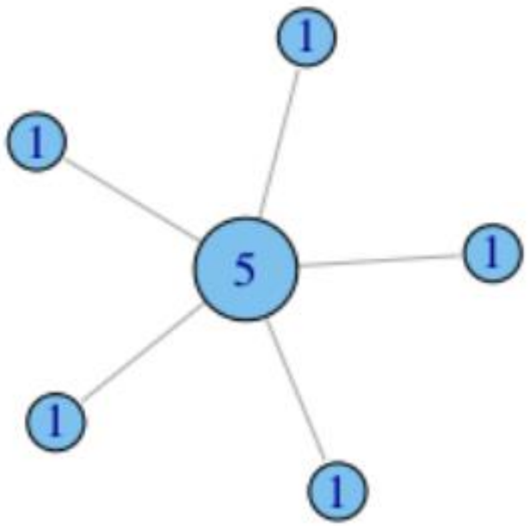
p_* - node with the largest centrality value

max - is taken over all graphs with the same number of nodes

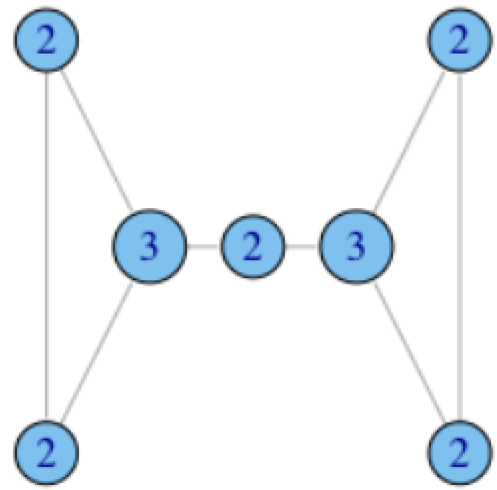
– for degree, closeness and betweenness, the most centralized structure is the star graph

Linton Freeman, 1979

Centralization (Network Centrality)



$C_D = 1.0$



$C_D = 0.167$

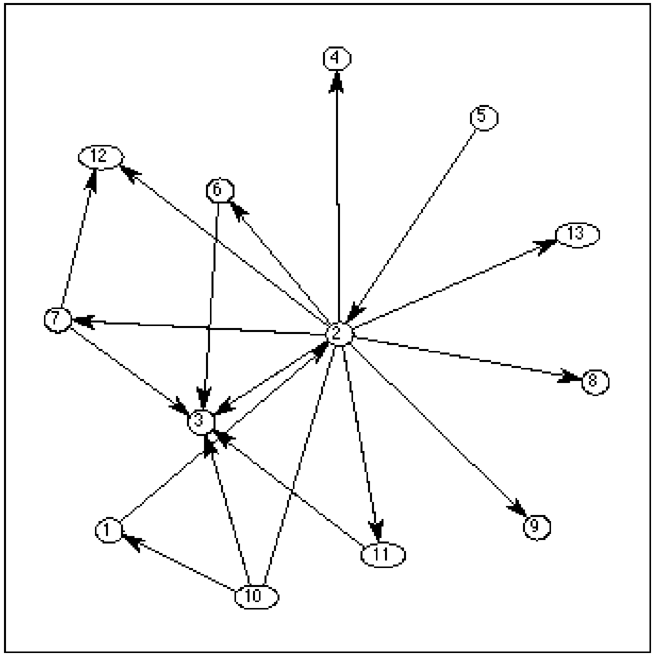
Degree Centrality

4+4+4+4+4/5*4
1+0+0+0+1/4*3=1/6
1+1+0+1+0+1+1/6*5=5/30

$C_D = 0.167$

Prestige

- ***Prestige*** - measure of node importance in directed graphs



  - Degree prestige $k_{in}(i)$
  - Proximity prestige (closeness)
  - Status or Rank prestige (Katz, Bonacich)

Metric Comparison

- **use PCC:** If you want to ***compare actual values*** of the centrality measures
- **use SRCC:** If you want to ***compare the actual rankings***

- ***Pearson correlation coefficient***

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Shows linear dependence between variables, $-1 \leq r \leq 1$ (perfect when related by linear function)

- ***Spearman rank correlation coefficient*** (Sperman's rho):
  - Convert raw scores to ranks - sort by score: $X_i \rightarrow x_i$, $Y_i \rightarrow y_i$

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2 - 1)}$$

Shows strength of monotonic association
(perfect for monotone increasing/decreasing relationship)

Spearman Rank Correlation

Sample Question:

The scores for nine students in physics and math are as follows:

Physics: 35, 23, 47, 17, 10, 43, 9, 6, 28

Mathematics: 30, 33, 45, 23, 8, 49, 12, 4, 31

Compute the student's ranks in the two subjects and compute the Spearman rank correlation.

Step 1: Find the ranks for each individual subject.

Physics	Rank	Math	Rank
35	3	30	5
23	5	33	3
47	1	45	2
17	6	23	6
10	7	8	8
43	2	49	1
9	8	12	7
6	9	4	9
28	4	31	4

Step 2: Add a third column, d, to your data. The d is the difference between ranks.

Physics	Rank	Math	Rank	d	d squared
35	3	30	5	2	4
23	5	33	3	2	4
47	1	45	2	1	1
17	6	23	6	0	0
10	7	8	8	1	1
43	2	49	1	1	1
9	8	12	7	1	1
6	9	4	9	0	0
28	4	31	4	0	0

https://www.statisticshowto.com/spearman-rank-correlation-definition-calculate/

Spearman Rank Correlation

Step 4: Sum (add up) all of your d-squared values.

4 + 4 + 1 + 0 + 1 + 1 + 1 + 0 + 0 = 12. You'll need this for the formula (the Σ d^2 is just "the sum of d-squared values").

Step 5: Insert the values into the formula. These ranks are not tied, so use the first formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

= 1 – (6*12)/(9(81-1))

= 1 – 72/720

= 1-0.1

= 0.9

The Spearman Rank Correlation for this set of data is 0.9.

Spearman's returns a value from -1 to 1, where:

+1 = a perfect positive correlation between ranks

-1 = a perfect negative correlation between ranks

0 = no correlation between ranks.

https://www.statisticshowto.com/spearman-rank-correlation-definition-calculate/

Correlation analysis among centrality measures for the Padgett Florentine families

	deg_B_S	close_B_S	betw_B_S	eigen_B..1..	page_B..1..
deg_B_S	1.0000	0.6976	0.6680	0.8620	0.8991
close_B_S	0.6976	1.0000	0.6905	0.7459	0.6611
betw_B_S	0.6680	0.6905	1.0000	0.5570	0.6963
eigen_B..1..	0.8620	0.7459	0.5570	1.0000	0.7000
page_B..1..	0.8991	0.6611	0.6963	0.7000	1.0000

Ranking Comparison

- The ***Kendall tau rank distance*** is a metric that counts the number of pairwise disagreements between two ranking lists
- ***Kendall rank correlation coefficient***, commonly referred to as ***Kendall's tau coefficient***

$$\tau = \frac{n_c - n_d}{n(n-1)/2}$$

n_c - number of concordant pairs, **n_d** - number of discordant pairs

- $-1 \leq \tau \leq 1$, perfect agreement $\tau = 1$, reversed $\tau = -1$
- Example

Rank 1	A	B	C	D	E
Rank 2	C	D	A	B	E

$$\tau = \frac{6-4}{5(5-1)/2} = 0.2$$