Department of Computer and IT Engineering
University of Kurdistan

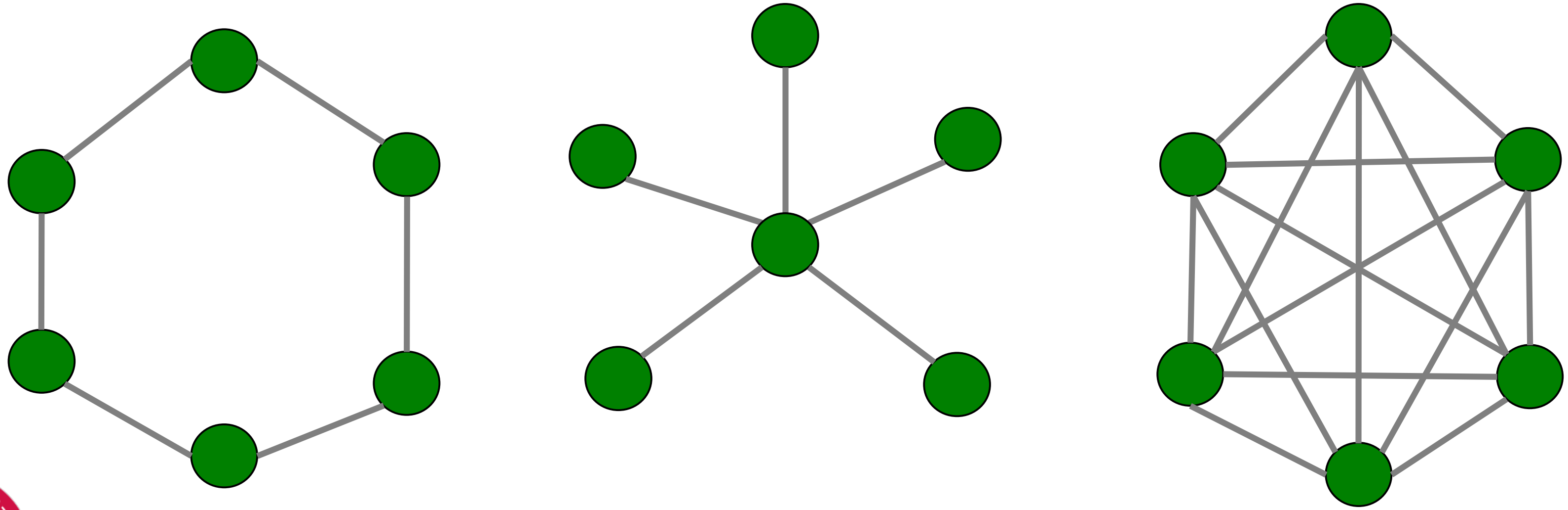
Complex Networks

Centrality

By: Dr. Alireza Abdollahpouri

Power in social networks

- Which vertices are **important**?
- How we answer this question depends on what exactly we mean by important, and there are several general approaches to answering it

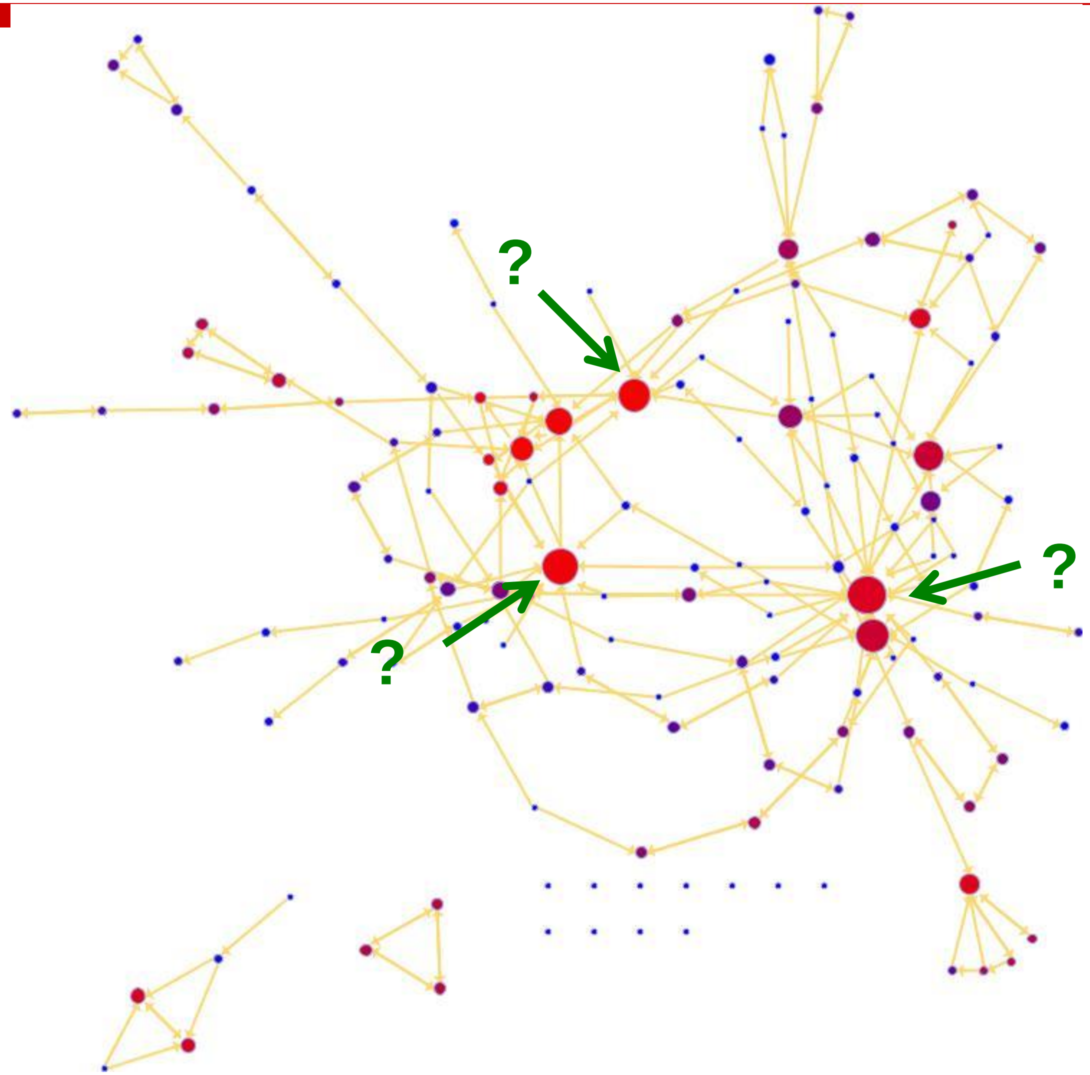


Centrality in social networks

- Centrality encodes the relationship between structure and power in groups
  Certain positions within the network give nodes more power or importance

- How do we measure importance?
  - Who can directly affect/influence others?
    - Highest *degree* nodes are "in the thick of it"
  - Who controls information flow?
    - Nodes that fall on shortest paths *between* others can disrupt the flow of information between them
  - Who can quickly inform most others?
    - Nodes who are *close* to other nodes can quickly get information to them

Characterizing networks: Who is most central?



Network centrality

- Which nodes are most ‘central’?

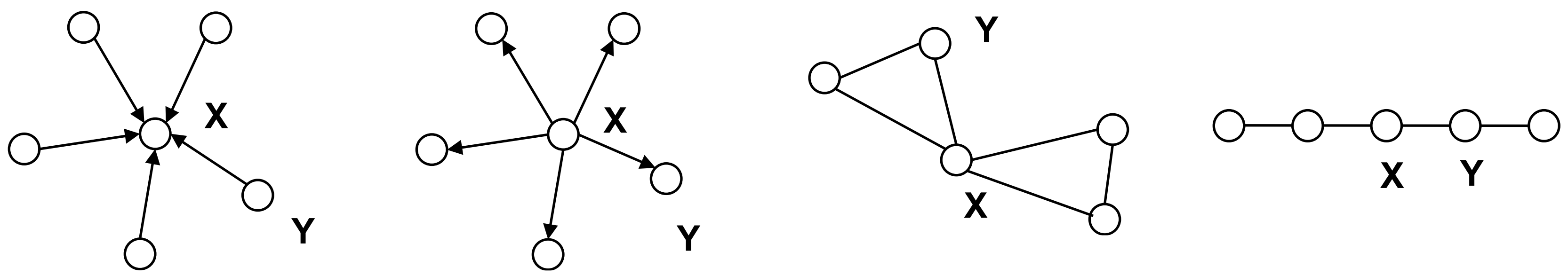
- Local measure:
  - degree

- Relative to rest of network:
  - closeness, betweenness, eigenvector (Bonacich power centrality), Katz, PageRank, …

- How evenly is centrality distributed among nodes?
  - Centralization, hubs and authorities, …

Centrality: who's important based on their network position

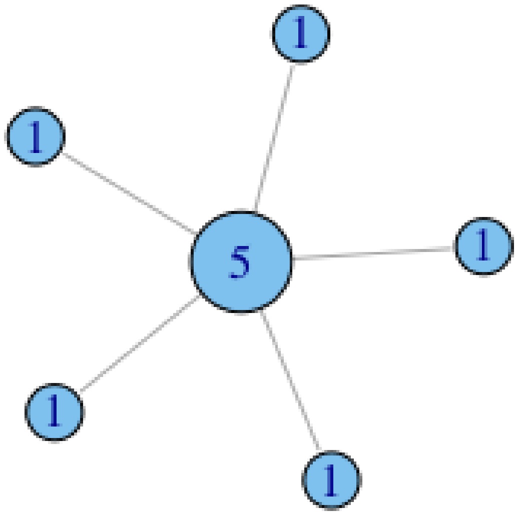
In each of the following networks, X has higher centrality than Y according to a particular measure



indegree **outdegree** **betweenness** **closeness**

Degree centrality (undirected)

He who has many friends is most important.

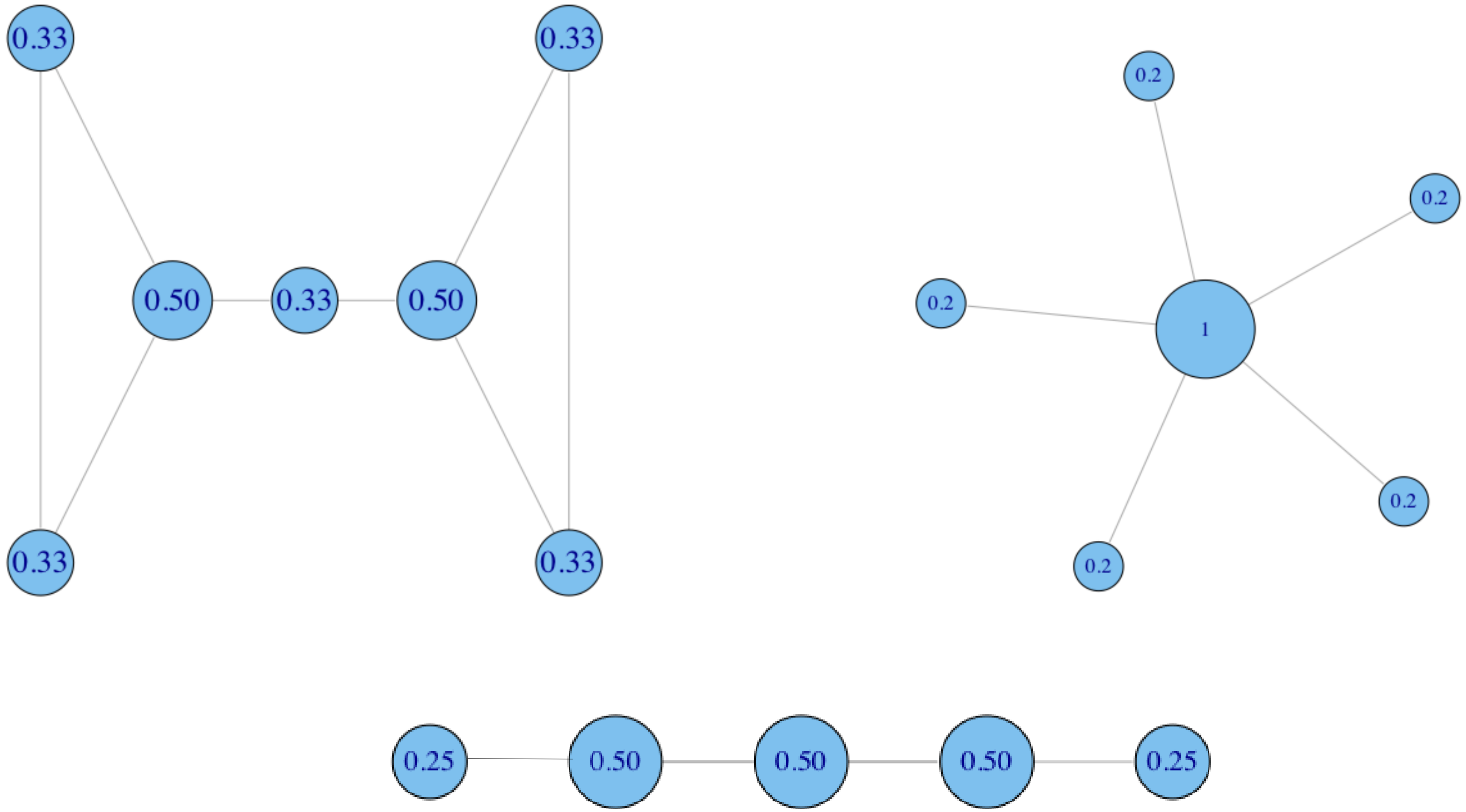


When is the number of connections the best centrality measure?

- **people who will do favors for you**
- **people you can talk to (influence set, information access, …)**
- **influence of an article in terms of citations (using in-degree)**

Degree: normalized degree centrality

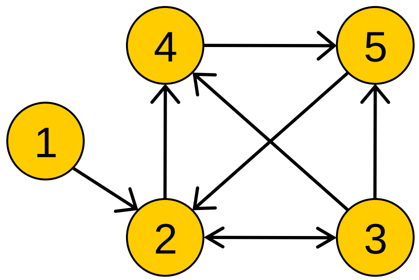
divide by the max. possible, i.e. (N-1)

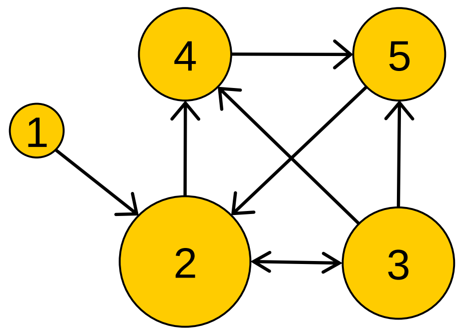


Degree centrality

- The number of others a node is connected to
  - Node with high degree has high potential communication activity

node	In-degree	Out-degree	Total degree
1	0	1	1
2	3	2	5
3	1	3	4
4	2	1	3
5	2	1	3



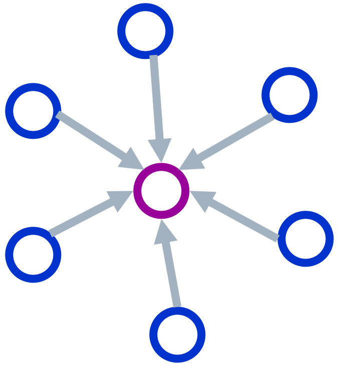


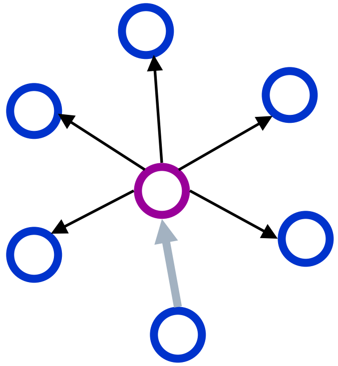
Extensions of undirected degree centrality - prestige

- degree centrality
  - indegree centrality

    - a paper that is cited by many others has **high prestige**

    - a person nominated by many others for a reward has **high prestige**





Centralization: how equal are the nodes?

How much variation is there in the centrality scores among the nodes?

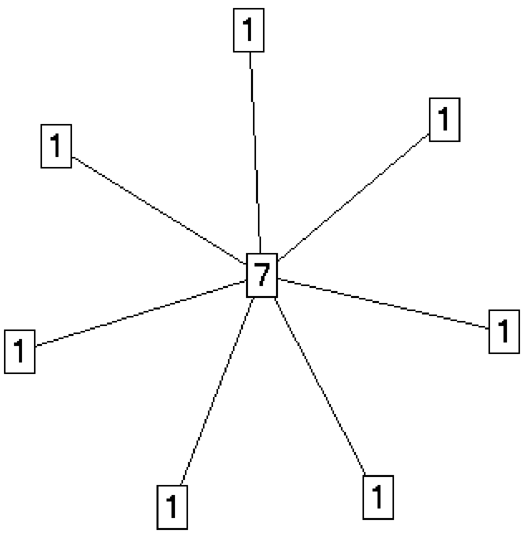
Freeman's general formula for centralization:

(can use other metrics, e.g. gini coefficient or standard deviation)

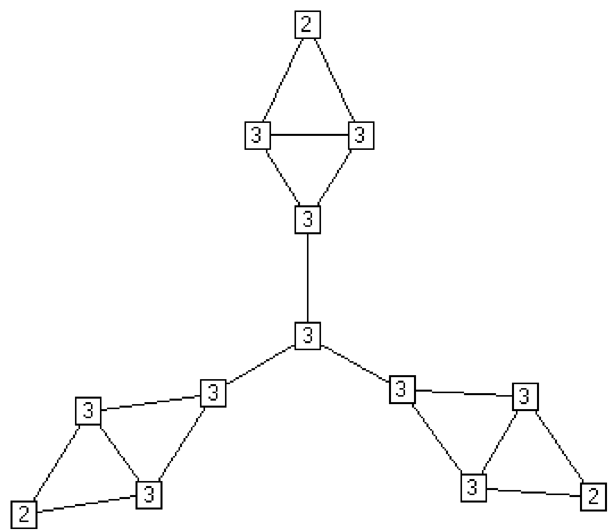
maximum value in the network

$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(i)]}{[(N-1)(N-2)]}$$

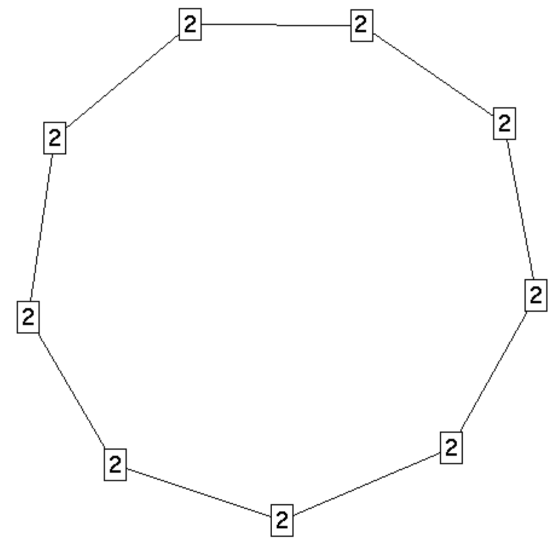
Degree Centrality in Social Networks



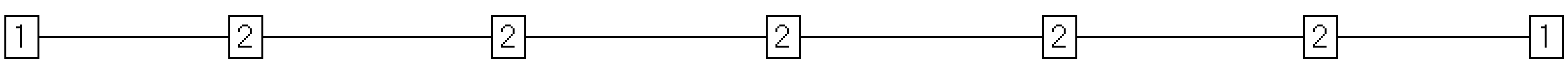
Freeman: 1.0
Variance: 3.9



Freeman: .02
Variance: .17

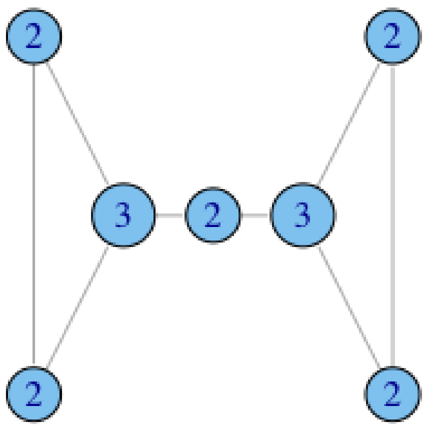


Freeman: 0.0
Variance: 0.0

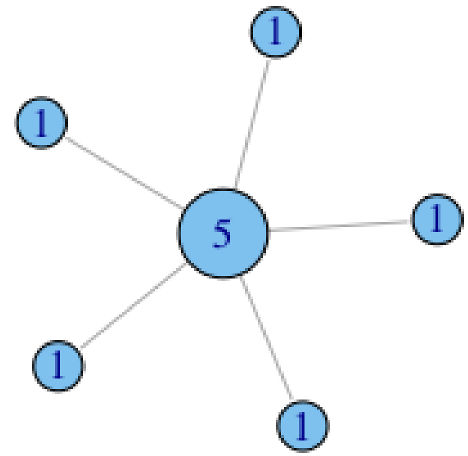


Freeman: .07
Variance: .20

Degree centralization examples



$C_D = 0.167$



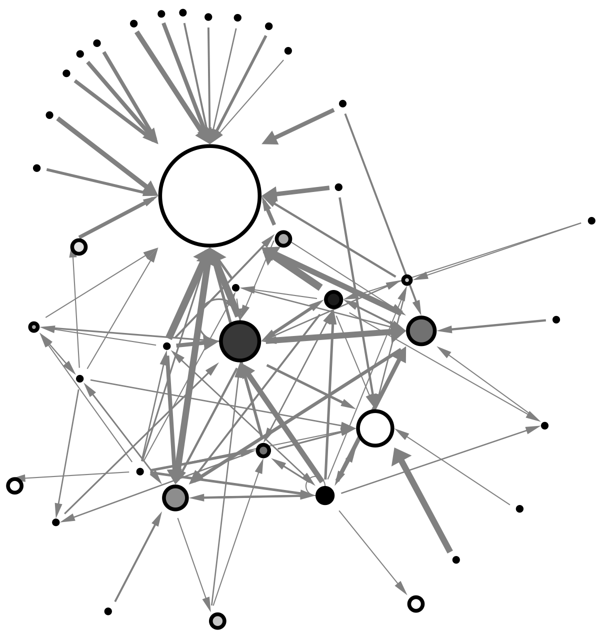
$C_D = 1.0$

$C_D = 0.167$

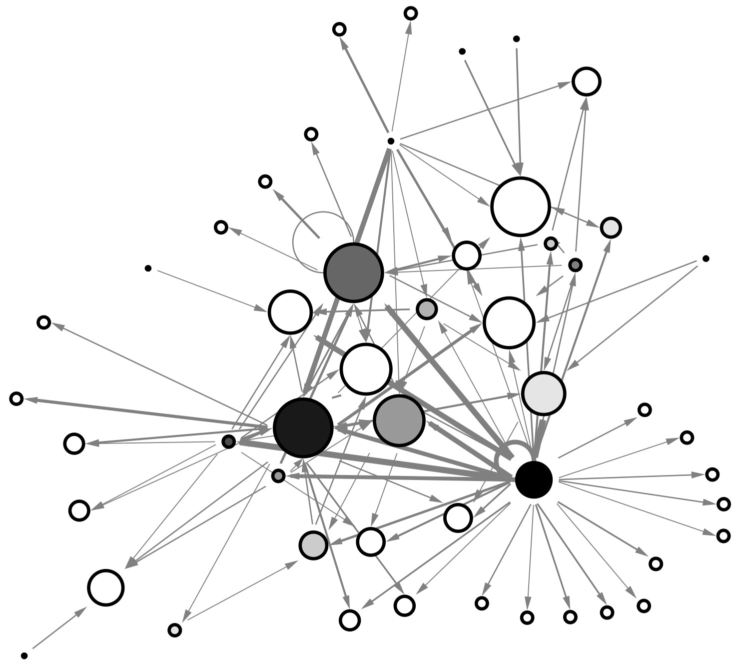
Centralization= Σ(C*-Ci) / Max Σ(C*-Ci)

Degree centralization examples

example financial trading networks



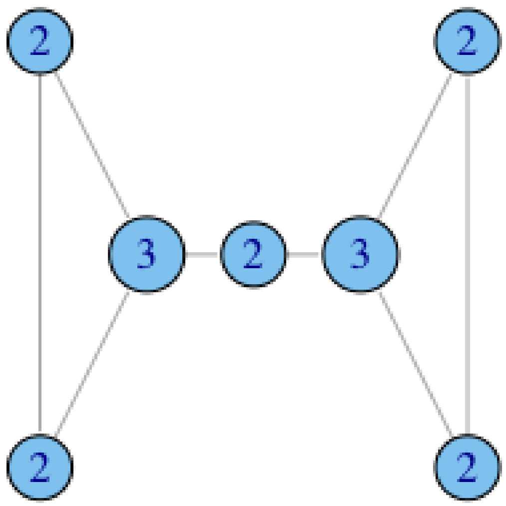
high centralization: one node trading with many others



low centralization: trades are more evenly distributed

When degree isn't everything

In what ways does degree fail to capture centrality in the following graphs?

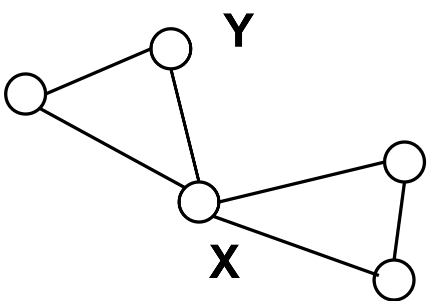


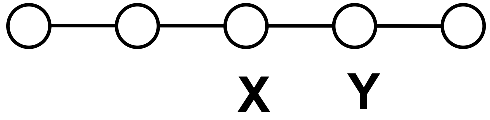
- ability to broker between groups
- likelihood that information originating anywhere in the network reaches you…

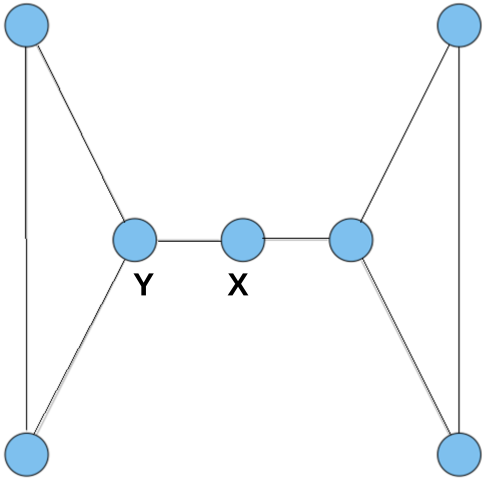
Betweenness: another centrality measure

- **intuition**: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?

- who has higher betweenness, X or Y?



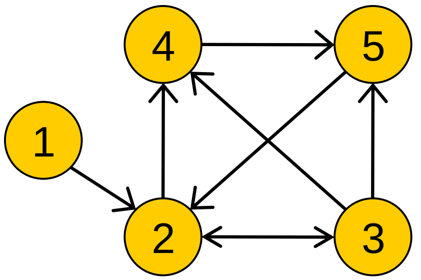


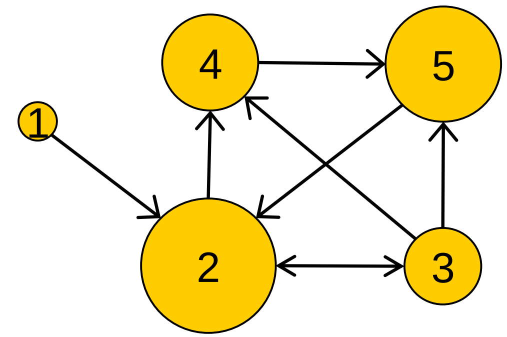


Betweenness centrality

- Number of shortest paths (geodesics) connecting all pairs of other nodes that pass through a given node
  - Node with highest betweenness can potentially control or distort communication

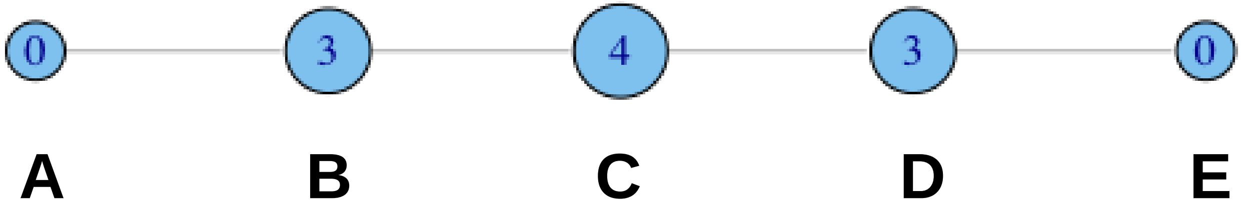
1→2	1→2→3	1→2→4	1→2→4→5
	2→3	2→4	2→4→5
	3→2	3→4	3→5
	4→5→2	4→5→2→3	4→5
	5→2	5→2→3	5→2→4





Betweenness on toy networks

- non-normalized version:



- A 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)

- note that there are no alternate paths for these pairs to take, so C gets full credit

Betweenness centrality: definition

betweenness of vertex i

paths between j and k that pass through i

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

all paths between j and k

Where g_{jk} = the number of geodesics connecting *j-k*, and g_{jk} = the number that actor *i* is on.

Usually normalized by:

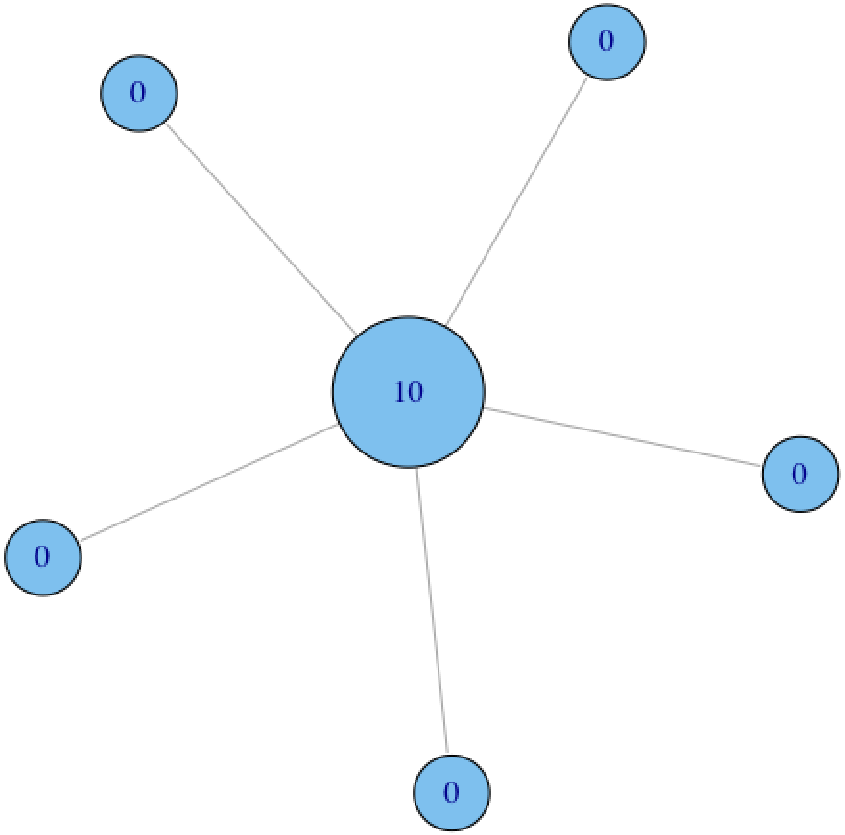
$$C'_B(i) = C_B(i) / [(n-1)(n-2)/2]$$

number of pairs of vertices excluding the vertex itself

For directed graph: (N-1)*(N-2)

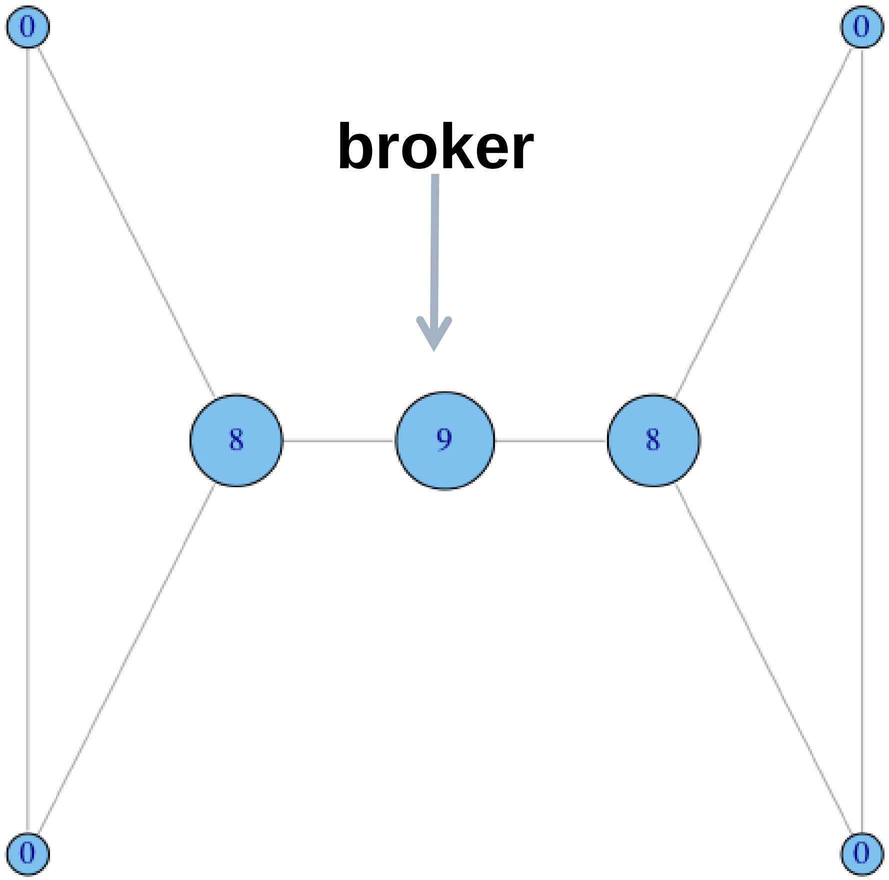
Betweenness on toy networks

- non-normalized version:



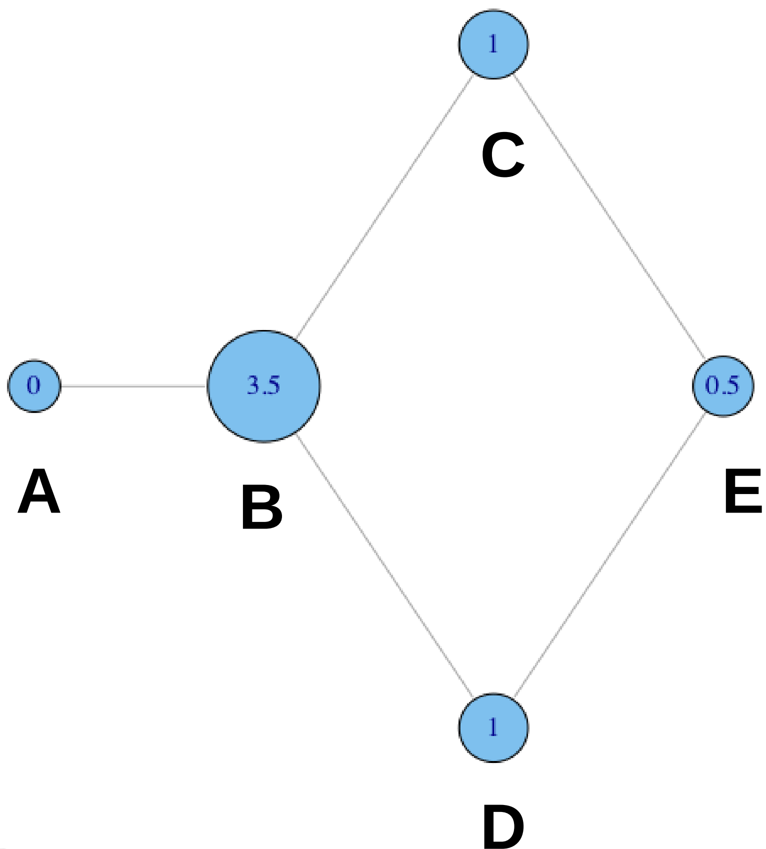
Betweenness on toy networks

- non-normalized version:



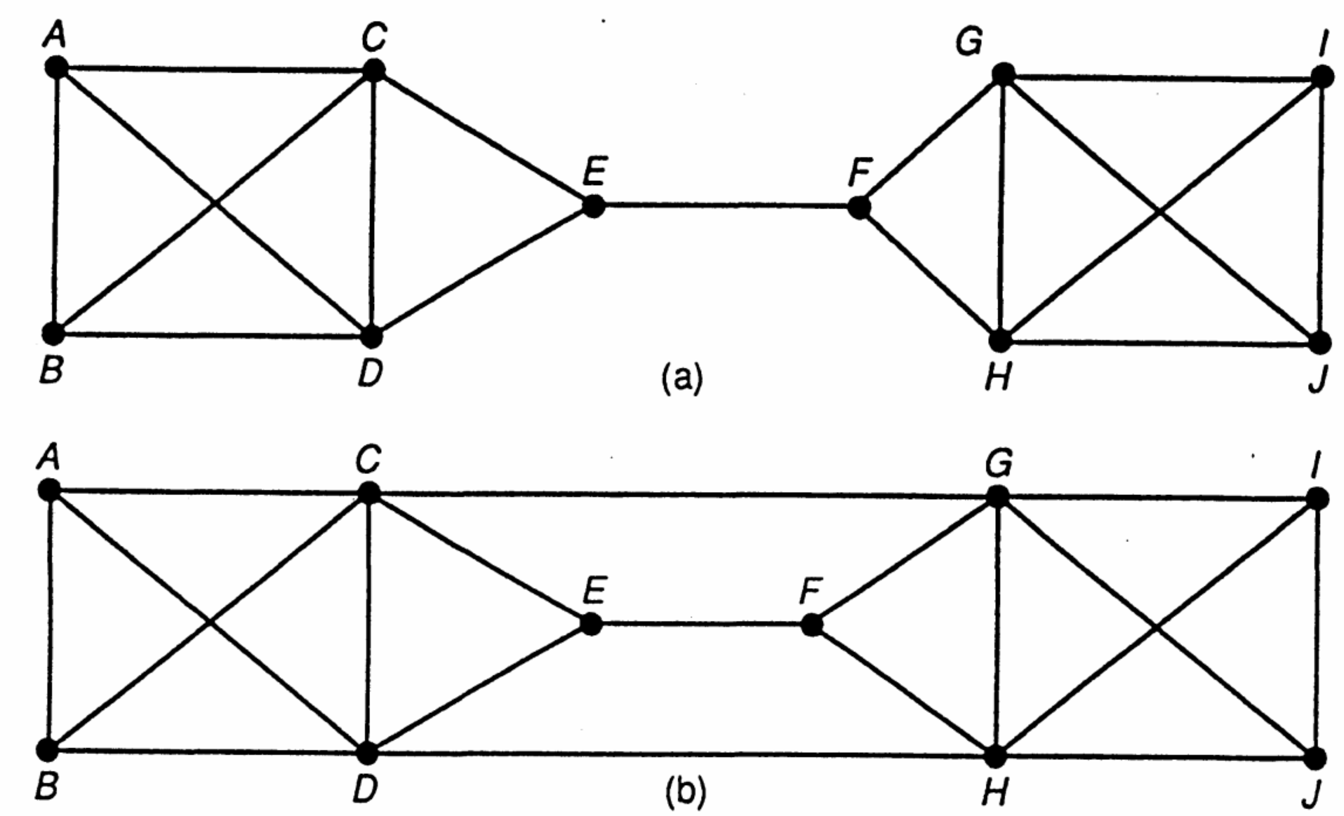
Betweenness on toy networks

- non-normalized version:



- 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?

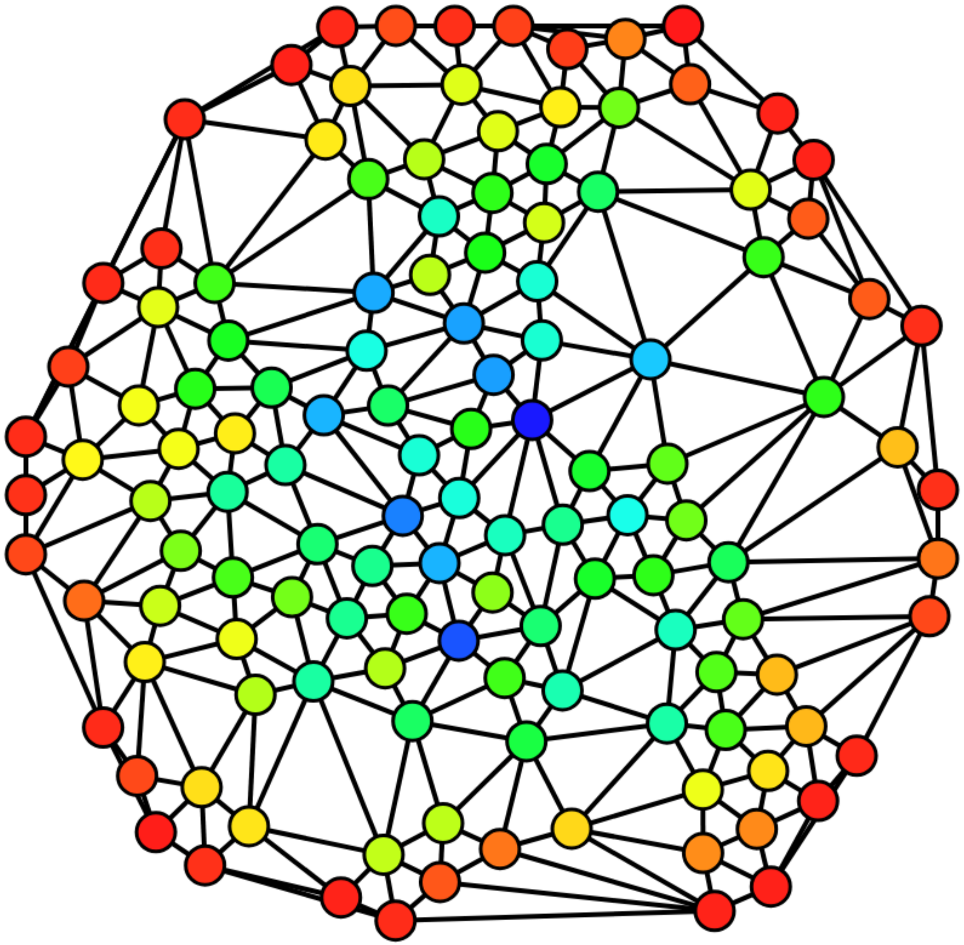
Betweenness centrality



If you add lines from C to G and from D to H, you remove the high betweenness centrality of E and F

University of Kurdistan

Betweenness centrality



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

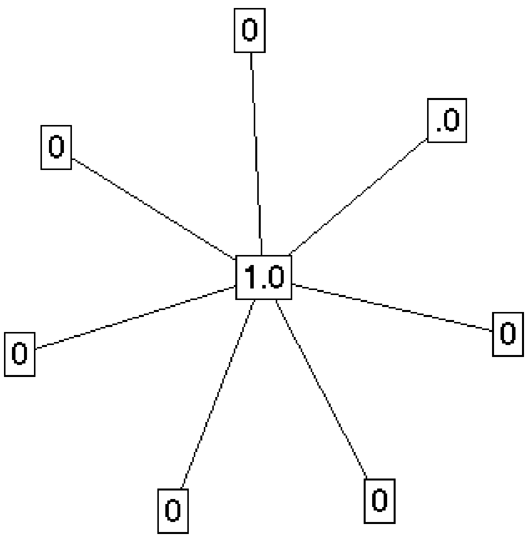
Centrality vs. Centralization

Centrality is a characteristic of an actor’s position in a network
Centralization is a characteristic of a network
Centralization indicates:
- how unequal the distribution of centrality is in a network or
- how much variance there is in the distribution of centrality in a network

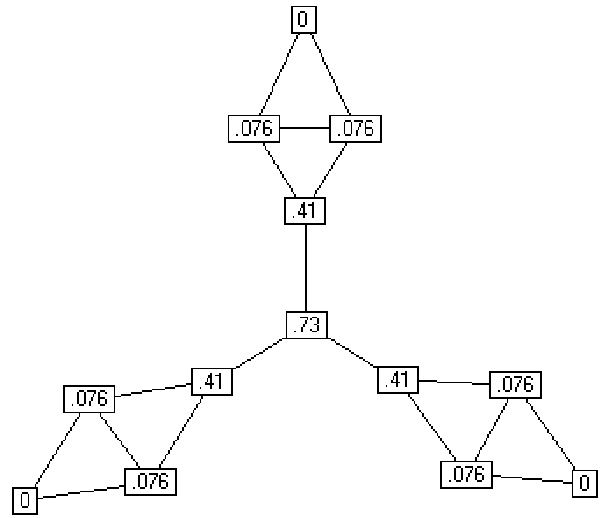
- Centrality is a micro-level measure
- Centralization is a macro-level measure

$$C_B(G) = \frac{\sum_{i=1}^{n}[C_B{}'(v^*) - C_B{}'(v_i)]}{(n-1)}$$

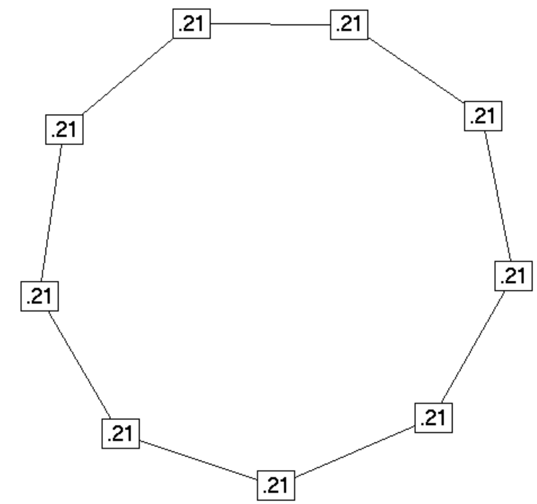
Betweenness Centralization (examples)



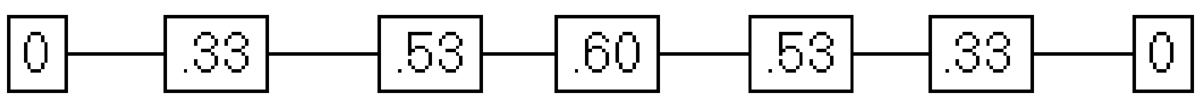
Centralization: 1.0



Centralization: .59



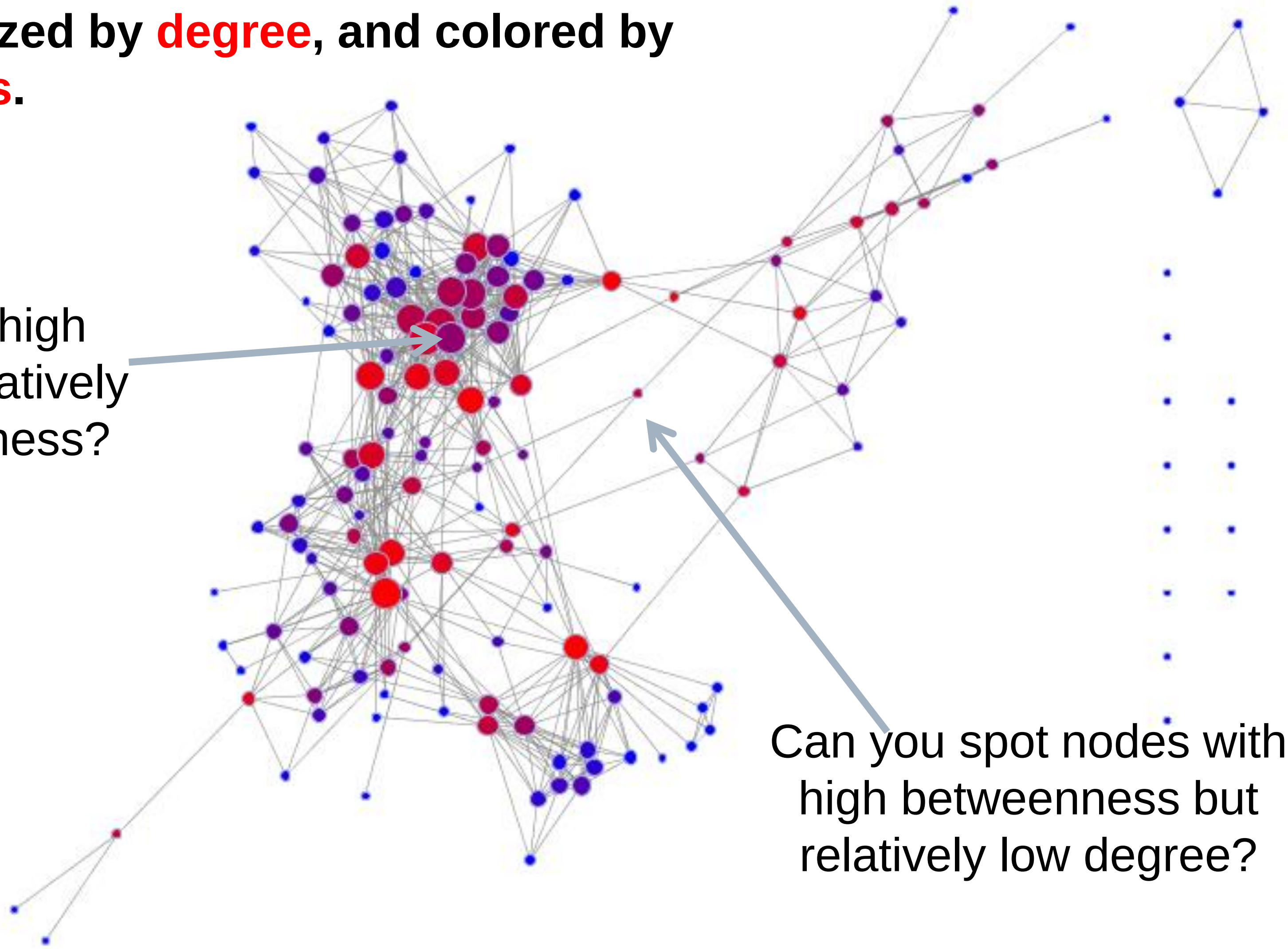
Centralization: 0



Centralization: .31

Comparison

Nodes are sized by degree, and colored by betweenness.



What about high degree but relatively low betweenness?

Can you spot nodes with high betweenness but relatively low degree?

Extending betweenness centrality to directed networks

- We now consider the fraction of all directed paths between any two vertices that pass through a node

betweenness of vertex i

paths between j and k that pass through i

$$C_B(i) = \sum_{j,k} g_{jk}(i) / g_{jk}$$

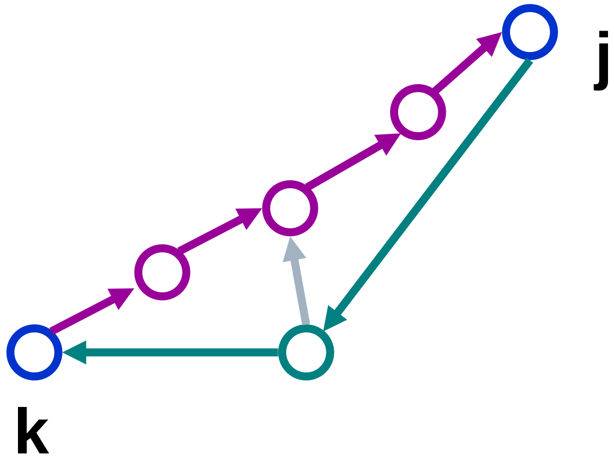
all paths between j and k

- Only modification: when normalizing, we have (N-1)*(N-2) instead of (N-1)*(N-2)/2, because we have twice as many ordered pairs as unordered pairs

$$C'_B(i) = C_B(i) / [(N-1)(N-2)]$$

Directed geodesics

A node does not necessarily lie on a geodesic from *j* to *k* if it lies on a geodesic from *k* to *j*



Closeness: another centrality measure

- What if it's not so important to have many direct friends?
- Or be "between" others
- But one still wants to be in the "middle" of things,
  - not too far from the center

Closeness centrality: definition

Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

Closeness Centrality:

$$C_c(i) = \left[\sum_{j=1}^{N} d(i,j) \right]^{-1}$$

depends on inverse distance to other vertices

Normalized Closeness Centrality

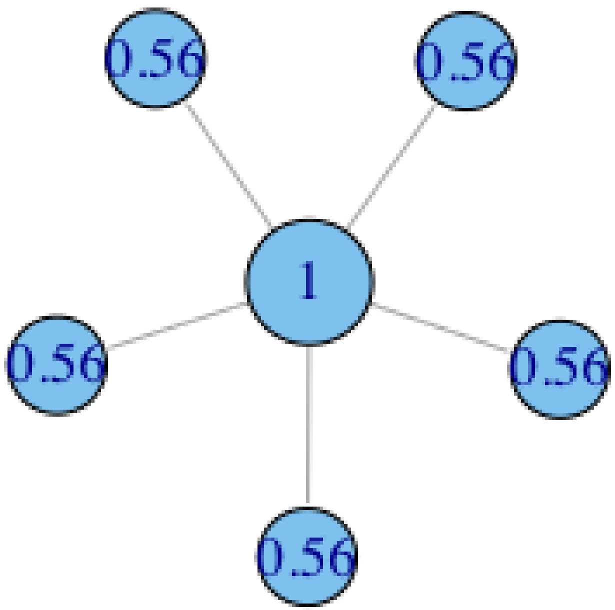
$$C_C^{'}(i) = (C_C(i)).(N-1)$$

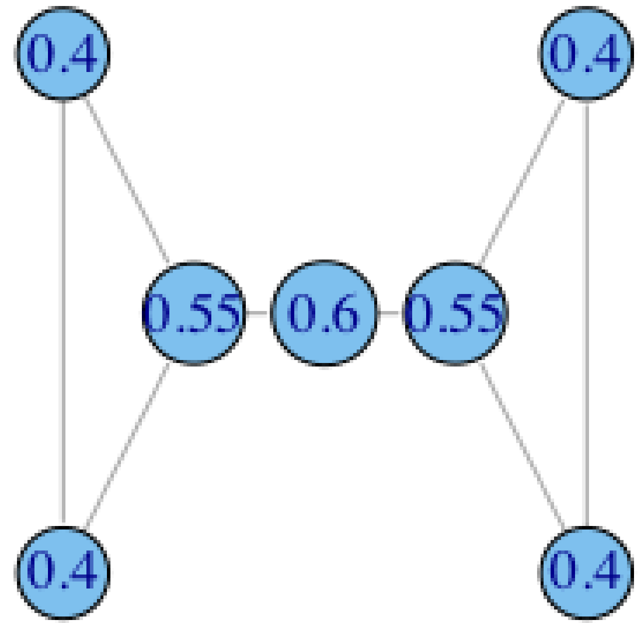
Closeness centrality: toy example

A B C D E

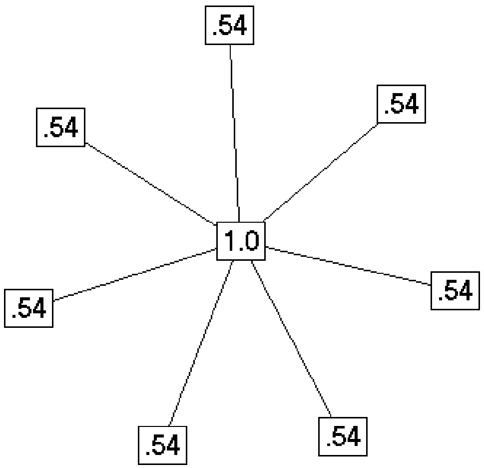
$$C'_c(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$

Closeness centrality: more toy examples

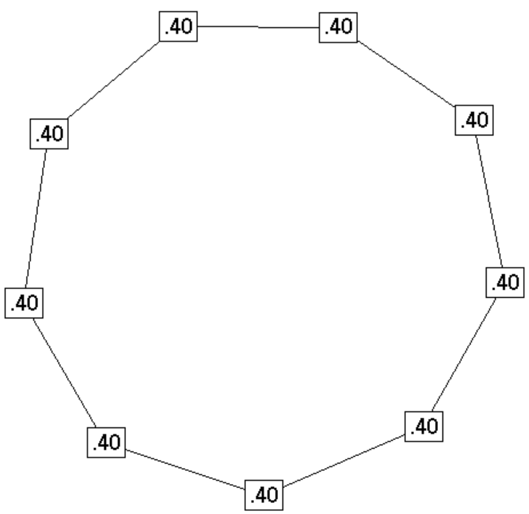




Closeness Centrality (examples)

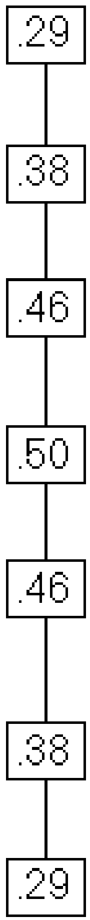


Distance	Closeness	normalized
0 1 1 1 1 1 1 1	.143	1.00
1 0 2 2 2 2 2 2	.077	.538
1 2 0 2 2 2 2 2	.077	.538
1 2 2 0 2 2 2 2	.077	.538
1 2 2 2 0 2 2 2	.077	.538
1 2 2 2 2 0 2 2	.077	.538
1 2 2 2 2 2 0 2	.077	.538
1 2 2 2 2 2 2 0	.077	.538



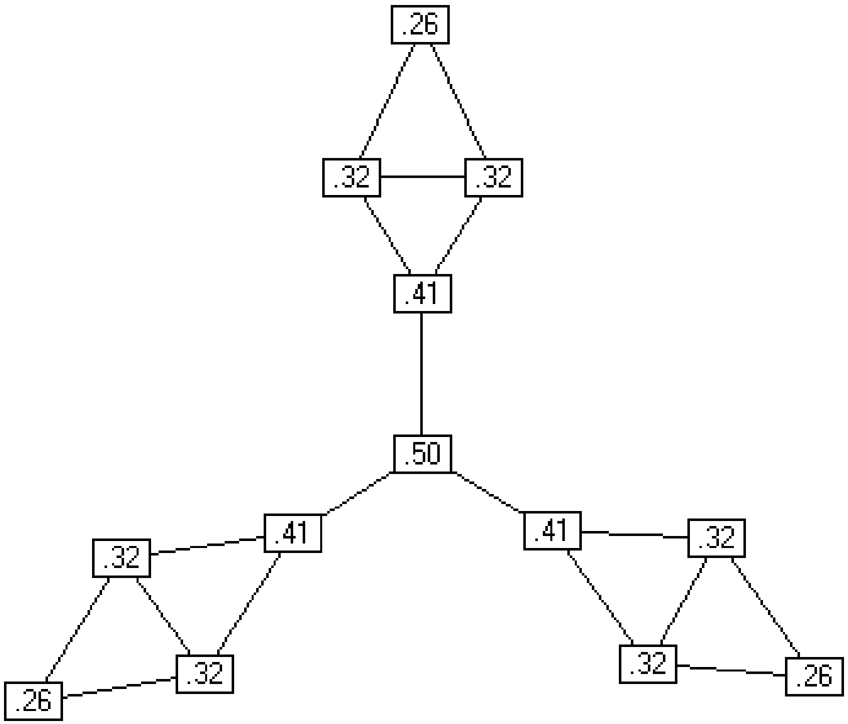
Distance	Closeness	normalized
0 1 2 3 4 4 3 2 1	.050	.400
1 0 1 2 3 4 4 3 2	.050	.400
2 1 0 1 2 3 4 4 3	.050	.400
3 2 1 0 1 2 3 4 4	.050	.400
4 3 2 1 0 1 2 3 4	.050	.400
4 4 3 2 1 0 1 2 3	.050	.400
3 4 4 3 2 1 0 1 2	.050	.400
2 3 4 4 3 2 1 0 1	.050	.400
1 2 3 4 4 3 2 1 0	.050	.400

Closeness Centrality in Social Networks



Distance	Closeness	normalized
0 1 2 3 4 5 6	.048	.286
1 0 1 2 3 4 5	.063	.375
2 1 0 1 2 3 4	.077	.462
3 2 1 0 1 2 3	.083	.500
4 3 2 1 0 1 2	.077	.462
5 4 3 2 1 0 1	.063	.375
6 5 4 3 2 1 0	.048	.286

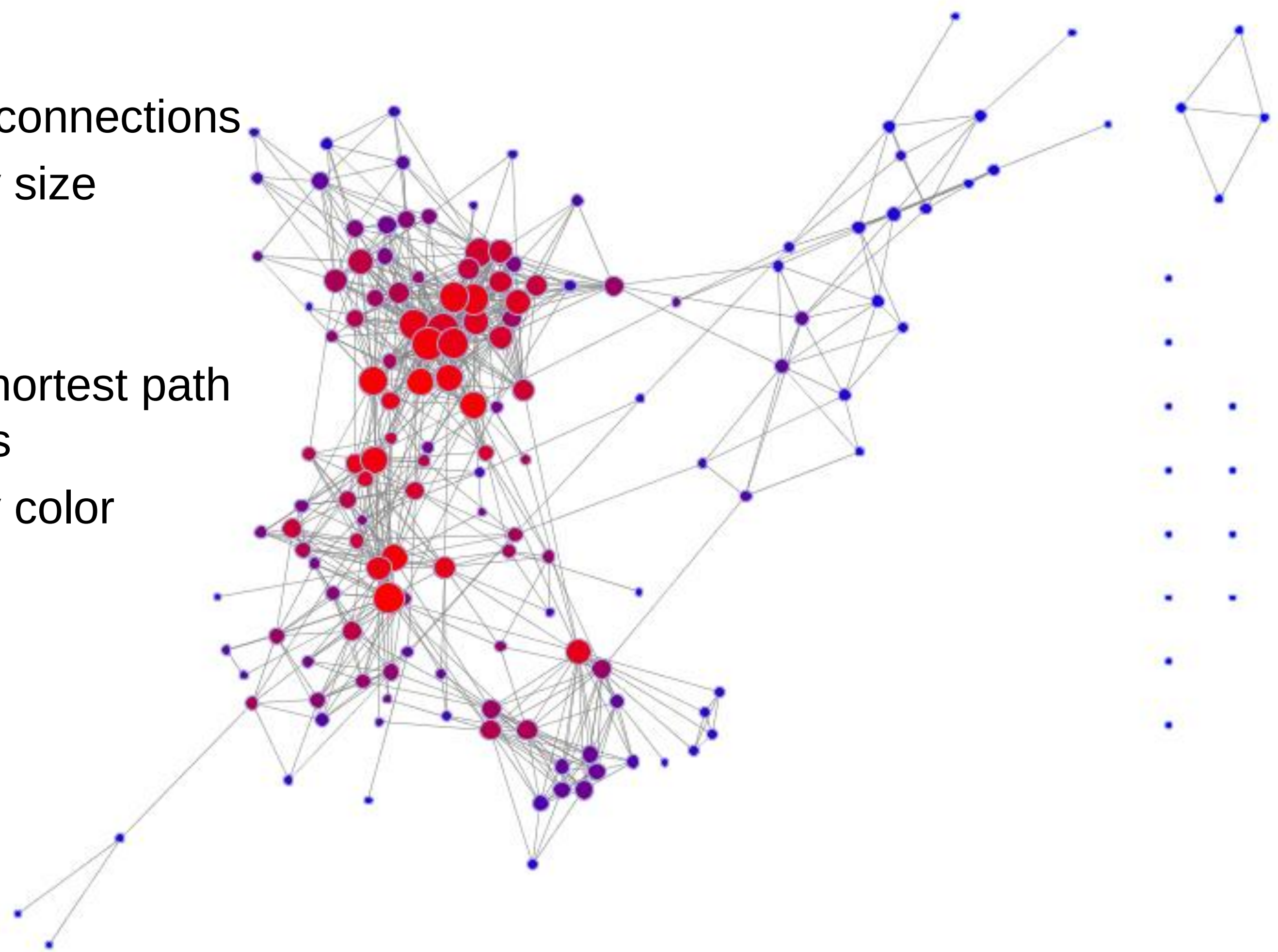
Closeness Centrality in Social Networks



Distance	Closeness	normalized
0 1 1 2 3 4 4 5 5 6 5 5 6	.021	.255
1 0 1 1 2 3 3 4 4 5 4 4 5	.027	.324
1 1 0 1 2 3 3 4 4 5 4 4 5	.027	.324
2 1 1 0 1 2 2 3 3 4 3 3 4	.034	.414
3 2 2 1 0 1 1 2 2 3 2 2 3	.042	.500
4 3 3 2 1 0 2 3 3 4 1 1 2	.034	.414
4 3 3 2 1 2 0 1 1 2 3 3 4	.034	.414
5 4 4 3 2 3 1 0 1 1 4 4 5	.027	.324
5 4 4 3 2 3 1 1 0 1 4 4 5	.027	.324
6 5 5 4 3 4 2 1 1 0 5 5 6	.021	.255
5 4 4 3 2 1 3 4 4 5 0 1 1	.027	.324
5 4 4 3 2 1 3 4 4 5 1 0 1	.027	.324
6 5 5 4 3 2 4 5 5 6 1 1 0	.021	.255

How closely do degree and betweenness correspond to closeness?

- **degree**
  - number of connections
  - denoted by size
- **closeness**
  - length of shortest path to all others
  - denoted by color

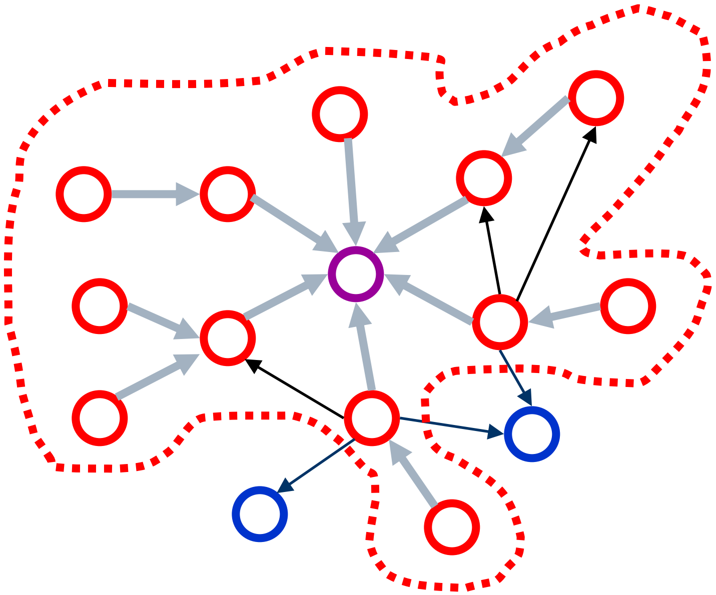


Closeness centrality

- Values tend to span a rather small dynamic range
  - typical distance increases logarithmically with network size
- In a typical network the closeness centrality C might span a factor of five or less
  - It is difficult to distinguish between central and less central vertices
  - a small change in network might considerably affect the centrality order
- Alternative computations exist but they have their own problems

Influence range

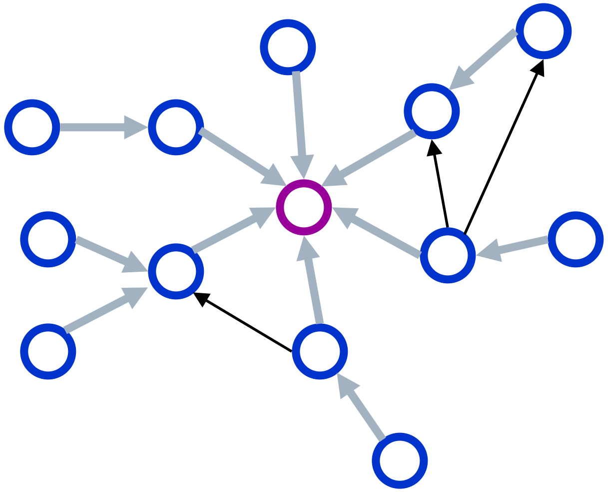
- The influence range of *i* is the set of vertices who are reachable from the node *i*



Extensions of undirected closeness centrality

- closeness centrality usually implies
  - all paths should lead to you
  - paths should lead from you to everywhere else

- usually consider only vertices from which the node *i* in question can be reached



Eigenvector Centrality

Idea: A central actor is connected to other central actors

A natural extension of the degree centrality

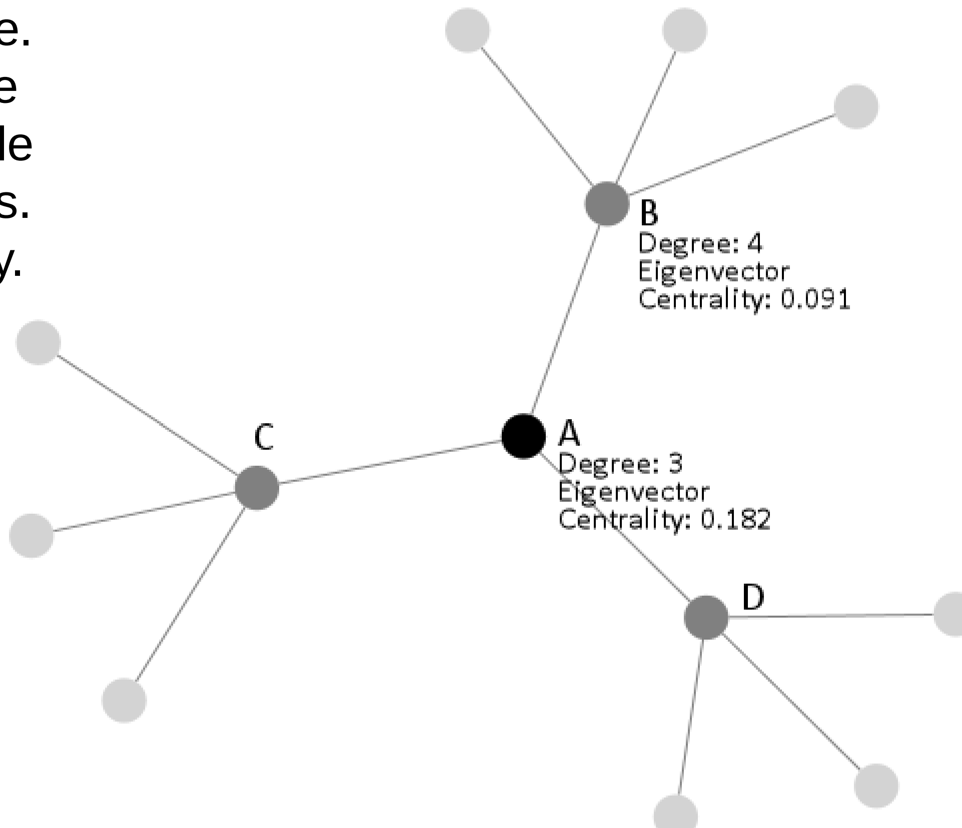
For a given graph G:=(V,E) with |V| number of vertices let A be the adjacency matrix. The centrality score of vertex v can be defined as:

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

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j$$

Eigenvector Centrality

Node **B** is more popular in the network if we only extend our vision out to a distance of 1 from each node. But **A** is connected to nodes that are connected to many other nodes, while **B** is connected to less-popular nodes. **A** has a higher eigenvector centrality.



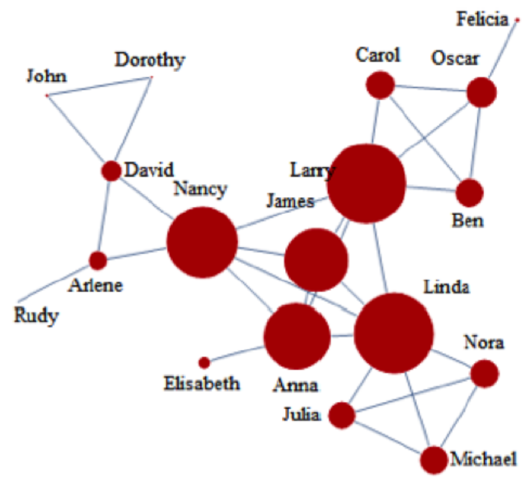
Eigenvector Centrality

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

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

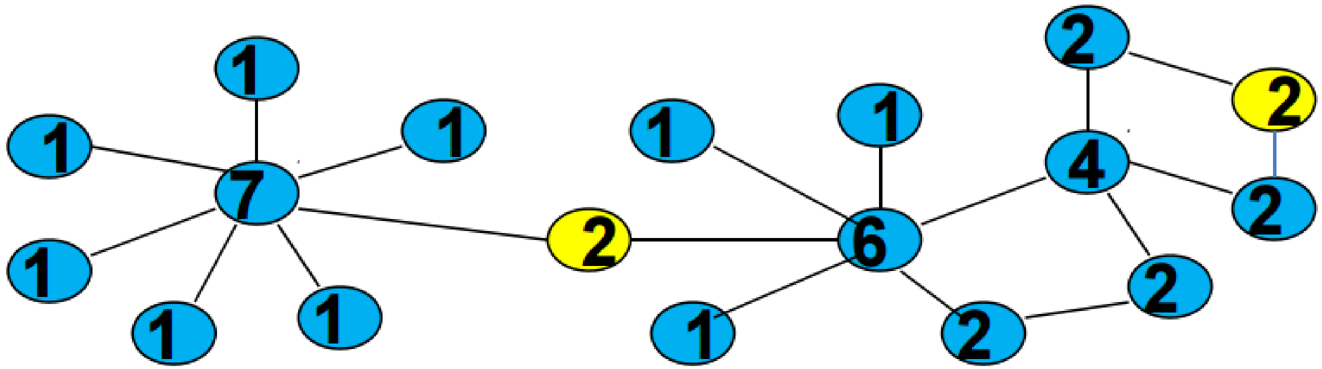
$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

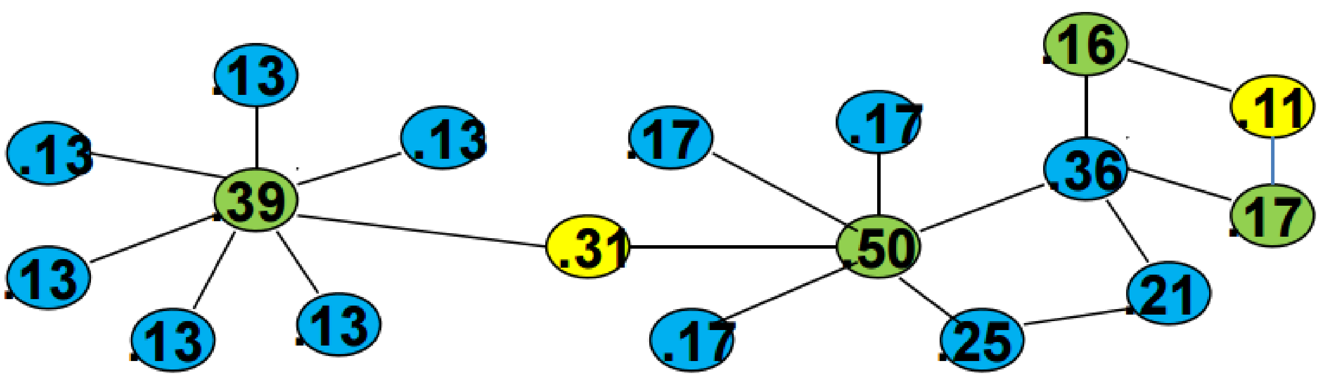
$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$



Select an eigenvector associated with largest eigenvalue $\lambda = \lambda_1$, $\mathbf{v} = \mathbf{v}_1$

Degree vs. Eigenvector Centrality

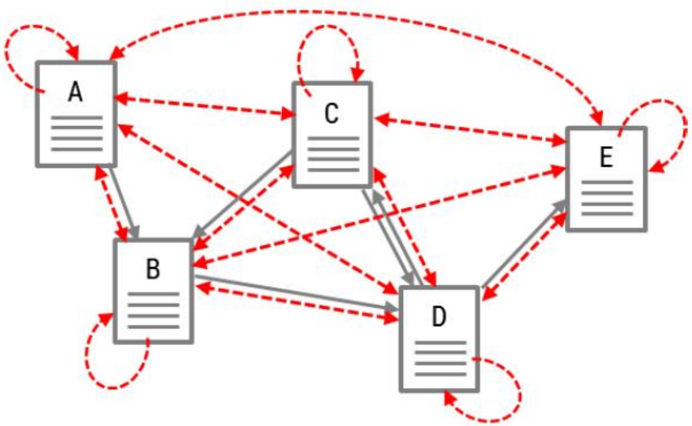


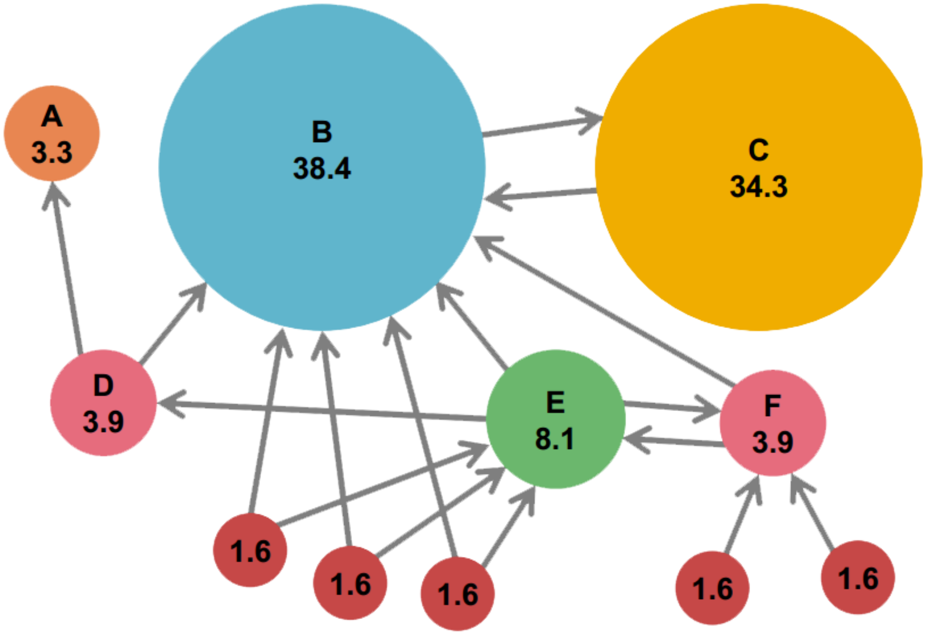


PageRank: Standing on the Shoulders of Giants

Key insights

- Analyzes the structure of the web of hyperlinks to determine importance score of web pages
  - A web page is important if it is pointed to by other important pages
- An algorithm with deep mathematical roots
  - Random walks
  - Social network theory





Page rank

Developed by Google founders to measure the importance of webpages from the hyperlink network structure.

- Link analysis approaches
  - Rank pages (nodes) by analyzing topology of the web graph
  - Idea: Links as votes
    - Page is more important if it has more links adjacent to it
  - Incoming links? Outgoing links?
  - Links from important pages have higher weight => recursive problem!

Page rank

n = number of nodes in the network

k = number of steps

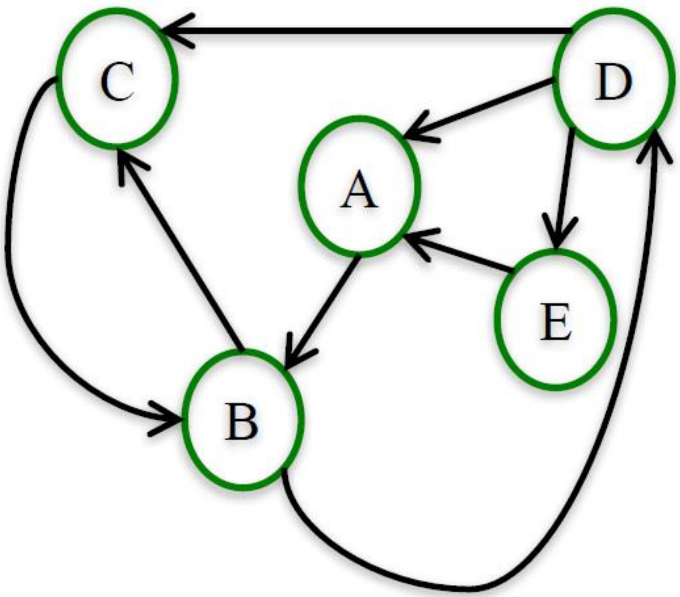
- 1. Assign all nodes a PageRank of $1/n$
- 2. Perform the *Basic PageRank Update Rule* *k* times.

Basic PageRank Update Rule: Each node gives an equal share of its current PageRank to all the nodes it links to.

The new PageRank of each node is the sum of all the PageRank it received from other nodes.

Page rank- Example

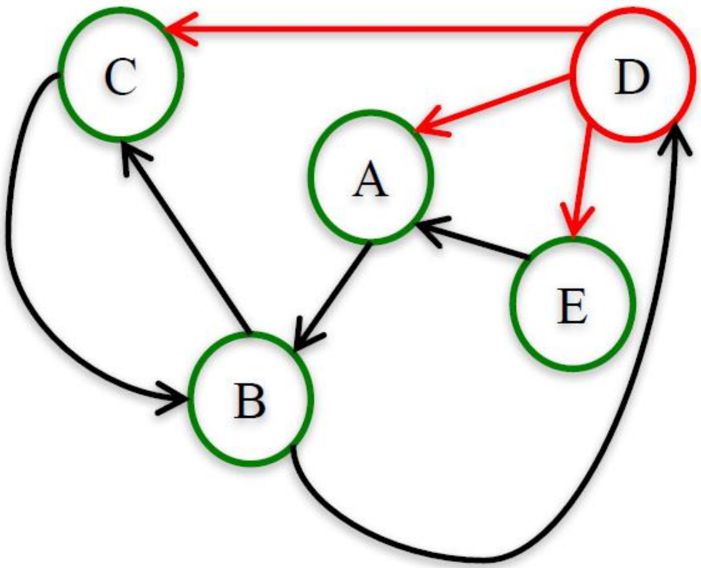
- Who should be the most “important” node in this network?
- Calculate the PageRank of each node after 2 steps of the procedure (k = 2).



Page Rank					
	A	B	C	D	E
	1/5	1/5	1/5	1/5	1/5

Page rank- Example

Page Rank (k = 1)					
	A	B	C	D	E
Old	1/5	1/5	1/5	1/5	1/5
New	4/15				



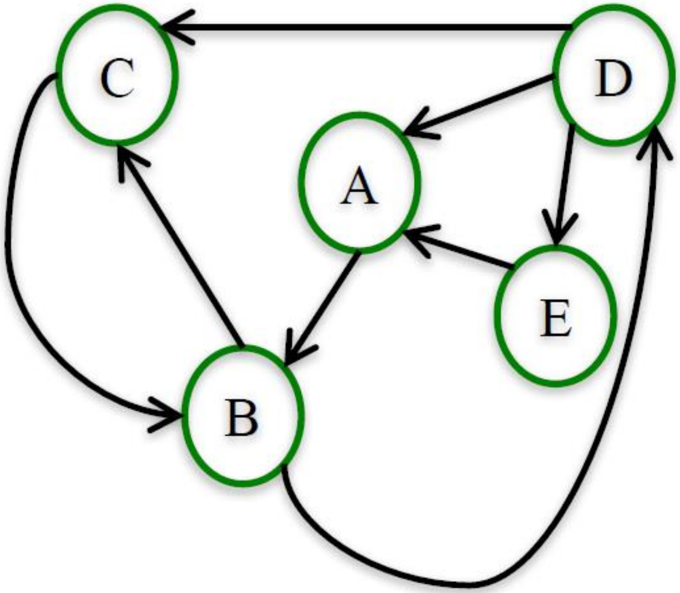
A: (1/3)*(1/5) + 1/5 = 4/15

From D   From E

University of Kurdistan

Page rank- Example

Page Rank (k = 1)					
	A	B	C	D	E
Old	1/5	1/5	1/5	1/5	1/5
New	4/15	2/5	1/6	1/10	1/15



A: (1/3)*(1/5) + 1/5 = 4/15

B: 1/5 + 1/5 = 2/5

C: (1/3)*(1/5) + (1/2)*(1/5) = 5/30 = 1/6

D: (1/2)*(1/5) = 1/10

E: (1/3)*(1/5) = 1/15

Page rank- Example

Page Rank (k = 2)					
	A	B	C	D	E
Old	4/15	2/5	1/6	1/10	1/15
New	1/10	13/30	7/30	2/10	1/30

A: (1/3)*(1/10) + 1/15 = 1/10
B: 1/6 + 4/15 = 13/30
C: (1/3)*(1/10) + (1/2)*(2/5) = 7/30
D: (1/2)*(2/5) = 2/10
E: (1/3)*(1/10) = 1/30

Page rank- Example

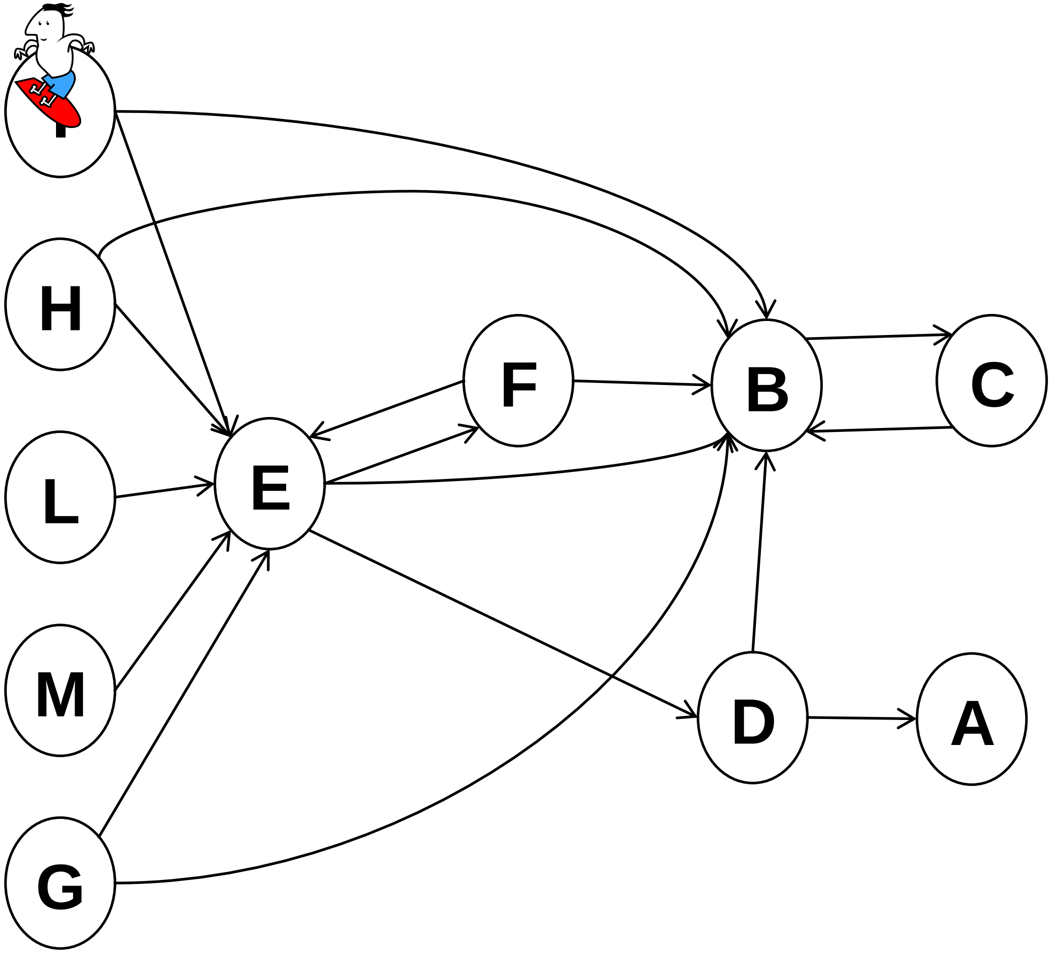
- What if continue with k = 4,5,6,...? For most networks, PageRank values converge

	Page Rank				
	A	B	C	D	E
k=2	1/10	13/30	7/30	2/10	1/30
k=2	.1	.43	.23	.20	.03
k=3	.1	.33	.28	.22	.06
k=∞	.12	.38	.25	.19	.06

PageRank and the Random Surfer

Random Surfer

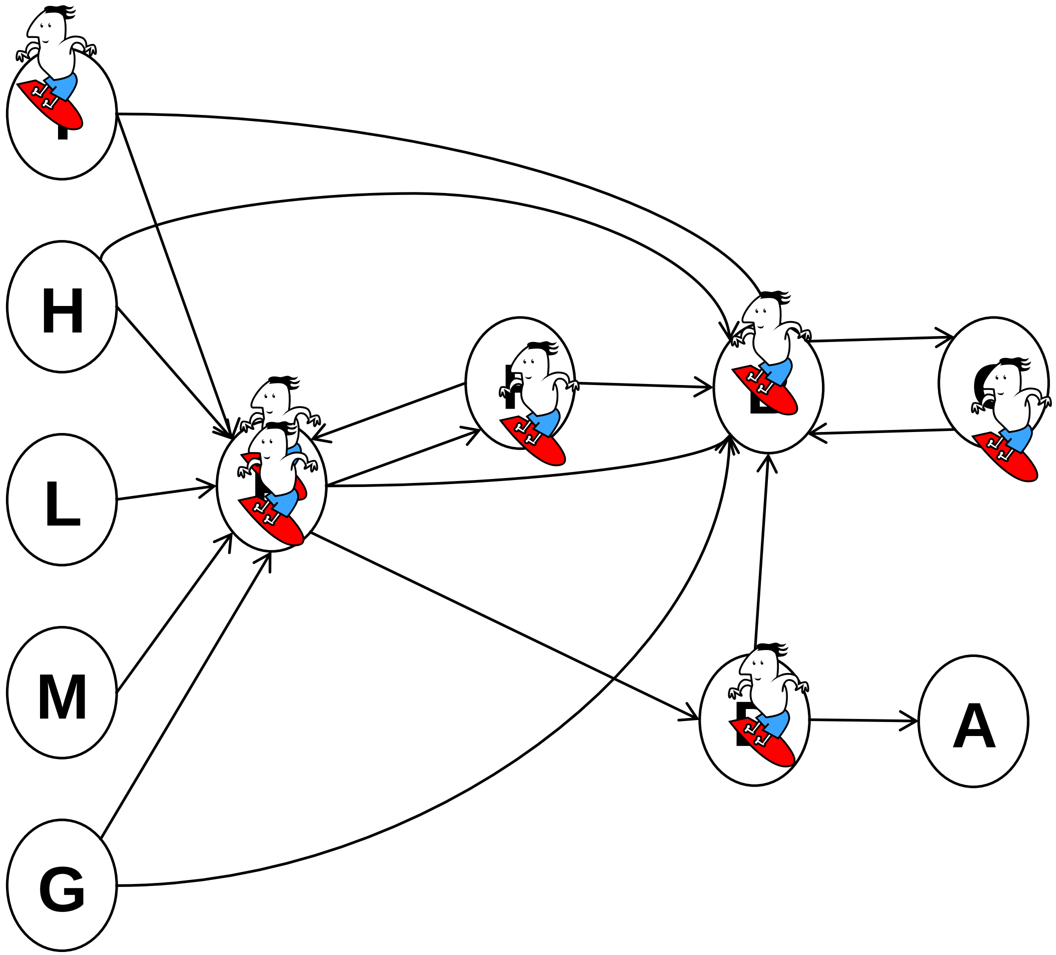
- Starts at arbitrary page



PageRank and the Random Surfer

Random Surfer

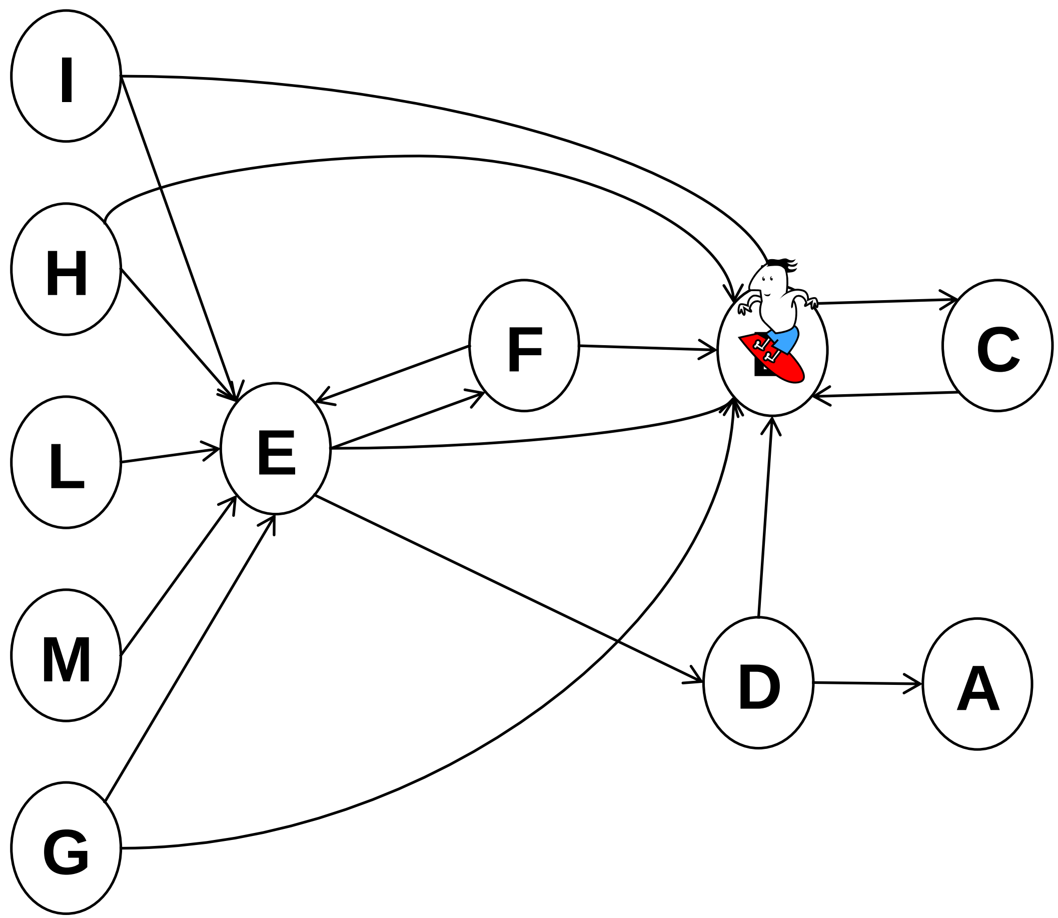
- Starts at arbitrary page
- Bounces from page to page by following links randomly



PageRank and the Random Surfer

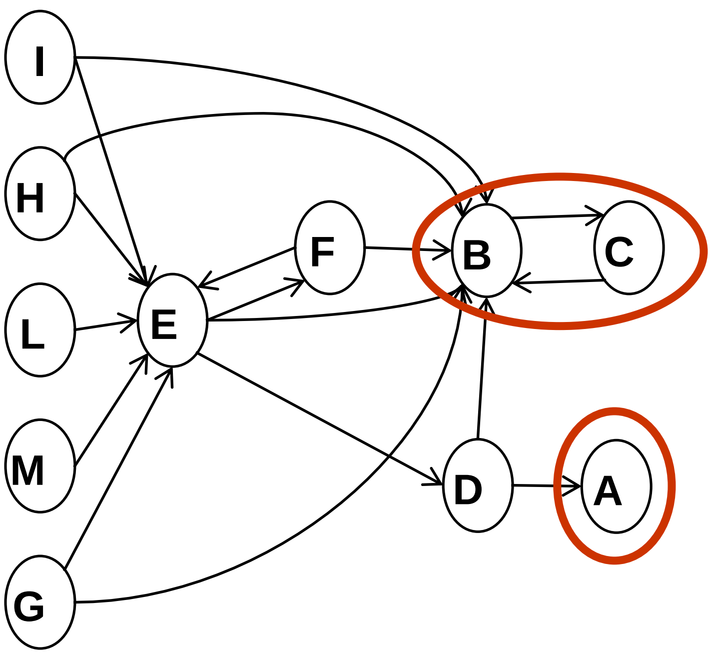
Random Surfer

- Starts at arbitrary page
- Bounces from page to page by following links randomly
- PageRank score of a web page is the relative number of time it is visited by the Random Surfer



But there are problems …

- Random Surfer gets trapped by dangling nodes! (no outlinks)
- Random Surfer gets trapped in buckets
  - Reachable strongly connected component without outlinks



Finally …

- Google matrix

  $$G = \alpha S + (1-\alpha) E$$

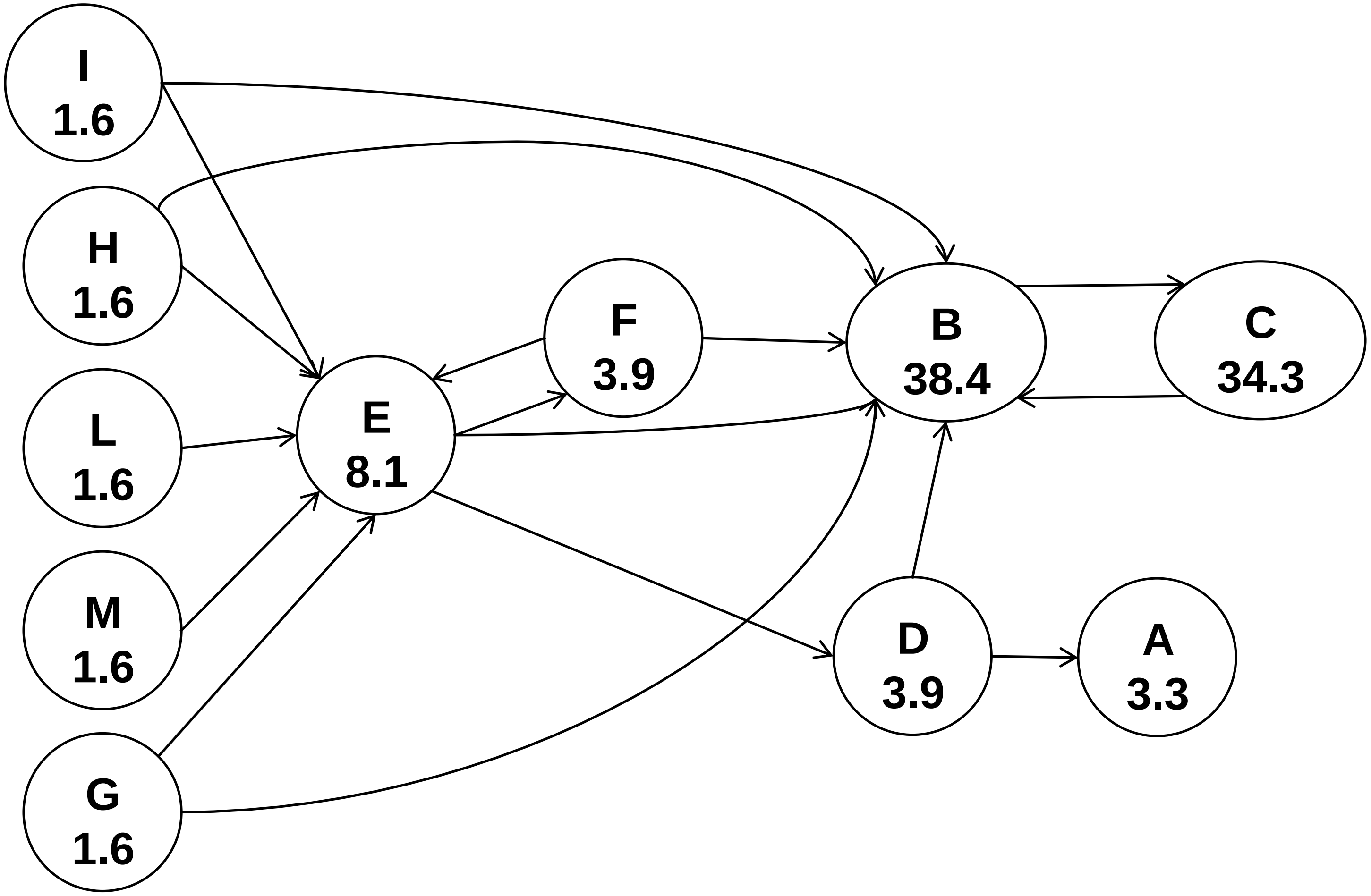
  - Where α is the damping factor
- Interpretation of G
  - With probability α, Random Surfer follows a hyperlink from a page (selected at random)
  - With probability 1-α, Random Surfer jumps to any page (e.g., by entering a new URL in the browser)

- PageRank scores are the solution of self-consistent equation

  $$\pi = \pi G$$

  $$= \alpha \pi S + (1-\alpha) u$$

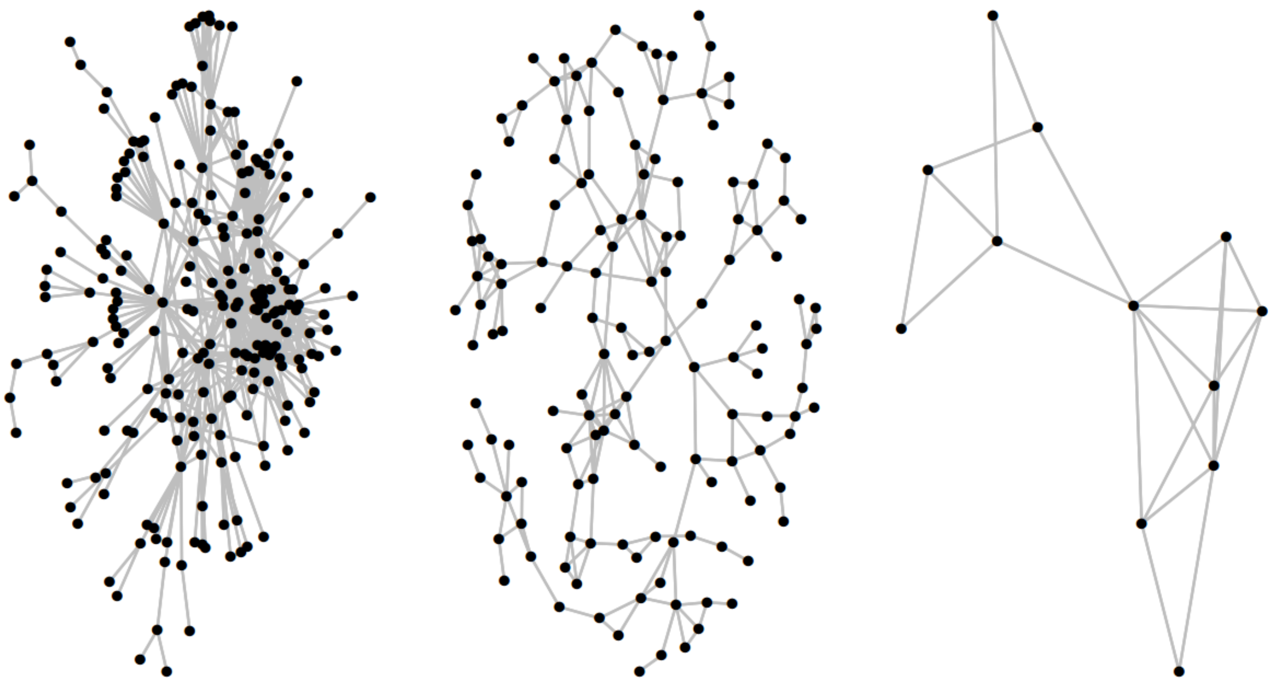
PageRank scores



Empirical study: Comparing centralization of different networks

Comparison of centralization metrics across three networks:

- butland ppi: binding interactions among 716 yeast proteins
- addhealth9: friendships among 136 boys
- tribes: positive relations among 12 NZ tribes



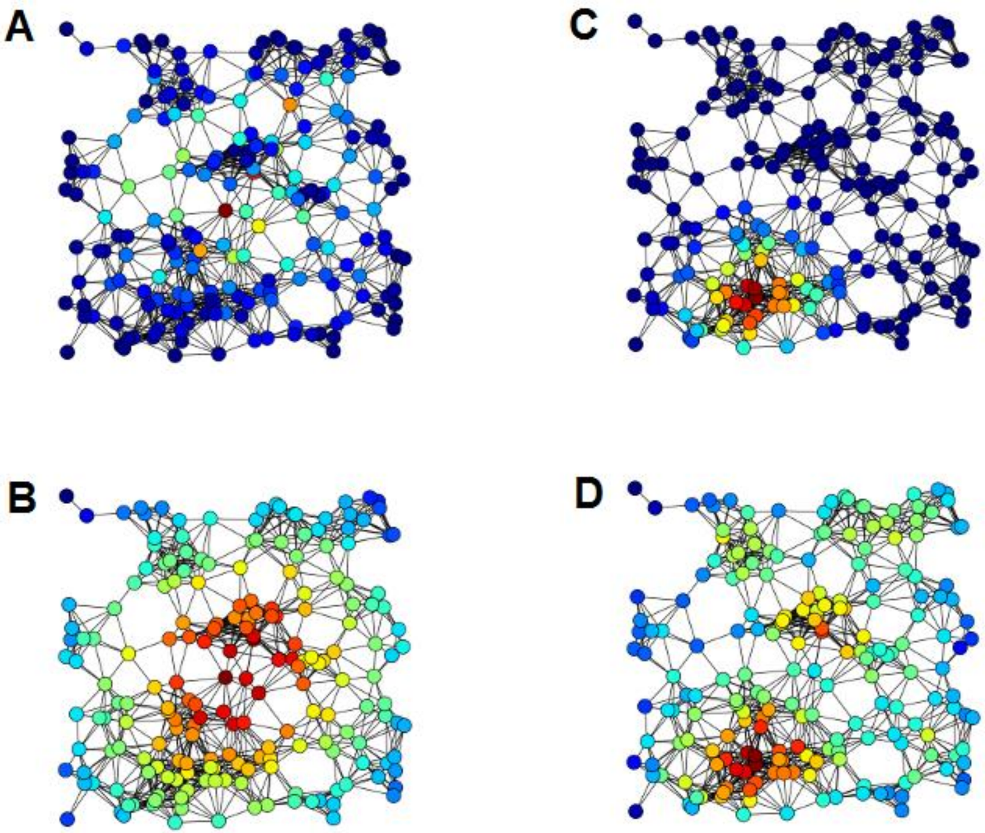
Empirical study: Comparing centralization of different networks

	degree	closeness	betweenness	eigenvector
ppi	0.13	0.26	0.31	0.35
addhealth	0.04	0.14	0.42	0.61
tribes	0.35	0.5	0.51	0.47

The protein network looks visually centralized, but

- most centralization is local;
- globally, somewhat decentralized.

The friendship network has small degree centrality (why?).
The tribes network has one particularly central node.



Examples of A) Betweenness centrality, B) Closeness centrality, C) Eigenvector centrality, D) Degree centrality, of the same graph.

Centralities in Python

```
import networkx as nx
import matplotlib.pyplot as plt
G=nx.read_edgelist("D:\\karate.txt")
nx.draw(G,with_labels = True)
plt.draw()
b = nx.edge_betweenness_centrality(G)
c = nx.closeness_centrality(G)
d = nx.degree_centrality(G)
e = nx.eigenvector_centrality(G)
k  = nx.katz_centrality(G)
```


Questions