University of Stuttgart
Germany

Complex Network Systems

Centrality metrics

Ilche Georgievski
ilche.georgievski@iaas.uni-stuttgart.de
Room: U38 0.353

2019/2020
Winter

Types

Graph-level metrics

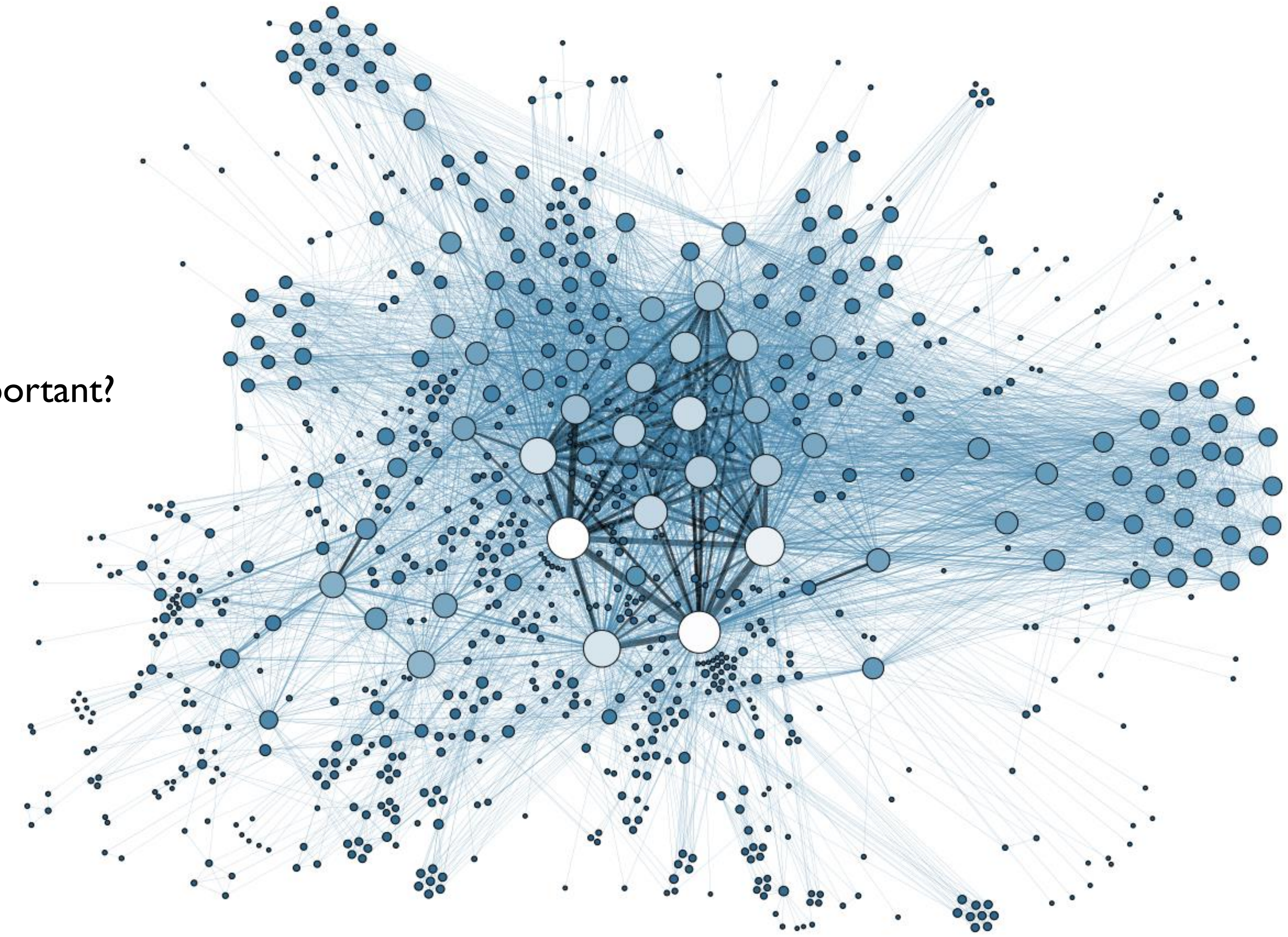
- Size
- Density
- Paths and distances
- Neighbourhoods
- Egocentric network
- Clustering coefficient
- Transitivity
- Cores
- Cliques
- Communities

Node-level metrics

- Closeness centrality
- Harmonic centrality
- Betweenness centrality
- Degree centrality
- Eigenvector centrality
- Katz centrality
- PageRank

Grandjean, M. (2014). La connaissance est un réseau. Les Cahiers du Numérique 10 (3): 37-54

Which nodes are important?



Why?

Centrality

- Which nodes are most *central*?
  - Definition of *central* varies by context/purpose

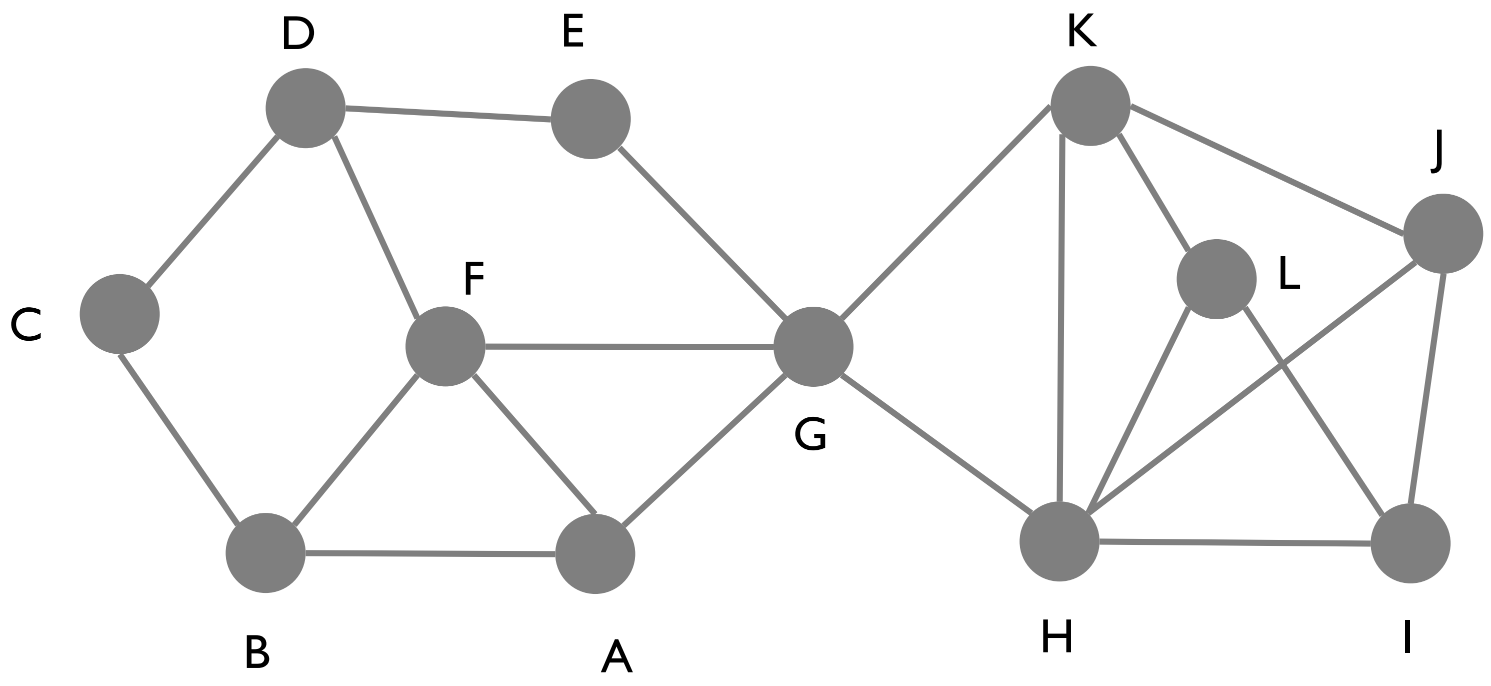
- Who is important based on their network position?

Centrality metrics

geometric	Closeness centrality
	Harmonic centrality
	Betweenness centrality
connectivity	Degree centrality
	Eigenvector centrality
	Katz centrality
	PageRank

What is closeness centrality, intuitively?

Closeness centrality



Which node is the most important one?

What is closeness centrality, intuitively?

The most important node is the one that is the most independent

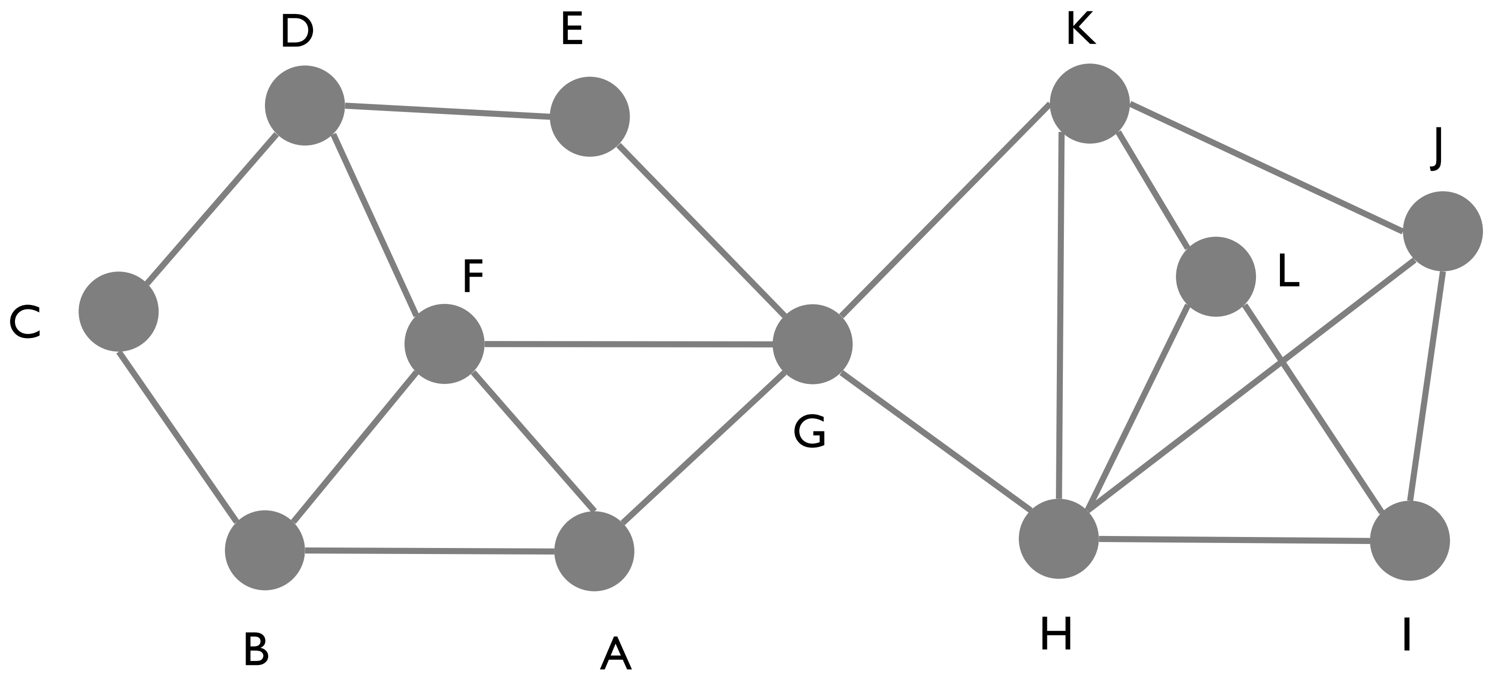
How close a node is to all other nodes

Closeness centrality

$$l_i = \frac{1}{n}\sum_j d_{ij} \quad \Longrightarrow \quad C_i = \frac{1}{l_i} \quad \Longrightarrow \quad C_i = \frac{n}{\sum_j d_{ij}}$$

Closeness centrality

$$C_i = \frac{n}{\sum_j d_{ij}} \qquad C_G = ? \quad C_F = ? \quad C_H = ?$$

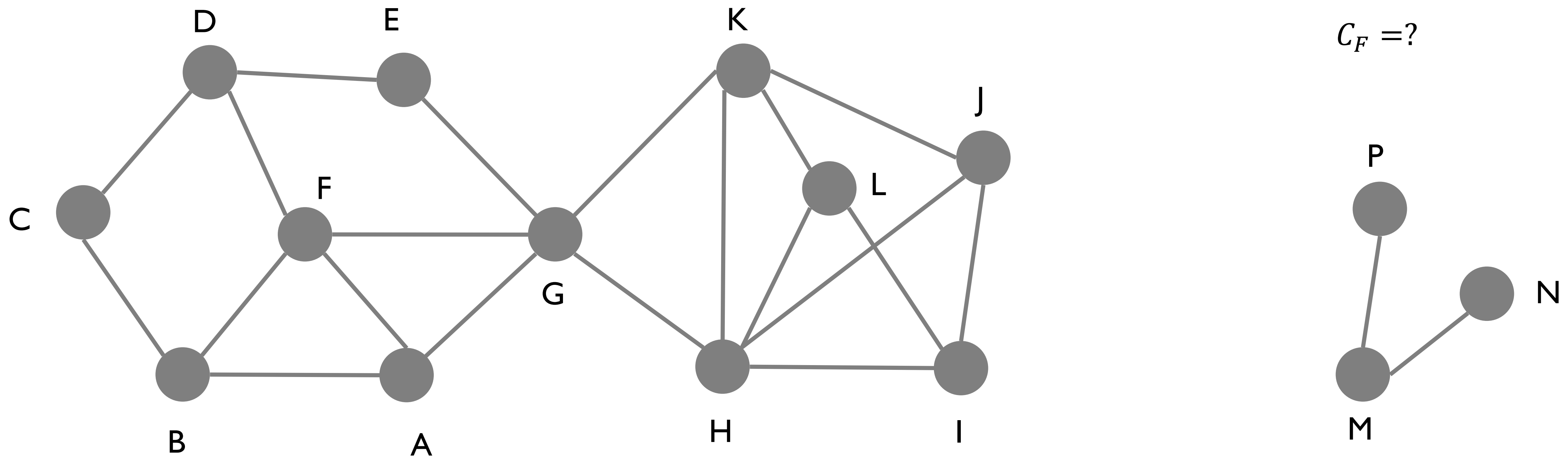


Node	d from G	d from F	d from H
A	1	1	2
B	2	1	3
C	3	2	4
D	4	1	3
E	1	2	2
F	1	0	2
G	0	1	1
H	1	2	0
I	2	3	1
J	2	3	1
K	1	2	1
L	2	3	1

Which nodes have the same closeness centrality?

$$C_G = \frac{12}{18} \quad C_F = \frac{12}{21} \quad C_H = \frac{12}{21}$$

Closeness centrality

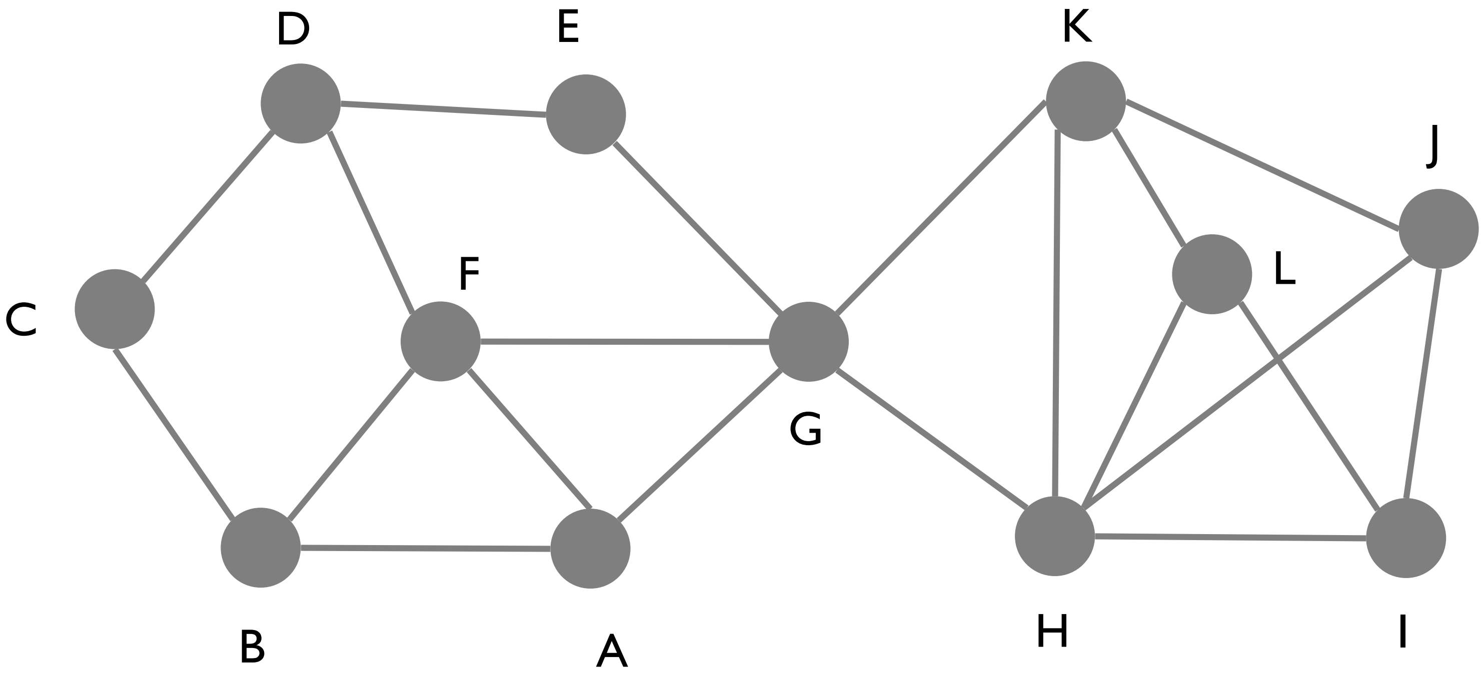


$C_F = ?$

What are the (two) practical problems with closeness centrality?

What would harmonic centrality be?

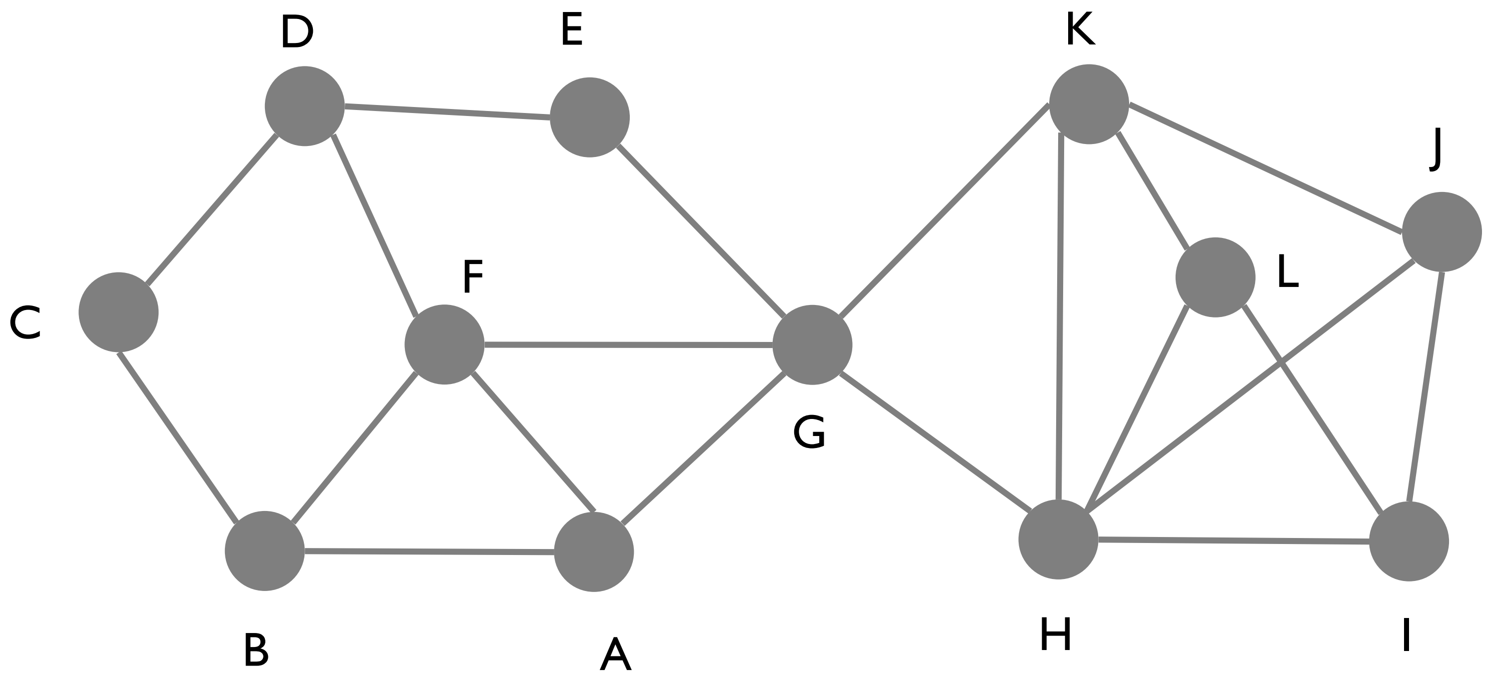
Harmonic centrality



Which node is the most important one?

Harmonic centrality

$$H_i = \frac{1}{n-1}\sum_{j(\neq i)} \frac{1}{d_{ij}} \qquad H_G = ? \qquad H_F = ? \qquad H_H = ?$$

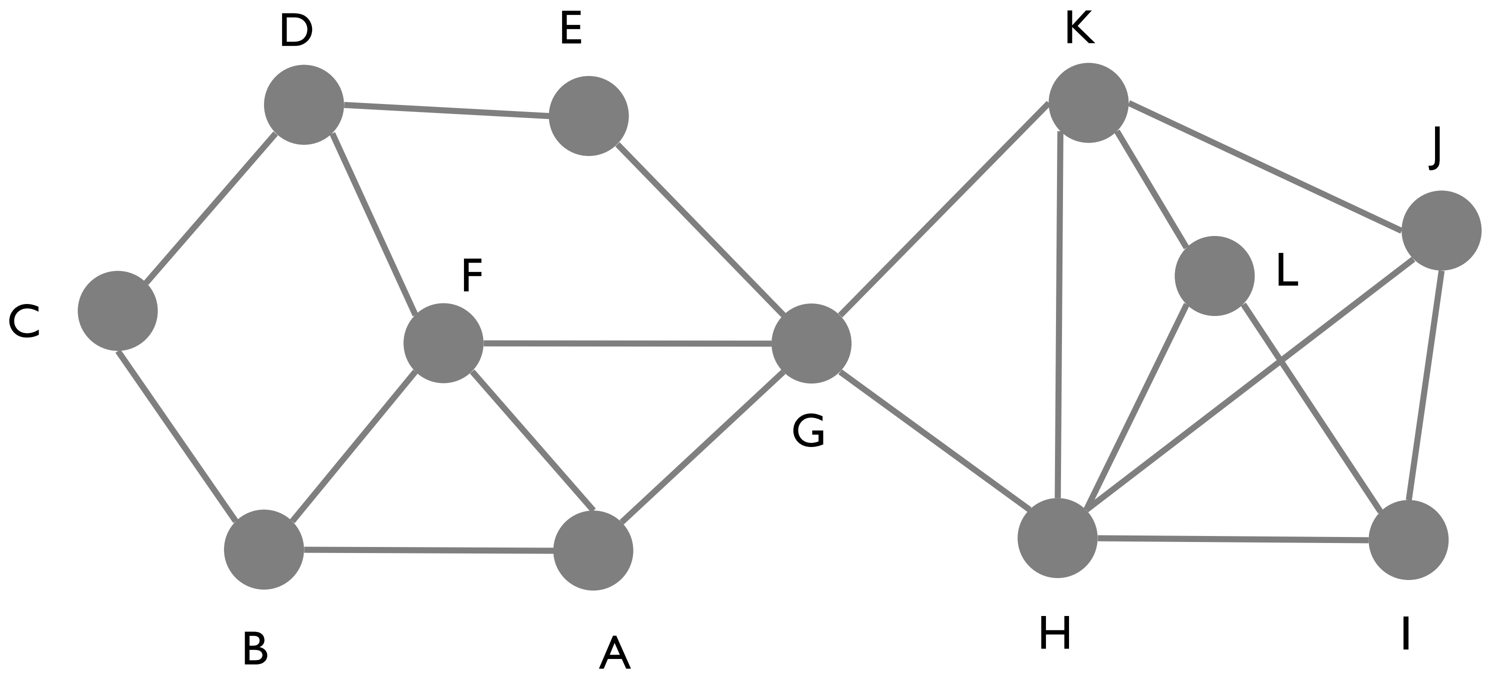


Node i	$1/(d$ from G)	$1/(d$ from F)	$1(d$ from H)
A	1	1	1/2
B	1/2	1	1/3
C	1/3	1/2	1/4
D	1/4	1	1/3
E	1	1/2	1/2
F	1	-	1/2
G	-	1	1
H	1	1/2	-
I	1/2	1/3	1
J	1/2	1/3	1
K	1	1/2	1
L	1/2	1/3	1

Which node has the second largest harmonic centrality?

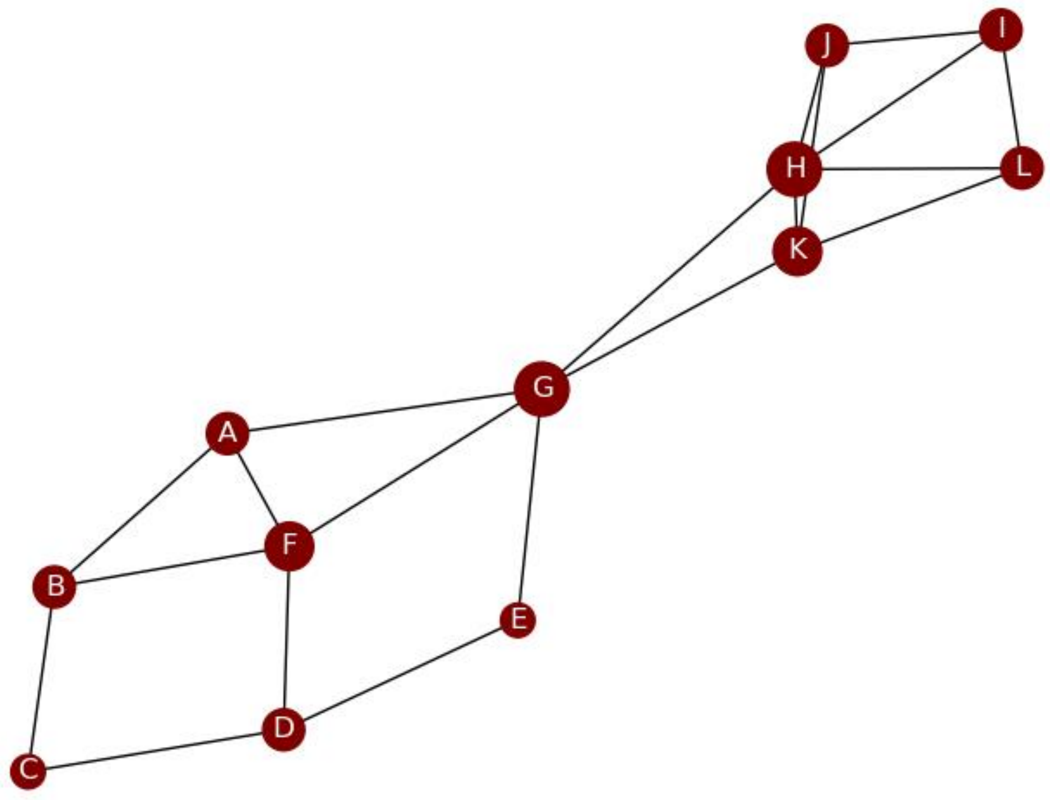
$$H_G = \frac{91}{132} \quad H_F = \frac{7}{11} \quad H_H = \frac{90}{132}$$

Closeness and harmonic centrality

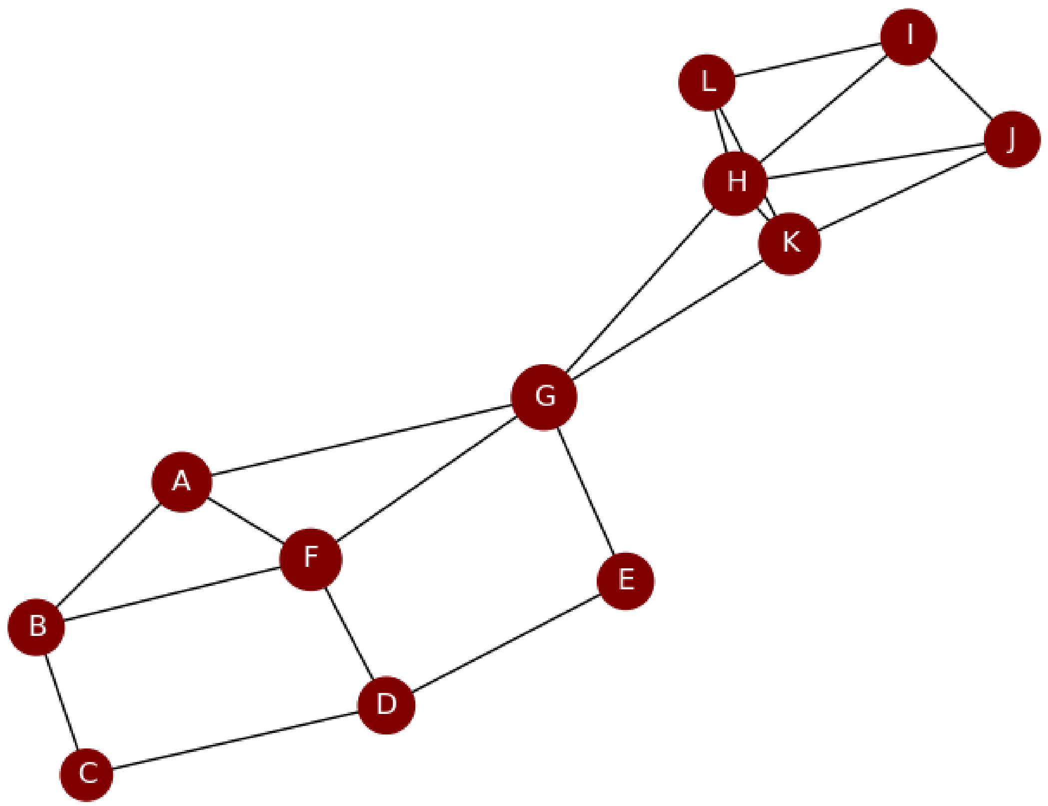


Node i	C_i	H_i
A	0.545	0.591
B	0.428	0.522
C	0.352	0.456
D	0.444	0.537
E	0.5	0.53
F	0.571	0.636
G	0.667	0.689
H	0.571	0.681
I	0.413	0.436
J	0.413	0.518
K	0.545	0.628
L	0.413	0.518

Degree

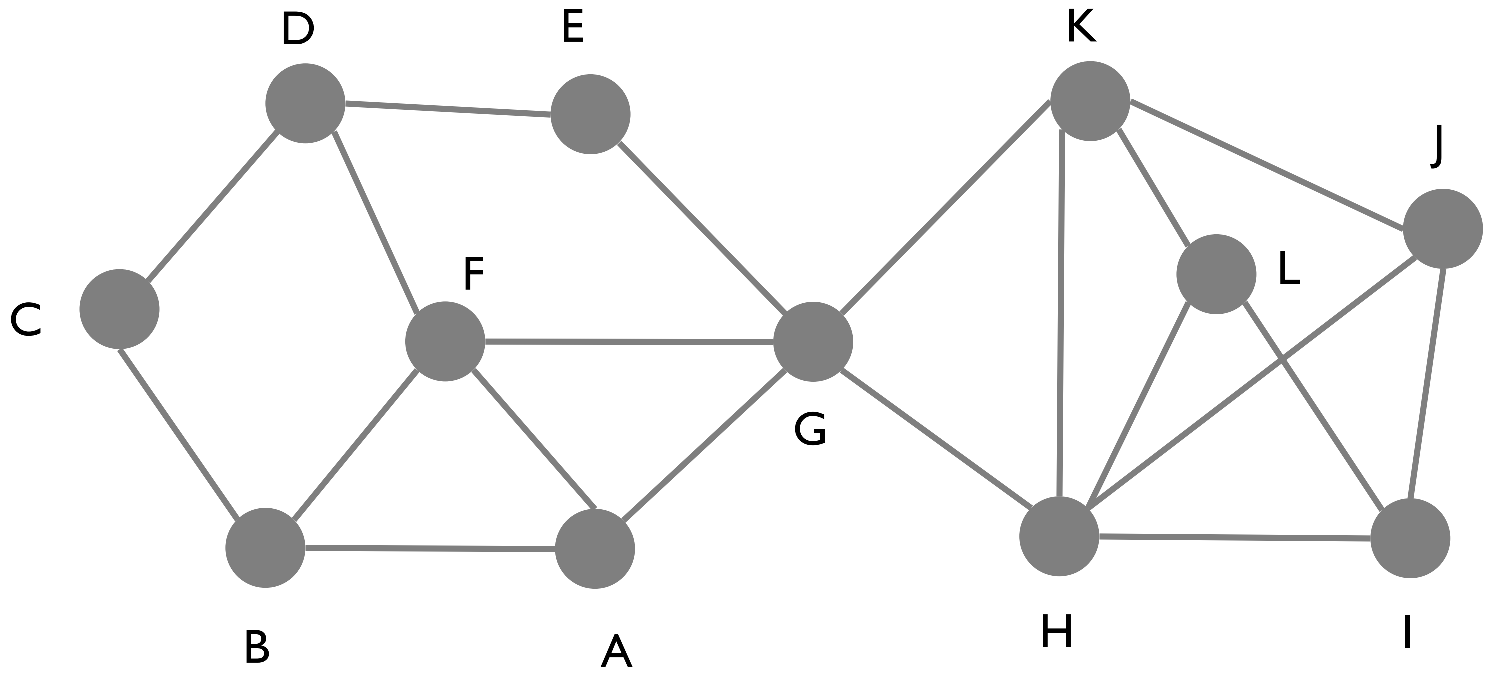


Harmonic centrality



What is betweenness centrality, intuitively?

Betweenness centrality



Which node is the most important one?

What is betweenness centrality, intuitively?

The most important node is the one with the most strategic location

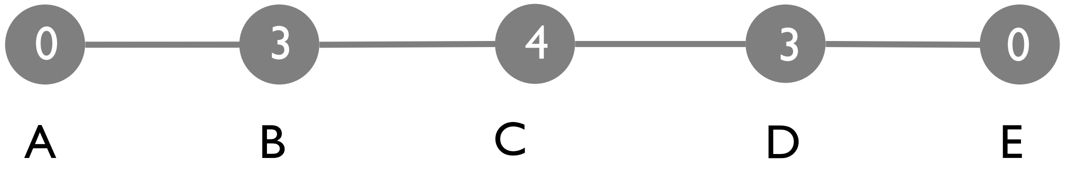
The frequency with which a node is on the shortest paths between all other nodes

Compute betweenness centrality

For node i

1. Select a pair of nodes
2. Find all the shortest paths between those nodes
3. Compute the fraction of those paths that include node i
4. Repeat steps 1-3 for every pair of nodes in the network
5. Sum up all the fractions computed

Betweenness centrality



Let's compute betweenness centrality for node B

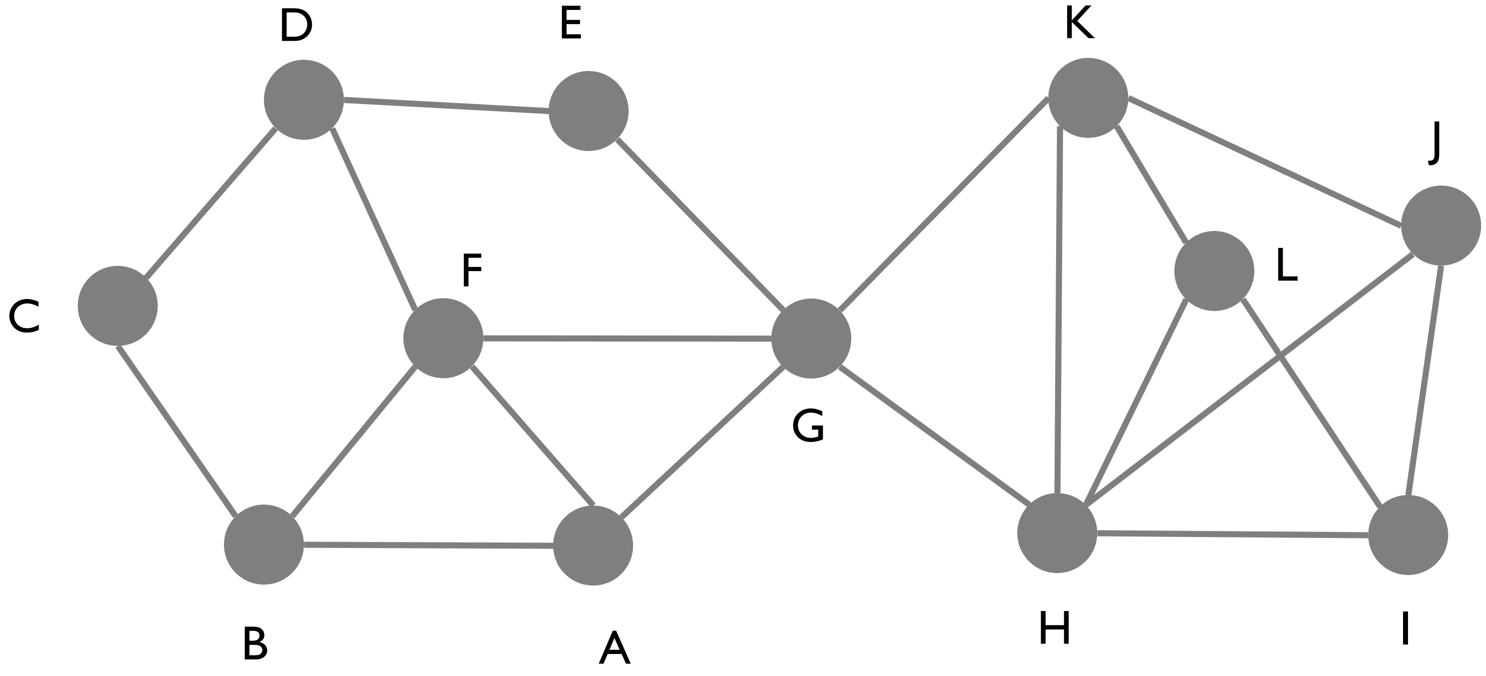
- There are 6 pairs to consider: AC, AD, AE, CD, CE, DE
- Fractions for AC, AD, AE are all 1
- Fractions for the remaining pairs are all 0
- $3 \times 1(A\ to\ all\ others) + 3 \times 0(all\ remaining\ pairs) = 3$
- Normalise the centrality: $3 \div 25 = 0.12$

Betweenness centrality

$$B_i = \sum_{jk} g_{jk}(i) \quad \Rightarrow \quad B_i = \sum_{jk} \frac{g_{jk}(i)}{g_{jk}} \quad \Rightarrow \quad B_i = \frac{1}{n^2} \sum_{jk} \frac{g_{jk}(i)}{g_{jk}}$$

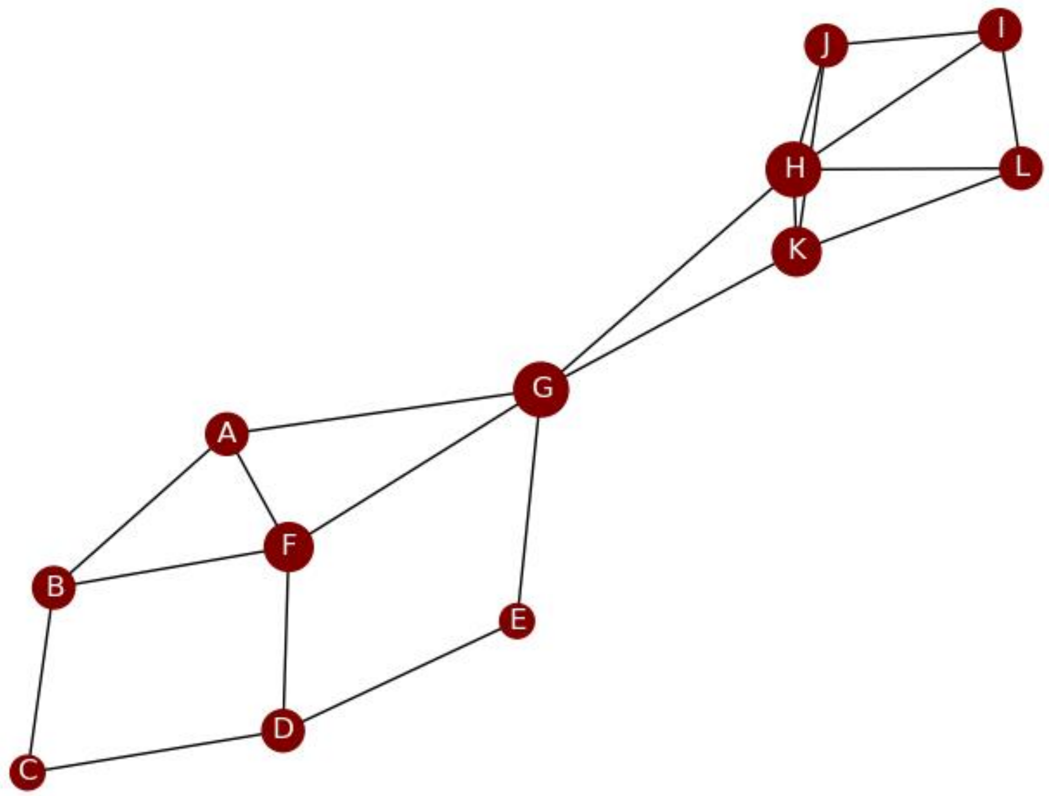
Betweenness centrality

$$B_i = \frac{1}{n^2} \sum_{jk} \frac{g_{jk}(i)}{g_{jk}}$$

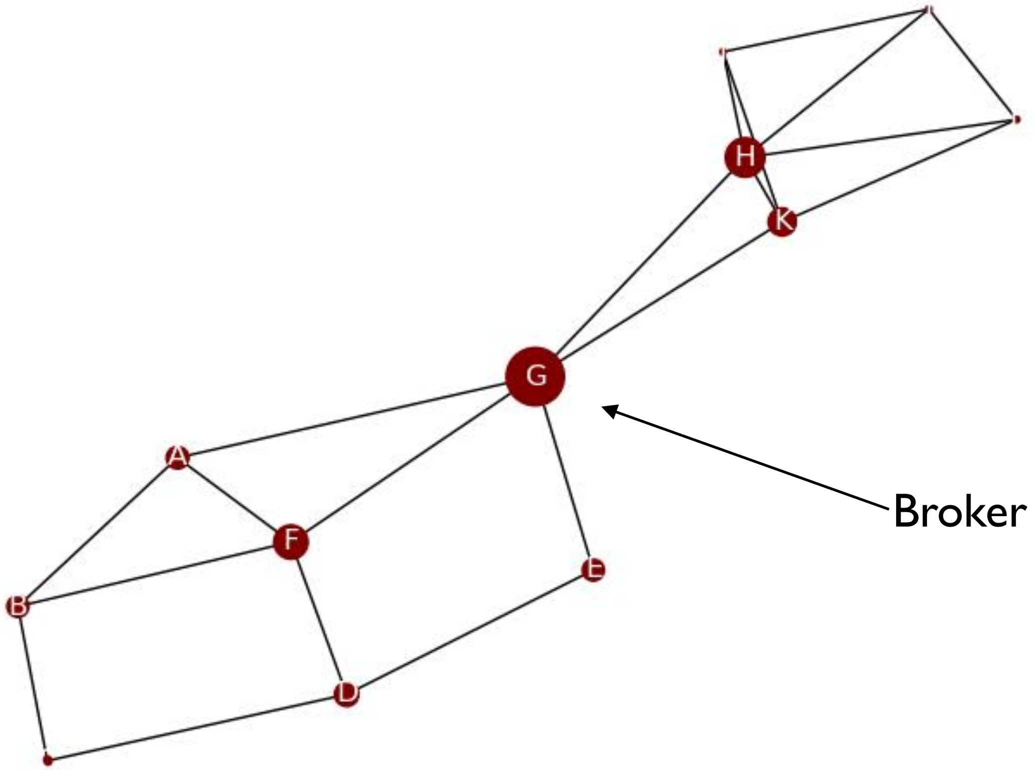


Node i	B_i
A	0.086
B	0.081
C	0.013
D	0.099
E	0.081
F	0.199
G	0.581
H	0.266
I	0.006
J	0.006
K	0.133
L	0.006

Degree

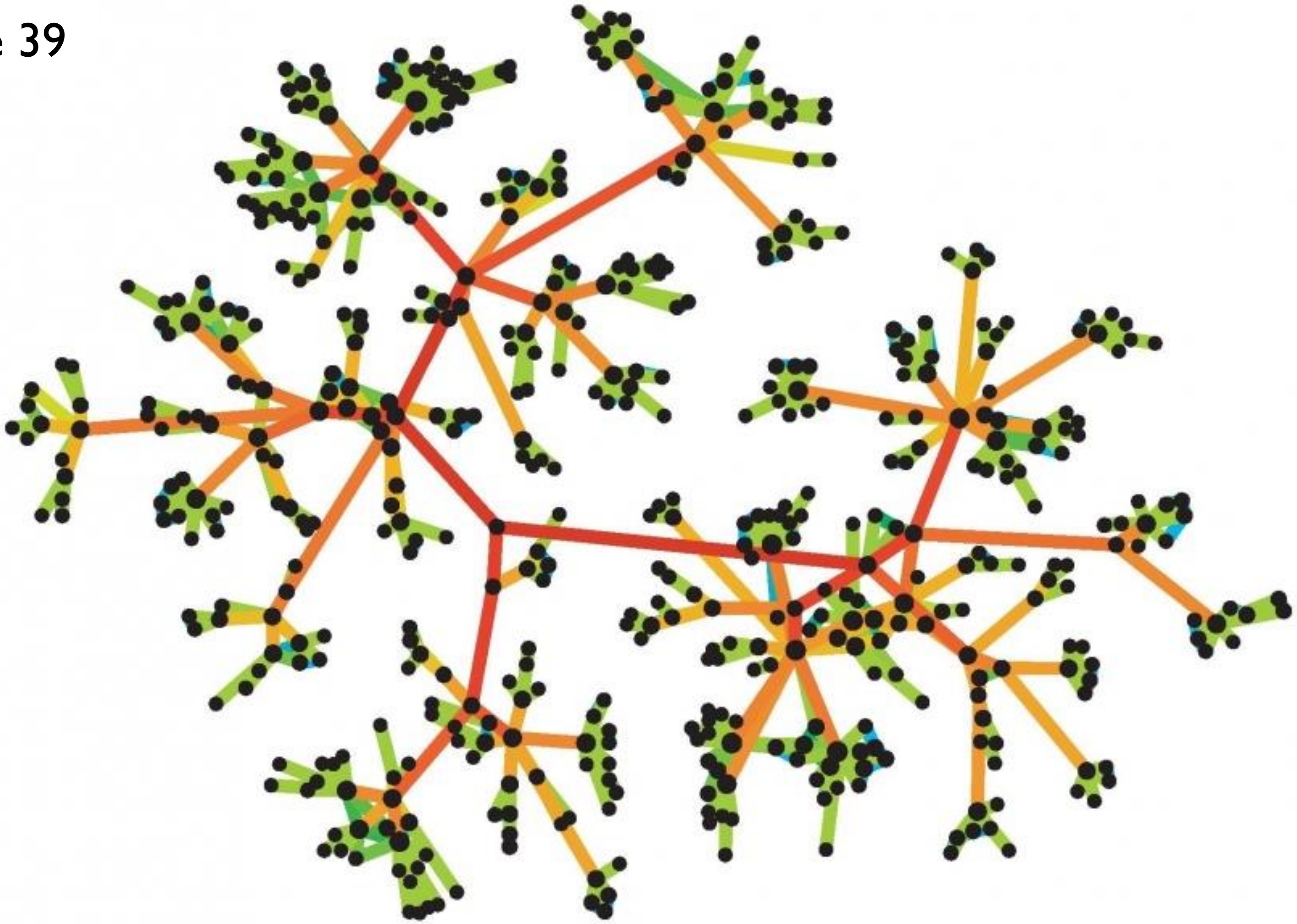


Betweenness centrality



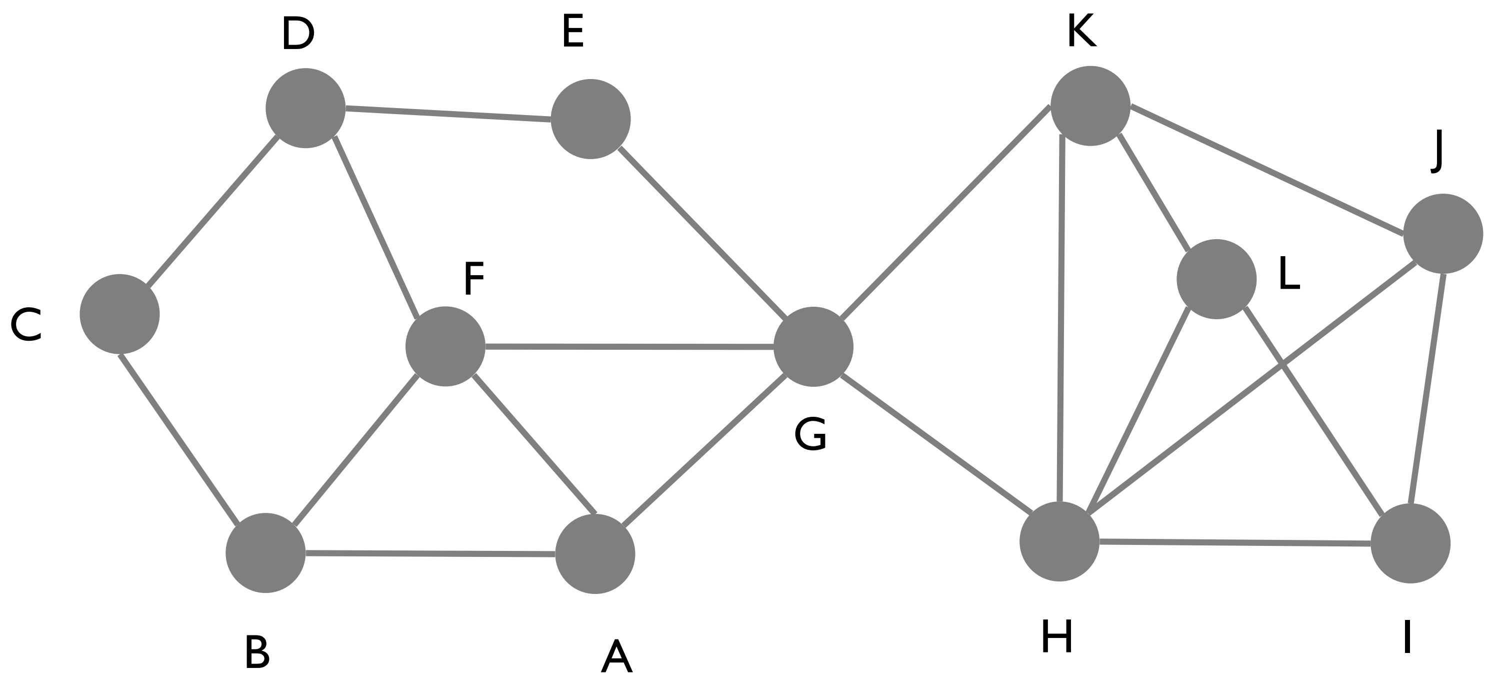
Communities and link weights

06 Communities, slide 39



What is degree centrality, intuitively?

Degree centrality



Which node is the most important one?

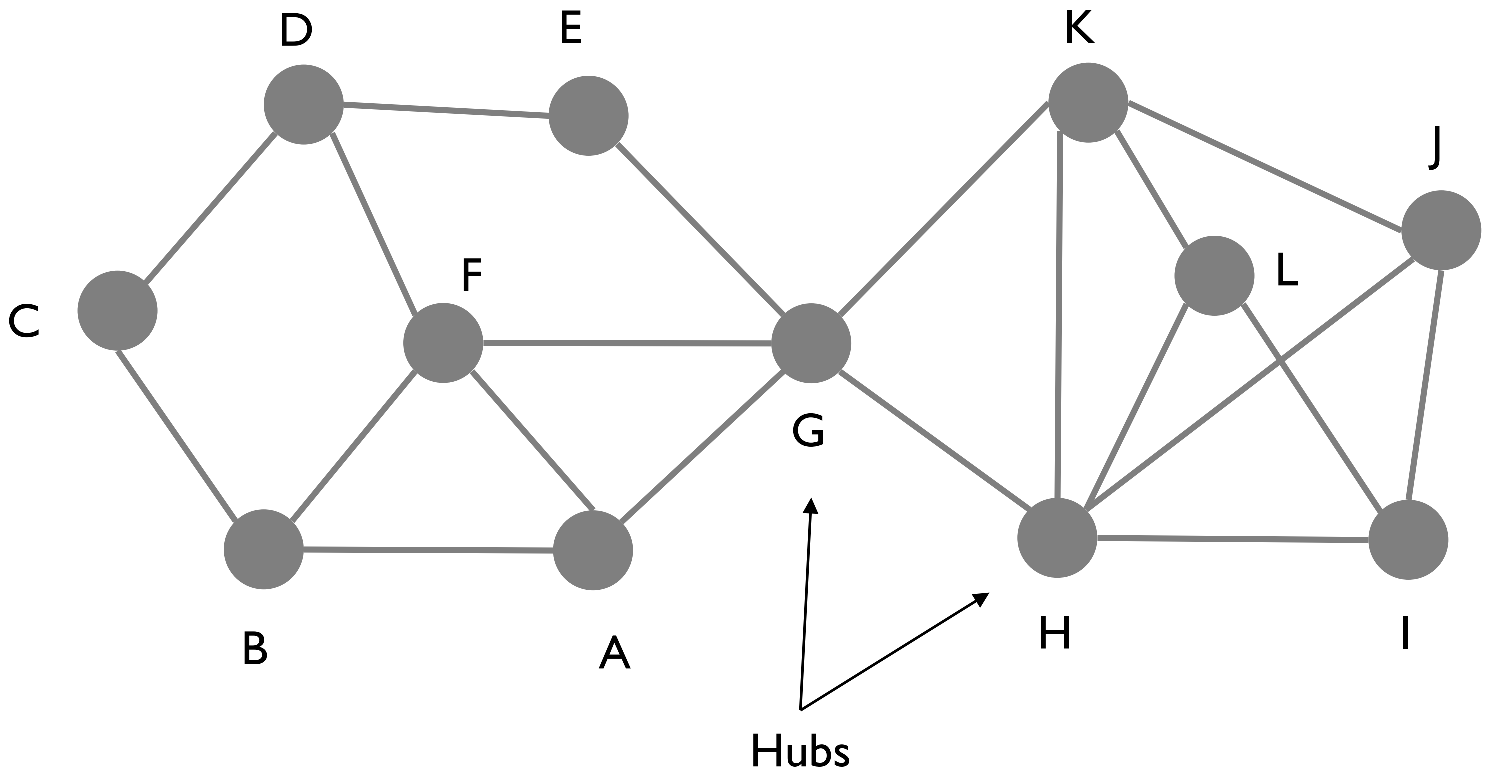
What is degree centrality, intuitively?

The most important node is the one with the most connections

It can be deceiving

Degree centrality

$\bar{k_i} = \frac{k_i}{n-1}$ – normalised degree centrality



Node	k_i	$\bar{k_i}$
A	3	0.272
B	3	0.272
C	2	0.181
D	3	0.272
E	2	0.181
F	4	0.363
G	5	0.454
H	5	0.454
I	3	0.272
J	3	0.272
K	4	0.363
L	3	0.272

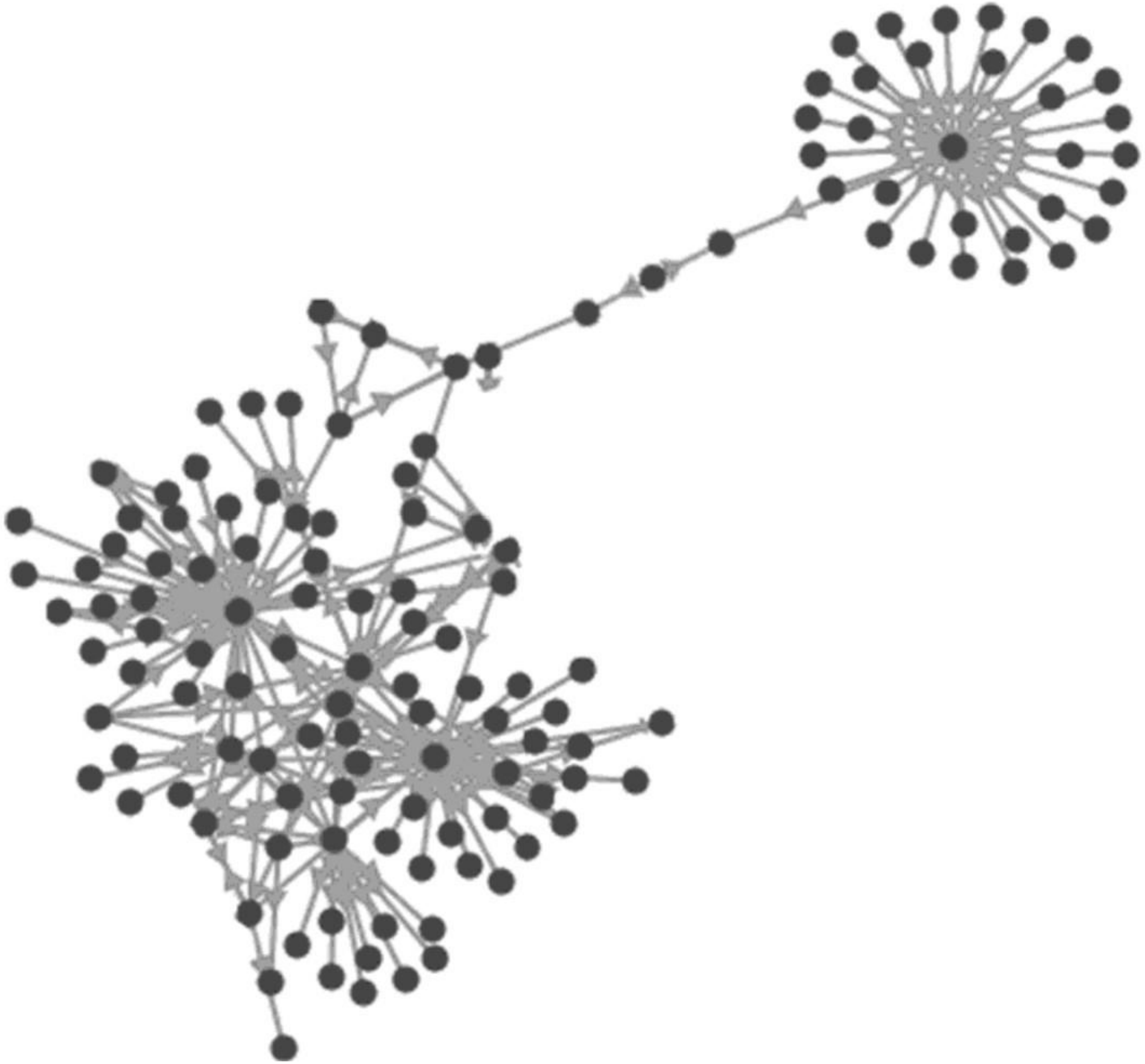
Comparison

Ordered nodes	C_i	H_i	B_i	$\overline{k_i}$
1st largest	G (0.667)	G (0.689)	G (0.581)	G (0.454)
2nd largest	F (0.571)	H (0.681)	H (0.266)	H (0.454)
3rd largest	H (0.571)	F (0.636)	F (0.199)	F (0.363)
4th largest	A (0.545)	K (0.628)	K (0.133)	K (0.363)

Comparison

Ordered nodes	C_i	H_i	B_i	$\overline{k_i}$
1st largest	G (0.667)	G (0.689)	G (0.581)	G (0.454)
2nd largest	F (0.571)	H (0.681)	H (0.266)	H (0.454)
3rd largest	H (0.571)	F (0.636)	F (0.199)	F (0.363)
4th largest	A (0.545)	K (0.628)	K (0.133)	K (0.363)

Golbeck, J. Analyzing the Social Web, Morgan Kaufmann, 2013.



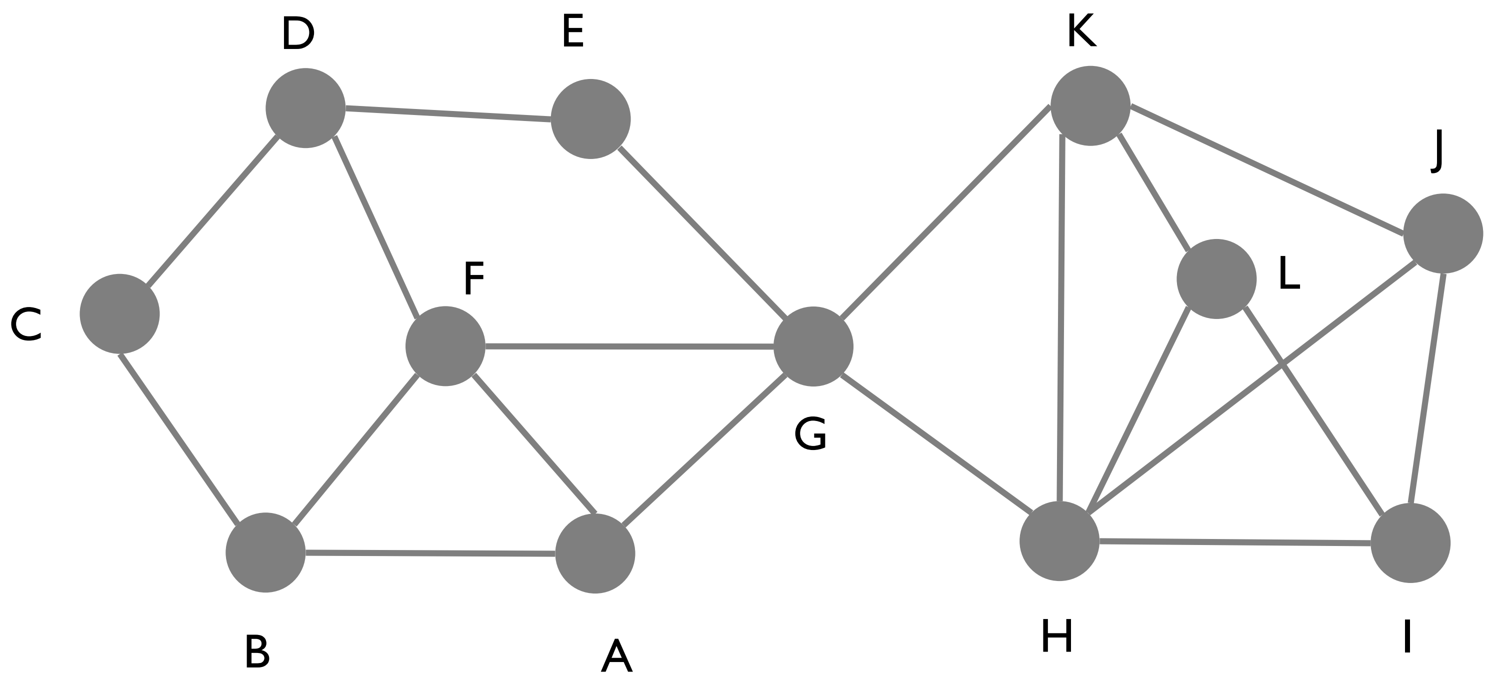
What kind of structure does this network have?

Generalisation

- Degree centrality captures only a local measure of importance
- Increase importance of nodes who are connected to other high-degree nodes

What is eigenvector centrality, intuitively?

Eigenvector centrality



Which node is the most important one?

What is eigenvector centrality, intuitively?

The most important node is the one that has the most information flowing through it

It cares if a node is a hub, but it also cares how many hubs it is connected to

How to compute the eigenvector centrality?

Power method

Method based on the largest eigenvalue

Eigenvector centrality

Element of the adjacency matrix

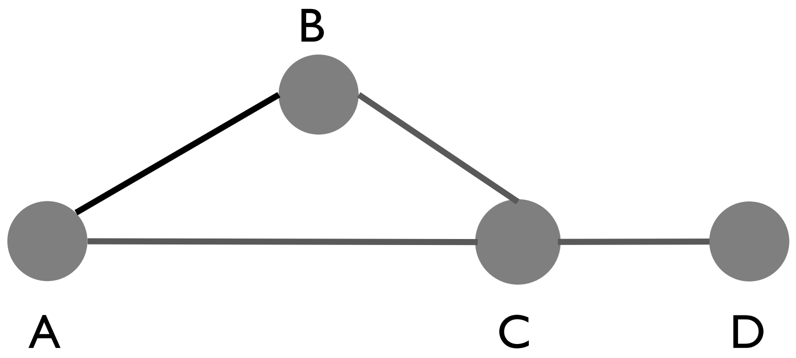
$$x_i^{(t+1)} = \sum_j A_{ij} x_j^{(t)}$$

Eigenvector at the previous step

$x_i^{(0)} = 1$ - initial condition for all i

Eigenvector at the current step

Find $x^{(1)}$

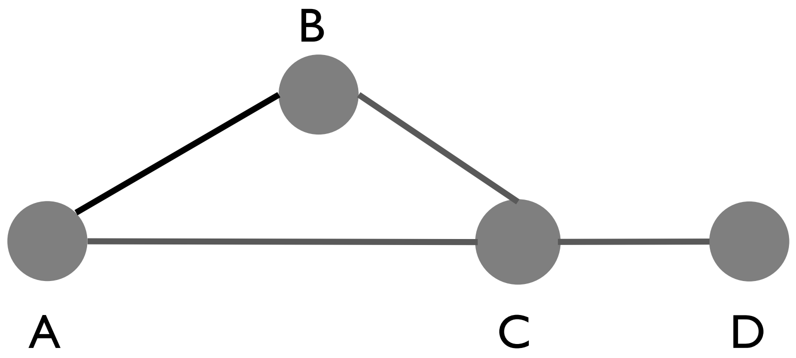


$$x_i^{(t+1)} = \sum_j A_{ij} x_j^{(t)}$$

$$x^{(1)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

What does it represent?

Find $x^{(2)}$



$$x_i^{(t+1)} = \sum_j A_{ij} x_j^{(t)}$$

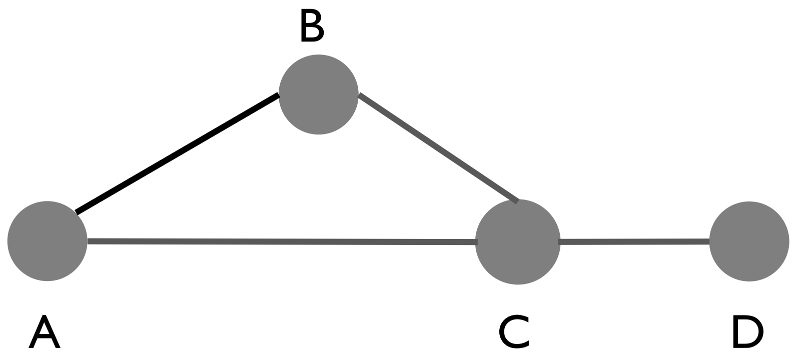
$$x^{(2)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 3 \end{bmatrix}$$

What does it represent?

Power method

- Start with some initial vector
  - $\mathbf{x}^{(0)} = 1$
- Iterate until the direction of the vector $\mathbf{x}$ stabilises or the maximum number of iterations has been reached
  - perform update $\mathbf{x}^{(t+1)} = \frac{\mathbf{A}\mathbf{x}^{(t)}}{\|\mathbf{A}\mathbf{x}^{(t)}\|}$

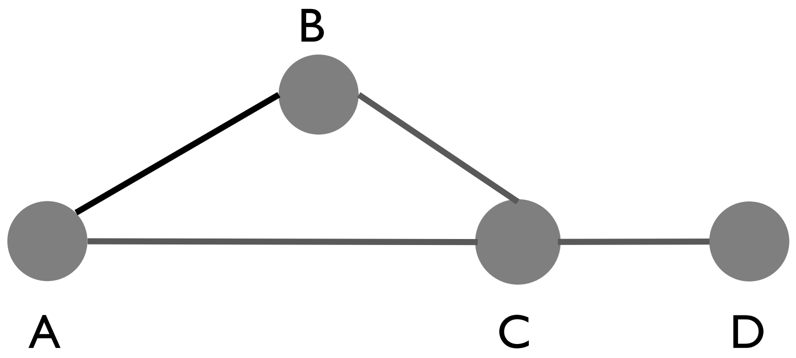
Find $\mathbf{x}^{(1)}$



$$\mathbf{x}^{(t+1)} = \frac{\mathbf{A}\mathbf{x}^{(t)}}{\|\mathbf{A}\mathbf{x}^{(t)}\|}$$

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.471 \\ 0.471 \\ 0.707 \\ 0.235 \end{bmatrix}$$

Find $\mathbf{x}^{(2)}$



$$\mathbf{x}^{(t+1)} = \frac{\mathbf{A}\mathbf{x}^{(t)}}{\|\mathbf{A}\mathbf{x}^{(t)}\|}$$

$$\mathbf{x}^{(2)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.471 \\ 0.471 \\ 0.707 \\ 0.235 \end{bmatrix} = \begin{bmatrix} 0.546 \\ 0.546 \\ 0.545 \\ 0.327 \end{bmatrix}$$

More iterations

Node i	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(10)}$	$\mathbf{x}^{(15)}$	$\mathbf{x}^{(20)}$
A	0.471	0.546	0.523	0.52275	0.52268	0.52267
B	0.471	0.546	0.523	0.52275	0.52268	0.52267
C	0.707	0.545	0.542	0.61117	0.61145	0.61154
D	0.235	0.327	0.284	0.28271	0.28234	0.28219

$\Uparrow$

Convergence at a fixed point

Comparison

Node i	$x^{(1)}$	$x^{(2)}$	$x^{(5)}$	$x^{(10)}$	$x^{(15)}$	$x^{(20)}$	k_i
A	0.471	0.546	0.523	0.52275	0.52268	0.52267	2
B	0.471	0.546	0.523	0.52275	0.52268	0.52267	2
C	0.707	0.545	0.542	0.61117	0.61145	0.61154	3
D	0.235	0.327	0.284	0.28271	0.28234	0.28219	1

Comparison

Node i	$x^{(1)}$	$x^{(2)}$	$x^{(5)}$	$x^{(10)}$	$x^{(15)}$	$x^{(20)}$	k_i
A	0.471	0.546	0.523	0.52275	0.52268	0.52267	2
B	0.471	0.546	0.523	0.52275	0.52268	0.52267	2
C	0.707	0.545	0.542	0.61117	0.61145	0.61154	3
D	0.235	0.327	0.284	0.28271	0.28234	0.28219	1

Perron-Frobenius theorem

Guarantees that when the network is an undirected, connected component, iterating

$$x_i^{(t+1)} = \sum_j A_{ij} x_j^{(t)}$$

will always converge on a fixed point equivalent to the principal eigenvector of the adjacency matrix

Eigenvector centrality

Largest eigenvalue of A

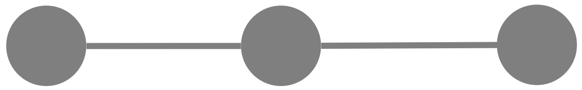
$$x_i = \sum_j A_{ij} x_j \quad \Rightarrow \quad A\mathbf{x} = \lambda_1 \mathbf{x}$$

Vector containing eigenvector centralities

Adjacency matrix

Can be solved using basic linear algebra

Example



$A\mathbf{x} = \lambda_1 \mathbf{x}$

$(A - \lambda_1 \mathrm{I})\mathbf{x} = 0$

1. Give the adjacency matrix A

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

2. Find the largest eigenvalue λ_1 of A

a. Find the determinant of the matrix $det(A - \lambda \mathrm{I})$

$$det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix}$$

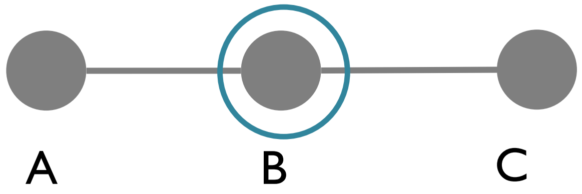
b. Solve for the values of λ such that $det(A - \lambda \mathrm{I}) = 0$

$$-\lambda^3 + 2\lambda = 0$$
$$\lambda(2 - \lambda^2) = 0$$

Eigenvalues are $(-\sqrt{2}, 0, \sqrt{2})$

c. Choose the largest eigenvalue $\lambda_1 = \sqrt{2}$

Example



$A\mathbf{x} = \lambda_1 \mathbf{x}$

$(A - \lambda_1 \mathrm{I})\mathbf{x} = 0$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

3. Find the eigenvector of A

$$\begin{bmatrix} -\sqrt{2} & 1 & 0 \\ 1 & -\sqrt{2} & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} -\sqrt{2}x_1 + x_2 = 0 \\ x_2 - \sqrt{2}x_3 = 0 \end{cases}$$

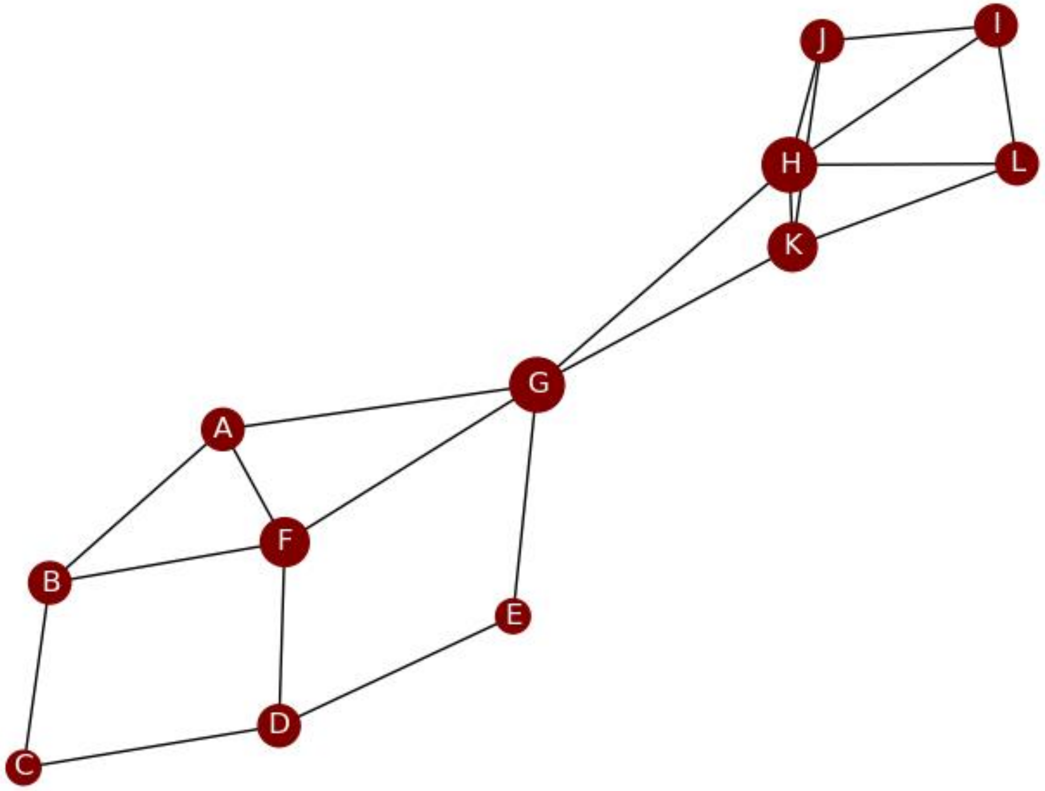
$$\mathbf{x} = \begin{bmatrix} 1/2 \\ \sqrt{2}/2 \\ 1/2 \end{bmatrix}$$

Eigenvector centrality

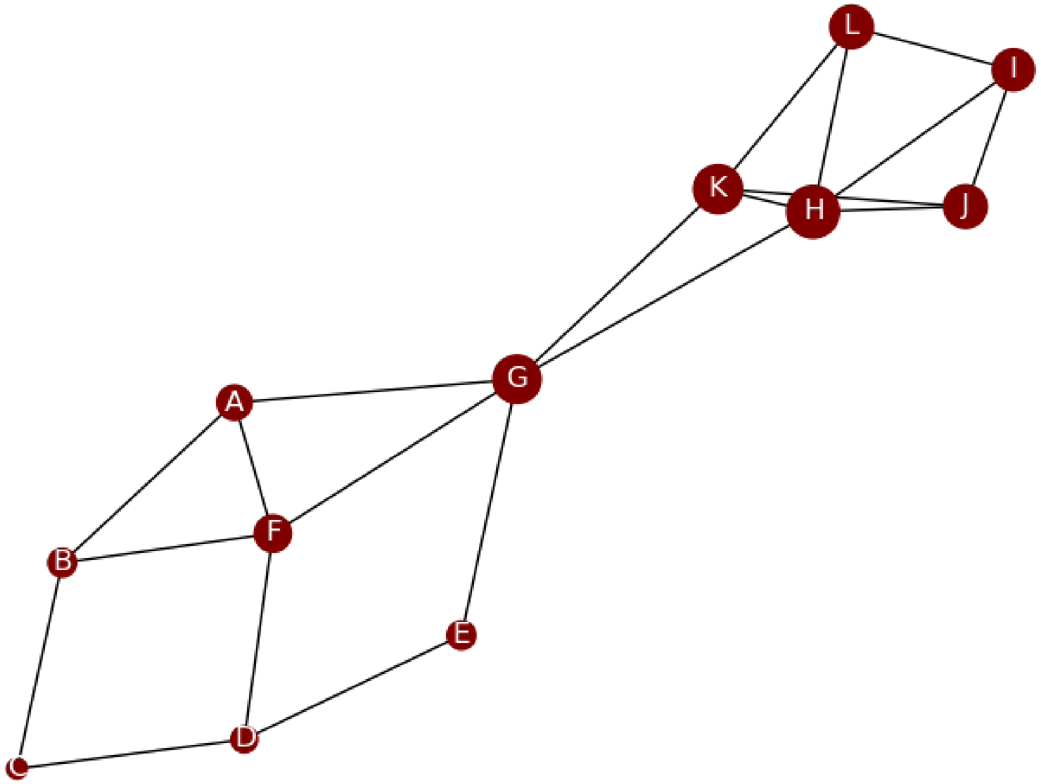
The eigenvector centrality of each node is given by the entries of the principal eigenvector

- Is the vector guaranteed to exist?
- Is it unique?
- Is the eigenvalue unique?
- Can we have negative entries in the eigenvector?

Degree



Eigenvector centrality

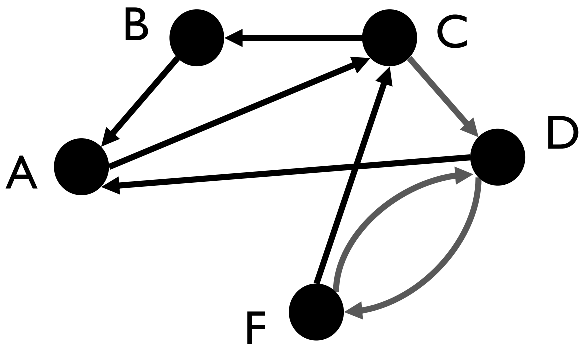


Comparison

Ordered nodes	C_i	H_i	B_i	$\overline{k_i}$	E_i
1st largest	G (0.667)	G (0.689)	G (0.581)	G (0.454)	H (0.473)
2nd largest	F (0.571)	H (0.681)	H (0.266)	H (0.454)	K (0.408)
3rd largest	H (0.571)	F (0.636)	F (0.199)	F (0.363)	G (0.396)
4th largest	A (0.545)	K (0.628)	K (0.133)	K (0.363)	J (0.320)

Exercise

Consider the following network:



a) Which node is the most important one? Why? (By intuition, no calculations)
b) Calculate the harmonic centrality of node D. Give the result in the form of a fraction
c) Using the adjacency matrix of the network, calculate $x^{(1)}$

Sources

- Zinoviev, D.. Complex Network Analysis in Python: Recognize, Construct, Visualize, Analyze, Interpret, The Pragmatic Bookshelf, 2018.
- Clauset, A. Network Analysis and Modeling, CSCI5352, Santa Fe Institute (2017), http://tuvalu.santafe.edu/~aaronc/courses/5352/
- Hagan, R., Feng, Y., and Bush, J., Centrality Metrics, University of Tennessee Knoxville, 2015.
- Golbeck, J. Analyzing the Social Web, Morgan Kaufmann, 2013.