University of Stuttgart
Germany

Complex Network Systems

Katz centrality and PageRank

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

2019/2020
Winter

Types

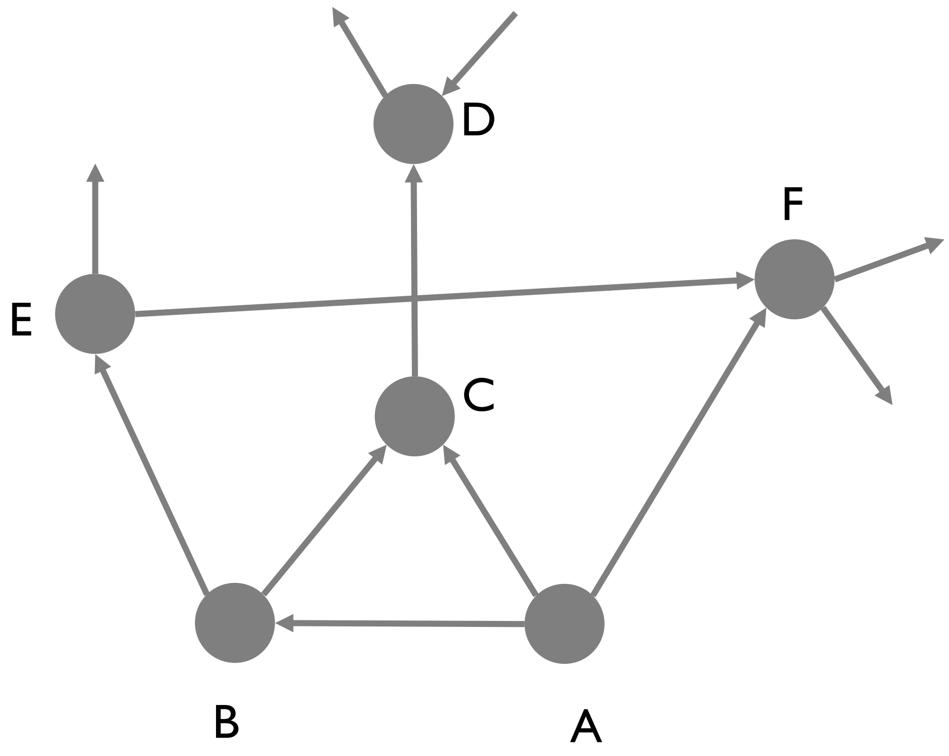
Graph-level metrics

- 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

What are the limitations of the eigenvector centrality?



What is the eigenvector centrality of A?

It has only outgoing edges

What about B?

It has an incoming edge, but it originates from A

Eigenvector centrality performs poorly for directed graphs

Issues with eigenvector centrality

- Poor performance in directed networks
- Zero centrality for all nodes not within a strongly connected component
- Only nodes with no outgoing edges will have non-zero centrality in a directed acyclic graph

Katz centrality

Eigenvector centrality

$$x_i = \alpha \boxed{\sum_j A_{ij} x_j} + \beta$$

controlling constant

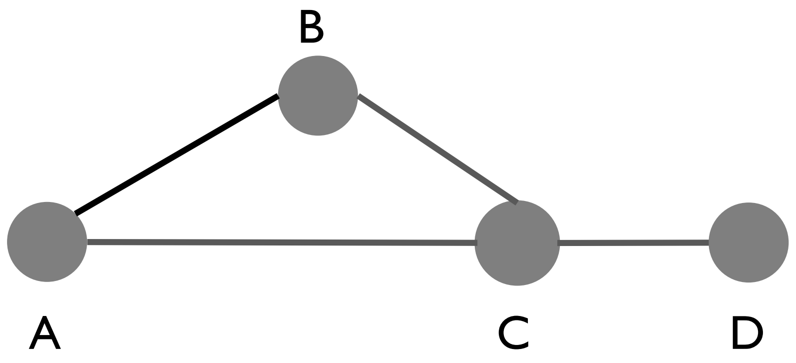
bias or free or uniform term

Matrix form

$$\mathbf{x} = \beta(I - \alpha A)^{-1}\mathbf{1}$$

in practice, we select $\alpha < 1/\lambda_1$

Example



$$\mathbf{x} = \beta(I - \alpha A)^{-1}\mathbf{1}$$

1. Give the adjacency matrix A

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

2. Find the largest eigenvalue λ_1 of A

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

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

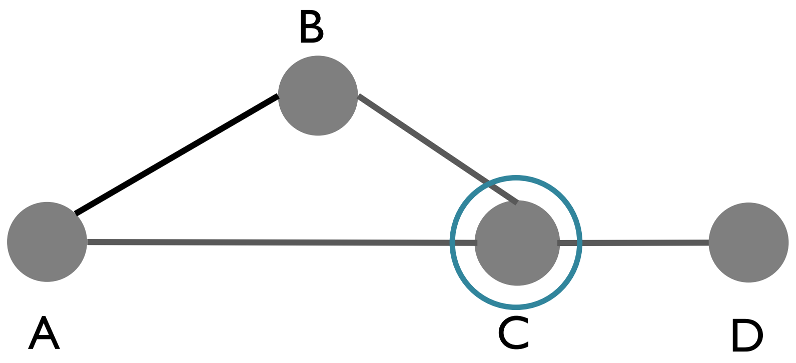
b. Solve for the values of λ such that $det(A - \lambda I) = 0$

$$\lambda^4 - 4\lambda^2 - 2\lambda + 1 = 0$$

Eigenvalues are $(-1.481, -1, 0.311, 2.17)$

c. Choose the largest eigenvalue $\lambda_1 = 2.17$

Example



$$\mathbf{x} = \beta(I - \alpha A)^{-1}\mathbf{1}$$

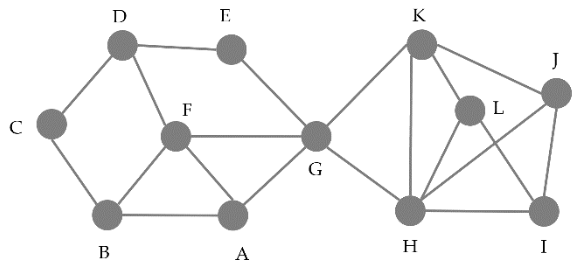
$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

4. Choose $\alpha = 0.3 < \frac{1}{2.17}$ and $\beta = 0.2$

5. Find the vector $\mathbf{x}$

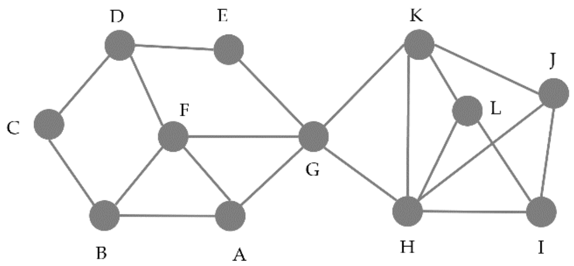
$$B^{-1} = \frac{1}{|B|}adj(B) = \frac{1}{|B|}C^{\mathrm{T}}$$

$$\mathbf{x} = 0.2\begin{bmatrix} 1 & -0.3 & -0.3 & 0 \\ -0.3 & 1 & -0.3 & 0 \\ -0.3 & -0.3 & 1 & -0.3 \\ 0 & 0 & -0.3 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.566 \\ 0.566 \\ 0.664 \\ 0.391 \end{bmatrix}$$



Katz centrality

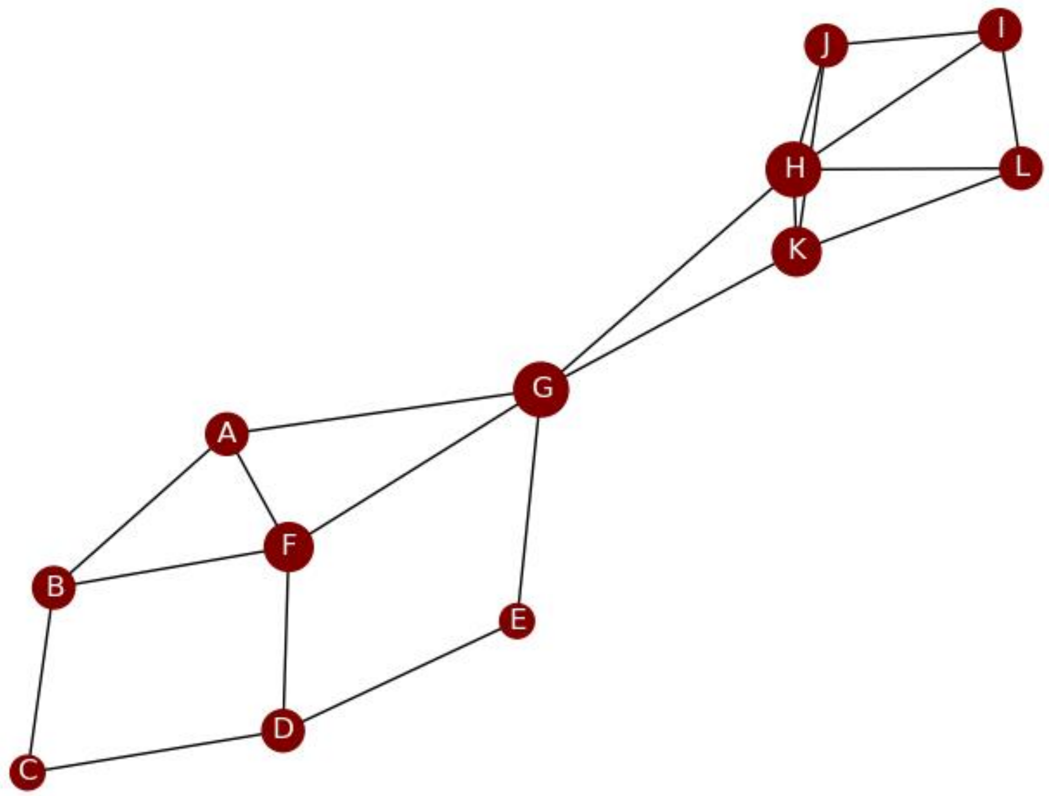
Node	$\alpha = 0.1, \beta = 1$	$\alpha = 0.1, \beta = 0.2$	$\alpha = 0.25, \beta = 1$	$\alpha = 0.25, \beta = 0.2$
A	0.28005441548424376	0.2800544425119089	0.23527499465958635	0.23527510895687903
B	0.27146816944113733	0.27146822842480306	0.1770764296806536	0.17707659423373143
F	0.30445376402396385	0.30445378929799183	0.2671285652976387	0.26712870633309543
G	0.33717926700590617	0.3371792248640899	0.39944214323710814	0.3994421499265649
C	0.24273042740923884	0.24273050844241534	0.10844929472211712	0.10844946003257827
D	0.2683928838971183	0.26839295125956253	0.15926786188132971	0.15926802966550838
E	0.24930152384627158	0.24930157206338463	0.16404071718333144	0.16404081977389845
H	0.3379463667326661	0.3379462867389788	0.4457607788670669	0.4457606513556881
K	0.31259400522413516	0.3125939459779813	0.3881107404786109	0.38811064281881286
I	0.2788760991930949	0.27887607957357974	0.28825021378324206	0.28825015020315803
J	0.28168592165838435	0.2816858935929373	0.30489363383198387	0.304893560385715
L	0.28168592165838435	0.2816858935929373	0.30489363383198387	0.304893560385715



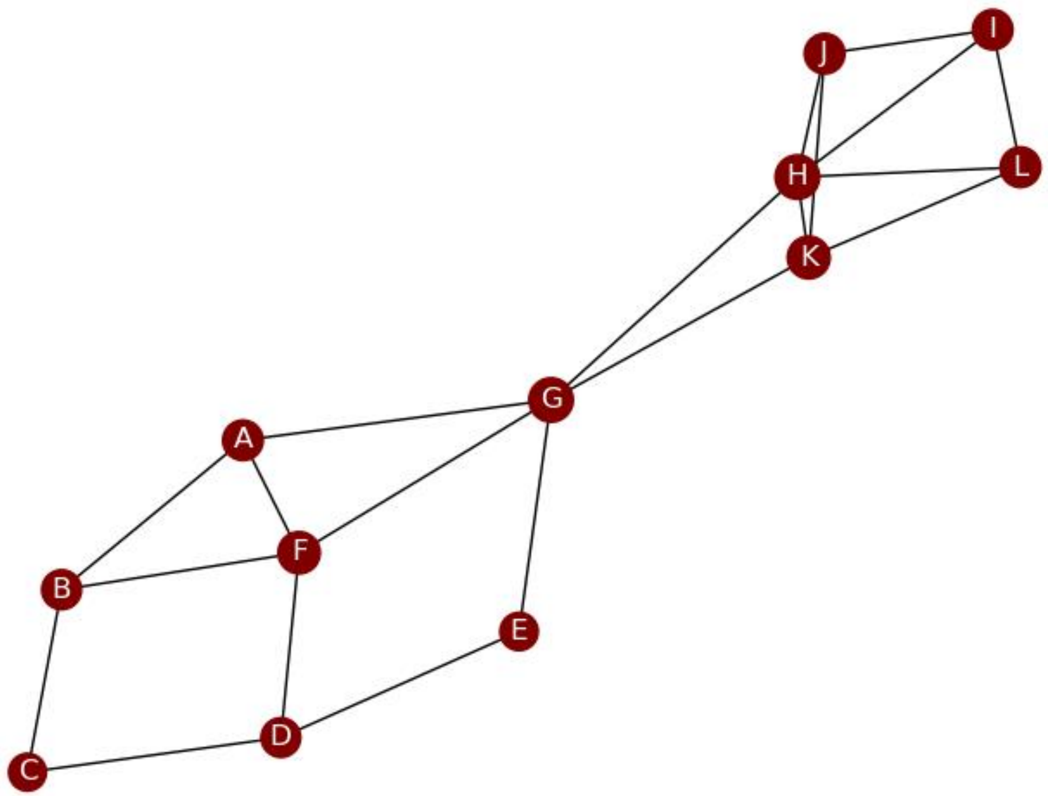
Comparison

Ordered nodes	Eigenvector centrality	Katz centrality
1st largest	H (0.473)	H (0.3379)
2nd largest	K (0.408)	G (0.3371)
3rd largest	G (0.396)	K (0.312)
4th largest	J (0.320)	F (0.304)

Degree

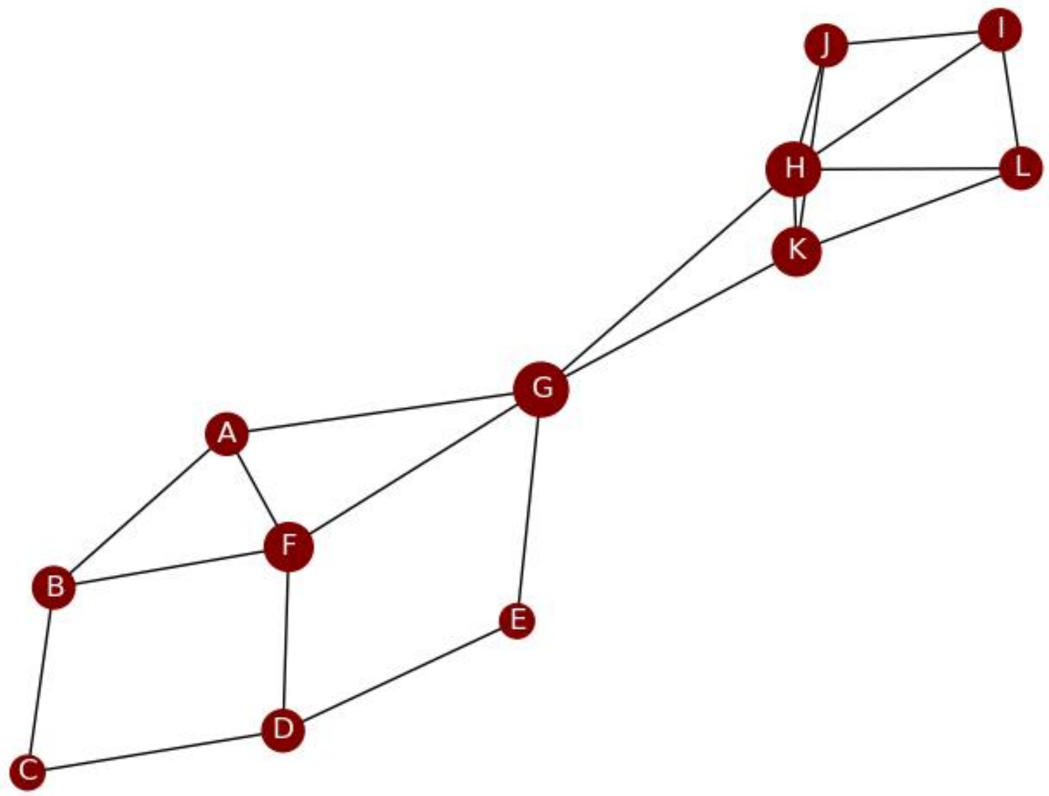


Katz centrality

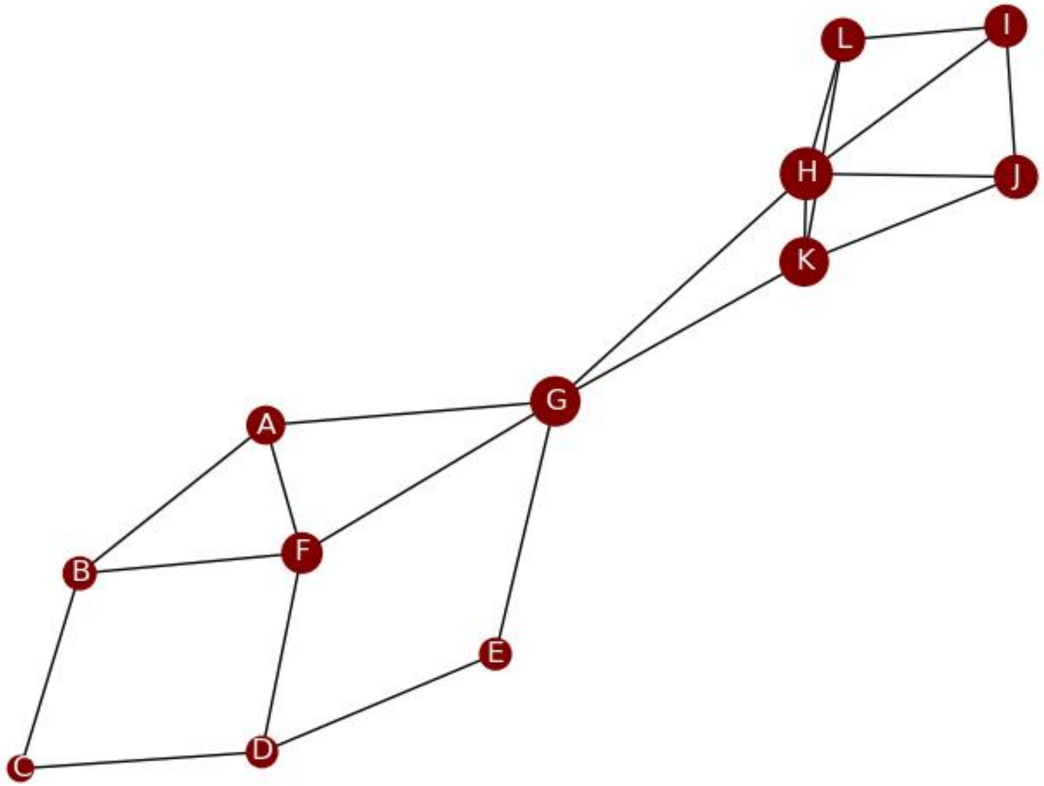


$\alpha = 0.1, \beta = 1$

Degree

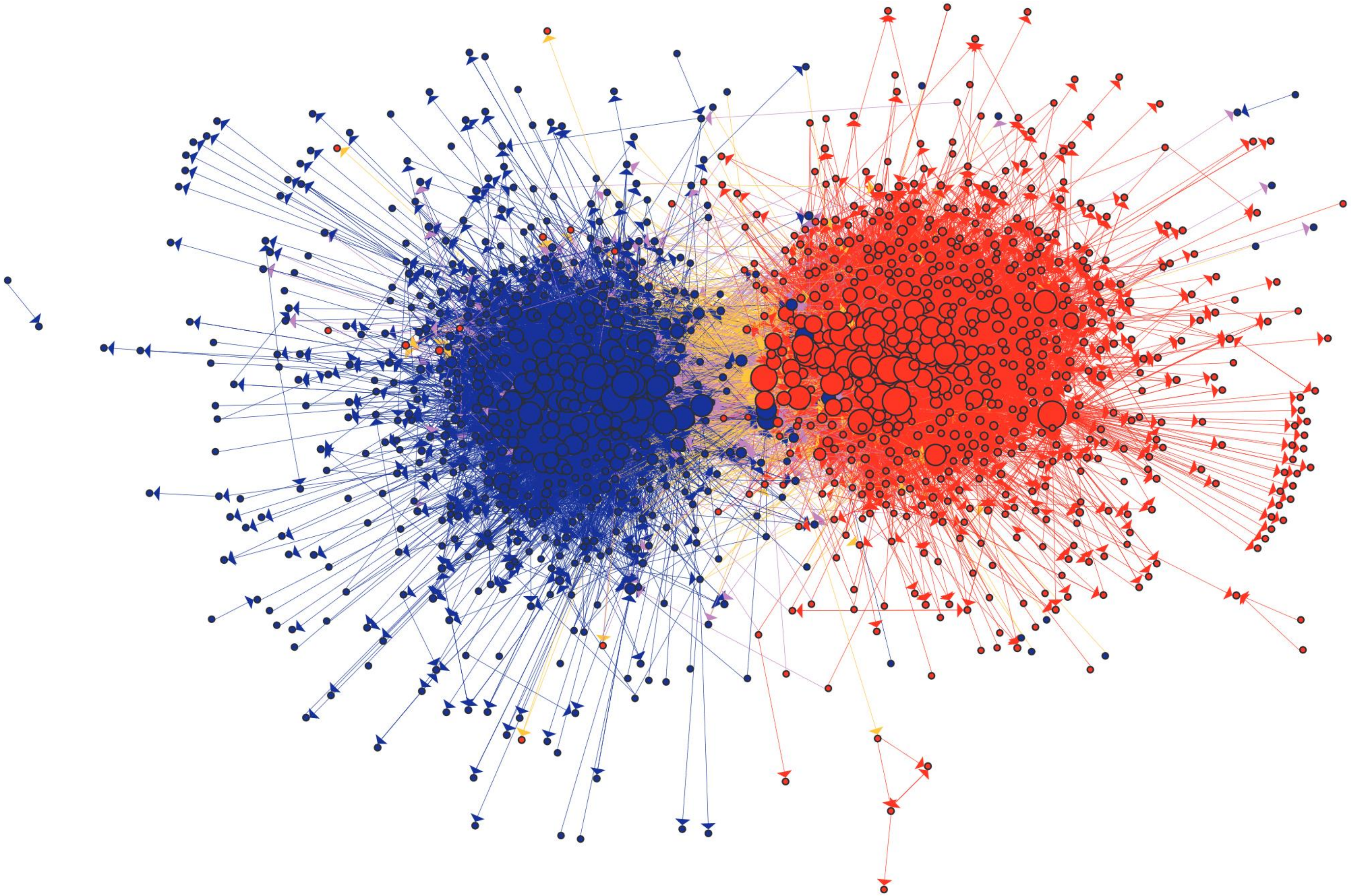


Katz centrality

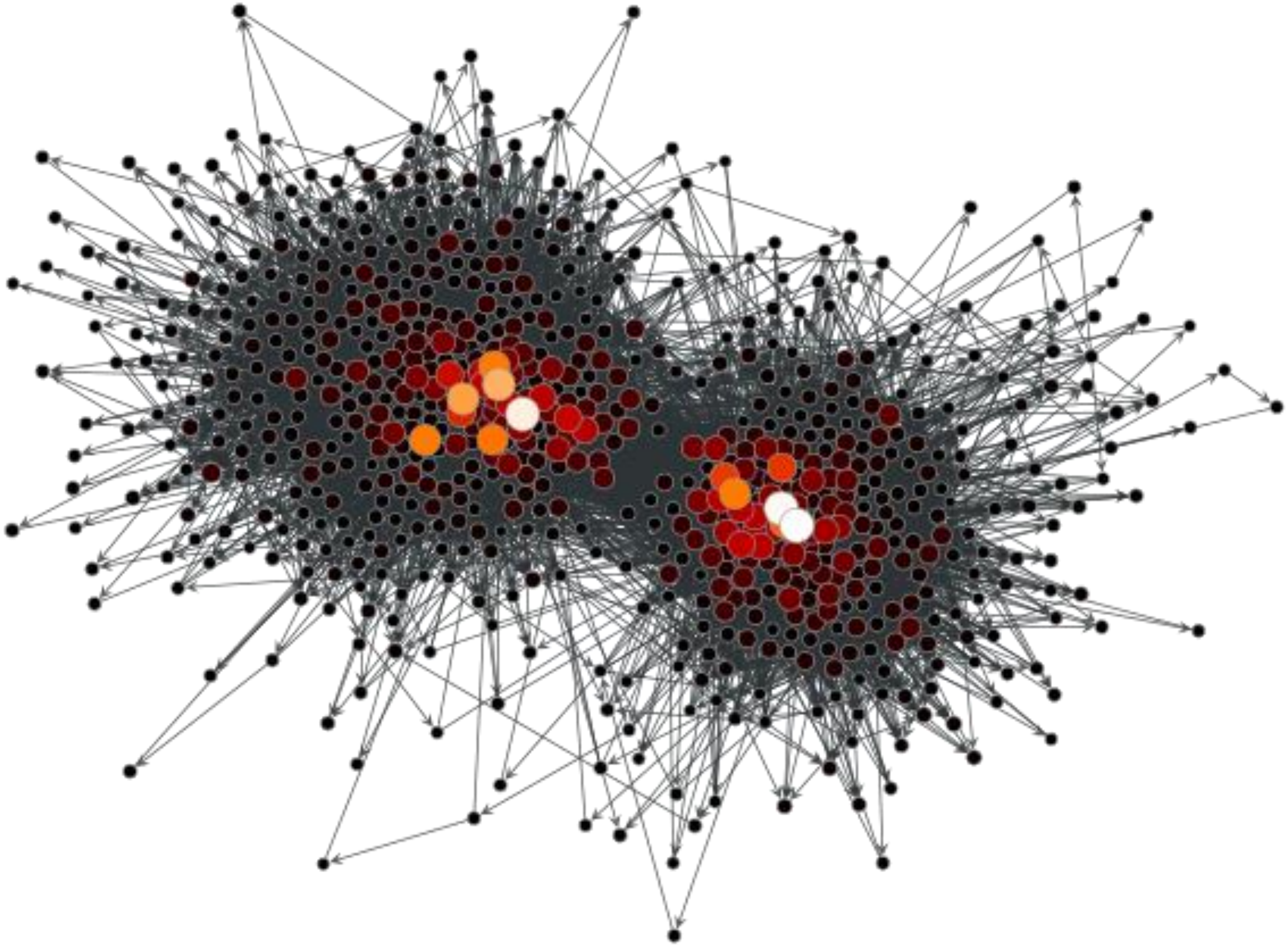


$\alpha = 0.25, \beta = 0.2$

Lada A. Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. I*nternational workshop on Link discovery*, pp36-43.



Graph tool - Centrality measures, https://graph-tool.skewed.de/static/doc/centrality.html



Assume that I have a personal webpage that is not that popular in general. Now assume that Wikipedia has a high Katz centrality score and that Wikipedia points to my personal webpage. What would be the Katz centrality of my page? Is that expected?

Not everyone known by a well-known person is well known

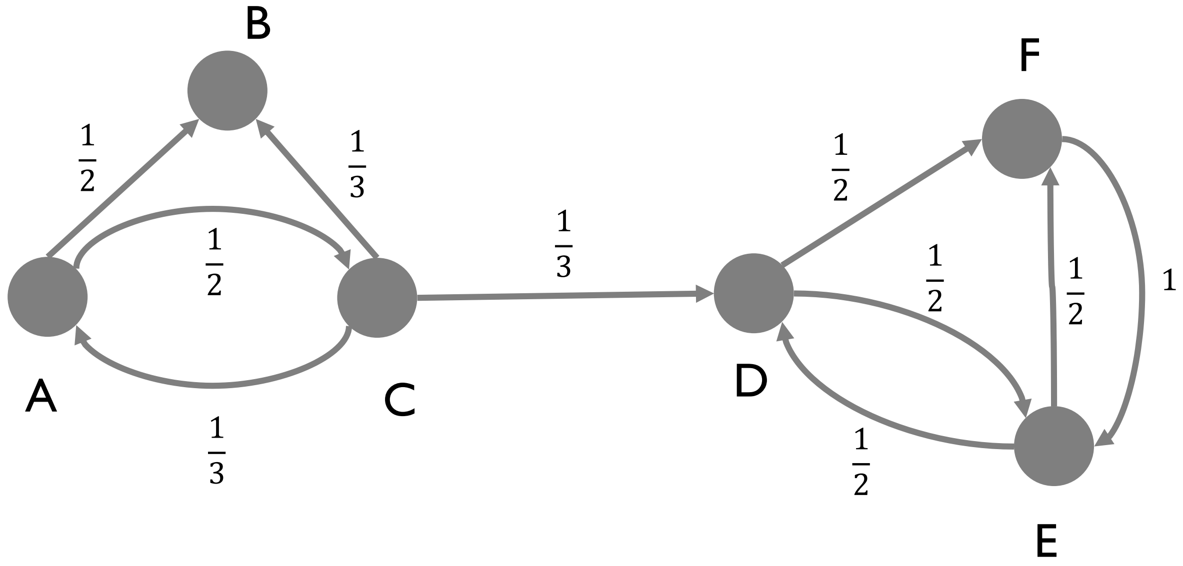
Random surfer

Let's assume that there is a "random surfer" who is given a page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page

Random surfer

Let's assume that there is a "random surfer" who is given a page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page

First attempt



$$r(P_i) = \sum_{P_j \in in(P_i)} \frac{r(P_j)}{|out(P_j)|}$$

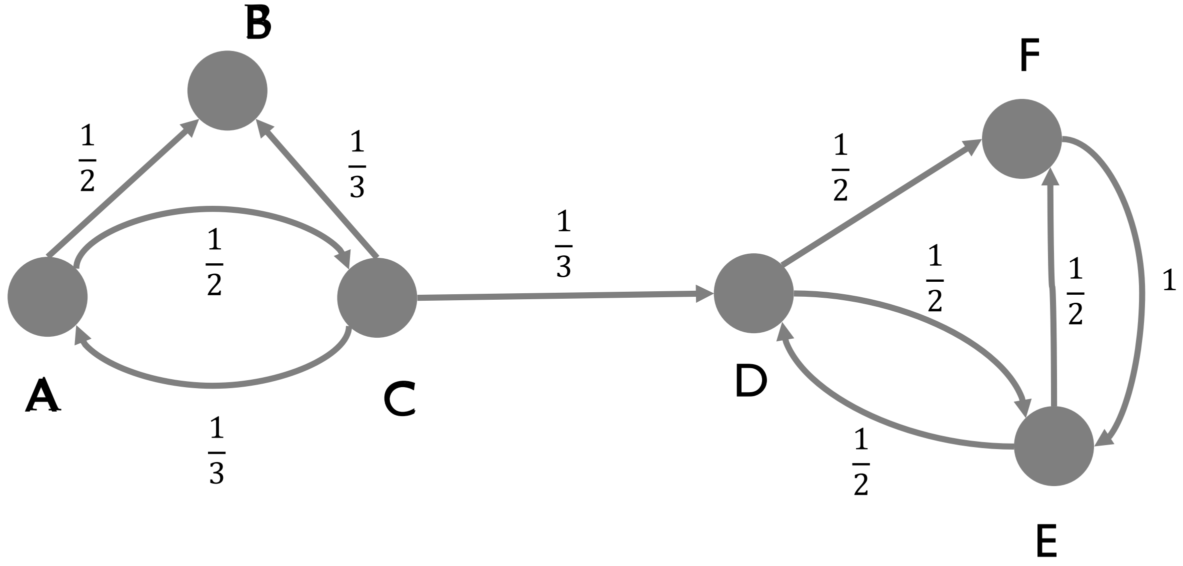
$$r^{(t+1)}(P_i) = \sum_{P_j \in in(P_i)} \frac{r^t(P_j)}{|out(P_j)|}$$

What is $r^0(P_i) =?$ $\quad \forall P_i \in V\colon r^0(P_i) = \frac{1}{n}$

$r^1(P_D) =?$ $\quad r^1(P_D) = \frac{\frac{1}{6}}{3} + \frac{\frac{1}{6}}{2} = \frac{5}{36}$ $\quad r^2(P_D) = \cdots$

Power method ✓

Graph as a matrix



$$H = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

How do we represent this in a matrix form?

Using the matrix

$$r^{(t+1)}(P_i) = \sum_{P_j \in in(P_i)} \frac{r^t(P_j)}{|out(P_j)|}$$

$$\forall P_i \in V:\ r^0(P_i) = \frac{1}{n}$$

$$\Rightarrow$$

$$\mathbf{x}^{(t+1)} = \mathbf{x}^t H$$

$\mathbf{x}^0$ is the initial rank vector , having all entries equal to $\frac{1}{n}$

Power method ✓

Matrix form

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

Claim λ_1=1 $\Longrightarrow$ $\mathbf{x} = H\mathbf{x}$

How do we know that H has an eigenvalue of 1 so that such a vector $\mathbf{x}$ exists?

Even if it does exist, will it be unique?

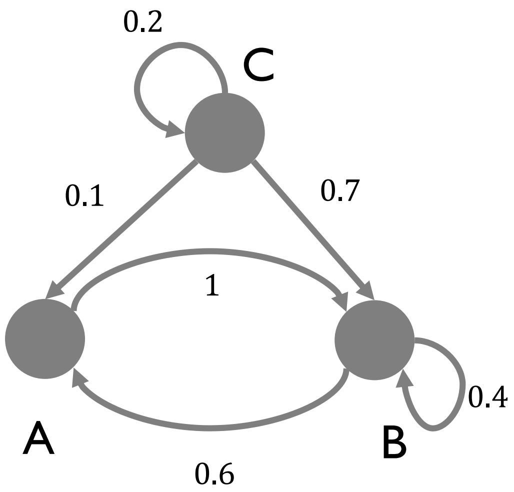
Markov chains (recap)

- A Markov chain is a stochastic process that moves between states
  - Set of states with size n
  - Transition probability matrix P made of all the transition probabilities
  - Matrix entry P_{ij} denotes $P(j|i)$: the probability of j being the next state given the current state is i

Markov chains (recap)

- P is a *stochastic* matrix if
  - $\forall i: \sum_j P_{ij} = 1$
- What is the largest eigenvalue of a stochastic matrix?
  - a stochastic matrix always has 1 as the largest eigenvalue
- P is *irreducible* if
  - there is a path from any state to any other (strongly connected component)
- P is *aperiodic* if
  - All states have a period $k = 1$
  - A state has a period k if k is GCD of the lengths of all the cycles that pass via the state

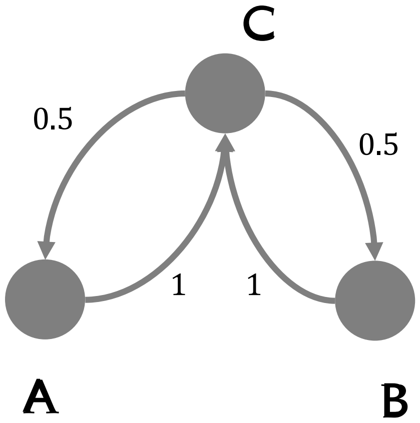
Transition probability matrix



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0.6 & 0.4 & 0 \\ 0.1 & 0.7 & 0.2 \end{bmatrix} \begin{matrix} 1 \\ 1 \\ 1 \end{matrix}$$

Is P stochastic?

Periodicity



What is the period of *C*?

$k_c = 2$

Markov chains

- Stationary distribution $\mathbf{x}$
  - Markov chain for which P does not vary over time

- Ergodic theorem
  - The power method applied to P for any starting vector converges to a unique positive vector (*stationary distribution*) if P is stochastic, irreducible and aperiodic

Random surfer and Markov chain

- Markov chains are abstractions of walks of a random surfer
  - Each state corresponds to a webpage
  - Each transition probability corresponds to the probability of going from one page to another

Matrix form

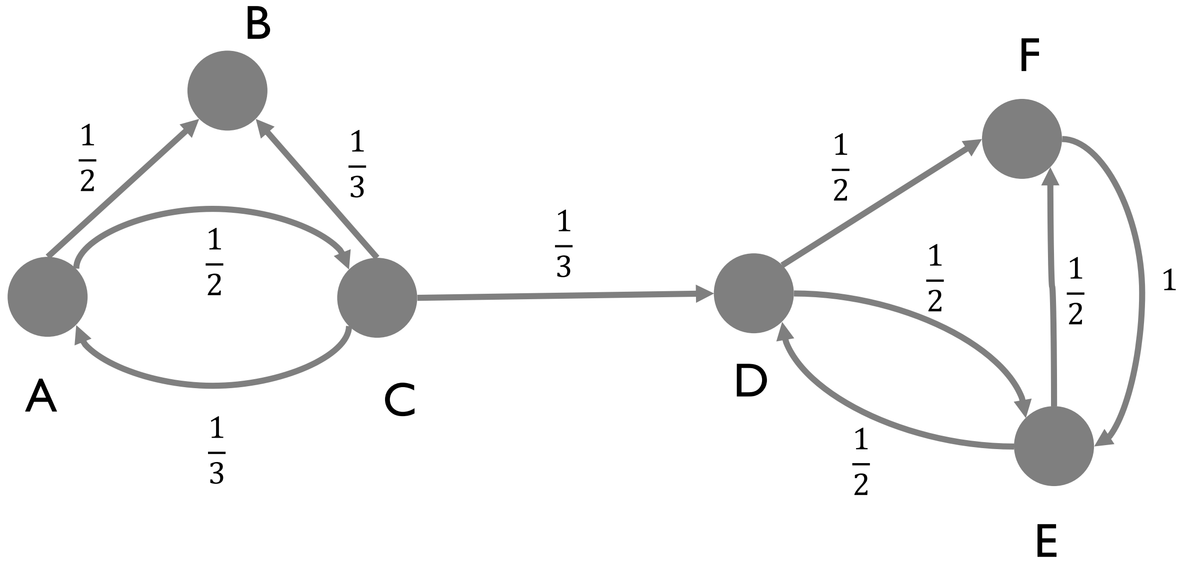
$H\mathbf{x} = \lambda_1 \mathbf{x}$

λ_1=1 $\Rightarrow$ $\mathbf{x} = H\mathbf{x}$

How do we know that H has an eigenvalue of 1 so that such a vector $\mathbf{x}$ exists? ✓

Even if it does exist, will it be unique? ✓

Transition probability matrix

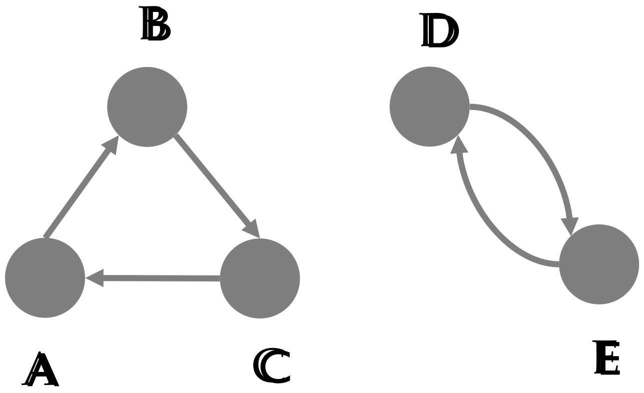


$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \boxed{0 & 0 & 0 & 0 & 0 & 0} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} = H$$

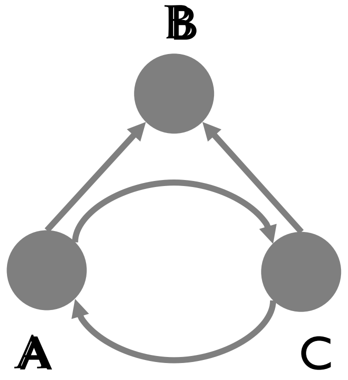
Is P stochastic?

Problems

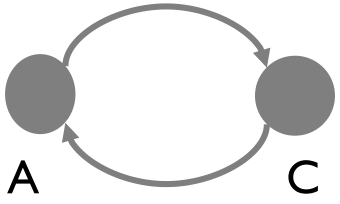
The Web is full of dead-ends



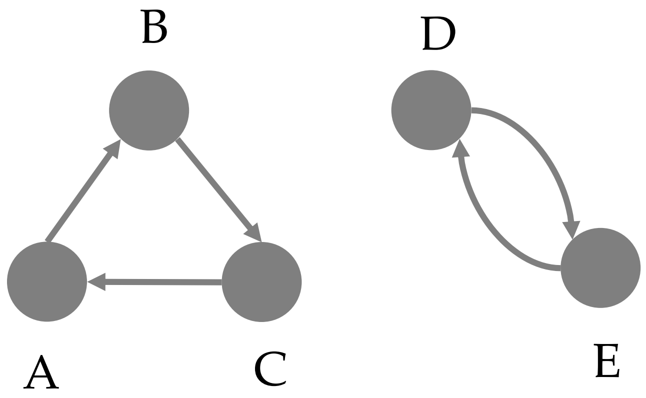
Disconnected components



Sink nodes or dangling links



Loops



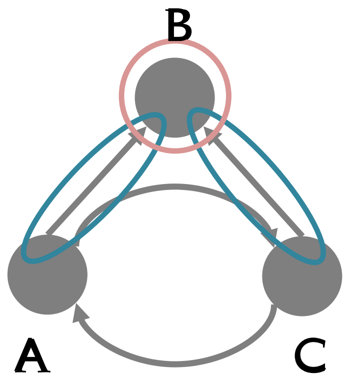
$$H = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Can you guess what might be the problem?

There are two eigenvectors of H with eigenvalue 1

$$\mathbf{x} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

These are opposite rankings!



$$r^0(P_A) = \frac{1}{3}$$

$$r^0(P_B) = \frac{1}{3}$$

$$r^0(P_C) = \frac{1}{3}$$

$$r^1(P_A) = \frac{\frac{1}{3}}{2} = \frac{1}{6}$$

$$r^1(P_B) = \frac{\frac{1}{3}}{2} + \frac{\frac{1}{3}}{2} = \frac{1}{3}$$

$$r^1(P_C) = \frac{\frac{1}{3}}{2} = \frac{1}{6}$$

$$r^2(P_A) = \frac{\frac{1}{6}}{2} = \frac{1}{12}$$

$$r^2(P_B) = \frac{\frac{1}{6}}{2} + \frac{\frac{1}{6}}{2} = \frac{1}{6}$$

$$r^2(P_C) = \frac{\frac{1}{6}}{2} = \frac{1}{12}$$

$$r^3(P_A) = \frac{1}{24}$$

$$r^3(P_B) = \frac{1}{12}$$

$$r^3(P_C) = \frac{1}{24}$$

Which node is a sink?

Which links are dangling links?

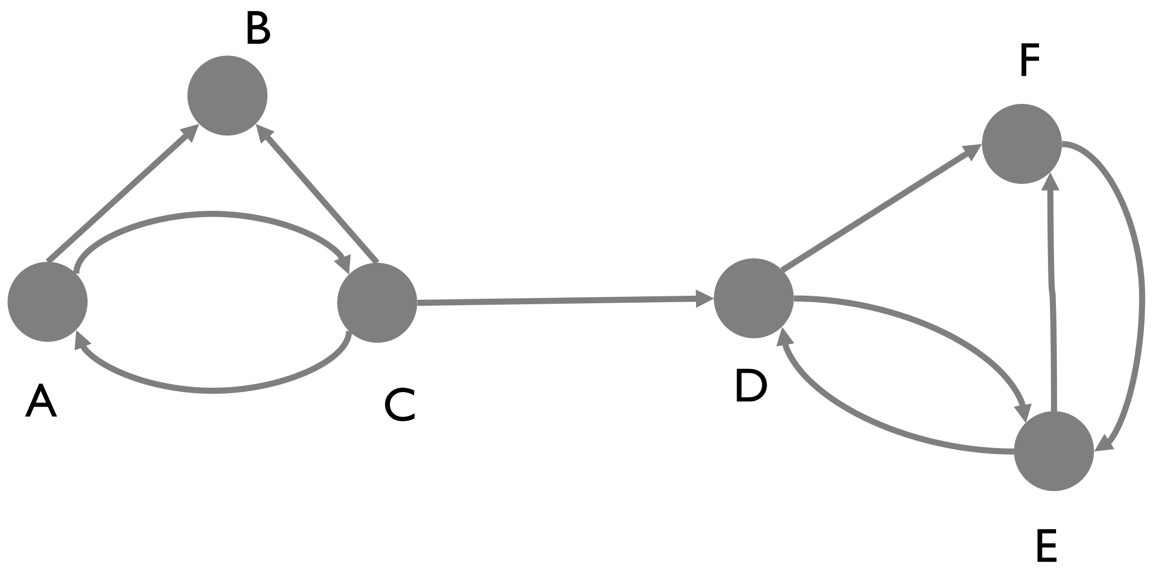
Node *B* leaks the PageRank

Correction (random jumps)

- At a dead-end (no out-links), the surfer jumps to a random page
  - Transition probability is $1/n$

- At any non-dead end (there is at least one out-link)
  - With probability α, the surfer jumps to a random page
    - Transition probability is α/n
  - With remaining probability, the surfer goes to a random link
    - Transition probability is $(1-\alpha)/n$

Correction example

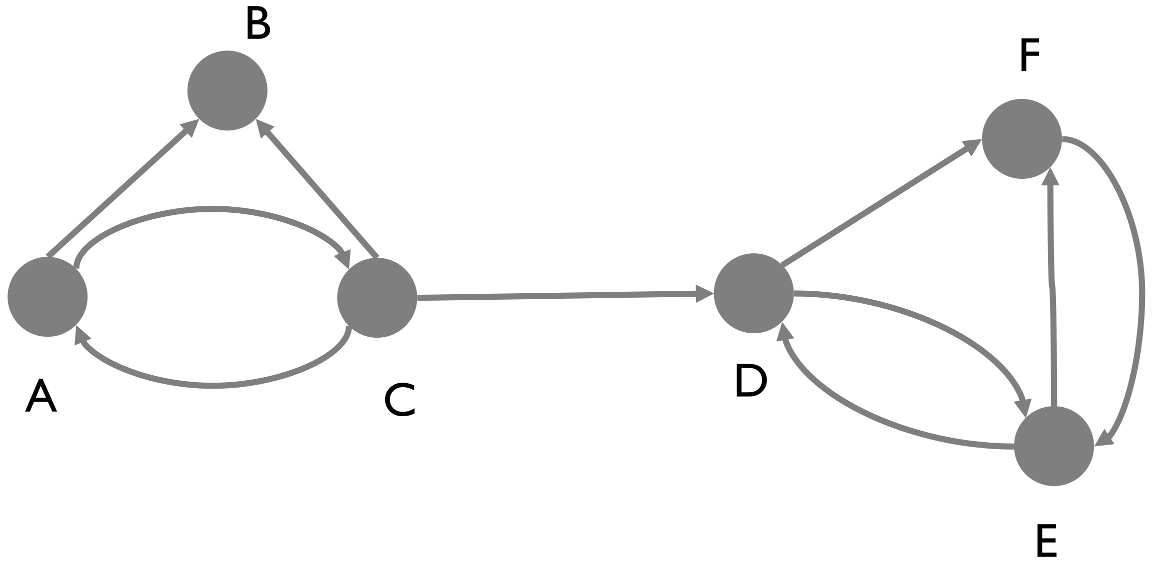


sink node

$$H = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The surfer randomly jumps to any other node with probability $1/n$

Correction example



$$H = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The surfer randomly jumps to any other node with probability $1/n$

H is a stochastic matrix

PageRank

$\alpha = 0.85$ is chosen as the best balance between the number of iterations needed and the importance of the link structure

2

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{t} G$$

$$G = \alpha H + (\alpha a + (1-\alpha)e)1/ne^{T}$$

Google matrix

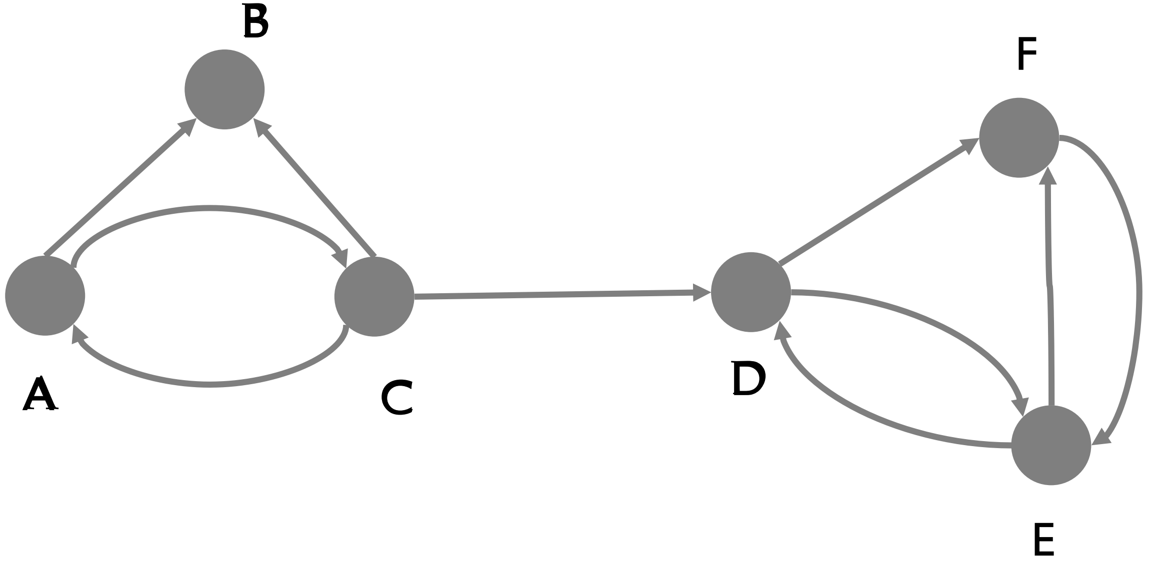
1

$$r(P_i) = \alpha \sum_{P_j \in in(P_i)} \frac{r(P_j)}{|out(P_j)|} + \frac{1-\alpha}{n}$$

3

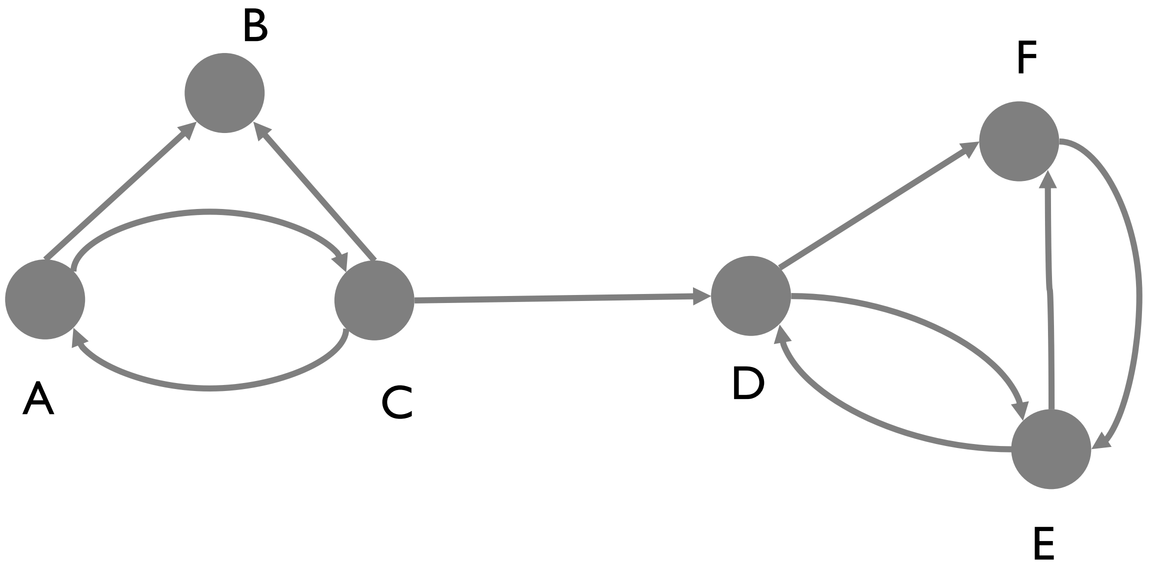
$$\mathbf{x} = D(D - \alpha A)^{-1}\mathbf{1}$$

$D = diag(|out(P_1)|, |out(P_2)|, \dots, |out(P_n)|)$ is a diagonal matrix of degrees with $D_{ii} = \max(|out(P_i)|, 1)$



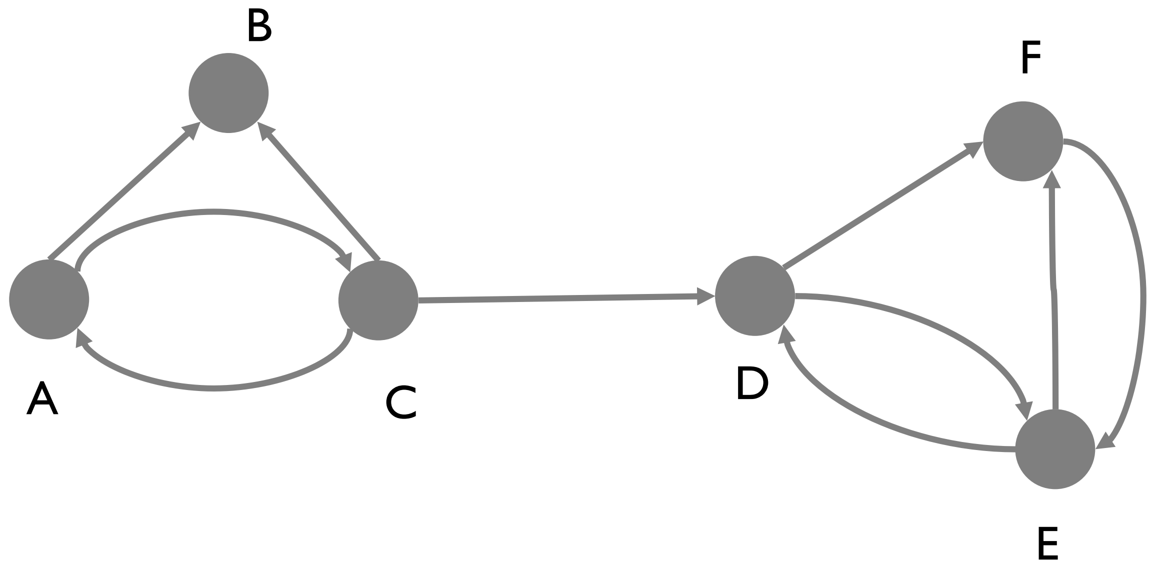
$$G = \alpha H + (\alpha a + (1-\alpha)e)1/ne^T$$

$$G = 0.85\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} + \left(0.85\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 0.15\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$



$$G = \alpha H + (\alpha a + (1-\alpha)e)1/ne^T$$

$$G = \begin{bmatrix} 0 & \frac{17}{40} & \frac{17}{40} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{17}{60} & \frac{17}{60} & 0 & 0 & \frac{17}{60} & 0 \\ 0 & 0 & 0 & 0 & \frac{17}{40} & \frac{17}{40} \\ 0 & 0 & 0 & \frac{17}{40} & 0 & \frac{17}{40} \\ 0 & 0 & 0 & \frac{17}{20} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \end{bmatrix} = \begin{bmatrix} \frac{1}{40} & \frac{18}{40} & \frac{18}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{37}{120} & \frac{37}{120} & \frac{1}{40} & \frac{1}{40} & \frac{37}{120} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{18}{40} & \frac{18}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{18}{40} & \frac{1}{40} & \frac{18}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{7}{8} & \frac{1}{40} \end{bmatrix}$$

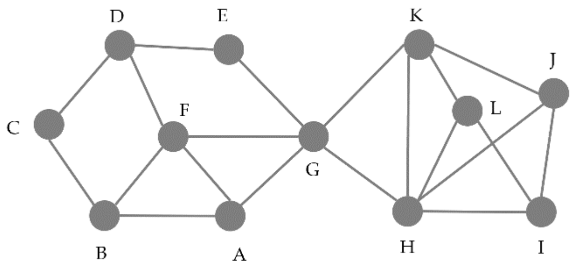


$$\mathbf{x}^{(t+1)} = \mathbf{x}^t G$$

$$\mathbf{x}^1 = \begin{bmatrix} \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{bmatrix} \begin{bmatrix} \frac{1}{40} & \frac{18}{40} & \frac{18}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{37}{120} & \frac{37}{120} & \frac{1}{40} & \frac{1}{40} & \frac{37}{120} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{18}{40} & \frac{18}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{18}{40} & \frac{1}{40} & \frac{18}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{7}{8} & \frac{1}{40} \end{bmatrix} = \begin{bmatrix} \frac{11}{120} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{5}{16} \end{bmatrix}$$

$$\mathbf{x}^2 = \cdots$$

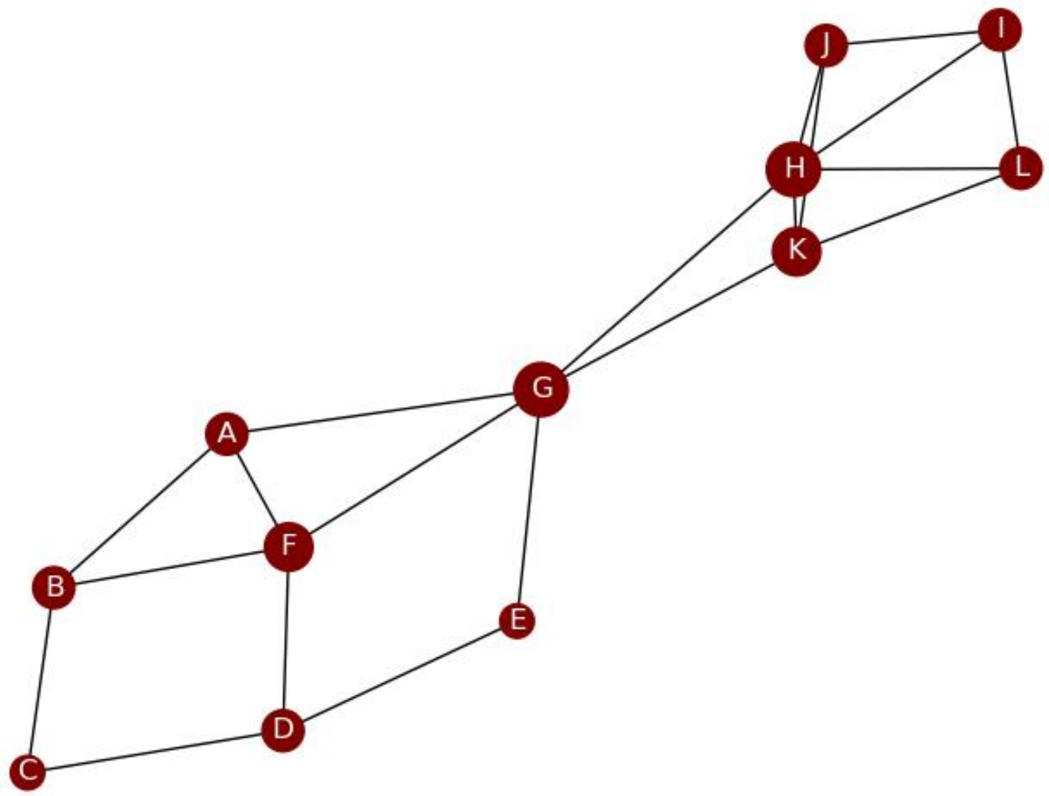
$$\mathbf{x} = \begin{bmatrix} 0.145 \\ 0.145 \\ 0.208 \\ 0.208 \\ 0.145 \\ 0.145 \end{bmatrix}$$



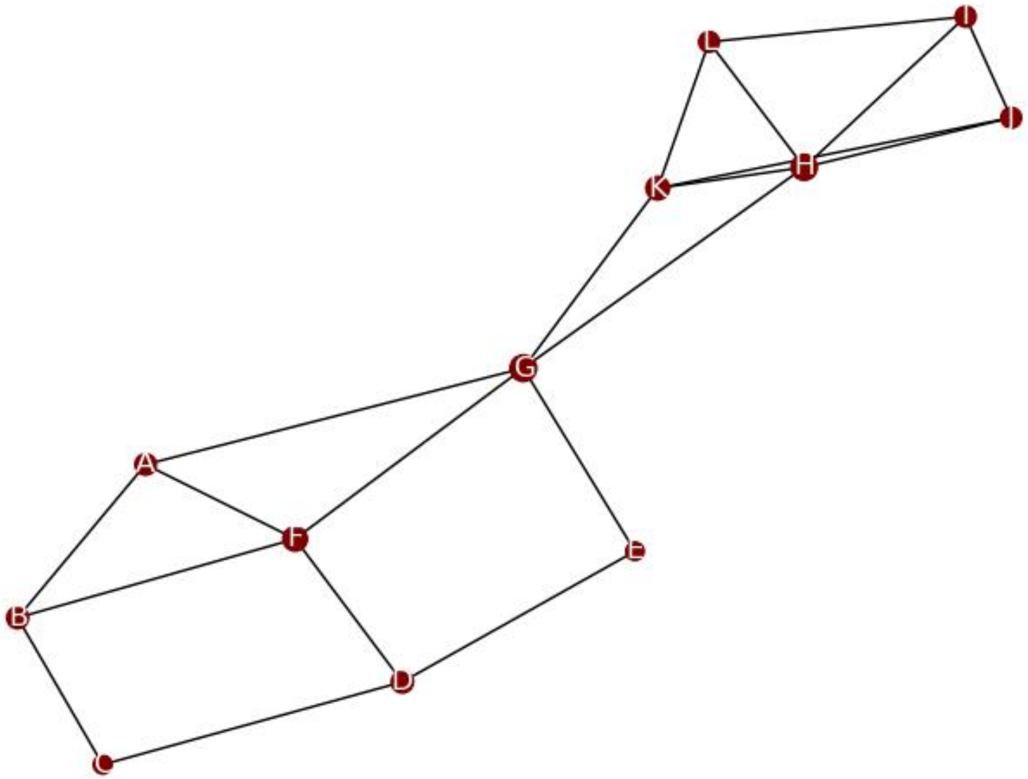
Comparison

Ordered nodes	Eigenvector centrality	Katz centrality	PageRank
1st largest	H (0.473)	H (0.3379)	G (0.118)
2nd largest	K (0.408)	G (0.337)	H (0.114)
3rd largest	G (0.396)	K (0.312)	F (0.101)
4th largest	J (0.320)	F (0.304)	K (0.093)

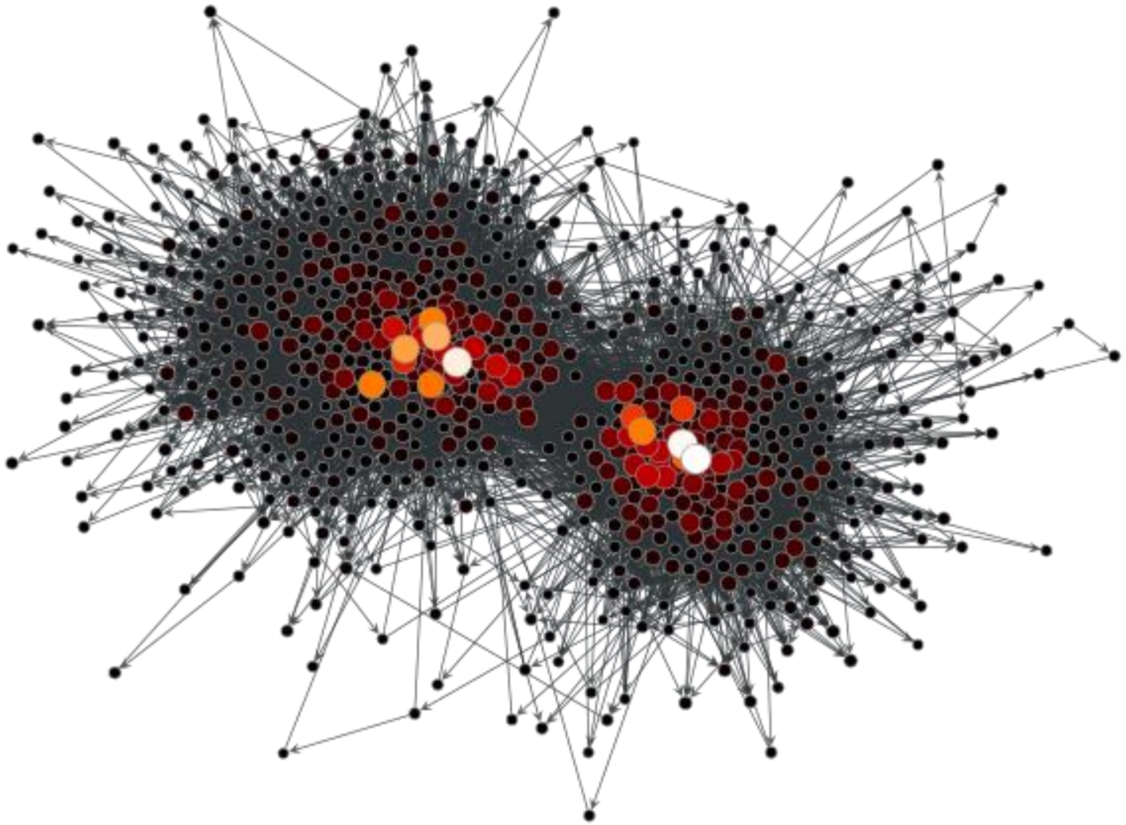
Degree



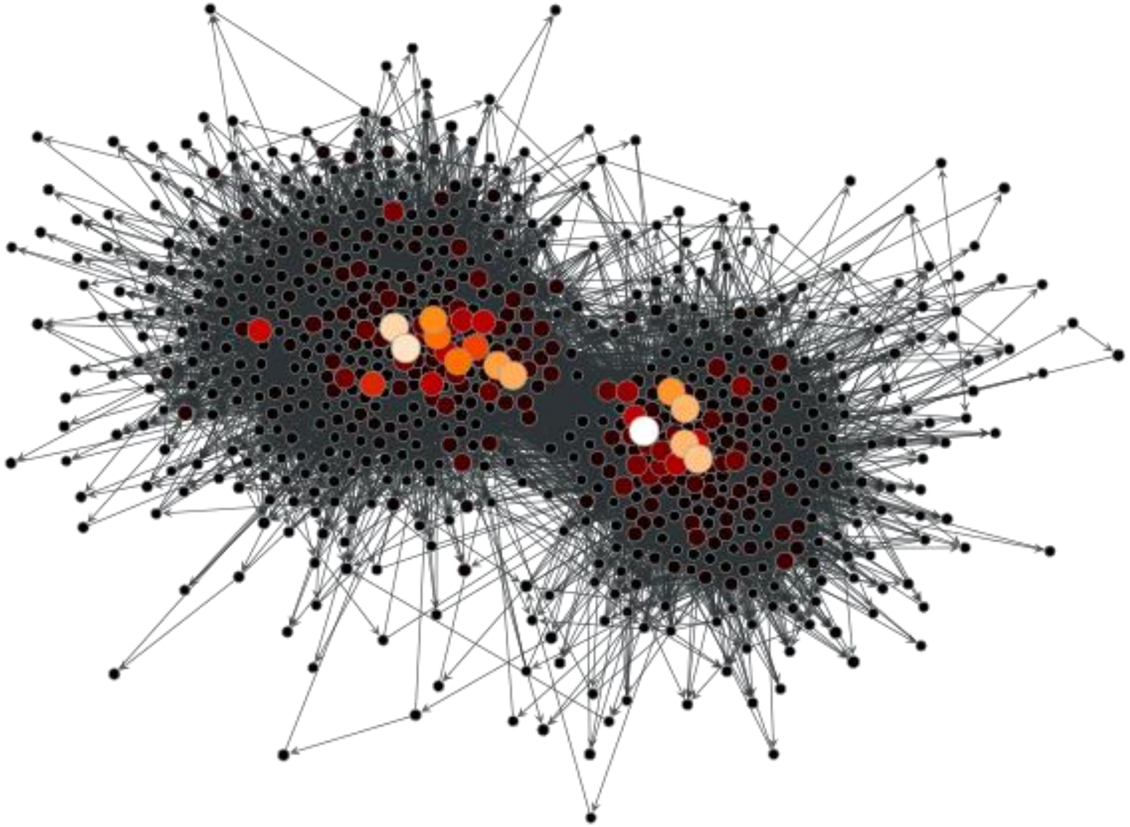
PageRank



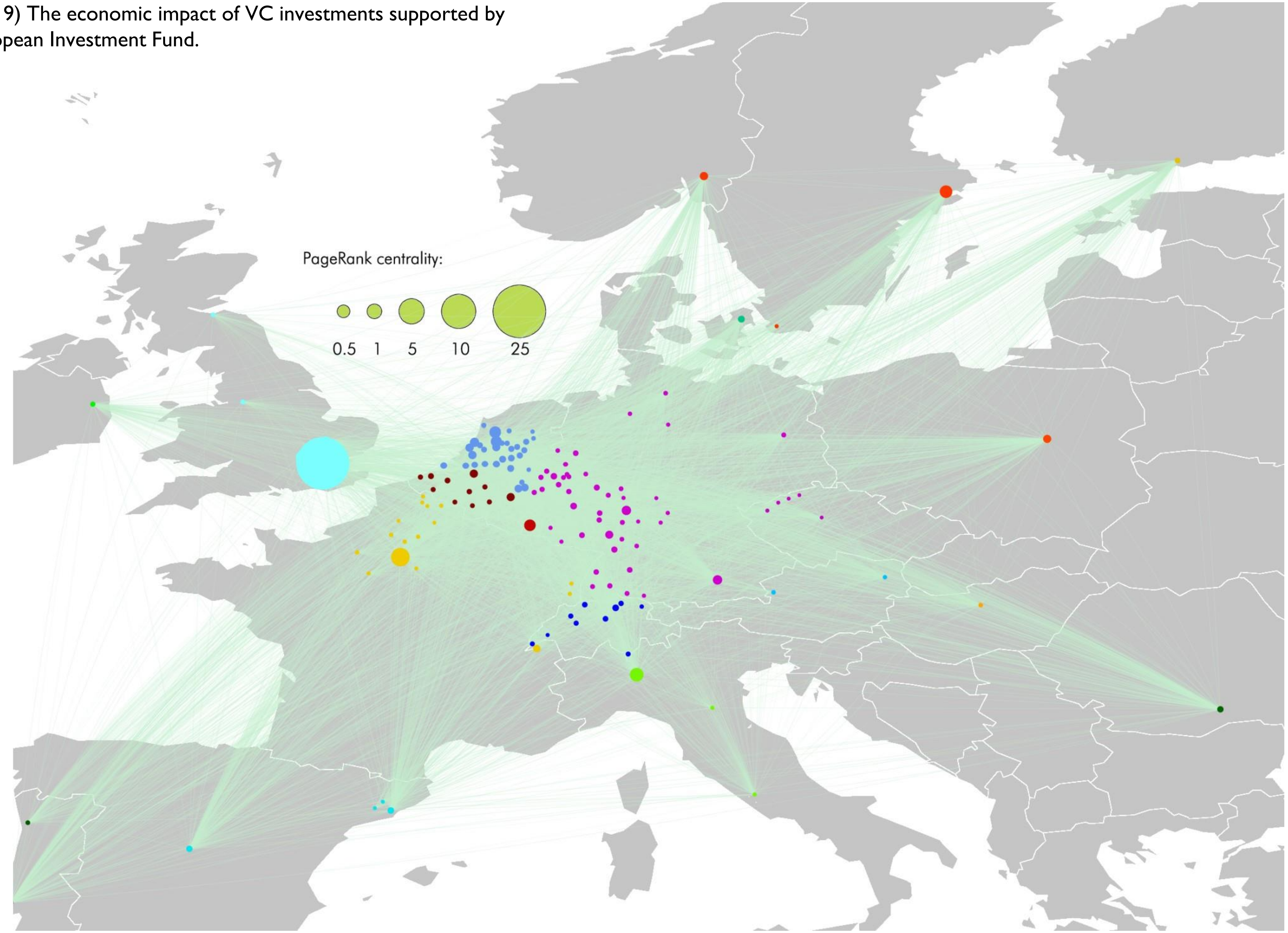
Katz centrality



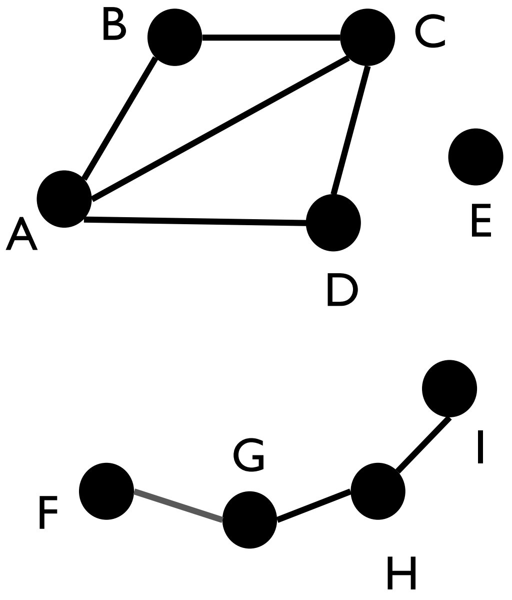
PageRank



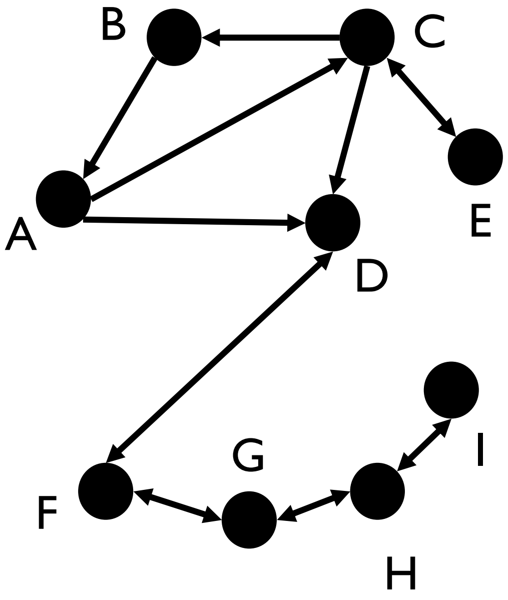
Pavlova, E. and Signore, S. (2019) The economic impact of VC investments supported by the EIF, Technical report, European Investment Fund.



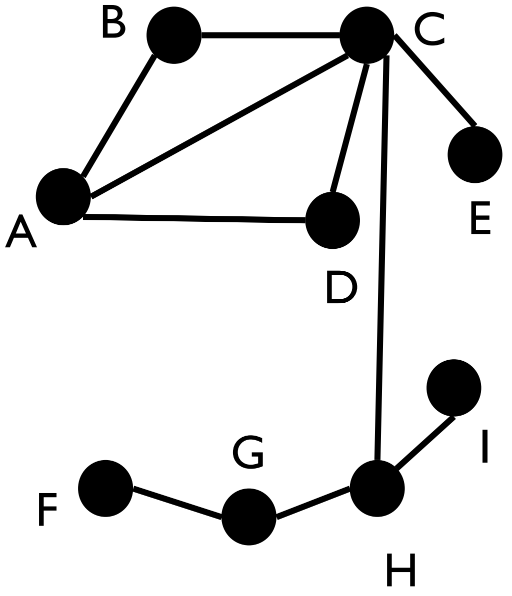
Name the centralities that can be computed for the following networks:



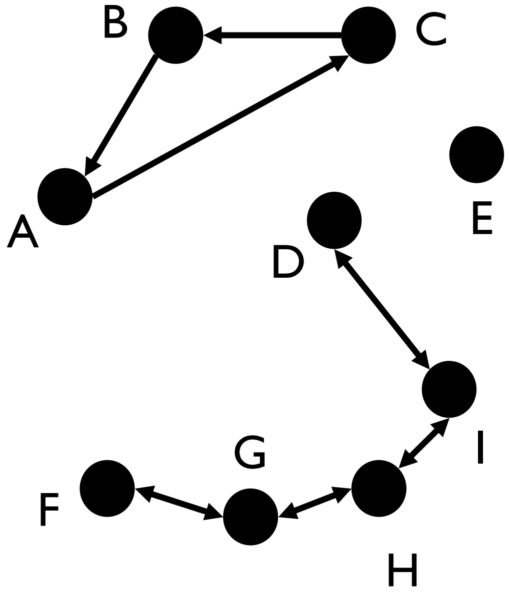
a)



b)



c)



d)

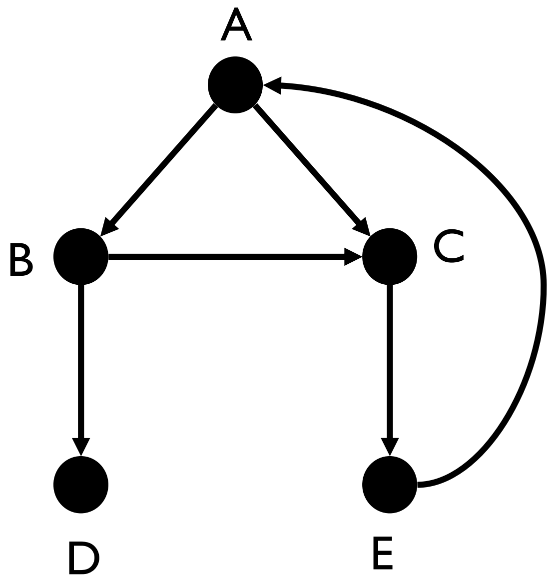
Answer the questions related to the following graph:

c) Is its Markov chain stochastic?

d) Is its Markov chain irreducible?

e) Is its Markov chain aperiodic?

f) Compute its Google matrix by hand.



Sources

- Clauset, A. Network Analysis and Modeling, CSCI5352, Santa Fe Institute (2017), http://tuvalu.santafe.edu/~aaronc/courses/5352/
- Tibshirani, R. Data Mining: PageRank, 36-462/36-662, Stanford University, 2012.
- Jauregui, J. Math 312: Markov chains, Google's PageRank algorithm, University of Pennsylvania, 2012.
- Zafarani, R., Abbasi, M. A. and Liu, H. *Social Media Mining: An Introduction*, Cambridge University Press, 2014.