Node centrality measures

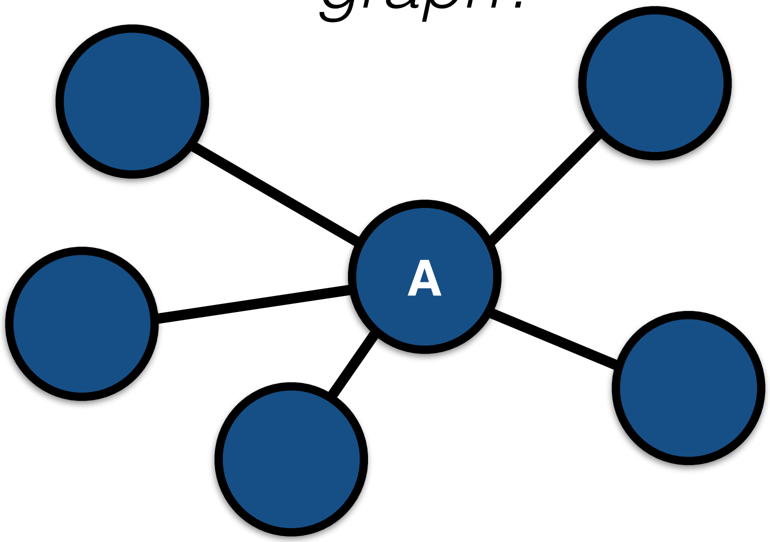
Charles Perez

Why studying centrality?

- Identify key nodes of a graph
  - Influent users
  - Celebrities, popular individuals
  - Experts
  - Key webpages (search engines)
- Applications : digital marketing, medecine, recommandation systems.

Measuring node importance

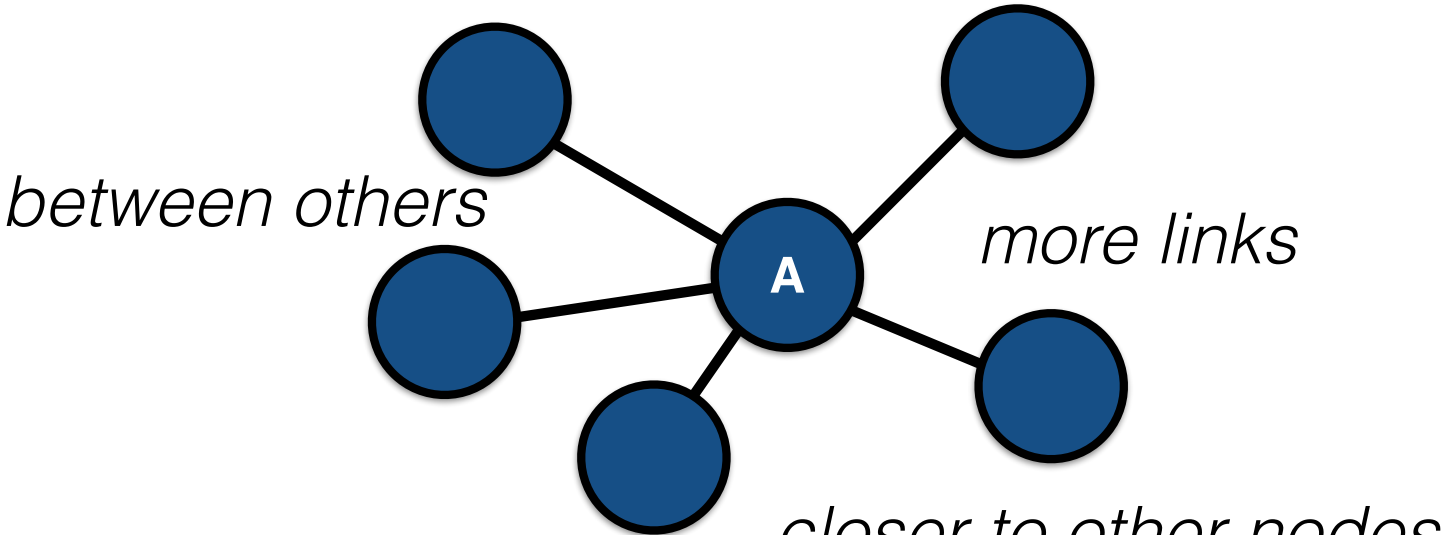
How to define important nodes from their position in the graph?



Freeman : Node **A** has three advantages on other nodes?

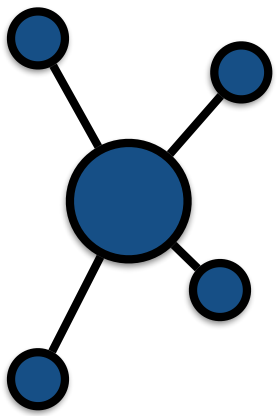
Measuring node importance

How to define important nodes from their position in the graph?

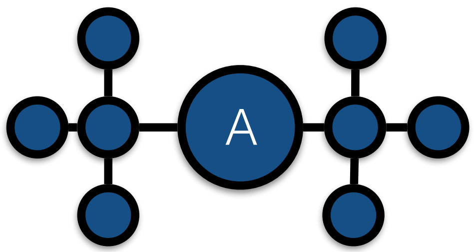


Node centrality

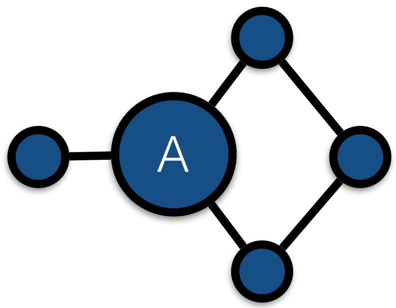
Degree centrality



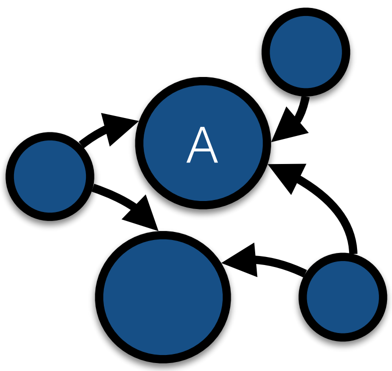
Betweeness centrality



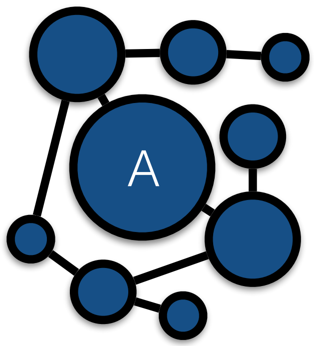
Closeness centrality



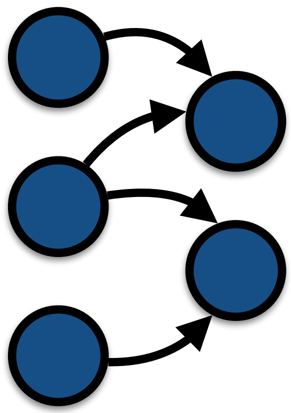
Page Rank



Eigenvalues centrality

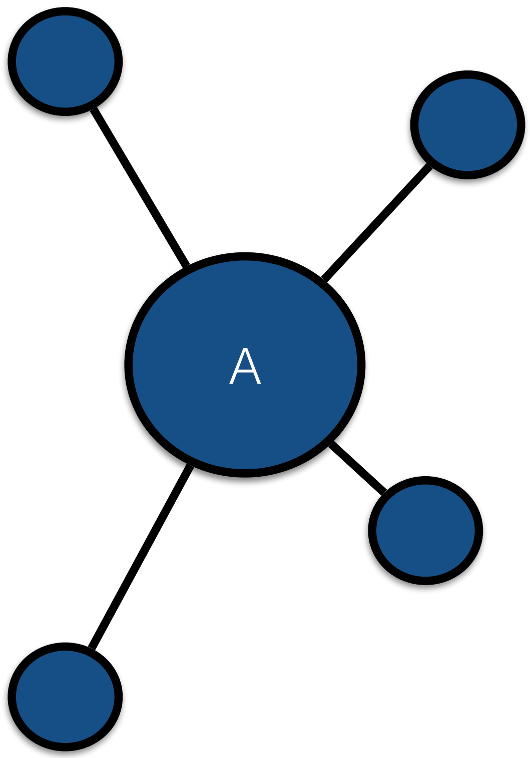


HITS



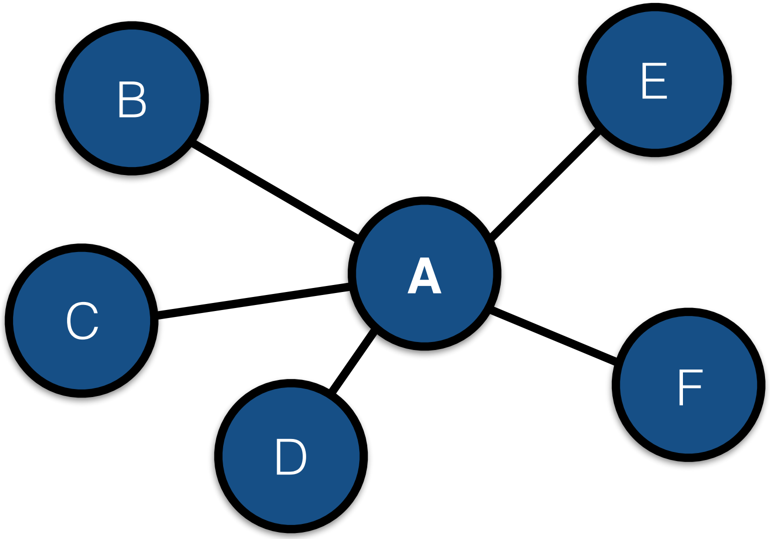
Degree centrality

- Hypothesis : More links a node has more important it is.

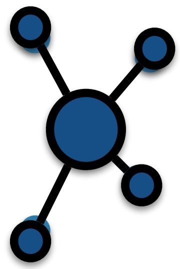


Degree centrality

- Definition : Degree equals to the number of links the node has (number of neighbours).



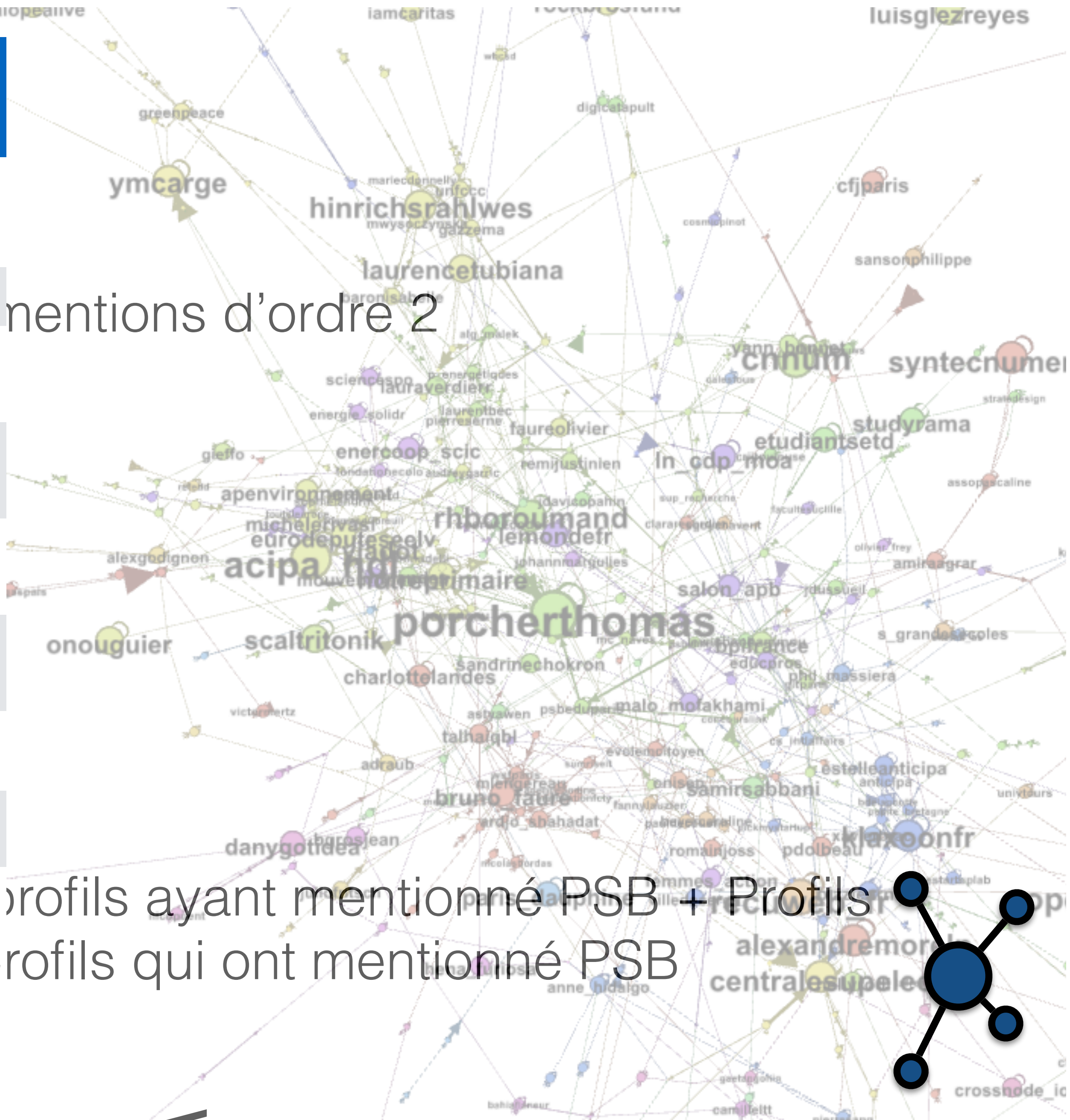
d_a=5
$d_b=1$
$d_c=1$
$d_d=1$
$d_e=1$
$d_f=1$



Case study

Pseudonyme	Degré
gouvernementfr	43
anne_hidalgo	41
afpfr	36
lemondefr	36
lesechos	32
le_figaro	32
frenchweb	31
franceinter	28
letudiantsalon	26

nentions d'ordre 2

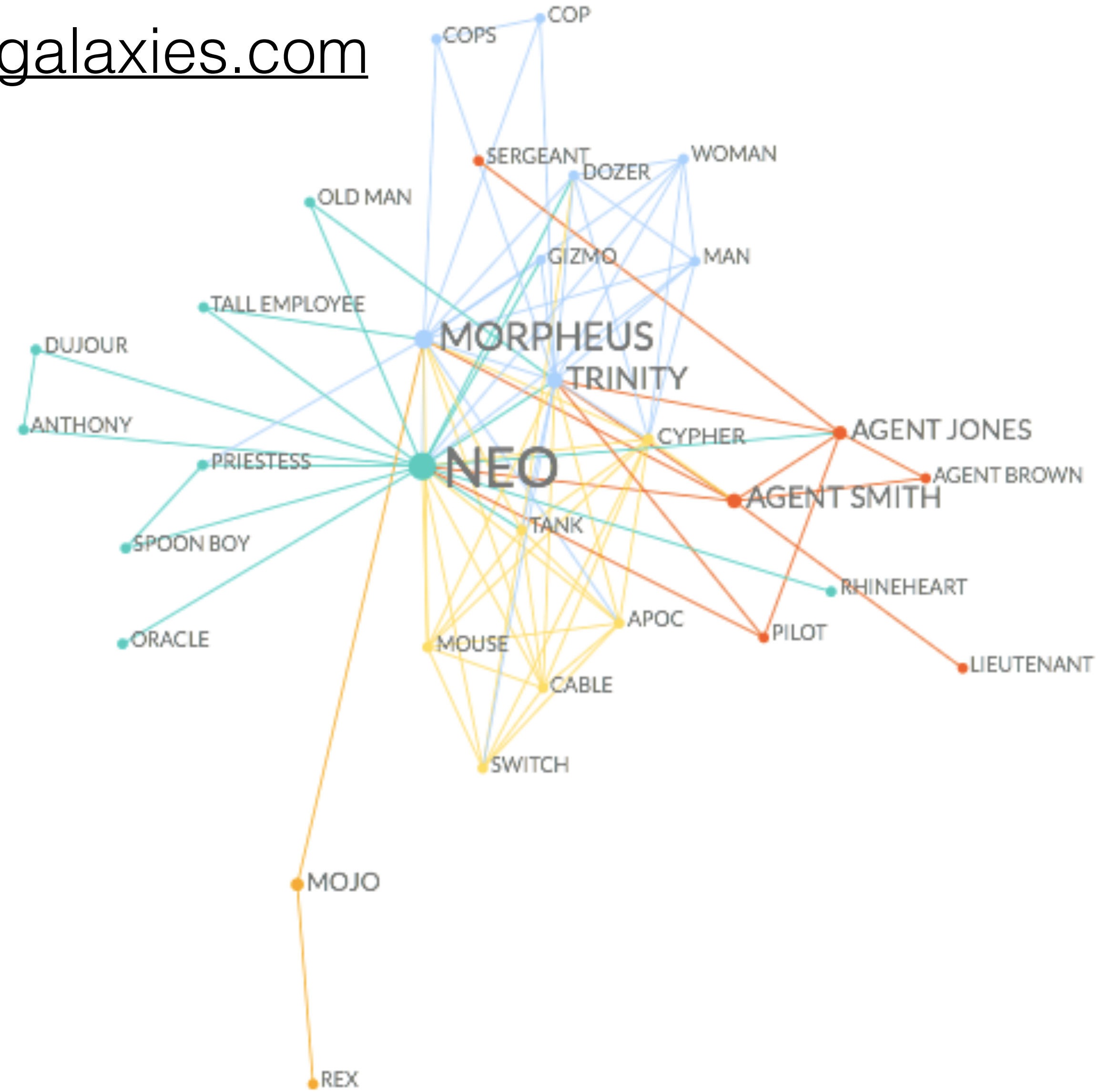


profils ayant mentionné PSB + Profils

ayant mentionné des profils qui ont mentionné PSB

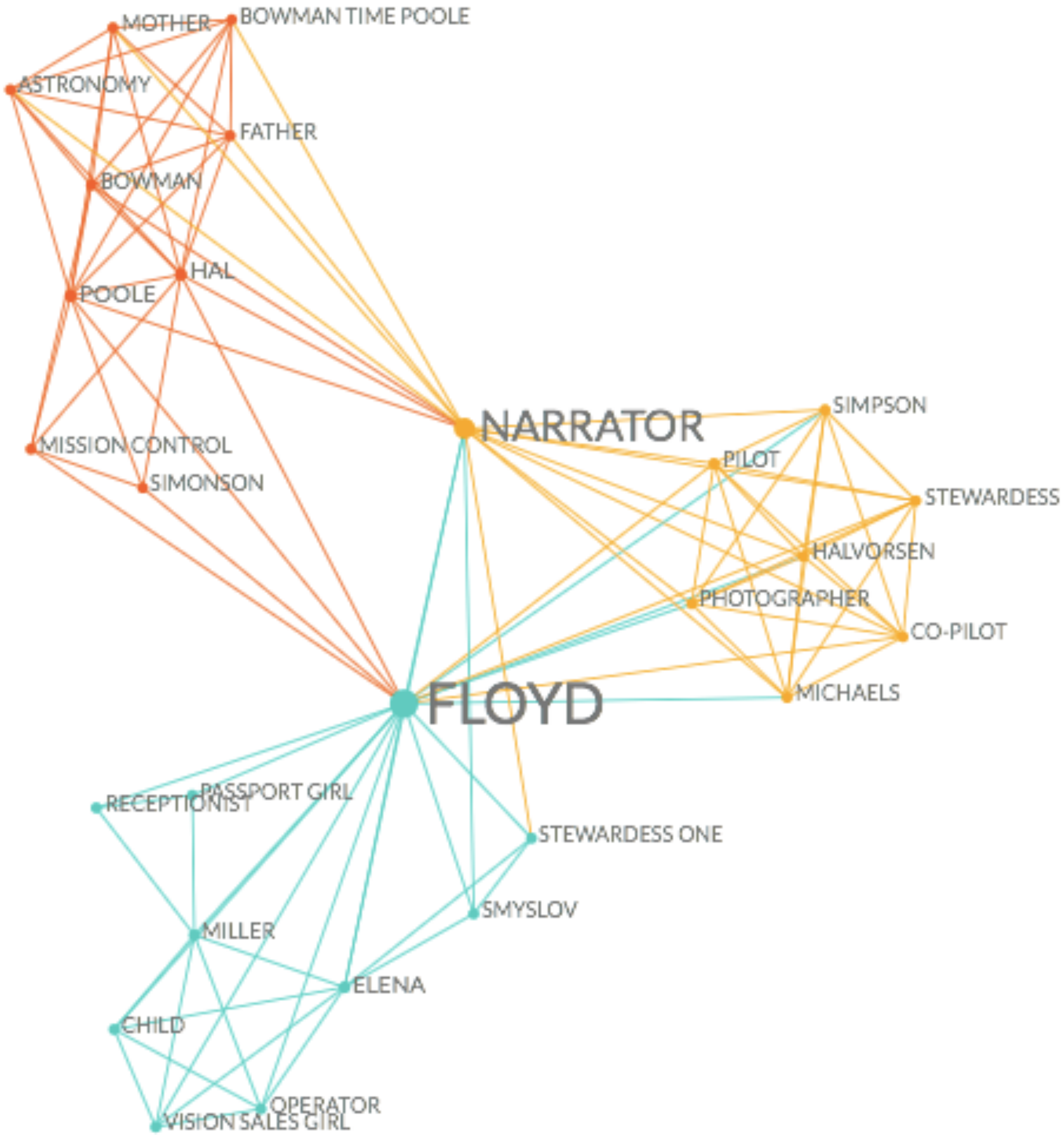
THE MATRIX 1999

http://www.moviegalaxies.com

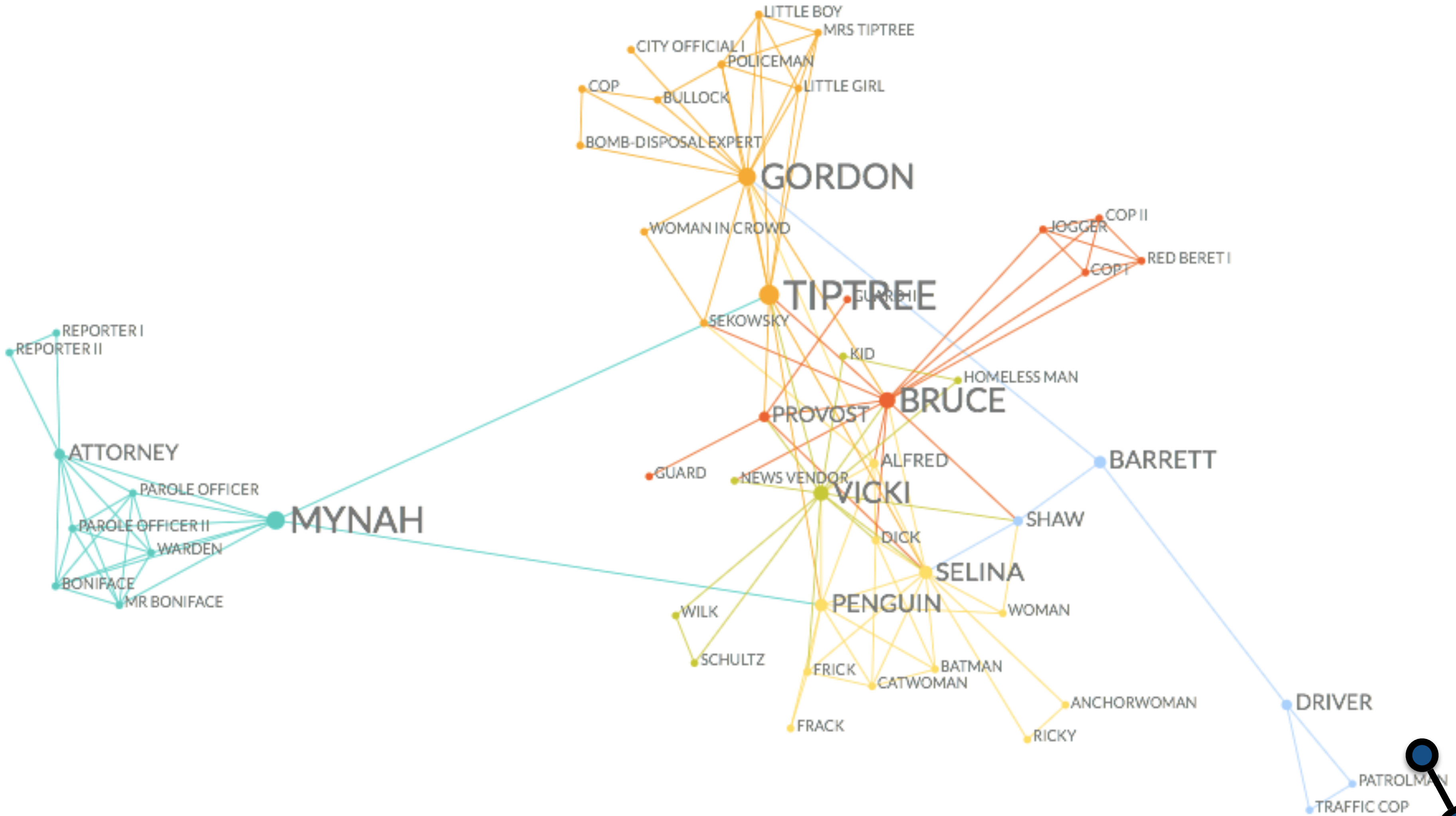


FORREST GUMP 1994

2001: A SPACE ODYSSEY 1968



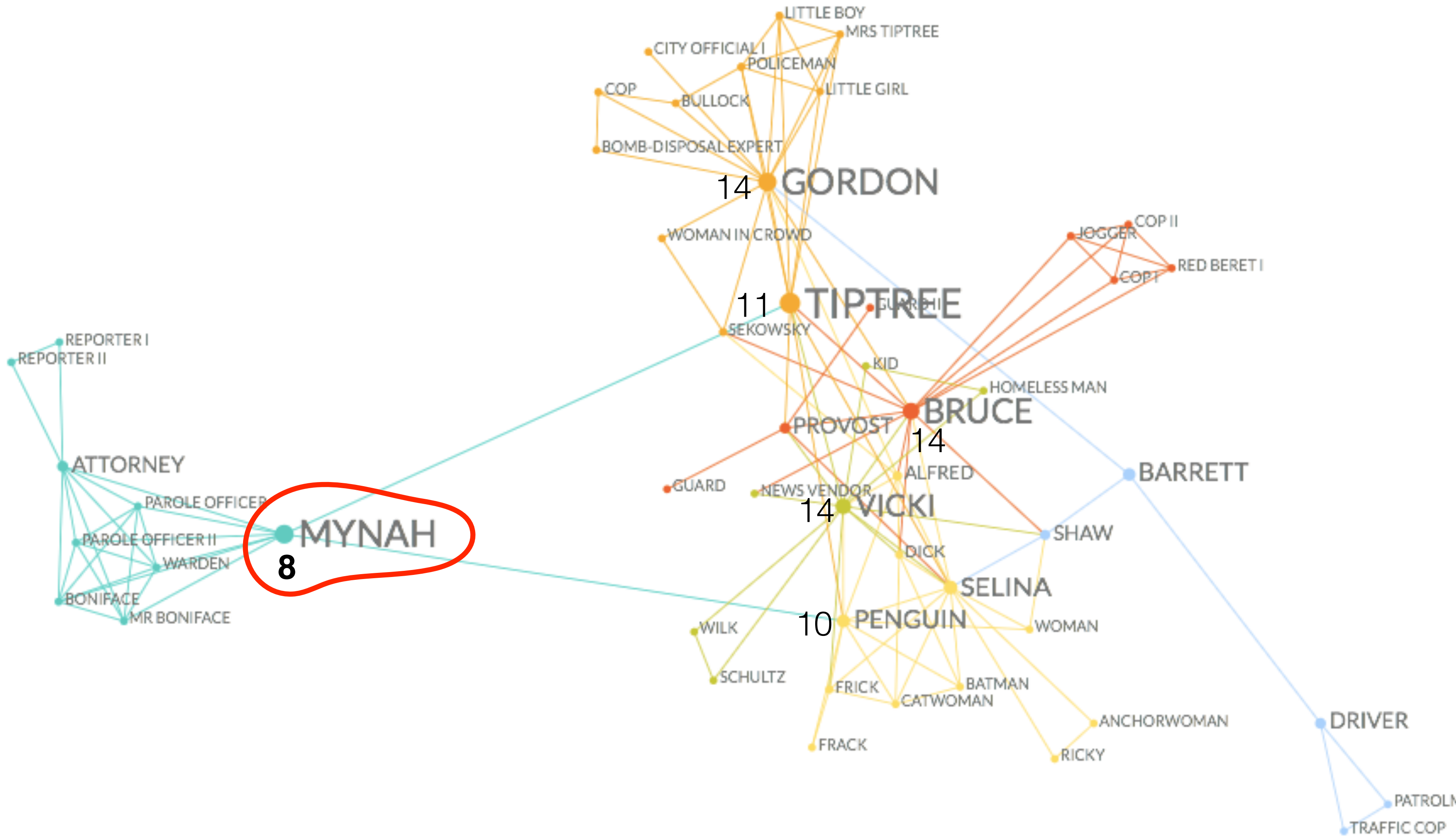
BATMAN RETURNS 1992



Who are the key players?

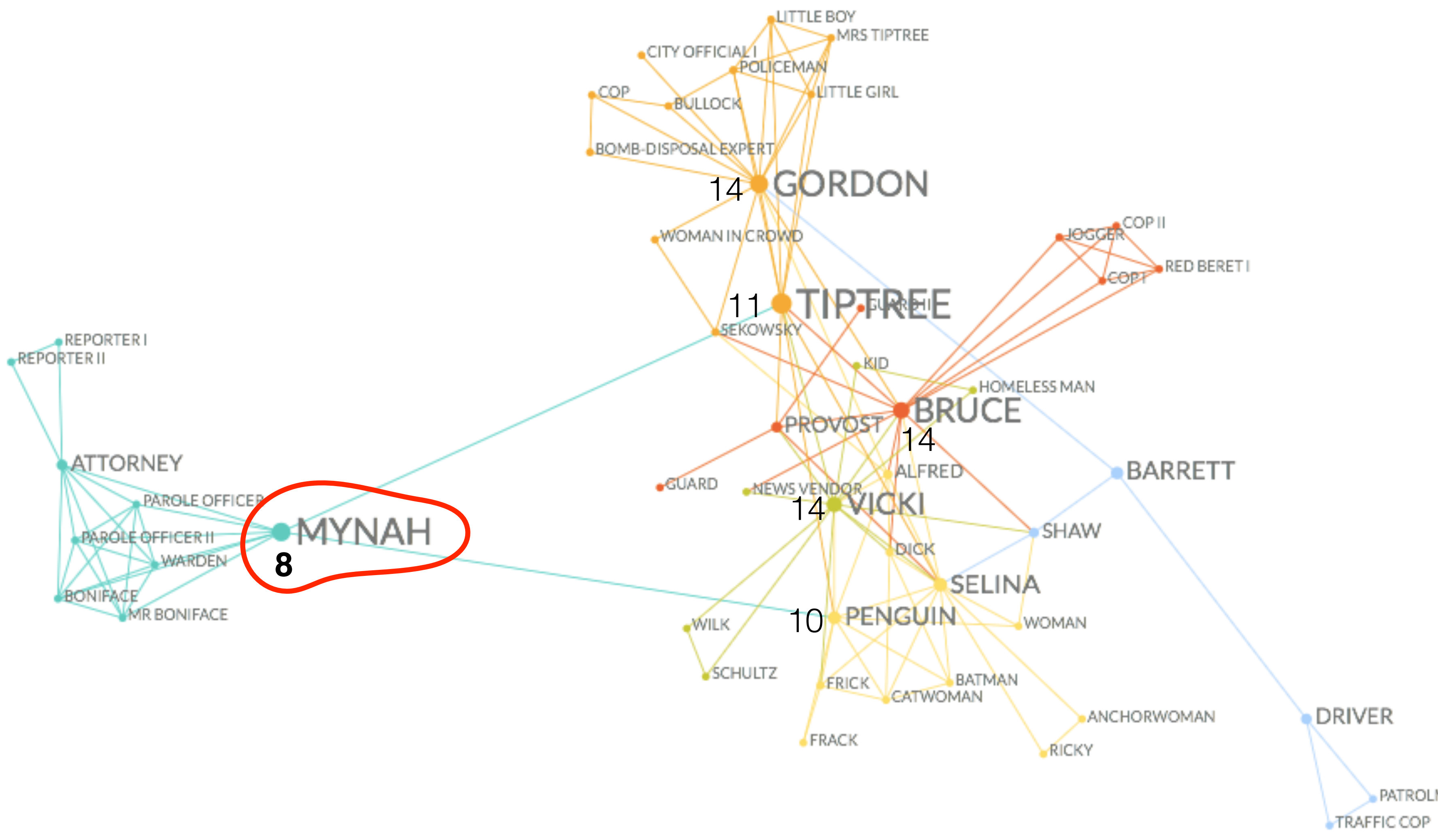
Number of nodes only?

Despite a low degree, Mynah has a key role between groups.



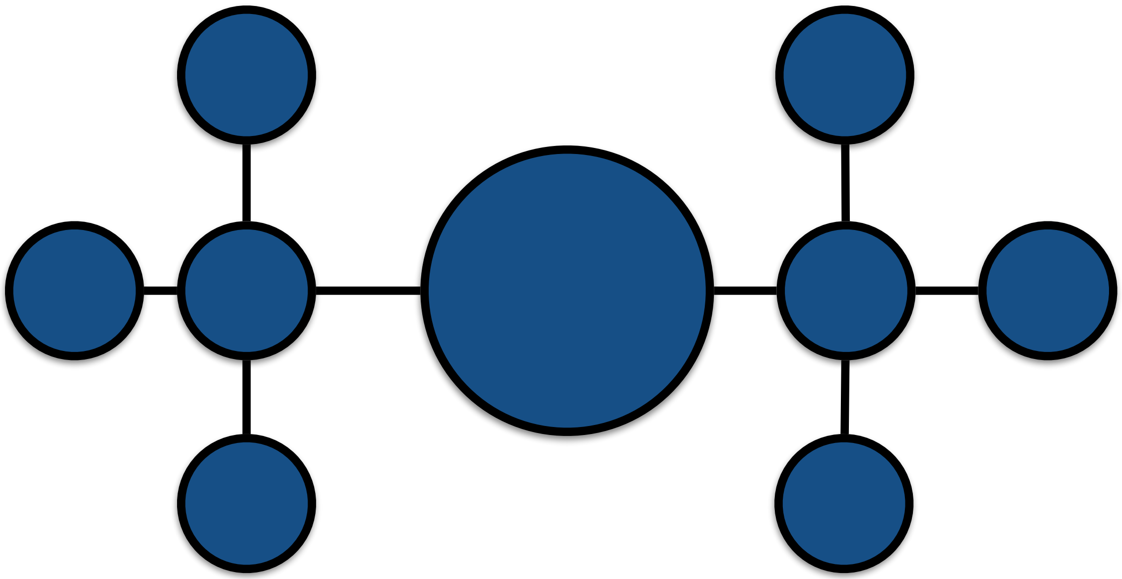
Number of nodes only?

How to measure this betweenness centrality?



Betweeness centrality

- <u>Hypothesis</u>: An important node appears in-between many other nodes (on the path between)



Betweeness centrality

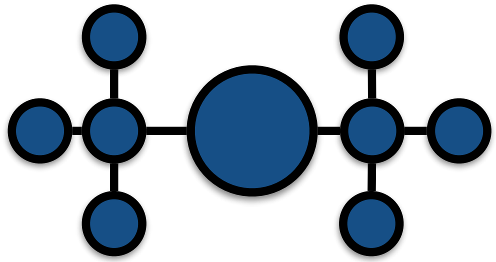
- <u>Definition</u>: Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

Number of shortest path between j and k passing through x

Number of shortest path between j and k

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

Sum over all couples of nodes

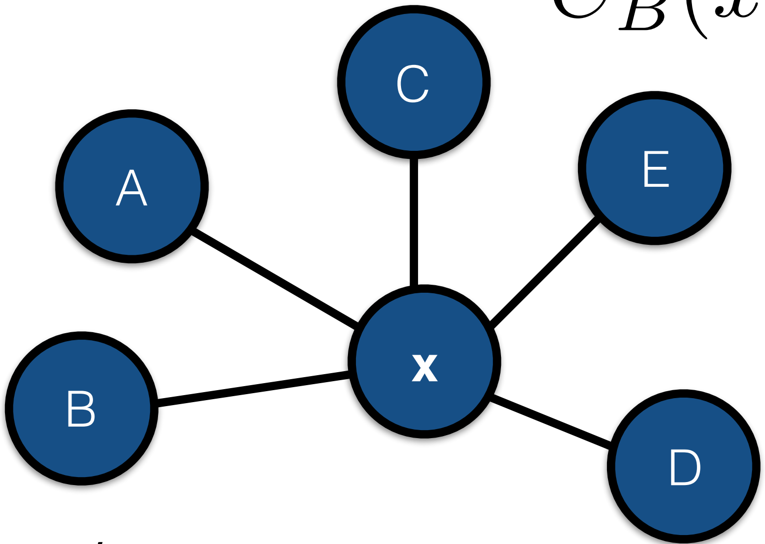


Betweeness centrality: example

Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

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

$C_B(x)=10$

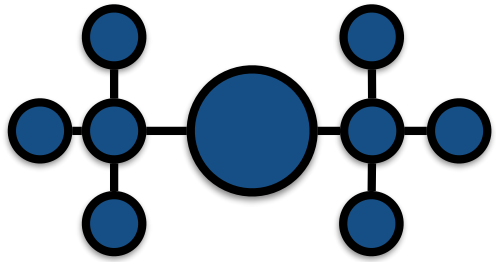


A->B : 1/1

A->C : 1/1 B->C : 1/1

A->D : 1/1 B->D : 1/1 C->D : 1/1

A->E : 1/1 B->E : 1/1 C->E : 1/1 D->E : 1/1

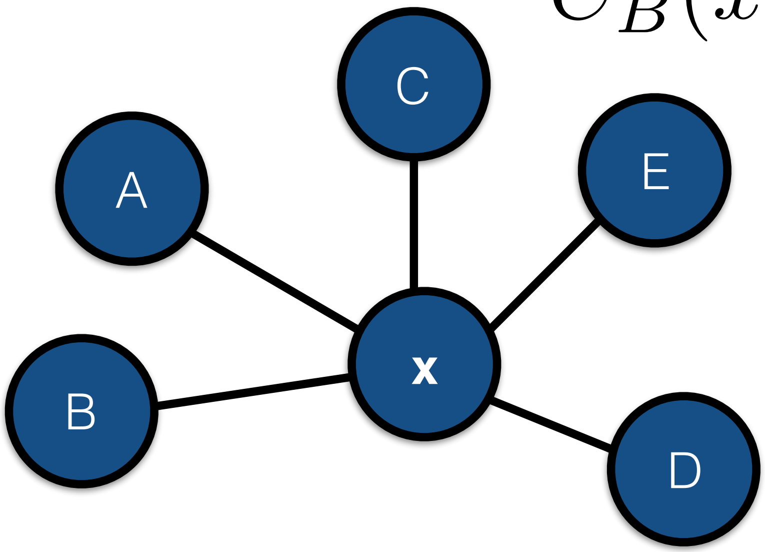


Betweeness centrality: example

Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

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

$C_B(A)=0$



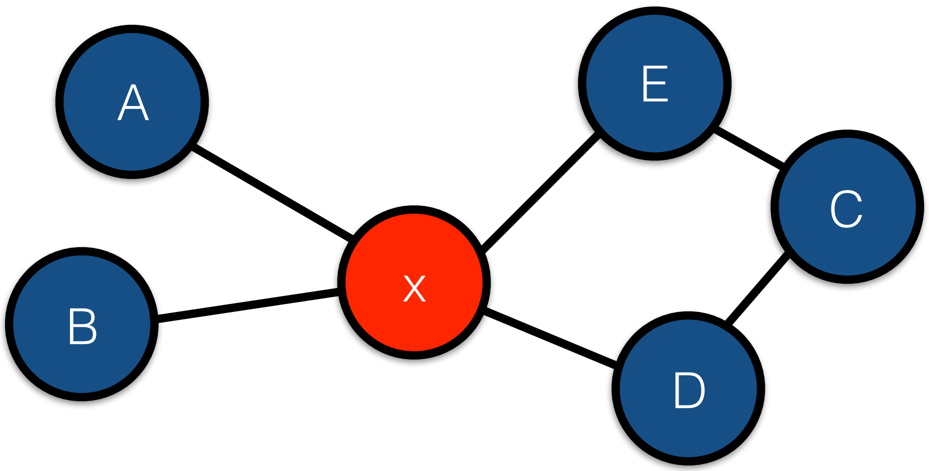
x->B : 0/1
x->C : 0/1 B->C : 0/1
x->D : 0/1 B->D : 0/1 C->D : 0/1
x->E : 0/1 B->E : 0/1 C->E : 0/1 D->E : 0/1

Betweeness centrality: example

Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

$C_B(x)$=7.5

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



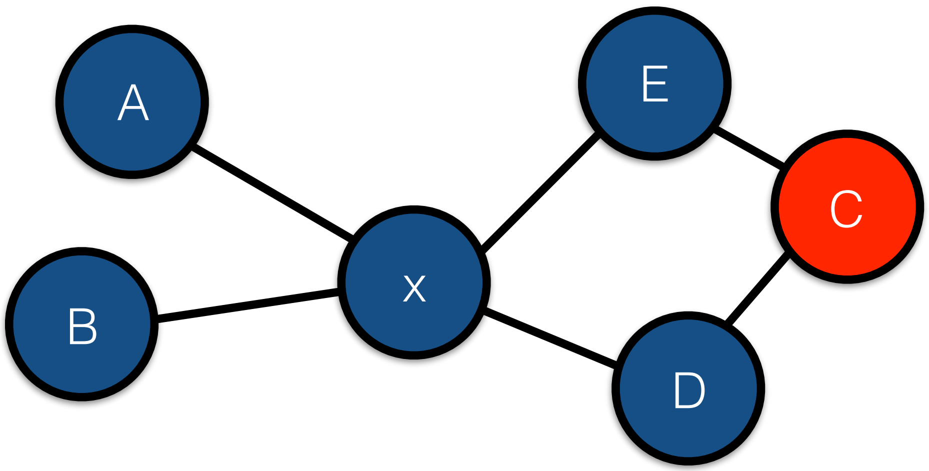
A->B : 1/1 D->E : 1/2
A->C :2/2 B->C : 2/2
A->D : 1/1 B->D : 1/1 C->D : 0/1
A->E : 1/1 B->E : 1/1 C->E : 0/1

Betweeness centrality: example

Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

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

$C_B(c) = 0.5$

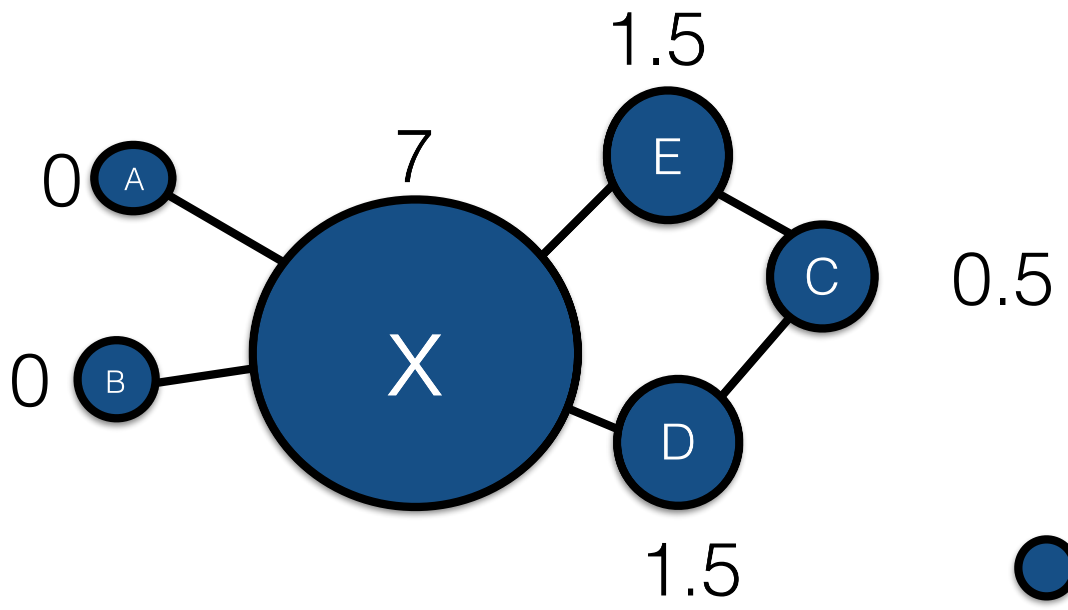


A->B : 0/1
A->X : 0/1 B->X : 0/1
A->D : 0/1 B->D : 0/1 X->D : 0/1
A->E : 0/1 B->E : 0/1 X->E : 0/1 D->E : 1/2

Betweeness centrality: example

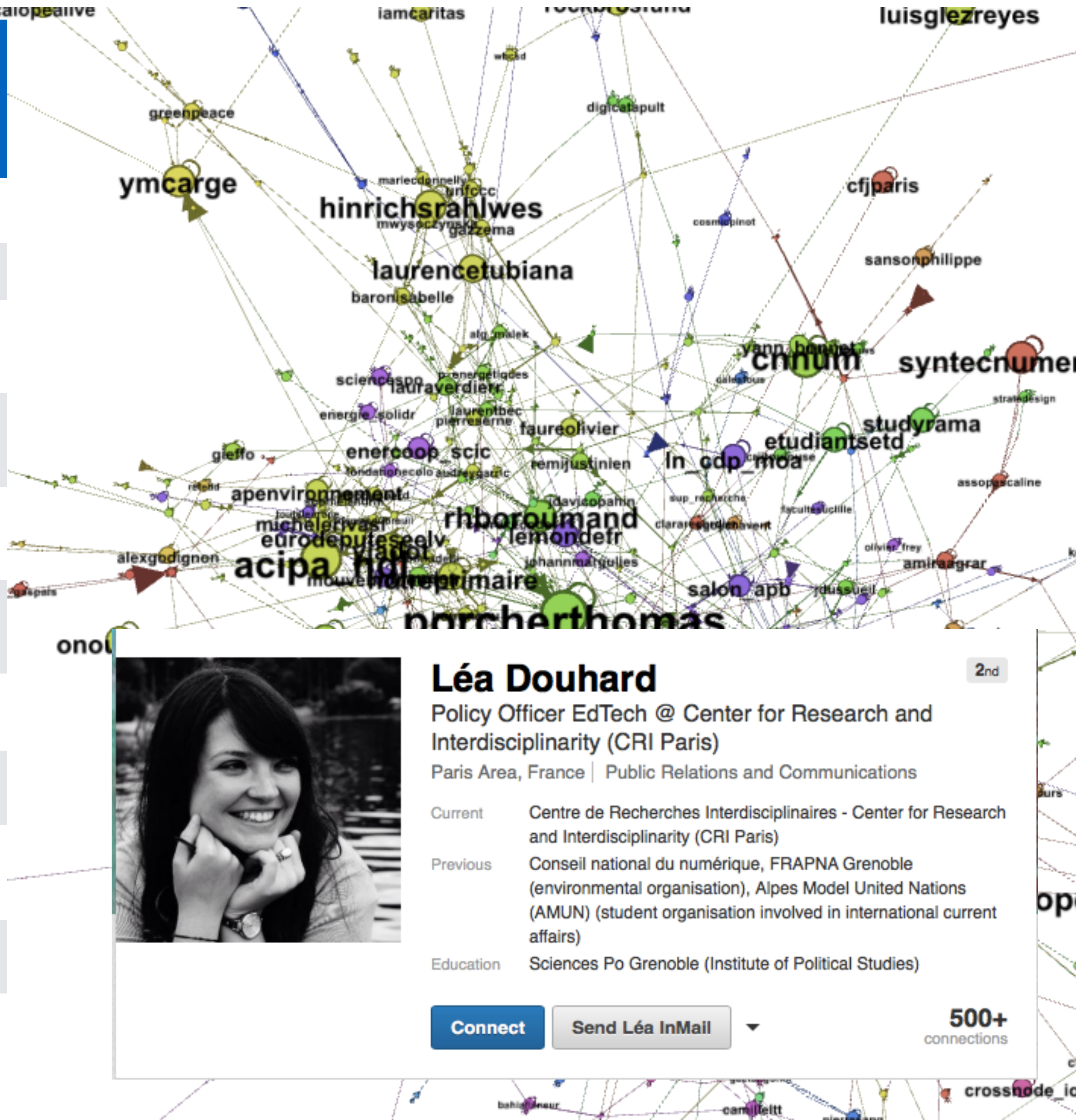
Betweeness centrality of a node i equals the the number of shortest path (between any couple of node) that passes through this nodes divided by the number of shortest path existing between these two nodes.

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



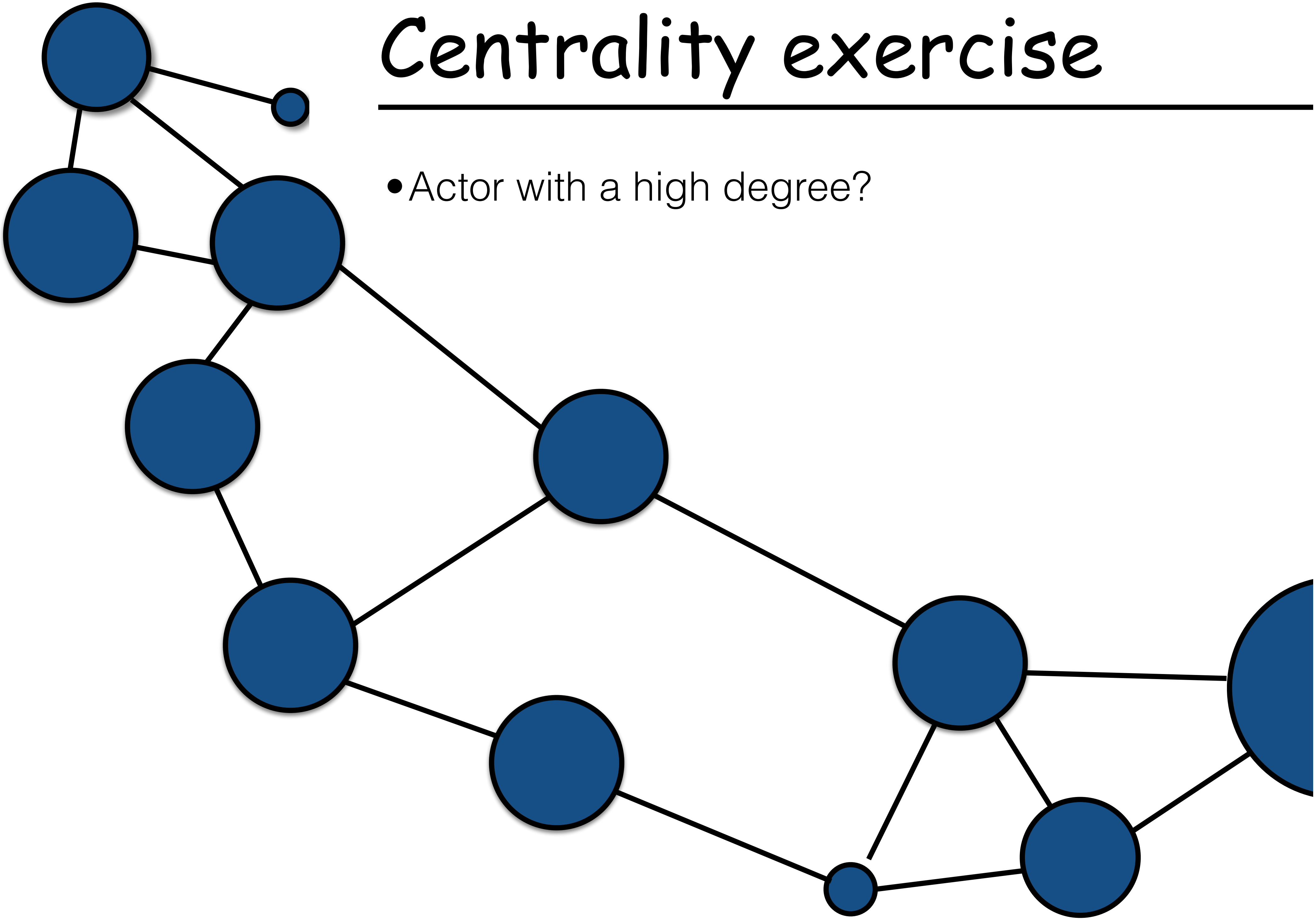
Case study

Pseudonyme	Centralité d'intermédiarité
leadouhard	142127
educpros	119220
cnnum	113837
sup_recherche	102324
jffiorina	101658
anne_hidalgo	91397
s_grandesecoles	76028
esilvparis	68383
conferencedesge	64594
figaro_etudiant	58903
educationfrance	54716



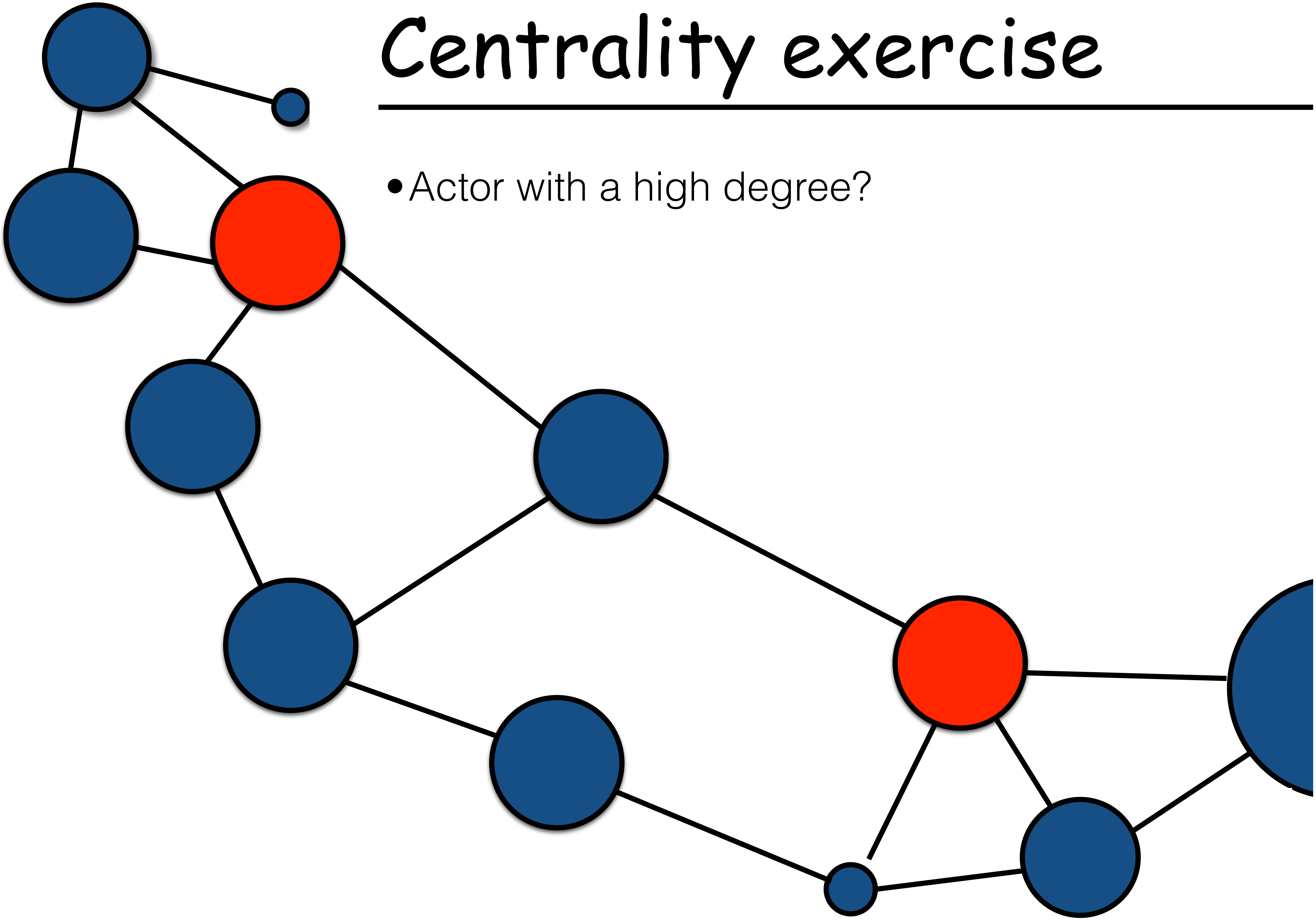
Centrality exercise

- Actor with a high degree?



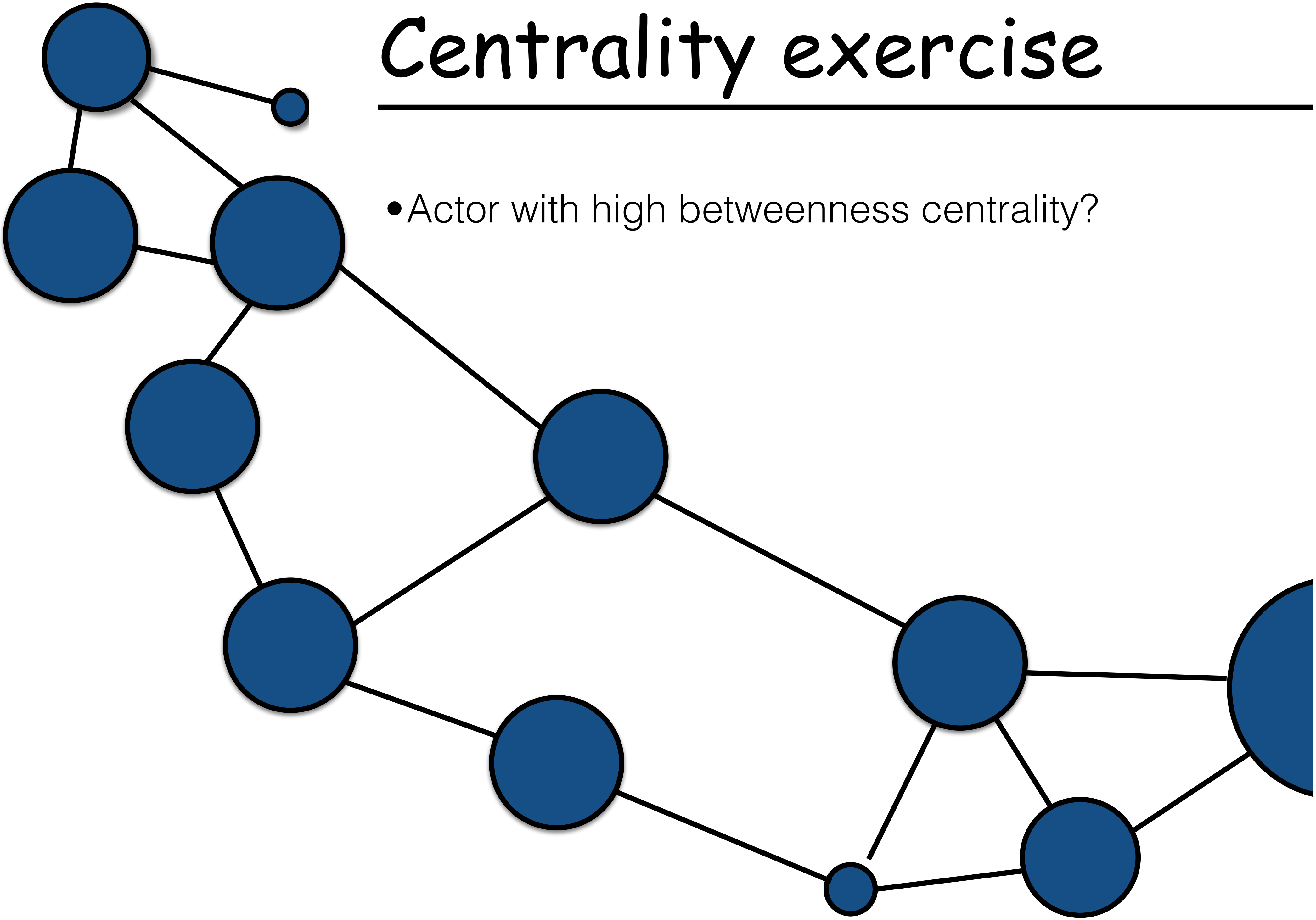
Centrality exercise

- Actor with a high degree?

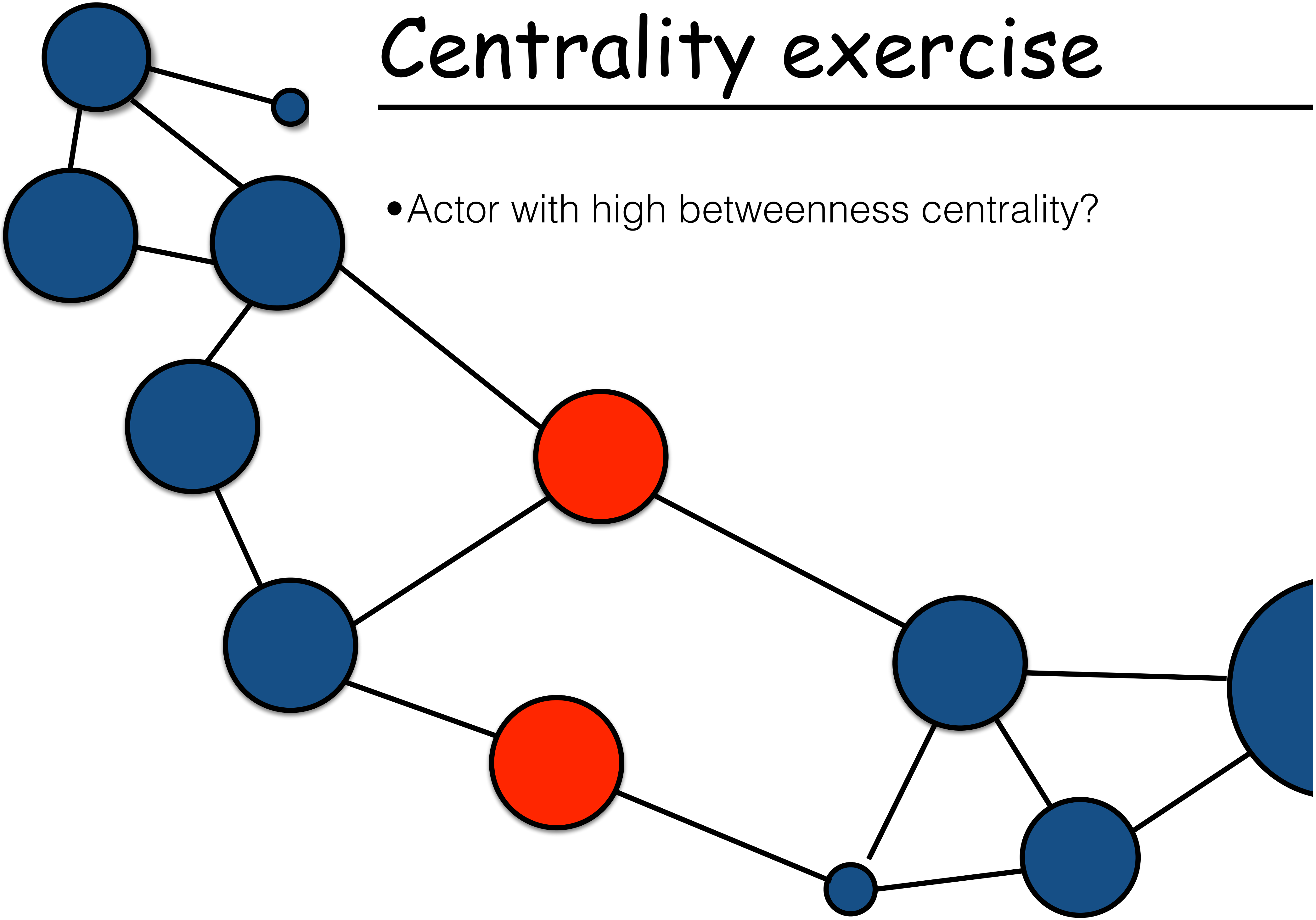


Centrality exercise

- Actor with high betweenness centrality?



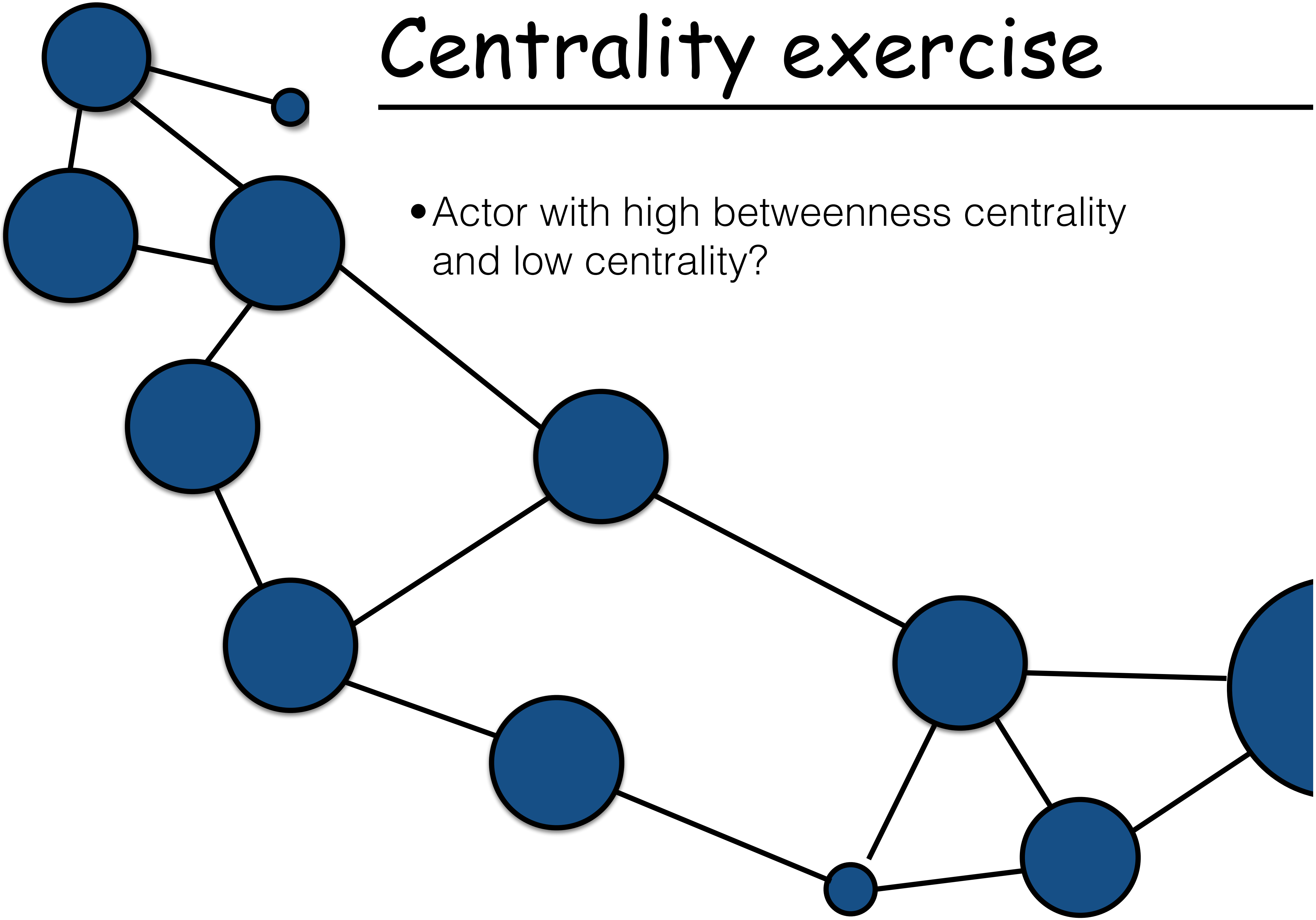
Centrality exercise



- Actor with high betweenness centrality?

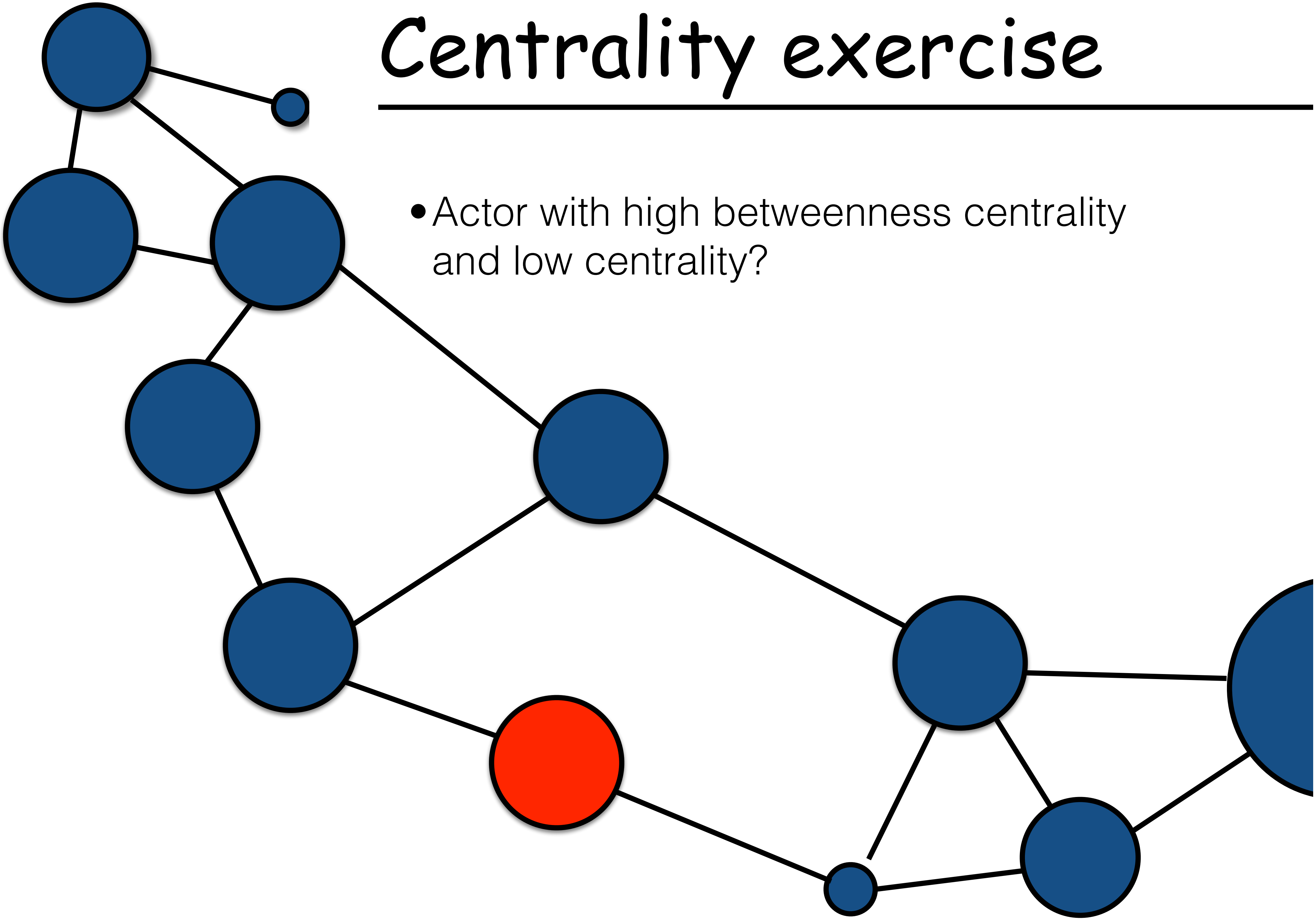
Centrality exercise

- Actor with high betweenness centrality and low centrality?



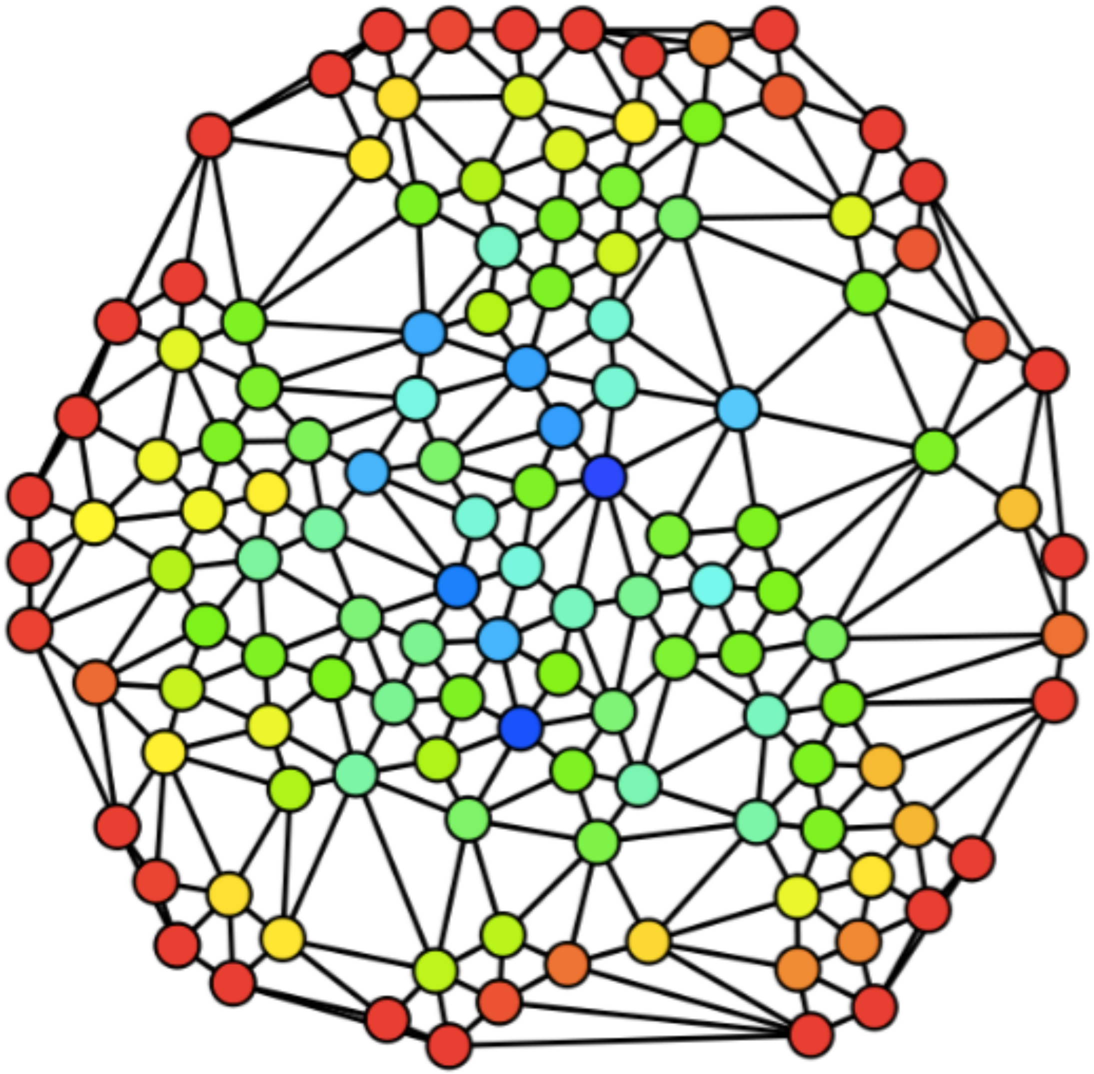
Centrality exercise

- Actor with high betweenness centrality and low centrality?



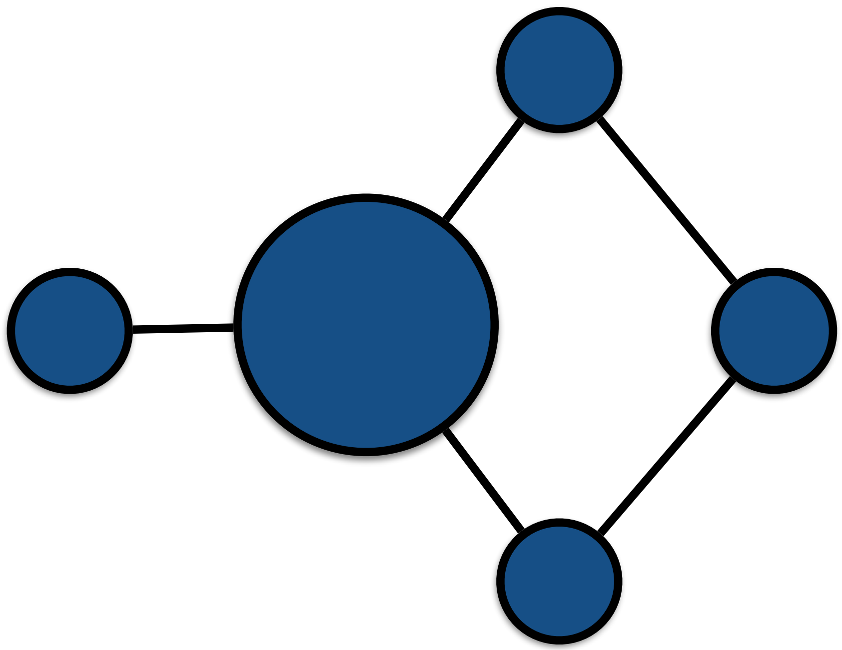
Centrality exercise

Actor with high betweenness centrality?



Closeness centrality

- <u>Hypothesis</u> : an important nodes is a node that is close to every other nodes



Closeness centrality

- Definition : closeness centrality of node i mesures the degree to which this node is close to other nodes.

by convention

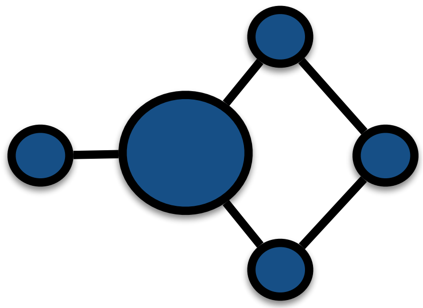
$$\frac{1}{\infty} = 0$$

minimum total distance

distance to other nodes

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

for all nodes

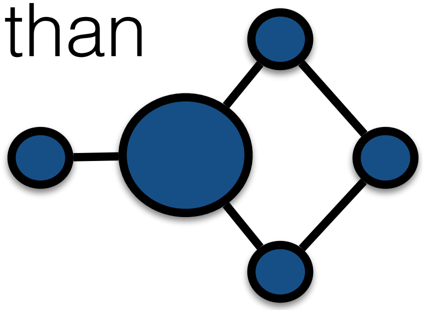


Closeness centrality

- <u>Definition</u> : closeness centrality of node i mesures the degree to which this node is close to other nodes.

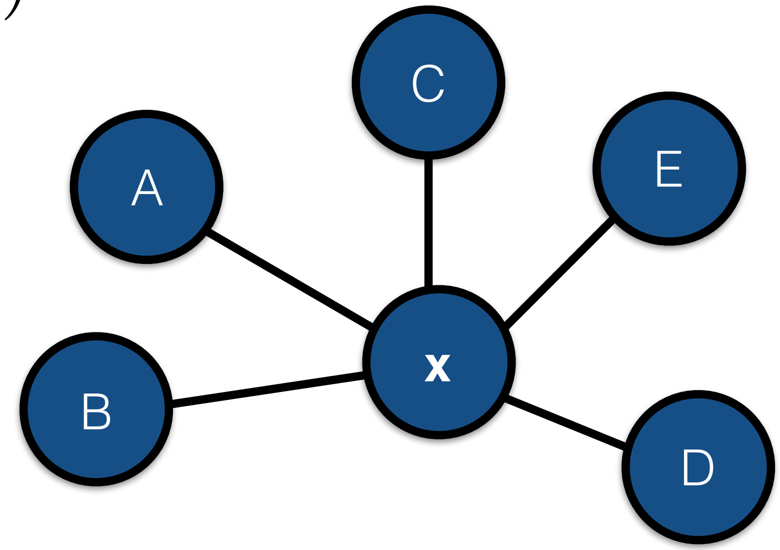
$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

- A node with high closeness centrality is closer to other nodes
- This node can reach any other node quicker than other nodes



Closeness centrality

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$



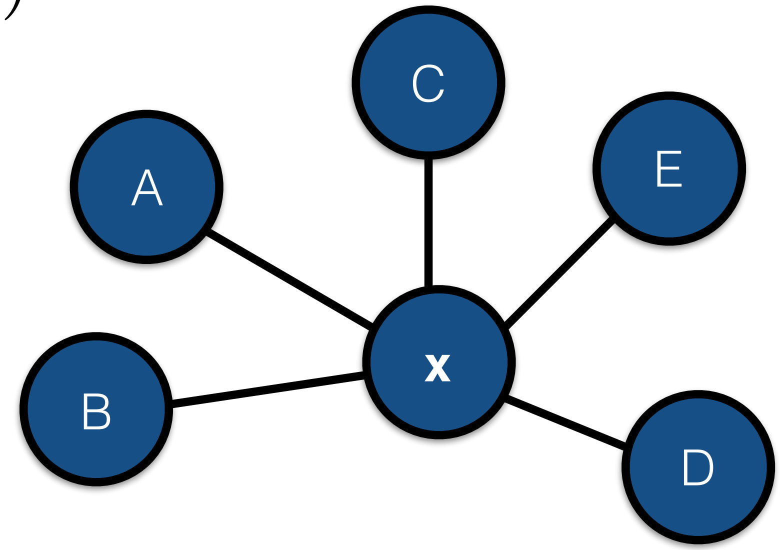
$C_P(x)=(6-1)*(1/1+1/1+1/1+1/1+1/1)$

$C_P(x)=5/5$

$C_P(x)=1$

Closeness centrality

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$



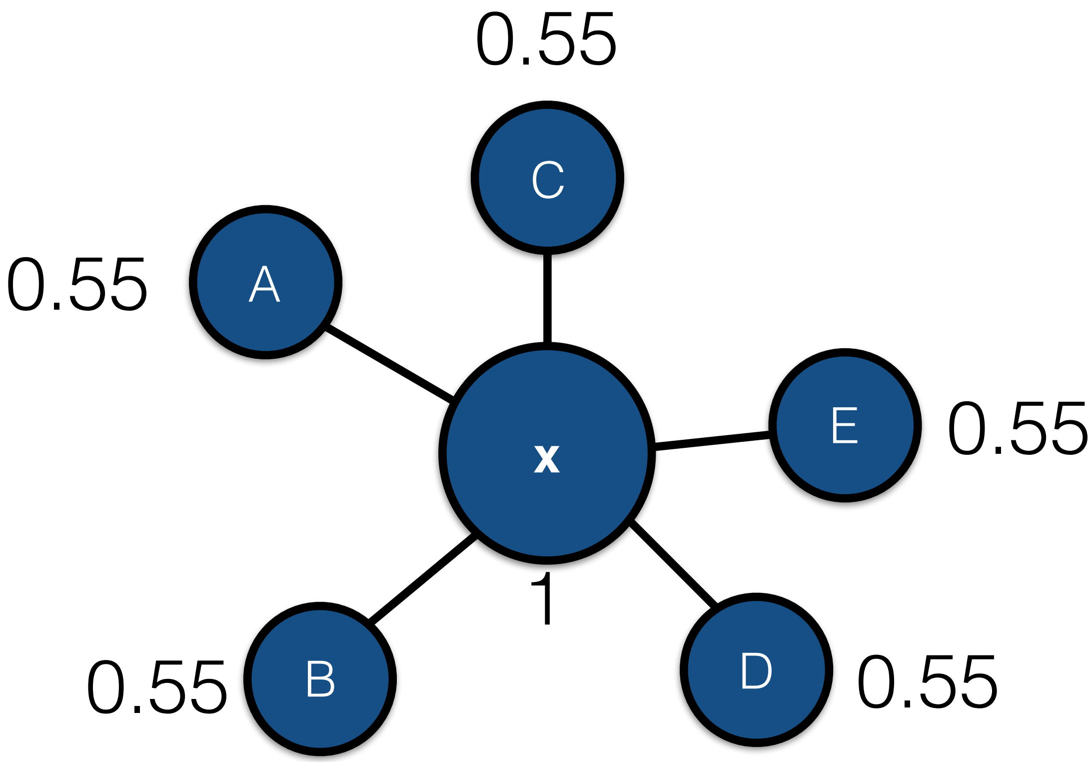
$C_P(A)=(6-1)*(1/1+1/2+1/2+1/2+1/2)$

$C_P(A)=5/9$

$C_P(A)=0.55$

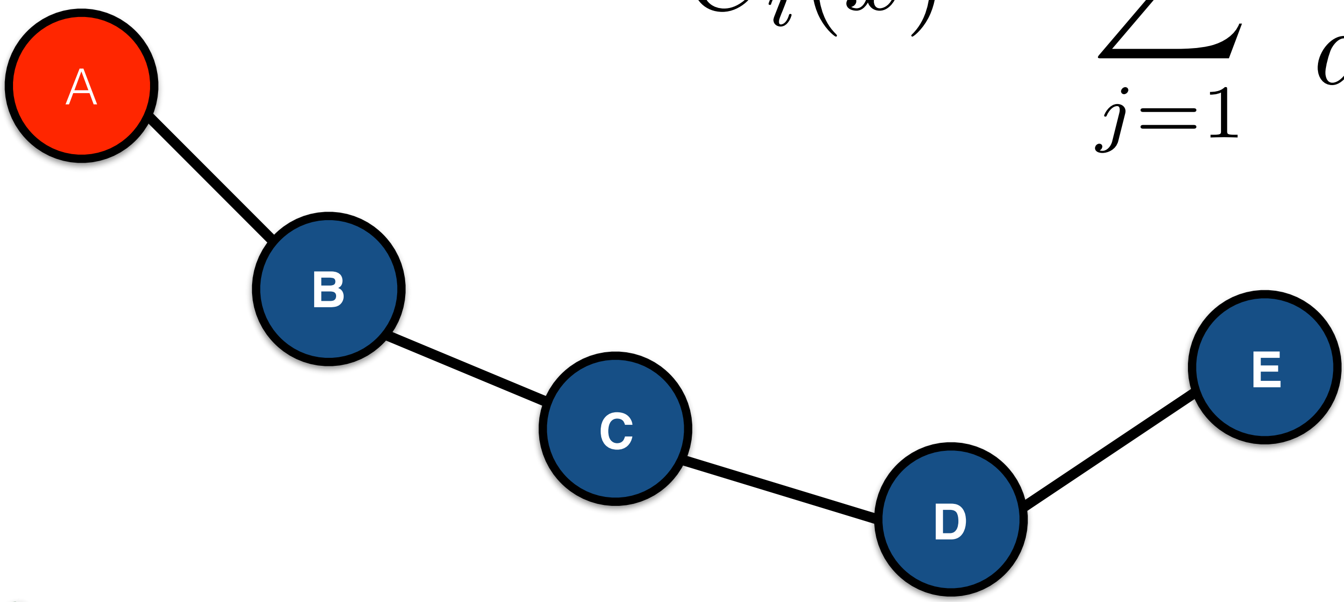
Closeness centrality

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$



Closeness centrality

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

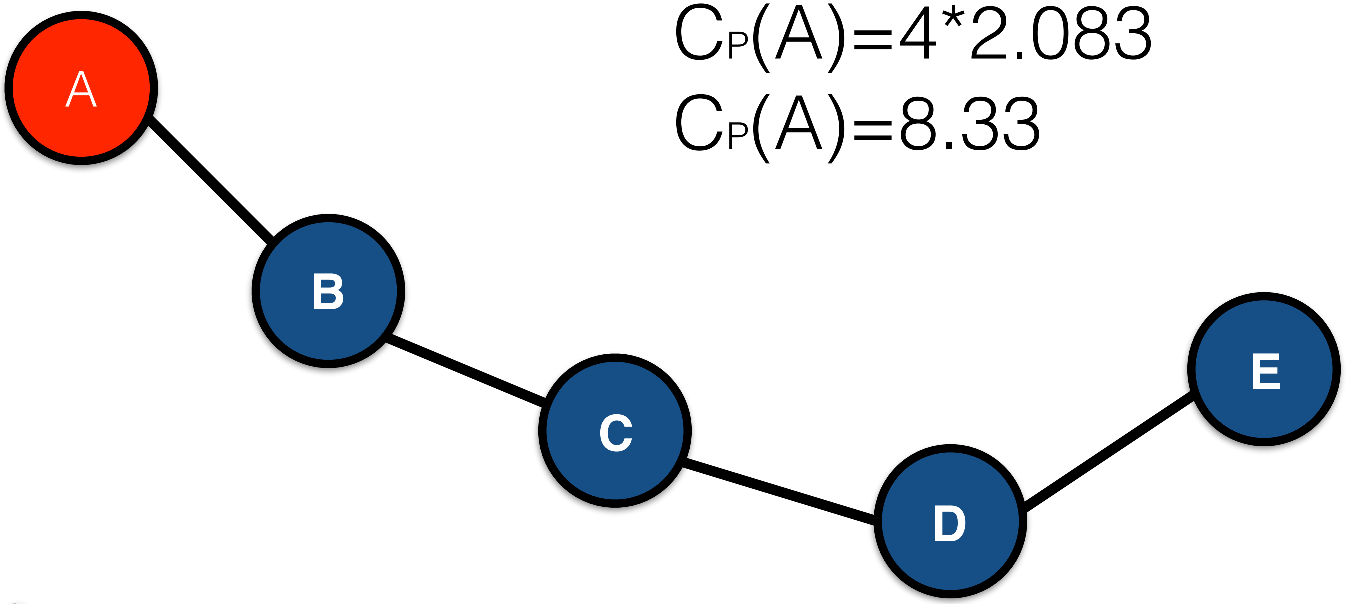


Closeness centrality

$C_P(A)=(5-1)*(1/1+1/2+1/3+1/4)$

$C_P(A)=4*2.083$

$C_P(A)=8.33$

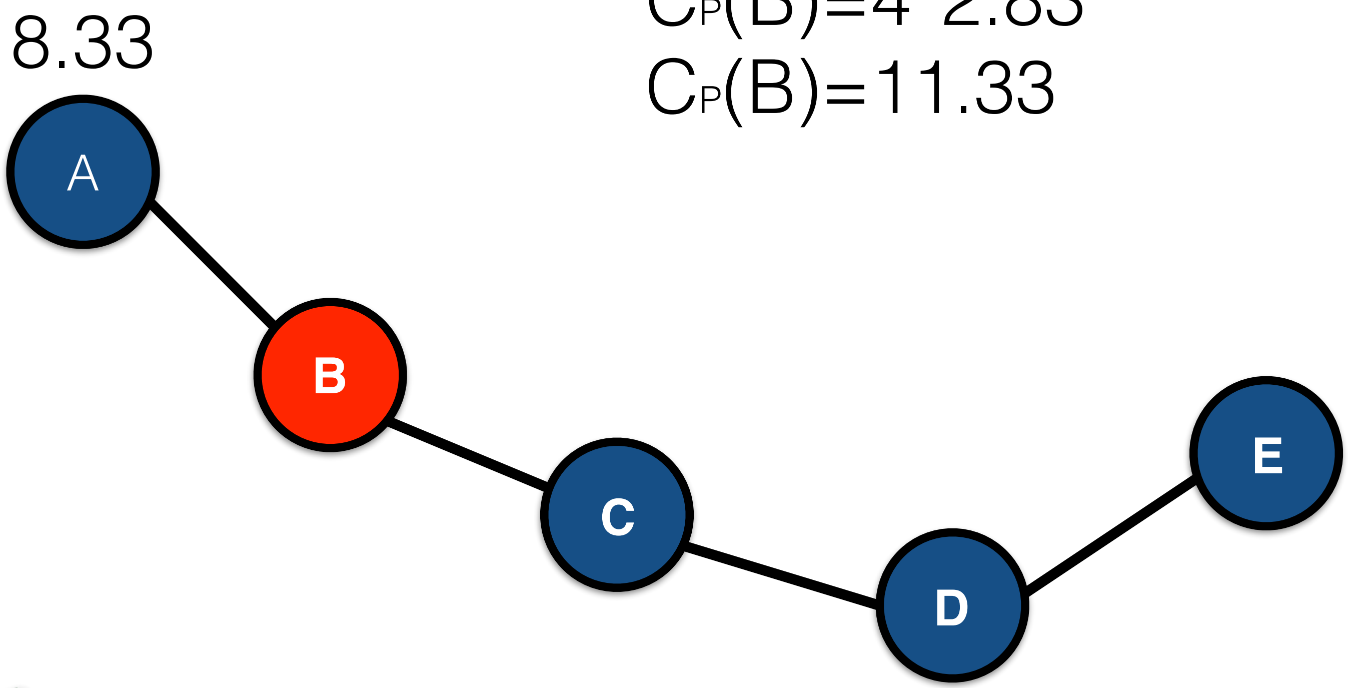


Closeness centrality

$C_P(B)=(5-1)*(1/1+1/1+1/2+1/3)$

$C_P(B)=4*2.83$

$C_P(B)=11.33$



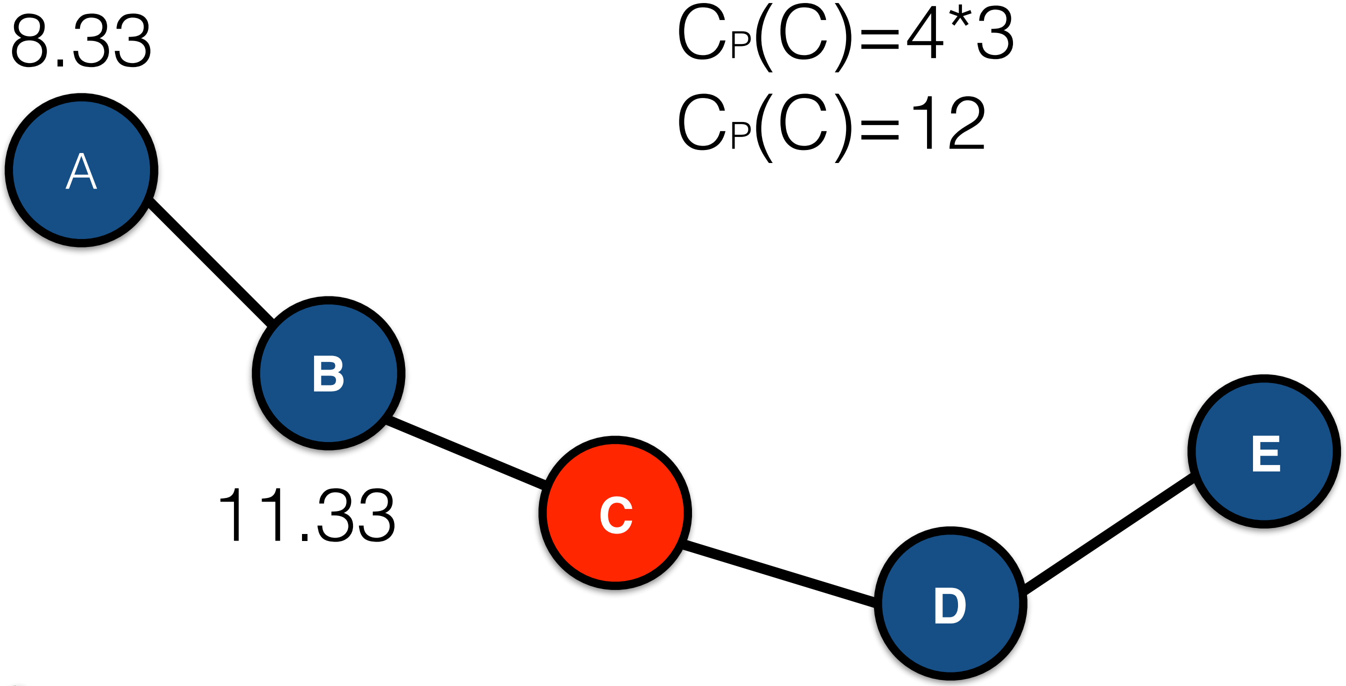
$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

Closeness centrality

$C_P(C)=(5-1)/(1/1+1/2+1/1+1/2)$

$C_P(C)=4*3$

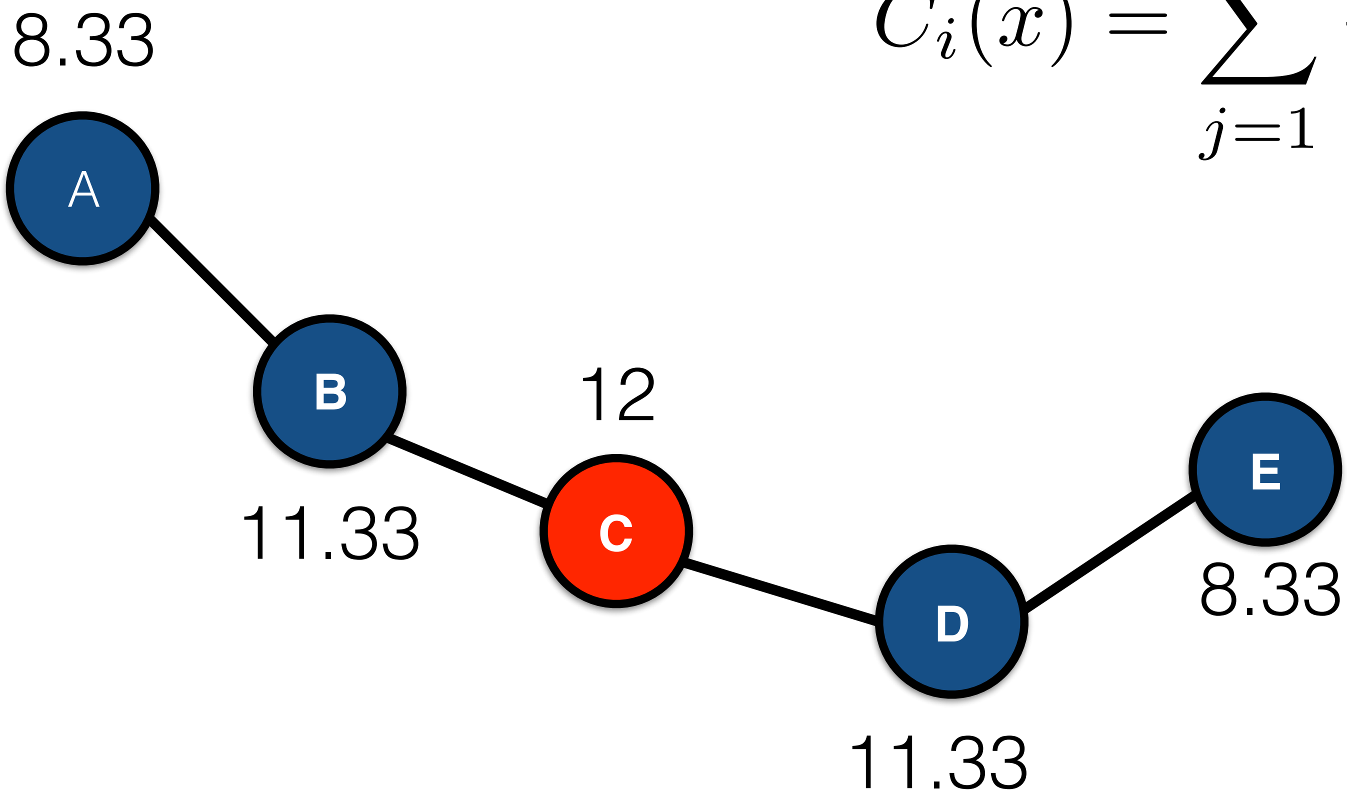
$C_P(C)=12$



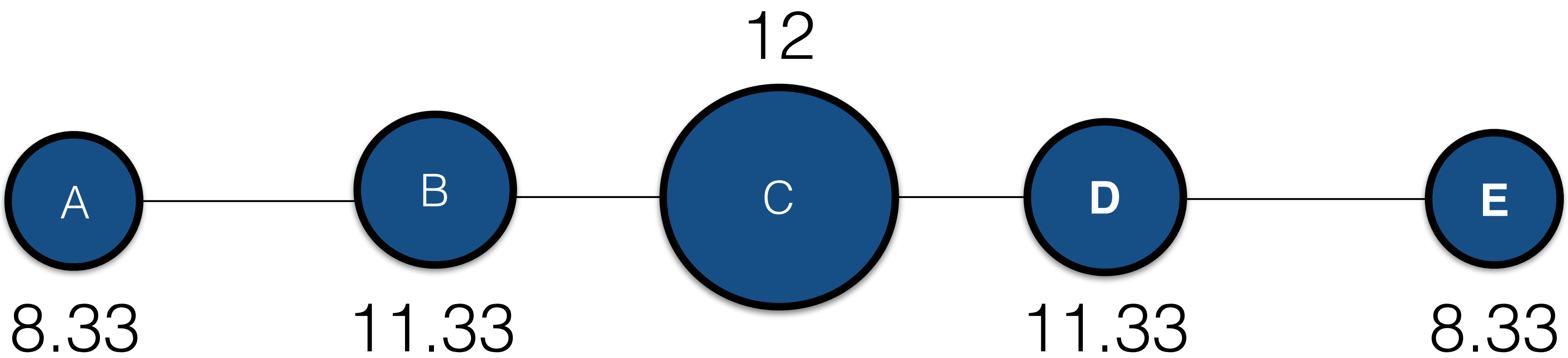
$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

Closeness centrality

$$C_i(x) = \sum_{j=1}^{N} \frac{(N-1)}{d(x,j)}$$

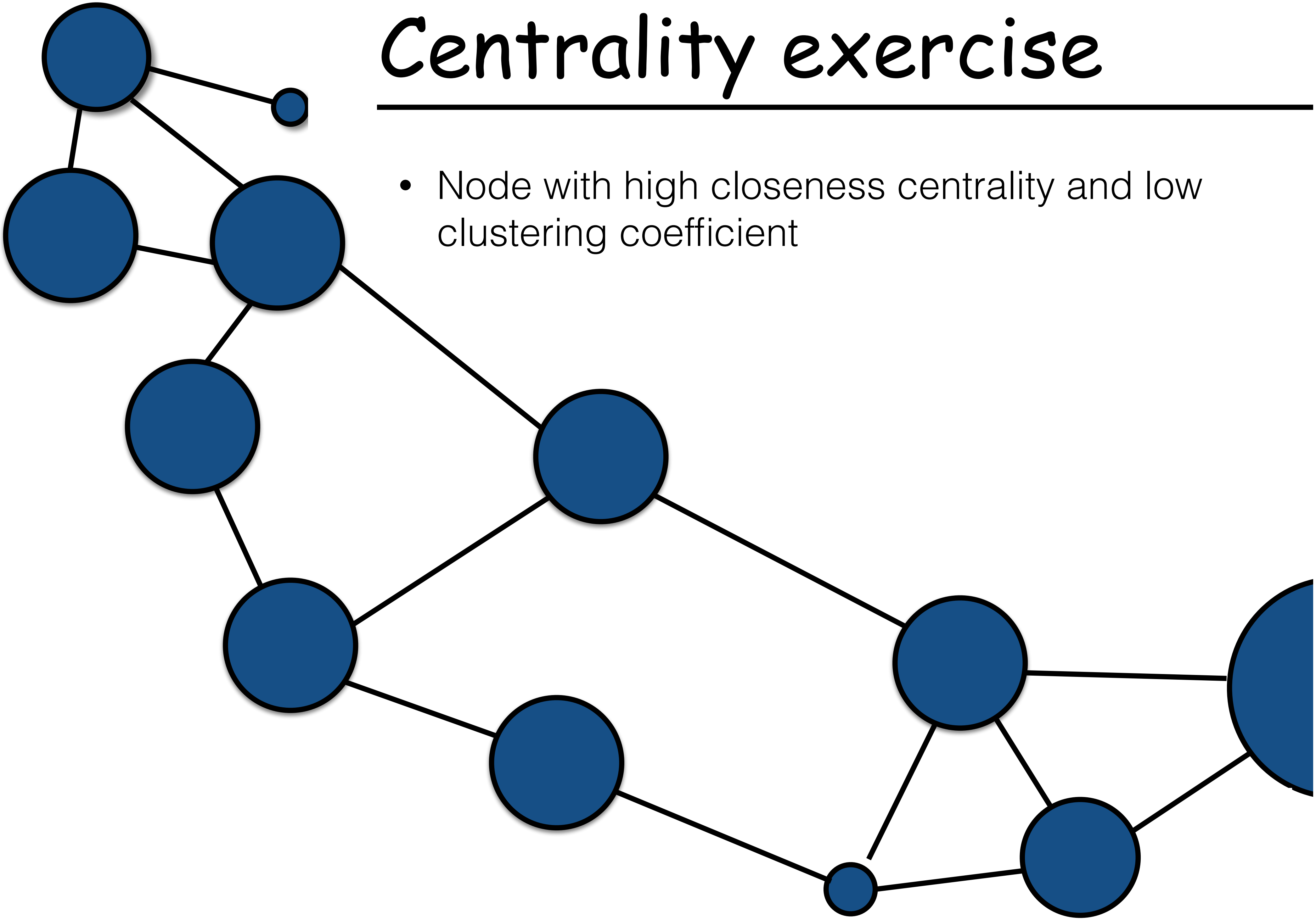


Closeness centrality



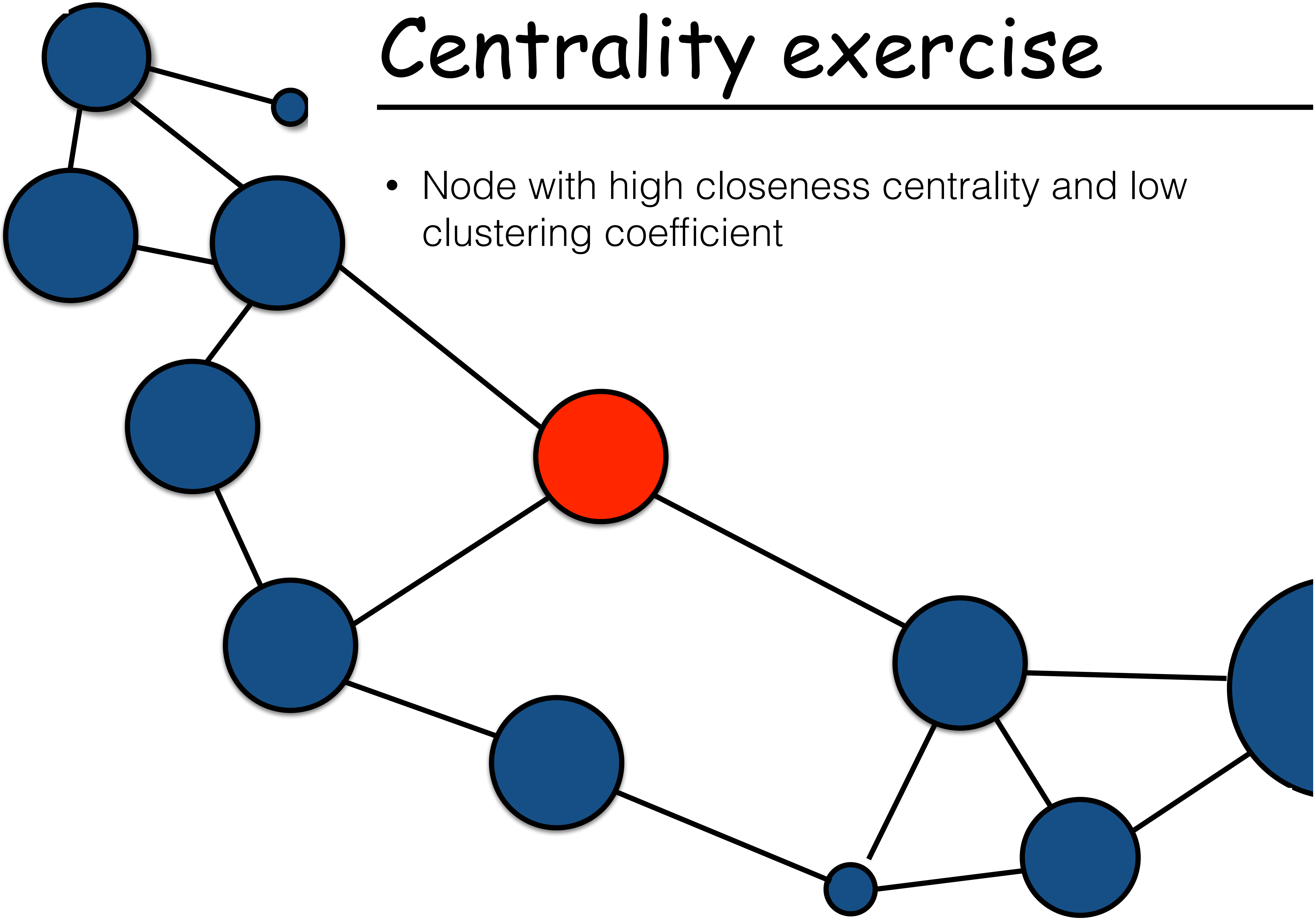
Centrality exercise

- Node with high closeness centrality and low clustering coefficient



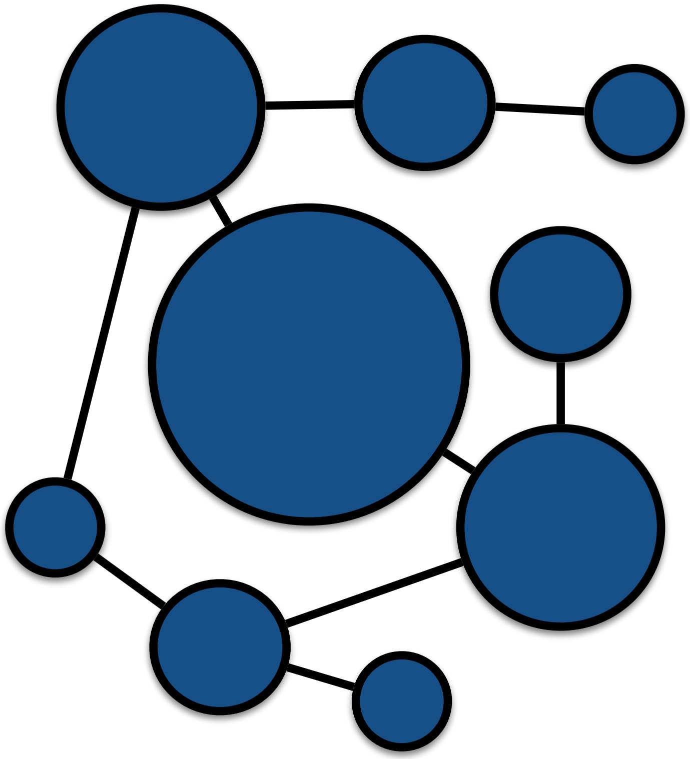
Centrality exercise

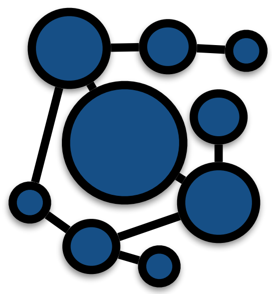
- Node with high closeness centrality and low clustering coefficient



Eigenvalues centrality

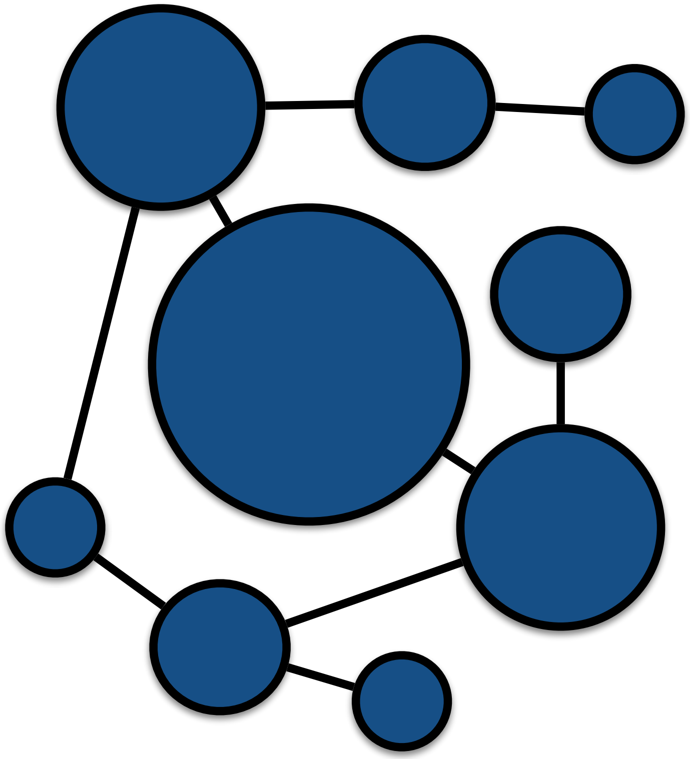
- <u>Hypothesis</u> : A node is central if it is connected to central nodes.

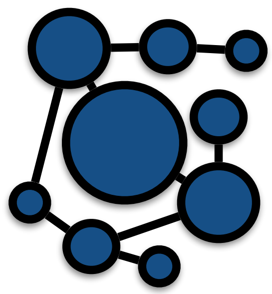




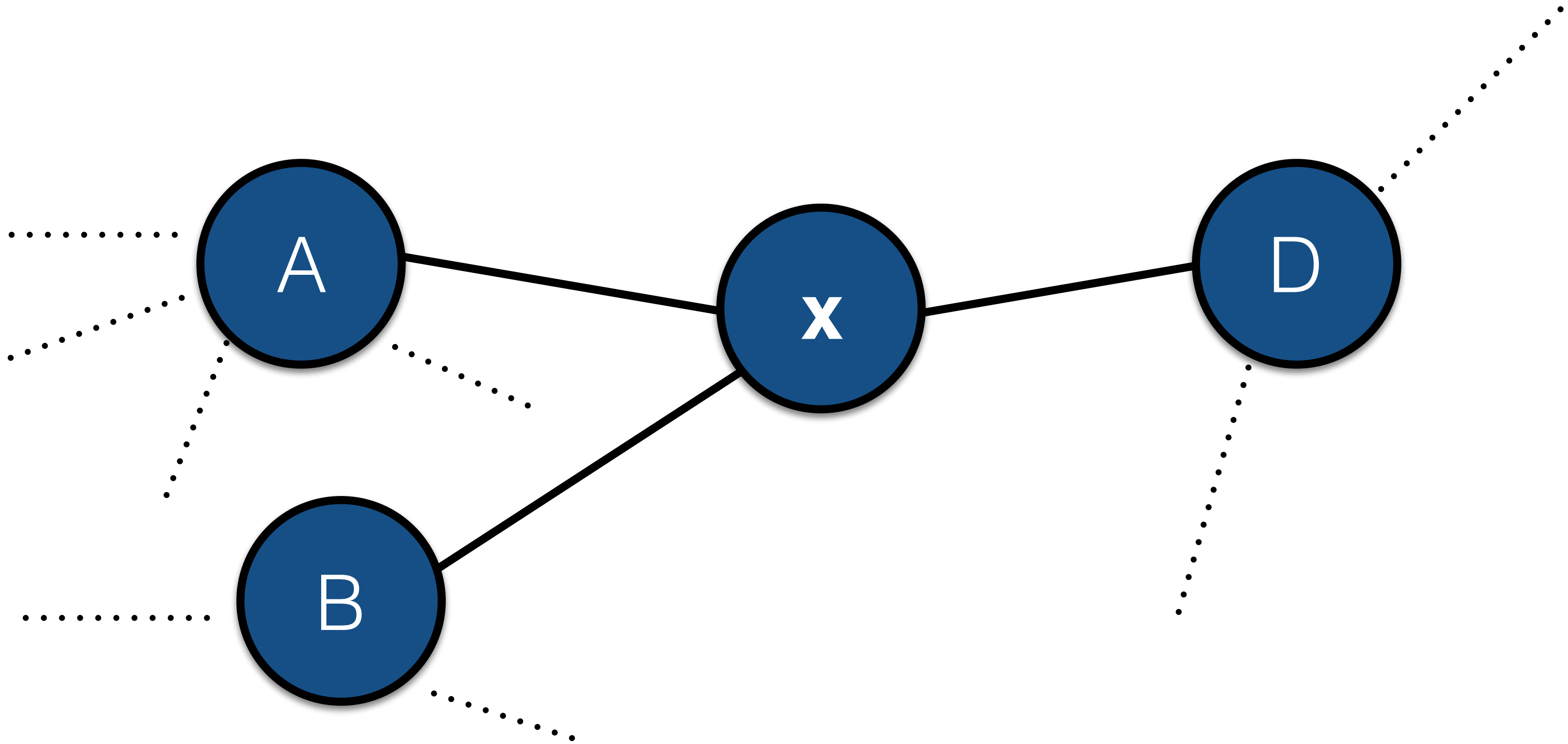
Eigenvalues centrality

- Similar to a vote mechanism where each page vote for the election of the best page. Each page gives its vote to its neighbours.

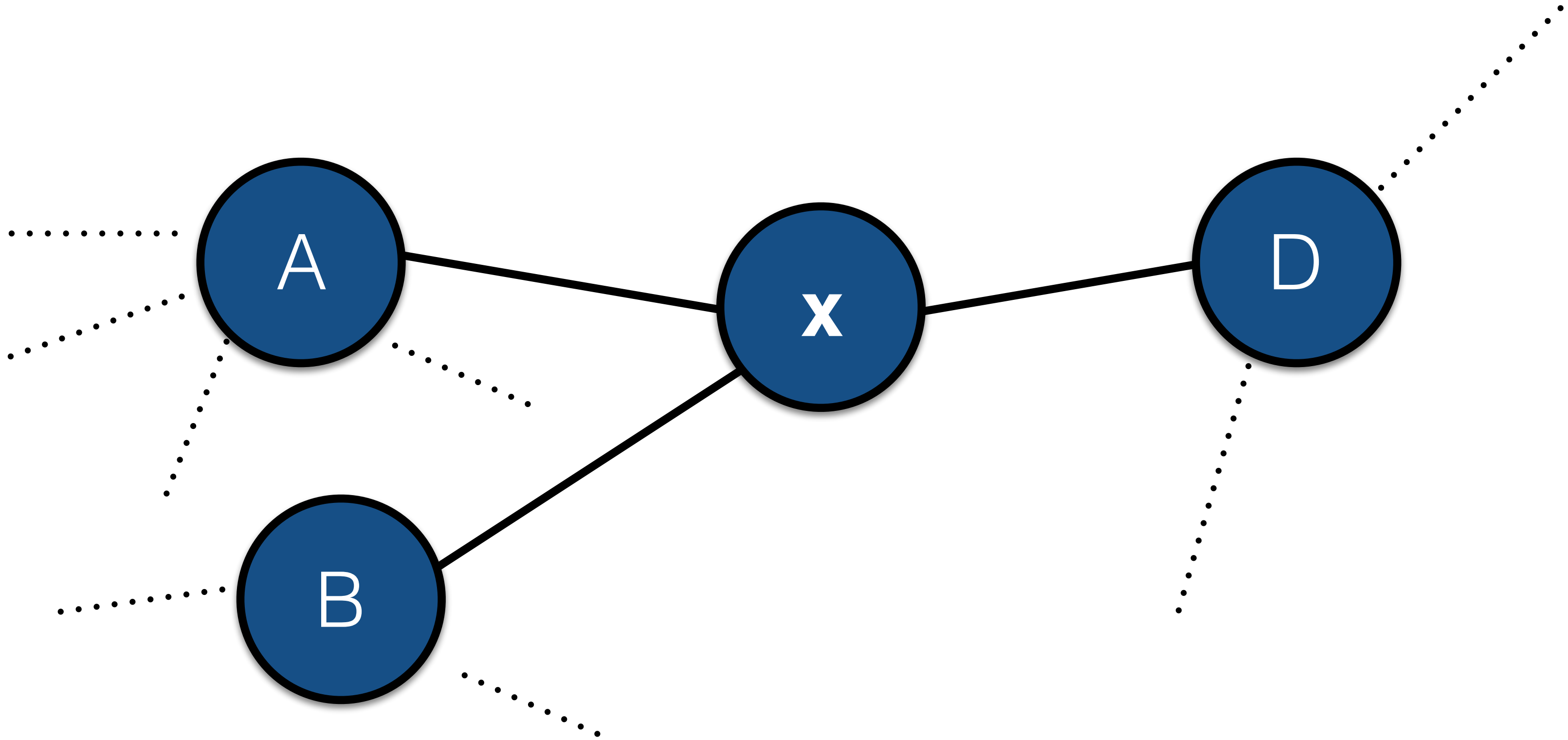




Eigenvalues centrality

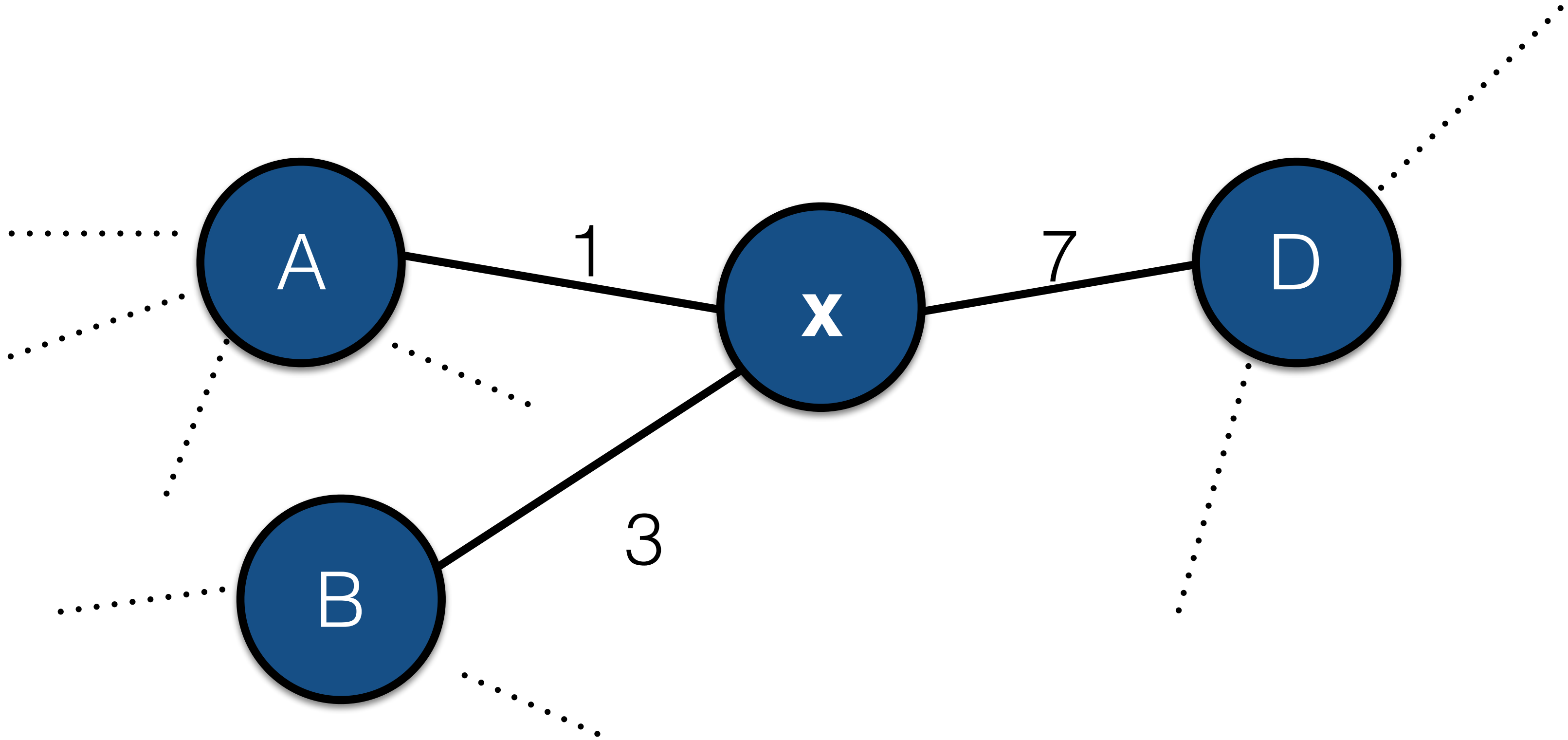


Eigenvalues centrality



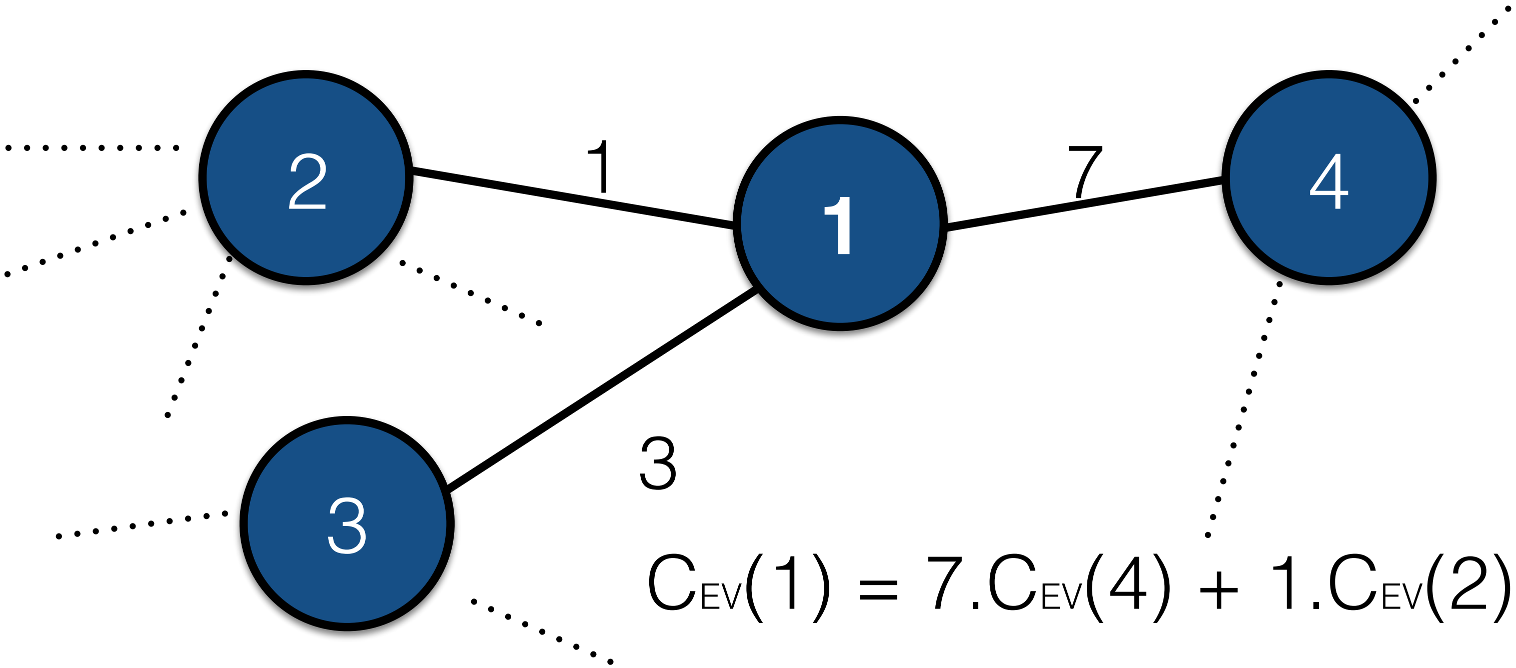
$$C_{EV}(X) = C_{EV}(D) + C_{EV}(A) + C_{EV}(B)$$

Eigenvalues centrality



$$C_{EV}(X) = 7.C_{EV}(D) + 1.C_{EV}(A) + 3.C_{EV}(B)$$

Eigenvalues centrality



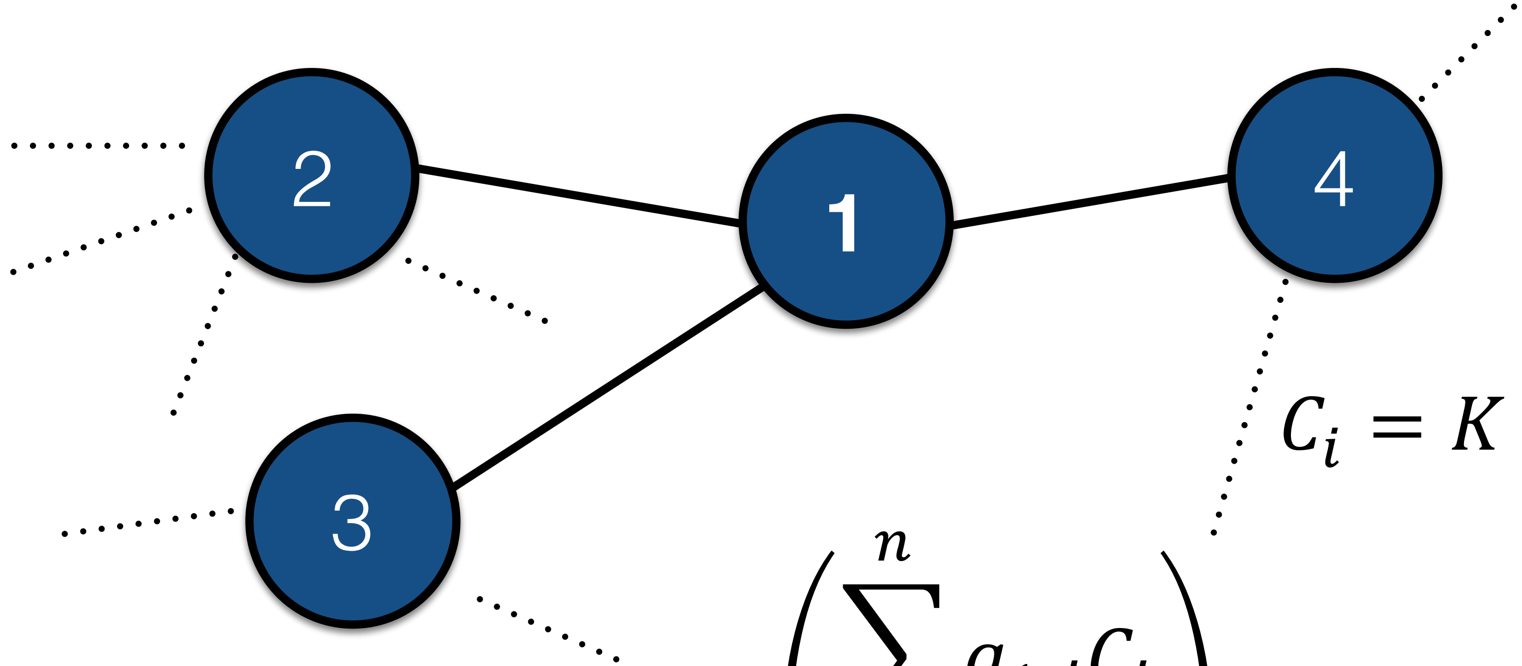
$$C_{EV}(1) = 7.C_{EV}(4) + 1.C_{EV}(2) + 3.C_{EV}(3)$$

$$C_{EV}(i) = K\,(\,a_{i,1}.C_{EV}(1) + a_{i,1}.C_{EV}(2) + \ldots + a_{i,n}.C_{EV}(n))$$

$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

Eigenvalues centrality

$$C_{EV}(1) = 7.C_{EV}(4) + 1.C_{EV}(2) + 3.C_{EV}(3)$$

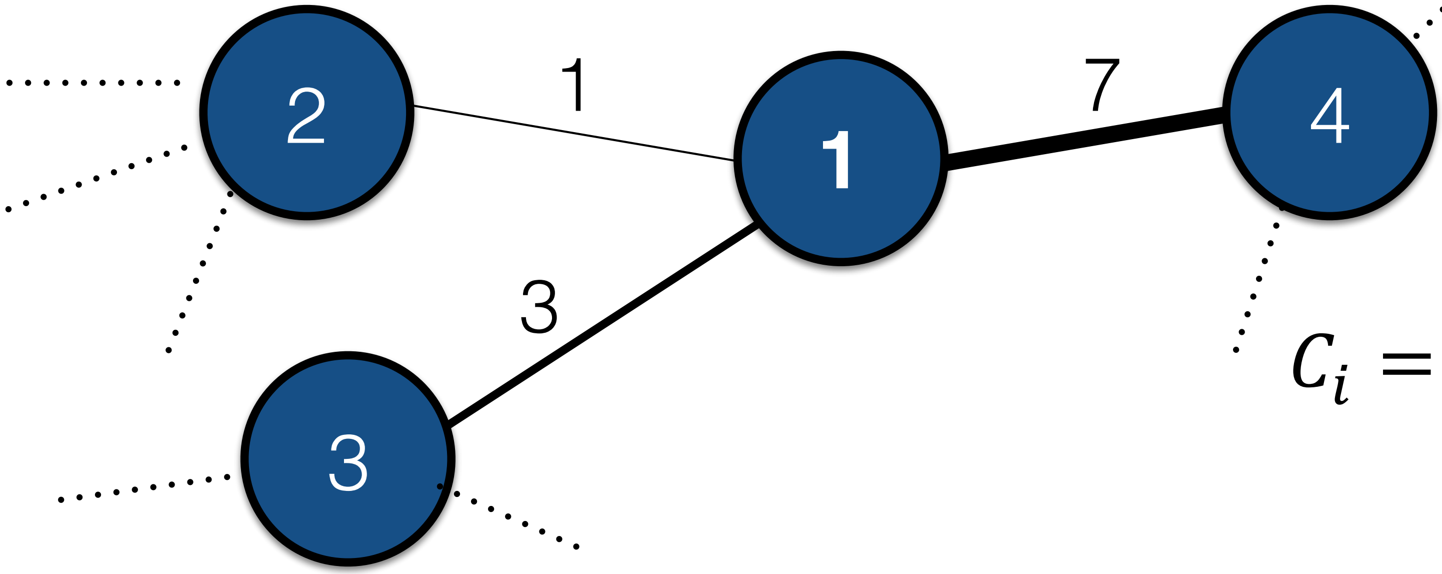


$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

$$\begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} = K \begin{pmatrix} \sum_{j=1}^{n} a_{1,j} C_j \\ \vdots \\ \sum_{j=1}^{n} a_{n,j} C_j \end{pmatrix}$$

Eigenvalues centrality

$C_{EV}(1) = 7.C_{EV}(4) + 1.C_{EV}(2) + 3.C_{EV}(3)$

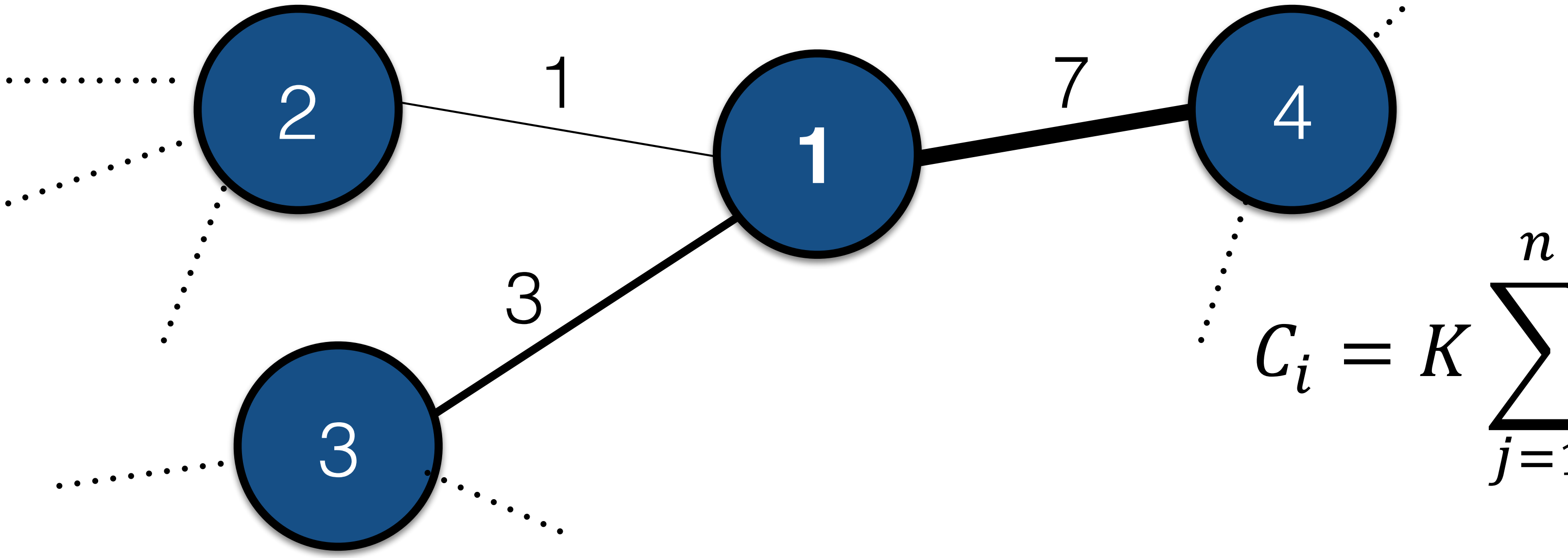


$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

$$\begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} = K \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix}$$

Eigenvalues centrality

$C_{EV}(1) = 7.C_{EV}(4) + 1.C_{EV}(2) + 3.C_{EV}(3)$



$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

$$C = K(A * C)$$

$$\frac{1}{K} C = AC$$

$$\lambda C = AC$$

Eigenvalues centrality

- **C** is an Eigenvector of **A** associated with the Eigenvalue **λ**
- We choose the highest Eigenvalue to make sure that every Eigenvector components are positive (Perron-Frobenius theorem)

$$\lambda C = AC$$

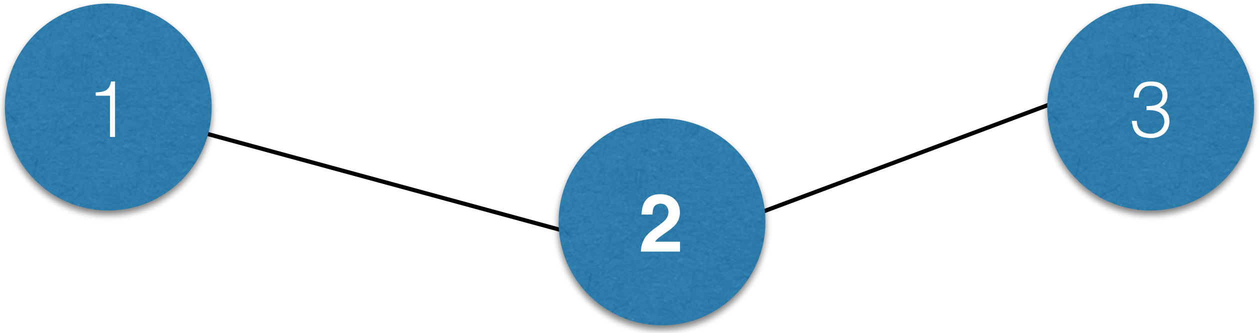
$$(A - \lambda I)C = 0$$

Eigenvalues centrality

- <u>Definition</u> : Eigenvalue centrality of a node *i* equals to the i^{th} component du Eigenvector associated with the highest Eigenvalues of the adjacency matrix of the graph.

$$C_{EV}(i) = x_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{i,j}x_j$$

Eigenvalues centrality

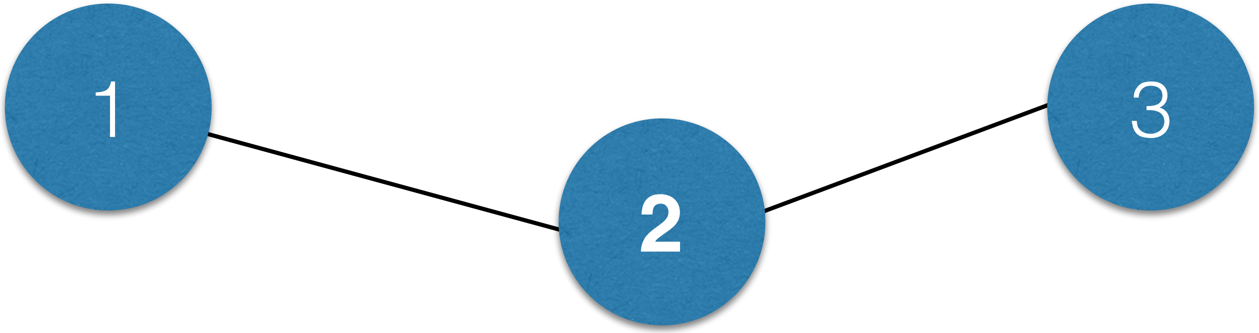


$$(A - \lambda I)C = 0$$

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

$$\begin{pmatrix} 0 - \lambda & 1 & 0 \\ 1 & 0 - \lambda & 1 \\ 0 & 1 & 0 - \lambda \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = 0$$

Eigenvalues centrality

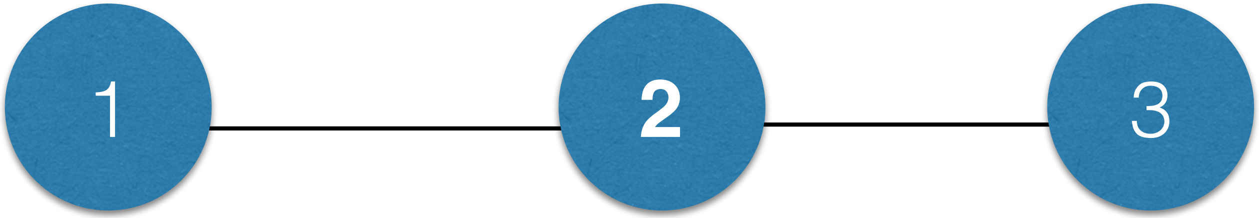


$$\begin{pmatrix} 0-\lambda & 1 & 0 \\ 1 & 0-\lambda & 1 \\ 0 & 1 & 0-\lambda \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = 0$$

Eigenvectors are the roots of the characteristic polynomial

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix} = 0$$

Eigenvalues centrality



$$\begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix} = 0$$

$$-\lambda * (\lambda^2 - 1) - (-\lambda) = 0$$

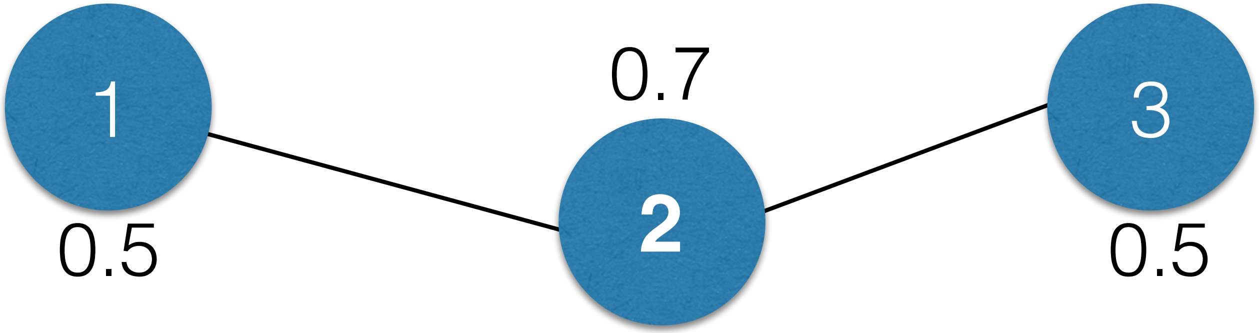
$$-\lambda^3 + 2\lambda = 0$$

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

$$\lambda = 0$$

$$\lambda = \pm\sqrt{2}$$

Eigenvalues centrality



$$\begin{cases} -\sqrt{2}C_1 + C_2 = 0 \\ C_1 - \sqrt{2}C_2 + C_3 = 0 \\ C_2 - \sqrt{2}C_3 = 0 \end{cases}$$

$$\begin{cases} C_3 = C_1 \\ C_2 = \sqrt{2}C_1 \end{cases}$$

$$\begin{cases} C_2 = \sqrt{2}C_1 \\ C_3 = \dfrac{1}{\sqrt{2}}C_2 \\ C_1 - \sqrt{2}C_2 + C_3 = 0 \end{cases}$$

$$c = \begin{pmatrix} C_1 \\ \sqrt{2}C_1 \\ C_1 \end{pmatrix}$$

$$c = \begin{pmatrix} 0.5 \\ 0.7 \\ 0.5 \end{pmatrix}$$

The PageRank algorithm

- Algorithm proposed by Lary Page
- Modification of the Eigenvalue centrality to fix some large scale issues
- Algorithm still partially used by Google and published in 1997

Eigenvalues centrality

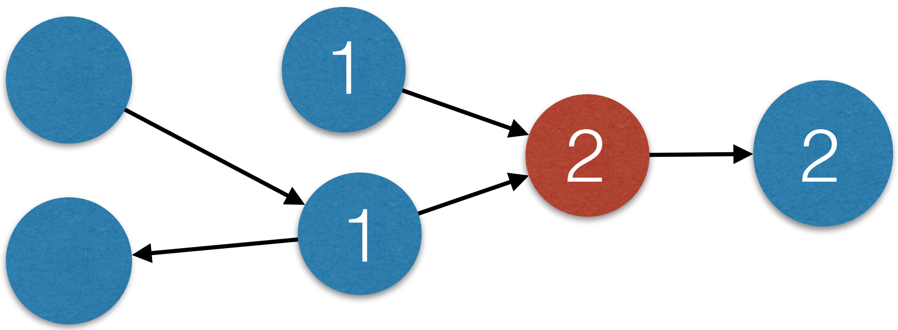
- Eigenvalues centrality

$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

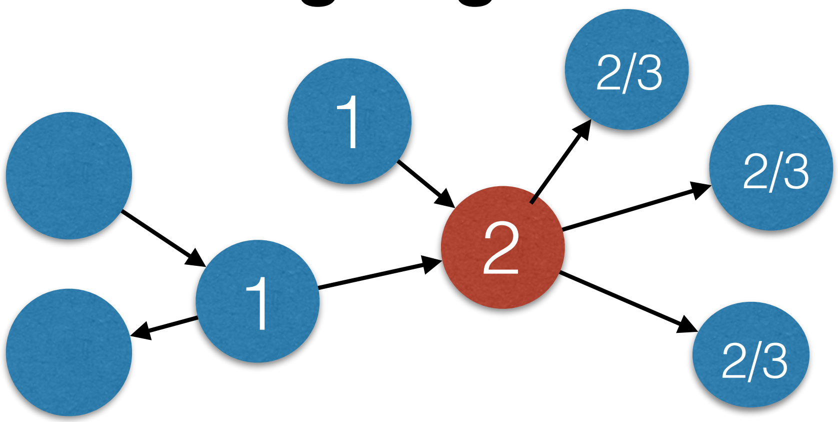
- Two modifications are made

Modification 1

- A page receives **less weight** from a page that has **a hight number of outgoing links**



- A page receives **more weight** from a page that has **a low number of outgoing links**



Modification 1

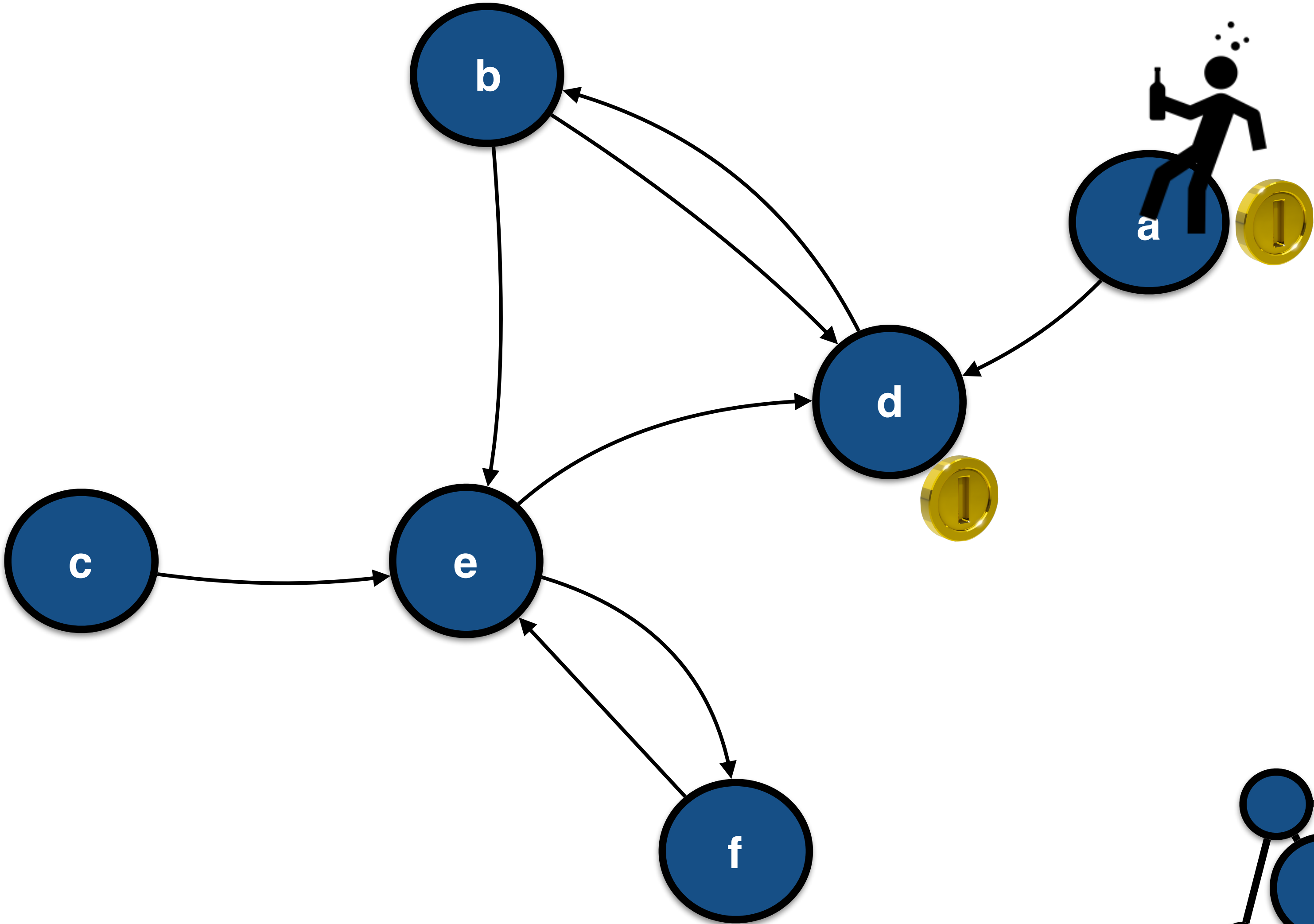
- Eigenvalue centrality

$$C_i = K \sum_{j=1}^{n} a_{i,j} C_j$$

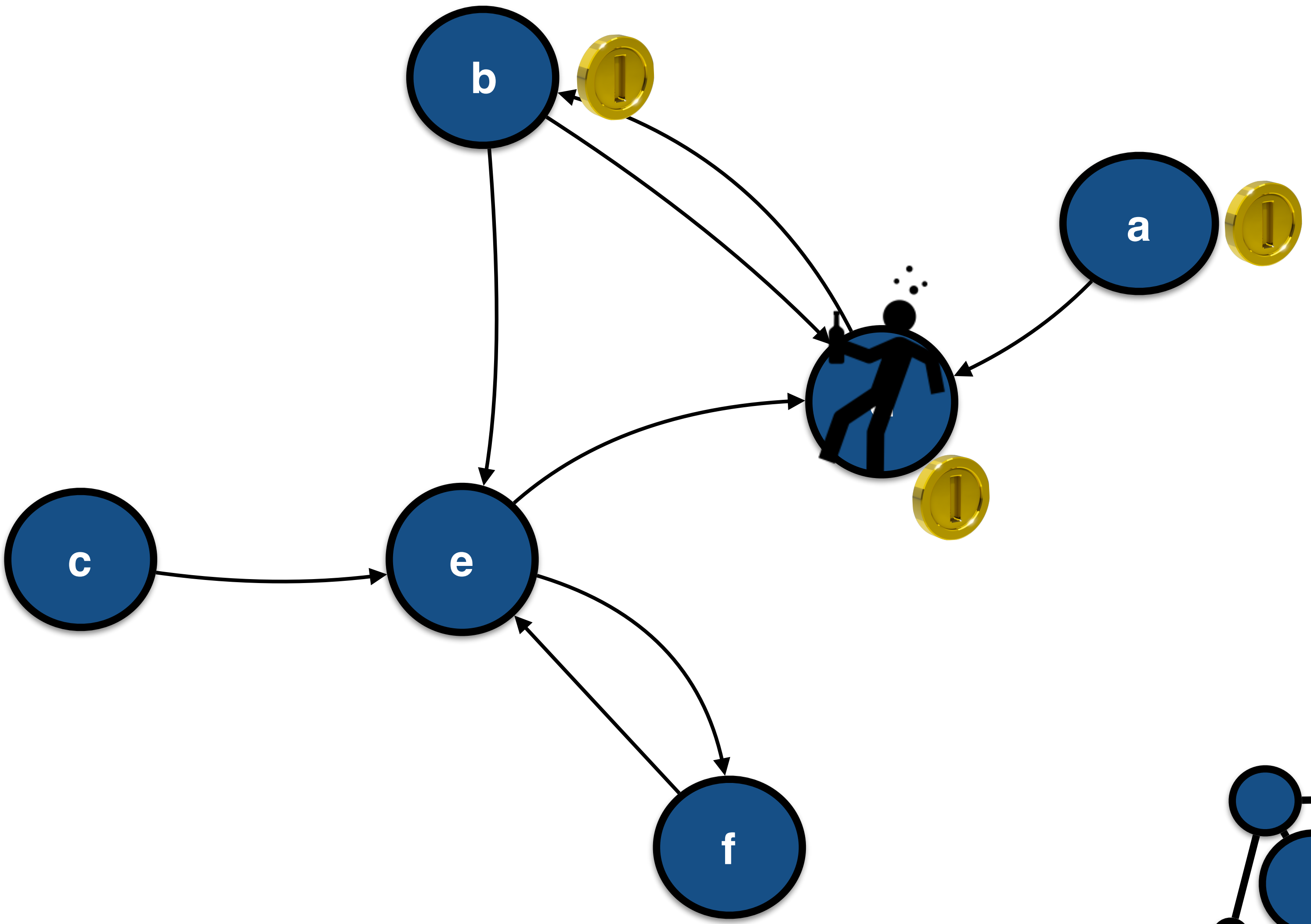
- PageRank : first modification

$$C_i = K \sum_{j=1..N} a_{j,i} \frac{C_j}{k_j^{out}}$$

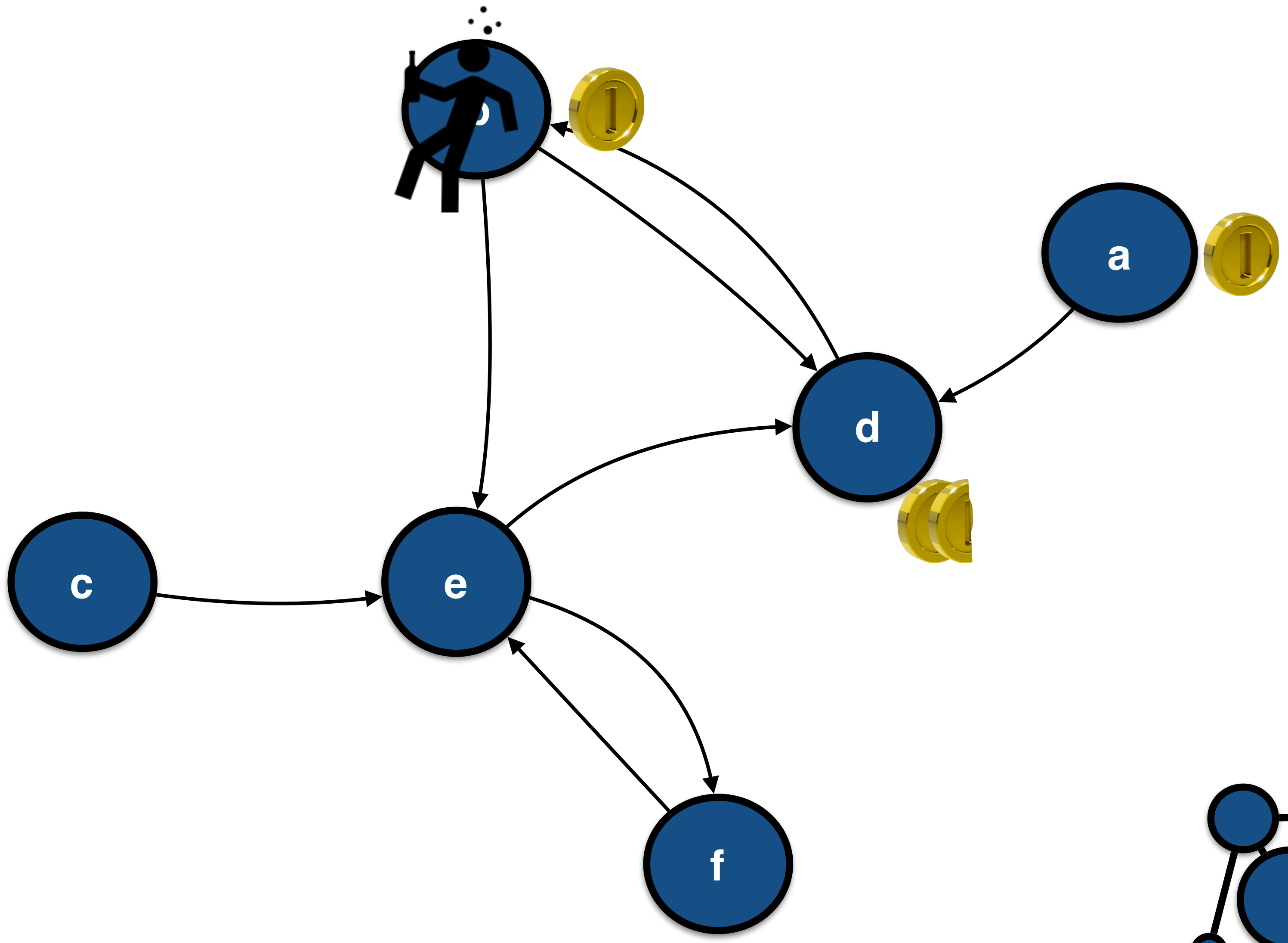
Random walk similarity



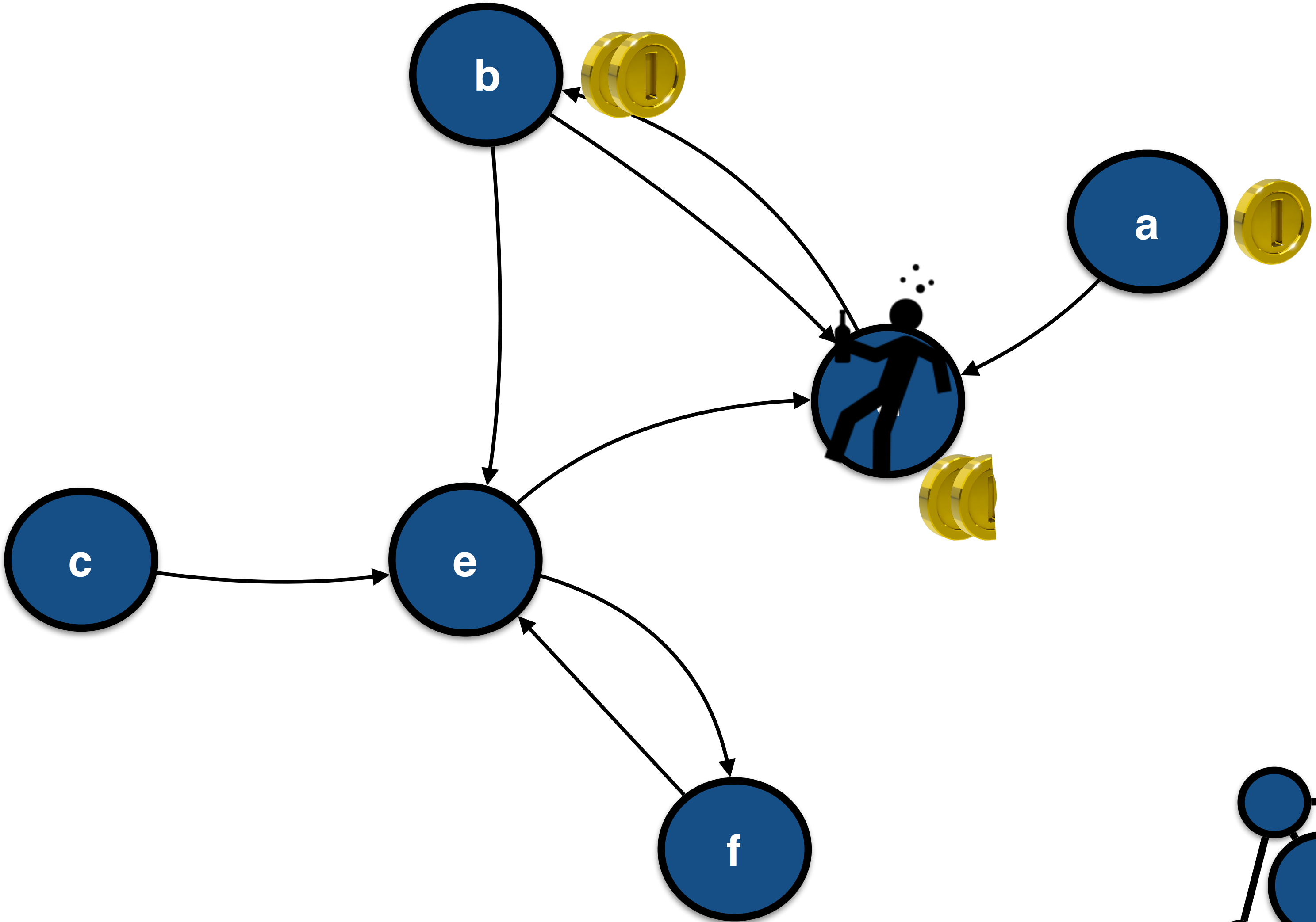
Random walk similarity



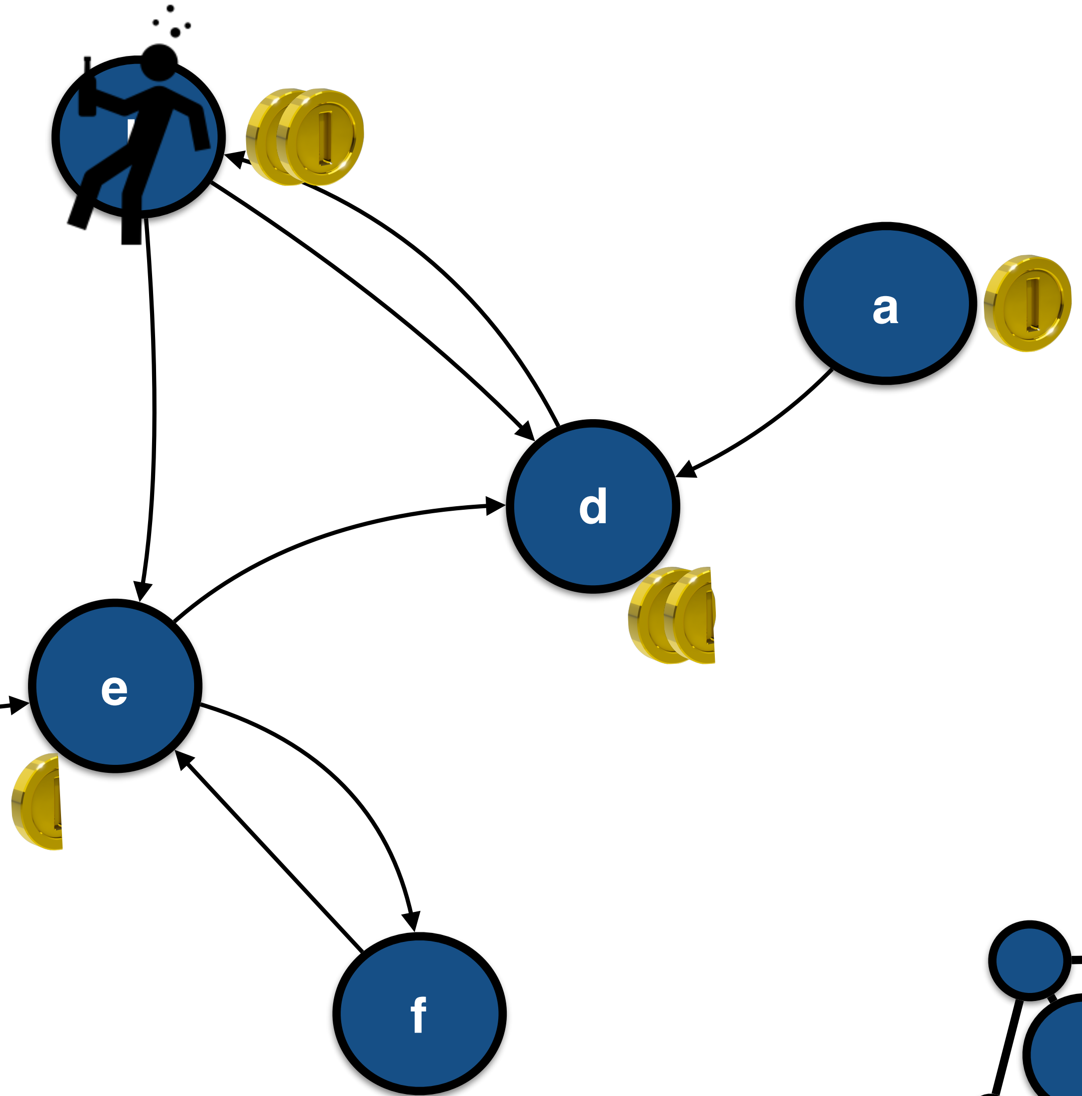
Random walk similarity



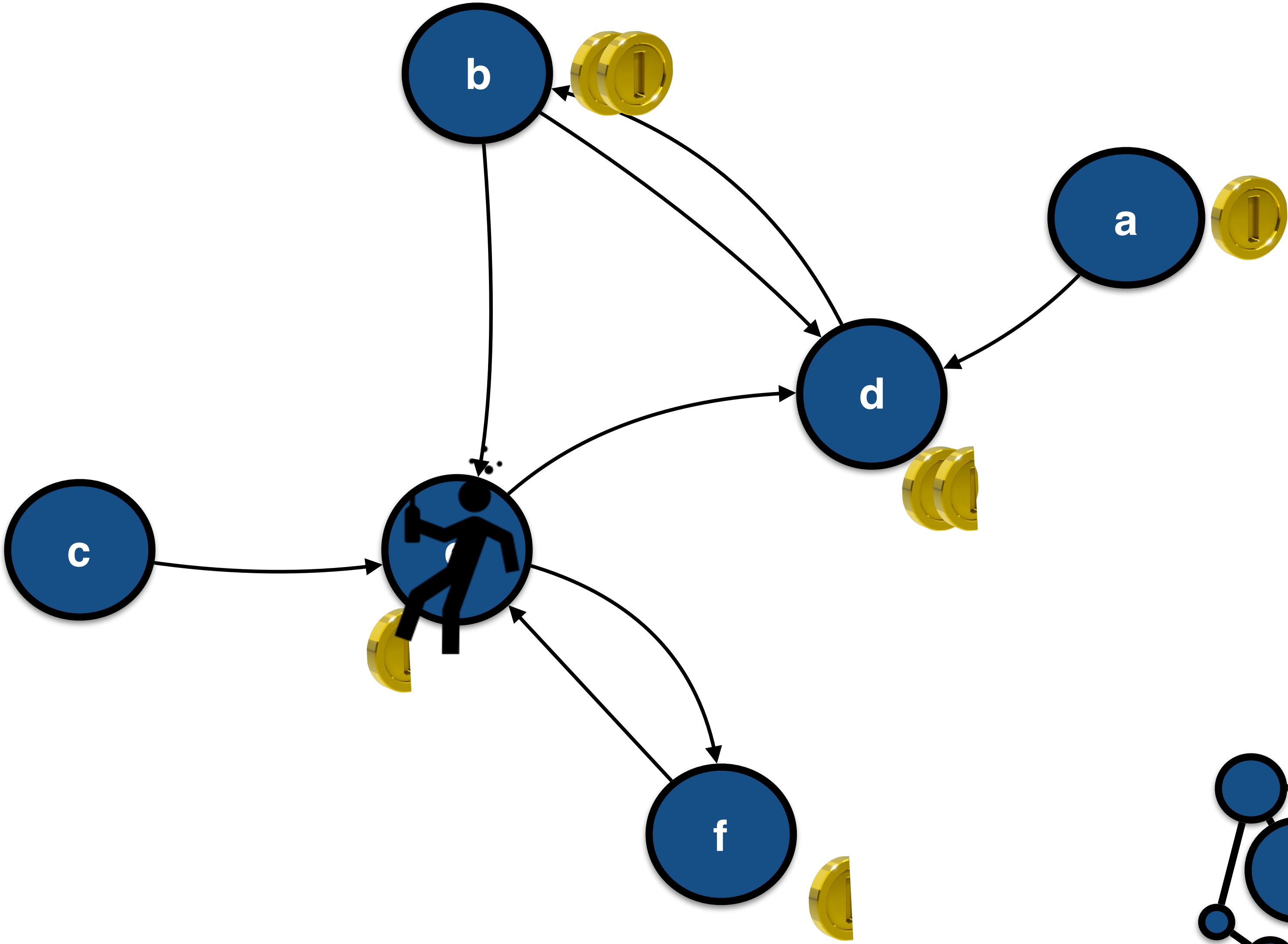
Random walk similarity



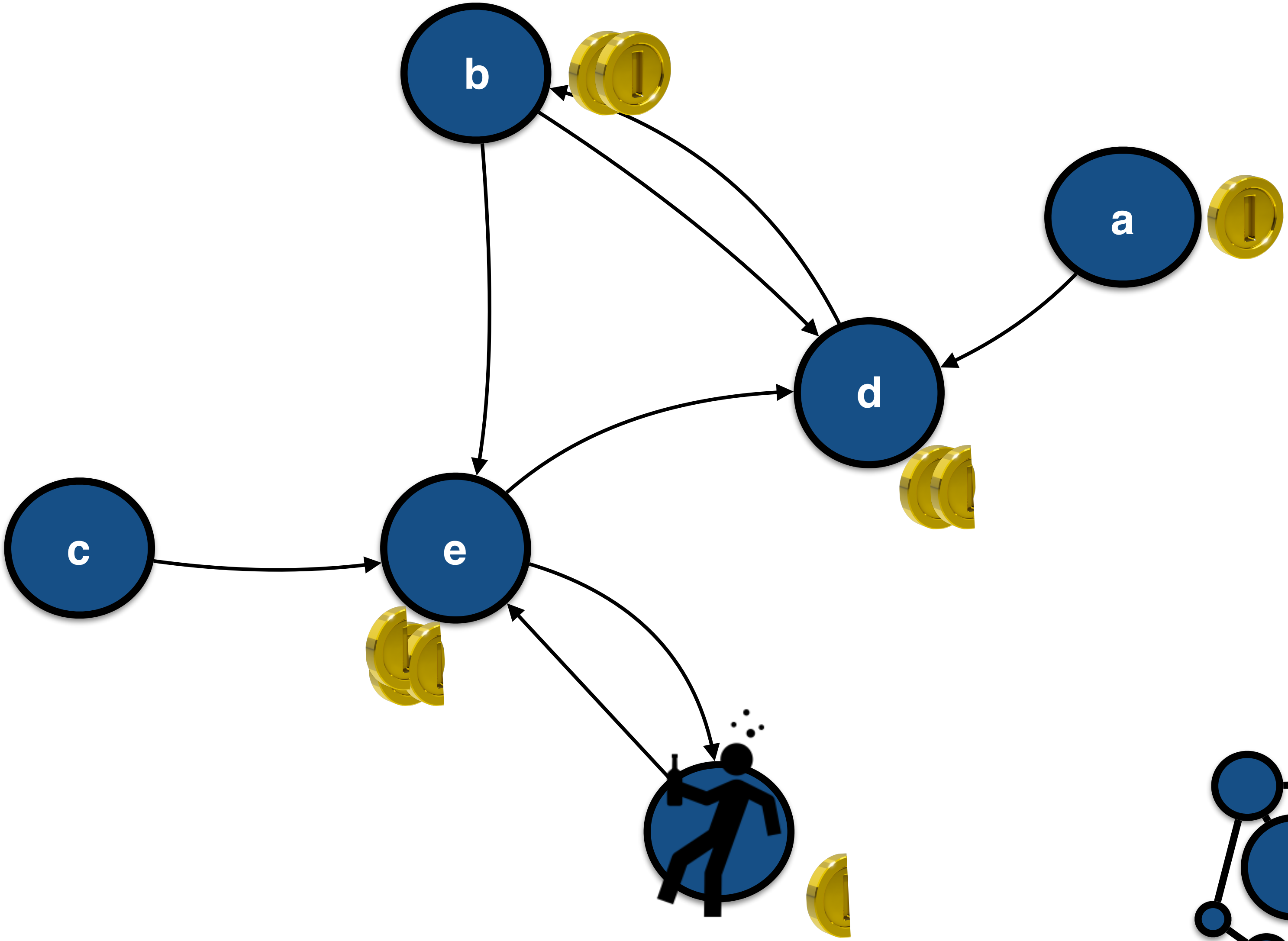
Random walk similarity



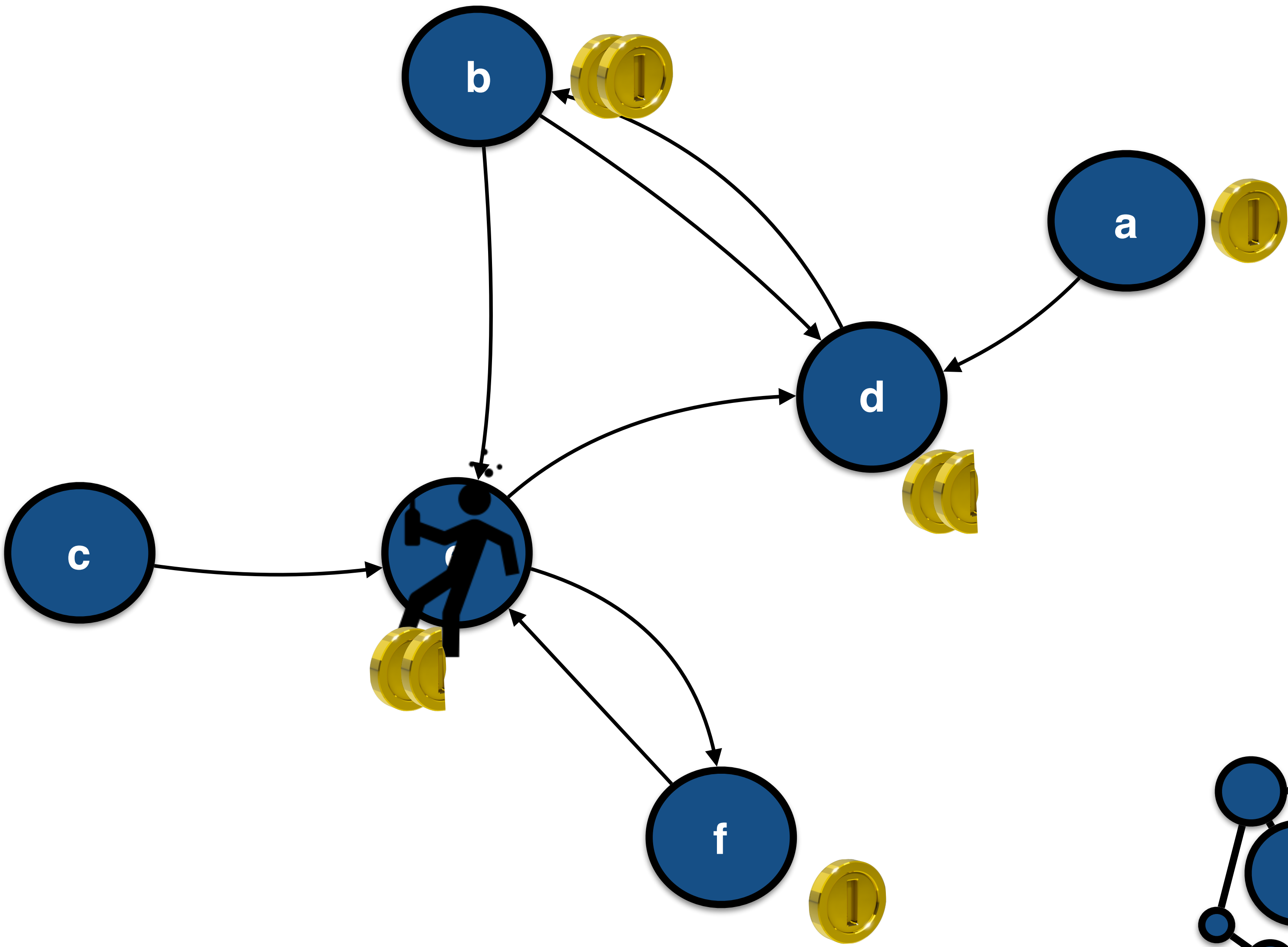
Random walk similarity



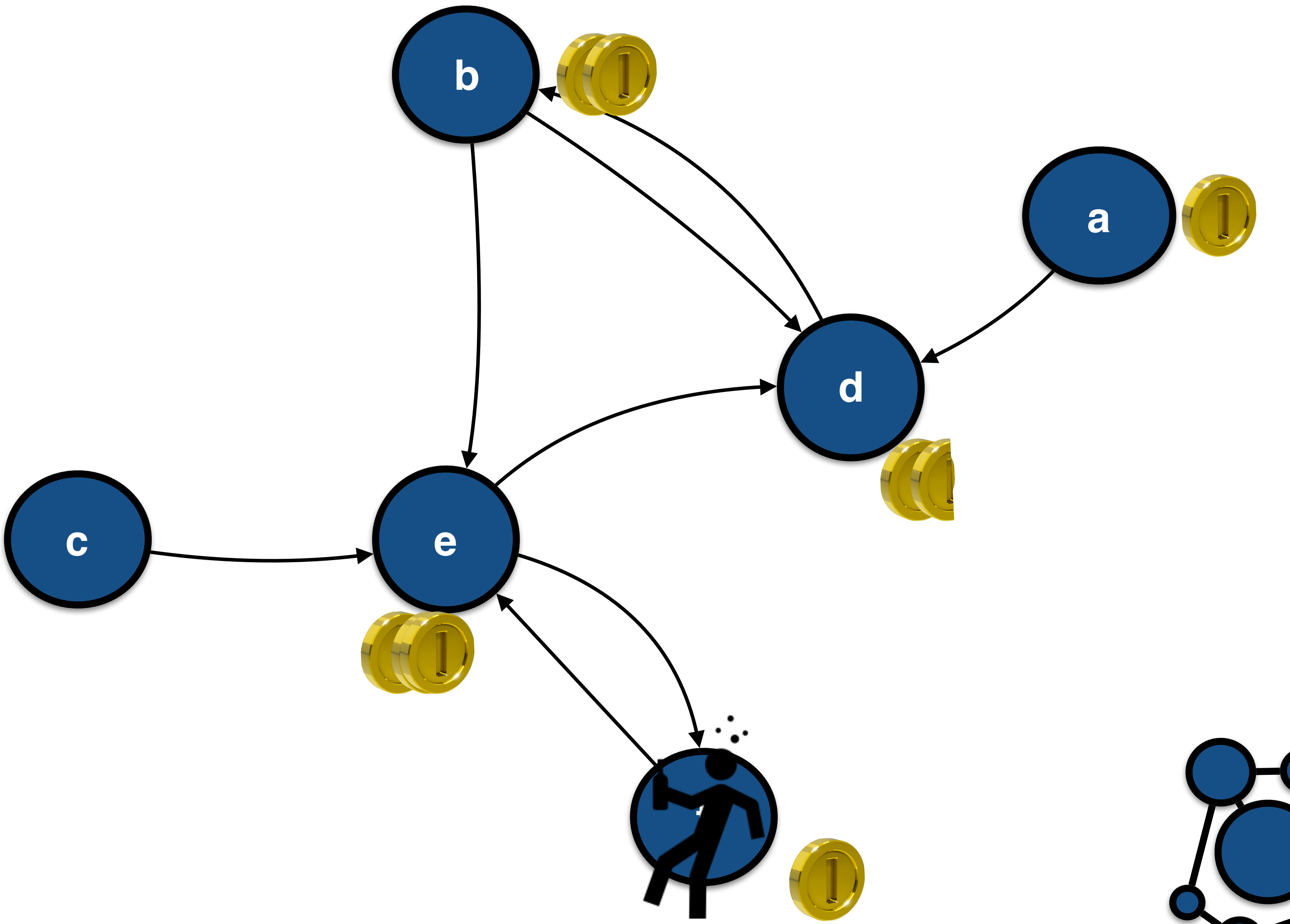
Random walk similarity



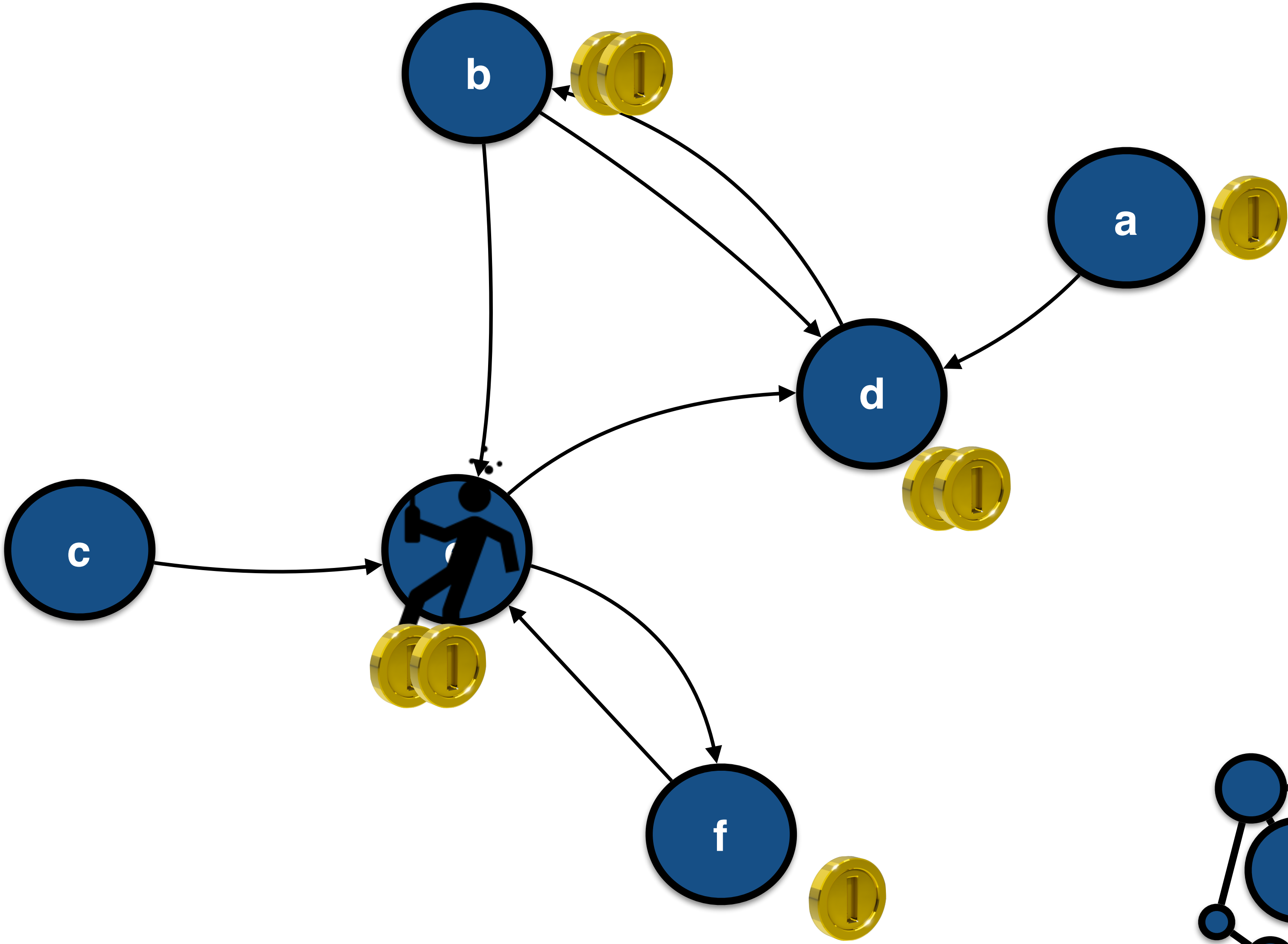
Random walk similarity



Random walk similarity

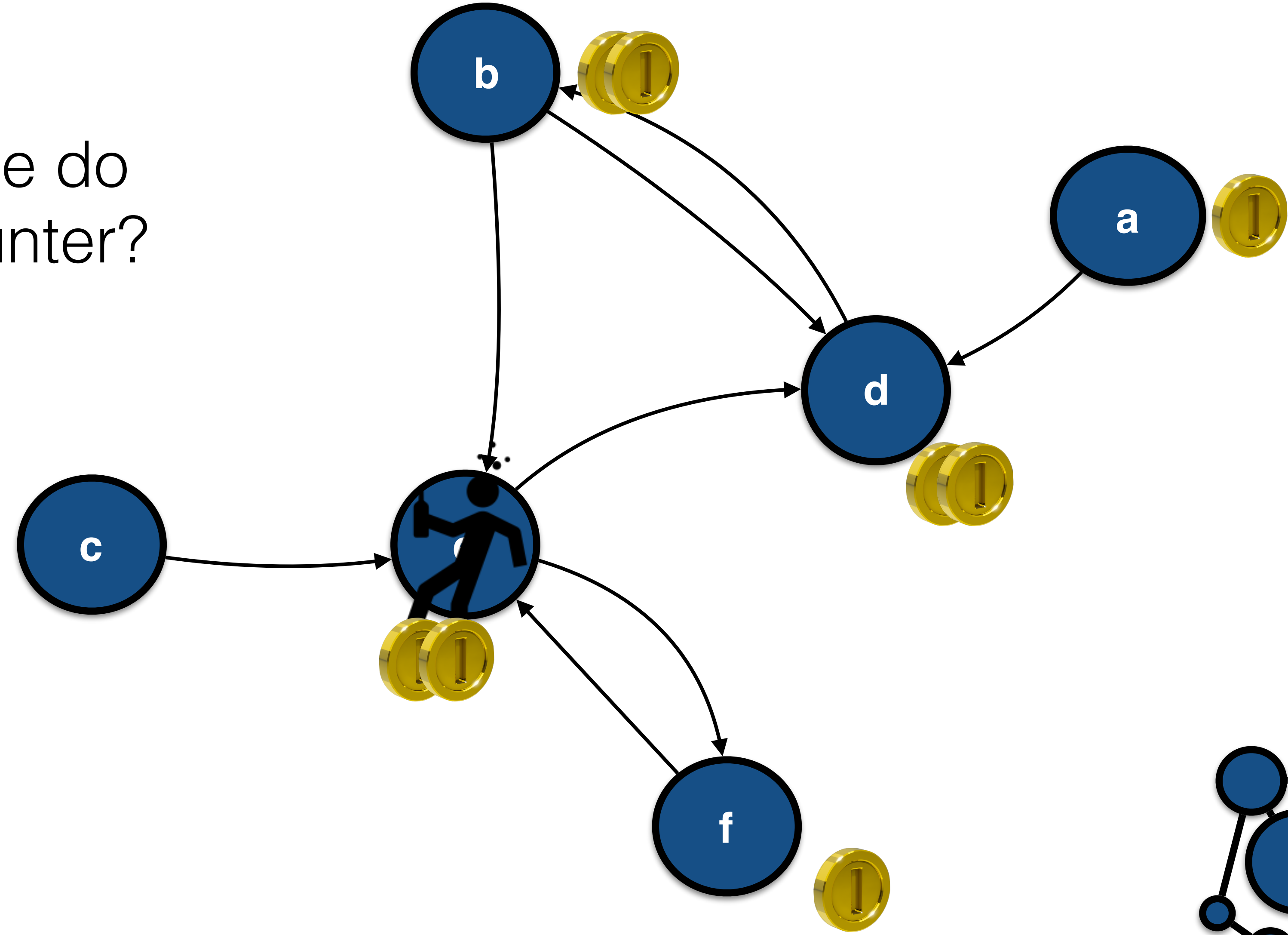


Random walk similarity



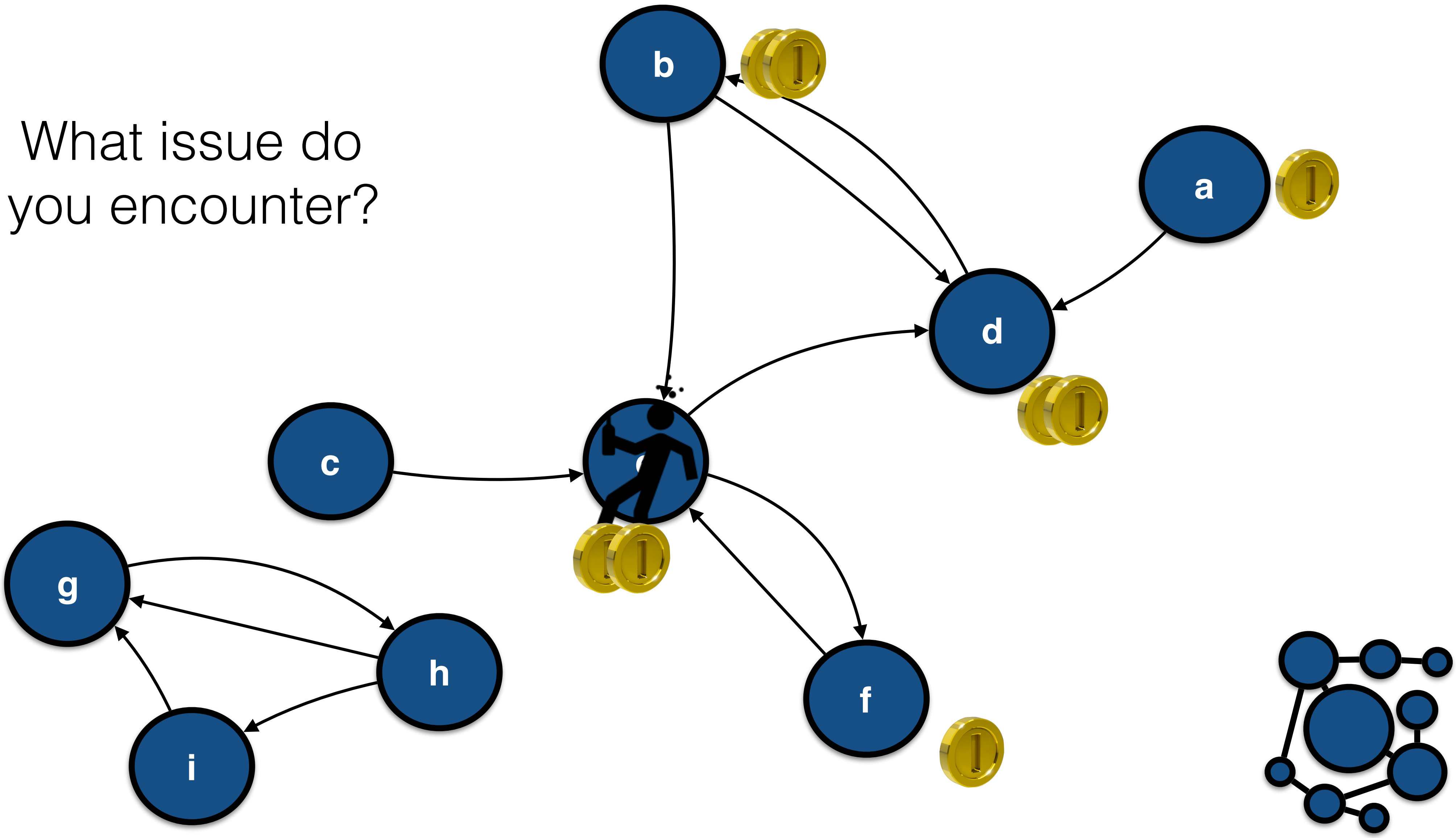
Random walk similarity

What issue do you encounter?



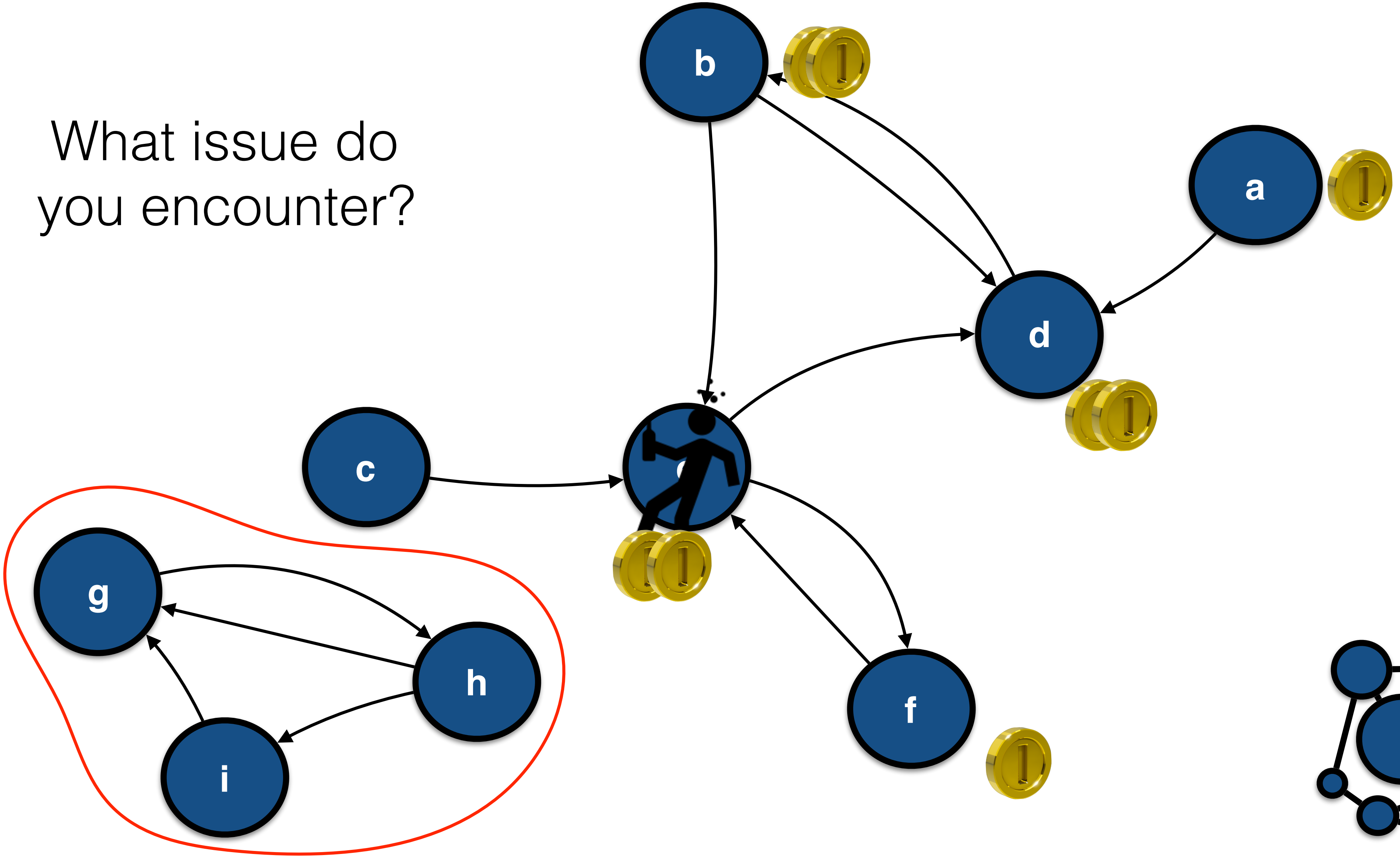
Random walk similarity

What issue do you encounter?



Random walk similarity

What issue do you encounter?



Modification 2

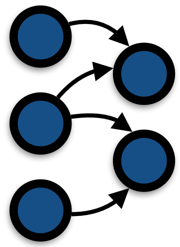
- It is now possible to move at any position of the graph with a certain probability

$$C_i = \frac{(1-d)}{N} + d \sum_{j=1..N} a_{j,i} \frac{C_j}{k_j^{out}}$$

- Usually d=0.85
- Teleportation probability equals 15%

The HITS algorithm

- Algorithme proposed by John Kleinberg
- Used for ranking webpages
- Issues :
  - Too many pages are results of a query
  - Identify only the most important pages



Starting point

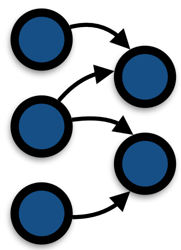
Starting point

Pages that contain the keywords searched

Starting point

Starting point

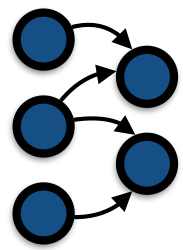
Enlarging to all pages citing or cited by these pages



Define the analysis graph

Subgraph(s,E,t,d)
s : query string
E : a text-based search engine.
t , d : natural numbers.

Let R denote the top t results of E on s.
  Set S := R
  For each page p of R
    Let $\Gamma^{+}(p)$ denote the set of all pages p points to.
    Let $\Gamma^{-}(p)$ denote the set of all pages pointing to p .
    Add all pages in $\Gamma^{+}(p)$ to S .
    If j $\Gamma^{-}(p)<d$ then
      Add all pages in $\Gamma^{-}(p)$ to S .
    Else
      Add an arbitrary set of d pages from $\Gamma^{-}(p)$ to S .



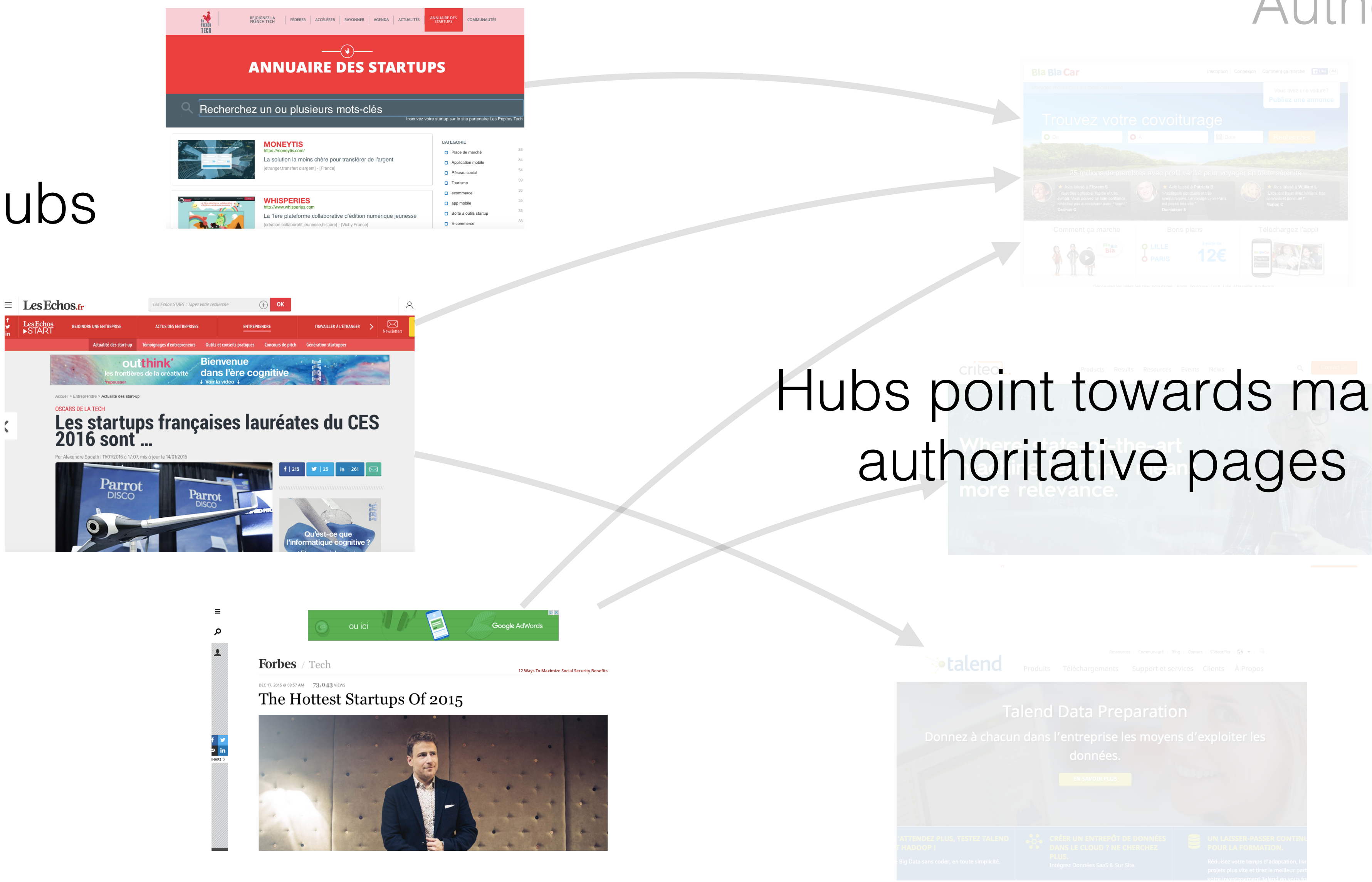
  End
Return S

Principle : Hubs et Authorities

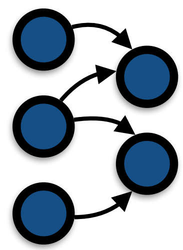
Principle : Hubs ?

Authorities

Hubs



Hubs point towards many authoritative pages

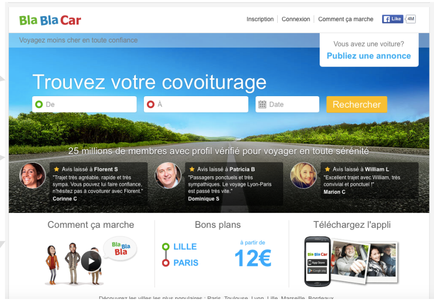


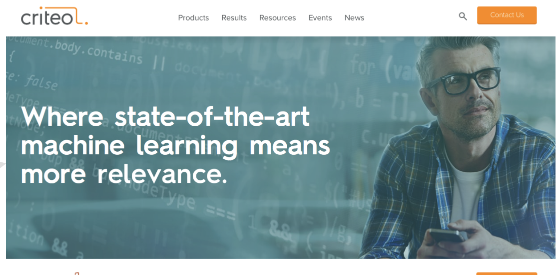
Principle : authorities?

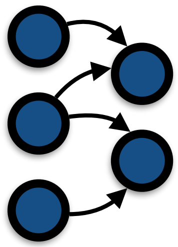
Authorities

Hubs

Authorities are pages that receives many incoming links from hubs



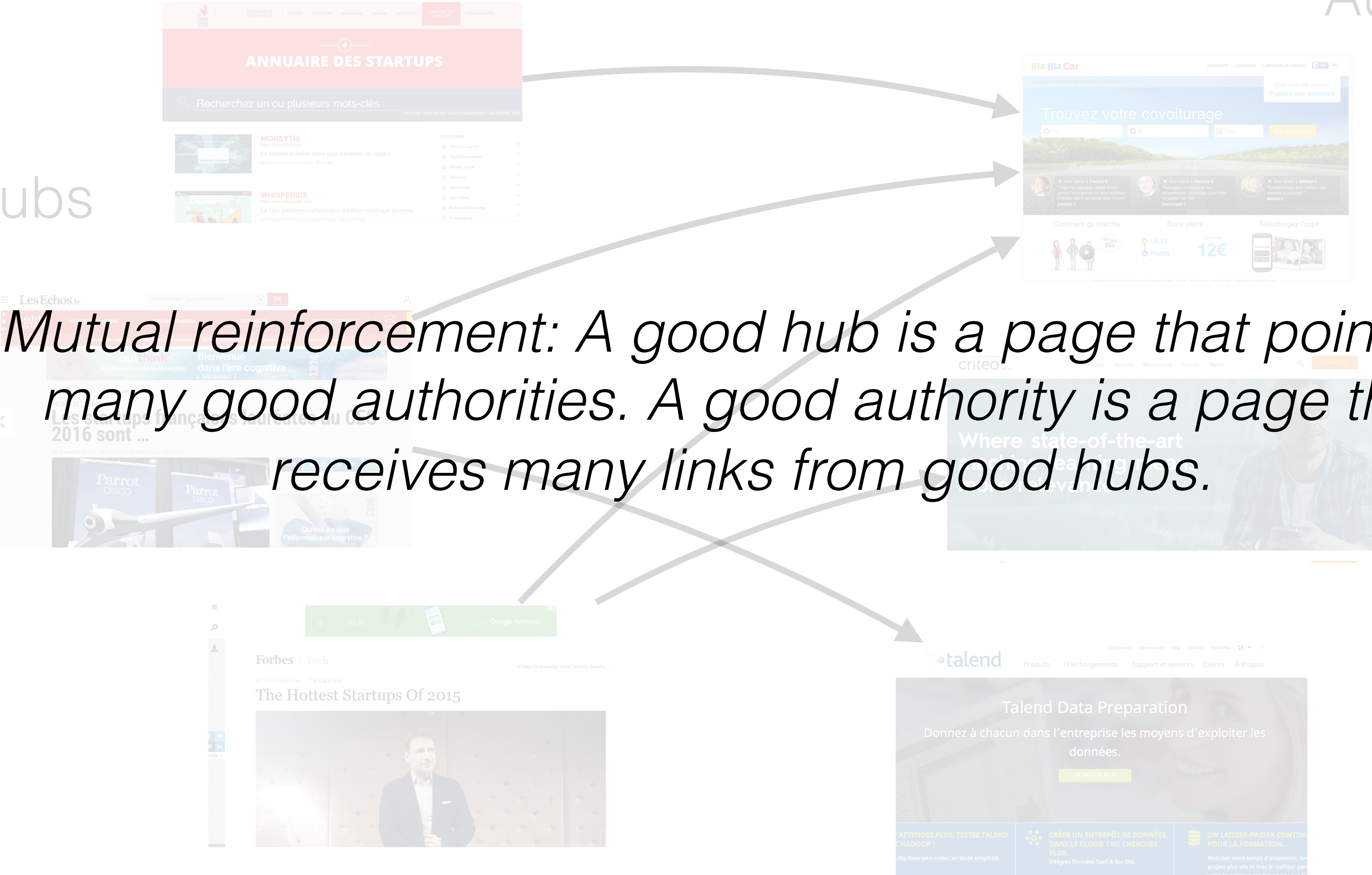




Algorithme paradigm

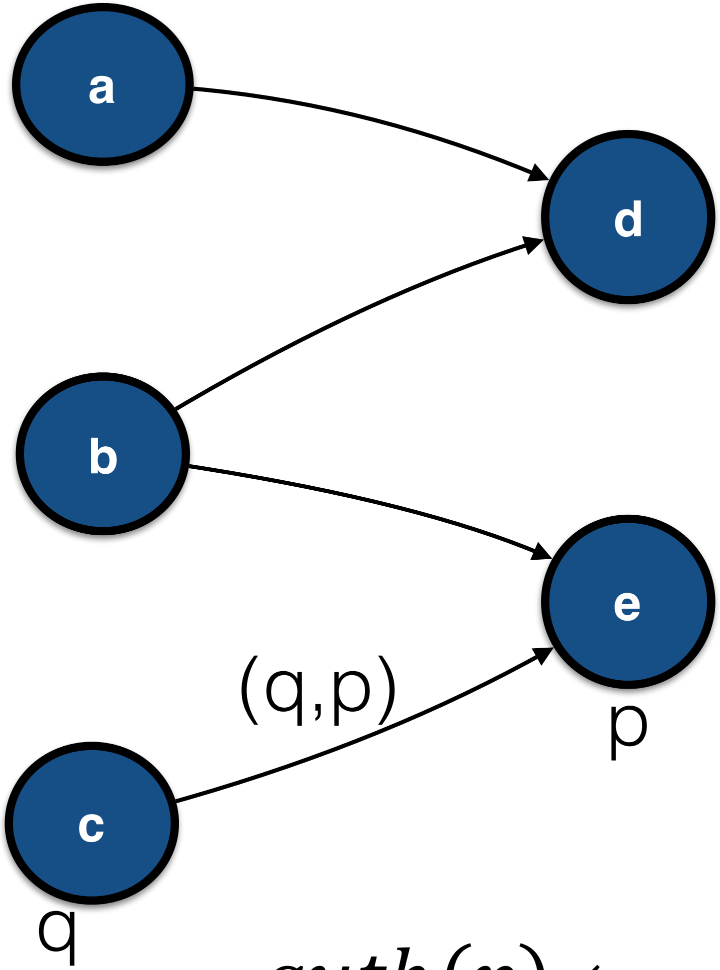
Autorités

Hubs



Mutual reinforcement: A good hub is a page that points to many good authorities. A good authority is a page that receives many links from good hubs.

Algorithm : authority score



auth(d)=hub(a)+hub(b)

auth(e)=hub(b)+hub(c)

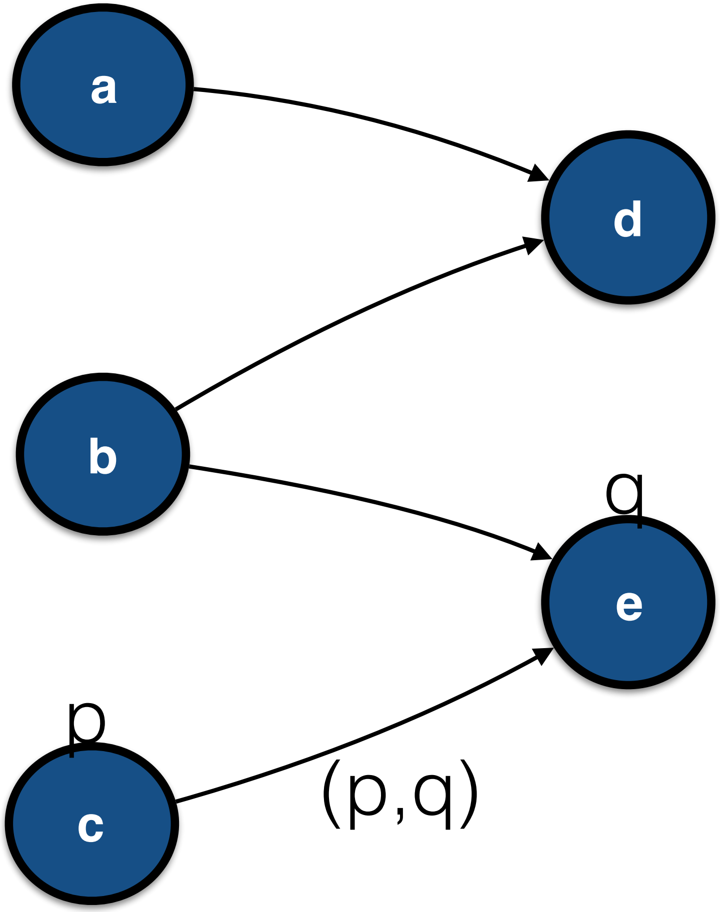
$$auth(p) \leftarrow \sum_{q:\{q,p\}\in E} hub(q)$$

Algorithm : hub score

hub(a)=auth(d)

hub(b)=auth(d)+auth(e)

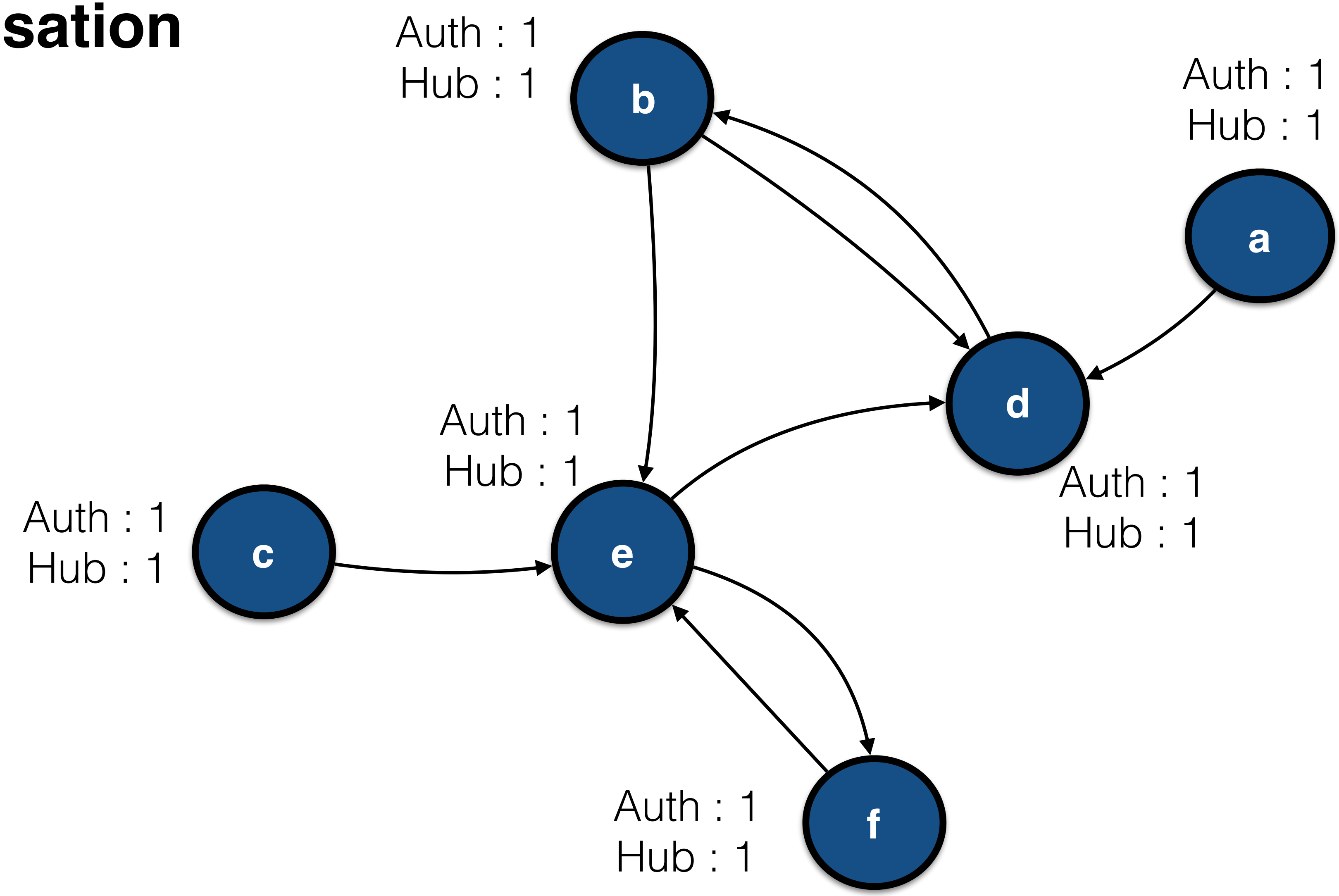
hub(c)=auth(e)



$$hub(p) \leftarrow \sum_{q:\{p,q\}\in E} auth(q)$$

Example

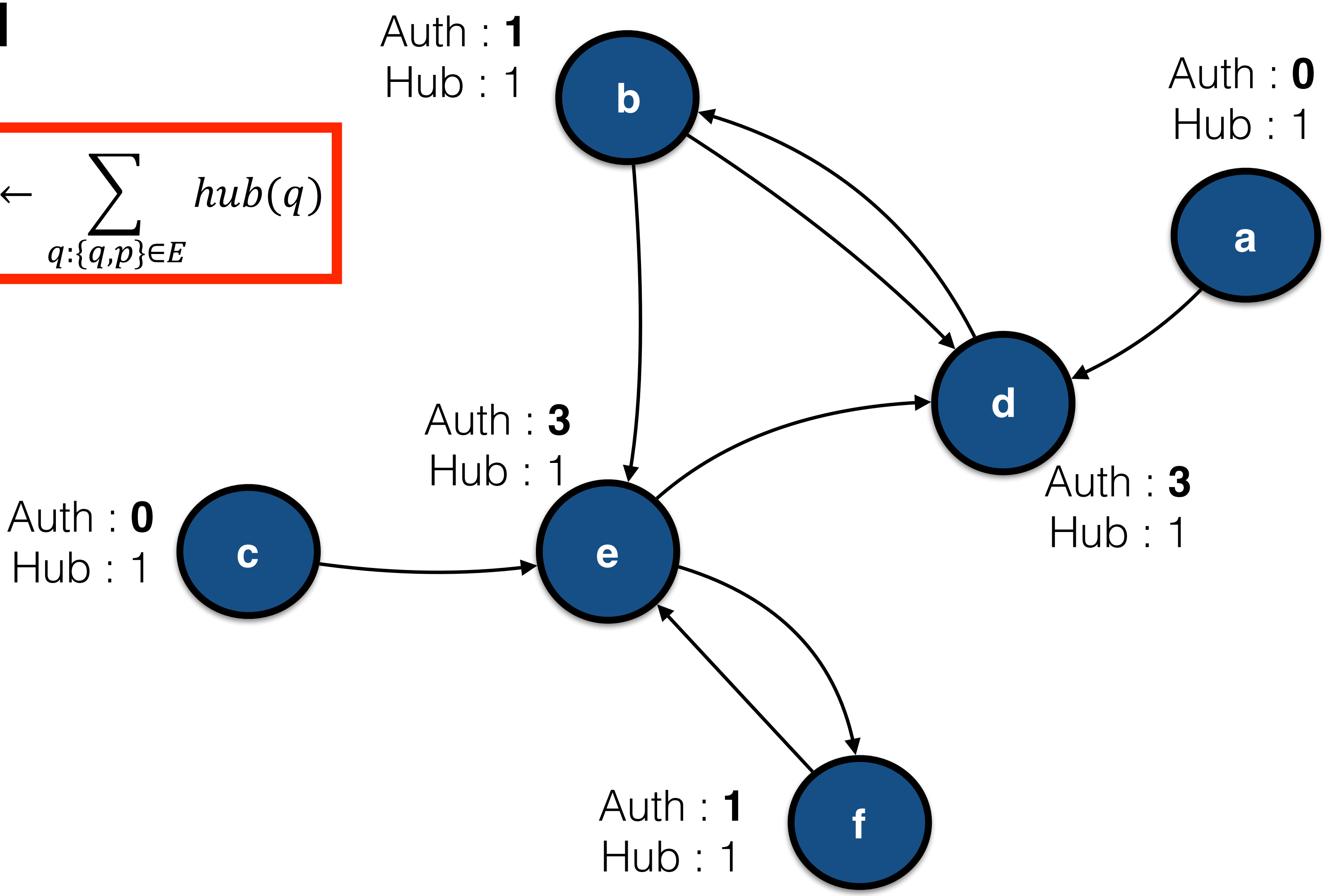
Initialisation



Example

auth_1

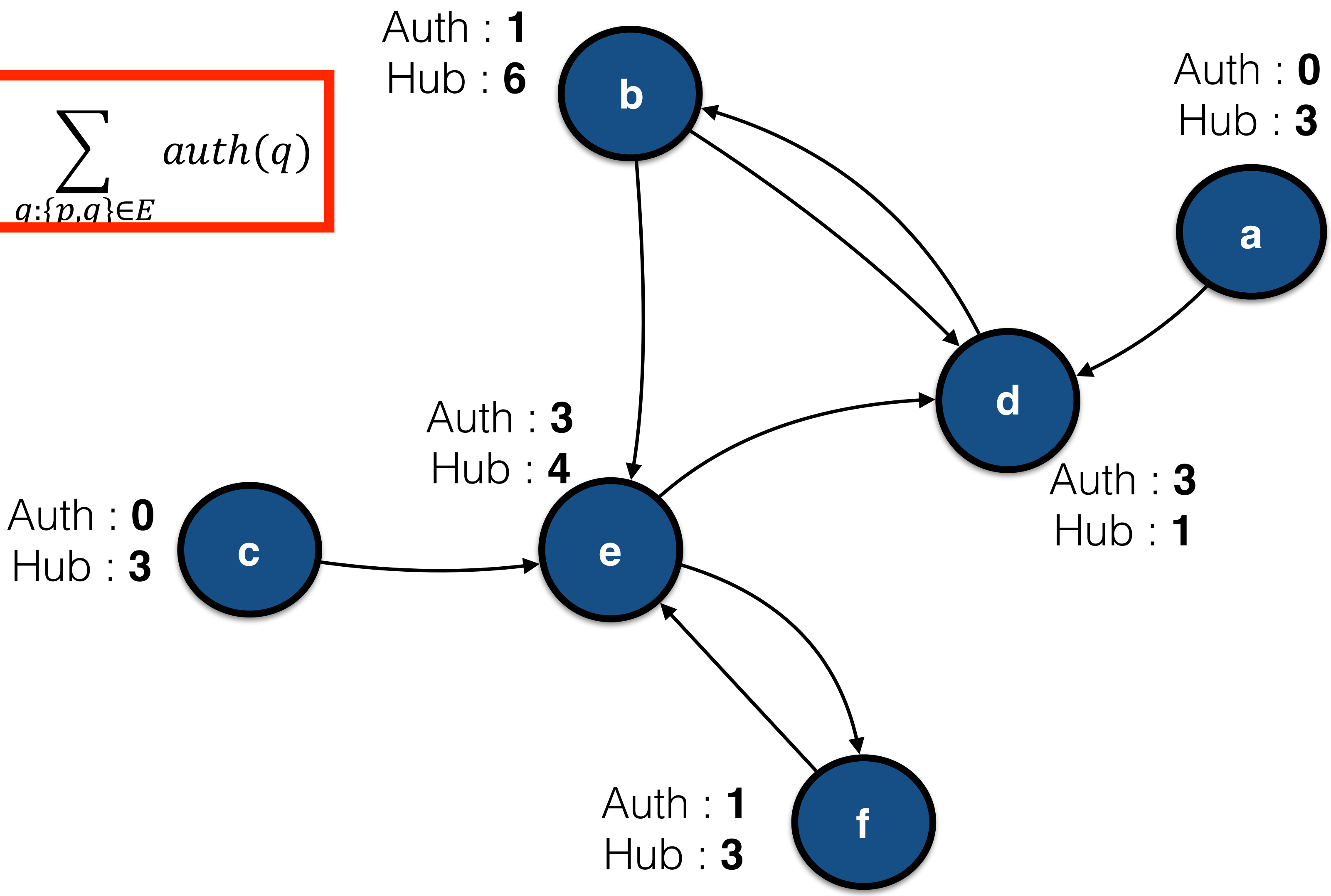
$$\boxed{auth(p) \leftarrow \sum_{q:\{q,p\}\in E} hub(q)}$$



Example

hub_1

$$\boxed{hub(p) \leftarrow \sum_{q:\{p,q\}\in E} auth(q)}$$

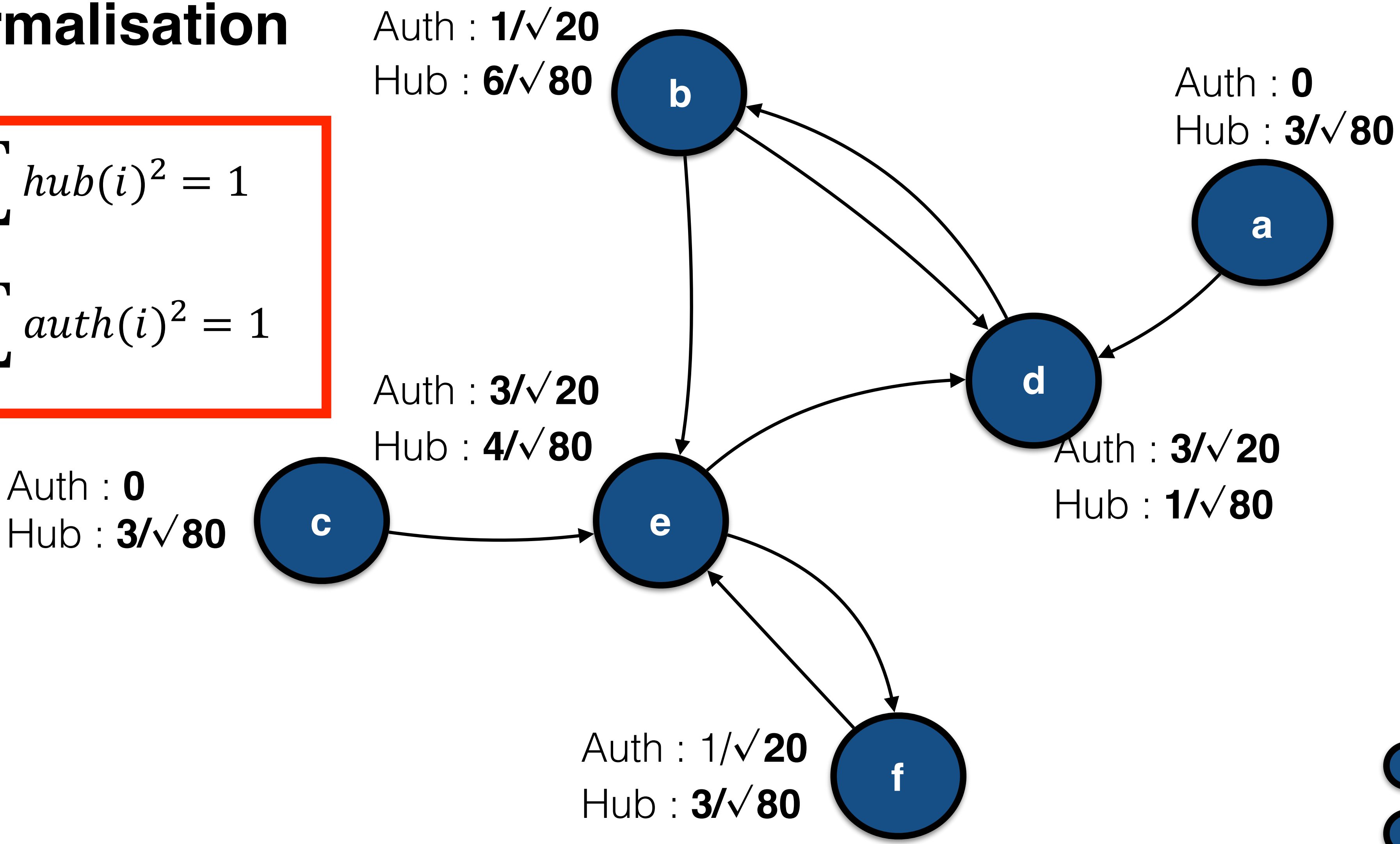


Example

Normalisation

$$\sum_i hub(i)^2 = 1$$

$$\sum_i auth(i)^2 = 1$$



Normalisation

$$\sum_i x'^2_i = 1$$

$$x'_i = \alpha x_i$$

Required for convergence

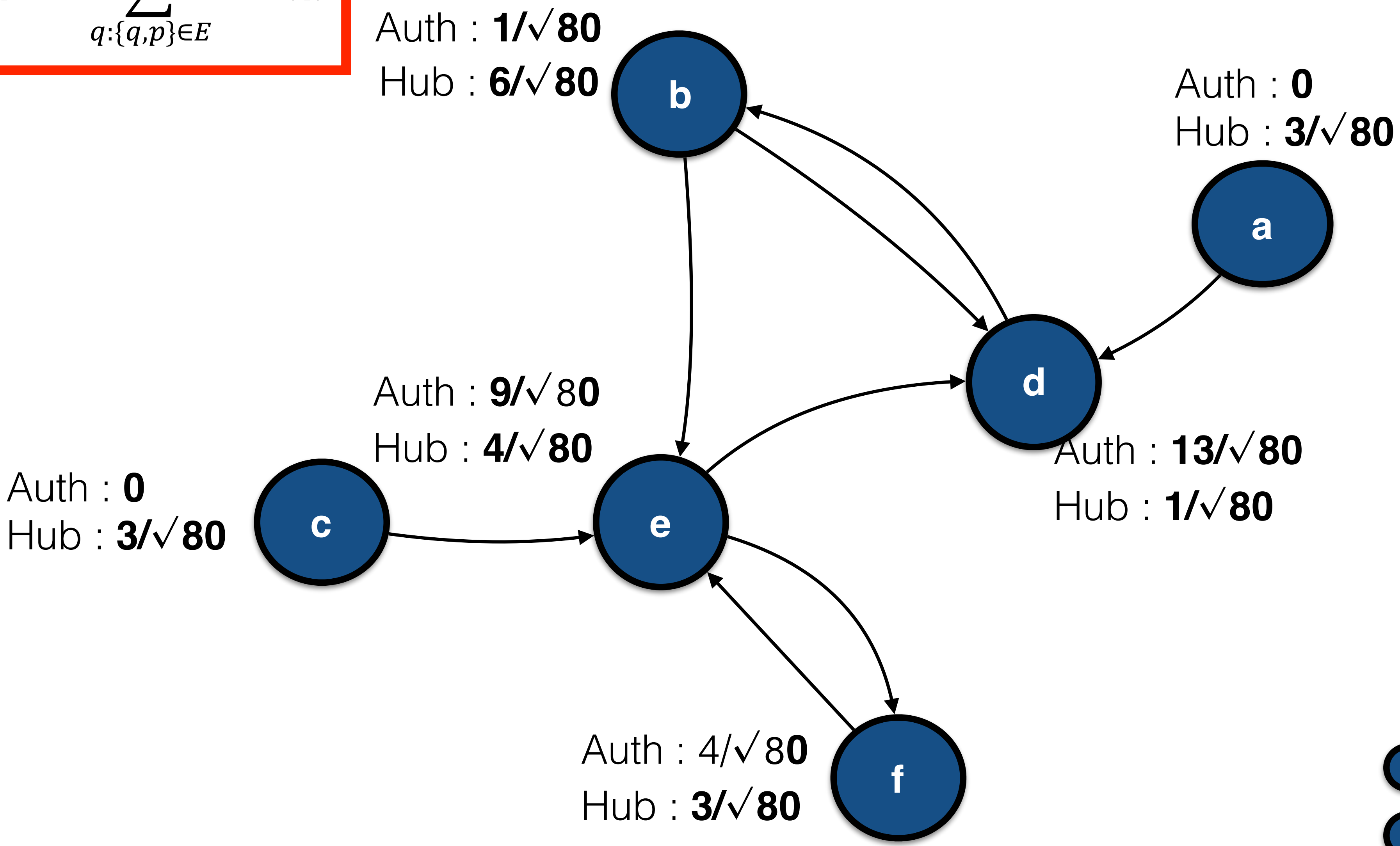
$$\sum_i x'^2_i = \sum_i (\alpha x_i)^2 \qquad \sum_i x'^2_i = \alpha^2 \sum_i {x_i}^2$$

$$\alpha^2 \sum_i {x_i}^2 = 1 \qquad \alpha = \frac{1}{\sqrt{\sum_i {x_i}^2}}$$

$$x'_i = \frac{x_i}{\sqrt{\sum_i {x_i}^2}}$$

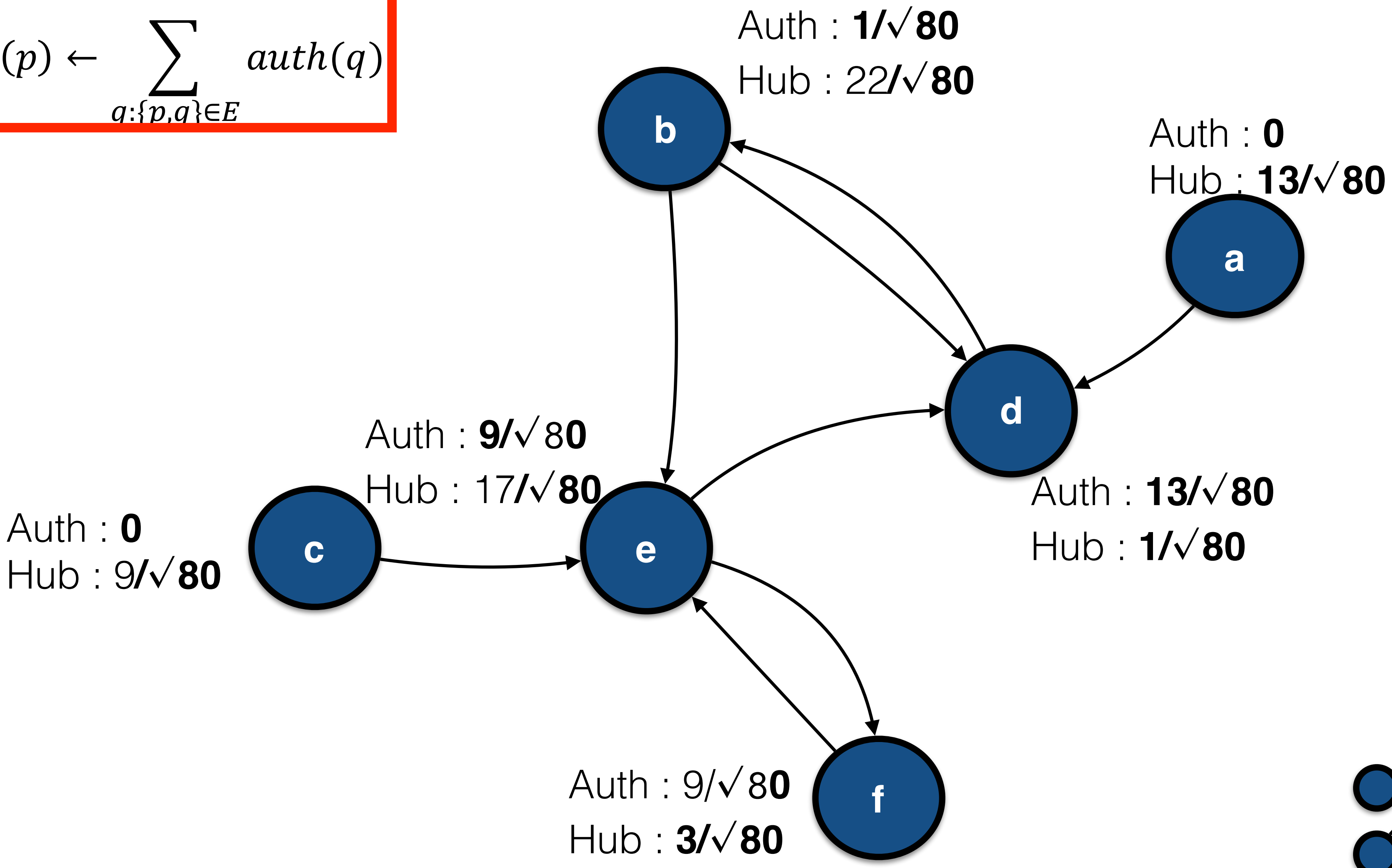
Example

$$auth(p) \leftarrow \sum_{q:\{q,p\}\in E} hub(q)$$



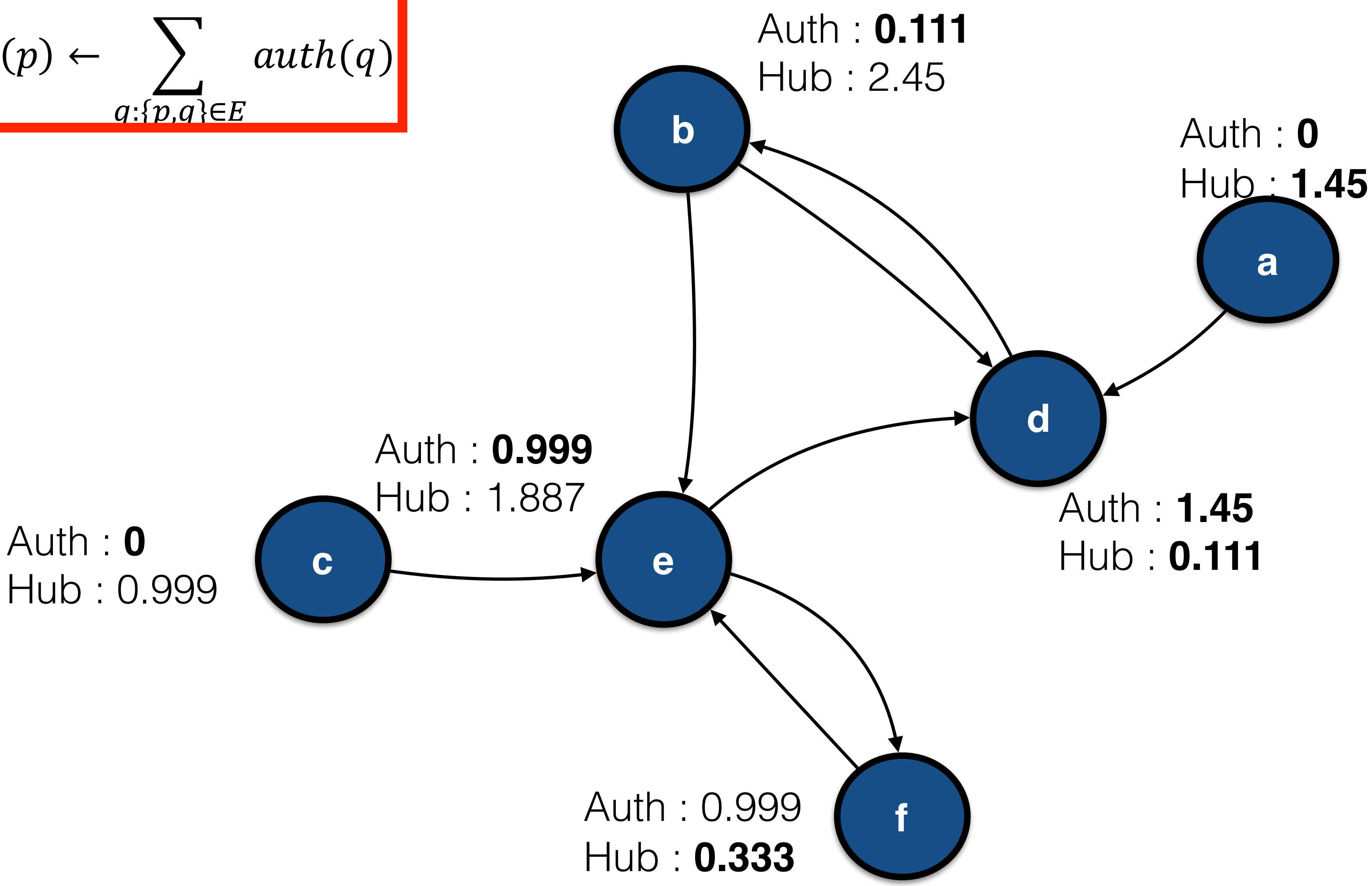
Example

$$\boxed{hub(p) \leftarrow \sum_{q:\{p,q\}\in E} auth(q)}$$



Example

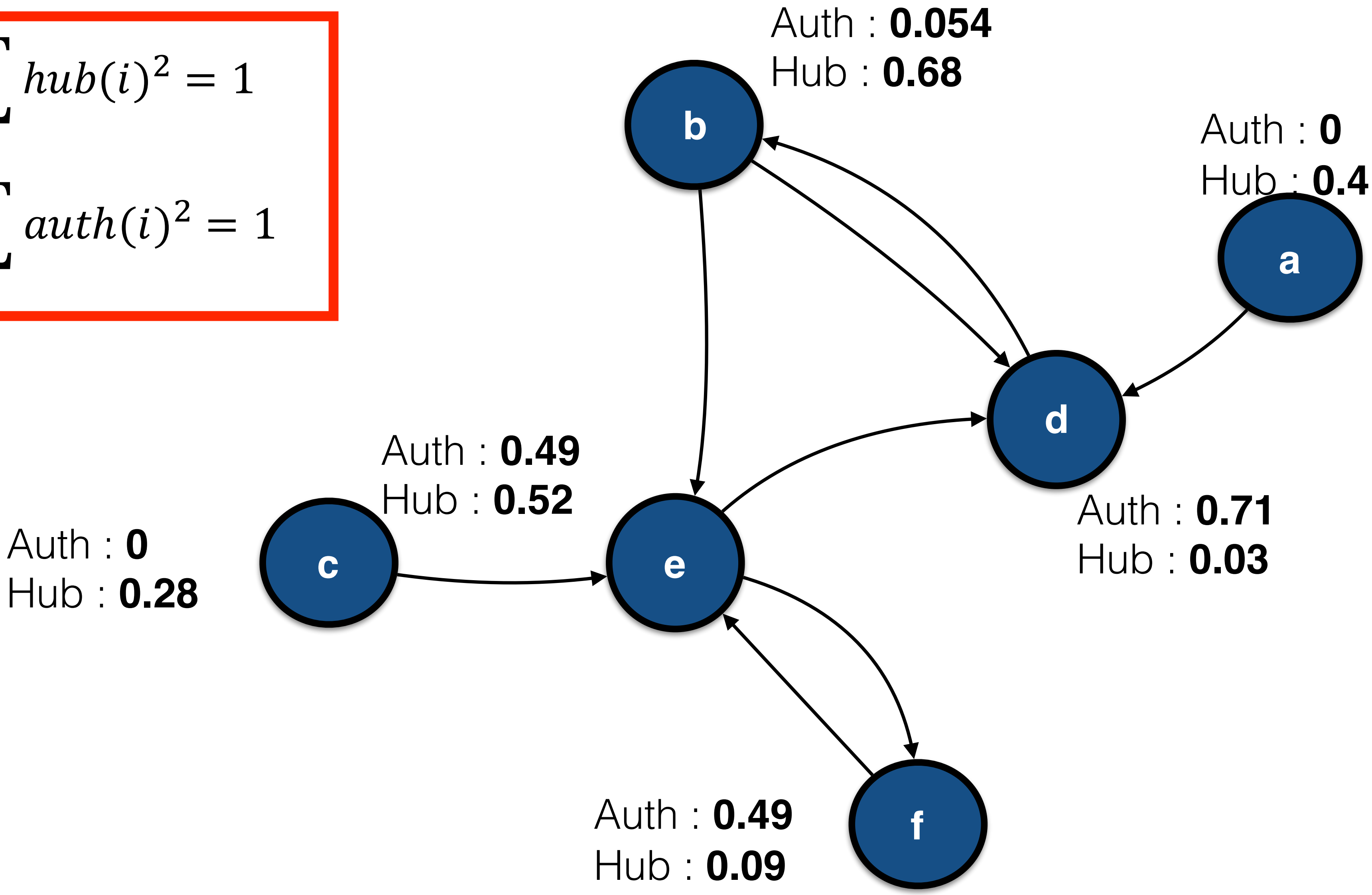
$$\boxed{hub(p) \leftarrow \sum_{q:\{p,q\}\in E} auth(q)}$$



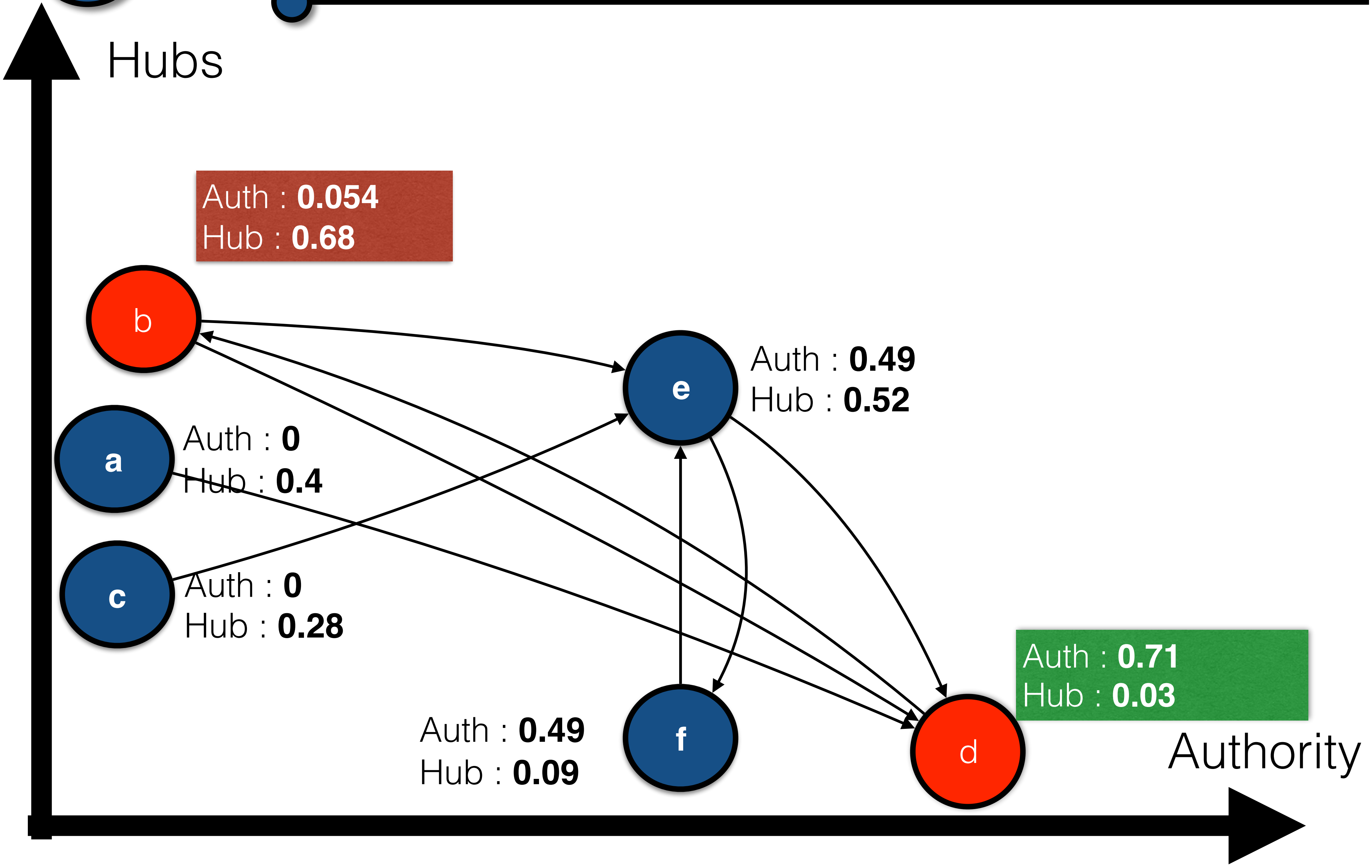
Example

$$\sum_i hub(i)^2 = 1$$

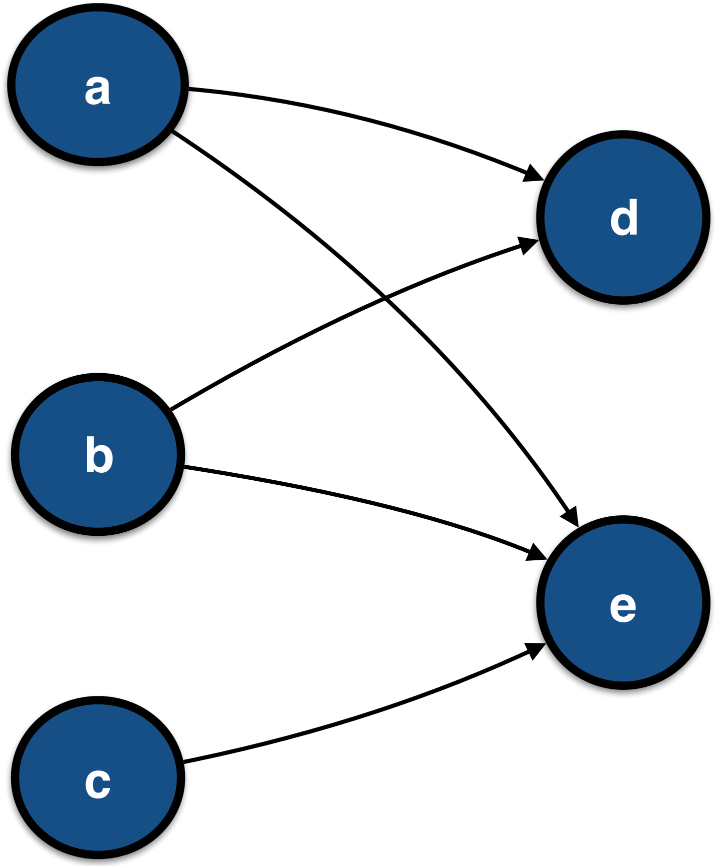
$$\sum_i auth(i)^2 = 1$$



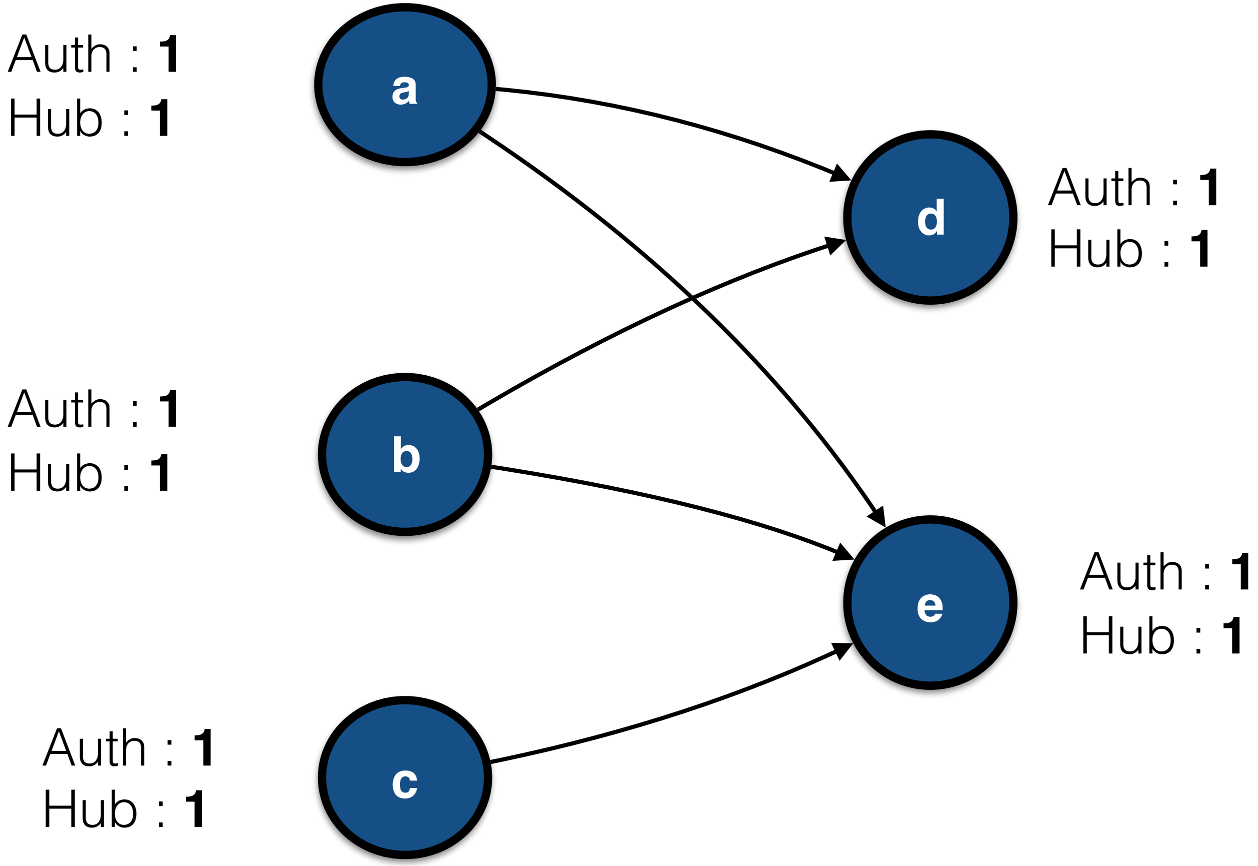
After two rounds



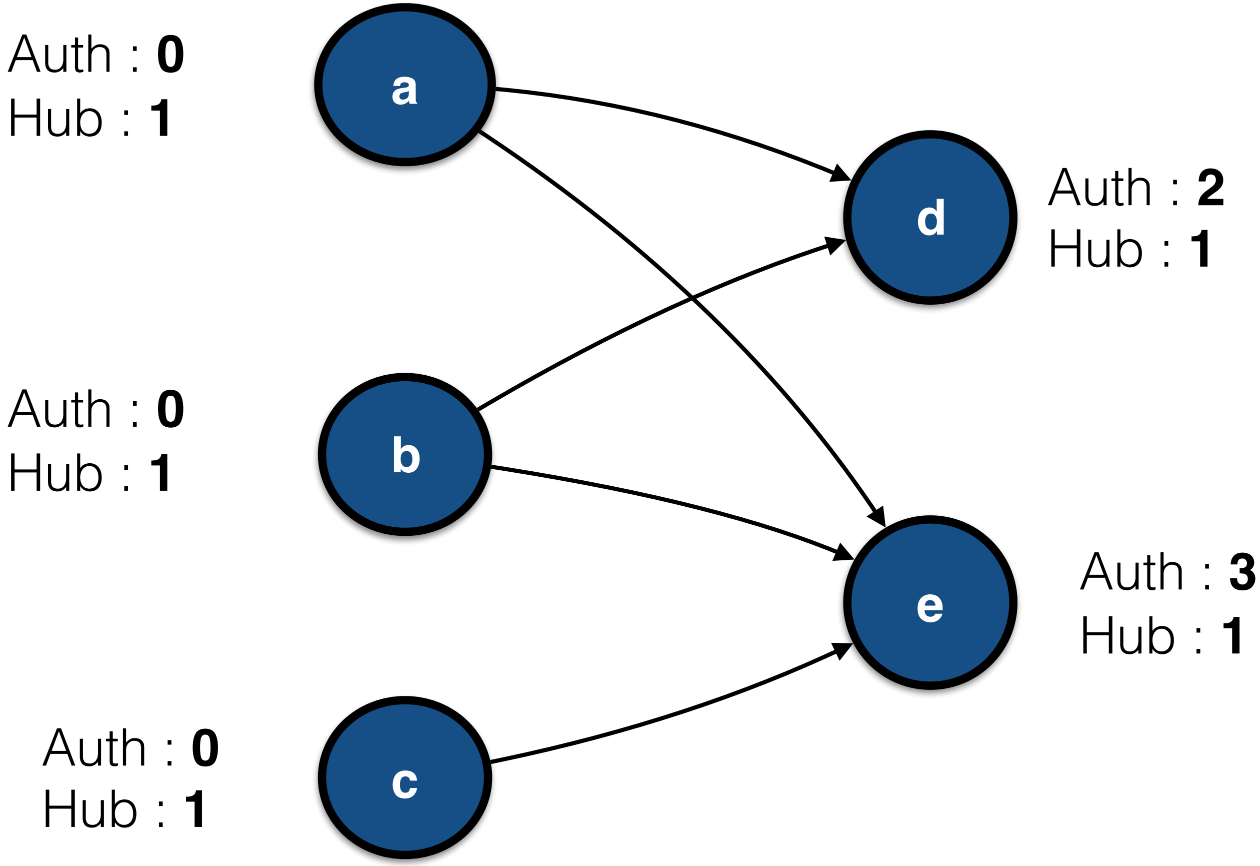
Second example



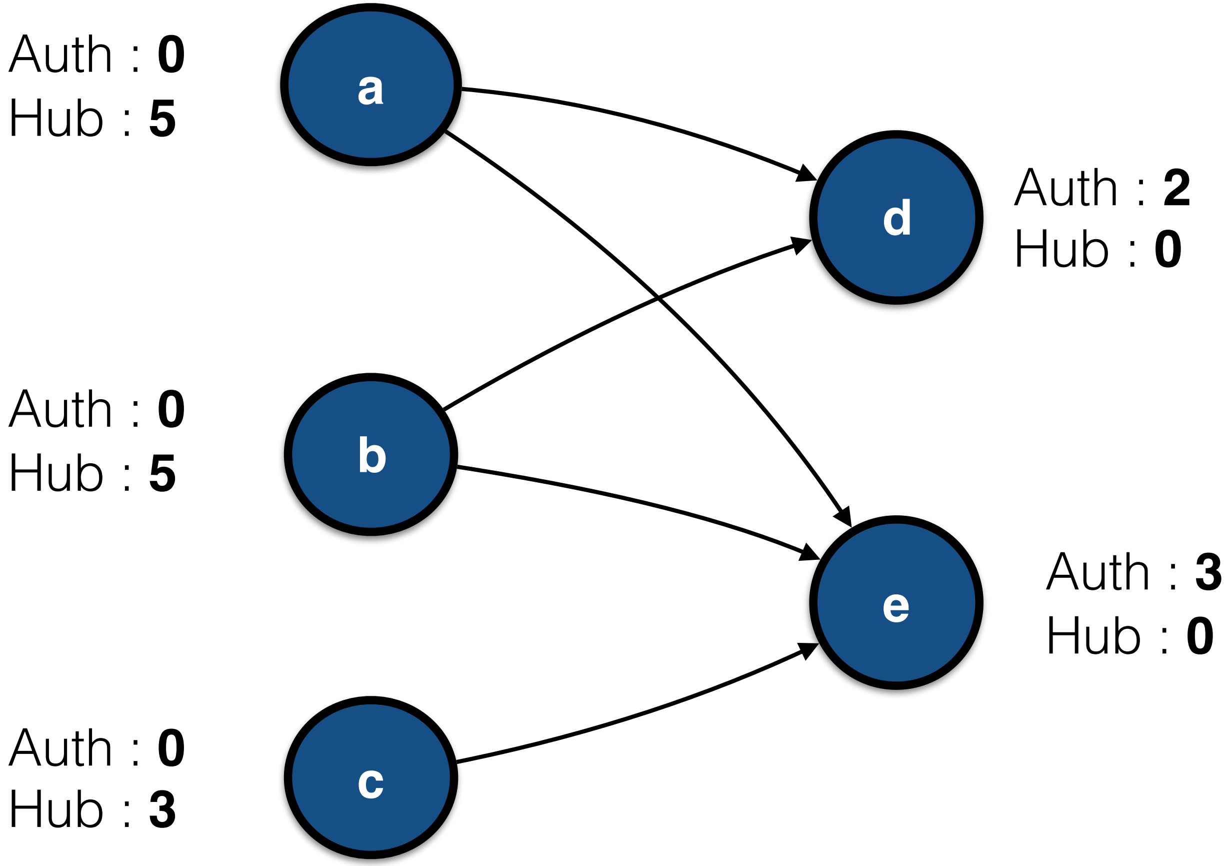
HITS Algorithm - 2 rounds



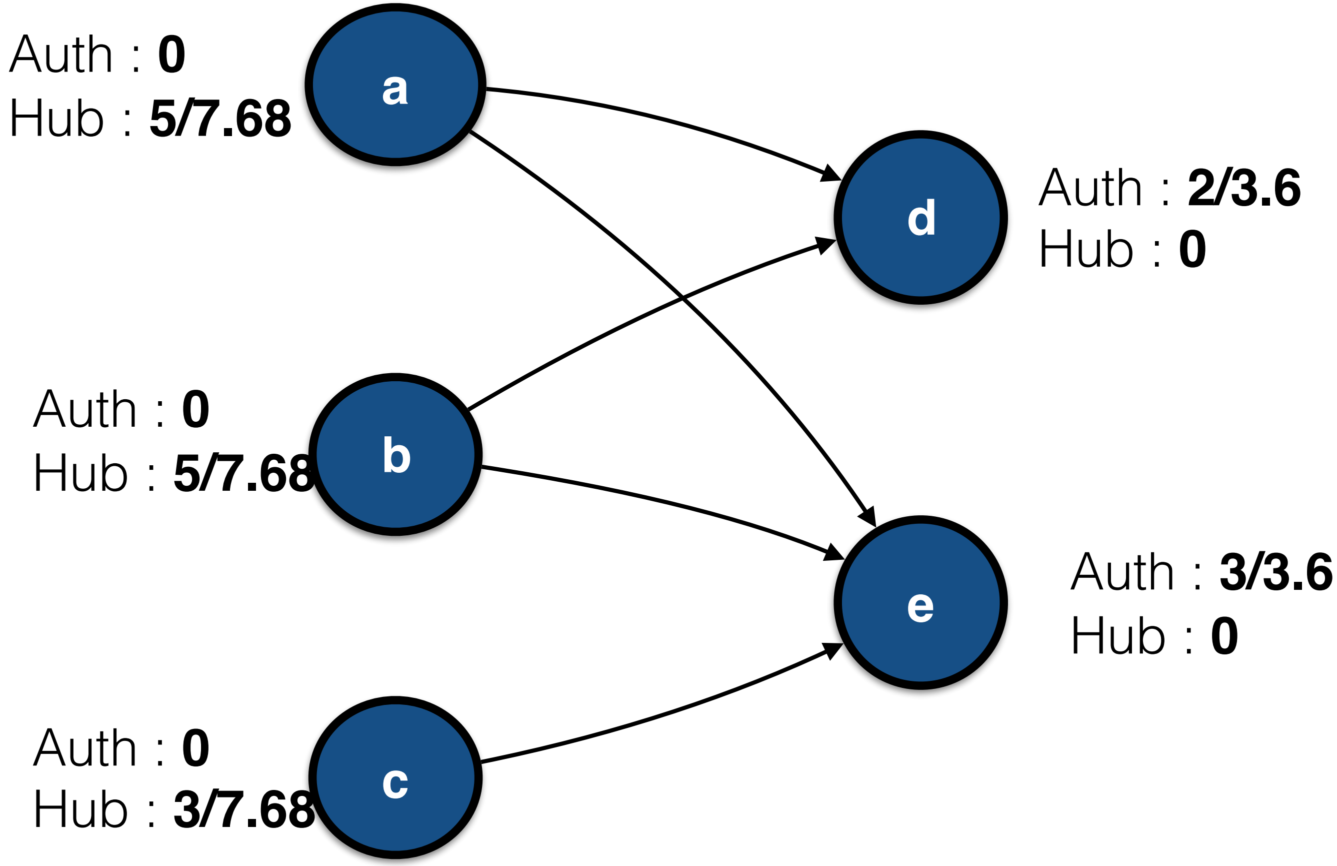
HITS - Auth1


Auth : **0**
Hub : **1**
a
Auth : **0**
Hub : **1**
b
Auth : **0**
Hub : **1**
c
d
Auth : **2**
Hub : **1**
e
Auth : **3**
Hub : **1**

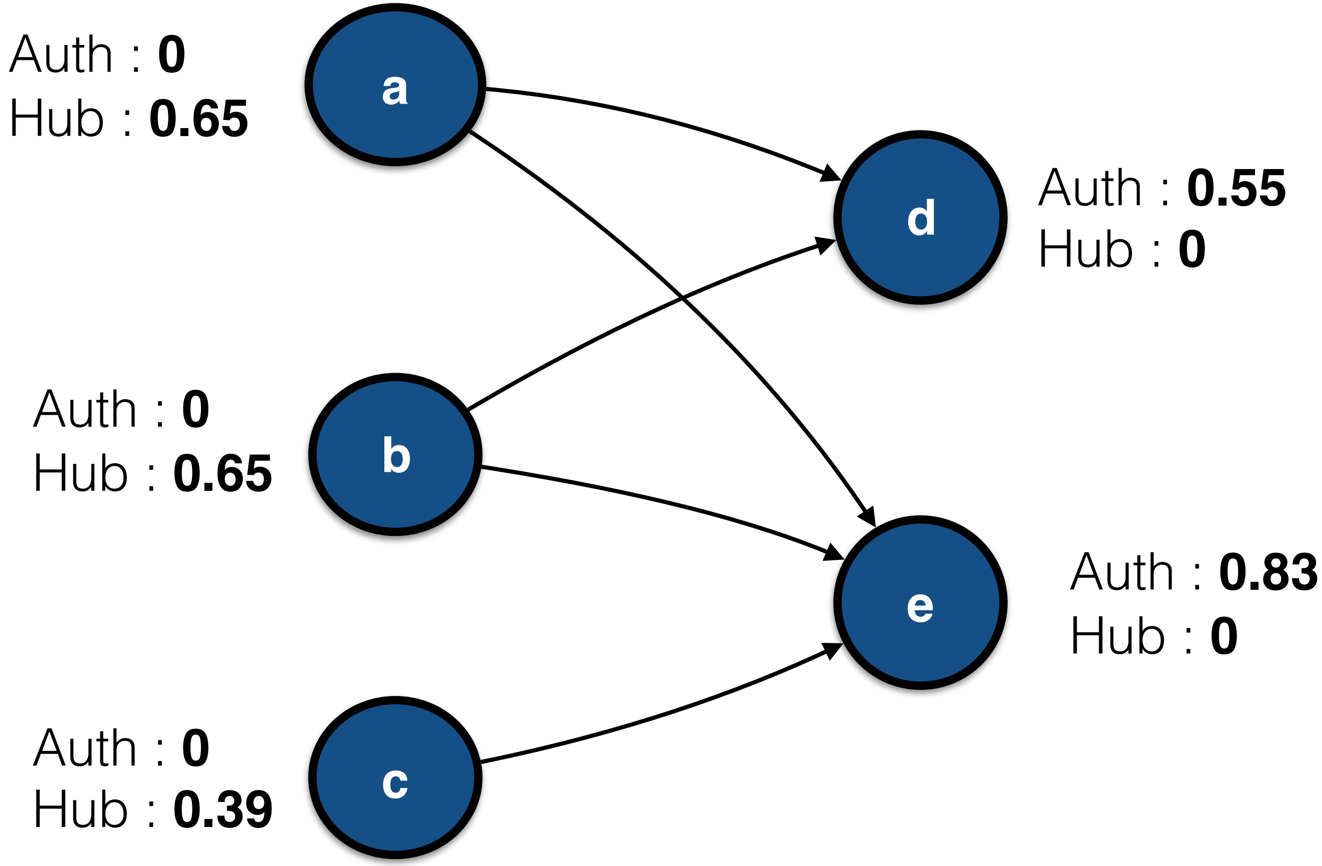
HITS - Hub1


Auth : 0
Hub : 5
a
Auth : 2
Hub : 0
d
Auth : 0
Hub : 5
b
e
Auth : 3
Hub : 0
Auth : 0
Hub : 3
c

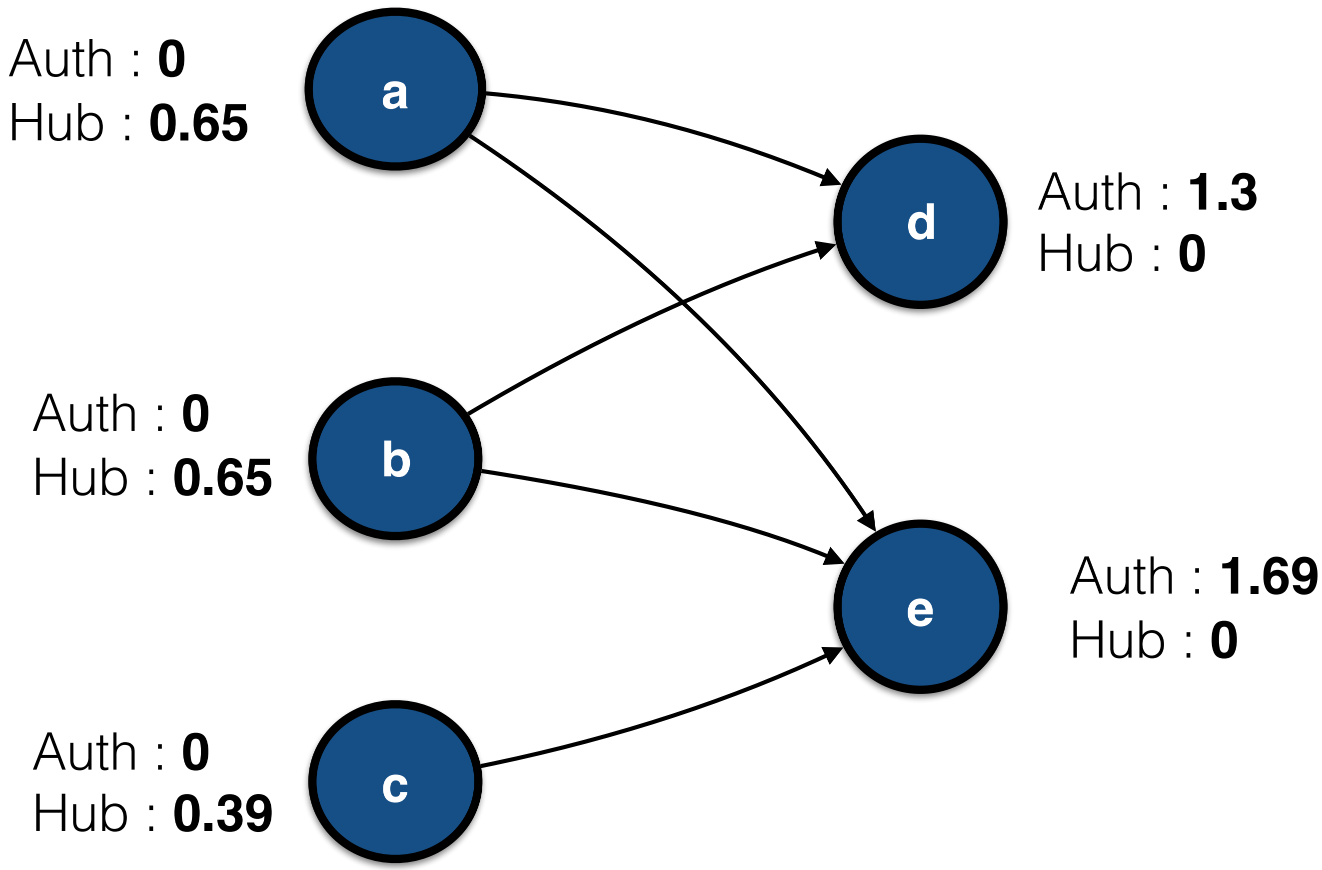
HITS - Normalisation



HITS - Normalisation

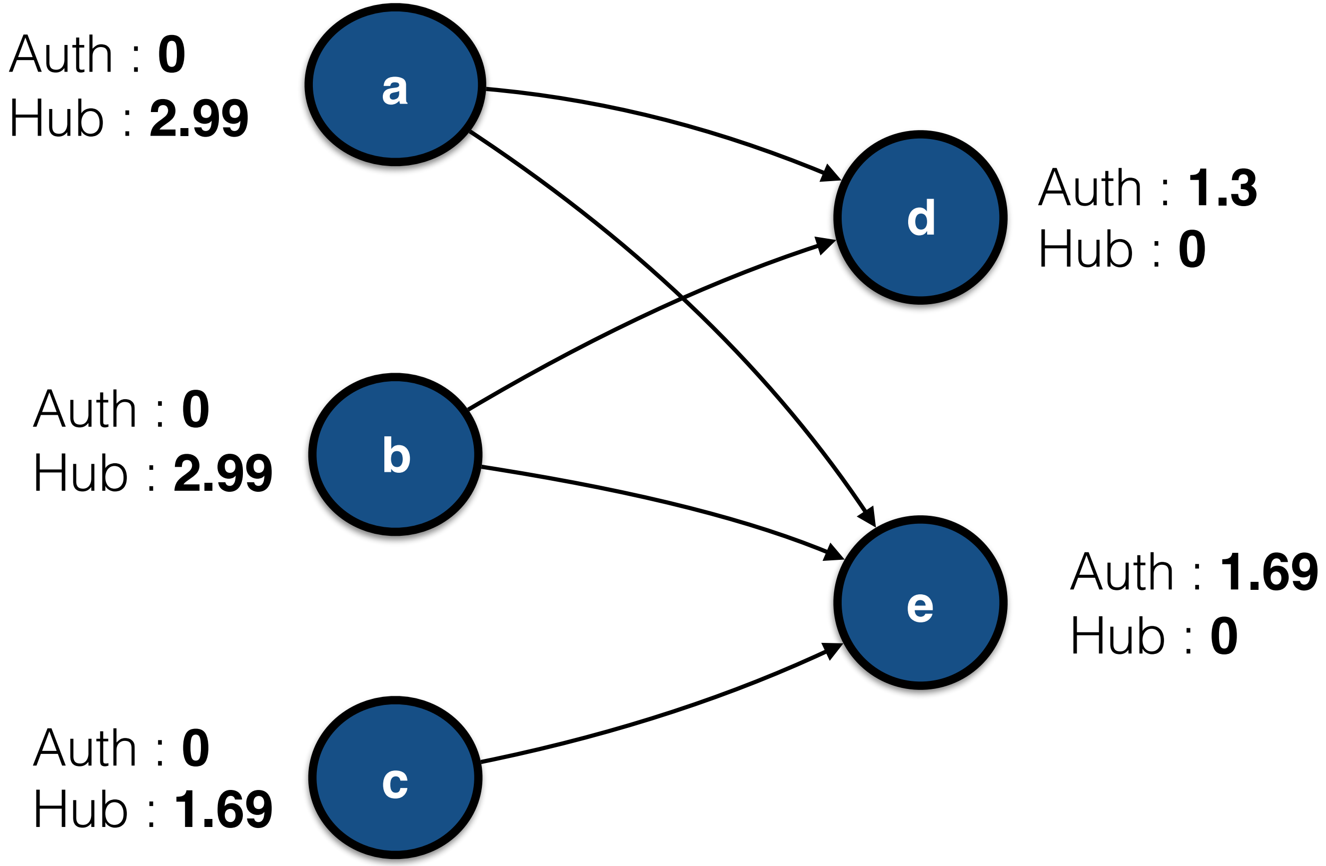


HITS - Auth2

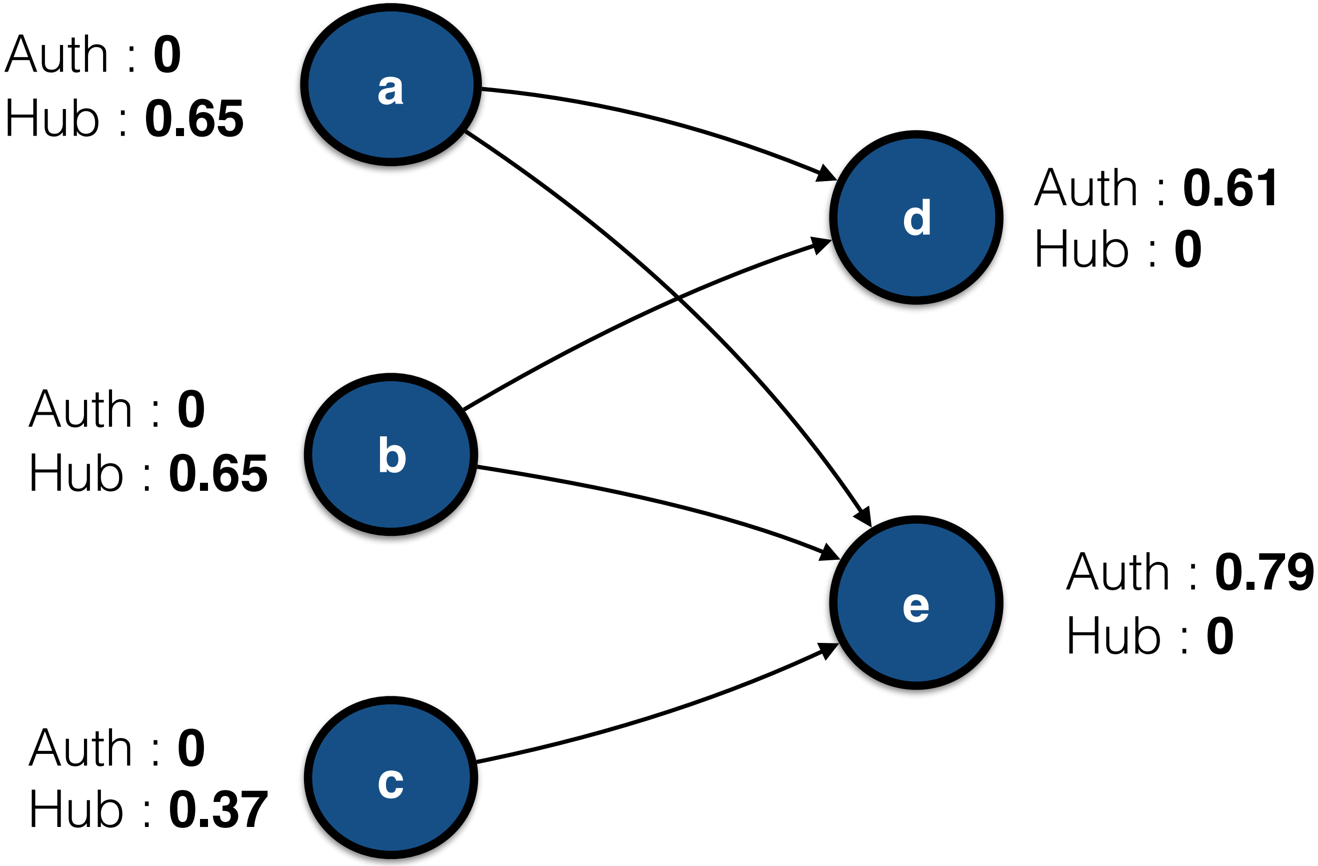


Auth : 0
Hub : 0.65
a
Auth : 0
Hub : 0.65
b
Auth : 0
Hub : 0.39
c
d
Auth : 1.3
Hub : 0
e
Auth : 1.69
Hub : 0

HITS - Hub2


Auth : **0**
Hub : **2.99**
a
d
Auth : **1.3**
Hub : **0**
Auth : **0**
Hub : **2.99**
b
e
Auth : **1.69**
Hub : **0**
Auth : **0**
Hub : **1.69**
c

HITS - Normalisation



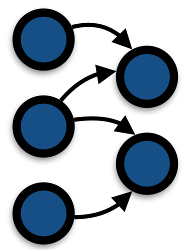
HITS Pseudo-algorithm

$$\mathcal{I} \quad auth(p) \leftarrow \sum_{q:\{q,p\}\in E} hub(q)$$

$$\mathcal{O} \quad hub(p) \leftarrow \sum_{q:\{p,q\}\in E} auth(q)$$

x = auth
y=hub

Iterate(G,k)
- G: a collection of n linked pages
- k: a natural number
- Let z denote the vector $(1, 1, 1, \ldots, 1) \in \mathbf{R}^n$.
- Set $x_0 := z$.
- Set $y_0 := z$.
- For $i = 1, 2, \ldots, k$
  - Apply the $\mathcal{I}$ operation to (x_{i-1}, y_{i-1}), obtaining new x-weights x'_i.
  - Apply the $\mathcal{O}$ operation to (x'_i, y_{i-1}), obtaining new y-weights y'_i.
  - Normalize x'_i, obtaining x_i.
  - Normalize y'_i, obtaining y_i.
- End
- Return (x_k, y_k).



Conclusion

- Many different approaches for measuring centrality
  - Degree : number of links
  - Betweenness : intermediary
  - Closeness : proximity to others
  - HITS : identify hubs & authorities
  - PageRank : mesure relative importance

- Choice of a metric depends on each situation