

ĐỘ ĐO TRUNG TÂM MẠNG

Biên soạn: **ThS. Nguyễn Thị Anh Thư**

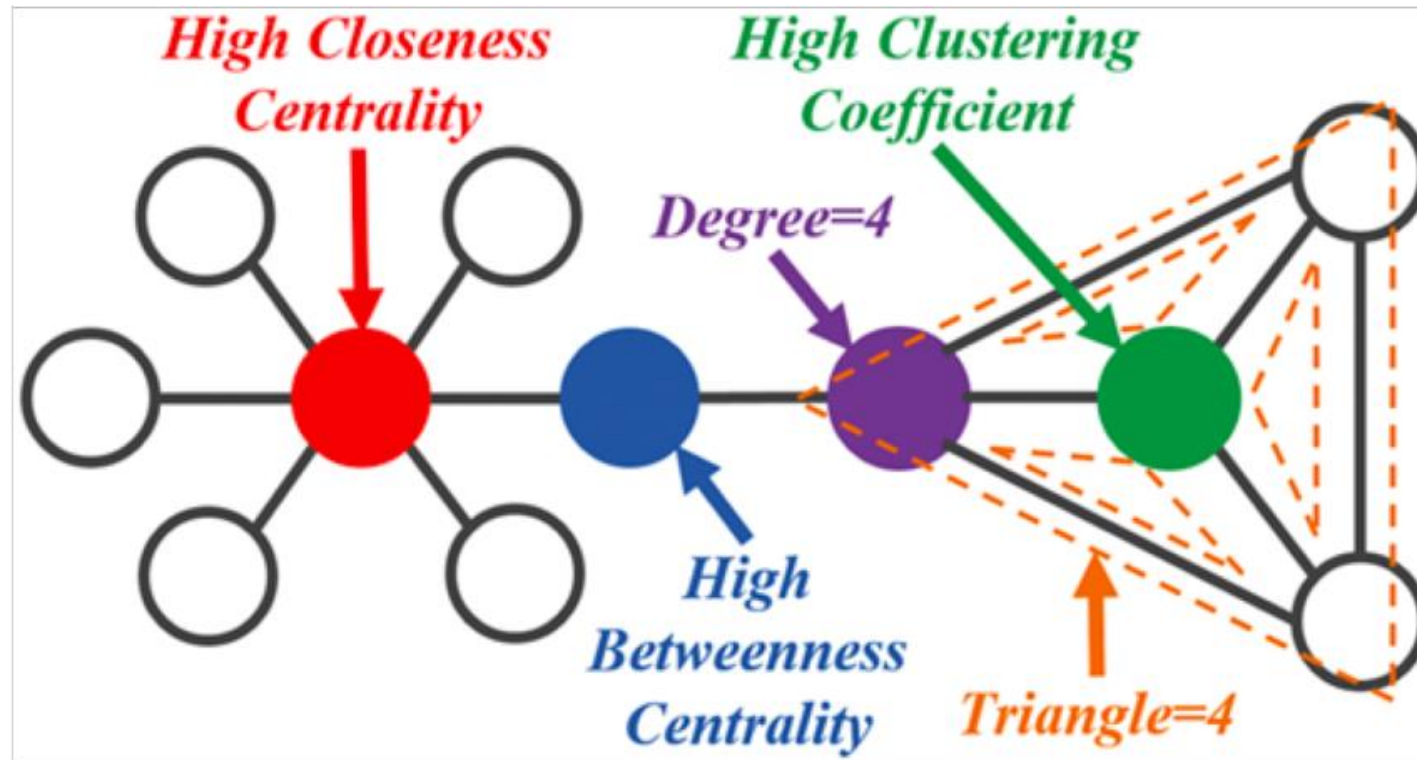
Email: thunta@uit.edu.vn

NỘI DUNG

1. Trung tâm mạng
2. Độ đo hình học
3. Độ đo dựa trên phổ đồ thị
4. Độ đo dựa trên đường đi

1. TRUNG TÂM MẠNG

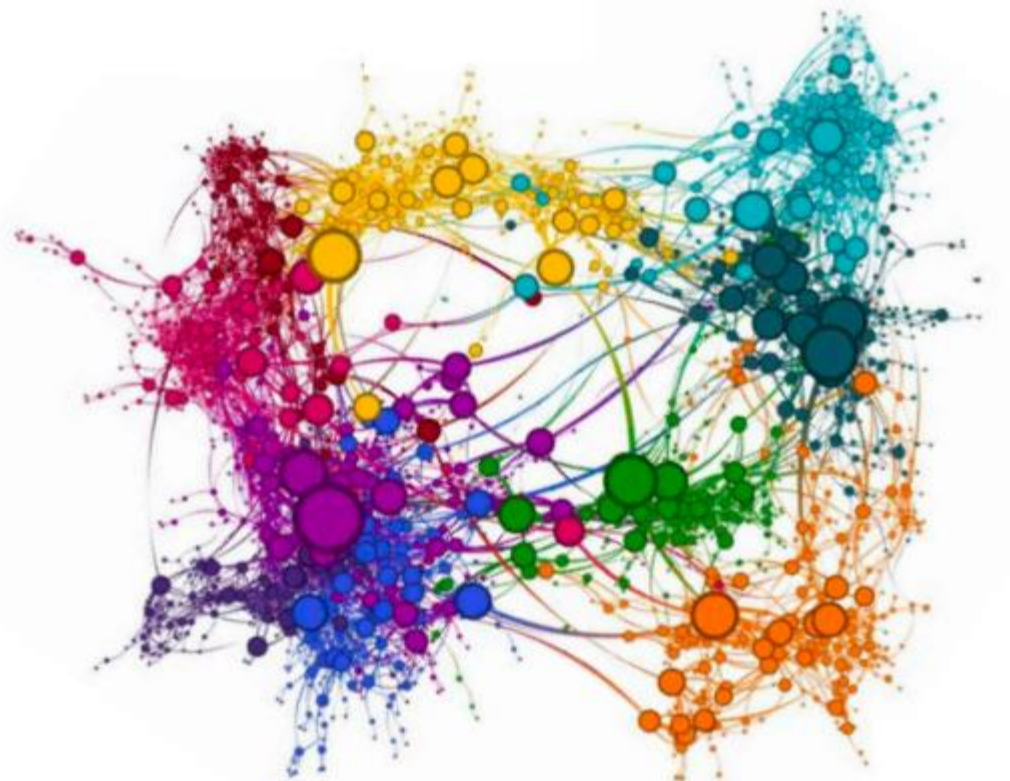
- Với một mạng xã hội, đỉnh nào **quan trọng hơn** hoặc **có ảnh hưởng hơn**?
- Các **độ đo trung tâm mạng** đã được đề xuất để giải thích tầm quan trọng của các đỉnh trong mạng.



1. TRUNG TÂM MẠNG

Tính trung tâm thường được sử dụng để phát hiện:

- **Mức độ ảnh hưởng** của một người trong mạng xã hội?
- **Đường được sử dụng tốt** như thế nào trong mạng lưới giao thông?
- Một trang web **quan trọng** như thế nào?



1. TRUNG TÂM MẠNG

Độ đo trung tâm mạng

- **Độ đo hình học**

- Tầm quan trọng là một *hàm tính khoảng cách* đến các đỉnh khác.

- **Độ đo dựa trên phổ đồ thị**

- Dựa trên *cấu trúc eigen* của một số ma trận liên quan đến đồ thị.

- **Độ đo dựa trên đường đi**

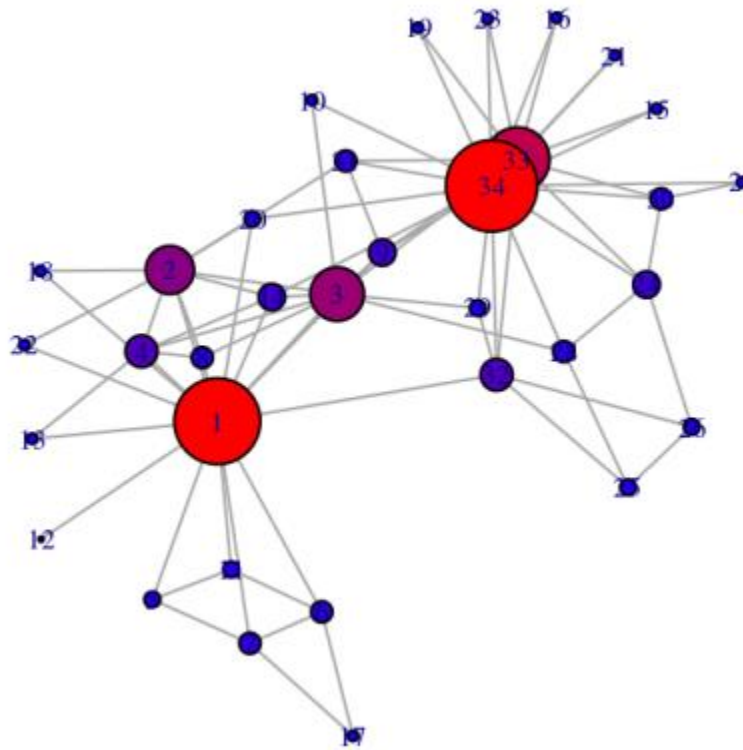
- Tính đến tất cả *các đường đi* vào một đỉnh.

2. ĐỘ ĐO HÌNH HỌC

(In-)Degree Centrality

- Số lượng liên kết đến.

$$c_{deg}(x) = d_{in}(x)$$



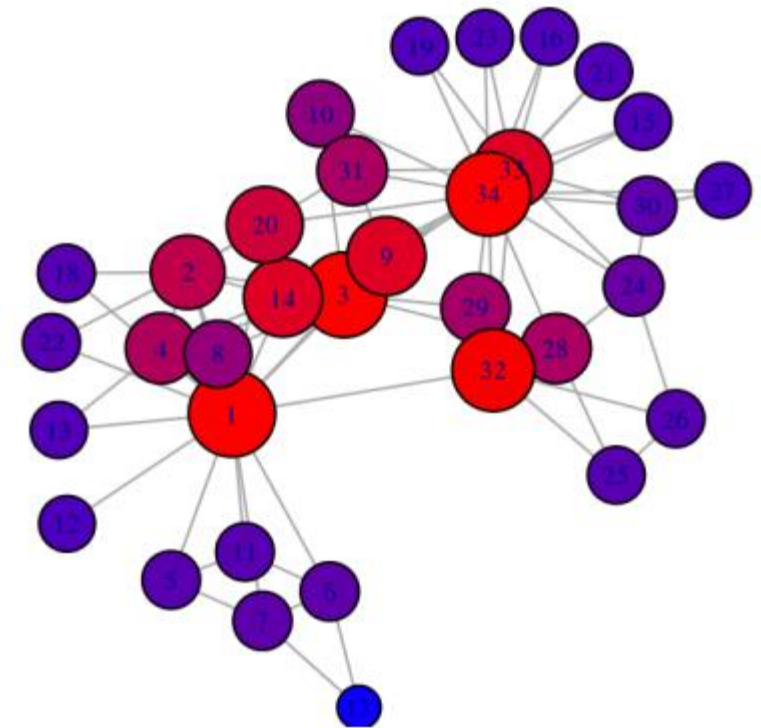
2. ĐỘ ĐO HÌNH HỌC

Closeness Centrality

- Các đỉnh ở gần trung tâm hơn thì có khoảng cách nhỏ hơn và độ trung tâm cao hơn.

$$c_{clos}(x) = \frac{1}{\sum_y d(y, x)}$$

- Trong đó:
 - $d(y, x)$: Độ dài đường đi ngắn nhất từ đỉnh y đến đỉnh x .



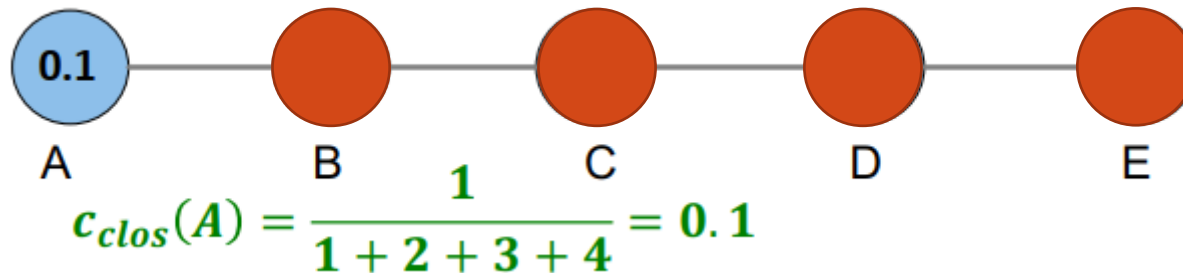
2. ĐỘ ĐO HÌNH HỌC

Closeness Centrality

- Các đỉnh ở gần trung tâm hơn thì có khoảng cách nhỏ hơn và độ trung tâm cao hơn.

$$c_{clos}(x) = \frac{1}{\sum_y d(y, x)}$$

- Mức độ một đỉnh có thể giao tiếp mà không cần phụ thuộc vào các bên thứ ba để thông điệp của anh ta được gửi đi.



2. ĐỘ ĐO HÌNH HỌC

Harmonic Centrality

- Thay khoảng cách trung bình bằng giá trị trung bình của mọi khoảng cách.

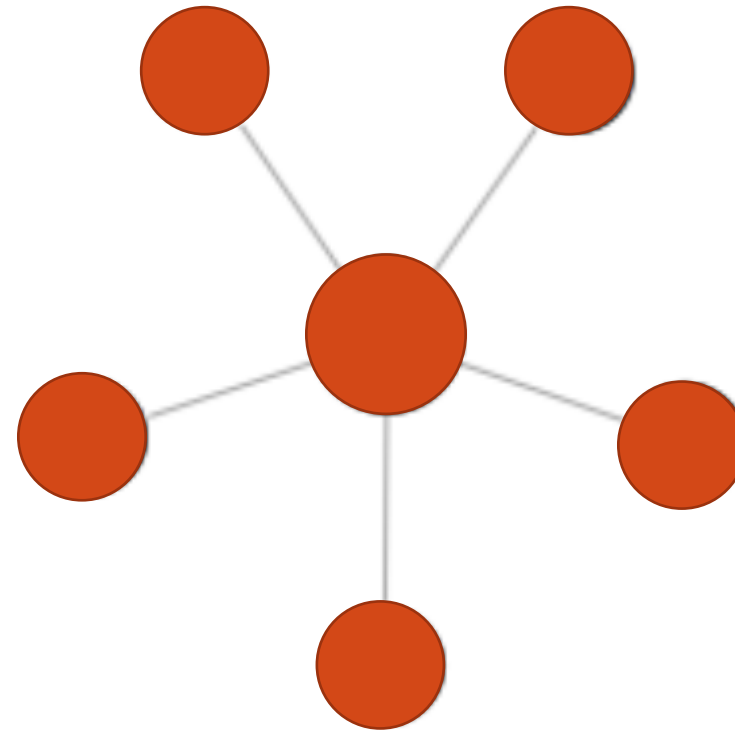
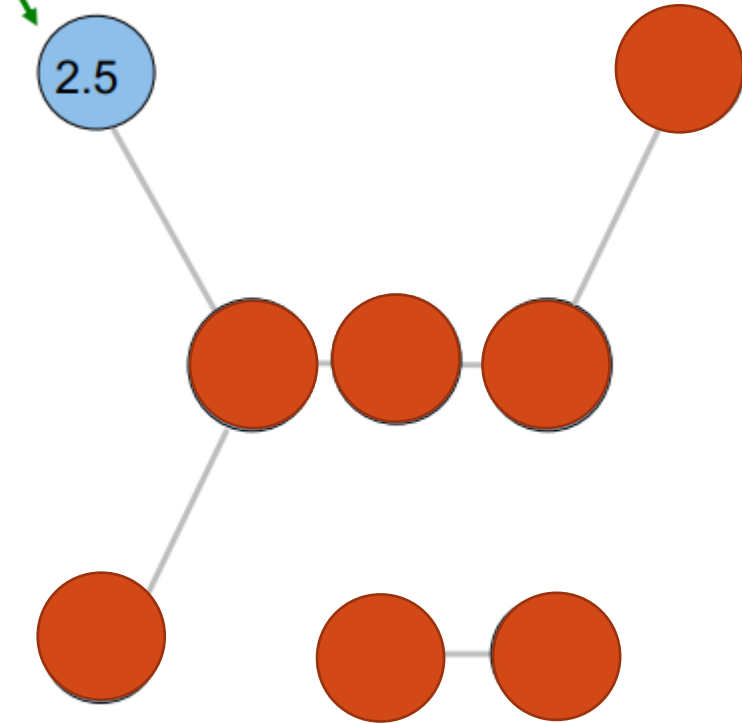
$$c_{har}(x) = \sum_{y \neq x} \frac{1}{d(y, x)}$$

- Đặc điểm:
 - Tương quan chặt chẽ với Closeness Centrality.
 - Có thể áp dụng cho các đồ thị không được kết nối mạnh.

2. ĐỘ ĐO HÌNH HỌC

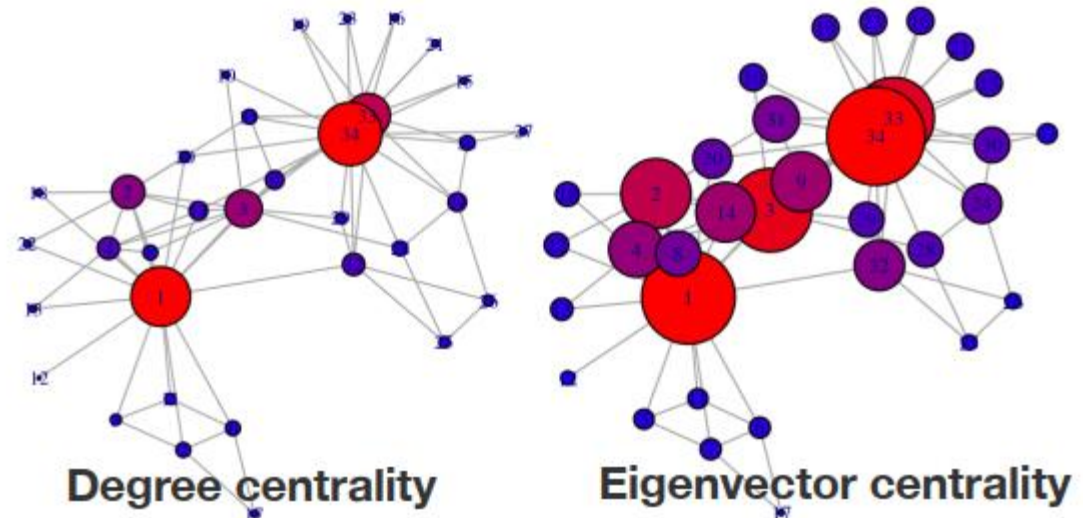
Harmonic Centrality

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



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

- **Tính toán dựa trên một số ma trận liên quan đến đồ thị.**
- **Ý tưởng:** Vị trí trung tâm của một đỉnh phụ thuộc vào vị trí trung tâm của các đỉnh liên kề.
 - Các đỉnh được kết nối với các đỉnh trung tâm có điểm trung tâm cao hơn các đỉnh được kết nối với các đỉnh không là trung tâm.
- **Eigenvector Centrality**
- **Katz's Index**
- **Page Rank**
- **Hits**



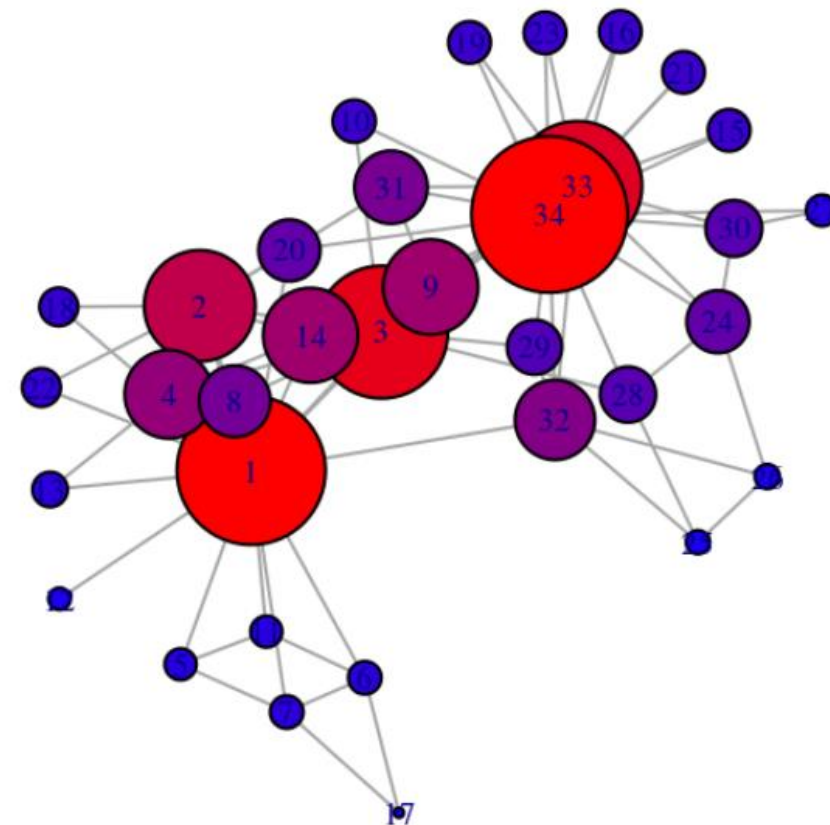
3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Eigenvector Centrality

- Đo lường mức độ ảnh hưởng của một đỉnh trong mạng.
- **Ý tưởng:** Mọi đỉnh đều bắt đầu với cùng một điểm và sau đó mỗi đỉnh sẽ cho điểm của nó cho người kế nhiệm.

$$c_{\text{eig}}(x) = \frac{1}{\lambda} \sum_{y \rightarrow x} c_{\text{eig}}(y)$$

Normalization constant = $\|c_{\text{eig}}\|_2$



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Eigenvector Centrality

- Đo lường mức độ ảnh hưởng của một đỉnh trong mạng.

$$c_{\text{eig}}(x) = \frac{1}{\lambda} \sum_{y \rightarrow x} c_{\text{eig}}(y)$$

- Trong đó:
 - c_{eig} : hội tụ đến vector eigen của ma trận kề A .
 - λ : hội tụ đến giá trị đặc biệt của vector eigen của ma trận kề A .
 - $\lambda = \|c_{\text{eig}}\|_2 = \sqrt{\sum_{r=1}^n |c_r|^2}$ với c là vector có các thành phần c_1, c_1, \dots, c_r .

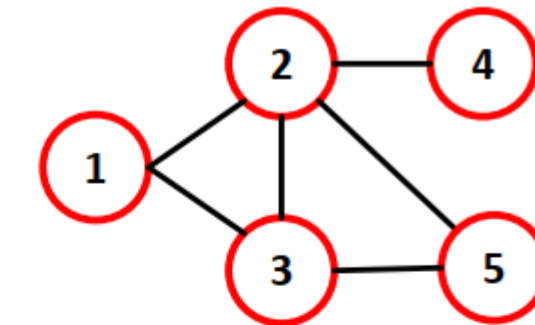
3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Eigenvector Centrality

■ Power Iteration:

- Set $\mathbf{c}^{(0)} \leftarrow \mathbf{1}, k \leftarrow 1$
- **1:** $\mathbf{c}^{(k)} \leftarrow A\mathbf{c}^{(k-1)}$
- **2:** $\mathbf{c}^{(k)} = \mathbf{c}^{(k)} / \|\mathbf{c}^{(k)}\|_2$
- **3:** If $\|\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\| > \varepsilon$:
- **4:** $k \leftarrow k + 1$, goto **1**

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{matrix}$$



$$\mathbf{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

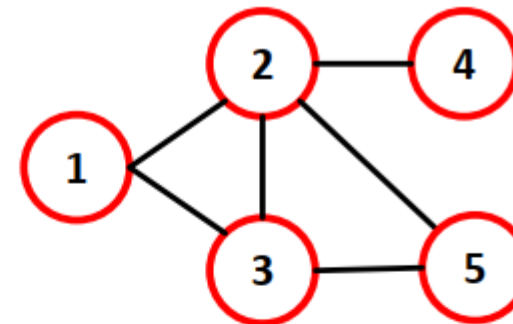
Eigenvector Centrality

■ Power Iteration:

Iteration 1

$$\begin{matrix}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\
 \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}
 \end{matrix} = \begin{matrix} \text{orange bar} \end{matrix} \equiv \begin{matrix} \text{orange bar} \end{matrix}$$

$A \quad c^{(0)} \quad c^{(1)} = Ac^{(0)} \quad c^{(1)} = c^{(1)} / \|c^{(1)}\|_2$



Iteration 2

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \text{orange bar} \end{matrix} = \begin{matrix} \text{orange bar} \end{matrix} \equiv \begin{matrix} \text{orange bar} \end{matrix}$$

Iteration 3

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \text{orange bar} \end{matrix} = \begin{matrix} \text{orange bar} \end{matrix} \equiv \begin{matrix} \text{orange bar} \end{matrix} \quad \dots \quad c = \begin{matrix} \text{orange bar} \end{matrix}$$

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Katz's Index

- Đo lường ảnh hưởng bằng cách tính đến tổng số đường đi giữa một cặp đỉnh.

$$c_{\text{katz}}(x) = \beta \sum_{k=0}^{\infty} \sum_{x \rightarrow y} \alpha^k \boxed{(A^k)_{xy}}$$

Total number of walks
of length k between x, y

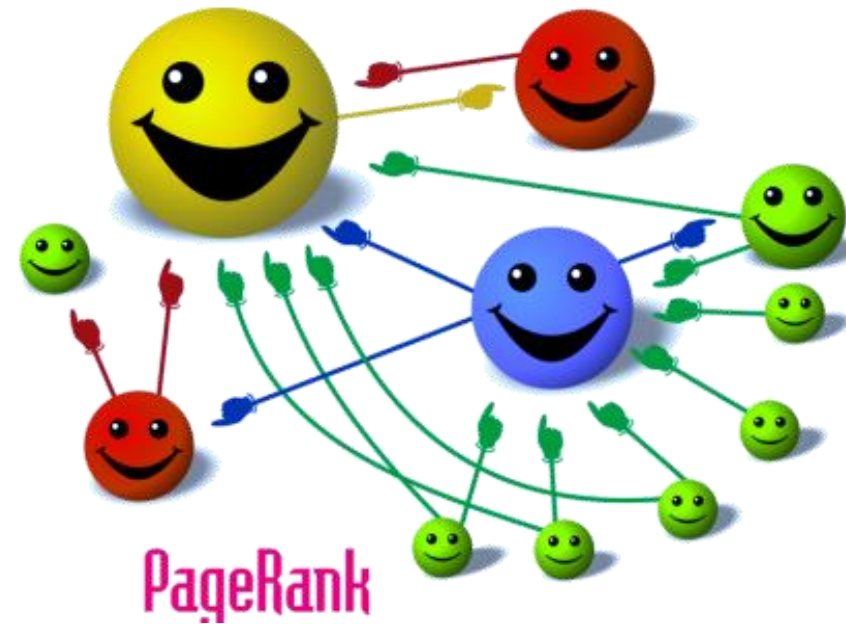
- Trong đó:
 - α là một giá trị suy giảm trong phạm vi $(0, \frac{1}{\lambda})$, với λ là giá trị lớn nhất của ma trận kề A .
 - β cho phép một đỉnh có nhiều đặc quyền hơn.
- **Katz's Index phù hợp với đồ thị vòng có hướng.**

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank

- Được Sergey Brin, Lawrence Page và các cộng sự đưa ra vào năm 1998.
- Một trong những thuật toán được dùng cho bộ máy tìm kiếm Google.

Ý tưởng thuật toán:
“Nếu có liên kết từ trang A đến trang B thì độ quan trọng của trang A cũng ảnh hưởng đến độ quan trọng của trang B.”



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank

- Công thức tính toán giá trị PageRank $P(i)$ của trang i :

$$P(i) = \sum_{j \in B(i)} \frac{P(j)}{O_j}$$

- Trong đó:
 - $B(i)$ là tập các trang có liên kết trở tới trang i .
 - O_j là số liên kết ra từ trang j (outbound link).

Mô hình duyệt ngẫu nhiên

- Vector Page Rank tỉ lệ với phân phối xác suất dừng của một quá trình ngẫu nhiên.
- Page Rank của một trang web chính là xác suất để 1 người ngẫu nhiên duyệt trang web đó.

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Random Surfer Model

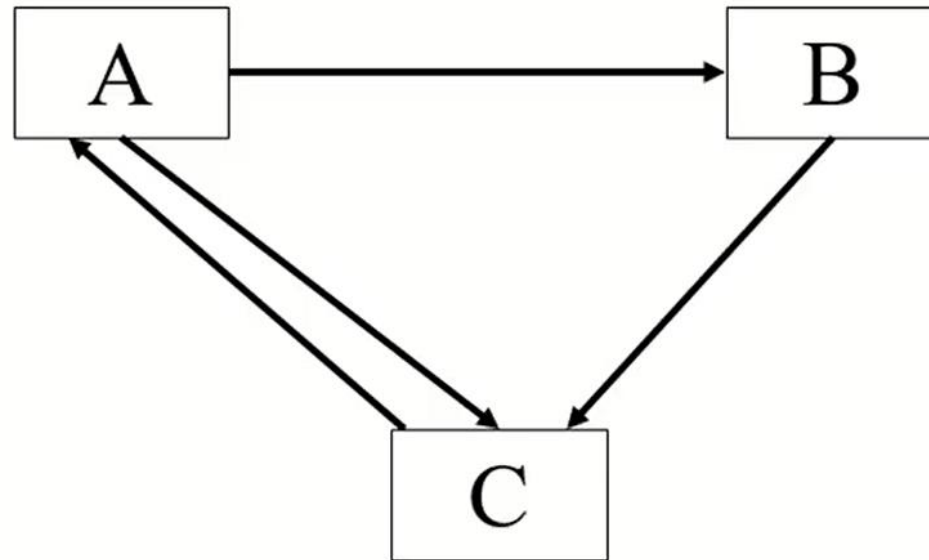
$$PR(x) = \frac{(1 - d) * 100}{N} + d \sum_{y \rightarrow x} \frac{PR(y)}{Out(y)}$$

- Page Rank của nút X là xác suất ở nút X tại thời điểm hiện tại.
- (1-d): Bước nhảy ngẫu nhiên. Xác suất kết thúc tại nút X vì một bước nhảy ngẫu nhiên từ một số nút, bao gồm cả nút X, là 1 / N.
- Tuy nhiên, bản thân một bước nhảy ngẫu nhiên như vậy có thể xảy ra với xác suất (1-d).
- Điều này tương đương với xác suất (1-d) / N ở nút X do một bước nhảy ngẫu nhiên.

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Random Surfer Model

- Độ đo Page Rank (PR) cho mỗi đỉnh là 1.



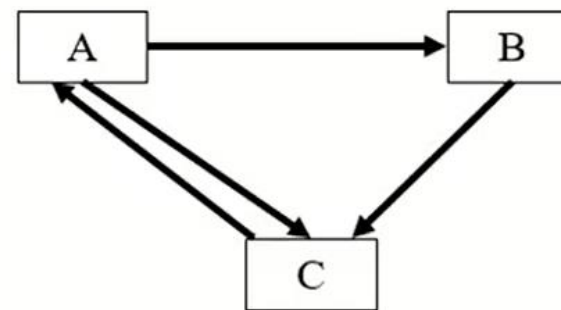
$$PR(A) = (1-d) + d(PR(T_i)/C(T_i) + \dots + PR(T_n)/C(T_n))$$

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Random Surfer Model

$$PR(A) = (1-d) + d(PR(T_i)/C(T_i) + \dots + PR(T_n)/C(T_n))$$

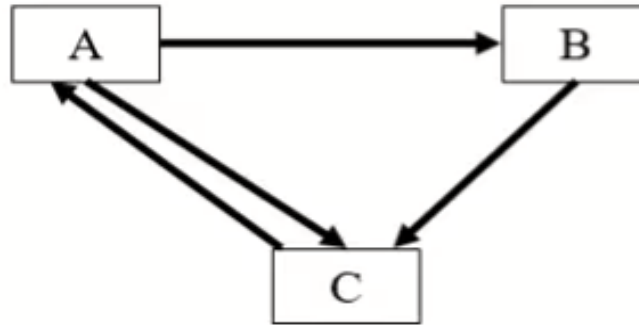
$$\begin{aligned} PR(A) &= (1-d) + d [PR(C) / C(C)] \\ &= (1-0.85) + 0.85 [1/1] \\ &= 0.15 + 0.85 [1] \\ &= 0.15 + 0.85 \\ &= 1 \end{aligned}$$



$$\begin{aligned} PR(B) &= (1-d) + d [PR(A) / C(A)] \\ &= (1-0.85) + 0.85 [(1) / 2] \\ &= 0.15 + 0.85 [0.5] \\ &= 0.15 + 0.425 \\ &= 0.575 \end{aligned}$$

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

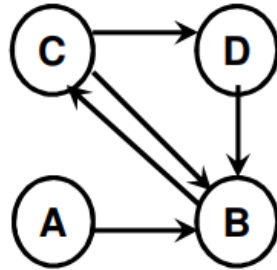
Thuật toán Page Rank – Random Surfer Model



Iteration	A	B	C
0	1	1	1
1	1	0.575	1.06375
2	1.0541875	0.598029688	1.106354922
3	1.090401684	0.613420716	1.134828324
4	1.114604075	0.623706732	1.153857454
5	1.130778836	0.630581005	1.16657486
6	1.141588631	0.635175168	1.175074061

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Random Surfer Model



Note that there are NO sink nodes
(nodes without any out-going links)

Assume damping
Factor $d = 0.85$

$$\begin{aligned} PR(A) &= (1-d) \cdot 100/4 \\ PR(B) &= (1-d) \cdot 100/4 + d \cdot [PR(A) + 1/2 \cdot PR(C) + PR(D)] \\ PR(C) &= (1-d) \cdot 100/4 + d \cdot [PR(B)] \\ PR(D) &= (1-d) \cdot 100/4 + d \cdot [1/2 \cdot PR(C)] \end{aligned}$$

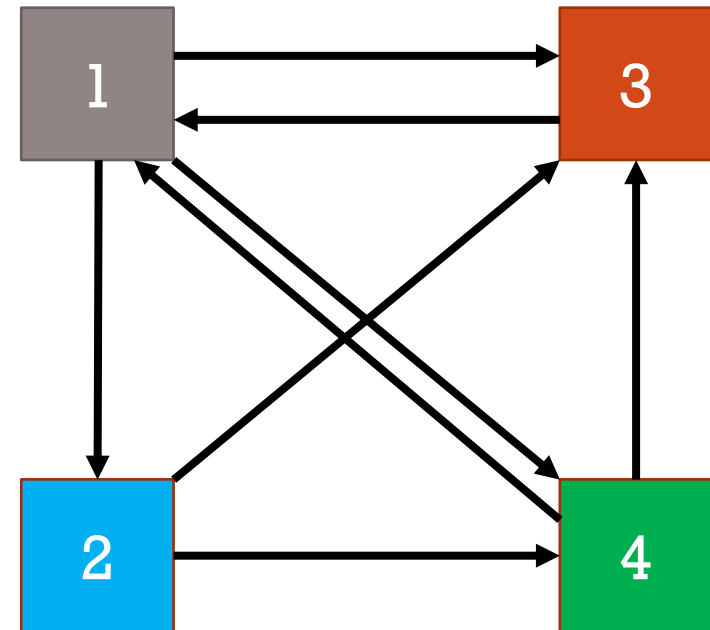
Initial $PR(A) = 25$ $PR(B) = 25$ $PR(C) = 25$ $PR(D) = 25$	It # 1 $PR(A) = 3.75$ $PR(B) = 56.88$ $PR(C) = 25$ $PR(D) = 14.38$	It # 2 $PR(A) = 3.75$ $PR(B) = 29.79$ $PR(C) = 52.10$ $PR(D) = 14.38$	It # 3 $PR(A) = 3.75$ $PR(B) = 41.30$ $PR(C) = 29.07$ $PR(D) = 25.89$	It # 4 $PR(A) = 3.75$ $PR(B) = 41.29$ $PR(C) = 38.86$ $PR(D) = 16.10$
It # 5 $PR(A) = 3.75$ $PR(B) = 37.14$ $PR(C) = 38.85$ $PR(D) = 20.27$	It # 6 $PR(A) = 3.75$ $PR(B) = 40.68$ $PR(C) = 35.32$ $PR(D) = 20.26$	It # 7 $PR(A) = 3.75$ $PR(B) = 39.17$ $PR(C) = 38.33$ $PR(D) = 18.76$	It # 8 $PR(A) = 3.75$ $PR(B) = 39.17$ $PR(C) = 37.04$ $PR(D) = 20.04$	It # 9 $PR(A) = 3.75$ $PR(B) = 39.71$ $PR(C) = 37.04$ $PR(D) = 19.49$
It # 10 $PR(A) = 3.75$ $PR(B) = 39.25$ $PR(C) = 37.5$ $PR(D) = 19.49$	Ranking B C D A			

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Dựa vào ma trận

- Độ đo Page Rank (PR) cho mỗi đỉnh là $1 / N$.
- Ma trận vote A (chiều đi, chiều cho độ quan trọng giữa các đỉnh).
- Ví dụ:

$$v = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}, A = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán Page Rank – Dựa vào ma trận

- Độ đo Page Rank (PR) cho mỗi đỉnh là $1 / N$.
- Ma trận vote A (chiều đi, chiều cho độ quan trọng giữa các đỉnh).
- Page Rank theo ma trận:

$$\begin{aligned} A \cdot \vec{v} &= \vec{v^2} \\ A \cdot \vec{v^2} &= \vec{v^3} \\ &\dots \end{aligned}$$

Dừng khi $\left\| \vec{v^k} - \vec{v^{k-1}} \right\| \leq \varepsilon$

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

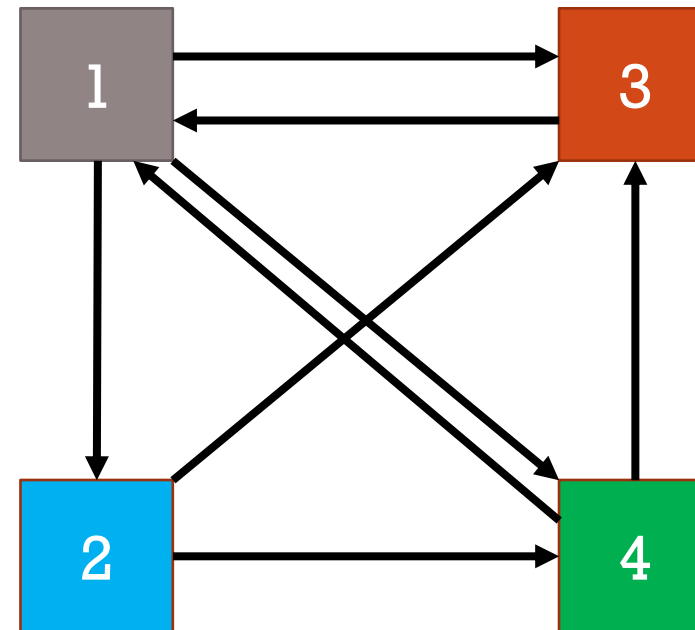
Thuật toán Page Rank – Dựa vào ma trận

▪ Ví dụ:

$$v = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}, A = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

Sau lần lặp thứ 8:

$$v = \begin{bmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{bmatrix}$$



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Các thuật toán cải tiến Page Rank:

Topic sensitive
PageRank
(2002)

Adaptive
PageRank
(2003)

Timed
PageRank
(2008)

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán HITS

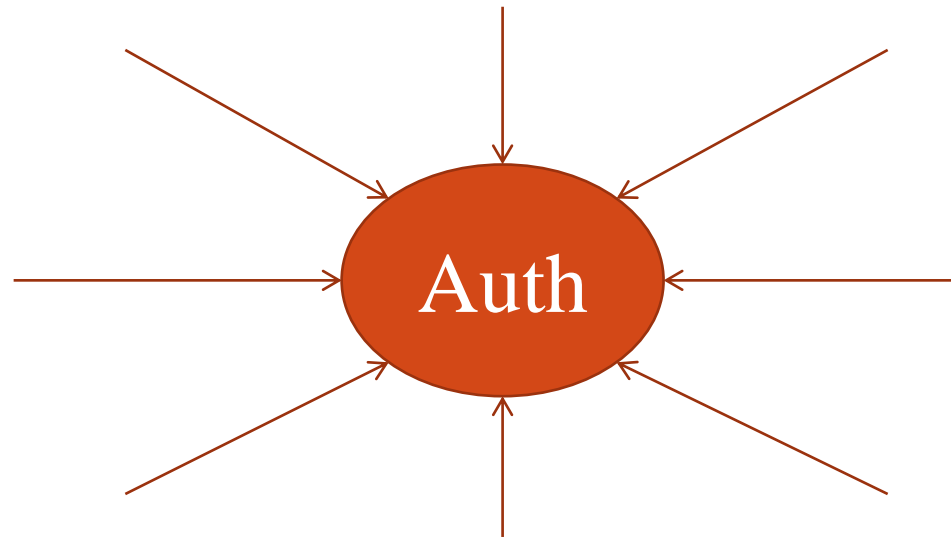
- HITS – **Hypertext Induced Topic Selection**
- Jon M. Kleinberg đề xuất, 1998 – 1999.
- Không giống như Page Rank – một thuật toán xếp hạng tĩnh, HITS phụ thuộc vào truy vấn tìm kiếm.
- Khi người sử dụng thực hiện một truy vấn tìm kiếm:
 - HITS đầu tiên mở rộng danh sách các trang có liên quan được trả về bởi **một công cụ tìm kiếm**.
 - Sau đó, tạo ra hai bảng xếp hạng của các thiết lập mở rộng của các trang, xếp hạng *authority* và *hub*.

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán HITS

Authority

- **Authority pages:** là những trang được xem là phù hợp nhất đối với mỗi câu truy vấn cụ thể nào đó.

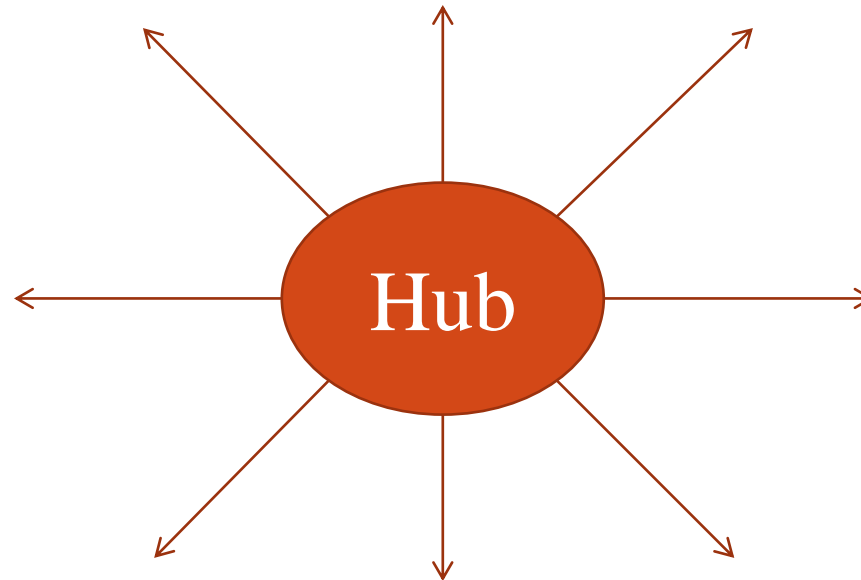


3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán HITS

Hub

- **Hub pages:** là những trang không cần có đặc tính “*authority*” nhưng lại trỏ tới nhiều trang có đặc tính “*authority*”.



3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

Thuật toán HITS

Ý tưởng

- Trang có hub tốt là trang có nhiều liên kết ra.
- Trang có authority tốt là trang có nhiều liên kết tới.
- Trang trở tới trang có authority cao thì trọng số hub càng cao, trang nào được nhiều trang có hub cao trở tới thì trọng số authority càng cao.

3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

So sánh Page Rank và HITS

Page Rank

Page Rank xếp hạng trang chỉ bởi các trang authority.

Page Rank được áp dụng cho toàn bộ trang web.

Là truy vấn độc lập.

HITS

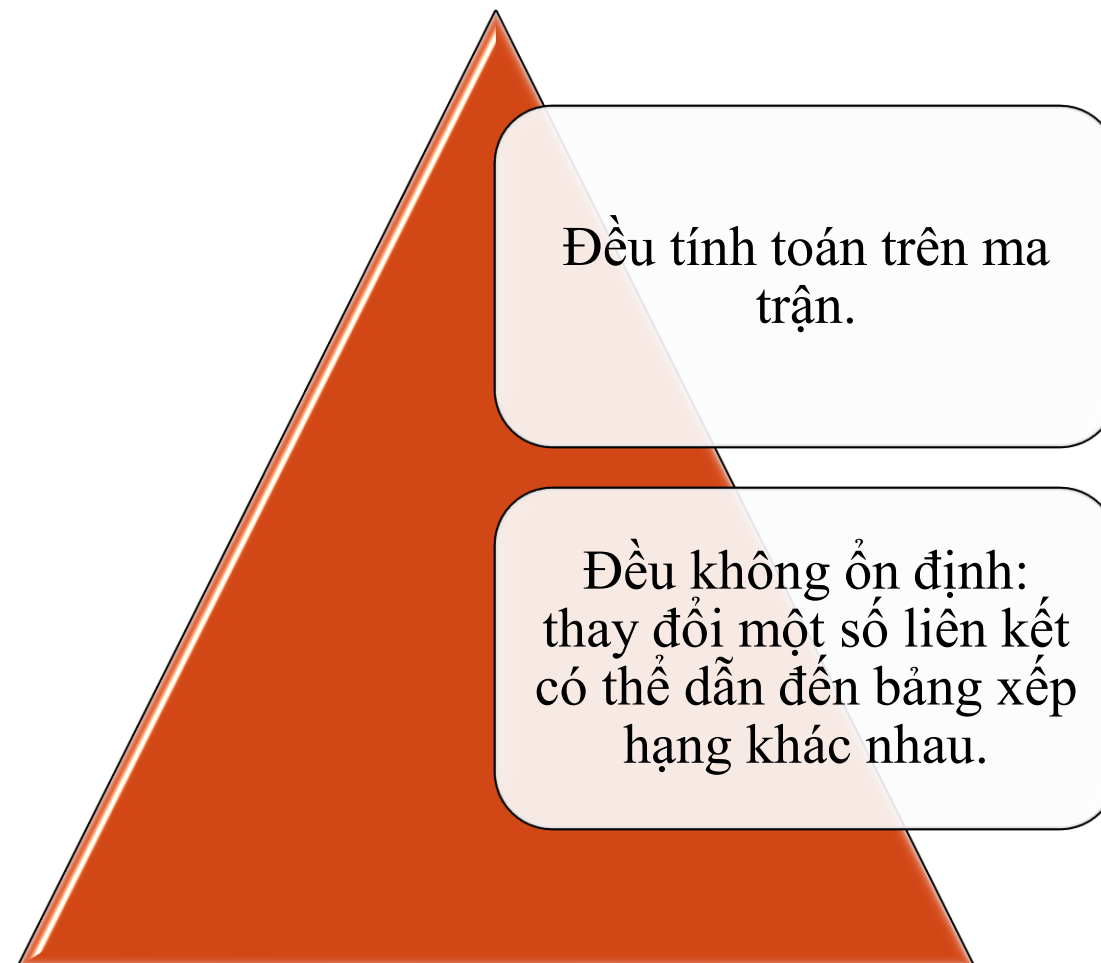
Nhấn mạnh tăng cường lẫn nhau giữa các trang authority và hub.

Được áp dụng cho các vùng lân cận của các trang xung quanh kết quả của một truy vấn.

Là truy vấn phụ thuộc

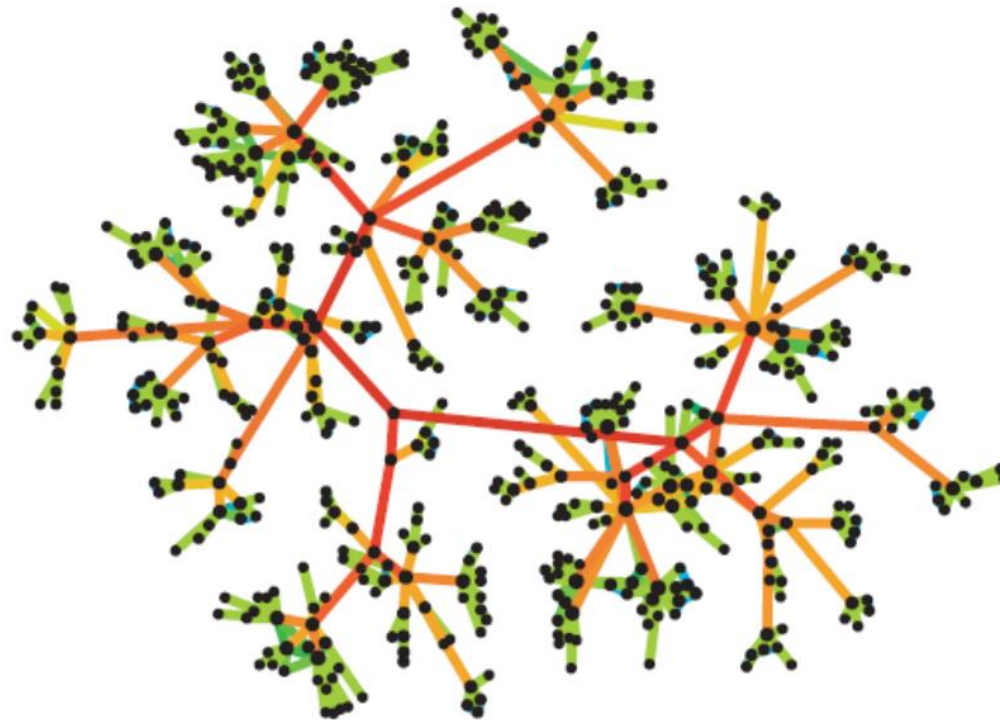
3. ĐỘ ĐO DỰA TRÊN PHỔ ĐỒ THỊ

So sánh Page Rank và HITS

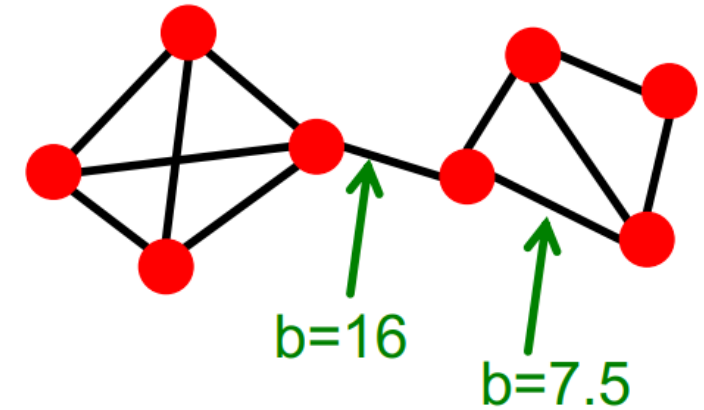


4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

- **Edge betweenness**: Số đường đi ngắn nhất đi qua cạnh.



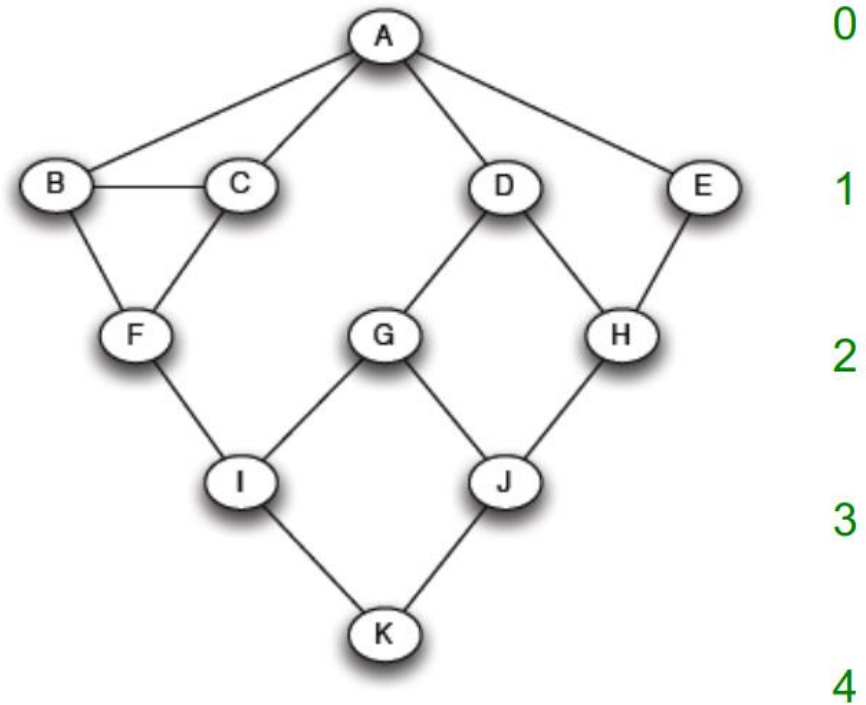
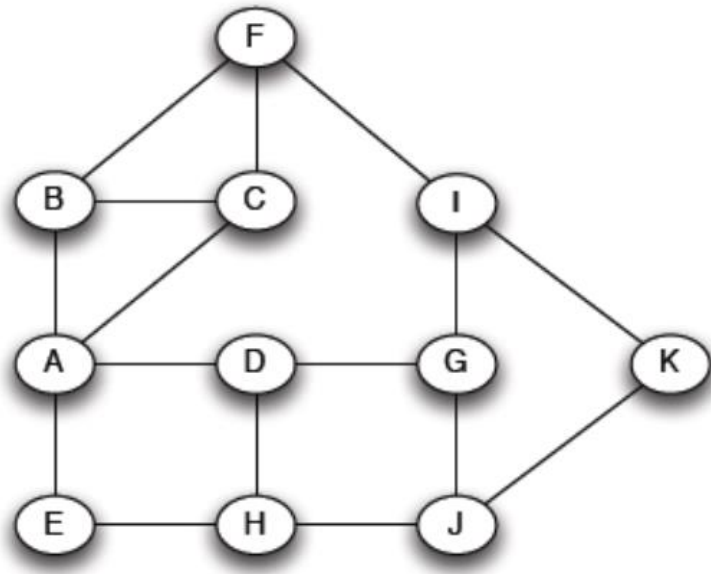
Edge betweenness
in a real network



4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Làm thế nào để tính toán Betweenness?

- Muốn tính toán Betweenness từ đỉnh $A \rightarrow$ **Breadth first search** từ đỉnh A .



4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Làm thế nào để tính toán Betweenness?

- **Forward step:** Đếm số đường đi ngắn nhất σ_{Ai} từ đỉnh A đến tất cả các đỉnh i trong mạng.
- **Backward step:** Tính toán Betweenness nếu có nhiều đường đi ngắn nhất.
- **Betweenness centrality** của đỉnh x là xác suất một đường đi ngắn nhất đi qua x .

4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Betweenness centrality

- Đo xác suất để một đường đi ngắn nhất ngẫu nhiên đi qua một đỉnh nhất định.
- Có thể được tính toán từ trung tâm của các cạnh:

$$c_{\text{bet}}(x) = \sum_{y,z \neq x, \sigma_{yz} \neq 0} \frac{\sigma_{yz}(x)}{\sigma_{yz}}$$

- Trong đó:
 - σ_{yz} : Số đường đi ngắn nhất từ y đến z .
 - $\sigma_{yz}(x)$: Số đường đi ngắn nhất từ y đến z đi qua x .

4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Betweenness centrality

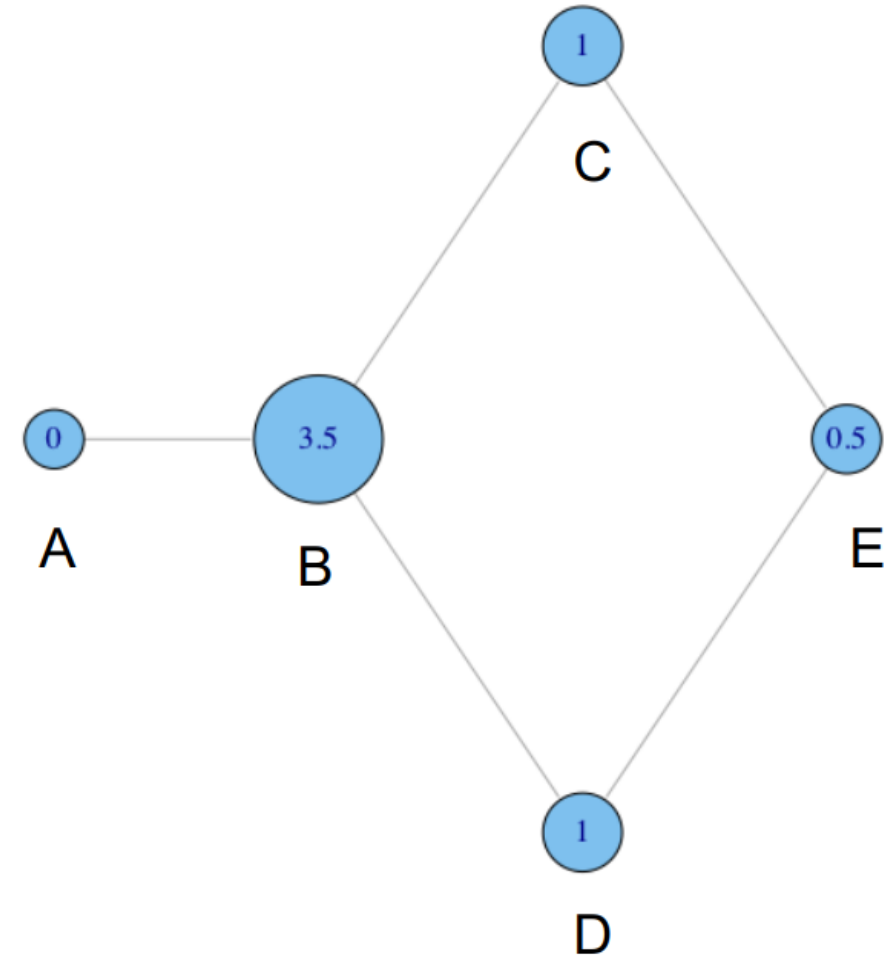
- Ví dụ 1: Tính **Betweenness** ($\sigma_{yz}(x)$).

- Tại sao C và D có **Betweenness** là 1?

- Chúng ta có đường đi ngắn nhất của (A,E) và (B,E) đều đi qua C và D. Vì thế, chúng phải chia Betweenness:

$$\frac{1}{2} + \frac{1}{2} = 1$$

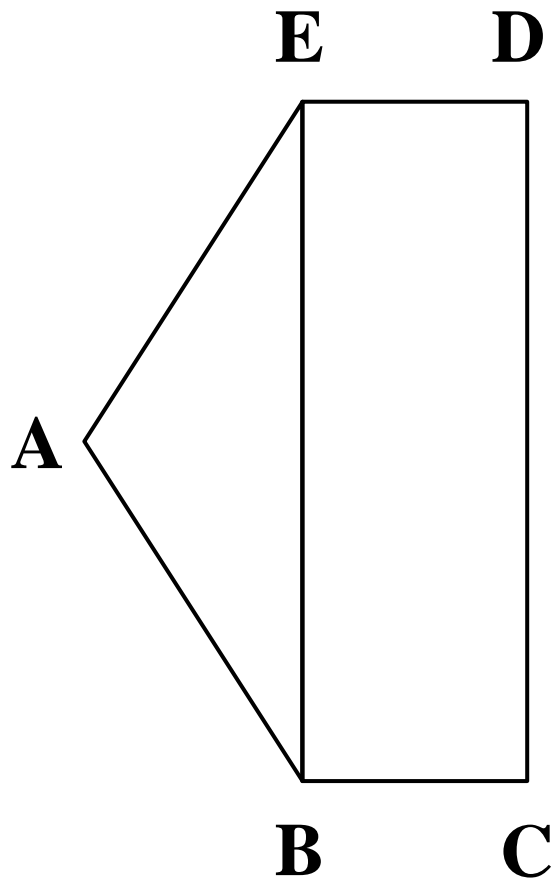
- Các em hãy tính **Betweenness** của các đỉnh còn lại (A, B, E).



4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Betweenness centrality

- Ví dụ 2: Tính $c_{bet}(A)$, $c_{bet}(B)$.



Đỉnh bắt đầu	Đỉnh đích	Đỉnh trung gian	Đường đi
A	B		AB
A	C	B	ABC
A	D	E	AED
A	E		AE
B	C		BC
B	D	C / E	BCD / BED
B	E		BE
C	D		CD
C	E	B / D	CBE / CDE
D	E		DE

4. ĐỘ ĐO DỰA TRÊN ĐƯỜNG ĐI

Betweenness centrality

▪ Ví dụ 2: Tính $c_{bet}(A)$, $c_{bet}(B)$.

▪ $\sigma_{yz} = 12$

▪ $\sigma_{yz}(A) = 0$

▪ $\sigma_{yz}(B) = 1 + \frac{1}{2} = 1.5$

$\Rightarrow c_{bet}(A) = 0$

$\Rightarrow c_{bet}(B) = 0.125$

Đỉnh bắt đầu	Đỉnh đích	Đỉnh trung gian	Đường đi
A	B		AB
A	C	B	ABC
A	D	E	AED
A	E		AE
B	C		BC
B	D	C / E	BCD / BED
B	E		BE
C	D		CD
C	E	B / D	CBE / CDE
D	E		DE



Q & A