

25 YEARS ANNIVERSARY  
SOICT

ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

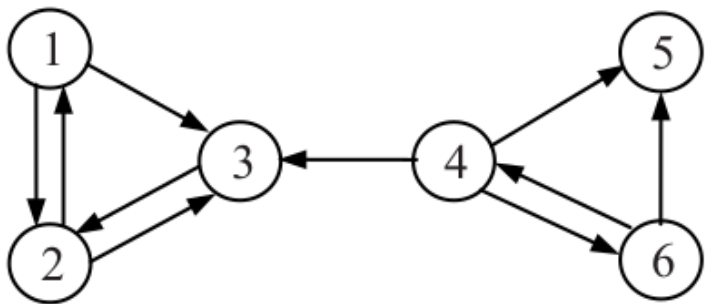
# Nhập môn Khoa học dữ liệu (IT4142)

# Nội dung môn học

- Lecture 1: Tổng quan về Khoa học dữ liệu
- Lecture 2: Thu thập và tiền xử lý dữ liệu
- Lecture 3: Làm sạch và tích hợp dữ liệu
- Lecture 4: Phân tích và khám phá dữ liệu
- Lecture 5: Trực quan hoá dữ liệu
- Lecture 6: Trực quan hoá dữ liệu đa biến
- Lecture 7: Học máy
- Lecture 8: Phân tích dữ liệu lớn
- Lecture 9: Báo cáo tiến độ bài tập lớn và hướng dẫn
- **Lecture 10+11: Phân tích một số kiểu dữ liệu**
- Lecture 12: Đánh giá kết quả phân tích

# Liên kết và Siêu văn bản

- Câu hỏi
  - Các liên kết (link) có biểu thị giá trị của các trang web được trỏ tới không? Liệu chúng có hữu dụng để xếp hạng web?
  - Khả năng trang chủ SOICT trỏ tới một trang web về Toán?



- Các miền ứng dụng
  - Web
  - Email
  - Mạng xã hội, ...

## Social network

From Wikipedia, the free encyclopedia

*This article is about the theoretical concept as used in the social and behavioral sciences. For social networking sites, see [Social networking service](#). For the 2010 movie, see [The Social Network](#). For other uses, see [Social network \(disambiguation\)](#).*

A **social network** is a [social structure](#) made up of a set of [social actors](#) (such as [individuals](#) or organizations), sets of [dyadic ties](#), and other [social interactions](#) between actors. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a

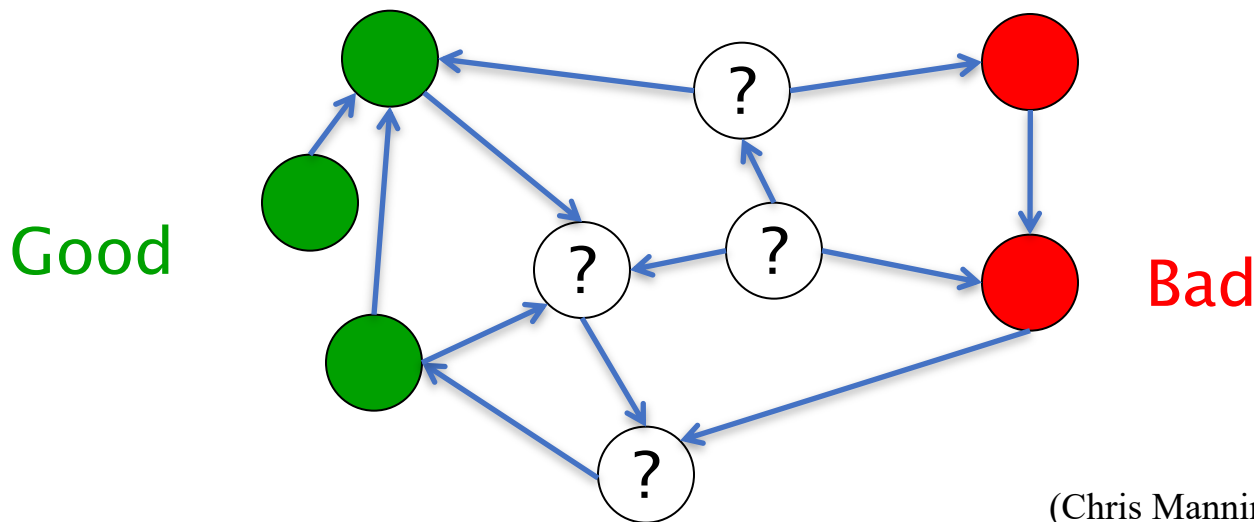
Part of a series on

**Sociology**



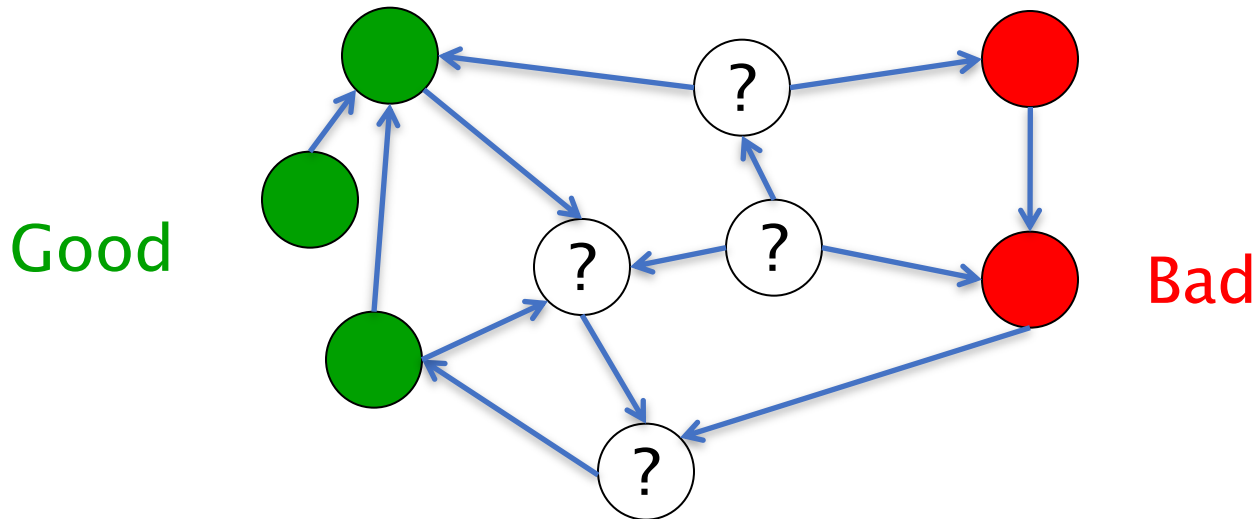
# Links có mặt khắp nơi

- Các nguồn về độ xác thực (authenticity) và tin cậy (authority)
  - Mail spam – những tài khoản email nào là spam?
  - Host quality – nguồn nào thường tệ?
  - Logs cuộc gọi, ...
- The **Good**, The **Bad** and The Unknown



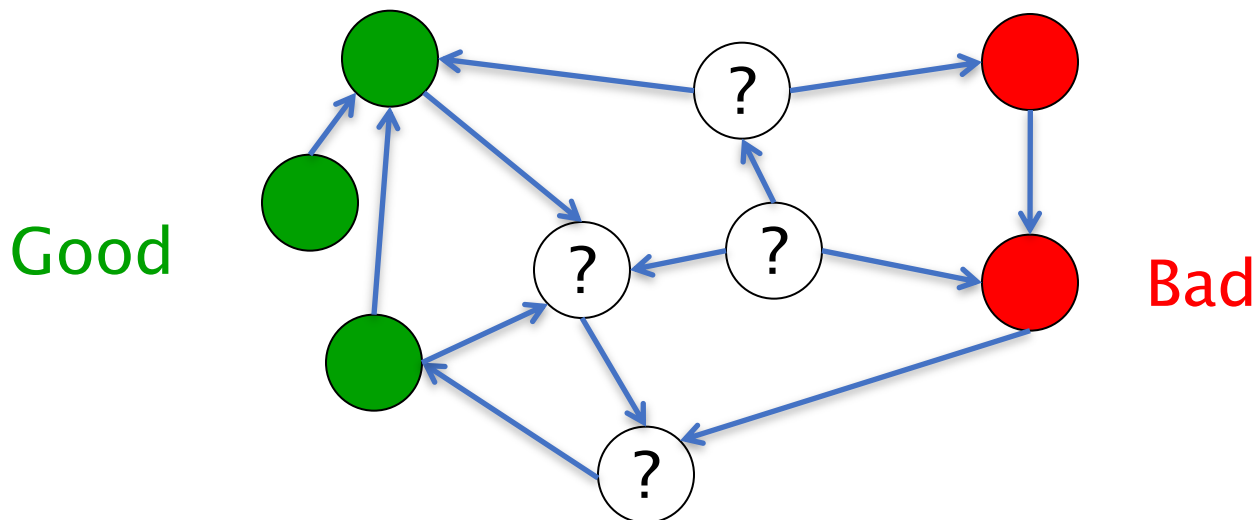
# Ví dụ: Good/Bad/Unknown

- The Good, The Bad and The Unknown
  - Đỉnh tốt sẽ không trở tới Đỉnh tồi
  - Còn các kết hợp khác đều hợp lý



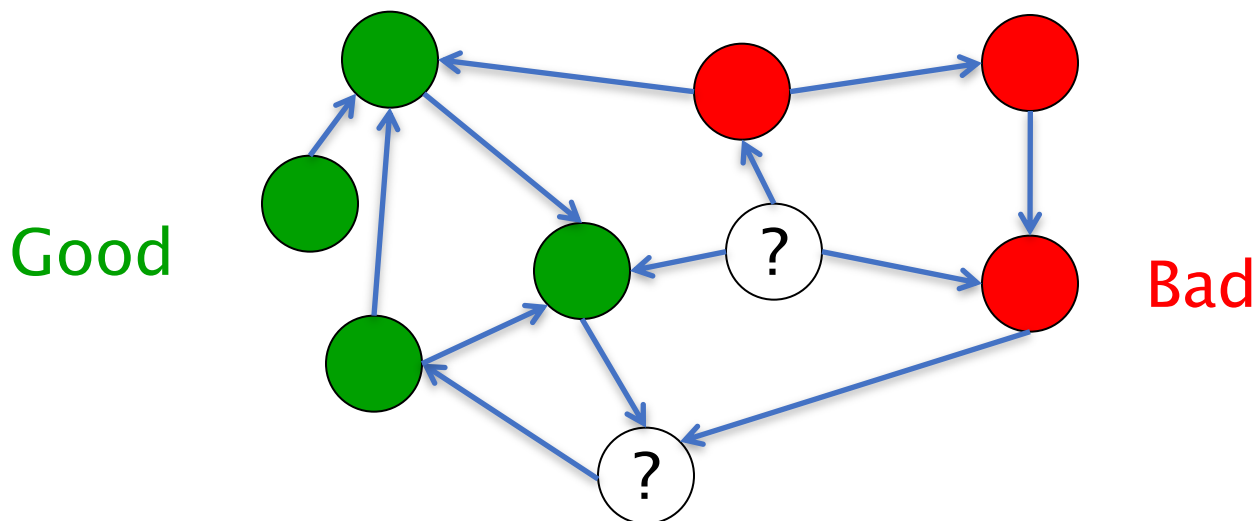
# Logic đơn giản

- **Đỉnh tốt** sẽ không trở tới **Đỉnh tồi**
  - Nếu bạn trở tới một **Đỉnh tồi**, bạn cũng **tồi**
  - Nếu một **Đỉnh tốt** trở tới bạn, bạn là **tốt**



# Logic đơn giản

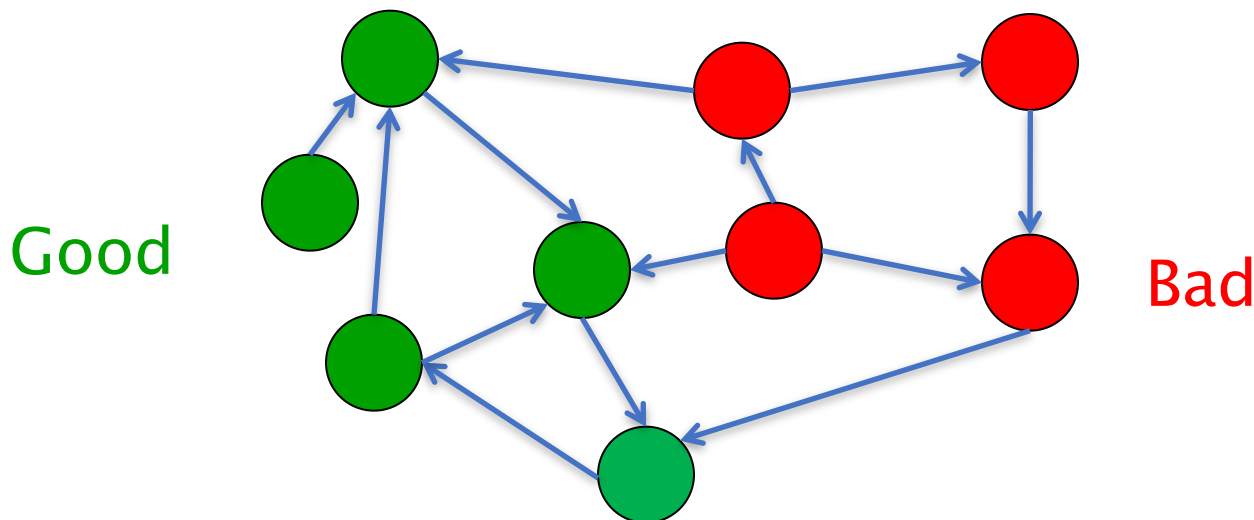
- **Đỉnh tốt** sẽ không trở tới **Đỉnh tồi**
  - Nếu bạn trở tới một **Đỉnh tồi**, bạn cũng **tồi**
  - Nếu một **Đỉnh tốt** trở tới bạn, bạn là **tốt**





# Logic đơn giản

- **Đỉnh tốt** sẽ không trở tới **Đỉnh tồi**
  - Nếu bạn trở tới một **Đỉnh tồi**, bạn cũng **tồi**
  - Nếu một **Đỉnh tốt** trở tới bạn, bạn là **tốt**



# Nhiều nhu cầu phân tích liên kết

- Phát hiện cộng đồng trong mạng xã hội
  - Mỗi cộng đồng sẽ chứa một số người mà có sở thích/hành vi giống nhau
- Thói quen mua sắm
  - Những khách hàng mà có bạn hay chi tiêu nhiều thì cũng hay chi tiêu nhiều
- Phân tích trích dẫn
  - Phát hiện các nghiên cứu ảnh hưởng lớn từ các trích dẫn

■ ...

## The reusable holdout: Preserving validity in adaptive data analysis

Cynthia Dwork,<sup>1\*</sup> Vitaly Feldman,<sup>2\*</sup> Moritz Hardt,<sup>3\*</sup> Toniann Pitassi,<sup>4\*</sup> Omer Reingold,<sup>5\*</sup> Aaron Roth<sup>6\*</sup>

Misapplication of statistical data analysis is a common cause of spurious discoveries in scientific research. Existing approaches to ensuring the validity of inferences drawn from data assume a fixed procedure to be performed, selected before the data are examined. In common practice, however, data analysis is an intrinsically adaptive process, with new analyses generated on the basis of data exploration, as well as the results of previous analyses on the same data. We demonstrate a new approach for addressing the challenges of adaptivity based on insights from privacy-preserving data analysis. As an application, we show how to safely reuse a holdout data set many times to validate the results of adaptively chosen analyses.

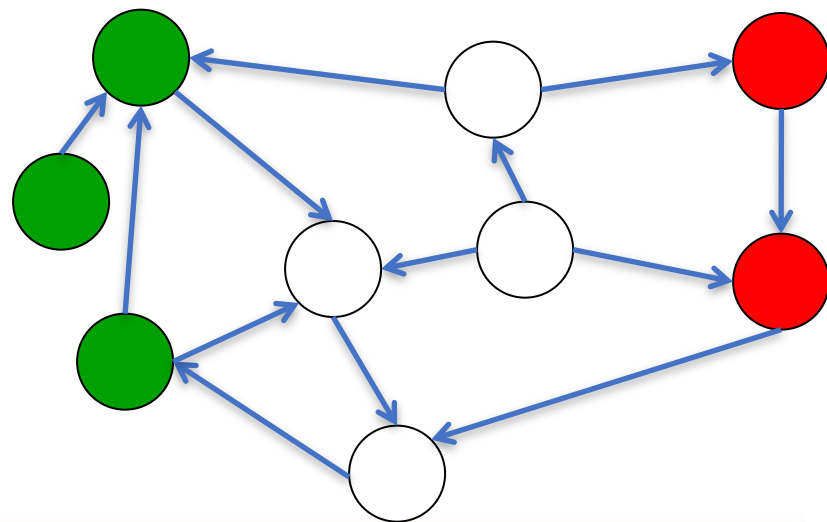
**T**hroughout the scientific community there is a growing recognition that claims of statistical significance in published research are frequently invalid. There has been a great deal of effort to understand and propose mitigations for this problem, largely focusing on statistical methods for controlling the false discovery rate in multiple hypothesis testing (1). However, the statistical inference theory surrounding this body of work assumes that a fixed proce-

of these methods focus on a single round of adaptivity—such as variable selection followed by regression on selected variables or model selection followed by testing—and are optimized for specific inference procedures [the literature is too vast to adequately cover here, but see chapter 7 in (5) for a starting point]. There are also procedures for controlling false discovery in a sequential setting where tests arrive one-by-one (6–8). However, these results crucially depend on

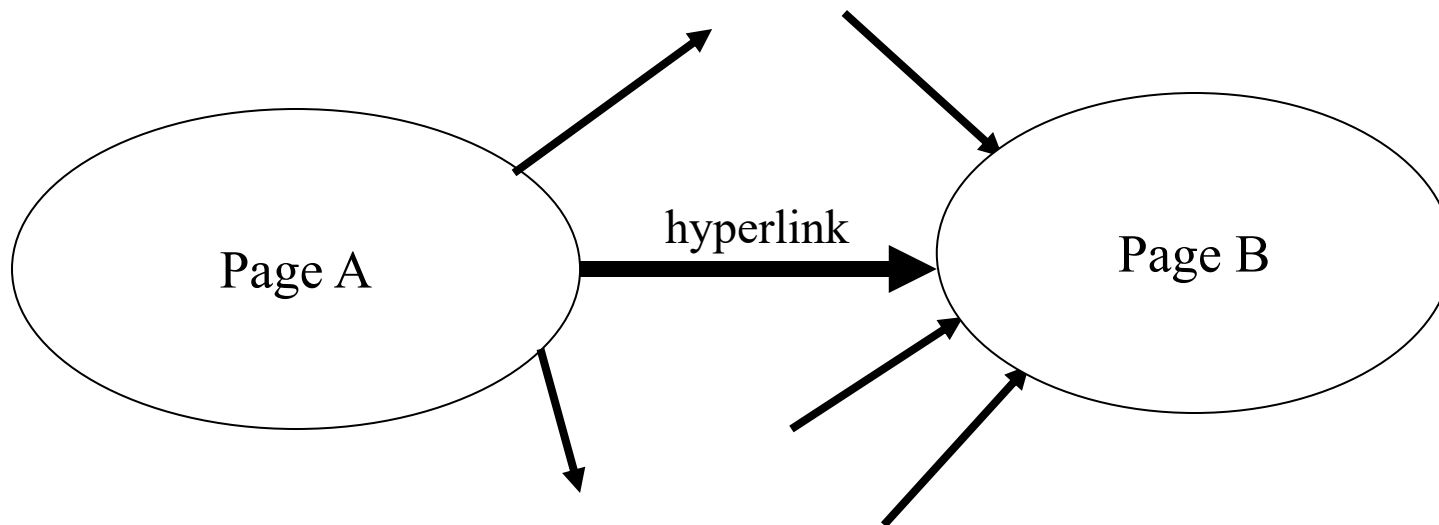
Trích dẫn đến các nghiên cứu khác

# Links và đồ thị

- **Đỉnh (vertex, node):** một đối tượng nào đó
  - VD: một người, trang web, tài liệu, công ty, ...
- **Cạnh (edge):** một liên kết (có hướng) từ đỉnh này tới đỉnh khác
- **Đồ thị (graph):**  $G = (V, E)$ 
  - V: tập tất cả các đỉnh
  - E: tập tất cả các cạnh



# World Wide Web là một đồ thị



**Giả thuyết:** mỗi liên kết giữa hai trang web có thể cung cấp minh chứng về độ tin cậy (chất lượng)

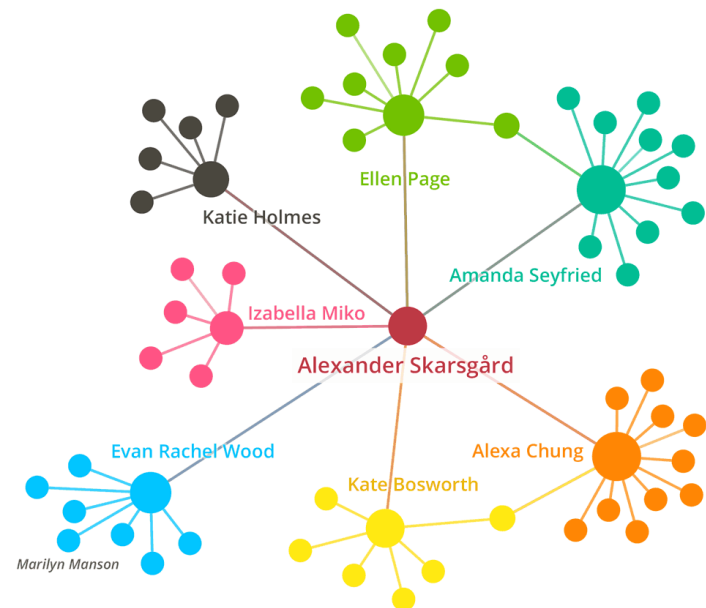
# Vài bài toán phân tích liên kết

- Xếp hạng các đỉnh (Vertex ranking)
- Phát hiện cộng đồng (Community detection)
- Phân loại đỉnh (Node classification)
- Phán đoán liên kết (Link prediction)

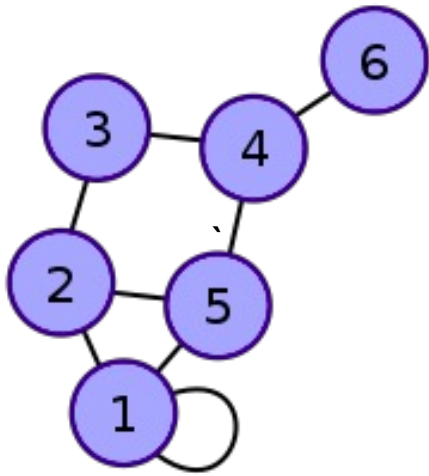
# Phân tích độ trung tâm

# Độ trung tâm (Centrality)

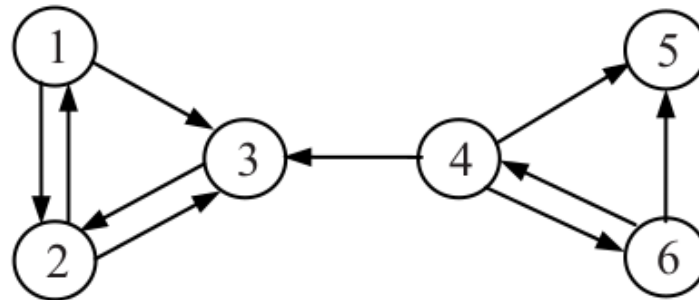
- Những đỉnh nào quan trọng?
  - Cần một độ đo
- Những đặc trưng nào mô tả một đỉnh quan trọng?
  - Độ trung tâm (Centrality)
- Ứng dụng:
  - Phát hiện những người có ảnh hưởng nhất trên mạng
  - Phát hiện những vị trí hạ tầng cốt lõi trên Internet hoặc mạng cục bộ
  - Phát hiện những điểm/nguồn siêu lây bệnh
  - ...



# Các loại đồ thị



a) Đồ thị vô hướng

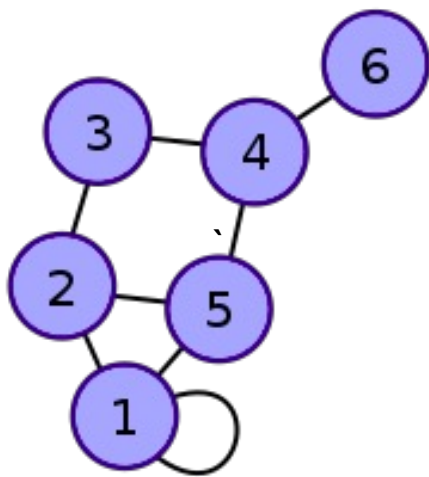


b) Đồ thị có hướng



# Ma trận kề

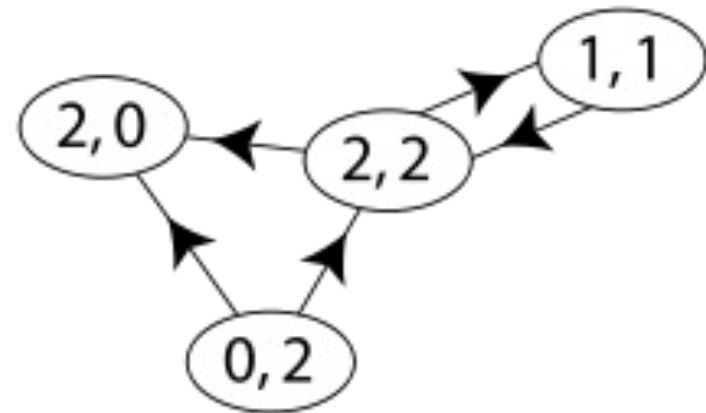
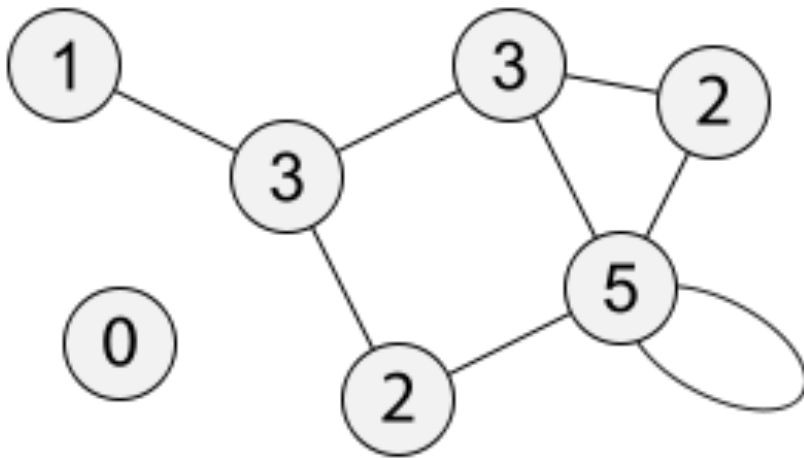
$$A(i, j) = \begin{cases} 1 & \text{nếu cạnh (i,j) tồn tại} \\ 2 & \text{nếu } i = j \text{ và có vòng lặp tại đỉnh } i \\ 0 & \text{ngược lại} \end{cases}$$



$$\begin{pmatrix} 2 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

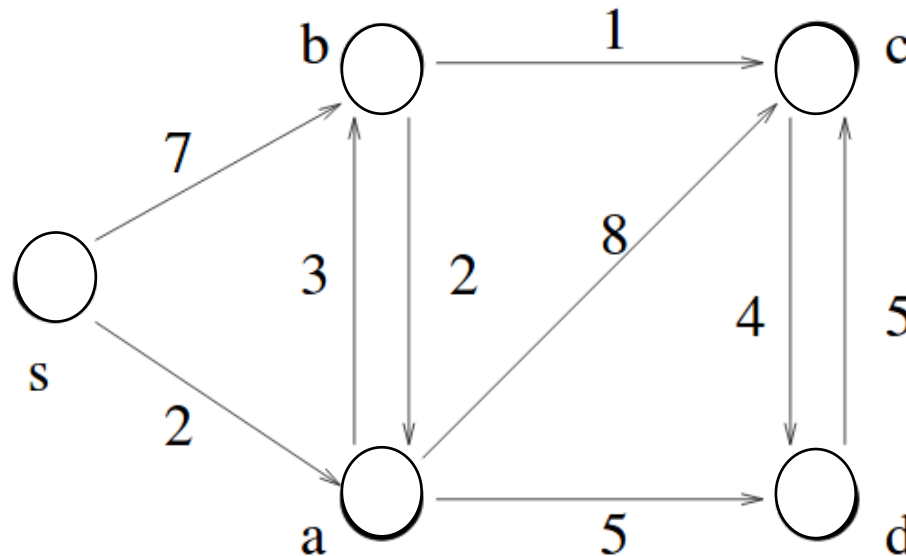
# Bậc của đỉnh (Vertex degree)

- $d_i(i)$  = số đỉnh trỏ tới  $i$
- $d_o(i)$  = số đỉnh mà  $i$  trỏ tới



# Đồ thị có trọng số

- Mỗi cạnh có một trọng số riêng
- Một đồ thị có thể được biểu diễn bằng một ma trận trọng số  $A$ 
  - $A(i, j) = 0$  nghĩa là không có cạnh nối đỉnh  $i$  tới  $j$
  - $A(i, j) \neq A(j, i)$  có thể diễn ra

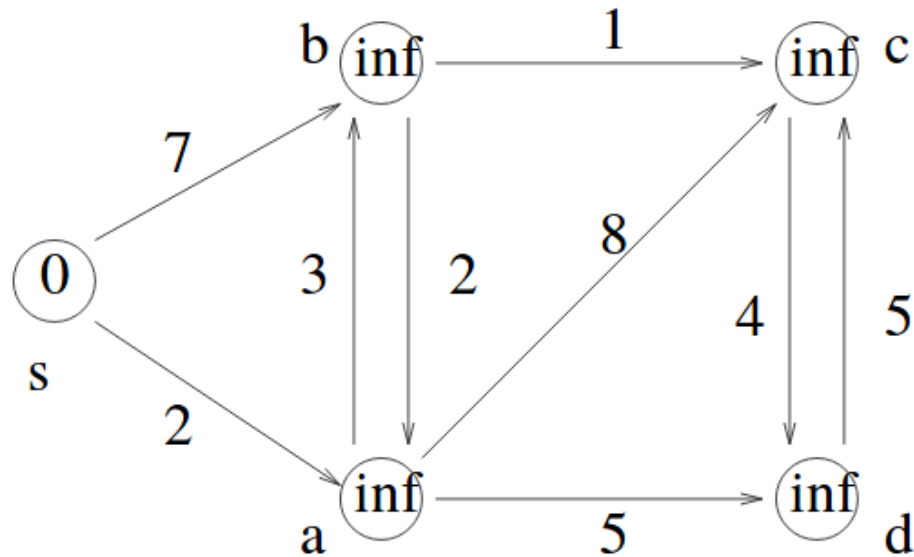


# Thuật toán Dijkstra

- Tìm đường đi ngắn nhất từ một đỉnh  $s$  tới các đỉnh còn lại của đồ thị
- $d(v)$ : Khoảng cách từ đỉnh  $s$  tới đỉnh  $v$ 
  - B1: Khởi tạo  $d(s) = 0$ ;  $d(v) = \infty$*
  - B2: Sắp xếp các đỉnh  $v$  theo một trật tự xác định trên hàng đợi  $Q$*
  - B3: Lấy một đỉnh  $u$  thuộc hàng đợi  $Q$  và cập nhật khoảng cách  $d(v)$  (nếu cần) với mỗi đỉnh  $v$  liền kề với  $u$*

*Quay lại B2 cho đến khi xử lý hết các đỉnh*

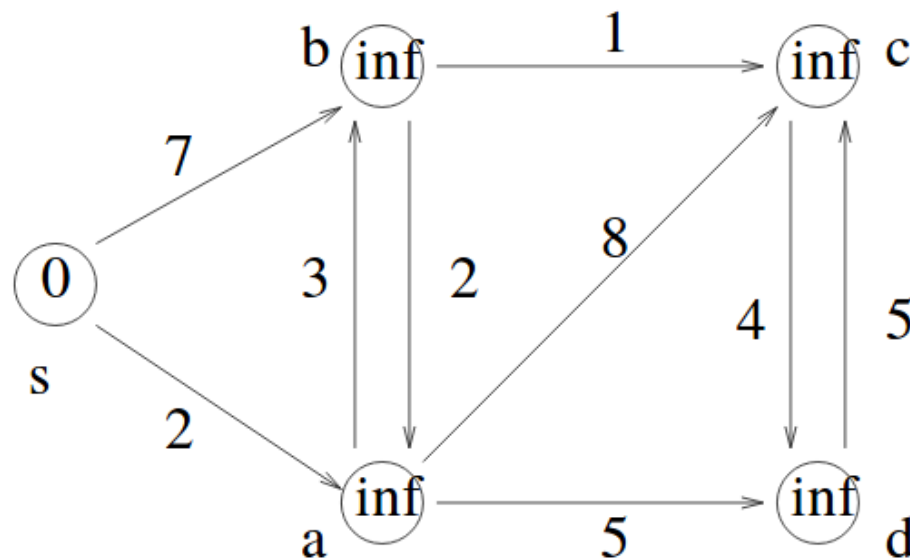
# Dijkstra: Ví dụ



# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	$\infty$	$\infty$	$\infty$	$\infty$
$pred[v]$	nil	nil	nil	nil	nil
$color[v]$	W	W	W	W	W

$v$	s	a	b	c	d
$d[v]$	0	$\infty$	$\infty$	$\infty$	$\infty$
					Q

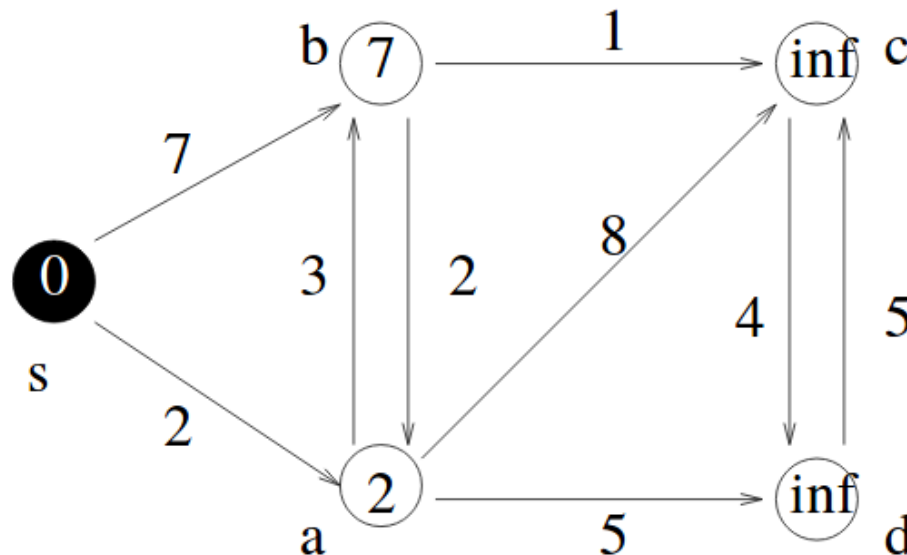


# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	2	7	$\infty$	$\infty$
$pred[v]$	nil	s	s	nil	nil
$color[v]$	B	W	W	W	W

$v$	a	b	c	d
$d[v]$	2	7	$\infty$	$\infty$

Q

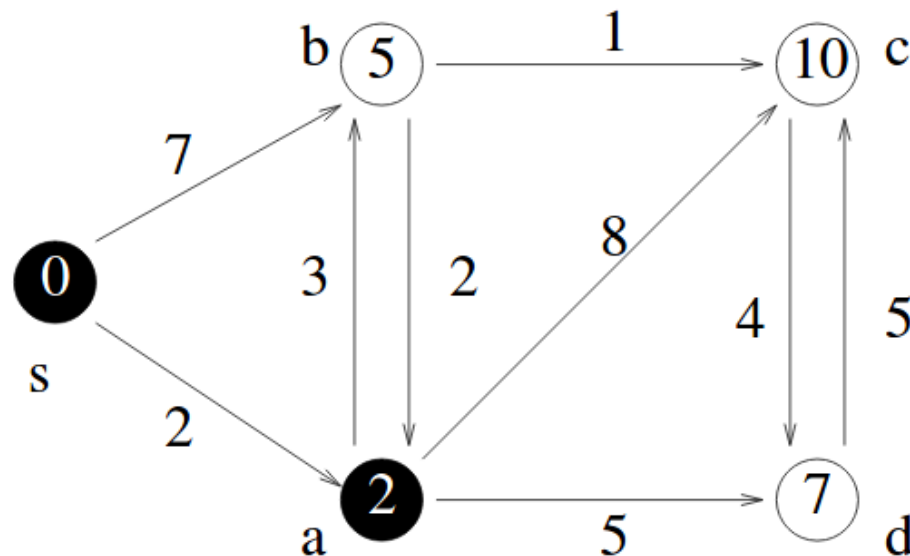


# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	2	5	10	7
$pred[v]$	nil	s	a	a	a
$color[v]$	B	B	W	W	W

$v$	b	c	d
$d[v]$	5	10	7

Q



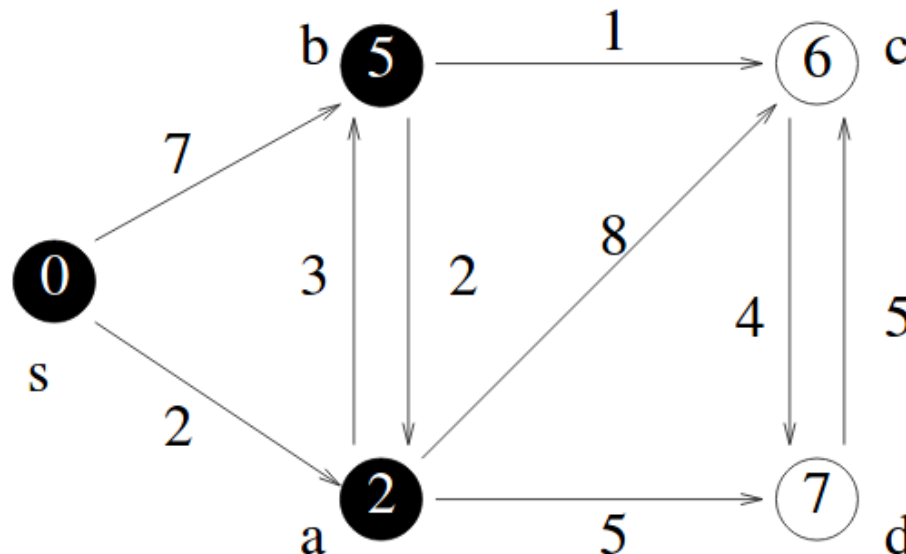


# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	2	5	6	7
$pred[v]$	nil	s	a	b	a
$color[v]$	B	B	B	W	W

$v$	c	d
$d[v]$	6	7

Q

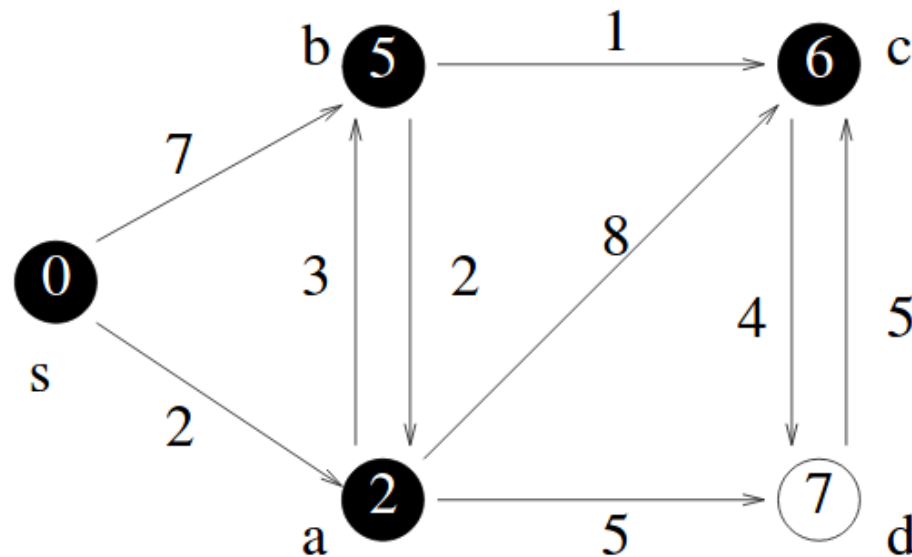


# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	2	5	6	7
$pred[v]$	nil	s	a	b	a
$color[v]$	B	B	B	B	W

$v$	d
$d[v]$	7

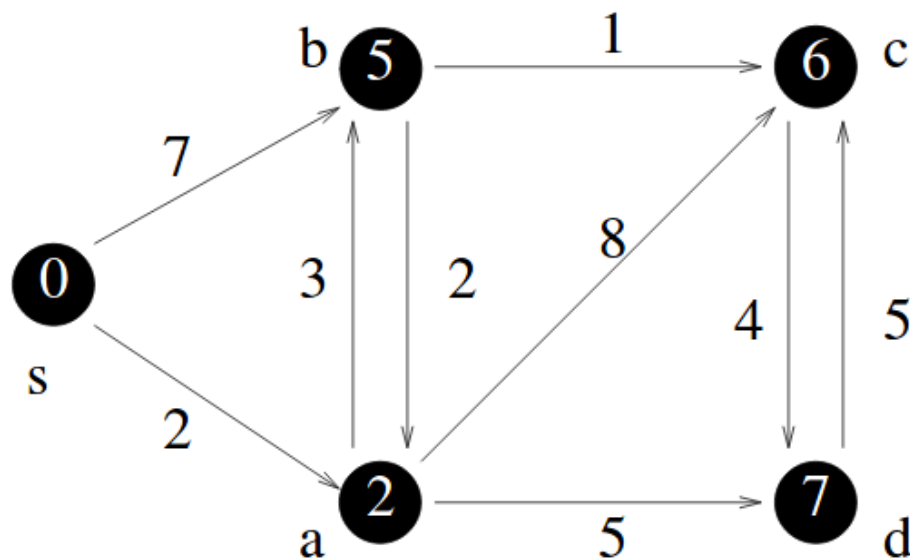
Q



# Dijkstra: Ví dụ (tiếp)

$v$	s	a	b	c	d
$d[v]$	0	2	5	6	7
$pred[v]$	nil	s	a	b	a
$color[v]$	B	B	B	B	B

$$Q = \emptyset$$



# Closeness centrality

- Giả sử ta đã có cách tính khoảng cách (ngắn nhất) từ đỉnh  $i$  tới đỉnh  $j$  (chẳng hạn, dùng thuật toán Dijkstra)
  - $d(i,j)$ : khoảng cách (ngắn nhất) từ đỉnh  $i$  tới đỉnh  $j$
  - $n$ : tổng số đỉnh trong đồ thị
- Độ trung tâm lân cận (Closeness centrality):

$$C_C(i) = \frac{n - 1}{\sum_{j=1}^n d(i, j)}$$

- Ngược lại mức độ xa
- Nghịch đảo của trung bình khoảng cách từ đỉnh  $i$  tới các đỉnh khác

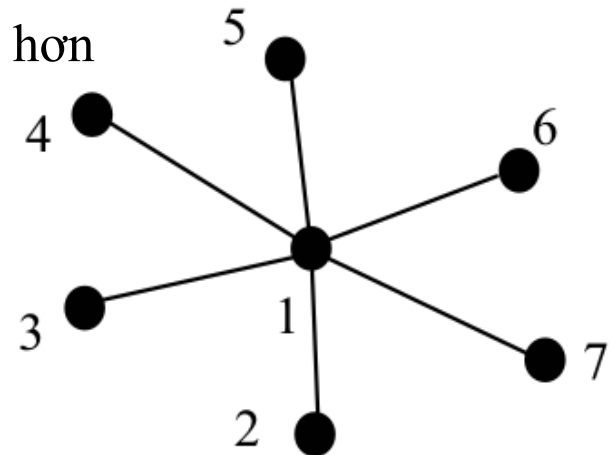
# Betweenness centrality

- Độ trung tâm trung gian (Betweenness centrality) của đỉnh  $i$ :

$$C_B(i) = \sum_{j \neq k \neq i} \frac{p_{jk}(i)}{p_{jk}}$$

- $p_{jk}(i)$ : số lượng đường đi ngắn nhất từ  $j$  tới  $k$  mà đi qua  $i$
  - $p_{jk}$ : số lượng đường đi ngắn nhất từ  $j$  tới  $k$
- Minh họa số lần mà đỉnh  $i$  làm **cầu nối** trên đường đi ngắn nhất giữa hai đỉnh khác
  - $C_B(i)$  cao hơn hàm ý đỉnh  $i$  có thể quan trọng hơn

$$\begin{aligned} C_B(1) &= 15, \\ C_B(2) &= C_B(3) = C_B(4) = 0 \\ C_B(5) &= C_B(6) = C_B(7) = 0 \end{aligned}$$



# Degree prestige

- Độ quan trọng theo bậc (Degree prestige) sử dụng bậc của một đỉnh để xem mức độ quan trọng

$$P_D(j) = \frac{d_i(j)}{n - 1}$$

- $d_i(j)$ : số lượng cạnh đi tới đỉnh  $j$
- $P_D(j)$  cao hơn hàm ý đỉnh  $j$  có thể quan trọng hơn

# Proximity prestige

$$P_P(i) = \frac{1}{n-1} \frac{|I_i|}{\sum_{j \in I_i} \frac{d(i,j)}{|I_i|}}$$

- $I_i$ : tập các đỉnh mà có thể đi tới  $i$
- $|I_i|$ : kích cỡ tập  $I_i$

# Thuật toán PageRank

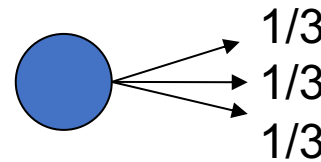


# PageRank

- PageRank được phát minh bởi Larry Page và Sergey Brin vào năm 1996,
  - (Có thể) là thành phần cốt lõi của máy tìm kiếm Google
- Việc xếp hạng dựa vào cấu trúc chung của toàn bộ đồ thị
  - Độ phổ biến của một trang web  $i$  có thể ẩn trong số lượng các trang khác liên kết tới  $i$
- Đối với các đồ thị lớn, xếp hạng có thể được xấp xỉ bằng cách dùng các bước đi ngẫu nhiên (random walk)

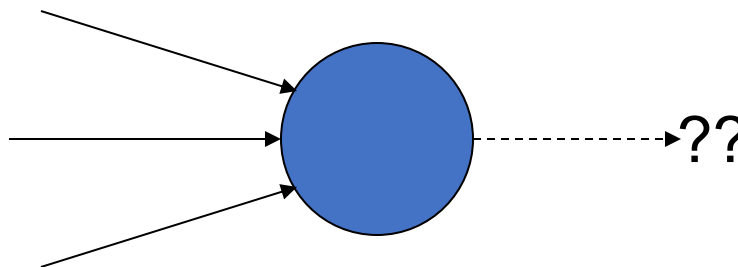
# Tính điểm trong PageRank

- Giả sử một người thực hiện một bước đi ngẫu nhiên trên các trang web:
  - Bắt đầu tại một trang nào đó
  - Tại mỗi bước, với xác suất như nhau, anh/cô ta đi tới một trang khác theo các đường links trên trang đang xem
- “Về lâu dài”, mỗi trang web có một tỉ lệ ghé thăm (visit rate)  
→ có thể coi như **điểm** của trang web đó (page's score).



# Nhưng chưa đủ

- Rất nhiều trang web chết (dead-end)
  - dead-end: không có liên kết tới trang web khác.
  - Lúc đó bước đi ngẫu nhiên sẽ bị tắc.
  - Tỷ lệ ghé thăm không còn ý nghĩa?



# Hồ biến

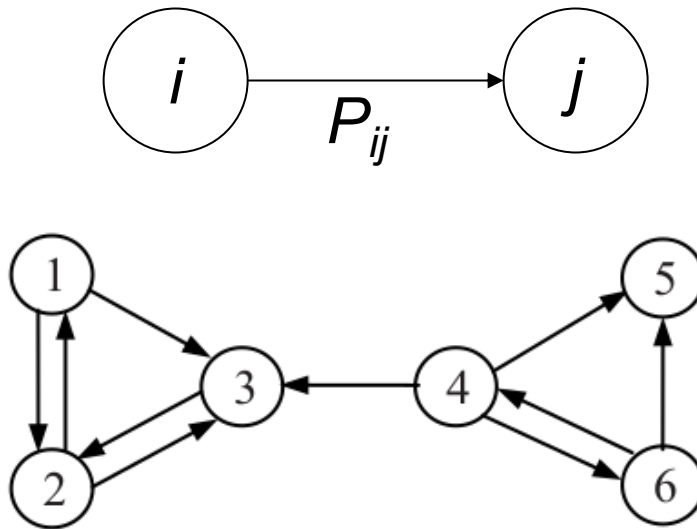
- Tại một trang web chết, nhảy ngẫu nhiên sang một trang web khác.
- Tại các trang khác, với xác suất 10%, nhảy ngẫu nhiên sang một trang web khác.
  - Với xác suất 90%, đi đến một trang web khác theo đường link đã có.
  - 10% - một tùy chọn (tham số).

# Kết quả “Hô biến”

- Không bao giờ bị tắc cục bộ.
- Tồn tại một tỉ lệ ghé thăm cho mỗi trang web.
- Làm sao để tính được tỉ lệ đó?

# Chuỗi Markov

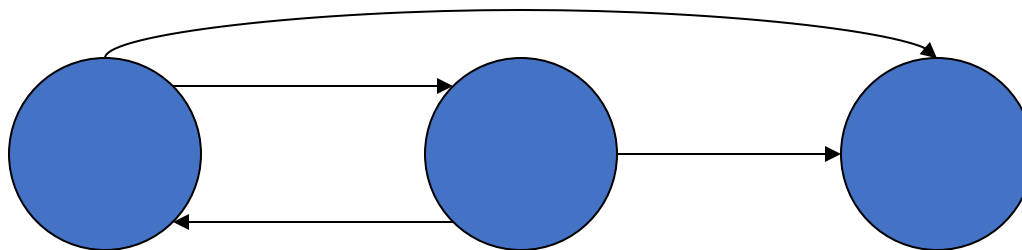
- Một chuỗi Markov bao gồm  $n$  trạng thái, và một ma trận xác suất chuyển  $\mathbf{P}$  có kích cỡ  $n \times n$ .
- **Tại mỗi bước, một người chỉ đứng ở một trạng thái nào đó.**
- Thành phần  $P_{ij}$  mô tả xác suất để di chuyển từ trạng thái  $i$  tới  $j$  ( $1 \leq i, j \leq n$ ).



$$\mathbf{P} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

# Chuỗi Markov

- Như vậy  $\sum_{j=1}^n P_{ij} = 1$ , với mỗi  $i$ .
- Chuỗi Markov là một ví dụ của “random walks.”



# Ergodic Markov

- Một chuỗi Markov được gọi là *ergodic* hay *irreducible* nếu nó cho phép di chuyển từ bất kỳ trạng thái nào tới một trạng thái khác với **xác suất dương**.
- Với các chuỗi Markov này, tồn tại duy nhất một tỉ lệ ghé thăm lâu dài cho mỗi trạng thái.
  - *Steady-state probability distribution* (phân bố xác suất trạng thái dừng)
- Dừng tỉ lệ này để đo đặc mức độ quan trọng?



# Véc tơ xác suất

- Một véc tơ xác suất (dạng hàng)  $\mathbf{x} = (x_1, \dots, x_n)$  mô tả nơi mà ta đang đứng tại một thời điểm.
  - Vd: (000...1...000) nói rằng ta đang ở trạng thái  $i$ .

1      i      n

- Về tổng quát,  $\mathbf{x} = (x_1, \dots, x_n)$  nói rằng ta đang đứng ở trạng thái  $i$  với xác suất  $x_i$ .

$$\sum_{i=1}^n x_i = 1$$

# Bước đi ngẫu nhiên

- Nếu có véctor xác suất  $\mathbf{x} = (x_1, \dots, x_n)$  ở bước hiện tại, bước tiếp theo sẽ ở đâu?
- Chú ý: hàng  $i$  trong ma trận chuyển  $\mathbf{P}$  nói về khả năng di chuyển từ  $i$  tới các trạng thái khác.
- Do đó từ  $\mathbf{x}$ , trạng thái tiếp theo được mô tả bởi  $\mathbf{xP}$ 
  - Tiếp theo nữa là  $\mathbf{xP}^2, \mathbf{xP}^3, \dots$
  - Liệu chuỗi bước đi này có hội tụ không?
  - Nếu có thì hội tụ đến đâu? (trạng thái dừng)

# Tìm điểm hội tụ

- Gọi  $\mathbf{a} = (a_1, \dots, a_n)$  là véc tơ xác suất dừng (steady-state probabilities).
- Nếu vị trí hiện tại được mô tả bởi  $\mathbf{a}$ , thì vị trí tiếp theo được mô tả bởi  $\mathbf{aP}$ .
- Do  $\mathbf{a}$  là trạng thái dừng nên  $\mathbf{a} = \mathbf{aP}$ .
- Giải phương trình này sẽ tìm được  $\mathbf{a}$ .
  - Như vậy  $\mathbf{a}$  là véc tơ riêng (eigenvector) của  $\mathbf{P}$ .  
(véc tơ riêng tương ứng với giá trị riêng lớn nhất của  $\mathbf{P}$ )

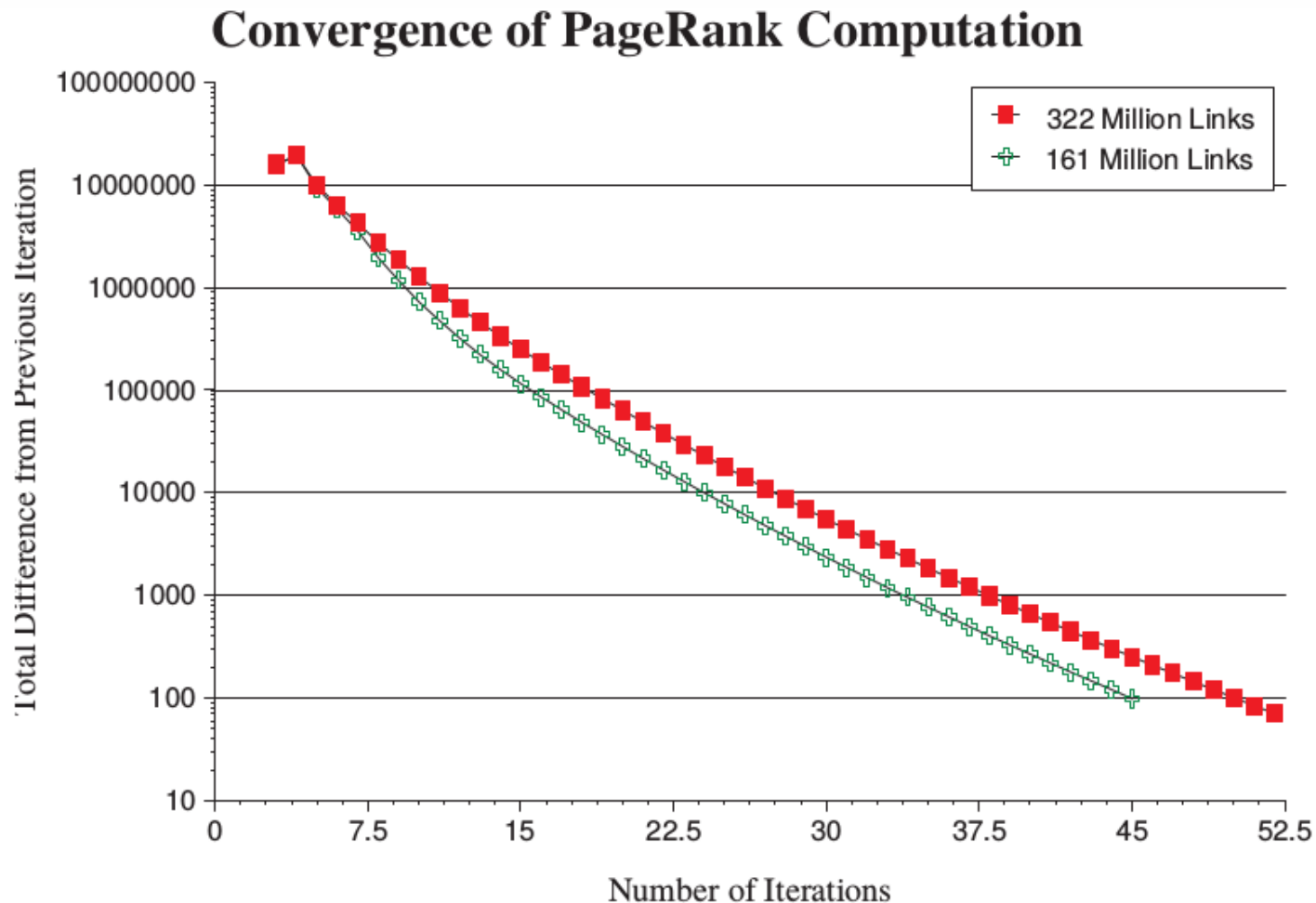
# Phương pháp mũ

- Để đảm bảo chuỗi Markov là ergodic, có thể thay  $\mathbf{P}$  bằng:

$$\hat{\mathbf{P}} = d\mathbf{P} + \frac{1-d}{n}\mathbf{1}$$

- $d$  là một tham số (**damping factor**),  $d \in [0,1]$ ;  $\mathbf{1}$  là ma trận các số 1
- Phương pháp mũ (Power method):
  - Khởi tạo  $\mathbf{x}^{(0)}$  (vị trí đầu tiên)
  - Tại bước lặp  $t$ , tính  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}\hat{\mathbf{P}}$
  - Cho đến khi  $\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\| < \epsilon$  (ngưỡng sai khác  $\epsilon$ )

# Tốc độ hội tụ



# Ứng dụng: Tìm kiếm Web



university top 100



All

News

Images

Videos

More

Tools

About 2,800,000,000 results (1.45 seconds)

## Colleges/Universities

From sources across the web



Stanford University  
Stanford



Harvard University  
Cambridge



University of Cambridge  
Cambridge



Massachusetts Institute of Technol...  
Cambridge



University of Oxford  
Oxford



California Institute of Technology  
Pasadena



University of California, Berkeley  
Berkeley



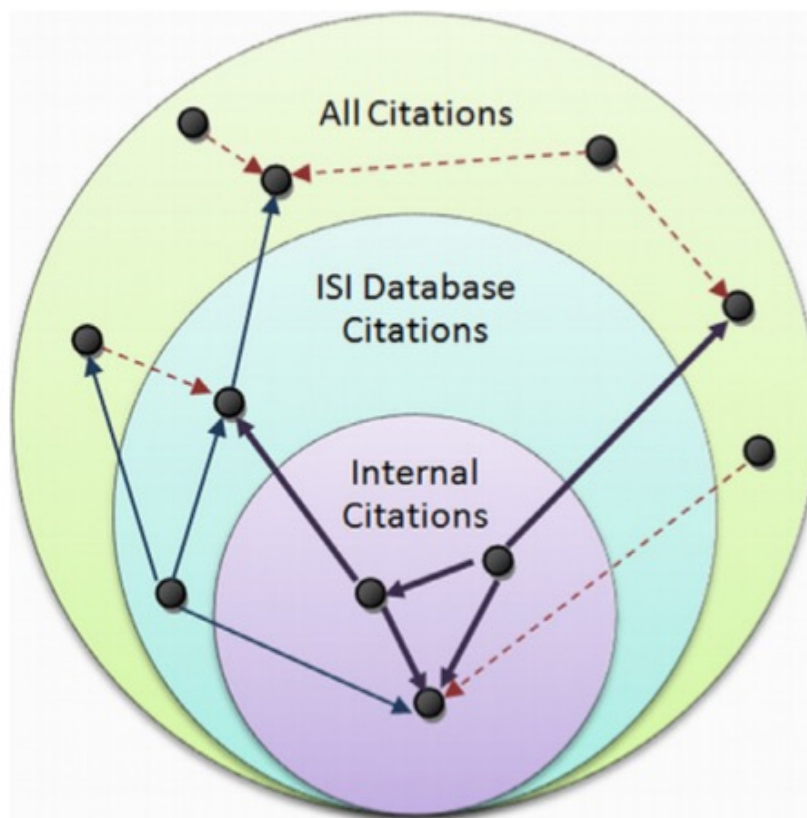
The University of Chicago  
Chicago



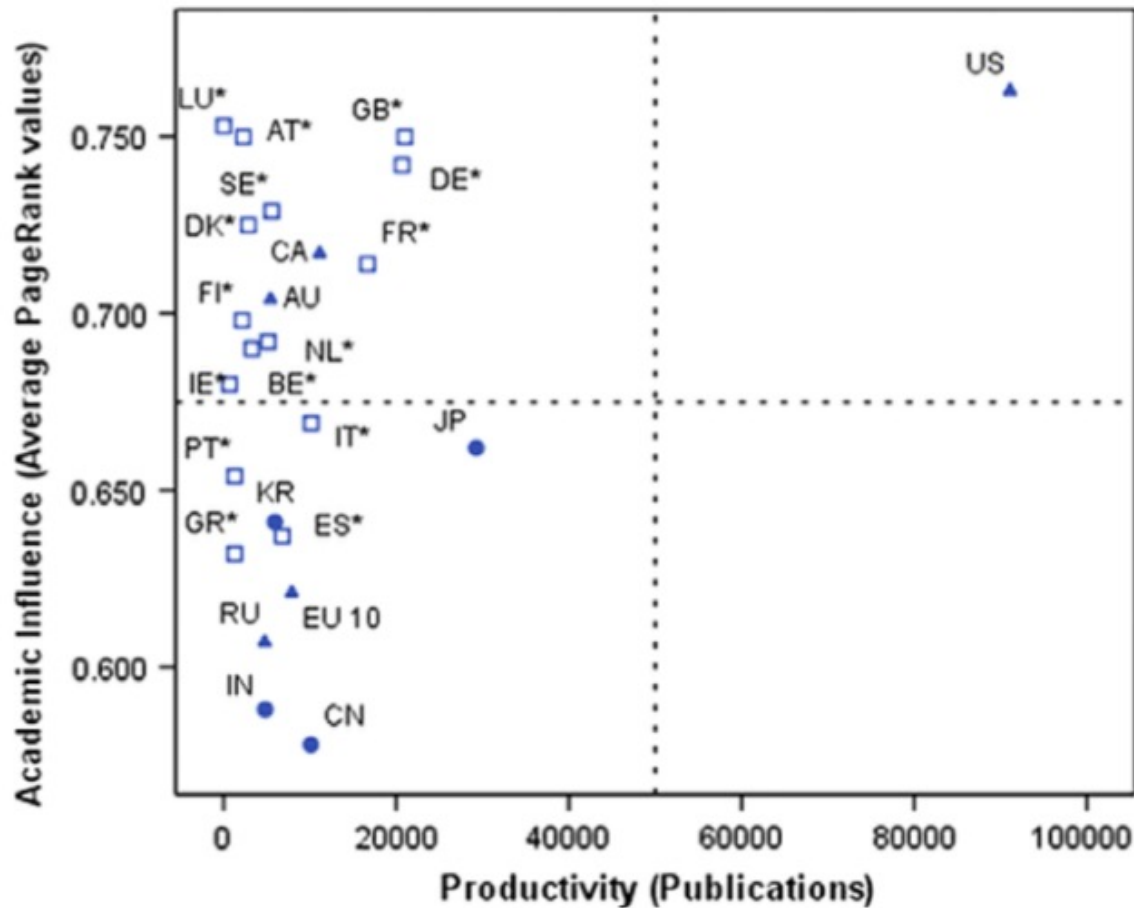
University College London  
London

# Ứng dụng: Phân tích trích dẫn

Guan et al. 2008. “*Bringing Page-Rank to the Citation Analysis*”



# Ứng dụng: Phân tích trích dẫn (tiếp)









# Thuật toán HITS

# HITS: Hypertext Induced Topic Search

Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46, no. 5 (1999): 604-632.

	Spam filtering	Query relevance	Execution
HITS			Online
PageRank			Offline

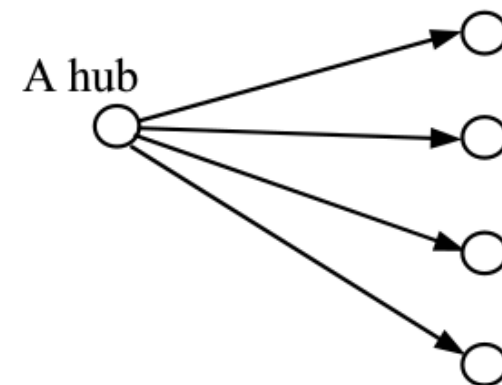
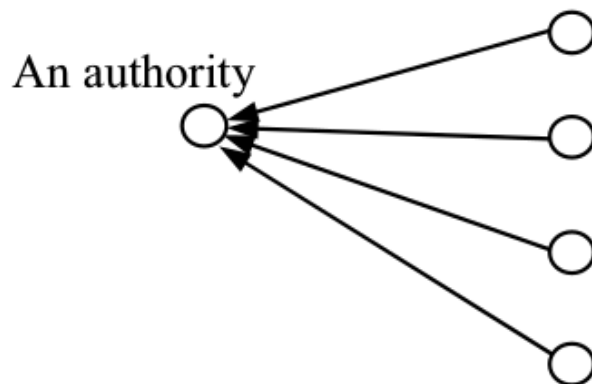
# Hubs and Authorities

- Một trang hub tốt đối với một chủ đề sẽ *trở tới* nhiều trang tin cậy (authoritative pages) về chủ đề đó.
- Một trang tin cậy (authority page) về một chủ đề *được trở tới* từ nhiều hub tốt.

# Authority/Hub

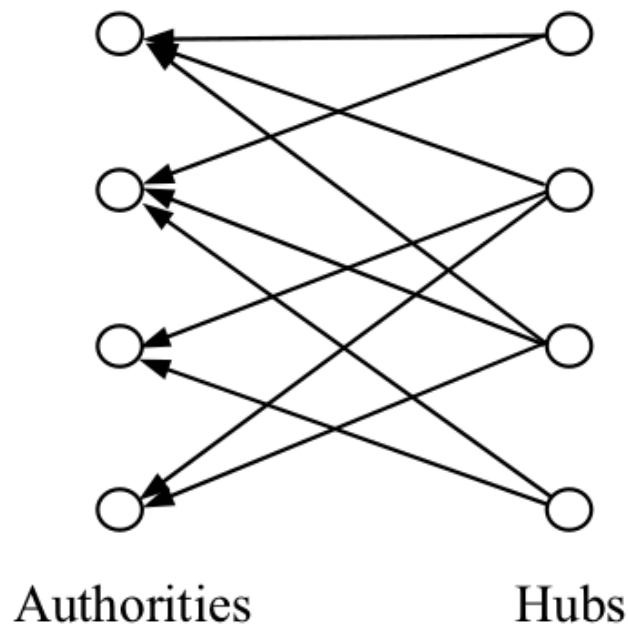
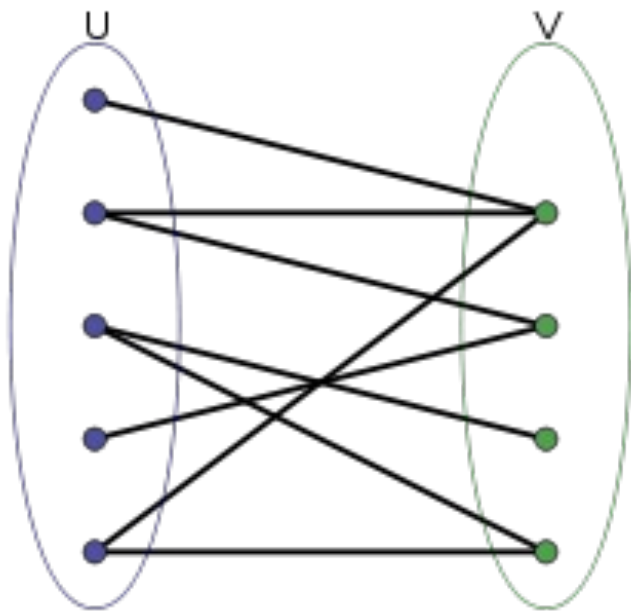
Authority: Trang được trỏ tới nhiều

Hub: Trang trỏ tới nhiều trang khác



# Bigraph

- Các đỉnh chia thành hai tập không giao nhau
- Mỗi cạnh đều nối hai đỉnh thuộc hai tập



# Thuật toán HITS

Đầu vào: Câu truy vấn  $q$

Đầu ra: Điểm authority và hub của các trang liên quan đến  $q$

Thuật toán:

1 - Truy hồi thông tin

2 - Mở rộng đồ thị

3 - Tính thứ hạng

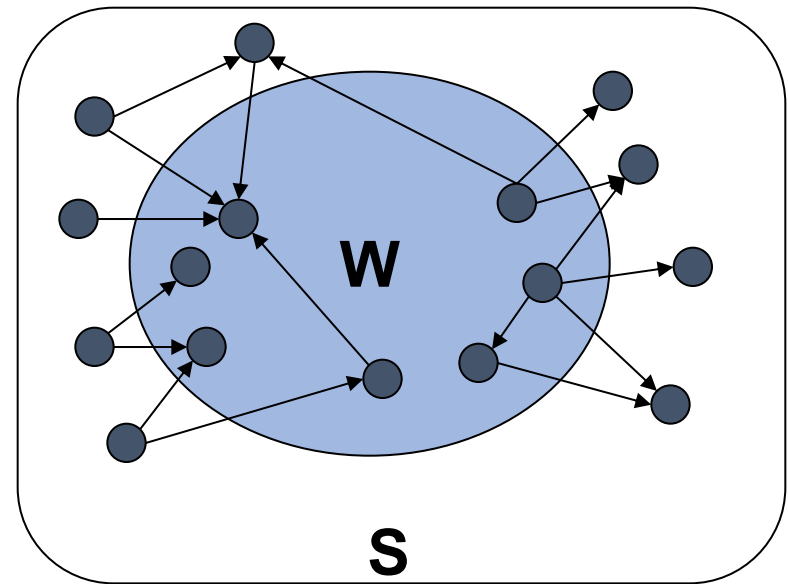
# 1-Truy hồi thông tin

- Đưa  $q$  vào máy tìm kiếm (Google, Coccoc, ...) và lấy về tập gốc  $\mathbf{W}$  gồm  $k$  trang liên quan nhất đến  $q$  (ví dụ  $k = 200$ )

## 2- Mở rộng đồ thị

Từ tập gốc  $W$ , tạo tập cơ sở  $S$  bằng cách

- Với mỗi trang  $p$  trong  $W$ 
  - Bổ sung các trang mà  $p$  trỏ tới
  - Bổ sung các trang trỏ tới  $p$





# 3- Tính thứ hạng

- Từ tập  $S$ , xác định tập nhỏ các hubs và authorities  
→ Thuật toán lặp

Điểm của Authority:  $a$

Điểm của Hub:  $h$

Cho trước đồ thị  $G = (V, E)$ , các điểm được tính như sau

$$\hat{a}(i) = \sum_{(j,i) \in E} h(j); \quad a(i) = \frac{\hat{a}(i)}{\|\hat{a}\|_1}$$

$$\hat{h}(i) = \sum_{(i,j) \in E} a(j); \quad h(i) = \frac{\hat{h}(i)}{\|\hat{h}\|_1}$$

# 3- Ranking (cont)

## HITS-Iterate( $G$ )

$\mathbf{a}_0 \leftarrow \mathbf{h}_0 \leftarrow (1, 1, \dots, 1);$

$k \leftarrow 1$

**Repeat**

$\mathbf{a}_k \leftarrow \mathbf{L}^T \mathbf{L} \mathbf{a}_{k-1};$

$\mathbf{h}_k \leftarrow \mathbf{L} \mathbf{L}^T \mathbf{h}_{k-1};$

$\mathbf{a}_k \leftarrow \mathbf{a}_k / \|\mathbf{a}_k\|_1; \quad // \text{normalization}$

$\mathbf{h}_k \leftarrow \mathbf{h}_k / \|\mathbf{h}_k\|_1; \quad // \text{normalization}$

$k \leftarrow k + 1;$

**until**  $\|\mathbf{a}_k - \mathbf{a}_{k-1}\|_1 < \varepsilon_a$  and  $\|\mathbf{h}_k - \mathbf{h}_{k-1}\|_1 < \varepsilon_h;$

**return**  $\mathbf{a}_k$  and  $\mathbf{h}_k$

$$\mathbf{a} = \mathbf{L}^T \mathbf{h}$$

$$\mathbf{h} = \mathbf{L} \mathbf{a}$$

$$L_{ij} = \begin{cases} 1 & \text{nếu } (i, j) \in E \\ 0 & \text{ngược lại} \end{cases}$$



25 YEARS ANNIVERSARY  
**SOICT**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

**Thank you  
for your  
attentions!**



[soict.hust.edu.vn/](http://soict.hust.edu.vn/)



[fb.com/groups/soict](https://fb.com/groups/soict)

