

Content

1. Social networks
2. PageRank algorithm
3. Community detection

Introduction to Social Network Analysis

Social network analysis
Social networks

- World wide web
- FB, Twitter, weibo, zalo
- Wikipedia
- Collaboration network, citation network
- Cellphone network



Elon Musk • @elonmusk
Boring
Joined June 2009

Tweets 3,780 Following 48 Followers 18.3M Likes 719

Tweets **Tweets & replies** **Media**

13 Elon Musk Retweeted
Robin Seemangal • @nova_road 11h My raw video of the #SpaceX Falcon Heavy static-fire at Kennedy Space Center. Come for the cloud plumes, stay for the sound.

A French space reporter just yelled "It's like the 4th of July!"

Gõ tiếng Việt
Trợ giúp

- Tự động [F9]
- Telex (T)
- VNI (T)
- VIQR (T)
- VIQR+
- Tất [F12]

Bé dấu kiểu cũ [F7]

Dùng chính tả [F8]

Công cụ

Các liên kết đến đây
Thay đổi liên quan
Các trang đặc biệt
Liên kết thường trực
Thông tin trang
Khoán mục Wikidata

In/xuất ra

Tạo một quyền sách
Tải về dưới dạng PDF
Bản để in ra

Tài liệu khác

Wikimedia Commons
MediaWiki
Meta-Wiki
Wikispecies
Wikibooks
Wikidata
Wikiquote

Bài viết chọn lọc

Sao Kim hay **Kim tinh**, còn gọi là **sao Thái Bạch**, là hành tinh thứ hai trong Hệ Mặt Trời, tự quay quanh nó với chu kỳ 224,7 ngày. **Trái Đất**. Xếp sau **Mặt Trăng**, nó là thiên thể tự nhiên sáng nhất trong bầu trời tối, với **cấp sao biểu kiến** bằng -4,6, dù sáng đẽ tạo nên bóng trên mặt nước. Bởi vì Sao Kim là hành tinh phia trong tinh tú Trái Đất, nó không bao giờ xuất hiện trên bầu trời mà qua xa **Mặt Trời**: góc ly giác đặt cực đại bằng 47,8°. Sao Kim đạt độ sáng lớn nhất ngay sát thời điểm hoàng hôn hoặc bình minh, do vậy mà dân gian còn gọi là **sao Hỏa**, khi hành tinh này mọc lên lúc **hoàng hôn**, và **sao Mai**, khi hành tinh này mọc lên lúc **bình minh**. Sao Kim được xếp vào nhóm **hành tinh đất đá** và đối với người ta còn coi nó là "hành tinh chí em" với Trái Đất do kích cỡ, giá tốc hấp dẫn, tham số quỹ đạo gần giống với Trái Đất. Tuy nhiên, người ta đã chỉ ra rằng nó rất khác Trái Đất trên những mặt khác. Mật độ không khí trong khí quyển của nó lớn nhất trong số bốn hành tinh đất đá, thành phần chủ yếu là cacbon dioxide. Áp suất khí quyển tại bề mặt hành tinh cao gấp 92 lần so với của Trái Đất. Với nhiệt độ bề mặt trung bình bằng 735 K (462 °C), Sao Kim là hành tinh nóng nhất trong Hệ Mặt Trời. Toàn bộ bề mặt của Sao Kim là một hoang mạc khô cằn với đá và bụi và có lẽ vẫn còn núi lửa hoạt động trên hành tinh này. ([xem tiếp...](#))

Mỗi chọn: Stephen Hawking • Trần Thành Tông • Imagine (bài hát của John Lennon)

Lưu trữ • Thêm bài viết chọn lọc • Ứng cử viên

Hình ảnh chọn lọc

Bạn có biết...

- ...Albert xứ Saxe-Coburg và Gotha và hồn thể tương lai của ông khi chào đời được trả sinh bởi cùng một bà mụ?
- ...cầu thành ngũ **Ichi-go ichi-e** khuyên mọi người tránh quấy những cuộc hội ngộ nguyên là lời cảm thán của bậc thầy trà đạo **Sen no Rikyu**?
- ...nhà văn **Hugo Gernsback** được mệnh danh là cha đẻ của khoa học viễn tưởng hiện đại?
- ...nhân vật chính Debbie Reynolds với con gái bà trong phim tài liệu **Bright Lights: Starring Carrie Fisher and Debbie Reynolds** qua đời trước khi phim lên sóng?

Từ những bài viết mới của Wikipedia

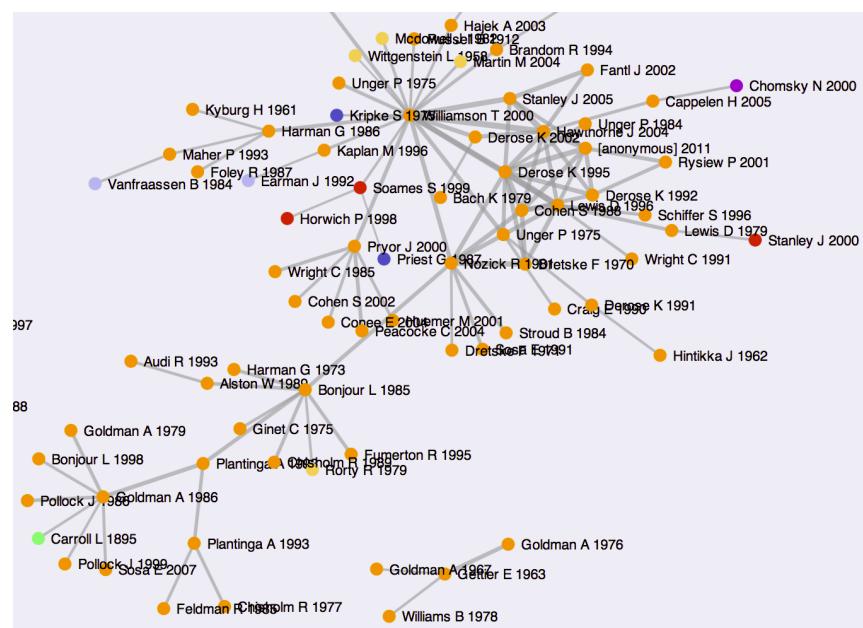
Tin tức

- **Người đẹp và thủy quái** giành chiến thắng cả bốn đề cử, trong đó có phim hay nhất tại lễ trao giải **BFCA** lần thứ 23
- Một vụ cháy xe buýt xảy ra ở khu vực tỉnh Aktobe, Kazakhstan, thiêu chết 52 người.
- Phim **Three Billboards Outside Ebbing, Missouri** giành nhiều chiến thắng nhất tại lễ trao giải **Quả cầu vàng** lần thứ 75.
- **George Weah** (hình) đắc cử **Tổng thống Liberia**.

Ngày này năm xưa

25 tháng 1: Ngày Tatiana tại Nga, ngày cử tri quốc gia tại Ấn Độ.

Cập nhật



Social network analysis Social networks

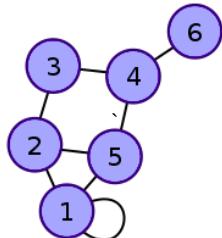
Social network analysis studies social agents and their relations (using only network typology)
– Bing Liu

- Main problems:
 - Network ranking
 - Community detection
 - Link prediction

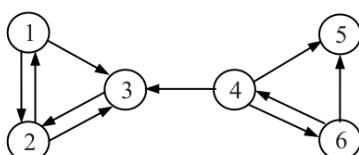
Social network analysis

Social networks

Graph fundamentals



a) Indirect graph



b) Direct graph

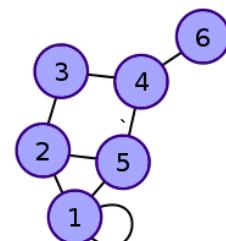
Social network analysis

Social networks

Graph fundamentals

- Adjacency matrix

- $a[i,j] = 1$ if $e(i,j)$ exists
 $= 0$ otherwise
- $a[i,i] = 2$ if $e(i,i)$ exists



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

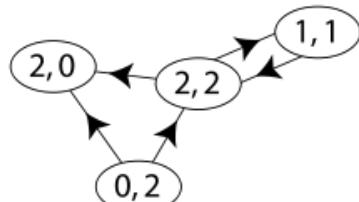
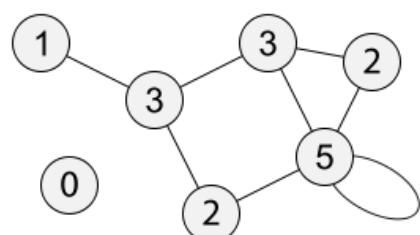
Social network analysis

Social networks

Graph fundamentals

- Node degree

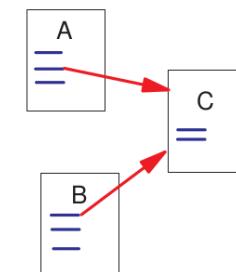
- $d_i(i)$: number of edges point to i
- $d_o(i)$: number of edges come from i



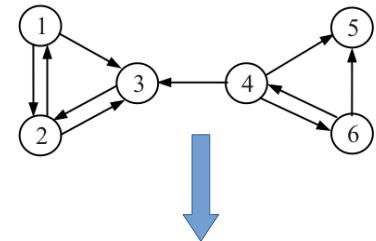
Social network analysis

PageRank

- Lawrence (Larry) Page et al. 1999. "The PageRank Citation Ranking: Bringing Order to the Web"
- 1999: 150M pages, 1.7B links



Social network analysis
PageRank
Transition matrix



$$A = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}.$$

Social network analysis
PageRank
Problem statement

$$\text{PR}(A) = (1 - d) / N + d * \sum_{B:(B,A) \text{ in } E} \text{PR}(B) / d_o(B)$$

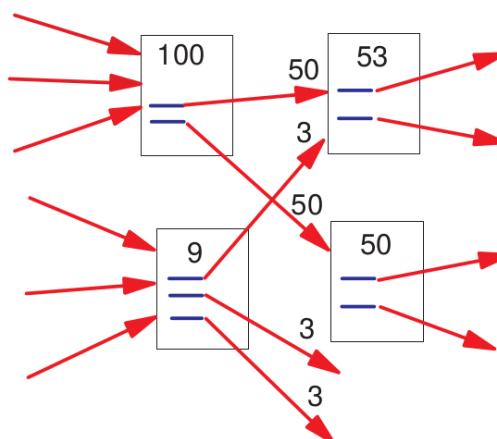
PR(A): Ranking of A
d: damping factor
N: number of vertices
(B,A) is a edge
 $d_o(B)$ is out-degree of B

Social network analysis
PageRank
Transition matrix

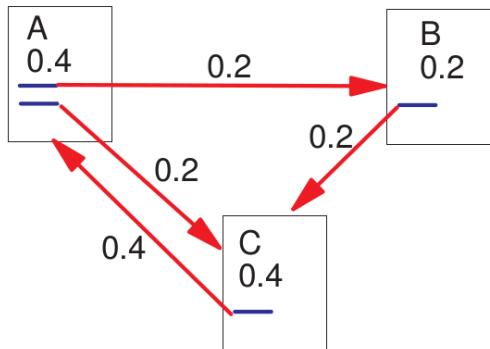
- Normalization

$$A = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} \xrightarrow{\text{Normalization}} \bar{A} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}.$$

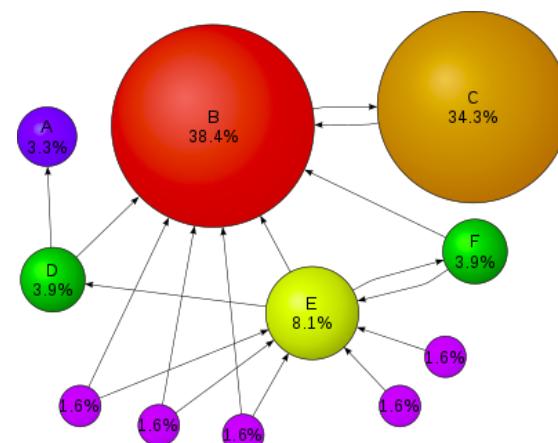
Social network analysis
PageRank
Example (d=1)



Social network analysis
PageRank
Example (d=1)



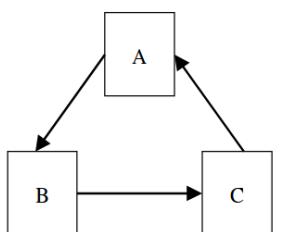
Social network analysis
PageRank
Example (d=.85)



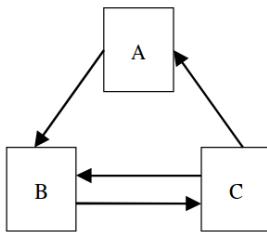
17

18

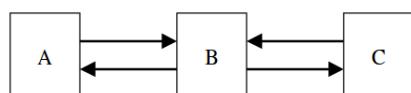
Social network analysis
PageRank
Practice (d=.7)



a)

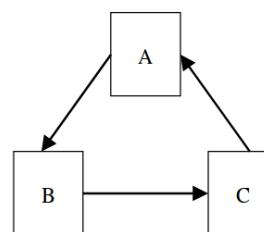


b)



c)

Social network analysis
PageRank
Practice (d=.7)



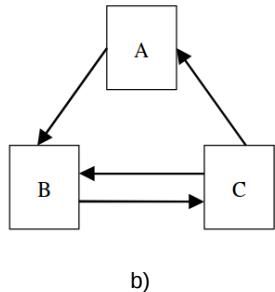
a)

$$\begin{aligned} PR(A) &= 1/3 = 0.33 \\ PR(B) &= 1/3 = 0.33 \\ PR(C) &= 1/3 = 0.33 \end{aligned}$$

19

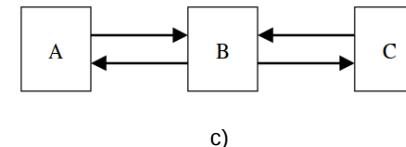
20

Social network analysis
PageRank
Practice (d=.7)



$$\begin{aligned} PR(A) &= 0.2314 \\ PR(B) &= 0.3933 \\ PR(C) &= 0.3753 \end{aligned}$$

Social network analysis
PageRank
Practice (d=.7)



$$\begin{aligned} PR(A) &= 0.2647 \\ PR(B) &= 0.4706 \\ PR(C) &= 0.2647 \end{aligned}$$

21

22

Social network analysis
PageRank
Brief implementation (in Matlab)

```
% Parameter M adjacency matrix where M_i,j represents the link from 'j' to 'i', such that for all 'j'
% sum(i, M_ij) = 1
% Parameter d damping factor
% Parameter v_quadratic_error quadratic error for v
% Return v, a vector of ranks such that v_i is the i-th rank from [0, 1]

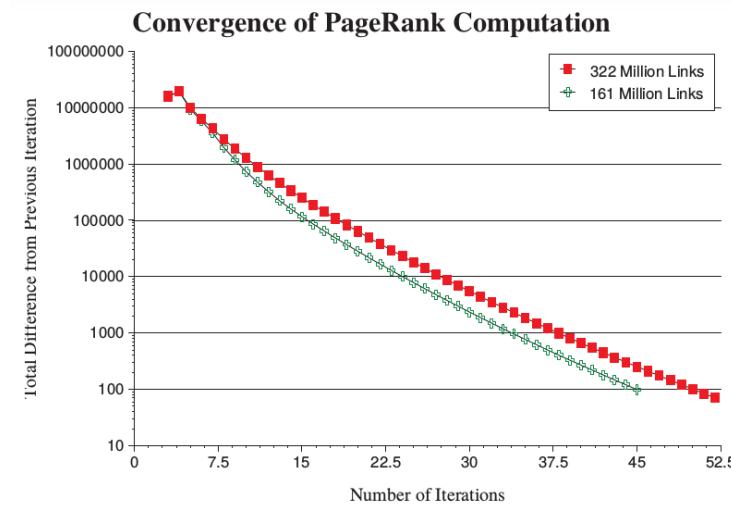
function [v] = rank2(M, d, v_quadratic_error)

N = size(M, 2); % N is equal to either dimension of M and the number of documents
v = rand(N, 1);
v = v ./ norm(v, 1); % This is now L1, not L2
last_v = ones(N, 1) * inf;
M_hat = (d .* M) + ((1 - d) / N) .* ones(N, N);

while(norm(v - last_v, 2) > v_quadratic_error)
    last_v = v;
    v = M_hat * v;
    % removed the L2 norm of the iterated PR
end

endfunction
```

Social network analysis
PageRank
Convergence



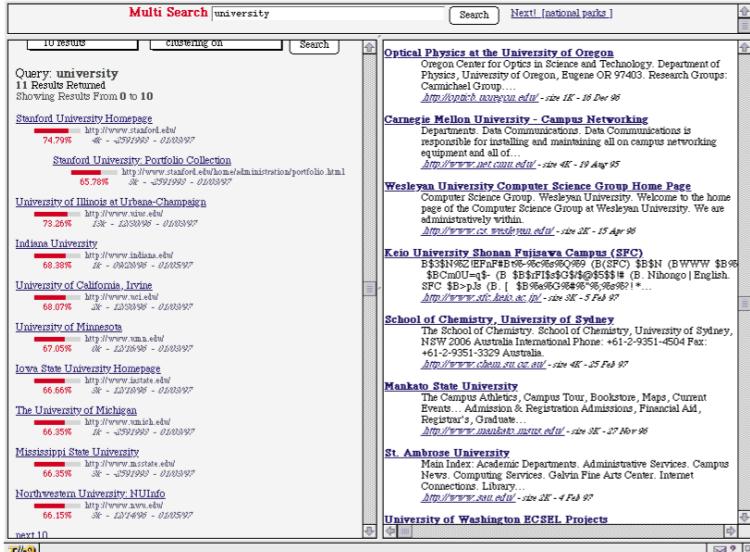
23

24

Social network analysis

PageRank

Application 1: Web search



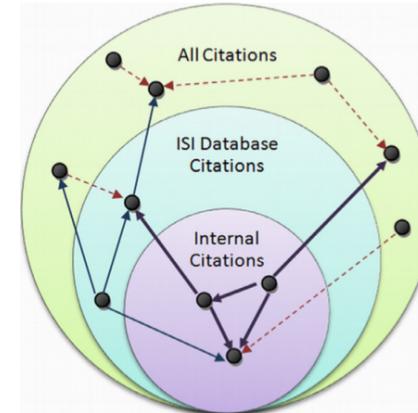
25

Social network analysis

PageRank

Application 2: Citation analysis

Guan et al. 2008. "Bringing Page-Rank to the Citation Analysis"

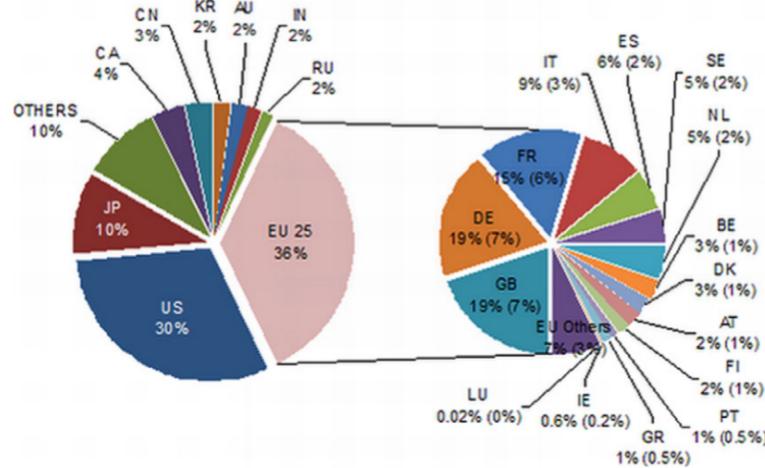


26

Social network analysis

PageRank

Application 2: Citation analysis

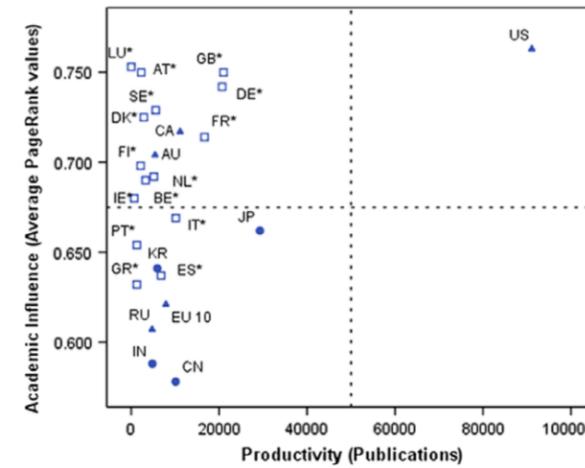


27

Social network analysis

PageRank

Application 2: Citation analysis



28

Social network analysis

PageRank

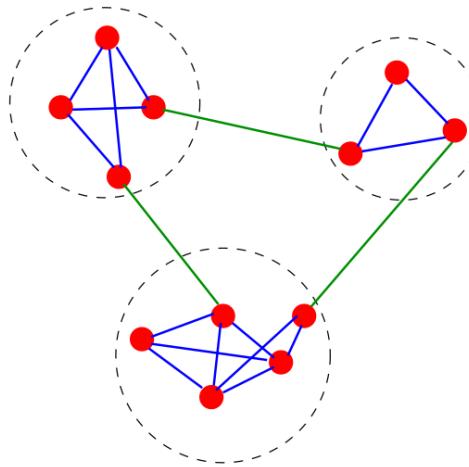
Application 2: Citation analysis

| Country | Number of publications | Internal citations per paper (ICPP) | Rank of ICPP | PageRank per paper (PRPP) | Rank of PRPP |
|--------------|------------------------|-------------------------------------|--------------|---------------------------|--------------|
| USA | 91,079 | 2.991 | 1 | 0.763 | 1 |
| Luxembourg* | 24 | 1.750 | 14 | 0.753 | 2 |
| UK* | 20,984 | 2.790 | 3 | 0.752 | 3 |
| Austria* | 2,287 | 2.798 | 2 | 0.750 | 4 |
| Germany* | 20,680 | 2.640 | 4 | 0.742 | 5 |
| Sweden* | 5,584 | 2.384 | 8 | 0.729 | 6 |
| Denmark* | 2,889 | 2.550 | 5 | 0.725 | 7 |
| Canada | 11,124 | 2.448 | 6 | 0.717 | 8 |
| France* | 16,674 | 2.404 | 7 | 0.714 | 9 |
| Australia | 5,442 | 2.096 | 10 | 0.704 | 10 |
| Finland* | 2,150 | 2.062 | 11 | 0.698 | 11 |
| Netherlands* | 5,144 | 2.132 | 9 | 0.692 | 12 |
| Belgium* | 3,283 | 2.048 | 12 | 0.690 | 13 |
| Ireland* | 692 | 2.036 | 13 | 0.680 | 14 |
| Italy* | 10,158 | 1.741 | 15 | 0.669 | 15 |
| Japan | 29,256 | 1.725 | 16 | 0.662 | 16 |
| Portugal* | 1,303 | 1.498 | 17 | 0.654 | 17 |
| South Korea | 5,944 | 1.457 | 19 | 0.641 | 18 |
| Spain* | 6,801 | 1.438 | 20 | 0.637 | 19 |
| Greece* | 1,285 | 1.487 | 18 | 0.632 | 20 |
| EU 10 | 7,891 | 1.187 | 21 | 0.621 | 21 |
| Russia | 4,757 | 1.042 | 22 | 0.607 | 22 |
| India | 4,821 | 0.790 | 23 | 0.588 | 23 |
| PR China | 10,111 | 0.729 | 24 | 0.578 | 24 |

29

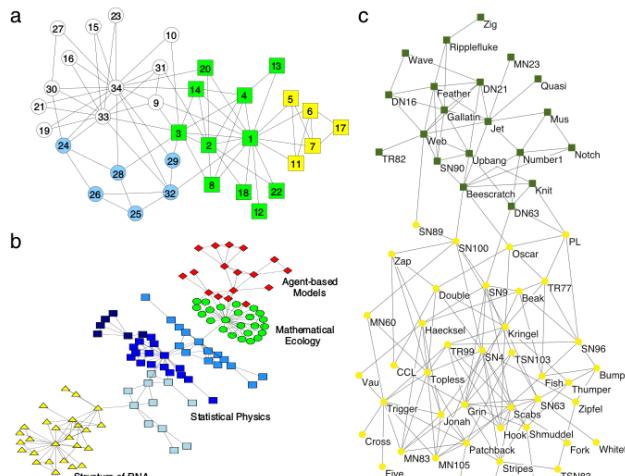
Social network analysis

Community detection



30

from Fortunato (2015)



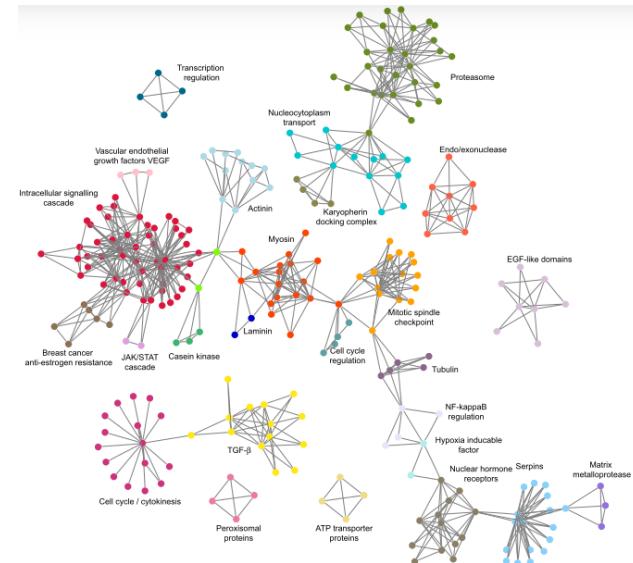
a) Zachary's karate club

b) Collaboration network between scientists working at the Santa Fe Institute

c) Lusseau's network of bottlenose dolphins

from Fortunato (2015)

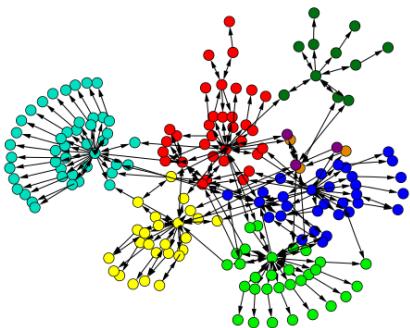
31



Community structure in protein–protein interaction networks

from Fortunato (2015)

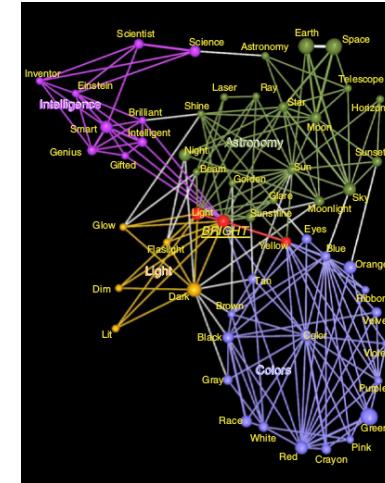
32



Community structure in technological networks

from Fortunato (2015)

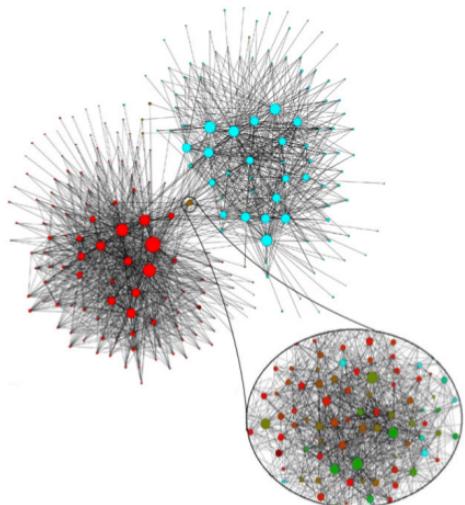
33



Overlapping communities in a network of word association

from Fortunato (2015)

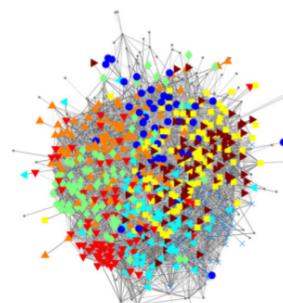
34



Community structure of a social network of mobile phone communication in Belgium

from Fortunato (2015)

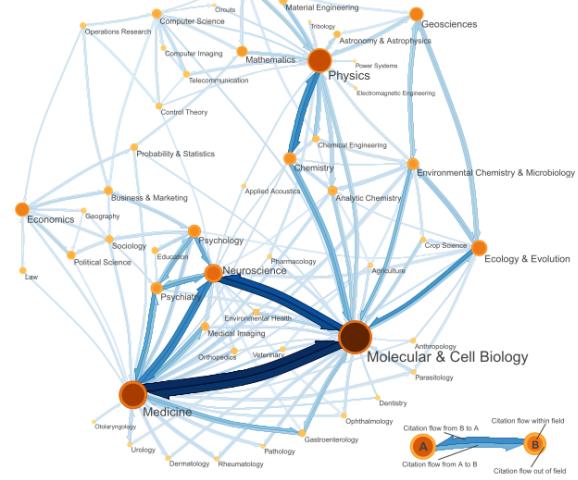
35



Network of friendships between students at Caltech

from Fortunato (2015)

36



Map of science derived from a clustering analysis of a citation network

from Fortunato (2015)

37

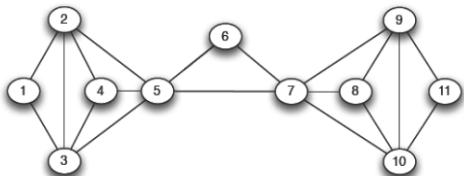
38

Social network analysis Community detection

"The aim of community detection in graphs is to identify the modules and, possibly, their hierarchical organization, by only using the information encoded in the graph topology." - Fortunato

Social network analysis Community detection **Girvan-Newman algorithm**

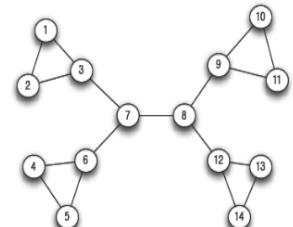
- Bridges connect communities in a graph
- If there are k shortest paths between two nodes, each path loads $1/k$ flow
- E.g: there are 2 shortest paths from 1 to 5, each equivalent to $\frac{1}{2}$ flow



39

Social network analysis Community detection **Girvan-Newman algorithm**

- Edge_betweenness(i) number of flows load by edge i
- E.g:
 - 7-8: 49
 - 3-7: 33
 - 1-3: 12
 - 1-2: 1



40

Algorithm:

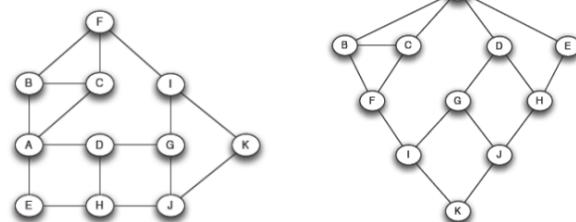
- 1) Calculate edge_betweenness of all edges
 - 2) Remove edges with highest edge_betweenness
 - 3) Recalculate edge_betweenness of affected edges
 - 4) Return to 2), stop when there is no more edge
- we have a *dendrogram*

- Calculate edge-betweenness based on breadth-first-search (BFS)
- For each vertex u
 - Step 1) Execute BFS from u
 - Step 2) Calculate the number of shortest paths from u to each other vertex in the graph
 - Step 3) From that calculate the number of flows from u to others vertices

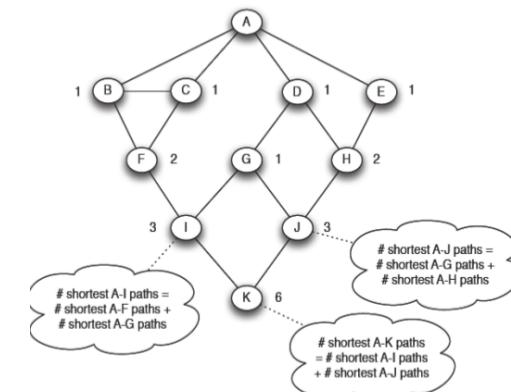
41

42

- Step 1)



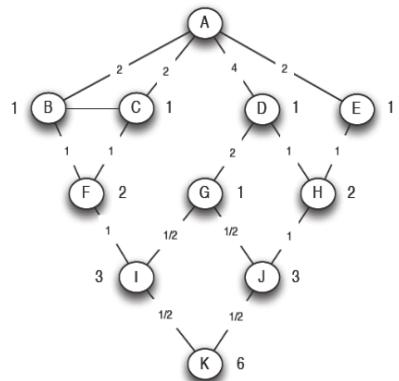
- Step 2)



43

44

- Step 3)



- Calculate edge_betweenness

Execute BFS for each vertex

Calculate edge_betweenness for each edge

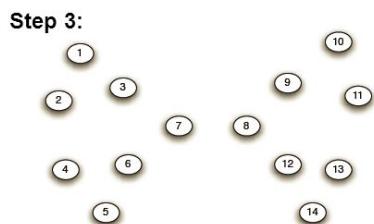
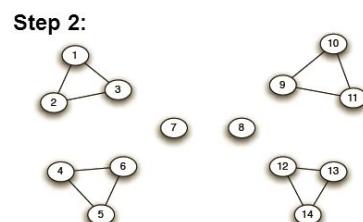
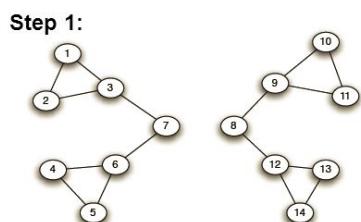
Divide by two (why?)

Apply Girvan-Newman algorithm

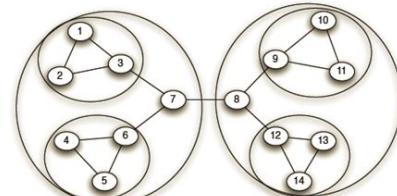
45

46

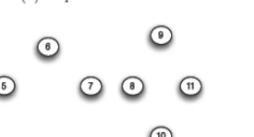
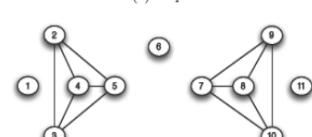
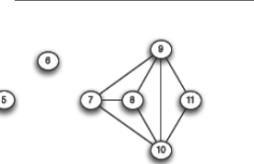
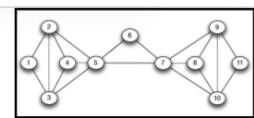
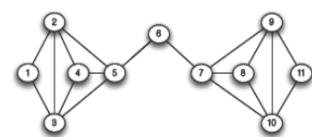
Girvan-Newman: Example



Hierarchical network decomposition:



Example 2



48

Q&A

Mail to: hieunk@soict.hust.edu.vn