Chapter 5
Node Centralities & Similarities

Xiaofan Wang
xfwang@sjtu.edu.cn

课程考核

- 平时表现（参与程度）：30%

11月19日之前把报告题目和参考文献发给：

wangyingdove@gmail.com

- 介绍别人的或者自己的工作，一定要有自己的观点
- 考核标准：**选题的品味、介绍的清晰、文章的规范**
- **严禁任何形式的抄袭！12月30日之前邮件发给助教**

 Xiaofan Wang

如何选题

- 参考课件中列出的一些方向和文献：cnc.sjtu.edu.cn
- 浏览研究人员主页：A-L Barabasi, Mark Newman, Jon Kleinberg, Sinan Aral...
- Google搜索或者顶级期刊搜索（关键词）
- Complexity Digest: http://comdig.unam.mx

Control Profiles of Complex Networks

Justin Ruths[1]* and Derek Ruths[2] SCIENCE VOL 343 21 MARCH 2014

Studying the control properties of complex networks provides insight into how designers and engineers can influence these systems to achieve a desired behavior. Topology of a network has been shown to strongly correlate with certain control properties; here we uncover the fundamental structures that explain the basis of this correlation. We develop the control profile, a statistic that quantifies the different proportions of control-inducing structures present in a network. We find that standard random network models do not reproduce the kinds of control profiles that are observed in real-world networks. The profiles of real networks form three well-defined clusters that provide insight into the high-level organization and function of complex systems.

TECHNICAL COMMENT *Science* **346**, 561 (2014)

NETWORK MODELS

Comment on "Control profiles of complex networks"

Colin Campbell,[1,2]* Katriona Shea,[2] Réka Albert[1,2]

Ruths and Ruths (Reports, 21 March 2014, p. 1373) find that existing synthetic random network models fail to generate control profiles that match those found in real network models. Here, we show that a straightforward extension to the Barabási-Albert model allows the control profile to be "tuned" across the control profile space, permitting more meaningful control profile analyses of real networks.

TECHNICAL RESPONSE *Science* **346**, 561 (2014)

NETWORK MODELS

Response to Comment on "Control profiles of complex networks"

Justin Ruths[1]* and Derek Ruths[2]

Campbell, Shea, and Albert propose an adaptation of the Barabási-Albert model of network formation that permits a level of tuning of the control profiles of these networks. We point out some limitations and generalizations of this method as well as highlight opportunities for future work to refine formation mechanisms to provide control profile tuning in synthetic networks.

 Xiaofan Wang

Locating the source of spreading in complex networks

- P. C. Pinto, P. Thiran, M. Vetterli, Locating the Source of Diffusion in Large-Scale Networks, Phys. Rev. Lett. 109 (2012) 068702.
- D. Brockmann, D. Helbing hidden geometry of complex, network-driven contagion phenomena， Science 342, 1337 (2013)
- F. Altarelli, et al., Bayesian Inference of Epidemics on Networks via Belief Propagation, Phys. Rev. Lett., 112(11) 118701, 2014

思考题：
情绪传染 Emotion Contagion

- 假设你是人人的研究人员，你可以经公司允许在人人上做实验以验证情绪是如何在人们之间传播的。
- 例如：如果一个人看到更多正面或者负面的帖子，是否自己也会变得更为正面或者负面？
- 请问你应该如何设计实验？

A. D. I. Kramer et al., Experimental evidence of massive-scale emotional contagion through social networks, PNAS, 111(24), 2014

- *We show, via a massive (N = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness.*
- *We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.*

- TED专题: NEED TO KNOW: ABOUT FACEBOOK’S EMOTIONAL CONTAGION STUDY
- Facebook“情绪感染”试验被指不道德
- 你的“情感”被Facebook这么玩弄，你造吗？
- 大数据背后的道德隐患
- Facebook的经验揭露了当代互联网的问题

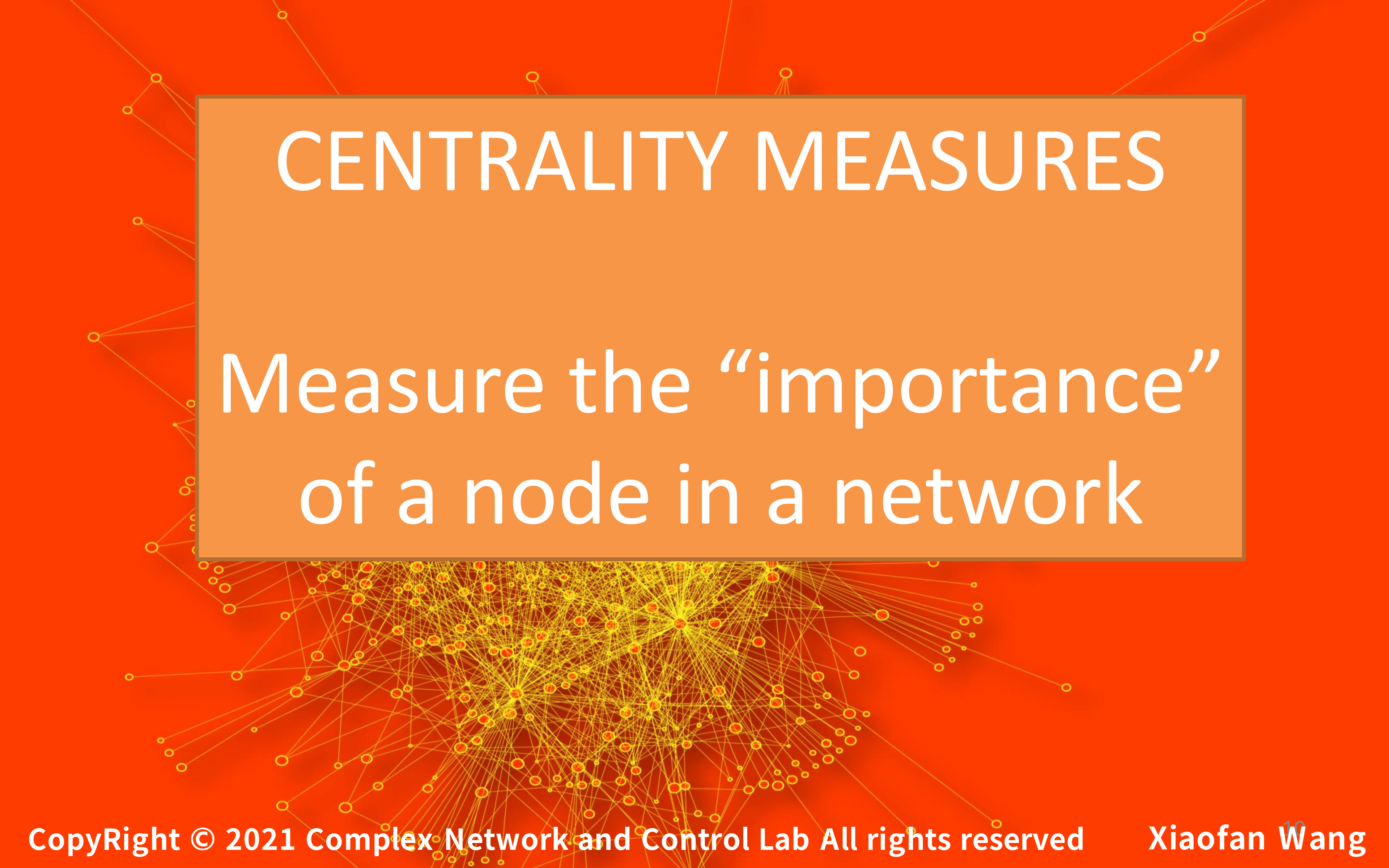
How Does Facebook Choose What To Show In News Feed?

News Feed Visibility =* I x P x C x T x R

Interest, Post, Creator, Type, Recency

Interest	Post	Creator	Type	Recency
Interest of the user in the creator	This post's performance amongst other users	Performance of past posts by the content creator amongst other users	Type of post (status, photo, link) user prefers	How new is the post

* This is a simplified equation. Facebook also looks at roughly 100,000 other high-personalized factors when determining what's shown.



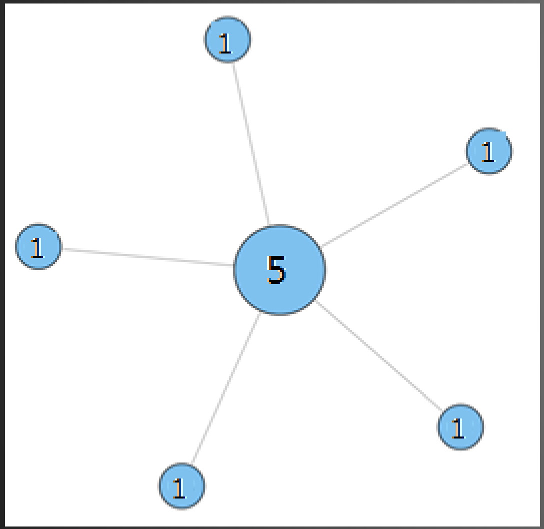
CENTRALITY MEASURES

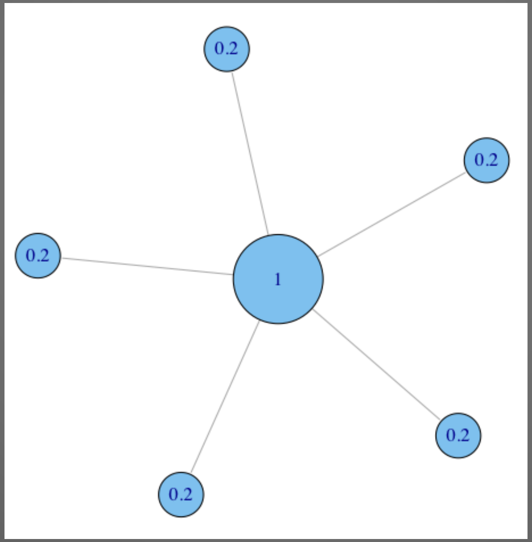
Measure the “importance” of a node in a network

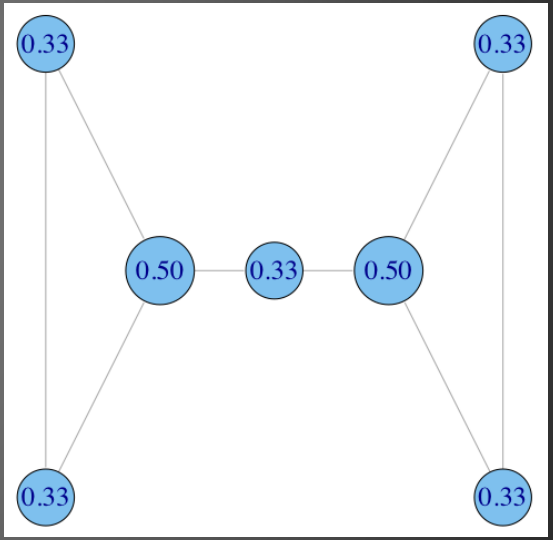
 Xiaofan Wang

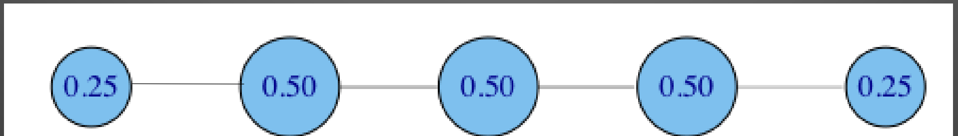
Degree Centrality

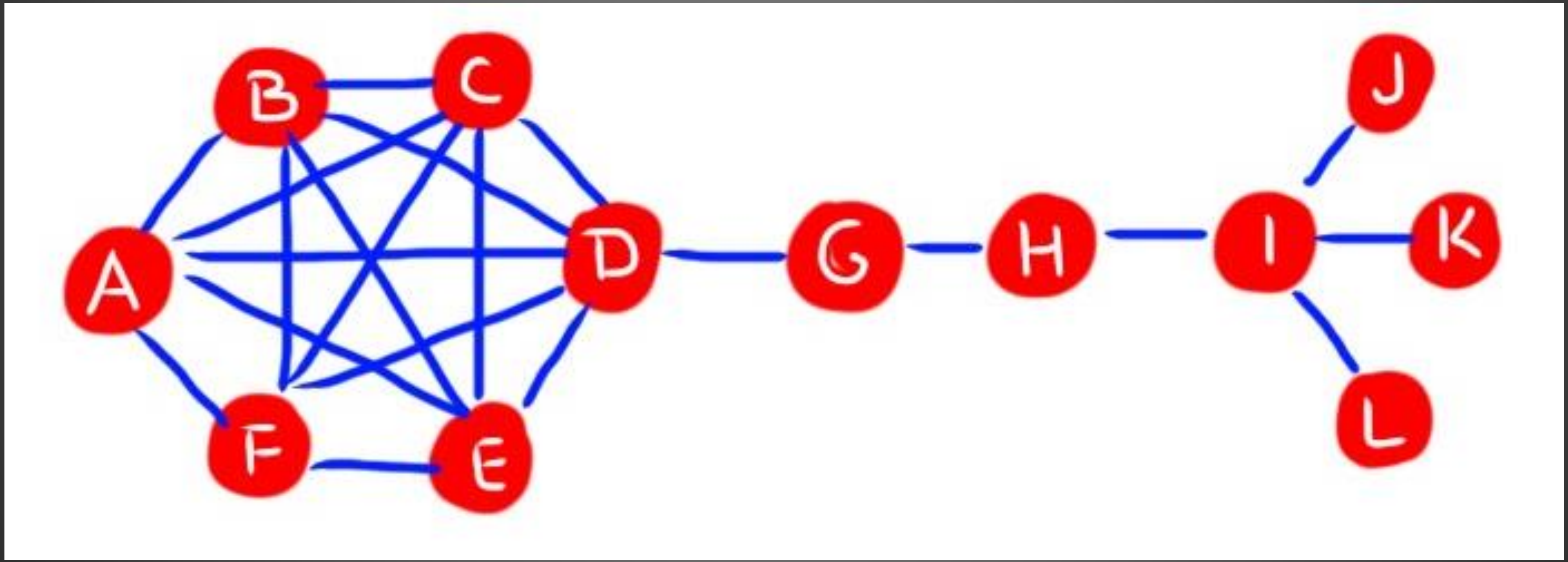
Normalized $$DC_i = \frac{k_i}{N-1}$$











 Xiaofan Wang

BETWENNESS CENTRALITY

number of shortest paths that go through a node

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}$$

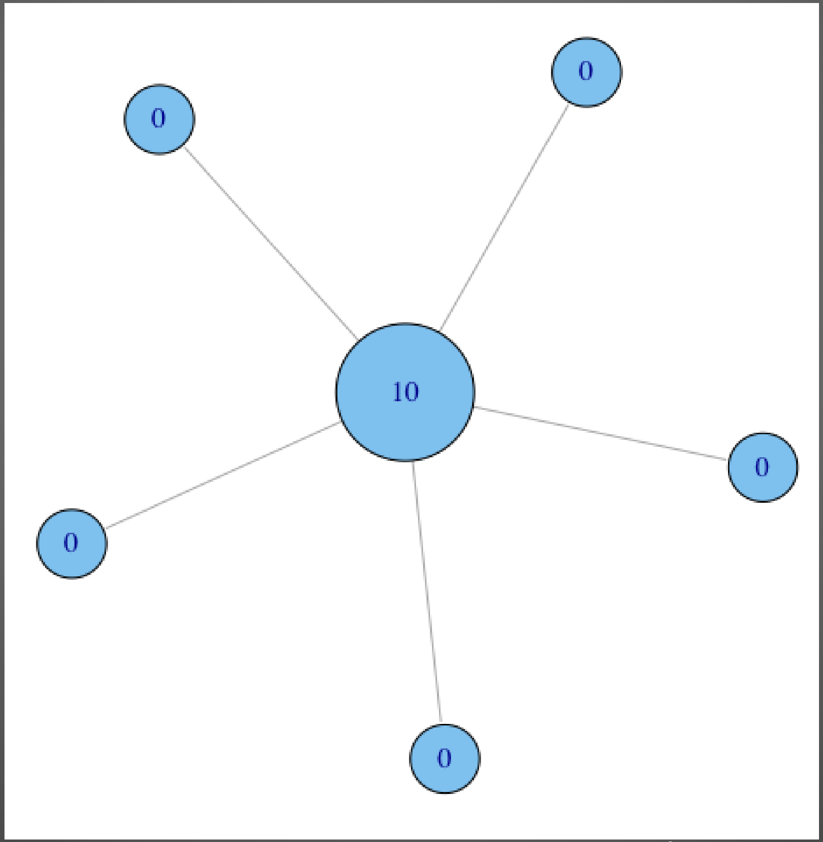
g_{st} = the number of shortest paths connecting *s* & *t*
n_{st} = the number that node *i* is on

$$BC_i = \frac{1}{(N-1)(N-2)/2} \sum_{s,t} \frac{n_{st}^i}{g_{st}}$$

Devided by number of pairs of vertices excluding node i

Betweenness on toy networks

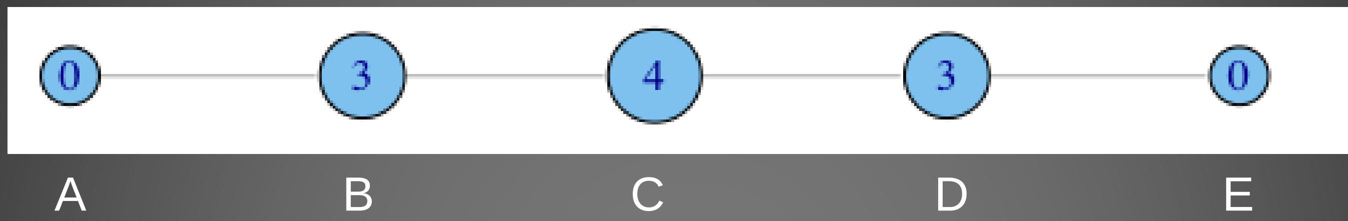
non-normalized version



Xiaofan Wang

Betweenness on toy networks

non-normalized version

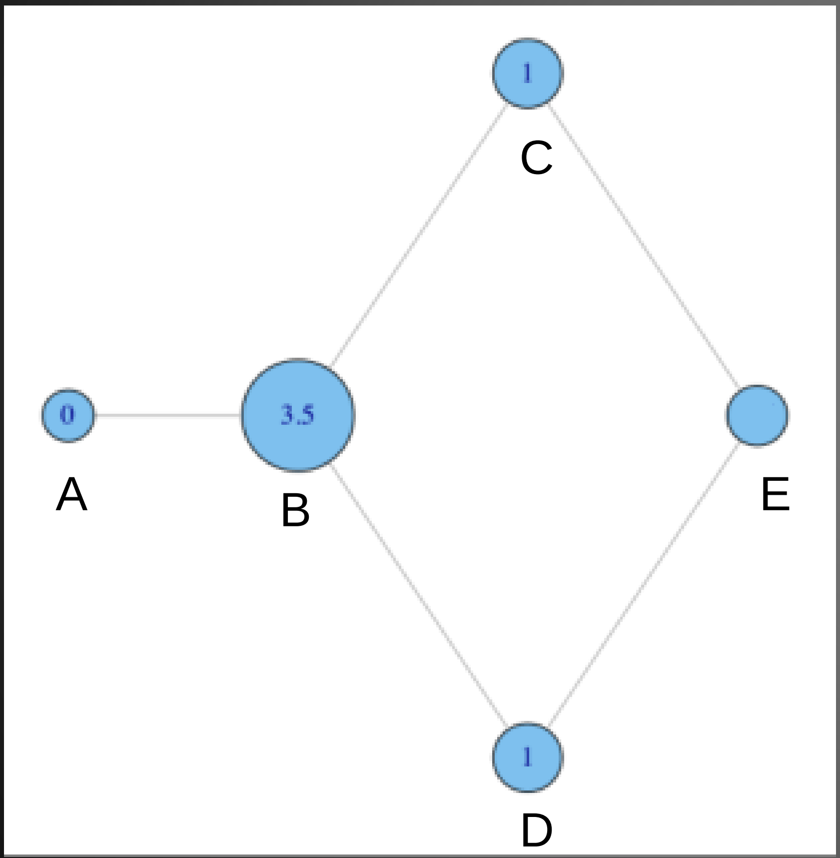


- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E),(B,D),(B,E)

- Note that there are no alternate paths for these pairs to take, so C gets full credit

 Xiaofan Wang

Quiz on Betweenness



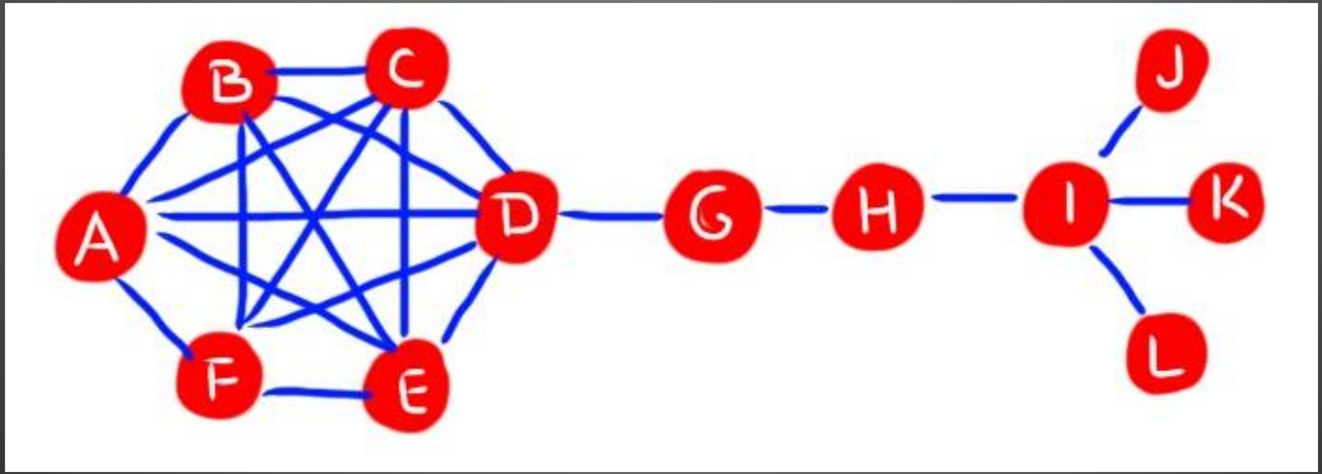
1. Why do C and D each have betweenness 1?
2. What is the betweenness of node E?

1. They are both on shortest paths for pairs (A,E), and (B,E), and so must share credit: ½+½ = 1
2. 0.5: E gets 1/2 of the credit for connecting C and D

Quiz on Betweenness

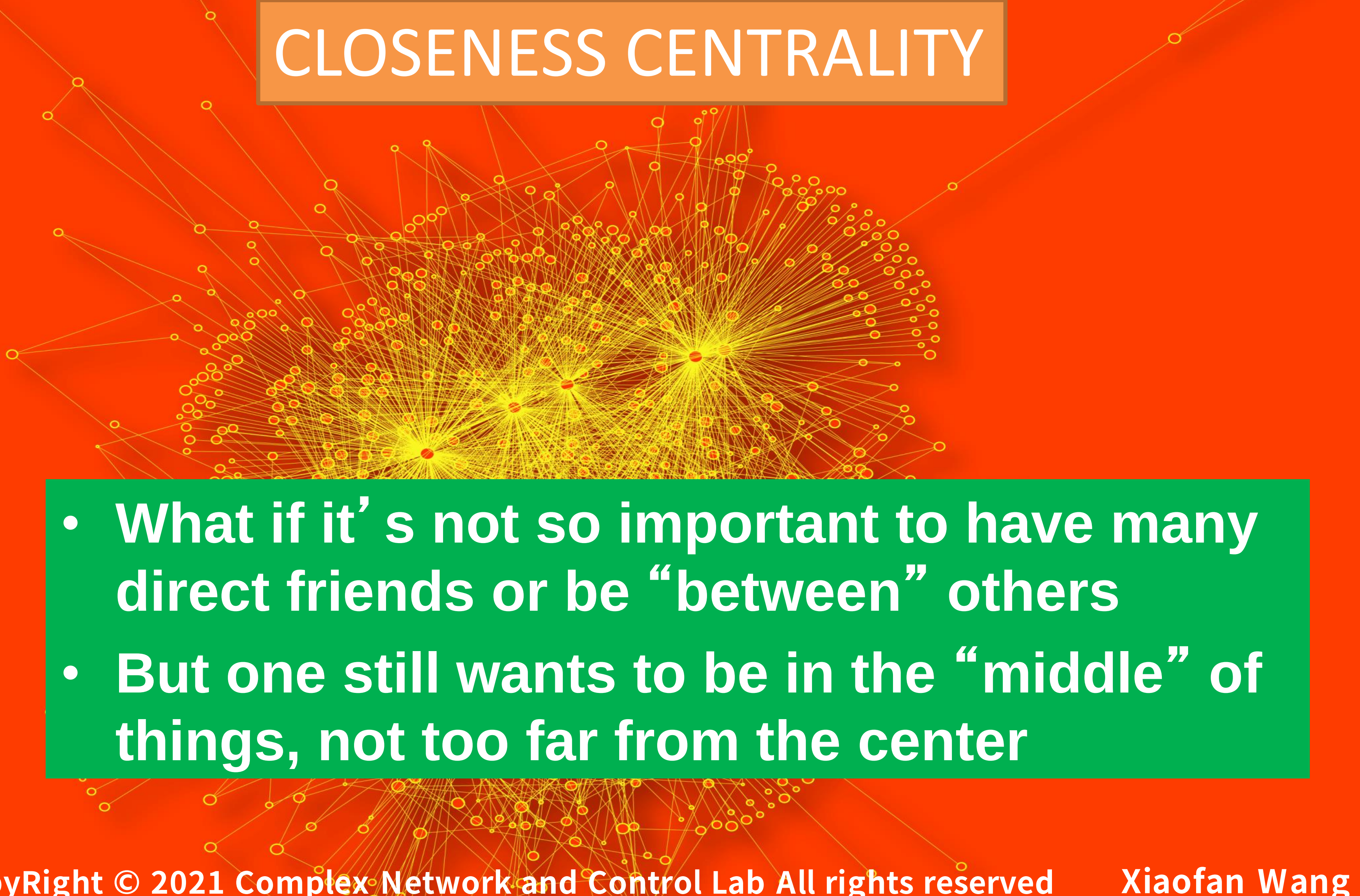
- Among the four nodes A, D, G, I:

1. Find a node that has high betweenness but low degree
2. Find a node that has low betweenness but high degree



 Xiaofan Wang

CLOSENESS CENTRALITY



- What if it’s not so important to have many direct friends or be “between” others
- But one still wants to be in the “middle” of things, not too far from the center

CLOSENESS CENTRALITY

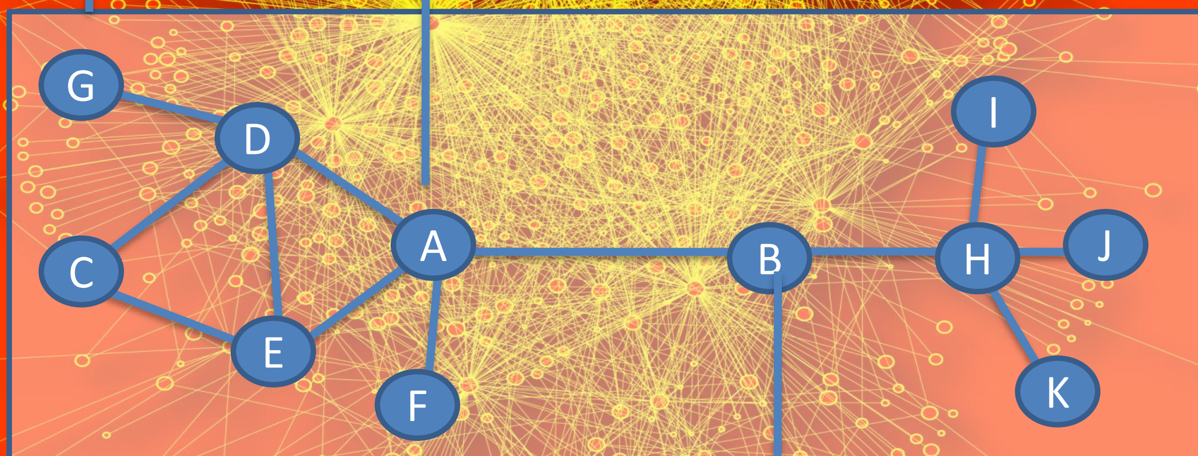
CC= Inverse of the average distance to all other nodes

d(G)=1/10(1+2*3+2*3+4+3*5)
CC(G)=1/3.2

d(A)=1/10(4+2*3+3*3)
CC(A)=1/1.9

$$d_i = \frac{1}{N-1}\sum_{j=1}^{N} d_{ij}$$

$$CC_i = \frac{1}{d_i}$$

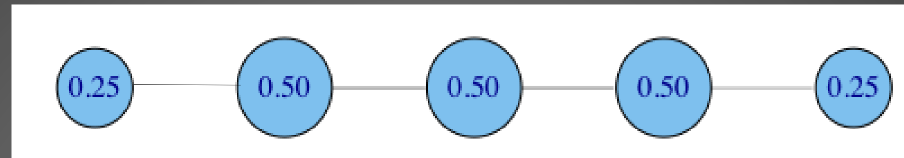


d(B)=1/10(2+2*6+2*3)
CC(B)=1/2

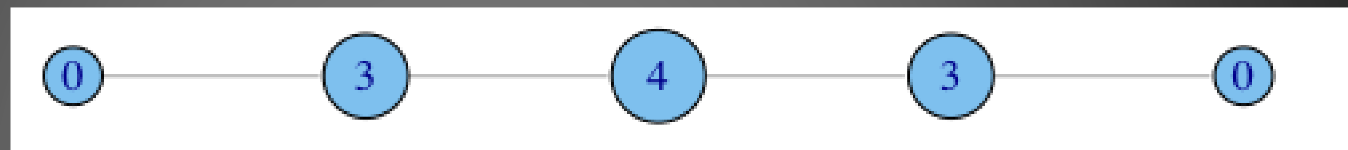
N=11

Examples

Degree



Betweeness



Closeness

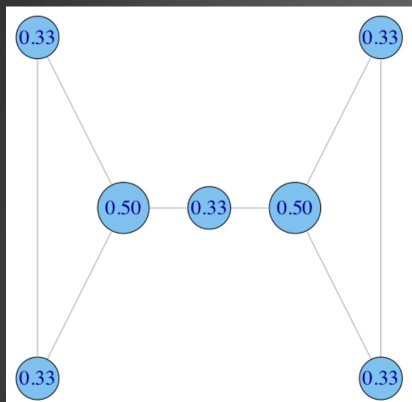
A B C D E

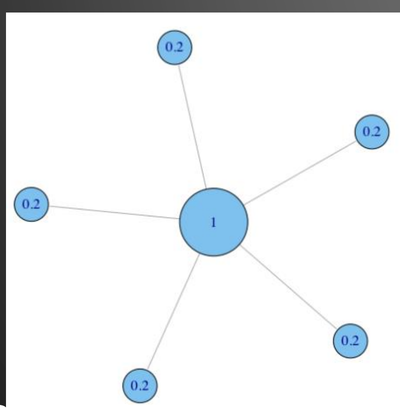
$$CC(A) = \left[\frac{\sum_{j=1}^{N} d(A, j)}{N-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$

 Xiaofan Wang

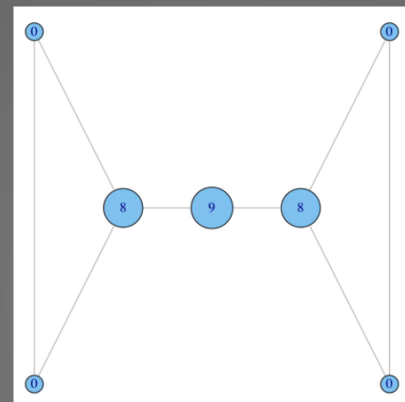
More Examples: Computation Issue (Local vs. Global)

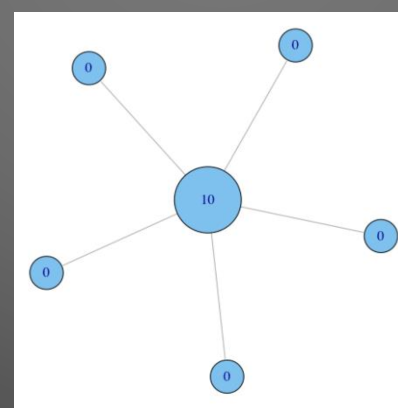
Degree



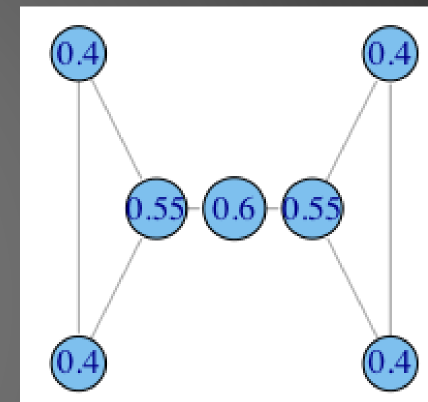


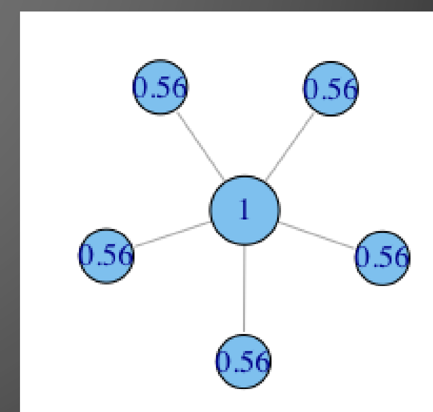
Betweeness





Closeness

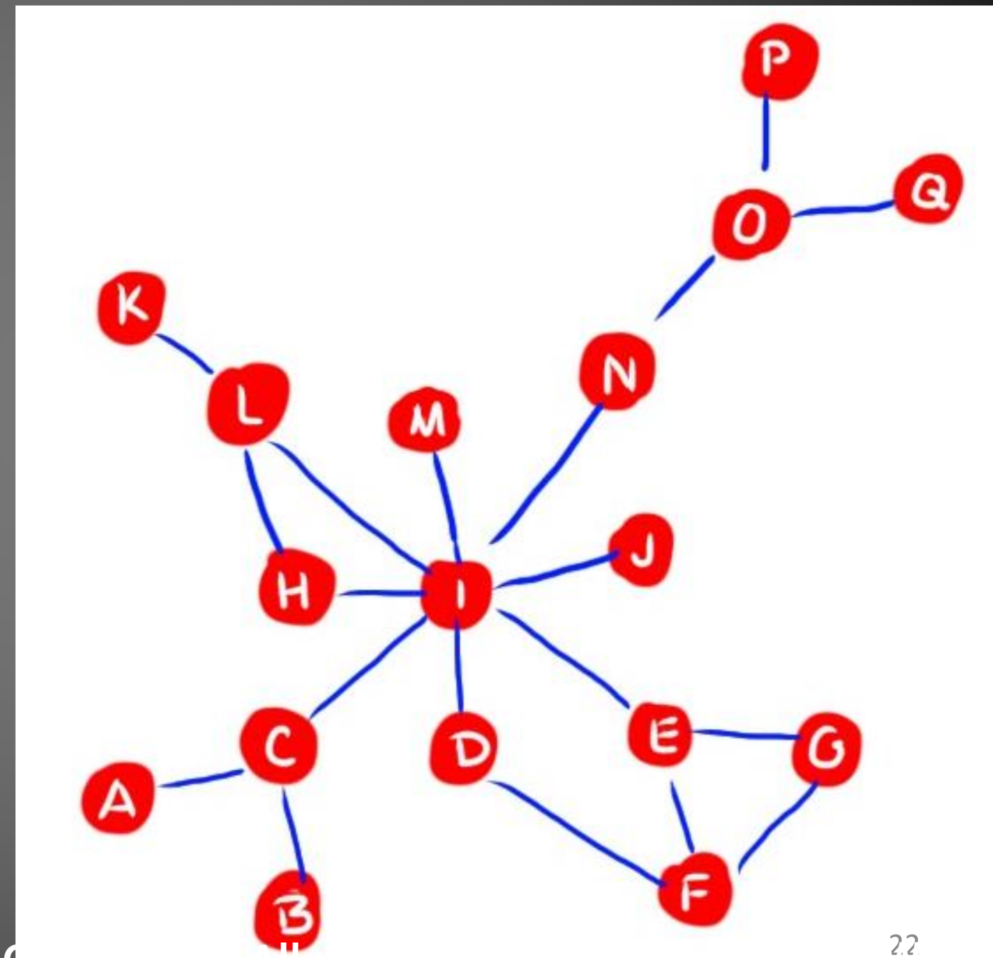




Quiz Q:

- Among four nodes: E, I, J, O

- Which node has relatively high degree but low closeness?



EIGENVECTOR CENTRALITY

- How central you are depends on how central your neighbors are

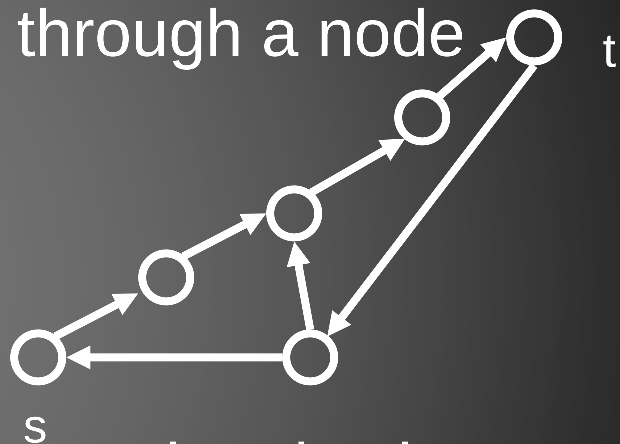
$$x_i = \frac{1}{\lambda_1} \sum_{j=1}^{N} a_{ij} x_j$$

Centrality in Directed Networks

Betweenness Centrality in Directed Networks

- We now consider the fraction of all directed paths between any two vertices that pass through a node

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}$$



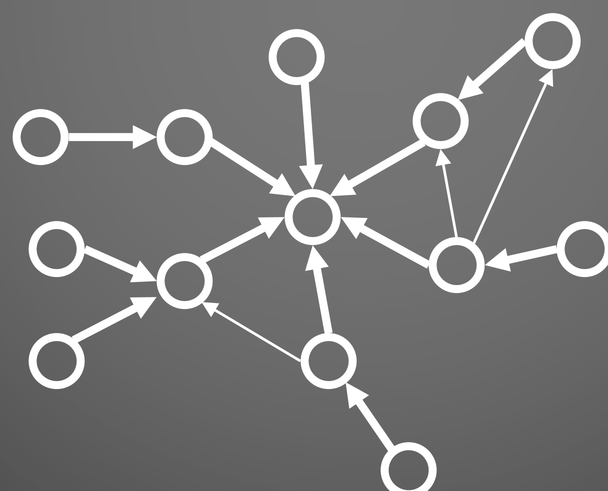
- Only modification: we have twice as many ordered pairs as unordered pairs

$$BC_i = \frac{1}{(N-1)(N-2)/2} \sum_{s,t} \frac{n_{st}^i}{g_{st}} \quad \Rightarrow \quad BC_i = \frac{1}{(N-1)(N-2)} \sum_{s,t} \frac{n_{st}^i}{g_{st}}$$

  Xiaofan Wang

Directed Closeness Centrality

- in-closeness & out-closeness
- usually consider only nodes from which node *i* can be reached



 Xiaofan Wang

Eigenvector Centrality in Directed Networks

- How central you are depends on how central your neighbors are

 Xiaofan Wang

How does Google know which pages are the most important?

Google 汪小帆 Search Advanced Search Preferences

Web

Tip: Try using pinyin for automatic Chinese keyword conversion. [Learn more about pinyin search]

<![CDATA[国家杰出青年科学基金获得者,国际重要学术兼职教授,]]>
File Format: Unrecognized - View as HTML
招生就业|校友会基金会|专题网站|网站地图]]>. 主页 > 师资队伍 > 师资介绍]]>. 152 汪小帆 ...
www.sjtu.edu.cn/staff/teachers/152.xml - Similar pages - Remove result

汪小帆教授受聘担任IEEE电路与系统汇刊编委 - [Translate this page]
电信学院自动化系汪小帆教授受聘于2004年1月起担任《IEEE Transaction on Circuits and ...
我校师生如有关于向该杂志投稿事宜需要咨询，可与汪小帆教授联系。 ...
www.sjtu.edu.cn/newsnet/newsdisplay.php?id=1684 - 7k -
Cached - Similar pages - Remove result
[More results from www.sjtu.edu.cn]

汪小帆专栏 - [Translate this page]
男，1967年2月7日出生，江苏句容人。1977年9月－1982年7月在句容县中学就读。1986年 和 1991年分别于苏州大学数学系和南京师范大学数学系获得理学学士和硕士学位。1996年 ...
www.paper.edu.cn/scholar/paper.jsp?name=wangguanghou - 15k -
Cached - Similar pages - Remove result

[PPT] SJTU 复杂网络中的同步 Synchronization in Complex Networks 汪小帆 ...
File Format: Microsoft Powerpoint - View as HTML
汪小帆. 上海交通大学自动化系. E-mail: xfwang@sjtu.edu.cn ... 汪小帆教授，李翔博士. 博士后 1名. 博士生8名. 硕士生10名. 研究课题：. 混沌控制、同步与反控制 ...
nlsc.ustc.edu.cn/.../ %B8%B4%D4%D3%CD%F8%C2%E7%D6%D0%B5%C4%CD%AC%B2%BD%A3%AD%CD%F4%D0%A1%B7%AB.ppt - Similar pages - Remove result

科技超市 - [Translate this page]
汪小帆，男，1967年2月生，1996年10月于东南大学自动化所控制理论与控制工程专业获博士 ... 汪小帆教授曾多次应邀在国内外学术会议和大学作大会报告和学术讲座。 ...

Earlier Search Engines: Inverted Index

	P1	P2	P3	P4
‘car’	1	0	4	0
‘toyota’	0	2	0	1
‘honda’	2	1	0	0

Pure True Age

Birth of Google， 1998

 Xiaofan Wang

Before 

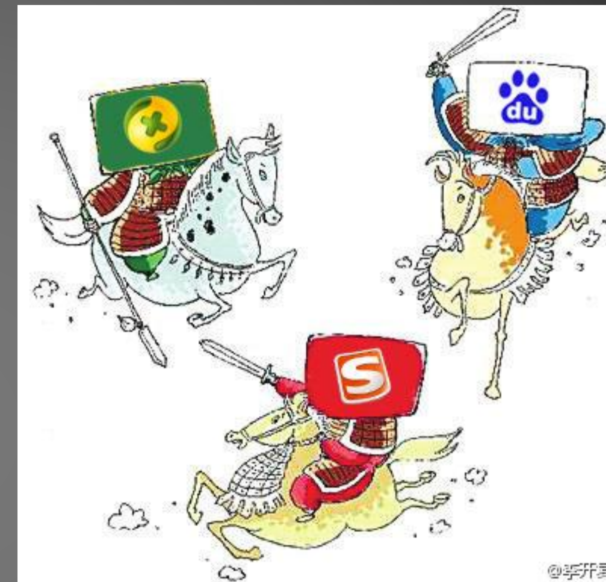
- Open Text (95-97)
- Magellan (95-01)
- Infoseek (95-01)
- Snap (97-01)
- Direct Hit (98-02)
- Lycos(94, reborn 99)
- WebCrawler(94, re 01)
- Yahoo (94, re 02)
- Excite (95, re 01)
- HotBot (96, re 02)
- Ask Jeeves (98, re 02)
- AltaVista (95-)
- LookSmart (96-)
- Overture (98-)
- AOL Search (97-)
- MSN Search (98-)

Google VS. Bing

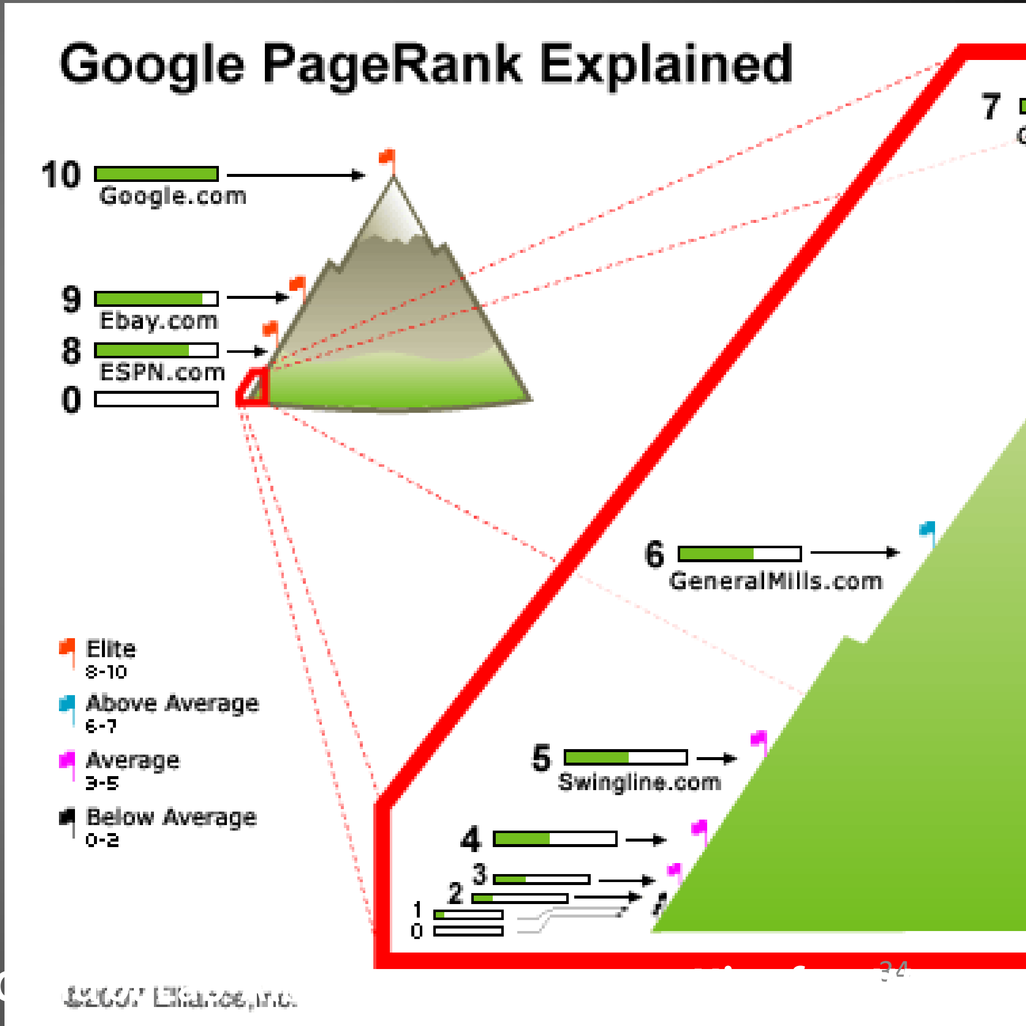
 Xiaofan Wang

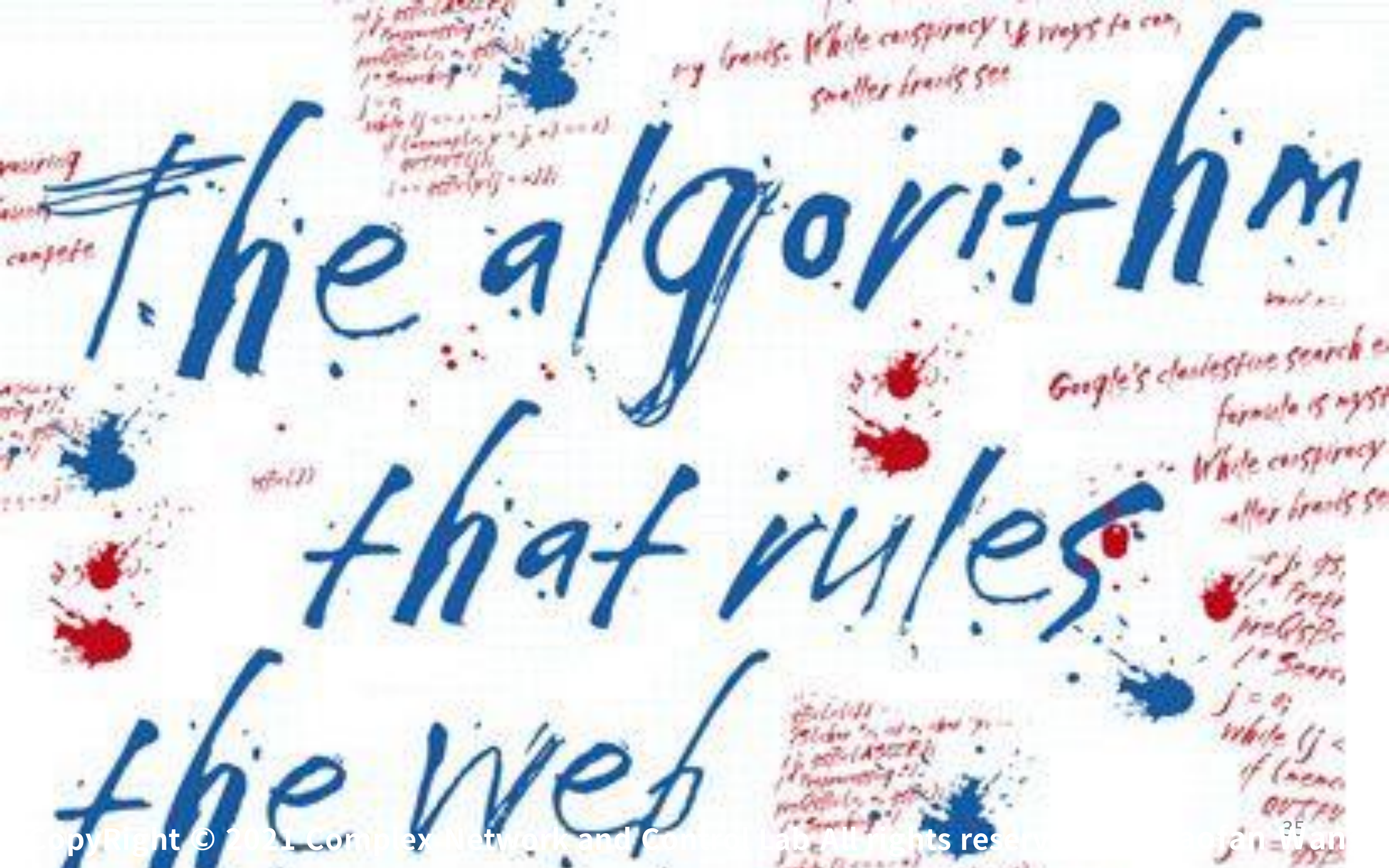
中国市场的搜索战争

- 百度
- 谷歌
- 必应
- 搜狗
- 腾讯搜搜
- 360综合搜索
- 即刻搜索

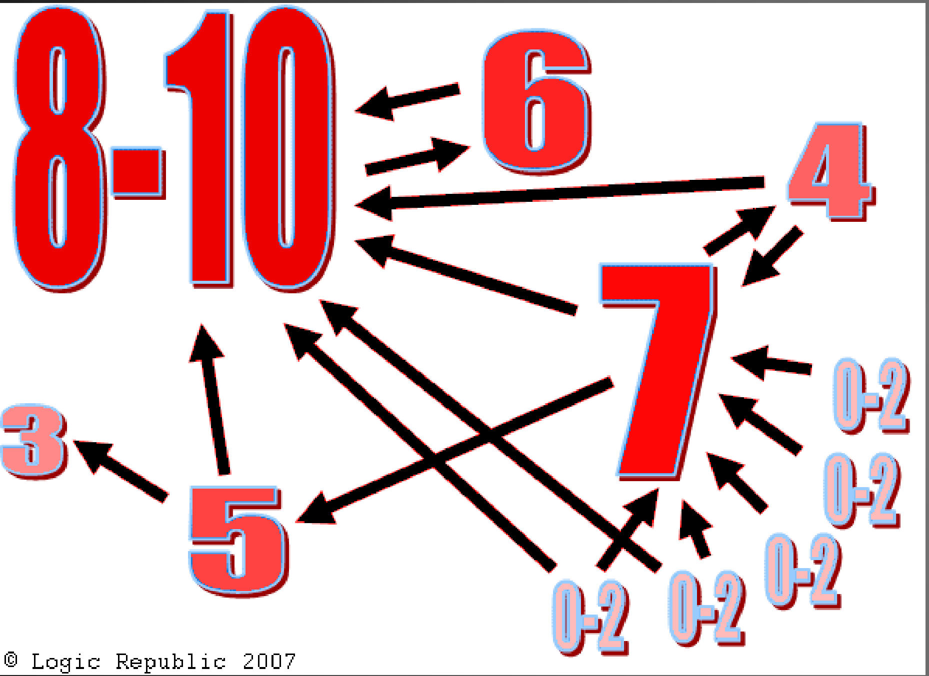


PageRank Tool




The algorithm
that rules
the web

Web Viewed as a Directed Network

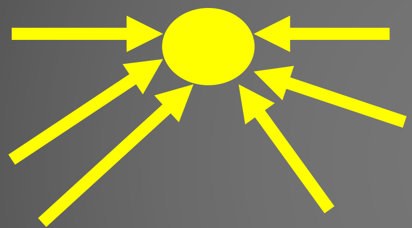


- Nodes: Webpages
- Edges: Hyperlinks

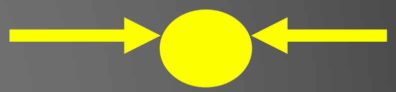
In-Degree of a Webpage

- Number of links point to the page

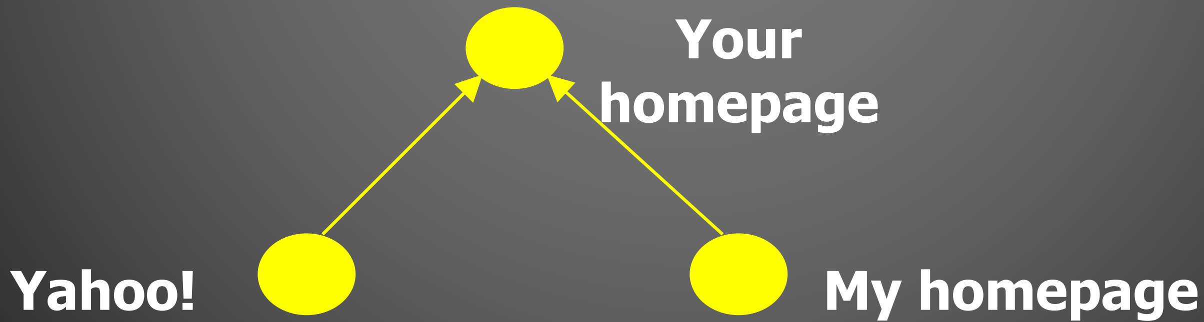
Page A: In-D=6



Page B: In-D=2



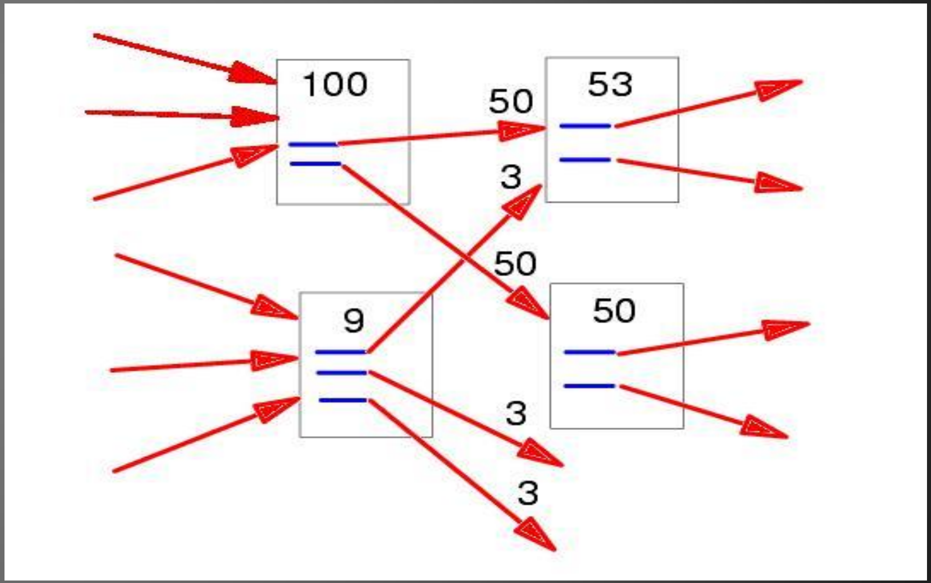
- Is page A more important than page B?



Basic PageRank

- The importance of a page is given by the importance of the pages that link to it

$$PR_i = \sum_{j=1}^{N} a_{ji} \frac{PR_j}{k_j^{out}} \equiv \sum_{j=1}^{N} \bar{a}_{ji} PR_j$$



An Example from Page et al. (1998)

 Xiaofan Wang

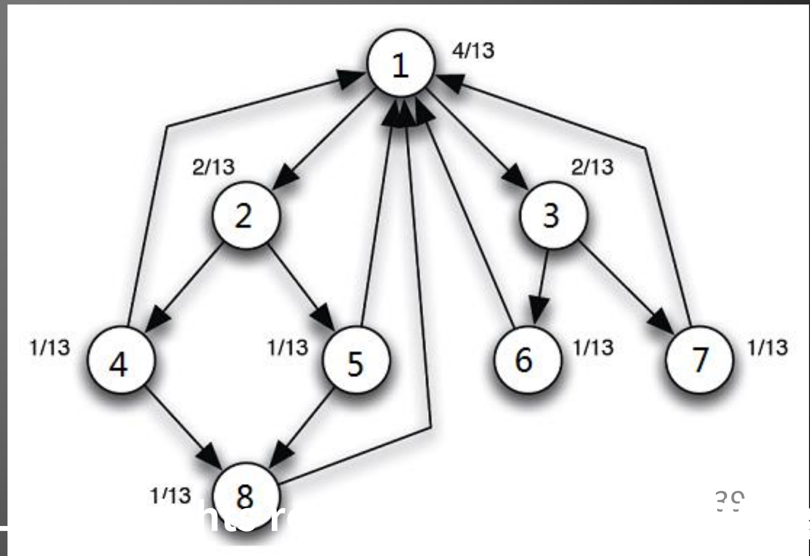
Basic PageRank Algorithm

$$PR_i = \sum_{j=1}^{N} \bar{a}_{ji} PR_j \qquad \sum_{i=1}^{N} PR_i(0) = 1$$

$$PR_i(k) = \sum_{j=1}^{N} \bar{a}_{ji} PR_j(k-1) \qquad PR(k) = \bar{A}^T PR(k-1)$$

- Power method

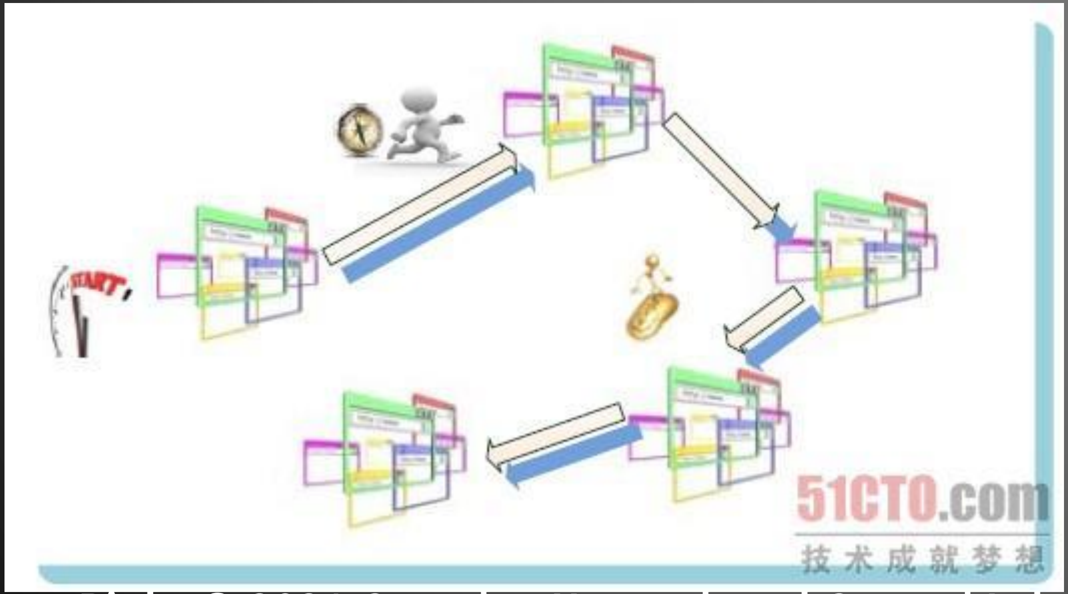
$$\bar{A} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

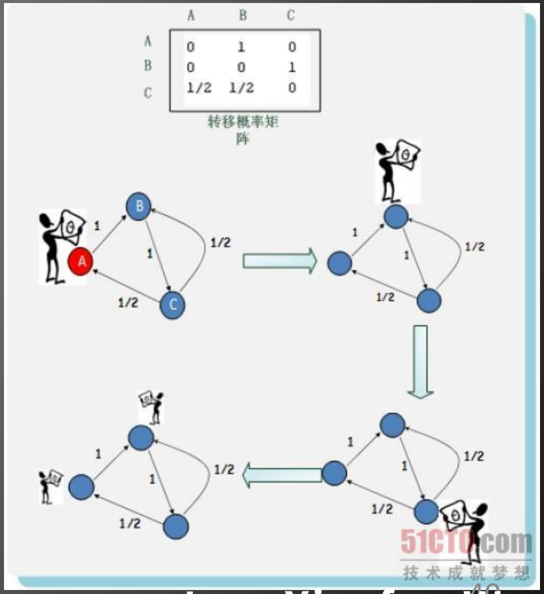


Random Surfer Model

$$PR_i(k) = \sum_{j=1}^{N} \bar{a}_{ji} PR_j(k-1)$$

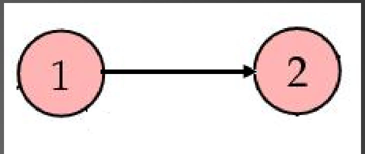
- $PR_i(k)$: Probability that the surfer will be on the webpage i at time k.





Dangling node

$$\bar{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad PR(0) = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$



$$PR(1) = \bar{A}^T PR(0) = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix} \qquad PR(2) = \bar{A}^T PR(1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\bar{a}_{ij} = \begin{cases} 1/k_i^{out} & \text{如果}k_i^{out}>0\text{且有从节点}i\text{指向节点}j\text{的边} \\ 0 & \text{如果}k_i^{out}>0\text{且没有从节点}i\text{指向节点}j\text{的边} \\ 1/N & \text{如果}k_i^{out}=0 \end{cases}$$

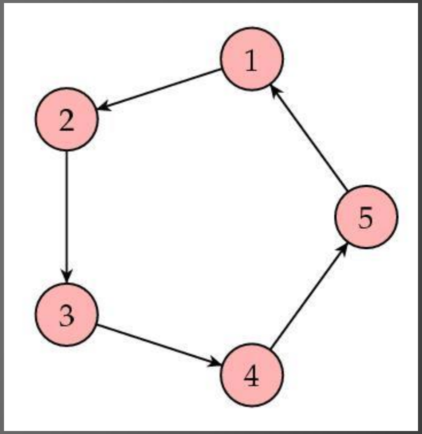
$$\bar{A} = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} \qquad PR^* = \begin{bmatrix} 1/3 & 2/3 \end{bmatrix}^T$$

 Xiaofan Wang

However

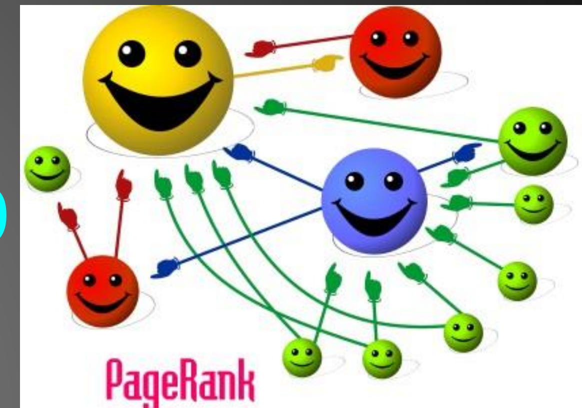
- The basic PR algorithm may still fail even if the network is strongly connected

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



PR(5)=PR(0)=$[1, 0, 0, 0, 0]^T$

PageRank Algorithm



- Basic PR Algorithm

$$PR_i(k)=\sum_{j=1}^{N}\bar{a}_{ji}PR_j(k-1) \qquad PR(k)=\bar{A}^T PR(k-1)$$

- PR Algorithm

$$PR_i(k)=s\sum_{j=1}^{N}\bar{a}_{ji}PR_j(k-1)+(1-s)\frac{1}{N}$$

$$PR(k)=\tilde{A}^T PR(k-1) \qquad \tilde{A}=s\bar{A}+(1-s)\frac{1}{N}ee^T \qquad e=[1 \;\; 1 \;\; \cdots \;\; 1]^T$$

 Xiaofan Wang

PageRank Algorithm Analysis

$$PR_i(k) = s\sum_{j=1}^{N} \bar{a}_{ji} PR_j(k-1) + (1-s)\frac{1}{N}$$

$$PR(k) = \tilde{A}^T PR(k-1) \qquad \tilde{A} = s\bar{A} + (1-s)\frac{1}{N}ee^T \qquad e = [1 \ \ 1 \ \ \cdots \ \ 1]^T$$

- The system matrix is positive
- Unique largest positive eigenvalue, unit eigenvector PR*
- If the matrix is row stochastic, then PR(k)→PR*

Google Algorithm

Google's Score

= (Keyword Usage Score * 0.3)

+ (Domain * 0.25)

+ (PR Score * 0.25)

+ (Inbound Link Score * 0.25)

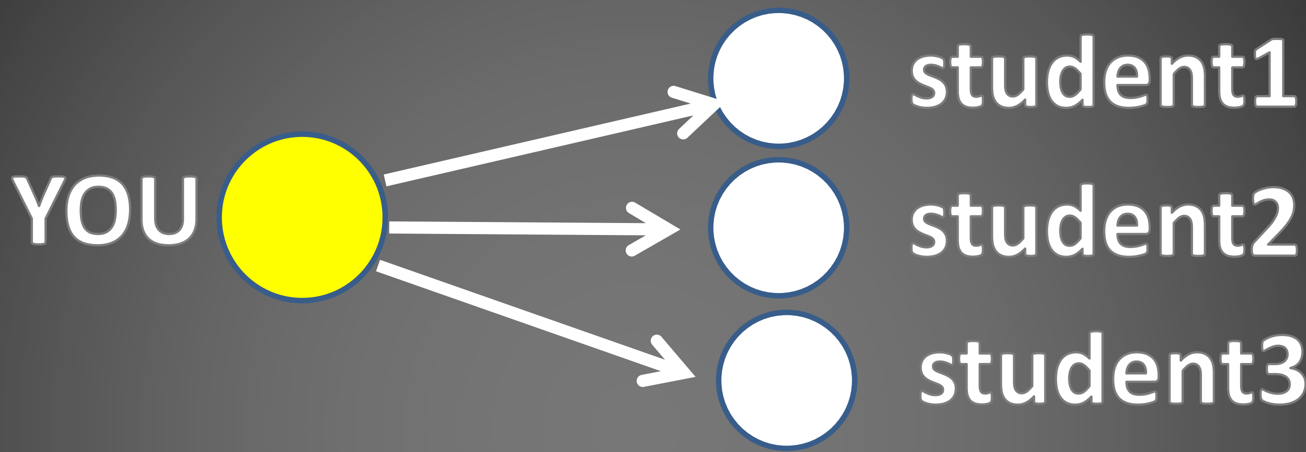
+ (User Data * 0.1)

+ (Content Quality Score * 0.1)

+ (Manual Boosts) - (Automated & Manual Penalties)

Websites that are clean, focused, compatible and fast will benefit.

Quiz Q: PageRank Competition



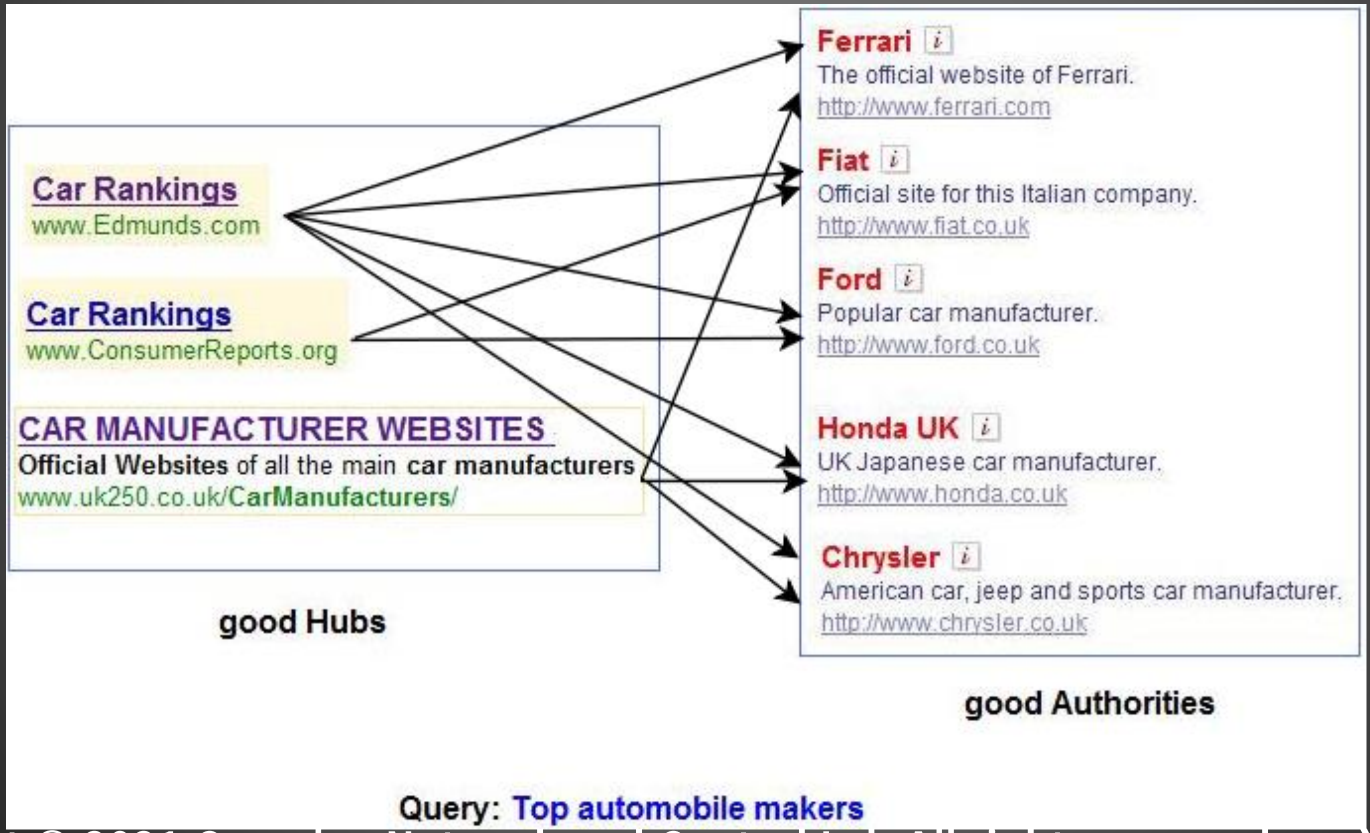
- Each submitted node will receive **3 points**. The node with the highest PageRank will receive **30 points**.
- He/she can distribute the points to anyone in the class. So basically it's a competition.
- Your objective is for you and your co-conspirators to achieve the highest PageRank for one of your nodes.

一位因Google算法调整而差点自杀的站长的话

- I can say that what happened to me this year with Google came close to suicide. I faced financial ruin. The only thing stopping me was not wanting to dump all of this onto my partner and leave my children. But there were many times I just wished I was gone. I could not cope with the desperation of not being able to pay our bills. It was horrendous. I am sorry if that breaks yet more rules or is unpalatable, but it is how it was.I honestly believe I was just collateral damage. I had never engaged in anything dodgy on my site. My competitors were wiped out too. They just turned up the dial on a couple of “brand” sites & the rest of us lost out. The consequences were devastating.
- I am sorry to anybody else who has been hit. I can say that for me, there has been a light at the end of the tunnel, and Google seem to like me again. Not so much with my competitors though. I still see them nowhere.

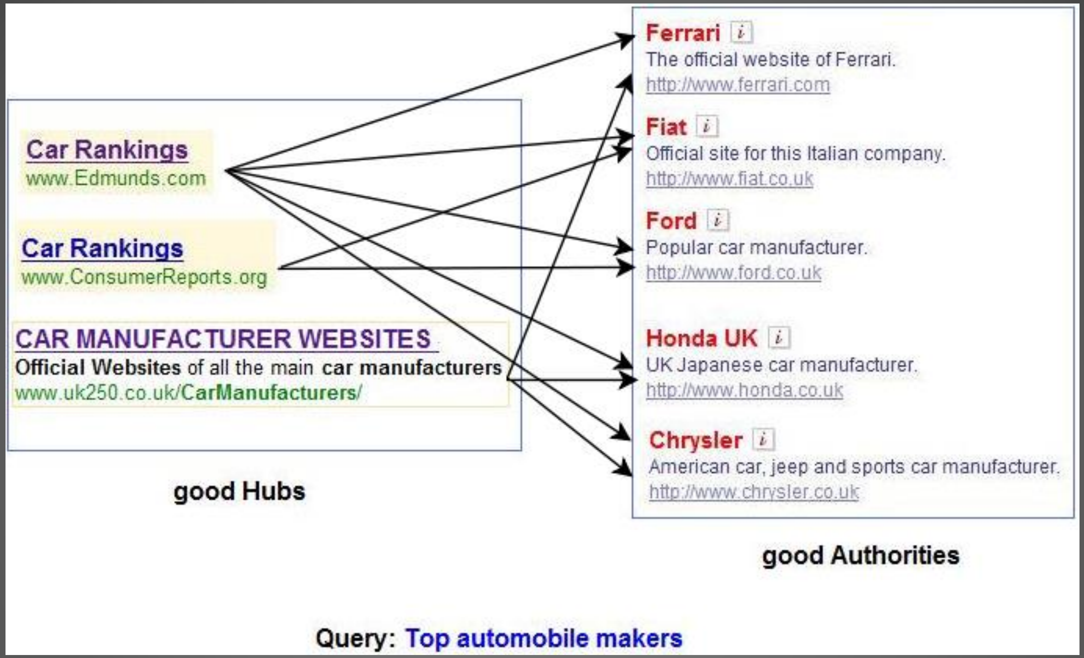
Authorities and Hubs

- An example: query "automobile makers"



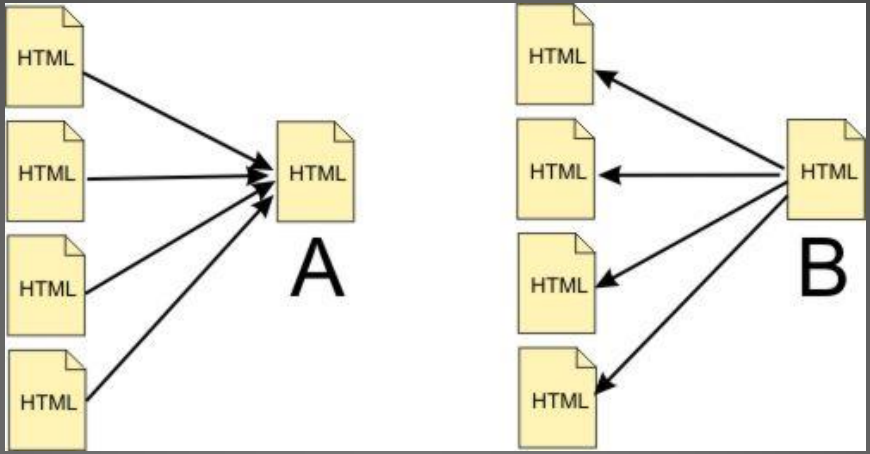
 Xiaofan Wang

Authorities and Hubs



- *Authority:* pages that provide an important, trustworthy information on a given topic
- *Hub:* pages that contain links to authorities

Authorities and Hubs



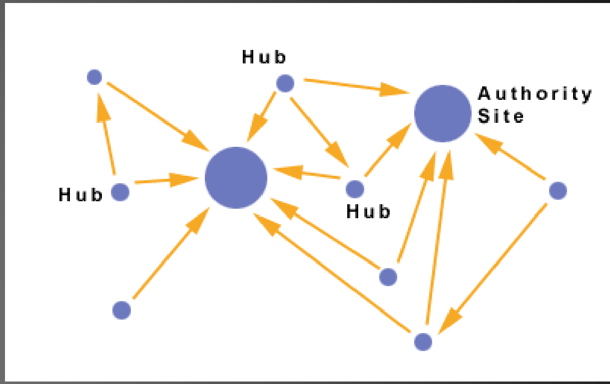
$$x_i = \sum_{j=1}^{N} a_{ji} y_j \qquad y_i = \sum_{j=1}^{N} a_{ij} x_j$$

- They exhibit a *mutually reinforcing relationship*:
- a better hub points to many good authorities
- a better authority is pointed to by many good hubs

HITS Algorithm

- Given x(0) and y(0)

$$x_i'(k)=\sum_{j=1}^{N}a_{ji}y_j(k-1) \qquad y_i'(k)=\sum_{j=1}^{N}a_{ij}x_j'(k)$$



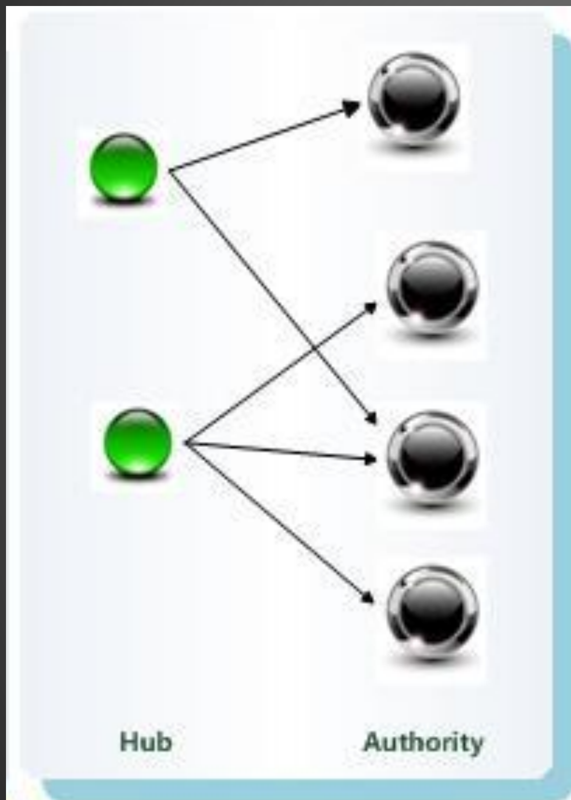
$$x_i(k)=\frac{x_i'(k)}{\|x'(k)\|} \qquad y_i(k)=\frac{y_i'(k)}{\|y'(k)\|}$$

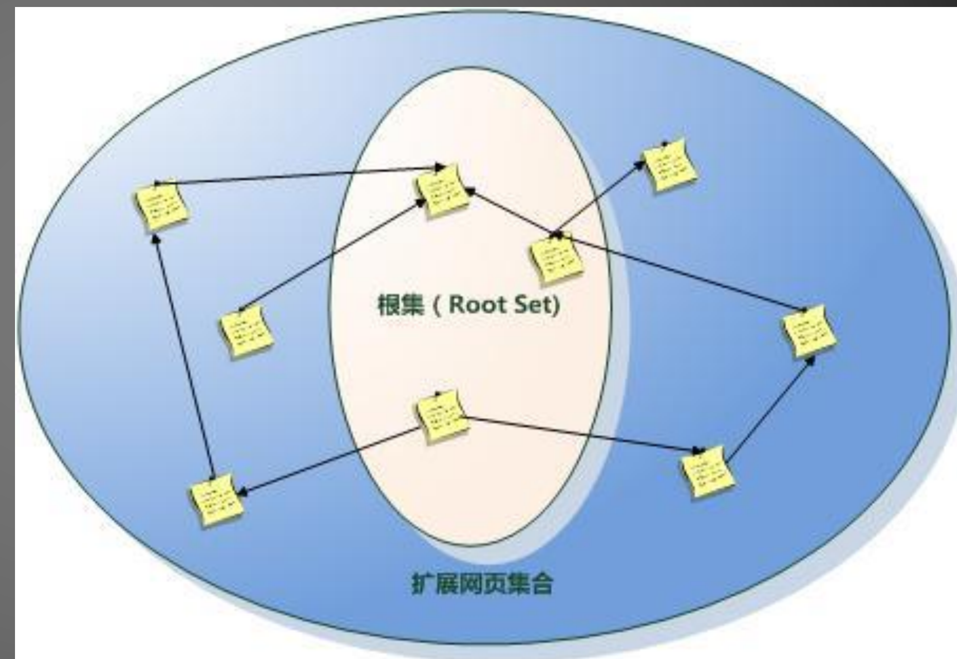
$$x(k)=\alpha_k\left(A^TA\right)x(k-1) \qquad y(k)=\beta_k\left(AA^T\right)y(k-1)$$

$$\lambda_1>\lambda_2\geq\lambda_3\geq\cdots\geq\lambda_N\geq 0$$

- ◆ The authority vector x* is an eigenvector of A^TA
- ◆ The hub vector y* is an eigenvector of AA^T

HITS Realization

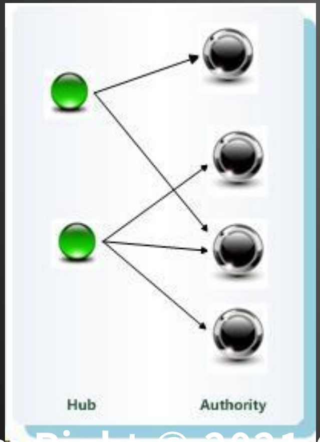


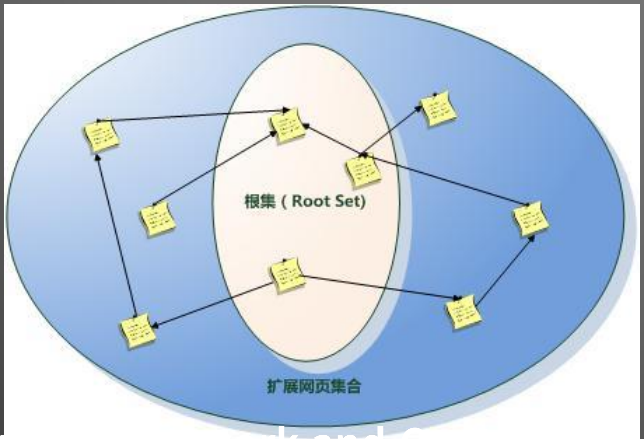


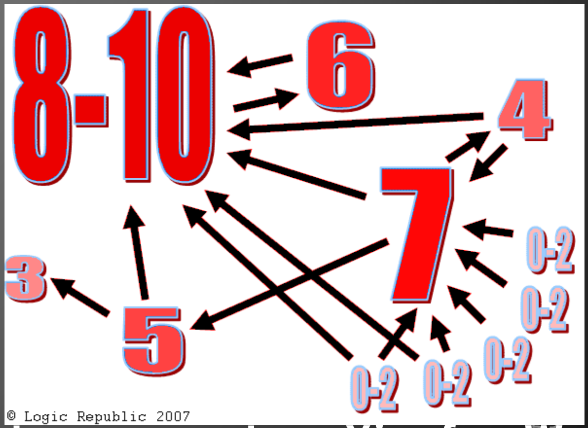
 Xiaofan Wang

HITS vs PageRank

- HITS emphasizes mutual reinforcement between authorities and hubs, while PageRank does not attempt to capture the distinction between hubs and authorities. It ranks pages just by authority.
- HITS is applied to the local neighborhood of pages surrounding the results of a query whereas PageRank is applied to the entire web
- HITS is query dependent but PageRank is query-independent



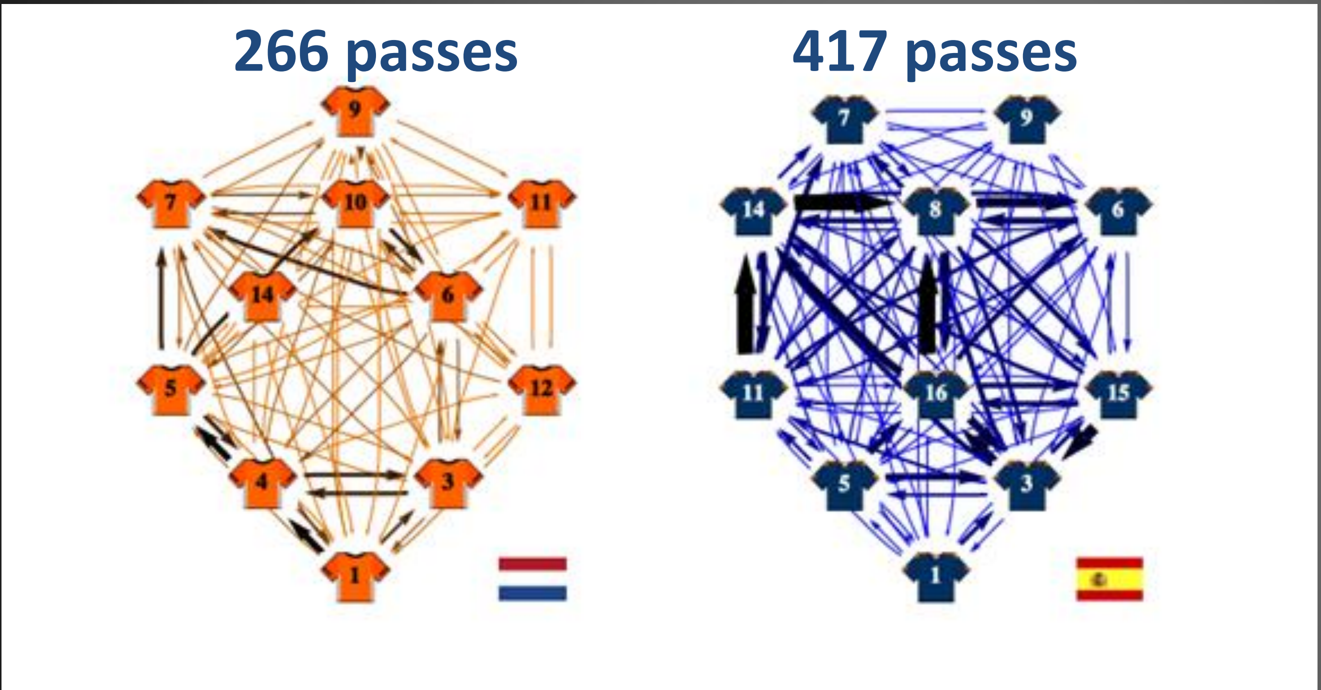




 Xiaofan Wang

A Network Theory Analysis Of Football Strategies

2010 World Cup in South Africa



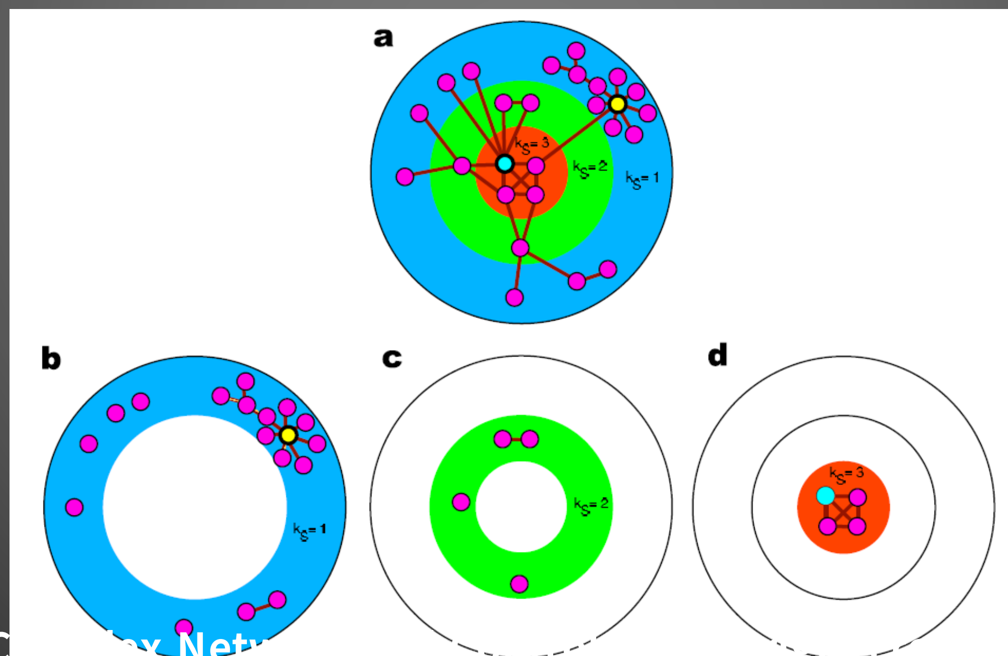
- Degree & CC:
  16 (Sergio Busquets)
  8 (Xavi)
- BC:
  11 Joan Capdevilla
  mainly feeds to 14(Alonso)
- PR:
  8 (Xavi)

- arxiv.org/abs/1206.6904

 Xiaofan Wang

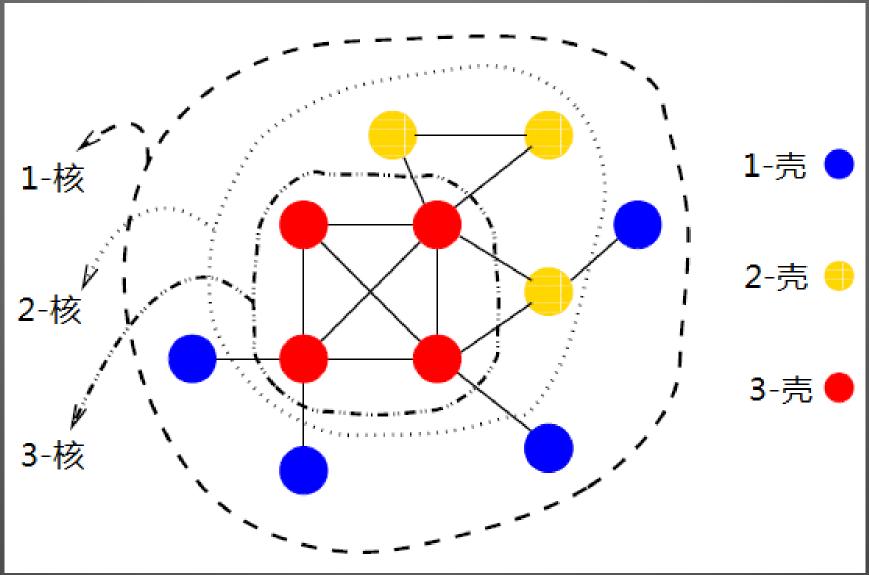
k-shell decomposition

- Start by removing all nodes with degree 1 only (with their links), until no more such nodes remain, and assign them to **the 1-shell.**
- In the same manner, recursively remove all nodes with degree<=k, creating the **k-shell.**



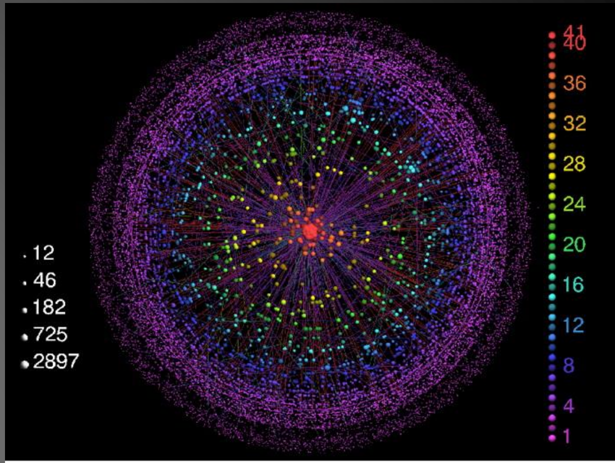
k-shell, k-core, k-crust

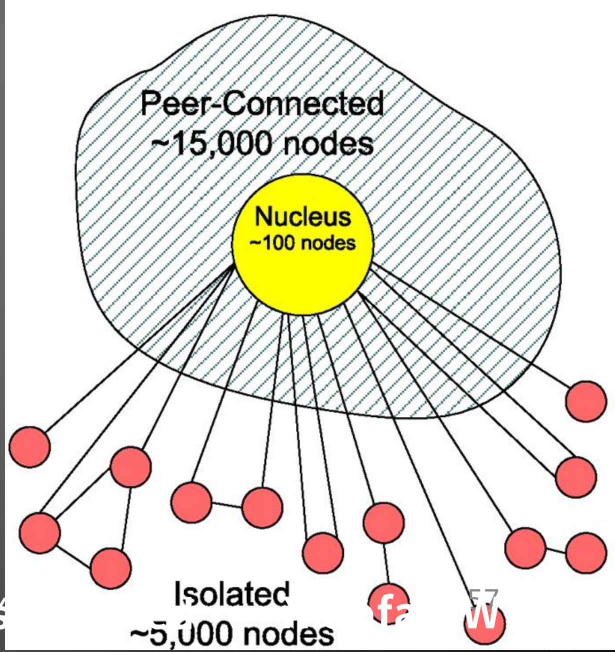
- The *k*-core is defined as the union of all shells with indices larger or equal to *k*.
- The *k*-crust is defined as the union of all shells with indices smaller or equal to *k*.



k-shell Decomposition of the Internet

- **Nucleus:** all nodes in the k_{max}-shell.
- **Peer-connected component:** nodes that belong to the largest connected component of the $(k_{max} - 1)$-crust.
- **Isolated component:** other nodes of the $(k_{max} - 1)$-crust, which belong to smaller clusters.



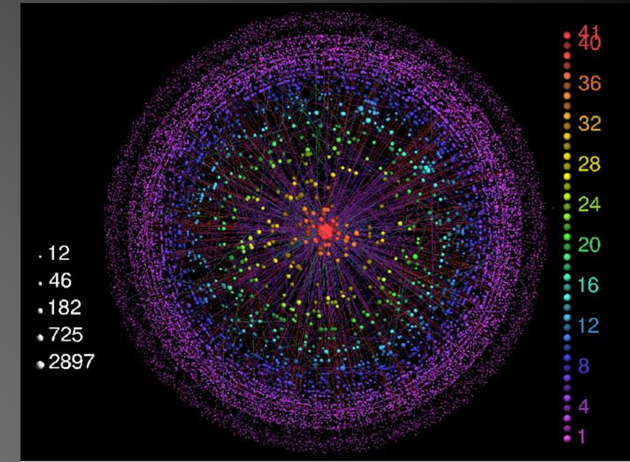


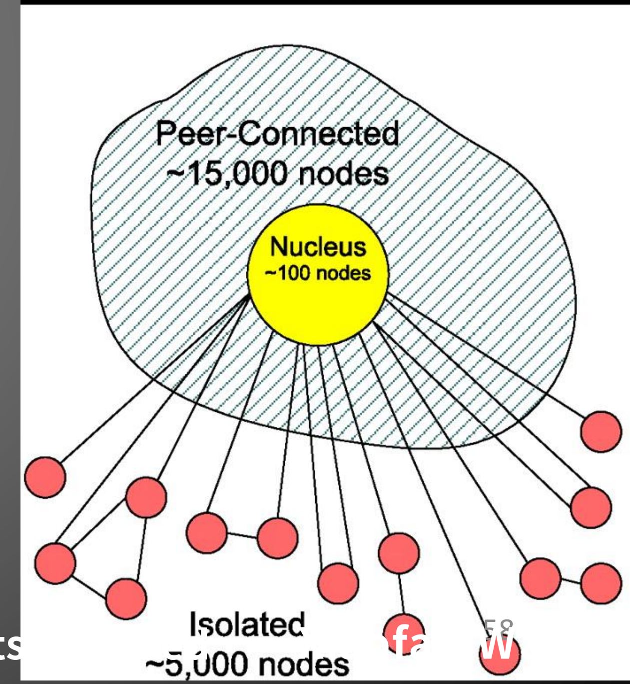
CARMI S, HAVLIN S, KIRKPATRICK S, et al. PNAS, 2007, 104(27): 11150-11155

k-shell Decomposition of the Internet

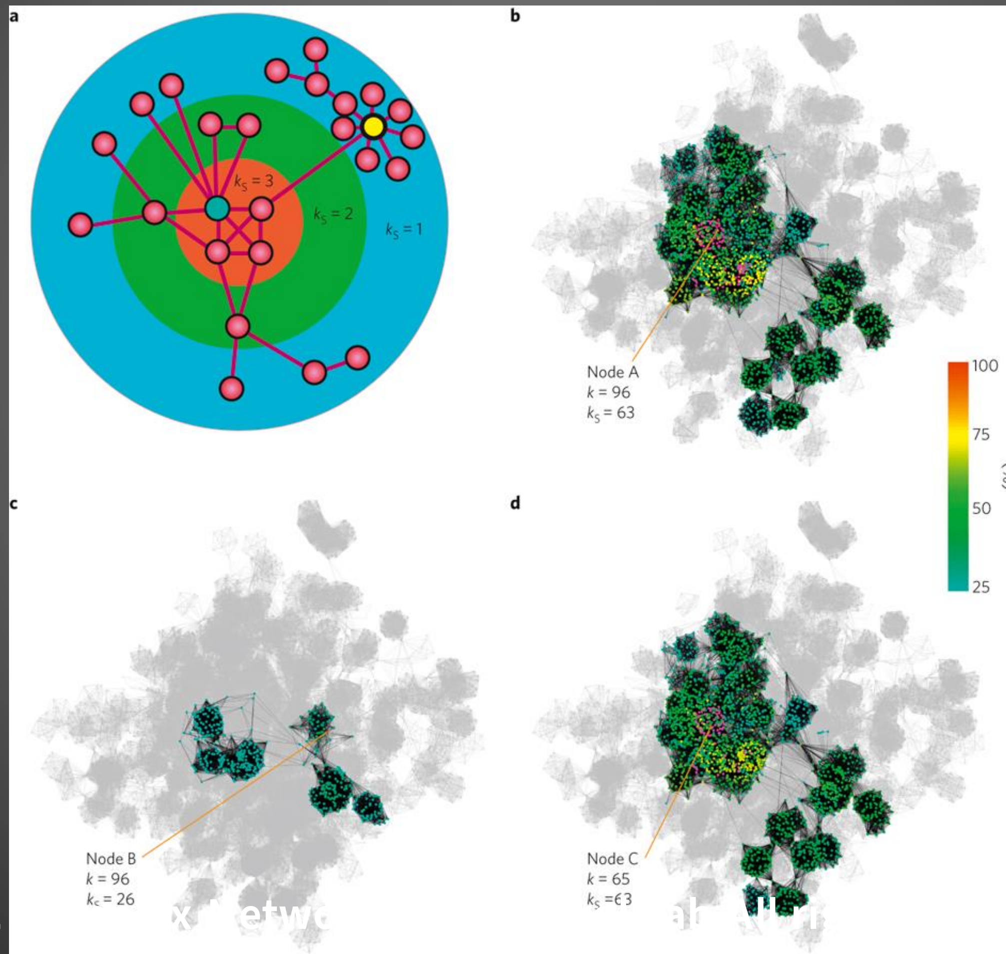
Nucleus:

- Unique, parameter-free, robust, easy to implement
- Degree ranged from >2,500 (ATT Worldnet) to as few as 50 carefully chosen neighbors, almost all within the nucleus (Google).
- The nucleus subgraph is redundantly connected, with diameter 2 and each node connected to ≈70% of the other nucleus nodes, which provides k_{max}-connectivity.



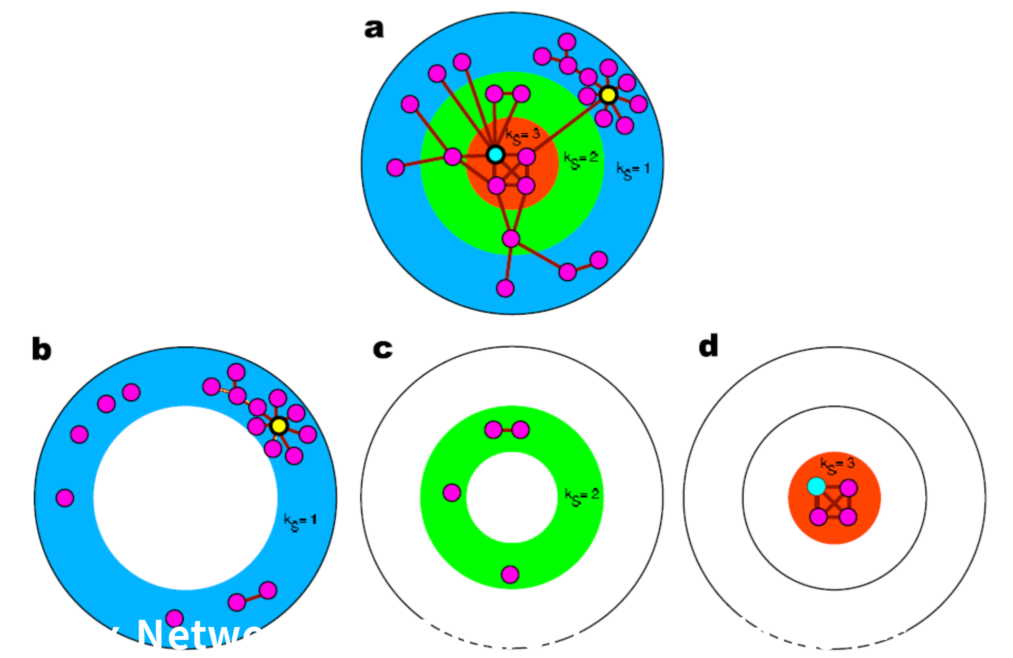


Kitsak M, Gallos L K, Havlin S, et al.
Identifying influential spreaders in complex networks.
***Nature Physics*, 2010, 6(11): 888-893**



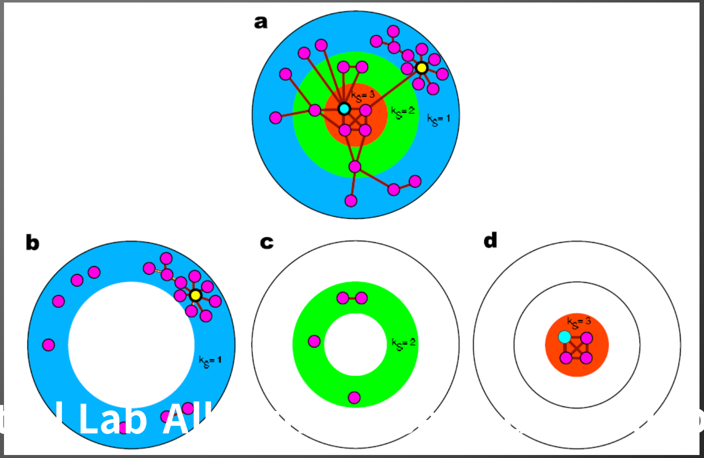
Problems with k-shell decomposition

- Start by removing all nodes with degree 1 only (with their links), until no more such nodes remain, and assign them to **the 1-shell.**
- In the same manner, recursively remove all nodes with degree<=k, creating the **k-shell.**



K-shell decomposition: Algorithm challenge

- Determine the k-shell index requires both global knowledge of the network topology and multiple iterations.
- Distributed k-shell decomposition achieved an 80 percent reduction in execution time, but still need iteration.
- A. Montresor, F. De Pellegrini, and D. Miorandi, “Distributed K-Core Decomposition,” *IEEE Trans. Parallel and Distributed Systems*, vol. 24, no. 2, 2013, pp. 288-300.



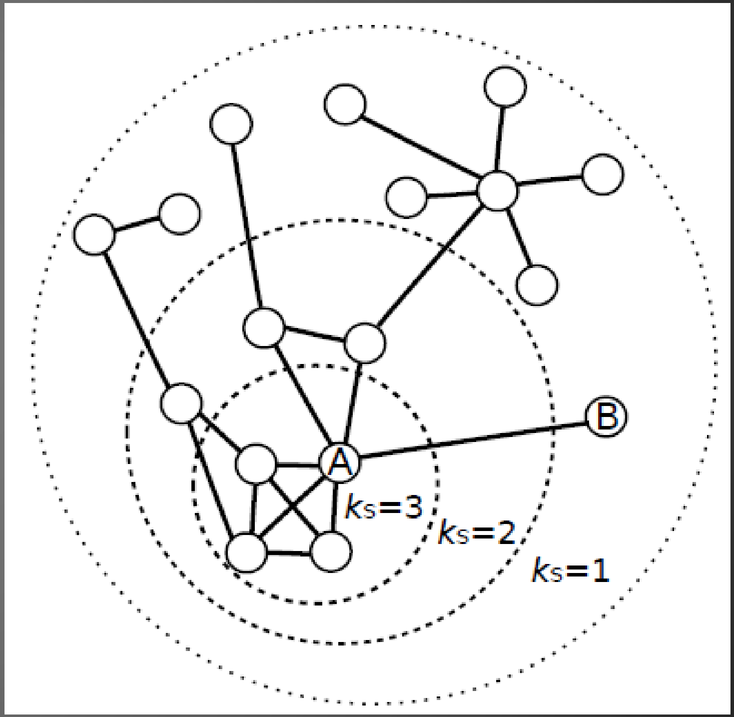
μ-PCI: A localized index

- μ-PCI of a node *v* is equal to *k*, such that there are up to μ $\times$ *k* nodes in the μ-hop neighborhood of *v* with degree >=*k*.
- The goal is to detect nodes located in dense areas of the network and thus likely influential spreaders.

- Basaras P, Katsaros, D., and Tassiulas L, Detecting Influential Spreaders in Complex, Dynamic Networks. IEEE Computer 46(4): 24-29 (2013)

A k-shell decomposition method for weighted networks

$$k_i' = \left[k_i^{\alpha} \left(\sum_{j}^{k_i} w_{ij} \right)^{\beta} \right]^{\frac{1}{\alpha+\beta}}$$



- $W_{AB}=3$
- $K_s(B)=2$

A. Garas, F. Schweitzer and S. Havlin, New J. Phys. 14 (2012) 083030

LeaderRank: Leaders in Social Networks, *the Delicious Case*

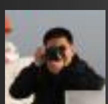
Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, Tao Zhou (2011), PLoS ONE 6(6): e21202

DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk

- Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G (2012). Scientific Reports, 2

Node Proximity & Link Prediction

- 新浪微博推荐：可能感兴趣的人
- 基本思想：两人的共同好友越多，两人就越相似

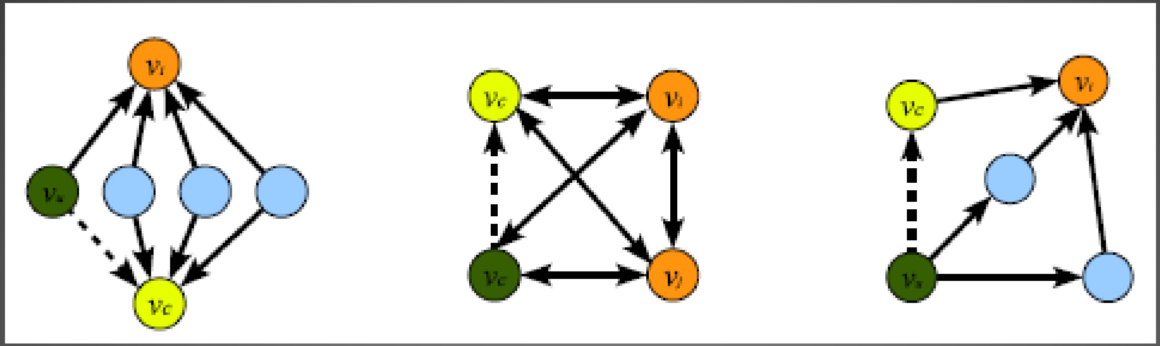


张鹏

我的好友中：谢耘耕、唐兴通、正结、王煜全、译言等7人也与他互相关注

我关注的人中：杜子建、vinW、龚斌Robin、段永朝、徐智明等16人也关注了他

Link Prediction in Microblogs



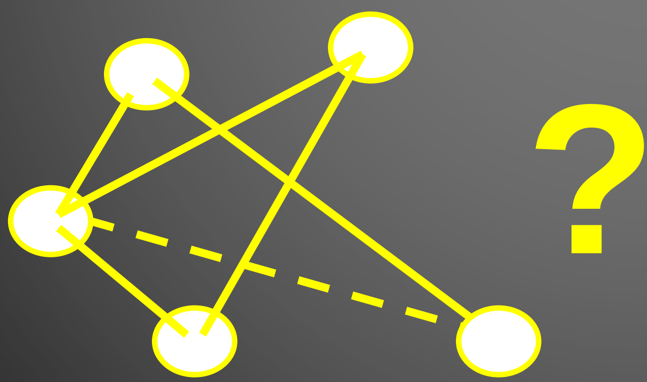
User *vu* may be interested in candidate *vc* because

- other similar users with *vu* are following *vc*.
- they may be friends in real life or other networks.
- *vu* is following other users which are following *vi* while *vc* is also following *vi*

Microbolgs calculate the probability that user *vu* follows user *vc*, rank candidate users in descending order, and recommend the top *N*

Link Prediction Problem

- Given a snapshot of a network at time t, we seek to predict the edges that will be added to the network during the interval (t, t')
- Based on "proximity" of nodes in a network
- **measures** of *proximity*



Link Prediction Verification

- Take a graph G=(V, E): G^T=(V, E^T), G^P=(V, E^P)
  E^P=(1, 3), (4, 5)
- Assign connection weight scores

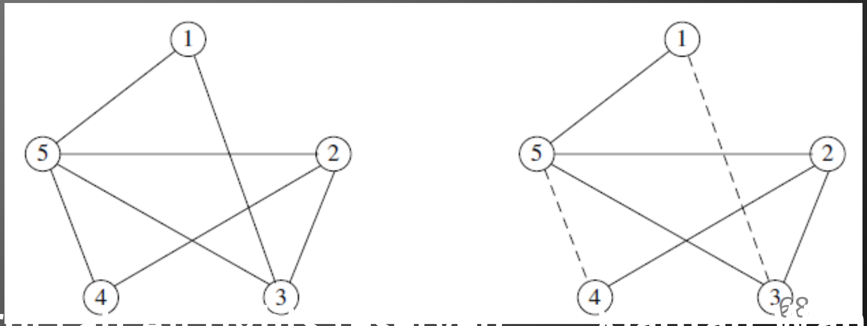
$$s_{12} = 0.4, s_{13} = 0.5, s_{14} = 0.6, s_{34} = 0.5, s_{45} = 0.6$$

- Verification

$$s_{13} > s_{12}, s_{13} < s_{14}, s_{13} = s_{34}, s_{45} > s_{12}, s_{45} = s_{14}, s_{45} > s_{34}$$

$$AUC = \frac{1}{6}(3\times1 + 2\times0.5) \approx 0.67$$

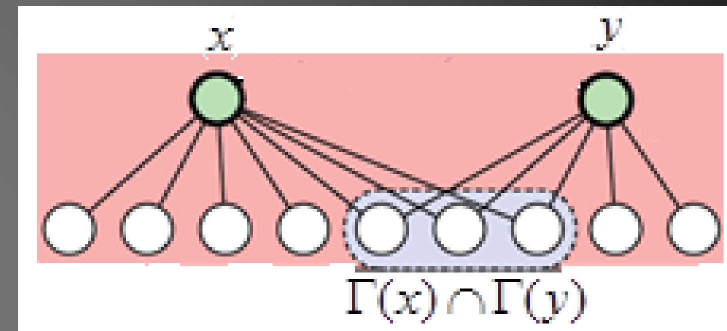
$$Precision = \frac{m}{L} = \frac{1}{2}$$



 Xiaofan Wang

Link Prediction based on Common Neighbors

$$s_{xy}^{CN} = \left|\Gamma(x) \cap \Gamma(y)\right|$$



$$s_{xy} = \frac{\left|\Gamma(x) \cap \Gamma(y)\right|}{\sqrt{k(x) \times k(y)}}$$

$$s_{xy} = \frac{\left|\Gamma(x) \cap \Gamma(y)\right|}{\left|\Gamma(x) \cup \Gamma(y)\right|}$$

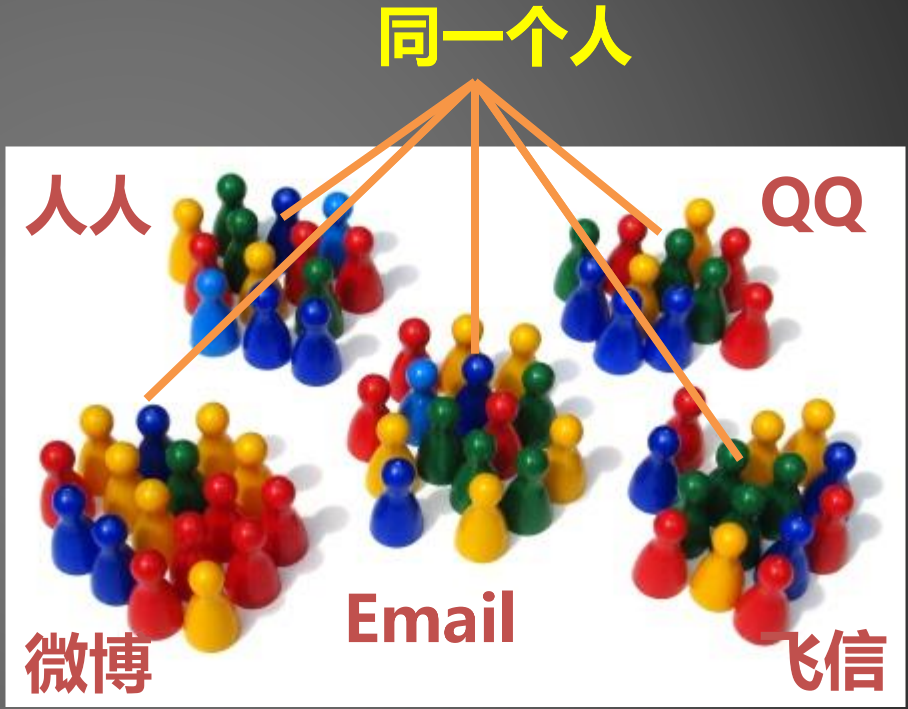
$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k(z)}$$

Adamic---Adar: weighting rarer neighbors more heavily

- Many other methods, but no single clear winner
- Many outperform the random predictor => there is useful information in the network topology

Network Alignment Problem

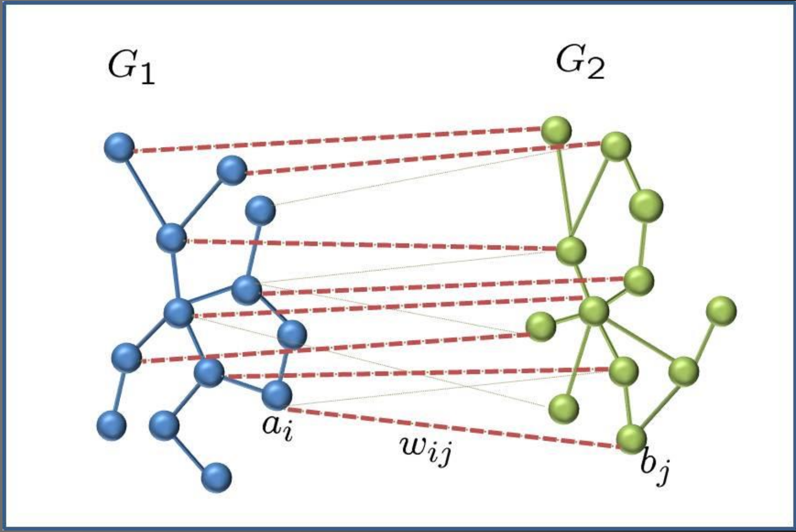
我们每一个人都出现在多个不同的网络中



 Xiaofan Wang

Network Alignment Problem

- How similar is each node in the first graph to each node in the second?
- constructing a similarity matrix W, where element $w_{i,j}$ denotes the similarity of node i in the first graph to node j in the second graph, depends on the specific measure of node similarity.



Thank You!

Xiaofan Wang

Shanghai Jiao Tong University

xfwang@sjtu.edu.cn