

Social Network Analysis

Chirayu Wongchokprasitti, PhD

University of Pittsburgh

Center for Causal Discovery

Department of Biomedical Informatics

chw20@pitt.edu

<http://www.pitt.edu/~chw20>

Overview

- Introduction
- Centrality
- Analysis techniques
 - PageRank
 - HITS
 - Community detection
- Example



Introduction

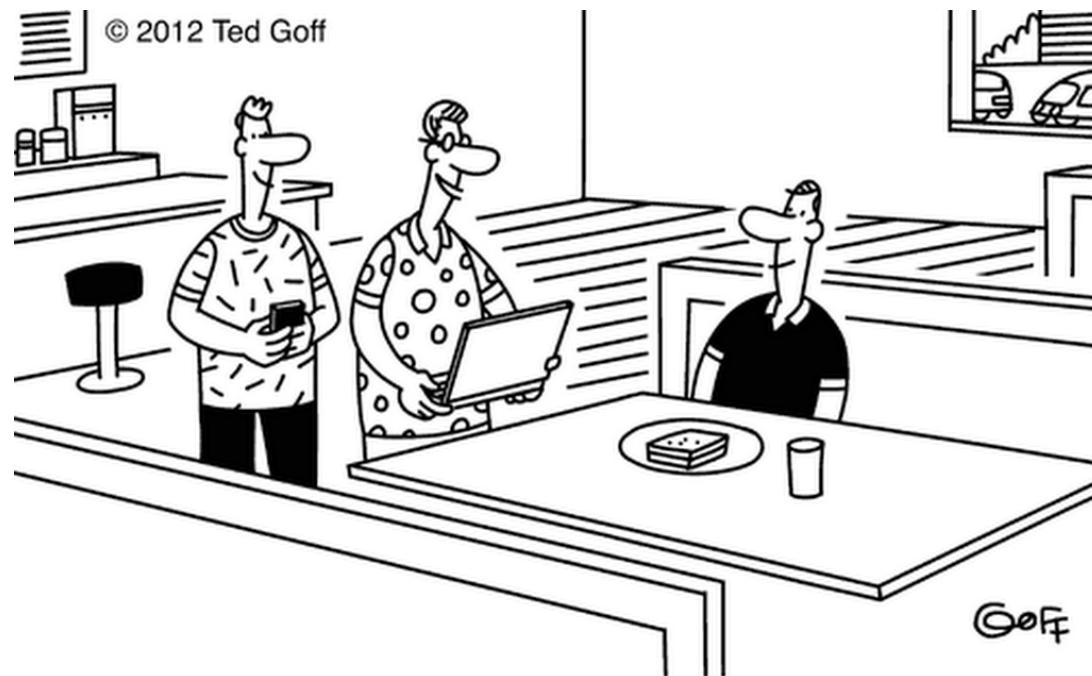
- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

What is a social network?

- **Study of social entities and their interactions and relationships**
- **These interactions could be represented as a network or graph**
- **Social networks could be analyzed using centrality and prestige**

- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

What is a social network?



“Twitter and Facebook can’t predict the election, but they did predict what you’re going to have for lunch: a tuna salad sandwich. You’re having the wrong sandwich.”

- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

Prominent online social networks



- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

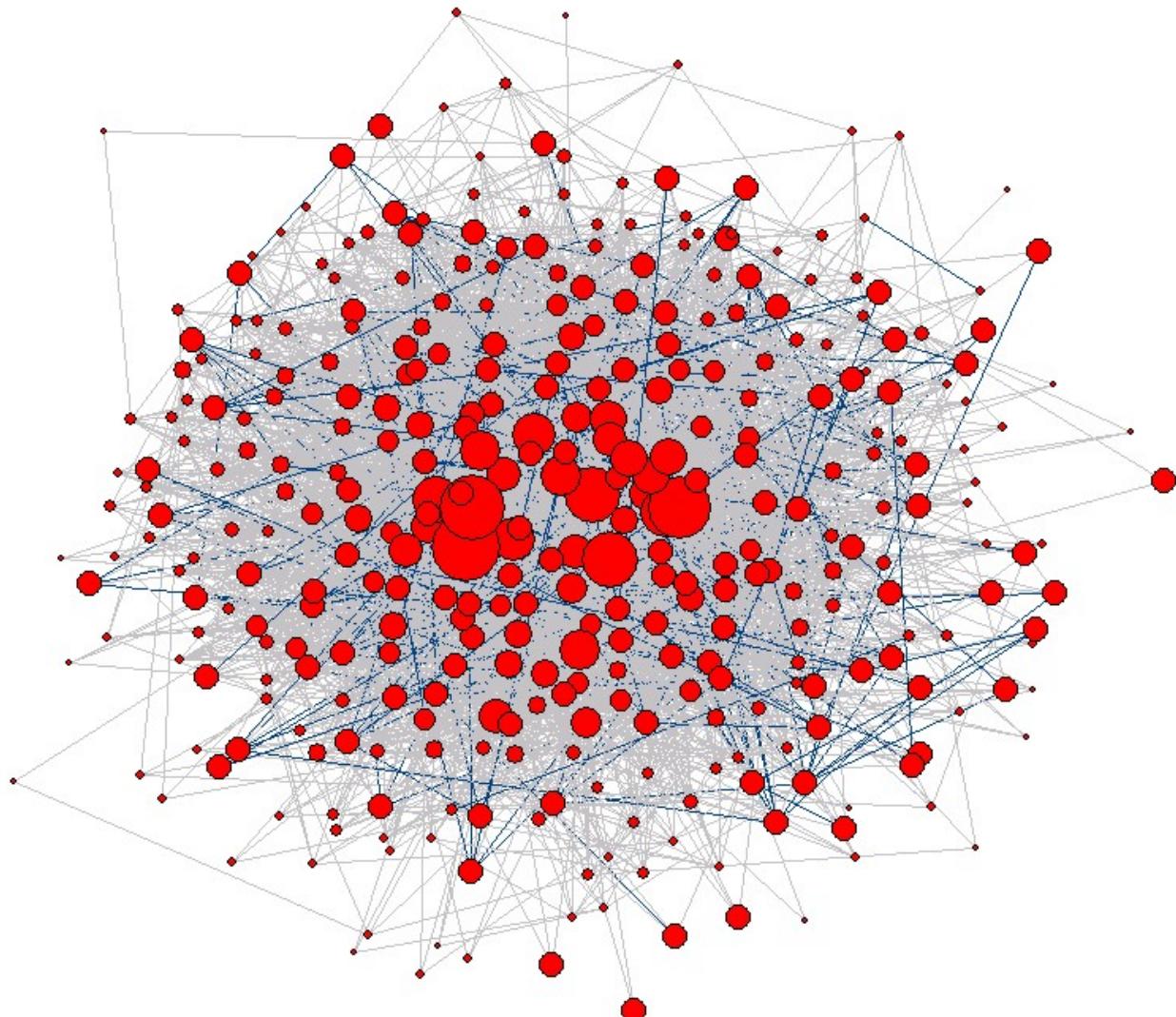
Applications of social networks

Typical applications of social network analysis and data mining:

- Detection of criminal activity, counter-terrorism, "homeland security," and intelligence
- Analysis of relationships within companies
- Sociological and anthropological studies
- Reciprocal trust schemes such as eBay ratings
- Recommended friends on Facebook
- Filter or recommend social media content
- ...

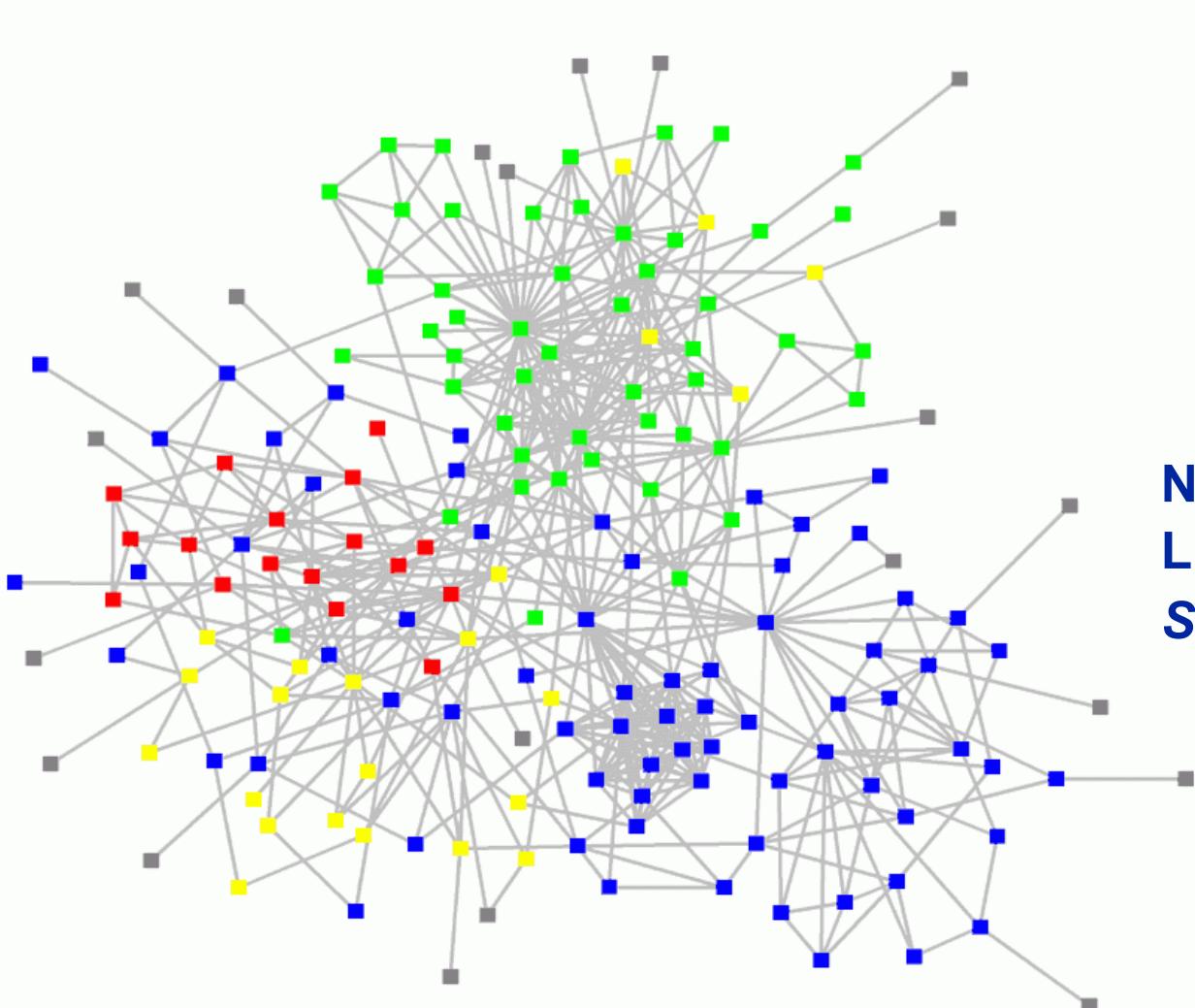
- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

Friendship network



- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

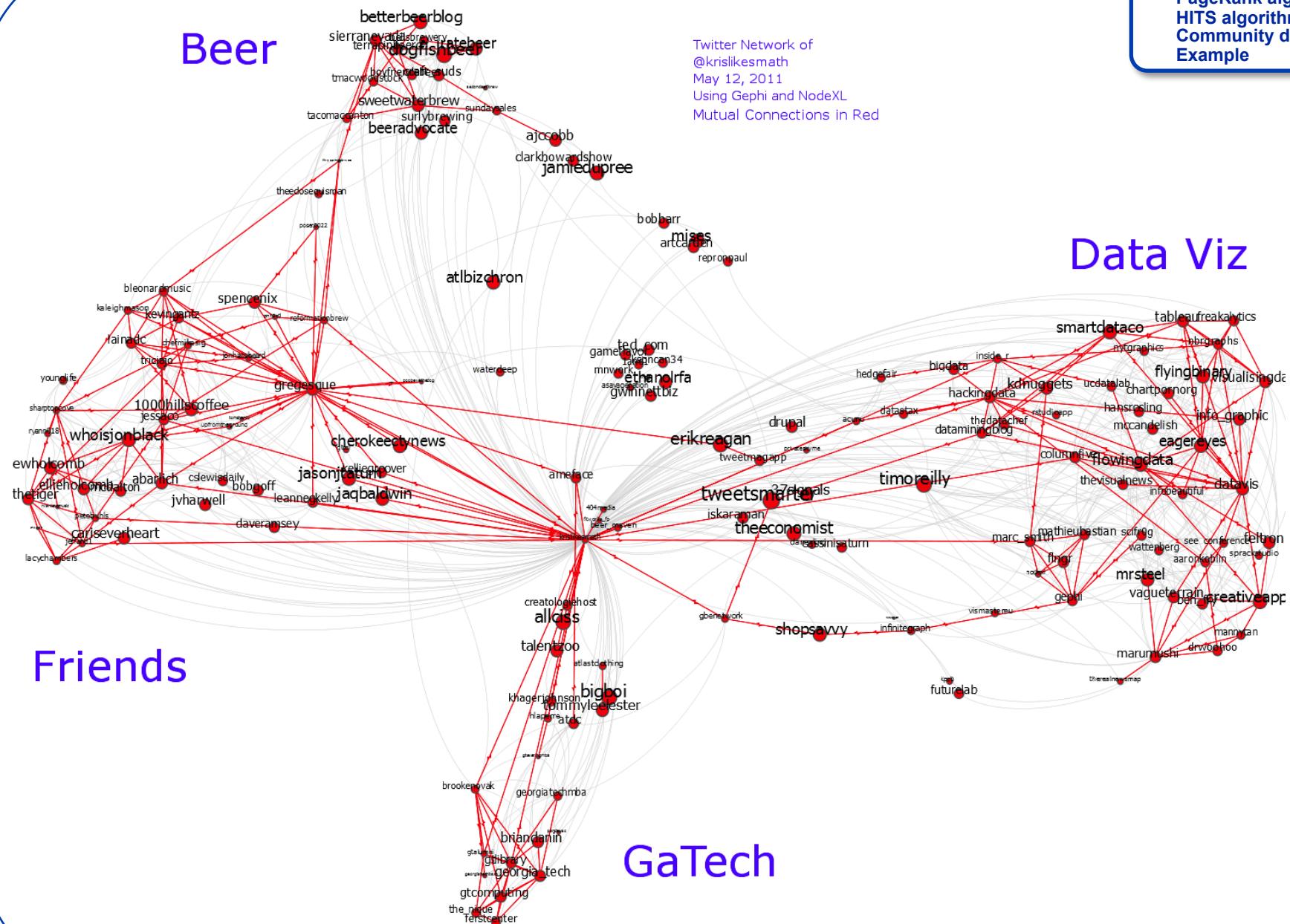
Email network



Nodes = People
Links = Emails
Source: orgnet.com

- Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
Example

Beer

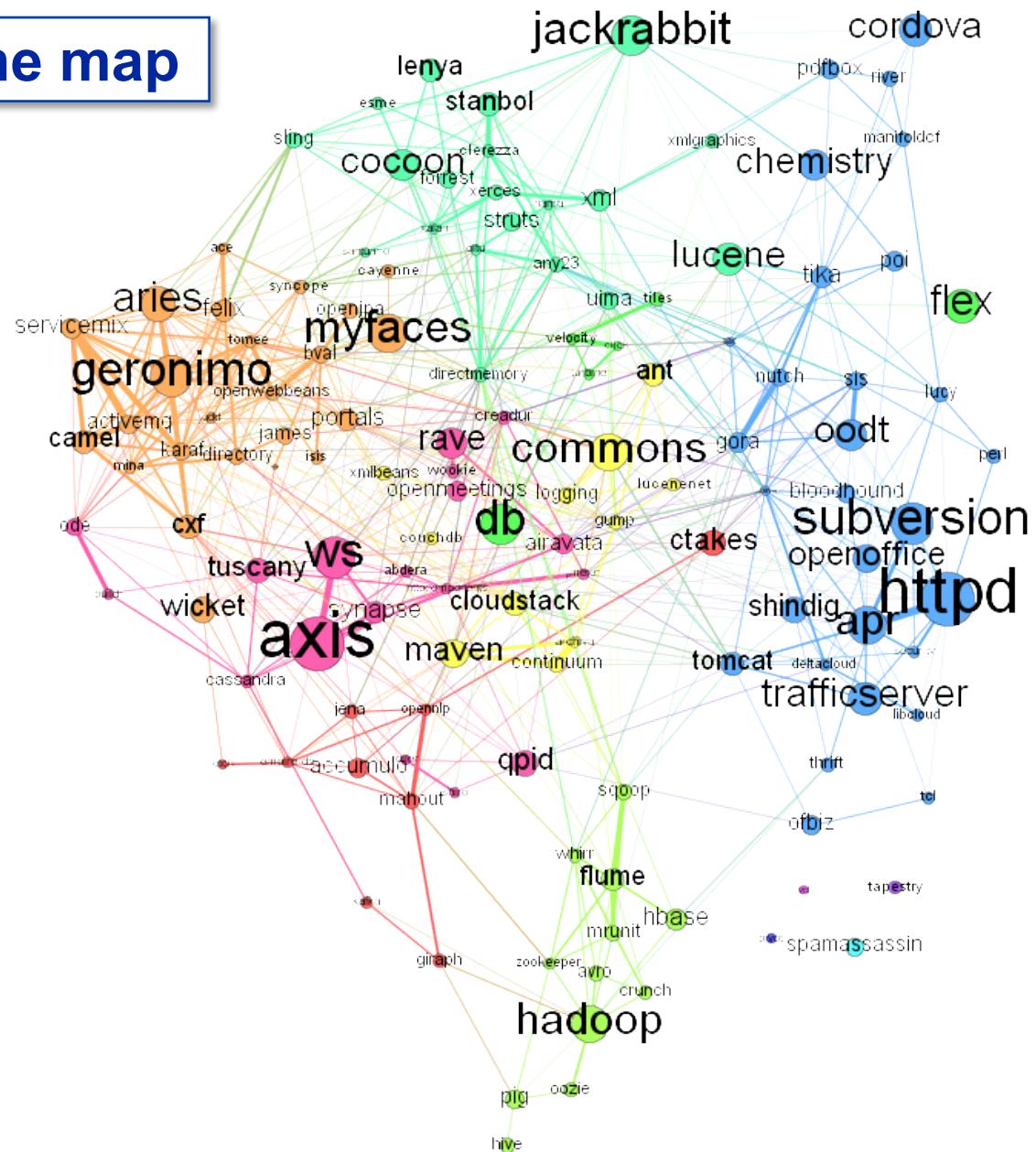


DSL
Decision Systems Laboratory

- Social Network Analysis

Apache map

- Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
Example



- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

A typical dataset for network analysis

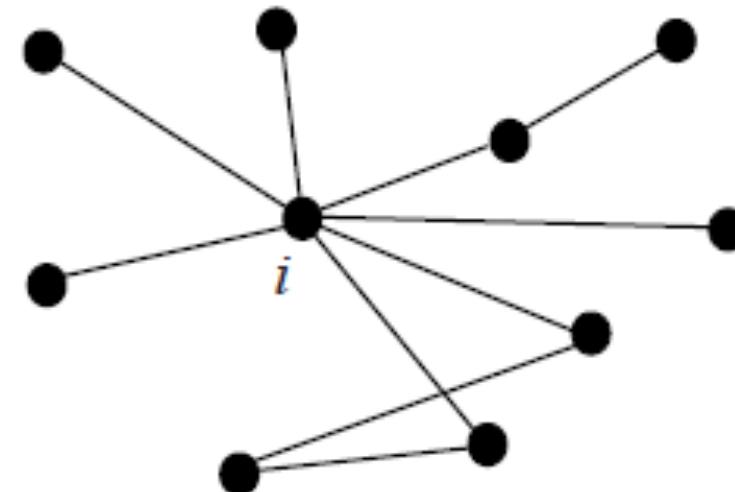
Sample

Year of publication	Data Set
2005	4 academic conferences, 500 participants, 3 years
2004	53 e-mail participants, 229 web-pages
2004	Buddy lists from LiveJournal, 25 days
2006	1 academic conference, 503 attendees,
2000	145 scientists, bibliography over 3 years
2000	1265 people, Friends listed on personal homepages in Stanford and MIT
2007	49897 photos from Flickr.com, 1015 days worth data
2000	108,676 academic papers from Citeseer, 13 years worth of data

Centrality

Centrality

- Assumption: Important Actors are Involved with Others Extensively.
- Each Node: An Actor
- Links (ties): Communication Between Actors



- Actor i is the most central actor in the above network fragment

Measuring network centrality

- In an undirected graph, the degree centrality of an actor i is given by

$$C_D(i) = \frac{d(i)}{n-1}.$$

$d(i)$ is the number of edges to the actor

$n-1$ is the maximum possible degree

Measuring network centrality

- In directed graphs, we distinguish between in-links and out-links
- In-links point towards the actor
- Out-links point way from the actor
- The degree of centrality here is based only on the out-degree (the number of out-links, $d_o(i)$)

$$C_D(i) = \frac{d_o(i)}{n-1}.$$

Closeness centrality

- Based on Closeness or Distance
- An actor is central if he can interact with all other actors
- It is a measure of HOW LONG it will take to get to all other nodes from a given node
- Useful in cases WHERE information transmission is of essence/interest
- Based on the distance measure between two actors, $d(i,j)$, defined as the shortest path between i and j

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

- Ranges between 0 and 1 (why?)
- Can be defined for directed graphs as well (needs to consider the direction of arcs)

Betweenness centrality

- Nodes that are located on communication paths between other nodes may have some control over the communication.
- Measures the control of an actor i over other pairs of actors
- Based on the number of shortest paths that pass through i divided by the number of all shortest paths in the network
- Betweenness could be computed even if the nodes are not connected

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

- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

Normalized betweenness centrality

The measure of betweenness can be normalized to the interval [0,1]

$$C'_B(i) = \frac{2 \sum_{j < k} \frac{p_{jk}(i)}{p_{jk}}}{(n-1)(n-2)}$$

Eigenvector centrality

- The main idea behind eigenvector centrality is that entities receiving many communications from other well connected entities, will be better and more valuable sources of information, and hence be considered central. The eigenvector centrality scores correspond to the values of the principal eigenvector of the adjacency matrix M .
- Formally, the vector v satisfies the equation

$$\lambda v = Mv$$

where λ is the corresponding eigenvalue and M is the adjacency matrix.

Prestige

Prestige

- Prestige is a more refined measure of prominence of an actor than centrality and deals with the importance of an actor in a network.
- Prestigious actor is the object of extensive ties as a recipient (has pervasive in-links).
- Prestige cannot be computed in undirected graphs because we are looking precisely at the direction of links
- The three measures of prestige are:
 - Degree prestige
 - Proximity, and
 - Rank prestige
- The third (rank prestige) forms the basis of most Web page link analysis algorithms, including PageRank and HITS

Degree prestige

- Degree prestige: Based on the number of incoming links to an actor i

$$P_D(i) = \frac{d_I(i)}{n-1},$$

- Normalized by the total possible number of incoming links
- Ranges between 0 and 1

Proximity prestige

- Proximity prestige generalizes prestige by considering both actors connected directly and indirectly to the actor i
- If I_i (called the influence domain of actor i) is the set of actors that can reach actor i , we can define proximity $d(j,i)$ as the shortest path distance from actor j to actor i
- Proximity prestige is based on the average distance, i.e.,

$$\frac{\sum_{j \in I_i} d(j,i)}{|I_i|}$$

where $|I_i|$ is the size of the set I_i

- Proximity prestige is based on the average distance, i.e.,

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

- It ranges from 0 to 1.0

Rank prestige

- An actor's prestige depends on the prestige of those actors that it is connected to

$$P_R(i) = A_{1i}P_R(1) + A_{2i}P_R(2) + \dots + A_{ni}P_R(n)$$

- This equation could be written in a matrix form

$$P = A^T P$$

- Web search algorithms are based on this equation
- These algorithms are PageRank and HITS

Graph statistics application

- **Closeness centrality**
 - Is this person central to the group?
 - Is your message likely to reach the audience?
- **Betweenness centrality**
 - Someone who has a high *betweenness centrality* is often a broker between others.
 - What happens if this person leaves the network?

PageRank Algorithm

PageRank algorithm

- **PageRank algorithm was developed by Brin and Page (Google founders) around 1997**
- **Exploits the hyperlink structure of the Web to rank pages according to their levels of “prestige” or “authority.”**
- **Emerged as the dominant link analysis model for web search (the reasons could be: query-independent evaluation of Web pages, ability to combat spamming, and Google’s business success ☺)**
- **Relies on the web’s vast link structure as an indicator of the quality of a page.**
- **Does not only accumulate the number of links to a page but also weight of those links**

PageRank algorithm

- **In-links of a page i :** Hyperlinks that points to the page from other pages
- **Out-links:** Hyperlinks that point to other pages, links to pages on the same sites are not included
- **Hyperlink from a page pointing to another page conveys authority to the target page**
- **A page with a higher prestige score pointing to a page i is more important than a page with a lower prestige score**

PageRank algorithm

- The importance of a page is determined by the sum of all PageRank scores of pages pointing to it.
- The prestige score of a page should be shared pages that it points to.
- The web is assumed to be a directed graph $G=(V,E)$, where V is the set of vertices and E is the set of directed edges.
- Hyperlinks are edges and web pages are the nodes.
- The PageRank score of page i (denoted by $P(i)$) is defined as

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

where O_j is the number of out-links of page j .

Graphs and networks

Basic Definitions

- **Graph $G = (V, E)$**
 - V : set of vertices / nodes
 - $E \subseteq V \times V$: set of edges
- **Adjacency matrix (sociomatrix)**
alternative representation of a graph

$$A_{i,j} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases}$$

- **Network:** used as a synonym to graph, a more application-oriented term

Graphs and networks

- We are dealing with a system of n equations with n unknowns:

$$P = A^T P .$$

- The solution to P is an eigenvector with the corresponding eigenvalue of 1
- A mathematical technique called power iteration could be used to find the P
- Alternatively, an enhanced form of the equation can be derived by means of Markov chains

Markov chain formulation of the Web

- Each web page or a node in the web graph is regarded as state
- A hyperlink is a transition which leads from one state to another with a state transition probability
- This models the web as a stochastic process
- Each transition probability is given by $1/k$, where k is the number of out-links from page i
- These transition probabilities compose into a state transition matrix

- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

Markov chain formulation of the Web

Transition matrix

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1n} \\ A_{21} & A_{22} & \cdot & \cdot & \cdot & A_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ A_{n1} & A_{n2} & \cdot & \cdot & \cdot & A_{nn} \end{pmatrix}$$

A_{ij} represents the probability that a surfer on page i will go to page j

Markov chain formulation of the Web

Given an initial probability vector that a surfer is on page

$$p_k = A^T p_{k-1}.$$

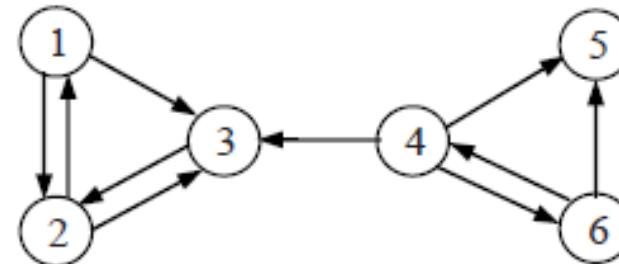
In general, the probabilities after k page transitions are given as

$$p_0 = (p_0(1), p_0(2), \dots, p_0(n))^T$$

After a series of transitions, the p_k will converge to a steady state probability vector p_i , regardless of the initial probability vector p_0

Markov chain formulation of the Web: Example

Hyperlink graph



The corresponding A matrix

Shows the probability of moving

From page i to page j

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

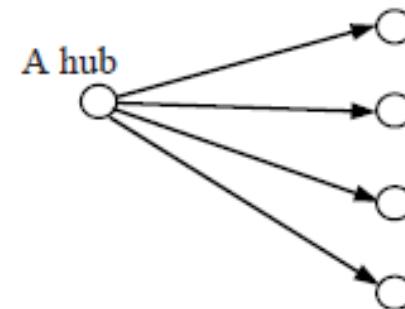
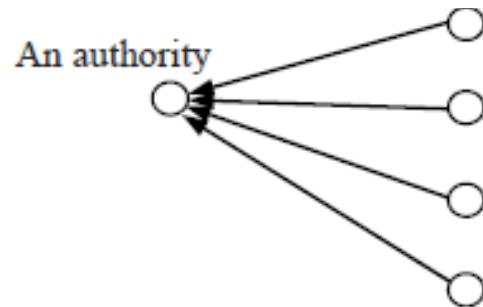
HITS Algorithm

Hyperlink Induced Topic Search (HITS)

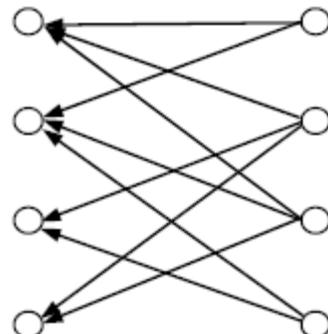
- Developed around 1998 by Jon Kleinberg
- Like PageRank, it exploits the hyperlink structure of the Web to rank pages according to their levels of “prestige” or “authority.”
- Unlike PageRank, HITS is search query-dependent
- HITS produces two rankings of expanded sets of pages, authority and hub ranking
- <http://www.cs.cornell.edu/home/kleinber/auth.pdf>

HITS algorithm: Bipartite graph representation of web pages

An authority page and a hub page



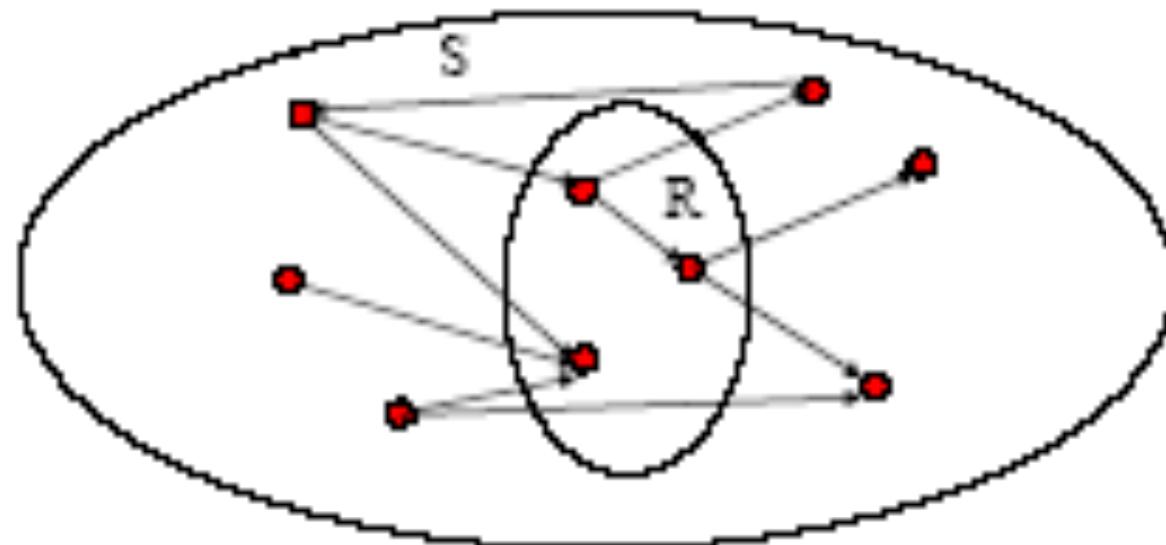
A densely linked set of hubs and authorities



- Authority is a page with many in-links
- A hub is a page with many out-links
- User's can get more information about other topics or pages when they visit a hub
- The idea of a hub is that a good hub points to good authorities and a good authority is pointed to by a good hub

HITS algorithm

- Determines a *base set S*
- Let set of documents returned by a standard search engine (in the original paper that they published it was 200 documents) be called the *root set R*
- Initialize *S* to *R*



HITS algorithm

- Add to S all pages pointed to by any page in R .
- Add to S all pages that point to any page in R .
- Maintain for each page p in S :

Authority score: ap (vector a)

Hub score: hp (vector h)

HITS algorithm

- For each node initialize the ap and hp to $1/n$
- In each iteration calculate the authority weight for each node in S

$$a(i) = \sum_{(j,i) \in E} h(j)$$

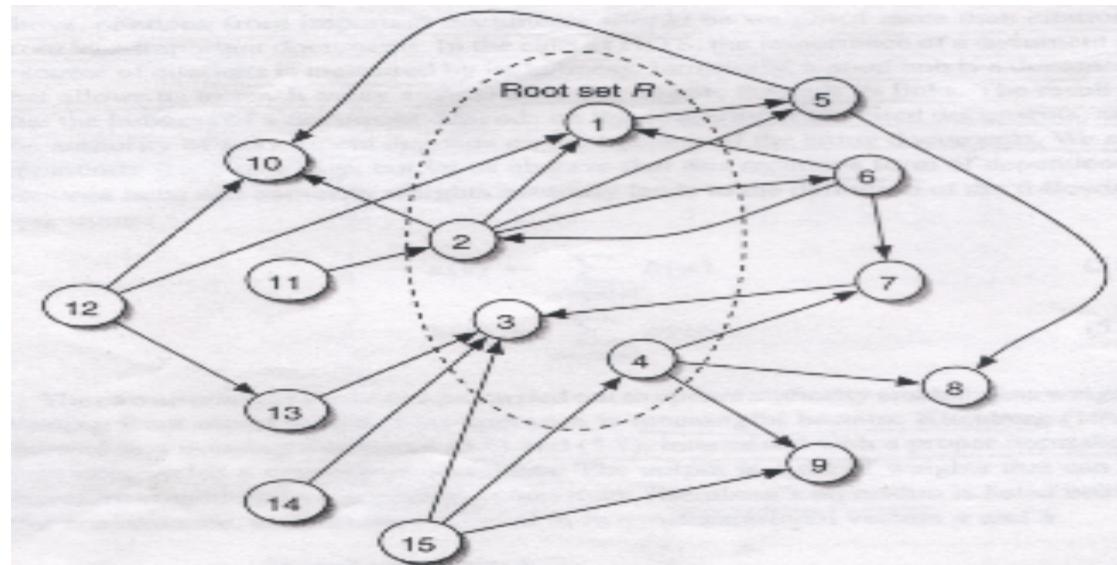
$$h(i) = \sum_{(i,j) \in E} a(j)$$

- Please note that the two are mutually reinforcing each other!

Convergence of HITS

- Let A be an adjacency matrix of S
- $A_{ij}=1$ for $i \in S, j \in S$ if and only if $i \rightarrow j$
- Authority and hub:
- $a_k = A^T A a_{k-1}$; $h_k = A A^T h_{k-1}$
- Iterate until
 $|a_k - a_{k-1}|$ and $|h_k - h_{k-1}|$ become smaller than a pre-set epsilon value

HITS algorithm: example



Root set R {1,2,3,4}
Extend it to form the base set S

Strengths and weaknesses of HITS

- HITS does not have anti-spam capability of PageRank
- Easy to influence by the addition of out-links to one's own page.
- Topic drift, in expanding the root it is possible to capture hub topics that are not related to the main topic.
- Getting the root set and then performing eigenvector computations are all drawbacks and time consuming operations

Community Detection

Community detection

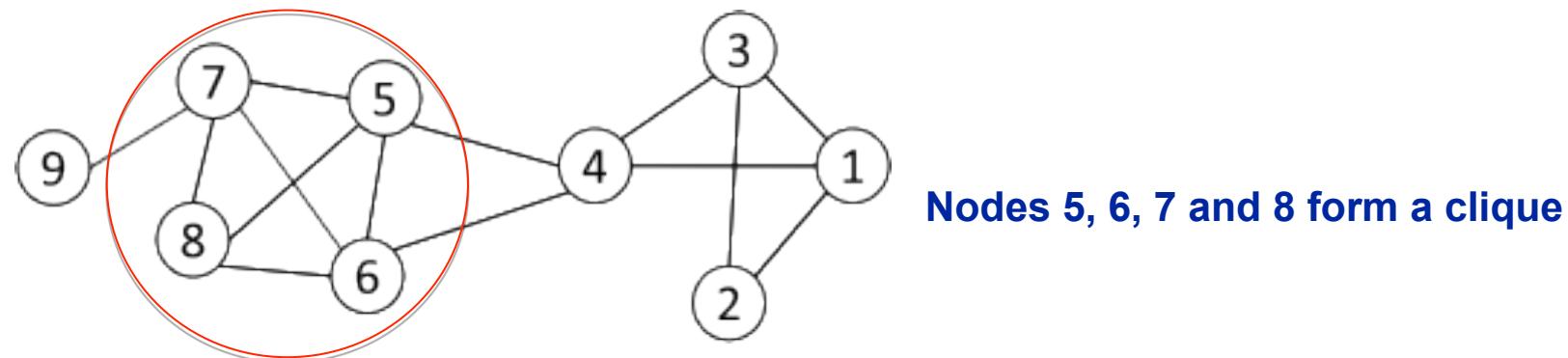
- Community detection methods can be divided into four non-exclusive categories:
- Node-centric community
 - Each node in a group satisfies certain properties
- Group-centric community
 - Consider the connections within a group as a whole. The group has to satisfy certain properties without zooming into node-level
- Network-centric community
 - Partition the whole network into several disjoint sets
- Hierarchy-centric community
 - Construct a hierarchical structure of communities

Node-centric community detection

- Nodes satisfy different properties
 - Complete Mutuality
 - » cliques
 - Reachability of members
 - » k-clique, k-clan, k-club
 - Nodal degrees
 - » k-plex, k-core
 - Relative frequency of Within-Outside Ties
 - » LS sets, Lambda sets
- Commonly used in traditional social network analysis
- Here, we discuss some representative ones

Complete mutuality: Cliques

- **Clique: a maximum complete subgraph in which all nodes are adjacent to each other**



- NP-hard to find the maximum clique in a network
- Straightforward implementation to find cliques is very expensive in time complexity

Group-centric community detection: Density-based groups

- The group-centric criterion requires the whole group to satisfy a certain condition
 - E.g., the group density \geq a given threshold
- A subgraph $G_s(V_s, E_s)$ is a γ -dense quasi-clique if

$$\frac{2|E_s|}{|V_s|(|V_s| - 1)} \geq \gamma ,$$

where the denominator is the maximum number of degrees.

- A similar strategy to that of cliques can be used
 - Sample a subgraph, and find a maximal γ -dense quasi-clique (size $|V_s|$)
 - Remove nodes with degree less than the average degree

$$|V_s|\gamma \leq \frac{2|E_s|}{|V_s|-1}$$

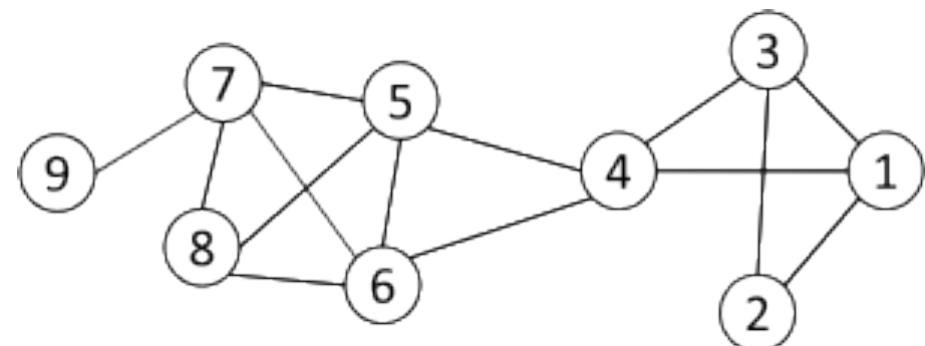
Network-centric community detection

- Network-centric criterion needs to consider the connections within a network globally
- Goal: partition nodes of a network into disjoint sets
- Approaches:
 - (1) Clustering based on vertex similarity
 - (2) Latent space models (multi-dimensional scaling)
 - (3) Block model approximation
 - (4) Spectral clustering
 - (5) Modularity maximization

Clustering based on vertex similarity

- Apply k-means or similarity-based clustering to nodes
- Vertex similarity is defined in terms of the similarity of their neighborhood
- Structural equivalence: two nodes are structurally equivalent iff they are connecting to the same set of actors

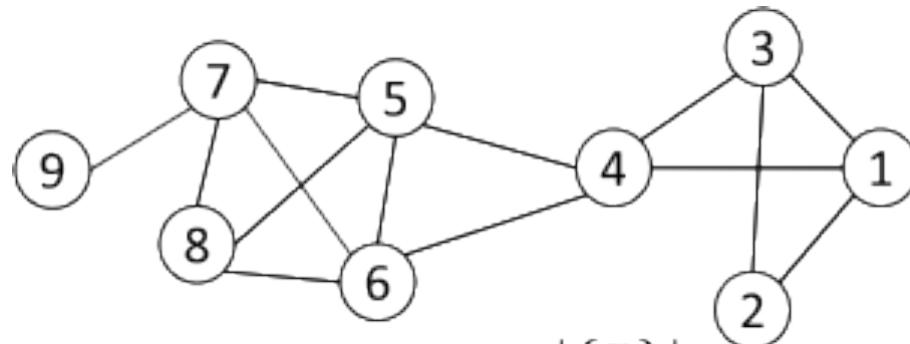
Nodes 1 and 3 are
structurally equivalent;
So are nodes 5 and 6.



- Structural equivalence is too restrictive for practical use.

Vertex similarity

- **Jaccard Similarity** $Jaccard(v_i, v_j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$
- **Cosine similarity** $Cosine(v_i, v_j) = \frac{|N_i \cap N_j|}{\sqrt{|N_i| \cdot |N_j|}}$



$$Jaccard(4, 6) = \frac{|\{5\}|}{|\{1, 3, 4, 5, 6, 7, 8\}|} = \frac{1}{7}$$

$$\cosine(4, 6) = \frac{1}{\sqrt{4 \cdot 4}} = \frac{1}{4}$$

Hierarchy-centric community detection

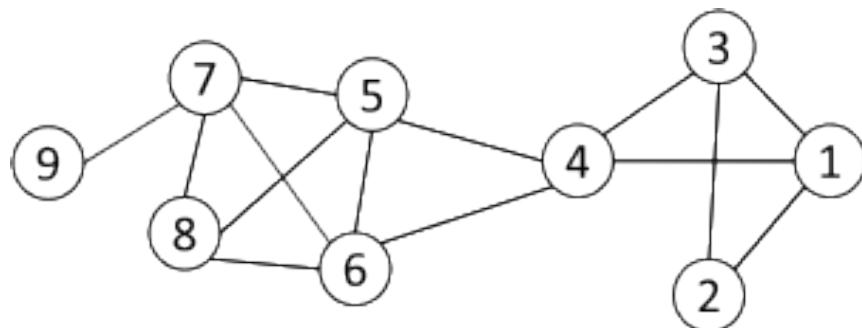
- Goal: build a hierarchical structure of communities based on network topology
- Allow the analysis of a network at different resolutions
- Representative approaches:
 - Divisive Hierarchical Clustering (top-down)
 - Agglomerative Hierarchical clustering (bottom-up)

Divisive hierarchical clustering

- **Divisive clustering**
 - Partition nodes into several sets
 - Each set is further divided into smaller ones
 - Network-centric partition can be applied for the partition
- One particular example: recursively remove the “weakest” tie
 - Find the edge with the least strength
 - Remove the edge and update the corresponding strength of each edge
- Recursively apply the above two steps until a network is decomposed into desired number of connected components.
- Each component forms a community

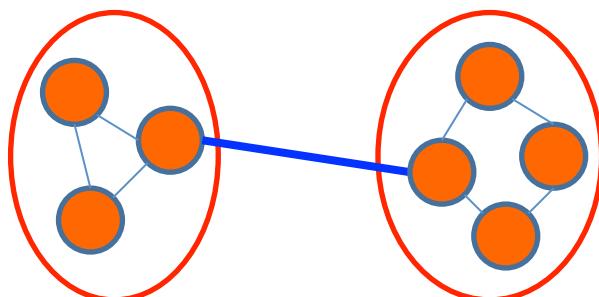
Edge betweenness

- The strength of a tie can be measured by edge betweenness
- Edge betweenness: The number of shortest paths that pass along with the edge



The edge betweenness of $e(1, 2)$ is 4 ($=6/2 + 1$), as all the shortest paths from 2 to $\{4, 5, 6, 7, 8, 9\}$ have to either pass $e(1, 2)$ or $e(2, 3)$, and $e(1,2)$ is the shortest path between 1 and 2

- The edge with higher betweenness tends to be the bridge between two communities.



Social Network Analysis

Semantic Web and social networks

- **Semantic Web:** having data on the Web defined and linked in a way that it can be used by people and processed by machines in a “wide variety of new and exciting applications”
- **SW and SN models support each other:**
 - Semantic Web enables online and explicitly represented social information
 - social networks, especially trust networks, provide a new paradigm for knowledge management in which users “outsource” knowledge and beliefs via their social networks

Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
Example

SW and SNA issues

- Knowledge representation.
 - Small number of common ontologies
- Knowledge management.
 - efficient and effective mechanisms for accessing knowledge, especially social networks, on the Semantic Web
- Social network extraction, integration and analysis
 - extracting social networks correctly from the noisy and incomplete knowledge on the (Semantic) Web
- Provenance and trust aware distributed inference.
 - manage and reduce the complexity of distributed inference by utilizing provenance of knowledge

Semantic Web and social networks

- **Drawbacks to Centralized Social Networks**
 - the information is under the control of the database owner
 - centralized systems do not allow users to control the information they provide on their own terms
- **The friend-of-a-friend (FOAF) project is a first attempt at a formal, machine processable representation of user profiles and friendship networks.**

Semantic Web and social networks

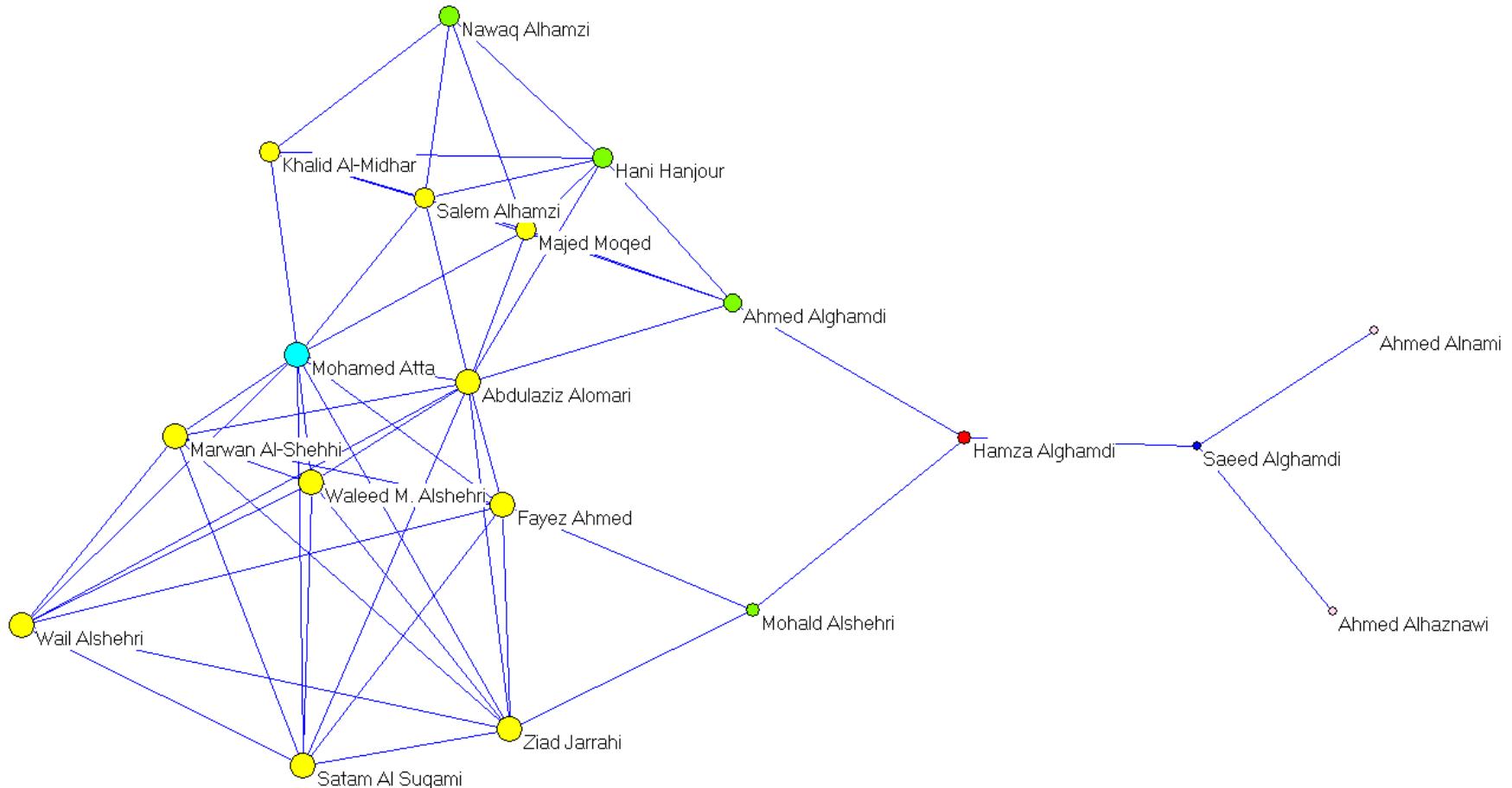
- The Swoogle Ontology Dictionary shows that the class *foaf:Person* currently has nearly one million instances spread over about 45,000 Web documents.
- The FOAF ontology is not the only one used to publish social information on the Web.
- For example, Swoogle identifies more than 360 RDFS or OWL classes defined with the local name “person.”

Example: 11 September 2001 attack Graphs

source: R. Feldman, Bar Ilan University

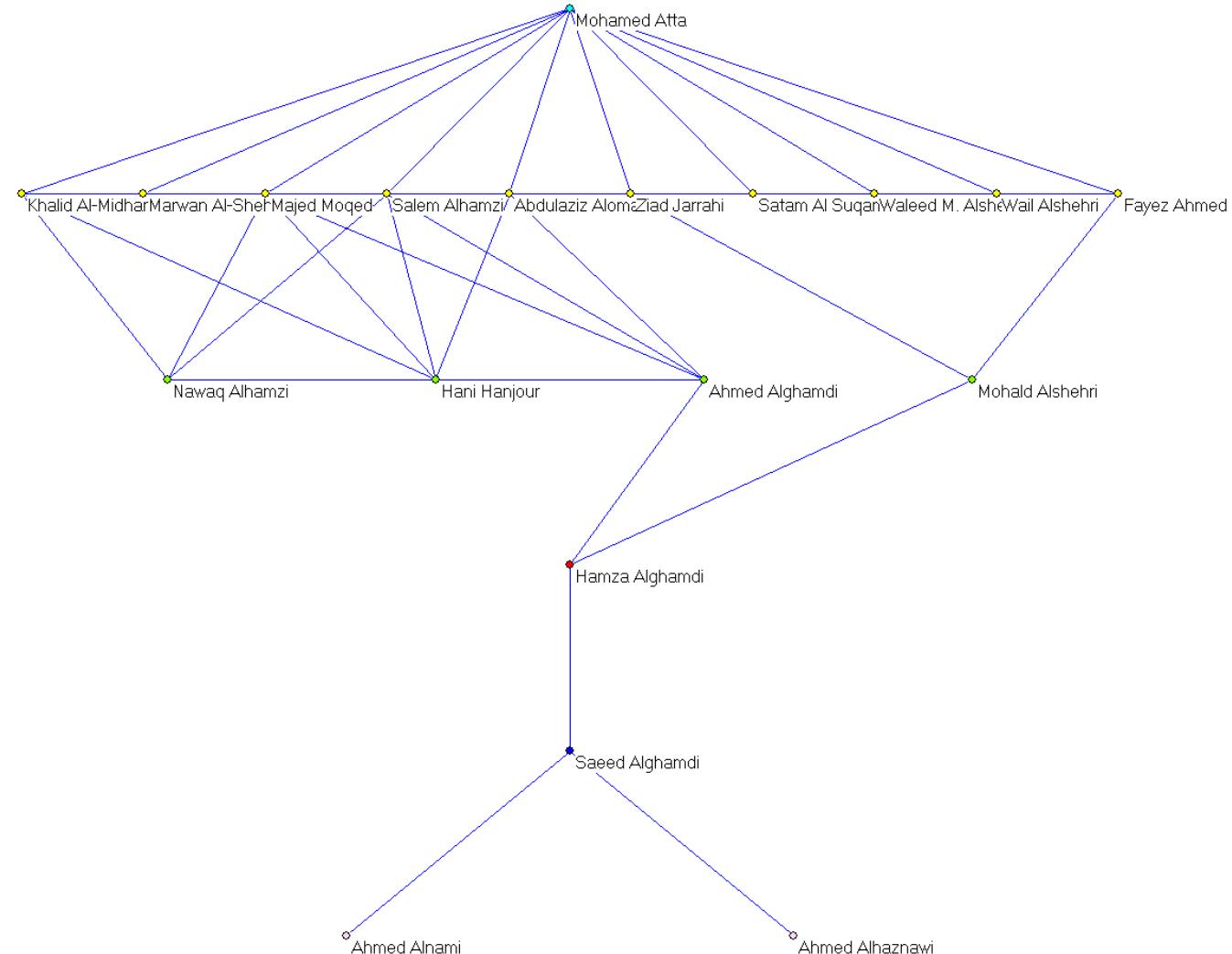
Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
Example

Finding the shortest path (from Mohamed Atta)



Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
● Example

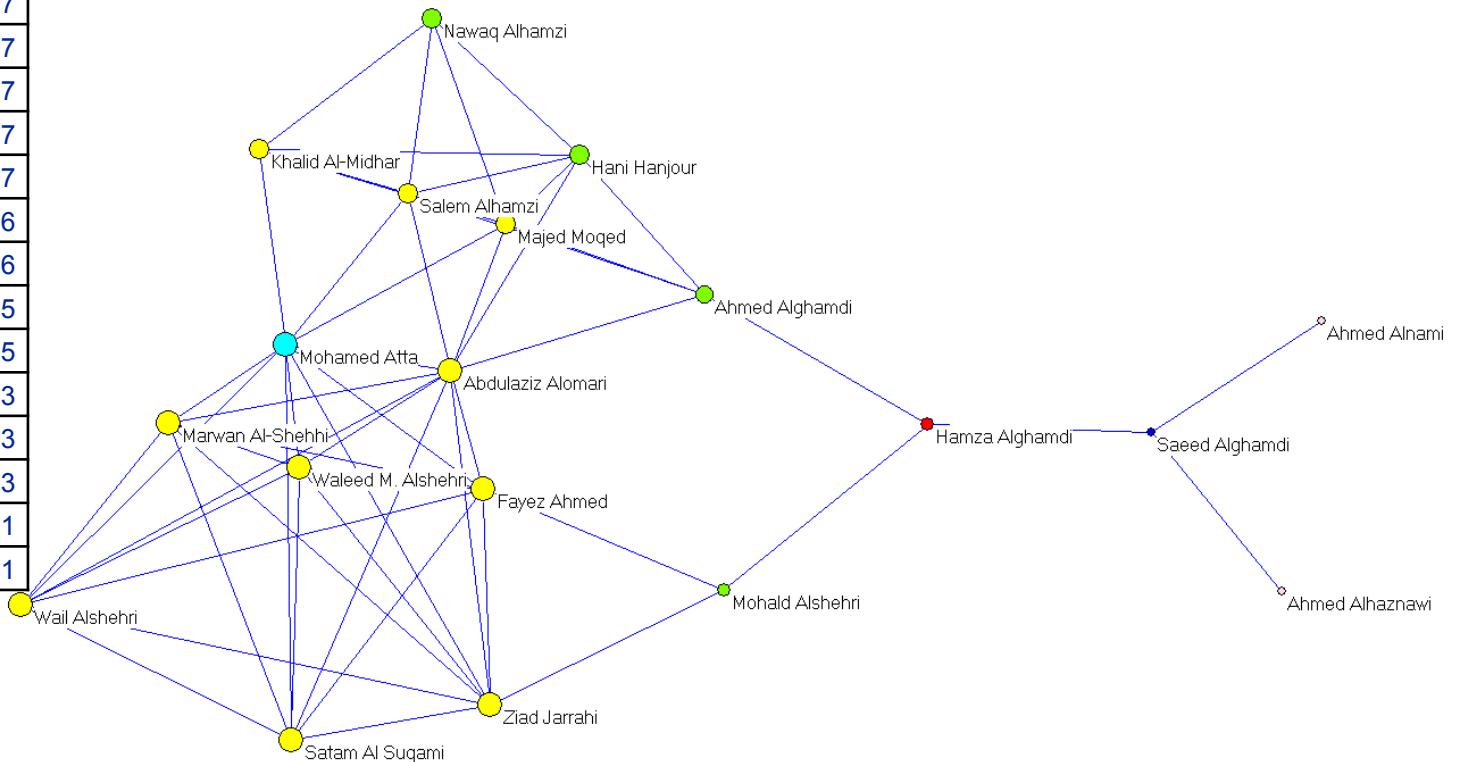
A better visualization



Introduction
 Centrality & prestige
 PageRank algorithm
 HITS algorithm
 Community detection
 ● Example

Centrality: Degree of the hijackers

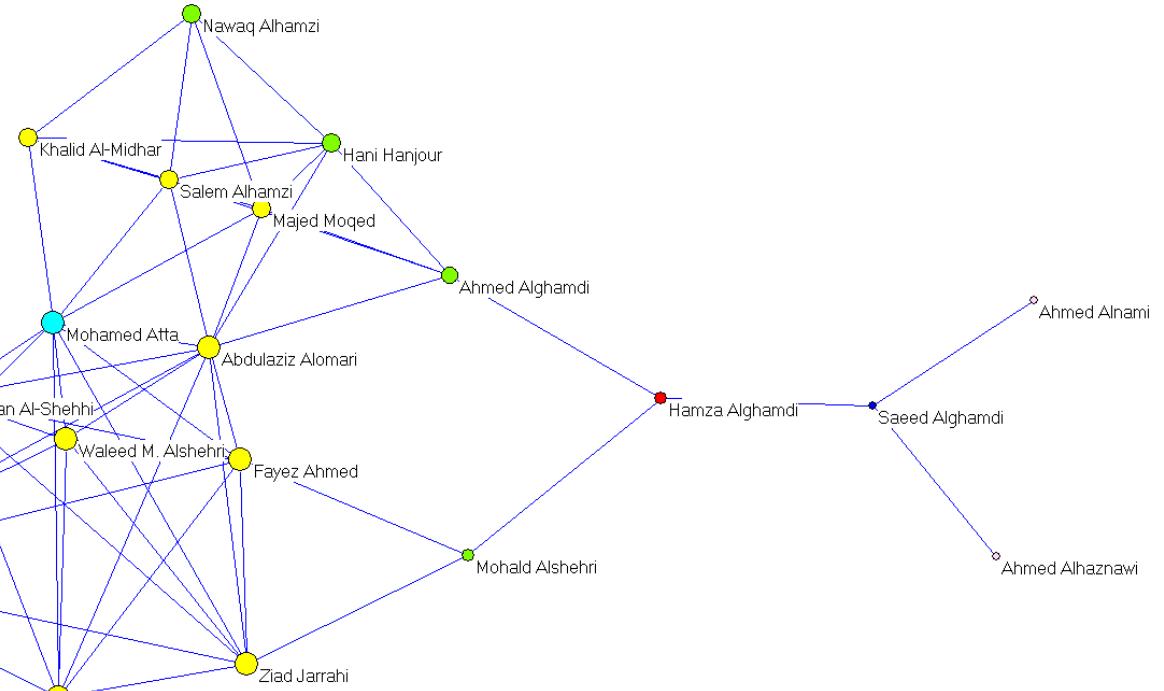
Name	Degree
Mohamed Atta	11
Abdulaziz Alomari	11
Ziad Jarrahi	9
Fayez Ahmed	8
Waleed M. Alshehri	7
Wail Alshehri	7
Satam Al Suqami	7
Salem Alhamzi	7
Marwan Al-Shehhi	7
Majed Moqed	7
Khalid Al-Midhar	6
Hani Hanjour	6
Nawaq Alhamzi	5
Ahmed Alghamdi	5
Saeed Alghamdi	3
Mohald Alshehri	3
Hamza Alghamdi	3
Ahmed Alnami	1
Ahmed Alhaznawi	1



- Introduction
- Centrality & prestige
- PageRank algorithm
- HITS algorithm
- Community detection
- Example

Centrality: Closeness of the hijackers

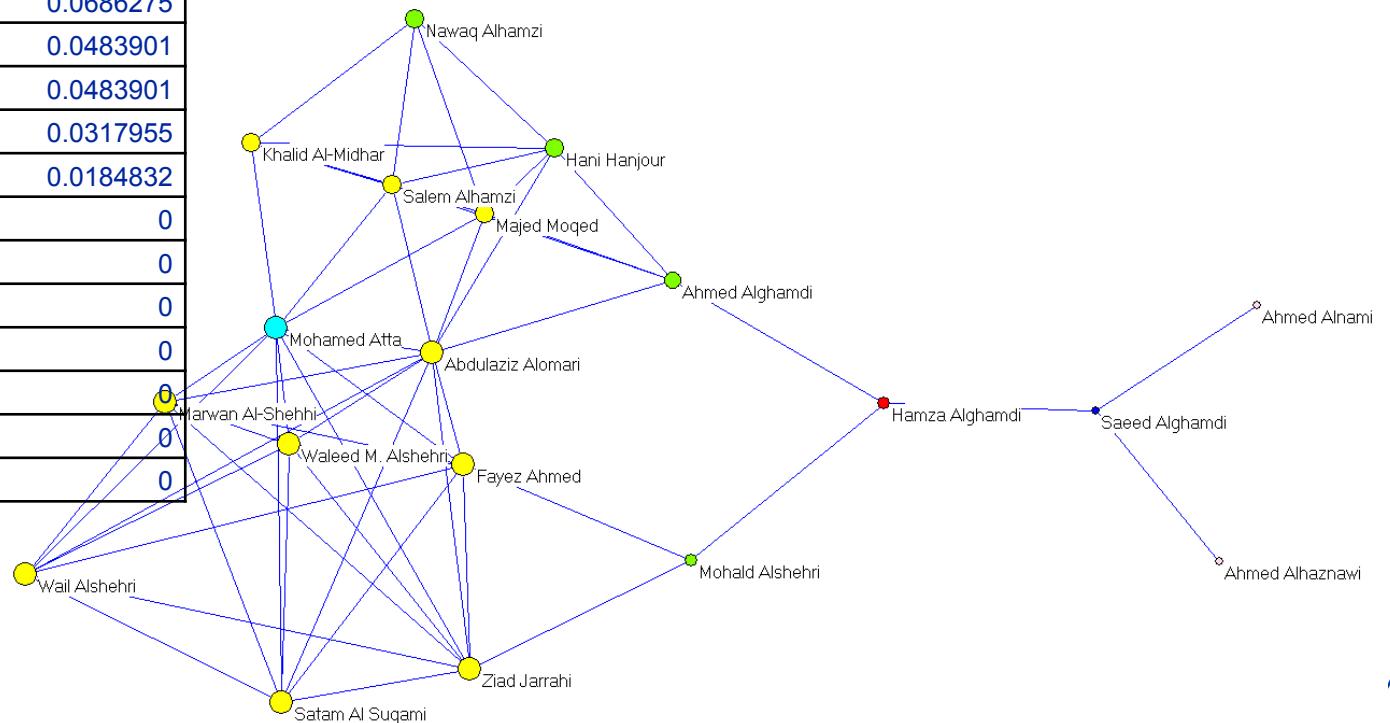
Name	Closeness
Abdulaziz Alomari	0.6
Ahmed Alghamdi	0.5454545
Ziad Jarrahi	0.5294118
Fayez Ahmed	0.5294118
Mohamed Atta	0.5142857
Majed Moqed	0.5142857
Salem Alhamzi	0.5142857
Hani Hanjour	0.5
Marwan Al Shehhi	0.4615385
Satam Al Suqami	0.4615385
Waleed M. Alshehri	0.4615385
Wail Alshehri	0.4615385
Hamza Alghamdi	0.45
Khalid Al Midhar	0.4390244
Mohald Alshehri	0.4390244
Nawaq Alhamzi	0.3673469
Saeed Alghamdi	0.3396226
Ahmed Alnami	0.2571429
Ahmed Alhaznawi	0.2571429



Introduction
 Centrality & prestige
 PageRank algorithm
 HITS algorithm
 Community detection
 ● Example

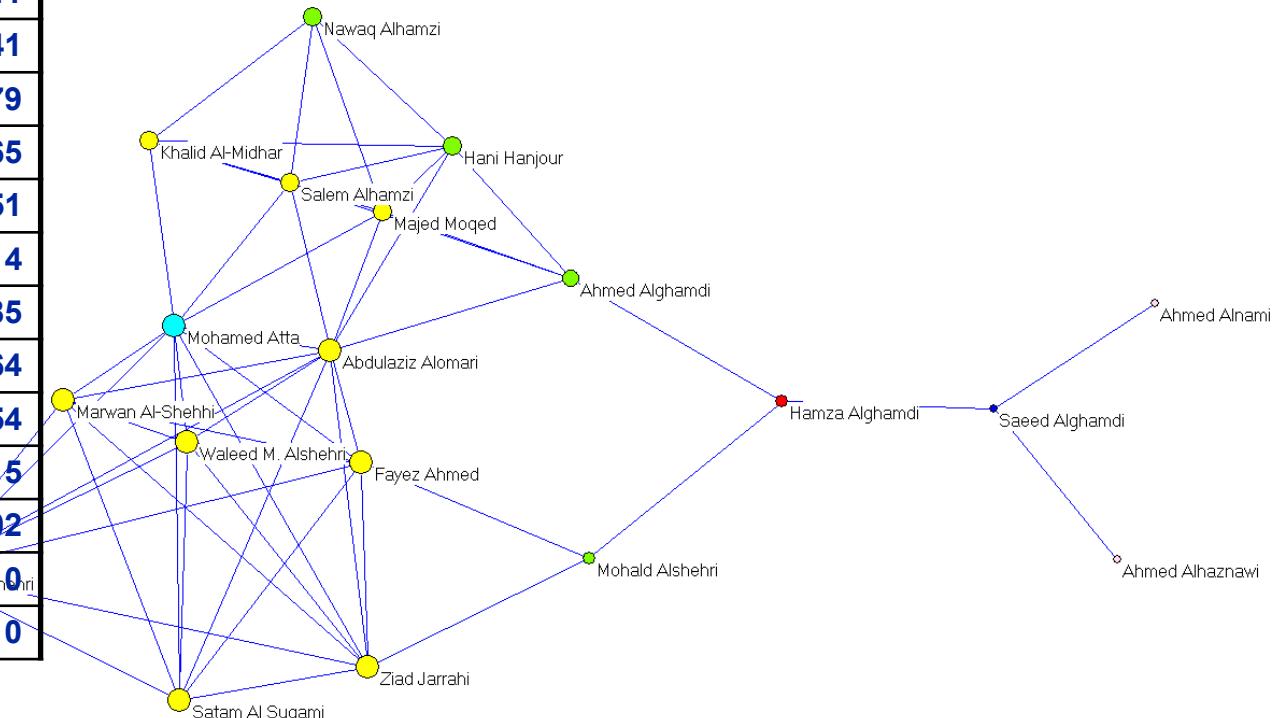
Centrality: Betweenness of the hijackers

Name	Betweenness (B_i)
Hamza Alghamdi	0.3059446
Saeed Alghamdi	0.2156863
Ahmed Alghamdi	0.210084
Abdulaziz Alomari	0.1848669
Mohald Alshehri	0.1350763
Mohamed Atta	0.1224783
Ziad Jarrahi	0.0807656
Fayez Ahmed	0.0686275
Majed Moqed	0.0483901
Salem Alhamzi	0.0483901
Hani Hanjour	0.0317955
Khalid Al-Midhar	0.0184832
Nawaq Alhamzi	0
Marwan Al-Shehhi	0
Satam Al Suqami	0
Waleed M. Alshehri	0
Wail Alshehri	0
Ahmed Alnami	0
Ahmed Alhaznawi	0



Centrality: Eigenvector centralities of the hijackers

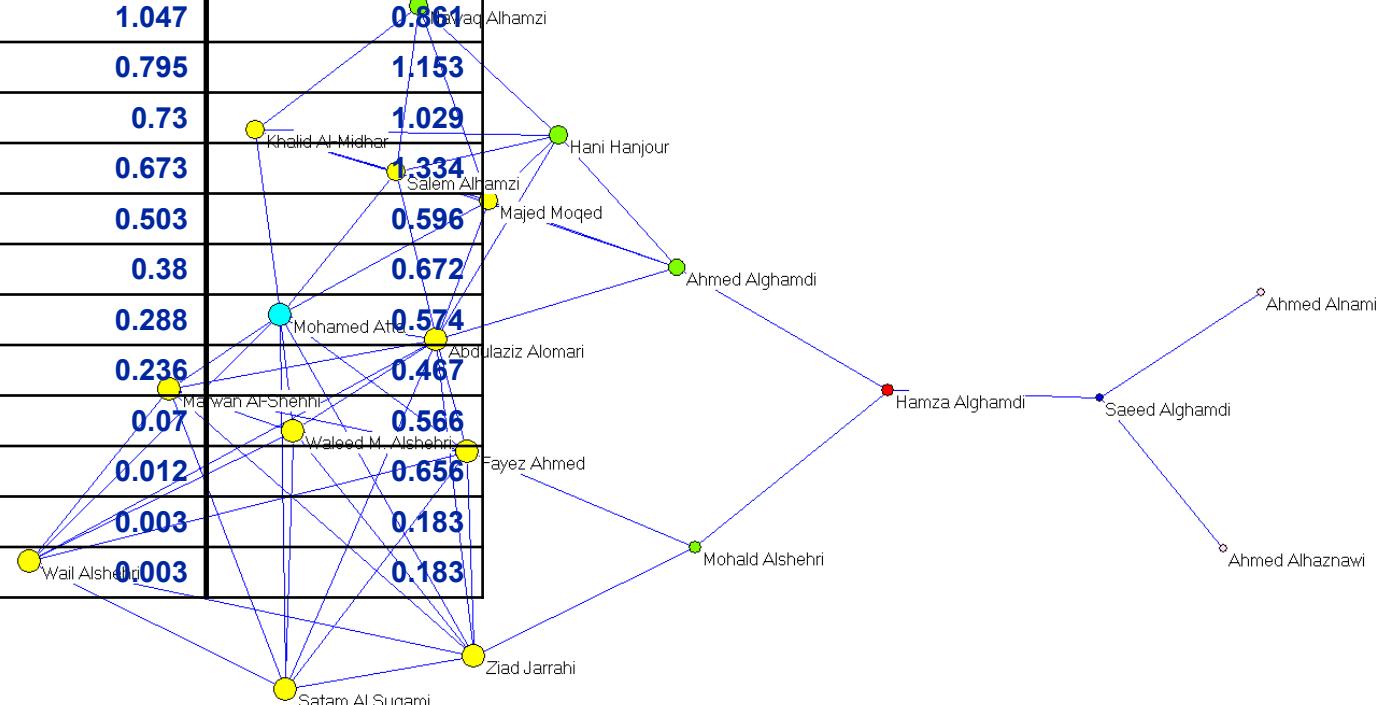
Name	E1
Mohamed Atta	0.518
Marwan Al-Shehhi	0.489
Abdulaziz Alomari	0.296
Ziad Jarrahi	0.246
Fayez Ahmed	0.246
Satam Al Suqami	0.241
Waleed M. Alshehri	0.241
Wail Alshehri	0.241
Salem Alhamzi	0.179
Majed Moqed	0.165
Hani Hanjour	0.151
Khalid Al-Midhar	0.114
Ahmed Alghamdi	0.085
Nawaq Alhamzi	0.064
Mohald Alshehri	0.054
Hamza Alghamdi	0.015
Saeed Alghamdi	0.002
Ahmed Alnami	0
Ahmed Alhaznawi	0



Introduction
 Centrality & prestige
 PageRank algorithm
 HITS algorithm
 Community detection
 ● Example

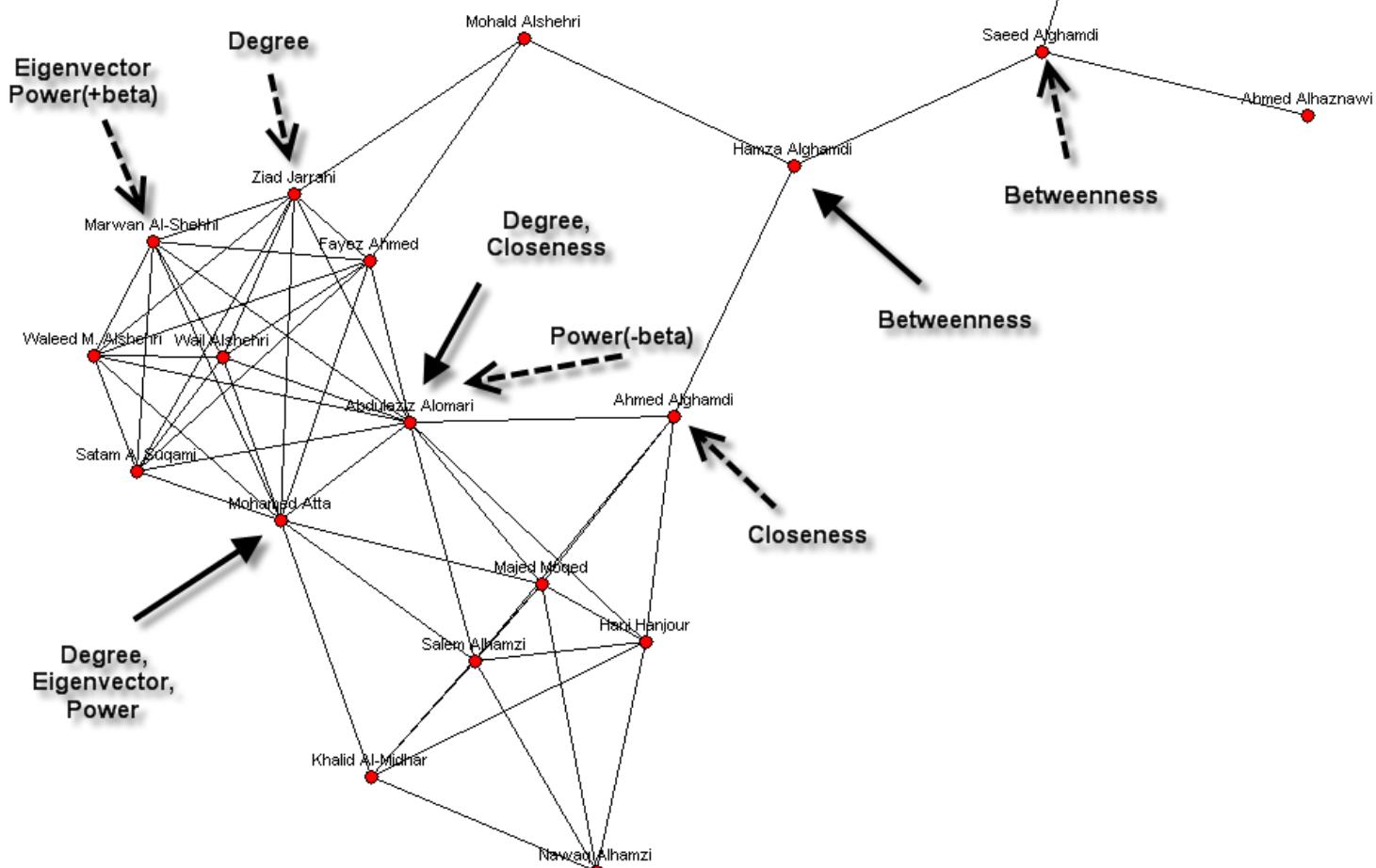
Power of the hijackers

	Power : $\beta = 0.99$	Power : $\beta = -0.99$
Mohamed Atta	2.254	2.214
Marwan Al-Shehhi	2.121	0.969
Abdulaziz Alomari	1.296	1.494
Ziad Jarrahi	1.07	1.087
Fayez Ahmed	1.07	1.087
Satam Al Suqami	1.047	0.861
Waleed M. Alshehri	1.047	0.861
Wail Alshehri	1.047	0.861
Salem Alhamzi	0.795	1.153
Majed Moqed	0.73	1.029
Hani Hanjour	0.673	1.334
Khalid Al-Midhar	0.503	0.596
Ahmed Alghamdi	0.38	0.672
Nawaq Alhamzi	0.288	0.574
Mohald Alshehri	0.236	0.467
Hamza Alghamdi	0.07	0.566
Saeed Alghamdi	0.012	0.656
Ahmed Alnami	0.003	0.183
Ahmed Alhaznawi	0.003	0.183



Introduction
 Centrality & prestige
 PageRank algorithm
 HITS algorithm
 Community detection
 ● Example

Summary diagram



**Introduction
Centrality & prestige
PageRank algorithm
HITS algorithm
Community detection
Example**

Example: Game of Thrones

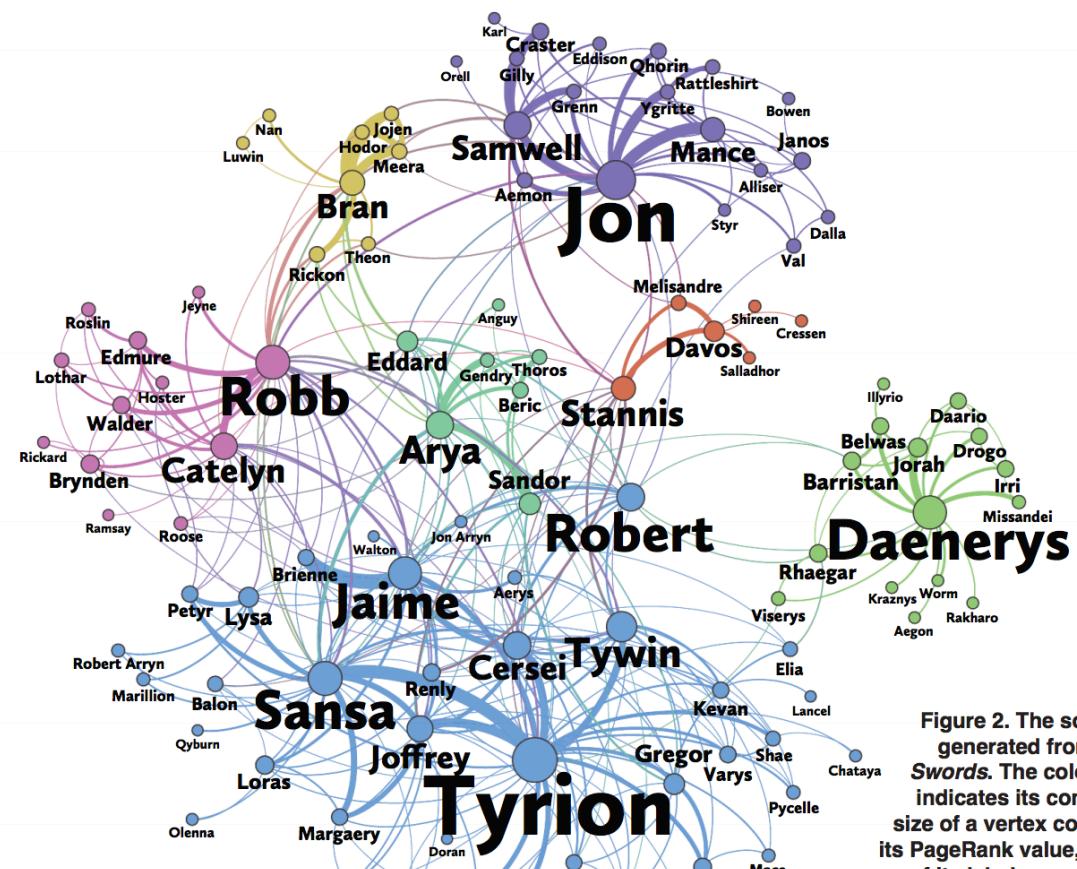
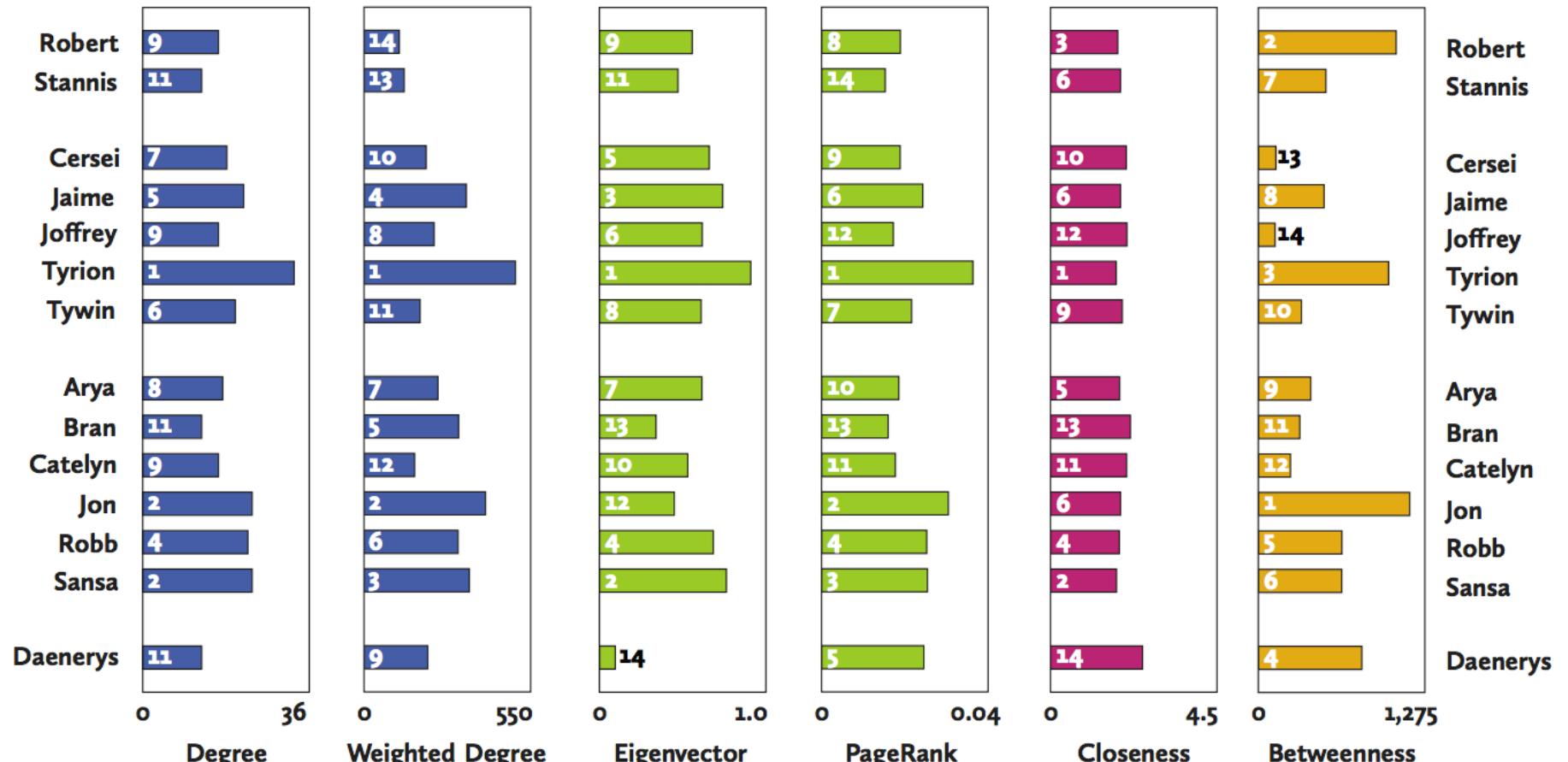


Figure 2. The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

<http://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20%281%29.pdf>

Example: Game of Thrones Cont'd



<http://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20%281%29.pdf>

Guest Lecture



**Nicholas Christakis on the
hidden influence of social
networks**

[http://www.ted.com/talks/
nicholas_christakis_the_hidden_influence_of_social_networks](http://www.ted.com/talks/nicholas_christakis_the_hidden_influence_of_social_networks)

Concluding remarks

- Social network analysis is one of the most important directions in “Big Data” analytics
- The field is very active and developing
- A number of measures and algorithms have been proposed but many more may still come
- Further Readings
 - <http://www.cs.cornell.edu/home/kleinber/networks-book/>
 - <http://arxiv.org/pdf/0906.0612.pdf>
 - <http://link.springer.com/article/10.1140/epjb%2Fe2004-00124-y>

