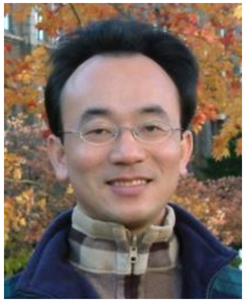
Social Media Marketing Analytics
社群網路行銷分析

Tamkang University

社會網路分析
(Social Network Analysis)

1032SMMA08
TLMXJ1A (MIS EMBA)
Fri 12,13,14 (19:20-22:10) D326



Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系
http://mail. tku.edu.tw/myday/

2015-05-22

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2015/02/27	和平紀念日補假(放假一天)
2	2015/03/06	社群網路行銷分析課程介紹 (Course Orientation for Social Media Marketing Analytics)
3	2015/03/13	社群網路行銷分析 (Social Media Marketing Analytics)
4	2015/03/20	社群網路行銷研究 (Social Media Marketing Research)
5	2015/03/27	測量構念 (Measuring the Construct)
6	2015/04/03	兒童節補假(放假一天)
7	2015/04/10	社群網路行銷個案分析 I (Case Study on Social Media Marketing I)
8	2015/04/17	測量與量表 (Measurement and Scaling)
9	2015/04/24	探索性因素分析 (Exploratory Factor Analysis)

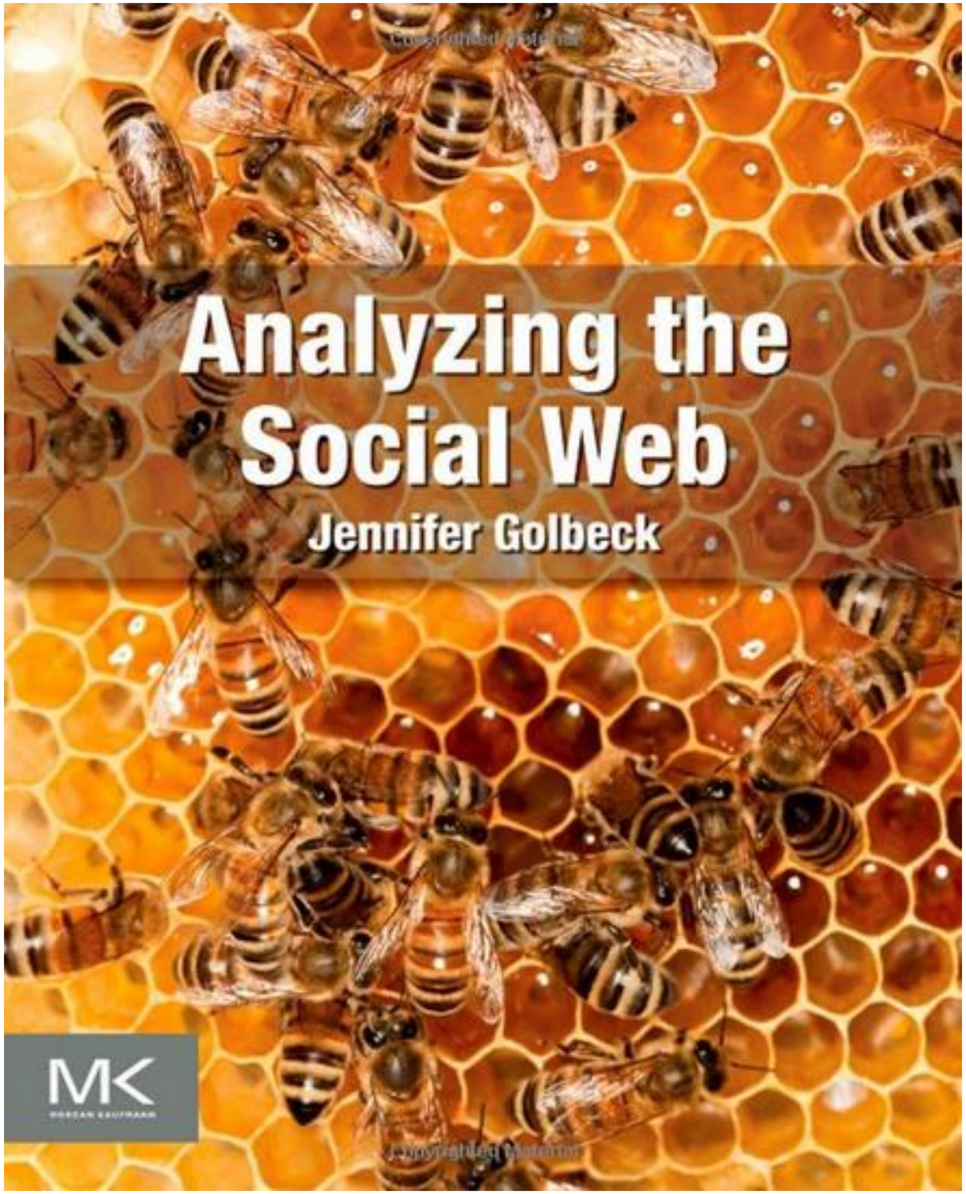
課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
10	2015/05/01	社群運算與大數據分析 (Social Computing and Big Data Analytics) [Invited Speaker: Irene Chen, Consultant, Teradata]
11	2015/05/08	期中報告 (Midterm Presentation)
12	2015/05/15	確認性因素分析 (Confirmatory Factor Analysis)
13	2015/05/22	社會網路分析 (Social Network Analysis)
14	2015/05/29	社群網路行銷個案分析 II (Case Study on Social Media Marketing II)
15	2015/06/05	社群網路情感分析 (Sentiment Analysis on Social Media)
16	2015/06/12	期末報告 I (Term Project Presentation I)
17	2015/06/19	端午節補假 (放假一天)
18	2015/06/26	期末報告 II (Term Project Presentation II)

Outline

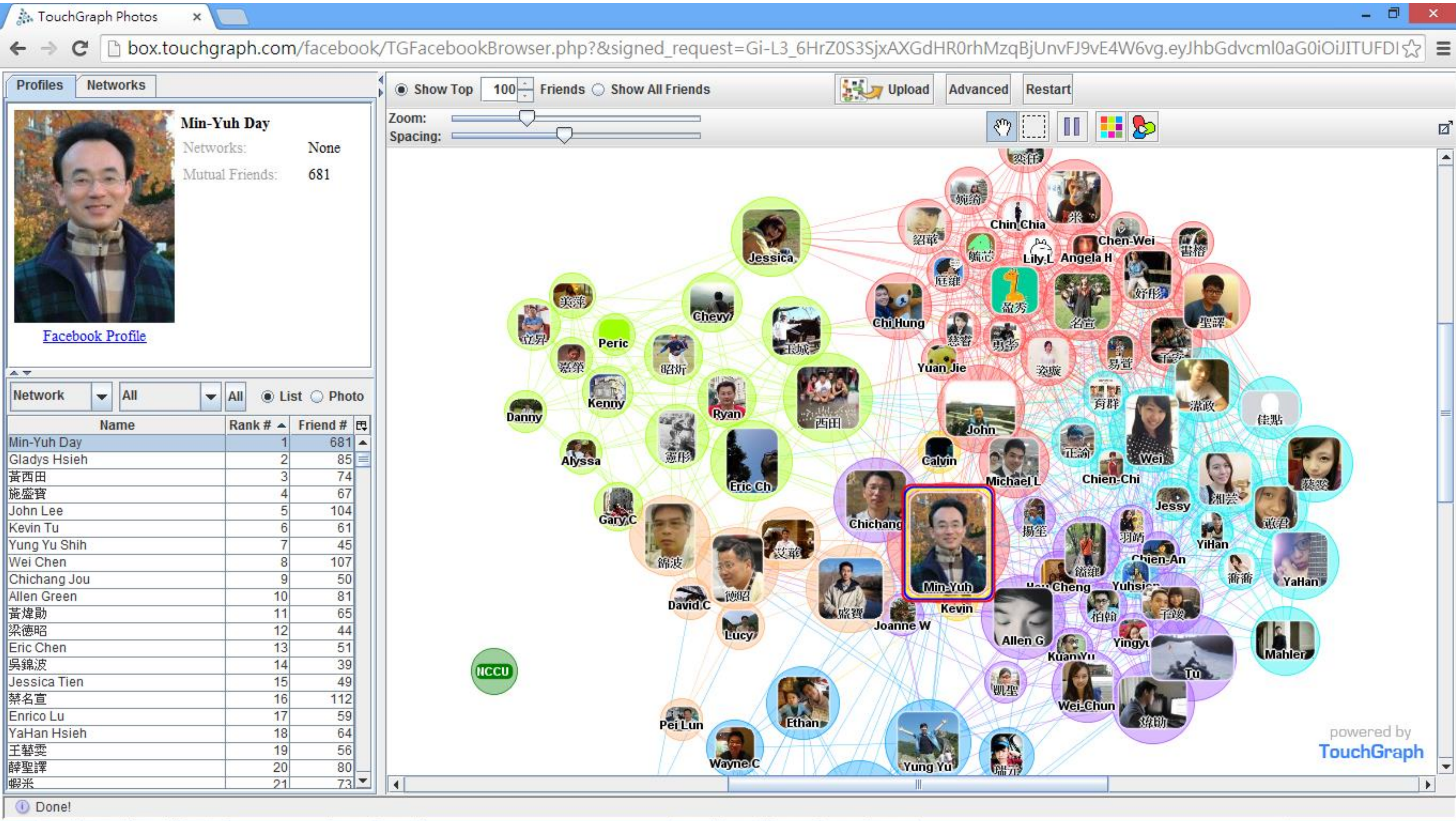
- Social Network Analysis (SNA)
  – Degree Centrality
  – Betweenness Centrality
  – Closeness Centrality
- Link Mining
- SNA Tools
  – UCINet
  – Pajek
- Applications of SNA

Jennifer Golbeck (2013), **Analyzing the Social Web**, Morgan Kaufmann

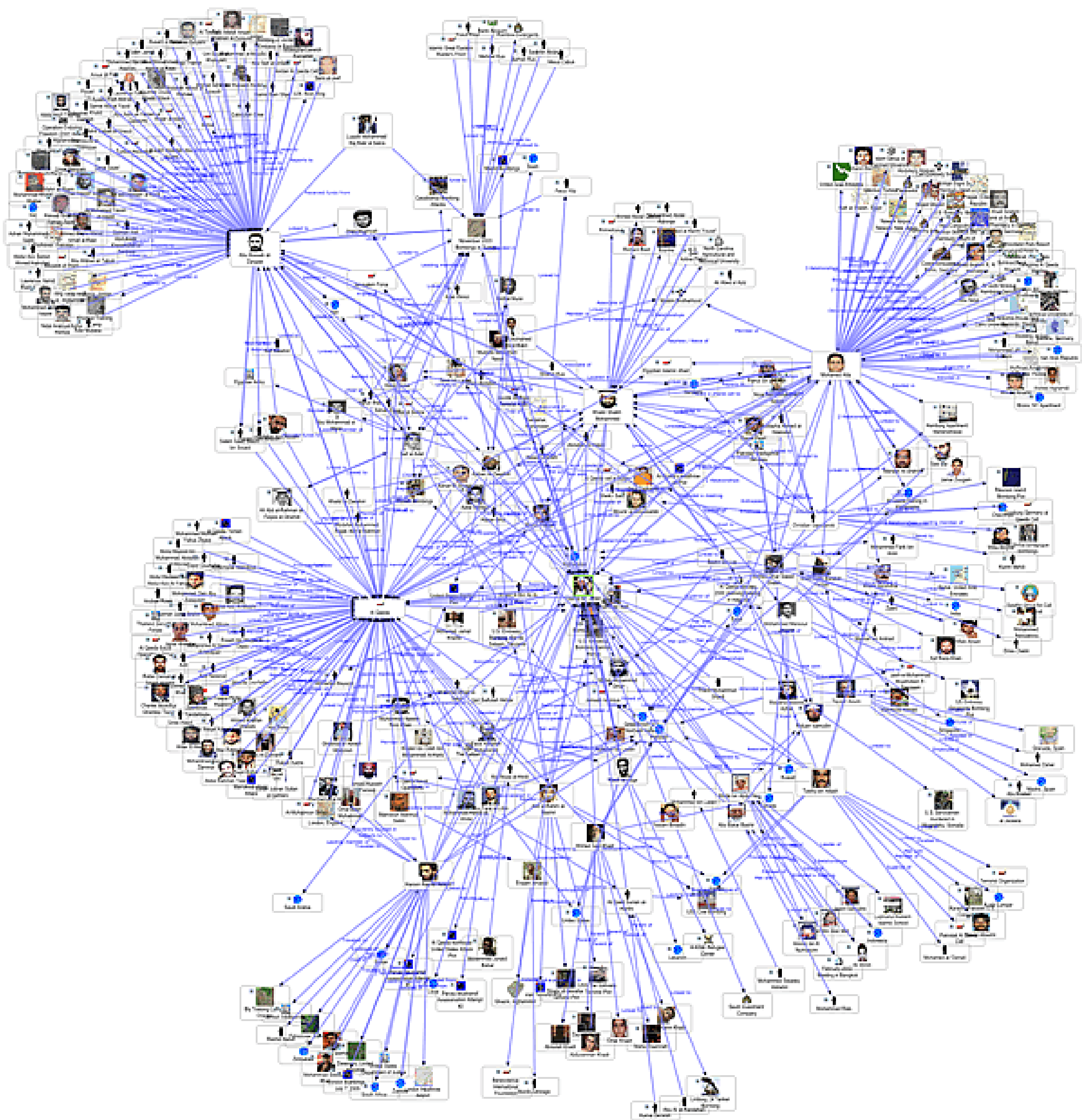


Source: http://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311

Social Network Analysis (SNA)
Facebook TouchGraph



Social Network Analysis



Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis

- A **social network** is a social structure of people, related (directly or indirectly) to each other through a common relation or interest
- **Social network analysis (SNA)** is the study of social networks to understand their structure and behavior

Source: (c) Jaideep Srivastava, srivasta@cs.umn.edu, Data Mining for Social Network Analysis

Social Network Analysis

- Using Social Network Analysis, you can get answers to questions like:
  - How highly connected is an entity within a network?
  - What is an entity's overall importance in a network?
  - How central is an entity within a network?
  - How does information flow within a network?

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis

- Social network is the study of social entities (people in an organization, called **actors**), and their interactions and relationships.
- The interactions and relationships can be represented with a network or graph,
  - each vertex (or node) represents an actor and
  - each link represents a relationship.
- From the network, we can study the properties of its structure, and the role, position and prestige of each social actor.
- We can also find various kinds of sub-graphs, e.g., **communities** formed by groups of actors.

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data"

Social Network and the Web

- Social network analysis is useful for the Web because the Web is essentially a virtual society, and thus a virtual social network,
  - Each page: a social actor and
  - each hyperlink: a relationship.
- Many results from social network can be adapted and extended for use in the Web context.
- Two types of social network analysis,
  - **Centrality**
  - **Prestige**

  closely related to hyperlink analysis and search on the Web

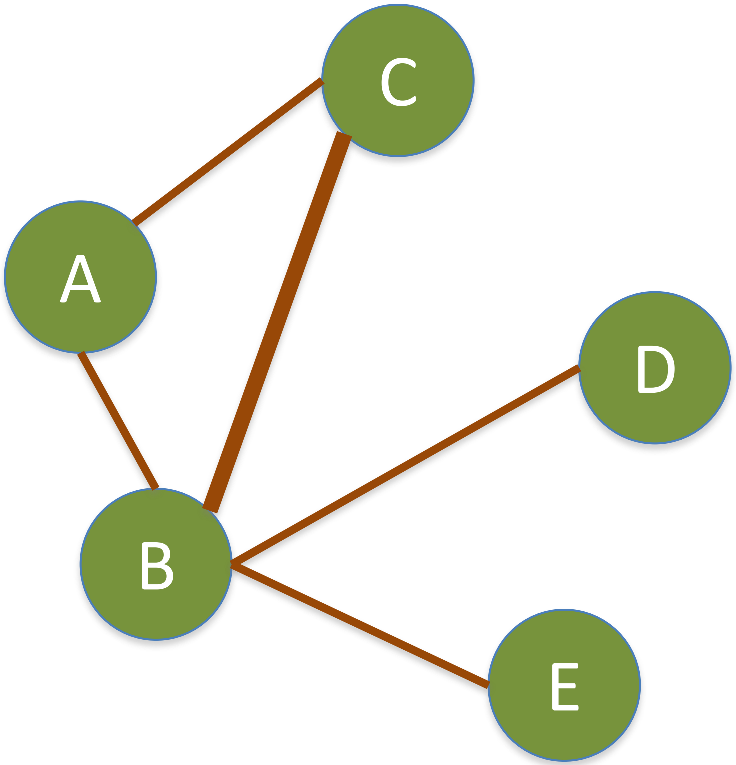
Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data”

Social Network Analysis (SNA)

Centrality

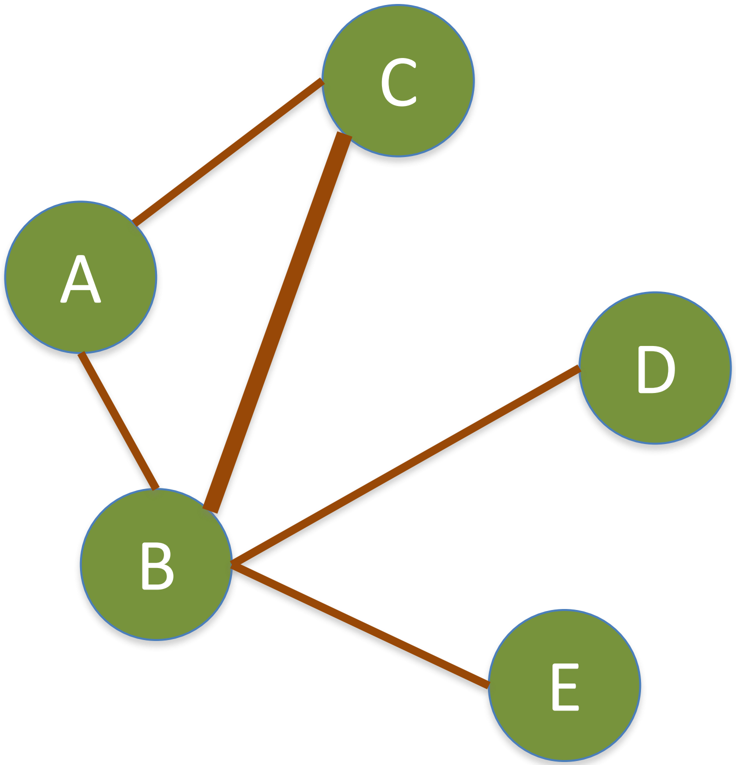
Prestige

Degree



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

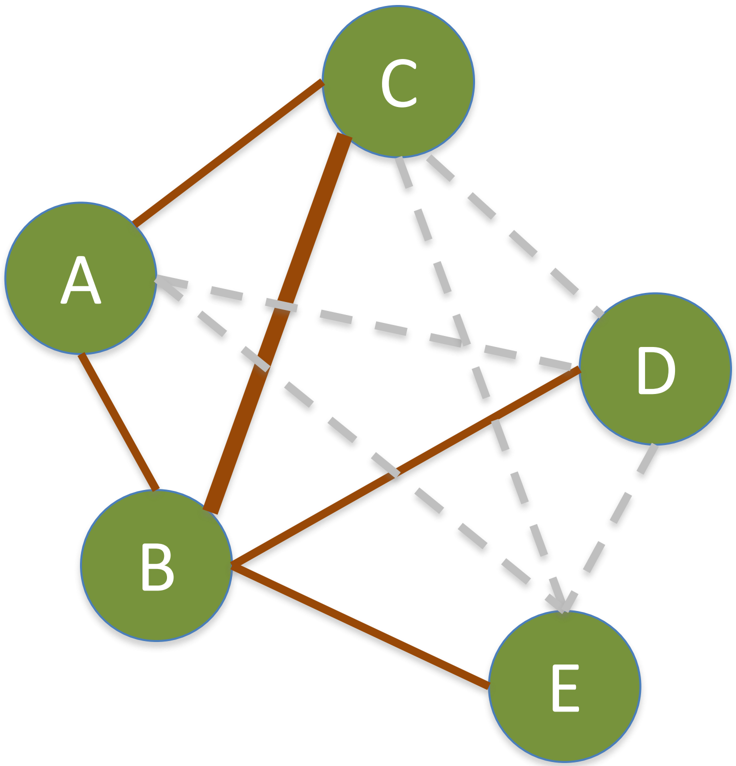
Degree



A: 2
<u>B: 4</u>
C: 2
D:1
E: 1

Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Density



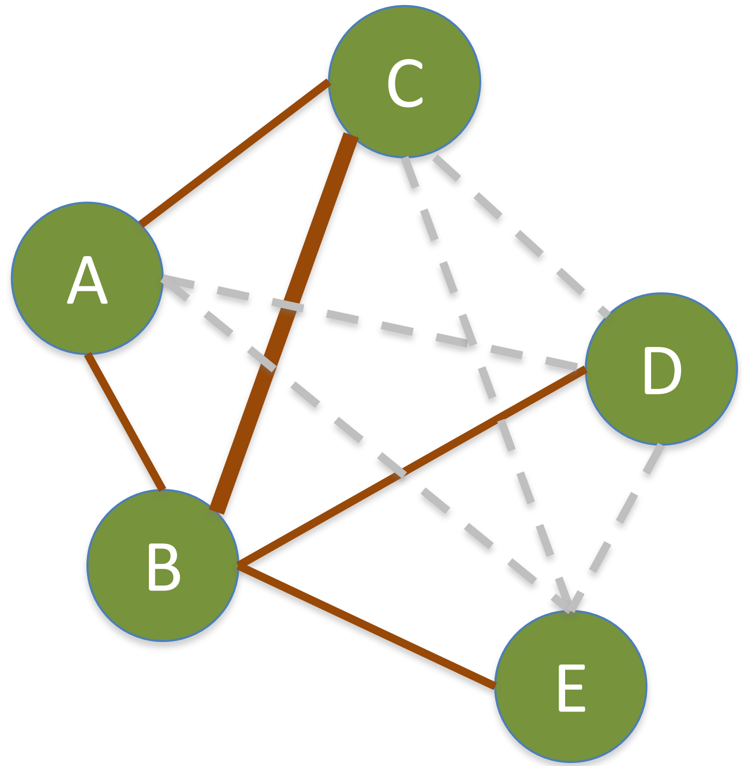
Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Density

Edges (Links): 5

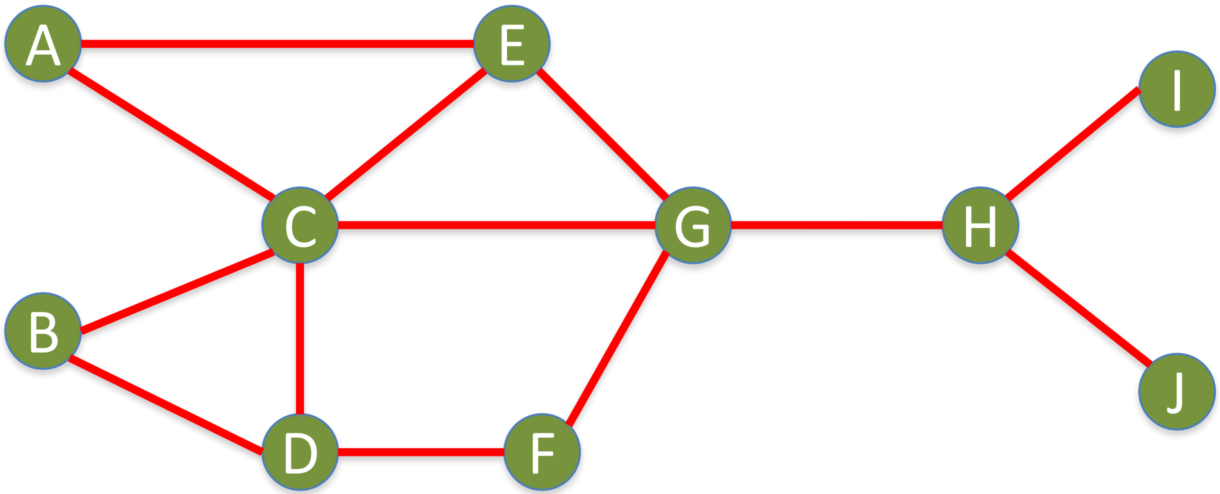
Total Possible Edges: 10

Density: 5/10 = 0.5



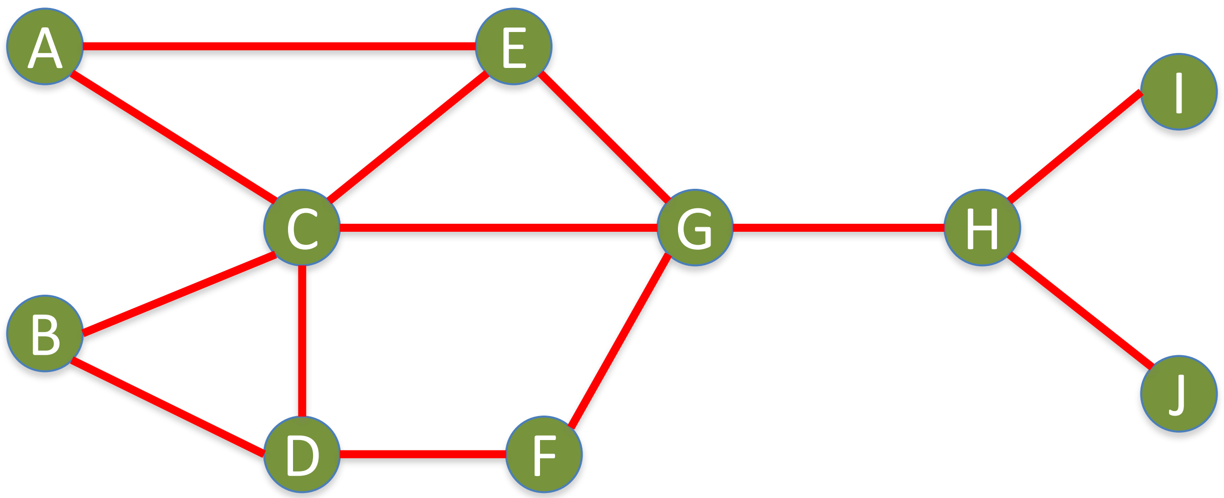
Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Density



Nodes (n): 10
Edges (Links): 13
Total Possible Edges: (n * (n-1)) / 2 = (10 * 9) / 2 = 45
Density: 13/45 = 0.29

Which Node is Most Important?



Centrality

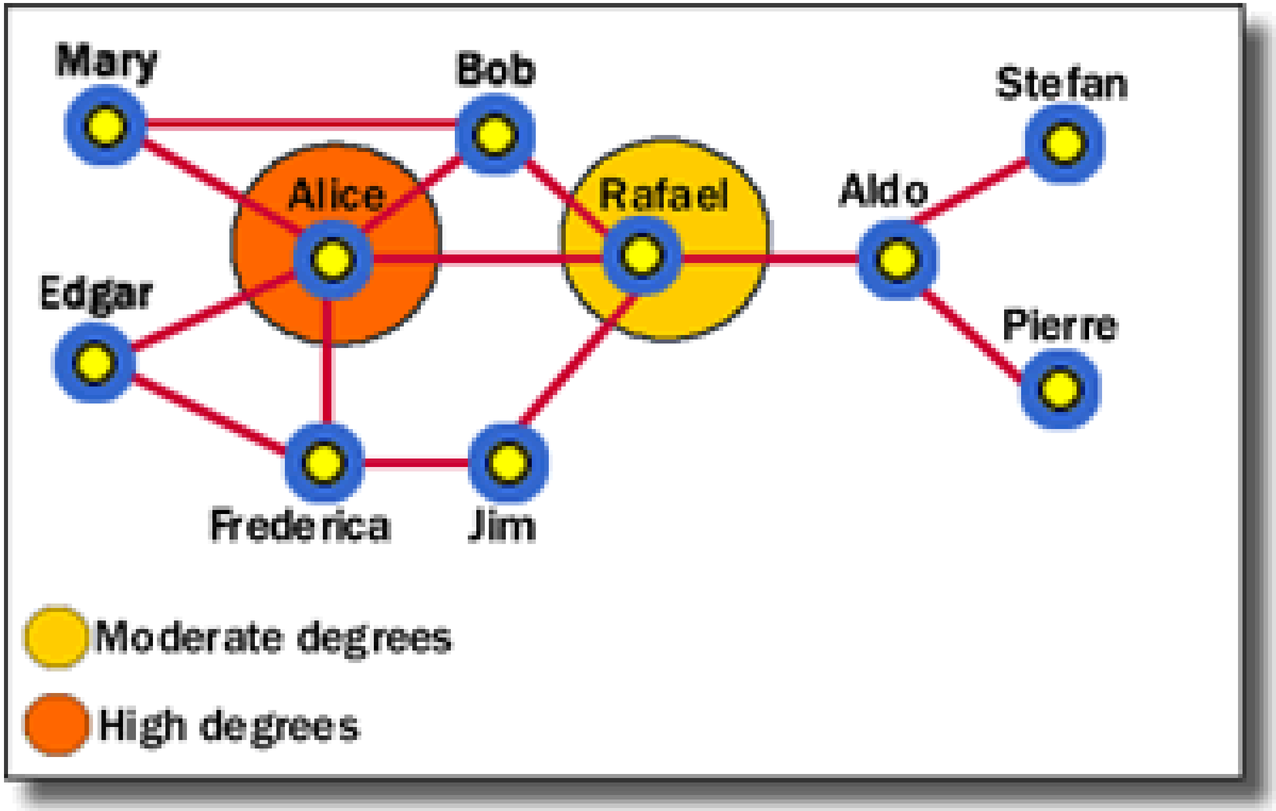
- Important or prominent actors are those that are linked or involved with other actors extensively.
- A person with extensive contacts (links) or communications with many other people in the organization is considered more important than a person with relatively fewer contacts.
- The links can also be called **ties**.
  A central actor is one involved in many ties.

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data"

Social Network Analysis (SNA)

- Degree Centrality
- Betweenness Centrality
- Closeness Centrality

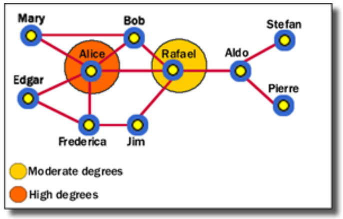
Social Network Analysis: Degree Centrality



Alice has the highest degree centrality, which means that she is quite active in the network. However, she is not necessarily the most powerful person because she is only directly connected within one degree to people in her clique—she has to go through Rafael to get to other cliques.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

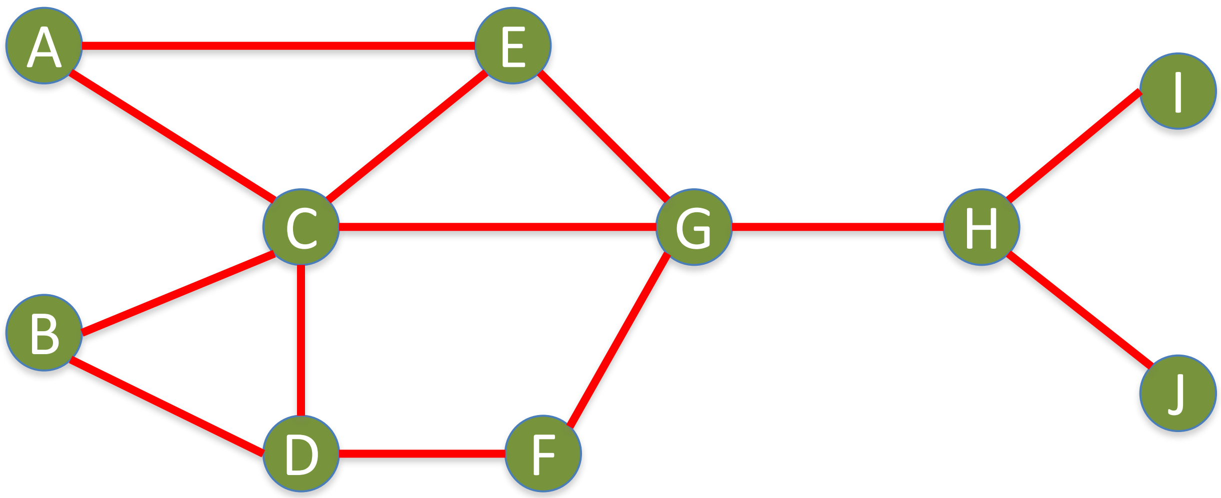
Social Network Analysis: Degree Centrality



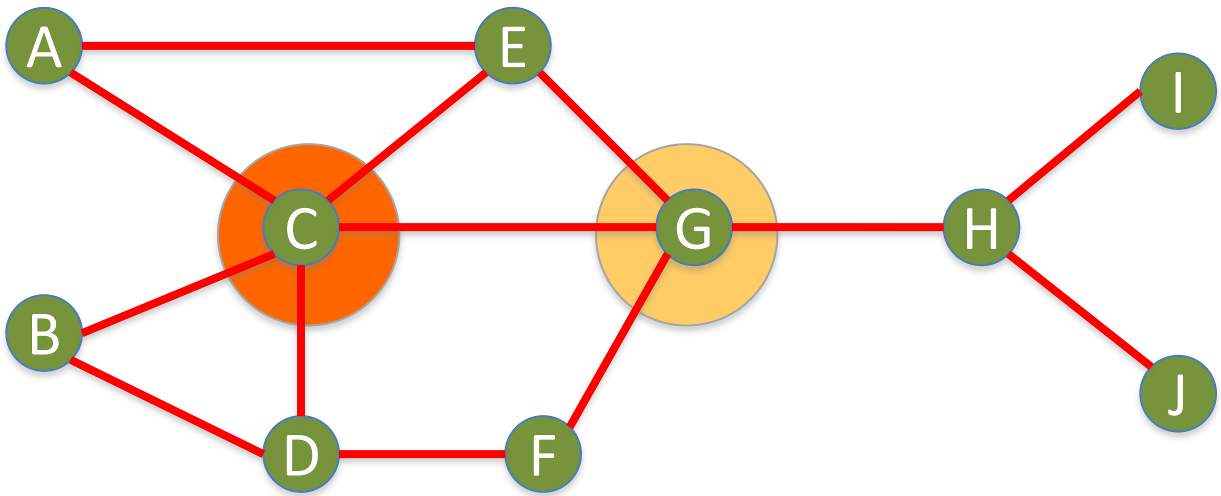
- Degree centrality is simply the number of direct relationships that an entity has.
- An entity with high degree centrality:
  - Is generally an active player in the network.
  - Is often a connector or hub in the network.
  - s not necessarily the most connected entity in the network (an entity may have a large number of relationships, the majority of which point to low-level entities).
  - May be in an advantaged position in the network.
  - May have alternative avenues to satisfy organizational needs, and consequently may be less dependent on other individuals.
  - Can often be identified as third parties or deal makers.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis: Degree Centrality

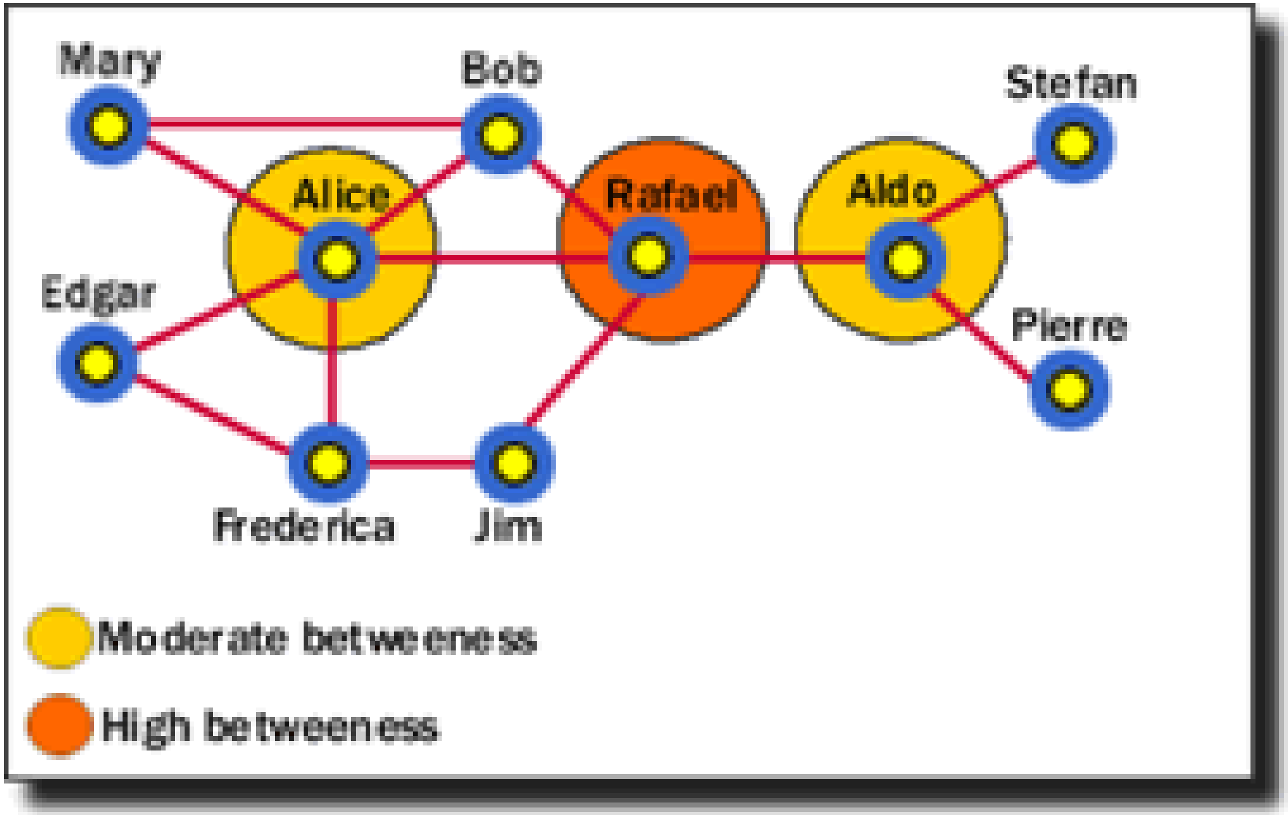


Social Network Analysis: Degree Centrality



Node	Score	Standardized Score
A	2	2/10 = 0.2
B	2	2/10 = 0.2
C	**5**	**5/10 = 0.5**
D	3	3/10 = 0.3
E	3	3/10 = 0.3
F	2	2/10 = 0.2
G	**4**	**4/10 = 0.4**
H	3	3/10 = 0.3
I	1	1/10 = 0.1
J	1	1/10 = 0.1

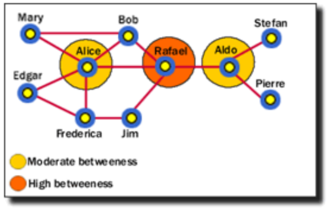
Social Network Analysis: Betweenness Centrality



Rafael has the highest betweenness because he is between Alice and Aldo, who are between other entities. Alice and Aldo have a slightly lower betweenness because they are essentially only between their own cliques. Therefore, although Alice has a higher degree centrality, Rafael has more importance in the network in certain respects.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis: Betweenness Centrality



- Betweenness centrality identifies an entity's position within a network in terms of its ability to make connections to other pairs or groups in a network.
- An entity with a high betweenness centrality generally:
  - Holds a favored or powerful position in the network.
  - Represents a single point of failure—take the single betweenness spanner out of a network and you sever ties between cliques.
  - Has a greater amount of influence over what happens in a network.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Betweenness centrality:

Connectivity

Number of shortest paths going through the actor

Betweenness Centrality

$$C_B(i) = \sum_{j<k} g_{ik}(i) / g_{jk}$$

Where g_{jk} = the number of shortest paths connecting jk

$g_{jk}(i)$ = the number that actor i is on.

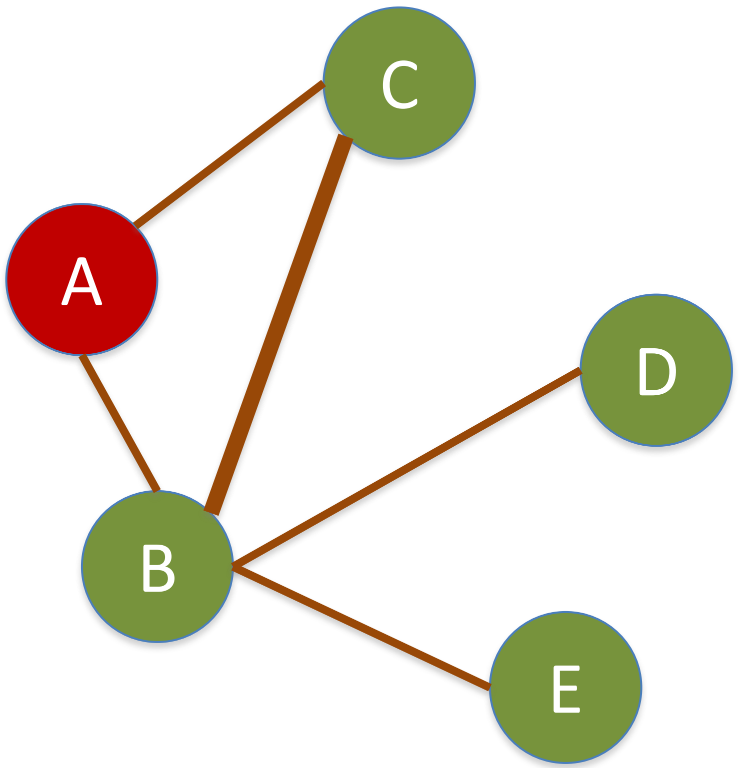
Normalized by

$$C'_B(i) = C_B(i) / \boxed{[(n-1)(n-2)/2]}$$

Number of pairs of vertices excluding the vertex itself

Source: https://www.youtube.com/watch?v=RXohUeNCJiU

Betweenness Centrality

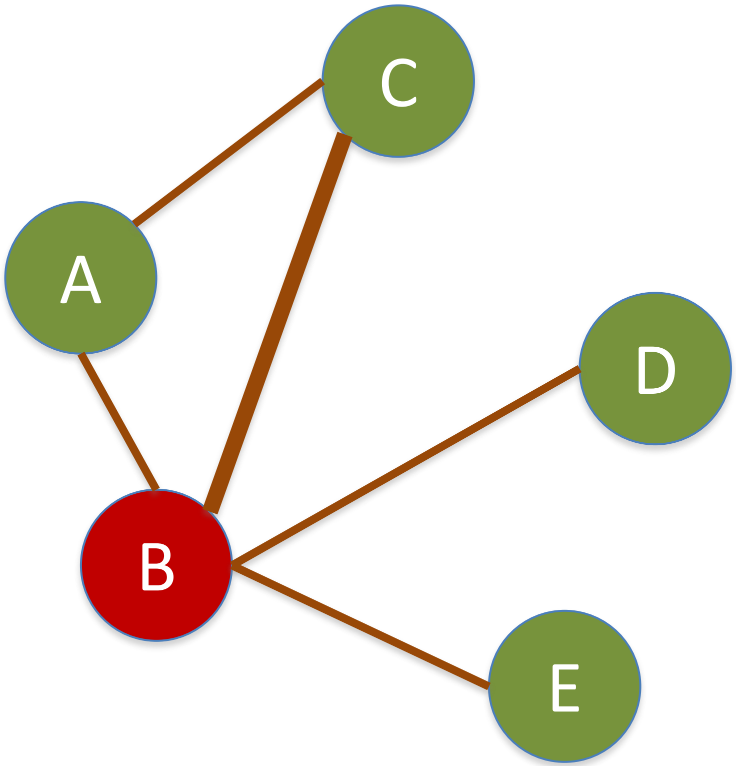


A:
B→C: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
C→D: 0/1 = 0
C→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

A: Betweenness Centrality = 0

Betweenness Centrality



B:

A→C: 0/1 = 0

A→D: 1/1 = 1

A→E: 1/1 = 1

C→D: 1/1 = 1

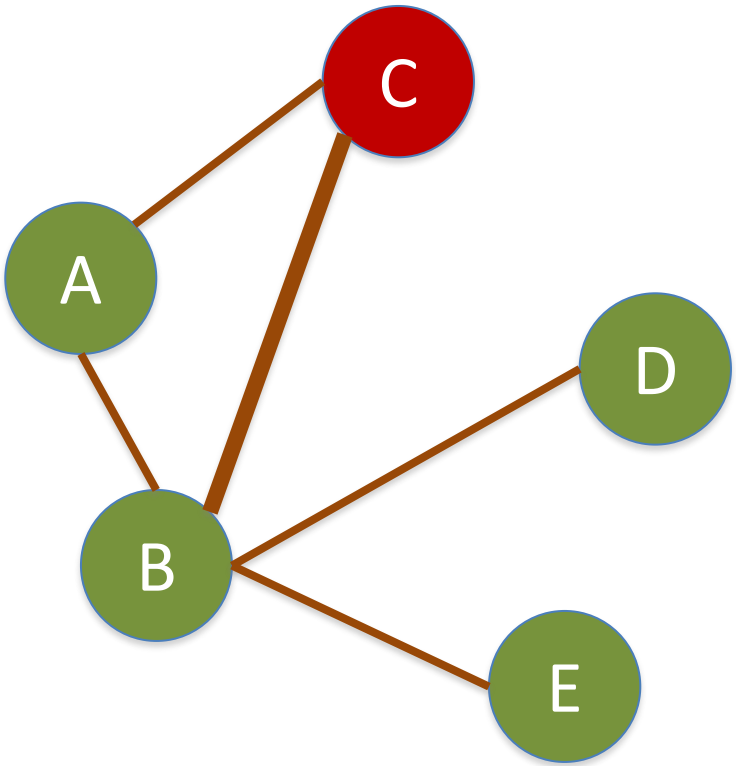
C→E: 1/1 = 1

D→E: 1/1 = 1

Total: 5

A: Betweenness Centrality = 5

Betweenness Centrality

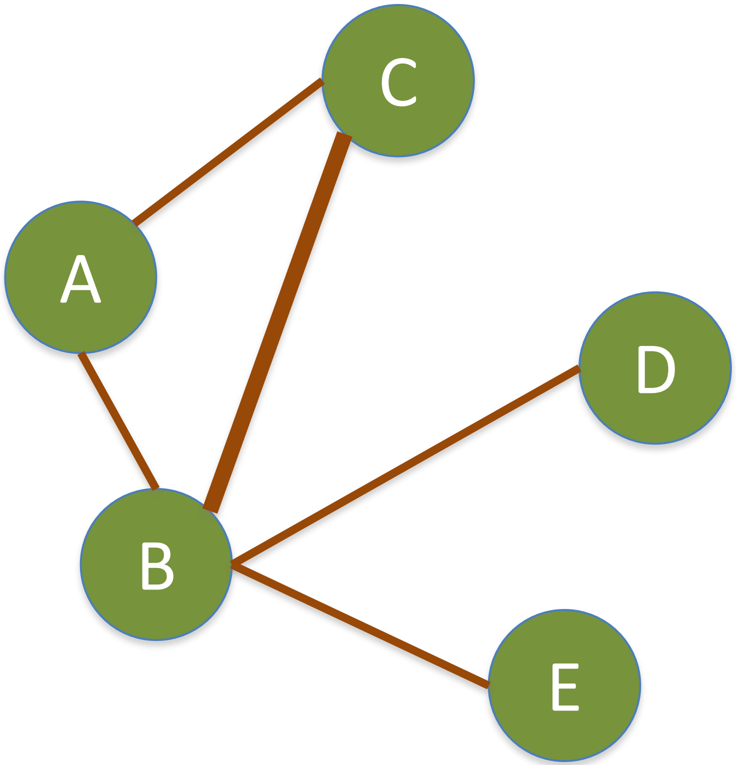


C:
A→B: 0/1 = 0
A→D: 0/1 = 0
A→E: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

C: Betweenness Centrality = 0

Betweenness Centrality



A: 0

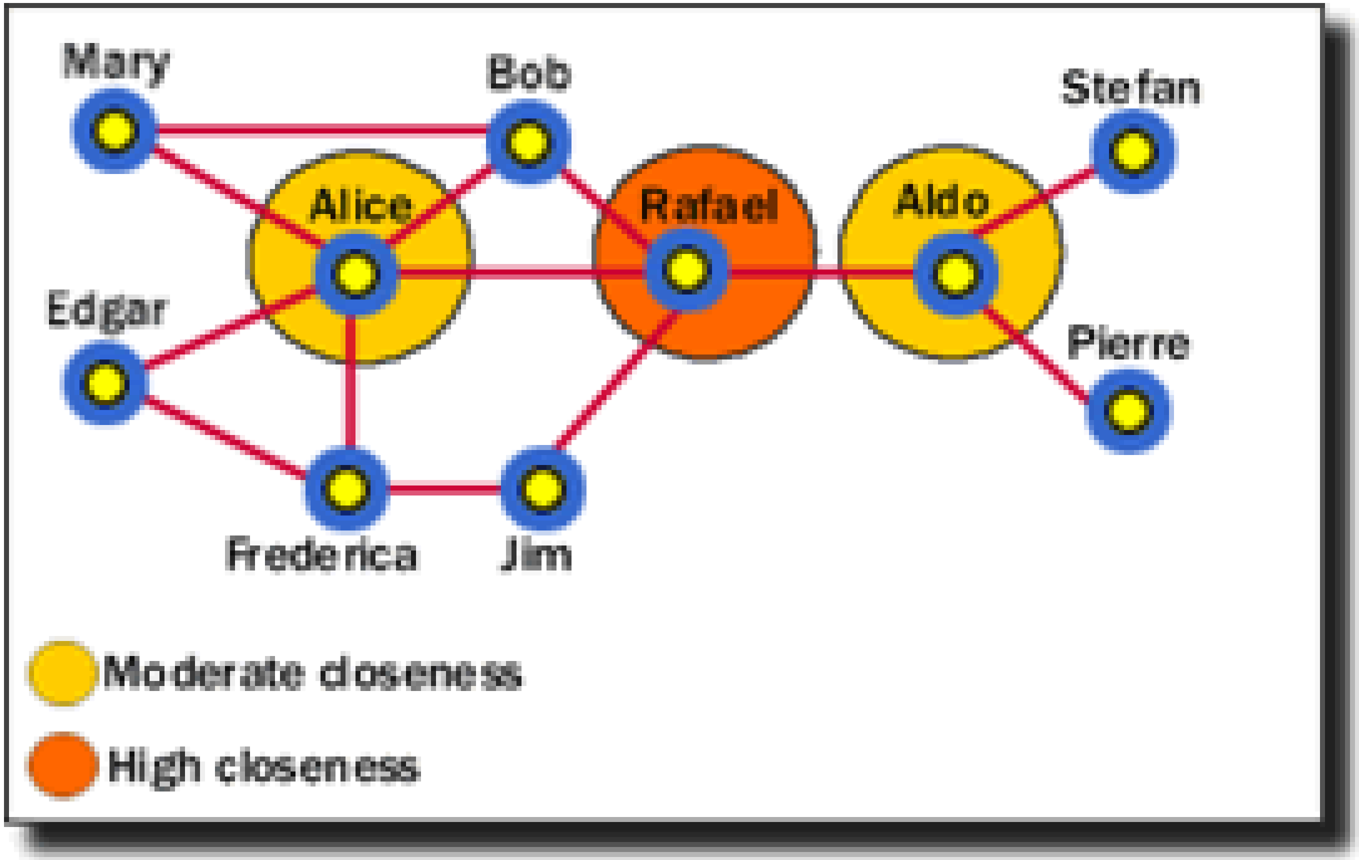
B: 5

C: 0

D: 0

E: 0

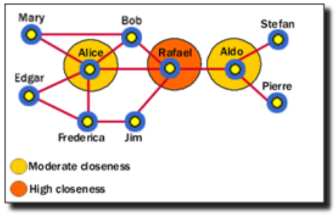
Social Network Analysis: Closeness Centrality



Rafael has the highest closeness centrality because he can reach more entities through shorter paths. As such, Rafael's placement allows him to connect to entities in his own clique, and to entities that span cliques.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

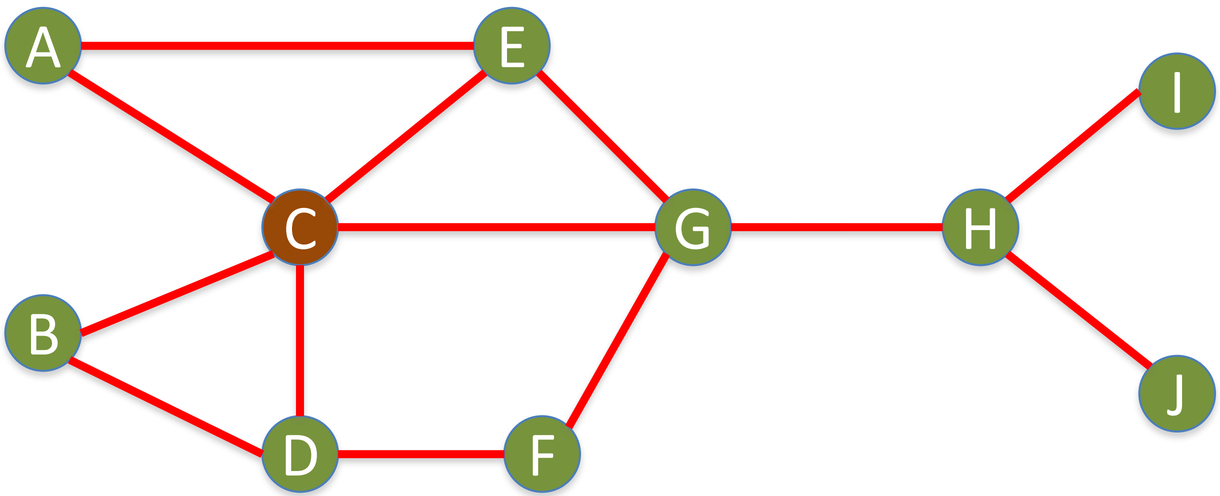
Social Network Analysis: Closeness Centrality



- Closeness centrality measures how quickly an entity can access more entities in a network.
- An entity with a high closeness centrality generally:
  - Has quick access to other entities in a network.
  - Has a short path to other entities.
  - Is close to other entities.
  - Has high visibility as to what is happening in the network.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis: Closeness Centrality

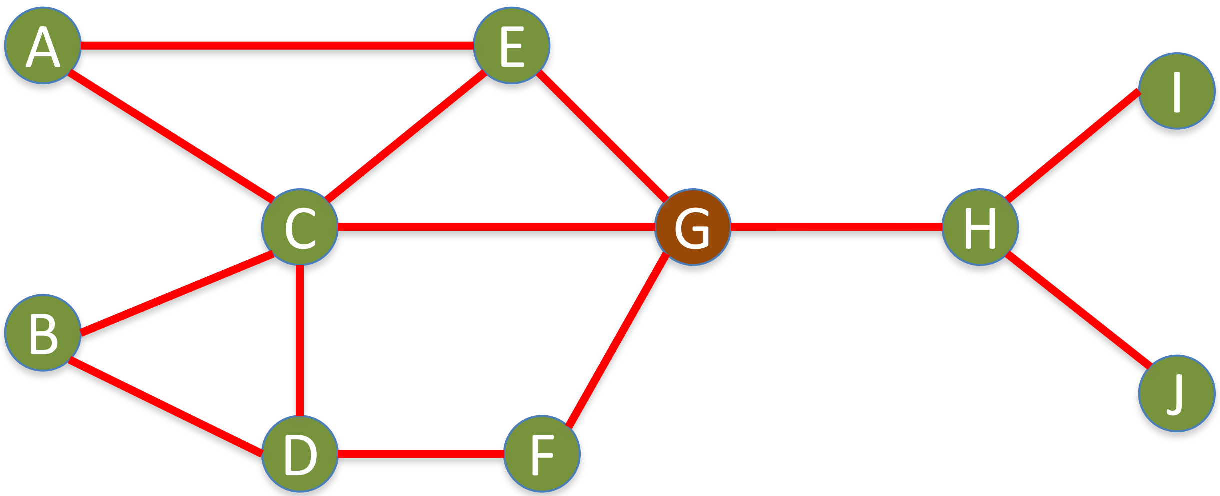


C→A: 1
C→B: 1
C→D: 1
C→E: 1
C→F: 2
C→G: 1
C→H: 2
C→I: 3
C→J: 3

Total=15

C: Closeness Centrality = 15/9 = 1.67

Social Network Analysis: Closeness Centrality

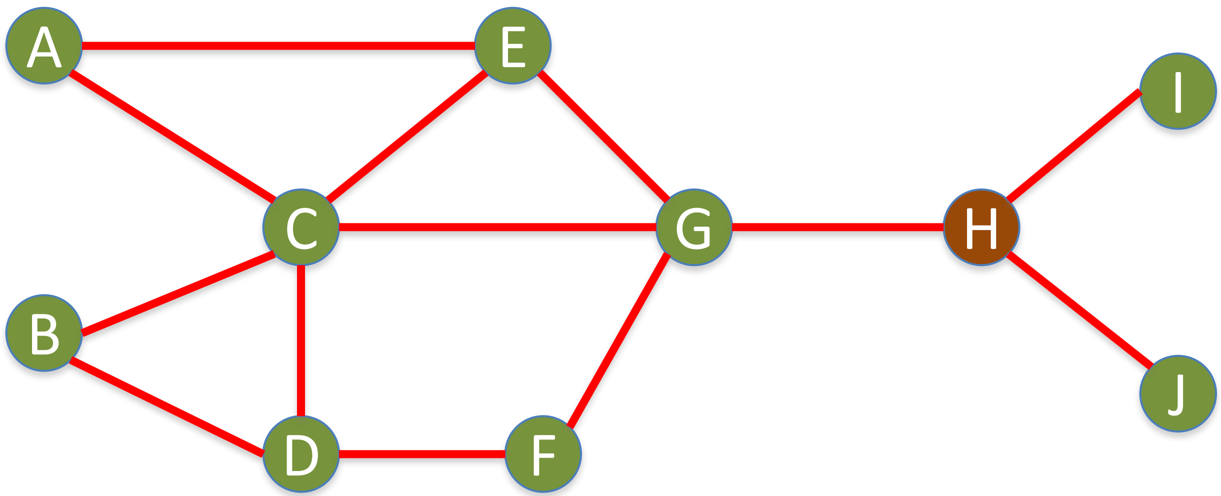


G→A: 2
G→B: 2
G→C: 1
G→D: 2
G→E: 1
G→F: 1
G→H: 1
G→I: 2
G→J: 2

Total=14

G: Closeness Centrality = 14/9 = 1.56

Social Network Analysis: Closeness Centrality

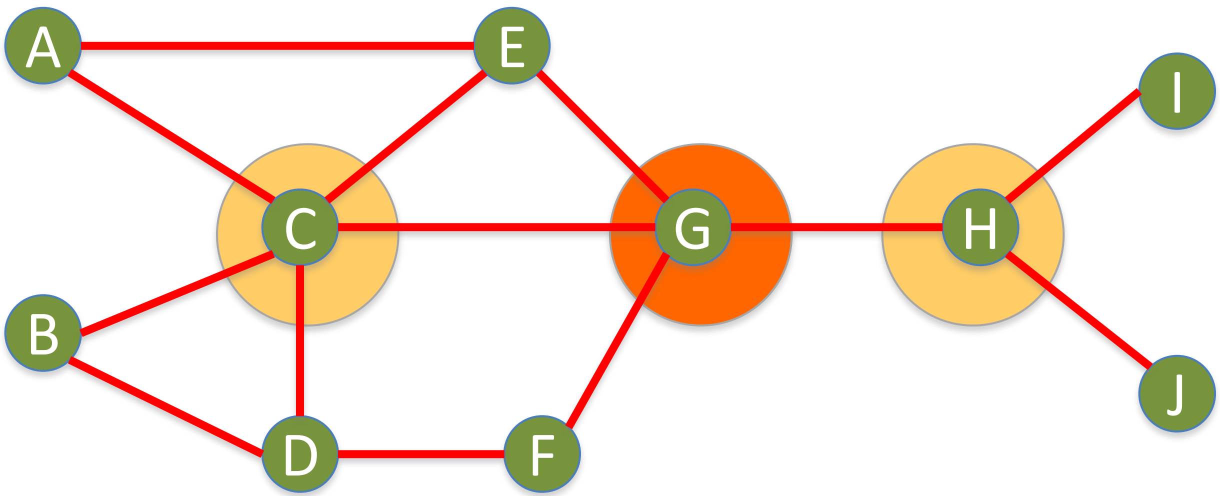


H→A: 3
H→B: 3
H→C: 2
H→D: 2
H→E: 2
H→F: 2
H→G: 1
H→I: 1
H→J: 1

Total=17

H: Closeness Centrality = 17/9 = 1.89

Social Network Analysis: Closeness Centrality

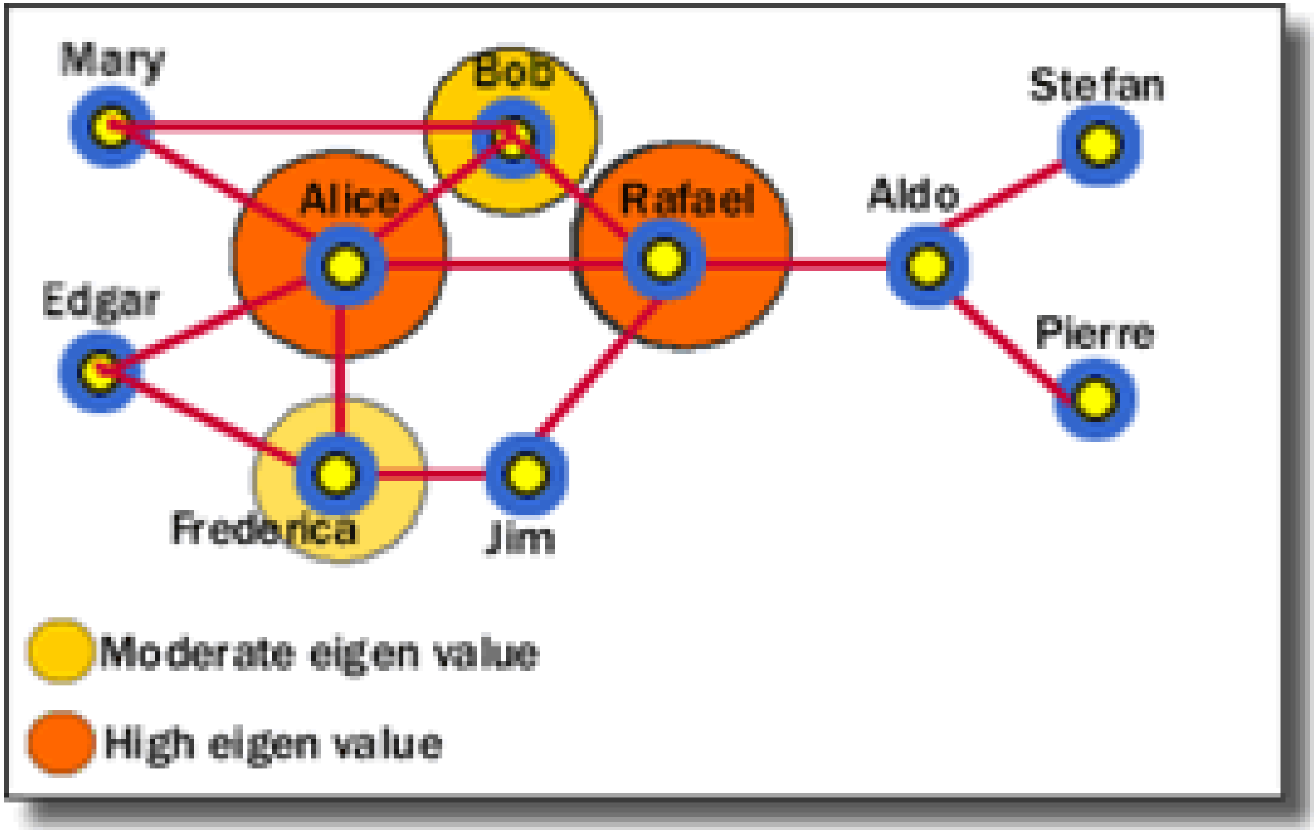


G: Closeness Centrality = 14/9 = 1.56 1

C: Closeness Centrality = 15/9 = 1.67 2

H: Closeness Centrality = 17/9 = 1.89 3

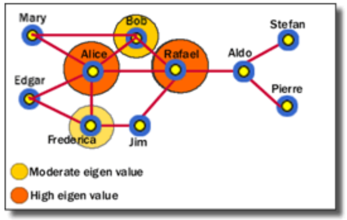
Social Network Analysis: Eigenvalue



Alice and Rafael are closer to other highly close entities in the network. Bob and Frederica are also highly close, but to a lesser value.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis: Eigenvalue



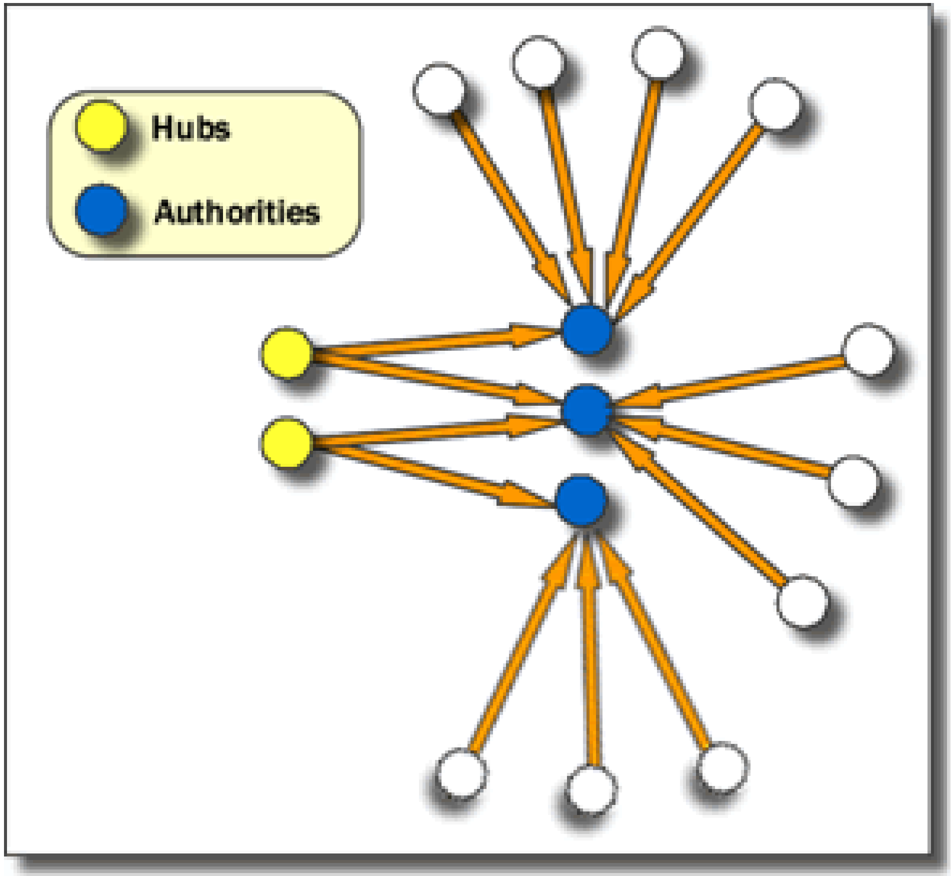
- Eigenvalue measures how close an entity is to other highly close entities within a network. In other words, Eigenvalue identifies the most central entities in terms of the global or overall makeup of the network.
- A high Eigenvalue generally:
  – Indicates an actor that is more central to the main pattern of distances among all entities.
  – Is a reasonable measure of one aspect of centrality in terms of positional advantage.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Eigenvector centrality:

Importance of a node
depends on
the importance of its neighbors

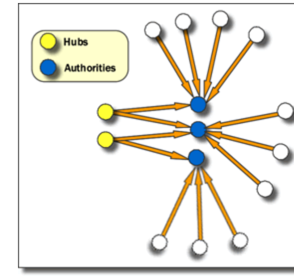
Social Network Analysis: Hub and Authority



Hubs are entities that point to a relatively large number of authorities. They are essentially the mutually reinforcing analogues to authorities. Authorities point to high hubs. Hubs point to high authorities. You cannot have one without the other.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis: Hub and Authority



- Entities that many other entities point to are called Authorities. In Sentinel Visualizer, relationships are directional—they point from one entity to another.
- If an entity has a high number of relationships pointing to it, it has a high authority value, and generally:
  - Is a knowledge or organizational authority within a domain.
  - Acts as definitive source of information.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Social Network Analysis

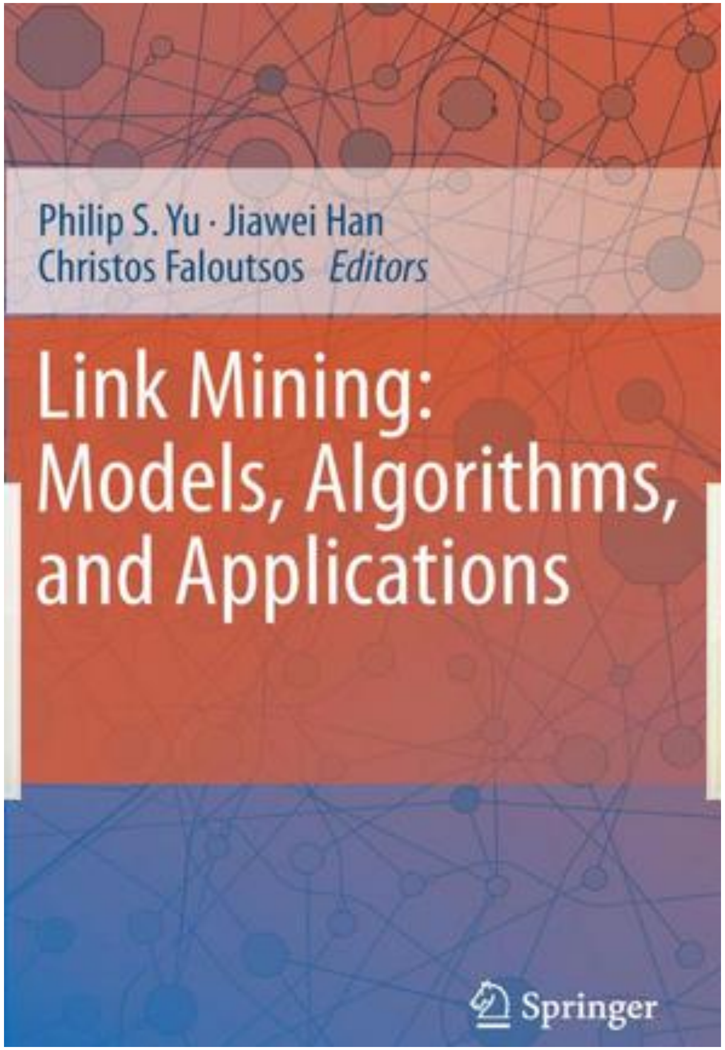
Network Metrics

Calculate ○ CardView ◉ TableView ☐ Group area Expand groups Collapse groups

Name	Type	Degree	Betweenness	Closeness	Eigenvalue	Hub	Authority
Osama bin Laden	Person	44	0.920492092358...	1	0.0271	0	0.011
Abdallah Al-Halabi	Person	2	0	0.654867256637...	0.0001	0	0
Abu Mussab al-Zarqawi	Person	84	0.934887847326...	0.869451697127...	0.7028	0.6572	0.1076
Al Qaeda	Terrorist Organiz...	85	1	0.962427745664...	0.0416	0.3941	0.0166
Ayman Al-Zawahiri	Person	14	0.045794908783...	0.716129032258...	0	0	0.0173
Enaam Arnaout	Person	4	0.031189325814...	0.656804733727...	0.0001	0	0
Imad Eddin Barakat Yarbas	Person	11	0.065049589038...	0.704016913319...	0.0015	0	0.0025
Khalid Shaikh Mohammed	Person	32	0.339916464724...	0.866059817945...	0.002	0	0.1528
Mohamed Atta	Person	61	0.666268740074...	0.820197044334...	0.0015	0	0.6816

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Link Mining



http://www.amazon.com/Link-Mining-Models-Algorithms-Applications/dp/1441965149

Link Mining
(Getoor & Diehl, 2005)

- Link Mining
  - Data Mining techniques that take into account the links between objects and entities while building predictive or descriptive models.
- Link based object ranking, Group Detection, Entity Resolution, Link Prediction
- Application:
  - Hyperlink Mining
  - Relational Learning
  - Inductive Logic Programming
  - Graph Mining

Characteristics of Collaboration Networks

(Newman, 2001; 2003; 3004)

- Degree distribution follows a power-law
- Average separation decreases in time.
- Clustering coefficient decays with time
- Relative size of the largest cluster increases
- Average degree increases
- Node selection is governed by preferential attachment

Social Network Techniques

- Social network extraction/construction
- Link prediction
- Approximating large social networks
- Identifying prominent/trusted/expert actors in social networks
- Search in social networks
- Discovering communities in social network
- Knowledge discovery from social network

Social Network Extraction

- Mining a social network from data sources
- Three sources of social network (Hope et al., 2006)
  - Content available on web pages
    - E.g., user homepages, message threads
  - User interaction logs
    - E.g., email and messenger chat logs
  - Social interaction information provided by users
    - E.g., social network service websites (Facebook)

Social Network Extraction

- IR based extraction from web documents
  - Construct an "actor-by-term" matrix
  - The terms associated with an actor come from web pages/documents created by or associated with that actor
  - IR techniques (TF-IDF, LSI, cosine matching, intuitive heuristic measures) are used to quantify similarity between two actors' term vectors
  - The similarity scores are the edge label in the network
    - Thresholds on the similarity measure can be used in order to work with binary or categorical edge labels
    - Include edges between an actor and its k-nearest neighbors
- Co-occurrence based extraction from web documents

Link Prediction

- Link Prediction using supervised learning (Hasan et al., 2006)
  - Citation Network (BIOBASE, DBLP)
  - Use machine learning algorithms to predict future co-authorship
    - Decision three, k-NN, multilayer perceptron, SVM, RBF network
  - Identify a group of features that are most helpful in prediction
  - Best Predictor Features
    - Keywork Match count, Sum of neighbors, Sum of Papers, Shortest distance

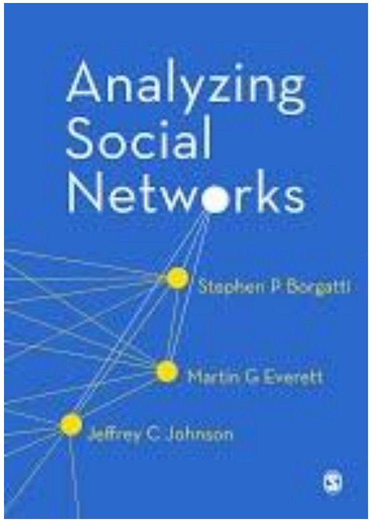
Identifying Prominent Actors in a Social Network

- Compute scores/ranking over the set (or a subset) of actors in the social network which indicate degree of importance / expertise / influence
  - E.g., Pagerank, HITS, centrality measures
- Various algorithms from the link analysis domain
  - PageRank and its many variants
  - HITS algorithm for determining authoritative sources
- Centrality measures exist in the social science domain for measuring importance of actors in a social network

Identifying Prominent Actors in a Social Network

- Brandes, 2011
- Prominence→ high betweenness value
- Betweenness centrality requires computation of number of shortest paths passing through each node
- Compute shortest paths between all pairs of vertices

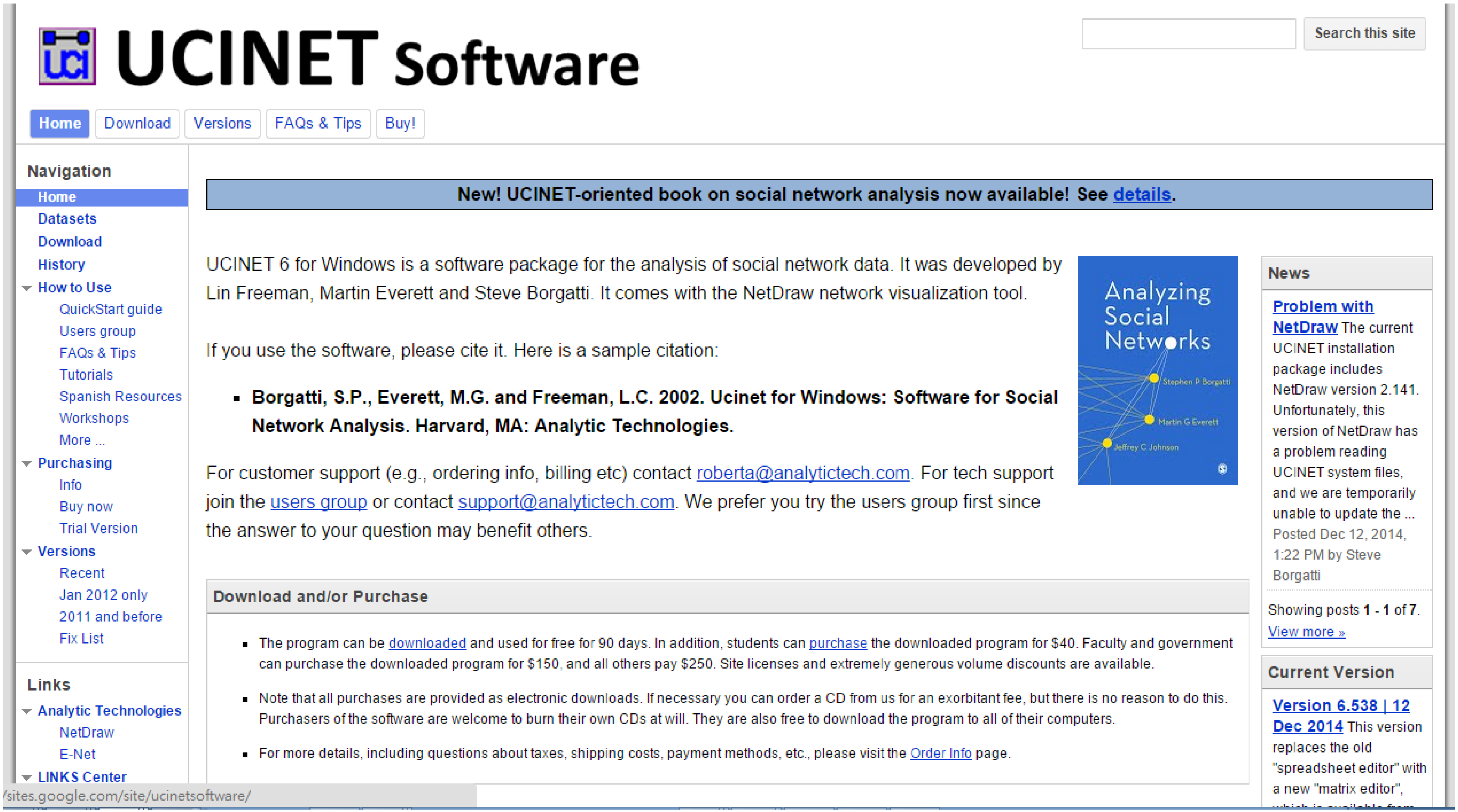
Social Network Analysis (SNA) Tools



- **UCINet**
- **Pajek**

SNA Tool: UCINet

https://sites.google.com/site/ucinetsoftware/home



UCINET Software

Search this site

Home | Download | Versions | FAQs & Tips | Buy!

Navigation

- Home
- Datasets
- Download
- History
- How to Use
  - QuickStart guide
  - Users group
  - FAQs & Tips
  - Tutorials
  - Spanish Resources
  - Workshops
  - More ...
- Purchasing
  - Info
  - Buy now
  - Trial Version
- Versions
  - Recent
  - Jan 2012 only
  - 2011 and before
  - Fix List

Links

- Analytic Technologies
  - NetDraw
  - E-Net
- LINKS Center

New! UCINET-oriented book on social network analysis now available! See details.

UCINET 6 for Windows is a software package for the analysis of social network data. It was developed by Lin Freeman, Martin Everett and Steve Borgatti. It comes with the NetDraw network visualization tool.

If you use the software, please cite it. Here is a sample citation:

- **Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.**

For customer support (e.g., ordering info, billing etc) contact roberta@analytictech.com. For tech support join the users group or contact support@analytictech.com. We prefer you try the users group first since the answer to your question may benefit others.

Analyzing Social Networks
Stephen P Borgatti
Martin G Everett
Jeffrey C Johnson

Download and/or Purchase

- The program can be downloaded and used for free for 90 days. In addition, students can purchase the downloaded program for $40. Faculty and government can purchase the downloaded program for $150, and all others pay $250. Site licenses and extremely generous volume discounts are available.
- Note that all purchases are provided as electronic downloads. If necessary you can order a CD from us for an exorbitant fee, but there is no reason to do this. Purchasers of the software are welcome to burn their own CDs at will. They are also free to download the program to all of their computers.
- For more details, including questions about taxes, shipping costs, payment methods, etc., please visit the Order Info page.

News

Problem with NetDraw The current UCINET installation package includes NetDraw version 2.141. Unfortunately, this version of NetDraw has a problem reading UCINET system files, and we are temporarily unable to update the ...
Posted Dec 12, 2014, 1:22 PM by Steve Borgatti

Showing posts **1** - **1** of **7**. View more »

Current Version

Version 6.538 | 12 Dec 2014 This version replaces the old "spreadsheet editor" with a new "matrix editor",

/sites.google.com/site/ucinetsoftware/

SNA Tool: Pajek

http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Networks / Pajek

Program for Large Network Analysis

In January 2008 this page was replaced by Pajek Wiki.

Pajek runs on Windows and is free for noncommercial use.

DOWNLOAD Pajek

Data: test networks, GPHs, GEDs, PDB files.

Screenshots; History; Manual (pdf); Papers/presentations; Applications; in News; Examples: SVG, PDF.

How to ? English / Slovene / Japanese (problems with IE - download and use Acrobat reader).
Pajek nicely runs on Linux via Wine, **Converting Excel/text into Pajek format**.
Pajek to SVG animation. WoS to Pajek.

Slides from **NICTA workshop**, Sydney, Australia, June 14-17, 2005.
Slides from **workshop at GD'05**, Limerick, Ireland, Sept 11-14, 2005.
Pajek workshop at **XXVIII Sunbelt Conference**, St. Pete Beach, Florida, USA, January 22-27, 2008: slides.
Network analysis course at **ECPR Summer School in Methods and Techniques**, Ljubljana, Slovenia, July 30 - August 16, 2008.

W. de Nooy, A. Mrvar, V. Batagelj: ***Exploratory Social Network Analysis with Pajek***, CUP, January 2005; ESNA page.
P. Doreian, V. Batagelj, A. Ferligoj: ***Generalized Blockmodeling***, CUP, November 2004.

Chapter about Pajek: V. Batagelj, A. Mrvar: ***Pajek - Analysis and Visualization of Large Networks***.
in Jünger, M., Mutzel, P., (Eds.) *Graph Drawing Software*. Springer, Berlin 2003. p. 77-103 / Amazon.

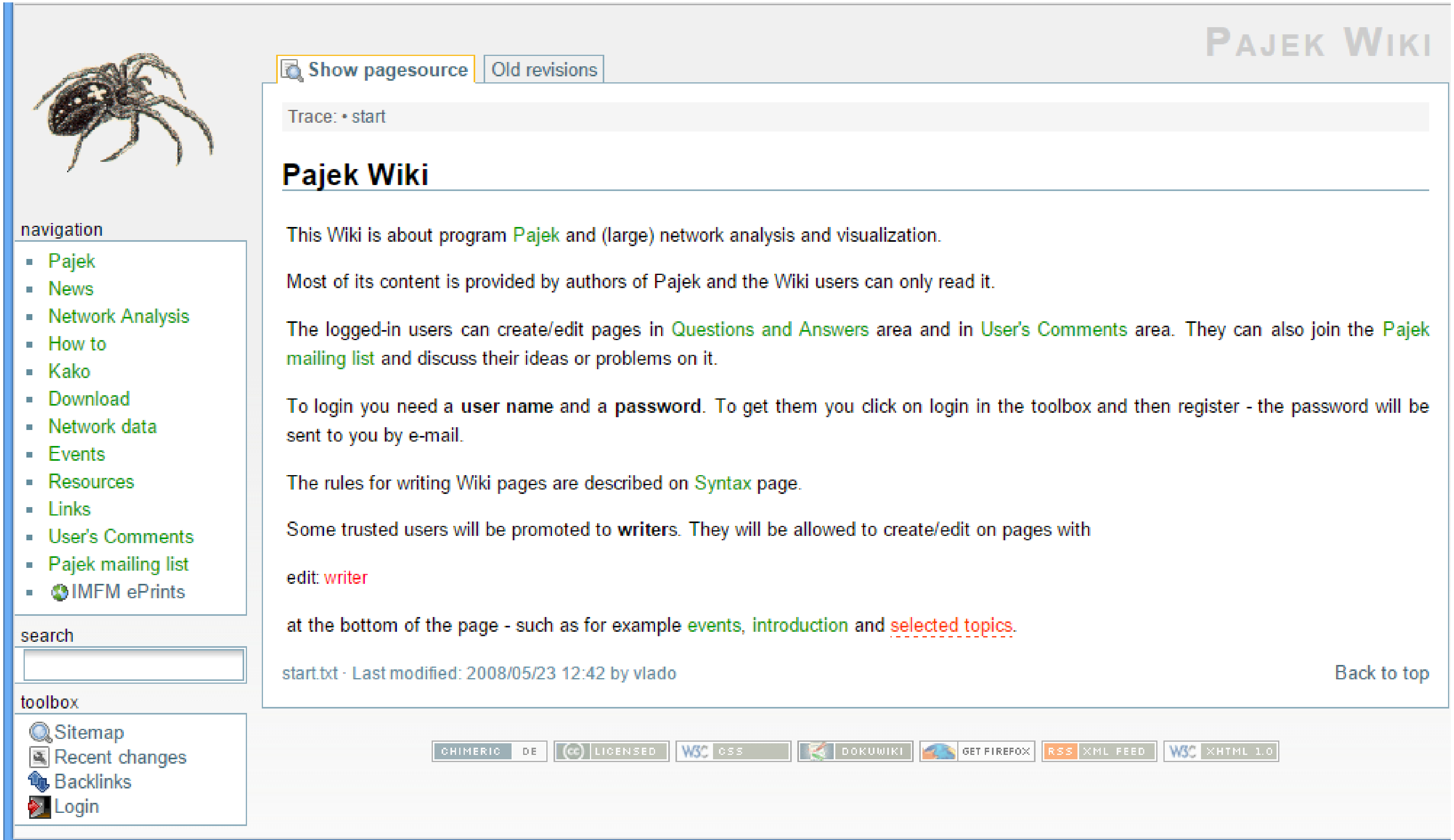
An improved version of the paper presented at Sunbelt'97 was published in Connections 21(1998)2, 47-57 - V. Batagelj, A. Mrvar: ***Pajek - Program for Large Network Analysis*** (PDF; PRISON.KIN).

Our layouts for ***Graph-Drawing Competitions***: GD95, GD96, GD97, GD98, GD99, GD00, GD01 and GD05.

Mladina (front page); Pajek in Koeln; PajekMan in Osoje (Ossiach, Austria);

SNA Tool: Pajek

http://pajek.imfm.si/doku.php



PAJEK WIKI

Show pagesource | Old revisions

navigation

- Pajek
- News
- Network Analysis
- How to
- Kako
- Download
- Network data
- Events
- Resources
- Links
- User's Comments
- Pajek mailing list
- IMFM ePrints

search

toolbox

- Sitemap
- Recent changes
- Backlinks
- Login

Trace: • start

Pajek Wiki

This Wiki is about program Pajek and (large) network analysis and visualization.

Most of its content is provided by authors of Pajek and the Wiki users can only read it.

The logged-in users can create/edit pages in Questions and Answers area and in User's Comments area. They can also join the Pajek mailing list and discuss their ideas or problems on it.

To login you need a **user name** and a **password**. To get them you click on login in the toolbox and then register - the password will be sent to you by e-mail.

The rules for writing Wiki pages are described on Syntax page.

Some trusted users will be promoted to **writers**. They will be allowed to create/edit on pages with

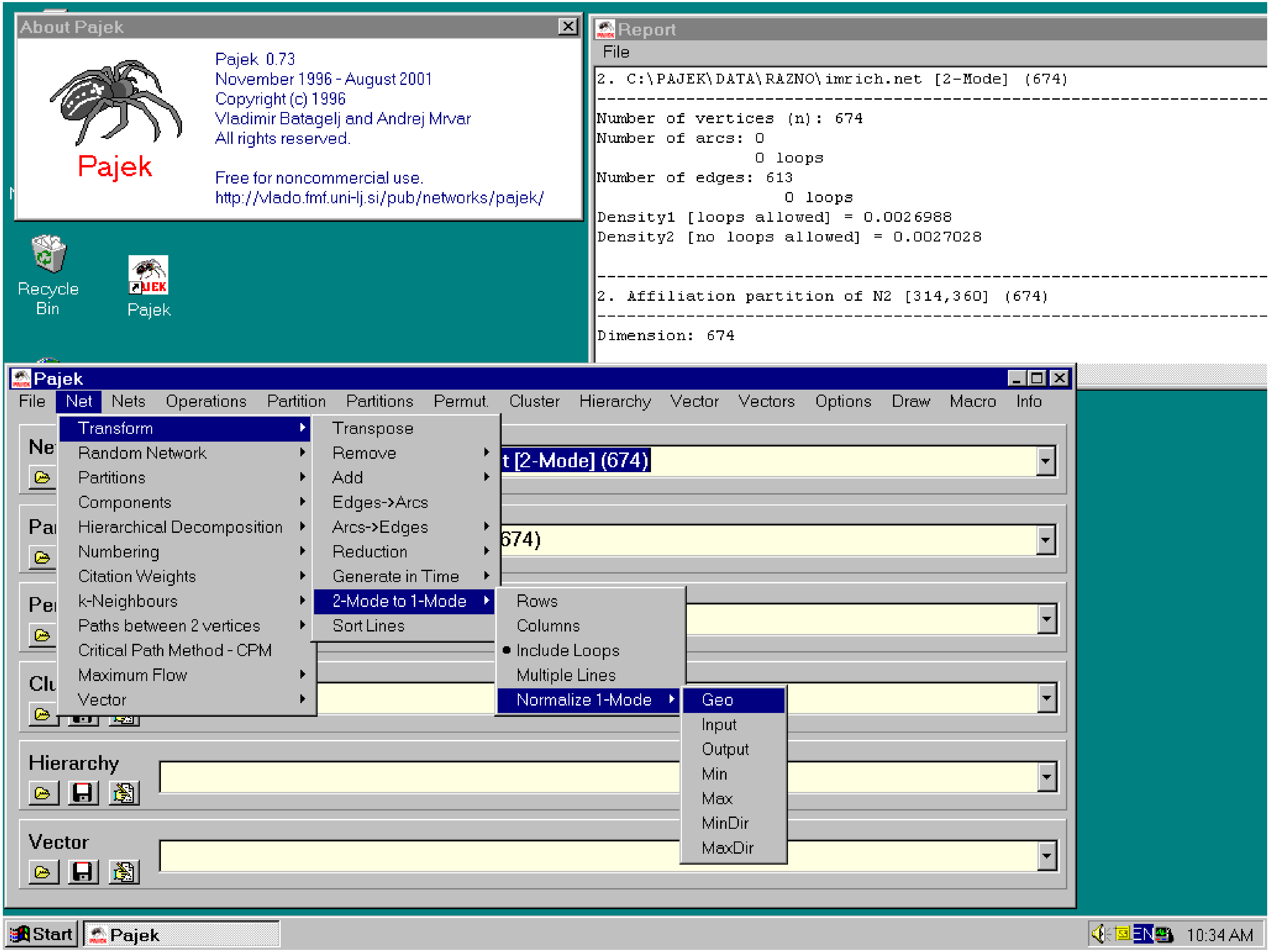
edit: writer

at the bottom of the page - such as for example events, introduction and selected topics.

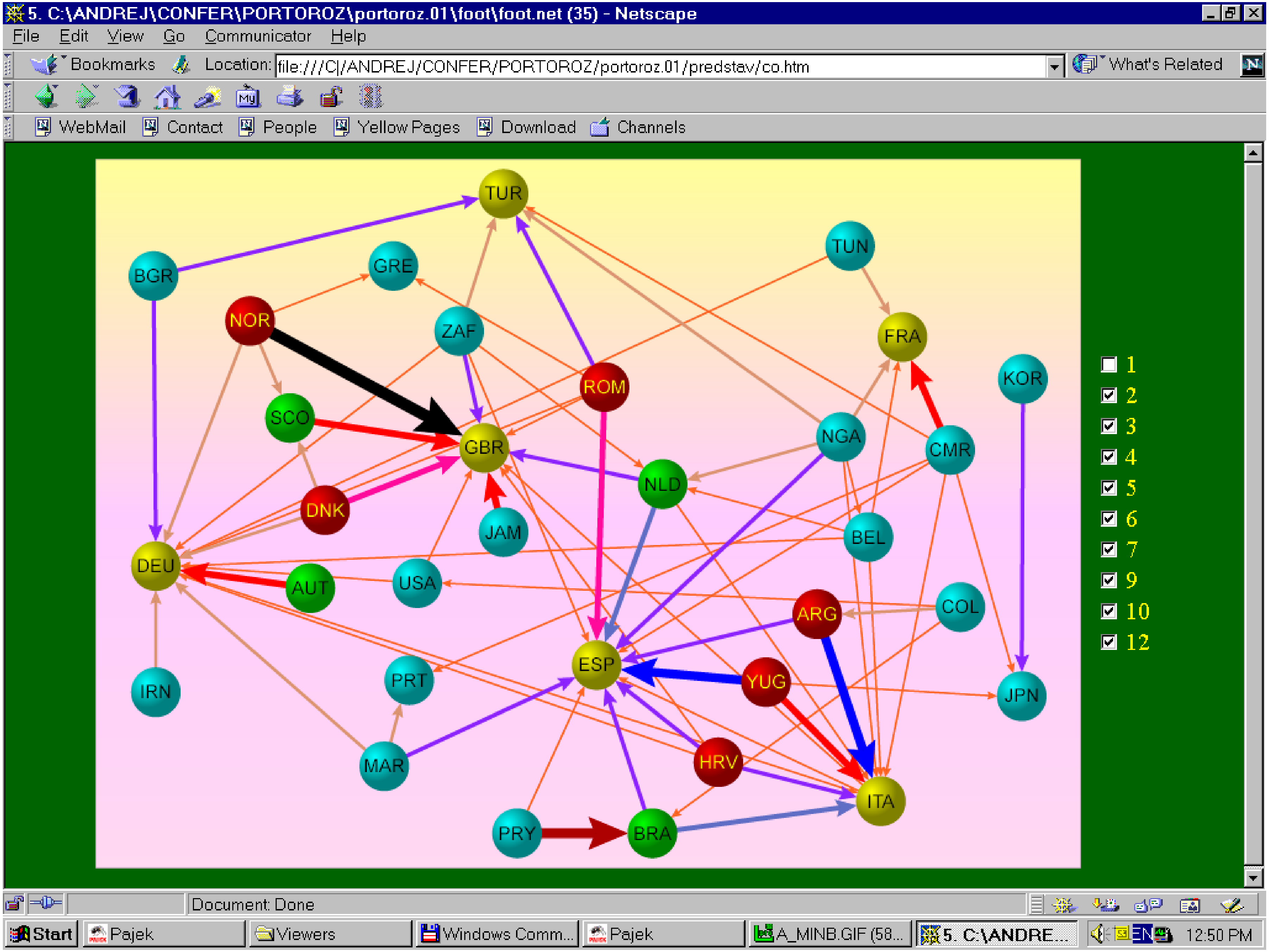
start.txt · Last modified: 2008/05/23 12:42 by vlado

Back to top

CHIMERIC DE | LICENSED | W3C CSS | DOKUWIKI | GET FIREFOX | RSS XML FEED | W3C XHTML 1.0



Source: http://vlado.fmf.uni-lj.si/pub/networks/doc/gd.01/Pajek9.png



Source: http://vlado.fmf.uni-lj.si/pub/networks/doc/gd.01/Pajek6.png

Application of SNA

Social Network Analysis

of

Research Collaboration

in

Information Reuse and Integration

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Example of SNA Data Source

home | browse | search | about

dblp
computer science bibliography

IRI 2010: Las Vegas, NV, USA

- **Proceedings of the IEEE International Conference on Information Reuse and Integration, IRI 2010, 4-6 August 2010, Las Vegas, Nevada, USA.** IEEE Systems, Man, and Cybernetics Society 2010
- Reda Alhajj, James B. D. Joshi, Mei-Ling Shyu: **Message from Program Co-Chairs.** 1
- Stuart Harvey Rubin, Shu-Ching Chen: **Forward.** 1
- Lotfi A. Zadeh: **Precisiation of meaning - toward computation with natural language.** 1-4
- Reda Alhajj, Shu-Ching Chen, Gongzhu Hu, James B. D. Joshi, Gordon K. Lee, Stuart Harvey Rubin, Mei-Ling Shyu, Lotfi A. Zadeh: **Panel title: Critical need for funding of basic and applied research in large-scale computing.** 1

Automation, Integration and Reuse across Various Apps

- László István Etesi, André Csillaghy, Lin-Ching Chang: **A message-based interoperability framework with application to astrophysics.** 1-6
- Awny Alnusair, Tian Zhao, Eric Bodden: **Effective API navigation and reuse.** 7-12
- Manabu Ohta, Ryohei Inoue, Atsuhiro Takasu: **Empirical evaluation of active sampling for CRF-based analysis of pages.** 13-18
- Qunzhi Zhou, Viktor K. Prasanna: **Workflow management of simulation based computation processes in transportation domain.** 19-24

Source: http://www.informatik.uni-trier.de/~ley/db/conf/iri/iri2010.html

Research Question

- RQ1: What are the scientific collaboration patterns in the IRI research community?

- RQ2: Who are the prominent researchers in the IRI community?

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011), "Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Methodology

- Developed a simple web focused crawler program to download literature information about all IRI papers published between 2003 and 2010 from IEEE Xplore and DBLP.
  - 767 paper
  - 1599 distinct author
- Developed a program to convert the list of coauthors into the format of a network file which can be readable by social network analysis software.
- *UCINet* and *Pajek* were used in this study for the social network analysis.

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Top10 prolific authors (IRI 2003-2010)

1. Stuart Harvey Rubin
2. Taghi M. Khoshgoftaar
3. Shu-Ching Chen
4. Mei-Ling Shyu
5. Mohamed E. Fayad
6. Reda Alhajj
7. Du Zhang
8. Wen-Lian Hsu
9. Jason Van Hulse
10. Min-Yuh Day

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011), "Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Data Analysis and Discussion

- **Closeness Centrality**
  - Collaborated widely
- **Betweenness Centrality**
  - Collaborated diversely
- **Degree Centrality**
  - Collaborated frequently
- **Visualization of Social Network Analysis**
  - Insight into the structural characteristics of research collaboration networks

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Top 20 authors with the highest closeness scores

Rank	ID	Closeness	Author
1	3	0.024675	Shu-Ching Chen
2	1	0.022830	Stuart Harvey Rubin
3	4	0.022207	Mei-Ling Shyu
4	6	0.020013	Reda Alhajj
5	61	0.019700	Na Zhao
6	260	0.018936	Min Chen
7	151	0.018230	Gordon K. Lee
8	19	0.017962	Chengcui Zhang
9	1043	0.017962	Isai Michel Lombera
10	1027	0.017962	Michael Armella
11	443	0.017448	James B. Law
12	157	0.017082	Keqi Zhang
13	253	0.016731	Shahid Hamid
14	1038	0.016618	Walter Z. Tang
15	959	0.016285	Chengjun Zhan
16	957	0.016285	Lin Luo
17	956	0.016285	Guo Chen
18	955	0.016285	Xin Huang
19	943	0.016285	Sneh Gulati
20	960	0.016071	Sheng-Tun Li

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011), "Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Top 20 authors with the highest betweeness scores

Rank	ID	Betweenness	Author
1	1	0.000752	Stuart Harvey Rubin
2	3	0.000741	Shu-Ching Chen
3	2	0.000406	Taghi M. Khoshgoftaar
4	66	0.000385	Xingquan Zhu
5	4	0.000376	Mei-Ling Shyu
6	6	0.000296	Reda Alhajj
7	65	0.000256	Xindong Wu
8	19	0.000194	Chengcui Zhang
9	39	0.000185	Wei Dai
10	15	0.000107	Narayan C. Debnath
11	31	0.000094	Qianhui Althea Liang
12	151	0.000094	Gordon K. Lee
13	7	0.000085	Du Zhang
14	30	0.000072	Baowen Xu
15	41	0.000067	Hongji Yang
16	270	0.000060	Zhiwei Xu
17	5	0.000043	Mohamed E. Fayad
18	110	0.000042	Abhijit S. Pandya
19	106	0.000042	Sam Hsu
20	8	0.000042	Wen-Lian Hsu

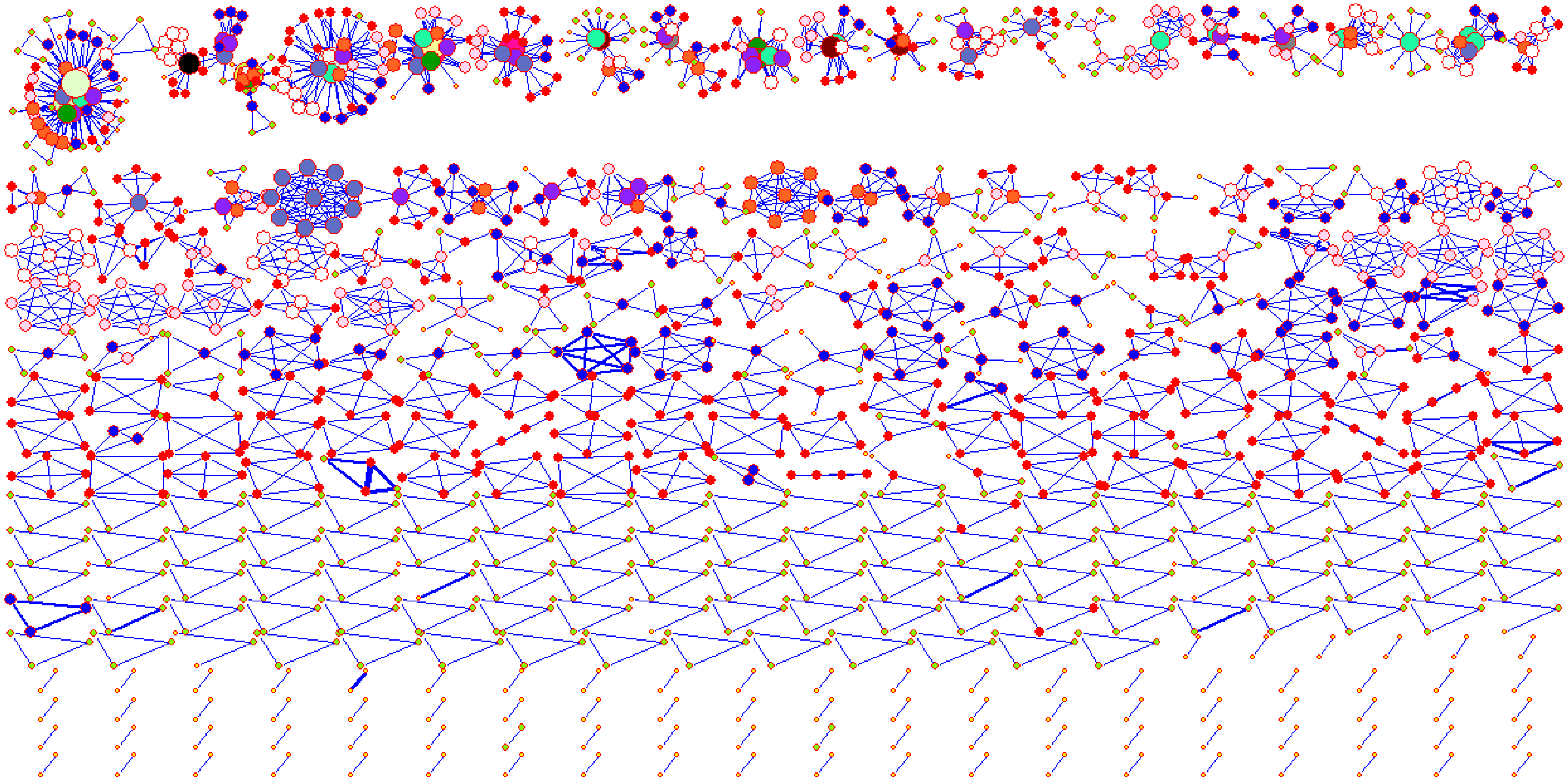
Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Top 20 authors with the highest degree scores

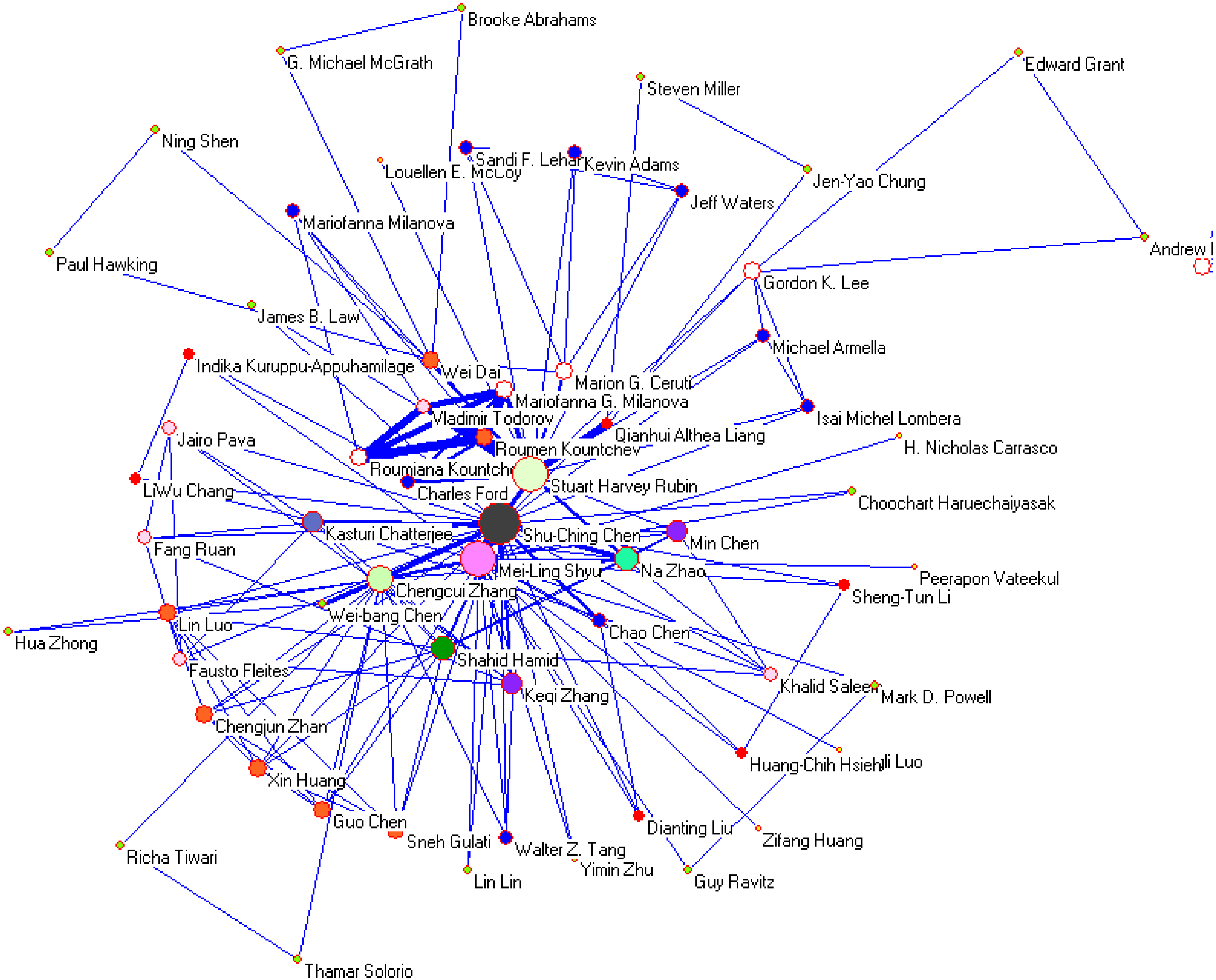
Rank	ID	Degree	Author
1	3	0.035044	Shu-Ching Chen
2	1	0.034418	Stuart Harvey Rubin
3	2	0.030663	Taghi M. Khoshgoftaar
4	6	0.028786	Reda Alhajj
5	8	0.028786	Wen-Lian Hsu
6	10	0.024406	Min-Yuh Day
7	4	0.022528	Mei-Ling Shyu
8	17	0.021277	Richard Tzong-Han Tsai
9	14	0.017522	Eduardo Santana de Almeida
10	16	0.017522	Roumen Kountchev
11	40	0.016896	Hong-Jie Dai
12	15	0.015645	Narayan C. Debnath
13	9	0.015019	Jason Van Hulse
14	25	0.013767	Roumiana Kountcheva
15	28	0.013141	Silvio Romero de Lemos Meira
16	24	0.013141	Vladimir Todorov
17	23	0.013141	Mariofanna G. Milanova
18	5	0.013141	Mohamed E. Fayad
19	19	0.012516	Chengcui Zhang
20	18	0.011890	Waleed W. Smari

Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Visualization of IRI (IEEE IRI 2003-2010) co-authorship network (global view)

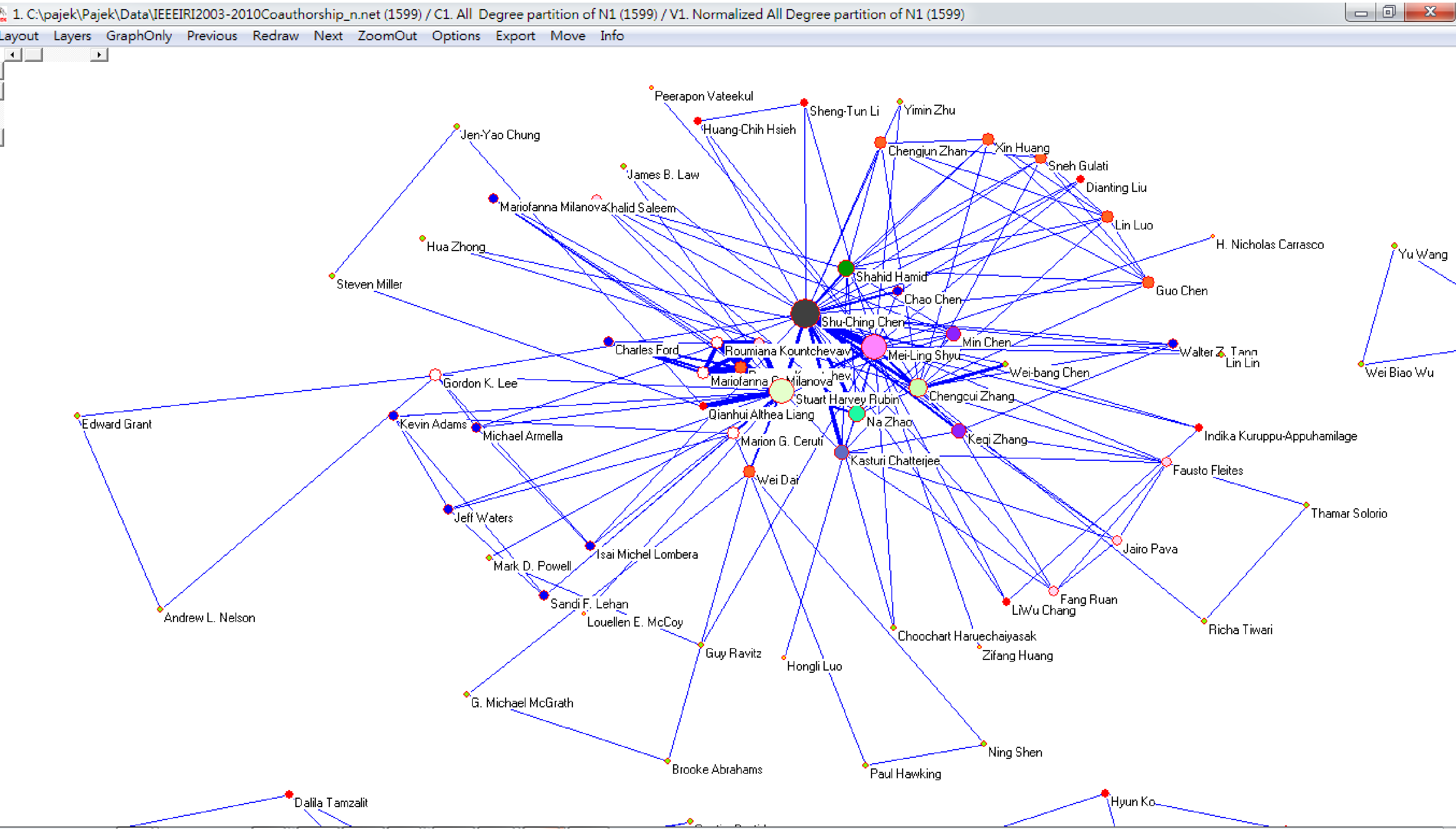


Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"



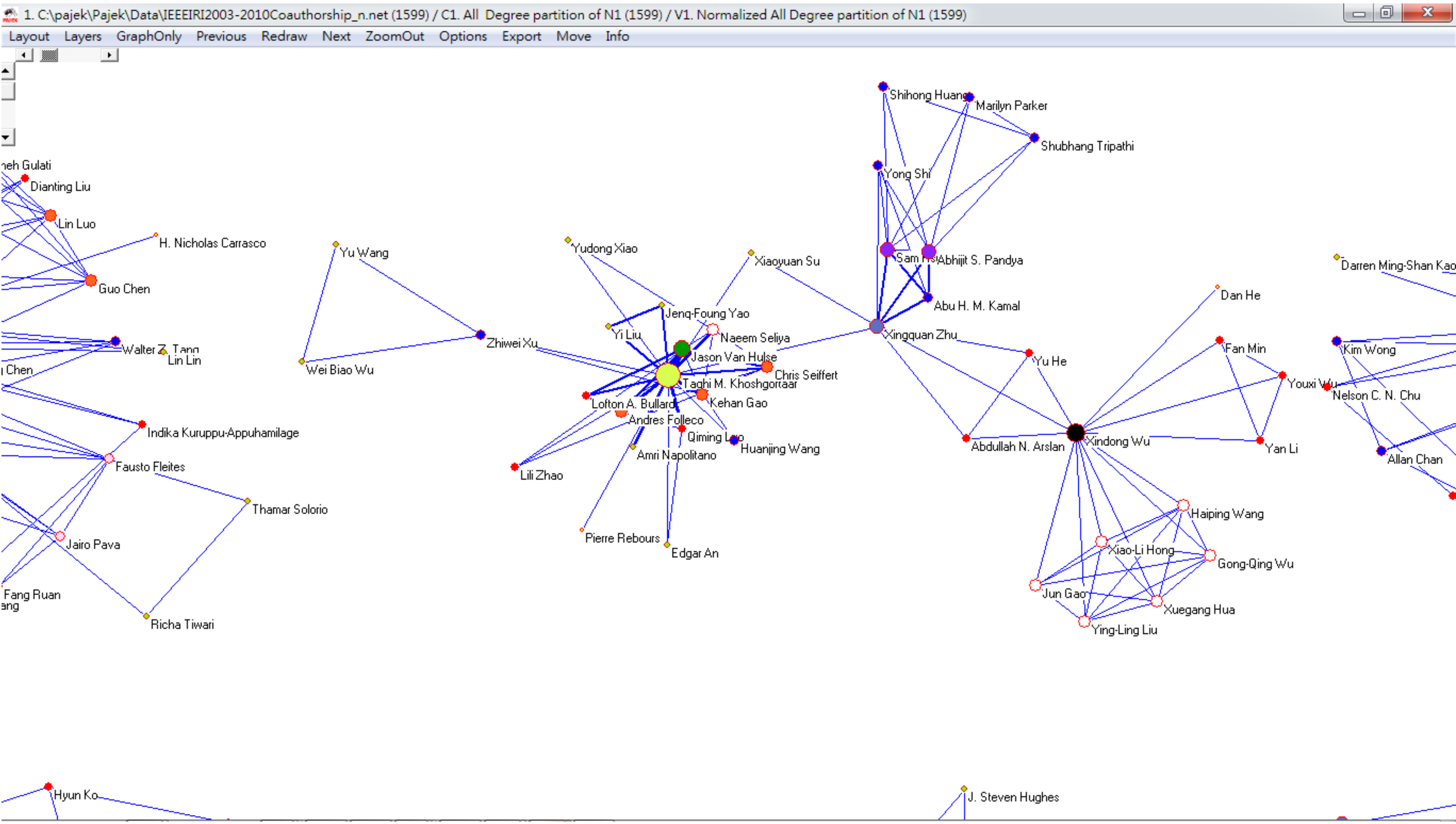
Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Visualization of Social Network Analysis



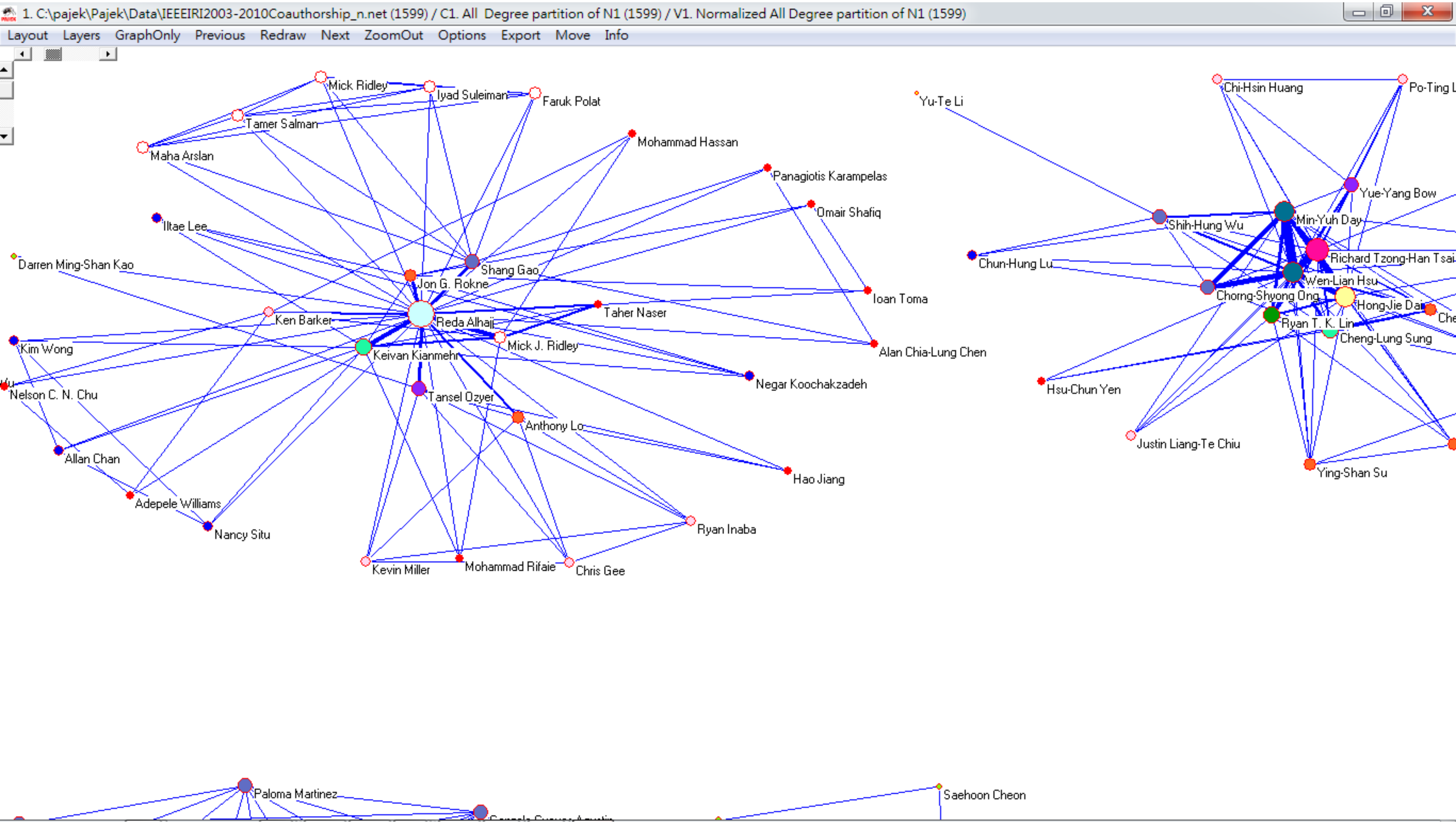
Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Visualization of Social Network Analysis



Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Visualization of Social Network Analysis



Source: Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011),
"Social Network Analysis of Research Collaboration in Information Reuse and Integration"

Summary

- Social Network Analysis (SNA)
  - Degree Centrality
  - Betweenness Centrality
  - Closeness Centrality
- Link Mining
- SNA Tools
  - UCINet
  - Pajek
- Applications of SNA

References

- Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," 2nd Edition, Springer. http://www.cs.uic.edu/~liub/WebMiningBook.html
- Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann. http://analyzingthesocialweb.com/course-materials.shtml
- Sentinel Visualizer, http://www.fmsasg.com/SocialNetworkAnalysis/
- Min-Yuh Day, Sheng-Pao Shih, Weide Chang (2011), "Social Network Analysis of Research Collaboration in Information Reuse and Integration," The First International Workshop on Issues and Challenges in Social Computing (WICSOC 2011), August 2, 2011, in Proceedings of the IEEE International Conference on Information Reuse and Integration (IEEE IRI 2011), Las Vegas, Nevada, USA, August 3-5, 2011, pp. 551-556.