Social Network Analysis with Gephi

Time: 13:00-16:00, 2018/5/14 (Monday)
Place: Engineering Building 5 (工程五館) A306, NCU
Host: Prof. Richard Tzong-Han Tsai (蔡宗翰 教授),
Department of Computer Science and Engineering, National Central University

Min-Yuh Day
戴敏育
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2018-05-14
戴敏育博士
(Min-Yuh Day, Ph.D.)
淡江大學資管系專任助理教授
中央研究院資訊科學研究所訪問學人
國立台灣大學資訊管理博士

Publications Co-Chairs, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in TExt (IEEE EM-RITE 2012-)

Workshop Chair, The IEEE International Conference on Information Reuse and Integration (IEEE IRI)
Outline

• Social Computing and Social Network Analysis (SNA)
• Social Network Analysis with Gephi
• Applications of SNA
Social Computing
Social Network Analysis (SNA)
Social Computing

• Social Network Analysis
• Link mining
• Community Detection
• Social Recommendation
Business Insights with Social Analytics
Analyzing the Social Web: Social Network Analysis
Devangana Khokhar (2015),
Gephi Cookbook, Packt Publishing

Source: http://www.amazon.com/Gephi-Cookbook-Devangana-Khokhar/dp/1783987405
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
Graph Theory
Graph
Graph

\( g = (V, E) \)
Vertex (Node)
Vertices (Nodes)
Edge
Edges
Arc
Arcs
Undirected Graph
Directed Graph

Source: https://www.youtube.com/watch?v=89mxOdwPfxA
Measurements of Social Network Analysis
Exploratory Network Analysis

1. see the network
   1st graph viz tool: Pajek (1996)
   Vladimir Batagelj, Andrej Mrvar

2. interact in real time
   Gephi prototype (2008)
   group, filter, compute metrics...

3. build a visual language
   size by rank, color by partition,
   label, curved edges, thickness...

Looking for a “Simple Small Truth”? What Data Visualization Should Do?

1. Make complex things **simple**
2. Extract **small** information from large data
3. Present **truth**, do not deceive

Measurements
Looking for Orderness in Data

Make varying 3 cursors simultaneously to extract meaningful patterns

at different levels

on multiple dimensions

at time scale

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf
“Zoom” cursor on Quantitative Data

Global
- connectivity
- density
- centralization

Local
- communities
- bridges between communities
- local centers vs periphery

Individual
- centrality
- distances
- neighborhood
- location
- local authority vs hub

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf
“Crossing” cursor on Quantitative Data

Social
- who with whom
- communities
- brokerage
- influence and power
- homophily

Semantic
- topics
- thematic clusters

Geographic
- spatial phenomena

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf
“Timeline” cursor on Temporal Data

Evolution of social ties
Evolution of communities
Evolution of topics

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf
# SNA Guideline

<table>
<thead>
<tr>
<th># nodes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 100</td>
<td>lists + edges in bonus, focus on qualitative data</td>
</tr>
</tbody>
</table>
| 100 - 1,000 | • easy to read, “obvious” patterns  
• focus on entities (in context)  
• metrics are tools to describe the graph (centrality, bridging...)  
• links help to build and interpret categories of entities  
  challenge: mix attribute crossing and connectivity |
| 1,000 - 50,000 | • hard to read, problem of “hidden signals”:  
  track patterns with various layouts and filtering  
• focus on structures  
• metrics are tools to build the graph (cosine similarity...)  
• categories help to understand the structure  
  challenge: pattern recognition |
| > 50,000  | require high computational power                                                             |

Degree

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

A: 2
B: 4
C: 2
D: 1
E: 1

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

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

Edges (Links): 5
Total Possible Edges: 10
Density: 5/10 = 0.5
Density

Nodes (n): 10
Edges (Links): 13
Total Possible Edges: \( \frac{n \times (n-1)}{2} = \frac{10 \times 9}{2} = 45 \)
Density: \( \frac{13}{45} = 0.29 \)
Diameter
Diameter
Geodesic Path (Shortest Path)

A → I : Diameter = 4
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.

Social Network Analysis (SNA)

• Degree Centrality
• Betweenness Centrality
• Closeness Centrality
Degree Centrality
Social Network Analysis: Degree Centrality
Social Network Analysis: Degree Centrality

<table>
<thead>
<tr>
<th>Node</th>
<th>Score</th>
<th>Standardized Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2/10 = 0.2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2/10 = 0.2</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>5/10 = 0.5</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>3/10 = 0.3</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>3/10 = 0.3</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>2/10 = 0.2</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>4/10 = 0.4</td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>3/10 = 0.3</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>1/10 = 0.1</td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>1/10 = 0.1</td>
</tr>
</tbody>
</table>
Betweenness Centrality
Betweenness centrality:

Connectivity

Number of shortest paths going through the actor
**Betweenness Centrality**

\[ C_B(i) = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}} \]

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

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

**Normalized Betweenness Centrality**

\[ C'_B(i) = \frac{C_B(i)}{\left[ \frac{(n-1)(n-2)}{2} \right]} \]

Number of pairs of vertices excluding the vertex itself

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

A: Betweenness Centrality = 0

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
Betweenness Centrality

B: Betweenness Centrality = 5

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
Betweenness Centrality

C: Betweenness Centrality = 0

C: B: 0/1 = 0
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
Betweenness Centrality

A: 0
B: 5
C: 0
D: 0
E: 0
Which Node is Most **Important**?
Which Node is Most Important?
Betweenness Centrality

\[ C_B(i) = \sum_{j<k} \frac{g_{ik}(i)}{g_{jk}} \]
Betweenness Centrality

A: Betweenness Centrality = 0

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
Closeness
Centrality
Social Network Analysis: Closeness Centrality

C: Closeness Centrality = 15/9 = 1.67
Social Network Analysis: Closeness Centrality

G: Closeness Centrality = 14/9 = 1.56
Social Network Analysis: Closeness Centrality

H: Closeness Centrality = 17/9 = 1.89
Social Network Analysis: Closeness Centrality

G: Closeness Centrality = 14/9 = 1.56
C: Closeness Centrality = 15/9 = 1.67
H: Closeness Centrality = 17/9 = 1.89
Social Network Analysis (SNA)
importance of neighbors
Eigenvector centrality
Eigenvector centrality:
Importance of a node depends on the importance of its neighbors
Social Network Analysis: Closeness Centrality

Sum of the reciprocal distances

\[ C_C(p_k) = \sum_{i=1}^{n} d(p_i, p_k)^{-1} \]

where \( d(p_j, p_k) \) is the geodesic distance (shortest paths) linking \( p_j, p_k \)
Social Network Analysis: Betweenness Centrality

$$C_B(p_k) = \sum_{i<j}^{n} \frac{g_{ij}(p_k)}{g_{ij}}; \quad i \neq j \neq k$$

where $g_{ij}$ is the geodesic distance (shortest paths) linking $p_i$ and $p_j$ and $g_{ij}(p_k)$ is the geodesic distance linking $p_i$ and $p_j$ that contains $p_k$.
Social Network Analysis: Degree Centrality

\[ C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k) \]

where \( a(p_i, p_k) = 1 \) if and only if \( p_i \) and \( p_k \) are connected by a line
0 otherwise

\[ C'_D(p_k) = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1} \]
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.

Tools of Social Network Analysis
Social Network Analysis (SNA) Tools

• NetworkX
• igraph
• Gephi
• UCINet
• Pajek
Tools of Social Network Analysis

• Focused Desktop Tools
  – Gephi
  – Ucinet
  – Pajek
  – NodeXL
  – Cytoscape

Source: https://www.researchgate.net/post/Which_open_source_software_is_best_for_network_data_analysis
Tools of Social Network Analysis

• Developer Tools
  – NetworkX
  – iGraph
  – SNAP
  – sigma.js
Gephi

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

Runs on Windows, Mac OS X and Linux.

Learn More on Gephi Platform »

Download FREE Gephi 0.9.1
Release Notes | System Requirements

Features
Quick start
Screenshots
Videos

Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:

Donate

APPLICATIONS
✓ Exploratory Data Analysis: intuition-oriented analysis by networks manipulations in real time.
✓ Link Analysis: revealing the underlying structures of associations between objects.
✓ Social Network Analysis: easy creation of social

Like Photoshop™ for graphs.
— the Community

LATEST NEWS
Gephi updates with 0.9.1 version

PAPERS

https://gephi.org/
UCINET Software

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:


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.

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.

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


Chapter about Pajek: V. Batagelj, A. Mrvar: Pajek - Analysis and Visualization of Large Networks.
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.

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

Pajek: analysis and visualization of large networks

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>32 bit</th>
<th>64 bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 10, 2016</td>
<td>4.10</td>
<td>Web Start</td>
<td>Web Start</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Install Shield</td>
<td>Install Shield</td>
</tr>
<tr>
<td>March 1, 2016</td>
<td>4.09</td>
<td>Install Shield</td>
<td>Install Shield</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Portable</td>
<td>Portable</td>
</tr>
<tr>
<td>Sept. 25, 2011</td>
<td>2.05</td>
<td>zip</td>
<td>zip</td>
</tr>
</tbody>
</table>

Pajek mailing list

Datasets

http://mrvar.fdv.uni-lj.si/pajek/
NodeXL

NodeXL: Network Overview, Discovery and Exploration for Excel

NodeXL Basic is a free, open-source template for Microsoft® Excel® 2007, 2010, 2013 and 2016 that makes it easy to explore network graphs. With NodeXL, you can enter a network edge list in a worksheet, click a button and see your graph, all in the familiar environment of the Excel window.

NodeXL Pro offers additional features that extend NodeXL Basic, providing easy access to social media network data streams, advanced network metrics, and text and sentiment analysis, and analyzing your data in Excel.
Cytoscape

http://www.cytoscape.org/
NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Features

- Python language data structures for graphs, digraphs, and multigraphs.
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g. text, images, XML records)
- Edges can hold arbitrary data (e.g. weights, time-series)
- Open source BSD license
- Well tested: more than 1800 unit tests, >90% code coverage
- Additional benefits from Python: fast prototyping, easy to teach, multi-platform

Documentation

- all documentation

Examples

- using the library

Reference

- all functions and methods

https://networkx.github.io/
igraph is a collection of network analysis tools with the emphasis on efficiency, portability and ease of use. igraph is open source and free. igraph can be programmed in R, Python and C/C++.

Recent news

R/igraph 1.0.0
June 24, 2015

Release Notes
This is a new major release, with a lot of UI changes. We tried to make it easier to use, with short and easy to remember, consistent function names. Unfortunately

http://igraph.org/redirect.html
SNAP for C++: Stanford Network Analysis Platform

Stanford Network Analysis Platform (SNAP) is a general purpose network analysis and graph mining library. It is written in C++ and easily scales to massive networks with hundreds of millions of nodes, and billions of edges. It efficiently manipulates large graphs, calculates structural properties, generates regular and random graphs, and supports attributes on nodes and edges. SNAP is also available through the NodeXL which is a graphical front-end that integrates network analysis into Microsoft Office and Excel.

Snap.py: SNAP for Python

Snap.py is a Python interface for SNAP. It provides performance benefits of SNAP, combined with flexibility of Python. Most of the SNAP C++ functionality is available via Snap.py in Python.

Stanford Large Network Dataset Collection

A collection of more than 50 large network datasets from tens of thousands of nodes and edges to tens of millions of nodes and edges. In includes social networks, web graphs, road networks, internet networks, citation networks, collaboration networks, and communication networks.

Tutorials

Tutorials on using SNAP, on methods to analyze large network data, on ways how to think about networks and how to model them at the level of network structure, and on methods to study evolution and dynamics of diffusion and cascading behavior in networks.

- Tutorial on Large Scale Network Analytics with SNAP will be held at WWW-15 conference, Florence, Italy, May 18, 2015. More info.

http://snap.stanford.edu/
Sigma is a JavaScript library dedicated to graph drawing. It makes easy to publish networks on Web pages, and allows developers to integrate network exploration in rich Web applications.

http://sigmajs.org/
Gephi

Gephi.app
Gephi: Social Network Analysis and Visualization
Network Analysis and Visualization with Gephi

Nodes and Edges
CSV Text Data for Gephi

Nodes1.csv
Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2

Edges1.csv
Source,Target
1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6

Source: http://www.martingrandjean.ch/gephi-introduction/
A = Degree centrality  
number of connexions

B = Closeness centrality  
closeness to the entire network

C = Betweenness centrality  
bridges nodes

D = Eigenvector centrality  
connection to well-connected nodes

Source: http://www.martingrandjean.ch/gephi-introduction/
Conference Participants

Source: http://www.martingrandjean.ch/analyse-de-reseau-thatcamp-et-communaute-des-humanites-numeriques-francophones/
Conference Participants

Source: http://www.martingrandjean.ch/analyse-de-reseau-thatcamp-et-communaute-des-humanites-numeriques-francophones/
Fruchterman Reingold

Source: http://www.martingrandjean.ch/gephi-introduction/
Force Atlas 2

Nodes’ color
Weighted In-Degree

Source: http://www.martingrandjean.ch/gephi-introduction/
Weighted In-Degree

Network Diameter
Betweenness Centrality
Closeness Centrality

Distance
The average graph-distance between all pairs of nodes. Connected nodes have graph distance 1. The diameter is the longest graph distance between any two nodes in the network. (i.e. How far apart are the two most distant nodes).

Betweenness Centrality: Measures how often a node appears on shortest paths between nodes in the network.
Closeness Centrality: The average distance from a given starting node to all other nodes in the network.
Eccentricity: The distance from a given starting node to the farthest node from it in the network.

Source: http://www.martingrandjean.ch/gephi-introduction/
Nodes’ color
Betweenness Centrality
## Gephi Supported Graph Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Edge List/Matrix Structure</th>
<th>XML Structure</th>
<th>Edge Weight</th>
<th>Attributes</th>
<th>Visualization Attributes</th>
<th>Attribute Default Value</th>
<th>Hierarchical Graphs</th>
<th>Dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DL Ucinet</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DOT Graphviz</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GDF</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GEXF</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GML</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>GraphML</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NET Pajek</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TLP Tulip</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VNA Netdraw</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Spreadsheet*</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Source: https://gephi.org/users/supported-graph-formats/
Gephi Supported Graph Formats

Do you need...

- GEXF
- Spreadsheet
- GraphML
- Guess GDF
- GML
- UCINET DL
- Netdraw VNA
- Graphviz DOT
- Pajek NET
- CSV
- Tulip TLP

Source: https://gephi.org/users/Supported-graph-formats/
The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

Runs on Windows, Mac OS X and Linux.

Learn More on Gephi Platform ➤

Download FREE Gephi 0.9.1

Release Notes | System Requirements

Features
Quick start
Screenshots
Videos

Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:

Donate

APPLICATIONS

- Exploratory Data Analysis: intuition-oriented analysis by networks manipulations in real time.
- Link Analysis: revealing the underlying structures of associations between objects.
- Social Network Analysis: easy creation of social

Like Photoshop™ for graphs. — the Community

LATEST NEWS

- Gephi updates with 0.9.1 version

PAPERS

https://gephi.org/
Download Gephi

Gephi is an open-source and multiplatform software distributed under the dual license CDDL 1.0 and GNU General Public License v3.

Official Releases
Release Notes | System Requirements | Installation instructions
Gephi 0.9.1 is the latest stable release.

Download Gephi for Mac OSX
Version 0.9.1

If you have an older Gephi on your computer, you should uninstall it first, see the installation instructions.

All downloads:
Download Gephi 0.9.1 for Mac OS X
Download Gephi 0.9.1 for Windows
Download Gephi 0.9.1 for Linux
Download Gephi 0.9.1 sources
Download Older Versions

Sources:
Gephi uses GitHub to host the source code and track issues. The trunk repository is the most up-to-date version but may be unstable. The last stable version is located in the release tab on GitHub.

Localization
Localization is available in French, Spanish, Japanese, Brazilian Portuguese, Russian, Chinese, Czech and German. In Gephi, simply go to Tools -> Languages to switch.

https://gephi.org/users/download/
Download Gephi

gephi-0.9.1-macos.dmg

Disk Image - 121.1 MB
Gephi 0.9.1
Gephi
Gephi:
New Project
Import
Nodes1.csv and Edges1.csv
to Gephi
Gephi New Project

Welcome to Gephi

Open recent

New Project

Samples

Les Miserables.gexf
Java.gexf
Power Grid.gml

Open at startup
Gephi Overview

1. Appearance
   - Nodes
   - Edges

2. Layout
   - Choose a layout

3. Graph

4. Context
   - Network Overview
     - Average Degree
     - Avg. Weighted Degree
     - Network Diameter
     - Graph Density
     - Modularity
     - PageRank
     - Connected Components
     - Node Overview
     - Avg. Clustering Coefficient
     - Eigenvector Centrality

5. Statistics
   - Dynamic
     - # Nodes
     - # Edges
     - Degree
     - Clustering Coefficient
   - Filters
Gephi Data Laboratory: Import Spreadsheet
Gephi Data Laboratory: Import Spreadsheet

Steps:
1. General options
2. Import settings

General options:
Choose a CSV file to import:

Separator: Edges table
As table:
Charset: UTF-8

Preview:

Invalid CSV file
Import Nodes1.csv to Gephi

Nodes1.csv
Id, Label, Attribute
1, John, 1
2, Carla, 2
3, Simon, 1
4, Celine, 2
5, Winston, 1
6, Diana, 2
Import Nodes1.csv to Gephi
Import Nodes1.csv to Gephi
Import Nodes1.csv to Gephi

Nodes1.csv

<table>
<thead>
<tr>
<th>Id</th>
<th>Label</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>John</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Carla</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Simon</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Celine</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Winston</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Diana</td>
<td>2</td>
</tr>
</tbody>
</table>
Import Edges1.csv to Gephi
Import Edges1.csv to Gephi

**Edges1.csv**

**Source, Target**

1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6

1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6
Import Edges1.csv to Gephi

Edges table
Import Edges1.csv to Gephi

Edges table
Import Edges1.csv to Gephi

Steps
1. General options
   2. Import settings

Import settings

- New columns are created with the specified type. A generated id is assigned if missing or already existing. Edges need 'Source' and 'Target' columns with the id of the nodes. If no 'Type' column is provided, all edges will be directed. If an edge already exists, attributes will be ignored, but the id will be preserved.

Imported columns:
- Source
  - String
- Target
  - String
- Create missing nodes

Finish button highlighted.
Import Edges1.csv to Gephi

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Type</th>
<th>Id</th>
<th>Label</th>
<th>Interval</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Directed</td>
<td>0</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Directed</td>
<td>1</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Directed</td>
<td>2</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Directed</td>
<td>3</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>Directed</td>
<td>4</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Directed</td>
<td>5</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Directed</td>
<td>6</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Directed</td>
<td>7</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>Directed</td>
<td>8</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>
Gephi Overview
Gephi Overview: Graph
Gephi Overview: Layout
Gephi Overview: Layout

Yifan Hu Proportional
Gephi Overview: Layout

Yifan Hu
Appearance: Nodes Color
Nodes Color / Attribute / Apply
Show Node Labels
Show Labels
Global Edges Labels
Labels
Labels Node Size
Labels Node Font Size
Labels Node Size
Labels Scaled
Labels Color
### Labels Color

<table>
<thead>
<tr>
<th>Font</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arial-BoldMT, 24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Font</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arial-BoldMT, 32</td>
<td></td>
</tr>
</tbody>
</table>
Gephi Statistics: Average Degree
Gephi Statistics: Average Degree

In-Degree Distribution

Out-Degree Distribution
Gephi Statistics: Avg. Weighted Degree

Weighted Degree Report

Results:
Average Weighted Degree: 1.500

Degree Distribution

Count

Value
Gephi Statistics: Network Diameter

Distance
The average graph-distance between all pairs of nodes. Connected nodes have graph distance 1. The diameter is the longest graph distance between any two nodes in the network. (i.e. How far apart are the two most distant nodes).

Directed
Undirected

Betweenness Centrality: Measures how often a node appears on shortest paths between nodes in the network.

Closeness Centrality: The average distance from a given starting node to all other nodes in the network.

Eccentricity: The distance from a given starting node to the farthest node from it in the network.

Graph Distance settings

Network Diameter
Network Overview
Average Degree
Avg. Weighted Degree
Connected Components
Node Overview
Avg. Clustering Coefficient
Eigenvector Centrality
Edge Overview
Avg. Path Length
Dynamic

# Nodes
# Edges
Degree
Clustering Coefficient

Node
Font: Arial-BoldMT, 24
Color:
Size:

Global Edges

Edge
Font: Arial-BoldMT, 32
Color:
Size:
Gephi Statistics: Network Diameter

Graph Distance Report

Parameters:
Network Interpretation: directed

Results:
Diameter: 1
Radius: 0
Average Path length: 1.0

Betweenness Centrality Distribution
Graph Density Report

Parameters:
Network Interpretation: directed

Results:
Density: 0.300

Network Overview:
- Average Degree: 3
- Avg. Weighted Degree: 1.5
- Network Diameter: 1
- Graph Density: 0.3

Modularity

PageRank

Connected Components

Node Overview:
- Avg. Clustering Coefficient
- Eigenvector Centrality

Edge Overview:
- Avg. Path Length: 1

Dynamic:
- # Nodes
- # Edges
- Degree
- Clustering Coefficient
Gephi Statistics: Modularity
Gephi Statistics: Modularity

Modularity Report

Parameters:
- Randomize: On
- Use edge weights: On
- Resolution: 1.0

Results:
- Modularity: 0.000
- Modularity with resolution: 0.000
- Number of Communities: 1

Size Distribution
Gephi Statistics: Connected Components

Connected Components settings

- Directed: Detects strongly & weakly connected components
- Undirected: Detects only weakly connected components

Network Overview:
- Average Degree: 3
- Avg. Weighted Degree: 1.5
- Network Diameter: 1
- Graph Density: 0.3
- Modularity: 0
- PageRank: Run

Connected Components: Run

Node Overview:
- Avg. Clustering Coefficient: Run
- Eigenvector Centrality: Run

Edge Overview:
- Avg. Path Length: 1

Dynamic:
- # Nodes: Run
- # Edges: Run
- Degree: Run
- Clustering Coefficient: Run
Gephi Statistics: Connected Components

Connected Components Report

Parameters:
Network Interpretation: directed

Results:
Number of Weakly Connected Components: 1
Number of Strongly Connected Components: 6

Size Distribution
Appearance Nodes Size
Appearance Nodes Size
Attribute / In-Degree
Appearance Nodes Size

Attribute / In-Degree / Min size / Max size / Apply
Appearance Edges
Attribute / Weight / Color
Appearance Edges
Attribute / Weight / Color / Apply
Gephi Data Laboratory
Gephi Preview: Show Labels
Gephi Preview: Default Straight
Gephi Preview: Default Straight
Open Gephi Samples
Gephi Samples
Gephi Samples
Les Miserables.gexf
Gephi Import Report

Source: Stream .gexf

Graph Type: Undirected
# of Nodes: 77
# of Edges: 254
Dynamic Graph: no
Dynamic Attributes: no
Multi Graph: no

INFO

More options...
- New graph
- Append Graph
Gephi Overview
Gephi Layout: Force Atlas
Gephi Layout: Contraction
Gephi Layout: Expansion
Gephi Layout: ForceAtlas 2
Gephi Layout: Fruchterman Reingold
Gephi Layout: Yifan Hu
Gephi Layout: Yifan Hu Proportional
## Gephi Data Laboratory: Nodes

<table>
<thead>
<tr>
<th>Id</th>
<th>Label</th>
<th>Interval</th>
<th>Modularity Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Myriel</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>Napoleon</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Labarre</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>Valjean</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Marguerite</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>MmeDeR</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>Isabeau</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Gervais</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Tholomyes</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>Listolier</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Fameuil</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>Blacheville</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>MlleBaptistine</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Favourite</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>Dahlia</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>Zepheine</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>Fantine</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>MmeThenardier</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>Thenardier</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>26</td>
<td>Cosette</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>Javert</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>28</td>
<td>Fauchelevent</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>Bamatabois</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>MmeMagloire</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>Perpetue</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>Simplice</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>Scauffaire</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>Woman1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>Judge</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>Champs Mathieu</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
## Gephi Data Laboratory: Edges

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Type</th>
<th>Id</th>
<th>Label</th>
<th>Interval</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>Undirected</td>
<td>0</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Undirected</td>
<td>1</td>
<td></td>
<td></td>
<td>8.0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>Undirected</td>
<td>2</td>
<td></td>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>Undirected</td>
<td>3</td>
<td></td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>Undirected</td>
<td>4</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>Undirected</td>
<td>5</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>Undirected</td>
<td>6</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>Undirected</td>
<td>7</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>Undirected</td>
<td>8</td>
<td></td>
<td></td>
<td>2.0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>Undirected</td>
<td>9</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>Undirected</td>
<td>10</td>
<td></td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>Undirected</td>
<td>11</td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>Undirected</td>
<td>12</td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>Undirected</td>
<td>13</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>Undirected</td>
<td>14</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>Undirected</td>
<td>15</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>Undirected</td>
<td>16</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>Undirected</td>
<td>17</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>Undirected</td>
<td>18</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>Undirected</td>
<td>19</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>Undirected</td>
<td>20</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>Undirected</td>
<td>21</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>Undirected</td>
<td>22</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>Undirected</td>
<td>23</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>Undirected</td>
<td>24</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>Undirected</td>
<td>25</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>Undirected</td>
<td>26</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>Undirected</td>
<td>27</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>Undirected</td>
<td>28</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>Undirected</td>
<td>29</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>
Gephi Data Laboratory: Export table to CSV file

<table>
<thead>
<tr>
<th>Id</th>
<th>Label</th>
<th>Interval</th>
<th>Modularity Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Myriel</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>Napoleon</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Labarre</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>Valjean</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Marguerite</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>MmeDeR</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>Isabeau</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Gervais</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Tholomyes</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>Listolier</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Fameuil</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>Blacheville</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>MileBaptistine</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Favourite</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>Dahlia</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>Zepheine</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>Fantine</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>MmeThenardier</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>Thenardier</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>26</td>
<td>Cosette</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>Javert</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>28</td>
<td>Fauchelevent</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>Bamatabois</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>MmeMagloire</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>Perpetue</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>Simplice</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>Scaufflaire</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>Woman1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>Judge</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>Champmathieu</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Gephi Preview
Gephi Overview: Text Labels

[Diagram of Gephi software interface with highlighted text label settings]
Gephi Overview: Text Labels
Comparison of Social Network Analysis (SNA) Tools
# General Comparison of SNA Tools

<table>
<thead>
<tr>
<th>Software</th>
<th>NETWORKX</th>
<th>IGRAPH</th>
<th>GEPHI</th>
<th>PAJEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE</td>
<td>LIBRARY</td>
<td>LIBRARY</td>
<td>STAND ALONE</td>
<td>STAND ALONE</td>
</tr>
<tr>
<td>PLATFORM</td>
<td>PYTHON</td>
<td>PYTHON\R\C LIBRARY</td>
<td>WINDOWS</td>
<td>WINDOWS</td>
</tr>
<tr>
<td>COMPUTATIONAL TIME</td>
<td>FAST</td>
<td>FAST</td>
<td>FAST</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>NO. OF NODES</td>
<td>1 MILLION</td>
<td>1 MILLION</td>
<td>0.15 MILLION</td>
<td>1 MILLION</td>
</tr>
</tbody>
</table>

Naheed Akhtar (2014)

### Network Types

**Supported by SNA Tools**

<table>
<thead>
<tr>
<th>Graph type</th>
<th>Networkx</th>
<th>IGraph</th>
<th>Gephi</th>
<th>Pajek</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Mode network</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2-Mode network Graph</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multirelational network</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>temporarily network Graph</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Naheed Akhtar (2014)

# Graph Layout

**Supported by SNA Tools**

<table>
<thead>
<tr>
<th>Layout</th>
<th>Networkx</th>
<th>IGraph</th>
<th>Pajek</th>
<th>Gephi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circular layout</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Random layout</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Spectral layout</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Spring layout</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Graphviz layout</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kamanda kawai</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Fruchterman reingold</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Force layout</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Naheed Akhtar (2014)

# Execution Time for SNA Features

<table>
<thead>
<tr>
<th>SNA Features</th>
<th>Networkx</th>
<th>IGraph</th>
<th>Gephi</th>
<th>Pajek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load time</td>
<td>54.67 sec.</td>
<td>3.707 sec</td>
<td>29 sec</td>
<td>3 sec</td>
</tr>
<tr>
<td>Degree centrality</td>
<td>58.57 sec</td>
<td>6.199 sec</td>
<td>4 sec</td>
<td>2 sec</td>
</tr>
<tr>
<td>Graph degree</td>
<td>60.87 sec</td>
<td>6.22 sec</td>
<td>4 sec</td>
<td>2 sec</td>
</tr>
<tr>
<td>Page rank</td>
<td>120.78 sec</td>
<td>9.81 sec</td>
<td>10 sec</td>
<td>No</td>
</tr>
<tr>
<td>Hits</td>
<td>57.23 sec</td>
<td>15.43</td>
<td>8 sec</td>
<td>No</td>
</tr>
<tr>
<td>Cliques</td>
<td>66.98 sec</td>
<td>9.35 sec</td>
<td>Na</td>
<td>No</td>
</tr>
<tr>
<td>Density</td>
<td>58.94 sec</td>
<td>3.302 sec</td>
<td>4 sec</td>
<td>No</td>
</tr>
<tr>
<td>Modularity</td>
<td>81.4 sec</td>
<td>9 sec</td>
<td>30 sec</td>
<td>6 sec</td>
</tr>
<tr>
<td>Network diameter</td>
<td>35 sec</td>
<td>3.51 sec</td>
<td>120 sec</td>
<td>No</td>
</tr>
<tr>
<td>Core</td>
<td>65.84 sec</td>
<td>6.532 sec</td>
<td>No</td>
<td>1 sec</td>
</tr>
<tr>
<td>Cohesion</td>
<td>No</td>
<td>8.943 sec</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>3303.99 sec</td>
<td>1800 sec</td>
<td>1200 sec</td>
<td>108 sec</td>
</tr>
<tr>
<td>Hub</td>
<td>76.57</td>
<td>5.831 sec</td>
<td>3 sec</td>
<td>No</td>
</tr>
<tr>
<td>Authority</td>
<td>Array is to big</td>
<td>6.783 sec</td>
<td>3 sec</td>
<td>No</td>
</tr>
</tbody>
</table>

### Comparative analysis of Social Networking Analysis tools

<table>
<thead>
<tr>
<th>Software</th>
<th>Pajek</th>
<th>Gephi</th>
<th>Social Network Visualizer</th>
<th>Netlytic</th>
<th>Graphviz</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Version</strong></td>
<td>1.26</td>
<td>0.7 alpha</td>
<td>1.56 Beta</td>
<td>Tier 1,2,3</td>
<td>2.38.0</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Stand-alone software</td>
<td>Stand-alone software</td>
<td>Stand-alone software</td>
<td>Stand-alone software</td>
<td>Stand-alone software</td>
</tr>
<tr>
<td><strong>Platform</strong></td>
<td>Windows</td>
<td>Java</td>
<td>Windows</td>
<td>Windows</td>
<td>Windows</td>
</tr>
<tr>
<td><strong>License</strong></td>
<td>Free</td>
<td>GNU GPL</td>
<td>Free</td>
<td>Tier 1,2 (Free)</td>
<td>Free</td>
</tr>
<tr>
<td><strong>Expectable Computing Time</strong></td>
<td>Fast(C)</td>
<td>Medium(JAVA)</td>
<td>Fast(C)</td>
<td>Medium(JAVA) Tier 3 (CS)</td>
<td>Fast(C)</td>
</tr>
<tr>
<td><strong>Tractable number of nodes</strong></td>
<td>500000 nodes</td>
<td>150000 nodes</td>
<td>100000 nodes</td>
<td>300000 nodes</td>
<td>1400000 nodes</td>
</tr>
<tr>
<td><strong>Time to load 10^6 nodes and 10^6 edges</strong></td>
<td>24 seconds</td>
<td>40 seconds</td>
<td>46 seconds</td>
<td>50 seconds</td>
<td>34 seconds</td>
</tr>
<tr>
<td><strong>File formats</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Pajek(.net)</strong></td>
<td>No</td>
<td>Important Only</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>GraphML</strong></td>
<td>Export only</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>DL</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>GEXF</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Graph types</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Two-mode graphs</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Multi-relational graphs</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Temporality</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Visualization layouts</strong></td>
<td>FruchtermanReingold</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Kamada Kawai</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Other spring layouts</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Indicators</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Degree centrality</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Betweenness centrality</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Closeness centrality</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Dyad census</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Triad census</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>HITS</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Page Rank</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Clustering Algorithms</strong></td>
<td>Edge Betweenness</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Walktrap</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Spinglass</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Dendogram Display</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

## Comparative analysis of Social Networking Analysis tools

<table>
<thead>
<tr>
<th>Software</th>
<th>Pajek</th>
<th>Gephi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>1.26</td>
<td>0.7 alpha</td>
</tr>
<tr>
<td>Type</td>
<td>Stand-alone software</td>
<td>Stand-alone software</td>
</tr>
<tr>
<td>Platform</td>
<td>Windows</td>
<td>Java</td>
</tr>
<tr>
<td>License</td>
<td>Free</td>
<td>GNU GPL</td>
</tr>
<tr>
<td>Expectable Computing Time</td>
<td>Fast(C)</td>
<td>Medium(JAVA)</td>
</tr>
<tr>
<td>Tractable number of nodes</td>
<td>5000000 nodes</td>
<td>1500000 nodes</td>
</tr>
<tr>
<td>Time to load $10^5$ nodes and $10^6$ edges</td>
<td>24 seconds</td>
<td>40 seconds</td>
</tr>
</tbody>
</table>

### File formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Pajek (.net)</th>
<th>GML</th>
<th>GraphML</th>
<th>DL</th>
<th>GEXF</th>
<th>Graph types</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Export only</td>
<td>Yes</td>
<td>No</td>
<td>Two-mode graphs</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Export only</td>
<td>Yes</td>
<td>No</td>
<td>Multi-relational graphs</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Export only</td>
<td>Yes</td>
<td>No</td>
<td>Temporality</td>
</tr>
<tr>
<td>Important Only</td>
<td>Yes</td>
<td>No</td>
<td>Export only</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

## Comparative analysis of Social Networking Analysis tools

<table>
<thead>
<tr>
<th>Software</th>
<th>Pajek</th>
<th>Gephi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visualization layouts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FruchtermanReingold</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Kamada Kawai</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Other spring layouts</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree centrality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dyad census</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Triad census</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HITS</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Page Rank</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Clustering Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge Betweenness</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Walktrap</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Spinglass</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Dendogram Display</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

A Survey of Tools for Community Detection and Mining in Social Networks
(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>Tools</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>4.1</td>
<td>0.9.1</td>
<td>1.0.1</td>
</tr>
<tr>
<td>Website</td>
<td>vlado.fmf.uni-lj.si/pub/networks/pajek/</td>
<td>gephi.org</td>
<td>igraph.org</td>
</tr>
<tr>
<td>Type</td>
<td>Software</td>
<td>Software</td>
<td>Library</td>
</tr>
<tr>
<td>Platform</td>
<td>Windows</td>
<td>Windows, Mac OS X, Linux</td>
<td>R, Python, C/C++</td>
</tr>
<tr>
<td>License</td>
<td>Free and Commercial</td>
<td>Open source and free</td>
<td>Open source and free</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>1 Billion</td>
<td>0.6 Million</td>
<td>&gt;1 Million</td>
</tr>
<tr>
<td>Computing Time</td>
<td>Medium</td>
<td>Fast</td>
<td>Fast</td>
</tr>
</tbody>
</table>

# A Survey of Tools for Community Detection and Mining in Social Networks

(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>Network Type</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-mode network</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2-mode network</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multi-relational network</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Temporal (dynamic) network</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

# A Survey of Tools for Community Detection and Mining in Social Networks

(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>File Format</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>.net</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>.gml</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>.graphml</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>.txt</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>.csv</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>.pajek</td>
<td>Yes</td>
<td>Import only</td>
<td>Yes</td>
</tr>
<tr>
<td>.dl</td>
<td>No</td>
<td>Yes</td>
<td>Import only</td>
</tr>
<tr>
<td>.graphdb</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>.dot</td>
<td>No</td>
<td>Yes</td>
<td>Export only</td>
</tr>
</tbody>
</table>
## A Survey of Tools for Community Detection and Mining in Social Networks
(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>Visualization Layout</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamada-kawai</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Fruchterman-Reingold</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Other spring layout</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Circular layout</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Random layout</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Force atlas layout</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Spectral layout</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Tree layout</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Visualization using igraph and Gephi

Visualization of American College Football using igraph with kamada-kawai layout.

Visualization of American College Football using Gephi with force-directed layout.

# A Survey of Tools for Community Detection and Mining in Social Networks

(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>Network Metrics</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree Centrality</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Closeness Centrality</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Network Diameter</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dyad Census</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Triad Census</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HITS</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Page Rank</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

A Survey of Tools for Community Detection and Mining in Social Networks
(Maivizhi et al., 2016)

<table>
<thead>
<tr>
<th>Community Algorithms</th>
<th>Pajek</th>
<th>Gephi</th>
<th>igraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Louvain Method</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Edge Betweenness</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Greedy Method</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Modularity Method</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Clique Percolation Method</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Label Propagation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Eigen Vector</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Random Walk</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Statistical Method</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Community Detection using Gephi and igraph

Community detection using **Gephi** with **Louvain method**.

Community detection using **igraph** with **Fast Greedy Algorithm**.

Application of SNA

Social Network Analysis of Research Collaboration in Information Reuse and Integration

Example of SNA Data Source

IRI 2010: Las Vegas, NV, USA


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?

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.

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

Data Analysis and Discussion

• **Closeness Centrality**
  – Collaborated widely

• **Betweeness Centrality**
  – Collaborated diversely

• **Degree Centrality**
  – Collaborated frequently

• **Visualization of Social Network Analysis**
  – Insight into the structural characteristics of research collaboration networks

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

### Top 20 authors with the highest betweenness scores

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

### Top 20 authors with the highest degree scores

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

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

Visualization of Social Network Analysis

Visualization of Social Network Analysis
Visualization of Social Network Analysis

Summary

• Social Computing and Social Network Analysis (SNA)
• Social Network Analysis with Gephi
• Applications of SNA
References

• Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann
  http://analyzingthesocialweb.com/course-materials.shtml
• Devangana Khokhar (2015), Gephi Cookbook, Packt Publishing
• 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.
  exploring and manipulating networks”, International AAAI Conference on Weblogs and
  Social Media.
• Agrawal, H., Thakur, A., Slathia, R., & Sumangali, K. (2015). A Comparative Analysis of
• Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International
  Conference on Communication Systems and Network Technologies (CSNT)
• Maivizhi, R., S. Sendhilkumar, G. S. Mahalakshmi (2016), "A Survey of Tools for
  Community Detection and Mining in Social Networks." In Proceedings of the International
Social Network Analysis with Gephi

Time: 13:00-16:00, 2018/5/14 (Monday)
Place: Engineering Building 5 (工程五館) A306, NCU
Host: Prof. Richard Tzong-Han Tsai (蔡宗翰 教授),
Department of Computer Science and Engineering, National Central University

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2018-05-14