Before starting...

A copy of these slides and all other materials you will need during this tutorial can be obtained from any of the DVDs or USB memory sticks being passed around right now

• Please register, download, and decompress Sci² from http://sci2.cns.iu.edu
  • If you have any problems, ask

• Please try opening PostScript test - chessboard.ps
  • You should see

  ➢ Additional Datasets http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets
  ➢ Additional Plugins http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
- Sci² tool visualizations
  - Bipartite networks
  - Map of Science
- Outlook and Q&A
- Adjourn
Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of datasets representing the structure and dynamics of complex systems.

Macroscopes provide a vision of the whole, helping us synthesize the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details.

Rather than making things larger or smaller, macroscopes let us observe what is too great, slow, or complex for the human eye and mind to notice and comprehend.
Plug-and-Play Macroscopes

While microscopes and telescopes are physical instruments, macroscopes are continuously changing bundles of software plugins.

Macroscopes make it easy to:
- Select and combine not only domain-specific algorithms and bridges to existing tools but also to meet the cross-cutting infrastructural requirements needed for a scientifically rigorous cyberinfrastructure.
- Put together plugins to create customized tools:
  - Share plugins via email, flash drives, or online
  - Simply drop plugins into the tool they appear in the menu, ready to use
  - Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube.

OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, modularized service platform.
  - Widely used in industry for over 10 years.
- CIShell provides “sockets” into which algorithms, tools, and datasets can be plugged using a wizard-driven process.
CIShell – Integrate New Algorithms

About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CIShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithm integration support is built in for Java and most other programming languages. Being Java-based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License. CIShell is the basis of Network Workbench, TaxTrend, Sci² and the upcoming Epic tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CIShell Features

A framework for easy integration of new and existing algorithms written in any programming language

Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

Learn More...

- CIShell Papers
- CIShell Powered Tools
- Algorithms
- Plugins (coming soon)
- Misc. Tool Documentation
- CIShell Web Services (coming soon)
- Screenshots

Getting Started...

- Documentation & Developer Resources
- Download

Getting Involved...

- Contact Us

CIShell Developer Guide is at http://cishell.wiki.cns.iu.edu

Additional Sci² Plugins are at http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins

OSGi/CIShell Adoption

- A number of other projects recently adopted OSGi and/or CIShell:
- Cytoscape (http://cytoscape.org) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).
- MAEviz (https://wiki.ncsa.uiuc.edu/display/MAE/Home) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.
- Taverna Workbench (http://taverna.org.uk) Developed by the myGrid team (http://mygrid.org.uk) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.
- TEXTtrend (http://texttrend.org) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corporuses with an inherently temporal component.
- DynaNets (http://www.dynanets.org) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
- As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
  - Sci² Tool – Advanced Topics
  - Outlook and Q&A
  - Adjourn
Science of Science (Sci²) Tool
http://sci2.cns.iu.edu

- Built on CShell/OSGi
- Explicitly designed for science of science research and practice
- Well-documented
- Easy to use
- Empowers average users to run common studies and expert users to perform novel research
- Advanced algorithms

- Effective visualizations
- Carefully designed and documented common workflows
- Full logging and data history for perfect replication of studies
- Free and open source software
- Anyone can review and extend the code, or use it for commercial purposes

Let’s make science metrics more scientific

To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says Julia Lane.

Sci² Tool

An OSGi/CShell-powered tool with NWB plugins and many new scientometric and visualization plugins

Sci² Tool Visualizations

Choropleth Map

Circular Hierarchy

Cognitive and Neuroscience at the NSF: 2007-2011
Sci² Tool Usage at National Institutes of Health

Mapping Transdisciplinary Tobacco Use Research Centers Publications: Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks. Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812

Sci² Tool now supports Web services and serves as a visual interface to publically available NIH RePORT Expenditure and Results. RePORTER/RePORTER data provided by NIH.
First time portrait of intramural research conducted by the U.S. Department of Agriculture (USDA) presented at the VIVO Conference 2012.
How did cognitive neuroscience of attention emerge from neurobiology and psychology, 1980–2005? Author co-citation analysis and Pfnet is used to trace prospectively the development of the field from its precursor disciplines: cognitive psychology, single cell neurophysiology, neuropsychology, and evoked potential research.
By 1990 a distinct cognitive neuroscience specialty cluster emerges, dominated by authors engaged in brain imaging research.


**Type of Analysis vs. Level of Analysis**

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of science, all of USA, all of research infrastructure</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in PNAS</td>
<td>113 Years of PNAS Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping an intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td></td>
<td></td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF author networkNIH’s scientific workforce</td>
<td></td>
<td>NIH’s scientific workforce</td>
</tr>
</tbody>
</table>
Type of Analysis vs. Level of Analysis
Covered Today

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping in 20-years of Physics</td>
<td></td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td></td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td></td>
<td>Ord/Topic maps of NIH funding</td>
<td></td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network NIH’s core competency</td>
<td></td>
</tr>
</tbody>
</table>

Sci² Tool – Supported Data Formats

**Input:**

**Network Formats**
- GraphML (*.xml or *.graphml)
- XGMML (*.xml)
- Pajek .NET (*.net)
- NWB (*.nwb)

**Scientometric Formats**
- ISI (*.isi)
- Bibtex (*.bib)
- Endnote Export Format (*.enw)
- Scopus csv (*.scopus)
- NSF csv (*.nsf)

**Other Formats**
- Pajek Matrix (*.mat)
- TreeML (*.xml)
- Edgelist (*.edge)
- CSV (*.csv)

**Output:**

**Network File Formats**
- GraphML (*.xml or *.graphml)
- Pajek .MAT (*.mat)
- Pajek .NET (*.net)
- NWB (*.nwb)
- XGMML (*.xml)
- CSV (*.csv)

**Image Formats**
- JPEG (*.jpg)
- PDF (*.pdf)
- PostScript (*.ps)

Formats are documented at [http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/2.3+Data+Formats](http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/2.3+Data+ Formats).
Sci² Tool – Supported Tools

Gnuplot
portable command-line driven interactive data and function plotting utility [http://www.gnuplot.info](http://www.gnuplot.info/).

GUESS
exploratory data analysis and visualization tool for graphs and networks. [https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS](https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS).

Simply add `org.textrend.visualization.cytoscape_0.0.3.jar` into your `/plugin` directory.
Restart Sci² Tool
Cytoscape now shows in the Visualization Menu

Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn
Sci2 v1.0 Alpha
Can be freely downloaded for all major operating systems from 
http://sci2.cns.iu.edu

Select your operating system from the pull down menu and download.
Unpack into a /sci2 directory.
Run /sci2/sci2.exe

Sci2 Manual is at 
http://sci2.wiki.cns.iu.edu

Cite as
Sci2 Team. (2009). Science of Science (Sci2) Tool, 
Indiana University and SciTech Strategies, 
http://sci2.cns.iu.edu

Sci2 v1.0 alpha
- Supports ASCII UTF-8 characters
- Web-based Yahoo! and offline geocoders
- New visualizations for Temporal, Topical, 
Geographical, and Bipartite data
- Customizable stop word lists
- New home page, wiki-based tutorial
- Reader for Google Scholar
- Gephi and R support
- Bug fixes, streamlined workflows

Sci2 runs on Windows, Mac, and Linux.
Decompress the archive and run sci2.exe
**Sci² Tool Interface Components**

*See also* [http://sci2.wiki.cns.iu.edu/2.2+User+Interface](http://sci2.wiki.cns.iu.edu/2.2+User+Interface)

**Use**

- **Menu** to read data, run algorithms.
- **Console** to see work log, references to seminal works.
- **Data Manager** to select, view, save loaded, simulated, or derived datasets.
- **Scheduler** to see status of algorithm execution.

All workflows are recorded into a log file (see `/sci2/logs/…`), and soon can be re-run for easy replication. If errors occur, they are saved in an error log to ease bug reporting.

All algorithms are documented online; workflows are given in tutorials, see Sci² Manual at [http://sci2.wiki.cns.iu.edu](http://sci2.wiki.cns.iu.edu)

---

**Tutorial Overview**

- **Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools**
- **Sci² Tool Basics**
  - Download and run the Sci² Tool
  - **Walkthrough: Load, analyze, and visualize a network**
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- **Sci² Tool – Advanced Topics**
- **Outlook and Q&A**
- **Adjourn**
Florentine families related through business ties (specifically, recorded financial ties such as loans, credits and joint partnerships) and marriage alliances.

Node attributes
- Wealth: Each family's net wealth in 1427 (in thousands of lira).
- Priorates: The number of seats on the civic council held between 1282-1344.
- Totalities: Number of business/marriage ties in complete dataset of 116 families.

Edge attributes:
- Marriage T/F
- Business T/F

“Substantively, the data include families who were locked in a struggle for political control of the city of Florence around 1430. Two factions were dominant in this struggle: one revolved around the infamous Medicis, the other around the powerful Strozzis.”


<table>
<thead>
<tr>
<th>Nodes</th>
<th>id</th>
<th>label</th>
<th>wealth</th>
<th>totalities</th>
<th>priorates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>&quot;Acciaiuoli&quot;</td>
<td>10</td>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&quot;Albizzi&quot;</td>
<td>36</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>&quot;Barbadori&quot;</td>
<td>55</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>&quot;Bischeri&quot;</td>
<td>44</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>&quot;Castellani&quot;</td>
<td>20</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>&quot;Ginori&quot;</td>
<td>32</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>&quot;Guadagni&quot;</td>
<td>8</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>&quot;Lamberteschi&quot;</td>
<td>42</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>&quot;Medici&quot;</td>
<td>103</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>&quot;Pazzi&quot;</td>
<td>48</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>&quot;Peruzzi&quot;</td>
<td>49</td>
<td>32</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>&quot;Pucci&quot;</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>&quot;Ridolfi&quot;</td>
<td>27</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>&quot;Salviati&quot;</td>
<td>10</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>&quot;Strozzi&quot;</td>
<td>146</td>
<td>29</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>&quot;Tornabuoni&quot;</td>
<td>48</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UndirectedEdges</th>
<th>source</th>
<th>target</th>
<th>marriage</th>
<th>business</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>1</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>
Pan: “grab” the background by holding left-click and moving your mouse.

Zoom: Using scroll wheel, press the “+” and “-” buttons in the upper-left hand corner, or right-click and move the mouse left or right. Center graph by selecting ‘View -> Center’.

Select to select/move single nodes. Hold down ‘Shift’ to select multiple.

Right click to modify Color, etc.
**Graph Modifier:**
Select all nodes in the Object drop-down menu and click Show Label button.

Select Resize Linear > Nodes > totalities drop-down menu, then type 5 and 20 into the From and To value box separately. Then select Do Resize Linear.

Select Colorize> Nodes>totalities, then select white and enter (204,0,51) in the pop-up color boxes on in the From and To buttons.

Select “Format Node Labels”, replace default text {originallabel} with your own label in the pop-up box Enter a formatting string for node labels.

**Interpreter:**
Uses Jython, a combination of Java and Python

Try: `colorize(wealth, white, green)`
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

42

Studying Four Major NetSci Researchers (ISI Data) using Database (section 5.1.4)

<table>
<thead>
<tr>
<th>Time frame:</th>
<th>1955-2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region(s):</td>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Topical Area(s):</td>
<td>Network Science</td>
</tr>
<tr>
<td>Analysis Type(s):</td>
<td>Paper Citation Network, Co-Author Network, Bibliographic Coupling Network, Document Co-Citation Network, Word Co-Occurrence Network</td>
</tr>
</tbody>
</table>

Thomson Reuter’s Web of Knowledge (WoS) is a leading citation database
Access it via the “Web of Science” tab at http://www.isiknowledge.com

(note: access to this database requires a paid subscription)

Along with Scopus, WoS provides some of the most comprehensive datasets for scientometric analysis
To find all publications by an author, search for the last name and the first initial followed by an asterisk in the author field

http://sci2.wiki.cns.in.edu/5.1.4+Studying+Four+Major+NetSci+Researchers+(ISI+Data)
Data Acquisition from Web of Science

In December 2007, we downloaded all papers by
- Eugene Garfield
- Stanley Wasserman
- Alessandro Vespignani
- Albert-László Barabási
from
- Science Citation Index Expanded (SCI-EXPANDED)--1955-present
- Social Sciences Citation Index (SSCI)--1956-present
- Arts & Humanities Citation Index (A&HCI)--1975-present

Comparison of Counts
No books and other non-WoS publications are covered.

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Age</th>
<th>Total # Cites</th>
<th>Total # Papers</th>
<th>H-Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eugene Garfield</td>
<td>82</td>
<td>1525</td>
<td>672</td>
<td>31</td>
</tr>
<tr>
<td>Stanley Wasserman</td>
<td>122</td>
<td>122</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Alessandro Vespignani</td>
<td>42</td>
<td>451</td>
<td>101</td>
<td>33</td>
</tr>
<tr>
<td>Albert-László Barabási</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dec 2007)</td>
<td>40</td>
<td>2218</td>
<td>126</td>
<td>47</td>
</tr>
<tr>
<td>(Dec 2008)</td>
<td>41</td>
<td>16920</td>
<td>159</td>
<td>52</td>
</tr>
<tr>
<td>(April 2011)</td>
<td>44</td>
<td>30102</td>
<td>201</td>
<td>68</td>
</tr>
</tbody>
</table>
Extract Co-Author Network


using \texttt{File > Load...}

A table of 361 unique records will be loaded into the Data Manager

Duplicates are removed and a log file is created

To extract the co-author network, select the \textit{361 Unique ISI Records} table and run \texttt{Data Preparation > Extract Co-Author Network}

using the ISI file format:

- The result is an undirected but weighted network of co-authors in the Data Manager.
- Run \texttt{Analysis > Networks > Network Analysis Toolkit (NAT)} to calculate basic properties: the network has 247 nodes and 891 edges.
- Use \texttt{Analysis > Networks > Unweighted and Undirected > Node Degree} to calculate the number of neighbors for each node independent of co-authorship weight.
- To view the complete network, select the \textit{Extracted Co-Authorship Network} and run \texttt{Visualization > Networks > GUESS}
- Network is loaded with random layout. In GUESS, run \texttt{Layout > GEM} and \texttt{Layout > Bin Pack} to improve layout. Run \texttt{Script > Run Script...} and select \texttt{yoursci2directory/scripts/GUESS/co-author-nw.py}
Use the GUESS Graph Modifier to change color and size coding.

Calculate node degrees in Sci²

An image editor can be used to add legends
Load... was selected. Found old-style ISI/Web Of Knowledge file. The original 361 records have been processed to remove duplicate unique ISI IDs leaving 361 records.

Extract Co-Author Network was selected. Input Parameters:
- File Format: isi

Network Analysis Toolkit (NAT) was selected.
- Nodes: 247
- Edges: 891

GUESS was selected.

Network Visualization:
Node Layout

Network Visualization:
Color/Size Coding by Data Attribute Values
Weak Component Clustering was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon

Input Parameters:
Number of top clusters: 10
3 clusters found, generating graphs for the top 3 clusters.

Node Degree was selected.
Documentation:
https://nwb.slis.indiana.edu/community/?n=AnalyzeData.NodeDegree

Network Visualization:
Giant Component

Network Visualization:
Color/Size Coding by Degree
Node Betweenness Centrality was selected.
Author(s): L. C. Freeman
Implementer(s): Santo Fortunato
Integrator(s): Santo Fortunato, Weixia Huang

Network Visualization:
Color/Size Coding by Betweenness Centrality

MST-Pathfinder Network Scaling was selected.
Input Parameters:
Weight Attribute measures: SIMILARITY
Edge Weight Attribute: weight

Network Visualization:
Reduced Network After Pathfinder Network Scaling
**Select Co-Author Network and run Blondel Community detection:**

![Network Visualization: Circular Hierarchy Visualization](image)

With parameter values

![Blondel Community Detection](image)

**Network Visualization: Circular Hierarchy Visualization**

Visualize resulting file using 'Visualization > Networks > Circular Hierarchy'

![Circular Hierarchy Visualization](image)

with parameter values
Network Visualization: Circular Hierarchy Visualization

Nodes that are interlinked/clustered are spatially close to minimize the number of edge crossings.

Node labels, e.g., author names.

Network structure using edge bundling.

Color coded cluster hierarchy according to Blondel community detection algorithm.

Note:
Header/footer info, legend, and more meaningful color coding are under development.

---

Paper-Citation Network Layout

➢ To extract the paper-citation network, select the 361 Unique ISI Records table and run Data Preparation > Extract Paper Citation Network
➢ The result is a unweighted, directed network of papers linked by citations, named Extracted paper-citation network in the Data Manager.
➢ Run NAT to calculate that the network has 5,342 nodes and 9,612 edges. There are 15 weakly connected components. (0 isolates)
➢ Run Analysis > Networks > Unweighted and Directed > Weak Component Clustering with parameters

![Weak Component Clustering](image)

Number of top clusters: 1

[OK | Cancel]

...to identify top-10 largest components. The largest (giant) component has 5,151 nodes.
General Network Extraction:
Weighted, Undirected Co-Occurrence Network

Author co-occurrence network

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Paper</td>
<td>Authors</td>
</tr>
<tr>
<td>2</td>
<td>P1</td>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>P2</td>
<td>A2:A6</td>
<td>P1</td>
</tr>
<tr>
<td>4</td>
<td>P3</td>
<td>A1:A3</td>
<td>P1:P2</td>
</tr>
<tr>
<td>5</td>
<td>P4</td>
<td>A1:A4:A5</td>
<td>P2</td>
</tr>
<tr>
<td>7</td>
<td>P6</td>
<td>A2:A6</td>
<td>P5</td>
</tr>
</tbody>
</table>

*Vertices 6
1 A1
2 A6
3 A2
4 A3
5 A5
6 A4

*Edges 6
2 3 2
1 4 1
1 5 1
5 6 1
1 6 1
2 5 1
General Network Extraction: Unweighted, Directed Bipartite Network

*Vertices 12
1 P1 bipartitetype "Paper"
2 A1 bipartitetype "Authors"
3 P2 bipartitetype "Paper"
4 A2 bipartitetype "Authors"
5 A6 bipartitetype "Authors"
6 P3 bipartitetype "Paper"
7 A3 bipartitetype "Authors"
8 P4 bipartitetype "Paper"
9 A4 bipartitetype "Authors"
10 A5 bipartitetype "Authors"
11 P5 bipartitetype "Paper"
12 P6 bipartitetype "Paper"

*Arcs
1 2
3 4
3 5
6 2
6 7
8 2
8 10
8 9
11 5
11 10
12 4
12 5

Paper-author bipartite (2-mode) network

General Network Extraction: Unweighted, Directed Paper-Citation Network

*Vertices 6
1 P1
2 P2
3 P3
4 P4
5 P5
6 P6

*Arcs
2 1
3 1
3 2
4 2
5 4
5 3
5 1
5 2
6 5
General Network Extraction: Unweighted, Directed Bi-Partite Network

*Vertices 11
1 P1 bipartitetype "Paper"
2 P2 bipartitetype "Paper"
3 P1 bipartitetype "References"
4 P3 bipartitetype "Paper"
5 P2 bipartitetype "References"
6 P4 bipartitetype "Paper"
7 P5 bipartitetype "Paper"
8 P4 bipartitetype "References"
9 P3 bipartitetype "References"
10 P6 bipartitetype "Paper"
11 P5 bipartitetype "References"

*Arcs
2 3
4 3
4 5
6 5
7 3
7 9
7 5
7 8
10 11

Mistake!

ISI Paper-Citation Network Extraction

Arcs from references to papers—in the direction of information flow
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn
Temporal bar graph for NSF projects

See 5.2.1 Funding Profiles of Three Universities (NSF Data)

Download NSF data

Visualize as Temporal Bar Graph

Area size equals numerical value, e.g., award amount.

Text

Start date  End date

NSF Awards Search via http://www.nsf.gov/awardsearch

Temporal Bar Graph of NSF projects
Download and load a dataset of your choice or load one of the sample data files, e.g.,
sampledata/scientometrics/nsf/Indiana.nsf
Run Visualization > Temporal > Temporal Bar Graph using
parameters:

Save visualized with Temporal Bar Graph as PS or EPS file
Convert into PDF and view
Zoom to see details in visualizations of large datasets, e.g., all NSF awards ever made

More NSF data workflows can be found in wiki tutorial:
5.1.3 Funding Profiles of Three Researchers at Indiana University (NSF Data)
5.2.1 Funding Profiles of Three Universities (NSF Data)
5.2.3 Biomedical Funding Profile of NSF (NSF Data)
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Geocoding and Geospatial Maps

Data with geographic identifiers
Geocode

Geolocated data
Aggregate (if necessary)

Geographic identifiers with data
Visualize

Region names + numeric data (Choropleth Map)

Geocoordinates + numeric data (Proportional Symbol Map)

http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder
Load File with Address and Times Cited Fields

Run File > Load...

Load the sample data table `sampledata/geo/usptoInfluenza.csv`

Let’s create a map showing influenza-related patent activity in the following countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Patents</th>
<th>Times Cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.16116</td>
<td>19.504959</td>
<td>0.0833333333</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>50.500992</td>
<td>4.47677</td>
<td>3.017857143</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>51.090839</td>
<td>10.4524</td>
<td>4.783333333</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>62.355739</td>
<td>-96.582092</td>
<td>5.392857143</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>59.461479</td>
<td>108.811779</td>
<td>0.266666667</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>47.69561</td>
<td>13.34577</td>
<td>4.2</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>52.108089</td>
<td>5.33033</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>46.813091</td>
<td>8.22414</td>
<td>0.507575758</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>23.599751</td>
<td>121.023811</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>-24.516201</td>
<td>133.39112</td>
<td>1.617857143</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>39.85</td>
<td>-98.38</td>
<td>75.9983889</td>
<td>220</td>
</tr>
<tr>
<td>12</td>
<td>46.712448</td>
<td>1.71832</td>
<td>2.20165501</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>-28.483219</td>
<td>24.679991</td>
<td>0.3333333333</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>37.487598</td>
<td>139.838287</td>
<td>15.99166667</td>
<td>39</td>
</tr>
<tr>
<td>15</td>
<td>31.389299</td>
<td>35.36124</td>
<td>3.5</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>54.313919</td>
<td>-2.23218</td>
<td>3.85</td>
<td>12</td>
</tr>
</tbody>
</table>

Yahoo Geocoder

Description

This algorithm converts place names or addresses into Latitude, Longitude co-ordinates. It accepts international addresses, countries, States of United States of America and ZIP codes of United States of America. All co-ordinates are obtained by querying Yahoo PlaceFinder service. Internet access must be available during geocoding.

Pros & Cons

1. The performance is slower than the Geocoder and may vary due to the network latency since the queries are requested through internet service. The benchmark test geocoded 470 unique locations per minute
2. Yahoo Geocoder supports address geocoding with international coverage which is not supported by Geocoder
3. To use Yahoo Geocoder, user has to obtain an application id through Yahoo registration. Save your application id and provide it when requested by the Yahoo Geocoder. Since each application id is allowed to geocode 60,000 locations per 24 hours, the user is encouraged to test on a small set of data first.

Applications

The plugin is useful for scientists who would like to visualize their data on a geographical map (geomap). User can obtain the geographical coordinates (Latitude and Longitude values) and feed them to the visualization plugin.

http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder
Run Analysis > Geospatial > Yahoo Geocoder

You can leave Application ID blank for trial purposes, but before heavy use, register later for your own personal Yahoo!
Application ID, see: http://developer.yahoo.com/geo/placefinder/

Aggregate by Country

Aggregate Data was selected.
Implementer(s): Chintan Tank
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data
Input Parameters:
Aggregate on column: Country
Delimiter for Country: |
Longitude: AVERAGE
Latitude: AVERAGE
Times Cited: SUM
Aggregated by "": All rows of Latitude column were skipped due to no non-null, non-empty values.
Aggregated by "": All rows of Longitude column were skipped due to no non-null, non-empty values.
Frequency of unique "Country" values added to "Count" column.
Right-click and **Save** map

Open this PostScript file to visualize

---

**Interpreting the Choropleth Map**

Header shows visualization type, data description, and creation date

Legend shows how data matches up with visual representation

---
Right-click and Save map

Open this PostScript file to visualize

Interpreting the Proportional Symbol Map

Legend shows how data matches up with visual representation

Header shows visualization type, data description, and creation date
For more information on creating geospatial visualizations, see Sci² Manual Section 5.2.4.1.

http://wiki.cns.iu.edu/display/SCI2TUTORIAL/5.2.4+Mapping+Scientometrics+%28ISI+Data%29

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn
Sci² Demo II:
Evolving collaboration networks

Load ISI formatted file

As CSV, file looks like:

Visualize each time slice separately:
Slice Table by Time

“Slice Into” allows the user to slice the table by days, weeks, months, quarters, years, decades, and centuries. There are two additional parameters for time slicing: cumulative and align with calendar. The former produces tables containing all data from the beginning to the end of each table’s time interval, which can be seen in the Data Manager and below.

Choosing “Years” under “Slice Into” creates multiple tables beginning from January 1st of the first year. If “Months” is chosen, it will start from the first day of the earliest month in the chosen time interval.
To see the evolution of Vespignani's co-authorship network over time, check 'cumulative'.

2. Extract co-authorship networks one at a time for each sliced time table using 'Data Preparation > Extract Co-Author Network', making sure to select "ISI" from the pop-up window during the extraction.

3. To view each of the Co-Authorship Networks over time using the same graph layout, begin by clicking on longest slice network (the 'Extracted Co-Authorship Network' under 'slice from beginning of 1990 to end of 2006 (101 records)') in the data manager. Visualize it in GUESS using 'Visualization > Networks > GUESS'.

4. From here, run 'Layout > GEM' followed by 'Layout > Bin Pack'. Run 'Script > Run Script …' and select 'yoursci2directory/scripts/GUESS/co-author-nw.py'.

5. In order to save the x, y coordinates of each node and to apply them to the other time slices in GUESS, select 'File > Export Node Positions' and save the result as 'yoursci2directory/NodePositions.csv'. Load the remaining three networks in GUESS using the steps described above and for each network visualization, run 'File > Import Node Positions' and open 'yoursci2directory/NodePositions.csv'.

6. To match the resulting networks stylistically with the original visualization, run 'Script > Run Script …' and select 'yoursci2directory/scripts/GUESS/co-author-nw.py', followed by 'Layout > Bin Pack', for each.
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
    - Sci² tool visualizations
      - Bipartite networks
      - Map of Science
- Outlook and Q&A
- Adjourn
R-Bridge

- Run ‘R > Create an R Instance’. You must set the ‘R Executable Directory’ parameter to be the path to the directory on your computer that contains Rgui.exe. Results in an ‘R Instance’ object in the Data Manager.
- To send a table from the data manager to an R Instance object, select the table and the R Instance object together then run ‘R > Send Table to R’. Select ‘R > Run Rgui’ and the table is available in the R environment using the variable name you specified as a parameter to the Import algorithm.
- To pull back data from an R Instance object to the Data Manager, select the R Instance object and run ‘R > Get Table From R’. Choose the name of the variable from the dropdown list.

Gephi Bridge

- Gephi must already be installed on the system for this bridge plugin to work.
- Select any network file, then choose Visualization > Networks > Gephi
- The selected network file will be opened in Gephi using the Import report
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
- Sci² tool visualizations
  - Bipartite networks
  - Map of Science
- Outlook and Q&A
- Adjourn

Map of Science

The Map of Science is a visual representation of 554 sub-disciplines within 13 disciplines of science and their relationships to one another.

- A set of journals can be overlaid onto this base map, showing the areas of science encompassed
- Each circle size represents the (weighted) proportion of the journal set mapped to that sub-discipline
- Detailed breakdown of each discipline
A bipartite network graph can be visualized using this plugin.

- Supports edge and node weights
- Node ordering can be sorted independently

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci\(^2\) Tool Basics
- Sci\(^2\) Tool – Advanced Topics
- Outlook and Q&A
- Adjourn
Scholarly Database: 25 million scholarly records
http://sdb.slis.indiana.edu

VIVO Research Networking
http://vivoweb.org

Information Visualization Cyberinfrastructure
http://iv.cns.iu.edu

Network Workbench Tool & Community Wiki
http://nwb.cns.iu.edu

Science of Science (Sci²) Tool
http://sci2.cns.iu.edu

Epidemics Tool & Marketplace
Forthcoming

Scholarly Database at Indiana University
http://sdb.wiki.cns.iu.edu

Supports federated search of 25 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

Register for free access at http://sdb.cns.iu.edu
Download automatically extracted networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent-citation

Download tables suitable for burst analysis
VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.

Funded by $12 million NIH award.

**Temporal Analysis (When)** Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level.

**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)
**Network Analysis (With Whom?)**  Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?

**Geospatial Analysis (Where)**  Where is what science performed by whom? Science is global and needs to be studied globally.
Download Data

General Statistics
- 36 publication(s) from 2001 to 2010 (.CSV File)
- 80 co-author(s) from 2001 to 2010 (.CSV File)

Co-Author Network
(GraphML File)

Save as Image (.PNG file)

Tables
- Publications per year (.CSV File)
- Co-authors (.CSV File)

Computational Epidemics
Forecasting (and preventing the effects of) the next pandemic.


TEXTrend adds WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at


Need Help? Ask an Expert!

https://sci2.cns.iu.edu/user/ask.php
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Thank you!

Q & A

Please complete the Post-Tutorial Questionnaire so that we can further improve these tutorials.

***

Bug reports and all comments are welcome.
All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience