
Video and paper are at: http://www.scivee.tv/node/27704

Taming Complexity
TTI VANGUARD
Conference in D.C.
2011.10.04

Find your way
Find collaborators, friends
Identify trends

Take terra bytes of data
Plug-and-Play Macroscopes
Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets 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 make things larger or smaller, macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.

While microscopes and telescopes are physical instruments, macroscopes resemble continuously changing bundles of software plug-ins. Macroscopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.
Plug-and-Play Macroscopes

Different datasets/formats. Diverse algorithms/tools written in many programming languages.

Finance
Smart Cities
Epidemiology
Crime
Health

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, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.
The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

The tool provides more 160 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

It has been downloaded more than 110,000 times.

Computational Proteomics
What relationships exist between protein targets of all drugs and all disease-gene products in the human protein–protein interaction network?


Computational Economics
Does the type of product that a country exports matter for subsequent economic performance?

Computational Social Science
Studying large scale social networks such as Wikipedia

Second Sight: An Emergent Mosaic of Wikipedian Activity,
The NewScientist, May 19, 2007

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


Sci2 Tool v0.5.1 Alpha (May 4th, 2011)
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
Indiana University and SciTech Strategies,
http://sci2.cns.iu.edu

<table>
<thead>
<tr>
<th>Type of Analysis</th>
<th>Level of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Micro/Individual</strong> (1-100 records)</td>
<td><strong>Meso/Local</strong> (101–10,000 records)</td>
</tr>
<tr>
<td>Statistical Analysis/Profiling</td>
<td>Individual person and their expertise profiles</td>
</tr>
<tr>
<td>Temporal Analysis (When)</td>
<td>Funding portfolio of one individual</td>
</tr>
<tr>
<td>Geospatial Analysis (Where)</td>
<td>Career trajectory of one individual</td>
</tr>
<tr>
<td>Topical Analysis (What)</td>
<td></td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF Co-PI network of one individual</td>
</tr>
</tbody>
</table>
Open Code for Replicable S&T Assessment

OSGi/CIShell powered tool, see [http://cishell.org](http://cishell.org)  

**Network Extraction: Examples**

**Author co-occurrence network**

**Paper-author 2-mode network**

**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.
Topic Mapping: UCSD Science Map

Data: WoS and Scopus for 2001–2005, 7.2 million papers, more than 16,000 separate journals, proceedings, and series

Similarity Metric: Combination of bibliographic coupling and keyword vectors

Number of Disciplines: 554 journal clusters further aggregated into 13 main scientific disciplines that are labeled and color coded in a metaphorical way, e.g., Medicine is blood red and Earth Sciences are brown as soil.

Geospatial Maps with Congressional Districts

Identify Congressional District, Latitude, Longitude

Aggregate/Count identical Congressional Districts

[Table and map showing congressional districts with coordinates and counts]
Evolving Collaboration Networks

Load isi formatted file

As csv, file looks like:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>Authors</td>
<td>Authors (Full Names)</td>
<td>Beginning</td>
<td>Book Title</td>
<td>Book Series</td>
</tr>
<tr>
<td>2</td>
<td>The systematic study of Colizza, V, El Barrat, A, Barthelemy, M, Vespignani, A</td>
<td>2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Uncovering the hidden role of Colizza, V, Flammini, A, Barthelemy, M, Vespignani, A</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Computer viruses can set Vespignani, A</td>
<td>135</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Mapping the Internet ge Field, A, Alvarez-Hamelin, J, Barrat, A, Vazquez, A, Vespignani, A</td>
<td>140</td>
<td>LEARNING SYSTEMS IN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Visualize each time slide separately:

Sci2 Tool Adoption

The Sci2 Tool is used by the National Science Foundation, the National Institutes of Health, and the US Department of Agriculture.

“As a new user, I am beginning with very little knowledge of the analyses and modeling techniques that Sci2 enables. I’ve been able to use my own dataset and follow through some of the workflows to the point of generating the first network and time horizon visuals. That was so exciting I stayed up far past bedtime to get to the visuals.”

Dr. Suzanne A. Pierce, Center for International Energy and Environmental Policy

Jackson School of Geosciences, The University of Texas at Austin
TEXTrend adds WEKA, UIMA, Wordij, CFinder, and more.
See the latest versions of TEXTrend Toolkit modules at [http://textrend.org](http://textrend.org)
OSGi/CIShell Adoption

A number of other projects recently adopted OSGi and/or CIShell:

- **Cytoscape** [http://cytoscape.org](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](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](http://taverna.org.uk) Developed by the myGrid team 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.

- **TEXTrend** [http://textrend.org](http://textrend.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 corpuses with an inherently temporal component.

- **DynaNets** [http://www.dynanets.org](http://www.dynanets.org) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.

- **SISOB** [http://sisob.ijc.unam.es](http://sisob.ijc.unam.es) An Observatory for Science in Society Based in Social Models. 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.
Maps created using Sci2 are travelling in the “Expedition Zukunft” science train visiting 62 cities in 7 months, 12 coaches, 300 m long. [http://www.expedition-zukunft.de](http://www.expedition-zukunft.de)

And they are part of the international Mapping Science exhibit: [http://scimaps.org](http://scimaps.org)
Taxonomy Visualization of Patent Data.

Bruce W. Herr, Russell Duhon, Elisha F. Hardy, Shashikant Penumarthy, and Katy Börner (2007)  
113 Years of Physical Review.
Examining the Evolution & Distribution of Patent Classifications


A Topic Map of NIH Grants 2007


This is the only mockup in this slide show. Everything else is available today.

References


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