

Sci2 Sample Workflows

1

Data Formats

GraphML (*.xml or *.graphml)
 XGMML (*.xml)
 Pajek .NET (*.net)
 Pajek .Matrix (*.mat)
 NWB (*.nwb)
 TreeML (*.xml)
 Edgelist (*.edge)
 Scopus csv (*.scopus)
 NSF csv (*.nsf)
 CSV (*.csv)
 ISI (*.isi)
 Bibtex (*.bib)
 Endnote (*.enw)

File

Load ...
 Load and Clean ISI File
 Read Directory Hierarchy
 Save ...
 View ...
 View with ...
 Merge Node and Edge Files
 Split Graph to Node and Edge Files
 Preferences
 Converter Graph
 Exit

2

Data Preparation

Database

ISI

Merge Identical ISI People
 Suggest ISI People Merges
 Merge Document Sources
 Create Document Source Merging Table
 Match References to Papers

Extract Authors
 Extract Documents
 Extract Keywords
 Extract Document Sources

Extract Authors by Year
 Extract References by Year
 Extract Original Author Keywords by Year
 Extract New ISI Keywords by Year

Extract Authors by Year for Burst Detection
 Extract Documents by Year for Burst Detection
 Extract Original Author Keywords by Year for Burst Detection
 Extract New ISI Keywords by Year for Burst Detection
 Extract References by Year for Burst Detection

Extract Longitudinal Summary

Extract Co-Author Network

Extract Author Citation Network
 Extract Document Citation Network (Core Only)
 Extract Document Citation Network (Core and References)
 Extract Document Source Citation Network (Core Only)
 Extract Document Source Citation Network (Core and References)

Extract Document Co-Citation Network (Core Only)
 Extract Document Co-Citation Network (Core and References)
 Extract Document Source Co-Citation Network (Core Only)
 Extract Document Source Co-Citation Network (Core and References)
 Extract Author Co-Citation Network

Extract Author Bibliographic Coupling Network
 Extract Document Bibliographic Coupling Network
 Extract Document Source Bibliographic Coupling

NSF

Merge Identical NSF People

Extract Investigators
 Extract Awards
 Extract Organizations

Extract Co-PI Network

General

Create Merging Tables
 Merge Entities

Custom Table Query
 Custom Graph Query

Extract Raw Tables from Database

Text Files

Remove ISI Duplicate Records
 Remove Rows with Multitudinous Fields

Extract Directed Network
 Extract Bipartite Network
 Extract Paper Citation Network
 Extract Author Paper Network

Extract Co-Occurrence Network
 Extract Word Co-Occurrence Network
 Extract Co-Author Network
 Extract Reference Co-Occurrence (Bibliographic Coupling) Network

Extract Document Co-Citation Network

Detect Duplicate Nodes
 Update Network by Merging Nodes

3

Preprocessing

General

Extract Top N% Records
 Extract Top N Records
 Aggregate Data

Temporal

Slice Table by Time

Geospatial

Extract ZIP Code

Topical

Lowercase, Tokenize,
 Stem, and Stopword Text

Networks

Extract Top Nodes
 Extract Nodes Above or Below Value
 Delete Isolates
 Extract Top Edges
 Extract Edges Above or Below Value
 Remove Self Loops
 Trim by Degree
 MST-Pathfinder Network Scaling
 Fast Pathfinder Network Scaling

Snowball Sampling (N nodes)
 Node Sampling
 Edge Sampling

Symmetrize
 Dichotomize
 Multipartite Joining

Merge 2 Networks

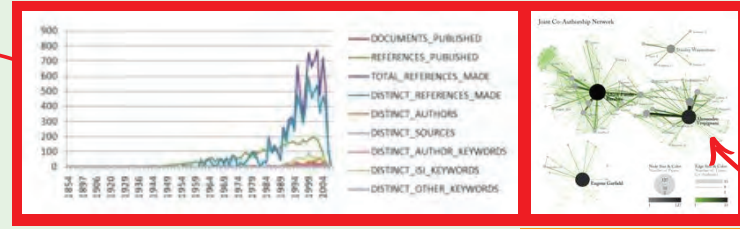
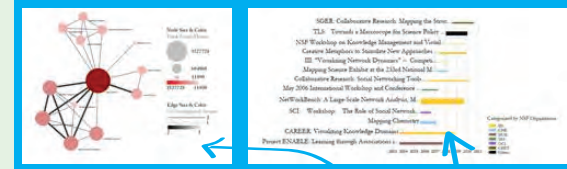
5

Modeling

Networks

Random Graph
 Watts-Strogatz Small World
 Barabási-Albert Scale-Free

TARL (Topics, Aging and Recursive Linking)



Micro

Individual Level Studies

NSF awards for one scholar are analyzed to understand temporal patterns and co-investigator network.

Meso

Institution Level Studies

ISI publication data from four network science researchers are studied to understand temporal patterns and co-author networks.

Macro

Global Level Studies

USPTO patents mentioning "influenza" downloaded from the Scholarly Database is geolocated by country of patent holder. Number of patents and patent citations are shown.

4

Analysis

Temporal

Burst Detection

Geospatial

Geocoder
 Yahoo Geocoder
 Congressional District Geocoder

Topical

Burst Detection

Networks

Network Analysis Toolkit (NAT)
 Unweighted & Undirected
 Node Degree
 Degree Distribution
 K-Nearest Neighbor (Java)
 Watts-Strogatz Clustering Coefficient
 Watts Strogatz Clustering Coefficient over K

Diameter
 Average Shortest Path
 Shortest Path Distribution
 Node Betweenness Centrality

Weak Component Clustering
 Global Connected Components

Extract K-Core
 Annotate K-Core-ness

HITS

Blondel Community Detection

Weighted & Undirected
 Clustering Coefficient
 Nearest Neighbor Degree

Strength vs Degree
 Degree & Strength
 Average Weight vs End-point Degree
 Strength Distribution
 Weight Distribution
 Randomize Weights

Node Betweenness Centrality

Blondel Community Detection

HITS

Unweighted & Directed

Node In-degree

Node Out-degree

In-degree Distribution

Out-degree Distribution

K-Nearest Neighbor

Single Node In-Out Degree Correlations

Dyad Reciprocity

Arc Reciprocity

Adjacency Transitivity

Weak Component Clustering

Strong Component Clustering

Extract K-Core

Annotate K-Core-ness

Node Betweenness Centrality

HITS

PageRank

Weighted & Directed

HITS

Weighted PageRank

6

Visualization

General

GnuPlot
 Image Viewer

Temporal

Horizontal Bar Graph

Geospatial

Geo Map (Circle Annotations)
 Geo Map (Colored-Region Annotations)

Topical

RefMapper

Science Map via 554 Fields
 Science Map via Journals

Networks

GUESS

Radial Tree/Graph (prefuse alpha)
 Radial Tree/Graph with Annotation (prefuse beta)
 Tree View (prefuse beta)
 Tree Map (prefuse beta)
 Force Directed with Annotation (prefuse beta)
 Fruchterman-Reingold with Annotation (prefuse beta)

DrL (VxOrd)
 Specified (prefuse beta)

Circular Hierarchy

Output formats

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

Sci²

A tool for science of science research & practice • <http://sci2.cns.iu.edu>

SCIENCE OF SCIENCE (SCI2) TOOL

Project Investigators: Katy Börner and Kevin W. Boyack (SciTech Strategies Inc.)

Programmers: Joseph R. Biberstine, Russell J. Duhon, Chin Hua Kong, Micah W. Linnemeier, Patrick A. Phillips, Thomas G. Smith, Chintan Tank

Users, Testers & Tutorial Writers: Scott Weingart, Hanning Guo, Katy Börner

Cyberinfrastructure for Network Science Center
School of Library and Information Science
Indiana University, Bloomington, IN
<http://cns.iu.edu>

Introduction

The Science of Science (Sci2) Tool is a modular toolset designed specifically for the study of science. It supports the temporal, geospatial, topical, and network analysis and visualization of datasets at the micro (individual), meso (local), and macro (global) levels. The web site at <http://sci2.cns.iu.edu> supports free registration, an online tutorial with workflow examples, and personalized support via the Ask an Expert feature.

Functionality

- Read datasets in many different formats including CSV, ISI, NSF
- Perform extensive data preprocessing: data cleaning, deduplication, filtering, network extraction
- Use advanced analysis algorithms
- Visualizations to interactively explore, understand, and communicate results. Many visualizations use an easy to read reference system, automatic legend design, and 'fine print' text on who created the visualization, when, and using what data source.
- Audit trail documentation to automatically record workflows
- New algorithm plugins can be easily installed by anyone

As of October 2011, more than 150 different preprocessing, analysis, modeling, and visualization algorithms are available, see listing on back.

System Architecture

The Sci2 Tool is built on the Cyberinfrastructure Shell (CIShell) (CIShell.org), an open source software framework for the easy integration and utilization of datasets, algorithms, tools, and computing resources. CIShell is based on the Equinox implementation of the OSGi R4 Specification (OSGi.org).

Four other tools:

- Network Workbench (<http://nwb.cns.iu.edu>)
- DynaNets (<http://www.dynanets.org>)
- TEXTrend (<http://www.textrend.org>)
- Epidemiology Tool (EpiC) (<http://epic.cns.iu.edu>)

are CIShell powered and plugins from these efforts/tools can be plug-and-played in the Sci2 Tool. Simply copy a *.jar file from the /plugin directory of one tool into the /plugin directory of another tool and the algorithm or tool becomes available in the menu system. As the functionality of OSGi/CIShell-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

Documentation & Tutorials

The tool comes with extensive documentation. A comprehensive online user manual shows how to use the tool for science of science research and science policy and there exist 12 two-hour tutorials originally designed for the National Institutes of Health.

Scott Weingart, Hanning Guo, Katy Börner, Kevin W. Boyack, Micah W. Linnemeier, Russell J. Duhon, Patrick A. Phillips, Chintan Tank, Chin Hua Kong, Thomas Smith, and Joseph Biberstine (2011) [Science of Science \(Sci2\) Tool User Manual](#). Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.

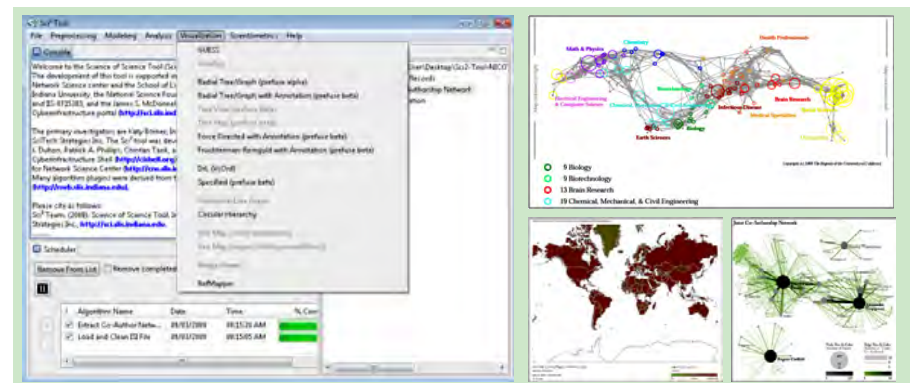
Katy Börner (2011) [Sci2: A Tool of Science of Science Research and Practice](#). 13th International Society for Scientometrics and Informetrics (ISSI) Conference, Durban, South Africa.

Katy Börner and Chin Hua Kong (2011) [Social Network Analysis Workshop CIShell Powered Tools: Network Workbench \(NWB\) & Science of Science \(Sci2\) Tool](#). Second Annual International Science of Team Science Conference, Chicago, IL.

Katy Börner and Angela Zoss (2010) [Plug-and-Play Macroscopes Tutorial](#). International Conference on Social Computing, Behavioral Modeling and Prediction, Bethesda, MD.

Agency Adoption

The tool is actively used in peer-reviewed research and agencies including the National Science Foundation, the National Institutes of Health, and the US Department of Agriculture.



Please cite as

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

Acknowledgments

This work is supported in part by the Cyberinfrastructure for Network Science Center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0513650, and the James S. McDonnell Foundation.