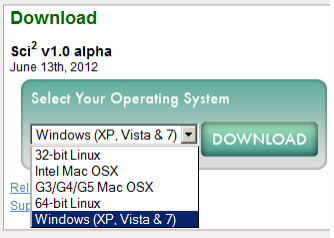
"Sci2: A Tool for Science of Science Research and Practice" Workshop at ARL

Michael Ginda
Data Analyst
Indiana University, Bloomington, Indiana, USA
mginda@Indiana.edu
http://cns.iu.edu

Adelphi, Maryland
Wednesday, June 17th, 2014 • 8:30am-3pm &
Thursday, June 18, 2014 • 8:30am-3pm

Software, Datasets, Plugins, and Documentation

- These slides
  http://ivl.slis.indiana.edu/km/pres/2014-ginda-sci2tutorial-arl.pdf
- Sci2 Tool Manual v0.5.1 Alpha, updated to match v1.0 Alpha tool release
  http://sci2.wiki.cns.iu.edu
- Sci2 Tool v1.0 Alpha (June 13, 2012)
  http://sci2.cns.iu.edu



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

- Make sure you have Java 1.6 (32-bit suffices) or higher installed or download from http://www.java.com/en/download. To check your Java version, open a terminal and run 'java -version'.
- Some visualizations are saved as Postscript files. A free Postscript to PDF viewer is at http://ps2pdf.com and a free PDF Viewer at http://www.adobe.com/products/reader.html.

CNS Cyberinfrastructure for Network Science Center

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- **Sci2 development and workflow design**
- **Geospatial Analysis: US and world maps**
- **Temporal Analysis: Horizontal line graph of NSF projects**
- **Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records**

12:00-12:45pm Break for Lunch

- **Topical Analysis: Visualize research profiles**
- **Network Analysis: Visualizing the Florentine Network**
- **Network Analysis: Word Co-occurrence Networks**

Q&A

3:00p Adjourn

3

CNS Cyberinfrastructure for Network Science Center

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- **Sci2 development and workflow design**
- Geospatial Analysis: US and world maps
- Temporal Analysis: Horizontal line graph of NSF projects
- Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records

12:00-12:45pm Break for Lunch

- Topical Analysis: Visualize research profiles
- Network Analysis: Visualizing the Florentine Network
- Network Analysis: Word Co-occurrence Networks

Q&A

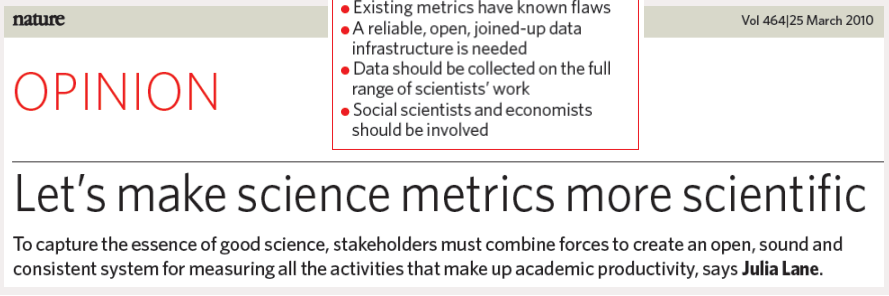
3:00p Adjourn

4

Science of Science (Sci2) Tool

http://sci2.cns.iu.edu

- Explicitly designed for SoS research and practice, well documented, easy to use.
- Empowers many to run common studies while making it easy for exports to perform novel research.
- Advanced algorithms, effective visualizations, and many (standard) workflows.
- Supports micro-level documentation and replication of studies.
- Is open source—anybody can review and extend the code, or use it for commercial purposes.



nature

Vol 464|25 March 2010

OPINION

SUMMARY

- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

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**.

5

Sci2 Tool v0.5.2 Alpha (Dec 19, 2011)

New Features

- Support new Web of Science format from ISI
- Support network overlay for geographical map
- Support Prefuse's visualizations on Macs OS

Improvements

- Improve memory usage and processing time of Extract top N nodes and Extract top N Edges algorithms
- Unify merging algorithms used by database

Bug fixes

- Fix legend boundary issue in geographical map
- Fix typo error on the output data label
- Fix slice by year algorithm

6

Sci2 Tool v1.0 Alpha (August 2013)

Major Release
featuring a Web services compatible CIShell v2.0 (http://cishell.org)

New Features
- Google Scholar citation reader
- New visualizations such as
  - geospatial maps
  - science maps
  - bi-modal network layout
- R statistical tool bridging
- Gephi visualization tool bridging
- Comprehensive online documentation

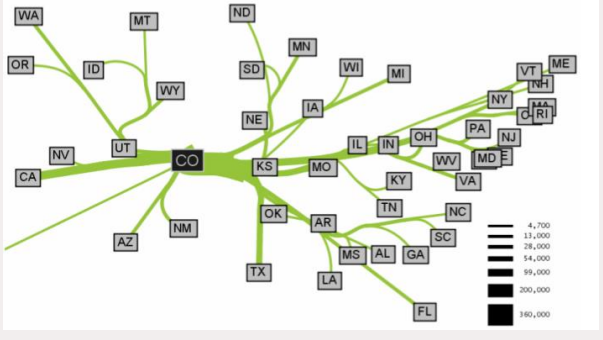
Release Note Details
http://wiki.cns.iu.edu/display/SCI2TUTORIAL/4.4+Sci2+Release+Notes+v1.0+alpha

7

Sci2 Tool v1.1 Alpha (August 2013)

New Features
- Twitter, Facebook, and Flickr readers
- Bing Geocoder
- Flow map visualization, see below
- Comprehensive online documentation

Bug fixes



8

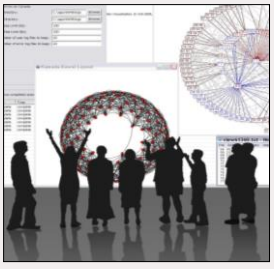
Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of the massive amounts of data that result from complex systems.

Rather than making things larger or smaller, macroscopes let us observe what is too great, slow, or complex for us to comprehend or sometimes even notice.

Microscopes

Telescopes



Macroscopes

Macroscopes (cont.)

Plug-and-Play Macroscopes

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

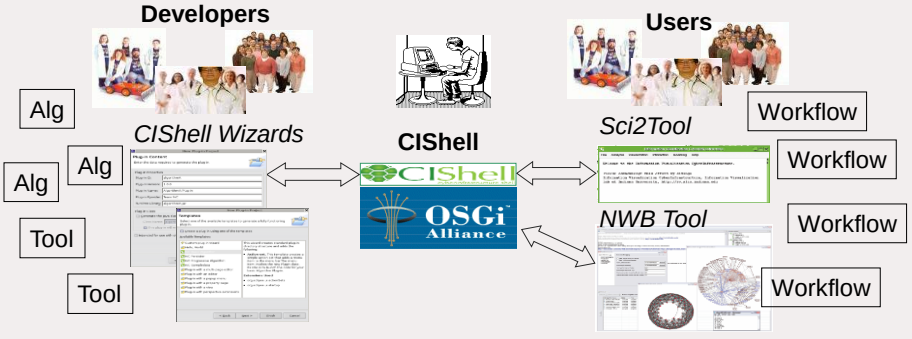
Macroscopes make it easy to

- Simply drop plugins into the tool and 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
- CIShell provides "sockets" into which algorithms, tools, and datasets can be plugged using a wizard-driven process

OSGi & Cyberinfrastructure Shell (CIShell)



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

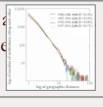
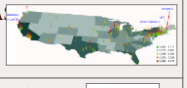
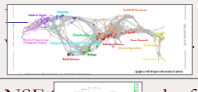
- JPEG (*.jpg)
- PDF (*.pdf)
- PostScript (*.ps)

Formats are documented at http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/2.3+Data+Formats.

13

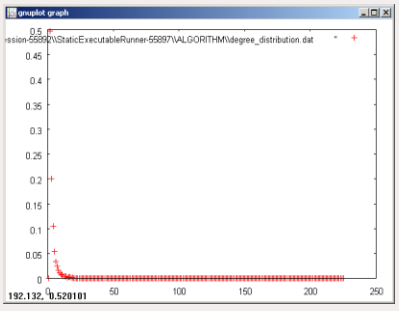
Types and Levels of Analysis

Types and Levels of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, a all of science 
Temporal Analysis (When)	Funding portfolio of one individual	 opic bursts of PNAS	113 Years of P Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a sta intellectual lan 	PNA 
Topical Analysis (What)		 flows in research	VxOrd/Topic NIH funding 
Network Analysis (With Whom?)	NSF ork of one in 	 netw 	NIH's core c  cy

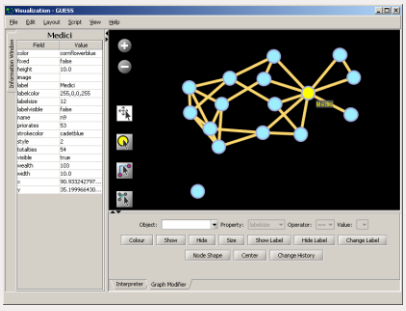
14

Sci² Tool – Supported Tools



Gnuplot

portable command-line driven interactive data and function plotting utility http://www.gnuplot.info/.



GUESS

exploratory data analysis and visualization tool for graphs and networks.

https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS.

Sci² Tool – Supported Tools

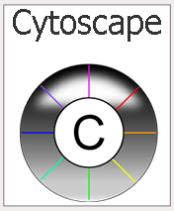
Adding more layout algorithms and network visualization interactivity

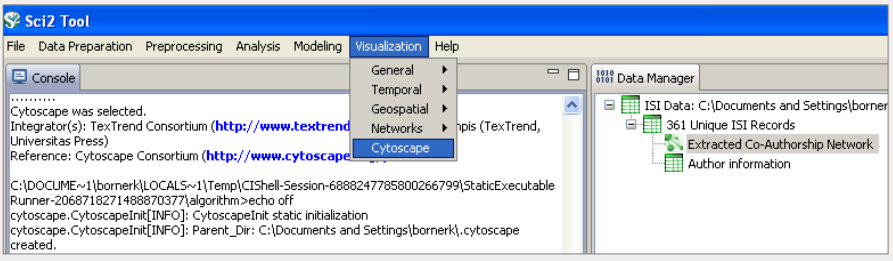
via Cytoscape http://www.cytoscape.org.

Simply add *org.textrend.visualization.cytoscape_0.0.3.jar* into your /plugin directory.

Restart Sci² Tool

Cytoscape now shows in the Visualization Menu

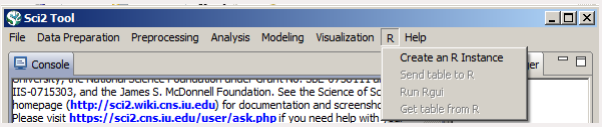




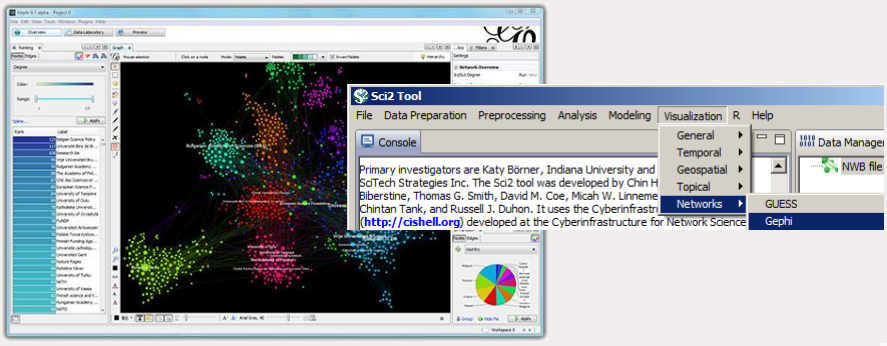
Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.

Sci² Tool – Bridged Tools

R statistical tool bridging



Gephi visualization tool bridging



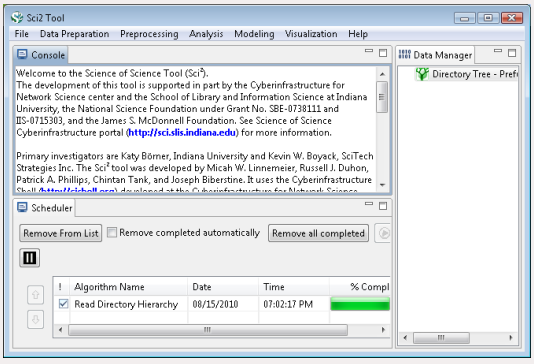
17

Sci2 Tool Interface Components

See also 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/…). If errors occur, they are saved in a 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

18

Questions?

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- Sci2 development and workflow design
- **Geospatial Analysis: US and world maps**
- Temporal Analysis: Horizontal line graph of NSF projects
- Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records

12:00-12:45pm Break for Lunch

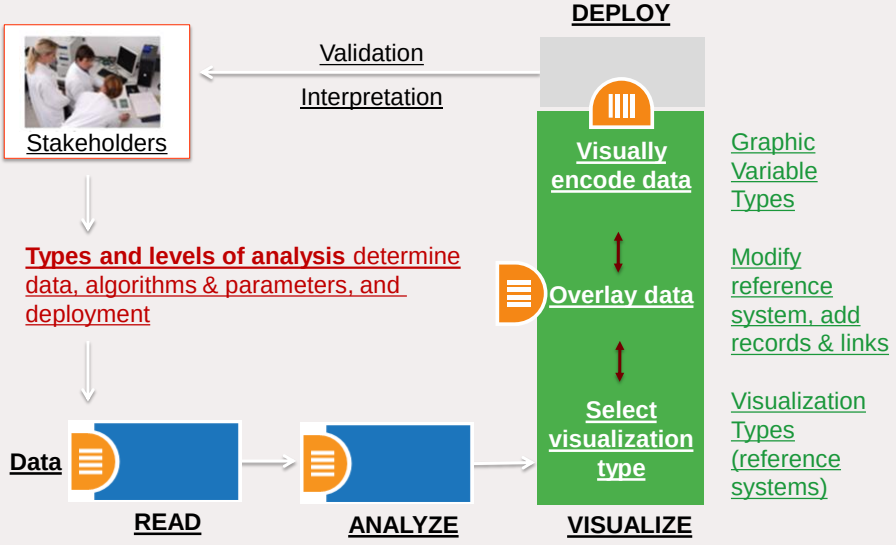
- Topical Analysis: Visualize research profiles
- Network Analysis: Visualizing the Florentine Network
- Network Analysis: Word Co-occurrence Networks

Q&A

3:00p Adjourn

CNS Cyberinfrastructure for Network Science Center

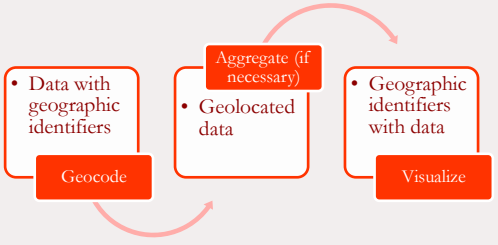
Needs-Driven Workflow Design using a modular data acquisition/analysis/ modeling/ visualization pipeline as well as modular visualization layers.

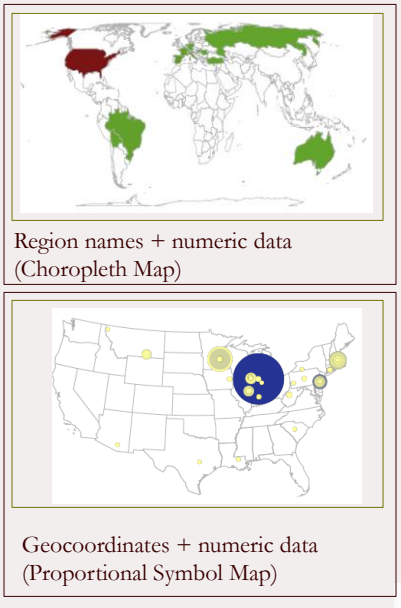


CNS Cyberinfrastructure for Network Science Center

Geocoding and Geospatial Maps

http://wiki.cns.iu.edu/display/CISHELL/Bing+Geocoder





Region names + numeric data (Choropleth Map)

Geocoordinates + numeric data (Proportional Symbol Map)

CIShell cyberinfrastructure shell **Bing Geocoder**

Edit Share Add Tools

Added by Mayur Masrani, last edited by Mayur Masrani on Apr 24, 2013

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 Bing geocoder service. Internet access must be available during geocoding.

Pros & Cons

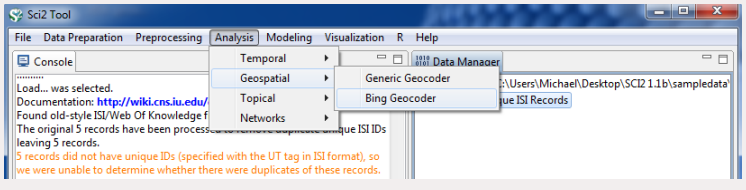
1. The performance is slower than the Geocoder and may vary due to the network latency since the queries are requested through internet service.
2. Bing Geocoder supports address geocoding with international coverage which is not supported by Geocoder.
3. To use Bing Geocoder, user has to obtain an API Keys from Bing Maps. Save your api keys and provide it when requested by the Bing Geocoder. Since each api key is allowed to geocode 50,000 locations per 24 hours, the user is encouraged to test on a small set of data first.

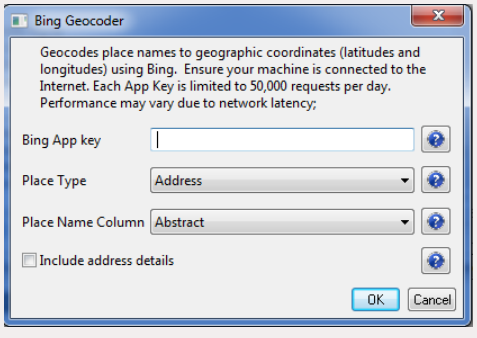
Applications

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

http://wiki.cns.iu.edu/display/CISHELL/Bing+Geocoder

Using Bing Geocoder



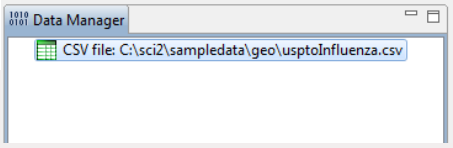


You can leave Application ID blank for trial purposes, but for heavy use, register for your own personal Bing Application ID, see: https://www.bingmapsportal.com

Load File with Address and Times Cited Fields

Run '*File* > *Load…*' and select the sample data table '*sampledata/geo/usptoInfluenza.csv*'
Create a map of influenza patents held by different countries.

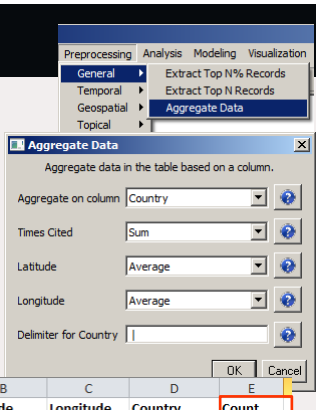
	A	B	C	D	E
1	Country	Latitude	Longitude	Patents	Times Cited
2	Hungary	47.16116	19.504959	0.083333333	4
3	Belgium	50.500992	4.47677	3.017857143	11
4	Germany	51.090839	10.45424	4.783333333	4
5	Canada	62.35873	-96.582092	5.539285714	21
6	Russia	59.461479	108.831779	0.266666667	2
7	Austria	47.69651	13.34577	4.2	17
8	Netherlands	52.108089	5.33033	1	2
9	Switzerland	46.813091	8.22414	0.507575758	6
10	Taiwan	23.599751	121.023811	2	3
11	Australia	-24.916201	133.393112	1.617857143	23
12	United States	39.83	-98.58	73.9983889	220
13	France	46.712448	1.71832	2.201165501	9
14	South Africa	-28.483219	24.676991	0.333333333	1
15	Japan	37.487598	139.838287	15.99166667	39
16	Israel	31.389299	35.36124	3.5	3
17	United Kingdom	54.313919	-2.23218	3.85	12



25

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.

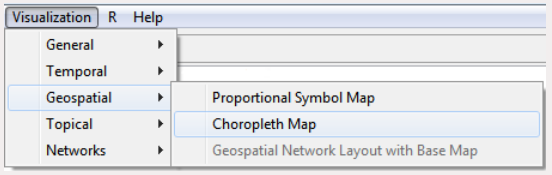


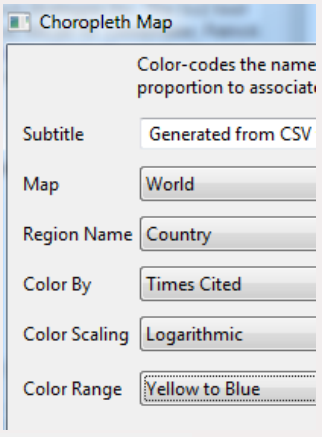
	A	B	C	D
1	Times Cited	Latitude	Longitude	Country
2	7	42.02946091	-87.68838501	United States
3	0			
4	0			
5	2	42.34999466	-71.08765411	United States
6	14	41.70074844	-86.23918915	United States
7	15	41.70074844	-86.23918915	United States
8	29	41.89422607	-87.61901855	United States
9	32	41.70074844	-86.23918915	United States
10	7	41.70074844	-86.23918915	United States
11	5	41.70074844	-86.23918915	United States
12	2	41.11500168	-85.73377991	United States
13	10	47.50622177	19.06481934	Hungary
14	44	41.70074844	-86.23918915	United States
15	0	47.50622559	19.06481934	Hungary
16	19	41.70074844	-86.23918915	United States

	A	B	C	D	E
1	Times Cited	Latitude	Longitude	Country	Count
2	14680	[41.10645f]	[-82.45309f]	United States	194
3	1802				57
4	398	[47.506226f]	[19.06482f]	Hungary	14
5	101	[37.25198f]	[127.08451f]	South Korea	4
6	18	[32.08439f]	[34.81297f]	Israel	1
7	57	[46.768517f]	[23.585135f]	Romania	2
8	55	[47.06615f]	[7.2015657f]	Switzerland	2
9	455	[47.977184f]	[2.2232702f]	France	12
10	92	[52.15457f]	[4.49463f]	Netherlands	5
11	21	[49.944717f]	[84.528114f]	Russia	2
12	1112	[41.545982f]	[1.7138832f]	Spain	13
13	1381	[43.352654f]	[12.727126f]	Italy	46
14	188	[-22.494667f]	[-45.4818f]	Brazil	3
15	56	[51.24459f]	[10.360385f]	Germany	2
16	0	[-16.49901f]	[-68.14626f]	Bolivia	1

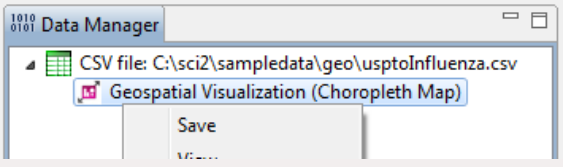
226

Choropleth Map



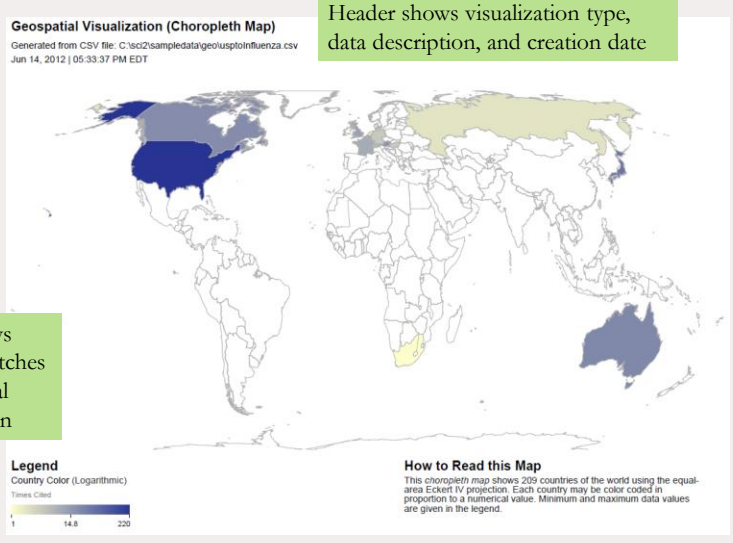


Right-click and **Save** map as PostScript file. Use PostScript Viewer or convert to pdf to view.



27

Reading the Choropleth Map

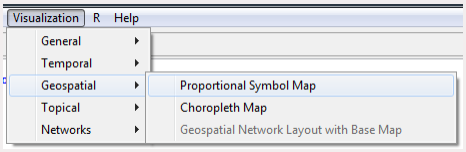


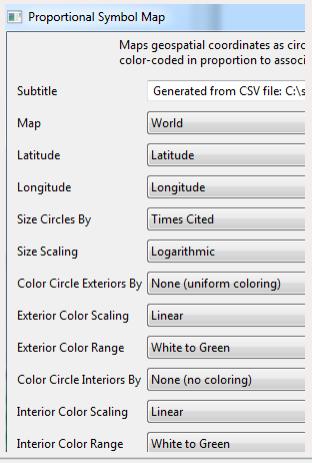
Header shows visualization type, data description, and creation date

Legend shows how data matches up with visual representation

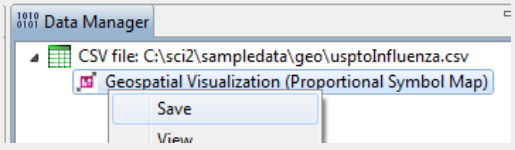
28

Proportional Symbol Map



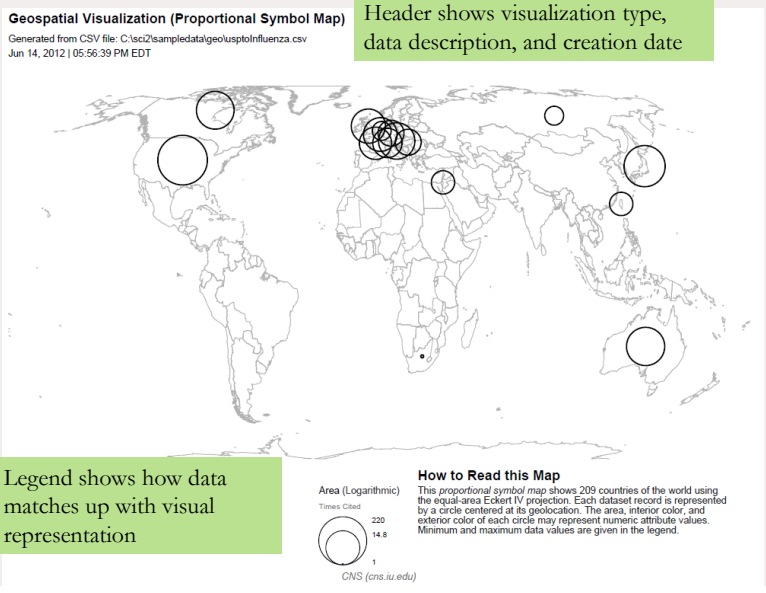


Right-click and **Save** map as PostScript file. Use PostScript Viewer or convert to pdf to view.



29

Reading the Proportional Symbol Map



Header shows visualization type, data description, and creation date

Legend shows how data matches up with visual representation

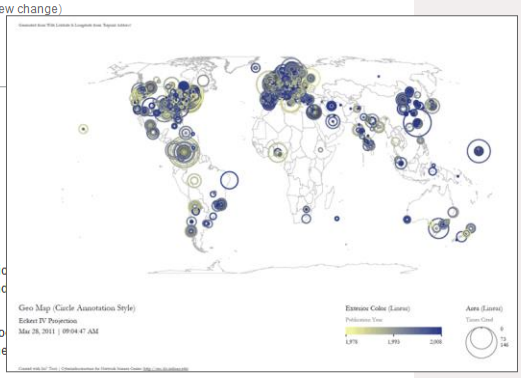
30

Relevant Sci2 Manual entry

Edit Add Tools

5.2.4 Mapping Scientometrics (ISI Data)

9 Added by Ted Polley, last edited by Ted Polley on Nov 14, 2011 (view change)



5.2.4.1 Document Co-Citation

Scientometrics.isi	
Time frame:	1978-2008
Region(s):	Miscellaneous
Topical Area(s):	Scientometrics
Analysis Type(s):	Document Co-Citation Network

Scientometrics is a discipline which uses statistical and computatic
science. Here we use ISI data from the journal "Scientometrics" and
Awards Search.

Download Scientometrics.isi. Load the file using '*File > Load*' and lo
document co-citation analysis, as the scale is large enough that the
similarity within the domain of scientometrics.

New ISI File Format

Web of Science made a change to their output format in September, 2011. Older versions of Sci2 tool may refuse to load these new files, with an error like "Invalid ISI format file selected."

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

31

CNS Cyberinfrastructure for Network Science Center

Questions?

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- Sci2 development and workflow design
- Geospatial Analysis: US and world maps
- **Temporal Analysis: Horizontal line graph of NSF projects**
- **Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records**

12:00-12:45pm Break for Lunch

- Topical Analysis: Visualize research profiles
- Network Analysis: Visualizing the Florentine Network
- Network Analysis: Word Co-occurrence Networks

Q&A

3:00p Adjourn

Introduction to Temporal Analysis

- Science evolves over time
- Temporal analysis seeks to study this evolution by examining patterns, trends, seasonality, outliers, and bursts of activity
- Time series data can be thought of as either discrete or continuous
- Many scholarly datasets can be understood as a discrete time series with events or observations (publications etc.) that happen at regularly spaced intervals (journal publication cycles etc.)

Temporal Analysis: Temporal Bar Graph

- Visualizes numeric data over time
- It accepts a CSV file as input, including NSF grant data
- Start and end dates for each record are necessary to use the temporal bar graph visualization algorithm
- The output of the visualization consists of labeled horizontal bars that correspond to records in the original dataset.

35

Temporal Analysis: Slice Table by Time

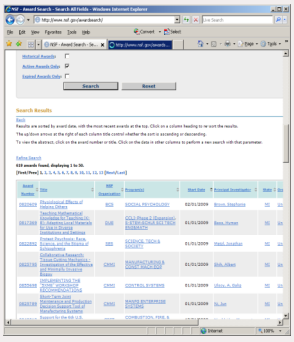
- Divides a table into new tables based on date/time column
- The column for date should have a single value for each row of data
- The output of this algorithm is separate tables so longitudinal analysis will require working with separate files, networks can be extracted from each of these tables to show evolution of a network over time
- The Slice Table by Time algorithm uses the Joda Time library extensively

36

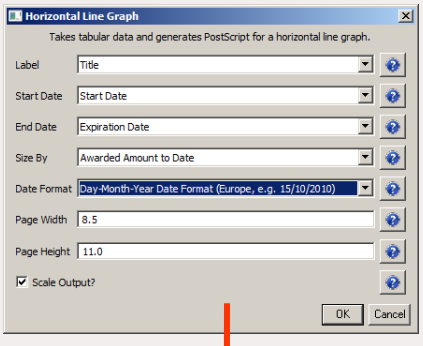
Horizontal line graph of NSF projects

See 5.2.1 Funding Profiles of Three Universities (NSF Data)

Download NSF data



Visualize as Horizontal Line Graph



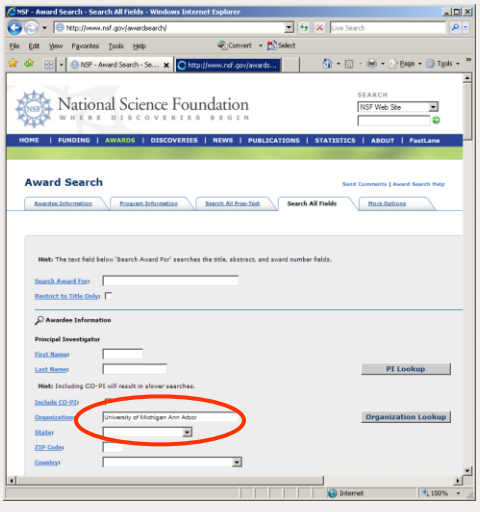
Area size equals numerical value, e.g., award amount.

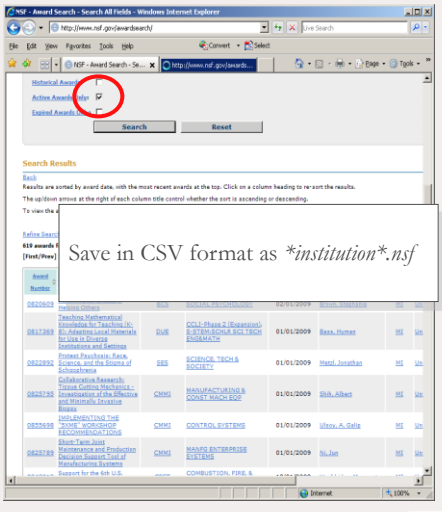
Text [redacted]

Start date — End date

Horizontal line graph of NSF projects

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



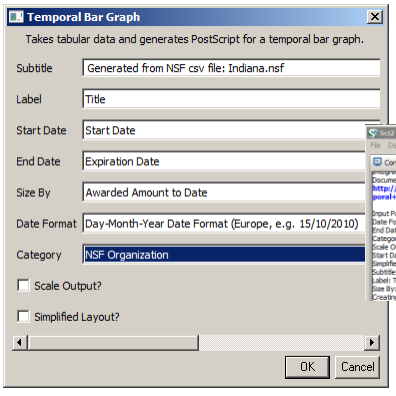


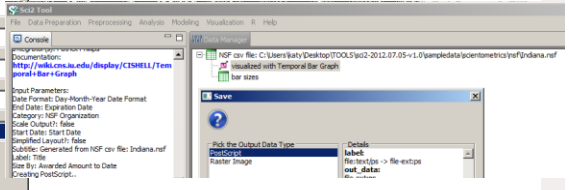
Save in CSV format as **institution*.nsf*

Temporal bar graph of NSF projects

Load a dataset of your choice from the NSF sample data files, e.g., ***'sampledata/scientometrics/nsf/Indiana.nsf.'***

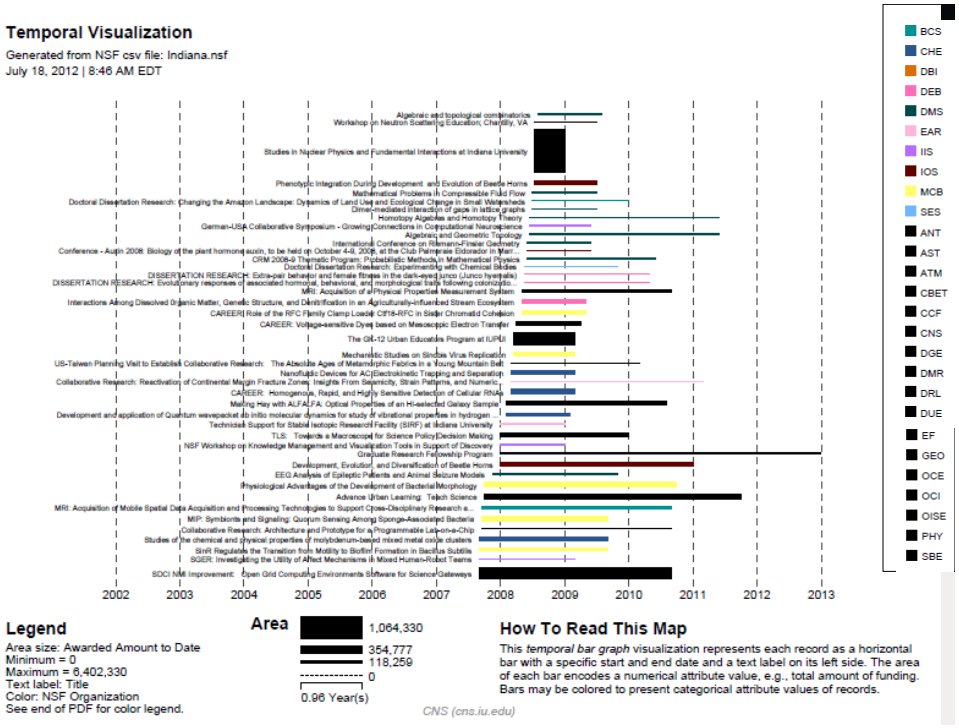
Run ***'Visualization > Temporal > Temporal Bar Graph'*** using parameters:

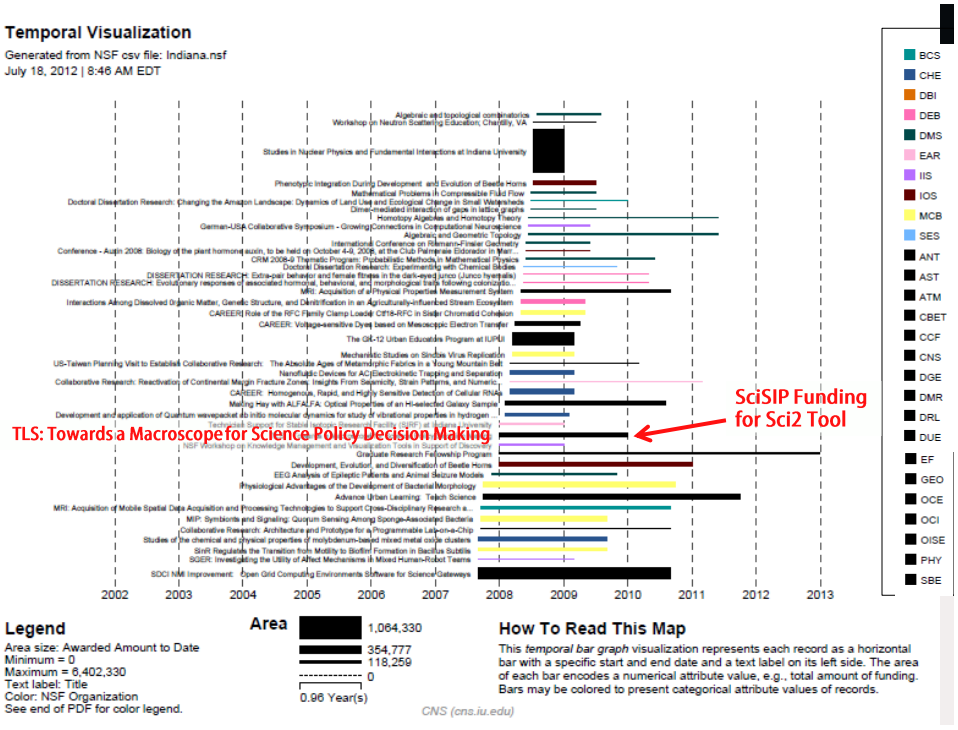


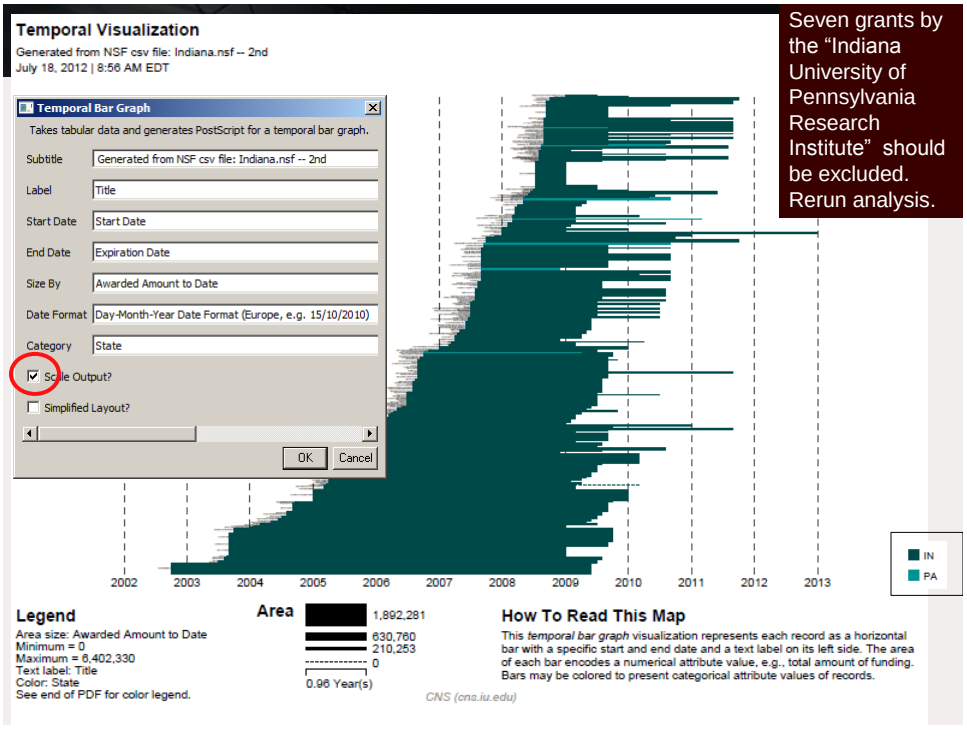


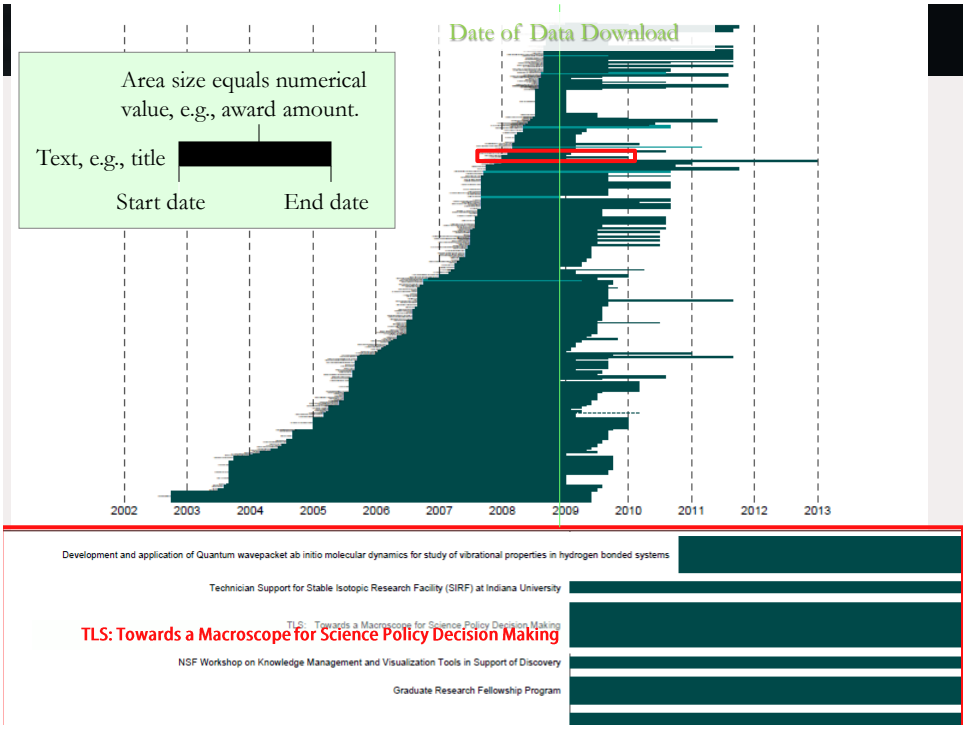
Save *'visualized with Horizontal Line Graph'* as ps or eps file. Convert into pdf and view.

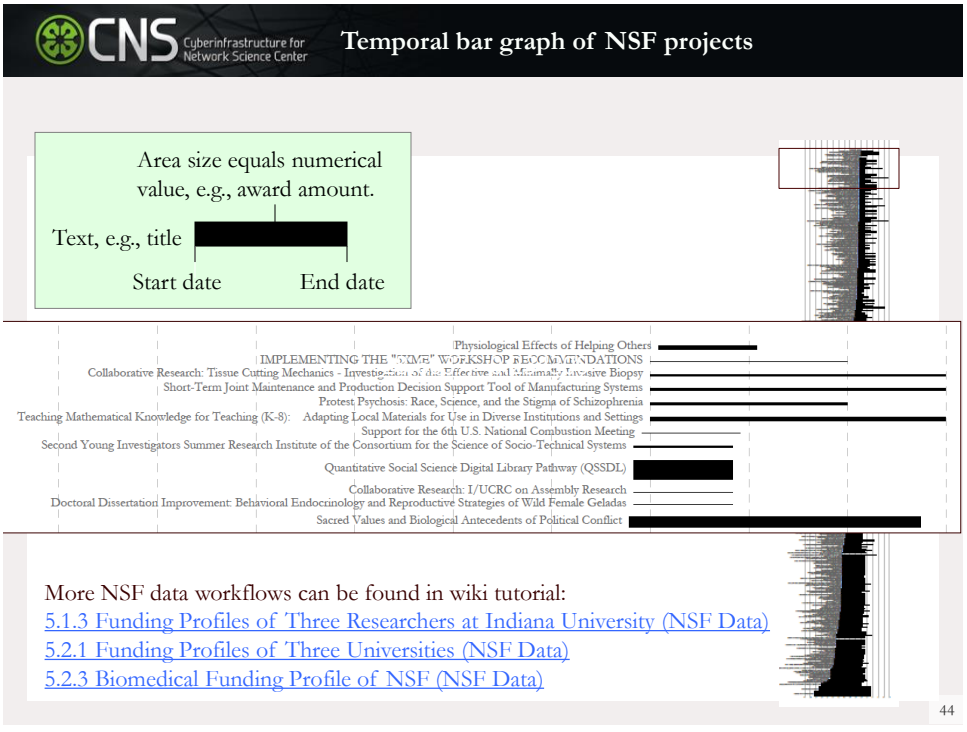
Zoom to see details in visualizations of large datasets, e.g., all NSF awards ever made.




Temporal Visualization
Generated from NSF csv file: Indiana.nsf
July 18, 2012 | 8:46 AM EDT
TLS: Towards a Macroscope for Science Policy Decision Making
SciSIP Funding for Sci2 Tool
BCS
CHE
DBI
DEB
DMS
EAR
IIS
IOS
MCB
SES
ANT
AST
ATM
CBET
CCF
CNS
DGE
DMR
DRL
DUE
EF
GEO
OCE
OCI
OISE
PHY
SBE
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Legend
Area size: Awarded Amount to Date
Minimum = 0
Maximum = 6,402,330
Text label: Title
Color: NSF Organization
See end of PDF for color legend.
Area
1,064,330
354,777
118,259
0
0.96 Year(s)
How To Read This Map
This temporal bar graph visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.
CNS (cns.iu.edu)


Temporal Visualization
Generated from NSF csv file: Indiana.nsf -- 2nd
July 18, 2012 | 8:56 AM EDT
Seven grants by the "Indiana University of Pennsylvania Research Institute" should be excluded. Rerun analysis.
Temporal Bar Graph
Takes tabular data and generates PostScript for a temporal bar graph.
Subtitle
Generated from NSF csv file: Indiana.nsf -- 2nd
Label
Title
Start Date
Start Date
End Date
Expiration Date
Size By
Awarded Amount to Date
Date Format
Day-Month-Year Date Format (Europe, e.g. 15/10/2010)
Category
State
Scale Output?
Simplified Layout?
OK
Cancel
IN
PA
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Legend
Area size: Awarded Amount to Date
Minimum = 0
Maximum = 6,402,330
Text label: Title
Color: State
See end of PDF for color legend.
Area
1,892,281
630,760
210,253
0
0.96 Year(s)
How To Read This Map
This temporal bar graph visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.
CNS (cns.iu.edu)


Date of Data Download
Area size equals numerical value, e.g., award amount.
Text, e.g., title
Start date
End date
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Development and application of Quantum wavepacket ab initio molecular dynamics for study of vibrational properties in hydrogen bonded systems
Technician Support for Stable Isotopic Research Facility (SIRF) at Indiana University
TLS: Towards a Macroscope for Science Policy Decision Making
NSF Workshop on Knowledge Management and Visualization Tools in Support of Discovery
Graduate Research Fellowship Program


CNS
Cyberinfrastructure for Network Science Center
Temporal bar graph of NSF projects
Area size equals numerical value, e.g., award amount.
Text, e.g., title
Start date
End date
Physiological Effects of Helping Others
IMPLEMENTING THE "WME" WORKSHOP RECOMMENDATIONS
Collaborative Research: Tissue Cutting Mechanics - Investigation of the Effective and Minimally Invasive Biopsy
Short-Term Joint Maintenance and Production Decision Support Tool of Manufacturing Systems
Protest Psychosis: Race, Science, and the Stigma of Schizophrenia
Teaching Mathematical Knowledge for Teaching (K-8):
Adapting Local Materials for Use in Diverse Institutions and Settings
Support for the 6th U.S. National Combustion Meeting
Second Young Investigators Summer Research Institute of the Consortium for the Science of Socio-Technical Systems
Quantitative Social Science Digital Library Pathway (QSSDL)
Collaborative Research: I/UCRC on Assembly Research
Doctoral Dissertation Improvement: Behavioral Endocrinology and Reproductive Strategies of Wild Female Geladas
Sacred Values and Biological Antecedents of Political Conflict
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)
44

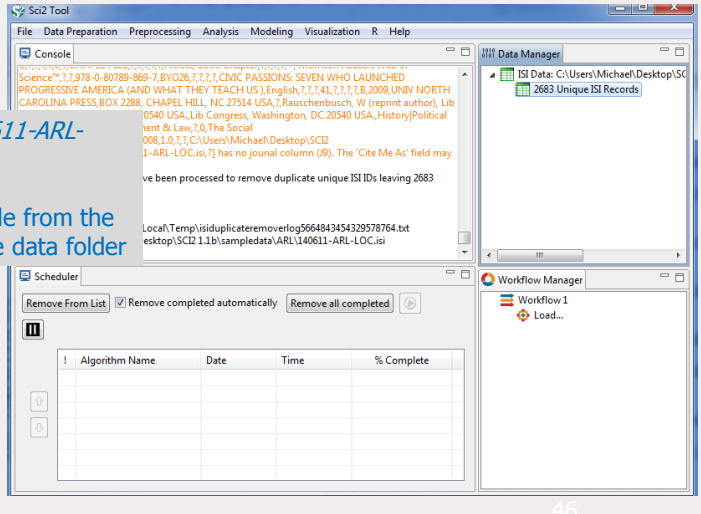
Temporal Analysis: Burst Detection

- Sci2 uses an implementation of Kleinberg's burst detection algorithm (Kleinberg 2002) to study bursts in usage of words in scholarly data
- Algorithm does not calculate the frequency of individual words
- Algorithm uses probabilistic model to determine the rate at which use of a word increases or decreases, identifying bursts in usage of a word

Kleinberg, J. (2002). Bursty and Hierarchical Structure in Streams. Proceedings from the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada: ACM.

45

Temporal Analysis: Burst Detection

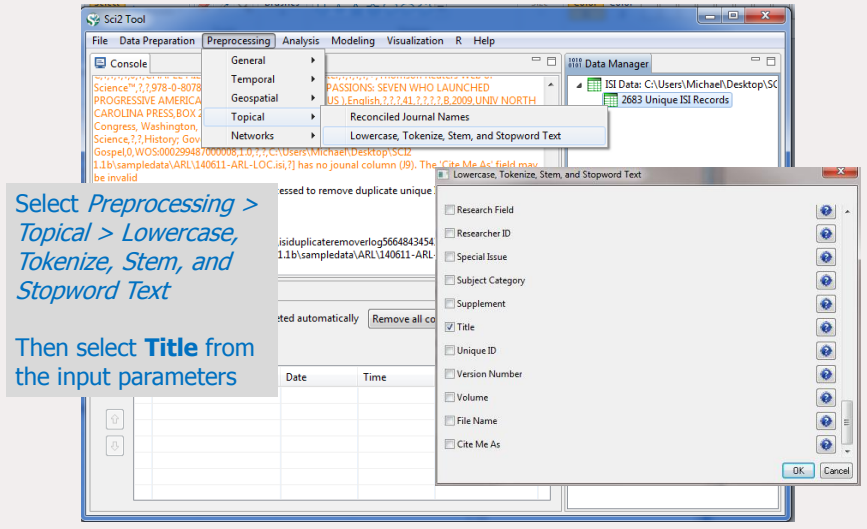


Load *140611-ARL-LOC.isi*

Load this file from the ARL sample data folder

46

Temporal Analysis: Burst Detection

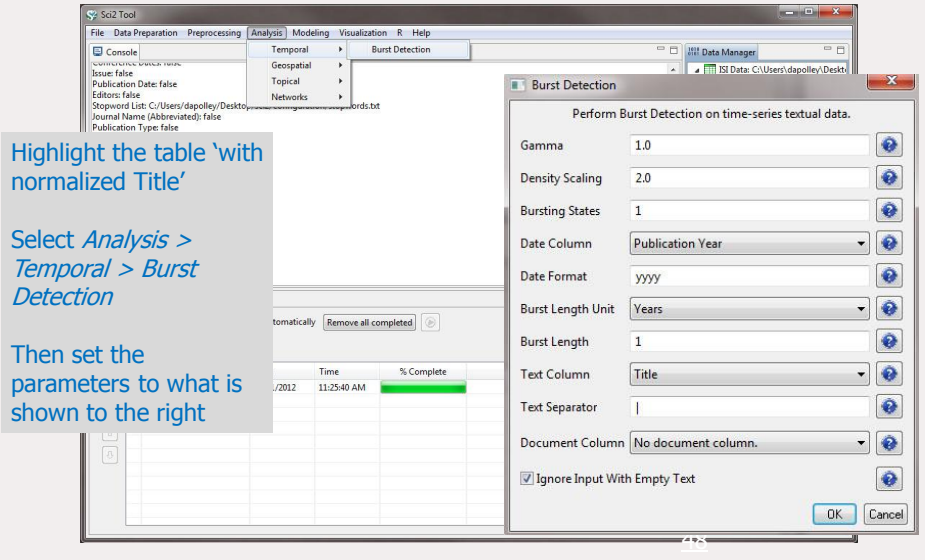


Select *Preprocessing > Topical > Lowercase, Tokenize, Stem, and Stopword Text*

Then select **Title** from the input parameters

47

Temporal Analysis: Burst Detection



Highlight the table 'with normalized Title'

Select *Analysis > Temporal > Burst Detection*

Then set the parameters to what is shown to the right

48

Temporal Analysis: Burst Detection

Gamma – the higher this value, the smaller the list of generated bursts.

Density Scaling – determines how much "more bursty" each level is beyond the previous one.

Bursting states – determines how many bursting states there will be, beyond the non-bursting states.

Date Column – name of the column in the original data with date/time when events/topics happen.

Date Format – specifies how the date column will be interpreted.

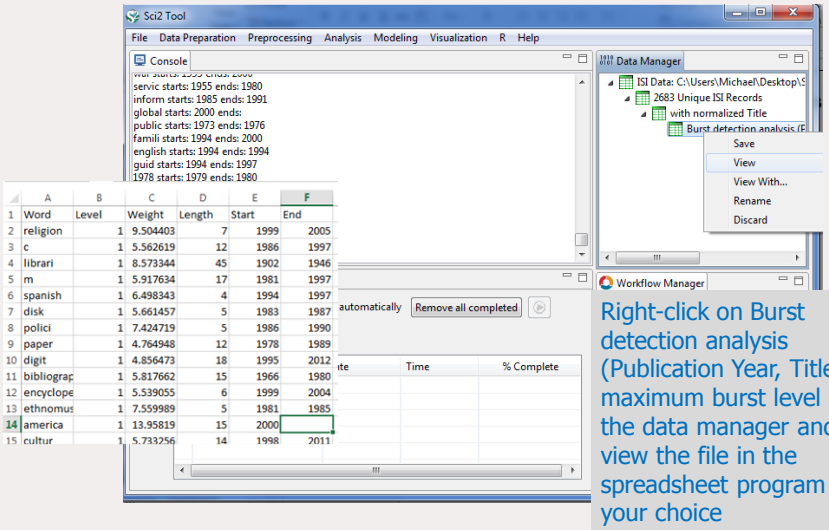
Burst Length Unit – specifies how to divide the date range into burstable units.

Burst Length – specifies the number of burstable units per burstable period.

Text Column – the name of the column with values (delimiter and tokens) to be computed for bursting results.

Text Separator – delimits the tokens in the text column.

Temporal Analysis: Burst Detection



Right-click on Burst detection analysis (Publication Year, Title): maximum burst level 1 in the data manager and view the file in the spreadsheet program of your choice

Temporal Analysis: Burst Detection

Missing end dates indicate the continuation of a burst in a given data set. Add the End date of 2014 to those records missing and End date.

	A	B	C	D	E	F
1	Word	Level	Weight	Length	Start	End
2	free	1	3.232962	3	2002	2004
3	critic	1	4.31613	6	1993	1998
4	complex	1	3.538345	6	2001	
5	transform	1	4.492169	6	1990	1995
6	sandpil	1	4.650639	3	1998	2000
7	approach	1	3.381684	4	1994	1997
8	self	1	3.764748	6	1993	1998
9	fractal	1	3.767573	8	1990	1997
10	network	1	12.33559	5	2002	
11	renorm	1	3.560887	5	1994	1998
12	fix	1	3.840594	6	1990	1995
13	absorb	1	3.049794	3	1998	2000

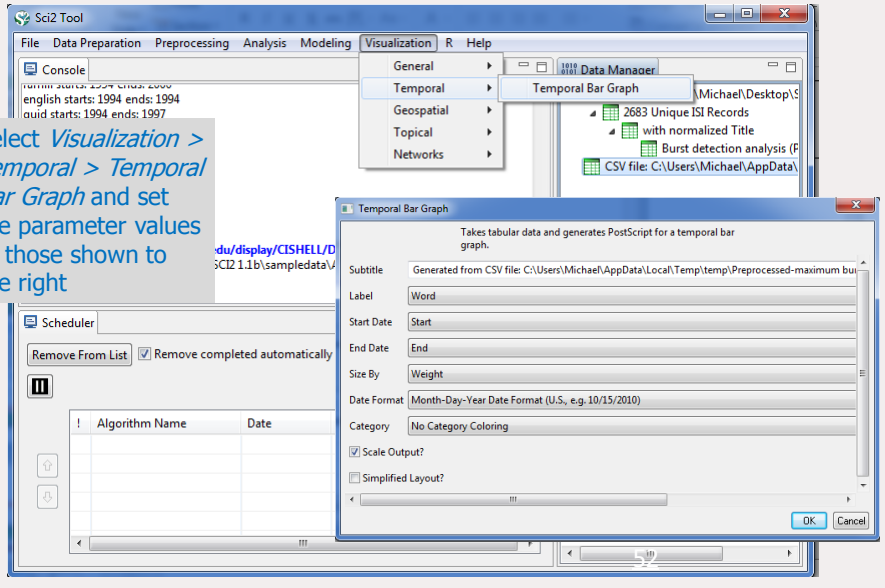
	A	B	C	D	E	F
1	Word	Level	Weight	Length	Start	End
2	free	1	3.232962	3	2002	2004
3	critic	1	4.31613	6	1993	1998
4	complex	1	3.538345	6	2001	2006
5	transform	1	4.492169	6	1990	1995
6	sandpil	1	4.650639	3	1998	2000
7	approach	1	3.381684	4	1994	1997
8	self	1	3.764748	6	1993	1998
9	fractal	1	3.767573	8	1990	1997
10	network	1	12.33559	5	2002	2006
11	renorm	1	3.560887	5	1994	1998
12	fix	1	3.840594	6	1990	1995
13	absorb	1	3.049794	3	1998	2000

Save the file as a .CSV file and load it back into Sci2, selecting the Standard CSV format

51

Temporal Analysis: Burst Detection

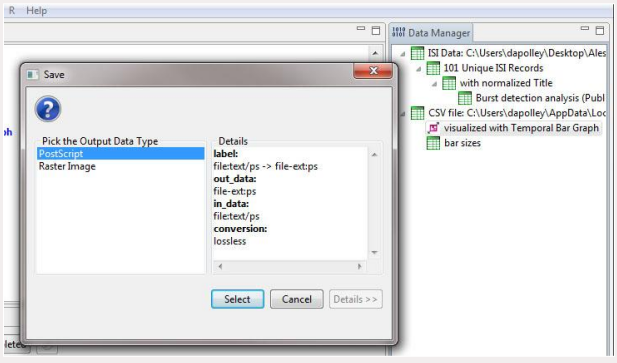
Select *Visualization > Temporal > Temporal Bar Graph* and set the parameter values to those shown to the right



52

Temporal Analysis: Burst Detection

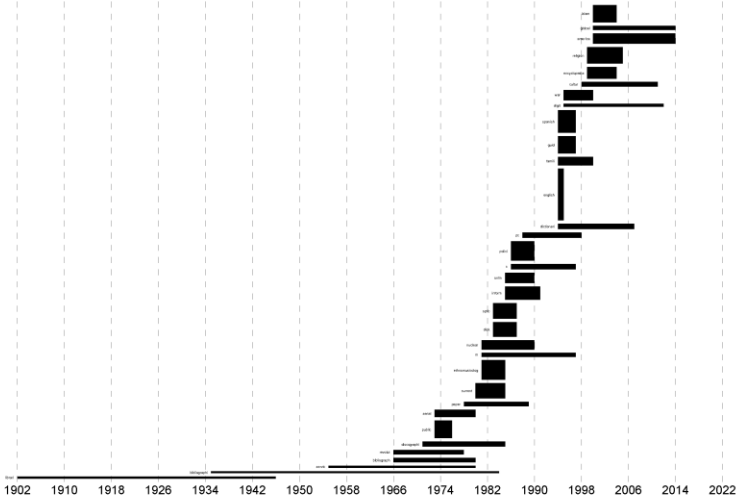
Right-click on the visualized with **Temporal Bar Graph** file in the Data Manager and save the PostScript file to your desired location



If you do not have a program to convert PostScript files look here.

Temporal Visualization

(Generated from CSV file: C:\Users\Michael\AppData\Local\Temp\temp\Preprocessed-maximum burst level 11510809905973920427-8253546097942883737.csv)
June 18, 2014 | 3:25 AM EDT



Legend
Area size: Weight
Minimum = 5
Maximum = 14
Text label: Word

Area
22
7
2
0
10.42 Year(s)

How To Read This Map
This *temporal bar graph* visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.

CNS (cns.iu.edu)

Questions?

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- Sci2 development and workflow design
- Geospatial Analysis: US and world maps
- Temporal Analysis: Horizontal line graph of NSF projects
- Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records

12:00-12:45pm Break for Lunch

- **Topical Analysis: Visualize research profiles**
- Network Analysis: Visualizing the Florentine Network
- Network Analysis: Word Co-occurrence Networks

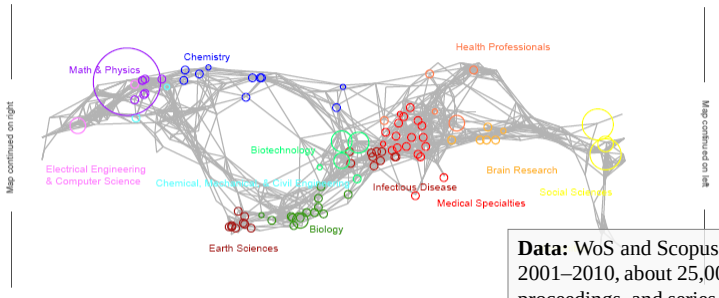
Q&A

3:00p Adjourn

Topical Analysis: Research Profiles

Topical Visualization

Generated from 361 Unique ISI Records of 4 NetSci Researchers
14 out of 109 publications were mapped to 94 subdisciplines and 12 disciplines.
June 05, 2012 | 05:39 PM EDT



2008 The Regents of the University of California and SciTech Strategie
Map updated by SciTech Strategies, OST, and CNS in 2011.

Legend
Circle area: Fractional Journal Count
Unclassified = 96
Minimum = 0
Maximum = 25
Color: Discipline
See end of PDF for color legend.

Area
29.09
16.19
2.8

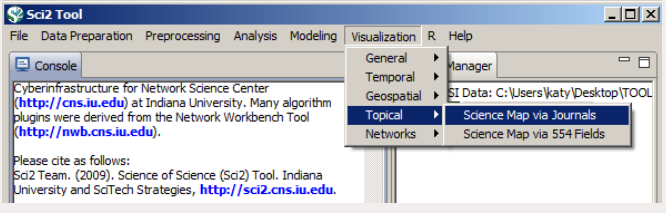
How To Read This Map
The *UCSD map of science* depicts
are aggregated to 13 main disciplin
color and is labeled. Overlaid are c
unique subdiscipline. Circle area is
assigned records. Minimum and m

CNS (cns.iu.edu)

Data: WoS and Scopus paper level data for 2001–2010, about 25,000 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 disciplines.
Börner, Katy, Richard Klavans, et al. (2012) Design and Update of a Classification System: The UCSD Map of Science. PLoS ONE 7(7): e39464. doi:10.1371/journal.pone.0039464

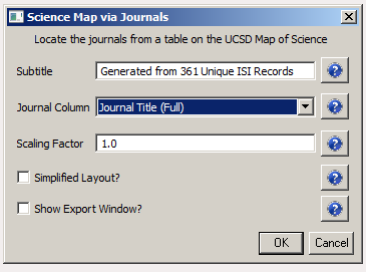
Research Profiles—Publication Data

Load an ISI (*.isi), Bibtex (*.bib), Endnote Export Format (*.enw), Scopus csv (*.scopus) file such as */sci2/sampledata/scientometrics/isi/FourNetSciResearchers.isi*



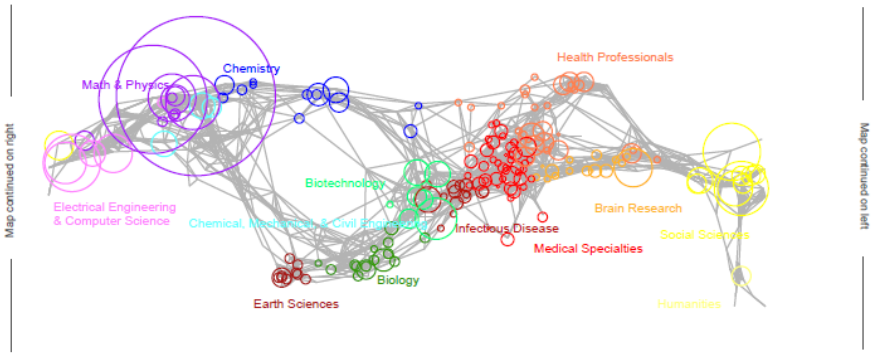
Run '*Visualization* > *Topical* > *Science Map via Journals*' using parameters given to the right.

Postscript file will appear in *Data Manager*. Save and open with a Postscript Viewer.



Topical Visualization

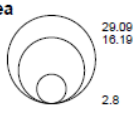
Generated from 361 Unique ISI Records
90 out of 112 publications were mapped to 182 subdisciplines and 13 disciplines.
June 24, 2012 | 04:04 PM EDT



Legend

Circle area: Fractional Journal Count
Unclassified = 22
Minimum = 0
Maximum = 98
Color: Discipline
See end of PDF for color legend.

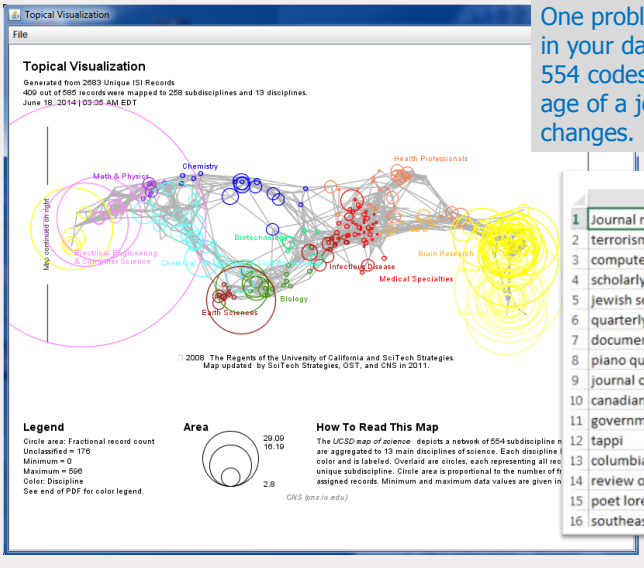
Area



How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)



One problem you may encounter in your data are journals lacking 554 codes. Reasons: relevance, age of a journal, and name changes.

	A	B
1	Journal name	Frequency
2	terrorism	6
3	computer performance	1
4	scholarly publishing	2
5	jewish social studies	1
6	quarterly journal of the library of con	74
7	document & image automation	1
8	piano quarterly	1
9	journal of the society of motion pictu	3
10	canadian slavonic papers-revue cana	1
11	government publications review	12
12	tappi	3
13	columbia journal of world business	1
14	review of public data use	1
15	poet lore	3
16	southeastern europe	2

Topical Visualization

Generated from 361 Unique ISI Records
90 out of 112 publications were mapped to 182 subdisciplines and 13 disciplines.
June 24, 2012 | 04:04 PM EDT

Biology
1 BMC EVOLUTIONARY BIOLOGY
1 NATURWISSENSCHAFTEN

Biotechnology
1 BMC BIOINFORMATICS
2 FEBS JOURNAL
1 GENOME RESEARCH
1 INTERNATIONAL MICROBIOLOGY
1 NATURE BIOTECHNOLOGY
3 NATURE GENETICS
1 NATURE REVIEWS GENETICS
1 NUCLEIC ACIDS RESEARCH
2 PROTEOMICS

Brain Research
5 JOURNAL OF MATHEMATICAL PSYCHOLOGY

Chemical, Mechanical, & Civil Engineering
1 JOURNAL OF CERAMIC PROCESSING RESEARCH
2 MATERIALS SCIENCE AND ENGINEERING A-STRUCTURAL MATERIA...
1 PHYSICS WORLD
1 SCIENTIFIC AMERICAN

Chemistry
1 COMPUTER PHYSICS COMMUNICATIONS
2 JOURNAL OF CHEMICAL INFORMATION AND COMPUTER SCIENCES
1 JOURNAL OF THE INDIAN INSTITUTE OF SCIENCE
1 PURE AND APPLIED CHEMISTRY

Earth Sciences
1 CURRENT SCIENCE

Electrical Engineering & Computer Science
1 ASIST 2003: PROCEEDINGS OF THE 66TH ASIST ANNUAL MEETING,...
1 CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REV...
5 IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION
1 INFORMATION TECHNOLOGY AND LIBRARIES
5 JOURNAL OF INFORMATION SCIENCE
3 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE
5 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENC...
2 LIBRARY QUARTERLY
1 LIBRI
1 PROCEEDINGS OF THE AMERICAN SOCIETY FOR INFORMATION SC...

Health Professionals
1 ANNALS OF BIOMEDICAL ENGINEERING
1 BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION
1 CROATIAN MEDICAL JOURNAL
2 JOURNAL OF APPLIED PHYSIOLOGY
1 JOURNAL OF PUBLIC HEALTH DENTISTRY
1 METHODS OF INFORMATION IN MEDICINE
1 PLASTIC AND RECONSTRUCTIVE SURGERY
1 TEXAS MEDICINE
1 UNFALLCHIRURG
1 WIENER KLINISCHE WOCHENSCHRIFT

Humanities
1 BULLETIN OF THE ATOMIC SCIENTISTS

Infectious Diseases
1 FEMS MICROBIOLOGY LETTERS
1 JOURNAL OF BACTERIOLOGY

Math & Physics
1 ADVANCES IN APPLIED PROBABILITY

CNS (cns.iu.edu)

Topical Visualization

Generated from 361 Unique ISI Records
90 out of 112 publications were mapped to 182 subdisciplines and 13 disciplines.
June 24, 2012 | 04:04 PM EDT

Math & Physics
10 APPLIED PHYSICS LETTERS
1 BRAZILIAN JOURNAL OF PHYSICS
3 CHAOS SOLITONS & FRACTALS
1 COMPLEXITY
1 COMPUTATIONAL MATERIALS SCIENCE
11 EUROPEAN PHYSICAL JOURNAL B
12 EUROPHYSICS LETTERS
2 INTERNATIONAL JOURNAL OF MODERN PHYSICS B
6 JOURNAL OF PHYSICS A-MATHEMATICAL AND GENERAL
1 JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT
1 JOURNAL OF STATISTICAL PHYSICS
1 JOURNAL OF THE KOREAN PHYSICAL SOCIETY
1 MATERIALS SCIENCE AND ENGINEERING B-SOLID STATE MATERIAL...
3 NATURE PHYSICS
3 NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH SEC...
12 PHYSICA A
5 PHYSICAL REVIEW A
2 PHYSICAL REVIEW B
45 PHYSICAL REVIEW LETTERS
2 REVIEWS OF MODERN PHYSICS

Medical Specialties
1 ANNALS OF INTERNAL MEDICINE
1 REVISTA DE INVESTIGACION CLINICA

Social Sciences
1 ADMINISTRATIVE SCIENCE QUARTERLY
1 AMERICAN BEHAVIORAL SCIENTIST
1 AMERICAN SOCIOLOGICAL REVIEW
1 ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL S...
1 ARBOR-CIENCIA PENSAMIENTO Y CULTURA
3 BRITISH JOURNAL OF MATHEMATICAL & STATISTICAL PSYCHOLOGY
1 JOURNAL OF CLASSIFICATION

Social Sciences
2 JOURNAL OF MATHEMATICAL SOCIOLOGY
3 JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2 PSYCHOLOGICAL BULLETIN
5 PSYCHOMETRIKA
1 RECHERCHE
5 SCIENTOMETRICS
1 SOCIAL FORCES
6 SOCIAL NETWORKS
3 SOCIOLOGICAL METHODS & RESEARCH

Multiple Categories
1 BRITISH MEDICAL JOURNAL
2 JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
1 JOURNAL OF THEORETICAL BIOLOGY
18 NATURE
44 PHYSICAL REVIEW E
5 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE ...
6 SCIENCE

Unclassified
1 ALGORITHMS AND MODELS FOR THE WEB-GRAPHS, PROCEEDINGS
2 AMERICAN DOCUMENTATION
2 ASIST 2002: PROCEEDINGS OF THE 65TH ASIST ANNUAL MEETING, ...
1 BIOLOGIYA MORYA-MARINE BIOLOGY
1 BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE
1 CHEMIKER-ZEITUNG
3 CHEMTECH
1 COMBINATORIAL AND ALGORITHMIC ASPECTS OF NETWORKING
7 CURRENT COMMENTS
3 CURRENT CONTENTS/LIFE SCIENCES
1 FEDERATION PROCEEDINGS
5 FRACTALS-AN INTERDISCIPLINARY JOURNAL ON THE COMPLEX GE...
1 FRONTIERS OF LIBRARIANSHIP-SYRACUSE UNIVERSITY

CNS (cns.iu.edu)

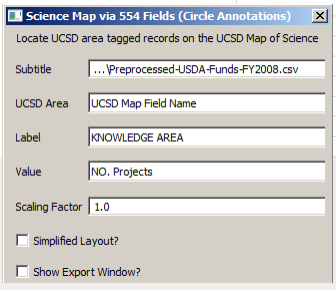
Research Profiles—Existing Classifications

In addition to using **journal names** to
- Map career trajectories
- Identify evolving expertise areas
- Compare expertise profiles

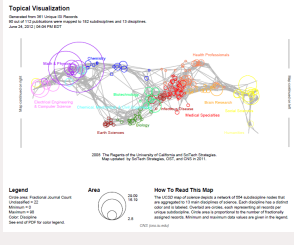
Existing classifications can be aligned and used to generate science map overlays.

B	C	D	E	F	G
KNOWLEDGE AREA	NO. Projects	USDA Staff Years	STATE APPR	TOTAL FUNDS	UCSD Map Field Name
101 Appraisal of Soil Resources					315
102 Soil, Plant, Water, Nutrient Relationships					227
103 Management of Saline and Sodic Soils and Salinity					158
104 Protect Soil from Harmful Effects of Natural Elements					120
111 Conservation and Efficient Use of Water					245
112 Watershed Protection and Management					245
121 Management of Range Resources					520
122 Management and Control of Forest and Range Fires					520
123 Management and Sustainability of Forest Resources					231
124 Urban Forestry					231
125 Agroforestry					231

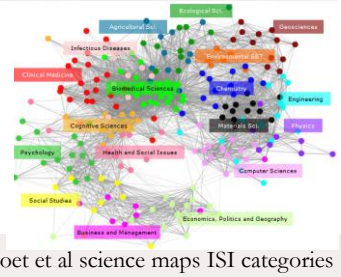
Run *Visualization > Topical > Science Map via 554 Fields* using parameters given to the right.
Postscript file will appear in *Data Manager*.
Save and open with a Postscript Viewer.



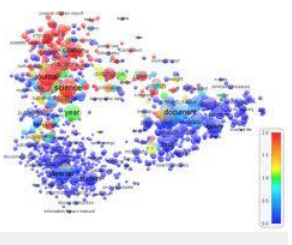
Align Science Basemaps using the Sci2 Tool



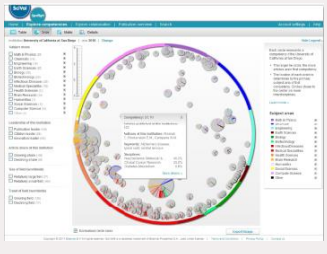
UCSD Map



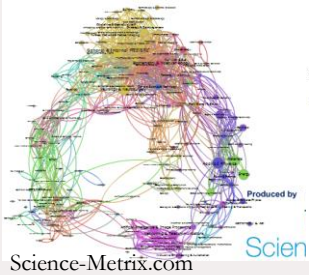
Loet et al science maps ISI categories



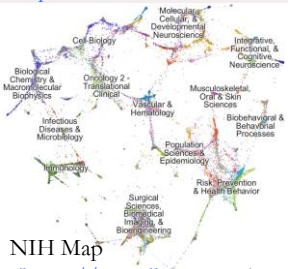
http://vosviewer.com



Elsevier's SciVal Map



Science-Metrix.com



NIH Map
(https://app.nihmaps.org)

CNS Cyberinfrastructure for Network Science Center

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- Sci2 development and workflow design
- Geospatial Analysis: US and world maps
- Temporal Analysis: Horizontal line graph of NSF projects
- Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records

12:00-12:45pm Break for Lunch

- Topical Analysis: Visualize research profiles
- **Network Analysis: Visualizing the Florentine Network**
- Network Analysis: Word Co-occurrence Networks

Q&A

3:00p Adjourn

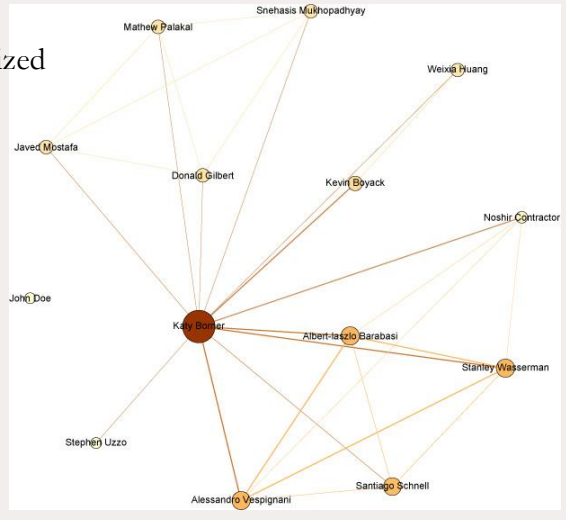
65

CNS Cyberinfrastructure for Network Science Center

Introduction to Network Analysis

What is a Network?

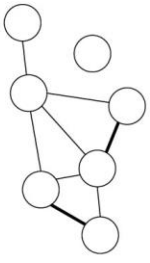
- Graph – network visualized
- Nodes (vertices)
- Edges (links)



66

66

Undirected Networks



Nodes:

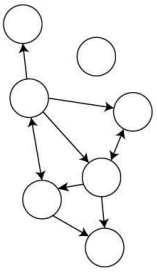
Edges:

Node Degree:
Number of edges connected to nodes

Isolates:
Nodes that are not connected to the rest of the network

Edge Weight:
Demonstrates relative importance of relationships

Directed Networks



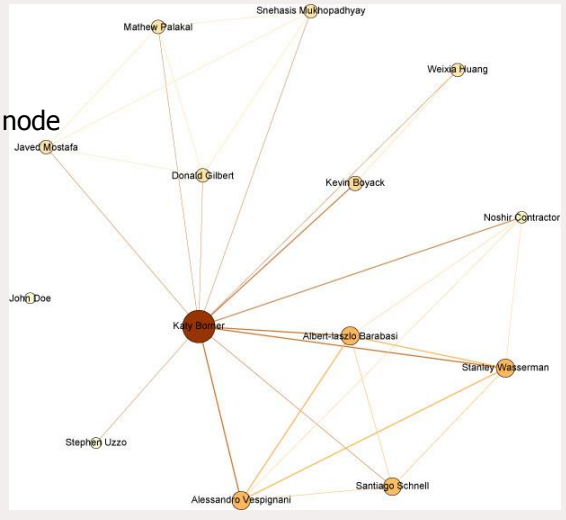
Directional relationship is represented by arrows

Number of incoming edges

Number of outgoing edges

Graph Metrics - Nodes

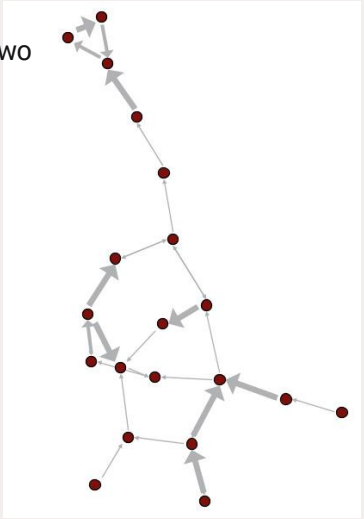
- Betweenness Centrality – number of shortest paths a node sits between
- Degree
- Isolates



68

Graph Metrics - Edges

- Shortest paths – shortest distance between two nodes
- Weight – strength of tie
- Directionality – is the connection one-way or two-way (in-degree vs. out-degree)?
- Bridge – deleting would change structure



69

This example will demonstrate how to visualize data using Sci2. In this workflow we will be working with Padgett's Florentine families dataset which includes 16 different Italian families from the early 15th century. Each family is represented by a node in the network and families are connected by edges that represent either a marriage or business/lending ties. Each node (family) has several attributes: wealth (in thousands of lira), number of priorates (seats on the civic council between 1282-1344), and total ties (total number of business ties and marriages in the dataset).

"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 Medici family, the other around the powerful Strozzis."

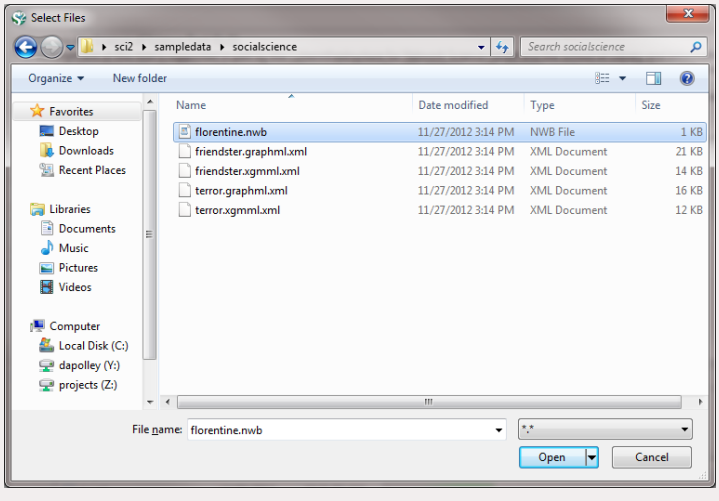
More info at http://svitsrv25.epfl.ch/R-doc/library/ergm/html/florentine.html

Visualizing the Florentine Dataset

```
*Nodes
id*int label*string wealth*int totalities*int priorates*int
1 "Acciaiuoli" 10 2 53
2 "Albizzi" 36 3 65
3 "Barbadori" 55 14 0
4 "Bischeri" 44 9 12
5 "Castellani" 20 18 22
6 "Ginori" 32 9 0
7 "Guadagni" 8 14 21
8 "Lamberteschi" 42 14 0
9 "Medici" 103 54 53
10 "Pazzi" 48 7 0
11 "Peruzzi" 49 32 42
12 "Pucci" 3 1 0
13 "Ridolfi" 27 4 38
14 "Salviati" 10 5 35
15 "Strozzi" 146 29 74
16 "Tornabuoni" 48 7 0
*UndirectedEdges
source*int target*int marriage*string business*string
9 1 "T" "F"
6 2 "T" "F"
7 2 "T" "F"
9 2 "T" "F"
5 3 "T" "T"
```

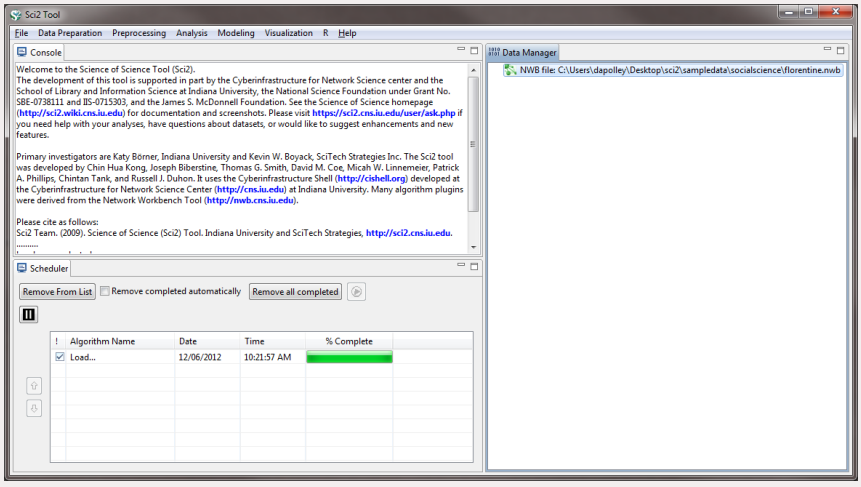

Visualizing the Florentine Dataset

First, load the florentine.nwb by following *File* > *Load* > ***yoursci2directory****/sampledata/scientometrics/endnote/florentine.nwb.*



Visualizing the Florentine Dataset

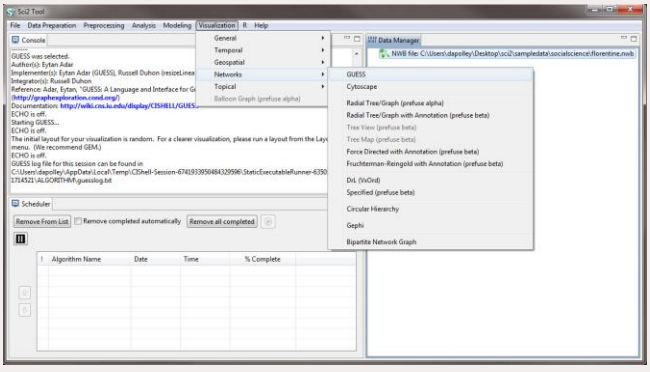
Once you have loaded the data in Sci2, it will appear in the Data Manager.



Visualizing the Florentine Dataset

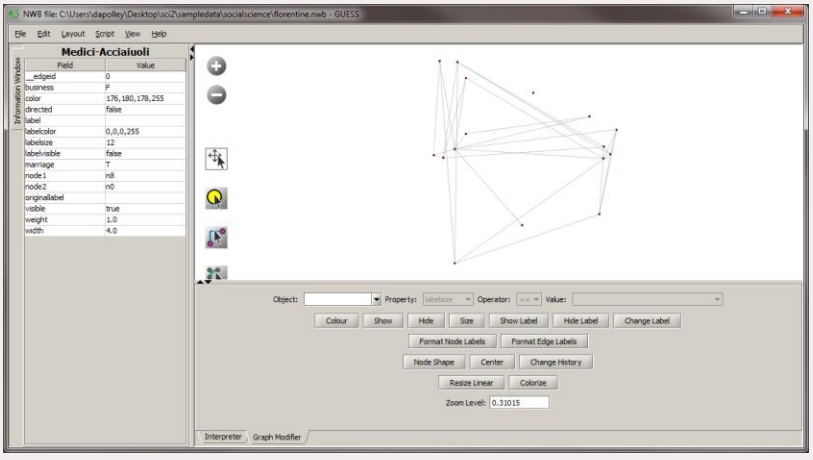
For this workflow we will skip straight to the visualization step, since the network file that we loaded already has the attributes we are interested in visualizing (wealth, priorates, and totalities). For other datasets, you will likely need to extract networks and run some type of analysis to answer the questions you are interested in.

To visualize this network select the file from the Data Manager and run *Visualization* > *Networks* > *GUESS*.



Visualizing the Florentine Dataset

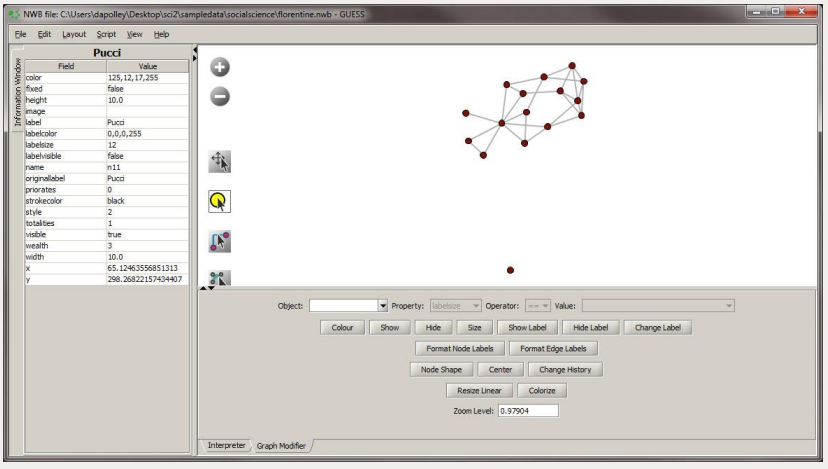
When the network is loaded in GUESS it will be laid out randomly.



75

Visualizing the Florentine Dataset

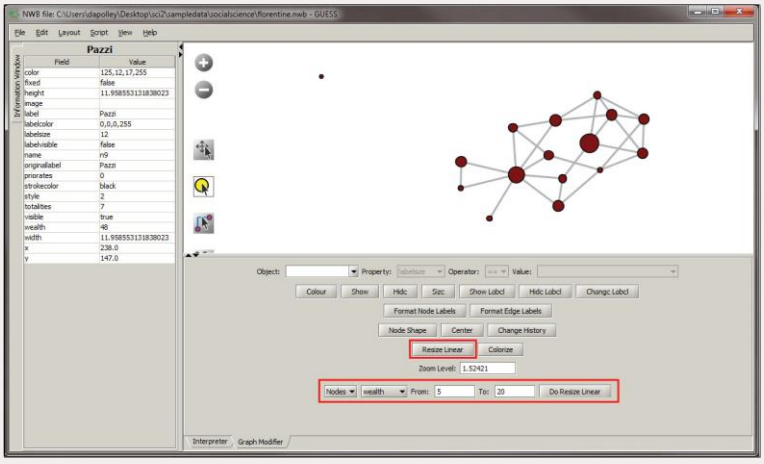
The first step in enhancing this network visualization is to apply a different layout. For this visualization we will use the GEM layout *Layout > GEM.* You will notice that the GEM layout is random, you can run it multiple times and the network will appear slightly different each time.



76

Visualizing the Florentine Dataset

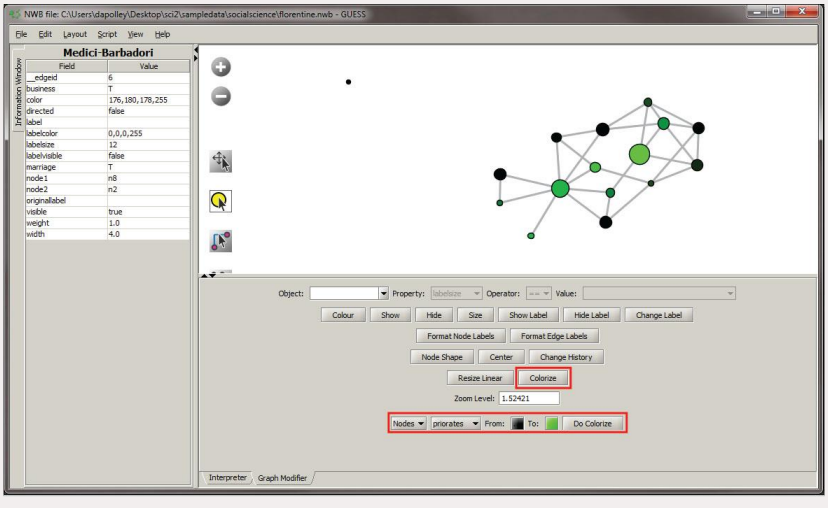
The next step will be to resize the nodes based on the wealth attribute. To do this resize select the *Resize Linear* button and set the parameters to those shown below.



77

Visualizing the Florentine Dataset

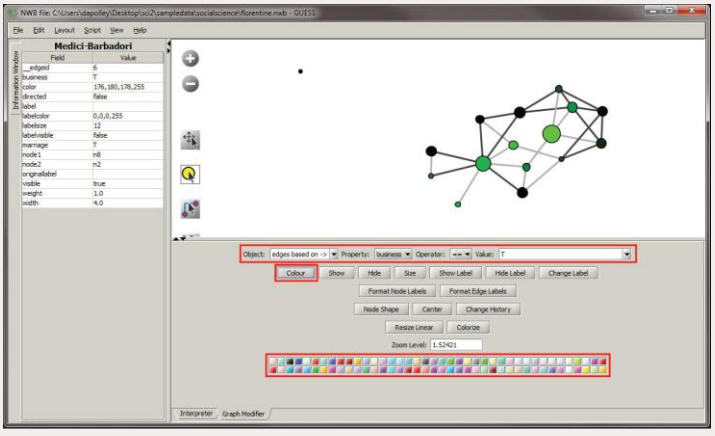
Next we will colorize the nodes based on priorates to add an additional dimension to this visualization.



78

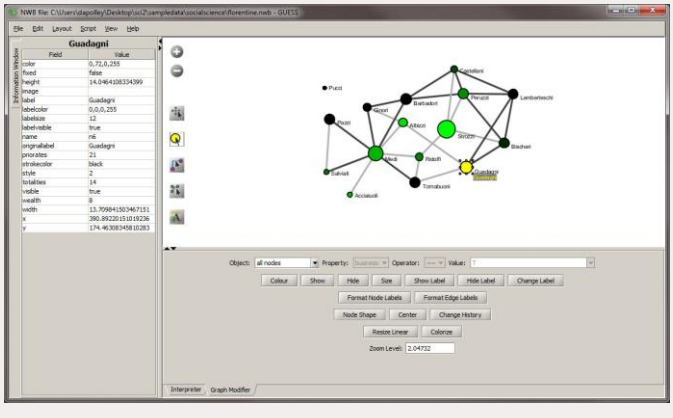
Visualizing the Florentine Dataset

Next we will color the edges to show the type of relationship between the families. To do this, you will need to select the Object edges based on ->, set the property to marriage, the operator to ==, and the value to T. Next, click the Color button and you can select the color of your choice from the pallet that will appear at the bottom of the Graph Modifier pane.



Visualizing the Florentine Dataset

You can repeat this process for the *business* property if you want to, or you can leave the edges that represent business ties the default color. In this workflow we will leave them the default color, light gray. The final step is to show all the labels. To do this, you will need to select the "Object" all nodes and the click the *Show Label* button and the labels will appear in the visualization.



Visualizing the Florentine Dataset

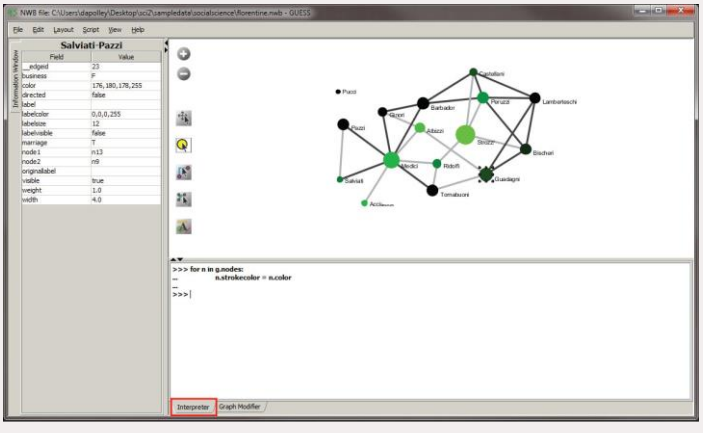
Since the GEM layout is random and all the nodes are spaced more or less evenly apart, you do not have to worry about disrupting the layout. However, other layout algorithms may space the nodes according to specific attributes of the network. Manually moving around nodes in this case would disrupt the layout of the network and distort the meaning of the visualization.

The last thing we want to do to our network is color the border of the nodes the same as the nodes themselves. This is not as crucial for networks with only a few nodes, but as the size of your network increases it can become difficult to read with the thick black lines around every node. To color those the same as the node go to the *Interpreter* tab at the bottom of the GUESS window and type in the following commands:

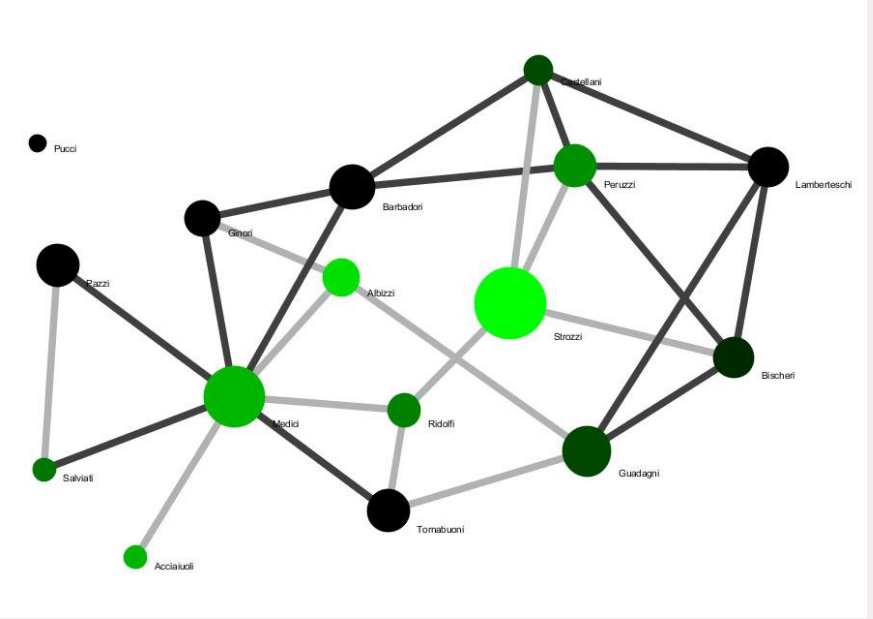
```
for n in g.nodes:
	n.strokecolor = n.color
```


Visualizing the Florentine Dataset

This code basically tells GUESS that for every node (*n*) in this graph of nodes (*g.nodes)* make the border color of the nodes (*n.strokecolor*) equal to the node color (*n.color)*. After you type the first line you will need to hit the "Tab" key before you start typing the second line of code.



Visualizing the Florentine Dataset



Questions?

Tutorial Overview

Welcome and Overview of Tutorial and Attendees
9:00 – 9:15 am Introductions and fill out the pre-questionnaire

9:15 a Sci2 Tool Hands-on

- Sci2 development and workflow design
- Geospatial Analysis: US and world maps
- Temporal Analysis: Horizontal line graph of NSF projects
- Topic/Temporal Analysis: Burst Detection using Library of Congress Web of Science Records

12:00-12:45pm Break for Lunch

- Topical Analysis: Visualize research profiles
- Network Analysis: Visualizing the Florentine Network
- **Network Analysis: Word Co-occurrence Networks**

Q&A

3:00p Adjourn

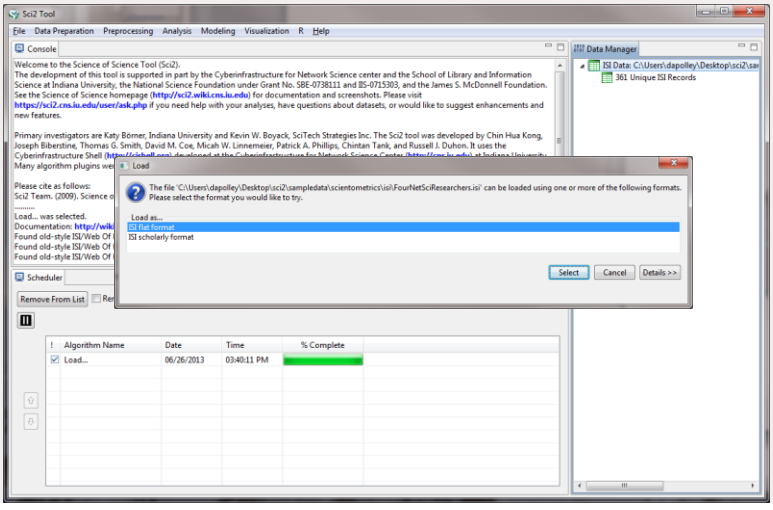
Word Co-Occurrence Network

The topic similarity of works (books, journal articles etc.) within a domain can be calculated via an analysis of the co-occurrence of words in associated texts. Works that share more words in common are assumed to have higher topical overlap and are connected via linkages and/or placed in closer proximity.

Sci2's Extract Word Co-Occurrence Network algorithm creates a weighted network where each node is a word and edges connect words to each other. The strength of an edge represents how often two words co-occur in the same body of text.

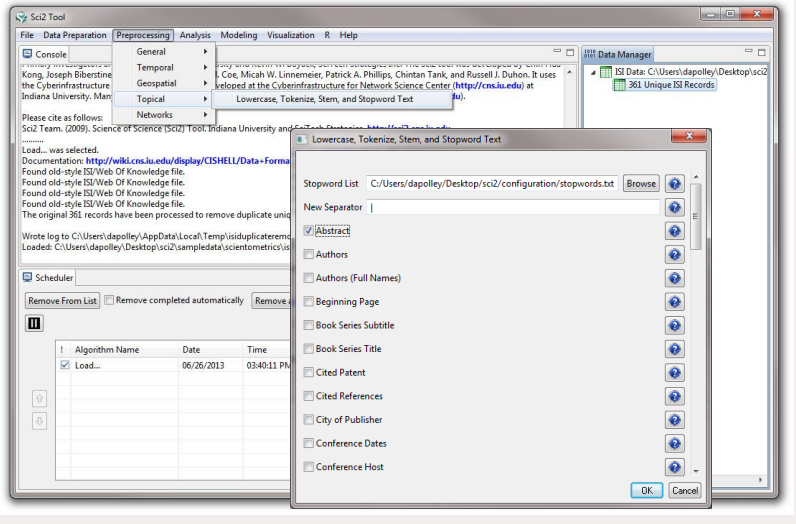
Word Co-Occurrence Network

Load the Four NetSci Researchers file (FourNetSciReseachers.isi) from the sample data folder in your Sci2 installation directory. Here is the path: *C:|Users|yourusername|Desktop|sci2|sampledata|scientometrics|isi*



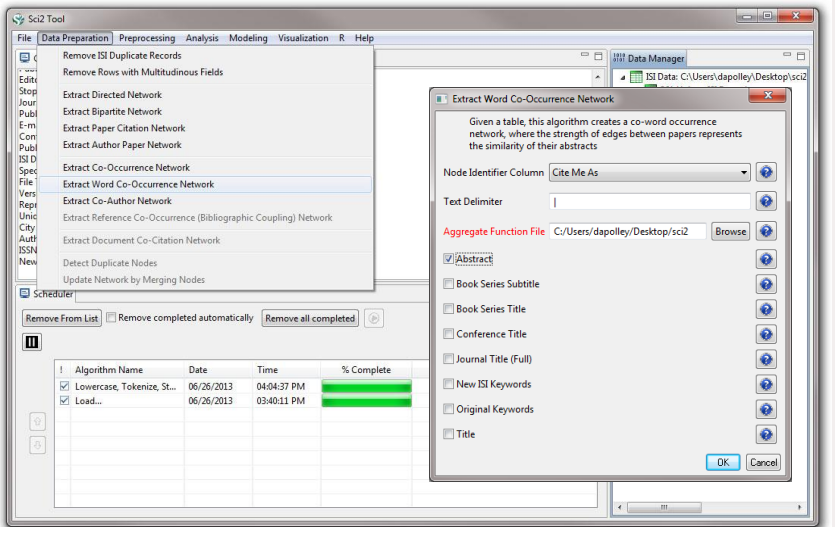
Word Co-Occurrence Network

Normalize the text of the abstract *Preprocessing* > *Topical* > *Lowercase, Tokenize, Stem, and Stopword Text*



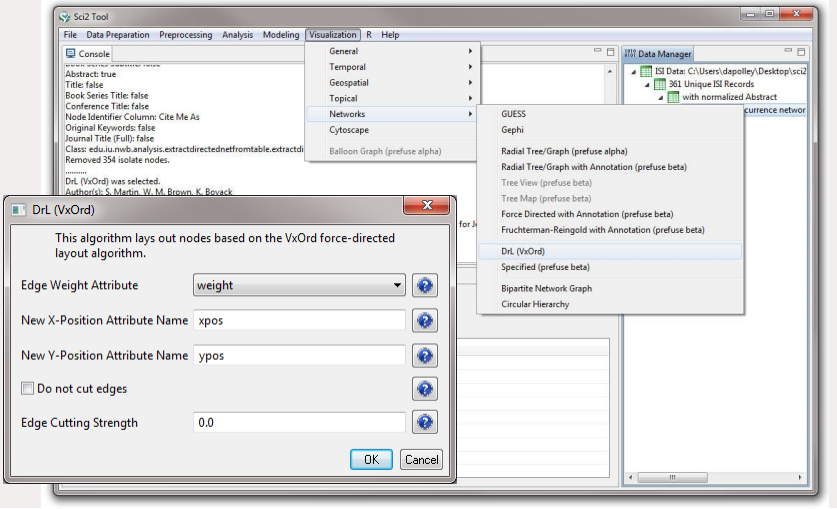
Word Co-Occurrence Network

Create the word co-occurrence network *Data Preparation > Extract Word Co-Occurrence Network*



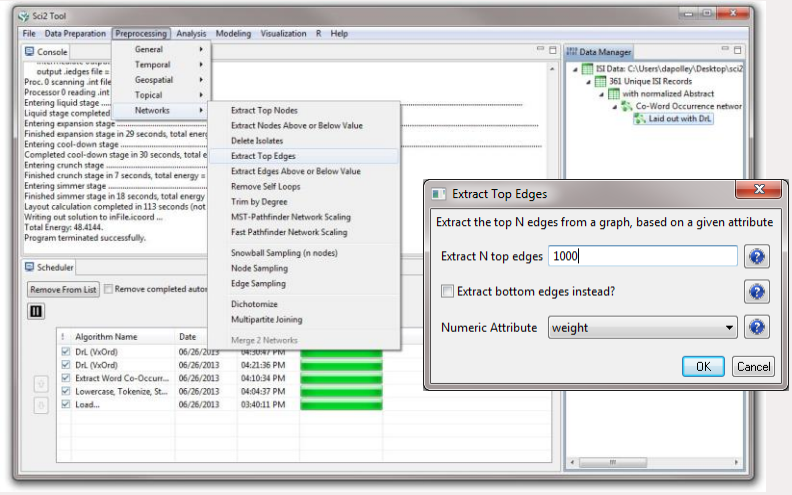
Word Co-Occurrence Network

Apply *Visualization > Networks > DrL (VxOrd)* and words that are similar will be plotted relatively close to each other.



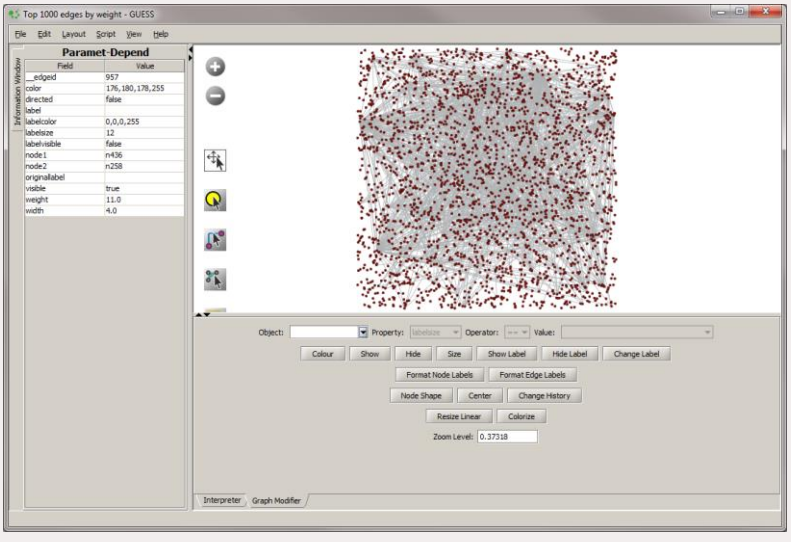
Word Co-Occurrence Network

Laying out the network with Drl (VxOrd) may take some time, but once the algorithm is complete you will want to keep only the strongest edges, so select the "Laid out with DrL" and select *Preprocessing > Networks > Extract Top Edges*



Word Co-Occurrence Network

Once edges have been removed, the network "top 1000 edges by weight" can be visualized by running *Visualization > Networks > GUESS.*



Word Co-Occurrence Network

In order to make use of the DrL (VxOrd) force directed layout we applied, we need to change to the interpreter at the bottom of the screen and type in the following commands:

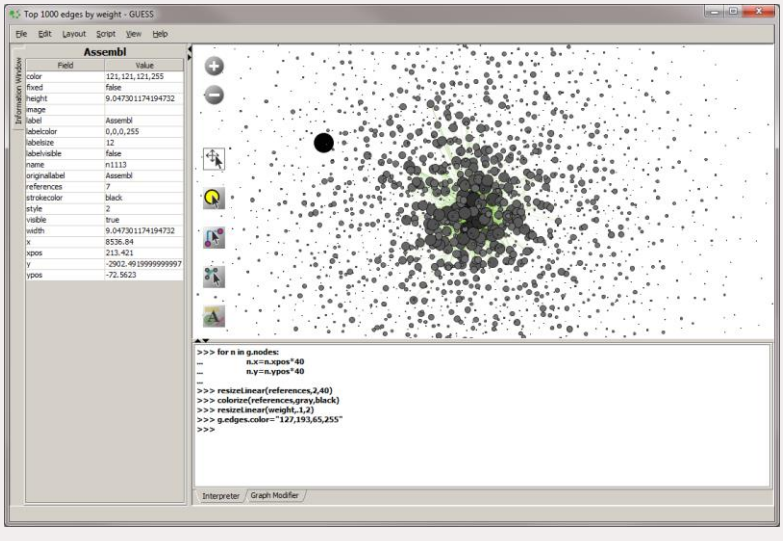
```
>>> for n in g.nodes:
...             n.x = n.xpos*40
...             n.y = n.ypos*40
...
>>> resizeLinear(references,2,40)
>>> colorize(references,gray,black)
>>> resizeLinear(weight,.1,2)
>>> g.edges.color = "127,193,65,255"
>>>
```

Interpreter | Graph Modifier

93

Word Co-Occurrence Network

Note, GUESS will not necessarily display the graph in the middle of the screen, you may have to scroll around the screen to find the graph.



94

Questions?