

"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



CNS Cyberinfrastructure for Network Science Center

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 <u>http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets</u>
- 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 <u>http://www.java.com/en/download</u>. 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 <u>http://ps2pdf.com</u> and a free PDF Viewer at <u>http://www.adobe.com/products/reader.html</u>.



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 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

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

SUMMARY

OPINION

nature

commercial purposes.

NS Cyberinfrastructure for Network Science Center

> 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

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

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

Vol 464|25 March 2010

- New Features
- Support new Web of Science format from ISI

Cyberinfrastructure for Network Science Center

- 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



Major Release

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

New Features

Google Scholar citation reader

Cyberinfrastructure for Network Science Center

- 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

Cyberinfrastructure for Network Science Center

- Bing Geocoder
- Flow map visualization, see below
- Comprehensive online documentation



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.



S Cyberinfrastructure for

Microscopes



Telescopes



Macroscopes

9



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

Sci2 Infrastructure

NS Cyberinfrastructure for Network Science Center

Shell (CIShell)

- CIShell (<u>http://cishell.org</u>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools
- It extends the Open Services Gateway Initiative (OSGi) (<u>http://osgi.org</u>), a standardized, modularized service platform
- CIShell provides "sockets" into which algorithms, tools, and datasets can be plugged using a wizard-driven process



CNS Cyberinfrastructure for Network Science Center

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

Suberinfrastructure for Network Science Center 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, all of science
Temporal Analysis (When)	Funding portfolio of one individual	n opic bursts i 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 f
Network Analysis (With Whom?)	NSF (ork of	netwo	NIH's core c



Sci² Tool – Supported Tools



Gnuplot

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



GUESS

exploratory data analysis and visualization tool for graphs and networks.

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

15



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



Sci² Tool – Bridged Tools

Gephi visualization tool bridging

NS Cyberinfrastructure for Network Science Center



Sci2 Tool Interface Components See also <u>http://sci2.wiki.cns.iu.edu/2.2+User+Interface</u>

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.

ile Da	ta Preparation Preprocessing	Analysis Mo	deling Visualizati	on Help			
🛛 Con	sole					ll Data Man	ager 🛄
'he dev letwori Iniversi	te to the Science of Science Tool elopment of this tool is supporter (Science center and the School o ty, the National Science Foundati 303, and the James S. McDonnell	d in part by the f Library and Inf on under Grant	formation Science No. SBE-0738111 a	at Indiana nd	* III	🍄 Direc	tory Tree - Pr
Cyberin Irimary Itrategi Iatrick / Isoll do Sche Remo	frastructure portal (http://sci.slis. investigators are Katy Börner, Ind es Inc. The Sci ² tool was develope A. Phillips, Chintan Tank, and Jose the/fcichell.org). developed at the	indiana.edu) fo liana University d by Micah W. I sph Biberstine. I a Cohorinfractur	and Kevin W. Boya .innemeier, Russell t uses the Cyberinf .churn for Notoroti	ck, SciTech J. Duhon, rastructure	•		
Syberin Irimary Itrategi Iatrick / Isoll (b)	frastructure portal (http://sci.slis. investigators are Katy Börner, Ind es Inc. The Sci ² tool was develope A. Phillips, Chintan Tank, and Jos the //ci.ball.org). developed at the /duler	indiana.edu) fo liana University d by Micah W. I sph Biberstine. I a Cohorinfractur	and Kevin W. Boya .innemeier, Russell t uses the Cyberinf .churn for Notoroti	ck, SciTech J. Duhon, rastructure	•		

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

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



Cuberinfrastructure for Network Science Center *http://wiki.cms.iu.edu/display/CISHELL/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 <u>Geocoder</u> 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.
- To use Bing Geocoder, user has to obtain an API Keys from <u>Bing Maps</u>. 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 <u>Geospatial Visualization</u>). User can obtain the geographical coordinates (Latitude and Longitude values) and feed them to the visualization plugin.

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

23

	Scyberinfrastructure for Network Science Center	Using B	ing	Geod	coder		
	Se Sci2 Tool						
	File Data Preparation Preprocessing	Analysis Modeling	Visualiz	ation R	Help		
	E Console	Temporal	• [1010 Data Manager	· · · ·	
	Load was selected. Documentation: http://wiki.cns.iu.edu/ Found old-style ISI/Web Of Knowledge f	Geospatial Topical Networks	•	Generic G Bing Geo		:\Users\Michael\Desktop\SCI2 1.1b\sampledat ue ISI Records	a
	The original 5 records have been process leaving 5 records. 5 records did not have unique IDs (specif we were unable to determine whether th	ied with the UT tag in I	ISI format), so			
Bing Geocod	er		x	1			
longitudes) Internet. Eac	ace names to geographic coordina using Bing. Ensure your machine is h App Key is limited to 50,000 requ e may vary due to network latency;	connected to the				ave Application ID bla ses, but for heavy use,	
Bing App key			0	fo	r your ov	vn personal Bing App	lication
Place Type	Address	•	0		, see: tps://ww	w.bingmapsportal.com	n
Place Name Co	lumn Abstract	•	0				_
🔲 Include addr	ress details	ſ	0				

OK Cancel

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

	А	В	С	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

CNS Cuberinfrastructure for Network Science Center

Data Manager	·\sci2\sampledata\g	eo\usptoInfluenza.csv
EIII COVINC. V	, (seiz (sumpredute (ge	co (disproximacinzales)

	PCN 2ate Data wa	S Cyberinf Network	rastructure for Science Center	Agg	regate by	y Count	try		Preprocessin General Temporal Geospatia	Extract To Extract To	p N% Records p N Records
0 0	nenter(s): Ch								Topical	•	
			s.iu.edu/disp	lav/CISHELL	/Aggregate+D	Data		🔳 Ag	gregate Data		
	Parameters:	1	· 1						Aggregate data i	in the table based o	on a column.
gre	gate on colur	nn: Country									
	ter for Cour							Aggre	gate on column	Icountry	
	ude: AVERA							Times	Cited	Sum	(
	le: AVERAG	ΈE								,	
	Cited: SUM							Latitu	de	Average	_ <u> </u>
					e to no non-null, lue to no non-nu			Longit	hude	Average	–
0 0			0	11		ill, non-empt	y values.	Longi	ude	Average	
que	ency or uniqu	ue Country v	values added to	o "Count" colun	10.			Delimi	ter for Country	1	(
					_	_					OK Can
	A	В	C	D	_	A	_	B	С	D	E
1	A Times Cited	-	C Longitude	D Country		Times Cite	d Latitud	le	Longitude	Country	E Count
_		Latitude	-	Country	2	Times Cite	d Latitud 0 [41.10	le	Longitude		E Count 194
2	Times Cited	Latitude	Longitude	Country	2	Times Cite 1468 180	d Latitud 10 [41.10	- le 645f]	Longitude [-82.45309f]	Country United States	E Count 194 57
2 3	Times Cited 7	Latitude	Longitude	Country	2 3 4	Times Cite	d Latitud 0 [41.10 12 18 [47.50	- le 645f] 6226f]	Longitude [-82.45309f] [19.06482f]	Country United States Hungary	E Count 194 57 14
2 3 4	Times Cited 7 0	Latitude 42.02946091	Longitude	Country United States	2 3 4 5	Times Cite 1468 180 39	d Latitud 0 [41.10 2 8 [47.50 1 [37.25]	- ie 545f] 5226f] 198f]	Longitude [-82.45309f] [19.06482f] [127.08451f]	Country United States Hungary South Korea	E Count 194 57 14 4
2 3 4 5	Times Cited 7 0 0	Latitude 42.02946091 42.34999466	Longitude -87.68838501 -71.08765411	Country United States	2 3 4 5 6	Times Cite 1468 180 39 10	d Latitud (41.10) (2) (47.50) (1) (37.25) (32.08)	- le 545f] 5226f] 198f] 439f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f]	Country United States Hungary South Korea Israel	E Count 194 57 14 4 1
2 3 4 5 6	Times Cited 7 0 0 2	Latitude 42.02946091 42.34999466	Longitude -87.68838501 -71.08765411 -86.23918915	Country United States United States	2 3 4 5 6 7	Times Cite 1468 180 39 10 10	 Latitud (41.10) (41.10) (47.50) (37.25) (32.08) (46.76) 	le 545f] 5226f] 198f] 439f] 8517f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f]	Country United States Hungary South Korea Israel Romania	E Count 194 57 14 4 1 1 2
2 3 4 5 6 7	Fimes Cited 7 0 0 2 14	Latitude 42.02946091 42.34999466 41.70074844 41.70074844	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915	Country United States United States United States	2 3 4 5 6 7 8	Times Cite 1468 180 39 10 10 10 10 10 10 10 10 10 10 10 10 10	 Latitud [41.10] [41.50] [47.50] [37.25] [37.25] [32.08] [46.76] [46.76] [47.06] 	le 545f] 5226f] 198f] 439f] 8517f] 515f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f]	Country United States Hungary South Korea Israel Romania Switzerland	E Count 194 57 14 4 4 1 2 2 2
2 3 4 5 6 7 8	Fimes Cited 7 0 0 2 14 15	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.89422607	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915 -87.61901855	Country United States United States United States United States	2 3 4 5 6 7 7 8 9 9	Times Cite 1468 180 39 10 10 10 5 5 5 45	 d Latitud (41.10) (47.50) (47.50) (37.25) (37.25) (32.08) (32.08) (46.76) (47.06) (47.97) 	le 545f] 5226f] 198f] 439f] 8517f] 515f] 7184f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f]	Country United States Hungary South Korea Israel Romania Switzerland France	E Count 194 57 14 4 4 1 2 2 2 2 2 12
2 3 4 5 6 7 8 9	Times Cited 7 0 2 14 15 29	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.89422607 41.70074844	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915 -87.61901855 -86.23918915	Country United States United States United States United States United States	2 3 4 5 6 7 7 8 9 9 10	Times Cite 1468 186 395 100 11 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	 Latitud Latitud [41.10] [47.50] [37.25] [37.25] [37.25] [37.25] [37.25] [37.25] [46.76] [46.76] [47.06] [47.97] [52.15] 	le 645f] 6226f] 198f] 439f] 8517f] 615f] 7184f] 457f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands	E Count 194 57 14 4 1 2 2 2 2 2 12 5
2 3 4 5 6 7 7 8 8 9 10	Times Cited 7 0 2 14 15 29 32	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.89422607 41.70074844	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915 -87.61901855 -86.23918915 -86.23918915	Country United States United States United States United States United States United States	2 3 4 5 6 7 8 9 9 1(1)	Times Cite 1468 188 180 10 10 1 1 2 2 4 5 2 4 5 2 4 5 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 3 3 3 5 10 10 5 1 4 5 8 1458 18 18 18 18 18 18 18 18 18 18 18 18 18	Latitud 41.100 41.100 12 14 15 167.061 17 18 19 11 137.251 18 137.251 18 137.251 137.251 14 147.061 15 147.061 15 147.061 15 147.061 14 149.944	de 645f] 6226f] 198f] 439f] 8517f] 615f] 7184f] 457f] 4717f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f] [84.528114f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands Russia	E Count 194 57 14 4 1 2 2 2 2 2 2 2 5 5 2
2 3 4 5 6 7 8 9 10 11	Times Cited 7 0 0 2 14 15 29 32 32 7	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.89422607 41.70074844 41.70074844 41.70074844	Longitude -87.68838501 -71.08765411 -86.23918915 -87.61901855 -86.23918915 -86.23918915 -86.23918915	Country United States United States United States United States United States United States	2 3 4 5 6 7 7 8 9 9 10 1 1 1	Times Cite 1468 180 39 10 10 5 5 5 6 45 6 6 5 1 2 111	Latitud 41.100 41.100 12 18 47.500 11 37.252 8 32.084 7 46.766 5 47.060 5 47.971 2 49.944 49.944 2 41.542	de 645f] 6226f] 198f] 439f] 8517f] 615f] 7184f] 457f] 4717f] 5982f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f] [84.528114f] [1.7138832f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands Russia Spain	E Count 194 57 14 4 4 1 2 2 2 2 2 2 2 2 2 2 2 3 2 3 3
2 3 4 5 6 7 8 9 10 11 12	Times Cited 7 0 0 2 14 15 29 32 32 7 5	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.89422607 41.70074844 41.70074844 41.70074844	Longitude -87.68838501 -71.08765411 -86.23918915 -87.61901855 -86.23918915 -86.23918915 -86.23918915 -86.23918915	Country United States United States United States United States United States United States United States United States	2 3 4 5 6 7 7 8 9 9 10 11 12 11	Times Cite 1468 180 39 10 5 5 5 6 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	d Latitud 00 [41.10] 12 1 18 [47.50] 18 [47.50] 19 [37.25] 8 [32.08] 17 [46.76] 15 [47.06] 15 [47.97] 12 [52.15] 12 [49.94] 12 [41.54]	de 645f] 645f] 5226f] 198f] 439f] 8517f] 615f] 7184f] 457f] 5982f] 2654f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f] [84.528114f] [1.7138832f] [12.727126f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands Russia Spain Italy	E Count 194 57 14 4 1 2 2 2 2 2 12 5 2 2 13 46
2 3 4 5 6 7 8 8 9 10 11 12 13	Times Cited 7 0 2 14 15 29 32 32 7 5 2 2	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.70074844 41.70074844 41.70074844 41.70074844 41.01500168 47.50622177	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -85.73377991 19.06481934	Country United States United States United States United States United States United States United States United States	2 3 4 5 6 7 7 9 10 12 12 12 12 12	Times Cite 1468 180 35 10 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	d Latitud 00 [41.10] 12 1 18 [47.50] 18 [47.50] 19 [37.25] 8 [32.08] 17 [46.76] 15 [47.06] 15 [47.97] 12 [52.15] 12 [49.94] 12 [41.54]	de 645f] 645f] 5226f] 198f] 439f] 8517f] 615f] 7184f] 457f] 5982f] 2654f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f] [84.528114f] [1.7138832f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands Russia Spain	E Count 194 57 14 4 1 2 2 2 2 2 2 2 2 2 12 5 5 2 13 3 46 6 3
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 	Times Cited 7 0 2 14 15 29 32 32 7 5 2 2 10	Latitude 42.02946091 42.34999466 41.70074844 41.70074844 41.70074844 41.70074844 41.70074844 41.70074844 41.01500168 47.50622177	Longitude -87.68838501 -71.08765411 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -86.23918915 -85.73377991 19.06441934 -86.23918915	Country United States United States United States United States United States United States United States United States United States United States	2 3 4 5 6 7 7 8 9 9 10 11 12 11	Times Cite 1468 180 10 10 10 10 10 10 10 10 10 10 10 10 10	d Latitud 00 [41.10] 12 1 18 [47.50] 18 [47.50] 19 [37.25] 8 [32.08] 17 [46.76] 15 [47.06] 15 [47.97] 12 [52.15] 12 [49.94] 12 [41.54]	de 645f] 6226f] 198f] 439f] 8517f] 615f] 7184f] 457f] 4717f] 5982f] 2654f] 44667f]	Longitude [-82.45309f] [19.06482f] [127.08451f] [34.81297f] [23.585135f] [7.2015657f] [2.2232702f] [4.49463f] [84.528114f] [1.7138832f] [12.727126f]	Country United States Hungary South Korea Israel Romania Switzerland France Netherlands Russia Spain Italy Brazil	E Count 194 57 14 4 1 2 2 2 2 2 12 5 2 2 13 46

S CNS	Cuberinfrastructure for Network Science Center	Channel at Man
Visualization R Hel	p	Choropleth Map
General Temporal		Color-codes the name proportion to associate
Geospatial	Proportional Symbol Map	
Topical	Choropleth Map	Subtitle Generated from CSV
Networks	 Geospatial Network Layout with Base Map 	Map World
PostScript fi	nd Save map as le. Use PostScript onvert to pdf to	Color By Times Cited Color Scaling Logarithmic Color Range Vellow to Blue
	1000 Data Manager	
	CSV file: C:\sci2\sampledata\geo\u Geospatial Visualization (Choro Save	
		2



			Proportional Symbol I	Мар
				s geospatial coordinates as circ r-coded in proportion to associ
ualization R Help			Subtitle	Generated from CSV file: C:\s
General 🕨			Мар	World
Temporal Geospatial	Proportional Symbol Map		Latitude	Latitude
Topical 🕨	Choropleth Map		Longitude	Longitude
Networks •	Geospatial Network Layout with	Base Map	Size Circles By	Times Cited
			Size Scaling	Logarithmic
				None (uniform coloring)
			Exterior Color Scaling	Linear
			Exterior Color Range	White to Green
			· · · · · ·	
			Interior Color Scaling	Linear
ight-click an	d Save map as	· · · · · · · · · · · · · · · · · · ·	Interior Color Range	White to Green
0	1	1010 Data Manager		
. *	. Use PostScript	CSV file: C:\	sci2\sampledata\geo	\usptoInfluenza.csv
newer or cor	overt to pdf to			portional Symbol Map)
ew.			ave	
				1







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

33

CNS Cuberinfrastructure for Network Science Center 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.)





- 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 <u>Joda Time</u> library extensively



CNS Cyberinfrastructure for Network Science Center Horizontal line graph of NSF projects

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

6F - Amard Search - Search All Fields - Windows Internet Explorer			NSF - Award Search - Search All Fields -1					_
U 12 12 12 1	🖬 👍 🗙 Live Search	P-	COO - O http://www.nd.gov/awardse	ardV		🕂 🗙 Une S	Search	
Edit Yow Favorites Iools Help 🔍 Convert - 🔊 Se			Elle Edit Yew Favorites Iools Help		🍕 Convert 👻 💽 Selec			
🐼 😳 🔹 🛞 NSE - Award Search - Se 🗙 💽 http://www.ref.gov/awards	🚹 • 🗟 - 🖶 • 🕑 Bagi	je = 🍈 Tgols = 🦥	😪 🐼 😁 🔹 💿 NSE - Award Search - Se	🗙 💽 htt	p://www.nsf.gov/awards	🔄 🔂 • 🛙	🗟 🔹 👼 🔹 🔂 Bage	• 🌀 Tgols
National Science Foundation	SEARCH NSF Web Ste	PastLane	Historical Awards C Active Awards Char Espired Awards C	ch	Reset			
			Search Results					
Award Search	Send Commenta Award :	Search Help	Back Results are sorted by award date, with the	most recent av	ards at the top. Click on a colum	in heading to re-so	ort the results.	
Awardee Information Program Information Search All Proc. Text	Search All Fields More Options		The up/down arrows at the right of each co	lumn title contr	ol whether the sort is ascending	or descending.		
	4		To view the a					
			619 anards 6 Satze 110	CSV	formatia	c *inc	titution	* 110
Hasts The search field balow "Search Award For" searches the tolic, abstract, and non- Search Award Ears Restrict to Title Onlys	ard number fields.			CSV	' format a	s *ins	titution	*.nsj
Search Annual For Restrict to Table Only: P Avandee Laformation	ard number fields.		(rist/rev) Annel Mannel Od20602 Technolog for Taccher (Kr. Od20602 Technolog for Taccher (Kr.	CSV	CCLI-Phase 2 (Expansion)		eren, bisihine	*.nsj
Inexh Annel Eas Inexh Annel Eas Inexh Annel Eas Annel Ease Inexh Annel Ease Inexh Ease Inexh Ease Inexh Ease Inexh	ard number fields.		100 and 1 [rest/may] Save in	83	DOLL-PLACE 2 (Excension), E-STEM SOLA BOLTECH ENSEMATH	02/01/2009	eren, bisihine	
Search Award Fav	ard number fields. PE Look	kup	Barendo (Institute) Barendo Ba	83	CCLI-Phase 2 (Expansion)	02/01/2009	eren, bisihine	
Inanth-Anned Ean Teacher Lis Fride Chinks		kup	Storenship (rest/iver) Sates S	RCS RVB	SOLIAL PERCHISION OCLI-Phase 2 (Possion) ENTEMISOR & SCITECH ENGENATE SCIENCE TECH & SOCIETY	02/01/2009 01/01/2009 01/01/2009	Basa, Human Matzi, Jonathan	HI HI
Sanch Anned Fare Enditor to Table Date		kup	19 method (received) Save in Mania 202022 million circles 202022 million circles 20202 million c	RCS RVB	COLI-Parae 2 (Pagasien) STATE SCHERK SCHECH ENGEMATH SCIENCE TECH 5	02/01/2009	Basa, Human Matzi, Jonathan	ы ы
Sauch Annel fan			Land the second provides of the second provid	828 R48 585 CH181	ECCAN PERFORMANCE OCLU-PERFORMANCE ENTERFORMENT FOR ENTERFORMENT FOR ENTERFORMENT ENTERFORMENT ENTERFORMENT ECCENT MANUFACTURING B CONST MACHEOP	02/01/2009 01/01/2009 01/01/2009 01/01/2009	Broot, Maria Inte Bass, Humas Metzl, Joseftan Shih, Albert	RI RI RI
tauch Anned Farr Textils the Table table T Proceeding to	PI Looi		Save in Save i	803 RVA 585	DOLIAL PERCENDION COLI-Phase 2 (Extension) E-FEMILONA SCITECH E-SCIENCE TECH & EOCIETY MANUFACTURING &	02/01/2009 01/01/2009 01/01/2009 01/01/2009	Basa, Human Matzi, Jonathan	ы ы ы
Isonih Anna Fare Institut A Talk Calve Anna Calve Anna Calve Anna Calve Anna Calve Anna	PI Looi		Barry Constraints and a second	828 R48 585 CH181	ECCAN PERFORMANCE OCLU-PERFORMANCE ENTERFORMENT FOR ENTERFORMENT FOR ENTERFORMENT ENTERFORMENT ENTERFORMENT ECCENT MANUFACTURING B CONST MACHEOP	02/01/2009 01/01/2009 01/01/2009 01/01/2009	Rana, Haman Rana, Haman Matul, Josephan Shih, Albara Uliroy, A. Galig	RI RI RI
Isanh Anna Far Tentistic Talle allow C Annada traditional Policy Journalized Test Annae High Looking CO-H (H reach to show searches) Jackda (2027) Discriticitum Di	PI Looi	n Lookup	Save in Save in Sav	253 255 555 555 555	COLLAGA PELOPOLICA COLLABOR 2 (Constant) PETER COLLA SCITICO PARAMATE SCIENCE TECHA SCIENCE TECHA SCIENCE TECHA SCIENCE TECHA SCIENCE TECHA SCIENCE SC	02/01/2009 01/01/2009 01/01/2009 01/01/2009 01/01/2009	Rana, Haman Rana, Haman Matul, Josephan Shih, Albara Uliroy, A. Galig	MI MI MI MI
Isanh Anna Far Tentistic Talle allow C Annada traditional Policy Journalized Test Annae High Looking CO-H (H reach to show searches) Jackda (2027) Discriticitum Di	PI Looi		Save in intervention Save in Save in	253 255 555 555 555	Excitic Second Color Color Page 2 (Second Color Extra Social Color Extra Social Color Excitic Color Excitic Color Color Social Social Color Social C	02/01/2009 01/01/2009 01/01/2009 01/01/2009 01/01/2009	Room, Balantara Rasa, Human Maszi, Janashan Shih, Albest Ulaoy, A. Salig Na, Jan	MI MI MI MI













CNS Cuberinfrastructure for Temporal bar graph of NSF projects







	Sci2 Tool	
	File Data Preparation Preprocessing Analysis Modeling Visualization R Help	
	🖾 Console 📃 🗖	🕴 🖓 Data Manager 👘 🗖
	Science***/2,978-0-80789-869-7,8YO25,222,CIVIC PASSIONS: SEVEN WHO LAUNCHED PROGRESSIVE AMERICA (AND WHAT THEY TEACH US Lengish,27,41,27,28,2009,INIV NORTH CARCILIAN PRESS BOX 288, CHAPEL HILL, INC. 27314 USA,748,2016.https://u.ku/u.ku/u.ku/u.ku/u.ku/u.ku/u.ku/u.	ISI Data: C:\Users\Michael\Desktop\SC
Load <i>1406.</i> <i>LOC.isi</i>	OS40 USA,Lib Congress, Washington, DC 20540 USA,Histon/Political werk & Law,70,The Social Ose1,0,2,7;C(User;Michael/Dektop)SCI2 I-ARL-OCia;J () as no jounal column (9). The 'Cite Me As' field may ve been processed to remove duplicate unique ISI IDs leaving 2683	
Load this file ARL sample	Local\Temp\isiduplicateremoverlog5664843454329578764.bxt	< <u> </u>
	Scheduler	🗘 Workflow Manager 📃 🗆
	Remove From List Remove completed automatically Remove all completed	Workflow 1
	! Algorithm Name Date Time % Complete	
	8	
		<u>46</u>

CNS Cuberinfrastructure for Network Science Center

Temporal Analysis: Burst Detection

Select Preprocessing > Topical > Lowercasc, Stem, and Stopword Text Then select Title from the input parameters Date Time Using III Removed III Processing Stopword Text				A 17.27.11.			- D X
Console General Femporal MiData Manager Femporal Midata Midata Midata Midata Midata Midata Manes Concerses, Manager Midata Manager Select Preprocessing > Topical > Lowercase, Altribute Midata Mill Midata Mall Midata Mill							
Science**.27978-0073 Temporal ASSIONS: SEVEN WHO LAUNCH ED PROGESSIVE AMERICA Geospatial ASSIONS: SEVEN WHO LAUNCH ED PROGESSIVE AMERICA Geospatial ASSIONS: SEVEN WHO LAUNCH ED Congress, Washington, Congress, Washington, Reconciled Journal Names Congress, Washington, Lowercase, Tokenice, Stem, and Stopword Text Select. Preprocessing > Topical > Lowercase, Tokenice, Stem, and Research Field Topical > Lowercase, Tokenice, Stem, and Besearch Field Stopword Text Besearch Field Then selecct Title from Itime the input parameters Date Version Number Version Number Version Number Version Number	· · · · · · · · · · · · · · · · · · ·			deling Visualization			
Select Pressiones Stones Stone	Console				8		
CAROLINA PERSION Topical Reconciled Journal Names Compress, Vashing Compress, Vashing Stopword Text Lowercase, Tokenize, stem, and Stopword Text Select Preprocessing > Topical > Lowercase, Tokenize, Stem, and esset to remove duplicate unique isiduplicateremovelog5648434st 11b/sampledata/ARL140611-ARL-LOC.isi/ Namo Journal Column (9). The Creat Matex Lipid room Image: Column (9). The Creat Matex Lipid room Select Preprocessing > Topical > Lowercase, Tokenize, Stem, and esset to remove duplicate unique isiduplicateremovelog5648434st 11b/sampledata/ARL140611-ARL-LOC.isi/ Namo Journal Column (9). The Creat Matex Lipid room Image: Column (9). The Creat Matex Lipid room Stopword Text esset to remove duplicate unique isiduplicateremovelog5648434st 11b/sampledata/ARL140611-ARL Research Field Image: Column (9). The Creat Matex Lipid room Topical > Lowercase, Tokenize, Stem, and subject Category Image: Column (9). The Date Image: Column (9). The Date Image: Column (9). The Date Then select Title from the input parameters Date Image: Column (9). The Date Image: Column (9). The Creat Matex			PASS				
Congress, Washington, Gorpeld, Wicking, J. History, Gov Networks Lowercase, Tokenize, Stem, and Stopword Text Listampledata/ARL14001-ARL-IC-CLSI() has no jound column (9). The Cite Me Ac fuel mov Isomercase. Tokenize, Stem, and Stopword Text Select Preprocessing > Topical > Lowercase, Stem, and Stopword Text Isomercase. Tokenize, Stem, and Stopword Text Isomercase, Stem, and Stopword Text Isomercase. Tokenize, Stem, and Subject Category Then select Title from the input parameters Date Isomercase Time Version Number Version Number Isomercase. Isomercase. Isomercase. Isomercase.			037,			2683 Unique ISI Re	cords
Select Preprocessing > Topical > Lowercase, Tokenize, Stem, and Stopword Text esset to remove duplicate unique isiduplicateremoverlog5664813454 11b/sampledata/ARL140611-ARL-10C/ast, 7han of good Then select Title from the input parameters ted automatically Removeallo Unique D Subject Category Date Time Vision Number Vision Number Vision Number		•					
Iberivalid Isoverase, Tokenize, Stem, and Stopword Text Select Preprocessing > Topical > Lowercase, Tokenize, Stem, and Stopword Text Isoverase, Tokenize, Stem, and Stopword Text Then select Title from the input parameters Isoverase, Tokenize, Stem, and Stopword Text	Gospel.0, WOS:000299487	000008.1.0.7.7.C:\U	sers\Michael	Desktop\SCI2			
Select Preprocessing > Topical > Lowercase, Tokenize, Stem, and Stopword Text Ibitampledat\ARU140611-ARL Subject Category isduplicateremoverlog566481454 Subject Category isduplicateremoverlog566481454 Then select Title from Unique ID Date Vision Number Vision Number Volume Cite Me As		0611-ARL-LOC.isi,	?] has no jou	nal column (J9). The	Cite Me As' field may Lowercase, Tokenize, Stem,	and Stopword Text	
Topical > Lowercase, Tokenize, Stem, and Stopword Text induplicateremovelog5648434st 11bisampledataARL140611-ARL Researcher ID Subject Category Subject Category Subject Category Subject Category Then select Title from the input parameters Date Unge ID Date Time Version Number Image: State of the input parameters Date Image: State of the input parameters Image: State of the input parameters Date Image: State of the input parameters Image: State of the input parameters Date Image: State of the input parameters Image: State of the input parameters Date Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters Image: State of the input parameters	Select Prenrocess	ina > "es	sed to remov	e duplicate unique			
Tokenize, Stem, and Stopword Text 11bisampledataARU40611-ARI subject Category Subject Category Then select Title from the input parameters Date Image: Displement Urge D Image: Displement Image: Displement Image: Displement Image: Displement Image: Displement<							
Stopword Text Subject Category Subject Category Subject Category Then select Title from the input parameters Unique ID Date Time Version Number Subject Category Image: Constraint of the input parameters Image: Constraint of the input parameters Image: Constraint of the input parameters Date Image: Constraint of the input parameters Image: Constraint of the input parameters	Topical > Lowerca				Researcher ID		
Stopword Text Subject Category ted automatically Subject Category Then select Title from the input parameters Time Date Time Vurine Pate Image: Subject Category Image: Subject Category	Tokenize, Stem, a	nd ¹¹	1b\sampledat	a\ARL\140611-ARL	Special Issue		•
ted automatically Remove all column Then select Title from the input parameters Date Date Time Vulume Vulume Image: State of the second	· · · · · · · · · · · · · · · · · · ·	-			Subject Category		•
Then select Title from the input parameters Date Time Version Number Volume Cite Me As	Stopword Text				Supplement		•
Then select Title from the input parameters Unique ID Date Time Version Number Image: Display the input parameters Image: Display the input parameters Date Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Date Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters Image: Display the input parameters		te	d automatica	Ily Remove all co	Title		0
Image: Second	Then select Title	from			🕅 Unique ID		
Volume Image: Crite Me As	the input paramet	ers a	Date	Time	Wersion Number		•
Image: Second se	and input parameter	0.0			🕅 Volume		0
Cite Me As					File Name		
					Cite Me As		
4Z							
47							OK Cancel
<u>47</u>							

CNS Cuberinfrastructure for Network Science Center

Temporal Analysis: Burst Detection

🛠 Sci2 Tool					
File Data Preparation Preprocessing	nalysis Modeling Visualizatio	on R Help			
Console	and the second	urst Detection		🗖 🗖 👯 Data Manager	- 8
Connectice Duck note Issue: False Publication Date: False Editors: False Stopword List: C2/Users/dapolley/Deskto _{wer}	Geospatial Topical Networks ords.bt		Burst Detection		Jsers\dapolley\Deskti
Journal Name (Abbreviated): false Publication Type: false			Perform B	urst Detection on time-se	ries textual data.
Highlight the table 'with	n		Gamma	1.0	
normalized Title'			Density Scaling	2.0	
			Bursting States	1	
Select Analysis >			Date Column	Publication Year	•
Temporal > Burst			Date Format		
Detection			Burst Length Unit	Vears	•
	tomatically Remove all co	mpleted	Burst Length	1	
Then set the			Durst Length		
parameters to what is	Time (2012 11:25:40 AM	% Complete	Text Column	Title	•][
shown to the right			Text Separator	1	
			Document Column	No document column.	•
			Ignore Input With	n Empty Text	[
					OK Ca



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.

CNS Cuberinfrastructure for Temporal Analysis: Burst Detection

			1	A											X-)
				🎭 Sci2 To											
				File Dat	a Preparatio	on Prepro	cessing A	Analysis Mod	eling	Visualiza	tion R	Help			
				Conse									- 8	1919 Data Manager	
				servic sta inform st global sta public sta famili sta english st guid start	rts: 1955 en arts: 1985 e arts: 2000 er arts: 1973 er rts: 1994 en tarts: 1994 en tarts: 1994 en ts: 1994 en ts: 1994 end	ds: 1980 nds: 1991 nds: nds: 1976 ds: 2000 ends: 1994 ls: 1997							*	ISI Data: C:\Users\ 2683 Unique IS with norm	SI Records
1	А	В		С	D	E	F								Rename
1	Word	Level		Weight	Length	Start	End								Discard
2	religion		1	9.504403	7	1999									
3	с		1	5.562619	12	1986	1997	7					+		
	librari		-	8.573344											,
	m			5.917634									- 8	🜔 Workflow Manager	- 0
	spanish			6.498343					lly Remove all completed			Distance in the			
	disk			5.661457				4			ea		Right-click on Burst		
8	polici		_	7.424719								detection analysis			
	paper		-	4.764948										detection a	lidiysis
	digit		-	4.856473				ite	Tim	e	9	% Compl	ete	(Publication	Vear Title
	bibliograp			5.817662										•	
	encyclope			5.539055										maximum b	ourst level 1
_	ethnomu	5		7.559989											
_	america			13.95819										the data ma	anager and
15	cultur		1	5.733256	14	1998	2011							view the file	a in the
					1									view the file	emule
				l	•								'	spreadshee	t program (
									_						
														your choice	,
														your choice	•



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.

24	A	В	С	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	В	С	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

<u>51</u>

CNS Cyberinfrastructure for Network Science Center

Temporal Analysis: Burst Detection

File Data Preparation Preprocessi	ng Analysis Modeling	y Visualiz	ation R Help			
Console			eneral 🔸	🗖 🗖 👬 Data Manar	ser 🗖 🗖	
english starts: 1994 ends: 1994 auid starts: 1994 ends: 1997		Ge	emporal +		Michael\Desktop\S 3 Unique ISI Records	
Select Visualization >			opical		with normalized Title Burst detection analysis (F	
Temporal > Temporal		198	etworks P	CSV fil	e: C:\Users\Michael\AppData\	
Bar Graph and set	1	Temporal	Bar Granh			×
	du/display/CISHELL/D	Subtitle	Takes tabu graph.	ular data and generates PostScr file: C:\Users\Michael\AppData	ipt for a temporal bar \Local\Temp\temp\Preprocessed-maxim	num bui
Scheduler		Start Date	Start			=
Remove From List Remove co	mpleted automatically	End Date	End			
		Size By Date Format	Weight Month-Day-Year Date	e Format (U.S., e.g. 10/15/2010)		
! Algorithm Name	Date	Category	No Category Coloring	9		
		V Scale Out	put?			
		Simplified Layout?				
					ОК	Cancel

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 <u>here</u>.

Temporal Visualization

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





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



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

Legend

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



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)





Topical Visualization

Generated from 381 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

4 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 CHEMITECH 1 COMBINATORIAL AND ALGORITHMIC ASPECTS OF NETWORKING 7 CURRENT COMMENTS 3 CURRENT CONTENTSUIFE SCIENCES 1 FEDERATION PROCEEDINGS 5 FRACTALS-AN INTERDISCIPLINARY JOURNAL ON THE COMPLEX GE... 1 FRONTIERS OF LIBRARIANSHIP-SYRACUSE UNIVERSITY

CNS (cns.iu.edu)



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.

В	С	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		E Scionc	o Man via EE4	Fields (Circle An	notations) 🗙	120
111 Conservation and Efficient Use of Water						245
112 Watershed Protection and Management		Locate UC	.SD area tagged	records on the UCS	SD Map of Science	245
121 Management of Range Resources		Subtitle	Prepro	cessed-USDA-Fund	s-FY2008.csv	520
122 Management and Control of Forest and Range Fires						520
123 Management and Sustainability of Forest Resources		UCSD Are	a UCSD Map	Field Name		231
124 Urban Forestry		Label	KNOWLED			231
125 Agroforestry		Laber	INNOWLED	JE AREA		231
Run Visualization > Topical > Science Map	via 554 Fi	elds ^{Value}	NO. Projec	ts		
using parameters given to the right.		Scaling Fa	ictor 1.0			

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

Simplified Layout? Show Export Window?





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 Cuberinfrastructure for Network Science Center

Introduction to Network Analysis







CNS Cuberinfrastructure for Visualizing the Florentine Dataset

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

CNS Cuberinfrastructure for Network Science Center

Visualizing the Florentine Dataset

71

CNS Cuberinfrastructure for Visualizing the Florentine Dataset

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

🕽 🔵 🗢 🕌 🔸 sci2 🔸	sampledata > socialscience	▼ 49	Search socialscience	م
Organize 👻 New fol	der		800 -	
☆ Favorites	Name	Date modified	Туре	Size
🧮 Desktop	florentine.nwb	11/27/2012 3:14 PM	NWB File	1 KB
🐌 Downloads	friendster.graphml.xml	11/27/2012 3:14 PM	XML Document	21 KB
🖳 Recent Places	friendster.xgmml.xml	11/27/2012 3:14 PM	XML Document	14 KB
	terror.graphml.xml	11/27/2012 3:14 PM	XML Document	16 KB
Ibraries Ibraries	terror.xgmml.xml	11/27/2012 3:14 PM	XML Document	12 KB
File	name: florentine.nwb	•	*.* Open I▼	▼ Cancel
Visualizing the Florentine Dataset

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

INS Cyberinfrastructure for Network Science Center

(83)

					- 0	110 Data Manager
e developm hool of Libr E-0738111 a ttp://sci2.w u need help tures. mary invest s develope Phillips, Ch e Cyberinfra re derived t	ray and Information 5C and IIS-0715303, and the viki.cns.iu.edu) for docc. p with your analyses, ha tigators are Katy Börner d by Chin Huak Kong, Jo initan Tank, and Russell d by Chin Huak Kong, Jo initan Tank, and Russell from the Network Work follows: 009). Science of Science	orted in part by the ience at Indiana Uni e James S. McDonn umentation and scr we questions about T. Indiana University sseph Biberstine, Th J. Duhon. It uses th J. Duhon. It uses th chench Tool (http://	versity, the Nationa ell Foundation. See te eenshots. Please visi datasets, or would I and Kevin W. Boyac mmas G. Smith, Dav c Cyberinfrastructu st/cms.iu.edu) at In (mvb.cms.iu.edu) at In (mvbristity and Scil	for Network Science center Science Foundation under Science Foundation under the Science Foundation under https://science.org/ documents/science.org/ documents/science.org/ documents/science.org/ fach Stategies, http://sci2. mpietes	and the Grant No. page er/ask.php if is and new e Sci2 tool eier, Patrick leveloped at ithm plugins	iifi Dita Manger
1	Algorithm Name	Date	Time	% Complete		
	Load	12/06/2012	10:21:57 AM			
₽						

UNS Cuberinfrastructure for 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.

 Station
 Constant
 Consta

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



 No.
 No.
 Operating of the state in the s

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

NS Cyberinfrastructure for Network Science Center

83

CNS Cuberinfrastructure for 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.

	Pucci	
Field color fixed height image label	Value	A
color	125, 12, 17, 255	
fixed	false	
height	10.0	
image		
label	Pucci	
labelcolor	0,0,0,255	
labelsize	12	• • • • • • • • • • • • • • • • • • •
labelvisible	false	♣
name	n11	
originallabel	Pucci	
priorates	0	
strokecolor	black	
style	2	
totalities	1	
visible	true	
wealth	3	
width	10.0	
×	65.12463556851313	•
Y	298.26822157434407	•
		Object: Property: Boddew Coperator:

75

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.

CNS Cyberinfrastructure for Network Science Center



NS Cuberinfrastructure for Visualizing the Florentine Dataset

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

Medi	ci-Barbadori	
Field Field business color directed label	Value	
edgeid	6	
business	т	
color	176,180,178,255	
directed	false	
label		
labelcolor	0,0,0,255	
labelsize	12	
labelvisible	false	
marriage	T	
node1	in8	
node2	n2	
originallabel		
visible	true	
weight	1.0	
width	4.0	e 💦
		Cóject: Property: dominan v Coperator: Colour Show Helde Size Show Label Helde Label O'Aunge Label Format Hode Labels, Format Edge Labels Node Shage Conter, Ourge Halory Resize Linew Coloree Zoom Level: 1.53421 Nodes y processes I from: To: Do Coloree Do Coloree
		Therester, Gaph Hodler

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.



CNS Cuberinfrastructure for 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.





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:

NS Cyberinfrastructure for Network Science Center

81

Cuberinfrastructure for 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.







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

85

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.

Cuberinfrastructure for Word Co-Occurrence Network

83

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*



NS Cuberinfrastructure for Word Co-Occurrence Network

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

file Data	Preparation Pr	reprocessing	Analysis I	Modeling Visuali	ization R Help
Consol		General	+	noucing visual	Date Manager
Kong, Jose the Cyberi	eph Biberstine infrastructure niversity. Man	Temporal Geospatial		Coe, Micah W. Lin sloped at the Cybe	And A state of the state o
	as follows:	Topical Networks		Lowercase, To	kenize, Stem, and Stopword Text
oad wa: Cound old Cound C	s selected. tation: http://wil i-style ISU/Web 01 -style ISU/Web 01 -style ISU/Web 01 -style ISU/Web 01 -style ISU/Web 01 al 361 records ha to C:\Users\dapolley' \Users\dapolley'	ki.cns.iu.edu/d Knowledge file Knowledge file Knowledge file Knowledge file Knowledge file we been proces billey\AppData\	isplay/CISH e. e. e. ssed to remo Local\Temp sampledata	IELL/Data+Forma ove duplicate uniq a\isiduplicateremc (scientometrics\is	Stopword List C/Users/dapolley/Desktop/sc2/configuration/stopwords.txt Browse
[Algorithm N	ame	Date	Time	🗖 Book Series Title 🚱
Ø	🗹 Load		06/26/2013	8 03:40:11 PN	Cited Patent
					City of Publisher
4					Conference Dates
4					
<u>ک</u>				_	Conference Host

Word Co-Occurrence Network

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

CNS Cyberinfrastructure for Network Science Center

(83)

Remove Romsvik Multitudinous Fields Remove Romsvik Multitudinous Fields Remove Romsvik Multitudinous Fields Extract Bipartite Network Extract Bipartite Network Extract Author Paper Network Extract Author Paper Network Extract Word Co-Occurrence Network Extract Word Co-Occurrence Network Extract Word Co-Occurrence Network Extract Word Co-Occurrence Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Bipartite Network Extract Document Co-Cutation Network Extract Bipartite Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Bipartite Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Bipartite Network Extract Bipartite Network Extract Bipartite Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Bipart Network	ts
Extract Directed Network Extract Bypartie Network Extract Bypartie Network Extract Bypart Network Extract Devork Extract Devork Extract Devork Extract Devork Extract Co-Occurrence Network Extract Co-Occurrence Network Extract Co-Occurrence Network Extract Devortence (Bibliographic Coupling) Network Extract Devortence Co-Occurrence (Bibliographic Coupling) Network Extract Devortence Network	ts
Etract Paper Citation Network Etract Autor Paper Network Etract Autor Paper Network Etract Autor Paper Network Etract Co-Occurrence Network Etract Co-Occurrence Network Etract Mence Co-Occurrence Network Etract Mence Co-Occurrence (Bibliographic Coupling) Network Etract Decime Co-Occurrence (Bibliographic Coupling) Network	• •
Estract Co-Occurrence Network Estract Co-Occurrence Network Estract Reference Co-Occurrence (Bibliographic Coupling) Network Estract Reference Co-Occurrence (Bibliographic Coupling) Network Estract Document Co-Citation Network	
Estract Co-Catabon Network Estract Reference Co-Occurrence (Bibliographic Coupling) Network Estract Document Co-Citation Network Estract Document Co-Citation Network	
Extract Co-Author Network Extract Co-Author Network Extract Reference Co-Occurrence (Bibliographic Coupling) Network Extract Document Co-Citation Network Extract Document Co-Citation Network	
Extract Document Co-Citation Network Aggregate Function File Cr/Users/dapolley/Desktop/sci2 Brow	
Extract Document Co-Citation Network	
	··· ·
	•
Update Network by Merging Nodes	•
Scheduler	
move From List Remove completed automatically Remove all completed	•
Conference Title	٠
Journal Title (Full)	•
! Algorithm Name Date Time % Complete	
Lowercase, Tokenize, St 06/26/2013 04:04:37 PM	•
✓ Load 06/26/2013 03:40:11 PM	•
Ŷ Title	
8	۲

NS Cuberinfrastructure for Word Co-Occurrence Network

89

90

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

File Data Preparation Preprocessing Analysis Modeling	Visualization R Help General	Data Manager	- 0
Abstract true Title false Book Series Title false	Temporal Geospatial Topical	ISI Data: C	:\Users\dapolley\Desktop\sci2 iique ISI Records th normalized Abstract
Conference Title: false Node Identifier Column: Cite Me As Original Keywords: false Journal Title (Full): false	Networks Cytoscape	GUESS Gephi	currence networ
Class: edu.iu.nwb.analysis.extractdirectednetfromtable.extractd Removed 354 isolate nodes. 	i Balloon Graph (prefuse alpha)	Radial Tree/Graph (prefuse alpha) Radial Tree/Graph with Annotation (prefuse beta) Tree View (prefuse beta)	
DrL (VxOrd)	Tree Map (prefuse beta) Force Directed with Annotation (prefuse beta) Fruchterman-Reingold with Annotation (prefuse be	eta)	
This algorithm lays out nodes based on the VxC layout algorithm.	Ord force-directed	DrL (VxOrd) Specified (prefuse beta)	
dge Weight Attribute weight		Bipartite Network Graph Circular Hierarchy	
lew X-Position Attribute Name xpos			
lew Y-Position Attribute Name ypos			
Do not cut edges			
dge Cutting Strength 0.0			

CNS cuberinfrastructure for 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*

File Data Preparation	Preprocessing Analysi	s Modeling Visualization R Help			
Console			- 8	III Data Manager	
output .iedges file = Proc. 0 scanning .int fil Processor 0 reading .int Entering liquid stage	e Geospatial + Topical +	1	<u> </u>	a 361 Unique ISI Rec	ords
Liquid stage completed Entering expansion stat	ge Networks P ge in 29 seconds, total ene	Extract Top Nodes Extract Nodes Above or Below Value Delete Isolates		Laid ou	
	stage in 30 seconds, total	e Extract Top Edges			
Entering simmer stage Finished simmer stage		Remove Self Loops Trim by Degree MST-Pathinder Network Scaling Fast Pathfinder Network Scaling Snowball Sampling (n nodes)	Extract Top Edges Extract the top N edges Extract N top edges		n a given attribute
	Remove completed auto	Node Sampling Edge Sampling Dichotomize	Extract bottom edge	es instead?	
1 Algorithm	n Name Date	Multipartite Joining	Numeric Attribute w	eight	- 📀
DrL (VxO DrL (VxO DrL (VxO Extract W	rd) 06/26/	2013 04:21:36 PM 2013 04:10:34 PM 2013 04:04:37 PM		<	OK Cancel

91

92

Suberinfrastructure for Word Co-Occurrence Network

(23)

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

Dara	met-Depend	4	and all the standards of the standard states and	
Field	Value	10	and the second	
_edgeid	957		and the set of the set	
color	176, 180, 178, 255		C. P. March & Section 1. State	
directed	false		and the second second second second second second second	
label		-	and the second state of the second state of the	
labelcolor	0,0,0,255		the second second the second second second	
labelsize	12		the second s	
labelvisible	false		Sent and the second	
node1	n436			
node2	n258	~ R	11、1963年代来及1215年代最大的1975年代。1995年代	
originallabel			the second s	
visible	true		a second seco	
weight width	11.0		くいちょう あい とうない あいまたい ないちき 感じる パー	
widen	4.0		Star Star and Warden Street and Street	
		Longer L	a three stands and the stand of the stand of the	
			The case of the second states and	
			a the state of the the second state of the s	
			a state and the second state of the second state of the second state of the	
		34	The Case of the State Price Price and	
		The second s	Constant Branch Strand Strand Strand	
			and a second second second second second	
			strainty to address of a real strain states	
		Objecti	Property: labeldze * Operator: * Value:	*
			Colour Show Hide Size Show Label Hide Label Change Label	
			Format Node Labels Format Edge Labels	
			Node Shape Center Change History	
			Resize Linear Colorize	
			Zoom Level: 0.37318	
		Interpreter Graph Modifier		

<text><text><text><code-block></code>

CNS Cuberinfrastructure for 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.





Questions?