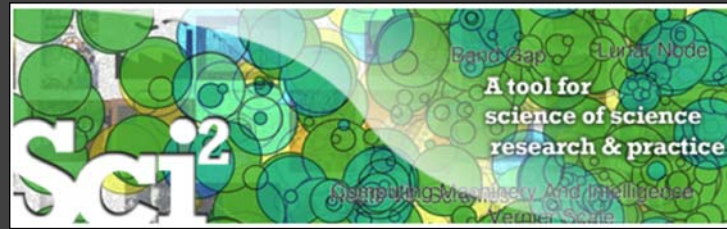


# Temporal Analysis using Sci2



**Ted Polley and Dr. Katy Börner**  
Cyberinfrastructure for Network Science Center  
Information Visualization Laboratory  
School of Library and Information Science  
Indiana University, Bloomington, Indiana, USA  
<http://cns.iu.edu>

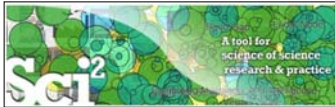
Please download Sci2 at <http://sci2.cns.iu.edu>  
See documentation at <http://wiki.cns.iu.edu/display/SCI2TUTORIAL>

Cyberinfrastructure for Network Science Center  
School of Library and Information Science  
Indiana University Bloomington  
LI001 Wells Library  
Monday October 15, 2012 – 6:00pm-7:00pm



## Overview

- Introduction to Temporal Analysis
  - Burst Detection
  - Slice Time by Table
  - Temporal Bar Graph
- Burst Detection Workflow
- Evolving Co-Authorship Network Workflow



## 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.)

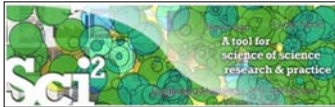
3



## Introduction to Temporal Analysis

- Demonstrate and discuss the utility of Kleinberg's burst detection algorithm as implemented in Sci2
- Design evolving network visualizations using Sci2

4

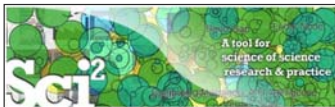


## Temporal Analysis: Burst Detection

- 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.

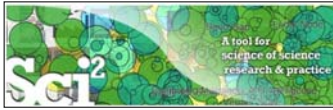
5



## 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

6



## 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.

7

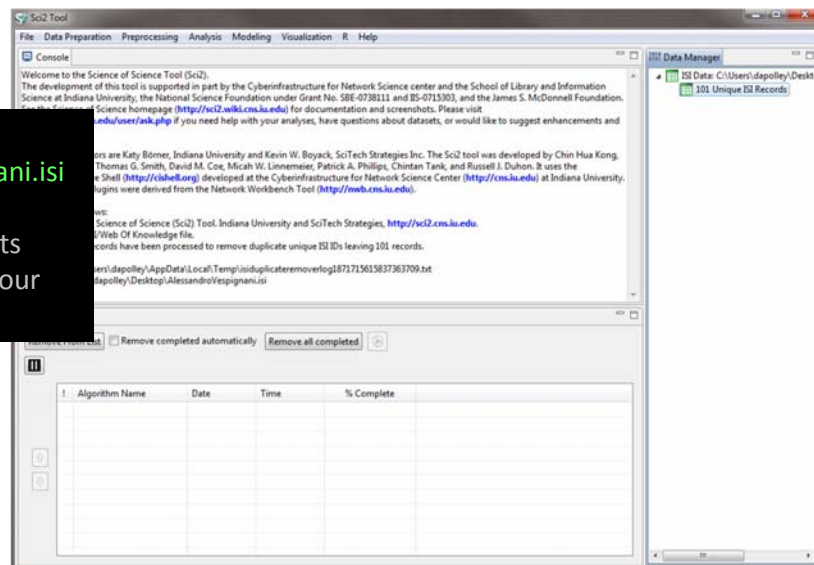


## Temporal Analysis: Burst Detection

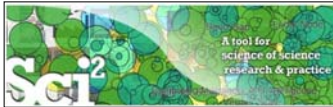
For this workflow we will be using the AlessandroVespignani.isi file, which is available from [2.5 Sample Datasets](#)

File > Load >  
AlessandroVespignani.isi

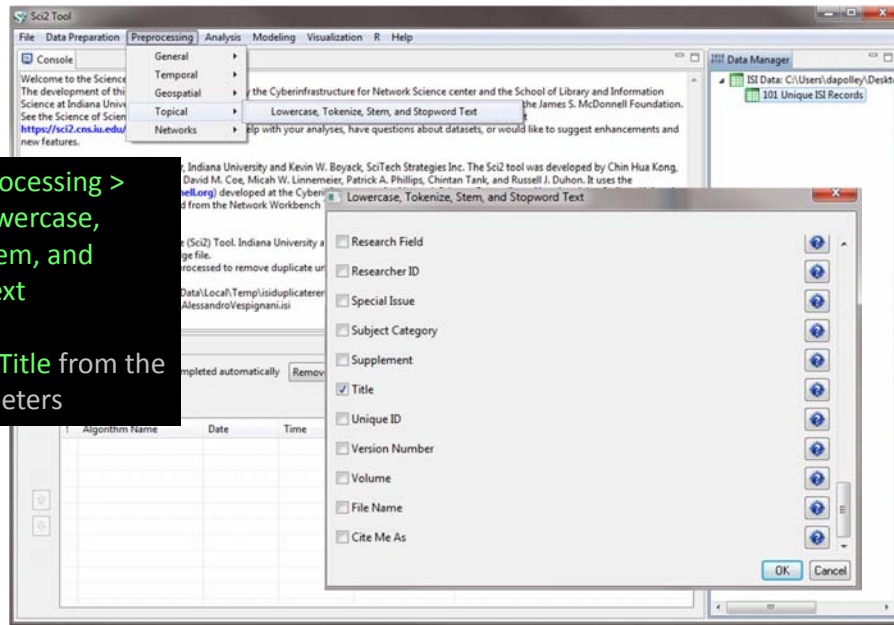
Load this file from its saved location on your computer



8



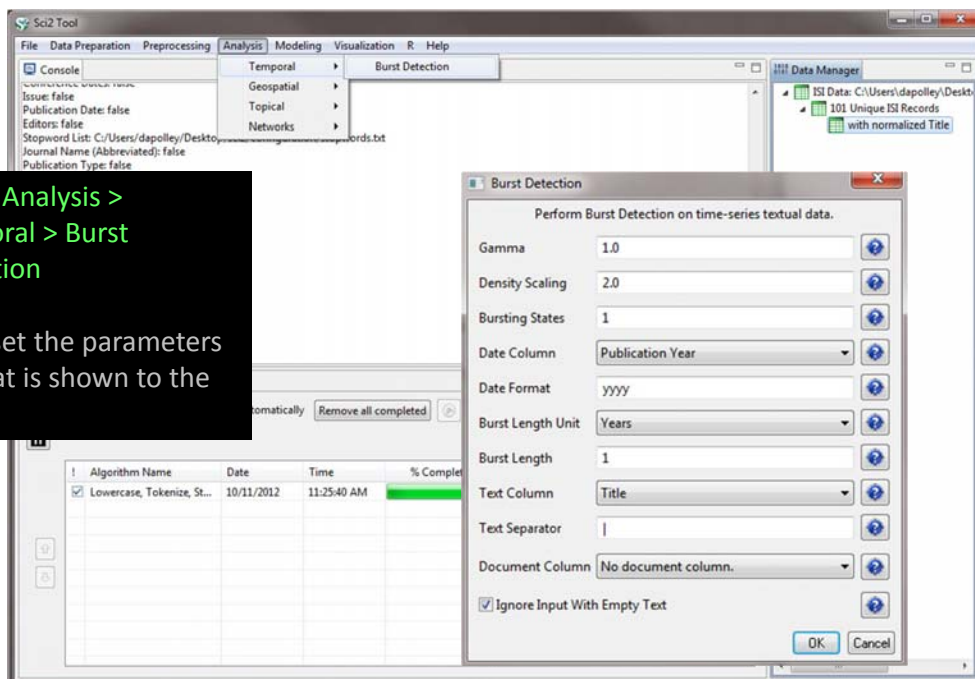
## Temporal Analysis: Burst Detection



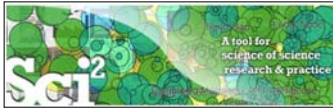
9



## Temporal Analysis: Burst Detection



10



## 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.

11



## Temporal Analysis: Burst Detection

Console

```

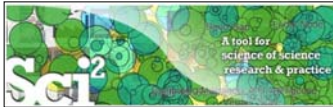
rout zapswalv1
Burst Lengths: 1
Ignore Input With Empty Text: true
Density Scaling: 2.0
Document Column: No document column.
Date Format: yyyy
Text Column: Title
  
```

	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

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

12





## Temporal Analysis: Burst Detection

Add the End date of **2006** to those records missing an 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

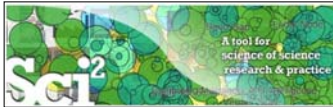
13



## Temporal Analysis: Burst Detection

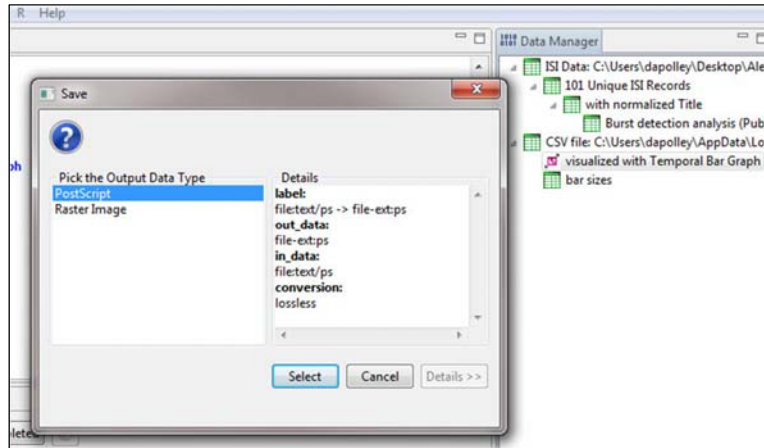
Select **Visualization > Temporal > Temporal Bar Graph** and set the parameter values to those shown to the right

14



# 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



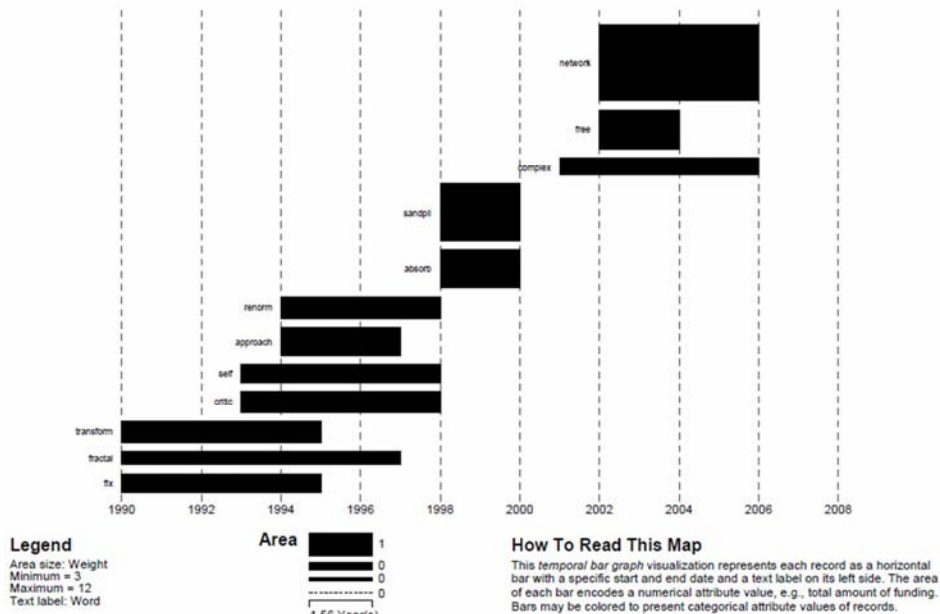
If you do not have a program to convert PostScript files look [here](#).



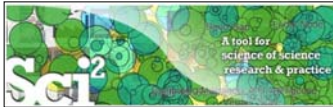
# Temporal Analysis: Burst Detection

## Temporal Visualization

(Alessandro Vespignani Word Burst in ISI Publication Titles 1990-2006)  
October 11, 2012 | 10:58 AM EDT





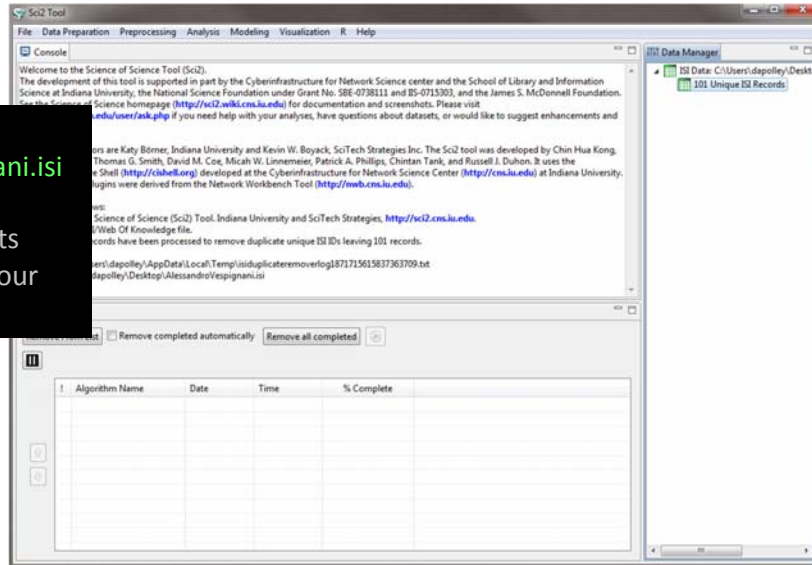


## Temporal Analysis: Evolving Co-Authorship Network

For this analysis we will be studying the evolution of Alessandro Vespignani's co-authorship network over time. We will see his network of collaborators grow from 1990 to 2006, giving us a sense of how his scholarly output has grown.

File > Load >  
AlessandroVespignani.isi

Load this file from its saved location on your computer

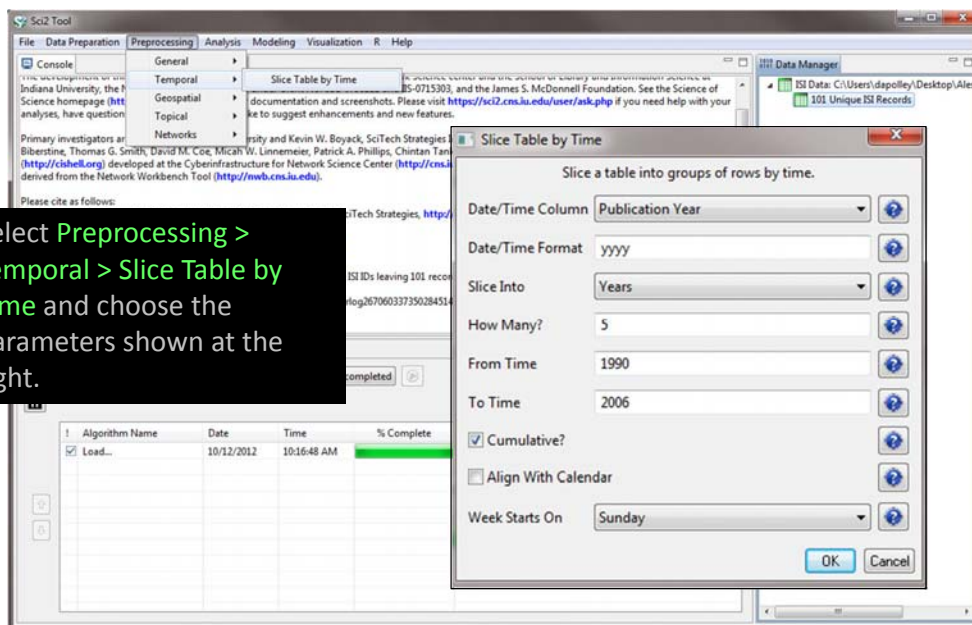


17

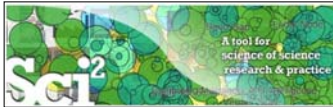


## Temporal Analysis: Evolving Co-Authorship Network

Select Preprocessing >  
Temporal > Slice Table by  
Time and choose the  
parameters shown at the  
right.



18



## Temporal Analysis: Evolving Co-Authorship Network

Now that the algorithm has been run, you will notice the original data set has been divided into four tables that cumulatively display the evolution of this data.

The screenshot shows the Sci2 Tool interface. The Data Manager window displays four data slices:

- 101 Unique ISI Records
- slice from beginning of 1990 to end of 2006 (101 records)
- slice from beginning of 1990 to end of 2001 (65 records)
- slice from beginning of 1990 to end of 1996 (26 records)
- slice from beginning of 1990 to end of 1991 (4 records)

The Scheduler window shows a table of algorithm results:

Algorithm Name	Date	Time	% Complete
Slice Table by Time	10/12/2012	10:34:18 AM	100%
Load...	10/12/2012	10:16:48 AM	100%

19



## Temporal Analysis: Evolving Co-Authorship Network

Select the first table and run **Data Preparation > Extract Co-Author Network**

Then set the file format to **isi**

Repeat this step for each of the tables in the Data Manager

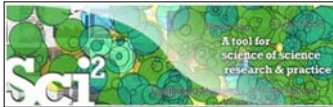
The screenshot shows the Sci2 Tool interface with the 'Extract Co-Author Network' dialog box open. The dialog box has the following text:

Extracts a co-authorship network from one of several supported file types.

File Format: **isi**

Buttons: OK, Cancel

20



## Temporal Analysis: Evolving Co-Authorship Network

Select the first extracted co-author network and run **Analysis > Networks > Unweighted & Undirected > Node Degree**

Then follow this step for the three other co-author networks you have extracted.

Algorithm Name	Date	Time	% Complete
Extract Co-Author Netw...	10/12/2012	00:38:14 PM	100%
Extract Co-Author Netw...	10/12/2012	00:38:09 PM	100%
Extract Co-Author Netw...	10/12/2012	00:38:04 PM	100%
Extract Co-Author Netw...	10/12/2012	00:37:54 PM	100%
Slice Table by Time	10/12/2012	00:37:17 PM	100%
Load...	10/12/2012	00:37:02 PM	100%

21

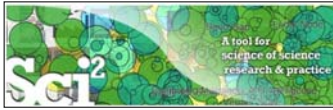


## Temporal Analysis: Evolving Co-Authorship Network

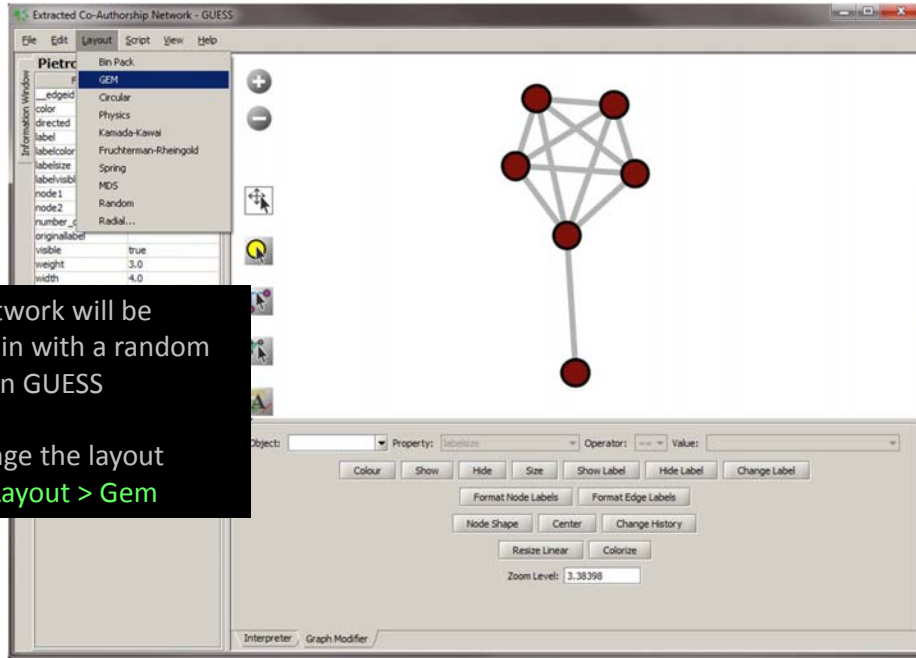
Select the first extracted co-author network and run **Visualization > Networks > GUESS**

Algorithm Name	Date	Time	% Complete
Extract Co-Author Netw...	10/12/2012	00:38:14 PM	100%
Extract Co-Author Netw...	10/12/2012	00:38:09 PM	100%
Extract Co-Author Netw...	10/12/2012	00:38:04 PM	100%
Extract Co-Author Netw...	10/12/2012	00:37:54 PM	100%
Slice Table by Time	10/12/2012	00:37:17 PM	100%
Load...	10/12/2012	00:37:02 PM	100%

22



## Temporal Analysis: Evolving Co-Authorship Network



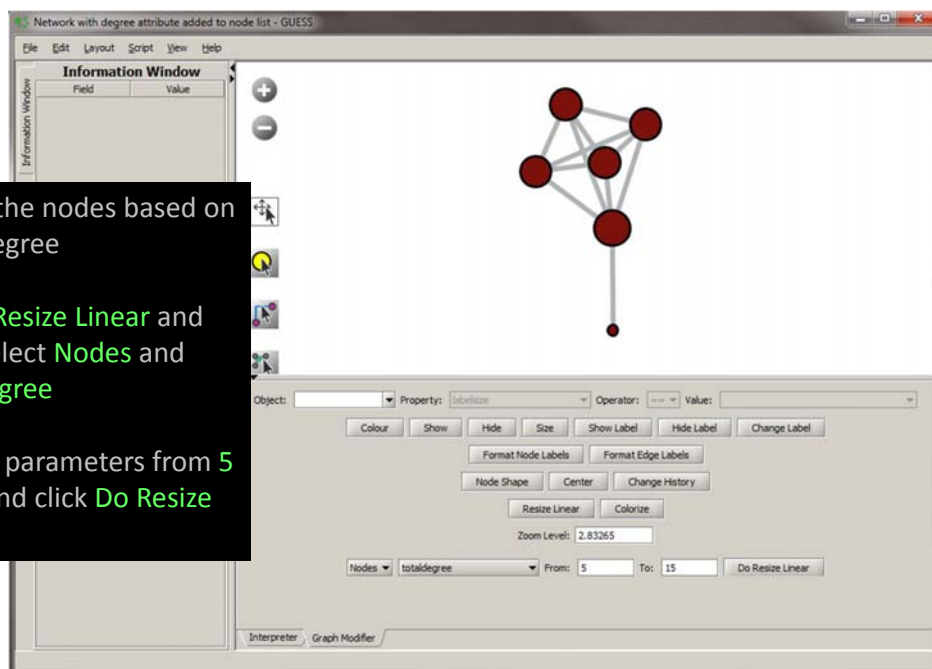
The network will be loaded in with a random layout in GUESS

To change the layout select **Layout > Gem**

23



## Temporal Analysis: Evolving Co-Authorship Network

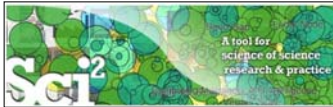


Resize the nodes based on total degree

Select **Resize Linear** and then select **Nodes** and **totaldegree**

Set the parameters from 5 to 15 and click **Do Resize Linear**

24



## Temporal Analysis: Evolving Co-Authorship Network

Network with degree attribute added to node list - GUESS

Field	Value
edges	1
color	176, 180, 178, 255
directed	false
label	
labelcolor	0, 0, 0, 255

Property: labelsize 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: 2.83265

Edges weight From: 1 To: 5 Do Resize Linear

Interpreter Graph Modifier

Resize the edges based on weight

Select **Resize Linear** and then select **Edges** and **weight**

Set the parameters from 1 to 5 and click **Do Resize Linear**

25



## Temporal Analysis: Evolving Co-Authorship Network

Network with degree attribute added to node list - GUESS

Field	Value
edges	0
color	176, 180, 178, 255
directed	false
label	
labelcolor	0, 0, 0, 255

Property: labelsize 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: 2.83265

Nodes totaldegree From: To: Do Colorize

Interpreter Graph Modifier

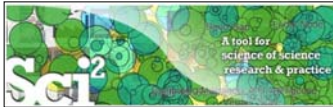
Colorize the nodes based on totaldegree

Select **Colorize** and then select **Nodes** and **totaldegree**

Set the parameters from  to  and click **Do Colorize**

26







## Temporal Analysis: Evolving Co-Authorship Network

Colorize the edges based on weight

Select **Colorize** and then select **Edges** and **weight**

Set the parameters from  to  and click **Do Colorize**

Network with degree attribute added to node list - GUESS

Deangelis, R	
Field	Value
color	51, 51, 255
fixed	false
height	13.228756555322...
image	
indegree	4

Property: labelsize 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: 2.83265

Edges weight From:  To:  Do Colorize

Interpreter Graph Modifier

27



## Temporal Analysis: Evolving Co-Authorship Network

If you want to remove the borders from the nodes, type the following commands into the interpreter:

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

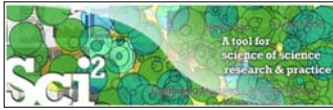
Network with degree attribute added to node list - GUESS

>>> for n in g.nodes:  
 n.strokecolor = n.color  
--  
>>>|

Interpreter Graph Modifier

28





## Temporal Analysis: Evolving Co-Authorship Network

Finally add the labels to the nodes by selecting  
Object: **all nodes**  
And then clicking: **Show Label**

Network with degree attribute added to node list - GUESS

Field	Value
_edged	4
color	0,204,0,255
directed	false
label	
labelcolor	0,0,0,255

Object: all nodes Property: labelsize 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: 2.83265

Interpreter Graph Modifier

29



## Temporal Analysis: Evolving Co-Authorship Network

To save the node positions of the current layout so that the layout is consistent across all time slices, run:  
**File > Export Node Positions**  
Save the file as a .csv file

Network with degree attribute added to node list - GUESS

File Edit Layout Script View Help

- Import Graph
- Export Image... Ctrl-I
- Export Graph... Ctrl-R
- Import Node Positions
- Export Node Positions** Ctrl-Q
- Save Logfile...
- Exit Ctrl-Q

labelvisible true

Save As

Save [n]: Desktop

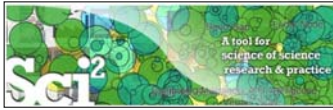
Admin	2012.10.15-Polley-flyer.pdf
interesting datasets	2012.10.15-Polley-flyer.pub
Map Archival	2012.10.15-Polley-flyer.zip
OSLOM2	AlessandroVespignani.isi
Presentations	ambassador.csv
Research	ambassador_geovis.jpg
sci2	crcStat.properties
VOSviewer_1.5.2_exe	collection.csv

File Name: node\_positions.csv  
Files of Type: All Files

Save Cancel

Interpreter Graph Modifier

30



## Temporal Analysis: Evolving Co-Authorship Network

When you load the next network in GUESS you run:

Layout > Gem

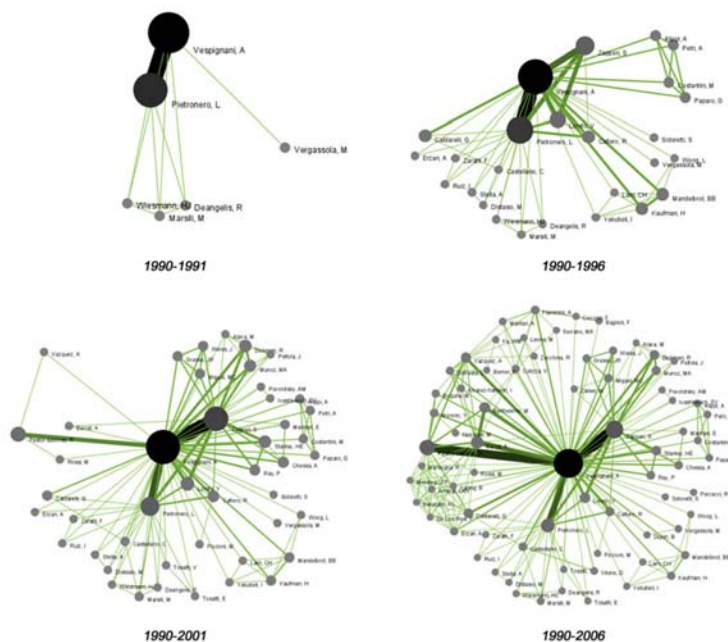
Then

File > Import Node Positions

31



## Temporal Analysis: Evolving Co-Authorship Network



32

Next session we will each download our own data to use for topical analysis. Please register for the Scholarly Database before the next session on November 19, 2012.

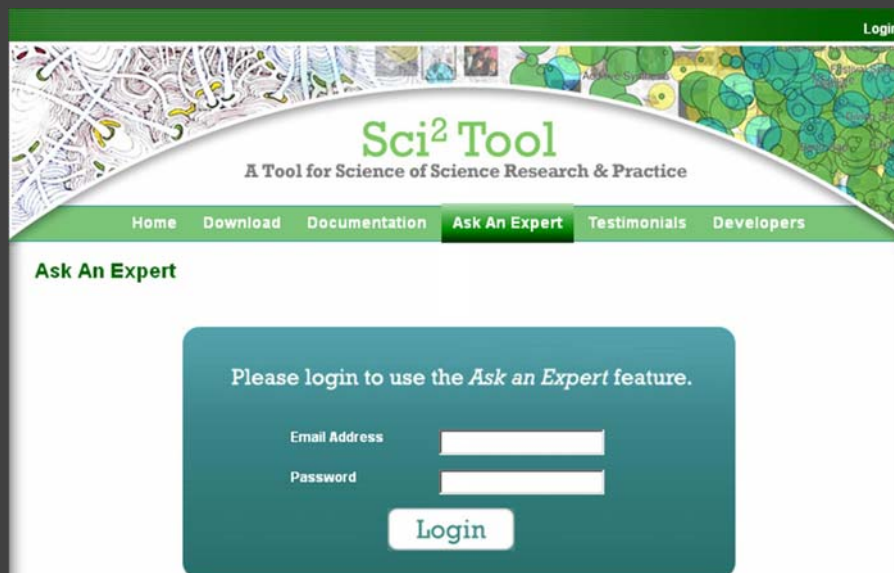


Register for the Scholarly Database at <http://sdb.cns.iu.edu/>

33



## Questions?



If you have questions feel free to email me [dapolley\[at\]indiana\[dot\]edu](mailto:dapolley[at]indiana[dot]edu)

34