

Workshop **Visualizing and Analyzing Scientific Literature with *CiteSpace***



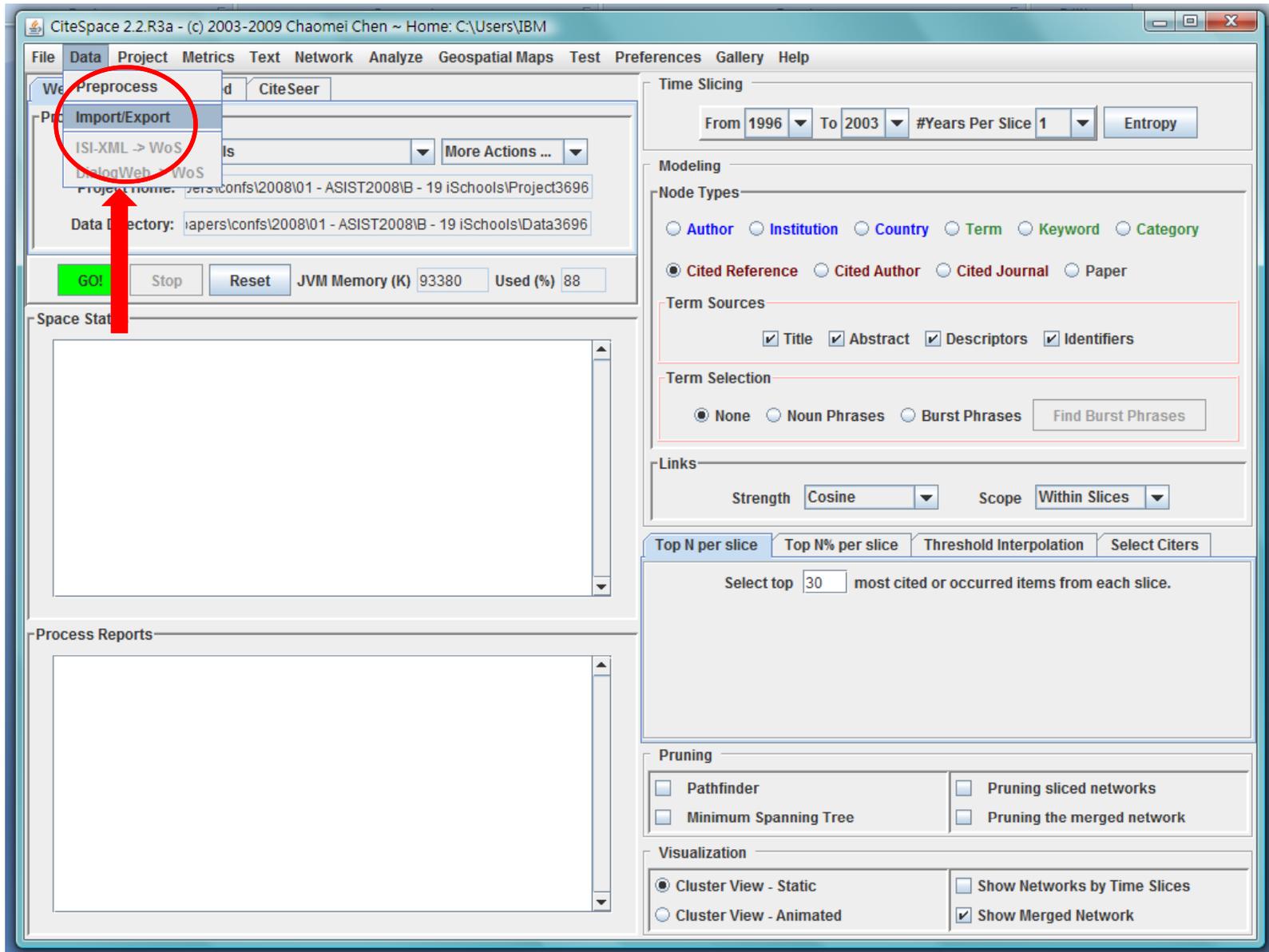
Chaomei Chen
College of Information Science and Technology
Drexel University

Updated: 11/16/2009

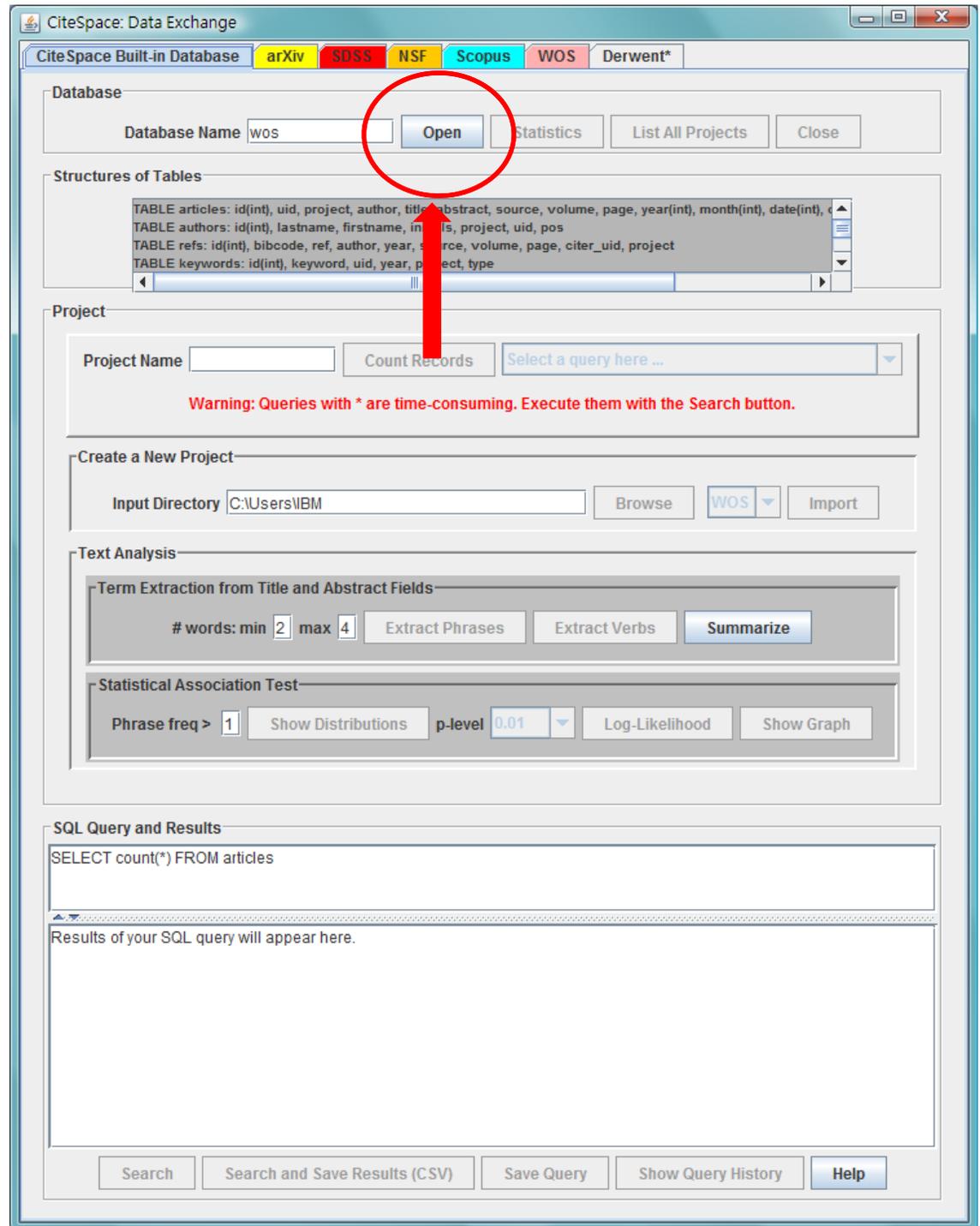
Email: chaomei.chen@cis.drexel.edu



Use the Built-in Database



To perform functions on this tab, open the database first by clicking on the Open button. Before you leave the page, remember to Close the database.



Show statistics of various tables in the database

The screenshot shows the CiteSpace: Data Exchange application window. At the top, there are tabs for different databases: arXiv, SDSS, NSF, Scopus, WOS, and Derwent*. The WOS tab is selected. Below the tabs, there is a 'Database' section with a text input field containing 'wos' and buttons for 'Open', 'Statistics', 'List All Projects', and 'Close'. The 'Statistics' button is circled in red, with a red arrow pointing to the 'Project' section below. The 'Structures of Tables' section lists several tables: articles, authors, refs, and keywords, each with its column definitions. The 'Project' section has a 'Project Name' field, a 'Count Records' button, and a dropdown menu for queries. A warning message states: 'Warning: Queries with * are time-consuming. Execute them with the Search button.' Below this is the 'Create a New Project' section with an 'Input Directory' field set to 'C:\Users\IBM', a 'Browse' button, a 'WOS' dropdown, and an 'Import' button. The 'Text Analysis' section includes 'Term Extraction from Title and Abstract Fields' with '# words: min 2 max 4' and buttons for 'Extract Phrases', 'Extract Verbs', and 'Summarize'. It also has a 'Statistical Association Test' section with 'Phrase freq > 1', 'Show Distributions' button, 'p-level 0.01', 'Log-Likelihood' button, and 'Show Graph' button. The 'SQL Query and Results' section shows a query: 'select distinct project from articles'. Below the query, a table of statistics is displayed, which is also circled in red. At the bottom, there are buttons for 'Search', 'Search and Save Results (CSV)', 'Save Query', 'Show Query History', and 'Help'.

Table Name	Column Definitions
articles	id(int), uid, project, author, title, abstract, source, volume, page, year(int), month(int), date(int), ...
authors	id(int), lastname, firstname, initials, project, uid, ...
refs	id(int), bibcode, ref, author, year, source, volume, page, citer_uid, project
keywords	id(int), keyword, uid, year, project, type

projects:	5
articles:	17433
authors:	31740
refs:	417563
keywords:	53874
phrases:	0
institutions:	20209
verbs:	0
DB file size:	0 (Kb)

List existing projects in the database

The screenshot shows the CiteSpace: Data Exchange application window. The 'Database' section has 'wos' selected in the 'Database Name' field. The 'List All Projects' button is circled in red, with a red arrow pointing to it. Below this, the 'Structures of Tables' section lists several tables: 'articles', 'authors', 'refs', and 'keywords'. The 'Project' section has a 'Project Name' field, a 'Count Records' button, and a 'Select a query here ...' dropdown menu. A warning message states: 'Warning: Queries with * are time-consuming. Execute them with the Search button.' The 'Create a New Project' section has an 'Input Directory' field set to 'C:\Users\IBM' and 'Browse', 'WOS', and 'Import' buttons. The 'Text Analysis' section has 'Term Extraction from Title and Abstract Fields' with '# words: min 2 max 4' and 'Extract Phrases', 'Extract Verbs', and 'Summarize' buttons. The 'Statistical Association Test' section has 'Phrase freq > 1', 'Show Distributions', 'p-level 0.01', 'Log-Likelihood', and 'Show Graph' buttons. The 'SQL Query and Results' section shows the query 'select distinct project from articles' and a list of projects: 'Conor', 'ISArticles', 'InfoSci', 'WittenE', and 'nmf'. The bottom of the window has 'Search', 'Search and Save Results (CSV)', 'Save Query', 'Show Query History', and 'Help' buttons.

CiteSpace: Data Exchange

CiteSpace Built-in Database arXiv SDSS NSF Scopus WOS Derwent*

Database

Database Name wos Open Statistics **List All Projects** Close

Structures of Tables

```
TABLE articles: id(int), uid, project, author, title, abstract, source, volume, page, year(int), month(int), date(int),
TABLE authors: id(int), lastname, firstname, initials, project, uid, pos
TABLE refs: id(int), bibcode, ref, author, year, source, volume, page, citer_uid, project
TABLE keywords: id(int), keyword, uid, year, project, type
```

Project

Project Name Count Records Select a query here ...

Warning: Queries with * are time-consuming. Execute them with the Search button.

Create a New Project

Input Directory C:\Users\IBM Browse WOS Import

Text Analysis

Term Extraction from Title and Abstract Fields

words: min 2 max 4 Extract Phrases Extract Verbs Summarize

Statistical Association Test

Phrase freq > 1 Show Distributions p-level 0.01 Log-Likelihood Show Graph

SQL Query and Results

```
select distinct project from articles
```

The database currently contains 5 projects:

- Conor
- ISArticles
- InfoSci
- WittenE
- nmf

Search Search and Save Results (CSV) Save Query Show Query History Help

Counting records in an existing project: InfoSci

1. Enter the project name
2. Click on the Count Records button

Results will be shown in the window at the bottom of the interface.

The SQL query used is shown in the SQL Query window above the results.

The screenshot shows the CiteSpace: Data Exchange application window. The 'Database' section is set to 'wos'. The 'Structures of Tables' section lists: TABLE articles: id(int), uid, project, author, title, abstract, source, volume, page, year(int), month(int), date(int), TABLE authors: id(int), lastname, firstname, initials, project, uid, pos, TABLE refs: id(int), bibcode, ref, author, year, source, volume, page, citer_uid, project, TABLE keywords: id(int), keyword, uid, year, project, type. In the 'Project' section, 'Project Name' is 'InfoSci' and the 'Count Records' button is highlighted with a red circle. A red arrow points from the 'Count Records' button to the 'SQL Query and Results' section. A warning message reads: 'Warning: Queries with * are time-consuming. Execute them with the Search button.' The 'SQL Query and Results' section shows the query: 'select distinct project from articles'. Below the query, a table of statistics is displayed, with the table area highlighted by a red box:

projects:	1
articles:	10853
authors:	17797
refs:	206180
keywords:	26778
phrases:	0
institutions:	11194
verbs:	0
DB file size:	0 (Kb)

At the bottom of the interface, there are buttons for 'Search', 'Search and Save Results (CSV)', 'Save Query', 'Show Query History', and 'Help'.

Show the most productive authors ...

The screenshot shows the CiteSpace: Data Exchange application interface. The 'Project' section has 'Most Productive Authors' selected in a dropdown menu, which is circled in red with an arrow pointing to it. The 'SQL Query and Results' section shows a query and a list of authors with their counts, also circled in red.

Database Name: wos

Structures of Tables:

- TABLE articles: id(int), uid, project, author, title, abstract, source, volume, page, year(int), month(int), date(int),
- TABLE authors: id(int), lastname, firstname, initials, project, uid, pos
- TABLE refs: id(int), bibcode, ref, author, year, source, volume, page, citer_uid, project
- TABLE keywords: id(int), keyword, uid, year, project, type

Project Name: InfoSci

Count Records: Most Productive Authors

Warning: Queries with * are time-consuming. Execute them with the Search button.

Create a New Project

Input Directory: C:\Users\IBM

Text Analysis

Term Extraction from Title and Abstract Fields

words: min 2 max 4

Statistical Association Test

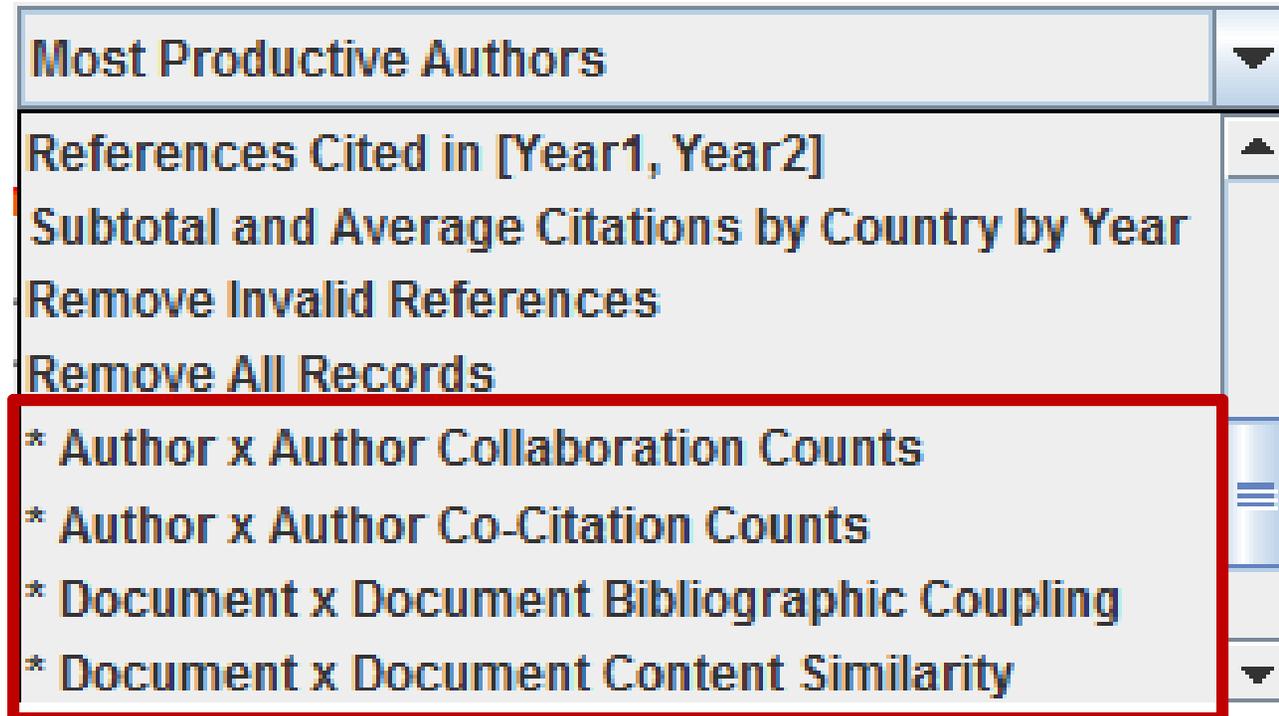
Phrase freq > 1

SQL Query and Results

```
select count(lastname + ',' + firstname), lastname + ',' + firstname from authors where project='InfoSci' group by lastname + ',' + firstname order by count(lastname + ',' + firstname) desc
```

800	[Anon],
98	Oppenheim, C
79	Thelwall, M
78	Egghe, L
75	Rousseau, R
74	Bawden, D
71	Boyce, BR
64	Fourie, I
64	Line, MB
62	Hernon, P
59	Spink, A

CiteSpace provides several ready-to-use queries. Note that queries with an * are likely to take much longer time to complete than those without an *.



1. You can enter your own SQL query and click on the Search button.
2. You may search and save the results to a file in CSV format in one function.
3. You can also save queries to a query history.

The screenshot shows a window titled "SQL Query and Results". The query text area contains the following SQL query:

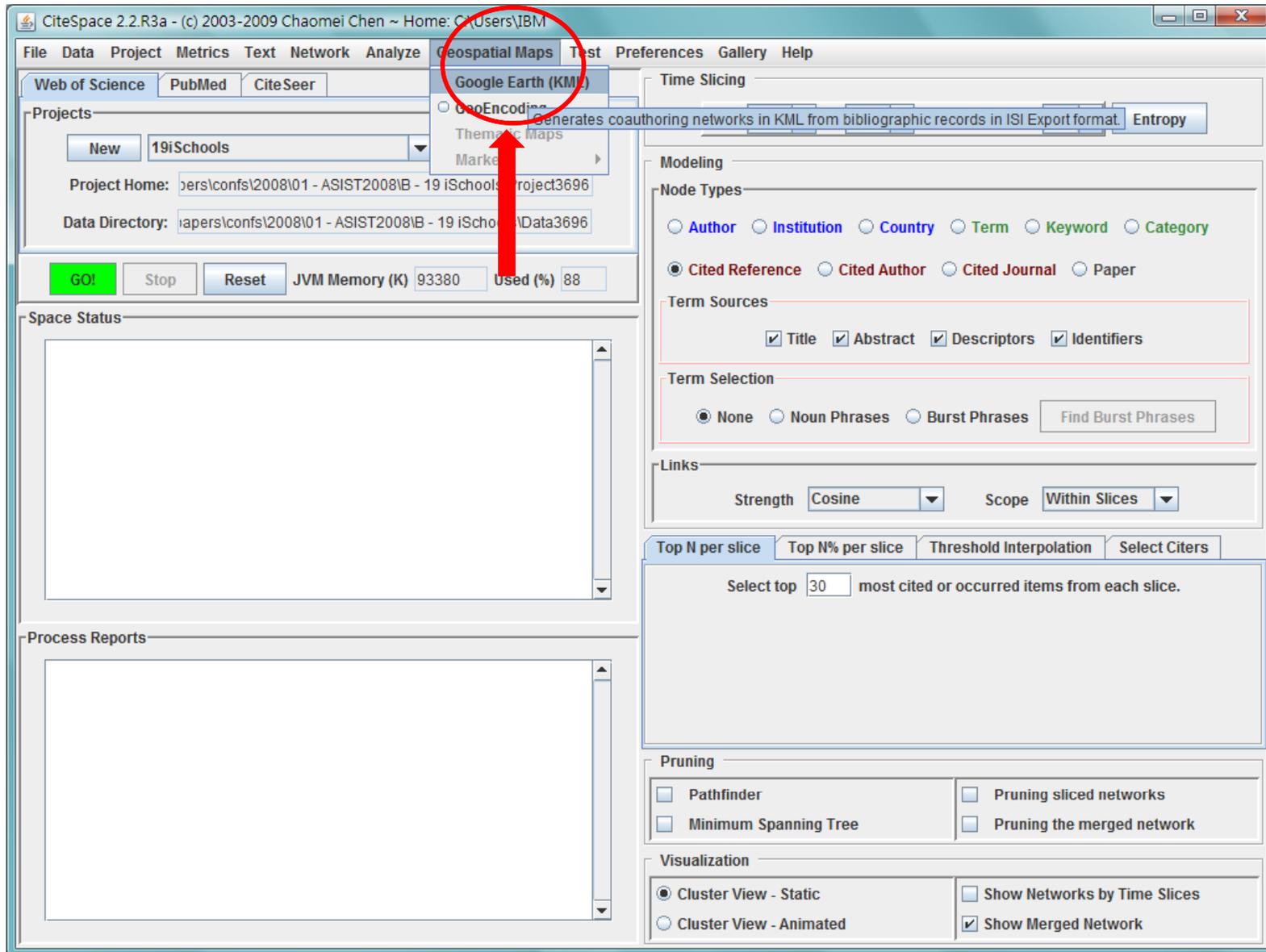
```
select count(lastname + ',' + firstname), lastname + ',' + firstname from authors where project='InfoSci' group by lastname + ',' + firstname order by count(lastname + ',' + firstname) desc
```

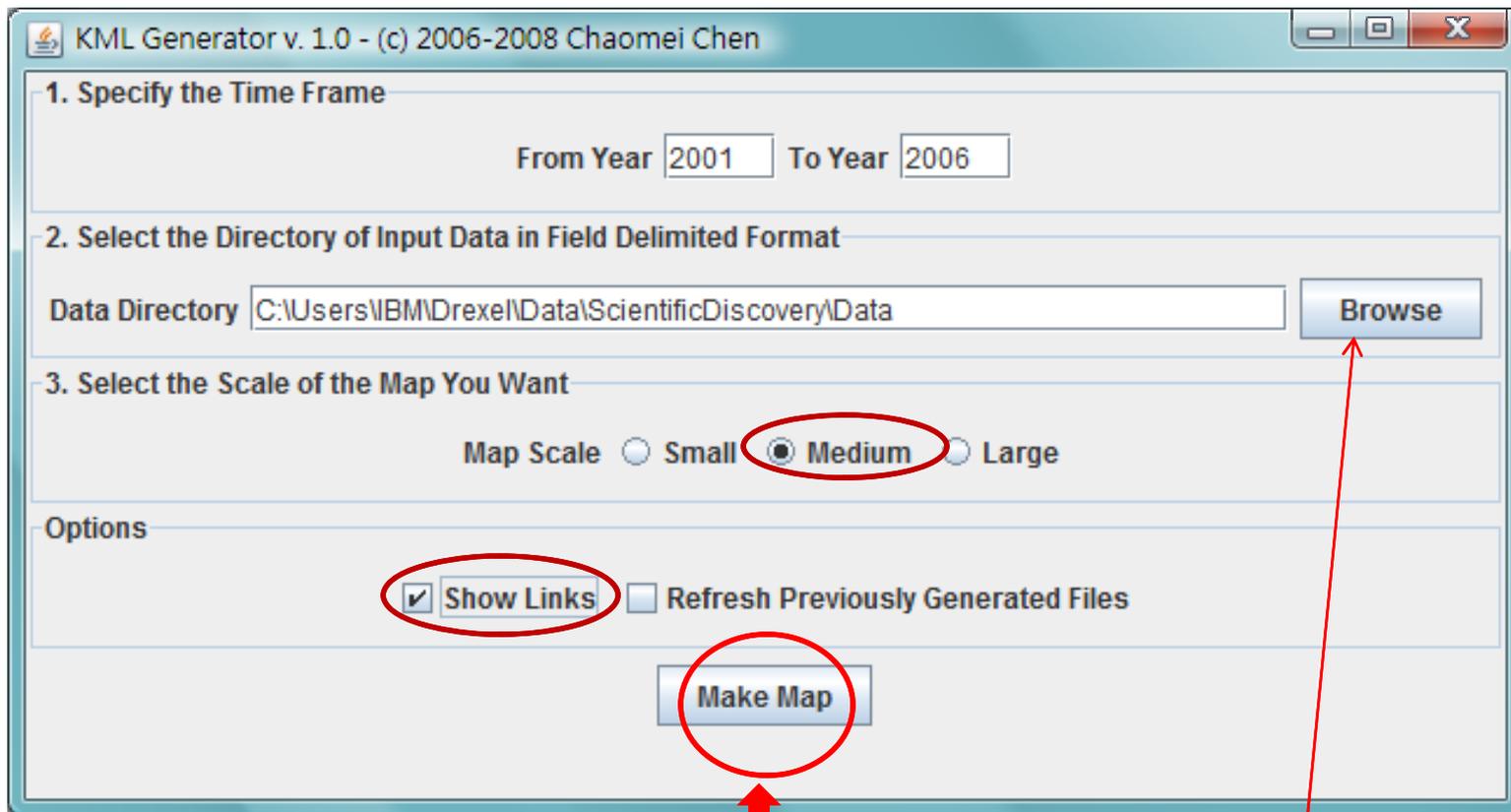
Below the query text is a table of results:

800	[Anon],
98	Oppenheim, C
79	Thelwall, M
78	Egghe, L
75	Rousseau, R
74	Bawden, D
71	Boyce, BR
64	Fourie, I
64	Line, MB
62	Hernon, P
59	Spink, A

At the bottom of the window are five buttons: "Search", "Search and Save Results (CSV)", "Save Query", "Show Query History", and "Help". The "Search" button is circled with a red circle containing the number 1. The "Search and Save Results (CSV)" button is circled with a red circle containing the number 2. The "Save Query" button is circled with a red circle containing the number 3.

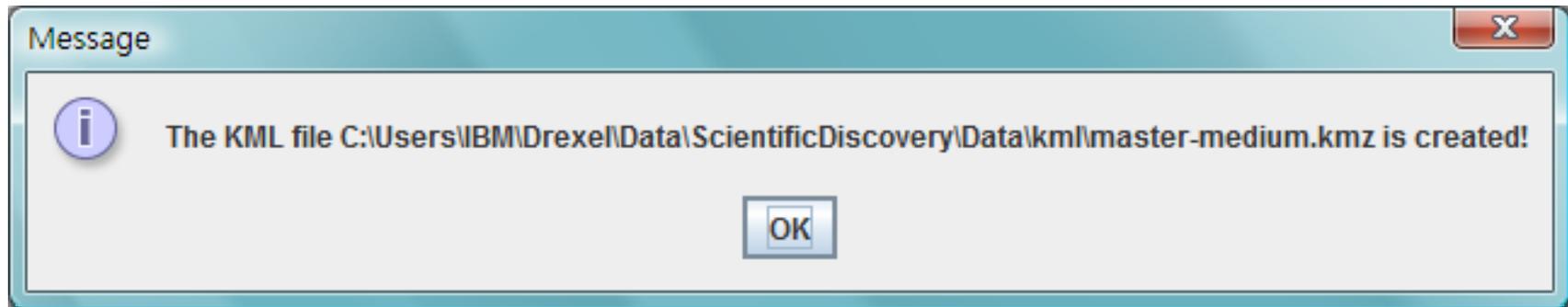
Generating Google Earth Overlays





Data directory as you would create for CiteSpace, i.e. download*.txt

wait until you see this:



view it in Google Earth or Google Map

A Collaboration Network of Scientific Discovery Researchers

Note: I will update the screenshot with an English version soon

The screenshot displays the Google Earth interface with a global network of red lines and dots. The interface includes a search bar, a location list, and a layer list. A red text overlay points to the layer list.

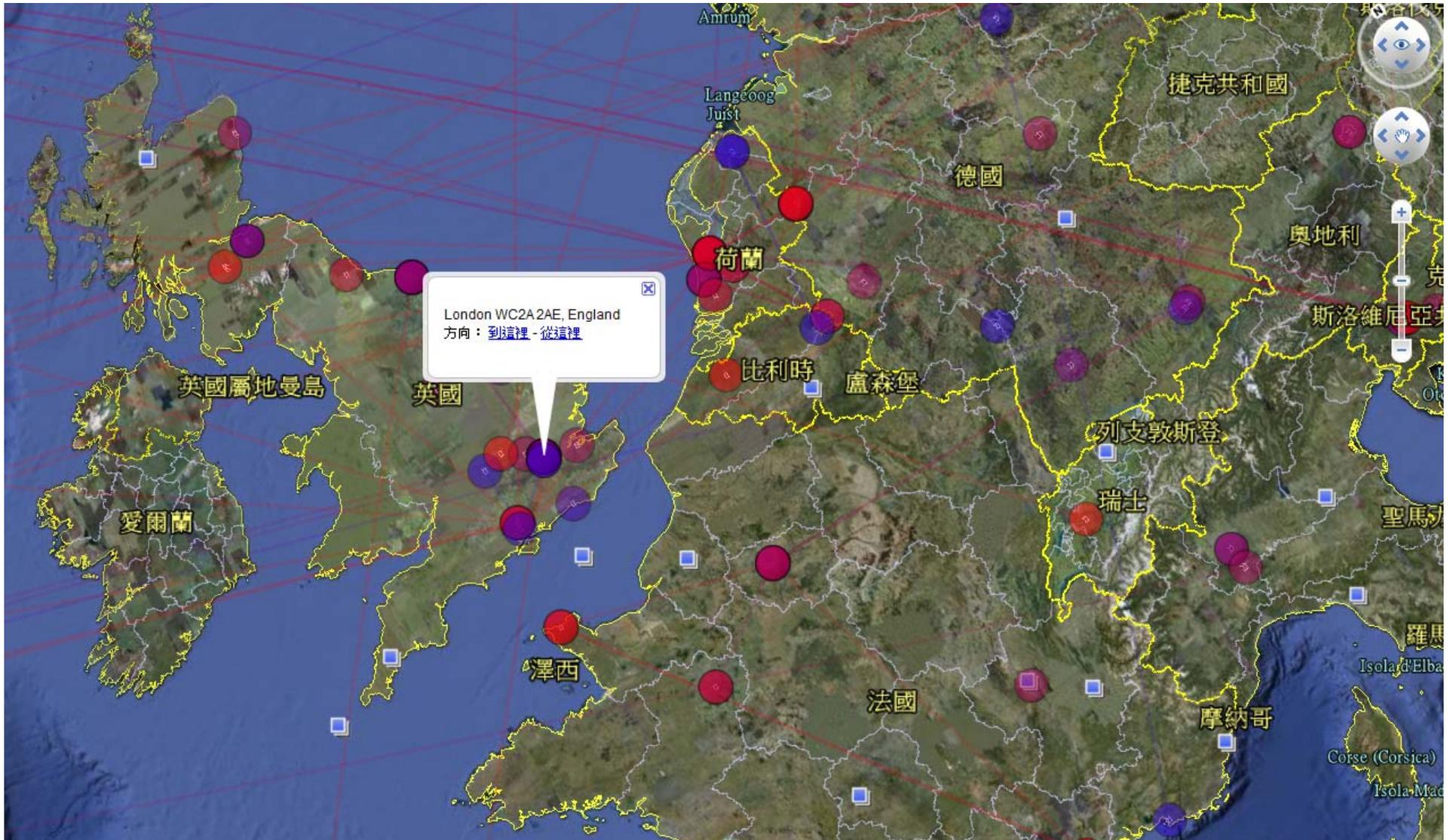
Select different layers here

© 2009 Europa Technologies
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
US Dept of State Geographer
© 2009 Tele Atlas

56°42'56.19" 北 54°53'21.37" 西 海拔高度 -8766 英尺
視角海拔高度 7987.38 英里

Coauthorship links between 2001 (blue) and 2006 (red) (showing Europe)

Accurate to the post code level



Further Readings on Geospatial Mapping

- **Chen, C.**, Song, I. Y., Yuan, X. J., Zhang, J. (2008) The Thematic and Citation Landscape of Data and Knowledge Engineering (1985-2007). *Data and Knowledge Engineering*, **67**(2), 234-259.
<http://cluster.cis.drexel.edu/~cchen/papers/2008/dke2008.pdf>
- **Chen, C.** (2007) Holistic sense making: Conflicting opinions, creative ideas, and collective intelligence. *Library Hi Tech*, 25(3), 311-327.
<http://cluster.cis.drexel.edu/~cchen/papers/2007/LibHiTech2007preprint.pdf>
- **Chen, C.**, Zhu, W., Tomaszewski, B., MacEachren, A. (2007) Tracing conceptual and geospatial diffusion of knowledge. [*HCI International 2007*](#). Beijing, China. July 22-27, 2007. LNCS, 4564. pp. 265-274.
<http://cluster.cis.drexel.edu/~cchen/papers/confs/hcii2007.pdf>

To be updated ...

Advanced Topics

Extracting Noun Phrases

Clustering and Labeling

Working with Pubmed

CiteSpace 2.2.R1 - (c) 2003-2009 Chaomei Chen ~ Home: C:\Users\IBM - COT

File Data Project Metrics Text Network Analyze Geospatial Maps Test Preferences Gallery Help

Web of Science PubMed CiteSeer

Projects

New Terrorism More Actions ...

Project Home: C:\Users\IBM\citespace\Examples\Projects\Terrorism

Data Directory: C:\Users\IBM\citespace\Examples\Data\Terrorism1990-2003

GO! Stop Reset JVM Memory (K) 275677 Used (%) 28

Space Status

CiteSpace is pre-processing data files. Please wait ...

Years: 14

Unique source records: 1732

Please wait while CiteSpace imports files and builds networks.
Note that counts in the space column include both citer and citee entries.
The process may take several minutes to complete.

1-year slices	criteria	space	nodes	links
1996-1996	top 30	655	30	86

Pruning configuration:

Process Reports

download1996_1x5unique.txt	5	5	
download1996_2x5unique.txt	5	5	
download1996_3x5unique.txt	5	5	
download1996_4x5unique.txt	5	5	
download1996_5x5unique.txt	5	5	
download1996_6x5unique.txt	5	5	
download1996_7x5unique.txt	5	5	
download1996_8x5unique.txt	5	5	
download1996_9x5unique.txt	5	5	
1997-1997	download1997_0x5unique.txt	5	5
download1997_10x5unique.txt	5	5	
download1997_11x5unique.txt	5	5	

Time Slicing

From 1996 To 2003 #Years Per Slice 1 Entropy

Modeling

Node Types

Author Term Keyword Category Institution Country

Cited Reference Cited Author Cited Journal

Term Sources

Title Abstract Descriptors Identifiers

Term Selection

None Noun Phrases Burst Phrases Find Burst Phrases

Links

Strength Cosine Scope Within Slices

Top N per slice Top N% per slice Threshold Interpolation Select Citers

Select top 30 most cited or occurred items from each slice.

Pruning

Pathfinder Pruning sliced networks

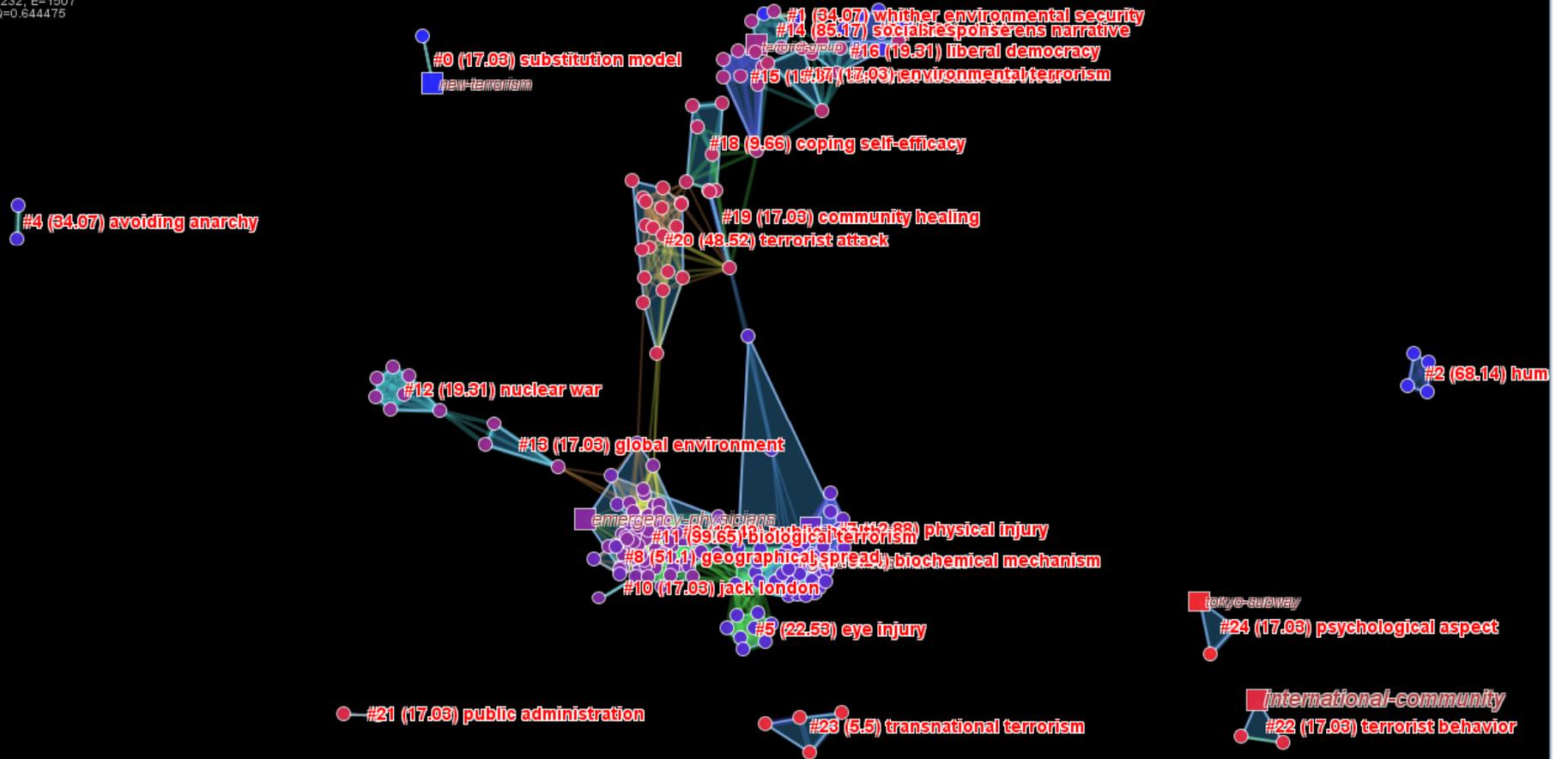
Minimum Spanning Tree Pruning the merged network

Visualization

Cluster View (Default) Show Networks by Time Slices

Time-zone View Show Merged Network

CiteSpace, v. 2.2.R1
March 9, 2009 11:11:48 AM EST
C:\Users\IBM\workspace\Examples\Data\Terrorism1990-2003
Timespan: 1996-2003 (Slice Length=1)
Selection Criteria: Top 40 per slice TC=[1, 211]
Network: N=232, E=1507
Modularity Q=0.644475





CiteSpace: Data Exchange

CiteSpace Built-in Database **arXiv** SDSS NSF **Scopus** WOS Derwent*

ADS Basic Search

Publication Date between (MM/YYYY) 1 2000 and (MM/YYYY) 12 2008

Title Search Text Search

Records Found **Count Records**

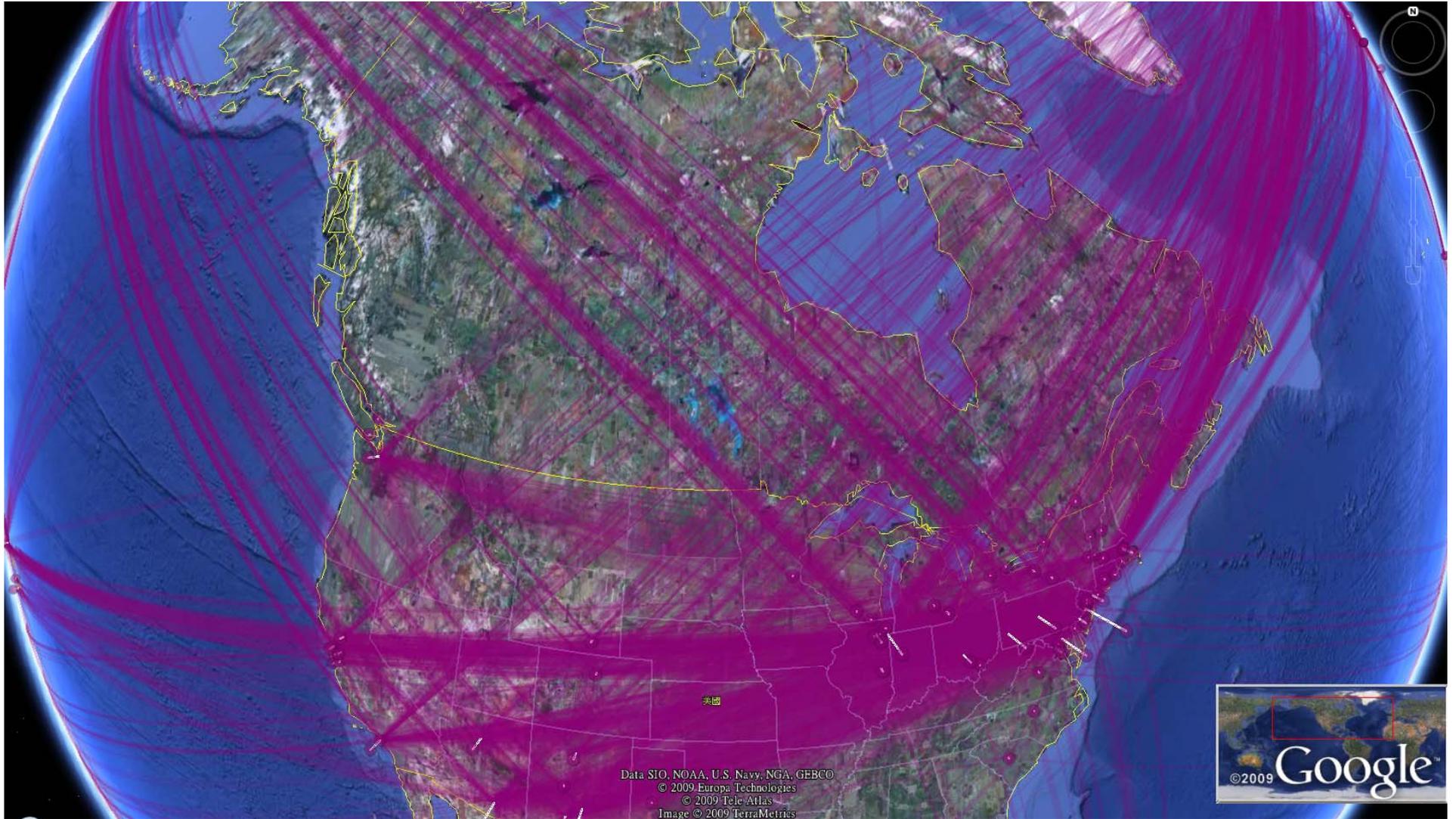
Show Previously Downloaded Files **Clear Downloaded Files** **Download** **Convert to WOS**

The scope of arXiv search:

- astro-ph
- cond-mat
- cs
- gr-qc
- hep-ex
- hep-lat
- hep-ph
- hep-th
- math
- math-ph
- nlin
- nucl-ex
- nucl-th
- physics
- quant-ph
- q-bio

Sun Mar 08 14:53:08 EDT 2009
Listing C:\Users\IBMADS...

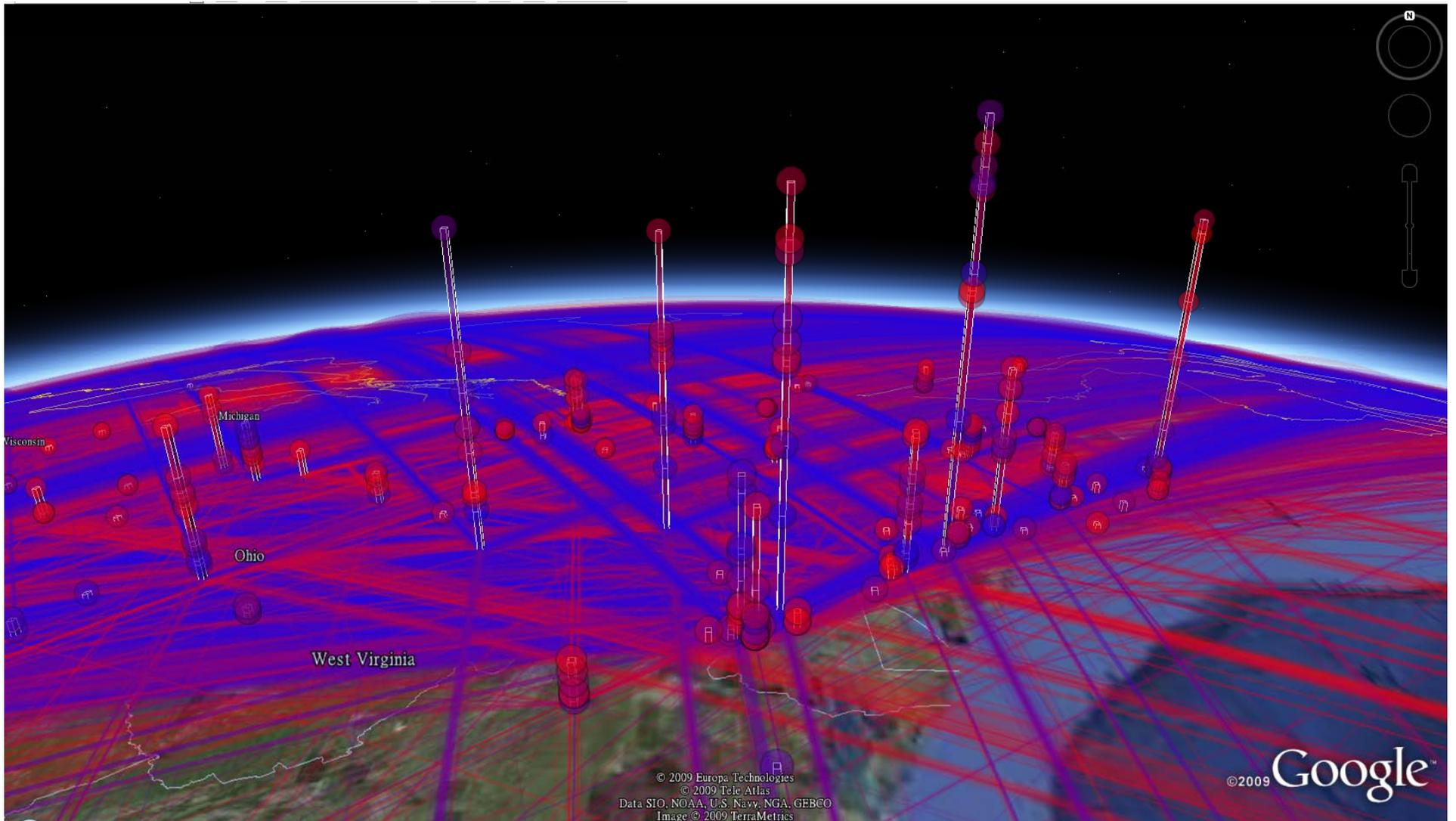
Clearing C:\Users\IBMADS...
Sun Mar 08 14:53:15 EDT 2009
Listing C:\Users\IBMADS...





Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2009 Europa Technologies
© 2009 Tele Atlas
Image © 2009 TerraMetrics





QUESTIONS AND DISCUSSIONS