The *CiteSpace* Manual

Version 0.65

Chaomei Chen
College of Computing and Informatics
Drexel University

Location of the manual:

http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf

Last Updated:

April 12, 2014

How to Cite:

Chen, Chaomei (2014) The *CiteSpace Manual*. http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf

Contents

1 What can I use CiteSpace for? 4
1.1 How should I cite CiteSpace? 4
2 Requirements to Run CiteSpace 5
2.1 Java Runtime (JRE) 5
2.2 How do I check whether Java is on my computer? 5
2.3 Do I have a 32-bit or 64-bit Computer? 7
3 How to Install and Configure *CiteSpace* 7
3.1 Where Can I download CiteSpace from the Web? 7
3.2 What is the maximum number of records that I can handle with CiteSpace? 8
3.3 How to configure the memory allocation for CiteSpace? 8
3.4 How to uninstall CiteSpace 9
4 Get Started with CiteSpace 10
4.1 Try it with a demonstrative dataset 10
4.1.1 The Demo Project 11
4.1.2 Clustering 14
4.1.3 Generate Cluster Labels 15
4.1.4 Where are the major areas of research based on the input dataset? 17
4.1.5 How are these major areas connected? 18
4.1.6 Where are the most active areas? 19
4.1.7 What is each major area about? Which/where re the key papers for a given area? 21
4.1.8 Timeline View 23
4.2 Try it with a dataset of your own 24
4.2.1 Collecting Data 24
4.2.2 Working with a CiteSpace Project 28
4.2.3 Data Sources in Chinese 29
5 Configure a CiteSpace Run 30
5.1 Time Slicing 30
5.2 Text Processing 30
5.3 Configure the Networks 30
5.4 Node Selection Criteria 31
5.5 Pruning, or Link Reduction 32
5.6 Visualization 32
6 Interactive with CiteSpace 33

6.1 Adding a Persistent Label to a Node 33
6.2 Using Aliases to Merge Nodes 33
7 Additional Functions 35
7.1 Menu: Data 35
7.1.1 CiteSpace Built-in Database 35
7.1.2 Utility Functions for the Web of Science Format 38
7.1.3 PubMed 39
7.2 Menu: Network 41
7.2.1 Batch Export to Pajek .net Files 41
7.3 Menu: Geographical 41
7.3.1 Generate Google Earth Maps 41
7.4 Menu: Text 44
7.4.1 Concept Trees and Predicate Trees 44
7.4.2 List Terms by Clumping Properties 47
7.4.3 Latent Semantic Analysis 48
8 References 49

1 What can I use CiteSpace for?

The default setting for the demo project will allow us to focus on the most important features in CiteSpace to answer questions such as the following:

- What are the major areas of research based on the input dataset?
- How are these major areas connected, i.e. through which specific articles?
- Where are the most active areas?
- What is each major area about? Which/where are the key papers for a given area?
- Are there critical transitions in the history of the development of the field? Where are the 'turning points'?

The design of CiteSpace is inspired by Thomas Kuhn's structure of scientific revolutions. The goal is to reveal patterns concerning scientific paradigms, including their structural and dynamic properties. Identifying transitions between an old paradigm and a new one is a central task supported by CiteSpace.

1.1 How should I cite CiteSpace?

The following three publications represent the core ideas of CiteSpace.

The 2004 PNAS paper is the initial publication on CiteSpace (Chen 2004). In hindsight, it could have been named CiteSpace I. The 19-page 2006 JASIST paper gives the most thorough and in-depth description of CiteSpace II's key functions (C. M. Chen, 2006), plus a follow-up study of domain experts identified in the visualizations. The 2010 JASIST paper is even longer with 24 pages (C. Chen, Ibekwe-SanJuan, & Hou, 2010), which is the third of the trilogy. It describes technical details on how cluster labels are selected and how each of the three selection algorithms in comparison with labels chosen by domain experts.

Citations (Google Scholar)	**Reference**
769	Chen, C. (2006). "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature." Journal of the American Society for Information Science and Technology **57**(3): 359-377.
383	Chen , C. (2004). "Searching for intellectual turning points: Progressive Knowledge Domain Visualization." Proc. Natl. Acad. Sci. USA **101**(Suppl.): 5303-5310.
152	Chen, C., et al. (2010). "The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis." Journal of the American Society for Information Science and Technology **61**(7): 1386-1409.

The most recent case study of a topic outside the realm of information science and scientometrics is a scienometric study of regenerative medicine (C. Chen, Hu, Liu, & Tseng, 2012).

> Chen, C., et al. (2012). "Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace." Expert Opinions on Biological Therapy **12**(5): 593-608.

2 Requirements to Run CiteSpace

2.1 Java Runtime (JRE)

CiteSpace is written in Java. It is a Java application. You should be able to run it on a computer that supports Java, including Windows or Mac.

CiteSpace is currently optimized for Windows 64-bit Java 7 (i.e. Java 1.7).

To run a Java application on your computer, you need to have Java Runtime (JRE) installed on your computer.

2.2 How do I check whether Java is on my computer?

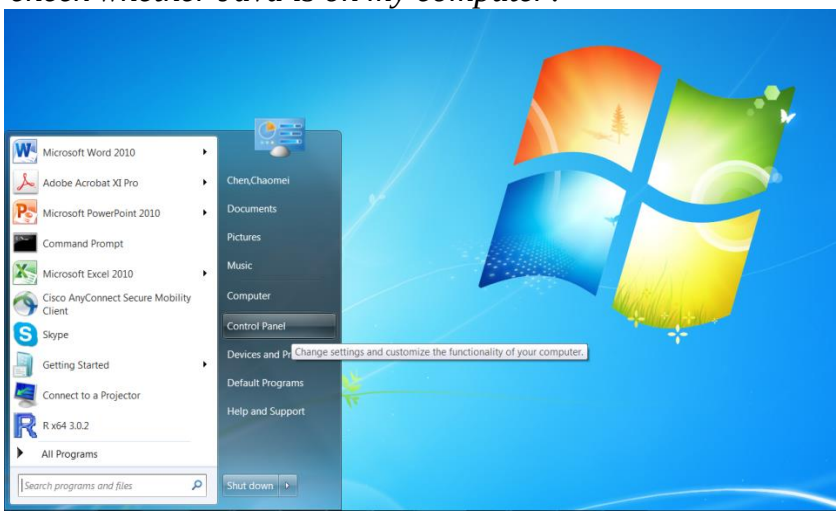


Figure 1. Select Control Panel.

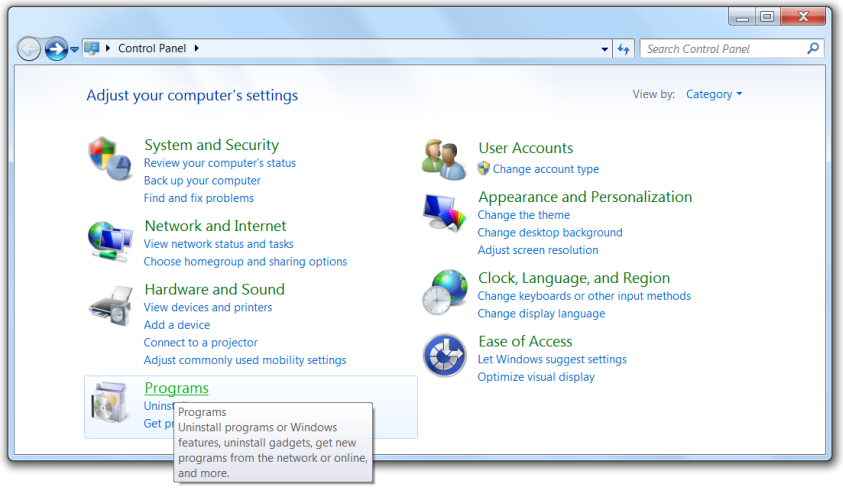


Figure 2. Click into the Programs category to find the Java control panel.

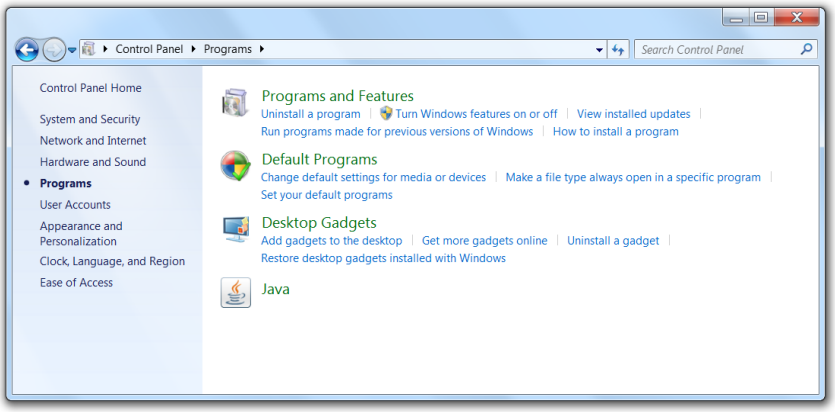


Figure 3. Locate the Java control panel.

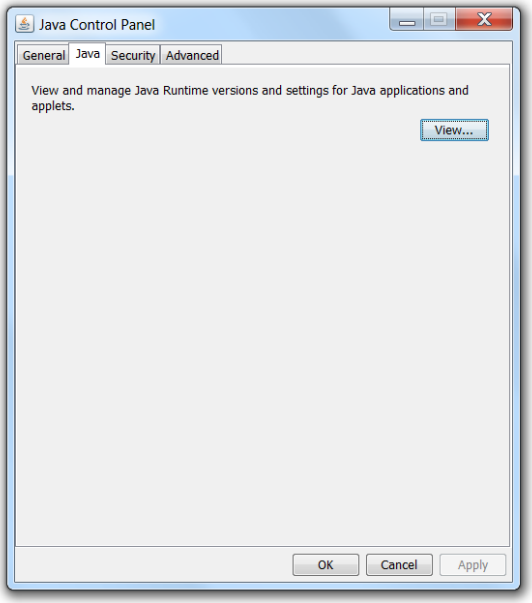


Figure 4. Java Control Panel. Choose the Java tab and press the View button to see more detail.

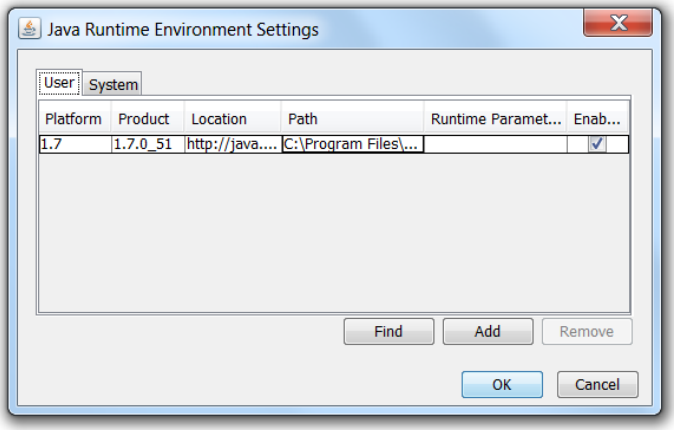


Figure 5. Java Runtime 1.7 is installed.

2.3 Do I have a 32-bit or 64-bit Computer?

You need to find out whether your computer has a 32-bit or a 64-bit operating system.

Go to Control Panel ►System and Security ►System. You will see various details about your computer. Under the System type, you will see whether you have a 32-bit or a 64-bit operating system.

Follow the link below for further instructions on how to install Java:

http://www.java.com/en/download/help/index_installing.xml

Once you have Java Runtime setup on your computer, you can proceed to install CiteSpace.

3 How to Install and Configure *CiteSpace*

CiteSpace is provided as a zip file for 64-bit and 32-bit computers. For Mac users, you need to download the 64-bit version.

3.1 Where Can I download CiteSpace from the Web?

You can download the latest version of CiteSpace from the following website:

http://cluster.ischool.drexel.edu/~cchen/citespace/download.html

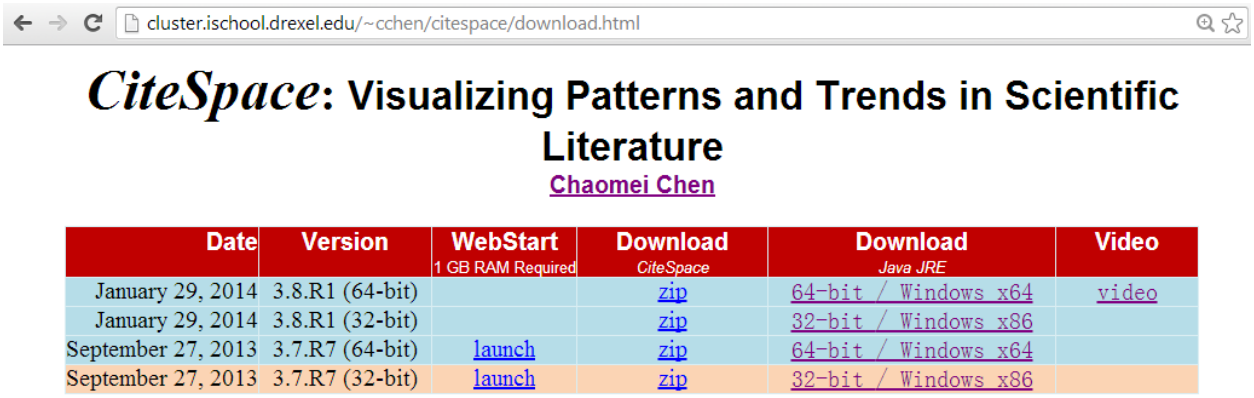


Figure 6.The download page of CiteSpace.

After you download the zip file to your computer, unpack the zip file to a folder of your choice.

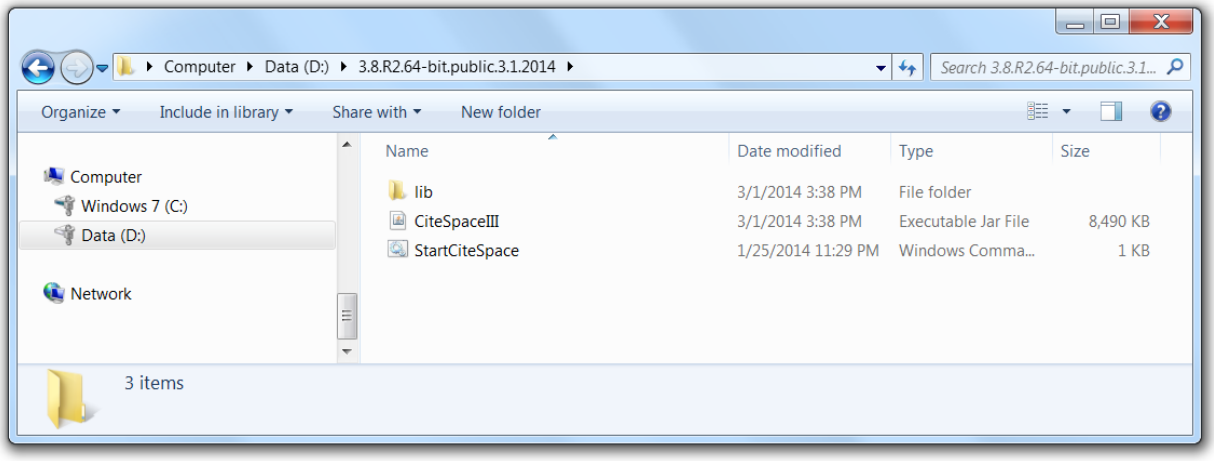


Figure 7. CiteSpace is unpacked to the D drive on a computer.

Now you can start CiteSpace by double clicking on the StartCiteSpace file.

If you need to modify the amount memory allocated for CiteSpace (more precisely for Java Virtual Machine on which CiteSpace to be running), you can edit StartCiteSpace as a plain text file with any text editor.

3.2 What is the maximum number of records that I can handle with CiteSpace?

This question needs to be answered at two levels: the number of records processed by CiteSpace and the number of nodes visualized, i.e. you can see and interact with them in CiteSpace.

The first number is the total number of records in your downloaded dataset. CiteSpace reads through each record in your download files.

The second number is determined by the selection criteria you specify and by the amount of memory, i.e. RAM, available on your computer. The more RAM you can make available for CiteSpace, the larger sized network you can visualize with a faster response rate.

The speed of processing is also affected by a few computationally expensive algorithms such as Pathfinder network scaling and cluster labeling. Empirically, the best options for Pathfinder network scaling would be 50~500 nodes per slice. With faster computers or if you can wait for a bit longer, you can raise the number accordingly.

The completion time of cluster labeling is related to the size of your dataset. If the entire timespan of your dataset is 100 years but you will only need to consider the most recent 10 years, it will be a good idea to carve out a much smaller dataset as long as it covers the 10 years of interest. It will reduce the processing time considerably.

3.3 How to configure the memory allocation for CiteSpace?

The performance of CiteSpace is influenced by the amount of memory accessible to the Java Virtual Machine (JVM) on which CiteSpace is running. To analyze a large amount of records, you should consider allocating as much as memory for CiteSpace to use.

You can modify the StartCiteSpace.cmd file to optimize the setting. More specifically, modify line 14 in the file. For example, -Xmx2g means that CiteSpace may get a maximum of 2GB of RAM to work with. Save the file after making any changes. And restart CiteSpace.

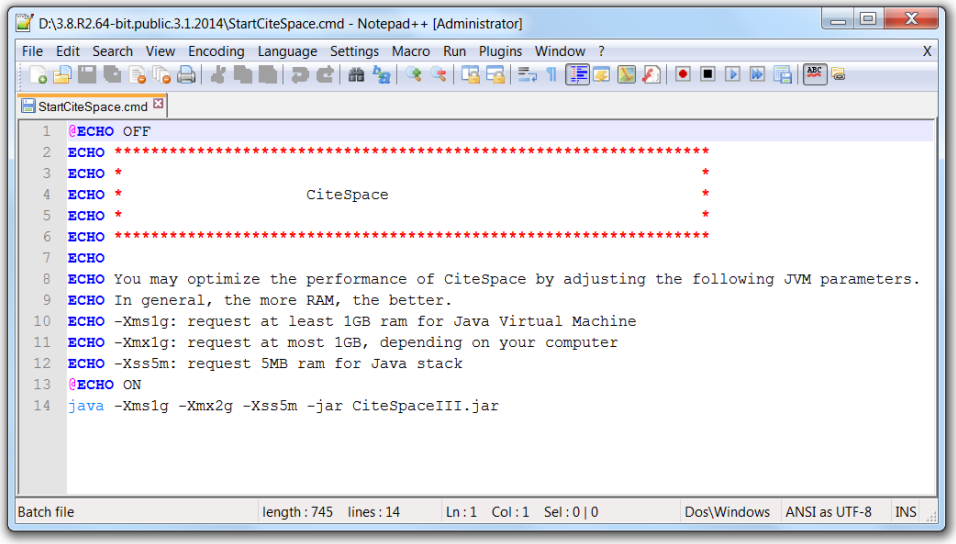


Figure 8. Configure the memory for Java in line 14.

3.4 How to uninstall CiteSpace

You can use the following steps to remove cached copies of CiteSpace from your computer.

Figure 9. In a Command Prompt window, type javaws –viewer.

When you see a list of cached copies of CiteSpace in the Java Cache Viewer, select the items that you want to remove and then click on the button with a red cross.

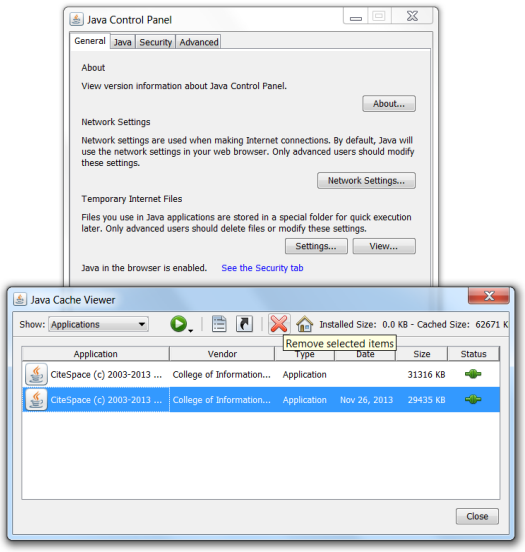


Figure 10. Select a cached copy of CiteSpace and remove the item.

4 Get Started with CiteSpace

4.1 Try it with a demonstrative dataset

When you installed CiteSpace for the first time, a demonstrative dataset on terrorism research is setup for you to play with and get familiar with the major analytic functions in CiteSpace.

If you have never used CiteSpace before, I strongly recommend you to start with this demo dataset.

To launch CiteSpace, double click on the StartCiteSpace.cmd file. You will see a command prompt window first. This window will also display various information on the status and any errors.

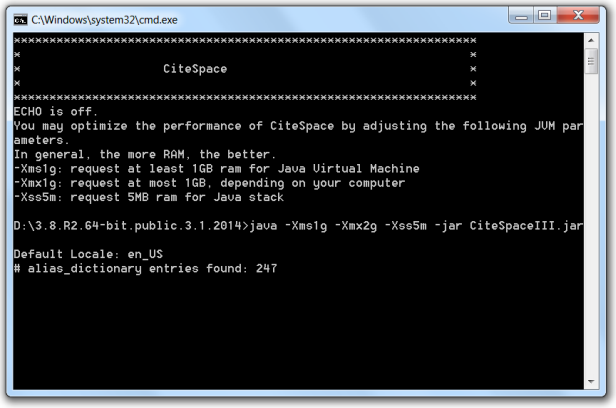


Figure 11. The command prompt window.

You will see another window of "About CiteSpace" – it displays system information of your computer, including the Java version.

To proceed, you need to click on the Agree button. CiteSpace may collect user driven events for research purposes.

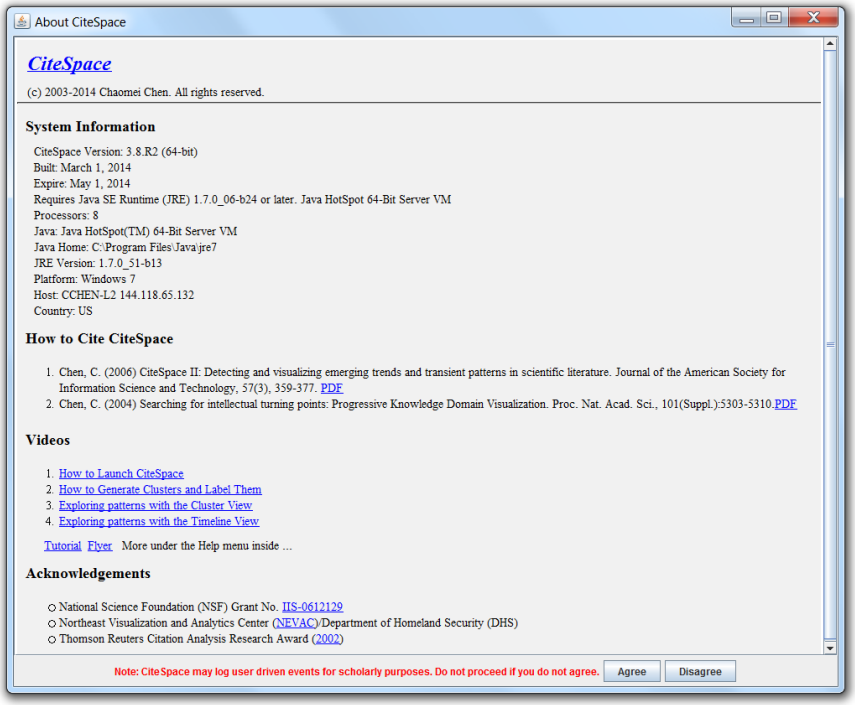


Figure 12. The "About CiteSpace" window. To proceed, click on the Agree button.

Next, you will see the main user interface of CiteSpace.

The user interface is divided into left and right halves. The left-hand side contains controls of projects (i.e. input datasets) and progress report windows. The right-hand side contains several panels for configuring the process with various parameters.

In a nutshell, the process in CiteSpace takes an input dataset specified in the current project, constructs network models of bibliographic entities, and visualizes the networks for interactive exploration for trends and patterns identified from the dataset.

The demo project contains a dataset on publications about terrorism research. These bibliographic records were retrieved from the Web of Science. See later sections on tips for how to construct your own dataset.

4.1.1 The Demo Project

We will start the process and explain how CiteSpace is designed to help you answer some of the key questions about a knowledge domain, i.e. a field of study, a research area, or a set of publications defined by the user.

Press the green GO! button to start the process.

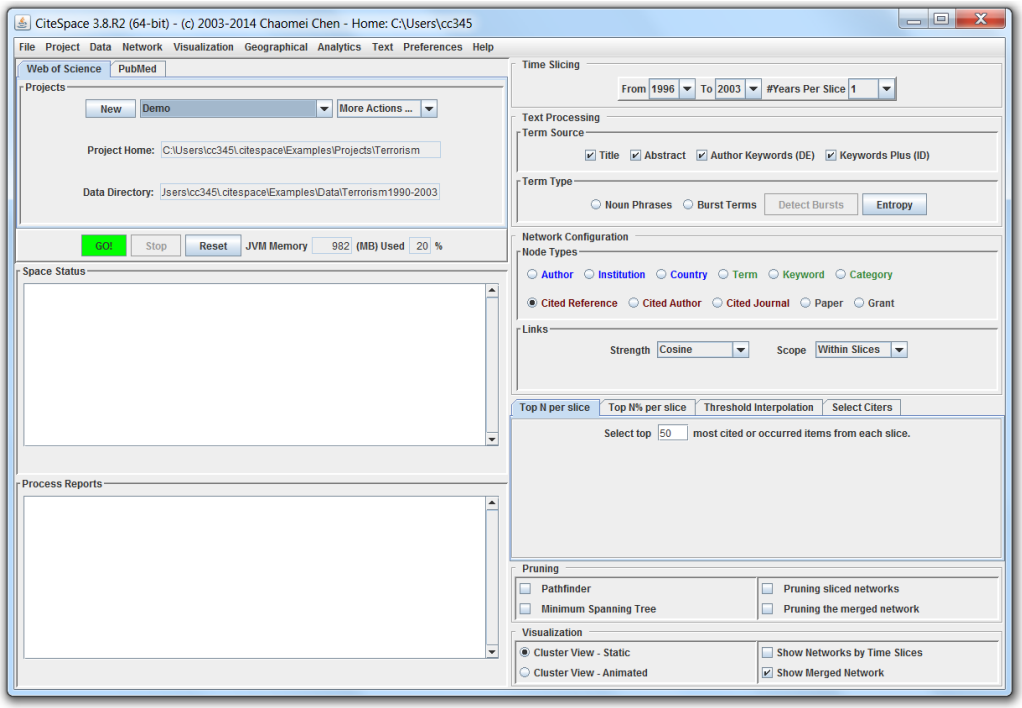


Figure 13. The main user interface of CiteSpace.

CiteSpace will read the data files in the current project (Demo) and report its progress in the two windows on the left-hand side of the user interface. When the modeling process is completed, you have three options to choose: Visualize, Save As GraphML, or Cancel.

Visualize:

This option will take you to the visualization window for further interactive exploration.

Save As GraphML:

This option will save the constructed network in a file in a common graph format. No visualization.

Cancel:

This option will not generate any interactive visualization nor save any files. It allows you to reconfigure the process and re-run the process.

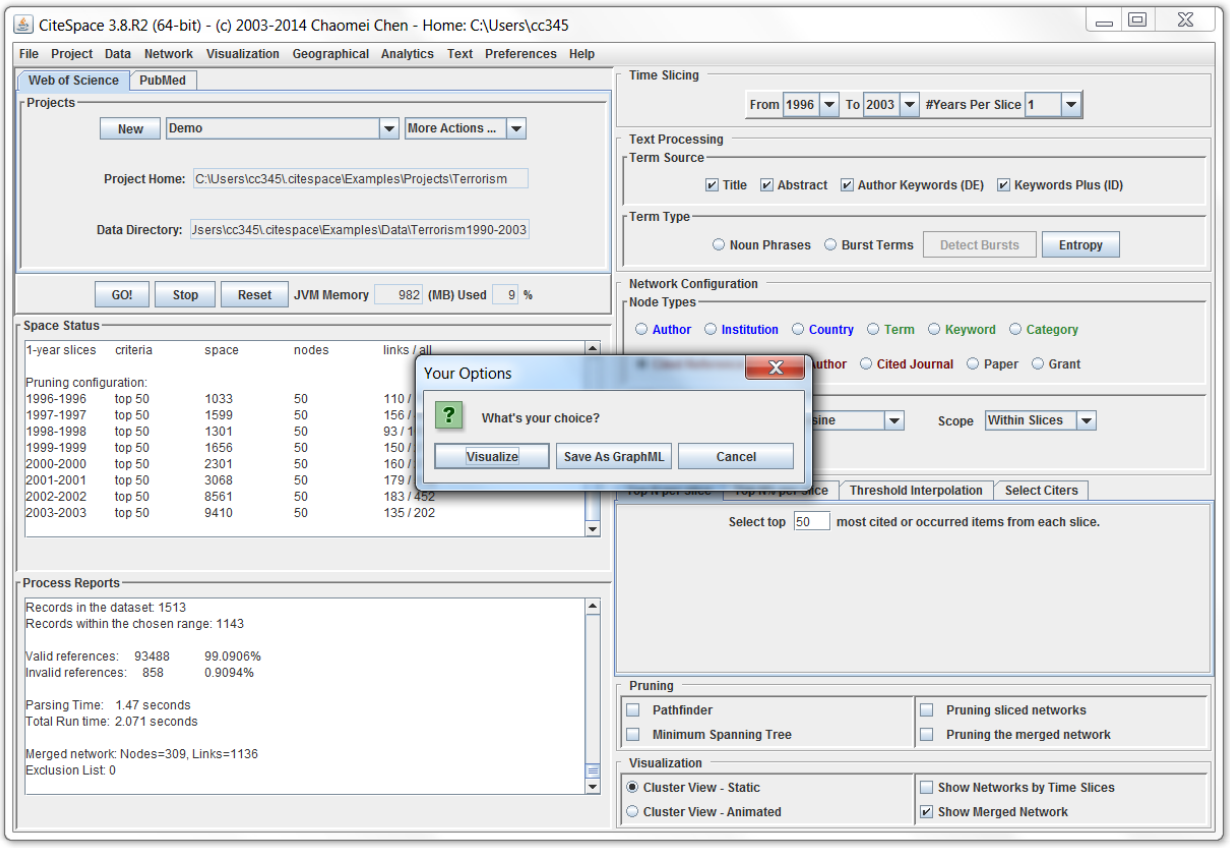


Figure 14. CiteSpace is ready to visualize the constructed network.

If you click on the Visualize button, a new window will pop up. This is the Visualization Window. Initially you will see some movements on your screen with a black background. Once the movements are settled, the background color turns to white.

Let's focus on what the initial visualization tells us and then explore what else we can find by using additional functions.

First, CiteSpace visualizes a merged network based on several networks corresponding to snapshots of consecutive years. In the Demo project example, the overall time span is from 1996 through 2003. The merged network characterizes the development of the field over time, showing the most important footprints of the related research activities. Each dot represents a node in the network. In the Demo case, the nodes are cited references. CiteSpace can generate networks of other types of entities. Here let's focus on cited references only for now. Lines that connect nodes are co-citation links; again, CiteSpace can generate networks of other types of links. The colors of these lines are designed to show when a connection was made for the first time. Note that this is influenced by the scope and the depth of the given dataset.

The color encoding makes it easy for us to tell which part of the network is old and which is new.

If you see that some references are shown with labels, then you will know that these references are highly cited, suggesting that they are probably landmark papers in the field. A list on the left side of the window shows all the nodes appeared in the visualization. The list can be sorted by the frequency of citations, Betweenness centrality, or by year or references as text. You can also choose to show or hide a node on the list.

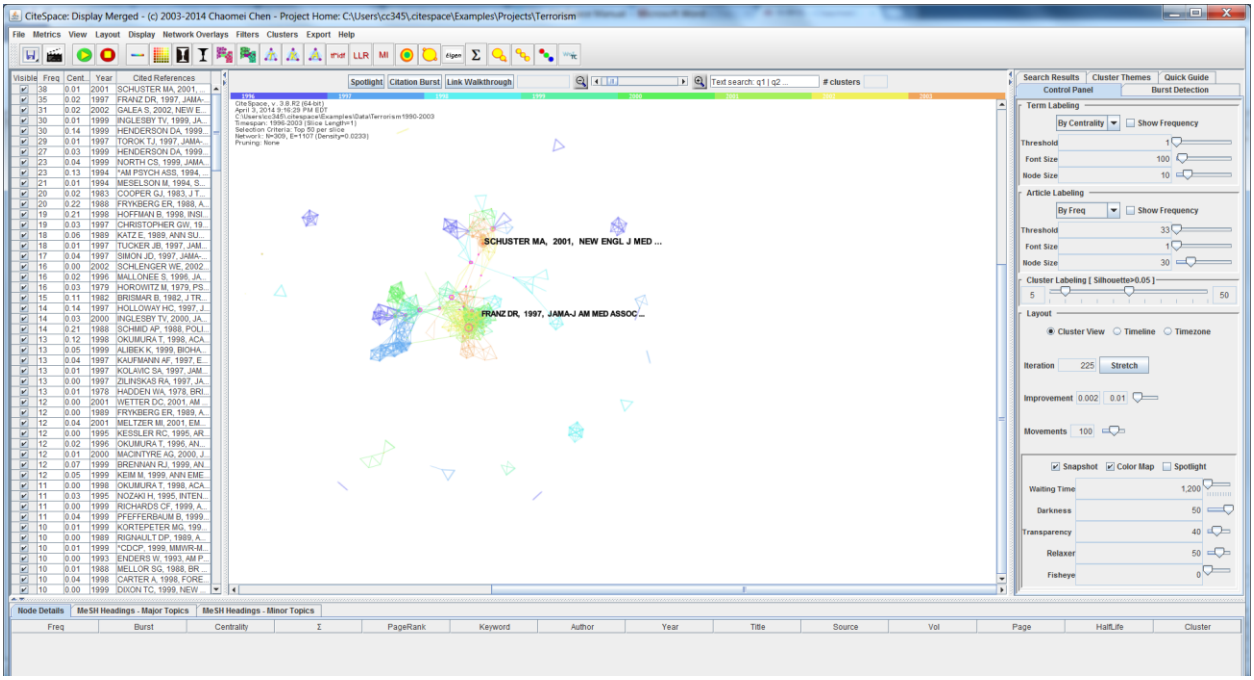


Figure 15. The Visualization window.

A control panel is shown on the right-hand side of the Visualization Window. You can change how node labels are displayed by a combination of a few threshold values through sliders. You can also change the size of a node by sliding the node size slider.

To answer the typical questions we asked before, let's use several functions in CiteSpace to obtain more specific information through clustering, labeling, and exploring.

4.1.2 Clustering

Although we can probably eyeball the visualized network and identify some prominent groupings, CiteSpace provides more precise ways to identify groupings, or clusters, using the clustering function.

Figure 16. Most frequently used functions for visual exploration in CiteSpace.

To start the clustering function, simply click on this icon  .

How do I know whether the clustering process is completed? You will see #clusters on the upper right corner of the canvas. In the Demo example, a total of 37 clusters of co-cited references are identified. Each cluster corresponds to an underlying theme, a topic, or a line of research.

The signature of the network is shown on the upper left corner of the display. In particular, the modularity Q and the mean silhouette scores are two important metrics that tell us about the overall structural properties of the network. For example, the modularity Q of 0.7141 is relatively high, which means that the network is reasonably divided into loosely coupled clusters. The mean silhouette score of 0.5904 suggests that the homogeneity of these clusters on average is not very high, but not very low either.

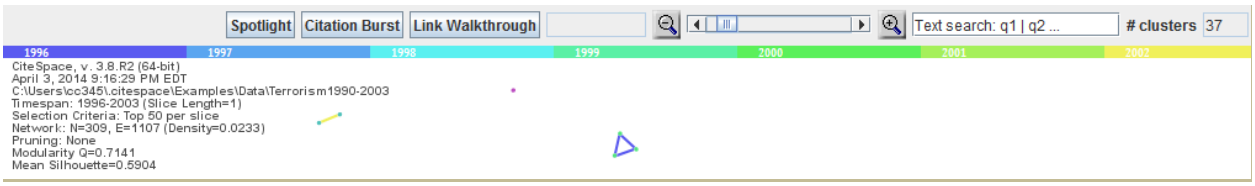
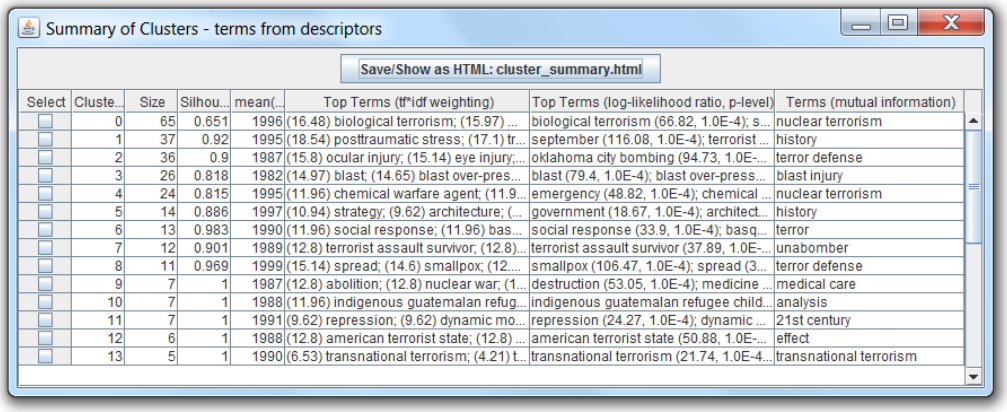


Figure 17. The clustering process is completed. 37 clusters are identified (#clusters shown in the upper right corner). Modularity and silhouette scores are shown in the signature of the network on the left.

You can inspect various measures of each cluster in a summary table of all the clusters using: Clusters ►4. Summarization of Clusters. The Silhouette column shows the homogeneity of a cluster. The higher the silhouette score, the more consistent of the cluster members are, provided the clusters in comparison have similar sizes. If the cluster size is small, then a high homogeneity does not mean much. For example, cluster #9 has 7 members and a silhouette of 1.00, this is most likely due to the possibility that all 7 references are the citation references of the same underlying author. In other words, cluster #9 may reflect the citing behavior or preferences of a single paper, thus it is less representative.

The average year of publication of a cluster indicates whether it is formed by generally recent papers or old papers. This is a simple and useful indicator.



Select	Cluste...	Size	Silhou...	mean(...	Top Terms (tf*idf weighting)	Top Terms (log-likelihood ratio, p-level)	Terms (mutual information)
	0	65	0.651	1996	(16.48) biological terrorism; (15.97) ...	biological terrorism (66.82, 1.0E-4); s...	nuclear terrorism
	1	37	0.92	1995	(18.54) posttraumatic stress; (17.1) tr...	september (116.08, 1.0E-4); terrorist ...	history
	2	36	0.9	1987	(15.8) ocular injury; (15.14) eye injury;...	oklahoma city bombing (94.73, 1.0E-...	terror defense
	3	26	0.818	1982	(14.97) blast; (14.65) blast over-pres...	blast (79.4, 1.0E-4); blast over-press...	blast injury
	4	24	0.815	1995	(11.96) chemical warfare agent; (11.9...	emergency (48.82, 1.0E-4); chemical ...	nuclear terrorism
	5	14	0.886	1997	(10.94) strategy; (9.62) architecture; (...	government (18.67, 1.0E-4); architect...	history
	6	13	0.983	1990	(11.96) social response; (11.96) bas...	social response (33.9, 1.0E-4); basq...	terror
	7	12	0.901	1989	(12.8) terrorist assault survivor; (12.8)...	terrorist assault survivor (37.89, 1.0E-...	unabomber
	8	11	0.969	1999	(15.14) spread; (14.6) smallpox; (12....	smallpox (106.47, 1.0E-4); spread (3...	terror defense
	9	7	1	1987	(12.8) abolition; (12.8) nuclear war; (1...	destruction (53.05, 1.0E-4); medicine ...	medical care
	10	7	1	1988	(11.96) indigenous guatemalan refug...	indigenous guatemalan refugee child...	analysis
	11	7	1	1991	(9.62) repression; (9.62) dynamic mo...	repression (24.27, 1.0E-4); dynamic ...	21st century
	12	6	1	1988	(12.8) american terrorist state; (12.8) ...	american terrorist state (50.88, 1.0E-...	effect
	13	5	1	1990	(6.53) transnational terrorism; (4.21) t...	transnational terrorism (21.74, 1.0E-4...	transnational terrorism

Figure 18. A summary table of clusters.

4.1.3 Generate Cluster Labels

To characterize the nature of an identified cluster, CiteSpace can extract noun phrases from the titles (T in the following icon), keyword lists (K), or abstracts (A) of articles that cited the particular cluster.

Let's ask CiteSpace to choose noun phrases from titles (i.e. select the T icon). This process may take a while as CiteSpace needs to compute several selection metrics. Once the process is finished, the chosen labels will be displayed. By default, labels based on one of the three selection algorithms will be shown, namely, tf*idf. Our study has found that LLR usually gives the best result in terms of the uniqueness and coverage.

Figure 19. Icons for performing Clustering and Labeling functions.

Cluster labels are displayed once the process is completed. The clusters are numbered in the descending order of the cluster size, starting from the largest cluster #0, the second largest #1, and so on.

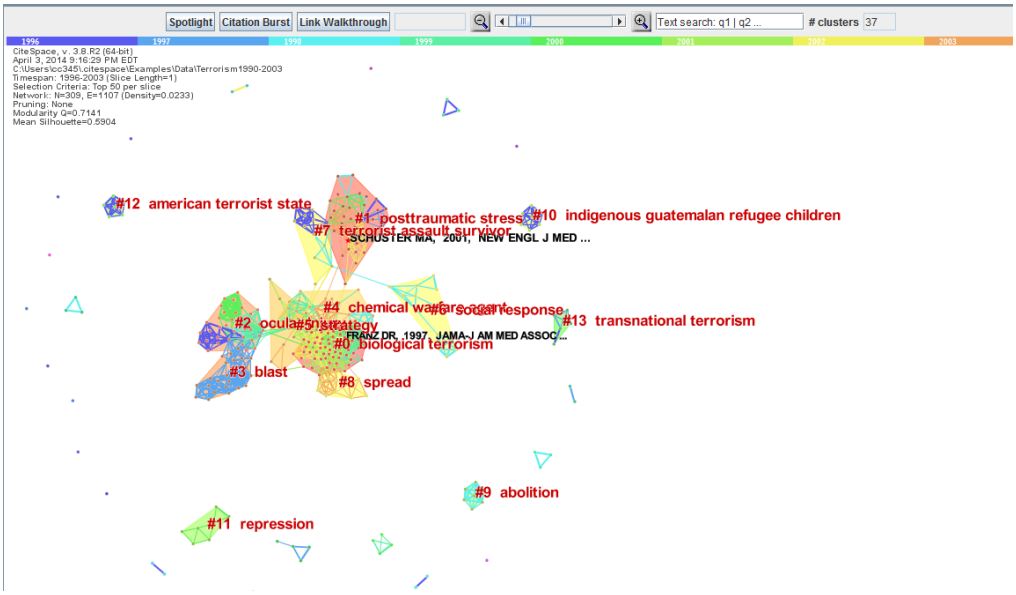


Figure 20. Cluster labels are generated and displayed.

To make it easier to see which clusters are the largest, you can choose to change the font size of the labels from the uniformed to proportional:

Display ►Label Font Size ►Cluster: Uniformed/Proportional

This is a toggle function. That means there are two states. Your selection will switch back and forth between the two states, i.e. either using a uniformed font size or proportional.

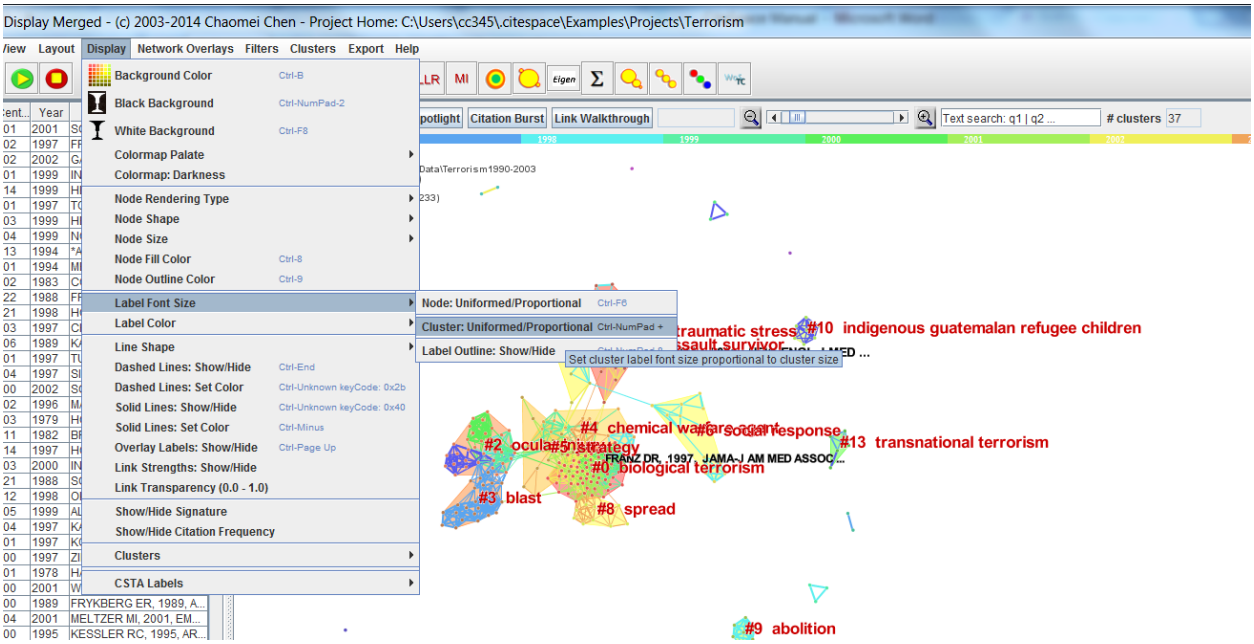


Figure 21. Set the cluster labels' font size proportional to their size.

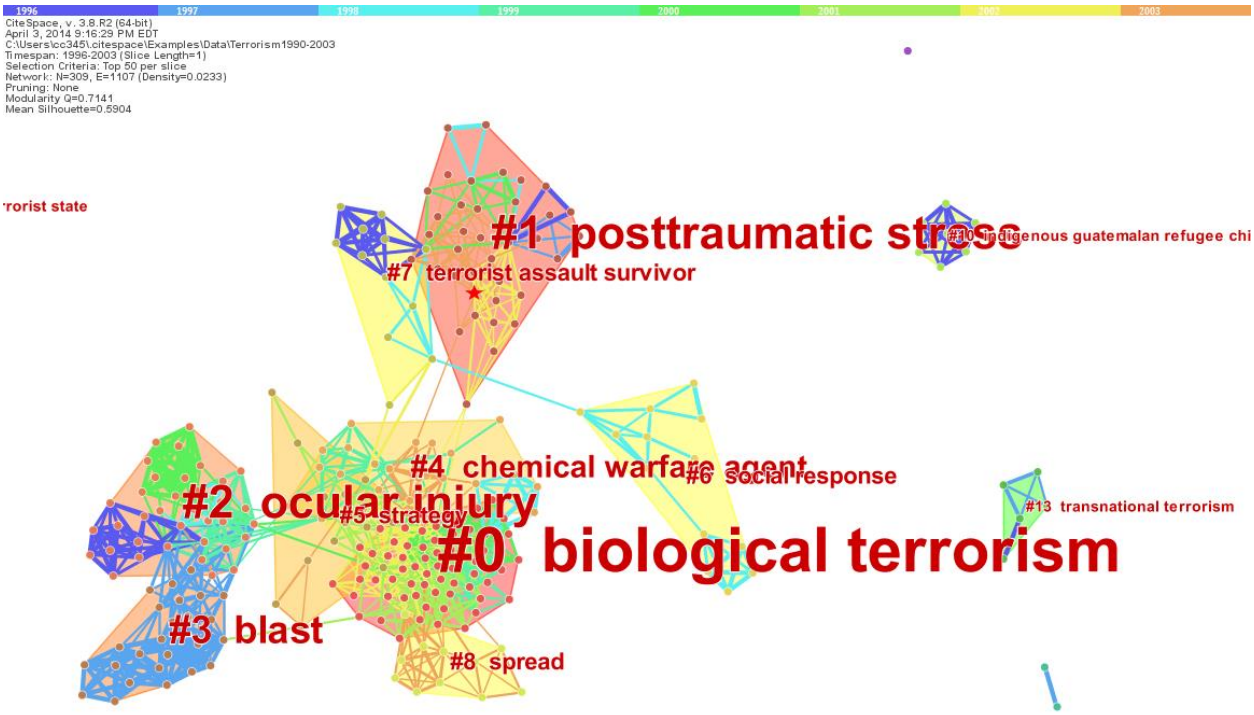


Figure 22. Cluster labels' font sizes are proportional to the size of a cluster. The largest cluster is #0 on biological terrorism.

4.1.4 Where are the major areas of research based on the input dataset?

This is one of the primary questions that CiteSpace is designed to answer. To answer this question, we will focus on the big picture of the collection of publications represented by your dataset. Let's make a few adjustments with the sliders in the control panel on the right so that the information of our interest will be shown clearly and information that is less relevant to this question right now will be temporarily hidden from the view.

1. *Node Size*

At this level, we don't really need to see the size of a node, although it provides rich information about the history of a node. Use the slider under Article Labeling ►Node Size ►[Slide to 0] (marked by the pointer #1 in the following figure).

2. *Cluster Label Size*

The font size of the cluster labels are controlled by a slider with two controls: one control the threshold for showing or hiding a label based on the size of the cluster (i.e. to make sure large-enough clusters are always labeled), and the other control the font size of the cluster labels (marked by the pointer #2 in the screenshot).

3. *Transparency of Links*

Detailed links would be useful later, but we can ignore them for now using the transparency slider to set all the links' transparency to the lowest level, i.e. invisible. In hindsight, a more accurate term would be completely transparent.

After making these minor adjustments, it will be straightforward to answer the question: Where are the major areas of research? Evidently, the largest area (cluster #0 with the largest number of

member references) is biological terrorism. The second largest is posttraumatic stress (cluster #1), i.e. PTSD. The third one is ocular injury (cluster #2). The fourth one is blast (cluster #3). And there are a few smaller clusters. So now we have a general idea what constituted terrorism research during the period of 1996 and 2003. You can repeat the process on a current dataset to get an up-to-date big picture.

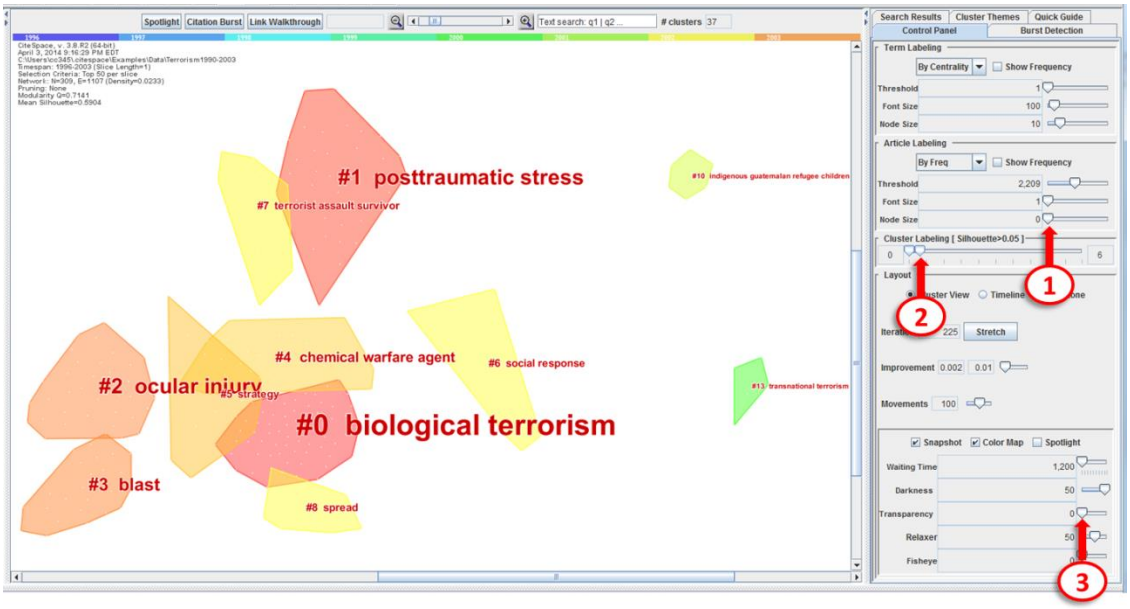


Figure 23. Adjust the appearance of the visualization with a few sliders. Pointers: 1) Node size control slider, 2) cluster label size, and 3) transparency of links.

4.1.5 How are these major areas connected?

To answer this question, we need to bring back the lines connecting nodes. Adjust the transparency slider to make the lines visible.

A useful indicator of how different clusters are connected is a type of nodes that have high betweenness centrality scores. In CiteSpace, betweenness centrality scores are normalized to the unit interval of [0, 1]. A node of high betweenness centrality is usually one that connects two or more large groups of nodes with the node itself in-between, hence the term betweenness. CiteSpace highlights nodes with high betweenness centrality with purple trims. The thickness of a purple betweenness centrality trim indicates how strong its betweenness centrality is. The thicker the stronger. Occasionally, a node with high betweenness centrality may appear at the center of a network component, but our interest is in the nodes that are truly in between.

To make see the purple rings, switch the node rendering mode to tree rings, which is the first icon shown in the following figure, i.e. concentric citation rings represent how many citations were made to the node in corresponding years. Remember that colors represent when citations were actually made.

Figure 24. Icons of node rendering controls.

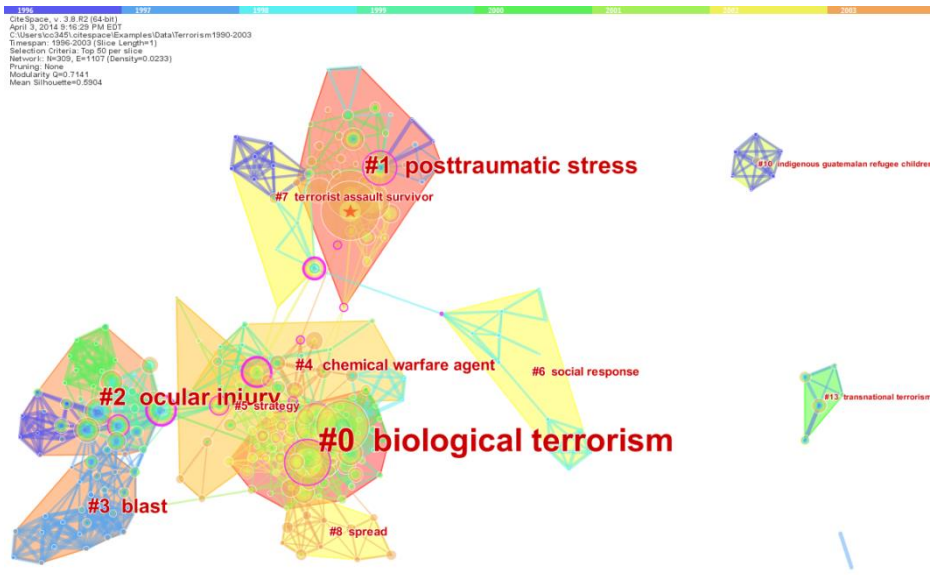


Figure 25. Nodes with purple rings are important in connecting different clusters.

4.1.6 Where are the most active areas?

Citation burst is an indicator of a most active area of research. Citation burst is a detection of a burst event, which can last for multiple years as well as a single year. A citation burst provides evidence that a particular publication is associated with a surge of citations. In other words, the publication evidently has attracted an extraordinary degree of attention from its scientific community. Furthermore, if a cluster contains numerous nodes with strong citation bursts, then the cluster as a whole captures an active area of research, or an emerging trend.

The burst detection in CiteSpace is based on Kleinberg's algorithm (Kleinberg, 2002).

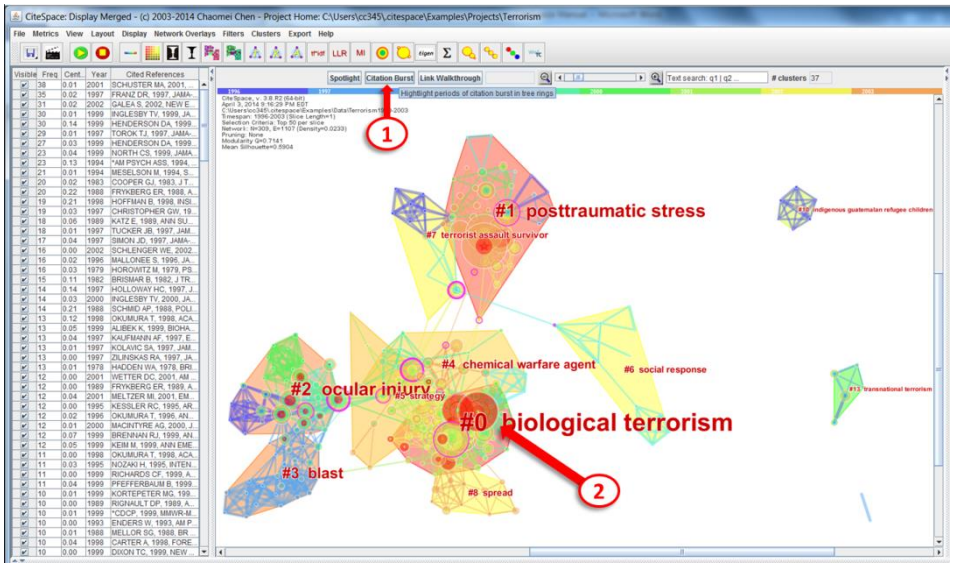


Figure 26. Citation bursts are indicators of most active areas.

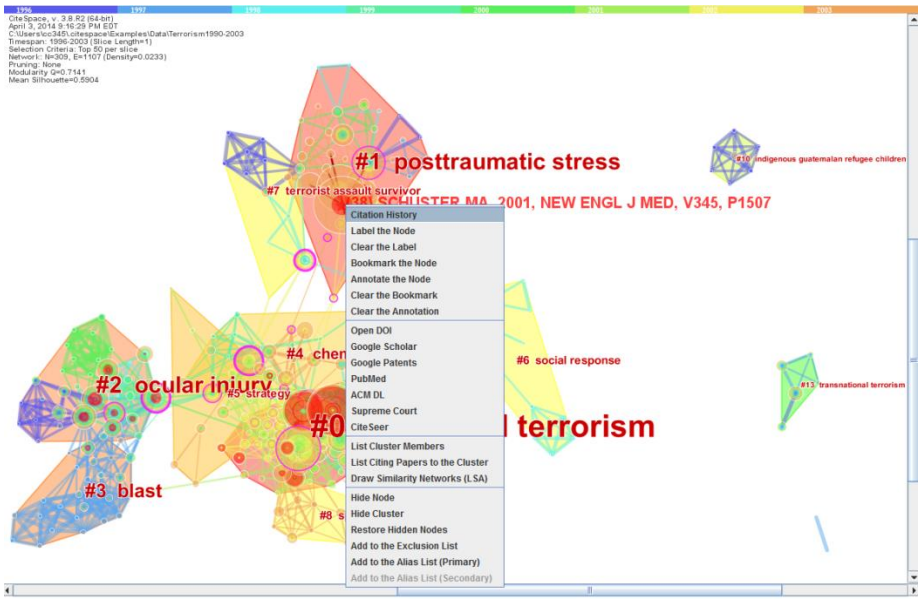


Figure 27. Right click on the node of interest and choose the Citation History of the node.

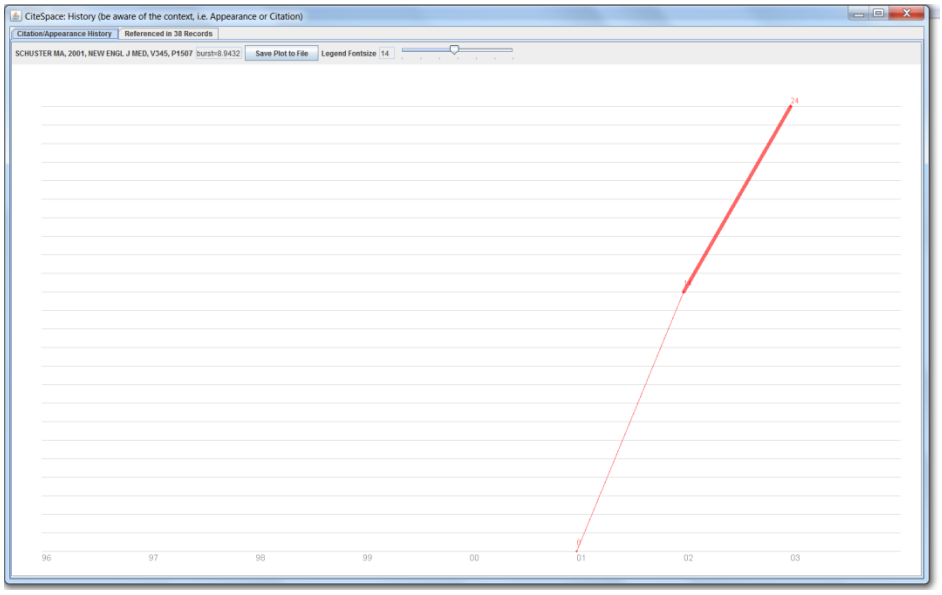


Figure 28. This is the citation history of an article that has a citation burst.

Using View ►Citation Burst History can generate a summary list of articles that are associated with citation bursts. This visualization shows which references have the strongest citation bursts and which periods of time the strongest bursts took place. For example, from the list, we can tell that Schuster et al. (2001) has the strongest bursts among articles published since terrorist attacks in 2001. It is also interesting to note that North et al. (1999) has the second strongest citation burst in the period of 2002 and 2003.

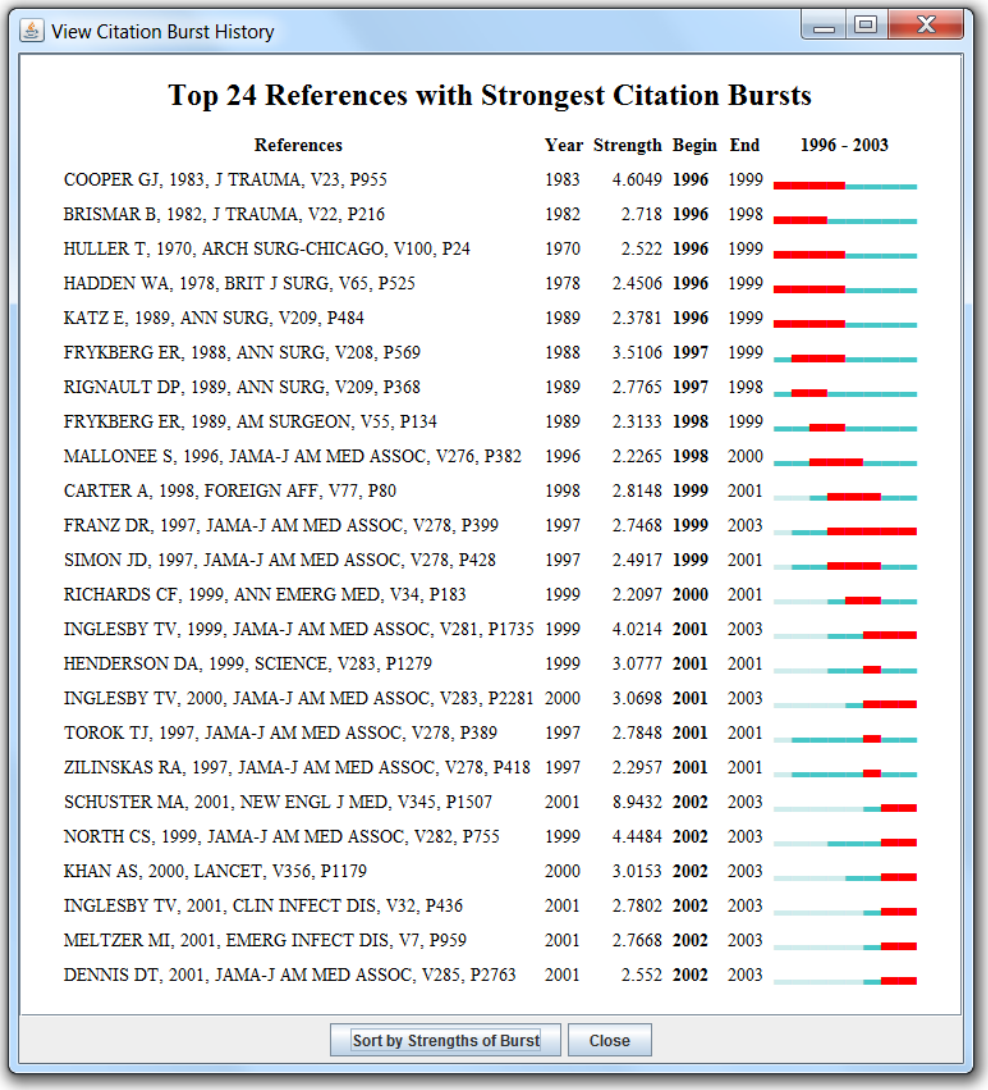


Figure 29. A summary list of references with citation bursts.

4.1.7 What is each major area about? Which/where re the key papers for a given area?

Cluster labels can tell us the context in which they are most cited because the label terms are extracted from citing articles' titles, keywords, or abstracts.

To explore these clusters in more depth, you should use the Cluster Explorer:

Clusters ►Cluster Explorer

The initial appearance of the Cluster Explorer shows four windows: 1) Clusters, 2) Citing Articles, 3) Cited References, and 4) Representative Sentences. Windows 2-3 are blank until you select a cluster in the Clusters window by checking the checkbox in front of each row of cluster information.

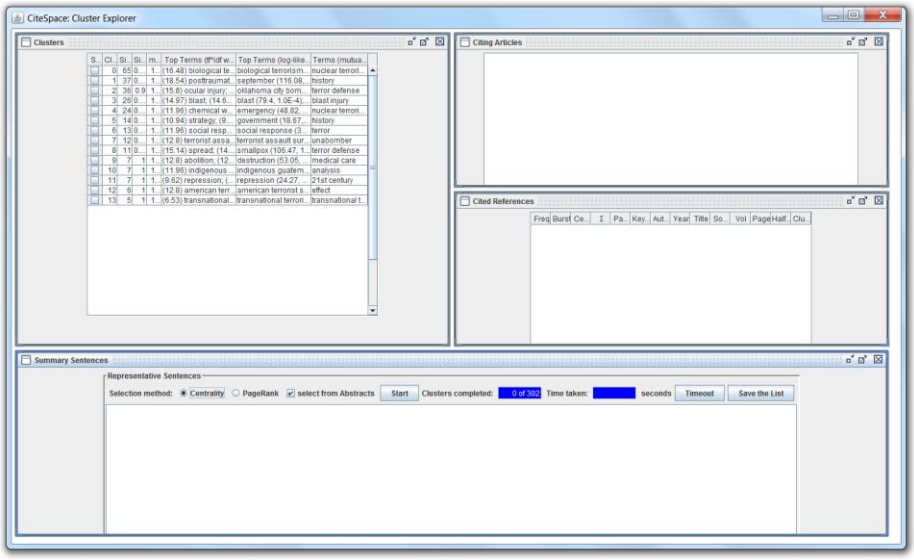


Figure 30. The initial appearance of the Cluster Explorer.

The following figure shows a screenshot after Cluster #0 was selected in the checkbox. As you can see, the Citing Articles window and the Cited References window are both populated accordingly. In the Citing Articles window, each entry is a citing paper, i.e. a paper that cites members of the cluster. The number in front of each entry shows the portion of the references cited by this particular article out of all the references in total. For example, Bak, SJ (2000) has a coverage of 0.28, i.e. 28% of the total 65 references in this cluster (you can find the 65 listed in the Clusters window's third column – Size).

The phrase biological terrorism was highlighted in yellow in the Citing Articles window. Note that the phrase is also the label of this cluster in the visualization. Furthermore, the phrase also appears in the Clusters windows' 7th column – Top Terms (log-likelihood ratio). For technical details, see (C. Chen et al., 2010).

The Cited References window shows the member references of this cluster. Each reference is listed with the number of citations, burstness if any, its centrality score, along with the name of the first author, the year of publication, source (i.e. journal or conference), volume number, and page number.

Regarding the centrality score, if the network size is greater than 350, then you will need to manually start the process to calculate the centrality scores using Metrics ►Compute Centrality.

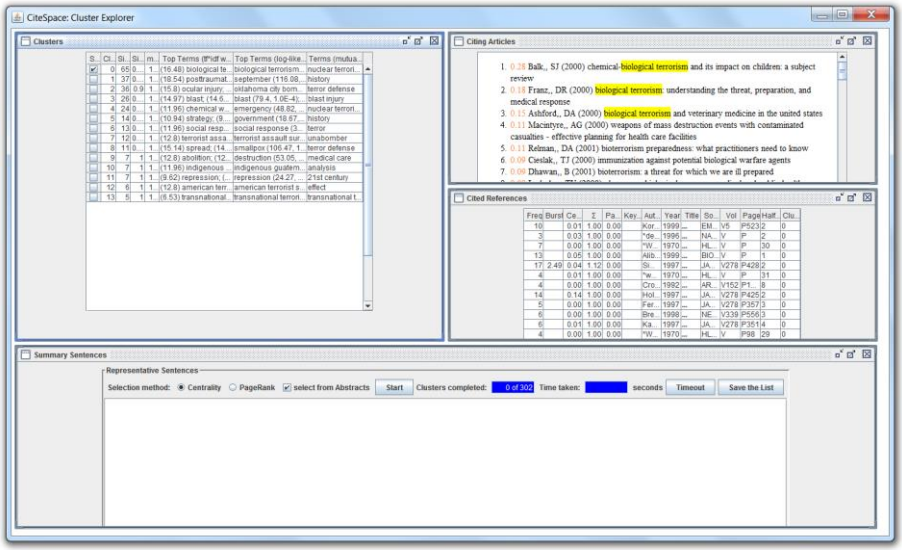


Figure 31. Cluster Explorer: Cluster #0 is selected in the checkbox.

In the Summary Sentences window, if you click on the Start button, CiteSpace will extract the most representative sentences from the abstracts of the citing articles to each cluster. A sentence is considered representative if it is either a sentence with a high degree centrality or a sentence with a high PageRank score.

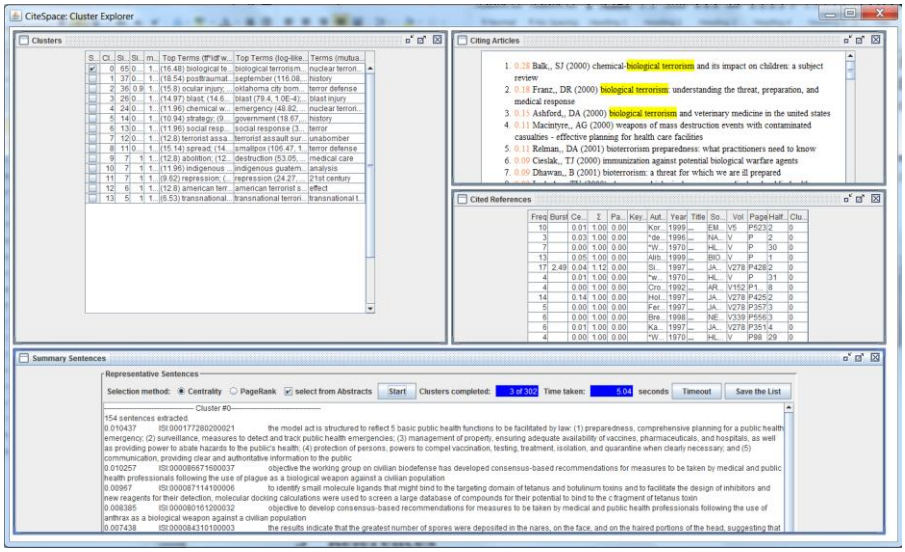


Figure 32. Representative sentences are displayed upon clicking on the Start button in the Summary Sentences window.

4.1.8 Timeline View

You can switch to a timeline view of the network by choosing the Timeline radio button in the Layout panel on the right (as pointed by the red arrow in the following figure). In a timeline view, each cluster is arranged on a horizontal timeline. The direction of time points to the right.

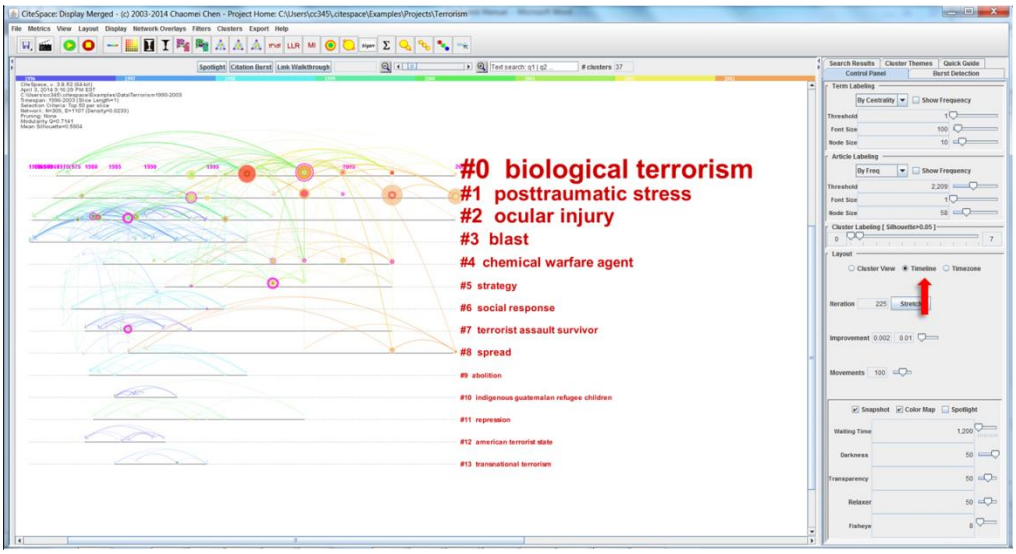


Figure 33. A timeline view of the network.

You have seen some of the basic moves. CiteSpace has many other features. We will introduce other features at more advanced levels.

4.2 Try it with a dataset of your own

4.2.1 Collecting Data

4.2.1.1 How to construct my own data from the Web of Science

The primary source of data for CiteSpace is the Web of Science.

Most importantly, the dataset should include cited references in order to maximize the potential of CiteSpace.

The Web of Science has several ways to search for bibliographic records. The most basic one is called, of course, basic search, which includes topic, author, and several other searchable fields. The following example shows a topic search for "CiteSpace" between the timespan of 2004 and 2014.

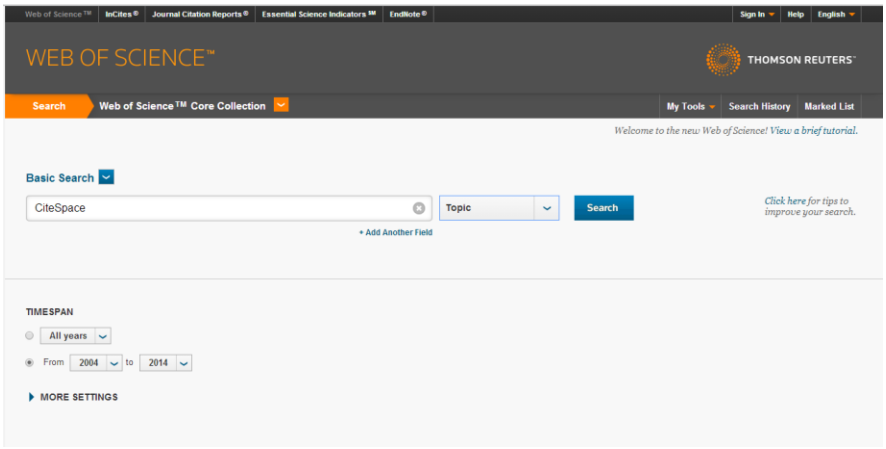


Figure 34. A topic search in the Web of Science.

The topic search found 16 results. The results are initially displayed in the chronological order of the publication date from the newest to the oldest. You can switch to a different order, for example, by the number of citations, from the highest to the lowest, so you can quickly narrow down to a small subset of the most highly cited records.

Figure 35. Sort the results by Times Cited – highest to lowest.

You will notice if the results are sorted by Times Cited – highest to lowest. The record with the highest times cited is the 2006 JASIST paper on CiteSpace II, with 185 citations. The topic search found 16 records. You can download these 16 records, however, that would be not representative. If you follow the Create Citation Report link, you will see you can expand the 16 records to about 220 records that cited the set of 16 records. We refer to this way to obtain more potentially relevant records as citation expansion. Since the only thing we know is that each record in the expanded set at least cited one of the original 16 records, it may turn out to be a less relevant record because of the diversity of how authors cite. Let's if we can do better than finding 220 records related by citation indexing.

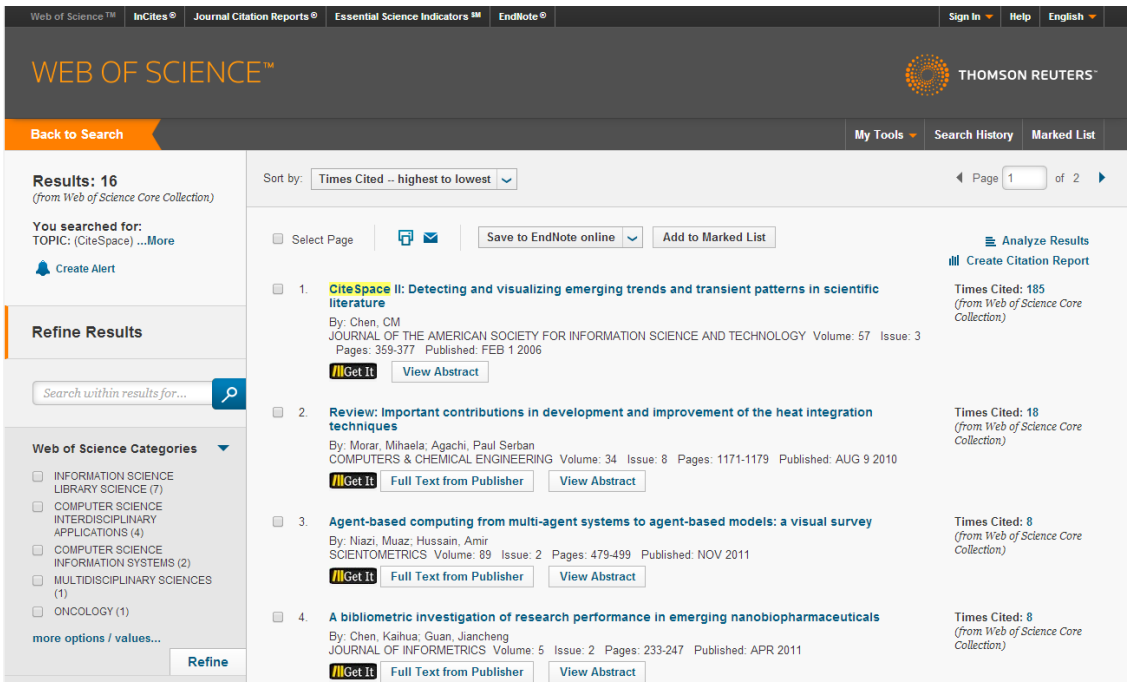


Figure 36. Results are now sorted by Times Cited from the highest to the lowest.

You may also notice that the 2004 PNAS and the 2010 JASIST paper on CiteSpace were NOT on the list, although they are certainly about CiteSpace and their citations would put them on the list too. Thus, this example shows that you should be careful when using the topic search along to construct your own dataset.

Under the Citation Network panel, the 104 Times Cited is a clickable link. If you click on it, it will bring you to the list of 104 records that cited the 2004 PNAS paper. The 2006 JASIST paper should be on the list. If we sort the list by Times Cited, then we will see the 2006 JASIST on the top.

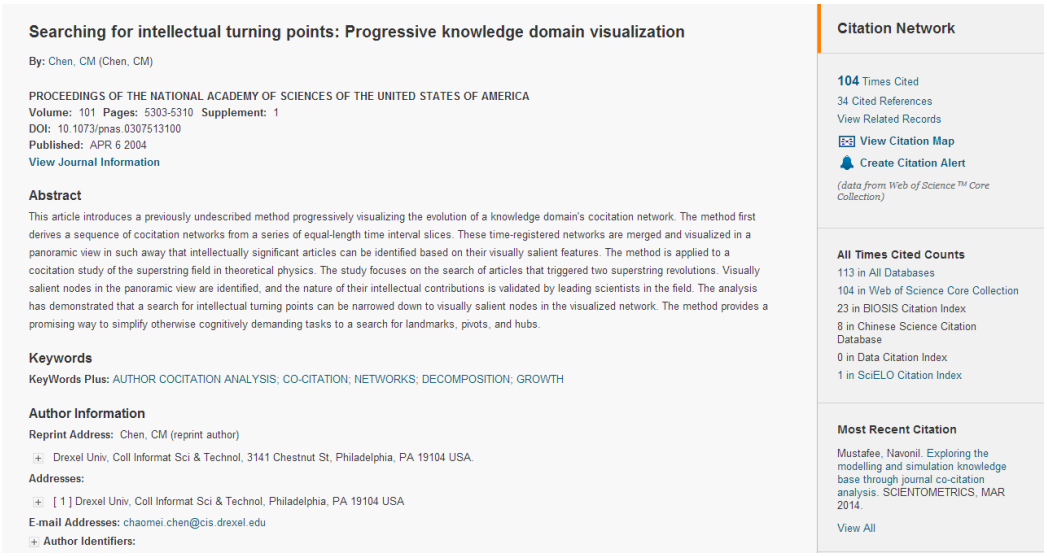


Figure 37. The 2004 PNAS paper is cited 104 times, but the topic search won't be able to find it because the term CiteSpace does not appear in its title, abstract, or the keywords.

Now if you click on the Create Citation Report on the right, you will get access to all the records that citing this lot, i.e. that would be the citation expansion we want.

Figure 38. Citing articles to the 2004 PNAS paper.

The Citation Report shows, among other things, 732 citing articles. These 732 articles would form the expanded set. In fact, you can go even further by adding your search results to the Marked List ►Create Citation Report ►Citing Articles. I will leave it to you to explore in the Web of Science.

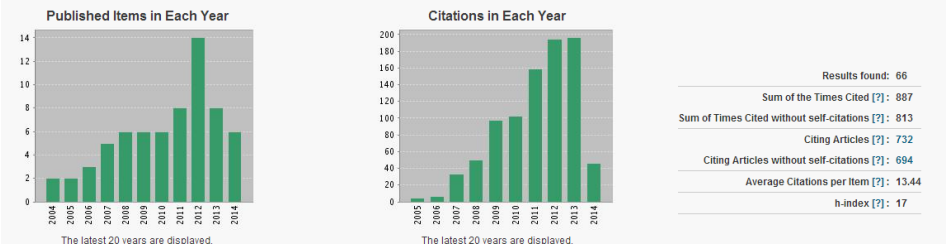


Figure 39. 732 Citing Articles will constitute the expanded set to download.

4.2.1.2 Download Records to Files

To download a set of records from the Web of Science, pull down the menu starting with Save to EndNote online and select Save to Other File Formats.

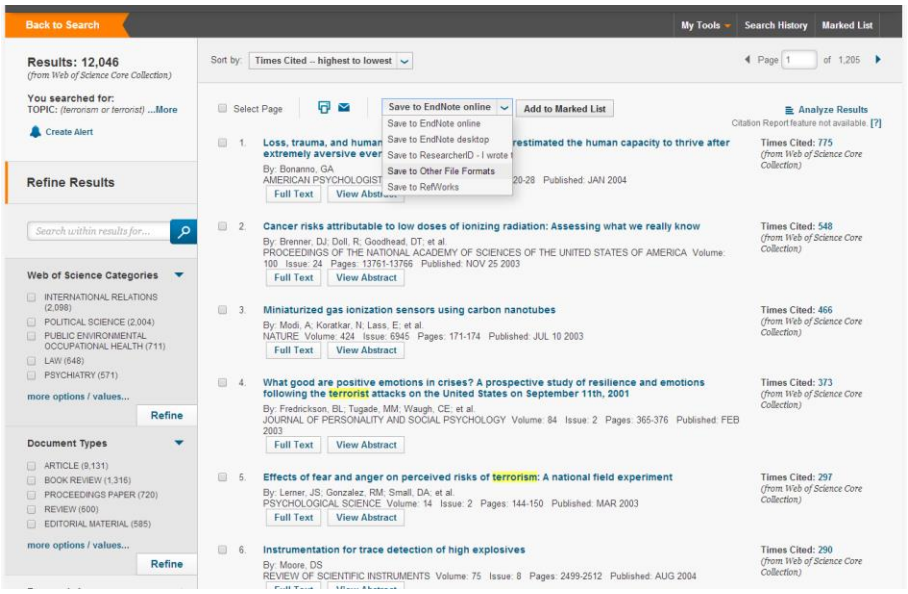


Figure 40. How to save records to other file formats.

Then you will need to enter the number of records, the content, and the file format in a dialog box like the following. For CiteSpace, include Full Record and Cited References and select Pain Text as the file format. When you save the file, make sure the file name starts with the word 'download' and the file extension is .txt. This naming convention will bring your more flexibility later on. For example, you can easily hide a file from CiteSpace by adding a prefix to the names of a few files you want CiteSpace to skip.

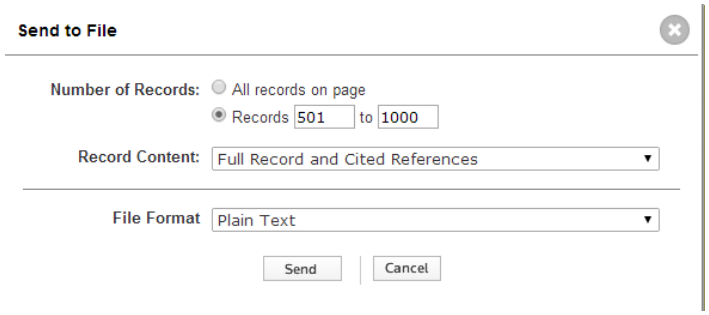


Figure 41. Download records 501-1000 in Plain Text. The Web of Science allows the maximum of 500 records each time to download. You may need to repeat the step multiple times.

4.2.2 Working with a CiteSpace Project

A CiteSpace project is designed to facilitate your analysis. Each project is associated with a dataset. You may analyze the dataset in many ways by selecting a variety of parameters and project properties. CiteSpace generates several types of intermediate files that you may want to inspect them in detail. You can handle most of these intermediate files directly.

4.2.2.1 Create a CiteSpace Project

You need to create two separate folders for a new project. One folder contains data files you just downloaded. We refer to it as the data folder. The other folder is the project folder, which will be used to store various intermediate files.

4.2.2.2 Edit an Existing Project

You can edit the properties of an existing project. To choose this function, full down the menu that shows "More Actions" next to the current project.

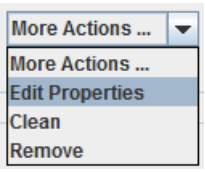


Figure 42. Edit the properties of an existing project.

You can edit several properties of an existing project based on your needs.

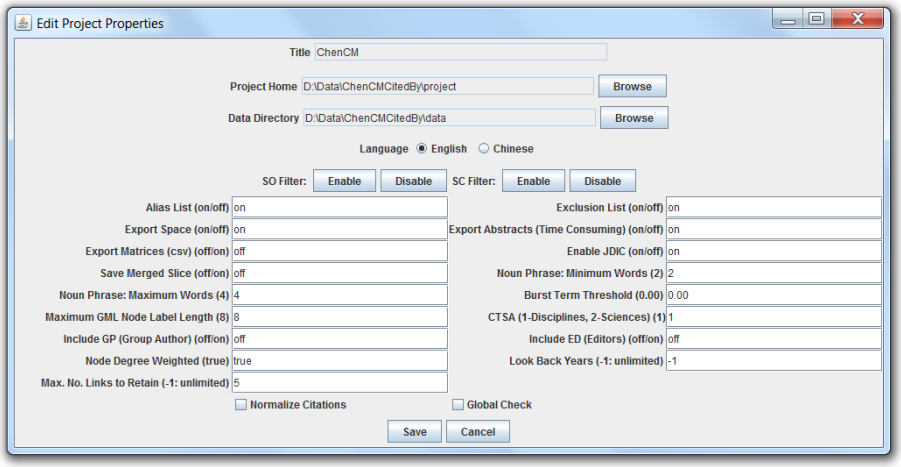


Figure 43. Properties of an existing project.

If you want to retain records from a specific set of journals in your dataset, you can enable the SO Filter function. First, you need to create a list of the names of journals in which those records you want to keep and save the list in an ASCII file as instructed below.

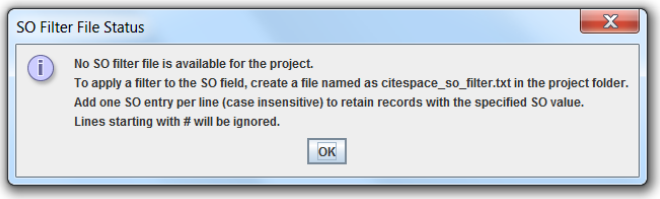


Figure 44. Instructions on creating an SO filter file.

You can similarly filter records based on their SC field, i.e. their subject categories.

Alias List: on/off

This property is used to enable or disable the feature of merging different variants of the same entity into a single node.

Exclusion List: on/off

This property is used to enable or disable CiteSpace to exclude a list of items to appear in the visualizations.

Look Back Years

This property controls the maximum length of a citation in terms of the difference between the publication dates of the citer and the cited reference. Set this property to -1 if you do not want any limit. For example, a value of 5 in this property means that citations made to references more than 5 years ago will be ignored.

This property is a simple link reduction method.

Max. No. Links to Retain

This property controls the maximum number of links to retain for each node in the network. Set this property to -1 if you do not want any limit.

For example, a value of 5 in this property means that up to 5 strongest links connecting to a node will be allowed. If the node has more than 5 connected neighbors, then they will be truncated, i.e. ignored.

This property is a simple link reduction method.

4.2.2.3 Clean a Project

This function will attempt to delete intermediate data files, for example, keyword extraction files, graph files in the graphml format, files of clusters, and files with the word citespace as the prefix of their filenames, which record how you configure your project.

CiteSpace will double check with you on some types of files to make sure you will not delete files that you may need.

4.2.2.4 Remove a Project

This function will remove the current project from CiteSpace, but it will leave the folders and files in these folders intact so that you can restore them by creating a new project and pointing to the existing folders.

4.2.3 Data Sources in Chinese

A Java utility application that can convert data in the CSSCI format to the WoS format is available for download at the following link:

http://cluster.ischool.drexel.edu/~cchen/citespace/utilities/CSSCIREC(new).jar

Store data files downloaded from CSSCI to a folder. To use the converter, open each data file and Save As them as utf-8 files. Then apply the converter to the data folder.

In order to use data files with Chinese encoding, use Preferences ►Chinese Encoding.

For more discussions in Chinese, see the following link:

http://blog.sciencenet.cn/blog-496649-427780.html

5 Configure a CiteSpace Run

A major process in CiteSpace is the network construction process. You can configure the process through a number of parameters. Your configuration will affect the results of the process.

5.1 Time Slicing

Given a dataset of bibliographic records, you need to choose the timespan that you want CiteSpace to analyze so that any records outside the timespan will be ignored. For example, your dataset may contain records from 1800s till 2014, you may choose to focus on the most recent 10 years or on a period in between. You can also include the entire dataset if you want to.

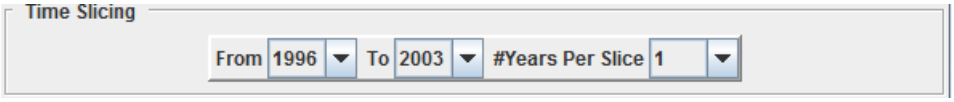


Figure 45. Configuring Time Slicing.

You can time slice the timespan in many ways by setting the value of #Years Per Slice. Typically, you would use 1-year slices and the number of networks will be the same as the number of years within the timespan. Alternatively, you could use k-year slices so that each slide represents data of k years. You can also make a single slice so that you will only deal with one network.

The default selection is to divide the timespan into multiple 1-year slices.

5.2 Text Processing

Each bibliographic record contains four textual fields. These fields provide unstructured text that can be processed and analyzed as part of a visual analytic process.

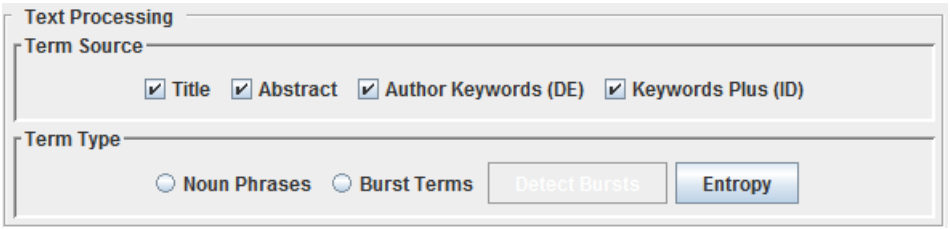


Figure 46. Settings for Text Processing.

You can skip the rest of this section if you are only interested in creating document co-citation networks, i.e. networks of cited references, or node types other than terms.

5.3 Configure the Networks

CiteSpace can generate several types of networks. The default node type is Cited References. In this case, the links are co-citation links. The networks are made of co-cited references.

CiteSpace allows you to choose a single node type or multiple concurrent node types. For example, you may select Author, Cited References, and Category to form networks of three

types of nodes and 6 types of links, i.e. Author-Author (collaborative), Reference-Reference (co-citation), Category-Category (co-occurrence), Author-Reference (author-cites-reference), Author-Category (author-publishes-in-category), and Category-Reference (paper-in-category-cites-Reference).

Document co-citation networks are built on the methods pioneered by Henry Small (Small, 1973), but extended from a single-slide equivalent to multiple-slice network analysis, i.e. a time series of networks in order to detect critical transitions over time more effectively.

Author co-citation networks are originated from (White & Griffith, 1981).

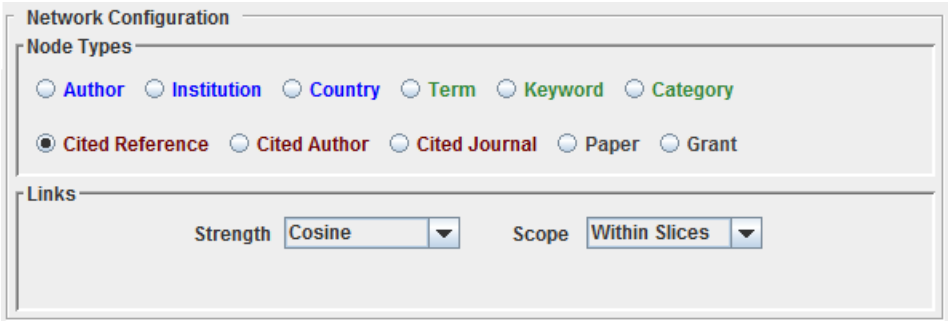


Figure 47. Network Configuration.

If you choose Paper as the node type, the similarity between papers will be calculated by their bibliographic coupling (Kessler, 1963).

Much of the attention in the design of CiteSpace has been devoted to document co-citation analysis due to the preferences that citation patterns of references provide particularly revealing insights into the structure and dynamics of scientific paradigms.

5.4 Node Selection Criteria

CiteSpace provides several ways to sample records to form the final networks. These criteria are known as node selection criteria.

The simplest and recommended one is the first tab Top N per slide. If you enter a value of 50, then CiteSpace will select the 50 most cited or occurred items from each slice to construct a network, depending on the node types you selected in the previous step. If you selected multiple node types, then these nodes will be ranked by the number of times they appeared in the records for each slice.

The second selection method is Top N% per slice. For example, you can select the top 15% most cited items or most frequent items per slice. You can also select the entire dataset by specifying top 100% (as long as you raise the upper limit value high enough, say, 10,000 per slice).

The third method is Threshold Interpolation. It selects both nodes and links. It is complex. I recommend you to explore other selection criteria before this one.

The fourth one needs to be used along with one of the above 3 methods – Select Citers. You can select records based on a distribution of citations. You can specify an interval of the citation distribution, for example, an interval of [5, max] will include records that have 5 or more citations. After the selection, you need to choose which one of the three selection methods you will need, namely, Top N, Top N%, or Threshold Interpolation.

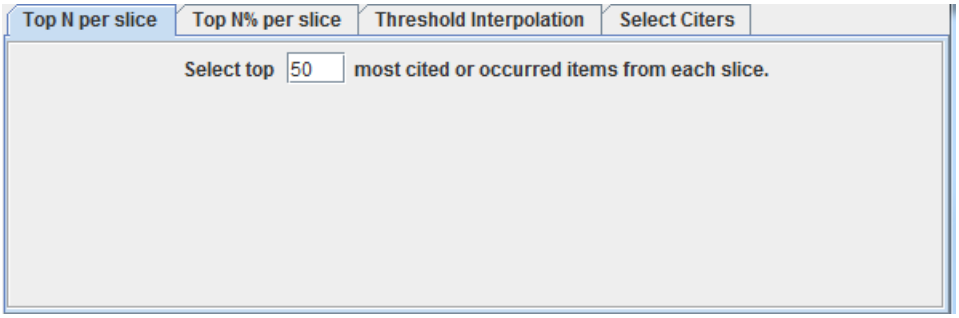


Figure 48. Node Selection Criteria.

5.5 Pruning, or Link Reduction

Bibliographic networks can be very dense with many links. The process to remove excessive links systematically is called network pruning or link reduction.

CiteSpace provides two ways for this purpose: Pathfinder and Minimum Spanning Tree. A comparison of the pros and cons of the two methods is detailed in a 2003 publication (Chaomei Chen & Morris, 2003). In a nutshell, Pathfinder is a theoretically better choice but it comes at a higher price.

I recommend you to start with networks without any pruning because sometimes pruning may reduce the characteristics of the natural groupings.

We are dealing with a time series of networks, i.e. sliced networks, and a merged network. When you select either Pathfinder or Minimum Spanning Tree, you will need to make another decision on whether you want to apply the pruning algorithm to all the individual sliced networks or the merged network only, or both. Since the merged network is resulted from what you do with the sliced networks, pruning sliced networks only will still lead to a merged network with reduced links. If you check both, then you will receive a merged network with the least number of links.

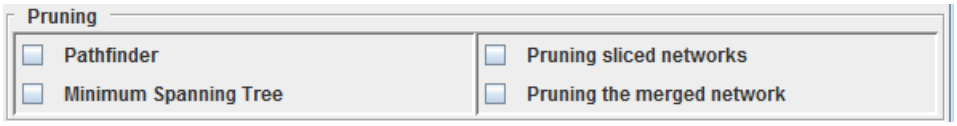


Figure 49. Pruning, or link reduction.

5.6 Visualization

By default CiteSpace will only show you the merged network. If you like, you can turn on the option to see networks of all the time slices. If you have 20 time slices, CiteSpace will open 20 extra windows for time sliced networks – you probably need to think twice before you do that!

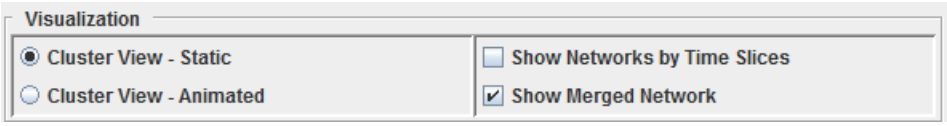


Figure 50. Visualization options.

6 Interactive with CiteSpace

6.1 Adding a Persistent Label to a Node

In addition to labels controlled by the citation or frequency sliders, you can add a label to any node you like. Right-click on the target node and choose Label the Node.

To clear the label, right-click on the node again and choose Clear the Label.

Similarly, you can "bookmark" a node. A "bookmark" will show as a red star at the center of the node, like the one for the Schuster 2001 paper.

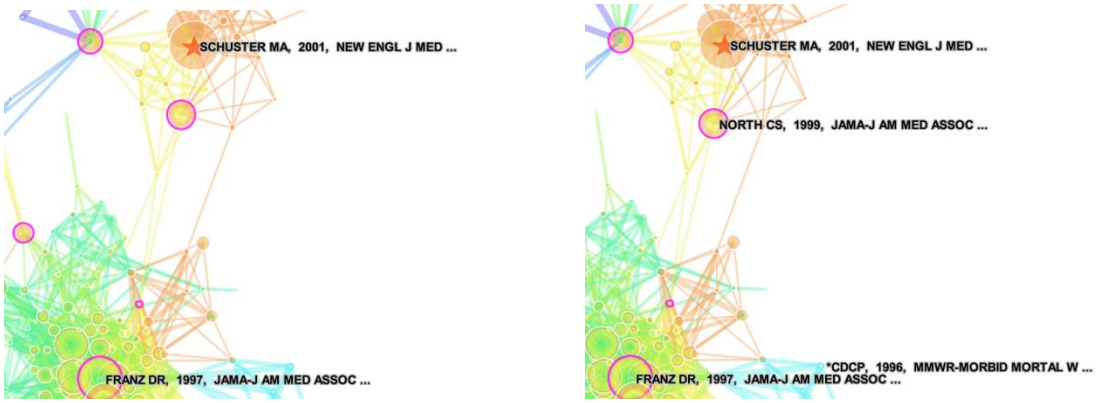


Figure 51. To add a persistent label to a node, right-click on the node and choose Label the Node.

6.2 Using Aliases to Merge Nodes

If you notice that some nodes in the network are in fact the variants of the same entity, you may use aliases to merge them so that they will appear as a single node. For example, in an author co-citation network below, CHEN CM and CHEN C are both from my own publications, so they should be merged into CHEN CM.

To use the alias function, first edit the properties of the current project and in particular make sure the Alias is on by typing an on and save.

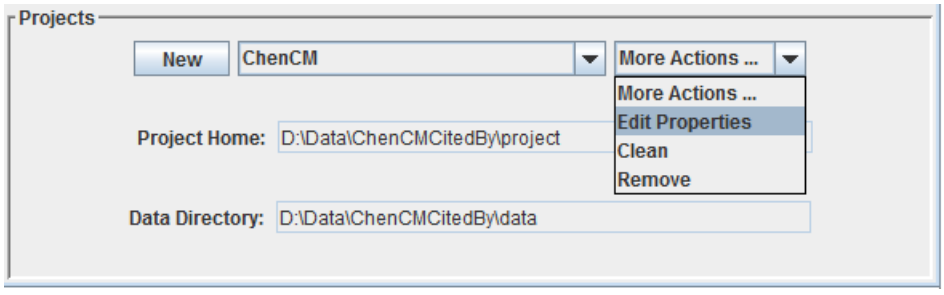


Figure 52. Edit the current project's properties.

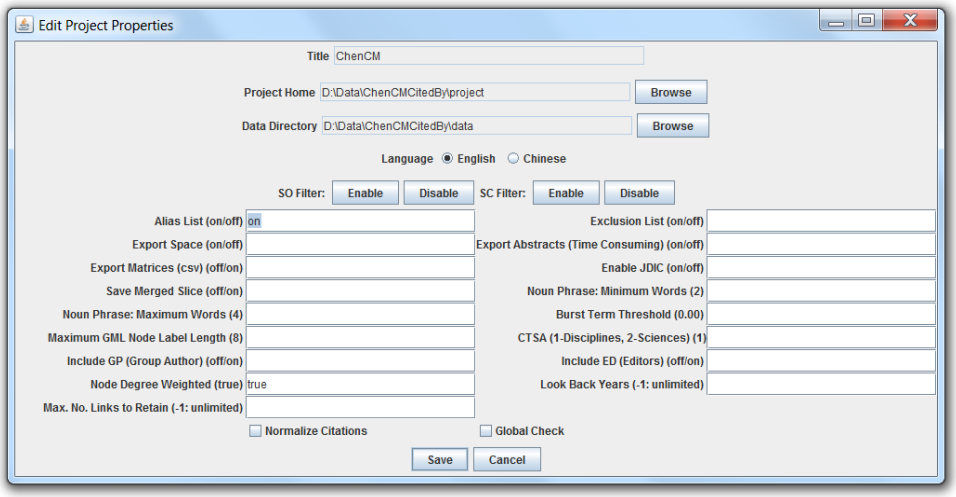


Figure 53. Make sure that the Alias List (on/off) is on. Type "on" in the field and save.

Right-click on the node CHEN CM and select it as the primary alias. Then right-click on the node CHEN C and select the secondary alias. CiteSpace will remind you that you need to re-run the process to see the changes.

Figure 54. Right-click on the node (574) CHEN CM and select "Add to the Alias List (Primary)" and select "Add to the Alias List (Secondary) for the node "(133) CHEN C."

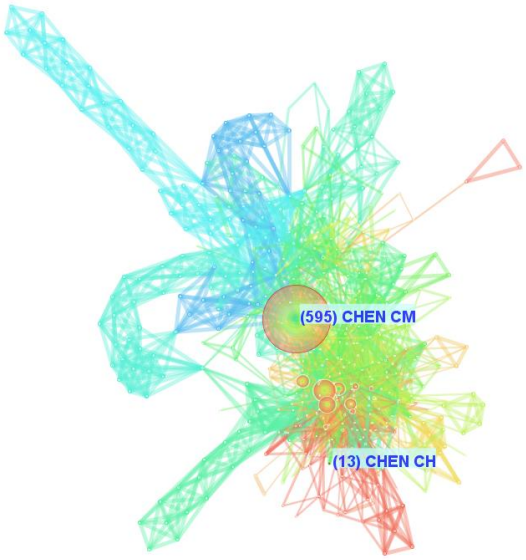


Figure 55. The visualized network after the CHEN CM and CHEN C are merged.

7 Additional Functions

The main menu provides access to additional functions.

7.1 Menu: Data

Data ►Import/Export

CiteSpace provides some utility functions to facilitate data import and export needs.

7.1.1 CiteSpace Built-in Database

CiteSpace provides a user interface to a MySQL database on localhost. The user interface provides various functions to import and export records in connection with the database.

Before you can use this group of functions, you need to set up your MySQL as follows.

On your computer, locate your own User folder and find the .citespace folder. Create a text file mysql.ini with the name-value pairs separated by a tab as the content:

host localhost
user *user_id*
pass *password*

where *user_id* and *password* are your user id and password for your own MySQL login.

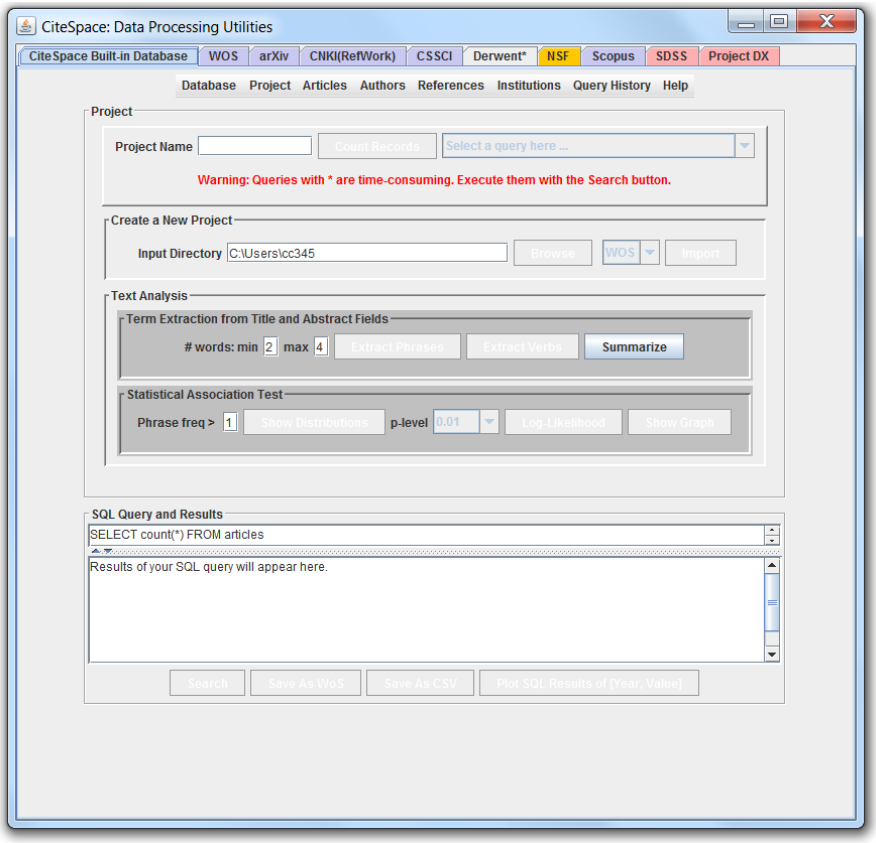


Figure 56. Data Processing Utilities.

After connecting to the database, you will see existing projects, i.e. projects that have been loaded to the database. Note the projects here are stored in the database and they are different from the projects appeared on the main interface of CiteSpace, which are file-based, i.e. the files you downloaded from the Web of Science. You can import the downloaded files to the database and edit them accordingly and export to files in the Web of Science format.

Since the database is a MySQL database on localhost, you can access the database directly with your own MySQL login. You can use this database to process your data before you apply visualization functions on them.

7.1.1.1 Structure of the Database

The name of the database is wos. It contains the following tables:

TABLE articles

id(int), uid, project, author, title, abstract, source, j9, volume, issue, bp, ep, page, dt, doi, year(int), month(int), date(int), citations(int), editor, tagged(boolean)

TABLE authors

id(int), lastname, firstname, initials, project, uid, pos

TABLE refs

id(int), bibcode, ref, doi, author, year, source, volume, page, citer_uid, project

TABLE keywords

id(int), keyword, uid, year, project, type

TABLE phrases

id(int), phrase, isTitlePhrase(booelan), project, uid, year(int), month(int), date(int), freq(int)

TABLE verbs

id(int), project, uid, verb, freq

TABLE bursts

id(int), project, term, weight(double), start(int), end(int)

TABLE institutions

id(int), name, country, uid, year(int), project

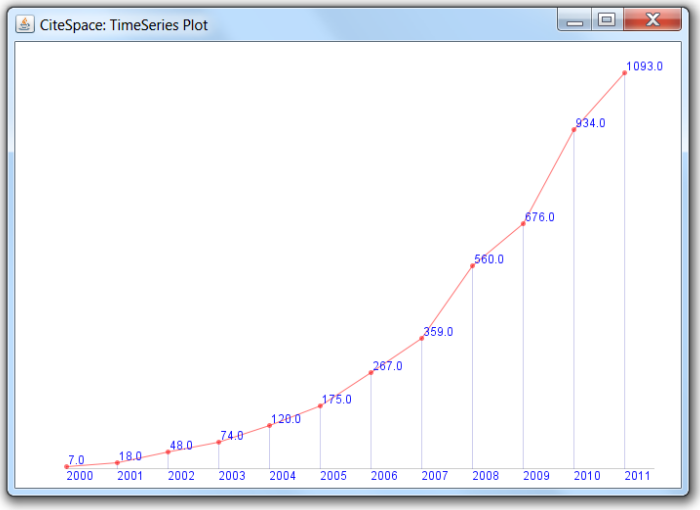


Figure 57. A plot from a project in the built-in database.

Articles ►Most Cited Articles

You can query the database with a few built-in functions on a loaded dataset. For example, you can find the most cited articles in the current project. The SQL query is displayed along with the results. It will help you to get familiar with the internal structure of the database.

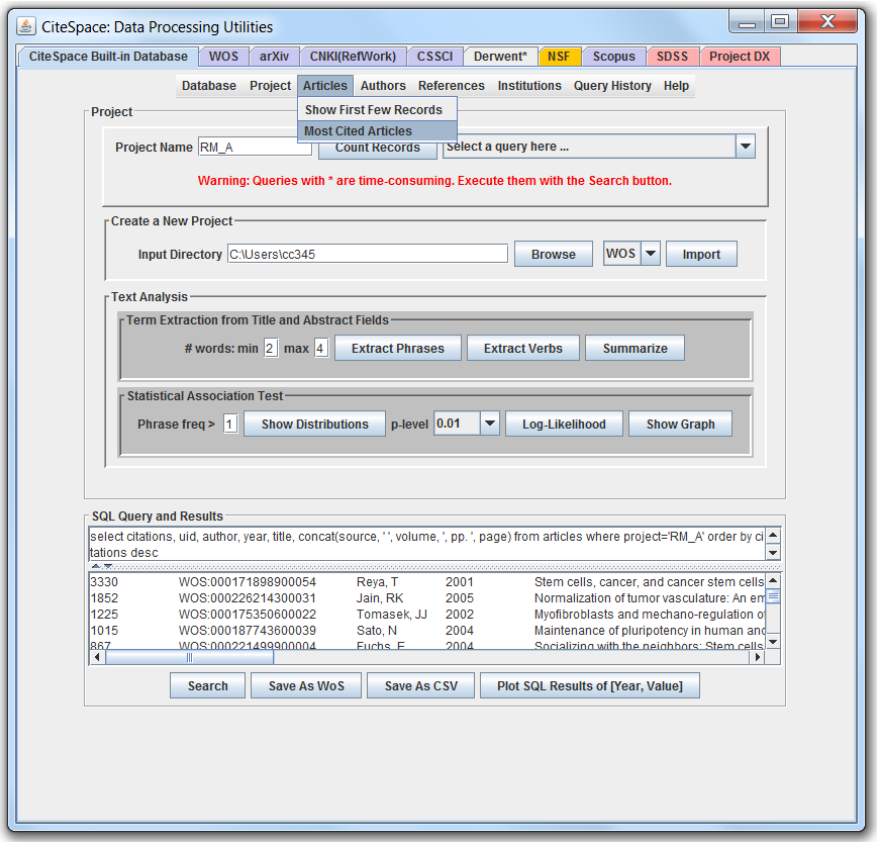


Figure 58. Using a built-in function to find the most cited articles with a SQL query.

7.1.2 Utility Functions for the Web of Science Format

7.1.2.1 Removing Duplicate Records

You can merge multiple datasets you have downloaded by merging the downloaded files to the same data folder. If some files have the same names, you will need to rename them first to resolve the conflicts before you move them together. The simplest way is to add a suffix to the names of the files. For example, if you have two datasets and each contains a file named download_500.txt, you can rename them to download_500_part1.txt and download_500_part2.

You will need to make sure that the merged files do not have duplicated records. CiteSpace has a utility function for this. Specify the input folder and the folder to save a copy of the dataset after duplicates are removed, then press the button "Remove duplicates (WoS)". Note the format of the input files must be in the Plain Text format of the Web of Science.

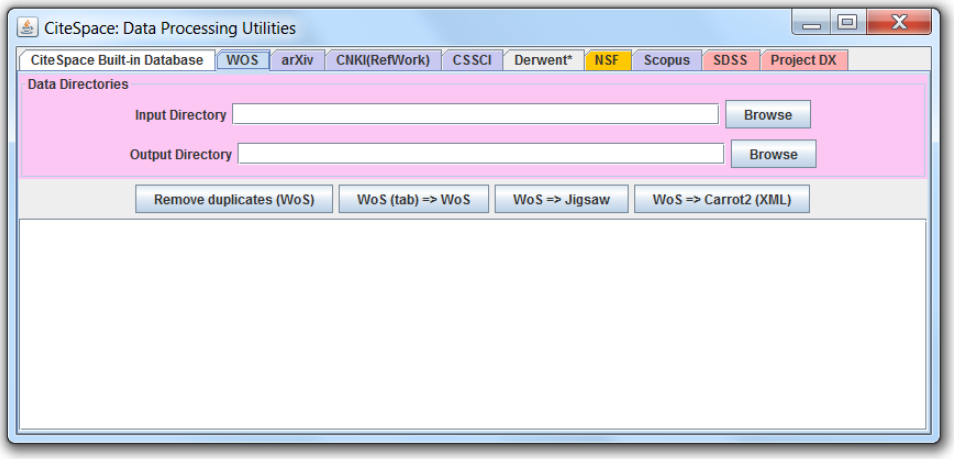


Figure 59. Utility functions for handling bibliographic records in the Web of Science format.

7.1.2.2 Convert the Tab Delimited WoS Format

You can convert the tab delimited WoS format to the Plain Text format (i.e. each field is marked by a two-letter code such as AU, TI, and AB) using another utility function "WoS(tab) ➔ WoS."

7.1.2.3 Convert the WoS Format for Jigsaw

You can convert files in the WoS format to a format that can be processed by Jigsaw – a visual analytic application, which is also freely available (Stasko, Gorg, & Liu, 2008).

7.1.2.4 Convert the Wos Format for Carrot2

You can convert files in the WoS format to a format that can be processed by Carrot2 – an open source text search and visualization tool ("Carrot2: Open source framework for building search clustering engines," 2012). The converted files are XML documents.

7.1.3 PubMed

CiteSpace allows you to retrieve bibliographic records from PubMed. For example, to retrieve records on hypertension based on MeSH headings you can use the query "hypertension [mh]" between 2008 and 201. You can specify the maximum number of records you want to retrieve each year. For illustrative purposes, we limit the maximum number to 25 per year. Retrieved records will be saved to a special folder $your_username\PubMed\SearchResults.

Once the data retrieval is completed, you need to switch to the Web of Science tab and analyze the data in the same way as you did with a dataset from the Web of Science.

Since PubMed records do not include information on cited references, it is not possible to perform citation analysis, i.e. you cannot choose the node types such as cited references, cited authors, or cited journals. Nevertheless, you can perform other analyses such as networks of collaborative authors, terms, keywords, and categories.

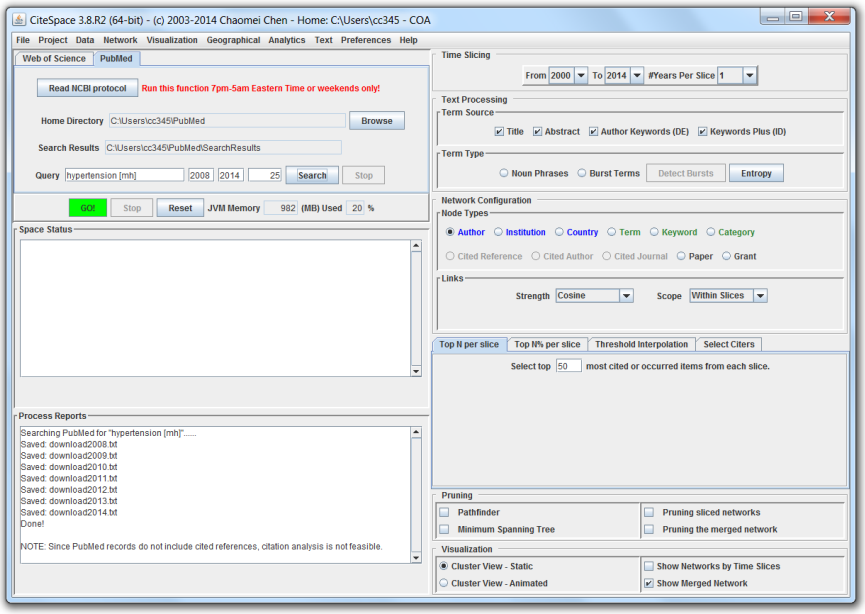


Figure 60. Retrieve bibliographic records from PubMed.

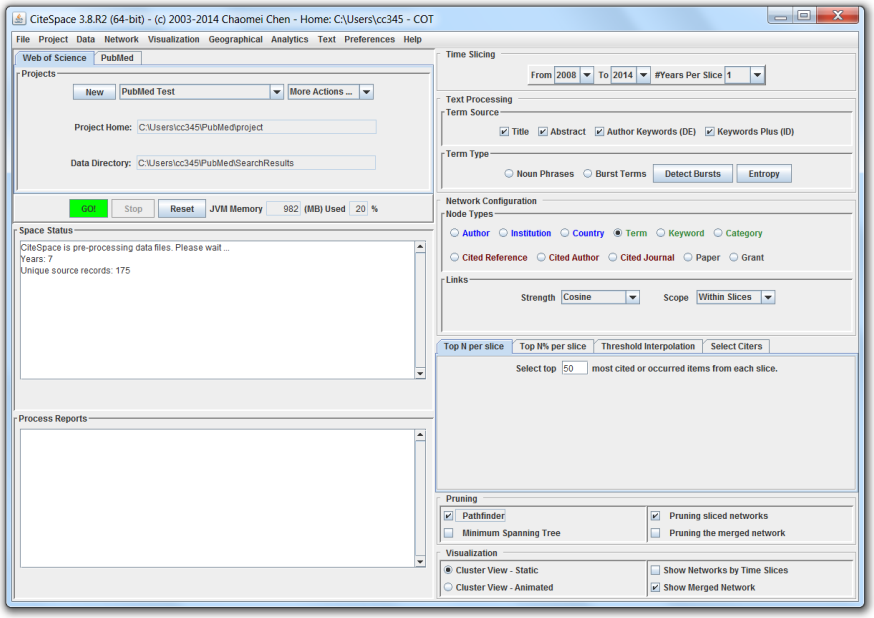


Figure 61. Analyzing the PubMed records …

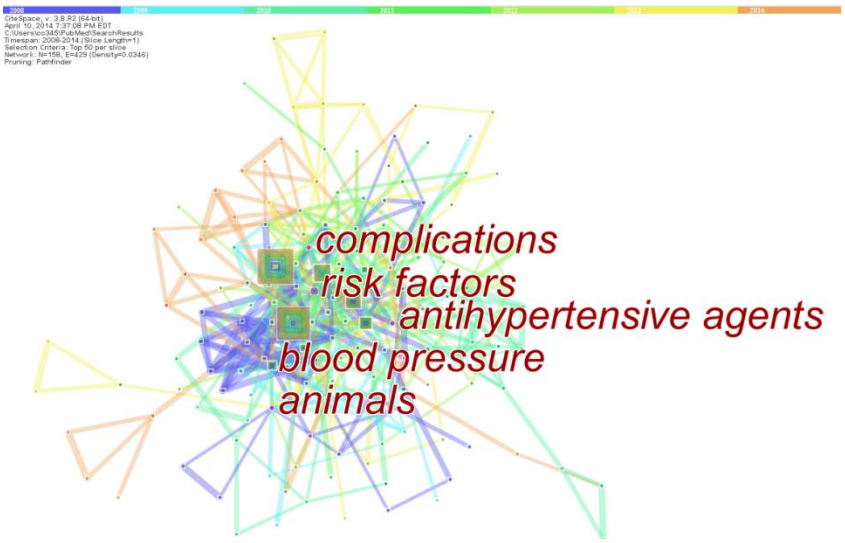


Figure 62. A network of co-occurring noun phrases on hypertension.

7.2 Menu: Network

7.2.1 Batch Export to Pajek .net Files

7.3 Menu: Geographical

7.3.1 Generate Google Earth Maps

Authors' geographic locations in their publication records can be mapped to a geospatial map in KML. You can use Google Earth as the interface to explore the authors' locations and links to their collaborators. You can also go to the original articles directly within Google Earth.

To generate the map file, you need to specify a data folder that contains bibliographic records in the Web of Science format (plain text), which is the same format for CiteSpace projects. This time we just need the data folder. A new folder will be automatically created under the data folder called kml. You will find the generated KML file in the kml folder when the geocoding process is completed.

The Google Earth map generator from CiteSpace needs to know the timespan you are interested, similar to the time slicing setup in the main interface of CiteSpace. Browse to the data folder of your data and click on the "Make Map" button. It may take a while for the process to complete.

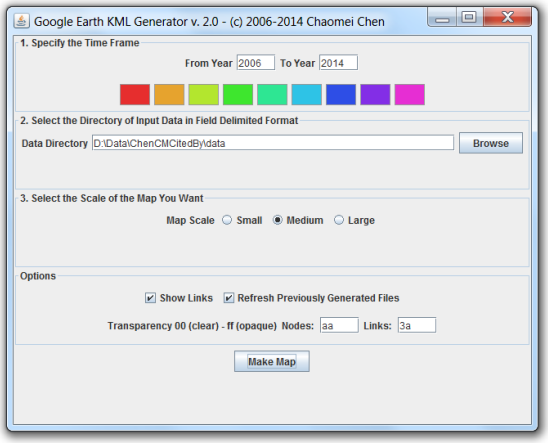


Figure 63. Google Earth KML Generator.

Once the map is generated, you will see a Message notifying you where the map file is, which is in kmz format, i.e. a compressed KML file.

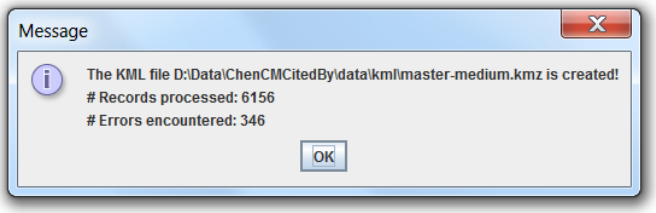


Figure 64. The map is generated.

If you see some errors reported by the generator, you may check the error log file – geocoding_log_tab.txt – and see if you can make corrections in the data and repeat the process afterwards. The map is stored in the master-medium.kmz file if you use the default scale of medium.

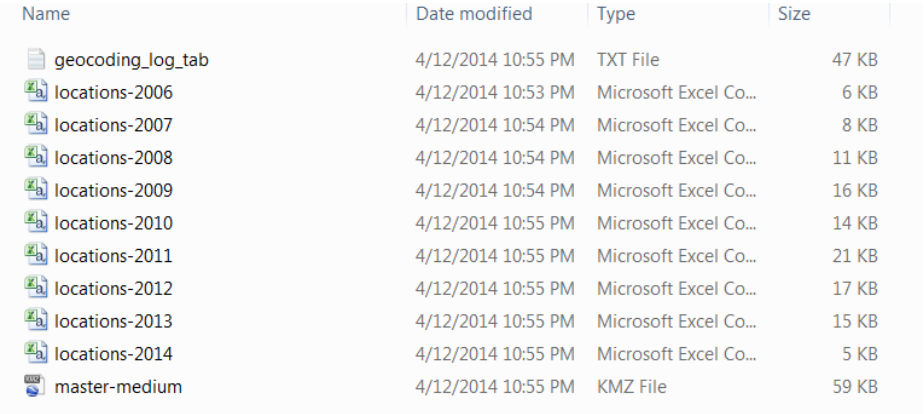


Figure 65. The generated files in the KML folder.

If you have Google Earth installed on your computer, you can double click on the kmz file.

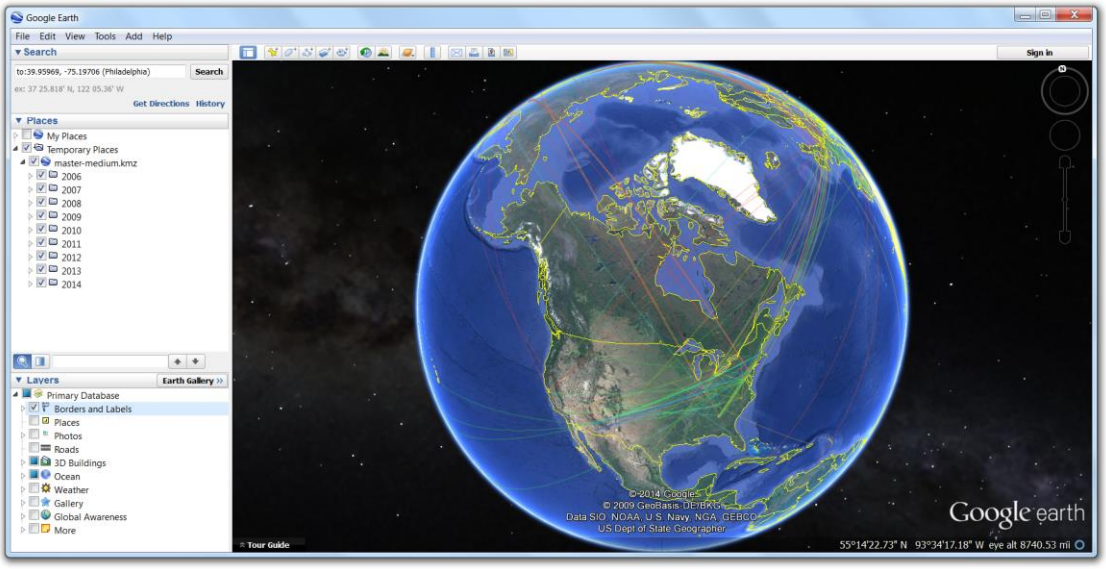


Figure 66. The author collaboration network is shown in Google Earth.

Under the Places, you will see a list of years as layers. You can select or unselect these layers by checking or unchecking the checkbox in front of them so that you can control which years of data you want to see. Coauthored papers in more recent years are linked by lines in red, whereas older collaborations are shown in green or blue lines.

You can drill down from a layer of a year to a location, then to a list of papers published by authors at that location. Each paper on the list is clickable. It will bring you to its full text via its DOI link. You need to have the right subscription to access papers in this way.

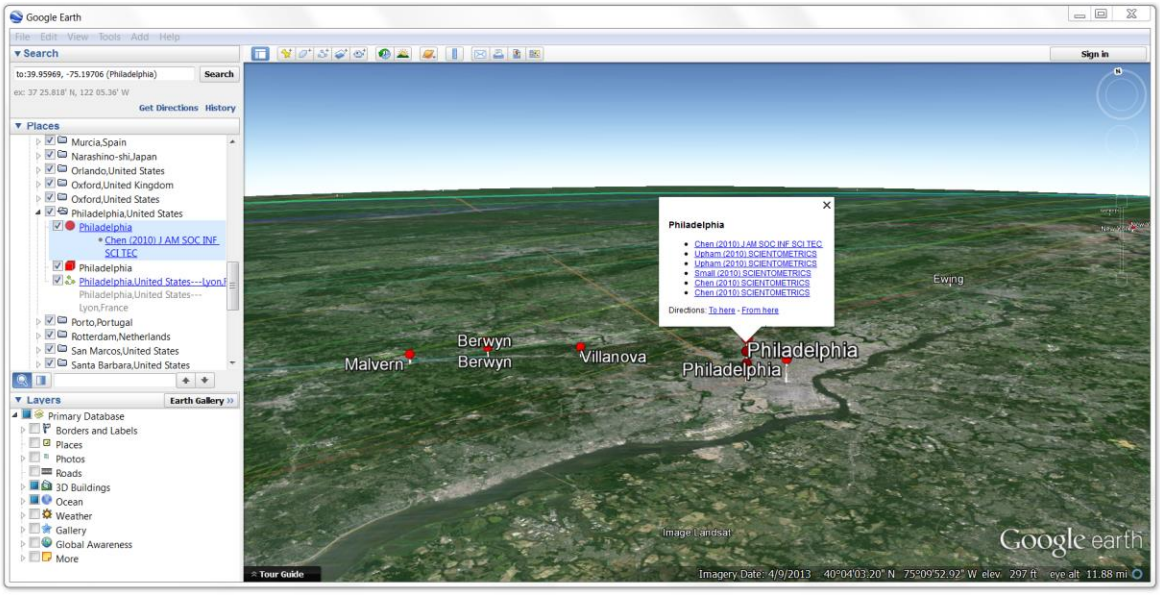


Figure 67. Unfold the list of places on the left and locate a city of interest – Philadelphia.

Click on any of the papers on the list to explore its content. Here let's click on our 2010 JASIST paper.

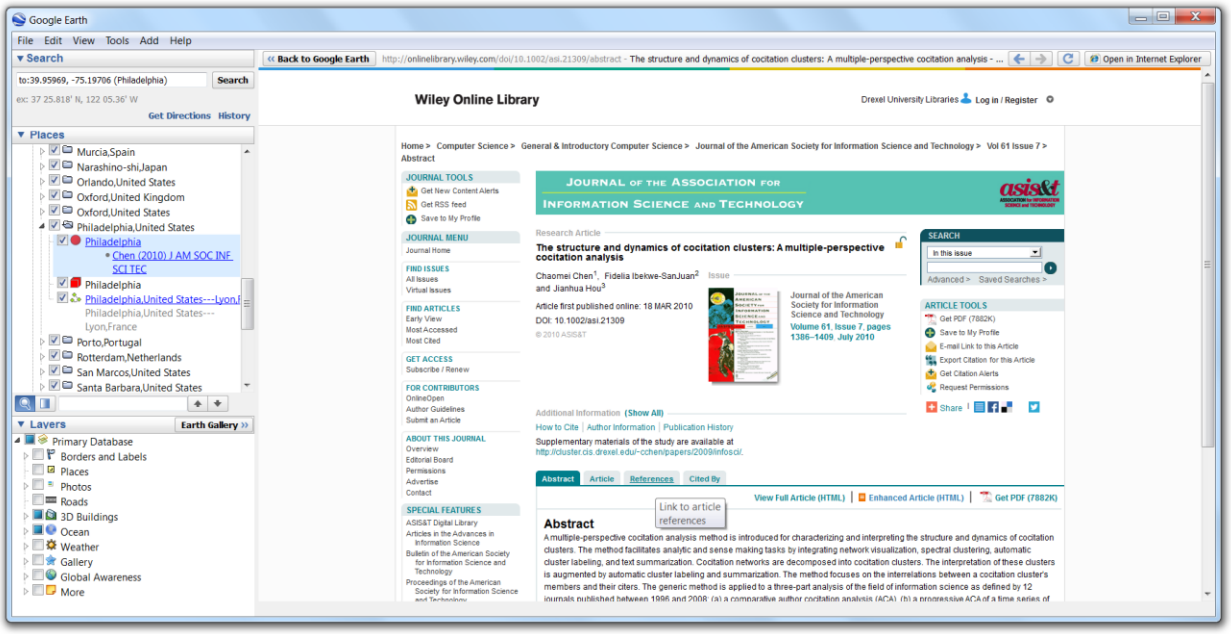


Figure 68. Clicking on the Chen (2010) link takes us to the publisher's page of the paper.

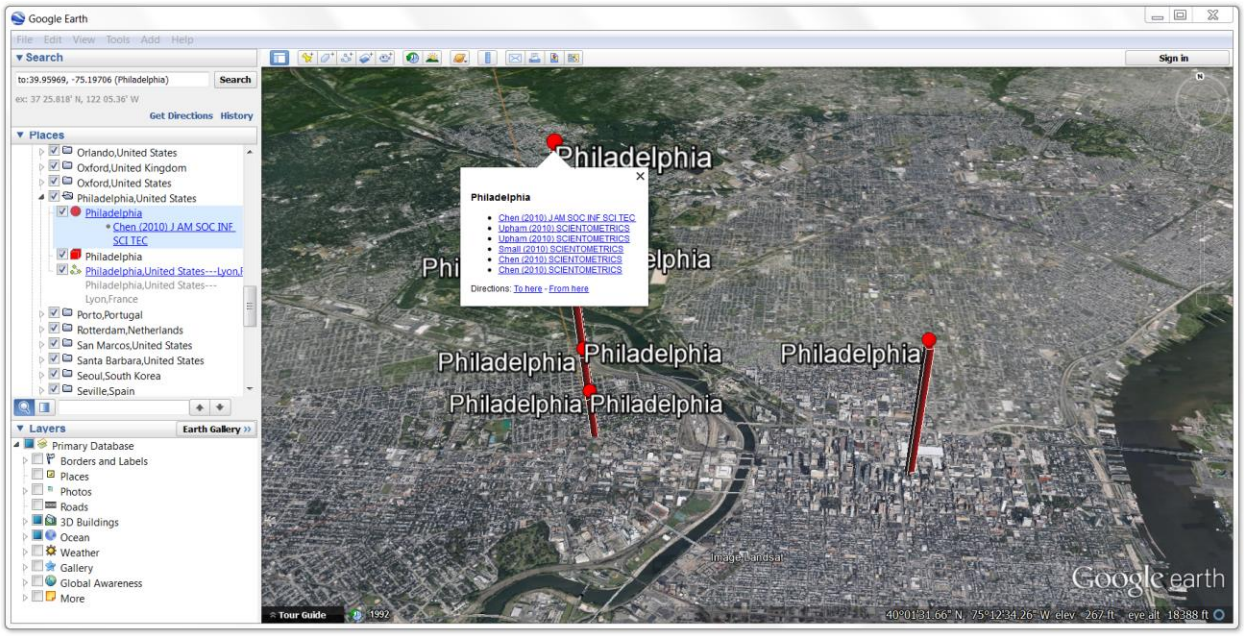


Figure 69. Here is a bird eye view of the downtown Philadelphia. The red bar on the left is on Drexel's main campus.

7.4 Menu: Text

7.4.1 Concept Trees and Predicate Trees

Concept trees and predicate trees in CiteSpace are generated from three types of unstructured text documents: 1) cut and paste text to an input window, 2) from full text files, and 3) from a folder of files in the WoS format, including the data files you downloaded directly from the WoS

and intermediate files saved to the project folder after you performed the clustering algorithm to the current network.

The following example shows how to generate concept trees and predicate trees from the records that cited the largest cluster in the Demo project (i.e. the terrorism research). First, set the Demo project as the current project. Then follow the menu Text ►Build Concept/Predicate Trees.

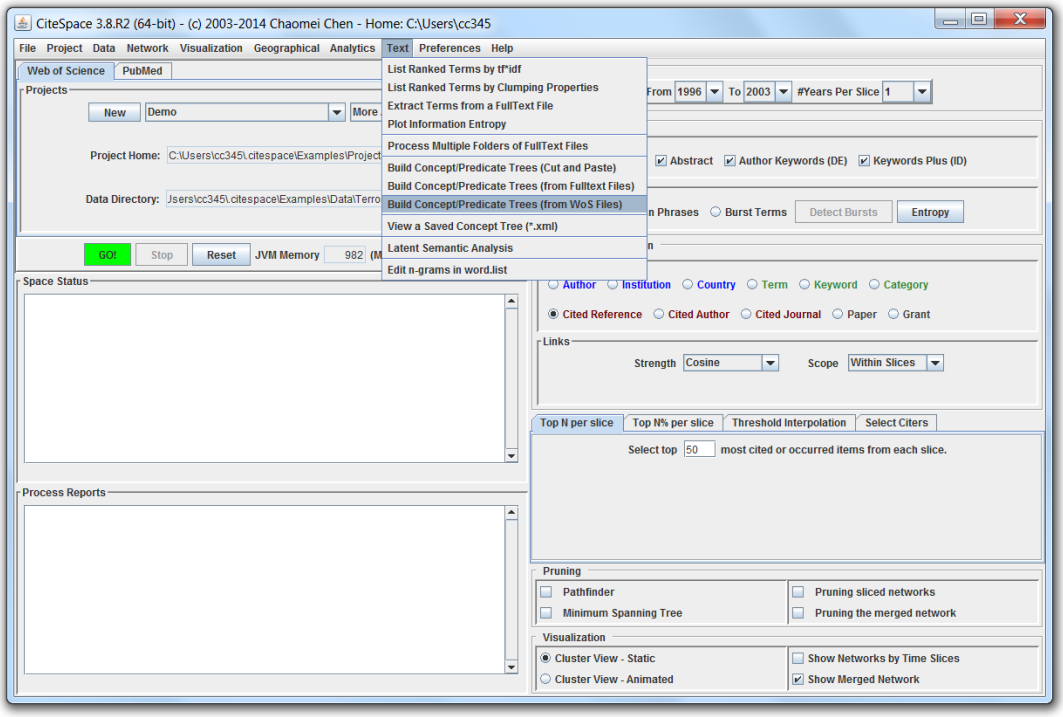


Figure 70.Generate concept trees and predict trees.

You will need to select the file that represents the citing articles to the largest cluster of the Demo project. CiteSpace will show you a list of folders and files. Select the folder **clusters**, then 0.txt, which corresponds to cluster #0, the largest cluster.

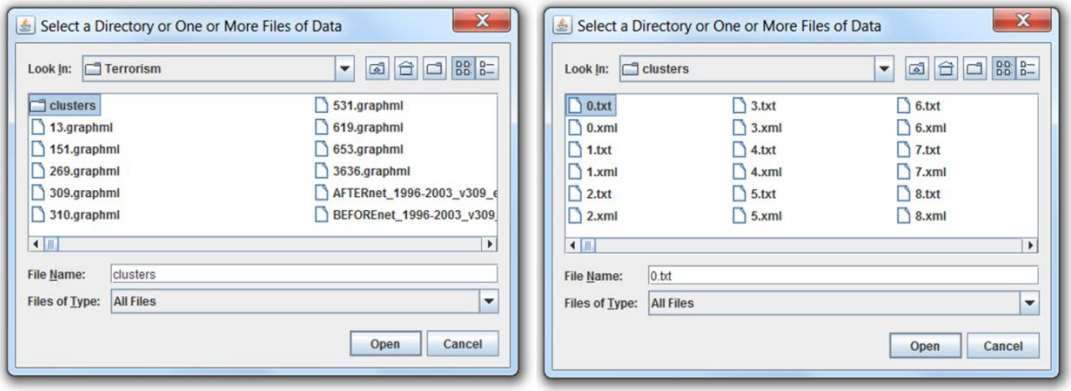


Figure 71. Select the clusters folder of the Demo project and the largest cluster #0.

The concept tree window has three panels. The tree window shows a visualized concept tree. The context window shows the sentences that contain a concept, i.e. the node in the concept tree. The example below shows when you move the mouse cursor over the bioterrorism node in the Tree

window. Different phrases that contain the term bioterrorism are shown as the children nodes of the concept, for example, threat (of) bioterrorism, weapons and agent (of) bioterrorism.

The nodes near the top of the tree are major concepts and major concerns of the cluster. Thus we know that the largest cluster in the Demo project is really about bioterrorism, United States, biological attack, and effective response. These concepts, taken together, give us a fairly focused sense of the nature of the cluster.

To pane the visualized tree, hold down the left button of your mouse and move it around.

To zoom the visualized tree, hold down the right button of your mouse and move it up (zoom out) or down (zoom in).

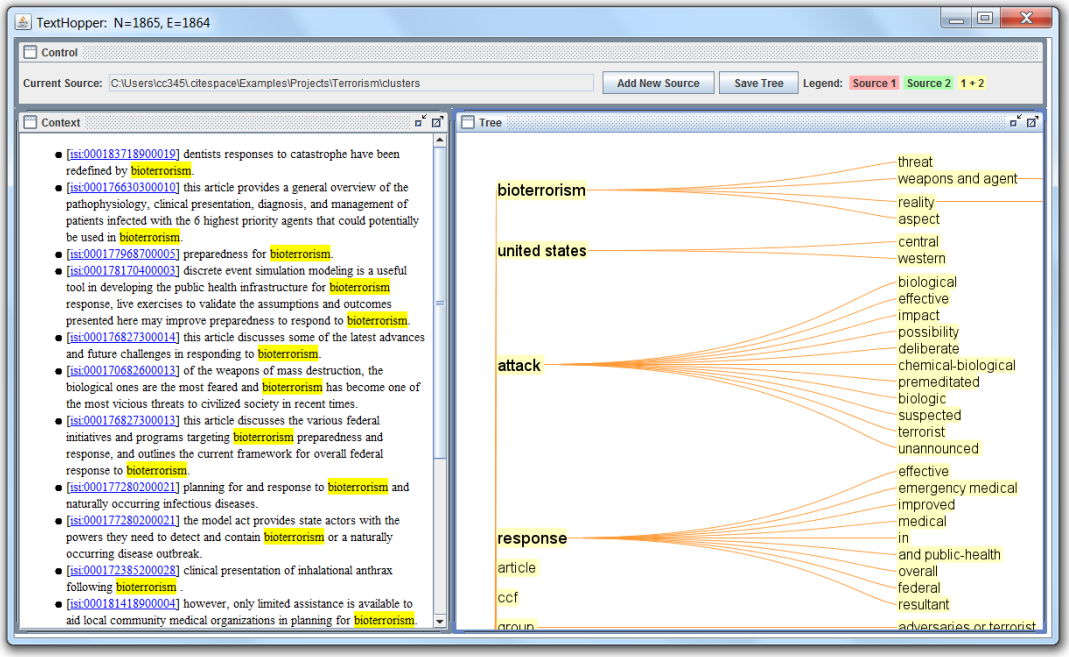


Figure 72. The concept tree of cluster #0 – bioterrorism in the Demo project.

In the Control window, you can add a new source to the existing concept tree. Here let's add the second largest cluster so that we can see what these two largest clusters have in common and where exactly they differ. Recall that the second largest cluster is labeled as PTSD – post traumatic stress disorder.

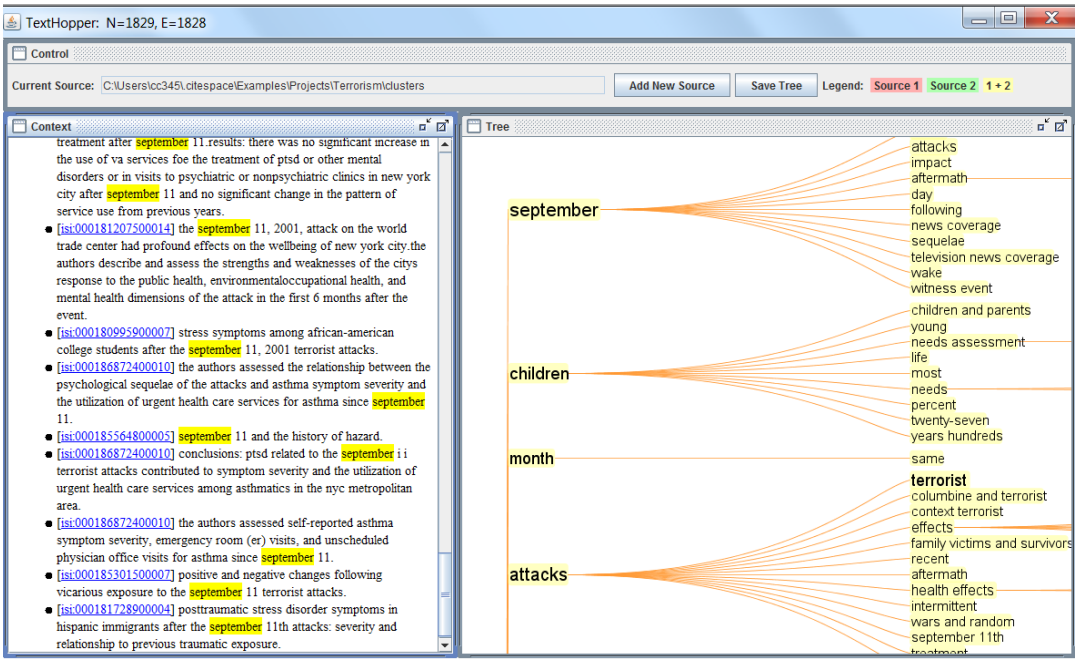


Figure 73. The PTSD cluster. Key concepts: September, children, same month, and terrorist attacks.

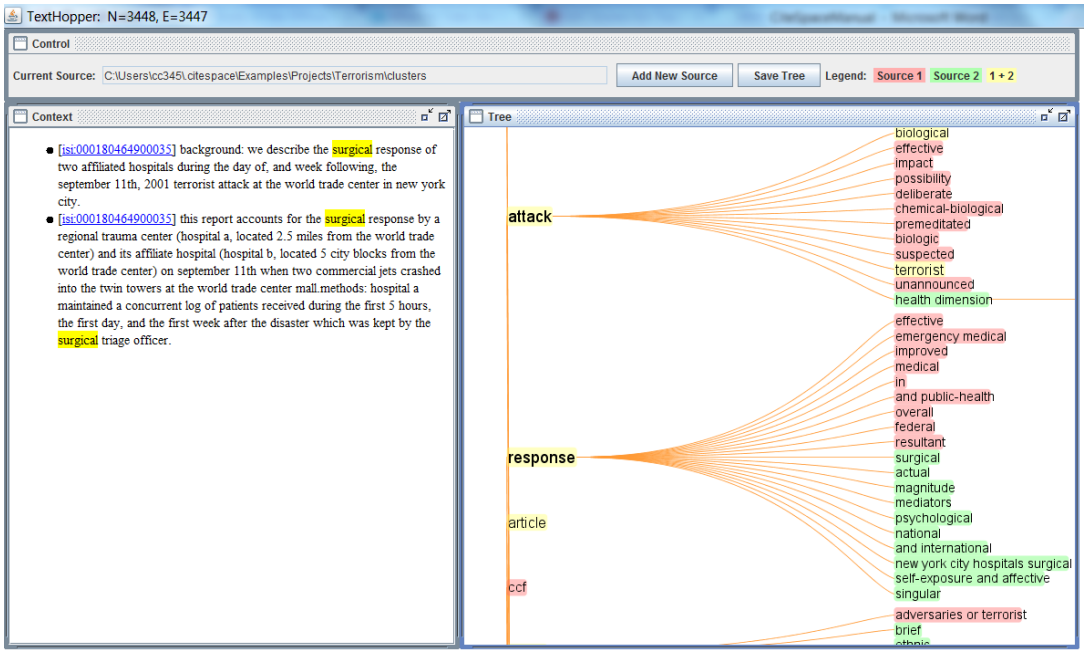


Figure 74. The concept tree of two sources. The bioterrorism cluster is in red. The PTSD cluster is in green. The overlap between the two is in yellow.

7.4.2 List Terms by Clumping Properties

Under the Text menu, you can find several functions dealing with text.

For example, Text ►List Ranked Terms by Clumping Properties, can sort terms by their clumping properties, i.e. how closely they tend to appear in text (Bookstein, Klein, & Raita, 1998). In the Demo project, the most prominent terms include terrorist attacks, world trade center, mass destruction, and biological terrorism.

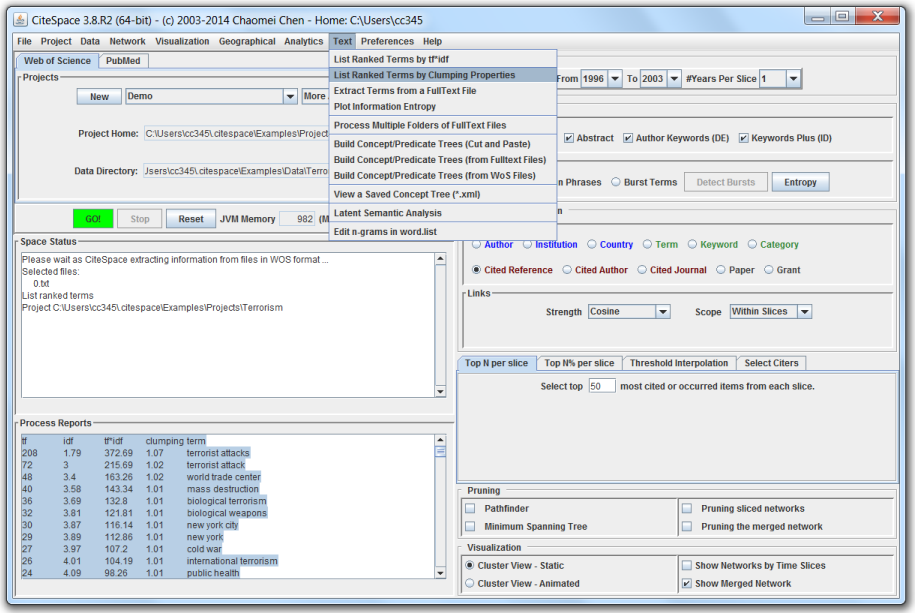


Figure 75. List ranked terms by clumping properties.

7.4.3 Latent Semantic Analysis

CiteSpace provides a somewhat underdeveloped Latent Semantic Analysis function under Text ►Latent Semantic Analysis. The Latent Semantic Analysis is based on a singular value decomposition of the term by document matrix. It is a dimension reduction method (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990).

Use the browse button to locate at least two data sources, i.e. folders of text files in plain full text or the WoS format. After select each data source, add it to the list using the button "Add to the List" then press the "Analyze" button. Then wait for it to finish …

Once it is done, five most representative words in each dimension are shown in the user interface.

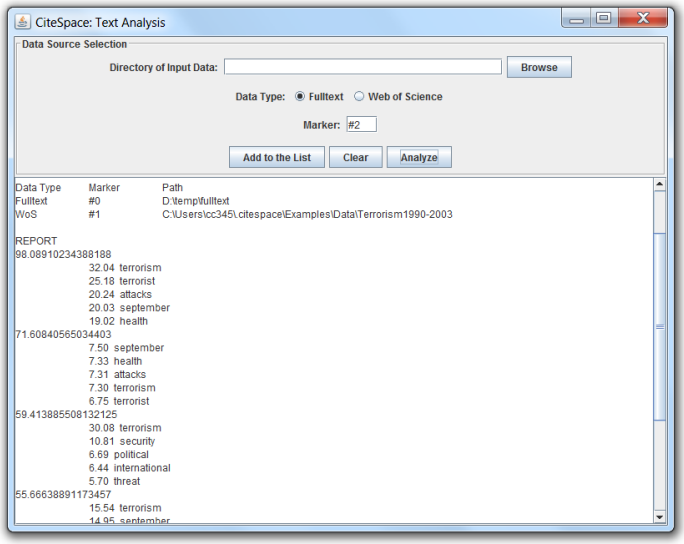


Figure 76. Latent Semantic Indexing.

Three coarse visualizations of the latent semantic space are provided for the three most prominent dimensions of the latent semantic space. Each visualization shows a mixture of terms and documents. You can zoom in and out, change the font size of labels, and the length of a label. That is about it. This function has been there for years, but it has not been actively developed.

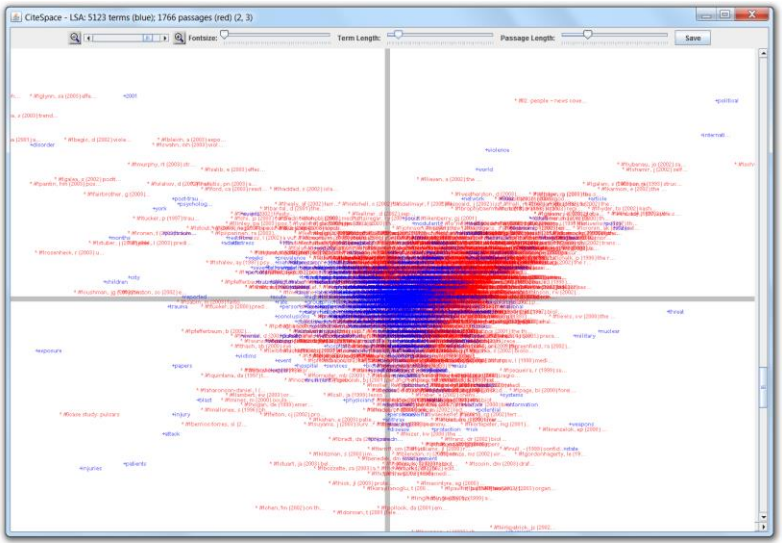


Figure 77. A visualization of the Latent Semantic Space, the 2nd and the 3rd dimensions.

8 References

Bookstein, A., Klein, S. T., & Raita, T. (1998). Clumping properties of content-bearing words. *Journal of the American Society for Information Science, 49*(2), 102-114.

Carrot2: Open source framework for building search clustering engines. (2012). from http://project.carrot2.org/

Chen , C. (2004). Searching for intellectual turning points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci. USA, 101*(Suppl.), 5303-5310.

Chen, C., Hu, Z., Liu, S., & Tseng, H. (2012). Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinions on Biological Therapy, 12*(5), 593-608.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology, 61*(7), 1386-1409. doi: 10.1002/asi.21309

Chen, C., & Morris, S. (2003, October 19-24, 2003). *Visualizing evolving networks: Minimum spanning trees versus Pathfinder networks.* Paper presented at the IEEE Symposium on Information Visualization, Seattle, Washington.

Chen, C. M. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359-377. doi: 10.1002/asi.20317

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*, 10-25.

Kleinberg, J. (2002). *Bursty and hierarchical structure in streams*. Paper presented at the Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada. http://www.cs.cornell.edu/home/kleinber/bhs.pdf

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*, 265-269.

Stasko, J., Gorg, C., & Liu, Z. (2008). Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization, 7*(2), 118-132.

White, H. D., & Griffith, B. C. (1981). AUTHOR COCITATION - A LITERATURE MEASURE OF INTELLECTUAL STRUCTURE. *Journal of the American Society for Information Science, 32*(3), 163-171.