# The CiteSpace Manual

Version 1.01

Chaomei Chen College of Computing and Informatics Drexel University



#### How to cite:

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

# Contents

1	Но	w ca	n I find the latest version of the CiteSpace Manual?	5
2	Wł	nat ca	an I use CiteSpace for?	5
	2.1	Wh	at if I have Questions	7
	2.2	Ноч	w should I cite CiteSpace?	7
	2.3	Wh	ere are the Users of CiteSpace?	8
3	Re	quire	ements to Run CiteSpace	10
	3.1	Jav	a Runtime (JRE)	10
	3.2	Hov	w do I check whether Java is on my computer?	10
	3.3	Do	I have a 32-bit or 64-bit Computer?	12
4	Но	w to	Install and Configure CiteSpace	12
	4.1	Wh	ere Can I download CiteSpace from the Web?	12
	4.2	Wh	at is the maximum number of records that I can handle with CiteSpace?	13
	4.3	Hov	w to configure the memory allocation for CiteSpace?	13
	4.4	Hov	w to uninstall CiteSpace	14
	4.5	On	Mac or Unix-based Systems	15
5	Ge	t Sta	rted with CiteSpace	19
	5.1	Try	it with a demonstrative dataset	19
	5.1	.1	The Demo Project	20
	5.1	.2	Clustering	23
	5.1	.3	Generate Cluster Labels	25
	5.1	.4	Where are the major areas of research based on the input dataset?	27
	5.1	.5	How are these major areas connected?	28
	5.1	.6	Where are the most active areas?	28
	5.1	.7	What is each major area about? Which/where are the key papers for a given are 36	ea?
	5.1	.8	Timeline View	38
	5.2	Try	it with a dataset of your own	39
	5.2	.1	Collecting Data	39
	5.2	.2	Working with a CiteSpace Project	43
	5.2	.3	Data Sources in Chinese	44
	5.2	.4	How to handle search results containing irrelevant topics	45
6	Co	nfigu	re a CiteSpace Run	47
	6.1	Tin	ne Slicing	47

	6.2	Tex	t Processing	. 48
	6.3	Cor	nfigure the Networks	. 48
	6.3	.1	Bibliographic Coupling	. 49
	6.4	No	de Selection Criteria	. 49
	6.4	.1	Do I have the right network?	. 50
	6.5	Pru	ning, or Link Reduction	. 50
	6.6	Vis	ualization	. 51
7	Int	eract	ing with CiteSpace	. 51
	7.1	Но	w to Show or Hide Link Strengths	. 51
	7.2	Ado	ding a Persistent Label to a Node	. 52
	7.3	Usi	ng Aliases to Merge Nodes	. 53
	7.4	Ho	w to Exclude a Node from the Network	. 55
	7.5	Ho	w to Use the Fisheye View Slider	. 55
	7.6	Ho	w to Configure When to Calculate Centrality Scores Automatically	. 56
	7.7	Ho	w to Save the Visualization as a PNG File	. 57
	7.8	Filt	ers: Match Records with Pubmed	. 58
8	Ad	ditio	nal Functions	. 62
	8.1	Me	nu: Data	. 62
	8.1	.1	CiteSpace Built-in Database	. 62
	8.1	.2	Utility Functions for the Web of Science Format	. 65
	8.1	.3	Scopus	. 66
	8.1	.4	PubMed	. 67
	8.2	Me	nu: Network	. 69
	8.2	2.1	Batch Export to Pajek .net Files	. 69
	8.3	Me	nu: Geographical	. 69
	8.3	.1	Generate Google Earth Maps	. 69
	8.4	Me	nu: Overlay Maps	. 72
	8.4	.1	Add an Overlay	. 73
	8.4	.2	Further Reading and Terms of Use	. 75
	8.5	Me	nu: Text	. 75
	8.5	5.1	Concept Trees and Predicate Trees	. 75
	8.5	5.2	List Terms by Clumping Properties	. 78
	8.5	5.3	Latent Semantic Analysis	. 79
9	Se	lected	d Examples	. 80

10	Metrics	and Indicators	82
1	0.1 I	nformation Theoretic	82
	10.1.1	Information Entropy	82
1	0.2 S	tructural	82
	10.2.1	Betweenness Centrality	82
	10.2.2	Modularity	82
	10.2.3	Silhouette	82
1	0.3 T	emporal	82
	10.3.1	Burstness	82
1	0.4 C	Combined	82
	10.4.1	Sigma	82
1	0.5 C	Cluster Labeling	83
	10.5.1	Term Frequency by Inversed Document Frequency	83
	10.5.2	Log-Likelihood Ratio	83
	10.5.3	Mutual Information	83
11	Referen	ces	83

# **1** How can I find the latest version of the CiteSpace Manual?

The latest version of the CiteSpace Manual is always at the following location:

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

You can also access the manual from CiteSpace: Help  $\blacktriangleright$  View the CiteSpace Manual (PDF). It will open up the PDF file in a new browser window.

Help	
View	Cite Space Homepage
View	the CiteSpace Manual (PDF)
Gloss	sary
Video	DS
What	t's New
Abou	t

Figure 1. The latest version of the CiteSpace Manual is accessible from CiteSpace itself.

# 2 What can I use CiteSpace for?

CiteSpace is designed to answer questions about a knowledge domain, which is a broadly defined concept that covers a scientific field, a research area, or a scientific discipline. A knowledge domain is typically represented by a set of bibliographic records of relevant publications. It is your responsibility to prepare the most appropriate and representative dataset that contains adequate information to answer your questions.

CiteSpace is designed to make it easy for you to answer questions about the structure and dynamics of a knowledge domain. Here are some typical questions:

- 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 central idea is that centers of research focus change over time, sometime incrementally and other times drastically. The development of science can be traced by studying their footprints revealed by scholarly publications.

Members of the contemporary scientific community make their contributions. Their contributions form a dynamic and self-organizing system of knowledge. The system contains consensus, disputes, uncertainties, hypotheses, mysteries, unsolved problems, and unanswered questions. It is not enough to study a single school of thought. In fact, a better understanding of a specific topic often relies on an understanding of how it is related to other topics.

The foundation of the CiteSpace is network analysis and visualization. Through network modeling and visualization, you can explore the intellectual landscape of a knowledge domain and discern what questions researchers have been trying to answer and what methods and tools they have developed to reach their goals.

This is not a simple task. Rather it is often conceptually demanding and complex. If you are about to write a novel, the word processor or a text editor can make the task easier, but it cannot help you to create the plot or enrich the character of your hero. Similarly, and probably to a greater extent, CiteSpace can generate X-ray photos of a knowledge domain, but to interpret what these X-ray photos mean, you need to have some knowledge of various elements involved.

The role of CiteSpace is to shift some of the traditionally labor-some burdens to computer algorithms and interactive visualizations so that you can concentrate on what human users are most good at in problem solving and truth finding. However, it is probably easier to generate some mysterious looking visualizations with CiteSpace than to fully understand what these visualizations tell you and who may benefit from such findings.



Figure 2. Hierarchically organized functions of CiteSpace, for example, GUI ▶ Pruning ▶ Pathfinder: true.

# 2.1 What if I have Questions

If you have a question regarding the use of CiteSpace, you should first check the manual whether your question is answered in the manual. You can do a simple search through the PDF file to find out.

If the manual does not get you anywhere, you can ask your questions on the Facebook page of CiteSpace:

https://www.facebook.com/pages/CiteSpace/276625072366558

You can also post questions to my blog on sciencenet:

http://blog.sciencenet.cn/home.php?mod=space&uid=496649

Please refrain from sending me emails because you will have a much better chance to get my response from either the Facebook or the sciencenet blog.

Generally speaking, thoughtful questions get answered quickly. Questions that you may be able to figure out the answer for yourself if you think a little bit more about it would have a lower priority in the answering queue; it is quite possible that some of them never get answered.

# 2.2 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 indepth 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	Reference
(Google Scholar)	
800	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.
394	Chen, C. (2004). "Searching for intellectual turning points: Progressive
	Knowledge Domain Visualization." Proc. Natl. Acad. Sci. USA
	<b>101</b> (Suppl.): 5303-5310.
157	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." <u>Expert Opinions on Biological Therapy</u> **12**(5): 593-608.

# 2.3 Where are the Users of CiteSpace?

In terms of the cities where CiteSpace were used, China, the United States, and Europe are prominent. Brazil, Turkey, and Spain also have many cities on the chart.



Figure 3. Cities with users of CiteSpace between August 2013 and March 2014 are shown on the map. The colors of markers depict the level of user intensity: green (1-10), yellow (10-100), red (100-1000), and the large red water drop shaped marker (1000+).



Figure 4. The use of CiteSpace in China (August 2013 – March 2014).

#### Chen, C. (2014) The CiteSpace Manual. http://cluster.ischool.drexel.edu/~cchen/citeSpace/CiteSpaceManual.pdf



Figure 5. The use of CiteSpace in the United States (August 2013 – March 2014).



Figure 6. The use of CiteSpace in Europe (August 2013 – March 2014).

# **3** Requirements to Run CiteSpace

# 3.1 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.

# 3.2 How do I check whether Java is on my computer?



Figure 7. Select Control Panel.



Figure 8. Click into the Programs category to find the Java control panel.



Figure 9. Locate the Java control panel.

🛓 Java Control Panel	
General Java Security Advanced	
View and manage Java Runtime versions and set applets.	tings for Java applications and
	View
	OK Const Later
	OK Cancel Apply

Figure 10. Java Control Panel. Choose the Java tab and press the View button to see more detail.

Java Runtime Environment Settings							
User Sy	stem						
Platform	Product	Location	Path	Runtime Paramet	Enab		
1.7	1.7.0_51	http://java	C:\Program Files\		1		
Find Add Remove							
OK Cancel							



# *3.3 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  $\triangleright$  System and Security  $\triangleright$  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.

# 4 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.

# 4.1 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

← → C 🗋 cluster.ischool.drexel.edu/~cchen/citespace/download.html

# *CiteSpace*: Visualizing Patterns and Trends in Scientific Literature

Chaomei Chen

Date	Version	WebStart 1 GB RAM Required	Download CiteSpace	Download <sub>Java JRE</sub>	Video
January 29, 2014	3.8.R1 (64-bit)		zip	<u>64-bit / Windows x64</u>	<u>video</u>
January 29, 2014	3.8.R1 (32-bit)		zip	<u>32-bit / Windows x86</u>	
September 27, 2013	3.7.R7 (64-bit)	<u>launch</u>	zip	<u>64-bit / Windows x64</u>	
September 27, 2013	3.7.R7 (32-bit)	<u>launch</u>	zip	<u>32-bit / Windows x86</u>	

Figure 12. The download page of CiteSpace.

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

Q 52

Computer > Data (D	:) ▶ 3.8.R2.64-bit.public.3.1.2014 ▶	•	← Search 3.8.R2.6-	4-bit.public.3.1 P
Organize   Include in library	Share with ▼ New folder			• 🗌 🔞
Somputer Solution (C:) Data (D:) Network	Name Iib CiteSpaceIII StartCiteSpace	Date modified 3/1/2014 3:38 PM 3/1/2014 3:38 PM 1/25/2014 11:29 PM	Type File folder Executable Jar File Windows Comma	Size 8,490 KB 1 KB
3 items				

Figure 13. 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.

# 4.2 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.

# 4.3 *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.

📔 D:\	D\38.R2.64-bit.public.3.1.2014\StartCiteSpace.cmd - Notepad++ [Administrator]									
File I	File Edit Search View Encoding Language Settings Macro Run Plugins Window ? X									
🔚 Sta	StartCiteSpace.cmd									
1	<b>ECHO</b> OFF									
2	ECHO ************************************	*****	*******	*****	****					
3	ECHO *				*					
4	ECHO *	CiteSpace			*					
5	ECHO *				*					
6	ECHO ************************************	*****	*******	*****	****					
7	ECHO									
8	ECHO You may optimize the	performance of C	iteSpace	by adjusting the	e following	JVM paramet	ers.			
9	ECHO In general, the more	RAM, the better.								
10	ECHO -Xms1g: request at 1	east 1GB ram for G	Java Virt	ual Machine						
11	ECHO -Xmx1g: request at m	ost 1GB, depending	g on your	computer						
12	ECHO -Xss5m: request 5MB	ram for Java stacl	c							
13	CHO ON									
14	14 java -Xms1g -Xms2g -Xss5m -jar CiteSpaceIII.jar									
Batch f	ile lengt	·745 lines · 14 Ir	1 Col:1	Sel:010	Dos\Windows	ANSLas LITE-8	INS			
battin	lengu	1.745 miles.14 Li		301.010	Dostaatingows	AND1 03 011-0	1113 H			

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

# 4.4 How to uninstall CiteSpace

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



Figure 15. 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.

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

	About			
	View version inform	nation about Java Control Pa	nel.	atura
	Network Settings			
	Network settings a use the network se these settings.	re used when making Interne ttings in your web browser.	et connections. By default, Ja Only advanced users should	va will modify
	Temporary Interne	t Files	Network Settin	95
	Files you use in Jan later. Only advance	va applications are stored in ad users should delete files o	a special folder for quick exe r modify these settings.	cution
			Settings View	6
	Java in the browse	r is enabled. See the Secu	irity teb	
Jawa Ca	ache Viewer			
Juru Ca	ache Treven		×	
how: Ap	plications •	<b>V.</b> E C .	Installed Size: 0	.0 KB - Cached Size: 6.
0.00	Application	Vendor	Type Date	Size Statu
1	eSpace (c) 2003-2013 .	College of Information	Application	31316 кв 🐠
The Cts	eSname (c) 2003-2013	College of Information	Application Nov 26, 201	1 20475 KI 🗰
1 Ch		and provide an end of the second second second	and a second second second second	
<u>م</u> ده				
Ch وي الم				
Chi Chi				
<u>م</u> لا م				

Figure 16. Select a cached copy of CiteSpace and remove the item.

# 4.5 On Mac or Unix-based Systems

The following example shows you the basic steps to get started with CiteSpace on a Mac. First, go to the CiteSpace homepage in a browser such as Chrome and download the latest 64-bit version.



Figure 17. On a Mac, go to the CiteSpace home page in a browser such as Chrome and download the latest 64-bit version.

Once the download is completed, follow the option "Show in Finder." It will take you to a list of files downloaded to your Mac. The most recent file should be the zip file for CiteSpace.



une	Date mounieu	Size
3.8.R7.64-bit.public.10.26.2014.zip	Today, 8:27 PM	25.4 MB
—		

Figure 19. The downloaded zip file is shown in your Finder.

Double-click on the zip file to unzip the file to a folder in the current folder.

3.8.R7.64-bit.p	ublic.11.5.2014	Nov 5, 2014, 11:35 PM	Folde
F	igure 20. The zip file	is unzipped to a new folder on the list	t <b>.</b>
Ser.		AN IN IN	1
	3.8.	R7.64-bit.public.11.5.2014	M
FAVORITES	Name	▲ Date Modified	Size Kind
🚇 All My Files	📄 CiteSpaceIII.jar	Nov 5, 2014, 11:32 PM	8.7 MB Java JAR
AirDrop	🕨 🚞 lib	Nov 5, 2014, 11:32 PM	Folder
A Applications	StartCiteSpace.cmd	Oct 26, 2014, 6:39 PM 74	45 bytes Unix E

Figure 21. The new folder contains CiteSpaceII.jar and a lib folder.

The simplest way to get started with CiteSpace is to open the CiteSpaceII.jar by clicking on it while holding the "Control" key on Mac. Select Open from the pop-up menu.

Documents
Downloads
Downloads
Movies
Movies
Music
Pictures
DEVICES
Remote Disc

TAGS



Figure 22. Click on the CiteSpaceII.jar while holding the "Control" key and select "Open."

Due to the Java security settings, you will see a dialog box with two options for Open or Cancel. Choose Open to proceed. It will not harm your computer.



Figure 23. Choose "Open" from the dialog box to proceed.

After you choose Open, CiteSpace is getting started on Mac. You will see its opening page as follows. Choose "Agree" to continue.



Figure 24. CiteSpace is now started on Mac.



Figure 25. Screenshots of running the Demo project of CiteSpace on Mac.

It is a good idea to get familiar with the basic functions of CiteSpace by going through the Demo project on terrorism, which is included in the zip file.

If you want to configure various Java Virtual Machine parameters in more detail than what is shown in the above example, you may generate a bash file for your Mac as follows.

The Mac equivalent of the StartCiteSpace.cmd would be a bash file, which should have a file extension of .sh and should be executable. Let's name the file as StartCiteSpace.sh to be consistent.

1. The content of the StartCiteSpace.sh file should have the following two lines:

```
#!/bin/bash
```

```
java -Xmslg -Xmx4g -Xss5m -jar CiteSpaceIII.jar
```

- 2. The following instruction turns the StartCiteSpace.sh file to an executable file: chmod +x StartCiteSpace.sh
- 3. To invoke the executable file, simply type its name or double click on it.

StartCiteSpace

# 5 Get Started with CiteSpace

# 5.1 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 26. 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 27. 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.

# 5.1.1 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.

a CiteSpace 3.8.R2 (64-bit) - (c) 2003-2014 Chaomei Chen - Home: C:\Users\cc345	
File Project Data Network Visualization Geographical Analytics Text Preferences Help Web of Science PubMed Projects	Time Slicing From 1996 V To 2003 V #Years Per Slice 1 V
New         Demo         More Actions            Project Home:         C:/Users/cc345\.citespace/Examples/Projects/Terrorism           Data Directory:         Jsers/cc345\.citespace/Examples/Data/Terrorism 1990-2003	Text Processing Term Source © Title @ Abstract @ Author Keywords (DE) @ Keywords Plus (ID) Term Type O Noun Phrases O Burst Terms Detect Bursts Entropy
SOL Stop Reset JVM Memory 982 (MB) Used 20 %	Network Configuration         Node Types         Author O Institution O Country O Term Keyword O Category
Process Reports	Top N per slice         Top N's per slice         Threshold Interpolation         Select Citers           Select top         50         most cited or occurred items from each slice.

Figure 28. 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.

🛓 CiteSpac	e 3.8.R2 (	64-bit) - (c) 2003-	2014 Chaom	nei Chen - Ho	ome: C:\Users\cc345		
File Project	Data Ne	twork Visualizatio	n Geographi	al Analytics	Text Preferences	Help	
Web of Scie	nce Put	bMed					Time Slicing
- Projects							From 1006 - To 2003 - #Voare Der Slice 1
Trojecto		1-					
	New	Demo		More	Actions		Text Processing
							r Term Source
	Project H	ome: C:\Users\cc34	5\.citespace\E	(amples\Projec	ts\Terrorism		✓ Title ✓ Abstract ✓ Author Keywords (DE) ✓ Keywords Plus (ID)
	Data Direc	tory: Jsers\cc345\.c	tespace\Exam	ples\Data\Terro	orism1990-2003		Cerrm Type
,	GO!	Stop Reset	JVM Memo	ory 982 (N	AB) Used 9 %		Network Configuration
Space Status							
1-vear slices	criteria	snace	nodes	links / all			o Rution o Instatution o country o fermi o heyword o category
Pruning confi	guration:	opuoo	10000	Yc	our Options		uthor O Cited Journal O Paper O Grant
1996-1996	top 50	1033	50	110/			
1997-1997	top 50	1599	50	156 /	? What's your ch	oice?	sine 💌 Scope Within Slices 💌
1998-1998	top 50	1301	50	93/1	_		
1999-1999	top 50	1656	50	150 / 1	Visualize	ave As	GraphML Cancel
2000-2000	top 50	2001	50	170 /			
2002-2002	top 50	8561	50	183/452		11	Threshold Interpolation Select Citers
2003-2003	top 50	9410	50	135/202		-	Select top 50 most cited or occurred items from each slice.
- Process Rep	orts —						
Records in th Records with	e dataset in the chos	1513 sen range: 1143					
Valid referen	ces: 934	88 99.0906% 8 0.9094%					
		0.000470					Pruning
Parsing Time	: 1.47 se	conds					Pathfinder Pruning sliced networks
Total Run tim	e: 2.071 s	econds					
Horaod poter	oric Modeo	-200 Linko-1126					Pruning the merged network
Exclusion Lis	t 0	=309, LINKS=1136				-	Visualization
						-	Cluster View - Static     Show Networks by Time Slices
							Cluster View - Animated

Figure 29. 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 30. 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.

#### 5.1.2 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 31. Most frequently used functions for visual exploration in CiteSpace.

To start the clustering function, simply click on this icon <sup>11</sup>

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.

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

Spotlig	t Citation Burst Link Walkthrough	Q I		Text search: q1   q2	# clusters 37
1996 1997 CiteSpace v. 3.8 R2 (64.bit)	1998	1999	2000	2001	2002
April 3, 2014 9:16:29 PM EDT C:\Users\cc345\.citespace\Examples\Data\Te Timespan: 1996-2003 (Slice Length=1) Selection Criteria: Top 50 per slice	rrorism1990-2003 •				
Network: N=309, E=1107 (Density=0.0233) Pruning: None Modularity Q=0.7141 Mean Silhouette=0.5904		$\diamond$			

Figure 32. 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.



Figure 33. Members of different clusters are shown in different colors.

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.

🛓 Sur	Summary of Clusters - terms from descriptors									
	Save/Show as HTML: cluster_summary.html									
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-				
	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 34. A summary table of clusters.

# 5.1.3 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 35. 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 36. 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 37. Set the cluster labels' font size proportional to their size.



Figure 38. Cluster labels' font sizes are proportional to the size of a cluster. The largest cluster is #0 on biological terrorism.

# 5.1.4 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  $\triangleright$  Node Size  $\triangleright$  [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 39. 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.

#### 5.1.5 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 40. Icons of node rendering controls.



Figure 41. Nodes with purple rings are important in connecting different clusters.

#### 5.1.6 Where are the most active areas?

#### 5.1.6.1 Burst Detection

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 42. Citation bursts are indicators of most active areas.



Figure 43. Right click on the node of interest and choose the Citation History of the node.

CiteSpace: History (be aware of the context, i.e. Appearance or Citation)		
ation/Appearance History Referenced in 38 Records		
RUSTER MA, 2001, NEW ENGL J MED, V345, P1507 burst-8 9432 Save Plot to File Legend Fontsize 14		
		24
	/	
	/	
	/	
	1	
	1	
96 97 98 99 00	01 02	03

Figure 44. This is the citation history of an article that has a citation burst.

Using View  $\triangleright$  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.

Siew Citation Burst History					- C X
Top 24 References with Stro	пде	st Cita	tion	Rm	rete
Top 24 References with Stro	nge.		n i	Du.	
References	Year	Strength	Begin	End	1996 - 2003
COOPER GJ, 1983, J TRAUMA, V23, P955	1983	4.6049	1996	1999	
BRISMAR B, 1982, J TRAUMA, V22, P216	1982	2.718	1996	1998	
HULLER T, 1970, ARCH SURG-CHICAGO, V100, P24	1970	2.522	1996	1999	
HADDEN WA, 1978, BRIT J SURG, V65, P525	1978	2.4506	1996	1999	
KATZ E, 1989, ANN SURG, V209, P484	1989	2.3781	1996	1999	
FRYKBERG ER, 1988, ANN SURG, V208, P569	1988	3.5106	1997	1999	_
RIGNAULT DP, 1989, ANN SURG, V209, P368	1989	2.7765	1997	1998	
FRYKBERG ER, 1989, AM SURGEON, V55, P134	1989	2.3133	1998	1999	
MALLONEE S, 1996, JAMA-J AM MED ASSOC, V276, P382	1996	2.2265	1998	2000	
CARTER A, 1998, FOREIGN AFF, V77, P80	1998	2.8148	1999	2001	
FRANZ DR, 1997, JAMA-J AM MED ASSOC, V278, P399	1997	2.7468	1999	2003	
SIMON JD, 1997, JAMA-J AM MED ASSOC, V278, P428	1997	2.4917	1999	2001	
RICHARDS CF, 1999, ANN EMERG MED, V34, P183	1999	2.2097	2000	2001	
INGLESBY TV, 1999, JAMA-J AM MED ASSOC, V281, P1735	1999	4.0214	2001	2003	
HENDERSON DA, 1999, SCIENCE, V283, P1279	1999	3.0777	2001	2001	
INGLESBY TV, 2000, JAMA-J AM MED ASSOC, V283, P2281	2000	3.0698	2001	2003	
TOROK TJ, 1997, JAMA-J AM MED ASSOC, V278, P389	1997	2.7848	2001	2001	
ZILINSKAS RA, 1997, JAMA-J AM MED ASSOC, V278, P418	1997	2.2957	2001	2001	
SCHUSTER MA, 2001, NEW ENGL J MED, V345, P1507	2001	8.9432	2002	2003	
NORTH CS, 1999, JAMA-J AM MED ASSOC, V282, P755	1999	4.4484	2002	2003	
KHAN AS, 2000, LANCET, V356, P1179	2000	3.0153	2002	2003	
INGLESBY TV, 2001, CLIN INFECT DIS, V32, P436	2001	2.7802	2002	2003	
MELTZER MI, 2001, EMERG INFECT DIS, V7, P959	2001	2.7668	2002	2003	
DENNIS DT, 2001, JAMA-J AM MED ASSOC, V285, P2763	2001	2.552	2002	2003	
Sort by Strengths of Burst		Close			

Figure 45. A summary list of references with citation bursts.

Burst detection and visualization can be applied to other types of nodes. For the node type of author, it will show you those authors who have rapidly increased the number of publications. Similarly, institutions will identify universities that are particularly active in the relevant research areas. For keywords, it will show you fast growing topics.

The general procedure is the same for different types of nodes. Here we illustrate the procedure with an example of detecting the burstness of keywords in publications of Drexel University between 2000 and 2014.

- 1. Select the node type: **Keyword**
- 2. Generate a network as usual: 2000-2014; Slice length: 1; Top N=100; GO o (N=392, E=3033)
- 3. Run the burst detection function: **Citation Burst**
- 4. Visualize the entities, i.e. nodes, that have bursts: View > Citation Burst History

Keywords	Year	Strength	Begin	End	2000 - 2014
molecular-sieves	2000	9.7663	2000	2003	
coronary-artery disease	2000	9.4559	2000	2003	
ischemia	2000	8.1566	2000	2003	
gene	2000	7.7435	2000	2002	
rat-brain	2000	7.1179	2000	2004	
enoxaparin	2000	6.5881	2000	2002	
randomized trial	2000	6.5616	2000	2003	
angioplasty	2000	6.4875	2000	2003	
artery disease	2000	6.4439	2000	2002	
unstable angina	2000	6.2517	2000	2003	
expression	2000	6.2063	2000	2000	
wave myocardial-infarction	2000	6.0161	2000	2003	
aging	2000	5.7598	2000	2002	
sol-gel process	2000	5.7502	2000	2002	
cosmology:observations	2000	6.5436	2001	2008	
randomized controlled trial	2000	6.4199	2001	2003	
hypoxia	2000	5.9587	2001	2004	
cardiomyopathy	2000	5.7931	2001	2004	
nitric oxide	2000	6.9119	2002	2006	
apoptosis	2000	9.6153	2003	2004	
spectroscopic target selection	2000	5.8507	2003	2008	
nanofibers	2000	8.1037	2005	2007	
active galactic nuclei	2000	6.6037	2008	2011	
quasars: general	2000	6.0817	2009	2014	
thrombelastography	2000	6.3714	2010	2012	
supercapacitors	2000	6.2642	2011	2014	
body-mass index	2000	6.2544	2012	2014	
exposure	2000	5.8107	2012	2014	
community	2000	5.752	2012	2014	
united-states	2000	9.8788	2013	2014	

# **Top 30 Keywords with Strongest Citation Bursts**

Figure 46. A visualization of the history of the burstness of keywords in publications of Drexel University (2000-2014). For example, cosmology: observations in Astronomy has the longest period of burst from 2001 till 2008.

#### 5.1.6.2 Burst Detection – Additional Controls

If the number of burst items is too many or too few, you can adjust several parameters for the burst detection algorithm.

I will illustrate the steps for detecting bursts of authors, i.e. authors who have published at a very fast rate, with the terrorism dataset that comes with the CiteSpace package. The procedure is the same for other node types.

New       Demo       More Actions         Project Home:       CUUsers/cC245/.ctespace/Examples/Drojects/Terrorism         Data Directory:       Jsers/cC345/.ctespace/Examples/Data/Terrorism         Data Directory:       Jsers/cC345/.ctespace/Examples/Data/Terrorism         Space Status       Image: Comparison of the status of the	Web of Scien	ce PubMed	6				Time Slicing
2000-2000         top 100         234         100         100/109           2000-2000         top 100         34/34         Top Ner slice         Threshold Interpolation         Select Citers           2003-2003         top 100         826         100         61/61         Top Ner slice         Top Ner slice         Threshold Interpolation         Select Citers           vocess Reports	Projects	New Der Project Home: ata Directory: criteria uration: top 100 top 100 top 100	C:Users\cc345\ Jsers\cc345\.ote space 95 130 124 191	citespacelExampl spacelExampl odes 95 100 100	More Actions      More Actions      More Actions      Projects/Terrorism     Mesi Data/Terrorism     1034 (ME) Used 11 %     Incs / all     119 / 119     104 / 104     146 / 126     123 / 123		From 1996       To 2003       PYears Per Silce 1         Text Processing         Term Source         © Title       Abstract       Author Keywords (DE)       Keywords Plus (D)         Term Type         • Noun Phrases       Burst Terms       Entropy         • Network Configuration       • Noue Phrases       Burst Terms       Entropy         • Note Types       • Author       • Cited Reference       Cited Author       Category         • Cited Reference       • Cited Author       • Cited Journal       Paper       • Grant         Links
Totocess Reports  download2003_0r5unique bt 5 5 0.001  download2003_0r5unique bt 5 5 0.001  Persent Time: 2.027 eaconds  Total Run Time: 1.033 seconds  Wargad network Nodes-717, Links-723  Wargad network Nodes-717, Links-723  Minimum Spanning Tree  Pruning the morged network	2000-2000 2001-2001 2002-2002 2003-2003	999-1999 top 100 101 100 123/123 000-2000 top 100 234 100 109/109 001-2001 top 100 390 100 34/34 002-2002 top 100 396 100 61/61 003-2002 top 100 971 100 64/64					Top N per slice Top IVs per slice Threshold Interpolation Select Citers Select top 100 most cited or occurred items from each slice.
Vetwork modeling ends at Mon Nov 10 15/47/43 EST 2014. Visualization	Records in the Records within Parsing Time: Total Run time Verged netwo Exclusion List Network mode	download20 download20 e dataset 1513 h the chosen ra 2.307 secon rc 1.033 secon ric Nodes=717 0 ling ends at M	03_8x5unique.bt 03_9x5unique.bt inge: 1513 ds ds . Links=723 on Nov 10 15:47.4	5 5	5 0.001 5 0.001	*	Pruning     Pathfinder     Iminumum Spanning Tree     Visualization

First, select Author in the Node Types panel and unselect other node types.

Figure 47. Select Author as the Node Type and unselect other types.

Wait until the Display window shows up. The layout process may continue to run for a while. Since our focus is on the burst detection, you can stop the layout process anytime you like by clicking on the Stop button (which is the one with a yellow square on a red background). Next, click on the Citation Burst button.

٨	CiteSpa	ce: Disp	olay M	erged - (c) 2003-2014	Cha	ei Chen - Project Hom	ne: C:\Users	\cc345\.cit	espace\Examples\Pr	ojects\Terrorism	
File	e Metric	s View	Layo	ut Display Network Ov	erlay	Iters Clusters Export	Help				
	<b>.</b>		0	) - 🛄 🖬 1		🍇 🛧 🔥	tfidf LLR	м	Ҫ Eigen Σ 🔍	<b>°€ °€</b> ™π́	
Vis	ible Fre	Cent.	. Year	Cited References			Spotlig	t Citation I	Rurst Link Walkthroug		
	16	0.00	1999	[Anon]			Spoulgi	chauon			
	16	0.00	1996	Mann P		1996	1997		Hightlight periods of c	itation burst in tree rings	2000
l l	12	0.00	1997	Pfefferbaum B	=	vember 10, 2014 3:47:31 Pl	M EST				
	/ 7	0.00	2002	Galea S		(Users\cc345\.citespace\Ex; nespan: 1996-2003 (Slice Le	amples\Data\Te enoth=1)	rrorism1990-2	:003		
	/ 7	0.00	2002	Vlahov D		lection Criteria: Top 100 per	slice				
	6	0.00	1996	Sandler T		uning: None	ty=0.0028)				
	6	0.00	1999	Waeckerle JF					~		
	∠ 5	0.00	1996	Shapira SC							
	5	0.00	1998	Noah DL					DA N		
	<b>/</b> 5	0.00	2002	Ahern J				-			
	5	0.00	1997	Kahana T							
	∠ 5	0.00	1997	Hiss J			$\triangleleft$				
	∠ 5	0.00	2000	Tonat K				$\boxtimes$			
	<b>/</b> 5	0.00	2002	Tucker P							
	✓ 4	0.00	2002	Kilpatrick D			4			44	
	✓ 4	0.00	2000	Johnson J			~	1 44			
	✓ 4	0.00	2000	Eitzen EM							
	✓ 4	0.00	1996	Phillips EH							
	✓ 4	0.00	2002	Gold J							
	✓ 4	0.00	1996	Enders W							
	✓ 4	0.00	2002	Bucuvalas M			AGA.				
	✓ 4	0.00	1998	Nixon SJ		· · · · · · · · · · · · · · · · · · ·	×				
	✓ 4	0.00	1997	Iwase H						📾 To ' /	

Figure 48. Once the layout process is completed, i.e. when you see the background turns to white, click on the Citation Burst button. You can also force the layout process to stop by click on the Stop button.

To visualize the distribution of items with bursts, select the View menu and then the Citation Burst History menu item. Although it is called the Citation Burst History, it covers the burst patterns of other types of events associated with different node types accordingly. For example, if the node type is Author, then a burst will be on how frequently the author publishes papers. If the node type is a Cited Reference, then a burst will be on how frequently the reference is cited.



Figure 49. Select the Citation Burst History under the View menu to display the durations and strengths of burst items.

A dialog box will ask you to input the type of nodes. In this example, you should input the word "Authors" and this word will be used as part of the heading in the following window.

Input	X
?	What is the node type? References Terms Keywords Cited Authors Citing Authors Institutions Countries Authors OK Cancel

Figure 50. Input the word "Authors" to generate an appropriate title of the display.

Next, you will have a chance to select the number of items to be included in the display.

Input	X
?	37 Authors have citation bursts. Enter how many Authors to be included (137): 20

Figure 51. Select the number of items to be included in the display. The number must be within the range shown in the dialog box.

The following figure shows the resultant display. You may sort the list in two ways: by the strength of a burst or by the beginning year of a burst.

Sview Citation Burst History											
Тор	20 Autho	ors v	with St	rong	gest	Citation Bursts					
	Authors	Year	Strength	Begin	End	1996 - 2003					
	Desivilya HS	1996	1.5466	1996	1996	<b></b>					
	Nagao M	1996	1.8417	1997	1998						
	Iwase H	1996	1.5373	1997	1998						
	Takatori T	1996	1.5373	1997	1998						
	Eitzen E	1996	1.6594	1998	2000						
	Waeckerle JF	1996	1.7193	1999	1999						
	Osterholm MT	1996	1.7072	1999	2000						
	Call JA	1996	1.4888	1999	1999	I					
	Hirshberg A	1996	1.4888	1999	1999	I					
	Kortepeter MG	1996	1.6724	2000	2001						
	Rosen J	1996	2.202	2001	2001	<b>_</b>					
	Vlahov D	1996	2.8452	2002	2003						
	Galea S	1996	2.8452	2002	2003						
	Greenfield RA	1996	2.5579	2002	2003						
	Bronze MS	1996	2.5579	2002	2003						
	Ahern J	1996	2.0264	2002	2003						
	Resnick H	1996	1.6358	2002	2003						
	Bucuvalas M	1996	1.6358	2002	2003						
	Gold J	1996	1.6358	2002	2003						
	Kilpatrick D	1996	1.6358	2002	2003						
Sort by Strengths of Burst Close											

Figure 52. Top 20 Authors with strongest bursts, i.e. who were most active in published papers according to the dataset.

If the initial burst detection only identifies a small number of items, you may adjust the parameters provided in the Burst Detection panel to increase or decrease the total number of burst items. For example, the following combination of parameters increases the number of authors with burst patterns from 37 to 85.



Figure 53. Four additional parameters on the Burst Detection panel can be adjusted to control the burst detection algorithm.

After clicking on the Refresh button, CiteSpace will re-calculate the burstness of all the items. Then you can use the steps described above to display the new results.



Figure 54. The re-calculated burst detection found 85 authors, a substantial increase from the 37 authors found by the default setting.

Siew Citation Burst History							
	Top 85 Auth	nrs w	rith Str	rong	est (	Citation Bursts	-
	Top 05 Auto		in su				
	Authors	Year	Strength	Begin	End	1996 - 2003	
	Desivilya HS	1996	2.0284	1996	1996		=
	Ayalon O	1996	1.7883	1996	1996		
	Gal R	1996	1.7883	1996	1996		
	Philips EH	1990	1.54/1	1996	1990		
	Nagao M	1990	2.2442	1997	1998		
	Elsayed NM	1990	1.7395	1997	1997		
	Takaton 1	1006	1.735	1997	1998		
	Iwase ri Matauda V	1006	1.755	1997	1998		
	Iviaisuda I	1006	1.4935	1997	1009		
	Nakaiima M	1990	1.4935	1997	1998		
	Deutch I	1990	1 4035	1997	1008		
	Fitzen F	1996	1.6616	1998	2000		
	Sharn TW	1996	1 4164	1008	1000		
	Clarencon D	1006	1 4164	1008	1000		
	Brennan R I	1996	1 4164	1008	1000		
	Burckhart MF	1996	1.4164	1998	1999		
	Ostroff SM	1996	1.4164	1998	1999		
	Galonnier M	1996	1.4164	1998	1999		
	Baubichon D	1996	1.4164	1998	1999		
	Lallement G	1996	1.4164	1998	1999		
	Kildew JA	1996	1.4164	1998	1999		
	Waeckerle JF	1996	2.0373	1999	1999		
		1007	0.0107	1000	2000		
		Sort by	Strengths	of Burs	t	Close	

Figure 55. The list shows the new results from a re-configured burst detection process.

#### 5.1.7 What is each major area about? Which/where are 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 56. 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' 7<sup>th</sup> 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.

Clastors	් ක් නි	Citing Articles	° 0' 🛛
	5. C.C. (26, 15, 16, 16, 16, 16, 16, 16, 17, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16	1.0.28 Bak, 52 (2000) chemical-biological humanits and its impact on children is subject review     2.0.18 Faura, 126 (2000) biological interesting, understanding the flavat, preparation, and     3.0.29 Addred, DA (2000) biological interesting, understanding the flavat, preparation, and     3.0.29 Addred, DA (2000) biological interesting and eventuary medicate in the united states     4.0.110 Addred, DA (2000) biological interesting and eventuary medicate in the united states     5.0.13 References, <i>DA</i> (2000) without a state interesting and explosible interesting interpretation events with communitation decimation of the state interesting and the state interesting interpretation with the state of the state interesting interpretation with the state of the state interesting interpretation with the state of the state of the state interesting interpretation with the state of the state of the state of the state interesting interpretation with the state of the state	•
	12 6 1 1. (12.8) americanter: americanterrants, effect	Cond Setemaces	र्वत छ
		$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
Summary S	entences		0'3' 2
	Representative Sentences	completed: 9 #300 Tree lakes excoudsTmeodSave the Lat	

Figure 57. 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.

lusters	of at 🖸 📄 Comp Articles		
	<ul> <li>G. (S. 9) S. (S. 9) S. (S. 9) S. (S. 9) Top Terms (Trift). Tops important investigation of the strangest investigation of the strangest investigation. A statistical strange investigation of the strangest invest investigation of the str</li></ul>	is impact on children a subject ing the threat, preparation, and inny molicine in the waited states on events with contaminand Aut practitioners need to know biological warfare agents he was relighterputed	
	12 6 1 1 12 8) smerican terr, american terroritis, effect     13 6 1 1 (5 5) transistonal, transistonal terrori, transistonal 1.		-
	3         0.00 1 100 000         109 - 1090           7         0.00 1 100 000         109 - 1090           70         0.00 1 100 000         109 - 1090           70         0.00 1 100 000         109 - 1090           71         0.00 1 100 000         109 - 1090           72         0.00 1 100 000         1090           74         0.00 1 100 000         1090           4         0.00 1 100 000         1090           6         0.00 1 100 000         1090           6         0.00 1 100 000         1090           6         0.00 1 100 000         1090           6         0.00 1 100 000         1090           6         0.00 1 100 000         1090	N.         V         P         2         0           N.         V         P         90         0           NO.         V         P         1         0           N.         V         P         1         0           N.         V/27         P/4582         0         1           A.         V/278         P/4552         0         1           A.         V/278         P/4573         0         0           M.         V/278         P/4573         0         0           M.         V/278         P/4563         0         0           M.         V/278         P/454         0         0           M.         V/278         P/454         0         0	
ummary 1	ry Sentences		o"
	Representative Sentences		
	Selection method: 🕷 Centrality 🔗 PageRank 🐼 select from Abstracts 🔄 Start Clusters completed: 🚺 3 of 302 Time taken: 5.04 seconds 🗌 Timeox	Save the List	
	Cubits # - Cubits # = - Cubits # =	Ing for a public heath dhospitals, as well (5) y medical and public s of inhibitors and lowing the use of	

Figure 58. Representative sentences are displayed upon clicking on the Start button in the Summary Sentences window.

## 5.1.8 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.

Spotlight Citation Berst Link Walkthrough	Q ( Int ) Q Test search; g1   g2 # clasters 37	Search Results Cluster Themes Quick Ga Control Panel Burst Detec
	<ul> <li>#0 biological terroris</li> <li>#1 posttraumatic stress</li> <li>#2 ocular injury</li> <li>#3 blast</li> <li>#4 chemical warfare agent</li> <li>#5 strategy</li> <li>#6 social response</li> <li>#7 terrorist assault survivor</li> <li>#8 spread</li> <li># aution</li> <li>#1 ingension</li> <li>#8 indianon gustamatian integes dilution</li> <li>#1 repeation</li> <li>#3 instruct strated abits</li> <li>#3 instruct strated abits</li> <li>#3 instructed interview</li> </ul>	Press Labels     Press     Pr

Figure 59. 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.

### 5.2 Try it with a dataset of your own

#### 5.2.1 Collecting Data

5.2.1.1 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 60. 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 61. 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.

Web of Science™ InCites® Journal Cit	ation Reports © Essential Science Indicators M EndNote ©	Sign In 🔻 Help English 🔻
WEB OF SCIENC	E™ ((	
Back to Search	My Tools 🛩	Search History Marked List
Results: 16 (from Web of Science Core Collection)	Sort by: Times Cited highest to lowest	A Page     1     of     2
You searched for: TOPIC: (CiteSpace)More	Select Page Save to EndNote online V Add to Marked List	
Create Alert	<ol> <li>CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific</li> </ol>	Times Cited: 185
Refine Results	INTERTURE By: Chen, CM JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volume: 57 Issue: 3 Pages: 359-377 Published: FEB 1 2006	(from Web of Science Core Collection)
Search within results for	Image: The second sec	Times Cited: 18
Web of Science Categories   INFORMATION SCIENCE	By: Morar, Mihaela; Agachi, Paul Serban COMPUTERS & CHEMICAL ENGINEERING Volume: 34 Issue: 8 Pages: 1171-1179 Published: AUG 9 2010	(from web of Science Core Collection)
LIBRARY SCIENCE (7) COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS (4)	Agent-based computing from multi-agent systems to agent-based models: a visual survey      By: Niazi Muazi Hussain Amir	Times Cited: 8 (from Web of Science Core
COMPUTER SCIENCE     INFORMATION SYSTEMS (2)     MULTIDISCIPLINARY SCIENCES     (1)	SCIENTOMETRICS Volume: 89 Issue: 2 Pages: 479-499 Published: NOV 2011 View Abstract View Abstract	Collection)
ONCOLOGY (1) more options / values Refine	A bibliometric investigation of research performance in emerging nanobiopharmaceuticals By: Chan, Kaihua: Guan, Jiancheng JOURNAL OF INFORMETRICS: Volume: 5 Issue: 2 Pages: 233-247 Published: APR 2011 IliCett: Full Text from Publisher View Abstract	Times Cited: 8 (from Web of Science Core Collection)

Figure 62. 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.

Searching for intellectual turning points: Progressive knowledge domain visualization	Citation Network
By: Chen, CM (Chen, CM)	ALC: NOTE: N
	104 Times Cited
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	34 Cited References
Volume: 101 Pages: 5303-5310 Supplement: 1	View Related Records
DOI: 10.1073/pnas.0307513100	Feel View Citation Man
Vublished: APK 6 2004	Create Citation Alast
Abstract	(data from Web of Science <sup>TM</sup> Core Collection)
This article introduces a previously undescribed method progressively visualizing the evolution of a knowledge domain's cocitation network. The method first	
derives a sequence of cocitation networks from a series of equal-length time interval slices. These time-registered networks are merged and visualized in a	
panoramic view in such away that intellectually significant articles can be identified based on their visually salient features. The method is applied to a	All Times Cited Counts
cocitation study of the superstring field in theoretical physics. The study focuses on the search of articles that triggered two superstring revolutions. Visually	113 in All Databases
salient nodes in the panoramic view are identified, and the nature of their intellectual contributions is validated by leading scientists in the field. The analysis	104 in Web of Science Core Collection
has demonstrated that a search for intellectual turning points can be narrowed down to visually salient nodes in the visualized network. The method provides a	23 in BIOSIS Citation Index
promising way to simplify otherwise cognitively demanding tasks to a search for landmarks, pivots, and hubs	8 in Chinese Science Citation Database
Keywords	0 in Data Citation Index
KeyWords Plus: AUTHOR COCITATION ANALYSIS; CO-CITATION; NETWORKS; DECOMPOSITION; GROWTH	1 in SciELO Citation Index
Author Information	
Reprint Address: Chen, CM (reprint author)	Most Recent Citation
+ Drexel Univ, Coll Informat Sci & Technol, 3141 Chestnut St, Philadelphia, PA 19104 USA.	Mustafee, Navonil. Exploring the
Addresses:	base through journal co-citation
🛨 [1] Drexel Univ, Coll Informat Sci & Technol, Philadelphia, PA 19104 USA	analysis. SCIENTOMETRICS, MAR 2014.
E-mail Addresses: chaomei chen@cis drexel edu	View All
+ Author Identifiers:	

Figure 63. 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 64. 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  $\triangleright$  Create Citation Report  $\triangleright$  Citing Articles. I will leave it to you to explore in the Web of Science.



Figure 65. 732 Citing Articles will constitute the expanded set to download.

#### 5.2.1.2 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.

Back to Search	My Tools 👻 S	earch History Marked List
Results: 12,046 (from Web of Science Core Collection)	Sort by: Times Cited – highest to lowest 🗸	Page 1 of 1.205
You searched for: TOPIC: (terrorism or terronst)More	Select Page Save to Emflition online Add to Marked List Save to Emflition online Save to Emflition online Save to Enditote desitop extremely aversive evert Save to Researcher0 - i untre I Bry Boarno, GA Save to Researcher0 - i untre I	Analyze Results ton Report feature not available. [?] Times Cited: 775 (from Web of Science Core Collection)
Refine Results	AMERICAN PSYCHOLOGIST Save to RefWorks 20-28 Published JAN 2004	
Search within results for	Cancer risks attributable to low doses of ionizing radiation: Assessing what we really know By Germen, DJ, Dol, R, Goothad, DT, et al PROCEEDINGS OF the (INTOTLER, ACACEMPY OF SCIENCES OF THE UNITED STATES OF AMERICA, Volume: PROCEEDINGS OF the (INTOTLER, ACACEMPY OF SCIENCES OF THE UNITED STATES OF AMERICA, Volume:	Times Cited: 548 (from Web of Science Core Collection)
Web of Science Categories	Full Text View Abstract	
INTERNATIONAL RELATIONS (2.096)     POLITICAL SCIENCE (2.004)     PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH (711)     LAW (540)	3. Miniaturized gas ionization sensors using carbon nanotubes By Mod. A: Koatika, N. Las, E: et al. NATURE Vulner: 424 House (545 Pages: 171-174 Published JUL 10 2003 Full Test View Abstract	Times Cited: 466 (from Web of Science Core Collection)
PSYCHATRY (571) more options / values	4. What good are positive emotions in crises? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001	Times Cited: 373 (from Web of Science Core
Refine	By Fredrickson, BL, Tugade, MM, Waugh, CE, et al. JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY. Volume: 84. Issue: 2. Pages: 365-376. Published: FEB. 2002.	Collection)
Document Types	Full Text View Abstract	
ARTICLE (9,131)     BOOK REVIEW (1.316)     PROCEEDINGS PAPER (720)     REVIEW (600)     EDITORNAL MATERIAL (565)	S. Effects of fear and anger on perceived risks of terrorism: A national field experiment By Lenex JP, Genzalez RM, Gaudi, CA, et al PSYCHOLOGORA, REGUED Volume 14 Issue 2 Pages: 144-150 Published: MAR 2003 Full Text View Abstract	Times Cited: 297 (from Web of Science Core Collection)
more options / values	5. Instrumentation for trace detection of high explosives	Times Cited: 290
Refine	By: Moore, DS REVIEW OF SCIENTIFIC INSTRUMENTS Volume: 75 Issue: 8 Pages: 2499-2512 Published: AUG 2004	(from Web of Science Core Collection)

Figure 66. 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.

Send to File		$\mathbf{x}$
Number of Records:	<ul> <li>All records on page</li> <li>Records 501 to 1000</li> </ul>	
Record Content:	Full Record and Cited References	¥
File Format	Plain Text	<b>T</b>
	Send Cancel	

Figure 67. 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.

#### 5.2.2 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.

#### 5.2.2.1 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.

#### 5.2.2.2 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.

More Actions	•
More Actions	
Edit Properties	
Clean	
Remove	

Figure 68. Edit the properties of an existing project.

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

Edit Project Properties				
	Title ChenCM			
1	Project Home D:\Data\ChenCMCitedBy\proje	ect	Browse	
Ε	Data Directory D:\Data\ChenCMCitedBy\data	1	Browse	
	Language 🖲 En	glish 🔾 Chinese		
	SO Filter: Enable Disable	SC Filter: Enable Dis	able	
Alias List (on/off)	on	Exclusio	n List (on/off)	on
Export Space (on/off)	on	Export Abstracts (Time Consu	ming) (on/off)	on
Export Matrices (csv) (off/on)	off	Enable	JDIC (on/off)	on
Save Merged Slice (off/on)	off	Noun Phrase: Minimu	ım Words (2)	2
Noun Phrase: Maximum Words (4)	4	Burst Term Thr	eshold (0.00)	0.00
Maximum GML Node Label Length (8)	8	CTSA (1-Disciplines, 2	Sciences) (1)	1
Include GP (Group Author) (off/on)	off	Include ED (Ed	itors) (off/on)	off
Node Degree Weighted (true)	true	Look Back Years (	-1: unlimited)	-1
Max. No. Links to Retain (-1: unlimited)	5			
	Normalize Citations	Global Check		
	Save	Cancel		

Figure 69. 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 70. 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.

### 5.2.2.3 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.

### 5.2.2.4 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.

### 5.2.3 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. Before using the format converter, make sure that the input data files with the ANSI encoding are saved to files with the UTF-8 encoding (Use any text editor and then Save As to files with the UTF-8 encoding). Then apply the converter to the data folder.

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

File name: Save as type:	*.txt  Text Documents (*.txt)
Alide Folders	Encoding: ANSI    Save Cancel
File name: Save as type: T	*.txt   Fext Documents (*.txt)
) Hide Folders	Encoding: UTF-8    Save Cancel

Figure 71. Save the downloaded CSSCI files in ANSI encoding to the UTF-8 encoding before using the format converter.

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

#### 5.2.4 How to handle search results containing irrelevant topics

You may realize that no matter how carefully you formulate your search query in the Web of Science or any other sources, it is always possible that your search results contain irrelevant topics.

I recommend you to consider the following strategy. Instead of refining your query endlessly, you take the dataset that may include irrelevant topics and let CiteSpace to differentiate various topics. Until then, you should keep an open mind. You can determine whether a topic is indeed irrelevant only after you have a chance to examine the visualized results.

In most of the cases, it becomes straightforward to spot irrelevant topics because they would end up in an isolated cluster all by themselves. If it appears hard to tell, then the "irrelevant topics" may not be that irrelevant as you thought after all!

Here is an example to illustrate this point. The dataset was collected by a topic search for 'hacker\*', intended to catch topics relevant to hackers, hacker behavior, and associated topics.

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



Figure 72. The largest connected component of a network of co-cited references.

The largest component shown above evidently contains relevant topics such as intrusion detection system, validating trust measure, cyber terrorism, and vulnerabilities. On the other hand, the visualization also reveals an interesting second largest component.



Figure 73. The second largest component of the network on 'hacker'.

The second largest component is not about the 'hacker' topics we wanted. Instead, these items were included in the original search result because the term Hacker is, as it turns out, the name of a prolific author – Hacker, P. M. S., a philosopher who published a number of articles in 1996 and 2003 and a book on Scepticism, Rules, and Language. The two clusters contained in the second largest component are essentially on topics irrelevant to the hackers in the context of computer security.

This example shows that it is a good idea to use a broader search query than a narrow one and defer the differentiation till the visual analytic process later on. The relevance of a topic would be much easier to detect at a later stage of the process.



Figure 74. Clusters of a network of co-occurring keywords. Set Node size = 0 and Transparency = 0.

# 6 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.

### 6.1 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 75. 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.

### 6.2 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.

Text Proces	ssing
	✓ Title ✓ Abstract ✓ Author Keywords (DE) ✓ Keywords Plus (ID)
Term Type-	
	O Noun Phrases O Burst Terms Detect Bursts Entropy

Figure 76. 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.

### 6.3 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).

Network Configuration
○ Author ○ Institution ○ Country ○ Term ○ Keyword ○ Category
Cited Reference      Cited Author      Cited Journal      Paper      Grant
Links
Strength Cosine  Scope Within Slices



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.

### 6.3.1 Bibliographic Coupling

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



Figure 78. An example of a network of citing articles based on their bibliographic coupling scores.

Source: Chen, Dubin, and Kim (2014) Emerging Trends and New Developments in Regenerative Medicine: A Scientometric Update (2000-2014). Expert Opinion On Biological Therapy. (In Press)

### 6.4 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.

Top N per slice	Top N% per slice	Threshold Interpolation	Select Citers	
	Select top 50	most cited or occurred item	ns from each slic	e.

Figure 79. Node Selection Criteria.

### 6.4.1 Do I have the right network?

Obviously the size of a visualized network influences the clarity and complexity of patterns we may learn from the visualization. The structure of a network is determined by the number of nodes selected for each time slice. It is unlikely that we will know in advance whether a Top N of 100 will generate a more desirable network than a Top N of 50.

Here are some suggestions:

First, begin with a Top N of 50 and generate a network visualization. Then check the modularity of the network, the number of clusters, and the average silhouette scores. We won't learn much from the network if there are only a couple of clusters. We won't get a big picture if there are hundreds of clusters either. A good range of the number of clusters would be about  $7\sim10$  major clusters with 10 or more members and each of the clusters has high silhouette values (e.g. > 0.70).

You can then try a Top N of 100 for each slice. If your computer is powerful enough, you can certainly try a Top N of 1,000 per slice or even higher.

You should start the process from a small network (although if you include many slices, even a Top N of 50 can accumulate to a large network), and then based on your initial assessment of the network enlarge the network accordingly.

Finally, note that the largest network is not necessarily the most informative one. Make clear the questions you want to answer first.

### 6.5 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 (C. 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.

Pru	Pruning					
	Pathfinder		Pruning sliced networks			
	Minimum Spanning Tree		Pruning the merged network			

Figure 80. Pruning, or link reduction.

### 6.6 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 81. Visualization options.

# 7 Interacting with CiteSpace

Most controls of visual appearance are under the Display menu.

### 7.1 How to Show or Hide Link Strengths

To turn on/off the display of the strength of a link, use Display ► Link Strength Show/Hide.



Figure 82. Showing the strength of each link in the display.

### 7.2 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 83. To add a persistent label to a node, right-click on the node and choose Label the Node.

### 7.3 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 cocitation 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.

Projects —				
	New Che	More Actions 💌		
				More Actions
	Desised			Edit Properties
	Project Home:	D:\Data\ChenCMCitedBy\project		Clean
				Remove
	Data Directory:	D:\Data\ChenCMCitedBy\data		

Figure 84. Edit the current project's properties.

Edit Project Properties				
	Title ChenCM			
	Project Home D:\Data\ChenCMCitedBy\pr	oject	Browse	
	Data Directory D:\Data\ChenCMCitedBy\da	ita	Browse	
	Language 🖲 F	nglish 🔾 Chinese		
	SO Filter: Enable Disable	SC Filter: Enable Di	sable	
Alias List (on/off)	on	Exclusio	on List (on/off)	
Export Space (on/off)		Export Abstracts (Time Consu	iming) (on/off)	
Export Matrices (csv) (off/on)		Enable	e JDIC (on/off)	
Save Merged Slice (off/on)		Noun Phrase: Minim	um Words (2)	
Noun Phrase: Maximum Words (4)		Burst Term Th	reshold (0.00)	
Maximum GML Node Label Length (8)		CTSA (1-Disciplines, 2	-Sciences) (1)	
Include GP (Group Author) (off/on)		Include ED (Ed	ditors) (off/on)	
Node Degree Weighted (true)	true	Look Back Years	(-1: unlimited)	
Max. No. Links to Retain (-1: unlimited)				
	Normalize Citations	Global Check		
	Save	Cancel		

Figure 85. 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 86. 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 87. The visualized network after the CHEN CM and CHEN C are merged.

In addition to merge nodes interactively through the graphical user interface, you can edit the citespace.alias file directly. The citespace.alias file needs to be located in the project folder for your target project. You can use any text editor to create and edit the file as long as the filename is citespace.alias.

The content of the file is formatted according to the following rules:

- 1. Each one contains a pair of node references separated by the '#' character. The node references can be cited references, cited authors, or institutions.
- 2. The primary form of the alias, i.e. the node you want to retain, should appear first, and followed by the '#' separator. The secondary alias, i.e. the node you want to merge into the primary alias node, should appear after the '#' separator.
- 3. Include as many lines (i.e. pairs) as you need.

Save the file and go back to CiteSpace. Make sure the Enable Alias field in the Project setting is on. Next, you can start the project with GO!

The easiest way to learn about the various detailed formats you can use is to try out with a few examples through the interactive mode. Then open the CiteSpace generated citespace.alias file with a text editor to get yourself familiar with the existing example.



Figure 88. The content of the citespace.alias file is editable.

### 7.4 How to Exclude a Node from the Network

You may exclude a node from the network by right-clicking on the node and select "Add to the Exclusion List". You need to re-run the GO function to re-calculate the network model.



Figure 89. Exclude a node from the network modeling process as well as the visualization process.

The exclusion list is saved in the file citespace.exclusion in the project folder. You may edit it directly as a text file, for example, adding new entries directly or removing existing entries. If you want to remove the exclusion list altogether, simply rename or delete the citespace.exclusion file.

### 7.5 How to Use the Fisheye View Slider

The fisheye view slider is provided for the timeline visualization so that you can see recent years are displayed with a larger screen estate than earlier years. As shown the in screenshots below, the majority of the publications are crowded in the recent few years in the original timeline visualization because some references are dated back as far as 1625, probably by philosophers.

Sliding the fisheye slider from 0 to 6 will help to spread out the crowded display so that we can each cluster's activity in more detail than before.

(f) CiteSpace: Display Merged - (c) 2003-2014 Chae	omei Chen - Project Home: D:\Data\WoS\Hacker\pr	oject for an and the second		
File Metrics View Layout Display Network Overlays	Filters Clusters Export Help			
😟 📾 📀 🖸 🗕 🔚 🛛 🎮	f 🎥 🔬 🔬 🕬 LLR MI 🙆 🥘	egen 🗴 🔩 🎭 🐝		
4	Spotlight Citation Burst Link Walkthrough	Q ( Teit search q1   q2	# clusters 103	Search Results Cluster Themes Quick Guide
CiteSpace v 38 87 (64bt)			100 500 204	Control Panel Burst Detection
November 1, 2014 10, 11:39 AM EDT D1Data1WoStHackeridata_category_Sitered554				The second secon
Timespan: 2000-2014 (Sice Length+1) Selection Criteria: Top 100 per sice Seture: Net1313, E+12274 (Density+0.0143)				By Centrality
Pruning: None Modularity QH0.5265				Threshold 1
	122000000000000000		1	Font Size 100
		010 (00 000 000000000000000000000000000	#0 intrusion detection system	Node Size 10
	Sector Sector	CALL DIGDLOROUSE	81 validating truet measure 82 cyber terrerism	Article Labeling
	and the second sec	And the second sec	83 vicerabilities 84 elucidation 85 intercidation	By Freq 🔄 Show Frequency
			86 information systems security 87 web security	Threshold 10
		and the second second second	#5 realm #9 leakage-resilient rsa-based authenticated key exchange protocol	Font Size 5
		- P. Martine Contraction	#10 hackers account #11 open source movement	Node Size 193
			#13 orwetlian id #14 considering systems risk	Cluster Labeling [ Silhouette>0.05 ]
	100	Contraction of the second seco	#15 realsecure case study #16 pointillat approach	0 0 0 10
			#17 paper wasp #18 authenticated data structure	Layout
	5	restation in	#20 resistance #21 w atermarking technique	Cluster View Timeline O Timezone
			#22 kulm #23 teaching information systems security	
		and the second	#24 statistical process-control     #25 attack detection	
		A STATE OF	820 viterta argentern 827 invasive mean 828 experien data experientes	Iteration 4.210 Stretch
		Contraction Status	#29 op2 web browser #30 mouse dynamic	
	4		#31 new webmaster #32 detection	and any one Com
			#34 probing security vulnerabilities #35 publics in information system security classification	improvement 0.077 0.001
			#36 hyterid honeypot framew ork #37 state e-gov-eroment website	
			#38 neural classifier #39 hiding scheme	Movements 100 💭
			840 entercing application-level web security policies 841 review	
	4	New Cal	#43 dynamic programming algorithm #44 lan security	
		et desi	#45 studying perception #46 peer-assisted carrying authentication	🗹 Snapshot 🗹 Color Map 🗔 Spotlight
		100 million	#47 stbck injection #48 managing vulnerabilities	
			#50 wireless internet #51 arabic baffetent capiche	Waiting Time 1,200
	in a state of the second s	A CONTRACTOR OF A CONTRACTOR OFTA CONTRACTOR O	#52 flow #53 environment	Dariness 50
		A REAL PROPERTY AND A REAL	#54 transformation #55 throating shrew attack	
			857 rext target 858 proventing machine-to-machine hijacking attack	Transparency 40 🗘
The second s		the second second	#59 ddos defense technique #60 sustam	
Contract - 14 <sup>-0</sup> Allow an and Contract - 14 <sup>-0</sup> Allow an and Contract - 10 <sup>-0</sup> Allow and Allow Allo		Line House and and		Relaxer 50 COn
Bases A Service Bases		All and high		
	Contraction of the second seco		Contraction of the Contraction	Fisheye 6
J I.L				E.
			Contraction of the second seco	
			A subscription of the subs	· · · · ·
			and the second s	
	El secolo del		and an and a second sec	
	Element of a second sec		No and team	
	E strategy E stra		Control of the second s	
	Af with any negative sector and a sector a sector and a sector a secto	-		
DEEODE	A construction of the second s	AFTER	14 Constanting of the second s	
BEFORE	A Part Annual A	ATTEN	A second se	
	Contraction of the second seco		Personal and a second s	
	Compared and a second and	-	Control of the second s	
		10 M	The second	

Figure 90. Use the Fisheye slider to adjust the timeline layout so that you can see recent years in more detail.

### 7.6 How to Configure When to Calculate Centrality Scores Automatically

As shown in the following figure, if the size of the network is more than 350 nodes, CiteSpace will turn off the automatic calculation of betweenness centrality scores in order to avoid an unnecessary delay.

Centrality(240): The calcuation of centrlaity is defereed due to the size of .	the
network (653>350). Use CiteSpace->Preferences to reset the parameter.	
Centrality(345): The network exceeds the centrality turn-off point (653>350).	Us
e CiteSpace->Preferences to reset the parameter.	

Figure 91. CiteSpace turned off the automatic calculation of betweenness centrality scores.

If the auto-calculation is turned off, you will see all the values in the Centrality columns are 0s.

Visible	Freq	Cent	Year	Cited References	
~	38	0.00	2001	SCHUSTER MA, 2001,	•
~	35	0.00	1997	FRANZ DR, 1997, JAMA	Г
~	31	0.00	2002	GALEA S, 2002, NEW E	
~	30	0.00	1999	INGLESBY TV, 1999, JA	l
~	30	0.00	1999	HENDERSON DA, 1999	ŀ
~	29	0.00	1997	TOROK TJ, 1997, JAMA	l
~	27	0.00	1999	HENDERSON DA, 1999	l
~	23	0.00	1994	*AM PSYCH ASS, 1994,	
~	23	0.00	1999	NORTH CS, 1999, JAMA	l
~	21	0.00	1994	MESELSON M, 1994, S	l

Figure 92. The values in the third column Centrality are all 0s because centrality scores were not automatically calculated in this example due to the 653-node network is greater than the 350 cut-off point.

To start the centrality calculation manually, go to menu Metrics Compute Centrality.



Figure 93. Manually start the centrality calculation.

Once the calculation is completed, you will see non-zero centrality values in the Centrality column.

Visible	Freq	Cent	Year	Cited References	
~	38	0.00	2001	SCHUSTER MA, 2001,	-
~	35	0.07	1997	FRANZ DR, 1997, JAMA	
~	31	0.05	2002	GALEA S, 2002, NEW E	=
~	30	0.05	1999	INGLESBY TV, 1999, JA	
~	30	0.07	1999	HENDERSON DA, 1999	
~	29	0.05	1997	TOROK TJ, 1997, JAMA	
~	27	0.05	1999	HENDERSON DA, 1999	
~	23	0.11	1994	*AM PSYCH ASS, 1994,	
	23	0.06	1999	NORTH CS, 1999, JAMA	

Figure 94. The Centrality column now has non-zero values.

You can change the default threshold to disable the automatic centrality calculation. Go to menu Preferences  $\triangleright$  Set the Turn-Off Point of Centrality Computation and enter a desirable number to the dialog box. For example, if you enter 1000, CiteSpace will use the new value next time to determine whether it will automatically calculate centrality scores or defer the calculation.

SiteSpace 3.8.R7 (64-bit) - (c) 2003-2014 Chaomei Chen - Home: C:\Users\cc34	5
File Project Data Network Visualization Geographical Overlay Maps Analytics Text	t Preferences Help
Web of Science PubMed	Set the Turn-Off Point of Centrality Computation
Projects	Show/Mute Visualization Window     03
New Demo	Chinese Encoding

Figure 95. Set the threshold value.

Central	ity Calculation
i	Cite Space will turn off the computation of centralities if the size of the network exceeds this parameter (350): 1000 OK Cancel

Figure 96. If you set the value to 1000, CiteSpace will automatically start to calculate centrality scores for networks with fewer than 1000 nodes.

### 7.7 How to Save the Visualization as a PNG File

You can save the visualization to a 300-dpi PNG file to the project folder on your computer. Click on the second icon under the menu bar.

ſ	🛓 Cite	eSpace	e: Disp	lay Mer	ged - (c)	2003-20	014 (	haon	nei Ch	en - I	Proje
	File M	etrics	View	Layout	Display	Network	( Over	lays	Filters	Clus	ters
		<b>E</b>		0			I			٨	Å
	Visible	Freq	Cent	valization Year	as a PNG Cited R	file Reference:	s				

Figure 97. Click on the second icon to save the current visualization to a 300-dpi PNG file.

🛓 Specify an	Image File (*.png) to Save	×
Save In:	Terrorism	
Clusters AFTERnet BEFOREne net_1996-	1996-2003_v309_e1107.png [_1996-2003_v309_e1107.png 2003_v309_e1107.png 2003_v309_e2039.png	
File <u>N</u> ame: Files of Type:	saved_image,png	
	· · · · · · · · · · · · · · · · · · ·	Save Cancel

Figure 98. The default folder is the project folder of the current project. The default filename contains information about the network, i.e. the number of nodes is 309 and the number of edges is 2039.

Name	Date modified	Туре	Size
saved_image	11/10/2014 3:43 PM	PNG Image	930 KB

Figure 99. The new PNG image file is saved to your computer.

### 7.8 Filters: Match Records with Pubmed

If the topic you are analyzing is medical related, there is a good chance that you can find at least some of the records in PubMed. CiteSpace can build the bridge between a visualized network and corresponding records on PubMed by providing you with direct links to the display of these records on PubMed. On PubMed, you can explore further, for example, similar papers and other information.

The basic steps are as follows:

- 1. Using bibliographic records (with cited references) to generate a network visualization in CiteSpace
- 2. Divide the network into clusters and label these clusters as usual
- 3. Pull down the Filters menu and choose Match records with PubMed, and wait for the process to complete (since CiteSpace is set to comply with NCBI's protocol, it will tell you in advance how long you need to wait)
- 4. Right-click on a node or a cluster of your interest, choose List Cluster Members or List Citers to the Cluster

🛓 Cit	teSpace	: Displ	ay Mer	ged - (c) 2003-2014 Chao	omei Chen - Project Home: D:\Data\WoS\CiteSpace_Context\project
File I	Metrics	View	Layout	Display Network Overlays	Filters Clusters Export Help
			0	- II I	<ul> <li>Spotlight</li> <li>Show Citation Burst</li> </ul>
Visible	e Freq	Centr	Year	Cited References	Show Publication Type     on Burst Link Walkthrough
~	197	0.04	2001	NEWMAN MEJ, 2001, P	Filter node labels using a filterlist file  2007 2008
~	185	0.00	1998	WATTS DJ, 1998, NATU	Match Records with PubMed
~	182	0.29	2002	ALBERT R, 2002, REV M	Timespan: 2004-2014 (Slice Length=1)
~	163	0.33	1973	SMALL H, 1973, J AM SO	Selection Criteria: Top Find matching PubMed records and add direct links to them.
~	161	0.27	2002	BARABASI AL, 2002, PH	Pruning: MST
~	150	0.03	2002	SCHLENGER WE, 2002,	Modularity Q=0.7847 Mean Silbouette=0.2115
~	149	0.03	2002	GALEA S, 2002, NEW E	meen on overe out to

Figure 100. Choose the function Match Records with PubMed from the Filters menu.

#0 terr	#2 scientific collaboration networ BARABASI AL (1999)	k
	To comply with the NCBI protocol, this function may take 10.2300005 seconds. Yes - to proceed No - to cancel the request and see more details           V         OK         Cancel	

Figure 101. To comply with the NCBI protocol, CiteSpace estimates the time to be taken to complete the process.

PubMed(338):	Searching for 31 records in PubMed
PubMed(383):	10% completed.
PubMed(383):	20% completed.
PubMed(383):	30% completed.
PubMed(383):	40% completed.
PubMed(383):	50% completed.
PubMed(383):	60% completed.
PubMed(383):	70% completed.
PubMed(383):	80% completed.
PubMed(383):	90% completed.
PubMed(389):	Found 12 matches of the 31 records.

Figure 102. The progress of the process is reported in the command line prompt window. As you can see here, among the 31 records in the example, there are only 12 matches on PubMed. Therefore, bear in mind this will be only useful if the topic is not completely outside the scope of PubMed.



Figure 103. Right-click on a node in Cluster #0 terrorist attack (Schulster 2001) and select List Cluster Members.



Figure 104. PubMed links of matched members of the cluster (#0) will be shown in the Links to PubMed panel under the Search Results tab.



Figure 105. Click on a link in the Links to PubMed panel will bring you to the PubMed page in a new browser window.



Figure 106. List Citing Papers to the Cluster will pop up a window that summarizes the citing papers of the cluster.

etwork Overlays Filte	rs Clusters Export Help	
HIM	🔥 📩 📩 πor LLR ΜΙ 🙆 🦲 των Σ 🔩 🎭 🛰	
eferences 4		Search Results
L, 1999, SGL	Spotlight Citation Burst Link Walkthrough	Control Pane
EJ, 2001, P Otel	2007 1004 1007 1004 1007 1018 1012 1012 1013 1014 1017 1018 1017 1018 1017 1018 1017 1018 1017 1018 1017 1018 1018	Links to PubMed
2002. REV M.	Toter 22, 2019 3-32.428 Mid 23 Millos SUlte Space Context data m2243	GALEA (2003) Tren
973, J AM SO Sele	ten Gineria Teo 10 per silice x teo 10 per sil	NORTH (1999) Psys
L. 2002, PH Prun	92, M97 61, 000 9847	SILVER (2002) Nati
02. NEV 🍰 Summa	ry of Cluster 0:	SCHUSTER (2001)
1965.5	9 1001 <del>N</del>	GALEA (2002) Psyc
MA, 200	4 BENEFIT	SCHLENGER (2002
2006, J A	4 ROLE	
N S. 195	3 AR4B	
2005. H	3 ASSOCIATION	
/, 2005.	3 EARELINGE SIGNAL SEELING	
1989, 101	3 ISRAELI-HEZBOLLAH WAR	
C, 1979	3 JEW	
2002.3	3 MALEETALUSS 3 NATONA STUDY	
2003. A	3 POSTTRAUMATIC GROWTH	
SEV SN	3 SEPTEMBER	
1999.3	3 TERROR ATTACK	
SEV SN Citing Titles		
1, 2010,	(5) Breslau Naomi 2010 the 9/11 terrorist attack and posttraumatic stress disorder revisited http://dx.doi.org/10.1097/hIMD.0b013e3181ea1e21	
03 AM	(a) Figueroa Roango A. 2010 psychological support for assaster vicinis: an evidence-based care model. http://dx.doi.org/ (c) Milanewich Tod. 2010 disaster in contact the affects of 2011 on unit distant from the affects. http://dx.doi.org/ 0.00000000000000000000000000000000000	
C 2012	(5) Nena Yuval 2010 long-term course of probable ptsd after the 8/11 attacks: a study in urban primary care http://dx.doi.org/10.1002/jts.20544	
3.2011	(5) Rimmerman Arie 2010 victims of hostile acts in israel and the united states: comparable policy review http://dx.doi.org/10.1080/13875981003714545	
REF L Z	(4) Vazquez Carmelo 2010 perceived benefits after terrorist attacks: the role of positive and negative emotions http://dx.doi.org/10.1080/17439761003630060	
	(3) Brucher Tim A 2010 male tentor and the us following the terrorist attacks of sectement 1, 2001 http://d.tot.org/10.1186/171-2458-10-273	
	(3) Hall Brian J. 2010 exploring the association between postfraumatic growth and ptsd a national study of jews and arabs following the 2006 israeli-hezbollah war http://dx.doi.org/10.10	
Ribliographic	Datala	
Bronograpine	<ul> <li>Neria Yuval, 2010, Long-Term Course of Probable PTSD After the 9/11 Attacks: A Study in Urban Primary Care, JOURNAL OF TRAUMATIC STRESS, V23, 9</li> </ul>	
	<ul> <li>[5] Figueroa Rodrigo A. 2010, Psychological support for disaster victims: An evidence-based care model, REVISTA MEDICA DE CHILE, V138, 9</li> </ul>	
	[5] Mijanovich Tod. 2010. Disaster in Context The Effects of 9/11 on Youth Distant from the Alacks. COMMUNITY MENTAL HEALTH JOURNAL, V45, 11 2010 June 2010, 101 Sector 10 Context The Effects of 9/11 on Youth Distant from the Alacks. COMMUNITY MENTAL HEALTH JOURNAL, V45, 11 2010 June 2010, 101 Sector 2010, 2010, 2010 June 2010, 2010	
	[3] Badministeriam, 2010, Atter a terror attack, is atterned ab protection as terrorises, and experiences, JOOKINAL OF SOCIAL AND PERSONAL RELATIONSHIPS, V27, 20 [5] Breslaw Macmini 2010, The 9/11 Terrorise Match and Pestfaumatalic Stress Disorder Revisited JOURNAL OF NERVOUS AND MENTAL DISEASE V198 5	
	[3] Hall Brian J., 2010, Exploring the Association Between Posttraumatic Growth and PTSD A National Study of Jews and Arabs Following the 2006 Israeli-Hezboliah War, JOURNAL OF N	
	<ul> <li>Bruckner Tim A, 2010, Male fetal loss in the US following the terrorist attacks of September 11, 2001, BMC PUBLIC HEALTH, V10, 6</li> </ul>	
	(4) Vazduzz Carmelo, 2010, Perceived behells and retronsist attacks. The folio of positive and negative emotions, Jourdoval, OF POSITIVE PSTCHOLOGY, V5, 10 Elitistic and an approximation of the state and the linked States Comparable Policy Review. DURINAL OF COMPARATIVE PDL (CV ANAL VSIX), V12, 24	
and the second s		
	· · · · · · · · · · · · · · · · · · ·	
4		1

Figure 107. The Summary of the current cluster (based on where you right clicked), including keywords, citing papers, and their DOI links. You can copy the URL of a DOI link and paste it to a browser to get the full text if you have an adequate subscription.

# 8 Additional Functions

The main menu provides access to additional functions.

### 8.1 Menu: Data

Data ► Import/Export

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

### 8.1.1 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.

```
Make sure that your MySQL server is on before you use this function in CiteSpace.
```

🛓 CiteSpace: Data Processing Utilities
CiteSpace Built-in Database WOS arXiv CNKI(RefWork) CSSCI Derwent* NSF Scopus SDSS Project DX
Database Project Articles Authors References Institutions Query History Help
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:IUsers\cc345 Browse WOS ¥ Import
Text Analysis
Term Extraction from Title and Abstract Fields
# words: min 2 max 4 Extract Parases. Extract Verbs. Summarize
Statistical Association Test           Phrase freq > 1         Snow Distributions           p-level         0.01           Uog-Likelinkood         Show Graph
⊂ SQL Query and Results
SELECT count(*) FROM articles
Results of your SQL query will appear here.
Search Save As WoS Save As CSV Plot SQL Results of [Year, Value]

Figure 108. 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.

### 8.1.1.1 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 109. 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.

SiteSpace: Data Processing Utilities							
CiteSpaceBuilt-in Database WOS arXiv CNKI(RefWork) CSSCI Derwent*	NSF Scopus SDSS Project DX						
Database Project Articles Authors References Institution	ions Query History Help						
Desired							
Project Show First Few Records							
Project Name RM_A Count Records Select a query he	ere 💌						
Warning: Queries with * are time-consuming. Execute them	with the Search button.						
Create a New Project							
Input Directory C:Users\cc345	irowse WOS 🔻 Import						
Text Analysis							
Term Extraction from Title and Abstract Fields							
# words: min 2 max 4 Extract Phrases Extract Ve	erbs Summarize						
Statistical Association Test							
Phrase freq > 1 Show Distributions p-level 0.01 v Log	Phrase freq > 1 Show Distributions p-level 0.01 v Log-Likelihood Show Graph						
SQL Query and Results							
select citations, uid, author, year, title, concat(source, '', volume, ', pp. ', page) from tations desc	, articles where project='RM_A' order by ci						
3330 WOS:000171898900054 Reya, T 2001 S	Stem cells, cancer, and cancer stem cells 🔺						
1852 WOS:000226214300031 Jain, RK 2005 N	Normalization of tumor vasculature: An en						
1225 WOS:000175350600022 Tomasek, JJ 2002 M	Ayotibroblasts and mechano-regulation of Maintenance of pluringtoney in human and						
867 WOS:00021499900004 Fuchs F 2004 S	Socializing with the neighbors: Stem cells						
Search Save As WoS Save As CSV Plot SQL	L Results of [Year, Value]						

Figure 110. Using a built-in function to find the most cited articles with a SQL query.

### 8.1.2 Utility Functions for the Web of Science Format

#### 8.1.2.1 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.

🗟 CiteSpace: Data Processing Utilities 📃 📼 📼	ζ
Cite Space Built-in Database WOS arXiv CNKI(RefWork) CSSCI Derwent* NSF Scopus SDSS Project DX	
Data Directories	
Input Directory Browse	
Output Directory Browse	
Remove duplicates (WoS) WoS (tab) => WoS WoS => Jigsaw WoS => Carrot2 (XML)	

Figure 111. Utility functions for handling bibliographic records in the Web of Science format.

#### 8.1.2.2 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)  $\rightarrow$  WoS."

### 8.1.2.3 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).

#### 8.1.2.4 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.

### 8.1.3 Scopus

The preferred format for Scopus data files is the RIS format. In Scopus, save your search results to one or more data files in RIS, e.g. my\_scopus\_search\_results.ris, to a folder on your computer. Specify the data folder and a folder where you want to have the converted files. And start the conversion process.

CiteSpace will tell you how many records in each RIS files have been converted and, more specifically, how many cited references in total are found in the data files and how many of them have been converted successfully. The 95.0% is a very decent successful rate, considering all the irregularities of the cited references.

If, for some reason, you have the Scopus data files in CSV instead of in RIS, you need to do a quick conversion from CSV to a tab delimited format using Excel before you can run the Tab Delimited Converter in CiteSpace.

SciteSpace: Data Process	ing Utili	ties								_ <b>D</b> X
CiteSpace Built-in Database	wos	Scopus	ADS	arXiv	CNKI	CSSCI	Derwent*	NSF	Project DX	
Instructions: In Scopus: 1. Save search results to RIS (ris) or CSV formatted files in a folder on your computer For CSV files, open them in Excel and save them as tab delimited files with filenames starting with the word 'download', e.g. download_tab.bt In CiteSpace:										
Data Directories	2. Configure the input and output folders accordingly 3. Start the conversion Note: The successful rate of each conversion varies with a possible loss of 5%-7% of the original cited references due to data irregularity. Data Directories									
Output Directo	ory D:\Da	ta\Scopus\	Converte	d					Brov	wse
		Scopus	(RIS) →	WoS	Scop	us (Tab De	elimited) → V	loS		
1036 records converted from sco 285 records converted from scop	pus.ris us_294ci	ters to cite:	space II p	paper.ris						
Total References: 57213 Valid References: 54555 (95.0%	)									

Figure 112. Converting Scopus data files to the WoS format for CiteSpace.

4,658 KB 1,176 KB

Figure 113. Converted Scopus data are saved in the designated output folder.

### 8.1.4 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.

CiteSpace 3.8.R2 (64-bit) - (c) 2003-2014 Chaomei Chen - Home: C:\Users\cc345 - CC la Project Data Network Visualization Generanbical Analytics Text Preferences Help	A				
Web of Science PubMed	Time Sticing From 2000 V To 2014 V Wears Per Stice 1 V				
Read MCB protocol Run this function /pm-Sam Eastern Time or weekeeds only! Home Directory CrUtersicc345PubMed Search Results CrUtersicc345PubMedSearchResults Genry Inpertnasion (mt) 2006 2014 28 Search Stop COC Stop Reset JvM Memory 982 (MB) Used 20 %	Text Processing Term Source Title R Abstract R Author Keywords (DE) Reywords Plus (D) Term Type Nous Phrases Burst Terms Detect Bursts Estrogy Network Configuration Node Types				
paro Status					
	Top II per slice Top IVs per slice Threshold Interpolation Select Citres Select top 50 most cited or occurred items from each slice.				
tocess Reports					
Saved: download2012.bt	Pruning				
Savet. download2014.bt Done!	Pathfinder     Pruning sliced networks     Minimum Spanning Tree     Pruning the merged setwork				
	Manafineties				
VOTE: Since PubMed records do not include cited references, citation analysis is not feasible.	Venalization				

Figure 114. Retrieve bibliographic records from PubMed.

le Project Data Network Visualization Geographical Analytics Text Preferences Help					
Web of Science Publied	Time Slicing				
New PubMed Test	From 2008 V To 2014 V Prears Per Sace 1 V				
	Text Processing     Term Source				
Project Home: C:Waers/oc345/PubMed/project	☑ Title ☑ Abstract ☑ Author Keywords (DE) ☑ Keywords Plus (ID)				
Data Directory: C:Usersicc345PubMed/SearchResults	Noun Phrases Burst Terms Detect Bursts Entropy				
GOL Stop Reset JVIII Memory 982 (MB) Used 20 %	Network Configuration				
pace Status-	Author Chattering Country & Term C Keyword C Category				
deSpace is pre-processing data files. Please wall	Change Casedoon Coondy & term Cheyword Conegory				
ears: 7 nique source records: 175	Cited Reference Cited Author Cited Journal Paper Grant				
	Strength Cosine  Scope Within Slices				
	Top N per slice Top N% per slice Threshold Interpolation Select Citers				
-	Select top 50 most cited or occurred items from each slice.				
recess Reports	_				
	Pruning				
	Pathfinder     Pruning sliced networks     Minimum Spanning Tree     Pruning the merged network				
	Visualization				
	Cluster View - Static     Show Networks by Time Slices				

Figure 115. Analyzing the PubMed records ...



Figure 116. A network of co-occurring noun phrases on hypertension.

### 8.2 Menu: Network

### 8.2.1 Batch Export to Pajek .net Files

### 8.3 Menu: Geographical

### 8.3.1 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.

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

1. Specify the Time Frame	and the second	
	From Year 2006 To Year 2014	
2. Select the Directory of In	put Data in Field Delimited Format	
Data Directory D:Data/Ch	en/CMCitedB/idata	Browse
Options		
	Show Links Refresh Previously Generated Files	
Transpa	arency 00 (clear) - ff (opaque) Nodes: 3a Links: 3a	
	Make Map	

Figure 117. 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.

Messag	je X
i	The KML file D:\Data\ChenCMCitedBy\data\kml\master-medium.kmz is created! # Records processed: 6156 # Errors encountered: 346
	OK

Figure 118. 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.

Name	Date modified	Туре	Size
geocoding_log_tab	4/12/2014 10:55 PM	TXT File	47 KB
🚳 locations-2006	4/12/2014 10:53 PM	Microsoft Excel Co	6 KB
💫 locations-2007	4/12/2014 10:54 PM	Microsoft Excel Co	8 KB
🚯 locations-2008	4/12/2014 10:54 PM	Microsoft Excel Co	11 KB
🚯 locations-2009	4/12/2014 10:54 PM	Microsoft Excel Co	16 KB
🚳 locations-2010	4/12/2014 10:55 PM	Microsoft Excel Co	14 KB
🔊 locations-2011	4/12/2014 10:55 PM	Microsoft Excel Co	21 KB
🔊 locations-2012	4/12/2014 10:55 PM	Microsoft Excel Co	17 KB
locations-2013	4/12/2014 10:55 PM	Microsoft Excel Co	15 KB
🔊 locations-2014	4/12/2014 10:55 PM	Microsoft Excel Co	5 KB
🖏 master-medium	4/12/2014 10:55 PM	KMZ File	59 KB

Figure 119. 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 120. 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 121. Unfold the list of places on the left and locate a city of interest - Frederick, Maryland, USA - on Ebola.

Click on any of the papers on the list to explore its content. Here is an example of what you will see after clicking on a link to our 2010 JASIST paper in Google Earth.

Intel work water       Intel work water <td< th=""><th>File Edit View Tools Add Help</th><th></th><th></th><th></th></td<>	File Edit View Tools Add Help			
Vertex	* Search	At Back to Canada Fauth   http://actually.com/doi.org/doi	In a second s	
	* Search	W BACK to Google Earth   http://orwinetorary.wiley.com/ dou	to toogenizziony and et - The structure and dynamics or cociation clusters: A multiple-perspectiv	e coccation analysis •
Places       Final-Stability       Final-Stability       Computer Science > General & Minductory Computer Science > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Sciency for Mindmanton Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Province of Cocket and > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the sum > Journal of the American Science and Technology > Vel of the science and Techn	to:39.95969, -75.19706 (Philadelphia) Search ec: 37 25.818' N, 122 05.36' W Get Directions History	Wiley Online Lib	Draxel Unive	rsity Libraries 📥 Log in / Register 🛛 O
Control.United States     Control.Unite	Places     P      Murcia.Spain     V     Narachine chi lanan	Home > Computer Science > Abstract	General & Introductory Computer Science > Journal of the American Society for Information Science	e and Technology > Vol 61 Issue 7 >
Normal Auford     Phaladelpha United States     Control CADI     Phaladelpha United States     Control CADI     Control	Variashino-ShUapan     Variashino-ShUapan     Viano-United States     Viano-United Kingdom     Viano-United States	JOURNAL TOOL S & Get New Carton Livers Get RSS feed Save to My Profile	JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	asist
SCITEC       Finishing       Characteristic       Science       Scienc	Philadelphia,United States     Philadelphia     * Chen (2010) J AM SOC INF.	JOURNAL MENU Journal Home	Research Article The structure and dynamics of cocitation clusters: A multiple-perspective of the structure and s	SEARCH In this assue
Poladela Autoria State:	SCI TEC	FIND ISSUES All tissues Virtual lissues	Chaomei Chen <sup>1</sup> , Fidelia Ibekwe-SanJuan <sup>2</sup> Issue and Jianhua Hou <sup>3</sup>	Advanced > Saved Searches >
Image: Control of the second seco	<ul> <li>Madelphia,United States Lyon J = Philadelphia,United States Lyon,France     </li> <li>Porto,Portugal</li> </ul>	FIND ARTICLES Early View Most Accessed Boot Cred	Article first published online. 18 MAR 2010 DOI: 10.1002lasi.21309 e.2010.45557	ARTICLE TOOLS Cet PDF (7862K) Save to My Profile E-mell Let to the Article
Porticipy Distance Survey     For Calculation       Porticipy Distance Survey     For Survey       Porticipy Distance Survey     For Porticipy Distance Survey       Porticipy Distance Survey	Rotterdam,Netherlands     San Marcos,United States	GET ACCESS Subscribe / Renew		Export Citation for this Article
Advert THE JOBANA     Primary Database     Pri	Santa Barbara,United States     * *	FOR CONTRABUTORS OnlineOpen Author Guidelines Submit an Article	Additional Information (Show All)	Share   🔜 🕅 🔐 💟
Protos     Conad     Conad     Protos     Roads     Sector KFATURES     Access (FFATURES     Access     Access (FFATURES     Access     Access     Access (FFATURES     Access	Primary Database     Pimary Database	ABOUT THE SOURNAL Overview Editorial Sourd Permissions Advertise	How to Care - Automic Internation - Production History Supplementary materials of the subdy are evaluable at http://cluster.cols.coveel.educonen/pspers/2008/infosol.	
Adda of Spanickeys     Adda of Spanickey	Roads	Contact SPECIAL FEATURES	Link to article View Full Article (NTML)	Article (HTML)   🛣 Get PDF (7882K)
If allery to internation Generation Contation networks are decomposed into contation dusters. The interpretation divides of the internation dinternation	Ceean     Gabel Awaraneer	ASSAT Digital Literary Actices in the Advances in Information Science Building of the American Society for Information Science and Technology	Abstract [references A multiple-perspective colution analysis method is introduced for characterizing and interpreting t clusters: The method facilities analytic and sense making tasks by integrating network visualization cluster tabeling, and test summarization. Coclatation networks are decomposed into contation duats is augomethed by automatic ducet tabeling and summarization. The method focuses on the inferre-	the structure and dynamics of cocitation on, spectral clustering, automatic iers. The interpretation of these clusters reations between a cocitation cluster's

Figure 122. Clicking on the Chen (2010) link takes us to the publisher's page of the paper.



Figure 123. Here is a bird eye view of the downtown Philadelphia. The red bar on the left is on Drexel's main campus.

### 8.4 Menu: Overlay Maps

Dual-map overlays are introduced in (C. Chen & Leydesdorff, 2014).

This function is made available for non-commercial and educational use. For commercial use, please contact me at <u>chaomei.chen@drexel.edu</u> directly for further detail.
(LiteSpace 3.8.R7 (64-bit) - (c) 2003-2014 Chaomei Chen - Home: C:\Users\cc345			
File Project Data Network Visualization Geographical Overlay Maps Analytics Text Prefe	erences Help		
Web of Science PubMed JCR Journal Maps	Time Slicing		
Projects	From 2000 💌 To 2014 💌 #Years Per Slice 1		
New ABCD Unique 💌 More Actions 💌	r Text Processing		
	Term Source		
Project Home: D:\Data\WoS\Regenerative medicine\Project ABCD unique P	✓ Title ✓ Abstract ✓ Author Keywords (DE) ✓ Keywords Plus (ID)		
	ј <sub>Г</sub> Тегт Туре ————————————————————————————————————		
Data Directory: D:/Data/WoS/Regenerative medicine/ABCD uniuqe y15 n713	Noun Phrases O Burst Terms Detect Bursts Entropy		
GOL Stop Reset JVM Memory 982 (MB) Used 14 %	Network Configuration		
	Node Types		
	Author Christiation Country Cherni Ckeyword Category		
	Cited Reference      Cited Author      Cited Journal      Paper      Grant		
	Links		
	Strength Cosine   Scope Within Slices		
	Top N per slice Top N% per slice Threshold Interpolation Select Citers		
	Select top 50 most cited or occurred items from each slice.		
Process Reports			
fi fi			
	Pruning		
	Pruning sided networks      Minimum Spanning Tree     Pruning sided network		
	Viewalization		
	Cluster View - Static     Show Networks by Time Slices		
	Cluster View - Animated		

Figure 124. The dual-map overlay maps function is accessible from the Overlay Maps menu.

# 8.4.1 Add an Overlay

The current version allows you to add up to 12 overlays. Each overlay is represented by a distinct set of bibliographic records in the WoS format. Start the process by clicking on the Add button.



Figure 125. Start the process of adding an overlay by clicking on the Add button.



Figure 126. Select 'w' for bibliographic records in the WoS format. The 'p' option is not available in the current release.

	Select a D	Livectory of Data Files in WoS Format
	] data_2284 ] kml	4
	_ ] project 10	)-19-2014
Fo	lder <u>n</u> ame:	D:DataWoS\Ebola - Topic Search n2284\data_2284
Fo	lder <u>n</u> ame: es of <u>T</u> ype:	D:\Data\WoS\Ebola - Topic Search n2284\data_2284 All Files

Figure 127. Select the data directory of a dataset in the same way as you create a CiteSpace project.

Input	
?	Choose a color for the new overlay: red, green, yellow, blue, pink, orange, black, white, darkgray, gray, cyan, magenta red
	OK Cancel

Figure 128. Select a color for the new overlay and wait for the display to update ...



Figure 129. The result of adding a set of bibliographic records in the WoS format.



Figure 130. The Border button controls the display of the border of a cluster and whether to fill these areas in color.

#### 8.4.2 Further Reading and Terms of Use

Technical details and several examples are provided in (C. Chen & Leydesdorff, 2014).

This function is made available for non-commercial and educational use. For commercial use, please contact me directly for further detail.

If you use the dual-map function in your publications, you should cite it as follows:

Chen, C., Leydesdorff, L. (2014) Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. Journal of the American Society for Information Science and Technology, 65(2), 334-351.

# 8.5 Menu: Text

#### 8.5.1 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.

🛓 CiteSpace 3.8.R2 (64-bit) - (c) 2003-2014 Chaomei Chen - Home: C:\Users\cc345							
File Project Data Network Visualization Geographical Analytics	Text Preferences Help						
Veb of Science PubMed Projects New Demo More	List Ranked Terms by th°idf List Ranked Terms by Clumping Properties Extract Terms from a FullText File Plot Information Entroov	From 1996 V To 2003 V #Years Per Slice 1 V					
Project Home: C:USers\cc345\citespace\Examples\Project	Process Multiple Folders of FullText Files Build Concept/Predicate Trees (Cut and Paste Build Concept/Predicate Trees (from Fulltext I	Abstract Z Author Keywords (DE) Z Keywords Plus (ID)					
Data Directory. Disersic:340(CitespaceExamples)Data(Terro	Build Concept/Predicate Trees (from WoS File View a Saved Concept Tree (*.xml)	n Phrases O Burst Terms Detect Bursts Entropy					
	Latent Semantic Analysis	n					
- Space Status	Edit n-grams in word.list	Institution       Country       Term       Keyword       Category         ence       Cited Author       Cited Journal       Paper       Grant         Strength       Cosine         Scope       Within Slices           Top N% per slice       Threshold Interpolation       Select Citers           Select top       50       most cited or occurred items from each slice.					
	Pruning     Pathinde     Minimum     Visualization     Cluster Vie     Cluster Vie	- Static - Animated  Pruning sliced networks Pruning the merged network Show Networks by Time Slices - Animated					

Figure 131.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.

Select a Directory or One or More Files	of Data	Select a Directo	ry or One or More Files	of Data
Look In: Terrorism	• <b>6 1 8 5</b>	Look In: 📑 cluste	rs	- G A C 885
clusters           13.graphmi           151.graphmi           269.graphmi           309.graphmi           310.graphmi           4	531.graphml     619.graphml     653.graphml     3636.graphml     AFTERnet_1996-2003_v309_€     BEFOREnet_1996-2003_v309.	0.txt     0.txt     0.txt     1.txt     1.txt     2.txt     2.xml	3.bxt         3.xmi         4.bxt         4.xmi         5.bxt         5.xmi	6.bxt     6.xmi     7.txt     7.xmi     8.bxt     8.xmi
File Name: clusters Files of Type: All Files	V Open Cancel	File <u>N</u> ame: 0.bt Files of <u>Type</u> : All Fi	les	▼ Open Cancel

Figure 132. 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 133. 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 134. The PTSD cluster. Key concepts: September, children, same month, and terrorist attacks.



Figure 135. 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.

#### 8.5.2 List Terms by Clumping Properties

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

For example, Text  $\triangleright$  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.

and the second second	oject Data Net	work Visua	lization Geographical Analytics	Text Preferences Help			
Vieb of Science PubMed Projects New Demo Vibors Project Home: C:Usersicc345:citespace/Examples/Project				List Ranked Terms by thidf			
				List Ranked Terms by Clumping Properties Extract Terms from a FullText File Plot Information Entropy Process Multiple Folders of FullText Files Build ConceptPredicate Trees (Cot and Paste) Build ConceptPredicate Trees (from Fulltext Files)		From 1996 V To 2003 V #Years Per Slice 1 V	
			▼ More				
			34E decession			Abstract Author Keywords (DE) Keywords Plus (ID)	
	basa birecii	ny. Joessa	345Collespaceexamples coatan reno	Build Concept/Predicate Trees (from WoS Files)		n Phrases O Burst Terms Detect Bursts Entropy	
-				View a Saved Concept Tree	(".xmi)		
	GOL	Stop	Reset JVM Memory 982 (M	Latent Semantic Analysis			
Proget Maras Consporte Instanting monitorian nom mes in vools roma Seeder Just pankot ferma Project C:Usersicc349; citespaceExamples:Projects(Terrorism				Cited Reference Cited Author Cited Journal Paper Grant			
			information of the second s		5	trength Cosine  Scope Within Slices	
					Ton N per slice	on N% ner slice   Threshold Internolation   Select Citers	
				×	Top N per slice	op IVs per slice   Threshold Interpolation   Select Citers   lect top  50   most cited or occurred items from each slice.	
Proces	s Reports			•	Top N per slice Te	op INs per slice   Threshold Interpolation   Select Citers   lect top 50 most cited or occurred items from each slice.	
Toces 1 208 72 48	s Reports Idf thidf 1.79 372 3 215 34 163	clumpin 89 1.07 58 1.02 26 1.02	g termi terronist attacks terronist attack wordt trauts center	•	Top N per slice To	op HN per slice   Threshold Interpolation   Select Citers	
roces f 208 72 48 40	s Reports lat thiat 1.79 3721 3 2151 3.4 1633 3.58 1433	diumpin 99 1.07 58 1.02 26 1.02 34 1.01	a termi terrotist attacks berrotist attack wordf stads center mass destruction	•	Top N per slice To Se	op IVs per slice Threshold Interpolation Select Citers	
roces f 208 72 48 40 36 32 30	s Reports idf third 1.79 372. 3 215 3.4 163. 3.58 143. 3.69 122. 3.81 121. 3.81 121. 3.87 116.	clumpin 59 1.07 58 1.02 52 1.02 34 1.01 8 1.01 51 1.01 14 1.01	g term terrotet uttacks terrotet attack word frada center mass destruction biological terrotem biological wespons new york db	•	Pruning Puthfinder Minimum Spann	op Ifly per slice Threshold Interpolation Select Citers  lect top 50 most cited or occurred items from each slice.	
Proces 17 208 72 48 40 30 30 29 27	s Reports idf thiaf 1.79 372: 3 216 3.4 163: 3.69 132: 3.61 132: 3.81 121: 3.87 116: 3.87 112: 3.87 112: 3.87 112: 3.92	clumpin 59 1.07 59 1.02 50 1.02 34 1.01 51 1.01 51 1.01 54 1.01 54 1.01 54 1.01	g termi terroisti attackij terroisti attacki wordi hada center masa destručani zelošgozi terroism zelošgozi terroism zelošgozi zeloš zeloš zelošgozi zelošgozi zelošgozi zeloš zelošgozi zeloška zeloška zeloš		Top N per slice Tr Se Pruning Pathfinder Minimum Spann Visualization	op IVs per slice Threshold Interpolation Select Citers Nect top 50 most cited or occurred items from each slice.	

Figure 136. List ranked terms by clumping properties.

## 8.5.3 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.

Data Sourc	e Selection						
Directory of Input Data:		ory of Input Data:				Browse	
		Data Type:  Fulltext  Web of Science Merker:  F2					
			Add to the List	Clear	Analyze		
Data Type	Marker	Path					
Fulltext	NO	D temp/fulltext	E%				
Wos	#1	C:\Users\cc34	51. citespace/Example:	Data\Terror	ism1990-2003		
REPORT							
98.08910234	4388188						
	32.04 terro	rism					
	25.18 terro	rist					
	20.24 attac	85					
	20.03 sept	emper					
14 000 1050	19.02 hean	un .					
1,0084050	7.50 conto	2222					
	7.30 septe	liver					
	7 31 stack						
	7 30 terrori	sm					
	6.75 terrori	st					
59.41388550	08132125						
	30.08 terro	rism					
	10.01 secu	irity					
	6.69 politic	al					
	6.44 interna	ational					
	5.70 threat						
55.6663889	1173457						
	15.54 terro	rism					



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 138. A visualization of the Latent Semantic Space, the 2<sup>nd</sup> and the 3<sup>rd</sup> dimensions.

# 9 Selected Examples

Here are some good examples I came across on the Internet. These examples are created by users with CiteSpace.

The following visualization is from blog.sciencenet.cn by Jie Li.



#### Chen, C. (2014) The CiteSpace Manual. http://cluster.ischool.drexel.edu/~cchen/citeSpace/CiteSpaceManual.pdf



Figure 140. An article published in Plos One, Figure 3. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018537



Figure 141. An article published in Liber Quarterly, Figure 4. https://liber.library.uu.nl/index.php/lg/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113638/8398

# **10 Metrics and Indicators**

# 10.1 Information Theoretic

### **10.1.1 Information Entropy**

CiteSpace computes information entropy based on noun phrases extracted from records to represent the diversity of a data set.

## 10.2 Structural

#### **10.2.1** Betweenness Centrality

The betweenness centrality of a node in a network measures the extent to which the node is part of paths that connect an arbitrary pair of nodes in the network (Brandes, 2001; C. M. Chen, 2006; Freeman, 1977).

#### **10.2.2 Modularity**

The modularity of a network measures the extent to which a network can be decomposed to multiple components, or modules. This metric provides a reference of the overall clarity of a given decomposition of the network (C. Chen et al., 2010).

The modularity change rate induced by a set of incoming information is considered to be a sign of a potentially important perturbation to a complex adaptive system (C. Chen, 2012).

#### 10.2.3 Silhouette

The silhouette value of a cluster measures the quality of a clustering configuration. Its value ranges between -1 and 1. The highest value represents a perfect solution. However, to ensure a sound interpretation in CiteSpace, it is recommended that you should balance the modularity and silhouette scores simultaneously (C. Chen et al., 2010).

#### 10.3 Temporal

#### 10.3.1 Burstness

The burstness of the frequency of an entity over time indicates a specific duration in which an abrupt change of the frequency takes place (Kleinberg, 2002). In CiteSpace, citation burst and occurrence burst are both supported.

#### 10.4 Combined

## 10.4.1 Sigma

This indicator measures the combined strength of structural and temporal properties of a node, namely, its betweeness centrality and citation burst (C. Chen et al., 2009).

# 10.5 Cluster Labeling

# **10.5.1** Term Frequency by Inversed Document Frequency

### 10.5.2 Log-Likelihood Ratio

# **10.5.3 Mutual Information**

# **11 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.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2), 163-177.
- Carrot2: Open source framework for building search clustering engines. (2012). from <u>http://project.carrot2.org/</u>
- Chen , C. (2004). Searching for intellectual turning points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci. USA, 101*(Suppl.), 5303-5310.
- Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, 63(3), 431-449. doi: 10.1002/asi.21694
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209.
- 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., & Leydesdorff, L. (2014). Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology*, 65(2), 334-351. doi: 10.1002/asi.22968
- 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.
- Freeman, L. C. (1977). A set of measuring centrality based on betweenness. *Sociometry*, 40, 35-41.
- 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.