

CiteSpace常见问题解答

李杰

首都经济贸易大学 安全与环境工程学院

个人主页: <http://blog.sciencenet.cn/u/jerrycueb>

写在前面

对于初学者使用CiteSpace来讲，会遇到一些常见的问题。笔者将比较重要的总结为以下几点，欢迎大家继续添加。

- ① 下载的WoS数据为什么不能做文献共被引分析？
- ② 图谱左上角的参数是什么意思？图谱参数在什么范围比较合理？
- ③ 可视化界面中的各个界面功能是什么？（包含节点属性、标签属性以及聚类方法的介绍）
- ④ 关于网络的布局问题，为什么重新运行后图谱整体的布局不一样了？
- ⑤ 名词性术语的提取，为什么提取不出来？
- ⑥ 网络中相同含义的词汇如何合并（单复数、英式和美式英语以及同义词合并）？
- ⑦ 在CiteSpace中关键文献如何确定？

1.下载的WoS数据为什么不能做文献共被引分析？

为了保证进行文献共被引分析，收集数据时包含参考文献信息是至关重要的。

可以按照下面步骤收集数据，或可参照详细版数据收集方法

http://cluster.ischool.drexel.edu/~cchen/citespace/doc/tutorial/how_to/1.download_from_WOS.pdf



发送至文件

记录数: ☐ 页面上的所有记录

☒ 记录 1 至 500

记录内容: 全记录与引用的参考文献

作者、标题、来源出版物
作者、标题、来源出版物、摘要
全记录
全记录与引用的参考文献

文件格式 2

发送

取消

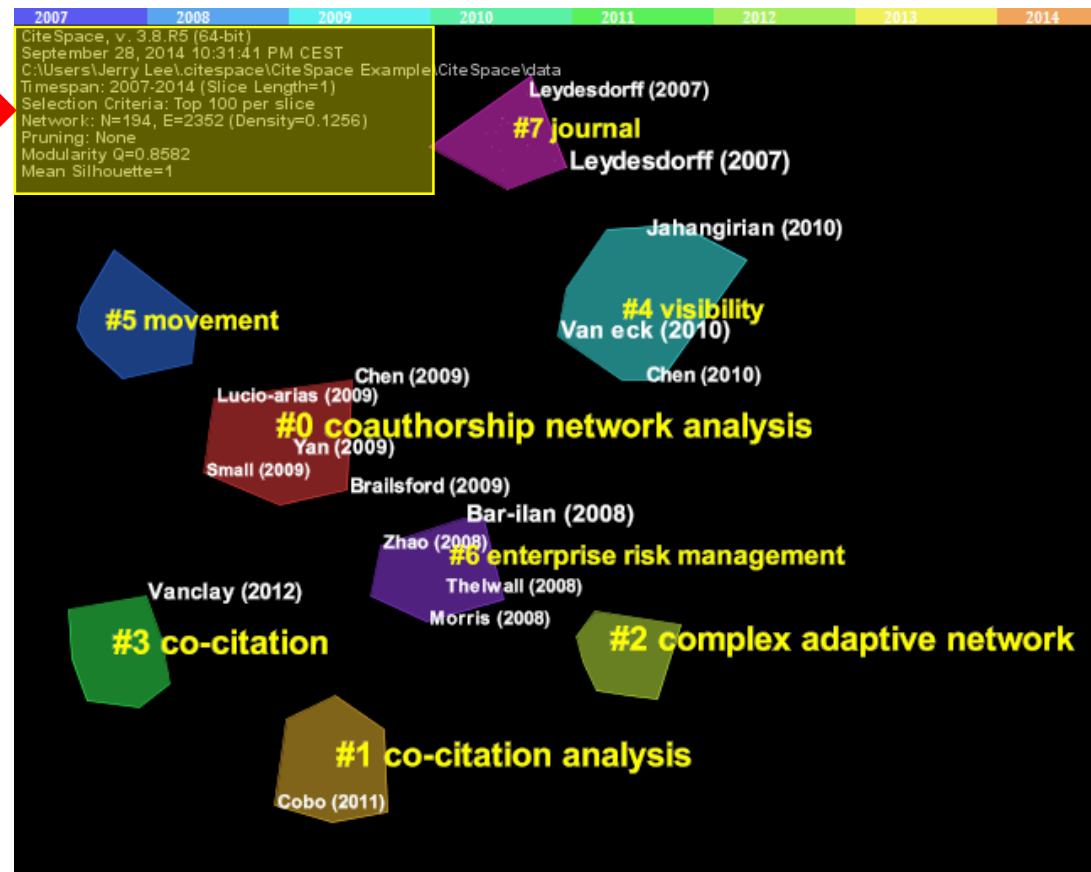


2.图谱左上角的参数是什么意思？图谱参数在什么范围比较合理？

```
CiteSpace, v. 3.8.R5 (64-bit)
September 28, 2014 10:31:41 PM CEST
C:\Users\Jerry Lee\.citespace\CiteSpace Example\CiteSpace\data
Timespan: 2007-2014 (Slice Length=1)
Selection Criteria: Top 100 per slice
Network: N=194, E=2352 (Density=0.1256)
Pruning: None
Modularity Q=0.8582
Mean Silhouette=1
```

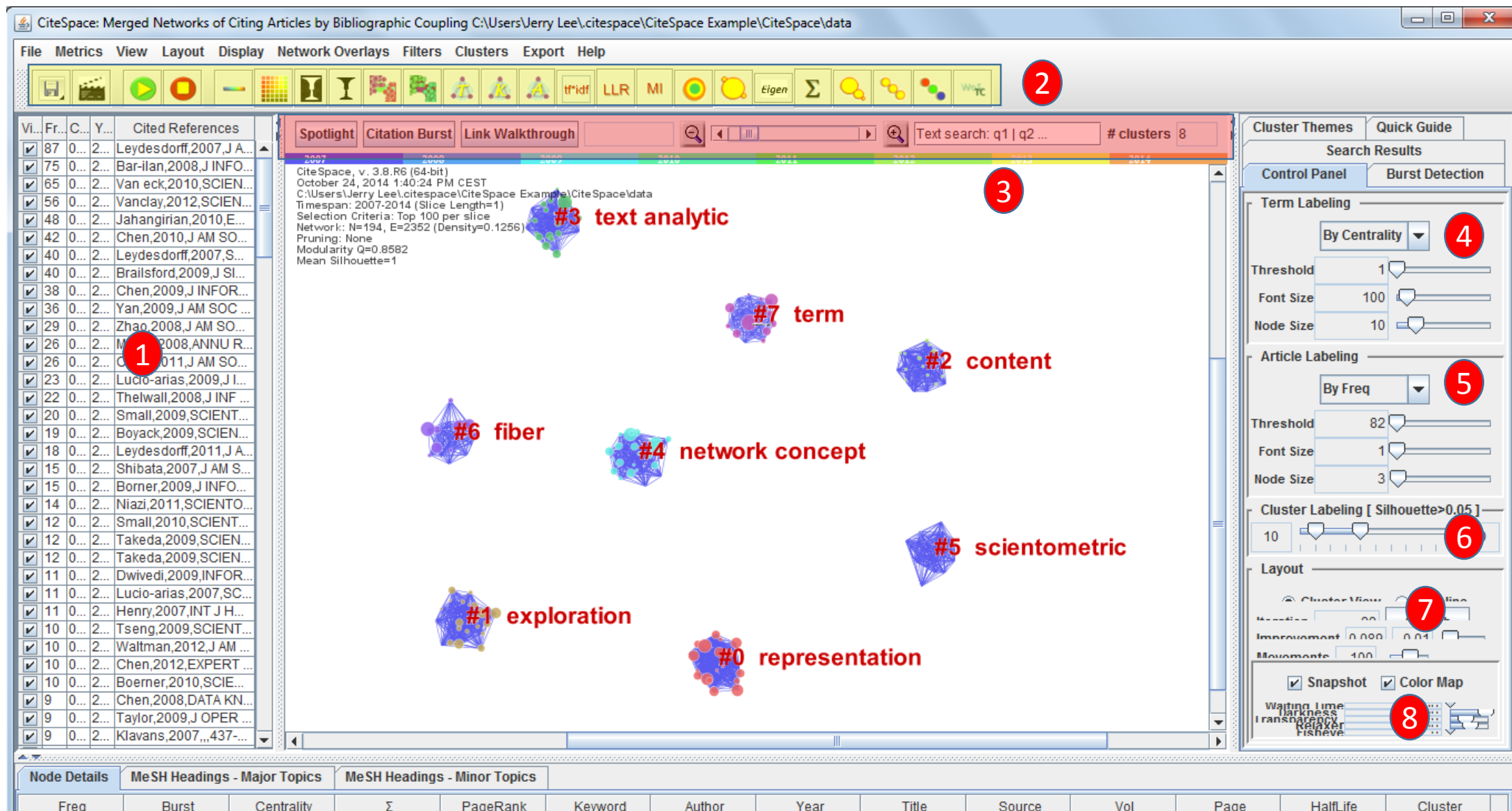


- ① CiteSpace, V.3.8 R5(64 bit)表示使用软件的版本信息
- ② September 28,2014 10:31:41PM CEST表示进行结果计算时的时间
- ③ C:\User\Jerry Lee\.CiteSpace... 表示数据所存放的文件夹位置
- ④ Time Span: 2007-2014(slice Length=1)表示所分析的时间区间，括号中代表的是时间切片。也就是说把这个时间区间按照多少年为一段进行切割。
- ⑤ Selection criteria: Top100 per slice表示的是提取了每个时间切片排名前100位的数据来生成最终的网络（这里选用的节点类型不同，top100的具体含义会有差异。如选择的是作者合作分析时，则提取的是这个时间段内发文量top 100的作者，做共被引分析时则提取的是被引频次在每个时间切片top100的数据）。
- ⑥ Network:N=194, E=2352（density=0.1256），N表示网络节点数量，E表示连线数量，Density则表示网络的密度

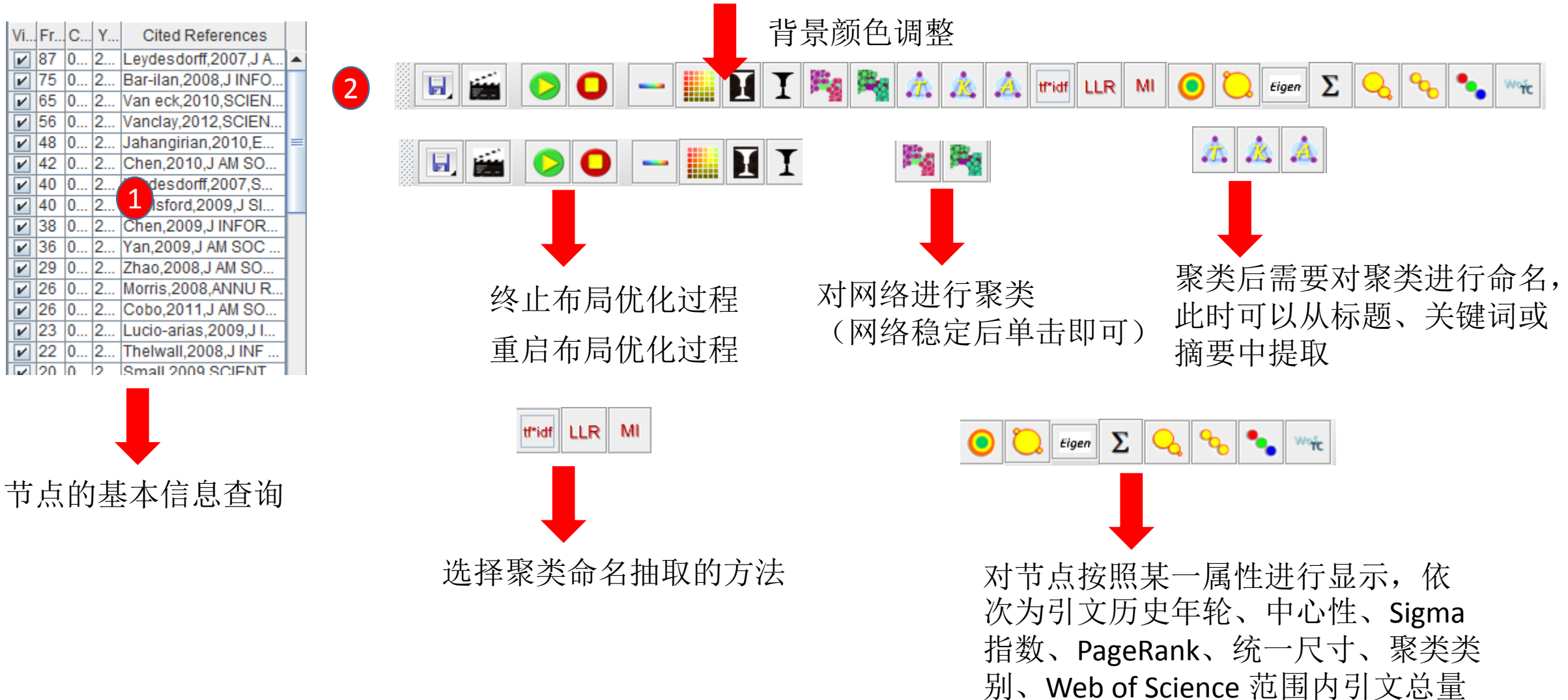


- ⑦ Pruning表示网络裁剪的方法，这里None表示没有剪裁。
- ⑧ Modularity表示网络的模块度，值越大表示网络的聚类结果越好。
- ⑨ Mean Silhouette=1, Silhouette值是用来衡量网络同质性的指标，越接近1，反映网络的同质性越高（注意Silhouette主要在聚类后来衡量某个聚类内部的同质性，但是在聚类内部成员很少时，这个值的信度会降低）

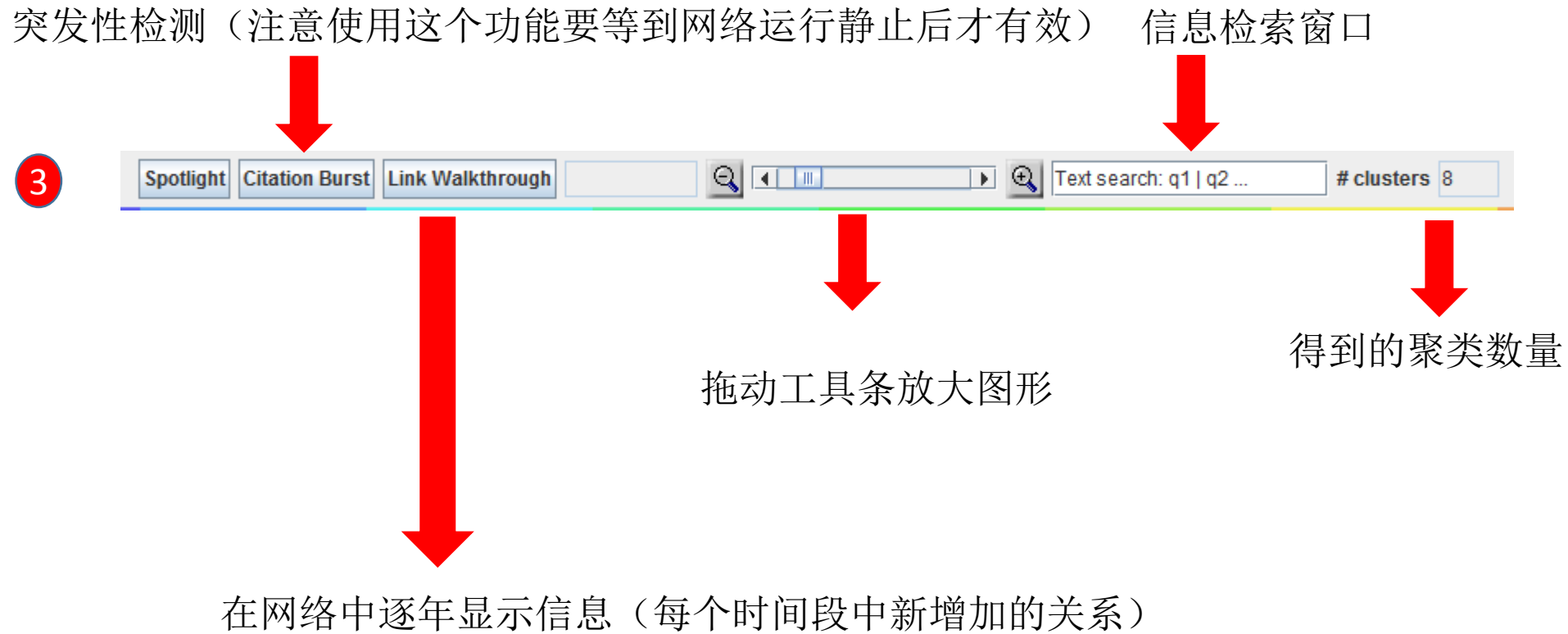
3.可视化界面中的各个界面功能是什么？（包含节点属性、标签属性以及聚类方法的介绍）



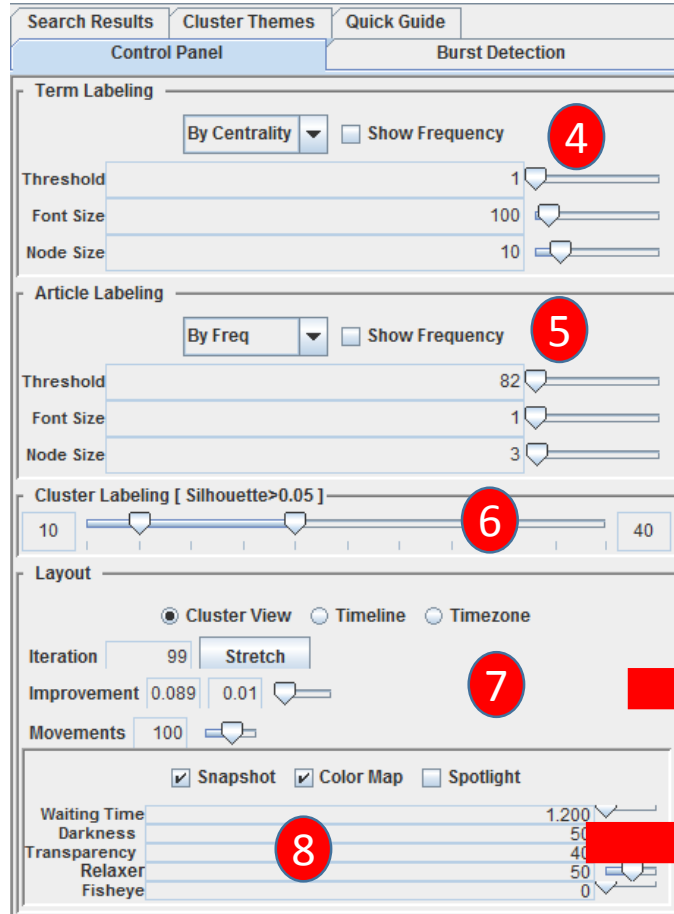
3.可视化界面中的各个界面功能是什么？（包含节点属性、标签属性以及聚类方法的介绍）



3.可视化界面中的各个界面功能是什么？（包含节点属性、标签属性以及聚类方法的介绍）



3.可视化界面中的各个界面功能是什么？（包含节点属性、标签属性以及聚类方法的介绍）



进行主题（Term）分析后，节点的属性和显示阈值需要通过此处调整

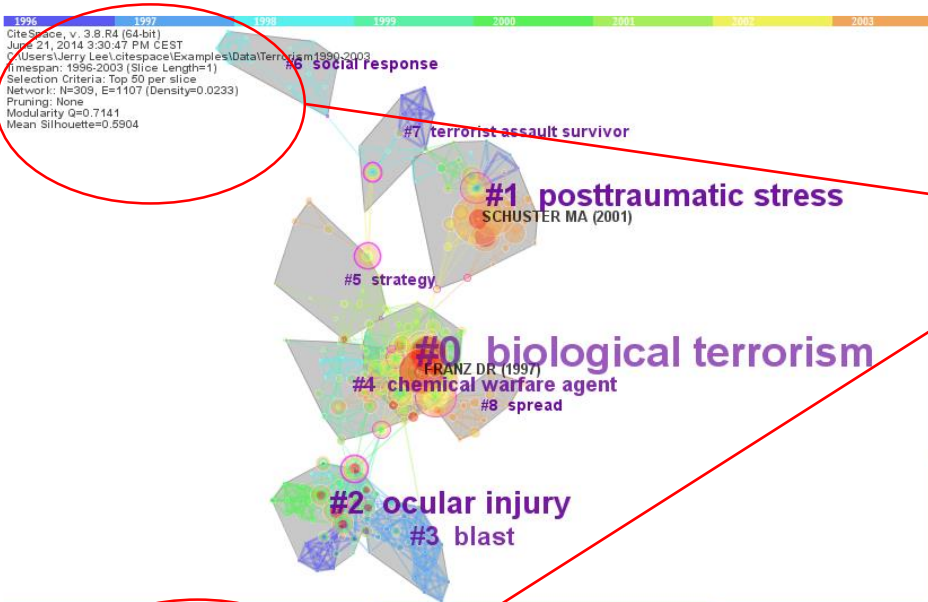
除了Term分析之外，其他节点的属性都是通过此处调整

对聚类标签的大小进行调整

对分析结果进行不同视图的展示

对网络连线的透明度等方面进行调整

4.关于网络的布局问题，为什么重新运行后图谱整体的布局不一样了？

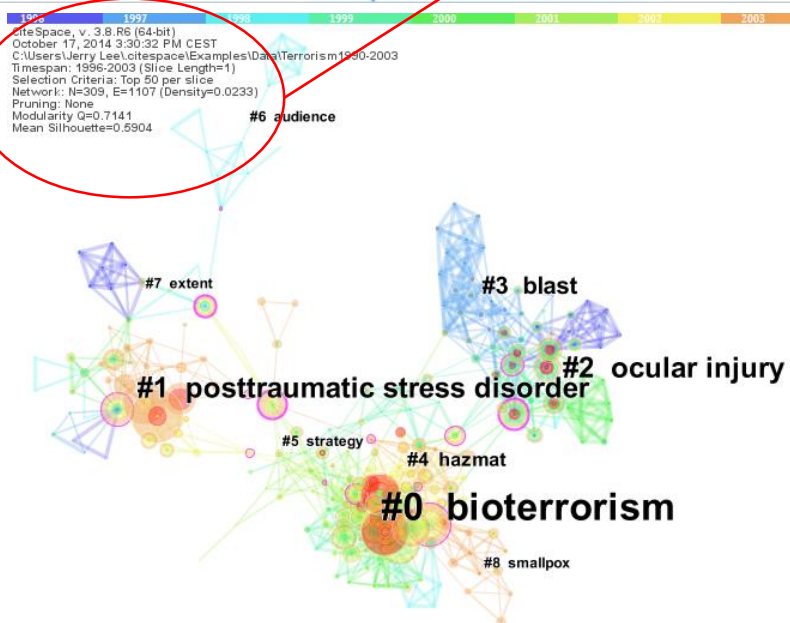


CiteSpace, v. 3.8.R6 (64-bit)
October 17, 2014 3:30:32 PM CEST
C:\Users\Jerry Lee\citespace\Examples\DT
Timespan: 1996-2003 (Slice Length=1)
Selection Criteria: Top 50 per slice
Network: N=309, E=1107 (Density=0.0233)
Pruning: None
Modularity Q=0.7141
Mean Silhouette=0.5904

左图使用的是不同版本CiteSpace对同一数据集的分析（当然使用相同版本软件分析相同数据集的结果类似）。可以看到节点和连线在空间的布局有一些不同，进一步检查网络的参数发现二者是相同的，这说明两个网络本质是一样的。

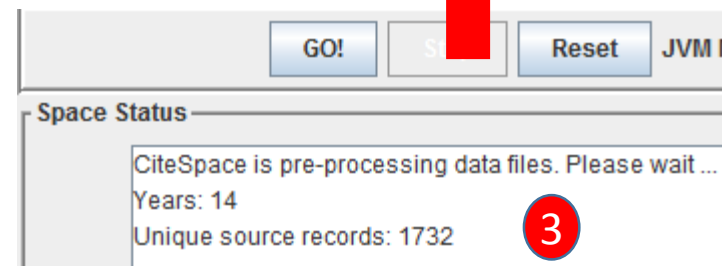
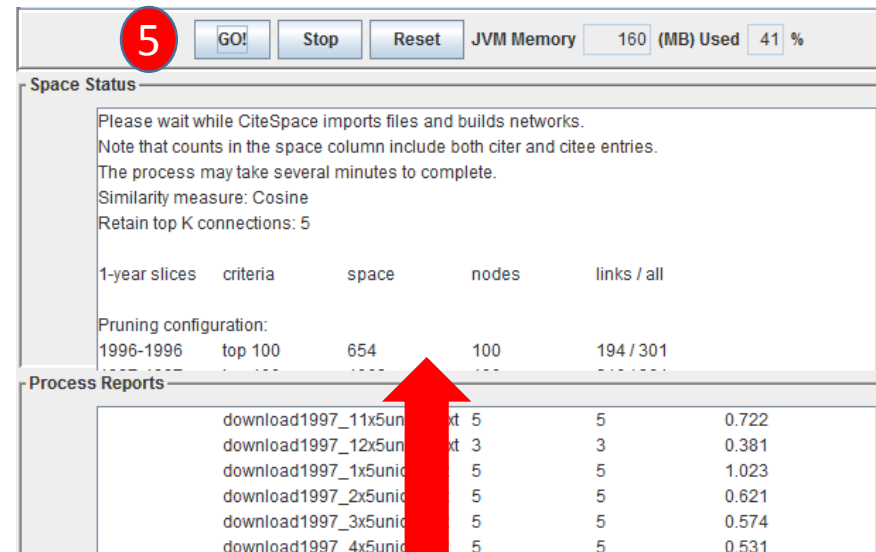
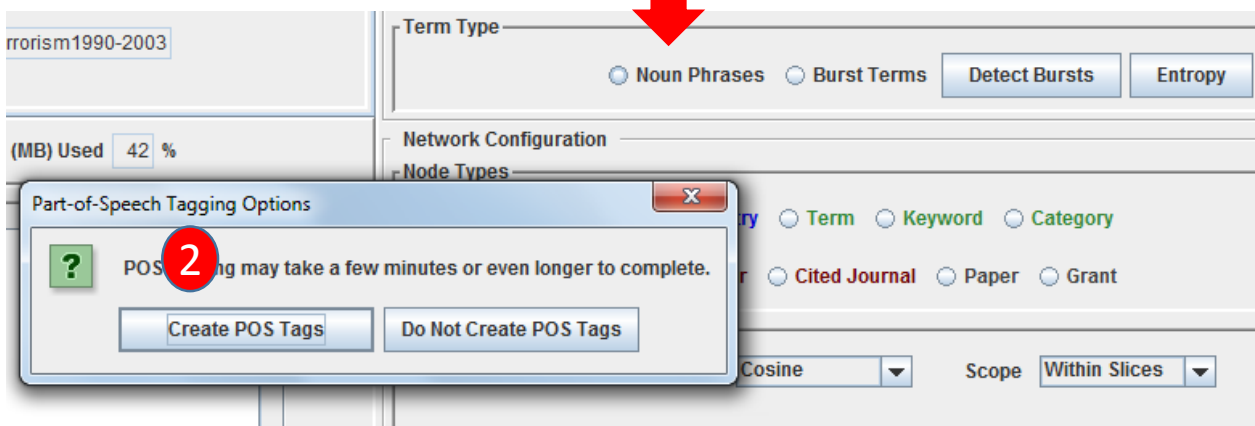
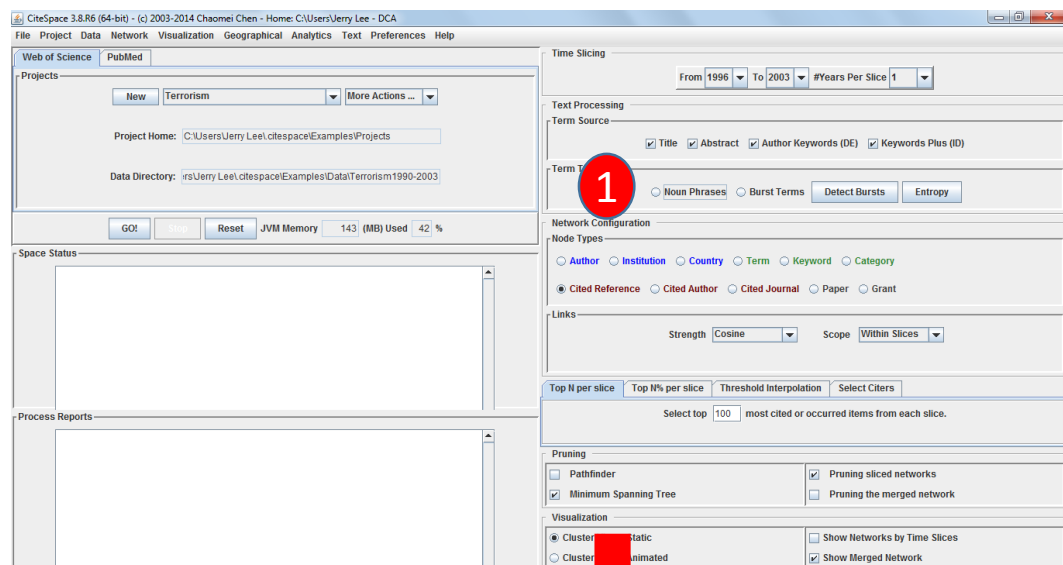
目前关于网络布局的方法（1）基于距离的布局，例如VOS方法、MDS方法、VxOrd以及Kopcsa-Schiebel；（2）基于关系的布局，例如Kamada-Kawai, Fruchterman-Reingold以及Pathfinder networks

因此只要在保证网络各项参数一致的条件下，即使使用CiteSpace会出现一些网络的layout细微的差别，这并不影响分析的结果。

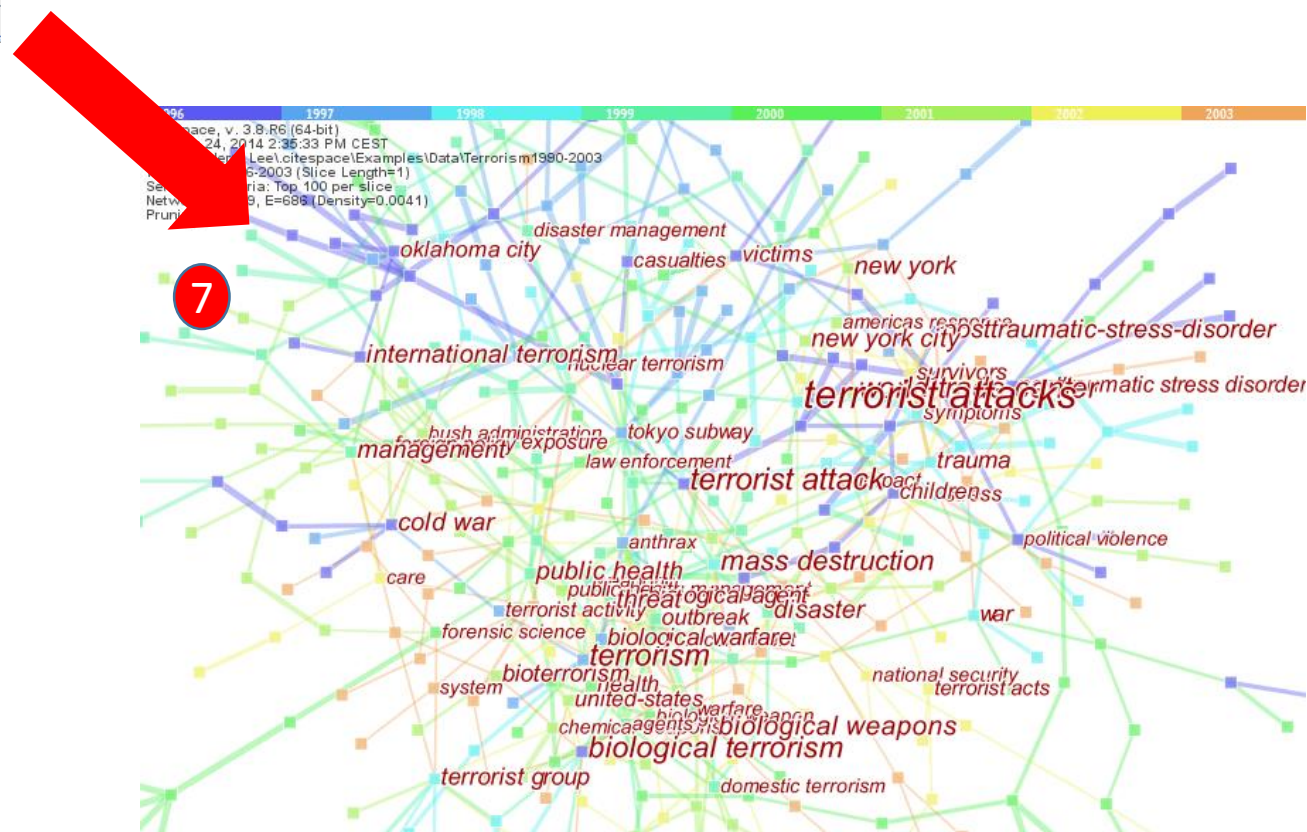
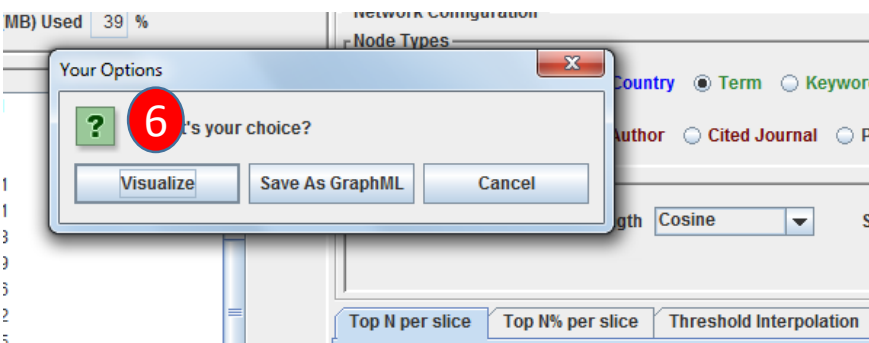


5.名词性术语的提取，为什么提取不出来？

通常的原因可能是操作步骤有误，名词性术语的分析按照下列步骤进行。



5.名词性术语的提取，为什么提取不出来？



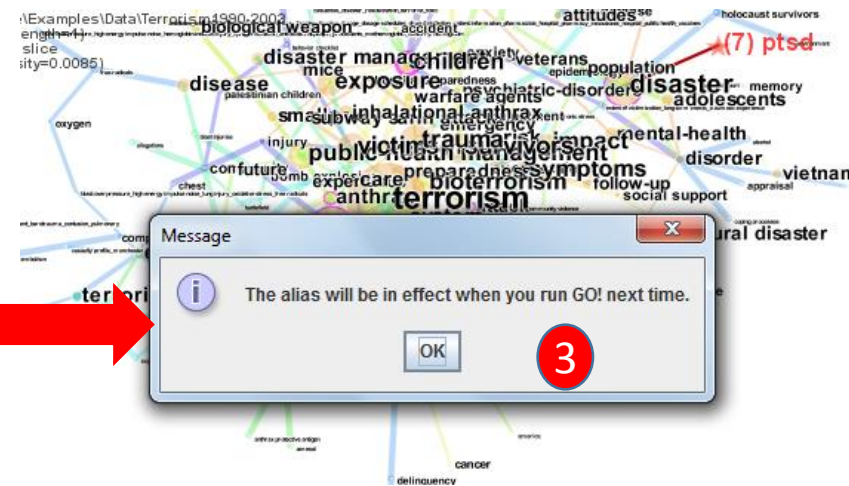
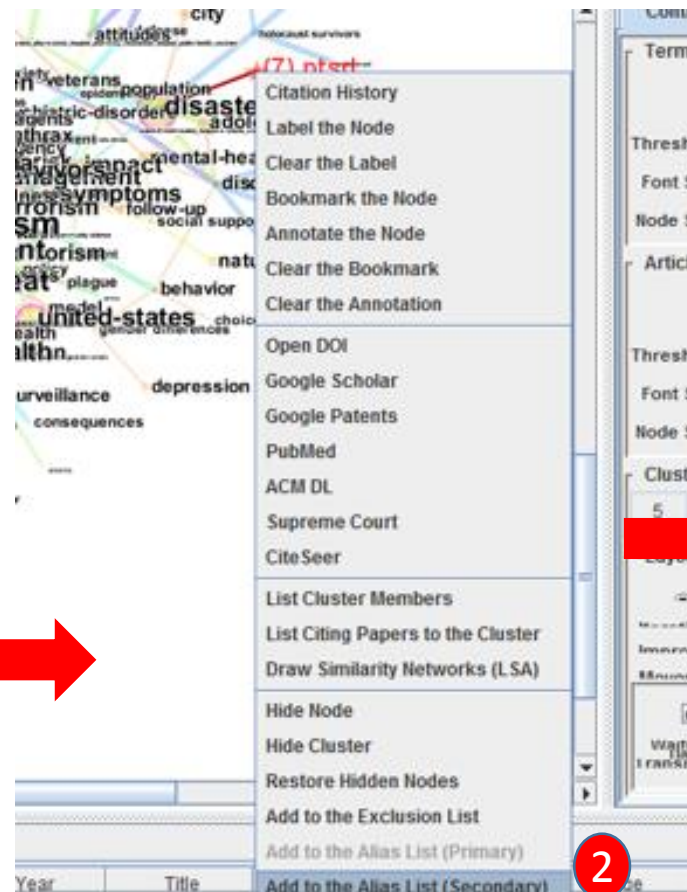
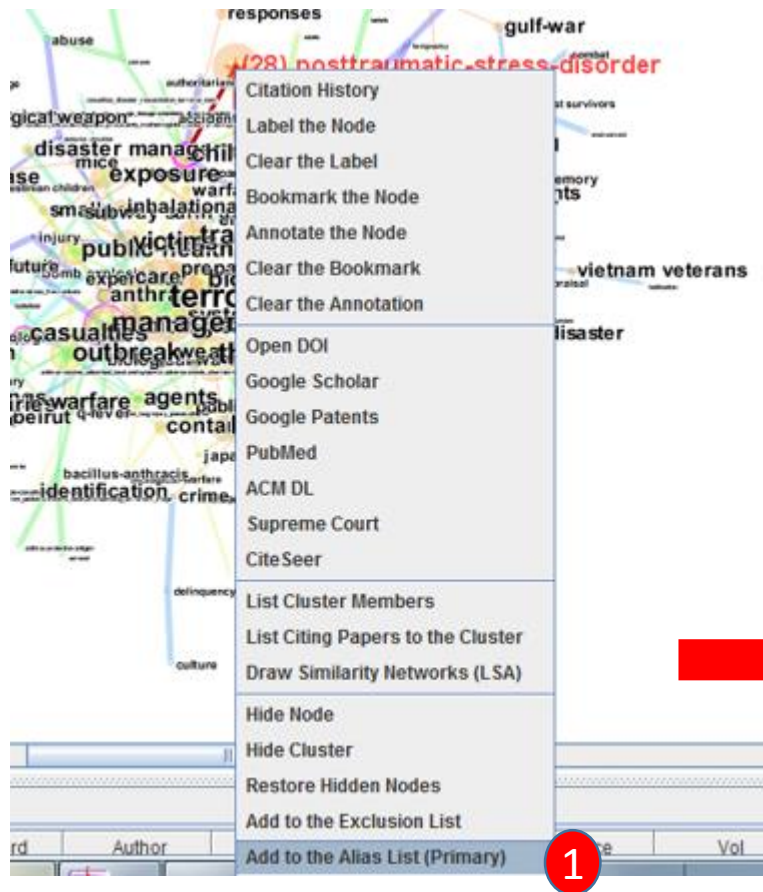
6. 网络中相同含义的词汇如何合并（单复数、英式和美式英语、简称和全称以及同义词合并）？



左图为使用CiteSpace自带的数据分析的恐怖主义研究的关键词共词网络。其中Posttraumatic-stress-disorder与ptsd明显是一个意思。

那么如何处理这样的问题呢？

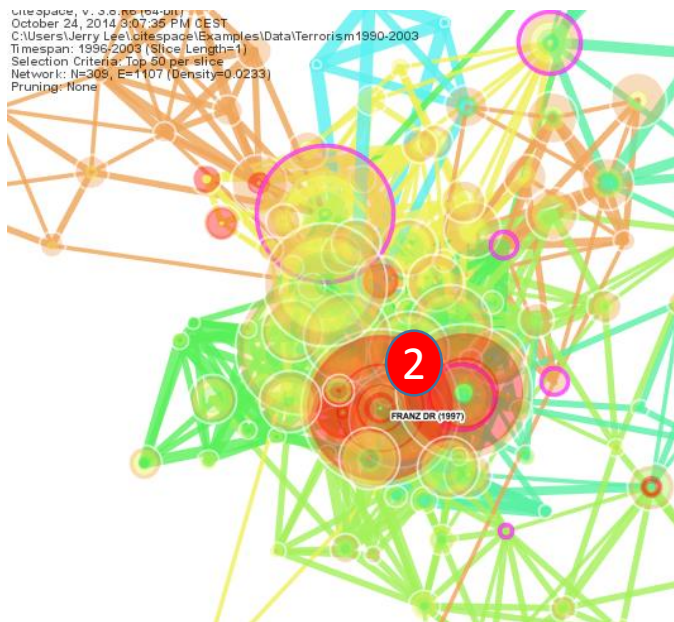
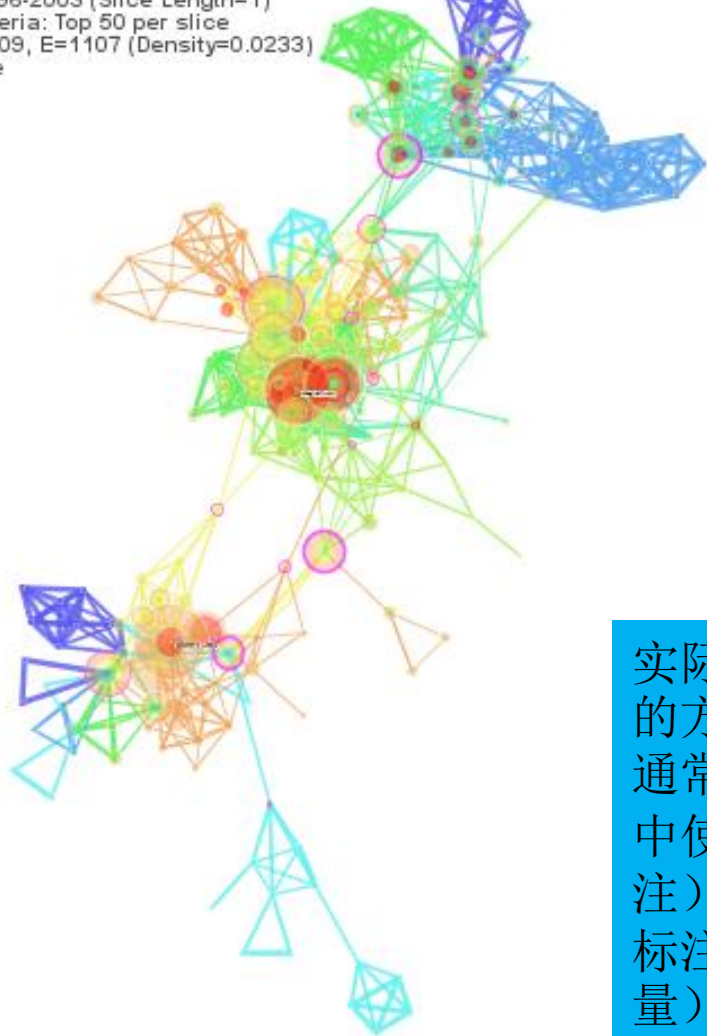
6. 网络中相同含义的词汇如何合并（单复数、英式和美式英语、简称和全称以及同义词合并）？



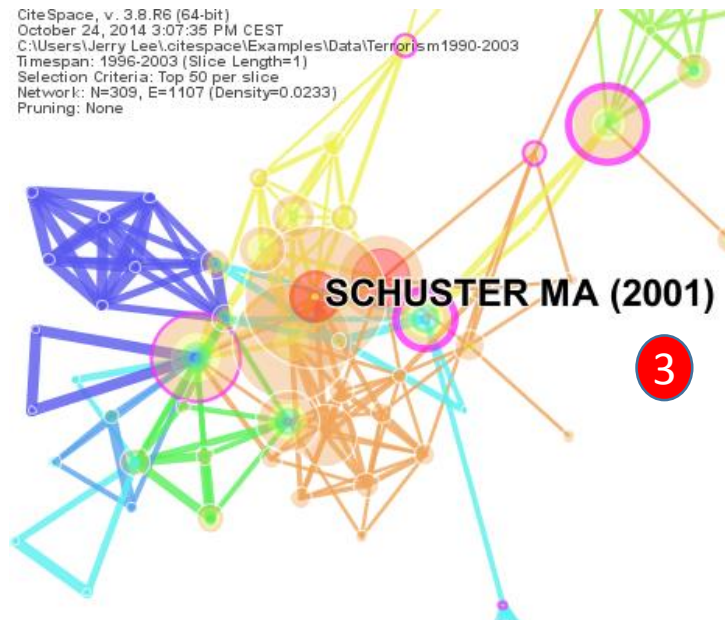
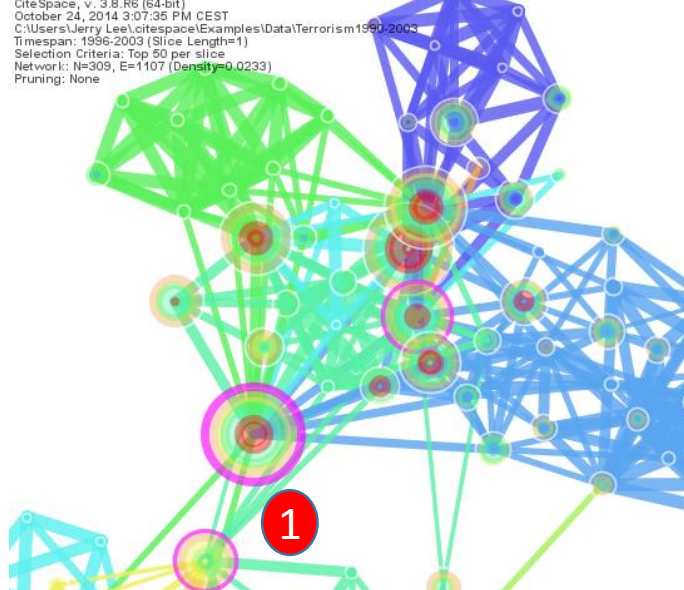
7.在CiteSpace中关键节文献如何确定？

1996 1997 1998 1999 2000 2001 2002 2003

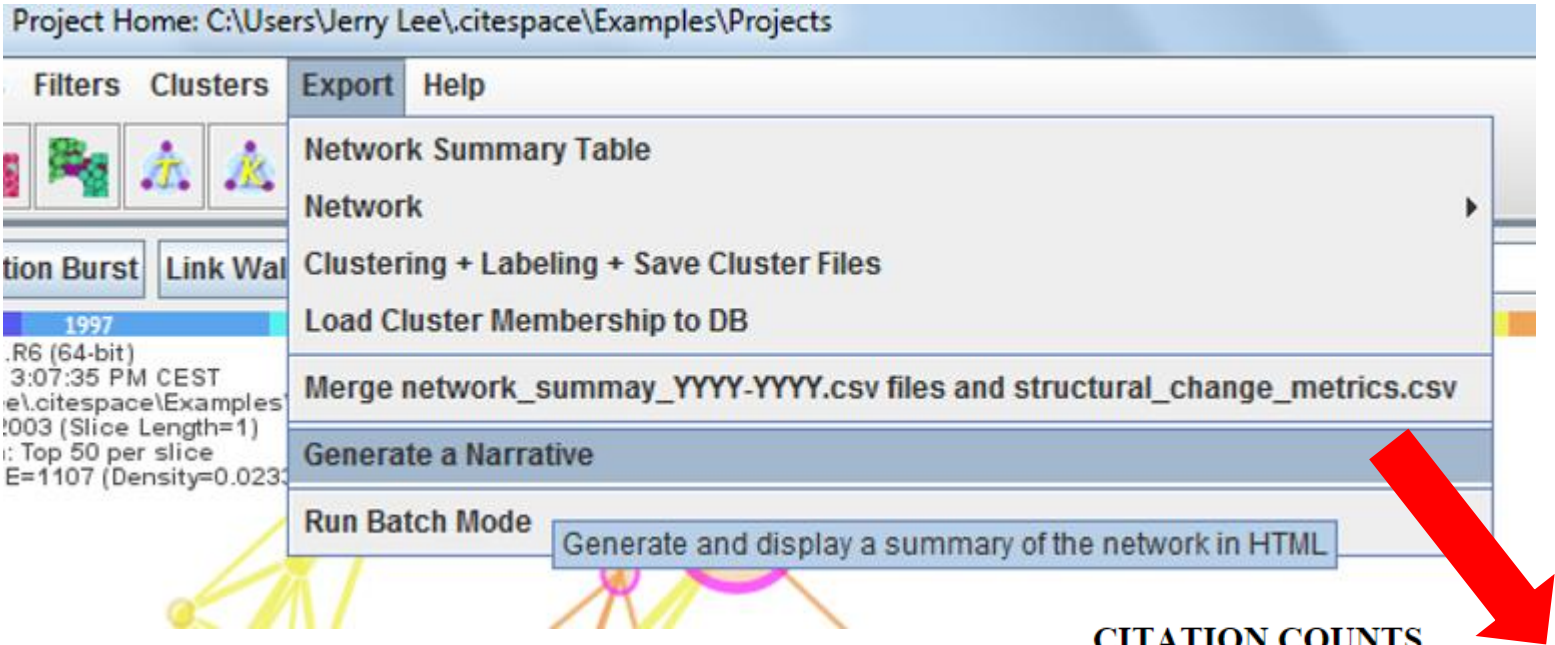
CiteSpace, v. 3.8.R6 (64-bit)
October 24, 2014 3:07:35 PM CEST
C:\Users\Jerry Lee\citespace\Examples\Data\Terrorism1990-2003
Timespan: 1996-2003 (Slice Length=1)
Selection Criteria: Top 50 per slice
Network: N=309, E=1107 (Density=0.0233)
Pruning: None



实际上那个文献是关键文献，通过不同的方法得到的结果也有所不同。通常可以使用中介中心性（在CiteSpace中使用紫色圈对此类文献进行了重点标注）、文献的爆发性（通过红色进行了标注）以及被引频次（通过节点大小衡量）来进行衡量。



7.在CiteSpace中关键文献如何确定？



CITATION COUNTS

The top ranked item by citation counts is Schuster MA (2001) in Cluster #1, with citation counts of 38. The second one is Franz DR (1997) in Cluster #1, with citation counts of 31. The 4th is Henderson DA (1999) in Cluster #0, with citation counts of 30. The 5th is Inglesby TV (1999) in Cluster #0, with citation counts of 29. The 7th is Henderson DA (1999) in Cluster #0, with citation counts of 27. The 8th is *am PSYCH ASS, 1994, DIAGN STAT MAN MENT, V, P in Cluster #0, with citation counts of 23. The 10th is Meselson M (1994) in Cluster #0, with citation counts of 21.

在CiteSpace中，点击Export，Generate a narrative可以直接得到重要文献列表信息

citation counts	references	cluster
38	Schuster MA, 2001, NEW ENGL J MED, V345, P1507	
35	Franz DR, 1997, JAMA-J AM MED ASSOC, V278, P399	
31	Galea S, 2002, NEW ENGL J MED, V346, P982	
30	Henderson DA, 1999, JAMA-J AM MED ASSOC, V281, P2127	
30	Inglesby TV, 1999, JAMA-J AM MED ASSOC, V281, P1735	
29	Torok TJ, 1997, JAMA-J AM MED ASSOC, V278, P389	
27	Henderson DA, 1999, SCIENCE, V283, P1279	
23	*am PSYCH ASS, 1994, DIAGN STAT MAN MENT, V, P	
23	Meselson M, 1994, JAMA-J AM MED ASSOC, V272, P755	

更多信息

- 如何使用CiteSpace的几个示范 [更新] <http://blog.sciencenet.cn/blog-496649-838067.html>
- 如何选取CiteSpace中的各项参数 <http://blog.sciencenet.cn/blog-496649-378974.html>
- 如何在CiteSpace中控制节点的取舍 <http://blog.sciencenet.cn/blog-496649-394695.html>
- 如何在CiteSpace中提取名词短语<http://blog.sciencenet.cn/blog-496649-675216.html>
- 如何增强CiteSpace处理数据的多少和快慢<http://blog.sciencenet.cn/blog-496649-455758.html>
- 如何用CiteSpace分析专利概况<http://blog.sciencenet.cn/blog-496649-433660.html>
- 关于CSSCI 的数据转换问题 （更新） <http://blog.sciencenet.cn/blog-496649-427780.html>
- 如何解决 “Could not create the Java virtual machine”的问题<http://blog.sciencenet.cn/blog-496649-396909.html>
- 如何用CiteSpace整理数据<http://blog.sciencenet.cn/blog-496649-386310.html>
- 如何解读CiteSpace产生的图谱<http://blog.sciencenet.cn/blog-496649-379517.html>
- CiteSpace中的Burst Detection<http://blog.sciencenet.cn/blog-496649-566289.html>
- Chen, C. (2014) The CiteSpace Manual. <http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf>