电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

《综合课程设计》报告

2019-2020-2

BACHELOR THESIS

论文 题目 基于科研社交媒体数据的

科技舆情形成及政策影响研究

学 院 公共管理学院

专 业 信息管理与信息系统

时 间 **2020/6/28**

小组成员及分工表

学号	姓名	主要工作
2017120101004	关仕浩	数据收集与数据预处理
2017120101029	魏玮浩	数据收集与数据预处理
2017120101010	黄宇星罗	**LDA** 关键词提炼与报告撰写
2017120101023	程天宇	共词分析与结果结论分析

摘　要

本课题“基于科研社交媒体数据的科技舆情形成及政策影响研究”，关键在于**“科研社交媒体数据”、“科技舆情形成”、“政策影响”**三个方面。实验选取了2018 年基因编辑婴儿事件作为采集数据的来源。针对基因编辑婴儿舆情六个时期的微博评论进行 LDA 主题建模，利用六个时期舆情评论进行共词分析。最终与政府、相关部门回应与政策进行比较，展现出基因编辑婴儿舆情变化与政策影响的结果。结果表明该方法能够有效识别基因编辑婴儿舆情主题。

关键词：数据挖掘；文本分析；科研社交媒体数据；科技舆情形成；政策影响；LDA 主题建模；共词分析

ABSTRACT

ABSTRACT

The key of this project "Research on the formation and policy impact of scientific and technological public opinion based on scientific research social media data" lies in the three aspects of "scientific research social media data", "scientific and technological public opinion formation" and "policy impact". In the experiment, we selected the event of gene editing baby in 2018 as the source of data collection. In this paper, LDA topic modeling is carried out for microblog comments in six periods of gene editing baby public opinion, and co word analysis is conducted by using the six periods of public opinion comments. Finally, compared with the response and policy of the government and relevant departments, it shows the results of the change of public opinion and the influence of policy. The results show that the method can effectively identify the theme of public opinion of gene editing infants.

Keywords: data mining; text analysis; scientific research social media data; formation of scientific and technological public opinion; policy influence; LDA topic modeling; CO word analysis

目录

摘　要.....I
ABSTRACT.....II
目录.....III
第一章 数据分析综合概况及选题设计.....1
1.1 选题及选题背景.....1
1.2 数据来源、数据获取方案.....1
1.3 数据结构.....2
1.4 理论模型.....2
1.5 实证模型.....3
第二章 网络数据采集与存储.....5
2.1 数据爬取.....5
2.2 数据分类存储.....7
第三章 实验数据的预处理.....9
3.1 数据预处理.....9
3.2 数据分词.....9
第四章 数据的统计分析与可视化.....11
4.1 数据统计分析.....11
4.2 数据的 LDA 主题提炼与可视化.....14
第五章 数据的共词分析.....23
5.1 数据的共词分析模型.....23
第六章 综合课程设计总结.....33
6.1 研究结果联系回应.....33
6.2 综合课程设计的不足之处与改进方法.....33
参考文献.....35

第一章 数据分析综合概况及选题设计

1.1 选题及选题背景

本小组选取了 2018 年基因编辑婴儿事件作为聚焦题目

原因：此热点事件能够激发公众讨论热情，从而引起舆论变化与可能引起的政策影响或得到回应

选题背景：近几年的热点事件、比如基因编辑、假疫苗、5G、学术不端，这里面有些内容引起了政府机构的重视，比如学术不端——翟天临事件后教育部发布了严厉的研究生学位论文评估，但是同一个事件舆情会有不同主题，有的会被回应有的可能就没有引起回应，就是没有产生政策影响。

所以研究的目的就是探寻舆情形成如何影响政府决策是否引起政府回应

关于科技型社会事件的相关网络舆情，对于实际的政策和回应形成有着较大的影响，其具体表现在紧跟事件发展过程，并存在全方位考量对事件本身处理、对个人本身、对社会影响的情况。

同时，科技类社会事件发生后的社会舆情，更多注重的是其对生活、对社会、对法律道德观念的影响，谈及技术本身，也往往注目于其必要性、合法性、可行性（不包括技术上困难的考量）。

研究此类数据的发展变化和其本身特点，对更好地与民情民意互联互通，充分体现人民民主，实现司法的公平正义和鉴定公民本身的科学文化素养等具有重大意义。

1.2 数据来源、数据获取方案

数据来源：

1.科研社交媒体数据：

所谓科研社交媒体，指的是网络中的节点大多由从事科研工作和进行学习实践的科研人员和学生组成的社交网络平台，发布的帖子、话题、讨论串等常常与科学研究中的问题和看法、对科技类热门社会事件的观点、关于今后个人发展方向的询问和其他内容有关。是大部分科研工作者和从事学习研究的学生等可能会大量地进行学习讨论时使用的社交媒体平台。平台的内容往往比一般 SNS 更具有专业性、建设性、科学性。但是这一优势同时也是一种巨大的劣势。基于这类内容在平台中的占比较大，参与讨论的人数可能并没有那么多，一个帖子往往只有 10 到 20 回复的数量，且发生科技类热门社会事件时，话题可能较为分散，若有重复性的回答其占总体回复的比例也较大，统计出的词频说服力较差。

故本小组考虑使用**微博平台上的科技类热门社会事件下方的评论**作为数据

采集对象，数据的数量和质量相对更有保证。

例如，本小组选取了 **2018 年基因编辑婴儿事件**作为采集数据的来源。

2.政策

政策，此处应视为**官方媒体、政府机关部门等发布的政策、回应说法、对现有政策法规的新解释**等等，方面比较多样。描述政策可以使用关键词提取后的关联分析或依存句法等，提取出政策的观点，并与现有舆情进行对比，讨论舆论是否在某种程度上影响了政策的形成。该影响同样可以利用比对舆论和政策声明的关键词频繁项集的重合率进行描述。

本小组收集了南方科技大学表态、国家科技部回应、国家卫建委回应、深圳市南山区人民法院一审公开宣判作为政策影响。

1.3 数据结构

利用 python 与八爪鱼抓取工具能得到以下数据，其格式为：

博文网址（微博网页地址）
发帖子人（发帖用户微博名）
博文发布时间（微博发布的时间）
来自（weibo.com）
博文内容（微博正文）
转发数（微博转发数量）
评论数（评论数量）
点赞数（点赞数量）
博文图片（配图）
博文视频（视频）
评论者（评论用户微博名）
评论内容（评论微博正文）
评论时间（用户评论的时间）
评论点赞（该评论获得的点赞数量）
二级评论人（二级评论用户微博名）
二级评论内容（二级评论微博正文）
二级评论时间（二级用户评论的时间）
二级评论点赞（二级评论获得的点赞数量）

1.4 理论模型

关于本课题，涉及的研究方向有下列几种：

1. 网络舆情对公共政策影响的研究

网络舆情正在逐渐演变为公民影响政治生活的重要路径，以约翰·W·金登为代表的众多学者均指出，突发公共事件是形成舆论，并引起官员注意的最重要方式之一，它常常能促使社会问题显性化，激起广泛的网络舆论讨论，并推动问题进入政策议程。

关于网络舆情通过什么方式影响，影响的因素有哪些，影响的结果会是什么，已经有多种方法理论存在。如有采用上海交通大学舆情研究实验室和信息安全计

算机科学技术研究团队合作研发的舆情数据实时抓取集成系统，探讨网络舆情分析传播中推进的公共政策的基本特点和网络舆情传播与公共政策推进的相互关系[4]，或者基于扎根理论研究中国网络舆情对影响政策出台的关键影响因素[5]，还有基于“间断—均衡理论”视角的新媒体环境下网络舆情对政策间断式变迁的影响研究[6]等。

2. 网络舆情的文本挖掘

在网络舆情已成为社会情报的一种重要表现形式的大背景下，挖掘技术为网上大量以非结构数据形式出现的舆情信息分析提供了方法和技术支持。

文本挖掘主要的技术包括：特征提取、文本分类、文本聚类、关联分析、文本总结和趋势预测。

1.5 实证模型

1.基于 LDA 的六个时期事件主题提炼

LDA（Latent Dirichlet Allocation，潜在狄利克雷分布）[2]是一种典型的文档主题生成模型[2]。该模型算法通过概率主题分布对科技文献进行文本语义分析，从中抽取出有价值的潜在主题信息，弥补了引文分析方法和词汇分析方法的不足，可实现内容主题层面的文献数据分析和影响力评价。

LDA 模型框架：

当前时期评论-去除停用词-去除高低频词-词性归并-LDA 模型

LDA 数据预处理[3]：

主要包括文档准备、去除停用词和高低频词、词形归并。

1)文档准备。将六个时期一级评论作为文献内容字段。各时期文献生成的内容文本，作为 LDA 主题模型中文档表达。

2）去除停用词和高低频词。

将各时期文档进行分词，根据停用词表去停用词，过滤掉文本中的标点符号、语气助词、副词、介词等没有实际含义的词汇。对过滤后的词汇进行词频计算，剔除出现频次小于 2 或者大于 1%的词汇。 （停用词表采用的是四川机器学习研究停用词表）

3）词形归并。词形归并主要涉及标点符号去除、语气词删去、相同名词合并等。

2.舆情共词分析

“共现”指文献的特征项描述的信息共同出现的现象，这里的特征项包括文献的外部和内部特征，如题名、作者、关键词、机构等。而“共现分析”是对共现现象的定量研究，以揭示信息的内容关联和特征项所隐含的知识。

共词分析的基本原理是通过统计文献集中词汇对或名词短语的共现情况，来反映关键词之间的关联强度，进而确定这些词所代表的学科或领域的研究热点、组成与范式，横向和纵向分析学科领域的发展过程和结构演化。该方法前提假设

是：词汇对在同一篇文献中出现的次数越多，则代表这两个主题的关系越紧密。由此，统计一组文献的主题词两两之间在同一篇文献出现的频率，便可形成一个由这些词对关联所组成的共词网络，网络内节点之间的远近便可反映主题内容的亲疏关系。虽然常用于图书情报学或是信息分析咨询的内容分析部分，也可以很好地运用于网络舆情的相关研究[7]。

本项目使用此方法进行自然语言处理的考量是，虽然 LDA 主题可以很好地帮我们概括网民们对于事件的关注点是什么，但却没法描述他们希望事件“如何处理”、“有怎样的结果”这一情况，而考量舆情对于政策形成的影响，必须要联系舆情关于如何回应处理的看法，因此共词分析这一方法的使用是必然的。

本课题主要使用了特征提取、文本分类和关联分析等方面的技术[1]

第二章 网络数据采集与存储

2.1 数据爬取

1.梳理在线科技舆情的主题事件形成的过程（基因编辑婴儿事件）：

1).新闻撤稿

2018 年 11 月 27 日消息，舆论引爆自人民网深圳频道一篇名为《世界首例免疫艾滋病的基因编辑婴儿在中国诞生》

2).删除视频

2018 年 11 月 25 日、26 日两天，贺建奎却以南方科技大学副教授身份分别在国内和海外视频平台优酷与 YouTube 上各上传了 5 段详细讲述该试验的解说视频。

3).公开数据

2018 年 11 月 26 日 23 时 30 分左右，贺建奎团队负责媒体事宜的工作人员陈远林表示，贺建奎将于本周三在香港会议上公开该项目数据。

4).“查封”办公室

2018 年 11 月 27 日晚间，在南方科技大学贺建奎所属办公室，发现办公室门外的介绍内容已摘除，并贴上印有“请勿入内，后果自负”南方科技大学盖章的提示条

5).现身港大

2018 年 11 月 28 日 12 时 50 分，贺建奎现身香港大学李兆

基会议中心。在讲台上站定后，其对自己的研究致歉，并在现场披露“露露娜娜已经健康出生”，“结果符合预期”。

6).暂停科研活动

2018 年 11 月 29 日，国家卫生健康委员会、科学技术部、中国科学技术协会等三部门负责人接受新华社记者采访表示：此次事件性质极其恶劣，已要求有关单位暂停相关人员的科研活动，对违法违规行为坚决予以查处。

7).初步查明

2019 年 1 月 21 日，从广东省“基因编辑婴儿事件”调查组获悉，现已初步查明，该事件系南方科技大学副教授贺建奎为追逐个人名利，自筹资金，蓄意逃避监管，私自组织有关人员，实施国家明令禁止的以生殖为目的的人类胚胎基因编辑活动。

8).案件宣判

2019 年 12 月 30 日，“基因编辑婴儿”案在深圳市南山区人民法院一审公开宣判。

根据事件脉络我们选取了六个时间点的微博热搜话题**[8]**进行数据采集

20181126 世界首例免疫艾滋病的基因编辑婴儿在中国诞生-评论

20181126 基因编辑婴儿 回应视频 相关评论

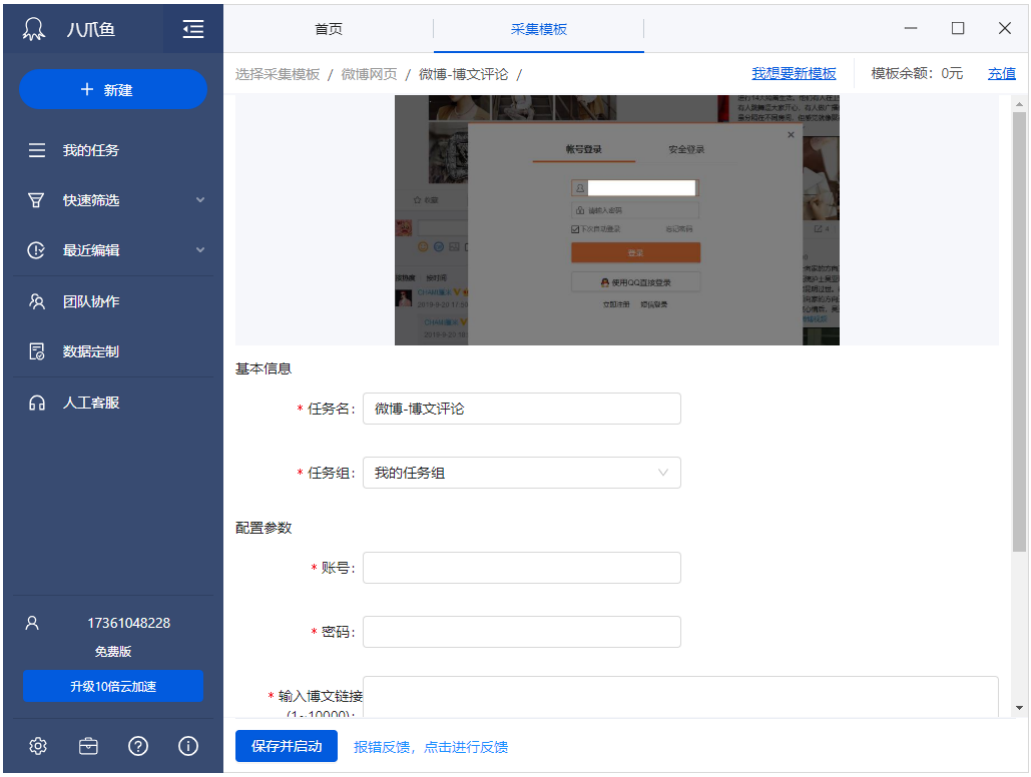
20181128 现身港大回应

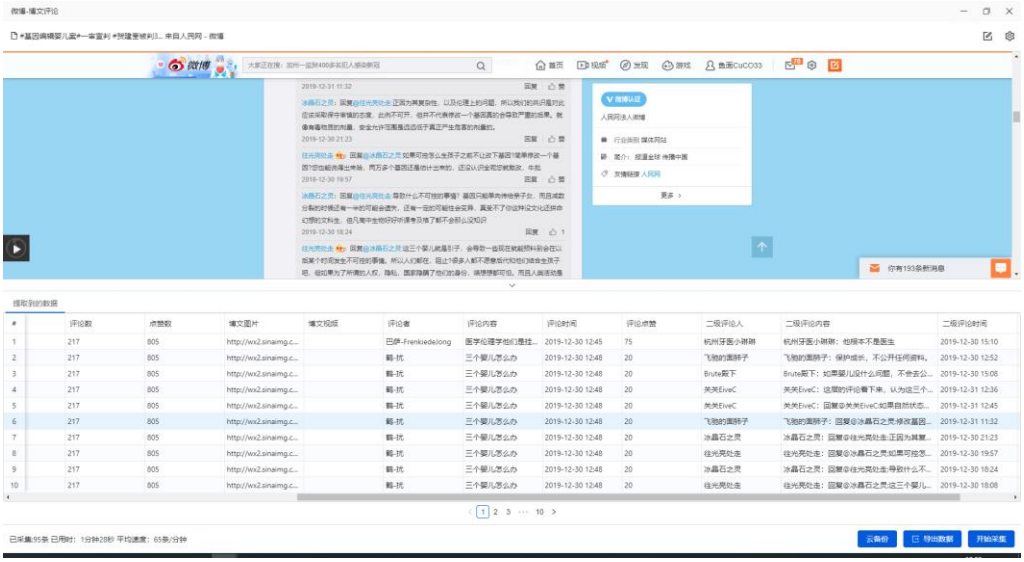
20181129 国家卫健委科技部回应

20191230 #人民网#基因编辑婴儿案#一审宣判

20191230 追究刑事责任

2.八爪鱼爬取界面：对任务名进行设置并且确定最大数据挖掘参数

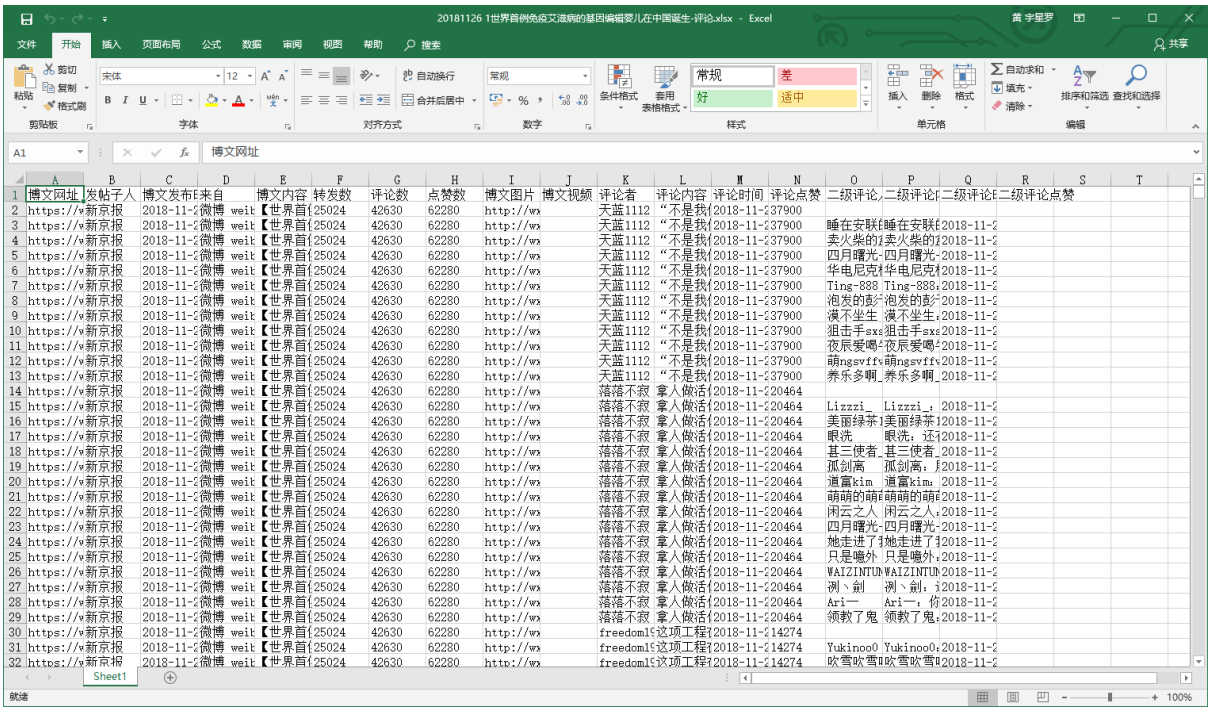




3.数据存放为 excel 表格文件

20181126 1世界首例免疫艾滋病的基因...	2020/5/14 18:25	Microsoft Excel ...	3,644 KB
20181126 2基因编辑婴儿 回应视频 相...	2020/5/14 16:17	Microsoft Excel ...	3,332 KB
20181128 现身港大回应.xlsx	2020/5/14 17:44	Microsoft Excel ...	3,924 KB
20181129 国家卫健委科技部回应.xlsx	2020/5/14 17:43	Microsoft Excel ...	3,650 KB
20191230 #人民网#基因编辑婴儿案#一...	2020/5/14 17:47	Microsoft Excel ...	23 KB
20191230 追究刑事责任.xlsx	2020/5/14 17:45	Microsoft Excel ...	361 KB

原始数据示例：



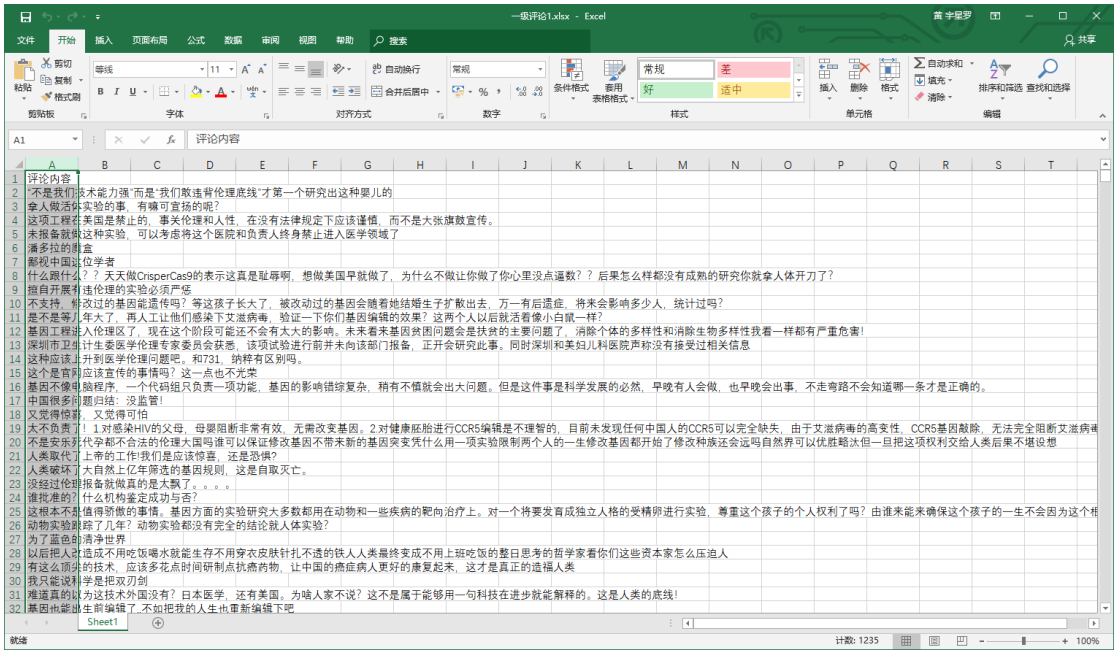
博文网址	发帖子人	博文发布时	来自	博文内容	转发数	评论数	点赞数	博文图片	博文视频	评论者	评论内容	评论时间	评论点赞	二级评论	二级评论	二级评论	二级评论点赞
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900					
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		睡在安联	睡在安联	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		卖火柴的	卖火柴的	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		四月曙光-	四月曙光-	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		华电尼克	华电尼克	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		Ting-888	Ting-888:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		泡发的彭	泡发的彭	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		漠不坐生	漠不坐生:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		狙击手sxs	狙击手sxs	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		夜辰爱喝	夜辰爱喝	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		萌ngsvffv	萌ngsvffv	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		天蓝1112	“不是我(	2018-11-237900		养乐多啊_	养乐多啊_	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464					
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		Lizzzi_	Lizzzi_:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		美丽绿茶	美丽绿茶	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		眼洗	眼洗；还	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		甚三使者_	甚三使者_	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		孤剑高	孤剑高；	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		道富kim	道富kim:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		萌萌的萌	萌萌的萌	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		闲云之人	闲云之人:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		四月曙光-	四月曙光-	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		她走进了	她走进了	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		只是噜外	只是噜外:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		WAIZINTUN	WAIZINTUN	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		洲丶剑	洲丶剑；	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		Ari一	Ari一：你	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		落落不寂	拿人做活(	2018-11-220464		领教了鬼	领教了鬼:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		freedom1	这项工程	2018-11-214274					
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		freedom1	这项工程	2018-11-214274		Yukinoo0	Yukinoo0:	2018-11-2	
https://w	新京报	2018-11-2	微博 weit	【世界首(	25024	42630	62280	http://ws		freedom1	这项工程	2018-11-214274		吹雪吹雪	吹雪吹雪	2018-11-2	

2.2 数据分类存储

将微博评论数据中的一级评论与二级评论分别提取出来进行存储

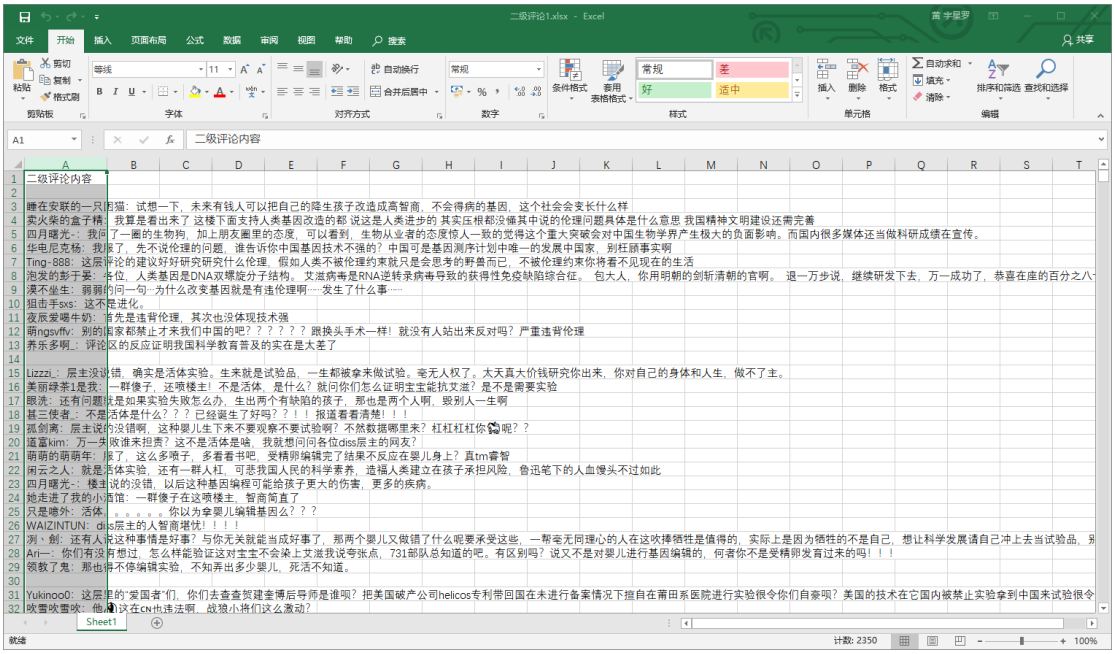
二级评论1.xlsx	2020/6/17 19:56	Microsoft Excel ...	57 KB
二级评论2.xlsx	2020/6/17 19:58	Microsoft Excel ...	42 KB
二级评论3.xlsx	2020/6/17 19:58	Microsoft Excel ...	44 KB
二级评论4.xlsx	2020/6/17 19:59	Microsoft Excel ...	14 KB
二级评论5.xlsx	2020/6/17 19:59	Microsoft Excel ...	12 KB
二级评论6.xlsx	2020/6/17 20:00	Microsoft Excel ...	60 KB
一级评论1.xlsx	2020/6/17 21:19	Microsoft Excel ...	63 KB
一级评论2.xlsx	2020/6/17 21:19	Microsoft Excel ...	98 KB
一级评论3.xlsx	2020/6/17 21:20	Microsoft Excel ...	78 KB
一级评论4.xlsx	2020/6/17 21:20	Microsoft Excel ...	17 KB
一级评论5.xlsx	2020/6/17 21:20	Microsoft Excel ...	13 KB
一级评论6.xlsx	2020/6/17 21:21	Microsoft Excel ...	112 KB

一级评论示例：



评论内容
"不是我们技术能力强"而是"我们敢违背伦理底线"才第一个研究出这种婴儿的
拿人做活体实验的事，有嘛可宣扬的呢？
这项工程在美国是禁止的，事关伦理和人性，在没有法律规定下应该谨慎，而不是大张旗鼓宣传。
未报备就做这种实验，可以考虑将这个医院和负责人终身禁止进入医学领域了
潘多拉的盒
鄙视中国这位学者
什么跟什么？？天天做CrisperCas9的表示这真是耻辱啊，想做美国早就做了，为什么不做让你做了你心里没点逼数？？后果怎么样都没有成熟的研究你就拿人体开刀了？
独自开展有违伦理的实验必须严惩
不支持，修改过的基因能遗传吗？等这孩子长大了，被改动过的基因会随着她结婚生子扩散出去，万一有后遗症，将来会影响多少人，统计过吗？
是不是等几年大了，再人工让他们感染下艾滋病毒，验证一下你们基因编辑的效果？这两个人以后就活看像小白鼠一样？
基因工程进入伦理区了，现在这个阶段可能还不会有太大的影响。未来看来基因贫困问题会是扶贫的主要问题了，消除个体的多样性和消除生物多样性我看一样都有严重危害！
深圳市卫生计生委医学伦理专家委员会获悉，该项试验进行前并未向该部门报备，正开会研究此事。同时深圳和美妇儿科医院声称没有接受过相关信息
这种应该上升到医学伦理问题吧。和731，纳粹有区别吗。
这个是官网应该宣传的事情吗？这一点也不荣荣
基因不像电脑程序，一个代码组只负责一项功能，基因的影响错综复杂，稍有不慎就会出大问题。但是这件事是科学发展的必然，早晚有人会做，也早晚会出事，不走弯路不会知道哪一条才是正确的。
中国很多问题归结：没监管！
又觉得惊喜，又觉得可怕
太不负责了！1.对感染HIV的父母，母婴阻断非常有效，无需改变基因。2.对健康胚胎进行CCR5编辑是不理智的，目前未发现任何中国人的CCR5可以完全缺失，由于艾滋病毒的高变性，CCR5基因敲除，无法完全阻断艾滋病毒
不是安乐死代孕都不合法的伦理大国吗谁可以保证修改基因不带来新的基因突变凭什么用一项实验限制两个人的一生修改基因都开始了修改种族还会远吗自然界可以优胜略汰但一旦把这项权利交给人类后果不堪设想
人类取代了上帝的工作!我们是应该惊喜，还是恐惧?
人类破坏了大自然上亿年筛选的基因规则，这是自取灭亡。
没经过伦理报备就做真的是大飘了。。。。。
谁批准的？什么机构鉴定成功与否？
这根本不是值得骄傲的事情。基因方面的实验研究大多数都用在动物和一些疾病的靶向治疗上。对一个将要发育成独立人格的受精卵进行实验，尊重这个孩子的个人权利了吗？由谁来能来确保这个孩子的一生不会因为这个
动物实验跑跨了几年？动物实验都没有完全的结论就人体实验？
为了蓝色的清净世界
以后把人改造成不用吃饭喝水就能生存不用穿衣皮肤针扎不透的铁人人类最终变成不用上班吃饭的整日思考的哲学家看你们这些资本家怎么压迫人
有这么顶尖的技术，应该多花点时间研制点抗癌药物，让中国的癌症病人更好的康复起来，这才是真正的造福人类
我只能说科学是把双刃剑
难道真的以为这技术外国没有？日本医学，还有美国。为啥人家不说？这不是属于能够用一句科技在进步就能解释的。这是人类的底线！
基因也能出生前编辑了,不如把我的人生也重新编辑下吧

二级评论示例：

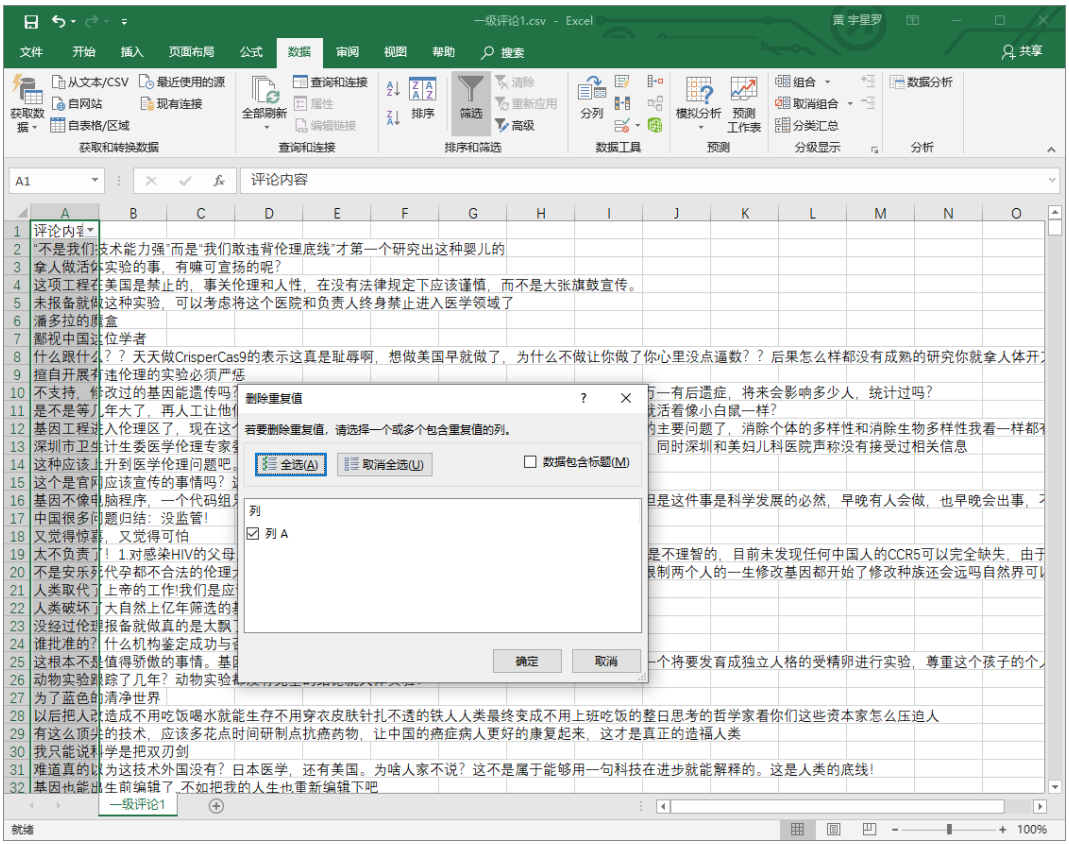


二级评论内容
睡在安联的一只胡猫：试想一下，未来有钱人可以把自己的降生孩子改造成高智商，不会得病的基因，这个社会会变长什么样
卖火柴的盒子精：我算是看出来了 这楼下面支持人类基因改造的都 说这是人类进步的 其实压根都没懂其中说的伦理问题具体是什么意思 我国精神文明建设还需完善
四月曙光-：我问了一圈的生物狗，加上朋友圈里的态度，可以看到，生物从业者的态度惊人一致的觉得这个重大突破会对中国生物学界产生极大的负面影响。而国内很多媒体还当做科研成绩在宣传。
华电尼克杨：我服了，先不说伦理的问题，谁告诉你中国基因技术不强的？中国可是基因测序计划中唯一的发展中国家，别枉顾事实啊
Ting-888：这层评论的建议好好研究研究什么伦理，假如人类不被伦理约束就只是会思考的野兽而已，不被伦理约束你将看不见现在的生活
泡发的彭于晏：各位，人类基因是DNA双螺旋分子结构。艾滋病毒是RNA逆转录病毒导致的获得性免疫缺陷综合征。 包大人，你用明朝的剑斩清朝的官啊。 退一万步说，继续研发下去，万一成功了，恭喜在座的百分之八
漠不坐生：弱弱的问一句···为什么改变基因就是有违伦理啊····发生了什么事····
狙击手sxs：这不是进化。
夜辰爱喝牛奶：首先是违背伦理，其次也没体现技术强
萌ngsvffv：别的国家都禁止才来我们中国的吧？？？？？？跟换头手术一样！就没有人站出来反对吗？严重违背伦理
养乐多啊_：评论区的反应证明我国科学教育普及的实在是太差了
Lizzzi_：层主没说错，确实是活体实验。生来就是试验品，一生都被拿来做试验。毫无人权了。太天真大价钱研究你出来，你对自己的身体和人生，做不了主。
美丽绿茶1是我：一群傻子，还喷楼主！不是活体，是什么？就问你们怎么证明宝宝能抗艾滋？是不是需要实验
眼洗：还有问题就是如果实验失败怎么办，生出两个有缺陷的孩子，那也是两个人啊，毁别人一生啊
甚三使者_：不是活体是什么？？？已经诞生了好吗？！！！报道看看清楚！！！
孤剑离：层主说的没错啊，这种婴儿生下来不要观察不要试验啊？不然数据哪里来？杠杠杠杠你[illegible]呢？？
道富kim：万一失败谁来担责？这不是活体是啥，我就想问问各位diss层主的网友？
萌萌的萌萌年：服了，这么多喷子，多看看书吧，受精卵编辑完了结果不反应在婴儿身上？真tm霉智
闲云之人：就是活体实验，还有一群人杠，可悲我国人民的科学素养，造福人类建立在孩子承担风险，鲁迅笔下的人血馒头不过如此
四月曙光-：楼主说的没错，以后这种基因编程可能给孩子更大的伤害，更多的疾病。
她走进了我的小酒馆：一群傻子在这喷楼主，智商简直了
只是嗜外：活体。。。。。。你以为拿婴儿编辑基因么？？？
WAIZINTUN：diss层主的人智商堪忧！！！！
冽、剑：还有人说这种事情是好事？与你无关就能当成好事了，那两个婴儿又做错了什么呢要承受这些，一帮毫无同理心的人在这吹捧牺牲是值得的，实际上是因为牺牲的不是自己，想让科学发展请自己冲上去当试验品，别
Ari—：你们有没有想过，怎么样能验证这对宝宝不会染上艾滋我说夸张点，731部队总知道的吧。有区别吗？说又不是对婴儿进行基因编辑的，何者你不是受精卵发育过来的吗！！！
领教了鬼：那也得不停编辑实验，不知弄出多少婴儿，死活不知道。
Yukinoo0：这层里的"爱国者"们，你们去查查贺建奎博后导师是谁呗？把美国破产公司helicos专利带回国在未进行备案情况下擅自在莆田系医院进行实验很令你们自豪呗？美国的技术在它国内被禁止实验拿到中国来试验很令
吹雪吹雪吹：他们这种建议在cn也违法啊，战狼小将们这么激动？

第三章 实验数据的预处理

3.1 数据预处理

经过筛选查看之后我们决定使用一级评论作为我们的实验数据
对一级评论数据进行去重并且删去空白值
去重示例：



3.2 数据分词并统计词频

利用 python jieba 分词包进行一级评论分词
分词示例：

In [5]:
```
# 分词函数
def chinese_word_cut(mytext):
    import jieba.posseg as psg
    checkarr = ['n']
    returnmsg = ''.join([(x.word) for x in psg.cut(mytext) if (x.flag in checkarr)])
    return returnmsg
```

In [6]:
```
 # 分词结果
df = df.content.apply(chinese_word_cut)
ff = df.apply(movestopwords)
ff = ff.dropna()

print(ff) #预览
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\ADMINI~1\AppData\Local\Temp\jieba.cache
Loading model cost 0.728 seconds.
Prefix dict has been built succesfully.
```

```
0          技术能力底线婴儿
1              人活体事
2        工程事关人性法律规定
3         医院负责人医学领域
4               魔盒
           ...
1228
1229    后果魔盒社有钱人世界人生
1230
1231           狗屁制度
1232
Name: content, Length: 1233, dtype: object
```

词频统计示例：

In [7]:
```
word_counts = collections.Counter(ff) # 对分词做词频统计

word_counts_top10 = word_counts.most_common(11) # 获取前11最高频的词

word_counts_top10 = word_counts_top10[1:10]
print (word_counts_top10)
```

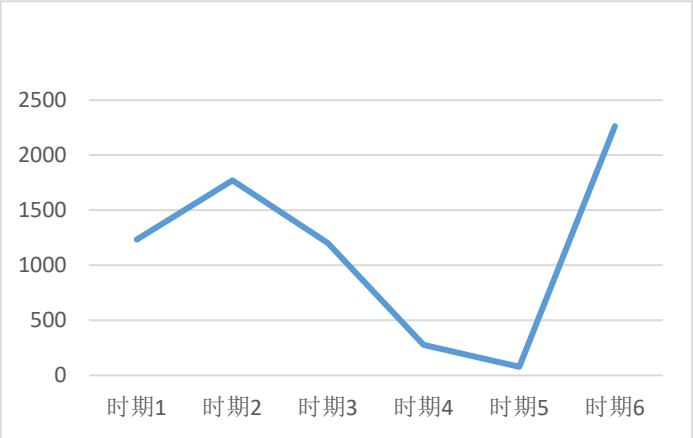
```
[('人', 18), ('人类', 13), ('科学', 9), ('底线', 9), ('魔盒', 8), ('孩子', 7), ('评论', 6), ('事', 6), ('科学家', 6)]
```

第四章 数据的统计分析与可视化

4.1 数据统计分析

1. 六个时期的一级评论数量

时期 1	1233
时期 2	1773
时期 3	1202
时期 4	276
时期 5	78
时期 6	2264

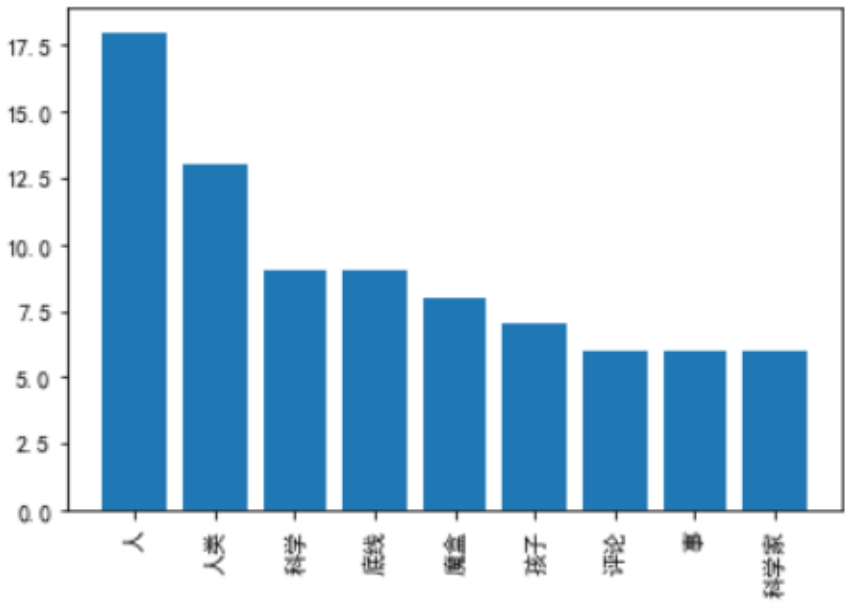


图表 4.1.1

2. 六个时期分词后词频统计

时期 1

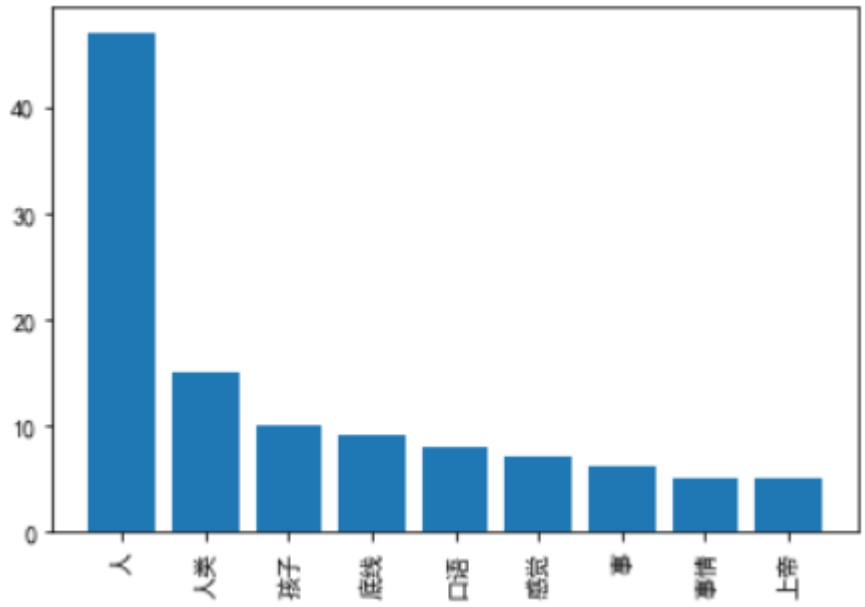
[('人', 18), ('人类', 13), ('科学', 9), ('底线', 9), ('魔盒', 8), ('孩子', 7), ('评论', 6), ('事', 6), ('科学家', 6)]



图表 4.1.2-1

时期 2

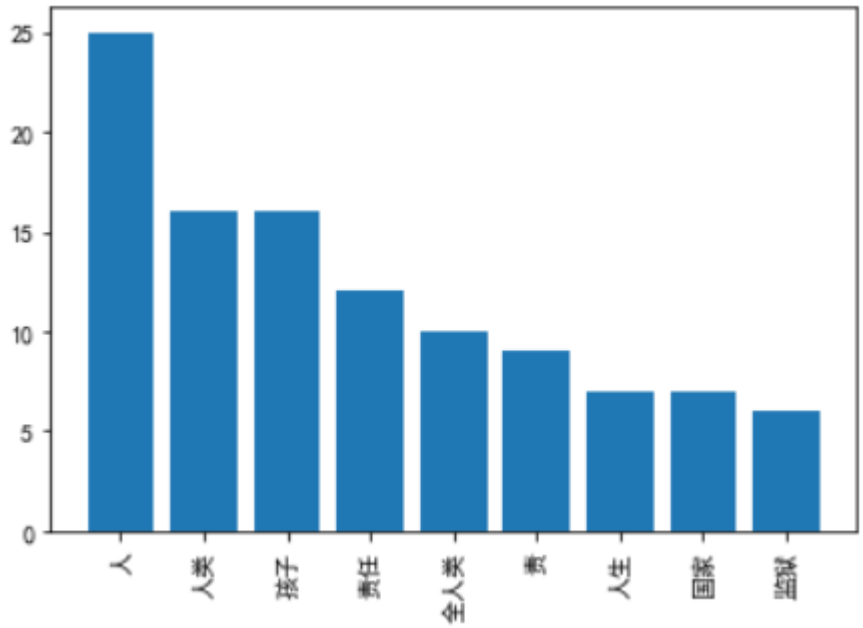
[('人', 47), ('人类', 15), ('孩子', 10), ('底线', 9), ('口语', 8), ('感觉', 7), ('事', 6), ('事情', 5), ('上帝', 5)]



图表 4.1.2-2

时期 3

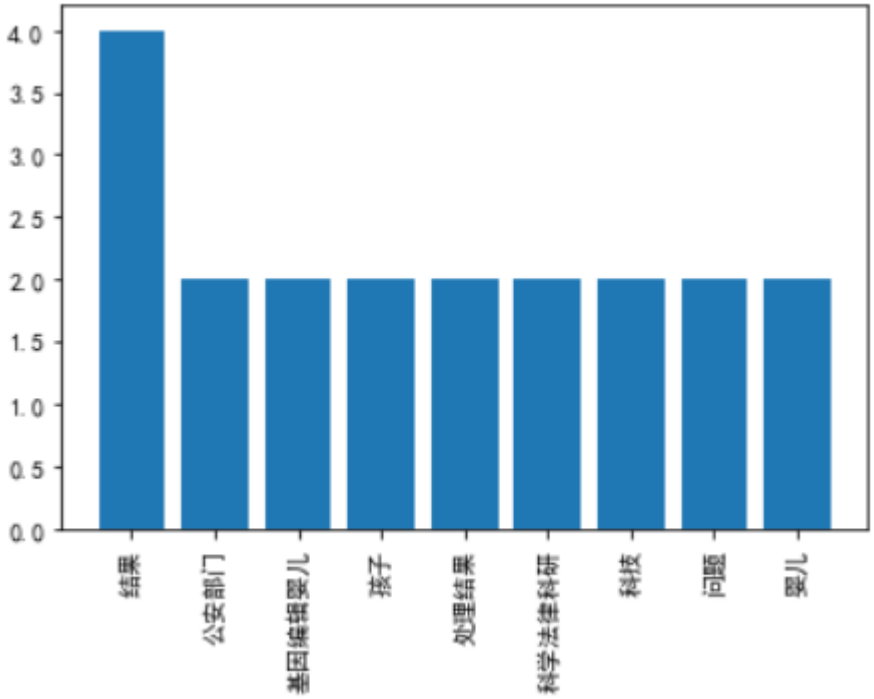
[('人', 25), ('人类', 16), ('孩子', 16), ('责任', 12), ('全人类', 10), ('责', 9), ('人生', 7), ('国家', 7), ('监狱', 6)]



图表 4.1.2-3

时期 4

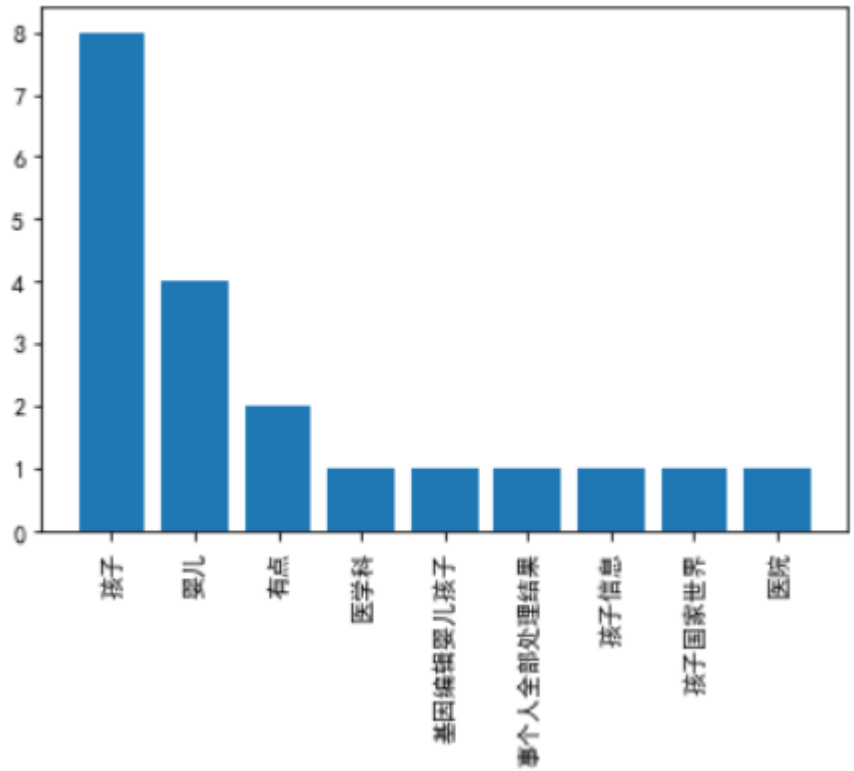
[('结果', 4), ('公安部门', 2), ('基因编辑婴儿', 2), ('孩子', 2), ('处理结果', 2), ('科学法律科研', 2), ('科技', 2), ('问题', 2), ('婴儿', 2)]



图表 4.1.2-4

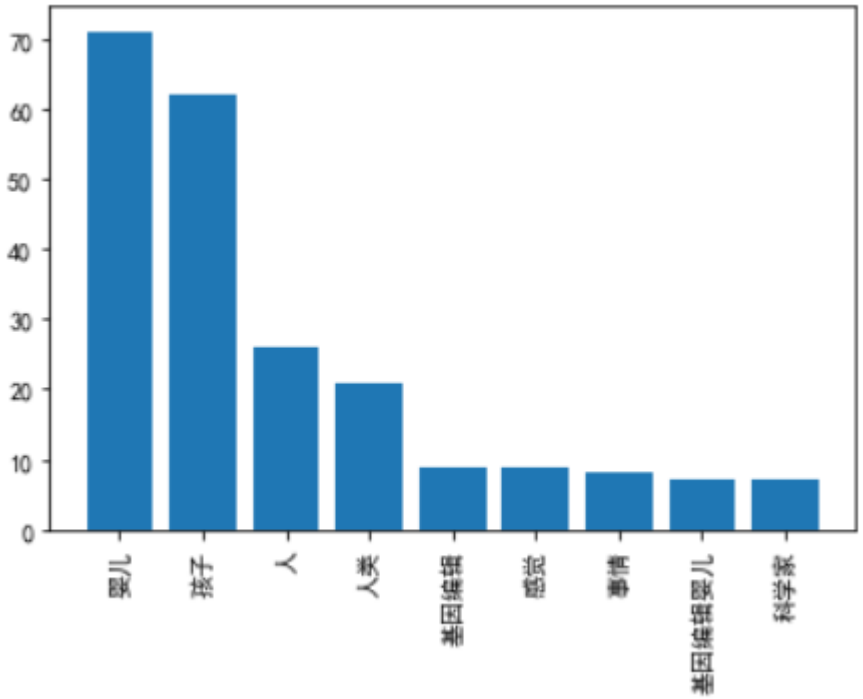
时期 5

[('孩子', 8), ('婴儿', 4), ('有点', 2), ('医学科', 1), ('基因编辑婴儿孩子', 1), ('事个人全部处理结果', 1), ('孩子信息', 1), ('孩子国家世界', 1), ('医院', 1)]



图表 4.1.2-5

时期 6
[('婴儿', 71), ('孩子', 62), ('人', 26), ('人类', 21), ('基因编辑', 9), ('感觉', 9), ('事情', 8), ('基因编辑婴儿', 7), ('科学家', 7)]



图表 4.1.2-6

4.2 数据的 LDA 主题提炼与可视化

1、主题提炼结果与可视化

时期 1

向量化提取特征关键词时

max_df=0.5 # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=3 # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1: 科学科技

前十关键词：科学 8.8 | 科技 6.33 | 评论 5.89 | 事情 4.69 | 感觉 4.08 | 真假 0.2 | 好事坏事 0.2 | 底线 0.2 | 基因 0.2 | 人性 0.2 |

Topic Nr.2: 编辑孩子技术有违人伦

前十关键词：孩子 6.91 | 技术 4.97 | 人伦 3.0 | 感觉 0.2 | 科技 0.2 | 事情 0.2 | 评论 0.2 | 科学 0.2 | 真假 0.2 | 好事坏事 0.2 |

Topic Nr.3: 科学家人性与基因技术

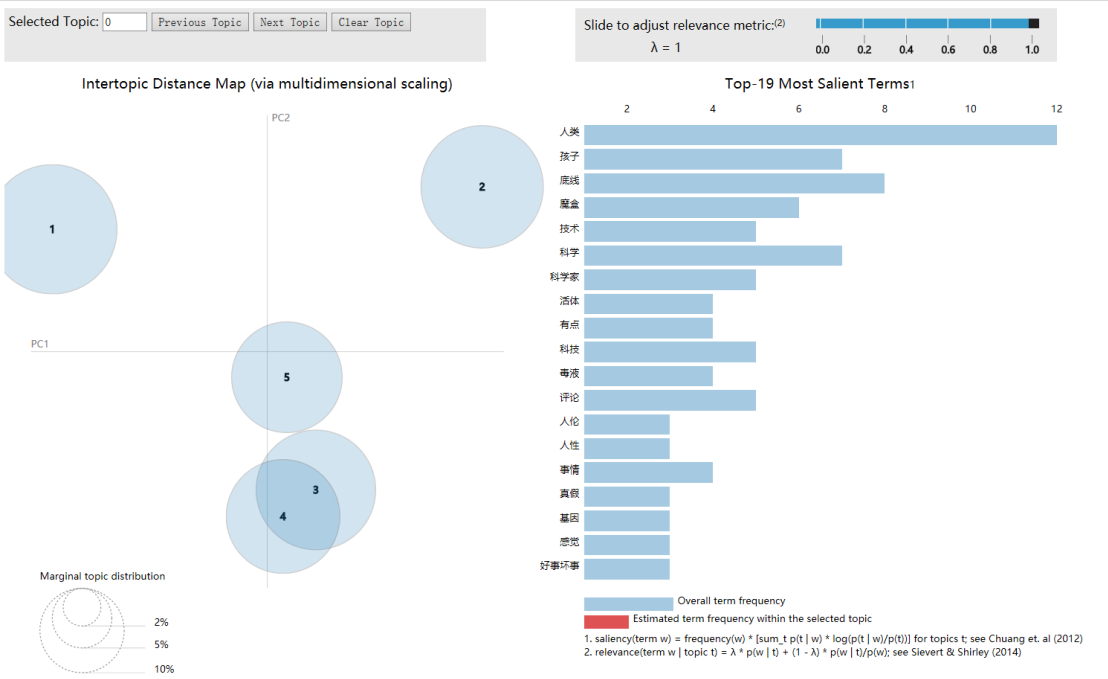
前十关键词：科学家 5.99 | 有点 5.01 | 毒液 4.73 | 人性 3.69 | 基因 3.12 | 感觉 0.2 | 事情 0.2 | 科技 0.2 | 评论 0.2 | 科学 0.2 |

Topic Nr.4: 人类活体实验

前十关键词：人类 12.64 | 活体 4.09 | 感觉 0.2 | 科技 0.2 | 事情 0.2 | 评论 0.2 | 科学 0.2 | 真假 0.2 | 好事坏事 0.2 | 基因 0.2 |

Topic Nr.5: 科学底线技术魔盒

前十关键词：底线 9.64 | 魔盒 7.85 | 真假 3.74 | 好事坏事 3.11 | 感觉 0.2 | 科技 0.2 | 事情 0.2 | 评论 0.2 | 科学 0.2 | 基因 0.2 |



可视化表 1

时期 2

向量化提取特征关键词时

max_df=0.5, # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=3 # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1: 疾病历史重演、扮演上帝还是疯子

前十关键词：事情 5.43 | 上帝 5.32 | 有点 5.22 | 历史 4.2 | 疯子 4.11 | 疾病 3.49 | 小孩孩子孩子人基因人类乱套有钱人孩子基因小孩孩子人类物种 3.03 | 细思 0.2 | 字幕 0.2 | 孩子 0.2 |

Topic Nr.2:口语感觉不错、科学家技术毒液

前十关键词：口语 8.05 | 感觉 7.33 | 评论 5.07 | 毒液 4.16 | 科学家 4.09 |

细思 0.2 | 字幕 0.2 | 小孩孩子孩子人基因人类乱套有钱人孩子基因小孩孩子人类物种 0.2 | 疾病 0.2 | 孩子 0.2 |

Topic Nr.3:遗传病魔盒、孩子编辑科技、天才与否

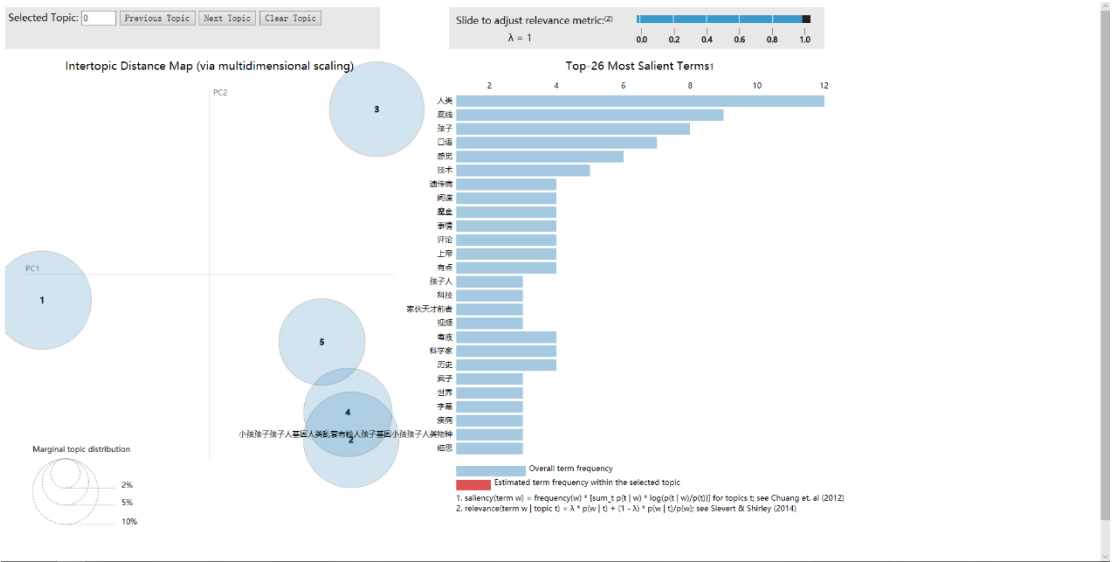
前十关键词：遗传病 3.99 | 魔盒 3.71 | 孩子人 3.1 | 科技 3.06 | 家伙天才前者 3.03 | 视频 3.02 | 细思 0.2 | 疾病 0.2 | 字幕 0.2 | 小孩孩子孩子人基因人类乱套有钱人孩子基因小孩孩子人类物种 0.2 |

Topic Nr.4:技术底线、间谍、世界问题

前十关键词：底线 10.14 | 技术 5.28 | 间谍 4.37 | 世界 3.17 | 细思 0.2 | 疾病 0.2 | 字幕 0.2 | 小孩孩子孩子人基因人类乱套有钱人孩子基因小孩孩子人类物种 0.2 | 孩子 0.2 | 历史 0.2 |

Topic Nr.5:人类、孩子

人类 14.92 | 孩子 10.22 | 字幕 4.13 | 细思 3.23 | 疾病 0.2 | 小孩孩子孩子人基因人类乱套有钱人孩子基因小孩孩子人类物种 0.2 | 历史 0.2 | 疯子 0.2 | 有点 0.2 | 事情 0.2 |



可视化表 2

时期 3

max_df=0.5, # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=3 # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1:关进监狱、没有资格扮演上帝

监狱 5.81 | 资格 3.93 | 资格能力 3.92 | 上帝 2.95 | 全人类 0.2 | 责任 0.2 | 国家 0.2 | 基因 0.2 | 人类 0.2 | 人生 0.2 |

Topic Nr.2:相关人员作为问题、没有底线

垃圾 5.91 | 问题 4.97 | 意思 3.94 | 底线 3.03 | 事情 3.01 | 人人生 2.94 | 全人类 0.2 | 国家 0.2 | 责任 0.2 | 基因 0.2 |

Topic Nr.3:孩子

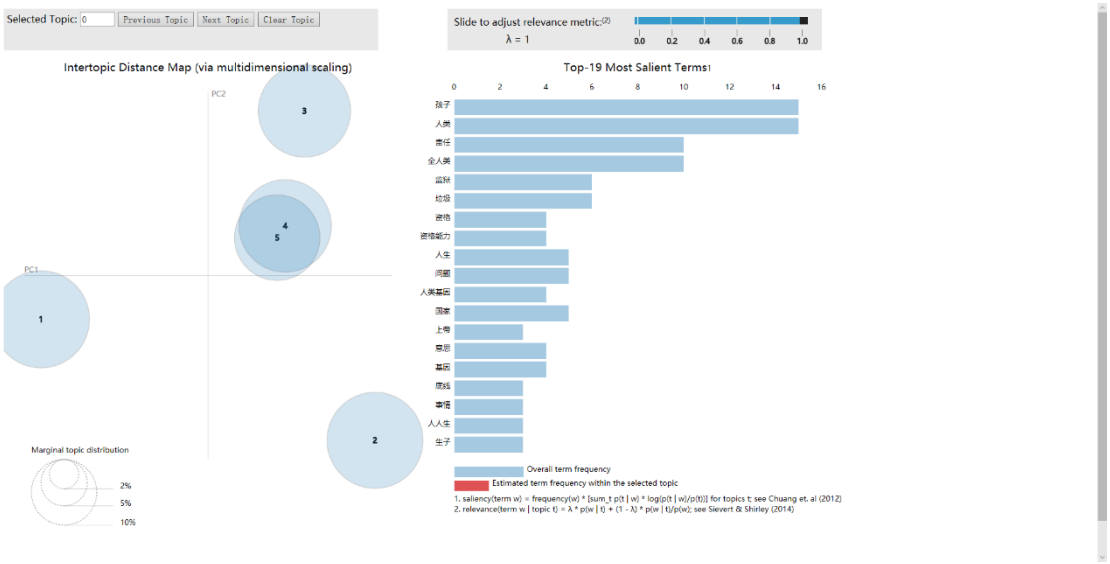
孩子 16.78 | 人类基因 4.98 | 生子 3.03 | 全人类 0.2 | 责任 0.2 | 国家 0.2 | 基因 0.2 | 人类 0.2 | 人生 0.2 | 底线 0.2 |

Topic Nr.4:人类

人类 18.42 | 人生 6.58 | 基因 4.83 | 全人类 0.2 | 责任 0.2 | 国家 0.2 | 生子 0.2 | 人类基因 0.2 | 底线 0.2 | 事情 0.2 |

Topic Nr.5:全人类责任

责任 14.63 | 全人类 14.47 | 国家 6.59 | 基因 0.2 | 人类 0.2 | 人生 0.2 | 生子 0.2 | 人类基因 0.2 | 底线 0.2 | 事情 0.2 |



可视化表 3

时期 4

max_df=0.5, # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=2) # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1: 基因编辑婴儿

基因编辑婴儿 1.62 | 人类 1.62 | 科技 1.6 | 问题 0.2 | 疯子 0.2 | 孩子 0.2 | 公安部门 0.2 | 科学法律科研 0.2 | 婴儿 0.2 | 处理结果 0.2 |

Topic Nr.2:处理结果

问题 1.62 | 处理结果 1.6 | 孩子 1.6 | 疯子 0.2 | 基因编辑婴儿 0.2 | 脑筋有点法则科学人们人类底线 0.2 | 人类 0.2 | 科学法律科研 0.2 | 婴儿 0.2 | 科技 0.2 |

Topic Nr.3:人类底线

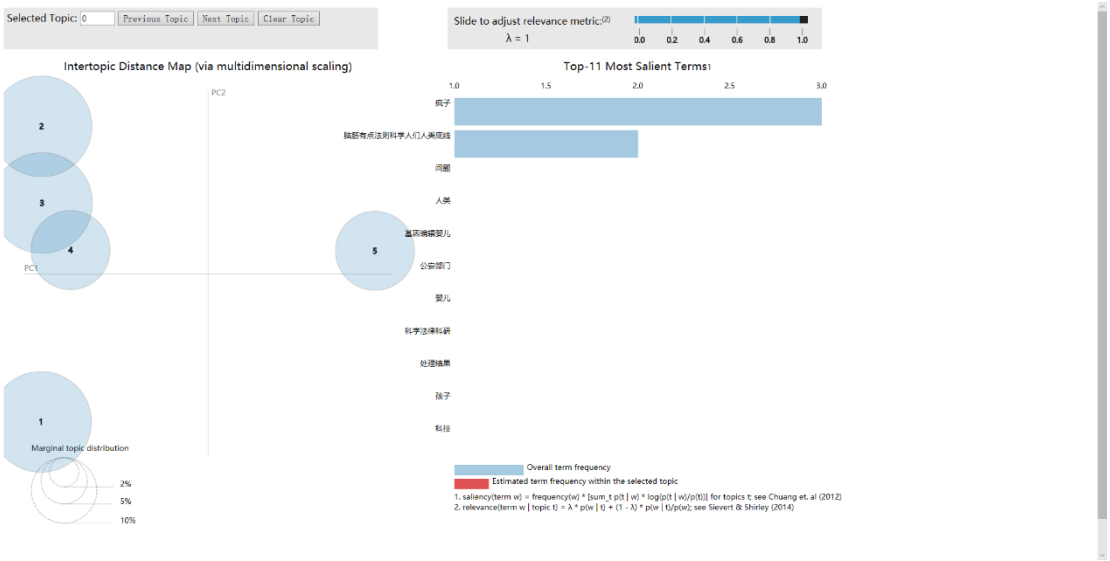
脑筋有点法则科学人们人类底线 1.64 | 科学法律科研 0.2 | 科技 0.2 | 孩子 0.2 | 处理结果 0.2 | 基因编辑婴儿 0.2 | 问题 0.2 | 疯子 0.2 | 人类 0.2 | 公安部门 0.2 |

Topic Nr.4:疯子

疯子 5.65 | 孩子 0.2 | 婴儿 0.2 | 科学法律科研 0.2 | 问题 0.2 | 人类 0.2 | 科技 0.2 | 公安部门 0.2 | 脑筋有点法则科学人们人类底线 0.2 | 处理结果 0.2 |

Topic Nr.5:婴儿

婴儿 1.62 | 科学法律科研 1.62 | 公安部门 1.62 | 科技 0.2 | 基因编辑婴儿 0.2 | 问题 0.2 | 孩子 0.2 | 疯子 0.2 | 脑筋有点法则科学人们人类底线 0.2 | 人类 0.2 |



可视化表 4

时期 5

max_df=0.5, # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=1 # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1:医生害人

医者医生人医生害人 1.2 | 事下文 1.2 | 下文 1.2 | 人类魔盒人类规则制度 1.2 | 发布会牢狱错孩子后果无辜孩子人权基因 1.2 | 名婴儿 1.2 | 基因编辑婴儿 1.2 | 基因编辑婴儿孩子 1.2 | 科研 1.2 | 硬道理产物 1.2 |

Topic Nr.2:事件丧病、医院问题

感觉丧尸 1.2 | 事结果 1.2 | 医院 1.2 | 婴儿问题人 1.2 | 魔盒 1.2 | 医生 1.2 | 时间媒体 1.2 | 人员人员 1.2 | 技术国家利益 1.2 | 人类罪 1.2 |

Topic Nr.3:超能力孩子

孩子 8.14 | 超能力人 1.2 | 科学贡献 1.2 | 热心网友 1.2 | 编辑基因方向全世界基因人类事情概率 1.2 | 行医实质量刑条件原告程度罪杀人法制 1.2 | 孩子信息 1.2 | 人类基因问题人评论 1.2 | 悲剧 1.2 | 婴儿事 1.2 |

Topic Nr.4:负面影响、人类敌人

婴儿婴儿编辑基因人类规律负面影响无法人类婴儿重点 1.2 | 全人类敌 1.2 | 步子蛋专家技术人孩子科学狂人 1.2 | 科学法律底线 1.2 | 基因编辑婴儿行医有点刑法规定基因编辑人类罪名 1.2 | 世界 1.2 | 理想阴影 1.2 | 技术角度意义人类绝症现实新人类人月球现实意义人人动手蠢事 1.2 | 超能力人 0.21 | 基因编辑婴儿孩子 0.21 |

Topic Nr.5:婴儿

婴儿 4.18 | 有点 2.19 | 老师理论区别学医 1.2 | 畜牲人驱逐出境 1.2 | 事个人全部处理结果 1.2 | 基因人 1.2 | 感觉事 1.2 | 孩子 0.22 | 医学科 0.21 | 医院 0.21 |

可视化表 5

时期 6

max_df=0.5, # 阈值如果某个词的 document frequence 大于 max_df，不当作关键词.float，词出现的次数与语料库文档数的百分比，int 出现次数

min_df=3) # 如果某个词的 document frequence 小于 min_df，则这个词不会被当作关键词

Topic Nr.1:孩子

孩子 63.88 | 医生 5.92 | 婴儿人权 3.18 | 事结果 3.08 | 人类人类 0.2 | 底线 0.2 | 成本 0.2 | 婴儿 0.2 | 科技 0.2 | 基因 0.2 |

Topic Nr.2:人类

人类 20.69 | 感觉 9.08 | 事情 7.93 | 基因编辑婴儿 6.93 | 基因 4.31 | 科技 3.26 | 意思 3.19 | 技术 3.19 | 基因编辑意思 3.17 | 太轻 3.07 |

Topic Nr.3:行医

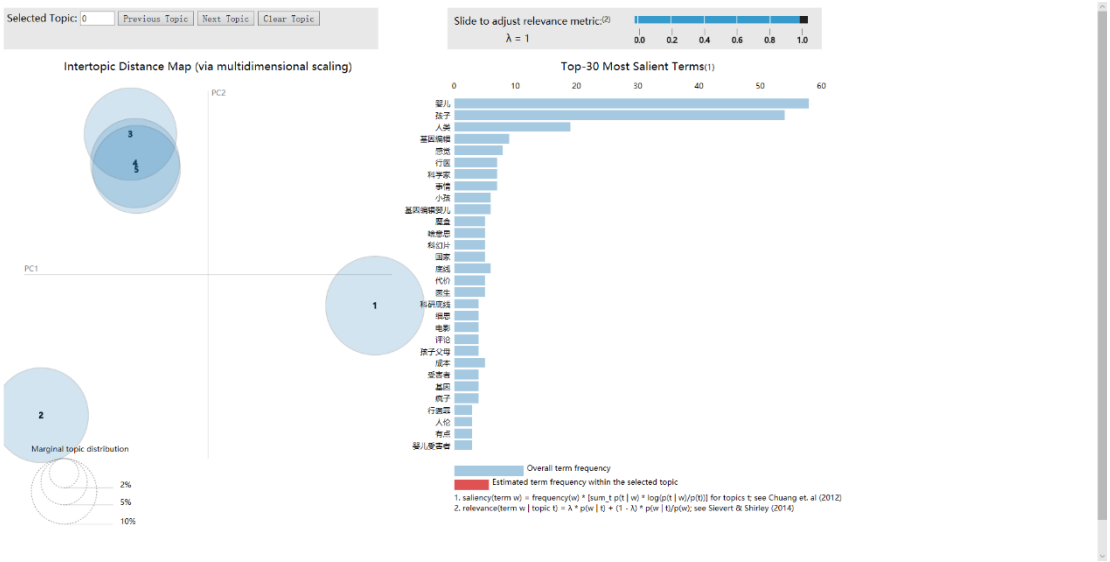
行医 7.17 | 小孩 6.45 | 魔盒 4.88 | 细思 4.25 | 电影 4.15 | 评论 4.0 | 受害者 3.71 | 疯子 3.64 | 行医罪 3.3 | 人伦 3.24 |

Topic Nr.4:婴儿

婴儿 74.21 | 底线 7.0 | 成本 6.0 | 人类人类 2.96 | 婴儿人权 0.2 | 事结果 0.2 | 科技 0.2 | 基因 0.2 | 基因编辑意思 0.2 | 医生 0.2 |

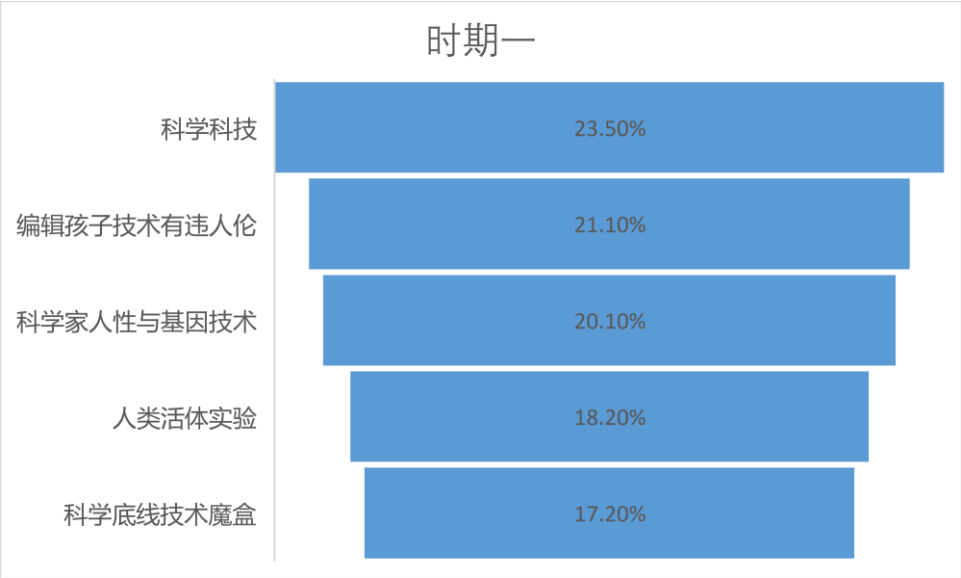
Topic Nr.5:基因编辑

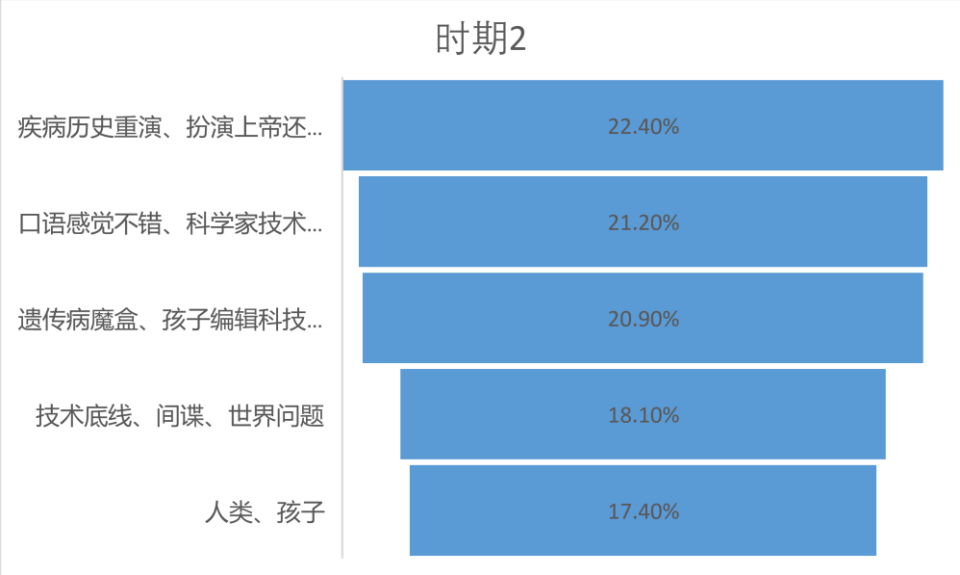
基因编辑 9.03 | 科学家 7.16 | 啥意思 5.07 | 科幻片 5.03 | 国家 5.02 | 代价 4.68 | 科研底线 4.57 | 孩子父母 4.19 | 建议 3.21 | 能力 3.16 |

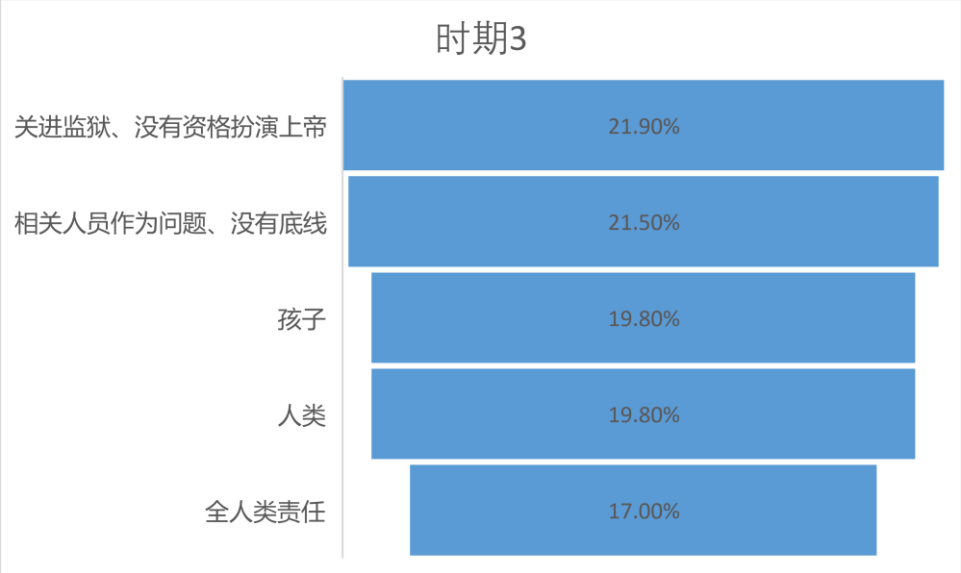


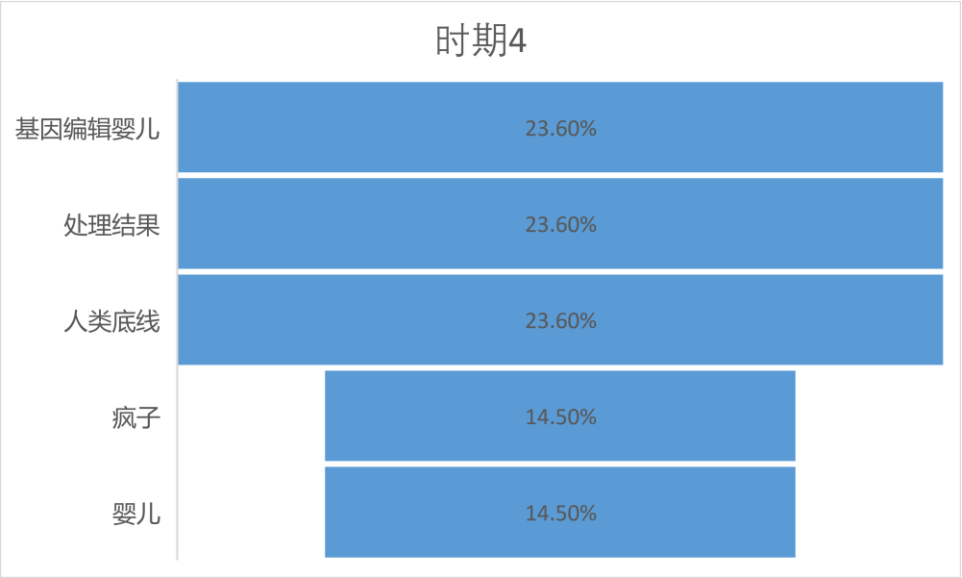
可视化表 6

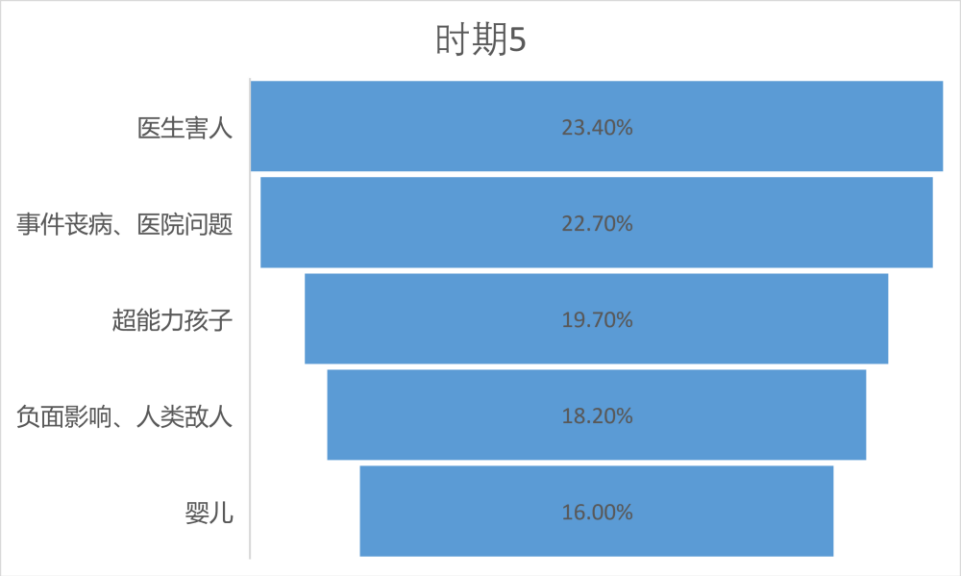
2、六个时期各话题占比可视化

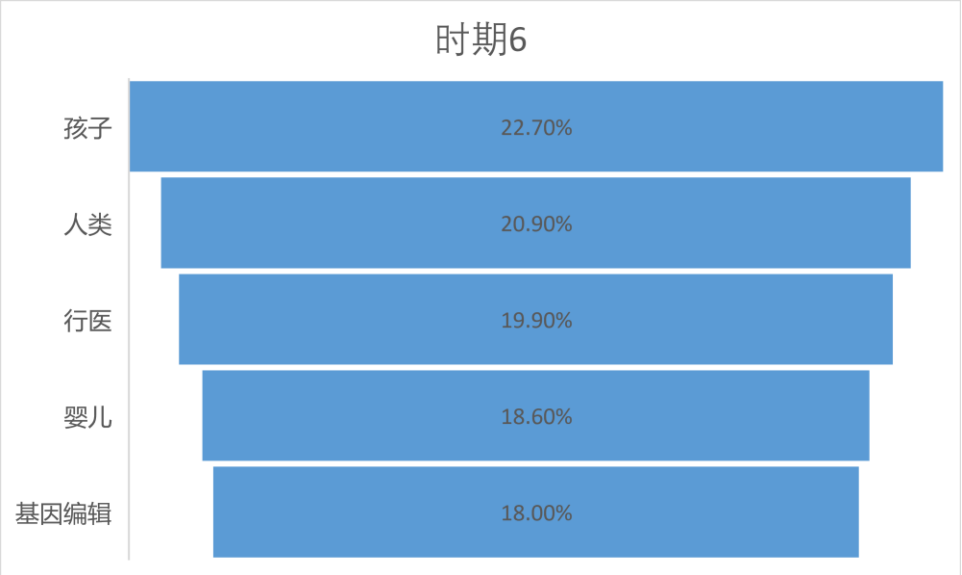







时期4
基因编辑婴儿
23.60%
处理结果
23.60%
人类底线
23.60%
疯子
14.50%
婴儿
14.50%


时期5
医生害人
23.40%
事件丧病、医院问题
22.70%
超能力孩子
19.70%
负面影响、人类敌人
18.20%
婴儿
16.00%


时期6
孩子
22.70%
人类
20.90%
行医
19.90%
婴儿
18.60%
基因编辑
18.00%

第五章 数据的共词分析

5.1 数据的共词分析模型

共词分析包括以下几个步骤：分词并去除停用词，共词矩阵构建，共词矩阵的可视化。

（1） 分词并去除停用词

使用 jieba 分词对原始语料进行分词，分词完毕后再利用一般停用词库对分词结果进行处理，目的是为了去除无意义的英文字符、数字、数学字符、标点符号及使用频率特高的单汉字和其他一些对理解语料其实并无太大价值的词语。

分词结果上文已有描述，此处不再做多余赘述。

代码截图：

In [1]:
```
import jieba
import xlsxwriter

# 待分词的文本路径
sourceTxt = 'D:/txt/First 1A.txt'
# 分好词后的文本路径
targetTxt = 'D:/txt/First 1B.txt'

# 对文本进行操作
with open(sourceTxt, 'r', encoding = 'utf-8') as sourceFile, open(targetTxt, 'a+', encoding = 'utf-8') as targetFile:
    for line in sourceFile:
        seg = jieba.cut(line.strip(' '), cut_all = False)
        # 分好词之后之间用空格隔断
        output = ' '.join(seg)
        targetFile.write(output)
        targetFile.write('\n')
    print('写入成功！')
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\toyot\AppData\Local\Temp\jieba.cache
Loading model cost 0.745 seconds.
Prefix dict has been built successfully.
```

```
写入成功！
```

In [ ]:

In [12]:
```
stop = []
standard_stop = []
text = []
after_text = []
file_stop = r'D:/txt/stopword.txt'  # 停用词表
file_text = r'D:/txt/First 1B.txt'  # 要处理的文本集合
with open(file_stop,'r',encoding='utf-8') as f :
    lines = f.readlines()  # lines是list类型
    for line in lines:
        lline  = line.strip('\n')     # line 是str类型
        stop.append(lline)        # 将stop 是列表形式

# stop 的元素是一行一行的 句子,需要进行转化为一个词一行,即下面:
for i in range(0,len(stop)):
    for word in stop[i].split():
        standard_stop.append(word)
# print(standard_stop)
```

```
In [12]: stop = []
         standard_stop = []
         text = []
         after_text = []
         file_stop = r'D:/txt/stopword.txt'  # 停用词表
         file_text = r'D:/txt/First 1B.txt'  # 要处理的文本集合
         with open(file_stop,'r',encoding='utf-8') as f :
             lines = f.readlines()  # lines是list类型
             for line in lines:
                 lline  = line.strip('\n')     # line 是str类型
                 stop.append(lline)        # 将stop 是列表形式

         # stop 的元素是一行一行的 句子,需要进行转化为一个词一行,即下面:
         for i in range(0,len(stop)):
             for word in stop[i].split():
                 standard_stop.append(word)
         # print(standard_stop)


         # 读取文本集,
         with open(file_text,'r',encoding='utf-8') as f :
             lines = f.readlines()
             print(lines)
             for line  in lines:
                 # lline = line.strip('\n')
                 # print(lline)
                 lline = line.split(' ')
                 # print(lline)
                 for i in lline:
                     if i not in  standard_stop:
                         after_text.append(i)
                         after_text.append(' ')


         print(' ',after_text,'/n')

         # 将结果保存在txt中
         with open(r'D:/txt/First 1B result.txt','w+',encoding='utf-8')as f :
             for i in after_text:
                 f.write(i)
```

结果：

First 1B result.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

技术 能力 强 违背 伦理 底线 第一个 研究 婴儿

做 活体 实验 事 宣扬

这项 工程 美国 禁止 事关 伦理 人性 法律 谨慎 宣传

未 报备 做 实验 医院 负责人 终身 禁止 医学 领域

潘多拉 魔盒

鄙视 中国 这位 学者

天天 做 CrisperCas9 耻辱 想 做 美国 早就 做 做 做 没点 逼数 后果 成熟 研究 人体 开刀

擅自 有违 伦理 实验 严惩

支持 修改 基因 遗传 孩子 长大 改动 基因 结婚 生子 扩散 后遗症 将来 影响 统计

几年 人工 感染 艾滋病毒 验证 基因 编辑 效果 两个 活着 小白鼠

基因工程 伦理 区 阶段 有太大 影响 未来 基因 贫困 扶贫 消除 个体 多样性 消除 生物 多样性

深圳市 卫生 计生委 医学 伦理 专家 委员会 获悉 该项 试验 前 并未 部门 报备 正 开会研究

医学 伦理 731 纳粹 区别

官网 宣传 事情 一点 光荣

基因 不像 电脑 程序 代码 组只 负责 一项 功能 基因 影响 错综复杂 稍有不慎 就会 出大 这件

（2） 共词矩阵构建

本文使用了一款名为 Co-Occurence3.9 的软件进行共词矩阵的构建，直接选择待分析的分词完毕的语料进行，考虑到大量的语料的影响，选择了最小的出现

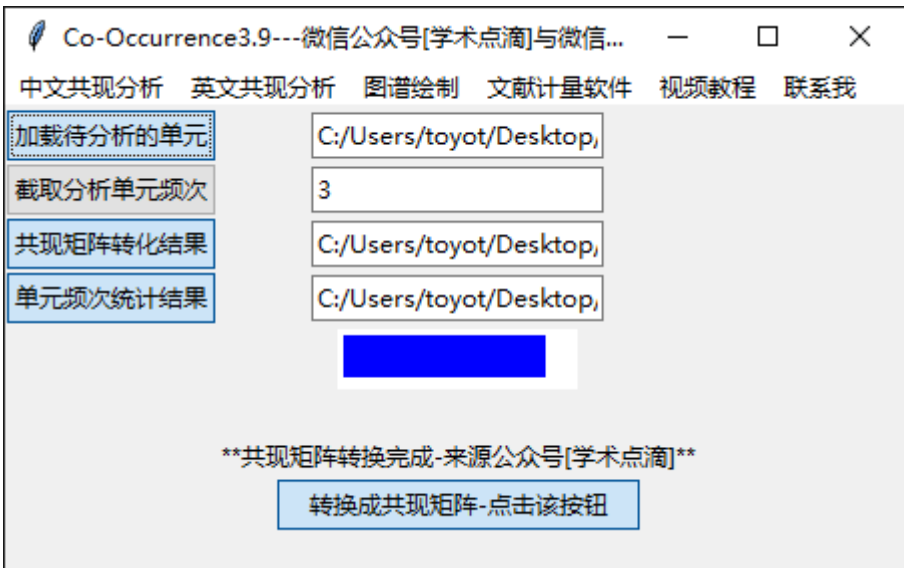
频次为 3，输出共词矩阵即可。[10]

鉴于此软件仅能识别分号作为分隔符，故将导出的分词结果进行如下处理：

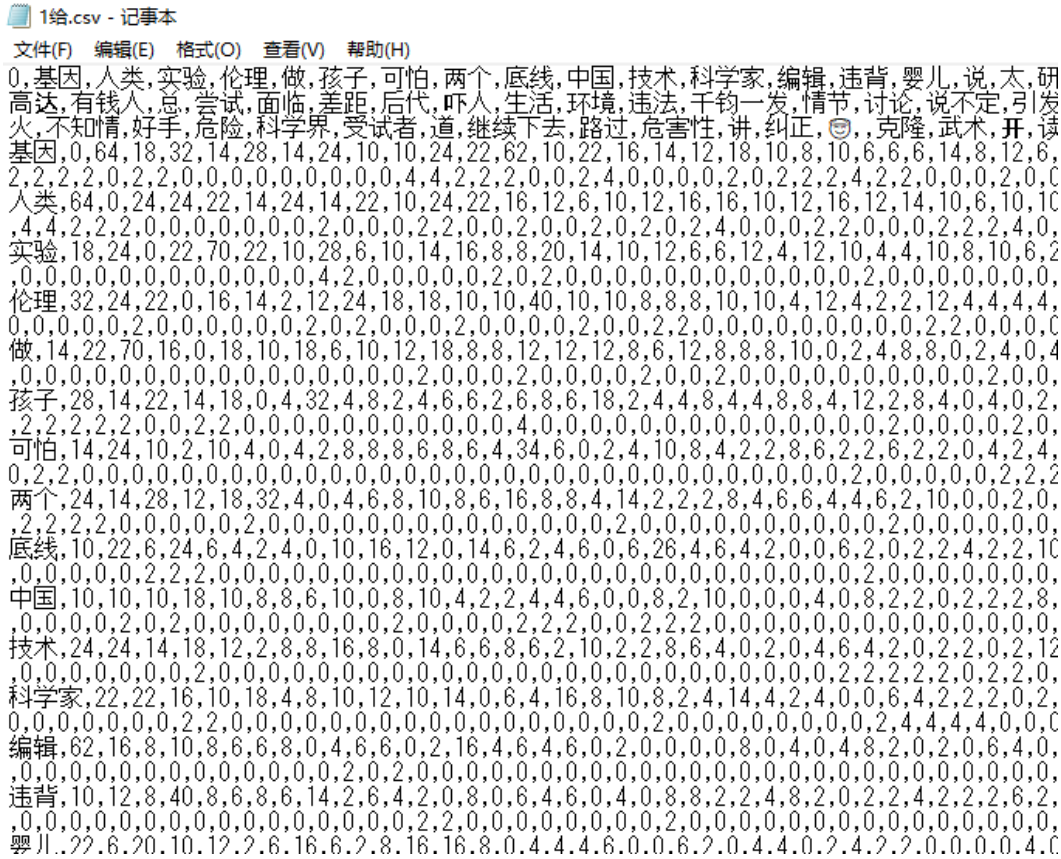
去除两行数据之间的空格（Excel 过滤）

去除头尾的分号（使用 Excel 的 LEN（）函数获取每一个字符串的长度后，利用 MID（text，start_num,num_chars）函数去除指定位置头尾的分号）

结果如下：



时间点 1：



0,基因,人类,实验,伦理,做,孩子,可怕,两个,底线,中国,技术,科学家,编辑,违背,婴儿,说,太,研
高达,有钱人,总,尝试,面临,差距,后代,吓人,生活,环境,违法,千钧一发,情节,讨论,说不定,引发
火,不知情,好手,危险,科学界,受试者,道,继续下去,路过,危害性,讲,纠正,😇,,克隆,武术,开,读
基因,0,64,18,32,14,28,14,24,10,10,24,22,62,10,22,16,14,12,18,10,8,10,6,6,6,14,8,12,6,
2,2,2,2,0,2,2,0,0,0,0,0,0,0,0,0,4,4,2,2,2,0,0,2,4,0,0,0,0,2,0,2,2,2,4,2,2,0,0,0,2,0,0
人类,64,0,24,24,22,14,24,14,22,10,24,22,16,12,6,10,12,16,16,10,12,16,12,14,10,6,10,10
,4,4,2,2,2,0,0,0,0,0,0,0,2,0,0,0,2,2,0,0,2,0,0,2,0,2,0,2,4,0,0,0,2,2,0,0,0,2,2,2,4,0,
实验,18,24,0,22,70,22,10,28,6,10,14,16,8,8,20,14,10,12,6,6,12,4,12,10,4,4,10,8,10,6,2
,0,0,0,0,0,0,0,0,0,0,0,0,4,2,0,0,0,0,0,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,
伦理,32,24,22,0,16,14,2,12,24,18,18,10,10,40,10,10,8,8,8,10,10,4,12,4,2,2,12,4,4,4,4,
0,0,0,0,0,2,0,0,0,0,0,0,2,0,2,0,0,0,2,0,0,0,0,2,0,0,2,2,0,0,0,0,0,0,0,0,0,2,2,0,0,0,0
做,14,22,70,16,0,18,10,18,6,10,12,18,8,8,12,12,12,8,6,12,8,8,8,10,0,2,4,8,8,0,2,4,0,4
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,2,0,0,0,0,2,0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,
孩子,28,14,22,14,18,0,4,32,4,8,2,4,6,6,2,6,8,6,18,2,4,4,8,4,4,8,8,4,12,2,8,4,0,4,0,2,
,2,2,2,2,2,0,0,2,2,0,0,0,0,0,0,0,0,0,0,0,4,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,2,0,
可怕,14,24,10,2,10,4,0,4,2,8,8,8,6,8,6,4,34,6,0,2,4,10,8,4,2,2,8,6,2,2,6,2,2,0,4,2,4,
0,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,2,2,2
两个,24,14,28,12,18,32,4,0,4,6,8,10,8,6,16,8,8,4,14,2,2,2,8,4,6,6,4,4,6,2,10,0,0,2,0,
,2,2,2,2,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,
底线,10,22,6,24,6,4,2,4,0,10,16,12,0,14,6,2,4,6,0,6,26,4,6,4,2,0,0,6,2,0,2,2,4,2,2,10
,0,0,0,0,0,2,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,
中国,10,10,10,18,10,8,8,6,10,0,8,10,4,2,2,4,4,6,0,0,8,2,10,0,0,0,4,0,8,2,2,0,2,2,2,8,
,0,0,0,0,2,0,2,0,0,0,0,0,0,0,0,0,2,0,0,0,0,2,2,2,0,0,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
技术,24,24,14,18,12,2,8,8,16,8,0,14,6,6,8,6,2,10,2,2,8,6,4,0,2,0,4,6,4,2,0,2,2,0,2,12
,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,2,2,2,0,2,2,0,
科学家,22,22,16,10,18,4,8,10,12,10,14,0,6,4,16,8,10,8,2,4,14,4,2,4,0,0,6,4,2,2,2,0,2,
0,0,0,0,0,0,0,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,2,4,4,4,4,0,0,0
编辑,62,16,8,10,8,6,6,8,0,4,6,6,0,2,16,4,6,4,6,0,2,0,0,0,0,8,0,4,0,4,8,2,0,2,0,6,4,0,
,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
违背,10,12,8,40,8,6,8,6,14,2,6,4,2,0,8,0,6,4,6,0,4,0,8,8,2,2,4,8,2,0,2,2,4,2,2,2,6,2,
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,
婴儿,22,6,20,10,12,2,6,16,6,2,8,16,16,8,0,4,4,4,6,0,0,6,2,0,4,4,0,2,4,2,2,0,0,0,0,4,0

共词矩阵 1

时间点 2：

2c.csv - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
0,基因,人类,做,孩子,说,实验,技术,中国,想,编辑,真的,两个,遗传病,伦理,可怕,科学家,道德,家庭,支持,底线,疾病,改变
基因,0,148,56,72,30,36,74,14,34,104,26,24,52,22,22,20,10,22,16,8,24,50,16,0,0,50,56,22,30,24,28,16,8,22,12,2
人类,148,0,64,44,30,44,54,10,24,36,24,16,30,22,16,28,14,24,10,8,28,32,18,0,0,18,38,28,16,36,24,20,4,22,14,0,
做,56,64,0,52,42,146,44,14,20,22,22,24,18,26,8,22,16,16,10,20,16,8,14,6,2,24,6,6,10,14,16,22,8,22,12,0,4,10,
孩子,72,44,52,0,32,46,20,12,22,20,16,62,26,10,4,10,2,28,8,2,14,10,10,0,0,8,24,4,16,2,12,10,10,6,8,0,4,4,6,16
说,30,30,42,32,0,20,20,44,26,6,18,14,14,12,6,16,4,8,2,6,4,2,10,2,32,2,2,0,4,0,12,8,12,6,0,38,8,10,6,12,12,6,
实验,36,44,146,46,20,0,26,14,20,10,10,32,6,14,8,18,14,4,8,16,10,2,10,4,0,22,10,4,10,4,8,26,4,12,4,0,4,0,6,4,
技术,74,54,44,20,20,26,0,10,12,28,20,10,28,26,12,14,12,16,14,10,8,14,12,0,0,12,8,8,10,24,12,6,8,12,6,0,0,4,4
中国,14,10,14,12,44,14,10,0,12,4,4,8,2,0,6,16,8,2,6,4,0,0,6,2,8,6,0,4,2,4,6,0,4,2,6,8,4,0,2,2,6,14,6,2,18,2,
想,34,24,20,22,26,20,12,12,0,14,12,18,4,6,0,8,6,2,8,6,2,2,4,2,4,8,10,0,6,4,6,2,8,4,4,2,2,6,10,4,4,4,4,8,4,8,
编辑,104,36,22,20,6,10,28,4,14,0,8,6,14,10,8,8,6,8,6,4,8,8,2,0,0,28,6,10,6,6,4,6,0,8,2,0,2,2,2,2,4,4,2,4,0,4
真的,26,24,22,16,18,10,20,4,12,8,0,4,6,6,12,2,2,10,8,6,12,6,2,2,4,6,8,0,2,8,2,2,2,8,4,0,4,8,12,4,4,4,0,8,2,4
两个,24,16,24,62,14,32,10,8,18,6,4,0,0,2,0,2,8,0,2,14,4,6,4,0,0,12,8,2,8,0,4,18,2,2,6,0,0,2,2,10,0,6,2,2,0,2
遗传病,52,30,18,26,14,6,28,2,4,14,6,0,0,4,0,4,2,20,4,2,4,10,0,0,0,6,4,0,4,4,4,4,2,10,0,0,0,6,4,10,2,4,0,2,0,
伦理,22,22,26,10,12,14,26,0,6,10,6,2,4,0,2,4,20,10,8,12,2,4,16,2,2,4,0,2,6,6,4,2,6,4,2,0,0,2,2,0,0,2,2,4,0,0
可怕,22,16,8,4,6,8,12,6,0,8,12,0,0,2,0,4,2,0,2,4,6,12,4,0,0,0,0,2,6,2,0,2,0,8,4,0,10,4,0,2,18,6,6,2,0,0,0,2,
科学家,20,28,22,10,16,18,14,16,8,8,2,2,4,4,4,0,2,2,0,8,8,6,8,0,2,0,0,12,2,2,6,2,2,6,0,2,2,0,2,0,4,12,0,0,0,2
道德,10,14,16,2,4,14,12,8,6,6,2,8,2,20,2,2,0,2,2,42,0,2,8,2,0,2,0,0,2,6,2,10,6,2,0,0,0,0,2,0,0,0,8,0,0,4,2,0
家庭,22,24,16,28,8,4,16,2,2,8,10,0,20,10,0,2,2,0,4,0,10,2,2,0,0,8,2,2,4,6,2,4,6,2,2,0,0,4,2,12,0,2,2,4,0,8,4
支持,16,10,10,8,2,8,14,6,8,6,8,2,4,8,2,0,2,4,0,0,4,0,6,0,0,0,6,0,0,12,2,4,2,6,2,0,2,0,2,0,0,2,4,4,0,4,0,0,0,
底线,8,8,20,2,6,16,10,4,6,4,6,14,2,12,4,8,42,0,0,0,2,0,8,4,0,0,0,2,4,2,2,12,4,2,2,0,0,0,2,0,0,6,0,2,0,4,2,0,
疾病,24,28,16,14,4,10,8,0,2,8,12,4,4,2,6,8,0,10,4,2,0,4,6,0,0,4,2,0,2,10,2,4,2,2,4,0,4,2,2,4,0,0,2,0,0,8,2,4
改变,50,32,8,10,2,2,14,0,2,8,6,6,10,4,12,6,2,2,0,0,4,0,2,0,0,2,2,8,6,4,4,4,0,4,2,0,0,12,0,4,6,0,2,0,0,0,2,4,
科学,16,18,14,10,10,10,12,6,4,2,2,4,0,16,4,8,8,2,6,8,6,2,0,2,0,2,4,2,4,6,6,2,2,2,8,0,0,0,0,2,0,2,2,2,0,0,4,0
抓,0,0,6,0,2,4,0,2,2,0,2,0,0,2,0,0,2,0,0,4,0,0,2,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,4,0,0,0
英语,0,0,2,0,32,0,0,8,4,0,4,0,0,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,6,0,2,0,0,2,0,0,0,0,0,0,0,0,0,0,
婴儿,50,18,24,8,2,22,12,6,8,28,6,12,6,4,0,0,2,8,0,0,4,2,2,0,0,8,4,4,6,2,6,0,4,4,0,4,2,4,2,0,0,2,6,0,4,0,0,
改造,56,38,6,24,2,10,8,0,10,6,8,8,4,0,0,0,0,2,6,0,2,2,4,0,0,8,0,10,6,2,0,4,2,2,2,0,0,2,0,8,2,2,2,2,0,4,0,0,0
世界,22,28,6,4,0,4,8,4,0,10,0,2,0,2,2,12,0,2,0,2,0,8,2,0,0,4,10,0,4,6,0,2,4,2,0,0,0,0,2,2,2,6,0,0,0,0,2,2,0,
未来,30,16,10,16,4,10,10,2,6,6,2,8,4,6,6,2,2,4,0,4,2,6,4,0,0,4,6,4,0,6,4,4,0,2,4,0,4,4,2,6,2,0,6,0,0,2,2,2,2
发展,24,36,14,2,0,4,24,4,4,6,8,0,4,6,2,2,6,6,12,2,10,4,6,0,0,6,2,6,6,0,4,0,4,4,6,0,4,2,0,4,2,2,12,2,0,2,2,2,
研究,28,24,16,12,12,8,12,6,6,4,2,4,4,4,0,6,2,2,2,2,2,4,6,2,0,2,0,0,4,4,0,4,0,4,0,4,4,0,2,4,0,2,0,4,6,6,0,6,2,2,0
生命,16,20,22,10,8,26,6,0,2,6,2,18,4,2,2,2,10,4,4,12,4,4,2,0,0,6,4,2,4,0,4,0,0,2,2,0,4,2,2,4,0,2,0,0,0,0,0,0
请,8,4,8,10,12,4,8,4,8,0,2,2,2,6,0,2,6,6,2,4,2,0,2,0,2,0,2,4,0,4,0,0,0,0,2,8,0,0,2,0,2,0,2,2,2,2,2,0,0,2,4,2
后果,22,22,22,6,6,12,12,2,4,8,8,2,10,4,8,6,2,2,6,2,2,4,2,0,0,4,2,2,2,4,4,2,0,0,0,0,2,2,2,0,2,2,4,0,0,4,0,6,2
事情,12,14,12,8,0,4,6,6,4,2,4,6,0,2,4,0,0,2,2,2,4,2,8,0,0,4,2,0,4,6,4,2,2,0,0,0,4,0,2,2,0,0,2,2,0,0,2,0,0,0,
中文,2,0,0,0,38,0,0,0,8,2,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,6,0,0,0,0,0,0,0,0,8,0,0,0,2,0,0,0,0,0,0,0,2,10,0,0,0,0,0,1
```

共词矩阵 2

时间点 3：

3c.csv - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
0,负责,孩子,基因,两个,人类,说,实验,半辈子,做,未来,想,露露,编辑,责任,真的,后代,人生,娜娜,科学家
间,HIV,开,骄傲,后悔,喷,不配,说句,多人,电影,明天,蔓延,付,p,可笑,管得,人,第一,审查,不想,负得,明
负责,0,154,64,130,82,100,64,78,54,48,30,20,18,20,24,28,52,18,22,50,26,28,36,18,12,12,18,34,34
孩子,154,0,64,158,30,38,46,20,46,40,16,10,22,12,18,24,18,8,18,8,12,14,14,4,6,4,16,26,6,6,10,1
基因,64,64,0,36,56,14,20,10,20,18,10,10,62,14,12,30,6,10,16,8,14,4,10,8,4,6,12,10,0,2,6,4,12,
两个,130,158,36,0,16,36,40,26,40,26,14,6,10,18,12,14,22,6,10,6,12,10,20,6,2,8,14,18,18,2,6,12
人类,82,30,56,16,0,20,24,4,22,30,8,6,20,16,16,14,6,6,10,2,10,6,6,0,2,6,6,2,2,0,6,10,6,8,8,4,0
说,100,38,14,36,20,0,16,12,28,14,8,6,8,14,6,6,10,6,10,2,10,6,6,6,0,4,10,14,6,0,2,10,6,6,6,8,4
实验,64,46,20,40,24,16,0,2,62,10,14,4,12,8,4,4,6,4,6,0,18,6,12,8,2,4,4,4,4,2,6,8,4,6,4,6,4,4,
半辈子,78,20,10,20,26,4,12,2,0,4,8,0,8,2,4,4,2,12,8,0,8,2,12,18,0,2,2,0,2,6,0,4,0,6,0,2,4,0,2,0,
做,54,46,20,40,22,28,62,4,0,18,6,10,8,6,10,6,2,10,8,2,12,10,6,10,0,6,16,4,4,4,6,6,8,4,6,12,6,
未来,48,40,18,26,30,14,10,8,18,0,4,14,8,2,14,6,0,14,10,6,4,0,4,0,4,2,4,2,6,0,6,2,6,4,2,12,4,4
想,30,16,10,14,8,8,14,0,6,4,0,4,6,2,10,4,2,4,2,6,4,4,2,6,2,2,2,2,4,2,4,2,0,2,0,2,4,6,0,0,0,2,
露露,20,10,10,6,6,6,4,8,10,14,4,0,4,0,2,6,2,60,2,4,2,0,2,0,0,2,4,8,0,0,0,2,0,0,8,0,2,2,0,2,
编辑,18,22,62,10,20,8,12,2,8,8,6,4,0,6,10,8,2,4,6,2,4,0,2,0,0,2,6,0,0,0,2,4,0,2,4,4,4,4,2,2,0
责任,20,12,14,18,16,14,8,4,6,2,2,0,6,0,2,2,8,0,2,6,0,0,2,0,0,18,2,0,2,0,2,2,2,2,0,2,0,6,0,18,
真的,24,18,12,12,16,6,4,4,10,14,10,2,10,2,0,4,4,2,4,2,0,4,2,2,0,0,8,2,10,0,8,4,4,4,4,4,4,0,2,
后代,28,24,30,14,14,6,4,2,6,6,4,6,8,2,4,0,2,4,2,0,2,2,2,2,2,2,6,10,0,2,0,0,8,4,2,4,2,10,6,0,2
人生,52,18,6,22,6,10,6,12,2,0,2,2,2,8,4,2,0,2,2,0,0,2,2,0,0,4,4,0,4,0,2,2,2,2,2,0,0,0,2,2,0,0
娜娜,18,8,10,6,6,6,4,8,10,14,4,60,4,0,2,4,2,0,0,4,2,0,2,0,0,2,4,8,0,0,2,0,2,0,0,8,0,2,2,0,2,2
科学家,22,18,16,10,10,10,6,0,8,10,2,2,6,2,4,2,2,0,0,2,4,0,0,6,2,2,0,0,2,0,0,4,0,4,2,2,0,4,4,2
全人类,50,8,8,6,2,2,0,8,2,6,6,4,2,6,2,0,0,4,2,0,0,2,0,0,0,2,4,2,2,0,2,2,6,0,2,4,0,4,0,4,0,2,0
研究,26,12,14,12,10,10,18,2,12,4,4,2,4,0,0,2,0,2,4,0,0,0,2,6,0,0,4,0,0,0,0,2,4,2,6,2,4,0,4,0,
死,28,14,4,10,6,6,6,12,10,0,4,0,0,0,4,2,2,0,0,2,0,0,2,0,0,0,0,4,6,0,0,2,4,0,0,0,0,2,0,0,0,0,2
一辈子,36,14,10,20,6,6,12,18,6,4,2,2,2,2,2,2,2,2,0,0,2,2,0,0,0,0,2,0,6,0,0,2,4,0,0,2,0,4,0,2,
中国,18,4,8,6,0,6,8,0,10,0,6,0,0,0,2,2,0,0,6,0,6,0,0,0,2,2,0,0,2,2,0,2,2,2,2,8,10,0,0,0,0,2,2
责,12,6,4,2,2,0,2,2,0,4,2,0,0,0,0,2,0,0,2,0,0,0,0,2,0,22,0,0,0,0,0,0,0,0,0,0,0,2,0,12,0,0,0,0
负,12,4,6,8,6,4,4,2,6,2,2,2,2,18,0,2,4,2,2,2,0,0,0,2,22,0,0,0,2,0,0,2,0,0,0,0,2,0,2,0,6,0,0,0,0
事,18,16,12,14,6,10,4,0,16,4,2,4,6,2,8,6,4,4,0,4,4,0,2,0,0,0,0,2,2,0,6,0,2,4,4,2,0,6,0,0,0,4,
结婚,34,26,10,18,2,14,4,2,4,2,2,8,0,0,2,10,0,8,0,2,0,4,0,0,0,0,2,0,0,0,0,0,8,0,0,2,0,4,0,0,2,
生命,34,6,0,18,2,6,4,6,4,6,4,0,0,2,10,0,4,0,2,2,0,6,6,2,0,2,2,0,0,0,2,0,6,0,0,0,0,0,0,2,2,0,2
抓,6,6,2,2,0,0,2,0,4,0,2,0,0,0,0,2,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,2,0,4,2,0,4,0,0,0,0,0,0,0,0,0,
科学,16,10,6,6,6,2,6,4,6,6,4,0,2,2,8,0,2,2,0,2,0,0,0,0,0,0,6,0,2,0,0,0,0,6,10,2,2,2,0,0,0,6,2
太,14,10,4,12,10,10,8,0,6,2,2,0,4,2,4,0,2,0,4,2,2,2,2,0,2,0,0,0,2,0,0,2,0,0,0,0,2,2,0,2,2,2
俩,22,36,12,2,6,6,4,6,8,6,0,2,0,2,4,8,2,2,0,6,4,4,4,2,0,0,2,8,6,0,0,2,0,2,2,2,4,2,0,0,0,0,0,0
底线,20,6,6,2,8,6,6,0,4,4,2,0,2,2,4,4,2,0,4,0,2,0,0,2,0,0,4,0,0,4,6,0,2,0,10,0,4,2,0,2,2,4,0,
伦理,18,12,4,10,8,6,4,2,6,2,0,0,4,0,4,2,2,0,2,2,6,0,0,2,0,0,4,0,0,2,10,0,2,10,0,4,2,2,0,0,0,4
```

共词矩阵 3

时间点 4：

4c.csv - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
0,基因,可怕,人类,婴儿,太,改变,严惩,科学,孩子,道德,做,两个,伦理,编辑,未来,发展,希
基因,0,2,4,8,2,4,1,1,0,0,1,0,0,6,3,1,0,2,0,0,2,2,1,0,0,1,1,1,2,3,2,1,0,1,0,0,0,(
可怕,2,0,2,2,7,2,2,0,1,1,1,1,0,0,0,0,0,0,0,0,1,2,0,0,0,1,0,0,1,2,0,1,1,0,0,0,(
人类,4,2,0,3,1,4,0,3,1,3,2,1,0,2,2,3,0,2,4,0,0,1,1,0,1,0,0,0,2,2,2,0,3,1,0,0,0,(
婴儿,8,2,3,0,2,3,0,1,0,0,0,3,0,5,4,1,0,2,0,0,0,2,0,0,0,1,3,0,2,2,2,1,0,0,0,0,0,(
太,2,7,1,2,0,1,2,1,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,2,0,0,0,0,0,1,0,(
改变,4,2,4,3,1,0,0,0,0,0,1,0,0,2,2,0,0,2,0,0,0,1,1,0,0,0,0,0,2,2,2,0,0,0,0,0,0,(
严惩,1,2,0,0,2,0,0,1,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,(
科学,1,0,3,1,1,0,1,0,0,5,0,0,2,1,1,4,3,0,3,2,1,1,0,0,0,1,0,0,0,0,0,1,3,0,0,2,0,(
孩子,0,1,1,0,0,0,0,0,0,1,2,3,1,1,1,0,1,0,1,0,0,0,2,0,1,0,1,0,0,0,0,1,1,1,0,0,1,(
道德,0,1,3,0,0,0,0,5,1,0,0,0,3,0,0,3,2,0,3,2,0,0,1,0,0,0,1,0,0,0,0,0,4,1,0,2,0,(
做,1,1,2,0,0,1,0,0,2,0,0,1,1,0,0,0,0,0,1,0,1,1,3,0,3,1,0,0,0,0,0,0,0,0,0,1,0,(
两个,0,1,1,3,1,0,0,0,3,0,1,0,0,0,1,0,0,0,1,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,(
伦理,0,0,0,0,0,0,0,2,1,3,1,0,0,0,0,0,2,0,1,3,0,0,1,0,0,0,0,0,0,0,0,0,0,0,3,1,
编辑,6,0,2,5,0,2,0,1,1,0,0,0,0,0,3,1,0,2,0,0,0,1,0,0,0,1,0,0,2,2,1,1,0,0,0,0,0,(
未来,3,0,2,4,0,2,0,1,1,0,0,1,0,3,0,1,0,2,0,0,0,1,0,0,0,1,0,0,2,2,1,0,0,0,0,0,0,(
发展,1,0,3,1,0,0,0,4,0,3,0,0,0,1,1,0,0,0,3,0,0,1,0,0,0,1,0,0,0,0,0,0,3,0,0,0,0,(
希望,0,0,0,0,0,0,2,3,1,2,0,0,2,0,0,0,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,2,0,(
新,2,0,2,2,0,2,0,0,0,0,0,0,0,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,1,0,0,0,0,0,0,0,(
违背,0,0,4,0,0,0,0,3,1,3,1,1,1,0,0,3,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,3,0,0,1,0,(
科研,0,0,0,0,1,0,0,2,0,2,0,0,3,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,2,(
说,2,0,0,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,0,0,0,0,0,0,(
国家,2,1,1,2,1,1,0,1,0,0,1,0,0,1,1,1,1,0,0,0,0,0,0,0,1,3,0,0,0,0,1,0,0,0,0,0,0,(
事情,1,2,1,0,0,1,0,0,2,1,3,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,1,0,0,1,(
违法,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,(
实验,0,0,1,0,0,0,0,0,1,0,3,1,0,0,0,0,0,0,1,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,(
技术,1,0,0,1,0,0,0,1,0,0,1,0,0,1,1,1,0,0,0,0,0,3,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,(
出生,1,1,0,3,0,0,0,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,1,0,0,0,(
赶紧,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,(
物种,2,0,2,2,0,2,0,0,0,0,0,0,0,0,2,2,0,0,2,0,0,0,0,0,0,0,0,0,0,0,2,1,0,0,0,0,0,0,(
变异,3,1,2,2,0,2,0,0,0,0,0,0,0,0,2,2,0,0,2,0,0,1,0,0,0,0,0,0,1,2,0,1,0,0,0,0,0,0,(
恐怖,2,2,2,2,2,2,0,0,0,0,0,0,0,1,1,0,0,1,0,0,0,1,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,(
懂,1,0,0,1,0,0,0,1,1,0,0,0,0,1,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,(
底线,0,1,3,0,0,0,0,3,1,4,0,0,0,0,0,3,0,0,3,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,(
这是,1,1,1,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,1,0,0,0,0,0,1,0,0,0,0,(
期待,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,(
```

共词矩阵 4

时间点 5：

5c.csv - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
0,孩子,婴儿,基因,人类,三个,编辑,事,太轻,三年,说,技术,贺建奎,判,保护,希望,几个
孩子,0,1,3,2,5,1,1,2,2,1,2,1,0,4,2,3,2,0,1,1,1,1,2,0,0,0,0,0,0,2,2,0,0,0,0
婴儿,1,0,4,2,4,4,1,0,0,0,0,0,0,1,1,1,1,1,0,0,1,0,1,0,3,0,0,1,0,0,1,0,2,1,1
基因,3,4,0,5,1,6,1,1,1,2,0,0,0,1,0,1,1,2,0,1,1,0,2,0,1,0,0,0,0,1,1,2,2,1,1
人类,2,2,5,0,0,3,1,1,1,2,2,1,0,0,1,0,1,1,1,3,1,0,1,0,1,0,0,1,1,1,1,2,1,1,1
三个,5,4,1,0,0,1,0,0,0,0,0,0,0,0,3,1,1,1,0,0,0,2,0,1,0,0,0,0,1,0,0,1,0,0,0,0
编辑,1,4,6,3,1,0,0,0,0,2,0,0,0,1,0,0,1,2,0,0,0,0,1,0,1,0,0,0,0,0,0,2,2,1,1
事,1,1,1,1,0,0,0,0,0,0,0,0,0,0,1,0,0,2,0,1,2,0,1,1,1,0,0,0,0,0,1,0,0,0,0
太轻,2,0,1,1,0,0,0,0,3,0,1,0,1,0,0,1,0,0,1,0,0,0,0,0,0,1,1,0,0,2,0,0,0,0,0
三年,2,0,1,1,0,0,0,3,0,0,2,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0
说,1,0,2,2,0,2,0,0,0,0,2,2,0,0,1,0,0,2,0,1,0,1,1,0,0,0,0,1,0,0,0,2,0,0,0
技术,2,0,0,2,0,0,0,1,2,2,0,2,0,0,1,0,0,0,1,1,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0
贺建奎,1,0,0,1,0,0,0,0,0,2,2,0,1,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0
判,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
保护,4,1,1,0,3,1,0,0,0,0,0,0,0,0,2,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0
希望,2,1,0,1,1,0,1,0,0,1,1,1,0,2,0,1,0,1,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0
几个,3,1,1,0,1,0,0,1,1,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0
估计,2,1,1,1,1,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0
这件,0,1,2,1,0,2,2,0,0,2,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,2,0,0,0
伦理,1,0,0,1,0,0,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0
做,1,0,1,3,0,0,1,0,0,1,1,1,0,0,1,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,1,1,0,0,0
想,1,1,1,1,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0,1,0,0,1,0,0,0,0
科学,1,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,1,0
可怕,2,1,2,1,1,1,1,0,0,1,0,0,0,1,0,0,0,0,0,1,1,0,0,0,0,0,0,1,0,0,1,0,0,0,0
👍,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
三名,0,3,1,1,0,1,1,0,0,0,0,0,0,0,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0
严惩不贷,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,
医生,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0
问,0,1,0,1,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0
有没有,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1
实在,2,0,1,1,0,0,0,2,2,0,1,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
健康,2,1,1,1,1,0,1,0,0,0,0,0,0,0,0,1,0,0,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0
事情,0,0,2,2,0,2,0,0,0,2,0,0,0,0,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
这三名,0,2,2,1,0,2,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0
搞,0,1,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1
非法,0,1,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,1,0
```

共词矩阵 5

时间点 6：

6c.csv - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
0,基因,婴儿,编辑,孩子,人类,三个,做,三年,说,判,太轻,真的,想,可怕
基因,0,96,206,48,87,40,22,3,16,8,5,18,24,15,32,13,18,14,11,16,1
婴儿,96,0,85,17,40,75,8,5,20,6,5,20,31,8,5,10,4,2,7,13,17,16,2,
编辑,206,85,0,32,56,22,20,5,9,8,4,11,16,8,29,7,18,13,10,14,14,2
孩子,48,17,32,0,26,90,20,9,12,8,9,22,11,12,8,14,6,3,10,4,17,5,2
人类,87,40,56,26,0,33,16,7,16,6,7,10,12,9,20,12,15,13,12,7,10,7
三个,40,75,22,90,33,0,12,9,11,7,7,15,14,11,2,9,6,2,6,3,15,4,2,2
做,22,8,20,20,16,12,0,5,8,3,7,6,7,7,12,8,7,2,8,2,4,2,3,3,3,3,9,
三年,3,5,5,9,7,9,5,0,2,12,8,3,1,4,2,4,3,2,1,1,3,6,4,3,3,3,0,0,1
说,16,20,9,12,16,11,8,2,0,2,2,5,5,1,4,6,4,0,2,1,5,3,3,3,5,6,2,1
判,8,6,8,8,6,7,3,12,2,0,36,2,1,3,2,4,3,0,2,6,0,1,0,2,1,3,0,0,2,
太轻,5,5,4,9,7,7,7,8,2,36,0,2,0,4,2,5,2,0,0,4,1,1,1,3,2,0,0,0,2
真的,18,20,11,22,10,15,6,3,5,2,2,0,4,9,3,10,4,3,5,3,11,3,2,7,3,
想,24,31,16,11,12,14,7,1,5,1,0,4,0,4,4,3,2,3,6,1,3,1,0,2,2,4,4,
可怕,15,8,8,12,9,11,7,4,1,3,4,9,4,0,3,21,4,4,0,0,1,1,2,4,3,1,2,
技术,32,5,29,8,20,2,12,2,4,2,2,3,4,3,0,2,6,5,3,2,8,3,7,6,4,3,1,
太,13,10,7,14,12,9,8,4,6,4,5,10,3,21,2,0,2,1,1,0,2,2,3,3,2,3,3,
伦理,18,4,18,6,15,6,7,3,4,3,2,4,2,4,6,2,0,12,4,1,7,2,9,1,3,5,3,
科学,14,2,13,3,13,2,2,2,0,0,0,3,3,4,5,1,12,0,6,1,3,3,8,3,3,3,1,
国家,11,7,10,10,12,6,8,1,2,2,0,5,6,0,3,1,4,6,0,2,6,3,2,4,2,1,4,
年,16,13,14,4,7,3,2,1,1,6,4,3,1,0,2,0,1,1,2,0,0,13,1,1,1,0,2,0,
希望,15,17,14,17,10,15,4,3,5,0,1,11,3,1,8,2,7,3,6,0,0,4,2,5,4,3
贺建奎,20,16,20,5,7,4,2,6,3,1,1,3,1,1,3,2,2,3,3,13,4,0,2,0,2,1,
底线,8,2,7,2,13,2,3,4,3,0,1,2,0,2,7,3,9,8,2,1,2,2,0,1,3,3,1,0,0
几个,13,15,7,16,7,2,3,3,3,2,3,7,2,4,6,3,1,3,4,1,5,0,1,0,8,2,1,2
未来,23,13,17,17,16,13,3,3,5,1,2,3,2,3,4,2,3,3,2,1,4,2,3,8,0,5,
基因库,24,17,13,8,46,15,3,3,6,3,0,8,4,1,3,3,5,3,1,0,3,1,3,2,5,0
```

共词矩阵 6

（3） 共词矩阵的可视化

使用 ucinet 进行可视化分析

将导出的 csv 格式共现矩阵转格式为 xls 格式，放入 ucinet 中进行分析

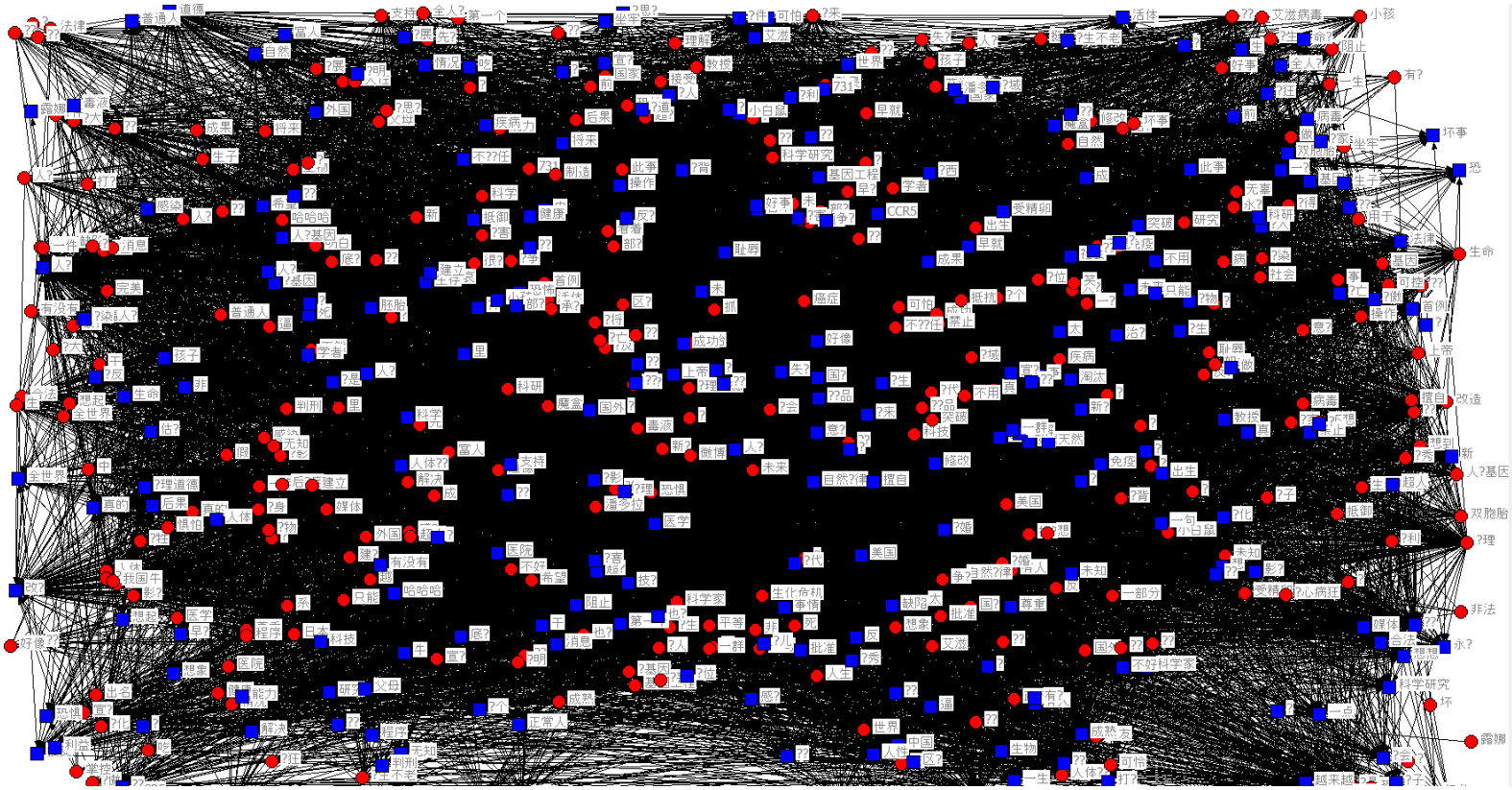
分析过程中发现各个时间点共现词对的频次差别较大，故又修改了几个时间点建立共现矩阵时词对准入门槛值。

此可视化网络主要讨论，当某一个高频词出现时，同时和它一同出现的会是哪个词。由此来体现当网民们讨论某一个话题时，常常会想到什么，会怎么认为。

上文所示的各个关联网络的词对准入门槛已尽量接近最优解，由于此值只能取整数，效果并不太好。

结果如下：

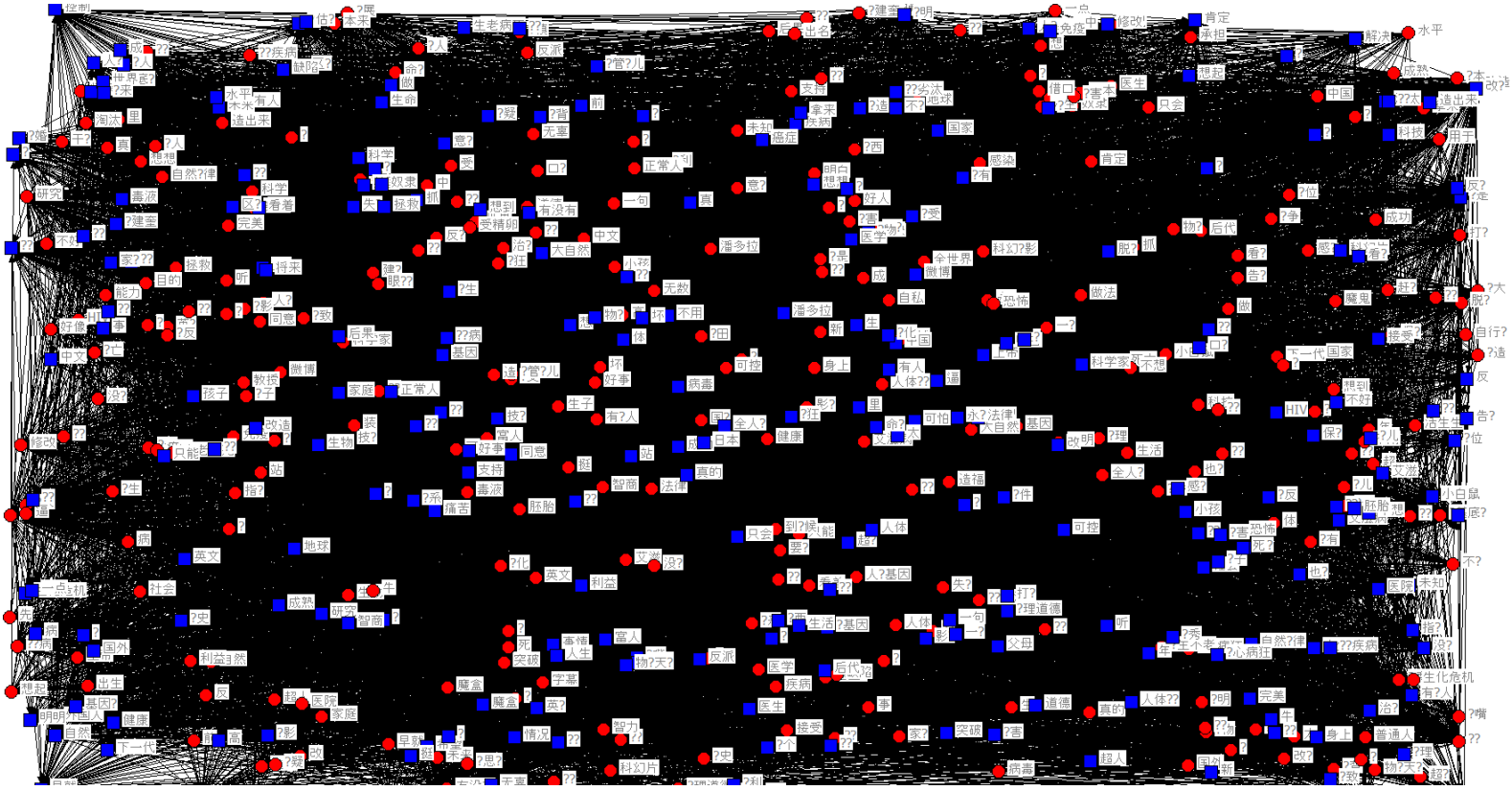
时间点 1（门槛值 8）：



共词可视化 1

时间点 1 位于中心位置的节点有科研、魔盒、未来、人类、失败、癌症、耻辱等等，在网络边缘可以发现如伦理、非法、生命、小孩等，这类可能更多论述基因编辑婴儿存在问题的词语反而位于边缘，可能与事件爆发迅猛，网民们的发言缺乏深度且充满情绪有关。

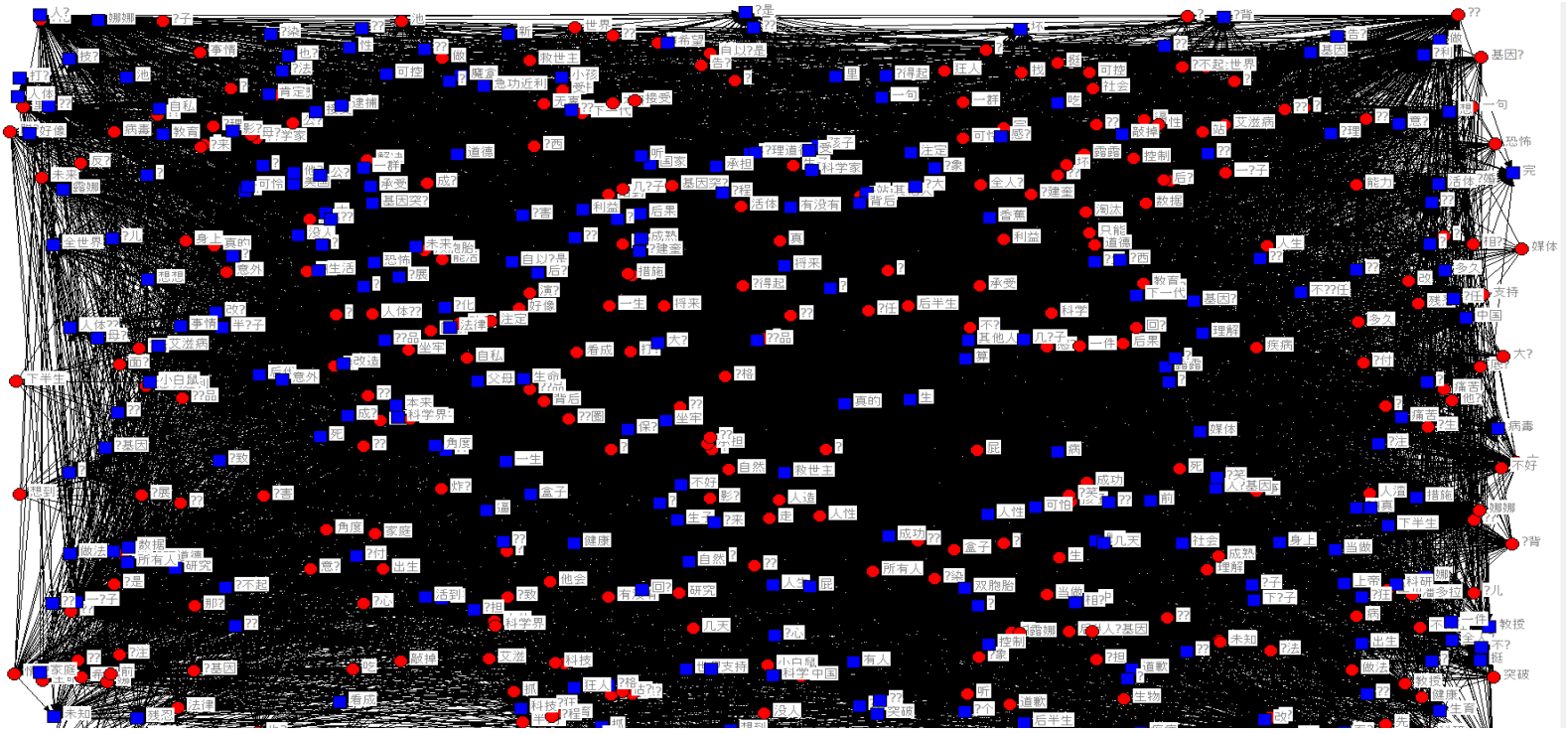
时间点 2（门槛值 13）：



共词可视化 2

时间点 2 位于中心位置的节点有病毒、可控、全人类、健康、可怕、大自然、艾滋、利益等，边缘词语有承担、生老病死、控制、修改、研究、超人等等。可以发现这一阶段的舆论更多地注目于基因编辑婴儿的未来意义，危险性和是否合乎自然规律等。

时间点 3（门槛值 8）：



共词可视化 3

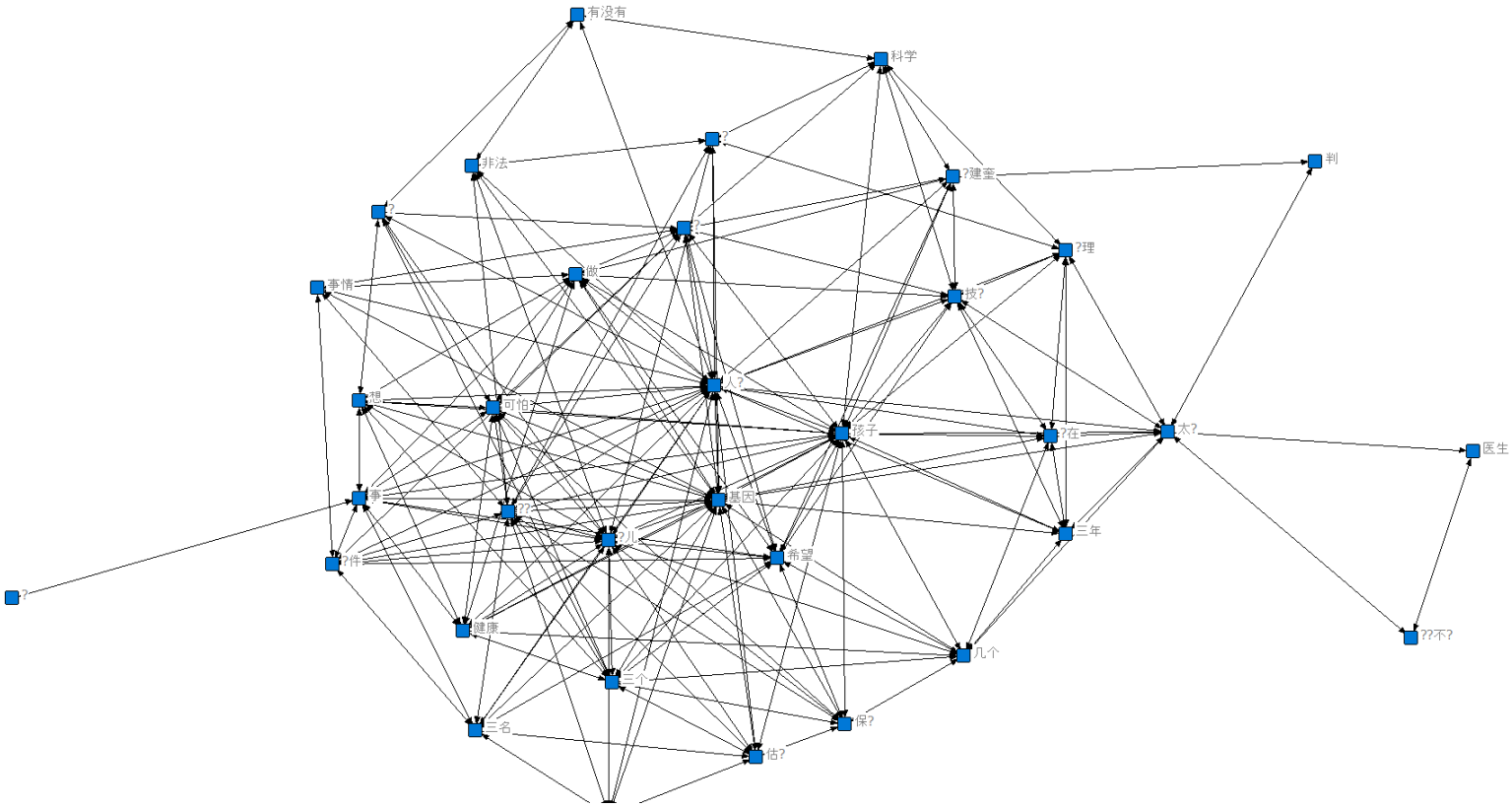
时间点 3 位于中心位置的节点有将来、坐牢、救世主、人造、后半生、承受、利益等词语。值得注意的是这个时间点贺健奎的名字比时间点 2 更加接近网络中心。考虑到这一时间点他现身港大，虽然道歉但向外界披露了项目的新进展，这一变化是符合常理的。网民的言论也更明显地关注贺健奎将来的处置，他发起这一项目的缘由等问题。

时间点 4（门槛值 3）：

共词可视化 4

时间点 4 数据量实在太少，缺乏说服力。

时间点 5（门槛值 3）：

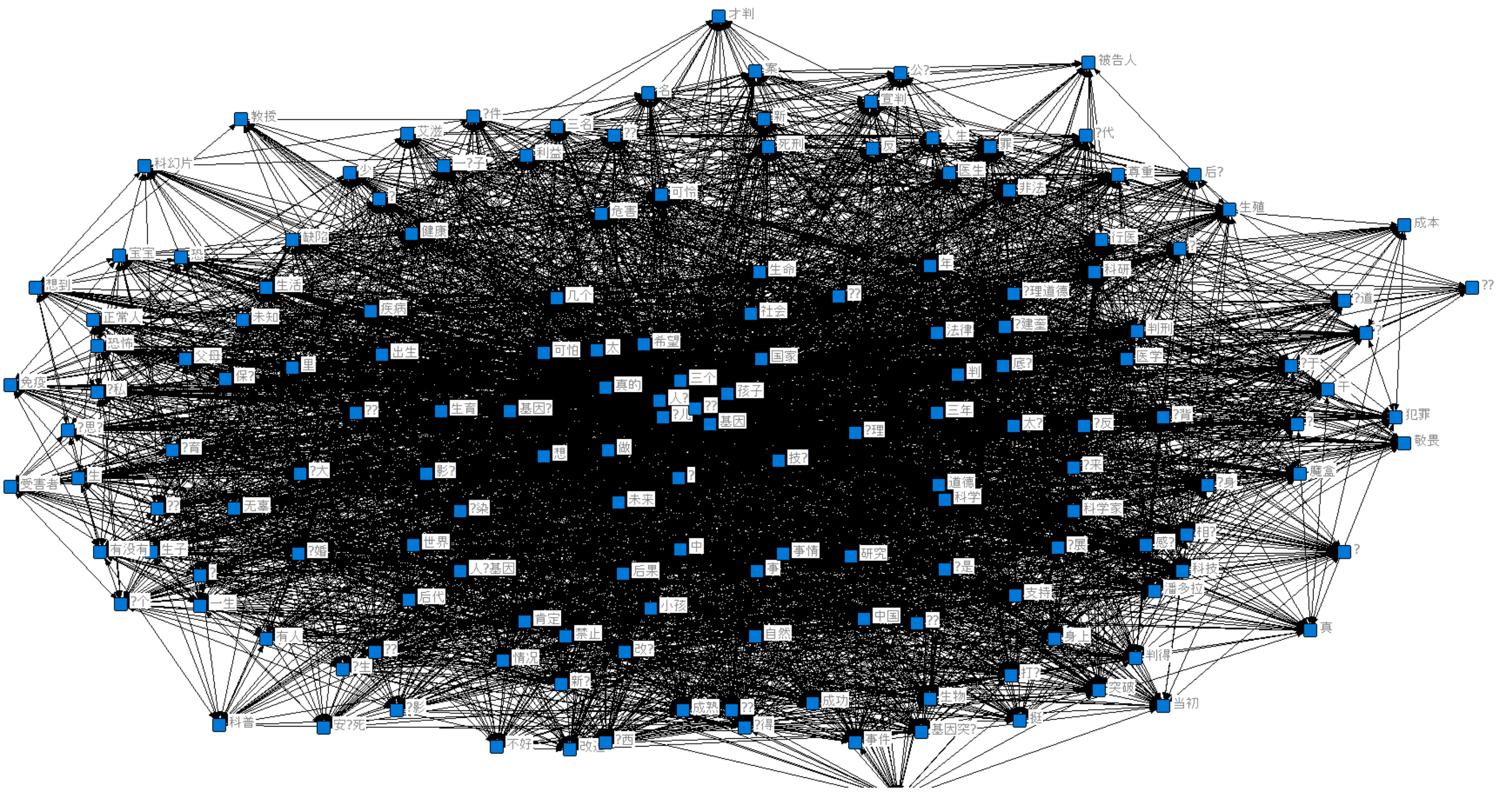


共词可视化 5

时间点 5，即一审宣判的时间。此网络的结构较为清晰，可以发现中心的节点主要为婴儿、孩子、基因、可怕、健康、三年等等。网民们主要关心宣判的结果和惩罚力度。贺健奎的名字与婴儿、基因、人类等相连，但是反而并未与官方

所说的伦理问题没有关系，考虑可能与网民这一阶段更在意预防今后类似事件再次发生，或者给此类事件定义有关。

时间点 6（门槛值 13）：



共词可视化 6

时间点6，贺健奎等三人被追究刑事责任后，出现在网络中心部分的节点有：孩子、希望、社会、国家、生命、道德、科学、可怜等，说明此时舆论更关心此事件的后续预防、相关法律法规发展完善，和此项目中诞生的婴儿今后该采取什么方式对待的问题。

第六章 综合课程设计总结

6.1 研究结果联系回应

联系政策回应情况

虑到数次官方回应内容并不太多，如果采用系统性的文本挖掘方法，在不利用人工智能和机器学习相关技术对语义进行深度挖掘的基础上，得到的结果并不好，投入产出比不高，故直接采用了人工总结的方法[9]。

关于法律部门、官媒、卫健委等等代表官方表态的发声口，主要有以下几次回应：

时间点 2，回应视频相关评论：

1. 南方科技大学表态：对于贺建奎副教授将基因编辑技术用于人体胚胎研究，生物系学术委员会认为其严重违背了学术伦理和学术规范。“我校将立即聘请权威专家成立独立委员会，进行深入调查，待调查之后公布相关信息。“

时间点 4，科技部与卫健委回应：

2. 国家科技部回应：该事件性质恶劣，科技部对此坚决反对，已全面暂停相关人员的科技活动，并将依据调查事实和事件定性，支持配合相关部门对涉事人员及机构依法依规进行严肃处理。同时将与有关部门一道，共同推动完善相关法律法规，健全包括生命科学在内的科研伦理审查制度。

3. 国家卫健委回应：该研究既违反中国目前的科研管理规则和伦理规范，同时也存在巨大的安全隐患。我们坚决反对这一研究，建议涉事单位、各级政府积极进行调查，采取必要手段厘清事实，对于违反法律法规的涉事人员予以严肃处理。

时间点 5，一审宣判：

4. 深圳市南山区人民法院一审公开宣判：3 名被告人因共同非法实施以生殖为目的的人类胚胎基因编辑和生殖医疗活动，构成非法行医罪，分别被依法追究刑事责任。在未取得医生执业资格的基础上，追名逐利，故意违反国家有关科研和医疗管理规定，逾越科研和医学伦理道德底线，贸然将基因编辑技术应用于人类辅助生殖医疗，扰乱医疗管理秩序，情节严重，其行为已构成非法行医罪。

可以看出官方的数次表态，均与“学术/科研/医学伦理“、”违反规范“等关键词有关，且处理方式在尚未定罪前以封禁查办为主，定罪后按一般司法程序处理。与科技舆情提及的”底线“、”伦理“、”利益“、”儿戏“等等词语有较大的关联度。但舆论时常关心的已出生婴儿的合法权利等，官方并无太多涉及，实在是遗憾。猜想可能是正常的放任处理，保证其正常合理的合法权益。但官方多次提及了借此机会对相关法律法规进行完善修订，反而是舆论涉及较少的。

6.2 综合课程设计的不足之处与改进方法

鉴于本研究采用的数据集来自中文微博，复杂性较大，在没有更新分词库或是停用词库等的基础上得到的结果必然与实际语义存在一定差异，且在有大量情绪化表达评论的基础上忽视情感和语义句式分析，这是未来可供进一步研究拓展的方向。

基于 LDA 的主题关键词提炼进行了多次的条件设置，最终得到了比较合适的筛选结果，但是仍然有着一定的优化空间。

同时，共现词可视化的原始数据，仍有进一步筛选过滤的可能，此次得到的各时间点结果并不十分理想，可能应考虑更换聚类方法。

参考文献

[1]. BLEI D M, JORDAN M I, et al. Latent Dirichlet allocation [J]. The Journal of Machine Learning Research,2003(3):993-1022.

[2]. MANN G S, MIMNO D, MCCALLUM A. Bibliometric impact measures leveraging topic analysis [C]//Digital Libraries, 2006. JCDI, Proceedings of the 6th AMC/IEEE-CS Joint Conference on IEEE, 2006: 65-74.

[3]. https://blog.csdn.net/leviopku/article/details/81388306

[4]. 黄扬, 李伟权. 新媒体环境下网络舆情对政策间断式变迁的影响研究——基于"间断-均衡理论"视角与案例分析[J]. 电子政务, 2018, 000(009):47-57.

[5]. 马静.基于扎根理论的中国网络舆情治理影响因素研究[J].河南社会科学,2018,26(11):103-107.

[6]. 黄扬,李伟权.新媒体环境下网络舆情对政策间断式变迁的影响研究——基于“间断-均衡理论”视角与案例分析[J].电子政务,2018(09):48-57.

[7]. 王倩倩.基于共词分析的国内在线商品评论研究热点探讨[J].现代情报,2017,37(10):158-164.

[8]. 李思雨. 微博热点话题情感计算技术研究与实现[D].河北科技大学,2015.

[9]. 万旋傲,谢耘耕.网络舆情传播对公共政策的影响研究[J].编辑之友,2015(08):5-11.

[10]. 包惠民,李智.CNKI 数据实现 ucinet 共现分析的方法及实证分析[J].软件导刊(教育技术),2012,11(01):91-93.