

新冠疫情研究可视化分析

刘金玲, 丁琪, 龚盛林, 马维炳, 梁焕城, 黎力溱

广西师范大学数学与统计学院, 广西 桂林

Email: 2871228767@qq.com, 2661981653@qq.com, 1835598458@qq.com, 1064856534@qq.com, 2990865512@qq.com, 2209040731@qq.com

收稿日期: 2021年5月21日; 录用日期: 2021年6月9日; 发布日期: 2021年6月24日

摘要

新冠疫情在我国各地爆发后, 关于“新冠疫情研究”主题的文献频繁出现在国内各类期刊上。为了探究新冠疫情研究的发展现状和趋势, 借助软件CiteSpace5.6.R5和中国知网的可视化知识图谱对该主题文献进行分析。通过对文献的发文作者、发文机构、关键词与主题、高被引文献以及热点演变进行分析, 对该主题研究进行阶段性概括描述, 旨在为后续研究者提供一定的参考资料。研究发现: 在应对突发公共卫生事件、疫情防控问题、经济复苏政策等方面, 仍然是新冠疫情研究未来需要发展的方向。

关键词

新冠疫情, 可视化分析, CiteSpace5.6.R5, 中国知网

Visualization Analysis of COVID-19 Research

Jinling Liu, Qi Ding, Shenglin Gong, Weibing Ma, Huancheng Liang, Liqin Li

School of Mathematics and Statistics, Guangxi Normal University, Guilin Guangxi

Email: 2871228767@qq.com, 2661981653@qq.com, 1835598458@qq.com, 1064856534@qq.com, 2990865512@qq.com, 2209040731@qq.com

Received: May 21st, 2021; accepted: Jun. 9th, 2021; published: Jun. 24th, 2021

Abstract

After the outbreak of COVID-19 across China, literature on the theme of “COVID-19 Research” has frequently appeared in variously domestic journals. In order to explore the developmental status and trend of COVID-19 research, the paper analyzed through the visual knowledge map of software CiteSpace5.6.R5 and CNKI. By analyzing the document author, publishing organization, key words and theme, highly cited literature and flashpoint evolution, the topic research is summarized, aiming to provide certain reference materials for subsequent researchers. The research has

found that in response to public health emergencies, epidemic prevention and control problems, economic recovery policy etc., it is still the future development direction of COVID-19 research.

Keywords

COVID-19, Visual Analysis, CiteSpace5.6.R5, China National Knowledge Infrastructure

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自新冠爆发以来，“新冠疫情”便成为了人们关注的社会热点。国家和政府纷纷出台了许多政策，为疫情的防控提供了保障，也构建了一个公开透明的疫情信息网络系统，让人们能够获得实时数据。在过去的一年里，不同领域的学者针对新冠疫情做了各方面的研究，在此之前，虽然也有学者对新冠疫情研究的可视化进行分析，但分析内容比较少，研究周期较短以及各领域对新冠疫情的研究没有形成系统性。本研究将依托可视化的研究方式，从发文学者、研究机构、关键词等方面制作知识图谱，通过分析，提出可行性建议，为未来新冠疫情研究提供一个更为科学的研究方向。

2. 数据的来源、筛选以及研究方法

2.1. 数据获取和整理

本文主要对主题为“新冠疫情研究”的文献进行可视化分析，为确保该研究分析的准确性及完整性，本文的数据来源于中国知网——具有广泛的数据库，并且实时更新，能够保障提取最新的数据。在中国知网中，以“新冠疫情研究”为主题，对该主题进行文献检索，并筛选出源于核心期刊的学术期刊文献，得到 1333 条数据，经过筛选，最终留下有效的数据有 1322 条。这些数据最早发表于 2020 年 1 月 1 日，最晚发表于 2020 年 12 月 15 日，故研究区间为 2020 年 1 月 1 日至 2020 年 12 月 15 日。

2.2. 研究方法

CiteSpace5.6.R5 引文空间是一款可以对发文学者、发文机构以及关键词等进行可视化分析的软件，主要提供共被引分析、共现分析、聚类分析等功能。CiteSpace5.6.R5 虽然提供共被引分析功能，但是却不支持对 CNKI 的文献进行共被引分析，为弥补 CiteSpace5.6.R5 的不足，因此本研究采用 CiteSpace5.6.R5 版本与 CNKI 计量可视化分析功能相结合的研究方法对主题进行研究。利用 CiteSpace5.6.R5 分析发文学者、发文机构以及关键词共现的网络图谱，同时利用中国知网的计量分析功能分析文献的共被引关系。通过可视化的手段呈现作者、机构、关键词、以及共被引关系的分布、结构和规律等，以此探索分析新冠疫情研究的变化趋势，研究热点与前沿话题的关系，为科学梳理我国新冠疫情研究进程、演化路径和探讨未来研究方向提供借鉴与参考[1]。

3. 研究结果和分析

3.1. 发文作者的可视化分析

对发文学者进行网络图谱分析，依据层次的差异性选取不一样的阈值，在 Node Types 中选择合作网

络分析中的 Author，得到发文作者聚类知识图谱如图 1。图中的节点数量与大小代表着核心作者群体共线的频次，节点越大表明该作者出现频率越高，反之则越低；线条多少与粗细则代表着作者间的合作强度与紧密性，连线越多越粗则表明作者间的联系程度越强[2]。



Figure 1. Clustering knowledge map of the authors of the new crown epidemic study from January 1, 2020 to December 15, 2020

图 1. 2020 年 1 月 1 日至 2020 年 12 月 15 日新冠疫情研究发文作者聚类知识图谱

从图 1 可以看出，新冠疫情研究发文作者聚类知识图谱中共有 79 个节点(N) 43 条连线(E)，共现密度 (density)为 0.014，大部分学者还处于分散状态，节点之间的连线较少，表明研究作者之间虽有共现关系，但相互合作程度有限，未形成有较大影响力的合作团体，成果共现度较弱。如表 1 统计显示，发文量较高的作者是段小丽、曹素珍、芦千文，各出现了 5 次，其次为方鹏骞、魏佳宁、樊胜根等，均出现 3 次及以上，发文量排名前 10 的作者一共有 25 篇文献，构成新冠疫情研究领域的重要研究力量。其中，段小丽、曹素珍、魏佳宁等作者出现了固定的合作关系，而来自中国社会科学院农村发展研究所的芦千文、华中科技大学的方鹏骞、中国农业大学樊胜根合作关系较为零散。表明目前学术界有关新冠疫情研究的作者基数较大，从整体上看相对分散，出现“小聚集”的研究趋势，这是因为各个团队之间受距离的影响，联系关系较弱。从研究团队上看，北京科技大学曹素珍 - 段小丽 - 魏佳宁组成的研究团队与郑州旅游职业学院、河南旅游研究院王璐 - 王少华 - 王梦茵组成的研究团队具有较强的专业性，发文数量高于其他学者，表明研究团队在组建方面具有专业性与高效性，但团队组建受地域等限制，综合性不强。对于孙一飞 - 朱杰 - 顾江 - 周恺 - 顾朝林组成的研究团队，团队成员来自不同高校或研究所，这一团队的出现，预示了在未来新冠疫情的研究中，学者间有倾向于不同地区不同领域间的交流合作的趋势，使新冠疫情的研究逐渐摆脱单一性、片面性，研究成果更具综合性、多元化，为人类抗击新冠疫情领域达到新的高度。

Table 1. Author of the top 10 publications from 1 January 2020 to 15 December 2020
表 1. 2020 年 1 月 1 日至 2020 年 12 月 15 日新冠疫情研究发文量排名前 10 的作者

序号	发文量	作者
1	5	段小丽
2	5	曹素珍
3	5	芦千文
4	4	方鹏骞
5	4	魏佳宁
6	3	樊胜根
7	3	金磊
8	3	王娇娥
9	3	吴丰昌
10	3	郑保卫

3.2. 合作机构分析

对发表在核心期刊上有关于国内新冠疫情研究文献的发文机构进行图谱分析, 得到发文机构网络图谱如图 2, 从发文机构的网络图谱中, 我们可以发现, 绝大部分的发文机构之间的连线相对比较少且稀疏, 有些机构仅为一个孤立点, 这意味着在“新冠疫情研究”这一领域, 国内大部分的研究机构之间的联系并不紧密, 但可以发现存在个别情况, 中国社会科学院、中国农业科学院、中国农业大学、中国人民大学之间有细线的连接, 彼此之间的联系较为密切, 形成了一个小网络。此外, 存在少数机构间有细线连接的现象, 但这些研究机构同属于某一个更大的研究机构, 如: 中国人民大学法学院与中国人民大学新闻学院均同属于中国人民大学, 这说明国内对新冠疫情的研究更加依赖于校内合作的研究模式。

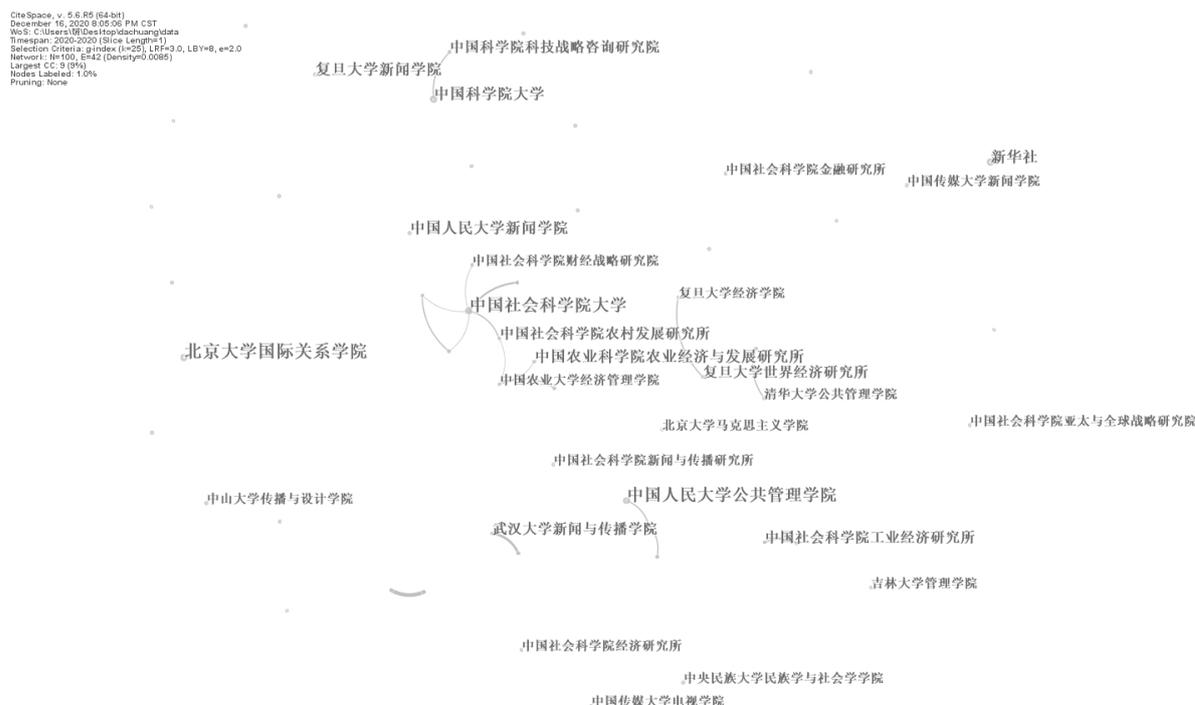


Figure 2. Mapping of the collaborative network of research institutions
图 2. 研究机构合作网络图谱

结合核心期刊中发文机构的网络图谱,对新冠疫情研究的相关文献进行数据统计如表 2,可以看到,在“新冠疫情研究”这一领域中,我国的主要研究力量包括:中国社会科学院、中国人民大学、中国科学院、复旦大学、武汉大学等,其中中国社会科学院在核心刊物上发文 61 篇,中国人民大学在核心期刊上发文 31 篇,中国科学院在核心期刊上发文 27 篇。这说明对新冠疫情进行研究的主力军多为研究所或 985、211 高校,且均分布在我国相对发达的地区;普通院校、欠发达地区的研究机构对新冠疫情的研究水平还有待提高。

Table 2. Number of communications from core agencies (TOP10)

表 2. 核心机构(TOP10)发文量

序号	发文量	发文机构
1	61	中国社会科学院
2	31	中国人民大学
3	27	中国科学院
4	22	复旦大学
5	20	武汉大学
6	16	清华大学
7	16	北京大学
8	12	四川大学
9	12	南京大学
10	11	吉林大学

3.3. 关键词和主题的研究

3.3.1. 关键词共现研究

关键词共现主要是运用文献的高频词汇对主题词语共同出现的情况,从而可以确定该文献所代表不同学科中各个不同的主题之间的联系,统计文献的主题词组在不同的文献之间出现的概率,由此形成的一个由这些主题词对关联词语所组成的共词网络图谱,利用软件 CiteSpace5.6.R5 对新冠疫情研究关键词进行关键词共现,在研究新冠疫情关键词的可视化图谱中,图谱里面的圈圈越大,并不是代表中心性的尺寸大小[3],而是表示该关键词出现的频次越多。其中,节点数就是图谱中关键词的个数,边数就是关键词之间的连线的数量,只要关键词在同一篇文章中出现,两者之间就会有一条连线。

根据图 3 可知,根据关键词共现的相对强度指标,现在对新冠疫情研究共现的主要是分为三大类,首先是:新冠肺炎疫情主要是围绕疫情发展情况的共现,其随之而来的有:全球化的发展、如何进行全球治理、人类命运共同体的发展、如何高效地进行复工复产等一系列的问题;其次是:新冠肺炎,这是围绕病毒本身的共现,其延伸出了:新型冠状病毒结构和性质等研究、如何应对突发公共卫生事件;最后是:疫情防控,主要是围绕如何对新冠肺炎疫情进行更科学地防控和把关的共现,其拓展出了对于疫情防控的应急措施和在重大疫情面前应该如何治理等措施。

3.3.2. 关键词聚类研究

运用 CiteSpace5.6.R5 软件对新冠疫情研究进行聚类可视化分析,得到关键词的可视化分析图谱。从图谱中可以得到聚类模块化 Q 值为 0.5338 大于 0.4,平均轮廓分数 S 值为 0.5937 大于 0.5,说明聚类的

结构层次是有效的，所利用对数极大似然率算法的方法置信度比较高[4]。如下图 4，我们发现显示了关键词动态的不断演化的过程，图中的每一个“#”号代表了一个聚类，顺序是从 0 到 8，数据越小，聚类中所包含的关键词越多，每个聚类是由几个紧密相关的词组成的，它们分别是：新冠肺炎疫情、新型冠状病毒肺炎、突发公共卫生事件、疫情防控、全球化、财政政策、防控措施、量化宽松、医学免疫等。



Figure 3. Common map of key words in the study of new crown epidemic

图 3. 新冠疫情研究关键词共现图谱

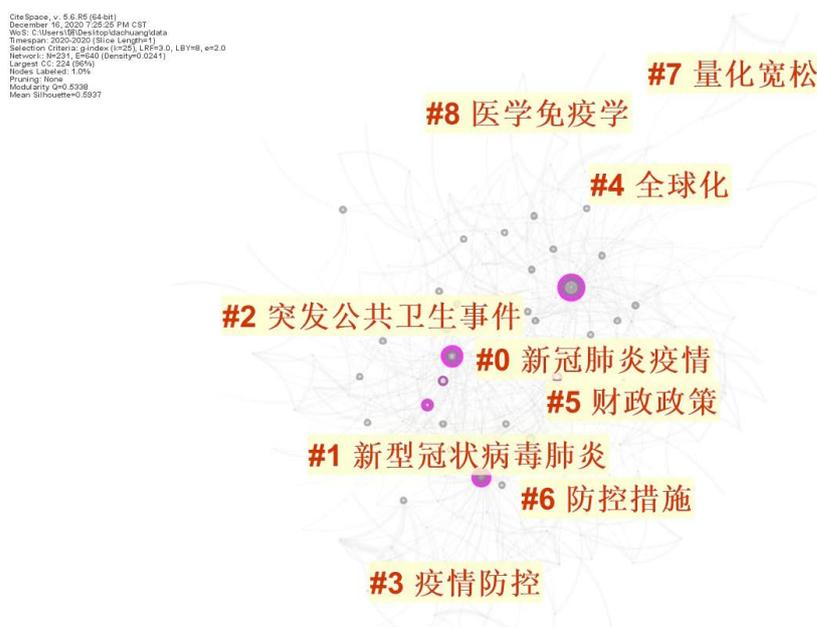


Figure 4. Clustering of Keyword in New Crown Epidemic Research

图 4. 新冠疫情研究关键词聚类图谱

“新冠肺炎疫情”成为该领域研究的最关键节点。第一类新冠肺炎疫情主要是对于疫情的发展情况进行实时的追踪,统计各地区每日疫情的确诊病例、疑似病例、境外输入、累计确诊、累计治愈、累计死亡、无症状等情况;第二类新型冠状病毒肺炎研究的是病毒的类型和变异、以及如何研制出病毒的抗体等;第三类突发公共卫生事件主要关注严重损害社会公众健康、安全的重大传染病疫情事件;第四类疫情防控的研究主要是对于疫情的蔓延,国家应该如何进行有效遏制疫情;第五类全球化主要是在全球化的背景之下,各国应该如何更好地应对这次疫情的爆发;第六类财政政策关注的是对于新冠疫情的爆发,财政的调控问题;第七类防控措施主要研究哪些措施既能尽快复工复产也能防止新冠疫情的再次爆发;第八类量化宽松主要关注中央银行实行较低的利率政策,降低储蓄量,提高人们的投资热情,加快资金的流动速度,促进经济高速平稳增长;第九类医学免疫主要研究利用人类的免疫系统产生抵抗疫情的抗体。从这九大聚类分析我们可以得到在新冠疫情研究这个领域中,最为关注的是对新冠疫情的实时情况进行追踪和分析。

3.4. 高被引论文

祝清松和冷伏海[5]认为高被引论文指在某个统计时间段内,被引用次数排在学科前列的论文,在一定程度上代表了学科的研究进展,具有重要的研究意义。因此选择从高被引论文的方向研究,有助于提高文献数据质量。

本文所采用的 1322 个文献中,论文被引用的次数比较难找到一个合适的断层点,因此需要寻找一个界定高被引论文的方法。在目前高被引用论文确定的几种方法中,本研究采用具有较强科学性和广泛性的百分比法[6]。刘雪立认为:可以利用文献计量学的“二八现象”,把被引用的论文按频次排序,以前 20%的论文作为高被引论文[7]。但是由于涉及新冠疫情研究还没到一个稳定期,处在前 20%位置的论文被引频次不高,故本研究采用前 15% (200 篇)的论文作为高被引论文。

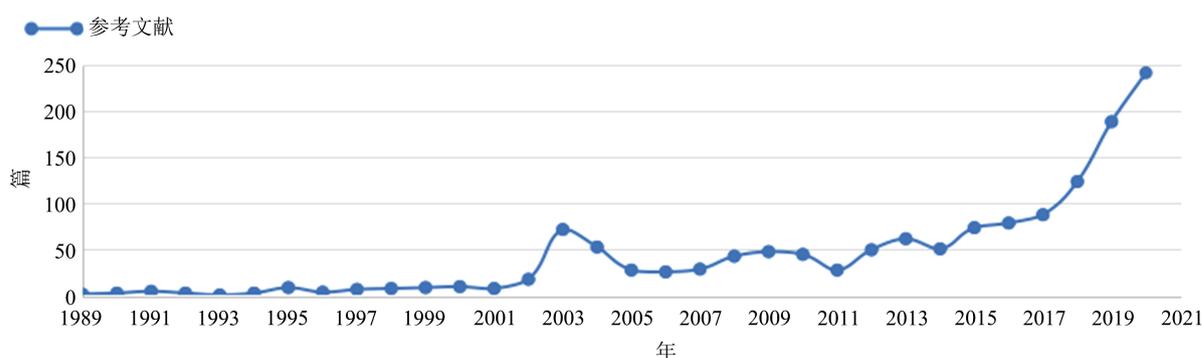


Figure 5. Distribution of references for high cited papers

图 5. 高被引论文参考文献的分布

由高被引论文作为研究文献数据,所得高被引论文参考文献数量随年份的分布如图 5 所示。高被引论文所参考的文献数量分布总体呈递增趋势,但是在 2003 年有一个突增点,原因是 2003 出现了传染性非典型肺炎,所以 2003 年存在比较多关于疫情的文献可以参考和借鉴。在 2018 年、2019 年、2020 年的参考文献增长幅度来看,这三年间高被引论文的参考文献逐期增长量大体上相等。通常来说,2020 年作为新冠疫出现的第一年,涉及新冠疫情研究主题的文献应该最多,但是逐期增长量和前两年的相差不大,这表明 2020 年的新冠疫情研究的进展似乎还没有到达成熟期。

从如下表 3 分析得到,作为高被引论文的两百篇文献的篇均参考数和篇均被引数相差不大。篇均被

引数为 10.5, 篇均参考数为 10.6, 表明为这 200 篇高被引论文中, 论文的被引和参考数刚刚满足一个内需的及格线; 再看下载被引比为 0.00251, 表 3 由于只保留两位小数所以显示为 0, 可以看出高被引论文总体影响不大或者质量相对不高。

Table 3. Indicators of high cited papers

表 3. 高被引论文指标

文献数	总参考数	总被引数	总下载数	篇均参考数	篇均被引数	篇均下载数	下载被引比
200	2120	2100	835,630	10.6	10.5	4178.15	0

3.5. 热点演变分析

基于中国知网 2020 年 1 月 1 日至 12 月 15 日主题为“新冠疫情研究”的学术期刊, 本次分析运用可视化软件 CiteSpace5.6.R5 挖掘文献的研究热点并探究热点的演变过程, 通过浏览此期间发布的文献的题目及关键词可以发现, 每一时间段人们研究的主题略显不同。因此, 为探究人们的研究热点的变化趋势, 我们把这一阶段分成初期(2020.1.1~2020.4.30)、中期(2020.5.1~2020.8.31)以及近期(2020.9.1~2020.12.15)三个阶段, 从初期的文献中选取发表时间为 2020.01.01~2020.02.15 的文献进行分析, 中期则选取发表时间为 2020.6.15~2020.6.25 的文献, 近期选择发表时间为 2020.11.20~2020.11.30 的文献。

3.5.1. 初期

疫情刚开始时, 面对武汉市市民被确诊为不明肺炎症状, 人们对此了解并不深, 武汉市政府甚至和民众说这种病毒没有人传人的危险, 但后来却发现越来越多的人出现类似的症状[8]。在此期间, 人们研究的热点主要是美元汇率、交通、不可抗力因素和线上线下教学等, 这是因为疫情在武汉爆发后不久便采取了封城的措施, 这一举措使人们无法随意走动, 导致贸易活动无法正常进行, 不利于经济增长, 进而影响汇率水平; 对于交通, 疫情在武汉严重起来之时, 多数社区在封城的基础上采取封社区的措施, 公交地铁停止运行; 疫情期间出于安全考虑, 大部分高校、中小学采取线上教学的措施, 部分疫情形势稍轻的地区则采取线上线下教学相结合的教学方法, 致使线上线下教学成为这一时期的重要特征之一。

如下图 6 为初期研究热点共现网络, 根据图中每一个圆的大小判断本期学者的研究热点, 圆越大表示其所代表的词出现的频率越高, 基于共线网络图我们发现这一时期学者关注的点主要是疫情防控, 基于疫情防控探究疫情对其他因素的影响等。从此图可以看出, 这一阶段各词组之间的联系密切, 圆与圆之间出现交叉的现象。

3.5.2. 中期

这一阶段, 中国的疫情形势渐趋稳定, 国外疫情形势则开始严峻起来。这一时期, 学者研究的热点可以分为以下几类: 中美关系、经济发展、新冠疫情、人类社会、疫情防控、受众等。疫情期间甚至美国疫情越来越严重之时, 美国仍不忘通过各种途径打击中国, 因此, 中美关系成为学者关注的焦点; 中国疫情渐趋稳定后, 各地掀起了复工复产的浪潮, 经济也在复工复产的浪潮下渐渐升温, 这时中国经济的发展成为学者关注的另一个焦点; 但是, 疫情形势渐趋稳定并不意味着我们可以高枕无忧, 通过国家卫健委公布的数据可以发现虽然中国的疫情得到了很好的控制, 但这一时期部分地区仍然有新增患者, 因此, 多数学者研究的方向依然是疫情防控。

如下图 7 为中期研究热点共现网络。从图中圆的大小可以看出, 这一时期学者的研究方向仍然是新冠疫情、疫情防控, 虽然疫情得到了一定程度的控制, 但我们不能掉以轻心, 依然要做好防护工作, 我想这应该也是学者以及国家所呼吁的。

上每年都有数以百万的大学生加入到求职人群中，找工作的人越来越多，人们的就业压力、工作压力越来越大；这次新冠肺炎疫情真实地反映出了人与自然之间的关系，人应与自然和谐相处，而不是为了自身的利益去伤害大自然，伤害大自然最后的结果可能伤害的却是人类自己；疫情期间，难免有人会现在的社会状况产生焦虑的心理，因此，也有一部分学者研究疫情期间人们的心理状况；因此，疫情背后隐含的原因是什么、如何促进大学生学习相关知识，成为学者探究的热点。

如下图 8 为近期研究热点共现网络，从此图可以看出这一时期词组与词组之间的联系较初期没有那么密切，圆与圆之间的重合部分较少。

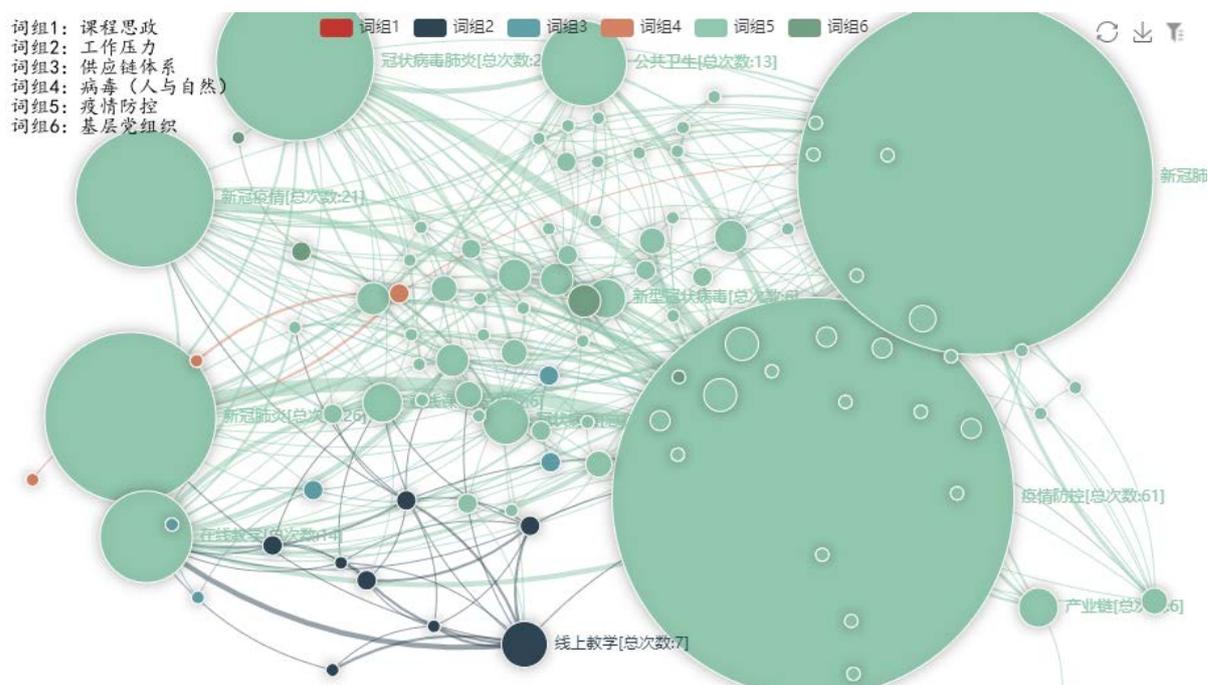


Figure 8. Recent hotspot co-operation network

图 8. 近期研究热点共现网络

4. 结论与讨论

在国内新冠疫情研究领域，研究机构之间、研究作者之间虽有一定的合作，但合作并不密切，未能建立起一个规模相对较大的合作网络，今后研究者们仍需打破校内小合作的格局，与交叉领域的学者一起交流、合作，以此促进新冠疫情研究领域的发展。

通过对新冠疫情研究的文献进行可视化分析，结果表明，新冠疫情研究的热点主要集中在应对突发公共卫生事件、疫情防控、经济复苏等方面。

此次疫情的爆发，给我们传递了一个不容忽视的信号——面对突发公共卫生事件，如何快速地进行科学、有效的处理？正所谓：构建科学的公共卫生应急管理制度体系，是增强应对突发公共卫生事件自信心的一剂“强心剂”。因此，各地医疗部门应总结经验，从疾病预防控制体系、公共卫生法治保障、应急物资保障体系、重大疫情防控救治体系和重大疾病医疗保险和救治制度等方面综合协调[9]，完善突发公共卫生应急管理体系。

本次新冠肺炎疫情研究中，学者对如何进行疫情防控的关注居高不下。在国内疫情防控常态化和国外疫情形势依旧严峻的时期，国内外政府和学者应持续关注疫情防控工作，完善疫情防控治理体系。只

有控制住疫情的蔓延趋势，找到防控和治疗方法，才能保障人民生产生活的正常运转。

正如前所述，各地医疗部门应总结经验，构建完善的公共卫生应急管理制度体系，在国内疫情防控常态化的基础上加强对疫情的研究，保证后续生产生活的正常运转。那么，如何在疫情防控常态化的基础上促进经济发展呢？

在致力于打赢疫情防控阻击战的同时，推出相关保障制度，着力推进经济发展。第一、刺激消费、增加内需。降低银行储蓄利率，刺激人们的购买欲望，稳定中等收入人群的数量，推出相关社会保障制度，增加居民收入，提升居民的消费能力；第二、政府宏观调控。政府继续实施积极的财政政策和货币政策，使民生等重要领域得到保障，此外，适当增加税率，使汇率国际收支基本平衡；第三、充分发挥社会主义制度集中力量办大事的优越性，提高自主创新能力。

本次研究基于 CiteSpace5.6.R5 对中国知网 2020 年 1 月 1 日至 12 月 15 日主题为“新冠疫情研究”的文献的发文作者、发文机构、关键词与主题、高被引文等进行可视化分析，挖掘学者的研究热点及趋势，为后续研究提供一些可供参考的方向。

基金项目

大学生创新创业训练计划项目资助(202010602210)，国家自然科学基金项目(11901124)；广西自然科学基金项目(2018GXNSFAA138164)；广西高校中青年教师基础能力提升项目(2021KY1798)。

参考文献

- [1] 胡志毅, 管陈雷, 杨天昊, 秦普艳, 陈艳. 中国旅游生态足迹研究可视化分析[J]. 生态学报, 2020, 40(2): 738-747.
- [2] 陈悦, 陈超美, 胡志刚, 王贤文. 引文空间分析原理与应用: CiteSpace 实用指南[M]. 北京: 科学出版社, 2014.
- [3] 董颖, 田剑. 基于 Citespace 的新零售研究可视化分析[J]. 商业经济研究, 2020(23): 13-16.
- [4] 杨永忠, 范亚南, 胡月明. 国内新型研发机构研究的文献计量与可视化分析[J]. 科技与创新, 2020(23): 20-22.
- [5] 祝清松, 冷伏海. 基于引文内容分析的高被引论文主题识别研究[J]. 中国图书馆学报, 2014, 40(1): 39-49.
- [6] Bornmann, L. (2014) How Are Excellent (Highly Cited) Papers Defined in Bibliometrics? A Quantitative Analysis of the Literature. *Research Evaluation*, 23, 166-173. <https://doi.org/10.1093/reseval/rvu002>
- [7] 刘雪立, 方红玲, 苗媛, 等. 五种综合性眼科学期刊论文下载量与被引量的关系及部分论文的量引背离现象[J]. 中国科技期刊, 2010, 21(5): 629-632.
- [8] 张敬苓. 重大公共卫生事件危机中政府治理能力困境及提升路——基于湖北省武汉市新冠肺炎疫情的观察[C]//福建商贸协会 2019 年座谈会论文集. 湖北大学政法与公共管理学院、湖北县域治理研究院, 2019.
- [9] 熊超. 科学构建防控体系是应对突发公共卫生事件关键[N]. 中国环境报, 2020-11-02.