“大数据”的文化建设战略价值：案例和实践

陈云松，严　飞，张　翼

[摘　要] 文章基于新近的社会科学大数据研究和地方治理实践，梳理出大数据对于国家和地方文化建设的六个方面战略价值：第一，展示地区软实力的“历史长卷”；第二，获得全球及地域的“比较视野”；第三，推进地区对外宣传的“有效传播”；第四，发现影响经济社会运行的“社会心态”；第五，为公共管理决策提供“科学依据”；第六，为经典理论提供“历史新证”。大数据能够以其超越传统调查数据的样本量和时空跨度，为研究者和管理者提供前所未有的海量数据、资料和信息，从而帮助研究者从过程性的历史视角来审视和验证重要理论问题，并为公共管理提供新的决策依据。无论是学界、智库还是党政部门的决策者，都应该进一步了解大数据，善于驾驭大数据，让大数据来助力人文研究和国家文化建设。

[关键词] 大数据；文化建设；国家战略；智库建设

[作者简介] 陈云松，南京大学社会学院副教授，南京市玄武区副区长，牛津大学社会学博士，江苏 南京 210046；严飞，清华大学社会学系助理教授，斯坦福大学 Walter H. Shorenstein 亚太研究中心博士后，北京 100084；张翼，中国社会科学院社会学所副所长，研究员，博士生导师，北京 100732

[中图分类号] G05　　[文献标识码] A　　[文章编号] 1004－4434(2016)06－0089－08
DOI:10.16524/j.45-1002.2016.06.019

大数据的出现，为社会科学学者研究现实生活中社会经济运行轨迹和文化行为、政治现象的互动提供了全新的前景。数年前，哈佛大学的加里·金在展望政治学的未来五十年时就预言，随着大数据的出现和使用，整个社会科学研究的实证基础将会出现重大的转化，研究问题的广度将伴随着海量数据的整合而得到极大拓展，甚至会加速定性与定量研究的大融合[1](P91-93)。

大数据主要是指在数量（Volume）、类型（Variety）、速度（Velocity）和价值（Value）等方面超过传统规模的海量数据资料，尤其是那些不是通过随机抽样方法所得到的调查数据[2]。对于文化建设而言，大数据的聚焦点主要是社会科学大数据，也即当代经济学、社会学、政治学、新闻和传播学等社会科学领域进行量化分析或可视化的海量资料[3]。从目前的来源看，这类大数据主要来自数量级以千亿词汇、万亿字节的数字化书籍、媒体、语料库、视频库、互联网文本、搜索引擎记录以及脸书、微博、微信等当代自媒体平台。不过，大数据概念虽热，但党委政府部门特别是地方基层组织在利用大数据为公共治理和执政党自身建设方面提供决策依据和参考的案例仍然不多。这是因为：一方面，大数据多为大型商业机构或专业机构持有，且往往各自为政，难以整合，现有大数据本身也并非为决策管理而专门设计。另外一方面，囿于传统工作对象和方法，决策者群体对大数据的了解、掌握和驾驭能力总体上仍需进一步提高。

基于新近的社会科学大数据研究和地方治理实践，我们梳理出“大数据”对于国家和地方文化建设的六个方面战略价值：（1）展示地区软实力的“历史长卷”；（2）获得全球及地域的“比较视野”；（3）推进地区对外宣传的“有效传播”；（4）发现影响经济社会运行的“社会心态”；（5）为公共管理决策提供“科学依据”；（6）提供经典理论的“历史新证”。

[基金项目] 江苏高校哲学社会科学研究重点项目“中国传统文化的全球知名度”（2015ZDIXM001）；江苏省社会科学基金重点项目“大数据视野中的江苏文脉研究”（15ZHA001）

一、利用大数据展示软实力的“历史长卷”

利用大数据来进行人文社科领域的量化研究，已经逐渐成为相关学科的前沿和热门领域。利用大数据独有的跨时空的宏大特征，可以深度挖掘潜藏在海量数据背后有意义的历史规律或信息，从而能够展示出以往研究中所无法呈现的超长时间、超大空间的历史画卷，让研究成果具有大视野、大跨度的特征，具备高度的科学性和说服力。我们以中国的“文化国际影响力”作说明。

党的十七大报告提出：“要加强对外文化交流，吸收各国优秀成果，增强中华文化国际影响力”；党的十八大报告更从建设社会主义文化强国的高度明确提出：“要开创全民族文化创造活力持续迸发、社会文化生活更加丰富多彩、人民基本文化权益得到更好保障、人民思想道德素质和科学文化素质全面提高、中华文化国际影响力不断增强的新局面。”因此，了解、分析和研究中国文化的国际知名度、影响力，是国家层面软实力研究的应有之义。但“文化国际影响力”难以衡量，而人们于“文化国际影响力”的评判分析绝大多数出于感性认识。例如，透过好莱坞电影的滥觞，人们感受到美式文化在全球的强力扩张；透过兵马俑在大英博物馆展陈的轰动，人们感受到中华文化在全球吸引力的不断增强。但这种当代影响力、知名度大到什么程度，处于何等地位，我们无法测量和比较。而要了解数十年、数百年来中国文化对全球影响力的历史轨迹，更是无法实现。不过，大数据已经开始改变这一困局，值得我们高度关注。

大数据之“大”，使得数据的性质发生了显著变化，其数据的获取和分析，往往需要有别于传统社会科学训练方法和工具。从 2004 年起，谷歌公司就开始对全球顶尖大学图书馆藏书进行数字化，目前已经数字化了 3 千万种，时间跨度从公元 1500 年到 2000 年，含 7 种语言，占古登堡印刷术以来全球书籍的 4%，词汇量达 5,380 亿[①]。2011 年，相关专家团队利用这一大数据，在《科学》杂志发表了《使用百万数字化书籍的文化定量分析》一文。这项研究以文化关键词在百万书籍中的历史频率变化，展示了 500 年来人类文化发展史中的重要趋势和现象[4]。利用这一技术，我们可以对近代人类知识总体中任何一个关键词的出现频率进行分析，也就为我们提供了分析、展示和比较“软实力”“文化国际影响力”的重要工具[②]。以国际知名度为例，利用 300 年来全球书籍中对某个人物、某个文化遗产、某个城市、某个省份或者某个国家的提及次数、频率等指标，我们就可以科学测量其相应的国际文化影响力。

学者还利用相关大数据对中国地级以上近 300 个城市的国际知名度、中国世界文化遗产、文化名人名著的国际影响力等进行了分析[5]。下面举两个案例以供参考。图 1 展示了近 300 年来中国城市的国际知名度前 20 强。经过测算，北京、香港、上海、广州、南京、澳门、天津、台北、重庆和拉萨稳居中国城市文化知名度的前 10 名，而南京无论是在哪个历史阶段，都位居全国前 5 名。利用大数据，笔者进一步总结了城市知名度形成的 5 个特征，并分析了内陆城市和殖民地城市知名度形成的不同模式。图 2 展示的是近 500 年来中国 34 处世界物质文化遗产的前 10 强。从图 2 中可见，长城、故宫、丝

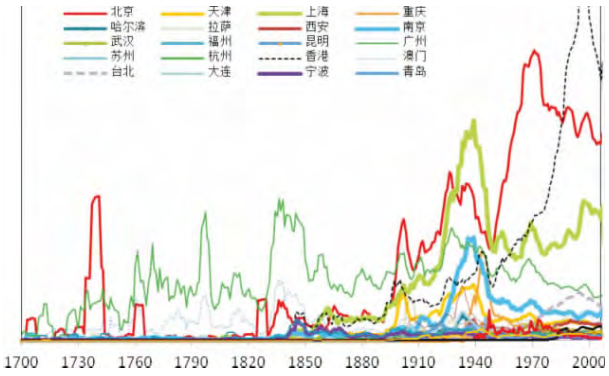


图 1　南京和近 300 年中国城市的国际知名度 20 强

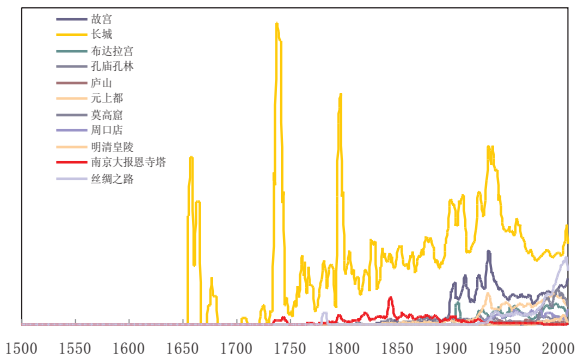


图 2　南京报恩寺塔和近 500 年中国世界物质文化遗产的国际知名度前 10 强

①Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman-Aiden, Jon Orwant, Will Brockman, and Slav Petrov. “Syntactic annotations for the Google Books Ngram Corpus”Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012.

②Acerbi, Alberto, Vasileios Lampos, Philip Garnett, and R. Alexander Bentley. “The expression of emotions in 20th century books”PLoS ONE, 2013, 8 (3); Hassanpour, Navid. “Tracking the semantics of politics: A case for online data research in political science”Political Science & Politics, 2013, 46(2).

绸之路、元上都遗址等的全球知名度名列前茅。而南京大报恩寺塔尽管因为在 19 世纪就已经焚毁而没有列入[6]，但其在 1700–1900 年之间的全球知名度相当高，甚至仅次于长城。

总体上，大数据之所以能为我们提供软实力的历史长卷，除了因为时间空间跨度大，更因其具有高度的代表性：第一，书面语言本身是承载人类观念、意识和价值观的最重要载体；第二，海量书籍报刊的语言词汇既反映作者、撰稿人的个人观点，更能反映和捕捉社会大众的整体思潮。从这个角度出发，通过数个世纪积累而成的海量英语书籍和报刊，实际上构成了国际社会知识、观念和经验的主流，只要语料库具有足够的代表性，我们就可以认为一个词汇在其中的出现频率能够近似地反映这个词汇本身及其意蕴的社会文化影响力和知名度，甚至折射出某种社会趋势、风尚或思潮[7]。而这类方法，显然可以扩展使用到人文社科领域的各个分支。同时，或许更为紧迫的，我国也应该抓紧对中文书籍进行类似的数字化建设，以在大数据领域获得重要的话语权（目前谷歌图书的中文数字化书籍仅有 30 万种）。

二、利用大数据获得国际国内“比较视野”

比较研究一直是人文社科领域重要的研究切入点，比较视野依托于“可比”的指标，而人文社科领域有很多用传统研究方法无法进行有效比较的议题。前文我们提及了城市、文化遗产等的国际知名度历史长卷，下面进一步以中国“长城”和埃及“金字塔”的比较来作直观的说明。我们都知道，这两处著名的世界物质文化遗产，分别是中国和埃及两个文明古国最为重要、最具有代表性、也是最具国际知名度和文化影响力的文明标志性建筑。那么，和埃及金字塔相比，长城的国际知名度究竟如何呢？以往的历史研究，无法直接回答这样的问题，而靠历史、考古等学者求诸于史料，靠当代社会学者、传播学者进行问卷调查，都是无法完成的任务。我们通过对大数据的分析，却能在很短时间使用很少人力物力获得一个全景的比较分析视野，从而回答这个问题。在图 3 中，我们同样利用谷歌百万书籍的语料库对长城和大金字塔的相关词频进行了可视化分析。不难发现，近 500 年来，埃及大金字塔的国际知名度（双实线）总体上不如长城（虚线），但在 1850–1890 年之间它超过了长城，究其原因，是和 19 世纪末金字塔相关考古进程息息相关。而对相关埃及考古学发展的关键词曲线，和我们的结论吻合得很好。

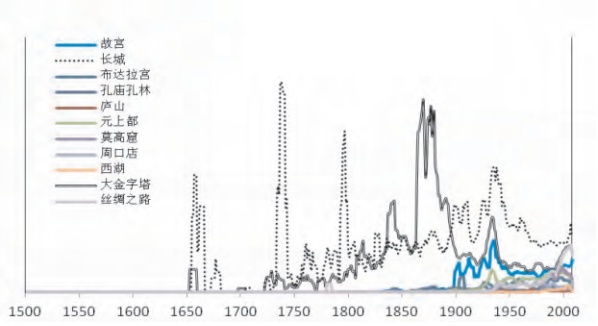


图 3　“大金字塔”和“长城”等中国世界物质文化遗产对比

我们再以江苏为例进行省份内的比较分析。图 4 展示的是江苏 13 个省辖市改革开放以来全球中文书籍的提及率（图 4）和英语书籍的提及率（图 5）。这些提及率，如我们前文所说，实际上反映的就是城市的总体国际国内知名度。有意思的是，如果我们把国际国内知名度进行对比分析可以发现，这些城市的国际国内知名度并不一定是完全对应的。例如，在国内知名度方面，南京、苏州、无锡的影响力数十年来一直领先，而其余城市竞争激烈，差距不大。扬州的国内知名度，在 2005 年之前落后于徐州、南通等城市，但 2005 年之后开始攀升并拉开与其他城市的差距，保持在地级市的第 4 位。而在国际知名度方面，南京遥遥领先，扬州在 2005 年左右开始超越无锡，到目前已经仅次于南京和苏州。

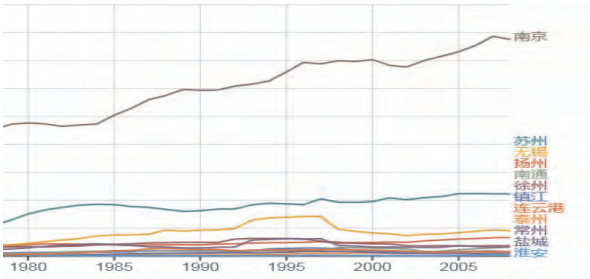


图 4　江苏城市的国内知名度

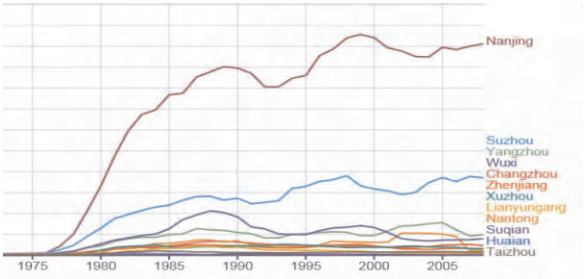


图 5　江苏城市的国际知名度

如果说数十年、数百年大跨度比较分析只是大数据中的“大历史”，利用同样来自书籍、平面媒

体以及国际互联网的大数据，我们还能够对相关的软实力、关注度、知名度、美誉度等指标进行当代比较研究，获得一个供决策者参考、分析比较的“当代大视野”，并有很多饶有兴味的发现。例如，我们发现，在英语世界中，网络搜索中国城市香港要远远高于大陆城市。此外，对上海的搜索甚至高于北京，这显然和我们的一般预期所有不同，值得进一步研究。图 6 展示了 2010–2015 年 5 年间 5 大中国城市的英语网络搜索对比，可以明显地看出层次性。图 7 则展示了南京的当代国际关注度地位：低于广州深圳拉萨，和西安非常接近。

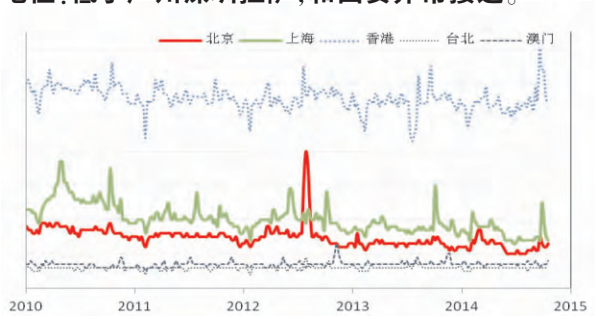


图 6　当代中国城市的国际互联网关注度前 5 强（2010–2015）

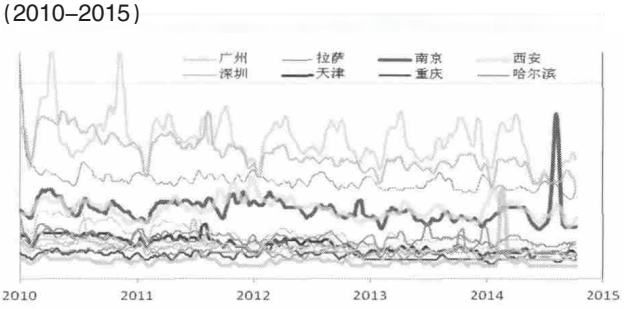


图 7　南京和其他主要大城市的国际关注度对比（2010–2015）

三、利用大数据推进对外宣传的“有效传播”

一个国家的国家形象在国际间的传播，一个地区的对外交流和宣传，是提升本国、本地区软实力和塑造国际品牌的重要手段。对于国际传播而言，要善于用国际受众听得懂的语言来说话，要善于讲中国故事。如果用受众无法接受、听不懂或者不熟悉的话语来传达我们的信息，则会事倍功半。而利用当代社会科学方法尤其是大数据来发现传播中的规律、症结，则有助于提高传播的“有效度”。我们以“南京大屠杀”和“南京大报恩寺塔”为例来分别对国家外交和城市传播作说明。

利用英语百万书籍大数据，我们对“南京大屠杀”（Nanjing/Nanking Massacre）、“犹太人大屠杀”（Holocaust）、“南京”（Nanjing/Nanking）、“奥斯维辛”（Auschwitz）等关键词的出现频率进行了分析。从图 8 中我们有两个发现：第一，南京大屠杀的国际知晓度仍然不高。和奥斯维辛、犹太人大屠杀两个词相比，半个世纪以来的英语书籍里基本没有提及。第二，犹太人大屠杀（Holocaust）这个词在 20 世纪 70 年代前默默无闻，但 70 年代末开始迅速为世界所知。也就在 70 年代末，奥斯维辛小镇的国际知名度开始超越南京。南京的规模远远大于奥斯维辛，且同样发生了骇人听闻的大屠杀事件，何以 70 年代末之后奥斯维辛的“犹太人大屠杀”和“奥斯维辛”的知名度急剧上升？我们提供的大数据解释就是文化传播的效度。而这种有效传播的载体，本身一点也不复杂。1978 年，当时尚未成名的美国著名影星梅尔·斯特里普（Meryl Streep）主演了一部反映纳粹迫害犹太人的 5 集电视短剧“Holocaust”。此片开始被西德拒播，但事件引起轰动，在美国以及西德播出后收视率非常高，这使得原意是“种族大屠杀”的单词“Holocaust”从此专指“犹太人大屠杀”。而“奥斯维辛”（Auschwitz）的词频曲线也随着大屠杀词频同步增长。

实际上，我们甚至发现，梅尔·斯特里普本人的知名度曲线和“犹太人大屠杀”词频曲线具有非常接近的特征：1978 年是增长速度的分水岭。相关的计量分析也表明了两者之间的关联。这一基于大数据的分析告诉我们，影视剧的跨国文化传播对于历史事件知晓度具有重要的意义。从公共外交的角度，影视传播既可以回避外交矛盾，又传播了特定的价值观，甚至可以形成舆论热点被全球关注。这个例子告诉我们，利用大数据对传播的历史细节进行描述和分析，有利于我们了解过去的不足，提高对外传播的有效性。

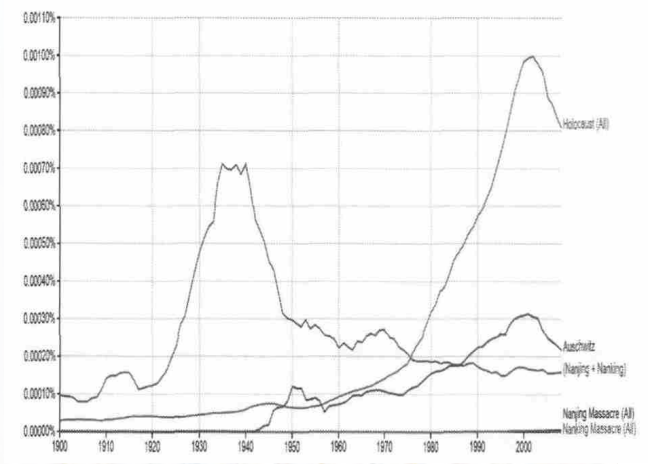


图 8　南京大屠杀、犹太大屠杀、南京和奥斯维辛的百年词频

同样，一个地区的对外传播也是如此。以南京为例，当我们向国际友人介绍南京时，较多地将南京介绍为“十朝古都”，并通过对“中山陵”“总统府”等我们心目中南京最具有代表性的建筑来加以推介。但实际上，除了少数中国通，国际受众对于中国历史、朝代等文化背景了解极少，大多数甚至不知孙中山为何人。因此，几十年来对南京这类城市的外宣内容和方法，往往忽视了国内和国际受众的知识结构差异。而通过大数据分析，我们发现，在国际受众中，中国瓷塔（Porcelain Tower），也即南京大报恩寺塔，是更多老外熟知的概念，安徒生童话里甚至都有提及。前文也已经提及，如果把大报恩寺的300年历史知名度曲线放置到中国世界物质文化遗产之中，在1700–1900年之间，它仅次于长城，比故宫还要知名。

为进一步说明大报恩寺塔让人吃惊的国际知名度，我们用图9来展示南京大报恩寺和西湖雷峰塔的国际知名度曲线。不难发现，在近百年来的英语书籍中，大报恩寺塔的提及率远远高于雷峰塔，两者几乎不在一个数量级上。而这几乎是让每一个初次见到这个图表的中国人十分讶异的结果：中国人更多地知道雷峰塔，而不知道南京大报恩寺塔。因此，这个例子同样告诉我们，利用大数据，了解大数据，可以更好地了解历史，了解中西差别，会让城市、历史的对外传播具有更高的效度。

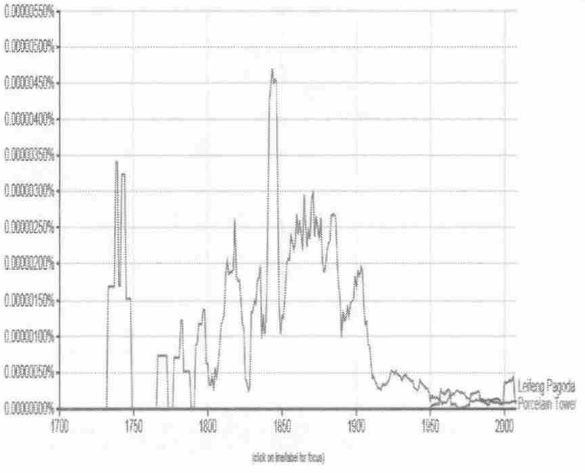
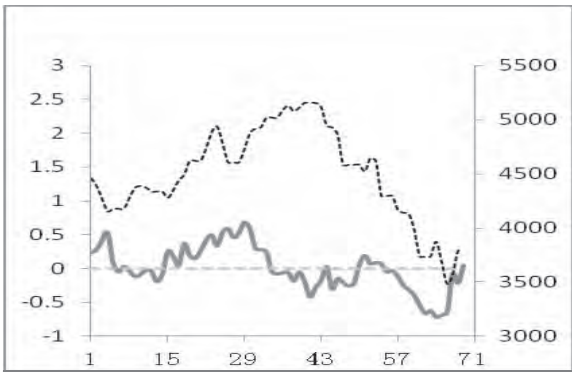


图9　南京大报恩寺塔与西湖雷峰塔的国际知名度对比（1700–2000）

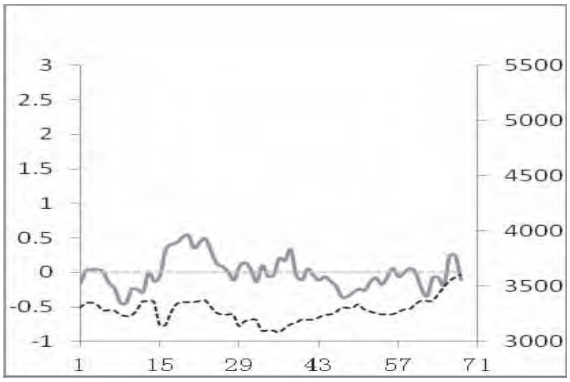
四、利用大数据发现影响经济社会运行的“社会心态”

大数据一直是舆情分析的强有力工具。基于大数据的舆情分析不仅仅可以用来监测、记录社会心态和舆情，更可以发现舆情和经济社会运行之间的重要关联。这是因为，舆情不仅仅被社会经济现象所催生，也会反作用于社会经济现象本身[8]。图10和图11展示的是2015年5–7月份、2–4月份的沪市K线（虚线）与基于省内新浪微博大数据的舆情信心指数曲线（实线）。具体而言，我们利用新浪微博中大量和股市有关的词汇，计算生成了代表舆情信心的指数。在代表股市震荡期的图10（5–7月）中，直观的视觉判断就能让我们大致感觉到两条曲线之间有明显的对应关系。实际上，时间序列分析发现，舆情信心可以用来预测三天内的股市波动行情。而这种关联，在股市平稳期并不存在（2–3月，图11）。这项研究的意义在于，新浪微博中每一天的股市信心指数代表了当日的全社会舆情，而不是哪一个网站、论坛或者某个地区、群体的舆情。因此，研究结果有力验证了社会舆论对于现实经济金融现象的影响。在覆盖面和科学精度上，这是传统的问卷调查所无法企及的。



2015年6–7月股市震荡期

图10　震荡期舆情与股市曲线



2015年2–3月股市平稳期

图11　平稳期舆情与股市曲线

再如，近年来网络出现了一些“尼玛”“逼格”“小婊砸”等粗鄙词汇，也让全社会产生了对网络文化、网络热词的担忧，甚至有学者呼吁国家对此类词汇进行干预。为进行分析，笔者搜集了2013年以来300多个网络热词在新浪微博省内出现的频率。通过对半衰期、半生期、峰值、谷值等的分析，大数

据研究发现，频繁出现在网络汇总的粗鄙热词占总体新生用语热词的比例仅仅在 6%左右，而且此类词汇的生命周期各有千秋，大多数在短暂的 1~2 个月高峰期后会下降、平稳甚至消失，很少能够成为真正的主流用语。图 12 展示了 20 个比较常见的粗鄙网络热词的 3 年使用曲线，除了“然并卵”“碧池”“碧莲”和“心机婊”等 4 个近半年内出现的新热词在大幅增长之外，其他的都呈稳定波动和下降趋势。实际上，在 3 年的尺度中，有的词汇实际上是昙花一现。我们还测试了相关的粗鄙词汇使用率汇总成单维曲线，发现其趋势总体上稳定并小幅波动，而并不呈上升状态①。

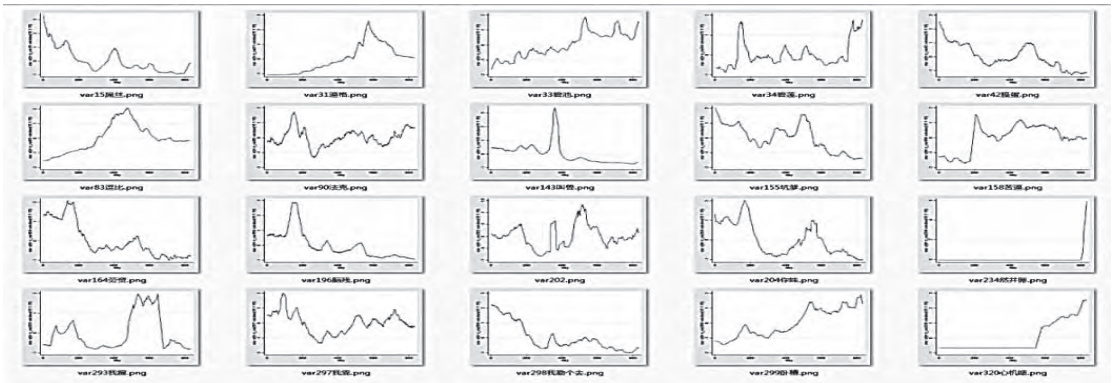


图 12　低俗网络词汇的三年趋势(2013–2015)

五、利用大数据为文保等公共管理决策提供“科学依据”

文物和文化遗址保护开放利用等公共管理过程，是城市文化建设的重要一环[9]，亦能从大数据中获取帮助。以南京市玄武区的实践为例，每年除夕，主城区内主要寺庙如鸡鸣寺、毗卢寺等都是人流密集地区，但香客数量的起伏并无一定规律，这给相应的警力配备提出了挑战。2015 年初，该区对除夕夜鸡鸣寺的警力配备进行了讨论分析，并利用大数据对比，估算出除夕夜香客数量将会是去年的 1.5 倍。主要的测算依据是，通过对除夕前数日内的微博中省内群众对“鸡鸣寺”“鸡鸣寺头香”等关键词的提及数量，以及在百度中对相关关键词的搜索数量。这是因为，除夕夜前几日，如果人们在网上提及或者搜素相关寺庙越多，则就表明有越多的人会来鸡鸣寺烧头香或者参观。

图 13 显示了微博中对鸡鸣寺相关词汇的提及次数，图 14 则是百度搜索中的检索次数。总体上，2015 年初的相关关键词数量是一年前的 1.5 倍多。因此可以类推，人流量也将达到 1.5 倍以上。因此，区政府配备了 500 人的各类警力和安保人员当晚进驻寺庙。结果除夕夜人流果然巨大，明显超过去年，事先的大数据分析，为公共决策提供了依据。

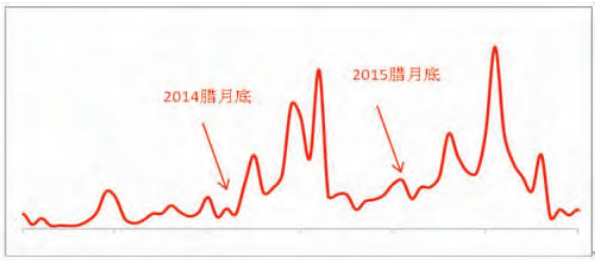


图 13　2015 年除夕前鸡鸣寺在微博提及数量

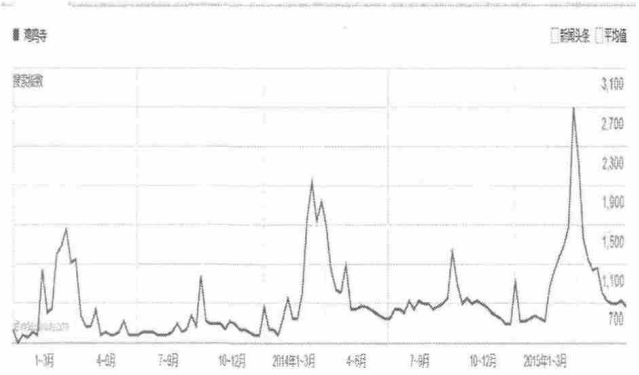


图 14　2015 年除夕前鸡鸣寺在百度被搜索数量

六、利用大数据提供经典理论的“历史新证”

马克思主义相关理论在 21 世纪是否仍然具有活力？这是国际人文社科领域争论不息的话题，也集中在理论分析、质性分析领域。而量化研究，囿于宏观数据、历史数据的无法获取，往往无力涉及这些话题，也就失去了进一步检验经典理论、延续理

①陈云松、朱灿然、张亮亮：《非代际文化反哺：概念、理论和实证》，南京大学社会学系工作论文。

论生命力的重要机会。而使用大数据，则可以为经典理论家的影响力、经典理论的现实解释力提供重要的依据。这是因为大数据能够以其超越传统调查数据的样本量和时空跨度，帮助研究者从过程性的历史视角来审视和验证经典的理论问题。学者发表于国际一流杂志 Social Science Research 的大数据研究，就展示了阶级意识理论的重要解释力[10]。

在图 15 中，深色曲线代表 1900–2000 年的 100 年中美国经济悲惨指数（通货膨胀率与失业率之和，一般用来衡量经济景气程度），浅色曲线代表着100 年中在美国出版的全部英语书籍中对“阶级/阶层”相关词汇的提及总量，是我们利用谷歌图书语料库对近百个关键词进行检索和主成分分析后获得。对这两组 100 年时间序列，我们进行了格兰杰因果分析、协整分析。计量分析的结果告诉我们，整个 20 世纪，美国的经济景气程度和美国人的“阶级意识”是密切相关的，而经济现象会对 3~5 年后的大众“阶级意识”产生因果影响。进一步的分析表明，经济景气程度和美国的基尼系数之间则不存在这种关系。我们接着对英国也进行同样的分析。总体上，研究结果表明，20 世纪美国的收入不公平（基尼系数）本身不影响社会的“阶级意识”，但通货膨胀和失业等经济现象则会产生阶级意识效应。而其原因，可能在于人们对全社会收入不平等的感知，不如对和自身直接相关的失业率、通货膨胀等那么敏感。该发现用跨度百年的大数据和以前无法获得的阶级关注度指标检验阶层理论，不仅是研究方法上的突破，实际上也更进一步拓展了阶级意识理论，展示了马克思经典理论的当代解释力。

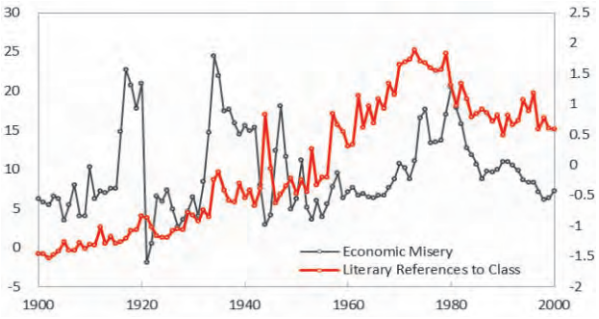


图 15　20 世纪美国的经济景气程度与民众“阶级意识”

在另外一项社会学领域内最早的大数据研究中，学者利用谷歌百万书籍比较了一系列最知名的社会学家的书籍提及率[11]。图 16 清晰地展示了马克思作为一名社会学家的重要地位。在社会学诞生以来的 150 年中，全球最为知名的社会学家排行榜里，马克思（深色双线）位居第一，韦伯（浅色双线）位居第二，美国的帕森斯（粗虚线）名列第三；而布迪厄（粗实线）在近 20 年来的知名度增长最为迅猛。

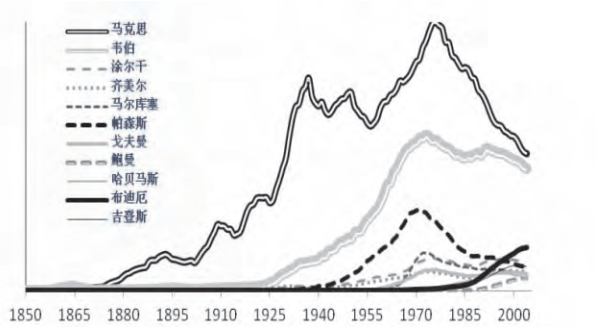


图 16　作为社会学家的马克思的文化影响力

七、结　语

在党的十八届五中全会上，中央决定实施“国家大数据战略”。实际上，在 2015 年 8 月的国务院常务会上，《促进大数据发展行动纲要》就已获通过，从政府系统的角度强调大数据作为推动经济转型发展的新动力、重塑国家竞争优势的新机遇以及提升政府治理能力新途径的重要价值。短短 2 个月后，党的中央全会进一步把国务院“大数据发展行动”提升到“国家大数据战略”层面。从执政党的视角来审视和关注大数据，具有重要的政治和组织意义。这是因为“大数据”具有国家和地方文化建设的重要战略价值，本身富含与执政能力、执政绩效和执政合法性息息相关的内容。此外，在执政党的组织架构中，宣传文化系统本身对人文社会科学战线的工作负有相关职责。因此，无论是从执政能力建设角度，还是从组织架构的工作职责角度，进一步认识大数据的战略价值，提升大数据意识，培养大数据思维，善于获得、分析和使用大数据，是执政党建设的应有之义。

在本文里，我们从国家文化建设的角度，基于目前的研究和资料，梳理了大数据尤其是社会科学大数据 6 个方面的战略价值。大数据能够以其超越传统调查数据的样本量和时空跨度，为研究者和管理者提供前所未有的海量数据、资料和信息，从而帮助研究者从过程性的历史视角来审视和验证经典的理论问题，并为公共管理提供新的决策依据。从这个角度出发，无论是学界、智库还是党政部门的决策者，都应该进一步了解大数据、善于驾驭大数据，让大数据来助力人文社科研究和国家治理。

我们认为，社会科学大数据的国家级项目、国家级实验室、分析中心，应该成为科研机制体制创新的重要内容。这是因为：第一，尽管大数据在全球

治理实践和国际学术研究中都属于前沿，但执政党、中央政府的高度重视，以及中国学者在大数据分析方面起步较早的特点，都使得我们没有后发劣势。第二，我国的体制特点和决策特征有利于集中部门、行业力量，整合多方数据，使得在大数据分析使用方面建立先发优势、拳头优势。第三，大数据的分析使用存在话语权和数据隐私等诸多问题，因此应该从国家层面重视该领域的投入、建设和管理。实际上，哈佛大学的政治学者甚至已开始利用网络大数据来实验、分析中国的互联网管理[12]；加州大学等开始使用 QQ 数据进行分析，而我们对国外大数据的分析研究仍然不够，缺乏“人类学”式的和国际分析的视野。

从智库建设的角度，我们还认为，社会科学大数据的搜集、分析和研究，应该成为国家和地方智库建设的重要内容。不过，从目前全国态势来看，尽管和大数据相关的研究机构、队伍在加快组建，基于可视化的大数据分析成果也在不断涌现，但围绕大数据的分析仍然不多，在大数据研究课题设置、量化分析水平和科学信度上仍然有待提高。更重要的是，围绕大数据的党委政府、智库、科研院所专家的交流共建仍然需要加强，尤其是大数据持有者、管理者和大数据研究者之间的沟通协作仍然不够。无论是党政机构、智库机构还是科研学术界，我们都应该进一步加强大数据学习，提升大数据意识，培养大数据思维，善于把大数据转化为研究、工作的重要手段，为文化建设和学术研究提供双赢的创新突破。

[参考文献]

[1] King, Gary. The changing evidence base of social science research [M]. King, G., Schlozman, K. L. and Nie, N. (Eds.), The Future of Political Science: 100 Perspectives. New York, NY: Routledge，2009

[2] Boyd, Danah and Kate Crawford. Critical questions for big data [J]. Information, Communication & Society，2012，15 (5).

[3] 陈云松，吴青熹，黄超.大数据何以重构社会科学[J].新疆师范大学学报(哲学社会科学版)，2015，(3).

[4] Michel, Jean-Baptiste, Yuan Kui Shen, and etc. Quantitative analysis of culture using millions of digitized books [J]. Science，2011，(331).

[5] 陈云松，吴青熹，张翼.近三百年中国城市的国际知名度——基于大数据的描述与回归[J].社会，2015，(5).

[6] 祁海宁，龚巨平.南京大报恩寺史话[M].南京：南京出版社，2008.

[7] Twenge, Jean M., Campbell, W. Keith, and Brittany Gentile. Increases in individualistic words and phrases in American books, 1960-2008[J].PLoS ONE，2012，(7).

[8] 陈云松，严飞.网络舆情是否影响股市行情？基于新浪微博大数据的 ARDL 模型边限分析[J].社会，2016(即发).

[9] 单霁翔.文化遗产保护与城市文化建设[M].北京：中国建筑工业出版社，2008.

[10] Chen, Yunsong, and Fei Yan. Economic Performance and Public Concerns about Social Class in Twentieth-Century Books [J]. Social Science Research，2016 (forthcoming).

[11] 陈云松.大数据中的百年社会学——基于百万书籍的文化影响力研究[J].社会学研究，2015，(1).

[12] King, Gary, Jennifer Pan, and Margaret E Roberts. Reverse-engineering censorship in China: Randomized experimentation and participant observation [J].Science，2014，(345).

[责任编辑：戴庆瑄]