

网络场域、文化认同与 劳工关注社群

基于话题模型与社群侦测 的大数据分析

社会
2017·2
CJS
第37卷

黄荣贵

DOI:10.15992/j.cnki.31-1123/c.2017.02.002

摘要:基于“网络与文化”和关系社会学的理论,本文指出网络场域与所沟通的文化内容之间存在互构关系。本文使用话题模型分析了关注劳工议题的用户所发布的51 288条博文,结果显示,劳工话题大致涉及以下几方面:文化与公益、工人的困境与问题、维权行动与工人组织、制度与劳工权。对用户间互动模式的社群侦测揭示了5个主要社群,结合社群关注的议题,作者将其称为:工人家园社群、工人维权社群、工人文化社群、劳工制度关注社群、工人权益关注社群。本文进一步指出劳工研究领域的两个新趋势:(1)随着新生代工人群体的壮大,工人文化与城市融入等议题受到较多关注;(2)微博促进研究者与劳工组织的互动,在一定程度上有助于后者从制度层面反思工人的境况。最后,本文讨论了大数据分析在网络文化与社会心态研究中的应用。

关键词:劳工 NGO 大数据 话题模型 互动社群 社会网络分析

Network Fields, Cultural Identities and Labor Rights Communities: Big Data Analytics with Topic Model and Community Detection

HUANG Ronggui

* 作者:黄荣贵 复旦大学社会学系(Author: HUANG Ronggui, Department of Sociology, Fudan University)E-mail: rghuang@fudan.edu.cn

** 本研究受到国家社科基金(12CSH043)、复旦大学社会发展与公共政策学院研究基金的支持。[This research is supported by Chinese National Social Science Foundation (12CSH043) and the Research Fund of the School of Social Development and Public Policy at Fudan University.]

本文曾于2016年5月在上海大学“数据科学与大都市研究中心”组织的“社会学中的大数据:应用与示例”工作坊上宣读,获得诸多宝贵建议,特此致谢。感谢匿名评审人的修改意见。文责自负。

Abstract: The Weibo platform is a social space for interaction and expression. It requires scholars to simultaneously examine communication patterns and communicated content among Weibo users. Based on theories of “network and culture” and relational sociology, this article contends that network fields and the communicated cultural meanings are mutually constituted. The author used Latent Dirichlet allocation (LDA) topic model and social network analysis techniques to examine 51 288 Weibo posts published by users concerned for labor rights. This article reveals the relationship between community structures and community focal topics. Specifically, the result of LDA topic modeling shows that the focal topics regarding labor issues can be categorized into four groups, namely workers’ culture and welfare, predicaments and problems, strikes (rights defending actions) and labor organizations, social institutions, and labor rights. Analysis of interaction patterns among users resulted in five major online communities. Based on the primary topics within communities, these five communities could be labeled as: community for workers’ homes in cities, community for workers’ culture, community for workers’ rights protection actions, community for workers’ rights and interests, and community for labor institutions. The results also showed two new trends in relation to labor issues: for one, labor culture and its integration into urban life have garnered increasing online attention with the growth of new generation workers; for another, the Weibo platform provides an interaction channel for labor researchers and labor NGOs, and such interaction facilitates the latter to critically reflect the current conditions or plights of workers from an institutional/structural perspective. This article concludes with a discussion about the significance of utilizing big data analytics to study online culture and social mentality.

Keywords: labor NGO, big data, topic model, interaction communities, social network analysis

一、问题的提出

互联网的出现为社会组织的发展提供了潜在的空间 (Yang, 2003), 其发展为资源匮乏的社会组织提供了自我呈现、网络建设、公众动员以及建构另类话语的工具, 具有赋权功能 (Sima, 2011)。现有研

究主要从三个视角检视非政府组织的新媒体使用与实践。一是从数字不平等的视角考察组织的在线影响力。非政府组织使用社交媒体的可能性及方式受组织策略、能力、治理结构及外部压力的影响(Nah and Saxton, 2013),它们所发布的博文的影响力在很大程度上取决于博文涉及的话题及其线上特征,但较少受到组织的线下特征的影响(黄荣贵、桂勇, 2014)。组织间在线链接的形成过程受组织合法性、离线合作关系、临近的地理位置及相同的关注领域等因素的影响(黄荣贵等, 2014)。第二个研究路径是从互联网政治的视角考察社交媒体对集体行动的赋权功能。现有的案例分析表明,非政府组织使用微博分享信息、直播事件发展、与传统媒体或其他用户进行互动、建构抗争话语、动员公众的支持,进而促进持续的集体行动,实现非政府组织的赋权(陈先红、张凌, 2015;陈韵博, 2014)。对“大爱清尘”的案例分析揭示了劳工组织在微博空间的话语竞争方式(Gleiss, 2015):一方面,发展另类话语来解读工人群体的境遇;另一方面,利用复调式(polyphonic)表达来合法化组织的工作,并将尘肺病议题政治化。第三个研究路径关注组织使用社会化媒体的方式。非政府组织可以使用社交媒体发布信息、建设社群、动员公众(Lovejoy and Saxton, 2012),国外的研究表明,在实践中它们主要使用推特进行单向互动(Lovejoy, *et al.*, 2012)。中国的非政府组织使用微博的方式与之类似,但更多地使用非正式语言且常发布与组织使命无关的博文(Zhou and Pan, 2016)。

尽管现有研究从不同的侧面分析了非政府组织与社交媒体使用之间的关系,但仍存在一些不足。首先,研究者倾向于将社交媒体看作一种工具性资源(影响力、动员手段),并且将用户沟通的形式与内容割裂(仅考察网络关系或话语竞争),未能展示组织间社群的整体图景。在社交媒体时代,中国的互联网用户内部存在着明显的分化,用户的社交媒体使用模式以及他们对社会、政治和外交等诸多议题的在线表达均呈现不可忽视的差异(deLisle, *et al.*, 2016),仅考察社交媒体使用的特定个案或特定方面可能会得到片面的结论。同时,社交媒体不仅是以信息为驱动的媒介,还是具有多重性和多中心特征的社会空间(Fenton, 2012; Cavanagh, 2007)。社交媒体的社会空间属性意味着研究者需要同时考察用户间的互动关系以及嵌入沟通过程的符号、认同与文化才能深入理解用户群及相应的在线公共空间。其次,以前的研

研究者主要采取案例研究和传统的内容分析方法,并没有充分利用社交媒体大数据,这也制约了他们对非政府组织及其社交媒体使用/在线表达的整体性把握。

基于上述讨论,本文以关注劳工议题的微博用户为研究对象,通过话题模型和社会网络分析技术分析微博文本大数据,初步呈现劳工关注社群的整体图景。具体的研究问题包括劳工关注用户形成了怎样的社群结构、他们之间主要就何种议题进行在线沟通与传播,跨社群互动与社群文化之间存在何种关联。对上述问题的大数据分析有多方面意义。首先,随着关注劳工议题的非政府组织的发展与分化,它们使用社交媒体的方式日益多样化,对其进行鸟瞰式描绘有助于今后的研究更准确地评估社交媒体对非政府组织的赋权作用。这既能避免过于强调社交媒体的特定使用方式与功能,也有助于发现新的趋势。其次,在方法上,本文初步展示了话题模型这一非监督性文本分析技术在研究网络文化方面的潜力,在应用层面探索了社会学拥抱大数据的可能性(孙秀林、施润华,2016)。本研究表明,除了使用现有的大数据语料库所提供的词频分析(陈云松,2015;陈云松等,2015),研究者还可以对社交媒体平台的原始文本进行语义分析来探索互联网空间所具有的文化意涵。最后,为了有效地解读、验证话题模型的结果,现有研究主要关注话题的时序变化趋势以及话题和文本来源(特征)之间的关系(DiMaggio, *et al.*, 2013; Levy and Franklin, 2014)。本文除了考察话题时序变化趋势,还将话题模型结果与用户的网络链接模式结合起来分析,这一做法既提供了解读和验证话题模型结果的思路,也初步将社会网络的结构分析与文化分析结合起来,避免了社会网络分析的形式化倾向(Fuhse and Mützel, 2011; 黄荣贵等, 2014: 58)。

二、微博空间中的网络场域与文化认同

本文将社交媒体看作一个社会空间,从社群结构和社群文化的角度考察新浪微博上的劳工关注用户群(Cavanagh, 2007)。在数字化时代,在线社群主要表现为用户间的链接模式(Cavanagh, 2007),而社群认同则表现为成员所使用的文字与符号(Tamburrini, *et al.*, 2015)。这一理解意味着本文可以借鉴关系社会学(*relational sociology*)的理论见解。关系社会学认为,文化形式与关系网络并不是两个自治性概

念,不应该对它们进行独立的测量,更不用说使用独立测量来考察两者之间的因果关系;相反,关系网络形成于沟通的文化过程(Mische, 2011; Mützel, 2009)。¹根据怀特的理论,行动者之间的关系及相关的话语互动或“故事”构成一个网络场域(netdoms),不同的网络场域之间存在竞争性关系,而认同则形成于互动过程的偶然性和网络场域之间的竞争(Mützel, 2009)。换言之,关系的结构维度与文化维度合二为一。当认同在不同网络场域“转换”(switching)时,认同将通过比较和反思而产生观念与意义。在该理论中,文化以流动的方式为社会群体所共享,而不是抽象地被设定(DiMaggio, 2011);而社会形态的出现则是行动者在不同情境下产生或解除关系所带来的结果。

对社交媒体大数据分析在一定程度上支持了上述理论。例如,政治家在意识形态上的相似性能够通过他们共有的推特粉丝进行测度(King, *et al.*, 2016)。对脸书的研究也表明,政治家及其支持者的在线关系能有效地反映出政治家及其他个人用户的意识形态倾向(Bond and Messing, 2015)。这些研究表明,在线互动关系与用户价值观之间存在密不可分的联系。我们可以使用关系社会学的术语将上述研究发现转述为:在线互动形成的社会关系构成了不同的网络场域,而用户所表达的话语符号在特定网络场域内以及不同网络场域的转换中获得具象的社会意义。对推特用户的语言使用的定量分析发现,词用法的频率和词尾频率在不同互动社群中存在显著差异,推特用户的语言风格也随着沟通对象的转换而发生改变(Tamburrini, *et al.*, 2015)。然而,这一研究仅证明了不同的在线社群存在差异化的文化与认同,但并没有从内涵上展示群体互动中产生的文化与观念。

随着计算社会学的发展,话题模型(LDA topic model)为研究文化与观念提供了可利用的手段(Blei, 2012; DiMaggio, *et al.*, 2013)。借助话题模型,研究者可以通过计算的方法从海量文本中提取有意义的话语框架(frames)。诺林(Nowlin, 2016)使用话题模型分析政策文本,提取出多维度、竞争性的议题定义框架,并进一步考察议题定义框架与外部政策环境变化之间的关系。由于话题模型能估算单个文本中各话题出现的概

1. 对网络与文化关系的另两种理解是:(1)网络是文化的传导器;(2)网络与文化是两个自主性的概念,两者之间相互影响(详见 Mische, 2011)。

率,研究者可以据此汇总出语料库(corpus)层面的话题流行率。在这个意义上,话题模型为研究者考察微观与宏观层面的议题定义框架之间的连接提供了经验基础(Nowlin,2016)。考虑到社群认同和文化的形成过程是各社群成员使用特定话语框架的累积性结果,且社群的互动关系的总和可以看作一个网络场域,话题模型与社群分析的结合将有助于深化我们对在线社群的文化形成与沟通过程的认识。

概言之,社交媒体的社会空间属性意味着,研究者可以采取网络场域与文化认同的双重视角来考察用户在线互动所形成的社会关系以及用户间所传递的话语符号。一方面,考察用户沟通所传播的话语符号有助于超越纯粹的社会网络结构分析,深入了解社群的文化和意义属性;另一方面,通过检视话语符号传播的网络场域,研究者可以结合互动情景来准确地解读话语符号的社会意义。在 Web 2.0 时代,海量的社交数据与文本分析技术的发展为社会学研究者分析在线文化、社会关系网络、行动者的地位和角色提供了难得的契机(Evans and Aceves, 2016)。本研究以关注劳工议题的微博用户为研究对象,考察用户间的社交关系和用户沟通所使用的话语符号,从而初步地呈现关注劳工议题的非政府组织的社交媒体使用/在线表达的整体图景。基于上述讨论,本文提出如下研究问题:

- Q1: 微博用户在互动过程中形成了哪些在线社群?
- Q2: 就各在线社群而言,社群成员之间主要就何种议题进行在线沟通与传播?
- Q3: 行动者的跨社群互动与社群文化之间存在何种关联?

三、数据及方法

本文选取了 14 个关注劳工议题的非政府组织作为种子用户,并在 SMP2015 微博数据集中筛选出含有种子用户的 51 288 条博文。笔者首先使用话题模型对筛选出来的博文进行机器学习以识别其所讨论的话题(Blei,2012; DiMaggio, *et al.*, 2013; Jacobi, *et al.*, 2016; Nowlin, 2016)。在该模型中,话题由一组词语的统计分布所定义,构成同一话题的词语更有可能同时出现在文本中,模型拟合结果将估计出每个话题对应的核心关键词及其概率,研究者可以根据关键词初步解读话题的含义。每一个文本均可能谈论多个话题,且每个话题的常见度由话

题概率来描述,因此话题模型是混合隶属(mixed membership)模型(Grimmer and Stewart,2013)。考虑到一条博文经常涉及两个或多个话题,混合隶属意味着研究者无需武断地做出是或否的二分决定,因此能更准确地反映文本的特征。研究者通过拟合好的话题模型估算出单个文本谈论各话题的概率,进而选择特定话题对应的典型文本,并据此评估模型的有效性。

拟合话题模型的步骤如下:首先,研究者对文本进行预处理,主要的工作是分词与特征词选择。最常见的特征词选择是删除停用词。结合前期的数据处理经验和他人的研究可知,剔除命名实体词(Burscher, *et al.*, 2016)和微博用户名有助于识别文本中的议题(否则舆论事件会被优先识别出来),因此笔者在文本预处理过程中过滤了这两类词语。接下来,研究者需要确定话题的个数并拟合话题模型。现有的方法论文献建议结合混乱度(perplexity)和一致性(coherence)指标以及话题模型的可解读性来确定话题数。其中较小的混乱度指标代表一个更优的模型,但该统计指标往往会选择一个过于复杂的模型;较大的一致性指标代表一个更优的模型,其中“c_v coherence”指标与人工评判之间具有较高的相关性(Röder, *et al.*, 2015)。最后,研究者对话题模型的有效性进行人工评估,并对有效度的话题进行后续分析。此时,研究者还需要结合研究目标来决定话题的精细程度:研究特定行动者的框架化策略时,需要较为精细的话题;而考察集体性议题定义时,话题则无需过于精细(Nowlin, 2015)。

在完成话题模型分析后,作者根据博文所记录的提及关系(直接@用户)来建构互动关系网。具体步骤如下:若用户A在发布或转发微博时直接@用户B,则认为互动关系网中存在一个由A指向B的网络关系。值得一提的是,这里仅计算直接的提及关系,不包括转发关系。由于每个用户在不同的博文中可以多次@其他用户,提及的次数则被定义为用户间互动关系的权重。根据上述规则处理博文后,本研究得到一个包括14 730位用户、41 202对关系的有向加权社会网。

由于低频互动关系可能反映随机的或不稳定的社群归属,本研究仅保留高频互动关系。为了选择一个有意义的分界点,作者分析了博文的时序分布。结果显示,有21个月的博文数超过1 000条,若将平均每月至少互动一次定义为持续而稳定的互动关系,则可以使用21作

为高频与低频互动关系的分界点。笔者以 21 为分界点删除低频互动关系后,绝大多数的(14 085)用户变为孤立点,少量(23)用户形成规模为 2 至 5 不等的网络成分,²其余 622 位用户则连结成一个大的网络成分。对这 622 位用户所构成的社会网络进行社群侦测 (community detection)³得到 5 个规模大于 50 的社群。

最后,为了理解各互动社群的议题关注度,本研究以社群成员是否出现在特定博文为标准将博文分为 5 组,根据话题模型计算出每条博文谈论各话题的概率,并汇总出各社群谈论各话题的平均概率。汇总的统计量为我们理解各社群的文化属性提供了实证基础。

本文使用 Python 进行数据处理与分析,其中话题模型使用 gensim 模型拟合,社会网络分析使用 python-igraph 模块完成。⁴

四、研究发现

本文所分析的博文发布于 2010 年 10 月至 2014 年 3 月。博文的时序分布显示(见图 1),2012 年之前发布的博文数量很少,月博文数均小于 400。自 2012 年起,博文数逐渐增加,2012 年 3 月的博文数超过 500。日渐活跃的趋势一直持续至 2013 年 9 月(4 662 条博文),此后博

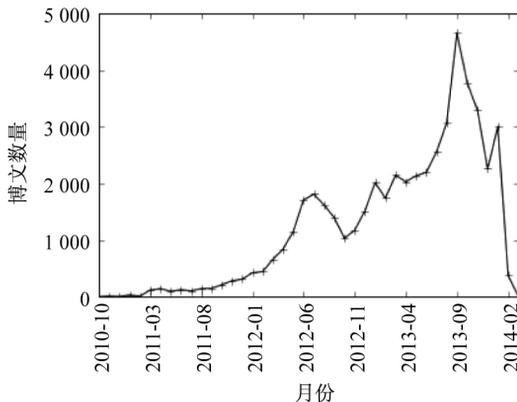


图 1: 博文数量月分布图

2. 如果一个有向网络可以分为若干部分,每个部分的节点之间存在关系(不管关系的方向性),而各个部分之间不存在任何关系,则这些部分称为弱成分(weak components)。

3. 关于社群侦测的含义,详见 Bruggeman, 2008。

4. Python 官方网站: <https://www.python.org/>; gensim 官方网站: <https://radimrehurek.com/gensim/index.html>; python-igraph 官方网站: <http://igraph.org/python/>。

文数逐渐下降,这一趋势可能与微博整体活跃度的下降有关。同时,2014年2月至3月博文数下降也可能受数据收集结束时点的影响,因此下文的分析不包含这两个月的数据。

(一) 话题模型的选择

为了确定话题数,笔者拟合了话题数为2至40的一系列模型,通过计算和比较各话题的混乱度来初步筛选备选模型。计算结果显示(见图2),随着话题数的增加,模型的混乱度持续下降,这与现有研究基本一致(Jacobi, *et al.*, 2016)。当话题数大于或等于12时,混乱度的变化幅度相对较小。结合混乱度的变化趋势,笔者进一步计算了话题数从4至30的一系列模型的一致性指标,结果表明话题数为7、9、13的模型具有较高的一致性。综合混乱度、一致性指标和这3个模型的可解读性,笔者决定采取话题数为13的模型作为分析结果进行汇报。⁵

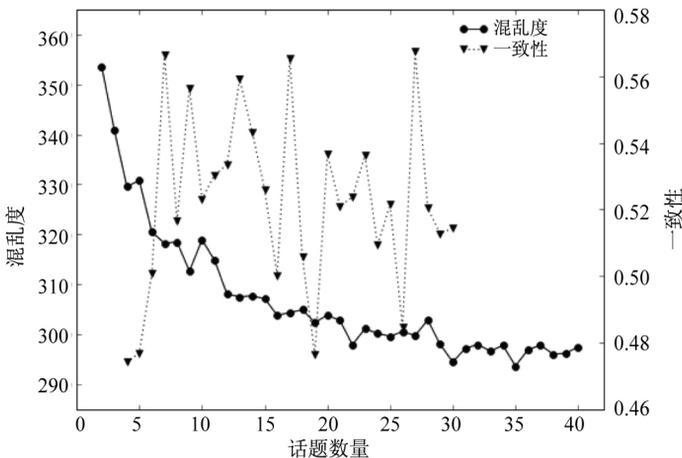


图 2: 话题模型比较

(二) 劳工社群关注什么议题?

使用话题模型分析含有特定关键词的文本可能会揭示出含有该关键词但实际上谈论无关话题的文本,这一特点应该被看作话题模型的优点(Jacobi, *et al.*, 2016)。考虑到中国非政府组织常常在微博上发布一些与组织目标无关的文字来吸引用户(Zhou and Pan, 2016),而本文的主要研究目标是理解劳工关注社群,并不是劳工组织的微博的一般

5. 感兴趣的读者可向作者索取话题数为7和9的模型结果。

性使用状况,下文仅汇报含义清晰且与劳工议题相关的 10 个话题(见表 1)。在汇报话题时,笔者既检视了话题模型给出的核心关键词,也提取并浏览了与话题相关的典型博文。为了进一步评估话题的质量(DiMaggio, *et al.*, 2013),笔者还关注了话题的时序活跃度,以便考察外部事件或环境变化与话题活跃度变化之间的关系。

表 1:话题模型结果(最重要的 10 个关键词)

1. 工人文艺	2. 春晚与公益	3. 职业病	4. 城市融入	5. 农民工问题
演出	春晚	白血病	他们	社会
劳动	公益	尘肺病	一个	农民工
艺术团	大家	生命	问题	问题
青春	支持	什么	自己	制度
音乐	孩子	他们	社会	术后
民谣	朋友	治疗	没有	一个
歌曲	一个	新年	生活	就是
分享	社区	求助	不是	学校
点击	儿童	帮助	城市	学生
一段时间	活动	好心人	农民工	工伤
6. 维权行动	7. 工人组织	8. 维权代表	9. 特定劳工事件	10. 制度性话题
(关键词 1)	工会	第一	自由	血小板
员工	谈判	医院	家属	AB 型
工厂	要求	这是	法律	国家
公司	工资	工伤	警方	民主
谈判	企业	他们	没有	社会
代表	职工	附属	相关	(关键词 6)
警察	(关键词 7)	(关键词 7)	有关	人民
老板	组织	家属	人权	(关键词 8)
劳动	代表	不再	公众	政治
要求	工作	辩护律师	代表	一个

注:这里仅汇报含义清晰且与劳工(组织)具有密切关系的话题。其中,相关度较低的话题包括:城管/小小/这个/没有/法庭/无罪/工伤/先行/暴力/不能/就是/没有/缺乏/天那水/总是/然而/权力/巡演/小时/自由;含义不清的话题为:派遣/劳务/不要/梦想/不是/建议/看到/政府/永远/投票。出于保护微博用户隐私的考虑,部分词以“(关键词 X)”的形式表示。

工人文艺与公益:话题 1 由“演出、劳动、艺术团、青春、音乐、民谣、歌曲、分享、点击”等关键词构成。⁶“演出、艺术团、音乐、民谣、歌曲”等词语表明话题 1 关注以打工者为受众或反映打工者生活的歌曲和演出。一位微博用户曾写道:“致我们伤痛倔强豪放的青春,唱给那些被

6. 话题的序号无实质意义。

收容的青春,被绝望的青春,被挤压着生长的青春。”⁷可以看出,“青春”是打工者文艺所反映的重要内容。与之相一致,“青春”是话题 1 的一个重要关键词。部分微博进一步反思了打工者的青春,并对社会现实进行批判。例如,一篇博文写道:“近来南方工人维权抗争,仿佛有些‘五四’时期的景象。这是前段时间我新写的一首歌《劳动·青春》,唱给我们那些被收容的青春……”也有演唱者在介绍歌曲时强调“唱出的是充满质感的现实”。“分享”“点击”这两个关键词则表明,新一代打工者习惯于通过社交媒体来分享和推广属于打工者群体的歌曲和演出。“趁热来听”“点击播放音乐”“演出视频随后分享”是常见的话语。从话题分布和趋势来看(见图 3),工人文艺话题呈现缓慢上升的趋势,但总体活跃度并不高。

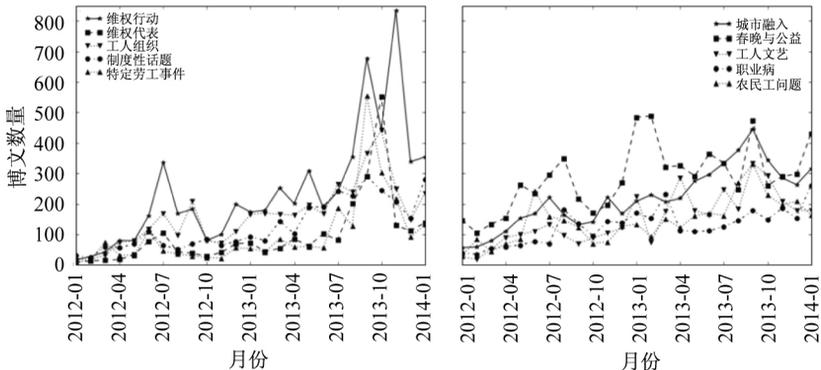


图 3:话题活跃趋势

话题 2 的核心关键词包括“春晚、公益、大家、支持、孩子、朋友、一个、社区、儿童、活动”。“春晚”这个词表明话题 2 的一个重要内容是讨论打工春晚活动。⁸有微博写道:“这是完整版的 2013 打工春晚,请朋友们广泛传阅,打工春晚——三亿打工者的春晚!”另一条微博则这样描述打工春晚的定位,“献给全国三亿劳动者的春节晚会盛装启航。新工人主体、倡导劳动价值、追求社会的公平和正义。打工春晚——劳

7. 本文所引用的微博均来自 SMP2015 微博数据集。笔者对部分引文加以微调以符合书面语言的表达规范。由于博文 ID 仅适用于数据集自身的索引,这里并不直接引用博文 ID 作为资料来源。

8. 打工春晚由草根组织“工友之家”发起,是以打工者为主题、反映打工者生活的民间春晚。

动者的春晚”。微博用户在传播打工春晚的同时向“大家”拜年,并号召“朋友们”广泛转发视频,因此“大家、朋友”也是话题2的核心关键词。⁹话题的变化趋势显示,2013年1月和2014年1月讨论打工春晚的博文数呈现局部峰值态势,这与打工春晚活动的举办时间基本一致。“公益”一词出现在本话题可能出于如下原因:打工春晚的组织者自我定位为公益组织,许多打工春晚节目的参与者平时也积极参与公益活动;此外,打工春晚也曾被“中国公益”栏目所报道。“儿童、孩子、社区”等均属于劳工公益的常见内容,这就解释了为什么“春晚、孩子”等关键词出现在同一话题中。

工人的困境与问题:就话题3而言,“白血病、尘肺病、生命、治疗”等关键词表明其讨论的主要内容是打工者的职业病问题,而“求助、帮助、好心人”等关键词则说明微博用户试图通过微博平台来呼吁社会援助。一条微博写道:“再转,让更多的人知道尘肺病农民工的悲惨遭遇!”@搜狐微博也曾写道:“在各媒体记者的陪同下,来到顺义一个城边村调研尘肺病农民工的生存状况……访谈和调研就在空地和路边进行。这种调研还真的别有一番滋味呢。”这些代表性微博与格莱斯(Gleiss, 2015)的研究发现大体上相呼应。然而,对话题的趋势分析显示,职业病话题所占的比例相对较低,并没有成为具有广泛影响力的劳工话题。

最能反映话题4内涵的关键词包括“社会、问题、城市、农民工、生活”,它们共同反映了城乡分割体系带来的社会问题以及农民工所面临的困境。一位微博用户这样写道:“我虽人在城市,可向往着农村,因为那儿有我的孩子、父母。”此外,有微博直接描述留守母亲的多重压力,或指出新工人返乡所面临的困境:“新工人返乡就业无外乎两种可能性,其一是务农,其二则是继续务工。倘若务农,新生代打工者面临着无地种、不会种的困境。”另一些微博则描述了新生代打工者渴望融入城市的愿望,“我们这代人,虽然户口在农村,但跟农村已经没有实质关系了……早已对城市产生了依赖……却没有归属感”。还有少量微博则从城乡关系的角度分析农民工融入城市的制度壁垒:“中国现代化对农民的歧视往往在无意识状态下发生。如城市化只要农民的土地,不要失地农民;工业化只要打工者的劳动力,不要打工者成为市民……”上述博文从多

9. 研究者在解读话题模型时需要关注文化的关系性意义(DiMaggio, *et al.*, 2013),结合具体的语境理解不同符号或话语同时出现在同一个话题中的原因。

个角度分析了打工者、城市和乡村之间的复杂关系,因此,本文将话题4定义为“城市融入”。趋势分析显示,城市融入话题呈现中等程度的活跃度,并且在2013年9月之前呈现较为明显的上升趋势。

话题5由“社会、农民工、问题、制度、学校、学生、工伤”等关键词构成。与话题4相比,话题5同样关注农民工所面临的问题,但讨论的角度不再侧重于城乡关系,而是从社会制度的角度进行探讨。其中,农民工子弟教育问题是讨论的焦点,这清晰地反映在2012年6月至7月的话题活跃度上。分析博文可知,北京同心民工子弟学校于2012年6月被政府叫停,引发了网民的在线讨论,他们呼吁保卫同心学校,同时也表达了对民工子弟后续安置的关注。有微博用户就同心学校停办事件发博文:“愿更多人可以关注外来务工人员群的权益保障和外来务工人员孩子的教育问题!”

工人组织与维权:话题6主要讨论工人维权行动。其中,“员工、工厂、代表、公司、老板、警察”等词语指涉了维权事件中劳工、资方、政府三方行动者,而“谈判、要求”等词语则描述了事件中工人与资方就维权诉求进行谈判和协商。微博话题趋势显示,维权话题在2013年的下半年呈明显的上升趋势,但在2014年的年初有所下降。分析话题8的博文可知,这一话题讨论的焦点在于支持与声援维权工人代表。在事件的后续发展中,工人代表被起诉,网民及工人代表的“辩护律师”在微博上发声。

“职工、工会、组织、代表”等关键词表明话题7讨论的主题是工人组织与代表。分析博文内容可知,“工会”一词表明工人在维权过程中希望工会能够代表工人的权益;若现有工会无法代表工人权益,劳工关注者则希望通过直选来选出新的工会或者通过推选工人代表来表达工人权益。针对深圳多家企业实行工会直选的事件,有博主撰文“提出若干思考和实施意见,旨在夯实工人的结社自由,防止工会直选演变为花瓶”。此外,工人、企业和工会是集体谈判的主要参与者这一事实解释了为什么“企业、工会”会成为本话题的关键词。“要求、工资、谈判、组织(动词)”等关键词表明话题7关注工人代表如何组织工人维权,或者就工资、加班、社保等问题与资方进行谈判。分析“谈判”这一关键词还可以发现如下关系:一方面,工人谈判能力不足促使工人诉诸行动来维护自身权益;另一方面,工人行动也有助于强化工人集体谈判的能力,

比如有博文曾提到,“(工人)可以考虑复工,但以老板答应其余诉求开始谈判为先决条件”。比较话题 6 和话题 7 可以发现,尽管两者均涉及工人维权,但前者侧重于维权行动与事件,而后者侧重于讨论工人组织和代表在权益维护过程中的重要性。有博文指出,“工人抱团来取暖,集体谈判争权益”,同时需要“注重策略,巩固基础,力争使劳资对话长效化”。图 3 显示,“工人组织”话题的活跃度在 2012 至 2013 年间呈现中度上升趋势,到 2013 年后期有所回落。

话题 9 主要讨论了某一起特定的劳工事件。深圳一家公司的工人代表因参与工人抗议于 2013 年被警方带走并拘留一年多,其家属与工友通过微博呼吁公众关注。“自由、法律、人权”等关键词较为清晰地反映出微博用户的讨论角度。

制度与劳工权:话题 10 的主要关键词包括“国家、民主、社会、人民、政治”等,表明本话题倾向于讨论制度与体制等政治性议题。¹⁰其中,部分博文直接讨论政治议题,但并没有将这些议题和工人议题相联系,另一部分博文则从制度的角度讨论工人问题。¹¹对话题分布的分析显示,话题 10 的活跃度相对较小。

(三) 社群结构及议题认同度分析

笔者对用户间的互动网络进行社群侦测发现了 5 个规模大于 50 的社群。社会网络可视化结果显示(见图 4),每个社群内部有较为紧密的互动关系,不同社群的成员之间的互动关系相对较少。笔者将社群侦测与话题模型结果进行综合分析(见图 5),得到各社群对不同话题的关注度,以了解各社群对议题的认同。

统计分析结果显示,第一个社群(图 4 的圆形节点)的规模为 138,笔者称其为工人家园社群。若将该社群作为一个独立的社会网进行分析,其网络传递性指标得分为 0.118(见表 2),高于工人文化社群和劳工制度关注社群的传递性,而低于工人维权社群和工人权益关注社群的传递性,表明该社群具有中等程度的群内凝聚力。与其他社群相比,本社群的成员更为关注打工者城市融入问题,同时也较为关注打工者

10. “血小板”“AB 型”这两个关键词出现在话题 10 可能有两方面原因:一是部分微博用户在积极参与讨论话题 10 的同时,也积极地为疾病患者转发求助信息;二是部分微博用户从制度的角度讨论职业病议题。

11. 考虑到保护微博用户的隐私,这里并不详细引用博文进行阐释。

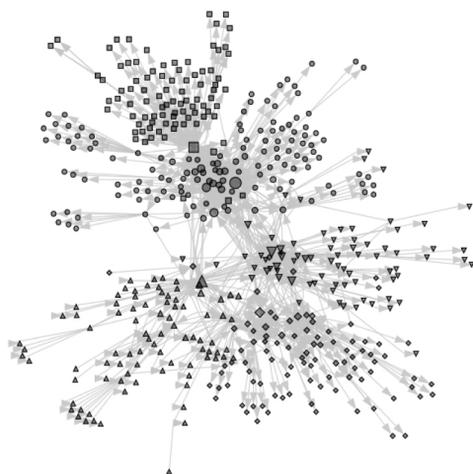


图 4: 劳工关注社群的社群侦测结果

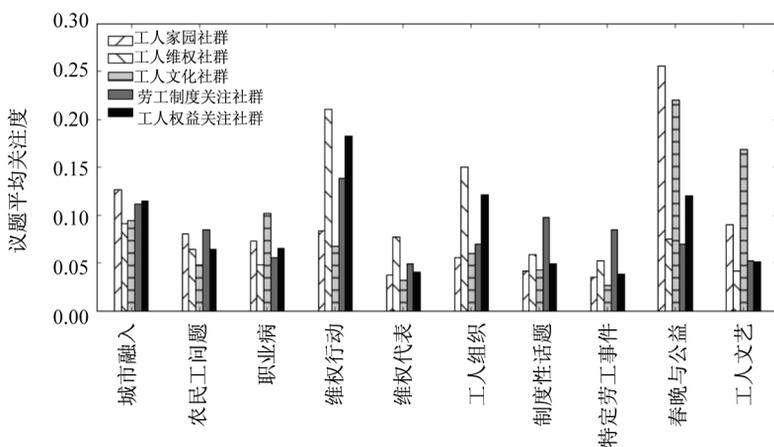


图 5: 劳工关注社群的议题关注度比较

所面临的问题。值得一提的是,该社群对春晚与公益、工人文艺、职业病等议题也较为关注。这可以从两方面进行理解:一是公益、职业病话题与城市融入话题具有较高的相容性;二是工人家园社群成员与下文所讨论的工人文化社群成员之间存在较密切的互动。对社群间互动模式的分析显示,从工人家园社群指向工人文化社群的互动次数为 1 467,明显高于其他跨社群互动的频数。由于工人文化社群较为关注

春晚、工人文艺等议题,这种跨网络场域的互动有助于促使工人家园社群的成员拓展其关注领域,延伸至与工人有关的文化话题(Mützel, 2009)。分析工人家园社群成员的构成可以发现,经常被提及的用户包括@SHWLDZGC¹²、@新工人网、@WDZKH、@同心实验学校、@工友之家工会,而总度数¹³最大的用户包括@SHWLDZGC、@WDZKH、@新工人网、@工友之家工会、@同心实验学校。根据用户的自我介绍可知,@SHWLDZGC是工人歌手、工友之家新工人艺术团的创办人;@同心实验学校是汇聚社会力量而成立的打工子弟公益学校;@新工人网则是北京同心互惠科贸公司所创办的网站;@WDZKH既是工友之家的职员,也是同心互惠公益商店的负责人;@工友之家工会自我定位为打工兄弟姐妹共同的家园。分析上述社群成员的自我描述可以发现:(1)与现有研究发现一致,在线互动社群的形成在很大程度上受社群成员线下关系的影响(黄荣贵等,2014);(2)社群的核心成员的身份在一定程度上影响了社群的议题认同倾向以及不同社群之间的互动模式。

第二个互动社群(图4的菱形节点)的规模为96,其成员特别关注维权行动、维权代表、工人组织等话题,因此笔者将其称为工人维权社群。分析社群成员的构成可知,经常被提及的微博用户包括@JJDCS、@ZFYWB、@集体谈判论坛小编、@HHG、@PPS,总度数较大的核心用户包括@ZFYWB、@中山打工族服务部、@GjZHWQC、@CDGM、@JJDCS。其中,@JJDCS在自我简介中宣称,自己愿意为劳工权利而战。@ZFYWB是广东番禺打工族服务部的工作人员,自我定位为维权人士。据财新网的报道,@ZFYWB曾参与组织多起工人行动。@HHG也是一位关注劳工权益的用户,主张通过集体谈判来建构劳工、资方和政府之间的关系。正如用户昵称所显示,@集体谈判论坛小编致力于推进集体谈判。由于工人维权或集体谈判经常涉及工会,@广东工会也经常被社群成员所提及。这一发现证明“工会”一词被识别为工人组织话题的关键词具有合理性。如果将维权社群看作一个独立的社会网,其网络传递性指标为0.154,仅次于工人权益关注社群,可见

12. 出于保护隐私的考虑,文中所有个人用户ID用字母缩写来表示。

13. 在有向社会网络中,入度(indegree)等于指向特定网络节点的关系总数,出度(outdegree)表示从特定节点指向其他网络节点的关系总数,总度数(degree)则等于入度和出度之和。

工人维权社群具有紧密的群内联系。进一步分析显示(见表2),工人维权社群主要与群内成员进行互动,与其他社群间互动的次数均小于1 000次;比较工人维权社群与其他社群的跨群互动模式可以发现,维权社群与权益关注社群的互动最频繁,其次是劳工制度关注社群,与工人家园社群、工人文化社群的互动则非常少。这一跨群互动模式在很大程度上受社群文化认同的相容度的影响;相容度越高,跨群互动就越有可能发生。综合上述分析可得到如下结论:(1)社群核心成员的自我定位、简介与基于话题模型所得的话题关注度具有较高的一致性,在一定程度上佐证了话题模型的效度;(2)与现有研究(陈韵博,2014)结论类似,劳工组织尝试通过微博平台进行维权,维权用户之间形成相对紧密的互动社群;(3)在线互动社群的边界兼具流动性和稳定性,即不同社群之间存在跨界互动,但跨界互动更有可能发生于文化认同兼容度较高的社群之间。

第三个互动社群(图4的方形节点)的规模为91,其成员特别关注工人文艺、春晚与公益等话题,因此笔者称之为工人文化社群。分析其成员构成可以发现,经常被提及的用户包括@XGRYST-XD、@OCSSSDNJ、@YjL、@YZWQ、@麻雀瓦舍,而总度数较高的用户包括@XGRYST-XD、@心跳在左边-民谣会、@HQ工作室、@OCSSSDNJ和@XjLANG7。其中,@XGRYST-XD是新工人艺术团的创始人。@心跳在左边-民谣会隶属于北京工友之家,是传播工人文化的一位代表性用户。@YjL是一位媒体人,曾主持过打工春晚节目,他所发表的博文偶尔会关注劳工维权与打工春晚等话题。@麻雀瓦舍是一个文艺汇演中心的官方微博,@HQ工作室是一个音乐/艺术工作室,尽管这两位微博用户不直接关注劳工议题,但他们所从事的文艺和演出与工人文艺传播/演出之间存在千丝万缕的联系。进一步分析表明,工人文化社群的网络传递性指标为0.010,明显低于其他社群,表明该社群的内部结构较为松散。换言之,尽管社群成员关注共同或类似的议题,但彼此之间的在线互动有限。对此有两种不同的理解:(1)工人文化的传播需要吸引广泛的受众,凝聚力强的小规模的在线社群反而是一个不利的影响因素;(2)文化的传播有赖于差异化的传播渠道和细分市场,这在一定程度上妨碍了社群核心用户之间的在线互动。这两种解释有待进一步的实证检验。

表 2: 社群间的互动关系

	(1)	(2)	(3)	(4)	(5)	网络传递性
工人家园社群(1)	265 (155 224)	4 (151)	21 (1 467)	4 (203)	15 (589)	0.118
工人维权社群(2)	5 (145)	175 (10 307)	1 (37)	7 (452)	17 (959)	0.154
工人文化社群(3)	16 (1 287)	0 (0)	105 (6 237)	1 (24)	2 (120)	0.010
劳工制度关注社群(4)	4 (243)	12 (1 016)	1 (23)	115 (6 249)	5 (297)	0.065
工人权益关注社群(5)	14 (429)	13 (408)	7 (224)	16 (600)	152 (5 989)	0.199

注: 第一个数字是发生跨社群互动的用户数(从行所在社群指向列所在的社群); 括号内的数字表示跨社群互动的次数。

第四个互动社群(图 4 的三角形节点)的规模为 88, 社群成员倾向于从制度视角讨论劳工话题。这也表明部分工人权益倡导者试图建构不同于主流的“另类话语”来解读工人的境况(Gleiss, 2015)。基于该社群所关注的议题, 笔者将其称为劳工制度关注社群。分析社群成员的构成可以发现, 不少核心成员是研究人员或律师。具体而言, 经常被提及的用户包括@WjS、@关注新生代农民工、@WKQ、@JXG、@XT, 而总度数较高的核心用户则包括@WjS、@SYLjW、@关注新生代农民工、@LJJ2012、@CNHKDC。值得一提的是, @关注新生代农民工是北京高校的 9 名学者集体设立的微博账户, 他们在富士康“九连跳”事件的背景下开通微博来探讨劳工问题。@XT 是一名律师, @CNHKDC 是一名关注劳工权益的学者, @JXG 则是一名媒体人。上述数据似乎表明律师、学者与媒体人的介入可能有助于劳工关注社群从制度的角度对工人现状进行归因与诊断。然而, 制度话题的活跃度比较低, 且劳工制度关注社群成员之间缺乏紧密的群内联系(网络传递性指数仅为 0.065)。跨社群互动分析显示, 劳工制度关注社群主要与工人维权社群进行互动, 但从制度关注社群指向维权社群的互动频率明显高于后者指向前者的频率, 可见对劳工制度的讨论并未引起其他劳工关注社群的共鸣和重视。总之, 尽管对劳工制度的在线讨论具有重要意义, 但多个网络分析指标均表明劳工制度关注社群的在线影响力较弱。

第五个互动社群(图 4 的倒三角形节点)的规模为 75, 社群成员的关注点较为分散, 相对而言对罢工维权、工人组织、城市融入等话题较为关注。尽管这个社群同样关注工人谈判和维权等话题, 但他们的关

注点侧重于维护工人的一般权益,因此笔者将其称为工人权益关注社群。对社群成员构成的分析显示,经常被提及的成员包括@城边村、@打工者中心、@QH17、@深圳小小草工友家园、@HBMG。其中,@打工者中心是由受伤工人所创立的旨在维护劳工权益和缓解劳资矛盾的民间机构;@深圳小小草工友家园是一家服务产业工人的公益机构,向工人提供文化与法律支援服务;@QH17是工人先锋网的义工编辑;@HBMG是一位在东莞工作的工人。工人权益关注社群的主要成员来自劳工组织最为发达的广东,具有较为明显的地域特征。此外,该社群还具有如下特点:(1)社群成员较为关注工人权益,但是他们与工人维权社群之间仅存在中等程度的互动(跨社群互动的次数仅600次)。这可能是因为工人权益关注社群更关注劳工组织的生存与发展,倾向于使用组织化而非直接行动的方式维护工人的权益。(2)该社群的网络传递性指标为0.199,高于其他4个社群,表明成员之间存在紧密的群内互动。一种可能的解释是,生存空间的挤压促使劳工组织之间相互支持和声援。

关系社会学的相关研究认为行动者之间的网络链接与所沟通的语言或符号之间具有互构关系(Tamburrini, *et al.*, 2015; Mützel, 2009)。这意味着同一社群成员互动时所沟通的话题与跨群互动时所沟通的话题具有显著差异。在此,笔者以工人家园社群和文化社群为例来检验互动模式与沟通内容的关联。不失一般性,笔者将城市融入和工人文化议题进行回归分析。¹⁴结果显示(见表3),与家园社群内部沟通相比,文化社群内部较少讨论城市融入话题(系数为-0.049且显著),家园社群与文化社群之间的跨界互动也较少讨论城市融入话题(系数为-0.057且显著)。进一步进行统计检验可知,家园社群与文化社群之间跨群讨论城市融入的可能性略低于文化社群成员之间讨论该话题的可能性($F=2.85, p=0.09$)。类似地,与文化社群内部的沟通相比,家园社群内部较少讨论文艺话题(系数为-0.183且显著),两个社群之间的沟通也较少讨论工人文艺话题(系数为-0.111且显著)。统计检

14. 笔者将博文作为分析单位,以博文讨论城市融入或工人文化议题的概率作为因变量、博文涉及的用户互动模式作为自变量进行回归分析,其中样本仅局限于涉及社群内部互动或两个社群成员跨群互动的16303条博文。回归分析的目的在于展示互动模式与讨论的话题之间存在的关联,本文并不对两者的联系作因果性解释。

验显示,跨群互动中讨论工人艺术的可能性显著高于家园社群内部讨论文艺话题的可能性($F=311.37, p<0.0001$)。总之,统计分析初步表明,行动者之间的在线链接模式与他们所沟通的话题之间存在紧密的关系。

表 3: 互动模式对沟通话题的影响: OLS 回归分析

互动模式	城市融入话题		工人文艺话题	
	系数	标准误	系数	标准误
常数项	0.140**	0.002	0.267**	0.002
工人家园社群内部互动	参照组	—	-0.183**	0.004
工人文化社群内部互动	-0.049**	0.004	参照组	—
家园与文化社群互动	-0.057**	0.004	-0.111**	0.005
样本量	16 303		16 303	
R ²	0.02		0.11	

注: ** 表示系数在 $p<0.05$ 水平下显著。

五、结论与讨论

本文通过大数据分析初步揭示了微博空间中劳工关注社群的互动结构以及各社群所关注的劳工话题。首先,笔者使用话题模型分析了 51 288 条博文,结果显示,劳工关注社群关注工人文艺、春晚与公益、职业病、城市融入、农民工问题、维权行动、工人组织、维权代表等诸多话题。部分话题反映了劳工组织和积极分子线上活动的新趋势。首先,随着新生代工人群体的壮大,劳工社群越来越关注工人文化与城市融入等话题。草根劳工组织和劳工积极分子通过歌唱、演出、摄影比赛等文化活动来反映工人的真实生活和社会处境,同时间接地塑造了工人群体的阶层意识。这一趋势意味着新生代农民工对自身处境有更为自觉的认识。在某种意义上,劳工组织和积极分子主动地选择与他们所代表的劳工群体相适应的倡导策略。在此分析脉络中,借鉴新生代农民工群体的研究有助于更好地理解劳工组织和积极分子的在线活动。其次,微博为劳工组织、工人积极分子、劳工研究者之间的在线互动和交流提供了平台。这种互动既为研究者了解劳工议题提供了窗口,也为劳工积极分子和草根组织了解学术研究成果、自觉地从制度和社会结构的角度反思自身处境提供了理论工具。微博用户对制度与劳工权利的讨论反映了这一新趋势。然而,讨论劳工制度的微博用户群的在线影响力相对较弱,他们对劳工行动的影响尚有待观察。以上分

析表明,利用大数据分析技术对社交媒体文本进行探索性分析有助于通过归纳的方式发现新话题和新趋势,减少现有知识的制约(DiMaggio, *et al.*, 2013),这在快速变化的新媒体生态中尤为重要。

通过对微博用户之间的互动模式进行社群侦测分析,本文揭示了5个主要的劳工关注社群。结合话题模型的结果,笔者进一步分析了各社群所关注的话题,从而窥视了各社群的文化认同。在此基础上,笔者将5个社群分别命名为工人家园社群、工人维权社群、工人文化社群、劳工制度关注社群、工人权益关注社群。对社群间互动模式的分析表明,工人家园社群与工人文化社群之间具有较密切的互动关系,这一方面是因为两个社群的核心成员拥有相同的组织归属,另一方面是因为两个社群的议题认同具有较高的相容性。工人家园社群关注新生代工人与城市之间的关系,而工人文化社群试图通过歌曲和表演等方式来呈现工人群体的日常生活,两者均关注工人在城市生活中的境况。其次,劳工制度关注社群与工人维权社群之间存在较为密切的跨社群互动,这可能是因为权利理念正逐渐成为工人行动的观念基础。值得一提的是,工人维权社群与工人权益关注社群之间并不存在密切的互动,表明劳工组织与工人维权之间可能并不存在紧密的联系。对此,有两种可能的解释:(1)除了直接的维权行动外,后者还关注城市融入、公益等与权益相关的话题,对权益的多元理解可能会影响它们的在线行动策略;(2)后者同时关注劳工组织的生存与发展,外部社会环境的制约使劳工组织倾向于避免组织和参与工人的直接行动。上述发现似乎表明,不同社群文化认同的相容度在很大程度上影响了社群之间在线互动的频率。当然,由于本文尚未系统地测度文化认同相似度与跨群互动之间的关系,这一结论有待进一步考察。

将微博平台看作用户互动与表达的社会空间意味着研究者应该同时考察沟通中形成的社会关系和沟通的实质性内容。这一研究路径有助于研究者就特定议题(比如劳工领域)的在线沟通与呈现勾勒出一幅整体的图景。随着中国社交媒体空间的日益分化(deLisle, *et al.*, 2016),整体性考察也变得越来越重要。在本研究中,笔者从网络场域与文化认同的角度分析在线互动社群与社群议题认同之间的互构关系,这一研究路径可以指导我们交互评估话题模型和社群侦测结果的效度——当社群结构有助于解读话题模型的结果且话题模型的结果亦

有助于对互动场域进行实质性定义时,研究发现具有较高的效度;反之,研究结果缺乏效度。同时,将社交媒体理解为社会空间也为我们理解在线数据与经验研究之间的关系提供了新的理论视角。本文认为,理解在线社群结构、社群的文化认同以及两者之间的关系是通过文本大数据分析推进社会学理论发展的一个可能的路径(类似观点可参见 Evans and Aceves, 2016)。随着海量文本数据的出现以及文本分析技术的发展,社会学家能够在中观、宏观层面测量文化环境以及不同类型的文化(Bail, 2014),从而更有效地研究不同文化与宏观环境之间的关系。以本研究为例,倘若将在线社群理解为中观的文化情境,那我们可以发现,不同情境中的行动者对同一个议题(如劳工权益)具有差异化的理解,而相同的行动者在不同的情境中也会谈论不同的话题。

从网络场域与文化认同互构的角度分析社交媒体大数据也有助于我们从新的视角理解在线空间与离线空间的关系。传统的分析视角倾向于将在线空间看作离线空间的延伸,同时会质疑线上数据的真实性及其可能存在的对因果机制的扭曲。从网络场域与文化认同互构的理论视角来看,用户所呈现的实质性内容取决于沟通过程所嵌入的场域,同一用户所沟通的实质性内容也随场域而变化,因此并不存在唯一“真实”的社会事实。一旦用户自我呈现的信息通过在线互动而传播,这些内容也会影响到用户之间的线上互动模式,并构成一个具有相对独立性的场域。一系列网络场域的累积性变迁甚至有可能带来离线社会的变迁。就此而言,沟通的实质性内容如何导致网络场域的出现、持续和消解,不同的网络场域又是如何影响行动者之间沟通的实质性内容才是分析的重点,而这一研究视角尤其适用于对互联网文化、网络社会心态等问题的研究。

在研究方法上,本文初步展示了话题模型在文本大数据分析中的应用。值得一提的是,话题模型和社群侦测技术的结合使研究者可以在经验上分析宏观结构与微观过程之间的连结。在微观层面,不同的行动者通过发布博文来表达自身对劳工议题的理解,个人层面的竞争性的议题定义方式汇总到宏观层面可得到若干占主导地位的议题,并以宏观结构的方式呈现。在此,话题模型为研究者考察话语的微观表达与宏观议题结构之间的联系提供了分析工具。同时,对劳工议题有不同理解的用户具有竞争或盟友关系,这种关系将抑制或促进用户互

动,而互动关系为进一步的议题定义提供了适宜的网络场域。因此,综合使用网络分析技术和话题模型有助于我们分析沟通模式与沟通内容之间的动态互构关系以及这一互构关系所引起的社群间的链接关系的变迁过程。

本研究仍存在一些不足:首先,本文所分析的数据在一定程度上取决于种子用户,今后的研究可以扩大或改变种子用户,并评估种子用户对研究发现的影响。其次,本文仅通过间接的方式评估了话题模型的效度,今后的研究可以结合监督性机器学习方法直接验证话题模型的效度。

参考文献(References)

- 陈先红、张凌. 2015. 草根组织的虚拟动员结构:“中国艾滋病病毒携带者联盟”新浪微博个案研究[J]. 国际新闻界(4):142-156.
- 陈韵博. 2014. 劳工 NGO 的微博赋权分析:以深圳“小小草”遭遇逼迁事件为例[J]. 国际新闻界(11):51-60.
- 陈云松. 2015. 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(1):23-48.
- 陈云松、吴青熹、张翼. 2015. 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 35(5):60-77.
- 黄荣贵、桂勇. 2014. 自媒体时代的数字不平等:非政府组织微博影响力是怎么形成的?[J]. 公共行政评论(4):133-152.
- 黄荣贵、桂勇、孙小逸. 2014. 微博空间组织间网络结构及其形成机制:以环保 NGO 为例[J]. 社会 34(3):37-60.
- 孙秀林、施润华. 2016. 社会学应该拥抱大数据[J]. 新视野(3):36-41.
- Blei, David M. 2012. “Probabilistic Topic Models.” *Communications of the ACM* 55(4):77-84.
- Bail, Christopher A. 2014. “The Cultural Environment: Measuring Culture with Big Data.” *Theory and Society* 43(3):465-482.
- Bond, Robert and Solomon Messing. 2015. “Quantifying Social Media’s Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.” *American Political Science Review* 109:62-78.
- Bruggeman, Jeroen. 2008. *Social Networks: An Introduction*. New York: Routledge.
- Burscher, Bjorn, Rens Vliegthart, and Claes H. de Vreese. 2016. “Frames beyond Words: Applying Cluster and Sentiment Analysis to News Coverage of the Nuclear Power Issue.” *Social Science Computer Review* 34(5):530-545.
- Cavanagh, Allison. 2007. *Sociology in the Age of the Internet*. Maidenhead: Open University Press.
- DiMaggio, Paul. 2011. “Cultural Networks.” In *The Sage Handbook of Social Network Analysis*, edited by John Scott and Peter J. Carrington. London: Sage:286-300.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. “Exploiting Affinities between Topic

- Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U. S. Government Arts Funding." *Poetics* 41(6):570—606.
- deLisle, Jacques, Avery Goldstein, and Guobin Yang. 2016. *The Internet, Social Media, and a Changing China*. Philadelphia: University of Pennsylvania Press.
- Evans, James and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42:21—50.
- Fenton, Natalie. 2012. "The Internet and Social Networking." In *Misunderstanding the Internet*, edited by James Curran, Natalie Fenton and Des Freedman. New York: Routledge:123—148.
- Fuhse, Jan and Sophie Mützel. 2011. "Tackling Connections, Structure, and Meaning in Networks: Quantitative and Qualitative Methods in Sociological Network Research." *Quality & Quantity* 45(5):1067—1089.
- Gleiss, Marielle Stigum. 2015. "Speaking up for the Suffering (Br) Other: Weibo Activism, Discursive Struggles, and Minimal Politics in China." *Media, Culture & Society* 37(4):513—529.
- Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* doi:10.1093/pan/mps028.
- Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. 2016. "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling." *Digital Journalism* 4(1):89—106.
- King, Aaron S., Frank J. Orlando, and David B. Sparks. 2016. "Ideological Extremity and Success in Primary Elections: Drawing Inferences from the Twitter Network." *Social Science Computer Review* 36(4):395—415.
- Levy, Karen E. C. and Michael Franklin. 2014. "Driving Regulation: Using Topic Models to Examine Political Contention in the U. S. Trucking Industry." *Social Science Computer Review* 32(2):182—194.
- Lovejoy, Kristen and Gregory D. Saxton. 2012. "Information, Community, and Action: How Nonprofit Organizations Use Social Media." *Journal of Computer-Mediated Communication* 17(3):337—353.
- Lovejoy, Kristen, Richard D. Waters, and Gregory D. Saxton. 2012. "Engaging Stakeholders through Twitter: How Nonprofit Organizations are Getting More out of 140 Characters or Less." *Public Relations Review* 38(2):313—318.
- Mische, Ann. 2011. "Relational Sociology, Culture, and Agency." In *The Sage Handbook of Social Network Analysis*, edited by John Scott and Peter J. Carrington. London: Sage: 80—97.
- Mützel, Sophie. 2009. "Networks as Culturally Constituted Processes: A Comparison of Relational Sociology and Actor-Network Theory." *Current Sociology* 57(6):871—887.
- Nah, Seungahn and Gregory D. Saxton. 2013. "Modeling the Adoption and Use of Social Media by Nonprofit Organizations." *New Media & Society* 15(2):294—313.
- Nowlin, Matthew C. 2016. "Modeling Issue Definitions Using Quantitative Text Analysis." *Policy Studies Journal* 44(3):309—331.
- Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." http://svn.aksw.org/papers/2015/WSDM__Topic_Evaluation/public.pdf, accessed on 14 October 2016.

- Sima, Yangzi. 2011. "Grassroots Environmental Activism and the Internet: Constructing a Green Public Sphere in China." *Asian Studies Review* 35(4):477-497.
- Tamburrini, Nadine, Marco Cinnirella, Vincent Jansen, and John Bryden. 2015. "Twitter Users Change Word Usage According to Conversation-Partner Social Identity." *Social Networks* 40:84-89.
- Yang, Guobin. 2003. "The Co-Evolution of the Internet and Civil Society in China." *Asian Survey* 43(3):405-422.
- Zhou, Huiquan and Quanxiao Pan. 2016. "Information, Community, and Action on Sina-Weibo: How Chinese Philanthropic NGOs Use Social Media." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 27(5):2433-2457.

责任编辑:冯莹莹