

# 英文维基百科的生育相关公共健康信息 趋势分析\*

王彦妍<sup>1,2</sup> 张进<sup>3</sup>

(1. 中国人民大学信息资源管理学院, 北京 100872;

2. 人文北京研究中心, 北京 100872;

3. 威斯康辛大学密尔沃基分校, 美国威斯康辛密尔沃基 53211)

**摘要:** [目的/意图] 生育话题是公众广泛关注的健康话题之一。随着全面二孩政策的推进, 近年来生育及其衍生话题日益成为社会公众关注的焦点。本研究基于英文维基百科, 对世界范围内生育相关主题、话题进行了探索, 并对其内在内容变化和外在关注度变化趋势进行了较为深入的分析。[方法/过程] 从英文维基百科采集 150 个生育健康相关词条, 包含 2010 至 2017 年的历史版本及编辑和浏览数据, 使用开放式编码、趋势分析、自组织映射、n-gram 文本分析等方法进行定性与定量相结合分析。[结果/结论] 分析结果显示, 生育健康相关主题包括计划生育和生殖健康、人类与环境、以及人口问题, 热点话题涉及生育健康、妇女保护、经济、环境等方方面面, 在内容上随着时间推移越来越丰富和多元; 外在关注度近年来整体呈下降趋势, 与国内趋势有明显区别; 维基百科内容生产者和用户对于相关内容的关注趋势也存在明显差异, 可见两个群体并非完全重叠。

**关键词:** 生育健康 公共健康 英文维基百科 社交媒体 趋势分析

**分类号:** G350

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2020.04.04

## 0 引言

2017 年全国生育状况抽样调查数据显示中国已处于低生育率近 20 年, 近年来妇女孕育年龄推迟, 一孩生育率下降。虽然随着单独二孩和全面二孩政策的推行, 二孩生育率上升, 但总体生

\* 本文系国家社会科学基金青年项目“媒体融合环境下用户健康信息搜索行为模式研究”(项目批准号: 20CTQ023)的研究成果之一。

[作者简介] 王彦妍 (ORCID: 0000-0002-5287-117X), 女, 讲师, 博士, 研究方向为信息检索、用户行为、用户健康信息, Email: yanyanwangruc@ruc.edu.cn; 张进 (ORCID: 0000-0002-6665-6606), 男, 美国威斯康辛大学密尔沃基分校博士生导师, 终身教授, 博士, 研究方向为信息检索、用户健康信息、信息可视化, Email: jzhang@uwm.edu。

育率却没有明显回升,部分学者认为持续低生育和少子化已成为中国人口新常态<sup>[1-2]</sup>。二孩政策全面放开的同时,生育及其衍生的医疗、教育、就业、社会福利、妇女儿童保护等相关话题也成为公众和专家学者讨论的焦点,在网络、尤其是社交媒体上得到了密切关注和讨论,而这些话题也是全世界广泛关注的问题。

近年来社交媒体已经成为用户产生、搜寻信息的重要渠道,它促使传统的单向、一对一或一对多的信息传播方式向双向、多对多的方式转变<sup>[3]</sup>。随着内容和用户的增多,社交媒体也成为健康信息产生、传播、搜寻和使用的重要平台。维基百科是用户协作生成知识的社交媒体平台,内容具有较高的准确性,被认为质量可媲美《大英百科全书》,其中有较为全面的疾病和健康信息<sup>[4-6]</sup>。

本研究旨在探索英文维基百科中生育相关主题、话题及其发展变化趋势,以此反映国际公众对生育健康及相关话题的关注度和关注点及其变化情况。选择维基百科作为数据来源,首先是因为维基百科准确性较高且健康信息较全;其次维基百科作为社交媒体平台,可反映一般公众对相关话题的认知和态度;第三维基百科所有词条的历史版本等数据都被较好、较完整地保存了下来,可以从时间维度呈现健康信息的长期变化。研究结果将有助于加深健康专家和一般公众对生育相关信息的了解以及对生育相关信息需求的了解,利于专家进行相关信息和知识的传播与普及,缩短专家和公众间的距离;其次健康信息生产者和健康网站可根据挖掘出的健康主题和话题,推送相关内容或制作相关专题,吸引用户访问和使用;此外,由于本研究聚焦英文维基,有助于国内学者了解国际上生育健康的关注趋势,也有助于公共卫生部门管理者基于国际和国内趋势进行决策。

## 1 相关研究

### 1.1 社交媒体健康信息相关研究

社交媒体是信息产生、共享、传播,以及用户交流的虚拟社区<sup>[7]</sup>。Kaplan 和 Haenlein 认为社交媒体是基于 Web 2.0 技术而产生的、允许用户生成和交换内容的互联网应用<sup>[8]</sup>。由此可见,社交媒体的最重要特点为基于互联网和 Web 2.0 技术,并且以用户生成内容为主。社交媒体的形式多种多样,Xie 和 Stevenson 将其分为八大类,分别为博客、微博客、播客、图片共享社区、视频共享社区、社交网络、RSS 和 Wikis。其中维基百科是 Wikis 最成功的实践成果,是当前最大的用户生成内容平台之一,不仅包含大量词条,也包括词条间的语义关系<sup>[9-10]</sup>。

社交媒体是健康信息产生和传播的重要媒介。当前各类社交媒体平台上的健康信息层出不穷,内容包含疾病、医疗、营养、保险、生活方式等方方面面,是健康信息用户获取所需信息的重要来源<sup>[11-12]</sup>。Dawson 的报告显示,81% 的欧洲用户和 63% 的美国用户相信社交媒体上的健康信息,Facebook、YouTube 和 Twitter 都被认为是可信的健康信息源<sup>[13]</sup>。Marar, Al-Madaney 和 Almousawi 发现 85% 的患者及其亲友利用社交媒体获取健康信息<sup>[14]</sup>。然而,虽然维基百科被认为准确性较高,但其他社交媒体平台上健康信息的真实性和准确性仍然存在争议。李月琳、张秀和王姗姗从真伪辨别的角度对社交媒体健康信息进行研究,发现伪健康信息存在标点和语法错误,内容上存在重复或不完整、诱导性强、主观性强等问题<sup>[15]</sup>。

社交媒体特殊的社交属性也正在改变患者、健康信息用户、健康专家和相关机构的信息行为, 重构这些个人、群体和组织机构之间的关系。健康专家使用社交媒体分享健康知识、增进技能和相互交流<sup>[16]</sup>。糖尿病患者及其亲友倾向于在 Facebook 上分享个人信息, 希望得到疾病治疗意见和反馈, 并获得情感支持<sup>[17]</sup>。张敏、车雨霏和张艳通过文献综述, 对社交媒体平台健康信息传播行为的理论基础、研究方法、研究主题和前沿热点进行了总结, 认为该领域的研究已经形成了系统性的分析框架和体系<sup>[18]</sup>。

网络百科是社交媒体中的重要类型, 其中有许多关于健康信息的内容。对于维基百科上健康信息的研究, 主要包括对于健康信息内容、质量, 以及信息用户使用的研究。例如 Wang 和 Zhang 通过对维基百科女性健康相关内容的分析, 发现主要包括歧视和暴力伤害、女性健康问题和风险、以及女性支持和保护等主题<sup>[19]</sup>。Fetahu 等发现内容生产者过往编辑词条次数可用于预测词条内容质量<sup>[20]</sup>。Huisman 等探讨了中老年人对维基百科上健康信息的看法, 发现他们在健康管理和决策中较少参考该平台的健康信息<sup>[21]</sup>。此外, 还有较多研究对教学和研究中使用维基百科健康信息的情况进行了探索<sup>[22]</sup>。

## 1.2 生育健康相关研究

大众面临着许多健康风险, 在 20 世纪末健康相关研究更侧重于对疾病和医疗的研究, 然而随着医学研究的不断进步, 专家学者们发现社会因素、行为习惯因素等会对人类的健康产生影响, 比如生活方式、家庭、社会服务、环境等, 公共健康理念从疾病治疗向预防保健转变, 相关研究的边界也得以拓展<sup>[23-25]</sup>。公共健康相关研究中生育健康是一个重要的话题, 涵盖生殖健康、计划生育、妇女和儿童保护等多方面的研究问题。

Kotch 发现自 1970 年起, 女性初婚初孕的平均年龄开始上升, 由于离婚率和未婚妈妈数量的上升, 单亲家庭数量也在增多<sup>[26]</sup>。孕妇面临着生理和心理疾病 (如寨卡病毒、妊娠期血糖高、产前和产后抑郁) 等多重风险, 其中单亲妈妈面临的此类风险更高<sup>[27-29]</sup>。受教育程度和经济状况都是影响孕产健康的重要因素<sup>[30]</sup>。

近年来中国生育相关研究的重点之一是“两孩政策”。2013 年中共十八届三中全会决定实施单独两孩政策, 而后在 2016 年启动全面两孩生育政策。政策调整后, 二孩出生人数呈现增长趋势, 生育率回升<sup>[31]</sup>。与此同时, 也催生了社会对高龄孕妇孕产风险和保健、女性二孩生育意愿等问题的讨论。祁静、茅倬彦和申小菊通过调查, 发现高龄二孩孕妇对高危妊娠咨询、紧急情况医疗转诊服务需求较高<sup>[31]</sup>。祁静、茅倬彦和顾宝昌发现单独二孩政策推行后, 北京地区 40% 左右的二孩是在孕妇意外怀孕下获得, 这也造成了潜在的风险<sup>[32]</sup>。对于女性二孩生育意愿的研究显示, 知识女性和其他女性在生育意愿上没有显著差别, 年龄、健康状况、医疗条件、工作状态、经济状况、相关福利、对生育的主观认知等因素都会影响女性的生育意愿<sup>[33-34]</sup>。

网络生育健康信息的相关研究主要从两个角度出发, 一是用户在网络上的生育相关健康信息行为, 二是网络生育健康信息本身。朱丽丽等人于 2017 年 9 月对各省级卫生和计划生育委员会官方微博的使用情况进行调查, 发现微博内容近 60% 为宣传科普, 且部分信息缺少科学依据, 总体来说官方微博运营管理水平较低, 传播力较差<sup>[35]</sup>。Arden 等人采集了育儿论坛中的 400 篇女性发布的帖子并进行内容编码分析, 将其归纳为 3 个主题, 分别为妊娠期体重控制意识和责

任、风险认知、以及她们收到的一些迷惑的信息<sup>[36]</sup>。

由以上分析可见,当前从公众视角对生育健康信息的研究主要集中于对微博上机构官方账号和普通公众所发布的信息,以及育儿论坛中女性发布的信息,对于网络百科上由广大公众共同生成的生育健康信息研究很少。然而网络百科作为公众协作生成知识的社交媒体平台,包含了大量生育健康相关的信息和知识,有极大的研究价值。此外,目前的研究主要关注网络生育健康信息内容和主题,而对其变化趋势和规律的研究很少。为了填补这方面的研究空白,本文聚焦维基百科平台上生育健康信息的变化情况,从内外部视角研究其变化发展的规律。

## 2 研究问题

本研究旨在从内部和外部两个角度探索英文维基百科平台上生育健康信息的演变,分别从外在关注度和内在内容进行深入探讨,研究问题为:(1)英文维基百科平台上生育相关信息在内容和结构上有什么特征?这一研究问题主要探索英文维基百科平台上生育相关主题和热门话题分别有哪些,以及它们之间如何相互关联。(2)英文维基百科平台上生育相关信息在内部和外部两方面如何变化?这一研究问题主要探索2010至2017年英文维基百科平台上生育相关主题和热门话题在关注度和内容上如何变化。

维基百科中一类相关词条拥有共同的主题,热门话题则是某个词条的关注点。生育相关词条按照主题分为不同大类,类与主题一一对应;而热门话题则从词条的内容中提取,词条与热门话题之间为多对多的关系。因此,主题与热门话题之间也呈现多对多的网状关系。图1展现了主题、词条和热门话题之间的关系。由该图可见,本文中词条是连接主题和热门话题两个概念的桥梁。

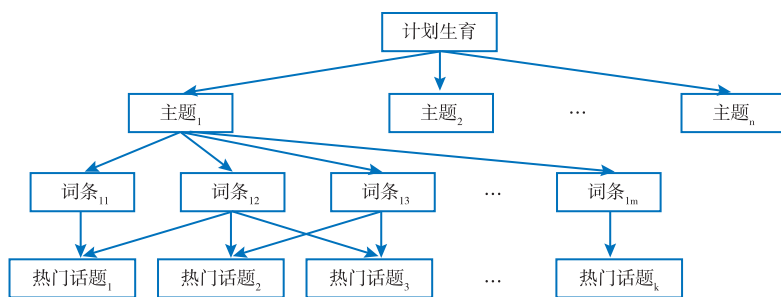


图1 主题、词条和热门话题的相互关系

为研究生育内容的演变,本文将观察期2010至2017年划分为四个时期,分别为2010至2011年(P1)、2012至2013年(P2)、2014至2015年(P3)、以及2016至2017年(P4)。在此基础上探索相关内容从一个时期到下一个时期的变化和长期的变化趋势。

## 3 研究方法

本研究以上文提及的三个概念,即主题、词条和热门话题为3种单元进行数据采集,采用定

性与定量相结合的研究方法进行分析。维基百科是当前最大的用户生成知识共享平台和网络百科全书之一, 所有当前和历史数据都向用户全面免费开放, 本研究所有数据都从英文维基百科网站和 Wikimedia Downloads 获得。

### 3.1 数据采集

#### 3.1.1 词条采集

生育相关词条通过两种手段搜寻和采集。一种是采用关键词搜索, 通过英文维基百科页面上的搜索框, 以“family planning”为检索词进行检索。以相关性对检索结果进行排序, 并对最相关的 100 个检索结果进行人工筛选, 由健康信息专家选取出与生育相关的词条。第二种是通过维基百科上每个词条的相关词条 (See also) 部分中的相关词条进行搜寻。首先以“Family planning”词条为中心, 从该词条的“See also”部分中找到相关词条, 作为一级相关词条; 继而从一级相关词条的“See also”部分找到二级相关词条。将以上两种方式获得的相关词条去除重复, 则得到本研究所需的生育相关词条。

#### 3.1.2 文本采集

维基百科中词条内容具有统一的格式, 基本上每个词条都包含标题、跳转词条 (即同义但标题不同的词条)、词条概述、目录、主体内容、See also 和参考文献。对于每个相关词条, 英文维基百科和 Wikimedia Downloads 保存其所有内容的当前和历史版本。根据所定义的四个时期, 采用 R 语言和 WikipediR 工具包采集每个词条在 2011、2013、2015 和 2017 年的最终版本<sup>[37]</sup>。

#### 3.1.3 关注度数据采集

与内在内容相对应的是外在的关注度, 本研究采用词条编辑次数和浏览次数测度各词条的关注度。编辑次数从内容生产者 (即维基百科编辑) 角度反映词条关注度, 而浏览次数从信息用户 (即维基百科浏览者) 角度反映关注度。词条编辑次数数据从英文维基百科网站采用 R 语言抓取, 浏览次数数据则从 Wikimedia Downloads 网站同样采用 R 语言获得。

### 3.2 数据分析

#### 3.2.1 主题提取

根据上文所述的词条采集方式所获取的相关词条涉及生育的方方面面, 但由于没有现成的类别划分, 研究人员采用开放式编码, 根据每个词条的内容将其归为不同大类, 并以人工浏览分析的手段提取每个大类的主题。如上文所言, 类和主题之间呈一一对应关系。

在开放式编码过程中, 由健康信息领域研究人员自行创建对概念的命名。首先根据每个词条的内容给予其标签, 而后不断比较不同词条的标签, 将相似标签合并, 形成类属, 从而实现主题提取<sup>[38]</sup>。在一位研究人员完成编码和分类后, 另一位研究人员根据最终形成的类再次进行编码。对两位研究人员的编码结果进行一致性检验, 得到的结果为 0.855, 一致性较高。

#### 3.2.2 文本处理

对于每个大类中的词条, 本文采用自组织映射方法将其聚为不同的簇, 使用 n-gram 模型发现高频词和短语, 再由人工分析提炼出热门话题。而进行文本分析的第一步则是对所采集的文本进行处理。

首先采用 R 语言和 tm 工具包对采集到的词条本文进行分词、去除停用词和日期、公式、特

殊符号等处理,获得词频,并将词频低于4的词去除,继而形成词条—词项矩阵,矩阵中单元格中的数字( $a_{ij}$ )代表词条 $i$ 中词项 $j$ 出现的频数。在获得了词条—词项矩阵之后,使用 $tf-idf$ 算法对矩阵中的数值进行进一步处理,使矩阵转换为 $tf-idf$ 数值矩阵<sup>[39]</sup>。该矩阵中的数值采用公式(1)进行计算,公式中 $a_{ij}$ 代表原词条—词项矩阵中的数值, $m$ 代表矩阵中包含的词条数, $e_j$ 代表所有词条中包含词项 $j$ 的词条数, $b_{ij}$ 代表 $a_{ij}$ 经过 $tf-idf$ 算法计算后所对应的数值。经过处理后,得到每个主题的 $tf-idf$ 数值矩阵。

$$b_{ij} = a_{ij} \times \text{Log}(m / e_j) \quad (1)$$

### 3.2.3 自组织映射分析

自组织映射方法是一种用以进行相似性计算的神经网络分析方法,整个分析是一种递归的回归过程<sup>[40]</sup>。本研究采用Kohonen的自组织映射算法,该算法将任意维度的输入层映射到一维或二维的输出层上<sup>[40]</sup>。在一次输入中,输出层的神经元经过竞争,一个神经元胜出,而后神经元随之进行微调,如此循环,最终将输入矩阵以有序的方式映射到输出层上。相似的输入单元对应同一个获胜神经元则代表它们被分为同一类。

本文采用各个主题所生成的 $tf-idf$ 数值矩阵作为分析的输入,输出聚类分析的可视化网格和U-matrix值。根据文本处理过程形成的输入矩阵为 $D=\{d_i: i=1, \dots, k\}$ ,其中 $d_i$ 代表词条 $i$ 的输入单元, $k$ 代表输入矩阵中的总词条数(即输入层的输入单元数)。输入单元 $d_i=\{b_{i1}, \dots, b_{iq}\}^T$ ,其中 $q$ 代表输入矩阵的总词项数。输出的可视化网格中,每个数字代表相应编码的词条,数字在网格上距离越近,则意味着其所代表的词条越相似。同时,网格背景中呈现的颜色由U-matrix值决定,U-matrix值越大则代表该区域为簇的边界,数值小则代表该区域可作为簇。

基于自组织映射方法的原理,本研究根据输出可视化网格上数字间的距离和背景颜色(即U-matrix值)得到聚类结果,聚类标准为:(1)位于同一节点的词条归为一个簇;(2)若词条位于不同节点,那么若其处于相邻或相隔一个节点的距离,且所处区域对应的U-matrix值小于最高值的一半,则同样被归为一个簇。基于这两个条件,每个主题中的词条通过聚类分析被归为若干个簇,未被归为某个簇的词条则单独存在。

### 3.2.4 热点话题分析

在开放式编码得到的大类以及聚类分析得到的簇的基础上,本研究采用R语言和n-gram语言统计模型,以及人工归纳总结进行热点话题的提取和热点话题演变分析。首先对于每个主题,采用自组织映射对2017年的最终版本进行聚类分析,得到每个大类的簇;再采用n-gram模型得到每个簇的高频词/短语,研究人员基于高频词/短语进行人工主题分析,提取每个簇的热点话题。基于这些分析结果和之前得到的相关词条、主题可发现三者之间的关联,从而解答第一个研究问题。本研究中的高频词/短语定义为每个簇中词频最高的15个词/短语或词频高于4的词/短语(当词频高于4的词/短语不足15个时)。

本文采用的n-gram是统计语言模型,用于计算某个文本中由 $N$ 个词组成的句子/短语出现的概率。该模型基于公式(2)计算概率,即:

$$p(S) = p(w_1 w_2 \dots w_n) = p(w_1) p(w_2 | w_1) \dots p(w_n | w_{n-1} \dots w_2 | w_1) \quad (2)$$

公式(2)中, $S$ 为由 $N$ 个词组成的句子/短语, $w_i$ 为组成 $S$ 的词, $p(S)$ 为该句子/短语在

文本中出现的概率。

热点话题的演变则通过对比 2011、2013、2015 和 2017 年四个不同时期的文本进行分析。首先采用 n-gram 模型分别对每个主题的 4 个时期的文本进行分析, 得到各个主题和时期的高频词/短语, 对同一词/短语在相邻时期的词频进行差值计算, 发现增减幅度最大的词/短语, 进而提取出 2010 至 2017 年间增长(词频持续增加)、消亡(词频持续减少)或存在波动但总体保持稳定(词频有增减但总体保持稳定)的话题。这些分析结果可反映生育相关主题和热门话题在内容上的变化, 即解答第二个研究问题中的内部变化问题。

### 3.2.5 趋势分析

除对内容本身进行分析之外, 对于关注度的分析基于编辑次数和浏览次数展开。维基百科和 Wikimedia Downloads 网站采集的数据通过统计得到各主题年度编辑次数和浏览次数, 以绘制趋势图的方式呈现关注度在这两个维度的变化, 从而解答第二个研究问题中的外部变化问题。

## 4 结果分析与讨论

### 4.1 各主题分析结果

根据上文所述研究方法, 本研究共从维基百科采集生育相关词条 150 个, 通过开放式编码将所有词条根据其主题分为三大类, 每个大类对应一个主题。三个主题分别为计划生育和生育健康、人类和环境、以及人口问题。表 1 列举了每个类别的主题、对于主题的描述以及该类别下的词条数。通过该表可看出, 关于计划生育和生育健康主题的词条远远多于其他两个主题, 表明公众对于该主题的相关政策法规、措施方法、机构、服务和研究等都更为关注。

表 1 计划生育相关主题和词条数

主题	描述	词条数
计划生育和生育健康	与计划生育和生育健康政策、法律、研究、方法、服务、人员、组织等相关的词条	95
人类和环境	与人权、人类生活、人类社会和环境等相关的词条	28
人口问题	与人口相关的词条	27

表 2 统计了从 2010 至 2017 年各年度计划生育相关词条的增长情况及每个主题下的新增词条数量。该表显示总体相关词条的增长主要体现在计划生育和生育健康主题相关词条的增加, 其中 2011 年和 2012 年增长最为迅速, 其他年份的增长较为平稳。人类和环境及人口问题两个主题相关词条从 2010 至 2017 年基本没有增长。表 2 反映从 2010 至 2017 年公众同样对计划生育和生育健康的相关内容更加关注, 且在 2011 年和 2012 年关注度暴增。

根据对 2010 至 2017 年各主题产生的词条的浏览发现, 计划生育和生育健康主题中的新词条主要是关于生育相关组织和机构(如 Society for Family Health Nigeria)、科学研究(如 Journal of Family Planning and Reproductive Health Care)、政策法规(如 Father's quota)、技术手段(如 Fertility monitor), 以及不同地区的计划生育历史、发展和现状(如 Family planning in the United

States), 涉及到不同方面。而人口问题主题的新词条主要关于组织机构和研究人員, 相对而言内容更为集中。

表2 各年度新增计划生育相关词条数量

主题	2010	2011	2012	2013	2014	2015	2016	2017
计划生育和生育健康	2	9	13	2	2	2	2	5
人类和环境	0	0	1	0	0	0	0	0
人口问题	0	1	0	0	3	0	0	0
总计	2	10	14	2	5	2	2	5

本研究对八年间的生育相关词条编辑次数按年度进行统计, 图2呈现了编辑次数的变化趋势。图2中横坐标代表年份, 左边纵坐标对应各主题编辑次数, 右边纵坐标对应总体编辑次数。该图显示总体编辑次数呈下降趋势, 除2011年编辑量突增之外, 其余年份的编辑量稳中有降。所有主题中, 仍然是计划生育和生育健康主题编辑量最高, 但下降趋势明显。人类与环境主题编辑量缓慢下降, 人口问题编辑量则在波动中略微上升。结合表2和图2可知, 从维基百科编辑者角度看, 对于生育的关注近年来逐步下降, 对于人类和环境及人口问题主题的关注度较低, 而对于生育健康的关注相对较高。

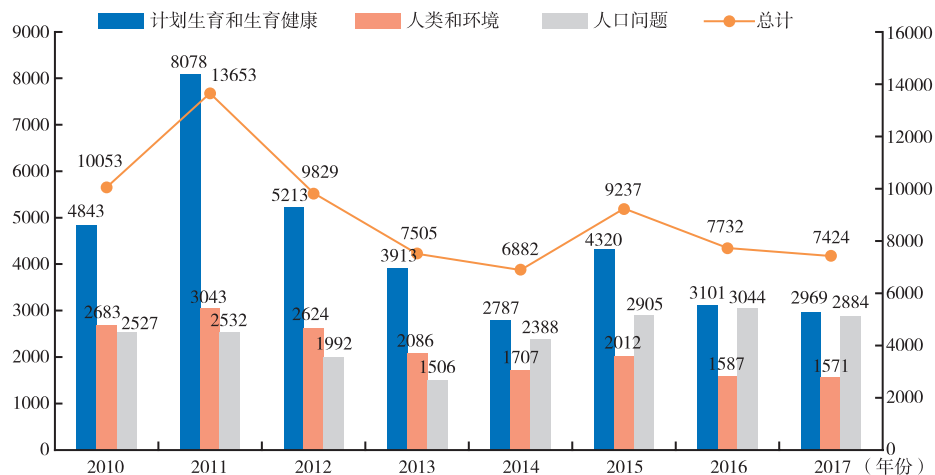


图2 2010至2017年各主题编辑次数

图3显示相关词条的浏览量变化趋势, 图中数字为科学计数法表示的浏览量(以百万为单位)。图3中横坐标代表年份, 左边纵坐标对应各主题浏览次数, 右边纵坐标对应总体浏览次数。由图3可见总体浏览量从2010至2013年呈上升趋势, 而自2013年起逐年下降。计划生育和生育健康主题浏览趋势与总体趋势一致, 且总体而言浏览量高于其他两个主题, 只在2016和2017年被人口问题主题的浏览量赶超。人口问题主题浏览量总体呈下降趋势, 但下降速度较计划生育和生育健康主题较为缓慢。人类和环境主题浏览量低于其他主题, 从2010至2013年浏览量有所



上升, 而此后逐年下降。该图反映维基百科用户对生育的关注度先升后降, 并在 2012 年后逐年下降。三个主题中, 一般用户前期更为关注计划生育和生育健康, 2016 和 2017 年更为关注人口问题。通过对比编辑次数和浏览次数可知, 维基百科编辑和一般用户的关注重点大致上保持一致, 但关注度的变化趋势有所不同, 说明两个群体存在差异。Wang 和 Zhang 对于女性健康相关话题的研究也得到了类似的结论<sup>[19]</sup>。

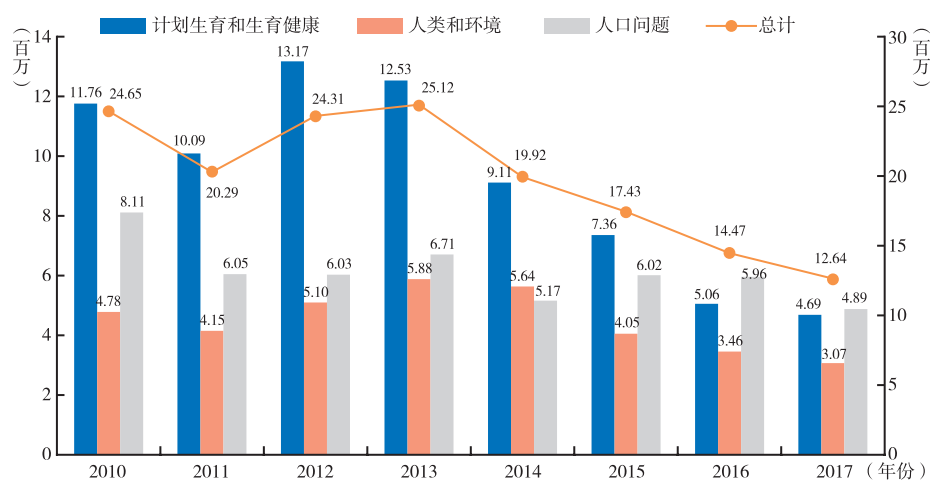


图 3 2010 至 2017 年各主题浏览次数

## 4.2 热点话题分析

除通过词条数量、编辑次数和浏览次数情况反映各主题受关注程度之外, 本文从内容角度对生育的相关主题进行了文本分析, 并提取出各主体的热点话题。基于上文中的三个主题及每个主题包含的相关词条, 本研究采用自组织地图映射方法对每个大类中的词条进行聚类, 图 4-6 显示聚类分析结果。每张图中每个数字代表一个词条, 每个词条所对应的编号在文末附录中 (其中第一列为主题, 第二列为每个主题所包含的词条, 每个词条按首字母顺序排序并编号)。图 4-6 中右边的色彩条中的不同颜色代表不同的 U-matrix 值, U-matrix 值越小则相似度越高。图中的矩形或多边形则代表所得到的簇, 每个矩形 / 多边形中的数字代表属于这个簇的词条。

基于聚类结果, 本研究采用 n-gram 模型对每个包含 3 个以上词条的簇中的词条进行文本分析, 从中提取高频词或短语, 并根据人工主题分析提炼出对应的热点话题。在此过程中研究人员针对每个热点话题的相关高频词 / 短语进行了进一步的分析, 归纳出了它们所涉及的热点话题的具体方面, 如高频词 female condom (女性避孕套) 是热点话题“计划生育和生殖”的“工具”。表 3-5 列出了三个主题的文本分析结果, 每个表的第一列为所得到的簇; 第二列为每个簇的高频词或短语, 每个词或短语之后括号中的数字为出现次数; 第三列为热点话题, 每个热点话题后的括号中的内容代表该话题所涉及的生育信息的具体方面。本文中只列出每个簇中出现频率最高的 15 个词 / 短语或词频高于 4 的词 / 短语 (当词频高于 4 的词 / 短语不足 15 个时)。

### 4.2.1 计划生育和生育健康主题热点话题

计划生育和生育健康主题的相关词条经过自组织映射方法聚类分析共得到 12 簇, 其中 9 个

簇包含3个以上词条，聚类结果在图4中呈现。由该图可知C1, C2, C5, C6, C8和C9彼此距离较近，且处于相连的蓝色区域，因此这些簇所包含的词条内容较为相似。C7虽然与C6等距离较近，但因C7与其他簇之间为黄色区域，因此相似性相对较低。C3和C4彼此相邻，但与其他簇相距较远，因此这两个簇相似度较高，但与其他簇相似度较低。

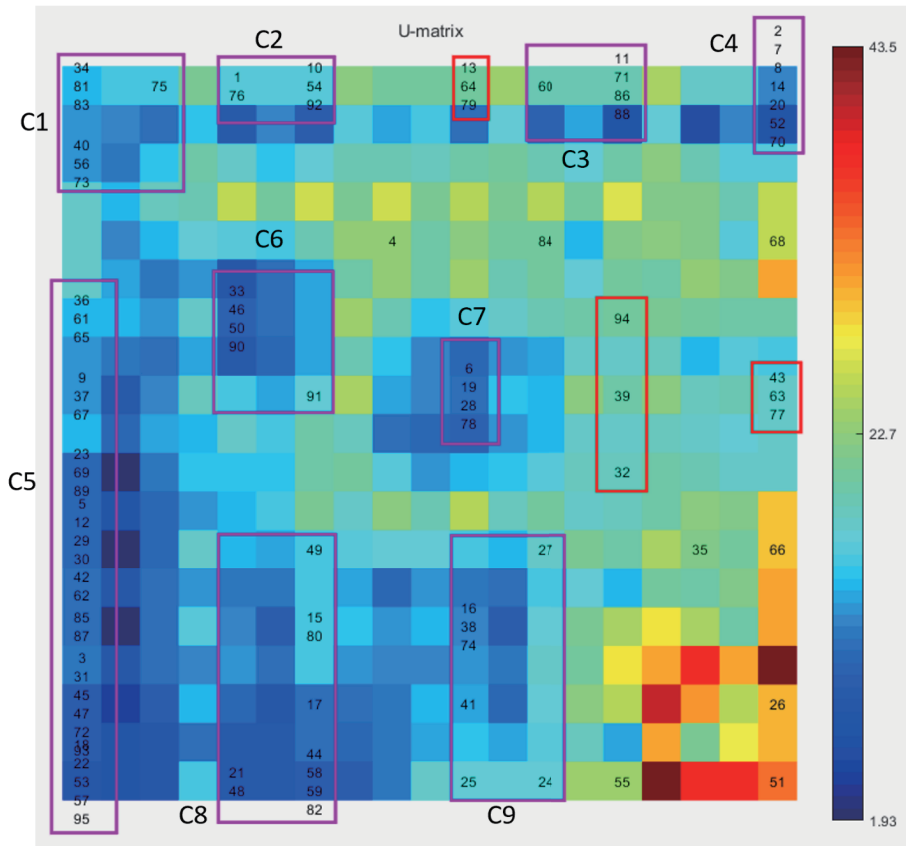


图4 计划生育和生育健康主题聚类结果

注：C5簇下端重叠的数字为18、93。

基于聚类分析结果，表3列出了每个簇中的高频词/短语以及相应的热点话题。该表显示计划生育和生殖话题出现在每个簇中，涉及到节育政策，机构，教育和研究，措施、技术和工具，以及宗教信仰相关内容。该现象说明公众对节育较为重视且关注面涉及较广，不仅仅限于实用范围。

另一热点话题为妇女保护，涉及到法律和新闻报道，其中关于堕胎的权力、法律和舆论的讨论较多，一定程度上反映出公众对妇女保护的关注以及相关工作的进步。而新闻报道是引发妇女问题获得社会关注和促进妇女保护事业进步的重要方式，也是对相关法律法规、社会事件等进行传播的主要途径。

计划生育和生育健康主题中也涉及人口相关问题，其中一个关注点是人口相关机构和组织，如 United Nations Fund for Population Activities。这些机构旨在促进人口发展，提升人们生活质量，

以及保护妇女儿童等。这个热点话题与另一个主题人口问题密切相关，也反映了一部分词条虽然主要内容聚焦于某一主题，但在部分内容上也会涉及到其他主题。

**表 3 计划生育和生育健康主题热点话题**

簇	高频词和短语	热点话题
C1	female condom (47), reproductive health (40), sperm donation (33), donor sperm (33), sperm bank (30), sperm donor (21), birth control (21), human rights (19), potential donors (17), domestic violence (17), artificial insemination (15), teen dating violence (15), partner violence (13), United States (13), donor insemination (12), family planning (12)	计划生育和生殖(措施, 工具), 虐待和暴力(家庭暴力, 恋爱暴力)
C2	reproductive rights (81), human rights (37), family planning (37), United States (33), oral health (29), health care (28), maternal mortality (24), birth control (22), maternal health (20), rhythm method (19), United Nations (19), health organization (19), sexual and reproductive health (17), maternal deaths (16), developing countries (14), calendar-based methods (14), standard days method (14)	计划生育和生殖(措施), 健康相关问题(问题, 机构), 妇女保护(法律)
C3	John Paul (61), birth control (53), family planning (51), Theology of the Body (39), Catholic Church (36), natural family (29), Supreme Court (27), natural family planning (22), United States (18), reproductive health (10), sexual intercourse (10), Humanae Vitae (9), Quiverfull adherents (9), periodic abstinence (8), the Bible (8), reproductive rights (8)	妇女保护(法律), 计划生育和生殖(宗教)
C4	birth control (364), United States (109), Margaret Sanger (92), New York (88), family planning (50), Planned Parenthood (36), reproductive health (35), health organization (33), induced abortion (23), latex condoms (18), first trimester (17), medical abortion (17), health care (17), unsafe abortion (17), Woman Rebel (16)	计划生育和生殖(机构, 措施, 工具), 健康相关问题(问题), 妇女保护(法律, 新闻报道)
C5	family planning (116), Title X (57), Planned Parenthood (44), health care (22), United States (22), reproductive health (19), sexual health (19), baby bonus (17), Hong Kong (16), family planning services (16), Family Planning Association (15), International Conference (14), sexuality education (14), Planned Parenthood Federation (14), sperm banks (12), public health (12), birth control (12)	计划生育和生殖(政策, 机构, 措施), 健康相关问题(服务)
C6	Planned Parenthood (22), family planning (21), United Nations (18), father's quota (16), parental leave (10), paternity leave (8), United States (8), reproductive health (7), Conservative Party (5), Population Fund (5), Nations Fund for Population (5), Jens Stoltenberg (4), Trivers-Willard hypothesis (4), coercive abortion (4), Democratic Party (4), human rights (4), State Department (4)	计划生育和生殖(机构, 政策, 措施, 研究), 人口相关问题(机构)
C7	family planning (63), domestic violence (32), sex education (30), birth control (13), Billings Ovulation Method (12), reproductive health (9), violence during pregnancy (9), married women (8), cervical mucus (8), health survey (7), public health (7), research and development (7), development solutions (6), Family Planning Association (6), Research and Development Solutions (6)	计划生育和生殖(教育, 措施, 政策), 虐待和暴力(家庭暴力), 健康相关问题
C8	family planning (41), reproductive health (13), leave of absence (12), Family Planning Commission (9), Population and Family Planning (9), Health and Family Planning (8), Rama Rau (6), commuted leave (6), Journal of Family Planning (6), Commodity Security (5), reproductive healthcare (5), contraceptive security (5), International Development (5), National Population (5), Sexual and Reproductive Healthcare (5)	计划生育和生殖(机构, 政策, 研究, 技术), 健康相关问题(服务), 人口相关问题
C9	reproductive health (44), family planning (39), Couple to Couple League (16), Deutsche Stiftung Weltbevölkerung (15), Essential Medicines (8), Iran's population (7), United Nations (7), fertility rate (7), population growth (7), birth control (6), natural family planning (6), Reproductive Health Supplies Coalition (6), growth rate (5), birth rate (5), German foundation (5), sexual and reproductive health (5)	计划生育和生殖(机构), 健康相关问题(治疗手段), 人口相关问题(机构)

另两个热点话题，健康相关问题与以上三个热点话题都相关，而虐待和暴力与生育、妇女保护相关。其中，虐待和暴力话题中频繁出现的家庭暴力（domestic abuse）是世界范围内都广泛存

在并受关注的问题。

### 4.2.2 人类和环境主题热点话题

人类和环境主题的相关词条经过聚类分析共得到2个簇，聚类结果在图5中呈现。由该图可知两个簇距离较近，但中间相隔区域为浅蓝色和黄色，表明两个簇中所包含的词条内容相似性不高。基于这些发现，表4列出了每类中的高频词/短语以及相应的热点话题，这些话题与上文中的热点话题存在很大不同。除人口问题这一热点话题与其他两个主题都相关之外，未来学、环境问题、社会因素、经济、军事、政治都是与人类命运休戚相关的较为宏观的问题。其中环境问题对于公众较为熟悉的一些问题都有涉及，比如全球变暖（global warming）和巨灾风险（catastrophic risks）。

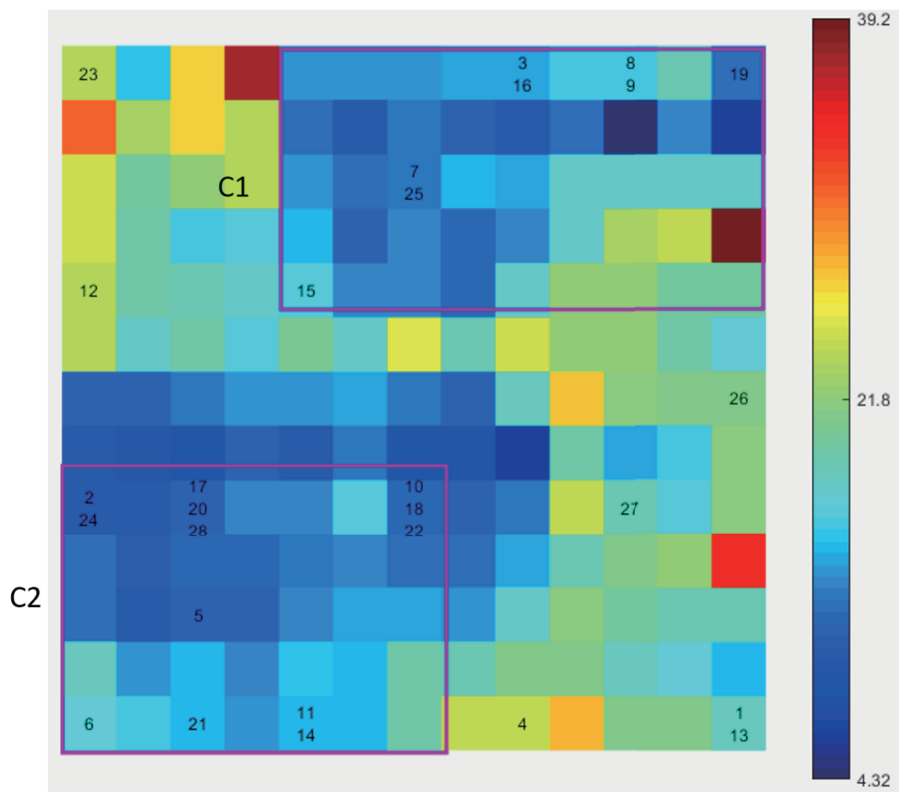


图5 人类和环境主题聚类结果

表4 人类和环境主题热点话题

簇	高频词和短语	热点话题
C1	futures studies (94), planetary boundaries (67), Gaia Hypothesis (62), climate change (46), tragedy of the commons (45), United States (28), earth system science (27), United Nations (21), carbon dioxide (16), global warming (14), existential risk (14), catastrophic risks (12), existential risks (11), artificial intelligence (11), population growth (11), human extinction (11)	未来学（技术），环境问题（研究，问题），人口相关问题
C2	United States (80), forced marriage (75), identity politics (42), financial crisis (35), fertility rate (31), human rights (25), reserve army (24), political demography (24), World Bank (23), men and women (23), maternity leave (22), people smuggling (21), New Feminists (21), migrant smuggling (19), Latin America (17)	社会因素，人口相关问题，经济，军事，政治

### 4.2.3 人口问题主题热点话题

人口问题主题经聚类分析得到 3 个簇, 结果如图 6 所示。该图显示 3 个簇距离较近且中间相隔区域为蓝色区域, 因此簇中所包含的词条内容较为相似。表 5 中列出了其中两个较大簇的文本分析结果, 从中可发现该主题下词条的相关热门话题与计划生育和生殖健康话题较为相关, 对节育的方式有一定关注。另一个热门话题为人口相关问题, 与其他主题中涉及到的组织机构不同, 这个主题中的相关词条内容主要聚焦于当前存在的问题以及问题的成因, 例如, 食品产量的增加和国际移民都被认为是造成部分地区人口增长的原因。

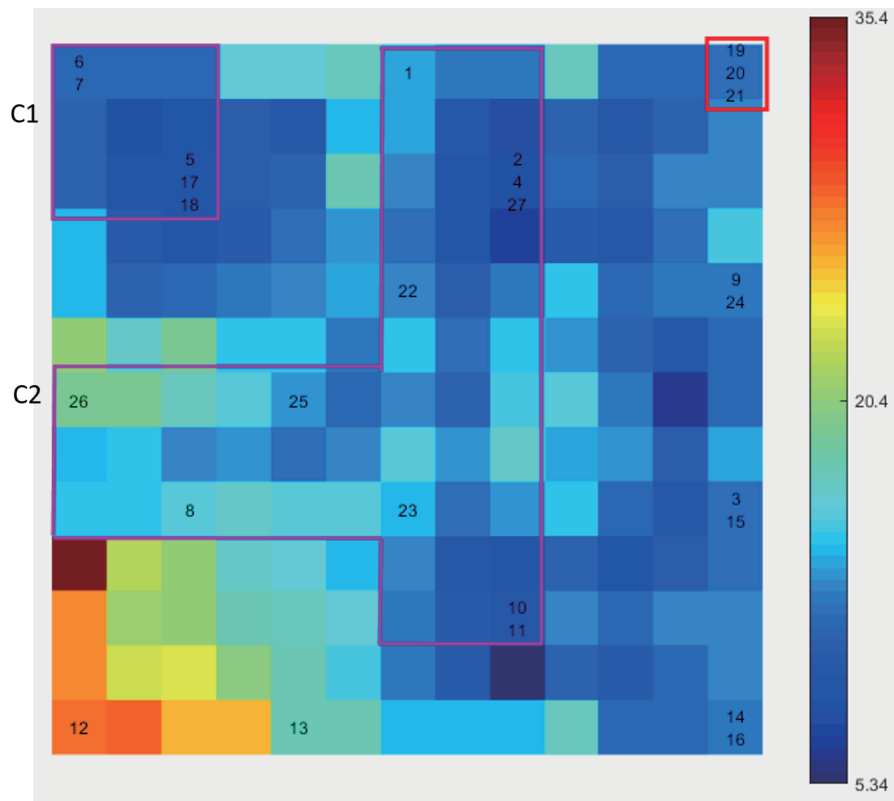


图 6 人口问题主题聚类结果

表 5 人口问题主题热点话题

簇	高频词和短语	热点话题
C1	population growth (75), world population (47), human population (33), United Nations (29), population planning (25), United States (21), birth control (16), family planning (16), developed countries (14), birth rates (14), carrying capacity (12), Club of Rome (12), developing countries (11), food production (11), population density (11), population control (11), international migration (11), human migration (11)	人口相关问题(问题, 原因), 计划生育和生殖健康(措施)
C2	world population (100), sex ratio (91), population growth (63), official population (54), data templates (44), United Nations (29), country's population (28), United States (25), official sources (22), population density (22), manual calculation (22), population is not accurate (22), automatically calculate today's population (22), human population (19), population prospects (19), template without providing justification (19)	人口相关问题(统计数据)

根据上文对每个主题相关词条的文本分析可见,主题与热点话题之间呈现网状关系,表6归纳了主题与热点话题的对应关系。由表6可见,计划生育和生殖健康主题、人类和环境主题的热点话题都非常多样化,彼此间的重合度也较小,侧面反映出两个主题在英文维基百科上的讨论内容差异较大。而人口问题主题的热点话题非常集中,且两个热点话题分别与前两个主题相关。

表6 热点话题分布

主题	热点话题					
	计划生育和生殖健康	虐待和暴力	健康相关问题	妇女保护	人口相关问题	
计划生育和生育健康	√	√	√	√	√	
人类和环境	×	×	×	×	√	
人口问题	√	×	×	×	√	
-	未来学	环境问题	社会因素	经济	军事	政治
计划生育和生育健康	×	×	×	×	×	×
人类和环境	√	√	√	√	√	√
人口问题	×	×	×	×	×	×

### 4.3 热门话题演变分析

为探索生育相关热门话题在2010至2017年间的演变,本研究将整个时间跨度分为4个时间段,分别为2010至2011年(P1)、2012至2013年(P2)、2014至2015年(P3),以及2016至2017年(P4)。首先对比每个主题不同时期文本,通过文本分析将变化最大的内容提取出来,再通过n-gram模型和人工主题分析将各个时期增加或减少的词或短语提取出来,并基于上文中发现的热门话题分析其演变。表7-9分别列出了3个主题从一个时期到下一个时期增减幅度最大的词/短语,以及由此得到的增长或消亡的热门话题。每个表的第一列表示不同时期的对比,第二列表示词频的增加或减少,第三列表示增加/减少的词/短语以及增加/减少的频数(显示在单词之后的括号中,负数代表词频减少,整数代表词频增加),第四列代表从第三列相应单元格中提炼出的热门话题,若词/短语频数持续减少则表示其对应的热门话题逐渐消亡;若词/短语频数持续增加则表示其对应的热门话题逐渐增长。

#### 4.3.1 计划生育和生殖健康热点话题演变

表7表示计划生育和生殖健康主题中热门话题在不同时期的演变。该表显示总体来说健康相关问题、计划生育和生殖话题和人口相关问题的内容增长最多,可见内容生产者对这些内容保持了兴趣。这些话题中关于健康和人口组织机构(如Planned Parenthood clinics和United Nations Population Fund),以及生育生殖政策法规(如亲子假parental leave,产假maternity leave和堕胎权力abortion rights)的关注逐步增长。而唯一在近年关注度下降的是宗教相关内容,该部分内容主要关注宗教对避孕和堕胎的限制,如只可采取自然手段避孕,并将堕胎等同于谋杀。这些发现反映了内容生产者对生育和生殖权力的关注上升,而对限制这些权力的宗教和传统关注减少。

表 7 计划生育和生殖健康热点话题演变

时期	高频词 / 短语		热点话题
P1 VS. P2	词频减少	Disease Control and Prevention (-46), the eugenics (-21), German Foundation for World (-14), World Population (-8), human heredity (-8), World War (-7), 20th century (-7), sperm donors (-6), Population Research (-6), birth control politics (-6), birth control movement (-6), teenage pregnancy (-5), Couple to Couple League (-5), University of California (-5), history of birth control (-5)	健康相关问题, 计划生育和生殖, 人口相关问题, 妇女保护
	词频增加	sex ratio (120), family planning (95), reproductive health (86), infant mortality (66), birth control (62), United States (59), Planned Parenthood (41), female condom (31), sex-selective abortion (28), reproductive coercion (27), domestic violence (25), health care (23), mortality rate (22), International Conference (19), United Nations (18), one-child policy (17), developing countries (17), health organization (16), prenatal care (16), ratio at birth (15)	计划生育和生殖, 健康相关问题, 虐待和暴力, 人口相关问题
P2 VS. P3	词频减少	reproductive coercion (-17), paid leave (-17), domestic violence (-12), birth control sabotage (-6)	计划生育和生殖, 虐待和暴力
	词频增加	family planning (67), sex ratio (41), parental leave (40), United States (40), Planned Parenthood (31), New York (31), birth control (29), maternity leave (25), birth sex (22), United Nations Population Fund (19), human rights (18), infant mortality (17), sex selection (17), women's health (17), sex-selective abortion (16), reproductive rights (14), health organization (13), men and women (13), ovulation method (13), Billings Ovulation Method (11)	计划生育和生殖, 健康相关问题, 人口相关问题, 妇女保护
P3 VS. P4	词频减少	Catholic church (-15), termination of pregnancy (-10), ovulation method (-6), population control (-6), Roman Catholic church (-5)	计划生育和生殖, 人口相关问题
	词频增加	family planning (82), Planned Parenthood (72), birth control (50), United States (42), reproductive health (37), maternity leave (32), one-child policy (31), health care (31), parental leave (30), sex education (29), human rights (23), United Nations Population Fund (20), Supreme Court (20), New York (16), mortality rate (16), infant mortality (14), sex-selective abortion (14), sex ratio (13), sexuality education (11), violence against women (11), Hong Kong (11)	计划生育和生殖, 健康相关问题, 人口相关问题, 虐待和暴力, 妇女保护

#### 4.3.2 人类和环境热点话题演变

表 8 展示了人类和环境主题各个时期的词频和热点话题变化, 可见随着时间推移该主题所包含的话题逐渐增多, 如人类发展和不平等与歧视等新的关注点分别在第三时期和第四时期产生, 这反映了内容生产者的关注点在逐渐变得更加多样化。社会因素话题虽有部分相关内容减少 (如家庭问题 family issue 和生活方式 lifestyle), 但总体而言相关内容在观察期间逐步增长。同时, 虐待和暴力、环境相关问题、人口相关问题、未来学等话题相关内容也保持持续增长。

表 8 人类和环境热点话题演变

时期	高频词 / 短语		热点话题
P1 VS. P2	词频减少	parenting styles (-9), Gaia hypothesis (-8), social protection (-5)	环境问题, 社会因素
	词频增加	a comedy series (28), planetary boundaries (22), forced marriage (21), United States (18), voluntary human extinction (16), New York (14), Jean Stapleton (13), financial crisis (12), global recession (11), Carroll O'Connor (10), United Nations (9), banking system (9), Rob Reiner (8), identity politics (7), maternity leave (7), Sally Struthers (7), New York Times (7), fertility rate (6), reserve army (6), middle class (5), conditional cash transfer (5)	社会因素, 暴力和虐待, 人口相关问题, 经济, 不平等和歧视, 计划生育和生殖, 军事

续表

时期	高频词 / 短语		热点话题
P2 VS. P3	词频减少	All in the Family (-6)	社会因素
	词频增加	forced marriage (38), earth system science (26), human rights (17), United States (15), Gaia hypothesis (11), human development (10), existential risk (7), United Nations (6), reserve army (6), New York Times (6), United Kingdom (6), gender equality (6), Futures Research (5), catastrophic risks (5), British Columbia (5)	暴力和虐待, 环境问题, 人类发展, 军事, 未来学, 社会因素
P3 VS. P4	词频减少	work-life balance (-9), New York Times (-6)	社会因素
	词频增加	United States (18), futures studies (16), identity politics (13), fertility rate (12), tragedy of the commons (8), forced marriage (6), United Nations (5), European countries (5), Gaia hypothesis (5), Great Recession (5), New Brunswick (5), strategic foresight (5)	未来学, 不平等和歧视, 计划生育和生殖, 社会因素, 环境问题, 经济, 人口相关问题

### 4.3.3 人口问题热点话题演变

人口问题相关内容词频和热点话题的演变如表9所示, 分析结果表明该主题只有一个热点话题, 即人口相关问题, 涉及内容包括存在的问题、政策和人口普查情况, 可见关注点较为集中, 其中对于人口普查的关注在观察期内逐步下降。

表9 人口问题热点话题演变

时期	高频词 / 短语		热点话题
P1 VS. P2	词频减少	list of countries (-6), Soviet Union (-5), population figures (-5)	人口相关问题
	词频增加	official population (24), sex ratio (27), birth sex ratio (18), population growth (14), world population (11), human sex (7), global population (7), population control (6), New York (6), human population (5), fertility rate (5)	人口相关问题
P2 VS. P3	词频减少	United States (-9), Census Bureau (-6), Population Clock (-5), 19th century (-5), 2008 census (-5), preliminary 2012 census (-5)	人口相关问题
	词频增加	population growth (24), human population (19), official population (14), population control (11), carrying capacity (8), world population (7), sovereign states (6)	人口相关问题
P3 VS. P4	词频减少	population control (-28), 2010 census (-23), 2012 census (-21), 2011 census (-21), dependent territories (-7), census result (-6), youth bulge (-5), 2014 census (-5)	人口相关问题
	词频增加	Country's population (22), population growth (19), world population (9), New York (9), list of countries (9), million people (9), United States (8), human overpopulation (7), Cambridge University (6), human population growth (6)	人口相关问题

本文对于生育健康热点话题的分类与 Wang 等<sup>[19]</sup>对女性健康话题的分类相似。Wang 等在进行女性健康主题演化分析时, 基于词频的变化将相关话题分为增长类、波动类和消亡类<sup>[19]</sup>。本研究将相关内容持续增加的话题归为增长类话题, 内容持续减少的话题归为消亡类话题, 而内容在观察期内有增有减则被归为波动类话题, 归纳总结的结果见表10。通过归纳总结发现, 生育相关内容所涉及的话题中主要有增长和波动两类, 而没有消亡类话题, 这反映了内容生产者对于计划生育的各个方面的关注变得更加多元、多样, 也间接反映了用户对于相关信息的需求变得更加多元化和多样化。



表 10 热点话题演变类型

演变类型	热点话题
增长	暴力和虐待, 经济, 计划生育和生殖, 未来学, 健康相关问题, 人类发展, 不公平和歧视, 军事, 人口相关问题, 社会因素
波动	环境问题, 妇女保护

#### 4.4 中外对比讨论

相较于上文中热点话题及其变化趋势的结果而言, 中国由于长期处于低生育水平, 目前对于节育的讨论和关注相对较少, 意外妊娠的案例较多, 如统计显示, 北京地区 40% 左右的二孩是在妇女没有备孕的情况下所得<sup>[31]</sup>。另外由于中国并非宗教国家, 对于堕胎权力的争议和讨论也较少, 反而对高龄产妇人工流产方面的健康风险较为重视<sup>[31]</sup>。同时, 对于生育福利和妇女劳动权益保障的关注也较多, 比如孕产假、生育保险、生育津贴、就业权利保护、特殊劳动保护等<sup>[34]</sup>。

根据对英文维基生育相关词条内容的分析可见, 该平台上明确涉及到中国生育相关的信息较少, 关注最多的还是独生子女政策 (one-child policy) 这一相对来说比较陈旧的政策, 对 2013 年就开始推行的二孩政策并没有予以相应关注, 对于中国生育相关政策的认知较为落后。实际上随着该政策的全面推进, 当前中国社交媒体平台上对于生育权, 尤其是被迫生育的关注明显增加, 而非独生子女政策下的“超生”问题<sup>[41]</sup>。

在二孩生育政策全面推行之后, 社交媒体平台上的相关讨论日益多元化, 探讨的话题包括生育意愿、独生子女压力与保障、女性健康与权益、养老问题、人口结构, 以及衍生的医疗、教育、就业、环境等诸多话题, 这些话题与英文维基百科上多元多样的话题有重合之处, 如环境问题、人口问题等<sup>[41]</sup>。中国对于性别平等的讨论较为突出, 如杨凤娇, 陈曦和锁菁研究发现近年来女性充分参与社交媒体上的社会公共话题讨论, 就女性社会地位、生育权、就业和职场发展等议题发声<sup>[41]</sup>。

英文维基百科中提到移民对人口增长的正面影响, 但也担忧随之产生的安全问题。中国对于流动人口对生育和人口影响也展开了研究, 但研究视角主要为流动人口生育意愿及影响因素。研究结果显示, 母亲受教育水平、夫妻结婚年龄和时长对生育意愿有负向影响, 家庭第一孩的性别对于流动人口是否生育两孩有显著影响<sup>[42]</sup>。

## 5 结语

生育是在世界范围内被广泛关注的公共健康话题之一, 在二孩生育政策全面推进的当下, 也是中国社会各界广泛关注的热点话题。该话题不仅仅涉及到婚育、生殖健康、医疗保健、教育、养老等密切相关的问题, 还涉及到人口结构、社会公平、国家发展等宏观问题。本研究基于英文维基百科的生育相关公共健康信息和数据, 对国外生育相关主题和热点话题及其变化趋势进行了较为全面和深入的分析。

研究结果显示国外公众主要关注 3 个相关主题, 分别为计划生育和生育健康、人类和环境、以及人口问题。随着时间的推移, 关注的热点话题变得日益多样化, 包括生殖健康、虐待和暴

力、妇女保护、经济、政治、军事、环境和未来学等。然而从关注度的角度分析可见,无论是内容生产者还是用户对此类话题的关注度都在下降,一方面是由于内容已经相对完善,且没有突发的新鲜话题和内容,无需大量更新;另一方面则说明公众的兴趣在稳定中下降。通过中外对比发现,中外公众在关注的主题和话题上有重合,但侧重点不同。英文维基百科上中国的生育相关内容较为陈旧,提醒相关人员在政策发布时需及时宣传,增进外国公众对我国现状的了解。

本研究的一个不足之处是分析热点话题演化时,对不同时期的维基百科词条文本进行对比,然而词条的编辑改动主要是对现有内容的勘误修改和扩展丰富,较少出现对词条内容的大面积删除,因此只发现了增长类和波动类话题,没有消亡类话题。这是由于所使用的数据的局限性造成的,在未来的研究中将采用更多其他数据对热点话题的演化进行探索。本次研究由于时间和篇幅有限,仅采用“family planning”作为关键词进行搜索,以“Family planning”词条作为中心词条进行相关词条的搜寻,仅对英文维基百科上的相关内容进行分析和讨论。未来将采用更多方式和手段进行数据采集,拓展研究范围,对多个社交媒体平台的不同公共健康话题进行研究和对比,从中发现更普遍的内外部发展变化规律。

#### 【参考文献】

- [1] 穆光宗. “全面二孩”政策实施效果和前景 [J]. 中国经济报告, 2017, (1):24-26.
- [2] 陈卫, 段媛媛. 中国近 10 年来的生育水平与趋势 [J]. 人口研究, 2019, 43(01):3-17.
- [3] HAWN C. Take two aspirin and tweet me in the morning: How Twitter, Facebook, and other social media are reshaping health care [J]. Health Affairs, 2009.
- [4] 黄令贺. 网络百科信息质量动态演进探索研究 [J]. 情报理论与实践, 2018, 41(02):127-131.
- [5] GILES J. Internet encyclopaedias go head to head [J]. Nature, 2005, 438(7070):900-901.
- [6] 陆泉, 江超, 陈静. 基于扩展疾病本体的电子病历大数据组织研究 [J]. 图书情报知识, 2019(01):109-118.
- [7] AHLQVIST T, BÄCK A, HALONEN M, et al. Social media road maps exploring the futures triggered by social media [J]. VTT Tiedotteita-Valtion Teknillinen Tutkimuskeskus, 2008, 2454:13.
- [8] KAPLAN A M, HAENLEIN M. Users of the world, unite! The challenges and opportunities of Social Media [J]. Business Horizons, 2010, 53(1):59-68.
- [9] MILNE D, WITTEN I H. Proceedings of the 17th ACM Conference on Information and Knowledge Management, October 26-30, 2008 [C]. New York: Association for Computing Machinery, 2008.
- [10] MILNE D, WITTEN I H. An open-source toolkit for mining Wikipedia [J]. Artificial Intelligence, 2013, 194:222-239.
- [11] ZHANG J, ZHAO Y. A user term visualization analysis based on a social question and answer log [J]. Information Processing & Management, 2013, 49(5):1019-1048.
- [12] ZHANG J, ZHENG Z, WANG Y, et al. A study on users' preference towards diabetes-related video clips on YouTube [J]. BMC Medical Informatics and Decision Making, 2020, 20(1).
- [13] DAWSON J. Doctors join patients in going online for health information [J]. New Media Age, 2010, 7.
- [14] MARAR S D, AL-MADANEY M M, ALMOUSAWI F H. Health information on social media. Perceptions, attitudes, and practices of patients and their companions [J]. Saudi Medical Journal, 2019, 40(12):1294-1299.
- [15] 李月琳, 张秀, 王姗姗. 社交媒体健康信息质量研究: 基于真伪健康信息特征的分析 [J]. 情报学报,

2018, 37(3):294-304.

[ 16 ] HAMM M P, CHISHOLM A, SHULHAN J, et al. Social media use by health care professionals and trainees: a scoping review [J]. *Academic Medicine Journal of the Association of American Medical Colleges*, 2013, 88(9):1376-1383.

[ 17 ] GREENE J A, CHOUDHRY N K, KILABUK E, et al. Online social networking by patients with diabetes: a qualitative evaluation of communication with facebook [J]. *Journal of General Internal Medicine*, 2011, 26(3): 287-292.

[ 18 ] 张敏, 车雨霏, 张艳. 社交媒体健康信息传播行为研究系统综述 [J]. *图书馆*, 2019, (5):33-40,79.

[ 19 ] WANG Y, ZHANG J. Investigation of women's health on Wikipedia—A temporal analysis of women's health topic [J]. *Informatics*, 2020, 7(3):22.

[ 20 ] HOLTZ P, FERAHU B, KIMMERLE J. Effects of contributor experience on the quality of health-related Wikipedia articles [J]. *Journal of Medical Internet Research*, 2018, 20(5):e171.

[ 21 ] HUISMAN M, JOYE S, BILTEREYST D. Health on Wikipedia: a qualitative study of the attitudes, perceptions, and use of Wikipedia as a source of health information by middle-aged and older adults [J]. *Information Communication and Society*, 2020.

[ 22 ] SMITH D A. Situating Wikipedia as a health information resource in various contexts: A scoping review. *PLoS ONE*, 2020, 15(2):1-19.

[ 23 ] FUEMMELE B F, BEHRMAN P, TAYLOR M, et al. Child and family health in the era of prevention: new opportunities and challenges [J]. *Journal of Behavioral Medicine*, 2016, 40(1):159-174.

[ 24 ] HALFON N, LARSON K, LU M, et al. Lifecourse health development: Past, present and future [J]. *Maternal & Child Health Journal*, 2014, 18(2):344-365.

[ 25 ] 翟绍果, 王昭茜. 公共健康治理的历史逻辑、机制框架与实现策略 [J]. *山东社会科学*, 2018, (7):95-101.

[ 26 ] KOTCH J. *Maternal and Child Health: Programs, Problems, and Policy in Public Health* [M]. Burlington, Jones & Bartlett Learning, 2005.

[ 27 ] BRASIL P, PEREIRA J P, MOREIRA M E, et al. Zika Virus Infection in Pregnant Women in Rio de Janeiro [J]. *New England Journal of Medicine*, 2016:2321.

[ 28 ] MARCUS S, FLYNN H A, BLOW F C, et al. Depressive symptoms among pregnant women screened in obstetrics settings [J]. *Journal of Women's Health*, 2003, 12(4):373-380.

[ 29 ] SUBRAMANIAM M, PRASAD R O, ABDIN E, et al. Single mothers have a higher risk of mood disorders [J]. *Annals of the Academy of Medicine Singapore*, 2014, 43(3):145-151.

[ 30 ] COHEN J J, BLEVINS M, MAPENZI A, et al. Overcoming the perceived barriers to health care access among single mothers in coastal Kenya [J]. *International Journal of Public Health*, 2014, 59(1):189-196.

[ 31 ] 祁静, 茅倬彦, 申小菊. 生育二孩妇女的孕产保健需求及其影响因素——以北京市为例 [J]. *人口与发展*, 2018, (1):85-93, 108.

[ 32 ] 祁静, 茅倬彦, 顾宝昌. 北京市生育二孩妇女孕产保健需求分析 [J]. *中国计划生育学杂志*, 2017, 25(2):88-92.

[ 33 ] 张永梅. 职业女性的两孩生育意愿及其政策启示——基于“知识女性”与“其他女性”比较的视角 [J]. *浙江社会科学*, 2018, (12):68-76, 59, 157.

[ 34 ] 韩振燕, 王中汉. 妇女福利政策对城市女性二孩生育意愿的影响研究——基于全国十地区城市育龄女性的调查 [J]. *中国人力资源开发*, 2017, (9):139-146.

[ 35 ] 朱丽丽, 尹文强, 宋春燕, 等. 我国省级卫生和计划生育委员会官方微博应用现状分析 [J]. *中华医院管理杂志*, 2018, (3):231-236.

[ 36 ] ARDEN M A, DUXBURY A M, SOLTANI H. Responses to gestational weight management guidance: a thematic

analysis of comments made by women in online parenting forums [J]. BMC Pregnancy and Childbirth, 2014, 14, 216.

[37] KEYES O. Package 'WikipediR' [EB/OL]. [2020-06-20] <http://www.stats.bris.ac.uk/R/web/packages/WikipediR/WikipediR.pdf>.

[38] 王念祖. 扎根理论三阶段编码对主题词提取的应用研究 [J]. 图书馆杂志, 2018, 37(05):74-81.

[39] BUCKLEY S C. Term-weighting approaches in automatic text retrieval [J]. Information Processing & Management, 1988, 24(5):513-523.

[40] KOHONEN T, KASKI S, LAGUS K, et al. Self organization of a massive document collection [J]. IEEE Transactions on Neural Networks, 2000, 11(3):574-585.

[41] 杨凤娇, 陈曦, 锁菁. 微博视域下女性在社会公共话题中的话语表达——基于新浪微博对“全面二孩”的讨论 [J]. 现代传播, 2017, 39(2):139-143.

[42] 王晓宇, 原新, 韩昱洁. 家庭生育决策与全面两孩政策——基于流动人口的视角 [J]. 南开经济研究, 2018, (2):93-109.

## 【附录】

主题	词条
Family planning and reproductive health: 95 个词条	(1) Abortion-rights movements, (2) Abortion, (3) American Family Planning, (4) Avabai Bomanji Wadia, (5) Baby bonus, (6) Billings ovulation method, (7) Birth control movement in the United States, (8) Birth control, (9) British Pregnancy Advisory Service, (10) Calendar-based contraceptive methods, (11) Catholic theology of sexuality, (12) Cecile Richards, (13) Childbirth in Sri Lanka, (14) Condom, (15) Contraceptive security, (16) Couple to Couple League, (17) Dhanvanthi Rama Rau, (18) Directorate General of Family Planning, (19) Domestic violence and pregnancy, (20) Eugenics, (21) Faculty of Sexual and Reproductive Healthcare, (22) Family Planning Association of India, (23) Family Planning Association, (24) Family planning in Bangladesh, (25) Family planning in Hong Kong, (26) Family planning in India, (27) Family planning in Iran, (28) Family planning in Pakistan, (29) Family planning in the United States, (30) Family planning policy, (31) Family Planning Queensland, (32) Family planning, (33) Father's quota, (34) Female condom, (35) Fertility and intelligence, (36) Fertility testing, (37) Forced abortion, (38) Forced pregnancy, (39) Gender inequality in Bolivia, (40) Genetic diagnosis of intersex, (41) German Foundation for World Population, (42) Individual Family Service Plan, (43) Infant mortality, (44) International Perspectives on Sexual and Reproductive Health, (45) International Planned Parenthood Federation, (46) Irish Family Planning Association, (47) Japan Family Planning Association, (48) Journal of Family Planning and Reproductive Health Care, (49) Leave of absence, (50) Lila Rose, (51) List of sovereign states and dependent territories by birth rate, (52) Margaret Sanger, (53) Marie Stopes International, (54) Maternal health, (55) Maternity leave and the Organisation for Economic Co-operation and Development, (56) Natalism, (57) National Alliance for Optional Parenthood, (58) National Health and Family Planning Commission, (59) National Population and Family Planning Commission, (60) Natural family planning, (61) Non-consensual condom removal, (62) Office of Population Affairs, (63) One-child policy, (64) Only child, (65) Onselling of sperm, (66) Parental leave, (67) Pharmaceutical fraud, (68) Planned Parenthood, (69) POPLINE, (70) Pregnancy from rape, (71) Quiverfull, (72) Red Triangle (family planning), (73) Reproductive coercion, (74) Reproductive Health Supplies Coalition, (75) Reproductive health, (76) Reproductive rights, (77) Sex-selective abortion, (78) Sex education in India, (79) Sex selection, (80) Society for Family Health Nigeria, (81) Sperm donation, (82) Studies in Family Planning, (83) Teen dating violence, (84) Teenage pregnancy, (85) The Family Planning Association of Hong Kong, (86) Theology of the Body, (87) Time bind, (88) Timeline of reproductive rights legislation, (89) Title X, (90) Trivers-Willard hypothesis, (91) United Nations Population Fund, (92) United States pro-choice movement, (93) US Family Health Plan, (94) Voluntary childlessness, (95) Yayasan Cipta Cara Padu

续表

主题	词条
Human and environment: 28 个词条	(1) All in the Family, (2) Conditional cash transfer, (3) Earth system science, (4) Family Day (Canada), (5) Family, (6) Forced marriage, (7) Futures studies, (8) Gaia hypothesis, (9) Global catastrophic risk, (10) Great Recession, (11) Identity politics, (12) International development, (13) Koyaanisqatsi, (14) New feminism, (15) Overconsumption, (16) Overexploitation, (17) Parenting, (18) People smuggling, (19) Planetary boundaries, (20) Political demography, (21) Pregnancy discrimination, (22) Reserve army of labour, (23) Societal collapse, (24) Sub-replacement fertility, (25) Tragedy of the commons, (26) Voluntary Human Extinction Movement, (27) What a Way to Go- Life at the End of Empire, (28) Work-life balance
Population problems: 27 个词条	(1) Antinatalism, (2) Behavioral sink, (3) Category-Overpopulation fiction, (4) Demographic trap, (5) Human migration, (6) Human overpopulation, (7) Human population planning, (8) Human sex ratio, (9) International Conference on Population and Development, (10) List of countries and dependencies by population density, (11) List of countries and dependencies by population, (12) List of countries by population growth rate, (13) List of countries by sex ratio, (14) List of organisations campaigning for population stabilisation, (15) List of people who have expressed views relating to overpopulation as a problem, (16) List of population concern organizations, (17) Malthusian trap, (18) Malthusianism, (19) Optimum population, (20) Overpopulation in domestic pets, (21) Overshoot (population), (22) Population ageing, (23) Population density, (24) Population ethics, (25) Population pyramid, (26) World population, (27) Zero population growth

注: 该表系本研究从维基百科采集的 150 个生育相关词条。

## A Trend Analysis of Reproductive-Health-Related Information on English Wikipedia

WANG Yanyan<sup>1,2</sup> ZHANG Jin<sup>3</sup>

- (1. School of Information Resource Management, Renmin University of China, Beijing 100872, China;
2. Humanities Beijing Research Center, Renmin University of China, Beijing 100872, China;
3. School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee WI 53211, USA)

**Abstract:** [ **Purpose/significance** ] Reproductive health is a widely discussed health topic all around the world. With the implementation of the two-child policy in China, the public has paid increasing attentions to reproductive health and related topics. This study explored the internal characteristics and external popularity of the reproductive health and relevant information on English Wikipedia. [ **Method/process** ] One hundred and fifty entries associated to reproductive health were retrieved and their history versions and page views and edits data from 2010 to 2017 were collected. The open coding, trend analysis, and Self-Organizing Map and n-gram approaches were employed for data analysis. [ **Result/conclusion** ] The findings show that the public focused on three themes of the investigated topic, which were the family planning and reproductive health theme, the human and environment theme, and the population problems theme. The popular subjects extracted

( 下转第 083 页 )