

· 综述 ·

维基百科研究综述

赵飞^{1,2}, 周涛^{1,3,4}, 张良^{1,2}, 马鸣卉^{1,5},
刘金虎^{1,6}, 余飞^{1,5}, 查一龙^{1,7}, 李睿琪^{1,7}

1. 电子科技大学互联网科学中心 成都 610054;
2. 电子科技大学经济与管理学院 成都 610054;
3. 中国科学技术大学近代物理系 合肥 230026;
4. 瑞士弗里堡大学物理系 弗里堡 1700;
5. 电子科技大学示范性软件学院 成都 610054;
6. 电子科技大学应用数学学院 成都 610054;
7. 电子科技大学国际化软件人才实验班 成都 610054)

【摘要】对维基百科的相关研究进行了综述。介绍了维基百科的历史发展与特点,宏观统计性质以及演化规律等方面,特别强调了复杂网络的思想和方法在分析维基百科中的应用。讨论了维基百科在社会、经济、文化、教育方面的意义和价值。最后提出了相关研究可能的发展方向,特别突出了复杂网络与人类动力学和维基百科研究可能的深入结合。

关键词 复杂网络; 演化规律; 统计性质; 维基百科

中图分类号 N941

文献标识码 A

doi:10.3969/j.issn.1001-0548.2010.03.001

Research Progress on Wikipedia

ZHAO Fei^{1,2}, ZHOU Tao^{1,3,4}, ZHANG Liang^{1,2}, MA Ming-hui^{1,5},
LIU Jin-hu^{1,6}, YU Fei^{1,5}, ZHA Yi-long^{1,7}, and LI Rui-qi^{1,7}

1. Web Sciences Center, University of Electronic Science and Technology of China Chengdu 610064;
2. School of Economy and Management, University of Electronic Science and Technology of China Chengdu 610054;
3. Department of Modern Physics, University of Science and Technology of China Hefei 230026;
4. Department of Physics, University of Fribourg Fribourg Switzerland 1700;
5. School of Software Engineering, University of Electronic Science and Technology of China Chengdu 610054;
6. School of Applied Mathematics, University of Electronic Science and Technology of China Chengdu 610054;
7. Experimental Class of International Software Professionals, University of Electronic Science and Technology of China Chengdu 610054)

Abstract The rapid development of web technology has promoted the emergence and organization of the collaborative Wiki systems. This paper introduces the Wikipedia's history, macro-level statistical properties, evolution regularities, and so on. Especially the application of the motivation and methods of complex network study in analyzing the Wikipedia is emphasized. Wikipedia's significance and impacts on society, economy, culture and education are also discussed. Finally, some open questions are outlined for future research; especially the connection between Wikipedia and the new development in complexity sciences, such as the studies of complex network and human dynamics.

Key words complex networks; evolution mechanism; statistical features; Wikipedia

维基模式及与其相关的Web2.0技术可能带来信息革命的新浪潮,其影响力不仅体现在各类与互联网相关的产品和服务中,还进一步渗透到实体经济中^[1-2]。维基百科是其中最成功的典范之一,在社会、经济、文化和教育方面都具有重大的影响和重要的意义。维基百科已经成为计算机科学、物理学、管

理科学、情报与图书馆学、教育学等多学科共同感兴趣的研究对象。

尽管国内已经有学者成功地将复杂性科学的一些思想应用于刻画和解释与Web2.0相关的技术和系统^[3],但是这些尝试主要还是停留在概念层面,针对如维基百科这类具体系统的定量研究还少见报

收稿日期: 2010-03-10

基金项目: 国家973计划(2006CB705500); 国家863计划(2007AA01Z440); 自然科学基金重大研究计划(90924011); 国家自然科学基金重点项目(10635040); 国家自然科学基金面上项目(60973069, 60973120, 60903073)。

作者简介: 赵飞(1985-),男,硕士生,主要从事复杂网络方面的研究。

道。相比而言,国际上已经有一些学者利用复杂网络的方法和手段对维基百科进行了深入细致的分析,得到了定性讨论难以企及的深刻见解。

本文综述了维基百科的历史发展、技术特点及演化规律研究,特别强调复杂网络的思想和方法在维基百科研究中的应用。统计物理学和复杂性科学的量化分析结果是本文不同于大部分国内相关论文的主要特征。文章最后讨论了维基百科在社会、经济、文化及教育等方面的意义和价值,提出了相关研究可能的发展方向,特别突出了复杂网络与人类动力学和维基百科研究深入结合的可能性。

1 维基百科的历史与特点

Wiki全称为WikiWiki,来源于夏威夷语的“*Wee kee Wee kee*”一词,原为“快点快点”的意思,中文多译为“维客”或者“维基”。Wiki是一种采用“多人协作”模式的超文本写作系统。2001年1月,由Jimmy Wales和Larry Sanger创建的维基百科英文版开始运行,http://www.wikipedia.org是目前世界上最大的Wiki网站,首页Logo见图1。维基百科在第一年就取得了飞跃式的成功,赢得了一批忠诚的追随者,制作了2万多个词条,提供了十几种语言的翻译版本^[4-5],之后逐渐发展成包含多种语言的网络百科全书家族。截止2010年3月,英文维基百科已有320万个条目,成为条目数最多的语言版本,而全球271种语言的总条目数达到1 500万个,总注册用户超过2 200万人,总编辑次数突破8.6亿次。维基百科各语言版本发展并不均匀,其中条目总数前15名语言版本的维基百科共占总条目数的71%^[6]。

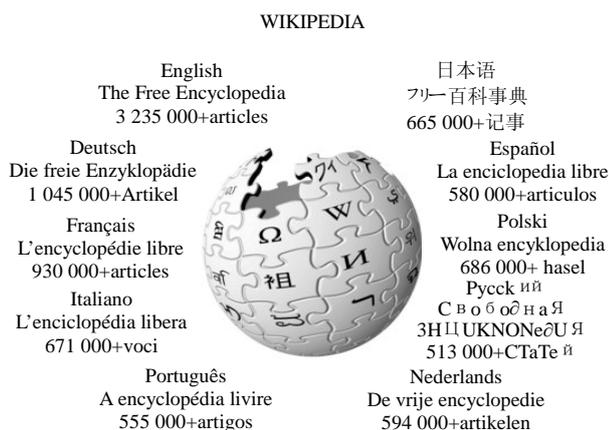


图1 维基百科 Logo^[5]

维基百科推出了自己独特的管理和编辑技术,主要包括:

(1) 保留记录:即使参与者将整个页面删掉,管

理者也可以方便地从记录中恢复以前的页面版本。

(2) 页面锁定:一些主要页面的编辑完善程度如果达到维基百科的标准,可以用锁定技术将内容锁定,其他用户就不能再随意编辑了。

(3) 版本对比:维基百科的每个页面都有更新记录,任意两个版本之间都可以进行对比,系统会自动找出它们的差别。

(4) 更新描述:在更新一个页面的时候用户可以在描述栏中说明更新内容的依据并与管理员进行对话等,有助于管理员及时知道页面改动的情况。

(5) IP禁止:尽管维基百科倡导“人之初,性本善”,人人都可参与,但破坏者、恶作剧者总是存在的。维基百科有记录和封存IP的功能,可记录破坏者的IP并禁止访问,使其不能再“胡作非为”。

(6) 沙箱(Sand Box)测试:维基百科建有一个沙箱页面(与普通页面一样),该页面让初次参与的人做编辑测试,从而熟悉维基百科网站的编辑环境,避免误操作。

作为web2.0技术与维基模式的成功典范,维基百科独特的闪光点^[7-8]包括:

(1) 开放共享:大部分页面都可以由任意用户使用浏览器进行浏览、修改、创建主题及条目等,且任何用户都可以免费下载、引用、收藏及分享它的内容。

(2) 互动协作:来自世界各地的用户都可以基于该平台针对某一主题内容展开交流研讨,通过不断地编写和修订,最终完善相应的词条与内容。

(3) 平等中立:任何用户在百科网站上享有同等的自由度,没有特权,并且被要求本着以客观事实为依据的原则进行信息发布、修改,尽量避免主观性的言语。维基百科不会倾向于某个观点,而是保持中立,相信“真理越辩越明”,让所有参与者去判断、思考。

(4) 简单快捷:维基百科操作简单,参与者可以非常方便地创建、编辑页面。与传统的百科全书相比,维基百科会在第一时间补充社会科技文化的新概念、新动态,保证知识的时效性。

(5) 信息全面、准确:维基百科是众人协作编写的百科全书,其内容涵盖地理、历史、社会、科学及教育等各个领域。而且大部分知识点都是经过不同的用户一次次编辑形成的,保证了内容的准确性。专家们在比较维基百科和大英百科全书的科学报道时,做了42项测试,大英百科全书出现3处不准确的报道,而维基百科也仅有4处^[9]。

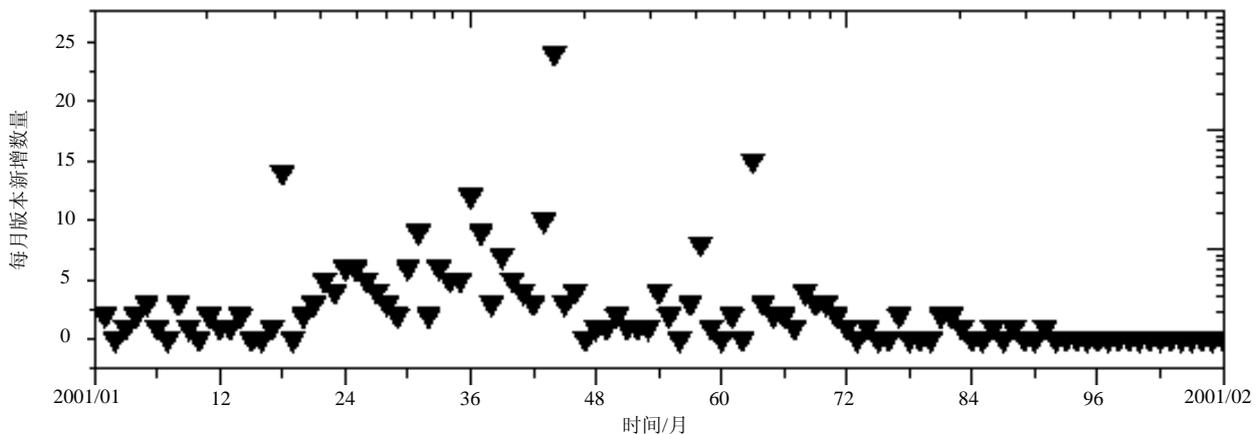
2 维基百科宏观统计性质和演化规律

研究维基百科的演化轨迹, 可以从纵向和横向两个方面考虑。前者考虑维基百科随时间的演化进程, 后者考虑不同版本维基百科之间的对比。本节先从不同语言版本发布数量的演化情况开始, 然后再简介一些维基百科的宏观统计量, 并挑选一些有代表性的统计量, 展示维基百科的宏观统计特性和演化规律。

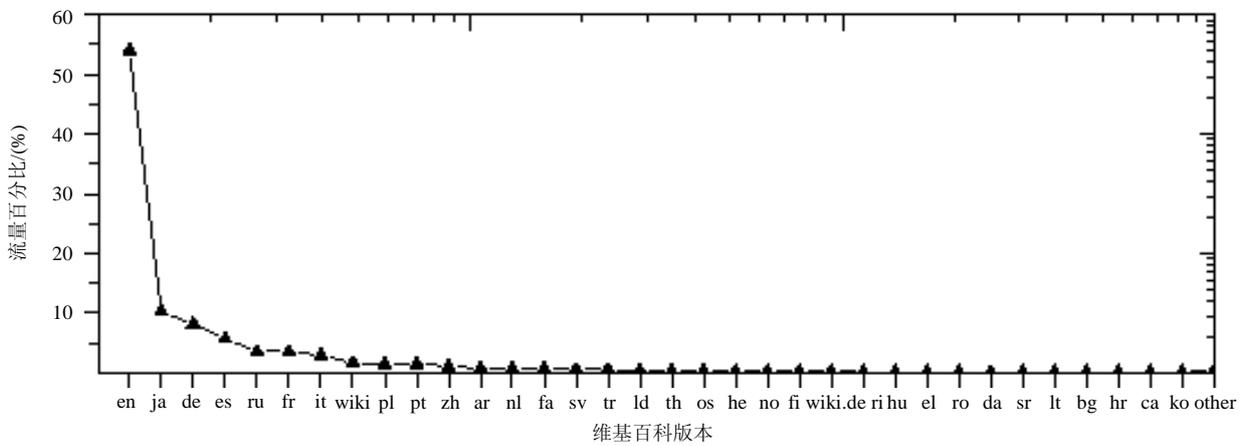
2.1 维基百科家族的发展情况

维基百科已经拥有 200 多个不同语言的版本, 是涵盖多种文化的多语言网络百科全书其演化情况如 2 所示。图 2a 回顾了 2001 年 1 月~2010 年 2 月

期间不同语言版本维基百科发布数量的变化情况, 从图中可以看出 2001 年~2007 年是各语言版本出现的高峰期, 其中 2004 年 9 月达到最多, 一共有 24 个新版本出现, 而 2007 年之后没有新的版本出现。根据 Alexa Internet 网路流量测量服务统计^[10], 目前在整个维基百科网站总流量中, 英文维基百科的流量总数已经占到 54%, 其他 46% 则由日文、德文等其他语言维基百科版本分享, 显示出英文维基百科在整个维基百科家族中的核心地位。与 2007 年在整个流量总数中占据的 45% 相比, 英文维基百科的比重仍然在增长。不仅如此, 流量百分比排名还具有显著的长尾效应, 如图 2b 所示。



a. 每月新增维基百科语言版本数量



b. 维基百科各版本网络流量百分比(2010 年 3 月数据)

图 2 维基百科家族演化情况

2.2 维基百科的宏观统计性质

为了帮助用户了解维基百科家族的演化状况, 维基媒体基金会提供了一个统计站点^[11], 负责独立发布其监控记录的各项维基百科运行数据。该站点主要发布 23 项维基百科的宏观统计量, 这些统计量的类别和含义见表 1。23 项统计量能综合反映维基百科演化的各个方面。本文着重考察正式文章数、贡献者、内部连接数以及每月编辑次数 4 项重要统

计量。将排名前 11 的语言版本的维基百科(按内部链接数量排)2001 年 1 月~2010 年 2 月的 9 年间该 4 项统计量每个月的增长率整合到同一张斑图上, 如图 3 所示。从横向时间维度看, 所有统计量在各版本的早期增幅较大, 基本上都是超过 20% 的增长; 但随着时间的推移, 增速减慢, 其中“内部链接”表现得最明显。从纵向版本对比看, 各个统计量演

化规律相似：(1) 除了俄语(RU)稍微异常外，其他版本各统计量的高增长周期长度一致；(2) 增长具有一定的同步性，统计量“每月编辑次数”表现最为明显，各版本要么同时增长快，要么同时增长慢。

表 1 维基百科的 23 项宏观统计量^[11]

类别	统计量	统计量含义
文章相关统计量	正式文章数	至少包含一个内部链接的文章的数量
	交互文章数	至少包含一个内部链接且可读单词数超过 200 的文章数量 数量
	文章编辑次数 *	每篇文章被编辑的总次数
	文章长度 *	每篇文章的可读单词数量
	每日新增文章数	--
	每日页面请求数 *	每篇文章每日的页面请求数
	大于 0.5 Kb 文章数	可读单词数超过 0.5 Kb 的文章数
用户相关统计量	大于 2 Kb 文章数	可读单词数超过 2 Kb 的文章数
	新用户	当月开始参与编辑且次数达到 5 次以上的用户
	贡献者	编辑次数累积超过 10 次以上的用户
	活跃用户	当月编辑次数达到 5 次以上
链接相关统计量	非常活跃用户	当月编辑次数达到 100 次以上
	用户编辑次数 *	用户总编辑次数
	内部链接数	指向维基百科内部的链接数量
	跨版本链接数	指向其它版本维基百科的链接数量
编辑相关统计量	外部链接数	指向非维基百科网站的链接数量
	重定向链接数	负责词条重定向功能的链接数量
	每月编辑次数	编辑次数包括对所有页面的编辑
数据量相关统计量	数据库大小	所有条目的总量
	总二进制数量	总的图片等多媒体数量
	总文字数量	总的可读单词数量
	每日访问量	每天点击维基百科的次数

表中，标记“*”的统计量是可排名统计量

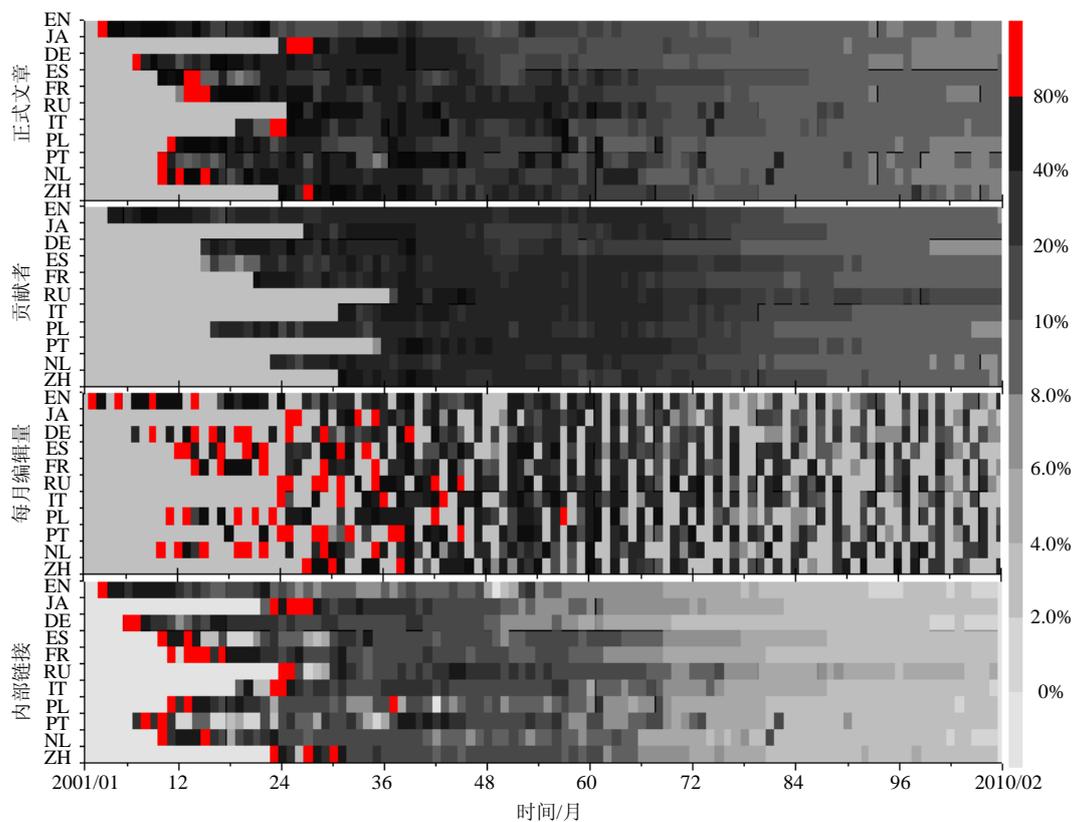


图 3 正式文章、贡献者、每月编辑总次数及内部链接 4 个统计量的每月增长率

其他研究人员关注了更多的指标。文献[12-13]关注了文章长度：前者发现其服从对数正态分布；后者发现其均值变化呈上升趋势，平均每页每年新增量在 400 个单词左右。文献[14-17]分别关注了用户相关指标下的各类用户的贡献变化。研究发现维基百科的贡献来源由活跃人群逐渐转移到低参与度人群。

在对表 1 中的 23 个可以直接获取的统计量仔细研究后，维基媒体基金会发现它们在衡量文章质量时都具有一定的局限性，于是引入了统计量“文章编辑深度”综合衡量维基百科文章质量^[18]，其定义为：

$$\text{Depth} = \frac{\text{Edits}}{\text{Articles}} \cdot \frac{\text{NonArticles}}{\text{Articles}} \cdot (1 - \frac{\text{Articles}}{\text{Total}})$$

式中 Edits、Articles、NonArticles 和 Total 分别

指各个版本的编辑总次数、文章总数量、非文章元素总数量和所有元素总数量。“文章编辑深度”是每篇文章平均编辑次数、每篇文章平均支撑页面(非文章页面)数量以及调整因子三者的综合。表 2 列出了文章数量在 10 000 以上的维基百科按照编辑深度排名的前 20 名以及相应的文章数量排名。从表中数据可以看出，英语(EN)版无论在文章数量还是质量上都排名第一。而其他版本的维基百科要么在数量上占优势而质量较差，要么在质量上排名靠前而数量不足。

表格中 20 个版本缩写分别对应于英语、马拉雅拉姆语、阿拉伯语、土耳其语、西班牙语、法语、波斯语、泰语、孟加拉语、俄语、德语、葡萄牙语、匈牙利语、意大利语、马其顿语、汉语、罗马尼亚语、拉托维亚语、波斯尼亚语和中国粤语。排名数据在该网页上会实时更新。

表 2 按编辑深度排名的前 20 名维基百科以及相应的文章数量排名(2010 年 4 月数据)^[19]

维基代码	EN	ML	AR	TR	ES	FR	FA	TH	BN	RU	DE	PT	HU	IT	MK	ZH	RO	IV	BS	ZH-YUE
文章编辑深度	506	376	186	172	163	136	108	101	95	88	87	79	77	73	68	65	64	60	55	55
文章数量排名	1	88	24	19	8	3	33	40	69	10	2	9	18	5	49	12	20	62	57	82

2.3 维基百科演化规律

是否存在统一的机制可解释维基百科所表现出的若干宏观统计规律呢？本小节分别就维基百科增长机制、宏观统计量的排名规律以及宏观统计量与网络规模间的相关关系，尝试初步回答该问题。

维基百科增长机制主要关注其文章的数量随时间的变化。如果能有一个简单的数值模型模拟其过去和未来的增长情况，无疑会对进一步认识维基百科提供帮助，甚至可以利用其进行预测。本文总结 4 个具有代表性的数值模型^[20]。

(1) 指数增长模型。该模型假设文章的增长具有固定的增长率，也就是呈现指数增长模式，因此采用指数函数模拟其增长过程，拟合的指数函数为：

$$N(t) = N(t_0)e^{t/\tau} \quad \tau = 500 \text{ days}$$

式中 $N(t)$ 为新文章增长数量； t 为时间； e 为自然对数。该模型对 2002 年 10 月~2006 年 10 月的增长情况拟合得很好，但是 2006 年 10 月之后的数据就不符合规律了。

(2) Logistic 模型。该模型假设更多的内容会造成更多的流量，反过来又会产生更多的新内容。而内容的增加又会导致潜在可写的内容变少，从而减少新内容的增加，两个因素共同作用影响文章数量的增长。基于该假设，Logistic 模型拟合的曲线为：

$$N(t) = 4 + \frac{33403}{1 + \left(\frac{t}{82}\right)^4}$$

与指数模型相比，该模型较好地模拟了新文章增长变缓的趋势，但是模型中新文章增量最终会趋于 0，以及增长趋势变缓过快，均与实际情况有很大不同。

(3) 扩展的增长模型。基于实际数据的分析，扩展的增长模型认为新文章数量既不会像指数模型那样一直持续增长，也不会像 Logistic 模型那样快速地衰减为 0，而是在达到最大值后，呈现出一种缓慢下降趋势，并且这个趋势会持续较长时间。

(4) 两阶段增长模型。模型将新文章增长数量 $N(t)$ 分为两个不同的指数函数，一个增长趋势，另一个则为下降趋势，拐点处出现在 2006 年 1 月(针对英文维基百科)。同时该模型还发现新文章增加数会有一个随季节变化的趋势。

宏观统计量中有些指标可以针对单个个体进行统计，然后根据统计结果进行排名(表 1 中标记“*”号的统计量)。文献[15]发现用户编辑次数服从 Zipf 定律。本文的进一步研究发现，几乎所有的排名指标都服从 Zipf 定律，即指标数量 N 与排名 r 具有 $N \propto r^{-b}$ 的比例关系。文章访问次数、文章最新编辑

次数、文章版本数量、用户编辑次数四个可排名统计指标的排名情况如图4所示。在对数-对数坐标上指数 b 在0.42~0.71之间。

文献[21]研究了不同语言版本的维基百科内部链接次数和每月编辑次数与文章总数 N 的关系,如图5所示。研究发现内部链接的数量与文章总数呈超线

性关系,如图5a所示。每月编辑次数则与文章总数呈亚近似线性关系,如图5b所示。该发现引出了一个重要的假说,即不同语言版本的维基百科实际上可以被看作是一个统一的演化机制在不同时间的表现。当然,该假说是否正确,以及在多大程度上正确,还需要进一步的研究。

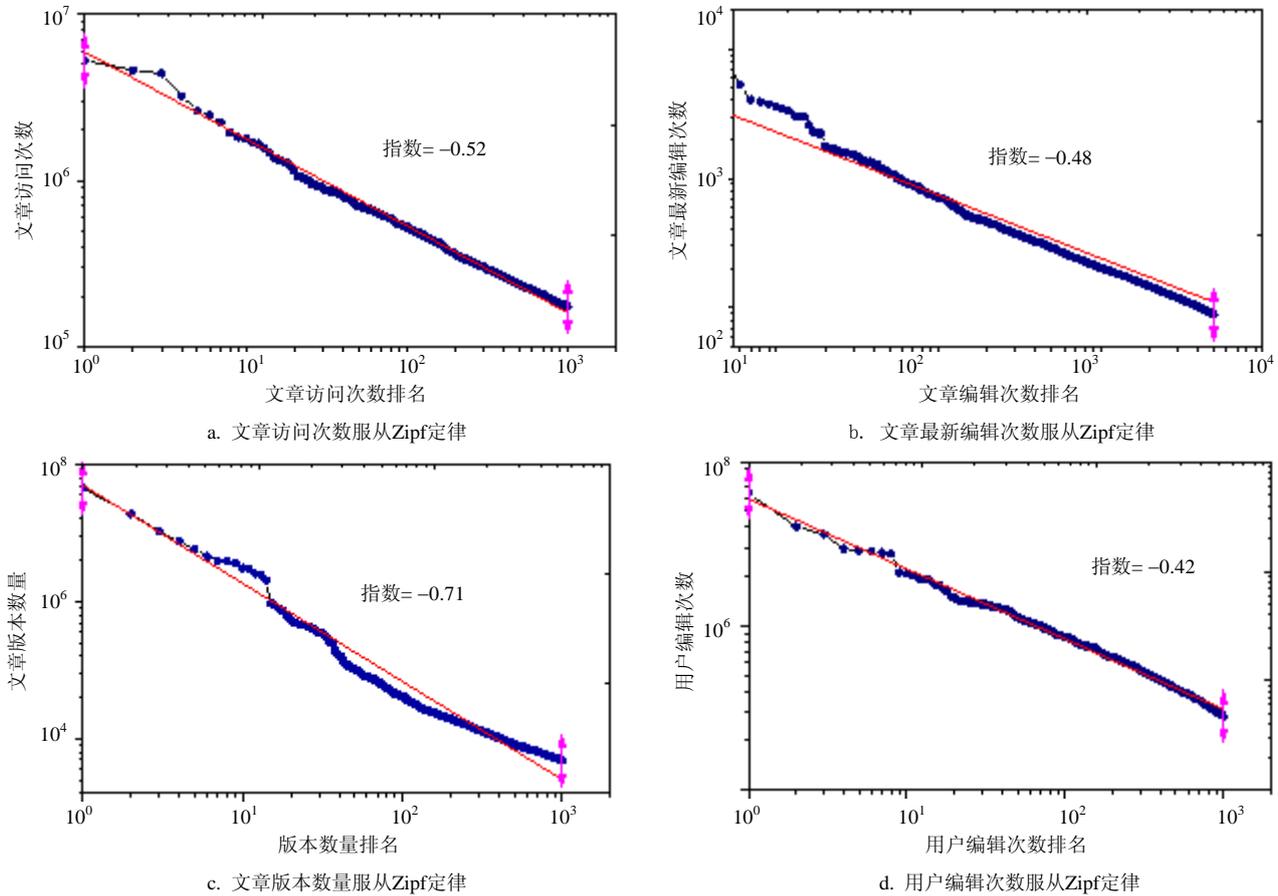


图4 可排名指标均服从 zipf 定律

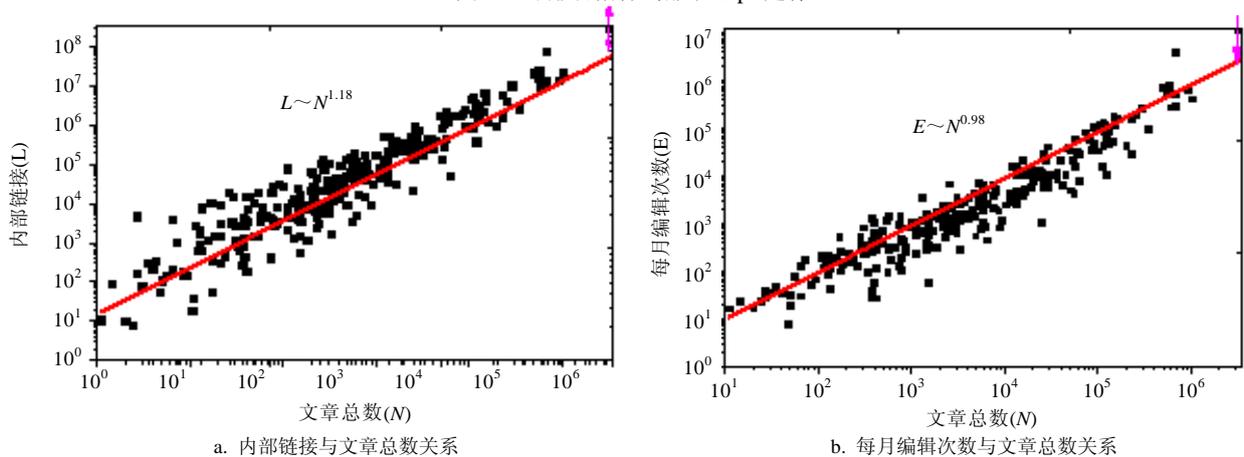


图5 不同语言版本维基百科个别指标与文章总数之间的关系

3 维基百科的复杂网络分析

近年来,复杂网络成为多学科关注的交叉研究

热点^[22-26]。随着web2.0网站的普及和海量数据分析方法的便利,复杂网络的分析方法被广泛应用于社交网络的研究。许多文献从维基百科的海量数据抽

象出各类网络结构并进行分析。从研究的角度看,可以分为3个层次:

- (1) 宏观层次上,研究网络的宏观物理统计量,评价网络的整体性质;
- (2) 中观层次上,研究网络的局部社区结构,揭示某些隐藏的局部信息;
- (3) 微观层次上,研究单个实体的演化情况,推断整个维基百科的发展。

从数据源来看,研究的网络包括文章-文章网络,语义类别-文章网络和用户-文章网络。下面,将从数据源的角度回顾有代表性的工作。

3.1 文章-文章网络

将维基百科中的每篇文章当作一个节点,文章之间的超链接当作有向边,就得到描述维基百科的文章-文章有向网络。该网络排除了维基百科中的分类、图片、多媒体、用户、帮助、讨论页面及已损坏的链接等,集中反映了文章与文章之间的关联性。由于文章是维基百科最主要的组成部分,文章之间的超链接的生成与文章内容密切相关,故研究该网络的特性,能够反映维基百科所涵盖的知识点之间的各种关联特性。文献[21]收集了2005年1月7日之前超链接数量排在前11的维基百科作为数据集,包括英语(EN)、德语(DE)、日语(JA)、法语(FR)、瑞典语(SV)、波兰语(PL)、荷兰语(NL)、西班牙语(ES)、意大利语(IT)、葡萄牙语(PT)和中文(ZH),横向比较了各个语言版本维基百科之间的各种网络结构特征(包括增长情况、整体拓扑结构、度分布、互惠性、聚类特性、度度相关性、平均最短路径以及三元子结构等),总结出不同语言版本维基百科之间的共性及个性。文献[27]在研究维基百科层次结构的同时,不仅重复了对23种不同语言百科的上述部分指标的测量,还进一步分析了入度出度比的分布情况、介数与点度数和网络大小的关系等。文献[28]考察了度-度相关性,并提出了比“优先链接”^[29]更加复杂的演化机制解释维基百科文章-文章网络的形成。

3.1.1 整体拓扑结构

网络整体拓扑结构分析的意义在于可以直观地了解网络是由哪几大部分组成,以及各部分之间的关系。文献[21]和[28]分别进行的分析发现,维基百科的文章-文章网络表现出与万维网^[30]类似的蝶形结构,如图6所示。在该结构中,最重要的部分强连接组件(strongly connected component, SCC)由那些两两之间都存在有向路径的最大的页面集合组成。IN、OUT区域分别表示只指向SCC和只由SCC指向

的页面。Tendrils部分既包括了那些由IN可达,但与SCC和OUT无关的页面;还包括了那些可达OUT,但与SCC和IN无关的页面。Tubes部分直接连接IN和OUT两个区域,而不经SCC。剩下的少数页面就是那些独立的部分(disconnected components, DISC)。研究发现SCC占的比例高达67%~89%(文献[21]为85%~95%,统计数据不一致的原因是两者所用的数据集不一样),表明不同版本的维基百科表现出很强的“内部互连特性”^[28]。

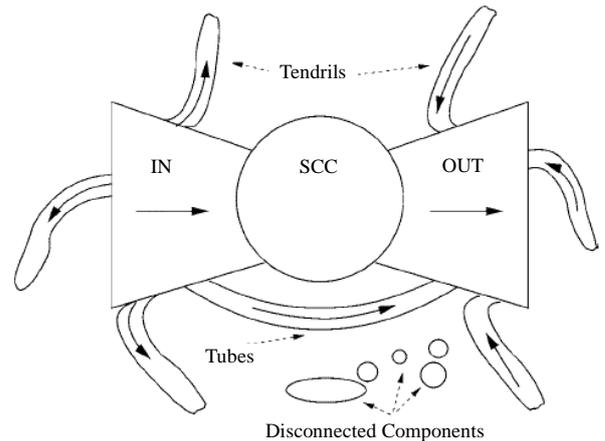


图6 维基百科的拓扑形状^[30]

3.1.2 度分布

调查度分布是复杂网络研究的基本内容^[31-32]。许多实际网络的度分布都具有无标度特性,即概率密度随度呈现幂律衰减形式 $p(k) \sim k^{-\tau}$ 。对任意页面(文章) i ,把从其他页面指向 i 的链接数量记为 $k_{i,in}$,称为 i 的入度;再把 i 指向其他页面的链接数量记为 $k_{i,out}$,称为 i 的出度。文献[21]和[28]分别统计了维基百科的入度和出度的分布情况,发现它们也服从幂律分布。在研究结果中,入度的统计结果一致,幂指数均在2.0~2.2之间;但出度分布差别较大,文献[21]所报道的幂指数约为2.6,而文献[28]的结果约为2.1。出现偏差主要是因为两者在计算出度分布时采用了不同的数据集和统计标准。

文献[28]通过实证验证了维基百科中文章链接的增长机制是否满足“优先链接”假设,结论是对于具有较小度的点($k \leq 100$),假设成立;而对于度大的点($k > 100$),由于“有限尺度效应”^[33],表现出了与“优先链接”不一致的行为。研究进一步指出,大量链接是在已经存在的文章中添加的。为了更加合理地描述这样的现象,在“优先链接”的基础上,提出了新的增长机制。该机制能够比“优先链接”更好地解释维基百科的文章增长情况。

文献[27]进一步研究了 $k_{in} > k_{out}$ 的点的 k_{in} / k_{out} 分布情况, 以及 $k_{out} < k_{in}$ 的点的 k_{out} / k_{in} 分布情况。发现两者均服从 $p(k) \sim p^{-1-\alpha}$ 的幂律分布。 $\alpha_{in/out}$ 在

1.1~1.7之间, $\alpha_{out/in}$ 在0.9~1.5之间。并且该二比值的分布比出度和入度分布更好地符合幂律分布形式, 如图7所示。

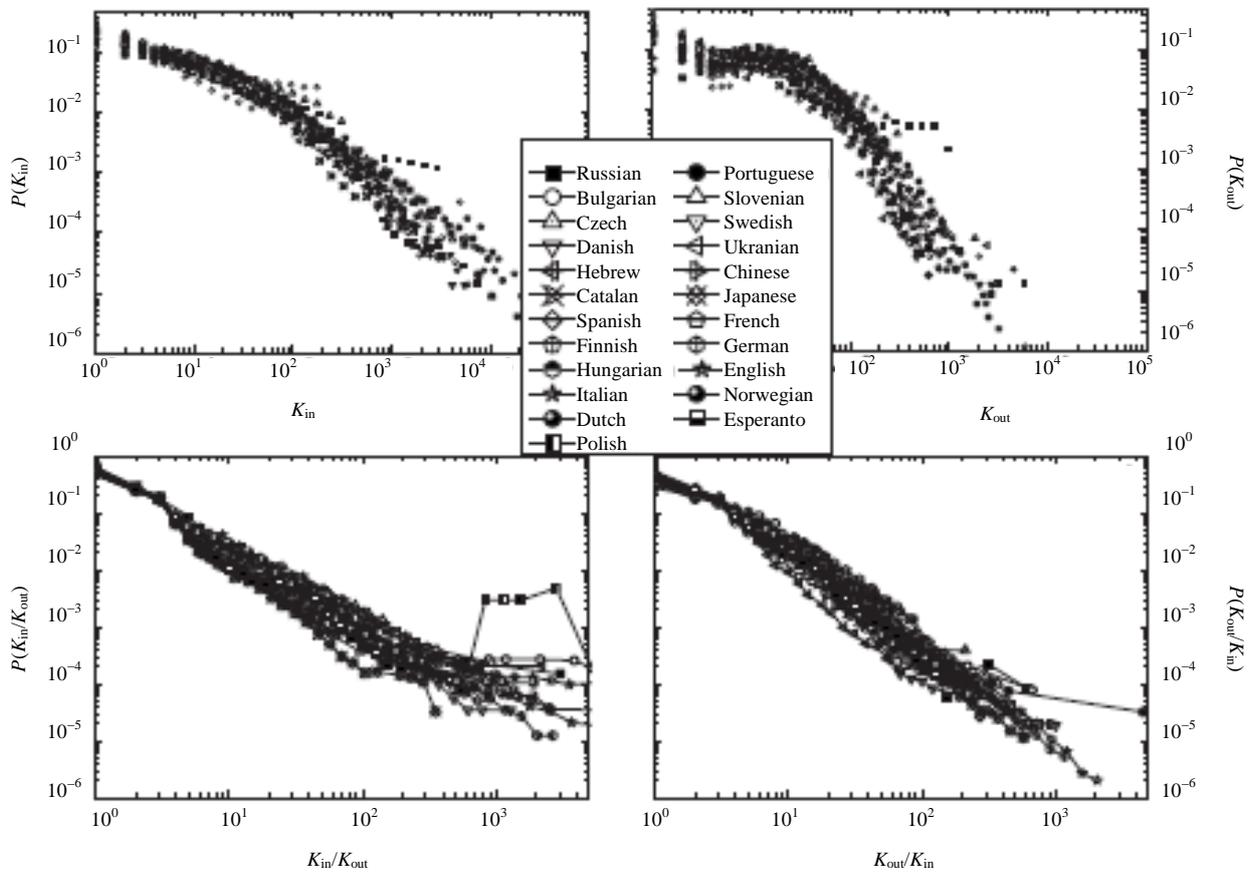


图7 不同版本维基百科的入度、出度及其比值的分布情况^[27]

3.1.3 互惠性

文献[34]引入无偏差的互惠性系数描述有向网络中的互惠连接性。所谓互惠连接是指在网络中若点 i 指向点 j , 则点 j 也指向点 i 。一个网络的互惠连接系数定义为:

$$\rho = \frac{L^{\leftrightarrow} - \bar{\alpha}}{L - \bar{\alpha}}$$

式中 L^{\leftrightarrow} 是互惠连接的数量; L 是总连接的数量; $\bar{\alpha} = L / (N(N-1))$ 是网络的连接密度。显然, ρ 越大, 则网络的相互连接越多。 $\rho > 0$ 和 $\rho < 0$ 分别表明网络的互惠连接数量与随机网络比起来是多还是少。文献 [21] 得到维基百科的互惠连接系数为 $\rho = 0.32 \pm 0.05$ 。该值大于随机网络 ($\rho = 0$), 少于万维网络 ($\rho = 0.52 \pm 0.0006$), 说明维基百科虽然存在着大量的互惠连接数量, 但与万维网网络相比仍略逊一筹。文献[27]发现不同版本的维基百科的所有连接中, 大约有23%~59%是互惠连接的。

3.1.4 聚类特性

聚类系数描述了节点的聚集特性, 最先是由文献[35]在描述网络的“小世界”特性时提出的。对具有度为 $k_i = k_{i,in} + k_{i,out}$ 的节点 i , 记其邻点集为 $N_i = \{v_j : e_{ij} \in E \wedge e_{ji} \in E\}$, 则 i 的聚类系数为:

$$C_i = \frac{|E_{jk}|}{k_i(k_i - 1)} : v_j \in N_i; e_{jk} \in E$$

而对于整个网络 G , 聚类系数定义为各个节点度的均值。文献[27]测量得到不同版本维基百科的聚类系数都在0.14~0.26之间, 还考察了节点聚类系数与度的关系, 发现聚类系数随着度的增大而增大, 在度大约为100时达到峰值, 之后又随着度的增大而减少。

3.1.5 同配性

文献[36]定义了有向图的同配性系数, 描述网络中度大的节点是更倾向于连接度大的节点, 还是更倾向于连接度小的节点, 该系数定义为:

$$r = \frac{\sum_i j_i k_i - M^{-1} \sum_i j_i \sum_i k_i}{\sqrt{\left[\sum_i j_i^2 - M^{-1} \left(\sum_i j_i \right)^2 \right] \left[\sum_i k_i^2 - M^{-1} \left(\sum_i k_i \right)^2 \right]}}$$

式中 j_i 和 k_i 分别是第 i 条有向边的两个端点的度数, M 为网络中边的条数。若同配型系数 $r > 0$, 则网络是同配的(assortative); 若 $r < 0$, 则网络是异配的(disassortative)。相关研究表明, 技术网络一般是异配的, 而社会网络一般是同配的。文献[21]研究发现除了波兰语表现出很高的同配性($\gamma = 0.38$), 其他各语言版本维基百科的文章-文章网络的同配性系数均为 $\gamma = -0.1 \pm 0.05$, 表现出弱的异配性, 表明自由编辑的维基百科内容同配性同技术网络相仿, 而波兰语特殊的编辑规则(文章必须采用包含大量指向日历页面的模板)会改变该特性。

3.1.6 平均最短路径

网络中两个节点 i 和 j 的距离定义为从 i 到 j 的最短路径上的边数, 而整个网络的平均最短路径是所有节点距离的均值。文献[27]统计了距离的分布情况, 发现各个语言版本的文章网络中, 距离为5的比例最多。文献[21]计算了各语言版本所对应网络的平均最短路径, 结果约为4.5, 且对网络规模不敏感, 表明文章-文章网络也存在“小世界”特性^[35]。

3.1.7 介数

一个节点的介数由网络中通过该节点的最短路径的数目决定^[37]。在有向网络 $G = (V, E)$ 中, 节点 v 的介数定义为:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

式中 $\sigma_{s,t}$ 是从节点 s 到节点 t 的最短路径的数量; 而 $\sigma_{s,t}(v)$ 是从节点 s 到节点 t 的最短路径中, 经过节点 v 的数量。文献[27]统计了各语言版本的介数与度的关系, 发现它们近似服从幂律关系, 其幂指数约为1.3。

3.1.8 三元模块分析

上述的统计指标主要刻画网络的全局结构特征。文献[38]提出用重要性剖面(significance profile, SP)方法分析网络的局部结构。重要性剖面指由少数几个点组成的各类子图结构在网络中出现频率的示意图, 如图8所示。横坐标显示了由3个点组成的13种重要性剖面, 称为三元组重要性剖面(triad significance profile, TSP)。计算SP的基本思想是将考察的实际网络与相对应的随机网络作对比, 得到每个待考察的子图结构的统计重要性:

$$Z_i = \frac{N_{real_i} - \langle N_{rand_i} \rangle}{std(N_{rand_i})}$$

$$SP_i = \frac{Z_i}{\left(\sum_i Z_i^2 \right)^{1/2}}$$

式中 N_{real_i} 和 N_{rand_i} 分别表示子结构 i 在实际网络和相应的随机网络中出现的次数, $\langle N_{rand_i} \rangle$ 和 $std(N_{rand_i})$ 分别表示 N_{rand_i} 的均值和标准差。文献[21]研究了维基百科文章-文章网络的三元组重要性剖面, 并将其与WWW网络比较, 惊奇地发现不同语言版本的网络的TSP分布与WWW网络的TSP分布极为相似。

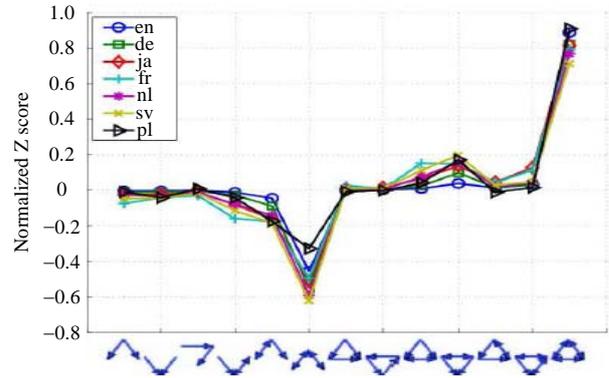


图8 各个语言版本维基百科的TSP剖面^[21]

3.2 语义类别-文章网络

维基百科为了便于管理, 同大多数实际的百科全书一样, 采用语义类别组织文章。语义类别指那些用户在编辑文章时产生的语义标签(如“Physics”、“Medical”等)。每一篇维基百科都会根据它的主题被划归到相应的类别下面。但与一般的百科全书不一样的是, 维基百科的文章和类别的关系不是简单的树型关系。因为在维基百科中, 一个类别下面可以有多篇文章, 一篇文章也可以隶属于不同的类别。严格地说, 维基百科中类别和文章的关系是有向无循环图的关系^[39]。这样的层次结构是人们在自由编辑文章时产生的, 修改和删除都要经过整个社区的同意。人们对语义类别的分析是从3个角度展开的:

(1) 研究语义类别的宏观统计特性, 试图通过这些全局的特性揭示知识组织的一般模式;

(2) 将语义分类对文章的分类结果作为判断文章-文章网络社区划分算法准确性的标准, 通过对比两种分类结果的重叠性对相应的分类算法打分, 寻找最佳的社区划分算法;

(3) 将语义类别看作是知识的层次结构, 或者说知识树, 然后设计一系列的算法分离隐藏在知识树中的潜在的层次结构, 并且对知识树中不同层次类别所包含的内容进行自动分类。

3.2.1 语义类别的宏观统计特性

文献[40]在研究语义分类所包含的子孙数量的分布情况时发现其服从幂律分布,且不同维基百科的幂指数都很接近(约为2.2),如图9所示。语义分类这样统一的分布情况不是偶然的,可以通过Yule过程^[41]和超临界树^[42-43]的性质加以解释。Yule过程解释导致幂律分布的基本机制;超临界树则通过随机分支过程解释分类的树形层次结构的演化,不同分支数量会影响幂律分布的指数。分析结果证明当分支数等于1时,幂指数为1.5;当分支数大于1时,幂指数为2。因此,维基百科的幂指数接近于2可以看作是超临界树的分支结构的结果^[44]。

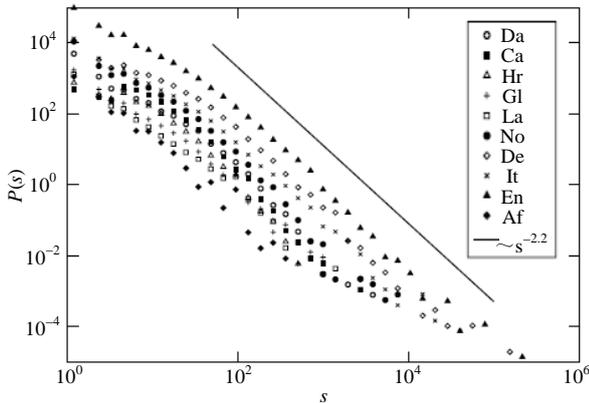


图9 不同版本维基百科的语义类别大小服从幂律分布,幂指数约为2.2^[40]

3.2.2 语义类别和社区结构分类

在对维基百科文章进行分类的问题上,除了通过语义类别对文章进行自然的分类外,还可以通过一系列算法来文章-文章网络进行社区结构划分,实现分类。语义类别形成的自然分类常常被看作分类算法好坏的判断标准。文献[40]用该方法检验马尔科夫聚类算法^[45](markov cluster algorithm, MCL)的分类情况,并采用“调整的随机指数(adjusted random index)”衡量两种分类之间的重叠性,但结果却显示“调整的随机指数”值似乎与文章-文章网络的大小没有关联性,说明文章的语义分类和通过MCL分类的结果没有必然联系,意味着后者实际上给出了文章-文章网络的新的完全不同的分类结果。不过另外一个有趣的结果是:通过MCL算法得到的分类结果的分布也服从幂指数约为2.2的幂律分布,暗示幂律分布的特性实际上是分类系统的一般特性,不能作为评估分类算法适应度的标准。

3.2.3 维基百科知识树

语义类别通常被称作维基百科知识树^[27]。对知识树研究的一个重点是在其中挖掘出隐藏信息结

构,因为基于知识树的现代浏览和搜索方法往往掩盖了信息结构。文献[27]提出用3个算法分离隐藏在知识树中的潜在的层次结构,并且对知识树中不同层次类别包含的内容进行自动分类。3个算法的基本思想是通过计算知识树中每个节点的3种不同指标,按照指标的大小判断两个节点的层次位置,指标越大节点的层次就越高,如果在预测的层次结构中两个节点的层次位置与知识树本身的层次位置一致,则认为预测正确。最后通过计算正确的百分比判断算法的好坏。

第一个算法计算层次中间性(hierarchical intermediacy),该指标对知识树中的每一个点*i*赋予一个值:

$$I(i) = \frac{C_B(i)}{\sqrt{[k_{in}(i)+1][k_{out}(i)+1]}}$$

式中 $C_B(i)$ 指点*i*的介数。实证显示该算法的平均正确率是82%。

第二个算法计算的是局部层次性(local hierarchy),该指标对每个点*i*的赋值只与其节点出入度相关定义为:

$$H(i) = \frac{k_{in}}{k_{in} + k_{out}} \sqrt{k_{in}} + \frac{k_{out}}{k_{in} + k_{out}} \sqrt{k_{out}}$$

该指标的平均正确率也是82%。

第三个算法综合了前两个算法,计算的指标是吸引子层次性(attraction basin hierarchy),定义为:

$$A(i) = \left(\sum_m \frac{N_{-m}(i)}{\langle N_{-m} \rangle} \alpha^{-m} \right) / \left(\sum_m \frac{N_m(i)}{\langle N_m \rangle} \alpha^{-m} \right)$$

式中 $N_{-m}(i)$ 和 $N_m(i)$ 是指向点*i*的点和*i*指向的点中最短距离为*m*的点的个数。当 $\alpha=2$ 时,该指标的平均正确率达到85%。

除了以上三种方法外,还可以采用PageRank^[46]和HITS^[47]算法的结果作为指标,平均正确率分别为83%和81%。

3.3 用户-文章网络

维基百科中的几乎所有内容都是由自由用户编辑产生,研究用户-文章关系有助于了解人们是如何进行协作编辑的。文献[12]统计了每篇文章独立编辑者的分布情况以及每个编辑者所编辑的独立文章数量分布情况,发现前者服从指数约为2.7的幂律分布,后者服从指数约为1.5的幂律分布。文献[48]最近将用户-文章关系当作二部图,用研究二部图派系的方法研究其模块结构,可以发现维基百科中的一些潜在的争论焦点和协同写作现象。例如,围绕着“全球变暖”词条,模块结构研究显示其包含了“太

阳变化”、“厄尔尼诺现象”、“二氧化碳”、“海平面上升”、“全球变暖争论”及“矿物燃料”等焦点词条。这些词条在维基百科知识树上,并不是

出于同一类别,显示了模块研究能发现潜在争论点的能力,如图10所示。

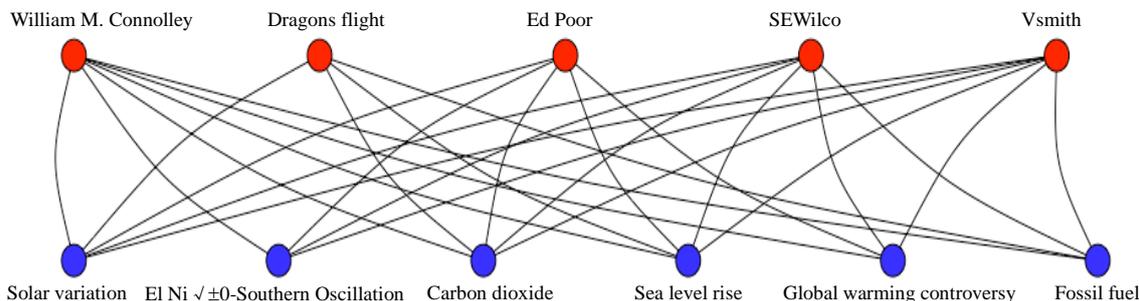


图10 模块结构研究能发现潜在的争论点和揭示编辑中的协调合作^[50]

再如,模块研究发现“元素周期表”中每一个词条都具有统一的风格,而事实上它们都是由少量的编辑者在WikiProjects^[49]的统一协调下修改完成的,显示了模块研究能发现潜在的协同写作现象的能力。此外,这些模块结构也反映出,词条的编辑往往由少数熟知该词条概念的编辑者合作完成,不同领域的词条编辑者往往不会重叠,这就解释了为什么维基百科的准确率堪比大英百科全书——每个词条都是由相应领域的“专家”合作完成的。

4 维基百科的影响及价值

维基百科发展至今,对社会、经济、教育科研等各方面都产生了一定的影响。因为维基百科所具有的开放、平等、自由等重要特征,彻底改变了传统百科全书科教的运作模式和传播方式,使它受到全球用户的肯定和喜爱。总结其对社会、经济、教育及其他方面的影响,并分析其在人类发展各方面表现出的价值,体现为:

首先,维基百科能促进非物质文化遗产的传播。许多非物质文化遗产要通过专人参与、集体协作实现完整传承,而维基百科自由平等的协作平台与传统百科相比,对非物质文化遗产的传承更有亲和力。维基百科为读者提供一个自由的客观的信息立场^[50],能成为社会舆论的重要组成力量,这是因为维基百科方便编写的特点,可以紧跟时代的发展,实现词条的快速添加和更新。

其次,维基百科的协同开发模式正在对经济生产模式产生着革命性影响^[2]。维基百科实现了知识的共享互动,使知识以更自由的方式快速地传播,有力地推动技术普及。维基百科作为一个客观的企业信息评价平台,对商业宣传与评价有客观引导价值。维基百科已经成为互联网上最流行的参考咨询

网站之一。

再次,与传统的百科全书相比,维基百科可以看成是一个在线的百科全书,用户在进行资料检索时,不必再寻找大量的书籍查找相关的内容,而是在维基百科上进行相关资料的查找,节省大量的精力与时间,同时,维基百科知识更新速度快,可以为人们提供最新最全面的信息,系统也会推荐一些相关知识点,使知识点融会贯通。因为维基百科涵盖的知识点丰富,高校的科研人员在进行学术研究时会选择在维基百科上进行相关资料的搜索,还能把维基百科当作科研成果展示平台。维基百科的开放共享等特性,也影响了新一代的教育方式。学生课后不单是可以进行独立的学习思考,还可以在网平台上进行群体性的学习交流。维基百科平等协作的理念让更多的人可以自由地参与整个资源知识系统的建设、管理,从而对建设一个完善的教学协作模式起到推动作用,提高教育质量^[51]。

除了以上三方面外,维基百科对互联网文化及其建设、知识以及情报的获取、新闻的传播与实证都有一定的影响。例如,美国现在正在积极打造“情报百科”,该体系被外界称为“维基百科”的翻版:美国情报特工进行注册后,可以把自己搜集的各类情报信息进行发布,但必须注明发布者,其他登录的情报机构人员有权限进行浏览和补充信息,可为情报部门提供可观而准确的实时信息,更加有利于高层做出正确决策^[52]。

维基百科的价值体现在它使用户成为信息制造的中心,拥有更多点对点传递信息的渠道:网络个人化、社会化、自组织等传播特点日益凸显。维基百科还打破了传统媒介信息准入的特权,使得消息更加趋向于真实、也更加精确,不再是官方“一言堂”。在知识爆炸的今天,如何过滤并且获取到可靠、

精确、有实效性的信息是问题之所在。虽然维基百科没有权威与中心,但在整个信息领域,它毫无疑问将是权威与中心。

5 结束语

维基模式已经极大地改变了社会经济生活,包括生产开发方式。维基百科作为最成功的代表之一,问世不足十年,却已经成为全球最著名的网络新贵之一。对于维基百科的研究,不仅仅是理解一种成功的模式,更重要的是通过对维基百科系统的分析,了解自组织系统从无到有、从小到大增长的规律,揭示大规模知识网络中各个知识点的关联特性,了解多人协作系统中不同人的角色和相互作用的模式。这些科学问题,是目前物理学、信息学、管理科学等多个学科共同关心的焦点。遗憾的是,大部分文献所关注的还是定性的讨论,国际上相关的定量研究很少,国内则几乎没有。本文除了详细介绍和进一步深入探讨维基百科的历史发展、技术特点以及其在社会经济教育文化等方面的意义,还尽所能整合集中了对于维基百科的量化研究,包括宏观统计特性、演化规律、复杂网络分析等等。文章特别强调了复杂网络的思想和方法在分析维基百科中的应用。

以下五个方面可能会成为近期维基百科研究的焦点和新兴增长点:

(1) 本文介绍了3类网络,事实上非常重要的一种网络是用户-用户网络,也就是把参与编辑过同一篇文章的用户连接起来形成的网络。维基百科的编辑历史提供了研究大规模协作所形成的社会网络的完美平台,这方面的研究必然是未来的亮点。

(2) 最近关于人类动力学的研究成为多学科交叉研究的共同焦点^[53-54],其中人与人相互作用的社会网络被认为对个人行为的时间统计特性有重要影响^[55]。遗憾的是,该假说没有得到有力的实证证明。维基百科每一次的新修改都会通知所有以前编辑过相应文章的用户,正好体现了社会相互作用,因此可以成为研究社会相互作用对个人行为时间特性的量化影响的实证平台。

(3) 基于用户-用户网络的分析结果和人类动力学的实证结果,有望辨别哪些用户是严谨的贡献者,哪些用户是恶意捣乱者,从而找到建立维基百科,甚至维基模式的信誉系统,打击可能的恶意甚至出于商业模式的蓄意欺骗行为。信誉系统的建立对于

互联网相关的服务和共享具有本质上的重要性^[56-57]。

(4) 分析维基百科语义关联有望在更细致准确的层面上揭示零散的知识点如何组成完整的知识系统的过程,还可以评估当前的知识点层次分类结构的有效性。

(5) 维基百科中很多内容还没有被分析过,例如用户讨论和图片信息等。这些内容可能很重要,例如在用户讨论区讨论过的用户一般而言都是资深的活跃用户,贡献很大。整合这些信息对于全面理解维基百科有重要作用。

本文研究工作得到电子科技大学“银杏黄”项目的资助,在此表示感谢。

参 考 文 献

- [1] 刘向晖. 互联网草根革命[M]. 北京: 清华大学出版社, 2007.
LIU Xiang-hui. The rebolution of internet grass-roots[M]. Beijing: Tsinghua University Press, 2007.
- [2] DON T, ANTHONY D W. Wikinomics: How mass collaboration changes everything[M]. New York: Portfolio Trade, 2006.
- [3] 张树人, 方美琪. Web2.0与信息系系统复杂性变革[M]. 北京: 科学出版社, 2008.
ZHANG Shu-ren, FANG Mei-qi. Web2.0 and complexity transformation of information system[M]. Beijing: Science Press, 2008.
- [4] 尹开国. 维基百科社群发展策略研究[J]. 图书情报知识, 2007, 117(3): 95-98.
YIN Kai-guo. Wikipedia community development strategy[J]. Library and Information Science, 2007, 117(3): 95-98.
- [5] 中文维基百科社区. 维基百科[DB/OL]. [2010-4-29]. <http://zh.wikipedia.org/wiki/维基百科>.
Wiki Community of China. Wikipedia[DB/OL]. [2010-4-29]. <http://zh.wikipedia.org/wiki/维基百科>.
- [6] 维基百科. 什么是维基百科[EB/OL]. [2010-4-29]. <http://wikipedia.jaylee.cn/>.
Wikipedia. What is Wikipedia[EB/OL]. [2010-4-29]. <http://wikipedia.jaylee.cn/>.
- [7] 张喜来. 对国内wiki发展的多元化思考[D]. 北京: 清华大学, 2006.
ZHANG Xi-lai. Ponder over the pluralism of domestic Wiki development[D]. Beijing: Tsinghua University, 2006.
- [8] 白崇远. 《维基百科》的特性及影响[J]. 辞海研究, 2009, 2(5): 67-72.
BAI Chong-yuan. The characteristics and influence of Wikipedia[J]. Word-Ocean Research, 2009, 2(5): 67-72.
- [9] GILES J. Internet encyclopaedias go head to head[J]. Nature,

- 2005, 438: 900-901.
- [10] Alexa Information Company. Wikipedia. Org [DB/OL]. [2010-4-29]. <http://www.alexa.com/siteinfo/wikipedia.org>.
- [11] Wikipedia Community. Wikipedia statistics site [DB/OL]. [2010-4-29]. <http://stats.wikipedia.org>.
- [12] VOSS J. Measuring Wikipedia[C]//Proceedings of the ISSI. Stockholm: ISSI Press, 2005: 1-12.
- [13] BURIOL L, CASTILLOC C, DONATO D, et al. Temporal evolution of the wikigraph[C]//Proceedings of the Web Intelligence Conference. Hong Kong: IEEE CS Press, 2006: 45-51.
- [14] ORTEGA F, BARAHONA J M G. Quantitative analysis of the Wikipedia community of users[C]//Proceedings of the 2007 international symposium on Wikis. New York: ACM, 2007: 75-86.
- [15] RODRIGO B, BARZAN M, JUNGHO C. On the evolution of Wikipedia[C]//ICWSM. Colorado: USA[S.n.], 2007: 1-8.
- [16] DENISE A, SEAN W S, WILLIAMSON T. Explaining quality in internet collective goods: Zealots and good samaritans, the case of Wikipedia[EB/OL]. [2010-4-29]. <http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf>.
- [17] KITTUR A, PENDLDTON B A, SUH B, et al. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie[C]//Proceedings of the 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007). San Jose, CA: ACM, 2007: 25-41.
- [18] Wikimedia Foundation. Wikipedia article editing depth[DB/OL]. [2010-4-29]. http://meta.wikimedia.org/wiki/Wikipedia_article_depth.
- [19] Wikimedia Foundation. Wikipedia article editing depth rating[DB/OL]. [2010-4-29]. http://meta.wikimedia.org/wiki/List_of_Wikipedias.
- [20] Wikimedia Foundation. Wikipedia growth mechanism [DB/OL]. [2010-4-29]. http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth.
- [21] ZLATIC V, BOZICEVIC M, STEFANCIC H, et al. Collaborative Web-based encyclopedias as complex networks[J]. Physical Review E, 2006, 74(5): 016115.
- [22] 吴金闪, 狄增如. 从统计物理学看复杂网络研究[J]. 物理学进展, 2004, 24(1): 18-46.
WU Jin-shan, DI Zeng-ru. Complex networks in statistical physics[J]. Progress in Physics, 2004, 24(1): 18-46.
- [23] 周 涛, 柏文洁, 汪秉宏, 等. 复杂网络研究概述[J]. 物理, 2005, 34(1): 31-36.
ZHOU Tao, BAI Wen-jie, WANG Bing-hong, et al. A brief review of complex networks[J]. Physics, 2005, 34(1): 31-36.
- [24] 方锦清. 迅速发展的复杂网络研究与面临的挑战[J]. 自然杂志, 2005, 27(5): 269-290.
FANG Jin-qing. Rapidly developing study of complex networks and its challenges[J]. Chinese Journal of Nature, 2005, 27(5): 269-290.
- [25] 汪小帆, 李 翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
WANG Xiao-fan, LI Xiang, CHEN Guan-rong. Complex network theory and its applications[M]. Beijing: Tsinghua University Press, 2006.
- [26] 何大韧, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2008.
HE Da-ren, LIU Zong-hua, WANG Bing-hong. Complex systems and complex network[M]. Beijing: Higher Education Press, 2008.
- [27] MUCHNIK L, ROYI I, SORIN S, et al. Self-emergence of knowledge trees: extraction of the Wikipedia hierarchies[J]. Physical Review E, 2007, 76(15): 016106.
- [28] CAPOCCI A, SERVEDIO V, COLAIORI F, et al. Preferential attachment in the wrowth of social networks[J]: The Internet Ncyclopedia Wikipedia. Physical Review E, 2006, 74(3): 036116.
- [29] JEONG H, NEDA Z, BARABQCI A L. Measuring preferential attachment in evolving networks[J]. EPL, 2003, 61(4): 567.
- [30] BRODER A. Graph structure in the Web[J]. Computer Networks, 2000, 33(1-6): 309-320.
- [31] BARABASI E A. Emergence of scaling in random networks[J]. Science 1999, 286(5439): 509-512.
- [32] GALDARELI G. Scale-free networks[M]. Oxford: Oxford University Press, 2007.
- [33] PELTOMATI M, ALAVA M. Correlations in bipartite collaboration networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2006(01): 01010.
- [34] GARLASCHELLI D, LOFFREDO M I. Patterns of link reciprocity in directed networks[J]. Physical Review Letters, 2004, 93(6): 268701.
- [35] WATTS D J, STROZATZ S H. Collective dynamics of "small-world" networks[J]. Nature, 1998, 393(6684): 440-442.
- [36] NEWMAN M E J. Mixing patterns in networks[J]. Physical Review E, 2003, 67(8): 026126.
- [37] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(35): 35-41.
- [38] MILO R, ITZKOVITZ S, KASHTAN N, et al. Superfamilies of evolved and designed networks[J]. Science, 2004, 303(5663): 1538-1542.
- [39] Wikipedia Community. Wikipedia categorization [DB/OL]. [2010-4-29]. <http://en.wikipedia.org/wiki/Wikipedia:Categorization>.
- [40] CAPOCCI A, RAO F, CALDARELLI G, et al. Taxonomy and clustering in collaborative systems: the case of the on-line encyclopedia Wikipedia[J]. EPL, 2008, 81(2): 28006.
- [41] YULE U G. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S[J]. Royal Society of London Philosophical Transactions Series B, 1925, 213(3): 21-87.

- [42] HARRIS T E. The theory of branching processes[M]. Berlin: Springer-Verlag, 1963.
- [43] RIOS P D L. Power law size distribution of supercritical random trees[J]. EPL, 2001, 56(6): 898.
- [44] CALDARELLI G, CARTOZO C C, LOSRIOS P D, et al. Widespread occurrence of the inverse square distribution in social sciences and taxonomy[J]. Physical Review E, 2004, 69(3): 035101.
- [45] ENRIGHT A J, DONGEN S V, OUZOUNIS C A. An efficient algorithm for large-scale detection of protein families[J]. Nucl Acids Res, 2002, 30(7): 1575-1584.
- [46] CHO J, GARCIA-MOLINA H, PAGE L. Efficient crawling through URL ordering[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 161-172.
- [47] KLEINBERG J M. Authoritative sources in a hyperlinked environment[J]. JACM, 1999, 46(5): 604-632.
- [48] JESUS R, SCHWARTZ M, LEHMANN S. Bipartite networks of Wikipedia's articles and authors: a meso-level approach[C]//Proceedings of the 5th International Symposium on Wikis and Open Collaboration. New York: ACM, 2009: 1-10.
- [49] Wikipedia Community. WikiProjects[DB/OL]. [2010-04-26]. <http://en.wikipedia.org/wiki/WikiProjects>.
- [50] 于嘉. 网络时代的维基百科研究[J]. 图书馆论坛, 2005, 25(4): 247-249.
YU Jia. Wikipedia research in network[J]. Library Forum, 2005, 25(4): 247-249.
- [51] 王小梅. 传播学视角看Wiki的教育应用[J]. 软件导刊(教育技术), 2008, 9(3): 38-40.
WANG Xiao-mei. Education of Wiki by view of communication[J]. Software Guide (Educational Technology), 2008, 9(3): 38-40.
- [52] 同创, 钟文. 美情报系统悄然掀起变革打造情报维基百科翻版[EB/OL]. [2009-4-9]. <http://world.huanqiu.com/roll/2009-04/428013.html>.
TONG Chuang, ZHONG Wen. Changes in the U.S.intelligence system to create intelligence quietly launched replica Wikipedia[EB/OL]. [2009-4-9]. <http://world.huanqiu.com/roll/2009-04/428013.html>.
- [53] 李楠楠, 周涛, 张宁. 人类动力学基本概念与实证分析[J]. 复杂系统与复杂性科学, 2008, 5(2): 15-24.
LI Nan-nan, ZHOU Tao, ZHANG Ning. Concept and empirical results of human dynamics[J]. Complex Systems and Complexity Science, 2008, 5(2): 15-24.
- [54] 汪秉宏, 韩筱璞. 人类行为的动力学与统计力学研究[J]. 物理, 2010, 39(1): 28-37.
WANG Bing-hong, HAN Xiao-pu. Study of dynamics and statistical mechanics for human behaviors[J]. Physics, 2010, 39(1): 28-37.
- [55] VAZQUEZ A. Impact of memory on human dynamics[J]. Physica A: Statistical and Theoretical Physics, 2007, 373(3): 747-752.
- [56] 周涛. 网络“水军”不可放任[EB/OL]. [2009-12-27]. <http://opinion.people.com.cn/GB/10658053.html>.
ZHOU Tao .Network “water army” is not a laissez-faire [EB/OL]. [2009-12-27]. <http://opinion.people.com.cn/GB/10658053.html>.
- [57] ZHANG Y C. Manifesto for the Reputation Society[J]. First Monday, 2004, 9(7): 1-25.

编辑 蒋晓