

图灵指数——学术大数据下的跨领域跨年代学者影响力评估

姚宇航, 欧俊杰, 李洋, 傅洛伊, 王新兵, 陈贵海

上海交通大学电子信息与电气工程学院, 上海 200240

摘要

随着学术界规模的扩大, 面对领域和年代的巨大差异, 现有的衡量学者影响力的量化指标很难消除偏差, 从而进行准确衡量。通过观察1865—2016年、310个领域、14 223 183位学者、126 438 664篇论文和533 556 856次引用的数据集, 得出不同领域、不同年代的新增引用量分布均遵循幂律分布, 因此提出图灵指数, 通过归一化消除指数膨胀的影响, 衡量学者的绝对影响力。通过对诺贝尔奖、菲尔兹奖、图灵奖等获得者的图灵指数的比较发现, 尽管他们在领域和年代上存在很大差异, 但是在各自领域却有着相似的影响力。图灵指数为跨领域、跨年代学者影响力评估提供了全新的思路, 也为国内外学者影响力评估和基金审理提供了参考。

关键词

影响力评估; 图灵指数; 大数据挖掘; 学术网络

中图分类号: TN 929.5

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2019039

Turing index: cross-domain and cross-generation metric of unraveling scholars' impact in academic big data

YAO Yuhang, OU Junjie, LI Yang, FU Luoyi, WANG Xinbing, CHEN Guihai

School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract

Gauging the impact of scholars, despite numerous quantitative indicators proposed, remains hard thanks to nowadays gigantic and inflating academic networks accompanied by huge disparity among domains and generations. Data collected from 14 223 183 scholars, 126 438 664 papers, 533 556 856 references across 310 domains with a time spanning of year 1865 to 2016, suggest universe power-law distributed yearly increased citations of scholars across domains and generations. Turing index then eliminates the inflation via normalizing different scaling parameters in those power law distributions to assess the absolute domain impact of each scholar. Comparison of Turing index among renowned scholars like Nobel, Fields Medal and Turing laureates confirms their equal significance to their dedicated domains despite the vast domain discrepancies in generations and citations. Turing index provides a new way of thinking for cross-domain and cross-generation impact evaluation, and can provide reference for scholars' impact assessment and national science foundations worldwide.

Key words

impact evaluation, Turing index, big data mining, academic network

1 引言

目前,越来越多的人投身于学术界,并在众多研究领域取得成就^[1]。如何量化众多学者的影响力已引起了全世界的关注,对学者影响力进行评估,对基金项目进行评审和学术奖项评选等都有很重要的意义,可以推动学术界进一步的发展。目前,最广泛使用的评价指标是引用量^[2-9],即学者发表论文的总被引用次数。另一个广泛使用的指标是H指数^[10-16],它根据研究者被引用最多的论文数目和被引用次数来衡量学者的产出和影响力。上述两种方法均未考虑领域和年代的影响,研究人员基于引用量和H指数提出了一些变体^[17-23],用于衡量学者的个人水平。

随着时间的推移,出现了很多解决学术影响力量化问题的方法^[24-32]。由于领域本身的迅速发展和不同领域的不平衡发展^[33],一些工作分别考虑了不同年代、不同领域对

影响力量化的影响^[34-36],并且基于引用量和H指数的归一化来平衡不同年代、不同领域学者的相对位置^[37-39]。在消除年代和领域的影响时,这些指标都没有合理考虑学术界规模迅速扩大造成的影响,导致评估结果并不理想。

2 学术界的指数膨胀

为了对不同领域、不同年代的学者进行影响力评估,笔者对学者和论文数据进行了统计和收集,数据来源于Acemap^[40]和微软学术^[9]。微软学术利用论文发表期刊、标题和正文等信息对论文所在领域进行层次划分,划分结果得到了学术界和工业界的广泛认可。本文利用Acemap整合微软学术从1865年到2016年的310个领域、14 223 183位学者、126 438 664篇论文和533 556 856次引用的数据集,进行学者影响力评估的探索。主要领域的学者、论文和引用量等统计结果见表1。

表1 数据集主要领域信息统计

| 主要领域 | 论文数/篇 | 引用量/次 | 学者数/位 | 子领域数/个 | 时间 |
|-------|-----------|-----------|-----------|--------|------------|
| 政治学 | 229 474 | 130 713 | 174 908 | 250 | 1928—2016年 |
| 药学 | 6 593 836 | 4 798 767 | 6 699 626 | 1 675 | 1875—2016年 |
| 哲学 | 1367 394 | 935 214 | 1 321 093 | 1 665 | 1874—2016年 |
| 化学 | 7 717 267 | 5 510 693 | 6 262 877 | 3 522 | 1865—2016年 |
| 经济学 | 3 928 782 | 2 801 510 | 3 313 730 | 2 349 | 1906—2016年 |
| 历史学 | 620 355 | 390 641 | 727 469 | 656 | 1890—2016年 |
| 数学 | 7 909 950 | 5 885 640 | 5 487 308 | 8 025 | 1889—2016年 |
| 生物学 | 7 751 483 | 5 619 200 | 7 038 316 | 4 174 | 1888—2016年 |
| 计算机科学 | 7 512 849 | 5 718 499 | 5 474 060 | 5 181 | 1889—2016年 |
| 材料科学 | 2 553 860 | 1 854 914 | 2 325 120 | 945 | 1902—2016年 |
| 社会学 | 3 107 882 | 2 130 742 | 2 755 201 | 2 048 | 1897—2016年 |
| 地质学 | 2 624 463 | 1 873 559 | 2 570 621 | 2 120 | 1895—2016年 |
| 心理学 | 4 354 577 | 3 145 985 | 4 278 842 | 2 430 | 1897—2016年 |
| 环境科学 | 441 410 | 289 355 | 585 570 | 178 | 1901—2016年 |
| 物理学 | 9 471 698 | 6 707 746 | 6 142 982 | 6 620 | 1865—2016年 |
| 工程学 | 5 819 259 | 4 323 780 | 4 374 996 | 2 689 | 1889—2016年 |
| 艺术学 | 456 271 | 320 508 | 548 785 | 422 | 1910—2016年 |
| 地理学 | 571 233 | 423 779 | 665 790 | 502 | 1912—2016年 |
| 商学 | 1 086 437 | 767 941 | 1 151 164 | 536 | 1894—2016年 |

如图1所示,历年新增论文、引用量和学者数目均呈指数级增长,可以看出学术界在飞速发展。从图1可以看到两个明显的凹陷,分别在1917年和1941年左右。笔者认为这是由两次世界大战造成的,可见军事对科学研究也会产生很大影响。

本文对各个主要领域内部的发展情况进行了统计,如图2所示,不同领域由于发展情况不同,线条的斜率和截距不同,但均呈指数增长。其中 c 为引用量。图3显示了指数膨胀的学术网络,图3中红线表示每年活跃学者的数量,网络中的彩色节点表示不同领域的活跃学者,每个簇代表一个领域,可见领域的膨胀是非常迅速的。图4为从1950—2014年16个物理子领域的历年引用量增长情况,该领域内部的子领域发展也是不均衡的。

3 指数膨胀效应对学者影响力评估的影响

随着学术界的指数膨胀,发表更多的文章、获得更多的引用量变得更加容易,在这种情况下,以往的文献引用和H指数等指标无法很好地描述学者在以下3种指数膨胀效应下的影响。

一是时间的膨胀效应。学者在不同时期的相同引用量是不等价的。由于学者和论文数量不断增加,现在的学者论文更容易被引用。1980年被引用100次的学者可能比2018年被引用100次的学者具有更高的影响力。同样,H指数也没有充分考虑时间效应。因此,仅仅采用引用量或H指数可能会导致一种错误的判断,即年长的学术巨头和年轻的学者对学术界具有相同的影响力。

二是领域的膨胀效应。一个重要但通

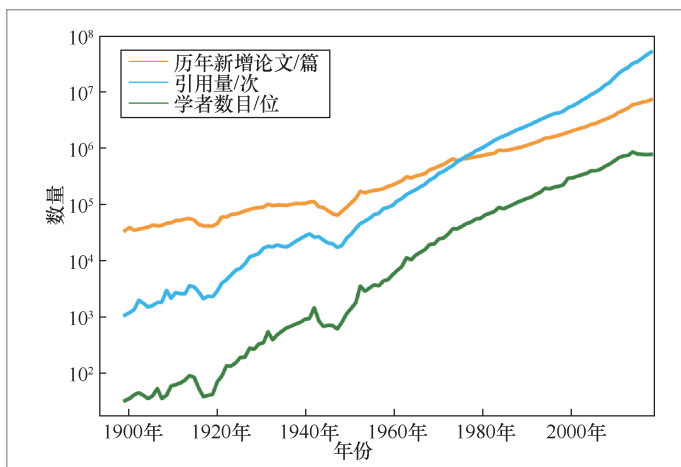


图1 历年新增论文、引用量和学者数目统计

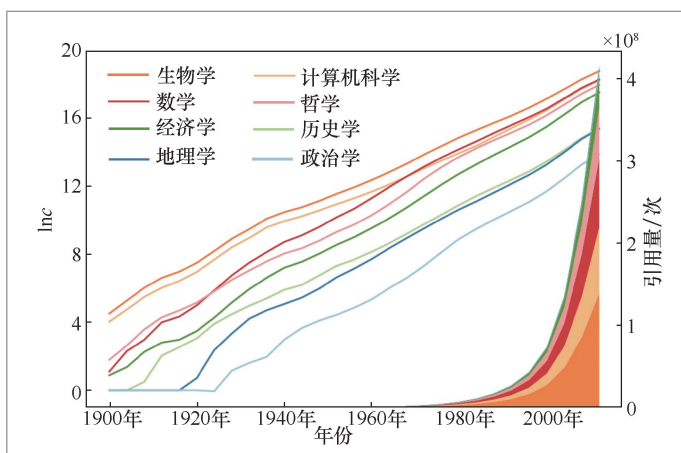


图2 主要领域历年引用量

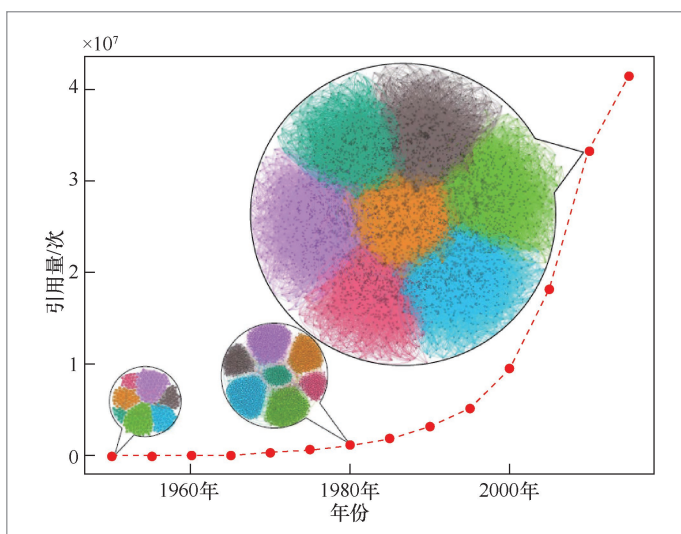


图3 学术网络指数膨胀可视化

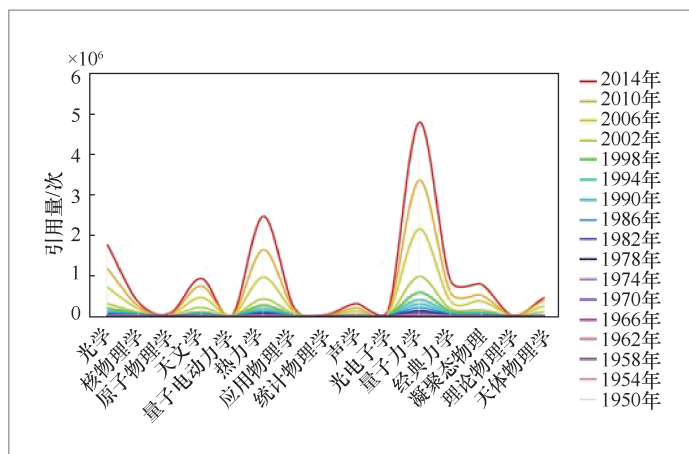


图4 物理子领域历年引用量增长

常被忽视的因素是研究领域发展水平不均衡带来的影响。虽然不同领域的引用量均呈指数增长,但其膨胀速度和状态不同,发展水平也不同。例如,生物学的年总被引次数大约是政治学的40倍。图3给出了活跃学者在不同领域的膨胀情况,领域的大小按照它们自己的节奏发展,即使在单个领域(如图4所示的物理学领域),其子领域的发展水平也是完全不同的。拥有相同引用量的不同领域的学者,他们在各自领域的影响力是不相等的。在蓬勃发展的热门领域和在小众领域得到相同的引用量是不等价的。

三是年代的膨胀效应。即使在同一时间、同一领域内拥有相同的引用量和H指数,但在学者们活跃年代不同的情况下,他们的影响力也可能存在差异。活跃年代是指学者在所属领域内被引频次数快速增长的时期。如图5所示,威廉·菲利普斯和迈克尔·弗莱施豪尔是量子光学领域的两位杰出物理学家。根据谷歌学术^[8]统计,截至2018年,威廉·菲利普斯共获得引用23 750次,H指数为66,而迈克尔·弗莱施豪尔的被引用次数和H指数分别为21 578次和60。虽然他们有相似的引用

和H指数,但威廉·菲利普斯(活跃年代为20世纪90年代)比迈克尔·弗莱施豪尔(活跃年代为21世纪初)有更早的活跃年代。也就是说,在一个全新的、学者较少的领域获得相同的引用通常比在一个成熟的领域难得多,而且领域的开创者比追随者影响力更大。因此,目前威廉·菲利普斯对量子光学的影响大于迈克尔·弗莱施豪尔,而实际中,威廉·菲利普斯获得了诺贝尔物理学奖,也验证了本文的观察。

然而,上述观察中得出的结论主要是基于人的主观经验。那么是否有一个量化指标可以消除领域和年代的膨胀效应,为学者提供科学客观的评价?作为一种大胆而又严谨的尝试,本文提出一种新的度量标准,即图灵指数(Turing index),以重新评估学者的影响。

4 图灵指数

为了消除不同领域、不同年代对学者影响力评价的影响,本文首先研究了各个领域每年学者的新增学者引用量分布,即学者的数量与特定年份内相应增加的引用分布情况。为了寻找合适的量化指标,本文从数据集中提取数据,并绘制历年学者新增引用量分布图。

图6为2015年物理领域的引文网络,图6中每个节点代表2015年在物理领域发表或引用论文的活跃学者,红色节点的度数较高,绿色节点的度数较低,从图6可以看出,极少数红色节点被绿色节点包围,由此笔者猜想引用量分布符合幂律分布,即红色节点获得大部分的引用次数,大多数的绿色节点被引用次数则少得多。为了验证这一猜想,基于对每年论文的新增引用量和这些论文的所属学者和领域,本文计算出1 228 960 765条四元组,

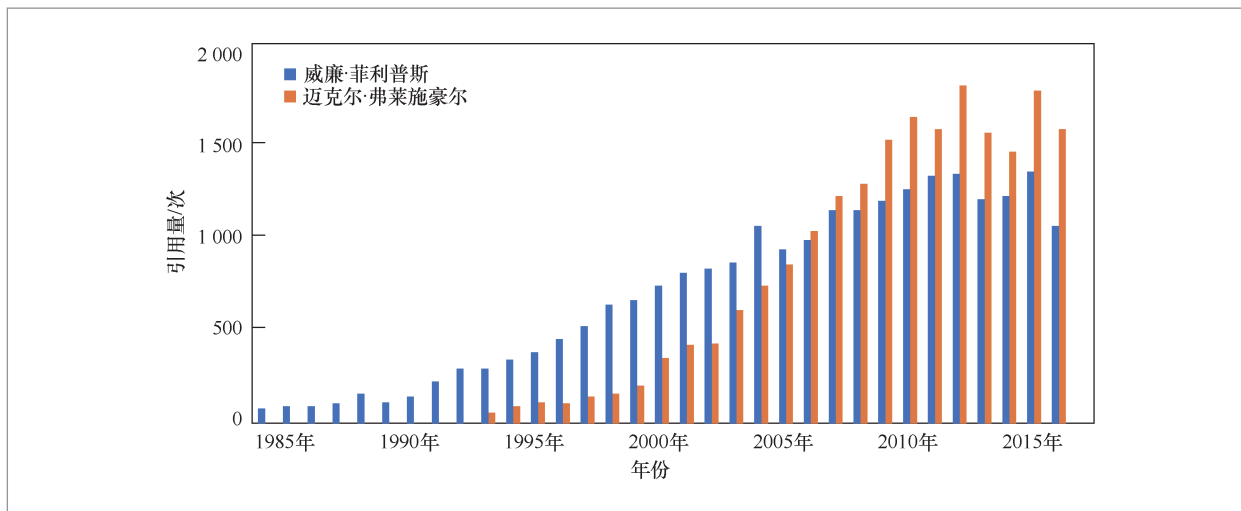


图5 两学者历年新增引用量比较

每一条四元组都由学者、领域、年份和引用组成，表示每年每个学者在不同领域的新增引用。

利用四元组，本文用对数图绘制学者的引用量分布情况，如图7所示，尽管不同领域的学者数量不同，但这些引用量分布都可以通过线性函数很好地拟合，说明它们都遵循幂律分布^[41]，其形式如式(1)所示。

$$n = N(k-1)c^{-k} \quad (1)$$

其中， n 表示每年被引用次数增加数值为 c 的学者数目， N 为当年被引用的活跃学者总数， k 为比例指数。

引用量分布的归一化是为了让不同幂律分布公式符合相同的参数。为此，本文假定目标归一化领域 f_0 有 N_0 个学者和标度系数 k_0 ，同时引用量为 c_0 的学者有 N_0 个，那么其引用分布为：

$$n_0 = N_0(k_0 - 1)c_0^{-k_0} \quad (2)$$

同样，本文假设未归一化领域 f_i ($i \geq 1$)的引用分布为：

$$n_i = N_i(k_i - 1)c_i^{-k_i} \quad (3)$$

为了使领域 f_i 具有与领域 f_0 相同的分布，本文需要在 $n_i = n_0$ 时，使得新增引用量 c_i 归一化为 c_0 ，其中， c_0 表示在领域 f_i 中引用

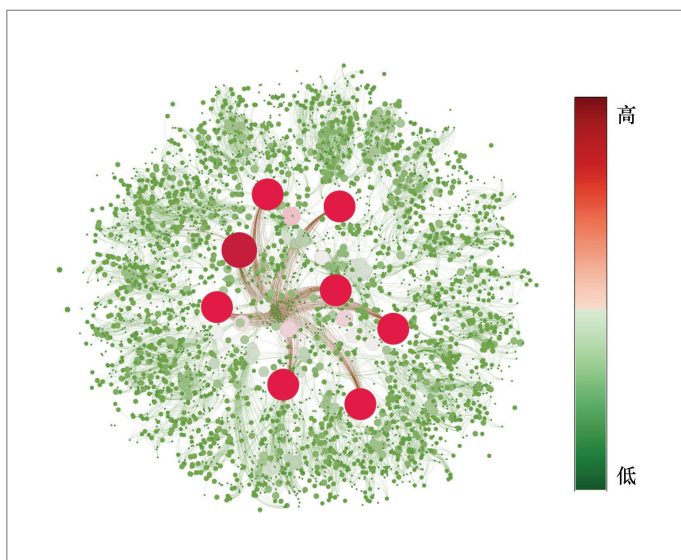


图6 2015年物理领域学者引文网络

量为 c_i 的学者 p 归一化后在领域 f_0 中的绝对位置。

然后有：

$$N_i(k_i - 1)c_i^{-k_i} = N_0(k_0 - 1)c_0^{-k_0} \quad (4)$$

因为 N_0 和 k_0 是常量，所以有：

$$c_0 \propto \frac{c_i^{\frac{k_i}{k_0}}}{N_i^{\frac{k_i}{k_0}}(k_i - 1)^{\frac{1}{k_0}}} = \frac{c_i^{\alpha k_i}}{N_i^\alpha (k_i - 1)^\alpha} \quad (5)$$

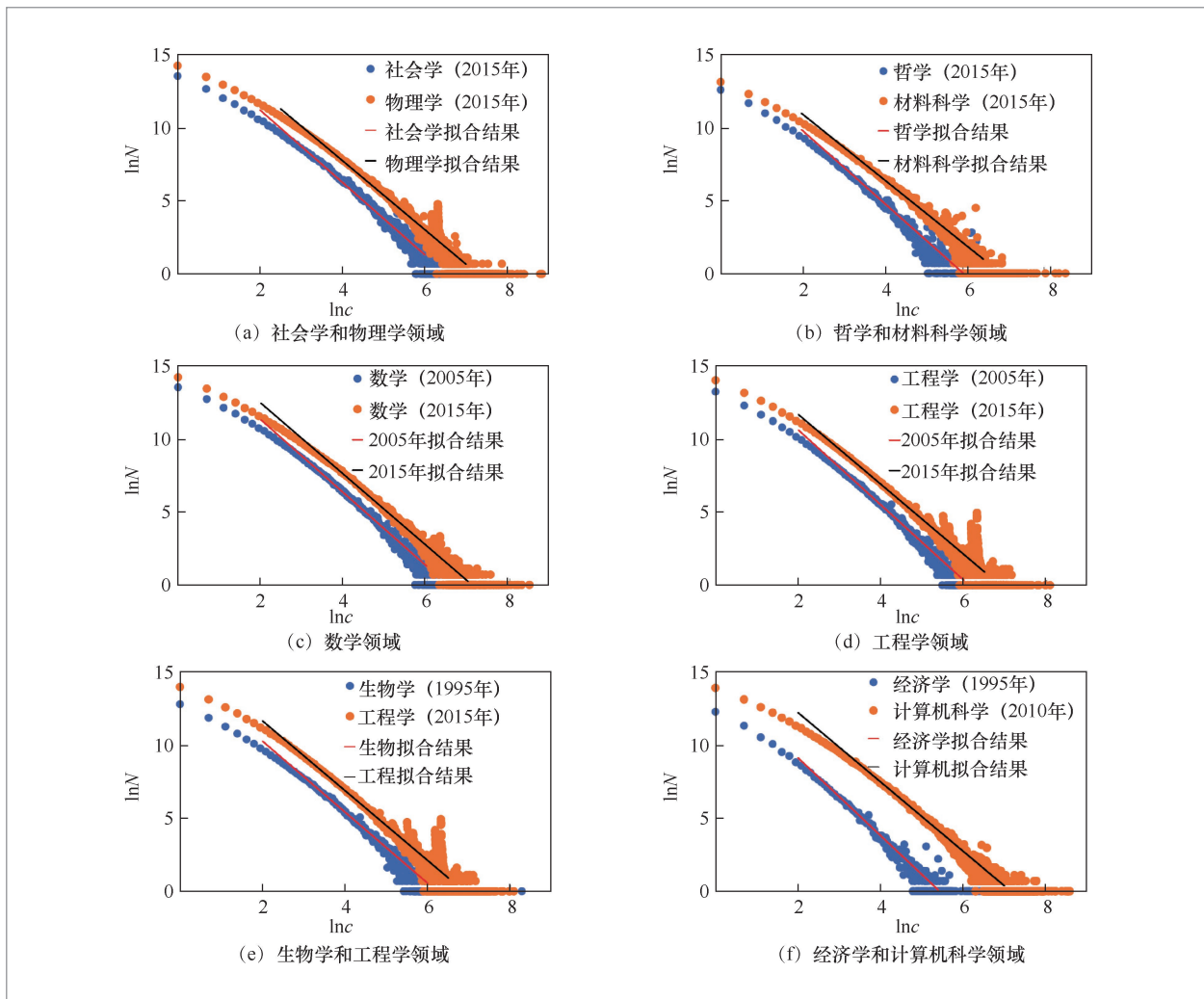


图7 不同领域不同年代新增引用量分布情况

其中, $\alpha = \frac{1}{k_0}$ 。针对归一化, 本文将 c_0 作为学者的影响力 I , 学者 p 在该领域的成就可表示为:

$$A = c^{\alpha k} \quad (6)$$

领域通胀水平为:

$$D = N^{\alpha} (k-1)^{\alpha} \quad (7)$$

领域的 k 值越高, 则在该领域越难得到同样的引用量 c , 因此, 相同引用量产生的影响力越大。活跃学者人数 N 的增加导致领域通胀水平 D 的增加, 从而刻画指数膨胀效应。 α 是一个常量, 可以看作

一个平衡项, 以防止归一化度量的欠收敛和过收敛。

图8显示了在3种情况下, 领域和年代的归一化过程: 相同年代不同领域、不同年代相同领域、不同年代不同领域。归一化后, 不同领域、不同年代的引用量分布基本一致。

基于式(5)和提出的指标 A 和 D , 学者 p 在领域 f 从 s 年到 t 年的总影响力的图灵指数 $T_{p,f,t}$ 为:

$$T_{p,f,t} = \sum_{i=s}^t I_{p,f,i} \quad (8)$$

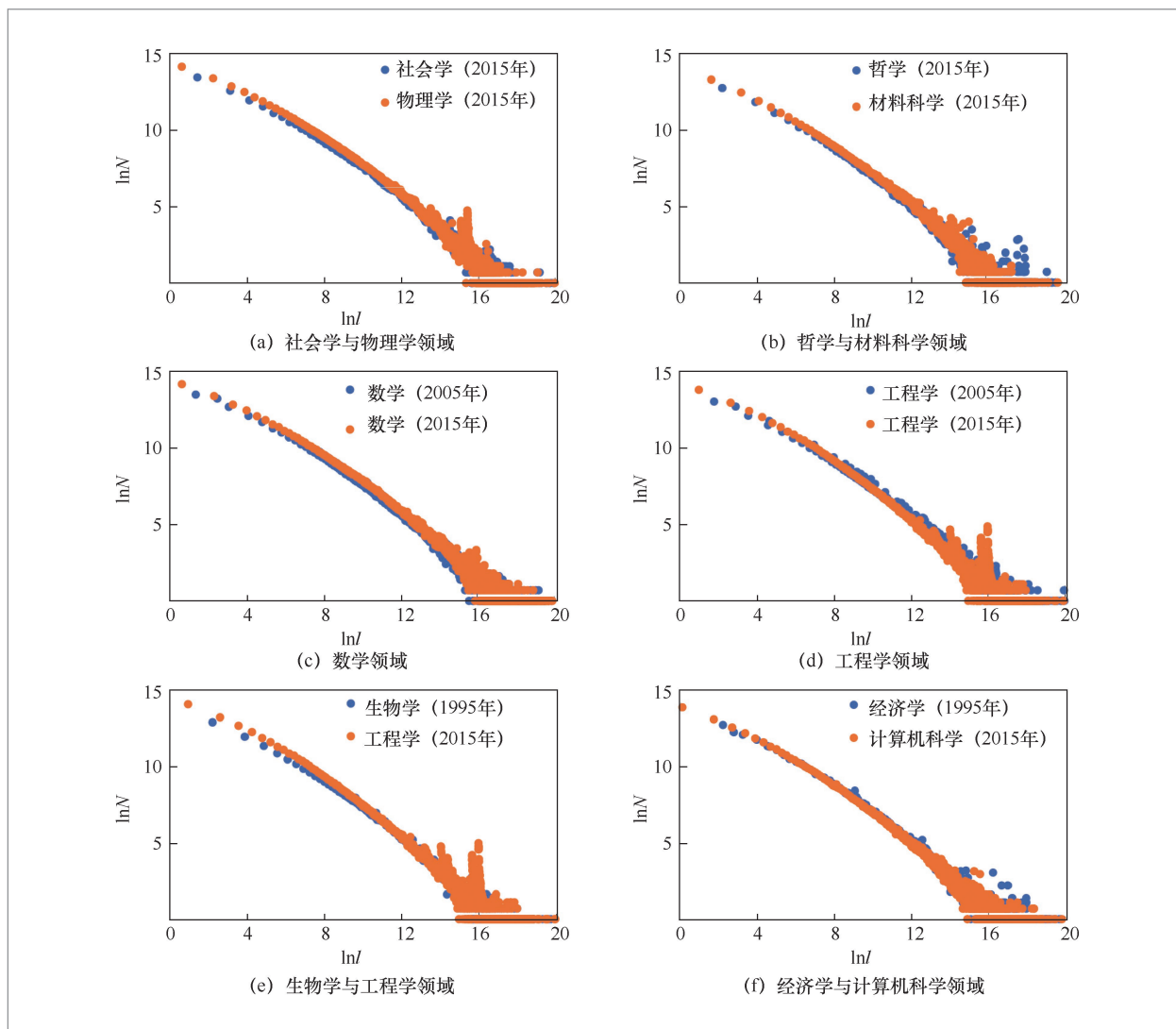


图8 不同领域不同年代新增引用量分布归一化结果

利用式(8),可以在相同条件下评估跨领域和跨年代学者的影响。

5 跨领域、跨年代学者影响力对比

本文建立了一个基于数据集的系统来计算和存储不同领域和不同年代学者的图灵指数。一个学者可以属于多个领域,通过对每年增加的引用量进行归一化处理,本文得到了1 228 960 765条包含学者、领域、年

份和图灵指数信息的四元组。为了更好地理解图灵指数如何消除通胀差异、评估不同领域不同年代的学者的影响力,本文分别基于引用量和图灵指数衡量诺贝尔奖、菲尔兹奖和图灵奖得主的影响力,并进行比较。截至2015年入选学者的获奖统计数据见表2。

本文对相同年代不同领域、不同年代相同领域和不同年代不同领域这3种情况进行对比。

(1) 相同年代不同领域

相同年代意味着相关学者的研究生涯

表2 学者获奖统计

| 姓名 | 获奖名称 | 年份 | 贡献 | 引用量/次 | 图灵指数 |
|---------------------|------------|-------|---------|--------|-------|
| Arieh Warshel | 诺贝尔化学奖 | 2013年 | 复杂化学系统 | 3 222 | 2 254 |
| Barbara McClintock | 诺贝尔生理学或医学奖 | 1983年 | 可移动遗传因子 | 3 809 | 2 245 |
| Bruce Beutler | 诺贝尔生理学或医学奖 | 2011年 | 先天免疫活化 | 15 650 | 4 558 |
| Edmund M Clarke | 图灵奖 | 2007年 | 模型检查 | 13 882 | 6 078 |
| Eric Betzig | 诺贝尔化学奖 | 2014年 | 荧光显微镜 | 6 422 | 4 509 |
| Jean Tirole | 诺贝尔经济学奖 | 2013年 | 市场权力及规管 | 3 031 | 5 650 |
| Johann Deisenhofer | 诺贝尔化学奖 | 1988年 | 光合作用中心 | 5 696 | 3 662 |
| John E Hopcroft | 图灵奖 | 1986年 | 算法设计 | 8 449 | 6 020 |
| Oliver E Williamson | 诺贝尔经济学奖 | 2009年 | 经济治理 | 5 131 | 5 237 |
| Paul Nurse | 诺贝尔生理学或医学奖 | 2001年 | 细胞周期调节器 | 11 048 | 2 613 |
| Pierrelouis Lions | 菲尔兹奖 | 1994年 | 偏微分方程 | 3 908 | 2 747 |
| Ralph M Steinman | 诺贝尔生理学或医学奖 | 2011年 | 树突细胞 | 18 683 | 5 820 |
| Roald Hoffmann | 诺贝尔化学奖 | 1981年 | 化学反应过程 | 1 737 | 2 261 |
| Shuji Nakamura | 诺贝尔物理学奖 | 2014年 | 蓝色发光二极管 | 4 412 | 2 832 |
| Susumu Tonegawa | 诺贝尔生理学或医学奖 | 1987年 | 抗体的多样性 | 17 419 | 5 150 |
| Sydney Brenner | 诺贝尔生理学或医学奖 | 2002年 | 基因调控 | 6 896 | 2 125 |
| Terence Tao | 菲尔兹奖 | 2006年 | 偏微分方程 | 9 706 | 3 750 |

大致是在同一时期开始的, 引用量的增长速度因领域而异。由于一个领域的学者数量可能是另一个领域的30倍, 所以领域的发展水平可能会形成巨大的对比。从图9(a)和图9(b)可以看出, 诺贝尔生理学奖得主Paul Nurse和菲尔兹奖得主Pierrelouis Lions之间引用量的显著差异并不意味着他们的影响力存在巨大差距, 因为他们不同的领域。对各自历年新增引用量进行归一化处理后, 两位学者的图灵指数表明, 他们对各自领域的影响是相当的, 发展轨迹也几乎相同。同样现象也出现在图9(c)和图9(d)中, 其中, 诺贝尔生理学奖获得者Ralph M Steinma在免疫学这一庞大的领域中被引用的次数较多, 而诺贝尔经济学奖获得者Jean Tirole在工业管理这一领域中被引用的次数较少。通过引用量分析, Steinma的影响力似乎比Tirole大得多, 但图灵指数显示两者的影响力相同, 甚至轨迹相同。另一个发现是, 由于

领域迅速发展导致的膨胀, 学者们可能会遵循不同的发展模式, 这些发展模式无法在传统的总引用量衡量中体现。例如, 在图9(e)和图9(f)中, Bruce Beutler的图灵指数呈线性稳步增长, 而Eric Betzig在开始发展时较平缓, 1992年后出现了一个跳跃, 这说明他可能在那一年取得了很大的突破。这一观察可以被证实, 在1992年, Eric Betzig发表了一篇关于显微镜^[42]的论文, 而这是他获得诺贝尔奖的主要贡献。

(2) 不同年代相同领域

不同年代意味着学者们在不同的年代开始研究。相同的领域表明, 相同年份的引用量增长具有相同的影响力。由于快速的领域膨胀, 一个领域中的不同年代具有完全不同的发展水平。例如, 在图9(g)和图9(h)中, 虽然John E Hopcroft和Edmund M Clarke都是图灵奖的获得者, 但是在后者的活跃年代, 其

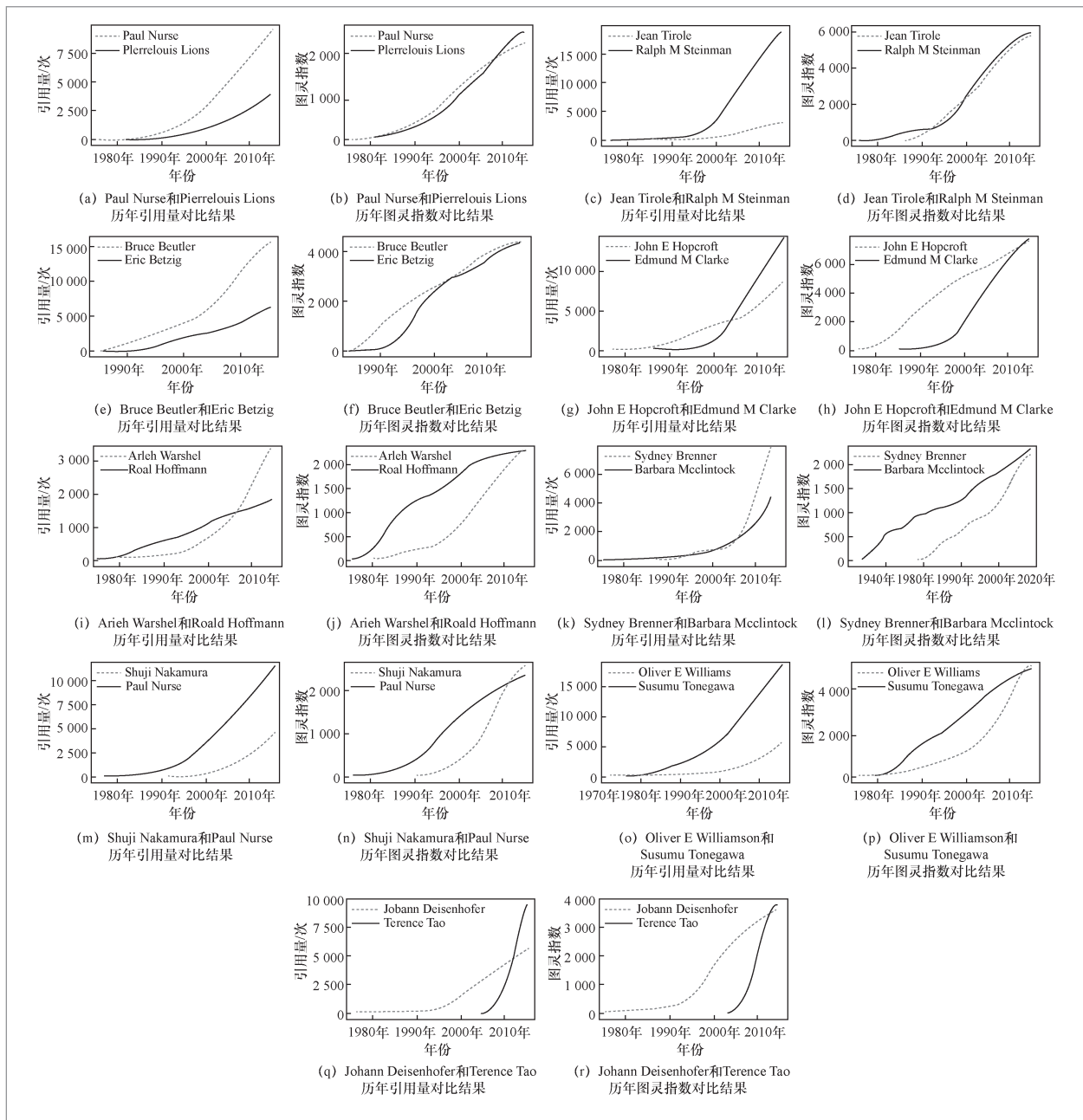


图9 学者的引用量和图灵指数对比结果

在算法领域获得的引用比前者多得多，因为随着时间的推移，领域膨胀越大，越容易获得高引用。然而，他们的图灵指数表明，他们具有相同的影响。同样的现象在图9(i)~图9(l)中也出现了，尽管前人的引用量可能比后来的学者少一半，

但是他们的影响力仍然是相等的。在此基础上，图灵指数在评价不同时代学者对领域的历史影响方面具有很强的优势。

(3) 不同年代不同领域

同样地，当同时考虑领域和年代时，

影响力的度量也遵循同样的规则,但是有更多的比较形式。如图9(m)和图9(n)所示,诺贝尔生理学奖得主Paul Nurse比诺贝尔物理学奖得主Shuji Nakamura有更早的活跃年代,被引用的次数也更多。虽然较早的年代和较高的引用量表明较高的影响,但是Paul Nurse的所属领域要比Nakamura的所属领域大,根据图灵指数,这两位学者有着相似的影响力。另一种相反的情况如图9(o)和图9(p)所示,Oliver E Williams有更早的活跃年代,并且所在领域比Susumu Tonegawa的领域小,Oliver E Williams的引用量仅为Susumu Tonegawa的三分之一左右,但图灵指数显示,前者的影响力在进入21世纪后增长更快,2010年左右变得和后者一样,这意味着前者的贡献在近几年得到了认可,这一现象可以再次得到事实印证。Oliver E Williams虽然比Susumu Tonegawa年长,但是他在2009年获得了诺贝尔奖,Susumu Tonegawa在1987年获得了诺贝尔奖,图灵指数因而能够发现这样的现象。同时,领域膨胀能够导致引用的快速增长,通过消除领域通胀的影响,图灵指数有助于发现真正的学术新星,在图9(q)和图9(r)中,菲尔兹奖获得者Terence Tao在通过图灵指数消除领域通胀影响后仍然增长迅速,这表明其影响力增长速度远远快于领域的通胀。综上所述,尽管在引用量、领域、年代等方面存在差异,但这些顶尖学者对其所属领域的影响是相似的。

从学者的发展历程来看,由于现有理论并不能准确地刻画真实情况下的学术网络,学者的生涯并不能无限延长,论文发表亦不能无限增长。图灵指数下的学术网络始终符合幂律分布,由于资深学者影响力较高,“富者愈富”现象会在较短的时间周期出现,与理论贴合。但随着网络演

进,经典的论文会保有其影响力,学者会更倾向于引用新近前沿的论文,图灵指数更高的学者在若干年后并不能因为过去图灵指数高而扩大优势,而是与当前的论文发表和学者工作的影响力息息相关,在过去图灵指数较低的学者,经过若干年的发展之后,有可能跟上甚至超越过去图灵指数更高的学者。

6 结束语

图灵指数可发现引用量增长的本质,不仅有助于确定和比较不同领域不同年代学者的历史影响,而且可为领域排名和科学奖项提供可靠的评价指标。从应用角度来看,由于学者影响评价的准确性与数据的可靠性密切相关,图灵指数未来的研究工作将侧重于提高学者数据的准确性和泛化性,同时尝试用图灵指数对学者生涯轨迹进行刻画。从理论化角度来看,由于学者的引用量和学者关系可以通过社交网络中的理论来刻画,因此可以试图采用网络演进理论对图灵指数的评估标准进行微观解释,并将图灵指数应用到更广泛的领域影响力评估中。

参考文献:

- [1] EVANS R. Open access and global participation in science[J]. Science, 2009, 323(5917): 1025.
- [2] GARFIELD E. “Science citation index” — a new dimension in indexing[J]. Science, 1964, 144(3619): 649–654.
- [3] EVANS J A, FOSTER J G. Metaknowledge[J]. Science, 2011, 331(6018): 721–725.
- [4] AZOULAY P. Research efficiency: turn the scientific method on ourselves[J]. Nature, 2012, 484(7392): 31.

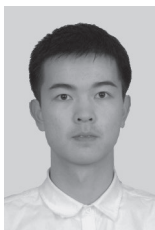
- [5] BARABASI A L, SONG C, WANG D. Publishing: handful of papers dominates citation[J]. *Nature*, 2012, 491(7422): 40.
- [6] OWENS B. Research assessments: judgement day[J]. *Nature*, 2013, 502(7471): 288.
- [7] HUTCHINS BI, YUAN X, ANDERSON J M, et al. Relative citation ratio (RCR): a new metric that uses citation rates to measure influence at the article level[J]. *PLOS Biology*, 2016, 14(9): e1002541.
- [8] HARZING A, VAN D W R. Google Scholar as a new source for citation analysis[J]. *Ethics in Science and Environmental Politics*, 2008, 8: 61-73.
- [9] SINHA A, SHEN Z, SONG Y, et al. An overview of microsoft academic service (MAS) and applications[C]// The 24th International Conference on World Wide Web, May 18-22, 2015, Florence, Italy. New York: ACM Press, 2015: 243-246.
- [10] HIRSCH J E. An index to quantify an individual's scientific research output[J]. *The National Academy of Science*, 2005, 102(46): 16569-16572.
- [11] BARILAN J. Which H-index? a comparison of wos, scopus and google scholar[J]. *Scientometrics*, 2008, 74(2): 257-271.
- [12] EGGHE L, ROUSSEAU R. An informetric model for the Hirsch-index[J]. *Scientometrics*, 2006, 69(1): 121-129.
- [13] LU L, ZHOU T, ZHANG Q M, et al. The H-index of a network node and its relation to degree and coreness[J]. *Nature Communications*, 2016, 7: 10168.
- [14] KE Q, FERRARA E, RADICCHI F, et al. Defining and identifying Sleeping Beauties in science[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(24): 7426-7431.
- [15] HIRSCH J E. Does the H-index have predictive power[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(49): 19193-19198.
- [16] HIRSCH J E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship[J]. *Scientometrics*, 2010, 85(3): 741-754.
- [17] LEO E. Theory and practise of the G-index[J]. *Scientometrics*, 2006, 69(1): 131-152.
- [18] ALONSO S, CABRERIZO F J, HERRERA V E, et al. H-index: a review focused in its variants, computation and standardization for different scientific fields[J]. *Journal of Informetrics*, 2009, 3(4): 273-289.
- [19] BORNMANN L, DANIEL H D. The state of H index research[J]. *EMBO Reports*, 2009, 10(1): 2-6.
- [20] BORNMANN L, MUTZ R, HUG S E, et al. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants[J]. *Journal of Informetrics*, 2011, 5(3): 346-359.
- [21] FRANCESCHINI F, GALETTO M, MAISANO D, et al. The success-index: an alternative approach to the h-index for evaluating an individual's research output[J]. *Scientometrics*, 2012, 92(3): 621-641.
- [22] PETERSEN A M, SUCCI S. The Z-index: a geometric representation of productivity and impact which accounts for information in the entire rank-citation profile[J]. *Journal of Informetrics*, 2013, 7(4): 823-832.
- [23] VANECK N J, WALTMAN L. Generalizing the H-and G-indices[J]. *Journal of Informetrics*, 2008, 2(4): 263-271.
- [24] GILES C L, COUNCILL I G, GRAY J N. Who gets acknowledged: measuring scientific contributions through automatic acknowledgment indexing[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(51): 17599-17604.
- [25] SINATRA R, WANG D, DEVILLE P, et al. Quantifying the evolution of individual scientific impact[J]. *Science*, 2016, 354(6312): 5239.
- [26] WANG D S, SONG C M. Quantifying long-term scientific impact[J]. *Science*,

- 2013, 342(6154): 127–132.
- [27] UZZI B, MUKHERJEE S, STRINGER M, et al. Atypical combinations and scientific impact[J]. *Science*, 2013, 342(6157): 468–472.
- [28] FORTUNATO S, BERGSTROM C T, BÖRNER K, et al. Science of science[J]. *Science*, 2018, 359(6379): 185.
- [29] VOLPP L. Complex networks: structure and dynamics[J]. *Physics Reports*, 2006, 424(4): 175–308.
- [30] SCHULZ C, UZZI B, HELBING D, et al. A network-based citation indicator of scientific performance[J]. *Computer Science*, 2018, arXiv: 1807. 04712.
- [31] LYU L Y, CHEN D, REN X L, et al. Vital nodes identification in complex networks[J]. *Physics Reports*, 2016, 650: 1–63.
- [32] CLAUSET A, ARBESMAN S, LARREMORE D B. Systematic inequality and hierarchy in faculty hiring networks[J]. *Science Advances*, 2015, 1(1): e1400005.
- [33] PAN R K, PETERSEN A M, PAMMOLLI F, et al. The memory of science: inflation, myopia, and the knowledge network[J]. *Journal of Informetrics*, 2018, 12(3): 656–678.
- [34] YIN Y, WANG D . The time dimension of science: connecting the past to the future[J]. *Journal of Informetrics*, 2017, 11(2): 608–621.
- [35] ALLESINA S, KORDING K P, ACUNA D E. Future impact predicting scientific success[J]. *Nature*, 2012, 489(7415): 201.
- [36] LU L, YANG W, ROBERTA S, et al. Hot streaks in artistic, cultural, and scientific careers[J]. *Nature*, 2018.
- [37] KAUR J, RADICCHI F, MENCZER F. Universality of scholarly impact metrics[J]. *Journal of Informetrics*, 2013, 7(4): 924–932.
- [38] RADICCHI F, FORTUNATO S, CASTELLANO C. Universality of citation distributions: toward an objective measure of scientific impact[J]. *Proceedings of the National Academy of Sciences of the United State of America*, 2008, 105(45): 17268–17272.
- [39] IGLESIAS J E, PECHARROM C. Scaling the h-index for different scientific ISI fields[J]. *Scientometrics*, 2007, 73(3): 303–320.
- [40] TAN Z, LIU C, MAO Y, et al. AceMap: a novel approach towards displaying relationship among academic literatures[C]// *The 25th International Conference Companion on World Wide Web*, April 11–15, 2016, Montreal, Canada. New York: ACM Press, 2016: 437–442.
- [41] ADAMIC L A, HUBERMAN B A. Power-law distribution of the world wide web[J]. *Science*, 2000, 287(5461): 2115
- [42] BETZIG E, TRAUTMAN J K. Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit[J]. *Science*, 1992, 257(5067): 189–195.

作者简介



姚宇航(1997-)，男，上海交通大学电子信息与电气工程学院博士生，主要研究方向为社交网络数据挖掘、影响力最大化和隐私保护。



欧俊杰 (1997-), 男, 上海交通大学电子信息与电气工程学院博士生, 主要研究方向为机器学习、大数据挖掘。



李洋 (1996-), 男, 上海交通大学电子信息与电气工程学院本科生, 主要研究方向为自然语言处理。



傅洛伊 (1987-), 女, 博士, 上海交通大学特别副研究员, 主要研究方向为社交网络大数据、信息传输与隐私保护以及知识发现与知识度量。



王新兵 (1975-), 男, 上海交通大学特聘教授, 电子信息与电气工程学院副院长, 国家自然科学基金委员会杰出青年基金获得者, ACM中国理事会副主席, 主要研究方向为物联网性能分析与优化设计。担任网络领域重要期刊 *IEEE/ACM Transactions on Networking*、*IEEE Transactions on Mobile Computing* 以及 *ACM Transactions on Sensor Networks* 的编委。2014—2018年连续5年入选爱思唯尔 (Elsevier) 中国高被引学者榜单。获宝钢优秀教师特等奖提名奖, 国家级教学成果二等奖 (排名第二), 上海市教学成果特等奖 (排名第二)。



陈贵海 (1963-), 男, 上海交通大学教授、博士生导师, 主要研究方向为对等计算、数据处理、传感器网络、路由算法、高性能计算机体系结构、组合数学等。发表的文章被Google Scholar引用10 000余次, SCI引用1 000余次, ESI高被引论文4篇, 11次获得国际会议论文奖。2006年获江苏省六大人才高峰称号, 2008年获国家杰出青年科学基金, 2011年获国务院政府特殊津贴, 2015年获教育部自然科学奖一等奖 (排名第一), 2017年入选中国计算机学会会士 (CCF Fellow), 2018年获江苏省科学技术奖一等奖 (排名第一)。现任中国计算机学会分布式计算与系统专业委员会主任委员, ACM SIGCOMM China副主席。

收稿日期: 2019-06-03