

基于小世界网络模型 建设学术期刊审稿专家数据库网络群*

史朋亮¹⁾ 吴晨²⁾

1) 中国科学院物理研究所《物理学报》编辑部, 100190; 2) 清华大学 84-48 信箱中国知网知识传播工程研究院, 北京

摘要 概括国际期刊业建设审稿专家数据库的动态, 总结当前期刊编辑部所使用的审稿专家数据库的来源, 分析中国期刊业在审稿专家数据库建设上面临的机遇。基于小世界网络模型的观点, 提出在中国知网(CNKI)作者库基础上构建审稿专家网络系统的设想和方案, 并对所能达到的目标作了展望。

关键词 审稿专家; 小世界网络; 数据库网络群; 动态更新

Construction of networks group of peer reviewer database for academic periodicals based on small-world-network model // SHI Pengliang, WU Chen

Abstract We give an overview of trends on establishment of peer reviewer database for academic periodicals in the world, conclude various resources of it, and sharpen current picture of challenge and chance on construction of a new peer reviewer database. Based on conception of a small-world-network model, we promote to build a system of networks group for peer reviewer database on authors in Chinese National Knowledge Infrastructure (CNKI), and give a prospecting eventually.

Key words peer reviewer; small-world-network; network group of database; dynamic update

First-author's address Acta Acta Physica Sinica Editorial Office, Institute of Physics, Chinese Academy of Sciences, 100190, Beijing, China

在一般的稿件流程中, 同行评议表现为来自特定领域的若干权威专家的对某稿件的审稿意见呈报给主编并由其作出最终的裁定。权威专家的评审意见很大程度上决定着编辑部是否采用该稿件。作为在知识领域的社会分工的产物, 同行评议的核心思想是权威认证。它深刻地塑造着绝大多数科技期刊, 也深远地影响着期刊业的整体发展^[1]。

正因为同行评议在期刊学术质量控制中的重要作用, 无论是著名科技期刊如《Nature》《Science》等, 还是国际大型期刊出版集团如汤森路透、爱思唯尔以及国内的中国知网(CNKI)和中国科学出版集团等, 都非常重视进行同行评议的必要条件——审稿专家数据库的建设。汤森路透利用 SCI 数据库, 开发出 ScholarOne 采编系统, 使用其内置的审稿专家库送审 1 次就需 20 ~ 30 美元。CNKI 基于其学术文献总库, 构建了大型学者

数据库, 同时将该库深度整合于 CNKI 腾云期刊协同采编系统之中, 实现了基于稿件内容的专家智能推荐功能, 很好地解决了编辑部获取并选择审稿专家的问题。

同时, 现有数据库都未能深度挖掘出专家之间的关系网络, 未能为编辑部选择审稿专家提供此方面的重要依据; 所以, 本文提出一种基于小世界网络 (small world network) 的专家数据库及其关系网络构建方法, 同时基于 CNKI 学术文献总库验证了构建此专家数据库及其关系网络的可行性, 并提出了应用专家关系网络来提升编辑部审稿专家选择质量的具体措施, 为编辑部更为有效地选择专家优化稿件审理过程创造条件。

1 审稿专家获取与选择的主要手段

编辑部遴选审稿专家的策略有多种^[2-5]。从打开一篇科技论文稿件到找到审稿专家可分为 2 大步骤, 第 1 步是从稿件提取出信息。能够提取的与审稿有关的稿件信息包括: 1) 作者 (和合作者) 的姓名和地址; 2) 题名、摘要及能提供研究领域和研究方向的关键词; 3) 分类码, 如《物理学报》使用的 PACS 号码; 4) 正文中的研究内容尤其是创新和发现; 5) 参考文献; 6) 课题得到的基金支持信息。第 2 步是从数据库找出与本稿件的关联, 并从中比较选择合适的审稿专家。数据库的来源可归结为: 1) 编委会名单; 2) 来稿作者数据库; 3) 各种正式出版的纸质专家名册, 如《中国科学技术论文评审专家名典》、《中国高校自然科学学报审稿专家名录》、《博士招生简章》等; 4) 各种基金的获得者名单、入选的学科带头人等; 5) 国内的期刊数据库, 如 CNKI、万方、维普; 6) 国际期刊和出版集团如汤森路透、爱思唯尔以及英国和美国各个学科学会所属系列期刊的数据库, 尤其以 SCI 数据库最为知名和权威, 而且已经整合进入期刊采编平台 ScholarOne; 7) CNKI 依托其海量文献数据库整理的学者数据库, 并且已经集成于 CNKI 腾云期刊协同采编系统中; 8) 期刊软件如马格泰克、勤云和西安三才等利用互联网搜集到的作者信息; 9) 中国教育和科研网以及入选“211 工程”的高校各系网站主页; 10) 著名搜索网站如谷歌、百度、互动百科等; 11) 预印本文库 (ArXiv.org) 中的作者群。

* 国家自然科学基金重点学术期刊专项基金资助 (11024807)

在目前普遍使用的采编系统中,已部分实现第1步到第2步的转换,已基本实现选择审稿专家的自动化,但在专家之间关系信息的挖掘与利用方面还较为欠缺。早在2003年,黎贞崇等^[6]就提出在期刊网站上建立审稿网的想法,但各大期刊网站却一直没能建立起一个普遍使用的审稿专家网。其原因可能是建设大型的审稿专家数据库并且实现数据库中信息的精准标引需要耗费大量的人力与物力,同时也需要多方协助。在如何从海量作者中抽提出合适的审稿专家及其可靠信息上,也缺乏仅依靠计算机即可实现的操作策略。

2 小世界网络和审稿专家数据库

20世纪60年代 Milgram^[7]和 Travers等^[8]发现,在现实社会关系网络中,某人到达他人的平均距离是6,进而提出小世界网络的概念。继开创性的研究工作^[9]之后,有关研究层出不穷,诸如“非典”^[10-11]等传染病传播网络、电影演员的合作网络、万维互联网、国家供电网^[12]、学术研究创新协作网^[13]、基因调控网络^[14]等。

小世界网络的出现,为编辑部依托海量学者数据库建设自身合理规模的审稿专家队伍并构建专家之间的合作与引用关系,提供了坚实的理论基础。小世界网络模型最典型的特征是节点之间的平均距离随远程连接的个数而呈指数下降^[11],这就保证了通过关联构建的审稿专家数据库能够快速找到合适的审稿专家。研究^[15]还发现:大多数小世界网络具有无标度的特征,即在一定范围内其网络结构的标度不因网络规模的大小而变化;期刊的合作者网络、文章引用网络都是典型的无标度小世界网络。现实世界的社会网络,规模太小和太大都会失去无标度特征。这条规律对某一期刊根据自身的定位保持一个适度规模的专家数据库提供了理论依据。稿件送审实践表明,专家数据库太大或太小都不是最优选择:规模太大,势必存在不适当的审稿专家,导致稿件质量把关不严;规模太小,可供选择的专家太少,会导致送审困难。如果只选最好的专家,还会遇到某些审稿专家会因为期刊定位和稿件水平而拒绝审稿。所以,一个规模适度且可变调控的审稿专家库是理想选择。而无标度特征的存在范围,就为这种弹性规模保持原有的审稿专家库的结构特征提供了指针。弹性规模对于数据库自身的新陈代谢及编辑部进行选题上的热点研究追踪转移,都提供了方便。

为进一步阐述小世界网络模型在审稿专家库建设中的作用,我们针对CNKI中全文文献的被引情况建立起的网络进行示例说明。表1给出了截至2009年5月CNKI学术期刊文献引文库中429万2452位作

者发表的1122万9951篇文献的统计结果,其中 N_C 表示某作者所有文章的被引频次, N_A 表示等于或大于该被引频次的作者总数。根据文献^[15]的研究,文献引用网络存在结构无标度性,即引用网络规模的变化并不改变其结构的显著特征。这意味着可据此来弹性地控制审稿专家队伍的规模以适应编辑部实际送审的需求。 $N_C > 500$ 的作者不到2000人,显示这些精英个个都是凤毛麟角,是期刊编委的最佳人选;但由于不同学科间文献引用情况存在较大差异,当然不能用本表中的 N_C 作为遴选专家的标准数。编辑部必须针对本刊的定位和本领域的文献重新统计,才能确定标准,从而得到一个数量合理的审稿队伍。

表1 截至2009年5月CNKI学术期刊文献引文库中429万2452位作者的1122万9951篇文献的被引频次 N_C 和作者总数 N_A 统计

N_C	N_A	N_C	N_A	N_C	N_A
1	2 934 420	30	173 952	500	1 787
2	2 042 950	40	119 743	600	1 213
3	1 581 350	50	88 435	700	842
4	1 289 900	60	68 457	800	609
5	1 086 670	70	54 619	900	463
6	936 261	0	44 878	1 000	352
7	819 929	90	37 635	2 000	47
8	727 860	100	31 891	3 000	11
9	652 629	200	10 325	4 000	2
10	590 225	300	5 005	5 000	2
20	283 323	400	2 891	6 000	1

从CNKI的全文数据库中提取出公认的专家,除了结合前文提到的社会认可和评选之外,还有一条与之并行不悖的道路,就是依照构建小世界网络的方法来提取出上述关联建设审稿专家关系网络。在以作者为节点的小世界网络模型里,不仅要看该节点的各个方面的连接数,如合作者数、文章被引数、文章引用数,还可以看到这些合作和引用的结构树,这种结构树还可以按照关键词和分类码以及创新点甚至被引用时的正文评价来构成,从而反映出作者研究的兴趣变化及与研究方向最接近的小同行的评价。如果对这些结构加以适当的数量度量和甄别,就能选出不同级别和性质的作者进入科研精英库来作为审稿专家库的备选。相对现今高校、科研院所对很多学科和研究方向实行SCI影响因子“一刀切”的评价体系,这种结构树更能体现一个作者对该研究方向真实的科研贡献,从而找到真正的审稿专家。相对很多编辑部单纯靠分类码或关键词来派送稿件去审稿,依靠这种结构树的推荐,将从审稿专家库中作出更好的挑选,进而能将稿件更准确地送到小同行的专家手里。在此理论的指导下,CNKI已经按照学科分类给出了一个总数达10万人的

审稿专家库,推荐给各个编辑部使用。

在有关小世界网络模型的研究中,还发现了一些有趣的节点:1)类似集线器(hub-like)的节点,其基本特征就是该节点与很多节点有联系;2)孤立集团(isolated cluster)的核心节点,该集团只通过核心节点与外界发生联系,而且核心节点拥有一大群孤立节点。显然,类似集线器节点对应于现实中的有广泛合作者的研究者,他们的作品也会有更多人关注和引用。这类专家一般会同时被若干单位任命聘用,交际广泛,活跃于不同领域的科研圈,最适合承当编委或者重点约稿人的角色,以便更好地由其推荐审稿专家或文章;但其中一些功成名就的学者由于繁忙和年龄反而并不一定是最佳审稿人的人选^[16],而孤立集团的核心节点一般对应博士生导师,是最常见的审稿专家。

把一篇稿件的诸多可提取信息放进已有的发表文献库中,不难发现如下关联:以参考文献自身为节点构建引用(和被引用)网络,以参考文献的作者为节点可自成网络,以被引用处的关键词为节点构成更精确的研究方向网络,以稿件的合作者或合作单位为节点可分别构建科研协作网络,以稿件的关键词或创新点为节点可组成研究方向网络,具有以相同或相近的研究分类代码为节点构建文献网络和作者网络。正是这些关乎审稿专家信息的诸多关联使得构建审稿专家数据库的网络群成为可能。

在稿件送审的实践中,有一些情形还需要特别回避。比如:同一稿件的合作者;与作者有可能利益攸关的审稿专家,如有相同导师的师兄师弟,或曾经的合作者,或出自学术上的亲密派别,甚至敌对派别等极端情形^[17]。第1种情形是应严格杜绝的,但事实上由于编辑部送审的人工操作失误却未必不会发生,由于师出同门,导致研究方向可能相当接近;第2种情形在精确送审到小同行手里时并不少见。在软件设计时加以考虑,只需从审稿专家库剔除合作者结构树的一级和二级链接,就可以很好地规避这2种情形。亲密派别之间形成所谓的学术双子孤岛现象,使得审稿专家的相互“放水”成为可能,而敌对派别之间的送审,又会使得优秀稿件拖延甚至冤死于无形。这些都会造成录取稿件的质量下降或优秀稿件的发表延误或流失。处理这些情形确实比较棘手。克服这种弊端,双盲送审是解决问题的法宝;但从根本上还是要拥有可弹性扩展的数据库来选择小同行。在小世界网络化的审稿专家库的网络群中,只需把结构树延伸到下一层,在更广的范围内就不难找到合适的审稿人选。

从长期来看,期刊保持一个常新的审稿专家库极其重要。审稿专家库的新陈代谢一直都是每个编辑部

头疼的大问题。有了基于动态管理的数据库群就可以实现自动更新:只要有新的文章发表,就会产生新的审稿专家库;一些慢慢退出科研第一线的审稿专家就自然而然地通过这种时时更新逐渐淡出审稿专家库,从而保证活跃在最前沿的作者始终囊括在审稿专家库中。

在软件实现方面,Java 技术提供的动态勾画网络的功能会改变很多编辑部的下拉列表式的呆板的选择,从而替代为一种可视化、层叠化、机动灵活的网络化的选择,送审稿件将变成一种愉悦的历程,甚至可以设想,将来可以实现通过人工预审的稿件由计算机全自动送审,最大程度地减少人工操作。

3 结论和展望

基于 CNKI 学术文献总库的资源和小世界网络模型理论,提取全文数据库的各种不同特质的关联构建审稿专家数据库的网络群,建设一个大型的审稿专家数据库系统,从而实现稿件的快速、准确送审。值得期待的是,该系统不但功能强大而且始终不断更新。不久的将来,CNKI 将按照学科分类给出一个总数超过 10 万人、能充分反映学者间关系的审稿专家库,该库随着期刊数据库的不断扩充而时时更新。这将会给编辑部提升稿件质量把关能力带来有益的帮助,并反映和促进中国科技水平的持续提高。

4 参考文献

- [1] 曾咏. 完善科学基金项目同行评议体系的探讨[J]. 研究与发展管理,2007,19(2):124-128
- [2] 任胜利. 科技论文的同行评议[J]. 中国有色金属学报,2004,14(11):封二
- [3] 傅佑丽. 从稿件本身入手准确遴选审稿人[J]. 编辑学报,2009,21(4):338-339
- [4] 黄劲松,彭超群,杨兵. 审稿专家的选择与管理. 编辑学报,2003,15(1):55-56
- [5] 杜秀杰,管咏梅. 高校学报网络审稿资源共享的探索[J]. 编辑学报,2006,18(增刊):20-21
- [6] 黎贞崇,唐莲英. 在期刊网站上建立审稿网的设想[J]. 编辑学报,2003,15(6):453-454
- [7] Milgram S. the small world problem[J]. Psychology Today, 1967,2:60-67
- [8] Travers J S, Milgram S. An experimental study of the small world problem[J]. Sociometry,1969,32(4):425-443
- [9] Watts D J, Strogatz S H. Collective dynamics of "small-world" networks[J]. Nature,1998,393:440-442
- [10] Shi P L, Small M. Modelling of SARS for Hong Kong[EB/OL]. [2011-06-01]. http://arxiv.org/PS_cache/q-bio/pdf/0312/0312016v1.pdf