



基于论文作者署名位置的广义 “陪护人效应”检测

柳娟¹, 许爽¹, 田文灿², 王贤文², 许小可^{1*}

(1. 大连民族大学信息与通信工程学院 辽宁 大连 116600; 2. 大连理工大学科学与科技管理研究所暨 WISE 实验室 辽宁 大连 116024)

【摘要】 学术论文中作者的署名顺序为研究作者的学术贡献提供了重要信息, 根据作者署名位置确定文章的重要作者并对论文发表过程进行分析是科学学领域一项有意义的研究课题, 如近年来提出的“陪护人效应”。该文分析了不同位置作者之间的差异性, 将不同署名位置的作者(即第一、二、三作者、最后作者、通讯作者、倒数第二作者)分别定义为论文重要作者, 对广义的陪护人效应进行了系统性检测和变化趋势研究。结果表明不同署名位置的作者都具有陪护人效应, 通过计算不同位置作者与通讯作者之间的相似系数对陪护人效应的差异性进行了解释, 如综合性期刊的最后作者和通讯作者重合性较大, 医学类期刊则是第一作者和通讯作者的重合性较大。该文研究有助于理解学术文章中作者的角色及贡献, 进一步探索基于作者署名位置的科学学理论与方法。

关键词 陪护人效应; 相似系数; 重要作者; 科学学; 作者署名位置
中图分类号 TP391 **文献标志码** A **doi**:10.12178/1001-0548.2020350

Detection of Generalized “Chaperone Effect” Based on Author's Signature Position

LIU Juan¹, XU Shuang¹, TIAN Wen-can², WANG Xian-wen², and XU Xiao-ke^{1*}

(1. College of Information and Communication Engineering, Dalian Minzu University Dalian Liaoning 116600;

2. WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology Dalian Liaoning 116024)

Abstract The sequence of authorship in academic papers provides important information for the study of their academic contributions. It is a significant research topic in the field of science of science to determine the important author of the article according to the author's signature positions, such as the chaperone effect proposed in recent years. The paper analyzes the differences among authors in different positions, defines the first author, second author, third author, last author, corresponding author and penultimate author in different signature positions as the important authors of the article to detect the chaperone effect systematically. The results show the generalized chaperone effect. At the same time, the paper analyzes the change trend and the differences of the chaperone effect of authors in different positions. We explain the differences by calculating the similarity coefficient among authors in different positions and corresponding authors. When there is a high coincidence among authors in different signature positions and corresponding authors, such as the coincidence between the last authors and corresponding authors in comprehensive journals is higher, the coincidence between the first authors and corresponding authors in comprehensive journals is higher. This study can help us to understand the role and the contribution of authors in academic articles, and further develop new scientific theories based on author's signature positions.

Key words chaperone effect; imilarity coefficient; important authors; science of science; signature position

近年来, 随着科学研究的迅速发展以及数据分析技术的广泛应用, 基于数据驱动的科学学家合作信息挖掘成为科学学领域的重要研究方向^[1-2]。科学家

合作是科研活动组织与科学信息传播的基础, 可推动科研工作的交流与创新, 对知识的创造和传播具有重要意义, 因此受到科学学者的广泛关注^[3-5]。学

收稿日期: 2020-09-21; 修回日期: 2020-12-25

基金项目: 国家自然科学基金(61773091, 61603073); 辽宁省“兴辽英才”计划(XLYC1807106); 辽宁省高等学校创新人才支持计划(LR2016070)

作者简介: 柳娟(1995-), 女, 主要从事社交网络分析和链路预测方面的研究。

通信作者: 许小可, E-mail: xuxiaoke@foxmail.com

术论文是科研成果的重要展示形式之一, 科学家合作最直接的表现形式就是合作发表论文。关于学术论文的发表, 大家经常会提出这样一个问题: “如果以前从未在著名期刊上发表过文章, 那么在该期刊上发表文章时是否需要已经在该期刊上发表过论文的知名作者提供帮助呢?” 针对此问题, 近期 PNAS 中的一项研究将最后作者作为文章的最重要作者, 研究了在同一期刊中没有经验的科学家如何在期刊上发表论文。研究表明如果科学家在顶级期刊上以普通作者的身份参与过论文的发表, 由于在论文发表过程中有重要作者的陪护, 那么该科学家今后在该期刊上以重要作者身份发表论文的可能性就越大, 称为“陪护人效应”^[6]。该效应强调了科学家在职业生涯初期的论文发表经验对其后来发表高水平论文的重要性。相反, 如果科学家从未以任何身份在某个期刊上发表过论文, 那么这位科学家以重要作者的身份在该期刊上发表论文的可能性较低。

陪护人原本指陪同年轻人出席社交场合的监护人, 在科学家合作中是指曾经在著名期刊上发表过论文的导师, 他们在科研工作中发挥重要的指导作用, 可以为年轻科学家发表高水平论文带来优势。即科学家发表论文往往是有一个学术“陪护”过程, 第一次在高水平期刊上发表论文的科学家往往都需要“大牛”导师的“陪护”, 而这些“大牛”导师可能已经在这些期刊上发表过多篇文章, 对相关研究和期刊审稿等各个过程比较熟悉。总的来说, PNAS 中“陪护人效应”的主要研究对象是文章作者署名中最重要的作者, 该文中指的是排名最后的作者, 他们对第一次在顶级期刊上发表文章的“配角”作者有重要的指导作用。因此, 定义文章的重要作者是对“陪护人效应”进行检测的前提条件。

在实际的科学研究过程中, 一些影响力较大的科研成果一般是由团队共同完成的^[7-10], 因此一些权威的学术文章通常也是由多个作者共同撰写的。与单一作者的论文相比, 多个作者合著论文的比例一直在增长, 文献 [11] 最早关注到多作者合作发文增多的现象, 并提出成果中应该描述团队中参与者各自承担的工作和他们之间的关系, 这样有助于明确作者角色及贡献。由于一些重要作者往往在多作者合著论文的撰写和发表过程中起着关键作用, 因此识别该类作者具有重要意义^[12-13]。

在多个作者合著的文章中, 论文作者的署名顺序为作者在项目中所作的贡献提供重要信息, 而占据不同署名位置的作者扮演着不同的角色^[14-16]。如

在生物和物理科学领域中, 论文的第一作者往往是该项目的主要执行者; 最后作者一般是项目的指导者, 负责指导项目的实施、构建论文的写作框架以及与期刊编辑联系等工作; 而处于第一作者和最后作者之间的中间作者通常负责一些数据支持或数据分析等工作^[17-18]。同时, 在科技期刊中标注通讯作者已经变得越来越普遍, 如《材料研究学报》、《科技通报(英文版)》和《力学学报》等, 通常认为文章的通讯作者是课题的总负责人, 负责指导课题及论文投稿等方面的工作, 是合著作者中较权威的作者^[19-21]。由此可见, 学术论文中的重要作者并不相同, 其可能是文章的第一作者、最后作者或通讯作者。根据大多数期刊都要求标记通讯作者的情况, 本文将通讯作者当作文章最重要作者, 为研究不同署名位置作者与通讯作者的差异性以及对差异性进行解释提供依据。

本文研究了 Web of Science 上 6 本著名期刊在 1984–2019 年间发表的近 10 万篇论文, 涉及自然科学和医学科学等领域。首先, 探究了不同署名位置作者与通讯作者之间是否存在差异性。然后将处于不同署名位置的第一作者、最后作者、通讯作者、倒数第二作者、第二作者和第三作者假设为文章的重要作者 (principal investigator, PI), 并将 PI 分成 3 类: 新 PI、有经验 PI 和资深 PI, 通过对比 6 种位置作者的 3 类 PI 比例随时间变化的趋势来对“陪护人效应”进行检测。最后, 计算了不同署名位置作者与通讯作者之间的相似性系数, 对他们与通讯作者之间的差异性进行解释。本研究将有助于理解不同署名位置作者的角色和贡献的差异性, 也将进一步加深对学术论文发表过程的认识。

1 数据说明

1.1 数据来源

本文实验利用了 Web of Science 中 6 本著名期刊数据, 包括《科学》、《自然》、《新英格兰医学杂志》、《柳叶刀》、《美国医学会杂志》和《英国医学期刊》。《自然》和《科学》属于世界最权威的综合类学术期刊, 涵盖科学研究的各个学科领域, 如生命科学、自然科学和物理化学等领域。其余 4 本期刊是顶级的医学领域期刊。6 本期刊涉及不同学科领域, 可使实验结果具有普适性。其中, 《自然》包含了 1998–2018 年的全部数据, 《科学》和其他 4 本医学期刊则包含了 1984–2019 年的全部数据。对于每一篇论文, 数据集包含了发

表时间、文章类型、文章作者及署名顺序、作者所属机构和文章通讯作者等字段信息。在这些期刊中设有各种不同的栏目,考虑到科学家合作主要是在研究性学术论文中进行的,因此仅考虑 Article 和 Review 这 2 种文献类型。由于实验数据不存在共同第一作者的情况,并且共同通讯作者的文章占比较少,在《自然》和《科学》中共同通讯作者的文章占比 0.046,而 4 本医学期刊中共同通讯作者的文章仅占比 0.003,综上本文将不考虑共同第一作者和共同通讯作者的情况。

1.2 数据预处理

由于原始数据存在冗余信息较多、数据格式不统一和姓名歧义等问题,因此在研究之前需要对原始数据进行数据预处理,提取研究所需的字段信息,包括发表时间、作者及通讯作者等信息。数据预处理包括作者姓名消歧^[22-23]和期刊合并两个过程。

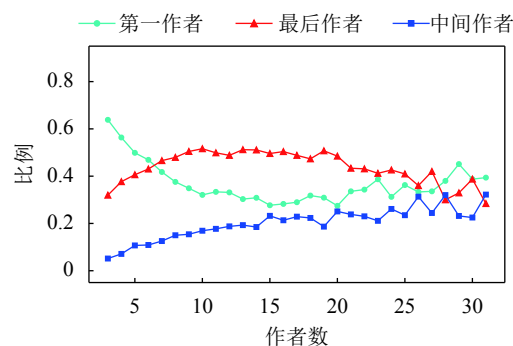
1) 姓名消歧。由于数据中作者的姓名由姓氏全称加上名字首字母缩写构成,这样就会存在不同作者的姓名相同的情况,于是本文对姓名进行了消歧处理,判断两篇文章中同名作者是否为同一人。消歧的原则为:当两篇文章存在姓名相同的作者时,如果在两篇文章中该作者存在相同的合著者,那么为同一作者,否则为不同作者。

2) 期刊合并。在发表高影响力科学成果时,经验往往起着至关重要的作用,这一点在顶级期刊论文上表现尤为明显。如果一位科学家在《自然》上发表过文章,那么在同级别期刊《科学》上发表文章的可能性就很大。同一类别的顶级期刊往往有多本,如《自然》和《科学》都属于同级别综合类期刊,而《新英格兰医学杂志》、《柳叶刀》、《美国医学会杂志》和《英国医学期刊》也是同级别医学类期刊,这样作者投稿这两大类期刊时不确定具体会投稿哪个期刊。即可认为作者向《自然》投稿与向《科学》投稿的可能性基本相同。如果只在每类期刊中选择一种期刊单独进行实验,那么数据量偏少而且无法充分考虑作者的论文发表情况。如果将同类期刊的多本期刊合并,那么在数据量增加的同时也会使实验结果更加稳定。因此本文在进行研究时,将《自然》和《科学》合并成一类期刊,将《新英格兰医学杂志》、《柳叶刀》、《美国医学会杂志》和《英国医学期刊》4 本顶级医学期刊合并成一类,这样就产生两组文章数分别约为 5 万的论文数据集,利用这两大类期刊的论文数据检测“陪护人效应”。

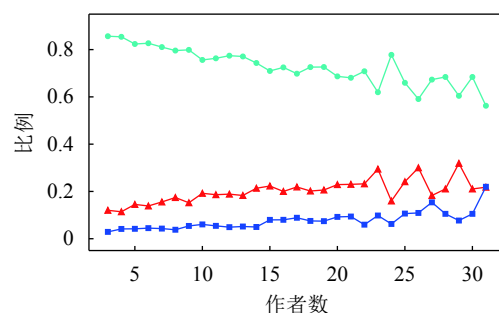
2 基于不同署名位置的“陪护人效应”检测

2.1 不同署名位置作者担任通讯作者的差异性

研究中将 1 篇论文的作者划分成 3 类:第一作者、最后作者和中间作者。中间作者是指该篇文章署名中位于第一作者和最后作者之间的作者,那么很明显两个作者的文章不存在中间作者。由于在单一作者的文章中,第一作者也是最后作者,并且不存在中间作者,所以 3 类作者之间明显不存在交集。综合以上因素,此处仅考虑作者数大于 2 的文章。根据文章的作者数分别统计了第一作者是通讯作者、最后作者是通讯作者和中间作者是通讯作者的比例,结果如图 1 所示,纵坐标表示 3 类作者是通讯作者的比例,横坐标表示文章的作者数,可以看出《自然》和《科学》与 4 本医学期刊的结果有一定的差异性。



a. 《自然》和《科学》中作者数与通讯作者比例关系曲线



b. 4 本医学期刊中作者数与通讯作者比例关系曲线

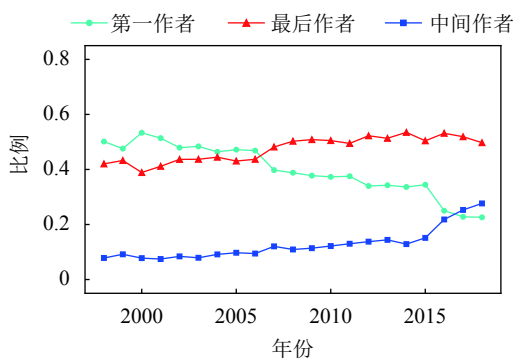
图 1 作者数与通讯作者比例关系曲线

在《自然》和《科学》中,第一作者是通讯作者的比例曲线由最初的比例最大缓慢下降至第二位,曲线呈下降后又慢慢上升的趋势。中间作者是通讯作者的比例曲线则整体呈上升趋势,这可能与中间作者的数量逐渐变多、因此通讯作者分布在中间位置的可能性变大有关。而最后作者是通讯作者的比例曲线由最初的第二位逐渐升至第一位,超过了第一作者是通讯作者的比例,曲线整体呈上升后又慢慢下降的趋势,该曲线比例最大约为 0.56,说

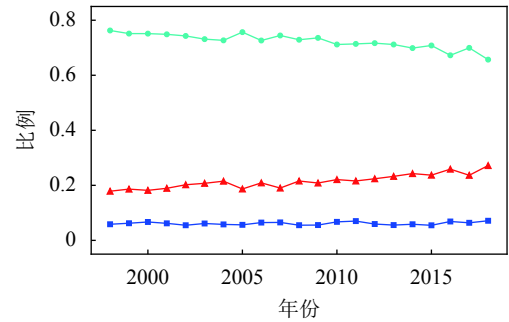
明在《自然》和《科学》中,最后作者和通讯作者之间的关联性较大。当文章的作者数大于25以后,3种情况的比例已基本相同,说明此时通讯作者可能分布在文章作者署名顺序的第一位置、最后位置和中间位置,并且分布在3个位置的可能性基本相近。

与《自然》和《科学》结果不同,在4本医学期刊中,第一作者是通讯作者的比例最大,即使随着作者数的增多,该曲线呈下降趋势,而另外两条曲线呈上升趋势,但是第一作者是通讯作者的比例仍然高于其他两种情况,该曲线比例最低也达到了0.6左右,而最后作者是通讯作者的比例则远小于第一作者是通讯作者的比例,比例最大也未超过0.3,中间作者是通讯作者的比例则最小。说明在医学期刊中,最后作者和中间作者担任通讯作者的可能性较小,而第一作者和通讯作者之间则存在较大的联系,三者论文中的角色并不一致。

通过上述分析得知,随着作者数不同,3类作者是通讯作者的比例会发生变化,并且这两类期刊的结果存在一定的差异性。随着时间推移,3类作者是通讯作者的比例也有变化,如图2所示。在《自然》和《科学》这两本期刊中,随着时间推移,第一作者是通讯作者的比例曲线由最高下降至最低,中间作者是通讯作者的比例曲线由最低缓慢上升,而最后作者是通讯作者的比例曲线由第二位上升至比例最高,最大比例约为0.54,说明最后作者与通讯作者之间的重叠性相比于另外两类作者要高。在4本医学期刊中,最后作者是通讯作者的比例曲线虽然有一定程度的上升,但是仍远小于第一作者是通讯作者的比例,该曲线比例最大仅为0.25左右;中间作者是通讯作者的比例最小,未超过0.1。医学期刊中最后作者和中间作者担任通讯作者的比例较小,而第一作者是通讯作者的比例最大,该比例最小也大于60%,说明第一作者和通讯作者之间的关联性较大。



a. 《自然》和《科学》中时间与通讯作者比例关系曲线



b. 4本医学期刊中时间与通讯作者比例关系曲线

图2 时间与通讯作者比例关系曲线

基于图1和图2中两类期刊的3曲线,可发现无论是随着文章作者数量还是时间变化,两类期刊的结果均存在明显的差异性。在《自然》和《科学》中,最后作者是通讯作者的比例占了一半以上,约为0.55左右。说明在该类综合性期刊中,最后作者与通讯作者之间存在很强的重叠性,而第一作者、中间作者与通讯作者的重叠性不高。但是在4本医学期刊中,第一作者和通讯作者之间存在较强重叠性,而最后作者和中间作者一般较少担任通讯作者,最后作者是通讯作者的比例远小于第一作者是通讯作者的比例,该比例最大也未超过0.3。综合两类期刊来看,第一作者、最后作者是文章通讯作者的概率较大,但是中间作者是通讯作者的可能性较小。

2.2 广义的“陪护人效应”检测

在PNAS的文章中,研究者基于最后作者是文章最重要作者(PI)的假设,发现了“陪护人效应”的存在^[6]。研究中将重要作者(PI)主要分成了3类^[6]: 1) 新PI指的是以前从未以任何作者身份在特定期刊上发表过文章的作者; 2) 有经验PI指的是以前只以普通作者的身份(非最后作者)在特定期刊上发表过文章,但从未当过重要作者(PI)的作者; 3) 资深PI指的是以前以重要作者身份(最后作者)在特定期刊上发表过文章的作者。

根据这3类定义可知,任何文章中的重要作者只能归于这3类中的一类。如2000年A在《自然》上首次以重要作者(最后作者)发表文章,但是2000年以前他从未以任何身份在《自然》上发表过文章,那么A将被归类为2000年在《自然》上的新PI。2006年,A再次以重要作者(最后作者)的身份在《自然》上发表文章,那么由于2000年他已经以重要作者(最后作者)的身份在

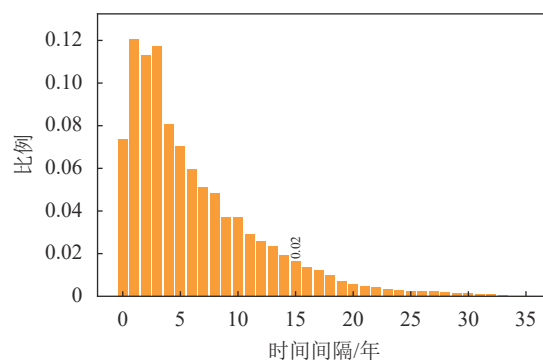
《自然》上发表过文章，于是在2006年他将被归类为《自然》上的资深PI。同时，如果B是这篇2006年文章的另外3个普通合著者之一(非最后作者)，如果2008年B以重要作者(最后作者)身份在《自然》上发表文章，那么B将被归类为2008年在《自然》上的有经验PI。

PNAS文章中最后作者作为论文的PI，分析了以上3类PI随时间的变化趋势，研究发现：如果一名科学家在事业早期没发过《自然》之类的顶级期刊，那么该科学家在成为PI后的论文也更容易登上同样的期刊。这一趋势在多学科期刊最为显著，随后依次是生物学、医学、化学、物理等学科。需要注意的是，不同学术领域中文章的重要作者可能并不相同，可能是文章的第一作者、最后作者，也可能是文章的中间作者。本文已经在两大类期刊中分别绘制了随着作者数和时间的变化，最后作者、第一作者和中间作者担任通讯作者的比例曲线。本文发现在《自然》和《科学》这样综合类期刊中，最后作者和通讯作者存在强重叠性；但是在4本顶级医学期刊中，第一作者和通讯作者之间存在最强重叠性。基于这些结果，本文尝试基于不同署名位置的作者角色对广义的“陪护人效应”进行检测，检测是否处于任何署名位置的作者都具有该效应，而不只有最后作者具有该效应。

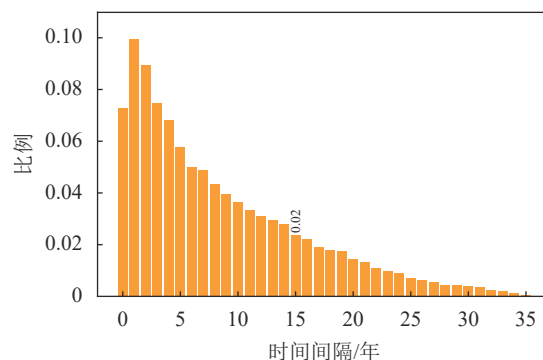
在PNAS论文中也检测了“陪护人效应”随着时间的变化情况，发现《自然》杂志从1990–2012年，由那些以其他身份发过论文的科学家担任末位作者的论文从16%增长到22%，而初次在《自然》杂志中以末位作者发布的论文比例从39%逐步下降到31%。尽管这种时变特性非常有意义，但是作者们在研究中没有对数据进行加时间窗处理。当不加时间窗进行限制时，随着时间的演化在一本期刊上发表的论文越来越多，在该期刊已发表过文章的作者数量将越来越多，而新作者数很自然会越来越少，这样有可能会使实验结果不准确。

为了排除以上因素的影响，本文对实验进行了加时间窗处理，并与不添加时间窗的结果对比，以便准确地探究“陪护人效应”随着时间的变化情况。为了选择合适的时间窗长度，本文首先绘制了论文作者的生存曲线，选择至少发表两篇文章的作者，计算作者的发文时间间隔。如果作者在期刊上

发表的第一篇文章的时间为 t_1 ，发表的第 i 篇文章的时间为 t_i ，最后发表的一篇文章的时间为 t_{end} ，那么该作者的发文时间间隔为 $t_{\text{end}} - t_1$ ，根据此规则统计所有作者发文时间间隔所占的比例。结果如图3所示，结合两类期刊的作者生存曲线可以看出，当发文时间间隔大于15年时所占的比例基本小于0.02，因此可认为作者在两类期刊上的生存时间约为15年，本研究中选择时间窗长度为15年。于是将选择两种实验条件进行检测，分别是加时间窗和不加时间窗，以验证不同条件下的实验结果是否存在差异。



a. 《自然》和《科学》中生存曲线



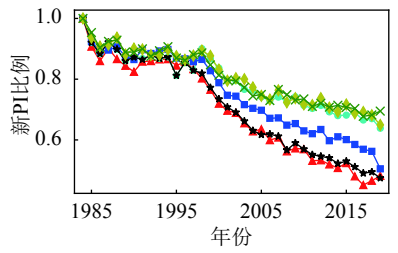
b. 4本医学期刊中生存曲线

图3 生存曲线方面的描述

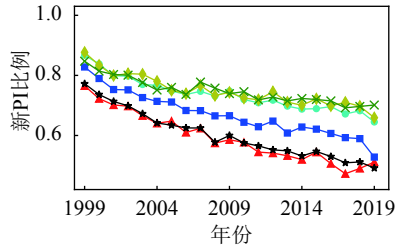
2.3 不同署名位置的“陪护人效应”检测

在“陪护人效应”的检测实验中，本文将假设通讯作者、第一作者、最后作者和中间作者(分别为第二作者、第三作者和倒数第二作者)作为文章的重要作者(PI)，以检测该效应是否存在。由于要计算多种中间作者的3类PI的比例，所以这里选择作者数大于2的文章进行数据分析。根据已经选好的时间窗长度，分别计算期刊中各个作者的3类PI比例随着时间的变化趋势，结果分别如图4和图5所示。其中横坐标表示年份，纵坐标表示3类PI的比例。

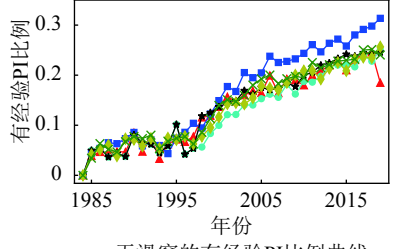
▲ 最后作者 ▲ 通讯作者 ● 第一作者
■ 倒数第二作者 ◆ 第二作者 × 第三作者



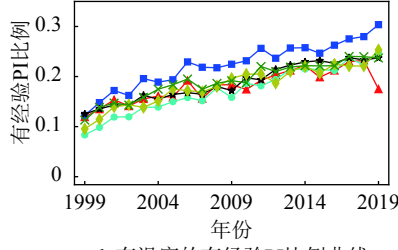
a. 无滑窗的新PI比例曲线



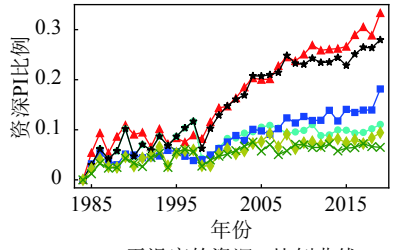
b. 有滑窗的新PI比例曲线



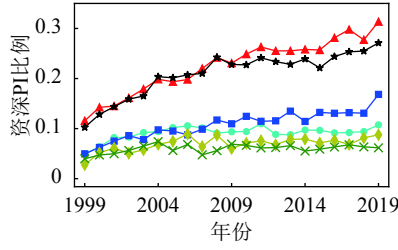
c. 无滑窗的有经验PI比例曲线



d. 有滑窗的有经验PI比例曲线



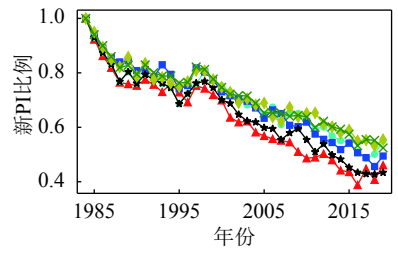
e. 无滑窗的资深PI比例曲线



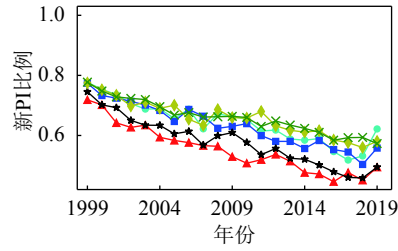
f. 有滑窗的资深PI比例曲线

图4 《自然》和《科学》6种作者的3类PI比例曲线

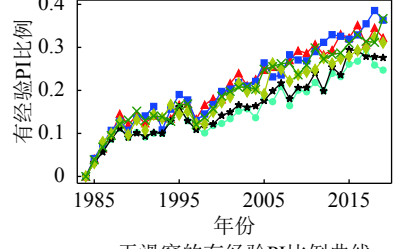
▲ 最后作者 ▲ 通讯作者 ● 第一作者
■ 倒数第二作者 ◆ 第二作者 × 第三作者



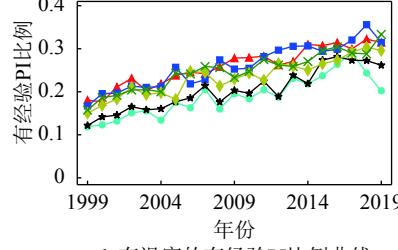
a. 无滑窗的新PI比例曲线



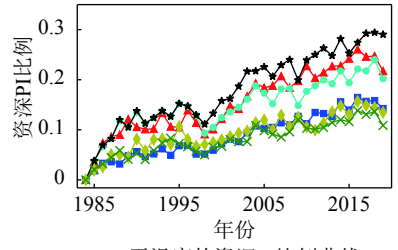
b. 有滑窗的新PI比例曲线



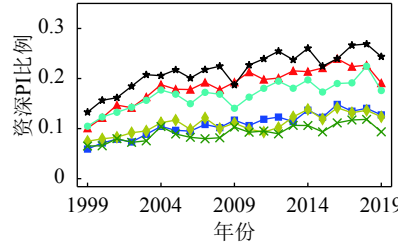
c. 无滑窗的有经验PI比例曲线



d. 有滑窗的有经验PI比例曲线



e. 无滑窗的资深PI比例曲线



f. 有滑窗的资深PI比例曲线

图5 4本医学期刊中6种作者的3类PI比例曲线

从图4和图5中的实验结果看出,在每类期刊中无论加窗与否,各种作者的3类PI比例都随时间变化而变化,并且整体变化趋势相似,说明在不同条件下的实验结果是一致的。但是同条件下的6种作者的“伴随效应”大小存在一定差异。在图4的《自然》和《科学》中,最后作者与通讯作者之间的比例相对来说很接近,说明最后作者和通讯作者之间存在一定关联性,这与前面小节中得出的最后作者与通讯作者之间的重叠性有关。但是第一作者、第二作者、第三作者和倒数第二作者这4种作者与通讯作者之间的比例相差很大,这说明他们与通讯作者之间存在较大差异。对比3种中间作者的实验结果,发现中间作者的“陪护人效应”存在一定差异,其中第二作者和第三作者的结果更加接近,两者都与倒数第二作者存在较大差异。

与综合性期刊《自然》和《科学》的结果不同,在图5的4本医学期刊中,第一作者与通讯作者之间的曲线相对比较接近,尤其是有经验PI比例,即图5c和图5d,说明在4本医学期刊中第一作者与通讯作者之间具有一定的关联性,该结果与前面分析中第一作者与通讯作者之间存在较强重叠性有关。但是最后作者、倒数第二作者、第二作者和第三作者与通讯作者之间的结果相差较大,说明他们与通讯作者在文章中所扮演的角色不同,对文章所做的贡献也不相同。对比3种中间作者可以看出,第二作者、第三作者和倒数第二作者之间的差异性并不大,这3种作者的“陪护人效应”的结果相对比较接近。

从整体结果来看,无论是《自然》和《科学》还是4本医学期刊,3类PI比例中6种角色的PI比例曲线随时间变化的趋势整体上是—样的,即新PI比例随时间变化呈下降趋势,有经验PI随时间变化整体呈上升趋势,资深PI随时间变化整体呈上升趋势。从3类PI的变化趋势来说看,正与PNAS论文中的“陪护人效应”相符合。但是,原始“陪护人效应”的研究发现实际是基于文章最重要作者(即最后作者)的,研究的是向顶级期刊发表论文时,重要作者对一般作者的“陪护”过程。本文发现不仅仅文章的最后作者才能具有该效应,即使是处于不同署名位置的作者,数据分析仍存在与“陪护人效应”相同的变化趋势,说明实际上存在广义的“陪护人效应”,即顶级学者只要参与到论文合作(顶级学者不一定非得是通讯作

者)就对一般作者有一定的“陪护”效应。

虽然不同署名位置的作者都存在“陪护人效应”,但是不同位置作者的“陪护”作用仍然有一定区别。综合3种PI比例变化可以看到,在图4的《自然》和《科学》中,倒数第二作者的陪护效果相对更好。在图5的四本医学期刊中,最后作者和倒数第二作者的陪护效果相对更好。

2.4 不同署名位置作者与通讯作者之间存在差异性的机理解释

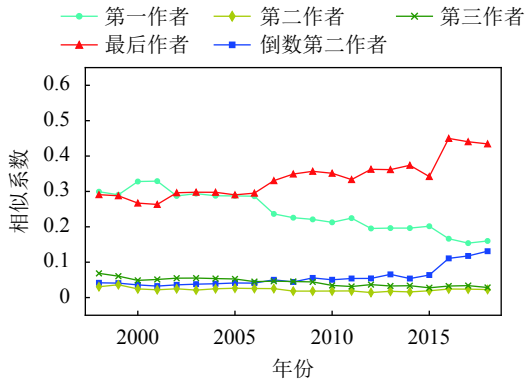
尽管不同署名位置作者之间存在明显的论文分工差异,但是通过基于不同位置作者对“陪护人效应”进行的检测表明,无论任何位置作者都存在“陪护人效应”的作用,但不同署名位置作者的陪护人效应存在一定的区别。

本节根据不同署名位置作者和通讯作者之间相似性的差别对这一区别进行解释,说明影响陪护人效应的内在原因。本文分别计算最后作者与通讯作者、第一作者与通讯作者、倒数第二作者与通讯作者、第二作者与通讯作者和第三作者与通讯作者之间的杰拉德相似系数,相似系数可定义为:

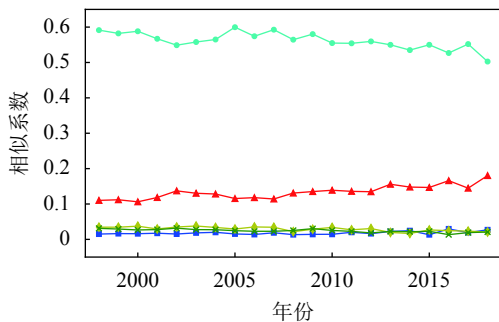
$$O_{ij} = \frac{|U_i \cap U_j|}{|U_i \cup U_j|} \quad (1)$$

式中, U_i 表示某一年中通讯作者的作者群; U_j 表示某一年中最后作者、第一作者、倒数第二作者、第二作者或第三作者的作者群;相似系数则为两个作者群交集与并集的作者数量比值。如两类期刊中每一年都会存在最后作者和通讯作者的作者群,若这两个作者群的相似性系数较大,那么说明大多数最后作者也是通讯作者。最后作者群与通讯作者群越相似,那么两者之间存在强关联性;若两个作者群的相似性系数较小,则说明最后作者和通讯作者之间的差异性较大。

如图6所示,横坐标表示论文的发表时间,纵坐标表示作者之间的杰拉德相似性系数。在《自然》和《科学》中,最后作者与通讯作者的相似系数随着时间呈上升趋势,相似性最大约为0.5左右。第一作者与通讯作者之间的相似性呈下降趋势;而倒数第二作者、第二作者和第三作者与通讯作者之间的相似性最小。这些结果说明在综合类期刊中,最后作者与通讯作者之间的相似性是最大的,这可以解释前述研究中本文发现综合类期刊中最后作者与通讯作者之间的强重叠性。



a. 《自然》和《科学》中通讯作者与另外5种作者之间的相似系数



b. 4本医学期刊中通讯作者与另外5种作者之间的相似系数

图6 通讯作者与另外5种作者之间的相似系数

与综合类期刊不同, 4本医学期刊中第一作者与通讯作者之间的相似性系数最大, 占到了50%以上。最后作者与通讯作者的相似系数远小于第一作者和通讯作者之间的相似系数, 仅为0.12左右; 而倒数第二作者、第二作者和第三作者与通讯作者之间的相似性系数最小, 未超过0.05。这说明在4本医学期刊中, 第一作者与通讯作者的相似性最大, 但最后作者、倒数第二作者、第二作者和第三作者与通讯作者之间的相似性很小, 说明他们与通讯作者之间存在较大差异。这也对前面得出的第一作者与通讯作者之间存在强重叠性、而另外4种作者与通讯作者之间存在差异性进行了解释。

3 结束语

本文基于不同位置作者对广义“陪护人效应”进行了检测, 将第一作者、最后作者、通讯作者和3种中间作者(即第二作者、第三作者和倒数第二作者)假设为文章的重要作者(PI), 并将文章的PI分成3类: 新PI、有经验PI和资深PI, 在两大类期刊中对6种作者的3类PI比例随时间的变化进行了分析。从整体变化趋势来看, 在两大类期刊中6种作者的新PI比例均随着时间的变化慢慢下降, 有经验PI比例随着时间上升, 资深PI比例随

随着时间的变化也慢慢上升。也就是说, 在3种PI比例曲线中, 6种位置作者的整体变化趋势是类似的, 都呈现出广义的“陪护人效应”。本文分别计算了第一作者、倒数第二作者、最后作者、第二作者和第三作者与通讯作者之间的相似性系数, 对不同位置的作者间的差异性进行了解释, 相似性系数较大的作者之间存在较强的重叠性, 而相似性系数较小的作者间则存在一定的差异。

本研究有助于深入理解不同署名位置的作者的角色和贡献, 为科学学领域的探索提供新的研究思路, 也可以为理解科学论文的动态发表过程以及年轻科学家选择科研导师带来一些启示。在进行创新性的科研工作和发表高水平的科研论文时, 年轻的科学家选择与资历较深的科学家合作, 能拓宽科研认知, 提高自身的科研水平, “站在巨人的肩膀上”增加在知名学术期刊上的发表机会。同时, 本文研究结果也可为科研管理部门和基金资助机构提供启示, 对科研团队的组建提供更多合理可行的指导和建议。如在团队组成中注重老中青研究者的合理构成, 有利于年轻一代学者更好更快地成长。在未来的研究中, 如何准确区分不同署名位置作者的角色及贡献, 深入理解高水平科研成果的产出过程与陪护人效应的内在机制是值得探讨的研究问题。

参 考 文 献

- [1] ZENG A, SHEN Z S, ZHOU J L, et al. The science of science: From the perspective of complex systems[J]. *Physics Reports*, 2017, 714: 1-73.
- [2] FORTUNATO S, BERGSTROM C T, BORNER K, et al. Science of science[J]. *Science*, 2018, 359(6379): eaao0185.
- [3] 刘岩, 刘亮, 罗天, 等. 基于子图的科学家合作网络家族辨识[J]. *科技管理研究*, 2019, 39(7): 249-255.
LIU Yan, LIU Liang, LUO Tian, et al. Family identification of cooperative network of scientists based on subgraph[J]. *Science and Technology Management Research*, 2019, 39(7): 249-255.
- [4] 刘亮, 罗天, 曹吉鸣. 基于复杂网络多尺度的科研合作模式研究方法[J]. *科研管理*, 2019, 40(1): 191-198.
LIU Liang, LUO Tian, CAO Ji-ming. A study of the multi-scale scientific collaboration patterns based on complex networks[J]. *Science Research Management*, 2019, 40(1): 191-198.
- [5] 李勇军, 刘尊, 于会. 基于最大熵模型的导师-学生关系推测[J]. *物理学报*, 2013, 62(16): 504-511.
LI Yong-jun, LIU Zun, YU Hui. Advisor-advisee relationship identification based on maximum entropy model[J]. *Acta Physica Sinica*, 2013, 62(16): 504-511.
- [6] SEKARA V, DEVILLE P, AHNERT S E, et al. The chaperone effect in scientific publishing[J]. *PNAS*, 2018,

- 115(50): 12603-12607.
- [7] WUCHTY S, JONES B F, UZZI B. The increasing dominance of teams in production of knowledge[J]. *Science*, 2007, 316(5827): 1036-1039.
- [8] JONES B F, WUCHTY S, UZZI B. Multi-university research teams: Shifting impact, geography, and stratification in science[J]. *Science*, 2008, 322: 1259-1262.
- [9] MILOJEVIC S. Principles of scientific research team formation and evolution[J]. *Proceedings of the National Academy of Sciences*, 2014, 111(11): 3984-3989.
- [10] KLUG M, BAGROW J P. Understanding the group dynamics and success of teams[J]. *Royal Society Open Science*, 2016, 3(4): 160007.
- [11] SMITH M. The trend toward multiple authorship in psychology[J]. *American Psychologist*, 1958(13): 569-599.
- [12] SONG B, JIANG G P, SONG Y R, et al. Rapid identifying high-influence nodes in complex networks[J]. *Chin Phys B*, 2015, 24(10): 100101.
- [13] SUN Y, YAO P Y, WAN L J, et al. Ranking important nodes in complex networks by simulated annealing[J]. *Chin Phys B*, 2017, 26(2): 020201.
- [14] LEHMANN S, JACKSON A D, LAUTRUP B E. Measures for measures[J]. *Nature*, 2006, 444(7122): 1003-1004.
- [15] SHEN H W, BARABASI A L. Collective credit allocation in science[J]. *PNAS*, 2014, 111(34): 12325-12330.
- [16] ZHOU J L, ZENG A, FAN Y, et al. Identifying important scholars via directed scientific collaboration networks[J]. *Scientometrics*, 2018, 114: 1327-1343.
- [17] VAN D D, MANOR O, CAREY L B. Publication metrics and success on the academic job market[J]. *Current Biology*, 2014, 24(11): R516-R517.
- [18] AUSTIN J. What it takes[J]. *Science*, 2014, 344(6190): 1422.
- [19] 金伟, 乔桢. 我国科技期刊合著作者署名的演变[J]. 辽宁师范大学学报(自然科学版), 2018, 41(4): 560-564.
JIN Wei, QIAO Zhen. Evolution of the signature of the collaborators of sci-tech journals in China[J]. *Journal of Liaoning Normal University(Natural Science Edition)*, 2018, 41(4): 560-564.
- [20] 孙丽娟. 科技论文作者署名排序与通讯作者[J]. 中国科技期刊研究, 2005, 16(2): 242-244.
SUN Li-juan. Order of authors and corresponding author in scientific papers[J]. *Chinese Journal of Science and Technical Periodicals*, 2005, 16(2): 242-244.
- [21] 王显苏. 科技论文中通讯作者及其相关问题探讨[J]. 科技创新导报, 2010(80): 241-243.
WANG Xian-su. Discussion on corresponding authors and related problems in scientific papers[J]. *Science and Technology Innovation Herald*, 2010(80): 241-243.
- [22] 范午攸. 一种针对已知作者的姓名消歧方法[J]. 图书馆杂志, 2018, 37(12): 56-63.
FAN Wu-you. Method to remove ambiguity of names of known authors[J]. *Library Journal*, 2018, 37(12): 56-63.
- [23] JINSEOK K. A fast and integrative algorithm for clustering performance evaluation in author name disambiguation[J]. *Scientometrics*, 2019, 120(2): 661-681.

编辑 蒋晓