arXiv:2603.23273v1 [cs.DL] 24 Mar 2026

计算机科学中的系统性性别化引文失衡：来自会议和期刊的证据

Kazuki Nakajima[1], Yuya Sasaki[2], Sohei Tokuno[3], 和 George Fletcher[4]

[1]东京都立大学[2]大阪大学[3]奈良先端科学技术大学院大学[4]埃因霍温理工大学

摘要

科学、技术、工程和数学（STEM）领域的性别失衡依然存在，计算机科学也不例外，研究人员人口统计特征、生产力、认可、招聘和职业晋升等方面均可见这种失衡。鉴于计算机科学迅速发展且影响力遍及全球，解决这一失衡对于扩大参与度和推动创新至关重要。尽管以期刊为导向的学科在引用实践中存在持续的性别失衡，但在以会议为中心的计算机科学文化中是否也存在类似模式，仍不清楚。在此，我们系统地考察了计算机科学会议论文和期刊论文中的引用性别失衡。我们发现，女性担任第一作者或末位作者的论文，获得的引用少于预期，部分原因在于同质性引用倾向（即作者倾向于引用具有特定共同属性的论文）。与期刊相比，这种失衡在会议论文中尤为显著，尤其是发表于顶级发表场所的论文。此外，我们发现，第一作者或末位作者的显著性及其局部合著网络的结构，可能是造成这些失衡的因素。通过探讨以会议为中心的出版实践如何加剧计算机科学中的系统性失衡，本研究提供了相关见解，有助于推动学术界实现更公平的代表性。

关键词：科学学、性别失衡、引用实践、引文网络

1 引言

性别失衡仍是全球学术界普遍存在的问题。尽管女性作者的比例一直在稳步上升，但与本科和研究生阶段的性别比例相比，女性在多个学科和地区的高级作者职位和领导职务中仍代表不足 [44, 31, 85]。此外，科研生产力 [104, 97, 33, 1], 学术奖项获得者 [49, 51], 教师招聘 [57, 77], 资金获取 [6,103], 以及作者获得的引用 [35], 方面的系统性差异依然存在，这些因素共同使女性研究人员在整个职业生涯中处于不利地位。在各个学科中都能观察到这些失衡，反映出学术生态系统中根深蒂固的、更广泛的不平等模式。解决此类系统性失衡，对于实现学术界的公平与可持续发展，以及促进科学多样性和创新至关重要 [88, 83]。

引用是“科学学”研究中许多测量指标的基础，可用于评估出版物、研究人员、机构和国家的影响力，并为关键

学术评价决策提供依据 [22, 54]。尽管引用为了解科学知识的结构与发展提供了宝贵见解，但引文统计往往受到一系列社会、结构和认识论因素的影响，从而造成系统性失衡。例如，不同学科的平均引用次数存在显著差异 [70], ，而许多期刊的大部分引用都来自少数出版场所 [41]。同样，科研活跃的国家获得的引用往往多于其发表论文所涉研究领域所预期的数量 [28]。在性别方面，近期研究发现，女性作者论文持续面临引用不足，部分原因在于同质性、失衡的引用实践，以及偏向男性作者的自引模式 [39, 18, 86]。这种失衡扭曲了科学知识的生产与流动，也损害了引用作为学术价值指标的中立性。

计算机科学在全球范围内尤为突出，其学生和研究人员群体规模庞大且不断扩大，科研产出丰富，投资力度也十分显著。首先，科学、技术、工程和数学（STEM）领域——包括计算机科学——的人才储备规模庞大且持续增长。仅在2020年，中国就授予了360万个新的STEM学位，印度为260万个，美国为82万个；巴西、墨西哥等国家的表现也超过了伊朗和日本 [62]。这一增长往往由计算机科学专业推动；例如，突尼斯和阿尔及利亚等北非国家的STEM毕业率较高，这得益于其扎实的计算与工程课程体系 [7]。计算机科学界也随着入学人数的增加而不断壮大；在欧盟，信息与通信技术行业从业者占研发人员总数的20%至50% [20]。其次，计算机科学出版物在全球科研产出中占有相当大的比重，约为十分之一 [89]；从2015年到2019年，其数量增长了21%，其中人工智能和机器人学等子领域的增长尤为迅速 [75]。第三，全球各国政府优先发展数字创新，并增加对计算和人工智能的资助，使计算机科学研究与教育的公共投资达到前所未有的水平 [75]。例如，美国国家科学基金会在相应财年向其计算机与信息科学及工程部拨款约10亿美元 2024 [90]。这些指标凸显了这一学科非凡的规模、增长和影响力，因此亟须考察其扩张是否伴随着公平代表性。

性别失衡影响计算机科学的诸多方面，包括作者署名、生产力、认可、研究职业生涯和学术职位。估计显示，全球计算机科学领域的女性作者仅占 15–30% [31, 24], ，且进展缓慢 [14]。此外，不同性别失衡程度因子领域而异 [24, 67], ，女性作者还面临额外的不利因素——例如职业生涯时长较短——因此，与男性同行相比，她们发表的出版物和获得的引用更少 [36, 33, 48]。随着职业阶段提升，这种失衡愈发明显；以美国为例，计算机科学领域的终身教职序列教师中，女性仅占 15 % [13, 102]。高级职位中女性代表性不足，可能源于留任率和晋升率存在差异，许多人在获得这些职位之前便离开了学术界 [82, 48]。鉴于该学科迅速发展且具有全球影响力，通过持续、以数据为依据的研究解决性别不平等至关重要且刻不容缓，这有助于确保计算机科学的未来公平、包容，并充分体现其人才储备。

研究实践（例如作者署名、出版场所和引用行为）及其与性别失衡的关联，在不同学科间差异显著。近期关于性别化引文失衡（即与作者性别相关的引用失衡）的研究，聚焦于单一学科的期刊论文，包括天文学 [8], 传播学 [101], 牙医学 [55], 经济学 [17], 环境研究 [61], 国际关系 [50], 医学 [11], 神经科学 [18], 物理学 [86],和植物科学 [66]。然而，计算机科学中的此类失衡程度仍鲜为人知。鉴于计算机科学独特的

出版文化，其他学科的研究发现无法直接推广到计算机科学。与大多数以期刊为主的学科不同，计算机科学家通常在会议上发表研究成果；会议采用固定日程和较短的同行评审周期，以加快成果传播 [98, 21]。此外，会议和期刊依据引文指标排名，排名靠前的出版场所被视为尤具声望 [26, 95, 99, 47]。顶级会议的迅速增长使组织委员会不堪重负，也引发了人们对同行评审严谨性的担忧 [98, 21, 80]。这种截然不同的出版文化凸显了考察以会议为中心的实践如何影响性别化引文模式的必要性。

在本研究中，我们系统地分析了一个独特的计算机科学论文引文网络，其中论文作者的二元性别已被指派。我们将性别化引文模式定义为观察到的引用行为与参照模型所预期行为之间的差异；这些参照模型考虑了网络的关键结构属性。本研究在我们先前的研究基础上进行了重要扩展。先前的研究仅关注计算机科学会议论文之间的引用 [59],；本研究则将期刊出版物纳入数据集，并探讨以下新研究问题：(i) 计算机科学领域的会议论文和期刊论文分别呈现出怎样的引文性别失衡模式？(ii) 会议与期刊中的引文性别失衡有何不同？(iii) 引用论文的哪些特征与性别化引用行为相关？

2 方法

2.1 新数据集的构建

我们构建了一个独特的数据集，涵盖发表于计算机科学会议和期刊的 394,432 篇论文以及 752,742 条引用。通过整合两个开放数据库中的出版数据，我们为每篇论文补充了元数据，包括书目信息、作者的二元性别、作者署名信息、引用、出版场所、研究主题和子领域。我们还利用外部排名数据库确定了会议和期刊的发表场所排名。由于研究性别化引用行为时，不仅需要考虑作者的性别，还须考量可能影响引用模式的各种作者层面和论文层面特征，因此，与我们先前的研究相比，这个经过补充的数据集能够支持对计算机科学领域性别化引用失衡开展更全面的分析 [59]。下文将介绍数据集的构建过程。

2.1.1 DBLP

DBLP计算机科学书目（最初是“数据库系统与逻辑编程”（DataBase systems and Logic Programming）的缩写，后来简称为 DBLP）收录了计算机科学领域的出版物，涵盖会议和期刊 [93, 45]。DBLP 一直是计算机科学领域的主要书目资源，并被广泛用于分析该学科的研究格局（例如，参考文献 [68, 71, 81, 9, 38, 72, 59]）。我们使用了其 2024年7月1日的快照，其中包括 1950 年至 2024 年间在 13,007 个会议和 2,031 种期刊上发表的 7,034,299 篇论文。每篇论文都提供其标题、发表年份、作者 ID、作者署名顺序，以及相关会议或期刊的名称。DBLP 使用专有算法为作者分配唯一 ID，共识别出 3,609,641 名作者。此前的评估显示，DBLP 的作者姓名消歧表现极具竞争力，这可能得益于其结合算法消歧与人工纠错的混合方法 [37]。每个 ID 均由作者全名构成；如果有多位作者同名，则后面会附上一个四位数 [92]。我们依据这项规则，从作者所分配的 ID 中还原每位作者的全名。尽管 DBLP 提供了精确且经过良好消歧的发表场所信息，但它缺乏更广泛的书目

覆盖范围和主题分类。

2.1.2 OpenAlex

OpenAlex 提供数亿条出版记录，涵盖多个学科，并附有丰富的元数据 [69]。我们使用了其 2024 年9月27日的快照，其中包含 1950 年至 2024 年间发表的 155,449,238 篇论文。对于每篇论文 v，我们提取了其标题、出版日期、主要研究主题和子领域、作者姓名及所属机构、作者署名顺序，以及其引用的论文列表 v。每篇论文至少列出一位作者，但有些论文缺少部分作者的所属机构信息。每篇论文都归入 4,516 个研究主题之一，这些主题比其所属的 252 个子领域划分得更细 [63]。尽管 OpenAlex 对出版物和引文数据提供了全面的跨学科覆盖 [16], ，我们发现，许多计算机科学出版物的会议和期刊名称要么缺失，要么未得到恰当的消歧。

2.1.3 OpenAlex 与 DBLP 之间的论文匹配

为弥补这两个数据库各自的局限，我们依据三项标准，将每篇 DBLP 论文与 OpenAlex 中对应的论文进行匹配：(i) 发表年份相同；(ii) 作者姓氏的顺序相符；(iii) 标题差异不超过 25%，该差异通过莱文斯坦距离除以较长标题的长度进行标准化后计算。此前已有研究使用这些标准来对齐多个出版数据集 [33]。莱文斯坦距离的 25% 阈值相对宽松，可涵盖数据库之间常见的差异，例如数学符号编码不同，以及是否包含副标题。我们将作者全名中按空格分隔的最后一个词视为其姓氏 [58]。采用这一方法，我们在 OpenAlex 中为 3,720,575 篇 DBLP 论文找到了唯一匹配项。对于每篇匹配的论文 v，我们使用 OpenAlex 元数据补充 DBLP 记录，包括标题、出版日期、主要研究主题和子领域、作者姓名及所属机构、作者署名顺序，以及其引用的论文列表 v。下文将这 3,720,575 篇补充了信息的论文称为 $\mathcal{D}$。 $\mathcal{D}$ 中的论文之间共有 23,274,023 次引用。

2.1.4 会议与期刊排名

我们使用以下排名体系来呈现计算机科学出版声望的结构。尽管有些资源使用经过整理的顶级发表场所列表（例如 CSRankings [15]），但我们的分析需要采用多层级分类，以考察广泛出版场所中的引用失衡情况。

我们使用2021年CORE会议排名 [91]来确定计算机科学会议的排名。该体系根据各会议的声誉以及一系列指标（例如论文和作者的引文统计、录用率）[1]对会议进行评估。我们重点关注被划分为四个等级——A*、A、B和C——的768个会议名称，其中A* 为最高等级，C为最低等级。

同样，我们使用 2021年 SCImago 期刊排名（SJR） [76] 对计算机科学期刊进行排名。SCImago 基于 Scopus 的数据，涵盖了来自 5,000 多家出版商[2]的 34,100 多种期刊名称。我们将筛选范围限定为 2021 年归入“计算机科学”学科领域的期刊。经过这一筛选过程，共有 1,734 种期刊被划分为第一季度、第二季度、第三季度或第四季度，其中第一季度排名最高，第四季度最低。尽管 SCImago 数据库收录了多个学科的期刊，

[1]https://drive.google.com/file/d/1DQixeK53tlq_jh6IspIHroiwu1pmM6-y/edit（访问于2024年4月）[2] https://www.scimagojr.com/aboutus.php（访问于2024年11月）

先前的分析表明，SJR 与计算机科学期刊的期刊影响因子呈高度相关 [78]。

由于 DBLP 经常缩写会议和期刊名称，许多条目无法与 CORE 和 SCImago 数据中的名称直接匹配。因此，对于 $\mathcal{D}$中出现的每个发表场所，第一作者和第二作者（KN 和 YS）手动指定发表场所排名（A*、A、B、C、Q1、Q2、Q3 或 Q4），或标记为“不适用”。这一过程得出了 60 个 A*级会议、130 个 A级会议、286 个 B级会议和 233 个 C级会议，以及 341 个 Q1分区期刊、301 个 Q2分区期刊、209 个 Q3分区期刊和 128 个 Q4分区期刊。

2.1.5 作者的显著性与合著者

构成 $\mathcal{D}$的论文中共有 2,430,109 个独一无二的作者 ID。对于每位作者 z，我们找出所有将 z 列为作者的论文（下文称为 z的论文），并计算两项特征：（i）显著性，定义为 z的论文收到的引用总数；（ii）合著者，即至少与 z共同署名过一篇论文的作者集合。我们使用整个数据集计算这些指标，并将其视为静态属性，用以体现作者的显著性和合著网络；这与我们采用会议和期刊静态排名的做法一致。

2.1.6 作者的所属国家与性别

我们根据作者出版物中列出的所属机构[33, 58]，为每位作者指定一个所属国家。具体而言，我们统计某位作者 z所属机构所在国家的出现频次；如果某个国家的频次独占最高，则将其指定为该作者的所属国家。如果没有国家的频次独占最高，我们便将该作者从数据集中排除。在数据集中的 2,430,109 位作者中，我们为 2,059,647 位指定了所属国家。

随后，我们采用参考文献 [58], 提出的方法，为每位作者指定二元性别。该方法综合考虑作者的所属国家、首次发表年份和名字。此方法旨在应对基于姓名的性别推断中已知的挑战，尤其是东亚姓名 [58],，但其分类仅限于二元分类（有关其他局限性，请参阅讨论部分）。

为了从作者的全名中提取名字候选项，我们根据其所属国家采用相应的命名惯例。对于所属国家为中国、日本或韩国，且姓名由 k 个以空格分隔的词组成（记作 $w_1, \ldots, w_k$）的作者，我们使用w_1、 w_1w_2、...和 $w_1 \cdots w_{k-1}$作为其名字的 $k-1$ 候选项。例如，“Gil Dong Hong”这一姓名的名字候选项为“Gil”和“Gil Dong”。对于其余作者，我们假设姓名中第一个以空格分隔的词是唯一的名字候选项。

我们将作者的名字候选项及其所属国家提供给 Gender API[3], 。该 API 曾用于此前关于性别失衡的研究 [18, 86]。对于每个输入的名字候选项，API 会返回三个标签之一（“女性”“男性”或“未知”），并同时提供准确率分数和样本量。

我们对准确率分数和样本量均设定阈值，以排除置信度较低的结果。如果某位作者有多个名字候选项，我们会比较女性分类准确率最高的结果与男性分类准确率最高的结果。哪一项分数较高，就将其性别判定为“女性”或“男性”；如果两项分数相同，则不判定性别。我们根据所属国家和首次发表年份设定阈值，以确保在带有真实性别标签的验证集上，分类准确率至少达到 90% [58]。根据所属国家设定不同阈值，旨在解决性别判定过程中可能存在的国家失衡问题，

[3]https://gender-api.com/en/（访问于2024年4月）

因为基于姓名的性别推断的准确性可能取决于姓名的原籍国 [74]。我们采用与参考文献 [58]相同的阈值（参见参考文献 [58]中的补充材料第S3节）。

我们总共为数据集中 2,430,109 位作者中的 1,257,016 位同时标注了所属国家和性别。

2.1.7 论文的所属国家和性别类别

我们参照文献 [58] 为每篇论文标注所属国家和性别类别。具体而言，我们聚焦于 $\mathcal{D}$ 中的 1,399,106 篇论文，这些论文满足以下两项标准之一：(i) 有一位已标注所属国家和性别的作者；或 (ii) 第一作者和末位作者的所属国家相同，且两人均已标注性别。随后，我们将单一作者或第一作者和末位作者的所属国家作为论文的所属国家。我们还根据第一作者和末位作者的性别，将每篇论文分为四种性别类别（“MM”“MW”“WM”“WW”） [18, 86]。具体而言，第一个字母（M 或 W）表示第一作者是男性还是女性，第二个字母则表示末位作者的性别。我们将男性或女性独著论文分别归为 MM 或 WW。我们假设第一作者和末位作者在计算机科学论文的研究与写作中发挥主要作用；这一做法符合许多子领域的惯例，尽管某些子领域采用按字母顺序排列的方式 [79]。

2.1.8 引用

我们聚焦于 $\mathcal{D}$ 中的 672,082 篇论文，每篇论文均被赋予一个所属国家、一个性别类别（MM、MW、WM 或 WW）和一个发表场所排名（A*、A、B、C、Q1、Q2、Q3 或 Q4），并仅考察这些论文之间发生的引用。对于观察到的从论文 u 到论文 v的引用，我们称 u “发出”了该引用，而 v “ 接收 ” 了该引用。若符合以下任一条件，我们便剔除该引用：（i） v 的出版日期比 u 的 $[18, 58]$, 早十年以上；（ii） u 与 v 至少有一位共同作者（自引）；（iii） u的第一作者或末位作者的合著者与 v的作者有重叠；或（iv） u的发表年份早于 1990年（因为 1980 年之前的 WW 论文数量很少）。关于标准（i），我们遵循了参考文献 [18, 58]；这一十年的引用时间窗口反映了当代引用行为，并避免我们的分析跨越作者人口统计特征显著不同的时期。随后，我们剔除既未发出也未接收任何引用的论文，最终得到 394,432 篇论文和 752,742 条引用。表1总结了数据的统计信息，并比较了初始匹配数据集（$\mathcal{D}$）与最终数据集。请注意，初始匹配数据集中未赋予作者性别和发表场所排名 $\mathcal{D}$。关于数据集的更多信息，见补充材料第S1节。

2.2 引文网络的参照模型

参照模型会对每篇论文所作的引用进行随机化，同时保留原始网络的某些结构属性。Dworkin 等人提出了随机抽样模型作为参照模型 [18], , 但该模型仅保留每篇论文所作的引用数量，因此无法充分捕捉其他网络属性对引文中的性别失衡的影响。为弥补这一不足，我们提出了一系列参照模型，它们保留每篇论文所作的引用数量，并额外保留最多两种属性 [58]: 引文同质性 [12, 106] 以及每篇论文收到的引文数量异质性 [19, 106]。表 2 概述了每种模型保留的属性。下文中，我们用 $V = \{v_1, \ldots, v_N\}$ 表示所有论文的集合，用 E 表示网络中的引用集合，用 $M = |E|$ 表示引用数量，其中 N 为论文数量。

T表 1：初始匹配数据集（$\mathcal{D}$）和最终数据集的基本统计数据。

初始匹配数据集	
含义	数量
论文	3,720,575
引用	23,274,023
作者	2,430,109
会议	10,735
期刊	2,004
所属国家	214
研究子领域	243
研究主题	4,351
最终数据集	
含义	数量
论文	394,432
引用	752,742
性别指定作者	285,406
会议	631
期刊	955
所属国家	140
研究子领域	238
研究主题	2,900
MM 论文	298,063 (75.6%)
MW论文	30,970（7.8%）
WM 论文	43,727（11.1%）
WW 论文	21,672（5.5%）
女性作者	54,229 (19.0%)
男性作者	231,177 (81.0%)
A级*排名会议	60
A级会议	124
B级会议	254
C级会议	193
Q1分区期刊	338
Q2分区期刊	294
Q3分区期刊	203
Q4分区期刊	120

表 2：各参照模型中要保留的属性。

模型	要保留的属性
随机抽取	• 每篇论文的引用次数
同质性抽取	• 每篇论文引用的次数 •引用模式中的同质性
偏好抽样	• 每篇论文引用的次数 •引用模式中的同质性 • 每篇论文收到的引文数量异质性

2.2.1 随机抽样模型

我们首先介绍随机抽样模型 [18]。令 $V_{\mathrm{RD}}(i)$ 表示在此模型下论文$v_i \in V$ 可能引用的论文集合。根据我们的引文标准， $V_{\mathrm{RD}}(i)$ 包含以下论文：(i) 发表时间比 v_i 至多早十年；(ii) 与 v_i 没有任何共同作者；(iii) 与 v_i 的第一作者或末位作者的合著者没有任何共同作者。对于每一条引用（v_i，v_j）$\in E$，我们以均匀概率随机抽取 $v_{j'} \in V_{\mathrm{RD}}(i)$，并进行有放回替换，将（$v_i$，$v_j$）替换为（$v_i$，$v_{j'}$）。随机抽样模型保留了原始网络中每篇论文的引用数量。

2.2.2 同质性抽取模型

我们将随机抽样模型扩展为保留引文同质性的模型，并称之为同质性抽取模型。令 S 表示一组可能与引文同质性相关的论文特征（性别类别除外）。本研究关注：(i) 所属国家；(ii) 主要研究主题；以及 (iii) 发表场所排名。请注意， S 并不限于这些属性，且可能因学科而异。

对于每条引用（v_i，v_j）$\in E$，令 $V\mathrm{HD}(i,j,S)$ 表示 $V\mathrm{RD}(i)$（来自随机抽取模型）的一个子集，其中包含与 v_j 在 S 中所有属性上都相同的论文。请注意， $V\mathrm{HD}(i,j,S)$ 包含v_j。在我们的分析中，$V\mathrm{HD}(i,j,S)$ 包含在所属国家、研究主题和发表场所排名方面与v_j 相匹配的论文。我们将每条引用（v_i，v_j）替换为（v_i，$v_{j'}$），并允许重复抽取，其中 $v_{j'}$ 从 $V\mathrm{HD}(i,j,S)$ 中均匀抽取。同质性抽取模型保留每篇论文发出的引用数量，以及所属国家、主题和发表场所排名两两组合之间的引用数量。

2.2.3 优先抽样模型

我们进一步扩展同质性抽样模型，使其近似保留每篇论文收到的引用数量的异质性。我们将此称为优先抽样模型；该模型在尊重 S 和优先连接动力学特征的同时，对每篇论文发出的引用进行随机化。

首先，我们按出版日期升序排列论文，记作 $\{v_{x_1},\ldots,v_{x_N}\}$，其中 v_{x_1} 最早， v_{x_N} 最新。设 $c_{i,l,\mathrm{PD}}$ 表示模型中论文 v_i 从论文 $v_{x_1},\ldots,v_{x_l}$ 获得的引用数量。我们初始化 $c_{i,l,\mathrm{PD}}=0$，对所有 i 和 l 均如此。对于每条引用（v_{x_l}，v_j）$\in E$，我们定义

$$\overline{V}_{\mathrm{PD}}(x_l,j,S)=\left\{\, v_{j'}\in \overline{V}_{\mathrm{HD}}(x_l,j,S)\;\middle|\; \lfloor \ln(c_{j',l-1,\mathrm{PD}}+1)\rfloor = \lfloor \ln(c_{j,l-1,\mathrm{PD}}+1)\rfloor \right\},$$

其中 $\lfloor x \rfloor$ 是向下取整函数。因此， $V_{\mathrm{PD}}(\overline{x}_l, j, S)$ 包含 $V_{\mathrm{HD}}(x_l, j, S)$ 中的论文，这些论文从 $\{v_{x_1}, \ldots, v_{x_{l-1}}\}$ 获得的引用数量与 v_j 大致相同。我们使用自然对数将论文分组到各个分箱中，使其对应于引用数量的对数尺度。这种对数分箱考虑了论文所获引用数量的重尾分布，并避免候选池（即 $V_{\mathrm{PD}}(x_l, j, S)$）变得过于稀疏，同时近似保持被替换论文所获的引用数量。

随后，我们依次随机化 v_{x_l} 在 $l = 2$ 至 N 期间所作的引用。对于每条引用（v_{x_l}， v_j） $\in E$，我们以均匀概率从 V_{PD}（x_l， j， S）中随机选择 $v_{j'}$，并以有放回抽样的方式，将（v_{x_l}， v_j）替换为（v_{x_l}， $v_{j'}$）。请注意， v_{x_1} 不作任何引用，因为它是最早的论文。这一顺序过程使引用次数（即 $c_{j,\ l-1,\ \mathrm{PD}}$）能够动态更新，并近似模拟优先连接动力学：在此过程中，较早积累引用的论文更有可能被后续论文引用。

与同质性抽取模型类似，该优先抽样模型保留了每篇论文所作引用的数量，以及在所属国家、研究主题和发表场所排名方面的引文同质性。此外，它还近似保留了每篇论文所获引用次数的原始分布。优先抽样模型的伪代码见补充材料第 S2 节。三个参照模型的结构属性比较见附录 A。

2.3 刻画引文中的性别失衡

网络分析中的标准做法是根据结构和动态特征，将原始网络与参照模型生成的随机化实例进行比较 [60, 53, 65, 23]。我们通过比较原始网络与参照模型中作者性别相关的引用模式差异，来刻画引文中的性别失衡。考虑两个论文子集，分别记为 V_{from} 和 V_{to}。我们衡量 V_{from} 中的论文对 V_{to} 中的论文在给定性别类别 $g \in \{\mathrm{MM}, \mathrm{MW}, \mathrm{WM}, \mathrm{WW}\}$ [18, 86]下的过度引用或引用不足程度。首先，我们统计原始网络中 V_{from} 中的论文对 V_{to} 中且属于 g 的论文所作的引用数量，记为 $n_{g,\ \mathrm{obs}}$。接着，我们使用参照模型独立生成 100 个随机化网络。我们将 100 个随机化实例中， V_{from} 中的论文对 V_{to} 中且属于 g 的论文所作引用数量的均值和标准差分别记为 $\mu_{g,\ \mathrm{rand}}$ 和 $\sigma_{g,\ \mathrm{rand}}$。我们将 V_{from} 中的论文对 V_{to} 中的论文所作的过度引用/引用不足定义为 $(n_{g,\ \mathrm{obs}} - \mu_{g,\ \mathrm{rand}})/\mu_{g,\ \mathrm{rand}}$ [18, 86]。

为检验过度引用/引用不足的统计显著性，我们根据（$n_{g,\ \mathrm{obs}} - \mu_{g,\ \mathrm{rand}}$）$/\sigma_{g,\ \mathrm{rand}}$计算 Z 分数。若 Z 分数大于 3.09（即 p 值小于 0.001），则认为 V_{from} 中的论文显著过度引用了 V_{to} 中的论文。同样，若 Z 分数小于 -3.09 （即 p 值小于 0.001），则认为 V_{from} 中的论文显著引用不足 V_{to} 中的论文。

2.4 引文中的性别失衡的匹配对分析

为考察哪些类型的论文会导致引文中的性别失衡，我们在控制关键论文层面特征 [18, 86]的同时，比较过度引用/引用不足模式。我们采用匹配对分析（此前有关性别失衡的研究也曾采用此方法 [33, 46], ），以确定被引用论文的哪些特征与其引用行为中更严重的性别失衡相关。

设 $\mathcal{M}$ 和 $\mathcal{M}'$ 是两个互不相交的论文子集。我们按如下方式生成一组匹配对。对于每篇论文 $u \in \mathcal{M}$，我们随机选择（不放回）一篇在以下特征上与其匹配的论文 $v \in \mathcal{M}'$ u ： (i) 发表年份，(ii) 所属国家，(iii) 主要

表3：分析中使用的匹配论文对数量。每一行显示一对论文子集 $\mathcal{M}$ 和 $\mathcal{M}'$，以及在两者之间生成的匹配对数量。

$\mathcal{M}$	$\mathcal{M}'$	匹配对
WW 论文	MM 论文	10,985
MM（会议）	MM（期刊）	49,860
MM（含知名作者）	MM（不含知名作者）	29,290
男性-男性（MA$_{or}$的前半部分）	男性-男性（MA$_{or}$的后半部分）	50,124
女性-女性（会议）	女性-女性（期刊）	1,215
女性-女性（包含知名作者）	女性-女性（不含知名作者）	741
女性-女性（MA$_{or}$的上半部分）	女性-女性（MA$_{or}$的下半部分）	1,166

研究子领域，以及（iv）论文引用的数量。我们重复此过程 100 次，生成 100 组相互独立的数据集，记为 $\{(\mathcal{M},\ \mathcal{M}'_i)\}_{i=1}^{100}$，其中 $\mathcal{M}'_i \subseteq \mathcal{M}'$ 取值为 $i = 1,\ \ldots\ldots,\ 100$。我们计算 $\Delta_{\mathcal{M},\ g}$，即 $\mathcal{M}$ 中的论文对性别类别为$g \in$ {男性-男性，男性-女性、女性-男性、女性-女性}的论文所做的过度引用/引用不足。类似地，对于每个 $i = 1,\ \ldots\ldots,\ 100$，我们计算 $\Delta_{\mathcal{M}'_i,\ g}$，即 $\mathcal{M}'_i$ 中的论文对 g中的论文所做的过度引用/引用不足。令 $\Delta_{\mathcal{M}',\ g}$ 和 $s_{\mathcal{M}',\ g}$ 分别表示这 100 个值的均值和标准差。检验统计量定义为

$$t = \frac{\bar{\Delta}_{\mathcal{M}',g} - \Delta_{\mathcal{M},g}}{s_{\mathcal{M}',g}/\sqrt{100}} \tag{1}$$

¯我们检验零假设：$\Delta_{\mathcal{M},\ g} = \Delta_{\mathcal{M}',\ g}$；与之相对的备择假设认为它们存在差异。如果所得的 p值低于0.001，我们就拒绝零假设，转而支持备择假设。

我们首先使用匹配的MM–WW论文对，检验引用作者的性别是否与引用失衡相关。接下来，我们考察引用论文的另外三项特征：(i) 发表场所类型（会议与期刊）、(ii) 是否有“知名”作者参与，以及 (iii) 局部合著网络。对于特征 (ii)，我们将各性别中作者声望排名前1%的作者定义为知名作者。对于特征 (iii)，我们采用男性作者过度代表度（MA$_{or}$），其定义为论文第一作者或末位作者的男性合著者占比 u减去数据集中男性作者的总体占比 [18]。作者声望和合著者的定义见第2.1.5节。针对每项特征，我们将MM论文划分为两个互不重叠的集合，具体而言，分别为会议发表的MM论文与期刊论文；有知名第一作者或末位作者的MM论文与两者皆无的MM论文；以及MA$_{or}$排名前半与后半的MM论文。随后，我们在这两个集合之间生成匹配对。对于WW论文，我们采用与MM论文相同的流程。表3报告了分析中使用的匹配论文对数量，包括MM–WW比较，以及基于发表场所类型、知名作者参与情况和局部合著网络的组内比较。

3 结果

3.1 量化计算机科学中的性别化引用失衡

我们构建了一个引文网络，其中包括发表在计算机科学会议和期刊上的 394,432 篇论文，以及这些论文之间的 752,742 条引用。我们将每篇论文归入

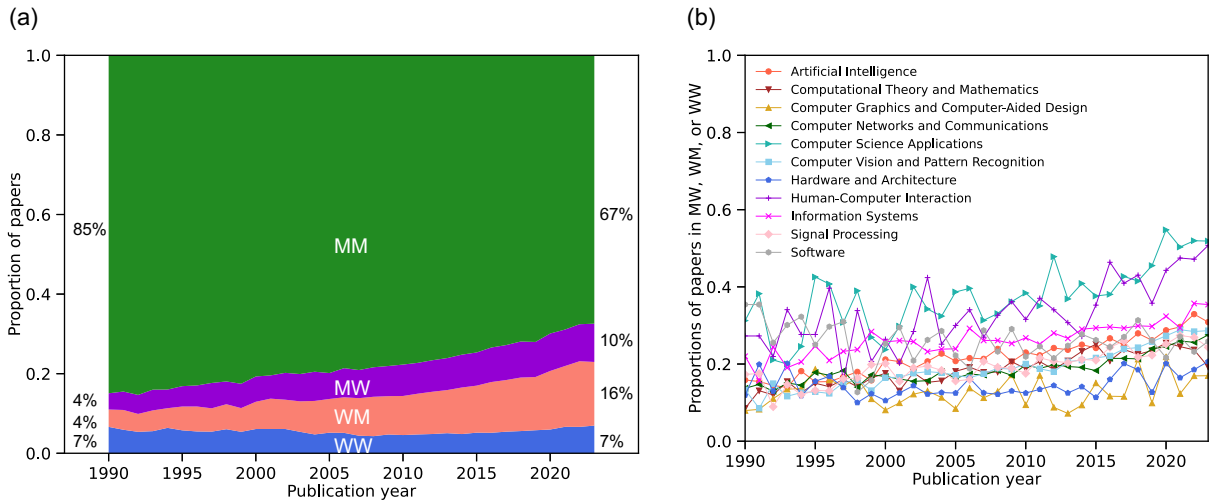

图 1：计算机科学领域按性别类别划分的已发表论文人口统计特征随时间的变化。(a) 各性别类别论文占比随时间的变化。(b) 计算机科学 11 个子领域中 MW、WM 或 WW 性别类别论文占比随时间的变化。

四种性别类别之一——男性-男性（MM）、男性-女性（MW）、女性-男性（WM）或女性-女性（WW）——分类依据是第一作者和末位作者的性别。首字母（M 或 W）表示第一作者的性别，第二个字母表示末位作者的性别。男性或女性独著的论文分别归为 MM 或 WW。我们的数据集包括 298,063 篇 MM 论文（75.6%）、30,970 篇 MW论文（7.8%）、43,727 篇 WM 论文（11.1%）和 21,672 篇 WW 论文（5.5%）。1990年至2023年，第一作者和/或末位作者为女性的论文占比从 15% 上升至 33%（图 1(a)）。尽管这一增幅在不同子领域有所差异，但总体趋势呈上升态势（图 1(b)）。

现在，我们量化计算机科学领域的性别化引用失衡。首先，统计一部分论文对各性别类别论文的引用次数。随后，将得到的引文计数与参照模型推导出的预期数量进行比较；该模型会在保留原始网络某些属性的同时随机化引用。我们将“超额/不足引用”定义为（观测值 –预期值）/预期值，用以衡量观测到的引用次数超出（或低于）预期值的程度（见方法）。这种方法采用了表征实证网络中子图或模式的标准做法，并将揭示无法由常见网络属性直接解释的引用失衡方面 [96, 65, 23,41]。

我们首先将随机抽样模型 [18], 作为参照模型，该模型保留每篇论文的引用数量。在此模型下，我们发现，数据集中的所有论文都显著过度引用 MM 论文，而对 WM 和 WW 论文的引用不足（见图 2(a)）。我们进一步按性别类别对论文所作的引用进行分类 [18, 86]；除非另有说明，我们主要关注 MM 和 WW 论文所作的引用，以区分男性作者和女性作者在引用实践上的差异。MM 论文显著过度引用 MM 论文，而对 MW、WM 和 WW 论文的引用不足（见图 2(b)）。相反，WW 论文显著减少对 MM 论文的引用，而过度引用 MW、WM 和 WW 论文（见图 2(c)）。这些结果与此前关于神经科学和物理学领域期刊论文引文中的性别失衡的研究一致 [18, 86]。

引文网络呈现出两种常见属性 [106]：引文同质性——论文倾向于引用具有相似特征的其他论文——以及每篇论文收到的引文数量异质性——大多数论文获得的引用很少，而少数论文积累了大量引用。

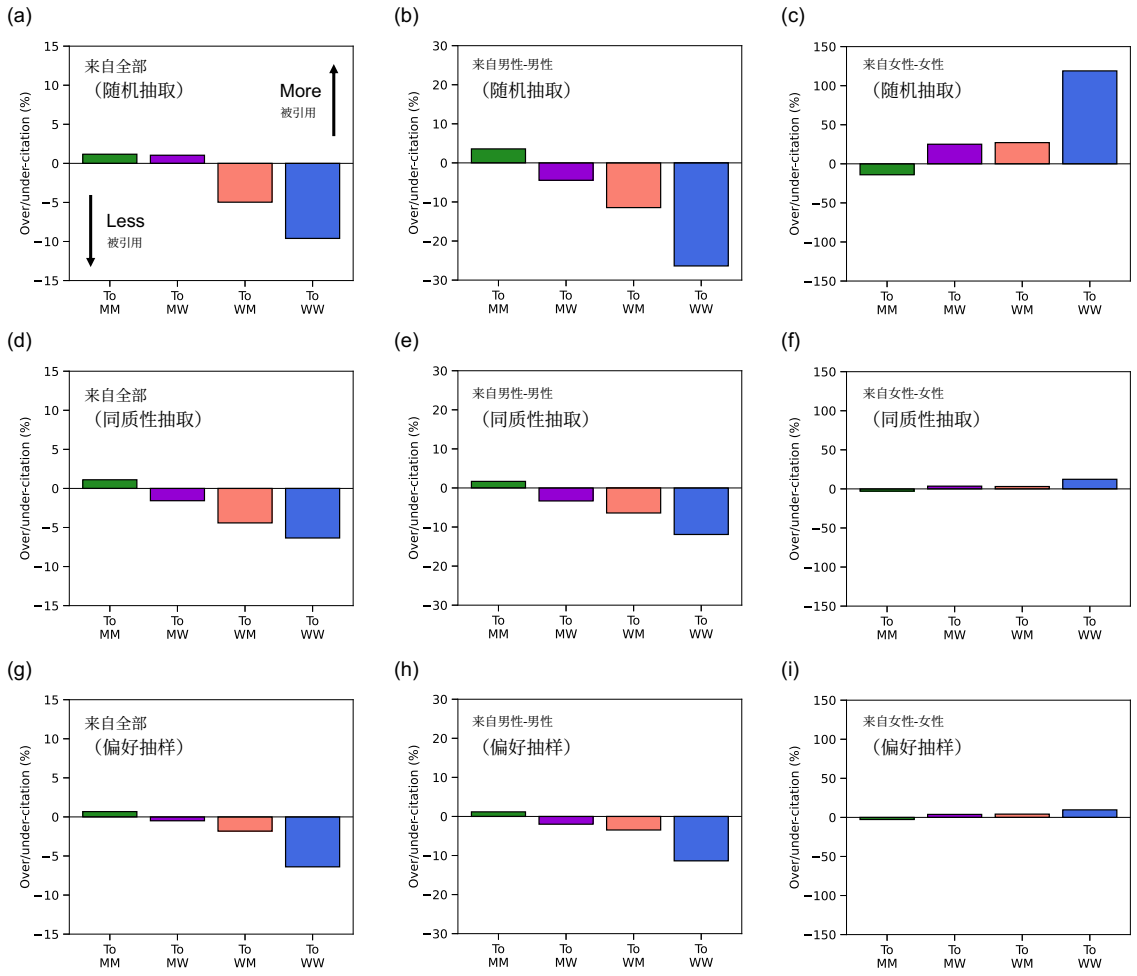

图2：计算机科学论文所作引用中的性别失衡。(a)、(d)和(g)展示全部论文的结果；(b)、(e)和(h)展示MM论文的结果；(c)、(f)和(i)展示WW论文的结果。(a)–(c)采用随机抽样模型，(d)–(f)采用同质性抽样模型，(g)–(i)采用优先抽样模型。关于这三种参照模型下过度引用/引用不足的统计显著性，请参见补充表S6–S8。

由于这些属性常常驱动引文网络中的结构性和动力学行为 [106]，我们假设它们会促成性别化引用失衡。为检验这一假设，我们将随机抽取模型扩展为一系列参照模型。随机抽取模型仅保留每篇论文所作引用的数量，却破坏了同质性和引用异质性；同质性抽样模型则保留以下方面的同质性：**(i)** 所属国家 [28, 58]，、**(ii)** 研究子领域和主题 [87]，以及 **(iii)** 发表场所排名 [95, 47]，；而优先抽样模型还进一步保留每篇论文收到的引文数量异质性。比较这些模型，可以考察这两种属性分别对性别化引用失衡的影响。

引用同质性导致全部论文所作引用中的过度引用/引用不足。事实上，在随机抽样模型中，WW论文所接收的过度引用/引用不足为 -10.1%；而在同质性抽样模型中，这一比例降至 -6.3%（见图2(a)和图2(d)）。我们观察到类似趋势，

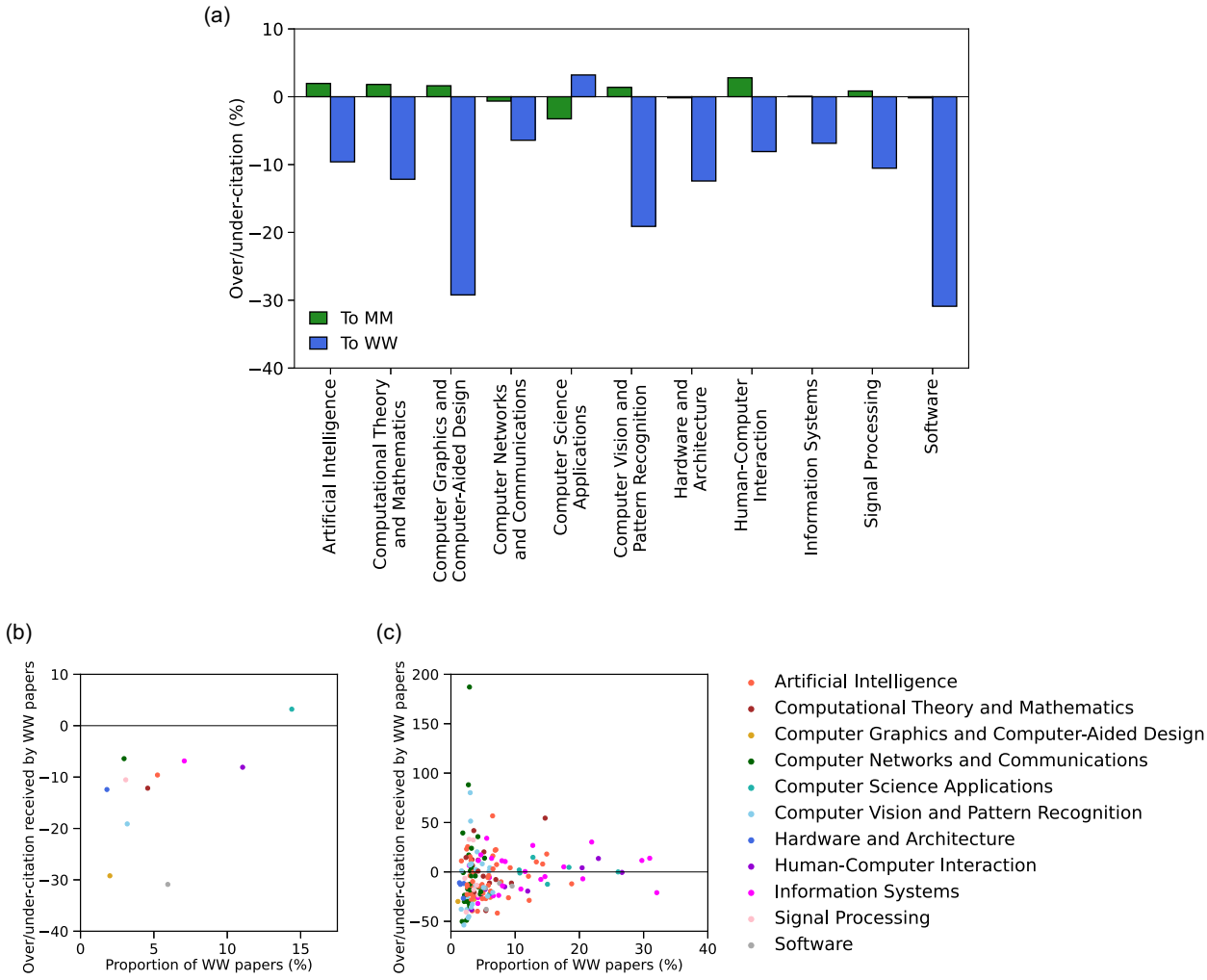

图3：计算机科学各子领域和主题中的性别化引用失衡。(a) 计算机科学各子领域中，全部论文对 MM 和 WW 论文的引用所呈现的引文中的性别失衡。过度引用/引用不足的统计显著性见补充表 S9。(b) 子领域中 WW 论文的比例与全部论文对这些论文的过度引用/引用不足之间的关系。(c) 主题中 WW 论文的比例与全部论文对这些论文的过度引用/引用不足之间的关系。

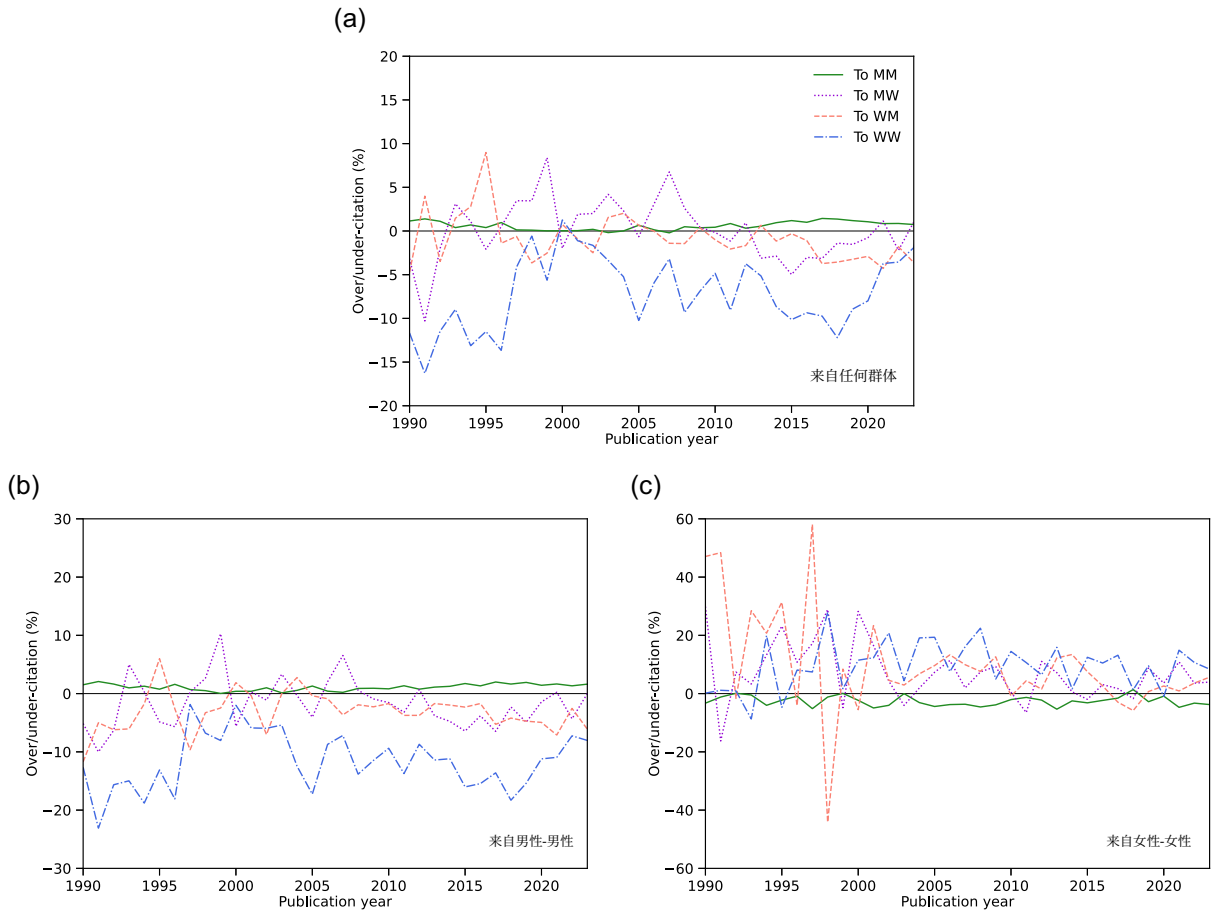

图4：计算机科学论文所引文献中性别失衡的时间趋势。(a)：全部论文。(b)：男性-男性论文。(c)：女性-女性论文。随时间变化的过度引用/引用不足的统计显著性，请参见补充图S3。

.

MM 论文的过度引用/引用不足（见图 2(b) 和 2(e)）；此外，WW 论文的过度引用/引用不足在很大程度上可由引文同质性解释（见图 2(c) 和 2(f)）。这表明，WW 论文呈现的性别化引文模式可能主要归因于结构性因素，例如女性研究人员集中于某些主题、国家和发表 venue 组合。相比之下，每篇论文收到的引文数量异质性影响甚微。事实上，在同质性抽取模型和优先抽样模型中，全部论文的过度引用/引用不足在数量上相近（见图 2(d) 和 2(g)）。我们观察到，MM 论文（见图 2(e) 和 2(h)）以及 WW 论文（见图 2(f) 和 2(i)）的过度引用/引用不足也呈现类似趋势。

总而言之，即使同时考虑引文同质性以及每篇论文收到的引文数量异质性，计算机科学领域依然存在持续的性别化引用失衡。引文同质性与这种失衡密切相关，而后者的影响较小。我们的研究结果表明：（i）解决引文实践中过度的同质性至关重要，包括减少所属国家相同的研究人员之间的过度引用 [2], ；（ii）仅仅关注高被引论文不太可能消除这种失衡。

除非另有说明，后续分析均采用优先抽样模型。该模型量化了无法直接通过每篇论文的引用数量、引文同质性或每篇论文收到的引文数量异质性来解释的引文中的性别失衡。此处我们排除了自引；若将其纳入，性别失衡会更加明显（详情见补充材料 S4 节），这与此前关于性别化自引行为的研究结果基本一致 [39, 18, 86]。此外，尽管我们的分析排除了孤立论文（即既不引用其他论文、也未获得任何引用的论文），但敏感性分析证实，关于过度引用/引用不足模式的主要发现从定性上看并未改变（详情见补充材料 S5 节）。

由于计算机科学子领域的性别失衡程度可能有所不同 [24, 67], ，我们分析了 OpenAlex 划分的计算机科学 11 个子领域。我们计算了各子领域中全部论文对 MM 和 WW 论文的过度引用/引用不足，并比较了这些数值。正如预期，这些子领域的性别化引用失衡程度各不相同（图3(a)）。这种差异可能与研究重点上的性别差异有关；近期证据表明，女性更多地分布在应用研究领域，但教职人员认为，与理论研究领域相比，这些领域的研究人员较不可能发表论文、获得终身教职或获得资助 [40]。研究重点上的此类差异可能是观察到的子领域层面引用失衡的混杂因素。图3(b)展示了各子领域中 WW 论文的比例（横轴）与全部论文对这些论文的过度引用/引用不足情况（纵轴）。我们发现，这两项指标之间存在中度正相关（斯皮尔曼秩相关系数 $= 0.455$）。

在主题层面，我们也观察到 WW 论文的占比与过度引用/引用不足之间存在类似的相关模式。具体而言，我们考察了计算机科学子领域中的 172 个主题，每个主题均包含至少 10 篇 WW 论文，且这些论文累计获得至少 10 次引用。图3(c)展示了各主题中 WW 论文的占比，以及全部论文对这些论文的过度引用/引用不足情况。我们观察到，不同研究主题的过度引用/引用不足程度存在显著差异，且呈现出较弱的正相关关系（斯皮尔曼秩相关系数 $= 0.202$）。

尽管每年发表的 WW 论文比例稳步上升（见图 1(a)），但 MM 论文所获的过度引用和 WW 论文所获的引用不足随时间呈现的趋势相对稳定（见图 4(a)）。按年份和性别对论文分类后，MM 论文的过度引用/引用不足随时间变化的趋势，与全部论文的相应趋势十分相似（见图 4(b)）。WW 论文始终对 MM 论文引用不足、对 WW 论文过度引用，这一模式也一直保持相对稳定（见图 4(c)）。这些发现与先前关于

神经科学和物理学学科的研究结果一致 [18, 86]，并表明引用实践会长期固化，不易随着作者群体日益多元化而改变。

3.2 会议与期刊

在计算机科学领域，国际会议通常是主要的出版渠道，承担着其他学科中期刊通常发挥的作用 [26, 25, 99]。其独特的出版文化以较短的评审周期及其他特征为标志，可能会影响作者的引用实践。事实上，长期以来，该学科一直存在关于会议与期刊的比较讨论[98, 21, 25, 99, 38]。在此，我们从性别化引用失衡的角度考察会议与期刊。

我们比较了全部论文对某一性别类别（男性-男性或女性-女性）中会议论文与期刊论文的过度引用/引用不足情况。男性-男性类别的会议论文被过度引用的比例是男性-男性类别期刊论文的两倍，而女性-女性类别的会议论文被引用不足的比例是女性-女性类别期刊论文的三倍（图 5(a)）。这一结果表明，引用会议论文的作者可能参考的研究人员范围更窄，这或许加剧了计算机科学领域观察到的性别化引用失衡。此外，这种差异表明，该领域独特的出版实践——研究人员往往在国际会议上而非期刊上发表成果[98, 21], ——可能会在出版场所之间形成一种声望等级体系，并与性别化引用失衡相互交织。这一观察结果总体上与此前关于性别失衡与研究机构、研究重点及出版场所声望之间交集的研究相符[102, 3, 33, 40]。

计算机科学领域的会议和期刊经常根据其发表论文的引文指标进行排名 [26]。排名最高的发表场所被视为尤其具有声望[95, 47]。我们探讨发表场所排名与性别化引用失衡程度之间的关系。一种常见的会议排名是 CORE 排名系统，该系统将顶级会议列为 A*，其后依次为 A、B 和 C。对于期刊，我们采用 SCImago 期刊排名，该排名将顶级期刊归为第一季度，其后依次为第二季度、第三季度和第四季度。我们计算全部论文对各发表场所排名和性别类别（男性-男性或女性-女性）所对应论文的过度引用/引用不足情况。

在计算机科学领域，发表场所声望与性别化引用失衡相关，而且这种关联在会议中比在期刊中更为显著（图 5(b)）。男性-男性论文在所有会议等级以及第一季度和第三季度期刊中均被过度引用；女性-女性论文则在 A* 和 B 级会议以及第一季度和第三季度期刊中被引用不足。在 A*级排名会议中，女性-女性论文被引用不足的程度是第一季度期刊中观察到的三倍；而在 A*级排名会议中，男性-男性论文获得的过度引用程度与第一季度期刊相当。

与顶级发表场所（第一季度期刊和 A*级排名会议）相比，中游发表场所（第二季度期刊和 A 级会议）中的性别化引用失衡较不明显，甚至呈现反转。这种非单调模式与先前研究的发现一致，即声望偏见往往在等级体系的顶层最为突出。例如，来自排名靠前机构的作者往往会因其机构声望而获得不成比例的优势，而来自中游或排名较低机构的作者，在同行评审流程中可能面临不同的评价机制[5, 84, 34]。同样，我们的结果表明，顶级会议和期刊中的性别化引用失衡可能更为严重，这或许是因为与中游发表场所相比，它们具有更高的可见度和更激烈的竞争。

我们考察了 120 个会议和 292种期刊中 WW 论文所获引用过度/不足的差异。每个会议或期刊均符合两项标准：（i）至少发表 10 篇 WW 论文；（ii）这些 WW 论文至少获得 10 次引用。在会议和期刊中，发表在该出版场所的 WW 论文所占比例与引用过度/不足几乎没有相关性——

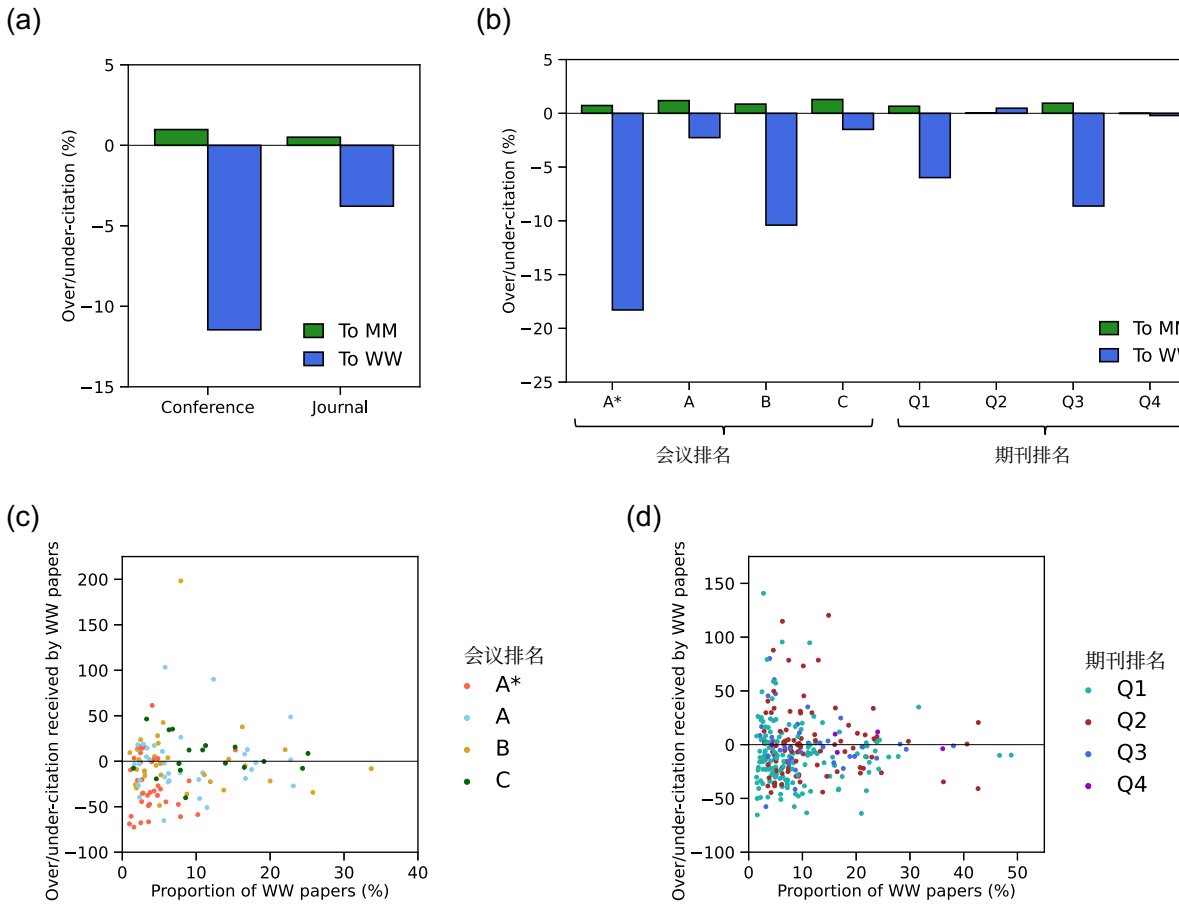

图 5：计算机科学不同出版场所中的性别化引用失衡。(a) 会议和期刊上发表的 MM 和 WW 论文收到的过度引用/引用不足。过度引用/引用不足的统计显著性见补充表S10。(b) 不同发表场所排名中的 MM 和 WW 论文收到的过度引用/引用不足。过度引用/引用不足的统计显著性见补充表S11。(c) 会议中 WW 论文的比例与这些论文收到的过度引用/引用不足之间的关系。(d) 期刊中 WW 论文的比例与这些论文收到的过度引用/引用不足之间的关系。

这些论文所接收的引用；会议的斯皮尔曼秩相关系数为0.092（图5(c)），期刊为0.026（图5(d)）。

总而言之，在计算机科学领域，会议和期刊中均存在性别化引用失衡，但会议论文中的失衡更为明显，尤其是在顶级会议上发表的论文。此外，会议和期刊之间性别化引用失衡程度的差异，很大程度上无法由报告作者中潜在的性别失衡解释。报告作者性别比例与实际引用实践之间的这种差异，可能源于系统性因素，例如会议主导文化 [98, 21], 知名作者的参与 [94, 84],以及合著网络 [52, 38]。

3.3 性别化引用失衡的潜在驱动因素

基于这些发现，我们考察了计算机科学领域性别化引用失衡的潜在驱动因素。为此，我们在控制关键论文特征 [18, 86]的情况下，比较论文所呈现的过度引用/引用不足模式。我们采用与此前关于研究职业生涯中性别失衡的研究 [33, 46], 相同的匹配对分析，以评估引用行为中呈现出更强性别失衡的论文是否具有某些相关特征。

首先，我们假设，正如图 2 所示的模式所表明的那样，作者性别会影响其论文引用中的性别失衡。为检验这一点，我们进行了匹配对分析，在控制性别类别以外的论文特征后开展比较（见方法）。我们分别构建 MM 论文和 WW 论文的独立匹配对组——每对论文在发表年份、所属国家、子领域和引用数量方面相匹配——随后比较这些匹配组中 MM 论文与 WW 论文的过度引用/引用不足情况。

MM 论文对各性别类别论文的过度引用/引用不足情况，与 WW 论文的情况存在显著差异（图 6(a) ）。具体而言，MM 论文过度引用 MM 论文，而对 MW 、 WM 和 WW 论文引用不足；WW 论文则对 MM 论文引用不足，而过度引用 MW、WM 和 WW 论文。这些发现与图 2(h) 和图 2(i) 一致；在控制发表年份、所属国家、子领域以及论文引用数量后，这些性别化引用模式依然存在。

接下来，我们考察 MM 和 WW 论文的哪些具体特征会导致这些论文所作引用中的性别失衡。我们聚焦于三个特征：(i) 发表场所类型（会议或期刊）；(ii) 是否有“知名”作者参与（即拥有高被引论文的作者）；以及 (iii) 局部合著网络。我们再次采用匹配对分析，检验每项特征是否会影响性别失衡（见方法部分）。首先，根据某一特征将 MM（或 WW）论文划分为两个互不重叠的集合；例如，在考察发表场所类型时，将 MM 会议论文与 MM 期刊论文分开。其次，我们生成相互独立的 MM（或 WW）匹配论文对集合，并控制发表年份、所属国家、子领域以及论文所作引用的数量。最后，我们计算各匹配集合中 MM（或 WW）论文的过度引用/引用不足，以评估该特征与性别化引用失衡之间的关联。

我们发现，这三个特征与 MM 和 WW 论文引用中的性别失衡之间存在若干关联。首先，MM 会议论文过度引用/引用不足的绝对幅度大于 MM 期刊论文（图 6(b)）。WW 会议论文和 WW 期刊论文对 MM 论文的引用不足程度相近，但 WW 会议论文对 WW 论文的过度引用程度高于 WW 期刊论文（图 6(b)）。其次，知名作者的参与对 MM 和 WW 论文呈现出不同的定性影响（图 6(c)）：有知名作者参与时，MM 论文过度引用/引用不足的绝对幅度更大，而 WW 论文的这一幅度则更小。

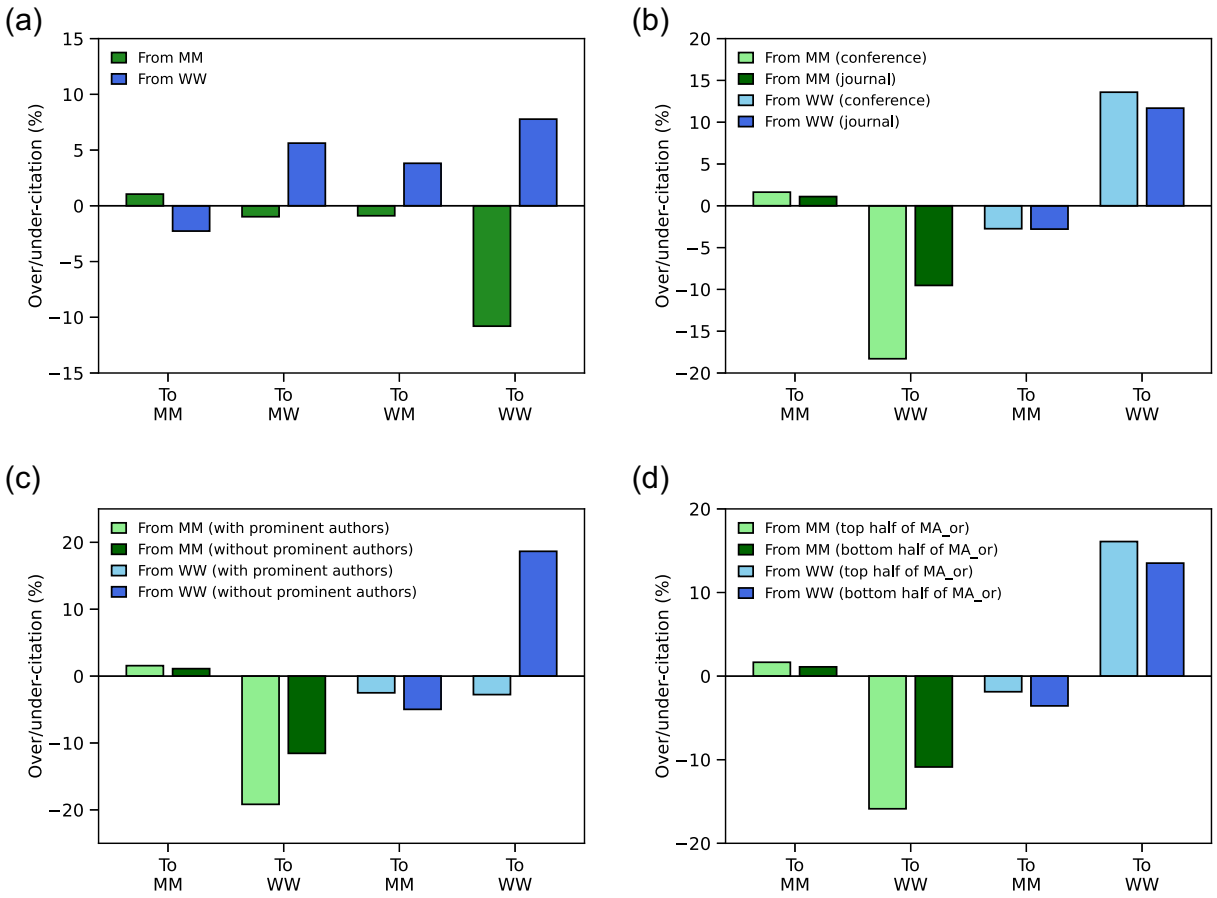

图6：论文引用中性别失衡潜在驱动因素的配对分析。(a) 作者性别。(b) 发表场所类型。(c) 知名作者的参与情况。(d) 本地合作署名网络。关于面板(a)、(b)、(c)和(d)中所示过度引用/引用不足的统计显著性，请参见补充表S14–S17。

这表明，尽管作者声望会增强男性同质性，却会削弱女性同质性。这一结果或许说明，声望较高的女性作者可能会采用占主导地位、以男性为中心的引用规范。尽管这一假设有待进一步审慎研究，但它揭示了作者声望与引用行为之间复杂的相互作用。最后，第一作者和末位作者的局部合著网络与其论文的过度引用/引用不足相关（图 6(d)），这与此前在神经科学领域的研究结果一致 [18]。具体而言，男性作者拥有较高比例男性合著者（即 MA_{or}范围的上半部分）的 MM 论文，其过度引用/引用不足的绝对幅度大于男性作者拥有较少男性合著者（即 MA_{or} 范围的下半部分）的 MM 论文。合著网络中男性代表比例较高（即 MA_{or} 范围的上半部分）的 WW 论文，对 WW 论文表现出更强的过度引用。这一结果或许表明，处于男性主导的合著网络中的女性可能更敏锐地意识到性别化引用失衡，并采用旨在纠正这种失衡的引用规范。

4 讨论

在本研究中，我们考察了计算机科学中性别化引用失衡的程度。我们发现，在计算机科学的各个子领域、主题和发表 venue 中，以女性为第一作者或末位作者的论文获得的引用更少。以往研究分析了计算机科学中研究人员人口统计特征、生产力、研究职业生涯和合作模式方面的性别失衡 [102, 36, 33, 43, 56, 48, 31, 100, 29]。我们的研究从引用实践的角度拓展了以往关于计算机科学中性别失衡的研究发现。我们的结果表明，计算机科学独特的以会议为导向的文化中的引用实践，可能会加剧性别化引用失衡。

我们发现，计算机科学会议论文比期刊论文更容易出现性别化引用失衡——无论是在进行引用还是获得引用方面。造成这种差异的因素可能有多个。其中之一是计算机科学会议高度压缩的同行评审周期，通常只有几周。虽然这种较短的周转时间能够加快决策和知识传播，但也引发了人们对评审深度和质量的担忧。Vardi 强调了会议程序委员会所承受的“极端的时间和工作量压力”[98]。Fortnow 认为，截止日期驱动的会议文化可能导致任意决策，并使速度优先于充分审查 [98, 21]。这些缩短的时间安排和仓促的评估可能会无意中助长选择性引用实践，从而加剧性别化引用失衡，但仍需进一步开展定量研究，以确定这种影响的确切程度。

我们的研究结果表明，计算机科学顶级会议论文中的性别化引用失衡尤为显著。除了固定的传统日程安排和简短的同行评审周期外，这些会议的迅速扩张已超出程序委员会的承载能力，显著增加了评审人员的工作负担 [98, 21]。早在2009年，Birman和Schneider就指出，顶级计算机系统会议已不堪投稿重负，导致评审人员过度劳累，论文评估也可能流于草率 [4]。2019年，计算机与通信安全旗舰会议（即所谓的CCS会议）的专家小组报告称，2020年四个顶级计算机安全会议的投稿量激增至3,039篇，这引发了人们对评审人员能否充分评估如此大量稿件的担忧 [80]。作为回应，若干顶级会议已开始为作者提供答辩机会 [32]。未来研究应考察顶级会议中的这些同行评审动态如何影响引用实践，从而确保计算机科学领域的学术认可公平公正。

计算机科学会议历来严格限制篇幅，以规范投稿并缩短

评审周期 [30]。尽管这类限制仍很常见，许多声望卓著的会议如今已允许提交篇幅更长、内容更丰富的论文。例如，人机交互领域的旗舰会议（即所谓的 CHI 会议）于 2016年取消了正式的页数限制，推动论文篇幅和参考文献数量双双增加 [64]。类似地，Geiger 对另一场顶级会议的分析表明，取消页数限制后，论文平均篇幅有所增长，反映出学界倾向于更全面报告研究的规范 [27]。这些变化旨在兼顾短评审周期下的效率、会议论文集的物理限制，以及更全面地报告研究的目标。尽管不断变化的页数限制政策似乎增加了参考文献数量，但这一趋势将如何影响计算机科学中的性别化引用失衡，仍不清楚。

会议与期刊在性别化引用失衡方面的显著差异，也可能部分源于引用实践中的“主题接近度”。物理学领域的近期研究表明，当作者引用自己不太熟悉的研究成果（即主题接近度较低）时，性别化引用失衡往往会加剧 [86]。在计算机科学中，一个可能的假设是，会议出版物的限制（例如紧迫的截止期限和页数限制）可能会导致作者在引用自身直接专业领域之外的研究时，优先选择高可见度或高引用量的论文。正如我们所展示的，由于计算机科学在历史上一直由男性主导，这种倾向可能会加剧性别化引用失衡，相较之下，期刊论文的作者可能拥有更多资源来开展全面的文献检索。未来的研究可以通过使用论文分类或文本嵌入来测量主题接近度，从而探讨这种可能的关联。

先前的研究采用参照模型来刻画引用行为。Uzzi 等人通过使用蒙特卡洛算法对引用进行随机化，考察了与论文影响力相关的非典型组合 [96]。Kojaku 等人使用一种考虑科学共同体和期刊规模的参照模型，考察了跨期刊的异常引用模式[41]。这些参照模型并非旨在量化论文所收到的引用中的失衡，因为它们保留了原始网络中每篇论文收到的引用数量。Dworkin 等人提出了一种用于量化论文所收到的引用中的性别失衡的参照模型 [18]。该模型并非用于研究网络属性与性别化引用失衡之间的关联，因为它只保留了原始网络中每篇论文所作出的引用数量。为解决这些局限，我们开发了一系列参照模型，在保留每篇论文所作出的引用数量的同时，还保留两项网络属性：引文同质性，以及每篇论文收到的引文数量异质性。

通过对引文网络使用参照模型，我们发现引文同质性是造成性别化引用失衡的重要因素。此前的研究也关注了与作者性别相关的引文同质性 [87, 107]。Tekles 等人报告称，在生物与医学领域，性别化引文模式主要由研究主题方面的同质性驱动 [87]。Zhou 等人发现，作者性别同质性在一定程度上解释了生命科学领域的引用行为[107]。通过控制与所属国家、研究主题和发表场所排名相关的引文同质性，我们确定了其对性别化引用失衡的影响。结合此前的研究，我们的发现凸显了在分析性别化引文模式时控制引文同质性的重要性。

通过进行匹配对分析并控制关键的论文层面特征，我们发现，第一作者或末位作者的显著性及其局部合著网络，会加剧论文所作引用中的性别失衡。这一观察与以往研究结果一致：这些因素在一定程度上塑造了研究者实践。事实上，在顶尖计算机科学会议上，在评审者知道作者姓名的同行评审中，有知名作者的稿件得分高于没有知名作者的稿件 [94, 84]。同样，局部合著网络会影响神经科学和物理学研究人员基于性别的引用模式 [18, 86]。通过使用 Yates 卡方检验 [1], , 我们发现这些特征并非相互独立（见补充材料第

S7 节）。这些因素之间的相互作用可能会加剧性别化引用失衡，因此有必要系统地研究它们的综合影响。

我们的研究存在若干局限。首先，我们采用的性别判定方法在中文姓名基准数据 [58]上的表现不佳。因此，我们未能为许多中国作者撰写的论文判定性别类别。鉴于中国对计算机科学研究 [28], 的贡献迅速增长，未来研究应考察中国研究人员撰写的论文中是否存在性别化引用失衡。其次，尽管我们假设第一作者和末位作者在每篇论文中都发挥主导作用，但这一假设并非总是成立。事实上，多位作者之间的贡献分配方式可能导致性别失衡 [73], ，而一些计算机科学子领域则采用按姓氏字母顺序排列的方式 [79]。未来还需进一步研究如何采用考虑性别的作者贡献分配方法，并评估这些子领域中的性别化引用失衡。第三，作者所属国家 [28], 、国籍 [58], 和种族 [42]也可能存在其他引用失衡。尽管我们关注的是与作者性别相关的引用失衡，但我们的定量框架并不限于此。例如，研究密集型国家与其他国家之间可能存在引用失衡 [28], ，可通过参照模型进行分析，并控制网络属性以及所属国家以外的论文特征。系统考察作者身份各个方面的引用失衡，是未来研究的重要方向。第四，我们承认，计算机科学正变得日益跨学科 [10]。这一趋势使得全面界定该领域的研究格局颇具挑战，我们的分析可能遗漏跨学科领域中一些相关的出版物、会议或期刊，而计算机科学在这些领域中发挥着主要作用。第五，尽管发表场所排名、作者声望和合著网络会随时间变化，我们的分析却将它们视为静态属性。未来有必要开展纳入这些时变属性的纵向分析，进一步研究作者或发表场所的声望随时间的演变会如何影响引用实践。

尽管存在这些局限性，我们的研究对计算机科学中的性别化引用失衡进行了系统且大规模的分析，涵盖会议和期刊。我们预计，研究结果将为正在开展的工作提供可操作的见解，以推动营造更加公平、包容且可持续的学术环境。

附录A 参照模型比较

我们将原始网络的结构属性与由三种参照模型生成的随机化网络的结构属性进行了比较。首先，所有模型都精确保留了每篇论文所作引用的分布（图7(a)）。其次，随机抽取模型破坏了同质性引用模式（包括会议排名、期刊排名、所属国家和研究主题），而同质性抽取模型和优先抽样模型则保留了这些模式（图7(b)–(f)）。第三，随机抽取模型无法较好地保留每篇论文收到的引文数量异质性，而同质性抽取模型和优先抽样模型能够更接近地近似该异质性，其中后者更为接近（图7(g)）。这些结果与我们的预期一致。

致谢

我们感谢 Naoki Masuda（纽约州立大学布法罗分校）提供反馈。本研究由日本科学技术振兴机构（JST）通过“面向创新研究生态系统的可持续合作伙伴关系计划（ASPIRE）”提供支持，项目编号为 JPMJAP2328。KN 感谢日本学术振兴会科研费的资助，项目编号为 24K21056。YS 感谢所获支持

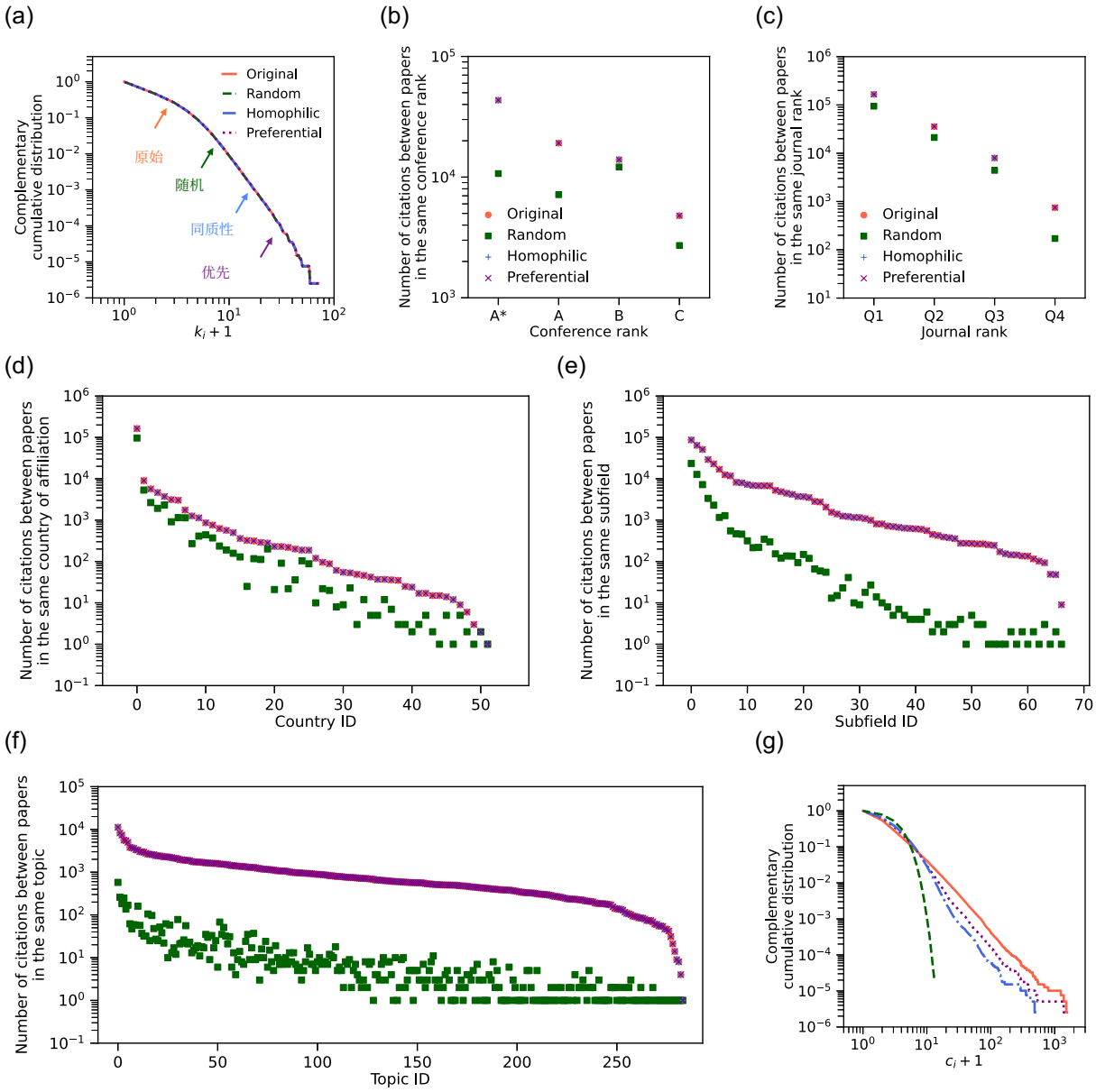

图7：原始网络与参照模型的结构属性比较。(a) 每篇论文引用数量的分布。(b)–(f) 同质性引用模式，分别针对 (b) 会议排名、(c) 期刊排名、(d) 所属国家、(e) 子领域和 (f) 主题。(g) 每篇论文收到的引用数量分布。“原始”表示原始网络，“随机”表示随机抽样模型，“同质性”表示同质性抽样模型，“优先”表示优先抽样模型。在 (d)–(f) 中，我们关注原始网络和各参照模型中至少出现过一次同质性引用的 52 个国家、67 个子领域和 284 个主题。完全或大部分重叠的曲线以箭头和标签标示。

由日本科学技术振兴机构 PRESTO资助，资助编号为JPMJPR21C5。

利益冲突

作者声明不存在任何利益冲突。

数据可用性

2024年9月的 OpenAlex 快照可从 https://docs.openalex.org/download-all-data/ openalex-snapshot 获取。2024年7月的 DBLP 快照可从 https://blog.dblp. org/tag/snapshot/ 获取。CORE排名数据可从 https://portal.core.edu.au/ conf-ranks/ 获取。SCImago 期刊排名数据可从 https://www.scimagojr.com/ journalrank.php?area=1700&year=2021 获取。Gender API 可付费使用，网址为 https:// gender-api.com/。OpenAlex 与 DBLP 整合的 3,720,575 篇论文（见第 2.1.3 节）的数据可从 https://doi.org/10.5281/zenodo.16537950 获取。

参考文献

[1] G. Abramo、D. W. Aksnes 和 C. A. D'Angelo。《各国之间及各国内部科研表现的性别差异：意大利与挪威的比较》。《信息计量学杂志》，15:101144，2021年。[2] A. Baccini 和 E. Petrovich。《1996—2019年各国自引的全球探索性比较》。《公共科学图书馆·综合》，18:e0294669，2023年。[3] M. H. K. Bendels、R. M¨uller、D. Brueggmann 和 D. A. Groneberg。《自然指数期刊揭示的高质量研究中的性别差异》。《公共科学图书馆·综合》，13:1–21，2018年。[4] K. Birman 和 F. B. Schneider。《观点：系统领域的程序委员会过载》。《通信》，52(5):34–37，2009年。[5] R. M. Blank。《双盲评审与单盲评审的影响：来自《美国经济评论》的实验性证据》。《美国经济评论》，第1041–1067页，1991年。[6] L. Born mann、R. Mutz 和 H.-D. Daniel。《基金同行评审中的性别差异：一项元分析》。《信息计量学杂志》，1:226–238，2007年。[7]K. Buchholz。《哪些国家的学生最积极参与科学、技术、工程和数学？》技术报告，世界经济论坛，2023年。https://www.weforum.org/stories/2023/03/which-countries-students-are-getting-most-involved-in-stem/ [访问于3月2025]。[8] N. Caplar、S. Tacchella 和 S. Birrer。《基于引文计数对天文学出版物中的性别偏见进行量化评估》。《自然·天文学》，1:0141，2017年。[9] A. Cavacini。《计算机科学期刊论文的最佳数据库是什么？》《科学计量学》，102:2059–2071，2015年。[10] T. Chakraborty。《跨学科性在计算机科学中的作用：量化、影响与生命轨迹》。《科学计量学》，114:1011–1029，2018年。

[11] P. Chatterjee 和 R. M. Werner。《高影响力期刊论文中的引用性别差异》。JAMA Network Open，4:e2114509–e2114509，2021年。[12] V. Ciotti、M. Bonaventura、V. Nicosia、P. Panzarasa 和 V. Latora。《引文网络中的同质性与缺失链接》。EPJ Data Science，5:7，2016年。[13] A. Clauset、S. Arbesman 和 D. B. Larremore。《教师招聘网络中的系统性不平等与等级体系》。Science Advances，1:e1400005，2015年。[14] J. M. Cohoon、S. Nigai 和 J. J. Kaye。《计算机会议论文中的性别与计算》。通信，54:72–80，2011年。[15]CSRankings。《CSRankings：计算机科学排名》。网址：https://csrankings.org/#/index?all&us [访问于1月 2026]。[16] L. Delgado-Quir´os 和 J. L. Ortega。《八个免费开放的学术数据库中出版物元数据的完整度》。Quantitative Science Studies，第1–19页，2024年。[17] M. L. Dion、J. L. Sumner 和 S. M. Mitchell。《政治学与社会科学方法论领域中的性别化引文模式》。Political Analysis，26:312–327，2018年。[18] J. D. D workin、K. A. Linn、E. G. Teich、P. Zurn、R. T. Shinohara 和 D. S. Bassett。《神经科学参考文献列表中性别失衡的程度及其驱动因素》。Nature Neuroscience，23:918–926，2020年。[19] Y.-H. Eom 和 S. Fortunato。《引文动态的特征分析与建模》。《公共科学图书馆·综合》，6:1–7，2011年。[20]Eurostat。《信息与通信技术行业——增加值、就业和研发》，2024年。https: //ec.europa.eu/eurostat/statistics-explained/index.php?title=ICT_sector_-_value_added,_employment_and_R%26D [访问于3月 2025]。[21] L. Fortnow。《观点：计算机科学该成熟起来了》。Communications of the ACM，52:33–35，2009年。[22] S. Fortunato、C. T. Bergstrom、K. B¨orner、J. A. Evans、D. Helbing、S. Milojevi´c、A. M. Petersen、F. Radicchi、R. Sinatra、B. Uzzi、A. Vespignani、L. Waltman、D. Wang 和 A.-L. Barab´asi。《科学学》。Science，359:eaao0185，2018年。[23] B. K. Fosdick、D. B. Larremore、J. Nishimura 和 J. Ugander。《固定度序列的随机图模型配置》。SIAM Review，60:315–355，2018年。[24] E. Frachtenberg 和 R. D. Kaner。《计算机系统研究中的女性代表性不足》。《公共科学图书馆·综合》，17:e0266439，2022年。[25] M. Franceschet。《会议发表在计算机科学中的作用》。Communications of the ACM，53:129–132，2010年。[26] J. Freyne、L. Coyle、B. Smyth 和 P. Cunningham。《计算机科学中期刊与会议出版物的相对地位》。Communications of the ACM，53:124–132，2010年。[27] R. S. Geiger。《短文的兴衰：ACM CSCW中论文篇幅的变化，2000—2018年》。Proceedings of the ACM on Human-Computer Interaction，3，2019年。

[28] C. J. Gomez、A. C. Herman 和 P. Parigi。全球科学领域的领先国家，其研究与其他国家相似，却日益获得更多引用。*Nature Human Behaviour*，6:919–929，2022年。[29] A. Hajibabaei、A. Schiffauerova 和 A. Ebadi。人工智能科学生态系统中的性别特定模式。《信息计量学杂志》，16:101275，2022年。[30] J. Y. Halpern 和 D. C. Parkes。期刊用于认证，会议用于快速传播。*Communications of the ACM*，54:36–38，2011年。[31] L. Holman、D. Stuart-Fox 和 C. E. Hauser。科学领域的性别差距：女性何时才能获得同等代表性？*PLOS Biology*，16:1–20，2018年。[32] J. Huang、W. bin Huang、Y. Bu、Q. Cao、H. Shen 和 X. Cheng。如何在计算机科学会议上成功答辩？关于社会互动的视角。《信息计量学杂志》，17:101427，2023年。[33] J. Huang、A. J. Gates、R. Sinatra 和 A.-L. Barab´asi。跨国家和学科的科学职业生涯性别不平等历史比较。《美国国家科学院院刊》，117:4609–4616，2020年。[34] A. E. Hultgren、N. M. Patras 和 J. Hicks。元研究：盲审可减少青年研究者奖申请初审中的机构声望偏见。*eLife*，13:e92339，2024年。[35] J. P. A. Ioannidis、K. W. Boyack、T. A. Collins 和 J. Baas。通过分析近580万名作者，考察不同时期各科学学科中高被引科学家的性别失衡。*PLOS Biology*，21:e3002385，2023年。[36] M. Jadidi、F. Karimi、H. Lietz 和 C. Wagner。科学领域的性别差异？男性和女性计算机科学家的退出、生产力、合作与成功。*Advances inComplex Systems*，21:1750011，2018年。[37] J. Kim。评估数字图书馆中的作者姓名消歧：以DBLP为例。《科学计量学》，116:1867–1886，2018年。[38] J. Kim。计算机科学中会议发表与期刊发表的基于作者的分析。*Journal of the Association for Information Science and Technology*，70:71–82，2019年。[39] M. M. King、C. T. Bergstrom、S. J. Correll、J. Jacquet 和 J. D. West。男性给自己引用得更高：跨领域、跨时期的性别与自引。*Socius*，3:2378023117738903，2017年。[40] S. Kleinberg 和 J. K. Marsh。女性在哪里：计算领域的性别失衡，以及教职人员对理论研究和应用研究的认知。*IEEE Access*，13:73520–73529，2025年。[41] S. Kojaku、G. Livan 和 N. Masuda。检测期刊网络中的异常引用群组。*Scientific Reports*，11:14524，2021年。[42] D. Kozlowski、V. Larivi`ere、C. R. Sugimoto 和 T. Monroe-White。科学领域的交叉性不平等。《美国国家科学院院刊》，119:e2113067119，2022年。

[43] N. Laberge、K. H. Wapman、A. C. Morgan、S. Zhang、D. B. Larremore 和 A. Clauset。美国计算机领域教职人员中的子领域声望与性别不平等。*Communications ofthe ACM*，65:46–55，2022年。[44] V. Larivi`ere、C. Ni、Y. Gingras、B. Cronin 和 C. R. Sugimoto。文献计量学：全球科学领域的性别差异。《自然》，504(7479):211–213，2013年。[45] M. Ley。DBLP计算机科学书目：发展历程、研究问题与展望。载于字符串处理与信息检索，第1–10页，2002年。[46] W. Li、S. Zhang、Z. Zheng、S. J. Cranmer 和 A. Clauset。厘清科学家生产力和显著性的网络效应。《自然·通讯》，13:4907，2022年。[47] X. Li、W. Rong、H. Shi、J. Tang 和 Z. Xiong。计算机科学会议排名系统的影响：比较回归分析。《科学计量学》，116:879–907，2018年。[48] H. Lietz、M. Jadidi、D. Kostic、M. Tsvetkova 和 C. Wagner。计算机科学中的个体与性别不平等：对1970至2000年各届队列的职业生涯研究。*Quantitative ScienceStudies*，第1–25页，2024年。[49] A. E. Lincoln、S. Pincus、J. B. Koster 和 P. S. Leboy。科学领域的马蒂尔达效应：美国20世纪90年代和21世纪初的奖项与奖品。《科学社会研究》，42:307–320，2012年。[50] D. Maliniak、R. Powers 和 B. F. Walter。国际关系领域的性别引用差距。《国际组织》，67:889–922，2013年。[51] L. I. Meho。2001至2020年国际知名研究奖项中的性别差距。*Quantitative Science Studies*，2:976–989，2021年。[52] G. V. Menezes、N. Ziviani、A. H. Laender 和 V. Almeida。计算机科学知识生产的地理分析。载于第*18*届万维网国际会议论文集，第1041–1050页，2009年。[53] R. Milo、S. Shen-Orr、S. Itzkovitz、N. Kashtan、D. Chklovskii 和 U. Alon。网络基元：复杂网络的简单构件。*Science*，298:824–827，2002年。[54] S. Milojevi´c。科学学。《科学计量学》，2025年。[55] L. B. Moreno、M. C. Franco、S. A. Karam、F. H. van de Sande 和 A. F. Montagner。过去40年牙科领域主要出版物中持续存在的性别差异：一项观察性研究。《临床流行病学杂志》，171:111386，2024年。[56] A. C. Morgan、S. F. Way、M. J. D. Hoefer、D. B. Larremore、M. Galesic 和 A. Clauset。为人父母对学术界产生的不平等影响。《科学进展》，7:eabd1996，2021年。[57] C. A. Moss-Racusin、J. F. Dovidio、V. L. Brescoll、M. J. Graham 和 J. Handelsman。科学领域教职人员不易察觉的性别偏见更青睐男学生。《美国国家科学院院刊》，109:16474–16479，2012年。[58] K. Nakajima、R. Liu、K. Shudo 和 N. Masuda。量化东亚学术界的性别失衡：科研生涯与引用实践。《信息计量学杂志》，第101460页，2023年。

[59] K. Nakajima、Y. Sasaki、S. Tokuno和G. Fletcher。量化计算机科学会议中的性别化引用失衡。载于*AAAI/ACM*人工智能、伦理与社会会议，第1011–1022页，2024年。[60] M. E. Newman、S. H. Strogatz和D. J. Watts。具有任意度分布的随机图及其应用。《物理评论*E*》，64:026118，2001年。[61] H. O.'Leary、T. Gantzert、A. Mann、E. Z. Mann、N. Bollineni和M. Nelson。引用即表征：环境研究出版物中的性别化学术引用政治依然存在。《环境研究与科学期刊》，14:525–537，2024年。[62]B. Oliss、C. McFaul和J. C. Riddick。科学、技术、工程和数学专业毕业生的全球分布：哪些国家领先？技术报告，安全与新兴技术中心，2023年。https://cset.georgetown.edu/article/the-global-distribution-of-stem-graduates-which-countries-lead-the-way/ [访问于3月 2025]。[63]OpenAlex。OpenAlex：主题分类端到端流程，2024年。https: //docs.google.com/document/d/1bDopkhuGieQ4F8gGNj7sEc8WSE8mvLZS/edit#heading=h.5w2tb5fcg77r [访问于4月 2024]。[64] J. Oppenlaender。人机交互中引用实践的过去、现在与未来。*arXiv*预印本 *arXiv:2405.16526*，2024年。[65] C. Orsini、M. M. Dankulov、P. Colomer-de Sim´on、A. Jamakovic、P. Mahadevan、A. Vah- dat、K. E. Bassler、Z. Toroczkai、M. Bogu˜n´a、G. Caldarelli、S. Fortunato和D. Krioukov。量化真实网络中的随机性。《自然·通讯》，6:8627，2015年。[66] S. Pandey和T. Burch-Smith。克服引用偏见是实现植物科学真正包容性的必要条件。《植物细胞》，36:10–13，2023年。[67] E. Peltonen。性别在国际普适计算与通信会议中的作用。《计算机科学前沿》，4:1008552，2022年。[68] V. Petricek、I. J. Cox、H. Han、I. G. Councill和C. L. Giles。在线计算机科学引文数据库比较。载于数字图书馆研究与先进技术，第438–449页，2005年。[69] J. Priem、H. Piwowar和R. Orr。OpenAlex：学术作品、作者、发表场所、机构和概念的完全开放索引。*arXiv*预印本 *arXiv:2205.01833*，2022年。[70] F. Radicchi、S. Fortunato和C. Castellano。引用分布的普适性：迈向科学影响力的客观衡量标准。《美国国家科学院院刊》，105:17268–17272，2008年。[71] F. Reitz和O. Hoffmann。分析DBLP数字图书馆中计算机科学子领域收录范围的演变。载于数字图书馆研究与先进技术，第216–227页，2010年。[72] A. Rosenfeld。DBLP是一个优秀的计算机科学期刊数据库吗？《*Computer*》，56:101–108，2023年。

[73] M. B. Ross、B. M. Glennon、R. Murciano-Goroff、E. G. Berkes、B. A. Weinberg和J. I. Lane。女性在科学领域获得的认可少于男性。《自然》，608:135–145，2022年。[74] L. Santamar´ıa和H. Mihaljevi´c。姓名性别推断服务的比较与基准评估。PeerJ计算机科学，4:e156，2018年。[75] S. Schneegans、J. Lewis和T. Straza。联合国教科文组织科学报告：为更智慧的发展与时间赛跑。执行摘要。技术报告，联合国教科文组织，2021年。https://unesdoc.unesco.org/ark:/48223/pf0000377250 [访问于3月 2025]。[76] SCImago。SCImago期刊与国家排名（SJR） [门户]。检索日期：2024年11月1日，网址：https://www.scimagojr.com/journalrank.php?area=1700&year=2021。[77] J. M. Sheltzer和J. C. Smith。生命科学领域的精英男性教职人员雇用的女性较少。《美国国家科学院院刊》，111:10107–10112，2014年。[78] M.-A. Sicilia、S. S´anchez-Alonso和E. Garc´ıa-Barriocanal。比较两个不同引文数据库的影响因子：以计算机科学为例。《信息计量学杂志》，5:698–704，2011年。[79] J. Solomon。程序员、教授与寄生虫：计算机科学中的认可与共同署名。《科学与工程伦理》，15:467–489，2009年。[80]A. Soneji、F. B. Kokulu、C. Rubio-Medrano、T. Bao、R. Wang、Y. Shoshitaishvili和A. Doup´e。“虽有缺陷，但就像民主一样，我们没有更好的制度”：专家对安全领域论文评审流程的见解。载于2022年IEEE安全与隐私研讨会（SP），第1845–1862页，2022年。[81] M. Song、G. E. Heo和S. Y. Kim。分析生物信息学中的主题演变：利用DBLP会议数据考察该领域的动态。《科学计量学》，101:397–428，2014年。[82] K. Spoon、N. LaBerge、K. H. Wapman、S. Zhang、A. C. Morgan、M. Galesic、B. K. Fosdick、D. B. Larremore和A. Clauset。美国教职人员中的性别与留任模式。《科学进展》，9:eadi2205，2023年。[83] C. R. Sugimoto和V. Larivi`ere。《科学领域女性公平》。哈佛大学出版社，马萨诸塞州剑桥，2023年。[84] M. Sun、J. Barry Danfa和M. Teplitskiy。双盲同行评审能减少偏见吗？来自一场顶尖计算机科学会议的证据。Journal of the Association for Information Science and Technology，73:811–819，2022年。[85] R. S´anchez-Jim´enez、P. Guerrero-Castillo、V. P. Guerrero-Bote、G. Halevi和F. De-Moya-Aneg´on。分析科学产出中作者署名的性别分布：全球视角。《信息计量学杂志》，18:101556，2024年。[86] E. G. Teich、J. Z. Kim、C. W. Lynn、S. C. Simon、A. A. Klishin、K. P. Szymula、P. Srivastava、L. C. Bassett、P. Zurn、J. D. Dworkin和D. S. Bassett。当代物理学中的引文不平等与性别化引文实践。《自然·物理学》，18:1161–1170，2022年。[87] A. Tekles、K. Auspurg和L. Bornmann。同性别引用并不表明存在显著的性别同质性偏见。《公共科学图书馆·综合》，17:e0274810，2022年。

[88] A. Llorens 等。学术界中的性别偏见：一个需要解决的终身问题。《神经元》，109:2047–2074，2021年。[89] 美国国家科学基金会。出版物产出：美国趋势与国际比较，2021年。https://ncses.nsf.gov/pubs/nsb20214/ [访问于3月 2025]。[90] 美国国家科学基金会。2024财年拨款，2024年。https://www.nsf.gov/ about/budget/fy2024/appropriations [访问于3月 2025]。[91] 澳大拉西亚计算研究与教育协会。CORE 2021年会议排名。https://portal.core.edu.au/conf-ranks/ [访问于4月 2024]。[92] DBLP团队。DBLP计算机科学书目。如何查找个人、出版物或数据流的标识符？2024年。https://dblp.org/faq/How+to+find+the+key+of+a+person+a+publication+or+a+stream.html [访问于12月 2024]。[93] DBLP团队。DBLP计算机科学书目。2024年4月月度快照发布，2024年。https://dblp.org/xml/release/dblp-2024-04-01.xml.gz [访问于4月2024]。[94] A. Tomkins、M. Zhang 和 W. D. Heavlin。单盲与双盲同行评审中的审稿人偏见。《美国国家科学院院刊》，114:12708–12713，2017年。[95] C.-F. Tsai。顶级计算机科学期刊的引文影响力分析及其排名。《信息计量学杂志》，8:318–328，2014年。[96] B. Uzzi、S. Mukherjee、M. Stringer 和 B. Jones。非典型组合与科学影响力。Science，342:468–472，2013年。[97] P. van den Besselaar 和 U. Sandstr¨om。科研表现中的性别差异及其对职业生涯的影响：一项纵向案例研究。《科学计量学》，106:143–162，2016年。[98] M. Y. Vardi。计算研究中的会议与期刊。Communications of the ACM，52(5)，2009年。[99] G. Vrettas 和 M. Sanderson。计算机科学领域的会议与期刊。Journal of the Association for Information Science and Technology，66:2674–2684，2015年。[100] L. L. Wang、G. Stanovsky、L. Weihs 和 O. Etzioni。计算机科学论文署名中的性别趋势。Communications of the ACM，64:78–84，2021年。[101] X. Wang、J. D. Dworkin、D. Zhou、J. Stiso、E. B. Falk、D. S. Bassett、P. Zurn 和 D. M. Lydon-Staley。传播学领域的性别化引文实践。《国际传播学会年刊》，45:134–153，2021年。[102] S. F. Way、D. B. Larremore 和 A. Clauset。计算机科学教职招聘网络中的性别、生产力与声望。载于国际万维网大会论文集，第1169–1179页，2016年。[103] H. O. Witteman、M. Hendricks、S. Straus 和 C. Tannenbaum。性别差距源于对申请者的评价还是对科学的评价？国家资助机构的一项自然实验。《柳叶刀》，393:531–540，2019年。

[104] Y. Xie 和 K. A. Shauman。科研生产力中的性别差异：关于一个老问题的新证据。《美国社会学评论》，63:847–870，1998年。[105] F. Yates。涉及小数值的列联表与 χ^2 检验。《皇家统计学会杂志增刊》，1:217–235，1934年。[106] A. Zeng、Z. Shen、J. Zhou、J. Wu、Y. Fan、Y. Wang 和 H. E. Stanley。科学学：复杂系统视角。《物理学报告》，714-715:1–73，2017年。[107] S. Zhou、S. Chai 和 R. B. Freeman。性别同质性：群体内引用偏好与性别劣势。《研究政策》，53:104895，2024年。

补充信息：
计算机科学中的系统性性别化引文失衡：来自会议和期刊的证据

Kazuki Nakajima、Yuya Sasaki、Sohei Tokuno 和 George Fletcher

S1 关于我们数据集的补充信息

S1.1 论文筛选流程图

图 S1 展示了论文筛选过程的流程图。

S1.2 对最终数据集代表性的分析

为检验最终数据集的代表性，我们在论文筛选过程中跟踪了四种性别类别（即“MM”“MW”“WM”和“WW”）所占的比例。我们发现，各类别的相对占比在整个过程中保持稳定：

- **按所属国家和性别类别划分的论文**（$N = 1{,}399{,}106$）：MM：75.2%，MW：7.4%，WM：10.7%，WW：6.6%
- **发表于排名会议和期刊的论文**（$N = 672{,}082$）：MM：76.0%，MW：7.3%，WM：10.5%，WW：6.1%
- **用于分析的最终集合**（$N = 394{,}432$）：MM：75.6%，MW：7.8%，WM：11.1%，WW：5.5%

我们还比较了初始匹配数据集 $\mathcal{D}$ 与最终数据集中发表场所类型和子领域的分布。最终数据集中会议论文的占比有所下降（从 51.8% 降至 39.3%），而期刊论文的占比有所上升。这一变化反映了我们的质量标准：仅纳入 CORE 或 SCImago 排名收录的发表场所；与期刊相比，初始匹配数据集中被排除的未排名会议更多。就子领域而言，筛选过程提高了 11 个计算机科学子领域的相对占比，同时降低了其他子领域的占比（从 51.9% 降至 41.6%；见表 S1）。这表明，我们的筛选使数据集更加集中于该学科的核心研究领域。

S1.3 纳入与排除作者的人口统计学比较

表 S2 比较了两组作者所属国家的分布：一组是已分配所属国家的作者，另一组是同时被分配所属国家和性别的作者。我们逐一列出各个国家，直到已分配所属国家的作者累计占比达到 90%，其余国家则归为“其他”。最明显的变化是来自中国的作者占比大幅下降。来自台湾、韩国、马来西亚、新加坡和泰国的作者占比也出现了类似下降。这些变化反映出这些地区的姓名具有较高的性别模糊性，因此往往无法达到我们的性别推断准确度阈值。

S1.4 论文元数据摘要

表 S3 总结了我们数据集中所纳入论文的元数据。

S1.5 补充统计数据

表 S4 显示了各子领域—性别类别组合中的论文数量。表 S5 显示了各发表 venue 排名—性别类别组合中的论文数量。图 S2 显示了各性别类别论文所引用和被引用文献的生存函数。

S2 优先抽样模型的伪代码

算法 S1 展示了优先抽样模型的伪代码。

算法 S1 优先抽样模型。

要求： 引文网络：$(V,\ E)$，以及论文特征集合：S。
确保： 随机化网络：$(V,\ E_{\text{rand}})$

按出版日期升序排列论文：v_{x_1}、……、v_{x_N}。
将 $c_{j,\ l,\ \text{PD}}$ 以及任意 $j = 1$、……、N和任意 $l = 1$、……、N初始化为零。
$E_{\text{rand}} \leftarrow$ 一个空列表。
对于 $l = 2, \ldots, N$ **do**
  对于每个引用 $(v_{x_l},\ v_j) \in E$ **执行**
    计算集合 $V_{\text{HD}}(x_l,\ j,\ S)$。
    $\overline{V}_{\text{PD}}(x_l,\ j,\ S) \leftarrow \{v_{j'} \mid v_{j'} \in \overline{V}_{\text{HD}}(x_l,\ j,\ S) \ \wedge \lfloor \ln(c_{j',\ l-1,\ \text{PD}} + 1) \rfloor = \lfloor \ln(c_{j,\ l-1,\ \text{PD}} + 1) \rfloor\}$。
    从集合 $V_{\text{PD}}(x_l,\ j,\ S)$ 中均匀随机抽取 $v_{j'}$。
    为 E_{rand} 添加引用（v_{x_l},$v_{j'}$）。
    对于 $z = l, \ldots, N$ **do**
      $c_{j',\ z,\ \text{PD}} \leftarrow c_{j',\ z,\ \text{PD}} + 1$。
返回 (V, E_{rand})

S3 过度引用/引用不足的详细结果

表 S6–S8 分别展示了随机抽取、同质性抽取和优先抽取模型下的过度/不足引用。表 S9 展示了所有论文对特定性别类别和特定子领域内论文的过度/不足引用。表 S10 展示了所有论文对特定性别类别和特定发表场所类型内论文的过度/不足引用。表 S11 展示了所有论文对特定性别类别和特定发表场所排名内论文的过度/不足引用。图 S3 展示了主文图 4 所示过度/不足引用值对应的 Z 分数随时间变化的趋势。

S4 自引对过度引用/引用不足的影响

我们构建了一个保留自引的引文网络，其中包含 418,090 篇论文和 820,258 次引用。请注意，如果施引论文 u 与被引论文 v的作者署名有任何重叠，则该引用被定义为自引。随后，我们计算了该网络中的过度引用/引用不足。表 S12 将这些数值与从排除自引的网络中计算出的数值进行了比较。

S5　孤立论文对过度引用/引用不足的影响

我们构建了一个保留孤立论文（即既不引用也未被引用的论文）的引文网络，其中包含 542,550 篇论文和 752,742 次引用。随后，我们计算了该网络中的过度引用/引用不足。表 S13 将这些数值与从排除孤立论文的网络中计算出的数值进行了比较。

S6 匹配实验的详细结果

表 S14–S17 分别展示了匹配实验的结果，涉及作者性别（MM 或 WW）、发表场所类型、知名作者的参与情况以及本地合著网络。

S7 论文三项特征之间的关系

我们考察三项特征之间的关系：发表场所类型、知名作者的参与情况以及本地合著网络。具体而言，我们分别对 MM 论文和 WW 论文使用耶茨卡方检验 [1]。固定一个性别类别 $g \in \{\text{MM}, \text{WW}\}$ 和两项特征 X 与 Y。对于每篇论文，每项特征都是具有两个可能类别的分类变量。我们用 $O_{g,\,i,\,j}$ 表示性别类别 g 中，特征 X 属于类别 i（$i = 1, 2$），且特征 Y 属于类别 j（$j = 1, 2$）的论文数量。令 $N_g = \sum_{i=1}^{2}\sum_{j=1}^{2} O_{g,\,i,\,j}$ 表示类别 g 中的论文总数。我们将检验统计量定义为

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}(|O_{g,i,j} - E_{g,i,j}| - 0.5)^2 / E_{g,i,j}, \tag{S2}$$

其中

$$E_{g,i,j} = N_g p_{g,i} q_{g,j}, \tag{S3}$$

$$p_{g,i} = \sum_{j=1}^{2} \frac{O_{g,i,j}}{N_g}, \tag{S4}$$

$$q_{g,j} = \sum_{i=1}^{2} \frac{O_{g,i,j}}{N_g}. \tag{S5}$$

零假设是，对于 g 中的论文，X 与 Y 相互独立。如果检验的 p 值小于 0.001，我们就拒绝零假设。

表 S18 展示了每种组合（g，X，Y）对应的检验统计量和 p 值。对于 MM 论文，每一对特征都不独立。对于 WW 论文，除“发表场所类型与局部合著网络”之外，每一对特征都不独立。

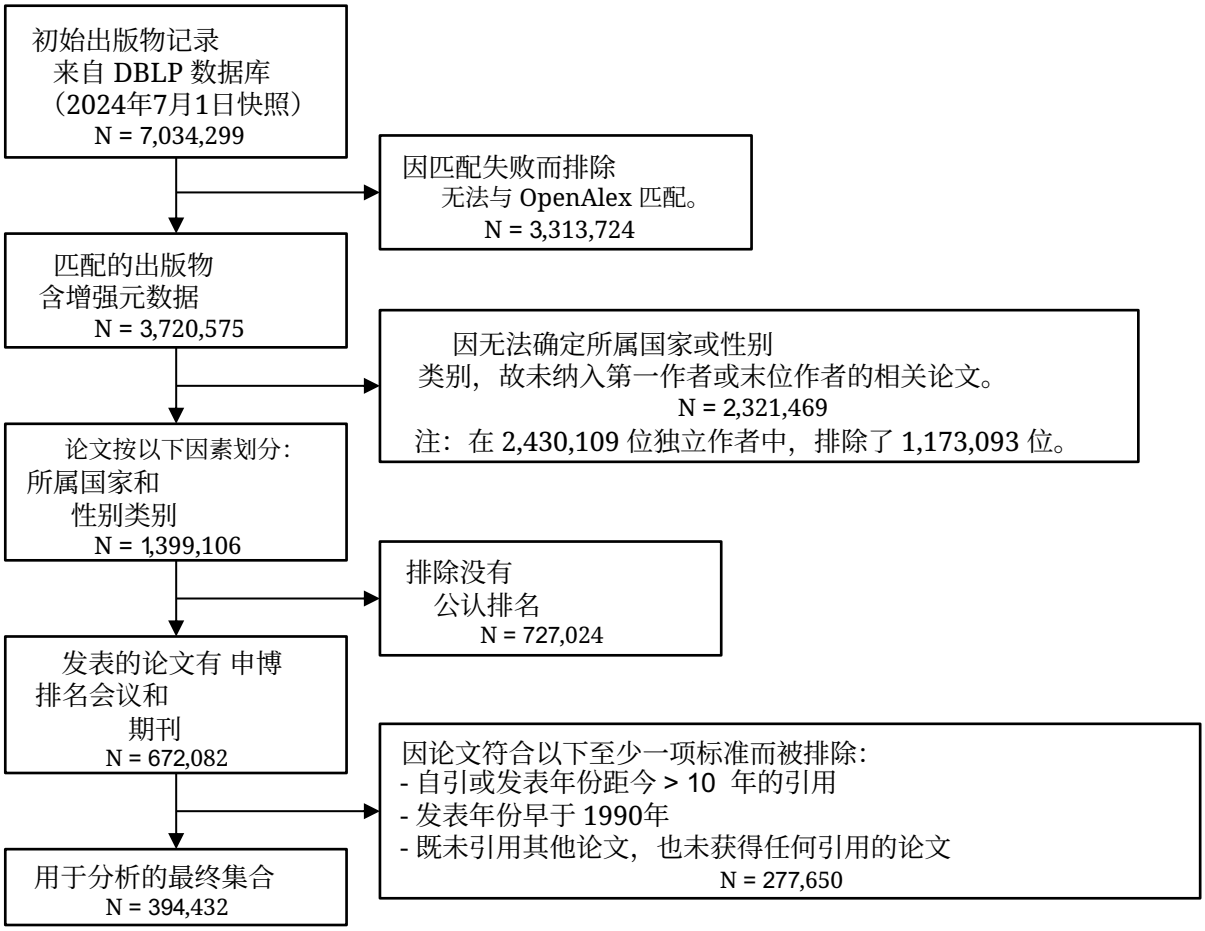

图 S1：论文筛选流程的流程图。该图展示了从 DBLP 数据库中的初始出版物记录逐步筛选至本研究分析所用论文最终集合的过程。每一步均标明了排除论文的原因和数量。我们还注明了因无法确定所属国家或性别而被排除的独立作者数量。

补充参考文献

[1] F. Yates. 涉及小数字的列联表及 χ^2 检验。《皇家统计学会杂志增刊》，1:217–235,1934年。

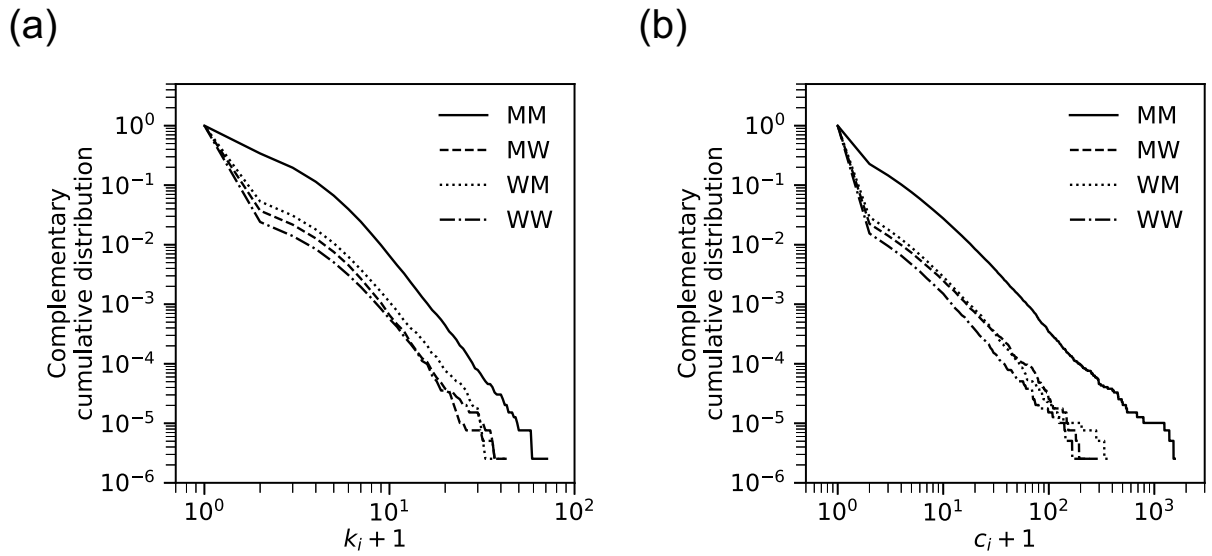

图 S2：各性别类别引用次数的生存函数。(a) 一篇论文引用的数量，记为 k_i，加一。(b) 一篇论文收到的引用数量，记为c_i，加一。

表 S1：初始匹配数据集与最终数据集的子领域分布比较。不属于 OpenAlex 中“计算机科学”领域的子领域归为“其他”。

子领域	初始集合（%）	最终集合（%）
人工智能	12.83	16.50
计算理论与数学	3.53	3.86
计算机图形学与计算机辅助设计	0.72	1.17
计算机网络与通信	9.76	11.63
计算机科学应用	0.87	1.35
计算机视觉与模式识别	8.25	9.04
硬件与体系结构	2.34	3.19
人机交互	1.10	1.71
信息系统	5.13	6.18
信号处理	2.78	2.66
软件	0.78	1.13
其他	51.92	41.59

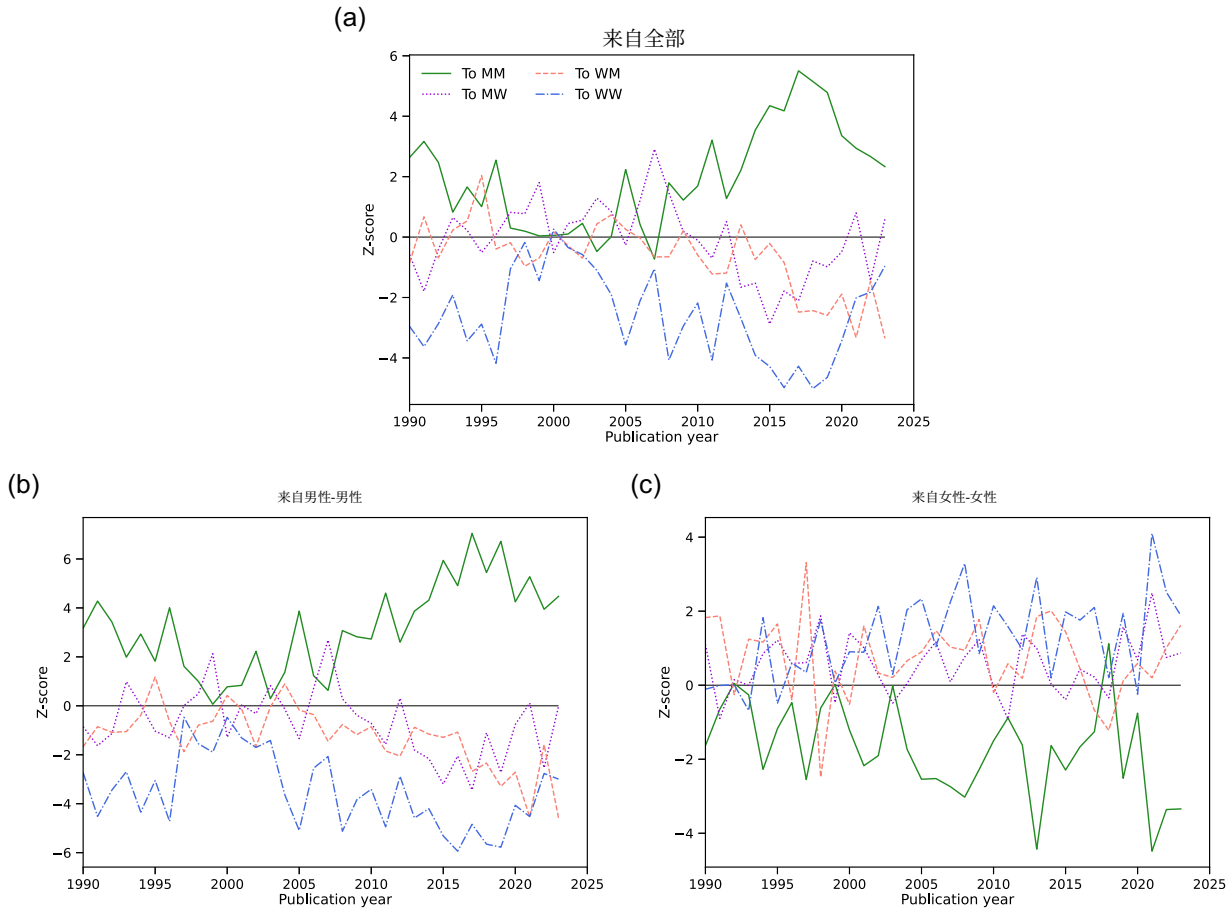

图 S3：计算机科学论文过度引用/引用不足的 Z 分数随时间的变化趋势。(a)：全部论文。(b)：男性-男性论文。(c)：女性-女性论文。

表 S2：对比被分配了国家（p_1）的作者与同时被分配了国家和性别（p_2）的作者所属国家分布。国家逐一列出，直到 p_1 的累计占比达到 90%，其余国家归为“其他”。

国家	p_1 (%)	p_2 (%)
美国	22.21	26.52
中国	16.22	1.77
德国	5.71	8.07
日本	4.98	6.00
印度	4.25	4.31
英国	4.22	5.20
韩国	3.41	2.30
意大利	3.15	4.10
台湾	2.95	0.56
加拿大	2.86	3.46
法国	2.67	3.55
巴西	2.21	3.19
澳大利亚	1.73	2.04
荷兰	1.50	1.76
Iran	1.26	1.73
西班牙	1.20	1.65
瑞士	0.96	1.30
希腊	0.84	1.13
瑞典	0.81	1.13
俄罗斯	0.81	0.90
马来西亚	0.79	0.68
比利时	0.73	0.99
奥地利	0.70	1.02
葡萄牙	0.70	1.02
芬兰	0.69	0.88
新加坡	0.68	0.44
以色列	0.66	0.73
巴基斯坦	0.58	0.79
泰国	0.49	0.37
丹麦	0.48	0.67
其他	9.55	11.74

表 S3：论文元数据摘要。

属性	数据来源
标题	OpenAlex
出版日期	OpenAlex
主要研究主题	OpenAlex
主要研究子领域	OpenAlex
作者姓名	OpenAlex
作者所属机构	OpenAlex
作者 ID	DBLP
作者署名顺序	DBLP
该论文引用的论文	OpenAlex
发表场所类型（即会议或期刊）	DBLP
发表场所名称	DBLP
发表场所排名	DBLP、CORE 和 SCImago
所属国家	OpenAlex 和 DBLP
性别类别	OpenAlex、DBLP、Gender API

表 S4：按子领域和性别类别划分的论文数量。

子领域	MM	MW	WM	WW
人工智能	49,852 (76.7%)	5,156 (7.9%)	6,653 (10.2%)	3,413 (5.2%)
计算理论与数学	12,383 (81.2%)	938 (6.2%)	1,224 (8.0%)	698 (4.6%)
计算机图形学与计算机辅助设计	4,039 (87.8%)	205 (4.5%)	261 (5.7%)	92 (2.0%)
计算机网络与通信	36,979 (80.5%)	3,102 (6.8%)	4,425 (9.7%)	1,362 (3.0%)
计算机科学应用	3,107 (58.5%)	571 (10.8%)	866 (16.3%)	764 (14.4%)
计算机视觉与模式识别	28,040 (78.6%)	2,552 (7.2%)	3,918 (11.0%)	1,135 (3.2%)
硬件与体系结构	10,790 (85.6%)	733 (5.8%)	851 (6.8%)	227 (1.8%)
人机交互	4,168 (62.0%)	730 (10.8%)	1,085 (16.1%)	743 (11.1%)
信息系统	17,497 (71.7%)	2,091 (8.6%)	3,081 (12.6%)	1,725 (7.1%)
信号处理	8,266 (78.9%)	750 (7.2%)	1,136 (10.8%)	323 (3.1%)
软件	3,321 (74.5%)	408 (9.2%)	461 (10.3%)	265 (6.0%)

表 S5：按发表场所排名和性别类别划分的论文数量。

排名 男性-男性		MW	WM	WW
A*	32,325 (79.3%)	3,062 (7.5%)	3,692 (9.1%)	1,680 (4.1%)
A	28,020 (75.8%)	3,064 (8.3%)	3,851 (10.4%)	2,045 (5.5%)
B	39,782 (76.4%)	4,283 (8.2%)	5,943 (11.4%)	2,091 (4.0%)
C	18,261 (73.1%)	2,101 (8.4%)	3,149 (12.6%)	1,475 (5.9%)
Q1	102,895 (75.9%)	10,381 (7.7%)	15,173 (11.2%)	7,068 (5.2%)
Q2	49,132 (72.7%)	5,509 (8.2%)	8,010 (11.8%)	4,962 (7.3%)
Q3	23,105 (76.0%)	2,134 (7.0%)	3,239 (10.7%)	1,909 (6.3%)
Q4	4,543 (74.5%)	436 (7.2%)	670 (11.0%)	442 (7.3%)

表 S6：根据随机抽样模型得出的过度引用/引用不足。我们展示一组论文（例如，属于某一性别类别的论文或全部论文）对另一组论文（同样按性别类别划分）的过度引用/引用不足情况。此处，n_{obs} 表示原始网络中观察到的引用次数，而 μ 和 σ 分别表示在 100 个随机化网络中计算得出的引用次数的均值和标准差。

作者	接收	n_{obs}	μ	σ	过度引用/引用不足引用（%）	Z 分数	p 值
MM	MM	454733	438991.5	304.9	3.6	51.63	< 0.001
MM	MW	38826	40627.3	179.8	-4.4	-10.02	< 0.001
MM	WM	47068	53171.2	234.7	-11.5	-26.01	< 0.001
MM	WW	21853	29689.9	170.9	-26.4	-45.86	< 0.001
MW	MM	45992	47472.8	106.1	-3.1	-13.96	< 0.001
MW	MW	5282	4594.3	64.7	15.0	10.64	< 0.001
MW	WM	6486	6157.7	72.9	5.3	4.50	< 0.001
MW	WW	3688	3223.2	47.2	14.4	9.85	< 0.001
WM	MM	64414	67369.8	131.7	-4.4	-22.44	< 0.001
WM	MW	7692	6777.6	89.4	13.5	10.23	< 0.001
WM	WM	10244	9189.0	94.3	11.5	11.19	< 0.001
WM	WW	5586	4599.6	65.3	21.4	15.10	< 0.001
WW	MM	27220	31674.4	86.4	-14.1	-51.54	< 0.001
WW	MW	3783	3007.8	55.2	25.8	14.06	< 0.001
WW	WM	5115	4021.0	62.6	27.2	17.48	< 0.001
WW	WW	4760	2174.8	41.4	118.9	62.46	< 0.001
All	MM	592359	585508.6	357.2	1.2	19.18	< 0.001
All	MW	55583	55006.9	217.8	1.0	2.65	0.004
All	WM	68913	72539.0	279.9	-5.0	-12.95	< 0.001
All	WW	35887	39687.6	202.0	-9.6	-18.82	< 0.001

表 S7：基于同质性抽样模型得出的过度引用/引用不足。

作者	接收	n_{obs}	μ	σ	过度引用/引用（%）	Z 分数	p 值
MM	MM	454733	447203.8	255.1	1.7	29.51	< 0.001
MM	MW	38826	40175.6	179.5	-3.4	-7.52	< 0.001
MM	WM	47068	50282.7	195.5	-6.4	-16.44	< 0.001
MM	WW	21853	24817.9	132.2	-11.9	-22.43	< 0.001
MW	MM	45992	45988.4	87.1	0.0	0.04	0.483
MW	MW	5282	5099.5	60.2	3.6	3.03	0.001
MW	WM	6486	6662.1	58.0	-2.6	-3.03	0.001
MW	WW	3688	3698.0	49.8	-0.3	-0.20	0.420
WM	MM	64414	64612.6	101.1	-0.3	-1.97	0.025
WM	MW	7692	7556.1	73.0	1.8	1.86	0.031
WM	WM	10244	10208.7	82.0	0.3	0.43	0.334
WM	WW	5586	5558.5	59.0	0.5	0.47	0.321
WW	MM	27220	28031.5	71.1	-2.9	-11.41	< 0.001
WW	MW	3783	3661.3	43.5	3.3	2.80	0.003
WW	WM	5115	4953.4	51.9	3.3	3.11	< 0.001
WW	WW	4760	4231.7	52.0	12.5	10.16	< 0.001
All	MM	592359	585836.3	324.6	1.1	20.10	< 0.001
All	MW	55583	56492.6	217.7	-1.6	-4.18	< 0.001
All	WM	68913	72106.9	242.3	-4.4	-13.18	< 0.001
All	WW	35887	38306.2	170.5	-6.3	-14.19	< 0.001

表 S8：基于优先抽样模型得出的过度引用/引用不足。

作者	接收	n_{obs}	μ	σ	过度引用/引用（%）	Z 分数	p 值
MM	MM	454733	449454.7	341.7	1.2	15.45	< 0.001
MM	MW	38826	39613.1	236.0	-2.0	-3.34	< 0.001
MM	WM	47068	48780.3	233.9	-3.5	-7.32	< 0.001
MM	WW	21853	24631.9	169.4	-11.3	-16.40	< 0.001
MW	MM	45992	46142.0	86.3	-0.3	-1.74	0.041
MW	MW	5282	5076.2	60.2	4.1	3.42	< 0.001
MW	WM	6486	6506.2	61.2	-0.3	-0.33	0.371
MW	WW	3688	3723.6	48.6	-1.0	-0.73	0.232
WM	MM	64414	64799.4	111.4	-0.6	-3.46	< 0.001
WM	MW	7692	7507.8	69.1	2.5	2.67	0.004
WM	WM	10244	10029.6	85.6	2.1	2.50	0.006
WM	WW	5586	5599.2	61.9	-0.2	-0.21	0.416
WW	MM	27220	27979.6	71.7	-2.7	-10.60	< 0.001
WW	MW	3783	3644.2	51.2	3.8	2.71	0.003
WW	WM	5115	4916.9	62.0	4.0	3.20	< 0.001
WW	WW	4760	4337.2	54.6	9.7	7.74	< 0.001
All	MM	592359	588375.6	427.6	0.7	9.32	< 0.001
All	MW	55583	55841.4	302.3	-0.5	-0.85	0.196
All	WM	68913	70233.1	288.2	-1.9	-4.58	< 0.001
All	WW	35887	38291.9	217.5	-6.3	-11.06	< 0.001

表 S9：全部论文对不同性别类别和子领域的论文的过度引用/引用不足。

子领域	性别类别	n_{obs}	μ	σ	过度引用/引用不足引用（%）	Z 分数	p 值
人工智能	MM	115843	113620.9	214.7	2.0	10.35	< 0.001
	WW	6324	7000.1	125.6	−9.7	−5.38	< 0.001
计算理论与数学	MM	21814	21427.3	57.6	1.8	6.71	< 0.001
	WW	951	1077.9	32.2	−11.8	−3.94	< 0.001
计算机图形学与计算机辅助设计	MM	11607	11426.6	34.7	1.6	5.20	< 0.001
	WW	158	223.2	17.1	−29.2	−3.80	< 0.001
计算机网络与通信	MM	79513	80036.4	145.9	−0.7	−3.59	< 0.001
	WW	2486	2653.5	73.6	−6.3	−2.28	0.011
计算机科学应用	MM	5980	6195.5	49.6	−3.5	−4.35	< 0.001
	WW	1407	1352.8	37.6	4.0	1.44	0.075
计算机视觉与模式识别	MM	69467	68487.4	176.9	1.4	5.54	< 0.001
	WW	1787	2208.7	72.6	−19.1	−5.81	< 0.001
硬件与体系结构	MM	24161	24200.2	55.3	−0.2	−0.71	0.239
	WW	517	588.2	22.8	−12.1	−3.12	< 0.001
人机交互交互	MM	9489	9235.9	56.5	2.7	4.48	< 0.001
	WW	1412	1535.1	47.4	−8.0	−2.60	0.005
信息系统	MM	39409	39386.1	104.3	0.1	0.22	0.413
	WW	3297	3539.7	61.9	−6.9	−3.92	< 0.001
信号处理	MM	14525	14405.2	55.9	0.8	2.14	0.016
	WW	376	422.3	19.5	−11.0	−2.37	0.009
软件	MM	7318	7336.6	48.5	−0.3	−0.38	0.351
	WW	464	677.9	25.0	−31.6	−8.56	< 0.001

表 S10：全部论文对不同性别类别和发表场所类型论文的过度引用/引用不足。

Type	性别类别	n_{obs}	μ	σ	过度引用/引用（%）	Z 分数	p 值
会议	MM	221681	219533.3	226.9	1.0	9.47	< 0.001
	WW	11520	13012.1	117.6	−11.5	−12.69	< 0.001
期刊	MM	370678	368820.3	316.5	0.5	5.87	< 0.001
	WW	24367	25325.7	148.2	−3.8	−6.47	< 0.001

表 S11：全部论文对不同性别类别和子领域论文的过度引用/引用不足。

Rank	性别类别	n_{obs}	μ	σ	过度引用/引用（%）	Z 分数	p 值
A*	MM	130750	129819.0	252.5	0.7	3.69	< 0.001
	WW	4391	5360.3	97.8	-18.1	-9.91	< 0.001
A	MM	55092	54469.0	150.5	1.1	4.14	< 0.001
	WW	3667	3759.5	82.0	-2.5	-1.13	0.130
B	MM	37248	36934.5	82.7	0.8	3.79	< 0.001
	WW	1633	1814.0	33.8	-10.0	-5.35	< 0.001
C	MM	14350	14169.3	54.6	1.3	3.31	< 0.001
	WW	1163	1184.6	28.0	-1.8	-0.77	0.220
Q1	MM	257718	256007.4	263.3	0.7	6.50	< 0.001
	WW	16138	17130.2	153.6	-5.8	-6.46	< 0.001
Q2	MM	70613	70598.2	125.3	0.0	0.12	0.453
	WW	6768	6748.6	83.5	0.3	0.23	0.408
Q3	MM	23349	23139.9	61.1	0.9	3.42	< 0.001
	WW	1827	1993.2	38.4	-8.3	-4.33	< 0.001
Q4	MM	3239	3238.2	16.6	0.0	0.05	0.481
	WW	300	301.6	9.1	-0.5	-0.18	0.430

表 S12：C　比较有自引和无自引的网络中过度/不足引用情况。

		有自引			无自引		
作者	接收	过度/不足引用（%）	Z 分数	p 值	过度/不足引用（%）	Z 分数	p 值
MM	MM	2.1	33.14	< 0.001	1.2	14.47	< 0.001
MM	MW	-6.2	-12.22	< 0.001	-2.0	-3.48	< 0.001
MM	WM	-7.4	-17.19	< 0.001	-3.5	-7.03	< 0.001
MM	WW	-15.8	-22.24	< 0.001	-11.4	-17.10	< 0.001
MW	MM	-3.3	-16.06	< 0.001	-0.3	-1.68	0.046
MW	MW	24.7	22.34	< 0.001	4.2	3.82	< 0.001
MW	WM	-0.6	-0.58	0.280	-0.2	-0.18	0.429
MW	WW	2.0	1.62	0.053	-1.5	-1.25	0.105
WM	MM	-3.1	-17.21	< 0.001	-0.6	-3.18	< 0.001
WM	MW	2.7	2.79	0.003	2.1	2.05	0.020
WM	WM	15.1	20.71	< 0.001	2.2	2.45	0.007
WM	WW	2.2	2.10	0.018	-0.2	-0.16	0.438
WW	MM	-6.4	-23.10	< 0.001	-2.7	-11.76	< 0.001
WW	MW	0.2	0.15	0.442	3.8	3.14	< 0.001
WW	WM	1.1	0.90	0.183	4.2	4.00	< 0.001
WW	WW	31.1	33.38	< 0.001	9.6	8.65	< 0.001
All	MM	0.8	11.50	< 0.001	0.7	8.90	< 0.001
All	MW	-1.4	-2.90	0.002	-0.5	-1.02	0.153
All	WM	-2.7	-6.97	< 0.001	-1.8	-4.33	< 0.001
All	WW	-5.0	-8.51	< 0.001	-6.4	-11.12	< 0.001

表 S13：有无孤立论文的网络中过度/不足引用的比较。

		有孤立论文			无孤立论文		
作者	接收	过度/不足-引用（%）	Z 分数	p 值	高于/低于-引用（%）	Z 分数	p 值
MM	MM	1.2	15.45	< 0.001	1.2	14.47	< 0.001
MM	MW	-1.4	-2.58	0.005	-2.0	-3.48	< 0.001
MM	WM	-3.7	-7.57	< 0.001	-3.5	-7.03	< 0.001
MM	WW	-12.5	-18.05	< 0.001	-11.4	-17.10	< 0.001
MW	MM	-0.4	-1.92	0.028	-0.3	-1.68	0.046
MW	MW	4.8	4.05	< 0.001	4.2	3.82	< 0.001
MW	WM	-0.0	-0.02	0.493	-0.2	-0.18	0.429
MW	WW	-1.9	-1.35	0.088	-1.5	-1.25	0.105
WM	MM	-0.7	-4.40	< 0.001	-0.6	-3.18	< 0.001
WM	MW	3.1	3.64	< 0.001	2.1	2.05	0.020
WM	WM	2.5	3.35	< 0.001	2.2	2.45	0.007
WM	WW	-0.7	-0.60	0.276	-0.2	-0.16	0.438
WW	MM	-3.1	-10.56	< 0.001	-2.7	-11.76	< 0.001
WW	MW	5.0	3.48	< 0.001	3.8	3.14	< 0.001
WW	WM	4.9	4.05	< 0.001	4.2	4.00	< 0.001
WW	WW	10.3	8.59	< 0.001	9.6	8.65	< 0.001
Any	MM	0.7	9.18	< 0.001	0.7	8.90	< 0.001
Any	MW	0.2	0.40	0.343	-0.5	-1.02	0.153
Any	WM	-1.9	-4.59	< 0.001	-1.8	-4.33	< 0.001
Any	WW	-7.2	-12.13	< 0.001	-6.4	-11.12	< 0.001

表 S14：C　MM 论文与 WW 论文的过度引用/引用不足比较　　.

接收	作者 男性-男性	作者 女性-女性	t-统计量	p 值
MM	1.24 ± 0.18	-2.27	194.7	< 0.001
MW	-1.13 ± 1.33	5.62	-50.6	< 0.001
WM	-2.40 ± 1.16	3.82	-53.6	< 0.001
WW	-10.69 ± 1.31	7.78	-141.1	< 0.001

表 S15：比较会议论文与期刊论文对男性-男性和女性-女性论文的过度引用/引用不足。

作者	接收	会议	期刊	t-统计量	p 值
MM	MM	1.63	1.10 ± 0.04	−117.8	< 0.001
	MW	−2.30	−3.14 ± 0.35	−24.1	< 0.001
	WM	−6.06	−3.84 ± 0.33	66.9	< 0.001
	WW	−18.30	−9.52 ± 0.45	194.2	< 0.001
MM	MM	−2.74	−2.79 ± 0.29	−1.6	0.114
	MW	2.09	7.63 ± 1.86	29.8	< 0.001
	WM	3.03	−0.43 ± 1.49	−23.2	< 0.001
	WW	13.59	11.68 ± 1.41	−13.5	< 0.001

表 S16：比较知名作者论文与非知名作者论文对男性-男性和女性-女性论文的过度引用/引用不足。

作者	接收	With 知名 作者	不含 知名 作者	t-统计量	p 值
MM	MM	1.56	1.11 ± 0.10	−44.2	< 0.001
	MW	−3.86	−2.04 ± 0.70	26.0	< 0.001
	WM	−5.10	−4.36 ± 0.73	10.3	< 0.001
	WW	−19.17	−11.55 ± 0.95	79.9	< 0.001
WW	MM	−2.50	−4.97 ± 0.63	−39.4	< 0.001
	MW	17.96	8.61 ± 3.30	−28.3	< 0.001
	WM	4.97	7.74 ± 3.30	8.4	< 0.001
	WW	−2.77	18.66 ± 3.01	71.2	< 0.001

表S17：比较 MM 和 WW 论文中，$\mathrm{MA_{or}}$ 值较高与 $\mathrm{MA_{or}}$ 值较低的论文所产生的过度引用/引用不足。

作者	接收	上半部分 $\mathrm{MA_{or}}$	下半部分 $\mathrm{MA_{or}}$	t-统计量	p 值
MM	MM	1.11	1.66 ± 0.04	122.6	< 0.001
	MW	−0.97	−5.13 ± 0.37	−113.1	< 0.001
	WM	−4.08	−6.35 ± 0.34	−67.3	< 0.001
	WW	−10.87	−15.87 ± 0.47	−106.0	< 0.001
MM	MM	−3.56	−1.87 ± 0.42	40.6	< 0.001
	MW	5.02	5.52 ± 2.47	2.0	0.044
	WM	2.56	−3.34 ± 1.86	−31.7	< 0.001
	WW	13.51	16.10 ± 1.64	15.7	< 0.001

表 S18：论文的三个分类变量之间关系的统计显著性。

性别类别	特征 X	特征 Y	χ2-统计量	p 值
MM	发表场所类型	知名作者的参与	4069.5	< 0.001
	发表场所类型	本地合作署名网络	568.8	< 0.001
	知名作者的参与	本地合作署名网络	924.2	< 0.001
WW	发表场所类型	知名作者的参与	270.1	< 0.001
	发表场所类型	本地合作署名网络	2.5	0.114
	知名作者的参与	本地合作署名网络	320.7	< 0.001