arXiv:2302.13054v1 [physics.soc-ph] 25 Feb 2023

绘制科学知识景观中的流动模式

ChakreshKumar Singh[1,2],Liubov Tupikina[2,3], FabriceL´ecuyer[4],MicheleStarnini[5,6] 和 Marc Santolini[1,2*]1巴黎西岱大学´´，法国国家健康与医学研究院，系统工程与进化动力学，F-75004，法国巴黎。2LearningPlanet研究所，F-75004，法国巴黎。3诺基亚贝尔实验室，法国。4索邦大学，法国国家科学研究中心，LIP6，F-75005，法国巴黎。5 CENTAI研究所，意大利都灵。6加泰罗尼亚理工大学物理系，北校区，08034，西班牙。

通讯作者。电子邮箱：
marc.santolini@cri-paris.org；

摘要

从小步迈进到大步飞跃，用空间流动性来比喻发现过程的说法比比皆是。在这里，我们通过系统研究科学知识流动模式，以形式化的方式阐释这些观点。我们运用低维嵌入技术，构建一个知识空间，其中包含来自物理学、计算机科学和数学领域的150万篇文章。通过分析个别研究人员的发表历程，我们发现知识流动模式与物理流动极为相似。总体而言，这些轨迹形成了可用引力模型描述的流动流：跳跃更可能发生在密度较高的区域，而跨越更长距离的跳跃则较少发生。根据研究人员各自的流动模式，我们识别出两类研究人员：开创新领域的跨学科研究人员 *explorers* ，以及 *exploiters* 更可能留在各自专业领域内的研究人员。我们的结果表明，空间流动分析是理解知识演化的一种有价值的工具。

关键词： 科学学、人类流动、社会动力学、知识探索

引言

量化知识演化对于理解过去并预测未来创新至关重要 [1], ，而创新最终推动社会进步。处于科学创新前沿的是那些重新组合思想、拓展已知边界的研究人员 [2, 3]。随着作者和出版物数量呈指数级增长 [4, 5], ，我们需要新的方法来呈现知识发展并提供相关洞见。

对大规模出版物数据集日益广泛的获取，为量化研究人员所作的选择并考察支配知识演化的因素提供了机会。通过研究研究人员在其出版物中的引用模式，相关研究测量了相互冲突的思想如何在研究人员之间得到追求，直至它们趋于共同共识 [6]或让位于新思想 [7]。其他研究则着重于识别研究中的“热门主题” [8],、量化知识流动模式 [9] 以及知识演化中的记忆效应 [10, 11], ，或预测研究人员最终的影响力 [12, 13]。同样，还可以利用出版物中的关键词和短语来追踪科学思想和领域的演化 [14, 15] ，或量化科学家如何随时间选择并调整其研究重点 [16, 17]。例如，美国物理学会发表的文章所采用的物理学与天文学分类方案（PACS），可用于研究探索研究领域边界与利用既有工作之间的“本质张力” [18]。最后，研究人员之间的科学荣誉及其相互之间的科学兴趣（分别通过论文之间的引文和关键词量化）可以结合起来，以改进对新的科学合作的预测 [19]。

因此，研究研究人员的出版轨迹，有助于识别知识演化背后多层面且复杂的过程。人们常以隐喻的方式谈论此类轨迹，例如将某些科学进展称为“巨大飞跃” [20]。在此，我们旨在借鉴人类流动性研究的见解，更正式地探索科学活动与人类流动之间的相似之处。以大规模真实人类轨迹数据为基础，先前的研究确实揭示了人类流动背后的若干规律。尽管人类的移动存在异质性，其流动仍呈现出反复出现的模式 [21, 22]。研究表明，这些模式会产生旅行距离分布的标度定律[23]。在宏观层面，两个地点之间由此产生的流动遵循引力模型 [24],，模拟了牛顿万有引力定律，即相隔一定距离的两个物体之间的吸引力。除跳跃距离之外，个体在整个轨迹层面也表现出可重复的特性。例如，根据个体返回同一地点或探索新地点的倾向，可将其分为两类：回归者 和探索者 [25]。更广泛地说，针对个体和集体流动数据集的研究提出了各种定量模型，用以解释人类流动动力学[23, 26–30]。重要的是，这些可重复的模式并非人类独有

流动 [31]。跨学科的多项研究发现，地理空间中的人类流动与动物觅食之间存在显著相似之处 [32],昆虫群体 [33],抽象环境中的搜索方法，例如记忆空间 [34], 组织学习 [35],和网络空间 [36–38]。

在知识演化的背景下，如今已有大量工具和数据来源可用于空间表示。自然语言处理和嵌入方法可结合出版物的元数据（如引文、关键词或摘要），利用研究出版物之间的相似性，生成科学景观的低维表示。这类表示已被用于量化研究人员探索的思想的认知范围 [39, 40] 以及期刊的结构分解 [41], ，从而深入了解知识的结构。

在这项研究中，我们将这些直觉转化为一个定量框架，用于表示科学知识，并借鉴人类流动的指标来描述研究轨迹。我们利用 arXiv预印本的出版物元数据，通过嵌入方法构建低维知识空间[42]。随后，我们根据消歧后的作者的出版记录，追踪他们在这一空间中的流动。我们发现，知识探索与人类在物理空间中的流动有着惊人的相似之处。首先，我们表明，科学知识空间中的流动遵循引力模型：跳跃更可能发生在密度较高的区域，而距离越远，跳跃发生的可能性越低。其次，我们发现，知识探索存在两种截然不同的类型：跨学科的科学探索者 更可能推动变革并开辟新领域，而开发者则倾向于深耕特定专业领域；这与空间流动中探索者和回归者之间的差异如出一辙。最后，我们探讨知识流动性分析在科学与创新研究中的用途，并讨论其局限性以及对未来研究的启示。

结果

arXiv知识空间中的科学轨迹

为了构建知识空间，我们利用 arXiv数据集，其中包含 1，456，403 篇于 1992年至2018年间在线发表的科学文章（参见方法、图 S1 和参考文献 [43]）。我们对该数据集的兴趣有两个方面。首先，它拥有清晰且稳定的领域标签本体，作者用这些标签来注明其文章涉及的相关研究领域。作者有充分的动机尽可能准确地标注这些标签，以便文章能够出现在目标科学共同体检索的正确 arXiv分区中，也能被收录在感兴趣的科学家可以订阅的相关每日电子邮件摘要中。其次，作为一个预印本服务器，它没有编辑门槛或出版成本，因此发表门槛较低。这使我们能够以精细的粒度追踪作者的出版历史，以作者认定文章完成之时为准，且不受文章被认为具有何种新颖性的影响。

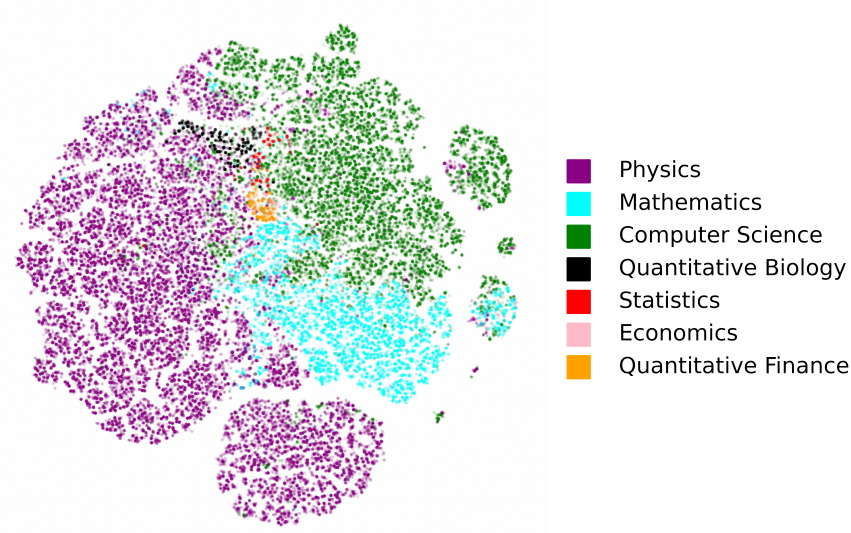

图1 知识空间的构建。我们使用了发布在 arXiv 上的 145 万篇文章的元数据，包括对应的文章领域标签、作者和时间戳。我们构建了一个高维的 175 维空间，每篇文章都通过对应于正交维度的领域标签唯一映射到该空间中。最后，我们使用 t-SNE算法将这个高维空间嵌入二维知识空间中。每个点代表一篇文章。颜色对应于 arXiv 中的主要学术领域，依据是文章的第一个标签（即主要领域）。

因此，arXiv预印本可以被视为以高分辨率追踪知识的步进，而无需满足新颖性阈值要求。

我们首先构建 arXiv预印本所形成的知识空间的空间表示。这个空间的结构由投稿作者用于为文章指定科学子领域的 175 个标签决定。文章可以被指定一个或多个标签。例如，一篇文章可以标记为社会与信息网络（cs.si）和物理学与社会（physics.soc-ph）。因此，一篇文章可以表示为高维 175 个子领域空间中的一个二进制向量X =(0，0，1，0，1，...，0)，其中$X_i = 1$ 表示如果该文章被指定了对应于科学领域 i 的标签。

由于文章很少会结合多个标签（见图S2），知识空间中的数据分布较为稀疏。此外，某些标签经常同时出现，造成信息冗余 [43]。基于这些观察，我们通过 t-SNE 算法将初始空间嵌入低维空间，从而降低其维度 [44, 45] （见方法）。本研究聚焦于二维嵌入，以便与人类地理流动的传统研究相对应。此外，我们在方法部分讨论了使用其他嵌入方法时结果的稳定性。图1展示了得到的知识空间，其中的文章以点表示，并按照其主要（第一个）领域标签着色。我们观察到，arXiv 中主要学科的文章聚集成界限清晰、各具特色的

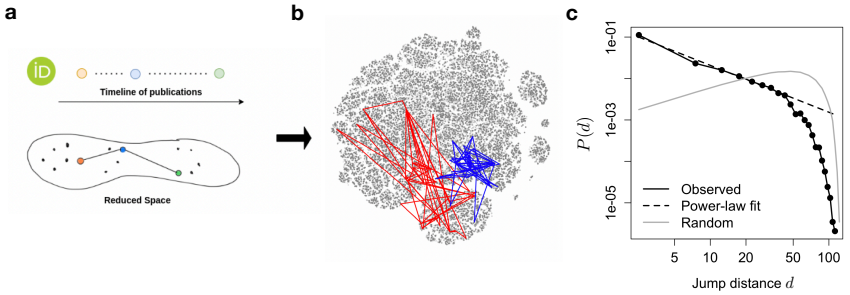

图 2 知识空间中的科学流动。a. 每位研究人员的 arXiv 预印本序列标示出其在知识空间中的独特科学轨迹。**b.** 我们展示了两条示例轨迹，呈现研究人员的不同行为。**c.** 至少发表 10 篇出版物的作者连续跳跃距离分布（$N = 11,826$）。黑色虚线是视觉参考线，表示幂律行为。灰线表示这样得到的跳跃距离分布：对每位作者，从所有可能访问过的地点中随机选择地点。

空间区域，跨学科领域（如定量生物学（q-bio）或定量金融学（q-fin））则位于相关学科之间的交界处。

作者发表文章的时间顺序构成了知识空间中的一系列地点，描绘出他们的 科学轨迹（图2a-b）。为获得高质量的轨迹，我们从 50，402 位经过消歧处理的研究人员中选取了 11，826 位样本；这些研究人员均拥有唯一的 ORCID 标识符，且至少发表过 10 篇文章。在一条轨迹中，连续两篇文章构成一次跳跃，其长度等于在嵌入空间中计算出的（欧几里得）距离，持续时间则等于两篇文章发表日期之间相隔的天数。如果作者在空间中的所有可能地点之间随机跳跃，那么根据纯扩散过程，跳跃分布将围绕某个较大的典型步长呈有界分布（图 2c，灰线）。但图 2c 显示，跳跃距离分布符合幂律函数形式，并因空间大小有限而在大距离处出现截断，这与扩散过程显著不同。重要的是，我们发现这一特征对于不同的嵌入技术均具有稳健性，见图 S4。这表明，尽管大多数跳跃距离较短，研究人员的活动范围围绕着某个相对集中的研究兴趣，但仍有一小部分跳跃延伸至知识空间的远处，意味着研究人员跨越了不同领域。下一节中，我们将探讨简单的人类流动模型是否能够解释观察到的行为。

科学流动的引力模型

观测到的跳跃距离分布呈现胖尾形式（带有截断），这让人联想到在人类流动中两个地点之间观察到的距离反比关系。这一观察催生了一种简单直观的

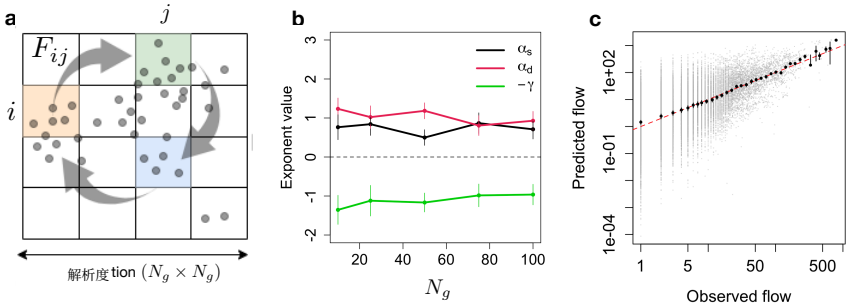

图 3 科学流动的引力模型。a. 将投影空间划分为网格后，我们可以计算聚合密度，并研究其对网格元素之间流动模式的影响。网格分辨率由任一维度上的单元格数量 N_g 固定。 **b.** 根据公式（1）拟合不同网格分辨率级别 $N_g = 10$、25、50、75、100 对应的指数。 **c.** 在网格尺寸分辨率为 $N_g = 10$ 时，对比预测与观测到的流动流。

人类流动性研究模型，即引力模型，其中两个地点 i 和 j 之间的通量 F_{ij}与 i 和 j 处的人口规模成正比，与两地之间的距离 d_{ij} 成反比。关于空间分布模型和城市建模的早期研究 [46–48] 表明，此类模型可根据统计力学的原理以及齐普夫定律等经验定律在功能上推导得出（引力定律的推导已注释掉）[49]。在像我们这样的起讫地设置中考察人口尺度流动时，引力模型自然会作为使两个地点之间流动熵最大化的分布的期望而出现[30]。

与城市和乡村景观类似，人口聚集在少数几个密集区域，即城市区域；低维知识空间中也存在较为密集的区域，对应于研究较多的领域。然而，与城市和行政区域不同，知识空间的边界并没有明确定义。在此，我们采用简单的盒式容器模型，将尺寸为 $N_g \times N_g$ 的网格覆盖整个知识空间，其中 N_g 是一个量化分辨率级别的参数，人口计数按网格进行聚合（图 3a）。接着，我们定义一个引力模型来预测观测到的流动 F_{ij} ，即在知识空间中两个网格地点 i 和 j 之间的流动，其定义为科学家在 5 年滚动时间窗口内从网格地点 i 跳转到地点 j 的人数：

$$\tilde{F}_{ij} = G\frac{V_i^{\alpha_s} V_j^{\alpha_d}}{d_{ij}^{\gamma}}, \tag{1}$$

其中， $\tilde{F}_{ij}$ 是地点 i 与 j之间的预测流量， G 是归一化常数， d_{ij} 是地点 i 与 j之间的距离，而 V_i 和 V_j（访问次数）分别是在地点 i发表过文章的作者人数

以及 j 在过去5年中的人数。指数 α_s、 α_d和 γ 引入了非线性缩放，例如访问次数中的拥挤效应；密度越高，流量的增长越可能是次线性（$\alpha < 1$）或超线性（$\alpha > 1$）。

图 3b 展示了通过将式（1）拟合到经验流 F_{ij} 在不同分辨率级别下的流量时得到的指数值 N_g （见方法）。总体而言，我们发现不同网格尺寸下的结果非常稳定，系数接近 1。拟合质量见图 3c，其中将预测流量与观测流量进行了比较，皮尔逊相关值为 $r = 0.58$，表明该模型解释了知识空间中流动流方差的 $r^2 = 33.6\%$。值得注意的是，观测到的相关性高于现实世界流动性中的相关性（其 r 介于 $0.03 - 0.49$，见 [26]）。最后，我们发现，除了 t-SNE 之外，引力模型也能够表示不同嵌入技术下的科学流动流，其指数在定性上相似（见图S6、S7）。

科学探索者与开发者

跳转到下一篇文章时，研究人员可以前往空间中的新区域，也可以回到先前去过的区域。也就是说，在我们的框架中，研究人员会在探索新的科学领域和利用此前积累的知识之间做出选择。虽然这两种行为可能产生相似的跳跃分布模式，但它们会影响有关完整轨迹的更广泛统计特征，例如覆盖的空间活动范围。以往研究揭示了人类流动模式中个体轨迹存在这种异质性，凸显了回归者与探索者之间的二分：回归者倾向于围绕少数地点活动，而探索者则更愿意前往新的地点。这些结果在空间 [25], 和虚拟 [23] 环境中均成立。在此，我们探究知识探索情境中是否也存在这种异质性。

为评估轨迹覆盖的活动范围，我们研究回转半径 R_g，其定义为访问过的地点到其质心的平均距离（见方法部分中的公式 (3)）。将分析范围限定于访问次数最多的前 k 个地点，即可定义相应的回转半径 R_g^k （公式 (4)），并将其与完整轨迹的回转半径 R_g 进行比较，以评估轨迹在多大程度上会返回少数（k）地点。图 4a 展示了研究人员的总回转半径 R_g 与 R_g^2 的比较。我们发现，研究人员大致可分为两类：利用者的 R_g^2 值与 R_g 相近（数据点位于对角线附近）；探索者的 R_g^2 则明显小于总回转半径 R_g（数据点更靠近 x 轴）。考察 $S_2 = R_g^2/R_g$ 的分布时，这两类人群的差异更加明显：分布呈现双峰，分别对应探索者和利用者两个人群（图 4b）。当考察更大的 k 值时，这种双峰性便消失了（图 S8），这支持了使用 $k = 2$ 区分这两类人群。

从设计上看，探索者与利用者之间的差异会随着时间推移，影响科学轨迹所涵盖的研究空间。回转半径考察的是研究人员围绕特定中心的引力

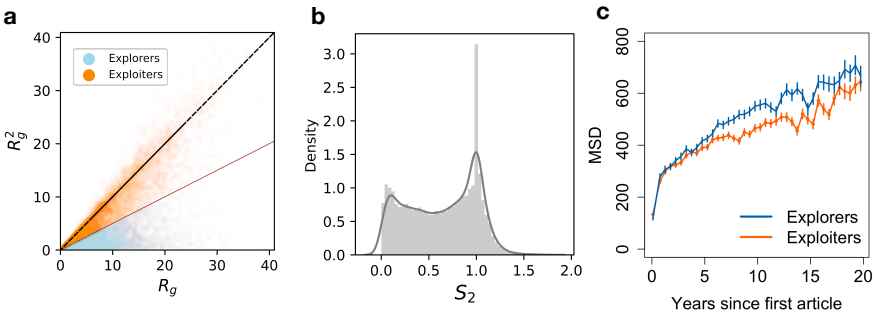

图 4 知识空间中的探索者与利用者。a. 回转半径的比较 R_g^2，其中质心根据访问次数最多的 2 个地点计算，并与完整回转半径进行比较 R_g（见方法）。我们使用平分线法对其进行分类，发现利用者（橙色）和探索者（蓝色）之间存在明显差异 [25]。 **b.** 分布$S_2 = R_g^2/R_g$，进一步凸显了这种差异：利用者（接近 1）和探索者（接近 0）呈双峰分布。 **c.** 探索者与利用者的均方位移随其发表首篇文章以来时间的变化对比。

（即重心）的吸引作用，其他指标则侧重于从初始起点出发的动态变化。在流动性分析中，这通常通过均方位移（MSD）来量化。 [50]，该指标追踪随时间推移从起点位置行进的平均距离（见方法）。之所以特别关注 MSD，是因为在齐次空间中，简单的扩散过程与时间呈函数尺度关系，即 MSD(t) $\sim t^{\beta}$，其中指数 β 表明该过程属于超扩散还是次扩散过程。在我们的研究中发现，尽管两类人群的跳跃幅度和持续时间相近（图 S9），且均方位移的次线性增长趋势相似，但职业生涯早期，探索者涉足的知识空间比例更大，表现为在 5 至 15 年间均方位移增长更快（图 4c）。这一差异在职业生涯后期（约 20 年）有所减小，这表明研究人员往往在学术生涯中期进行更多探索，而资深科学家则更倾向于利用以往的研究成果。这一发现印证了先前的观察：随着年龄增长，科学家造成颠覆性影响的可能性降低，也会对新兴研究持更批判的态度 [51]。

最后，除了流动模式上的差异，我们还想探究是否存在其他能够区分利用者与探索者的特征。为此，我们通过逻辑回归，根据若干属性预测某位研究人员是否属于探索者。为控制轨迹长度和领域特定行为，我们将发表文章总数以及作者发表文章最多的研究领域（图 1）纳入控制。图5 显示，与此前的观察一致，探索者会探索更广阔的空间区域，这以其职业生涯中达到的最大均方位移衡量；而利用者则倾向于停留在同一地点，这以其跳跃距离为 0 的比例衡量。

此外，探索者在文章内部和不同文章之间涉足的学科更多，而且这些学科在认知上往往相距较远，也就是说，它们在领域

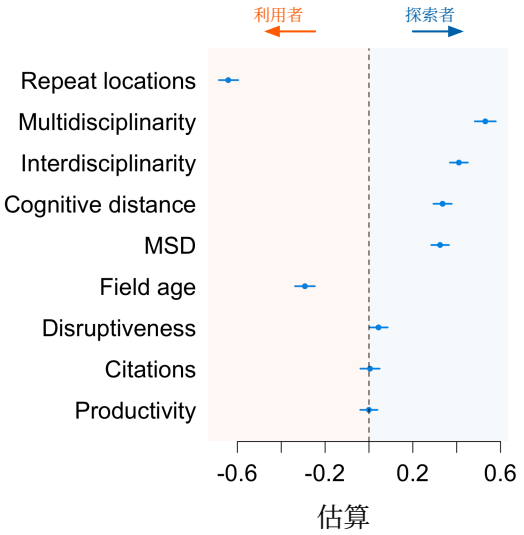

图 5 探索者的特征。我们对二元变量进行逻辑回归 y，该变量指示某个个体是探索者（$y = 1$）还是利用者（$y = 0$），并考察研究人员的不同特征。对于每项属性（详见方法部分），我们展示回归标准化系数的估计值及其 95% 置信区间，并控制研究人员的文章数量和主要研究领域。除颠覆性（10^{-5}）外，两类之间的差异均具有显著性，p 值小于 $p = 0.04$；引文和生产力的差异则不显著（$p > 0.05$）。重复位置对应于大小为 0 的跳跃所占比例。多学科性是文章中使用的不同领域标签总数，跨学科性则是每篇文章的领域标签平均数。认知距离是研究人员涉足的最大学科距离；领域年龄是研究人员发表的文章所涉及领域的最小归一化年龄。这两个量的定义见 [43]。MSD 表示研究人员职业生涯中达到的最大均方位移。颠覆性是研究人员各篇文章的颠覆性指数中达到的最高百分位数，以 arXiv 全部文章为参照。引文数对应于研究人员文章收到的最高引文数的对数。最后，生产力是研究人员年均发表文章数。

标签共现网络 [43]。考察与科学领域特定发展阶段的关联时 [43], ，我们发现探索者在某个领域演进的早期阶段发表文章，这表明他们从事开创性活动和创新性工作。与这一观察相符的是，我们发现探索者的颠覆性略高（$p = 0.04$）。颠覆性是衡量创新性工作的定量指标，用于量化引用某篇目标文章的文章同时引用其来源（颠覆性低）或不引用其来源（颠覆性高）的程度 [52],；此处通过其颠覆性最高的文章所处的百分位数进行衡量。最后，我们观察到，探索者和利用者的影响力相近（以其某篇文章获得的最高引文数衡量），年度生产力也相似。不过需要注意的是，基于引用的指标结果应谨慎看待，因为引文网络仅考虑 arXiv 内部的引用，因此并不完整 [53]，而且会受到领域特定习惯的影响。

讨论

在本研究中，我们表明，将流动性分析方法应用于知识空间的低维表示，有助于理解研究人员的科学流动。我们利用涵盖 30 年、来自预印本存储库 arXiv 的 150 万篇文章数据，发现研究人员的流动模式与人类流动性研究中的模式相似。知识空间不同区域之间的流动遵循引力模型，与距离成反比。这一结果并非特定表示方式造成的假象，因为在不同嵌入参数和方法下，这一结果均成立（见图 S4、S6、S7）。此外，该模型的准确性优于人类流动性研究中的经验结果 [26], ，表明尽管模型简单，却有望成为未来研究的坚实基础。此外，通过分析个体轨迹，我们发现研究人员可分为利用者和探索者两类：前者的轨迹局限于知识空间中的特定区域，后者则跨越边界，开拓新领域。这种二分法令人联想到科学研究中传统与创新之间的“本质张力”：探索新的、有前景的领域的愿望，与利用过去已有成果的需要相互制衡 [18, 54]。在此，我们通过观察到的双峰性，揭示了两种知识流动模式，从而识别出这种张力 R_g^2/R_g。

考察知识空间中科学轨迹的性质时，我们观察到，探索者和开发者的流动模式均呈现亚扩散区间。从理论上看，当考察个体在齐次空间中的流动时，例如在初始超立方体或正则格中，如果步长分布的二阶矩和等待时间概率分布的一阶矩均为有限值，则均方位移会呈现线性区间 [50]。在我们的案例中，观测到的均方位移偏离线性（图4）可能是由两组研究人员的重尾等待时间概率分布所导致的 [50] （图S9）。另一种可能性是，知识空间作为一种复杂且不断演化的认知构念所具有的多尺度特性，可能导致了这种类似受困的行为。还需要进一步研究，以确定哪一种解释更适合说明所观察到的均方位移行为的非线性。

我们基于研究人员的完整轨迹，将其划分为稳定的利用者和探索者类别，但他们的职业生涯中可能会出现变化。例如，我们观察到轨迹中存在老化模式：探索者和利用者的均方位移在首次发表论文后的前5年内数值相近，此后探索者的均方位移则显著更大（图4c）。这可能表明，探索者会经历两个阶段：第一阶段，他们主要停留在少数几个最常去的地点；之后则转向探索其他地点。这种行为可以用随机过程中的“间歇行为”概念加以形式化刻画 [55]。未来研究可以考察研究轨迹中的这类时间模式，例如采用时间窗口或凸包法，以更精细的尺度分析动态特征，并评估这些阶段是否

可能与制度约束有关，某些环境可能会促使个体发展出更具探索性的模式。

我们的框架依赖于定义知识空间所采用的方法。使用嵌入方法时并不存在真实标准，而且每种方法都可能使结果偏向某些特定的独有属性。然而，我们已经表明，多种参数方法（t-SNE）和非参数方法（UMAP、PaCMap）得出的结果在定性上相似，无论是长尾跳跃分布（图 S4）还是引力模型拟合（图 S6-S7）。这表明，尽管空间结构本身会带来一些差异，但本文揭示的总体流动模式并非空间表示伪影。

我们的研究聚焦于一个由 arXiv预印本构成的数据集。该数据集能够高精度地识别子领域，这对于构建知识空间以及计算领域年龄等特征 [43] 都很有用。不过，该数据集在总体规模、涵盖的领域以及引文网络的完整性方面存在局限。因此，我们的发现能否推广至其他学科，例如人文学科或社会科学，目前尚不明确。未来的研究应利用更大且更多样化的数据集，探索我们的发现能否重复验证，并采用其他领域识别方法，例如使用自然语言处理的方法 [56]。

尽管我们的研究着重于描述个体轨迹，但大多数文章由团队共同署名，[57], 而引导模式对科学生涯至关重要 [58]。因此，未来的研究可以探讨个体轨迹之间的耦合，从而形成相关模式，并最终产生集体流动。此外，还可以扩展引力模型，将与局部属性相对应的变量纳入其中，例如影响力（如通过引文体现）、领域年龄、专项资助等。这些特征可能会充当偏置力，引导集体流动朝向知识空间中的某些区域。在宏观层面上，这些场域能够影响流动性，就像力场影响物理学中粒子的轨迹一样。基于深度学习的新方法，例如结合更广泛的引文和资助数据的深度引力模型 [26], ，可能有助于拓展我们的研究。

总体而言，通过在知识空间中运用流动性分析所获得的洞见，有助于研究政策对知识探索与利用的影响，并可应用于资助机构以及更广泛的研究评估。

方法

arXiv数据集概览

在本研究中，我们使用了此前发表的一个数据集，其中包含来自 arXiv预印本库 [43] 的文章元数据。该数据集包含 1，456，403 篇发表于 1992年 至 2018年 的文章，主要涉及物理学、数学和计算机科学，较少涉及定量生物学、统计学、金融学、经济学和工程学。我们特别注意到

过去十年间计算机科学领域文章数量显著增长，按发表数量计算，很快就会超过物理学领域的文章；而物理学曾是 arXiv 早期使用的核心领域（图 S1）。

上传文章时，投稿作者会选择一个用于标识核心学科的主要领域标签，并可按需选择次要标签。在大多数情况下，作为质量控制机制，arXiv 要求首次向某个学科类别投稿论文的作者获得一位资深 arXiv 作者的推荐。标签涵盖 175 个预定义子领域，例如量子代数 (math.QA) 或信号处理 (eess.SP)，这些子领域均列于网站主页上。随着时间推移，这些子领域一直相对稳定 [43]。此外，作者有很强的动力选择恰当的领域，因为 arXiv 提供每日摘要邮件订阅服务，可自动接收包含特定领域标签的新投稿文章。因此，标签系统与投稿者所面向的相关受众直接挂钩，从而鼓励作者准确地自我标注。

低维嵌入

为了降低初始 175 维领域空间的维度，我们采用 t-SNE 算法。这是一种无监督、参数化的降维技术，能够在潜在空间中保留数据的局部结构 [44, 45]。t-SNE 方法能够捕捉高维数据的大部分局部结构，同时揭示不同尺度上簇的存在等全局结构。图 1展示了将 arXiv 知识空间中的嵌入映射到二维空间后的可视化结果。每个点对应 arXiv 文章中观察到的 49,575 种领域标签组合。需要注意的是，标签的不同排列会映射到同一个点，因此我们的分析不受标签顺序影响。

对于 t-SNE算法的实现，我们使用`scikit-learn`Python 包 [59]。嵌入空间的维度设为 2。嵌入方法的主要参数，如学习率、迭代次数和早期夸大参数，均采用默认值。为了测试 t-SNE 嵌入对参数变化的稳健性，我们针对不同的困惑度水平 p （代表最近邻）和学习率参数 LR 生成了算法的 t-SNE 映射，并绘制了不同设置下随机抽样点之间的成对距离分布（图S3）。我们发现，文中所述 t-SNE 的各种参数 [60]，（包括困惑度水平）变化时，结果仍表现出显著稳定性，这表明选择不同的 t-SNE 参数不会对结果产生显著影响。

嵌入方法的稳健性

为评估结果的稳健性，我们测试了不同嵌入方法、参数以及数据子样本对

跳跃分布（图S4）。除 t-SNE 外，我们还使用 PaCMAP 和 UMAP 嵌入评估了分析的稳健性。

统一流形近似与投影（UMAP） [61] 的理论基础是流形理论和拓扑数据分析。从宏观层面来看，UMAP 使用局部流形近似和模糊单纯集合，构建数据在高维和低维空间中的拓扑表示。随后，它通过最小化这两种拓扑表示之间的交叉熵，寻找最优的低维表示。UMAP 也可以被理解为一种基于 k 近邻的图学习算法，用于寻找加权图在低维空间中的最佳表示。

成对控制流形近似（PaCMAP）算法[60] 也是一种基于图的技术，它会识别三组点对，即邻居对、中近邻对和远邻对。随后，它使用定制的梯度下降算法系统地优化损失函数，以找到既能保留局部结构又能保留全局结构的低维表示。

我们发现，这两种方法的跳跃距离分布均呈现相似的长尾衰减（图S4a）。此外，我们发现，引力模型在 UMAP（图S6）和 PaCMAP（图S7）情境下的结果稳定，不过我们发现距离的指数整体较小，更接近于 $\gamma \simeq 0.5$。

引力模型的拟合过程

为了拟合引力模型，我们对对数变换变量进行了线性回归。对于每个分辨率级别，我们首先计算了从 1997 年开始的每一年中，源单元格与目标单元格之间的跳跃次数 i ，以及此前 5 年内每个单元格发表的文章数量 j (其中$i \neq j$)。为考虑样本量较小的情况，我们在原始访问计数中加上 1 作为伪计数：
$V_i \leftarrow V_i + 1$。随后，我们计算所有量的自然对数，并使用这些对数变换后的值进行回归分析。由于小流量值的数量远多于大流量值（图3），我们采用分箱技术，避免模型对低流量值过拟合。具体而言，我们将得到的对数流量值分成 100 个包含相同数量数据点的箱，并合并断点相同的箱，最终得到 42 个箱。随后，我们对这些箱内经对数变换的特征（流量、访问次数、距离）取平均值，并使用这些平均值拟合引力模型，调用 R 4.2.2 中的`lm` 函数。模型的残差如图 S5 所示，且呈正态分布。

回转半径

回转半径衡量个体轨迹所覆盖活动范围的典型大小。要计算它，我们首先定义轨迹的质心

在各地点间的 R_{cm} 质心 $i = 0$，1，.....， n：

$$R_{cm} = \frac{\sum_{i=1}^{n} M_i R_i}{\sum_{i=1}^{n} M_i}, \tag{2}$$

其中， M_i 是每个地点 i的访问频率，即个体访问地点 i 的次数； R_i 是位置矢量，用于表征该地点在知识空间中相对于所选坐标中心的位置。回转半径定义为距质心的特征距离：

$$R_g = \sqrt{\frac{\sum_{i=1}^{n} M_i (R_i - R_{cm})^2}{\sum_{i=1}^{n} M_i}}. \tag{3}$$

为了估计少数地点对轨迹的影响，我们仅考虑访问次数最多的前 k 个地点，并据此定义第 k个回转半径：

$$R_g^k = \sqrt{\frac{\sum_{i=1}^{k} M_i (R_i - R_{cm}^k)^2}{\sum_{i=1}^{k} M_i}}, \tag{4}$$

其中， R_{cm}^k 是基于访问次数最多的前 k 个地点计算的质心。

均方位移

轨迹在时间 t 的均方位移（MSD）定义为行走者（在本研究中指研究人员）的位置相对于参考位置随时间的偏差：

$$\mathrm{MSD}(t) = \langle |x(t) - x(0)|^2 \rangle \tag{5}$$

其中， $x(t)$表示研究人员自发表第一篇文章以来，在时间 t 时所处的位置； $x(0)$ 表示轨迹的起点。

探索者与利用者的逻辑回归

为了探索图中与探索者相关的特征 5，我们进行了逻辑回归，因变量为 y_i，这是一个二元变量，用于表示个体 i是探索者（$y_i = 1$）还是开发者（$y_i = 0$），自变量为各项个体特征 x_i。我们控制了作者发表文章所涉及的主要领域 F_i（根据其文章中出现频率最高的领域标签确定），以及研究人员发表的文章数量 N_i。各领域被编码为因子。我们使用 R 中的`glm` 函数拟合模型 $y_i \sim x_i + N_i + F_i$，并将参数 `family = binomial`设为二项分布，以进行逻辑回归。我们使用 R 中jtools 包里的summ 函数生成回归摘要，并设置参数 `scale=T`，通过对输入数据进行缩放和均值中心化来标准化回归系数；设置参数 `confint=T` 以获得 95% 置信区间。

创新、颠覆性与影响力

为了衡量一篇作品的创新水平，我们采用了两种方法。首先，我们计算每篇文章在其提及的领域中出现得有多早。为此，我们按照 重标度时间 （RT）的方法，计算其相关领域标签中的最小值，具体方法见 [43]。重标度时间是一个归一化量，使我们能够将文章与某个领域的发展阶段（早期、峰值或晚期）相关联，即使不同领域的兴起和衰退时长差异很大。随后，我们计算每位研究人员的所有文章中达到的最小 RT 值，并将其定义为“领域年龄”。

其次，我们采用另一种独立方法评估文章的创新潜力。该方法通过比较一篇文章获得的关注度与其所引用文章获得的关注度，来衡量该文章的颠覆性。引文数据取自 [53]。随后，我们使用 [62] 的方法计算每篇文章的颠覆性指数（DI）。对于每位作者，我们计算其所有文章中的最高 DI 值。最后，我们计算该值在所有文章中的百分位数，以此衡量作者的颠覆性。

最后，对于每位作者 i，我们计算其整个职业生涯中任一文章获得的最高引文数 $\hat{c}_i$。由于引文数呈重尾函数分布，我们使用变换 $\log(\hat{c}_i+1)$ 来量化影响力。

认知距离

我们从图中观察到 5，与利用者相比，探索者每篇文章使用的领域标签更多，其所有文章中使用的独有标签也更多。然而，就研究领域而言，有些标签之间的关联可能比其他标签更密切，而用于线性估计所用标签数量的简单测量方法无法区分这种差异。为考虑这一影响，我们采用 [43], 提出的基于网络的认知距离测量方法，其中认知距离 C_{ij} 是领域标签 i 与 j 之间沿最短路径的加权距离，该路径位于标签共现网络中i 和 j 之间。

致谢

感谢 Emma Barme 在该主题早期讨论中的参与，也感谢 Ariel Lindner 和 Robert Ward 就相关文献提供建议。感谢 Bettencourt Schueller 基金会的长期合作伙伴关系，本研究部分得到了 Marc Santolini 所获 CRI 研究奖学金的支持。

补充图

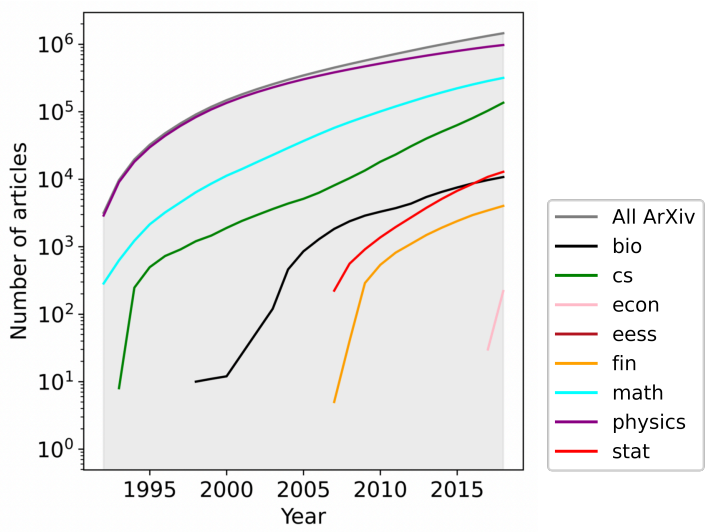

图 S1 1992年至2018年间，向 arXiv 各主要领域提交的文章数量增长情况。我们展示了 arXiv 中 9 个主要领域的文章累计数量随时间的变化，总计 1456403 篇文章。这些领域包括：定量生物学（bio）、计算机科学（cs）、经济学（econ）、电气工程与系统科学（eess）、定量金融学（fin）、数学（math）、物理学（physics）和统计学（stat）。为了更直观地呈现较新领域的文章数量，我们采用了对数缩放。除了早期占主导地位的物理学（arXiv 最初便兴起于该领域）之外，我们还可以看到，计算机科学在21世纪初迅速呈指数增长。

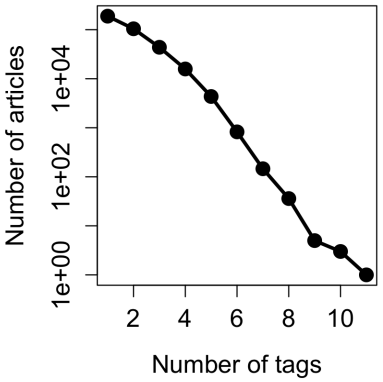

图 S2 高维空间的稀疏性。 arXiv 中具有给定标签数量的文章数。大多数论文拥有不超过 10 个标签，这远少于该空间的维度（$N = 175$），这意味着空间中大多数可能的地点都没有被占据。

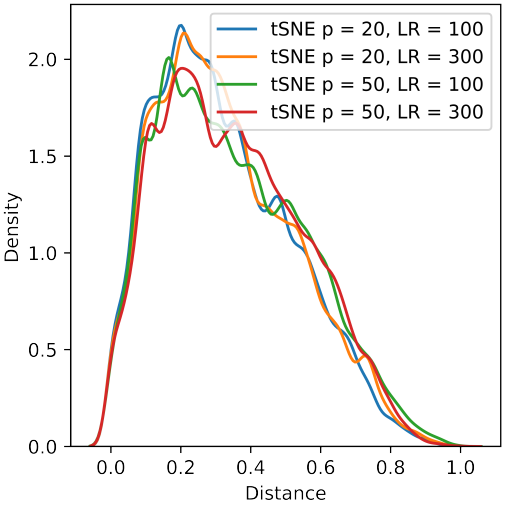

图 S3 不同 t-SNE 参数下成对距离分布的比较。 不同 t-SNE 嵌入中随机抽取的点对之间距离的密度分布，参数显示在图例中。

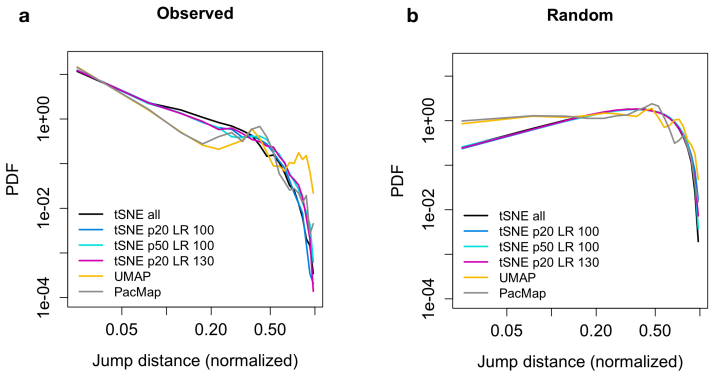

图 S4 不同嵌入方法的跳跃分布比较。a.我们展示了使用不同参数进行 t-SNE 嵌入（p 为困惑度，LR 为学习率）以及采用其他嵌入方法（PaCMap 和 UMAP）时跳跃分布的变化。为了比较不同嵌入结果，每种嵌入的跳跃距离均按其最大距离进行归一化。**b.** 对于每位作者，我们从嵌入中所有可访问的点（即数据集中现有的唯一地点）随机选取地点，并绘制不同嵌入中相应的随机化跳跃分布。

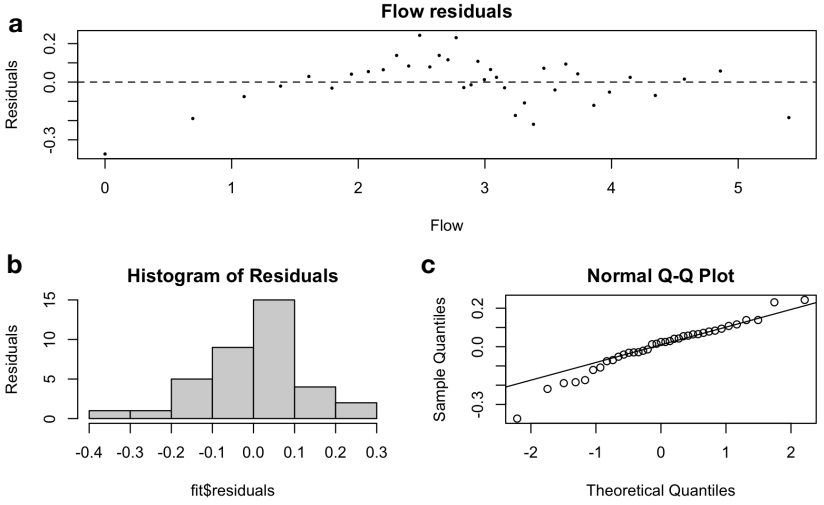

图 S5 引力模型的残差图。我们检验回归分析中残差服从正态分布这一假设是否成立。**a** 残差值随拟合的（对数）流量值的变化。我们发现，流量值非常低时会略微被高估，这可能是由所采用的伪计数方法导致的。**b** 残差直方图。**c** 残差呈正态分布，这可通过 Q-Q图进行评估。

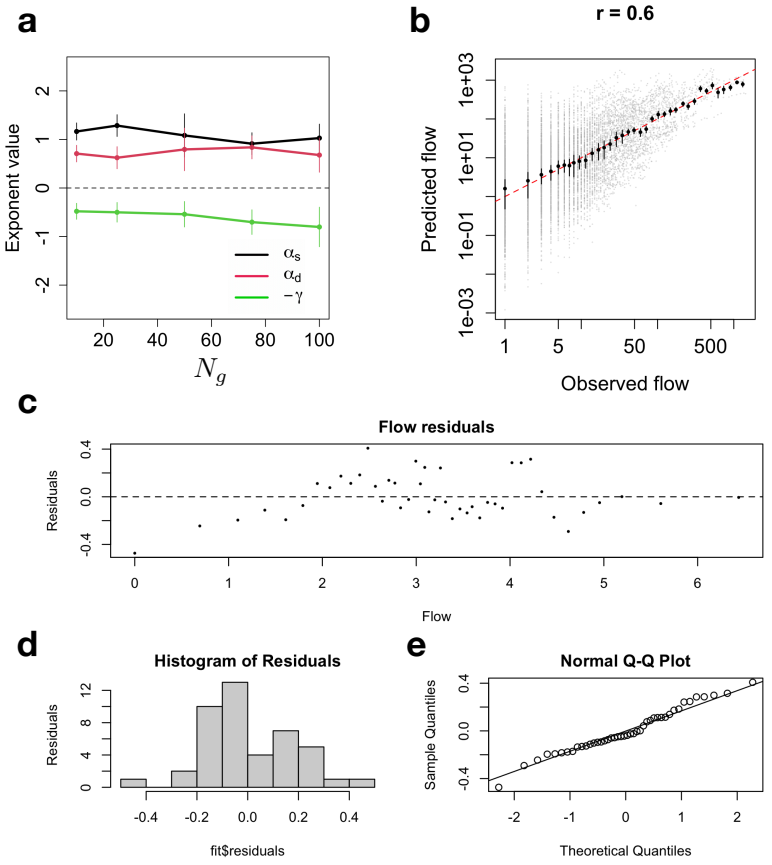

图 S6 UMAP 嵌入的引力模型拟合结果。a-b. 与图3b-c. **c-e.** 与图 S5。

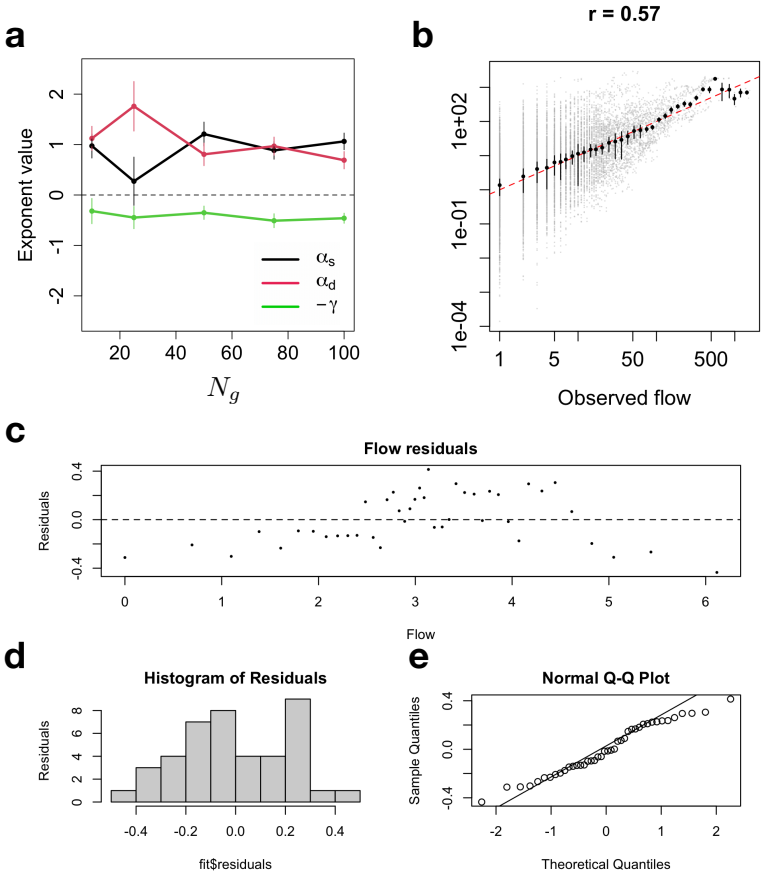

图 S7 PaCMAP 嵌入的引力模型拟合结果。a-b. 与图3b-c. **c-e.** 与图S5相同。

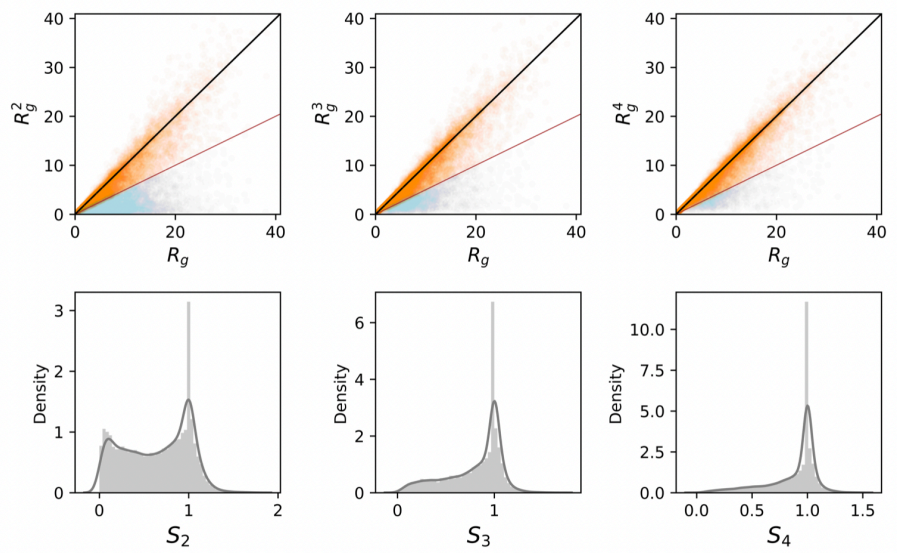

图 S8 较大值下的探索者和利用者 k。顶行：与图 4a 相同，适用于 $k = 2$、3、4。红线对应于 R_g^k- $R_g/2 = 0$，线以上的点（橙色）对应于利用者，线以下的点（蓝色）对应于探索者。底行：与图 4b 相同，适用于 $k = 2$、3、4。$k = 2$ 所观察到的双峰性在 k的较大值下消失。

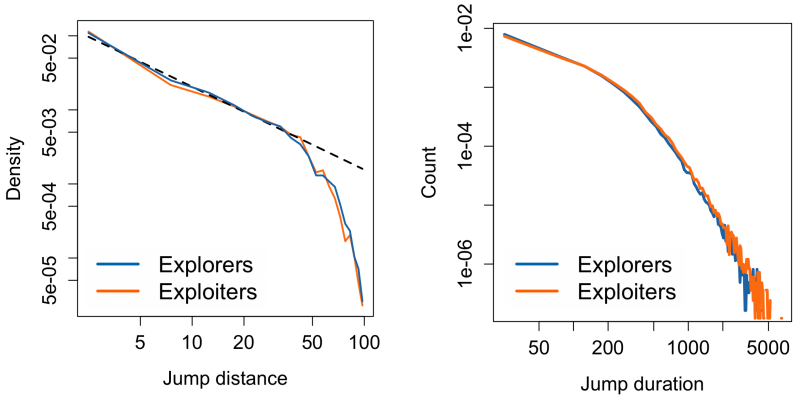

图 S9 比较探索者和利用者之间的跳跃分布。探索者和利用者之间的跳跃距离（左）和跳跃持续时间（右，以天为单位）分布相似（双对数图）。

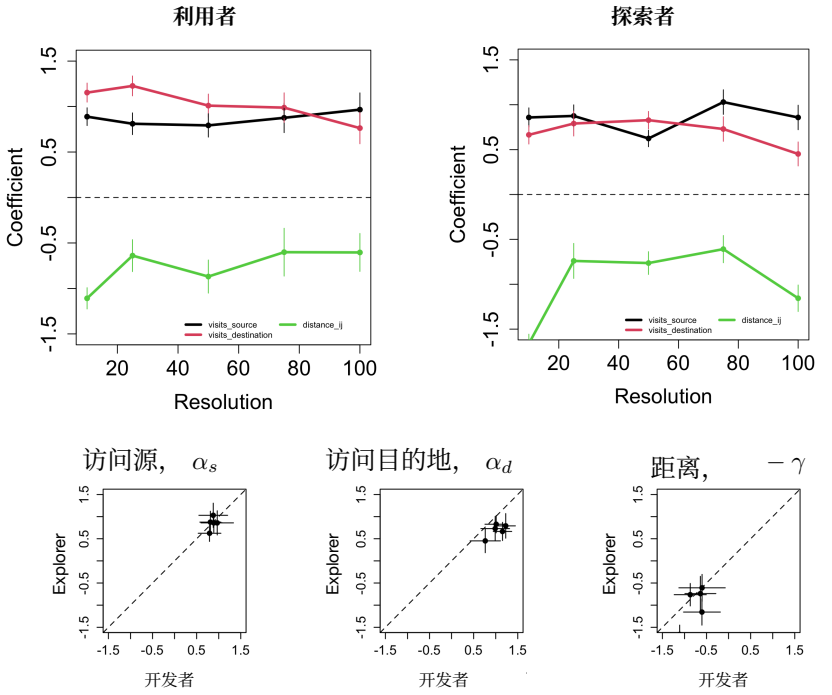

图 S10 探索者和开发者的引力模型。我们展示了仅应用于利用者（左上）和仅应用于探索者（右上）的引力模型拟合过程结果。随后，我们比较这两类人在不同分辨率下访问源、访问目的地和距离的指数（下方）。误差条表示回归的标准误差。

参考文献

[1] Belikov, A.V., Rzhetsky, A., Evans, J.：《从文献中预测稳健的科学事实》。《自然·机器智能》4(5)，445–454（2022年）[2] Iacopini, I., Milojevi´c, S., Latora, V.：《创新过程的网络动力学》。《物理评论快报》120(4)，048301（2018年）。 https://doi.org/10.1103/PhysRevLett.120.048301。 arXiv:1707.04239 [物理学][3] Ferreira, M.R., Reisz, N., Schueller, W., Servedio, V.D.P., Thurner, S., Loreto, V.：载于：La Porta, C.A., Zapperi, S., Pilotti, L.（编）：《科学演化中的量化适应性转用》，第55–68页。施普林格，尚（2020年）。 https://doi.org/10.1007/978-3-030-45784-6 5[4] Bornmann, L., Haunschild, R., Mutz, R.：《现代科学的增长率：采用潜在分段增长曲线方法，对来自既有文献数据库和新文献数据库的出版物数量进行建模》。《人文与社会科学传播》8(1)，1–15（2021年）[5] Fortunato, S., Bergstrom, C.T., B¨orner, K., Evans, J.A., Helbing, D., Milojevi´c, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B.等：《科学学》。《科学》359(6379)，0185（2018年）[6] Shwed, U., Bearman, P.S.：《科学共识形成的时间结构》。《美国社会学评论》 75(6)，817–840 （ 2010 年）。 https://doi.org/10.1177/0003122410388488。 出版商：SAGE 出版公司[7]Lin, Y., Evans, J.A., Wu, L.：《科学的新方向源于断联与不一致》。《信息计量学期刊》16(1)，101234（2022年）[8] Liu, L., Wang, Y., Sinatra,R., Giles, C.L., Song, C., Wang, D.：《艺术、文化与科学生涯中的高光期》。《自然》559(7714)，396–399（2018年）[9] Sun, Y., Latora, V.：《现代物理学领域内部及跨领域的知识演化》。《科学报告》10(1)，1–9（2020年）[10] Yin, Y., Wang, D.：《科学的时间维度：连接过去与未来》。《信息计量学期刊》11(2)，608–621（2017年）。 https://doi. org/10.1016/j.joi.2017.04.002[11] Pan, R.K., Petersen, A.M., Pammolli, F., Fortunato, S.：《科学的记忆：膨胀、短视与知识网络》。《信息计量学期刊》12(3)，656–678（2018年）[12] Wang, D., Song, C., Barab´asi, A.-L.：《量化长期科学影响力》。《科学》342(6154)，127–132（2013年）

[13] Sinatra, R., Wang, D., Deville, P., Song, C., Barab´asi, A.-L.：《量化个人科学影响力的演化》。《科学》354(6312)，5239（2016年）[14] Chavalarias, D., Cointet, J.-P.：《科学演化中的系统发育模因模式——科学领域的兴衰》。《公共科学图书馆·综合》8(2)，54847（2013年）[15] Battiston, F.,Musciotto,F., Wang, D., Barab´asi, A.-L., Szell, M., Sinatra, R.：《开展物理学普查》。《自然评论：物理学》1(1)，89–97（2019年）[16] Jia, T., Wang, D., Szymanski, B.K.：《量化研究兴趣演化模式》。《自然·人类行为》1(4)，1–7（2017年）[17]Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H.E., Havlin, S.：《科学家在不同主题之间转换的趋势日益增强》。《自然·通讯》10(1)，3439（2019年）。https://doi.org/10.1038/ s41467-019-11401-8。 编号：1 出版商：自然出版集团。访问日期：2023年2月21日[18] Aleta, A., Meloni, S., Perra, N., Moreno, Y.：《谨慎探索：绘制物理学中科学兴趣的演化图谱》。《EPJ数据科学》8(1)，1–15（2019年）。https://doi.org/10.1140/epjds/s13688-019-0205-9。 编号：1 出版商：SpringerOpen。访问日期：2023年2月21日[19]Tuninetti, Marta, Aleta, Alberto, Paolotti, Daniela, Moreno, Yamir, Starnini, Michele：《通过多重网络预测新的科学合作》。《EPJ数据科学》10(1)，25（2021年）。https://doi.org/10.1140/ epjds/s13688-021-00282-x[20] Holden, C.：《科学家联合会计划“大跃进”》。《科学》185(4145)，47–47（1974年）。https://doi.org/10.1126/science.185.4145. 47.a。 出版商：美国科学促进会。访问日期：2022年12月21日[21] Wu, L., Hasan, S., Chung, Y., Kang, J.E.：《理解人类流动模式的异质性：用户特征与出行方式偏好》。《可持续发展》13(24)，13921（2021年）。 https://doi.org/10.3390/ su132413921。 编号：24 出版商：多学科数字出版研究所[22] Ubaldi, E., Monechi, B., Chiappetta, C., Loreto, V.：《流动模式的异质性与隔离》。《熵、复杂性与空间动力学手册》，486–509（2021年）。 ISBN：9781839100598 出版商：爱德华·埃尔加出版社 章节：《熵、复杂性与空间动力学手册》

[23] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M.,

Louail, T., Menezes,R., Ramasco, J.J., Simini, F., Tomasini,M.：人类流动：模型与应用。《物理学报告》**734**，1–74（2018年）。
https://doi.org/10.1016/j.physrep.2018.01.001

[24] Schl¨apfer,M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., West, G.B.：人类流动的普适访问定律。《自然》**593**(7860)，522–527（2021年）

[25] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barab´asi, A.-L.：人类流动中的回访者与探索者二分法。《自然·通讯》6(1)，1–8（2015）[26] Simini, F., Barlacchi, G., Luca, M., Pappalardo, L.：用于生成流动流的深度引力模型。《自然·通讯》12(1)，1–13（2021）[27] Alessandretti, L., Aslak, U., Lehmann, S.：人类移动尺度。《自然》587(7834)，402–407（2020）[28] Schneider, C.M., Belik, V., Couronn´e, T., Smoreda, Z., Gonz´alez, M.C.：解析日常人类移动模式。《英国皇家学会界面杂志》10(84)，20130246（2013）[29] Simini, F., Gonz´alez, M.C., Maritan, A., Barab´asi, A.-L.：移动与迁徙模式的通用模型。《自然》484(7392)，96–100（2012）[30]Wilson, A.G.：空间分布统计理论。《交通研究》1(3)，253–269（1967年）。https://doi.org/10.1016/0041-1647(67)90035-4[31] Hills, T.T., Todd, P.M., Lazer, D., Redish, A.D., Couzin, I.D., Group, C.S.R.，等：空间、心智与社会中的探索与利用。《认知科学趋势》19(1)，46–54（2015）[32] Hills, T.T.：动物觅食与目标导向认知的演化。《认知科学》30(1)，3–41（2006）[33] Bonabeau, E., Dorigo, M., Theraulaz, G., Theraulaz, G.：群体智能：从自然系统到人工系统（1）（1999）[34]Hills, T.T., Jones, M.N., Todd, P.M.：语义记忆中的最优觅食。《心理学评论》119(2)，431（2012）[35] March, J.G.：组织学习中的探索与利用。《组织科学》2(1)，71–87（1991年）。出版商：美国运筹学与管理科学研究所[36] Zhao, Z.-D.,Huang, Z.-G.,Huang, L.,Liu, H., Lai,Y.-C.：网络空间与物理空间中人类运动的尺度关系与相关性。《物理评论

E **90**（5），050802（2014）

[37] Hu, T., Luo, J., Liu, W.: “矩阵”中的生活：网络空间中的人类流动模式。载于：第十二届国际 AAAI 网络与社交媒体会议（2018年）[38] Barbosa, H.S., de Lima Neto, F.B., Evsukoff, A., Menezes, R.: 网页浏览行为中的回访者与探索者二分法——人类流动方法，第173–184页。施普林格，尚（2016年）。https://doi.org/10.1007/ 978-3-319-30569-1 13[39]Milojevi´c, S.: 量化科学的认知范围。《信息计量学期刊》9(4), 962–973（2015年）[40] Milojevi´c, S., Sugimoto, C.R., Yan, E., Ding, Y.: 图书馆与信息科学的认知结构：文章标题词分析。《美国信息科学与技术学会期刊》62(10), 1933–1953（2011年）。https://doi.org/10.1002/asi.21602[41] Peng, H., Ke, Q., Budak, C., Romero, D.M., Ahn, Y.-Y.: 学术期刊的神经嵌入揭示了复杂的学科组织结构。《科学进展》7(17), 9004（2021年）[42]Ying, Q.F., Venkatramanan, S., Chiu, D.M.: 科学景观中的学者流动建模与分析。arXiv。arXiv:1502.00523[物理学]（2015年）。http://arxiv.org/abs/1502.00523[43] Singh, C.K., Barme,E.,Ward,R.,Tupikina, L., Santolini,M.: 量化科学领域的兴衰。《公共科学图书馆·综合》17(6), 0270131（2022年）[44] Van der Maaten, L., Hinton, G.: 使用 t-SNE 可视化数据。《机器学习研究期刊》9(11)（2008年）[45] Van Der Maaten, L.: 通过保留局部结构学习参数化嵌入。载于：人工智能与统计学，第384–391页（2009年）。PMLR[46]Wilson, A.G.: 空间分布模型的统计理论。《交通研究》1(3), 253–269（1967年）。https://doi.org/10.1016/ 0041-1647(67)90035-4[47] Senior, M.L.: 从引力建模到熵最大化：教学指南。《人文地理学进展》3(2), 175–210（1979年）[48] Wilson, A.: 城市与区域建模中的熵（劳特利奇重印版）。劳特利奇（2013年）

[49]Ribeiro, F.L., Rybski, D.: 解释城市规模缩放定律起源的数学模型：综合评述。arXiv。arXiv:2111.08365[物理学] （2021年）。https://doi.org/10.48550/arXiv.2111.08365。http://arxiv.org/abs/2111.08365 访问日期：2022年12月26日[50] Klafter, J., Sokolov, I.M.: 随机游走入门：从工具到应用（2011年）。https://doi.org/10.1093/acprof:oso/9780199234868. 001.0001[51] Cui, H., Wu, L.,Evans, J.A.: 科学家老化与进展放缓。arXiv。arXiv:2202.04044 [计算机科学]（2022年）。https://doi.org/10.48550/arXiv.2202.04044。http://arxiv.org/abs/2202.04044 访问日期：2023年2月21日[52] Park, M., Leahey, E., Funk, R.J.: 论文和专利随着时间推移的颠覆性日益减弱。《自然》613(7942), 138–144（2023年）。https://doi.org/10.1038/s41586-022-05543-x。访问日期：2023年1月13日[53]Clement, C.B., Bierbaum, M., O'Keeffe, K.P., Alemi, A.A.: 论将 arXiv 用作数据集。arXiv。arXiv:1905.00075 [物理学] （2019年）。https://doi.org/10.48550/arXiv.1905.00075。http://arxiv.org/abs/1905.00075 访问日期：2022年12月23日[54] Kuhn, T.S.: 本质张力：科学传统与变革研究选集。芝加哥大学出版社，美国伊利诺伊州芝加哥（1979年）。https://press.uchicago.edu/ucp/books/book/chicago/E/bo5970650.html 访问日期：2023年2月21日[55]Lanoisel´ee, Y., Grebenkov, D.S.: 用局部凸包方法揭示单粒子轨迹中的间歇性特征。《物理评论E》96(2), 022144（2017年）。https://doi.org/10.1103/PhysRevE.96.022144。出版者：美国物理学会。访问日期：2022年12月21日[56]Beltagy, I., Lo, K., Cohan, A.: SciBERT：科学文本预训练语言模型。载于：2019年自然语言处理实证方法会议暨第九届国际自然语言处理联合会议（EMNLP-IJCNLP）论文集，第3615–3620页。计算语言学协会，中国香港（2019年）。https://doi.org/10.18653/v1/D19-1371。https: //aclanthology.org/D19-1371 访问日期：2023年2月20日[57] Wuchty, S., Jones, B.F., Uzzi, B.: 团队在知识生产中的主导地位日益增强。《科学》316(5827), 1036–1039（2007年）[58]Sekara, V., Deville, P., Ahnert, S.E., Barab´asi, A.-L., Sinatra, R., Lehmann, S.: 科学出版中的伴护效应。《美国国家科学院院刊》115(50), 12603–12607（2018年）

[59]Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,M., Prettenhofer, P., Weiss, R., Dubourg, V., Vander- plas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.：Scikit-learn：Python 机器学习。《机器学习研究杂志》**12**(null), 2825–2830 (2011年)

[60] Wang, Y., Huang, H., Rudin, C., Shaposhnik, Y.：了解降维工具的工作原理：一种用于解读 t-sne、umap、trimap 和 pacmap 数据可视化的实证方法。《机器学习研究杂志》**22**(201), 1–73 (2021年)[61] McInnes, L., Healy, J., Melville, J.：Umap：用于降维的统一流形近似与投影。arXiv 预印本 arXiv:1802.03426 (2018年)[62] Wu, L., Wang, D., Evans, J.A.：大型团队推动科学与技术发展，小型团队颠覆科学与技术。《自然》**566**(7744), 378–382 (2019年)。https://doi.org/10.1038/s41586-019-0941-9。编号：7744。出版方：自然出版集团。访问日期：2021-03-19