使用谷歌学术估算文献计量联系：移民与住房的半系统性文献映射

Boyana Buyuklieva博士，boyana.buyuklieva@ucl.ac.uk
Juste Raimbault博士，j.raimbault@ucl.ac.uk

摘要

随着高等教育不断扩张，实证研究数量空前增长；然而，由于相关研究面临可发现性危机，人口研究的理论发展受到阻碍。医学领域通过各种结构化综述系统地运用既有研究，这种做法十分普遍。然而，尽管社会科学采用此类方法潜力巨大，这种做法仍不常见，尤其是在移民等跨学科领域。我们使用谷歌学术，通过广泛的文献考察移民研究中住房研究这一细分领域。本元分析有三方面贡献。首先，我们展示了移民与住房相关文献语料库中关键词之间的关联，并绘制了自20世纪60年代以来迁移文献的增长情况。其次，我们利用网络测度指出关键的桥接文献。最后，我们估算了文献之间的阅读时间距离。研究结果表明，移民与住房交叉领域的既有研究语料库包含许多累加式案例研究，这说明在不断扩展的知识图景中，仍缺乏可复制的整合式方法。

1.1 引言

本文的元分析概述了移民研究的发展，重点关注内部迁移与住房。Ravenstein（1889）的研究被视为具有奠基意义的研究成果之一，也标志着移民领域的开端（Levy、Pisarevskaya 和 Scholten，2020年，第2页；Pisarevskaya 等，2020年，第457页）。一个多世纪后，人们对该领域的兴趣不断增加。关于这一主题的期刊和学位项目数量有所增长（King，2020年，第4页），国际研究网络也日益增多（Levy、Pisarevskaya 和 Scholten，2020年，第20页）。自20世纪90年代中期以来，相关出版物的数量以及专门发表这些研究成果的出版渠道一直呈“指数级增长”（Pisarevskaya 等，2020年，第477页）。

以往关于迁移的文献计量研究几乎完全基于Scopus或科学网等订阅制学术数据库。这些数据库的文献覆盖范围也各不相同：1990年代至2018年间有N = 1 193篇文献（Piguet、Kaenzig和Guélat，2018年，第357页），1945年至2015年间有N = 2 750 篇文献（Nestorowicz和Anacka，2018年，第291页），2000年至2016年间有N = 21千篇文献（Sweileh等，2018年，第5页），1959年至2017年间有N = 30千篇文献（Pisarevskaya 等，2020年，第461页），1975年至2018年间有N = 49千篇文献（Levy、Pisarevskaya和Scholten，2020年，第6页）。本文将在这些研究的基础上，将数值分析的规模扩大一个数量级，同时聚焦于英格兰和威尔士迁移研究中住房这一更狭窄的检索范围交集。

1.2 认识共同体

认识共同体是由从事“特定概念子集”研究的研究人员组成的群体（Roth和Bourgine，2005年，第108页）。认识共同体聚集之处，便形成“话语空间”（Levy、Pisarevskaya和Scholten，2020年，第16页），作者在其中依靠相似的体裁知识，在概念层面定位自身。这一认识论文轨迹的作用和目的颇为有趣。一种观点认为，科学家希望其原创贡献获得认可（Cozzen于1981年提出的默顿假说，第3页）。另一种观点则认为，关键不在于获得赞赏，而在于归属于更广泛的学术社会化体系（Cozzen于1981年提出的卡普兰假说，第2页）。按照这一思路，参考文献引用可以作为一种说服手段，使当前研究通过与另一来源建立关联而定位于既有证据之中。因此，参考文献引用是一种嵌入意义的象征性行为（Cozzens，1981年，第4–5页）。然而，所有路径最终都指向同一结果：引文在任何领域的学术成果之间建立了切实的联系，其中也包括迁移领域。这种切实的联系可以用于量化一个领域的影响范围，之后再梳理其反复出现的构成主题。迁移研究中存在若干大型认识共同体。其中尤为突出的是构成国际迁移和内部迁移集群的不同群体。

1.2.1 国际迁移与内部迁移

移民研究是社会科学中一个成熟且独立的分支；然而，其国际与内部研究范围之间存在一道“鸿沟”（King，2020年，第4页）。历史上，这道鸿沟体现为“不同的文献、概念、方法和政策议程”（King 和 Skeldon，2010年，第1619页；Nestorowicz 和 Anacka，2018年，第284页）。国际迁移与内部迁移学者所构成的认识共同体，在术语和关键词词汇方面存在差异（Korgelli，1994年，第152页；Nestorowicz 和 Anacka，2018年，第294页）。由于实证研究核心的数据来源在议题和范围方面存在问题，迁移二分“有其充分理由”（Nestorowicz 和 Anacka，2018年，第285页）。

尽管国际移民推动了迁移研究的发展，但内部迁移在全球更为普遍，因此“在数量上更为重要”（King，2020年，第3页）。此外，尽管这两种研究方法在词汇和相互引用方面存在很大差异，内部迁移和国际移民都是可行的策略，可在个体层面相互替代（King，2011年，第137页）。内部迁移作为规模较小的知识共同体（Nestorowicz 和 Anacka，2018年，第293页），可以借鉴前者的理论和实证方法。然而，国际移民与内部迁移显然是两个不同的研究领域。由于住房受国家管辖权约束，更为重要的知识共同体是有关内部迁移的规模较小的文献群体。

1.3 体裁知识

不同学科各有其语言使用方式——语言学研究这一现象，并将其称为体裁知识，或称“基于社群的知识赖以构建的智力支架……专业人士在学科共同体中进行交流所需的知识”（Berkenkotter and Huckin，2016，第24–25页）。如表1-1所示，移民是多个学术学科共同研究的主题。该表改编自Brettell and Hollifield（2015，第4/24页）编著的论文集中有关迁移的学科研究方法的比较性综述。表1-1列出了地理学、经济学、历史学和社会学等关键领域，以说明各领域与移民的关联。

Discipline	Scale of Analysis	Units of Analysis	Common Independent Variable	Common Dependent Variable	Relevance to this research
Geography	Macro, Meso	Distance, People counts	Spatial, environmental, or socio-economic geographical delinations	Migration decisions: resulting flows	**Quantifying migration and delinating areas spatially**
Demography	Micro, Meso, Macro	People counts often by groups *(often ethno-racial or foreign populations)*	Ethno-racial diversity, characteristics of migrant or context of reception *(often in biological terms: age-structure, mortality, fertility ect.)*	Size of migration flows, degree of similarity between migrants and locals	**Quantifying life-course stages**
Economics	Micro	Individual's choices *(often using decision theory)*	Income differentials, Human capita, demand/ supply, push/pull factors	Migration decisions: economic impact of migration flows	**Framing individual choices algorithmically** *(e.g need thresholds and migration decisions)*
Sociology	Macro, Meso	Groups *(often by social class)*	Social capital, networks or enclaves	Migrant behaviour:, including experiences societal cohesion	**Framing neighbourhood characteristics** *(e.g community definitions)*
History	Temporal	Timespans	Social or historical context	Migrant experience: periodisation	**Setting initial conditions for housing context**
Political Science	Macro	Population groups	Institutions, rights and responsibilities, interest groups	Policy output or outcome Political incorporation and civil engagement of migrants	**Political theory, specifically definitions and evaluation of exogenous factors** *(e.g Land and human capital, equality and equity)*

表*1-1.* 跨学科迁移

因此，迁移研究者在不同学科共同体之间进行交流时所使用的语言，值得探讨。这是因为体裁知识也是“个人针对反复出现的情境所形成的一套符合情境的应对方式”（Berkenkotter and Huckin，2016，第ix页）。从这个意义上说，专业知识既能助力人们在既有研究基础上继续推进，也有助于与同行建立联系。然而，它同时也具有局限性：某一案例研究中的任何“恰当应对”都受其经验语境在时间和空间上的限制，未必能推广到其他情境。

1.3.1 迁移中的地理学

地理学家是迁移领域内另一个重要的认识共同体。地理学在这一领域中的作用可以说遭到了“低估”，因为研究工作并未得到整合，而是分散在许多子领域中（King，2020年，第1–7页）。因此，地理学在迁移研究中发挥了至关重要的作用，将不同的认识共同体联系起来：“我们的[移民] 专家一直认为，地理学家是推动跨学科发展的关键人物”（Levy、Pisarevskaya 和 Scholten，2020年，第19页）。

然而，作为一个关键支点，地理学在移民研究中的发展一直被描述为“一个描述事实、图表和区域专著的描述性学科”——缺乏“批判锋芒”（King，2020年，第6页）。当研究受到经验数据的限制时，这种情况并不意外：内部迁移研究在地理分布上偏向全球北方，来自非洲或拉丁美洲的研究成果寥寥无几（Carling，2015）。更具体地说，移民研究由针对美国、英国和中国背景的案例研究主导，这些研究从宏观经济术语的角度来阐释这一过程，涉及城市增长、就业／失业和收入（Nestorowicz 和 Anacka，2018年，第293页）。经济地理学——本身就是经济学领域中的一个小众分支——

也带有一种“只管去做”的风气，“理论、方法乃至实质内容上的‘转向’似乎正以越来越高的频率被宣布”（Foster 等，2007年，第1页）。这一领域的反复无常表明，迁移的空间研究采取了一种缺乏协调、支离破碎的路径。

考察2008年研究评估活动（RAE）中“地理学与环境研究”领域的出版数据（N= 4 590 项书面成果），作者得出了相同的结论。他们指出，“自然地理学家和人文地理学家在选择期刊时都相当广泛”（Richards 等，2009年，第234页）。地理学家的知识共同体（具备体裁知识与技能，能够研究建成环境这一问题）始终致力于跨越较大的学科领域，尽管住房的作用——这必然是一个空间问题——可能被忽略了。现代移民研究从各个不同领域发展而来，成为一项跨学科研究（Levy等，2020年，第23页）。因此，了解这些发展有助于理解与建成环境相关的移民问题。

1.4 迁移文献的增长

Levy等（2020年，第20页）指出，在1975年至2018年期间，迁移文献语料库（N = 49k）中仅出现一次的引文来源增加了14倍。这引发了人们对普通移民研究可能参与并交流的认识共同体的担忧。实证研究局限于特定的情境和时间范围，其性质可能使研究发现难以推广，因而也难以与其他研究建立关联。或许正因如此，越来越多的移民研究成果与其他学科的研究相比，获得的引文分数较低。

平均而言，社会科学领域，尤其是移民领域的一篇学术论文，最多会引用 30–40 个来源（Levy、Pisarevskaya 和 Scholten，2020年，第 20 页）。随着移民相关出版物不断增加，考虑到出版物篇幅有限，而可能相关的既有研究数量庞大，我们也会预期被共同引用的研究会越来越少。例如，仅在人口学期刊中发表的移民文章，据估计，截至 2017年每年就超过 330 篇；而半个多世纪前，每年仅有 0–9 篇（van Dalen，2018年）！

大多数移民研究以论文或专著的形式发表（Pisarevskaya 等人，2020年，第 457 页），刊载于许多跨学科期刊上。此类学术成果的篇幅因来源而异。不过，论文稿件平均可按 25–40 页双倍行距文本估算，其中包括必要的数据呈现（Borja，2015年）。保守估计的话，可以将平均研究文档视为一篇正文 20 页的论文，另有 5 至 20 页用于图表和参考文献。双倍行距文本每页约 250 词（ HCC 图书馆，2020 年）；因此，保守估计一份文档的篇幅约为 5,000 词。假设阅读速度正常（Duggan 和 Payne 估计为每分钟 225 词（2009年，第 15 页）），且不会感到疲劳，那么阅读一份 5,000 词的研究成果至少需要 22 分钟。要人工读完本文后续分析的论文语料库，至少需要 90 年——这意味着必须终身专心阅读，且没有假期！

1.4.1 文献映射

文献报告科学——或将文献作为数据加以利用——在过去半个世纪里，随着计算技术的发展和研究成果的数字化而不断进步（Fortunato 等，2018年，第359页）。

其中一些重要里程碑包括：1961年科学引文索引（CSI）的创立（Cozzens，1981年，第16页），推动了文献计量学——即对书面交流进行"定量研究"——以及科学计量学的发展；后者通过引文研究"科学活动"，例如科学社会学（Hood 和 Wilson，2001年，第298–299页）。文献计量分析的不同方法，可表示为严谨性（可重复性）与投入（手动或程序化）之间的函数关系。图1.1展示了审阅学术文献的多种方式；若将可重复性纳入考量，曲线则会陡然上升。尽管迁移研究具有跨学科性质，因此很适合采用文献计量分析，但直到近年才出现少数纳入这一方法的先例（Nestorowicz 和 Anacka，2018年；Piguet、Kaenzig 和 Guélat，2018年；Sweileh 等，2018年；van Dalen，2018年；Levy、Pisarevskaya 和 Scholten，2020年；Pisarevskaya 等，2020年）。这可能是因为社会科学的索引不如自然科学明晰，而且迁移研究还是一个相对年轻的领域。

在医学领域，关于相关学术研究批判性评估的文献报告最为成熟。这是因为，不同群体之间的受控实验以及详细的文章索引编制，非常适合遵循预先确定的综述方案。在自然科学之外，由于缺乏两种资源，遵循方案的"真正"系统综述并不总是可行。首先，这类综述通常由至少三人组成团队完成，以便在决定哪些文章相关时加入一名打破平局的评审者；而单人研究并不总能做到这一点。其次，在不同数据库之间缺乏整合的元信息、因而无法进行穷尽且可重复搜索的情况下，综述方案也无法预先制定（Roth，2020年）。在这种情况下，可以采用包含更多判断的替代方法。这些方法显示在图1.1的左侧。最左端是快速综述或前沿综述，通常在五周以内完成，用于收集有关某一问题的证据（Roth，2021b）。再向右则是传统的（叙述性）综述，其特点是对既有工作的独特但不可复现的描述和评价，以便为新的研究提供背景（Roth，2020年）。在此之外，研究便开始进入利用数据驱动的计算方法评估既有文献的领域。

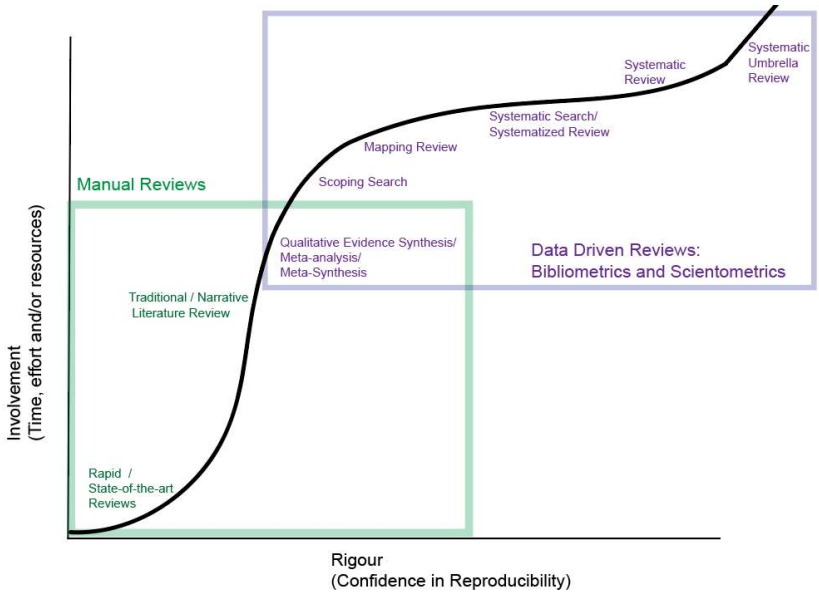

图*1.1.* 文献报告的方法论路径

图1.1最右端是正式的“系统综述”，其采用“可问责的严谨研究方法”（Gough、Oliver和Thomas，2017年，第2页），并要求“至少两名评审者，另加一名评审者负责打破平局”（Roth，2020年）。这一谱系一端关于学术文献报告的例子，包括构成Cochrane综述的方案家族（Higgins 等，2019年），以及系统综述和荟萃分析优先报告条目（Preferred Reporting Items for Systematic Reviews and Meta-Analyses）——PRISMA提供的指南（Moher 等，2009年）。比这些更高一层的是“伞式综述”，即“将多项综述综合为一份易于获取和使用的文件”（Grant和Booth，2009年，第95页）。在数据驱动模型中，这类综述将通过自动化管道生成：该管道把与某一查询相匹配的论文存储在相关文献数据库中。一个半自动化的例子是将气候变化与移民研究整合到CLIMIG数据库中的项目（Piguet、Kaenzig和Guélat，2018年，第375页；CLIMIG，2020年）。严谨程度最高的是元综合，可以是定量的，也可以是定性的。这些方法的共同点在于：对于“荟萃分析”，它们大体上都依赖研究者对数值一致性以及统计结果可接受敏感性的判断（Deeks、Higgins和Altman，2019年，第241页）；或者，对于“定性证据综合”，依赖研究者对主题兼容性和概念一致性的判断与解释（Grant和Booth，2009年，第94页；Roth，2021a）。用于理解移民的多种方法论路径和数据覆盖范围（DeWind，2020年，第5页），可能使单独进行元综合变得困难，并可能产生误导。

在这类情况下，经过叙述性综述界定范围后，“文献映射”可提供一种替代方式来报告相关文献。这种方法旨在“描绘并归类现有文献”（Grant 和 Booth，2009年，第94页），适用于该领域不遵循严格的出版物关键词分类，或研究成果本身具有多样性、因而索引编制也各不相同的情况。文献映射有助于追溯研究谱系的发展，并构成“系统化综述”的基础；当无法开展系统综述时，系统化综述可针对某一主题提供“最佳证据综合”（Grant 和 Booth，2009年，第94页）。

1.5 方法论

本节将说明如何收集迁移与住房交叉领域论文所采用的具体方法。本分析所用的书目数据语料库包含 N = 445k 份文献，数据来自开放获取的谷歌学术数据库，时间覆盖范围为1885年至2020年。选择这一时间跨度，是从 Ravenstein 的第一篇论文起始，以便概述迁移研究的发展。图1.2总结了本分析所用论文语料库的构建流程。以下内容将逐步说明图表中的各项步骤。

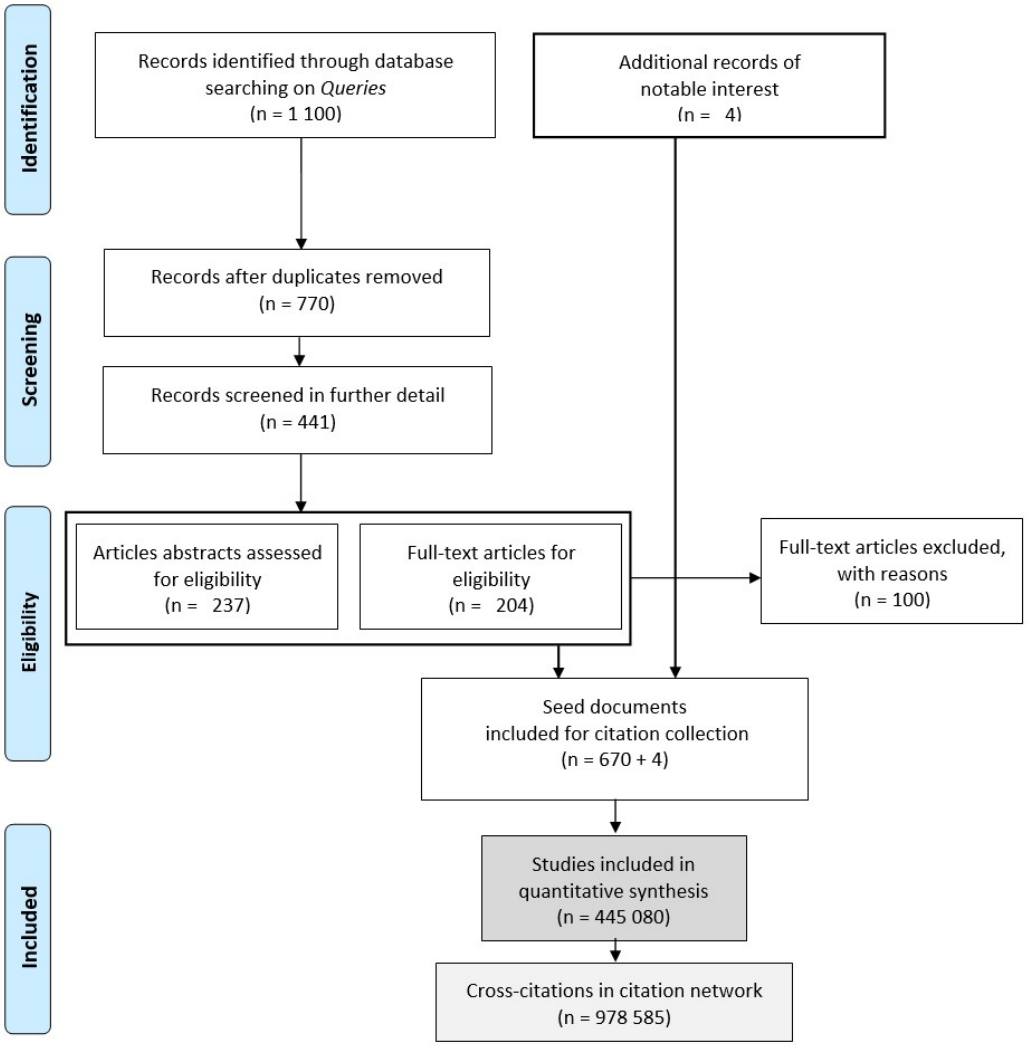

图 *1.2.* 根据 *PRISMA* 指南纳入论文的流程图

1.5.1 谷歌学术数据库

主要的书目数据库提供商包括 Scopus（爱思唯尔）、Web of Science（WoS）以及可免费使用的谷歌学术)16年，第115页）。与另外两个主要引文数据库相比，谷歌学术的覆盖范围更广，尤其是在 Humanities 和社会科学领域（Martín-Martín 等，2018年，第20页）。使用谷歌学术数据库的一个优势是，它对各种学科以及

专著的覆盖范围都很广，而不仅限于期刊文章。这一点至关重要，因为书籍很难通过结构更严谨的文献计量方法进行收录，但它们仍是地理学和建成环境研究中常见的成果形式（Richards 等，2009年，第12页）。不过，引文数据通常受许可约束，并由专有算法收集，这意味着此类数据要么无法像 Scopus 和 WoS 中的数据那样免费获取，要么虽可开放获取，但存在元数据访问限制（Martín-Martín、Thelwall 和 López-Cózar，2020年，第6页）。

1.5.1.1 抓取谷歌学术

书目数据由我们与 Juste Raimbualt 博士共同创建。我们使用网页抓取工具，模拟在谷歌学术上进行匿名请求的手动搜索——也就是说，搜索时未登录用户账户，也未使用网站 Cookie，详见 Raimbault（2017，第 5 页）。对于每个查询（图 1.3-1），数据收集的第一步是获取每个查询对应的前 100 条论文检索结果。图 1.3-2 展示了谷歌学术排名算法如何对排名靠前的结果进行排序。第一步创建了一个范围界定语料库，随后通过人工检查，从中筛选出论文种子语料库。对结果进行人工合理性检查后，接下来的数据收集步骤是获取引用种子语料库中每篇文档的前 100 篇文档（图 1.3-3）。Raimbault（2017）在"数据库构建"部分完整介绍了用于协作收集数据的工具。

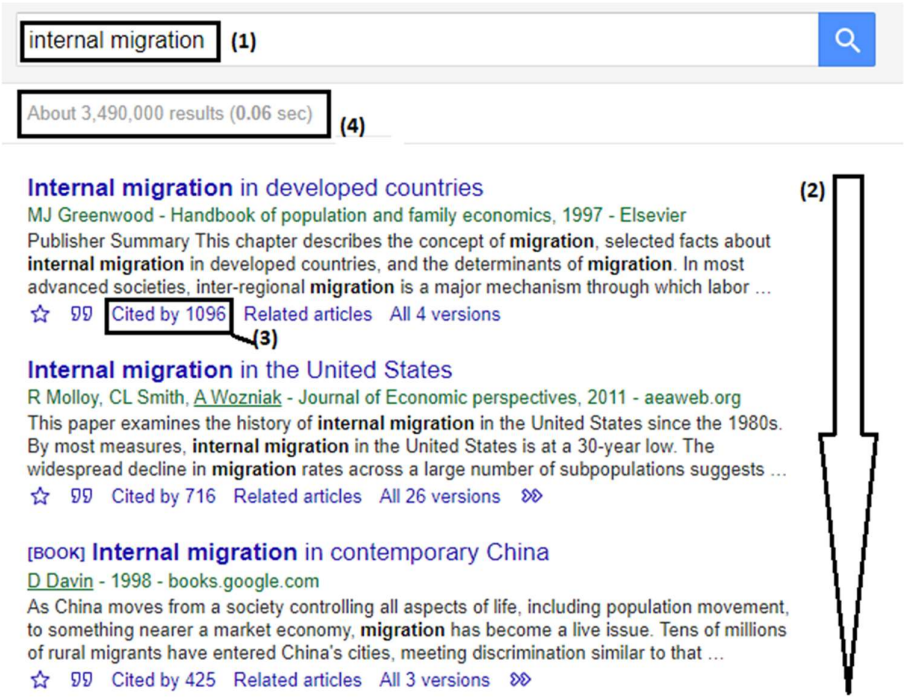

图*1.3* 检索式与论文的关系

我们使用了十一个检索式（如图1.6所示），以全面收集与移民和住房相关的论文，并建立一个用于人工检查的论文范围界定语料库。由于检索式数量众多，而谷歌学术允许的请求次数有限，收集论文耗时数周。

最耗时的阶段是数据收集的第二步——收集施引论文——因为论文之间的链接呈指数级增长。引文收集分为两层（深度= 2）。这意味着收集的论文是那些最多沿参考文献追溯两层、引用种子语料库的论文（见图 1.3-3）。

第一层是纳入引用种子语料库中每篇论文的排名前 100 篇文章（见图 1.3-3）后形成的。第二层引文则以类似方式，通过收集引用第一层文献的排名前 100 篇论文而形成。两层之间的关系以图 1.4 中的示例说明。在这一阶段收集的所有文献共同构成了本分析所用的文献引文语料库。

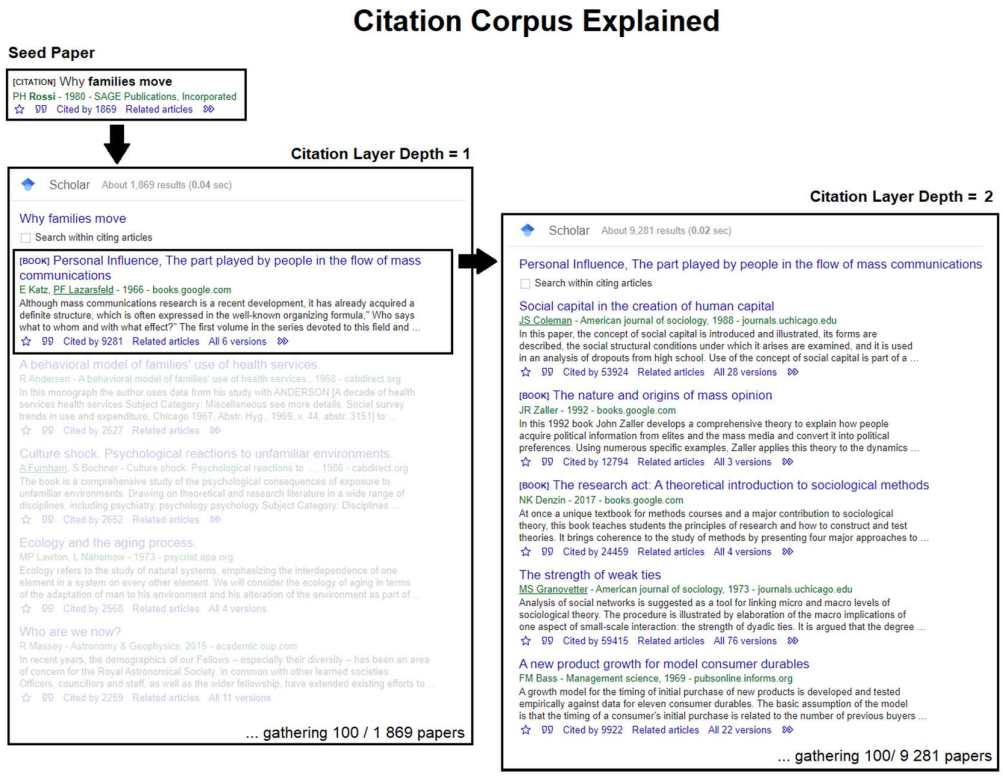

图*1.4* 引文语料库的层次

1.5.2 检索式

方法论中最重要的部分是选择检索式和统一的关键词，因为这些因素最终决定了纳入最终语料库的论文范围。本节接下来将介绍选择关键词、检索式以及四篇重要参考文献的逻辑。

1.5.2.1 检索式关键词的选择

需要注意的是，第一阶段的目标是在研究范围内全面收集论文。因此，尽管谷歌学术没有叙词表分类（Gibson，2020年），但参考其他数据库对关键词的分类仍然很有帮助。检索式条目的措辞（如图1.3-1所示）参考了人工文献探索以及对各种图书馆数据库叙词表的考察，其中尤其包括由英国数据档案馆编制的人文与社会科学电子叙词表（HASSET）（英国数据档案馆，2020年）。

HASSET 同义词层级结构中值得关注的关键词如图1.5所示。这些关键词表明，“内部迁移”（IM）和“居住流动”（RM）都是“人口迁移”下的狭义术语。有趣的是，作为关键词组合的 RM 在关于内部迁移的学术研究中属于一个小众子术语，

不过，它也作为一个突出的主题，直接归属于有关“生命史”的更广泛学术研究，这表明 RM 汇聚了认识共同体，为迁移研究增添了纵向视角。图1.5中还值得注意的一点是，RM 是通向住房相关术语的枢纽节点。它与“住房需求”“搬迁费用”相关，也用于“搬家”和“住房史”。RM 还带有一丝人口学色彩——因为人会随着时间推移而变老——其狭义研究领域关注青年流动和“离家”。

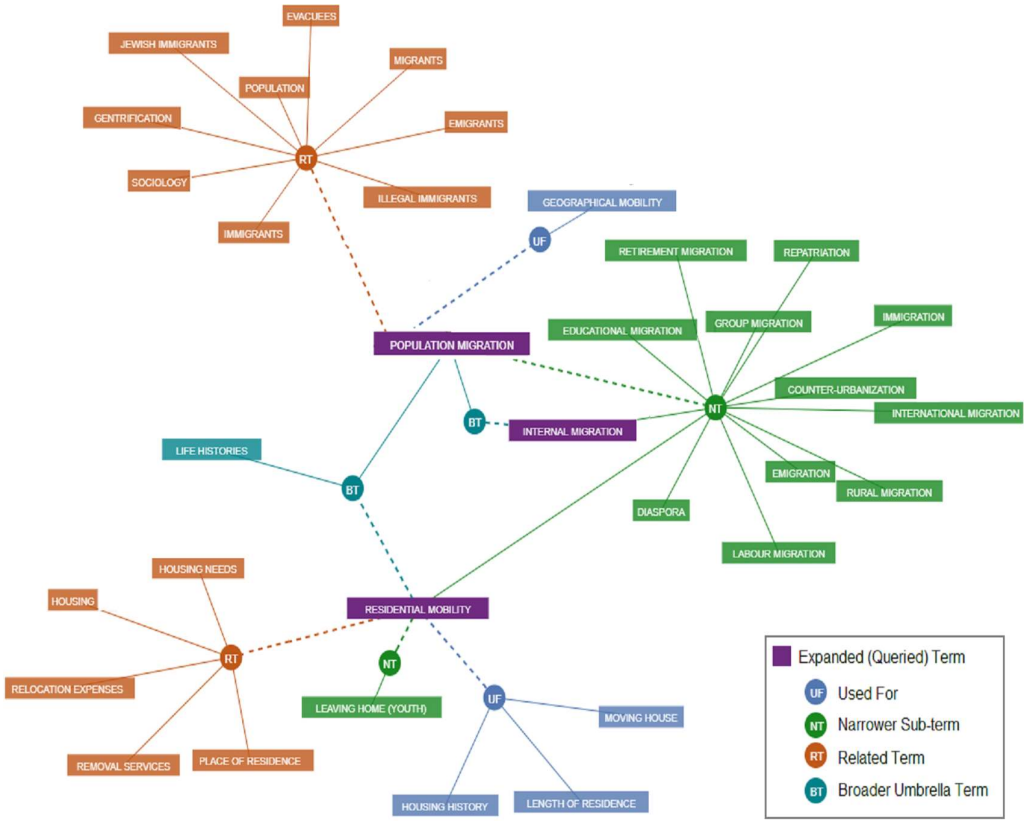

图*1.5. Hasset*词库中“居住流动”作为关键词的关系图。

注：为使各元素适应页面，图中缩短了距离。

查阅英国数据档案馆的人文与社会科学电子叙词表（HASSET），有助于指引关键词组合和相关文献的检索。它印证了通过无指导阅读形成的直觉，并为下一节构建了检索式。

1.5.2.2 查询概述

本分析中的迁移文献涵盖四种理论层面的细微差异，即体裁知识子领域，如图1.6所示。范围最宽泛的检索式是“内部迁移与住房”。原本预计该检索式的结果将以英语语系国家案例研究为主，但其占比究竟有多大尚不明确。因此，特意将这一检索式设定得较为宽泛。

其余三个术语分别是：“居住流动”——带有社会学意味，暗示研究聚焦于个体层面；“内部迁移”——英国文献偏好的术语；以及“城市迁移”——在北美学术研究中，尤其是在涉及城市住房相关迁移的语境下使用的同义术语。城市

迁移之所以被纳入，是因为研究者预计城市环境是内部迁移与住房相互关联的场所。然而，由于“城市迁移”在全球南方与北方所指的现象大相径庭，因此进一步通过主题和地理限定词缩小了检索范围。未纳入“人口迁移”，因为在谷歌学术中占比过高的自然科学与占比较低的社会科学之间，该术语的含义模糊且有所不同。

上述三个与理论相关的检索式进一步缩小范围，以便聚焦并穷尽所有相关论文。为此，在检索式中加入了限定词，指定相关语境（Britain OR UK OR England）、感兴趣的数据集（“人口普查”）以及住房子主题（当单独使用“住房”过于宽泛、无法检索到相关结果时，则使用“住房市场”）。最终形成的查询字符串见图1.6。

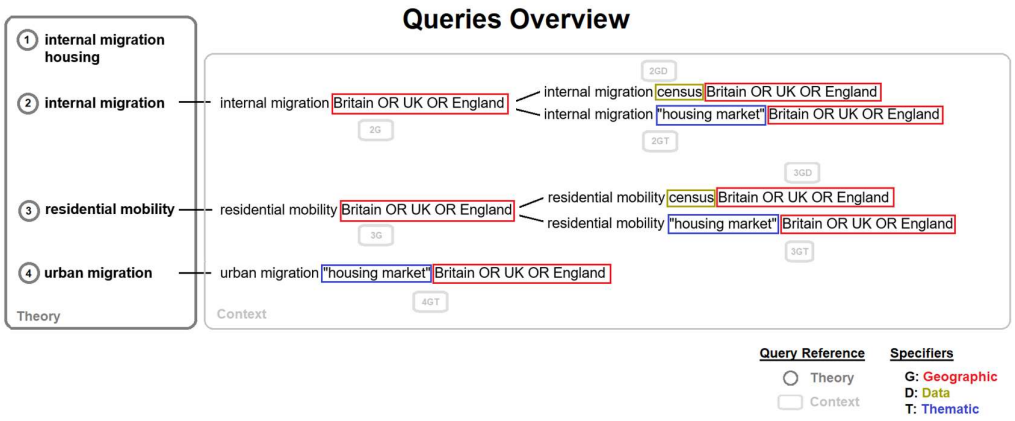

图*1.6*。检索式概览

1.5.3 手动筛选

选定检索式后，第二个关键步骤是对范围界定语料库中的论文进行人工评估——该语料库由首次抓取的文档组成。通过这一步更细致的人工审查，从范围界定语料库的760篇论文中筛除100篇。论文筛除分为三个步骤。第一步仅查看标题，将其分为五组，并按相关程度排序：“4 - 待查看”（n=107）、“3 - 合适”（n = 212）、“2 - 核查”（n=137）、“1 - 边缘相关”（n = 81）、“0 – 不太可能相关”（n = 223）。最后三组被合并，进行更耗时的评估，包括查看摘要（n = 237）或全文（n = 204）。评估重点放在最后三组，因为与漏掉检索结果中可能有用的论文相比，剔除无关论文更为有利。例如，“2 - 核查”组包含一些案例研究论文，这些论文因其文献综述的质量而被纳入，尤其是其对移民理论的参考。此次人工评估中，接受全文审查的标题略少于三分之一。

筛选并精简符合条件的论文集（N = 660）后，如图 1.2 所示，这些论文被用作收集论文、构建最终语料库的主要“种子文献”。种子论文还补充了第 1.5.4 节讨论的四篇其他重要参考文献，因此种子论文总数为 N = 664。对于被引用次数超过 100 次的种子论文（如图 1.3-3 所示的示例），仅纳入按“被引用次数”排名前 100 的论文。这样

便构建出一个引文网络，其中包含经过人工筛选的符合条件的论文，以及这些种子文献学术谱系中最受欢迎的论文。两层足以实现广泛覆盖，原因有二：引文网络会随着每篇新论文的加入而呈指数级增长；此外，重要参考文献中还包括较早的论文，例如 Ravenstein（1889年）。

1.5.4 重要参考文献

除图1.6中的检索式外，本项目框架构建所需的四篇关键论文也被纳入，作为最终语料库的种子文献。这些论文列于表1-2。尽管在此阶段纳入更多论文颇具诱惑力，但我们谨慎地选择了感兴趣的参考文献。这是因为引文具有指数级增长的特性，每增加一篇论文，都可能使最终语料库增加多达100 000篇论文（两个引文层级各有100篇论文，即图1.4所示的100×100篇）。

纳入的四篇论文体现了一个特定主题，且均为在叙述性文献综述研究过程中经探索选出的高被引论文。之所以纳入Ravenstein（1889），是因为这篇论文被广泛视为现代移民研究的起点，同时也是定量人口地理学论文脉络中的首篇（毕竟Ravenstein本人就是一位地理学家！）。Sjaastad（1962）同样重要，他在宏观层面以定量方法考察移民。然而，与Ravenstein不同，Sjaastad从鲜明的经济视角出发，将移民置于资源配置和收入差异的框架下进行探讨。Sjaastad（1962）也是内部迁移和国际移民领域最重要的五篇论文之一（Nestorowicz和Anacka，2018年，第295页）。

为补充自上而下理论，最后两篇参考文献——Lee（1966）和 Rossi（1955）——被纳入其中，以体现理解移民特征所需的微观层面方法。Lee（1966）提出了针对个体的“中间障碍”这一概念，拓展了理论学术研究；理解移民这一具有显著人口学差异的经历，这一概念至关重要。迄今提及的三篇论文按此顺序被列为移民研究的“经典著作”（King，2020年，第4页）——在本文收集数据后回顾时才注意到这一有趣的巧合。最后一篇值得关注的参考文献是 Rossi（1955）。这部专著为社会学探究家庭单位在移民过程中随时间推移所发挥的作用提供了起点。

	标题	作者	*Year*	期刊	引文
(1)	《移民法则》	EG Ravenstein	1885	伦敦统计学会期刊 伦敦	5012
(2)	人类迁移的成本与收益 人口迁移	LA Sjaastad	1962	政治经济学杂志	6057
(3)	迁移理论	ES Lee	1966	《人口学》	6592
(4)	家庭为何迁移	PH Rossi	1980	SAGE 出版社	1833

注：被引次数截至*2020*年*8*月表*1-2* 关键论文

1.5.5 自然语言处理

文档标题统一转换为小写，并清除数字及所有标点符号，但保留复合词中的连字符。研究采用词袋方法识别出现频率最高的词项：将文档标题切分为单个词项，再汇总在一起。此外，为了突出最重要的词项，研究从 SMART 和 Snowball 词表中选取了 174 个常见英语停用词，并

将其删除。这些词包括“the”、“of”、“from”等（quanteda，2019），它们几乎无法提供主题方面的见解。

使用词项-文档矩阵表示法来计算词语之间的相关性得分，即两个词语在同一标题中共同出现的可能性。这是一个二元矩阵，每个文档标题表示为一列，语料库中的所有可能词语则列为行。由于这会生成一个非常稀疏的矩阵（445k 篇文档中有 213k 个不重复词项），因此仅纳入出现次数超过 50 次的词项，并使用 Porter 算法对词语进行词干提取（Porter，2006）。词干提取是通过去除常见词尾来规范化词项，使词语仅保留词根或“词干”，从而将同一词语的不同变体归为一组。通过这一过程降低词项的分辨率，仅对评估相关性得分有用。这是因为该语料库中几个不同的关注词具有相同的词干；具体而言，“house”“housing”和“household”都会被简化为“hous”。

我们使用了谷歌紧凑型语言检测器 2 的 R 语言实现来识别标题的语言。这是一种概率模型，通过搜索某种语言特有的二元组和三元组组合，从而识别 83 种语言之一（CLD2Owners，2021）。尽管其算法已被神经网络模型（CLD3）取代，但研究发现，早期实现的表现更好，在包含 44.5 万份文档的引文语料库中，成功翻译标题的覆盖范围超过 90%。

1.6 结果：

1.6.1 范围界定标题

1.6.1.1 顶级标题

根据谷歌学术专有但公开可用的排名算法，抓取技术会返回最相关的前 100 篇论文。打个比方：返回的排名前一百篇文章，只是与搜索词相关的海量文档中的冰山一角。每个查询的排名第一标题如下所示。

		Queries										
		1_IMH	2_IM				3_RM				4_UM			
Title	*Year*			2G	2GD	2GT		3G	3GD	3GT		4GT	*Authors*	*Cited*
Does land use planning shape regional economies? A simultaneous analysis of housing supply, internal migration and local employment growth in the Netherlands	2009	1				•							Vermeulen and Ommeren	63
Internal migration and ethnic groups: evidence for Britain from the 2001 Census	2008	•	•	3	1	2							Finney and Simpson	132
Population movement within the UK	2005			•	•	1							Champion	117
Intra-urban migration and housing submarkets: Theory and evidence	2004										•	1	Jones and Leishman	99
Migration in a mature economy: emigration and internal migration in England and Wales 1861-1900	2003		•	1	•								Baines	396
Tied down or room to move? Investigating the relationships between housing tenure, employment status and residential mobility in Britain	2002					•	•		1	2	•		•	Böheim and Taylor	241
Internal migration in developed countries	1997		1	•									Greenwood	1090
Spatial mobility, tenure mobility, and emerging social divisions in the UK housing market	1987									1			Forrest	82
Residential satisfaction as an intervening variable in residential mobility	1974						1						Speare	876
Life-cycle, housing tenure and intra-urban residential mobility: A causal model	1973							•	1	•		•	Pickvance	86
The intra-urban migration process: a perspective	1970										1		Brown and Moore	1223

Note: Cited by counts are correct as of August 2020; Ranks 1-3 included otherwise shown as •

表*1-3.* 每个单个查询的顶级标题

图1.6显示，当排名靠前的文献对应多个检索式时，检索式的主题范围较窄。图中还显示，最早的论文（1970年）同时也是“城市迁移”（4_UM）检索式排名最高的结果，其引文数最多，共有 1 223 条引文。此外，图1.6中的理论检索式，其结果数量级几乎都至少比按地理背景筛选后的排名最高结果高一个数量级。唯一的例外是“内部迁移与住房”（1_IMH）。该检索式的首篇文献既是列表中最新的论文，也是引文数最少的论文，这表明关于内部迁移与住房相结合的学术研究是一个相对较新的子领域。

1.6.1.2 论文筛除

上一节介绍了相关论文的类型学，本节则将讨论哪些论文相关性较低，因此在开始引文收集之前被剔除。范围界定语料库（n=760）中的十一种检索式均在手动筛选过程中剔除了部分论文。被剔除的论文大致分为三类，包括：医学论文（例如，“差异性居住流动导致的选择偏差，作为电线编码与儿童癌症关联的一种解释”）；过于具体、无法与英国进行比较的描述性外国案例研究（例如，“马来西亚巴生谷的内部迁移：问题与影响”）；以及1800年以前的历史研究（例如，“聚落中有什么？青铜时代早期英格兰南部的家庭实践与居住流动”）。本阶段论文筛除所采用的逐步方法将在1.5.3节讨论。

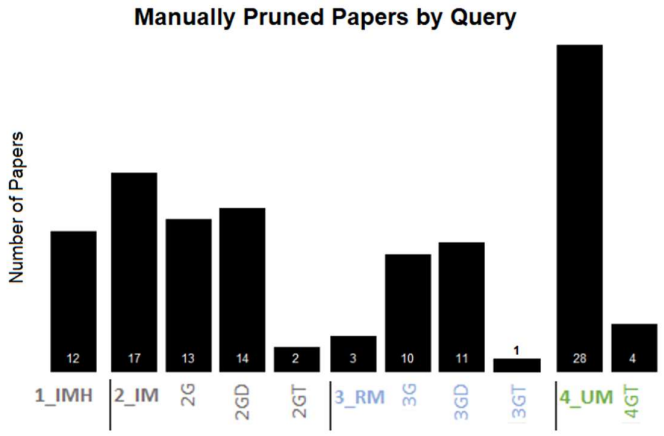

图*1.7* 按检索组手动筛选的论文

不出所料，图1.7显示，最宽泛的检索式——即与理论相关的检索式——包含数量最多的切题性较低的论文。其中，城市迁移的遗漏量最大。这一关键词组合通常与发展中国家的案例研究文献相关。相比之下，将作为主题（T）的“住房”与相关地理限定词（G）结合后返回的论文几乎全部都被保留。

有趣的是，作为搜索词的“居住流动”与少数几篇使用同位素分析测量移民的历史论文相关（例如《英国北约克郡西赫斯勒顿的盎格鲁-撒克逊人居住流动》

该论文题为《基于氧和锶同位素分析》，发表于一部关于等离子体质谱法的编著中）。在这一阶段被剔除的最有趣的异常标题来自民族音乐学领域，题为《远离寂静：秘鲁高原的音乐与城市迁移体验》。这一阶段共系统剔除了一百篇论文（详见第1.5.3节的讨论），最终种子论文语料库中剩余 N = 660 篇论文。

1.6.1.3 四种主题类型

不同学科可能会以不同方式使用语言来赋予概念意义并构建研究框架。体裁知识中的术语由此可以界定认识共同体，并为推进移民研究创建学科特定的封闭领域。这意味着，要全面识别相关论文，就需要考虑移民研究的不同主题类型。用于创建范围界定语料库的主题类型对应于图1.6中的四个理论组。

图1.8是维恩图，展示了图1.6所示的四个范围最广的理论检索式所检出的种子论文。图中清楚表明，关于居住流动的热门文献形成了一个独立的学术研究领域，与内部迁移和城市迁移相区别。图中还显示，住房这一主题仅出现在有关内部迁移的热门且影响力最大（按引文数衡量）论文中的一小部分：在内部迁移检索结果排名前一百的文献中，只有25篇论文（25%）提及住房。最后，图1.8显示，与回返迁移类似，城市迁移也较少与内部迁移领域的认识共同体联系起来。

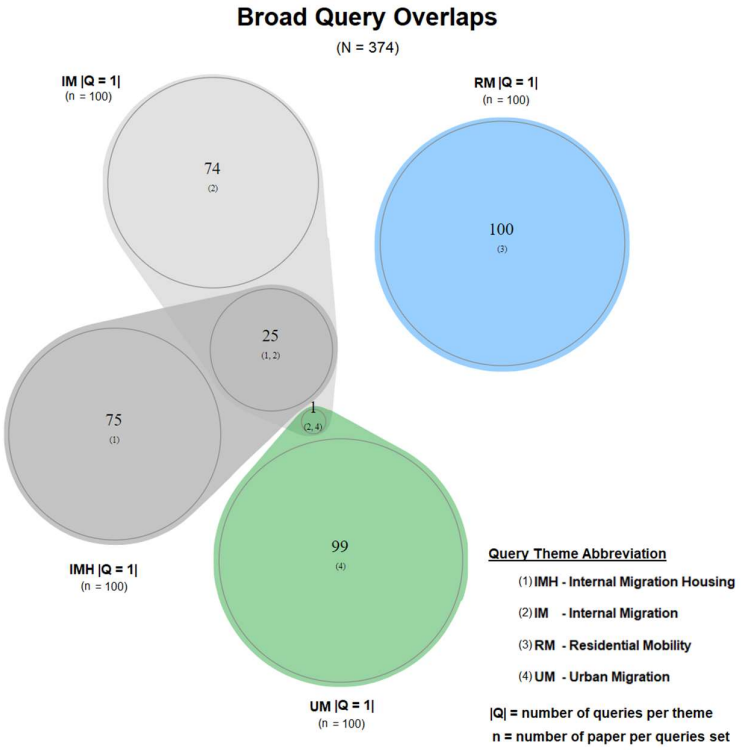

图*1.8*。关键搜索词的种子结果

最重要的是，图1.8显示，从回返迁移和城市迁移视角研究流动性的成果，与内部迁移领域的一般学术研究相距甚远。然而，相距甚远并不意味着彼此独立。这是因为，只要考察的论文足够多，引文的指数级增长意味着任意两篇论文最终都会产生关联。

1.6.1.4 主题类型的重叠

为了收集第1.6.1.1节所讨论的各类型交叉领域的论文，我们进一步按照地理学、数据和主题对其进行了筛选，如图1.6所示。值得注意的是，范围语料库中独一无二的论文共760篇，而非1100篇；如果十一个检索式各自返回100篇互不重叠的论文，结果本应是1100篇。

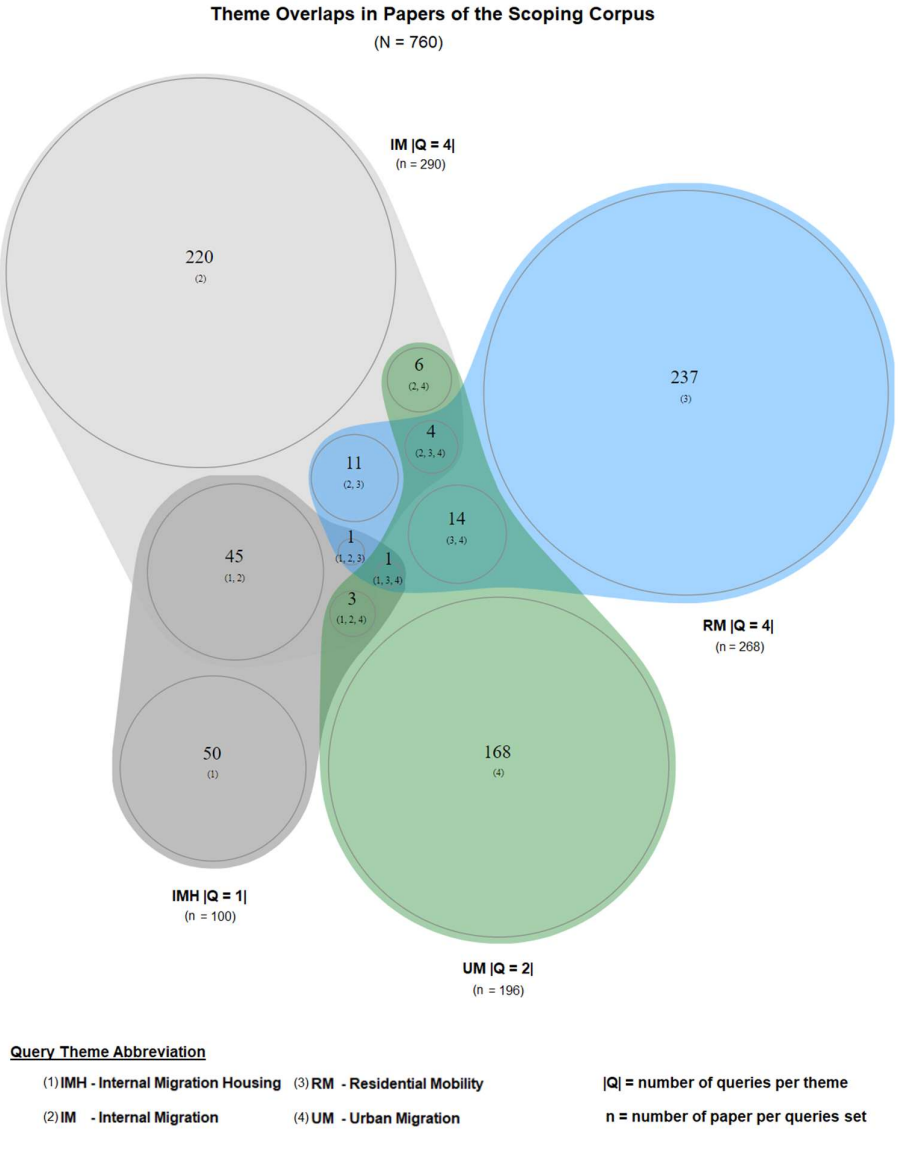

图*1.9* 种子结果：论文主题的重叠

设置十一个检索式的目的是做到全面。然而，考察十一个检索组会显得繁琐，因此这里按其主题类型展示。按理论类别划分的范围界定语料库总量见图1.9。这与图1.8所示结果一致：即使增加了检索限定符，

根据最热门论文的衡量结果，四个认识共同体彼此相距甚远。维恩图展示了被多个检索组覆盖的论文数量。一篇论文可能同时出现在多个检索式中——也就是说，它可能在多次谷歌学术搜索中都成为排名靠前的结果。如果将内部迁移和 IHM 合并考虑，那么理论主题之间仅有 40/760 篇论文存在重叠，占文献总数的 5%。

1.6.1.5 查询重叠

图1.10按主题类型展示了种子语料库中的论文在全部十一个检索式之间的重叠情况。对于所有论文，可以将每篇论文视为相交查询集中的一个元素。然而，由于十一个集合难以在维恩图中清晰呈现并加以理解，图1.10以列的形式展示查询交集。图1.10中的行表示查询字符串。

每列顶部的条形图显示符合相应查询交集模式的论文总数（这相当于维恩图中的面积）。查询行按照其理论框架（参见图1.6）分组并置于方框中。因此，每行要么是四个理论查询缩写之一（以星号标示），要么是这些缩写的较窄搜索限定词版本（图1.10右下方）。

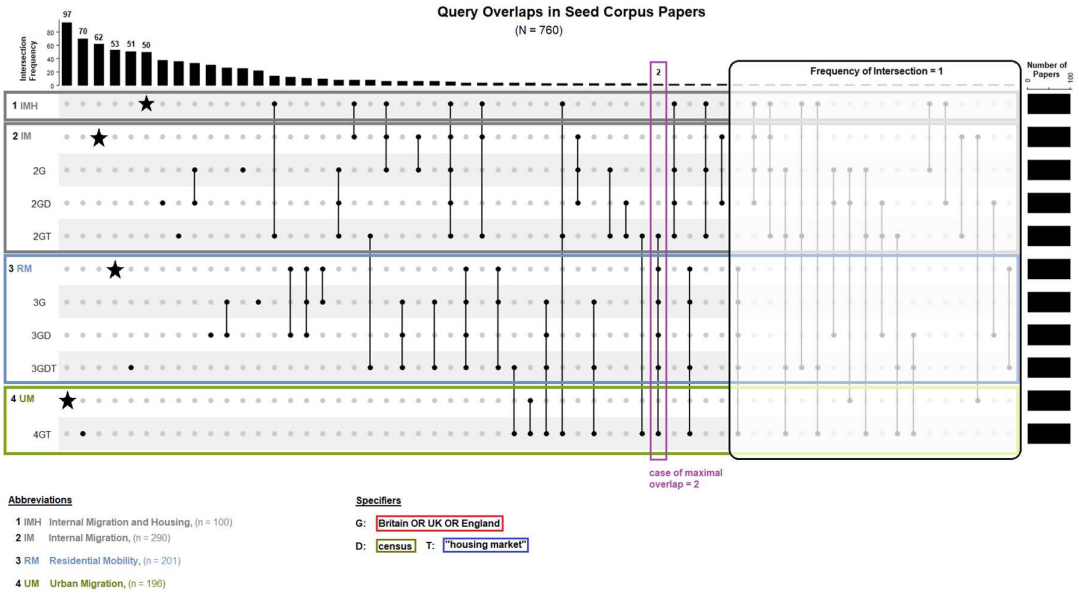

图*1.10* 按论文频次划分的查询交集

图1.10首先值得注意的是，正如预期，范围最广的四个理论检索式（IHM、RM、内部迁移和城市迁移，以星号标示）检索到的独有论文数量最多。不过，值得注意的是，“城市迁移”（UM）最具独特性（100篇中有97篇为独有论文）；这体现了该短语在全球北方和南方文献中的预期含义差异：在全球北方，它指城市内部的迁移；在南方，它指迁往城市地区。独有论文数量次多的一组是“内部迁移”（IM）的检索结果（100篇中有70篇为独有论文）。这是因为，按定义，内部迁移受国家边界限制，而相关研究可能不涉及或未纳入与英国背景的比较。另一个值得注意的情况是

最大重叠是两篇论文同时匹配六个检索式（图1.10中以紫色标示）；这两篇论文都涵盖了所有与RM相关的检索式。

1.6.1.6 最大重叠论文

图1.11展示了检索式之间重叠数量的详细情况，表明双重重叠最为常见。这是采用穷尽式方法所致。双重重叠对应于图1.6所示的“理论查询”匹配，以及将同一理论查询进一步缩小范围，使其包含与英国或人口普查数据的关联或比较。表1-4列出了存在重叠的论文。

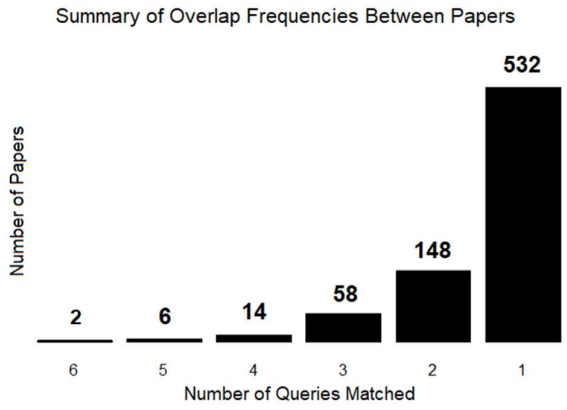

图*1.11*。十一项查询的重叠

标题	Hits	Year	引用次数	作者
住房市场区域的定义与战略规划	6	2002	113	Jones
搬迁住所，营造家园：探索居住流动性	6	2002	182	Winstanley 等 al.
维多利亚时期城市中的两次人口普查期间的流动	5	1977	44	Dennis
唐人街的进出：纽约市华人的居住流动与隔离 纽约市华人	5	1991	173	Zhou 和 Logan
土地利用规划会影响区域经济吗？一项同步 住房供给、内部迁移与地方就业的分析 荷兰的增长	5	2009	63	Vermeulen 和 Ommeren
东欧地区的区域间迁移与住房结构 转型国家：立陶宛视角，2001—2008年	5	2009	29	Bloze
族群隔离与居住流动：少数 荷兰族裔群体的迁居	5	2010	138	Bolt 和 Kempen
远距离-逆-城市迁移至边缘地区的解释 丹麦的区域	5	2011	43	Andersen

表*1-4* 最大重叠的论文

表1-4显示，同时被许多检索式涵盖的论文寥寥无几。例如，Jones（2002）在对格拉斯哥进行实证分析时指出，住房市场区域地理划定的关键变量应当是“迁移自我封闭性”（2002，第549页）。值得注意的是，他回顾了学者（可能为政策提供建议）和规划人员（将政策指导转化为实地决策）对住房市场区域的不同理解和界定。相比之下，Winstanley等人（2002）采用定性民族志方法，通过访谈新西兰基督城的家庭，探讨“房屋”和“家”这两个主题随时间变化的含义，并凸显了其复杂性

以及流动决策的复杂性。本文借鉴Rossi（1955）的学术研究，运用轶事来讨论作为家庭住房需求函数的调整性迁移，从而承认生命历程轨迹可能存在多样性。

表1-4中的论文展现了将地方纳入考量的各种方式。大体而言，有三种方法。首先，Jones（2002年）——与 Vermeulen 和 Ommeren（2009年）、Bloze（2009年）及 Andersen（2011年）一样——通过对特定且有名称的地理区域进行实证研究，在关注空间细微差异的同时采用生态宏观方法。相比之下，Winstanley等人（2002年）——与 Zhou 和 Logan（1991年）一样——采用轶事性访谈方法，在这种方法中，地方的重要性取决于个人经验。这两种方法分别代表了呈现地理学作用的两个极端：一是聚焦于“有名称的”地方，从而更深入地了解地方的特征；二是聚焦于“有名称的”人，从而更深入地了解个人的动机。

最后，将地方与社会过程联系起来的折中路径，是把地方作为一种类别。这需要一种差异度量，可应用于建成环境，根据其特征划分地理区域，并依据对地方的概括性描述，将这些地方归为大致“相同”或“不同”。在这种方法中，关注的重点不再是“有名称的”地方或人，而是地方和人的“类型”。这种折中方法通常采用包含丰富特征的人口普查数据或家庭面板数据；Dennis（1977年）和 Bolt 与 van Kempen（2010年）的研究分别采用了这两类数据。

1.6.2 种子标题

1.6.2.1 主题及其子查询

假设采用极为保守的费米估算，即按照第 0 节推导的结果，每 22 分钟阅读一份文档，那么包含 N = 660 篇论文的种子语料库，以每周 37 小时的速度“阅读”，将需要近 7 周。因此，为了更高效地了解种子语料库，本节将根据文档标题的措辞，简要概述其中的关键主题。

V1_10172022

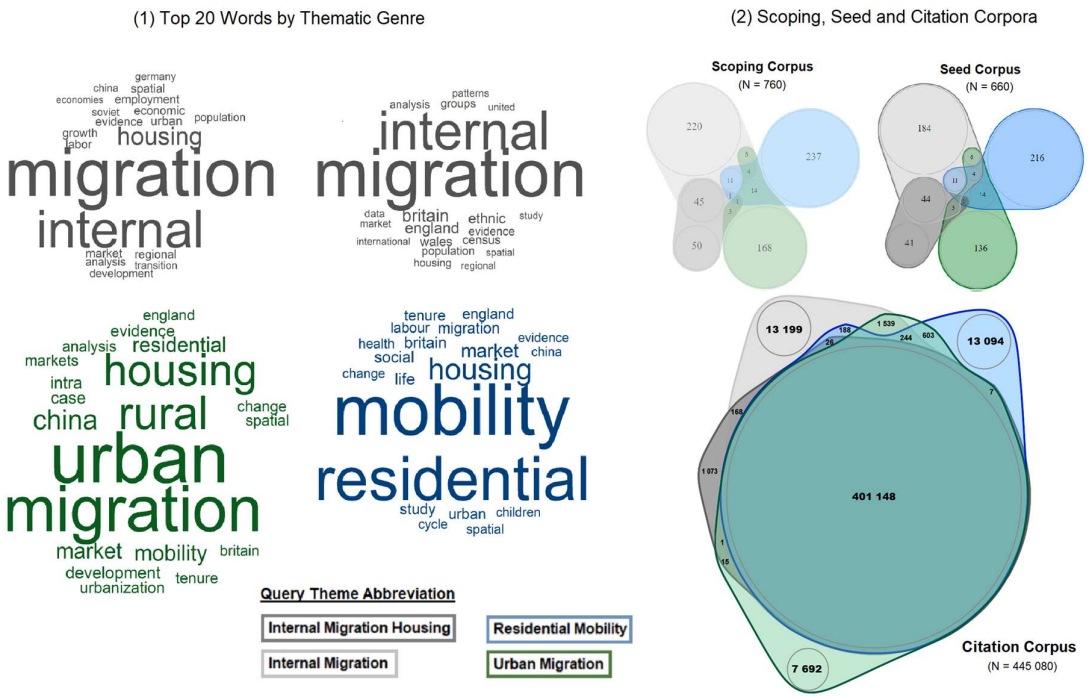

图*1.12* 四种主题类型概览

图1.12概述了四种主题类型：一方面以关键词形式呈现（1），另一方面展示了在构建相关论文语料库的各个阶段中论文所体现的主题类型（2）。维恩图展示了每个步骤的语料库，其中最值得注意的是最终的论文引文语料库，各主题类型之间的重叠程度更高。图中还列出了种子语料库中论文标题里出现频率最高的前二十个关键词，各词的字号取决于其出现频率。首先要注意的是，所有主题都包含住房（因为图1.6所示的检索式对此作了规定）；不过，这一术语的重要程度在 IHM 与内部迁移之间差异最为明显。IHM 学术研究呈现出对前东欧集团研究的侧重，关键词包括“苏联”“转型”和“德国”。这种侧重可能是因为这些地方从社会主义向民主治理模式转型时，住房领域也发生了剧烈变化：在铁幕倒下、人员得以自由流动的同时，住房从几乎完全由国家所有转向私有化。另一方面，内部迁移研究并未被强制聚焦于住房，其高频关键词包括“英国”“英格兰”，以及作为主题的“族裔”“群体”和“人口”。这表明，内部迁移论文脉络主要关注以盎格鲁为中心的隔离案例研究。正如预期，回返迁移通过“生活”“周期”“儿童”和“社会”等词语，引入了家庭和家族层面的个体主题。相比之下，城市迁移则引入了宏观空间主题，相关词语包括“农村”“发展”“变化”和“城市化”。

1.6.3 引文语料库

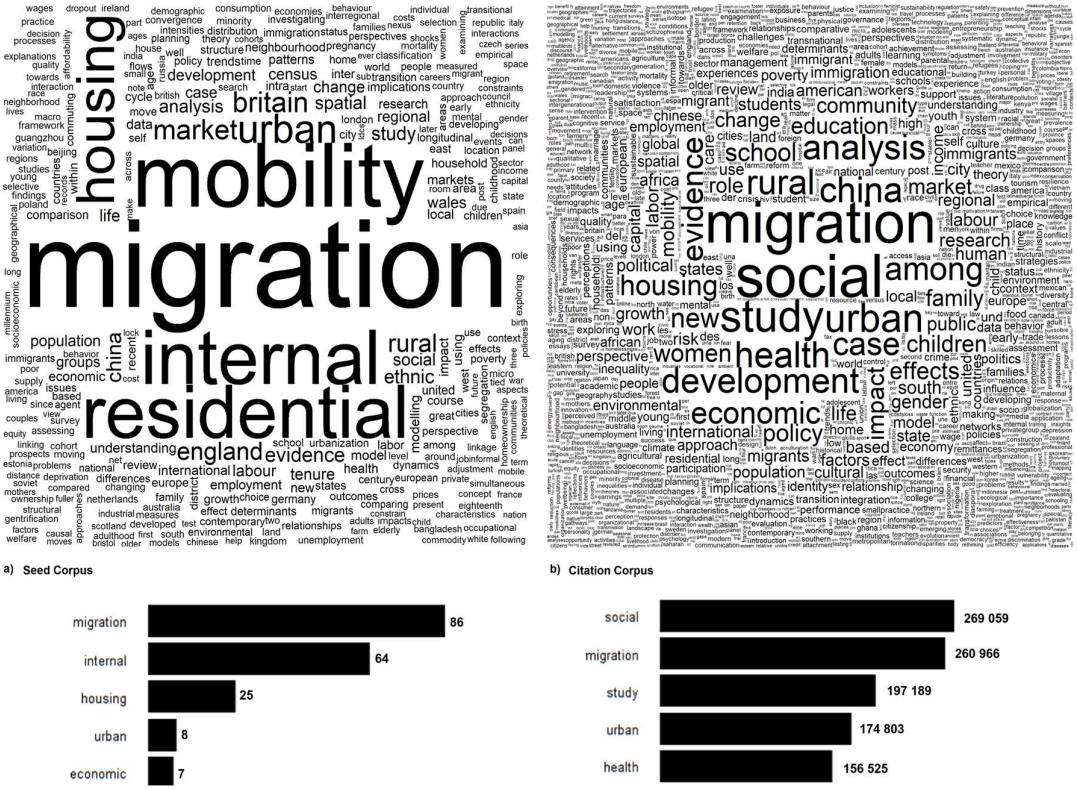

图*1.13* 论文标题中使用频率最高的前*15*个词

图1.13 并列比较了种子语料库和引文语料库，并考察了两组论文标题中最突出的前五个术语。引文语料库规模大得多，因为它包含种子语料库，以及引用这些论文、最多相隔两步的最热门前100篇论文（如图1.4所示）。这就是图1.13-b）中出现更多术语的原因。并列比较中值得注意的是，在引文语料库中，“社会”一词比“移民”更为突出。这表明，移民与住房研究在研究者的认识共同体之外也被广泛引用，以帮助理解其他社会过程。考察文档标题中与“社会”关联最密切的术语，可以了解这些过程的本质。图1.14-a）表明，已有研究倾向于将移民与特定社会过程联系起来，尤其是与个体层面收益有关的过程，例如社会资本、网络和支持；在较小程度上，也涉及可能跨空间显现的过程，如排斥和凝聚力。

图1.13-b）中值得注意的是，“住房”并未成为排名靠前的术语。这是因为它在引文语料库中的受欢迎程度仅排在第15th 位。“住房”的出现频率低于“健康”（排名第5th位），但与“经济”（排名第12th 位，提及次数为13.2万）处于同一数量级。仔细考察这一术语（如图1.14-b）所示，便不难理解这一点：住房这一主题因其市场层面的内涵而尤为突出，涉及可负担性、价格，在较小程度上还涉及住房产权形式。这两个术语的相关性得分相对较低——这反映了词语对共同出现的可能性——说明相关术语多样，且与之结合使用的主题广泛而多元。

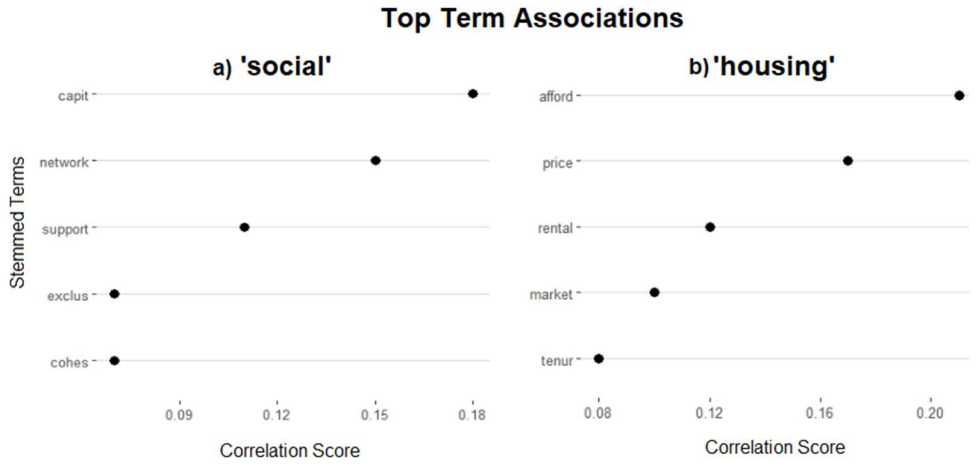

图*1.14* “社会”和“住房”的词干化术语关联

1.6.3.1 国际化

图1.15显示，为这篇系统化综述收集的论文语料库在各年份间增长了三个数量级，呈指数级增长。另一个值得注意的事实是，英语以外语言撰写的论文所占比重几乎持续上升，在整个语料库中占全部文献的17%。这清楚表明，移民与住房交叉领域的学术研究具有全球视野。

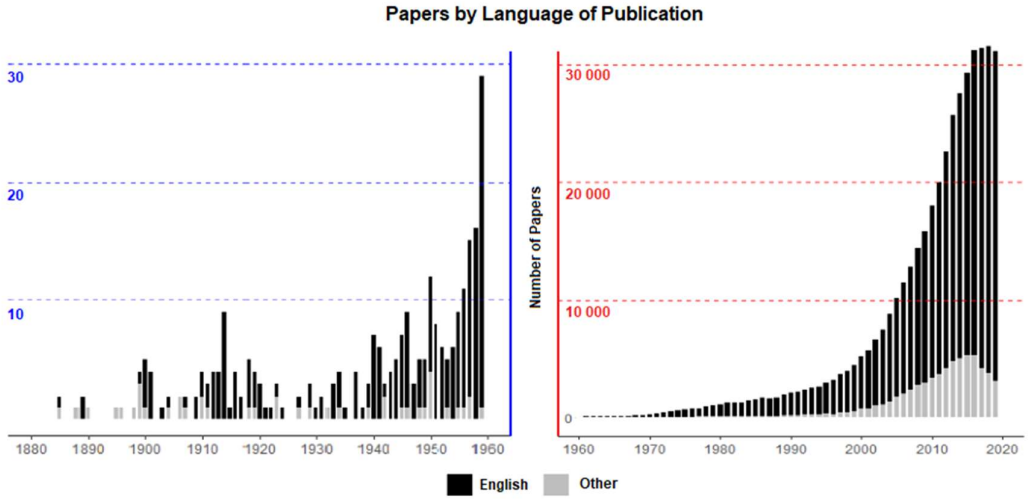

图*1.15* 不同时期按发表语言划分的论文

有趣的是，相关研究中第二常见的语言是简体中文（占论文的3%）。“中国”一词在排名最靠前的一百篇论文（见图1.12）和最终语料库中都十分突出，在后者的词频排名中位列6th 。这并不意外，因为中国是人口最多的国家，2020年人口略高于全球总人口的18%。

1.6.3.2 桥接文献

尽管总体而言这些网络较为稀疏，但其中的文献并非都同等重要。中介中心性较高的文献，通常位于语料库中其他文献之间的最短路径上。这些文献十分重要，因为它们发挥桥梁作用，使整个学术研究体系保持连通。最终语料库中排名靠前的桥接论文见图1.16，住房子网络中相应的论文见图1.17。

值得注意的是，在两个网络中，以较大幅度占据主要桥接文献地位的都是书籍。尽管住房引文网络中存在三个清晰的桥接点，但所考察的完整论文语料库包含若干连接度很高的节点。图1.16显示，de Hass等人（2020年）关于国际移民的多版本书籍以相当大的优势成为最重要的文献。该书认为，截至20世纪90年代，移民——尤其是国际移民——可能是全球变化中最重要的单一因素，并具有社会政治影响。这一论点在1993年至2020年间的不同版本中不断修订和完善，以反映新出现的问题，例如2008年经济危机的影响，以及对非西方地区更为广泛的讨论。《移民时代》之后，排名靠前的是其他探讨国际迁移、城乡区分以及更广泛意义上的发展问题的书籍和编著文集。

相较于年代更久、篇幅更长的文献，有两篇论文格外突出。它们分别是发表于《城市研究》的Champion（1994年）和发表于《国际移民评论》的Schewel（2019年）。Champion（1994年）主张应采用人口学视角来解读国际移民。因此，他敦促推动有关这一主题的协调统一的统计数据；近三十年后，这一呼吁依然具有共鸣，因为移民数据仍是一个颇具争议的问题。图1.16中出现的另一篇重要桥接文献，是Schewel（2019年）关于移民另一面的立场论文——不流动性。她认为，迁移或留下的能力与意愿可能处于一个连续谱上，因此，多样化的方法论策略可以通过捕捉个体随时间变化的重心，为稳定性提供线索。重要的是，这篇论文指出，留下与迁移同样是对更广泛结构性力量的重要回应。这一关于流动性与不流动性的概念框架被纳入本论文第6章。

考察住房子集（图1.17）可以发现三篇桥接文献。与更大的语料库不同，其中只有排名第一的参考文献是一本书，排名第二和第三的是两篇论文。排名靠前的桥接论文分别是 Dietz 和 Haurin（2003）关于住房自有的议程设定型综述，以及 Clark 等人（2003）的美国案例研究；后者表明，1968年至1993年间，成年人的质量、价格和住房产权形式总体呈上升趋势。排名靠前的桥接文献中缺少专著，这一点本身或许就表明住房在移民研究中的发展尚不充分。事实上，排名最前的文献中，只有一篇是书籍：该书由 William A. V. Clark 教授与 Frans Dieleman 教授合作完成，最初于1998年出版，并于2012年再版（Clark and Dieleman，2012），重点探讨美国和荷兰背景下的家庭选择与结果。这部专著的重要性在于，它能够将家庭与住房存量的特征联系起来，例如住房产权形式和规模（或称住房层级，这是该书中的表述）。这表明，在不断变化的社会经济与政治现实中，家庭为理解移民与住房结果提供了一个重要视角。

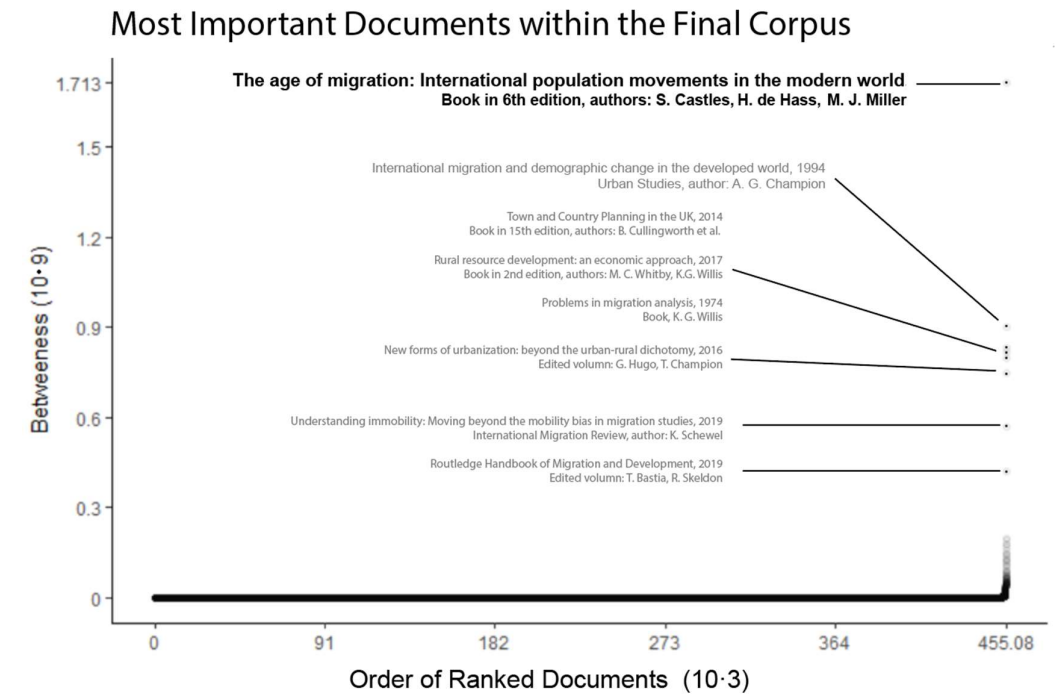

图*1.16* 最终语料库的桥接文献

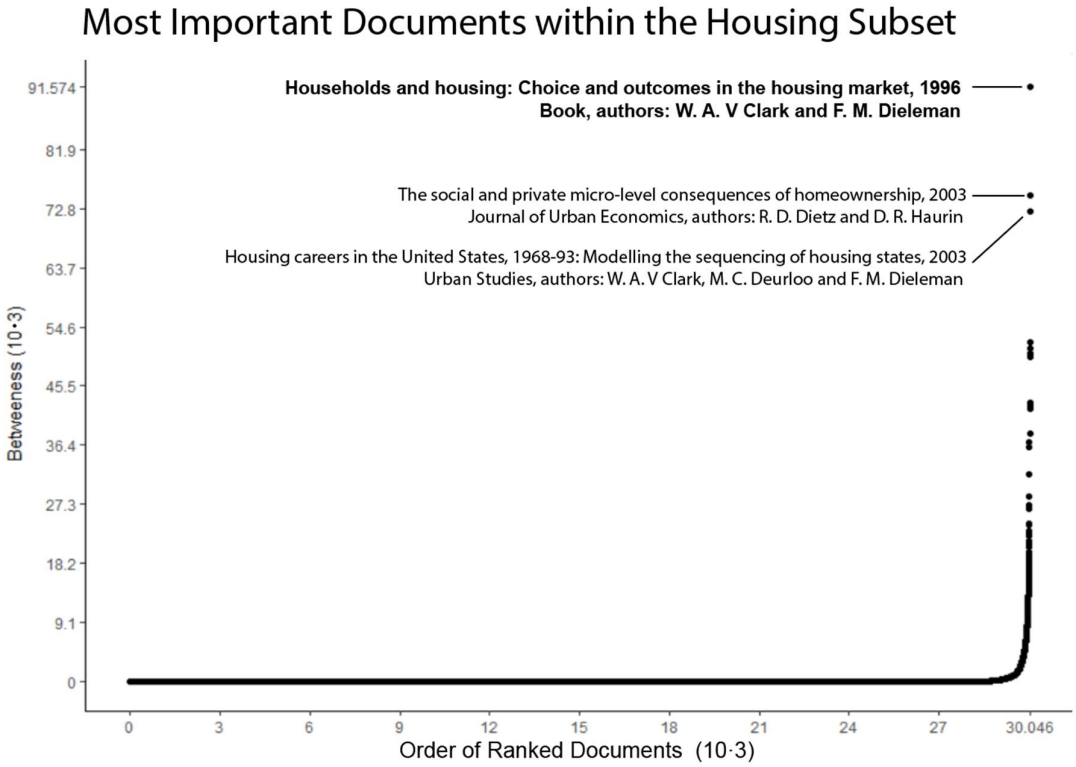

图*1.17* 住房子集的桥接文献

1.7 局限性

本分析梳理了探讨移民与住房交叉议题的文献，并对这两个领域之间的联系进行了反思和量化。尽管提供了全球概览，但引文率这一指标仍需谨慎看待，原因有以下几点。有时，质量存疑的论文也可能被大量引用，因为与一篇发表于新期刊的类似论文相比，其发表期刊的质量被认为更高（Richards 等，2009年，第7–10页）。然而，更重要的是，引文具有滞后性，这会让较早发表的文献相比新近文献占据明显优势。此外，文献的受欢迎程度也取决于不同的索引标准。值得注意的是，学科内部索引有利于在预先界定的关键词和概念范围内积累知识，也使相关内容更容易被检索到。这对医学领域，尤其是对照临床试验而言颇具优势；但要跟上不断演变的社会范式，则较为困难。

1.8 讨论

本文利用谷歌学术（GS）对有关移民与住房的广泛学术研究进行了系统化综述。谷歌学术收录了多种来源，因此是规模最大的书目数据库（Guse

- 一种描述性方法，着眼于“命名的”地方的特征• 一种因果／主观方法，基于“命名的”个人对地方的体验或判断• 以及一种以“类型”的地方为重点的对比性方法

对最终论文语料库的考察表明，大多数文献都是以实证研究为主的案例研究：图1.13-b）中排名第三的热门关键词是“研究”。研究的突出地位值得关注，因为这再次证实，移民与住房交叉领域的文献依赖实证案例研究和归纳推理。这类研究是对现实的观察；其最佳成果是通过概括形成理论和范式，并使它们相互联系，从而在某一主题领域日渐成熟时经受住时间的考验。鉴于有关移民与住房的学术研究横跨众多学科，本文的研究结果表明，理解移民与住房是一项将多元观点融入共同整合框架的任务。

引文网络有助于广泛识别“该领域的本质”（Foster 等人，2007 年，第 3 页），因为学术成果的数量已逼近人类阅读时间的极限。鉴于与移民相关的研究成果呈指数级增长——本分析证实了这一点，Pisarevskaya 等人（2020 年，第 463 页）也观察到了这一趋势——案例研究的概括能力正变得日益重要；尤其是随着国际案例研究的增加，以及由此引发的第二波比较研究，这一点更为迫切。在语料库中，通过该搜索方法返回的排名靠前的文章中，有 17% 并非使用英语撰写。此外，并非所有以英语发表的出版物都涉及英语语境。因此，所引用的百分比只是对英国和美国主导语境之外研究的保守估计。此外，纳入语料库分析的论文仅是谷歌学术中被高度引用的论文，因此不太可能代表实际的

全球范围内的研究活动规模。因此，可复现的研究——即能够跨地方进行概括的研究——至关重要。在数据与计算时代，这类研究甚至可能优先于传统的描述性或轶事性叙述，成为更重要的贡献；在这个时代，我们被数据淹没，却渴望理解。

最后，通过引文相互关联的最终论文语料库，可以看出文献类型的广泛性，以及与移民和住房相关主题的多样性。这种分散性与此前关于主题广度的观察一致：研究评估活动（RAE）2008年“地理学与环境研究”类别中的出版数据（N= 4 590 篇书面成果）显示，Richards 等人也得出了相同结论。他们指出：“自然地理学家和人文地理学家在选择期刊时都不拘一格”（2009年，第234页）。

1.9 结论

尽管文献计量分析可以通过引文数识别并评估回顾性的“影响力”，但人工同行评审或许更能贴近当下对“质量”的看法。然而，评审和文献计量分析都是推动认识论探索的参与方式。期刊文章——专业学术数据库中收录最广泛的内容——并非知识传播的唯一形式，事实上也并非普遍意义上最佳的形式。书籍至关重要，因为它们经过协作与编纂，能够“产出具有相当影响力且历久弥新的基准分析”（Richards 等人，2009年，第3页）。不过，工作论文或博客、白皮书或政策文件以及教学等更及时的成果，可能传播范围更广，对一个领域的发展贡献也更为深远。

本分析考察了关键词重叠情况（图1.12、图1.13和图1.14），以展示移民与住房交叉领域文献中的研究主题。重要的是，本分析将移民与住房领域的学术研究置于整体背景中，从而反思二者之间联系的量化方式。通过考察文献之间的相互关联，本文凸显了住房主题在移民研究中相对缺乏关注。本文考察的书目数据集包含445 080条记录，其中大多数文献为案例研究。这类研究对现实进行观察；若能做到最好，便可将观察结果推广，形成理论与范式，随着各研究主题领域日渐成熟，进一步加强研究成果之间的联系。鉴于文献数量随时间呈指数级增长，未来有关移民与住房交叉领域的研究可遵循一项指导原则：采用整合式而非系加式的方法，使研究能够适用于不同情况。如此一来，这些领域中不断涌现的庞大数据集或许能更好地完全析采，建设都

参考文献

Anders en, H.S.（2011年）《丹麦边缘地区远距离逆城市化迁移的解释》 ,
Populati on, 空间与地方，17(5)，第 627–641 页。链接：https://doi.org/10.1002/psp.568。

Berkenkotter, C. 和 Huckin，智力支架之上 (2016) 学科交流中的体裁知识：认知／文化／权力。劳特利奇。链接：https://doi.org/10.4324/9781315538747。

Bloze, G. (2009)《东欧转型国家的区域间迁移与住房结构：立陶宛 2001–2008 年概况》，《波罗的海经济学杂志》，9(2)，第 47–66 页。链接：https://doi.org/10.1080/1406099X.2009.10840461。

Bolt, G. 和 van Kempen, R.（2010）《族群隔离与居住流动：荷兰少数族裔群体的迁居》，《族群与移民研究杂志》，36(2)，第 333–354 页。可访问：https://doi.org/10.1080/13691830903387451。

Brettell, C.B. 和 Hollifield, J.F.（2015）《移民理论：跨学科对话》。劳特利奇。

Carling, J.（2015）《谁是迁移研究领域的重要人物：值得了解的107个名字》，*Jørgen Carling*，6月1日。可访问：https://jorgencarling.org/2015/06/01/who-is-who-in-migration-studies-108-names-worth-knowing/（访问日期：2021年1月13日）。

Champion, A.G.（1994年）《发达国家的国际移民与人口变化》，《城市研究》，31(4–5)，第 653–677 页。可访问：https://doi.org/10.1080/00420989420080621。

Clark, W.A.V., Deurloo, M.C. 和 Dieleman, F.M.（2003）《美国的住房历程，1968–93：住房状态序列的建模》，《城市研究》，40(1)，第143–160页。网址：https://doi.org/10.1080/00420980220080211。

Clark, W.A.V. 和 Dieleman, F.M.（2012）《家庭与住房：住房市场中的选择与结果》。新不伦瑞克：Transaction Publishers。

CLD2Owners（2021）“CLD2Owners/cld2”。CLD2Owners。网址：https://github.com/CLD2Owners/cld2（访问日期：2021年1月26日）。

CLIMIG（2020）“Climig | 移民、气候变化与环境数据库”。网址：https://climig.com/（访问日期：2021年1月13日）。

Cozzens, S. (1981年)《衡量科学：引文理论综述》，载于国际知识社会学学会。

van Dalen, H. (2018年)《移民仍是人口学的继子女吗？》，*DEMOS*，34，p. 8。

Deeks, J.J.、Higgins, J.P. 和 Altman, D.G. (2019)《分析数据并开展荟萃分析》，载于《干预措施系统综述科克伦手册》。John Wiley & Sons, Ltd，pp. 241–284。获取地址：https://doi.org/10.1002/9781119536604.ch10。

Dennis, R.J. (1977)《维多利亚时期城市中的两次人口普查期间的流动》，英国地理学家学会会刊，2(3)，pp. 349–363。获取地址：https://doi.org/10.2307/621836。

DeWind, J.（2020年）《盲人摸象：对迁移研究领域的一种看法》，《比较移民研究》，8(1)，第39页。来源：https://doi.org/10.1186/s40878-020-00191-4。

Dhammi, I.K. 和 Haq, R.U.（2016）《什么是索引编制》，《印度骨科杂志》，50(2)，第115–116页 .
来源 网址：https://doi.org/10.4103/0019-5413.177579。

Dietz, R.D. 和 Haurin, D.R.（2003）《住房所有权的社会和个人微观层面后果》，《城市经济学杂志》，54(3)，第401–450页。来源：https://doi.org/10.1016/S0094-1190(03)00080-9。

Fortunato, S. 等（2018年）“科学学”，《科学》，359(6379)，第 eaao0185 页。可访问：https://doi.org/10.1126/science.aao0185。

Foster, J. 等（2007）“经济地理区域的流通：经济地理学中的引文模式与引文行为，1982–2006”，英国地理学家学会会刊，32(3)，第 295–312 页。可访问：https://doi.org/10.1111/j.1475-5661.2007.00239.x。

Gibson, N.P.（2020）研究指南：借助谷歌学术获得更佳结果：在谷歌学术中进行搜索。参见：https://uark.libguides.com/c.php?g=78885&p=509014 （访问日期：2020年8月26日）。

Gough，D.、Oliver, S. 和 Thomas, J.（2017）系统综述导论。第2版。洛杉矶：SAGE。

Grant, M.J. 和 Booth, A.（2009）《综述类型学：对14种综述类型及相关方法学的分析》，《健康信息与图书馆学期刊》，26（2），第91–108页。参见：https://doi.org/10.1111/j.1471-1842.2009.00848.x。

Gusenbauer, M. (2019) ‘谷歌学术会令其他搜索引擎黯然失色吗？比较 12 个学术搜索引擎和书目数据库的规模’, 《科学计量学》, 118(1), pp. 177–214. 获取地址：https://doi.org/10.1007/s11192-018-2958-5。

de Haas, H., Castles, S. 和 Miller, M. (2020) 《移民时代》第六版：现代世界的国际人口流动。 吉尔福德出版社。

Higgins, J. 等 (eds) (2019) ‘《干预措施系统综述科克伦手册》’. 科克伦。获取地址：www.training.cochrane.org/handbook（访问日期：2020年8月24日）。

Hood，W.W. 和 Wilson, C.S. (2001) ‘《文献计量学、科学计量学和信息计量学文献》’，第 24 页 .

Jones, C.（2002年）《住房市场区域的定义与战略规划》，《城市研究》，39(3)，第549–564页。可通过以下网址获取：https://doi.org/10.1080/00420980220112829。

King, R.（2011年）《地理学与迁移研究：回顾与展望：地理学与迁移研究：回顾与展望》，《人口、空间与地方》，18(2)，第134–153页。可通过以下网址获取：https://doi.org/10.1002/psp.685。

King, R. (2020)《论移民、地理学与认识共同体》，《比较移民研究》，8(1)，第35页。可通过以下网址获取：https://doi.org/10.1186/s40878-020-00193-2。

King, R. 和 Skeldon, R.（2010年）《“注意差距！”整合国内移民与国际移民研究方法》，《族群与移民研究杂志》，36(10)，第1619–1646页。可通过以下网址获取：https://doi.org/10.1080/1369183X.2010.489380。

Korgelli, P.（1994年）《关于国内移民与国际移民之间的相互关系》，《创新：欧洲社会科学研究期刊》，7(2)，第151–163页。参见：https://doi.org/10.1080/13511610.1994.9968397。

Lee, E.S.（1966年）《迁移理论》，《人口学》，3(1)，第47–57页。参见：https://doi.org/10.2307/2060063。

Levy, N.、Pisarevskaya, A. 和 Scholten, P.（2020年）《在碎片化与制度化之间：迁移研究作为一个研究领域的兴起》，《比较移民研究》，8(1)，第24页。参见：https://doi.org/10.1186/s40878-020-00180-7。

Martín-Martín, A. 等（2018年）《谷歌学术、科学网与 Scopus：对252个学科类别中引文的系统比较》，《信息计量学期刊》，12(4)，第1160–1177页。参见：https://doi.org/10.1016/j.joi.2018.09.002。

Martín-Martín, A.、Thelwall, M. 和 López-Cózar, E.D.（2020年）《谷歌学术、微软学术、Scopus、Dimensions、科学网和 OpenCitations 的 COCI：通过引文比较覆盖范围的多学科研究》，第 36 页。

Moher，D. 等（2009年）《系统综述和荟萃分析优先报告条目：PRISMA 声明》，《英国医学杂志》，339。获取地址：https://doi.org/10.1136/bmj.b2535。

Nestorowicz，J. 和 Anacka，M.（2018年）《注意鸿沟？量化移民文献中两种传统之间的相互联系》：《国际移民评论》 [预印本]。获取地址：https://doi.org/10.1177/0197918318768557。

Piguet，E.、Kaenzig，R. 和 Guélat，J.（2018年）《"环境移民"研究的不均衡地理分布》，《人口与环境》，39(4)，第 357–383 页。网址：https://doi.org/10.1007/s11111-018-0296-4。

Pisar evskaya，A. 等（2020年）《绘制移民研究图谱：一项关于这一新兴研究领域逐渐成熟的实证分析 rese研究领域》，《迁移研究》，8(3)，第 455–481 页。网址：https://doi.org/10.1093/migration/mnz0 31.

Porter，M.（2006年）*Porter*词干提取算法。网址：https://tartarus.org/martin/PorterStemmer/（访问日期：2021年1月26日）。

quanteda（2019）"停用词"。网址：https://github.com/igorbrigadir/stopwords（访问日期：2021年1月26日）。

Raimbault，J.（2017）《跨学科科学景观探索》，《科学计量学》，119（2），第617–641页。获取地址：https://doi.org/10.1007/s11192-019-03090-3。

Ravenst ein，E.G.（1889年）《移民法则》，《皇家统计学会期刊》，52（2），第241–305页 . 可获取 地址：https://doi.org/10.2307/2979333。

Richards, K. 等（2009年）《地理学与环境研究中的出版与评估性质：来自2008年研究评估活动的证据》，*Area*，41（3），第231–243页。获取地址：https://doi.org/10.1111/j.1475-4762.2009.00908.x。

Rossi, P.H. (1955) 家庭为何迁移. 第2版. 比佛利山：SAGE 出版社。

Roth, C. 和 Bourgine, P. (2005)《认识共同体：描述与层级分类》，数学人口学研究，12(2)，第107–130页。网址：https://doi.org/10.1080/08898480590931404。

Roth, S. (2020年) 研究指南：系统综述及其他综述类型：什么是系统综述 ? 可查阅 阅：https://guides.temple.edu/c.php?g=78618&p=4178713 （访问日期：2020年8月24日）。

Roth, S. (2021a) 研究指南：系统综述及其他综述类型：什么是元综合？可访问：https://guides.temple.edu/c.php?g=78618&p=4178716 （访问日期：2021年1月13日）。

Roth, S. (2021b) 研究指南：系统综述及其他综述类型：什么是快速综述？可访问：https://guides.temple.edu/c.php?g=78618&p=4156608 （访问日期：2021年1月13日）。

Schewel, K. (2019) '理解不流动性：超越迁移研究中的流动性偏见'，《国际移民评论》，54(2)，第328–355页。可访问：https://doi.org/10.1177/0197918319831952。

Sjaastad, L .A. (1962) '人口迁移的成本与收益'，《政治经济学杂志》，70(5, Part 2)，第 . 80–93。

Sweileh, W.M. 等（2018年）《同行评审文献中全球移民健康研究的文献计量分析（2000-2016）》，*BMC*公共卫生，18(1)，第777页。可查阅：https://doi.org/10.1186/s12889-018-5689-x。

英国数据档案馆（2020年）《受控词表》。可查阅：https://www.data-archive.ac.uk/managing-data/standards-and-procedures/controlled-vocabularies/（访问日期：2020年8月26日）。

Vermeulen, W. 和 van Ommeren, J.（2009年）《土地利用规划是否影响区域经济？对荷兰住房供给、内部迁移和地方就业增长的同步分析》，住房经济学杂志，18(4)，第294–310页。可查阅：https://doi.org/10.1016/j.jhe.2009.09.002。

Winstanley, A., Thorns, D.C. 和 Perkins, H.C.（2002年）《搬迁住所，营造家园：探索居住流动性》，住房研究，17(6)，第813–832页。可访问：https://doi.org/10.1080/02673030216000。

Zhou, M. 和 Logan, J.R.（1991）《进出唐人街：纽约市华人的居住流动与隔离》，社会力量，70(2)，第387–407页。可访问：https://doi.org/10.2307/2580245。