arXiv:2601.16990v1 [cs.DL] 7 Jan 2026

pyBiblioNet：用于全面基于网络的文献计量分析的 Python 库

Mirko Lai[1*†]、Salvatore Vilella[1†]、Federica Cena[2]，Giancarlo Ruffo[1]

[1*]皮埃蒙特东方大学技术科学与创新系，Teresa Michel 大道 11 号，亚历山德里亚，15121，皮埃蒙特，意大利。[2]都灵大学计算机科学系，Svizzera 大道 185 号，Via Pessinetto 12 号，都灵，10149，皮埃蒙特，意大利。

*通讯作者。电子邮箱：mirko.lai@uniupo.it；贡献作者：salvatore.vilella@uniupo.it；federica.cena@unito.it；giancarlo.ruffo@uniupo.it；

†这些作者对本研究的贡献相同。

摘要

文献计量分析是理解科学研究结构、动态和影响力的重要工具。传统方法往往难以捕捉科学文献中复杂的关系和不断演变的趋势。为弥补这一空白，我们推出了 pyBiblioNet，这是一个旨在促进全面的基于网络的文献计量分析的 Python 库，可为引文网络、合著网络和关键词共现网络提供洞见。该库与 OpenAlex 集成。OpenAlex 是全球科研体系中广受欢迎且开放的目录，使用户能够轻松地预处理、可视化和分析文献计量数据。其关键功能包括主题选择、通过 OpenAlex API 自动下载数据、创建用于分析的手稿根集和基础集、构建引文网络和合著网络、网络可视化工具，以及一套用于计算网络中心性、聚类和社区检测的算法；这些功能均针对文献计量领域进行了定制。此外，该库还支持使用 NLP 技术分析关键主题和概念。我们以跨学科的“15分钟城市范式”为例开展文献计量分析，展示该库的主要功能，从而证明 pyBiblioNet 在不同科学领域中发现隐藏模式和新兴趋势方面的实用价值。pyBiblioNet 能够为研究人员、图书馆员和政策制定者提供强大且易用的工具，以提升其文献计量分析能力并支持数据驱动决策。

关键词：文献计量分析、网络分析、Python 库、15 分钟城市

1 引言与相关工作

文献计量分析是探索科学研究结构、动态和影响力的重要方法论。通过系统地量化和分析学术成果，文献计量技术使研究人员能够发现引文实践、共同署名关系以及研究主题演变中的模式。尽管所有科学领域在开展书目研究时都有各自的传统、习惯和惯常做法，但仍存在一些跨领域工具 Moral-Mu˜noz 等(2020年)，能够帮助研究人员检索信息、整理和探索他们偏好的书目来源。鉴于每年发表的论文数量持续高速增长，并且自20世纪90年代以来呈指数级增长，这类工具正变得至关重要 Fire 和 Guestrin(2019年)，这使科学家越来越难以及时掌握其研究领域的最新进展，也让他们更难涉足自身专业之外的领域。

在此背景下，整合新的方法以促进文献计量任务或许具有重要意义，例如纳入网络分析，它能够提供与文献计量研究相关的补充指标。正如Donthu 等(2021所指出的，网络分析是现代文献计量研究的重要组成部分：众多涉及各类领域的文献综述都运用它来揭示研究领域与子领域、作者、研究机构，以及最重要的思想和概念之间的关系，例如Liu 等(2015)；Andrikopoulos 和 Economou(2016)；Tunger 和 Eulerich(2018)；Rossetto 等(2018)；Cisneros 等(2018)；Baker 等(2020)；Andersen (2021)；Ruffo 等(2023)。

事实上，网络分析的设计本身就允许通过各种网络类型考察出版物之间的关系。引文网络展示出版物如何通过引文彼此关联，而合作网络则揭示研究人员之间的合著模式。此外，语义网络能够突出出版物中的共同术语，从而帮助洞察文献中的主题关联。共被引和书目耦合等技术通过识别相关成果的聚类及新兴研究领域，进一步丰富了这类分析。VOSviewer 等工具Van Eck 和 Waltman(2010采用基于距离的方法创建引文关系和合著网络的可视化呈现；GephiBastian 等(2009)也常用于可视化这些网络，以图形方式呈现网络，从而加深对复杂文献计量关系的理解Van Eck 和 Waltman(2014)。这种多方面的方法不仅有助于识别有影响力的作者和期刊，还有助于描绘研究领域随时间的演变。Research Rabbit 等其他解决方案Cole 和 Boutet(2023)则提供参考文献管理工具，并可视化引文网络和合著网络。

近年来，文献计量工具的发展势头日益增强，涌现出多种软件库，以便开展数据收集和分析Moral-Mu˜noz 等

（2020年）。其中值得关注的是 BibliometrixAria 和 Cuccurullo（2017年），这是一个功能丰富的 R 软件包，可用于文献计量分析，包括从各种数据库中提取数据以及进行可视化。另一项重要成果是 pyBibXPereira 等（2025年），它将人工智能功能（包括嵌入向量、主题建模、文本摘要及其他通用自然语言处理任务，采用 Sentence-BERT、BerTopic、BERT、ChatGPT 和 PEGASUS 等模型）融入文献计量分析，并支持探索性数据分析以及引文和合作分析等网络功能。

从更广泛的背景来看，多位作者（例如 Salman 等人（2025年）；Tosi （2025年））指出，越来越需要采用自动化工具来加快研究工作流程，通过可靠的检索和摘要方法，减少文献综合所耗费的时间。尽管 ChatGPT 和 Perplexity 等生成式人工智能工具带来了令人期待的机遇，但也引发了伦理方面的担忧，例如抄袭、偏见以及过度依赖人工智能生成内容的风险。因此，将人工智能的效率与人类研究人员的批判性思维和监督相结合至关重要。从这一角度来看，与解释和可解释性相关的挑战变得尤为重要，不应被忽视。基于网络的文献计量分析提供了一个稳健的理论框架，在该框架中，每项指标都能得到清晰的理解、探究和可视化。我们认为，短期内，基于网络的文献计量分析将与生成式人工智能工具相辅相成，而非被其取代，尤其是在应对科学文献中的信息过载问题方面。

1.1 我们的贡献

如此多样的软件解决方案为促进文献计量研究提供了不同的途径 Moral-Mu˜noz 等 (2020年）。其中大多数，例如 bibliometricx Aria 和 Cuccurullo (2017年）和 pybibx Pereira 等 (2025年），都包含网络分析。然而，这类分析通常仅用于可视化，旨在大致了解引文格局，并识别论文和研究人员构成的凝聚集群；而且，它们往往需要从 Web of Science 或 Scopus 等来源数据库手动收集数据，通常以 CSV 或 BibTeX 格式下载数据，以便后续处理。

我们开发了 pyBiblioNet，这是一个专门用于增强基于网络的文献计量分析的 Python 库。pyBiblioNet 由两个主要模块组成：一个专注于数据收集，另一个旨在进行社区检测和网络分析。第一个模块利用 OpenAlex Priem 等人 （2022年）的应用程序接口，便于收集文献计量数据并构建引文图（参见附录中该模块的文档 `openalex`，附录 A）。虽然可以纳入其他数据源，但图对象是第二个模块所实现分析的核心；该模块负责计算网络指标并识别相关节点和连接。此外，它还专注于社区检测，实现了多种常用算法，例如 Louvain Blondel 等人 （2008）、Infomap Rosvall 和 Bergstrom（2008）、随机块模型 Holland 等人（1983）（详见表格 3），允许用户通过可自定义的可视化来展示聚类之间的连接（参见附录中模块 文献计量分析 的文档，附录 B）。由于 pyBiblioNet 以

开源软件[1], 形式发布，因此用户可以集成自己的数据源、分析和可视化，使其成为一个极其灵活的基于网络的文献计量研究工具。

在 pyBibX 库发布之际，Pereira 等 Pereira 等 (2025年) 对 15 种不同的文献计量工具进行了深入分析，并为每种工具详细总结了其关键功能、数据源和运行环境。本文将我们的系统 pyBiblioNet 与根据其分析满足最多已识别关键功能的三种工具进行比较，即 pyBibXPereira 等 (2025年)、Bibliometrix Aria 和 Cuccurullo (2017年) 和 Scientopy Ruiz- Rosero 等(2019年)。表格 1 展示了 pyBiblioNet、pyBibX、Bibliometrix 和 Scientopy 之间的比较分析。

Tool	来源	数据分析	网络分析	人工智能
bibliometrix	Cochrane Dimensions PubMed Scopus WoS	数据处理 数据分析 词云 投影 演化图 桑基图 条形图	引文分析 合作分析 相似性分析 全球合作分析	-
pyBibX	PubMed Scopus WoS	数据处理 数据分析 词云 N 元语法 投影 演化图 桑基图 矩形树图 条形图	引文分析 合作分析 相似性分析 全球合作分析	主题建模 嵌入 摘要 ChatGPT
Scientopy	Scopus WoS	数据处理 数据分析 词云 演化图 条形图	引文分析 合作分析 相似性分析 全球合作分析	-
pyBiblioNet	OpenAlex	聚类网络 N 元语法 演化图 条形图	引文分析 合作分析 聚类分析 基础集分析 中心性分析	主题建模

表 1：文献计量工具在数据源、数据分析能力、网络分析功能和人工智能集成方面的比较。

在所分析的库中，pyBibX 提供的功能数量最多，其中包括采用大语言模型的高级功能。对于 pyBiblioNet，我们的目标并不是复现现有库已经实现的功能，而是提供新颖且独特的功能。

首先，pyBiblioNet 是唯一使用 OpenAlex 作为数据源的工具。它不仅是一个替代方案，还整合了多个数据源，包括 ORCID、ROR、DOAJ、Unpaywall、PubMed、PubMed Central 和 ISSN 国际中心。此外，它提供的文章数量显著更多（243

[1] https://github.com/mirkolai/pybiblionet和 https://pypi.org/project/pybiblionet/

百万），相比之下，Scopus（8700 万）、Web of Science（核心合集）（8700 万）和 Dimensions（1.35 亿）[2]。

此外，尽管我们当前的分析主要聚焦于引文网络和合著网络，但我们提出了通过纳入基础集节点来扩展网络的可能性——也就是纳入那些最初不满足搜索筛选条件、但与根集网络相连接的文章——从而拓宽分析范围。另一项重要创新是将聚类技术应用于引文网络，以对文章群组进行主题分析。类似的聚类方法也可应用于合著网络，以研究国家之间的合作模式。最后，我们使中心性指标更易于获取，从而能够依据超越简单引文计数的标准（例如合作数量）识别高影响力文章。

为了避免再创建一个孤立的库，我们还致力于通过支持以兼容其他库的格式导出从 OpenAlex 获取的元数据，确保与现有工具实现互操作。此外，引文网络和合著网络也可以标准化格式导出，以便与外部网络分析工具配合使用。

2 基于网络的文献计量学

如上一节所述，基于网络的文献计量学是一套分析方法和工具，运用网络科学研究并量化作者、文章、期刊或机构等科学实体之间的关系及其影响。这种方法论植根于图论和社会网络分析的原理，科学文献计量网络通常表示为有向图或无向图 以及加权图或无权图，其中科学实体表示为节点，并由边 连接，以表示它们之间的交互。有向有向图 $G = (V, E)$由一组节点 V和一组有向边 E构成，其中每条边都有一个方向，用有序对（u, v）$\in E$表示，意即存在一条从节点 u 指向节点 v的有向连接。相反，无向图中的边没有方向，两个节点 u和 v 之间的连接仅用无序对 $\{u, v\}$表示。当实体之间的关系是相互的时，这种表示方式很合适，例如研究人员之间的共同著作。此外，加权图 是指每条边都赋有权重的图，权重通常表示两个节点之间关系的强度或程度。数学上，加权图可表示为 $G = (V, E, w)$，其中$w: E \to \mathbb{R}$ 是一个权重函数，为每条边赋予一个实数值权重。对于文献计量网络而言，这些权重可能反映例如两位作者合作成果的数量。另一方面，无权图假定所有关系都相同，不区分连接的强弱；例如，文章之间的引文网络便属于这种情况，因为一篇文章对另一篇文章至多只能引用一次。

[2] https://openalex.org/about#comparison

2.1 中心性指标

网络分析的强大之处在于能够以简单且定量的方式识别关键 参与者。这可以通过计算多种中心性指标来实现，其中一些已在该库中实现，并列于表 2中。这些指标超越了对论文引文数或作者 h指数的简单统计，以更精细的方式量化实体在不同角色中的重要性。有些实体在不同科学共同体中广受欢迎；相反，有些实体可能极受欢迎，但仅限于其子领域；有些实体在连接不同社群方面发挥关键作用，使研究人员能够从一组论文转向另一组论文；还有些实体倾向于被权威实体引用，或与之合作。网络分析能够捕捉并量化这些及其他细微差异，而且在文献计量分析中十分有效，正如 Diallo 等人在 Diallo 等人 （2016年）所示。

指标	定义	相关性
度中心性	$C_D(v) = \deg(v)$，其中 $\deg(v)$ 是与其相连的边的数量节点 v。	识别被频繁引用的论文或高产 作者， 表明 直接影响力或生产力。
中介中心性	$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$，其中 σ_{st} 是最短路径总数从 s 到 t，而 $\sigma_{st}(v)$是经过这些路径的数量经过 v。	突出显示充当中间节点的节点中介， 促进 the flow of 在不同部分之间传递信息的网络，例如跨学科论文。
接近中心性	$C_C(v) = \frac{1}{\sum_{u \neq v} d(v,u)}$，其中 $d(v, u)$是最短路径距离介于 v 和 u之间。	识别具有战略意义的节点能够高效地获取或分发信息至整个网络。
特征向量中心性	$C_E(v) \propto \sum_{u \in V} A_{vu} C_E(u)$，其中 A 是该网络。	识别具有更广泛系统性影响的节点，例如开创性论文被引用 by 其他 高度 有影响力的成果。
PageRank	$PR(A) = \frac{1-d}{N} + d \sum_{i=1}^{k} \frac{PR(B_i)}{L(B_i)}$其中， d是阻尼因子， N是页面总数（共网络中的页面总数， B_i 是链接到页面 A的页面，而 $L(B_i)$表示来自.....的出站链接数量第 B_i页。	突出权威性或影响力较大的成果，即使它们并非拥有最高的直接引文数。

表2：常见的基于网络的中心性指标及其与文献计量网络的相关性。

2.2 社区检测

社区检测提供了一种强有力的方法，能够超越基于科学域或学科的传统分类。它揭示了系统的自然结构，展现出科学知识与合作的内在组织方式。传统分类（例如基于预先定义的科学领域或主题领域的分类）往往受限于僵化的边界，可能无法体现现代研究的动态性和跨学科特征。另一方面，社区检测

则依据网络中的实际关系，识别联系紧密的节点聚类；这些关系可以是引文链接、合著关系或其他形式的互动。这些聚类代表了自然形成的科研活动群体，反映了知识如何自然地组织起来并彼此相连。

通过检测这些社群，我们能够深入了解科学版图的真实拓扑结构。与主题建模从文本内容中提取语义主题、依据词语使用情况对文档分组不同，社区检测依靠结构关系（引文或共同著作）来识别紧密相连的实体聚类。因此，它捕捉的是科学活动的结构与合作模式，而非单纯的主题相似性。这两种方法相辅相成，为知识的产生与组织方式提供了各具特色又相互交叠的视角。例如，与其将人工智能研究视为一个单一领域，社区检测不如揭示机器学习、自然语言处理和计算机视觉等不同子领域，以及它们与神经科学或伦理学等相关领域之间的联系。同样，常常被传统分类方式遗漏的跨学科研究领域，也可能形成独立社群，凸显其对科学生态系统的独特贡献。我们还可以研究社群如何随时间演变，从而识别新兴领域或研究重点的变化。例如，引文网络中不断扩大的聚类可能表明某个领域正快速创新或受到越来越多的关注，为科学研究趋势提供早期信号。此外，检测桥接节点（即连接不同社群的节点）的能力，有助于识别促进知识转移与跨学科合作的关键参与者或成果。

社区检测有大量方法可应用于各种领域，各有优缺点。关于该领域的全面概述，请参阅相关科学综述 Fortunato (2010年)；Li 等人 (2024年)；Jin 等人 (2024年）。我们在表 3中简要介绍一些常用技术，这些技术均已在该库中实现，并说明它们与文献计量分析的相关性。

3 pyBiblioNet：通过网络分析探索文献计量数据

pyBiblioNet 的结构总结见图 1。该结构由两个主要部分组成：数据收集和文献计量分析。文献计量分析模块由核心分析功能和可视化模块构成，可生成图以便于解释。本节将分析该软件的主要组成部分。

3.1 引文网络的根集/基集结构

pyBiblioNet 的方法论基于对网络指标的分析，以及对以下概念的运用：**网络中的根集/基集**，该概念由 Jon Kleinberg 于 1999年开发的 HITS（超链接诱导主题搜索）算法提出Kleinberg (1999年)。广义而言，根集代表图中连接紧密且最具相关性的部分，通常包含对查询至关重要的节点。

方法	定义	相关性
Louvain Blondel 等人 (2008)	A 基于模块度的方法 that 迭代地优化社群结构以实现最大模块度。	适用于分析——大型文献计量网络，识别——诸如研究聚类 学科 or 主题 子领域基于 on 模块化结构。
Girvan-Newman Girvan 和 Newman (2002)	使用边介数 ness to 迭代地移除 the most 边，将网络分割成若干部分。构建网络。	有助于识别层级结构，例如引文或合著网络 ，例如嵌套子字段 or 跨学科桥梁。
Infomap Rosvall 和 Bergstrom（2008）	划分网络以最小化描述长随机……的数量长度游走长度。	捕捉 自然的分组 in 引文……或合作网络，揭示知识如何流动之间 主题 or 社群。
谱聚类 Ng 等人(2001年)	划分 网络利用特征值图拉普拉斯矩阵以识别聚类。	适合 for 检测广泛的学科边界或层级结构 高 pat- 文献计量网络中的模式 文献计量网络。
随机块模型Holland 等人（1983）	假设 社群 nities are 潜在变量 生成观测到的网络结构。	有效的 for 概率 建模文献计量学 net- 成果。

表 3：社区检测的常用方法及其与文献计量网络的相关性。

基集反而包括连接较少的节点，这些节点通常位于网络边缘，例如被根集文章引用、但自身并不处于中心位置的文章。包括基集可以构建一个更全面的图，从而深入了解间接关系和更广泛的连接；而排除基集则会将分析集中于根集，通常形成更紧凑且更具分析针对性的网络。伪代码 1展示了这一过程。

它由两个主要步骤组成：`FetchArticles` 和`CreateCitationGraph`。`FetchArticles`会根据用户定义的搜索条件，从书目数据库（例如 OpenAlex）中检索**根集** 和基础集文章。对于每篇检索到的文章，系统都会提供各种元数据，并额外包含该文章引用的文章列表以及引用该文章的文章列表。这会创建一个简洁的数据集，用于表示特定主题或感兴趣的领域。数据收集过程可能需要多次查询，主要是因为需要检索被引用文章或引用文章的列表，并且一个

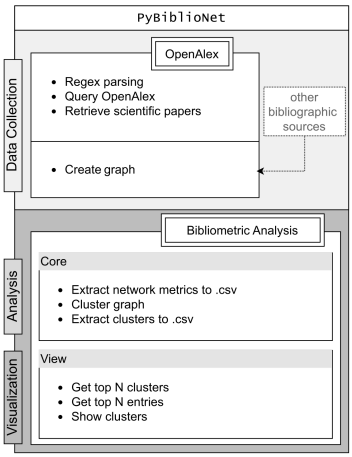


图 1：pyBiblioNet 示意图。该库分为两个主要模块：第一个模块用于数据收集，目前从 OpenAlex 收集数据，但用户也可以将其他文献来源集成其中；最终输出为引文图，可使用第二个模块中用于文献计量分析的函数进行探索。

由于预设了超时以防止 API 负载过高，这需要花费大量时间。此外，OpenAlex 允许每位用户每天最多发起 100,000 次请求。因此，建议通过将参数 缓存=`True` 来缓存查询。

`CreateCitationGraph` 将检索到的文章转换为图表示，其中节点对应文章，边表示引文关系。如果基集 设为`False`，图中只包含符合搜索条件的文章（即**根集** 文章）。反之，当基集 设为`True`时，图会扩展至包含涉及**根集**之外文章的引文，从而生成内容更丰富、具有上下文感知能力的图，突出根集与基础集节点之间的交互。在引文网络中，**根集** 是围绕特定主题或领域形成的紧密关联文章群。这些文章通常被广泛引用，彼此之间也存在密切联系。另一方面，**基集**由关联较弱的文章组成；这些文章可能引用根集中的文章，或被根集中的文章引用，但并非网络的核心。纳入基集能够更全面地呈现引文格局，从而揭示根集文章对外部成果的影响，以及外部成果对根集文章的影响。

同样，对于合著网络，`CreateCoAuthorshipGraph`过程可以创建一个无向网络，其中节点代表作者，两位作者共同发表一篇文章时，两者之间便存在一条边。当基集为 `True` 时，会考虑所有文章；否则，仅纳入根集文章的共同署名关系。

例如，假设一位研究人员正在研究“网络科学”领域。`FetchArticles` 过程会检索严格聚焦于该领域的文章，例如奠基性论文和近期的重要研究，

将基集设为`False` 后，`CreateCitationGraph` 过程会构建一张图，其中节点是这些根集文章，边表示它们之间的引文，从而形成一个主题明确、结构紧密的网络。

如果研究人员将基集设置为`True`，图将扩展为纳入引用根集文献或被根集文献引用的基础集文章。例如，来自“社会网络分析”或“复杂系统”等相关领域的文章可能会出现在图中。这一更广阔的视角有助于揭示网络科学与相邻领域之间的跨学科影响和连接。

3.2 数据收集：查询 OpenAlex API

pyBiblioNet 使用来自 OpenAlex 的数据进行文献计量分析Priem 等人 (2022年）。OpenAlex 是一个开放获取数据库，旨在提供有关学术成果的全面、结构化信息，包括学术出版物、作者、机构、期刊和研究概念的元数据。它是 Scopus 和 Web of Science 等专有学术数据库的免费替代方案。OpenAlex 中的数据来自多个数据存储库，包括 CrossRef、ORCID 和 PubMed，从而确保广泛的覆盖范围和定期更新。该数据库由多个实体组成，这些实体总结于表 4中；这些实体相互连接，从而支持复杂查询和关系分析。OpenAlex 提供 RESTful API，以便以编程方式访问这些数据，促进其与文献计量工具和分析的集成。通过利用 API 的功能，用户可以针对多个条件执行复杂的布尔查询，筛选结果，搜索关键词和短语等。

实体	**描述**
成果	期刊文章等单篇学术成果，书籍和会议论文。
作者	研究人员的详细信息，包括所属机构和发表历史。
来源	期刊、会议及其他出版渠道。
机构	大学、研究组织及其他关联机构。
主题	主题和研究领域，可用于主题建模和分析。
出版商	出版商列表及相关信息。
资助机构	研究资助机构列表，信息来自 CrossRef，以及通过 Wikidata 和 ROR 的数据进行补充。
Geo	用于对学术数据进行地理分类的信息，利用联合国数据将全球划分为大洲和地区，以便更轻松地筛选数据。

表4：OpenAlex数据库实体概览

仅举几个例子，https://api.openalex.org/works?search=dna查询成果实体，用于搜索标题、摘要或全文中包含搜索词“dna”的成果；
https://api.openalex.org/authors?filter=display name.search:einstein搜索作者实体，用于查找姓名中包含“爱因斯坦”的所有作者；
https://api.openalex.org/works?filter=title.search:cubist

算法 1 展示该过程的伪代码。

获取文章：

```
过程 FetchArticles(search criteria)
    文章 ←从 OpenAlex 获取（搜索条件）      -
    对 文章 in 文章 执行
        被引用 ←获取引文 -(文章)
        文章.被引用 ←被引用       -
        引用 ←获取文章的引文
        文章.引用 ←引用
    结束循环
    返回 文章
结束过程
```

创建合著图：

```
过程 CreateCoAuthorshipGraph(articles, baseset)
    图 ←初始化空图()    -       -
    已完成 = 集合()
    对于 文章，遍历文章执行
        已完成.添加(文章)
        对于文章中的作者1执行
            将节点添加到图中（graph，作者1）
            对于文章中的作者2 执行
                将节点添加到图中（graph，作者2）
                将边添加到图中（graph，作者1，作者2）
            结束循环
        结束循环
        如果基础集合 == 为真则
            遍历文章的被引用文章 ∪ 文章引用执行
                如果文章 ∉=已完成则
                    done.add(文章)
                    遍历作者1 在文章中执行
                        将节点添加到图中（图，作者1）
                        对于作者2，在文章中 执行
                            将节点添加到图中（图，作者2）
                            将边添加到图中（图，作者1，作者2）
                        结束循环
                    结束循环
                结束条件判断
            结束循环
        结束条件判断
    结束循环
    返回 图
结束过程
```

创建引文图：

```
过程 CreateCitationGraph(articles, baseset)
    图 ←初始化空图()    -       -
    对于 文章，遍历文章执行
        向图中添加节点(图, 文章)
    结束循环
    对于文章中的每篇文章执行
        对于文章的被引用项中的每个被引用项 执行
            如果基础集合 ==False 则
                如果文章中的被引用项则
                    将边添加到图中（图、被引用项、文章）
                结束如果
            else
                如果被引用 ∉ 图则
                    将节点添加到图中(graph, 被引用) -
                结束 if
                将边添加到图中(graph, 被引用, 文章)
            结束 if
        结束 for
        对于文章.引用中的引用执行
            如果基础集合 ==False 则
                如果引用位于文章中则
                    将边添加到图中(图, 文章, 引用)
                结束如果
            else
                如果引用 ∉ 图则
                    将节点添加到图中（graph, cite）
                结束 if
                将边添加到图中（graph, article, cite）
            结束 if
        结束 for
    结束 for
    返回 图
结束过程
```

使用以下查询搜索 成果 实体：查找标题中包含“cubist”的成果；
`https://api.openalex.org/works?search=(elmo AND "sesame street") NOT(cookie OR monster)` 则搜索提及“elmo”和“sesame street”，但不包含“cookie”或“monster”的成果 。

这些查询均会返回一个包含结果的 JSON 对象；每条结果由许多字段组成，包含大量有关作者和成果的信息。有关可用字段的完整说明，请参阅 OpenAlex API 技术文档[3]。

至于该库的结构，负责数据收集和存储的模块包含以下功能：

• **与书目 API 交互**。借助自定义封装器，用户可以与 OpenAlex API 交互，从而简化学术数据的检索和分析流程。最重要的是，该封装器允许指定查询类型为“搜索”、“引用”或“被引用”，还可以选择特定于出版物的时间窗口。该功能还通过缓存减少冗余 API 调用，避免触及速率限制。根据查询类型，该功能将按表 5 所示，以 JSON 格式返回结果。

查询类型	OpenAlex 的 API	返回
`search`	*works?filter=title_and_abstract.search:*	搜索 *works* 实体，查找特定的标题和摘要中的查询内容。
`cite`	*works?filter=cites:*	检索所有明确引用该项的成果查询参数。
`cited by`	*works?filter=cited_by:*	检索所有被某篇特定文章引用的成果。

表 5：pyBiblioNet 查询类型、所使用的 OpenAlex API 及其相应返回结果。

•**检索文章与构建图**函数检索文章使用户能够执行不同类型的查询，例如搜索文章、检索引用特定成果的文章，或查找某篇文章引用的成果。该函数负责处理构造 API 请求和解析响应的复杂工作（表 6。为了便于执行复杂的搜索查询，封装器包含一个实用函数，轻量正则表达式字符串生成器，可根据简化的正则表达式模式生成查询字符串列表。这样便能根据包含多个备选项的模式，高效生成多个搜索查询。该创建引文图函数根据检索到的文章创建有向图，其中节点代表文章，边代表引文关系。此外，创建合著图函数根据检索到的文章创建无向图，其中节点代表作者；两位作者共同发表一篇文章时，两者之间便会存在一条边。这些图可用于可视化和分析引文网络的结构。这些函数利用 NetworkX 库Hagberg 等人(2008构建图，也可以将其

[3]https://docs.openalex.org/how-to-use-the-api/api-overview，最后访问日期：2025年5月15日

导出为 GML 格式，以便在图分析工具中进行进一步分析。导出功能经过专门设计，可确保与 Gephi 工具（0.10.1 版本）兼容，从而能够在外观菜单中使用分区选项（例如语言、文章类型、国家等）。在此，用户还可以指定是否只将根集纳入图中，还是也将基集文章纳入图中。

方法	Task	描述
retrieve_articles	与以下内容交互 书目 数据库	根据搜索查询获取文章， 用于检索引文并导出 结果到 CSV 文件。
create_citation_graph	构建 引文网络	构建文章及其引文关系的有向图 利用 NetworkX 处理引文关系， 可视化与分析。
create_coauthorship_graph	构建 合著 网络	构建作者及其关系的无向图 共同署名关系，利用 用于可视化和分析的 NetworkX。

表 6：用于构建引文或共同著作图的 pyBiblioNet 公共方法概览。

•**导出文章和作者的元数据**：函数将文章导出为 CSV 和将作者导出为 CSV 允许将检索到的文章和作者的元数据导出到 CSV 文件中（表 7）。用户还可以指定要导出的字段，以及是否仅包含根集文章，或同时包含基集文章。该库还通过方法将出版场所导出为 CSV 和将机构导出为 CSV支持导出文章发表场所信息和作者隶属机构。最后，函数将文章导出至 Scopus 支持元数据导出，包括以兼容 Scopus 的 CSV格式导出文章级元数据，从而可以将通过 OpenAlex 封装器检索的数据无缝用作其他文献计量分析工具的输入。 - -

方法	Task	描述
export_article_to_csv	导出元数据	以CSV格式导出文章元数据。
export_author_to_csv	导出元数据	以CSV格式导出作者元数据。
export_institution_to_csv	导出元数据	以CSV格式导出机构元数据。
export_venue_to_csv	导出元数据	以 CSV格式导出出版场所元数据。
export_article_to_scopus	导出元数据	以类似 Scopus 的 CSV 格式导出文章元数据格式。

表7：用于导出文章和作者元数据的 pyBiblioNet 公共方法概览。

•**数据分析** 表格 8 列出了为支持数据分析而设计的函数。这些方法使用户能够进行时间分析和排名分析，从而有助于解释出版动态、主题演变和学术影响力。具体而言，时间数据分析函数

例如绘制文章趋势、绘制主题趋势和绘制关键词趋势，可视化文章产出、主题变化和关键词显著性的纵向模式。像绘制顶尖作者和根据摘要绘制高频关键词 这样的排名函数，有助于识别有影响力的作者以及文章摘要中最常见的词语，还可选择考虑引文计数和外围文章的作用。

方法	Task	描述
`plot_article_trends`	时间数据 分析	绘制一个堆叠面积图，以可视化 根集文章和非根集合文章的数量 随时间变化，并按指定的时间间隔聚合。
`plot_topic_trends`	时间数据 分析	绘制分组堆叠条形图，展示 基于主题聚合的数据。
`plot_keyword_trends`	时间数据 分析	绘制一张显示高频 关键词随时间变化趋势的折线图。
`plot_top_authors`	排名分析	绘制一张显示排名前列 作者及其文章或引文按以下方式划分的堆叠条形图 主题。
`plot_top_keywords` `_from_abstracts`	排名分析	绘制提取出的出现频率最高的关键词 这些关键词来自文章摘要，包括基集 当 `show periphery=True` 时。

表8：用于数据分析的 pyBiblioNet 公共方法概览。

3.3 数据分析

在分析引文网络之前，获取由数据收集模块收集的文章语料库的描述性统计数据可能会有所帮助。可以通过文献计量分析.charts模块中的一组函数来实现。该模块包含以下函数：

- 时间分析：三种公共方法（即绘制文章趋势、绘制主题趋势和绘制关键词趋势）可对已发表文章数量、涵盖的主题及相关关键词进行时间趋势分析。

- 排名分析：公共方法绘制顶尖作者 和根据摘要绘制热门关键词 分别用于评估最多产（或被引用最多）的作者，以及从所有检索到的文章摘要中提取出的最相关关键词。

该主题是 OpenAlex 数据库中的一个实体，可以从“领域”或“域”的角度进行分析；域是“域、领域、子领域、主题”这一体系中的最高层级。在关键词提取方面，我们采用了 KeyBERT Grootendorst (2020年) 来识别最相关的术语。KeyBERT 利用 BERT 嵌入，从给定文本中提取最相关的词语或短语。该方法能够检测摘要中的关键词随时间的变化，从而揭示其时间演变情况。

3.4 基于网络的文献计量学

文献计量分析模块的核心以多种方式分析引文图和合著图。该模块支持考察结构属性、识别高影响力文章和作者，使用户能够提取中心性指标并执行聚类。

方法	Task	说明
extract_metrics_to_csv	提取网络指标	计算中心性指标（例如，介数中心性、接近中心性、PageRank），用于引文网络中的节点，并将结果导出到 CSV 文件。
show_graph_statistics	可视化网络指标	显示中心性指标的分布（例如，介数中心性、接近中心性、PageRank）通过直方图展示相关性热图，呈现中心性度量的相关性，并提供一份包含引文图统计数据的表格。
cluster_graph	文章聚类	使用以下算法为网络中的文章分配聚类：Louvain、Girvan-Newman 和 Infomap，使人们能够探索主题分组。
extract_clusters_to_csv	导出聚类结果结果	将聚类分配及相关节点信息导出到 CSV 文件，以便进一步分析。
show_clustered_graph	可视化聚类	在文献计量图中进行可视化，使用饼图展示实体分布，并提供自定义呈现选项。
show_cluster_statistics	可视化聚类规模	生成条形图，以展示网络中的聚类规模分布。

表格 9：pyBiblioNet 用于文献计量分析和聚类的公共方法概览。

表格9 概述了可用函数及其功能，包括：

•**提取网络指标**：构建引文图后，`extract metricsto csv` 函数会计算引文网络或合著网络中节点（即文章或作者）的各种中心性指标，并将这些指标与其他选定字段一起保存到 CSV 文件中。目前支持的指标包括中介中心性、接近中心性、特征向量中心性、PageRank、度、入度和出度（表格 2）。这些指标对于识别引文网络中有影响力的论文至关重要。例如，较高的中介中心性可能表明某篇论文连接了不同的研究聚类，而 PageRank 则有助于识别被有影响力的成果频繁引用的论文。

• **可视化网络统计数据**：show graph statistics 方法会生成定制图表，帮助用户更详细地探索网络统计数据。可视化内容包括网络整体统计数据摘要、展示各项中心性指标分布的直方图，以及呈现中心性指标之间关系的相关矩阵。

• **执行社区检测**：聚类图函数使用表3中列出的某种社区检测算法，为引文网络中的文章或共同著作网络中的作者分配聚类。该函数会返回一张图，其中每个节点都标注了聚类标识符，便于探索数据集中的研究领域或主题分组。目前支持的算法包括 Louvain、Girvan-Newman、Infomap、谱聚类和随机块模型；根据第2.2节所述的需求，也可以轻松纳入其他算法。还可以通过将聚类提取到 CSV方法将节点与聚类的对应关系导出为 CSV，以便进一步分析。

• **可视化社群**：pyBiblioNet 的一项关键功能是将文献计量图的社群结构可视化。该可视化功能围绕显示聚类图 展开，能够直观呈现文献计量网络中的文章或作者聚类，并通过饼图展示每个聚类中的主题（针对文章）或国家（针对作者）的分布。此功能还提供多种自定义选项，可根据具体需求调整可视化效果，例如更改节点和主题的颜色与大小，以及饼图的厚度。

4 用例：将*pyBiblioNet* 应用于“15 分钟城市”查询

本节将展示如何使用*pyBiblioNet* 库对查询“15分钟城市”进行文献计量分析。15分钟城市是一种城市规划理念，倡导居民步行或骑行15分钟即可从家到达基本服务设施 Moreno 等人 (2021）。近年来，这一主题因与城市规划和可持续发展理念相关而日益受到关注。此外，不同学科和视角的学者都对其进行了研究，因此我们可以更直观地展示聚类分析的有效性。

分析包括几个不同阶段：相关文献检索与引文网络构建，随后提取关键文献计量指标，并对引文图和合著图进行聚类。下文将详细介绍这些步骤的执行过程。

4.1 文献检索与引文图构建

在以下示例中，我们通过检索符合文献计量查询条件的文章，并使用表格 6 中所述的方法创建引文网络，来执行 pyBiblioNet 典型应用的第一步。以下代码片段展示了这一过程：

```
import json

from pybiblionet.openalex.core import string_generator_from_lite_regex, retrieve_articles,
     create_citation_graph,create_coauthorship_graph

if __name__ == "__main__" :
```

```
queries = string_generator_from_lite_regex("(15)( )(minute|min)( )(city)")

mail = "your_email@example.com"
from_publication_date = "2019-01-01"
to_publication_date = None
periphery = True

json_file_path = retrieve_articles(queries, mail, from_publication_date, to_publication_date)
#the articles are saved in a json file inside the query_result folder
articles = json.load(open(json_file_path))
G_citation=create_citation_graph(articles,"citation_graph.GML",periphery=True)
print(G_citation.number_of_nodes(),G_citation.number_of_edges())
G_coauthorship=create_coauthorship_graph(articles,"coauthorship_graph.GML",periphery=True)
print(G_coauthorship.number_of_nodes(),G_coauthorship.number_of_edges())
```

此脚本使用 OpenAlex API，根据查询“15 minute/min city”搜索文章。可以看到，我们采用了上一节介绍的一些方法，这些方法将解析查询、调用相应的 OpenAlex API 函数，以及获取并解析正确元数据等操作封装起来，以便据此构建引文图。

实际上，查询由轻量正则表达式字符串生成器 函数生成,_ 因此可以灵活组合查询词。在此示例中，我们指定日期范围从 2019年1月1日开始。随后，使用这些参数执行检索文章函数，以收集相关文章（包括基集中的文章），并将它们存储在查询结果文件夹中的 JSON 文件里。

还可以通过指定所需字段（例如，导出路径 = ‘‘15minute.csv’’），将检索到的文章导出为 CSV 文件。随后，使用创建引文图 函数，根据检索到的文章构建引文图。在此示例中，构建图时包含了基础集文章。生成的图以 GML 格式保存，非常适合在下一步中进行进一步分析。作为一种标准且广泛支持的格式，GML 也可用作 Gephi 等外部网络分析工具的输入，从而便于与现有分析工具无缝集成。

4.2 元数据导出

该库还支持将收集的数据导出为 CSV格式，方便那些觉得 `retrieve articles` 方法的 JSON 输出不便于浏览的用户使用。此功能可导出与文章、作者、机构和出版场所相关的所选元数据。对于机构和出版场所，还会包含它们在根集和基集中的出现次数。这些导出功能通过核心模块中的四种方法实现，汇总见表格 7 (有关这些方法的更多详情，请参阅附录 B)。用户可以选择特定字段并指定文件名，自定义输出内容。为增强与现有文献计量分析库和工具的互操作性，我们还提供了一项功能，可将文章导出为兼容Scopus的格式；这些格式常用作 Bibliometrix 等工具的输入，见 Aria 和 Cuccurullo (2017年),ScientopyRuiz-Rosero 等(2019年), and VOSviewer Van Eck 和 Waltman(2010年)。尽管 OpenAlex 提供的数据字段并不涵盖 Scopus 中的所有字段，我们仍将所有兼容元素进行映射，以最大限度地利用可用信息并支持

与外部工具的无缝集成。以下代码片段展示了如何导出文章元数据：既可以通过选择要包含的特定字段，导出为用户自定义的 CSV格式，也可以导出为类似 Scopus 的 CSV 格式，之后可将其用作 Python 库 pybibx 的输入，见 Pereira 等 (2025年):

```
from pybiblionet.openalex.core import export_articles_to_csv,  export_authors_to_csv
from pybibx.base import pbx_probe

if __name__ == "__main__" :


    json_file_path= "query_results/
     query_result_1ea020320b230de5a973a39682eaa53dce89a9bb026b441a5f825232.json"

    export_articles_to_csv(json_file_path,
                           fields_to_export=["venue","type","id", "doi",  "title",  "publication_date"
     "authorships_display_name", "language"],
                           export_path="15minute_articles.csv"
                           )

    export_articles_to_scopus(json_file_path,
        export_path="scopus/scopus.csv",
        include_periphery=True
    )

    filename = "scopus/myscopus.csv"
    bibfile = pbx_probe(file_bib=filename,db="scopus",del_duplicated=True)

    report= bibfile.eda_bib()
```

4.3 数据分析

表 8 介绍了该库提供的用于分析通过公共方法检索的文章时间分布的方法检索文章（更多详细信息见附录 B）。这些方法支持多种类型的可视化，包括展示根集和基集文章在指定时间间隔内分布情况的堆叠面积图、用于聚合基于主题的数据的分组堆叠条形图、展示顶尖作者及其主要主题贡献分布的堆叠条形图、基于摘要的关键词频率分析，以及随时间变化的关键词趋势分析。算法中给出的代码片段表明，可以从使用检索文章 在第 4.1 节生成的同一文件中加载文章，并生成图2所示的图表。生成这些图的方式在样式方面具有可定制性，用户可以通过调整不同级别的时间聚合或限制显示条目的数量来实现个性化可视化。

```
from pybiblionet.bibliometric_analysis.charts import plot_topic_trends,  plot_article_trends,  \
    plot_top_authors,  plot_top_keywords_from_abstracts,  plot_keyword_trends
from datetime import datetime
import matplotlib.pyplot as plt
from matplotlib import colors
import numpy as np
import json
if __name__ == "__main__" :


    json_file_path= "query_results/
     query_result_1ea020320b230de5a973a39682eaa53dce89a9bb026b441a5f825232.json"
    articles = json.load(open(json_file_path))
```

```
plot_article_trends(articles,
                    date_from=datetime(2020, 1, 1),
                    interval="quarter",   # Change to "month" or "quarter" or "year" as needed
                    date_to=datetime(2023, 12, 31),
                    num_ticks=20)
plot_topic_trends(
    articles,
    top_n_colors= [colors.to_hex(c) for c in  plt.get_cmap('tab10')(np.linspace(0, 1, 10))],
    show_core=True,
    show_periphery=True,
    interval="month",   # Change to "quarter" or "year"
    date_from=datetime(2021, 9, 1),
    date_to=datetime(2022, 12, 31),
    top_n=10,
    )
plot_top_authors(
    articles,
    date_from=datetime(2020, 1, 1),
    date_to=datetime(2022, 12, 31),
    num_authors=10,
    by_citations=True,
    show_periphery=False,
    n_colors=["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd"]
)

plot_top_keywords_from_abstracts(articles)


plot_keyword_trends(
    articles=articles,
    date_from=datetime(2020, 1, 1),
    date_to=datetime(2023, 1, 1),
    show_core=True,
    show_periphery=True,
    top_n=10,
    ngram_range=(1, 2),
    interval="quarter"
)
```

图 2a 展示了根集文章和基集文章在堆叠面积图中的分布情况。根集文章是指标题或摘要中明确包含搜索字符串“15分钟”的文章，而基集文章则不包含该字符串。根据这一搜索标准识别出的第一篇文章 [4] 可追溯至2020年1月。尽管“15分钟城市”范式最初由Carlos Moreno于2016年提出，但直到2020年才广受欢迎；当时巴黎市长Anne Hidalgo在连任竞选期间倡导采纳这一理念。事实上，自2020年起，根集文章数量不断增加，基集文章数量也随之增长。值得注意的是，出版物数量在1月达到峰值，这与科学出版中广为人知的趋势相符。一种可能的解释是，免除费用在新年伊始重置，或者引文从次年开始计算。因此，年初发表的文章可能拥有更长的引用寿命，从而提高其整体影响力。这一趋势表明，该主题的文章不仅引用早期成果，也经常引用当前的研究。图 2b 和 2c 进一步揭示了相关情况。Carlos Moreno是根集文章中被引次数很高的作者，常与Allam和Chabaud等合著者合作；主要主题领域为社会科学和物理科学

[4]需要注意的是， `retrieve articles` 也配置了日期筛选器，将结果限制为2019年1月1日之后的出版物。因此，检索数据集中不包含在此日期之前发表的根集文章

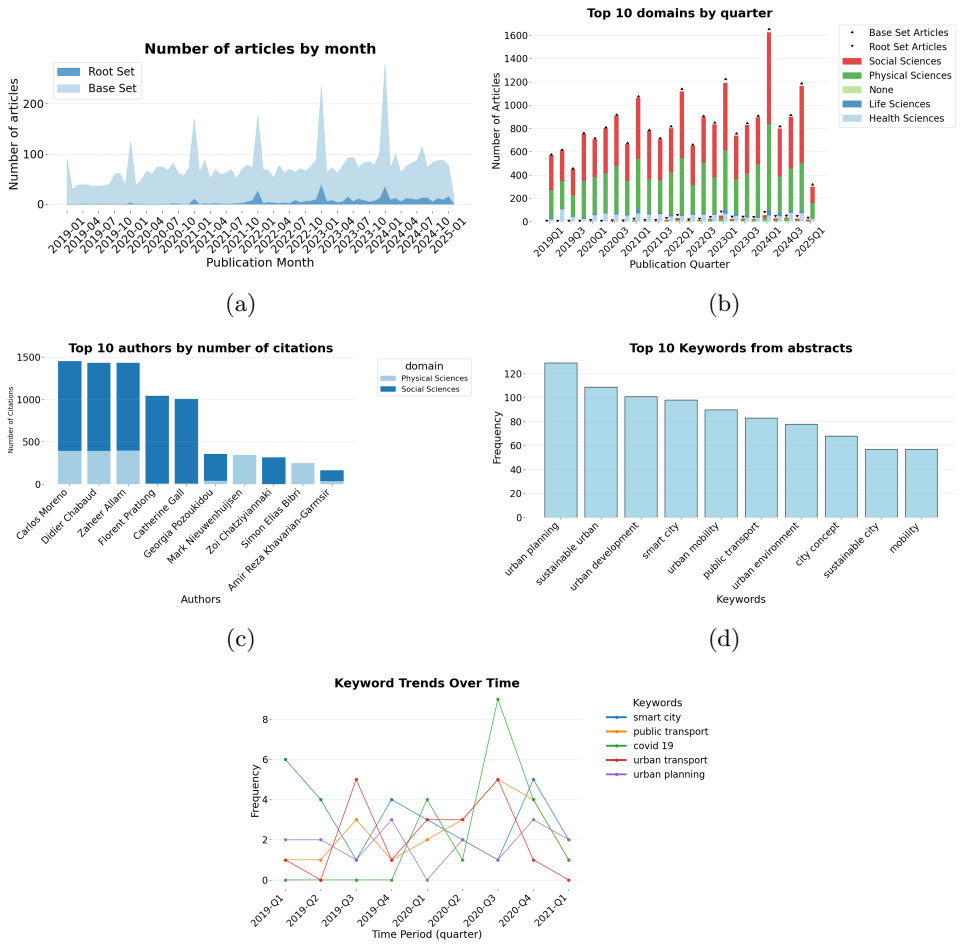


图 2：已发表文章和研究趋势的时间分析。（a）每季度发表的文章数量。（b）按月统计的文章中出现频率最高的概念。（c）按已发表文章数量排名的前 10 位作者。（d）从摘要中提取的最高频关键词。（e）高频关键词随时间的趋势。

Science。这一发现并不令人意外，因为 Moreno 是首位提出 15分钟范式的人。图 2b进一步表明，基集文章包含大量聚焦于 COVID-19 流行病学研究的出版物。这些文章在大流行病爆发后不久开始出现，并经常被根集文章引用。这种相关性可能源于 15分钟城市概念与大流行病时期的城市规划需求高度契合，有助于实现基本服务的本地化获取，并支持封锁措施。除与大流行病相关的研究外，基集文章涵盖的大多数主题都涉及

出行、城市规划和智慧城市，凸显了 15分钟城市范式在这些研究领域中的广泛相关性。

图2d 和2e 展示了使用 KeyBERT 进行关键词提取的结果[5]。具体而言，图2d以简明的条形图展示了根集文章中出现频率最高的关键词，而图2e 还展示了包括基础集文章在内的时间趋势。尽管关键词的总体频率随时间推移而上升，但“城市新冠”和“城市概念”等特定术语的增长更为显著。这一初步分析阶段为了解该主题相关话语随时间的演变、参与度最高的作者以及相关子主题提供了宝贵见解

4.4 基于网络的文献计量学

获取与 OpenAlex 查询相关的所有信息后，我们便以此构建引文图和合著图，并可利用 pyBiblioNet 中实现的方法（详见第3.4节），开展文献计量分析、社区检测，并对结果进行可视化（方法概要见9）。

```
import networkx as nx
import numpy as np
import matplotlib.pyplot as plt
from pybiblionet.bibliometric_analysis.core import extract_metrics_to_csv, cluster_graph,
     extract_clusters_to_csv
from pybiblionet.bibliometric_analysis.charts_network import show_clustered_graph,
     show_cluster_statistics,show_graph_statistics
from matplotlib import colors


if __name__ == "__main__":
    print("Reading graph...")

    json_file_path="query_results/
     query_result_1ea020320b230de5a973a39682eaa53dce89a9bb026b441a5f825232.json"

    network_file_name= "15minute_citation_graph.GML"

    G = nx.read_gml(network_file_name)


    print("Extracting metrics...")
    metrics = ["betweenness_centrality", "closeness_centrality", "page_rank", "in_degree", "
     out_degree"]
    fields = ["id", "title"]
    csv_file_path = "metrics_and_fields.csv"


    extract_metrics_to_csv(G, metrics, fields, csv_file_path)

    show_graph_statistics(G,csv_file_path)


    print("\nClustering the graph...")
    clustered_graph = cluster_graph(G, algorithm='louvain',)
    csv_file_path = "cluster_and_fields.csv"
    extract_clusters_to_csv(clustered_graph, fields, csv_file_path)
```

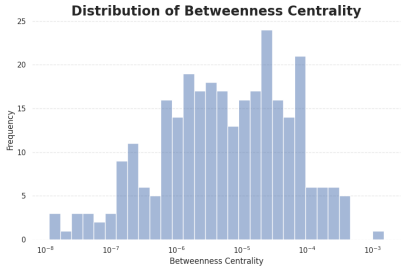
[5]由于缺少词形还原，KeyBERT 可能会返回形态相似的关键词，并将语义等同的单复数形式视为不同词语（例如 *smart city* 和 *smart cities*）。

```
print("Visualizing clusters...")
show_clustered_graph(clustered_graph, image_size=(800, 800),
                     n_clusters=5,
                     topics_level="field",
                     )

show_cluster_statistics(csv_file_path, image_size=(800, 800),
              n_clusters=5)
```

在这段代码中，我们首先从第4.1 节导出的 GML 文件中加载引文图，然后提取各种中心性度量，例如中介中心性、接近中心性和 $PageRank$（调用函数将指标提取到 CSV）。这些指标与文章的关键信息（例如标题、摘要、作者）一同存储在 CSV 文件中，从而揭示引文网络中不同文章的重要性。函数显示图统计数据 可视化一般网络统计数据、各项指标的分布以及指标的相关性热图（图3）。虽然合著网络也可以进行相同的操作，但需要注意的是，该网络是无向的。因此，入度和出度指标不适用。同样，每位作者可导出的字段类型也与文章可导出的字段不同。导出与网络度量相关联的文章元数据，即使没有高级计算方面的专业知识，也能识别出每项指标中最相关的文章。用户只需使用能够读取 CSV 文件的程序（例如 LibreOffice Calc）对数据进行排序，即可为文章排名。除了识别被引用次数最多的根集文章或基础集文章外，还可以考察特征向量中心性最高的文章，例如，这有助于发现具有更广泛系统影响力的节点，例如经常被其他高影响力成果引用的开创性论文。

然后，我们可以利用 聚类图 方法对图进行聚类，将已实现的聚类技术之一作为参数传入。在本例中，选择了 *Louvain* 算法。聚类结果连同文章元数据一起导出到另一个 CSV 文件中，以便进一步检查。可以使用 显示聚类 函数将聚类图可视化，该函数会突出显示规模最大的前 5 个聚类。输出结果见图 4a。每个聚类都附有其研究领域构成的可视化表示（展示前 5 个最大聚类中识别出的出现频率最高的 5 个域）。不同聚类中关键主题的分布表明，“15分钟城市”这一主题具有跨学科性质；然而，在某些聚类中，特定的主要主题出现得更为频繁。例如，与其他聚类相比，其中一个聚类包含更多关于“计算机科学”和“工程学”的文章。“环境科学”相关文章的大量出现并不令人意外，因为15分钟城市范式被视为减少私家车使用、进而在以可持续性和福祉为核心的框架内降低城市污染的一种方式。图 4c 展示了各个聚类规模的条形图。图 4b和 4d 展示了由合著网络生成的同一组图表，其中显示了前 5 个最大聚类中出现频率最高的 5 个国家。两者的主要区别在于，引文网络中呈现的实体对应于聚类中文章的主要主题，而在合著网络中，则对应于所属国家

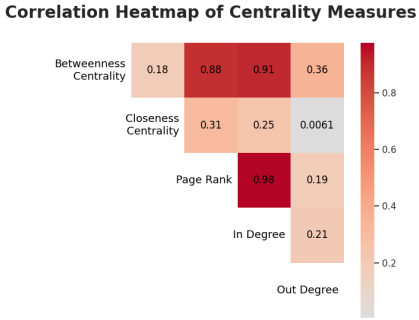


(a)

Statistic	Value
Number of Nodes	9372
Number of Edges	13855
Density	1.58e-04

Statistic	Mean	Variance
Betweenness Centrality	1.18e-06	4.04e-10
Closeness Centrality	2.18e-03	6.59e-05
Page Rank	1.07e-04	2.43e-07
In Degree	1.58e-04	1.59e-06
Out Degree	1.58e-04	7.17e-07

(b)



(c)

图 3: （a）展示采样指标的分布，其中以介数中心性作为示例。（b）使用 pyBiblioNet 查询“15分钟城市”获得的引文图统计信息。在该图中，节点代表科学论文；如果两篇论文之间存在传出或传入的引文关系，则通过一条边将它们连接起来。（c）中心性度量的相关性热图。

作者的所属机构。合著网络中的聚类显得更加相互分离，因为大多数作者聚类之间不存在共同署名关系。此外，来自不同国籍作者之间的合作显然仅限于少数几个国家，且这些国家主要位于西方，尤其是欧洲。法国机构隶属关系的作者数量较多，可以解释为该模型在巴黎获得了较高的知名度。

4.5 结果与讨论

通过 pyBiblioNet 的不同方法检索到的信息，有助于我们重建围绕“15分钟城市”这一主题的有趣文献图景。通过爬取自2019年起、针对“15 minutes/mincity”查询的数据，我们得以构建引文图和合著图，其基本统计数据汇总于图3a中。如图所示，我们共检索到 9372 篇论文；据 OpenAlex 数据，这些论文均符合用户查询条件。通过构建图并分析论文之间的连接

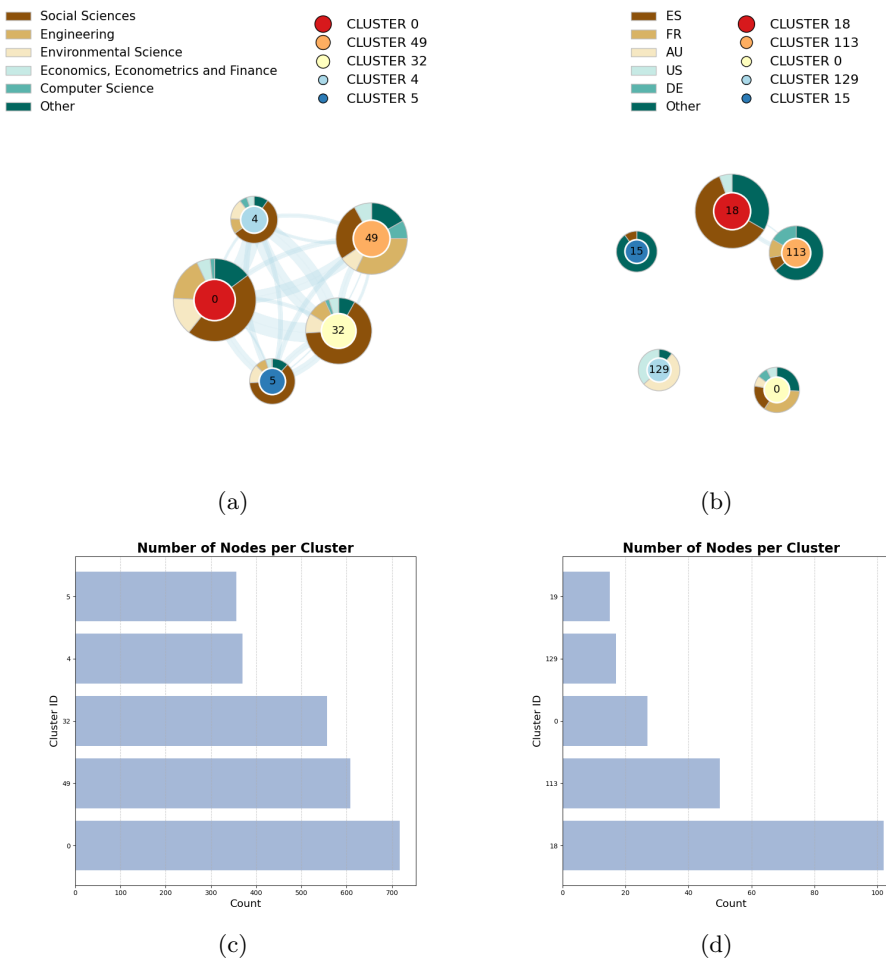


图4：(a) 和 (b) 分别通过函数显示聚类图 对引文图和合著图中的聚类进行可视化。每个编号聚类周围都标注了该聚类中论文的主要研究领域，或作者的国家隶属关系。(c) 和 (d) 分别通过函数 显示聚类规模图表 对引文图和合著图中的聚类规模进行可视化。

，我们能够识别出最核心的论文，并探索彼此联系更紧密的研究成果聚类，最终目标是缩小庞大的结果范围，根据我们的目标和需求找出最相关的成果。

简单分析表明，在这个非常稀疏的网络中，不同中心性指标之间具有较高的相关性，这说明无论采用哪种具体定义来衡量中心性，对于基础论文重要性的判断总体上都相当一致。

然而，测量中心性之后，对节点进行排名，并根据不同的中心性定义选出最有影响力的节点，可能会很有意义。这样可以分析具体差异，并了解网络中是否只有少数几篇有影响力的论文发挥着更重要的作用，还是在类型学和研究领域方面存在多样性，并量化这种多样性的程度。

通过图中所示的社区检测结果，可以进一步深入分析。4a。Louvain 算法识别出的最大聚类由各类论文构成，已经体现出较强的跨学科特征，这一点可从图中所示的聚类组成看出。4a。此外，通过查看各个社区中基于中心性的节点排名，我们可以识别出每个群体中最突出的论文。在本例中，我们仅考虑两个广为人知的指标，即介数中心性和接近中心性。表格10 和11 显示，除了少数几项在不同指标下始终表现突出的研究成果外，采用多个指标便能根据所选指标，让用户关注到不同的论文。这些论文中有许多也不同于仅通过查看表中所列的被引次数最多的研究而检索到的论文。表格 12。社区的组成在主题 和研究领域方面也并非简单明了，这表明我们面对的是一个高度跨学科的查询；在对该主题进行文献计量分析时，必须考虑到这一点。

最后，表格13 展示了合著网络中依据五种中心性指标排名前五的作者：度、介数中心性、接近中心性、特征向量和 PageRank。采用多种指标有助于捕捉影响力和合作模式的不同方面，而这些方面仅通过合作者数量无法体现。例如，B. B¨uttner 在几乎所有指标中排名都很高，这表明他不仅合作广泛，而且在网络中占据着核心且结构上至关重要的位置。相比之下，K. Geurs 或 G. Pozoukidou 等作者只出现在介数中心性排名中。这说明他们的影响力源自充当原本彼此分离群体之间的连接者，而非直接与许多人合作。特征向量和 PageRank 指标则提供了另一种视角：它们突显了与其他联系广泛或处于核心位置的合作者相连的作者。在 PageRank 排名前五的作者中，Z. Allam 和 C. Moreno 脱颖而出，这反映了他们对该主题发展的奠基性贡献；此后，该主题吸引了该领域其他领先研究人员的广泛关注与合作。总的来说，中心性度量揭示了学术人员在合作网络中的不同层面，这比单纯统计合作次数能够提供更全面、更细致的理解。

我们可以将通过 pyBiblioNet 收集和整理的研究项目，与诸如 Papadopoulos等人 （2023年）的系统性文献综述进行比较。在该研究中，作者未限定时间范围，而是使用关键词“15 min* cit*”和“20 min* cit*”从 Scopus 中识别出 78 篇会议论文和期刊研究论文。在筛选英语论文并应用全文资格标准后，他们保留了 40 篇文章。随后，他们同时采用向后与向前滚雪球法（即通过查阅所选文章的参考文献以及引用这些文章的论文，识别其他相关成果），最终形成了包含 45 篇文章的最终语料库。相比之下，使用 pyBiblioNet 进行“15minut* city”查询，并将搜索范围限定为自 2019年1月1日以来发表的文章，

聚类	标题	概念	Rank	介数中心性
0	引入 “15分钟城市”：可持续性, 韧性 and 地方 认同 in 未来 后— 大流行城市Moreno 等人(2021)	交通运输 (城市 交通 与可达性)	1°	1.48e-03
	15分钟 城市： 解析 the New 城市 规划乌托邦 Pozoukidou 和 Chatziyiannaki (2021)	交通运输 (城市 交通 与可达性)	9°	2.71e-04
	A 15分钟城市案例研究 概念 新加坡2040年陆路交通总体规划：20-分钟城镇与45分钟城市 Manifesty与 Park (2022年)	交通运输 (城市 交通 和可达性)	101°	1.94e-05
49	剖析 the ‘15分钟 城市’ via 6G、物联网, 与数字孪生：迈向新的叙事，以 提升 城市效率、韧性和可持续性 性 Allam 等人 (2022c)	媒体技术（智慧城市与 技术）	8°	3.02e-04
	The 理论, 实践, and 技术 15分钟城市的基础 模型：邻近性- 邻近性及其环境、社会与经济 可持续性效益 Allam 等人 (2022a)	媒体技术（智慧城市与 技术）	16°	1.65e-04
	基于邻近性的 规划 与 “15分钟 城市”：一种可持续 未来城市模型 未来Allam 等人 (2023年)	交通运输 (城市 交通 与可达性)	17°	1.62e-04
32	A 综合 X分钟城市骑行可达性 指标及其在评估空间和社会 经济不平等中的作用——以乌得勒支为例，荷兰 荷兰 Knap 等人(2023年)	交通运输 (城市 交通 与可达性)	23°	1.19e-04
	将15分钟城市的理念延伸至城市核心区以外: 可达性和公共交通在其中的作用 荷兰Poorthuis 和 Zook (2023年)	交通运输 (城市 交通 与可达性)	28°	8.77e-05
	理解X分钟城市的决定因素 政策：对北美和澳大利亚 城市的 规划文件 Lu 和 Diab (2023年)	交通运输 (城市 交通 与可达性)	29°	8.75e-05
4	衡量对15分钟城市概念的遵循情况: 最新进展、主要组成部分及进一步要求 Papadopoulos等人 (2023年)	交通运输 (城市 交通 与可达性)	4°	3.61e-04
	15分钟城市：城市规划与设计 致力于打造可持续社区的努力 Khavarian-Garmsir等人 (2023年)	交通运输 (城市 交通 与可达性)	6°	3.36e-04
	人人享有的15分钟城市？– 个体与时间 维度的步行可达性差异 Willberg等人(2023年)	交通运输 (城市 交通 与可达性)	10°	2.35e-04
5	城市转型与邻里规划 的回归：质疑邻近综合征与 15分钟 City Marchigiani 和 Bonfantini (2022年)	管理, 监测, 政策 及法律（城市规划与评估 ）	11°	2.24e-04
	15分钟城市视角下的巴塞罗那：绘制 基于可达性和邻近潜力的娱乐主管 步行出行时间Ferrer-Ortiz 等 (2022年)	交通运输 (城市 交通 与可达性)	12°	2.16e-04
	以邻近为名的城市未来: 城市更新的新视角: 意大利公营住房郊区在泛 疫情D'Onofrio 和 Trusiani (2022年)	管理, 监测, 政策 和法律（城市规划与估值—— 值）	35°	7.86e-05

表10：各聚类中介数中心性最高的论文

我们检索到 409 篇根集文章，其中 285 篇是以英语撰写的期刊或会议论文。计入基集后，总数增至 2615 篇文章。值得注意的是，我们发现该方法能够收集数量更多、更加显著的文章；当然，被引文献中列出的论文是经过广泛人工筛选流程得出的，研究人员对论文进行了仔细整理和评估，以确保其相关性与质量。事实上，构建文献综述并不完全取决于通过基于关键词的查询检索到的文章数量；然而，结果数量过于庞大可能会显著增加处理难度，使得针对特定主题开展系统综述既耗时又繁琐。为应对这一挑战，pyBiblioNet 提供了帮助研究人员优先处理最相关文章的机制。具体而言，可以依据文章在引文网络中的中心性度量对其进行排名，或者识别出由共同著作网络中处于结构中心位置的人员撰写的文章。此外，主题的可用性

聚类	标题	概念	Rank	接近中心性
0	介绍 “15分钟城市”：可持续性，韧性 and 地方 认同 in 未来 后-大流行城市Moreno 等人(2021)	交通运输与可达性）（城市 交通	1°	1.32e-01
	居民对适宜步行街区的偏好和 (2017年)	交通运输与可达性)(城市 交通	27°	8.03e-02
	混合土地利用对实现社会资本Nabil 和 Eldayem(2015)	社会学 and 政治 Science（城市、邻里和隔离-治理研究）	33°	8.03e-02
49	重新定义智慧城市：文化、代谢与治理 Allam 和 Newman （2018）	媒体技术（智慧城市与技术）	6°	8.34e-02
	《城市》与数字革命Allam(2020年)	媒体技术（信息与通信技术影响与政策）	7°	8.29e-02
	关于大数据、人工智能与智慧城市 Allam 和 Dhunny (2019年)	媒体技术（智慧城市与技术）	8°	8.24e-02
32	15分钟步行可达社区：测量—社会不平等及其对构建—建设健康社区的中国城市Weng等人 (2019年)	交通运输与可达性）（城市 交通	2°	9.29e-02
	实践中的可达性：20分钟城市作为可持续发展规划目标—可持续发展规划目标 Capasso Da Silva等人 (2019年)	交通运输与可达性）（城市 交通	3°	8.65e-02
	COVID-19 及随后实施的社会隔离对出行行为的影响De Vos （2020年）	交通运输与可达性）（城市 交通	19°	8.06e-02
4	可持续城市规划中的邻近性协商：瑞典案例 Gil Sol´a 和 Vilhelmson (2018)	交通运输与可达性)(城市 交通	78°	4.04e-02
	可达性是否终于迎来了它的时代? Handy （2020年）	交通运输和可达性）(城市 交通	105°	2.74e-02
	可持续交通范式Banister （2008）	交通运输与可达性）（城市 交通	113°	2.44e-02
5	步行友好型城市及邻近环境的重要性，对巴塞罗那日常出行的影响，城市 Marquet 和 Miralles-Guasch(2015)	交通运输与可达性）（城市 交通	4°	8.50e-02
	伦敦的交通与社会排斥 Church 等 (2000)	交通运输与可达性）（城市 交通	73°	4.30e-02
	15分钟城市视角下的巴塞罗那：绘制基于可达性与邻近潜力的步行出行时间Ferrer-Ortiz 等人（2022年）	交通运输与可达性）（城市 交通	110°	2.49e-02

表 11：各聚类中接近中心性最高的论文

建模可以根据主题内容筛选语料，而应用社区检测算法则能够识别出由文章或作者构成的连贯群组。随后可利用这些社群，将分析聚焦于特定子领域或研究聚类，从而更有针对性、更深入地探索科学领域。此外，通过 pyBiblioNet 识别的文章（可导出为 CSV 格式）可以轻松按语言、文档类型（例如，仅期刊论文和会议论文）以及是否纳入基础集合进行筛选。这大大简化了研究人员为开展系统综述而识别、整理和评估相关文章的工作。

5 结论、局限性与未来工作

pyBiblioNet 是一个 Python 软件包，旨在轻松收集特定查询的书目数据，并基于网络科学开展文献计量分析。为此，它包含多种方法，可完成解析查询、收集和导出书目数据，以及构建和分析引文图与合著图所需的全部步骤。虽然目前它依赖 OpenAlex 来收集文献计量数据，但 pyBiblioNet 的模块化结构允许用户加入自己的一组函数，以纳入更多数据来源，例如爱思唯尔研究产品

聚类	标题	概念	Rank	引文
0	引入“15分钟城市”：可持续性，韧性 and 地方 认同 in 未来 后大流行城市 Moreno 等人（2021）	交通运输 （城市 交通与可达性）	1°	992
	15分钟 城市： 分解 New 城市规划乌托邦 Pozoukidou 和 Chatziyiannaki（2021）	交通运输 (城市 交通与可达性）	2°	318
	15分钟 City 量化 使用 出行数据Abbiasov 等人（2022年）	交通运输 （人类 移动性与基于位置的分析）	59°	14
49	“15分钟城市”理念可以塑造净零城市未来 Allam 等人(2022年b)	《媒体技术》（智慧城市与技术）	7°	105
	解读 “15分钟 城市” 通过6G、物联网，与数字孪生：迈向……的新叙事	媒体技术（智慧城市与技术）	9°	87
	提升 城市效率、韧性与可持续性Allam 等人（2022年c）			
	The 理论， 实践， and 技术 15分钟城市的基础 模型：邻近性及其环境、社会与经济	媒体技术（智慧城市与技术）	17°	56
	对可持续性的益处 Allam 等人（2022a）			
32	15分钟城市是否 触手可及？评估前往杂货店的步行和骑行可达性	交通运输 （城市 交通与可达性）	15°	57
	在温哥华Hosford 等人（2022年）			
	A 综合 X分钟城市骑行可达性指标及其在评估空间和社会	交通运输 （城市 交通与可达性）	25°	36
	经济不平等中的作用——以乌得勒支为例，荷兰Knap 等人（2023年）			
	了解X分钟城市政策：北美和澳大利亚城市的综述	交通运输 (城市 交通与可达性)	30°	32
	澳大利亚《城市》的 规划文件 Luand Diab（2023年）			
4	15分钟城市：城市规划与设计创建可持续社区的努力	交通运输 (城市 交通与可达性)	5°	128
	Khavarian-Garmsir等人（2023年）			
	15分钟城市：解读这一模式，以提升城市韧性Abdelfattah 等人（2022年）	交通运输 （城市 交通与可达性）	10°	86
	人人享有的15分钟城市？– 测量个体和时间上的步行可达性变化	交通运输 （城市 交通与可达性）	11°	76
	Willberg等人（2023年）			
5	15分钟城市视角下的巴塞罗那：绘制基于步行出行时间的可达性与邻近潜力	交通运输 （城市 交通与可达性）	8°	98
	步行出行时间 Ferrer-Ortiz 等人（2022年）			
	城市转型与邻里规划。质疑邻近综合征与15分钟 City Marchigiani 和 Bonfantini（2022年）	管理、 监测、 政策与法律（城市规划与 Val-评价）	26°	34
	迈向采用15分钟城市的路径：我们能否理解 of 气候 政策的 可接受性	传播学（社交媒体与政治学）	62°	13
	解释针对 X分钟城市计划的反弹——方案？ Marquet 等(2024年)			

表12：按聚类划分的根集论文引文数排名

应用程序接口[6] 或 PubMed 应用程序接口[7],只要该模块的最终输出采用便于执行文献计量分析的模块读取的结构即可。

pyBiblioNet 在探索高度跨学科的科学领域方面尤其有帮助，这一点可从第 4 节的用例中看出。以网络分析指标为基础开展文献计量分析非常有效，因为这使我们能够评估出版物、作者和研究主题之间的结构性关系。通过利用网络分析指标，pyBiblioNet 能够识别关键作者、有影响力的出版物和新兴研究聚类。这种方法能够深入揭示科学领域内的协作模式、学科交叉和知识演化，展现出传统文献计量方法可能忽略的趋势和有影响力的节点。此外，在这样的跨学科环境中，使用这一技术有助于研究人员探索不熟悉的领域。

[6]https://dev.elsevier.com/ [7]
https://www.ncbi.nlm.nih.gov/home/develop/api/

Rank	度		介数中心性		接近中心性	
1	B. B¨uttner	1.40e-1	B. B¨uttner	1.10e-2	B. B¨uttner	1.40e-1
2	M. T. B. Larriva	1.03e-1	K. Geurs	1.87e-3	A. Gorrini	1.12e-1
3	J. Nofre	1.02e-1	G. Pozoukidou	9.51e-4	D. Presicce	1.12e-1
4	J. Carpio-Pinedo	1.02e-1	J. Nofre	6.34e-4	F. Messa	1.12e-1
5	A. Gorrini	1.02e-1	L. Abdelfattah	6.34e-4	M. T. B. Larriva	1.12e-1
Rank	**特征向量**		**PageRank**			
1	B. B¨uttner	1.01e-1	Z. Allam	3.24e-3		
2	M. T. B. Larriva	1.00e-1	B. B¨uttner	2.98e-3		
3	A. Gorrini	1.00e-1	C. Moreno	2.51e-3		
4	D. Presicce	1.00e-1	B. Murgante	2.49e-3		
5	F. Messa	1.00e-1	O. Marquet	2.40e-3		

表13：按不同中心性度量排名前 5 的作者

值得注意的是，我们发现，我们的方法能够收集数量更多且更加多样化的来源；当然，系统性文献综述中呈现的论文清单，是经过大量人工筛选的结果。在这一过程中，论文会经过仔细整理和评估，以确保其相关性和质量。相比之下，pyBiblioNet 是一种辅助工具，定位于研究人员工作流程的初始阶段，旨在以最少的努力将广泛的成果纳入考量。通过利用网络分析，pyBiblioNet 能够从与查询相关的众多主题中识别出有影响力的出版物。这种方法最大限度地减少了研究人员先入之见可能造成的偏见，促进对该领域进行更加全面且公正的探索。因此，它增强了研究人员获取更广泛且可能更丰富的资源池的能力，而这些资源随后还可以在后续阶段接受更细致的人工筛选。这种自动化与人工综述流程的结合，优化了文献综述的严谨性和广度，推动形成对科学领域更加异质化的认识。

不过，我们也必须承认 pyBiblioNet 当前实现存在一些局限。具体而言，分析范围仅限于引文网络和合著网络。还可以探索其他网络结构，例如作者间引文网络。此外，我们采用了一组选定的中心性度量和社区检测算法；若纳入更广泛的指标和社区检测算法，则可以深化文献计量分析。另外，通过支持以标准化格式导出数据（例如 Web of Science 、 Dimensions 或 PubMed 使用的格式；这些格式常被许多文献计量分析工具接受为输入），可以提升 collection 模块与 Python 和 R 中其他工具或库之间的互操作性。此外，主题建模模块目前相当基础，并且没有实现任何缓存机制，这可能会显著影响其在大型数据集上的性能。最后，生成的可视化内容是静态的，且可定制程度有限；引入交互式、可配置的可视化输出，将使该工具更加灵活，也更便于更多用户使用。我们计划在该库的未来版本中解决其中一些局限。

附录 A openalex 模块文档

本附录提供 openalex 模块的文档，详细介绍其公共方法及其功能。

A.1 核心函数

A.1.1 轻量正则表达式字符串生成器 _

根据简化的正则表达式模式生成查询字符串列表。**参数：**

- 查询 正则表达式 (str)：表示简化正则表达式模式的字符串。

返回：

- 列表 [字符串]：查询字符串列表。

A.1.2 查询 OpenAlex

根据给定参数查询 OpenAlex API。**参**

数：

- API 类型(str)：API 查询类型（'搜索'、'引用'、'被引用'）。
- 参数 (可选[str])：用于"引用"和"被引用"查询的文章 ID。 _
- 邮件 (str)：用于识别 API 请求的电子邮件地址。
- 起始出版日期(Optional[str])：用于筛选的开始日期（YYYY-MM-DD）。
- 截止出版日期(Optional[str])：用于筛选的结束日期（YYYY-MM-DD）。
- 缓存 (bool)：使用缓存。

返回：

- 列表 [字典 [字符串，字典]]：与查询匹配的结果列表。

A.1.3 检索文章 _

根据查询列表和可选参数，从 OpenAlex 检索根集文章。

参数：

- 查询(列表[字符串])：搜索查询。
- 电子邮件 (字符串)：用于识别 API 请求的电子邮件地址。
- 起始出版日期(可选[字符串])：开始日期 (YYYY-MM-DD)。
- 至出版日期（可选[str]）：结束日期（YYYY-MM-DD）。

返回：

- str：JSON 结果的文件路径。

A.1.4 将文章导出为 CSV _ _

将文章数据从 JSON 文件导出为 CSV。**参数：**

- JSON_文件路径 (str)：JSON 文件的路径。
- 要导出的字段（可选[List[str]]）：要包含的字段。
- 导出路径 (O可选 l[str]): Path for th导出的 CSV .

返回：
- 无：将选定的字段保存到 CSV 文件。

A.1.5 将作者导出为 CSV_ _

将作者数据从 JSON 文件导出为 CSV。**参数：**

- JSON_文件路径 (str)：JSON 文件的路径。
- 要导出的字段(Optional[List[str]])：要包含的字段。
- 导出路径 (Optional[str])：导出 CSV 的路径。

返回值：
- 无：将选定字段保存到 CSV 文件。

A.1.6 将机构导出为 CSV _ _

将机构数据从 JSON 文件导出为 CSV。**参数：**

- JSON_文件路径 （str）：JSON 文件的路径。
- 要导出的字段（可选[List[str]]）：要包含的字段。
- 导出路径 （可选[str]）：导出 CSV 的路径。

返回值：
- None：将选定字段保存到 CSV 文件。

A.1.7 将出版场所导出为_CSV

将 JSON 文件中的出版场所数据导出为 CSV。**参数：**

- JSON_文件路径 （str）：JSON 文件的路径。
- 要导出的字段（可选[列表[字符串]]）：要包含的字段。
- 导出路径 （可选[字符串]）：导出 CSV 的路径。

返回：
- 无：将所选字段保存到 CSV 文件。

A.1.8 将文章导出为 Scopus 格式

将 JSON 文件中的作者数据导出为类似 Scopus 的 CSV。**参数：**

- JSON_文件路径 （str）：JSON 文件的路径。

• 包含外围_（可选[布尔值]）：包含基集。
• 导出路径（可选[str]）：导出 CSV 的路径。

返回：
• None：将选定字段保存到 CSV 文件。

A.1.9 创建引文图

根据表示引文的文章数据创建有向图。**参数：**

• 文章（Dict[str, Dict]）：文章数据。
• 导出路径（可选[str]）：导出图的路径。**返回：**

• nx.DiGraph：无向合著图。

A.1.10 创建合著图

创建表示共同署名关系的无向图。

参数：
• 文章 (Dict[str, Dict]：文章数据。
• 导出路径 （可选[str]）：用于导出图的路径。
• 外围 （bool）：包含基础集文章。**返回：**

• nx.Graph：无向合著图。

附录 B bibliometric analysis 模块文档

B.1 核心函数

B.1.1 将指标提取到 CSV

从图中提取指定的指标和字段，并将其保存到 CSV 文件。**参数：**

• G (nx.DiGraph|nx.Graph)：NetworkX 有向图或无向图。
•指标 (List[str])：要计算的指标列表（可用项：'中介中心性'，'接近中心性'，'PageRank'，'入度'，'出度'，'度'）。
• 字段(List[str])：要包含在 CSV 中的字段列表。
• CSV_文件路径 (str)：用于保存结果的 CSV 文件路径。

返回：
• 无：将选定的字段和聚类保存到 CSV 文件。

B.1.2 将聚类导出到 CSV _ _

从图中提取聚类和选定的字段，并将其保存到 CSV 文件。**参数：**

- G (nx.DiGraph|nx.Graph)：一个 NetworkX 有向图或无向图。
- 字段(List[str])：要包含在 CSV 中的字段列表。
- CSV_文件路径(str)：用于保存结果的 CSV 文件路径。

返回：
- None：将选定的字段和指标保存到 CSV 文件。

B.1.3 聚类图 _

使用指定的聚类算法对图中的节点进行聚类，并返回一个以聚类编号作为节点属性的新图。

参数：
- G (nx.DiGraph|nx.Graph)：NetworkX 有向图或无向图。
- 算法 (str)：要使用的聚类算法（int）：使用谱聚类时要创建的聚类数量。

返回：
- 一个 NetworkX 图，其中聚类编号作为新的节点属性。

B.2 图表函数

B.2.1 绘制文章趋势 _

生成堆叠面积图，展示根集和基础集文章数量随时间的变化，并按指定的时间间隔（月、季度、年）汇总。

参数：
- 文章 （字典）：包含文章及其元数据的字典。
- 时间间隔(str)：横轴标签的时间间隔。可以是"月"、"季度"或"年"。
- 核心颜色(可选[str])：图中根集文章的颜色。
- 外围颜色(可选[str])：基集文章的颜色。
- 起始日期(可选[datetime])：聚合范围的开始日期。
- 结束日期(可选[datetime])：聚合范围的结束日期。
- 刻度数(可选[int])：要显示的 x 轴刻度最大数量。

返回值：
- 无：该函数生成并显示堆叠面积图。

B.2.2 绘制主题趋势_

生成分组堆叠条形图，以可视化主题或概念随时间变化的聚合数据。文章可以拆分为根集文章和基础集文章，并可选择按特定主题或概念进行筛选。

参数：

- 文章 (字典)：包含带有元数据的文章的字典。
- 时间间隔 (可选 [字符串])：x 轴的聚合时间间隔（“month”、“quarter”、“year”）。
- 前 n 种颜色(可选 [列表])：用于条形的颜色列表。
- 显示外围(可选 [布尔值])：若为 `True`，则显示基础集文章。
- 显示核心(可选 [布尔值])：若为 `True`，则显示根集文章。
- 字段键(可选 [字符串])：用于筛选主题的键（“field”或 “domain”）。
- 起始日期(可选 [日期时间])：用于筛选的开始日期。
- 结束日期(可选 [日期时间])：用于筛选的结束日期。
- 前 n 个 (可选 [整数])：要显示的主题/概念数量。
- 刻度数 (可选 [整数])：x 轴上的最大刻度数。

返回：

- 无：该函数会生成并显示分组堆叠条形图。

B.2.3 绘制顶尖作者

生成堆叠条形图，展示顶尖作者及其文章，并按主题或概念分类。可根据引文数或出版物数对作者进行排名。

参数：

- 文章 (`dict`)：包含文章及其元数据的字典。
- 字段键(可选 [`str`])：用于筛选主题的键（“领域” 或 “域”）。
- 起始日期(可选 [`datetime`])：筛选的开始日期。
- 结束日期(可选 [日期时间]): 用于筛选的结束日期。
- 作者数量 (整数): 要显示的顶级作者数量。
- 按引文排序 (布尔值): 如果为 `True`，则按引文数量而非出版物数量对作者进行排名。
- 显示外围(布尔值): 如果为 `True`，则在分析中包含基础集文章。
- 显示核心 (布尔值): 如果为 `True`，则在分析中包含根集文章。
- 颜色数量 (可选 [列表]): 图表条形的颜色列表。

返回值：

- 无：该函数会生成并显示堆叠条形图。

B.2.4 从摘要中绘制热门关键词图

生成图表，以可视化从文章摘要中提取的最高频关键词。

参数：

- 文章 （字典）： 包含带有元数据的文章的字典。
- 起始日期（可选 [日期时间]）： 用于筛选的开始日期。
- 结束日期（可选日期时间]）： 用于筛选的结束日期。
- 显示核心 （布尔值）： 如果为`True`，则在分析中包括根集文章。
- 显示外围（布尔值）： 如果为`True`，则在分析中包含基础集文章。

- 前 n 个（整数）：要显示的高频关键词数量。
- n 元语法范围（元组）：用于提取关键词的 n 元语法范围。

返回：

- 无：该函数会生成并显示图表。

B.2.5 绘制关键词趋势

生成图表，以可视化高频关键词随时间变化的趋势。**参数：**

- 文章（字典）：包含带有元数据的文章的字典。
- 起始日期（可选[日期时间]）：用于筛选的开始日期。
- 结束日期(可选[日期时间])：用于筛选的结束日期。
- 显示核心 (布尔值)：如果为 `True`，则在分析中包含根集文章。
- 显示外围 (布尔值)：如果为 `True`，则在分析中包含基础集文章。
- 前 n 个(整数)：要显示的高频关键词数量。
- n 元语法范围(元组)：用于提取关键词的 N 元语法范围。
- 时间间隔(字符串)：聚合的时间间隔（“月”、“季度”、“年”）。
- 聚类数(可选[列表[字符串]])：颜色列表

返回：

- 无：该函数会生成并显示图表。

B.3 ChartsNetworks 函数

B.3.1 显示聚类图

在图中可视化聚类，并使用饼图表示每个聚类中条目的分布。它还支持将文章信息导出为 CSV 文件。

参数：

- G(`nx.DiGraph`)：包含带有聚类数据的节点的 NetworkX 有向图。
- 聚类数 n (可选[整数])：要显示的前列聚类数量。
- 条目数 m (可选[整数])：饼图中要显示的前列条目数量。
- 聚类颜色数 n (可选[列表[字符串]])：聚类颜色列表。
- 条目颜色数 m (可选[列表[字符串]])：饼图条目的颜色列表。
- 最小节点半径（可选[浮点数]）：节点的最小半径。
- 最大节点半径 （可选[浮点数]）：节点的最大半径。
- 最小饼图半径（可选[浮点数]）：饼图的最小半径。
- 最大饼图半径 （可选[浮点数]）：饼图的最大半径。
- 图例标记大小（可选[整数]）：图例标记的大小。
- 图例字体大小（可选[整数]）：图例字体的大小。
- 节点字体大小（可选[int]）：节点注释的字体大小。
- 最小边宽度（可选[int]）：聚类之间边的最小宽度。
- 最大边宽度 （可选[int]）：聚类之间边的最大宽度。
- 边颜色 （可选[str]）：边的颜色。
- 要显示的前 m 个条目 (可选[列表[字符串]])：要显示的前列条目列表。

• 用于锚定图例条目的边界框 (可选 [元组 [浮点数，浮点数，浮点数，浮点数]])：图例条目的边界框。
• 用于锚定图例簇的边界框 (可选 [元组 [浮点数, 浮点数, 浮点数, 浮点数]])：聚类图例的边界框。

• 图像尺寸(可选 [元组 [整数, 整数]])：输出图像的尺寸。
• 主题层级 (可选 [字符串])：要考虑的主题层级（“领域” 或“域”）。
• PNG 导出路径 (可选 [字符串])：将图形导出为 PNG 的路径。

返回值：
• 无：该函数会生成并显示带有饼图的聚类图。

B.3.2 显示聚类统计信息

显示一张水平条形图，展示每个聚类中的节点数量，并附有一张包含聚类统计信息的表格。

参数：
• CSV_文件路径 (`str`)：包含聚类数据的 CSV 文件路径。
• 颜色(可选 [`str`])：条形图的颜色。
• 图像尺寸 (可选 [元组 [`int, int`]])：输出图表的图像尺寸。
• 聚类数量 (可选 [整数])：要显示的顶级聚类数量。

返回：
• 无：该函数会生成并显示条形图和统计表。

B.3.3 显示图统计信息

显示图的各种统计信息，包括节点数量、边数量、图密度以及各种中心性度量。此外，还会生成相关矩阵和中心性度量的分布直方图。

参数：
• `G` (`nx.Graph`)：要显示其统计信息的图。
• CSV_文件路径 (`str`)：包含中心性度量数据的 CSV 文件路径。
• 标题颜色(`str`，可选)：表格标题的颜色。

返回：
•`None`：该函数会生成并显示各种统计信息，包括表格、相关性热图和直方图。

参考文献

Abbiasov T、Heine C、Glaeser EL 等（2022年）使用出行数据量化15分钟城市。技术报告，国家经济研究局

Abdelfattah L, Deponte D, Fossa G（2022）15分钟城市：阐释这一模型，发掘城市韧性。《交通运输研究进展》60:330–337。
https://doi.org/https://doi.org/10.1016/j.trpro.2021.12.043，安全出行的新情景

城市地区安全出行《第二十五届城市生活与步行国际会议论文集》（LWC 2021），2021年9月9日至10日，意大利布雷西亚。Allam Z（2020）《城市与数字革命》。施普林格。Allam Z, Dhunny ZA（2019）大数据、人工智能与智慧城市。《城市》89:80–91。
https://doi.org/https://doi.org/10.1016/j.cities.2019.01.032Allam Z, Newman P（2018）重新定义智慧城市：文化、代谢与治理。《智慧城市》1(1):4–25。Alla m Z, Bibri SE, Chabaud D, 等（2022a）15分钟城市模型的理论、实践与技术基础：邻近性及其对可持续性的环境、社会与经济效益。《能源》15(16):6042。Allam Z, Bibri SE, Chabaud D, 等（2022b）“15分钟城市”概念能够塑造城市净零未来。《人文与社会科学传播》9(1):126。https://doi.org/10.1057/s41599-022-01145-0Allam Z, Bibri SE, Jones DS, 等（2022c）通过6G、物联网和数字孪生解读“15分钟城市”：构建提升城市效率、韧性与可持续性的新叙事。《传感器》22(4):1369。Allam Z, Moreno C, Chabaud D, 等（2023）基于邻近性的规划与“15分钟城市”：面向未来城市的可持续模型。施普林格国际出版，瑞士查姆，第1523–1542页。
https://doi.org/10.1007/978-3-031-01949-4 178及KB（2017）居民对适宜步行街区的偏好。《城市设计杂志》22(1):44–58。
https://doi.org/10.1080/13574809.2016.1234335Andersen N（2021）绘制外派人员研究文献图谱：对1998年至2017年该领域的文献计量综述及当前研究前沿的识别。《国际人力资源管理杂志》32(22):4687–4724。Andrikopoulos A, Economou L（2016）金融经济学中的合著与次级署名模式。《国际金融分析评论》46:12–19。Aria M, Cuccurullo C（2017）Bibliometrix：用于综合科学知识图谱分析的R工具。《信息计量学杂志》11(4):959–975。Baker HK, Pandey N, Kumar S, 等（2020）董事会多元化的文献计量分析：现状、发展与未来研究方向。《商业研究杂志》108:232–246。Banister D（2008）可持续交通范式。《交通政策》15(2):73–80。
https://doi.org/https://doi.org/10.1016/j.tranpol.2007.10.005，城市交通规划的新进展

Bastian M、Heymann S、Jacomy M（2009）Gephi：用于探索和操纵网络的开源软件。载于：国际人工智能促进协会网络与社交媒体会议论文集，第361–362页。Blondel VD、Guillaume JL、Lambiotte R 等（2008）大型网络中社群的快速展开。《统计力学杂志：理论与实验》2008(10):P10008。Capasso Da Silva D、King DA、Lemar S（2019）实践中的可达性：20分钟城市作为可持续性规划目标。《可持续性》12(1):129。Church A、Frost M、Sullivan K（2000）伦敦的交通与社会排斥。《交通政策》7(3):195–205。https://doi.org/https://doi.org/10.1016/S0967-070X(00)00024-XCisneros L、Ibanescu M、Keen C 等（2018）1939年至2017年间家族企业继承的文献计量研究：作者网络的绘制与分析。《科学计量学》117:919–951。Cole V、Boutet M（2023）Researchrabbit。《加拿大健康图书馆协会期刊》44(2):43–47。https://doi.org/10.29173/jchla29699，URL https://doi.org/10.29173/jchla29699，© Cole 和 Boutet。未声明存在利益冲突。，JCHLA-44-043De Vos J（2020）COVID-19及随后实施的社交距离措施对出行行为的影响。《交通研究跨学科视角》5:100121。https://doi.org/https://doi.org/10.1016/j.trip.2020.100121Diallo SY、Lynch CJ、Gore R 等（2016）通过网络中心性度量识别期刊中的关键论文。《科学计量学》107:1005–1020。Donthu N、Kumar S、Mukherjee D 等（2021）如何开展文献计量分析：概述与指南。《商业研究杂志》133:285–296。D'Onofrio R、Trusiani E（2022）以邻近为名的城市未来：大流行病之后意大利公营住房郊区城市更新的新视角。《可持续性》14(3):1252。Ferrer-Ortiz C、Marquet O、Mojica L 等（2022）15分钟城市视角下的巴塞罗那：基于步行出行时间绘制可达性和邻近潜力。《智慧城市》5(1):146–161。Fire M、Guestrin C（2019）学术出版指标的过度优化：观察古德哈特定律的实际作用。《GigaScience》8(6):giz053。Fortunato S（2010）图中的社区检测。《物理报告》486(3):75–174。

Gil Sol´a A、Vilhelmson B（2018）可持续城市规划中的邻近性协商：瑞典案例。《可持续性》11(1):31。Girvan M、Newman ME（2002）社会网络和生物网络中的社群结构。《美国国家科学院院刊》99(12):7821–7826。Grootendorst M（2020）KeyBERT：基于BERT的轻量级关键词提取。https://doi.org/10.5281/zenodo.4461265，URL https://doi.org/10.5281/zenodo.4461265Hagberg A、Swart PJ、Schult DA（2008）使用NetworkX探索网络结构、动态和功能。技术报告，洛斯阿拉莫斯国家实验室（LANL），美国新墨西哥州洛斯阿拉莫斯。Handy S（2020）可达性这一理念是否终于迎来了它的时代？《交通研究D辑：交通与环境》83:102319。https://doi.org/https://doi.org/10.1016/j.trd.2020.102319

Holland PW、Laskey KB、Leinhardt S（1983）随机区块模型：初步探索。《SocialNetworks》5(2):109–137

Hosford K、Beairsto J、Winters M（2022）15分钟城市是否触手可及？评估温哥华杂货店的步行和骑行可达性。《交通研究跨学科视角》14:100602。https://doi.org/https://doi.org/10.1016/j.trip.2022.100602

Jin Y 等（2024）基于属性和交互的社区检测方法。《国际社会网络挖掘杂志》10(1):1–25

Khavarian-Garmsir AR、Sharifi A、Sadeghi A（2023）15分钟城市：致力于打造可持续社区的城市规划与设计工作。《城市》132:104101。https://doi.org/https://doi.org/10.1016/j.cities.2022.104101Kleinberg JM（1999）超链接环境中的权威来源。《美国计算机协会杂志》46(5):604–632。https://doi.org/10.1145/324133.324140Knap E、Ulak MB、Geurs KT 等（2023）X分钟城市骑行可达性综合指标及其在评估空间与社会经济不平等中的作用——以荷兰乌得勒支为例。《城市交通杂志》3:100043。https://doi.org/https://doi.org/10.1016/j.urbmob.2022.100043Li J、Lai S、Shuai Z 等（2024）图中社区检测的综合综述。《神经计算》600:128169。https://doi.org/https://doi.org/10.1016/j.neucom.2024.128169，网址 https://www.sciencedirect.com/science/article/pii/S0925231224009408

Liu Z, Yin Y, Liu W 等（2015）创新系统研究的知识结构与演变可视化：文献计量分析。《科学计量学》103：135–158

Lu M, Diab E（2023）理解X分钟城市政策的决定因素：北美和澳大利亚城市规划文件综述。《城市交通杂志》3：100040。https://doi.org/https://doi.org/10.1016/j.urbmob.2022.100040Manifesty OR, Park JY（2022）新加坡2040年陆路交通总体规划中的15分钟城市概念案例研究：20分钟城镇与45分钟城市。《国际可持续交通技术期刊》5(1)：1–11 Marchigiani E, Bonfantini B（2022）城市转型与邻里规划的回归：质疑邻近综合征与15分钟城市。《可持续性》14(9)：5468 Marquet O, Miralles-Guasch C（2015）步行友好型城市与邻近环境对巴塞罗那日常出行的重要性。《城市》42：258–266。https://doi.org/https://doi.org/10.1016/j.cities.2014.10.012Marquet O, MojicaL, Fern´andez-N´u˜nez MB 等（2024）迈向15分钟城市的采用：我们对气候政策可接受性的理解能否解释针对X分钟城市计划的反弹？《城市》148：104878。https://doi.org/https://doi.org/10.1016/j.cities.2024.104878Moral-Mu˜noz JA, Herrera-Viedma E, Santisteban-Espejo A 等（2020）用于开展科学文献计量分析的软件工具：最新综述。《信息专业》´on 29(1) Moreno C, Allam Z, Chabaud D 等（2021）介绍“15分钟城市”：未来后疫情城市的可持续性、韧性与地方认同。《智慧城市》4(1)：93–111 Nabil NA, Eldayem GEA（2015）混合土地利用对实现社会资本的影响。《HBRC期刊》11(2)：285–298。https://doi.org/https://doi.org/10.1016/j.hbrcj.2014.03.009，网址 https://www.sciencedirect.com/science/article/pii/S1687404814000303Ng AY, Jordan MI, Weiss Y（2001）论谱聚类：分析与算法。载于：《神经信息处理系统进展》 Papadopoulos E, Sdoukopoulos A, Politis I（2023）衡量对15分钟城市概念的遵循情况：最新进展、主要组成部分及进一步要求。《可持续城市与社会》104875页 Pereira V, Basilio MP, Santos CHT（2025）Pybibx——一个由人工智能工具赋能的文献计量与科学计量分析Python库。《数据技术与应用》59(2)：302–337。https://doi.org/10.1108/DTA-08-2023-0461

Poorthuis A, Zook M（2023）将15分钟城市拓展至城市核心区之外：可达性与公共交通在荷兰的作用。《交通地理学杂志》110：103629。
https://doi.org/https://doi.org/10.1016/j.jtrangeo.2023.103629

Pozoukidou G, Chatziyiannaki Z（2021）15分钟城市：解构新城市规划乌托邦。《可持续性》13（2）：928 Priem J, Piwowar H, Orr R（2022）OpenAlex：学术成果、作者、出版场所、机构和概念的完全开放索引。arXiv预印本 arXiv:220501833 Rossetto DE, Bernardes RC, Borini FM 等（2018）过去60年创新研究的结构与演变：通过引文和共引分析回顾商业领域的研究及未来趋势。《科学计量学》115（3）：1329–1363

Rosvall M, Bergstrom CT (2008) 复杂网络上的随机游走图揭示了社群结构。《美国国家科学院院刊》105(4):1118–1123

Ruffo G, SemeraroA, Giachanou A, et al (2023年) 研究假新闻传播、极化动态及机器人操纵：网络与语言的故事。《计算机科学评论》47:100531

Ruiz-Rosero J,Ramirez-Gonzalez G, Viveros-Delgado J (2019年) 软件综述：Sci-entopy，一款用于分析科学出版物主题趋势的科学计量工具。《科学计量学》121(2):1165–1188

Salman HA、Ahmad MA、Ibrahim R 等（2025）生成式人工智能工具在学术研究和同行评审中的整合：系统性分析。《Online Journal of Communication and Media Technologies》网址
https://api.semanticscholar.org/CorpusID:275403583Tosi D（2025）生成式人工智能文献综述与人工主导的系统性文献综述比较：大数据研究案例。《IEEE Access》13：56210–56219。网址
https://api.semanticscholar.org/CorpusID:277359598Tunger D、Eulerich M（2018）德语国家公司治理研究的文献计量分析：使用定制数据库将文献计量学应用于商业研究。《科学计量学》117：2041–2059。Van Eck N、Waltman L（2010）软件调查：VOSviewer，一款用于文献计量映射的计算机程序。《科学计量学》84（2）：523–538。Van Eck NJ、Waltman L（2014）文献计量网络可视化。载于：《衡量学术影响力：方法与实践》。施普林格，第285–320页

Weng M、Ding N、Li J 等（2019）15分钟步行可达社区：测量、社会不平等，以及对在中国城市建设健康社区的启示。《Journal of Transport & Health》13：259–273。https://doi.org/https://doi.org/10.1016/j.jth.2019.05.005

WillbergE, Fink C, Toivonen T（2023年）人人享有的15分钟城市？——测量步行可达性的个体差异和时间变化。《交通地理学杂志》106:103521。
https://doi.org/https://doi.org/10.1016/j.jtrangeo.2022.103521