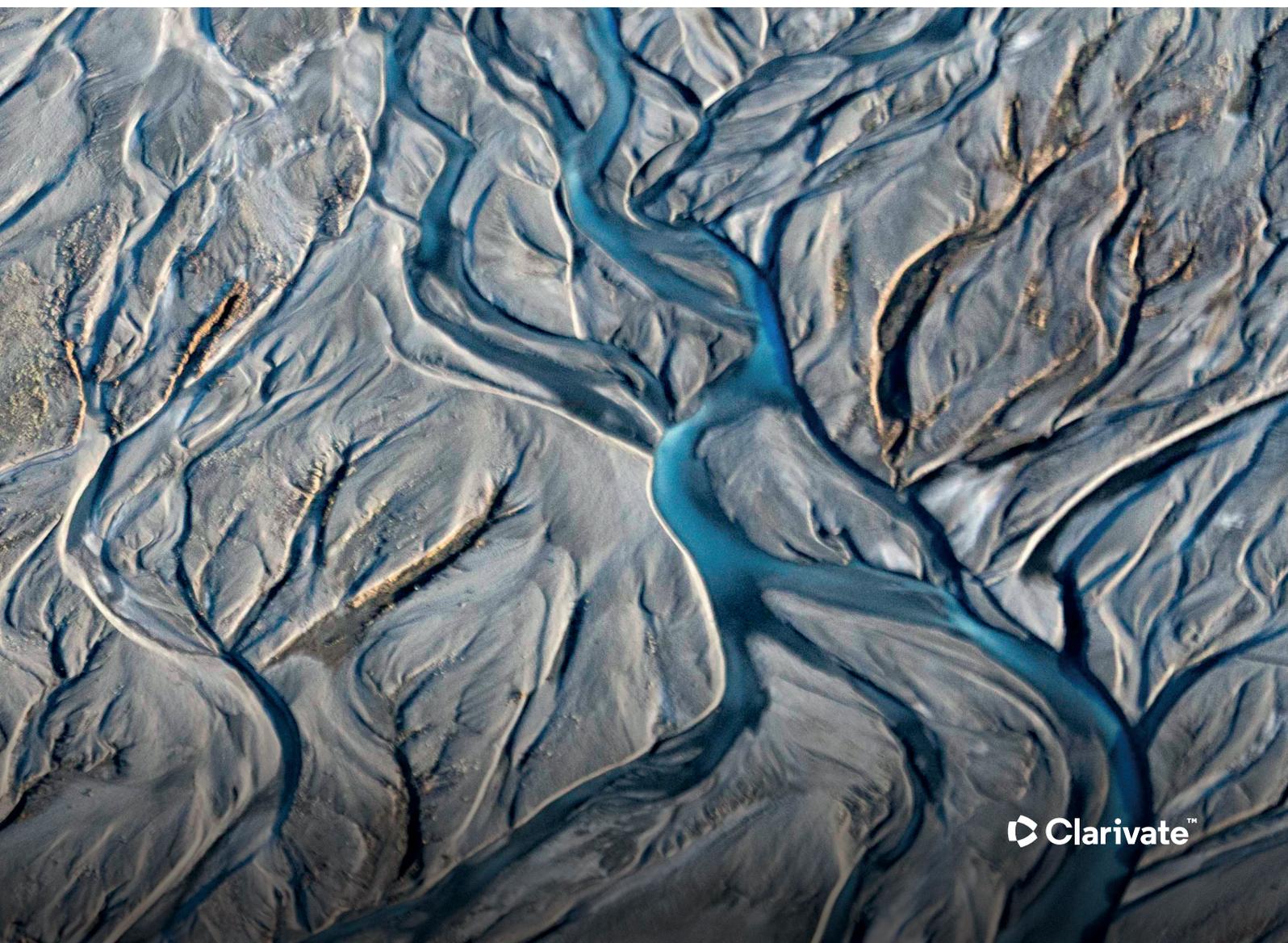


2020 年 9 月

全球研究報告 識別 Web of Science 中的 研究前沿：從指標到意義

Martin Szomszor、David Pendlebury 和 Gordon Rogers



作者簡介

Martin Szomszor 博士為現任 Institute for Scientific Information (ISI) 總監暨 ISI 科研數據分析主管。他與英國高等教育資助委員會 (HEFCE) 合作創建了 REF2014 影響力案例研究資料庫 (REF2014 impact case studies database)，因此被評為 2015 年英國資訊時代 (Information Age) 前 50 名資料領導者。

David Pendlebury 是 ISI 的 研究分析主管。自 1983 年以來，他一直致力於使用 Web of Science 資料來探尋科學研究的結構和動態。他與 ISI 創辦人 Eugene Garfield 共事多年，並與 Henry Small 共同開發了 Essential Science Indicators (ESI) 資料庫。

Gordan Rogers 是 ISI 資深數據科學家。他在書目計量學及數據分析領域有超過 10 年的經驗，協助全球客戶評估研究組合及策略。

奠基過去，展望未來

關於 Institute for Scientific Information (ISI)

科睿唯安旗下的 Institute for Scientific Information (ISI) 自成立伊始，半個多世紀以來一直引領著全球科學資訊的研究。今天，ISI 致力於推動研究誠信，改善科學資訊的檢索、解釋和應用。作為科睿唯安學術研究業務的知識研究機構，ISI 透過活

動、會議與出版物對外進行知識傳遞，同時進行基礎研究，讓知識研究庫可以持續擴展和更臻完善。更多資訊請參考：<https://clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/>

執行摘要

本報告鼓勵研究人員和研究團隊經理，利用 Web of Science 的研究前沿資料，以及專業領域的結構和動態地貌圖，來深入評估學術研究。

研究評估和政策制訂經常要參考量化指標，以出版和引用作為基礎，輔助傳統的專家同儕審查流程。研究社群的人士大部分都很熟悉標準指標，例如引用次數、Web of Science 期刊影響因子 (Journal Impact Factor™) 或 H-Index 指數。分數和排名有其用處，但研究活動面向多重，貢獻也有不同的維度，難以藉此方式呈現。除此之外，現在已有評鑑方法可以提供更多資訊 – 只是採用率仍然極低。

由於超大型資料集在處理和視覺化方面有所進步，我們已能對文獻進行科學圖譜化 (science mapping)，藉此瞭解並掌握科學與學術研究的前線。這類圖譜通常會針對研究的領域和主題，提供平面或立體全景圖，其組成元素為連結不同出版品的引用網絡，以及共同術語。文獻的相似度決定全景圖中的接近程度，出版品的密度則會決定結構，例如會形成知識的「山群」或「島嶼」。分析師可以在全景圖中找出個人、機構、贊助者和期刊，評估不同領域的組織參與情形，以及時間變化。如此可更加瞭解當前的活動，包括識別關鍵參與者、熱門主題、新興主題。

由於超大型資料集在處理和視覺化方面有所進步，我們已能對文獻進行科學圖譜化 (science mapping)，藉此瞭解並掌握科學與學術研究的前線。

書目計量法與研究評估

研究評估檢視過往的研究歷程，包括：投入 (經費及其他的可能資源)、活動 (計畫)、產出 (通常是編輯過的文件，例如學術論文或產業專利)，以及成果 (論文引用，以及逐漸受到重視的社會和經濟效益)。大部分的評估都沒有妥善處理最後一項成果，因為活動結束後緊接著就要進行評估，尚未產出明顯效益。

因此，想要深入綜合評估就不免要往後回顧，必須檢視許久以前的研究，方能評估過去幾年的投資成效。

若研究投資無法達到贊助者期待的成果，將造成難以彌補的損失。但若研究人員無法掌握機會募集資金，就必須盡可能符合效率，將資源用在刀口上，以利發揮最大成效。

這表示研究人員不僅要支持在同儕審查中最被看好的研究案例，還要從最有可能創新的領域中做出選擇，盡可能產出明確的社會和經濟效益。

學科正規化的引文影響力 (CNCI) 是廣為使用的一項傳統指標。若有學術論文參考引用先前的論文，先前發表的論文就會累積引用次數。

一般認為，較常被引用的論文比無人引用的論文產生更多影響，或更具學術「影響力」。然而引用次數除了會隨時間累積，其累積頻率也會受到學科和論文類型的影響。平均而言，生命科學論文的引用比率較高，勝過科技和應用科學論文，而同時期的文獻綜述 (review article) 引用次數通常也高過其他類型。有鑑於這些差異，任何文獻的引用次數都必須與同年度發表、屬於相同研究領域的同類論文全球平均引用次數相比。文獻引用次數與全球平均引用次數的比率就是 CNCI。

文獻的 CNCI 為即時產生的數據，平均 CNCI 經常被視為判斷指標，用來顯示某個國家、機構或研究團體的文獻影響力。這項指標的解讀能力具有潛在缺點，大部分的使用者應

該都已注意到。但我們會發現，由於引用次數需要時間來累積，所以 CNCI 指數一定要經過一段時間，才能計算出來。

相對較簡單的傳統回顧式指標，顯然無法滿足研究管理上的主要訴求。因此眾人呼籲，必須開發更即時的研究活動檢視方法，幫忙解決這樣的困境。

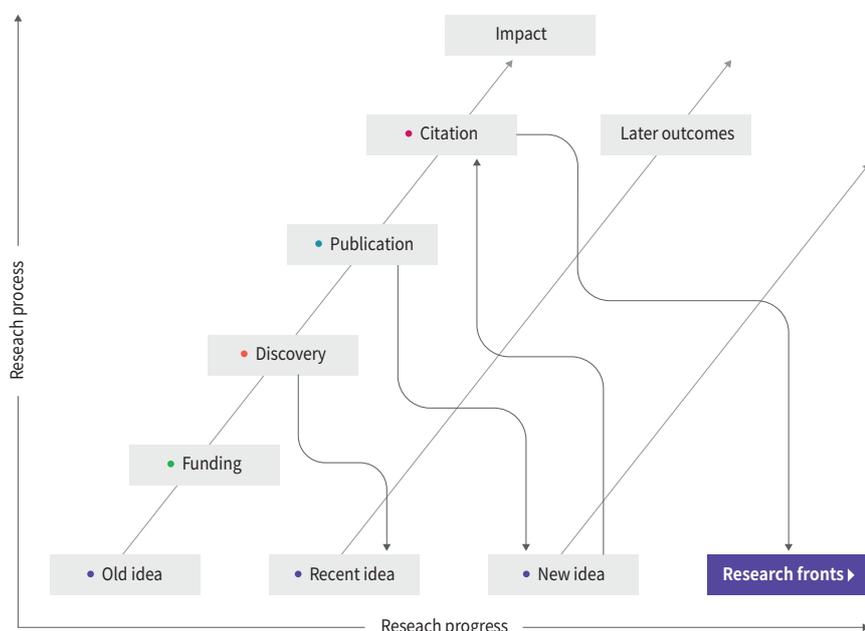
關鍵在於，要從評估研究歷程，轉換到評估研究進展。傳統指標把焦點放在歷程：基本上，這些指標分析的是研究計畫的結果。但計畫都是來自於更大的分支，大的支幹代表該研究領域的進展。我們必須評估自己與主幹的距離。

計畫支線發展、新興知識、主幹的概念演進，這三者之間存在許

多回饋循環，會透過新舊出版品的交叉引用呈現出來。美國科學資訊研究所 (Institute for Scientific Information, ISI) 的 Eugene Garfield，便是以此為基礎，開發出 Science Citation Index™ (SCI)，稱此為「*聯結想法的索引*」(an *association-of-ideas index*)。他認為引用聯結可以串起科學主題、概念和方法：「引用能精準明確地代表某個主題，不需解讀，不受術語改變影響。」(1955) 引用本身橫跨不同領域，引用網絡的連結不受限於單一或數個領域，而是自然地在研究全景上隨處漫遊。

Garfield 認為引用資料能提供素材，幫助建立科學研究架構的樣貌，並且描繪出研究的範疇。只要能透過引用來連結論文的索引，我們即可據此判斷論文之間的知識關係，Derek de Solla Price (1965) 也表示：「書目式參考文獻能指出科學研究前沿的本質。」此模式為我們提供了一張地圖，讓我們找出研究出版品的位置，並且得出反映知識歷程發展方向的時間軸。我們可以找出研究主題在哪裡，以及相關主題研究的發展方向。但在 Price 身處的年代，尚無法實現他心目中的全球科學圖譜。

圖 1： 研究流程與研究進展。引用回饋循環會增添資訊，讓我們更瞭解發展過程。我們留意獲得許多論文引用的高影響力研究，如此將有可能加快進展。



研究前沿是什麼？

Price 認為研究有一個定義明確且可依據引用模式找出的「前沿」。根據他的描述，相較於早期的文獻，近期文獻會有「一系列」或不成比例的引用叢集，稱為「即時因素」(immediacy factor)。他表示：「由於先前的文獻只有一小部分彙總為新年度論文集，人們可能會將這一小部分視為某種生長點或表皮層，當作具有活力的研究前沿。」(Price 1963)

這一個世紀以來，研究前沿的相關文獻穩定增加，並在最近二十年快速發展。現在「研究前沿」是公認的詞彙，經常與研究趨勢、增長領

域、新興的產業和主題相關聯，突顯出我們希望而且有可能識別創新和改變的焦點。這個詞彙本身也包含創新的概念，不只是構想上的創新，也是領域上的創新。因此，現有的類型學或分類可能經常不適當，甚至有可能妨礙人們找出這類的創新。

研究前沿的近期論文經常討論視覺化，強調偵測新興主題。視覺化將描述研究前沿的研究，連結至更廣泛的整體學術知識映射研究。首先，關鍵問題在如何繪製這些地圖，其次則是如何在地圖上找出關鍵點。

現在「研究前沿」是公認的詞彙，經常與研究趨勢、增長領域、新興的產業和主題相關聯。

我們如何描繪科學圖譜？

如果沒有電腦運算能力、儲存能力、龐大的資料，就只能以人工的方式，有選擇性地從出版品和引用資料來進行研究前沿分析。有許多方式能蒐集研究文獻並彙總為單一叢集，然後形成特定領域或網絡。Web of Science 使用期刊分類法，但不會指明各類別之間的遠近關係。

針對個別出版品，我們可以運用文本來分類 (例如參考摘要的相似度或共同關鍵字)，但文本分析會是一項繁重的工作，需要有詳盡的語料庫，

因為同一個字在不同領域，可能會有完全不同的意思。其他詮釋資料包括文件的參考資料書目或引用清單。Kessler (1963) 提出書目耦合的技巧，根據文件的共同引用文獻頻率來評估不同文件的主題相似度。

1973 年，ISI 的 Henry Small 將 Kessler 的方法顛倒過來：

「有一種新的文獻耦合方式，稱為共被引，意思是兩份文件同時被引用的頻率。想要知道兩篇科學論文

的共被引頻率，可以比較 SCI 資料庫裡的引用文獻清單，計入符合的條目。共被引論文網絡可從特定科學專門領域產生……共被引論文叢集為專業領域科學架構提供了新的研究方法。」

俄羅斯資訊科學家 Irena V. Marshakova-Shaikovich 也在同一時間提出共被引分析的概念，但她和 Small 都不曉得對方的研究 – 也就是科學社會學家 Robert K. Merton 所說的「重複發現」。

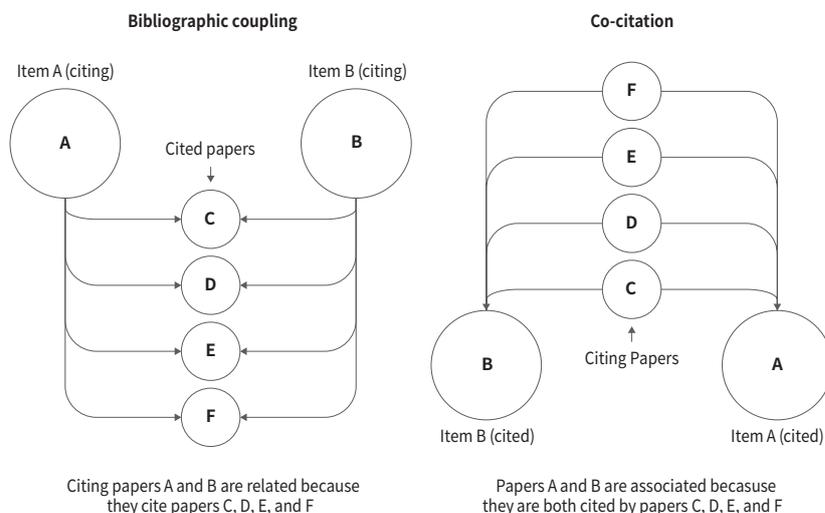
Small 用兩份文件的共被引次數來評估文件的相似度：此為文件的共被引頻率。他分析粒子物理學的論文後發現，從共被引模式可以看出「主題相似度的概念」以及「概念的關聯性或同時共現性」。他指出，經常被引用的論文反映出關鍵的概念、方法或實驗，可以作為共被引分析的起點，用來破譯專業領域中社會和知識結構的主題。在 Price 的研究前沿中，緊密交織的近期論文群組相對來說會比較小，Small 也發現，共被引分析主要針對專業領域，專業領域成為自然的研究單位，而非符合傳統定義的大型領域。他還發現，這類組織單位可以隨著時間演進繼續研究。

隨後 Small 與 Belver C. Griffith (費城卓克索大學) 為共被引分析的專業領域定義，打下了基礎。Small and Griffith (1974; Griffith et al., 1974) 指出，個別研究前沿之間的相似度可以評估，進而形成專業領域的核心。他們用多元尺度法來映射，並以相似度來繪製兩個維度之間的接近程度。Price (1979) 大力稱讚，認為此法具有「革命性的意義」。

1981 年 Garfield 將 Small 和 Griffith 的基礎研究轉變成資訊產品「ISI Atlas of Science: Biochemistry and Molecular Biology, 1978/80」。Atlas 包括 102 項研究前沿，每一項都有核心論文的地圖，以及用多元尺度法標示的論文關係。共 102 項研究前沿，依照彼此的相似度，繪製於一張大型的折疊地圖。ISI Atlas of Science 並未傳承下來，但 Garfield 和 Small 繼續研究科學圖譜化的技術。後來 Small (1985) 針對研究前沿的定義，做出重大修正：部分共被引叢集。他根據引用論文的文獻清單長度來計算部分引用頻率，調整不同領域平均引用頻率之間的差異。以數學為例，整體計算讓這類強勢學科，始終沒有被充分呈現。Small 也證明研究前沿的相似度叢

圖 2：

Kessler 的書目耦合 (左) 與 Small 和 Marshakova 的共被引分析 (右) 有何差異



集，層級高於個別前沿的群組。他和 Garfield (1985) 彙總這些進展，並且結合 SCI 和 Social Sciences Citation Index™ (SSCI) 資料庫的資料，進而推出一份全球科學圖譜。

我們必須強調，建立研究出版品集群沒有所謂最好的方式。將研究「資訊」分門別類時會面臨到的挑戰，就是我們沒有黃金準則，也沒有檢驗正確性的絕對方法可以參考。我們擁有的是諸多研究人員的文化觀點，而這些觀點會受他們的來源地、訓練、經驗和不斷發展的領域觀點以及他人觀點所影響。對化學家來說，數學領域的主題分野模糊不清。對歷史學家來說，奈米科技在化學、材料和數學的跨幅，就像拜占庭史那麼錯綜複雜。

透過共被引分析識別專業領域，可以找出橫跨熟悉領域、在知識上互有關聯的研究主題。想要為研究管理和未來決策提供更實用的指引，就必須將專業領域置於更廣的地圖上，這份地圖要能顯示可辨識的主要和次要研究領域。唯有那時我們才能完整解讀挑選出來的主題。

現在全世界有許多學術中心，運用各式各樣的技術和工具，致力於鑽研科學圖譜化的領域。印第安納大學教授 Katy Borner (2010) 的 Atlas of Science 摘述了這些後續發展。其中，卓克索大學 Chaomei Chen (2006) 開發的 CiteSpace 和萊頓大學科學技術中心 Nees-Jan Van Eck 和 Ludo Waltman (2010) 開發的 VOSviewer 尤為重要。

關於科學映射背景的更多詳細資訊，本文參考 Eugenio Petrovich 的最新文獻綜述 (2020)，以及最新科技指標手冊中的兩篇概述 (Boyack 和 Klavans 2019, Thijs 2019)。

Web of Science 研究前沿的運用及價值

在知識全景圖上，找出卓越研究的「高峰」，可提供重要資訊。這些高峰代表高被引論文，在研究前沿中互相連結，其重要程度可以加重計算。引用交叉連結至當前最受關注的主題，可能會為現有領域帶來突破，或是成為具有創新度、可能橫跨不同學科的新興研究領域。

當知識網絡確定找出研究前沿，就會出現重要的管理機會，範疇遠超研究績效指標得出的資訊。

- **研究人員**
識別研究前沿可為研究工作指引明燈。作者可找出當前活動，瞭解研究工作與研究前沿的距離。
- **學術機構**
研究團隊經理可判斷組織產出在知識全景圖上的分布情形，篩選出近期或長期的機會，然後評估研究叢集與研究前沿之間的關係。也可以針對競爭機構進行比較評估。
- **研究贊助人**
識別資助計畫出版品分布情形，可讓研究機構瞭解投資的研究是否位於或接近研究前沿，也可以考慮轉投資符合研究前沿主題的計畫。

- **政策制訂者**
一個國家的研究在知識全景圖中如何分布，可放在國際研究中比較，也會受到研究前沿所涉國家的關注，尤其是與政策優先目標相關的領域。
- **出版商**
期刊內容在全景圖中的位置，不僅可放在較廣的學科範疇中觀察，也可以從中瞭解，期刊內容與研究前沿最熱門主題的關係。編輯政策可在適當機會隨之調整。

中國和日本的國立研究機構證實，識別研究前沿本身具有重要的政策價值，可為投資決策提供資訊，並指出新的機會所在。

中國科學院 (CAS)

CAS 為何仰賴研究前沿

- CAS 發現 ESI 研究前沿提及的專業領域，與透過其他管道找出的熱門研究方向相符合。
- 領域專家也證實，研究前沿的核心論文大多是研究領域的經典研究文獻。因此，研究前沿可當作導覽工具，讓研究人員更瞭解某個研究領域。

- 研究前沿產生自共被引分析，CAS 運用研究前沿來分析核心論文，找出專業研究領域的關鍵參與者。
- 檢視 CAS 的引用論文不僅能追蹤最新進展，還能瞭解特定領域的發展方向。
- CAS 對研究前沿關鍵用途案例的分析
 - 自 2014 年起推出並發表中文和英文版的研究前沿年度報告。
- CAS 也根據研究前沿報告中的熱門和新興主題，開發一套指數，用以評估全球主要國家的研究活動，並從 2017 年起發表《研究前沿熱度指數》(Research Fronts – Active Fields, Leading Countries) 年度報告。
- CAS 用研究前沿分析特定研究領域：
 - 數學與物理學的科學發展圖譜：供中國國家自然科學基金委員會運用，針對數學和物理學領域進行策略研究。

- 奈米科學研究的研究前沿分析：與中國國家奈米科學中心合作。
- 中國土地科學與技術進展：中國土地資源部。
- 農業機械研究與技術進展：中國科學技術部。
- 印刷與造紙業分析報告：中國製漿造紙研究院。
- CAS 受研究前沿啟發，針對特殊領域舉辦研討會。
- 2017 年合成生物學研究前沿研討會；

– 2018 年阿茲海默症、系外行星、鈣鈦礦材料研究前沿研討會。

CAS 如何運用研究前沿

- 用關鍵字搜尋研究前沿，找出與研究領域相關的研究前沿，並且分析核心論文和研究前沿的引用論文。由領域專家解讀分析結果。
- 撰寫研究前沿年度報告。
- 具備領域知識的 CAS 分析師檢視 Web of Science 研究前沿，最後選出十大熱門研究前沿，並分別從十大領域中，各挑選出一項新興研究前沿。

· 依照選出的研究前沿分別研究其核心論文，根據領域知識來重新命名選出來的研究前沿。

- 分析並展示每一年的熱門研究前沿分布情形。
- CAS 開發兩項指標，在各項大的範疇裡挑選關鍵的熱門研究前沿和新興研究前沿，以利進一步解讀。
- 分析國家與組織對關鍵熱門研究前沿中核心論文和引用論文的貢獻。
- 解讀關鍵新興研究前沿的內容、研究活動和持續趨勢。

日本科學技術振興機構 (Japan Science & Technology Agency, JST)

JST 為何仰賴研究前沿

- 傳統的同儕審查科學出版領域，版圖正在快速擴大。有一部分原因是，中國這類主要經濟體大量出版科學刊物，顯示該國的科學社群正快速茁壯。
- 科學論文數量猛增，讓科學家無法像以前那樣，調查某個研究領域的所有科學論文。

- 因此，我們需要適當縮小焦點，以書目計量分析來避免因人偏誤。
- 研究前沿叢集源自於高共被引論文，將 Web of Science 基本科學指標 (Essential Science Indicators™, ESI) 領域與出版年份中引用排名前 1% 的 10,000 篇論文縮減至 3,000 份有必要加以分析的文件。

JST 對研究前沿關鍵用途案例的分析

- JST 運用研究前沿，在頂尖科學文獻中找出關鍵與新興主題，作為審查和贊助的候選主題。
- 此外，JST 透過社會需求分析來排定主題優先順序，評估研究主題的科學定位。

JST 如何運用研究前沿

- JST 團隊調查核心論文 (高被引論文) 的標題和摘要，以人工方式為研究前沿標記。
- 標籤有助於劃定分析範圍，包括國際標竿分析、國內組合分析以及關鍵人才識別。
- 除了標準指標 (例如：核心論文數量、平均出版年)，JST 還加入原始指數 (例如：中國作者出現頻率、核心論文的自然期刊指數 (Nature Index) 百分比)。
- 為了彌補引用式分析的時間落差，JST 也留意科睿唯安找出的熱門論文 (近幾個月的高被引論文)。

研究領域的圖譜是否仰賴映射方法？

在研究知識全景圖中，將特定出版品和相關出版群組的位置視覺化，能幫助我們順利解讀並真正瞭解某項知識發現的進展。我們可以辨識已建立主題的熟悉叢集，找出高被

引論文，追蹤研究前沿中這類論文的連結網絡（往往橫跨不同的主題領域），並進一步判斷相近程度，例如我們自己的論文和我們的組織相近度如何。

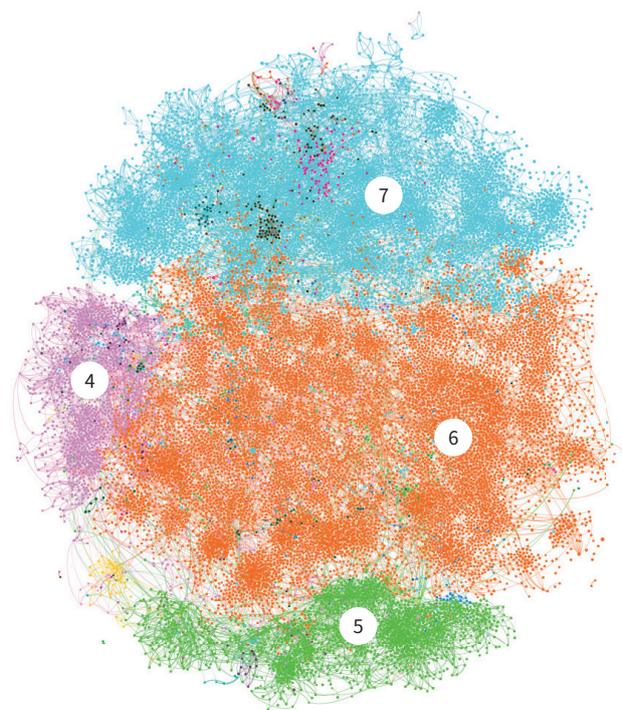
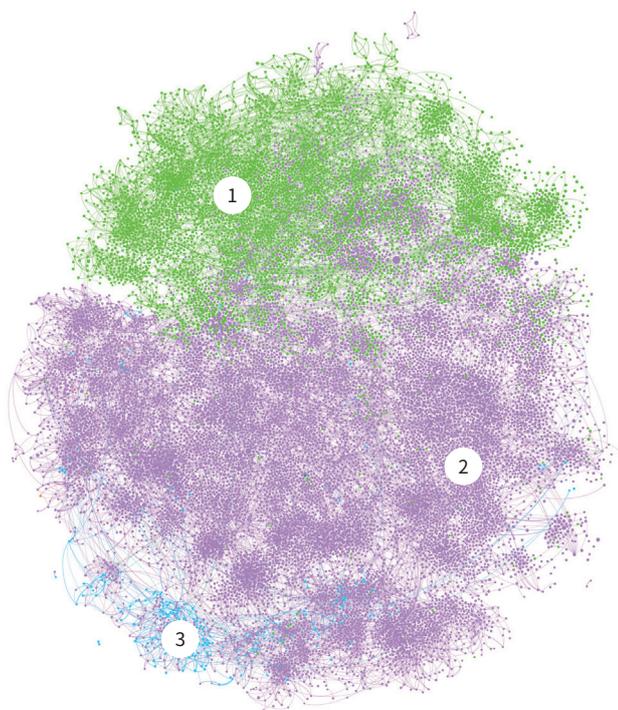
大家通常會問，壓縮成兩項維度，來建立作者較為熟悉的全景圖，進而產生一份圖譜（亦即，論文在圖形中的空間配置），這麼做是否具有效力，可否重複操作。

圖 3：

針對由第三種流程（主題建模方法）決定之出版領域（地圖），比較兩項不同分類流程所辨識出來的主題（ESI 期刊分類和 CWTS 主題引用集群）

Essential Science Indicators

萊頓大學科學技術中心



1. 物理學
2. 太空科學
3. 地球科學
4. 火星、起源、演化、表面、月球、動力學、太陽系、大氣、模型、任務
5. 電漿、亂流、模型、波、傳輸、托卡馬克、太陽：日冕、太陽：磁場、動力學
6. 活躍星系核、演化、數位巡天、銀河：演化、銀河：活躍、放射、方法：數值分析、恆星形成、恆星方法：資料分析
7. 模型、IHC、重力、QCD、宇宙、標準模型、搜尋、廣義相對論、質量

我們可以用論文的子集合來說明確實如此。以這個例子來說，我們運用 2016 年在 Web of Science 天文學和天文物理學類別，所發表的總共 19,000 篇論文，然後用文本分析作為依據，進一步判斷論文的相對位置（實際上是論文標題和摘要的相似度），並為了作圖需要畫成圓盤形。

然後我們在這個簡單的圓盤圖譜上，找出兩種不同的獨立分類系統中相同的論文，並為其標色：一種系統是 ESI 的領域類別，以期刊為基礎；另一種是由萊頓大學科學技術中心 (CWTS) 開發的分類系統，以直接引用連結為基礎。圖片顯示，這兩種系統建立的類別叢集，與我們最初以不相干方法建立的全景圖前後一

致。一系列論文的詮釋資料來源，在找出類別和主題方面本身即具有一致性。在建立這些「科學圖譜」的跨類別效度後，我們便可進一步檢視特定的研究前沿圖譜。

我們的全球圖譜如何產生？

共有兩個步驟：由期刊提供框架；由核心與共同引用論文提供細節。

美國科學資訊研究所 ISI 利用期刊引用資料，為映射提供了一套框架。他們採用現代機器學習網絡分析運算法 Node2Vec (Grover & Leskovec 2016)，根據期刊引用剖面圖（例如，期刊引用其他期刊的比率）為每一份期刊建立摘要特徵向量。我們可以把任何期刊放進壓縮過的特徵空間，用多重投射運算法 UMAP 在二維座標上標示期刊位置 (McInnes & Healy 2018)。我們可以用這種技術來產生一張地圖，將知識上具有相似性（即共同引用）的期刊畫在鄰近的位置（局部接近度），同時維持跨領域和跨學科的整體發展情形（整體方位）。

因為有這一系列的參考點（例如期刊的位置），所以只要建立被引用或引

用文獻的剖面圖，並利用 UMAP 投射法將其置入地圖，就有可能畫出任何論文或論文的位置。參考文獻範圍非常狹隘的論文（例如只參考一兩份期刊）會被集中在其他同類論文的附近。參考多元文獻的論文（參考各式各樣的期刊）會依照其他參考資料，來決定是否從主要叢集中取出，放置到地圖上的另外一個區域。舉例來說，討論「邏輯」的論文會放在數學和電腦科學的交界處，成為地圖上這兩項主要領域的連接橋梁。

框架用的是熱力圖，看起來很像島嶼圖：藍色海洋環繞綠色島嶼，有灰色和白色的山峰。

這是貼切的類比，因為這些的確是知識的島嶼，四周被相對平靜的海洋所環繞。島嶼的山峰高度取決於各個位置上期刊的相對數量，以及期刊的知識接近程度。我們可以檢

視每座島嶼的人口組成，然後為其貼上「部落」標籤 (Becher and Trowler, 2001)；在這個例子是找出 ESI 類別，有些例子則是找出 Web of Science 裡的期刊叢集。

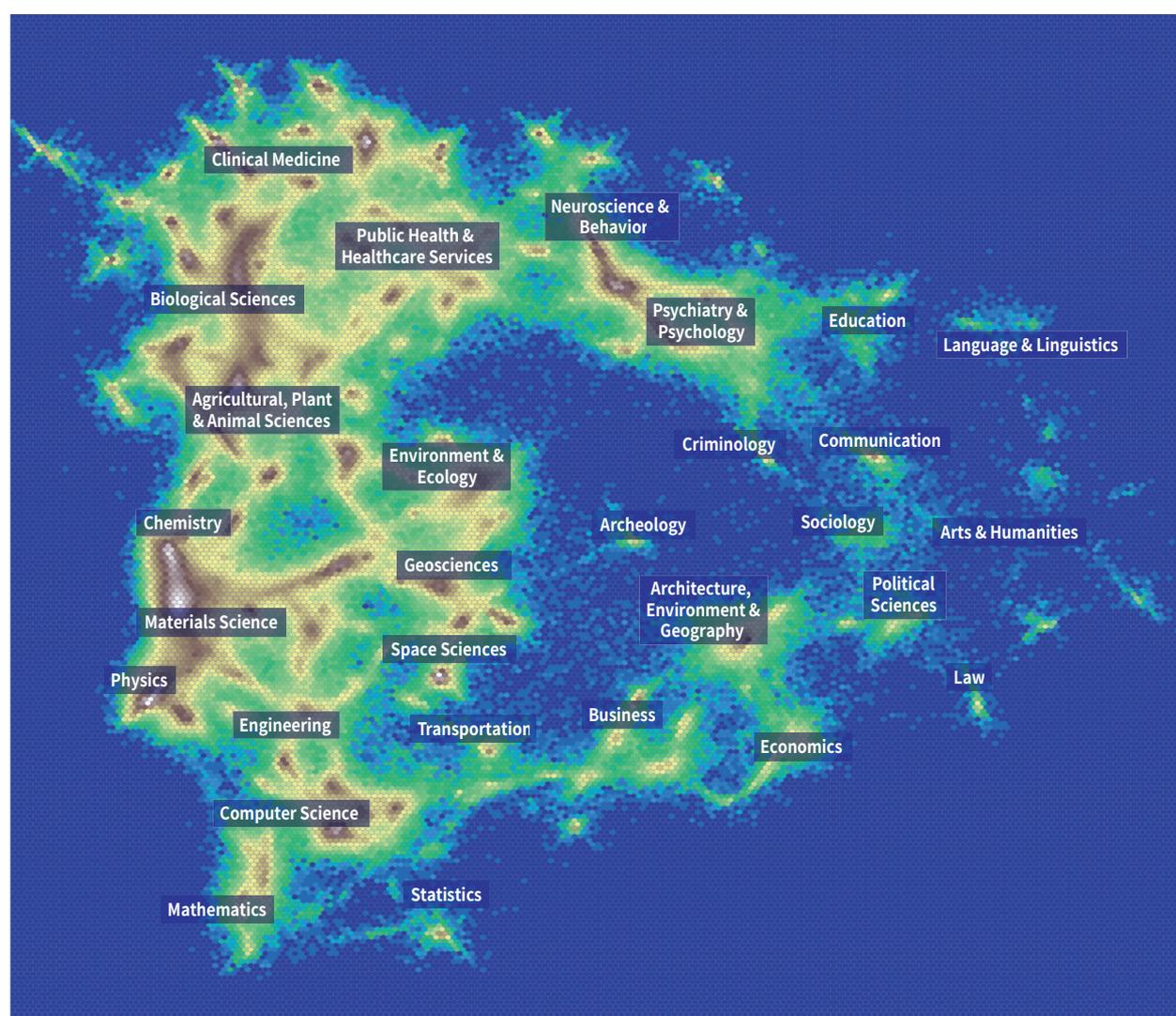
研究前沿是知識的島嶼，四周被相對平靜的海洋環繞

地圖的西北方聚集了許多生物醫學與健康領域的部落，與西邊的核心科學以及科技學門相連。材料科學位於山峰，鄰近化學，但有一條主要支脈伸向工程；運輸是全景圖上介於工程學與商學之間的利基區域；以此類推。

我們可以重複運用研究文獻，仰賴研究文獻來產生有意義的主題地圖，而且可以產生依循直覺解讀的部落領域地形圖。這就表示，當我們建立起能夠反映實際研究活動的框架，在這個框架裡找出關鍵出版品（例如研究前沿），我們就能更瞭解其他分析得出的結果。

圖 4：

所有研究前沿論文的熱力圖 (2014 至 2019 年)，以 ISI 映射框架繪製而成。海拔較高的地區 (依序為黃色、棕色、白色)，對應出版品高度集中的領域。標籤表示地圖上主要學科領域的位置。



研究前沿是如何建立的？

在 Small 和 Griffith 的原始概念中，研究前沿包含 (1) 高被引論文群組，其共被引次數超過相似度優勢門檻，以及 (2) 相關的引用論文。由於研究前沿包括共被引核心論文和引用論文，所以本質究竟為何會受解讀所影響；共被引核心論文可視為基礎或啟發後續研究的突破論文，而引用論文時期比較接近現在，因而具有領先優勢。

- 我們將高被引論文當作地標，來建立研究前沿。共被引分析的基礎為挑選某個領域和年份當中最常被引用的前 1%，因為這些出版品的引用歷史顯示它們是具有影響力的文獻，有可能代表特定專業領域的關鍵概念，或研究前沿。選出 ESI 的近期文獻 (當年度及前五年) 子集合進行分析。

- 共被引論文組透過單一連接叢集互相連結，意思是只要有一條共被引連結，就能將共被引論文組與另一個共被引論文組彼此連結 (例如，若論文 B 和 C 具有共被引關係，則共被引論文 A 和 B 與共被引論文 C 和 D 具有關聯性)。

- 根據論文的共被引相似度，將論文集結成研究前沿。

今天我們使用的方法，可以處理比以前更龐大的研究前沿，並運用更多現代技術，去建立更精準的叢集。我們用萊頓運算法 (Traag et al 2019) 集結論文，因為它提供了可以微調的解析參數 (因此可以建立細緻一些的公式)，還能將更多高被引論文加入研究前沿 (從 43% 提升到 99%)。

有了高被引論文叢集，我們就能為研究前沿建立一組核心論文，附上一系列比較近期和具有領先優勢的共同引用論文。引用論文的標題告訴我們研究前沿的意義，但標記是很主觀的工作，會隨解讀過程而有所改變。我們對核心與共同引用論文的標題和摘要進行文字探勘，以 TextRank 演算法搜尋顯著的術語，為研究前沿訂定標籤。

重複試驗顯示這些程序能一致產生有意義的研究前沿。因此，在調整方面有了革命性的進展，但建立 ESI 研究前沿的基本方式，依然是 Small 和 Griffith (1974) 最初為 ISI 建立的方法。

找出研究前沿的位置

核心與引用論文的全景熱力圖 (圖 4) 建立自 2014 至 2019 年研究前沿論文資料庫，提供了背景參考基礎，可說明個別的研究前沿。立刻呈現出所涵蓋研究領域的知識散播情形，繼續追蹤下去，還能顯示主題在部落之間的移動時互有消長。有鑑於此，我們排除顯示在熱力圖上的島嶼與海洋，並將相同的背景色套用在文獻上，顯示出叢集的明暗度。在通通套用的灰色地圖上，可以為

某一項研究前沿加入標示色彩，看看色塊出現在哪裡。

我們的第一個例子是 CRISPR (clustered regularly interspaced short palindromic repeats) 的研究前沿。CRISPR 是熱門研究文獻和學術期刊裡經常出現的術語，屬於細菌 DNA 序列的一種，成因是被噬菌體感染，而產生 DNA 斷片。Cas9 酵素用這些序列，來辨識並阻斷對

CRISPR 序列具有輔助作用的特定 DNA 股。這是有機體基因編輯技術的基礎，因此相關研究受到極大的關注，而且應用廣泛。

強調 CRISPR 研究前沿的圖譜告訴我們，主要集中區域為農業科學，往上延伸到基礎生物科學，並向外擴展至環境科學。研究前沿中也有些論文落在神經科學和化學，甚至有論文落在遠處的法律學門。

這些非常有用。我們掌握主要的研究主題，是現代生命科學研究裡能幫助我們前進的研究方法，但它不限於單一主要領域，甚至不限於傳統全景圖上的某個連續網絡。經過套色突顯的地圖能讓研究人員掌握領域情況 – 這是以出版品分類資料做階層式分析，所辦不到的事情。

在研究前沿中，除了透過共同引用來分析發展的擴散情形，我們也會想要瞭解，核心論文的位置在哪裡。當共同引用論文參考在其之前截然不同的創新活動，這麼做能為我們提供重要的見解。

為了進一步說明，用此種方式描繪出來的研究前沿分析能立即提供的資訊，我們要再看兩個可能比較陌生的例子：2D 材料和全球能源系統轉型。如圖 6 所顯示。

2D 材料研究前沿穩坐於自然科學的高地：化學、材料科學及物理。有一些比較鬆散的支脈，向工程和電腦科學延伸。農業與植物科學叢集自成一格，向我們顯示該區域有新興的知識連結。

全球能源系統轉型研究前沿有兩個集中的叢集，由零星的點連接在一

起。最大的叢集是建築、環境與地理學，與商學有一點關聯。比較小的叢集位於材料科學和工程之間較不密集的区域。

這是我們特別關心的區域，因為我們在尋找充分吸引注意的研究主題，希望能藉此辨識出研究前沿，但研究前沿的形式和結構還不能發展到可以明確識別該項研究的程度。從化石燃料轉型到可再生能源是活躍的新興領域，已經引起政策制訂者的關注，可能會繼續發展，為日後的研究投資帶來可觀的報酬。

圖 5：
生物醫學 CRISPR 研究前沿論文的全球領域圖譜 (圖 4)

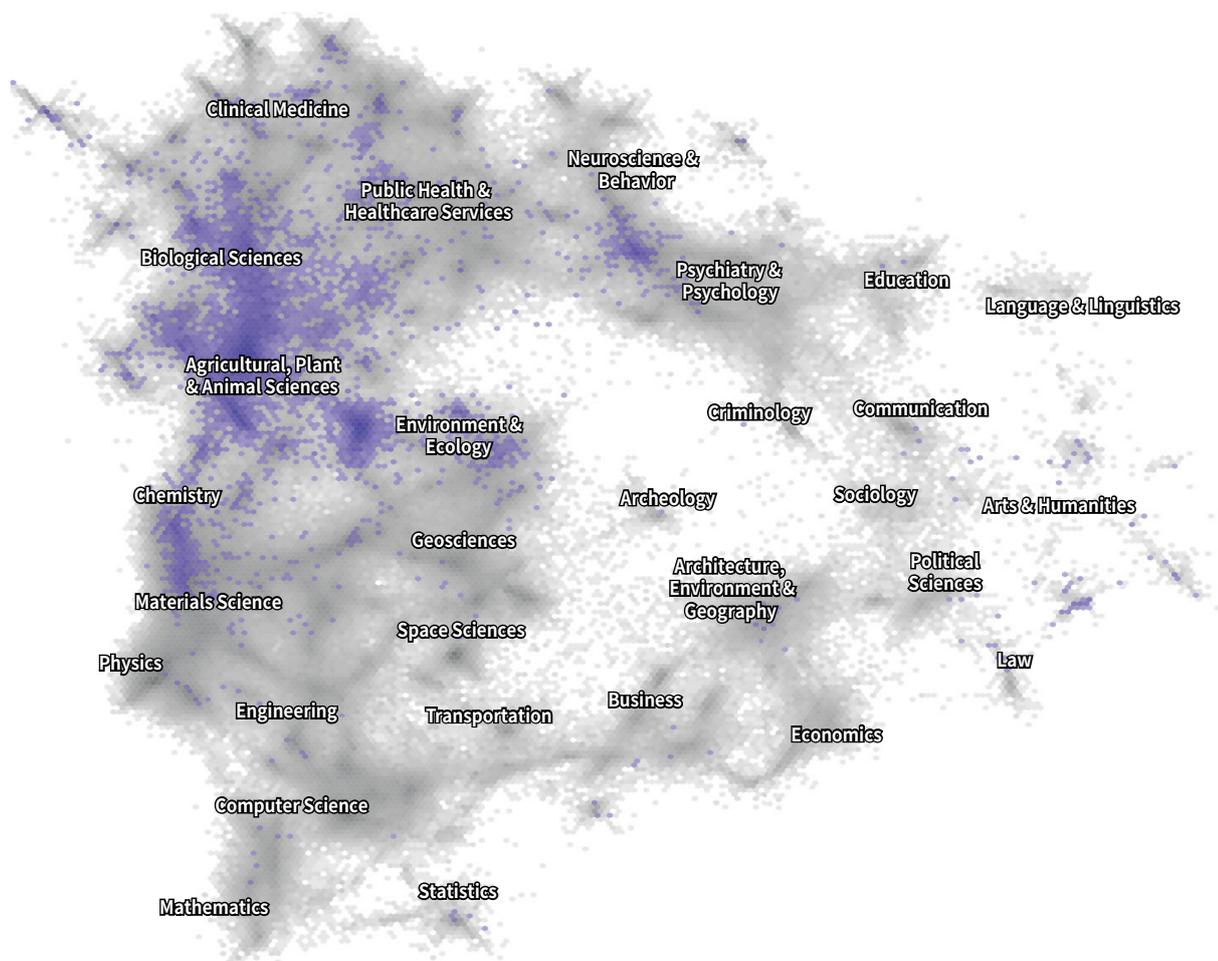
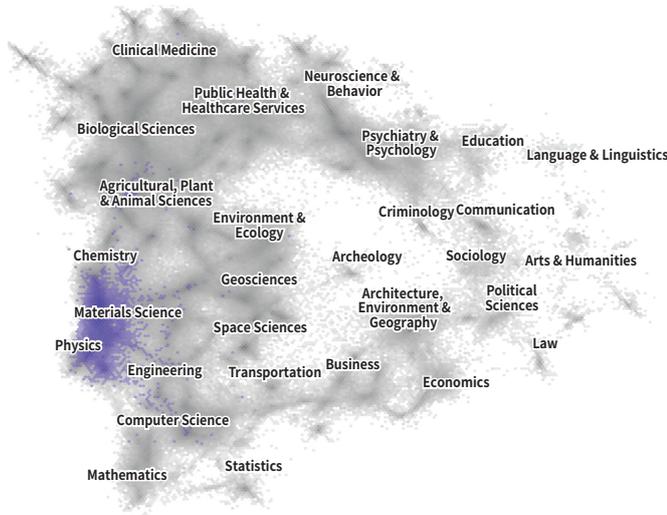


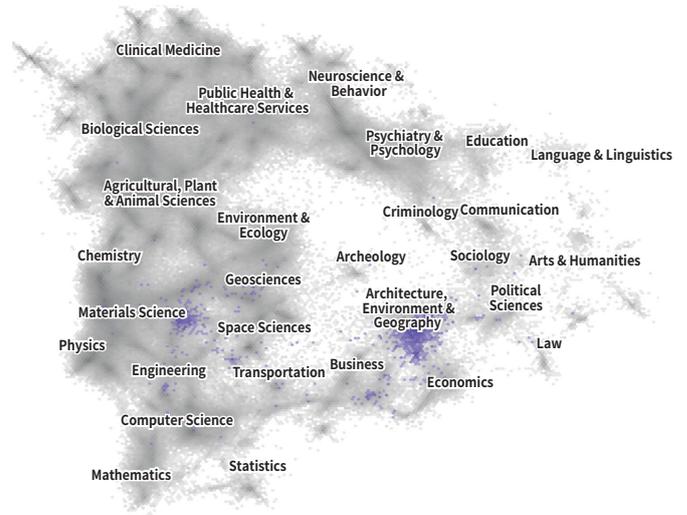
圖 6 :

兩項科技研究前沿論文在全球領域圖譜 (圖 4) 的分布情形

2D 材料



全球能源系統轉型



使用研究前沿地圖

從前述例子可立即看出研究前沿的明顯用途。CRISPR 研究前沿主要確認了該領域中許多人已經知曉的事，但政策制訂者可以藉此確認這項技術非常重要。2D 材料研究前沿顯示出自然科學和生物科學某項領域之間有所連結，大部分的人可能並不清楚，卻能開啟新的契機。全球能源研究前沿告訴我們，持續監督一項新興主題，能為研究贊助人提供重要的投資指引。

這項資訊來自於考量整體研究前沿以及主題位置，或是在全球學科圖譜上的位置。我們可以運用與出版

紀錄有關的其他編輯詮釋資料，加入其他層面的意義，進一步探究和解讀。

幾乎每一篇期刊論文都有作者的地址資訊，讓我們將論文連結至一個或多個組織與國家。表示我們可以用研究前沿分析得出的主題地圖，找出參與研究的組織，也可以突顯單一組織的論文，查看該組織是否與研究前沿有所關聯。

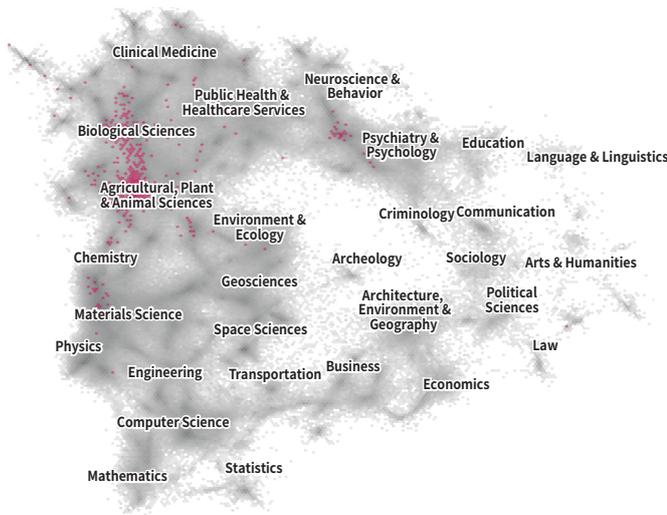
針對 CRISPR 研究前沿，我們擷取出與兩間大型研究密集組織有關的資訊：哈佛大學和中國科學院

(CAS)。在這個例子裡，我們突顯這兩間組織的 CRISPR 研究前沿出版物，包含核心及共同引用論文 (圖 7)。至於研究密集度較低的組織，我們可能會從整體地圖開始，先瞭解他們的研究是否鄰近特定的研究前沿。

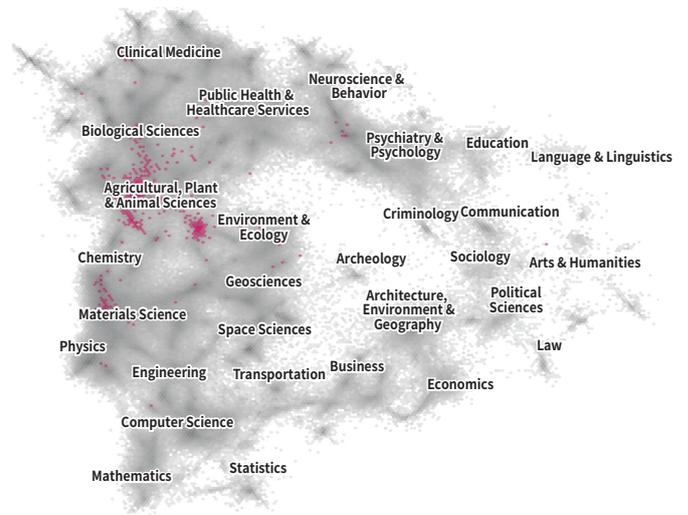
圖 7：

CRISPR 研究前沿論文 (圖 5) 在全球領域圖譜 (圖 4) 中的位置，作者或作者群屬於頂尖的美國與中國研究機構

哈佛大學



中國科學院



圖面顯示哈佛大學的 CRISPR 研究焦點為有機體生命科學，有一條長支脈往上延伸到有機生物科學，外圍還有一群很有意思的叢集，落在神經科學與行為。CAS 地圖也有紮實的有機體生命科學叢集，但該領域的密集度分布情形稍有不同，而且有第二個紮實的叢集，偏向環境與生態學。

這些細微卻很重要的相對分配位置，需要靠專家來詳細解讀，還要嚴格檢視個別出版品的例子。我們能從圖面立刻得知的資訊是，在研究前沿中，不同的組織研究組合，有不一樣的觀點存在。研究團隊經理可能會因此很想瞭解，自己與競爭組織的焦點有何差異。他們有沒有錯

過什麼機會？是否應該尋求合作？

研究贊助機構可能會從研究前沿分析中找出某項價值，讓他們的諮詢機構仔細監督研究地貌，以及優先目標可能會落在哪些區域。他們可能想要找出與現有研究中，有什麼是受到關注的主題，然後評估該項主題是否鄰近研究前沿。他們可以直接問：「我們最近贊助的研究，有多少與研究前沿有關？」Igami 和 Saka (2016) 指出，針對日本研究進行的分析顯示，全國出版活動的差異性正在下降。因此，研究機構可以判斷自己是否有必要針對優先領域的研究前沿進行研究，以及是否有可能投入資金推廣該項研究前沿，因應從社會或政策分析判斷出來的

挑戰。

不需要單獨分開考量個別研究前沿。可以採取不一樣的分析方法，去針對範圍較廣的研究領域，例如整個 ESI 類別，去瞭解所有列為研究前沿的主題。例如，國家贊助機構的地球科學研究，可能會想瞭解所有相關的研究前沿及其動態：規模多大，日期多近，是否橫跨不同領域？還有，當然也想瞭解，有誰參與其中？

下圖 (圖 8) 不在全球領域圖譜中的位置來顯示研究前沿的位置，而是參考相關論文的平均年份，以及那些論文所屬 ESI 領域的多樣化程度，將研究前沿集中在一起。

可以用顏色來表示，研究前沿各有各的主要 ESI 領域：多數色塊指向地球科學。跨領域差異度增加，不同顏色就有可能變多，例如綠色代表環境與生態學，棕色代表工程，黃色代表化學，紫色代表物理學。

也可以為主題設定臨時標籤。標記主題和類別一直都很具挑戰性，因為識別主題有時很受主觀影響，就連專家也不例外。有時候，個人對特定主題的認知會隨著探索過程，在思考主題的內容後而有所改變。此時我們先從一系列論文，擷取標題和關鍵字中最常出現的術語，單純當作標記使用，希望之後使用者

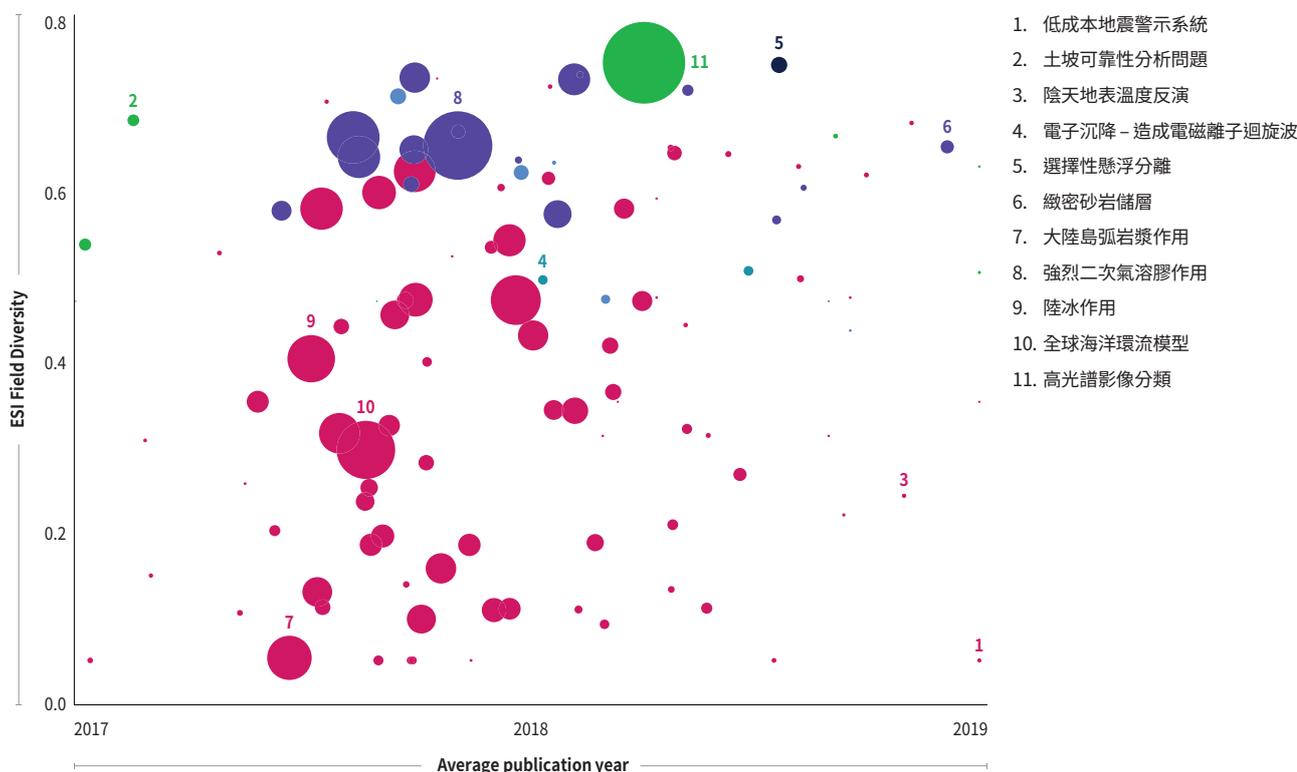
會在取得更多資訊後重新解讀。

這張圖告訴我們地球科學研究前沿的主題、年份、規模、多樣性。同時也顯示出新的資訊，因為我們可以看見，隨著時間發展研究前沿正在茁壯。就像一顆橡樹，從小樹苗開始長大。這對找尋早期跡象，判斷是否可能發展成新興研究領域特別有用。不是所有的主題幼苗都會在長期發展下成為大樹。有些主題會合併或再次併吞，有些則會消失無蹤。但這絕對能提供對管理有用的實際資訊，協助討論未來的投資標的。

雖然研究前沿分析在這些層面上，有可能提供非常豐富的管理支援，但不是只能用於募資計畫。研究前沿分析對學術研究部門也很有用，事實上，也可以幫助個別研究人員規劃職業發展動向。在評估階段，評估者和個別研究人員可以依據圖譜進行討論，提供較單純評分更深入的資訊。圖譜能在評估過程發揮作用，幫助完成形成性評量和總結性評量。表明自己正在參與研究前沿，這麼做本身具有價值，有助研究人員取得職位、繼續發展其職業生涯。

圖 8：

地球科學的研究前沿，根據出版品的平均年份 (x 軸) 以及學科多樣性 (y 軸，ESI 領域多樣性的辛普森指數)。標籤對應研究前沿的名稱與摘要。圓圈大小表示研究前沿的論文數量



證明自己參與這些重要主題，也有益於推廣部門的研究計畫。部門主管可透過出版品的地址資訊，立即掌握部門的研究計畫在全球圖譜的所處位置、在哪裡特別凸出，同時評估與研究前沿凸出主題的相對距離。「我們是否位於新興領域，還是我們獨立於受關注的主題之外？」如此一來，部門可在獲得充分資訊下討論策略方針，也有可能指出未來應該招募的人才，鞏固團隊的實力或補足不齊之能力。

研究人員可以先關注某個特別重要的研究前沿，然後予以拆解，檢視共同引用論文在文本中提及核心論

文的方式(方法、概念、資料?)，並追蹤核心論文及其使用論文的源頭。由這些專業人士來解讀領域的建構和發展，再適合不過了。

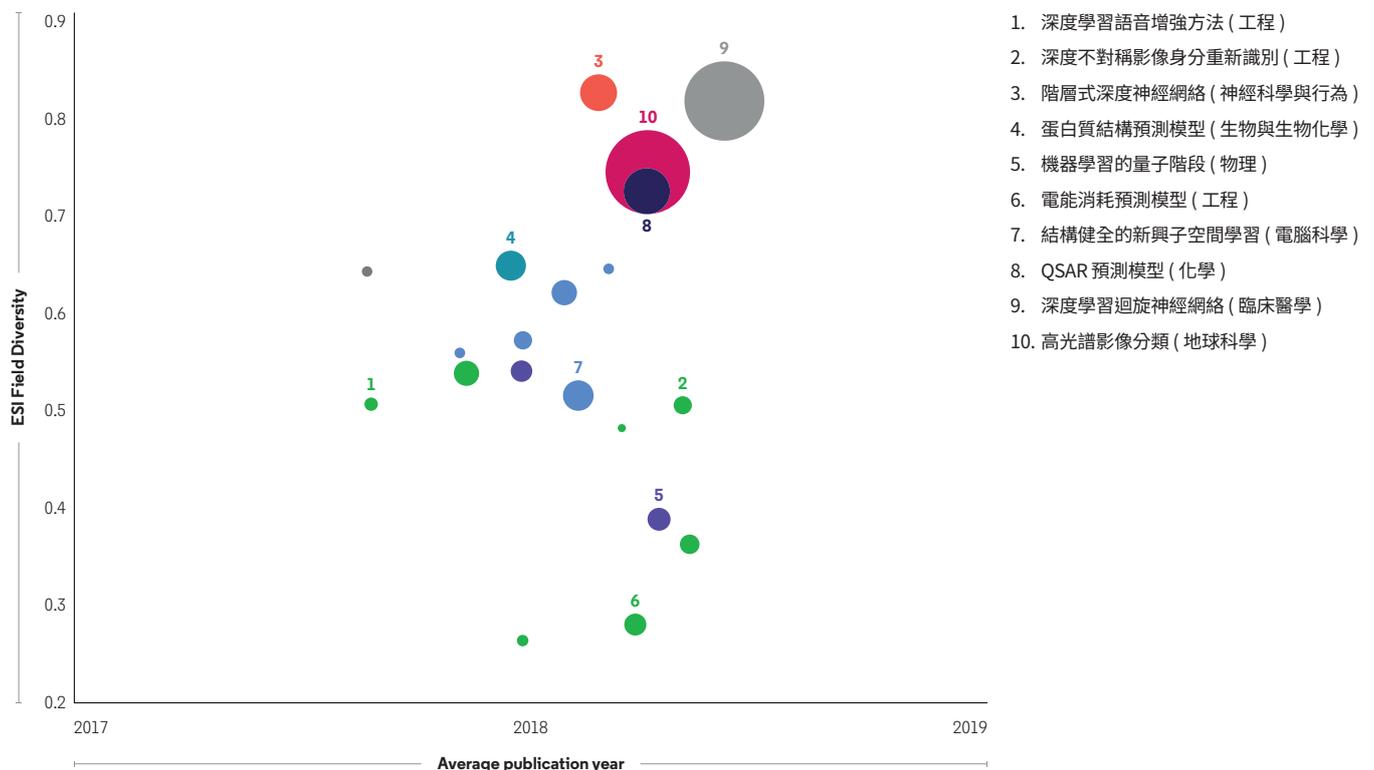
研究人員也可以在地貌上隨處漫遊，從自己目前所在位置開始，接著思考如何邁向新的研究前沿(我的領域接下來可能會發生什麼?創新軌跡是什麼?)，或是進入完全尚未探索的領域。

另外也可以問：我的研究運用在哪裡?例如，機器學習在如今可使用多樣化與多源資料庫而日漸重要，可應用於許多不同的領域。

第一步是找出與「機器學習」這項主題有關的所有研究前沿，可以運用高頻率關鍵字，或與機器學習有關聯的術語詞彙(圖9與10)。下圖說明相關研究前沿的規模與新近程度，依照ESI類別套用的相同顏色編碼顯示應用廣泛。第二張圖則是標示出研究前沿在我們全球領域圖譜上的位置。

圖 9：

機器學習的研究前沿，根據出版品的平均年份(x軸)以及學科多樣性(y軸，ESI領域多樣性的辛普森指數)。標籤對應研究前沿的名稱、摘要，以及重要的ESI領域。圓圈大小表示研究前沿的論文數量

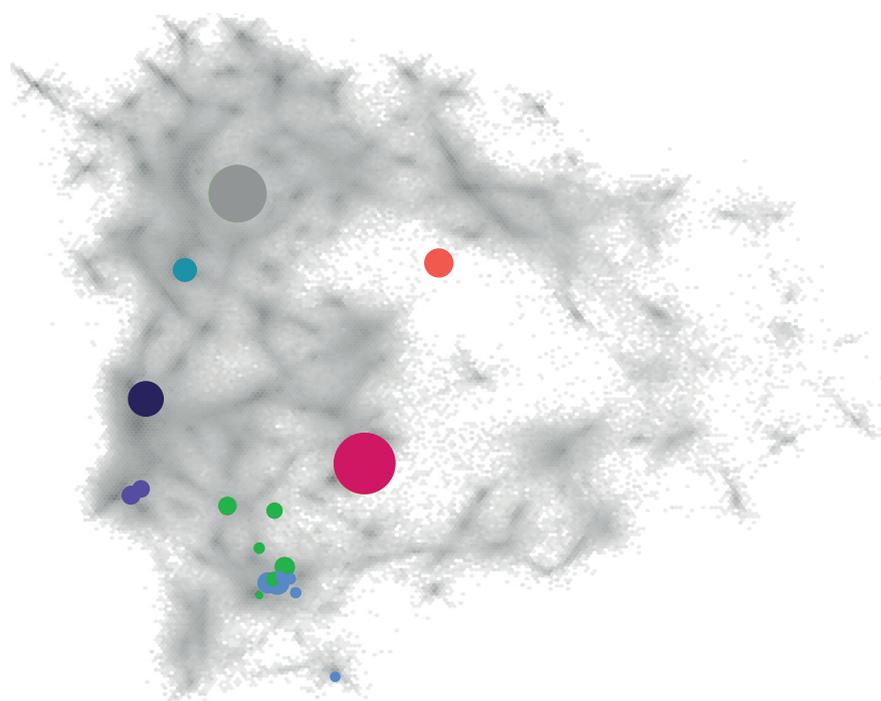


從這裡可清楚看出，機器學習是橫跨各式各樣最新研究前沿的元素。它出現在臨床醫學的大型近期主題叢集，有另外一個以地球科學為焦點的叢集，在工程和電腦科學領域有大量發展，也應用於物理、化學、精神病學與心理學、生物與生物化學。年輕研究人員顯然擁有許多不同的職業發展機會。

年輕研究人員顯然擁有許多不同的職業發展機會。

圖 10：

機器學習研究前沿，置於全球領域圖譜中。雖然研究前沿包含橫跨不同地區的論文，但我們擷取核心與引用論文的平均座標，在同一個位置上總結這些論文。圖片顯示機器學習運用於各種場合，包括臨床醫學、化學、物理學、工程與地球科學。



踏出下一步

ISI 鼓勵分析師參與研究評估與研究政策，不僅將引用網絡視為指標工具，更將其視為不斷發展的框架，瞭解這能反映出研究的變遷歷程。透過研究前沿的映射與分析，我們證明可以找出符合主題的跨學科研究

領域，並追蹤領域在研究生態系裡發展、成熟的進程。科睿唯安專業服務團隊持續精進，確保研究前沿能為學術界、業界與公部門客戶提供客製化的研究計畫，幫助客戶深入瞭解自身的研究組合定位、與同儕

之間的相對表現，並且提供投資與策略規劃的相關知識。

www.webofsciencegroup.com/isi

References

- Becher, T. and Trowler, P.R. (2001). *Academic Tribes and Territories* (second edition, pp 238). Open University Press, Milton Keynes UK. ISBN: 978-0335206278
- Börner, K. (2010). *Atlas of Science – Visualizing What We Know*. MIT Press, Cambridge MA. ISBN: 978-0262014458
- Boyack, K.W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79, 27-44. DOI: 10.1007/s11192-009-0402-6
- Boyack, K.W. and Klavans, R. (2010). Co-citation analysis, bibliographic coupling and direct citation: which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61, 2389-2404. DOI: 10.1002/asi.21419
- Boyack, K.W. and Klavans, R. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge. *Journal of the Association of Information Science and Technology*, 68, 984-998. DOI: 10.1002/asi.23734
- Boyack, K.W. and Klavans, R. (2019). Creation and analysis of large-scale bibliometric networks. In *Springer Handbook of Science and Technology Indicators*, W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (eds.), Springer, 187-212. ISBN: 978-3030025106
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57, 359-377. DOI: 10.1002/asi.20317
- Chen, C. (2013). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, second edition, Springer. ISBN: 978-1447151272
- De Bellis, N. (2009). Maps and paradigms: Bibliographic citations at the service of the history and sociology of science. In *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*, Scarecrow Press, 143-179. ISBN: 978-0810867130
- Garfield, E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122, 108-111. DOI: 10.1126/science.122.3159.108
- Glänzel, W. and Thijs, B. (2012). Using ‘core documents’ for detecting and labelling new emerging topics. *Scientometrics*, 91, 399-416. DOI: 10.1007/s11192-011-0591-7
- Griffith, B.C., Small, H.G., Stonehill, J.A. and Dey, S. (1974). Structure of scientific literatures. II: toward a macrostructure and microstructure for science. *Science Studies*, 4, 339-365. DOI: 10.1177/030631277400400402
- Grover, A. and Leskovec, J. (2016). Node2vec. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD’ 16: www.doi.org/10.1145/2939672.2939754
- Igami, M. and Saka, A. (2016). Decreasing diversity in Japanese science, evidence from in-depth analyses of science maps. *Scientometrics*, 106, 383-403. DOI: 10.1007/s11192-015-1648-9
- Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14, 10-25. DOI: 10.1002/asi.5090140103
- Marshakova-Shaikevich, I.V. (1973). System of document connections based on references. *Nauchno Tekhnicheskaya, Informatsiya Seriya 2, SSR*, [Scientific and Technical Information Serial of VINITI], 6, 3-8.
- McInnes, L. and Healy, J. (2018) UMAP: *Uniform Manifold Approximation and Projection for Dimension Reduction*, ArXiv e-prints 1802.03426, 2018
- Noyons, E.C.M., Moed, H.F. and Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50, 115-131. DOI: 10.1002/(SICI)1097-4571(1999)50:2<115::AID-ASIS3>3.3.CO;2-A
- Noyons, E.C.M. (2004). Science maps within a science policy context. In *Handbook of Quantitative Science and Technology Research*, H.F. Moed, W. Glänzel, U. Schmoch (eds.), Springer, 187-213. ISBN: 978-1402027024
- Pendlebury, D.A. (2013). *Research Fronts: In search of the structure of science*. In *Research Fronts 2013: 100 Top-Ranked Specialties in the Sciences and Social Sciences*, C. King and D.A. Pendlebury (eds.), Thomson Reuters, 26-31.
- Petrovich, E. (2020). Science mapping and science maps. *ISKO Encyclopedia of Knowledge Organization*, B. Hjørland and C. Gnanli (eds.) www.isko.org/cyclo/science_mapping
- Price, D de S. (1965). Networks of scientific papers. *Science*, 149, 510-515. DOI: 10.1126/science.149.3683.510
- Price, D de S. (1979). Foreword. In E. Garfield, *Essays of an Information Scientist, 3, 1977-1978*, Institute for Scientific Information, v-ix.
- Price, D de S. (1986). *Little Science, Big Science... and Beyond*. (reprint edition of 1963 *Little Science, Big Science*). Columbia University Press. ISBN: 978-0231049566
- Small, H. (1973). Co-Citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265-269. DOI: 10.1002/asi.4630240406
- Small, H. (1997). Update on science mapping: creating large document spaces. *Scientometrics*, 38, 275-293. DOI: 10.1007/BF02457414
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50, 799-813. DOI: 10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASIS9>3.0.CO;2-G
- Small, H. (1999). A passage through science: crossing disciplinary boundaries. *Library Trends*, 48, 72-108.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68, 595-610. DOI: DOI: 10.1007/s11192-006-0132-y
- Small, H. and Garfield, E. (1985). The geography of science: Disciplinary and national mappings, *Journal of Information Science*, 11, 147-159. DOI: 10.1177/016555158501100402
- Small, H. and Griffith, B.C. (1974). Structure of scientific literatures. I: identifying and graphing specialties. *Science Studies*, 4, 17-40. DOI: 10.1177/030631277400400102
- Small, H. and Sweeney, E. (1985). Clustering the Science Citation Index using co-citations. I. A comparisons of methods. *Scientometrics*, 7, 391-409. DOI: 10.1007/BF02017157
- Thijs, B. (2019). Science mapping and the identification of topics: Theoretical and methodological considerations. In *Springer Handbook of Science and Technology Indicators*, W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (eds.), Springer, 213-233. ISBN: 978-3030025106
- Traag, V. A., Waltman, L. and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1). DOI DOI: 10.1038/s41598-019-41695-z
- Van Eck, N.J. and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523-538. DOI: 10.1007/s11192-009-0146-3
- Waltman, L., Van Eck, N.J. and Noyons, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4, 629-635. DOI: 10.1016/j.joi.2010.07.002

關於 ISI 全球研究報告系列

ISI《全球研究報告》(Global Research Reports) 利用我們獨特的行業見解，提供分析、想法和評論，以啟發更多的深入討論。

每一份報告展現了科研資料擁有的巨大潛力，可以為科研評估和科研政策中的管理問題提供資訊，並加速全球科研體系的發展。

關於科睿唯安

科睿唯安 (Clarivate™) 是全球領先的專業資訊服務提供者，我們致力於提供值得信賴的資料與卓越的洞見，幫助客戶加速創新步伐。我們的專業知識和解決方案涵蓋創新的各個環節 — 從基礎研究，到保護創新，直至實現創新成果的商業化。今天，科睿唯安正在銳意進取、推陳出新，協助客戶用創新賦能美好生活。我們擁有全球眾多備受信賴的品牌，包括

Web of Science™、Cortellis™、Derwent™、CompuMark™、MarkMonitor™ 和 Techstreet™。更多資訊請參考 clarivate.com。

科睿唯安旗下的 Web of Science™ 集結全球科學研究資訊，以全球最大、出版社中立的引文索引及研究情報平台為基礎，協助學術機構、企業、期刊出版社、政府單位加速科學研究步伐。

webofsciencigroup.com/isi