

Title	OpenAlexを用いた国際共著実態可視化システムの構築と分析例
Author(s)	小柴, 等; 佐々木, 達郎; 北島, 謙生
Citation	年次学術大会講演要旨集, 39: 664-669
Issue Date	2024-10-26
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19441
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

OpenAlex を用いた国際共著実態可視化システムの構築と分析例

小柴 等 (NISTEP), ○ 佐々木達郎 (NISTEP), 北島謙生 (NISTEP)

1 はじめに

情報処理技術の進展や、国際環境の変化に伴い、研究者の国境を越えた共同研究の実態は大きく変化している。こうした実態を定量的に把握するため、論文の共著関係を用いた各種の分析や、そのための可視化手法も提案され、有益な知見をもたらしている（例えば、文献 [Borner2005]）。一方、これらの分析は政策実務の観点では、継続的な報告が行われていなかったり、分野や分析期間が必ずしも実務のニーズとマッチしなかったり、という課題もある。たとえば研究という面から考えると、手法を開発し、その新規性・有用性を示すための分析を行うまでがひとつの単位で、業績にも数えやすい。しかしその後、その手法を用いて年度単位などでクロールングするだけでは研究上の新規性や有用性が乏しくなりがちであり、成果に繋がりにくい。また、分野や分析期間についても、紙面の関係上、ある程度限定せざるを得ない。したがって、継続的に調査するインセンティブに繋がりにくく、上記の課題は構造的に解決しがたい。

今回、我々は継続的な分析と、より手軽かつ柔軟な仕組みの構築を念頭に、既存手法の一つに着目し、オープンな論文書誌データベース“OpenAlex”[openalex2022]^{*1}のデータを用いた Web サービスを構築した。本稿では、このサービス（システム）と、当該システムを用いて分析した国際共著の分野別動向例について紹介する。

2 方法論

2.1 分析・可視化手法

分析・可視化手法は Okamura[okamura2023] の提案手法を用いる。この手法では、まず共著関係をベースに、国・地域、あるいは機関の間の距離を Jaccard 係数ベースで算出する。その上で、この距離に基づきクラスタリングし、可視化する。参考文献では階層クラスタリング（Ward 法）を用い、結果をデンドログラムとして可視化する方法を提案している。これを、任意の分野、期間で実行することで国際関係の変化を視覚的にもわかりやす

く可視化できている。メリットとしては、複数の国・地域のクラスタを観察できること、ハードクラスタリングであるので所属クラスタに対する解釈の曖昧性が無いこと、などが挙げられる。

参考文献ではこの手法を国・地域よりも更に粒度の細かい機関単位にも適用し、機関とそれに紐づく国・地域レベルでの国際共著分析も実現している。これらの特徴に加えて、データソースに OpenAlex を用いていることから、再現性などを確認しやすいという面もあり、本手法を採用することとした。

2.2 データ

分析・可視化対象のデータも、文献 [okamura2023] と同じく OpenAlex のデータを用いた。OpenAlex は 2019 年公開の比較的新しい論文書誌データベースで、その名の通りオープンに利用できる。そのため、ペイウォールの問題がなく、第三者による検証可能性・再現性が高く、システムの公開運用や安定的な更新が容易である。

商用の論文書誌データベースである Web of Science (WoS) や Scopus と比較した場合、OpenAlex は会議録 (Proceedings) やプレプリント、書籍のデータも含んでおり、2022 年時点で、WoS, Scopus より 1 桁多く、約 2.4 億レコードを有している。また、2015 年から 2022 年のレコードに限って DOI ベースで WoS, Scopus との重複率を見たところでは 7 割程度となっており [Culbert2024]、十分なデータ量を有している。データソースは Microsoft Academic Graph (MAG), Crossref を中心に、ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, The ISSN International Centre, Internet Archive, arXiv や Zenodo 他の分野別レポジトリ、Web クローリング結果、に基づく^{*2}とされる。

分析において最も重要となる国・地域の情報は、著者の所属機関データに合わせて整備されている。具体的には、著者ごとに割り振られてた機関コードには機関の住所として国・地域が紐付いている。これらのデータを用いて関係性の分析を行う。

^{*1} <https://openalex.org/>

^{*2} <https://help.openalex.org/hc/en-us/articles/24397285563671-About-the-data> (2024.09.12 Last accessed.)

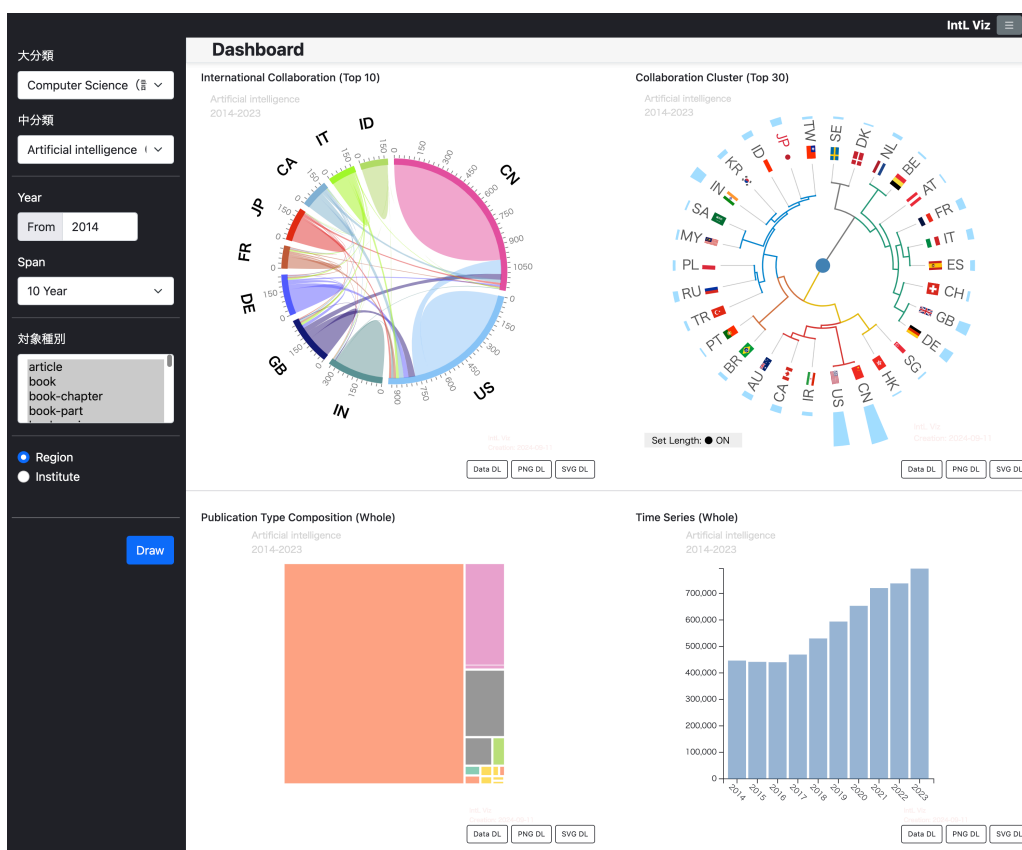


図 1: システムの概観（国別比較，AI 分野，2014 – 2023 年）

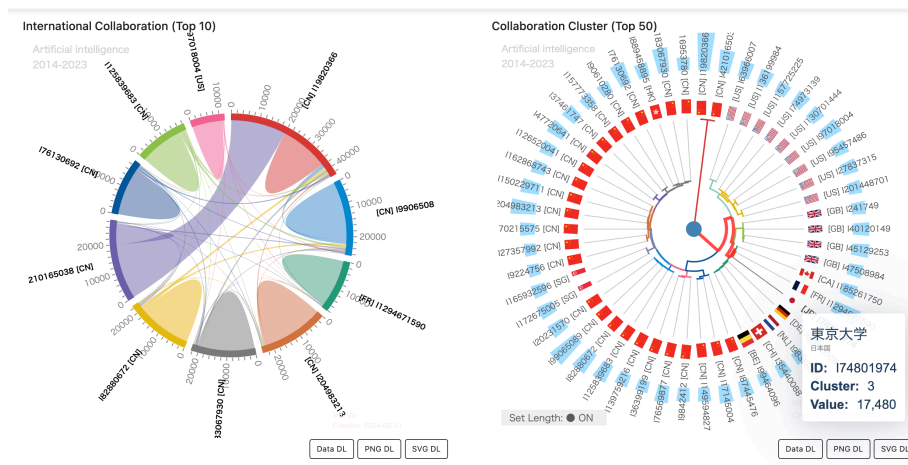


図 2: 機関別比較の例（AI 分野，2014 – 2023 年）

分野については，WoS などではジャーナル単位での割り振りが基本であるのに対し，OpenAlex では機械学習を用いて個々のレコード単位で割り振っている。また，分野 (Concepts) は 6 階層，約 6.5 万件の分野から複数割り振られる^{*3}。

データの利用方法には，API を通じて都度取得する方

法と，スナップショットのバルクデータをダウンロードして用いる方法がある。今回の用途では個別の論文の著者所属情報をそれぞれ収集・解析する必要があることから，API 利用は実質的に不可能であり，ダウンロードの方式を採る。現状の実態ベースでは，おおよそ 1 ヶ月単位でスナップショットが更新されている^{*4}。システム側

^{*3} Concepts は今後 Topics という新区分に変更される予定となっており，現在は移行期間として両方が設定されている。

^{*4} https://openalex.s3.amazonaws.com/RELEASE_NOTES.txt (2024.09.05 Last accessed.)

では処理の都合上、概ね 2-3 ヶ月単位でデータの差し替えを行っている。

2.3 システム

構築したシステムは試験的に公開している^{*5}。

分野については 6 階層約 6 万種類では選択が困難であること、最高位の 19 分類ではやや粗すぎることから、上から 2 階層目の分野単位で分析できるようにしている。

分析期間については 2024 年 8 月時点において 1970 年以降、2023 年まで^{*6}、1,3,5,10 年の粒度で分析できる。

また、国単位か機関単位かを切り替えて分析できる。論文、書籍等の出版種別についても選択できる。

画面のイメージを図 1, 2 に示した。

図 1 に示す 4 つのグラフは左上から時計回りに、成果数上位 10 カ国（地域を含む）間の共著関係、上位 30 カ国の共著関係に基づくクラスタ、当該分野・期間における年ごとの成果数推移、当該分野・期間全体における成果種別（論文、書籍など）、を示している。

機関単位での分析の場合は、共著関係とクラスタの部分が、図 2 のとおり機関単位となる。機関のグラフに表示されている I00000 などの文字列は OpenAlex 上の機関コードに対応しており、マウスオーバーにより機関名が確認できる。機関国籍については国旗により簡易にも識別できる。

この他、各グラフの右下には「Data DL」「PNG DL」「SVG DL」のボタンがあり、数値データ、または画像データをダウンロードできる。

3 分析事例

3.1 AI 分野

計算機科学 (Computer Science) の下位分野にある AI 分野について、1994 年から 2023 年まで 10 年単位で 3 期間取得した結果について示す。

3.1.1 国・地域別

国別での分析を図 3 から 5 に示す。

例えば図 3 に示した 1994 年からの 10 年の期間を見ると香港 (HK) を含む中国のクラスタが独立している。また、米英加の 3 国のクラスタ、日本と韓国のクラスタ等があり、これらはさらに近い位置で結合するなど、米英

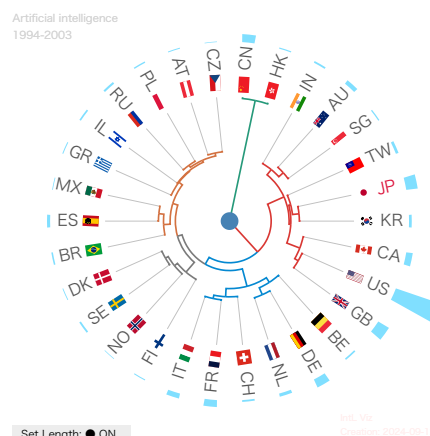


図 3: AI 分野（国・地域別 1994 – 2003 年）

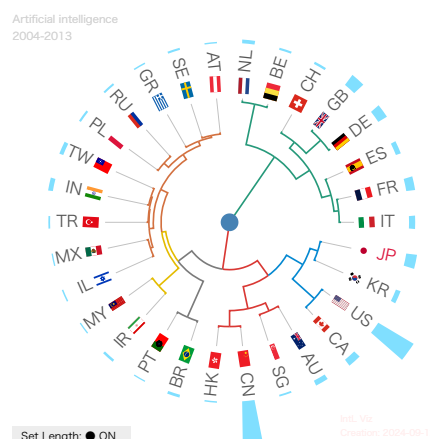


図 4: AI 分野（国・地域別 2004 – 2013 年）

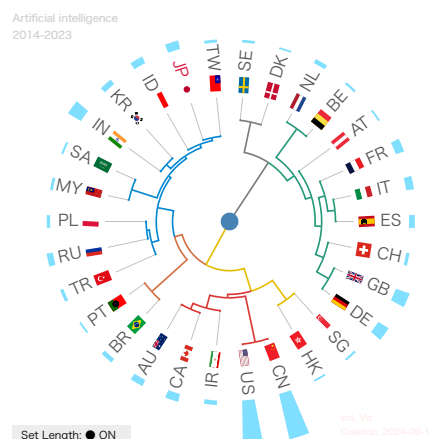


図 5: AI 分野（国・地域別 2014 – 2023 年）

^{*5} <https://devgru.nistep.go.jp/intl/>（ライセンス等はサイトで確認のこと。）OpenAlex では現在も活発に過去データを含めて適時修正が行われており、本システムでも定期的にそれらのデータを取り込んで更新しているため、そのデータの違いに起因して違いが出る可能性があるものの、後述する事例は基本的にすべて、本システム上で再現できる。

^{*6} スナップショットには 2024 年分データも含まれるが、年単位での分析のため切り捨てている。

と日本がやや近い場所に位置している。青のバーで示した成果数では米国が突出しているが、日本も比較的多い。

図 4 の 10 年では、英国が欧米クラスターに移動し、米国とカナダの北米グループが結束を強めている。日本と

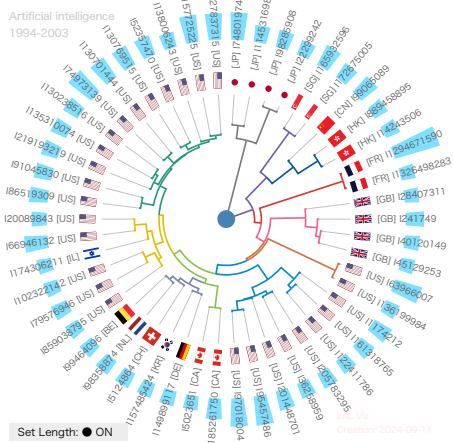


図 6: AI 分野 (機関別 1994 - 2003 年)

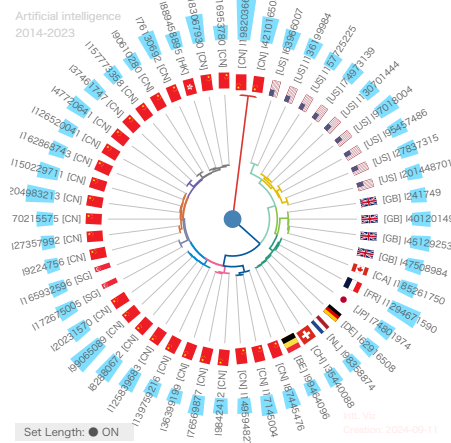


図 8: AI 分野 (機関別 2014 - 2023 年)

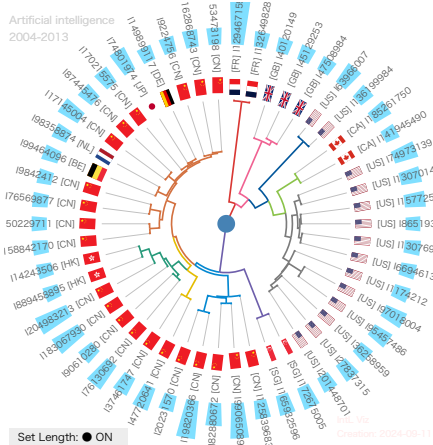


図 7: AI 分野 (機関別 2004 - 2013 年)

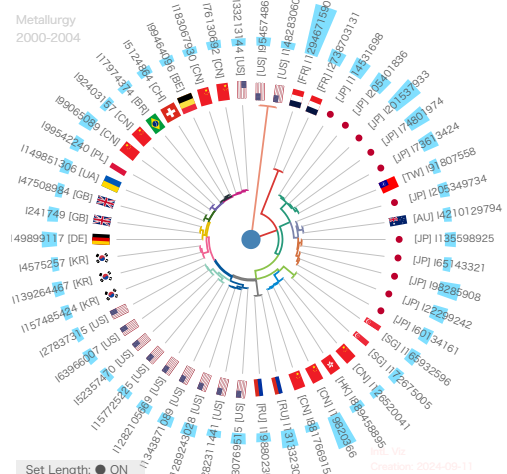


図 9: 金属工学分野 (機関別 2000 - 2004 年)

韓国もこれらのクラスターの近くに位置している。

図 5 の直近 10 年では、様相が大きく変わり、米国と中国が接近している。日本は中国を除く「その他アジア」クラスターに属している他、ヨーロッパクラスターも観察でき、米中を除くと地理的に近い国・地域で固まっているように読める。

3.1.2 機関別

機関別での分析を図 6 から 8 に示す。

例えば図 6 に示した 1994 年からの 10 年の期間は上位を米国が占めており、かつ米国内にも 2 つの勢力があること。また、日本も複数機関でコミュニティを築いていることがわかる。

図 7 の 10 年では、中国が一気に台頭し半数近くを占めている。日本は 1 機関のみでドイツや中国とクラスターを形成している。

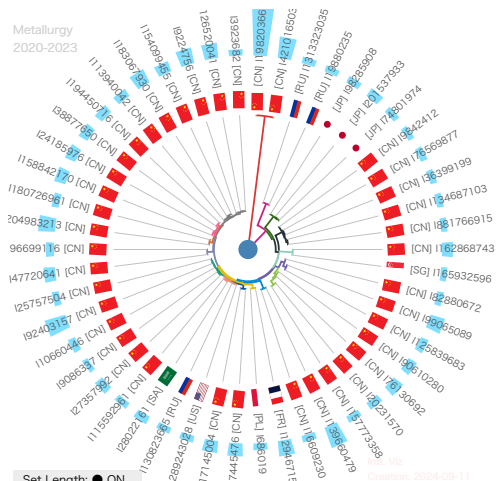
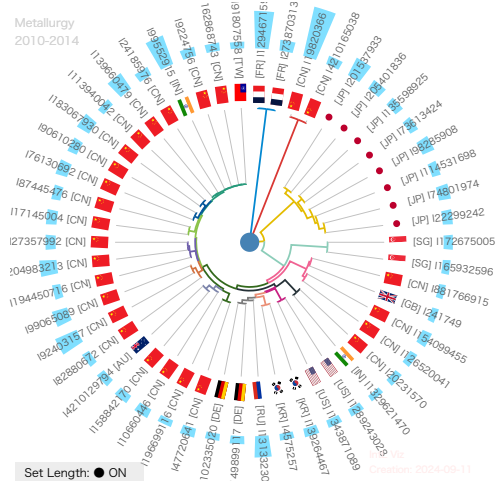
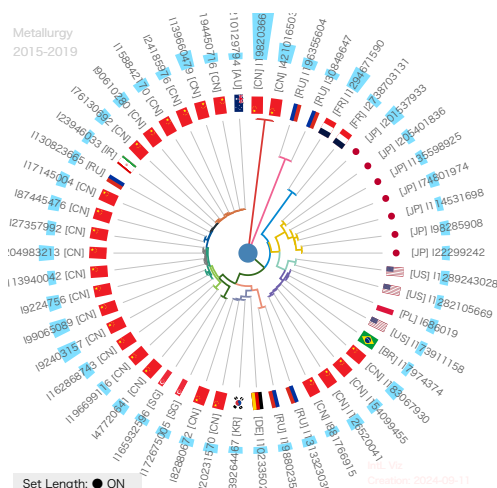
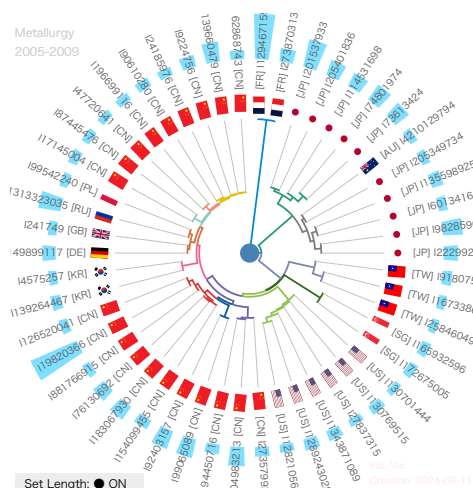
図 8 の直近 10 年では、中国が単独で過半数を超え、日本からは引き続き 1 機関のみが登場している。クラスターの構成としては、中国クラスターの中にシンガポ-

ールの機関が割り込んでいる点が興味深い。また、これら中華圏クラスターと欧米を基本とするそれ以外が綺麗に分かれており、欧米系の非中華圏も米国クラスター、英国クラスター、その他に 3 分されている。日本は「その他」クラスターに属し、唯一のアジア勢となっている。

実際に AI 関連のトップカンファレンスでも中国からの発表が首位になっており、既存の動向と一致する。さらに、共著関連では米英中がそれぞれ国内機関の連携を強める中、フランスやドイツ、カナダ、日本などは複数機関でアライアンスを組んでいる様子が見える点は興味深い。

3.2 金属工学分野

材料科学は伝統的に日本が強みを持つと言われてきた研究分野のひとつである [CRDS2023]。ここでは材料工学 (Materials Science) の下位分野である金属工学 (Metallurgy) について、機関に着目し、2000 年から 2023



年まで5年単位^{*7}で5期間を可視化した。結果を図9から13に示す。

2000 年からの 5 年を示した図 9 をみると、米国からは 12 機関がランクインしており最大勢力、続いて日本が 11 機関となっている。また、米国は 9 機関からなるクラスタが形成されており、内部凝集性が高い。日本のクラスタは基本的に 2 つに別れており、さらに台湾、オーストラリアとのクラスターも存在する。

続く 2005 年からの 5 年を示した図 10 をみると、米国に変わり中国が急激に場を席巻している。また、台湾もわずかながら機関数を増やしている。日本は 10 機関がランクインしており、勢力を維持している。

2010 年からの図 11 では米国が更に数を減らす。

2015 年からの図 12 では、日本も 7 機関までに減少し

ている。興味深いことにロシアから4機関がランクインする。

2020 年からの図 12 では、中国が上位 50 機関中 39 機関と大勢を占めている。次点は日本とロシアの各 3 機関で、米国は 1 機関のみとなっている。

3.3 計量経済学分野

ここまでは、中国の急激な台頭の目立つ分野を見てきた。ここでは経済学 (Economics) の下位分野にある計量経済学 (Econometrics) 分野について、機関に着目し、1994 年から 2023 年まで 10 年単位で 3 期間取得した結果について、図 14 から 16 に示す。

これまで紹介した AI、金属工学分野と比較して、計量経済学分野では、1994 年からの図 14 と、2014 年からの図 16 の間での変化が少ない。もちろん、図 16 では中国が出現はしているものの、まだ 7 機関程度で限定的である。他方で、この分野においては米国の存在感はほぼ変

*7 この期間からも読み取れるとおり、最後のみ 2020 年から 2023 年の 4 年。

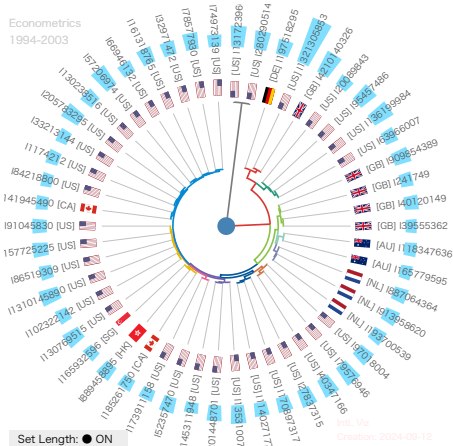


図 14: 計量経済学分野（機関別 1994 – 2003 年）

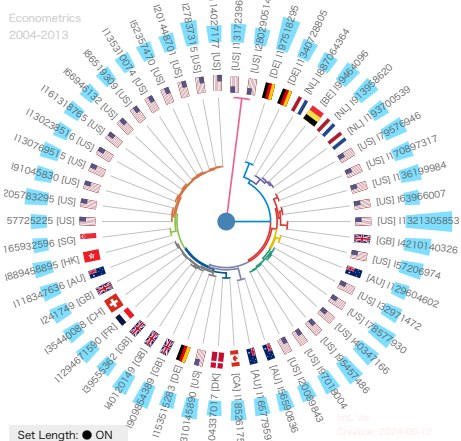


図 15: 計量経済学分野（機関別 2004 – 2013 年）

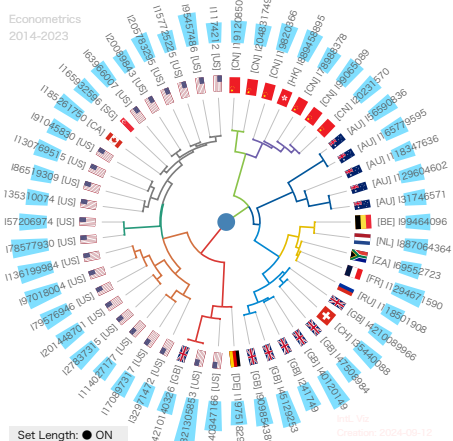


図 16: 計量経済学分野（機関別 2014 – 2023 年）

化しておらず大勢を占めている。興味深い点としては、南アフリカ（国コード：ZA）が登場していることが挙げられる。

4 おわりに

本稿ではオープンな論文書誌 DB のデータを元に、国際・機関共著関連の状況を分析・可視化する既存手法を実装したシステムと、それを用いたごく簡単な分析事例を紹介した。

事例でも示したとおり、本システムを用いることで、興味のある分野・機関についての国際・機関連携の実態を手軽に確認できる。これらと、社会・政治環境、研究トレンドを合わせて見ていくことで、研究における国・地域の関係性を的確に理解できると期待される。

情報技術の発展・普及に伴って、情報の地理的格差は低減している。さらにオープンアクセスやオープンデータをはじめとするオープンサイエンスの動向とも関連して、科学研究における国境は比較的低くなっていると考えられる。これにより、研究上の流行り廃りや、社会的・政治的な状況の変化などにより連携状況が変化しやすくなっているとも言える。状況の理解はもちろん、制御するには計測が必須要素であり、こうしたシステムを利用して適時、適切に状況を把握していくことが必要である。発表では、本システムを用いた分析の事例を中心に、国際・機関連携の状況について紹介する。

謝辞

本研究の実施に当たって、分析手法の提案者である岡村圭祐氏より、各種のご助言を頂いた。記して感謝する。

参考文献

- [Borner2005] Katy Börner and Luca Dall'Asta and Weimao Ke and Alessandro Vespignani : Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams . *arXiv*, (preprint) 2005. <https://doi.org/10.48550/arXiv.cond-mat/0502147>
- [openalex2022] Jason Priem, Heather Piwowar and Richard Orr : OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts . *arXiv*, (preprint) 2023. <https://doi.org/10.48550/arXiv.2205.01833>
- [Culbert2024] Jack Culbert, Anne Hobert, Najko Jahn, Nick Haupka, Marion Schmidt, Paul Donner, Philipp Mayr : Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus . *arXiv*, (preprint) 2024. <https://doi.org/10.48550/arXiv.2401.16359>
- [okamura2023] Keisuke Okamura : A half-century of global collaboration in science and the ‘Shrinking World’ . *Quantitative Science Studies*, Vol.4, No.4, pp.938–959. 2023. https://doi.org/10.1162/qss_a_00268
- [okamura2024] Keisuke Okamura : Atlas of Science Collaboration, 1971–2020. *SN Computer Science*, Vol.5, No.640. 2024. <https://doi.org/10.1007/s42979-024-02973-4>
- [CRDS2023] JST CRDS : ナノテクノロジー・材料分野（2023 年）. 研究開発の俯瞰報告書. CRDS-FY2022-FR-05. 2023. <https://www.jst.go.jp/crds/report/CRDS-FY2022-FR-05.html>