

レポート

米国における研究動向の調査研究

－ NSF を事例とした共起ネットワーク分析から見る研究動向－

第2調査研究グループ 上席研究官 荒木 寛幸

概要

米国における研究動向を調査するため米国国立科学財団（NSF）のファンディング情報から研究動向の調査を行った。ARAKIシステムを用いたデータ分析のもと共起ネットワーク分析を行いサブグラフ検出（modularity）、中心性（媒介）、共起パターンの変化（相関）の3つの結果を考察したところ、NSFにおける2019～2020年の採択課題は環境やAIに関するキーワードの研究が多いことが分かった。なかでも「レジリエンス」「インフラストラクチャー」などのキーワードが中心性が高く、研究のハブ・キーワードとなっていた。2020年は2019年に比較してコロナウイルスに関するキーワードの研究課題が増加していた。

キーワード：科学技術動向，EBPM，ARAKIシステム，政策立案，データサイエンス

1. はじめに

2021年3月、2021年度から5年間を計画期間とする第6期科学技術・イノベーション基本計画（以下「第6期基本計画」という）が閣議決定された。第6期基本計画では、気候変動問題をはじめとするグローバルな課題の克服への貢献等を目的としており、国際社会における我が国のプレゼンスを高めていくことを目指す内容となっている¹⁾。

科学技術・学術政策研究所（NISTEP）では、国内外の研究動向を把握するための調査を行っており、近年ではEBPM（エビデンス・ベースト・ポリシー・メイキング。証拠に基づく政策立案）を推進するためビッグデータを活用したデータサイエンスによる調査研究にも取り組んでいる。第2調査研究グループでは、科学研究費補助金の採択データや学会発表の抄録等情報を用いた国内における先端研究動向調査として、研究助成プログラムの分析と可視化を目的としたAdvanced Research Analysis in Keen-keyword Investigationシステム（以下「ARAKIシ

ステム」という）を開発し、研究動向など状況把握を行ってきた²⁾。

しかし、上述した第6期基本計画にあるグローバルな課題の克服への貢献等を進めるためには、国内の研究動向ばかりではなく海外の状況を把握する必要がある。そこで今回、海外における研究動向調査として米国国立科学財団（NSF）におけるファンディング情報から研究動向の調査を行うこととした。

2. 調査方法

今回はARAKIシステムを用いてNSFの採択データから共起ネットワーク図を作成し研究動向を調査することとした。

採択データは、NSFのウェブサイト^注からダウンロードした。それをデータクリーニングし、さらに、加工した後ARAKIシステムに組み込んだ。今回は試行のため、2019年及び2020年に採択されたデータを利用している。ただし、科学技術研究に対応す

注 <https://www.nsf.gov/awardsearch/download.jsp>

る内容を把握するため、今回は Standard Grant と Continuing Grant を対象とした分析を行うこととした。

NSF の採択データにはキーワードとなる単語が付与されていない。そのような場合は、課題名や概要等から単語の抽出を行うのが一般的であるが、一般的なテキストマイニングの手法を用いた場合、わかち書きで得られる単語が科学技術用語と一致しにくく、その単語を用いた分析の結果から得られる情報では科学技術動向を考察することが難しい。そこで今回は各課題に対し国立研究開発法人科学技術振興機構 (JST) の JST 科学技術用語シソーラスに準拠したキーワードの付与を行った³⁾。

なお、共起ネットワーク図を作成する方法として、今回は、GUI の操作を行うことで R 言語を用いた共起ネットワーク図の作成ができる KHCoder⁴⁾ を利用している。

3. 基礎データ

2019~2020 年に採択された NSF の課題の基礎データを調べた。

採択件数は 25,145 件 (2020 年 :12,568 件、2019

年 :12,577 件)、うち Standard Grant は 20,013 件 (2020 年 :10,032 件、2019 年 :9,981 件) であった。1 課題当たりの平均キーワード数は 30.5 個 (2020 年 :31.0 個、2019 年 :30.0 個) であった。

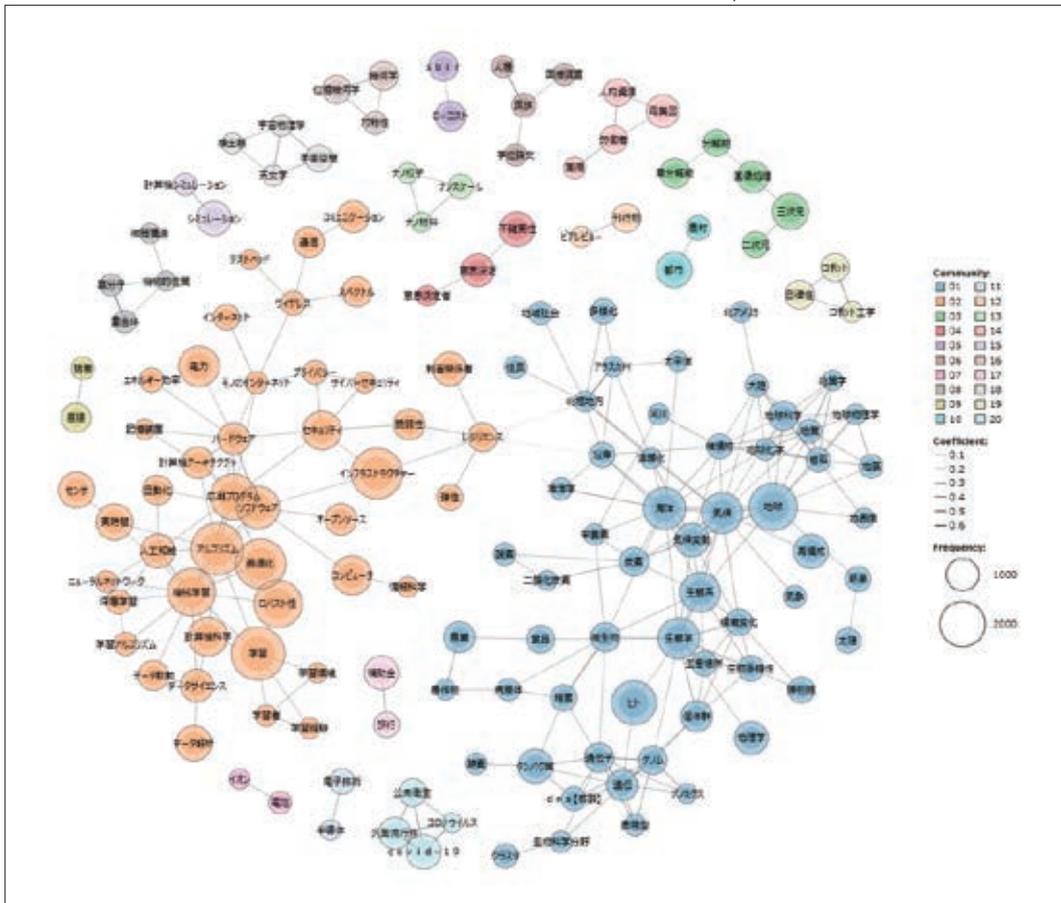
4. 分析結果

共起ネットワーク分析によりサブグラフ検出 (modularity)、中心性 (媒介)、共起パターンの変化 (相関) について図を作成した。キーワードの出現数は、最小出現数 350 に設定したところ 271 個のキーワードが該当した。このキーワードを用いて共起ネットワーク分析を行った結果、Node 数 155、Edge 数 271 の図が作成できた。この図から考察を行う。

4-1 《サブグラフ検出》 図表 1

Community01 は「地球」「海洋」「生態学」など環境に関するキーワードが多い。Community02 は「アルゴリズム」や「機械学習」など AI に関するキーワードが多いことが分かる。特にこの 2 つの Community は全体に対して占める割合が大きく見て取れ、NSF の採択課題は環境や AI をテーマとしたものが多いことが分かる。

図表 1 サブグラフ検出 (modularity)



4-2 《中心性》 図表 2

中心性の高いキーワードを確認すると「レジリエンス」や「インフラストラクチャー」などのキーワードが中心性が最も高く、広く使われていることが分かる。つまり、これらのキーワードがハブとなる研究が行われていることが分かる。

次いで「ソフトウェア」や「機械学習」、「セキュリティ」などのキーワードが中心媒介性が高い。サブグラフ検出で確認した Community01 の中には「沿岸」や「気候変動」など環境に関するキーワードの中心性が高いことが分かる。

4-3 《共起パターンの変化》 図表 3

共起パターンの変化は、採択年度の相関係数を計算し、そのキーワードが 2020 年に正の相関があれば赤く、負の相関があれば青くなっている。

ここでは、「covid-19」などのキーワードが赤くなっており、2020 年の採択課題はコロナ関連の研究が行われていることが分かる。

米国の研究開発政策動向に関しては JST 研究開発戦略センター (CRDS) が報告書を公開しており、2021 年版では、トランプ政権とバイデン新政権における研究開発戦略が詳細にまとめられている。米国に

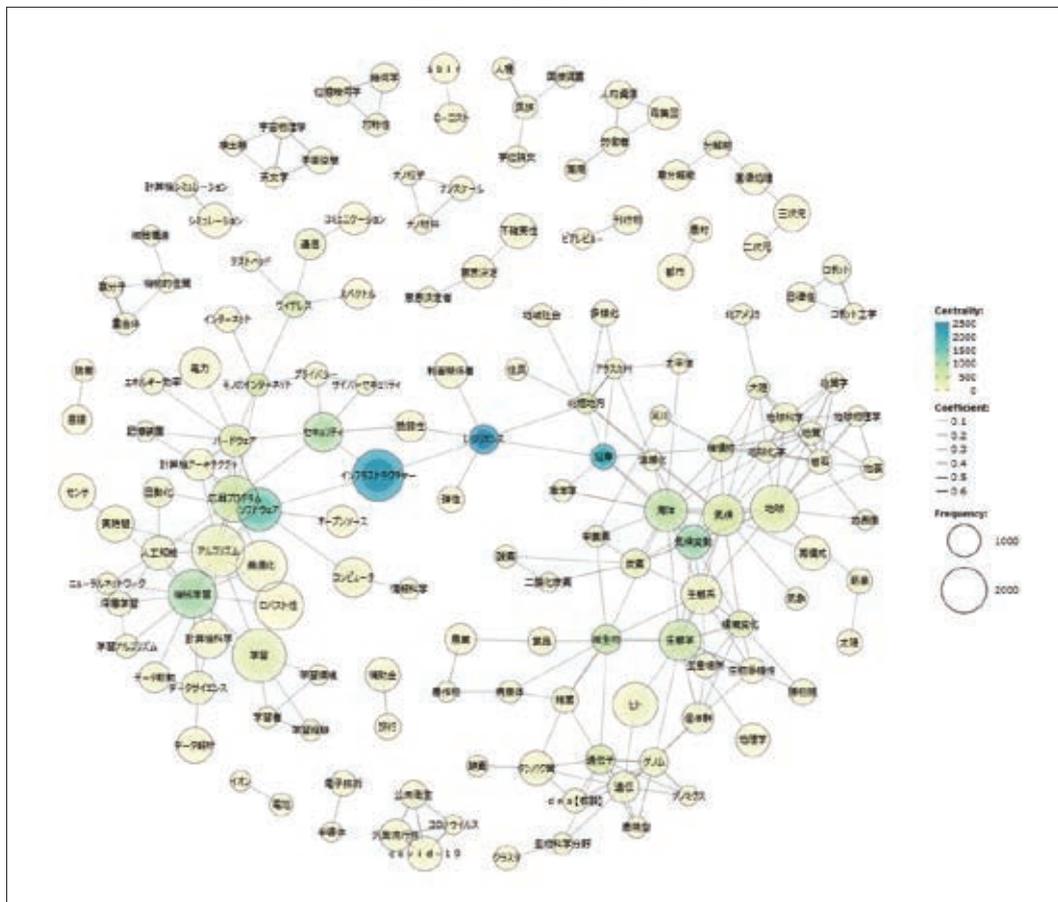
おいては環境や AI をテーマとした研究やコロナ禍^かに対応する研究開発にも力を入れていることが分かり、本分析結果と合致している⁵⁾。

5. まとめ

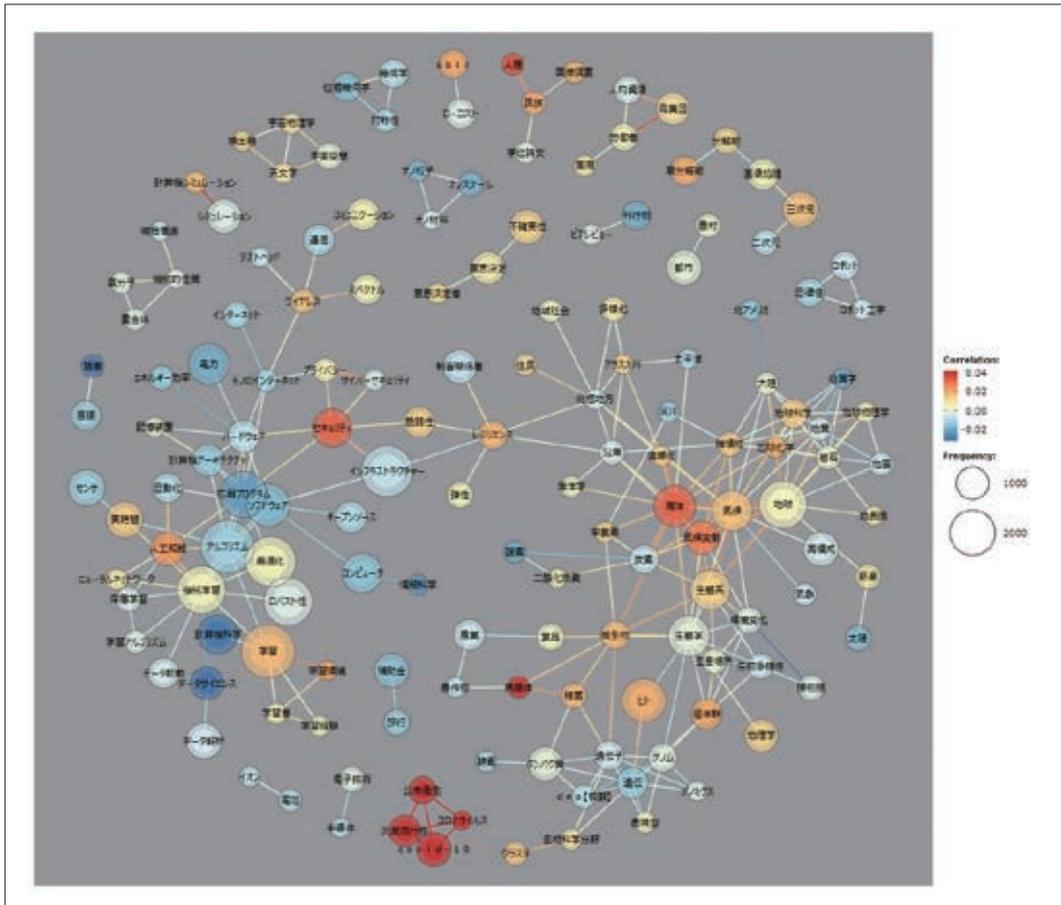
本研究では、ARAKI システムを用いた分析結果による共起ネットワーク図から研究動向の考察を行った。サブグラフ検出により Community における研究の動向を読み取ることができ、環境や AI に関するテーマが主流であることがうかがえた。また、中心性により、ハブとなるキーワード「レジリエンス」「インフラストラクチャー」が確認できた。最後に共起パターンの変化で採択年による流行のキーワードを確認できた。

要約すると、近年ではコロナウイルスに関するキーワードの研究課題が多くなってきており、2019～2020 年の採択課題は環境や AI に関するキーワードの研究が多く見て取れ、「レジリエンス」「インフラストラクチャー」などのキーワードがハブとなった研究が多いことが分かった。この結果は CRDS の報告書に記載されていた米国の政策動向に沿った内容が得られたといえるだろう。

図表 2 中心性 (媒介)



図表 3 共起パターンの変化 (相関)



最後に、テキストマイニングで良質な分析結果を得るには、分析する対象となるデータから抽出する単語が重要であると考えられる。特に科学技術に関する分析となれば、それに相当する単語に特殊性があるため、科学技術用語を整備している JST 科学技術用語シソーラスがなければ今回のような結果を得ること

ができなかった。日本における科学技術政策のための EBPM 推進には、こうした用語辞書を始め、各種科学技術に特化した辞書整備が重要な役割を担っていると見える。また、科学技術は日々進歩しており、それに対応するためにも切れ目ない継続した整備が求められる。

参考文献・資料

- 1) 内閣府, 第 6 期科学技術・イノベーション基本計画, (オンライン) 2021 年 3 月 26 日, <https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html>
- 2) 荒木寛幸. EBPM のための研究プログラムの分析 (科研費を事例として) - Advanced Research Analysis in Keen-Keywords Investigation -. 文部科学省科学技術・学術政策研究所 (NISTEP). 2020, STI Horizon, Vol.6, No.1
- 3) 科学技術振興機構 知識基盤情報部, JST 科学技術用語シソーラスの改訂. 情報管理. 2015, vol. 58, no. 1, p. 70-73. doi: <https://dx.doi.org/10.1241/johokanri.58.70>
- 4) 樋口耕一. テキスト型データの計量的分析 —2 つのアプローチの峻別と統合— 『理論と方法』. 数理社会学会. 2004, 19(1), p.101-115.
- 5) 国立研究開発法人科学技術振興機構 研究開発戦略センター. 研究開発の俯瞰報告書 主要国の研究開発戦略(2021 年). CRDS-FY2020-FR-05. 2021.