

ネットワーク指標を用いた学際的な論文の抽出

Extraction of Interdisciplinary Papers by Network Indicators

梶川裕矢, 森純一郎

Yuya KAJIKAWA, Junichiro MORI

東京大学大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

〒113-8656 東京都文京区弥生2-11-16工学部9号館

E-mail: kaji@ipr-ctr.t.u-tokyo.ac.jp

我々が直面している持続可能性, 少子高齢化といった複雑な課題は単一の学問領域では解決できず, 領域間に跨った学際的な研究が重要であるとされている. 本研究では, 引用ネットワーク分析を用いて, そのような学際的な論文を抽出する手法を提案する. 学際性を測定するネットワーク指標を用いて, 持続可能性学をケースとして分析を行った結果, 複数の指標の組み合わせにより, 学際的な論文を効率的に抽出可能であることが分かった.

Complex issues we face such as sustainability and aging society cannot be solved within an academic discipline and interdisciplinary research is important and therefore needed. In this paper, we propose a methodology to extract interdisciplinary papers by citation network analysis. Using network indicators measuring the extent of interdisciplinary, we performed a case study in sustainability science. We found that combining indicators is an effective way to extract interdisciplinary papers.

キーワード: 引用分析, ネットワーク分析, 学際性, 分野横断性

citation analysis, network analysis, multidiscipline, interdisciplinary

1 緒言

引用ネットワーク分析は, 日々膨大な数が出版される学術論文の全体の構造を俯瞰し, 目的に応じて適切な論文を抽出するための有効な手法である. 引用ネットワーク分析においては, 共引用, 書誌結合, 直接引用といった手法を用いてリンクを形成したのち, 被引用数などの統計的集計量の算出, ネットワークの可視化やクラスタリングが行われる^[1,2]. また, 時

系列分析や論文の平均出版年の算出により萌芽的な研究領域を検出することが可能である^[3,4].

それらの分析により, 複雑な学術研究の全体像を粗視的に把握することが可能である. しかし, 膨大な量の情報を単純化することによる情報の見落としは避けられない. 従って, 利用者の目的に応じた情報の縮約の方法を構築する必要がある.

本研究の目的は, 学際的な論文の抽出を行うための手法を構築することである.

本研究では学際性を測定する複数の指標を設定し,その妥当性と有効性を評価する.分析の対象としてはサステナビリティ学を取り上げる.持続可能性や少子高齢化といった複雑な課題は単一の学問領域では解決できず,領域間に跨った学際的な研究が重要であるとされている^[5,6].Kajikawaらは引用分析を用いて,サステナビリティ学が,農業,森林,水,エネルギーなど持続可能とすべき対象によって分かれていることを明らかにしている^[5].

しかし,それらは本来相互に関連している.例えば,森林の存在が農業の生産性に影響を与えるし,造水にはエネルギーが必要だ.本研究では,近年,計量書誌学の分野で研究が盛んになってきた学際性を測る研究^[7,8]において用いられている複数の指標を,サステナビリティ学に対して適用し評価したので報告する.

2 手法

媒介中心性はネットワークの中で他のノードを結びつける機能を測定する指標としてよく用いられる.Leydesdorffはこの媒介中心性を引用ネットワーク分析に適用し,媒介中心性(Betweens Centrality, B)を用いて学際性を評価できるとしている^[7].また,Rafolsは学際性を測る指標として,当該文献が引用している文献の引用文献(RefRef)が属している,カテゴリー数(N),Shannonの情報量(H),Simpsonの多様性(I),Strling(S)を用いて評価している.各指標の定義は以下の通りである.

$$H = -\sum_i p_i \ln p_i$$

$$I = 1 - \sum_i p_i^2$$

$$S = 1 - \sum_{i,j} s_{i,j} p_i p_j$$

ただし, p_i はカテゴリー*i*に属するRefRefの割合, $s_{i,j}$ はカテゴリー*i, j*間の距離である.

我々のサステナビリティ学分野における既存研究では,引用ネットワークをNewman法^[9]法を用いてクラスタリングを行い,可視化することで全体像の分析を行っている.従って,本研究でも同様にNewman法を用いたクラスタリングを行い,各クラスタをカテゴリとした.また, $s_{i,j}$ としては,クラスタ間のリンク数を各クラスタのノード数で正規化したものを用いた.

上記指標に加え,クラスター分散度(Participation coefficient, P)^[10],次数中心性(Degree, D)を測定した.

3 結果と考察

表1に各指標間のPearsonの相関係数を示す.Rafolsの提案している,N, H, I, Sはどれも互いに相関性が高い.Pは他の指標との相関性が低く,Leydesdorffの提案しているBはDとの相関性が高い.これは,Rafolsの指標群はどれも属しているクラスターの数,すなわち,Nの影響を強く受けるためと考えられる.また,Bはしばしばネットワークにおける媒介性の指標として広く使われるが,本質的には中心性の指標であり,リンク数が多いほど中心性も高くなるためBとDの相関性が高くなったといえる.一般に引用ネットワークのクラスタリングにより得られた各クラスタの特徴は各クラスタにおいてハ

ブとなっている論文の影響を色濃く受ける。従って、媒介中心性を学際性の指標とすることは少なくとも今回の場合適切でないように思われる。

表1 各指標間の相関性

	N	H	I	S	P	B	D
N	1	0.89	0.82	0.82	0.37	0.22	0.29
H		1	0.99	0.99	0.43	0.16	0.2
I			1	1	0.42	0.14	0.18
S				1	0.42	0.14	0.18
P					1	0.21	0.22
B						1	0.74
D							1

そこで、NとPに対し、どちらの方が学際性の高い論文を抽出する精度が高いか評価した。評価のためにまず、各クラスタの特徴付けを行った。特徴付けには、そのクラスタに含まれる論文が出版されているジャーナル、そのクラスタのハブ論文のタイトルとアブストラクトの情報を引用著者が行った。NとPの値が高いそれぞれ20論文を、A:クラスタの特徴と一致、B:他クラスタの特徴と一致、C:学際性が高い、という3つのカテゴリに分類した。結果、Nに対しては、A=5、B=9、C=6、Pに対しては、A=1、B=9、C=10、という結果が得られた。すなわち、少なくとも今回のサステイナビリティ学に対しては学際性の指標としてクラスタ分散度が有効であることが分かった。

4 結言

引用ネットワークのクラスタリングや可視化は学術分野の全体像を把握するための有効な手法であるが、しばしば萌芽

的な分野、学際性の高い分野の存在を見落としがちである。本研究では引用ネットワーク中の各論文のトポロジカルな位置を表す複数の指標をサステイナビリティ学における論文集合に対して適用した。各指標間の相関性を分析することで、各指標を3つのカテゴリに整理した。これらのうち、媒介中心性はリンク数との相関が強く学際性の測定には不適切であることが示唆された。属するクラスタ数とクラスタ分散度の比較では、クラスタ分散度の方が学際性の評価指標として優れていることが示唆された。

参考文献

- [1] Small, H.: "Visualizing science by citation mapping", Journal of the American Society for Information Science, Vol.50, pp.799-813, 1999.
- [2] Börner, K.; Chen, C.; Boyack, K.W.: "Visualizing knowledge domains", Annual Review of Information Science & Technology, Vol.37, pp.179-255, 2003.
- [3] Small, H.: "Tracking and predicting growth areas in science", Scientometrics, Vol.68, pp.595-610, 2006.
- [4] Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K.: "Detecting emerging research fronts based on topological measures in citation networks of scientific publications" Technovation, vol.28, pp.758-775, 2008.
- [5] Kajikawa, Y.; Ohno, J.; Takeda, Y.; Matsushima, K.; Komiyama, H.: "Creating an academic landscape of sustainability science: an analysis of the citation network", Sustainability Science, Vol.2, pp.221-231,

2007.

[6] Kajikawa, Y.: "Research core and framework of sustainability science", Sustainability Science, Vol.3, pp.215-239, 2008.

[7] Rafols, I., Meyer, M.: "Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience", Scientometrics, in press.

[8] Leydesdorff, L.; "Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals",

Journal of The American Society For Information Science and Technology, Vol.58, pp.1303-1319, 2007.

[9] Newman, M.E.J.: "Fast algorithm for detecting community structure in networks", Physical Review E, Vol.69, artNo.066133, 2004.

[10] Guimera, R.; Amaral, L.A.N.: "Functional cartography of complex metabolic networks". Nature, Vol.433, pp.895-900, 2005.