

学術文献データから構成される共著ネットワークを用いた 有望な研究者の探索

Detecting Expected Researchers using Co-Authorship Networks
from the Academic Paper Database

藤田 正典^{*1} 青木 健^{*1} 井ノ上 寛人^{*2} 寺野 隆雄^{*1}

Masanori Fujita^{*1}, Ken Aoki^{*1}, Hiroto Inoue^{*2}, and Takao Terano^{*1}

^{*1} 東京工業大学

^{*2} 東京電機大学

^{*1} Tokyo Institute of Technology

^{*2} Tokyo Denki University

In this paper, we conducted a case study in the field of life science to propose a method to detect expected researchers by analyzing co-authorship networks from the academic paper database of Japan Science and Technology Agency (JST). We define the researchers with high centralities in the co-authorship networks as excellent researchers, then we propose a method to detect expected researchers by analyzing growth of the centralities in the course of time. In addition, we aim to clarify the characteristics of expected researchers by analyzing relationship between these researchers and their research fields and/or institutions they belong.

1. はじめに

組織がイノベーションを推進するにあたり、研究者は不可欠な存在であり、将来の成長が期待される有望な研究者の確保はイノベーションを推進しようとする組織の最重要課題の一つである。しかしながら、有望な研究者を発掘することは容易ではなく、有望な研究者を客観的に判別できる指標の開発が望まれる。一方、研究者の研究分野ごとの研究体制は様々であり、数学等の理論的な研究手法を主体とする分野では個人が単独で、生物学や物理学等の実験的な研究手法を主体とする分野ではグループで組織的に研究する傾向にある(Newman(2004))。また、生命工学や生命情報学は生物学と工学や情報学が融合した学際分野であり、複数の分野の研究者や研究組織が共同で研究をすることが望まれる。したがって、研究者の評価基準も、一律なものではなく、研究分野によって異なると考えられる。

本研究では、生命科学分野を事例として、複数研究者がグループで組織的に研究する分野において、有望な研究者を探索する手法を提案する。この手法では、①研究者の組織的研究活動の能力の指標として、組織におけるリーダーシップやチームワーク等個人の特性を考慮し、研究者の論文とその共著関係から構築されるソーシャルネットワーク(共著ネットワーク)の中心性に注目する。また、②有望な研究者(将来に優れた成果を生み出すことが期待される研究者)を予測するため、研究者の中心性の時間的な推移を分析する。つまり、本手法では、研究者の中心性の時間的上昇過程を分析ことにより、有望な研究者を探索する。

2. 関連研究

2.1 研究者の評価指標

研究者の評価の基準として、文部科学省(2014)の「研究者等の業績に関する評価に関する調査・分析報告書」では、研究

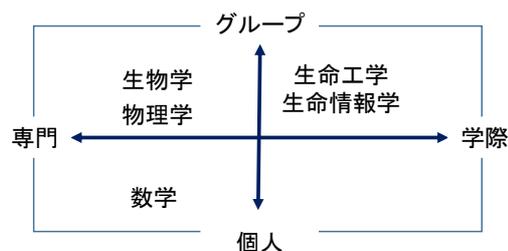


図 1: 研究分野ごとの研究特性

者の学会発表・講演、論文・解説、等が研究者の評価基準の上位に挙げられており、論文が研究者の評価にとって重要な位置づけであることが分かる。研究者の評価のために使われる論文に関する指標としては、Impact Factorやh-indexなどが挙げられることが多いが、これらは研究成果である論文の評価指標であり、今後成長が期待される研究者自身の特性、例えばリーダーシップやチームワーク等について考慮されていない。研究分野を考慮した評価指標としては、大学評価・学位授与機構(2015)の「教育・研究水準の学系別評価基準のあり方にかかる調査研究報告書」が、研究分野の特性を踏まえ、研究分野を7つの研究分野に分けて評価の基準の参考例を示している。この報告書では、学術面の観点に加え、社会・経済・文化面の観点も含めているが、これらも研究成果についての評価であり、研究者自身の特性を踏まえた評価指標とは言えない。

2.2 論文の引用数の関する評価指標

研究者の評価のためには、執筆した論文に関する評価指標が活用されることが多い。論文の評価指標としては、執筆論文数に加え、Impact Factorやh-indexなどが挙げられる。これらの指標は論文の引用数をもとに求められる。Impact Factorは、各ジャーナルの掲載論文の引用数をもとにした指標であり、論文が掲載されるジャーナルの評価を示す指標であるため、研究者や論文そのものを直接評価する指標でない。一方、h-indexは、ある研究者の論文について被引用数がh以上である論文の数

が h 以上であることを満たす最大の数値と定義されるが、Impact Factor と同様に、論文という研究成果についての評価指標であり、研究成果を生み出す研究者自身の特性についても考慮が必要と考えられる。

2.3 研究者のコンピテンシー

Spencer (1993) は、「ある職務または状況に対し、基準に照らして効果的、あるいは卓越した業績を生む原因として関わっている個人の根源的特性」を「コンピテンシー」として定義し、ある個人が持つ表層的な知識やスキルに加え、個人の態度や価値観等の隠された特性の重要性を論じている。Spencer (1993) は、コンピテンシーとして、達成・行動(達成思考等)、援助・対人支援(対人理解等)、インパクト・対人影響力(インパクト等)、管理領域(チームワーク・リーダーシップ等)、知的領域(分析的思考等)、個人の効果性(自己管理等)、等を挙げている。また、McClelland (1973) は、「Testing “Competence” Rather Than “Intelligence”」と述べ、知性に加え、コンピテンシーの重要性を論じている。

本研究では、組織的に複数の研究者が共同で研究する分野における研究者の評価基準として、チームワークやリーダーシップといったコンピテンシーを考慮し、研究者の論文とその共著関係から構築されるソーシャルネットワーク(共著ネットワーク)の中心性に注目する。共著ネットワークの中心性は、研究者が論文を通して発表した学術的成果(Intelligence)と研究者の組織の中でのチームワークやリーダーシップ(Competency)を併せ持った指標と考えられるからである。

2.4 共著ネットワーク

共著ネットワークとは、論文の共著関係をもとに構築したソーシャルネットワークの一つである。Newman (2004) は、生物学、物理学、数学の3分野の論文による共著ネットワークの構造分析を行い、論文あたりの著者数は、研究様式によって異なり、理論的な研究手法を主体とする分野(数学等)よりも、実験的な研究手法を主体とする分野(生物学、物理学等)の方が多い傾向にあることを示した。篠田(2011)は、人工知能学会論文誌に掲載された論文の共著ネットワークを作成して人工知能学会における中心人物の特定を行い、また、その中心人物の移り変わりをみることで、人工知能学会の系譜の作成を行った。宮西(2012)は、高エネルギー物理理論の論文の共著ネットワークを用い、時系列ネットワークから将来的に重要となる研究者の順位付け手法を提案している。森(2014)は、コンピュータ分野の論文を網羅している ACM Digital Library の論文データベースを用い、シードとして特定した研究者から共著関係を逐次的にたどることで作成される共著ネットワークにより、萌芽領域の中心的研究者の予測手法を提案している。しかし、これらの研究では、今後成長が期待される有望な研究者の抽出やその属性について研究されていない。本研究では、有望な研究者の探索することを目的とし、生命科学分野の論文の共著ネットワークを用いて、中心性が上昇した研究者を特定するとともに、有望な研究者の特性を明らかにすることを旨とする。

3. 対象データ

本研究では、科学技術振興機構(JST)が提供する学術文献データベース(JSTPlus)を対象とする。JSTPlus は医学を含む科学技術全分野に関する文献情報を 2500 万件以上収録しており、本研究では、JSTPlus から表 1 に示すような生物科学分野(JST 分類コード E)を抽出して分析した。

表 1: JST 分類コード E (生物科学) の分野

EA 生物科学一般	EG 微生物学, ウイルス学
EB 生化学	EH 植物学
EC 遺伝学, 進化論	EJ 動物学
ED 生体防御と免疫系	EK 放射線生物学
EE 生態学, 環境生物学	EL 生体工学
EF 細胞学	

4. 手法

4.1 有望な研究者の定義

本研究では、有望な研究者の測定指標として、共著ネットワークにおける中心性に注目する。

(1) 論文から構築されるソーシャルネットワーク

論文から構築されるソーシャルネットワークとしては、以下のようものが挙げられる。

- 引用論文ネットワーク: 論文の引用関係から構築される論文をノードとするソーシャルネットワーク
- 共著ネットワーク: 論文の共著関係から構築される著者をノードとするソーシャルネットワーク

本研究においては、組織におけるチームワークやリーダーシップ等のコンピテンシーを考慮し共著ネットワークを採用する。

(2) ネットワークの中心性

ネットワークの中心性の指標としては、以下のようなものが挙げられる。

- 次数中心性: ネットワーク内で他の研究者とどのぐらい多くのリンクを持つかを示す
- 近接中心性: 他の研究者への最短距離の総和の逆数であり、ネットワーク内の全ての研究者とどのぐらい近いのかを示す
- 媒介中心性: ある研究者が他の研究者間の最短経路上に位置する程度を示し、他の研究者同士をつなぐ働きをする研究者を示す
- Page Rank: 重要度が高い研究者とリンクを持つ研究者は重要度が高いことを示す

本研究においては、他の研究者とのつながりを考慮し、媒介中心性を採用する。

(3) 中心性の時間的推移

本研究では、共著ネットワークにおける中心性の高い研究者を「優秀な研究者」と定義するとともに、中心性が時間的に高くなる研究者を「有望な研究者」と定義する。その上で、中心性の時間的推移による研究者の成長過程を分析することで「有望な研究者」を探索する。

4.2 有望な研究者の抽出プロセス

本研究では、以下のプロセスで有望な研究者を抽出する。

- ① 学術文献データベースから、分析対象とする研究領域の論文を発行年ごとに分割して抽出。
- ② 抽出した論文の共著関係をもとに、発行年ごとに共著ネットワークを構築。
- ③ 共著ネットワークにおいて、各著者の中心性を算出。
- ④ 中心性の時間的推移を分析し、「有望な研究者」を抽出。

5. 結果

4.2 のプロセスに沿って、共著ネットワークの中心性の時間的推移を分析し、「有望な研究者」の抽出や特定を行った。

5.1 発行年ごとの分析対象論文の抽出

JST が保有する学術論文データベース(JSTPlus)から分析対象とする生物科学分野(JST 区分 E)の論文を、表 2 に示す通り発行年ごとに 2006 年から 2015 年まで約 134 万件抽出した。

表 2: 発行年ごとの論文数

発行年	2006	2007	2008	2009	2010	
論文数	122,191	132,722	130,656	138,432	137,717	
発行年	2011	2012	2013	2014	2015	合計
論文数	138,405	131,785	147,125	136,272	126,518	1,341,823

5.2 共著ネットワークの構築

抽出した論文から、発行年ごとに、その共著関係をもとに共著ネットワークを構築した。2006 年度の共著ネットワークのクラスタの分布を図 2 に示す。図 2 の横軸はクラスタ数、縦軸はクラスタサイズの対数を示しており、共著ネットワークは、円の部分で示された一つの大きなクラスタを含む複数のクラスタで構成されていることが判る。

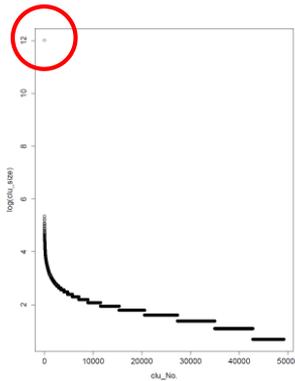


図 2: 共著ネットワークのクラスタの分布

また、著者ネットワークにおけるある 1 つのクラスタの例を図 3 に示す。図 3 の円の部分は、媒介中心性が高い著者を示している。

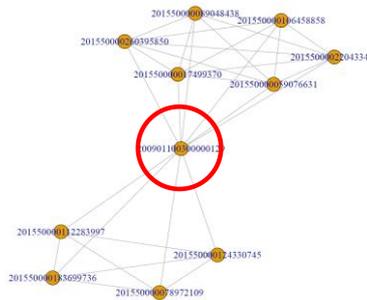
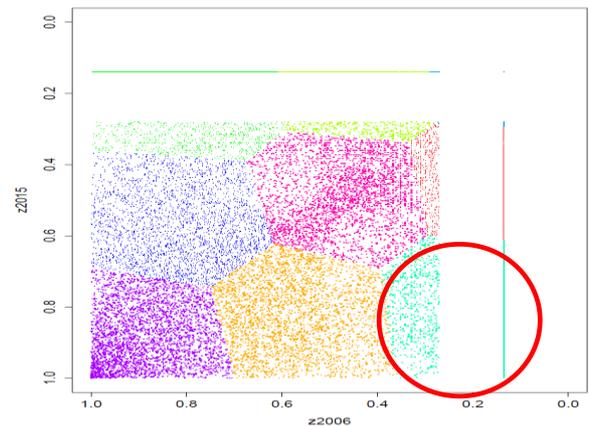


図 3: 共著ネットワークの例

5.3 中心性の算出

共著ネットワークの著者の中心性の値を発行年ごとに算出し、抽出した中心性の値をもとに発行年ごとに著者の中心性の順位付けを行った。図 4 は、2006 年と 2015 年の著者の中心性の順位を散布図(横軸:2006 年の中心性の順位, 縦軸:2015 年の中心性の順位)に示し、クラスタ化(k-means 法:k=9)して色分けしたものである。図 4 の円の部分は、2006 年度の中心性の順位が低く、2015 年度の中心性の順位が高い著者のクラスタを示しており、中心性が時間と共に上昇した有望な研究者を抽出できることが判る。

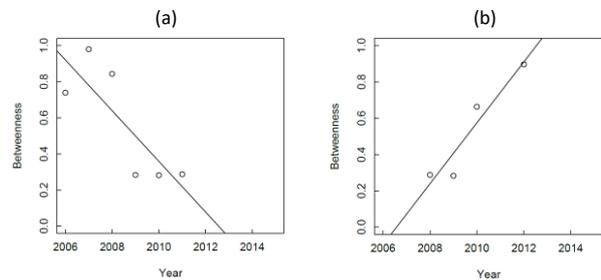


(横軸: 2006 年, 縦軸: 2015 年)

図 4: 共著ネットワークの中心性順位

5.4 発行年ごとの分析対象論文の抽出

各著者の中心性の順位の時間的推移を線形回帰分析した結果の例(中心性の出現回数 ≥ 4 , p-value <0.05)を図 5 に示す。ここで、図 5 (a)は中心性の順位が低下した研究者を示している。また、図 5 (b)は中心性が上昇した研究者を示しており、中心性が時間と共に上昇する有望な研究者を特定できることが判る。



(中心性が低下した著者例)

(中心性が上昇した著者例)

図 5: 研究者の中心性の推移(線形回帰)

6. 議論

6.1 JSPS 特別研究員の中心性

前章で示した結果を評価するため、「有望な研究者」のベンチマークとして、「日本学術振興会 (JSPS) 特別研究員」を取り上げる。JSPS 特別研究員は、我が国の学術研究の将来を担う優れた若手研究者として採用される研究者である。生物科学分野の約 136 万件の論文のうち、所属機関名に「日本学術振興会」、「特別研究員」、「JSPS」が含まれる論文の著者を抽出し、JSPS 特別研究員の中心性の時間的推移を分析した。

(1) JSPS 特別研究員の中心性の上昇

JSPS 特別研究員の中心性の順位の時間的推移を線形回帰分析した結果を図 6 に示す。図 6(a), (b), (c) は、それぞれ 2006 年、2008 年、2010 年に JSPS 特別研究員として初めて論文を発行した著者 (109 名、116 名、155 名) の場合の分析結果 (傾き = 0.019, 0.017, 0.027, p-value < 0.05, 0.01, 0.05) である。有望な研究者とされる JSPS 特別研究員の共著ネットワークの中心性は有意に上昇していることが判る。

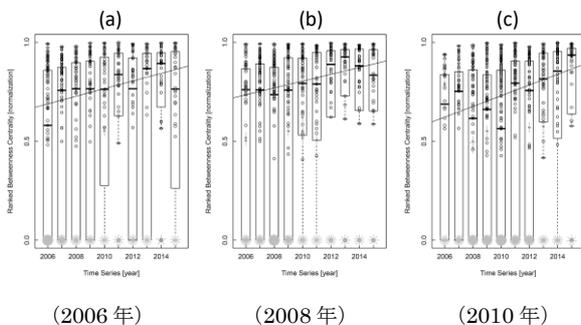


図 6: JSPS 特別研究員の中心性の推移 (線形回帰)

このように、有望な研究者と中心性の上昇の間には相関関係が存在しており、有望な研究者の指標として研究者の中心性の推移に注目し、これを分析することが有効であると考えられる。

(2) JSPS 特別研究員の中心性と全研究者の中心性の比較

図 7 は、2008 年に JSPS 特別研究員として初めて論文を発行した著者 (116 名) と 2008 年に初めて論文を発行した全著者の

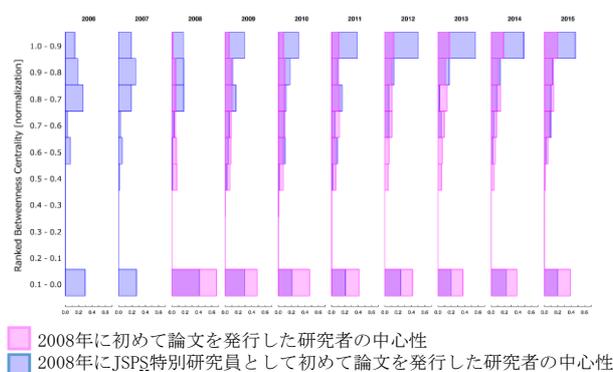


図 7: JSPS 特別研究員と全研究者の中心推移の比較

各発行年の中心性の順位の十分位数を、発行年ごとに重ねて示し、横に並べたものである。一般に、研究者は研究を重ねることにより成長し、研究者の中心性も研究者の年齢とともに上昇すると考えられる。一方、図 7 から、全研究者の中心性と比較し、有望な研究者とされる JSPS 特別研究員の中心性は、時間の推移とともに大きく上昇していることが判る。

このように、有望な研究者は、中心性が上昇している研究者の分析により探索できると考えられる。

6.2 今後の課題

生命工学や生命情報学分野のような学際分野においては、一つの研究分野における組織的な研究活動だけでなく、複数の研究分野の研究者や研究組織が融合的に研究をすること有効であると考えられる。したがって、今後、著者の中心性の上昇と、その著者の研究分野の拡大や所属機関の変化等との時間的先行・遅行関係を分析することにより、有望な研究者とこれらの因果関係を明らかにすることを目指す。

7. おわりに

本稿では、生命科学分野を事例として、科学技術振興機構 (JST) が提供する学術文献データから構成される共著ネットワークにおける中心性の高い研究者の時間推移を分析することで有望な研究者を探索する手法について述べ、日本学術振興会 (JSPS) 特別研究員の分析結果をもとに、この手法が有効であることを論じた。今後、著者の中心性の上昇と、その著者の研究分野や所属機関等の属性との因果関係を明らかにすることを目指す。

謝辞

本研究の実施にあたっては、JST より提供された文献データを利用して、JST 関係者の皆様に感謝の意を表する。

参考文献

- [Newman 2004] Newman, Mark E. J.: "Coauthorship networks and patterns of scientific collaboration." Proceedings of the National Academy of Sciences, Vol. 101, suppl. 1, pp. 5200-5205, 2004.
- [文部科学省 2014] 文部科学省: 研究者等の業績に関する評価に関する調査・分析報告書, 2014.
- [大学評価・学位授与機構 2015] 大学評価・学位授与機構: 教育・研究水準の学系別評価基準のあり方にかかる調査研究報告書, 2015.
- [Spencer 1993] Spencer, L. M., and Spencer, S. M.: Competence at Work, Wiley, 1993.
- [McClelland 1973] McClelland, D. C.: "Testing for 'Competence' Rather Than for 'Intelligence'", American Psychologist, January, pp. 1-14, 1973.
- [篠田 2011] 篠田孝祐: 日本における人工知能研究の系譜, 人工知能学会誌, Vol.26, No. 6, pp. 584-589, 2011.
- [宮西 2012] 宮西大樹, 関和広, 上原邦昭: リンク予測を基にした時系列ネットワーク中でのオブジェクトランキング, 人工知能学会論文誌, Vol.27, No. 3, pp. 223-234, 2012
- [森 2015] 森 純一郎, 原 忠義, 榊 剛史, 梶 川 裕矢, 坂田 一郎: 大規模学術論文データの共著ネットワーク分析に基づく萌芽領域の中心研究者予測に関する研究, 2015 年度人工知能学会全国大会, 2015.