

# 学術文献データの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<sup>1</sup> 青木 健<sup>1</sup> 井ノ上 寛人<sup>2</sup> 寺野 隆雄<sup>1</sup>

Masanori Fujita<sup>1</sup>, Ken Aoki<sup>1</sup>, Hiroto Inoue<sup>2</sup>, and Takao Terano<sup>1</sup>

<sup>1</sup>東京工業大学大学院総合理工学研究科 <sup>2</sup>東京電機大学未来科学部  
<sup>1</sup>Tokyo Institute of Technology <sup>2</sup>Tokyo Denki University

**Abstract:** 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。然しながら、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様式は様々であり、数学等の理論的な研究手法を主体とする分野では単独で、生物学等の実験的な研究手法を主体とする分野では組織的に複数研究者が共同で研究する傾向にある。また、研究者が所属する組織の研究目的も様々で、大学や公的研究機関においては学術的または社会的な成果が、そして産業界の企業においては経済的な利益が挙げられる。従って、研究者の評価基準は、組織ごと、研究分野ごとに異なる。

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、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、学術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。これらの指標は論文の引用数をもとに求められるが、引用に関する慣習は研

究分野により異なるため、異なる研究分野の研究者間の比較は難しいとされており、また、これらは論文の評価指標であり、研究者自身の評価指標ではない。

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、これらは研究成果についての評価であり、リーダーシップやチームワークといった、研究成果を生む研究者自身の特性についても考慮が必要と考えられる。

本研究では、生命科学分野を事例として、組織的に複数の研究者が共同で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。有望な研究者の指標として、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）の中心性に注目し、中心性の時間の推移による研究者の成長過程を分析することにより、有望な研究者を探索する。またこれらの研究者と研究分野や所属機関等との関係を分析することにより、有望な研究者の特性を明らかにすることを目指す。

## 2. 関連研究

### 2.1. 研究者の評価指標

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るが、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の論文そのものを直接評価する指標でない。また、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。しかし、引用に関する慣習は研究分野により異なるため、Impact Factor と h-index はともに、同じ研究分野における研究者同士の比較のための指標としてのみ使用されるべきであると指摘されている。これらは論文の評価指標であって研究者自身の評価指標ではなく、また、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。

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)、を挙げている。また、McClelland (1973) は、“Testing ‘Competence’ Rather Than ‘Intelligence’” と述べて、コンピテンシーの重要性を論じている。

イノベーションを推進するにあたっては、分野融合の重要性や、企業や業種間の人材交流(オープンイノベーション)の重要性が指摘されており、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する必要があると考えられる。本研究では、研究者の評価指標として、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、研究者が発表した論文の学術性(Intelligence)と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。

### 2.2. 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)の方が多く傾向にあることを示した。また、物理学では研究者同士で綿密なネットワーク上の共著関係を構築する傾向にあり、生物学では影響力のある研究者を中心とした放射状の共著関係を構築する傾向があることを述べている。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とその属性について研究されていない。原田(2005)は、研究者ネットワークでの優位性と研究費との関係を明らかにすることを目的とし、計算機利用技術分野の論文の共著ネットワークの分析をしている。本研究では、有望な研究者の探索とその属性を明らかに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、中心性が成長した研究者と特定するとともに、これらの研究者と研究分野や所属機関等との関係を分析する。

## 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(約1000万件の科学技術文献データを収録)を利用し、表1に示す生物科学分野(JST分類コードE)を抽出して分析する。また、

表1: JST 分類コード E に含まれる分野

生物科学一般	生化学	遺伝学・進化論	生体防御・免疫系
生態学・環境生物学	細胞学	微生物・ウイルス学	植物学
動物学	放射線生物学	生体工学	

JST が提供する人名名寄せデータを用いて、研究者の氏名表記の揺らぎを抑える。

## 4. 手法

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。また、中心性の時間の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」を探索するとともに、これらの研究者と研究者の研究分野や所属機関等の時間的先行・遅行関係を分析することにより、有望な研究者の特性を明らかにすることを目指す。

### 4.1. 優秀な研究者の定義

本研究では、優秀な研究者の測定指標として、論文の共著関係から構築される共著ネットワークにおける中心性を取り上げる。論文から構築されるソーシャルネットワークとしては、他に、論文の引用関係から構築される引用論文ネットワークも考えられるが、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こととした。ネットワークの中心性の指標としては、①ネットワーク内でより多くの研究者とのリンクを持つ研究者を高く評価する「次数中心性」、②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くらい近いのかを示す「近接中心性」、③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す「媒介中心性」、④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「Page Rank」等が挙げられる。

### 4.2. 分析のプロセス

本研究においては、分析プロセスを以下の通り構想している。

- ① 論文ベースから分析対象とする研究領域の論文を抽出する。
- ② 抽出した論文を発行年ごとに分割する。
- ③ 各年の論文をその共著関係を元に共著ネットワーク構築し中心性を算出する。
- ④ 中心性の時間的推移をもとに成長した研究者を抽出する。
- ⑤ 抽出された研究者の中心性の時間推移と研究者の以下の属性の時間推移の関係を分析する。
  - ・ 研究分野
  - ・ 所属研究機関
  - ・ 論文の被引用数
  - ・ 競争的研究費獲得状況

・ 所属研究機関での職位、など  
 なお、⑤については、研究者と研究分野や所属機関等の研究者の属性の時間的先行・遅行関係を分析することにより、有望な研究者とその属性の因果関係を明らかにすることを目指す。

## 5. 予備実験

### 5.1. 予備実験の内容と結果

本研究では、予備実験として、下のステップで有望な研究者を抽出し分析を行った。

- ① JST 学術論文ベース (約 1000 万件) から分析対象とする研究領域 (生物科学分野: JST 区分 E) の論文 (2003 年から 2010 年: 約 101 万件) を抽出した。
- ② 抽出した論文を発行年ごとに分割した (表 2)。
- ③ 分割した論文の発行年ごとに、その共著関係をもとに共著ネットワーク構築し (図 1)、中心性を算出した (表 3)。ネットワークは一つの大きなクラスタで構成されている。また、ネットワークのリンクは共著関係を表し、例えば 2003 年の場合、95,275 人の研究者が存在し、研究者間で 279,866 の共著関係があることが分かる。更に、②で分割した全ての発行年に存在する研究者を抽出し、抽出した各研究者

表 2: 発行年ごとの論文数

年	2003	2004	2005	2006
論文数	115,041	118,545	119,051	122,191
年	2007	2008	2009	2010
論文数	132,722	130,656	138,432	137,716

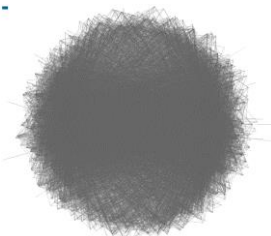


図 1: 共著ネットワーク

表 3: 共著ネットワークの研究者数とリンク数

年	2003	2004	2005	2006
研究者数	95,275	98,438	104,331	106,088
リンク数	279,866	297,434	341,651	357,806
年	2007	2008	2009	2010
研究者数	113,644	114,033	118,337	119,797
リンク数	387,048	397,972	419,227	433,100

の発行年ごとの媒介中心性を算出した。

- ④ ③で算出した中心性が 2003 年に下位 20%に位置し、2010 年に上位 20%に位置する研究者 (40 人) を有望な研究者として抽出した。図 2 に、2003 年と 2010 年における媒介中心性に基づいた研究者の散布図を示す。図 2 の横軸と縦軸は、各発行年における媒介中心性に基づいた順位を示しており、値が小さいほど順位が高いことを表す。図 2 の右下に位置する研究者が、前述した有望な研究者 40 人に該当する。この 40 人には、国立研究機関や大学教員及び民間企業の研究所研究員が含まれていることが分かった。

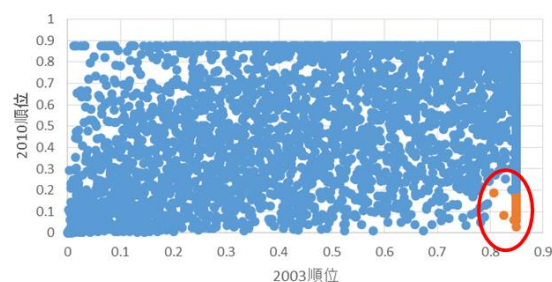


図 2: 2003 年と 2010 年における媒介中心性に基づいた研究者の分布

## 5.2. 予備実験の課題と今後の対応

今回の予備実験の課題と今後の対応は以下の通り。

- (1) (予備実験では、全ての発行年に存在する研究者のみを抽出したが) 全ての発行年に存在しない研究者も含めた「有望な研究者」探索手法の検討
- (2) 「有望な研究者」の抽出結果の評価
  - ① 今回抽出した 2010 年に上位 20%になった研究者と、2010 年に下位 20%のままの同数の研究者、及びランダムに抽出した同数の研究者について、論文の被引用数や競争的研究費獲得状況などを比較
  - ② 今回抽出した 2010 年に上位 20%の研究者の評価について、有識者へヒアリング
- (3) 抽出された研究者と、その研究者の特性 (研究分野, 所属機関, 等) との関係の分析
  - ① 中心性の時間推移と上記特性の時間推移の先行・遅行関係の有無を分析
  - ② ①で先行特性が認められた場合は、先行指標による中心性の予測
- (4) 組織のニーズに応じた有望な研究者の要件の調査 (研究者を確保したい組織へのアンケートの実施など)

## 6. 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の構想を述べ、予備実験の結果と今後の課題について論じた。今後これらの研究者と研究分野等との関係を分析することにより、有望な研究者の特性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いる。JST 関係者の皆様に感謝の意を表す。

## 参考文献

- [1] 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)
- [2] 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)
- [3] Spencer, L. M., and Spencer, S. M.: *Competence at Work*, Willy, (1993) (梅津祐良・成田攻・横山哲夫訳: コンピテンシー・マネジメントの展開, 生産性出版, (2001))
- [4] McClelland, D. C.: "Testing for 'Competence' Rather Than for 'Intelligence'", *American Psychologist*, January, pp. 1-14. (1973)
- [5] 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)
- [6] 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)
- [7] 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)
- [8] 森 純一郎, 原 忠義, 榎 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)
- [9] 原田泰輔 吉川厚 寺野隆雄; ネットワーク分析を用いた計算機利用技術分野の変遷, 計測自動制御学会第 8 回社会システム部会研究会資, Vol.8, pp.117-126, (2015)