

## 複雑ネットワークを用いた学術研究の動向に関する研究

## Complex network approach to the trend of Academic Research

原田 泰輔<sup>\*1</sup>  
Taisuke Harada周 劼亮<sup>\*1</sup>  
Keiryō SyuThawiworadilok Chaiwat<sup>\*1</sup>  
Thawiworadilok Chaiwat吉川 厚<sup>\*1</sup>  
Atsushi Yoshikawa寺野 隆雄<sup>\*1</sup>  
Takao Terano<sup>\*1</sup> 東京工業大学  
Tokyo Institute of Technology

We analyze the trend of relation between researchers' network and research fund in 10 year's time span using the academic discipline data. Also, the pattern analysis has been applied to classify the grow process of researchers who has won large amount of fund based on their affiliation and field information.

## 1. はじめに

## 1.1 背景

研究者が研究活動を行う上で、研究資金は欠かせないものである。一方で、研究活動は多くの場合、研究者個人だけではなく、大学の研究室や研究グループといった集団で行われる。

こうした背景から、研究者間の関係において優位な一を占めている研究者は、研究費の獲得においても優位であると考えられる。そこで、本研究では、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、研究者ネットワークでの中心性と科研費の取得率について分析を行う。さらに、研究者の動向を分析し、研究費を獲得できる研究者の特徴の抽出を行う。

## 1.2 関連研究

西澤らは、学術賞受賞者の科研費取得履歴を分析することにより、大型研究の採択に至るまでのパターン、分野の違いによる特徴的な採択パターンについて分析を行っている[西澤2006]。柿沼らは、研究者のキャリアパスを年齢により4つのステージに分け、各ステージにおける研究者の科研費額、学問分野間の科研費額の隔たりについての分析を行っている[柿沼2007]。しかし、これらは、科研費取得における共同研究者の要因や研究領域内でのポジションには着目していない。また、篠田は、ある研究領域における第一人者や貢献者を、論文の共著関係から見つけ出し、それらの研究者の系譜の作成を行っている[篠田2011]。しかし、これらの研究では、その領域の第一人者や貢献者を見つけ出すことに主眼を置いており、研究の成長過程に着目はしていない。

## 2. 使用データ

## 2.1 JST 学術文献データ

本研究では、科学技術振興機構の提供する、JST科学技術文献データをもとに、2003年から2013年までの『計算機利用技術分野』を対象として分析を行った。

JST科学技術文献データには、2003年から2013年までの論文タイトル、著者、論文の分野、発行年月日、発行国、著者の所属機関の情報が含まれている。JST科学技術文献データにおける計算機利用技術分野には、表1に示す分野が含まれている。計算機利用技術分野の論文数、著者数、ネットワーク密度、連

連絡先:原田泰輔, 東京工業大学, 〒226-8502 神奈川県横浜市緑区長津田町 4259 J2 棟 1705 号室,  
taisuke-harada@trn.dis.titech.ac.jp

表2 計算機利用技術分野に含まれる分野

計算機利用技術分野			
計算機利用技術一般	数値計算	記号処理	図形・画像処理
図形・画像処理[*81-'92]	図形・画像処理一般	医用画像処理	音声処理
自然語処理	パターン認識	人工知能	CAI
CAD, CAM	計算機シミュレーション	医用情報処理[*81-'92]	医用情報処理
事務・経営情報処理[*81-'92]	その他の計算機利用技術	情報処理	情報処理一般
事務・経営情報処理	分子・遺伝情報処理	その他の情報処理	

表2 計算機利用技術分野の共著者ネットワーク推移

	2003-2006	2004-2007	2005-2008	2006-2009	2007-2010	2008-2011	2009-2012	2010-2013
ノード数	69629	77151	82724	87841	90257	90121	89536	82421
エッジ数	192203	224639	250099	271661	278845	273459	266320	221437
ネットワーク密度	7.93E-05	7.55E-05	7.31E-05	7.04E-05	6.85E-05	6.73E-05	6.64E-05	6.52E-05
連結成分数	8023	8703	9207	9497	9816	10123	10383	10018
最大連結成分の占める割合	45.74%	45.89%	46.25%	46.40%	46.64%	45.58%	43.75%	40.86%

結成分数の数の変化はそれぞれの期間について表2のようである。

本研究では、著者の氏名表記の揺れを抑える為、JST 科学技術文献データの人名名寄せデータを用いてネットワークの作成を行った。人名名寄せデータは、各著者に固有のIDを付与したものである。

## 2.2 科学研究費助成事業データベース(KAKEN)

研究者の科研費取得状況を把握するため、科学研究費助成事業データベース(KAKEN)より科研費情報の取得を行った。KAKENには、当初採択時のデータ(採択課題)、研究成果の概要(研究実施状況報告書、研究実績報告書、研究成果報告書概要)、研究成果報告書及び自己評価報告書が収録されている。本研究では、JST 学術文献データより抽出した研究者名に対し、2009年から2012年までの採択の有無と研究種目をKAKENより取得した。

## 3. 分析手法

ここでは、共著者ネットワークの作成手法、科研費と共著者ネットワークとの中心性指標による順位の関係についての分析を行う。

## 3.1 共著者ネットワークの作成

共著者ネットワークの作成は下記の手順で行った。

- (1) 各論文をその発行年をもとに 第1期(2003-2006)、第2期(2004-2007)、第3期(2005-2008)、第4期(2006-2009)、第5期(2007-2010)、第6期(2008-2011)、第7期(2009-2012)、

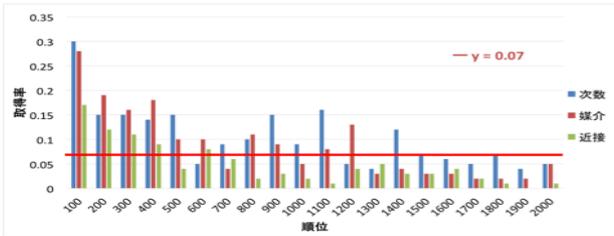


図 3 中心性順位と科研費取得率

表 4 中心性での順位と科研費取得率の相関

	次数順位-科研費取得率	媒介順位-科研費取得率	近接順位-科研費取得率
標本数	20	20	20
自由度	18	18	18
r 値	-0.71	-0.82	-0.80
p 値	4.4E-06	8.1E-06	1.9E-05

第8期(2010-2013)の8つの期間に分け、各期間での論文の著者データから、共著者間全てにリンクを張る。

- (2) (1)をもとに、複数回の共起を考慮しない、重みなし無向グラフを作成する。

### 3.2 研究者の順位付け

3.1 で作成した、各期間での共著者ネットワークを用い、次数中心性、媒介中心性、近接中心性の値を全著者に対し求め、その値により各中心性での順位付けを行った。用いた中心性指標は、次数中心性、媒介中心性、近接中心性の三つである。

### 3.3 研究者のクラスタリング

第1期間と第7期の中心性指標での順位を用い、2003年から2012年の7期間で存在する、研究者5658人を対象とし研究者のクラスタリングを行う。混合分布モデルを適用し、BIC値(ベイズ情報基準)から、混合分布モデルとクラスタ数の決定を行った。

## 4. 分析結果

### 4.1 中心性指標での順位と科研費

ここでは、科研費と共著者ネットワークの中心性指標による順位との関係について分析を行った。

第7期で各中心性 Top2000位以内の研究者について、科研費(基板B以上)の取得状況の調査を行った。ここでは、各中心性指標での順位を100名ずつ20区間に区切り、各順位区間の研究者の科研費の取得率を求めた。その結果が、図1である。また、グラフ内の赤線は、1000人の無作為に抽出した研究者の科研費の取得割合である。

図1から、各中心性順位の上位者は、科研費を取得している割合が高いことがわかる。各中心性順位と科研費の取得率との相関係数をまとめたものが、表3である。表3から、いずれの中心性指標でも科研費の取得割合と相関があることがわかる。

### 4.2 研究者のクラスタリング

図2から、第1期、第7期の両方で高順位の研究者が多く、逆に両期間で共に低順位となっている研究者は少ないことがわかる。本研究では、中心性順位の高い研究者の特徴抽出を行うため、図2の円で囲んだ4つのクラスタに着目し、分析、比較を行った。各クラスタに属する研究者数は表4であった。

[リーダー]の研究者は、全期間で高い順位を維持しており、この領域の中心的役割を果たしている研究者の集団である。逆に、[底辺]の研究者は、低い順位で推移しており、この領域の研究に貢献しているとは言えない。[成長]の研究者は、第1期では順

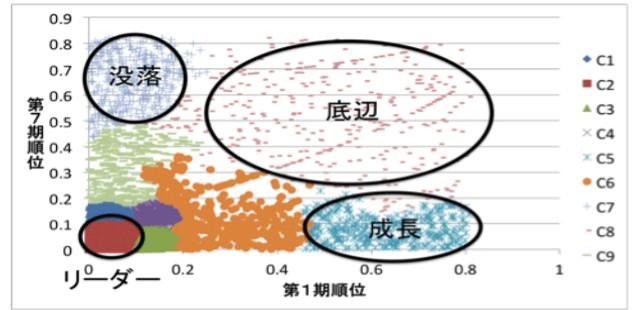


図 2 中心性での順位と科研費取得率の相関

表 3 各クラスタの人数と全体に占める割合

	リーダー	成長	没落	底辺
人数	2174	319	420	380
全体に占める割合	38.4%	5.6%	7.4%	6.7%

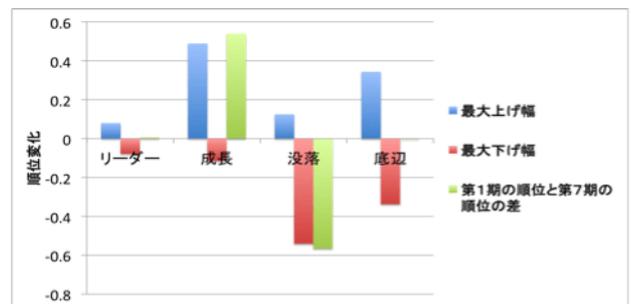


図 2 各クラスタの順位の変化幅

位が低かったものの、第7期では、高順位となっており、成長を遂げた研究者である。また、[没落]の研究者は、高順位から低順位へ推移した集団である。

以降では、この研究者のクラスタリング結果を用い、媒介中心性順位推移、所属機関、研究分野についてクラスタ間の比較を行い、クラスタの特徴抽出を行った。

### 4.3 各クラスタの媒介中心性順位推移

ここでは、第1期～第7期の全7期間を対象とし、前後期間での順位差を調査した。前後期間での順位差のうち、最も順位上がったものを最大上げ幅、最も順位低下したものを最大下げ幅とし、クラスタ間での比較を行った。また、各クラスタで第1期での順位と第7期での順位差についても比較を行った。また、各クラスタで第1期での順位と第7期での順位差についても比較を行った。

各クラスタの最大上げ幅と最大下げ幅の平均をまとめたものが図3である。この結果より、[リーダー]では、順位の変動は小さく、安定していることがわかる。一方で、[成長]では、徐々に順位が上昇していくのではなく、ある期間に急激に順位の上昇が起こっていることがわかる。[没落]でも、ある期間に急激な順位の上昇が起こっていることがわかる。また、[底辺]では、急激な順位の変動はあるものの、最終的には始めとほぼ同じ順位となっている。

### 4.4 所属情報を用いた特徴抽出

ここでは、所属情報を用いた分析についての結果を記す。所属情報を用いた分析では、所属機関の変更の有無と、所属機関変更前後での媒介中心性順位の変動についてクラスタ間での比較を行った。

所属変更前と所属変更後の順位差の取り方については、図4

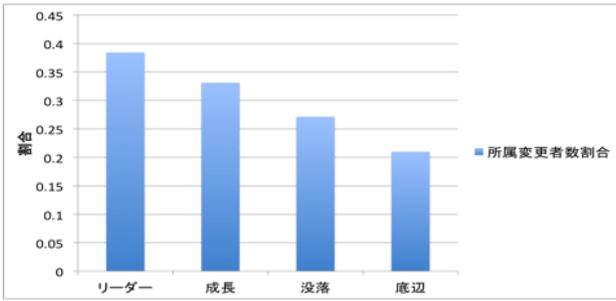


図 5 所属変更者の割合

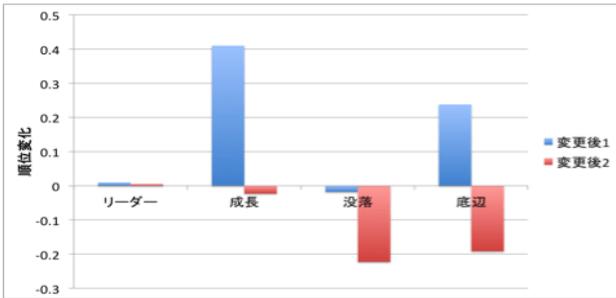


図 6 所属変更と順位変化

のように期間を定め、所属変更前と所属変更後①の差を「変更後1」、所属変更後①と所属返納後②との差を「変更後2」とし、測定を行った。



図 4 所属変更期間の取り方の例

各クラスタでの所属変更者割合が図5である。[リーダー]が最も変更者割合が高く、[成長]、[没落]、[底辺]の順で変更者割合が高いことがわかる。最も高い[リーダー]と最も低い[底辺]では、変更者割合に約2倍の差が見られる。

次に、各クラスタの所属変更前後での媒介中心性の順位の変化が図6である。[リーダー]、[没落]では、変更後1では順位の変動がほぼないことがわかる。一方、[成長]、[底辺]では、所属変更後に大きな順位の上昇があることがわかる。また、所属変更2では、[リーダー]、[成長]には大きな変化はみられず、[没落]、[底辺]は順位が大きく下がっている。

#### 4.5 研究分野情報からの特徴抽出

ここでは、研究分野情報を用いた分析の結果を記す。

まず、計算技術利用分野に含まれる小分類に対しクラスタリングを行う。JST 学術文献データでは、各論文に対し、一つまたは複数の小分野分類が付与されている。ここでは、各論文に付与された小分野分類情報を用い、小分野の共起回数によるクラスタリングを行う。これは、計算技術利用分野には、23の研究分野が含まれているが、各研究者の研究分野の動向と分析する上で、研究者の動向を把握し易くするためである。また、クラスタリング手法としては、階層的クラスタリングを用い、ワード法により、距離の計算を行った。研究分野のクラスタリング結果が表5である。計算機技術利用分野は大きく分けて5つにわけられることがわかる。しかし、クラスタ1に関しては、本研究で分析対象としている期間では一度も付与されていない分野であるため、クラスタ2から5の4分野として以降の分析を行った。

次に、各研究者の研究主分野、副分野の特定を行い、各研

表 5 研究分野のクラスタリング結果

クラスタ	1	2	3	4	5
クラスタ1	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野
クラスタ2	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野
クラスタ3	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野
クラスタ4	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野
クラスタ5	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野	計算機技術利用分野

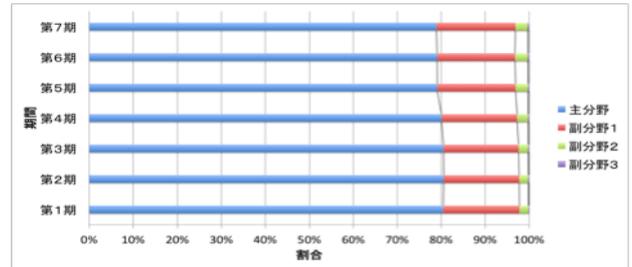


図 7 リーダーの分野推移

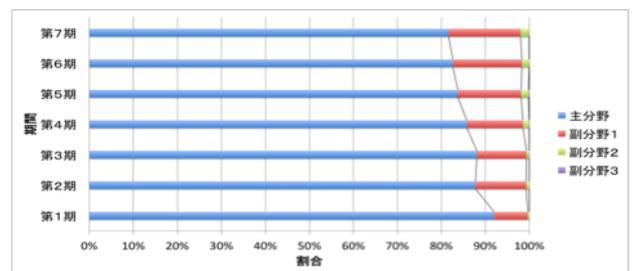


図 8 成長の分野推移

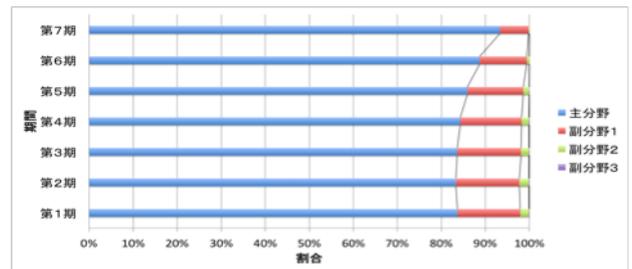


図 9 没落の分野推移

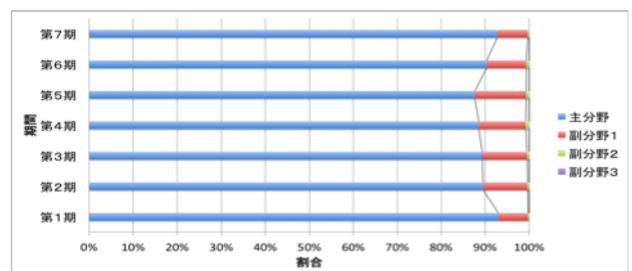


図 10 底辺の分野推移

究者クラスタの特徴抽出を行う。ここでは、各期間での研究者の全ての論文に付与された分野のうち、最も多い分野を主分野、次に多い分野を副分野1とした。

研究者の論文各期間での研究者の主分野、副分野の割合が図7から図10である。[リーダー]では全期間で主分野が約80%、副分野1が約20%となっている。[成長]では第1期では主分野の割合が90%以上あるが、期間が進むにつれ副分野1の割合が大きくなり、第7期では、主分野と副分野1の割合が[リーダー]と同程度となっている。一方、[没落]では、第1期では、主分野の割合が80%程度であるが、第7期では、90%以上となっている。また、[底辺]では、どの期間も主分野の割合が90%前後で

推移していることがわかる。

## 5. まとめ

本研究では、JST 学術文献データと科研費情報を使用し、計算機利用技術分野を対象に、共著者ネットワーク上での中心性と、科研費の取得率との関係について分析を行い、その関係を明らかにした。さらに、共著者ネットワーク上優位な位置に存在する研究者の特徴について、中心性指標での順位、所属、研究分野情報を用いた分析を行い、特徴を抽出した。

得られた各クラスタの研究者の特徴は以下である。

### a) リーダー

- ・順位の急激な変動がほぼない
- ・所属変更が多く、初速変更時の変化が小さい
- ・分野変更者が少ない

### b) 成長

- ・短期間に急激な順位の上昇がある。
- ・所属変更直後に順位の大きな上昇がある。

### c) 没落

- ・短期間に急激な順位の下降がある。
- ・所属変更後の期間で順位の大きな下降がある。
- ・主分野変更者が多い。

### d) 底辺

- ・ある程度大きな順位の上昇があるが、同程度の下降もある。
- ・所属変更直後に順位の上昇があるが、その後には下降する。

## 参考文献

- [西澤 2006] 西澤正己, et al.: 科学研究費データベースによる採択と研究成果のパターン分析, 情報知識学会誌, **16**-2, 1/6 (2006)
- [柿沼 2007] 柿沼澄男, et al.: 科学研究費補助金による研究助成の効果に関する調査—研究キャリアステージにおける科学研究費補助金取得パターンの分析—, 情報知識学会誌, **17**-2, 111/116 (2007)
- [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, **26**-6 (2011)