

共著ネットワークによる 高被引用論文と撤回論文の系譜の比較

Comparative analysis of genealogy of highly-cited papers and retracted papers through co-author networks

藤田 正典¹, 奥戸 嵩登², 寺野 隆雄³, 長根 裕美⁴

Masanori Fujita¹, Takato Okudo², Takao Terano³, Hiromi Nagane⁴

¹ 政策研究大学院大学 ² 総合研究大学院大学 ³ 千葉商科大学 ⁴ 千葉大学
¹National Graduate Institute ²the Graduate University ³Chiba University ⁴Chiba
for Policy Studies for Advanced Studies of Commerce University

Abstract

In this paper, we propose a method to show the characteristics of knowledge inheritance and diffusion based on the coauthor networks of academic papers. We construct coauthor networks from highly cited papers and retracted papers in academic literature database, and create genealogies of researchers using authorship information. Then we compare and analyze the characteristics of the genealogies using indexes such as extension / elongation, researcher's discipleship / collegueship, etc. between each paper groups. In addition, based on the results of analysis by the proposed method, we will discuss research activities from the organizational perspective.

1. はじめに

1.1. 研究における知識結合と知識移転

科学の発展やイノベーションの実現には、優れた研究者や優れた研究活動が必要である。高度で複雑に発展してきた現代の科学技術によるイノベーションを実現するためには、研究者単独の研究活動に加え、様々な研究者の英知を結集することも重要である。英知の結集にはいくつかの方法があるであろう。たとえば、複数の研究者が協業関係をつくり知識を結合して新たな知識をつくり出すことができる。また、研究者が師弟関係をつくり知識を継承して知識を発展させることである。時間的な観点では、前者はコンテンポラリーな活動であり、後者はパーマネントな活動である。それでは、研究活動の業績や結果と上記の二つの研究活動の方法(協業関係, 師弟関係)との間にはどのような関係があるのだろうか。

1.2. 研究における公正と不正

近年、研究におけるデータのねつ造、改ざん、盗用など、研究不正 (Research Misconduct) の増加に伴い、研究公正 (Research Integrity) の確保に向けての取り組みが図られている。例えば、2018年12月に公布された「科学技術・イノベーション創出の活性

化に関する法律」では「研究開発等の公正性の確保等」が織り込まれており[1]、2019年6月に閣議決定された「統合イノベーション戦略2019」においても「研究公正の推進」が謳われている[2]。また、文部科学省では「研究活動における不正行為への対応等に関するガイドライン」を定め、同省が予算を配分する研究機関に対する研究体制などの調査などを実施している[3]。研究不正の発生要因としては、研究者の理解不足や研究機関の管理体制の不備が挙げられているが、不正発生のメカニズムについては十分に明らかになっていないといえない。

1.3. 知識結合と知識移転のメカニズム

本稿では、知識結合と知識移転のメカニズムについて、研究活動の同僚関係や師弟関係に焦点をあて、高被引用論文と撤回論文とを比較しつつ分析する。上記の通り、不正発生のメカニズムについては、定量的に十分に明らかにはなっておらず、例えば、例えば、以下のような疑問が湧くであろう。

- (1) ある研究者で不正研究が発生した場合、協業する研究者でも研究不正が発生する可能性が高いであろうか。
 - (2) ある研究者の不正研究は別の研究者へ継承され、研究不正には系譜があるであろうか。
- 上記のような疑問に対して、本稿では、不正研究が

最大の原因とされる撤回論文を取り上げ、Clarivate Analytics 社が提供する Web of Science (WoS) に収録された撤回論文群、インパクトファクターが比較的高い Nature 誌 (Nature Biotech 誌および Nature Material 誌) に採録された論文群について、それぞれ共著ネットワークを構築し、著者および著者間の特徴を比較分析する。

2. 関連研究

2.1. 研究不正の研究

研究不正について、Steen(2011)は、世界的医療論文データベースである PubMed に収録された 2000 年から 2010 年までの 724 件の撤回論文の理由を調べた結果、73.5%が Error であるのに対して、26.6%のみが Fraud (ねつ造) であるが、不正論文は増加しているとした[4]。その後 Fang(2012)は、PubMed に収録された 2,047 件 (2012 年 5 月時点) の撤回論文の理由を調べた結果、Error は 21.3%のみであり、Misconduct (不正) が 67.4%で、その内訳は Fraud/Suspected Fraud (ねつ造・ねつ造の疑い) 43.4%、Duplicate Publication (多重出版) 14.2%、Plagiarism (盗作) 9.8%であり、1975 年の約 10 倍になっているとした。[5]。撤回論文が増加した背景について、Fanelli(2013)は、WoS に収録された 2,294 件の撤回論文を調査し、論文撤回の理由が、不正そのものの増加による根拠はなく、不正論文の撤回によるものであるとしている[6]。Wang(2019)は、2017 年時点でオープンアクセスジャーナルから抽出した Biomedical 分野の撤退論文 621 件の理由を調査し、Error が 148 件、Misconduct が 341 件 (plagiarism : 142, duplicate publication : 101, fraud/suspected fraud : 98) あり、不正論文論文の大半が、インパクトファクターの低い論文誌からであることを示した[7]。Campos-Varela (2019)は、PubMed に収録された 2013 年から 2016 年までの 1,082 件の撤回論文を調べ、不正が 65.3%で主な理由であること、イラン、エジプト、中国などで頻度が高いことを明らかにした[8]。Kuroki(2018)は、WoS に収録された 2016 年 1 月までの撤回論文 3,427 件と PubMed に収録された 2012 年 5 月までの撤回論文 2,047 件を分析し撤回論文にはべき乗則が適用され、一部の著者が多くの論文を撤回していることを明らかにした[9]。

これらの研究は、研究不正をはじめとした撤回論文を区分して記述したものであるが、撤回論文が発生した理由 (mechanism) について明らかにしたものではない。本研究では、不正研究の原因について、組織的研究活動の観点からの定量的に分析する。

2.2. 組織的研究活動

研究活動において、組織的活動および組織の中におけるリーダーは重要である。Crane は、科学知識の成長には関連分野の共同研究集団が影響し合うこと

を明らかにした上で、これらの集団のネットワークを「見えざる大学 (invisible university)」と呼び、科学知識の成長には研究者間の交流や共同研究集団におけるリーダーシップが重要な影響を与えること、そしてリーダーを通じてアイデアが伝播することを説いている[10]。さらに、Allen は、高度な技術的能力を持つとともに、組織の内外の研究者との接触頻度が高く、外部情報を組織内部に翻訳し伝達する仲介者を、コミュニケーションの中核を担うスター研究者として「ゲートキーパー」と呼び、その存在が研究開発の成果を向上させるとした[11]。

本研究では、組織的な研究活動やリーダーシップがその成果に影響を及ぼすことから、研究不正についても影響を及ぼす可能性があることを想定し、これを定量的に分析する。

2.3. ネットワーク分析

科学技術を定量的に研究する科学計量学においては、科学技術文献から構築されるソーシャル・ネットワークを分析することが多く、代表的なソーシャル・ネットワークとして、知識のネットワークである引用ネットワークや、共同研究者のネットワークである共著ネットワークが挙げられる。

引用ネットワークは、文献の引用関係から構築され、文献をノード、引用関係をエッジとするソーシャル・ネットワークであり、サイエンス・マップやパテントマップなどに応用されている。また、既に発行された文献をその後に発行される文献が引用するという関係性により、時間概念をもつ有向グラフとして示される。Leydesdorff ら (2010) は、ジャーナルの引用情報をもとに、ISI (現 Clarivate Analytics 社) の Subject Categories に基づく科学分野の関係を示すサイエンス・マップを作成し、サイエンスの分野間のつながりを可視化した[12]。

一方、共著ネットワークとは、文献の共著関係から構築され、著者をノード、共著関係をエッジとするソーシャル・ネットワークであり、リサーチマップなどに応用されている。一般にはある時点または期間におけるコラボレーション関係を示す無向グラフとして示される。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。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[13]。また、研究者の系譜による知識の流れの研究として、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[14].

2.4. 本研究の目的

本稿では、知識結合と知識移転のメカニズムについて、研究活動の師弟関係や同僚関係に焦点をあてて分析するとともに、不正研究が最大の原因とされる撤回論文を取り上げ、撤回論文と高被引用論文とを比較分析する。

3. 手法

3.1. 分析モデル

共同研究における研究者の役割、研究者間の関係は図1の実線で示すような師弟関係と、図1の破線で示すような同僚関係が考えられるであろう。

本稿では、撤回論文群や高被引用論文群などの共著関係をもとに、著者をノード、共著論文をエッジとして、研究者間の関係を以下のように定めて、共著ネットワークを構築し、共著ネットワークの特徴を師弟関係や同僚関係などのモデルを使って分析する。

(1) 著者順 (Authorship)

物理学などの一部の分野を除き、共同研究において、論文のオーサーシップ (著者順) は研究活動において、特定の役割を示すことが多いとされる。

- ① 第一著者 (First Author : FA) : 協業研究における担当研究者。以下に示す LA から知識を移転されることが多いと考えられる。
- ② 最終著者 (Last Author : LA) : 協業研究のスーパーバイザー。FA に知識を移転することが多いと考えられる。
- ③ 中間著者 (Middle Author : MA) : 協業研究における FA/LA 以外の協力研究者とし、FA から知識を移転されるものとする。

(2) 研究者間の関係

研究者間の関係には、知識の継承が行われる師

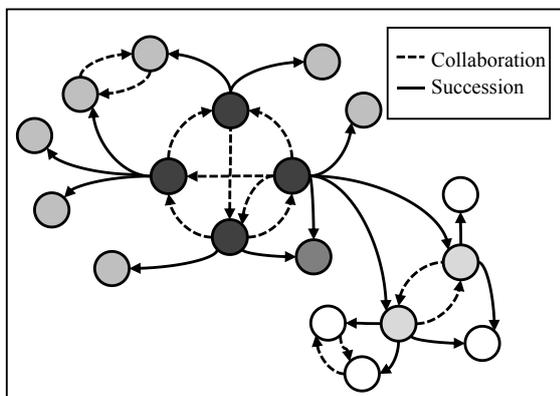


図2 組織的研究における研究者の位置付け

経営課題にAIを!ビジネスインフォマティクス研究会 (第17回)

JSAI Special Interest Group on Business Informatics (SIG-BI #17)

弟型と知識を結合する同僚型が考えられる。

① 師弟型: 二人の研究者の関係で、二人が含まれる共著論文のなかで、一方が第一著者もう一方が最終著者の論文が存在するが、逆の論文が存在しない関係

② 同僚型: 二人の研究者の関係で、師弟型を除く関係

(3) 研究者の役割・機能

師弟型の研究者間の関係において、各研究者が果たす主な役割・機能は以下の通り様々である。

① 孤立型: 単独で研究を行う

② 受入型: 知識の移転される

③ 創出型: 知識を移転する

④ 伝達型: 知識を伝達する

⑤ 媒介型: 複数の知識を移転され、複数の知識を移転する

(4) 研究者の系譜

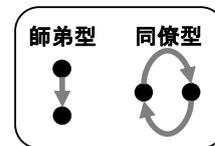
研究者の役割・機能によって、研究者の系譜の型は以下の通り変化する。

① 伸張型: 知識の移転が長く伸びる

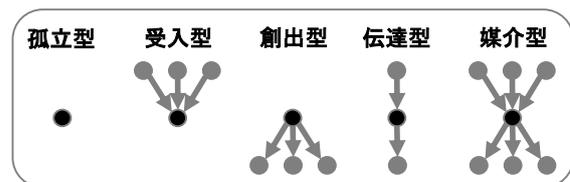
② 拡張型: 知識の移転が広く広がる

③ 凝縮型: 多くの知識が凝縮される

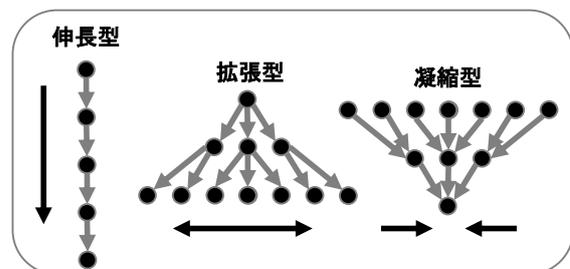
本研究では、論文の著者順に基づき研究者の関係を設定し、この関係を用いて研究者の系譜ネットワークを構築した上で、系譜ネットワークの特徴量



(a) 研究者の関係



(b) 研究者の役割



(c) 研究の系譜

図1 研究者の関係、役割・機能、系譜

と著者の特徴量を算出し分析する。

3.2. 分析対象データ

今回分析対象とするのは、WoS に収録されている発行期間が 1990 年から 2019 年までの論文のうち、以下に示す論文群である。

- (1) 撤回された論文群 (“Retraction” に分類された撤回論文群)
WoS_Retraction : 10494 件
- (2) 高被引用論文群 (インパクトファクターが高いジャーナルに掲載された高被引用論文群)
WoS_Nature Biotech : 9819 件
WoS_Nature Materials : 5049 件

3.3. 分析プロセス

本稿の分析ステップを以下に示す。3.2.項で示したそれぞれの論文群に対して、これらのステップで分析する。

Step 1. 論文の著者順に基づく研究者関係の設定
共著論文の著者順に基づき、研究者の関係を以下の通り定義し、エッジリストを作成する。

- (A) LA を始点とし、FA と MA を終点とするエッジで結ばれた関係
- (B) FA を終点とし、LA と MA を始点とするエッジで結ばれた関係
- (C) LA を始点とし、FA を終点とするエッジで結ばれた関係 (MA は無視する)
- (D) FA を始点とし、LA を終点とするエッジで結ばれた関係 (MA は無視する)

Step 2. 研究者の系譜ネットワークを構築

3.1.項で対象とした各論文群について、Step 1.で定義した (A)(B)(C)(D)のエッジリストを用いて、以下のような共著ネットワークを作成する。本稿ではこれを「系譜ネットワーク」と呼ぶ。

- (1) 無向共著ネットワーク
 - ① 全ての共著関係からなる重み有り無向ネットワーク
 - ② 全ての共著関係からなる重み無し無向ネットワーク
- (2) 有向共著ネットワーク
 - ③④⑤⑥ (A)(B)(C)(D)のそれぞれの共著関係から成る重み有り有向グラフ
 - ⑦⑧⑨⑩ (A)(B)(C)(D)のそれぞれについて、2つのノード間の有向エッジの重みの差分を新たなエッジリストを(A1)(B1)(C1)(D1)とし、各々の共著関係からなる重み有り単方向有向グラフ (なお、この場合、重みが逆方向で等しい場合は、エッジリストがなくなる)
 - ⑪⑫⑬⑭ (A1)(B1)(C1)(D1)のそれぞれの共著関係から成る重み無し有向グラフ

Step 3. 系譜ネットワークの特徴量および著者の特徴量の算出

Step 2.の系譜ネットワークの特徴量および著者の特徴量を算出する。各論文群に対して、論文または著者ごとに、以下のネットワーク特徴量、研究者間特徴量、研究者特徴量を算出する。

(1) ネットワーク特徴量

- 基本特徴量
無加重無向グラフにおいて、
 - ネットワーク密度
 - クラスタ数、クラスタサイズ、クラスタネットワーク密度
- 伸長度・拡張度
全加重、差分加重、差分無加重有向グラフにおいて、
 - 最大系譜伸長度 (Maximum Elongation) : あるノードから伸びている有向グラフの最大のパスの長さ
 - 最大系譜拡張度 (Maximum Expansion) : あるノードから到達できる有向グラフの全ての終端ノードの数
- 師弟度・同僚度
全加重有向グラフにおいて、
 - 師弟頻度率 : 全ての著者の組み合わせに対する、弱結合 (片方向結合) の数
 - 同僚頻度率 : 全ての著者の組み合わせに対する、強結合 (双方向結合) の数

(2) 研究者特徴量

- 基本特徴量
無加重無向グラフにおいて
 - 中心性 (度数中心性、近接中心性、媒介中心性、固有値中心性)
- 師弟度・同僚度
全加重有向グラフについて、
 - 師弟頻度率 : 全ての著者の組み合わせに対する、弱結合 (片方向結合) の数
 - 同僚頻度率 : 全ての著者の組み合わせに対する、強結合 (双方向結合) の数
 - 次数 (入次数、出次数、出入次数差分)
 - ゲートキーパー度 (入次数 x 出次数)
- 伸長度・拡張度
無加重無向グラフにおいて
 - 最大系譜伸長度 (Maximum Elongation)
 - 最大系譜拡張度 (Maximum Expansion)

Step 4. 特徴量の分析

Step 3.の系譜ネットワークの特徴量および著者の特徴量について、論文群間で比較する。

4. 結果

前述の 3.3 の条件に基づき、試行的分析を行った結果の一部を表 1 に示す。

表 1 より、例えば、以下のようなことが分かる。

- Nature Biotech, Nature Materials について、とも

表1 系譜ネットワークの特徴

		Nature Biotech		Nature Materials		WoS Retraction	
		伸長度	拡張度	伸長度	拡張度	伸長度	拡張度
ladg_weight	(A) 全加重	28	2,307	28	2,901	21	2,465
ladg_diff_weight	(A) 差分加重	18	1,892	29	2,315	22	2,320
ladg_diff_unweight	(A) 差分無加重	18	1,892	29	2,315	22	2,320
fadg_weight	(B) 全加重	11	640	23	1,217	22	3,345
fadg_diff_weight	(B) 差分加重	8	463	20	1,087	19	3,086
fadg_diff_unweight	(B) 差分無加重	8	463	20	1,087	19	3,086
l2fdg_weight	(C) 全加重	10	50	4	33	13	106
l2fdg_diff_weight	(C) 差分加重	9	40	4	33	13	106
l2fdg_diff_unweight	(C) 差分無加重	9	40	4	33	13	106
f2ldg_weight	(D) 全加重	10	32	4	16	13	172
f2ldg_diff_weight	(D) 差分加重	9	24	4	15	13	168
f2ldg_diff_unweight	(D) 差分無加重	9	24	4	15	13	168

に、LAを始点としFA/MAを終点とするネットワークの伸長度、拡張度が大きい。

- Retractionについても、Nature Biotech, Nature Materialsと同様の伸長度、拡張度が存在している。

5. 考察

5.1. 知識の系譜

今回の試行的分析の結果、Nature Biotech, Nature Materialsの論文群には系譜が存在することが分かった。また、Retractionの論文群にも、Nature Biotech, Nature Materialsのようなインパクトファクターの高い論文と同様に系譜が存在し、知識移転は、高被引用論文だけでなく、撤回論文においても発生していることが推察される。

5.2. 今後の対応

本稿で示したのは、前述の3.3の条件に基づいて試行的な分析を行った結果である。今後、以下のような対応を行う予定である。

- (1) 一般の論文との比較
本稿では、インパクトファクターの高いフォン分としてNature Biotech及びNature Materialsを取り上げ、また撤回された論文群と比較したが、これら以外に、被引用数が高くない論文群やランダムに抽出した論文群などの分析を行う。
- (2) 国・地域間の比較
今回分析した論文データには国・地域の情報が含まれており、その特徴を分析する。
- (3) 時間推移の分析
撤回論文は増加傾向にあり、増加の実態につい

て、年ごとまたは一定期間ごとの推移を分析する。

(4) 分野ごとの分析

今回は、Biotechnology分野及びMaterials分野について分析したが、今後他の分野についても分析を行う。著者順の慣習は分野によって異なるとされるが、その特徴についても分析する。

6. まとめ

本稿では、知識の継承と拡散の特徴を、学術論文の共著ネットワークをもとに、実証的に示す手法を提案した。具体的には、学術論文データベースにおいて被引用数の高い論文群、撤回された論文群、などの論文群について、それぞれ共著ネットワークを構築し、著者順情報を利用して研究者の系譜を作成した上で、論文群間で、系譜の伸長性・拡張性、研究者の師弟性・同僚性などの特徴を比較分析する。また、Nature Biotech誌、Nature Materials誌に掲載された論文群、及びWeb of Scienceで撤回された論文群の間で、系譜の伸長性・拡張性について試行的に分析した結果、Nature Biotech, Nature Materialsには系譜が存在すること、また、撤回された論文群にも、同様に系譜が存在し、知識移転が発生している可能性があることが分かった。今後さらに分析を進める予定である。

謝辞

本研究はJSPS 科研費(18H00840)の助成を受けたものである。本研究においては、政策研究大学院大学の隅蔵教授の協力を得た。感謝申し上げます。

参考文献

- [1] 文部科学省, 科学技術・イノベーション創出の活性化に関する法律, 2018, https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=420M60001000015
- [2] 内閣府, 統合イノベーション戦略 2019, 2019, https://www8.cao.go.jp/cstp/togo2019_honbun.pdf
- [3] 文部科学省, 研究活動における不正行為への対応等に関するガイドライン, 2014, https://www.mext.go.jp/a_menu/jinzai/fusei/index.htm
- [4] Steen, R. G., Retractions in the scientific literature: is the incidence of research fraud increasing?, *Journal of Medical Ethics*, Vol. 37, No. 4, pp. 249-253, 2011.
- [5] Fang, F. C., Steen, R. G., and Casadevall, A., Misconduct accounts for the majority of retracted scientific publications, *PNAS* 109 (42) 17028-17033, 2012.
- [6] Fanelli, D., Why Growing Retractions Are (Mostly) a Good Sign, *PLOS Medicine*, Vol. 10, Issue 12, 2013.
- [7] Wang, T., Xing, Q., Wang, H, Chen, W., Retracted Publications in the Biomedical Literature from Open Access Journals, *Sci Eng Ethics*, 25:855–868, 2019.
- [8] Campos-Varela, I., Ruano-Ravi, A., Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors, *Gac Sanit.*, Vol. 33, No. 4, pp. 356–360, 2019.
- [9] Kuroki, T., Ukawa, A., Repeating probability of authors with retracted scientific publications, *Accountability in Research*, Vol. 25, No. 4, pp.212-219, 2018.
- [1 0] D. Crane, *Invisible Colleges*, The University of Chicago (1972).
- [1 1] Thomas J. Allen, *Managing the Flow of Technology*, MIT Press (1979).
- [1 2] Leydesdorff, L., Rafols, I., A Global Map of Science Based on the ISI Subject Categories, *Journal of the American Society for Information Science and Technology*, 60(2), 348-362 (2009).
- [1 3] Mark E. J. Newman, Coauthorship networks and patterns of scientific collaboration, *Proceedings of the National Academy of Sciences*, 101(suppl. 1), 5200-5205 (2004).
- [1 4] 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).