JAIST Repository
https://dspace.jaist.ac.jp/

Title	リサーチフロントの形成と個々の研究者の研究パフォーマンスの関係
Author(s)	七丈，直弘
Citation	年次学術大会講演要旨集，24：625-629
Issue Date	2009-10-24
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/8709
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｅ０２

リサーチフロントの形成と個々の研究者の研究パフォーマンスの関係

○七丈　直弘（東京大学）

1. はじめに

経済発展の結果、先進国のみならず新興国においても、製造業から知識産業に代表されるようなサービス業へと産業重点領域のシフトが起きつつある。新興国での科学技術研究の進展は著しく、特に、中国では公刊される論文数が指数関数的に増加している(Zhou and Leydesdorff 2006)。このように世界規模で加速する研究開発競争を背景として、新しい研究領域を自ら創出し、学術的発見を土台として、世界に先駆けてイノベーションを普及させ市場化を通じた収益化を社会全体で実現していくことが、研究者と研究評価者に与えられた課題となっている。この実現に向け、新しい研究分野が生まれていくプロセスをモデル化することで、新たな研究領域の創生を予見できるようなシステムの提案が求められている。

de Solla Price(1965)は、その萌芽的な計量書誌分析の研究の中で、公刊される論文の中のごく少数が多数の論文から引用されるという現象を発見し、多くの論文から同時に引用（共引用）される少数の論文をリサーチフロントと呼んだ。この手法は萌芽的研究領域を同定するために利用できる可能性があることから、トムソン社(Thomson Scientific)は同社の書誌データベースESI(Essential Science Indicators)に含まれる論文を用い、高被引用論文間の共引用関係をダイアグラムとして可視化し、Science Watch 誌の中で公開してきた。日本でもリサーチフロントと同様の手法を用いて俯瞰的に研究領域を先導する特徴的な研究成果をピックアップするために用いようという取り組みが行われてきている(阪 彩香, 伊神 正貫 et al. 2008)。

このようなリサーチフロントが評価される背景には、先導的研究は旧来の学術領域の境界を超え、分野横断的な取り組みから生まれる傾向があるという現象が背景にある(Metzger and Zare 1999)。また、高い影響力を持った学際的研究に牽引され、多くの関連研究が蓄積されていくことで、学術領域として認識されていき、新たな研究分野が生まれたのだと人々は認識し、最終的に新しい論文誌の創刊につながる場合もある。実際、公刊されている学術論文誌の数は急速に増加[1]している。論文誌の数の増加の要因としては、研究論文の増加により、コミュニケーションを円滑に進めるためには、より細分化した領域で論文誌を編纂する必要が生じているという可能性と、既存の論文誌に収まらない新たな研究領域が創出されている可能性が考えられる(Noyons and van Raan 1998)。

だが、研究の学際性を測ることは難しい。ある特定の研究が学際的であるかを判断するための基準としては、それが掲載された論文誌に付与されたカテゴリの組み合わせをみるという方法がよく用いられている(van Leeuwen and Tijssen 2000; Rinia, van Leeuwen et al. 2002)。この手法では、論文誌単位での学際性しか測ることができないため、当該論文が引用している論文が掲載されている論文誌のカテゴリの組み合わせや、当該論文を引用している論文が掲載されている論文誌のカテゴリの組み合わせを考慮することによって発展させ、論文単位や研究者単位で学際性を定量化する手法も考えられている(Qin, Lancaster et al. 1997; Rinia, van Leeuwen et al. 2002)。このような指標は計算が比較的容易であるため、大集団に対して適用し、学術研究のダイナミクスのマクロ的把握ができるものの、学際性の判断が既存のカテゴリの区分（研究領域によってカテゴリの細分化に大きな差がある）に依存しているということ、さらにその区分が比較的粗いという問題を抱えており、細分化された研究領域における学際性が研究深化のダイナミクスに与える影響を仔細に分析するには適さない。トムソン社のリサーチフロントもこのカテゴリに属する。

これらに対し、Noyons ら(1998)は論文誌の分類に関する先験的な情報を用いずに、キーワードや論文分類コード(Journal Classification

[1] Ulrich's Periodical Directory では、1970 年代には 4,634 誌、1980 年代には 5,673 誌、1990 年代には 6,245 誌、2000 年代には 5,264 誌が新規の peer-review による学術誌として登録されている。

Codes)[2]の共出現関係のみから学術領域の構造とその時間的変化を導くことで、非経験論的に学術構造を分析できる可能性を示した。本研究では、共引用関係によって得られた論文間の関係性ネットワークをクラスタリング手法によってドメイン分割し、細分化された学術領域として捉える。このドメインに対して、Small が高被引用論文に対して適用した手法(Small 2006)を参考にしながら、時系列的変化についても可視化することで、特定学術領域内におけるミクロな領域間での知識の相互関係を明らかにする。また、研究者ごとに、研究論文のドメイン間分布（論文ポートフォリオ）は大きくことなっていることから、論文ポートフォリオの集中指数と研究パフォーマンスの関係についても、考察を試みる。

2. データ

本研究では、論文書誌情報として Thomson Scientific 社が作成している学術論文書誌データベースである SCI-EXPANDED を利用した。[3]

まず、分析対象となる学術領域を規定するために、以下のようにして論文サンプルの抽出を行った。最初に全文検索により「photocatal*」（語尾の*はワイルドカードであり、光触媒に関連した語に適合する）を文章中に含む論文を抽出した。その結果抽出された論文の集合 S と定義する。次に、引用情報を用いて S に含まれる論文を引用している論文を抽出し、この集合を C と定義する。S に含まれる論文は 6,992 本であり、C に含まれる論文は 25,651 本であった。後で説明されるように、C の論文が持つ引用情報によって S がクラスタリングされることになる。

また、研究パフォーマンスの比較を行うために、研究者ごとの論文数の集計を行った。S に属する論文を有する研究者は 6,763 名のうち，1 回以上引用を受けている論文を 5 本以上有する研究者を分析対象とした。このような条件を課すことにより，学生などのように派生的成果として論文を出版した者を排除することが可能となり 684 名の研究者が抽出された。このうち 9 名は、上記の研究者内での共著関係を持たなかったため[4]、研究者間の連携による効果を評価することができず、今回は分析対象外とした。最終的に、共著関係を有する 635 名の研究者が以下の分析の対象となった。

[2] 代表的なものとして、経済学における JEL コード、物理学における PACS コードなどがある。

[3] 1970 年から調査時点である 2004 年までに収録されたデータを対象とした。

[4]本研究では、研究者間のコラボレーションの影響を分析していくため、この 684 人の中における共著関係に着目する。

3. 手法・結果

分野の時系列的な発展を知るために S を対象としたクラスタリングを行う。なお、クラスターの時間的発展を知るために、直近の 5 年間に出版された論文を対象としてクラスタリングを行うこととした。具体的には、y 年のクラスタリング対象は，y-4 年から y 年に引用された S に含まれる論文の中で、その後 1 回以上他の論文から引用を受けたものとなる。各年のクラスタリング対象となった論文数の変化を図 1 に示す。

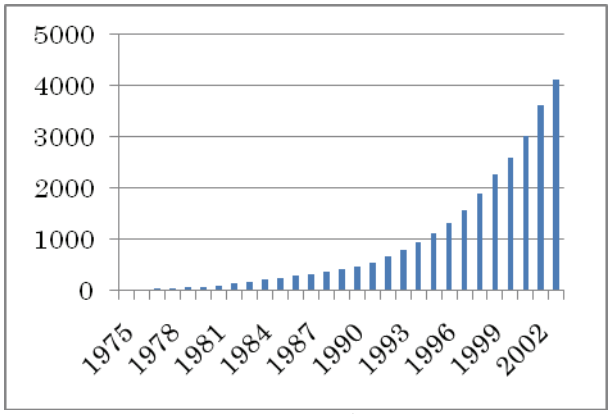

図 1 クラスタリングの対象となった論文数の時間的推移

次にクラスタリングの手法について述べる。

書誌結合分析や語分析で扱うデータは，論文の参考文献リストや本文といった，論文が出版された時点で確定されるデータであり，どの時点で分析を行っても分析結果は変わらない。一方，共引用分析では，新しく引用を受けるたびに分析結果がダイナミックに変化する。この性質は，知識構造の進化プロセスを観察するのに有効であると考える。共引用分析は，最新の論文が分析対象に含まれないという短所があるが，知識構造の進化プロセスを観察することを目的とする場合，最新の知識構造は必ずしも必要ない。

また，共引用分析による分析結果は，当該分野の知識構造に対する研究者の分析時点での認識といえる。研究者の研究戦略を分析するためには，研究者の意思決定には当時の認識が影響を与えているので，過去にさかのぼって共引用分析を行って，当該分野の知識構造に対する研究者の当時の認識を捉える必要がある。そこで、以下に述べるように重みつき Newman 法によって分析を行うこととした。

まず、y年の論文ネットワークの重み付き隣接行列 $\mathbf{R}(y)$ は，以下のように共引用回数を用いて表現されている。

$$R(y) = \begin{bmatrix} \ddots & & \\ & C_{ij}(y) & \\ & & \ddots \end{bmatrix}$$

$C_{ij}(y)$ = 論文 i と論文 j が $y-4$ 年から y 年までに同一の論文によって引用された件数

ただし，$\mathbf{R}(y)$の対角成分はすべて 0 とした。

クラスタリング手法は，重み付き Newman 法(榊，松尾 et al. 2006)を用いるが、これはクラスタリング問題を以下によって定義される量 Q の最小化問題ととらえるものである。

$$Q = \sum_i (e_{ii} - a_i^2)$$

$$e_{ij} = \frac{\text{クラスター}i\text{と}j\text{との間のエッジの重みの和}}{\text{全エッジの重みの和}}$$

$$a_i = \sum_j e_{ij} = \frac{\text{クラスター}i\text{の頂点と結びついたエッジの重みの和}}{\text{全エッジの重みの和}}$$

このようにしてクラスタリングを行うことで得られたクラスター分割の結果を可視化したものを図 2 に示す。

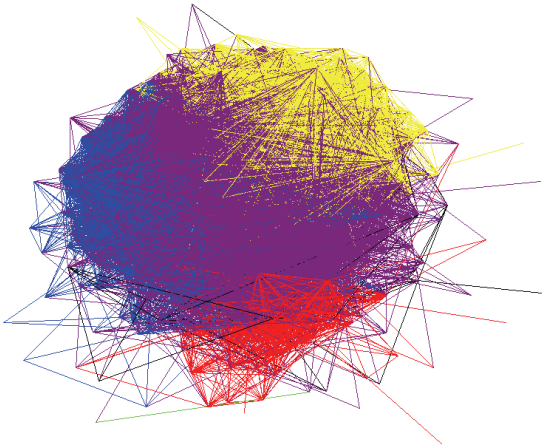

図 2　2003 年の共引用ネットワークのクラスタリング結果
論文は点で示され、論文間を結ぶ辺の色は所属するクラスターを意味する。

クラスターの時系列変化を図 3 に示す。図において，赤い丸は各年の各クラスターを示し[5]，丸の大きさはそのクラスターに属する論文数の大きさを意味する。丸と丸を結ぶ線は，各クラスターに属する論文が次の年にどのクラスターに属するかを示しており，線の太さはその論文数を示している。この手法は，(Small 2007) による。

[5] 視認性確保のため，各年，所属論文数の上位 10 クラスターのみを描画している.

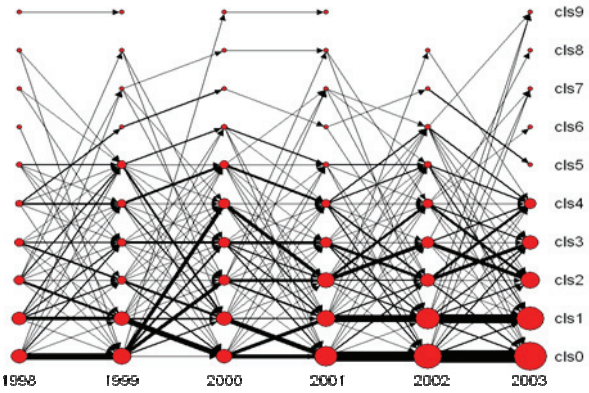

図 3　クラスター間の継承関係の時系列的変化

既にクラスターの時系列変化を図示したが，クラスターは複雑に融合したり分裂したりしていることが分かった。そこで，本項では，知識の融合と分裂を表すマクロ的指標としてハーフィンダール指数を導入し，この値の時系列変化を追う。クラスター数をk，クラスターjに属する論文数をa_iとした場合のハーフィンダール指数HHIは以下の式で与えられ，この値が大きいほど知識構造が集中している（少数の大きなクラスターで知識構造が形成されている）といえ，この値が小さいほど知識構造が分散している（多数の小さなクラスターで知識構造が形成されている）といえる。

$$HHI = \frac{\sum_{i=1}^{k} a_i^2}{\left(\sum_{i=1}^{k} a_i\right)^2}$$

そして，HHIの値を年ごとに算出した結果，HHIの時系列変化は図 4 のようになった。

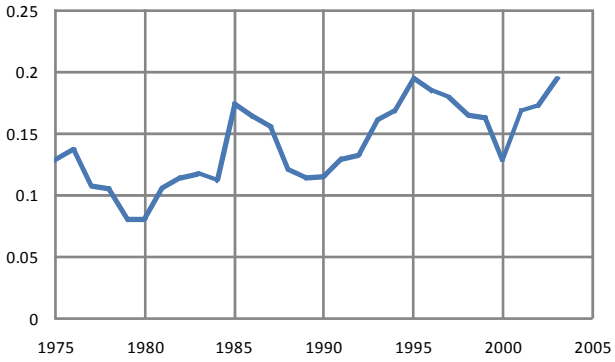

図 4　ハーフィンダール指数（知識構造の集中度）の時系列変化

図 4 から，ハーフィンダール指数は下降と上昇を繰り返していることが認識される。この現象は，知識の細分化と融合化が繰り返し起きている結果であると推察される。

次に、学術領域の細領域の各々に対する個々の研究者の貢献をベクトルとしてとらえ、研究ポートフォリオとして捉える。このような研究ポートフォリオが研究生産性に与える影響について調べていくこととする。

分析の対象とされた 635 名の研究者に対して、年（1975 年から 2003 年）ごとに論文ポートフォリオを作成する。論文ポートフォリオは，各クラスターに属する論文をどれだけ持っているかを示す。

研究者 i の y 年の論文ポートフォリオは，前章で作成した y 年の各論文クラスターに含まれる論文のうち，著者欄に i が含まれる論文の本数を並べたベクトルで表す。例えば，y 年のクラスター数が 3 で，研究者 i の論文が 3 つのクラスターにそれぞれ，3，2，0 本ずつ含まれているとした場合，研究者 i の y 年の論文ポートフォリオベクトル $\boldsymbol{P}_i(y)$を，

$$\boldsymbol{P}_i(y) = (3,2,0)$$

と表すこととする。

また，各研究者の研究内容の散らばり度合いを示す指標として，1－ハーフィンダール指数を用いる。論文ポートフォリオベクトルの成分を $P_{i,k}$ とすると，研究者 i の y 年のこの指数は，

$$hi_i = 1 - \sum_k \frac{p_{i,k}^2}{\boldsymbol{P}_i^2}$$

と表わすことができる。この値が大きいほど，研究内容が幅広いことを示す。

研究内容の遠い研究者と共同研究を行うことは，新たな知識の獲得に有効であり，研究内容の近い研究者と共同研究を行うことは，知識の深化に有効であると考えられる。そこで，共著者間の研究内容の遠近が共著以降の研究パフォーマンスに与える影響をみるため，共著者間の論文ポートフォリオの類似度を算出する。y 年に研究者 i と j とが共著論文を出した場合の共著者間の論文ポートフォリオの類似度 $S_{ij}(y)$は，y－1 年時の 2 者の論文ポートフォリオの余弦で表す。この値が大きいほど，2 者の研究内容は近いことを示す。

$$S_{ij}(y) = \frac{\boldsymbol{P}_i(y-1) \cdot \boldsymbol{P}_j(y-1)}{|\boldsymbol{P}_i(y-1)||\boldsymbol{P}_j(y-1)|}$$

また，研究者によっては 1 年間に複数の研究者と共著論文を出す者もいる。その場合，共著者の中で最も研究内容の遠い者との論文ポートフォリオの類似度 $cosMin_i(y)$をパネルデータ分析で用いる。

$$cosMin_i(y) = \min_j S_{ij}(y)$$

$cosMin_i(y)$の値が小さいほど，離れた研究者と共同研究を行っているといえる。表 1 に被説明変数を y+1 年における論文数とし、説明変数として過去 5 年間における論文数、総被引用数、共著論文数、共著者数、論文ポートフォリオの Herfindahl 指数、共著者との研究ポートフォリオの相違度を用い、パネル回帰分析を行った結果を示す。

表 1 パネル回帰分析による推計結果

被説明変数: 論文数（y+1）	
説明変数:	
論文数	0.186***
	(0.022)
総被引用数	-0.001***
	(0.000)
共著論文数	-0.047*
	(0.025)
共著研究者数	0.031***
	(0.008)
hi	-0.349**
	(0.159)
cosineMin	-0.170**
	(0.085)
定数高	0.467***
	(0.093)
χ^2	1066.61
N	4012

パネル回帰分析の結果、自らの研究領域が幅広いこと自体は論文数に負の影響を与え、研究内容が異なる研究者と共同研究を行うことが論文数に正の影響を与えることが判明した。この結果は Qin(1997)と整合的である。

4. 終わりに

本研究によって、特定研究領域におけるドメインのミクロ構造を共引用分析によって合理的に行えることが判明した。また、Small(2006)の手法を適用することで、ドメイン間のダイナミクスを可視化することができた。また、得られたドメイン構造から、研究者の研究ポートフォリオを定義し、研究パフォーマンスとの比較をパネル分析によって行った。手法上の限界として、今回の分析では光触媒という限定された領域内での分析

であるため、観測の対象外となった関連領域との間の相互作用（外部性）については考慮していない。また、光触媒という分野そのものが、研究の進展の結果確立したものであって、対象とした論文の中の比較的古いものは、外部性の影響を大きく受けていることが予想される。この問題を克服するためには、query expansion などの手法で、できるだけ関連領域を広く含めるように工夫する必要がある。また、リサーチフロントが対象とするようなマクロレベルでのダイナミクスと、本論文が対象としたミクロレベルでのダイナミクスとの連関についても今後考察したい。

謝　辞

本研究は、科学技術研究費補助金若手(B) #21730290 による支援の成果が含まれる。また、加毛 誠氏、馬場靖憲教授から研究内容についてアドバイスを受けた。

引用文献

Metzger, N. and R. N. Zare (1999). "Science policy - Interdisciplinary research: From belief to reality." Science **283**(5402): 642-643.

Noyons, E. C. M. and A. F. J. van Raan (1998). "Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research." Journal of the american society for information science **49**: 68-81.

Price, D. J. D. (1965). "NETWORKS OF SCIENTIFIC PAPERS." Science **149**(3683): 510-&.

Qin, J., F. W. Lancaster, et al. (1997). "Types and levels of collaboration in interdisciplinary research in the sciences." Journal Of The American Society For Information Science **48**(10): 893-916.

Rinia, E., T. van Leeuwen, et al. (2002). "Measuring knowledge transfer between fields of science." Scientometrics **54**(3): 347-362.

Small, H. (2006). "Tracking and predicting growth areas in science." Scientometrics **68**(3): 595-610.

Small, H. (2007). "Citation Structure of an Emerging Research Area: Organic Thin Film." Proceedings of ISSI: 718-725.

van Leeuwen, T. and R. Tijssen (2000). "Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows." Research Evaluation **9**: 183-187.

Zhou, P. and L. Leydesdorff (2006). "The emergence of China as a leading nation in science." Research Policy **35**(1): 83-104.

阪 彩香, 伊神 正貫, et al. (2008). サイエンスマップ 2006―論文データベース分析（2001 年から 2006 年）による注目される研究領域の動向調査―, 科学技術政策研究所.

榊, 剛., 豊. 松尾, et al. (2006). "制約付きクラスタリングを用いた論文分類." The 20th Annual Conference of the Japanese Society for Artificial Intelligence.