

共著相手の特性が研究者の論文生産性に及ぼす影響に関する分析

芳鐘 冬樹 (大学評価・学位授与機構
野澤 孝之 (大学評価・学位授与機構
渋井 進 (大学評価・学位授与機構

評価研究部, E-Mail: fuyuki@niad.ac.jp)
評価研究部, E-Mail: nozawa@niad.ac.jp)
評価研究部, E-Mail: shibui@niad.ac.jp)

Abstract

共著相手の特性が研究者のその後の論文生産性に及ぼす影響についての知見を得ることを目的として、計算機科学分野の新規参入者を対象に、共著相手の過去の研究活動実績(発表論文数・ネットワーク上の重要度) と、新規参入者のその後の発表論文数との関連を調べた。結果、共著相手の過去実績と、新規参入者のその後の論文数との相関は認められないこと、ただし、参入後も引き続き 1 編以上論文を発表している研究者と、参入時の 1 編だけで終わってしまっている研究者を比較すると、前者の方が、共著相手の過去実績が高いことが明らかになった。

1 はじめに

共同研究者間の特性の相関を調査した研究は、論文生産性に注目したもの、所属機関に注目したものなど多数存在するが (e.g., Kretschmer, 1994; 1997; Kretschmer & Gupta, 1998; Kundra & Kretschmer, 1999; Bahr & Zemon, 2000; 安田, 2004) , そこで見られているのは同一の時期の特性であり、時間が経過した後の状況との関係という観点から、共同研究者間の特性の関連性に関する定量的な分析を行った研究は、これまであまりなされていない。本研究は、共著ネットワークに着目して、共著相手の論文生産・ネットワーク上の特性が研究者のその後の論文生産性に及ぼす影響についての知見を得ることを目的に、両者の関連の有無を明らかにするものである。

研究の高度化・多様化が進む今日、研究協力の重要性はますます強くなってきており、共同研究を通じたネットワーク形成・情報伝達が研究活動全体に与える影響は非常に大きい。特に、未だ研究のスタイルが十分確立していない段階の研究者の場合、誰の共同研究者になるかによって、その後の研究者としての方向性が大きく変わるものと考えられる。ある研究者がある時期に共同研究を行った共同研究者の特性と、その研究者のその後の研究活動の状況との関係についての傾向を明らかにできれば、誰を共同研究者とするか (例えば、どの指導教員の下につくか) を決める際の参考になりうる。また、学生 (大学院生) の教育を行う過程で/行った成果として、学生と指導教員が共著で論文を発表することが一般的な分野では、教育効果 (研究者養成を目的

とした教育の効果) の予測にも活用できるという点でも、本研究は意義を持つと考える。

2 分析対象およびデータ

計算機科学分野を分析の対象とした。共同研究が活発な分野であり、ネットワークを考慮する必要性が大きいことが、計算機科学を分析対象に選んだ理由である。前述のように、共同研究者が与える影響は、初期の段階の研究者に対して特に大きく現れるものと考え、この分野のコアジャーナルに第 1 著者として初めて論文を発表した、すなわち、新規にこの分野への本格的な参入を果たした研究者を対象に、彼のその後の生産性と、新規参入時の共著相手のそれ以前の研究活動実績の関連を調べることにした。また、著者クレジットの最後に名前を連ねる最終著者は、その研究の指導者・監修者としての役割を担う者であることが多く¹、共同研究者に与える影響は大きいと考え、新規参入時の共著相手の中でも、特に、最終著者に注目する。

分析対象とする研究者の論文発表状況、および共著ネットワークを観察するための情報源として、Thomson ISI の SCI (Science Citation Index) の Disc 版を用いる。SCI は、質的な基準²を満たす雑誌のみを収載するという方針を採っている (根岸ら, 2004) 。その選定を信用して、SCI に収載され、主題カテゴリ別収載誌一覧³で、‘computer science, theory & methods’ に含まれている雑誌⁴を、計算機科学分野のコアジャーナルと見なすことにした⁵。それらの雑誌に 1998 年に論文 (‘Document type’ が ‘Article’ の文献) を発表しており、かつ、それより過去 7 年間 (1991 年から 1997 年まで) に 1 編も論文を発表していない研究者を、1998 年時点の新規参入者と見なす。対象がある程度均質になるよう、初出の 1998 年に発表している論文が、当人が第 1 著者である共著論文 1 編のみの研究者に限る。彼らの 1998 年より後の論文発表状況と、彼らの共

¹ 計算機科学分野の共著ネットワークにおいても、その特徴は観察されている (Yoshikane ら, 2006) 。

² <http://www.isinet.com/selection/>

³ <http://www.thomsonscientific.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=K>

⁴ *Journal of Algorithms* など 37 誌が含まれる。

⁵ 計算機科学分野は応用領域も盛んであるが、その中核は理論領域にあると考え、今回は ‘theory & methods’ を対象としている。

著相手の 1998 年より前の実績を把握するために、1998 年の前後それぞれ 7 年間をとり、1991 年から 2005 年までの、それらの雑誌に掲載されている文献のデータを SCI から抽出した。

書誌データベースを利用して著者の集計を行う場合、同姓同名や表記上のゆれの処理(名寄せ)が必要になる。所属機関などの著者識別の手がかりになる情報を利用する方法や、ミドルネームの有無の統一を行う方法などがある(e.g., Hayashi & Tomizawa, 2006)。SCI は、各著者と所属機関の 1 対 1 の対応が取られる形式にはなっておらず、また、1 つの著者が複数の機関に所属しているケースや、そもそも所属機関の情報がないケースもあり、著者識別の手がかりとするには不完全である。高頻度で出現する著者については、手作業でチェックを行うが、それ以外については、全体的な傾向を見るにおいては、大きな影響はなく誤差の範囲内であると考え、所属組織の情報に基づく処理は加えないことにした。ただし、表記上のゆれについては、ミドルネームのイニシャルの有無と、大文字小文字のゆれを統合した⁶。

NP	TA	DA	A_{av}	P_{av}	DA_{new}
29820	69240	34374	2.32	2.01	641

Table 1: データの基本的数量

表 1 にデータの基本的数量を示す。NP は論文の総数を、TA は延べ著者数を、DA は異なり著者数(名寄せ済み)を、 A_{av} は 1 論文あたりの平均著者数(= TA/NP) を、 P_{av} は 1 著者あたりの平均発表論文数(= TA/DA) を、それぞれ表している。前述の条件を満たす 1998 年時点の新規参入者は、641 人存在する(DA_{new})。

3 分析手法

3.1 指標

対象とする研究者の参入後の状況を見る指標として、初出年より後の 7 年間(1999 年から 2005 年まで)の発表論文数を用いる。完全計数(CMP)の他、貢献の分量を考慮した活性度を見るために、著者数で規格化する調整計数(ADJ)を、研究の代表者としての活性度を見るために⁷、第 1 著者である論文だけを数える第 1 著者計数(FST)を、それぞれ用いて発表論文数を算出する。一方、参入時の共著相手(最終著者)の過去実績を見る指標とし

⁶ 2179 の異表記を統合した。ミドルネームの表記がある著者となし著者が別人である可能性もあるが、ほとんどの場合において同一人物であるものと仮定している。ただし、ミドルネームが異なる複数の著者が存在する場合は、対応を一意に定めることができなため、処理は加えず、すべて別人として集計している。

⁷ 第 1 著者は代表者として研究の設計を行う者であり、それ以外の共著者とは異なる特別な役割を担っている、という前提に基づく。計算機科学分野のガイドライン(Zobel, 1999)にも、そうした記載があり、ある程度は前提は妥当と考える。

ては、過去 7 年間(1991 年から 1997 年まで)の発表論文数、および共著ネットワーク上の重要度を用いる。完全計数(SCMP)、調整計数(SADJ)、第 1 著者計数(SFST)の他、指導者・監修者としての実績を見るために、最終著者である共著論文の数(SLST)も見る。

ネットワーク上の重要度に関しては、ノウハウや人脈などは、間接的な関係を介しても伝わる、したがってネットワークの大域的な構造も考慮すべき、と考える立場をとる。第 1 著者を到達点、それ以外の共著者を出発点とする有向グラフ、かつ、共著関係の強度を考慮に入れた重み付きグラフを想定し、直接的な関係のみを考慮する指標として、入次数 SD_{in} と出次数 SD_{out} を、間接的な関係まで反映させる指標として、次に述べる Yoshikane ら(2006)の SC_l と SC_f を用いる。 SD_{in} と SC_l が代表者(第 1 著者)としての重要度、 SD_{out} と SC_f が支援者(その他の共著者)としての重要度に対応する。

$$SC_l(n_i) = \sum_{j=1}^g a_{ji} SC_f(n_j) \quad (1)$$

$$SC_f(n_i) = \sum_{j=1}^g a_{ij} SC_l(n_j) \quad (2)$$

ここで、 g は研究者の数を指す。また、 a_{ij} はネットワークの隣接行列 A の成分を指し、研究者 n_j から研究者 n_i に向けた結合の強度、ここでは n_i が第 1 著者になり n_j と共著した頻度、を値としてとる⁸。ここで置いている仮定は、「重要な代表者を支えている支援者は、支援者として重要な役割を担っており、重要な支援者をまとめている代表者は、代表者として重要な役割を担っている」という相互の依存関係である。式(1)(2)で、再帰的な代入を繰り返すことにより、ネットワークの大域的な構造が各々の研究者の重要度に反映する。HITS アルゴリズム(Kleinberg, 1998)と同様のステップで、代入とベクトルの規格化を再帰的に繰り返すことにより(10 回ループ)、 $SC_l(n_i)$ と $SC_f(n_i)$ を計算する。

3.2 関連性の分析

共著相手の過去の研究活動実績と、新規参入者のその後の論文発表状況との関連性を、以下の 2 つの観点から調べる。

1. 新規参入者のその後の発表論文の多寡と、共著相手の過去実績の高低との間に相関はないか
2. 参入時の 1 編のみで終わってしまっている者と、その後少なくとも 1 編は発表しており、研究者として論文生産を継続している者として、共著相手の過去実績に差はないか

⁸ 共著者が多いほど、1 人 1 人の関係は希薄になると仮定し、芳鐘 & 野澤(2006)と同じく、著者数で規格化した共著頻度を使う。

例えば、教育効果の予測への応用において、論文の多寡というよりも(数は少なくとも質は高いかもしれないので)、その後も研究者として論文生産を継続できるかどうかに関心を置く立場も想定されるため、2. に関して明らかにすることにも意味があると考える。

1. については、新規参入者のその後の発表論文数を表す指標群 $[CMP, ADJ, FST]$ と、共著相手の過去実績を表す指標群 $[SCMP, SADJ, SFST, SLST, SD_{in}, SD_{out}, SC_l, SC_f]$ の各組み合わせについて、相関係数を算出する。ただし、これらの指標は、想定される分布はベキ則分布であり⁹、正規分布ではないため、スピアマンの順位相関係数を用いる。2. については、 $CMP = 0$ の新規参入者と、 $CMP \geq 1$ の新規参入者、それぞれ、共著相手の $[SCMP, SADJ, SFST, SLST, SD_{in}, SD_{out}, SC_l, SC_f]$ の平均値を求めて比較し、ウィルコクソンの順位和検定により、その差の検定を行う。

4 分析結果

新規参入者 641 人の発表論文数 $[CMP, ADJ, FST]$ と、共著相手の過去実績 $[SCMP, SADJ, SFST, SLST, SD_{in}, SD_{out}, SC_l, SC_f]$ の相関係数を表 2 に示した。第 1 著者としての論文数(代表者としての活性度) FST と相対的に高い関連がある共著相手の過去実績は、同じく第 1 著者としての論文数 $SFST$ である、一方、第 1 著者が否かを区別しない論文数 CMP, ADJ と相対的に高い関連があるのは、共著相手の支援者としての大域ネットワーク上の重要度 SC_f である、といった傾向は認められるものの、すべての組み合わせで、相関係数は 0.1 未満であり、新規参入者の発表論文数と共著相手の過去実績の間に、ほとんど相関はなかった¹⁰。また、指標の値の散布図からも関連の傾向は読み取られなかった(一例として、図 1 に、その後の論文数(完全計数) CMP と、共著相手の過去論文数(完全計数) $SCMP$ の散布図を示した)。

	CMP	ADJ	FST
$SCMP$	0.085	0.083	0.079
$SADJ$	0.083	0.082	0.079
$SFST$	0.046	0.049	0.084
$SLST$	0.046	0.041	0.020
SD_{in}	0.020	0.020	0.050
SD_{out}	0.072	0.067	0.050
SC_l	0.021	0.020	0.051
SC_f	0.090	0.084	0.073

Table 2: 参入後の状況と共著相手の特性の相関

⁹ Lotka (1926) の法則などがそれを表している。

¹⁰ その後(1999年-2005年)の状況同士では、新規参入者と共著相手の論文発表状況は、ある程度(0.3前後)の相関が見られた。その後も共同研究を続けており両者が連動するケースによる寄与と考えられる。

	CMP	ADJ	FST
$SCMP$	0.020	0.007	0.051
$SADJ$	0.005	0.002	0.046
$SFST$	0.011	0.031	0.106
$SLST$	0.040	0.008	-0.016
SD_{in}	0.016	0.015	0.094
SD_{out}	0.047	0.011	0.023
SC_l	0.015	0.013	0.094
SC_f	0.077	0.034	0.061

Table 3: 参入後の状況と共著相手の特性の相関 ($CMP \geq 1$ の新規参入者のみ)

表 3 は、参入後、少なくとも 1 編は論文を発表している ($CMP \geq 1$) 研究者 232 人に対象を限定して、彼らの発表論文数と共著相手の過去実績との相関係数を示したものである。全体を対象にしたとき(表 2)よりも、さらに相関係数が低いという結果が得られたことから、共著相手の過去実績は、むしろ、新規参入者がその後 1 編も発表しないか(できないか)どうかに関連している可能性がある。

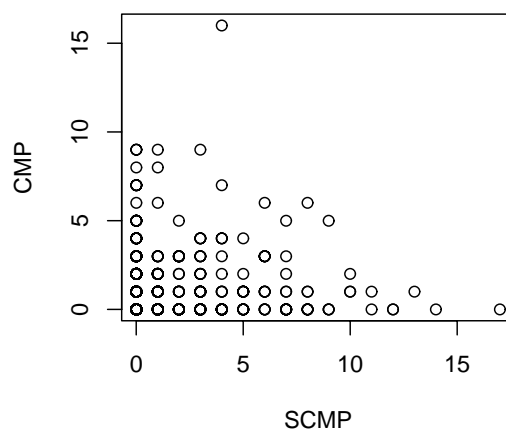


Figure 1: 参入後の発表論文数と共著相手の発表論文数

次に、参入時の 1 編のみで終わってしまっているグループ 409 人 ($CMP = 0$) と、その後、少なくとも 1 編は発表しており、論文生産を継続しているグループ 232 人 ($CMP \geq 1$) に分けて、それぞれ共著相手の過去実績を調べた。各指標について平均値を求めた結果が表 4 である。 SC_l 以外は、すべて、論文生産を継続しているグループの共著相手の方が大きい (SC_f は約 2.8 倍、その他は 1.3 倍前後) ことが確認できる。特に、第 1 著者が否かを区別しない論文数 $SCMP, SADJ$ と、支援者としての大域ネットワーク上の重要度 SC_f に関しては、1 編の

みで終わってしまっているグループの共著相手との間に有意差が見られる ($p < .05$)

支援者としての活性度,あるいは重要度という点では, $SLST$, SD_{in} , SC_f は共通しているが, 2グループ間の有意差があったのは, 上に述べたとおり, それらのうち SC_f だけであった。 SC_f は, 他の2つと異なり, 支援相手の重要度まで反映させる指標であり, 「研究代表者として活躍する研究者を支援してきた実績」を表すものと言える。その点に関する過去実績(に基づくノウハウや人脈など)が共著相手のその後の論文生産に影響を及ぼしている可能性が示唆される。

	$CMP \geq 1$	$CMP = 0$
$SCMP$	*1.51	1.17
$SADJ$	*0.63	0.52
$SFST$	0.49	0.38
$SLST$	0.70	0.59
SD_{in}	0.59	0.42
SD_{out}	0.88	0.65
$SC_i(\times 10^7)$	2.04	3.50
$SC_f(\times 10^6)$	*3.99	1.43

*は, 有意差あり ($p < .05$)

Table 4: 継続の有無と共著相手の特性

5 おわりに

共著相手が及ぼす影響についての知見を得ることを目的に, 共著相手の特性と新規参入者のその後の論文生産性の関連を調べた。共著相手の過去実績と新規参入者の論文数は無相関に近いが, 参入後も引き続き論文を發表しているか, 参入時の1編だけで終わってしまっているかという点においては, 共著相手の過去実績との関連が認められた。共同研究者のノウハウや人脈などに接する機会を得ることは, 研究者として論文生産を続ける術を身に付けることに, ある程度繋がるが, 活発に活躍できるか否かは, 別の要因の影響が大きく, 共同研究者の特性にほとんど因らない, という推測も成り立つだろう。ただし, 今回新規参入者と見なした中には, もともと継続の意志はなく, 例えば周辺領域から一時的に発表の場を求めて現れただけの研究者なども含まれると考えられる。それらをうまく分別できれば, 本来見たい新規参入者の特徴をより明確に観察できるかもしれない。また, 今回は論文生産性だけを見たが, 今後は, ネットワーク形成や論文のトピックといった面での関連性も見ること, 共同研究者間の影響関係についてより詳細な分析を進めたい。

謝辞

本研究の一部は科学研究費補助金若手研究(B)の助成によるものであり(研究課題番号18700244),ここに謝意を表します。

References

- Bahr, A. H. and Zemon, M. (2000) "Collaborative authorship in the journal literature: perspectives for academic librarians who wish to publish," *College & Research Libraries*, vol. 61, no. 5, p. 410-419.
- Hayashi, T. and Tomizawa, H. (2006) "Restructuring the Japanese national research system and its effect on performance," *Scientometrics*, vol. 68, no. 2. (to appear)
- Kleinberg, J. M. (1998) "Authoritative sources in a hyperlinked environment," *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, p. 668-677.
- Kretschmer, H. (1994) "Coauthorship networks of invisible-colleges and institutionalized communities," *Scientometrics*, vol. 30, no. 1, p. 363-369.
- Kretschmer, H. (1997) "Patterns of behaviour in coauthorship networks of invisible colleges," *Scientometrics*, vol. 40, no. 3, p. 579-591.
- Kretschmer, H. and Gupta, B. M. (1998) "Collaboration patterns in theoretical population genetics," *Scientometrics*, vol. 43, no. 3, p. 455-462.
- Kundra, R. and Kretschmer, H. (1999) "A new model of scientific collaboration Part 2: collaboration patterns in Indian medicine," *Scientometrics*, vol. 46, no. 3, p. 519-528.
- Lotka, A. J. (1926) "The frequency distribution of scientific productivity," *Journal of the Washington Academy of Sciences*, vol. 16, no. 12, p. 317-323.
- 根岸 正光・Ann S. Okerson・伊藤 義人・Raym Crow・佐藤 寛子・James Testa・安達 淳・土屋 俊・早瀬 均 (2004) 『電子図書館と電子ジャーナル: 学術コミュニケーションはどう変わるか (情報学シリーズ, 8)』, 丸善: 東京, 157p.
- 安田 雪 (2004) 『人脈づくりの科学』, 日本経済新聞社: 東京, 253p.
- Yoshikane, F., Nozawa, T. and Tsuji, K. (2006) "Comparative analysis of co-authorship networks considering authors' roles in collaboration: differences between the theoretical and application areas," *Scientometrics*, vol. 68, no. 3, p. 643-655.
- 芳鐘 冬樹・野澤 孝之 (2006) "著者の役割を考慮した共著ネットワークの比較分析: HITS アルゴリズムに基づく手法の改善," 『大学の諸活動に関する測定指標の調査研究報告書』, 大学評価・学位授与機構: 東京, p. 131-143.
- Zobel, J. (1999), *Guideline on Research Practice in Computer Science*. Retrieved September 14, 2006 from: <http://goanna.cs.rmit.edu.au/~jz/conduct.html>.