

研究者の論文生産に関わる諸属性の相関に関する分析 ：論文生産性に影響を与える要因の把握に向けて

芳鐘 冬樹 (大学評価・学位授与機構 評価研究部, E-Mail: fuyuki@niad.ac.jp)

Abstract

研究者のどの属性が、本人や共著相手の論文生産性と高い関連性があるかを特定することを目的に、属性間の相関について調査した。同時期の発表論文数に関しては、高い相関を示す指標も存在したが、後の発表論文数に関しては、本人のものにも、共著者のものにも、特に高い相関を示す指標はなかった。ただし、自身が第1著者であるときの共著者に限定した場合、ネットワークの大域的構造を反映した代表者としての重要度が、共著者のその後の発表論文数と、比較的高い相関を持つことが明らかになった。

1 はじめに

本研究は、共著ネットワーク(研究協力ネットワーク)の観察に基づき、研究者個人の生産性とネットワーク上の重要度について、本人と共著相手の状況の相関や、ある時点とその後との状況の相関という視点も含め、論文生産の状況に関する属性間の関連性についての傾向を明らかにすることを目的とする。特に、論文生産性との相関を中心に分析する。それらの属性間の関連性を把握できれば、対象分野における知識生産活動の特徴や、論文生産性に影響を与える要因についての知見を得ることができる。実際的な応用面でも、尺度の相関・指標の代替可能性を明らかにすることを通じて、研究評価の効率化の是非に関する示唆を与えたり、また、研究協力のパートナーを選択する際の参考にもなりうるといった点で有用である。

共同研究者間の属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(e.g., Kretschmer, 1994; 1997; Kretschmer & Gupta, 1998; Kundra & Kretschmer, 1999; Bahr & Zemon, 2000)。また、研究者個人の論文の発表数や引用率と、共著ネットワーク上の重要度が正の相関を示すことを報告した研究もある(安田, 2004)。しかしながら、それらの中で見られているのは同一の時期の属性のみであり、時間が経過した後の状況との関係まで含めて、研究者の属性間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。Yoshikaneら(2007)は、新規参入の研究者を対象に、彼らの共著者の過去実績と、彼ら自身の論文生産性との相関に関する分析を行っているが、本研究では、それを拡張し、ある時期の研究者の諸属性と、

(1) 本人の同時期の生産性、(2) 共著者の同時期の生産性、(3) 本人のその後の生産性、(4) 共著者のその後の生産性、それぞれについて、相関の有無を調べる。

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。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、SCI (Science Citation Index) を用いる。SCIに収載され、主題カテゴリ別収録誌一覧¹で、'computer science, theory & methods'に含まれている雑誌²を、計算機科学分野のものを見出すことにした。それらの雑誌に1996年から2000年までの間に、少なくとも1編、論文('Document type'が'Article'の文献)を共著で発表している研究者を、分析の対象とする。共同研究のスタイルをとる研究者に注目し、共著者の属性も含めて、研究者の属性間の関連性を調べたいため、共著論文の著者に対象を限定している。彼ら、および彼らの共著者の、1996年から2000年までの状況(論文生産性とネットワーク上の重要度)と、その後の2001年から2005年までの状況を把握するために、1996年から2005年まで、計10年間の、上述のコアジャーナル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。しかし、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。

表1にデータの基本的な数量を示す。NPは論文の

¹ <http://www.thomsonscientific.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=K>

² *Journal of Algorithms* など 37誌が含まれる。

	<i>NP</i>	<i>TA</i>	<i>DA</i>	A_{av}	P_{av}
'96-'00	10181	23162	14483	2.28	1.60
'01-'05	12772	31273	19792	2.45	1.58

Table 1: データの基本的数量

総数を, *TA* は延べ著者数を, *DA* は異なり著者数を, 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。

1996 年 ~ 2000 年に現れる著者 14483 人 ($= DA$ ('96-'00)) のうち, 今回, 分析対象とする著者, すなわち少なくとも 1 編共著論文を発表している著者は 13059 人存在した。90% 以上が共著論文の著者ということになり, 計算機科学分野が共同研究主体の分野であることを確認できる。また, それら 13059 人の著者のうち, 第 1 著者として少なくとも 1 編論文を発表している著者は 5518 人存在した。

3 分析手法

3.1 指標

データを前半 (1996 年 ~ 2000 年) と後半 (2001 ~ 2005 年) に分け, 「前半における研究者の論文生産性および共著ネットワーク上の重要度」と, 「同時期すなわち前半, および, その後すなわち後半の, 彼らおよび彼らの共著者 (前半に共著した相手) の論文生産性」との関連を調査する。

論文生産性は, 量的側面, すなわち発表論文数を考え, 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³, 第 1 著者である論文だけを数える第 1 著者計数 (*FST*) を用いる。前半と後半の区別, そして対象研究者と彼らの共著者の区別は, 添字を付すことで示す (前半: 1, 後半: 2, 研究者本人: *r*, 共著者: *c*)。

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 *DI* と出次数 *DO* 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 *CL* と *CF* を用いる。*DI* と *CL* が代表者 (第 1 著者) としての重要度, *DO* と *CF* が支援者 (その他の共著者) としての重要度に対応する。

³ 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。

$$CL(n_i) = \sum_{j=1}^g a_{ij} CF(n_j) \quad (1)$$

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

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⁴。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), $CL(n_i)$ と $CF(n_i)$ を計算する。

3.2 関連性の分析

研究者の属性のうち, 何れが, 本人および共著者の, 同時期およびその後の発表論文数と相関が高いかを明らかにする。具体的には,

1. 本人の, 同時期/その後の発表論文数を表す指標群 [$CMP_{r1/2}$, $ADJ_{r1/2}$, $FST_{r1/2}$]
2. 共著者の, 同時期/その後の発表論文数を表す指標群 [$CMP_{c1/2}$, $ADJ_{c1/2}$, $FST_{c1/2}$]

1. と 2. それぞれ, 研究者の諸属性を表す指標群 [CMP_{r1} , ADJ_{r1} , FST_{r1} , DI_{r1} , DO_{r1} , CL_{r1} , CF_{r1}] と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, 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. に関しては, 共著者が複数存在する場合を考慮し, 指標の平均値, 最小値, 最大値, 各々との相関係数を求める。また, 特に, 研究の代表者 (第 1 著者) が支援者 (その他の共著者) に与える影響に関する知見を得るため, 対象の研究者が第 1 著者であるときの共著者に絞った場合の相関係数も, 同様に求めることにする。

4 分析結果

前半に少なくとも 1 編共著論文を発表している研究者について, 前半の発表論文数と, 本人/共著者の同時期/その後の属性を表す指標群との相関係数を表 2 に示した。

まず, 同時期における本人の属性間の関係に注目すると, CMP_{r1} と ADJ_{r1} の相関が高いことが確認できる。ある程度は, 単純な完全計数で調整計数を代替できると言える。一方, FST_{r1} と $CMP_{r1} \cdot ADJ_{r1}$ との相関はさほど高くはなく, 研究代表者

⁴ 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。

	CMP_{r1}	ADJ_{r1}	FST_{r1}	DI_{r1}	DO_{r1}	CL_{r1}	CF_{r1}
CMP_{r1}	1.000	0.769	0.412	0.254	0.412	0.289	0.380
ADJ_{r1}	0.769	1.000	0.476	0.265	0.166	0.382	0.226
FST_{r1}	0.412	0.476	1.000	0.905	-0.545	0.911	-0.448
CMP_{r2}	0.419	0.370	0.208	0.144	0.182	0.169	0.200
ADJ_{r2}	0.416	0.376	0.214	0.144	0.171	0.171	0.191
FST_{r2}	0.296	0.286	0.298	0.237	-0.009	0.261	0.035
CMP_{c1} {平均}	0.378	0.261	0.168	0.121	0.147	0.246	0.390
CMP_{c1} {最小}	0.040	0.215	0.117	0.037	-0.147	0.183	0.065
CMP_{c1} {最大}	0.440	0.224	0.151	0.123	0.237	0.220	0.446
ADJ_{c1} {平均}	0.343	0.503	0.254	0.153	0.030	0.324	0.327
ADJ_{c1} {最小}	-0.027	0.465	0.199	0.075	-0.275	0.238	-0.049
ADJ_{c1} {最大}	0.448	0.397	0.211	0.145	0.179	0.277	0.431
FST_{c1} {平均}	0.270	0.304	-0.294	-0.360	0.540	-0.280	0.687
FST_{c1} {最小}	-0.087	0.245	-0.201	-0.265	0.103	-0.199	0.232
FST_{c1} {最大}	0.386	0.180	-0.261	-0.302	0.618	-0.248	0.726
CMP_{c2} {平均}	0.288	0.218	0.133	0.099	0.126	0.179	0.281
CMP_{c2} {最小}	-0.065	0.157	0.087	0.021	-0.169	0.136	-0.013
CMP_{c2} {最大}	0.374	0.193	0.123	0.107	0.223	0.161	0.342
ADJ_{c2} {平均}	0.287	0.225	0.132	0.095	0.126	0.174	0.277
ADJ_{c2} {最小}	-0.066	0.161	0.087	0.020	-0.170	0.135	-0.015
ADJ_{c2} {最大}	0.373	0.202	0.121	0.101	0.223	0.156	0.340
FST_{c2} {平均}	0.288	0.190	0.023	-0.007	0.242	0.030	0.350
FST_{c2} {最小}	-0.065	0.134	-0.010	-0.060	-0.063	0.006	0.046
FST_{c2} {最大}	0.354	0.184	0.031	0.010	0.303	0.034	0.389

Table 2: 各属性と、本人および共著者の発表論文数との相関(対象数 13059)

としての活性度を見る際には、完全計数や調整計数でなく、第1著者計数を用いる必要性が強い。ネットワーク上の重要度との相関については、代表者としての重要度 $DI_{r1} \cdot CL_{r1}$ が、第1著者としての論文数 FST_{r1} と、かなり高い相関がある一方、支援者としての重要度 $DO_{r1} \cdot CF_{r1}$ は、論文数と高い相関を示していない。この観点から評価を行う場合、論文数を指標にして測ることは難しいと言える。

前半における属性と、本人の後半の論文数との関係を見ると、すべての組合せで相関係数は0.5未満であり、特に高い相関があるものはない。後半の論文数とある程度(0.3 ~ 0.4程度)の相関を示しているのは、完全計数による論文数 CMP_{r1} であり、他の属性よりも、その後の生産性に比較強く関わっている(FST_{r2} との相関については、 FST_{r1} も同程度)。貢献の分量を考慮した活性度や、第1著者としての活性度を予測する場合も含み、全体として、その後の生産性の予測に、相対的に最も役立つと言えたのは、単純な完全計数という結果だった。

次に、共著者の発表論文数との関係を見ると、同時期の論文数に関しては、かなりの相関を示す指標も存在したが(CF_{r1} が FST_{c1} {平均} に対して0.687, など)、その後の論文数との相関係数は、すべて0.4未満であり、特に高い相関があるものはない。

表3は、支援者(第2以降の著者)との関係に限定した場合の相関係数である。共著者のその後の生産性との関係に着目すると、平均・最小・最大、何れで見ても、ネットワークの大域的構造を反映した、

代表者としての重要度 CL_{r1} が、最も高い相関を示している(特に、 CMP_{c2} {平均}, ADJ_{c1} {平均} に対しては0.5を超える)。入次数 DI_{r1} も、同じく、代表者としてのネットワーク上の重要度であるが、こちらは、支援者のその後の発表論文数と、ほとんど相関がない(平均で0.1 ~ 0.2)。支援者のその後の生産性の予測を考えるならば、ネットワークの大域的構造まで考慮する必要があることを示す結果となった。

5 おわりに

研究者のどの属性が、本人や共著相手の論文生産性と高い関連性があるかを特定することを目的に、属性間の相関について調査した。同時期の発表論文数に関しては、高い相関を示す指標も存在したが、後の発表論文数に関しては、本人のものにも、共著者のものにも、特に高い相関を示す指標はなかった。ただし、支援者(自身が第1著者であるときの共著者)に限定した場合、Yoshikaneらの指標 CL が、彼らのその後の発表論文数と、比較的高い相関を持つことが分かった。 CL は、共著相手の重要度まで反映させる指標であり、「支援者として活躍する研究者をまとめてきた実績」を表すものと言える。そのような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生産性に影響を及ぼしている可能性が示唆される。

分野によって、研究スタイル、特に共同研究の慣習は大きく異なるため、こうした属性間の関連性の

	CMP_{r1}	ADJ_{r1}	FST_{r1}	DI_{r1}	DO_{r1}	CL_{r1}	CF_{r1}
CMP_{r1}	1.000	0.858	0.756	0.268	0.777	0.421	0.772
ADJ_{r1}	0.858	1.000	0.696	-0.166	0.625	0.447	0.623
FST_{r1}	0.756	0.696	1.000	0.324	0.281	0.450	0.293
CMP_{r2}	0.464	0.412	0.369	0.148	0.390	0.264	0.391
ADJ_{r2}	0.453	0.412	0.367	0.129	0.373	0.257	0.374
FST_{r2}	0.336	0.317	0.328	0.117	0.225	0.221	0.231
CMP_{c1} {平均}	0.359	0.281	0.306	0.141	0.271	0.784	0.291
CMP_{c1} {最小}	0.155	0.253	0.074	-0.310	0.148	0.526	0.162
CMP_{c1} {最大}	0.394	0.256	0.355	0.297	0.288	0.785	0.307
ADJ_{c1} {平均}	0.327	0.415	0.296	-0.157	0.231	0.800	0.253
ADJ_{c1} {最小}	0.084	0.381	0.030	-0.663	0.081	0.441	0.094
ADJ_{c1} {最大}	0.387	0.365	0.365	0.095	0.266	0.824	0.287
FST_{c1} {平均}	0.297	0.214	0.221	0.182	0.257	0.424	0.270
FST_{c1} {最小}	0.074	0.175	-0.028	-0.304	0.119	0.196	0.127
FST_{c1} {最大}	0.328	0.209	0.268	0.284	0.269	0.441	0.283
CMP_{c2} {平均}	0.246	0.202	0.225	0.123	0.186	0.525	0.205
CMP_{c2} {最小}	0.036	0.170	-0.020	-0.336	0.064	0.330	0.076
CMP_{c2} {最大}	0.298	0.189	0.293	0.288	0.211	0.535	0.229
ADJ_{c2} {平均}	0.245	0.208	0.229	0.111	0.180	0.510	0.199
ADJ_{c2} {最小}	0.034	0.173	-0.020	-0.347	0.061	0.322	0.072
ADJ_{c2} {最大}	0.296	0.197	0.296	0.270	0.203	0.522	0.222
FST_{c2} {平均}	0.226	0.158	0.217	0.192	0.157	0.310	0.169
FST_{c2} {最小}	0.021	0.127	-0.037	-0.277	0.053	0.137	0.061
FST_{c2} {最大}	0.257	0.158	0.260	0.279	0.171	0.324	0.183

Table 3: 各属性と、本人および共著者（第2以降の著者）の発表論文数との相関（対象数 5518）

傾向は、分野に依存するものと考えられる。今回は計算機科学分野のみを扱ったが、今後は、他の分野も対象に加えて分析を行い、傾向を比較したい。

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

1. 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.
2. Kleinberg, J. M. (1998) "Authoritative sources in a hyperlinked environment," *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, p. 668-677.
3. Kretschmer, H. (1994) "Coauthorship networks of invisible-colleges and institutionalized communities," *Scientometrics*, vol. 30, no. 1, p. 363-369.
4. Kretschmer, H. (1997) "Patterns of behaviour in coauthorship networks of invisible colleges," *Scientometrics*, vol. 40, no. 3, p. 579-591.
5. Kretschmer, H. and Gupta, B. M. (1998) "Collaboration patterns in theoretical population genetics," *Scientometrics*, vol. 43, no. 3, p. 455-462.
6. 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.
7. 安田 雪 (2004) 『人脈づくりの科学』, 日本経済新聞社: 東京, 253p.
8. 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.
9. 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.
10. Yoshikane, F., Nozawa, T., Shibui, S. and Suzuki, T. (2007) "An analysis of the connection between researchers' productivity and their co-authors' past attributions, including the importance in collaboration networks," *Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics*, vol. 2, p. 783-791.
11. Zobel, J. (1999), Guideline on Research Practice in Computer Science. Retrieved September 14, 2006 from: <http://goanna.cs.rmit.edu.au/~jz/conduct.html>.