

## 研究者の論文生産に関わる諸特性の相関に関する分析（要旨）

芳鐘 冬樹

本研究は、研究者の活動状況に関する特性を異なる観点から観察・評価したとき、それらの間に何らかの関連性が認められるかについて、計算機科学分野を事例に、傾向・特徴を明らかにすることを目的とする。具体的には、(1) 共著ネットワークを考慮した重要度と、(2) 研究者個人の論文生産性との相関を中心に、本人と共著相手の状況の相関や、ある時点とその後の状況の相関という視点も含め、論文生産活動の状況に関する観点間の関連性についての傾向を明らかにした。これらの研究活動の状況に関する観点間の関連性を把握できれば、対象分野・領域における知識生産活動の特徴、特に論文生産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と考えられる。

これまで、同時期の特性間の相関を分析する研究は多くなされてきた。例えば、共著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し、研究者個人の生産性と共著ネットワーク上の重要度の相関を報告した研究もある。しかしながら、時間が経過した後の状況との相関に着目して、研究者の特性間の関連性に関する定量的な分析を行った研究は、これまでほとんどなされていない。また、共著ネットワークの分析において、直接的な結合相手の数のみに基づいた研究は、数多く存在するものの、より大域的なネットワーク構造を考慮に入れた結合相手の重要度や、間接的な結合関係に着目して、研究者の特性を分析した研究は、未だ十分になされているとは言い難い。HITS アルゴリズムを応用した指標によって、共著ネットワークの大域的な構造を反映した特性を測り、さらに間接的な関係の媒介性にも目を向け、それらも含めて、研究者の特性についてその後の状況との相関を分析している点に、本研究の特色がある。

計算機科学分野の理論領域と応用領域という2つの下位領域を分析の対象とした。計算機科学分野を含む自然科学系分野における最も包括的な書誌データベースのひとつである *SCI (Science Citation Index)* 論文データベースを情報源として、1991年から2005年までの当該分野・領域の論文書誌データを抽出し分析に用いた。データの基本的な数量は、雑誌数 68 (うち理論領域 35, 応用領域 33), 論文数 68,352 (うち理論領域 28,485, 応用領域 39,867), 延べ著者数 175,055 (うち理論領域 64,666, 応用領域 110,389) である。また、名寄せ処理を施した後の異なり著者数は、理論領域で 31,692, 応用領域で 59,771 である。本研究では、分野内の研究者の相対的關係 (生産性の高低, 共著傾向の強弱など) を内包する抽象的な論文生成源を潜在的母集団として想定する。観察される現象 (生産性分布, 共著ネットワークなど) は、生成源のポテンシャルティ (潜在性) が、その外側の条件 (論文産出サイクルや発表媒体の数量など, そしてそれらによって規定される論文発表の機会, 論文数) に応じて現実化したものと捉える。すなわち、用いるデータは、現実態としては全数調査に基づくが、母集団そのものを構成するものでなく標本と位置付ける。

研究者の活動状況を見る観点として、(A) 個人の論文生産性と、(B) 研究協力ネットワークを考慮した重要度を設定した。(B) については、(B)-1. 研究協力を行った相手の数だけ見る観点、(B)-2. 相手の重要度も含めて考慮する観点、(B)-3. 相手の所属の多様性、および、(B)-4. 間接的な経験・知識の媒介性に細分化した。(A) は、完全計数法、調整計数法、第1著者計数法、最終著者計数法による論

文数 ( $CMP, ADJ, FST, LST$ ) と単著論文数 ( $SIN$ ) を指標に用いた (論文数指標族と呼ぶ)。(B)-1. は、支援を受けた支援者の異なり数と延べ数 ( $C_{in}, C'_{in}$ ), 支援した代表者の異なり数と延べ数 ( $C_{out}, C'_{out}$ ), 共著パートナーの異なり数と延べ数 ( $C_{deg}, C'_{deg}$ ) を指標に用いた (次数中心性指標族と呼ぶ)。(B)-2. は、HITS アルゴリズムを応用して算出する代表者としての協力実績・人脈の重要度 ( $CL, CL'$ ), 支援者としての協力実績・人脈の重要度 ( $CF, CF'$ ), 役割を区別しない協力実績・人脈の重要度 ( $CC, CC'$ ) を指標に用いた (HITS 指標族と呼ぶ)。そして、(B)-3. は共著相手機関の異なり数 ( $V_{org}$ ) と共著相手国の異なり数 ( $V_{cnt}$ ) を、(B)-4. は媒介する研究者ペア数 ( $CB$ ) を指標に用いた (それぞれ、所属多様性指標族、媒介中心性指標と呼ぶ)。(B) 研究協力ネットワークを考慮した重要度のうち、 $C'_{in}, C'_{out}, C'_{deg}, CL, CF, CC$  は協力実績を重視するものであり、一方、 $C_{in}, C_{out}, C_{deg}, CL', CF', CC', V_{org}, V_{cnt}, CB$  は、重みなしグラフに基づき、実績よりも人脈の構成に注目するものである。また、(B)-1. 次数中心性指標族と (B)-2. HITS 指標族のうち、 $C_{in}()$ ,  $CL()$  は代表者としての重要度に、 $C_{out}()$ ,  $CF()$  は支援者としての重要度に、そして、 $C_{deg}()$ ,  $CC()$  は役割を区別しない重要度に対応する。

研究者の論文生産に関わる特性間の関連性の分析では、現実には観察された論文数を重視し、現実の要因・制約のもとでの対象の特徴として、データそのものの特徴を指標で測る。標本の信頼性への配慮は、同じ論文生成源が同じ条件のもとで現実化する (同じ潜在的母集団から同じ量の標本が再抽出される) 場合に生じうるランダムな誤差のみを想定した統計的有意性の確認にとどめる。ただし、特性間の関連性の分析結果をより適切に解釈できるようにするための下地として、各研究領域の研究者の生産性および共著傾向の基本的な状況については、その潜在的な特徴にも目を向け、標本をもとにした現象の再構築、論文数に応じた現象の変化の分析を行う。具体的には、各々の領域について、2,000 ずつに区切った標本量ごとに、ランダムサブサンプリングを 10,000 回繰り返すことで指標の平均値を求めるモンテカルロシミュレーションを行い、標本量の変化に伴う指標の挙動を観察する。

1991 年～1995 年と 1996 年～2000 年、および 1996 年～2000 年と 2001 年～2005 年を、それぞれ対にして、対象研究者の状況と共著者の状況、そして特定の時期の状況とその後の状況という視点も含めて各指標の相関を調査した。相関の強さは、影響関係を直接示すものではないが、その可能性の有無、程度についての示唆が得られるものと考え。また、因子分析などの手法を採らず、相関係数に基づく分析を行う理由は、複数の変数の背後に潜む要因ではなく、設定した観点そのものが重要と考える立場を本研究はとり、それらの観点に対応する指標同士の関連性が見たいからである。

分析結果から明らかにした点のうち主要なものを以下にまとめる。まず、全般的な状況の動的傾向に関しては、領域全体の論文の数が等しいという条件のもとでは、理論領域に属する研究者の方が、応用領域の研究者より、平均して多くの論文を生産しており、平均論文数の成長率を見ても、理論領域の方が一貫して高い値を保ち続けている。つまり、応用領域に比べて、理論領域は、領域の論文生産数が増えていく過程で、新たな周辺的な著者が加わる傾向は弱く、同じ著者が繰り返し論文を生産する傾向が強い。平均共著パートナー数の成長率の変化では、1995 年以前の応用領域の減衰が遅いことが、最も顕著な点として観察された。1995 年以前の応用領域は、論文の生産数を増やすのに、多くの新たなパートナーを要する傾向が相対的に強いと推測される。

同時期の特性間の相関に関しては、完全計数法による論文数  $CMP$  は、全般的に他の指標との相関がある程度は高かった。特に、調整計数法による論文数  $ADJ$  との相関が比較的高かったが、それでも

両指標の相関係数は 0.70 を下回っており、研究パフォーマンスの評価の場面などにおいて、完全計数法で調整計数法を代替することは妥当とは必ずしも言い難い。本人のその後の特性を測る指標に対しては、完全計数法による論文数  $CMP$  および媒介ペア数  $CB$  が、全般的にある程度高い相関を示していた。その後の論文数  $CMP$  に対して相関が最も高いのは、同じ指標  $CMP$  だったが、その後のネットワーク関係基本指標  $C_{deg}, V_{org}$  に対しては、それらの指標自身でなく、媒介ペア数  $CB$  の相関が高かった。媒介ペア数  $CB$  は、パートナーを変えつつ論文の発表を繰り返す傾向に関わる指標であるが、そうした傾向が、単なるパートナー数の多寡よりも、その後のネットワーク構築に影響している可能性もあると考える。共著者のその後の特性を測る指標に対しては、相関が最も高い指標は、HITS 指標族  $CC(^{\cdot}), CL(^{\cdot}), CF(^{\cdot})$  だった。HITS 指標族によって測られるところの「隣接する研究者とのつながりの多さだけでなく、それら周囲の研究者が形成する関係をも含めた、より広範囲なつながりの構成」が、共著した相手のその後の状況に、ある程度強く影響している可能性が窺える。

理論領域の方が、応用領域よりも、指標間の相関が強い傾向が観察された。統計量の標本量依存性に起因して、比較する条件（各々の領域内の論文数）が変われば、相関係数は変化する。両領域の比較結果を一般化するに際して注意を要する。ただし、例えば、完全計数法による論文数  $CMP$  に関わる相関については、 $CMP$  の平均値の動的傾向から、等しい条件のもとで比べると理論領域の方が一般的に相関が強くなると推測しうる。