

共著ネットワークの成長：4分野の比較分析

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

これまで、多くの研究が「共著ネットワーク」を分析して、そのパターンの記述を試みてきた。しかしながら、それらのほとんどはネットワークの静的な記述が中心であり、著者生産性データの特性 - 統計量の標本量依存性 - を、必ずしも考慮したものではなかった。本研究では、標本量の増加にともなうネットワークの成長、すなわちネットワークの動的特性に着目し、それを考慮に入れて共著の実態に関する分野間の差異を明らかにすることを試みる。

1 はじめに

今日、学術研究の高度化・複雑化、研究者の専門分化、そして学際的分野の発達を背景に、複数の研究者による研究協力の重要性がますます増加してきている。それは、例えば、自然科学分野を中心として、多くの分野で1論文あたりの平均著者数、あるいは全論文に占める共著論文の比率の増加が報告されていることからもうかがえる (e.g. Devilliers, 1984; Drenth, 1998; O'Neill, 1998)。そのような状況のもと、研究協力 (特に、産学連携や大学間の協力など) の体制整備が重要な課題になっていると言えるが、そのためには、まず、現状における研究協力の実態を把握する必要がある。

研究協力の実態は、その成果である共著論文の分析を通して、ある程度測ることができる。これまで、多くの研究が「共著ネットワーク」を分析して、そのパターンの記述を試みてきた (e.g. Abdel Kader et al., 1998; Elalami et al., 1992; Okubo et al., 1992)。しかしながら、それらのほとんどはネットワークの静的な記述が中心であり、著者生産性データの特性 - 統計量の標本量依存性 - を、必ずしも考慮したものではなかった。本研究では、標本量の増加にともなうネットワークの成長、すなわちネットワークの動的特性に着目し、それを考慮に入れて共著の実態に関する分野間の差異を明らかにすることを試みる。

2 分析の観点と指標の選択

本研究では、ネットワークの成長という観点から、共著の実態を分析する出発点として、共著機会の増加にともなう、個々の著者 (ノード) の共著関係 (リンク) の変化に注目する。具体的には、共著パートナーの異なり数 (ノードから伸びるリンクの数) を観察することで、ネットワークの量的な成長を、各々のパートナーとの結びつき (リンクの強度) の

偏りを観察することで、ネットワークの質的な変化を調べる。

前者については、パートナーの異なり数 V で直接計り、後者については、各々のパートナーとの共著頻度をもとに算出したジニ係数 G を指標として用いることにした。 G は次の式で求められる。

$$G = \sum_{i=1}^V \sum_{j=1}^V \frac{|f_i - f_j|}{2\mu V^2} \quad (1)$$

ここで、 f_i はパートナー $a_i (1 \leq i \leq V)$ との共著頻度を、 μ は共著頻度の平均を表す。

偏り、あるいは不平等の指標が数多く提案されている (cf. Egghe & Rousseau, 1991) 中で、 G を選んだ理由は、 G の2つの特徴にある。2つの特徴とは、事象数に対して感度を持たないこと、そして、特定の事象に対して高い感度を持つということはないこと、である (芳鐘, 2000)。本研究では、ネットワークの質的特性として、事象数 (パートナーの異なり数) を考慮した「集中・分散」ではなく、純粋な「偏り」を観察したいため、1つめの特徴は望ましく (事象数は、量的特性として、もう1つの指標 V で観察する)、そして、特定の事象 (例えば最頻パートナー) との共著頻度に注目するのではなく、総体としての偏りを観察したいため、2つめの特徴もまた望ましいと考えた。

3 データと分析対象

国立情報学研究所が提供する学会発表データベースから抽出した4分野 (電気・EE, 情報処理・IP, 高分子・PS, 農芸化学・BC) の論文データをもとに、分野の特徴を比較する。それぞれの分野について、1992年から1997年までの発表論文を抽出し¹、これを標本データとした。4分野の標本それぞれの基本的な数量を表1に示した。 P は論文数、 A_{token} は延べ著者数、 A_{type} は異なり著者数を表す。

	P	A_{token}	A_{type}	C_L
EE	19784	75685	25230	0.241
IP	27047	79372	24267	0.225
PS	21505	76411	16820	0.213
BC	17782	71974	21315	0.229

Table 1: 4分野の標本の基本的数量

¹ 電気学会, 情報処理学会, 高分子学会, そして日本農芸化学会を開催学会とする発表論文を抽出した。

あわせて、それぞれの標本の損失係数 C_L も表 1 に示した。 C_L は、標本としての不十分さを示すものである。 C_L は、現実に観測された異なり著者数と、母集団出現確率を標本相対頻度で推定して求めた異なり著者数の差から導かれる (Chitashvili & Baayen, 1993)。すべての標本において、20%以上の差が生じていることから、これらの標本をもとに計算した統計量から信頼できる母集団推定は行えないことが分かる。また、 C_L が大きな値をとるとき、ほとんどすべての統計量が標本量に依存して系統的に変化し (Tweedie & Baayen, 1998)、本研究が用いる指標 V, G についても、それは例外ではない。

すなわち、本研究が、ネットワークの静的な記述ではなく、その変化を追う理由は、「ネットワークの変化 (成長) という視点から特徴を捉える方が、対象をより豊かに記述できること」にあると同時に、「理論的に、そもそも母集団を直接比較することが困難なこと」にもある。

4 分析手法

標本量の増加にともなうネットワークの変化を観察する場合、標本抽出の単位が問題になる。表 2 に示したとおり、分野によって、1 論文あたりの平均著者数 A_{num} や共著論文の比率 C_{per} に差があるため、論文単位で標本抽出を行うと、当然、その影響が大きく及ぶ。本研究では、 A_{num}, C_{per} に大きく左右される「共著を行う機会」の多寡というよりも、機会が訪れた際の、パートナー選択の特徴そのものを観察するために、論文をいう枠を外した、1 対 1 の「ペア」を標本抽出の単位とした²。つまり、本研究では、抽象的な「共著機会の数」とでも言うものを標本量としている。4 分野の標本量 (延べペア数) N も、表 2 に示したおいた。

	A_{num}	C_{per} (%)	N
EE	3.83	95.3	130668
IP	2.93	85.0	103092
PS	3.55	94.4	118539
BC	4.05	95.9	132608

Table 2: 4 分野の共著関係の数量

これらの標本をもとに、5000 ずつに区切った標本量ごとに、無作為部分標本抽出を 1000 回繰り返すモンテカルロ実験を行い、リンク数 V とリンク偏り G という 2 つの指標の、標本量 N の変化にともなう値の変化を観察する。(1) 指標の値の著者 1 人あたりの平均から、分野のマクロな特徴を、(2) 指標の値ごとの著者数の分布から、分野のより詳細な特徴を明らかにする。

² 例えば、A, B, C という 3 人の共著が 1 つあった場合、A-B, A-C, B-C という 3 つの標本に分解する。

5 分析結果：4 分野の特徴

標本量 (共著機会) N の増加にともなうリンク数 (パートナー数) $V(N)$ とリンク強度 (共著頻度) の偏り $G(N)$ の変化 (著者 1 あたりの平均) を図 1 に、それぞれの成長率を図 2 にプロットした。標本量の増加にともなって、 $V(N), G(N)$ が増加していくこと、そしてその変化の特徴は分野によって異なることが観察できる。各分野の特徴を簡単に整理すると以下ようになる。

まず、情報処理 (IP) と高分子 (PS) は、 $V(N), G(N)$ ともに成長率の低下が速い。これらの分野は、比較的速くコアとなるパートナー集団の形成がされ、その後は新しいパートナーの出現が少なく、コアとの共著を同じような頻度で繰り返す傾向が強いという解釈が可能である。ただし、情報処理の $V(N)$ は、初期段階においても、それほど大きく増加していないため、他の 3 分野と比較すると、一貫して最も低い値で推移している。また、高分子の $G(N)$ は、初期段階における増加が非常に大きいため、成長率の低下が速いといっても、観察範囲内では逆転は生じず、一貫して最も高い値で推移している。

一方、電気 (EE) と農芸化学 (BC) は、上述の 2 分野に比べて、 $V(N), G(N)$ の成長率の低下が遅い。例えば、農芸化学は、 $V(N), G(N)$ ともに、初期段階では高分子よりも値が低いが、その後の成長率が高いため、標本量が増加する過程で逆転している。このことから、電気と農芸化学は、コアとなるパートナーとの共著を繰り返す一方で、その他のパートナーも着々と増やしていく、そして両者 (コアとその他) の共著頻度の差が大きくなっていくと解釈することができよう。

各分野のより細かい特徴は、図 3, 4 に示したリンク数 (パートナー数) ごと、リンク強度 (共著頻度) の偏りごとの著者数の分布からうかがえる^{3, 4}。例えば、高分子の最も顕著な特徴は、非常に多く (18 人以上) のパートナーを持つ著者が、かなり早い段階で多数現れるが、初期段階を過ぎると、そのような著者はそれほど増えないことである。この特徴からも、図 1 における高分子の挙動を説明することができるだろう。

6 おわりに

本研究では、ネットワークの成長という観点から、共著の実態を分析する出発点として、共著機会の増加にともなう、個々の著者の共著関係の変化に注目し、その分野による違いを明らかにした。本研究で示した共著の実態に関する分野間の差異は、基本的には、それぞれの分野が持つ研究の性格上の必要性から生じていると考えられるが、一方で、何らかの制約に望ましからざる影響を受けている可能性もある。例えば、所属機関の違いが共著ネットワークを広げる障害になっていることも考えられるだろう。

³ 図 3, 4 では、著者数 F の軸を対数でとっている。

⁴ $V(N)$ については 2 きざみ (18 以上はまとめる)、 $G(N)$ については 0.2 きざみ (0.18 以上はまとめる) で区間をとり、そこに含まれる著者数を数えている。

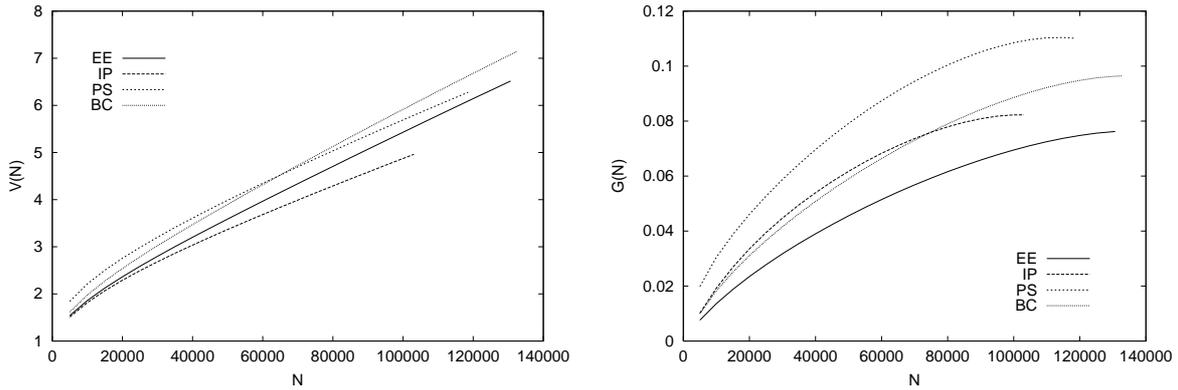


Figure 1: リンク数 (左図)・偏り (右図) の変化

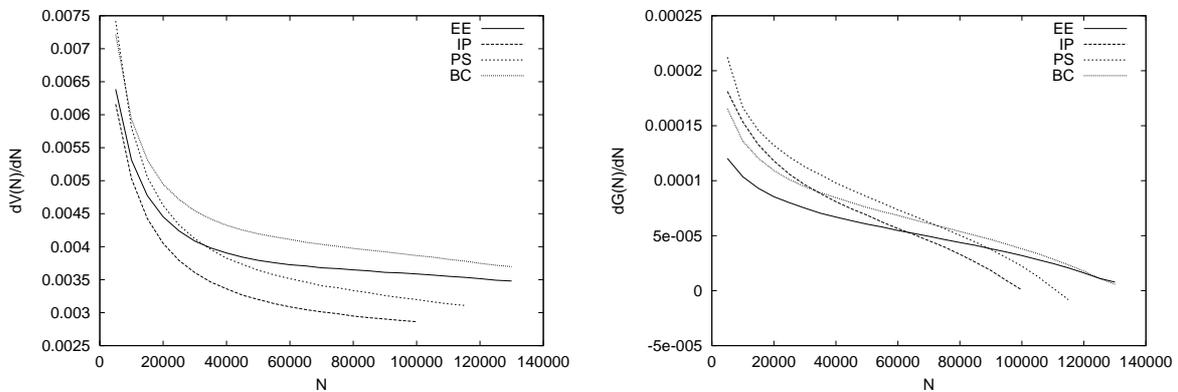


Figure 2: リンク数 (左図)・偏り (右図) の成長率

今後は、所属機関など、著者の属性も考慮に入れて、共著ネットワークに関する分野の特徴について、さらに詳細な分析を進めたいと考えている。

References

- Abd el Kader, M., Ojasoo, T., Miquel, J. F., Okubo, Y. and Dore, J. C. (1998) "Hierarchical author networks: an analysis of European Molecular Biology Laboratory (EMBL) publications," *Scientometrics*, Vol. 42, No. 3, p. 405-421.
- Chitashvili, R. J. and Baayen, R. H. (1993) "Word frequency distributions." In: Hrebicek, L. and Altmann, G. (eds.) *Quantitative Text Analysis*. Trier: Wissenschaftlicher Verlag. p. 54-135.
- Devilliers, F. P. R. (1984) "Publish or perish: the growing trend towards multiple authorship," *South African Medical Journal*, Vol. 66, No. 23, p. 882-883.
- Drenth, J. P. H. (1998) "Multiple authorship: the contribution of senior authors," *Journal of the American Medical Association*, Vol. 280, No. 3, p. 219-221.
- Egghe, L. and Rousseau, R. (1991) "Transfer principles and a classification of concentration measures,"

Journal of the American Society for Information Science, Vol. 42, No. 7, p. 479-489.

Elalami, J., Dore, J. C. and Miquel, J. F. (1992) "International scientific collaboration in arab countries," *Scientometrics*, Vol. 23, No. 1, p. 249-263.

Okubo, Y., Miquel, J. F., Frigoletto, L. and Dore, J. C. (1992) "Structure of international collaboration in science: typology of countries through multivariate techniques using a link indicator," *Scientometrics*, Vol. 25, No. 2, p. 321-351.

O'Neill, G. P. (1998) "Authorship patterns in theory based versus research based journals," *Scientometrics*, Vol. 41, No. 3, p. 291-298.

Tweedie, F. J. and Baayen, R. H. (1998) "How variable may a constant be? :measures of lexical richness in perspective," *Computers and the Humanities*, Vol. 32, p. 323-352.

芳鐘 冬樹 (2000) "計量書誌学的分布における集中度：集中度の概念と指標の特徴," 『日本図書館情報学会誌』, Vol. 46, No. 1, p. 18-32.

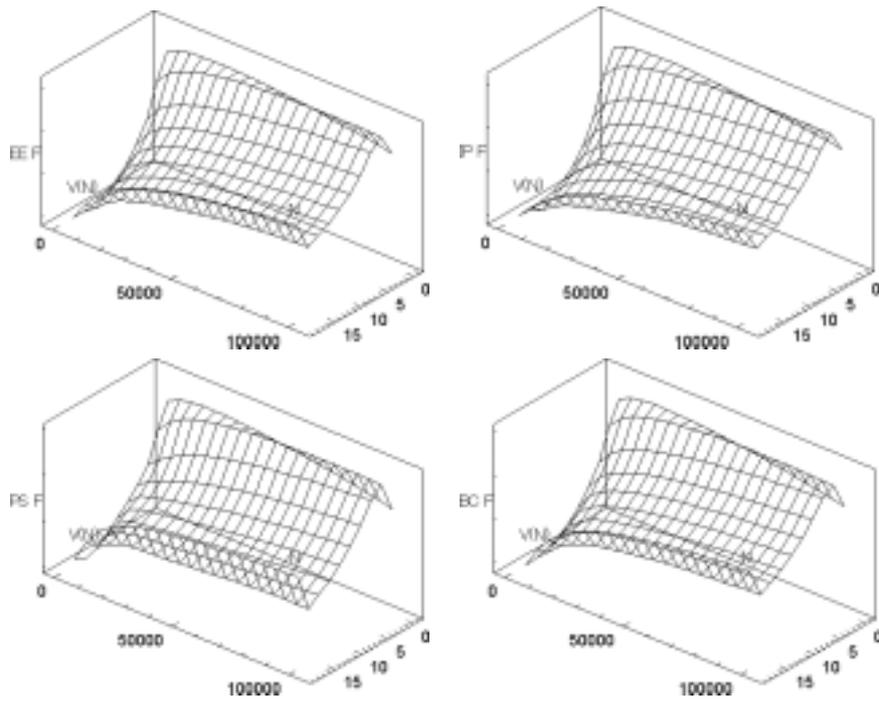


Figure 3: リンク数ごとの著者数の変化

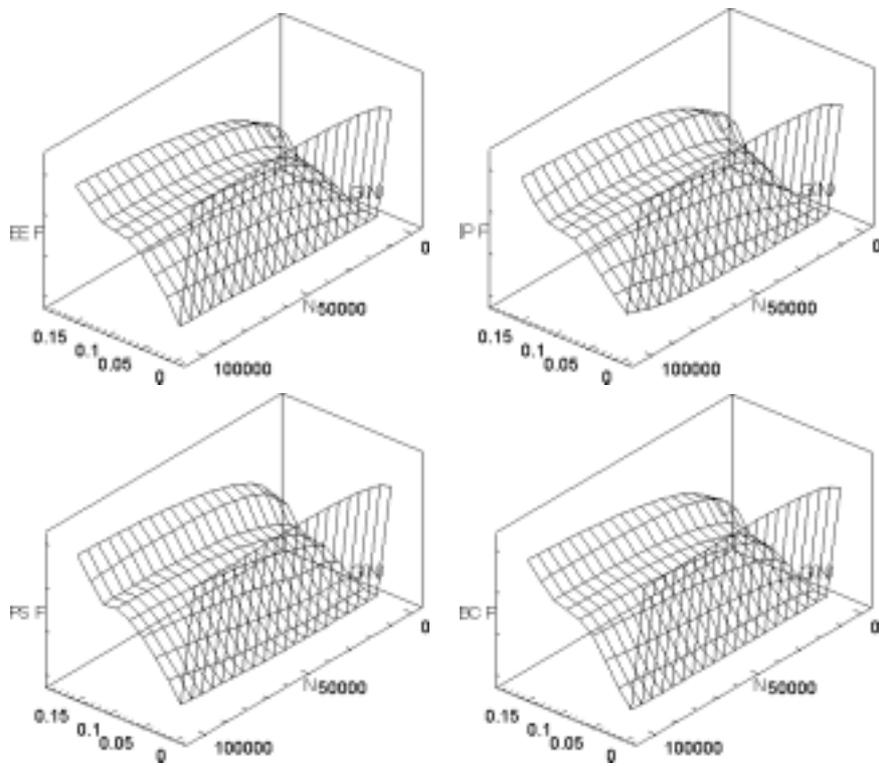


Figure 4: リンク偏りごとの著者数の変化