

共著ネットワークの分析：視点と手法

芳鐘 冬樹 (東京大学大学院教育学研究科, E-Mail: fuyuki@p.u-tokyo.ac.jp)

影浦 峯 (国立情報学研究所, E-Mail: kyo@nii.ac.jp)

Abstract

これまで、多くの研究が、ミクロな視点、あるいはマクロな視点から共著という現象を分析し、そのパターンの特徴付けを行ってきた。しかしながら、用いられる指標や手法は様々であり、共著分析に関する知見が体系化されているとは言い難いのが現状である。そこで、本研究では、共著分析に関する研究を整理し、分析の視点、用いる指標・手法といった観点から体系化を試みた。

1 はじめに

Lotka (1926) が著者の生産性に関する法則を発見した時代、ほとんどの論文は、1人の著者によって著された単著であった。しかし、時代が降につれ、複数の著者による共著の比率が大きくなってきていることが、様々な分野で報告されている。「共著」という現象は、一面では、パートナー選択に関する研究者個人の社会的振舞いの現れであり、また、所属機関や国の政策を反映するマクロな側面も併せ持つ。そのような考えに基づき、研究者同士、あるいは研究者が属する機関や国同士のつながりという、学術コミュニティにおけるネットワーク(見えざる大学)の現れとして、共著を解釈する研究は多い。

共著を観察し、ネットワークのパターンを記述した研究のほとんどは、著者の属性と、共著頻度をもとに計算する何らかの指標の値との関係を分析したものである。分析の視点によって注目する属性は異なり、また、用いられる指標や分析手法も様々である。以下では、まず、代表的な研究を紹介し、それらの分析の視点について説明する。次に、それらの研究が用いた指標・手法を整理する。そして、最後に、これまでの研究の問題点を指摘したい。

2 共著分析の視点

表1に、共著ネットワークの分析を行った代表的な研究をまとめた。これらは、どのような属性の著者とどのような属性の著者(例えば、どの国の研究者とどの国の研究者)の間に共著が多いかという傾向を分析した研究である。

これらの研究で行われている共著分析の視点は、大きく分けて2つある。研究者個人への振舞いに着目する視点と、所属する集団としての振舞いに着目する視点である。その違いは、例えば、共著頻度の数え方に現れる。前者の場合、1論文内に組合せの重複があっても、その延べ数を数えるが(例えば、フランスの著者1人と日本の著者3人の共著論文が1つあるとき、フランス・日本間の共著頻度は3)、後者の場合は重複が省かれる(同じ例で、フランス・日本間の共著頻度は1)。所属する研究者個人への振舞いの集合として、その属性グループ(国)の傾向を観察する方法と、属性グループ(国)の振舞いを直接観察する方法との違いと言える。それは、つまり、共著という現象の捉え方の違いによるものである。

注目する属性と分析の視点の関係を簡単にまとめると次のようになる。

著者の生産性、性別など：

第1の視点-共同研究のパートナー選択において、研究者個人の振舞いに影響を及ぼす社会的要因を分析

著者の所属国・機関など：

第2の視点-研究者個人というよりも、国や機関という集団としての政策や、集団同士の協力関係を分析

ただし、所属国については、第2の視点から分析した研究だけでなく、第1の視点から、地理的近接性や歴史的・文化的要因などが、パートナー選択において及ぼす影響を分析した研究もある。

	対象属性	指標	分析手法
Abd el Kader et al. (1998)	個人	-	階層ネットワーク化
Arunachalam et al. (1994)	国	協力係数, 親和係数	マトリクス化
Bahr & Zemon (2000)	性別	共著頻度	マトリクス化
Balog (1980)	機関	共著頻度	マトリクス化
Braun et al. (1992)	国	協力係数	ネットワーク図示
Cunningham & Dillon (1997)	地域, 機関, 性別	共著頻度	マトリクス化
Elalami et al. (1992)	国	共著頻度	FCA
Godin & Ippersiel (1996)	国, 都市	共著係数	ネットワーク図示
Kim (1999)	国, 機関	共著頻度	マトリクス化
Kretschmer (1994)	生産性	相対共著頻度	属性値差との相関分析
Kretschmer (1997)	生産性	相対共著頻度	マトリクス化
Kretschmer & Gupta (1998)	生産性	相同係数	属性値差との相関分析
Kundra & Kretschmer (1999)	生産性	相対共著頻度	マトリクス化
Lewison et al. (1993)	国	協力係数	ネットワーク図示
Miquel & Okubo (1994)	国	共著頻度	マトリクス化
Narin et al. (1991)	国	共著係数	ネットワーク図示
Okubo et al. (1992)	国	共著頻度	FCA, MST
Russell (1998)	国	共著頻度	マトリクス化

表 1: 代表的な共著ネットワーク研究

3 指標

ほとんどの研究において, ネットワークの分析は, 各々のノード(属性)間のリンクの強度(結びつきの強さ, すなわち共著しやすさ)の測定から始まっている。そのとき使われる指標には, 次のようなものがある。

共著頻度 n_{ij} :

属性 i の著者と属性 j の著者が共著で論文を著す頻度であり, リンクの絶対強度を示す指標として使われる。最も基本的な尺度であるこの指標は, 他の指標を計算する際にもとになる指標でもある。

協力係数 (cooperation index) r_{ij} :

(Arunachalam et al., 1994)

$$r_{ij} = \frac{n_{ij}}{\sqrt{n_i n_j}} \quad (1)$$

共著頻度 n_{ij} を, i の論文数¹ n_i と, j の論文数 n_j で補正した指標である。

¹ n_{ij} を延べ数で数える場合は, i に所属する著者の延べ出現数を, n_{ij} を異なり数で数える場合は, i に所属する著者が少なくとも 1 人いる論文の総数を指す。

単純な共著頻度 (= 絶対強度) では, 属性グループの生産性に左右される(当然, 論文数が多いグループとの共著頻度は高くなりやすい)。協力係数は, その影響を除くために, 共著頻度を論文の総数(2つのグループの幾何平均)で割っており, リンクの相対強度を示す指標として使われる。

相同係数 (homophily index) h_{ij} :

(Kretschmer & Gupta, 1998)

$$h_{ij} = \frac{n_{ij}G}{G_i G_j} \quad (2)$$

N 個のグループが存在するとして, G は $n_{lm} (l = 1, \dots, N, m = 1, \dots, N)$ の幾何平均, G_i は $n_{im} (m = 1, \dots, N)$ の幾何平均, G_j は $n_{lj} (l = 1, \dots, N)$ の幾何平均を表す。2つのグループの生産性に応じた補正が加えられており, 相同係数も, 協力係数と同様, リンクの相対的強度を示す指標であると言える。

相対共著頻度 z_{ij} :

(relative frequencies of coauthorships)

(Kundra & Kretschmer, 1999)

4 分析手法

前節で紹介した指標を使って、すべてのノード（グループ）の組合せについて、リンク強度を計算し、2次元マトリクス I_{ij} ($i = 1, \dots, N, j = 1, \dots, N$) を得る。このマトリクスをもとに、共著ネットワークの傾向が調べるのが、最も基本的な分析手法である。リンク強度のマトリクスは、表形式で示されることが多いが、3次元グラフ ($x = i, y = j, z = I_{ij}$) にプロットして、視覚的な分析を行った研究もある (Kretschmer, 1997; Kretschmer & Gupta, 1998; Kundra & Kretschmer, 1999)。

グループ数が少ない場合、マトリクス表示でも傾向を読みとることができるが、グループ数が多くなると、それだけではネットワーク全体の傾向をつかむのが困難になってくる。そこで、多数のグループを扱う研究では、ノード（グループ）と、ノードを結ぶリンクを、それぞれ点と線で表したネットワーク図が作られている。リンク強度 (I_{ij}) をリンクの線の太さに反映させた研究が多い。あるいは、一定値以上の強度を持つリンクのみを描き込む方法もある (Braun et al., 1992)。また、有向のリンク強度を示す指標 c_{ij} , a_{ij} を使った研究は、矢印でリンクの向きを表している。

Abd el Kader et al. (1998) は、ノードとノードを線で結ぶだけでなく、階層構造を組み入れたネットワークの図示を行っている。その手順は次のとおりである。(1) 一度でも共著したことがある著者同士を結んでクラスターを作る。(2) 最も生産性が高い(多数の論文を著している)著者をクラスターから除くと、いくつかのリンクが切れてサブクラスターに分割される。(3) それぞれのクラスターについて、著者が一人一人ばらばらになるまで2を繰り返す。(4) 同心円を描き、最初に除いた著者(最も生産性が高い著者)を中心に、後から除いた著者を除いた順に外側に描き込み、順々に線で結ぶ(円の中心から外側に広がるツリーができる)。すると、図の中心に共著ネットワークの中核を担う著者が、外側には生産性の低い周辺的な著者が現れることになる。

生産性(論文数)のような数量的な属性に基づくグループの場合、属性値とリンク強度の相関からグループの共著の傾向を調べることが可能である。Kretschmer (1994), Kretschmer & Gupta (1998) は、2人の著者の論文数の差と、彼らの間のリンク強度との相関を分析し、生産性が近い著者同士か、生産性に極端な差がある著者同士が共著しやすいと

$$z_{ij} = \frac{n_{ij}}{\sum_{k,l} n_{kl}} \div \frac{n_i}{\sum_k n_k} \frac{n_j}{\sum_k n_k} \quad (3)$$

Kundra & Kretschmer が「相対共著頻度」と呼んだ指標は、単なる共著の相対頻度ではなく、 i と j の共著が現れる確率について、観察値と期待値の比を求めたものである²。

相対共著頻度が1より大きければ、2つのグループの生産性から期待される確率よりも、実際にそれらのグループが共著する確率が高い、つまり共著しやすい傾向があると解釈される。

共著係数 (coauthorship index) c_{ij} :

(Narin et al., 1991)

$$c_{ij} = \frac{n_{ij}}{n_i} \quad (4)$$

協力係数と同様に、共著頻度をグループの論文数で割っているが、こちらは2つのグループの平均ではなく、どちらか一方のグループの論文数で割っている。あるグループ (i) の論文全体に占める、特定のグループ (j) との共著の割合を表す指標が共著係数である。したがって、共著頻度や協力係数とは異なり、一方的な(有向の)リンクの強度を示す指標と言える。

親和係数 (affinity index) a_{ij} :

(Arunachalam et al., 1994)

$$a_{ij} = \frac{n_{ij}}{\sum_{k \neq j} n_{ik}} \times 100 \quad (5)$$

親和係数の意味は、あるグループとの共著数と、その他のグループの共著数の総計の比である。つまり、一方のグループ (i) から見て、もう一方のグループ (j) とのリンクが、その他のグループとのリンクに比べてどれだけ強いを示す指標である。意味するところは共著係数と似かよっており、この指標も、一方的な(有向の)リンクの強度を示している。

上に挙げた6つの指標のうち、共著頻度 n_{ij} は絶対リンク強度を示し、それ以外は相対リンク強度を示す。また、共著係数 c_{ij} と親和係数 a_{ij} は有向のリンク強度を、それ以外は無向のリンク強度を示す指標である。

² 任意の共著者ペアをとってきたとき、それが i と j の共著である確率は、 i, j の共著頻度 n_{ij} を、共著者ペアの延べ数 $\sum_{k,l} n_{kl}$ で割ることで得られる。一方、共著者ペアがランダムな組合せであると仮定したとき、 i と j の共著が現れる確率は、それぞれの出現確率の積 $\frac{n_i}{\sum_k n_k} \frac{n_j}{\sum_k n_k}$ が期待値となる。

いう結果を示している。

FCA (Factorial Correspondence Analysis) を使って、リンク強度のマトリクスを多次元座標に投影する手法もある。Elalami et al. (1992) は、FCA で決定した要因のうち最大の 2 つを座標にとって、その 2 次元座標上にそれぞれのグループをプロットし、グループ同士の近接性を可視化している(共著相手の傾向が似かよったグループ同士が、近くにプロットされる)。

MST (Minimum Spanning Tree) は、ノード(グループ)のプロフィールを多次元空間に投影したときのノード間の最短グラフを生成する手法である。例えば、Okubo et al. (1992) は、各国の国際共著の分野内訳をプロフィールにして、そのプロフィールから見た近接性を MST を使って分析し、国際協力における国同士の近接性を調べている。

5 おわりに

本研究では、共著ネットワークを分析した研究を紹介し、それらの視点や用いられた指標・手法を整理した。最後に、まとめとして、これまでの研究の問題点をいくつか指摘しておきたい。

まず、様々なリンク強度の指標が提案されているにも拘わらず、分析の目的に適した指標が必ずしも選ばれていないという点が挙げられる。相対強度を調べるべきところで絶対強度の指標を用いたり、有向のリンク強度を調べるべきところで無向のリンク強度の指標を用いている研究が少なくない。例えば、Bahr & Zemon (2000) は、著者の性別の組合せから見た図書館学分野の共著傾向を分析し、女性だけによる共著が最も多い(共著頻度が高い)という結果から、女性同士が共著しやすいという結論を導いている。しかし、もともと著者に女性が多いという分野の特徴を考慮に入れるならば、その影響が除かれる協力係数や相同係数などを用いた方が妥当であったと考えられる。

もうひとつの問題点は、各ノード間のリンク強度を求めた後の分析手法が、それほど洗練されていないことである。リンク強度をマトリクス表やネットワーク図に示すにとどまる研究が大半である。より客観的な分析結果を引き出せる手法の研究が今後必要であろう。例えば、2 つのネットワークを比較する場合、グラフのマッチング手法が応用できれば、視覚的な比較だけでは分からない傾向が読みとれるのではなからうか。

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.
- Arunachalam, S., Srinivasan, R. and Raman, V. (1994) "International collaboration in science: participation by the Asian giants," *Scientometrics*, Vol. 30, No. 1, p. 7-22.
- 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.
- Balog, C. (1980) "Multiple authorship and author collaboration in agricultural research publications," *Journal of Research Communication Studies*, Vol. 2, No. 3, p. 159-168.
- Braun, T., Gomez, I., Mendez, A. and Schubert, A. (1992) "International coauthorship patterns in physics and its subfields: 1981-1985," *Scientometrics*, Vol. 24, No. 2, p. 181-200.
- Cunningham S. J. and Dillon, S. M. (1997) "Authorship patterns in information systems," *Scientometrics*, Vol. 39, No. 1, p. 19-27.
- Elalami, J., Dore, J. C. and Miquel, J. F. (1992) "International scientific collaboration in arab countries," *Scientometrics*, Vol. 23, No. 1, p. 249-263.
- Godin, B. and Ippersiel, M. P. (1996) "Scientific collaboration at the regional level: the case of a small country," *Scientometrics*, Vol. 36, No. 1, p. 59-68.
- Kim, M. J. (1999) "Korean international co-authorship in science 1994-1996," *Journal of Information Science*, Vol. 25, No. 5, p. 403-412.
- 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.
- Lewis, G., Fawcett-Jones, A. and Kessler C. (1993) "Latin American scientific output 1986-91 and international co-authorship patterns," *Scientometrics*, Vol. 27, No. 3, p. 317-336.
- Lotka, A. J. (1926) "The frequency distribution of scientific productivity," *Journal of the Washington Academy of Sciences*, Vol. 16, No. 12, p. 317-323.
- Miquel, J. F. and Okubo, Y. (1994) "Structure of international collaboration in science 2: comparisons of profiles in countries using a link indicator," *Scientometrics*, Vol. 29, No. 2, p. 271-297.
- Narin, F., Stevens, K. and Whitlow E. S. (1991) "Scientific cooperation in Europe and the citation of multi-nationally authored papers," *Scientometrics*, Vol. 21, No. 3, p. 313-323.
- 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.
- Russell, J. M. (1998) "Publishing patterns of Mexican scientists: differences between national and international papers," *Scientometrics*, Vol. 41, No. 1-2, p. 113-124.