

【学術賞受賞者講演】

所属情報と論文共著情報による 研究者評価指標の提示とその有効性検証

－ 拡張型媒介中心性指標を用いた主要研究者の特定 －

[Excellent Paper Award (International Society of Management Engineers) 「A Study of the Use of Co-authored Information, Authors' Affiliation Information of Scientific Literature Using Centralities -The researchers who act as mediators between organizations-」 平成 27 年 9 月 1 日]

日大生産工 ○水上 祐治、小田部 明

1. はじめに

文部科学省は、各学問分野の研究の更なる発展のため、国内外の研究者による大学共同利用機関の資源と資産の共同利用促進を目指している。本稿は、その共同利用の促進とその効果確認について、客観的に評価できる新たな研究者評価指標が必要との立場にて、論文共著情報をもとにした新たな指標を定義、そして、その効果を統計的見地から分析するものである。

本稿では、文部科学省の統計数理研究所に所属する研究員が、2013 年に著者となる文献を対象として、論文共著情報を拡張型媒介中心性にて分析した。分析の結果、各研究分野の中心研究者の特定とその特徴を得ることができた。

2 従来研究

2.1 研究評価の分類

研究評価の方法には、文献の発表数を対象に分析する論文数調査の系統と、論文への引用度数や論文刊の引用による結合関係を対象に分析する引用統計・分析の 2 系統がある[1]。研究評価の分類を表 1 に示す。

前者の「論文数調査」の目的は、論文の生産性指数、研究活動の規模指数を求めることであり、主に単純集計の手法が用いられる。一方、後者の「引用統計・分析」の目的は、論文の消費指数、研究活動の品質指数を求めることであり、引用分析、共著分析等の手法が用いられる。

[1][2]。また、少数ではあるが、謝辞分析、共語分析、供分類分析等の手法が用いられる[2]。

分析手法において、前者の「論文数調査」は、学術雑誌掲載論文について、分野別、年次別、国別、機関別等にて集計すること、さらに、経済統計指数との相関分析等が行われている[1]。後者の「引用統計・分析」では、引用分析が代表例として挙げられる。この分析は、論文当たりの引用度数を、雑誌別や国別に集計して比較するものである。

2.2 共著分析とネットワーク中心性

本稿の共著分析では、ネットワーク中心性を指標として用いる。ネットワーク中心性とは、ネットワークを構成するノードが、そのネットワーク内でどの程度重要かを定量的に表す指標であり、次数中心性、近接中心性、媒介中心性の 3 つが代表例として挙げられる[3]。次数中心性とは「線が集まる点ほど中心性が高い」とする中心性の指標である。次の近接中心性とは「他の点と距離が近いほど中心性が高い」と

表 1：研究評価の分類

系統	目的	分析手法
論文数調査	論文の生産性指数 研究活動の規模指数	単純集計
引用統計・分析	論文の消費指数 研究活動の品質指数	引用分析 共著分析 (謝辞分析) (共語分析) (共分類分析)

根岸・山崎[1]と藤垣ら[2]をもとに筆者作成

Applying Affiliation Information and Literature's Co-authored Information, in
Making Researcher's Metrics

- Specifying Key Researchers by using Extended Betweenness Centrality Index -

Yuji MIZUKAMI and Akira OTABE

$$r_i = \sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i(i_s i_t)}{N_{i_s i_t}} \quad (2)$$

$$b_i = \frac{r_i}{c_i} \quad (1)$$

$$c_i = \begin{cases} \frac{(N-1)(N-2)}{2}, & \text{if 無向グラフ} \\ (N-1)(N-2), & \text{if 有向グラフ} \end{cases} \quad (3)$$

b_i : 媒介中心性 r_i : 媒介数
 c_i : 総リンク数 (ノード(i)を除く) i : ノード(i)
 i_s : 接続元ノード i_t : 接続先ノード
 $g_i(i_s i_t)$: ノード(i)を通る最短経路の数
 $N_{i_s i_t}$: ノード(i_s)とノード(i_t)の最短経路の数

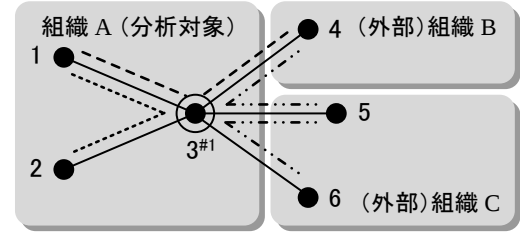
する中心性の指標である[3]。そして、本稿が分析で用いる媒介中心性とは「その点を通る経路が多いほど、中心性が高い」とする中心性の指針である[3]。媒介中心性が高いほどネットワーク内での情報や資源の流れに頻繁に関与しているノードになる。媒介中心性の式を式 1、式 2、式 3 に示す。

Abbasi ら[4]は、1999 年から 2009 年までの鉄骨に関する論文において、その共著情報を中心性指標にて分析した。分析の結果、コミュニケーションと情報の流れを司る著者(媒介中心性の高い人々)は、多くのつながりを持つ著者(次数中心性の高い人々)と多くの直接つながりを持つ著者(近接中心性の高い人々)に比べ、新しい共著者をより多く増やすことができるとしている。さらに、当該分野における媒介中心性の高い著者は、学生を指導する立場の人物であり、学生との共著、さらに、同僚との共著を行う人物ではないかと示唆している。

3. 研究フレームワーク

3.1 コミュニケーション理論の適用

Rogers[5]は、組織などのグループが、新たに革新的な知識、サービス、技術などのイノベーションを採用する場合の各段階とそれらに関連する人々の特徴を示した。そして、それら人々の特徴整理に、コスモポリト性、ローカ



#1: 媒介中心性を求める
 ローカル・リンク
 - - - - - コスモポリタン・リンク
 - - - - - ローカル & コスモポリタン・リンク

図 1: 拡張型媒介中心性指標の構図

ライド性の 2 つの概念を用いた。コスモポリト性とは、グループの外の情報に敏感である性向である。次に、ローカライド性とは、グループ内の情報に敏感である性向である。

普及の初期段階では、コスモポリト性の高い人々が、グループの外の情報をグループ内に持ち込む(コスモポリタン・リンク)。次に、第 2 段階に入ると、コスモポリト性とローカライド性の両方が高い人々が、グループの外の情報をさらに持ち込み、そして、グループ内に広める。そして、第 3 段階に入ると、ローカライド性の高い人々が、グループ内で普及をさらに加速する(ローカル・リンク)としている。

本稿では、国内外の研究者による大学共同利用機関の資源と資産の共同利用の客観的評価を目的として、媒介中心性指標にコスモポリト性とローカライド性の 2 つの概念を導入した拡張型媒介中心性指標を定義する。拡張型媒介中心性指標の構図を図 1、その計算式を式 4 から式 11 に示す。

拡張型媒介中心性指標は、式 1 の媒介中心性を式 4 の組織間媒介中心性(bd_i)、式 5 の組織外媒介中心性(boo_i)、式 6 の組織内外媒介中心性(bio_i)、式 7 の組織内媒介中心性(bii_i)に分割して個別に集計するものである。最初の組織間媒介中心性(bd_i)は、研究者(i)が、外部組織の研究者(i_s)と他の外部組織の研究者(i_t)を媒介す

$$bd_i = \frac{rd_i}{c_i} \quad (4) \quad boo_i = \frac{roo_i}{c_i} \quad (5) \quad bio_i = \frac{rio_i}{c_i} \quad (6) \quad bii_i = \frac{rii_i}{c_i} \quad (7)$$

$$rd_i = \sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i(i_s i_t; O_{i_s} \neq O_{i_t})}{N_{i_s i_t}} \quad (8) \quad roo_i = \sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i(i_s i_t; O_{i_s} \neq O_{i_t}; O_{i_s} \neq O_{i_s})}{N_{i_s i_t}} \quad (9)$$

$$rio_i = \sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i(i_s i_t; O_{i_s} \neq O_{i_t}; O_{i_s} = O_{i_s} \text{ or } O_{i_t} = O_{i_t})}{N_{i_s i_t}} \quad (10) \quad rii_i = \sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i(i_s i_t; O_{i_s} = O_{i_t}; O_{i_s} = O_{i_s})}{N_{i_s i_t}} \quad (11)$$

bd_i : 組織間媒介中心性 rd_i : 組織間媒介数 boo_i : 組織外媒介中心性 roo_i : 組織外媒介数
 bio_i : 組織内外媒介中心性 rio_i : 組織内外媒介数 bii_i : 組織内媒介中心性 rii_i : 組織内媒介数
 $g_i(i_s i_t; O_{i_s} \neq O_{i_t})$: O_i : ノード(i)の組織
 $g_i(i_s i_t; O_{i_s} \neq O_{i_t}; O_{i_s} \neq O_{i_s})$: O_{i_s} : ノード(i_s)の組織
 $g_i(i_s i_t; O_{i_s} \neq O_{i_t}; O_{i_s} = O_{i_s} \text{ or } O_{i_t} = O_{i_t})$: O_{i_t} : ノード(i_t)の組織
 $g_i(i_s i_t; O_{i_s} = O_{i_t}; O_{i_s} = O_{i_s})$: $g_i(i_s i_t)$: if $O_{i_s} = O_{i_t}$ and $O_i = O_{i_s}$

表 2：媒介中心性上位者の比較 (2013)

順位 著者	1 A	2 B	3 C	4 D	5 E
媒介中心性(b_i)	0.828	0.456	0.244	0.065	0.057
累積比率[%]	43.361	67.279	80.273	83.655	86.625
近接中心性	5E-6	5E-6	5E-6	5E-6	5E-6
次数中心性	59	39	18	9	9
共著論文数(編)	13	8	5	3	5

るケースである。図 1 では、ノード 4-3-5 が一例として挙げられる。次の組織外媒介中心性(boo_i)は、研究者(i)が、外部組織の研究者(i_s)と研究者(i_t)を媒介するケースである。図 1 では、ノード 5-3-6 が挙げられる。このように、組織間媒介中心性(bd_i)と組織外媒介中心性(boo_i)は、外部組織の研究者との共著に着目した指標であり、研究者(i_s)の外部組織との接触の度合いを示すと考えられる。

続いて、組織内外媒介中心性(bio_i)は、研究者(i)が、同組織の研究者(i_s)と外部組織の研究者(i_t)を媒介するケースである。または、その逆の外部組織の研究者(i_s)と同組織の研究者(i_t)も含める。図 1 では、ノード 1-3-4 が一例として挙げられる。この指標は、組織の内外の研究者を媒介するケースであり、本稿の目的である「大学共同利用機関の資源と資産の共同利用の促進とその評価」への適用が考えられる。

最後の組織内媒介中心性(bii_i)は、研究者(i)が、同組織の研究者(i_s)と研究者(i_t)を媒介するケースである。図 1 では、ノード 1-3-2 が挙げられる。この指標は、組織内の共同研究の度合いの指標としての適用が考えられる。

3.2 分析データと分析方法

分析データは、トムソンロイター Science Citation Index をデータソースとする統計数理研究所 ISM+Neo4J 論文検索システムを用いる。データ検索条件は、基本検索、検索ワードは「Inst Stat Math」、カテゴリは「著者所属」である。また、タイムスパンは「2013-2013」であり、502 名の研究者が含まれていた。

表 3：媒介中心性上位者の所属とその分類

順位 著者	所属 (2013 年)	分類
1 A	中部地方国立大学大学院医学系研究科 教授、統数研リスク解析戦略センター客員教授	指導的 立場
2 B	統数研 データ科学研究系軽量科学グループ 助教、リスク解析戦略センター 助教(兼)	複数部門 を兼務
3 C	統数研 名誉教授	指導的 立場
4 D	統数研 データ科学研究系構造探索グループ 助教、調査科学研究センター 助教(兼)	複数部門 を兼務
5 E	統数研 数理・推論研究系学習推論グループ 教授	指導的 立場

統数研：統計数理研究所

分析は 2 段階で行う。最初の段階では、Honda ら[7]の分析フレームワークに則り、媒介中心性が高い著者を特定して、それら人物の身分を確認する。次に拡張型媒介中心性にて、それら人物のコミュニケーションの特徴を示す。また、拡張型媒介中心性の追加の評価項目として、組織への影響度、組織への帰属度を用いる。組織への影響度は、媒介中心性全体(b_i)における組織内のコミュニケーションに係る媒介中心性($bio_i + bii_i$)の割合で定義、組織への帰属度は、組織内外媒介中心性(bii_i)と組織内媒介中心性(bio_i)の比率で定義した。

4 分析結果と考察

4.1 媒介中心性上位者の特定

統計数理研究所 2013 では、媒介中心性が 0 より大きい研究者は、51 名であり全体の 10.16%であった。表 2 に媒介中心性上位者(5 位まで)の各指標の比較を示す。表 2 では、媒介中心性とその累積比率を中心にして、近接中心性、次数中心性、共著論文数を示す。

最も高い媒介中心性を示したのは A 氏であり、媒介中心性が 0.828、全媒介中心性の 43.361%であった。また、上位 5 位までの著者の媒介中心性の総和は、全媒介中心性の 86.625%を占めており、本稿では、上位 5 位までの著者を分析対象とした。そして、近接中心性は、変化が小さいが、他の指標は、媒介中心

表 4：拡張型媒介中心性による上位者のコミュニケーション方法の比較 (2013)

順位 名前			1 A	2 B	3 C	4 D	5 E
媒介中心性		b_i^{-1}	0.828	0.456	0.244	0.065	0.057
拡張型 媒介中心性	組織外(&間)媒介中心性	$boo_i + bd_i$	0.673	0.369	0.195	0.052	0.051
	組織内外媒介中心性	bio_i	0.148	0.083	0.047	0.012	0.006
	組織内媒介中心性	bii_i	0.006	0.004	0.003	0.000	0.000
評価項目 [%]	組織への影響度	$(bio_i + bii_i)/b_i$	18.941	18.985	20.175	18.870	9.951
	組織への帰属度	bii_i/bio_i	4.362	4.489	5.917	0.000	0.000

性の高まりと共に増加する傾向を示した。(各中心性指標の相関は、Uddin ら[6]が詳しい。)

媒介中心性上位 5 位の人物について、各個人の身分を確認、さらに、それら情報をもとに属性の分類を行った。表 3 に媒介中心性上位者の所属とその分類を示す。なお、統計数理研究所以外の所属に関しては仮名を用いる。1 位の A 氏、3 位の C 氏、5 位の E 氏は、「指導的立場」、2 位の B 氏、4 位の D 氏は、統計数理研究所内の 2 つのグループに助教として所属しており、「複数部門を兼務」であった。

4.2 上位者のコミュニケーションの特徴

媒介中心性上位 5 位の人物について、拡張型媒介中心性指標を用いてそのコミュニケーションの特徴を分析した。表 4 に媒介中心性上位者の各拡張型媒介中心性指標の比較を示す。また、表 4 では、各種拡張型媒介中心性指標と共に、追加の評価項目として、組織への影響度、組織への帰属度を示す。

各拡張型媒介中心性指標は、媒介中心性の順位と共に高まる傾向を示している。

組織への影響度は、その平均が 17.336[%]、標準偏差が 3.729、最大は C 氏で 20.175[%]、最小は E 氏で 9.951[%]であった。この結果より、最大で 28.822[%](平均+3 標準偏差)は、組織内または組織内外のコミュニケーション、最小で 71.178[%]は、組織外または、組織間のコミュニケーション(著者(i)以外の組織内研究者は関わらない)と考えられる。次に、組織への帰属度は、その平均が 2.954[%]、標準偏差が 2.473、最大は C 氏で 5.917[%]、最小は D 氏と E 氏で 0.000[%]であった。この結果より、最大で 10.372[%]はローカル・リンク、最小で 89.628[%]はコスモポリタン・リンクと考えられる。つまり、媒介中心性の 71.178[%](最大)は、著者(i)以外の組織内研究者が関わらないものであった。一方、組織内研究者が関わるものにおいて、25.833 [%] (28.822 [%] * 89.628 [%]) は、著者(i)以外の組織内研究者が外部と関わるものであった。

5 まとめ

本稿は、文部科学省が進める大学共同利用機関の共同利用の促進とその効果確認について、客観的に評価できる新たな研究者評価指標が必要との立場にて、論文共著情報をもとにした拡張型媒介中心性指標を新たに定義、そして、その効果について、統計数理研究所(2013)を用いて統計的見地から分析するものである。

分析の結果、Abbasi ら[4]が示唆した「媒介中心性の高い著者は、学生を指導する立場の人物であり、学生との共著、さらに、同僚との共著を行う人物である。」に関して、本稿では、これをサポートする「指導的立場」を検証することと、新発見事項として、新たに「複数部門を兼務」という属性を示すことができた(表 3 参照)。さらに、本稿では、拡張型媒介中心性指標を用いて、それら著者のコミュニケーションの特徴を明らかにした。つまり、それら著者のコミュニケーションにおいて、71.178[%](最大)は、他の組織内研究者が関わらないものであった。一方、他の組織内研究者が関わるものにおいて、25.833 [%] (28.822 [%] * 89.628 [%]) は、他の組織内研究者が外部の研究者と関わるのを仲介するものであった。

謝辞

本研究は、統計数理研究所共同研究プログラム (27-共研-1028 学術データベースを利用した評価指標の開発)の助成を受けたものです。

本研究を進めるにあたり、ご助言を賜りました統計数理研究所 中野純司教授、本多啓介博士、情報通信医学研究所 鈴木重徳博士には、誠意を表わすとともに厚く御礼申し上げます。

「参考文献」

- [1] 根岸正光、山崎茂明、「研究評価—研究者・研究機関・大学におけるガイドライン」、丸善、2001
- [2] 藤垣裕子、平川秀幸、富澤宏之、調麻佐志、林隆之、牧野淳一郎、「研究評価・科学論のための科学計量学入門」、丸善、2004
- [3] L. Freeman, "Centrality in social networks: Conceptual clarification," Vol.1, Issue 3, pp. 215-239, 1979
- [4] A. Abbasi, L. Hossain, L. Leydesdorff, "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks", Journal of Informetrics, Vol. 6, Issue 3, pp. 403-412, 2012
- [5] M. Rogers, Diffusion of innovations (5th ed.), New York: Free Press, 2003
- [6] S. Uddin, L. Hossain, A. Abbasi, K. Rasmussen, "Trend and efficiency analysis of co-authorship network", Scientometrics, Vol.90, Issue2, pp.687-699, 2012
- [7] K. Honda, Y. Mizukami, S. Suzuki and J. Nakano, "An approach for evaluating academic research performance using betweenness centrality in authors networks", 3rd International IBM Cloud Academy Conference 2015, Proc. p.34, Hungary