

研究者情報サーバの構築

– ネットワーク構造可視化と解析の試み –

高久雅生[†] 相澤彰子[‡] 大山敬三[‡]

[†] 情報・システム研究機構 新領域融合研究センター

[‡] 国立情報学研究所 コンテンツ科学研究系

{masao, aizawa, oyama}@nii.ac.jp

概要

筆者らは、国立情報学研究所が提供している科学研究費補助金データベースの成果報告書、実績報告書を元にした研究者情報提供サーバの構築を行っている。本稿では特に、これらのデータセットから得られる、科研費研究代表者、研究分担者による共同研究ネットワークを取り上げ、これらのネットワークの持つ特性の分析結果の報告、およびその構造の可視化ブラウジング機能について述べる。

1 はじめに

筆者らは、科学研究費補助金データベース^{*1}を元にした研究者情報サーバの構築を行っている。2006年10月には、そのプロトタイプである研究者ブラウジングツールの開発について報告した [1]。本稿では、現在開発中の本プロトタイプの研究者ネットワーク可視化機能を取りあげ、その可視化手法とともに、研究者ネットワーク全体の構造、特性について考察する。

研究者ネットワークの分析の源流は、図書館情報学の一分野である計量書誌学にもとめられ、書誌流通における論文の引用分析などを中心として、論文間の関連や研究者や研究機関、掲載論文誌などとの関連を明らかにするものであった。一方、近年には、これらの計量書誌学の知見に加え、ネットワーク理論および様々な分野の情報流通において、複雑なネットワークの持つ特性に注目があつまり、スモールワールド現象、スケールフリーネットワークなどの複雑ネットワークの分析についての研究が盛んに進められている [2][3][4]。例えば、スモールワールド現象の説明として頻繁に取り上げられる「エルデシュ数」は、著名な数学者ポール・エル

デシュとの共著関係を持つ研究者をネットワークとしてとらえ、その最短距離を見つけるものである。Newman[5] は実際の引用論文から生成される研究者ネットワークを取り上げ、いくつかの分野の引用文献データベースのデータを用いて、そのネットワーク特性について分析している。

一般に、人と人との関係は家族や友人、知人といった個人的な関係から、職種、興味などに応じた関係まで幅広く、それぞれの人が持つ他の人々との関連は社会的ネットワークと呼ばれ、多くの研究がなされてきた。とりわけ近年では、先に述べた複雑ネットワークの特性分析における応用としても注目されている。また、インターネット上でのソーシャルネットワークの構築を支援するツールとして、ソーシャルネットワークサービス (SNS) と呼ばれるサービスも普及し、日本では SNS サイトの代表格 Mixi^{*2}が 800 万人を越える会員数を得るなど、一般的なツールとして認識が進んでいる [6]。また、松尾らは研究者のソーシャルネットワークを Web 上の情報から自動生成するシステム POLYPHONET を開発している [7]。

学術研究の分野でも古くから、このような人と人とのつながりに着目し、ソーシャルネットワークを

^{*1} <http://seika.nii.ac.jp/>

^{*2} <http://mixi.jp/>

含む、研究者間の関係性に着目した研究が行われている。最も広く知られているものの一つに、共著関係や引用論文の分析がある。共著関係の分析は、書誌データベースなどから論文、著書を共同執筆したものの同士の関係を抽出し、ネットワークとしてモデル化することで、当該研究者の持つ人的関係や、周りの研究者とで作り出すネットワーク構造により、分野ごとの研究者のつながりを発見することができ、研究者が持つインパクトや、ひいては当該研究者が執筆した論文のインパクトとして、科学研究の発展の方向性や流過程を分析することができる。また例えば、ある研究テーマが与えられたときに、その分野において主要な研究者やキーとなる論文を特定し、その分野において顕著な業績を挙げたものを並べたりすることにより、その影響の範囲を探ることができ、科学研究の進展の理解がたやすくなる。

科研費は政府文部省系列の研究助成金であり、質・量ともに国内最大規模である。科研費においては、研究者一人一人につき固有の研究者番号が申請時に付与される枠組みがあり、報告書の基本データとの対応を取ることにより、上記の基本情報のうち、氏名、所属、職位、研究分野、発表文献と、各課題との対応を得ることができる。さらに、研究者個人に留まらず、各研究課題における研究分担者の情報を合わせて得ることにより、ある研究課題に参画した研究者間の関係を得ることができる。

このような研究者ネットワークを見る際には、まず大域的なネットワーク構成の様子と、局所的な各個人レベルでのネットワークの様子それぞれを眺める必要がある。しかし、ある程度以上の規模のネットワークの表示は、人がその構造を把握するのも困難であり、また、計算機によるそのネットワークを見易いように自動的にレイアウトしていくのも理論的にも技術的にも難しい問題であることが知られている。

そこで、本研究のプロトタイプシステムでは、ばねモデルに基づくインタラクティブな操作可能なネットワーク可視化システムを採用して個人間の関係の提示を行い、人物情報と、ネットワークとを

シームレスにつなぐことを考え、システムを構築した。次章以降では、研究者ネットワークのモデルとその有用性について議論し、プロトタイプシステムの概略を述べる。

2 研究者ネットワーク

科研費の研究課題において研究者がある研究課題に参画する場合、2つの種類の役割を持つ。「研究代表者」と「研究分担者」である。各研究課題においては研究代表者1名が主たる研究を進める責任者として研究課題の統括を行って、研究課題の進捗に責任を持つとともに、交付される研究補助金の管理も行う。研究分担者は、代表者および他の分担者と連携しながら研究を進め、共同研究としての研究に参画する者であり、課題によっては所属や分野の異なる複数の共同研究者から構成されることもある。これらの代表者や分担者は、研究の進捗や研究組織の異動などの事情に伴って、年度毎もしくは年度途中で入れ替わりが生じることもある。また、科研費申請のルール上、一年間に同時に応募できる研究種目が限られていたり、「若手研究」などのように研究代表者が一人だけで行い、分担者を必要としない研究種目があるなどの制度上の制約も存在する。

これらの代表者・分担者からなる研究者間の関係は、研究課題を介した研究者ネットワークとしてとらえることができる。本研究では、研究者をノード、各研究課題において共同研究を行った研究者同士の間リンクが張られるような、研究者ネットワークを想定する。この関係をネットワークととらえる際には、以下の2種類の研究者ネットワークのモデルが考えられる。

1. 参画研究者集合モデル: 研究課題に参画した研究者は、代表者・分担者の別なく同等にネットワークを形成するモデル。
2. 研究代表者集中モデル: 各研究課題において、研究代表者が文字通り共同研究を主導する立場にあり、それぞれの分担者は代表者を通じてのみ、つながるモデル。

ただし、本稿では分析の都合上、前者の集合モデル

表 1 全データにおける連結成分のノード数

課題数	研究者 (n)	連結成分	成分のノード数 (1 位 ~ 10 位) ($n_{c_1}, \dots, n_{c_{10}}$)										n_{c_1}/n		
			108,485	34	31	30	26	26	21	20	20	19			
247,745	133,067	2,911													81.53%

のみ扱うこととし、以下ではこれに対する分析、可視化結果を示す。

また、いずれのモデルも研究者間の関係を示すリンクには、過去のいくつかの課題で研究者間の関係が持たれたかに応じてリンクに重みを付けることができる。すなわち、複数の課題を通じて、2 人の研究者がたびたび共同研究の関係を持っていたとすれば、単一の課題でのみ結び付けられるような研究者間の関係に比べて、強い共同研究関係にあると言える。

3 研究者ネットワークの分析

まず、研究者ネットワークの分析として、連結成分からなるサブネットワークのサイズによる分析結果について報告する。

対象としたデータは主として、科学研究費補助金の 1989 年度から 2003 年度まで約 15 年分の実績報告書および成果報告書である。これら報告書に記載された代表者・分担者情報を手掛かりに、ネットワークモデルを構築した。この際、着任直後やもともと番号の割り当てられない研究者等の事情にともない、報告書に研究者番号が付与されない研究者のあることがあるが、研究者番号が振られていない研究者の分の情報については無視することとした。こうしてできた研究者ネットワークは、計 133,067 名分の研究者からなる。

連結成分は、任意の研究者ノードから出発し、当該研究者が持つリンクから辿れる範囲のサブネットワークを抽出したものである。表 1 に、生成されたネットワークの各連結成分が持つノード（研究者）数が上位 10 位までのネットワークを示す。驚くことに、最も大きな連結成分からなるサブネットワークは、研究者数 10 万人を越え、本データセット中の約 81.5% の研究者を含む、巨大なネットワークとしてつながっていることが分かる。

表 2 連結成分ノード数上位 2 位から 10 位まで

i	n_{c_i}	サブネットワーク内の研究者所属
2	34	和歌山県立医科大, 山梨医科大
3	31	日本歯科大学
4	30	愛知医科大学
5	26	岐阜大学
5	26	自治医科大学, 筑波大
7	21	東京慈恵会医科大
8	20	長崎大歯学部
8	20	佐賀医科大
10	19	島根医科大, 島根大学
10	19	東京慈恵会医科大
10	19	北里大, 広島大, 厚生連尾道総合病院

また、成分ノード数 2 位から 10 位までの連結成分を目視により確認したところ、研究者の所属がほぼ全て医歯薬学系の学部・研究科等からのみ構成されることが分かった（表 2）。なお、科研費全体を分野別に見た場合、医学系の占める割合が最も高いこともあり、これらの特性が医歯薬学系の研究者のネットワークに特有の現象を示している可能性もあるが、これらの結果の詳細な分析はまだ行っており、詳細は不明である。ネットワーク全体の分析から、分野特有のなんらかの興味深い構造が発見できる可能性も示唆される。

さらに、これらを時系列で分析してみる。まず、単年度ごとの報告書のみを取り出し、それぞれの各年度のみにおけるネットワークを生成し、連結成分を抽出した結果を表 3 に示す。また、これを年度累積で見た例も同様に表 4 に示す^{*3}。単年度だけをとってみても、40% から 45% 程度を最大のサブネットワークが占めている構成になるのが分かる。

^{*3} 表中の研究者数、課題数は表 1 の値と異なるが、これは年度集計の際には 1989-2003 年度の範囲外であった一部課題を除去したためである

表3 単年度ごとの研究者ネットワークにおける連結成分のノード数の推移

年度	課題数	研究者 (n)	連結成分	成分のノード数 (1位~10位) ($n_{c_1}, \dots, n_{c_{10}}$)										n_{c_1}/n
1989	12,766	26,034	3,456	11,048	38	37	28	28	28	25	21	21	20	42.44%
1990	14,269	28,603	3,577	12,960	117	32	28	26	25	25	23	23	23	45.31%
1991	15,138	29,681	3,739	13,249	54	49	39	35	34	32	31	30	30	44.64%
1992	16,228	30,821	3,951	13,554	125	72	39	31	26	25	24	24	24	43.98%
1993	22,531	35,938	4,239	14,434	71	52	52	34	32	30	29	28	27	40.16%
1994	26,726	39,437	4,652	15,414	44	37	35	33	28	27	27	26	25	39.09%
1995	30,338	43,351	4,939	17,789	58	35	31	31	31	30	29	28	27	41.03%
1996	31,368	45,022	4,986	19,473	53	44	39	38	36	32	32	30	30	43.25%
1997	31,441	45,422	5,428	18,008	97	72	60	58	48	46	37	33	32	39.65%
1998	35,970	49,295	5,531	19,538	41	40	40	39	38	37	37	34	31	39.63%
1999	36,138	51,154	5,451	22,029	53	42	42	38	35	32	32	28	25	43.06%
2000	38,100	52,636	5,389	23,744	53	50	47	45	34	33	32	31	30	45.11%
2001	39,221	54,189	5,492	24,470	36	36	31	29	28	27	26	26	26	45.16%
2002	44,143	55,852	5,631	25,853	56	49	47	43	42	39	36	34	31	46.29%
2003	45,005	57,682	5,804	26,521	53	44	37	37	34	33	32	31	31	45.98%

表4 年度累積の研究者ネットワークにおける連結成分のノード数の推移*3

年度	課題数	研究者 (n)	連結成分	成分のノード数 (1位~10位) ($n_{c_1}, \dots, n_{c_{10}}$)										n_{c_1}/n
1989	12,766	26,034	3,456	11,048	38	37	28	28	28	25	21	21	20	42.44%
1989-1990	21,669	36,115	3,625	19,708	32	26	26	26	26	26	25	22	22	54.57%
1989-1991	30,499	43,216	3,658	26,362	32	29	27	27	27	26	25	23	23	61.00%
1989-1992	40,505	49,770	3,670	32,602	46	28	28	24	23	22	22	22	21	65.51%
1989-1993	56,115	59,073	3,676	39,283	30	29	28	27	23	22	22	20	20	66.50%
1989-1994	75,297	67,868	3,710	45,878	35	25	25	24	24	21	20	19	19	67.60%
1989-1995	96,506	76,476	3,642	53,129	27	26	25	22	22	20	20	19	19	69.47%
1989-1996	116,487	84,057	3,617	59,902	30	26	24	23	22	22	19	19	18	71.26%
1989-1997	136,071	90,991	3,535	66,437	28	25	23	23	20	20	19	18	18	73.01%
1989-1998	152,985	97,826	3,400	73,112	25	24	23	23	21	21	21	19	19	74.74%
1989-1999	170,677	104,782	3,213	80,310	30	25	23	22	22	21	21	20	19	76.64%
1989-2000	187,824	111,576	3,074	86,860	35	30	24	23	21	20	20	19	19	77.85%
1989-2001	205,691	118,524	2,987	93,675	30	26	23	21	21	20	19	18	18	79.03%
1989-2002	227,767	125,755	2,940	100,585	30	26	23	23	21	20	19	18	18	79.98%
1989-2003	247,705	133,043	2,836	107,681	30	29	26	26	21	20	20	19	19	80.94%

また、年度累積でも、最初の3年度分のデータによるだけでも、全体の6割の研究者を持つ巨大なネットワークが構成されることが分かる。

上記の結果から、科研費採択課題を持つ研究者の大半は、なんらかのつながりで他の研究者全体とつながりを持っており、それらの関係をつないでいくだけでも、大半の研究者を辿れるという構造を持つことが分かる。

4 研究者ネットワークの可視化

本節では、研究者情報サーバにおける、研究者ネットワークの可視化およびその情報を理解しやすくするためのブラウジング機能の強化について述べる。

図1に研究者情報サーバの画面を示す。研究者情報サーバの詳細については、以前の融合センター研究会で発表済であるので、説明は省略する [1]。

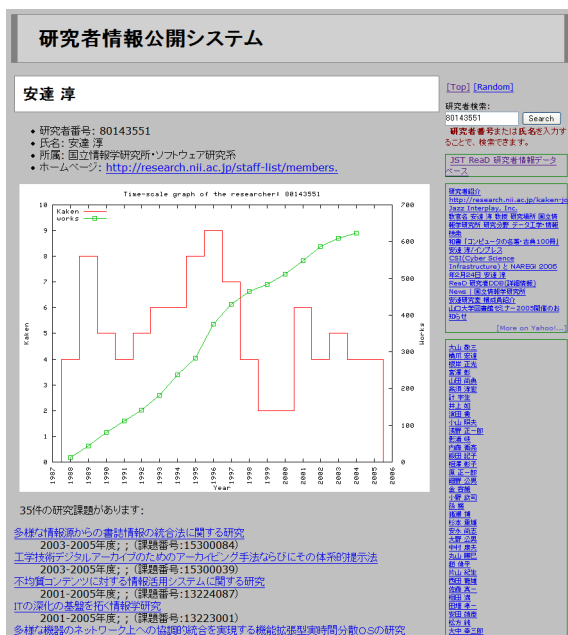


図1 研究者情報表示画面

共同研究者ネットワークの表示機能として、共同研究ネットワークにおいて隣接する研究者の一覧をその重みの順にならべ、画面右側に表示している(図1参照)。これらの研究者の一覧を見ることにより、当該研究者が他のどの研究者と多く共同研究を行っているかを眺めることができ、また、その氏名のリンクをたどることで、実際にその研究者の情報を知ることができる。また、可視化機能として、これらのリストの末尾に「共同研究者ネットワーク」と書かれたリンクをたどると、研究者ネットワークを可視化したグラフを見ることができる。

研究者ネットワークの可視化はJava アプレットにより実装し、ばねモデルにもとづくネットワークレイアウトにより、動的にネットワークの構成を確認することができる(図2)。各研究者の氏名が示されたノードをクリックして選択し、そのままドラッグして動かすことで、その研究者のノードを移動することができ、その研究者に付随するリンクがある場合は、それらのネットワークも合わせて適宜配置が変更されるため、ある一人の研究者の情報を選択して、それらを動かすことにより、その研究者の影響の大きさが直感的に把握できる。

また、研究者情報サーバで保持している研究者ネットワークの情報は GraphML[8] として XML 出力し、外部アプリケーションと容易に連携できるような機能を開発している。これにより、これらを外部から呼び出して動的なグラフ可視化を行うツールを作成できる。

5 考察

本稿では、科研費を通じた研究者間のつながりとその可視化の可能性を示した。

科研費における研究課題を通じた研究者ネットワークのつながりが単一分野ごとに隔絶せず、多くの分野の研究者をつなげているなど、興味深い構造をしていることが分かった。これらの構造は科研費などの競争的資金獲得のための研究活動の結果、生まれてきたものと推定できる。今後、これらの構造のさらなる分析を通じ、科研費を通じた研究者ネットワークの特性と、その要因をさぐっていく予定である。また、これらのネットワークの特性は、科研費の枠を越えた国内の研究者間の共同研究そのものの分析にも資するものであり、今後、これらのデータを活かして研究者ネットワークの利用による共同研究先機関等の推薦といったより研究活動を活性化するような機能を拡充していきたい。また、本稿で示したようなネットワークのマクロな特性だけでなく、細部における構造も今後の課題である。個々の研究者におけるつながりがどのように全体のネットワークにつながっているかを適切に提示し、それぞれの研究者の位置付けを示すような方向での展開を考えている。また、これらの他にも、研究者間のつながりを別のデータベース・情報源のものと統合する試みなども重要となってくる。

可視化においても、ブラウジング機能と連携し、さらに外部アプリケーションからも利用できるような機能を開発した。現在これらはプロトタイプ段階であるが、より多くの情報を適切に表示できるような表示上の工夫を加えるとともに、ユーザの操作に応じて、指定された、表示したいネットワーク箇所だけを細部に渡って展開したり折り畳んだりできるようなインタラクティブ性の向上に取り組む予定で

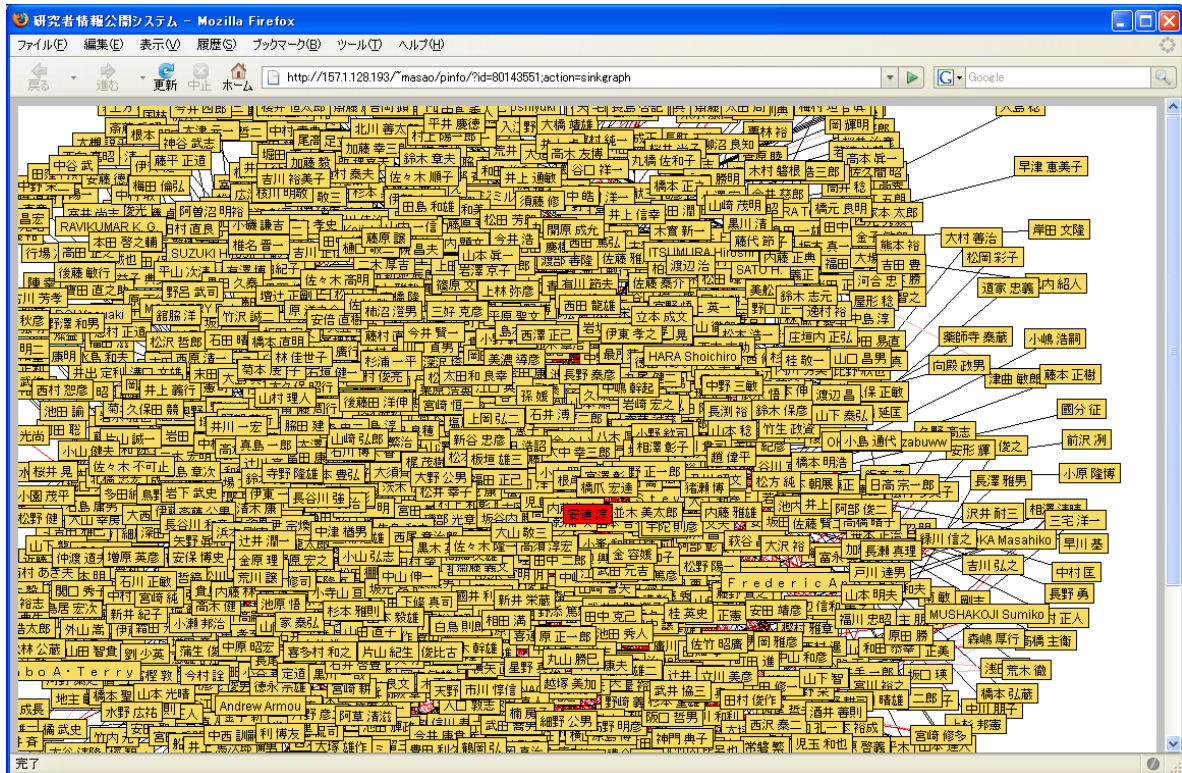


図 2 研究者ネットワーク可視化画面

ある。これらの詳細については、研究会においてデモを行う予定である。

参考文献

- [1] 高久雅生, 相澤彰子, 大山敬三. 科研費データベースにもとづく研究者情報ブラウジングツール. 「大規模データ・リンケージ、データマイニングと統計手法」研究会, pp. 89–96, 東京, 統計数理研究所, October 2006.
- [2] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, Vol. 45, No. 2, pp. 167–256, 2003.
- [3] アルバート・ラスロ・バラバシ. 新ネットワーク思考 世界のしくみを読み解く. NHK 出版, 2002. 青木薫 (訳) (ISBN 4-14-080743-1).
- [4] 松尾豊. スモールワールドとチャンス発見. 人工知能学会誌, Vol. 18, No. 3, pp. 288–294, 2003.
- [5] M. E. J. Newman. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E*, Vol. 64, No. 1, p. 016131, Jun 2001.
- [6] ミクシィ. SNS 『mixi』、ユーザー数が 800 万人を超える. オンライン, 2007-01-29. <http://press.mixi.co.jp/press-070129.html>.
- [7] Yutaka Matsuo et al. POLYPHONET: An advanced social network extraction system from the web. In *Proceedings of the 15th International Conference on World Wide Web*, pp. 397–406, 2006.
- [8] Ulrik Brandes, Markus Eiglsperger, and Jürgen Lerner (Eds.). GraphML Primer. オンライン, (accessed on 2007-02-13). <http://graphml.graphdrawing.org/primer/graphml-primer.html>.

Construction of a Researcher Information Server

– An Attempt to Analyse and Visualize Researcher Networks –

Masao Takaku[†] Akiko Aizawa[‡] Keizo Oyama[†]

[†]Transdisciplinary Research Integration Center,
Research Organization of Information and Systems
[‡]Digital Content and Media Sciences Research Division,
National Institute of Informatics
{masao, aizawa, oyama}@nii.ac.jp

Abstract

Today, researcher information is essential for academic communication. The Grants-in-aid for Scientific Research (*Kakenhi*) is one of the largest research fund in Japan. National Institute of Informatics (NII) has built and served Kakenhi database service based on the outcome reports on Kakenhi fund. We have built a researcher information server based on the DB, and reported its browsing features at the previous symposium. In this paper, we report an analysis on a network of researchers in Kakenhi DB, and a visualizing feature in our server.

From our analysis on connected components, it is shown that the largest component has most of researchers (81.5%). In an analysis on subset for each year, each of the largest component has 40–45% of researchers for that year. This result shows interesting nature of Kakenhi DB.

At our server, a visualization feature on researcher networks is also developed. This feature includes simple graph layout tools and browsing capabilities. Demonstration on these features will be shown at the symposium.