

【原著論文】

引用データに基づく自然科学領域における学術研究分野間の関係

天野晃<sup>\*</sup>, 児玉閑<sup>\*\*</sup>, 柴田大輔<sup>\*\*\*</sup>, 小野寺夏生<sup>\*\*\*\*</sup>

<sup>\*, \*\*, \*\*\*, \*\*\*\*</sup> 筑波大学

<sup>\*</sup> 農業生物資源研究所, 理化学研究所 BRC

<sup>\*\*</sup> 東邦大学

<sup>\*\*\*</sup> 帝京平成大学

<sup>\*</sup>kamano@affrc.go.jp, <sup>\*\*</sup>kodamat@mnc.toho-u.ac.jp, <sup>\*\*\*</sup>daisuke.shibata@thu.ac.jp

<sup>\*\*\*\*</sup>onodera@slis.tsukuba.ac.jp

要約 目的: 本研究では, 自然科学領域における各分野間の影響関係を, 大きな粒度の分野分類に基づいて探る. ここでの自然科学領域とは, 工学, 医学, 農学等の応用分野も含む. 本研究で用いる分類では, 分野間の引用が全体の引用の 3 割程度であり, 本研究では, この分野間の引用関係について詳細な分析を行うことにより, 自然科学領域の研究構造の理解を深めることを目的とする. データ: 本研究での対象は, Journal Citation Reports (JCR) 2004 Science Edition CD-ROM 版に収録されている雑誌の掲載論文の引用データを, 分野間の被引用/引用行列に編集しなおしたデータである. 該当する雑誌は, 2000 年から 2004 年に出版され, 2004 年に少なくとも 1 回の引用を受けた論文を含む雑誌 5964 誌となる. 方法: まず, 既存の 22 分野の分類を修正し, 21 分野の区分を作成した. これは, 既存の分類では非常に小さな分野が存在したためである. 21 分野の区分をもとに, 分野間の被引用/引用行列を作成し, このデータを次の分析の対象とした. 1. 各分野の他分野引用/被引用の程度. 2. 分野間の階層クラスタ関係. 3. 分野間の影響関係 (分野間の相互結合と非対称的な影響関係). 4. 各分野が他の分野に及ぼす全体的な影響強度 (Eigenfactor の手法を応用した全領域における影響度指標を用いた分析). 結果: 論文数規模は, 医学生物学関連の分野が大きく, さらに, 論文あたり被引用数との Spearman 順位相関の分析等から, 論文数規模と影響度の相乗効果が認められた. また, 論文数規模と自分分野引用を除いた被引用/引用比の間にはわずかに相関が認められた. 各分野の被引用傾向に基づいた分野間階層クラスタ関係からは, 分析化学が農学に近いなどの, 物理系, 化学系というような従来型の関係とはやや異なる様相が確認できた. それぞれの分野間の相互結合の強さも, 概ね, 階層クラスタ関係を支持する結果となったが, 物性物理-化学は, 階層クラスタ関係では近い位置にあるが, 相互結合としては強いものではないなど, やや異なる傾向も見られた. Eigenfactor の手法を応用した影響強度の分析では, 論文数の寄与と引用の寄与を調節するいわゆる  $\alpha$  を変動させたが, これによる順位変動が見られ, 分野という大きな区分を用いても, このパラメーターの影響を受けることが解った.

Interdisciplinary Relations among Scientific Research Fields in The Natural Science Area  
Based on Citation Data

Kou AMANO\*, Tadashi KODAMA\*\*, Daisuke SHIBATA\*\*\*, Natsuo ONODERA\*\*\*\*

\*,\*\*,\* \*\*\*,\*\*\*\*University of Tsukuba

\*National Institute of Agrobiological Sciences, RIKEN BRC

\*\*Toho University

\*\*\*Teikyo Heisei University

\*kamano@affrc.go.jp , \*\*kodamat@mnc.toho-u.ac.jp , \*\*\*daisuke.shibata@thu.ac.jp

\*\*\*\*onodera@slis.tsukuba.ac.jp

**Abstract** Purpose: We analyze the mutual impact relationships among various research fields in the Natural Sciences using a field classification with a large granularity. Here, Natural Sciences include application fields such as engineering, medicine, and agriculture. On the basis of the classification used in this study, inter-field citations account for approximately 30% of the total citations. The objective of this study is to reveal the research structure of the Natural Science field by analyzing in detail the inter-field citations. Data Source: The data source of this study is Journal Citation Reports (JCR) 2004 Science Edition (CD-ROM). From this data source, we selected 5964 journals that contained articles published in 2000–2004 and cited at least once in 2004. Using the selected journals, we constructed an inter-journal citation matrix, from which we then constructed an inter-field citation matrix. Method: We used the field classification consisting of 21 fields; this classification is the same as that developed in our previous study, except that this classification united an extremely small field to the nearest field. The inter-field citation matrix based on these 21 fields was used for the analyses of the following indexes: 1. The extent that citations in the individual fields were made from/toward other fields. 2. Hierarchical relation of the 21 fields. 3. Impact strength between each field pair (symmetric/asymmetric linkage). 4. Total influence by the individual fields on other fields (based on the field Eigenfactor indicator). Results: In general, biomedical fields had a larger size than others, in terms of the number of articles. A synergistic effect was recognized between the size of a field and the impact of the articles in the field, by considering the Spearman rank correlation between them. There was a weaker synergistic effect between the size of a field and the cited/citing ratio of the field with other fields. The result of hierarchical clustering among the fields based on their citation patterns revealed interesting information about the inter-field relationships; for example, analytical chemistry was found to be nearer to agriculture than to other chemical fields. The result of symmetric linkage mostly supported the hierarchical structure; however, we observed some differences; for instance, the similarity between the condensed matter physics field and the chemistry field was the highest in the hierarchical structure but not so high in the symmetric linkage. In the analysis of the field influence based on the Eigenfactor, a large influence of biomedical fields on all other fields was observed. Additionally, we analyzed the effect of  $\alpha$ , which tunes the influence ratio between publications and citations, and observed that  $\alpha$  is an important parameter for determining the influence ranking of the fields.

## 1 背景・目的

本研究では、自然科学領域における各分野間の影響関係を、分野間の引用関係から探る。ここでの自然科学領域とは、物理学、化学、生物学等の基礎的自然科学以外に、工学、医学、農学等その応用分野も含むものとする。

分野間の影響関係を論ずるにはいくつかの方法が考えられるが、引用の流れを分析することが最も直接的である。Zhang ら [1] は、引用分析は分野間の情報の流れや、科学の構造とその動的变化を捉えるのに有用であると述べている。

分野間の被引用/引用関係を調査した先行研究はいくつかある [2, 3, 4, 5, 6, 7] が、その被引用/引用関係を検討する分野区分は、Thomson Reuters 社の Journal Citation Reports が定める約 200 の主題カテゴリーをそのまま用いるなど、非常に細かく設定される場合が多く、構造の細部を見ることにおいては有利であるものの、自然科学の全体的な構造に対して十分な理解が得られているとは言いがたい。

分野間の関係を大局的に論ずるには、より粒度の大きい分野区分が必要である。その場合の最適な分野の数を論理的に定めることは難しいが、多くとも 30 を超えない程度が、直観的理解が可能な上限であると考えられる。分野間影響関係だけでなく、自然科学の全体的構造の把握、国や機関の活動度比較などの研究でも、そのような粒度の大きい分野区分が適当である。Thomson Reuters 社の Essential Science Indicators (ESI) における分類、Scopus の第一階層雑誌分類で用いられる区分が 20 余りであることも、このことを裏付けていると考えられる。

このように比較的大きな粒度で分野分類を行った先行研究には以下のものがある。Granzel ら [8] は、「認知的」アプローチにより雑誌に対し分野決定を行った。Zhang ら [1] は、雑誌の被引用/引用行列に基づきアルゴリズム的に各分野を決定する方法を提案している。Chen [9]、および、Rosvall ら [10] は、被引用/引用のネットワークを解析するこ

とにより、雑誌の分類を行った。天野ら [11] は、雑誌の被引用数をもとに特徴ベクトルを算出し、これをクラスタリングすることによって分類を行った。Leydesdorff ら [12] は、階層分類法を用い、人文社会科学各分野の雑誌と主題とのマッピングを行った。以上のうち、Granzel ら [8] 以外は被引用/引用データに基づき分野分類を行っている。

以上の理由から、本研究で用いる分野分類は、その区分数を 20-30 程度とし、分類の根拠には引用関係を用いる。紹介した先行研究のなかで、利用可能と思われる分類はいくつかあるが、天野ら [11] が研究で用いた分野分類のデータが入手可能であったため、本研究ではこれを用いることとした。詳細は後述するが、天野らの分類によれば引用全体の約 7 割が分野内の引用であり、残りの約 3 割が分野間の引用である。分野内引用が 7 割を占めることは、この分類がかなり妥当であることを示すと考えられるが、3 割の他分野引用は無視できる数字ではない。本研究では、この分野間の引用関係について詳細な分析を行うことにより、自然科学領域の研究構造の理解を深めることを目的とする。具体的には、以下に挙げる項目を分析する。

1. 各分野の他分野引用/被引用の程度
2. 分野間の階層クラスタ関係
3. 分野間の影響関係
4. 各分野が他の分野に及ぼす影響強度

## 2 データ

本研究での対象は、Journal Citation Reports (JCR) 2004 Science Edition CD-ROM 版に収録されている雑誌 (これらは、冒頭に述べた自然科学領域にほぼ合致すると考えられる) の収載論文の引用データを、分野間の被引用/引用行列に編集しなおしたデータである (3.1, 3.2 にて詳述)。

該当する雑誌は、2000 年から 2004 年に出版され、2004 年に少なくとも 1 回の引用を受けた論文を含む雑誌 5964 誌となる。これらは、天野らの分野分類 [11] に用いられたものと全く同じである。

該当する分野は、天野らの分類による 22 分野をほぼそのまま用いたが、後述するように、信頼性工学の分野を数理工学の分野に統合し 21 分野とした。

### 3 方法

#### 3.1 天野らの分類

ここで、天野らによる雑誌分類の方法 [11] を概観する。

天野らは、まず、JCR 2004 Science Edition のデータに基づき、雑誌間被引用/引用行列を算出した(被引用/引用の関係が特定され得る雑誌、5964 誌が得られた)。基になるデータは、各雑誌の 2000-2004 年に出版された論文が 2004 年に得た引用数である。上記行列における各被引用ベクトルに注目し、これをもとに非階層クラスタリングを行った。クラスタリングに用いたベクトルは、元の被引用ベクトルを基に、自誌引用の影響を緩和する操作(自誌被引用数を、次に大きい他誌からの被引用数に置き換えた)を行い、さらに全てのベクトルの要素を要素和に対する比で表したものの(被引用比率ベクトル)である。クラスタリングでは、コサイン距離(コサイン類似度を 1.0 から引くことにより距離(非類似度)に換算した指標)を用い k-means ベースの Self-organizing clustering(SOC)[13, 14, 15]を利用した。

次に、各クラスタに対し、そこに含まれる雑誌名とその雑誌に付される JCR カテゴリを参考に、適切と思われる分野名を決定した。

#### 3.2 分野の決定と分野間被引用/引用行列の作成

天野らの分類によれば、22 分野が得られたが、このなかには非常に小さな信頼性工学のクラスタ(4 誌)が含まれていた。筆者らは、これを(引用関係に基づいて)最も近い分野である数理工学に統合し、全分野数を 21 分野とした。すなわち、本研究における当該 5964 誌の 21 分野への割り当ては、信頼

性工学の分野を数理工学の分野に統合したことを除けば、天野らによるそれと全く同じである。

この 21 分野は次の通りである。(1) 数学、(2) 数理工学、(3) 電子情報工学、(4) 化学工学・エネルギー工学、(5) 機械・土木建築、(6) 応用化学、(7) 金属・セラミックス、(8) 物性物理、(9) 化学、(10) 素粒子・天体物理、(11) 分析化学、(12) 農学・食品学、(13) 地球科学、(14) 気象・環境、(15) 水産・海洋、(16) 動植物学、(17) 獣医学・感染症、(18) 心理・神経科学、(19) 臨床医学 1、(20) 臨床医学 2、(21) 生化学・生理学。

臨床医学は 2 つの分野に分かれているが、それぞれのクラスタに含まれる雑誌名を確認したところ、臨床医学 1 はどちらかといえば外科系、臨床医学 2 はどちらかといえば内科系である。

さらに、この 21 分野の分類にしたがい、天野らの 5964 誌の雑誌間被引用/引用行列を 21 分野間の被引用/引用行列に計算しなおした。ここで、行列の各行は被引用分野、各列は引用分野であり、第  $i$  行第  $j$  列の要素は、分野  $i$  が分野  $j$  から引用される度数である。以降、この行列を、被引用/引用行列と呼ぶこととする。この被引用/引用行列を表 1 に示す。

#### 3.3 分析の方法

前述の 21 分野間の被引用/引用行列を次の 3.3.1 から 3.3.4 の分析に用いた。

##### 3.3.1 各分野の他分野引用/被引用の程度

各分野について次の統計値を求めた: 論文数、雑誌数、自分分野引用を除く被引用数と引用数及び被引用/引用比、他分野被引用率と他分野引用率、論文あたり被引用数(自分分野引用を含める場合と含めない場合)。

##### 3.3.2 分野間の階層クラスタ関係

自分分野引用の影響を緩和した被引用ベクトルに基づき階層クラスタ関係を求めた。

表 1: 分野間被引用/引用行列

被引用分野	引用分野																				
	数学	数理工学	電子情報工学	化学工学	機械・土木建築	応用化学	金属・セラミックス	物理	化学	素粒子・天体物理学	農学・食品学	地球科学	気象・環境学	水産・海洋学	動植物学	獣医学・感染症学	心理学・神経科学	臨床医学 1	臨床医学 2	生化学・生理学	
被引用分野	数学	28,792	529	3,766	115	949	19	44	2,011	110	1,293	4	13	118	76	12	265	17	32	42	22
	数理工学	618	20,155	3,560	316	391	51	450	1,829	95	23	85	53	48	544	83	371	62	183	260	848
	電子情報工学	3,283	3,325	88,752	946	1,761	109	520	3,916	508	192	223	81	132	673	53	319	60	1,623	1,230	593
	化学工学	106	144	554	30,056	1,550	2,520	1,471	1,818	5,991	69	1,160	1,168	284	2,449	46	215	12	17	53	25
	機械・土木建築	982	192	1,951	1,878	33,686	1,033	3,095	3,243	309	307	202	42	310	1,704	257	90	16	442	487	59
	応用化学	24	23	68	2,783	1,034	77,685	3,887	4,418	20,113	55	2,720	1,982	137	598	70	308	463	361	2,092	425
	金属・セラミックス	22	287	379	1,549	2,491	3,025	67,463	18,194	11,878	44	984	59	373	309	10	13	4	10	271	11
	物理	2,394	1,824	4,727	2,374	3,350	7,039	24,046	403,763	50,665	25,433	4,829	190	4,540	3,607	129	337	269	751	1,146	99
	化学	96	53	479	7,580	507	24,422	11,849	39,515	571,546	744	15,556	4,926	1,613	3,722	268	692	858	1,068	959	422
	素粒子・天体物理学	1,002	70	366	68	291	128	129	19,773	1,435	241,828	213	7	824	3,426	2	11	3	12	225	13
	分析化学	9	60	282	1,276	189	2,261	939	3,592	14,115	124	85,268	4,323	628	2,955	219	598	464	860	418	1,369
	農学・食品学	13	42	78	1,618	23	2,337	81	146	6,428	14	6,214	83,165	621	4,885	2,329	7,157	6,449	1,018	969	4,423
	地球科学	72	0	26	195	78	170	294	1,222	1,219	605	833	254	58,837	8,998	788	1,490	38	24	99	14
	気象・環境学	210	270	555	3,465	1,514	872	383	3,100	4,643	2,460	6,240	4,795	10,549	137,796	6,357	8,191	570	425	384	2,227
	水産・海洋学	3	10	26	21	222	68	12	91	335	15	416	1,773	1,090	6,297	43,839	8,428	1,113	1,143	117	65
被引用分野	動植物学	218	115	333	163	74	394	23	562	934	17	1,011	8,916	2,100	8,033	9,838	188,946	2,839	1,756	144	293
	獣医学・感染症学	22	49	91	10	11	792	9	375	1,970	0	789	6,834	125	680	1,236	2,735	140,776	3,549	5,175	17,796
	心理学・神経科学	64	178	2,825	35	496	627	11	765	4,254	58	1,192	1,355	173	1,095	1,723	1,918	4,739	343,063	13,873	28,591
	臨床医学 1	29	204	1,800	65	616	4,630	863	1,389	2,535	199	894	1,303	110	691	167	212	5,445	13,273	239,889	37,875
	臨床医学 2	9	789	1,061	93	98	1,267	18	269	3,457	16	2,978	8,719	84	4,516	366	635	28,083	32,923	55,372	414,878
生化学・生理学	402	740	5,138	732	581	11,894	1,596	30,036	88,621	4,115	20,303	38,548	8,527	17,038	12,174	60,318	67,304	123,705	78,594	135,450	1,703,555

緩和とは，具体的には，被引用数において自分分野からの引用数が最大の場合，これを次に大きい他分野からの被引用数に置き換える処理である．

### 3.3.3 分野間の影響関係

分野間の影響関係を表す指標として，相互結合係数，および，影響係数を考案した．相互結合係数は，影響を与える/与えられるに係わりなく結び付きの強さを評価するものである．影響係数は，分野間で影響を与えているか与えられているかを評価するものである．それぞれの算出法を以下に述べる．

**相互結合係数** 分野間の影響の強さは，引用の多少によって測ることができるが，単純に引用数を用いたのでは，当然ながら規模の大きい分野が影響も強いという結果になってしまう．そこで，相互結合係数においては，分野規模の影響を取り除いたかたちでの影響の強さを考える．具体的には，それぞれの分野における全体的な影響（与えるものも受けるものも）を同じ値とすることでこの問題を解決する．

被引用/引用ともに分野規模の影響を除くため以下の条件を設ける．

1. 各分野における被引用ベクトルを比で表す．各分野の被引用の合計は 1 になる．
2. 各分野における引用ベクトルを比で表す．各分野の引用の合計は 1 になる．

この 2 つの条件をともに満たすため，被引用/引用行列に対し，以下の 1, 2 を交互に繰り返すと，やがて各行/列ともに合計が 1 に近づく [16]．

1. 行列の各要素値をその行和（各分野の総被引用数）で割る
2. 行列の各要素値をその列和（各分野の総引用数）で割る

十分に収束するまで繰り返し適用して得られた行列を，規格化被引用/引用行列と呼ぶこととする．

以上により得られた規格化被引用/引用行列より，相互結合係数を以下のように算出する．

- 相互結合係数: 規格化被引用/引用行列の各要素値と，その転置行列の対応する各要素値の和を，当該の 2 要素間の相互結合係数とする．すなわち，規格化被引用/引用行列の要素を  $a_{ij}$  とするとき，分野  $i$  と分野  $j$  の間の相互結合係数  $b_{ij}$  は， $b_{ij} = b_{ji} = (a_{ij} + a_{ji})$  で与えられる．

これは，影響を与える/与えられるにかかわらず，2 つの分野間の相互作用が強いかどうかを評価する指標（つまり，symmetric linkage）となる．

**影響係数** 2 つの分野の間で相互に及ぼし合う非対称的な影響関係 (asymmetric linkage) を評価するため，相互の引用数の比を考える．比をとることにより，分野の規模の影響は除かれる．この指標，影響係数をもとの被引用/引用行列より以下のように算出する．

- 影響係数: 被引用/引用行列の各要素値と，その転置行列の対応する各要素値の比を，当該の要素から対応する要素への影響係数とする．すなわち，被引用/引用行列の要素を  $h_{ij}$  とするとき， $c_{ij} = h_{ij}/h_{ji}$  を分野  $i$  から分野  $j$  への影響係数と呼ぶことにする． $c_{ij}$  と  $c_{ji}$  は互いに逆数関係にあり ( $c_{ij} = 1/c_{ji}$ )， $c_{ij} > 1$  なら，分野  $i$  から分野  $j$  への影響が分野  $j$  から分野  $i$  への影響よりも大きく， $c_{ij} < 1$  ならその逆である．

### 3.3.4 各分野が他の分野に及ぼす影響強度

各分野がそれ以外の全分野に及ぼす影響強度を，雑誌の影響強度指標である Eigenfactor を求める手法 [17, 18, 19, 20] を応用して求めた．具体的な算出法を以下に述べる．

まず， $n$  個の分野間の被引用/引用行列  $H$  が与えられているとする．また，各要素値が各分野の論文

数に比例したベクトル  $p; \sum_i p_i = 1$  があるとする．この行列  $H$  を以下のように調整し，行列  $G$  を算出する．

1. 各対角項を 0 にする．すなわち自分分野引用を無視する (これを  $H^*$  とする)
2.  $H^*$  の各要素  $h_{ij}$  を，分野  $j$  の ( $H^*$  における) 全引用数  $\sum_i h_{ij}$  で規格化する (各列の和が 1 になる)
3. 2 で得られた行列全体に係数  $\alpha$  ( $0 < \alpha < 1$ ) を掛ける (各列の和が  $\alpha$  になる)
4. 各引用ベクトル (列ベクトル) の各要素に  $p_i(1 - \alpha)$  を足す．これを行列  $G$  とする

つまり，行列  $G$  の各要素  $g_{ij}$  は次式で表される．

$$g_{ij} = p_i(1 - \alpha) + \alpha h_{ij} / \sum_i h_{ij} \quad (1)$$

行列  $G$  に対する (第一) 固有ベクトルを  $\pi$  とすると (式 (2))<sup>1</sup>，その各要素  $\pi_i$  が分野  $i$  の影響を表すスコアとなる (以降，これを分野 Eigenfactor と呼ぶこととする)．

$$\begin{aligned} G\pi &= \pi \\ e^T \pi &= 1 \end{aligned} \quad (2)$$

分野 Eigenfactor の高い順番が影響度の高い順番となる．式 (1) の要素  $g_{ij}$  は，分野  $j$  の論文を見ている読者が，次に別の分野  $i$  の論文に移る確率を表しており， $\alpha$  の項は引用を辿って移る確率， $(1 - \alpha)$  の項は引用に関係なく  $i$  の論文数に比例して移る確率である． $\alpha$  を大きくするほど，引用による影響の比率が高まる．詳細は，文献 [17] を参照されたい．

<sup>1</sup>行列  $G$  は最大固有値 1 を持つ．式 (2) の最初の式は，その固有値に対する固有ベクトルが  $\pi$  であることを意味する．また，式 (2) の 2 番目の式により， $\pi$  の要素値の和が 1 になるように規格化している ( $e^T$  はベクトル  $(1, 1, 1, \dots)$  を表す)．詳細は文献 [19] を参照されたい．

## 4 結果

### 4.1 各分野の他分野被引用/引用の程度

各分野の，他分野被引用率/引用率およびその他の基本的な統計値を表 2 に示す．

論文数における各分野の規模を見ると，生化学・生理学が最大で，次いで物性物理，化学，臨床医学 2 (内科系と思われる)，臨床医学 1 (外科系と思われる)，心理・神経科学となる．全体として医学生物学関連の分野が上位を占める．

他分野引用率，他分野被引用率は，概ね 0.2 から 0.5 の範囲内にある．素粒子・天体物理分野はどちらも 0.1 程度で極めて低く，この分野が孤立していることを窺わせる．一方，農学・食品学，化学工学・エネルギー工学，獣医学・感染症等の分野は，他分野との結びつきが強い．全体的には，学際的な引用ネットワークの形成にかかわる他分野間の引用は全引用の 30% 程度である．

各分野の自分分野引用を除いた被引用/引用比を見ると，その値が 0.5 から 1.6 程度で，データの規模が大きいことを考慮するとばらつきは大きく，いくつかの分野間においては，その差は明確である．被引用が，引用に比べてくに多い分野は，生化学・生理学分野 (比: 1.59)，臨床医学 2 (1.23)，数理工学 (1.17) である．臨床医学 1 と臨床医学 2 では，同じ臨床医学分野であっても，被引用/引用比に大きな差が見られる．

自分分野引用を除いた 1 論文あたりの被引用数を見ると，唯一，生化学・生理学のみが 1 を超える値 (1.19) であり，他の全ての分野は 1 を下回る結果となった．次に数値の大きい分野は臨床医学 2 (0.95) である．分野の論文数と論文あたり被引用数の間の Spearman 順位相関係数は，自分分野引用を含める場合 0.67，含めない場合 0.50 である．つまり，全体的に，規模の大きな分野が論文あたり被引用数も多い傾向があり，規模と影響度の相乗効果が認められる．一方，他分野被引用率および他分野引用率と論文数との間の Spearman 順位相関係数は，それぞれ，-0.17，-0.30 である．つまり，規模の大きな分

表 2: 各分野における引用数等

分野	論文数	雑誌数	被引用数 (*)	引用数 (*)	被引用/ 引用 (*)	他 分 野 引用率	他 分 野 被 引 用 率	1 論文あ たり被引 用数	1 論文あ たり被引 用数 (*)
数学	77,345	216	9,594	9,578	1.00	0.25	0.25	0.5	0.12
数理工学	57,812	182	10,398	8,904	1.17	0.31	0.34	0.53	0.18
電子情報工学	145,987	435	21,060	28,065	0.75	0.24	0.19	0.75	0.14
化学工学・エネ ルギー工学	75,438	173	19,756	25,282	0.78	0.46	0.40	0.66	0.26
機械・土木建築	69,050	187	16,809	16,226	1.04	0.33	0.33	0.73	0.24
応用化学	93,381	163	46,395	63,658	0.73	0.45	0.37	1.33	0.50
金属・セラミッ クス	131,554	209	40,113	49,720	0.81	0.42	0.37	0.82	0.30
物性物理	384,818	248	142,806	136,264	1.05	0.25	0.26	1.42	0.37
化学	345,502	315	147,894	219,615	0.67	0.28	0.21	2.08	0.43
素粒子・天体物 理	110,448	78	29,097	35,783	0.81	0.13	0.11	2.45	0.26
分析化学	81,244	130	42,969	66,846	0.64	0.44	0.34	1.58	0.53
農学・食品学	122,403	243	64,483	85,341	0.76	0.51	0.44	1.21	0.53
地球科学	55,707	165	17,925	32,406	0.55	0.36	0.23	1.38	0.32
気象・環境	144,004	308	64,800	72,296	0.90	0.34	0.32	1.41	0.45
水産・海洋	57,635	148	25,433	36,127	0.70	0.45	0.37	1.20	0.44
動植物学	152,543	427	62,616	94,303	0.66	0.33	0.25	1.65	0.41
獣医学・感染症	144,683	277	83,704	118,808	0.70	0.46	0.37	1.55	0.58
心理・神経科学	201,248	374	158,080	183,175	0.86	0.35	0.32	2.49	0.79
臨床医学 1	243,966	416	125,362	161,910	0.77	0.40	0.34	1.50	0.51
臨床医学 2	296,555	519	283,151	230,620	1.23	0.36	0.41	2.35	0.95
生化学・生理学	591,277	751	705,816	443,334	1.59	0.21	0.29	4.08	1.19

\*自分分野引用を除いた値で計算

野は相対的には自分分野内での引用/被引用の比重が高くなる。しかし、自分分野引用を除いた被引用/引用比と論文数との間の Spearman 順位相関係数は 0.26 であり、規模が大きな分野の方が他分野から影響を受けるよりも与える方向があることが判る。

#### 4.2 分野間の階層クラスタ関係

図 1 に全 21 分野のデンドログラムプロットを示す。

同図より、全ての分野は、以下の 6 つの上位の分野 (系) に統合することができる。

- 情報・数学系:  
数学, 数理工学, 電子情報工学
- 物質・工学系:  
化学工学・エネルギー工学, 機械・土木建築, 応用化学, 金属セラミックス, 物性物理, 化学
- 素粒子・天体物理系:  
素粒子・天体物理
- 生医学系:  
獣医学・感染症, 心理・神経科学, 臨床医学 1, 臨床医学 2, 生化学・生理学
- 農学系:  
分析化学, 農学・食品学



## 引用データに基づく自然科学領域における学術研究分野間の関係



図 1: 全分野のデンドログラムプロット

データには自己引用の影響を緩和した被引用ベクトル，距離にはコサイン距離，クラスタリング法には群平均法を用いた．

表 3: 各分野間の相互結合係数

被引用分野	引用分野	分野																				
		数学	数理工学	電子情報工学	化学工学・エネルギー工学	機械・土木建築	応用化学	金属・セラミックス	物性物理	化学	素粒子・天体物理	分析化学	農学・食品学	地球科学	気象・環境	水産・海洋	動植物学	獣医学・感染症	心理・神経科学	臨床医学 1	臨床医学 2	生化学・生理学
数学		1.714	0.04	0.114	0.006	0.051	0.001	0.001	0.03	0.001	0.024	0.	0.	0.004	0.003	0.	0.005	0.	0.001	0.001	0.	0.001
数理工学		0.04	1.694	0.133	0.014	0.017	0.001	0.015	0.03	0.001	0.001	0.003	0.002	0.001	0.011	0.002	0.006	0.002	0.004	0.005	0.014	0.004
電子情報工学		0.114	0.133	1.56	0.021	0.053	0.002	0.009	0.032	0.003	0.003	0.005	0.001	0.002	0.008	0.001	0.004	0.001	0.019	0.015	0.006	0.009
化学工学・エネルギー工学		0.006	0.014	0.021	1.491	0.082	0.081	0.05	0.026	0.072	0.001	0.036	0.039	0.009	0.063	0.001	0.004	0.	0.	0.001	0.001	0.002
機械・土木建築		0.051	0.017	0.053	0.082	1.553	0.03	0.088	0.04	0.004	0.006	0.006	0.001	0.006	0.033	0.01	0.002	0.	0.007	0.009	0.001	0.002
応用化学		0.001	0.001	0.002	0.081	0.03	1.448	0.069	0.043	0.145	0.001	0.045	0.037	0.004	0.01	0.002	0.004	0.009	0.004	0.033	0.006	0.025
金属・セラミックス		0.001	0.015	0.009	0.05	0.088	0.069	1.467	<u>0.174</u>	0.084	0.001	0.019	0.001	0.008	0.005	0.	0.	0.	0.006	0.	0.002	
物性物理		0.03	0.03	0.032	0.026	0.04	0.043	<u>0.174</u>	1.285	0.121	0.109	0.031	0.001	0.024	0.018	0.001	0.002	0.002	0.003	0.006	0.001	0.019
化学		0.001	0.001	0.003	0.072	0.004	0.145	0.084	0.121	1.313	0.004	0.094	0.034	0.011	0.019	0.003	0.003	0.007	0.007	0.006	0.004	0.062
素粒子・天体物理		0.024	0.001	0.003	0.001	0.006	0.001	0.001	0.109	0.004	1.805	0.002	0.	0.01	0.025	0.	0.	0.	0.	0.001	0.	0.005
分析化学		0.	0.003	0.005	0.036	0.006	0.045	0.019	0.031	0.094	0.002	1.502	0.088	0.016	0.057	0.008	0.009	0.008	0.009	0.006	0.015	0.042
農学・食品学		0.	0.002	0.001	0.039	0.001	0.037	0.001	0.001	0.034	0.	0.088	1.35	0.009	0.06	0.049	0.091	0.087	0.01	0.011	0.044	0.084
地球科学		0.004	0.001	0.002	0.009	0.006	0.004	0.008	0.024	0.011	0.01	0.016	0.009	1.662	<u>0.159</u>	0.029	0.027	0.001	0.001	0.001	0.	0.015
気象・環境		0.003	0.011	0.008	0.063	0.033	0.01	0.005	0.018	0.019	0.025	0.057	0.06	<u>0.159</u>	1.287	0.115	0.07	0.006	0.004	0.004	0.017	0.026
水産・海洋		0.	0.002	0.001	0.001	0.01	0.002	0.	0.001	0.003	0.	0.008	0.049	0.029	0.115	1.55	<u>0.152</u>	0.023	0.018	0.002	0.002	0.032
動植物学		0.005	0.006	0.004	0.004	0.002	0.004	0.	0.002	0.003	0.	0.009	0.091	0.027	0.07	<u>0.152</u>	1.497	0.026	0.011	0.001	0.002	0.083
獣医学・感染症		0.	0.002	0.001	0.	0.	0.009	0.	0.002	0.007	0.	0.008	0.087	0.001	0.006	0.023	0.026	1.498	0.028	0.043	0.127	0.131
心理・神経科学		0.001	0.004	0.019	0.	0.007	0.004	0.	0.003	0.007	0.	0.009	0.01	0.001	0.004	0.018	0.011	0.028	1.518	0.07	0.113	<u>0.172</u>
臨床医学 1		0.001	0.005	0.015	0.001	0.009	0.033	0.006	0.006	0.006	0.001	0.006	0.011	0.001	0.004	0.002	0.001	0.043	0.07	1.459	<u>0.197</u>	0.121
臨床医学 2		0.	0.014	0.006	0.001	0.001	0.006	0.	0.001	0.004	0.	0.015	0.044	0.	0.017	0.002	0.002	0.127	0.113	<u>0.197</u>	1.267	<u>0.184</u>
生化学・生理学		0.001	0.004	0.009	0.002	0.002	0.025	0.002	0.019	0.062	0.005	0.042	0.084	0.015	0.026	0.032	0.083	0.131	0.172	0.121	0.184	0.979

表 4: 各分野間の影響係数

引用分野 被引用分野	数学	数理工 学	電子情 報工学	化学工 学・エネ ルギー 工学	機械・土 木建築 学	応用化 学	金 属 セ ラ ミッ クス	物性物 理	化学	素粒子・天 体物理	分析化 学	農学・食 品学	地球科 学	気象・環 境	水産・海 洋	動植物 学	獣医学・感 染症	心理・神 経科学	臨床医 学 1	臨床医 学 2	生化学・ 生理学
数学	1.	-	1.15	-	0.97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
数理工学	-	1.	1.07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
電子情報工学	0.87	0.93	1.	-	0.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
化学工学・エネルギー工学	-	-	-	1.	0.83	0.91	0.95	-	0.79	-	-	-	-	0.71	-	-	-	-	-	-	-
機械・土木建築	1.03	-	1.11	1.21	1.	-	1.24	-	-	-	-	-	-	-	-	-	-	-	-	-	-
応用化学	-	-	-	1.1	-	1.	1.28	-	0.82	-	-	-	-	-	-	-	-	-	-	-	-
金属・セラミックス	-	-	-	1.05	0.8	0.78	1.	0.76	1.	-	-	-	-	-	-	-	-	-	-	-	-
物性物理	-	-	-	-	-	-	<u>1.32</u>	1.	1.28	1.29	-	-	-	-	-	-	-	-	-	-	-
化学	-	-	-	1.27	-	1.21	1.	0.78	1.	-	1.1	-	-	-	-	-	-	-	-	-	0.37
素粒子・天体物理	-	-	-	-	-	-	-	0.78	-	1.	-	-	-	-	-	-	-	-	-	-	-
分析化学	-	-	-	-	-	-	-	-	0.91	-	1.	0.7	-	0.47	-	-	-	-	-	-	-
農学・食品学	-	-	-	-	-	-	-	-	-	-	1.44	1.	-	1.02	-	0.8	0.94	-	-	-	0.51
地球科学	-	-	-	-	-	-	-	-	-	-	-	-	1.	0.85	-	-	-	-	-	-	-
気象・環境	-	-	-	1.41	-	-	-	-	-	-	2.11	0.98	1.17	1.	1.01	1.02	-	-	-	-	-
水産・海洋	-	-	-	-	-	-	-	-	-	-	-	-	-	0.99	1.	0.86	-	-	-	-	-
動植物学	-	-	-	-	-	-	-	-	-	-	-	1.25	-	0.98	1.17	1.	-	-	-	-	0.41
獣医学・感染症	-	-	-	-	-	-	-	-	-	-	-	1.06	-	-	-	-	1.	-	-	0.63	0.62
心理・神経科学	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.	1.05	0.87	0.76
臨床医学 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.96	1.	0.68	0.68
臨床医学 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.58	1.15	<u>1.46</u>	1.	1.05
生化学・生理学	-	-	-	-	-	-	-	-	2.72	-	-	1.96	-	-	-	2.43	1.62	<u>1.31</u>	1.48	0.95	1.

表 3 において値が 0.05 以上であるものに対応する部分のみ表示

#### ● 環境・地球系:

地球科学, 気象・環境, 水産・海洋, 動植物学

#### 4.3 分野間の影響関係

相互結合係数 分野間の相互結合係数を算出したものが表 3 である。

以下の 6 分野間で相互結合係数の値が 0.15 を超え (表 3 の下線を付した値), とくに結び付きが強いことが示された。

- 臨床医学 1-臨床医学 2(0.197)
- 臨床医学 2-生理学・生化学 (0.184)
- 物性物理-金属・セラミックス (0.174)
- 心理・神経科学-生理学・生化学 (0.172)
- 気象・環境-地球科学 (0.159)
- 動植物学-水産・海洋 (0.152)

影響係数 分野間の影響係数を算出したものが表 4 である。この表では, 相互結合係数 (表 3) の値が 0.05 以上である分野間についてのみ値を示している。同表からは, 多くの分野間で影響の対称性が壊

れていることが解かる。前述の相互結合関係の強い分野間のうちでとくに影響関係に非対称性が見られるものは, 以下の 3 分野間であった (表 4 の下線を付した値)。

- 臨床医学 2 → 臨床医学 1 (1.46)
- 物性物理 → 金属・セラミックス (1.32)
- 生理学・生化学 → 心理・神経科学 (1.31)

#### 4.4 各分野が他の分野に及ぼす影響強度

分野 Eigenfactor を求める手法による各分野の影響強度を,  $\alpha$  を 0.5 から 0.9 の範囲で変動させながら繰り返し法により求めた。これによる影響度順位を表 5 に示す。 $\alpha$  による影響は,  $\alpha$  が小さいほど, すなわち  $(1 - \alpha)$  が大きいほど, 論文数の寄与が大きくなる。 $\alpha$  を積極的に変動させて用いれば, 論文寄与を調整するためのパラメーターとして利用できる。表 5 では, 実際に,  $\alpha$  の変動, すなわち論文寄与の変動による順位の変化が見られる。

#### 5 考察・まとめ

表 5: 分野 Eigenfactor による影響度順位

	0.5	0.6	0.7	0.8	0.9	順位変動	被引用数	論文数	被引用数順位	論文数順位
数学	18	18	19	20	20		9,594	77,345	21	16
数理工学	20	21	21	21	21		10,398	57,812	20	19
電子情報工学	12	12	12	14	15		21,060	145,987	16	8
化学工学・エネルギー工学	16	16	16	17	17		19,756	75,438	17	17
機械・土木建築	17	17	17	18	18		16,809	69,050	19	18
応用化学	14	14	13	12	11		46,395	93,381	11	14
金属・セラミックス	11	11	11	11	13		40,113	131,554	13	11
物性物理	3	3	3	5	6		142,806	384,818	5	2
化学	4	4	4	4	5		147,894	345,502	4	3
素粒子・天体物理	13	13	14	15	14		29,097	110,448	14	13
分析化学	15	15	15	13	12		42,969	81,244	12	15
農学・食品学	10	10	10	9	9		64,483	122,403	9	12
地球科学	21	20	20	19	19		17,925	55,707	18	21
気象・環境	9	9	9	10	10		64,800	144,004	8	10
水産・海洋	19	19	18	16	16		25,433	57,635	15	20
動植物学	8	8	8	8	8		62,616	152,543	10	7
獣医学・感染症	7	7	7	7	7		83,704	144,683	7	9
心理・神経科学	5	5	5	3	3		158,080	201,248	3	6
臨床医学 1	6	6	6	6	4		125,362	243,966	6	5
臨床医学 2	2	2	2	2	2		283,151	296,555	2	4
生化学・生理学	1	1	1	1	1		705,816	591,277	1	1

分野の規模・全体における影響の強さ 表 2 より、概しては論文数の大きな分野が論文あたりの被引用数も大きくなる傾向にある。また、このような分野では被引用数/引用数比（自分分野引用を除く）も大きい。すなわち、規模の大きい分野は分野全体としてのみならず、個々の論文単位でも、他の分野に対し大きい影響を与えると推察される。

分野 Eigenfactor による影響度順位 (表 5) を見ると、やはり、分野規模の大きい前述の 5 分野が順位も上であるが、 $\alpha$  により 5 分野の中でも変動が見られる。臨床医学 1 は、 $\alpha$  が 0.9 のときの順位が、単純な被引用数順位、論文数順位よりも高くなっている。逆に、この作用により、化学、物性物

理は、順位を落としている。 $\alpha$  が高い（つまり論文の寄与よりも引用の寄与が大きい）とき、分野 Eigenfactor は、他分野からの被引用数が多い分野で高くなると同時に、被引用数の多い分野からの引用の割合が高い分野ほど高くなるという性質を持つ。表 1 から、臨床医学 1 分野は、生化学・生理学分野や臨床医学 2 分野のような被引用数の多い分野からの引用割合が多いため、その順位が押し上げられたと考えられる。

Eigenfactor や類似の手法を利用する際、 $\alpha$  をどう設定するかについては、本論文では議論しないが、分野という大きな粒度で解析を行った際にも、これが重要な要素となることが確認された。

分野間の引用関係 被引用傾向の近さによる階層クラスタ関係 (図 1) を見ると、素粒子・天体物理は単独でクラスタを作り、その他の物理系分野が含まれないことや、分析化学が他の化学関連の分野より農学・食品学に近いことは興味深い。また、生化学・生理学は医学関連の分野と近く、他の生物・農学関連の分野はむしろ地球科学等に近いことも興味深い。これらは環境にかかわるテーマにより結び付きが強くなっていると予想され、現代の自然科学の様相をよく表していると言える。

相互結合係数 (表 3) からは、概ね階層クラスタ関係に見られる各分野間の近さが支持されるものの、やや異なる傾向も見られた。階層クラスタ関係に見られる物性物理-化学の組は、相互結合係数では、物性物理に注目すれば金属・セラミックスとの結びつきが強く、化学に注目すれば応用化学との結びつきが強い。また、数理工学-電子情報工学、物性物理-化学、臨床医学 2-生化学・生理学、地球科学-気象・環境の各組は、階層クラスタ関係では同程度の近さを示しているが、相互結合係数では数値に開きがある。階層クラスタ関係は、被引用ベクトル間の類似度 (すなわち当該の 2 分野以外の各分野との関係の類似度) に基づいているのに対し、相互結合係数は当該 2 分野間の直接の関係の強さを表している。前述の差異はこの違いが現れたものである。

影響係数 (表 4) からは、物性物理 → 金属・セラミックス、生理学・生理学 → 心理・神経科学のように基礎分野から応用分野への影響と考えられる関係もあるが、臨床医学 2 → 生理学・生理学 (1.05) のようなやや意外な関係も見られた。

まとめ 本研究で明らかになったことを以下にまとめる。

- 学際的と言える引用関係、つまり分野間引用は、引用全体の 3 割程度である
- 論文数規模と影響度には相乗効果が認められ、医学生物学関連の分野の規模と影響度が大きい
- 分野 Eigenfactor による各分野の影響強度も、

医学生物学関連の分野が高い

- 各分野を被引用傾向に基づきさらに上位の分野に分類した結果、物理系、化学系というような従来型の関係とはやや異なる分野間の結びつきが確認できた
- 被引用傾向に基づく階層クラスタ関係と相互結合係数とを比較した結果、最下層のクラスタにおいては、相互結合係数から見える関係とは、やや異なる傾向が見られた

謝辞

本研究は、Thomson Scientific 社 (現 Thomson Reuters Scientific) が提供する Journal Citation Reports (JCR) 2004 Science Edition CD-ROM 版に収録されているデータを利用して行われました。

本研究の一部は、農林水産省 農林水産技術会議 科学技術計算システムを利用して行われました。

本研究の一部は、RIKEN Integrated Cluster of Clusters (RICC) を利用して行われました。

尚絅大学 角田裕之教授 (現鶴見大学) には研究の全般において貴重なアドバイスを頂きました。感謝いたします。

注・文献

- [1] Zhang, L.; Janssens, F.; Liang, L.; Glanzel, W. Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics*. 2010, Vol. 82, p. 687–706.
- [2] Leydesdorf, L. Various methods for the mapping of science. *Scientometrics*. 1987, Vol. 11, Nos 5–6, p. 295–324.
- [3] Leydesdorf, L. Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports. *Journal of Documentation*. 2004, Vol. 60, No. 4, p. 371–427.
- [4] Leydesdorf, L. Indicators of structural change in the dynamics of science: Entropy statistics of the SCI Journal Citation Reports. *Scientometrics*. 2002, Vol. 53, No. 1, p. 131–159.

- [5] Leydesdorf, L. Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*. 2006, Vol. 57, No. 5, p. 601–613.
- [6] Shildt, H. A.; Mattsson, J. T. A dense network sub-grouping algorithm for co-citation analysis and its implementation in the software tool Sitkis. *Scientometrics*. 2006, Vol. 67, No. 1, p. 143–163.
- [7] Bassecoulard, E.; Zitt, M. Indicators in A Research Institute: A Multi-level Classification of Scientific Journals. *Scientometrics*. 1999, Vol. 44, No. 3, p. 323–345.
- [8] Glanzel, W.; Schubert, A. A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*. 2003, Vol. 56, No. 3, p. 357–367.
- [9] Chen, C.-M. Classification of Scientific Networks Using Aggregated Journal-Journal Citation Relation in the journal Citation Reports. *Journal of the American Society for Information and Technology* 2008, Vol. 59, No. 14, p. 2296–2304.
- [10] Rosvall, M.; Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2008, Vol. 105, No. 4, p. 1118–1123.
- [11] 天野晃, 児玉関. 引用にもとづく雑誌クラスタリング法の開発. *情報メディア研究*. 2009, Vol. 7, No. 1, p. 63–73.
- [12] Leydesdorf, L.; Hammarfelt, B.; Salah, A. The Structure of the Arts & Humanities Citation Index: A Mapping on the Basis of Aggregated Citations Among 1,157 Journals. *Journal of the American Society for Information Science and Technology*. 2010, Vol. 62, No. 12, p. 2414–2426.
- [13] 天野晃. 自己組織化による DNA シークエンスの分類. *情報メディア学会第 5 回研究会 発表資料*. 2003, p. 5–8.
- [14] Amano, K.; Nakamura, H.; Ichikawa, H. Self-organizing clustering: A novel non-hierarchical method for clustering large amount of DNA sequences. *Genome Informatics*. 2003, Vol. 14, p. 575–576.
- [15] Amano, K.; Ichikawa, H.; Nakamura, H.; Numa, H.; Fukami-Kobayashi, K.; Nagamura, Y.; Onodera, A. Self-organizing Clustering: Non-hierarchical Clustering for Large Scale DNA Sequence Data. *IPSIJ Digital Courier*. 2007, Vol. 3, p. 193–197.
- [16] Darroch, J. N.; Ratcriff, D. Generalized Iterative Scaling for Log-Linear Models. *The Annals of Mathematical Statistics*. 1972, Vol. 43, No. 5, p. 1470–1480.
- [17] Eigenfactor<sup>TM</sup> Score and AI<sup>TM</sup> Score: Detailed methods. <<http://www.eigenfactor.org/methods.pdf>> (accessed 2012-06-06).
- [18] 角田裕之, 小野寺夏生. 論文と研究者のインパクトに対する新しい計量書誌学的指標 -論文引用グラフの固有ベクトル解析-. *情報メディア研究*. 2006, Vol. 5, p. 1–20.
- [19] Google Page Rank の数理, Langville, Amy N.; Meyer, Carl D. 岩野和生, 黒川利明, 黒川洋 訳. 共立出版, 296 ページ.
- [20] 小野寺夏生. 新しい雑誌評価指標アイゲンファクターとは. *現代化学*. 2009, Vol. 465, p. 38–42.

(yyyy 年 mm 月 dd 日 受付)

(yyyy 年 mm 月 dd 日 採録)

(yyyy 年 mm 月 dd 日 出版)