Scientometrics (2023) 128:3507–3534
https://doi.org/10.1007/s11192-023-04711-8

New concept of the affinity between research fields using academic journal data in Scopus

Ryo Takahashi[1] · **Kenji Kaibe**[2] · **Kazuyuki Suzuki**[3] · **Sayaka Takahashi**[4] · **Kotaro Takeda**[5] · **Marc Hansen**[6] · **Michiaki Yumoto**[7]

Received: 24 August 2022 / Accepted: 16 April 2023 / Published online: 8 May 2023

Abstract

We propose a new concept for measuring the affinity between fields of academic research. The importance of interdisciplinary research has been increasingly emphasized in recent years. The degree of interdisciplinarity of a research article can be determined using bibliographic information from the cited literature. However, the properties of the affinity of each field to other fields have not yet been discussed. Therefore, we employ our method to quantify the affinity between 27 research fields using academic journal data from the citation and abstract database Scopus. We show that the affinity between fields should be viewed from two perspectives: the affinity of other fields to the field of interest, and the affinity of the field of interest to other fields. We identify the fields of “Arts and Humanities” and “Social Sciences”, and “Earth and Planetary Sciences” and “Environmental Sciences”, as those with the highest bidirectional affinity. We also demonstrate that affinity to “Medicine” is particularly high, with seven fields of interest having the highest affinity to this field: “Biochemistry, Genetics and Molecular Biology”, “Immunology and Microbiology”, “Neuroscience”, “Pharmacology, Toxicology and Pharmaceutics”, “Nursing”, “Dentistry”, and “Health Professions”.

Keyword Interdisciplinary research · Affinity between fields · ASJC · Co-assigned field

Ryo Takahashi
r.takahashi@g.tohoku.ac.jp

[1] Graduate School of Science, Tohoku University, Sendai city, Japan

[2] Advanced Institute for Materials Research, Tohoku University, Sendai city, Japan

[3] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai city, Japan

[4] Graduate School of Life Sciences, Tohoku University, Sendai city, Japan

[5] Graduate School of Engineering, Tohoku University, Sendai city, Japan

[6] University Research Administration Center, Tohoku University, Sendai city, Japan

[7] Institute for Materials Research, Tohoku University, Sendai city, Japan

Introduction

When setting policies for science and technology innovation and planning research strategies, it is important to analyze the current research capabilities of researchers and research institutes and to predict future trends in related research fields. In discussions on science and technology innovation in Japan, interdisciplinarity is considered to be particularly important for research conducted at universities. To investigate interdisciplinary research for their policy and strategy planning, Japanese government agencies use analyses based on indicators of interdisciplinary research, such as cited references, from the National Institute of Science and Technology Policy (NISTEP) [Cabinet Office of Japan, 2016, 2019, 2020; Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), 2018].

In a study measuring interdisciplinarity, Porter et al. (2007) proposed two indices: integration and specialization. Integration describes the degree to which a research article cites articles from other subject categories in Web of Knowledge (the precursor to Web of Science), whereas specialization measures the spread of subject categories in which a body of research (such as the work of a given author in a set time period) is published. Porter & Rafols (2009) then investigated changes in the degree of interdisciplinarity in six research areas from 1975 to 2005 and found that although scientific research is becoming more interdisciplinary, progress is small and citations are largely from neighboring fields. Leydesdorff & Rafols (2011) studied three indicators of interdisciplinarity: the Shannon entropy (Shannon, 1948), the betweenness centrality (Freeman, 1977, 1978/1979), and the Rao-Stirling measures (Rao, 1982; Stirling, 2007). Although no single indicator clearly measured the interdisciplinarity of research, they each captured a different aspect of interdisciplinary field integration. Other studies used Hill-type true diversity to measure the diversity of cited references (Jost, 2006, 2007, 2009; Leinster & Cobbold, 2012), and it has been suggested that this indicator may be a useful measure of the degree of interdisciplinarity (Zhang et al., 2016). A more detailed discussion of these indices can be found in Wagner et al. (2011), and a review of the broader context of scientometrics is available in Mingers & Leydesdorff (2015). Kim et al. (2022) have presented a methodological framework for analyzing topic-based interdisciplinarity. The framework is not discipline-specific, but serves as a guide for identifying the characteristics of and relationships among topics that are actively discussed in highly interdisciplinary research areas.

To determine the impact of interdisciplinary research, Lariviere & Gingras (2010) conducted a citation rate study, defining the degree of field integration for individual academic papers as the percentage of citations to journals in other disciplines. From that study, the authors noted: (i) across all disciplines, there is no clear correlation between the degree of interdisciplinarity and citation rates; (ii) the more specialized a discipline is, the higher the citation rate of specific fields; and (iii) papers with higher degrees of specialization and interdisciplinarity have smaller scientific (citation) impact. From an impact perspective, there is an optimal degree of field integration. In a study investigating citations at the journal level, Silva et al. (2013) tried to quantify field integration by conducting entropy measurements of the diversity of the fields of journals that cite a particular journal. The result indicated that scientific fields are becoming increasingly interdisciplinary, and the degree of interdisciplinarity (entropy) is strongly correlated with the impact factor (IF) of journals (high entropies were obtained for journals with very wide readership). Kong & Wand compared the differences in citation counts and Altmetric scores between covered and non-covered articles published in Nature (Kong & Wang, 2020). They showed in their empirical study that the number of citations is significantly higher than that of noncited papers, but the difference has become smaller in recent years. They also noted that in the biological sciences, physical sciences, and other interdisciplinary areas, papers with high citation counts and high Altmetric scores were relatively more interdisciplinary. In a recent study, Chen et al. (2021) have indicated that variety rather than balance and disparity is likely the most important interdisciplinary factor for citation impact.

Other studies limit their assessment of interdisciplinarity to specific disciplines. One notable example is a study of the NASA Astrobiology Institute at the University of Hawaii that was based on bibliometric methodology and machine learning algorithms (Gowanlock & Gazan, 2013). Another notable example is a study predicting the degree of interdisciplinarity in nanotechnology (Jang et al., 2018). The authors proposed a framework using the Glänzel-Schubert-Schoepflin model to estimate the citation rate using stochastic processes. As a study of approaches to increasing interdisciplinarity, Levy et al. (2005) have noted elements that they believe will help to further enhance interdisciplinarity in life course research. In the field of bio-nanoscience, a case study assessing the Leinster-Cobbold diversity indices noted that the various measurements of field integration are special cases of these indices (Mugabushaka et al., 2016). In the field of major medical subspecialties, one interdisciplinarity score was positively correlated with the journal's impact factor, indicating that interdisciplinary research has greater impact (Petterson et al., 2021). In terms of social science research, Brink et al. (2020) focused on interdisciplinary measures of sustainability, reviewing them and investigating whether measurement issues differ among environmental, economic, and social. In interdisciplinary research across the human and natural sciences, Pittman et al. (2016) have presented the experience of 20 years of motivating scientists to engage in interdisciplinary research, providing an environment for learning across disciplines, and structuring research programs to advance knowledge for decision making on global change.

Other studies on the degree of field integration in academic research have discussed assessment indices, the impact of field integration, and case studies on specific research institutions/disciplines. However, there have been no specific and quantitative discussions on the state of realization of field integration in a particular field, including a discussion of the other fields with which it is, or is not, integrated. In the above-mentioned "adjacent" areas by Porter et al. (2007), the degree of integration between a particular field and other fields has yet to be shown quantitatively. The disparity, which is one of the Rao-Stirling measures, is defined as the average distance between fields. This index is calculated for a set of publications using cited references.

In this study, we propose a concept of the "affinity" between fields in academic research. We define affinity as the number of journals assigned to two fields divided by the number of journals assigned to one of those two fields. To calculate the affinity between fields, we use information on the research fields of all academic journals registered in Scopus. A difference between affinity and disparity is the analysis target. The analysis target of affinity is a set of journals, while disparity utilizes individual publications. Thus, citation relations affect the size of disparity, but not affinity. We also show that there is two-way information on affinity between fields, that is, percentages of field of interest assigned in other fields and percentages of other fields assigned in the field of interest. This kind of directedness is an aspect that is not included in the concept of disparity but identified as a critical factor in the context of interdisciplinarity.

As a University Research Administrator (URA), we are often exposed to top-down research evaluation indicators in the form of policies and university research strategies, etc. We often feel that the quantitative aspects of research interdisciplinarity are unclear in discussions. One of the motivations for this research is to try to improve this point from the URA's point of view.

This paper is organized as follows: In Section "Affinity between fields", we define the affinity between fields. In Section "Data and the method of analysis", we explain the data and the method of analysis and then visualize the affinity. In Section "Discussion", we present detailed results and a discussion. In Section "An example of applications", we give an example of application. In Section "Limitations", we outline the limitations of our analyses and results. Section "Summary", we summarize our results.

Affinity between fields

In measuring the interdisciplinarity of research, many studies based on citation analyses of articles show that some fields of research are more likely to be interdisciplinary than others, without taking into account the academic affinity between fields (Jost, 2006, 2007, 2009; Leinster & Cobbold, 2012; Leydesdorff & Rafols, 2011; Porter & Rafols, 2009; Porter et al., 2007; Zhang et al., 2016). The question arises: Can we treat all intrinsic affinities between fields equally? For example, we may infer a strong relationship between the fields of chemistry and chemical engineering from the keyword 'chemistry', but it is more difficult to find such relationships for the field of arts and humanities. This means that before discussing the degree of interdisciplinarity, it is necessary to measure the affinity between fields. Therefore, in this study, we develop a new approach for quantifying the affinity between fields.

Methods of network analysis are often used in analyses of the interdisciplinarity of research (*e.g.* for determining relationships between keywords in academic research). To construct bibliometric networks, four main bibliometric techniques are used: co-citation, bibliographic coupling, co-author, and co-word (Cobo et al., 2011). By contrast, our approach for quantifying affinity between fields introduces a new technique: co-assigned fields. The unit of analysis is journals, and relationships are characterized by co-assigned field. These points are different from disparity of the Rao-Stirling measures. Disparity looks at individual papers, and relationships are characterized by citations (referenced literature). This choice of technique, unit of analysis, and relationship characterization underpin our method to quantify the affinity between fields. We also visualize our results while keeping the quantitative aspect, the ease of reproducibility, and enhancement of visibility. We return to this point later.

We define the affinity between fields using the information about fields that is assigned to academic journals. Our definition is applicable in cases where academic journals are assigned to one or more fields. We define the affinity of field i for field j ($i \neq j$), A_{ij}, as

$$A_{ij} \equiv \frac{\text{Number of journals assigned to } \varepsilon i \varepsilon \text{ and} \varepsilon j \varepsilon}{\text{Number of journals assigned to } \varepsilon i \varepsilon} \times 100 \tag{1}$$

Note that A_{ij} is not equal to A_{ji}; A_{ij} represents the affinity of field i for field j, while A_{ji} represents the affinity of field j for field i. The factor of 100 is multiplied for convenience. The calculation and properties of the affinity are shown in the next section. This affinity is not a direct measure of the interdisciplinarity of research, but rather serves as a weighting of the potential 'distance' between fields to the interdisciplinarity measures. We present this argument in Section "An example of applications", using similarity as an example.

Data and the method of analysis

To derive the affinity between fields, we use the field information for all academic journals in the Scopus database from Elsevier. This abstract and citation database of peer-reviewed literature covers all academic fields and contains journals, conference proceedings and books. All academic journals registered in Scopus have one or more fields assigned from the All Science Journal Classification (ASJC). The top level of this classification system has 27 fields (Table 1); the bottom level has 334 subcategories. Furthermore, these fields belong to four larger subject areas (*i.e.*, Life Sciences, Social Sciences, Physical Sciences, and Health Sciences), except the field General, which includes journals covering the field of science in general. We use Scopus data as the number of journals and 27 fields are sufficient to illustrate our new concept of the affinity between fields. Moreover, Scopus data can be downloaded by anyone, thus enabling reproducibility of our results.

An analysis using more detailed classifications could be valuable, but expansion beyond the 27 fields used here would require care, as some of the ASJC subcategories have confusing labels (*e.g.* similar fields like "Linguistics and Language" and "Language and

Table 1 ASJC codes and subject areas for the 27 fields assigned to academic journals in Scopus

ASJC code	Field (Abbreviation)	Subject area
1000	General (GENE)	–
1100	Agricultural and Biological Sciences (AGRI)	Life Sciences
1200	Arts and Humanities (ARTS)	Social Sciences
1300	Biochemistry, Genetics and Molecular Biology (BIOC)	Life Sciences
1400	Business, Management and Accounting (BUSI)	Social Sciences
1500	Chemical Engineering (CENG)	Physical Sciences
1600	Chemistry (CHEM)	Physical Sciences
1700	Computer Science (COMP)	Physical Sciences
1800	Decision Sciences (DECI)	Social Sciences
1900	Earth and Planetary Sciences (EART)	Physical Sciences
2000	Economics, Econometrics and Finance (ECON)	Social Sciences
2100	Energy (ENER)	Physical Sciences
2200	Engineering (ENGI)	Physical Sciences
2300	Environmental Science (ENVI)	Physical Sciences
2400	Immunology and Microbiology (IMMU)	Life Sciences
2500	Materials Science (MATE)	Physical Sciences
2600	Mathematics (MATH)	Physical Sciences
2700	Medicine (MEDI)	Health Sciences
2800	Neuroscience (NEUR)	Life Sciences
2900	Nursing (NURS)	Health Sciences
3000	Pharmacology, Toxicology and Pharmaceutics (PHAR)	Life Sciences
3100	Physics and Astronomy (PHYS)	Physical Sciences
3200	Psychology (PSYC)	Social Sciences
3300	Social Sciences (SOCI)	Social Sciences
3400	Veterinary (VETE)	Health Sciences
3500	Dentistry (DENT)	Health Sciences
3600	Health Professions (HEAL)	Health Sciences

Linguistics"), as pointed out in Wang & Waltman, (2016). Such expansion is beyond the scope of this study, but we hope to conduct an expanded study in future. We also note that our method for quantifying the affinity between fields can use classification systems other than ASJC to define fields.

Dataset

The quantification of affinity between fields is based on a dataset of 39,743 journals registered in Scopus as of September 2019. The journals were published from 1924 to September 2019. The dataset was retrieved from Scopus by downloading the Source title list (XLSX format; https://www.elsevier.com/solutions/scopus/how-scopus-works/content).

Each academic journal has at least one assigned field. To measure the affinity of a certain field for other fields, we aggregated the data using a symmetric matrix based on the 27 ASJC fields. An advantage of using a matrix instead of a network visualization is that the matrix simply represents the frequency of links between fields. The matrix is constructed as follows:

- If a given journal has only one assigned field, a default value of 1 is added to the relevant diagonal element (*e.g.,* If only the field "General" is assigned, 1 is added to the diagonal element in both the "General" row and column).
- If a given journal has more than one assigned field, the default value of 1 is added to all diagonal elements corresponding to those fields.
- If a given journal has two assigned fields, 1 is added to the four elements of the matrix representing the combinations of the two fields (*e.g.*, If "General" and "Chemistry" are assigned, 1 is added to the element in the "General" row and "General" column (the diagonal element), in the "Chemistry" row and "Chemistry" column (the diagonal element), in the "General" row and "Chemistry" column, and in the "Chemistry" row and "General" column).
- If a given journal has three assigned fields, 1 is added to the elements representing all nine pair-wise combinations of the three fields (*e.g.*, If "General," "Chemistry," and "Energy" are assigned, 1 is added to the element in the "General" row and "General" column, in the "Chemistry" row and "Chemistry" column, in the "Energy" row and "Energy" column, in the "General" row and the "Chemistry" column, in the "Chemistry" row and the "General" column, in the "General" row and the "Energy" column, in the "Energy" row and the "General" column, in the "Chemistry" row and the "Energy" column, and in the "Energy" row and the "Chemistry" column).
- If a given journal has four or more assigned fields, 1 is added to the corresponding diagonal and off-diagonal elements in the same way as is used for journals with three assigned fields.

We performed these operations for all 39,743 journals to complete one matrix (see Fig. 7 in the Appendix). Using this method, we obtained the total number of journals by field, which enables us to see the distribution of other fields with respect to a certain discipline.

We then normalized the values to take into account the number of journals in each discipline (Fig. 8 in the Appendix). The diagonal components of the 27 × 27 symmetric matrix obtained in Fig. 7 normalized to 100. Each matrix element in Fig. 8, $A_{ij}^{(2)}$, is based on the values in Figure 7, $A_{ij}^{(1)}$, as

$$A_{ij}^{(2)} = \frac{A_{ij}^{(1)}}{A_{ii}^{(1)}} \times 100$$

where i and j represent elements of the matrix rows and columns, in numerical order of their ASJC classification code. In this study, we call $A_{ij}^{(2)}$ the affinity between fields. The data for obtaining this affinity is updated on the Scopus approximately once every 3 months. Therefore, if a precise analysis using affinity is to be continued, it is necessary to update the data approximately once every 3 months.

It should be noted that this affinity between fields is normalized so that the diagonal elements are 100; unlike Figs. 7, 8 is not a symmetric matrix, and the meaning of the off-diagonal elements is different: $A_{ij}^{(1)} = A_{ji}^{(1)}$, but $A_{ij}^{(2)} \neq A_{ji}^{(2)}$ because $A_{ii}^{(1)} \neq A_{jj}^{(1)}$. $A_{ij}^{(2)}$ is a quantity that indicates the percentage of the field j among all journals that have been assigned to the field i when i is the field of interest, and $A_{ji}^{(2)}$ is a quantity that indicates the percentage of the field i among all journals that have been assigned to the field j. For example, where column 'MEDI' (i = Medicine) intersects with row 'NEUR' (j = Neuroscience) in Fig. 8 $A_{ji}^{(2)}$ = 58.2550. This means that about 58.3% of the academic journals assigned to Neuroscience were also assigned to Medicine. By contrast, the row of a field of interest, i, contains values representing the affinity of i to each other field, j, indicated in the relevant columns. For example, the value at the intersection of the 'MEDI' row and the 'NEUR' column in Fig. 8 is $A_{ij}^{(2)} = 3.2110$. This means that about 3.2% of the academic journals assigned to Medicine were also assigned to Neuroscience. It is also important to note that the affinity cannot determine which field individual articles within these journals belong to.

In terms of the bibliometric network, each field corresponds to a node and each off-diagonal element of the matrix is a link. When one tries to show weights on the links, the magnitudes of off-diagonal elements can be used. We also found that for a given field of interest, two types of affinity between fields can be identified: the affinity of other fields to the field of interest (shown in columns in Fig. 8), and the affinity of the field of interest to other fields (shown in rows). In general, the values of these two types of affinity differ even for pairs of fields. Therefore, in discussing affinity between fields, it is necessary to clarify the field of interest, i, and the type of affinity ($A_{ij}^{(2)}$ or $A_{ji}^{(2)}$). Thus, it is possible to understand the bidirectional affinity between fields as the flows from other fields to the field of interest as well as from the field of interest to other fields (Fig. 1). The degree of affinity is also an important factor.

These degrees of affinity can be illustrated using a Sankey diagram, as shown in Fig. 2 (all Sankey diagrams for each field can be downloaded from https://data.mendeley.com/datasets/gx8g4mfk7x/draft?a=aeb9aad2-0b12-4c2b-8e0a-1aea5b90f522). Map-type networks or cyclized maps (*e.g.* as shown in Boyac et al., 2005; Boyac & Klavans, 2014; Klavans & Boyac, 2006, 2009, 2011; Börner & Scharnhorst, 2009; Börner et al., 2012) may also be used. Although map-type networks and cyclized maps can provide an overall picture of relationships, they provide less quantitative information than our Sankey diagrams because we use one diagram for each field of interest.

Discussion

For each of the 27 fields, we determined the top three fields with the highest affinity to the field of interest as well as the top three fields to which the field of interest has the highest affinity (Table 3 in the Appendix).

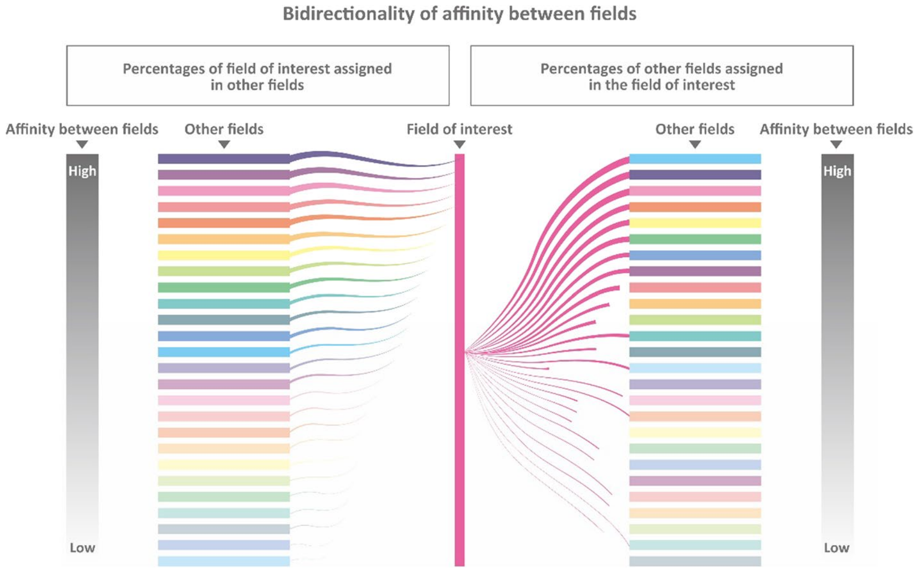

Fig. 1 The affinity between fields is bidirectional: the percentage of the field of interest assigned in other fields is not necessarily the same as the percentage of journals of other fields assigned in the field of interest. The center line corresponds to the field of interest

Of the 5 fields in the subject area of Life Sciences, “Immunology and Microbiology” has the highest affinity (28.8) to “Biochemistry, Genetics and Molecular Biology”, indicating that 28.8% of journals in the field of “Immunology and Microbiology” are also in the field of “Biochemistry, Genetics and Molecular Biology”. (Note that other affinity values shown below should also be interpreted as the percentage of journals assigned to a field.) “Biochemistry, Genetics and Molecular Biology” also has the highest affinity (8.40) to “Immunology and Microbiology”. Four of the five fields in the Life Sciences have the highest affinity to “Medicine” (“Biochemistry, Genetics and Molecular Biology” (43.1), “Immunology and Microbiology” (57.5), “Neuroscience” (58.3), and “Pharmacology, Toxicology and Pharmaceutics” (43.8), which is categorized in Health Sciences.

Of the 6 fields in Social Sciences, “Arts and Humanities” has the highest affinity (58.1) to “Social Sciences”, and “Social Sciences” has the highest affinity (35.5) to “Arts and Humanities”. These values are also the highest from the standpoint of each field. Three fields have the highest affinity to “Social Sciences”: “Arts and Humanities” (58.1), “Business, Management and Accounting” (25.0), and “Psychology” (37.1).

Of the 10 fields in Physical Sciences, “Chemistry” and “Chemical Engineering” have the highest affinity to each other (27.2 and 32.3, respectively), and “Environmental Sciences” and “Earth and Planetary Sciences” have the highest affinity to each other (33.6 and 37.6, respectively). About the "Chemical Engineering", "Environmental Sciences", and “Earth and Planetary Sciences”, the affinities for "Chemistry" (32.3), "Earth and Planetary Sciences" (33.6), and "Environmental Sciences" (37.6) are also the highest ones from the standpoint of each field, respectively.

Of the 5 fields in Health Sciences, “Nursing”, “Dentistry”, and “Health Professions” have the highest affinity to “Medicine” (62.5, 32.1, and 67.2, respectively). By contrast, “Medicine” as the field of interest has low affinity for fields in Health Sciences.

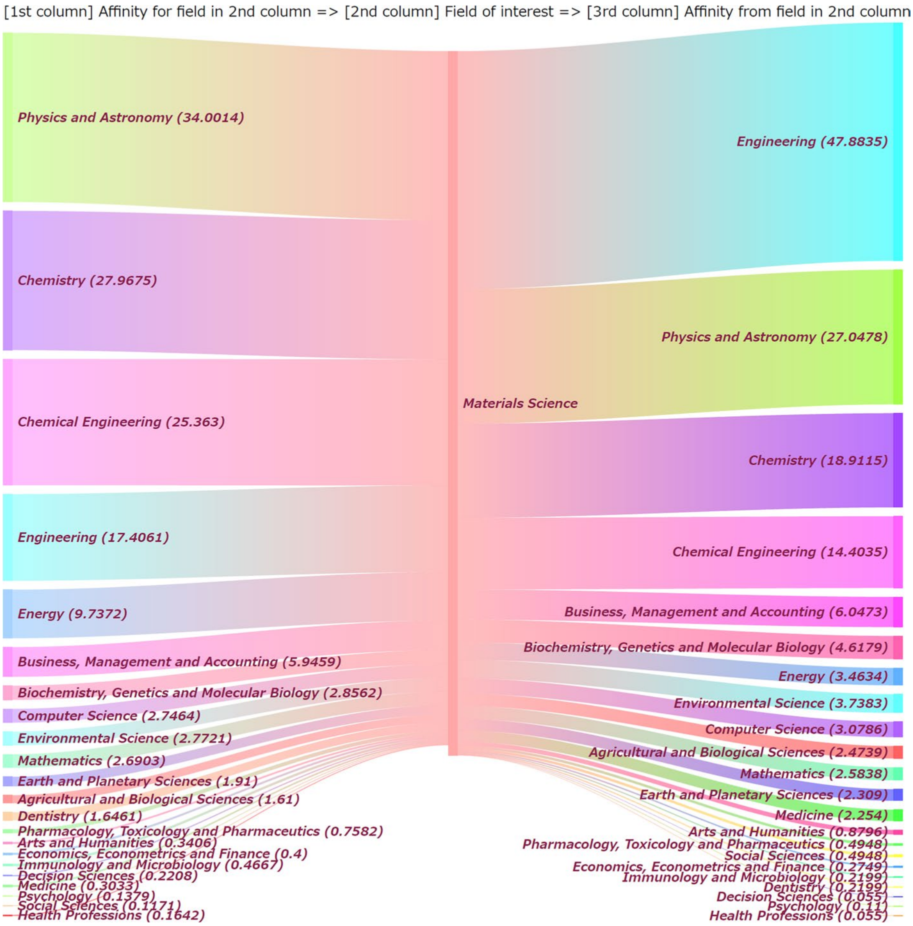

Fig. 2 A Sankey diagram for Materials Science as the field of interest. The field of interest is represented as the center column. To the left are other fields with affinity to Materials Science; to the right are fields to which Materials Science has affinity

In some cases, the field with the highest affinity to the field of interest is also the field to which the field of interest has the highest affinity. This is the case for nine fields of interest (presented as field of interest: field with highest affinity):

- "Arts and Humanities": "Social Sciences"
- "Chemical Engineering": "Chemistry"
- "Earth and Planetary Sciences": "Environmental Sciences"
- "Economics, Econometrics and Finance": "Business, Management and Accounting"
- "Engineering": "Materials Science"
- "Environmental Sciences": "Earth and Planetary Sciences"
- "Social Sciences": "Arts and Humanities"
- "Veterinary": "Agricultural and Biological Sciences"
- "Dentistry": "Medicine"

Affinity to “Medicine” is particularly high, as this field appears in the top position for seven fields of interest: “Biochemistry, Genetics and Molecular Biology”, “Immunology and Microbiology”, “Neuroscience”, “Pharmacology, Toxicology and Pharmaceutics”, “Nursing”, “Dentistry”, and “Health Professions”.

Based on our values for the affinity between fields, we calculated the mean and median values for other fields’ affinity to the field of interest (matrix elements in columns in Fig. 8) and for the affinity of the field of interest for other fields (matrix elements in rows), which are shown in Fig. 3 and Fig. 4, respectively. Each figure includes box plots of affinity for each of the 27 fields of interest; circles indicate the other fields’ affinity to the field of interest (Fig. 3 and Table C) or the affinity of the field of interest to other fields (Fig. 4 and Table D), while crosses represent the mean value.

Within each subject area in Fig. 3, the mean and median affinity are highest for the fields of “Biochemistry, Genetics and Molecular Biology”, “Social Sciences”, “Engineering”, and “Medicine”. These fields, which have the largest number of academic journals in each subject area, are positioned as core fields and can be clearly differentiated from other fields in Fig. 3. In Fig. 4, there is no marked difference between these four core fields and the remaining fields, which means that there is no marked difference in the intermediate layers of the affinity of each field of interest to other fields (the positions of the third and first quartiles, and the mean and median values). Meanwhile, high affinity to core fields is often seen as outliers for each field of interest, such as the affinity of “Arts and Humanities” to “Social Sciences” (58.1), of “Materials Science” to “Engineering” (47.9), and of “Nursing” and “Health Professions” to “Medicine” (62.5 and 67.2, respectively). The affinities of two fields of interest to “Medicine”, “Immunology and Microbiology” (57.5) and “Neuroscience” (58.3), also represent outliers from another subject area (Life Sciences).

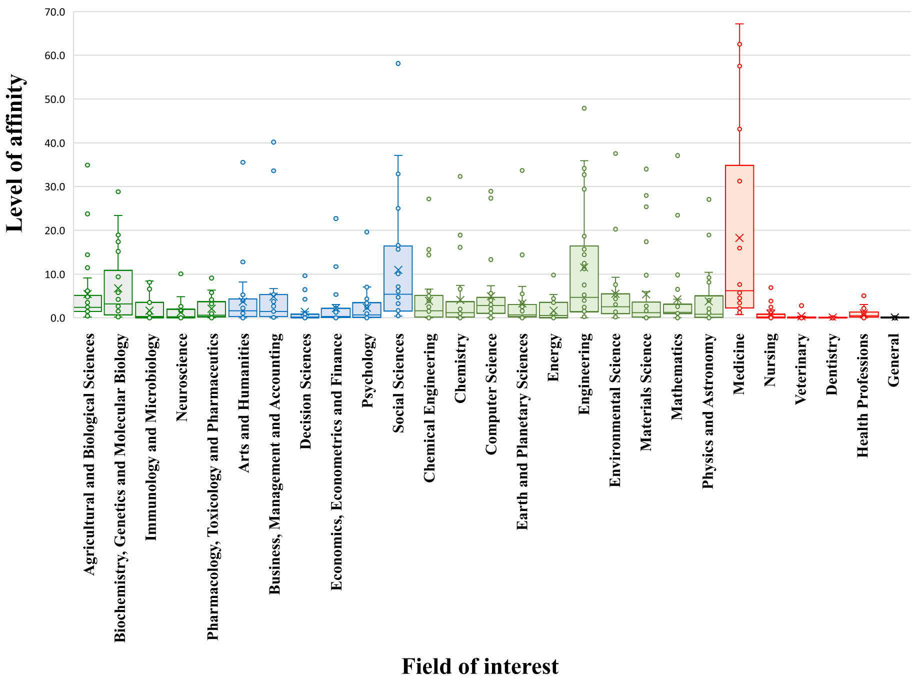

Fig. 3 Mean and median affinity of other fields to the field of interest (using values from columns in Fig. 2)

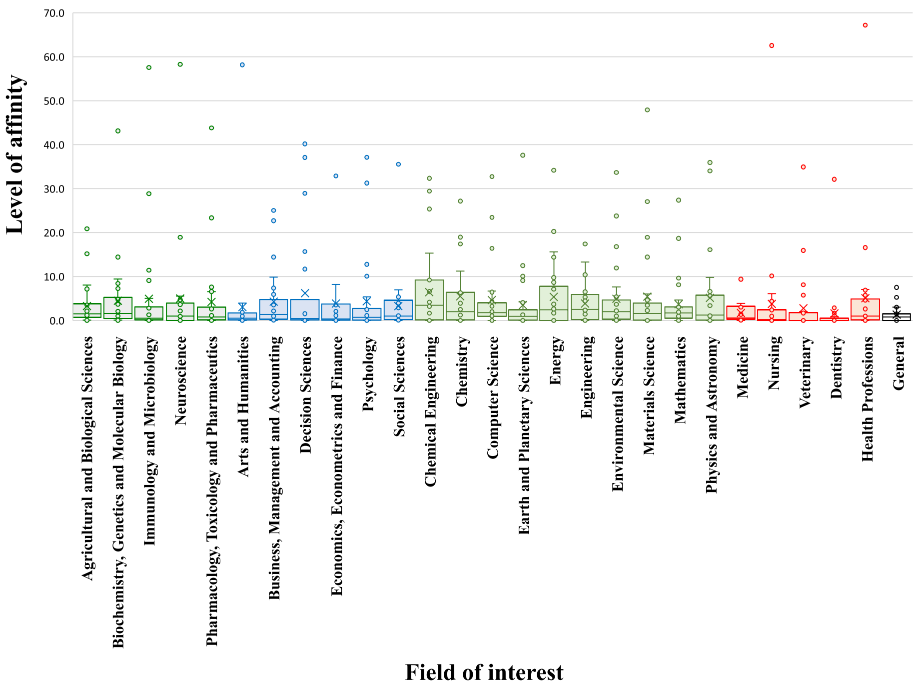

Fig. 4 Mean and median affinity of the field of interest to other fields (using values from rows in Fig. 2)

These points can be demonstrated quantitatively by measuring correlations between the numbers of journals and affinity in each field. Figures 5 and 6 show the correlations between the numbers of journals of each field (using values in Fig. 7) and values of affinity of other fields to the field of interest (Fig. 5: using values in Fig. 3) and of the field of interest to other fields (Fig. 6: using values in Fig. 4), respectively. The upper and lower plots represent the correlations between the numbers of journals and mean and median values of affinity, respectively. Each point in the figures corresponds to each field. Figure 5 shows that there is a correlation between the number of journals in each field and both the mean and median of affinity. Their correlation coefficients are 0.933 ($p \cong 1.35 \times 10^{-12}$) and 0.882 ($p \cong 1.21 \times 10^{-9}$), respectively, indicating a strong positive correlation. On the other hand, Fig. 6 shows that there is no strong correlation between them. In fact, correlation coefficients between the number of journals and the mean and median of affinity in each field are -0.385 ($p \cong 0.0472$) and 0.00177 ($p \cong 0.993$), respectively, indicating that the number of journals and the mean of affinity have a weak negative correlation.

An example of application

Since the Salton-cosine similarity (Salton & McGill, 1983) is one of the most fundamental measures of interdisciplinarity, we examine the extent to which the affinity obtained in Section "Data and the method of analysis" affects the similarity. To demonstrate this clearly, it is appropriate to consider a simple sample of data. Therefore, we use the following table (matrix) for the three fields of Medicine (MEDI), Neuroscience

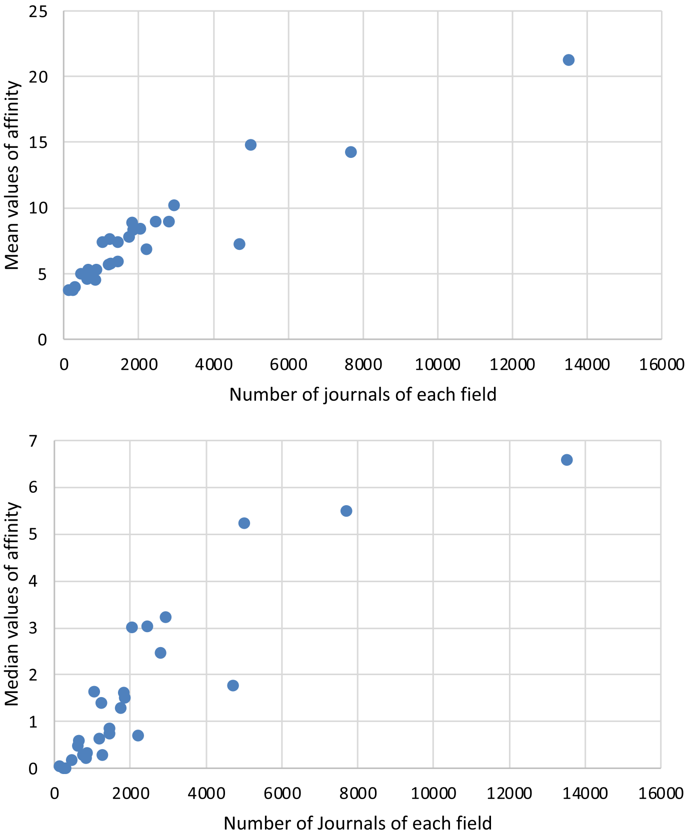

Fig. 5 Correlations between the numbers of journals of each field (using values in Fig. 7) and mean (upper figure) and median (lower one) values of affinity of other fields to the field of interest (using values in Fig. 3)

(NEUR), and Nursing (NURS) as an example. We randomly generate the following number of papers for the three fields to see the impact of the affinity as follows:

Fields of target papers/Fields of cited papers from target papers	MEDI	NEUR	NURS
MEDI	1268	4145	5476
NEUR	2058	2848	6195
NURS	6842	2993	4884

Here, the labels (fields) in the rows of the matrix above indicate the assigned fields of a certain set of target papers, and those papers will usually cite several papers. The labels (fields) in the columns of the matrix indicate the fields of the papers cited by those papers. The values in the table (matrix) indicate the number of papers, and we randomly generate those values by utilizing "randbetween(0,10,000)" in the Microsoft Office Excel. Using the table, non-trivial ($\neq 1$) similarities (S_{ij}: i and j stand for fields, in this case, MEDI, NEUR, and NURS) among the three target fields are calculated as follows,

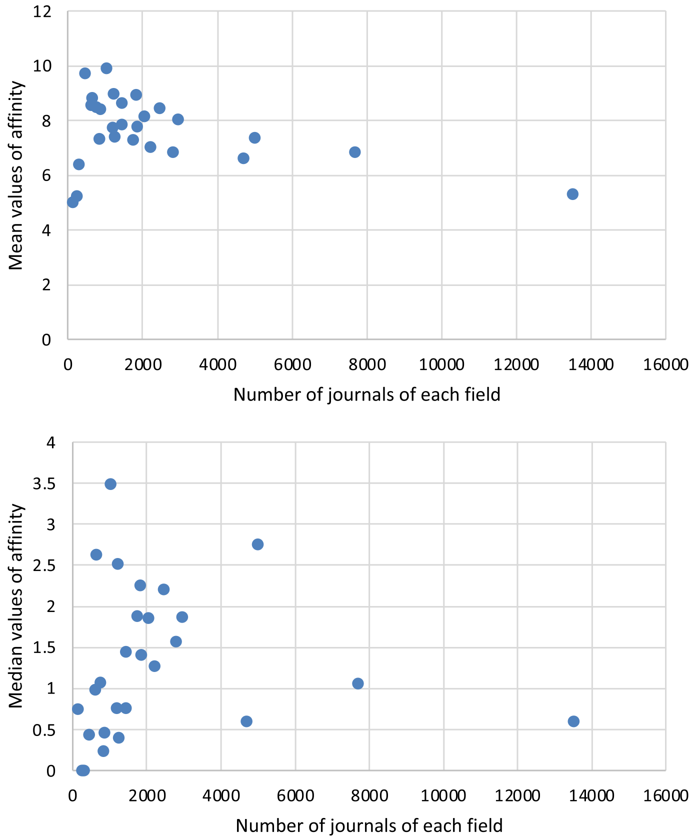

Fig. 6 Correlations between the numbers of journals of each field (using values in Fig. 7) and mean (upper figure) and median (lower one) values of affinity of the field of interest to other fields (using values in Fig. 4)

$$S_{\text{MEDI,NEUR}} \fallingdotseq 0.972, S_{\text{MEDI,NURS}} \fallingdotseq 0767, S_{\text{NEUR,NURS}} \fallingdotseq 0.795, \tag{2}$$

where the similarity is defined as

$$S_{ij} \equiv \frac{\sum_f^p x_{if} x_{jf}}{\sqrt{\sum_f^p x_{if}^2 \sum_f^p x_{jf}^2}}. \tag{3}$$

The similarity is invariant with respect to the interchange of i and j (symmetric with respect to i and j: $S_{ij} = S_{ji}$). Also, i, j, and f denote the fields, and in the present case, the labels are assigned to the 27 ASJC fields shown in Table 2. In addition, p is all the fields under consideration, and if all ASJC fields are considered, the sum is taken for the 27 fields, or in the current example, MEDI, NEUR, and NURS. In this example, we find

$$S_{\text{MEDI, NEUR}} > S_{\text{NEUR, NURS}} > S_{\text{MEDI, NURS}}. \tag{4}$$

Here, we consider how to consider the influence of the affinity $A_{ij}^{(2)}$ obtained in Section "Data and the method of analysis" on this similarity. We propose two intuitively understandable and simple methods as examples. They are (i) multiplying the similarity by the affinity and (ii) adding the similarity and the affinity. In (i), the affinity plays

Table 2 Summary of three examples for the similarities

Fields i, j	Similarity:S_{ij}	Similarity × Affinity:S_{ij}^{M}	Similarity + Affinity/100:S_{ij}^{A}
MEDI, NEUR	0.972	3.12	1.00
NEUR, MEDI	0.972	56.6	1.55
MEDI, NURS	0.767	2.94	0.806
NURS, MEDI	0.767	48.0	1.39
NEUR, NURS	0.795	1.39	0.812
NURS, NEUR	0.795	1.24	0.810

a role as a weighting factor for the similarity, and in (ii), the similarity and the affinity can be understood as equivalent information to be added together (However, as will be shown later, the affinity is to be treated as a frequency distributed between 0 and 1 as in the similarity, so the affinity divided by 100 is to be added. This means simply removing the 100-fold factor introduced for convenience in Eq. (1).)

First, we consider (i) multiplying the similarity by the affinity and define as a new similarity S_{ij}^{M} as

$$S_{ij}^{\mathrm{M}} \equiv A_{ij}^{(2)} S_{ij.} \tag{5}$$

Note that while the similarity given in Eq. (3) is invariant with respect to the interchange of i and j, the affinity calculated in Section "Data and the method of analysis" has bidirectional information (asymmetric with respect to i and j: $A_{ij}^{(2)} \neq A_{ji}^{(2)}$), and thus, the new similarity defined in Eq. (5) has also bidirectional information, $S_{ij}^{\mathrm{M}} \neq S_{ji}^{\mathrm{M}}$. Following this definition and calculating a new nontrivial ($\neq 1$) similarity S_{ij}^{M} for the three fields MEDI, NEUR, and NURS as above, we obtain

$$\begin{aligned} &S_{\mathrm{MEDI,NEUR}}^{\mathrm{M}} \fallingdotseq 3.12, S_{\mathrm{NEUR,MEDI}}^{\mathrm{M}} \fallingdotseq 56.6, \\ &S_{\mathrm{MEDI,NURS}}^{\mathrm{M}} \fallingdotseq 2.94, S_{\mathrm{NURS,MEDI}}^{\mathrm{M}} \fallingdotseq 48.0, \\ &S_{\mathrm{NEUR,NURS}}^{\mathrm{M}} \fallingdotseq 1.39, S_{\mathrm{NURS,NEUR}}^{\mathrm{M}} \fallingdotseq 1.24, \end{aligned} \tag{6}$$

and thus, we fined

$$S_{\mathrm{NEUR,\,MEDI}}^{\mathrm{M}} > S_{\mathrm{NURS,\,MEDI}}^{\mathrm{M}} > S_{\mathrm{MEDI,\,NEUR}}^{\mathrm{M}} > S_{\mathrm{MEDI,\,NURS}}^{\mathrm{M}} > S_{\mathrm{NEUR,\,NURS}}^{\mathrm{M}} > S_{\mathrm{NURS,\,MEDI}}^{\mathrm{M}}, \tag{7}$$

where we use for the affinity $A_{ij}^{(2)}$ as $A_{\mathrm{MEDI,NEUR}}^{(2)} \fallingdotseq 3.21$, $A_{\mathrm{NEUR,MEDI}}^{(2)} \fallingdotseq 58.3$, $A_{\mathrm{MEDI,NURS}}^{(2)} \fallingdotseq 3.84$, $A_{\mathrm{NURS,MEDI}}^{(2)} \fallingdotseq 62.5$, $A_{\mathrm{NEUR,NURS}}^{(2)} \fallingdotseq 1.74$, $A_{\mathrm{NURS,NEUR}}^{(2)} \fallingdotseq 1.57$ obtained in Section "Data and the method of analysis" (Fig. 3).

Next, we consider (ii) adding the similarity and the affinity and define as another new similarity S_{ij}^{A} as

$$S_{ij}^{\mathrm{A}} \equiv S_{ij} + A_{ij}^{(2)}/100. \tag{8}$$

Note that 100-fold factor in the second term is introduced to be the affinity treated as appropriate degree for the similarity S_{ij}, which is in the range from 0 to 1, and S_{ij}^{A} has also bidirectional information, $S_{ij}^{\mathrm{A}} \neq S_{ji}^{\mathrm{A}}$. Following this definition and calculating another similarity S_{ij}^{A} for the three fields MEDI, NEUR, and NURS as above, we obtain

$$
\begin{aligned}
&S_{\mathrm{MEDI,NEUR}}^{\mathrm{A}} \fallingdotseq 1.00, S_{\mathrm{NEUR,MEDI}}^{\mathrm{A}} \fallingdotseq 1.55,\\
&S_{\mathrm{MEDI,NURS}}^{\mathrm{A}} \fallingdotseq 0.806, S_{\mathrm{NURS,MEDI}}^{\mathrm{A}} \fallingdotseq 1.39,\\
&S_{\mathrm{NEUR,NURS}}^{\mathrm{A}} \fallingdotseq 0.812, S_{\mathrm{NURS,NEUR}}^{\mathrm{A}} \fallingdotseq 0.810,
\end{aligned}
\tag{9}
$$

and thus, we finedwhere we use for the same values of affinity as calculating for Eq. (7).

$$
S_{\mathrm{NEUR,\ MEDI}}^{\mathrm{A}} > S_{\mathrm{NURS,\ MEDI}}^{\mathrm{A}} > S_{\mathrm{MEDI,\ NEUR}}^{\mathrm{A}} > S_{\mathrm{NEUR,\ NURS}}^{\mathrm{A}} > S_{\mathrm{NURS,\ MEDI}}^{\mathrm{A}} > S_{\mathrm{MEDI,\ NURS}}^{\mathrm{A}}, \tag{10}
$$

The results of the three different calculations of similarities for the test data up to this point are summarized in Table 1. As mentioned above, when neuroscience and nursing are the fields of interest, the affinities with medicine are high, which also influences the similarity calculations for the sample data conducted in this study. It is also possible to quantitatively reflect the bidirectional nature of affinity (i.e., the difference between which is the field of interest) in the similarity. In our two examples of applying affinity to similarity, we can see the following differences and commonality. S_{ij}^{M}, which is affinity multiplied to similarity, strongly reflects the relationship between the size of the affinity and amplifies (strongly weighted by the affinity). On the other hand, S_{ij}^{A} is determined by the balance between the similarity and affinity. This can be seen from the difference in the order of $S_{\mathrm{MEDI,NURS}}^{\mathrm{M}}$, $S_{\mathrm{NEUR,NURS}}^{\mathrm{M}}$, and $S_{\mathrm{NURS,MEDI}}^{\mathrm{M}}$ in Eqs. (7) and (8). The commonality is that the actual values calculated from the conventional similarity Eq. (3) (Eq. (2)) and the relationship among them (Eq. (4)), as described above, have been extended to the values (Eqs. (6) and (9)) and the relationships among them (Eqs. (7) and (10)) calculated from the definition including bidirectionality (Eqs. (5) and (8)), and resulted in a higher resolution of similarity. Furthermore, in the case presented here, reflecting the potential high affinity of neuroscience and nursing for medicine, $S_{\mathrm{NEUR,MEDI}}^{\mathrm{M}}$, $S_{\mathrm{NURS,MEDI}}^{\mathrm{M}}$ and $S_{\mathrm{NEUR,MEDI}}^{\mathrm{A}}$, $S_{\mathrm{NURS,MEDI}}^{\mathrm{A}}$ appear as having higher similarity than $S_{\mathrm{MEDI,NEUR}}^{\mathrm{M}}$ and $S_{\mathrm{MEDI,NEUR}}^{\mathrm{A}}$, respectively (note that $S_{\mathrm{MEDI,NEUR}}$ has shown the highest similarity in the conventional relationship, Eq. (4)).

The two new similarities proposed in this study extend the relationships among fields that can be measured by the similarity (doubling the resolution of the similarity) due to the bidirectional nature of the affinity, and also incorporate a feature of the affinity among fields derived from the number of journals. When discussing similarity due to researchers' activities (citations etc.), the conventional similarity S_{ij} (Eq. (3)) is used, and when discussing similarity reflecting the field distribution of journals, S_{ij}^{A} (Eq. (8)) is used. And when discussing similarity that most strongly reflects the field distribution of journals, the use of S_{ij}^{M} (Eq. (5)) is appropriate. Thus, it becomes possible to use different methods according to the analyst's intention. In this study, we took the similarity as an example of an interdisciplinary research indicator that is affected by the affinity and showed its quantitative impact. The affinities derived from the field distribution in academic journals and their bidirectionality can be incorporated into other interdisciplinary research indicators to measure them more precisely and from a broader perspective.

Limitations of our analyses and results

In proposing a new concept of affinity between fields, we have demonstrated the idea and calculation using information about the 27 academic fields assigned to journals in Scopus. All academic journals in Scopus are assigned at least one academic field from the ASJC classification system (however, we note that the method of this field classification has not been clarified). It should be noted that these fields we used are not the actual academic fields, but the classifications presented by Scopus.

The ASJC system classifies all fields, except for "General", into four broader subject areas. In addition, the 27 fields have 334 more detailed subcategories. This classification of these fields was an appropriate size for demonstrating our concept, but it provided a coarser-resolution analysis than we might have achieved using the 334 subcategories. It is possible to calculate affinity in the same manner using these subcategories, but in that case, it would be necessary to consider the problems inherent in ASJC classification system. For example, field classifications are sometimes confusing, and some fields appear to be very similar. For this reason, we did not extend our calculation of affinity values to the 334 subcategories in this paper.

As Scopus assigns ASJC fields to a journal, its classification does not take into account whether the articles published in that journal actually fit into that field. In addition, if a journal is assigned multiple fields, we counted the fields equally; we did not take into account how the collection of articles published in the journal might be biased towards one field.

Scopus is not the only large database of scholarly publications; Web of Science by Clarivate Analytics (previously Web of Knowledge) comprises a suite of databases of citation data in different disciplines. It uses approximately 250 Research Areas to classify content, as well as 22 broad research disciplines in its Essential Science Indicators tool. Other field classification systems include the US National Science Foundation classification system, the Science-Metrix classification system, the University of California San Diego classification system, the Australian and New Zealand Standard Research Classification, and the Chinese Library Classification. Our concept of affinity and the calculation method presented in this paper can be used with other classification systems. However, because the academic fields are different for each classification system, it is not possible to directly compare the affinity values presented in this paper with values based on other classification systems.

We summarize our results

Here, we have proposed a new concept: the affinity between fields in academic research. We define the affinity as the number of journals assigned to two fields, divided by the number of journals assigned to one of those two fields. The affinity should be examined from two perspectives: the affinity of other fields to the field of interest, and the affinity of the field of interest to other fields. To derive the affinity, we have used information on the academic fields of all journals in Scopus, the largest database of peer-reviewed literature, which covers all academic fields and contains journals, conference proceedings and books. All academic journals in Scopus have one or more fields assigned from a portfolio of the 27 fields. With the exception of the field "General", these fields are all categorized into four larger subject areas. Scopus data and the 27 fields in the ASJC classification system are an appropriate size for demonstrating our new concept of the affinity between fields. Moreover, Scopus data is freely downloadable by anyone, thus making it easier for others to reproduce our results.

Our detailed analyses reveal the affinity between fields. In the Life Sciences, “Immunology and Microbiology” is the field with the highest affinity (28.8) to “Biochemistry, Genetics and Molecular Biology”, which indicates that 28.8% of the academic journals in “Immunology and Microbiology” are also assigned to “Biochemistry, Genetics and Molecular Biology”. Similarly, “Biochemistry, Genetics and Molecular Biology” has the highest affinity (8.40) to “Immunology and Microbiology”. In the subject area of Social Sciences, “Arts and Humanities” has the highest affinity (58.1) to the field of “Social Sciences”, and “Social Sciences” has the highest affinity (35.5) to “Arts and Humanities”. These values are also the highest from the standpoint of each field. In Physical Sciences, “Chemistry” and “Chemical Engineering” have the highest affinity for each other (27.2 and 32.3, respectively), as do “Environmental Sciences” and “Earth and Planetary Sciences” (33.6 and 37.6, respectively). In Health Sciences, “Nursing”, “Dentistry” and “Health Professions” each have the highest affinity to “Medicine” (62.5, 32.1 and 67.2, respectively). The fields with the highest bidirectional affinity are “Arts and Humanities” and “Social Sciences”, and “Earth and Planetary Sciences” and “Environmental Sciences”. Medicine is the field that most often has the highest affinity to the field of interest, securing the top position for seven fields (the first four of which are in Life Sciences): “Biochemistry, Genetics and Molecular Biology” (43.1), “Immunology and Microbiology” (57.5), “Neuroscience” (58.3), “Pharmacology, Toxicology and Pharmaceutics” (43.8), “Nursing” (62.5), “Dentistry” (32.1), and “Health Professions” (67.2).

The affinity plays a weighting role in the indicators calculated from citation relationships etc., for example, as shown in Section “An example of applications”. And this means that the discussion of the degree of interdisciplinary research derived from it will be more reflective of the current relations among fields. When discussing similarity due to researchers’ activities (citations etc.), the conventional similarity S_{ij} (Eq. (3)) is used, and when discussing similarity reflecting the field distribution of journals, S_{ij}^{A} (Eq. (8)) is used. And when discussing similarity that most strongly reflects the field distribution of journals, the use of S_{ij}^{M} (Eq. (5)) is appropriate. Thus, it becomes possible to use different methods according to the analyst’s intention. In addition, by reflecting the bidirectional nature of the affinity, it is possible to give a new bidirectional view to characteristics (e.g., similarity) that have been expressed only in a single direction so far. The concept of the affinity between fields might also be used when researchers (and/or university research administrators) are considering an extension of their research theme and to analyze research trends. And it could be used to help researchers find suitable academic journals for submitting their work. From the perspective of analyzing the research activities of research institutions and other organizations, the introduction of the affinity concept is expected to improve the accuracy of measuring the interdisciplinarity of research, thereby contributing to the validation of the effectiveness of interdisciplinary research and the discussion of research strategy formulation. Concrete applications, drill-down on smaller subcategories, and analysis of other journal subject classifications are subject to further discussions.

Appendix

See Figs. 7 and 8; Tables 3, 4, and 5.

Row\ Column	GENE	AGRI	ARTS	BIOC	BUSI	CENG	CHEM	COMP	DECI	EART	ECON	ENER	ENGI	ENVI	IMMU	MATE	MATH	MEDI	NEUR	NURS	PHAR	PHYS	PSYC	SOCI	VETE	DENT	HEAL
GENE	133	2	7	1	2	2	2	4	0	1	1	0	10	0	0	0	4	1	0	0	0	2	0	7	0	0	1
AGRI	2	2795	33	424	42	44	59	25	7	200	44	20	117	583	98	45	33	226	21	51	37	21	16	113	103	1	1
ARTS	7	33	4697	11	80	2	7	88	8	32	102	1	114	45	0	16	39	119	29	20	2	15	185	2731	0	0	28
BIOC	1	424	11	2941	5	158	214	66	1	12	2	15	154	108	247	84	55	1268	141	22	277	61	22	30	17	7	39
BUSI	2	42	80	5	1850	46	18	136	182	13	420	24	267	65	1	110	26	40	2	5	5	5	78	463	1	0	6
CENG	2	44	2	158	46	1033	334	48	1	36	3	101	304	96	43	262	35	68	2	0	33	95	1	17	0	0	1
CHEM	2	59	7	214	18	334	1230	41	0	31	2	48	138	57	4	344	36	81	0	0	78	233	1	15	0	1	6
COMP	4	25	88	66	136	48	41	2039	131	32	26	26	667	28	5	56	478	75	35	1	13	81	38	334	0	0	21
DECI	0	7	8	1	182	1	0	131	453	2	53	2	71	7	0	1	168	11	0	1	1	4	11	71	0	0	2
EART	1	200	32	12	13	36	31	32	2	2199	4	93	274	826	3	42	28	39	1	1	0	89	1	221	0	1	0
ECON	1	44	102	2	420	3	2	26	53	4	1250	17	14	54	0	5	53	15	1	1	3	0	19	411	1	0	0
ENER	0	20	1	15	24	101	48	26	2	93	17	647	221	131	1	63	18	11	0	0	1	56	0	46	0	0	0
ENGI	10	117	114	154	267	304	138	667	71	274	14	221	5004	293	13	871	326	226	8	2	31	520	7	304	0	2	19
ENVI	0	583	45	108	65	96	57	28	7	826	54	131	293	2453	26	68	25	187	7	2	90	21	3	412	6	1	5
IMMU	0	98	0	247	1	43	4	5	0	3	0	1	13	26	857	4	11	493	27	4	78	1	0	4	24	2	5
MATE	0	45	16	84	110	262	344	56	1	42	5	63	871	68	4	1819	47	41	0	0	9	492	2	9	0	4	1
MATH	4	33	39	55	26	35	36	478	168	28	53	18	326	25	11	47	1747	38	9	2	8	142	17	82	0	0	7
MEDI	1	226	119	1268	40	68	81	75	11	39	15	11	226	187	493	41	38	13516	434	519	520	49	453	441	47	78	409
NEUR	0	21	29	141	2	2	0	35	0	1	1	0	8	7	27	0	9	434	745	13	31	1	146	41	0	0	16
NURS	0	51	20	22	5	0	0	1	1	1	1	0	2	2	4	0	2	519	13	830	8	0	36	84	1	0	42
PHAR	0	37	2	277	5	33	78	13	1	0	3	1	31	90	78	9	8	520	31	8	1187	0	9	15	5	2	35
PHYS	2	21	15	61	5	95	233	81	4	89	0	56	520	21	1	492	142	49	1	0	0	1447	0	26	0	0	7
PSYC	0	16	185	22	78	1	1	38	11	1	19	0	7	3	0	2	17	453	146	36	9	0	1450	538	1	0	43
SOCI	7	113	2731	30	463	17	15	334	71	221	411	46	304	412	4	9	82	441	41	84	15	26	538	7686	6	1	101
VETE	0	103	0	17	1	0	0	0	0	0	1	0	0	6	24	0	0	47	0	1	5	0	1	6	295	0	1
DENT	0	1	0	7	0	0	1	0	0	1	0	0	2	1	2	4	0	78	0	0	2	0	0	1	0	243	1
HEAL	1	1	28	39	6	1	6	21	2	0	0	0	19	5	5	1	7	409	16	42	35	7	43	101	1	1	609

Fig. 7 Number of academic journals in Scopus to which each field was assigned

Row \ Column	GENE	AGRI	ARTS	BIOC	BUSI	CENG	CHEM	COMP	DECI	EART	ECON	ENER	ENGI	ENVI	IMMU	MATE	MATH	MEDI	NEUR	NURS	PHAR	PHYS	PSYC	SOCI	VETE	DENT	HEAL
GENE	100.00	1.50	5.26	0.75	1.50	1.50	1.50	3.01	0.00	0.75	0.75	0.00	7.52	0.00	0.00	0.00	3.01	0.75	0.00	0.00	0.00	1.50	0.00	5.26	0.00	0.00	0.75
AGRI	0.07	100.00	1.18	15.17	1.50	1.57	2.11	0.89	0.25	7.16	1.57	0.72	4.19	20.86	3.51	1.61	1.18	8.09	0.75	1.82	1.32	0.75	0.57	4.04	3.69	0.04	0.04
ARTS	0.15	0.70	100.00	0.23	1.70	0.04	0.15	1.87	0.17	0.68	2.17	0.02	2.43	0.96	0.00	0.34	0.83	2.53	0.62	0.43	0.04	0.32	3.94	58.14	0.00	0.00	0.60
BIOC	0.03	14.42	0.37	100.00	0.17	5.37	7.28	2.24	0.03	0.41	0.07	0.51	5.24	3.67	8.40	2.86	1.87	43.11	4.79	0.75	9.42	2.07	0.75	1.02	0.58	0.24	1.33
BUSI	0.11	2.27	4.32	0.27	100.00	2.49	0.97	7.35	9.84	0.70	22.70	1.30	14.43	3.51	0.05	5.95	1.41	2.16	0.11	0.27	0.27	0.27	4.22	25.03	0.05	0.00	0.32
CENG	0.19	4.26	0.19	15.30	4.45	100.00	32.33	4.65	0.10	3.48	0.29	9.78	29.43	9.29	4.16	25.36	3.39	6.58	0.19	0.00	3.19	9.20	0.10	1.65	0.00	0.00	0.10
CHEM	0.16	4.80	0.57	17.40	1.46	27.15	100.00	3.33	0.00	2.52	0.16	3.90	11.22	4.63	0.33	27.97	2.93	6.59	0.00	0.00	6.34	18.94	0.08	1.22	0.00	0.08	0.49
COMP	0.20	1.23	4.32	3.24	6.67	2.35	2.01	100.00	6.42	1.57	1.28	1.28	32.71	1.37	0.25	2.75	23.44	3.68	1.72	0.05	0.64	3.97	1.86	16.38	0.00	0.00	1.03
DECI	0.00	1.55	1.77	0.22	40.18	0.22	0.00	28.92	100.00	0.44	11.70	0.44	15.67	1.55	0.00	0.22	37.09	2.43	0.00	0.22	0.22	0.88	2.43	15.67	0.00	0.00	0.44
EART	0.05	9.10	1.46	0.55	0.59	1.64	1.41	1.46	0.09	100.00	0.18	4.23	12.46	37.56	0.14	1.91	1.27	1.77	0.05	0.05	0.00	4.05	0.05	10.05	0.00	0.05	0.00
ECON	0.08	3.52	8.16	0.16	33.60	0.24	0.16	2.08	4.24	0.32	100.00	1.36	1.12	4.32	0.00	0.40	4.24	1.20	0.08	0.08	0.24	0.00	1.52	32.88	0.08	0.00	0.00
ENER	0.00	3.09	0.15	2.32	3.71	15.61	7.42	4.02	0.31	14.37	2.63	100.00	34.16	20.25	0.15	9.74	2.78	1.70	0.00	0.00	0.15	8.66	0.00	7.11	0.00	0.00	0.00
ENGI	0.20	2.34	2.28	3.08	5.34	6.08	2.76	13.33	1.42	5.48	0.28	4.42	100.00	5.86	0.26	17.41	6.51	4.52	0.16	0.04	0.62	10.39	0.14	6.08	0.00	0.04	0.38
ENVI	0.00	23.77	1.83	4.40	2.65	3.91	2.32	1.14	0.29	33.67	2.20	5.34	11.94	100.00	1.06	2.77	1.02	7.62	0.29	0.08	3.67	0.86	0.12	16.80	0.24	0.04	0.20
IMMU	0.00	11.44	0.00	28.82	0.12	5.02	0.47	0.58	0.00	0.35	0.00	0.12	1.52	3.03	100.00	0.47	1.28	57.53	3.15	0.47	9.10	0.12	0.00	0.47	2.80	0.23	0.58
MATE	0.00	2.47	0.88	4.62	6.05	14.40	18.91	3.08	0.05	2.31	0.27	3.46	47.88	3.74	0.22	100.00	2.58	2.25	0.00	0.00	0.49	27.05	0.11	0.49	0.00	0.22	0.05
MATH	0.23	1.89	2.23	3.15	1.49	2.00	2.06	27.36	9.62	1.60	3.03	1.03	18.66	1.43	0.63	2.69	100.00	2.18	0.52	0.11	0.46	8.13	0.97	4.69	0.00	0.00	0.40
MEDI	0.01	1.67	0.88	9.38	0.30	0.50	0.60	0.55	0.08	0.29	0.11	0.08	1.67	1.38	3.65	0.30	0.28	100.00	3.21	3.84	3.85	0.36	3.35	3.26	0.35	0.58	3.03
NEUR	0.00	2.82	3.89	18.93	0.27	0.27	0.00	4.70	0.00	0.13	0.13	0.00	1.07	0.94	3.62	0.00	1.21	58.26	100.00	1.74	4.16	0.13	19.60	5.50	0.00	0.00	2.15
NURS	0.00	6.14	2.41	2.65	0.60	0.00	0.00	0.12	0.12	0.12	0.12	0.00	0.24	0.24	0.48	0.00	0.24	62.53	1.57	100.00	0.96	0.00	4.34	10.12	0.12	0.00	5.06
PHAR	0.00	3.12	0.17	23.34	0.42	2.78	6.57	1.10	0.08	0.00	0.25	0.08	2.61	7.58	6.57	0.76	0.67	43.81	2.61	0.67	100.00	0.00	0.76	1.26	0.42	0.17	2.95
PHYS	0.14	1.45	1.04	4.22	0.35	6.57	16.10	5.60	0.28	6.15	0.00	3.87	35.94	1.45	0.07	34.00	9.81	3.39	0.07	0.00	0.00	100.00	0.00	1.80	0.00	0.00	0.48
PSYC	0.00	1.10	12.76	1.52	5.38	0.07	0.07	2.62	0.76	0.07	1.31	0.00	0.48	0.21	0.00	0.14	1.17	31.24	10.07	2.48	0.62	0.00	100.00	37.10	0.07	0.00	2.97
SOCI	0.09	1.47	35.53	0.39	6.02	0.22	0.20	4.35	0.92	2.88	5.35	0.60	3.96	5.36	0.05	0.12	1.07	5.74	0.53	1.09	0.20	0.34	7.00	100.00	0.08	0.01	1.31
VETE	0.00	34.92	0.00	5.76	0.34	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.00	2.03	8.14	0.00	0.00	15.93	0.00	0.34	1.69	0.00	0.34	2.03	100.00	0.00	0.34
DENT	0.00	0.41	0.00	2.88	0.00	0.00	0.41	0.00	0.00	0.41	0.00	0.00	0.82	0.41	0.82	1.65	0.00	32.10	0.00	0.00	0.82	0.00	0.00	0.41	0.00	100.00	0.41
HEAL	0.16	0.16	4.60	6.40	0.99	0.16	0.99	3.45	0.33	0.00	0.00	0.00	3.12	0.82	0.82	0.16	1.15	67.16	2.63	6.90	5.75	1.15	7.06	16.58	0.16	0.16	100.00

Fig. 8 Affinity between fields: Data from the 27×27 symmetric matrix in Fig. 6 are normalized so the diagonal elements become 100

Table 3 Level of affinity between fields by field of interest

Life sciences				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Agricultural and Biological Sciences	1. Veterinary	34.9	1. Environmental Science	20.9
	2. Environmental Science	23.8	2. Biochemistry, Genetics and Molecular Biology	15.2
	3. Biochemistry, Genetics and Molecular Biology	14.4	3. Medicine	8.09
Biochemistry, Genetics and Molecular Biology	1. Immunology and Microbiology	28.8	1. Medicine	43.1
	2. Pharmacology, Toxicology, and Pharmaceutics	23.3	2. Agricultural and Biological Sciences	14.4
	3. Neuroscience	18.9	3. Pharmacology, Toxicology, and Pharmaceutics	9.42
Immunology and Microbiology	1. Biochemistry, Genetics and Molecular Biology	8.40	1. Medicine	57.5
	2. Veterinary	8.14	2. Biochemistry, Genetics and Molecular Biology	28.8
	3. Pharmacology, Toxicology, and Pharmaceutics	6.57	3. Agricultural and Biological Sciences	11.4
Neuroscience	1. Psychology	10.1	1. Medicine	58.3
	2. Biochemistry, Genetics and Molecular Biology	4.79	2. Psychology	19.6
	3. Medicine	3.21	3. Biochemistry, Genetics and Molecular Biology	18.9
Pharmacology, Toxicology and Pharmaceutics	1. Biochemistry, Genetics and Molecular Biology	9.42	1. Medicine	43.8
	2. Immunology and Microbiology	9.10	2. Biochemistry, Genetics and Molecular Biology	23.3
	3. Chemistry	6.34	3. Environmental Sciences	7.58

Table 3 (continued)

Social Sciences

Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Arts and Humanities	1. Social Sciences	35.5	1. Social Sciences	58.1
	2. Psychology	12.8	2. Psychology	3.94
	3. Economics, Econometrics and Finance	8.16	3. Medicine	2.53
Business, Management and Accounting	1. Decision science	40.2	1. Social Sciences	25.0
	2. Economics, Econometrics and Finance	33.6	2. Economics, Econometrics and Finance	22.7
	3. Computer science	6.67	3. Engineering	14.4
Decision Sciences	1. Business, Management and Accounting	9.84	1. Environmental Science	37.6
	2. Mathematics	9.62	2. Engineering	12.5
	3. Computer science	6.42	3. Social Sciences	10.1
Economics, Econometrics and Finance	1. Business, Management and Accounting	22.7	1. Business, Management and Accounting	33.6
	2. Decision science	11.7	2. Social Sciences	32.9
	3. Social sciences	5.35	3. Arts and Humanities	8.16
Psychology	1. Neuroscience	19.6	1. Social Sciences	37.1
	2. Health Professions	7.00	2. Medicine	31.2
	3. Social Sciences	7.00	3. Arts and Humanities	12.8
Social Sciences	1. Arts and Humanities	58.1	1. Arts and Humanities	35.5
	2. Psychology	37.1	2. Psychology	7.00
	3. Economics, Econometrics and Finance	32.9	3. Business, Management and Accounting	6.02

Table 3 (continued)

Physical sciences				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Chemical Engineering	1. Chemistry	27.2	1. Chemistry	32.3
	2. Energy	15.6	2. Engineering	29.4
	3. Materials Science	14.4	3. Materials Science	25.4
Chemistry	1. Chemical Engineering	32.3	1. Materials Science	28.0
	2. Materials Science	18.9	2. Chemical Engineering	27.2
	3. Physics and Astronomy	16.1	3. Physics and Astronomy	18.9
Computer Sciences	1. Decision science	28.9	1. Engineering	32.7
	2. Mathematics	27.4	2. Mathematics	23.4
	3. Engineering	13.3	3. Social Sciences	16.4
Earth and Planetary Sciences	1. Environmental Sciences	33.6	1. Environmental Sciences	37.6
	2. Energy	14.4	2. Engineering	12.5
	3. Agricultural and Biological Sciences	7.16	3. Social Sciences	10.1
Energy	1. Chemical Engineering	9.78	1. Engineering	34.2
	2. Environmental Sciences	5.34	2. Environmental Sciences	20.2
	3. Earth and Planetary Sciences	4.23	3. Chemical Engineering	15.6
Engineering	1. Materials Science	47.9	1. Materials Science	17.4
	2. Physics and Astronomy	35.9	2. Computer Sciences	13.3
	3. Energy	34.2	3. Physics and Astronomy	10.4
Environmental Sciences	1. Earth and Planetary Sciences	37.6	1. Earth and Planetary Sciences	33.7
	2. Agricultural and Biological Sciences	20.9	2. Agricultural and Biological Sciences	23.8
	3. Energy	20.3	3. Social Sciences	16.8

Table 3 (continued)

Physical sciences				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Materials Science	1. Physics and Astronomy	34.0	1. Engineering	47.9
	2. Chemistry	28.0	2. Physics and Astronomy	27.0
	3. Chemical Engineering	25.4	3. Chemistry	18.9
Mathematics	1. Decision Science	37.1	1. Computer Sciences	27.4
	2. Computer Science	23.4	2. Engineering	18.7
	3. Physics and Astronomy	9.81	3. Decision Sciences	9.62
Physics and Astronomy	1. Materials Science	27.1	1. Engineering	35.9
	2. Chemistry	18.9	2. Materials Science	34.0
	3. Engineering	10.4	3. Chemistry	16.1
Health sciences				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Medicine	1. Health Professions	67.2	1. Biochemistry, Genetics and Molecular Biology	9.38
	2. Nursing	62.5	2. Pharmacology, Toxicology, and Pharmaceutics	3.85
	3. Neuroscience	58.3	3. Nursing	3.84
Nursing	1. Health Professions	6.90	1. Medicine	62.5
	2. Medicine	3.84	2. Social Sciences	10.1
	3. Psychology	2.48	3. Agricultural and Biological Sciences	6.14
Veterinary	1. Agricultural and Biological Sciences	3.69	1. Agricultural and Biological Sciences	34.9
	2. Immunology and Microbiology	2.80	2. Medicine	15.9
	3. Biochemistry, Genetics and Molecular Biology	0.578	3. Immunology and Microbiology	8.14

Table 3 (continued)

Health sciences				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
Dentistry	1. Medicine	5.77	1. Medicine	32.1
	2. Biochemistry, Genetics and Molecular Biology	0.238	2. Biochemistry, Genetics and Molecular Biology	2.88
	3. Immunology and Microbiology	0.233	3. Materials Science	1.65
Health Professions	1. Nursing	5.06	1. Medicine	67.2
	2. Medicine	3.03	2. Social Sciences	16.6
	3. Psychology	2.97	3. Nursing	6.90

*General field				
Field of interest	Top 3 fields with highest affinity to the field of interest	Level of affinity	Top 3 fields to which the field of interest has highest affinity	Level of affinity
General	1. Mathematics	0.229	1. Engineering	7.52
	2. Computer Sciences	0.196	2. Arts and Humanities	5.26
	2. Engineering	0.200	3. Social Sciences	5.26

Table 4 Mean and median affinity of other fields to the field of interest (using values from columns in Fig. 2)

Subject area	Field	Mean	Median
Life Sciences	Agricultural and Biological Sciences	8.95	2.47
	Biochemistry, Genetics and Molecular Biology	10.2	3.24
	Immunology and Microbiology	5.31	0.325
	Neuroscience	4.93	0.285
	Pharmacology, Toxicology and Pharmaceutics	5.71	0.638
Social Sciences	Arts and Humanities	7.27	1.77
	Business, Management and Accounting	8.36	1.50
	Decision Sciences	5.01	0.170
	Economics, Econometrics and Finance	5.81	0.290
	Psychology	5.90	0.748
	Social Sciences	14.3	5.50
Physical Sciences	Chemical Engineering	7.41	1.64
	Chemistry	7.66	1.41
	Computer Science	8.44	3.01
	Earth and Planetary Sciences	6.88	0.703
	Energy	5.28	0.598
	Engineering	14.8	5.24
	Environmental Science	8.98	3.03
	Materials Science	8.87	1.61
	Mathematics	7.79	1.28
	Physics and Astronomy	7.38	0.856
Health Sciences	Medicine	21.3	6.58
	Nursing	4.50	0.221
	Veterinary	4.02	0
	Dentistry	3.77	0
	Health Professions	4.64	0.484
-	General	3.77	0.0455
Mean		7.68	1.62

Table 5 Mean and median affinity of the field of interest to other fields (using values from rows in Fig. 2)

Subject area	Field	Mean	Median
Life Sciences	Agricultural and Biological Sciences	6.84	1.57
	Biochemistry, Genetics and Molecular Biology	8.04	1.87
	Immunology and Microbiology	8.43	0.467
	Neuroscience	8.50	1.07
	Pharmacology, Toxicology and Pharmaceutics	7.73	0.758
Social Sciences	Arts and Humanities	6.63	0.596
	Business, Management and Accounting	7.79	1.41
	Decision Sciences	9.71	0.442
	Economics, Econometrics and Finance	7.41	0.400
	Psychology	7.86	0.759
	Social Sciences	6.85	1.07
Physical Sciences	Chemical Engineering	9.91	3.48
	Chemistry	8.97	2.52
	Computer Science	8.16	1.86
	Earth and Planetary Sciences	7.04	1.27
	Energy	8.83	2.63
	Engineering	7.38	2.76
	Environmental Science	8.45	2.20
	Materials Science	8.95	2.25
	Mathematics	7.28	1.89
	Physics and Astronomy	8.62	1.45
Health Sciences	Medicine	5.32	0.599
	Nursing	7.34	0.241
	Veterinary	6.38	0
	Dentistry	5.24	0
	Health Professions	8.55	0.985
-	General	5.01	0.752
Mean		7.67	1.31

Acknowledgements This research was supported by the cooperation program of research institutes in Tohoku University, S-R2-515 of The Watanabe Memorial Foundation for the Advancement of Technology, and JSPS KAKENHI (Grant No. 20K02932).

References

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., Larivière, V., & Boyack, K. W. (2012). Design and update of a classification system: The UCSD map of science. *PLoS ONE, 7*(7), e39464.

Börner, K., & Scharnhorst, A. (2009). Visual conceptualizations and models of science. *Journal of Informetrics, 3*(3), 161–172.

Boyack, K. W., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Science and Technology, 65*(4), 670–685.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351–374.

Brink, M., Hengevelda, G. M., & Tobi, H. (2020). Interdisciplinary measurement: A systematic review of the case of sustainability. *Ecological Indicators, 112*, 106145.

Cabinet Office of Japan (2016), "The 5th Science and Technology Basic Plan", https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf (in Japanese).

Cabinet Office of Japan (2019), "Basic Policy on Economic and Fiscal Management and Reform 2019 ~A New Era of Reiwa: Challenges toward society 5.0~", https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/2019/2019_basicpolicies_en.pdf.

Cabinet Office of Japan (2020), "Integrated Innovation Strategy 2020", https://www8.cao.go.jp/cstp/togo2020_honbun.pdf (in Japanese).

Chena, S., Qiua, J., Arsenaultb, C., & Larivière, V. (2021). Exploring the interdisciplinarity patterns of highly citedpapers. *Journal of Informetrics, 15*, 101124.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology, 62*(7), 1382–1402.

National Institute of Science and Technology Policy in Ministry of Education, Culture, Sports, Science and Technology of Japan (2018), Science map 2016, NISTEP REPORT No.178, October 2018.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239.

Gowanlock, M., & Gazan, R. (2013). Assessing researcher interdisciplinarity: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics, 94*, 133–161.

Jang, W., Kwon, H., Park, Y., & Lee, H. (2018). Predicting the degree of interdisciplinarity in academic fields: The case of nanotechnology. *Scientometrics, 116*, 231–254.

Jost, L. (2006). Entropy and diversity. *Oikos, 113*(2), 363–375.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology, 88*(10), 2427–2439.

Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics, 68*(4), 925–928.

Kim, H., Park, H., & Song, M. (2022). Toward a consensus map of science. *Journal of Informetrics, 16*, 101255.

Klavans, R., & Boyack, K. W. (2006a). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology, 57*(2), 251–263.

Klavans, R., & Boyack, K. W. (2006b). Quantitative evaluation of large maps of science. *Scientometrics, 68*(3), 475–499.

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology, 60*(3), 455–476.

Klavans, R., & Boyack, K. W. (2011). Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology, 62*(1), 1–18.

Kong, L., & Wang, D. (2020). Comparison of citations and attention of cover and non-cover papers. *Journal of Informetrics, 14*, 101095.

Lariviere, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology, 61*(1), 126–131.

Leinster, T., & Cobbold, C. A. (2012). Measuring diversity: The importance of species similarity. *Ecology, 93*(3), 477–489.

Levy, R., Ghisletta, P., Goff, J. L., Spini, D., & Widmer, E. (2005). Incitations for interdisciplinarity in life course research. *Advances in Life Course Research, 10*, 361–391.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*, 87–100.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research, 246*, 1–19.
Mugabushaka, A. M., Kyriakou, A., & Papazoglou, T. (2016). Bibliometric indicators of interdisciplinarity: The potential of the Leinster-Cobbold diversity indices to study disciplinary diversity. *Scientometrics, 107*, 593–607.
Petterson, M. B., Longhurst, C., & Yu, J. J. (2021). Measuring interdisciplinarity of biomedical research, medical specialty performance, and implications for radiology: A retrospective review of 2.6 million citations. *Clinical Imaging, 80*, 322–328.
Pittman, J., Tiessen, H., & Montan, E. (2016). The evolution of interdisciplinarity over 20 years of global change research by the IAI. *Current Opinion in Environmental Sustainability, 19*, 87–93.
Porter, A. L., Cohen, A. S., Roessner, J. D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics, 72*(1), 117–147.
Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719.
Rao, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology, 21*(1), 24–43.
Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill Book Co.
Shannon, C. E. (1948), "A mathematical theory of communication", Bell System Technical Journal, 27, 379–423 and 623–356.
Silva, F. N., Rodrigues, F. A., & OliveiraCosta, O. N. L. F., Jr. (2013). Quantifying the inter disciplinarity of scientific journals. *Journal of Informetrics, 7*, 469–477.
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface, 4*(15), 707–719.
Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics, 165*, 14–26.
Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of web of science and Scopus. *Journal of Informetrics, 10*, 347–364.
Zhang, L., Rousseau, R., & Glanzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257–1265.