Asian Journal of Innovation and Policy (2016) 5.003:251-274
DOI: http//dx.doi.org/10.7545/ajip.2016.5.3.251

Quantitative Definitions of Collaborative Research Fields in Science and Engineering

Mathew Schwartz[*], **Kwisun Park**[**], **Sung-Jong Lee**[***]

Abstract Practical methodology for categorizing collaborative disciplines or research in a quantitative manner is presented by developing a Correlation Matrix of Major Disciplines (CMMD) using bibliometric data collected between 2009 and 2014. First, 21 major disciplines in science and engineering are defined based on journal publication frequency. Second, major disciplines using a comparing discipline correlation matrix is created and correlation score using CMMD is calculated based on an analyzer function that is given to the matrix elements. Third, a correlation between the major disciplines and 14 research fields using CMMD is calculated for validation. Collaborative researches are classified into three groups by partially accepting the definition of pluri-discipline from peer review manual, European Science Foundation, inner-discipline, inter-discipline and cross-discipline. Applying simple categorization criteria identifies three groups of collaborative research and also those results can be visualized. Overall, the proposed methodology supports the categorization for each research field.

Keywords Collaborative research, inner-disciplinary research, interdisciplinary research, cross-disciplinary research, quantification method, correlation matrix of major disciplines(CMMD), bibliometric data

I. Introduction

In recent decades, complementary collaboration between disciplines, as well as the re-purposing and reuse of methodologies or theoretical foundations between disciplines has become increasingly normal in academia (Aboelela et al., 2007; Bourke and Butler, 1998; Broto, Gislason and Ehlers, 2009; Cameron R., 2016; Huutoniemi et al., 2010; National Academies, 2004; Roco, 2008).

Submitted, Devember 13 , 2016; 1st Revised, December 27; Accepted, December 27

[*] Advanced Institutes of Convergence Technology, Seoul National University, Suwon, 443-270, Korea; New Jersey Institute of Technology, College of Architecture and Design, New Jersey, USA; umcadop@gmail.com

[**] National R&D Planning Team, National Research Foundation of Korea, Daejeon, 34113, Korea; kwisun_park@nrf.re.kr

[***] Corresponding author, Biotechnology Team; chris@nrf.re.kr

Although interdisciplinary collaboration has traditionally been associated with technological applications, it has proven a valuable practice in research, contributing to the development of several fields. As a result, interdisciplinary characteristics have emerged as a major issue in research and development (R&D) funding and knowledge production (Bruun et al., 2005; Han and Kyung 2011; Huutoniemi et al., 2010; Lee and Choi, 2010; Mansilla, 2005; Van Rijnsoever and Hessels, 2011; Sá, 2008).

The National Research Foundation of Korea (NRF), one of the largest science and engineering (S&E) funding agencies in the world, organized the Division of Interdisciplinary Research, under the supervision of the directorate for basic research in science and engineering, in early 2000. The division's funding is mandated to support the development and success of research requiring significant collaboration between S&E disciplines; while this mandate is specific to the NRF, funding agencies around the world continue to go through similar challenges. Additionally, the division is tasked with developing a convergence research support framework to promote creative and transformative research (Park et al., 2012/2013).

For many years, this division has categorized research applications as interdisciplinary based on applicant proposals' self-reported status as well as partial sorting by reviewers. This process has brought up a few important questions: are proposals interdisciplinary enough? How can interdisciplinarity be defined? Are there indicators for interdisciplinary fields that can lead to general agreement? If so, what are these key factors and how can they be analyzed?

In order to answer these questions and develop a better approach to collaborative research funding, a more meaningful definition of a 'discipline' is required. Both researchers and funding agencies have made attempts to create this distinction, with varying degrees of success (Beers and Bots, 2009; Brandt et al., 2013; Bruun et al., 2005; Fagerberg, Landström and Martin, 2012; Huutoniemi et al., 2010; Klein, 2006; Rafols et al., 2012; Rinia et al., 2001; Tijssen, 1992). In fact, attempts to define and classify individual fields have met with difficulty since the earliest division of classical disciplines; because there are so many topics and several of them overlap, it is difficult to definitively and rigidly categorize them (Mansilla, 2005; Repko, 2008). For example, mathematics is originally defined by the Oxford English Dictionary as:

> "Originally: (a collective term for) geometry, arithmetic, and certain physical sciences involving geometrical reasoning, such as astronomy and optics; spec. the disciplines of the quadrivium collectively. In later use: the science of space, number, quantity, and arrangement, whose methods involve logical reasoning and usually the use of symbolic notation, and

> which includes geometry, arithmetic, algebra, and analysis; mathematical operations or calculations" (Stevenson, 2010).

This definition, relying mostly on exemplum, shows how early distinctions were created through the combination of smaller fields. As science advances and new fields emerge, it is difficult to keep to such qualitative definitions. There must be a limit in place when accounting for these new fields. Instead, as shown in Repko's 2008 study, the definitions become social constructs that have evolved and changed through additional qualitative measures. Sugimoto provides an in-depth discussion on how disciplines are perceived through various conceptual frameworks, but fails to give a specific metric or quantitative definition (Sugimoto and Weingart, 2015).

While many mappings of the academic landscape have been created, they problematically rely on selecting disciplines subjectively (Klavans and Boyack, 2009) or utilizing existing map classifications (Rafols, Porter and Leydesdorff, 2010), making it difficult to continuously update the ever-changing landscape. As the work of Klavans and Boyack combines and finds common features among 20 existing maps, the resulting disciplines are abstracted twice. For a funding agency relying on transparent and consistent evaluations, an automated process for defining these main disciplines based on academic institutions and not subjective discourse is required. As such, rather than consolidating multiple expert opinions each year, department names of the academic institutions provide a more direct categorization method. While it has not always been easy to parse this data, the improvements to databases have allowed access to author addresses, including department names. Furthermore, this method allows for a democratized system, where if many academic institutions were to consider a topic such as graphene to be regarded a major discipline, hence creating a Department of Graphene, this would automatically be included in the yearly discipline categorization.

In this study, the researchers present novel and practical methodology for categorizing collaborative disciplines, such as inter-disciplines and cross-disciplines, in a quantitative manner by developing a Correlation Matrix of Major Disciplines (CMMD) using bibliometric data. Research shows that collaborative disciplines can be identified quantitatively by first defining the major disciplines and then measuring the correlation between the major disciplines and a compared discipline. Disciplines were categorized as major disciplines, inner-disciplines of a specific major discipline, inter-disciplines or cross-disciplines of the major disciplines in this research, or emerging disciplines in which the data has yet to stabilize.

The conceptual framework of the proposed methodology was introduced by the NRF in the 2013 Society of Research Administrators International annual meeting, in a session designed for sponsors and agencies (Park, Cha and Lee, 2013). In a previous study, the authors presented the possibility that major disciplines could be differentiated based on knowledge productivity, such as the number of publications in a specific discipline, while other disciplinary categories such as inner-discipline, inter-discipline, multidiscipline, and trans-discipline might be explained by their relationship with the predefined major disciplines.

In this paper, preceding work was refined and simplified to identify collaborative disciplines and classify disciplines based on types and levels of collaboration or academic dependency for practical implementation. The suggested approach could be an effective complimentary tool to manage the proposal review process in early stages and could simplify proposal classification and review panel organization for both funders and proposal applicants.

II. Background

Since the mid-to-late 1990s, the literature has recorded the frequent combination of knowledge between disciplines, including multi-, inter-, cross-, and trans-disciplinarity (Aboelela et al., 2007; European Science Foundation, 2011, National Academies, 2004; Huutoniemi et al., 2010; Nordmann, 2004). The prefixes multi, inter, cross and trans have been applied to the word discipline; however, it is difficult to find a quantitative definition that enables academics around the world to find shared meaning in these terms (Mansilla, 2005; Rafols et al., 2012). NRF defines interdisciplinary research as any research that breaks down disciplinary boundaries (Song and Seol, 1999; Seol and Song, 1999; Park et al., 2012/2013). NRF has used the national science and technology (S&T) standard classification system to measure interdisciplinarity of R&D proposals statistically by mapping multiple S&T classifications, but it still remains subjective manner by applicants (Song and Seol, 1999; Seol and Song, 1999). In order to rank the relevance of a proposal's topic to the category of funding in detail, subjective measures are used. For example, in some cases, reviewers are asked to do the ranking themselves. This can cause problems for both the grant applicants and the funding foundation since many categories are not clear for specific topics, especially in interdisciplinary fields (Park, Cha and Lee, 2013; Porter and Rossini, 1985).

Disciplines as a whole can be valuable references in identifying inter-disciplinarity (Broto, Gislason and Ehlers, 2009). A discipline refers "to a particular branch of learning or body of knowledge" (Repko, 2008) and can be categorized by its social organizations, based on factors such as mutual dependence and uncertainty (Broto, Gislason and Ehlers, 2009). Researchers have used existing data to create a journal-journal citation matrix on document sets in relation to each other, but have not arrived at a quantifying value (Leydesdorff, Rafols and Chen, 2013). As such, the publication trends of major disciplines were used for the CMMD in this study.

Many conceptual definitions of interdisciplinary research have been suggested by scholars (Aboelela et al., 2007; Apostel, 1972; European Science Foundation, 2011, National Academies, 2004; Huutoniemi et al., 2010; Klein, 1990/ 1996; Lattuca, 2001; Repko, 2008; Rosenfield, 1992; Stember, 1991). Wagner et al. (2011) gives an in-depth analysis of past analysis methods of interdisciplinary research. Among these definitions, the European Science Foundation (ESF) provides relatively comprehensive meanings for practical use and a guideline for peer review (European Science Foundation, 2011). ESF suggested pluri-disciplinary research as the contrary term of mono-discipline (European Science Foundation, 2011). Pluri-discipline was broken down into four categories: multidisciplinary, interdisciplinary, crossdisciplinary and transdisciplinary research. Each of them are then defined as follows:

> Multidisciplinarity is concerned with the study of a research topic within one discipline, with support from other disciplines, bringing together multiple dimensions, but always in the service of the driving discipline. Disciplinary elements retain their original identity. It fosters wider knowledge, information and methods.
>
> Interdisciplinarity is concerned with the study of a research topic within multiple disciplines, and with the transfer of methods from one discipline to another. The research topic integrates different disciplinary approaches and methods
>
> Crossdisciplinarity is concerned with the study of a research topic at the intersection of multiple disciplines, and with the commonalities among the disciplines involved
>
> Transdisciplinarity is concerned at once with what is between, across and beyond all the disciplines with the goal of understanding the present world under an imperative of unity of knowledge (European Science Foundation, 2011).

For the purpose of this research, we accepted these proposed definitions of inter-disciplinary and cross-disciplinary research and specified our own term, ‘inner-discipline,’ which is a discipline or field that belongs to a larger discipline but has yet to qualify as a major discipline.

Because the characteristics of inter-disciplinarity are not fully known, several scholars have attempted to identify underlying factors of such collaboration, although these are mostly qualitative studies (Broto, Gislason and Ehlers, 2009; Ho, Choi and Lee, 2013; Van Rijnsoever and Hessels, 2011). Some studies sought factors associated with disciplinary and interdisciplinary research collaboration by quantifying individual researchers’ characteristics, such as global innovativeness, work experience, dynamics of the scientific fields, and gender collected through surveys (Van Rijnsoever and Hessels, 2011). Others used indicators of inter-disciplinarity typology, such as the scope of inter-disciplinarity, type of inter-disciplinary interaction, and type of goals (Huutoniemi et al., 2010). These studies still remain largely qualitative and act as conceptual guidelines, while some are designed using empirical analysis for practical use for funders and policy-makers.

Past literature has shown a recent trend toward defining inter-disciplinarity in a quantitative way (Bourke and Butler 1998; Kaur et al., 2012; Pan et al., 2012; Schoolman et al., 2012; Tijssen, 1992; Xie et al., 2015; Yang and Heo 2014; Yang, Park and Heo, 2010). Bibliometrics is a useful tool for identifying publication trends, authors’ academic fields, and authors’ co-workers. Combined with citation and bibliometric analysis, which uses citation data, these methods provide an effective way to better examine the nature or characteristics of inter-disciplinary activity. Recent studies have utilized citation information and constructed a network analysis to quantitatively and visually measure inter-disciplinarity (Kaur et al., 2012; Schoolman et al., 2012; Small, 2009; Xie et al., 2015; Yang and Heo, 2014). Although this approach offers some value in individual measurement, it is difficult to apply in macro scale analysis, including the S&E field. Furthermore, previous studies initially assumed the inter-disciplinarity of certain research and used such assumptions as basis for later analyses (Kaur et al., 2012; Schoolman et al., 2012; Small, 2009; Xie et al., 2015).

Other quantitative assessments of inter-disciplinarity in science and technology have been conducted based on co-occurrence of publication in predefined classifications using bibliometric information (Börner et al., 2012; Bourke and Butler, 1998; Pan et al., 2012; Tijssen, 1992). This approach is quite practical and simple; however, it has limitations, such as the lack of generalization of co-occurrence, absence of quantitatively defined core disciplines that can be compared with other fields of study, disregard for diversity of publication size in different disciplines, and the restriction of testing to specific fields.

This paper proposes a method for quantifying the relevance of a research topic to the major disciplines within S&E, where a topic can be considered a type of research field or discipline that does not fall within the list of major disciplines. Instead of starting from a boundary condition that initially assumes a certain discipline or research topic that is inter-disciplinary, the major disciplines are identified by the NRF in a quantitative manner using bibliometric data and lists of R&D fields in S&E, and correlation over the major disciplines is calculated using the CMMD. Both the frequency of publications and co-occurrences of specific journal publication are used as a function representing collaboration of and relation to the major disciplines. Through correlation, these disciplines are then shown to be inner-disciplines, cross-disciplines or inter-disciplines. This classification is made by defining the major disciplines, creating a function representative of journal frequency, and analyzing the correlation of a compared discipline over the major disciplines. The correlation values describe the relation between a research topic and the major discipline. For example, if a recent research topic such as graphene shows high correlation with major disciplines such as physics, material science and electrical engineering, it can be considered crossdisciplinary, as it is strongly related to those fields. Conversely, algebra can be considered an inner-discipline due to its high correlation with mathematics only.

As a funding agency, the motivation for this work is not in performance evaluations of researchers, but instead in the interest of appropriating funds designated to emerging and collaborative research topics. The ability for a funding agency to make decisions based on repeatable and quantitative methods is vital to a transparent and fair system. While subjective measures can judge aspects not easily seen in data, they are also a problem when comparing several reviewers for the same funds. This work presents a quantitative method of classifying emerging and collaborative research topics that can supplement the subjective measures currently used. However, it is not meant to replace the entire process. Finally, while numerous researchers have developed maps of the academic landscape, the methods that follow a completely quantitative approach are difficult to implement and lack methodology within the literature, an important aspect of this work and the likelihood of implementation.

III. Methods

1. Major Discipline Definition

In order to correlate either major disciplines to each other, or different research topics to major disciplines, it is necessary to quantitatively define what a major discipline is. As the main funding agency for South Korea, the NRF has built a list over the last five years of fields in which it funds (NRF, 2016). The list has over one thousand fields broken into three categories; chief review board, review board, and sub review board. From this list, all entries from chief review boards and review boards, the more general categories, were used as the candidates for a major discipline. This list of fields is not determined by a subjective measure; rather, it is a culmination of all fields and topics utilized in funded applications. Additions to the list in the future are possible, and would easily be analyzed for inclusion of major disciplines, but within the immediate term no additions would seem likely to contribute as a major discipline based on the following criteria.

With a large initial set of keywords, the online database from the Web of Science (WOS) was used to determine the presence of each keyword in the academic community based on publications. The search filters available for the WOS were used to find articles published that were relevant to a specific field. Specifically, the option to search for author address was used. By using the address, the list of major discipline candidates were compared against the word the Universities use to define their department, i.e. Department of Chemistry. The number of queries returned for each candidate discipline for the entire period that had been stored in WOS was recorded. In order to define a limitation to the number of major disciplines while including a significant number, a value of 300,000 articles was decided as the cut-off point for a major discipline. This number provided 22 major disciplines and did not leave any obvious major disciplines out of the group, which was reduced further to 21 major disciplines by the criteria in the following subsection due to duplications in the dataset.

2. Data Collection

All data used in this research was accessed either through the Web Of Science online or the API (Web of Knowledge Web Services v. 3.0) using the python programming language. When using data from the WOS database, it is important to note the methods in which the data is stored and searched. For most disciplines, the full name of the discipline is not stored in the WOS database; instead, a concatenated version is kept in place. For example, Chemistry is stored as Chem. This introduces errors when a word is concatenated to a different less common word. However, the full string Chemistry is still allowed to be used in the search, only the user must be aware of the automatic concatenation done by the database as other terms, such as Chemical Engineering must be searched as Chem* Eng*. For this research, the search term for address is used as the WOS search tag AD, and when searching for keywords as the topic, the search tag TS is used.

2.1 Major Disciplines

After defining the major disciplines total search results, all data from the years 2009 through 2014 were downloaded for each discipline with full inclusion. Using the most recent six years of publication data is both more representational of the current academic landscape and a reasonable limit to the amount of data collected. In order to download the most accurate image of the database, a python script was used to both query the search terms and store the full raw data through the University account. Although the WOS API has a download limit of 100,000 results, the search for each discipline is split into yearly data. On a few disciplines, the yearly data exceeds 100,000, yet is under 200,000. In this case, the data is sorted in reverse order, downloaded, and parsed for duplicates in combination with the first set of 100,000, allowing any search term under 200,000 to be downloaded.

For the search of major disciplines, a single keyword was used in the address search field. In this case, for a given search term, i.e. Chemistry, the results Department of Chemistry and Department of Chemistry and Chemical Engineering are both valid results. Additionally, multiple authors may have different addresses, resulting in a publication counting for two different search terms. The presence of multiple authors with different departments, or departments with multiple keywords, creates duplicate publications in each field. In order to check the overlaps, the total duplicates in one field to another were found. In the case of Chemical Engineering, the resulting data was included at a rate of 99.36% in Chemistry. This is caused by both the WOS concatenating search terms and the frequency of which departments are named as Chemistry and Chemical Engineering. The second most duplicated data was Biology containing 35.6% of publications from Biochemistry. Finally, the removal of Chemical Engineering due to the lack of unique data reduced the total major disciplines to 21.

After deciding on the final 21 disciplines, all duplicate data was removed from the database to give a correlation based on a basic frequency of publications in a specific journal. This removes duplicate publications through database errors as well as collaborative papers. The motivation behind this is two folds. First and foremost, the data seems to agree with conventional ideas of disciplines when the duplicates are removed. Second, the use of duplicate data creates a strong weight on co-authored papers, which could be analyzed separately for their own metrics.

Table 1 lists the final major disciplines and the total number of papers collected over each year. The data collection period occurred between May

13th and May 15th 2015. Each discipline displayed here had an overall publication number over 300,000 when searching through all years. Table 1 represents the full data records collected for each of the past six years.

Table 1 Yearly data downloaded for each major discipline

Discipline	Code	2009	2010	2011	2012	2013	2014
Agriculture	A	35183	37350	42936	45599	48809	51872
Anatomy	B	12788	13120	13406	13758	13552	13235
Biochemistry	C	38592	39922	43170	44564	45884	45025
Biology	D	116802	123181	132868	141010	144436	146917
Chemistry	E	160299	161809	176172	182105	194227	199700
Computer Science	F	18197	18911	20605	22690	24553	26535
Dentistry	G	17667	19061	20067	21514	22019	21155
Electrical Eng.	H	20930	22265	24644	26715	28681	31113
Food	I	19263	19930	23005	24519	27853	29417
Internal Medicine	J	23796	25619	27170	29434	30417	29704
Material Science	K	29647	30925	34596	36678	40572	44363
Mathematics	L	48438	49655	53165	57853	61249	62425
Mechanical Eng.	M	18959	20039	21810	23362	26210	28475
Nutrition	N	17166	18086	20184	20872	24423	22863
Obstetrics	O	14207	15223	16513	17510	17792	18154
Pathology	P	45851	33400	50618	53791	53822	54168
Pharmacology	Q	26457	26710	28298	29055	30674	29209
Physics	R	110782	113486	121529	124636	132182	137450
Physiology	S	25850	26262	27037	28107	29111	28690
Psychiatry	T	23512	24503	26240	27514	28576	29147
Surgery	U	61888	67653	72212	77397	83092	85911

2.2 Minor Disciplines

After creation of CMMD, any research topic can be used as a keyword to find the collaboration category. A large challenge to quantifiable categories is the lack of prior work using quantifiable measures. In order to compare the results of this research to outside opinion, multiple keywords from various known sources were collected. In this study, these keywords refer to minor disciplines. Table 2 shows the selected 14 minor disciplines, the source and defined category from the source, as well as the total number of papers for each minor discipline of the six years. The search terms used for research topics are in the supporting information.

Table 2 Source, category and number of publications downloaded per year

Keyword	Code	Source	Category	2009	2010	2011	2012	2013	2014
3D Printing	AA	Gatner, 2013	Collaborative or Emerging	143	123	168	234	379	762
Algebra	BB	NRF, 2016	Inner of Mathematics	3497	3513	3665	3764	3915	3885
Artificial Intelligence	CC	ESF, 2011	Transdisciplinary	589	595	669	754	807	826
Atomic Physics	DD	NRF, 2016	Inner of Physics	849	1159	1541	1586	927	634
Climate Change	EE	European Environment Agency, 2014	Collaborative or Emerging	8150	9676	11585	13281	15550	16582
Cognition	FF	ESF, 2011	Transdisciplinary	4974	5296	5996	6814	7738	8156
Combinatorics	GG	NRF, 2016	Inner of Mathematics	261	269	291	281	285	290
Differential Equations	HH	NRF, 2016	Inner of Mathematics	6678	6652	6963	7441	8085	8116
Graphene	II	Sanchez *et al.*, 2012	Collaborative or Emerging	2134	3506	5683	8477	11892	16554
Internet of Things	JJ	Gartner, 2011	Collaborative or Emerging	41	61	144	202	412	623
Molecular Machines	KK	Schlick, 2010	Collaborative or Emerging	310	309	287	324	322	357
Quantum Mechanics	LL	NRF, 2016	Inner of Physics	1789	1826	1825	1913	1988	2009
Robotics	MM	ESF, 2011	Interdisciplinary	905	972	1024	1069	1230	1200
Synthetic Biology	NN	ESF, 2011	Transdisciplinary	424	515	641	851	912	1040

3. Analyzer Score

The correlation of two disciplines is defined by the common journals in which each discipline publishes. By parsing the data collected through publications, the total publications within a specific journal are counted, referred to here as journal frequency. The journal frequency is used as a function to correlate the importance of a specific journal in one field to another.

With new data sets, the analyzer function can be derived by the same method each time, without the need for future model fitting. The analyzer function is then defined by the average of the most frequent journal frequency of all major disciplines, which can be written as Equation 1.

$$a(x)=\frac{\sum_{i=1}^{n} x(P_i)}{n} \tag{1}$$

Where n is the total number of major disciplines, P is the set of disciplines with elements P_i, where P_i is the set containing the journal frequency of different journals within a discipline and $x(P_i)$ is the journal frequency of the *xth* most frequent journal in a discipline.

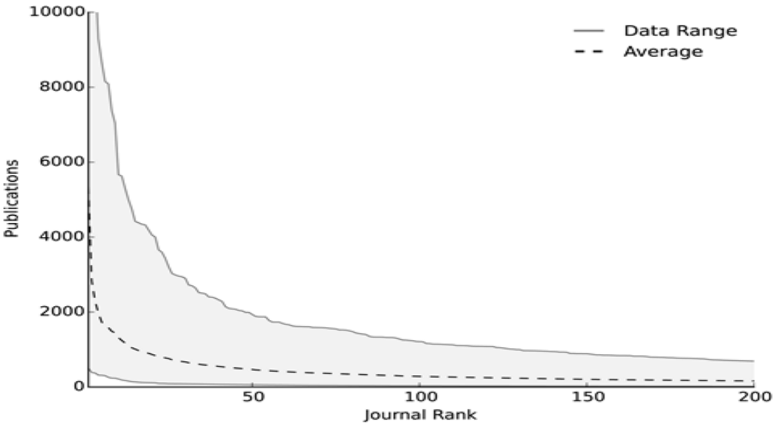

Figure 1 *xth* most frequent journal frequency (journal rank)

The minimum and maximum data is displayed in grey. This graph shows the core part of the graph with extents of 10,000 and 200 respectively. The middle dotted line is the calculated average using Equation 1.

4. Correlation Matrix

The correlation of major discipline A to major discipline B, is referred to as $C_{a,b}$, where $C_{a,b}$ is a row-column pair of the $n\times n$ matrix C. To calculate a discipline pair A and B, a normalized analyzer score is defined as k in Equation 2.

$$k=\sum_{i=1}^{n} a(i)^2 \tag{2}$$

Where i is the argument for the analyzer function in equation 1. The value for an individual pair of journals in each discipline given ordered sets $A=\{x_1,x_2,x_3,\ldots,x_j : x_i \prec x_j \Leftrightarrow i<j\}$ and $B=\{x_1,x_2,x_3,\ldots,x_j : x_i \prec x_j \Leftrightarrow i<j\}$ where elements $x_i \ldots x_j$ are journal names ordered by journal frequency within discipline A and B. The resulting values from the analyzer function $a(i)$ are multiplied. Defined in Equation 3.

$$h(x)=(a(i):x_i \in A)\times(a(i):x_i \in B) \tag{3}$$

The final correlation value for $C_{a,b}$ is given through Equation 4, where p is the title of a journal in both discipline A and B.

$$r_{a,b}=\sum_{p\in A\cap B} h(p) \tag{4}$$

$$C_{a,b}=\frac{r_{a,b}}{k}*100$$

For the minor discipline to major discipline correlation, an additional factor is applied to reduce the influence of a particular field on the correlation value. For a given major discipline a to minor discipline m, referred to as $\mathrm{C}_{\mathrm{a,m}}$, a normalizer value t acts as the divisor for the final normalized score $\mathrm{N}_{\mathrm{m,a}}$ defined in Equation 5.

$$t=\sum_{i=0}^{n} C_{a,i} \tag{5}$$

$$N_{m,a}=\frac{C_{m,a}}{t}$$

Where $C_{a,i}$ is the correlation score defined in Equation 4, of discipline A to n total major disciplines.

5. Categorization

Due to limited data of recently emerging fields, categorization through bibliometric data may be unreliable. In this case, the additional tag of *Emerging* is given to the research topic if the number of publications has doubled in the past 6 years, which can be found from Table 2.

A function $max_i(x)$ is defined as the maximum value of the set X, where i is the ith maximum value such that $max_1(x) > max_2(x)$. Finally, the categorization for a minor discipline set of normalized correlations to the major disciplines, N_m, is defined as:

$$f(N_m) = \begin{cases} inter & \text{if } x \leq 5 \ \forall\ x \in N_m \\ inner & \text{if } max_1(N_m) - max_2(N_m) > 4 \ \wedge max_i(N_m) - max_{i+1}(N_m) \ \leq 2\ \forall\ i > 1, i < |N_m| \\ cross & \text{default} \end{cases} \tag{6}$$

These values are defined based on the definitions given in the Introduction section for *inter-discipline* and *cross-discipline,* for *inter-discipline* being a research topic, which is not central to any one specific field, and hence will have a lower overall correlation, which in this research was chosen as a value of 5. In contrast, *cross-discipline* is defined to be the default case such that a research topic is not inter-discipline and is also not considered inner-discipline, requiring a high correlation (above 5) to more than one major discipline. Further research into more research topics in the future would help in understanding if these values can be kept as constant or must be determined each year. While this is one drawback of the proposed method, it is acceptable for the intended purpose of correlating research topics to the major disciplines and being able to compare these topics against each other.

IV. Results and Discussion

1. Major Discipline

This research took a bottom up approach to defining collaborative fields by using quantitative measures for each step of the process. The first step was in defining what is considered a major discipline. After finding the major disciplines, a correlation matrix between the major disciplines is created (Table 3).

Table 3 Correlation matrix of the major disciplines rounded to the nearest integer

Code	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
A	*	33	38	43	10	10	11	2	33	30	3	3	2	20	7	20	18	3	33	13	11
B	33	*	44	47	3	10	19	2	10	37	1	3	1	25	13	37	35	4	58	21	19
C	38	44	*	71	21	11	20	2	21	47	6	3	2	29	13	35	43	5	58	15	16
D	43	47	71	*	10	12	17	2	16	39	3	4	2	25	10	32	33	5	54	16	14
E	10	3	21	10	*	2	3	5	18	3	36	1	10	4	1	2	8	15	5	1	1
F	10	10	11	12	2	*	3	21	3	9	2	11	6	4	2	5	6	4	11	5	4
G	11	19	20	17	3	3	*	1	5	24	3	1	1	13	9	16	15	2	20	8	11
H	2	2	2	2	5	21	1	*	1	1	22	6	17	1	0	1	1	23	2	1	1
I	33	10	21	16	18	3	5	1	*	8	5	1	2	27	3	7	11	2	11	3	3
J	30	37	47	39	3	9	24	1	8	*	1	3	1	30	15	41	32	3	47	17	25
K	3	1	6	3	36	2	3	22	5	1	*	1	29	1	0	1	2	38	2	0	1
L	3	3	3	4	1	11	1	6	1	3	1	*	5	1	1	2	2	8	3	1	1
M	2	1	2	2	10	6	1	17	2	1	29	5	*	1	0	1	1	14	2	0	1
N	20	25	29	25	4	4	13	1	27	30	1	1	1	*	12	16	24	2	35	9	10
O	7	13	13	10	1	2	9	0	3	15	0	1	0	12	*	14	10	1	13	6	10
P	20	37	35	32	2	5	16	1	7	41	1	2	1	16	14	*	21	2	26	9	18
Q	18	35	43	33	8	6	15	1	11	32	2	2	1	24	10	21	*	3	52	18	11
R	3	4	5	5	15	4	2	23	2	3	38	8	14	2	1	2	3	*	5	2	1
S	33	58	58	54	5	11	20	2	11	47	2	3	2	35	13	26	52	5	*	23	17
T	13	21	15	16	1	5	8	1	3	17	0	1	0	9	6	9	18	2	23	*	6
U	11	19	16	14	1	4	11	1	3	25	1	1	1	10	10	18	11	1	17	6	*

It is useful to see the correlation of these fields within S&E for funding decisions as the major disciplines of researchers collaborating can be quantified in terms of closeness through past journal publications. Easily visible through Figure 2 is the tendency for Engineering and Medical fields to be more correlated to their own group, a sign that the methodology produces intuitively agreeable results.

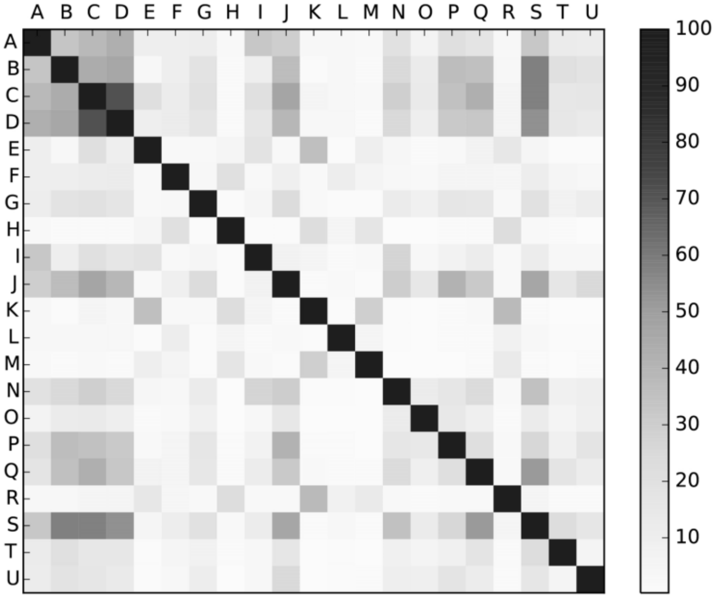

Figure 2 Correlation of the major disciplines

The color map shows the correlation values from 0 to 100 of the major Disciplines. This map makes it easy to understand which fields have high correlation, which is most common in similar type fields.

2. Minor Discipline

The final result of this research is the categorization of any given keyword or research topic in relation to its collaboration within major disciplines. Fourteen minor disciplines have been analyzed, with the correlation matrix shown in Table 4. The resulting correlation between the referenced minor disciplines and the suggested major disciplines using the CMMD is shown in Table 5.

Table 4 Minor disciplines to the major disciplines rounded to the nearest tenth

Code	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
AA	.6	.6	.6	.6	2.3	1.1	.8	3.5	.6	.5	2.5	.6	4.1	.4	.4	.4	.4	1.5	.5	.5	.9
BB	.1	.0	.0	.1	.1	1.4	.0	.8	.0	.0	.2	17.6	.5	.0	.0	.0	.0	2.7	.0	.1	.0
CC	.4	.3	.2	.3	.3	8.0	.2	3.1	.3	.3	.3	2.9	3.5	.2	.2	.2	.2	.4	.2	.3	.2
DD	.1	.1	.2	.3	1.4	.5	.1	4.7	.1	.1	4.1	2.1	3.5	.0	.1	.1	.1	8.9	.1	.1	.1
EE	2.2	1.5	1.3	2.1	.6	1.5	1.0	.5	.9	1.4	.3	1.2	.7	.8	.8	1.1	1.0	1.0	1.3	1.5	1.2
FF	1.7	2.0	1.4	1.7	.4	1.6	1.3	.3	.7	1.6	.1	1.0	.2	1.1	1.0	1.2	1.5	.5	1.9	6.1	1.4
GG	.0	.0	.1	.1	.2	3.8	.0	.5	.1	.0	.1	12.3	.2	.0	.0	.0	.0	.9	.0	.0	.0
HH	.2	.1	.1	.1	.3	1.4	.1	1.7	.1	.1	.4	26.3	3.1	.1	.1	.1	.1	1.9	.1	.1	.1
II	.2	.1	.2	.2	5.2	.3	.2	4.1	.6	.0	6.7	.9	3.7	.1	.1	.1	.1	8.7	.1	.1	.1
JJ	.1	.1	.1	.1	.2	4.3	.1	2.7	.2	.1	.2	1.2	.8	.1	.1	.1	.1	.3	.1	.1	.1
KK	.7	.6	1.3	1.2	6.6	1.3	.5	.6	.7	.5	1.9	1.0	.8	.4	.4	.6	.7	1.8	.7	.6	.5
LL	.2	.1	.3	.3	2.4	.9	.1	1.4	.3	.1	1.7	5.8	1.1	.1	.1	.1	.2	7.8	.2	.1	.1
MM	.2	.3	.2	.2	.2	2.2	.2	1.8	.1	.2	.3	.7	2.3	.1	1.0	.3	.2	.4	.2	.3	1.1
NN	1.2	1.0	1.6	1.8	2.9	1.5	.8	.4	1.0	.9	.9	1.0	.5	.6	.6	.9	.9	.9	1.0	1.0	.8

Table 5 Categorization of the minor disciplines using CMMD

Research Topic	Code	Category	Emerging	Agreement with Reference
3D print	AA	Inter	Yes	Yes
Algebra	BB	Inner		Yes
Artificial Intelligence	CC	Cross		Yes
Atomic Physics	DD	Inner		Yes
Climate Change	EE	Inter	Yes	Yes
Cognition	FF	Inner		No
Combinatorics	GG	Cross		No
Differential Equations	HH	Inner		Yes
Graphene	II	Cross	Yes	Yes
Internet of Things	JJ	Inter	Yes	Yes
Molecular Machines	KK	Inner		No
Quantum Mechanics	LL	Cross		No
Robotics	MM	Inter		Yes
Synthetic Biology	NN	Inter	Yes	Yes

When comparing against the categorization of the NRF disciplines, Algebra, and Differential Equations agree as inner-disciplines of Mathematics. However, the NRF has categorized Quantum Mechanics as an inner discipline of Physics, while this research has found it to have high correlations with both

Mathematics and Physics, easily seen in Figure 3. Suggesting there is much more input from the mathematics discipline than previously thought. This color map shows the high correlation to Mathematics and Physics, suggesting a Cross-discipline rather than an Inner-discipline of Physics.

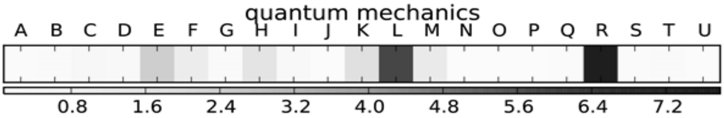

Figure 3 Correlation color map of quantum mechanics

Atomic Physics was categorized as an inner discipline of Physics by the NRF database, but it seems to have strong correlation with Mathematics, Mechanical Engineering, and Physics as shown in Figure 4. However, it falls into the inner-discipline of Physics based on the suggested categorization criteria, equation (6).

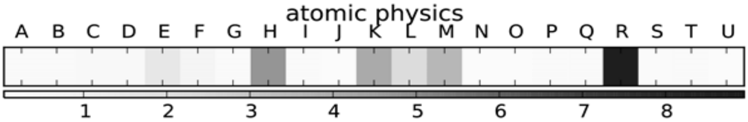

Figure 4 Correlation color map of atomic physics

As the ESF manual uses *Transdisciplinary*, not being in the categories proposed here, it is considered either Inter or Cross disciplinary to validate the results (i.e. not Inner-disciplinary). In this view, most of the minor disciplines were categorized similarly to the referenced categorization of minor disciplines with a few exceptions such as *Molecular Machines*. The results in Figure 5 suggest it is still strongly correlated with *Chemistry*, leading it to be categorized as an Inner discipline. This color map shows the high correlation to Chemistry, with very little correlation to the other major disciplines, suggesting an Inner-discipline rather than a Collaborative discipline.

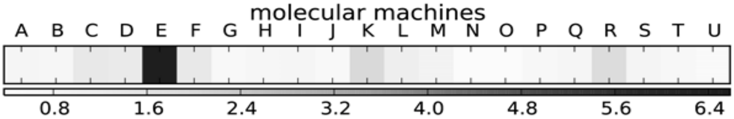

Figure 5 Correlation color map of molecular machines

Finally, a good representation of how this research represents the definitions proposed in the ESF manual in a quantitative way is the research field Robotics. As the ESF defines Robotics as an inter-disciplinary field due to its application in a variety of disciplines, with no one major discipline taking ownership, an even correlation among many disciplines is to be expected, as seen in Figure 6. This color map shows the relatively similar correlation among many disciplines, suggesting a research topic that is neither cross nor inner disciplinary.

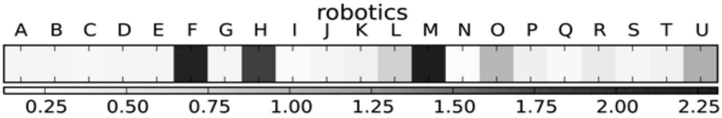

Figure 6 Correlation color map of robotics

The highest correlation value in Robotics can be seen as 2.3 with Mechanical Engineering, an unsurprising result. Additionally, Electrical Engineering and Obstetrics both have correlation values of approximately 1.8 and 1.0 respectively. While the former is again unsurprising, the latter is more unique, most likely caused by the large use of robots and robotic technology in the discipline.

10 of out 14 minor disciplines agreed with the referenced categorization; 3D print, Algebra, Artificial Intelligence, Atomic Physics, Climate Change, Differential Equations, Graphene, Internet of Things, Robotics, Synthetic Biology. Cognition and Molecular Machines are in disagreement with the references, as the proposed methodology suggests they are inner-disciplines of Psychiatry and Chemistry, respectively. Additionally, while the reference material refers to Molecular Machines as both collaborative and emerging, the growth rate of bibliometric data does not suggest this. However, it is possible the low quantity of data has skewed the results. Quantum Mechanics and Combinatorics are considered inner-disciplines by the reference material, but are shown to have high correlations to more than one discipline. Quantum Mechanics is highly correlated to both Mathematics and Physics, while Combinatorics is highly correlated to both Mathematics and Computer Science.

Based on the criteria stated in Table 2, the research topics 3D Print, Climate Change, Graphene, Internet of Things, and Synthetic Biology are emerging topics. As these minor disciplines were identified by Gartner (Gartner, 2011/2013) as emerging topics, the results are in agreement.

V. Discussion

The purpose of this research is to introduce a quantitative method of categorizing disciplines and research topics in terms of their correlation to one another. The values produced from this correlation can be used to quantitatively define frequently used terms such as Inter-disciplinary. While this paper uses a variety of sources to ensure the accuracy of results, it is important to note that these references act as only a guide and cannot be used to truly test the validity of the results, as the current methods used by organizations such as the NRF and ESF are still qualitative.

An initial search for publication frequency on a larger dataset of keywords in S&E would be beneficial. These keywords, though, had not been predefined as major or interdisciplinary, and were solely used as a source list for possible search terms without access to the raw database and numbers of keywords.

Further improvement of the work can be done in data collection. Yearly data could be expanded to many more years, which would assist in the development of the analyzer function. Issues with the WOS database do introduce errors, although the error is believed to be negligible, especially as more yearly data is collected. Additional sources of data for publication records would also improve the accuracy of the correlations.

This work is meant to be a starting point for the quantification of disciplines, a necessary step to improve funding allocations and allow academics to better understand new fields and collaborations. Additionally, the work can be applied as an indexing tool for academic institutions focusing on interdisciplinary or convergence research. In planning and development, institutes looking for new collaborations between seemingly unrelated disciplines can use this method to detect lowly correlated disciplines and focus on the possible convergence between them. As with any new method of categorization, there will be disagreements among academics on the way a discipline is categorized, in both the major disciplines and the categorization of minor disciplines. The authors acknowledge this issue and encourage further development of this method as a means for more accurately describing a discipline and more precisely (quantifiably) defining the use of words such as inner-,multi-,cross-,inter-, and trans-disciplinary.

References

Aboelela, S.W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S.A., and Haas, J. (2007) Defining interdisciplinary research: conclusions from a critical review of the literature, Health Services Research, 42(1), 329-346.

Apostel, L. (1972) Interdisciplinarity Problems of Teaching and Research in Universities, Organisation for Economic Cooperation and Development.

Beers, P.J. and Bots, P.W.G. (2009) Eliciting conceptual models to support interdisciplinary research, Journal of Information Science, Sage Publications.

Börner, K., Klavans, R., Patek, M., Zoss, A.M., Biberstine, J.R., Light, R.P. and Larivière, V. (2012) Design and update of a classification system: the UCSD map of science, PloS One, 7(7).

Bourke, P. and Butler, L. (1998) Institutions and the map of science: matching university departments and fields of research, Research Policy, 26(6), 711-718.

Brandt, P., Ernst, A., Gralla, F., Luederitz, C., Lang, D.J., Newig, J., and Reinert, F. (2013) A review of transdisciplinary research in sustainability science, Ecological Economics, 92, 1-15.

Broto, V.C., Gislason, M. and Ehlers, M.H. (2009) Practising interdisciplinarity in the interplay between disciplines: experiences of established researchers, Environmental Science & Policy, 12(7), 922-33.

Bruun, H., Hukkinen, J., Huutoniemi, K. and Klein, J.T. (2005) Promoting interdisciplinary research: the case of the Academy of Finland, Academy of Finland.

Cameron, R. (2016) An interdisciplinary approach to industry-based complex problem-solving: sustainable policy solutions to the Malaysian water crisis, Asian Journal of Innovation and Policy, 5(1), 55-77.

European Environment Agency (2014) Assessment of global megatrends-global megatrend 9: increasingly severe consequences of climate change, European Environment Agency.

European Science Foundation (2011) European Peer Review Guide - Integrating Policies and Practices into Coherent Procedures, European Science Foundation.

Fagerberg, J., Landström, H. and Martin, B.R. (2012) Exploring the emerging knowledge base of 'the knowledge society', Research Policy, 41(7), 1121-1131.

Gartner (2011) Gartner Identifies the Top 10 Strategic Technologies for 2011, http://www.gartner.com/newsroom/id/1826214.

Gartner (2013) Gartner Identifies the Top 10 Strategic Technologies for 2014, http://www.gartner.com/newsroom/id/2603623.

Han, S.H. and Kyung, J.U. (2011), Empirical analysis for promotion of integarted research: focused on the difference of recognition between all disciplines, Journal of the Korea Association for Policy Studies, 20(1), 151-179.

Ho, D.N., Choi, K.Y. and Lee, S.J. (2013) Bibliometric analysis of theranostics: two years in the making, Theranostics, 3(7), 527-531.

Huutoniemi, K., Klein, J.T., Bruun, H. and Hukkinen, J. (2010) Analyzing interdisciplinarity: typology and indicators, Research Policy, 39(1), 79-88.

Kaur, J., Hoang, D.T., Sun, X., Possamai, L., JafariAsbagh, M., Patil, S. and Menczer, F. (2012) Scholarometer: a social framework for analyzing impact across disciplines, PloS One, 7(9).

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

Klein, J.T. (1990) Interdisciplinarity: History, Theory, and Practice, Wayne State University Press.

Klein, J.T. (1996) Crossing Boundaries: Knowledge, Disciplinarities, and Interdisciplinarities, University of Virginia Press.

Klein, J.T. (2006) Afterword: the emergent literature on interdisciplinary and transdisciplinary research evaluation, Research Evaluation, 15(1), 75-80.

Lattuca, L.R. (2001) Creating Interdisciplinarity: Interdisciplinary Research and Teaching Among College and University Faculty, Vanderbilt University Press.

Lee, J.M. and Choi, M.S. (2010) Social network analysis on interdisciplinary collaboration of convergence technologies specialists, Journal of the Korea Contents Association, 10(6), 415-428.

Leydesdorff, L., Rafols, I. and Chen, C. (2013) Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations, Journal of the American Society for Information Science and Technology, 64(12), 2573-2586.

Mansilla, V.B. (2005) Assessing student work at disciplinary crossroads, Change, Magazine of Higher Learning, 37(1), 14-21.

National Academies (2004) Facilitating Interdisciplinary Research, Washington, DC: The National Academies Press.

Nordmann, A. (2004) Converging Technologies-shaping the Future of European Societies, Interim Report of the Scenarios Group, High Level Expert Group 3, European Commission.

National Research Foundation of Korea (2016) NRF Review Boards, http://www.nrf.re.kr/nrf_tot_cms/show.jsp?show_no=183&check_no=178&c_relation=0&c_relation2=0.

Pan, R.K., Sinha, S., Kaski, K. and Saramäki, J. (2012) The Evolution of Interdisciplinarity in Physics Research, Scientific Reports 2, Nature Publishing Group.

Park, K.S., Cha, E.J. and Lee, S.J. (2013) Transformative Convergence Research Promotion Strategy and Review Processes in Basic Science and Engineering of National Research Foundation of Korea (NRF), 2013 SRA International Annual Meeting.

Park, K.S., Shin, S.K., Song, Y.D. and Cha, E.J. (2012) developing the underpinning strategies and NRF's convergence research-specific support system to promote creative and transformative research in S&E, Proceedings of the of the Korea Technlogy Innovation Society Conference 2012, 266-78.

Park, K.S., Shin, S.K., Song, Y.D., Lee, S.J., Lee, E.K. and Cha, E.J. (2013) Developing a convergence research support framework to promote creative and transformative basic research, Journal of the Korea Contents Association, 13(12), 221-234.

Porter, A.L. and Rossini, F.A. (1985) Peer review of interdisciplinary research proposals, Science, Technology and Human Values, 10(1), 33-38.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. and Stirling, A. (2012) How journal rankings can suppress interdisciplinary research: a comparison between innovation studies and business & management, Research Policy, 41(7), 1262-1282.

Rafols, I., Porter, A.L. and Leydesdorff, L. (2010) Science overlay maps: a new tool for research policy and library management, Journal of the American Society for Information Science and Technology, 61(9), 1871-1887.

Repko, A.F. (2008) Interdisciplinary Research, Process and Theory, Sage.

Van Rijnsoever, F.J. and Hessels, L.K. (2011) Factors associated with disciplinary and interdisciplinary research collaboration, Research Policy, 40(3) 463-472.

Rinia, E.J., Van Leeuwen, T.N., Van Vuren, H.G. and Van Raan, A.F.J. (2001) Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research, Research Policy, 30(3), 357-361.

Roco, M.C. (2008) Possibilities for global governance of converging technologies, Journal of Nanoparticle Research, 10(1), 11-29.

Rosenfield, P.L. (1992) The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences, Social Science & Medicine, 35(11), 1343-1357.

Sá, C.M. (2008) 'Interdisciplinary strategies' in US research universities, Higher Education, 55(5), 537-552.

Sanchez, V.C., Jachak, A., Hurt, R.H. and Kane, A.B. (2012) Biological interactions of graphene-family nanomaterials: an interdisciplinary review, Chemical Research in Toxicology, 25(1), 15-34.

Schlick, T. (2010) Molecular Modeling and Simulation: An Interdisciplinary Guide, Springer.

Schoolman, E.D., Guest, J.S., Bush, K.F. and Bell, A.R. (2012) How interdisciplinary is sustainability research? analyzing the structure of an emerging scientific field, Sustainability Science, 7(1), 67-80.

Small, H. (2009) Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy, Scientometrics, 83(3), 835-849.

Seol, S.S. and Song, C.H. (1999) The theoretical study on classification system of R&D activity, Journal of Korea Technology Innovation Society, 2(3), 19-33.

Song, C.H. and Seol, S.S. (1999) The study on establishing underpinning philosophy and structure on the national S&T standard classification system, Journal of Korea Technology Innovation Society, 2(3), 34-47.

Stember, M. (1991) Advancing the social sciences through the interdisciplinary enterprise, The Social Science Journal, 28(1), 1-14.

Stevenson, A. (2010) Oxford Dictionary of English, Oxford Dictionary of English.

Sugimoto, C.R. and Weingart, S. (2015) The kaleidoscope of disciplinarity, Journal of Documentation, 71(4), 775-94.

Tijssen, R.J.W. (1992) A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research, Research Policy, 21(1), 27-44.

Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J. and Rafols, I. (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature, Journal of Informetrics, 5(1), 14-26.

Xie, Z., Duan, X., Ouyang, Z. and Zhang, P. (2015) Quantitative analysis of the interdisciplinarity of applied mathematics, PloS One, 10(9).

Yang, C.H. and Heo, J. (2014) Network analysis to evaluate cross-disciplinary research collaborations: the human sensing research center, Korea, Science and Public Policy, 41(6), 734–49.

Yang, C., Park, H. and Heo, J. (2010) A network analysis of interdisciplinary research relationships: the Korean government's R&D grant program, Scientometrics, 83(1), 77-92.