Disciplinary knowledge production and diffusion in science

Erjia Yan[1]

College of Computing and Informatics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA. Phone: (215)895-1459. Fax: (215) 895-2494. Email: ey86@drexel.edu

Abstract

This study examines patterns of dynamic disciplinary knowledge production and diffusion. It uses a citation data set of Scopus indexed journals and proceedings. The journal-level citation data set is aggregated into 27 subject areas and these subjects are selected as the unit of analysis. A three-step approach is employed: the first step examines disciplines' citation characteristics through scientific trading dimensions; the second step analyzes citation flows between pairs of disciplines; and the third step uses ego-centric citation networks to assess individual disciplines' citation flow diversity through Shannon entropy. Results show that measured by scientific impact, the subjects of Chemical Engineering, Energy, and Environmental Science have the fastest growths. Furthermore, most subjects are carrying out more diversified knowledge trading practices by importing higher volumes of knowledge from a greater number of subjects. The study also finds that the growth rates of disciplinary citations align with the growth rates of global R&D expenditures, thus providing evidences to support the impact of R&D expenditures on knowledge production.

Introduction

Knowledge has powered economic growth and profoundly perpetuated the conditions of our existence (Knorr-Cetina, 1999). The knowledge societies that we live in are characterized by the proliferation of knowledge-intensive communities, specialized in knowledge production and reproduction, knowledge learning and exchange, and the use of information technologies (David & Foray, 2002). Knowledge societies are propelled by the investment of intangible capitals, typically in the form of education and research and development (R&D) expenditures (David, 2000).

R&D expenditures have a significant impact on economic growth and society's well-being (Lane, 2009). They are becoming an unassailable investment for governments worldwide, developed and developing alike (Grueber et al., 2011). United Nations Educational, Scientific and Cultural Organization (UNESCO) estimated that as of 2009, 1.77% of the world GDP or 1,277 billion PPP$ (purchasing power parity) are spent on R&D (UNESCO, 2011). With these vast investments, there is the need to assess the accountability of R&D expenditures, to justify "the national investment in terms of returns that the taxpayer can appreciate" (Holton, 1978, p. 200). The general public should be informed on the impact of these investments (Lane, 2009): in the short-term, what research centers are established and what papers are published; in the longer-term, how will the investments produce new knowledge, create new jobs, and build new economy.

While the inputs of the investments can be quantitatively assessed, evaluations of the outputs are less accessible, due to the fact that knowledge is fundamentally unobservable (Jaffe, Trajtenberg, &

[1]Corresponding author

Henderson, 1993). Thus, alternative instruments, such as surveys, ethnographic methods, and quantitative methods, have been employed to study the subject matters. Among these methods, the citation-based quantitative methods has gained popularity, because these methods can examine the “full externalities of science” that others were incapable of (Adams & Griliches, 1996, p.12664). Citations serve as a valuable instrument to study knowledge transfer in science and technology. In the citation representation, a paper, patent, journal, author, or institution is a research entity and a citation link denotes the transfer of knowledge from the cited entity to the citing entity (Stigler, 1994; Lockett & McWilliams, 2005; Yan, Ding, Cronin, & Leydesdorff, 2013).

While prior citation-based studies have revealed patterns of inter-organizational, interdisciplinary, and international knowledge transfer, there lacks a holistic and dynamic examination of disciplinary knowledge flows. Consequently, we have limited understandings of how disciplinary knowledge is used and diffused. Furthermore, an overview of the relationship between science investments and knowledge production at the discipline-level is largely inadequate from the literature. To fill these gaps, this study investigates patterns of the dynamic disciplinary knowledge production and diffusion through a citation data set that comprehensively indexes journals and proceedings in life sciences, social sciences, physical sciences, and health sciences. The following questions will be addressed:

- What are the characteristics of disciplines measured by scientific trading dimensions (i.e., trading impact, exported and imported ratio, disciplinary self-dependence, and trading dynamics);
- What disciplinary citation flows have the highest growths during the past five citation windows from 1997/2000 to 2009/2011;
- How diversified are disciplines’ citation practices measured by Shannon entropy and what are their dynamics; and
- What is the relationship between R&D expenditures and the volume of knowledge production operationalized by the number of citations?

The answers to these questions will provide large-scale, empirical evidences on knowledge production and transfer in science. Relating obtained findings with economy statistics on R&D expenditures, this study will also help inform the understanding the impact of science investment to knowledge production and innovation.

Literature review

Citations are typically employed in quantitative studies of knowledge transfer, predicated on the observation that citations imply knowledge flows from the cited documents to the citing ones. In this context, two types of document citations can be distinguished: patent citations and paper citations. Patent citations have been sought to examine factors that contribute to effective knowledge diffusion between different sectors, industries, and geographic locations. Studies that employed paper citations, on the other hand, are largely focused on addressing issues related to disciplinarity and interdisciplinarity. Related studies are reviewed in the following sections.

Patent citations and knowledge flows

Patent citations manifest knowledge flows—particularly for public research (Roach & Cohen, 2013) and serve as an expedient instrument to quantitatively study knowledge production and innovation (Yan, 2014). In a pioneering research on the relationship between geographical distances and patent citation intensity, Jaffe, Trajtenberg and Henderson (1993) found that domestic patents are more likely to cite other domestic patents. Likewise, patent citations were also regulated by country boundaries, organizational boundaries, and patent classes (Jaffe & Trajtenberg 1999). These factors can be broadly fit into the proximity framework. Proximity provides an accessible way to make inferences into innovation and diffusion; it is often seen as having multiple dimensions: for example, distinctions have been made between cognitive, organizational, social, institutional, and geographical factors (Boschma, 2005). By applying the proximity framework to patent citations, a series of observations have been made. For instance, studies have shown that knowledge diffusion is enhanced by physical and technological proximity (MacGarvie, 2005; Bacchiocchi & Montobbio, 2007) and is inhibited by geographical distances and organizational barriers (Breschi & Lissoni, 2009) and country boundaries (Belenzon & Schankerman, 2013; Li, 2014).

Recent advances on network theories and methods have prompted more "linked" perspectives on studies of patent citations. In network representations, proximities are typically modeled as node attributes with the goal to assess the association of these proximities to network topology. For instance, the geographic proximity was modeled in a patent citation network to examine the impact of U.S. granted, Chinese applicants owned patents to knowledge spillovers in China (Yu & Wu, 2014). Through a social network analysis, Cassi and Plunket (2014) identified a demonstrable relationship between the physical proximity and the likelihood of establishing technological collaborations. Nomaler and Verspagen (2008) employed a sector-to-sector matrix to identify inter-sectoral knowledge diffusion patterns. They found that the indicator of citations to science literature per patent has effectively captured inter-sectoral knowledge diffusion. In the same vein of research, a technological knowledge flow matrix was constructed to represent knowledge flows among technology classes and explore the inter-class coherence between technology classes and industrial sectors (Ko, Yoon, & Seo, 2014).

In addition to network-based approaches, statistical methods have gained popularity to model patent citations—largely attributed to their inference abilities. Representative statistical methods in this thread of research include the linear regression model (e.g., Singh & Marx, 2013), the probit model (e.g., Geroski, 2000; Fier & Pyka, 2014), Markov Chain (e.g., Parent & LeSage, 2012), and epidemic models (e.g., Hethcote, 2000; Vitanov & Ausloos, 2012). Through a regression model, Singh and Marx (2013) confirmed that country and state borders have independent effects on knowledge flows in addition to geographic proximity measured by distances. The probit model—a special case of the regression model that only takes two values—fits patent citations smoothly because these citations are binary in nature (i.e., the present or absent of citations are two-valued). Using a probit model, for instance, Geroski (2000) postulated that knowledge adoption is dependent with types of organizations; Fier and Pyka (2014) found cultural closeness has promoted patent citations between industries. Aside from the regression models, the dynamic changes can also be modeled by Markov Chain which posits that the next state is only dependent on the current state. This unique feature has posed opportunities to examine many real-world processes including patent citations. For instance, through a Bayesian spatial Markov Chain Monte Carlo model, Parent and LeSage (2012) identified several factors that can lead to patent production and citations, including human resources, research infrastructure, investments, science policies, and regional industry structure.

Paper citations and knowledge flows

Paper citations have been employed to study knowledge flows in science in addition to their well-known role in scientific evaluations. Because paper citations can be aggregated into several higher levels, studies have explored journal-, institution-, and field-level knowledge flows via journal citations, institution citations, and field citations. At the journal-level, it is found that journal knowledge flows in library and information science is frequent (Zhao & Wu, 2014) and a few library and information science journals heavily cited communication science journals (Borgman & Rice, 1992). Similarly, Leydesdorff and Probst (2009) revealed that communication science journals received citation flows from political science and social psychology journals.

At the institution-level, through a study of the spatio-temporal changes of 500 most cited research institutions, it is found that the intensity of institutional citations is dependent on the distance: the log of the citation counts has an inverse linear relationship with the log of the distance (Börner, Penumarthy, Meiss, & Ke, 2006). This finding was verified and extended by an analysis of institutional citations in library and information science in that the number of citations between institutions is becoming less dependent on country boundaries and physical distances (Yan & Sugimoto, 2011). Besides geographic distances, structural holes and degree centrality of researchers were also associated with knowledge diffusion at the institution level (Liu et al., 2014).

At the field-level, efforts have been made to describe the global-level knowledge flows (e.g., Van Leeuwen & Tijssen, 2000; Rinia, Van Leeuwen, & Bruins, 2001; Naumis & Phillips, 2012; Yan et al., 2013; Zitt & Cointet, 2014). Zitt and Cointet's (2014) found a steady drop in variances of normalized impact and relative growth in science using a Web of Science data set from 1999 to 2008. Studies have also found that publications in one discipline tended to cite papers in adjacent disciplines (Van Leeuwen & Tijssen, 2000) and citations to publications of the own discipline occurred sooner than citations to papers in other disciplines (Rinia, Van Leeuwen, & Bruins, 2001). Diachronically, it is evident that a global-level epistemological change took place around 1960 which gradually reshaped the structure of disciplinary knowledge flows (Naumis & Phillips, 2012).

In the meantime, there is also renewed interest in understanding the between-field knowledge flows, stemming from prior qualitative research on disciplinarities (Carnap, 1955; Cole, 1983). The trading metaphor has laid a useful framework to interpret field-level citations in this regard: it makes analogies with concepts from international trade in that a field serves as a trading unit and can export knowledge by receiving citations and importing knowledge by sending citations; a field is a noticeable knowledge exporter if it enjoys a knowledge surplus, the fact that it exports more knowledge than it imports, and an importer if it has a knowledge deficit (Cronin & Meho, 2008; Yan et al., 2013). Using the trading metaphor, Cronin and Meho (2008) found that information science has become a more successful exporter of knowledge by receiving citations from computer science, engineering, business and management, and education. Similarly, Levitt, Thelwall, and Oppenheim (2011) found that library and information science grew the fastest in interdisciplinarity between 1990 and 2000 among all social science fields. Statistical models have enriched the field-level diffusion studies, methods such as epidemic models (Kiss et al., 2010), main path (Xiao et al., 2014), and dynamic network models (Gao & Guan, 2012; Rosas et al., 2013) were introduced to identify knowledge flow patterns in the fields of data quality research (Xiao et al., 2014), *h*-index research (Gao & Guan, 2012), kinesin research (Kiss et al, 2010), and National Institutes of Health (NIH) HIV/AIDS clinical trials (Rosas et al., 2013).

Lastly, there are attempts to jointly study patent and paper citations. An earlier research found that citations from U.S. patents to U.S. research papers have tripled over a six-year period from 1987 to 1993 (Narin, Hamilton, & Olivastro, 1997). Although research has assessed the impact of journal publications on patents (Fabrizio & Di Minin, 2008; Azoulay, Ding, & Stuart, 2009), these studies used small sets of papers and patents and did not provide an extensive picture on the mutual engagement of publications and patents at a higher and more abstract level.

While the previous literature has laid a solid theoretical and methodological foundation for quantitative studies of knowledge diffusion, there lack a holistic and dynamic examination of disciplinary knowledge flows. The goal of this research is to fill this gap by probing into the dynamic characteristics of sciences and social sciences at three integrated levels that include disciplines, disciplinary citation flows, and disciplinary ego-centric networks. Data, methods, and results are presented in the following sections.

Data

Scopus data

The Data section first introduces the Scopus data set used in this study; it then discusses the limitations of using the data set to study disciplinary knowledge flows. The data set was awarded by the Elsevier Bibliometrics Research Program. The intermediary data file was a journal-to-journal citation matrix for all indexed journals and proceedings in Scopus with a two-year citation window; that is, citations in year t to papers published in year t-2. Data on the following cited/citing years were obtained: 1997/1999, 2000/2002, 2003/2005, 2006/2008, and 2009/2011. The journal-to-journal citation data were aggregated to the discipline level using Elsevier's journal classification schema named All Science Classification Codes (ASJC). The schema comprises journals and proceedings, 307 minor subject areas, 27 major subject areas, and four top-level domains. A journal is typically assigned into one or a few minor subject areas; these minor subject areas are grouped into one of the major subject areas and these in turn are grouped into one of the four top-level domains: Life Sciences, Social Sciences, Physical Science, and Health Sciences (Figure 1). In this study, we will focus on the analysis of the 27 major subject areas and the four top-level domains.

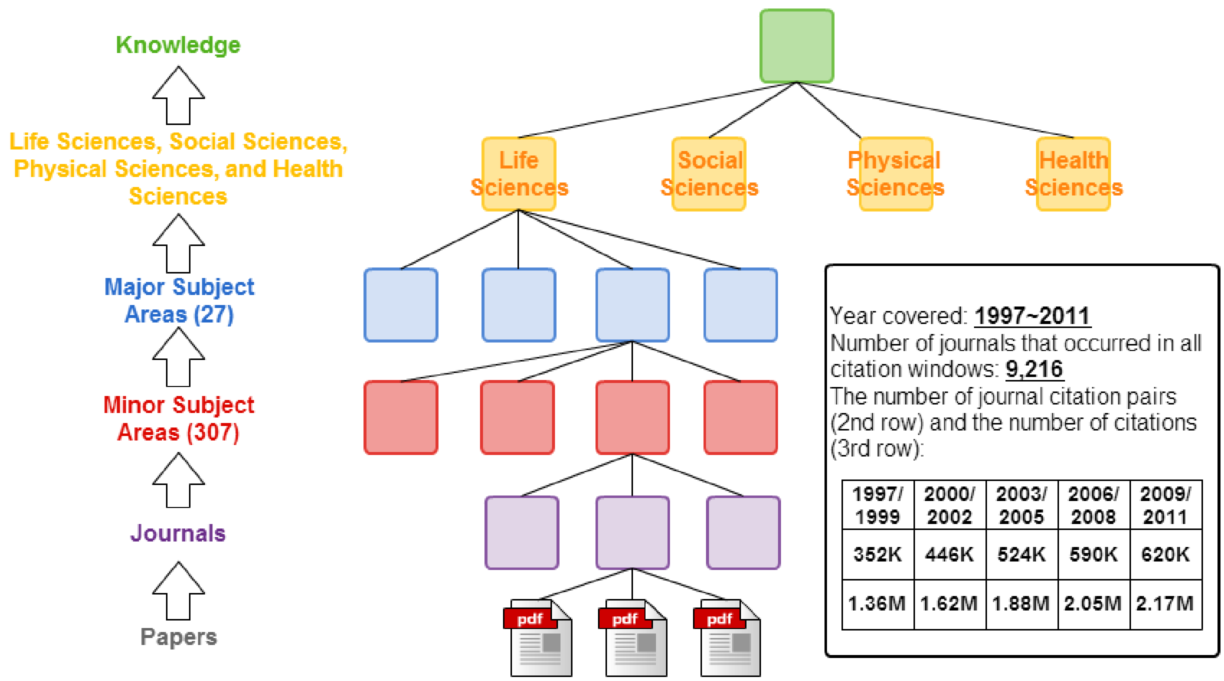
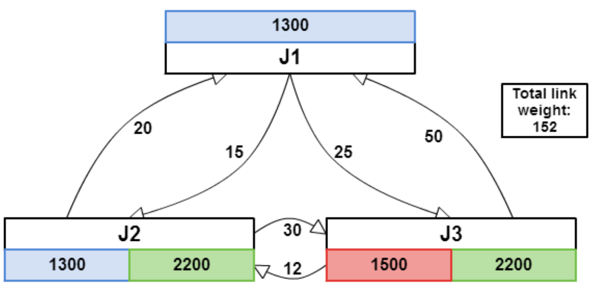


Figure 1. All Science Classification Code (ASJC) by Elsevier

There are in total 9,216 journals that occurred in all five citation windows and these journals were aggregated into the major subject level based on ASJC. Figure 1 shows the number of journal citation pairs formed by these journals and the aggregated number of citations for each citation window. In the 2009/2011 citation window, the data set comprises 619,753 journal citation links and 2,167,594 aggregated citations among the 9,216 journals. When aggregating these journals, fraction counting was considered in that if a journal is associated with multiple major subject areas, its citations are divided among these subjects. Figure 2 shows an example of the aggregation.

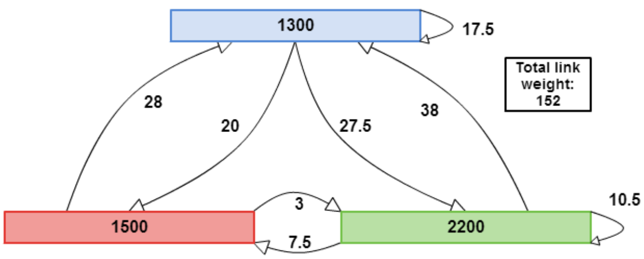
a



b*

	Citing	Cited	Citing Sub	Cited Sub	C.	N1	N2	N1*N2	F.C.
1	J1	J2	1300	1300	15	1	2	2	7.5
2	J1	J2	1300	2200	15	1	2	2	7.5
3	J2	J1	1300	1300	20	2	1	2	10
4	J2	J1	2200	1300	20	2	1	2	10
5	J1	J3	1300	1500	25	1	2	2	12.5
6	J1	J3	1300	2200	25	1	2	2	12.5
7	J3	J1	1500	1300	50	2	1	2	25
8	J3	J1	2200	1300	50	2	1	2	25
9	J2	J3	1300	1500	30	2	2	4	7.5
10	J2	J3	1300	2200	30	2	2	4	7.5
11	J2	J3	2200	1500	30	2	2	4	7.5
12	J2	J3	2200	2200	30	2	2	4	7.5
13	J3	J2	1500	1300	12	2	2	4	3
14	J3	J2	1500	2200	12	2	2	4	3
15	J3	J2	2200	1300	12	2	2	4	3
16	J3	J2	2200	2200	12	2	2	4	3

c



d

Number of journals (2nd row) that are assigned to different numbers of major subject areas (1st row)

1	2	3	4	5	6	7	8	Total
4,830	3,200	940	192	41	8	3	2	9,216

* C.: Number of citations; N1: the number of assigned subjects for the citing journal; N2: the number of assigned subjects for the cited journal; N.C.: the fractional number of citations=C./(N1*N2); Sum(N.C.)=152

Figure 2. An example of aggregating journal-level citations to the major subject area level

The example illustrates the citation relations among three sample journals J1, J2, and J3. Based on ASJC, J1 is associated with one major subject areas 1300 (Biochemistry); J2 is associated with two major subject areas 1300 (Biochemistry) and 2200 (Engineering); J3 is associated with two major subject areas 1500 (Chemical Engineering) and 2200 (Engineering). The fractional counting considers three factors: the number of citations from one journal to another (C), the number of associated subjects for the citing journal (N1), and the number of associated subjects for the cited journal (N2). The resulted fractional number of citations is C/(N1*N2). For instance, J2 cited J3 30 times and both J2 and J3 are associated with two subjects; the following four subject citation links are formed: Biochemistry (1300)-Chemical Engineering (1500), Biochemistry (1300)-Engineering (2200), Engineering (2200)-Chemical Engineering (1500), and Engineering (2200)-Engineering (2200), all of which has a fractional number of citations of 30/(2*2)=7.5. Thus, this fractional counting addresses the need to include journal multi-assignments while keeping the subject-level citations uninflated.

The numbers of journals that are associated with different numbers of major subject areas are illustrated in Figure 2(d). The numbers show that up to 52% of journals and proceedings are associated with one major subject area and up to 35% are associated with two major subject areas. Two sources are associated with eight major subject areas: *Bulletin of Mathematical Biology* and *Materials Letters*.

Limitations

The limitations of this study are primarily derived from the journal citation data and the employed journal classification scheme ASJC. First, although Scopus tends to have a more comprehensive coverage on journals and proceedings than the Web of Science database (Klavans & Boyack, 2007; Meho & Yang, 2008; Leydesdorff, de Moya-Anegón, & Guerrero-Bote, 2010; Leydesdorff, de Moya-Anegón, & de Nooy, 2014), it is not expected to contain all important scholarly literature—compared with biomedical related disciplines, social science and humanities may still have an inequitable visibility in Scopus (de Moya-Anegón et al., 2007). Second, the two-year citation window tends to favor subjects with high immediacy and penalize subject with lower immediacy (Schubert & Glänzel, 1986). For instance, in reference to the tendency of scientists to cite recent work, Stephen Cole (1983) argued that "[i]n highly codified fields we should find a faster rate…than in fields with lower levels of codification" (p. 125). Some of the more highly codified fields, according to Zuckerman and Merton (1973), include physics, biophysics, and chemistry which "show a larger share of reference to recent work; they exhibit a greater 'immediacy'" (p. 508). Third, in regards to ASJC, similar to other classification schemes, it takes into consideration several factors—citation patterns, editorial judgments, and managerial decisions (Garfield, Pudovkin, & Istomin, 2002). Moreover, ASJC seems to have more elaborate hierarchies for biomedicine as up to 10 major subject areas relate closely to this field, while general social science fields are grouped into one major subject area. Citations of social sciences are sparser and largely stay within the social sciences (Yan, 2014); thus the unbalanced representation may make it more difficult to inclusively capture the citation flows within social sciences. Realizing these limitations, in this study, major subject areas in ASJC was employed as a proxy to study disciplinary knowledge flows. Comparisons with other classification approaches, such as the Map of Science (Boyack, Börner, & Klavans, 2005; Börner et al., 2012) and modularity-based methods (Waltman & Van Eck, 2012; Zhang et al., 2010) are thus recommended as future work.

This work makes a "ceteris paribus" assumption that knowledge flows are the only changing variable here while all other factors remain constant. This assumption helps interpret the citation numbers but clearly subjects to some alternative accounts because knowledge production and diffusion is a complex social process that draws strengths from a variety of factors. Because of the observability and complexity issues, studies on knowledge transfer typically made ceteris paribus assumptions (e.g., Bresman, Birkinshaw, & Nobel, 1999; Contractor & Ra, 2002; Mu, Tang, & MacLachlan, 2009).

Another noticeable boundary of this work lies on the level of interpretations. This work presents rich descriptive findings; however, to further understand the numbers, one needs to examine more nuanced disciplinary citation practices, such as reference length, publishing frequency, and community size. One also needs to use additional data sources such as those on funding decisions and science policies to determine the latent mechanisms that may lead to the dynamic changes of disciplinary characteristics.

Methods

Characteristics of disciplines measured by scientific trading dimensions

To reveal disciplinary citation practices at node-, link-, and network-levels, a three-step approach is espoused. The first step examines individual disciplines' citation characteristics through scientific trading dimensions (i.e., incoming citations, cited/citing ratios, self-citation ratios, and citation dynamics). The second step analyzes citation flows between two disciplines. The third step involves the ego-centric analysis of individual disciplines' citation flow diversity measured by Shannon entropy. These efforts deliver dynamic, comprehensive perspectives to disciplinary citation patterns. Each step is elaborated in the following paragraphs.

Yan and colleagues' (2013) scientific trading dimensions were adopted: incoming citations signify scientific trading impact; cited/citing ratios leverage the balance between exported and imported knowledge; and self-citation ratios denote disciplinary self-dependence.

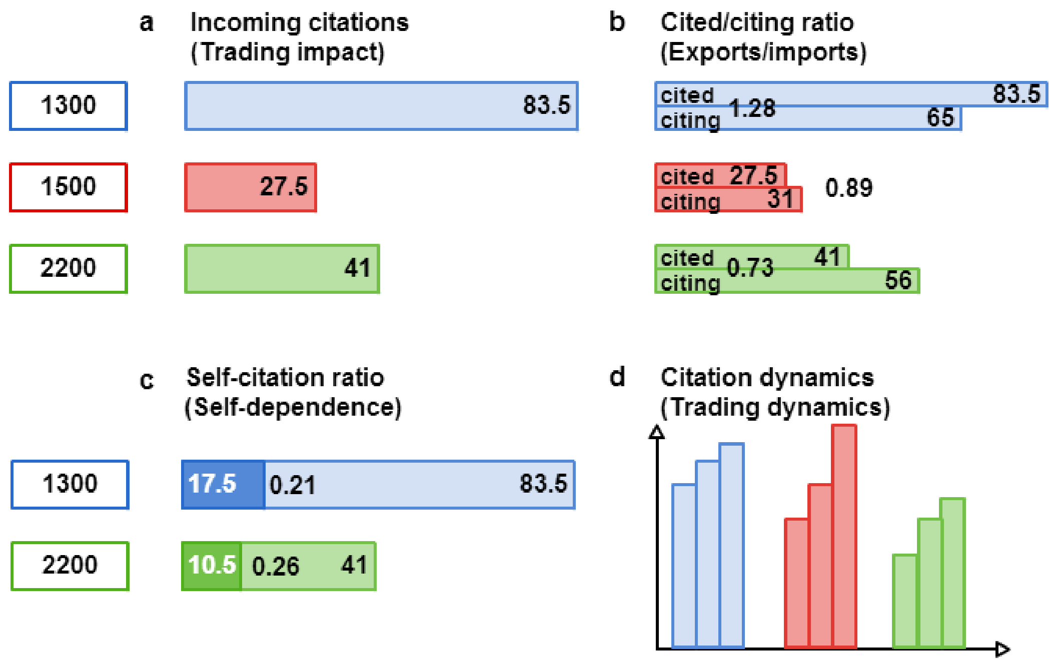


Figure 3. An example of calculating scientific trading dimensions

Using the same example introduced in Figure 2, Figure 3 illustrates the calculation of incoming citations (Figure 3(a)), cited/citing ratios (Figure 3(b)), self-citation ratios (Figure 3(c)), and citation dynamics (Figure 3(d)) for Biochemistry (1300), Chemical Engineering (1500), and Engineering (2200). The weighted directed subject-to-subject citation network can be represented as G=(V, A) where A represents the weighted directed link set and V represents the vertex set of subjects.

- Incoming citations (Trading impact): $trading\_impact_k = \sum_{i=1}^{n} G_{ik}$, for subject area k where G_{ik} is the incoming citations from subject i to k and n is the number of subject areas. In this study, n equals 27.

- Cited/citing ratio (Exports/imports): $\mathrm{export/import_k} = \frac{\sum_{i=1}^{n} G_{ik}}{\sum_{j=1}^{n} G_{kj}}$, for subject area k. In scientific trading, if a discipline exports more knowledge than it imports, it is a knowledge exporter; if a discipline imports more knowledge than it export, then it is a knowledge importer. Cited/citing ratio leverages the relationship between the exported and imported knowledge: a cited/citing ratio of one suggests a trading balance, a ratio greater one suggests a trading surplus, and a ratio smaller than one suggests a trading deficit.

- Self-citation ratio (Self-dependence): $\mathrm{self\_citation\_ratio_k} = \frac{G_{kk}}{\sum_{i=1}^{n} G_{ik}}$, for subject area k. Self-citation ratios have been proven to be an effective measure of disciplinary self-dependence (e.g., Borgman & Rice, 1992; Leydesdorff, 2011). Prior studies suggested that independent disciplines tend to have higher self-citation ratios and are likely to be those having established educational systems and distinctive scholarly communication channels whereas dependent disciplines tend to have lower self-citation ratios and are likely to be newer or less-established ones (e.g., Yan et al., 2013).

- Citation dynamics (Trading dynamics): $\mathrm{trading\_dynamics_k} = \mathrm{slope}(\mathrm{x_{k,t}}, \mathrm{x_{k,t+1}}, \dots)$, for subject area *k*. Ideally, *x* can be any of the incoming citations, cited/citing ratios, or self-citations ratios; however, as results have shown that the slopes of cited/citing ratios and self-citation ratios for most subjects in this study were not significant at the 0.05, only the slopes of incoming citations were reported in this study. Incoming citations were fit into a one-independent variable linear regression where the citation windows are the independent variable ranging from 1 to 5 (*t*: [1,5]) and the normalized incoming citations $\frac{trading\_impact_{k:t}}{\sum_t^5 trading\_impact_{k:t}} = \frac{\sum_{i=1}^n G_{ik:t}}{\sum_{t=1}^5 \sum_{i=1}^n G_{ik:t}}$ for subject *k* are the dependent variable. Slopes from each individual linear regression can thus be obtained. The normalization makes it possible to compare slopes across different subjects.

Disciplinary citation flows

This subsection introduces the link-level approach to examine disciplinary citation flows. The creation of knowledge is not independent, but rather, it is dependent on the transfer of knowledge from one to another. To examine the dynamic aspect of such disciplinary knowledge flows, slopes of each citation link over the past five citation windows were obtained using the same normalization method: $G_{ik}^{norm} = \frac{G_{ik:t}}{\sum_{t=1}^5 G_{ik:t}}$, where G_{ik}^{norm} is the weight of the normalized citation link from the citing subject *i* to the cited subject *k*. The dynamics of disciplinary citation flows can be expressed as:

$$flow\_dynamics_{ik} = slope(G_{ik,t}^{norm}, G_{ik,t+1}^{norm}, \dots)$$

Because weaker links are more susceptible to change than established ones, citation links are examined separately based on link weights in the 1997/1999 citation window: those between 100 and 1,000, those between 1,000 and 10,000, and those heavier than 10,000.

Disciplinary ego-centric networks measured by Shannon entropy

This subsection introduces the ego-centric approach to examine disciplinary citation diversity. Disciplines vary greatly in their ability to export and import knowledge: some are more permeable while others are more self-dependent. To quantify such interdisciplinary diversity, Shannon entropy was applied. Shannon entropy has been widely used in evaluating signal transmissions (e.g., Lin, 1991). In the context of scientific trading, it measures, for each subject, the proportions of each incoming or outgoing citation sources (i.e., major subject areas). Shannon entropy therefore effectively assesses the knowledge flow diversity for each subject (e.g., Zhang et al., 2010). Citation diversity as measured by Shannon entropy:

$$H_{k:incoming} = -\sum_{i=1}^{n} \frac{G_{ik}}{\sum_{j=1}^{n} G_{jk}} \ln \frac{G_{ik}}{\sum_{j=1}^{n} G_{jk}}$$

where $H_{k:incoming}$ is the Shannon entropy for subject *k* measured by incoming citations to *k*, G_{ik} is the incoming citations from subject *i* to *k*, $\frac{G_{ik}}{\sum_{j=1}^{n} G_{jk}}$ is the proportion of incoming citations from *i* to *k* over the total incoming citations of *k*, and *n* is the number of subject areas. The Shannon entropy for subject *k* measured by outgoing citations from *k* can thus be expressed as:

$$H_{k:outgoing} = -\sum_{i=1}^{n} \frac{G_{ki}}{\sum_{j=1}^{n} G_{kj}} \ln \frac{G_{ki}}{\sum_{j=1}^{n} G_{kj}}$$

Shannon entropy can also be applied to the top-level domains for each subject area. The Shannon entropy for subject *k* measured by incoming citations from *k* can thus be expressed as: $H_{k:incoming} = -\sum_{i=1}^{m} \frac{\sum_{a=1 \cap a \in i}^{n} G_{ik}}{\sum_{j=1}^{n} G_{jk}} \ln \frac{\sum_{a=1 \cap a \in i}^{n} G_{ik}}{\sum_{j=1}^{n} G_{jk}}$, where *m* is the number of top-level domains—in this study, m equals 4 (*i*=1 for Life Sciences, 2 for Social Sciences, 3 for Physical Science, and 4 for Health Sciences), *n* is still the number of subject areas, and $\sum_{a=1 \cap a \in i}^{n} G_{ik}$ determines the number of citations subject *k* received from all subjects that are assigned into domain *i*. The Shannon entropy for subject *k* as measured by outgoing citations from *k* can thus be expressed as: $H_{k:outgoing} = -\sum_{i=1}^{m} \frac{\sum_{a=1 \cap a \in i}^{n} G_{ki}}{\sum_{j=1}^{n} G_{kj}} \ln \frac{\sum_{a=1 \cap a \in i}^{n} G_{ki}}{\sum_{j=1}^{n} G_{kj}}$.

Results

This section first introduces results on the characteristics of disciplines measured by scientific trading dimensions; it then reports the characteristics of disciplinary citation flows; it lastly presents results on disciplinary flow diversity through Shannon entropy.

Characteristics of disciplines

We reports results on four disciplinary trading dimensions: trading impact (Figure 4), cited/citing ratios (Figure 5), disciplinary self-dependence (Figure 6), and trading dynamics (Table 1). In Figure 4, the *y*-axis denotes the $trading\_impact_{k:t}$ of subject *k* in citation window *t*.

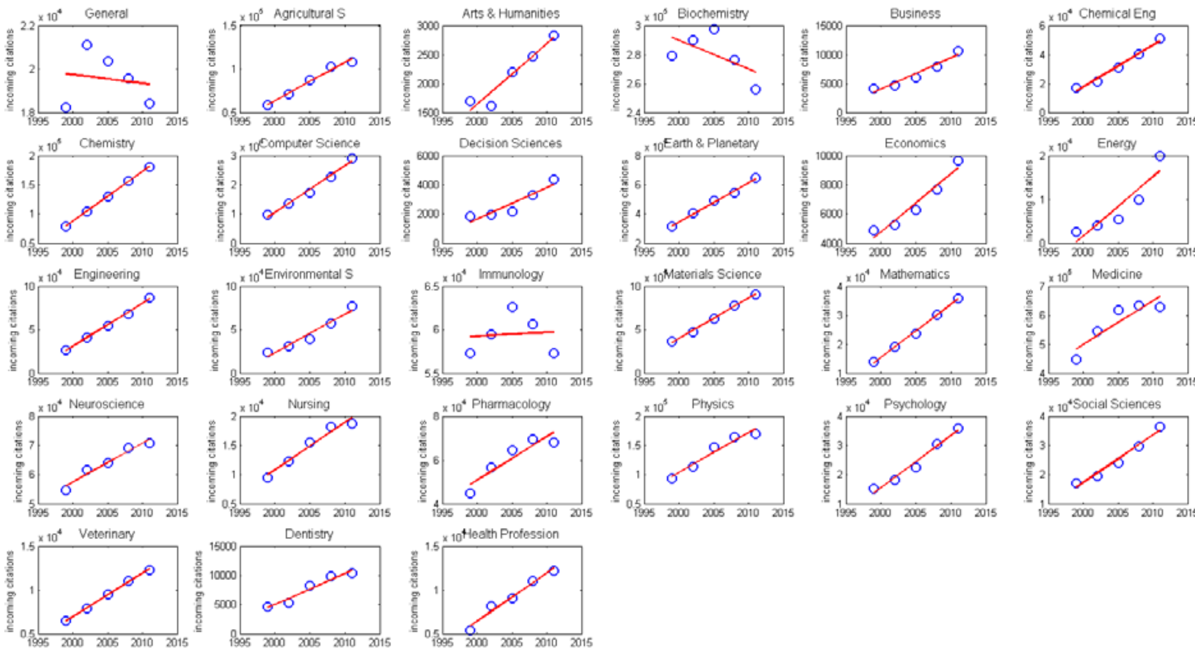


Figure 4. Incoming citations of the 27 subjects

Except for General and Biochemistry, all other disciplines received more scientific impact during the past five citation windows. The increasing rate varies across disciplines: while Chemical Engineering, Energy, and Environmental Science gained a significant amount of trading impact, Immunology had a narrower gain. These dynamic characteristics will be further examined in Table 1.

Results on cited/citing ratios are illustrated in Figure 5 where the *y*-axis denotes the $export/import_{k:t}$ of subject *k* in citation window *t*.

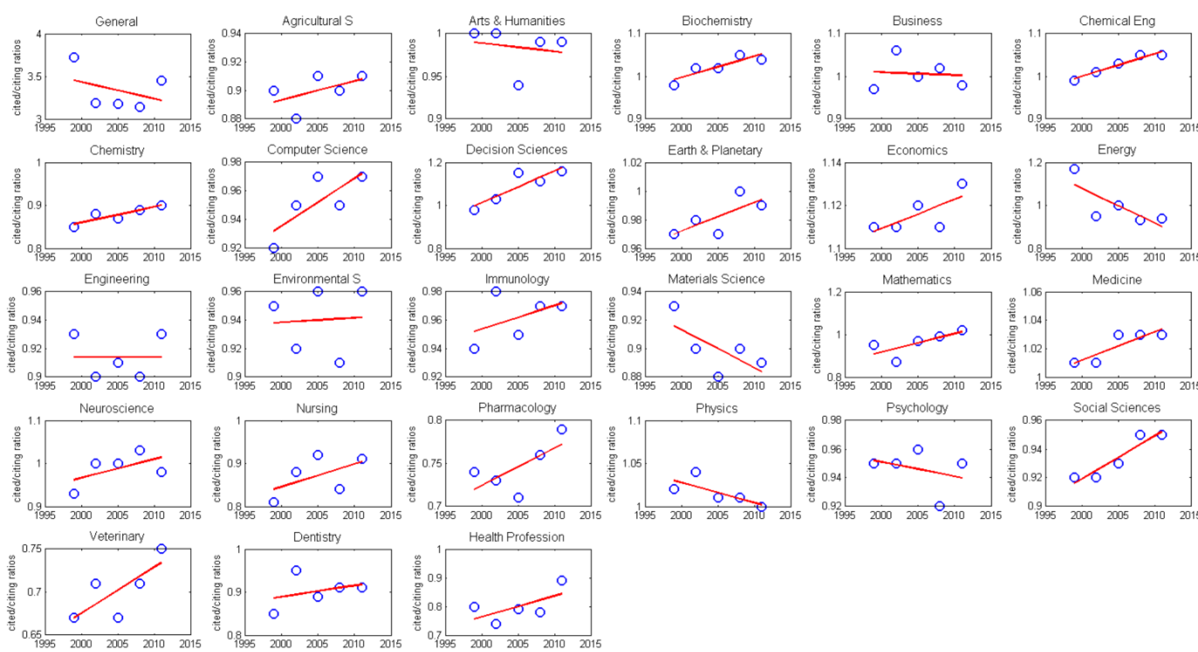


Figure 5. Cited/citing ratios of the 27 subjects

We see from Figure 5 that most subjects maintained relatively stable cited/citing ratios during the past five citation windows, with the exception of General, Energy, and Materials Science in that a noticeable drop can be found and Computer Science, Decision Sciences, Pharmacology, Social Sciences, and Veterinary in that a visible increase is present. The results suggest that the former group became importer-oriented, while the latter group had a tendency to become exporter-oriented.

Diachronical patterns of self-citation ratios for the 27 subjects are reported in Figure 6 where the *y*-axis denotes the $self\_citation\_ratio_{k:t}$ of subject *k* in citation window *t*.

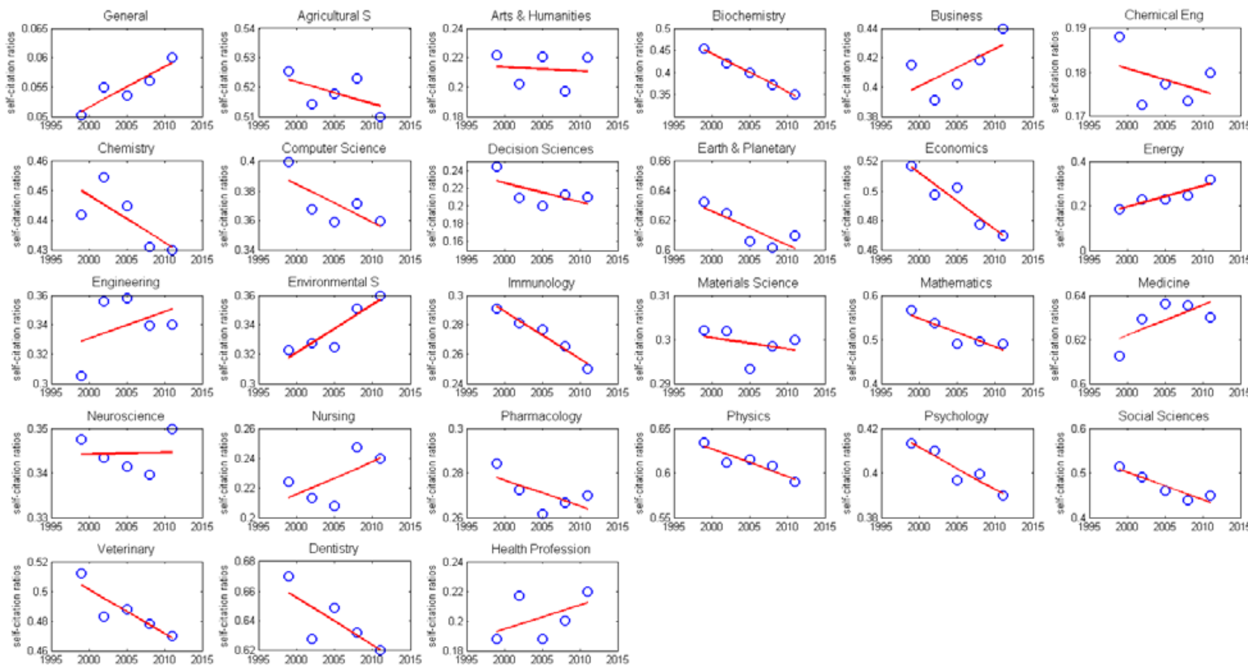


Figure 6. Self-citation ratios of the 27 subjects

While most subjects in Figure 6 exhibited moderately declining self-citation ratios, there are subjects that have discernable ratio increase, including General, Business, Energy, Environmental Science, and Nursing, indicating that these are the subjects that became more sustained on its own disciplinary knowledge. Subject that have noticeable ratio decrease include Biochemistry, Economics, Immunology, Psychology, Social Sciences, Veterinary, and Dentistry, implying a growing interdisciplinary dependency. Subdomain analyses are necessary to further understand these dynamic characteristics.

Table 1 lists incoming citations, cited/citing ratios, and self-citation ratios in 2009/2011 as well as the slopes for normalized citations.

Table 1. Incoming citations, cited/citing ratios, and self-citation ratios for the 27 subjects

	Incoming citations (*rank*)	Slope of norm. citations (*rank*)	Cited/citing ratio (*rank*)	Self-citation ratio (*rank*)
General	63,584 (11)	-0.0045 (27)	3.45 (1)	0.06 (27)
Agricultural and Biological Sciences	98,695 (5)	0.0318 (19)	0.91 (20)	0.51 (5)

Arts and Humanities	2,806 (27)	0.0289 (20)	0.99 (9)	0.22 (23)
Biochemistry, Genetics and Molecular Biology	266,485 (2)	-0.0011 (26)	1.04 (5)	0.35 (15)
Business, Management and Accounting	10,486 (23)	0.0477 (8)	0.98 (11)	0.44 (10)
Chemical Engineering	53,276 (14)	0.0562 (3)	1.05 (4)	0.18 (26)
Chemistry	162,392 (4)	0.0413 (13)	0.90 (23)	0.43 (11)
Computer Science	28,202 (18)	0.0537 (5)	0.97 (13)	0.36 (13)
Decision Sciences	5,021 (26)	0.0544 (4)	1.16 (2)	0.21 (25)
Earth and Planetary Sciences	63,767 (10)	0.0345 (18)	0.99 (9)	0.61 (3)
Economics, Econometrics and Finance	10,904 (22)	0.0362 (15)	1.13 (3)	0.47 (7)
Energy	18,747 (19)	0.0912 (1)	0.94 (18)	0.32 (18)
Engineering	80,802 (6)	0.0533 (6)	0.93 (19)	0.34 (17)
Environmental Science	73,599 (8)	0.0585 (2)	0.96 (15)	0.36 (13)
Immunology and Microbiology	55,429 (12)	0.0013 (25)	0.97 (13)	0.25 (21)
Materials Science	80,546 (7)	0.0430 (9)	0.89 (24)	0.30 (19)
Mathematics	36,685 (15)	0.0504 (7)	1.02 (7)	0.49 (6)
Medicine	647,218 (1)	0.0166 (23)	1.03 (6)	0.63 (1)
Neuroscience	69,694 (9)	0.0151 (24)	0.98 (11)	0.35 (15)
Nursing	17,085 (20)	0.0355 (17)	0.91 (20)	0.24 (22)
Pharmacology, Toxicology and Pharmaceutics	53,589 (13)	0.0230 (22)	0.79 (26)	0.27 (20)
Physics and Astronomy	170,679 (3)	0.0282 (21)	1.00 (8)	0.59 (4)
Psychology	33,977 (17)	0.0430 (9)	0.95 (16)	0.39 (12)
Social Sciences	34,472 (16)	0.0401 (14)	0.95 (16)	0.45 (9)
Veterinary	9,192 (25)	0.0359 (16)	0.75 (27)	0.47 (7)
Dentistry	9,350 (24)	0.0427 (11)	0.91 (20)	0.62 (2)
Health Professions	10,911 (21)	0.0426 (12)	0.89 (24)	0.22 (23)

In regards to incoming citations, subjects such as Medicine, Biochemistry, Physics, Chemistry, and Agricultural Sciences are the ones with the highest trading impact. Energy, Environmental Science, Chemical Engineering, Decision Sciences, and Computer Science had the fastest growths, indicating that their disciplinary knowledge became more visible among others. Highly visible disciplines such as Biochemistry, Immunology, Neuroscience, Medicine, and Pharmacology, on the other hand, had the least fast growth.

As for cited/citing ratios, the subject of General received a cited/citing ratio far above one. This subject contains journals such as *Science* and *Nature*—their papers were more intensively cited than the others, thus resulting in a high cited/citing ratio. Other knowledge exporters include Decision Sciences, Economics, Chemical Engineering, and Biochemistry. Subjects such as Veterinary Science and Pharmacology had a cited/citing ratio smaller than one, suggesting a knowledge deficit.

Except for General, all other subjects retained high self-citation ratios, ranging from 0.18 to 0.63. Because General primarily contains multidisciplinary journals, its low self-citation ratio comes as no surprise. Disciplines such as Medicine, Dentistry, and Earth and Planetary Sciences had the highest self-citation ratios, suggesting that they possessed a more distinctive cognitive core than the others. Chemical Engineering and Decision Sciences, on the other hand, were more permeable and did not yet form more distinguishable cognitive bases.

Dynamic patterns of disciplinary citation flows

We now further our analysis from disciplines to disciplinary citation flows. Figure 7 shows three sets of citation links that had the highest increases: link weights between 100 and 1,000 (first ten images, in blue), between 1,000 and 10,000 (middle ten images, in green), and greater than 10,000 (last ten images, in red).

The y-axis shows the percentage of citations in citation window t over the sum of citations from all windows $\frac{G_{ik:t}}{\sum_{t=1}^{5} G_{ik:t}}$.

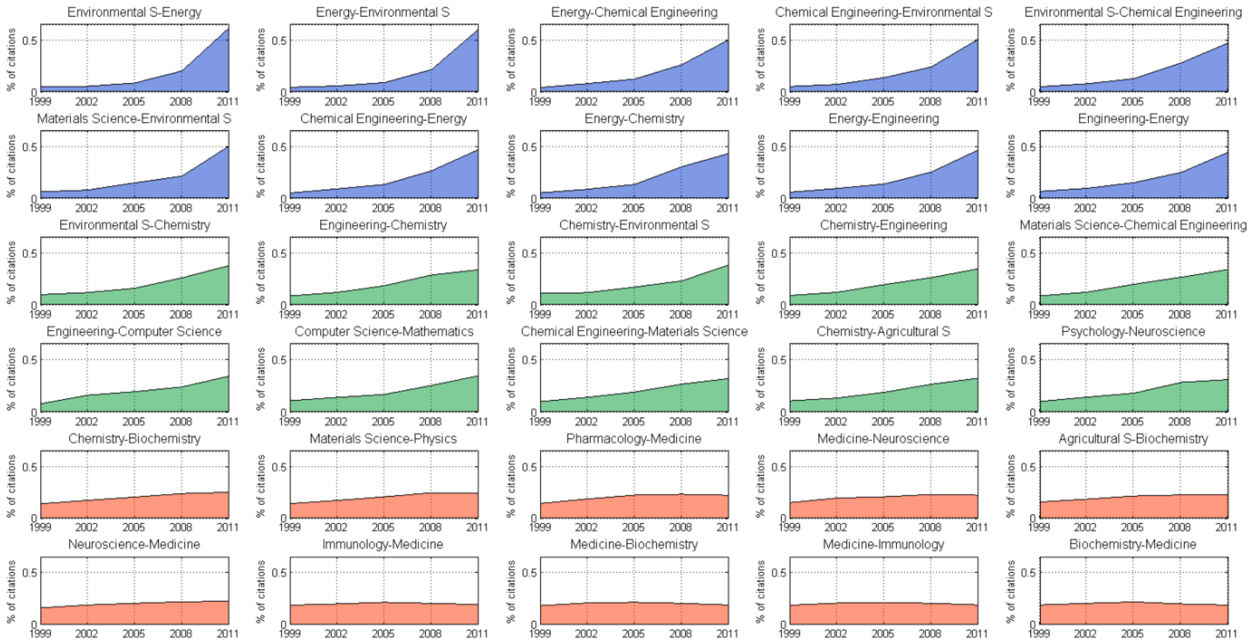


Figure 7. Top citation flows that had the highest increase for three link weight levels: 10e2~10e3 (blue), 10e3~10e4 (green), and 10e4~ (red)

Overall, among the top citation flows illustrated in Figure 7, those in the 10e2~10e3 category grew faster than those in the 10e3~10e4 category and the latter grew faster than those in the 10e4~ category, suggesting that more established citation flows are less susceptible to change. Energy stands out in the 10e2~10e3 category: it formed stronger exporting and importing relationships with Environmental Science, Chemical Engineering, Chemistry, and Engineering. In the 10e3~10e4 category, Chemistry strengthened its connections with Environmental Science, Engineering, and Agricultural Sciences; other heightened citation flows in this category include the ones from Materials Science to Chemical Engineering, from Engineering to Computer Science, from Computer Science to Mathematics, from Chemical Engineering to Materials Science, and from Psychology to Neuroscience. In the 10e4~ category, Medicine enhanced its relationship with Chemistry, Pharmacology, Neuroscience, and Immunology. These escalated citation flows suggest that the connected subjects became more inter-dependent and more absorptive towards each other's knowledge.

Dynamic patterns of the diversity of disciplinary citation practices

In this subsection, we report results obtained from disciplinary ego-centric network analyses. These results help depict the diversity of disciplinary citation practices. Area maps were employed to render visualizations on the diversity of incoming citations (Figure 8) and outgoing citations (Figure 9). Both figures adopted the same color coding scheme: different shades of blue for subjects in Life Sciences, shades of green for subjects in Social Sciences, shades of yellow for subjects in Physical Sciences, and shades of red for subjects in Health Sciences (as seen in the legends).

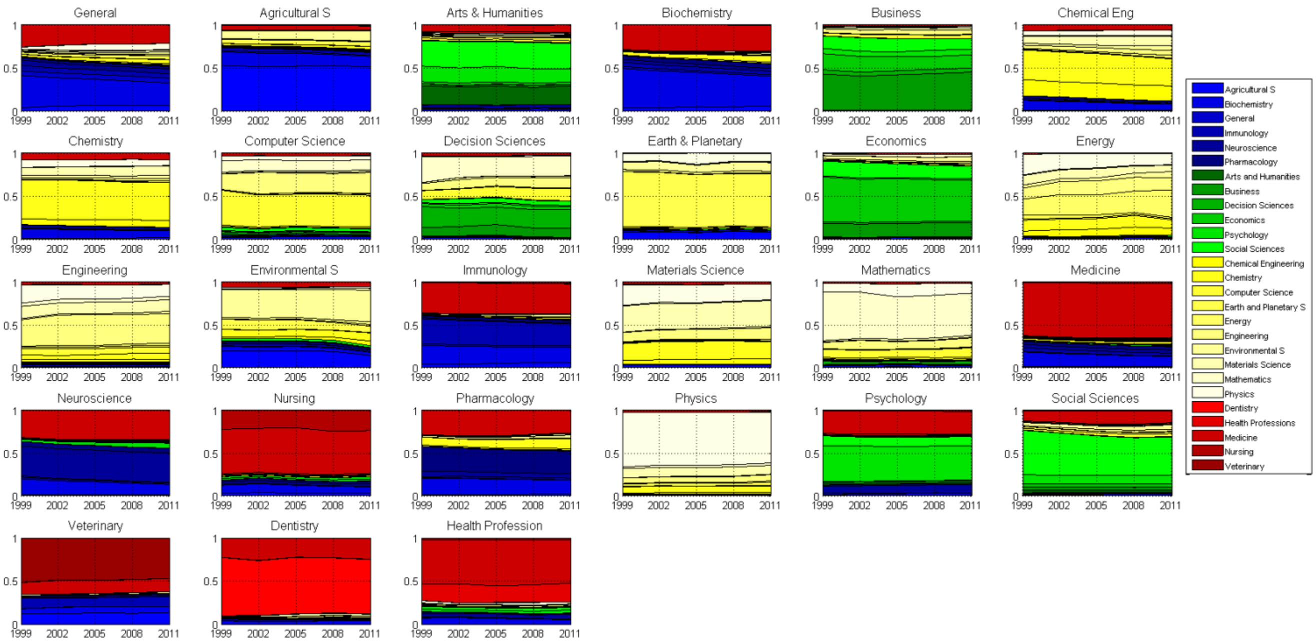


Figure 8. An area map representation of sources of incoming citations for the 27 subjects

Most subjects had a dominant source of incoming citations, typically the subjects themselves. A few subjects, however, were largely dependent on subjects other than themselves; for instance, Arts and Humanities received most citations from Social Sciences; Immunology, Nursing, and Health Professions received most citations from Medicine. In addition, some subjects maintained two or a few equally important knowledge importers; for instance, Biochemistry had Medicine and itself; Computer Science had Mathematics, Engineering, and itself; Neuroscience had Medicine and itself; Pharmacology had Medicine and itself; and Psychology had Medicine and itself. Diachronically, ratios of sources of incoming citations remained relatively stable, though there was a gradual percentile decline of the primary knowledge importer for some disciplines (e.g., Physics and Veterinary). Such dynamic changes will be captured through a time series analysis through Shannon entropy.

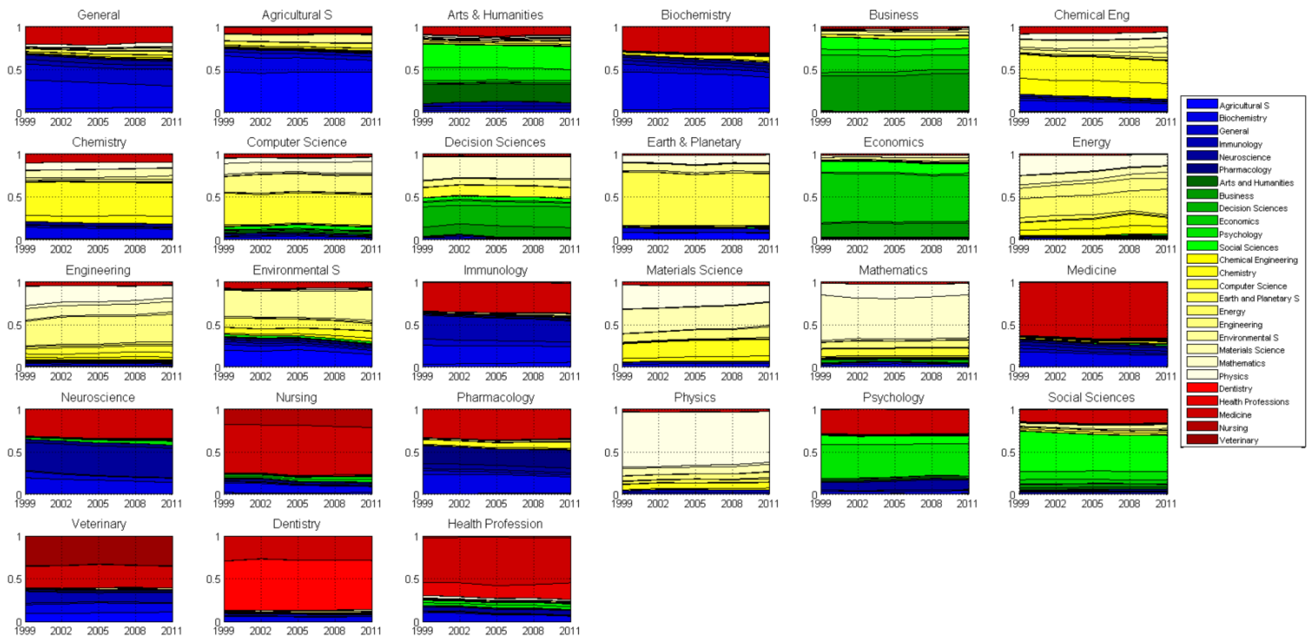


Figure 9. An area map representation of sources of outgoing citations for the 27 subjects

Overall, sources of outgoing citations for each subject have similar patterns with incoming citations, in that the majority of subjects imported most knowledge from themselves, a few imported the most from subjects other than themselves, and a few imported an equal amount from two or more subjects.

Table 2 lists the Shannon entropy for incoming citations and outgoing citations in 2009/2011.

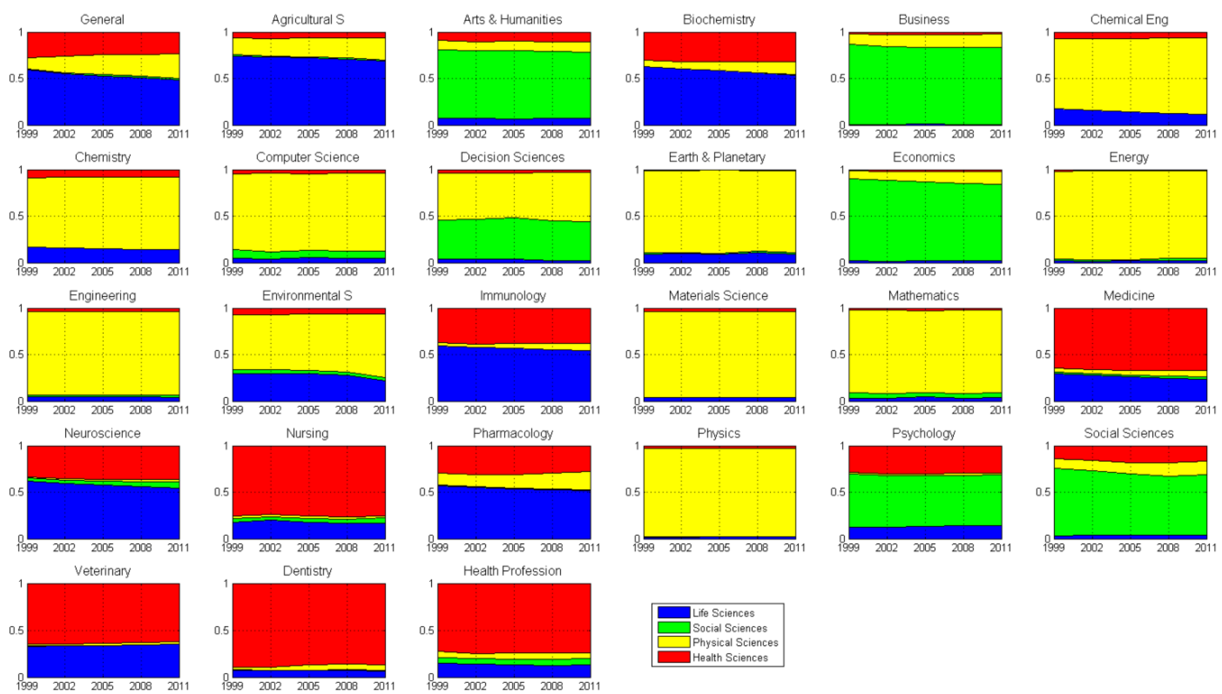
Table 2. Shannon entropy for incoming citations and outgoing citations for the 27 subjects

	Incoming citations		Outgoing citations	
	Shannon Entropy (*rank*)	Slope (*rank*)	Shannon Entropy (*rank*)	Slope (*rank*)
General	3.47 (1)	0.0087 (1)	3.29 (5)	0.0074 (1)
Agricultural and Biological Sciences	2.64 (17)	0.0026 (13)	2.78 (14)	0.0026 (10)
Arts and Humanities	3.20 (4)	0.0014 (20)	3.38 (1)	0.0018 (15)
Biochemistry, Genetics and Molecular Biology	2.82 (13)	0.0082 (2)	2.74 (15)	0.0074 (1)
Business, Management and Accounting	2.77 (14)	0.0013 (21)	2.80 (13)	0.0017 (16)
Chemical Engineering	3.17 (6)	0.002 (18)	3.33 (3)	0.0016 (18)
Chemistry	2.87 (11)	0.0024 (15)	3.02 (10)	0.0013 (20)
Computer Science	3.00 (9)	0.0006 (25)	3.12 (7)	0.0001 (25)
Decision Sciences	3.20 (4)	0.0017 (19)	3.10 (8)	-0.0005 (26)
Earth and Planetary Sciences	2.12 (26)	0.0024 (15)	2.21 (24)	0.0021 (13)
Economics, Econometrics and Finance	2.60 (19)	0.0057 (6)	2.40 (22)	0.0052 (5)
Energy	3.11 (7)	-0.0008 (26)	3.25 (6)	0.0009 (22)
Engineering	3.22 (3)	0.0013 (21)	3.37 (2)	0.0022 (11)
Environmental Science	3.23 (2)	0.001 (24)	3.31 (4)	0.0012 (21)
Immunology and Microbiology	2.72 (15)	0.0068 (3)	2.68 (17)	0.0059 (4)
Materials Science	2.87 (11)	0.0026 (13)	3.00 (11)	0.0043 (6)
Mathematics	2.72 (15)	0.0051 (7)	2.69 (16)	0.0022 (11)
Medicine	2.24 (25)	0.003 (10)	2.11 (26)	0.0017 (16)
Neuroscience	2.50 (20)	0.0029 (12)	2.47 (20)	0.0016 (18)
Nursing	2.38 (22)	0.0011 (23)	2.18 (25)	0.0002 (24)
Pharmacology, Toxicology and Pharmaceutics	2.93 (10)	0.0051 (7)	2.82 (12)	0.0029 (8)

Physics and Astronomy	2.29 (24)	0.0047 (9)	2.37 (23)	0.0067 (3)
Psychology	2.63 (18)	0.0021 (17)	2.64 (18)	0.0019 (14)
Social Sciences	3.02 (8)	0.0058 (5)	3.08 (9)	0.004 (7)
Veterinary	2.35 (23)	0.003 (10)	2.57 (19)	0.0009 (22)
Dentistry	1.77 (27)	0.0064 (4)	1.87 (27)	0.0027 (9)
Health Professions	2.49 (21)	-0.0008 (26)	2.46 (21)	-0.0021 (27)

The subjects of General, Environmental Science, Engineering, Decision Sciences, and Arts and Humanities had the highest Shannon entropy for incoming citations; Arts and Humanities, Engineering, Chemical Engineering, Environmental Science, and General had the highest Shannon entropy for outgoing citations—these subjects are thus the most diversified and interdisciplinary at the discipline level. On the other hand, Dentistry, Earth and Planetary Sciences, Medicine, Physics, and Veterinary had the lowest Shannon entropy for incoming citations; Dentistry, Medicine, Nursing, Earth and Planetary Sciences, and Physics had the lowest Shannon entropy for outgoing citations. Dynamically, except for three subjects, Computer Science, Energy, and Health Professions, all other subjects gained entropy. General, Biochemistry, and Immunology possessed the highest entropy growth (for both incoming and outgoing citations), suggesting that they made the greatest effort in diversifying their scientific trading practices.

Subject-level citations can be further aggregated into top domains. Sources of domain-level incoming (Figure 10) and outgoing citations (Figure 11) for each subject are illustrated[2]. Both figures adopted the same color coding scheme: blue for Life Sciences, green for Social Sciences, yellow for Physical Sciences, and red for Health Sciences.



[2] The subject of General is not assigned to any of the four domains in ASJC; its citations are therefore not included in calculating domain-level citations for each subject.

Figure 10. An area map representation of the aggregated sources of incoming citations for 27 subjects

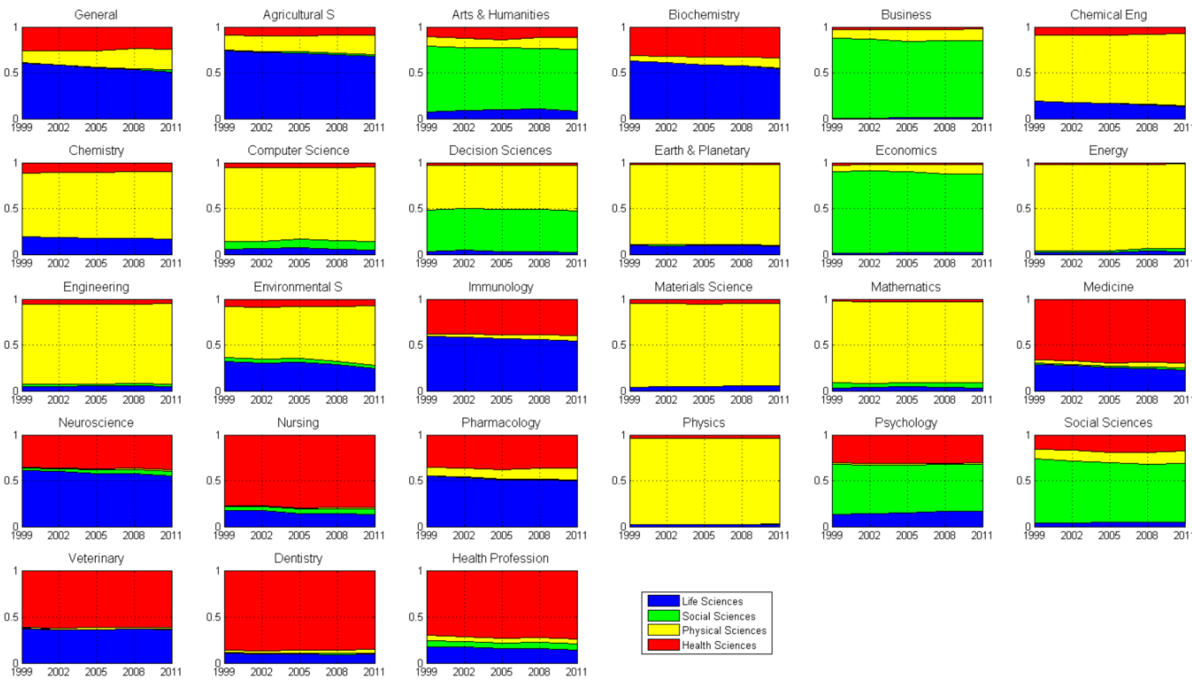


Figure 11. An area map representation of the aggregated sources of outgoing citations for 27 subjects

Both figures show that for most subjects, the majority of incoming and outgoing citations took place within one particular top-level domain. For instance, most incoming and outgoing citations of Arts and Humanities, Business, Economics, and Social Sciences occurred within the top-level domain of Social Sciences; most incoming and outgoing citations of Chemical Engineering, Chemistry, Computer Science, Earth & Planetary Science, Energy, Engineering, Materials Science, Mathematics, and Physics occurred within the top-level domain of Physical Sciences; most incoming and outgoing citations of Medicine, Nursing, Dentistry, and Health Profession occurred within Health Sciences. Subjects assigned to the top-level domain of Life Sciences pertained to less homogeneity: incoming and outgoing citations of subjects such as Agricultural Sciences, Biochemistry, Immunology, Neuroscience, and Pharmacology had a noticeable cross over with Health Sciences and Physical Sciences. Additionally, two Social Sciences assigned subjects Decision Sciences and Psychology also had mixed domain-level trading practices—the former had marked knowledge trading with Social Sciences and Physical Sciences, and the latter with Social Sciences and Health Sciences.

Table 3 lists the Shannon entropy for incoming citations and outgoing citations aggregated at the top-level domains in 2009/2011.

Table 3. Shannon entropy for aggregated incoming citations and outgoing citations for the 27 subjects

	Incoming citations		Outgoing citations	
	Shannon Entropy (*rank*)	Slope (*rank*)	Shannon Entropy (*rank*)	Slope (*rank*)
General	1.62 (1)	0.0076 (6)	1.59 (1)	0.0075 (4)
Agricultural and Biological Sciences	1.21 (13)	0.0042 (10)	1.25 (11)	0.0049 (11)

Arts and Humanities	1.33 (7)	0.0036 (11)	1.43 (5)	0.0042 (12)
Biochemistry, Genetics and Molecular Biology	1.45 (4)	0.0083 (3)	1.39 (6)	0.008 (3)
Business, Management and Accounting	0.81 (20)	0.0064 (8)	0.78 (19)	0.0061 (7)
Chemical Engineering	0.85 (19)	-0.0103 (27)	0.94 (18)	-0.0087 (27)
Chemistry	0.99 (16)	-0.004 (25)	1.10 (15)	-0.0037 (25)
Computer Science	0.86 (17)	-0.003 (23)	0.96 (17)	-0.0019 (22)
Decision Sciences	1.27 (9)	-0.0039 (24)	1.30 (9)	-0.0026 (23)
Earth and Planetary Sciences	0.62 (23)	0 (19)	0.66 (24)	-0.0013 (21)
Economics, Econometrics and Finance	0.86 (17)	0.014 (1)	0.75 (21)	0.0093 (1)
Energy	0.39 (26)	-0.0003 (20)	0.50 (26)	0.0086 (2)
Engineering	0.60 (24)	-0.0022 (22)	0.72 (22)	-0.0003 (19)
Environmental Science	1.27 (9)	-0.0049 (26)	1.35 (7)	-0.0045 (26)
Immunology and Microbiology	1.32 (8)	0.0072 (7)	1.28 (10)	0.0066 (6)
Materials Science	0.47 (25)	0.0005 (18)	0.58 (25)	0.006 (8)
Mathematics	0.68 (22)	0.0011 (15)	0.68 (23)	0.0009 (18)
Medicine	1.27 (9)	0.0023 (13)	1.21 (12)	0.0021 (17)
Neuroscience	1.41 (6)	0.0082 (5)	1.35 (7)	0.0071 (5)
Nursing	1.09 (15)	0.0011 (15)	1.00 (16)	-0.0006 (20)
Pharmacology, Toxicology and Pharmaceutics	1.53 (2)	0.0044 (9)	1.48 (3)	0.0041 (13)
Physics and Astronomy	0.34 (27)	0.0008 (17)	0.41 (27)	0.0026 (15)
Psychology	1.51 (3)	0.0022 (14)	1.54 (2)	0.0025 (16)
Social Sciences	1.44 (5)	0.0083 (3)	1.44 (4)	0.005 (10)
Veterinary	1.11 (14)	0.0033 (12)	1.12 (14)	0.0027 (14)
Dentistry	0.72 (21)	0.0114 (2)	0.76 (20)	0.0051 (9)
Health Professions	1.22 (12)	-0.0007 (21)	1.20 (13)	-0.0034 (24)

General, Pharmacology, Psychology, Biochemistry, Social Sciences, and Arts and Humanities were the most diversified subjects when considering incoming and outgoing citations at the four top-level domains. Meanwhile, Physics, Energy, Materials Science, Engineering, Mathematics, and Earth and Planetary Sciences were the least diversified. Diachronically, eight subjects (i.e., Chemical Engineering, Chemistry, Computer Science, Decision Sciences, Energy, Engineering, Environmental Science, and Health Professions) became less diversified as knowledge exporters while nine subjects (i.e., Chemical Engineering, Chemistry, Computer Science, Decision Sciences, Earth and Planetary Sciences, Engineering, Environmental Science, Nursing, and Health Professions) became less diversified as knowledge importers. When comparing the results with the subject-level Shannon entropy, it is found that although disciplines became more interdisciplinary-oriented at the subject-level, cross top-domain level knowledge transfer was less practiced.

Discussions

Disciplinarity and interdisciplinarity

Modern science is organized through disciplines (Klein, 1990). They vary greatly in permeability and cognitive autonomy (e.g., Klein, 1996; Yan et al., 2013; Yan, 2014). This study shows that disciplines such as earth and planetary science, medicine, and dentistry have the highest self-citation ratios and are thus highly specialized and strongly dependent on their own knowledge bases. These disciplines typically have established educational institutions (e.g., Department of Geology, School of Medicine, and School of Dentistry), professional societies (e.g., in the U.S.: Geological Society of America, American Medical Association, and American Dental Association), and scholarly communication channels that newer, less established disciplines may be unable to maintain (Yan et al., 2013). The results can also be explained by the dependent relationship of applied science upon basic science: while basic science fields are more self-dependent, applied science fields tend to cite basic science fields (Narin, Pinski, & Gee, 1976; Boyack et al., 2014), or in Narin and colleagues' words "[t]he basic research journals and basic research fields are highly influential: their citation-influence measures are significantly greater than the measures for the clinical journals and fields" (p. 43).

Diachronically, based on the results on Shannon entropy, the trading practices of most disciplines are increasingly becoming more diversified (with the exception of Energy, Decision Sciences, and Health Professions), signifying that they are more inclined to import higher volumes of knowledge from a greater number of disciplines. The findings are somewhat different from Zitt and Cointet's (2014) study in that a steady drop in variances of normalized impact and relative growth were found for the Web of Science discipline-level data from 1999 to 2008. The difference may be attributed to the "citing-side normalization" used in their study that weights citation links proportionally to the average outgoing links by a node. The normalization thus corrected the "absolute growth or the average impact over science… [and is] in contract with long-range analyses…which focus on volumes of publications and citations" (p. 7). Historically, there was a tendency towards a unified science (Neurath, 1996). This tendency was largely driven by the unity of language (e.g., Hyland, 2004) and the unity of laws (e.g., Carnap, 1955). Recently, the need of the data-driven research may transform the research landscape. Data-drive research allows scientists and scholars to collaborate on the same data sets and apply their own expertise. This juxtaposition of expertise over data has progressively changed the characteristics of science—scientists and scholars no longer need to collaborate with others who share the same expertise but team up with those with diversified expertise toward certain problem-solving (e.g., Wuchty, Jones, & Uzzi, 2007). This mode in turn facilitates more diversified tangible (i.e., publications and citations) and intangible (i.e., informal knowledge sharing) knowledge transfer.

R&D expenditures and knowledge production

UNESCO publishes economic indices that allow us to compare R&D expenditures with citation statistics. Table 4 lists the world total R&D expenditures in billions PPP$ for years 2002, 2007, and 2009. These three time points correspond to the three citation windows of this study: 2000/2002, 2006/2008, and 2009/2011.

Table 4. Global R&D expenditures and number of citations

Year	R&D expenditures (in billions PPP$)*	Increase in percentage	Citation window	No. of subject-level citations	Increase in percentage
2002	787.7	-	2000/2002	1,623,863	-
2007	1,155.4	46.68%	2006/2008	2,045,946	25.99%
2009	1,276.9	10.52%	2009/2011	2,167,594	5.95%

* PPP (Purchasing Power Parity) conversion factor (local currency per international $): World Bank; World Development Indicators, as of September 2011 and UNESCO Institute for Statistics (UIS) estimations.

The R&D expenditures have increased by 46.68% from 2002 to 2007 and 10.52% from 2007 and 2009; in the meantime, the number of citations for all disciplines has increased by 25.99% from 2000/2002 to 2006/2008 and 5.95% from 2006/2008 to 2009/2011. The increment of R&D expenditures is seemingly proportioned to the increment of citations. Here, we do not intend to make inaccurate causation inferences and we acknowledge the delay of R&D expenditures on actual knowledge creation. Nonetheless, the results do indicate that growth rates of R&D expenditures are commensurate with the growth rates of knowledge production in the form of citations. The citations numbers in turn may be explained by the growing number of publications.

At the discipline level, different disciplines exhibit varied “fundability” characteristics: while for some, funding is indispensable, but for the others, the importance of funding ranges from essential to desirable. In addition, the output of funded research may not relate directly to the investment inputs. Science and innovation maintain a nonlinear relationship in nature and thus the output of funded research can vary substantially between disciplines (Lane, 2009). This nonlinear feature calls for new data to characterize disciplinary differences and new methods to capture varied forms of knowledge production and transfer.

While the statistics of global R&D expenditures are reported by UNESCO, detailed discipline-level investment data are only available for some countries, largely attributed to the varied science classification schemes that different countries adopted. This made it difficult to synthesize country-level numbers to the global level. As an alternative, in this study, discipline-level statistics of U.S. R&D expenditures at universities and colleges were used to proximate global statistics, because of U.S.’s highest share of R&D expenditures among all countries—30.63% as of 2009 (UNESCO, 2011) and also because of its comprehensive investments on all disciplines. One should be reminded that despite this effort, discrepancies may exist between the U.S. and the globe in disciplinary funding allocations.

The U.S. Census Bureau reports R&D expenditures in the following disciplines: physical sciences, environmental sciences, mathematical sciences, computer sciences, life sciences, psychology, social sciences, engineering, and other sciences (U.S. Census Bureau, 2012). Some of them match the major subjects of this study and some match the top-level domains. For instance, environmental sciences, mathematical sciences, computer sciences, and engineering were reported separately with physical sciences in the Census Bureau statistics whereas they were all categorized into Physical Sciences in ASJC. Thus, for a fair comparison, incoming citations of the corresponding major subjects were excluded from the top-level domain of Physical sciences. We chose R&D expenditures reported at two time points: 2000 and 2009 which correspond to the two citation windows of this study, as shown in Table 5.

Table 5. U.S. R&D expenditures at universities and colleges (in million$) and number of citations for several disciplines

Disciplines	2000 (2000/2002)		2009 (2009/2011)	
	R&D	Citations	R&D	Citations
Physical sciences*	2,713 (10.63%)	340,259 (20.95%)	4,294 (9.28%)	586,874 (27.07%)
Environmental sciences	1,766 (6.92%)	28,598 (1.76%)	2,940 (6.35%)	73,599 (3.40%)
Mathematical sciences	342 (1.34%)	16,558 (1.02%)	553 (1.19%)	36,685 (1.69%)
Computer sciences	877 (3.44%)	12,828 (0.79%)	1,592 (3.44%)	28,202 (1.30%)
Life sciences**	17,471 (68.44%)	1,116,147 (68.73%)	32,791 (70.85%)	1,243,247 (57.36%)
Psychology	517 (2.03%)	17,407 (1.07%)	979 (2.12%)	33,977 (1.57%)
Social sciences***	1,300 (5.09%)	33,991 (2.09%)	2,075 (4.48%)	65,791 (3.04%)
Engineering	543 (2.13%)	36,954 (2.28%)	1,060 (2.29%)	80,802 (3.73%)
Total	25,529	1,623,863	46,284	2,167,594

* Excluding the incoming citations of Environmental science, Mathematics, Computer science, and Engineering.

** Combining the incoming citations of Life Sciences and Health Sciences.

*** Excluding the incoming citations of Psychology.

Through this comparison, the following observations can be made. First, a few disciplines’ shares of citations exceeded their share of R&D expenditures, including physical sciences and engineering. In the meantime, for environmental sciences and computer sciences, their shares of R&D expenditures exceeded their share of citations. Shares of citations of mathematics, life sciences, psychology, and social sciences roughly matched with their shares of R&D expenditures. Dynamically, shares of R&D expenditures of different disciplines remained relatively stable; the share of citations of life sciences, however, decreased from 68% to 57% while shares of other disciplines increased.

The results presented here are part of a larger effort to understanding the impact of R&D investments on knowledge production and diffusion. In addition to knowledge production in the form of publications and citations, there are other intangible forms, such as training students and researchers and organizing workshops (e.g., Cohen, Nelson, & Walsh, 2002). To systematically evaluate the impact of R&D investments, one also needs to leverage its outcome on the market (e.g., new drugs and new tools been developed) as well as on the societal institutions of people's day-to-day life (Lam, 2000).

Conclusion

This study has examined the patterns of disciplinary knowledge production and diffusion through a comprehensive citation data set of Scopus indexed journals. A three-step approach has been adopted: the first step has involved the examination of disciplines' citation characteristics through scientific trading dimensions; the second step has analyzed citation flows between two disciplines; and the third step has used ego-centric citation networks to assess individual disciplines' citation flow diversity through Shannon entropy.

This study has found that except for General and Biochemistry, all other disciplines received more scientific impact during the past five citation windows; in particular, Chemical Engineering, Energy, and Environmental Science have the fastest growths. Measured through cited/citing ratios, Decision Sciences, Economics, Chemical Engineering, and Biochemistry are noticeable knowledge exporters. Meanwhile, most subjects retained high self-citation ratios, among which Medicine, Dentistry, and Earth and Planetary Sciences had the highest self-citation ratios (above 0.6). Through an investigation of disciplinary citation flows, the study has found that while weaker citation flows gained faster growths, more established citation flows are less susceptible to change. Disciplines such as Energy, Environmental Science, Chemistry, and Computer Science are the ones that strengthened their citation flows with other subjects.

Measured through Shannon entropy, this study has revealed that Environmental Science, Engineering, Decision Sciences, and Arts and Humanities are the most diversified and interdisciplinary; Dentistry, Earth and Planetary Sciences, Medicine, Physics, and Veterinary, on the other hand, are the least diversified. Except for Computer Science, Energy, and Health Professions, all other subjects gained entropy and are thus committed to more interdisciplinary trading practices. Results from Shannon entropy have also suggested that although disciplines became more interdisciplinary-oriented at the subject-level, cross top-domain level knowledge transfer was less practiced.

By matching the disciplinary citation data with R&D expenditures, the study has shown that the growth rates of disciplinary citations align with the growth rates of global R&D expenditures, thus providing evidences to support the impact of R&D expenditures on knowledge production.

The analytical units of this study include individual disciplines, pairs of disciplines, and ego-centric networks of disciplines. The next step in this direction will examine each discipline in the context of a flow network and using network theories and methods to capture the more latent “linked” aspect of disciplinary knowledge production and diffusion. Additionally, studies will also benefit from cross-referencing data sources on science policy, research communities, and other social-technical factors to understand the mechanisms that may lead to the dynamics changes of disciplines.

Acknowledgement

The data set used in this paper is supported by the Elsevier Bibliometric Research Program (EBRP).

References

Adams, J., & Griliches, Z. (1996). Measuring science: An exploration. *Proceedings of the National Academy of Sciences*, *93*(23), 12664-12670.

Azoulay, P., Ding, W., & Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *Journal of Industrial Economics, 57*(4), 637-676.

Bacchiocchi, E., & Montobbio, F. (2009). Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations. *The Journal of Technology Transfer*, *34*(2), 169-181.

Belenzon, S., & Schankerman, M. (2013). Spreading the word: Geography, policy, and knowledge spillovers. *Review of Economics and Statistics*, *95*(3), 884-903.

Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science, 43*(6), 397-411.

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

Börner, K., Penumarthy, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major U.S. research institutions. *Scientometrics, 68*(3), 415-426.

Boschma, R.A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61-74.

Boyack, K. W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, *79*(1), 45-60.

Boyack, K. W., Patek, M., Ungar, L. H., Yoon, P., & Klavans, R. (2014). Classification of individual articles from all of science by research level. *Journal of Informetrics*, *8*(1), 1-12.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flow. *Journal of Economic Geography, 9*(4), 439-468.

Bresman, H., Birkinshaw, J., & Nobel, R. (2010). Knowledge transfer in international acquisitions. *Journal of International Business Studies*, *41*(1), 5-20.

Carnap, R. (1955). Logical foundations of the unity of science. In O. Neurath, R. Camap, and C. Morris (Eds.), *International Encyclopedia of Unified Science* (*Reprint*, Vol. 1, pp. 42-62). Chicago: University of Chicago Press.

Cassi, L., & Plunket, A. (2014). Proximity, network formation and inventive performance: in search of the proximity paradox. *The Annals of Regional Science*, *53*(2), 395-422.

Chen, C., & Hicks, D. (2004). Tracing knowledge diffusion. *Scientometrics*, *59*(2), 199-211.

Chen, C., Zhang, J., & Vogeley, M. S. (2009). Visual analysis of scientific discoveries and knowledge diffusion. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics* (ISSI 2009). Rio de Janeiro, Brazil, July 14-17, 2009. pp. 874-885.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: The Influence of public research on industrial R&D. (In A. Geuna, A. J. Salter, & W. E. Steinmueller, Eds.) *Management Science*, *48*(1), 1-23.

Cole, S. (1983). The hierarchy of the sciences? *American Journal of Sociology, 89*(1), 111-139.

Contractor, F. J., & Ra, W. (2002). How knowledge attributes influence alliance governance choices: a theory development note. *Journal of International Management*, *8*(1), 11-27.

Cronin, B., & Meho, L. I. (2008). The shifting balance of intellectual trade in information studies. *Journal of the American Society for Information Science & Technology*, *59*(4), 551-564.

David, P. A. (2000). Understanding digital technology's evolution and the path of measured productivity growth: present and future in the mirror of the past. In E. Brynjolfsson & B. Kahin (Eds.), *Understanding the digital economy* (pp. 49-95). Cambridge, MA: MIT Press.

David, P. A., & Foray, D. (2002). An introduction to the economy of the knowledge society. *International social science journal*, *54*(171), 9-23.

de Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, *73*(1), 53-78.

Fabrizio, K. R., & Di Minin, A. (2008). Commercializing the laboratory: Faculty patenting and the open science environment. *Research Policy, 37*(5), 914-931.

Fier, H., & Pyka, A. (2014). Against the one-way-street: analyzing knowledge transfer from industry to science. *The Journal of Technology Transfer*, *39*(2), 219-246.

Gao, X., & Guan, J. (2012). Network model of knowledge diffusion. *Scientometrics*, *90*(3), 749-762.

Garfield, E & Pudovkin, A. I, & Istomin, V. S. (2002). Algorithmic citation-linked historiography: Mapping the literature of science. *Proceedings of the American Society for Information Science and Technology Annual Meeting, 39*(1), 14-24.

Geroski, P. A. (2000). Models of technology diffusion. *Research policy*, *29*(4), 603-625.

Grueber, M. (2011). 2012 global R&D funding forecast. Advantage Business media. Retrieved September 9, 2013 from: http://www.rdmag.com/articles/2011/12/2012-global-r-d-funding-forecast

Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. *SIAM Review, 42*(4), 599-653.

Holton, G. (1978). Can Science Be measured? In *Scientific Imaginations: Case Studies* (pp. 199-228). Cambridge, UK: Cambridge University Press.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, *90*(2), 215-218.

Jaffe, A. B., Trajtenberg, M., & Henderson, A. D. (1993). Geographical localization of knowledge spillovers by patent citations. *Quarterly Journal of Economics, 108*(3), 577-599.

Kiss, I. Z., Broom, M., Craze, P. G., & Rafols, I. (2010). Can epidemic models describe the diffusion of topics across disciplines? *Journal of Informetrics*, *4*(1), 74-82.

Klavans, R., & Boyack, K. W. (2007). Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases. In *Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics* (Vol. 1, pp. 437-448).

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice*. Detroit, MI: Wayne State University Press.

Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities*. Charlottesville, VA: University of Virginia Press.

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Ko, N., Yoon, J., & Seo, W. (2014). Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. *Expert Systems with Applications, 41*(4), 1955-1963.

Lam, A. (2000). Tacit knowledge, organizational learning and societal institutions: an integrated framework. *Organization studies*, *21*(3), 487-513.

Lane, J. (2009). Assessing the impact of science funding. *Science*, *324*(5932), 1273-1275.

Levitt, J. M., Thelwall, M., & Oppenheim, C. (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. *Journal of the American Society for Information Science and Technology*, *62*(6), 1118-1129.

Leydesdorff, L. (2011). “Structuration” by intellectual organization: The configuration of knowledge in relations among structural components in networks of science. *Scientometrics, 88*(2), 499-520.

Leydesdorff, L., & Probst, C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. *Journal of the American Society for Information Science and Technology, 60*(8), 1709-1718.

Leydesdorff, L., de Moya-Anegón, F., & de Nooy, W. (2014). Aggregated journal-journal citation relations in Scopus and Web-of-Science matched and compared in terms of networks, maps, and interactive overlays. *arXiv preprint arXiv:1404.2505*.

Leydesdorff, L., de Moya-Anegón, F., & Guerrero-Bote, V. P. (2010). Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, *61*(2), 352-369.

Li, Y. A. (2014). Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age?-Evidence from Patent Citations. *European Economic Review, 71*, 152-172.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *Information Theory, IEEE Transactions on*, *37*(1), 145-151.

Liu, X., Jiang, S., Chen, H., Larson, C. A., & Roco, M. C. (2014). Nanotechnology knowledge diffusion: measuring the impact of the research networking and a strategy for improvement. *Journal of Nanoparticle Research*, *16*(9), 1-15.

Lockett, A., & McWilliams, A. (2005). The balance of trade between disciplines: do we effectively manage knowledge? *Journal of Management Inquiry, 14*(2), 139-150.

MacGarvie, M. (2005). The determinants of international knowledge diffusion as measured by patent citations. *Economics Letters, 87*(1), 121-126.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Mu, J., Tang, F., & MacLachlan, D. L. (2010). Absorptive and disseminative capacity: Knowledge transfer in intra-organization networks. *Expert Systems with Applications*, *37*(1), 31-38.

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy*, *26*(3), 317-330.

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American society for information science*, *27*(1), 25-45.

Naumis, G. G., & Phillips, J. C. (2012). Diffusion of knowledge and globalization in the web of twentieth century science. *Physica A: Statistical Mechanics and its Applications*, *391*(15), 3995-4003.

Neurath, O. (1996). Unified science as encyclopedic integration. In S. Sarkar (ed.), *Logical Empiricism at Its Peak: Schlick, Carnap, and Neurath*, (pp. 309-335). Boston, MA: Harvard University.

Nomaler, Ö., & Verspagen, B. (2008). Knowledge flows, patent citations and the impact of science on technology. *Economic Systems Research*, *20*(4), 339-366.

Parent, O., & LeSage, J. P. (2012). Determinants of knowledge production and their effects on regional economic growth. *Journal of Regional Science*, *52*(2), 256-284.

Rinia, E. J., Van Leeuwen, T. N., & Bruins, E. E. W. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics, 51*(1), 293-309.

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management science*, *59*(2), 504-525.

Rosas, S. R., Schouten, J. T., Cope, M. T., & Kagan, J. M. (2013). Modeling the dissemination and uptake of clinical trials results. *Research Evaluation, 22*(3), 179-186.

Schubert, A., & Glänzel, W. (1986). Mean response time: A new indicator of journal citation speed with application to physics journals. *Czechoslovak Journal of Physics*, *36*(1), 121-125.

Shi, X., Tseng, B., & Adamic, L. (2009). Information diffusion in computer science citation networks. *Proceedings of the International Conference on Weblogs and Social Media ICWSM 2009*, 319-322.

Singh, J., & Marx, M. (2013). Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. *Management Science*, *59*(9), 2056-2078.

Stigler, S. M. (1994). Citation patterns in the journals of statistics and probability. *Statistical Science, 9*(1), 94-108.

U.S. Census Bureau (2012). Science & technology: Expenditures, research development. Retrieved September 9, 2013 from http://www.census.gov/compendia/statab/cats/science_technology/expenditures_research_development.html

UNESCO (2011). Regional totals for R&D expenditure (GERD) and researchers, 2002, 2007 and 2009. Retrieved September 9, 2013 from http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx?IF_ActivePath=P,54

Van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: Analysis of cross-disciplinary citation flows. *Research Evaluation*, *9*(3), 183-187.

Vitanov, N. K., & Ausloos, M. R. (2012). Knowledge epidemics and population dynamics models for describing idea diffusion. In *Models of Science Dynamics* (pp. 69-125). Springer Berlin Heidelberg.

Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392.

Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2010). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37-47.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Xiao, Y., Lu, L. Y., Liu, J. S., & Zhou, Z. (2014). Knowledge diffusion path analysis of data quality literature: A main path analysis. *Journal of Informetrics*, *8*(3), 594-605.

Yan, E. (2014). Finding knowledge paths among scientific disciplines. *Journal of the Association for Information Science and Technology*, *65*(11), 2331-2347.

Yan, E., & Sugimoto, C. R. (2011). Institutional interactions: Exploring social, cognitive, and geographic relationships between institutions as demonstrated through citation networks. *Journal of the American Society for Information Science and Technology, 62*(8), 1498-1514.

Yan, E., Ding, Y., Cronin, B., & Leydesdorff, L. (2013). A bird's-eye view of scientific trading: Dependency relations among fields of science. *Journal of Informetrics, 7*(2), 249-264.

Yu, F., & Wu, Y. (2014). Patent citations and knowledge spillovers: an analysis of Chinese patents registered in the USA. *Asian Journal of Technology Innovation*, *22*(1), 86-99.

Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics, 4*(2), 185-193.

Zhao, R. Y., & Wu, S. N. (2014). The network pattern of journal knowledge transfer in library and information science in China. *Knowledge Organization, 41*(4), 276-287.

Zuckerman, H. A., & Merton, R. K. (1973). Age, aging, and age structure in science. In *The Sociology of Science*, *by R. K. Merton* (pp. 497-559). Chicago: University of Chicago Press.