

Measuring the Diffusion of an Innovation: A Citation Analysis

Yujia Zhai

Department of Information Resource Management, The Business School, Nankai University, 94 Weijin Road, Nankai District, Tianjin 300071, China.

Telephone: (571) 342-1377

E-mail: zhaiyujiachn@gmail.com

Ying Ding

School of Informatics & Computing, Indiana University, Bloomington, IN, 47405, USA;

School of Information Management, Wuhan University, Wuhan, Hubei 430072, China;

University Library, Tongji University, Shanghai 200092, China.

Telephone: (812) 855-5388

E-mail: dingying@indiana.edu

Fang Wang (*Corresponding Author*)

Department of Information Resource Management, The Business School, Nankai University, 94 Weijin Road, Nankai District, Tianjin 300071, China;

Center for Network Society Governance, The Business School, Nankai University, 94 Weijin Road, Nankai District, Tianjin 300071, China.

Telephone: (86-22) 2350-2800

E-mail: wangfangnk@nankai.edu.cn

Measuring the Diffusion of an Innovation: A Citation Analysis

Abstract

Innovations transform our research traditions and become the driving force to advance individual, group, and social creativity. Meanwhile, interdisciplinary research is increasingly being promoted as a route to advance the complex challenges we face as a society. In this paper, we use Latent Dirichlet Allocation (LDA) citation as a proxy context for the diffusion of an innovation. With an analysis of topic evolution, we divide the diffusion process into five stages: testing and evaluation, implementation, improvement, extending, and fading. Through a correlation analysis of topic and subject, we show the application of LDA in different subjects. We also reveal the cross-boundary diffusion between different subjects based on the analysis of the interdisciplinary studies. Results show that as LDA is transferred into different areas, the adoption of each subject is relatively adjacent to those with similar research interests. Our findings further support researchers' understanding of the impact formation of innovation.

Introduction

An innovation is defined as an idea, practice, or object that is perceived as having new values by an individual or other unit of adoption (E. M. Rogers, 1962). Innovations are complex, uncertain, disorderly, and subject to change (Kline & Rosenberg, 1986). Widely seen as the driving force of economic growth and social creativity, these advancements are a central plank of national and local policies and consume billions of dollars of investment worldwide. In both academic and practitioner communities, it is commonly perceived that organizations should innovate to be effective, or even survive.

Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system (E. M. Rogers, 1962). The value of innovation can only be reflected in the adoption and diffusion processes. For scientific research, diffusion of innovations is particularly important. Because scientific innovation is very difficult and time consuming, such as the invention of drugs requires a combination of many skills and a long period of repeated experiments. Without effective adoption and diffusion, research achievements will lose their value and do not fully come into play. So understanding the diffusion of innovations is central to promoting new ideas, bringing necessary progress, and staying competitive.

When an innovation has obvious advantages, it is not always diffused and adopted rapidly, however. For example, the diffusion of hybrid seed corn in Iowa led to an agricultural revolution in farm productivity during the 1930s to 1950s. But only two of 259 farmers in one study (Ryan & Gross, 1943) had adopted hybrid corn during this period (1928 to 1941). If innovations can be adopted sooner and diffuse faster, our economic conditions and living environments may improve dramatically. In order to foster promotion and strengthen competitiveness, it becomes important to understand the diffusion of innovations.

Citations of academic publications document the diffusion process and trajectory of innovation. As Latour (1987) argues that citations from a paper reinforce its arguments and connect it to an intellectual lineage. It also may indicate existing knowledge upon which the current publication builds (Cole, 2000). Although scholars have continually investigated questions about the diffusion of innovations, such as the process of diffusions (Chatterjee & Eliashberg, 1990; Hagerstrand, 1967), the rate of idea adoption (Abrahamson & Rosenkopf, 1997; Mintrom & Vergari, 1998), and the decision-making process in the diffusion of innovations (E. M. Rogers & Shoemaker, 1971), most data are collected via semi-structured interviews and surveys used for descriptive purposes, which is a limited approach that has bias. With large-scale citation data, we can quantitatively analyze the scientific diffusion of innovation, where publications and citations provide especially good footprints to trace the pathway of scientific and technical progress.

Cross-disciplinary knowledge flows are often constructed when one discipline cites another. A citation issued by a biology research article to a chemistry field work can be defined as a cross-disciplinary citation. This kind of citations contains a wealth of detailed empirical data on the diffusion of knowledge. As Desouza (2009) argues that diffusion and implementation of innovations require knowledge and the ability to apply that knowledge in novel ways and across a variety of disciplines. For instance, Latent Dirichlet Allocation (LDA) is a topic model and was first proposed in Computer Science field in 2003 (Blei, Ng, & Jordan, 2003). When it is introduced into other disciplines, the original algorithm can be upgraded and reconstructed into new models (Wang et al., 2011) or be used for solving a new problem (Au Yeung & Jatowt, 2011). The exchanging and recombination process of concepts in these cross-disciplinary knowledge flows show us how innovative ideas and technologies from different areas can be adopted, integrated, and ameliorated.

In this paper, we quantitatively measure the diffusion of a particular innovation, building on previous work that defines the trail of progress in terms of citations. With a case study of the evolution of LDA, we demonstrate the sequence and progress of the growth of innovation, explore the patterns of applying or improving innovative ideas, and reveal the cross-boundary knowledge flow between different subjects.

Literature Review

Diffusion process of innovations

Rogers popularized the diffusion of innovations theory, where he synthesized studies from over 500 diffusion studies across numerous fields in his seminal book (1962). During 1970s, the theoretical model gradually integrated the organizational changes and innovation diffusion. The Stage theory model, which suggests that innovation adoption goes through various stages that require discipline-specific strategies, became the primary model to describe the phenomenon of innovation diffusion (Cooper & Zmud, 1990; Kwon & Zmud, 1987; Lewin, 1947; McFarlan, McKenney, & Pyburn, 1983). These Stage models tend to view this process as linear and sequential, whereas in practice it is more likely to be iterative and recursive (Saren, 1984). For example, Rogers (1962) proposed five stages, including agenda-setting, matching, redefining/restructuring, clarifying, and routinizing, where later stages in the innovation

process cannot be undertaken until earlier stages have been completed. Kwon and Zmud (1987) proposed that IT implementation follows six-stages, including initiation, adoption, adaptation, acceptance, routinization and infusion. There are also many studies attempt to discover which stages in the development process or features of a new product are most critical to achieving market success and wide adoption (Henard & Szymanski, 2001). In a recent study, Yasir Mehmood et al. (2016) proposed a stochastic framework for modeling user adoption and different stages of innovation diffusion. Continuing with Roger's theory, their model explained the propensity of new projects to spread through the population and the rate at which adoption occurs at various stages.

Meanwhile, mathematical modeling of innovation diffusion researches confirmed the existence of a statistical bell shaped curve for the frequency of adoption plotted against time, and a S-shaped curve for the cumulative number of adopters. Several diffusion models have been proposed to study the diffusion phenomenon, such as, the internal logistic curve influence model (Mahajan & Muller, 1979), the internal influence model (Mansfield, 1961), and the Bass model (Bass, 1969). The Bass model characterizes the diffusion of an innovation as a contagious process that is initiated by mass communication and propelled by word-of-mouth. The model is widely used in market analysis and demand forecasting of innovation diffusion in various areas.

With the increase of social network data, diffusion researches have moved their focus onto the relationship between the entities that are diffused (Goyal, Heidari, & Kearns, 2013; J. Tang, Sun, Wang, & Yang, 2009; Weng, Flammini, Vespignani, & Menczer, 2012), as well as the role of individual users through the diffusion process of innovation in a network (Backstrom, L., Huttenlocher, D., Kleinberg & Lan, 2006; Rong & Mei, 2013; Ugander, Backstrom, Marlow, & Kleinberg, 2012). For example, Montanaria and Saberib (2010) found that innovation spreads quickly in locally connected networks and high-degree nodes slow down the diffusion process. Rong and Mei (2013) regard algorithms in computer science as the nodes of an innovation network and citations as links to study two different interrelationships among innovations (competition and collaboration) affect users' adoption behavior. The limitation of these studies is that they cannot reveal the evolution process of the innovation only by modeling the cumulative number of adoption.

Cross-disciplinary knowledge flows

Cross-disciplinary knowledge flows are specifically associated with the co-occurrence of innovation diffusion (e.g. some of the domain-specific methods are introduced into new disciplines, original concepts are reassembled and new knowledge is generated). As argued by Kusiak (2016, p. 255), "A building block of innovation science is connecting seemingly unrelated ideas. We are flooded with discoveries in isolated domains. Making quick connections between, for instance, biology and technology, could lead to bigger ideas and redirect research and development."

Generally considered as the indicator of knowledge flow, citations are often adopted to examine patterns of dynamic disciplinary knowledge production and diffusion (Cronin & Meho, 2008; Kiss, Broom, & Rafols, 2009; Levitt, Thelwall, & Oppenheim, 2011; Yan & Yu, 2015; Zhao & Wu, 2014). Using citation analysis, studies have found that scientific works in

one discipline tended to cite publications from adjacent disciplines (Leeuwen & Tijssen, 2000) and citations to publications of the own discipline occurred sooner than citations to papers in other disciplines (Rinia, Leeuwen, Bruins, & Vuren, 2001). When looking at specific areas of research, it is found that a few library and information science journals heavily cited communication science journals (Borgman & Rice, 1992) and journal knowledge flows in library and information science is frequent (Zhao & Wu, 2014). Similarly, Leydesdorff and Probst (2009) revealed that communication science journals have a strong connection with political science and social psychology journals.

Another thread of citation based knowledge diffusion studies use citation data to explore the knowledge path cross various disciplines. For example, using part of the Journal Citation Index data from 1969, Narin et al. (1972) proposed a cross-field model that utilizes the relationships between journal citations. They found that the fields of science and nature function as a link for physics and biology, and there is a knowledge path through the fields of biology – biochemistry – chemistry – physics – mathematics. Rorissa and Yuan (2012) use citation data for 10 years (2000–2009) and find top five disciplines that contribute to information retrieval are computer science, library and information science, engineering, telecommunications, and management. Yan and Yu (2015) built a discipline-level citation network based a journal-to-journal citation matrix for all journals and proceedings indexed in the Scopus database with a 2-year citation window. They used MST (Maximum Spanning Tree) algorithm to find knowledge paths and found that Medicine served as the largest exporter of knowledge and several STEM connected paths (e.g. Medicine - Biochemistry – Agriculture Sciences – Environmental Science and Medicine – Biochemistry – Chemistry – Materials Science – Physics – Earth and Planetary Science).

Citation-based innovation diffusion

Based on the assumption that authors cite the works that influence them, some studies have specially used citation as a proxy for innovation diffusion. Jaffe and Trajtenberg (1996), for example, use citations of all United States patents granted since 1963 to measure the diffusion of knowledge on the geographic portrait. The results show that patents granted to United States inventors are much more likely to cite previous United States patents than are patents granted to inventors of other countries. Mowery and Ziedonis (2002) interpret patent citations as measures of the importance of the contribution to inventive knowledge and find higher citation rates among patents originating from university labs.

Indeed, citations cannot represent directly adoption of innovation and scientists have pointed to a number of concerns about citation analysis, including biases in citation patterns, motivations to cite that extend beyond intellectual influence, wrong or misleading citations, variation between specialties, and authors' ignorance of some relevant literature (MacRoberts & MacRoberts, 1996). Edge (1979) also maintains that publication citations only capture the influence of other formal publications, ignoring informal communications and tacit knowledge that may be far more important as influences. However, the data from citation-based approaches still have much more advantages for innovation diffusion research comparing with the data gathered from conventional approaches, including interviews, questionnaires, and in-depth case studies, and also social network data.

First of all, citations are with high quality and credibility after peer review which means

the publications get admitted by the scientific community. Second, citations are easily accessible in electronic form for revealing the content of the diffusion process. As Small (1978) argues that citations are considered as part of the cognitive process of producing written discourse. Moreover, publications are linked to inventiveness and contain a trace of what knowledge they build upon through the citation of prior art. Last of all, the traditional method can only collect data for a period of time with a small number of sample. But citation is a cumulative number and the duration is very long, so we can map the entire diffusion process of the innovation.

Citations have become intellectual linkages across academic and professional disciplines and can be used to study the nature and the development of different domains (Zhang, Ding, & Milojević, 2013). Tang (2004) use citation analysis to study the scholarship maturity of LIS and find that LIS is a highly interdisciplinary field that exchange knowledge with a variety of disciplines from the domains of science, social science, and the humanities. They also find identified the most common disciplines to which LIS exports ideas: computer science, communication, education, management, business, and engineering. Consisted with this research, Cronin and Meho (2008) also found IS has become successful exporter of ideas. By analyzing interdisciplinary bridges between pairs of disciplines. Levitt et al. (2011) found that library and information science grew the fastest in interdisciplinarity between 1990 and 2000 among all social science fields. Yan (2015) found that the subjects of chemical engineering, energy, and environmental science have the fastest growth.

In summary, previous studies have been more concerned about offering a quantitative proxy of the diffusion of innovations and generally do not consider how innovations are adapted and improved over time. In this study, we choose LDA as a research instance and reveal the diffusion process of LDA by analyzing its citation history. The citations we identify not only show the trajectory of the innovative idea diffusing and evolving, but also reveal the knowledge flows across disciplines.

Methodology

Dataset

Latent Dirichlet Allocation (LDA) is a generative probabilistic model used for clustering sets of discrete data (Blei, Ng, & Jordan, 2003). Due to its scalability and meaningful results, LDA has become an important tool for scientific research as well as many academic and business fields, such as sentiment analysis (Pang & Lee, 2008), image retrieval (Hare, Lewis, Enser, & Sandom, 2006), and social network analysis (Mccallum, Wang, & Corrada-Emmanuel, 2005). According to Scopus, the citation of LDA has reached more than 6,800 in 2015. Because of the large number of citation and a wide range of application types, we choose LDA's citations in the case study.

Although Google Scholar tends to have the most comprehensive data on journals and proceedings, it can't show the full dataset because of its search mechanism. Only part of the citation data can be achieved and the quality of the dataset is not carefully controlled like commercial databases. Scopus includes a more expanded spectrum of journals than PubMed and Web of Science (Falagas, Pitsouni, Malietzis, & Pappas, 2008; Klavans & Boyack, 2009;

L Leydesdorff & Moya-Anegón, 2014; Meho & Yang, 2007). The search result of citation number of LDA is 6,822 in Scopus and only 3,797 in Web of Science between 2003 and 2015. So in this study, we use LDA citation extracted from Scopus. The citations are taken from articles, review articles, and proceeding papers published between 2003 and 2015, and the metadata contains title, abstract, year, authors, and keywords.

Academic disciplines are simply particular branches of knowledge and taken together they form the whole or unity of knowledge that has been created by the scientific endeavor (Krishnan & Krishnan, 2009). Many studies choose to use Subject Categories to represent scientific disciplines and to study knowledge flow and diffusion (L Leydesdorff & Rafols, 2009; Rafols & Leydesdorff, 2009; Yan, 2015a). Scopus uses journal-to-journal citation data to predefine its own journal classification schema—ASJC (All Science Journals Classification). According to the journal’s assigned subject, a paper is typically associated with one or more subjects and these, in turn, are grouped into one of the four subject areas: Life Sciences, Social Sciences, Physical Science, and Health Sciences. In this study, we use subject as a proxy as the discipline and focus on the analysis of the diffusion process of LDA.

TABLE 1. LDA’s Citation Distribution in each Subject Area

| Subject Areas | Subject | Number of Citations | Subject Areas | Subject | Number of Citations |
|--------------------------|------------------------------|---------------------|--------------------------------------------|----------------------------------------------|---------------------|
| Physical Sciences | Computer Science | 5576 | Social Sciences | Social Sciences | 638 |
| | Engineering | 1374 | | Decision Sciences | 365 |
| | Mathematics | 1320 | | Arts and Humanities | 326 |
| | Physics and Astronomy | 105 | | Business, Management and Accounting | 250 |
| | Earth and Planetary Sciences | 72 | | Psychology | 55 |
| | Material Science | 46 | | Economics, Econometrics and Finance | 19 |
| | Environmental Science | 25 | Life Sciences | Biochemistry, Genetics and Molecular Biology | 131 |
| | Chemical Engineering | 8 | | Neuroscience | 111 |
| | Chemistry | 8 | | Agricultural and Biological Sciences | 55 |
| | Energy | 3 | | Immunology and Microbiology | 8 |
| Health Sciences | Medicine | 209 | Pharmacology, Toxicology and Pharmaceutics | 4 | |
| | Health Professions | 50 | Multidisciplinary | 36 | |
| | Nursing | 4 | | | |

(a) Subject Area Classification by Scopus

| | | | | | | | | |
|--------------------|------|------|-----|----|----|---|---|-------|
| numbers of subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| Number of papers | 3758 | 2395 | 502 | 99 | 61 | 6 | 1 | 6822 |

(b) Number of papers (row 2) assigned to different numbers of subject (row 1)

There are a total of 6,822 citations that cited LDA from 2003 to 2015. These citations have been assigned to 25 subjects according to Scopus’ ASJC (All Science Journals Classification). The largest subject is Computer Science with 5,576 articles, followed by Engineering. Energy gets the smallest number of papers, with three articles.

The numbers of papers that are associated with different numbers of subjects are illustrated in Table 1(b). Unlike the subject area “multidisciplinary” which is predefined by Scopus, in our study, an interdisciplinary study is considered as a paper signed into two or more subjects. The numbers of papers that are associated with different numbers of subjects are illustrated in Table 1(b). It shows that up to 55 percent (3758/6822) of papers are associated with one major subject and up to 45 percent (3064/6822) are interdisciplinary studies. One paper, “Negative Example Selection for Protein Function Prediction: The NoGO Database” (Youngs, Penfold-Brown, Bonneau, & Shasha, 2014, published by PLoS Computational Biology (Youngs, Penfold-Brown, Bonneau, & Shasha, 2014), is associated with seven subjects: Computer Science, Mathematics, Medicine, Biochemistry, Genetics and Molecular Biology, Neuroscience, Agricultural and Biological Sciences, and Environmental Science.

To illustrate the relationships between different subjects, we build a two-mode network based on links of paper-subject (Figure 1) consisting of 6,822 articles and 25 subjects. Each article belongs to one or more subjects, where the size of the label for each subjects is determined by the number of associated articles. For the visualizations, we use Gephi (M. Bastian, S. Heymann, 2009) to construct the network and apply Hu’s (2005) layout method to map the network structure.

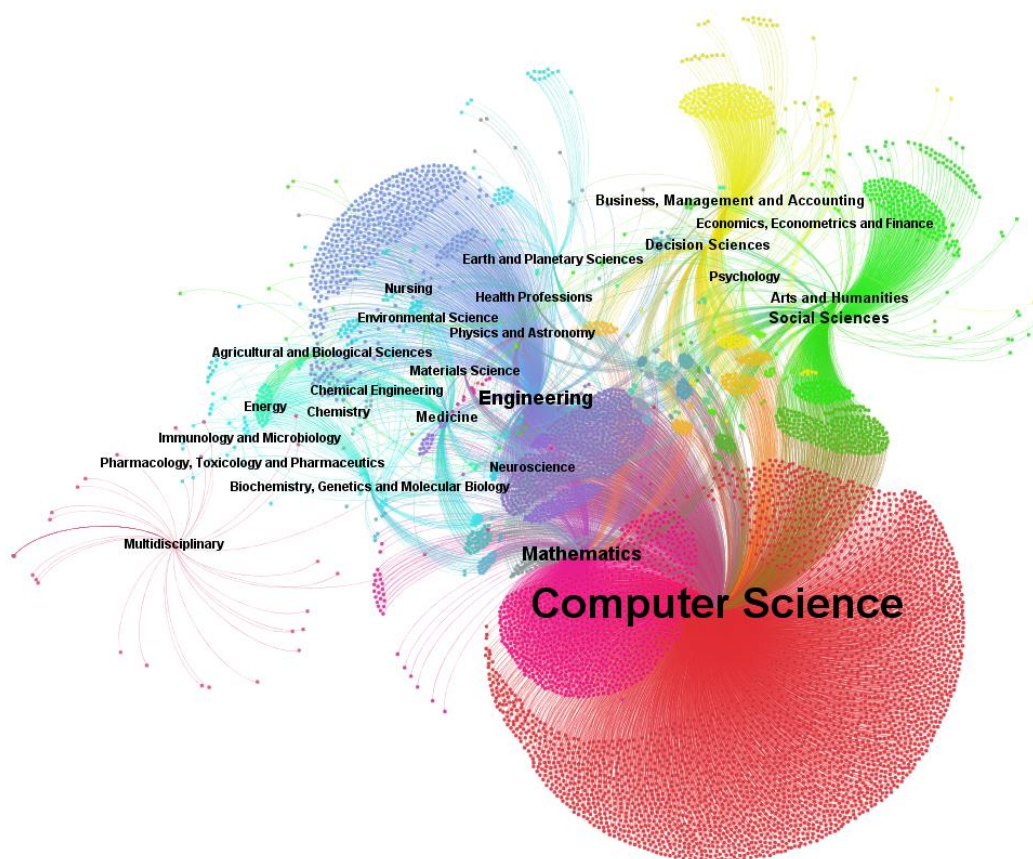


FIG. 1. Paper-Subject Network for LDA Citations

Methods

To reveal the diffusion process of LDA, a three-step approach is employed. The first step

examines the topic evolution in the whole citation history of LDA to answer questions about which research topic attracts the most attention and what topics have been proposed during the process of adopting LDA. In order to understand the chronological order of the adoption of LDA in different subjects, the second step builds the subject-topic matrix and then extracts the most relevant topics for each subject. The third step proposes a keyword extraction method to analyze the concept exchange and recombination in the interdisciplinary researches, which gives us a deeper look into the interdisciplinary bridges where innovation can be found. These efforts provide a dynamic and comprehensive understanding of the diffusion of LDA.

Titles and abstracts of LDA citations are pre-processed: (1) All words are converted to lowercase and the plural changed its singular form; and (2) A stop word list is used to filter the common words, and words that have fewer than three letters are removed.

Topic modeling. The LDA topic discovery model (Blei, Ng, & Jordan, 2003) is an unsupervised algorithm for performing statistical topic modeling using a “bag of words” approach that treats each document as a vector of words. Each document is represented as a probability distribution over certain topics, where each topic is a probability distribution of words.

With a corpus of M documents $\{w_1, w_2, \dots, w_m\}$ containing words from a vocabulary of N terms, LDA assumes that documents are generated from a set of K latent topics. In a document, each word w_i is associated with a hidden variable $z_i \in \{1, \dots, K\}$ indicating the topic from which w_i is generated. The probability of word w_i can be expressed as:

$$P(w_i) = \sum_{j=1}^K P(w_i|z_i = j)P(z_i = j)$$

where $P(w_i|z_i = j) = \beta_{ij}$ is a probability of word w_i in topic j and $P(z_i = j) = \theta_j$ is a document-specific mixing weight indicating the proportion of topic j in the document.

The multinomial parameters β and θ are sampled respectively as latent random variables from a Dirichlet prior with parameters α and η . Each document is obtained using the following generative process (Figure 2): (1) Sample a K -vector θ of document specific mixing weights from the Dirichlet distribution $P(\theta|\alpha)$; and (2) For each word, sample topic assigns j according to mixing weights $P(z) = \theta$ and draws a word according to $P(w|z) = j$.

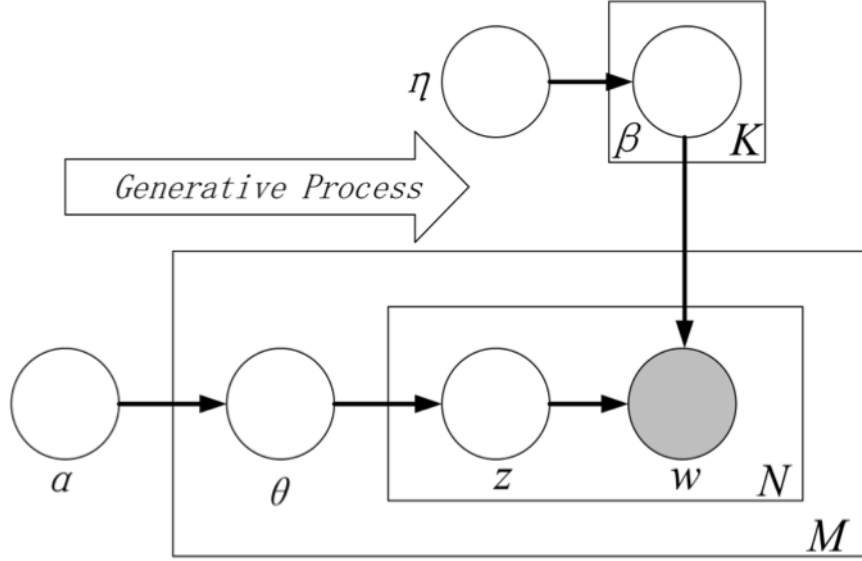


FIG. 2. Graphic Model Presentation of LDA

The Gensim library is used for the LDA topic modeling (Řehůřek & Sojka, 2010), where we apply standard parameters provided by Gensim (alpha='symmetric', eta=None, decay=0.5, eval_every=10, iterations=50, gamma_threshold=0.001, update_every=1). Considering the disciplinary differences in relation to the size, diversity, and duration of the current data set, the number of topics is set at 30 for this study.

Topic popularity. Based on previous studies (Griffiths & Steyvers, 2004), popular topics are found to be those with high topic proportions among a number of articles. Topic popularity is calculated through θ_d , the per-document topic proportion for document d . For example, as illustrated in Table 2, five papers are assigned to three topics. For each topic j , the popularity of topic $\text{Pop}(j)$ can be calculated through aggregating $\theta_{d,j}$.

TABLE 2. An Example of Topic Popularity for Three Topics

| | Doc 1 | Doc 2 | Doc 3 | Doc 4 | Doc 5 | Popularity |
|---------|-------|-------|-------|-------|-------|------------|
| Topic 1 | 0.31 | 0.02 | 0.09 | 0.11 | 0.02 | 0.55 |
| Topic 2 | 0.22 | 0.12 | 0.39 | 0.04 | 0.08 | 0.85 |
| Topic 3 | 0.01 | 0.80 | 0.22 | 0.03 | 0.43 | 1.49 |

The topic popularity for topic j in year t can be expressed as:

$$\text{Pop}(j|t) = \sum_{d|py(d)=t} \theta_{d,j}$$

where $py(d)$ denotes the publication year of document d .

Subject topic. With the topic distribution of each paper, we can obtain the most-related topic for each subject by manipulating the Paper-Subject matrix and Topic-Paper matrix. Figure 3 shows the way to form the Topic-Subject matrix. We define the subject topic as the topic with the maximum value.

Paper-Subject matrix

| | Doc 1 | Doc 2 | Doc 3 | Doc 4 | Doc 5 |
|---------|-------|-------|-------|-------|-------|
| Topic 1 | 0.31 | 0.02 | 0.09 | 0.11 | 0.02 |
| Topic 2 | 0.22 | 0.12 | 0.39 | 0.04 | 0.08 |
| Topic 3 | 0.01 | 0.8 | 0.22 | 0.03 | 0.43 |

Topic-Paper matrix \times

| | Subject A | Subject B | Subject C |
|-------|-----------|-----------|-----------|
| Doc 1 | 1 | 0 | 0 |
| Doc 2 | 0 | 0 | 1 |
| Doc 3 | 1 | 1 | 0 |
| Doc 4 | 0 | 1 | 0 |
| Doc 5 | 0 | 1 | 1 |

Topic-Subject matrix =

| | Subject A | Subject B | Subject C |
|---------|-----------|-----------|-----------|
| Topic 1 | 0.4000 | 0.2200 | 0.0400 |
| Topic 2 | 0.6100 | 0.5100 | 0.2000 |
| Topic 3 | 0.2300 | 0.6800 | 1.2300 |

FIG. 3. The formation of a Topic-Subject Matrix

Results

Diffusion stages of an innovation

In this section, we analyze the research topics generated from the citations of LDA. Table 2 lists 30 topics labeled by the top five words with the highest associations for each topic. These topics give us a general understanding of LDA-related researches. We further divide these topics into three categories:

Technology-related topics (1 to 12), which mainly focus on LDA algorithm, evaluation, parameter setting, and extension/improvement;

Application-related topics (13-23), which consist of text mining, topic modeling, system design, automatic translation, information retrieval, user recommendation, sentiment analysis, opinion mining, and image annotation; and

Data-related topics (24-30), which include biological data, scientific data, medical data, image data, and social media data.

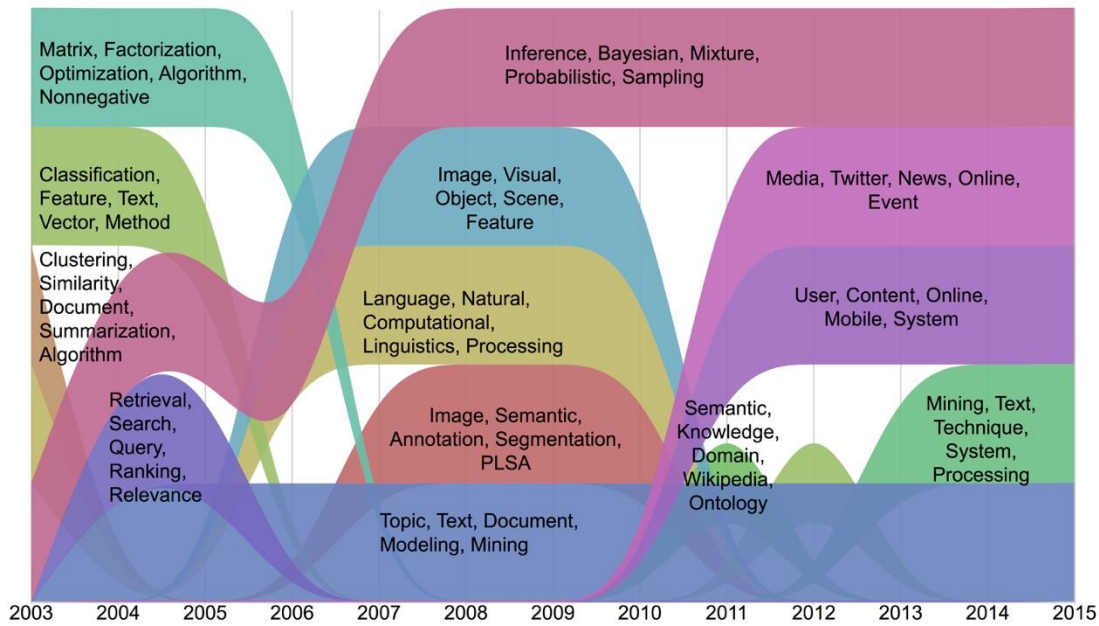
Obviously, these three types are interrelated. For example, sentiment analysis (application-related) mainly uses social media data (data-related) and inextricably links with natural language processing (technology-related). LDA originally applied in topic modeling and text mining (Blei, Ng, & Jordan, 2003) is found to produce more technology-related and application-related derivatives.

TABLE 3. 30 Topics in LDA Citations

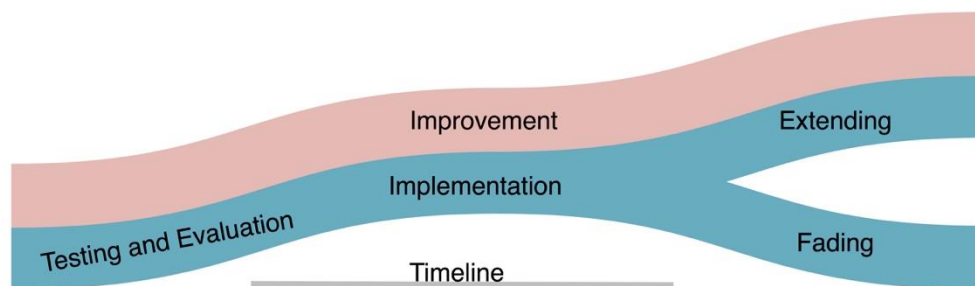
| Topic ID | Topic Keywords | Topic ID | Topic Keywords |
|----------|----------------|----------|----------------|
|----------|----------------|----------|----------------|

| | | | |
|----|---------------------------------------------------------------|----|----------------------------------------------------------|
| 1 | Language, Natural, Computational, Linguistics, Processing | 16 | Recognition, Human, Translation, Action, Crosslingual |
| 2 | Learning, Machine, Semisupervised, Sparse, Artificial | 17 | Retrieval, Indexing, Correlation, Multimodal, Multimedia |
| 3 | Supervised, Label, Classification, Generative, Discriminative | 18 | Community, Service, Discovery, Detection, Matching |
| 4 | Software, Source, Code, Quality, Programming | 19 | Recommendation, Collaborative, Filtering, System, User |
| 5 | Inference, Bayesian, Mixture, Probabilistic, Sampling | 20 | Sentiment, Prediction, Regression, Online, Forecasting |
| 6 | Clustering, Similarity, Document, Summarization, Algorithm | 21 | Retrieval, Search, Query, Ranking, Relevance |
| 7 | Matrix, Factorization, Optimization, Algorithm, Nonnegative | 22 | Review, Opinion, Visualization, Product, Mining |
| 8 | Dirichlet, Allocation, Hierarchical, Statistics, Topic | 23 | Image, Semantic, Annotation, Segmentation, PLSA |
| 9 | Classification, Feature, Text, Vector, Method | 24 | Human, Expression, Gene, Functional, Urban |
| 10 | Network, Graph, Link, Influence, Structure | 25 | Scientific, Digital, Author, Paper, Citation |
| 11 | Time, Temporal, Dynamic, Evolution, Blog | 26 | Video, Detection, Patterns, Activity, Behavior |
| 12 | Semantic, Knowledge, Domain, Wikipedia, Ontology | 27 | User, Content, Online, Mobile, System |
| 13 | Topic, Text, Document, Modeling, Mining | 28 | Media, Twitter, News, Online, Event |
| 14 | Mining, Text, Technique, System, Processing | 29 | Image, Visual, Object, Scene, Feature |
| 15 | System, Automated, Requirements, Reports, Design | 30 | Medical, Clinical, Health, Risk, Support |

Next, we use the topic popularity in each year ($\text{Pop}(j|t)$) to determine the top five popular topics per year, as shown in Figure 5(a).



(a) Evolution of Top 5 Topics in Each Year



(b) Scientific Innovation Growth Model

FIG. 5. Diffusion Stages of Innovation

Rogers (1962) identified five stages of the innovation diffusion process from the perspective of the users' adoption decision: **knowledge** – person becomes aware of an innovation and has some idea of how it functions; **persuasion** – person forms a favorable or unfavorable attitude toward the innovation; **decision** – person engages in activities that lead to a choice to adopt or reject the

innovation; **implementation** – person puts an innovation into use; **confirmation** – person evaluates the results of an innovation-decision already made. The weakness of this stage model is that the decision-making process is only part of the diffusion process and it does not show the entire innovation cycle, for example, innovation may change during the diffusion process. What's more, diffusion of scientific innovation involves cross-domain applications, which are different from products that have long been used to solve a particular problem/task and remain in a stable form.

In the context of citation-based diffusion of innovation, citations represent indirect diffusion of innovation which means that, first, innovation has been formed, and secondly, diffusion of innovation can be expressed as the transmission of knowledge and not necessarily adoption, and finally, the form of innovation is not immutable. Based on the topic evolution of LDA and follow Rogers' diffusion model (1962), we further divide the diffusion process of innovation into five phases, as shown in Figure 5(b): Testing and Evaluation, Implementation, Improvement, Extending, and Fading.

Testing and Evaluation is the first stage of diffusion and deals principally with the assessment of the performance and efficiency of the innovation. In the early period of the diffusion of LDA (2003-2006), topics are mainly technology-related and application-related, which means that adopters question the performance of the innovation and use different ways to evaluate its algorithms and outcomes. Corresponding to knowledge and persuasion stage of Rogers' diffusion model, testing and evaluation can accelerate the diffusion of innovation and help scientists in other disciplines understand the advantages of innovation.

Implementation, which occurs in the early period of innovation diffusion, means that researchers begin to put the innovation into practical applications. In this stage, innovation is widely accepted by scientists in the originally proposed field and transformed into other forms to solve other types of problems. For instance, LDA is upgraded into Multimodal Multi-Instance Multi-Label LDA (M3LDA) for image annotation task (Nguyen, Zhan, & Zhou, 2013). We can also find the rapid growth of topic23 {Image, Semantic, Annotation, Segmentation, PLSA} and topic29 {Image, Visual, Object, Scene, Feature} during 2004 to 2006 in Figure 5(a).

Extending, which plays a crucial role in the whole diffusion process, is the stage when the innovation is applied across the disciplinary boundary into other domains. Here, innovation has been recognized and trusted, and adopters are no longer skeptical about the efficiency of the innovation. Scientists no longer care about the technical problems of the innovation, but begin to apply the innovation to solve practical problems in different areas. The type of task which the innovation is originally designed for and the form of the innovation itself are not changed. In our case, more data-related topics (topic 27 and topic 28) are beginning to emerge at the end of the diffusion.

Fading shows that the original innovation is replaced by new ones, which does not occur once, but will constantly circulate along with the stage of implementation. This means that the process of innovation diffusion will not end until an enormous innovation completely replaces the existing methods. As shown in Figure 5(a), topic23 {Image, Semantic, Annotation, Segmentation, PLSA} and topic29 {Image, Visual, Object, Scene, Feature} start to decline from 2009 to 2011. This decline is due to neural networks gradually replacing topic modeling in these two topics (Li, Su, Xing, & Fei-fei, 2010).

Improvement, which persists throughout the whole diffusion process, refers to enhancing the algorithms and compensating for the deficiencies (i.e. determining the number of the topics, and updating the existing parameters). Related to methodology improvement and optimization, topic5 {Inference, Bayesian, Mixture, Probabilistic, Sampling} has always been an important topic in LDA studies.

In this growth model, five stages of innovation diffusion exist simultaneously and attach to each other. In other words, when the innovation is being tested, other researchers could also apply it, improve it, and extend it. Note that we do not consider the stage of idea generation because our analysis of the growth process is based on the precondition that the innovation has already formed.

Subject topic

First of all, we need to figure out the key topic of each subject during the diffusion of LDA, that is, we identify how LDA is applied in each subject. According to the introduction to Topic-Subject matrix in the methods section, we extract the corresponding topic to each subject with the maximum value in the matrix. As seen in Figure 6, the left side shows each subject and the right side shows the corresponding topic, from which we can clearly verify the research themes of distinct areas.

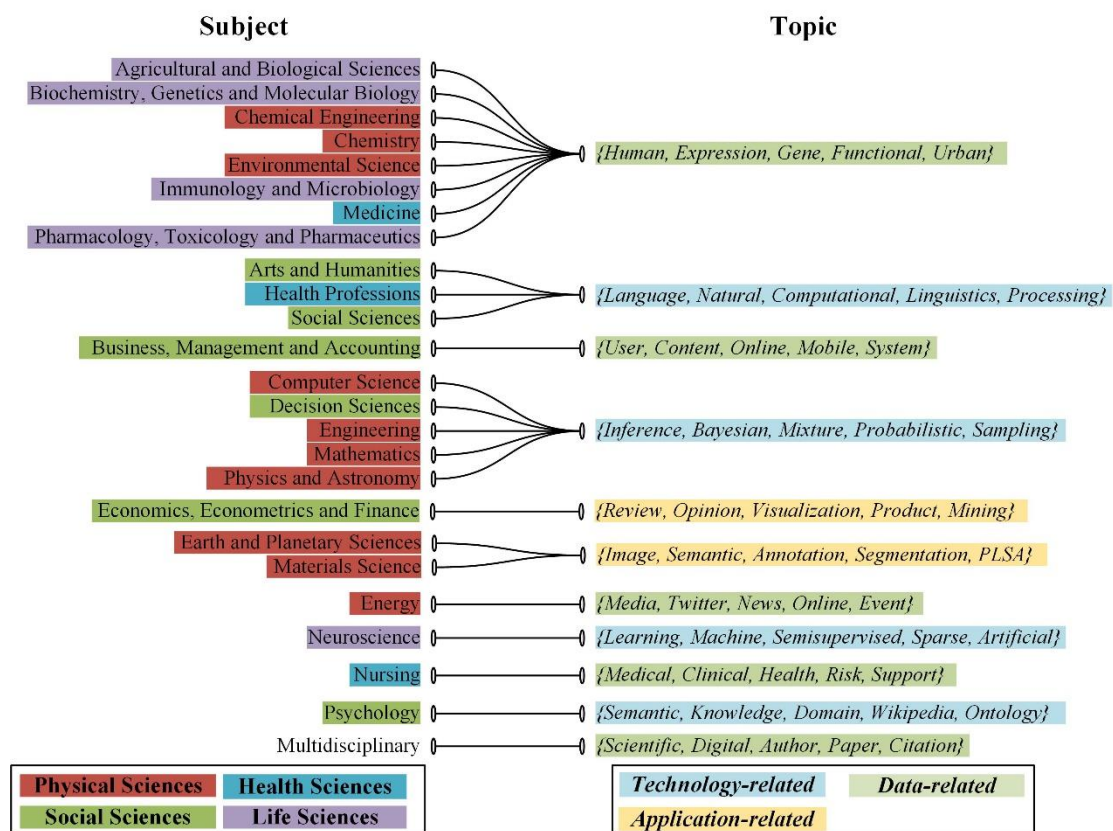


FIG. 6. Topics from Different Subjects (left side is subject and right side is topic)

Physical Sciences covers all three kinds of topics: technology-related, data-related and application-related. On the one hand, technology-focused subjects, such as Computer Science and Mathematics, connect with the topic {Inference, Bayesian, Mixture, Probabilistic,

Sampling} (technology-related). On the other hand, Earth and Planetary Science and Materials Science concentrate on the image annotation (application-related). Unexpectedly, the results show that Energy is highly related to topic {Media, Twitter, News, Online, Event}. When we take a close look at the original dataset, we find that there are only three Energy papers citing LDA and only one of them adopts LDA, in this case to extract topics in safety reports for maintenance action recommendation (Das, 2013).

Medical informatics has become a big growth area with the increased collaboration between medicine and data mining (Quackenbush, 2006). In the Health Sciences, Medicine, Nursing, and Health Professions get three relevant topics, which are natural language processing (technology-related), medical disease (data-related), and human genes (data-related). Some of the Health Sciences researchers apply LDA on text data directly, such as health checkup questionnaires (Hatakeyama et al., 2015), and others extend LDA to the medical study and knowledge discovery tasks. For example, Wang and Ding et al. (2011) described an algorithm called Bio-LDA that uses extracted biological terminology to automatically identify latent topics, and provides a variety of measures to uncover putative relations among topics and bio-terms.

In Life Sciences (Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Immunology and Microbiology, Pharmacology, Toxicology and Pharmaceutics), the main topic is concentrated on the human genome, which refers to {human, expression, gene, functional, urban}. One of the emerging Life Sciences subjects, Neuroscience mainly focuses on machine learning-related topics, and often reveals textual information from the physical point of view while paying more attention to human behavior analysis. For example, to analyze the variations in language based on personality, gender, and age, Schwartz and Eichstaedt et al. (2013) used 700 million words, phrases, and topic instances collected from the Facebook messages of 75,000 volunteers that give a comprehensive exploration of language and distinguish people and finds connections that are not captured with traditional closed-vocabulary word-category analyses.

Social Sciences covers the most diverse topics. Opinion mining and sentiment analysis are the most-related research themes. Arts and Humanities attaches importance to the natural language processing(NLP) in topic modeling for questions like understanding the history of cognition (Cohen& Austerweil, 2015); Business, Management and Accounting focuses on analyzing information behavior of online users; Decision Sciences gives substantial attention to optimization model and algorithm improvements; Economics, Econometrics and Finance develops the method of LDA to extract the opinions from user reviews; and Psychology concentrates on organizing knowledge on a semantic level.

In multidisciplinary researches, LDA is usually used in the forecast and foresee applications for the evolution of science and the development of subjects. Overall, the corresponding topics of each subject (Figure 6) provides an in-depth understanding of the diffusion of innovation.

Cross-boundary diffusion

Interdisciplinary works give us opportunities to better understand how an innovation is diffused across disciplinary boundaries. Diffusion is a process that develops along the time. Subjects can be viewed as adopters, and the time they receive an innovation can be expressed in terms of the publication time of the first article citing the innovation. The first article of the

subject is likely to be interdisciplinary, for example an article signed into subject A and B. When A has already referenced LDA before B appears, A can be seen as a bridge where LDA diffuse to B. This kind of bridge is not the same as the citation-based knowledge flow. On the one hand, the individual adopter of the innovation, that is, the researcher, determines the source of the innovation-related information. Scientists may learn LDA by reading literature, listening to lectures, and communicating with other scholars, so we cannot determine the channels by which subject receive LDA. Here, the bridge between subjects that we build represents the path of innovation diffusion. It is also a process of knowledge accumulating. On the other hand, the citation-based knowledge flows use all citations between subjects, and here we only focus the diffusion of LDA.

We depict the adoption time for different subjects during LDA's diffusion (Figure 7). If the first article of the subject is interdisciplinary study, we add a bridge between the signed subjects. Subjects are coded with colors that represent the subject areas the subject belongs to.

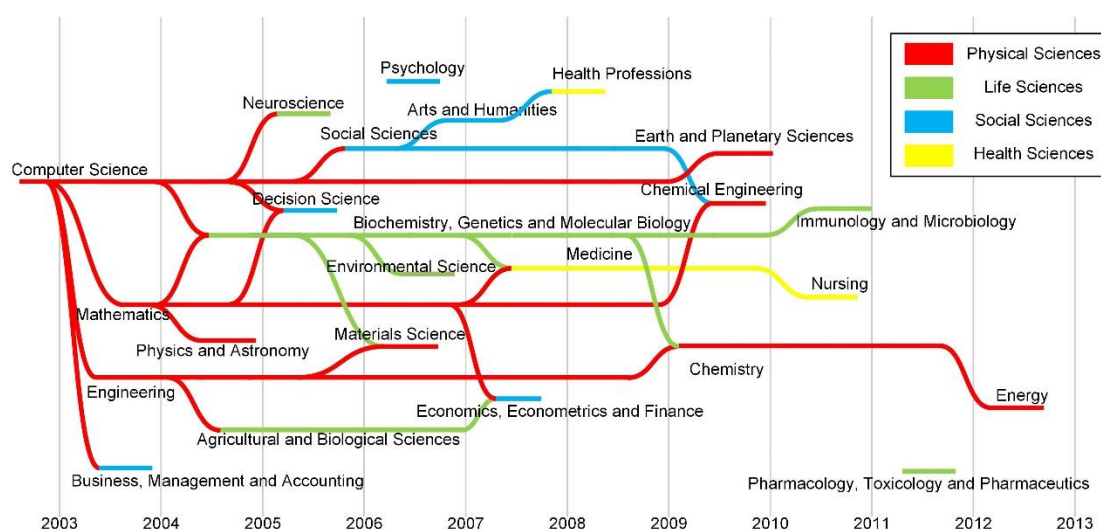


FIG. 7. A Subject Timeline for LDA Diffusion

Interdisciplinary researches are very important intermediary for innovation diffusion. In Figure 7, most of the subjects first cited LDA are through interdisciplinary research, except Psychology and Pharmacology, Toxicology and Pharmaceutics. LDA needs to go through the interdisciplinary researches to reach the subjects. In addition, subjects also play an important mediator role in the diffusion of LDA. Computer Science, Mathematics, Engineering and Biochemistry, Genetics and Molecular Biology build the major branches in this diffusion map.

At the very beginning of the diffusion process, innovation spreads out towards the adjacent subjects. Starting from Computer Science, diffusion of LDA goes in sequence through Engineering, Business, Management and Accounting, Mathematics, Physics and Astronomy, all of which belong to the Physical Sciences. This result is also consistent with previous findings, that publications in one discipline tend to cite papers in adjacent disciplines (van Leeuwen & Tijssen, 2000).

The distribution of the adoption time for each subject area is quite dispersed. As shown in Figure 7, Computer Science and Engineering are the earliest subjects to cite/adopt LDA.

Although both Energy and Engineering belong to the Physical Sciences, they first cite LDA in 2003 and 2013. Similarly, in Life Sciences, LDA is first adopted by Agricultural and Biological Sciences in 2004, and is finally diffused to Pharmacology, Toxicology, and Pharmaceutics in 2012. The interdisciplinary diffusion order does not therefore depend on whether the subjects directly belong to the same subject areas

By combining the subject topics (Figure 6) and adoption times (Figure 7), one finds that the adoption times for different subjects are relatively adjacent and continuous in years when the subjects have similar research interests. For instance, Computer Science, Engineering, Mathematics, Physics and Astronomy, and Decision Science are mostly related to topic {Inference, Bayesian, Mixture, Probabilistic, Sampling}. Their adoption times are all concentrated in the period from 2003 to 2005. Similarly, Arts and Humanities, Health Professions, and Social Sciences, which adopt LDA in 2006 and 2007, are all interested in topic {Language, Natural, Computational, Linguistics, Processing}. The reason behind this phenomenon is that, people with similar interests are more likely to operate, communicate, and work together. As innovation diffuses to different disciplines, the disciplinary boundary is broken, and a social network with potentially strong cohesion is established between researchers from different disciplines. Some theories also emphasize that the evolution of disciplines is driven by the formation of social groups of scientists (Bettencourt, Kaiser, & Kaur, 2009; Guimera, Uzzi, Spiro, & Amaral, 2005; Hagstrom & Crane, 1973).

When first using the LDA model, scientists who are unfamiliar with programming find it difficult to operate. The training efficiency and quality of outcomes also need to be further examined and evaluated. With the maturity of LDA, scientists have developed topic modeling software that can be used directly, such as Genism (Řehůřek & Sojka, 2010), MALLET (McCallum, 2002), and Stanford Topic Modeling Toolbox (2009), which help LDA diffuse to more disciplines. Researchers from other disciplines gradually started to apply it directly to different datasets, such as mining daily life activity for patients (Seiter, Derungs, Schuster-Amft, Amft, & Tröster, 2015).

Conclusion

This study analyzes the diffusion of innovation by using LDA citations as a proxy. It highlights the diffusion stages of LDA using topic evolution and examines the cross boundary diffusion process between different subjects. First, a topic modeling technique is applied to reveal the topic evolution of innovation between subjects. We further divide the diffusion process of an innovation into five stages, including testing and evaluation, implementation, improvement, extending, and fading. The stages develop simultaneously and attach to each other. These findings thus contribute to the literature on the dynamic phases found in innovation diffusion. Also, topic-level studies are found to play a critical role in bringing more granular perspectives to the existing co-occurrence-based analyses. In summarizing the corresponding topics of each subject, from the adoption time of LDA in different subjects, innovation is seen to first diffuse to adjacent subjects, and then transfers along to others with the evolution of topics. These findings give us a better understanding of the diffusion process through subjects. Furthermore, interdisciplinary researches and subjects are both play an

intermediary role in innovation diffusion.

Here we address some limitations and future works of this study. One limitation is that Scopus is not expected to contain all important scholarly literature which necessarily underestimate the true amount of citation among a set of journals—compared with biomedical related disciplines, social science and humanities may still have an inequitable visibility in Scopus (de Moya-Anegón et al., 2007). The limited size of the dataset is also one of the shortcomings. Although focusing on one particular topic like LDA can give us deeper insights and closer views of the evolution process involved in innovation diffusion and transfers into different disciplinary areas, we need more cases to test our stage model's general explanations on innovation diffusion regarding to other innovations. Finally, we only consider the direct citation, while ignoring the indirect and inner citations which can also represent the flow of knowledge. Moreover, there are some articles that adopt the innovation directly but are not indicated in the reference literature which are also implicit diffusion of innovation. In order to solve these problems, we need to use content-based citation analysis to restore a more complete diffusion process. Future works will focus on exploring factors that contribute to topic dynamics and discovered patterns of the diffusion of innovation. This requires further analysis of the background and context of the citations (e.g. social network of the authors, and the papers' full text) and identifying how LDA was adopted, transferred, and improved during its diffusion phases.

Acknowledgments

This work was supported by National Social Science Foundation of China (14ZDA063)

References

- Abrahamson, E., & Rosenkopf, L. (1997). Social Network Effects on the Extent of Innovation Diffusion: A Computer Simulation. *Organization Science*, 8(3), 289-309. doi:10.1287/orsc.8.3.289
- Au Yeung, C. M., & Jatowt, A. (2011, October). Studying how the past is remembered: towards computational history through large scale text mining. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 1231-1240). ACM.
- Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006, August). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 44-54). ACM.
- Bass, F. M. (1969). A new product growth model for consumer durables. *Management*, 15(5), 215-227.
- Bettencourt, L. M. A., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, 3(3), 210-221. doi:10.1016/j.joi.2009.03.001
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- 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.
- Chatterjee, R. a., & Eliashberg, J. (1990). The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach. *Management Science*, 36(9), 1057-1079. Retrieved at doi:10.1287/mnsc.36.9.1057

- Cole, J. R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, 281-300.
- Cooper, R. B., & Zmud, R. W. (1990). Information Technology Implementation Research: A Technological Diffusion Approach. *Management Science*, 36(2), 123-139. Retrieved at doi:10.1287/mnsc.36.2.123
- Cronin, B., & Meho, L. (2008). The Shifting Balance of Intellectual Trade in Information Studies. *Journal of American Society for Information Science and Technology*, 59(4), 551-564.
- Das, S. (2013). Maintenance action recommendation using collaborative filtering. *International Journal of Health Policy and Management*, 4(2), 7-12.
- Desouza, K. C., Dombrowski, C., Awazu, Y., Baloh, P., Papagari, S., Jha, S., & Kim, J. Y. (2009). Crafting organizational innovation processes. *Innovation: Management, Policy & Practice*, 11(1), 6-33. doi:10.5172/impp.453.11.1.6
- Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of science*, 17(2), 102-134.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, 22(2), 338-342.
- Goyal, S., Heidari, H., & Kearns, M. (2014). Competitive contagion in networks. *Games and Economic Behavior*.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228-5235. Retrieved at doi:10.1073/pnas.0307752101
- Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722), 697-702.
- Hägerstrand, T. (1967). *Innovation Diffusion as a Spatial Process*. Translated by A. Pred. Chicago: Chicago University Press.
- Hagstrom, W. (1973). Reviewed Work: *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. *Contemporary Sociology*, 2(4), 381-383. Retrieved from <http://www.jstor.org/stable/2062040>
- Hare, J. S., Lewis, P. H., Enser, P. G., & Sandom, C. J. (2006, January). Mind the Gap: Another look at the problem of the semantic gap in image retrieval. In *Electronic imaging 2006* (pp. 607309-607309). International Society for Optics and Photonics. Retrieved at doi:10.1117/12.647755
- Hatakeyama, Y., Miyano, I., Kataoka, H., Nakajima, N., Watabe, T., Yasuda, N., & Okuhara, Y. (2015). Use of a Latent Topic Model for Characteristic Extraction from Health Checkup Questionnaire Data. *Methods of information in medicine*, 54(6), 515-521.
- Henard, D. H., & Szymanski, D. M. (2001). Why some new products are more successful than others. *Journal of marketing Research*, 38(3), 362-375.
- Hu, Y. (2005). Efficient, High-Quality Force-Directed Graph Drawing. *Mathematica Journal*, 10(1), 37-71. doi:10.3402/qhw.v6i2.5918
- Jaffe, A. B., & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *proceedings of the National Academy of Sciences*, 93(23), 12671-12677.
- 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. (2009). Toward a consensus map of science. *Journal of the American*

- Society for information science and technology*, 60(3), 455-476.
- Kline, S. J., & Rosenberg, N. (1986). An Overview of Innovation. *European Journal of Innovation Management*, 38, 275–305. doi:10.1108/14601069810368485
- Krishnan, A. (2009). What are academic disciplines. *University of Southampton, NCRM E Prints Repository eprints. ncrm.ac.uk/783/1/what_are_academic_disciplines.pdf*.
- Kusiak, A. (2016). Put innovation science at the heart of discovery. *Nature*, 530(7590), 255. doi:10.1038/530255a
- Kwon, T. H., & Zmud, R. W. (1987, April). Unifying the fragmented models of information systems implementation. In *Critical issues in information systems research* (pp. 227-251). John Wiley & Sons, Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=54915>
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard university press.
- 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. doi:10.1002/asi.21539
- Lewin, K. (1947). Group decision and social change. *Readings in Social Psychology*, 3(2), 213–231. doi:10.1177/0021886306297004
- 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., & 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. doi:10.1002/asi.21052
- Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348-362.
- Li, L. J., Su, H., Fei-Fei, L., & Xing, E. P. (2010). Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Advances in neural information processing systems* (pp. 1378-1386).
- M. Bastian, S. Heymann, M. J. et al. (2009). Gephi: an open source software for exploring and manipulating networks. *Proceedings of International AAAI Conference on Web and Social Media*, 361–362. Retrieved at <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/154/1009/>
- MacRoberts, M., & MacRoberts, B. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.
- Mahajan, V., & Muller, E. (1979). Innovation diffusion and new product growth models in marketing. *The Journal of Marketing*, 43(4), 55–68. doi:10.2307/1250271
- Mansfield, E. (1961). Technical Change and the Rate of Imitation. *Econometrica*, 29(4), 741–766. doi:10.2307/1911817
- McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit. <http://Mallet.Cs.Umass.Edu>. Retrieved at <http://mallet.cs.umass.edu>
- Mccallum, A., Wang, X., & Corrada-Emmanuel, A. (2005). Topic and Role Discovery in Social Networks. *Proceedings of the 19th International Joint Conference on Artificial Intelligence - IJCAI '05*, 786–791. Retrieved at <http://dl.acm.org/citation.cfm?id=1642293.1642419>

- McFarlan, F., McKenney, J., & Pyburn, P. (1983). *The information archipelago -- Plotting a course*. *Harvard Business Review* (Vol. 61).
- Mehmood, Y., Barbieri, N., & Bonchi, F. (2016). Modeling adoptions and the stages of the diffusion of innovations. *Knowledge and Information Systems*, 48(1), 1-27.
- 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.
- Mintrom, M., & Vergari, S. (1998). Policy networks and innovation diffusion: The case of state education reforms. *The Journal of Politics*, 60(1), 126–148. Retrieved at <http://www.jstor.org/stable/2648004>
- Montanari, A., & Saberi, A. (2010). The spread of innovations in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47), 20196–20201. doi:10.1073/pnas.1004098107
- Mowery, D. C., & Ziedonis, A. A. (2002). Academic patent quality and quantity before and after the Bayh-Dole act in the United States. *Research Policy*, 31(3), 399-418.
- Narin, F., Carpenter, M., & Berlt, N. C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science*, 23(5), 323–331. doi:10.1002/asi.4630230508
- Nguyen, C. T., Zhan, D. C., & Zhou, Z. H. (2013, August). Multi-modal image annotation with multi-instance multi-label LDA. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence* (pp. 1558-1564). AAAI Press.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135. doi:10.1561/15000000011
- Priva, U. C., & Austerweil, J. L. (2015). Analyzing the history of Cognition using topic models. *Cognition*, 135, 4-9.
- Quackenbush, J. (2006). Microarray analysis and tumor classification. *New England Journal of Medicine*, 354(23), 2463-2472.
- Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. *Journal of the Association for Information Science and Technology*, 60(9), 1823-1835.
- Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Retrieved at https://radimrehurek.com/gensim/lrec2010_final.pdf
- Rinia, E. D., Van Leeuwen, T., Bruins, E., Van Vuren, H., & Van Raan, A. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, 51(1), 293-309.
- Rogers, E. M. (1962). *Diffusion of innovations*. New York: The Free Press.
- Rogers, E. M., & Shoemaker, F. (1971). *Communication of innovations: a cross-cultural approach*. New York: Free Press, 476. Retrieved from <http://eric.ed.gov/?id=ED065999>
- Rong, X., & Mei, Q. (2013, October). Diffusion of innovations revisited: from social network to innovation network. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 499-508). ACM.
- Rorissa, A., & Yuan, X. (2012). Visualizing and mapping the intellectual structure of information retrieval. *Information processing & management*, 48(1), 120-135.
- Ryan, B., & Gross, N. C. (1943). The Diffusion of Hybrid Seed Corn In Two Iowa Communities. *Rural Sociology*, 8(1), 15–24. doi:citeulike-article-id:1288385

- Saren, M. a. (1984). A classification and review of models of the intra-firm innovation process. *R&D Management*, 14(1), 11–24. doi:10.1111/j.1467-9310.1984.tb00504.x
- Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. (2013) Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE* 8(9): e73791. doi:10.1371/journal.pone.0073791
- Seiter, J., Derungs, A., Schuster-Amft, C., Amft, O., & Tröster, G. (2015). Daily life activity routine discovery in hemiparetic rehabilitation patients using topic models. *Methods of information in medicine*, 54(3), 248-255.
- Small, H. G. (1978). Cited documents as concept symbols. *Social studies of science*, 8(3), 327-340.
- Tang, J., Sun, J., Wang, C., & Yang, Z. (2009, June). Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 807-816). ACM.
- Tang, R. (2004). Evolution of the interdisciplinary characteristics of information and library science. *Proceedings of the American Society for information Science and Technology*, 41(1), 54-63.
- Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16), 5962-5966.
- Van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows. *Research Evaluation*, 9(3), 183–187. doi:10.3152/147154400781777241
- Wang, H., Ding, Y., Tang, J., Dong, X., He, B., Qiu, J., & Wild, D. J. (2011). Finding complex biological relationships in recent PubMed articles using Bio-LDA. *PloS one*, 6(3), e17243. doi:10.1371/journal.pone.0017243
- Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific reports*, 2.
- Yan, E. (2015a). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology*.
- Yan, E. (2015b). Research dynamics, impact, and dissemination: A topic-level analysis. *Journal of the Association for Information Science and Technology*, 66(11), 2357–2372. doi:10.1002/asi.23324
- Yan, E., & Yu, Q. (2015). Using path-based approaches to examine the dynamic structure of discipline-level citation networks: 1997-2011. *Journal of the Association for Information Science and Technology*, 14(4), 1943–1955. doi:10.1002/asi.23516
- Youngs, N., Penfold-Brown, D., Bonneau, R., & Shasha, D. (2014). Negative Example Selection for Protein Function Prediction: The NoGO Database. *PLoS Computational Biology*, 10(6), e1003644. doi:10.1371/journal.pcbi.1003644
- Zhang, G., Ding, Y., & Milojević, S. (2013). Citation content analysis (cca): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology*, 64(7), 1490-1503.
- Zhao, R., & Wu, S. (2014). The network pattern of journal knowledge transfer in library and information science in China. *Knowledge Organization*, 41(4), 276–287.