Scientometrics
https://doi.org/10.1007/s11192-018-2693-y

The mutually beneficial relationship of patents and scientific literature: topic evolution in nanoscience

Yashuang Qi[1] · Na Zhu[1] · Yujia Zhai[2] · Ying Ding[3]

Received: 27 August 2017

Abstract Patent and scientific literature are the fundamental sources of innovation in knowledge creation and transfer activities. Establishing and understanding the complex relationships between them can help scientists and stakeholders to predict and promote the innovation process. In this paper, we consider the domain of nanoscience, using a large scale collection of patents and scientific literature to find evolution patterns and distinctive keywords of each topic in a particular period. By extracting the semantic-level topics from a dataset of nearly 810,000 scientific literature from Web of Science and 160,000 patents from Derwent, the results reveal that the degree of topic popularity for both innovative platforms shows a totally different situation during the last 20 years from 1995 to 2015. In addition, the top keywords of patents and scientific literature, representing the topic content of concern, have changed respectively as time went on. Not only our analysis can be used for confirming existing topics in nanoscience, but it also gives new views on the relationship between scientific literature and patents.

Keywords Nanoscience · Lead–lag analysis · Patent · Scientific literature

Mathematics Subject Classification 68T10

JEL Classification D830

✉ Yujia Zhai
zhaiyujiachn@gmail.com

1 Department of Information Resource Management, School of Business, Nankai University, Tianjin 300071, China

2 Department of Information Resource Management, School of Management, Tianjin Normal University, Tianjin 300387, China

3 School of Informatics and Computing, Indiana University, Bloomington 47401, USA

Published online: 26 February 2018

Introduction

Measuring and assessing innovation is a prominent issue in accelerating economic growth and social development. As the major innovation carriers, scientific literature and patents can play key roles in detecting the trajectories of innovation diffusion and evaluating the value of innovation. Among both scientists and economists, it is widely accepted scientific advances are the wellspring of technical innovations. However, in practice, science may be influenced by technological innovation, such as the exploration of new fundamental questions in the practical process of solving problems (Mansfield 1991). For example, when an engineer presents a solution and implements it, the technology generates empirical evidences awaiting the explanation from the scientists.

Although quite a bit of works have examined the relationships between scientific research and technological innovation(Gibbons and Johnston 1974; Narin et al. 1997; Meyer 2000a; Hullmann and Meyer 2003; Glänzel and Meyer 2003; Hsu et al. 2015), the underlying mechanism of interaction between them is still a mystery. Establishing and understanding the complex relationships can help us to predict and promote the long-term development of innovation, and also help the experts and stakeholders to ensure the innovations adequately reflect the public interest as new technologies are developed and deployed.

Science and technology are the fundamental sources of innovation in knowledge creation and transfer activities. The growth trajectory of both is not linear but has changed to a chain-linked relation (Caraça et al. 2009). Through measuring the citation linkage between research papers and patents, Meyer (2000b) concluded that nanoscience and nanotechnology are separate activities that are related to each other in a mediated manner. While empirical findings based on citation links still await substantial interpretation, it should be stressed that a considerable limitation of the evidences available so far is that of being ignored the content of scientific literature and patent.

Nowadays, the debate about the driven force of innovation gives rise to the effects of demand-pull and technology-push, or societal, economic, political and technological factors (Nemet 2009; Di Stefano et al. 2012), but ignores the internal relationship between scientific literature and patents. Questions like which one is the leader of innovation and how they support each other on a long-term scale are still difficult to answer. By analyzing the topic evolution patterns (e.g., the trend and duration of topic popularity) in scientific literature and patents, this work can help us better understand the interactivity of innovation paradigms and the role of each platform in the dissemination of novel knowledge.

Nanotechnology has been recognized as a key technology of the 21st century. Due to its implications and rapid development, nanotechnology has already been identified as a critical indicator of a country's technological competence. Various governments have prioritized nanotechnology in their national agenda for science and technology development (Paull et al. 2003; Roco and Bainbridge 2005).

The rest of this paper is organized according to the following structure. "Literature Review" section reviews previous studies including the relationship between patent and scientific literature, lead–lag analysis, topic modeling, and a comparison between our research and related works. In the Method section, we present in detail how the method of Latent Dirichlet Allocation Modeling and Regression Modeling work. The section "Results" follows, identifies hotspots in nanoscience, the growth patterns of topics in two kinds of publications. By comparing the keywords of topics between patents and scientific literature, which represent the concerned knowledge concepts, we further illustrate the reason behind the lead–lag relationship. Finally, we conclude our current research, noting limitations, and put forward possible directions for future work.

Literature review

Relationship between patent and scientific literature

Publications are considered as the outcome of scientific research, and as such represent science. At the same time, patents are regarded as the output of technological development(Verbeek et al. 2002; Meyer 2002; Bassecoulard and Zitt 2004). Scientists communicate their research findings through publications which are definitely an important component of knowledge generation and dissemination. Similarly, patents are not just legal documents which give a temporary monopoly to an inventor in exchange for disclosing the information of its invention; they are also regarded as a paper trail of technology advancement (C. Huang et al. 2011). The relationship between them in the emerging nanotechnology field has been extensively studied in the literature (Bhattacharya et al. 2003; Meyer 2000a), but which undertake the lead/lag position in the domain should be considered.

Today, over 50,000 nanotechnology articles have been published annually worldwide in recent years, and more than 2500 patents are filed at major patent office such as the European Patent Office (Logothetidis 2012). This large volume of data is desirable for rigorous quantitative analysis. Many prior works in citation analysis (Meyer 2000a, 2000b; Tussen et al. 2000; Branstetter and Ogura 2005; Igami and Okazaki 2007; Li et al. 2014) have been conducted to explore the linkage between patent and scientific research papers. Meyer (2000a) contended that a patent citation link in the field of nanotechnology should be considered as an indication of the multifaceted interplay between science and technology. Igami and Okazaki (2007) argued that nanoscience research provides an increasingly important foundation for nanotechnology innovation. Zhang et al. (2017) analyzed the scientific literature cited by the patent in the four fields of pharmaceuticals, nanotechnology, transgenic and energy. They studied the time delay between scientific research and technology patent in these four fields and try to find an optimal model for patent application. By comparing four classic journal citation models (i.e. the Bernal negative exponential model, the B–K equation, the corrected B–K equation and the transfer function model), it is found that the transfer function model is better than the other models, which is suitable for calculating the time delay between research achievements and technical patents.

Although extensive works have been done, few study has analyzed the issues which will be the lead one when a new topic appears, and how will the topic evolve in the two kinds of publishers in the following years. Topic modeling can help scientists to detect the evolving trend of the topics in a document collection without requiring the availability of citation data (Mei and Zhai 2005; Shaparenko et al. 2005; Nallapati et al. 2011). Scientific publication and patent data can be linked together by matching similar topics. Based on the hypothesis that novel ideas are disseminated through the creation and propagation of new or newly emphasized keywords, the lead/lag relation can be further estimated by tracking word usage across the timeline. In this paper, we will apply LDA (Latent Dirichlet allocation) to identify different topic evolution patterns between research papers and patents, find out the key topics in a collection of documents and how their popularity changes over time.

Lead–lag analysis

Temporal pattern analysis has been used on a variety of corpora, such as news and email datasets, and refined to identify bursts that are characterized by topics that appear, grow intensely, and then fade away based on time sequence with the combined approach of content analysis and time-series analysis (Shaparenko et al. 2005). It is successful in identifying papers that contained new and influential ideas. For example, Swan and Jensen (2000) developed TimeMines using robust, flexible techniques to determine significant keywords in documents and to judge their temporal significance in the context of a corpus. Another statistical method, labeled Temporal Text Mining (TTM), was introduced by Mei and Zhai (2005) and used to identify temporal patterns through the discovery of latent topics from text by constructing an evolution graph of topics, and to analyze life cycles of each topic for discovering globally interesting topics in each time periods.

Now, scientists conduct studies which compared topic evolution of research articles from the same domain but that were published in different (i.e., formal or informal) channels to find out the leader or the follower between them. Shi et al. (2010) proposed a general method for lead–lag estimation based on the LDA model and time series analysis. The result showed that the lead–lag of research papers, with respect to research grants, is topic specific. Security and Cryptography research papers lead grant proposals by 2 years, while grants related to the topic of Neural Network lead research papers by 3 years. In a later research work(Nallapati et al. 2011), they validated the hypothesis that lead/lag of outlets can be captured by tracking usage of keywords by testing a simple text based TF-IDF model on ACM Journals versus Conferences outlet pairs where general agreement on lead/lag behavior exists. The results showed that the model on three different outlet pairs presents interesting insights into their topic-specific lead/lag behavior. More recently, Hu et al. (2015) applied LDA and regression analysis to conduct a lead–lag analysis to identify different topic evolution patterns between preprints and papers from arXiv and Web of Science (WoS) in astrophysics. They found that arXiv and WoS share similar topics in a given domain, but differ in evolution trends. Topics in WoS lose their popularity much earlier and their durations of popularity are shorter than those in arXiv. It demonstrates that open access preprints have stronger growth tendency as compared to traditional printed publications.

In this paper, we use LDA to identify the strength variation trends of various topics and to compare the relative strengths of 50 different topics over time. Then, we apply regression modeling to analysis the lead–lag dynamic pattern. In the end, we plot the changes of top five words determined by word frequency along with different time durations and further explore the mutually beneficial relationship between patent and scientific literature in nanoscience.

Methodology

Dataset

Nanoscience has been greatly developed in recent years because of its enormous economic value (Kostoff et al. 2007). The goal of nanoscience research and development is to understand and create products to improve people's living standards (Guan and Zhao 2013). As a priority field in the strategic planning of the economic development, nanoscience has gained a lot of investment and important support (Hassan 2005). The reasons why we choose nanoscience is because it is considered as a promising field of science and technology (Meyer 2006). Scientists in this field have a disciplinary advantage in publishing and applying patents to their research findings. The scientific research and technological development of nanoscience have aroused wide attention of scholars and decision-makers all over the world.

Researchers have conducted in-depth analyzes on how to collect patent data or scientific literature of nanoscience (Zitt and Bassecoulard 2006; Li et al. 2007; Porter et al. 2008; Li et al. 2009). According to the research question waiting to answer, we emphasize a relatively larger scope and pay more attention to the recall than the precision in the data collection process. As stated by Porter et al. (2008), for a large and scattered field such as nanoscience, there is no absolute standard to measure the recall and precision. In short, the opinion about what should be included is different, because the scope is too wide. We need to have a clear understanding of the intention of the usage of nanoscience dataset. This will be reflected in the selection of a broader dataset or a more compact focus, and a trade-off between the accuracy of the precision rate or the sensitivity of the recall error. As our main purpose is to analysis the relation between scientific researches and patents, a high level of recall for both datasets is more important than high precision. In this way, although patent codes such as B82 can help us get the related patents of nanoscience more accurately (Ozcan and Islam 2017), it will lead to mismatch with the retrieval results of paper, which will lead to a deviation in the analysis of the relationship between them. Therefore, we only use the keyword-based retrieval method to get both datasets.

The dataset of this study includes patents retrieved from the Derwent Innovation Index (DII) database and academic periodical papers retrieved from the Web of Science (WOS) database. As the DII database has the elaborated classification scheme and lexical fields of patent, and the WOS database has an extensive collection of academic periodicals which contains a large amount of academic research findings (Sampat and Ziedonis 2004).

In reference to related studies (Li et al. 2007; Li et al. 2009), a list of keywords which are related to and can be representative of the broad topic of nanoscience are used to search and retrieve patents from DII database, including “Atomic force microscope”, “Atomic force microscopic”, “Atomic force microscopy”, “Atomic-force-microscope”, “Atomic-force-microscopy”, “Atomistic simulation”, “Biomotor”, “Scanning tunneling microscope”, “Scanning tunneling microscopic”, “Scanning tunneling microscopy”, “Scanning-tunneling-microscope”, “Scanning-tunneling-microscopy”, “Molecular device”, “Molecular electronics”, “Molecular modeling”, “Molecular motor”, “Molecular sensor”, “Molecular simulation”, “Quantum computing”, “Self assembly”, “Self assembled”, “Self assembling”, “Self-assembly”, “Self-assembled”, “Self-assembling”, Nano*, (Quantum dot*), (Quantum effect*) or (Self assembl*) in the ‘‘Title’’ and ‘‘Abstract’’. Search queries for WOS also used the same keywords to search by ‘‘Topic’’ to collect papers.

The patent date we searched for is based on the Derwent database’s patent ID, which is a unique ID for each patent family. The patent ID document a record with the time of publication. It is a six-digit serial number (YYYY–NNNNNN) where YYYY indicating the publication year of the patent in Derwent. The scientific literature is also retrieved based on the publication date.

We carried out our retrieving processes from February 21 to 29, 2017. Finally, a total of 164,183 patents and 813,624 papers related to nanoscience during the period of 1995 and 2015 were retrieved. For each paper and patent, year, title, abstract and author information were collected.

Methods

To explore the dynamic relationships between patents and scientific literature, we here adopt a four-step method as follows: Firstly, due to the large dataset and wide research field of nanoscience, we use LDA to extract 50 topics from the whole texts, which contain the title and abstract of all the patents and scientific publications. Then, we calculate the popularity of each topic in each independent year from 1995 to 2015 both for patents and scientific literature, which help us to understand the evolving trend of each topic in these two kinds of publications. After this, we use regression modeling to find out the dynamic evolving patterns of each topic respectively as time went on. The curve of topic popularity further demonstrates the lead–lag relationship between patents and scientific literature in nanoscience. Finally, we extract the top five keywords for each kind of publications in a 5-year window size.

Latent dirichlet allocation

The LDA topic discovery model (Blei et al. 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 $\left\{w_{1,}w_{2,}\ldots w_m\right\}$ 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 $z_i \in \{1,\ldots,K\}$ 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 (Fig. 1): (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 $P(w|z) = j$.

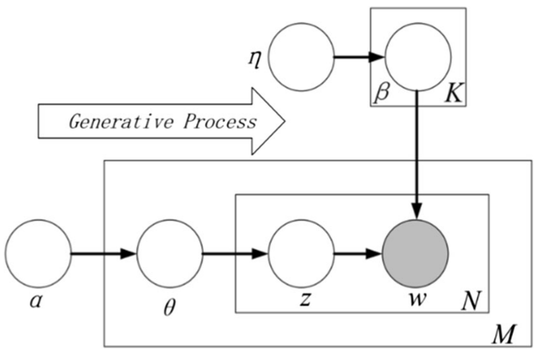


Fig. 1 Graphic model presentation of LDA

The Gensim library is used for the LDA topic modeling (Řehůřek and 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). The following data preprocessing was implemented: (1) title words that have less than three letters were removed from titles; (2) 30 most frequently occurred words were removed from titles; (3) those words that occurred in less than three publications were removed; (4) publications whose titles have less than three words were removed from the data set; and finally, (5) the number of topics was set as 50.

Topic popularity

Based on previous studies (Griffiths and 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 1, five papers are assigned to three topics. For each topic j, the popularity of topic $\mathrm{Pop}(j)$ can be calculated through aggregating $\theta_{\mathrm{d},j}$.

The normalized popularity for topic j is dividing the popularity of topic j by the sum of the popularity of all topics, which can be expressed as:

$$N\mathrm{Pop}(j) = \frac{\sum_{d} \theta_{\mathrm{d},j}}{\sum_{j} \sum_{\mathrm{d}} \theta_{\mathrm{d},j}}$$

Through a similar way, the normalized yearly popularity for topic j in year t can be expressed as:

$$\mathrm{Yearly\_}N\mathrm{Pop}(j,t) = \frac{\sum_{\mathrm{d}|py(\mathrm{d})=t} \theta_{\mathrm{d},j}}{\sum_{j} \sum_{\mathrm{d}|py(\mathrm{d})=t} \theta_{\mathrm{d},j}}$$

where $py(\mathrm{d})$ denotes the publication year of document d.

Since the large difference exists between the numbers of scientific literature and patents, we normalize the popularity of the topics in order to clearly compare the relationship between them. Although it's clear that the number of scientific literature is much larger than the patents, bigger does not mean more popular. The reason behind this is the practitioners in both field are not exactly the same, that is, scientists and patent applications of nanoscience are not the same group of people.

Regression modeling

The dynamic patterns of lead–lag are analyzed by regression modeling. The popularity of each topic can be treated as the dependent variables, while time T can be treated as the independent variable. Due to the limited number of research works and patents in the early

Table 1 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

stage of nanoscience field, we count 1995 as the first year ($t = 1$) and simulate the topic evolution process between 1995 and 2015 for 20 years.

We adopt curvilinear regression to fit the topic popularity curves. According to the definition of Weierstrass approximation theorem, there must be a polynomial approximation which can get arbitrarily close to any continuous function as the polynomial order is increased. Let $f(x)$ be continuous on a real interval I. Then for any $\epsilon > 0$, there exists an n th-order polynomial $P_n(f, x)$, where n depends on ϵ, such that

$$|P_n(f, x) - f(x)| < \epsilon$$

for all $x \in I$. Thus, any continuous function can be approximated arbitrarily well by means of a polynomial. As the popularity of topics always changing over the years, we can view the trends of each topic as a curve which can be fitted by curvilinear regression. In curvilinear regression, usually an intrinsic linear model is assumed and data are fitted to this model using polynomial regression (Cohen et al. 2013). As we make $x = t$, then the popularity of publications of scientific research and patents would be the functions of t, the polynomial can be represented as $f(t, t^2, t^3, \ldots)$. The original model of the popularity trends of topics could be:

$$Y = \alpha + \sum_{i=1}^{m} \beta_i \left(t^i\right) + \varepsilon$$

In this model, the Y-intercept α indicates the first year popularity of each topic, β indicates the slope of the curve and m is the power of the polynomial. To find the accurate value of m, we first set $m = 3$ and the significant level as 0.05. Then, we fit the model by ordinary least squares (OLS). If all the variables pass the test, we then select the current setting of the power value. If any of the variables does not pass the test, we increase m by 1 and refit the model.

Results

Overview

Fifty topics were extracted from the combination of scientific literature and patent using the LDA model. Each topic was labeled using the top five ranked words (i.e., words with high probability in this topic). Table 2 shows these 50 topics, which demonstrate the major research topics as well as the technology topics in nanoscience.

Using an ontology of nanoscience proposed by Tanaka (2013) and with the help of experts in nanoscience, we further categorized the 50 topics into five major fields including:

- Nano-Physics (in purple, 12 topics) (e.g., Atomic Molecular, Condensed Matter, Fluids and Plasma, General, and Phenomenology)
- Nano-Chemistry (in red, 16 topics) (e.g., Chemical analysis, Chemical kinetics, Chemical reactions, Disperse systems, and Electrochemistry)
- Nano-Materials Science (in green, 12 topics) (e.g., Corrosion, Crystal growth, Materials preparation, Thin film deposition, and Treatment of materials)
- Nano-Biological (in blue, 5 topics) (e.g., Biomolecular structure, Biosensors, Molecular biophysics, and Subcellular structures)

Table 2 Fifty topics of scientific literature and patent in nanoscience

Topic ID	Keywords	Topic ID	Keywords
Topic 01	Titanium, Particle, Cathode, Nanoporous, Cylinder	Topic 26	Material, System, Sensor, Titanate, Generator
Topic 02	Ammonium, Block, Vacuum, Valve, Ester	Topic 27	Chloride, Magnesium, Antioxidant, Black, Thermal
Topic 03	Polymer, Catalyst, Alcohol, Sulfate, Polyvinyl	Topic 28	Solution, Aqueous, Calcium, Carbonate, Nanostructure
Topic 04	Carrier, Medium, Barrier, Antibiosis, Magnetic	Topic 29	Energy, Secondary, Natural, Nanorod, Purification
Topic 05	Heat, Phosphate, Additive, Insulation, Electrolyte	Topic 30	Polyurethane, Retardant, Yarn, Elastomer, Waterproof
Topic 06	Nanocomposite, Electric, Outer, Graphite, Cement	Topic 31	Particles, Ceramic, Nanomaterial, Nanofiltration, Fluoride
Topic 07	Zinc, Inorganic, Polyester, Etching, Immersing	Topic 32	Antibacterial, Nanowires, Lipid, Disease, Food
Topic 08	Precursor, Surfactant, Nanopowder, Apparatus, Dispersant	Topic 33	Resin, Source, Molecule, Beam, Nanocarbon
Topic 09	Tube, Nickel, Ethanol, Ether, Distilled	Topic 34	Nano, Metal, Carbon, Ethylene, Nanofibers
Topic 10	Fluorescent, Activated, Nanosheet, Infrared, Electron	Topic 35	Solar, Nitride, Micronano, Cloth, Cells
Topic 11	Alloy, Curing, Silicate, Cavity, Nanocalcium	Topic 36	Silver, Semiconductor, Nitrate, Nanoparticles, Cancer
Topic 12	Treatment, Carbide, Layers, Fuel, Zirconium	Topic 37	Acid, Electrode, Dissolving, Nano, Adhesive
Topic 13	Nanoparticle, Copolymer, Hydrogen, Electrodes, Guide	Topic 38	Porous, Metallic, Precipitate, Reflection, Tire
Topic 14	Hole, Silicone, Ultrasonic, Deposition, Plasma	Topic 39	Fiber, Forming, Body, Glycol, Nanosilver
Topic 15	Manufacturing, Machine, Integrated, Cable, Selfassembly	Topic 40	Substrate, Silicon, Organic, Pipe, Fabric
Topic 16	Composite, Manganese, Erric, Hybrid, Lead	Topic 41	Crystal, Diameter, Pressure, Doped, Nanoscale
Topic 17	Oxide, Powder, Iron, Lithium, Polypropylene	Topic 42	Epoxy, Circuit, Nanotitanium, Terephthalate, Nanosilica
Topic 18	Water, Dioxide, Liquid, Gold, Nanoparticles	Topic 43	Acrylic, Coreshell, Ring, Polylactic, Selenium
Topic 19	Carbon, Nanotube, Heating, Graphene, Shell	Topic 44	Aluminum, Magnetic, Cell, Copper, Nanoparticles
Topic 20	Quantum, Wire, Dots, Channel, Direction	Topic 45	Nanowire, Polyethylene, Alumina, Nanosilicon, Wave
Topic 21	Optical, Sulfide, Cobalt, Cadmium, Microspheres	Topic 46	Base, Anode, Separation, Peroxide, Mobile
Topic 22	Filter, Transparent, Nanocrystalline, Lubricant, Tungsten	Topic 47	Glass, Steel, Insulating, Nitrogen, Conducting
Topic 23	Rubber, Fluid, Imaging, Starch, Cotton	Topic 48	Portion, Drug, Nanoparticles, Medical, Wood
Topic 24	Hydroxide, Molybdenum, Nanostructures, Stearate, Amine	Topic 49	Solvent, Conductive, Salt, Silica, Nanoparticles
Topic 25	Nanofiber, Acetate, Ethyl, Catalytic, Voltage	Topic 50	Cellulose, Flame, Medicine, Sintering, Cutting

- Nano-Engineering (in gray, 5 topics) (e.g., Electrical engineering, Materials science for electrical and electronic engineering, and Mechanical engineering).

We see from the results that most of the topics belong to Nano-Chemistry. It shows that Nanoscience is the most closely related to Chemistry, followed by Nano-Physics and Nano-Materials Science. Because the field of nanoscience is not only focused on a single discipline, but also reflects the convergence of all scientific disciplines, it promises to have broad implications on society (Stevens et al. 2009). In the broadest sense, Nano-Chemistry employs the tools of synthetic chemistry and materials chemistry to make nanomaterials with size, shape, and surface properties that are designed to evoke a specific function and orchestrated to target a particular end use (Ozin and Cademartiri 2009). It is also being used in materials, physical, engineering, biological and medical applications. The multiple sharing areas between all of the nanoscience fields indicate the same core concepts, although the usages of those concepts are different.

Figure 2 shows the topic distribution of these 50 topics for scientific literature and patent over the last 20 years. We find that during the last 20 years from 1995 to 2015, the degree of topic popularity of patent and scientific literature represents a totally different situation.

In scientific literature, the topic distribution was comparably stable with lower amplitude oscillations. The proportion of topic 30 is the biggest but only gain 2%, and others' are almost at the same level. In patents, topics were distributed with topic 1 as the one sharing the most proportion which is three times as popular as it is in the scientific literature, then followed by topic 6, 21, 29, 34, 46, 50. Among these, a large proportion belongs to Nano-Materials Science which indicates that Nano-Materials Science not only receives more attention but also has more powerful ability to create economic value. In 2013, more than

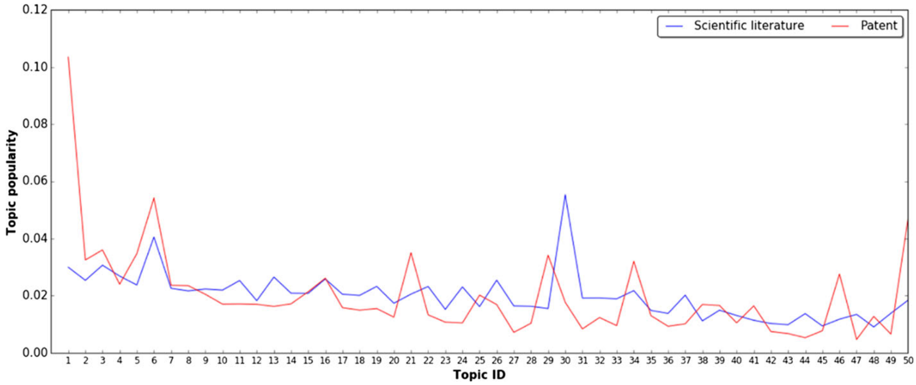


Fig. 2 Overview of the topic distribution for nanoscience scientific literature and patent (1995–2015). *Note*: horizontal axis represents topics, and vertical axis represents the topic distribution probability

1300 manufacturers around the world produced commercial products contain nanomaterials with a market value of 1.6 trillion US dollars and annual growth of 49% between 2009 and 2013 (Wohlleben et al. 2014). Topic 6 {Nanocomposite, Electric, Outer, Graphite, Cement} is an exception that both of patent and scientific literature paid a high attention to it.

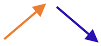
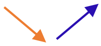
Topic popularity growth pattern

Due to the huge difference between the number of patents (164,183) and scientific literature (813,624), we are concerned about the changes in the popularity degree of the 50 topics in the last 20 years. The large gap of numbers between patents and scientific literature comes from the difference in a number of practitioners on the one hand, on the other hand, is the imbalanced development between science and technological innovation. In this case, we analyze the growth patterns of each topic by showing the percentage change over time periods. With the aim of showing a more clearly topic growth pattern, the regression algorithm is applied. Based on the regression algorithm, 50 pairs of fitted equations for the 50 topics are extracted.

From the fitted equation curves, four growth pattern types of topic popularity have been identified. As can be seen in Table 3, the dynamic trends of patent and scientific literature are complementary to each other. Most of the topics present the trend that the growth pattern of patent is in downward while scientific literature is in upward, others tend to be both in an upward trend or both in a downward trend, the situation of patent in upward while scientific literature in the downward trend is rare.

75% of the 11 topics in type A are both in upward trend through the last 20 years. The rest of them are downward first then go upward in patents, while the topics of scientific literature are in upward trend continuously. These topics contain mature, relatively well-characterized materials and technologies, and also represent a promising direction in Nanoscience and nanotechnology. For instance, titanium dioxide (TiO_2), as an important part of topic 1, has been used as a pigment for decades and has been studied in its nanoparticulate form since the 1980s. Demonstrated as topic 5, 29 and 35, energy and environment is one of the emerging topics for future nanoscale science and engineering

Table 3 Classifications for growth patterns of topic popularity

Type	Growth pattern	Classification	Trend	Example
A	Both in upward trend	Both in upward trend through the last 20 years		Topic 1, 5, 6, 16, 25, 29, 34, 50
		Patent in downward first then go upward versus scientific literature in upward trend		Topic 35, 45, 46
B	Both in downward trend	Both in upward first then go downward		Topic 17, 18, 20, 23
		Patent in downward trend versus scientific literature in upward first then go downward		Topic 11, 22, 24, 36, 43, 47, 48
C	Patent in upward versus scientific literature in downward trend	Patent in upward versus scientific literature in downward trend through the last 20 years		Topic 30
		Patent in downward first then go upward versus scientific literature in upward first then go downward		Topic 15, 41
D	Patent in downward versus scientific literature in upward trend	Patent in downward versus scientific literature in upward trend through the last 20 years		Topic 2, 3, 4, 7, 8, 9, 13, 14, 21, 32, 39, 42
		Patent in upward first then go downward versus scientific literature in upward trend		Topic 10, 12, 19, 26, 27, 28, 31, 33, 37, 38, 40, 44, 49

Left line indicates the patent and right is scientific literature

research reported in the WTEC (World Technology Evaluation Center) Panel Report on Nanotechnology research directions for societal needs in 2020 (Roco et al. 2011).

The declining popularity trends of both patent and scientific literature in type B indicate that most of these topics have become mature nano-materials and technologies after a long research and application history while it has been much more difficult to obtain obvious breakthrough innovations. For example, iron/lithium for use in thermal batteries (17), silver (topic 36), rubber (topic 23), and insulating glass (topic 47). Among these, silver is the most common nanomaterial used in products, followed by carbon-based nanomaterials and metal oxides (McIntyre 2012). Cellulose from rubber and cotton is the most abundant natural biopolymer and is readily available from renewable resources (Habibi et al. 2010).

In type C, there are only three topics that exhibit an upward trend in patent and a downward trend in scientific literature. The topics fitted this trending are primarily engineering-related. Obviously, the topic 15 {manufacture, machine, integrated, cable, selfassembly} is about nano-engineering, and others like fire retardant and heat resistant yarns (topic 30) and crystal diameter control (topic 41) are all related with nanotechnologies.

The topic popularity trend in type D is downward in patents while upward in the scientific literature. This type contains the largest proportion, accounting for 50% of the total number of topics. From the point of research data, the amount of scientific literature has reached nearly 810,000, far more than 160,000 patents, besides, the first scientific literature in the field of nanoscience was published much earlier than patents. As stated by Hullmann and Meyer in 2003, scientists do produce more, yet not necessarily "nanoscience", knowledge that is relevant to nanotechnology than can be traced by patent citations from nanoscience papers to nanotechnology patents (Hullmann and Meyer 2003).

Moreover, when the application of the scientific literature has been saturated and inventors can hardly find new ideas from the scientific literature for technical innovation, the patent number will gradually decline. In the meanwhile, in order to get new ideas and the latest progress, people would shift their focus to scientific research (Hu et al. 2007), so as to provide basic knowledge and theoretical support for technical innovation (Huang et al. 2015). In Mansfield's survey (1991), he estimates that 10% of industrial innovation would not have occurred, or would have occurred with great delay, without the contribution of academic research. Science is the soil of technology, but it never fully reveals how technology support science in turn. Scientific discovery is often inspired by goal oriented enterprises. The close combination of science and technology, as well as the feedback mechanism it implies, has a self-tuning mechanism. Technical failures often encourage the development of basic science (David 1972).

Another reason behind this is the safety concerns of nanotechnology. Although topics in type D belong to the various parts of the nano-domain, some of them are closely related to nanobiology, especially for the medical usage, for example, magnetic nanoparticles can be used for drug/gene delivery (topic 4) (Panyam and Labhasetwar 2003; Dobson 2006). With the increasing number of discussion on the safety of nanoparticles in medicine, scientists have pointed out that further research is needed to examine the cytotoxicity of nanoparticles against human cells and further incorporate them into therapeutic use (Wolfram et al. 2015).

The lead–lag relationships of patent and scientific literature

In order to demonstrate the four types of topic popularity growth patterns more intuitively, we draw the fitted curves as follows. In this research, we take topic 29, topic 36, topic 15 and topic 9 as examples, which can be seen in Fig. 3.

Integrating the above examples in Fig. 3, the trending lines of topic popularity are intersected. For type A, when the scientific literature tends to be constant, patents topic popularity will gradually rise and then flatten out. This kind of evolution pattern is the practical process of applying scientific research to technology development, then into patent products.

According to the fitted curves of all the topics, we can further identify the leader between patent and scientific literature for each topic and also the transcending times in the timeline. As shown in Table 4, lead means the topic is more popular in one of the publishers compared to the other at the time.

In most topics of nanoscience, the patent is the lead in the early stage of development. This is because nanoscience lays greater emphasis on applications than understanding. Above all, nanoscience and nanotechnology provide a new generation of scientific and technological method, as well as research and clinical tools and equipment. It creates and utilizes molecules for material assembly at the particle dimensional scale. Furthermore, the history of nanotechnology traces the development of the concepts and experimental work such as the invention of the scanning tunneling microscope in 1981 (Selin 2007). This determines that nanoscience is closely linked with engineering, and explains why the patent in most topics is leading at the beginning.

Nanoscience and nanotechnology are built for practical application and the enterprise is the key subject of transforming knowledge into products. In order to consolidate the commercial profits brought by technological innovation, the most common practice is to

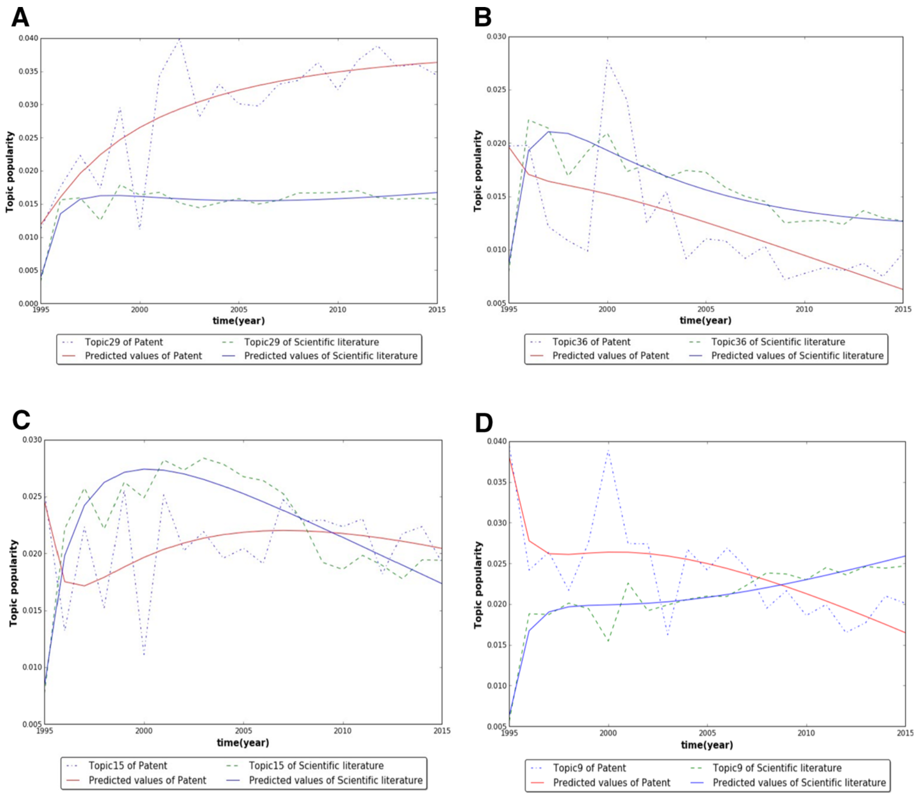


Fig. 3 Fitted curves for topic 29, topic 36, topic 15 and topic 9. *Note*: Y-axis shows the degree of popularity of patent and scientific literature while X-axis shows the time from 1995 to 2015

first document the latest research results on patent for the protection of intellectual property rights, especially for the novel research. On the contrary, scientific literature was not used as an innovation carrier at the beginning because of its openness to the public and lack of knowledge protection. It is noteworthy that most of the transcending occurred in both 1996–1997 and 2009–2013. In the early 1990s and after 2005, North America, the European Union, and other governments have introduced a series of policies and gave heavy funding to support and encourage academia for the field of nano-innovation and development (Paull et al. 2003; Hullmann 2008; National Research Council 2012). Meanwhile, nanotechnology firms rapidly increased in number between 1995 and 2007, with a total of 215 entries during that time period (Maine et al. 2014). For most topics, science occupied the lead position in the end. On the one hand, an innovation that can bring economic value is difficult to create. Related materials have not been commercialized due to the expensive production techniques employed, relative stability of the formed product and lifetime. On the other hand, the publication period of scientific literature is short and the cost is low. In order to establish their academic position in the field, scientists from university and research institutions will be more inclined to choose research papers to publish their research findings.

As a further step in the understanding of the interaction between patents and scientific literature, we analyze the changing keywords from both publishers. Taking topic 23 as an

Table 4 The leading relationship of topic popularity

Topics	Lead at beginning	Times of transcending	Transcending years	Lead in the end
1	Patent	0	–	Patent
2	Patent	0	–	Patent
3	Patent	0	–	Patent
4	Patent	1	2009	Scientific literature
5	Patent	0	–	Patent
6	Patent	2	1996, 1998	Patent
7	Patent	2	2005, 2010	Patent
8	Patent	1	2012	Scientific literature
9	Patent	1	2009	Scientific literature
10	Scientific literature	2	1997, 2005	Scientific literature
11	Patent	1	1996	Scientific literature
12	Patent	1	1999	Scientific literature
13	Patent	1	1997	Scientific literature
14	Patent	1	1996	Scientific literature
15	Patent	2	1996, 2009	Patent
16	Patent	1	1996	Scientific literature
17	Patent	1	1996	Scientific literature
18	Patent	1	1996	Scientific literature
19	Patent	1	1997	Scientific literature
20	Scientific literature	0	–	Scientific literature
21	Patent	0	–	Patent
22	Scientific literature	0	–	Scientific literature
23	Patent	1	1996	Scientific lite0072ature
24	Patent	1	1996	Scientific literature
25	Patent	0	–	Patent
26	Patent	1	1996	Scientific literature
27	Patent	1	1999	Scientific literature
28	Patent	1	1997	Scientific literature
29	Patent	0	–	Patent
30	Scientific literature	3	1997, 2003, 2013	Patent
31	Patent	1	1996	Scientific literature
32	Patent	1	1998	Scientific literature
33	Patent	1	1996	Scientific literature
34	Patent	0	–	Patent
35	Patent	1	1996	Scientific literature
36	Patent	1	1996	Scientific literature
37	Patent	1	1997	Scientific literature
38	Patent	1	1996	Patent
39	Patent	1	2013	Scientific literature
40	Patent	3	1996, 2000, 2009	Scientific literature
41	Patent	2	1996, 2003	Patent
42	Patent	1	2000	Scientific literature
43	Patent	1	1996	Scientific literature

Table 4 continued

Topics	Lead at beginning	Times of transcending	Transcending years	Lead in the end
44	Scientific literature	0	0	Scientific literature
45	Patent	2	2003, 2014	Patent
46	Patent	0	–	Patent
47	Patent	1	1996	Scientific literature
48	Patent	0	–	Patent
49	Patent	1	2001	Scientific literature
50	Patent	0	–	Patent

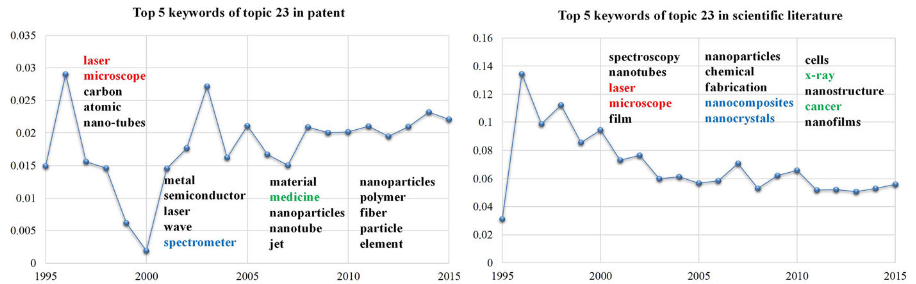


Fig. 4 Top 5 keywords of topic 23 in patent and scientific literature from 1995 to 2015. *Note*: Y axis shows the degree of popularity of patent and scientific literature while X axis shows the four time periods from 1995 to 2015

example, we divide the past 20 years into four periods with 5 years each. The top five frequency words of topic 23 are shown with the trending line of topic popularity.

Scientific literature is not as sensitive as patents on detecting the new ideas, that is to say, the content of patents can better reflect the frontier of nanoscience and nanotechnology. As can be seen in Fig. 4, there is a 5 years' time interval between the popular concepts of scientific research and patents. The keywords of patents are just a time period earlier (5 years) than that of scientific literature, which has been marked with the same color. This phenomenon indicates that patents can better represent the technology innovation. Although lack of economic impetus, scientific literature can continue to explore the forefront concepts identified by patents. For instance, it is suggested that when technical solutions proceed theoretical understanding, technology offers an enormous amount of empirical evidence awaiting scientists' explanation. Therefore, in order to create radical technological breakthroughs, shorten the product development cycle, and meet the growing demand for industrial innovation, research alliances have become increasingly important to industry engineers to absorb external scientific knowledge. Enterprise researchers must work with the university laboratory to absorb the latest knowledge to speed up the process of research and development.

The unconverted knowledge needs more attention, and the inventor of patents should pay more attention to the inquiry of scientific literature when the bottleneck of technology appears. There is still a big difference between patents and scientific literature, suggesting that a lot of knowledge can be transformed to the patents, but this transformation cannot be solved by a single field. From the contents of topics, although we can roughly categorize the topics, most topics are interdisciplinary and belong to multiple domains, which is determined by the nature of nanoscience. Nanotechnology is not only a technology but a set of techniques, some of the patents belong to the field of nanostructured materials, others are related to optoelectronics and medicine, there are also cases at the borderline with other fields, such as biotechnology and microelectronics (Meyer 2000a). Therefore, the research results also show that to re-examine the scientific research from the perspective of economic benefits and practical applications is an important basis for future scientific projects.

In most cases, technology awaits scientific breakthroughs. Analyzing the lead–lag between patent and scientific literature with keywords can be seen as the evolution of knowledge where each topic represents a subfield gathering similar concepts. Engineers often seek innovative solutions through the principle of knowledge similarity. In other words, they search for solutions that are near to existing solutions instead of searching for new ones. This limits the engineers to the barriers of self-awareness and their own knowledge systems without breaking through the boundaries of the field. Therefore, it reinforces the status quo and is unlikely to produce changes in inventive performance (Makri et al. 2010). Under these circumstances, interdisciplinary cooperation or multi-disciplinary integration of knowledge can be a multi-faceted way to promote technological innovation.

Conclusion

In this paper, we apply topic modeling and regression analysis to conduct a lead–lag analysis and identify different topic evolution patterns for huge collections of patents and scientific literature over the last 20 years (1995–2015) in the nanoscience. The attribution of topics shows that the most closely related field of nanoscience is Chemistry, followed by Physics and Materials Science. However, Nano-Materials Science receives more attention, which also implies that it has the economic potential of stimulating innovation.

Basic research provides a strong theoretical support for technology, and technology can be recognized to help and predict the future direction of research. As early as 1992, Rip proposed a two-branched model to reflect the interrelationship and emphasize the interrelated development and exploration process between science and technology. He regards science and technology as “dance partners”. Progressing at the same time, but in the course of development, they follow different steps. In this article, we use the data in the field of nanoscience to revisit the questions behind this dynamic relationship. We capture the lead–lag relationship between patent and scientific literature in nanoscience from the fitted equation curves. In most topics of nanoscience, patent is the lead in the early stage of development and science occupied the lead position in the end. From a domain point of view, it is because that nanoscience and nanotechnology are more emphases on the application and method of development according to the definition and history. However the underlying reason is that science and technology are driven by different motives. The pressure of the product market and service promoted the development of technology. To give a clear response to customer needs or competitors’ competitive behavior, the first thing to do is to understand what they need to produce. In contrast, science is driven by the interests and goals of the researchers. Science does not focus on specific problems, it cannot be used as a simple direct investment in the development of technology (Nightingale 1998). More specifically, scientific discoveries provide important basic knowledge for additional but more direct research. As stated by Meyer (2000a), science plays an important but indirect role for technology development.

Our research can help scientists to analyze the transform patterns of innovation and give us a better understanding of the development of nanoscience and nanotechnology. Moreover, the findings promote the understanding of the chain-linked innovation model that the low degree of early nano-patenting could provide a basis for improving scientific investigation which in turn stimulate technological development.

One of the limitations of this study is that we overpass the potential internal links between scientific literature and patents, such as the citations and author-inventor relationship. In the future, we will add the citation and author- inventor information both from scientific literature and patents to investigate the ways in which the transformation of scientific discovery into technological innovation occurs. Also, we plan to apply emerging topic detection separately on each dataset which could further reveal the concept evolution process in both publishers over time.

Acknowledgements This work is supported in part by Major Projects of National Social Science Foundation of China (No. 14ZDA063).

References

Bassecoulard, E., & Zitt, M. (2004). Patents and publications. In *Handbook of quantitative science and technology research* (pp. 665–694). Netherlands: Springer.

Bhattacharya, S., Kretschmer, H., & Meyer, M. (2003). Characterizing intellectual spaces between science and technology. *Scientometrics, 58*(2), 369–390.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3,* 993–1022.

Branstetter, L., & Ogura, Y. (2005). *Is academic science driving a surge in industrial innovation? Evidence from patent citations* (No. w11561). National Bureau of Economic Research.

Caraça, J., Lundvall, B. Å., & Mendonça, S. (2009). The changing role of science in the innovation process: From Queen to Cinderella? *Technological Forecasting and Social Change, 76*(6), 861–867.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. London: Routledge.

Di Stefano, G., Gambardella, A., & Verona, G. (2012). Technology push and demand pull perspectives in innovation studies: Current findings and future research directions. *Research Policy, 41*(8), 1283–1295.

Dobson, J. (2006). Gene therapy progress and prospects: Magnetic nanoparticle-based gene delivery. *Gene Therapy, 13*(4), 283.

Gibbons, M., & Johnston, R. (1974). The roles of science in technological innovation. *Research Policy, 3*(3), 220–242.

Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations. *Scientometrics, 58*(2), 415–428.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl 1(1)), 5228–5235.

Guan, J., & Zhao, Q. (2013). The impact of university–industry collaboration networks on innovation in nanobiopharmaceuticals. *Technological Forecasting & Social Change, 80*(7), 1271–1286.

Habibi, Y., Lucia, L. A., & Rojas, O. J. (2010). Cellulose nanocrystals: Chemistry, self-assembly, and applications. *Chemical Reviews, 110*(6), 3479–3500.

Hassan, M. H. (2005). Small things and big changes in the developing world. *Science, 309*(5731), 65–66.

Hsu, C.-W., Lien, Y.-C., & Chen, H. (2015). R&D internationalization and innovation performance. *International Business Review, 24*(2), 187–195.

Hu, D., Chen, H., Huang, Z., & Roco, M. C. (2007). Longitudinal study on patent citations to academic research articles in nanotechnology (1976–2004). *Journal of Nanoparticle Research, 9*(4), 529–542.

Hu, B., Dong, X., Zhang, C., Bowman, T. D., Ding, Y., Milojević, S., et al. (2015). A lead–lag analysis of the topic evolution patterns for preprints and publications. *Journal of the Association for Information Science and Technology, 66*(12), 2643–2656.

Huang, C., Notten, A., & Rasters, N. (2011). Nanoscience and technology publications and patents: A review of social science studies and search strategies. *The Journal of Technology Transfer, 36*(2), 145–172.

Huang, M.-H., Yang, H.-W., & Chen, D.-Z. (2015). Industry–academia collaboration in fuel cells: A perspective from paper and patent analysis. *Scientometrics, 105*(2), 1301–1318.

Hullmann, A. (2008). *European activities in the field of ethical, legal and social aspects (ELSA) and governance of nanotechnology* (p. 1). Brussels: European Commission.

Hullmann, A., & Meyer, M. (2003). Publications and patents in nanotechnology. *Scientometrics, 58*(3), 507–527.

Igami, M., & Okazaki, T. (2007). Capturing nanotechnology's current state of development via analysis of patents. *OECD science, technology and industry working papers.*

Kostoff, R. N., Koytcheff, R. G., & Lau, C. G. Y. (2007). Technical structure of the global nanoscience and nanotechnology literature. *Journal of Nanoparticle Research, 9*(5), 701–724.

Li, R., Chambers, T., Ding, Y., Zhang, G., & Meng, L. (2014). Patent citation analysis: Calculating science linkage based on citing motivation. *Journal of the Association for Information Science and Technology, 65*(5), 1007–1017.

Li, X., Hu, D., Dang, Y., Chen, H., Roco, M. C., Larson, C. A., et al. (2009). Nano mapper: An Internet knowledge mapping system for nanotechnology development. *Journal of Nanoparticle Research, 11*(3), 529–552.

Li, X., Lin, Y., Chen, H., & Roco, M. C. (2007). Worldwide nanotechnology development: A comparative study of USPTO, EPO, and JPO patents (1976–2004). *Journal of Nanoparticle Research, 9*(6), 977–1002.

Logothetidis, S. (Ed.). (2012). *Nanostructured materials and their applications*. New York: Springer.

Maine, E., Thomas, V. J., Bliemel, M., Murira, A., & Utterback, J. (2014). The emergence of the nanobiotechnology industry. *Nature Nanotechnology, 9*(1), 2–5.

Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal, 31*(6), 602–628.

Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy, 20*(1), 1–12.

McINTYRE, R. A. (2012). Common nano-materials and their use in real world applications. *Science Progress, 95*(1), 1–22.

Mei, Q., & Zhai, C. (2005, August). Discovering evolutionary theme patterns from text: An exploration of temporal text mining. In *Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining* (pp. 198–207). ACM.

Meyer, M. (2000a). Does science push technology? Patents citing scientific literature. *Research Policy, 29*(3), 409–434.

Meyer, M. (2000b). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics, 49*(1), 93–123.

Meyer, M. (2002). Tracing knowledge flows in innovation systems. *Scientometrics, 54*(2), 193–212.

Meyer, M. (2006). Knowledge integrators or weak links? an exploratory comparison of patenting researchers with their non-inventing peers in nano-science and technology. *Scientometrics, 68*(3), 545–560.

Nallapati, R., Shi, X., McFarland, D. A., Leskovec, J., & Jurafsky, D. (2011, July). LeadLag LDA: Estimating topic specific leads and lags of information outlets. In *ICWSM.*

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

National Research Council. (2012). *Rising to the challenge: US innovation policy for the global economy.* Washington DC: National Academies Press.

Nemet, G. F. (2009). Demand-pull, technology-push, and government-led incentives for non-incremental technical change. *Research Policy, 38*(5), 700–709.

Nightingale, P. (1998). A cognitive model of innovation. *Research policy, 27*(7), 689–709.

Ozcan, S., & Islam, N. (2017). Patent information retrieval: Approaching a method and analysing nanotechnology patent collaborations. *Scientometrics, 111*(2), 1–30.

Ozin, G. A., & Cademartiri, L. (2009). Nanochemistry: What is next? *Small (Weinheim an der Bergstrasse, Germany), 5*(11), 1240–1244.

Panyam, J., & Labhasetwar, V. (2003). Biodegradable nanoparticles for drug and gene delivery to cells and tissue. *Advanced Drug Delivery Reviews, 55*(3), 329–347.

Paull, R., Wolfe, J., Hébert, P., & Sinkula, M. (2003). Investing in nanotechnology. *Nature Biotechnology, 21*(10), 1144–1147.

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research, 10*(5), 715–728.
Řehůřek, 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* (pp. 45–50).
Rip, A. (1992). Science and technology as dancing partners. In *Technological development and science in the industrial age* (pp. 231–270). Dordrecht: Springer.
Roco, M. C., & Bainbridge, W. S. (2005). Societal implications of nanoscience and nanotechnology: Maximizing human benefit. *Journal of Nanoparticle Research, 7*(1), 1–13.
Roco, M. C., Mirkin, C. A., & Hersam, M. C. (2011). Nanotechnology research directions for societal needs in 2020: Summary of international study. *Journal of Nanoparticle Research, 13*(3), 897–919.
Sampat, B. N., & Ziedonis, A. A. (2004). Patent citations and the economic value of patents. In *Handbook of quantitative science and technology research* (pp. 277–298). Netherlands: Springer.
Selin, C. (2007). Expectations and the emergence of nanotechnology. *Science, Technology and Human Values, 32*(2), 196–220.
Shaparenko, B., Caruana, R., Gehrke, J., & Joachims, T. (2005). Identifying temporal patterns and key players in document collections. In *Proceedings of the IEEE ICDM workshop on temporal data mining: Algorithms, theory and applications (TDM-05)* (pp. 165–174).
Shi, X., Nallapati, R., Leskovec, J., McFarland, D., & Jurafsky, D. (2010). Who leads whom: Topical lead–lag analysis across corpora. In *NIPS workshop.*
Stevens, S. Y., Sutherland, L. M., & Krajcik, J. S. (2009). *The big ideas of nanoscale science and engineering*. NSTA press.
Swan, R., & Jensen, D. (2000, August). Timemines: Constructing timelines with statistical models of word usage. In *KDD-2000 workshop on text mining* (pp. 73–80).
Tanaka, M. (2013). Toward a proposed ontology for nanoscience. In *Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI*, October.
Tussen, R., Buter, R., & Van Leeuwen, T. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics, 47*(2), 389–412.
Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics, 54*(3), 399–420.
Wohlleben, W., Kuhlbusch, T. A., Schnekenburger, J., & Lehr, C. M. (Eds.). (2014). *Safety of nanomaterials along their lifecycle: Release, exposure, and human hazards*. Boca Raton: CRC Press.
Wolfram, J., Zhu, M., Yang, Y., Shen, J., Gentile, E., Paolino, D., et al. (2015). Safety of nanoparticles in medicine. *Current Drug Targets, 16*(14), 1671–1681.
Zhang, G., Feng, Y., Yu, G., Liu, L., & Hao, Y. (2017). Analyzing the time delay between scientific research and technology patents based on the citation distribution model. *Scientometrics, 111*, 1–20.
Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management, 42*(6), 1513–1531.