The Role of Articles in Science–Technology Relationship: A Topic Analysis of Non-patent Literature (NPL) References

Article *in* Serials Review · September 2022

DOI: 10.1080/00987913.2022.2127403

Ali Mansouri
University of Isfahan

Serials Review

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usrv20

The Role of Articles in Science–Technology Relationship: A Topic Analysis of Non-patent Literature (NPL) References

Fatemeh Fadavi Hoseini & Ali Mansouri

To cite this article: Fatemeh Fadavi Hoseini & Ali Mansouri (2022): The Role of Articles in Science–Technology Relationship: A Topic Analysis of Non-patent Literature (NPL) References, Serials Review, DOI: 10.1080/00987913.2022.2127403

To link to this article: https://doi.org/10.1080/00987913.2022.2127403

Published online: 30 Sep 2022.

SERIALS REVIEW
https://doi.org/10.1080/00987913.2022.2127403

The Role of Articles in Science–Technology Relationship: A Topic Analysis of Non-patent Literature (NPL) References

Fatemeh Fadavi Hoseini and Ali Mansouri

Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran

ABSTRACT

Patents with non-patent literature (NPL) references indicate how the link between science and technology interact. Using topic modeling, this paper investigated the thematic relationship between patents and their cited articles in the field of Nanotechnology. For this purpose, patents in the field of nanotechnology (IPC Class: B82) were obtained from the United States Patent and Trademark Office from 1985 to 2019. Then, NPL references listed in "Other References" section of the patents was extracted and abstract of the NPL references was retrieved from Scopus database. R software, topic modeling, and Latent Dirichlet Allocation algorithm were used to analyze the data. Results showed that most of the subclasses in Nanotechnology use few NPL references and are more dependent on patents. In total, NPL references account for only 36% of patent citations. The topics of the NPL references in this field (nanotechnology) belonged to six categories: Physics, Electricity, Chemistry, Cellular and Molecular Biology, Medicine, and Nanotechnology. Consequently, it seems that nanotechnology patents are more technology-driven, and a medium to low relationship exists between science and nanotechnology. The topic modeling of NPL references uncovered that nanotechnology patents have been more influenced by non-nano scientific.

KEYWORDS
Citation analysis; nanotechnology; non-patent literature; patents; topic modeling

Introduction

Science and technology jointly determine the evolutionary path of scientific innovation, with an increasingly close relationship between them (Xu et al., 2020). University–industry collaboration is a widely adopted strategy of R&D collaboration with external partners (Lin & Yang, 2020). In other words, university and industry are two effective institutions that contribute to the social, cultural, political and economic development of society. In addition to educational and research activities, the university has established its role in economic development through joint R&D activities (Etzkowitz & Leydesdorff, 2000). The university provides the research needs and manpower of the industry by training specialists and publishing scientific outputs. The contractual link between universities and the experimental approaches of technology are mutually desirable and complementary. The former reinforces problem-oriented thinking and the latter fosters solution-oriented thinking, both of which are useful in innovation activities (Nakagawa et al., 2017). These approaches have led to the formation of concepts such as the relationship between industry and academia, research and development, science and technology, etc.

Various models have been proposed to assess how the university and industry relate. The triple helix model (TH1, TH2, TH3) is one of the most widely used models in this field. In the TH1 model, industry and academia (the university) are under the control of a third party, the strong government. In the case of TH2, three separate entities are established and there are limited relationships between them (Etzkowitz & Leydesdorff, 2000). In the TH3 model, these institutions overlapped and interacted with each other (Etzkowitz, 2007). The triple helix model welcomes the presentation of the design to extend the model to more than three turns and states that the N-tuple of helices can be considered for it (Leydesdorff, 2012). Accordingly, with the addition of the "media-based and culture-based public" and "civil society" to this helix by (Carayannis & Campbell, 2010), the Quadruple helix model was formed. A Quintuple Helix model followed with the addition of the environment (natural environment) to the previous models. This model clearly shows that the knowledge production and utilization as well as innovation must be in

CONTACT Ali Mansouri a.mansouri@edu.ui.ac.ir Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran.

the context of the natural environment of society or similar to it (Carayannis & Campbell, 2010).

In a knowledge-based economy, the relationship between university and industry is based on the relationship between science and technology. Opinions and models have been presented on the interaction and action of science and technology. Based on the science precedence model over technology, or the linear model of the science–technology relationship, significant advances by academic researchers are almost immediately transformed into industry innovation (Meyer, 2000). However, empirical evidence refutes this theory for three reasons: (1) the instrumental role of technological progress in inferring scientific interpretations; (2) the importance of technology-based tools in scientific research; and (3) experimental evidence that technological change is often the result of initiative and experimentation rather than scientific theories and methods (Verbeek et al., 2002). By rejecting this model, a two-stream model was proposed that emphasizes the interaction of technology on science and science on technology. Price likens this kind of interaction to "dancing partners" and Toynbee to a "pair of dancers" (de Solla Price, 1965). The same attitude of independent but coordinated activity of science and technology became the basis of the two-branched Rip model. This model provides a specific model for scientific exploration and exploitation and sees a beneficial relationship between science and technology as a result of the simultaneous evolution of the two (Rip, 1992). Based on the complexity of the science–technology interaction, Brooks (1994) argues that this relationship is different in different fields and that it is thought of as two strands of DNA that can be independent but cannot function until they mate. Explaining that science and technology are two parallel streams of cumulative knowledge that are highly interdependent and interconnected their internal connection is stronger than the interconnectedness. The complexity of this relationship extends to the Network Model, which has increased participation in the supply and demand and exploitation of knowledge between producers and users of knowledge. As actors involved in technology grow and become heterogeneous (Verbeek et al., 2002).

In addition to the above, researchers such as Meyer (Meyer, 2000; Verbeek et al., 2002) and Leydesdorff (2004) concluded that the relationship between science and technology in various fields takes place in different ways. This connection is strong in some areas and weak in others. Sometimes the advancement of science drives technology forward, and sometimes technology drives scientific advancement. The scientific dependence of technologies in some fields is high and in some is low. Consequently, the result of one pattern cannot be generalized to the rest. According to the above, in this research, an attempt has been made to study the relationship between science and technology in the new field of nanotechnology. To measure this type of relationship, guidelines and indicators have been developed so far. The most important indicators were prepared by the Organisation for Economic Co-operation and Development (OECD) and became the basis for the development of indicators by other organizations and countries. These indicators are composed of various elements such as Research and Development (Frascati Manual 2002, 2007), Innovation (OECD/Eurostat, 2005), Human Resources (OECD/Eurostat, 1995), and Patents.

Publications and patents are the key outcomes of invention and research. Publications are the main channel for documentation of scientific findings, and patents reflect the development of new technologies. They provide important information on scientific, technical and conceptual development in an area (Kumari et al., 2021; Sun & Ding, 2018). In the context of exploring the evolution of technology, patents have a history of being used as proxies for both technology and innovation. The underlying assumption is that most R&D activities generate innovations, which are then protected by innovators as their intellectual assets in the form of a patent (Pereira & Quoniam, 2017; Smojver et al., 2020; Watanabe et al., 2001).

In the triple helix model, patents are considered as positioned in terms of the three social coordination mechanisms of (1) wealth generation on the market by industry; (2) legislative control by government; and (3) novelty production in academia. Whereas patents are output indicators for science and technology, they function as input into the economy (Leydesdorff, 2012). According to this model, patents, especially their citations, are an important indicator of the transformation and evaluation of the relationship between science and technology.

The volume of knowledge contained in patents, as a type of scientific literatures, is substantial, with the current number of patents over six million and increasing by roughly 150,000 patents per year (Smojver et al., 2020). Patents have many different characteristics from papers. One difference is related to citations. Therefore, the analysis methods cannot be the same as paper (Lu et al., 2020). Regarding the importance of citation, in Price model, science and technology are to be seen as dancing partners and their interactions as dances, patent citations tell the observing audience where the dance could take place (de Solla Price, 1965). The nature of the citations indicates the strengths and weaknesses of the relationship and the prevalence of each (Meyer, 2000) and is one of the most common indicator for measuring patent quality and knowledge flow from science to technology, and from technology to technology (Alcácer et al., 2009; Chang et al., 2017; Du et al., 2019; Ke, 2018; Meyer, 2000; Verbeek et al., 2002).

Due to the importance of citations, the US Patent Law required the citation of previous sources to prove novelty and other technical criteria for patent registration based on which it determines the degree of overlap, relevance, limitation, rejection or approval and registration (Alcácer et al., 2009). Mentioning previous sources or the same patent cited in the patent is called citation to the past (backward citation) and is like citing an article.(Harhoff et al., 2003) The wider the scope and domain of the patent, the more resources these citations contain (Harhoff et al., 2003). These citations, which represent the patent's previous literature, are divided into two categories: patent references and non-patent literature (NPL) references (Ke, 2018). Patent Reference are patents that are related to the topic of the citing patent and can complement it, and non-patent Reference generally involve scientific publications, technical standards, conference proceedings, clinical trials, and books related to the field of the citing patent.

Scholarly works cited by patents, especially journal references, highlight research that focuses on the extent to which technological developments are situated within the vicinity of scientific knowledge, notwithstanding the questions indicating and interpreting the influence from science to technology (Harhoff et al. 2003; Jefferson et al. 2018; Qu & Zhang 2020). Patents with NPL references contain more complex and fundamental knowledge and are of significantly higher quality (Cassiman et al., 2008; Qu & Zhang, 2020). The technological impact of scientific works can be studied through the citations they receive from patents, just like the scientific impact of articles can be analyzed through the citations (Gerrero-Bote et al., 2019). In other words, NPL references determine how the link between science and technology interacts and interprets (Verbeek et al., 2002). "Science-based" patents are believed to contain a relatively high number of non-patent references (Harhoff et al., 2003). In fact, a large part of the patented inventions are based on scientific advances, often published in scientific journals (Gerrero-Bote et al., 2019). These references are available in the OTHER REFERENCE section of United States Patent and Trademark Office (USPTO) patents. Patent references include previous patents related to the current patent and are mostly related to the specifications of patents created based on previous technologies.

As an invaluable source of comprehensive empirical information, patent documents uncover the technical specifications of the inventions and may contain explicit references to the non-patent resources, which indicate a proxy measure for the industrial relevance of research (Tijssen, 2001). By studying the nature and distribution of citations in patents, we can provide very useful information about the impact of science on technology, whether those patents are knowledge-based or technology-based. For this reason, to examine the relationship between science and technology and the flow of knowledge, the analysis of NPL references is one of the most common methods (Chang et al., 2017; Jefferson et al., 2018; Ke, 2018; Li et al., 2014; Meyer, 2000; Verbeek et al., 2002). In addition, the results of numerous studies in this field indicate the importance of studying citations in patents: examining the time lag between turning an article into a patent (Du et al., 2019), studying international and organizational collaboration, identifying the inventor-author, and the scientific productions of the inventors (Breschi et al., 2007; Callaert et al., 2006; Leydesdorff, 2004), research on policy and decision-making in the field of science and technology and the identification of emerging core technologies using citation network analysis (Cho & Shih, 2011; Uzun, 2006), and the commercial value of patents by studying the hidden impact of patent citations, technological strength, and the market value of patents (Chen et al., 2007; Hall et al., 2005; Smith, 2014). Due to the limited studies on the topic analysis of citations from patents to NPL, in the present study, according to two important indicators of article and patent for evaluating the relationship between science and technology (Meyer, 2000), we investigate the thematic relationship of Nanotechnology patents with their non-patent references (cited articles).

Research methodology

In this research, two groups form the research population:

1. All patents in the field of Nanotechnology registered in USPTO from 1985 to 2019, data were searched and downloaded by using the query

Table 1. Logarithm of the correctness of topic modeling based on the number of different subjects.

K	30	40	50	60	70	*80
LL	−22,723,523	−22,332,744	−22,152,955	−21,976,503	−21,715,095	−21,557,374

ICL/B82$. The class B82 is related to Nanotechnology,[1] which has several subclasses, so the $sign at the end of the query was used to retrieve all the patents in this field. The result of this search strategy led to 8324 patents. We used a researcher made software (Patent Extractor Final) to download patent information and extract citation data.

2. The articles cited in the "Other Reference" section of the patents are the second group of the research community. Unlike the patents, which were not sampled, the articles were extracted from the Scopus database by random sampling, and 3305 articles comprised the sample articles for this study. A topic modeling method was used to identify the topics of the articles cited.

Since the most critical source of information in patents is presented in the form of tacit knowledge (Durack, 2004), application of a topic model can interpret information of such unstructured documents and improve ability to understand the pattern and distribution of the topics and keywords from a corpus of documents (Kim et al., 2016). Topic modeling is useful when dealing with large collections of documents and not able to find topics manually in the body of text (Akhtar et al., 2019, p. 22). A topic model, in the form of an unsupervised algorithm, can be used to expose hidden topics from the set of documents (Suominen et al., 2017). To be more precise, this method is a statistical model for discovering hidden topics that can automatically generate clusters of similar documents and then compare the results with the actual structure. The topic model is widely used in computer science to apply text-mining techniques to analyze text documents (Kumari et al., 2021). The most common and popular method for topic modeling is the Latent Dirichlet Allocation (LDA). LDA is a form of topic model that calculates certain numbers of topics by considering the probability distribution of terms associated with them (Blei et al., 2003). It is a three-layer Bayesian model and a possible generator model that is widely used in the field of data recovery. The basic idea of this model is that each document in a set is a random combination of hidden topics, and each hidden subject is identified by a distribution of a word or term (Yau et al., 2014).

For topic modeling and clustering, a combination of R software, LDA and K-means algorithms were used. By extracting information about each patent, the following steps were performed to identify the topics of NPL references using the LDA algorithm:

A. Importing data and reading it by the program and converting the text to a bag of words and forming a text corpus;
B. Pre-processing step which includes deleting dots and symbols, numbers, extra spaces, stop words;
C. Creating document-word matrix: for topic modeling, the text must be converted to a standard format. This standard format is the document-word matrix that is generated using the "Document Term Matrix()" function and the frequency of given word in each document is determined. In this structure, word order is not important and depends on the Bag of Words (Ponweiser, 2012).
D. Topic modeling: In the topic modeling process, determining the right number of clusters is very important. However, no fixed way to choose the correct number of clusters exists. One of the common methods is to estimate the logarithm in Log likelihood (LL) and to plot the growth of the logarithm for different clusters. LL grows rapidly for small clusters and declines with increasing number of clusters. The number of clusters (K) is appropriate that indicate the end of rapid LL growth and can provide an appropriate interpretation of the topics (Kosinski et al., 2016). In this study, Log likelihood and Gibbs method were used. To obtain the appropriate K, models with the number of clusters of 30, 40, 50, 60, 70, and 80 were implemented and based on the LL values (Table 1), the best number of topics (i.e. 80) was considered.

The results of the implementation of the topic modeling algorithm are visible as two matrices (Figure 1):

1. Word-topic assignment matrix in which one dimension is the matrix of words in corpus and the other is the subjects, which their number is determined by the researcher. The numbers in the

[1]For more information, B82 category, see https://www.wipo.int/classifications/ipc.

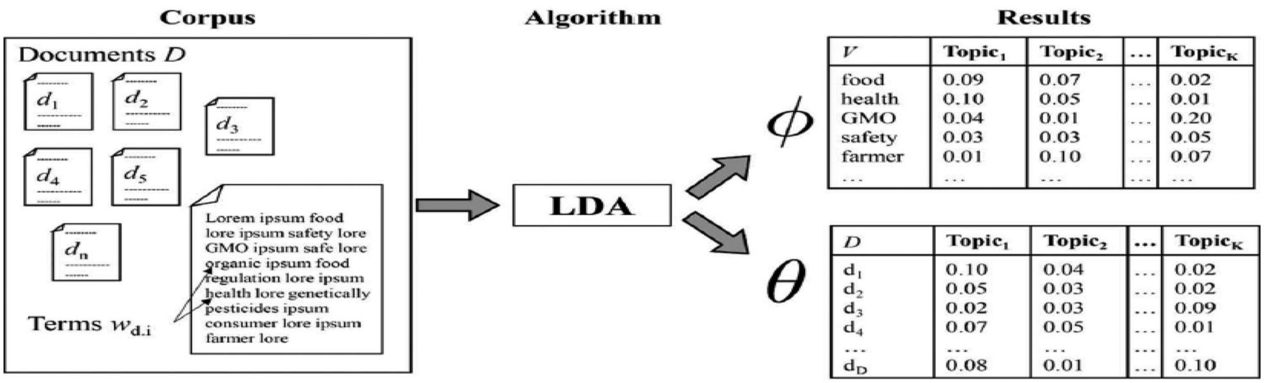

Figure 1. Input and output of the LDA algorithm (Maier et al., 2018).

matrix indicate the probability of a given word in each subject. In this way, the researcher can identify the top words in each topic and thus interpret the topics. In this study, high-frequency words and the top 15 words in each topic were used to determine a title for each topic.

2. Document-topic assignment matrix; this matrix includes two dimensions of the document (the number of documents in the corpus) and topics (the number of the subjects specified by the researcher). The numbers in each cell indicate the probability of the topics in each document and determine what topics each document has and with what probability. In order to obtain φ and θ, it is necessary to consider the previous distribution for each. The previous distribution β is selected for φ and α for θ. These two previous distributions affect how the word-subject and document-subject are distributed (Maier et al., 2018). Regarding the parameters required for the implementation of the algorithm, the closer α is to 1 and β to 0, the fewer subjects the evidence belongs to, and the better the separation of subjects (Bittermann & Fischer, 2018). To run the LDA algorithm, the values for the α and β parameters were set to 1 and 0.001, respectively.

Findings

Topic relationship between nanotechnology patents and their NPL references

At this stage, the retrieved patents were divided according to the International Patent Classification (IPC), and the number of NPL references, patent references, and the determination of the link between science and technology in each class was determined according to Table 2. The relation between science and technology were calculated according to Narin's formula (Narin, 2000):

$$\text{Technology linkage} = \text{the number of technical documents cited in the patents}/\text{Total References} \times 100$$

$$\text{Science linkage} = \text{number of NPL references}/\text{Total References} \times 100$$

The findings showed that:

- Subclasses Footwear (A43), Construction of Roads, Railways, or Bridges (E01), Signaling (G08), and Generating or Transmitting Mechanical Vibrations in General (B06), have a 100% technology link and do not have NPL references.
- The subclass Treatment of Water, Waste Water, Sewage, Or Sludge (C02) has used both references equally. Patents in subclasses Foods or Foodstuffs (A23), Separation Of Solid Materials Using Liquids Or Using Pneumatic Tables Or Jigs; Magnetic Or Electrostatic Separation Of Solid Materials From Solid Materials Or Fluids; Separation By High-Voltage Electric Fields (B03), Organic Chemistry (C07), Refrigeration Or Cooling; Combined Heating And Refrigeration Systems; Heat Pump Systems; Manufacture Or Storage Of Ice; Liquefaction Or Solidification Of Gases(F25), Drying (F26), Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants (F02), and Checking-Devices (G07) cited NPL references more than patent references. Meanwhile, the subclass Drying (F26), with 69% scientific linkage, has the most citations to NPL references. For the rest of the patents, patent references were greater than NPL references; and the highest citation to patent references occurred in the subclass Animal Or Vegetable Oils, Fats, Fatty Substances Or Waxes;

Table 2. Patent references and NPL references in subclasses.

IPC	Number of patents	citations	NPL references	Patent references	Link with science	Link with technology
A43	1	3	0	3	0	100
E01	2	16	0	16	0	100
G08	1	8	0	8	0	100
B06	1	9	0	9	0	100
C11	1	43	1	42	2	98
C02	13	422	209	213	50	50
A23	2	183	93	90	51	49
C07	187	14,926	7461	7285	51	49
G07	1	35	18	17	51	49
F02	4	136	76	60	56	44
B03	10	206	126	80	61	39
F25	1	9	6	3	67	33
F26	1	13	9	4	69	31

Fatty Acids Therefore; Detergents; Candles (C11), with 98% technology linkage. In total, 36% of patent citations are NPL references and 64% are patent references (Table 2).

Identifying the topics of NPL references

After clearing the text and implementing the algorithm, it was revealed that the high-frequency words in the abstract were graphene, surface, color, light, cell, carbon nanoparticles, metal, quantum, DNA strand, clarity, electron, density, and energy. With a frequency of 2794, grapheme was the most widely used word. By implementing topic modeling and selecting 80 topics, the words with the most similarity are in the same topic and by using these words, the general meaning of the topic can be understood and a title can be determined for it (Table 3). Therefore, for each subject, 15 top terms (15 terms that had the most repetition in each topic) were selected and subclasses of the Nanotechnology were identified (Table 4).

Topic trend of articles

The following steps were taken to find topics that have become more popular over the years (hot topics) or reduced (cold topics) in Nanotechnology patents; topics were determined first; the average probability of each topic per year was then measured; next, based on the mean probability of the subject-document and based on linear regression, a graph was drawn; finally, hot and cold topics were identified based on the ascending and descending slope of the linear trend. Hot topics were identified by the highest linear range.

Top or hot topics with a positive slope of the linear trend (Figure 2) are: Carbon Nanotube Conduction (2), Combining & Separating Liquids (18), Medication (38), Genetic (60), Materials Engineering (31), Nanotechnology in Print & Art (27), Memory Design &Construction (69), Medicine (Cancer & Therapy) (30), Chemistry (23), and Food & Foodstuffs (29). The title and the top 15 words related to each topic are shown in Table 3.

Cold topics or topics with a sharply decreasing linear trend (negative slope) are Print & Image (22), Organic Chemistry (33), Quantum Physics (63), Electronic Microscopes & X-ray Topics (32), Biochemistry (70), Mineral Chemistry (7), Cellular Immunity (Antigen & Antibody) (48), Nanoparticles (65), Optic (1), and Electromagnetic Radiation (50) (Figure 3).

In order to investigate the growth or decline of topic areas and to explore the issue of paying much attention to some topic areas (hot topics) and lack of attention or little attention to other subject areas (cold topics) in a period of time, it is necessary to use methods and tools suitable for determining the "average probability of each topic" (θ) in a time interval. The linear regression method is a suitable tool for investigating the topic trending. The results of using this method can help identify hot and cold topics. This method, which is a supervised learning algorithm, models a predicted value according to independent variables and finds the relationship between those variables and the predicted value. However, regression models depend on the relationship between independent and dependent variables, as well as the number of variables they use.

Table 5 shows the regression analysis data of 10 hot topic areas. In this table, the results of calculating the average quantity of theta (θ) as the average probability of assigning each document to each topic in each year are presented. In this table, the number of topics is given in the first row and the year under review is given in the first column. Using the information in this table, it is possible to understand how the weight factor of each specific issue has changed during the period from 2011 to 2018. In addition, it is possible to examine the changes of this variable for

Table 3. Eighty topics of NPL references used by nanotechnology inventors.

No	Topic	No	Topic
1	Optic	41	Macromolecules
2	Carbon Nanotube Conduction	42	Enzymes & Amino Acids
3	Chemical Reactions	43	Power Control Devices
4	Melting Of Covalent Solids	44	Antibacterial
5	Solar Energy	45	Ionizing Radiation, Solar Cells
6	Artificial Fiber	46	Graphene & Electrical & Thermal Conductivity
7	Mineral Chemistry	47	Chemical Synthesis
8	Magnetic Waves	48	Cellular Immunity (Antigen & Antibody)
9	Cell Biology	49	Conductive & Semi-Conductive
10	Metals	50	Electromagnetic Radiation
11	Gases	51	Nanostructures & Electronics
12	Nanostructures & Nanoparticles	52	Tumor & Its Treatment
13	Disease & Treatment	53	Medication
14	Cellular Staining With Fluorescent	54	Electricity
15	Electrical Oscillation Measurement	55	Electric Current & Magnetic Field
16	Semiconductor	56	Superconductivity
17	Tumors & Circulation Of Nanoparticles In Its Treatment	57	Nanostructures In Electric Current
18	Combining & Separating Liquids	58	Heating
19	Viral Infection	59	Amino Acids & Catalysts
20	Measuring & Measuring Devices	60	Genetics
21	Molecular Biology	61	Radiation (Spectroscopy)
22	Print & Image	62	Nanotechnology In The Treatment Of Leukemia
23	Chemistry	63	Quantum Physics
24	Colored Glasses	64	Identify Cancer Cells &Treat
25	Peptide (Amino Acid)	65	Nanoparticles
26	Physical &Chemical Processes	66	Nanostructures &Peptide Synthesis
27	Nanotechnology In Print & Art	67	Electric Current &Heat
28	Molecular Interaction	68	Radiation (Optic)
29	Food & Foodstuffs	69	Memory Design &Construction
30	Medicine (Cancer & Treatment)	70	Biochemistry
31	Materials Engineering	71	Nanoparticles &The Immune System
32	Electronic Microscopes &Xray Topics	72	Electrolyte Processes
33	Organic Chemistry	73	The Human Immune System
34	Cell &Transfer Materials To It	74	Antibody
35	Polymers	75	Electromagnetic Waves
36	Electronic	76	Water &Its Properties (Pressure, Absorption, Etc.)
37	Electrostatic	77	Human Body (Organs)
38	Medication	78	Oxidation
39	Nanostructure	79	Electrophoresis &Its Therapeutic Role
40	Magnetism	80	Atomizer (Using Liquids For Surfaces)

Table 4. Top 15 words for top 10 topics.

Topic	Term	Title
Topic 2	Carbon, Nanotube, CNT, Density, SWNTS (single wall carbon nanotubes), SERS (surface-enhanced Raman scattering), Semiconducting, Raman, Walled, Vertically, Multi, Orders, Multiwalled, Electrically, Magnitude.	Carbon Nanotube Conduction
Topic 18	Morphology, Ionic, Nanocrystal, Copolymers, Solid, Liquid, Colloidal, Separate, Composition, MOL, Liquids, Excess, Morphologies, PIL (polymerized ionic liquid), BIS (budesonide inhalation suspension).	Combining & Separating Liquids
Topic 38	drug, strategies, therapeutic, leading, lipid, models, drugs, cause, response, liposomes, loaded, greater, rates, diseases, estimated.	Medication
Topic 60	DNA, Sequence, Sequencing, Complexes, Hybridization, Cyclodextrin, Origami, Kinetics, Level, Nanotechnology, Thiol, Complementary, Immobilized, Presence, Measurement.	Genetic
Topic 31	Composites, Composite, Presence, Application, Article, Polymerization, Fiber, Medium, Fibers, Compound, Described, Soft, Dimensional, Pre, Coefficient.	Materials Engineering
Topic 27	Nano, Paper, Physical, micro, Generation, Finally, Improvement, Providing, Systemic, Literature, Mechanism, Art, Targeting, Subsequently, Functionalization	Nanotechnology in Print & Art
Topic 69	Memory, Edge, Broad, Environment, Techniques, Make, Traditional, Base, Applied, Architectures, Requires, Critical, Robust, Processes, Currently.	Memory Design & Construction
Topic 30	Cancer, Cell, Expression, Human, Clinical, Vivo, Therapy, Breast, Disease, Treatment, Target, Resistance, SIRNA (Small interfering RNA), Death, Survival.	Medicine (Cancer & Therapy)
Topic 23	Phase, Liquid, Crystalline, Dynamic, Ethylene, Phases, Furthermore, Gel, Regime, Exhibits, Nematic, Depth, Experiment, anisotropic, decreases.	Chemistry
Topic 29	Induced, Promising, Constant, Known, Suggests, Leads, Irradiation, Considered, Damage, Means, Find, Interference, Evaluated, bone, Elastic.	Food & Foodstuffs

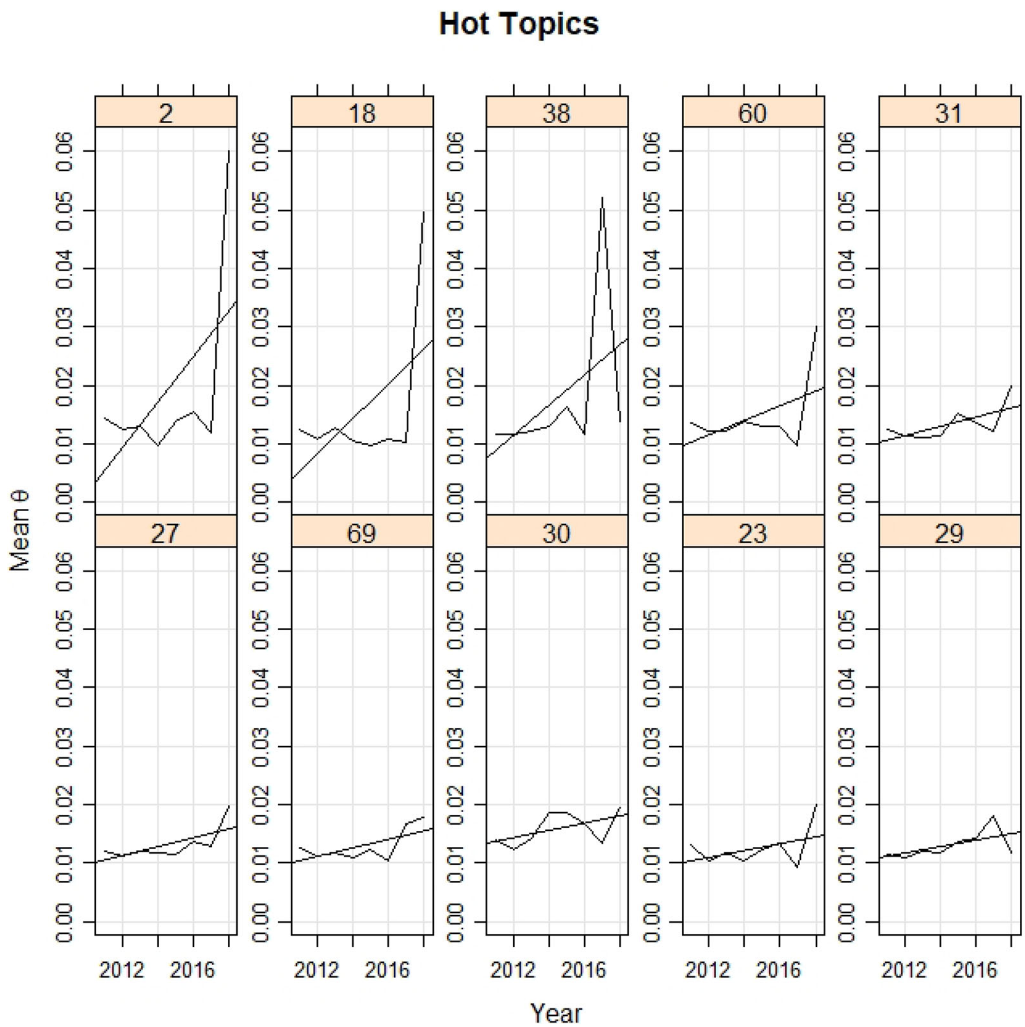

Figure 2. Ascending linear trend for 10 hot topics based on linear regression.

different topics in a given year. For example, Topic 38 in 2017 has the highest average probability and has received more attention. Besides, Topic Number 2 in 2018 has the highest theta value, so in 2018 it has been assigned a higher weighting factor and received more attention. Moreover, the examination of changes in each year showed that in 2018 subjects 2, 15, 18, 20, 23, 25, 27, 30, 31, 36, 43, 47, 53, 57, 60, 68, and 69 had the highest average probability (θ), and the diversity of topics in 2018 was more than other years.

By examining nonlinear regression, the subjects that did not have the same slope in increasing and decreasing the average probability of the subject-document probability and had successive peaks and declines were classified as nonlinear trends (Figure 4) and their downward and upward trends was determined. These topics include Metals (10) and Nanostructure (39) with a peak in 2012 and a decline in 2013; Polymers (35) and Nanostructures and Electronics (51) with the highest probability in 2013 and the lowest in 2011 and 2014, respectively; Cellular Immunity (Antigen and Antibody) (48) with the highest probability in 2014 and lowest in 2012; Magnetic Waves (8), Solar Energy (5) and Electromagnetic Waves (75) with a peak in 2015 and a decline in 2017, 2012, and 2018, respectively; Electrostatic (37), Gases (11), Tumors and Circulation of Nanoparticles in its Treatment (17), and Heating (58) with the highest probability in 2016 and the lowest in 2018; Macromolecules (41), Measuring and Measuring Devices (21), Identify Cancer Cells and Treat (64), Electric Current and Magnetic Field (55), Cell Biology (9), Nanoparticles and the Immune System (71), The Human Immune System (73), and Chemical Reactions (3) with the maximum popularity in 2017 and the maximum decrease in 2018; and Nanostructures in Electric Current (57), and Measuring and Measuring Devices (20) with a peak in 2018 and a decline in 2015 and 2016, respectively.

Topics that culminated in two points included Enzymes and Amino Acids (42) (peak points in 2015 and 2012), Medication (53) (2012 and 2018), Covalent Solids Melting (4) (2013 and 2017), Electrophoresis and its Therapeutic Role (79) (2014 and 2017), Ionizing Radiation, Solar Cells (45) (2016 and 2013), 68 (2016 and 2018), and Tumor and its Treatment (52) (2014 and 2016). The topic Power Control Devices (43) with peak points in 2011 and 2018 and a sharp decline in 2014 and 2016 showed the most fluctuations in the average probability of the topic-document. In total, the data analysis shows that many areas of nanotechnology have experienced growth and decline in different time periods, and the amount of attention paid to minor areas of this general area has been different according to time conditions. Some topics, which have been disregarded for many years, have been re-considered and the course of development of such topics has not been linear.

Discussion and conclusion

The relationship of scientific knowledge development to technological development is widely recognized as

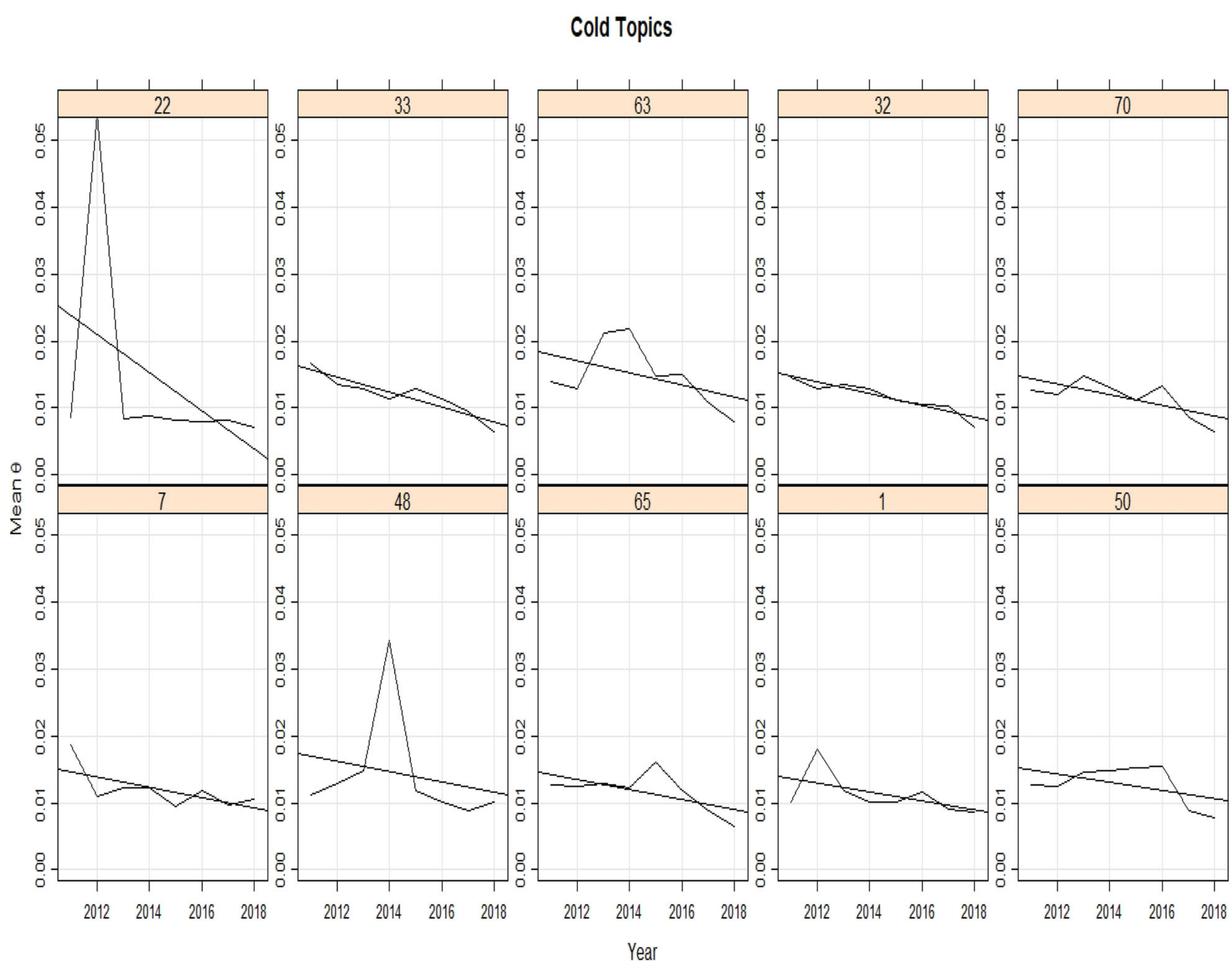

Figure 3. Descending linear trend for 10 cold topics based on linear regression.

Table 5. Average probability of document-subject of the top topics per year.

	2	18	38	60	31	27	69	30	23	29
2011	0.014	0.012	0.012	0.014	0.012	0.012	0.013	0.014	0.013	0.012
2012	0.013	0.011	0.012	0.012	0.011	0.011	0.011	0.012	0.010	0.011
2013	0.013	0.013	0.012	0.012	0.011	0.012	0.012	0.014	0.012	0.012
2014	0.010	0.010	0.013	0.014	0.011	0.012	0.011	0.019	0.010	0.012
2015	0.014	0.010	0.016	0.013	0.015	0.011	0.012	0.019	0.012	0.014
2016	0.015	0.011	0.012	0.013	0.013	0.014	0.010	0.017	0.013	0.014
2017	0.012	0.010	**0.052**	0.010	0.012	0.013	0.017	0.013	0.009	**0.018**
2018	**0.060**	**0.050**	0.014	**0.030**	**0.020**	**0.020**	**0.018**	**0.019**	**0.020**	0.012

Highest average probability (h) are the maximum values in each column which are mostly belongs to 2018 (indicated in **bold**).

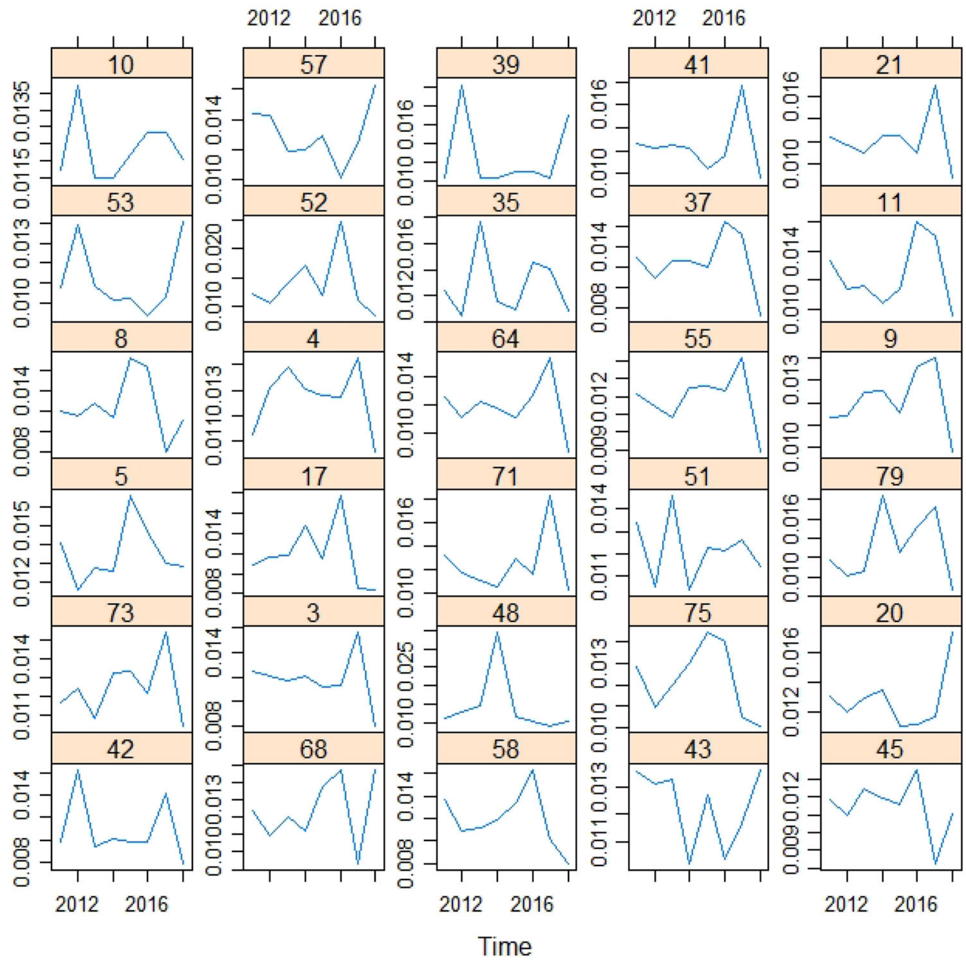

Figure 4. Nonlinear trend of topics based on nonlinear regression.

one of the most important and complex aspects of technological revolution (Han & Magee, 2018). Scientific publications and patents are usually considered good barometers of basic research and technical development, respectively (H. Y. Xu et al., 2017). Therefore, building topic linkages between scientific publications and patents for the purpose of understanding the relationships between science and technology is increasingly important (S. Xu et al., 2019). Bibliometric methods based on paper and patent analysis are the main methods to study the relationships between science and technology. However, the research on the interrelated and interactive topics between science and technology is still relatively rare (Sun & Ding, 2018). Moreover, NPL as an indicator of interaction between science and technology has recently been the subject of some debate (Qu & Zhang, 2020).

Citation analysis of patents is one of the methods to study how science affects knowledge. In this regard, NPL references are interpreted as the relationship between science and technology (Verbeek et al., 2002). The greater the number, the better and stronger the connection between science and technology (Callaert et al., 2006) and the scientific dependence of patents (Karki, 1997). Considering the importance of citation in determining the relationship between science and technology in a field, this paper investigated the thematic relationship between patents and their NPL references in the field of Nanotechnology (IPC Class: B82). To accomplish this research, NPL references were extracted from the database and analyzed using text mining and topic modeling. Some of the key findings of the present study are listed below.

In total, 36% of patent citations were NPL references and 64% were patent references. This result indicates that the link with technology is greater than the link with science and a weaker link exists between science and nanotechnology and it is consistent with the results of Meyer (2000), Hullmann and Meyer (2003) and C. Lee et al. (2015). This part of the result shows that the nanotechnology patents use less of the results of scientific publications (articles) and are usually dependent on technological findings. Meyer concluded that there is a very close interaction between science and technology and the interaction of citing scientific articles in patents reflects the closeness of the relationship. However, subsequent research has shown that this interaction is greatly reduced (Meyer, 2000). The second finding of the mentioned research is also true in the present study. The results indicate that in the field of nanotechnology, research outputs in the form of articles have less impact on the formation of technology in this field. In other words, the development path of this field in the industry has been different from its development path in the university.

A review of the first patent in the field of nanotechnology showed that it first entered the USPTO nanotechnology category in 2004; therefore, most citations were expected to be NPL references, and the results showed the opposite. Of course, this result is in line with the usual rate of citation of NPL references in patents, which shows that the rate of NPL references in patents is about 30% (S. Xu et al., 2019). In contrast to the expectation of citing more NPL references in the field of nanotechnology to reality, the country of the patent source may also be an influence (S. S. Lee et al., 2012; Michel & Bettels, 2001). However, another independent study suggested considering the impact of some citations made by the examiner, which may not be very relevant to the patent (Jaffe et al., 2000).

Abstracts of patented citations were analyzed using LDA-based topic modeling and 80 partial thematic clusters were uncovered. The topics of the NPL references in this field (nanotechnology) belonged to six categories: Physics, Electricity, Chemistry, Cellular and Molecular Biology, Medicine, and Nano. Examination of the scientific fields used in patents showed that this trend has experienced a downward and upward trend in the research period, at times it has received a lot of attention and at other times its importance has diminished. Some of the 80 minor areas, including the Discovery of Carbon Nanotubes (Topic 2) with special properties such as low toxicity and biocompatibility for use in biological applications and its optical and electronic properties are used for catalysts and conversion of light into electricity. This application has been emphasized in the research conducted by Bayda et al., (2019) and has caused the peak of this field (hot topic). In addition, due to its importance in nanotechnology and its application in most fields, also the high citation rate of articles on this subject (Braun et al., 1997) is one of the reasons for its peak. Topic 22 (Photography, Printing), with the lowest average probability of subject-document 0.007 in 2018, has lost its popularity and is considered the coldest subject. Given the importance of the role of tunnel and survey microscopes in the advent of nanotechnology, it can be understood that articles related to imaging and printing devices and everything used to better represent the Nanospace, at first, it was very important and gradually decreased and became one of the cold topics. In the nonlinear process, the electrical control devices (Topic 43) showed the most fluctuations in the mean probability of the subject-document.

In general, the reasons for the rise and fall of topics can be attributed to the needs of society and the life cycle of technology. For cold topics, the technology is nearing the end of its life, and for hot topics, the technology has reached its growth and maturity stage. For topics with a nonlinear trend, it can be said that their necessity and importance in society vary and fluctuate at different times. If universities consider the need for technology in society and address it in their research works, it is likely that the research outputs of universities will be used in patents, hence strengthening the process of communication between science and technology. On the other hand, with the development of the relationship between industry and universities, the needs of the industry sector are transferred to the university, and the influence of science on technology expands.

In the current research, it has been shown that this relationship is currently weak and needs to be strengthened while investigating the reasons. In addition to the aforementioned, the results of this research can be used in the selection and collection of resources, especially publications related to hot topics in libraries, hoping that libraries can introduce hot topics to inventors and industrialists and make them more aware of those topics, and publishers can further publish resources related to hot topics. Moreover, the results of the current research can be useful for policy makers in the field of technology and industry.

ORCID

Fatemeh Fadavi Hoseini http://orcid.org/0000-0001-8371-8939
Ali Mansouri http://orcid.org/0000-0003-4047-2697

References

Akhtar, N., Javed, H., & Ahmad, T. (2019). Hierarchical summarization of text documents using topic modeling and formal concept analysis. In *Data management, analytics and innovation* (pp. 21–33). Springer. https://doi.org/10.1007/978-981-13-1274-8_2

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, *38*(2), 415–427. https://doi.org/10.1016/j.respol.2008.12.001

Bayda, S., Adeel, M., Tuccinardi, T., Cordani, M., & Rizzolio, F. (2019). The history of nanoscience and nanotechnology: From chemical-physical applications to nanomedicine. *Molecules*, *25*(1), 112. https://doi.org/10.3390/molecules25010112

Bittermann, A., & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift Für Psychologie*, *226*(1), 3–13. https://doi.org/10.1027/2151-2604/a000318

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(1), 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Braun, T., Schubert, A., & Zsindely, S. (1997). Nanoscience and nanotechnology on the balance. *Scientometrics*, *38*(2), 321–325. https://doi.org/10.1007/BF02457417

Breschi, S., Lissoni, F., & Montobbio, F. (2007). The scientific productivity of academic inventors: New evidence from Italian data. *Economics of Innovation and New Technology*, *16*(2), 101–118. https://doi.org/10.1080/10438590600982830

Brooks, H. (1994). The relationship between science and technology. *Research Policy*, *23*(5), 477–486. https://doi.org/10.1016/0048-7333(94)01001-3

Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics*, *69*(1), 3–20. https://doi.org/10.1007/s11192-006-0135-8

Carayannis, E. G., & Campbell, D. F. J. (2010). Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other? A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology. *International Journal of Social Ecology and Sustainable Development*, *1*(1), 41–69. https://doi.org/10.4018/jsesd.2010010105

Cassiman, B., Veugelers, R., & Zuniga, P. (2008). In search of performance effects of (in)direct industry science links. *Industrial and Corporate Change*, *17*(4), 611–646. https://doi.org/10.1093/icc/dtn023

Chang, Y. W., Yang, H. W., & Huang, M. H. (2017). Interaction between science and technology in the field of fuel cells based on patent paper analysis. *The Electronic Library*, *35*(1), 152–166. https://doi.org/10.1108/EL-10-2015-0195

Chen, D. Z., Lin, W. Y. C., & Huang, M. H. (2007). Using Essential Patent Index and Essential Technological Strength to evaluate industrial technological innovation competitiveness. *Scientometrics*, *71*(1), 101–116. https://doi.org/10.1007/s11192-007-1655-6

Cho, T. S., & Shih, H. Y. (2011). Patent citation network analysis of core and emerging technologies in Taiwan: 1997-2008. *Scientometrics*, *89*(3), 795–811. https://doi.org/10.1007/s11192-011-0457-z

de Solla Price, D. J. (1965). Is technology historically independent of science? A study in statistical historiography. *Technology and Culture*, *6*(4), 553. https://doi.org/10.2307/3101749

Du, J., Li, P., Guo, Q., & Tang, X. (2019). Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. *Journal of Informetrics*, *13*(1), 132–148. https://doi.org/10.1016/j.joi.2018.12.004

Durack, K. T. (2004). Tacit knowledge in patent applications: Observations on the value of models to early US Patent Office practice and potential implications for the 21st century. *World Patent Information*, *26*(2), 131–136. https://doi.org/10.1016/j.wpi.2003.11.001

Etzkowitz, H. (2007). *University-industry-government: The triple helix model of innovation* [Proceedings of 51-St EOQ Congress]. European Organization of Quality. http://triplehelix.stanford.edu/etzkowitz_pubs

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and “mode 2” to a triple helix of university-industry-government relations. *Research Policy*, *29*(2), 109–123. https://doi.org/10.1016/S0048-7333(99)00055-4

Frascati Manual 2002. (2007). Frascati manual 2002. Organisation for Economic Co-Operation and Development. https://doi.org/10.1787/9789955682684-lt

Guerrero-Bote, V. P., Sánchez-Jiménez, R., & De-Moya-Anegón, F. (2019). The citation from patents to scientific output revisited: A new approach to Patstat/Scopus matching. *El Profesional de la Información*, *28*(4), 1–12. https://doi.org/10.3145/epi.2019.jul.01

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, *36*(1), 16–38.

Han, F., & Magee, C. L. (2018). Testing the science/technology relationship by analysis of patent citations of scientific papers after decomposition of both science and technology. *Scientometrics*, *116*(2), 767–796. https://doi.org/10.1007/s11192-018-2774-y

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, *32*(8), 1343–1363. https://doi.org/10.1016/S0048-7333(02)00124-5

Hullmann, A., & Meyer, M. (2003). Publications and patents in nanotechnology: An overview of previous studies and the state of the art. *Scientometrics*, *58*(3), 507–527. https://doi.org/10.1023/B:SCIE.0000006877.45467.a7

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, *90*(2), 215–218. https://doi.org/10.1257/aer.90.2.215

Jefferson, O. A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U., Ballagh, A., Moe, J., Dicuccio, M., Ward, K., Bilder, G., Dolby, K., & Jefferson, R. A. (2018). Mapping the global influence of published research on industry and innovation. *Nature Biotechnology*, *36*(1), 31–39. https://doi.org/10.1038/nbt.4049

Karki, M. M. S. (1997). Patent citation analysis: A policy analysis tool. *World Patent Information*, *19*(4), 269–272. https://doi.org/10.1016/S0172-2190(97)00033-1

Ke, Q. (2018). Comparing scientific and technological impact of biomedical research. *Journal of Informetrics*, *12*(3), 706–717. https://doi.org/10.1016/j.joi.2018.06.010

Kim, J., Lee, J., Kim, G., Park, S., & Jang, D. (2016). A hybrid method of analyzing patents for sustainable technology management in humanoid robot industry. *Sustainability*, *8*(5), 474–14. https://doi.org/10.3390/su8050474

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, *21*(4), 493–506. https://doi.org/10.1037/met0000105

Kumari, R., Jeong, J. Y., Lee, B. H., Choi, K. N., & Choi, K. (2021). Topic modelling and social network analysis of publications and patents in humanoid robot technology. *Journal of Information Science*, *47*(5), 658–676. https://doi.org/10.1177/0165551519887878

Lee, C., Kang, B., & Shin, J. (2015). Novelty-focused patent mapping for technology opportunity analysis. *Technological Forecasting & Social Change*, *90*(PB), 355–365. https://doi.org/10.1016/j.techfore.2014.05.010

Lee, S., Lee, H. J., & Yoon, B. (2012). Modeling and analyzing technology innovation in the energy sector: Patent-based HMM approach. *Computers & Industrial Engineering*, *63*(3), 564–577. https://doi.org/10.1016/j.cie.2011.12.002

Leydesdorff, L. (2004). The university-industry knowledge relationship: Analyzing patents and the science base of technologies. *Journal of the American Society for Information Science and Technology*, *55*(11), 991–1001. https://doi.org/10.1002/asi.20045

Leydesdorff, L. (2012). The triple helix, quadruple helix, …, and an n-tuple of helices: Explanatory models for analyzing the knowledge-based economy? *Journal of the Knowledge Economy*, *3*(1), 25–35. https://doi.org/10.1007/s13132-011-0049-4

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. https://doi.org/10.1002/asi.23054

Lin, J. Y., & Yang, C. H. (2020). Heterogeneity in industry–university r&d collaboration and firm innovative performance. *Scientometrics*, *124*(1), 1–25. https://doi.org/10.1007/s11192-020-03436-2

Lu, Y., Xiong, X., Zhang, W., Liu, J., & Zhao, R. (2020). Research on classification and similarity of patent citation based on deep learning. *Scientometrics*, *123*(2), 813–839. https://doi.org/10.1007/s11192-020-03385-w

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2-3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Meyer, M. (2000). Patent citations in a novel field of technology - What can they tell about interactions between emerging communities of science and technology? *Scientometrics*, *48*(2), 151–178. https://doi.org/10.1023/A:1005692621105

Michel, J., & Bettels, B. (2001). Patent citation analysis: A closer look at the basic input data from patent search reports. *Scientometrics*, *51*(1), 185–201. https://doi.org/10.1023/A:1010577030871

Nakagawa, K., Takata, M., Kato, K., Matsuyuki, T., & Matsuhashi, T. (2017). A university–industry collaborative entrepreneurship education program as a trading zone: The case of Osaka University. *Technology Innovation Management Review*, *7*(6), 38–49. https://doi.org/10.22215/timreview1083

Narin, F. (2000). *Tech-line background paper*. CHI Research. Inc.

OECD/Eurostat. (1995). Measurement of scientific and technological activities: Manual on the measurement of human resources devoted to S&T - Canberra manual. In *The measurement of scientific and technological activities*. OECD Publishing. https://doi.org/10.1787/9789264065581-en

OECD/Eurostat. (2005). Oslo manual: Guidelines for collecting and interpreting innovation data. In *The measurement of scientific and technological activities* (3rd ed.). OECD Publishing. https://doi.org/10.1787/9789264013100-en

Pereira, S. D. A., & Quoniam, L. (2017). Intellectual property and patent prospecting as a basis for knowledge and innovation – A study on mobile information technologies and virtual processes of communication and management. *RAI Revista de Administração e Inovação*, *14*(4), 301–310. https://doi.org/10.1016/j.rai.2017.07.006

Ponweiser, M. (2012, May). *Latent Dirichlet allocation in R* (pp. 2–21). http://epub.wu.ac.at/3558/

Qu, Z., & Zhang, S. (2020). References to literature from the business sector in patent documents: A case study of charging technologies for electric vehicles. *Scientometrics*, *124*(2), 867–886. https://doi.org/10.1007/s11192-020-03518-1

Rip, A. (1992). Science and technology as dancing partners. In P. Kroes & M. Bakker (Eds.), *Technological development and science in the industrial age. Boston studies in the philosophy of science* (Vol. 144, pp. 231–270). Springer. https://doi.org/10.1007/978-94-015-8010-6_10

Smith, D. (2014). A citation-based patent evaluation framework to reveal hidden value and enable strategic business decisions. *Technology Innovation Management Review*, *4*(6), 4–13. https://doi.org/10.22215/timreview799

Smojver, V., Štorga, M., & Zovak, G. (2020). Exploring knowledge flow within a technology domain by conducting a dynamic analysis of a patent co-citation network. *Journal of Knowledge Management*. Advance online publication. https://doi.org/10.1108/jkm-01-2020-0079

Sun, X., & Ding, K. (2018). Identifying and tracking scientific and technological knowledge memes from citation networks of publications and patents. *Scientometrics*, *116*(3), 1735–1748. https://doi.org/10.1007/s11192-018-2836-1

Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technological Forecasting and Social Change*, *115*, 131–142. https://doi.org/10.1016/j.techfore.2016.09.028

Tijssen, R. J. W. (2001). Global and domestic utilization of industrial relevant science: Patent citation analysis of science-technology interactions and knowledge flows. *Research Policy*, *30*(1), 35–54. https://doi.org/10.1016/S0048-7333(99)00080-3

Uzun, A. (2006). Science and technology policy in Turkey. National strategies for innovation and change during the 1983-2003 period and beyond. *Scientometrics*, *66*(3), 551–559. https://doi.org/10.1007/s11192-006-0040-1

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. https://doi.org/10.1023/A:1016034516731

Watanabe, C., Tsuji, Y. S., & Griffy-Brown, C. (2001). Patent statistics: Deciphering a "real" versus a "pseudo" proxy of innovation. *Technovation*, *21*(12), 783–790. https://doi.org/10.1016/S0166-4972(01)00025-6

Xu, H., Winnink, J., Yue, Z., Liu, Z., & Yuan, G. (2020). Topic-linked innovation paths in science and technology. *Journal of Informetrics*, *14*(2), 101014. https://doi.org/10.1016/j.joi.2020.101014

Xu, H. Y., Yue, Z. H., Wang, C., Dong, K., Pang, H. S., & Han, Z. (2017). Multi-source data fusion study in scientometrics. *Scientometrics*, *111*(2), 773–792. https://doi.org/10.1007/s11192-017-2290-5

Xu, S., Zhai, D., Wang, F., An, X., Pang, H., & Sun, Y. (2019). A novel method for topic linkages between scientific publications and patents. *Journal of the Association for Information Science and Technology*, *70*(9), 1026–1042. https://doi.org/10.1002/asi.24175

Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, *100*(3), 767–786. https://doi.org/10.1007/s11192-014-1321-8