RESEARCH

Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works

Tahereh Saheb[1]* and Tayebeh Saheb[2]

*Correspondence:
t.saheb@modares.ac.ir
[1] Management Studies
Center, Tarbiat Modares
University, Tehran, Iran
Full list of author information
is available at the end of the
article

Abstract

Big data innovation is a key instrument for economic and social development and for the creation of new business opportunities. This study analyzes the patenting activities of global jurisdictions in the field of big data; as well as the scholarly cited works. We collected 13,112 patent applications between 1992 and 2019, and 642 cited scholarly works by the patents. Our findings report on the development trends in big data technologies, as well as on the link between patenting activities and the cited scientific works. It also analyzes and visualizes the social networks embedded in patents and the cited scientific works; thus, it reveals the patenting activities of global jurisdictions on big data technologies, the strength of the interaction of various agents, such as inventors and applicants within social networks, and the link between patents and the scientific world. This study shows that most of the big data patent applications filed fall within the IPC category of information retrieval, database structures, and file system structures. The majority of the applicants and inventors of the patents filed are Chinese companies and individuals. Scientific fields with stronger connections within the network of co-fields are computer science and medicine. There is a weak link between inventions and scientific works.

Keywords: Innovation evolution, Technological development trend, Big data, Patent analysis, Social network analysis, Bibliometric analysis, Citation linkage

Introduction

With the development of technologies capable of generating high data volumes at a rapid pace, a global effort has been made to develop solutions for better storage, processing and analysis of big data. Big data is considered to be a vital element in the economic and social development of businesses and societies [1]. With the machine subordination and growing use of social media, consumers are increasingly generating data about their behaviors and attitudes that could be purchased and recirculated like a commodity within the debt economy and may be used for marketing and other business strategies [2]. Many organizations are shifting towards re-modeling into data-driven enterprises as companies that distinguish themselves as data-driven function appropriately on "objective measures of financial and operational results" [3]. More than 80% of technology

information is patented [4]. This information may be used for a variety of purposes, such as discovering market trends or predicting future technological development.

I It is estimated that worldwide revenues from the big data market for software and service sectors alone will increase from $42B in 2018 to $103B in 2027 [5]; verifying the emergence of new market opportunities for big data industry. In 2014, the UK government identified big data as one of the "eight great technologies" that will lead the UK into economic prosperity [6]. Big data is a technology with a very broad landscape, and as Matt Truck Company reports, it encompasses a range of innovations on infrastructure, analytics, applications, data resources, data sources, APIs, and open sources. The growing impact of big data on global development and its large landscape has led to the introduction of a large number of big data innovations worldwide. It is therefore important for both academia and industry to recognize the emerging patterns of big data technologies, which are of primary importance in the growth of data-driven enterprises. In this context, while very few studies have evaluated big data innovations in different jurisdictions, to our knowledge, there is a lack of academic effort to provide a comprehensive picture of the innovative big data activities of companies and countries over time. The goal of this paper is to provide a comprehensive overview of the evolution of big data technology in order to track and characterize its trends over time and across jurisdictions, as well as to characterize its linkage and interaction with the scientific world.

Patent data is a valuable resource for understanding the dynamics and activities of the invention ecosystem. Patent data are considered to be result-based indicators of innovation and the reflection of technological and scientific changes and inventive processes [7]; and are capable of "appropriately" describing the diffusion of a technology [8] or assessing technology management [9, 10].

Nonetheless, some studies emphasize that patent data cannot represent the entire invention ecosystem; thus, there is a need to incorporate other sources of information, such as scientific literature [7, 11]. Interactions between universities and industries are vital in the development of innovation systems [12]; as recommended, inventors should cite scientific papers in order to file patent documents [13]. Many jurisdictions require applicants to provide a complete and clear description of their invention, including prior art; that is, any invention disclosed or made available to the public anywhere in the world by any person and by any means prior to filing (priority data). It is therefore assumed that the patent documents describe both the technical features of the invention and the recent scientific and academic developments in the relevant field of industry, so that patent data can also represent scientific advances of a technology.

In order to extract predictive insights about the development of a technology out of patent data, quantitative methods, such as bibliometric and social network analysis examine patents as primary sources of information to uncover technological trends, its linkage with scientific data and to discover hidden patterns of innovation ecosystem [14]. Previous studies argue that social network analysis as a data mining visualization method [15] is a superior method compared to other conventional techniques [16]. In particular, the analysis of the social network identifies connections and relationships (edges) between actors (nodes). In a patent analysis, nodes represent individuals, such as applicants and inventors or entities, such as patent documents or fields of study. The edges between the nodes can also be their cooperative activities, such as their citation

links [17]. Social network analysis enables multi-dimensional analysis of similarities or dissimilarities between actors in low-dimensional spaces. In this method, actors who are closer and more similar to each other in input data are closer in space; and actors who are less similar or closer to each other are further apart in space [18].

In this study, patent applications were analyzed using techniques for bibliometric and social network visualization. Bibliometric analysis is incorporated to analyze scientific trends in a variety of topics [19, 20]. Using descriptive bibliometric analysis techniques, we aim to (1) explore the temporal evolution of patent productivity on big data; (2) measure the productivity of the big data industry on the basis of the most productive inventors, applicants, authors, jurisdictions, institutions, companies, patent classifications and scientific fields of study; and (3) to measure the interaction of patents with scientific scholarly works. Using social network analysis techniques, we aim at (4) analyzing and visualizing interactions and connections within the networks of co-applicants, co-inventors, co-authors, co-appearances of scientific keywords and scientific fields of study; and (5) uncovering hidden patterns and trends within the scientific and inventive communities of big data.

In order to track the trends in big data technologies, we analyzed 13,112 patents and 642 scientific journals cited in the patents. We used the techniques of bibliometric and social network analysis to provide (1) a descriptive bibliometric analysis of patents and of the cited scholarly works; (2) a citation linkage analysis of patents to scholarly works; (3) a social network analysis of big data invention activities; and (4) a social network analysis of scientific journals cited by big data patents.

Preceding patent analysis studies have proposed a variety of analysis and visualization techniques, such as natural language processing [21], semantic analysis [22, 23] or neural networks analysis [24–26]. Some of the past research also combined patent citation data with external data [11, 27–29]. Previous studies on the patent analysis of big data technology explored the patent abstract analysis of Chinese big data [30]; hot classified fields of big data technology [31]; technology valuation methods using quantitative patent analysis for technology transfer in big data marketing in Europe [32]; or business interests and activities around big data [33].

This paper consists of the following sections: first, a review of previous studies that carried out a bibliometric analysis of patents. This is followed by a detailed description of our research method, the data collection procedure and analytical methods, and the tools we used. The findings and results of our analysis are described in the next section. We conclude the paper by discussing the results, the theoretical and practical contribution of the paper, the limitations of our research and some possible future works.

Theoretical framework

Patents as an indicator to measure innovation

Output-based indicators are one of the categories for measuring innovation [34, 35]. These indicators are made up of the results of innovation activities, such as patents. Patent data is a valuable resource for understanding the dynamics and activities of the invention ecosystem. Patent data presents insightful windows for inventors, engineers, companies and decision makers [36–38]; and can be adopted as a tool to model and explain the growth of inventions across countries [39]. Patent data are indicators

of representing the techno-scientific shifts and inventive processes [7]; and are capable of "appropriately" describing the diffusion of a technology [8] or assessing technology management [9, 10]. A patent is an intellectual property right that should be novel, adequately described and claimed as an inventive activity; and refers to the prior art, such as scientific works. A patent shall be granted for an inventive activity to an individual, a company, a university or any public or private enterprise [40].

One of the essential requirements for the protection of an invention is its novelty. Under the TRIPS Agreement, members are required to submit applications that they are novel, involve an inventive step and are capable of industrial application [41]. The agreement does not, however, define the term novelty. The definition of novelty has therefore been delegated to the Member Governments. In the Iranian legal system, for example, novelty means that the invention is not predicted in the prior art. The term "prior art" is also defined as anything that has been disclosed in any part of the world by means of written or verbal communication, or by practical use, or in any other way, prior to the date of application or priority date (Clause 4 of the Patent, Industrial Designs, and Trademarks Act of Iran, 2008). Thus, in this general sense, the prior art concerns both patent applications and scholarly works. Publishing the information on the invention in scientific articles would eliminate the novelty of the invention and, consequently, the invention is not patentable [42]. For this reason, certain legal systems oblige the applicant to describe the invention, which is one of the documents annexed to the application.

However, the use of patents to portray the whole picture of innovation is being challenged [43]. Some studies believe that some innovations are not patentable [7]; or patented inventions do not end up in an innovation or does not indicate whether their new technical knowledge has a positive economic value [44]. Moreover, countries and organizations may value patenting activities differently [45]. For example, in China, Article 22 of the patent law, as amended in 2008, stipulates that "novelty means that the invention or utility model concerned is not an existing technology"; so "existing technologies mean the technologies known to the public both domestically and abroad before the date of application". Article 36 of the law provides that, where the applicant requests a substantive examination, he/she shall submit the reference material relating to the invention. In China, however, the non-disclosure of information has no legal effect; and, to date, the Chinese Patent Office has not declared any application void or invalid or withdrawn as a result of the non-disclosure of information [46]. On the other hand, the US Chapter 2000 Manual of Patent Examining Procedure (MPEP) dealt in detail with the obligation of disclosure, since failure to disclose the patent may ultimately render the patent unenforceable. I In the European Patent Office, which has the highest reference rate for scientific works, Article 42 of the European Patent Convention provides that the description should indicate the background art; otherwise the application will be refused (Art 97) or the patent will be revoked (Art 101). In the jurisdiction of the WIPO referred to in Article 5 of the Patent Cooperation Treaty, all descriptions of patents should refer to prior art information.

In view of the limitations of patent data, previous studies argue that there is a need to integrate other sources of information, such as scientific literature, and citations linkage of patents to scholarly works [7, 11]. These studies argue that patent citations may act as an indicator of the value of innovation [47]. The patent data and the cited scholarly

works are still a commonly used tool for studying innovation trends and technological developments [48].

Patent as a measure of big data technologies' evolution

Big data patent analysis has been performed in very few studies as outlined in Table 1. In one study [31]; the authors explore trends in hot classified big data technology fields from 2004 to 2012. Their network analysis shows a weak collaborative publishing network; also, as their study shows, the acquisition of big data is at a relatively higher level of research than other fields. In another study, the authors studied big data patents in China from 1980 to 2016. Their analysis shows that the development of big data patents in China increased after 2005; and patent applicants are mainly universities and companies [30]. Another study focused on the analysis of big data patents in Europe from 1989 to 2013. This study shows that patents are strongly linked to big data sub-technologies; and are dependent on each of them. This study shows that the top-level keywords are image, layout, and object [32]. Studies by [33] reveals that there is a limited co-occurrence among leading big data institutions; and the main topics in big data fields are business intelligence, cloud-based services, customer experience, social media, and healthcare and web services.

Materials and methods

In this study, we have integrated bibliometric and social network analysis techniques to uncover the key dynamics that define the development patterns of big data innovations concealed through unstructured patent and scientific literature texts; to measure the performance of big data invention activities; and to measure the interaction strength of different agents inside social networks. We used the methods of social network visualization to identify and visualize the communities and clusters within the Big Data Inventiveness Network and the cited scholarly works; the nodes and actors, and the links and edges (the strength of interaction) between and between the nodes.

Data collection

We have collected data from the Lens Open Source Platform. As described in the Lens website, the Lens database is an "open global cyber infrastructure" for cartography innovation. The Lens database contains 95% of the global patent documents and links to the majority of scholarly literature. The patent data sources incorporated into the Lens are DocDB bibliographic records of the European Patent Office from 1907, USPTO Applications from 2001, USPTO Grants from 1976, USPTO Assignments, European Patent Office (EP) Grants from 1980, WIPO PCT Applications from 1978, and Australian Patent Full Text from IP Australia. Scholarly datasets are also integrated with PubMed, CrossRef and Microsoft Academic.

In the patent search, the keyword "big data" ~0 was searched in Title OR Abstract OR Claims. We did not exclude any dates, jurisdiction, and type of document. However, we founded the classification on the basis of the IPC classification. Of the options for full text, one doc per family, and stemming, we chose stemming. The query language was English as well. We did the query on January 5th, 2019. The search showed that 642 scholarly works were cited by patents; the number of patents cited was 291.
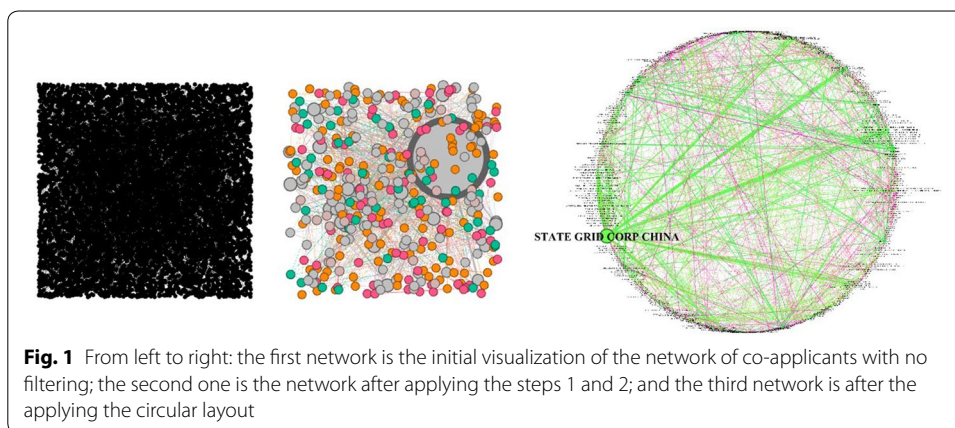
Table 1 A summary of previous studies on big data patenting activities

Study	Method of analysis	Technology	Scope	Patent + literature	Some of findings
The current study	Bibliometric; social network analysis	Big data	Global	Patent and literature	Please check the result section of this paper
[31]	Bibliometric, patent metrics, statistical analysis and other methods of related laws	Technical classification fields system of big data (9 classifications)	Global	Patents	Data mining, classification, big data, clustering, large data sets, machine learning, cluster analysis, algorithm, execution, and biological information are the ten big data fields with high occur frequency
[30]	Word segmentation models on the basis of TF-IDF principle (text mining)	Chinese big data	China	Abstracts of patents	Development of big data patents in china after 2005; Patent applicants are mainly universities and enterprises with Huawei having the most qualities of patent applications; an unbalance situation in the development of big data technologies in China
[32]	Text mining Social network analysis	European big data marketing	Europe	Patents	Patents are strongly connected so the sub- technologies of big data are depended to each other The top-ranked keywords are also image, layout, and object
[33]	Tech mining	To measure "hot topics" in industry with regard to big data	Unspecified	Patents and publication	There is little co-occurrence among leading big data institutions The main topics in big data fields are business intelligence, cloud-based services, customer experience, social media, and healthcare and web services

We used the following software to analyze and visualize data: Gephi 0.9 software, PatCite, Patent and Scholarly works on the Lens platform. As an open global cyber infrastructure, Lens acts as a public resource for global patents and scientific knowledge and is a platform for innovation cartography. Its aim is to make the problem solving more accessible, secure and inclusive. We used the Lens platform for our descriptive bibliometric analysis and the Gephi software to analyze and visualize the social networks. This approach is novel. As far as our knowledge is concerned, we have not found any study that integrates the following platforms for the analysis of patents and the works cited by scholars.

After importing the CSV file extracted from the Lens platform, we checked the items of "creating links between applicants", "removing duplicates", and "removing self-loops in cases when an agent is connected to itself". The graph type for the network of co-applicants was undirected with 7830 nodes and 2383 edges. We used the average weighted degree as our statistical method. Calculation of the average weighted degree indicates how many times the edge is passed between a pair of nodes. The higher the weight of the node, the higher the connection is compared to the low weight of the node [49]. The average degree will calculate the average number of edges connected to the node; whereas the average weighted degree will be the average sum of the weights of the edges connected to the node. We ran the average degree formula under the network overview. The average degree score was 0.609 and the average weighted degree was 0.743. The closer this score is to 1, the more the network is connected [50]. We then filtered the network based on their weighted range of degrees and set the number between 3.122 and 446.0. On the size of the nodes menu, we chose the weighted degree and set the minimum size to 20 and the maximum size to 200. We chose the Circular Layout as the third step (Fig. 1). The final version of the map is described in more detail in the results section of this paper.

We followed the same procedure to analyze and visualize the co-inventor network. The network consists of 23,977 nodes and 77,170 edges, and the graph type is undirected. The average weighted degree was 7.703 and the average degree was 6.437, showing a highly connected network compared to the network of co-applicants.

Fig. 1 From left to right: the first network is the initial visualization of the network of co-applicants with no filtering; the second one is the network after applying the steps 1 and 2; and the third network is after the applying the circular layout

Because of the high density of the network, the filtering range is set to 57.94 and 347 in order to have a clear representation of the network (Fig. 2).

As a next step, we included only the following jurisdictions: Japan, the United States, the EP, the Republic of Korea, the WIPO, France, Germany, the United Kingdom and Canada to showcase their network of co-applicants and co-inventors. We excluded China in order to present a clearer version of the activities of the other active jurisdictions. We also excluded jurisdictions with very few big data patenting activities. We followed the same procedure for the analysis and visualization of the networks.

We also analyzed and visualized the social networks of scientific works cited in the patents. We used modularity to detect communities for the network of co-occurrence of keywords. We set a resolution to 1.0. Modularity was 0.213; the number of communities was 8. Modularity looks for nodes that are more densely connected together than the rest of the network [51]. We followed the same procedure for the network of co-authors and the network of fields of study.

Results

Descriptive bibliometric analysis

We analyzed 13,112 patents and 642 scientific scholarly works cited in the patents. These documents have been collected from the Lens Platform.

As regards the temporal evolution of scientific and invention productivity, as shown in Fig. 3, the productivity of big data inventions increased sharply in 2014 with 491 applications, followed by a constant increase until 2018. The number of patent applicants increased to 1266 in 2015 and increased sharply in 2016 with 24 29 applications. The same growth continued with 4184 applications in 2017. The number of patents in 2018, however, did not increase significantly, as only 4546 applications were indexed in 2018. Of the total number of patent applicants, 1158 patents were granted, and 1247 patent applications were limited.

With regard to scholarly works on big data, as shown in Fig. 3; the interaction of patents with scholarly works increased sharply in 2012 with 74 patents, followed by a steady decrease; the number of cited works reached to 66 in 2013; to 45 in 2014; and to 28 in 2015. The number of patents in 2018, however, did not increase significantly, as only

Fig. 2 From left to right: the first network is the initial visualization of network of co-inventors with no filtering; the second one is the network after applying steps 1 and 2; and the third network is after the applying the circular layout

Fig. 3 Temporal evolution of scientific and invention productivity about big data: patent applicants per year (left), and timeline of cited works based on publication year (left)

Fig. 4 The IPC classification of the inventions of the most productive applicants
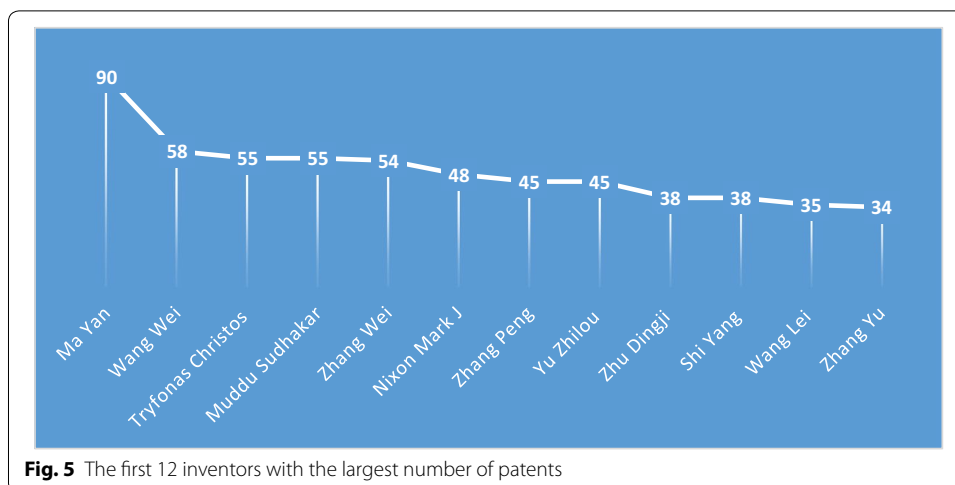
**Fig. 5** The first 12 inventors with the largest number of patents

4546 applications were indexed in 2018. Of the total number of patent applicants, 1158 patents were granted, and 1247 patent applications were limited. The patent search showed that the patent with the earliest priority data had been submitted in 1989. It was filed in 1989 and was published in 1991. The applicant is Prochazka Miroslav Ing. The oldest big data patent was a limited patent under the IPC classification G06F13/16; G refers to physics, G06 to computing, and G064 refers to digital electrical data processing. The title of the patent was "Automatic Analyzer of Image". This invention offered a design for collecting and processing of big data files.

With regard to inventors with the largest number of inventions, as shown in Figs. 4 and 5, Ma Yan has the largest number of patents with 90 inventions. This analysis helps us to identify not only the largest inventors, but also their topic of invention. Based on the IPC classification, most of his works fall under the G06F17/30 classification, which refers to information retrieval and database structures. He was the applicant for most of his inventions (36 inventions). Out of the other applicants of his works, the major ones are Shenzhen Boxinnuoda Economic Relations & Trade Consultants Co Ltd, Shenzhen Boxinnuoda Economic and Trade Consulting Co Ltd, and State Grid Corp China. All of his applicants are Chinese companies.

Wang Wei is the second inventor with the highest number of patents. His inventions are half of the inventions of the first inventor (58 inventions). Most of his patents are categorized under the IPC classification of G06F17/30, which applies to the processing of information and the structure of the database. Out of the 10 applicants for his work are State Grid Shandong Electric Power Co, Ztesoft Tech Co Ltd and State Grid Corp China. He is not the applicant of any of his inventions.

Two inventors have 57 inventions. Most of the works of Muddu Sudhakar as the third inventor with the highest number of inventions fall within the category IPC of H04L29/06; which involves interconnecting or transferring information or other signals between memories, input/output devices or central processing units. Most of the works of Tryfonas Christos, also as the third inventor with the highest number of inventions (57 inventions) are classified under the electricity and transmission of digital information. Splunk Inc is the significant patent applicant for Christos and Sudhakar.

Fig. 6 Map of patent applications filed in various jurisdictions
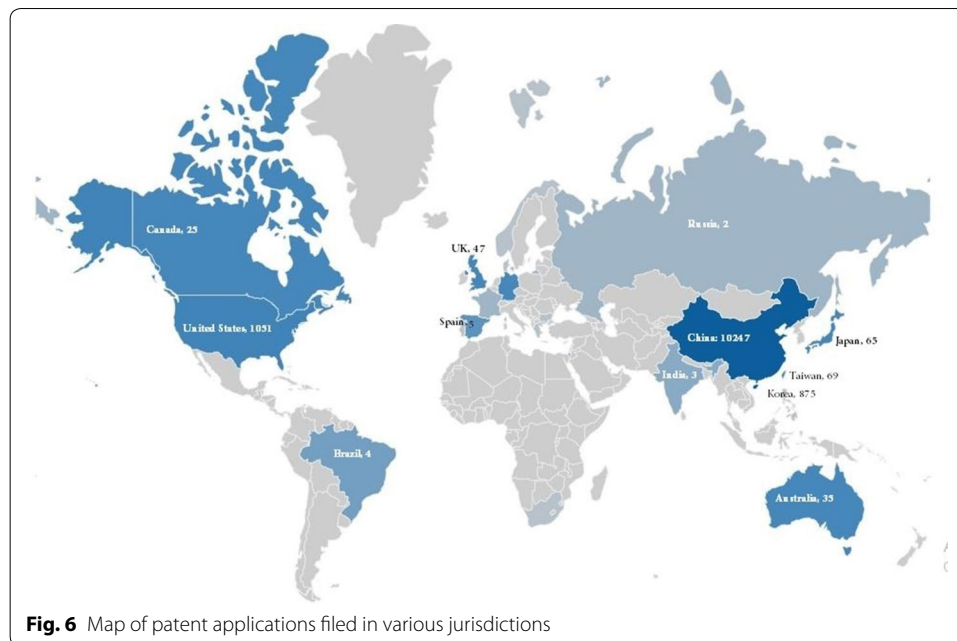
Zhang Wei is the fourth inventor. Major patents are listed under the IPC classification of G06F17/30, which refers to the retrieval of information and the structure of the database. Out of the tenth applicants listed on his portfolio, companies like Wuhu Yueruisi Information Consulting Co Ltd, Guangxi Power Grid Corp Electric Power Res Inst, State Grid Corp China and Oracle Int Corp are listed.

The fifth inventor with the highest number of patents is Nixon Mark J. His major works are listed under the IPC classification of G05B19/418, which applies to computer integrated manufacturing and integrated manufacturing systems. The only applicant for Nixon Mark J is the Fisher Rosemount Systems Inc.

We also traced each applicant's patent counts, revealing that China has filed more big data patent applications than the US; with companies like the State Grid Corp China; securing the first rank. This Chinese company has filed 289 patents; followed by IBM with 121 patents, Inspur Group Co LTD with 85 patents, Huawei Tech Co Ltd with 78 patents, ZTE Corp with 73 patents, Alibaba Group Holding with 63, Splunk Inc with 59 patents, University South China Tech with 55 patents, University Southeast with 50 patents, and Zhengzhou Yunhai Information Tech Co with 50 inventions. Except the State Grid Corp China, most of the companies, such as IBM are leading players in the information communication technologies industry. Inspur Group Co is a leader in the cloud computing and big data services; while ZTE Corp or Huawei are leading in the telecommunication industry.

Of the patent applications of the State Grid Corp China, 7% fall are limited patents and 93% are at the pending state. The major applications of the company are under the IPC classification of G06Q50/06; which refers to data processing systems and methods. The company submitted its first application in 2012 under the title "Realizing Method of Supervisory Control and Data Acquisition (scada) History Data Distribution-type Storage Facing Power Grid".
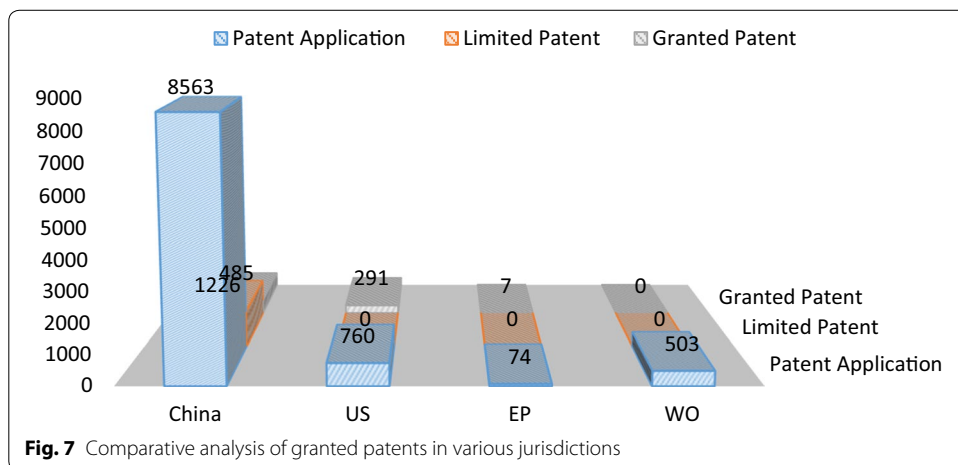
The major US big data patent applicants are IBM, Splunk Inc, Fisher-rosemount Systems Inc, Electronics and Telecommunications Research Institute, American Express Travel Related Services Company Inc, Microsoft Technology Licensing Llc, and Sap Se.
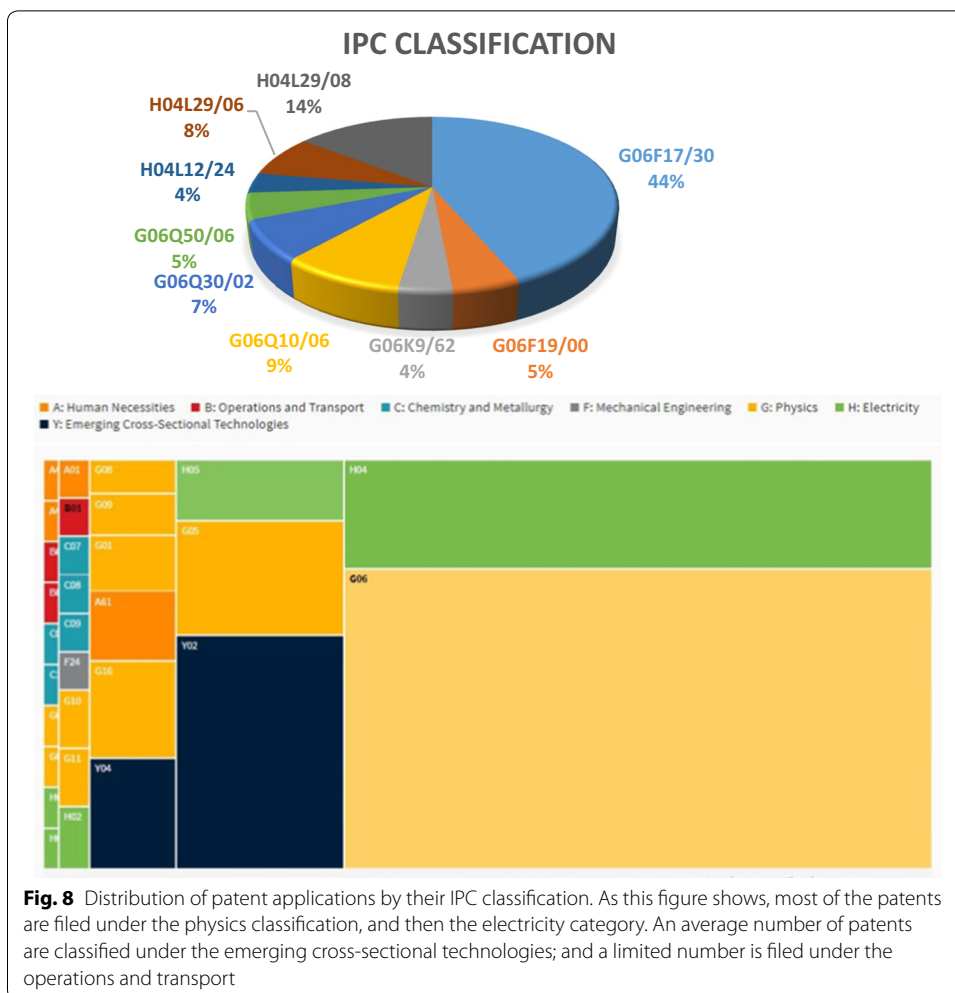
In the European jurisdiction, the major applicants are respectively Huawei, Tata Consultancy Services, Honeywell, Alibaba, STE, and GEOTAB Inc. The number of patents also increased from 2 patents in 2013 to 8 in 2014, and then to 14 in 2016; and the number of patents increased to 51 in 2018.

Shenzhen Boxinnuoda Economic Relations & Trade Consultants Co Ltd, Huawei and ZTE Corp are the key applicants in the jurisdiction of the WIPO. The IPC classification of major patents of the company is G06F17/30.

The analysis of jurisdictions with the highest number of patent applications filed (Fig. 6) shows that China is in the first place and in the lead; that it has filed a total of 10,247 big data patent applications (i.e. 78% of global patent applications). Applications from the USA came second with 1051 patent applications (i.e. 8% of global applications), followed by the Republic of Korea, with 875 applications representing 7% of global patent applications. The list suddenly falls to WIPO with 517, which (i.e. 4% of the global patent application) and the European patent with 107 filings (i.e. 1% of the global application). The list continues with Taiwan with 69 applications; Japan with 65 applications; Australia with 35 applications; Singapore with 11 applications; and India with 3 applications. However, this analysis (Fig. 7) shows that 83% of China's 10,274 patent applications are still pending, 12% are limited patents, and only 5% are granted. In the US jurisdiction, 72% of patent applications are pending and 28% of patents are granted. In the WIPO jurisdiction, 98% of patent applications are pending and 2% search reports have been issued.

As regards the classification of patents (Fig. 8), this analysis shows that most of the patent applications are filed under the IPC classification of G06F17/30. In this classification, G refers to physics, and G06F refers to the processing of electrical digital data. In this subclass, "handling" means processing or transporting data and "data processing equipment" is also included. IPC defines this subclass as follows: "Electrical arrangements or processing means for the performance of any automated operation using empirical data in electronic form for classifying, analyzing, monitoring, or carrying out calculations

Fig. 7 Comparative analysis of granted patents in various jurisdictions

Fig. 8 Distribution of patent applications by their IPC classification. As this figure shows, most of the patents are filed under the physics classification, and then the electricity category. An average number of patents are classified under the emerging cross-sectional technologies; and a limited number is filed under the operations and transport

on the data to produce a result or event". G06F17 refers to digital computing or data processing equipment or methods, specially adapted for specific functions (information retrieval, database structures or file system structures); and G06F17/30 to information retrieval and database structures and file system structures; which refers to G 06F16/00. The main classifications of the patent family were G, H and Y. In this classification, G refers to physics, H to electricity, and Y to emerging cross-sectional technologies.

Citation linkage analysis of patents to scholarly works

First, in this section we present statistics on citations to scientific works on the basis of global patents. We then look at the jurisdictions of China, the United States, the European Patent Office and the WIPO, which have the highest rate of receiving big data patent applications.

This analysis shows that only 291 (2.2% of patent applications) of the 13,112 patent applications cited were scholarly works. The number of cited works is 642. The major cited institutions are respectively Microsoft, IBM, Stanford University, MIT,

University of California, Berkeley, Chinese Academy of Sciences, Carnegie Mellon University, Hewlett-Packard, Nanyang Technological University and the University of British Columbia. Therefore, the major cited institutions are American companies and top-ranked American Universities.

The oldest paper cited in the patent applications was published by a Russian scholar in 1970. The title of the paper is "Heuristic self-organization in problems of engineering cybernetics" [52]; and was published in the field of engineering cybernetics and mathematical optimization and was cited by 6 patents.

The highly cited paper, with 2523 patent citations was published by some scholars from Harvard University, MIT and Ohio State University in 1999; and was titled "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring" [53]. Of the patents cited this work, 1818 patents are at the state of granted patent and the major applicant of these patents are Genetech Inc, Hoffmann La Roche and Squibb Bristol Myers Co. Genetech as a member of the Roche Group, is a biotechnology company for discovering, developing, manufacturing, and commercializing medicines to treat patients with serious or life-threatening medical conditions. Roche Holding is also a Swiss multinational healthcare company which is active under two divisions of pharmaceutical and diagnostics.
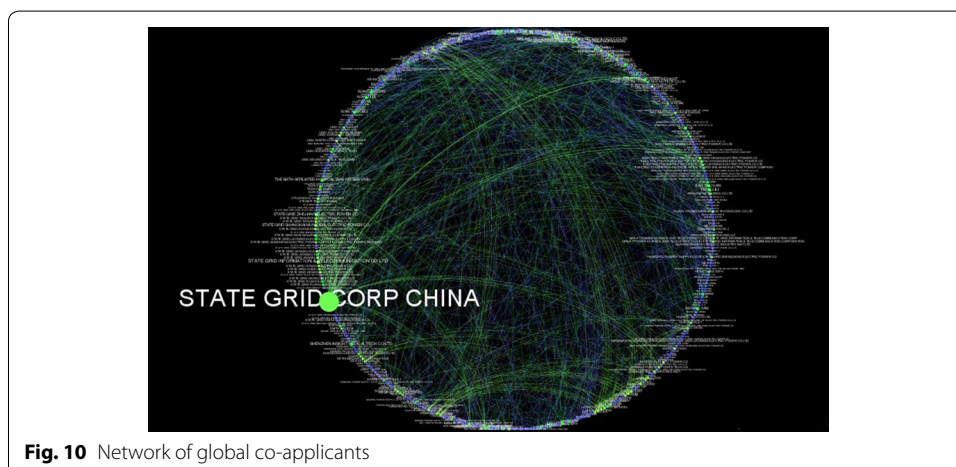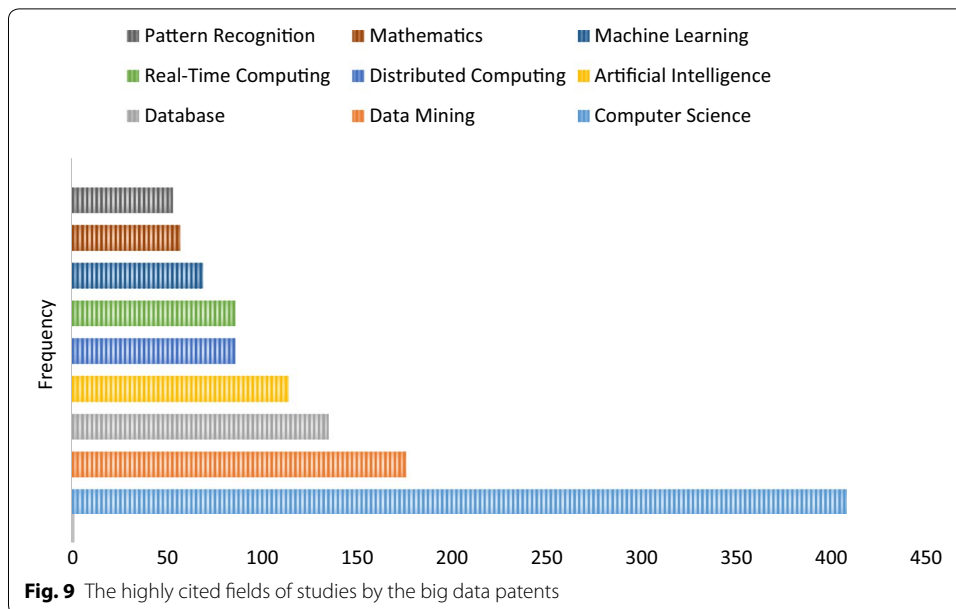
As far as co-citation analysis of patent applications and scholarly work by jurisdiction is concerned, this analysis shows that only 82 applications (i.e. less than 1%) have been referred to scientific work in China as the top jurisdiction with the largest number of big data patent applications. The number of scientific articles cited is 103.

The United States is the second largest jurisdiction with the highest number of applications filed. In that jurisdiction, of the 1051 applications, only 146 statements (i.e. 13.89%) referred to scientific work. Out of 517 applications, only 38 (i.e. 7.35%) of the jurisdiction of the WIPO referred to scientific articles. The number of articles referred to is 83. Of the 107 applications received by the European Patent Office, only 16 (i.e. 14, 95%) refer to scientific works. The number of articles cited was 45.

With regard to the citation of patents and scholarly works by major applicants, of the 289 applications filed by the State Grid Corp China, only 2 applications cited 2 scholarly works; showing a weak link between patents and scholarly works of the State Grid Corp China. Out of the 121 applications filed by the IBM, 29 of the applications cited 59 scholarly works.

Table 2 The most productive big data applicants; and their linkage with scholarly works

Applicant	Major applications under the IPC Classification	Number of applications	Granted patent %	% of patents citing the scholarly works
State Grid Corp (China)	G06Q50/06	289	0	0.69
IBM (US)	G06F17/30	121	31	23
Inspur Group (China)	G06F17/30	85	0	1
Huawei (China)	G06F17/30	78	1	6
ZTE (China)	G06F17/30	73	3	0
Alibaba (China)	G06F17/30	63	2	1
Splunk (US)	H04L29/06	59	36	7

Fig. 9 The highly cited fields of studies by the big data patents

Fig. 10 Network of global co-applicants

Thus, as Table 2 shows, IBM, as an American company, has developed stronger links between patents and scholarly works, with 23% of patent applications citing scholarly works. Despite the fact that the State Grid Corp China have submitted the largest number of patent applications, it has a very weak link with scholarly works, with less than 1% of the patent applications citing scholarly works.

With regard to the fields of study (Fig. 9), some of the fields of study widely cited by patents include computer science, data mining, database, artificial intelligence, and distributed computing. In computer science, the main subjects were also software, information systems, general computer science, computer science applications, theoretical computer science, computer networks and communications, and hardware and architecture, computational theory and mathematics, computer vision and pattern recognition, and artificial intelligence.

Social network analysis of invention activities on big data

In this part of the paper, we analyzed a network of global co-applicants and co-inventors. We have also visualized this network for major jurisdictions other than China. In the next step, we visualized the network of co-authors, the co-occurrence of the keywords of the cited scientific works, and the network of fields of study of the cited scientific works.

As mentioned in the Methodology section, the network of co-applicants is visualized based on the degree of weighted agent. Figure 10 is a co-applicant network with 398 nodes and 1090 edges; after the filtering has been applied. The blue lines indicate agents with a low weighted degree and a weak connection; and the green lines indicate agents with a higher weighted degree and a stronger connection between the nodes. The analysis shows that 6020 (78.88%) elements had a weighted degree of 0; indicating a weak connection of the nodes within the network. One element, the State Grid Corp China, has a weighted degree of 446, indicating the strong link between the applicant and the other network applicants.

The network of co-applicants shows that IBM has a weighted rating of 10 as the second applicant with the highest number of patent applications, which is very low compared to the State Grid Corp China. The weighted degree of the other major applicants is as follows: Huawei (10), Alibaba (15); University South China Tech (5); University South East (13); and Microsoft (13). The analysis shows that the majority of applicants with the highest weighted degree are Chinese companies. Some major applicants within the jurisdiction of the United States have the weighted degree as follows: (1) American Express (1.0) and Sap Se (3); while Splunk and Fisher-Rosemount Systems have zero weighted degrees. The weighted degree of the major applicants in the jurisdiction of the European patent and the WIPO was also zero.

With regard to the co-inventor network (Fig. 11), this analysis shows that Wang Wei (347); Wang Jian (290); Wang Lei (271); Zhang Pen (250); Zhang Wei (213) are the first
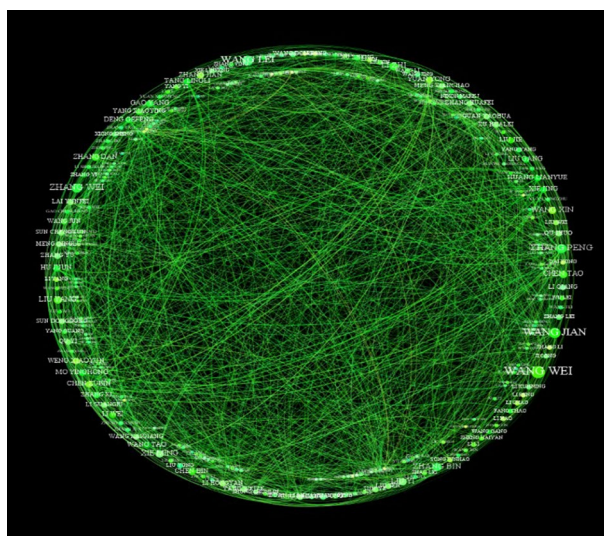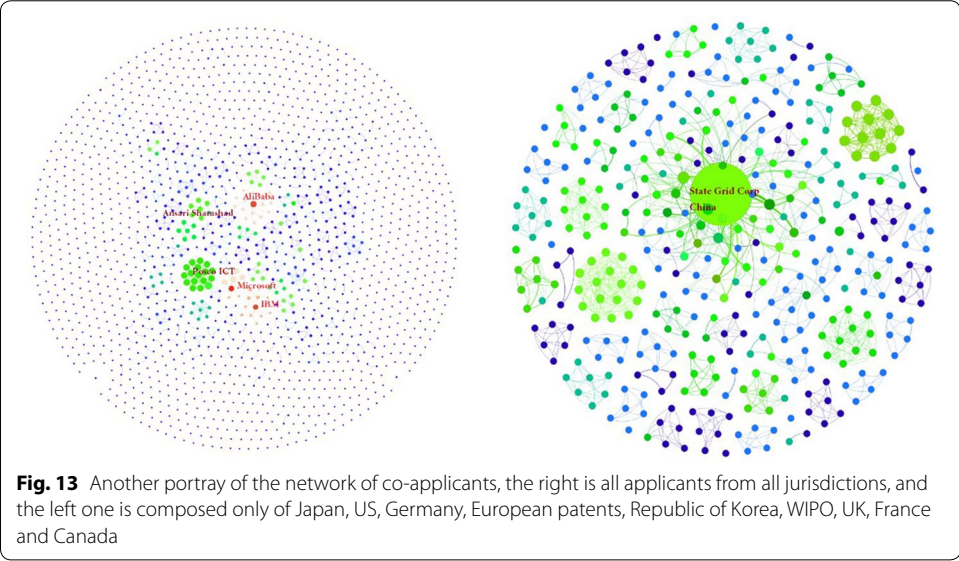

Fig. 11 Network of global co-inventors

Fig. 12 Network of co-applicants of major jurisdictions except China. As the size of nodes and the line of edges show there is not a significant difference among the applicants in terms of their interaction. We highlighted IBM, Microsoft and Alibaba manually

Fig. 13 Another portray of the network of co-applicants, the right is all applicants from all jurisdictions, and the left one is composed only of Japan, US, Germany, European patents, Republic of Korea, WIPO, UK, France and Canada

five inventors with the highest average weighted degree in the network. The weighted degree of Miroslav Prochazka, who filed the first big data invention in 1989, is 0.

As far as inventors with the highest number of patents are concerned, Wang Wei, Ma Yan, Muddu Sudhakar and Tryfonas Chritos have the largest number of inventions, as we mentioned earlier. Wang Wei, however, has a very large weighted degree of 374; showing the network's strong connection. Ma Yan has a weighted degree of 64; which shows a relative average network connection. Both Muddu Sudhakar and Tryfonas Christos have a weighted degree of 19. Zhang Wei has a very large weighted degree of

250 as the fifth inventor with the highest number of patents; which also shows his strong connection in the global network.

In this part of the paper, we excluded China as one of the jurisdictions. We have included the following jurisdictions with the highest number of patents: Japan, United States, Germany, European patents, Republic of Korea, WIPO, United Kingdom, France and Canada (Fig. 12).

This analysis shows that that the highest weighted degree belongs to Posco ICT Co with a degree of 19. Alibaba Group Holding has a degree of 15; Microsoft has a degree of 13; and IBM has a degree of 12. Posco ICT is a company established in 2010 in South Korea. Unlike the global network, the major applicants have relatively similar network connections and we have not seen a significant difference between the major applicants. Figure 13 is another portrait of the network. The blue lines and circles show weaker connections; while green lines and circles indicate stronger connections. We highlighted Alibaba, Microsoft and IBM, which also have stronger connections in the network.

The right network in the Fig. 13 is the global network of co-applicants with the State Grid Corp China as a major applicant with higher degree of interaction inside the network. There is also a sharp difference between the company and the other major applicants in terms of their interaction within the network. Most of the nodes have formed small communities, most of which are disconnected from other communities. The left network in Fig. 13 represents a network of co-applicants in major jurisdictions with the exception of China. This network consisted of 1654 nodes and 848 edges; this shows a weak interaction within the network; however, there was no significant difference between the major applicants in terms of their interaction. Most of the nodes are disconnected from each other, and we observe a few small interconnected communities around applicants like IBM, Microsoft and Alibaba.

In our next step, we analyzed the network of co-inventors among Japan, US, Germany, European patents, Republic of Korea, WIPO, UK, France and Canada. T This network consists of 4073 nodes and 6669 edges (Fig. 14), which shows a relatively strong interaction within the network. The average weighted degree of the network is 3.955; and the average degree is 3.275. Figure 14 shows the visualization of the network based on
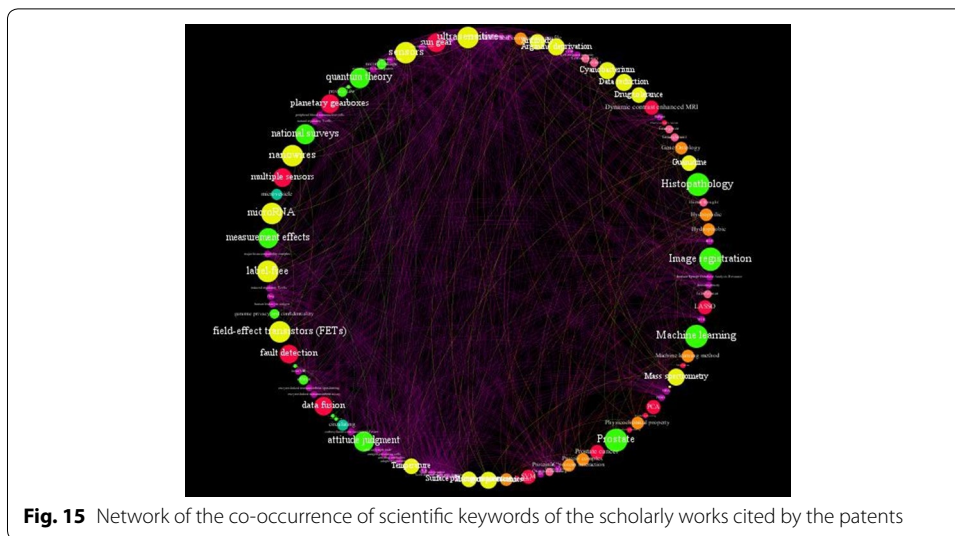

Fig. 14 Two layouts of the network of co-inventors of the major jurisdictions except China
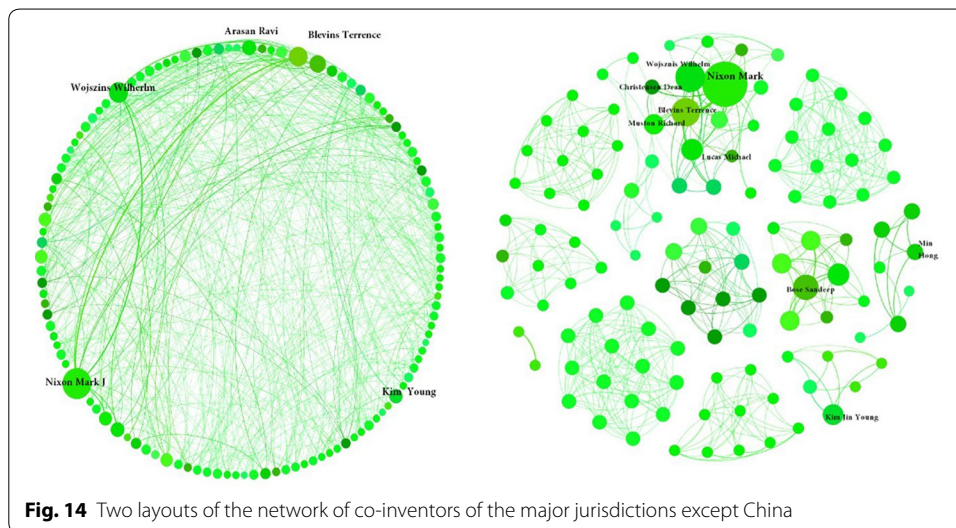
Fig. 15 Network of the co-occurrence of scientific keywords of the scholarly works cited by the patents

Table 3 Clusters of the co-occurrence of keywords of scientific works cited by the patents

Clusters	Keywords of the cluster
Cluster Purple with a weighted degree of 49 (28.26%)	Major histro-compatability complex, induced regulatory T cells, anti- drug antibodies, T Cell receptor, antigen-presenting cells, immunogenicity, cell-mediated immunity, sport-forming cells, peripheral blood…, unmodified original epitopes, recombinant Fc Fusion protein, immune epitope database analysis resource, thymus dependent, enzyme linked immunosorbent spot-forming, human leukocyte antigen, natural regulatory T cells, adaptive regulatory t cells, artificial lymph node, thymus, carboxyl fluorescein succinmidyl ester, quality by design, T cell, factor VIII
Cluster Orange with a weighted degree of 8 (9.78%)	Physicochemical property, gene ontology, hydrophobic, sequence alignment, machine learning method, amino acid composition profile, protein to protein interaction, hydrophilic, protein complex
Cluster light purple with the weighted degree of 6 (7.61)	Quantum modeling, context, concept theory, entanglement, interference, emergence,
Cluster of dark yellow with the weighted degree of 5 (17.39)	Antibody, field-effect transistors, mass spectrometry, label free, cyanobacterium, shotgun proteomics, drug tolerance, nanowires, temperature, guanidine, arginine deprivation, ultrasensitive, data reduction, sensors, microRNA, surface Plasmon resonance
Cluster red with the weighted degree of 4 (16.3%)	Fault detection, LASSO, scientific workflows, electronic health records, SVM, map reduce, prostate cancer, multiple sensors, dynamic contrast enhanced MRI, PCA, data fusion, planetary gearboxes, predictive modeling, sun gear, parallel computing
Cluster green with the weighted degree of 3 (17.39%)	Histopathology, classifier, attitude judgment, genome privacy and confidentiality, classifications, national surveys, proteins, measurement effects, machine learning, record linkage, ethics, privacy law, image registration, quantum theory, family, prostate
Cluster dark green with the weighted degree of 1 (2.17%)	Microvesicle, circulating
Cluster light yellow with the weighted degree of 0 (1.09%)	Non-programmatic

two different layouts; which is filtered based on the weighted degree of 18 and 123. This analysis shows that Nixon Mark has the highest weighted degree of 123; then Wojszins Wilhelm with 76; and Velvins Terrence with 72.

Social network analysis of scientific works cited by patents

In order to conduct a social network analysis of the scholarly works cited by the patents; we first conducted an analysis of the co-occurrence of keywords. This network is composed of 118 nodes and 1382 edges (Fig. 15). Our clustering layout was based on the modularity algorithm used to detect communities of keywords.
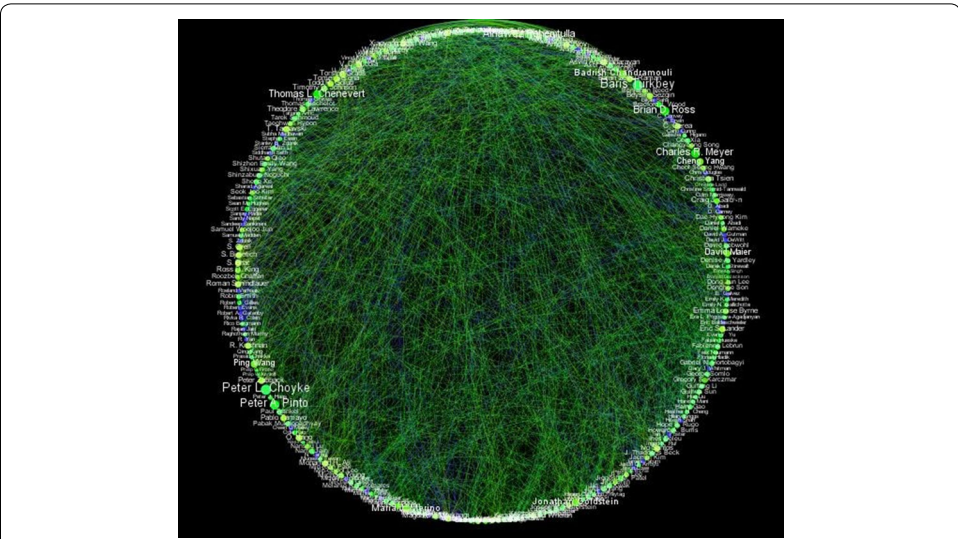
Fig. 16 Network of co-authors of scientific works cited by the patents
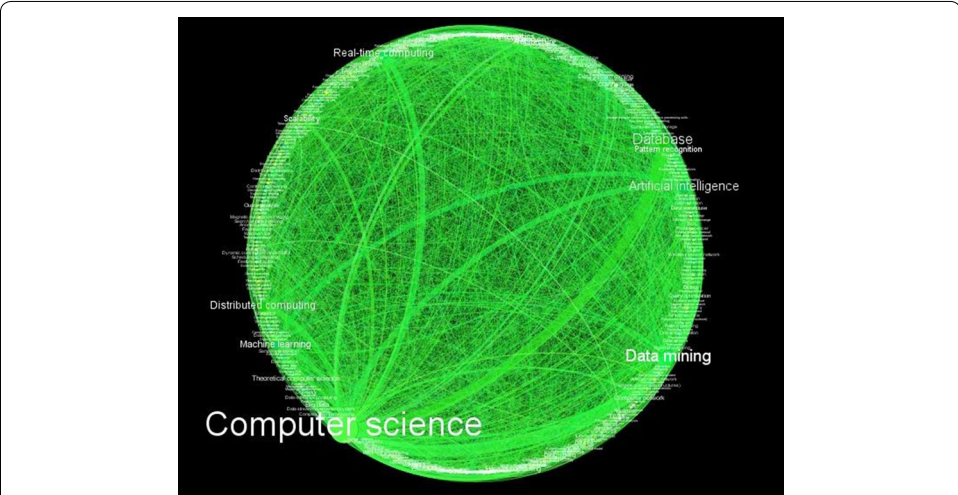

Fig. 17 Network of scientific fields of studies of the scholarly works cited by the patents

Some of the items in each cluster are shown in Table 3. Our clustering layout was based on a modularity algorithm used to detect keyword communities.

In this part of the paper, we also analyzed and visualized the network of co-authors (Fig. 16), consisting of 2298 nodes and 6363 edges. The average weighted degree of the network is 5.812; and average degree of 5.538. This analysis shows that the authors with the highest weighted degree are Baris Turkbey and Peter Choyke (42); Peter Pinto (38); and Charles Meyer, Brian Ross, Alnawas Rehemtulla, Thomas Chenevert (36). Ismail Baris Turkbey is associate research physician at the National Cancer Institute. Peter L Choyke is also the program director of the Molecular Imaging Program at the same center; Peter Pinto is also the head of the prostate cancer section at the same center. Charles R Meyer is a professor emeritus of radiology at the University of Michigan. Brian Ross is also from the University of Michigan and his field of study is radiology and molecular imaging. Alnawas Rehemtulla is the professor of radiation oncology and

Table 4 Some of fields of studies with the highest weighted degree

Field of study	Average weighted degree	Field	Average weighted degree	Field of study	Average weighted degree	Field of study	Average weighted degree	Field of study	Average weighted degree
Computer science	3773	Data mining	1642	Database	1288	Artificial intelligence	1177	Distributed computing	887
Real-time computing	843	Machine learning	732	Mathematics	554	Pattern recognition	553	Scalability	526
Architecture	509	Theoretical computer science	664	Cloud computing	438	Big data	437	Parallel computing	398
Computer network	359	Data stream mining	345	Computer vision	310	Engineering	286	Analytics	283
Data modeling	279	Query optimization	262	SQL	218	Data warehouse	218	Server	217
Statistics	217	Medicine	215	Biology	212	Cluster analysis	209	Data set	201
…	…	Bioinformatics	150	Pathology	131	Radiology	129	Physics	117
Econometrics	100	Cancer	100	Prostate cancer	93	Internet privacy	89	Chemistry	81
Molecular biology	81	Immunology	77	Analytical chemistry	75	Diffusion MRI	39	Biosensor	38

director of molecular imaging division. Thomas Chenevert is also professor of radiology and his interest is quantitative MRI in the assessment of treatment response.

Network of field of studies also ended up in 2110 nodes and 20,802 edges (Fig. 17). The average weighted degree is 27.041; and the average degree is 19.718. This network shows that the various scientific fields of big data have strong connections within the network. In Table 4, we have listed some of the fields with the highest weighted degree. Most of these fields are linked with computer science and engineering; however, as the table and the network shows, scientific fields such as medicine, biology, radiology, pathology, bioinformatics, physics, chemistry, cancer, chemistry, immunology and molecular biology have high interactions inside the network as well. The network of co-occurrence of keywords also confirmed the influential role of medicine and health in the scientific and invention productivity of big data.

Discussion and conclusion

This study incorporated bibliometric and social network analysis methods in order to discover the development trends of patenting activity on big data and the linkage of patents with the cited scientific literature over time; and also the interaction of agents, such as inventors, applicants, and the cited authors inside the social networks. To the best of our knowledge, this study is the first scholarly work to present a comprehensive and global comparison of the evolution of big data innovation, the link between patents

and the scientific world, and the strength of the connectivity of agents within social networks.

This analysis shows that China is at the forefront of filing global applications for patents on big data technology. US applications came a distant second; and only two US companies are on the list of top ten big data patent applicants; and Chinese firms and universities took the other eight top ten spots. A Chinese company is at the top of the ranking, the applications of which are more than twice the patent applications of IBM, which ranks second among the top ten. By January 2019, however, only 5% of patent applications in the jurisdiction of China are patented, while 28% of patents filed in the United States are patented.

This analysis also shows that the first, second and fifth inventors with the highest number of Big Data inventions (respectively 90, 58 and 55 inventions) are all Chinese; it shows that Chinese inventors have remained world leaders in the filing of Big Data patent applications. The inventor holding the third place is Indian with 57 inventions and the applicant for all of his inventions is the American Splunk Company. The company is ranked among the top ten major data applicants (i.e. the seventh rank). Consequently, this analysis shows the dominance of Chinese companies in the filing of patent applications and the dominance of the US in the granting of patents. This analysis also shows that the majority of inventors are not applicants for their inventions; they mainly work for large firms.

The analysis of jurisdictions with the highest number of patent applications filed also shows that China ranks first with 10,247 patent applications filed under that jurisdiction (i.e. 78% of global patent applications). Second, the United States has 1051 patent applications (i.e. 8% of global applications).

As far as the classification of patents is concerned, this analysis shows that most of the patent applications are filed under the IPC classification of G06F17/30, which refers to the retrieval of information and the structure of the database. This study shows the patent with the earliest priority date was submitted in 1989, with a hike of 491 patent applications in 2014; making this year as the most prolific year in the filing of patent applications on big data.

As far as the citation of scholarly works is concerned, this analysis shows a weak link between inventions and scientific works; as only 2.2% of global patent applications have cited scientific works. Patents that are filed in the US and Europe justifications have the highest linkage with the scholarly works compared to the other jurisdictions (respectively US with 15% citation and Europe jurisdiction with 14% citation to scholarly works). With an immense difference in the number of cited scholarly works, China cited less than 1% of patent applications. In addition, the highest number of citations occurred in 2012. The number of scholarly works cited was halved in 2014, which is also the peak of big data patent applications. T As a result, there appears to be a strong association between inventions and scholarly works before 2014, which has decreased in 2015. Although there is a mandatory rule in all jurisdictions for referring to prior art, including scholarly works, in China due to lack of enforcement for non-disclosure of information and the practices of the patent office, the number of references to scholarly works is very limited.

Fig. 18 A summary of main findings of the research
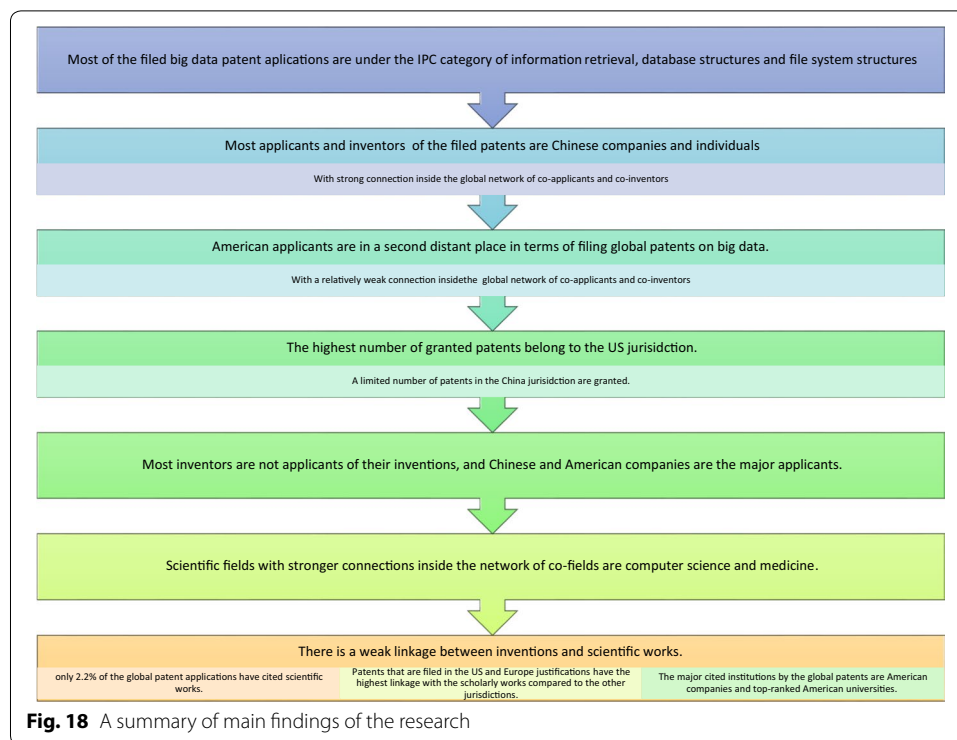
This analysis also shows that the State Grid China is very strongly linked to the other applicants in the network of co-applicants; whereas IBM, as the second applicant with the highest number of patent applicants, has a weak connection in the global network; compared to the State Grid, which is ranked first in terms of patent applications. This analysis shows that most Chinese applicants and inventors have more connections within the network than the American applicants; whereas applicants in the European and WIPO jurisdictions do not have connections within the network. In the network of Japan, US, Germany, European patents, Republic of Korea, WIPO, UK, France and Canada jurisdictions; however, most applicants have a same degree of connection with no huge difference among them. In this network, some applicants like Alibaba, Pepco ICT, IBM and Microsoft have formed small communities while most of the applicants have zero degrees and are disconnected from the network. Unlike the network of co-applicants in these specific jurisdictions, the network of co-inventors has strong connection among its applicants. The co-occurrence of keywords also shows that the majority of keywords belong to the fields of medicine and computer science. Authors with stronger connections within the network of co-authors are experts in areas such as cancer, radiology and molecular imaging. Scientific fields with stronger connections within the network also fall within the fields of computer science and engineering and medicine. As this research shows, one of the most promising areas of big data inventions is cancer treatment; suggesting that companies working on other complex diseases can also benefit from big data inventions; and shifting their R&D towards the creation of new big data technologies (Fig. 18).

Limitations and future research directions

There are some limitations to this study that should be acknowledged. In the first place, the data for this research included only patents and scholarly works cited in the patents. We did not have access to scholarly works that cited patents. Data on the cited scholarly works were also available until 2015; therefore, the link between patents and the cited scholarly works could not be measured after 2015. Third, big data has a very complex and wide landscape. We measured its innovation evolution in the broad term and did not focus on a specific layer of big data, such as analytics, software or hardware. The fourth limitation was that we did not measure each jurisdiction's patenting activities in detail; instead, we focused on a global comparison. In the light of the following limitations, future research may reduce the scope of analysis to specific layers of the big data landscape or to a specific jurisdiction or company. One of the possible future research also could be why China is leading filing big data patent.

Author details
[1] Management Studies Center, Tarbiat Modares University, Tehran, Iran. [2] Faculty of Law, Tarbiat Modares University, Tehran, Iran.

References
1. Lohr S. The age of big data. 2012. https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/GENPRESS/N120211L.pdf. Accessed 15 Jan 2019.
2. Gibson J. The logic of innovation: intellectual property, and what the user found there. 2016. https://content.taylorfrancis.com/books/download?dac=C2015-0-85030-2&isbn=9781317025214&format=googlePreviewPdf. Accessed 19 Oct 2019.
3. McAfee A, Brynjolfsson E. Big data: the management revolution. Harv Bus Rev. 2012;90(10):60–8.
4. Guadamuz A, Cabell D. Data mining in UK higher education institutions: law and policy, HeinOnline. 2014. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/qmjip4&section=4. Accessed 19 Oct 2019.
5. Columbus L. 10 Charts that will change your perspective of big data's growth, Forbes. 2018. https://www.forbes.com/sites/louiscolumbus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/#65a5f3c29268. Accessed 15 Jan 2019.
6. I. Property Office Patent Informatics Team. Eight great technologies big data a patent overview. 2014. http://www.ipo.gov.uk/informatics. Accessed 15 Jan 2019.
7. Archibugi D. Patenting as an indicator of technological innovation: a review. Sci Public Pol. 1992. https://doi.org/10.1093/spp/19.6.357.
8. Ernst H. The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. Small Bus Econ. 1997;9:361–81. https://doi.org/10.1023/A:1007921808138.
9. Ernst H. Patent information for strategic technology management. World Pat Inform. 2003;25:233–42. https://doi.org/10.1016/S0172-2190(03)00077-2.
10. Han Q, Heimerl F, Codina-Filba J, Lohmann S, Wanner L, Ertl T. Visual patent trend analysis for informed decision making in technology management. World Pat Inform. 2017;49:34–42. https://doi.org/10.1016/J.WPI.2017.04.003.
11. Comai A. Beyond patent analytics: insights from a scientific and technological data mashup based on a case example. World Pat Inform. 2018. https://doi.org/10.1016/j.wpi.2018.10.002.

12. de Carvalho Pereira F, Costa HG, Pereira V. Patent filings versus articles published: a review of the literature in the context of multicriteria decision aid. World Pat Inform. 2017;1(50):17–26.
13. Ke Q. Comparing scientific and technological impact of biomedical research. J. Inform. 2018;12:706–17. https://doi.org/10.1016/J.JOI.2018.06.010.
14. Dereli T, Profile S, Kusiak A, Altuntas S, Kusiak A. Forecasting technology success based on patent data. Technol Forecast Soc Change. 2015. https://doi.org/10.1016/j.techfore.2015.03.011.
15. Daim TU, Rueda G, Martin H, Gerdsri P. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol Forecast Soc Change. 2006;73:981–1012. https://doi.org/10.1016/j.techfore.2006.04.004.
16. Kim YG, Suh JH, Park SC. Visualization of patent analysis for emerging technology. Exp Syst Appl. 2008;34:1804–12. https://doi.org/10.1016/J.ESWA.2007.01.033.
17. Sternitzke C, Bartkowski A, Schramm R. Visualizing patent statistics by means of social network analysis tools. World Pat Inform. 2008;30:115–31. https://doi.org/10.1016/J.WPI.2007.08.003.
18. Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge: Cambridge University Press; 1994.
19. Saheb T, Saheb M. Analyzing and visualizing knowledge structures of health informatics from 1974 to 2018: a bibliometric and social network analysis. Healthc Inform Res. 1974;25(2019):61–72. https://doi.org/10.4258/hir.2019.25.2.61.
20. Saheb T, Izadi L. Paradigm of IoT big data analytics in the healthcare industry: a review of scientific literature and mapping of research trends. Telemat Inform. 2019;41:70–85. https://doi.org/10.1016/j.tele.2019.03.005.
21. Cascini G, Fantechi A, Spinicci E. Natural language processing of patents and technical documentation. Berlin: Springer; 2004. p. 508–20. https://doi.org/10.1007/978-3-540-28640-0_48.
22. Bergmann I, Butzke D, Walter L, Fuerste JP, Moehrle MG, Erdmann VA. Evaluating the risk of patent infringement by means of semantic patent analysis: the case of DNA chips. R&D Manag. 2008;38:550–62. https://doi.org/10.1111/j.1467-9310.2008.00533.x.
23. Gerken JM, Moehrle MG. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. Scientometrics. 2012;91:645–70. https://doi.org/10.1007/s11192-012-0635-7.
24. Chen Y-S, Chang K-C. Using neural network to analyze the influence of the patent performance upon the market value of the US pharmaceutical companies. Scientometrics. 2009;80:637–55. https://doi.org/10.1007/s11192-009-2095-2.
25. Lai Y-H, Che H-C. Modeling patent legal value by extension neural network. Expert Syst Appl. 2009;36:10520–8. https://doi.org/10.1016/J.ESWA.2009.01.027.
26. Lamirel JC, Al Shehabi S, Hoffmann M, François C. Intelligent patent analysis through the use of a neural network. In: Proceedings of the ACL-2003 workshop on Patent corpus processing, vol 20. Association for Computational Linguistics; 2003, pp. 7–23. https://doi.org/10.3115/1119303.1119305.
27. Fukuzawa N, Ida T. Science linkages between scientific articles and patents for leading scientists in the life and medical sciences field: the case of Japan. Scientometrics. 2016;106:629–44. https://doi.org/10.1007/s11192-015-1795-z.
28. Franceschini F, Maisano D. The evolution of large-scale dimensional metrology from the perspective of scientific articles and patents. Int J Adv Manuf Technol. 2014;70:887–909. https://doi.org/10.1007/s00170-013-5317-y.
29. de Moura AMM, Caregnato SE. Co-autoria em artigos e patentes: um estudo da interação entre a produção científica e tecnológica, Perspect. Em Ciência Da Informação. 2011;16:153–67. https://doi.org/10.1590/S1413-99362011000200010.
30. Yang Z, Zhang Z, Zhang S, Wang J, Lin H, Zeng B. Patent abstract analysis on Chinese big data. In: 2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE; 2017, pp. 2116–22.
31. Gui B, Liu Y, Bai X, Zhang J. Longitudinal patent analysis for big data technology. In: 2017 portland international conference on management of engineering and technology (PICMET). IEEE; 2017, pp. 1–8.
32. Jun S, Park S, Jang D. A technology valuation model using quantitative patent analysis: a case study of technology transfer in big data marketing. Emerg Mark Financ Trade. 2015;51:963–74. https://doi.org/10.1080/1540496X.2015.1061387.
33. Huang Y, Youtie J, Porter AL, Robinson DK, Cunningham SW, Zhu D. Big Data and Business: Tech mining to capture business interests and activities around Big Data. In: 2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom) (BDCloud-SocialCom-SustainCom). IEEE; 2016, pp. 145–50. http://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.32.
34. M. Rogers, The definition and measurement of innovation. Melbourne Institute of Applied Economic and Social Research. 1998. https://scholar.google.com/scholar_lookup?title=Thedefinitionandmeasurementofinnovation&author=M.Rogers&publication_year=1998. Accessed 3 Feb 2019.
35. Dangelico RM, Garavelli AC, Petruzzelli AM. A system dynamics model to analyze technology districts' evolution in a knowledge-based perspective. Technovation. 2010;30:142–53. https://doi.org/10.1016/J.TECHNOVATION.2009.09.006.
36. Akers L. The future of patent information—a user with a view. World Pat Inform. 2003;25:303–12. https://doi.org/10.1016/S0172-2190(03)00086-3.
37. Trappey AJC, Trappey CV. An R&D knowledge management method for patent document summarization. Ind Manag Data Syst. 2008;108:263–5577. https://doi.org/10.1108/02635570810847608.
38. Narin F, Noma E, Perry R. Patents as indicators of corporate technological strength. Res Policy. 1987;16:143–55. https://doi.org/10.1016/0048-7333(87)90028-X.
39. Furman JL, Porter ME, Stern S. The determinants of national innovative capacity. Res Policy. 2002;31:899–933. https://doi.org/10.1016/S0048-7333(01)00152-4.
40. Adams DO. Patents demystified: an insider's guide to protecting ideas and inventions. 2015.

41. United Nations Conference on Trade and Development. International centre for trade and sustainable development. UNCTAD-ICTSD Project on IPRs and Sustainable Development. Resource book on TRIPS and development. Cambridge: Cambridge University Press; 2005.
42. World Intellectual Property Organization. WIPO patent drafting manual. Geneva: World Intellectual Property Organization; 2007.
43. Albino V, Ardito L, Dangelico RM, MesseniPetruzzelli A. Understanding the development trends of low-carbon energy technologies: a patent analysis. Appl Energy. 2014;135:836–54. https://doi.org/10.1016/j.apenergy.2014.08.012.
44. Acs ZJ, Audretsch DB. Entrepreneurship, innovation and technological change. Found Trends Entrep. 2005;1:149–95. https://doi.org/10.1561/0300000004.
45. Bessen J. The value of US patents by owner and patent characteristics. Res Policy. 2008;37:932–45. https://doi.org/10.1016/j.respol.2008.02.005.
46. Zhang T, Fang J. Mining ideas for diamonds : comparing China and US IP practices from invention selection to patent monetization; 2016.
47. Trajtenberg M. A penny for your quotes: patent citations and the value of innovations. RAND J Econ. 1990;21:172. https://doi.org/10.2307/2555502.
48. Abraham BP, Moitra SD. Innovation assessment through patent analysis. Technovation. 2001;21:245–52. https://doi.org/10.1016/S0166-4972(00)00040-7.
49. Ayyappan G, Nalini C, Kumaravel A. A study on SNA: measure average degree and average weighted degree of knowledge diffusion in Gephi. Int J Phram Technol. 2016;8. http://www.arnetminer.org. Accessed 17 Jan 2019.
50. Khokhar D. Gephi cookbook : over 90 hands-on recipes to master the art of network analysis and visualization with Gephi; 2015.
51. Emilien G, Weitkunat R, Lüdicke F. Consumer perception of product risks and benefits. Berlin: Springer; 2017. https://books.google.com/books?id=ag9hDgAAQBAJ&source=ttb. Accessed 20 Jan 2019.
52. Ivakhnenko AG. Heuristic self-organization in problems of engineering cybernetics. Automatica. 1970;6:207–19. https://doi.org/10.1016/0005-1098(70)90092-0.
53. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286:531–7.

Publisher's Note