TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT
https://doi.org/10.1080/09537325.2020.1815693

Research status and collaboration analysis based on big data mining: an empirical study of Alzheimer's disease

Rongrong Li[a,b], Xuefeng Wang[b], Yuqin Liu[c], Shuo Zhang[b] and Omer Hanif[b]

[a]School of Economics and Management, China University of Petroleum (East China), Qingdao, People's Republic of China; [b]School of Management & Economics, Beijing Institute of Technology, Haidian District, People's Republic of China; [c]School of Journalism and Publication, Beijing Institue of Graphic Communication, Beijing, People's Republic of China

ABSTRACT

This paper employs text mining techniques that aimed to facilitate technology information. First, this paper used patent data to monitor technological development trends systematically to show the technology research status from perspectives of country, institution, technology fields, and subjects. Secondly, this study explores the cooperation network institutions and inventors by applying the data mining approaches, social network analysis,. Additionally, the sequence analysis is applied to reveal a more comprehensive and objective appearance of cooperative relationships, partners, and centrality. The empirical findings reveal four significant observations. (1) The R&D centres have been mainly influenced by the United States and other developed countries. (2) All technological fields in both B IPC and Derwent manual codes are concentrated around pharmaceutical activities. (3) 1-6c alkyl, pharmaceutical composition, and central nervous system et al. are traditional research and core subjects. 2-6c alkenyl, amino acid sequence, and 1-3c alkoxy et al. are the hot subjects. (4) The influential institutions are HOFFMANN LA ROCHE & CO AG F (degree centrality is 0.0872), ASTRAZENECA AB, MERCK SHARP & DOHME CORP, PFIZER INC and UNIV CALIFORNIA, INCYTE GENOMICS. (5) The influential inventors are WANG Y, BACHER G, and PETERS D.

ARTICLE HISTORY
Received 4 February 2020
Revised 29 June 2020
Accepted 20 August 2020

KEYWORDS
Research status; data mining; Alzheimer's disease; technology trend

1. Introduction

Alzheimer's disease (AD) is a type of primary senile dementia and the most common type of dementia. It is a general term used to classify memory loss symptoms or other severe enough to interfere with daily life, predominantly in the elderly population (Alzheimer's Association 2018; Selkoe 2019; Veitch et al. 2019). Today, someone in the World develops Alzheimer's disease every three seconds (Alzheimer's Disease International 2019). The reported deaths due to Alzheimer's disease and other dementias has been increased from 800,000 in 2010–1.99 million in 2016, depicting a fifth leading cause of global death in 2016 from 14th place in 2000 (World Health Organization 2018). The death rate was 97% of Alzheimer's patients over 60 years of age from 2010 to 2016. Recently, China is ranked first globally, with more than 7 million Alzheimer's patients (Jia et al. 2018), and 3.21% of those patients are over 65 years of age (Jia et al. 2014), thereby indicating the fastest growth rate in the world (Figure 1).

CONTACT Xuefeng Wang wxf5122@bit.edu.cn School of Management & Economics, Beijing Institute of Technology, Haidian District, Beijing, People's Republic of China

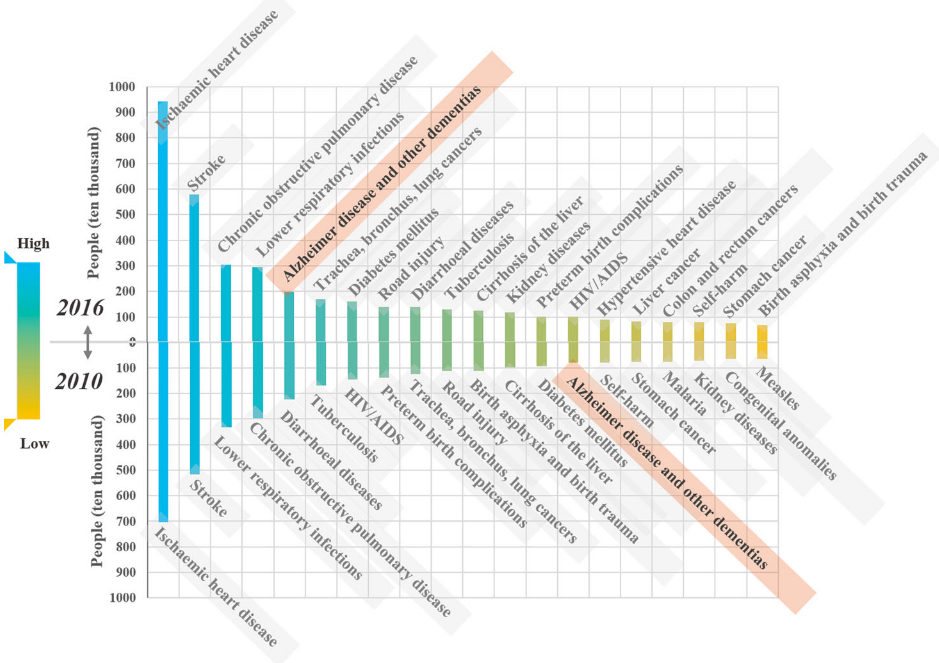

Figure 1. The top 20 causes of death worldwide in 2010 and 2016.

The disability from disability Alzheimer's disease brings a high mortality rate of up to 55%, with a relatively higher need for medical and personal care for an expended time. A variety of drugs are currently available to alleviate the symptoms of Alzheimer's disease; however, there is no way to cure or slow the progression of the disease. At present, the success rate of drug development projects is very low, and successful R&D of drugs to combat Alzheimer's disease faces many challenges. Till now, only four drugs concerning Alzheimer's disease have been approved out of 146 drugs that remained unsuccessful in clinical trials between 1998 and 2017as presented in Figure 2. Accordingly, only one research project generates a new drug out of approximately 37 substitutions – the success rate is only 2.7% [43].

Research on Alzheimer's disease brings significant empirical and theoretical contributions. The etiology and pathogenesis of Alzheimer's disease are still uncertain, which further backed by the low success rate of drug development. It is not easy to achieve significant technological innovation and breakthroughs by relying on a single research and development institution. Accordingly, strengthening the cooperation innovation demands urgency to achieve technological innovation and breakthroughs. A systematic understanding and monitoring of technological development trends, and analysing the cooperation network more comprehensively and objectively can accelerate more support for cooperation innovation. This paper comprehensively and objectively presents the research status and trend of Alzheimer's disease research from the perspective of applied research to demonstrate the cooperation network from the macro and micro levels to facilitate the support from partners and directions for cooperation innovation. Therefore, data mining approaches, including text analysis, adopts software, and sequence analysis, are employed.

2. Literature review

The first category focused on improving the technology development status and trend of Alzheimer's disease research. Ansari, Gul et al. used bibliometric analysis to explore the distribution of literature

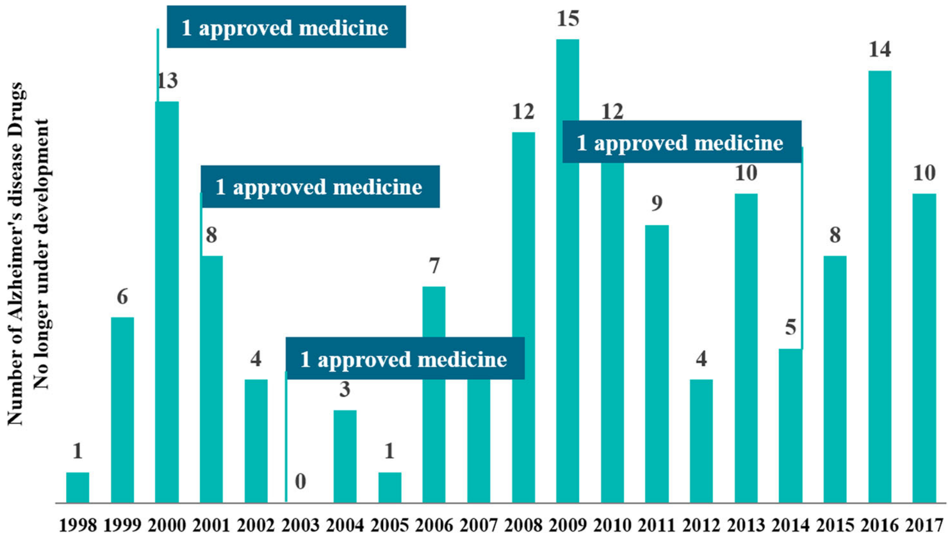

Figure 2. Unsuccessful investigational drugs for Alzheimer's disease (1998–2017) ('Alzheimer's Medicines: Setbacks and Stepping Stones' 2018).

about Alzheimer's disease collected from Index Medicus from different parameters like country, authorship, and production. (Ansari, Gul, and Yaseen 2006). Asghar, Cang et al. examined the characteristic of research activities on assistive technology of dementia using literature collected from Scopus on country perspective and found that the USA and UK were working extensively in AT research (Asghar, Cang, and Yu 2017). Chen, Wan et al. focused on presenting the development of cholinesterase inhibitors on Alzheimer's disease and sorting out the Sequence of drugs that were most tolerated or more effective in AD treatment (Chen et al. 2014). Dong, Wang et al. compared the reach and influence of Alzheimer's disease using research publications in China employing Web of Science and PubMed databases during 1988–2017 (Dong et al. 2019). Xu, Kong et al. applied a patent citation network method to form multiple integrated technology clusters and discover the technology flow of anti-AD medicines, i.e. the evolution of technology by 329 US patents from 1978 to 2013 (Xu et al. 2014). Theander and Gustafson applied the quantitative bibliometric analysis to show the development of publications related to dementia in Medline from 1974 to 2009 (Theander and Gustafson 2013). Pettersson, Stepan et al. focused on reading and reviewing the patents concerning g-Secretase modulators in the period 2010–2012. They concluded that a higher percentage of potent GSM chemical matter, used in central nervous System drug space, may utilise in testing the GSM mechanism of action (Pettersson et al. 2013).

The second perspective described the cooperation in Alzheimer's disease. Sorensen, Seary, et al. explored the co-authorship network analytics of Alzheimer's disease research and revealed the two-step collaboration networks of co-authors and co-authors of co-authors to locate the bridge (Sorensen, Seary, and Riopelle 2010). Song, Heo et al. explored the metadata of 96,081 articles from PubMed and presented the analysis of Alzheimer's disease literature at two levels: at the macro-level including author, journal, and institution analysis of the literature; and micro-level network analysis (keyword co-occurrence frequency) (Song, Heo, and Lee 2015). Ivinson, Lane et al. subjectively described the cooperation possibility between senior industry researchers and academic investigators on the drug discovery and development of Alzheimer's disease. Carrillo, Blennow et al. reported that international collaboration was critical for the acceleration of biomarker standardisation efforts and the efficient development of improved diagnosis and therapy (Carrillo et al. 2013). Jones-Davis and Buckholtz exercise the Alzheimer's Disease Neuroimaging Initiative 2 as an example to examine the effect of public-private partnerships on pushing the boundaries of clinical and basic science research on Alzheimer's disease (Jones-Davis and Buckholtz 2015).

Besides, the software ItgInsight, applied in this paper, is the only software developed by China in the visualisation field, which can facilitate the analysis of big data, refining the term recognition, and introduce Sequence to distinguish contribution compared with other visualisation tools. The visualisation tools, such as UCNET, Pajak, and Vxinsight, are mainly used to analyse structured data; therefore, they are applied with other data mining tools for analysing scientific literature data. True-Teller, VosViwer, Vantage-Point, Thomson Data Analyser and ItgInsight are visual analysis software based on semantics, and capable of performing data mining functions, realising the text analysis, and visual display of structured and unstructured text data in the field of intelligence analysis.

To summarise, it is obvious that the earlier related literature' perspective focused on technology development on macro level. As far as the cooperation research is involved, it mainly sorts out the current cooperation situation and partner subjectively in the field, while lacking quantitative analysis. Therefore, this paper employed data mining tools, and visualisation software to comprehensively and objectively explore the research status and trend of Alzheimer's disease research, and demonstrates the cooperation network from the macro and micro levels to facilitate support from the perspective of partners and directions for cooperation innovation.

3. Data source and methodology

3.1. Data source

The Derwent World Patents Index (DWPI) is a comprehensive database of global patents covering all technical fields (Madani, Daim, and Weng 2017; Sampaio et al. 2018). The data in the DWPI database comes from 48 patent issuing agencies and two literary sources worldwide, including patent data from the Asia Pacific, Europe, the Middle East, and North and South America (Oppenheim 1981). Currently, DWPI consists of over 23 million unique inventions (basic records/patent series), covering more than 50 million patent documents, which are used by thousands of organisations and 40 patent offices worldwide, presenting it the most trustworthy and authoritative patent database (de Oliveira et al. 2018). The DWPI database presents the most comprehensive view of activity in emerging markets by providing the world's most comprehensive English patient data set – more than 50 authoritative institutions and covering more than 30 languages (Emmerich 2009) – and nearly 86% of DWPI summary records are from English patient records.

Given the DWPI advantages mentioned-above, this paper employed the DWPI database as a data source. The Metathesaurus of Unified Medical Language System reveals 75 synonyms of Alzheimer's disease. Accordingly, the abbreviations are observed to be prone to ambiguity rather than proprietary abbreviations. Subsequently, ambiguous words and abbreviations were eliminated to customise the search strategy, as shown in Table 1. The data was updated to June 7, 2019, and a total of 48,268 patent data was collected.

Table 1. The search strategy of Alzheimer's disease.

Search strategy	Search results
TS = ((Alzheimer* (disease* or dementia)) or Alzheimer or (Alzheimer (Dementia* or (Sclerosis or Syndrome) or (Type Dementia) or (Alzheimer Type Senile Dementia))) or (Dementia ((Alzheimer's type) or Alzheimer* or (of The Alzheimer* Type) or (Alzheimer's type) or (in Alzheimer's disease) or (of Alzheimer's Type))) or (Primary Senile Degenerative Dementia) or (Senile Dementia of The Alzheimer Type) or (Senile Dementia) or (simple senile dementia)) AND PY = (2000–2018)	48,268

3.2. Methods

3.2.1. Bibliometric analysis

The bibliometric analysis is retrieval and exploration of relevant scientific literature indexed in a recognised database of scientific articles (Bugge, Hansen, and Klitkou 2016; Davidse and Van Raan 1997). Bibliometric is a set of methods to analyse scientific and technological literature quantitatively (Nicolaisen 2010; Raan 2005; Wang and Chen 2010). Bibliometric aims to quantify the outcome and interconnection of scientific activity. For instance, the number of publications is the commonly used measure of scientific output, and the number of citations is the most popular indicator of scientific impact, a measurable aspect of quality (Haunschild, Bornmann, and Marx 2016). As a more unambiguous definition given by White and McCain (Godin 2006), bibliometrics is the quantitative study of literature as reflected in bibliographies; its task is to provide evolutionary models of science, technology, and scholarship (Wang and Li 2016). Bibliometrics is a useful tool to map the literature around a research field. It refers to research methodology employed in library and information sciences, which utilises statistics and quantitative analysis methods to describe distribution patterns of articles with a given topic, field, institute, or country (Braun, Glänzel, and Grupp 1995).

3.2.2. Social network

A social network is characterised as a set of people or groups each of which form connections to some or all the others. In the language of social networks, the people or groups are referred to as actors (Reuther 2006). Lo Re applied the network analysis techniques to identify different sectors and at several stages of the value chain to contextualise a set of isolated elements (Lo Re and Veglianti 2017). Monarca, Umberto, et al. discovered and identified networks' main metrics from the perspective of economic between manufacturing and other industries in China and Italy (Monarca et al. 2019). Lo Re provided a conceptual framework to classify the studies and the recent applications of network analysis in the economic field (Lo Re 2018). In this study, social network analysis (SNA) is employed as the main method, including the network structure, for example, drawing the collaboration maps to analyse author collaborative situations. Social networks have been the subject of empirical and theoretical studies in the social sciences for at least 50 years, partly because of inherent interest in the patterns of human interaction, and because their structure has important implications for the spread of information (Newman 2001). In this method, actors in the network are positioned to represent nodes, and the relationships between them represented the links (Cross, Borgatti, and Parker 2007; Wang et al. 2014). Similarly, a collaboration network analysis is one kind of social network analysis and a social network is a network of social relations, reflecting a relationship between actors. In this way, a collaboration network is one kind of social network analysis.

In this paper, we mainly use the ITGinsight, software developed by Yu-qin Liu, to conduct our social network analysis. It is a generic approach to detecting and visualising emerging trends and transient patterns in scientific literature. The basic principles of collaboration analysis and research topic evolution analysis performed in the software are described as follows. The collaboration analysis includes two dimensions: the establishment of the collaboration relationship matrix, and clustering analysis. The research topic evolution analysis mainly uses data mining techniques.

(1) Principle of collaboration analysis

Step 1: Build collaboration relationship matrix

The specific implementation of this method is described as follows:

(a) Identify the countries in the patent document, using the national name dictionary to standardise and merge the same countries, and establish standardised membership matrix B between the country and the document.

$$B = \begin{bmatrix} & P_1 & P_2 & \cdots & P_j & \cdots & P_m \\ B_1 & b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1m} \\ B_2 & b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ B_i & b_{i1} & b_{i2} & \cdots & b_{ij} & \cdots & b_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ B_n & b_{n1} & b_{n2} & \cdots & b_{nj} & \cdots & b_{nm} \end{bmatrix}$$

Where, $b_{ij} = 1$ indicates whether the document P_j belongs to the country i, and $b_{ij} = 0$ indicates whether the document P_j does not belong to the country i respectively.

(a) Record the sequence of each academic subject that appears in each document. The cooperation relationship matrix's elements represent the number of countries cooperating, and the sum of the rows or columns represents the number of documents issued by the country. Simultaneously, the sequence of the country depends on the sequence of the author belongs to the country. For the convenience of research, this paper regards the countries where the authors of the literature belong to the third place and their follow-up countries as the third place.
(b) This relationship is mapped to the nodes and connections in the network graph. The thickness of the connection indicates the number of collaborations, while the size of the node indicates the number of documents published by the country.
(c) With the membership matrix B between the country and the document, building the national cooperation matrix BB'.

Step 2: Cluster analysis based on the construction of cooperation matrix

Shotorbani et al. used K-means clustering algorithm and LDA (Latent Dirichlet Allocation) topic modelling technology to extract topic patterns in the manufacturing corpus (Shotorbani et al. 2016).

Data set X can be composed of n samples, that is, $X = \{x_1, x_2, \ldots, x_n\}$. There are m sets $C_1, C_2, \ldots, C_m$, which are split from the data set X, and the following two conditions should be met (Jain and Dubes 1988):

$$\bigcup_{i=1}^{m} C_i = X, i = 1, 2, \ldots, m \qquad \forall C_i \neq \emptyset$$

$$C_i \cap C_j = \emptyset, \ \mathrm{i, j} = 1, 2, \ldots, \mathrm{m} \qquad \forall \mathrm{i} \neq \mathrm{j}$$

Then $C_1, C_2, \ldots, C_m$ are called clusters based on the data set X. It can be seen from the constraints that in the clustering process, each sample point must belong to only one set cluster, and the clusters should not intersect.

In this study, we use K-means clustering algorithm to achieve cluster analysis. K-means clustering algorithm is the most commonly used algorithm in cluster analysis. The steps of the clustering process are as follows (Anandarajan, Hill, and Nolan 2019; Gaitani et al. 2010):

(a) Select k objects from the original data set as random cluster centres;
(b) Calculate the distance between each object in the original data set and the cluster centre, then assign it to the nearest cluster centre;
(c) Calculate the mean of the objects in each cluster as the new cluster centre;
(d) If the cluster centres change, return to step 2 until k collection classes' cluster centres no longer change.

When calculating the distance between the object and the cluster centre in b, Euclidean distance is usually used. The Euclidean distance between two objects in a dataset refers as the square root of the sum of the squares of the differences in the dimensions between the two objects, calculated as (Agrawal, Faloutsos, and Swami 1993):

$$d_{ED} = \sqrt{\sum_{i=1}^{N} (x_{1i} - x_{2i})^2}$$

Among them, d_{ED} represents the Euclidean distance between two objects, $i = 1, 2, \ldots, N$ represents the dimension. Besides, the Manhattan distance (d_{MD}) is often used in K-means clustering analysis. The calculation formula is (Medrano-Marqués and Martín-del-Brío 1999):

$$d_{MD} = \sum_{i=1}^{N} |(x_{1i} - x_{2i})|$$

(2) Principle of research topic evolution analysis

(a) Read text data;
(b) Text preprocessing: Use the ending symbol of the sentence as a unit to split the text data into a single sentence and remove the irrelevant stopwords;
(c) Part-of-speech tagging: segment each single sentence in the pre-processed sentence and divide it into words or phrases;
(d) Use the vocabulary in the constructed field and the similarity-based merge command to perform preliminary merging of words or phrases. This step can remove numbers, stop words, and many general words that are not related to the analysis topic. Then form a keyword list.
(e) Count the frequency of keywords every year. The high-frequency words finally obtained can represent the research topic.

4. Analyses and results

4.1. Temporal distribution of research on Alzheimer's disease

Import 48,268 patent data into VantagePoint analysis software to analyse the trend of patent number in the field of Alzheimer's disease over the years (575 patent data lack year information), as shown in Figure 3. Figure 3 shows the trend of patent productivity in the field of Alzheimer's disease during

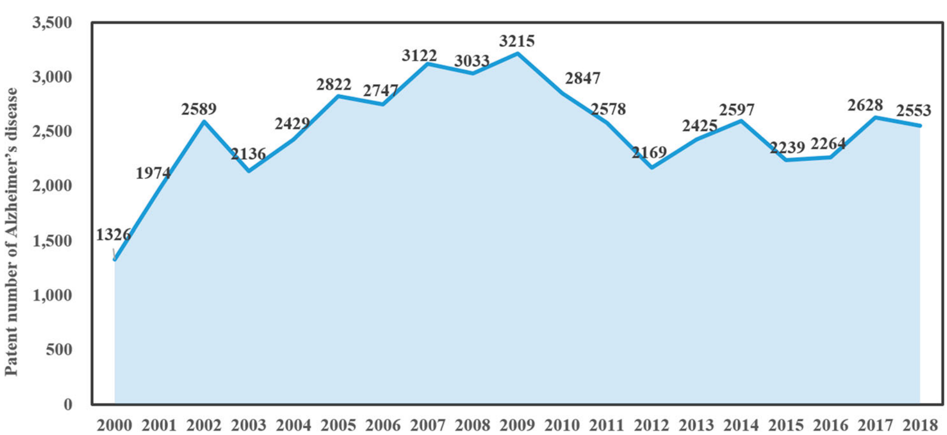

Figure 3. Temporal distribution of research on Alzheimer's disease during 2000–2018.

2000–2018. It was observed that the research of Alzheimer's disease has significantly fluctuated from 2000 to 2018 and divided into two periods. The patent number shows an overall growth trend from 2000 to 2009 reaching a peak of 3215 in 2009, since 2009, the number of global patent applications appears volatile decrease, a falling to 2,553 in 2018.

4.2. Spatial distribution of research from countries/territories, institution

Figure 4 reflects the top ten patent-granting countries in the field of Alzheimer's disease, which are respectively WIPO, the United States, China, Japan, South Korea, Spain, Germany, France, the United Kingdom, and India. The red nodes represent the patent-granting country, and the yellow nodes represent the patent-granting organisation. The size of the nodes shows the patent number. Reflecting the technology R & D strength of patentees in patent-granting countries, the patent number granted by WIPO, the United States, and China are 27829, 10007 and 3971 respectively, accounting for 58%, 21% and 8% of the total patents, the rest seven countries account for only 11% of the total patents. It means that over 50% of patents are granted through the PCT.

In the Derwent database, about 2,100 companies worldwide have a unique 4-character code as the patentee code, which is considered as a standard company; its subsidiaries and related institutions share the same patentee code. Universities and research institutions also have a unique 4-character code as the patentee code. For a very small number of non-standard companies and individuals, a non-standard 4-character code will also be assigned as the patentee code. This code is not unique. Additionally, the top ten institutions (including companies, universities, and research organisations) are all standard companies.

We use the Vantagepoint software to clean the patent data and acquire the patentee codes. Table 2 presents the top ten patentee codes and the patent number. The patentee codes were ranked by the patent number. The top ten patentees are all large pharmaceutical companies. The most significant contributor, MERCK SHARP & DOHME CORP (MERI-C), owns 1058 patents. HOFFMANN LA ROCHE & CO AG F (HOFF-C), PFIZER INC (PFIZ-C), and ASTRAZENECA AB (ASTR-C) come next, respectively, contributing 943, 847and 784. Then, followed by BAYER HEALTHCARE AG (FARB-C), TAKEDA PHARM CO LTD (TAKE-C), GLAXO GROUP LTD (GLAX-C), BRISTOL-MYERS SQUIBB CO (BRIM-C), NOVARTIS AG (NOVS-C), and WYETH (AMHP-C). In this field, MERI-C and HOFF-C have strong R & D capabilities. At the same time, among the top ten patentees, four companies are from the United States, two from Switzerland, and remaining from Germany, Japan, Sweden, and the United Kingdom.

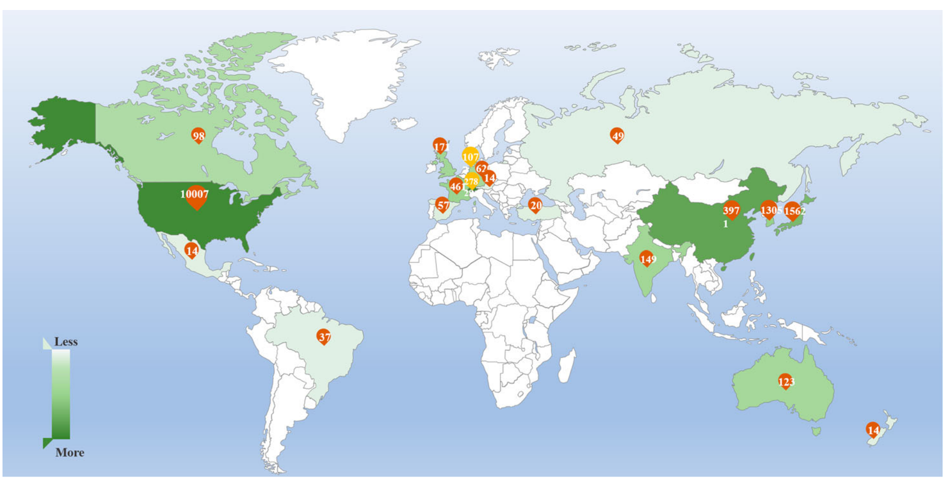

Figure 4. Countries/territories distribution of research on Alzheimer's disease during 2000–2018.

Table 2. Top ten patentees (institutions) in Alzheimer's field.

Number	Patent number	Patentee code	Company name	Country
1	1058	MERI-C	MERCK SHARP & DOHME CORP	United States
2	943	HOFF-C	HOFFMANN LA ROCHE & CO AG F	Switzerland
3	847	PFIZ-C	PFIZER INC	United States
4	784	ASTR-C	ASTRAZENECA AB	Sweden
5	752	FARB-C	BAYER HEALTHCARE AG	Germany
6	736	TAKE-C	TAKEDA PHARM CO LTD	Japan
7	668	GLAX-C	GLAXO GROUP LTD	United Kingdom
8	537	BRIM-C	BRISTOL-MYERS SQUIBB CO	United States
9	497	NOVS-C	NOVARTIS AG	Switzerland
10	481	AMHP-C	WYETH	United States

4.3. Main technology fields of research on Alzheimer's disease

(a) IPC codes
(b) Derwent manual codes

The International Patent Classification (IPC) is a contemporary international universal classification and search tool for patent documents. The unique classification code and indexing system are Derwent manual codes. The DWPI database uses Derwent manual codes to manually classify and index patents in major patent granting agencies and all technical fields in the world. Both IPC and Derwent manual codes can recover the technology fields. Figure 5(a) indicates the main technology fields under IPC codes are focused on A61K, A61P, and C07D, respectively accounts for 26.70%, 21.12% and 13.20%. They stand for preparations for medical, dental, or toilet purposes, specific therapeutic activity of chemical compounds or medicinal preparations, and heterocyclic compounds. The more specific information of top ten technology fields under IPC codes is shown in Table 3.

In addition, the Derwent manual codes reveal more specific technology fields. Figure 5(b) shows the main technology fields under Derwent manual codes are focused on B14-J01, B14-N16, B14-J01A4, respectively accounts for 2.23%, 2.18%, and 1.97%. They stands for CNS active general and other Covers terms such as cerebroprotective and neuroprotective, Brain and spinal cord Including stroke, meningitis, encephalitis and other prion type diseases and Alzheimer's, Huntington's, senility, senile dementia, cognitive enhancer, ant amnesia, nootropics. After them, the other top technology fields are B14-C03, B14-H01, B14-S04, B14-S01, B14-F02, B14-N03, and B14-F01. It is worth mentioning that the top ten technology fields all belong to B14, which means pharmaceutical activities. The more specific information of the top ten technology fields under Derwent manual codes is shown in Table 4. The IPC and Derwent manual codes are both concentrated on pharmaceutical activities.

4.4. Main subjects of research on Alzheimer's disease

Two steps were followed to further refine the understanding of research subject in this field. First, we employed the software ITGinsight developed by co-author Yuqin Liu to acquire the top keywords reflecting the subjects and the matrix of co-occurrence keywords; the top 50 keywords with the high-frequency words of the abstract were extracted. Second, we applied the Unicet software to draw the co-occurrence analysis of main subjects. The node represents the subject; the node's size presents the centrality, bigger the size of the node, higher the centrality. The line presents the co-occurrence and correlation.

The top 50 subjects are all related to pharmaceutical ingredient research, forming an interconnected and closely linked network. Firstly, Figure 6 reveals an intuitive observation that firstly 1-6c alkyl, pharmaceutical composition, neurodegenerative disease, multiple sclerosis, neurodegenerative diseases, rheumatoid arthritis, neurodegenerative disorder, inflammatory diseases, cognitive

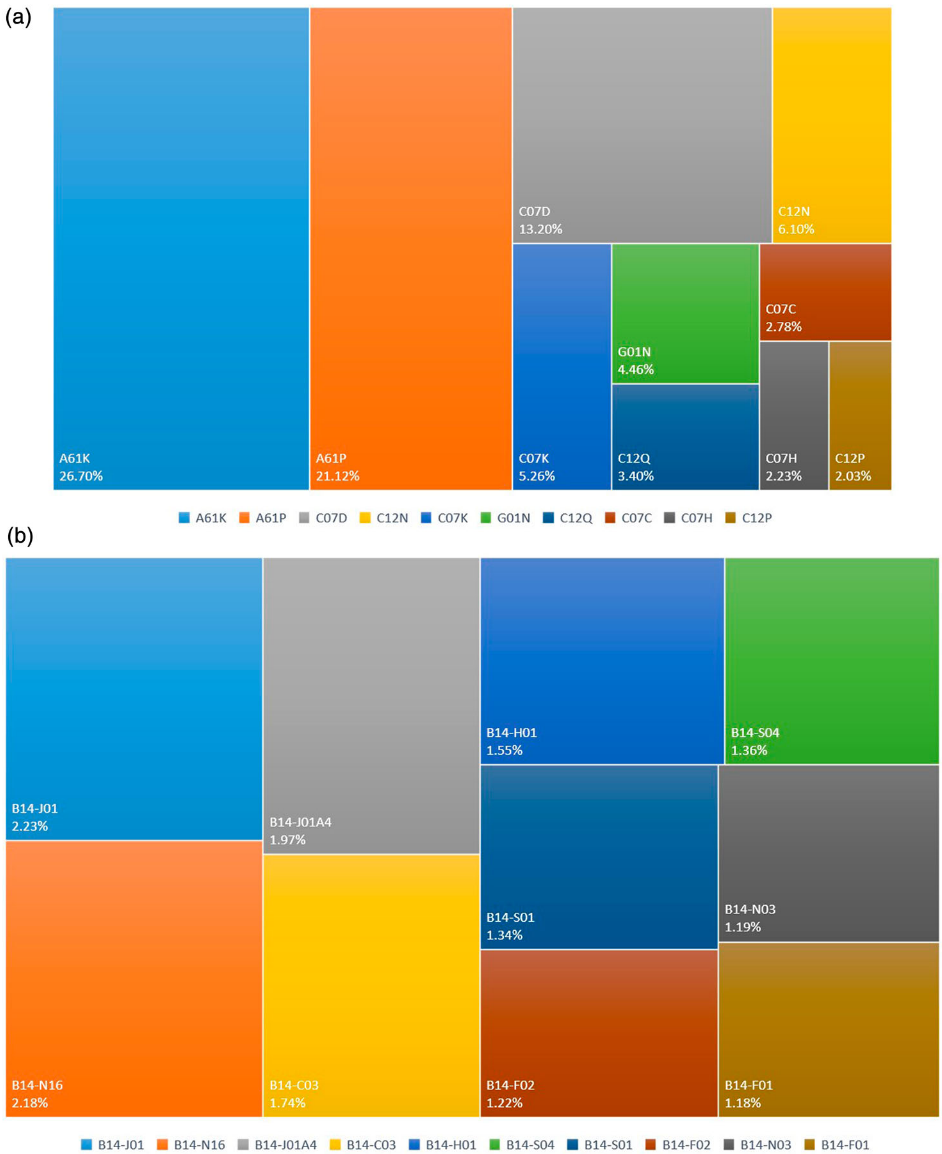

Figure 5. Main technology fields of research on Alzheimer's disease under IPC and Derwent manual codes.

impairment, autoimmune diseases, and central nervous system are at the centre of the network and have a higher degree of centrality, implying that these keywords and other keywords appear in the same document most frequently. These keywords represents the core of Alzheimer's research, i.e. other research fields are centred around core subjects. In addition, 1-4c alkyl, 1-6c alkoxy, lower alkyl, test compound, 3-8c cycloalkyl, nucleic acid, nucleic acid molecule, 3-7c cycloalkyl, 1-10c alkyl, active ingredient, and heterocyclic compounds are in the middle of the network, which act as the bridge connecting the edges and core subjects of the network. Thirdly, 2-6c alkenyl, amino acid sequence, 3-6c cycloalkyl, 1-4c alkoxy, 2-6c alkynyl, 1-3c alkyl, 1-8c alkyl, biological sample, 6-10c aryl, acid sequence, nucleotide sequence, 1-6c haloalkyl, therapeutic agent, degree c, host cell, lower alkoxy, stem cell, heterocyclic ring, 3-10c cycloalkyl, inflammatory disease, 1-5c alkyl, neurological disorder, autoimmune disease, senile dementia, pathological condition, amino acids, candidate

Table 3. The specific information of top ten technology fields under IPC codes.

IPC codes	Specific information of technology field
A61K	Preparations for medical, dental, or toilet purposes (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms a61j 3/00; chemical aspects of, or use of materials for deodorisation of air, for disinfection or sterilisation, or for bandages, dressings, absorbent pads or surgical articles a61l; soap compositions c11d)
A61P	Specific therapeutic activity of chemical compounds or medicinal preparations
C07D	Heterocyclic compounds (macromolecular compounds c08)
C12N	Microorganisms or enzymes; compositions thereof (biocides, pest repellants or attractants, or plant growth regulators containing microorganisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, microorganisms or animal material a01n 63/00; medicinal preparations a61k; fertilisers c05f); propagating, preserving, or maintaining microorganisms; mutation or genetic engineering; culture media (microbiological testing media c12q 1/00)
C07K	Peptides (peptides containing β-lactam rings c07d; cyclic dipeptides not having in their molecule any other peptide link than those which form their ring, e.g. Piperazine-2,5-diones, c07d; ergot alkaloids of the cyclic peptide type c07d 519/02; single cell proteins, enzymes c12n; genetic engineering processes for obtaining peptides c12n 15/00)
G01N	Investigating or analysing materials by determining their chemical or physical properties (measuring or testing processes other than immunoassay, involving enzymes or microorganisms c12m, c12q)
C12Q	Easuring or testing processes involving enzymes, nucleic acids or microorganisms (immunoassay g01n 33/53); compositions or test papers therefor; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes
C07C	Acyclic or carbocyclic compounds (macromolecular compounds c08; production of organic compounds by electrolysis or electrophoresis c25b 3/00, c25b 7/00)
C07H	Sugars; derivatives thereof; nucleosides; nucleotides; nucleic acids (derivatives of aldonic or saccharic acids c07c, c07d; aldonic acids, saccharic acids c07c 59/105, c07c 59/285; cyanohydrins c07c 255/16; glycals c07d; compounds of unknown constitution c07g; polysaccharides, derivatives thereof c08b; dna or rna concerning genetic engineering, vectors, e.g. Plasmids, or their isolation, preparation or purification c12n 15/00; sugar industry c13)
C12P	Fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture
A61K	Preparations for medical, dental, or toilet purposes (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms a61j 3/00; chemical aspects of, or use of materials for deodorisation of air, for disinfection or sterilisation, or for bandages, dressings, absorbent pads or surgical articles a61l; soap compositions c11d)
A61P	Specific therapeutic activity of chemical compounds or medicinal preparations

Table 4. The specific information of top ten technology fields under Derwent manual codes.

Derwent manual codes	Specific information of technology field
B14-J01	Cns active general and other covers terms such as cerebroprotective and neuroprotective.
B14-N16	Brain and spinal cord including stroke, meningitis, encephalitis and other prion type diseases
B14-J01A4	Alzheimer's, huntington's, senility, senile dementia, cognitive enhancer, antiamnesia, nootropics
B14-C03	Antiinflammatory general this code is used for treatment of general oedema or inflammation. Specific inflammation treatments are coded elsewhere when possible e.g. Bronchitis is coded b14-k01 only, colitis as b14-e10c only etc.
B14-H01	Anticancer general and other
B14-S04	Diabetes this code is used when a drug targets the symptoms and associated dissequences. Hypoglycaemic is coded b14-f09
B14-S01	Multiple sclerosis treatment, demyelinating diseases
B14-F02	Circulatory active general and other
B14-N03	Eye dissequence treatment
B14-F01	Cardioactive general and other

compound and 1-3c alkoxy are at the edge of the network, which exhibit that they are the hot subjects of Alzheimer's research.

4.5. Collaboration network of research on Alzheimer's disease

We use the software ITGinsight developed by Yu-qin Liu to identify the collaboration network and research topic evolution analysis. The map was built using the top 50 institutions, inventors, where

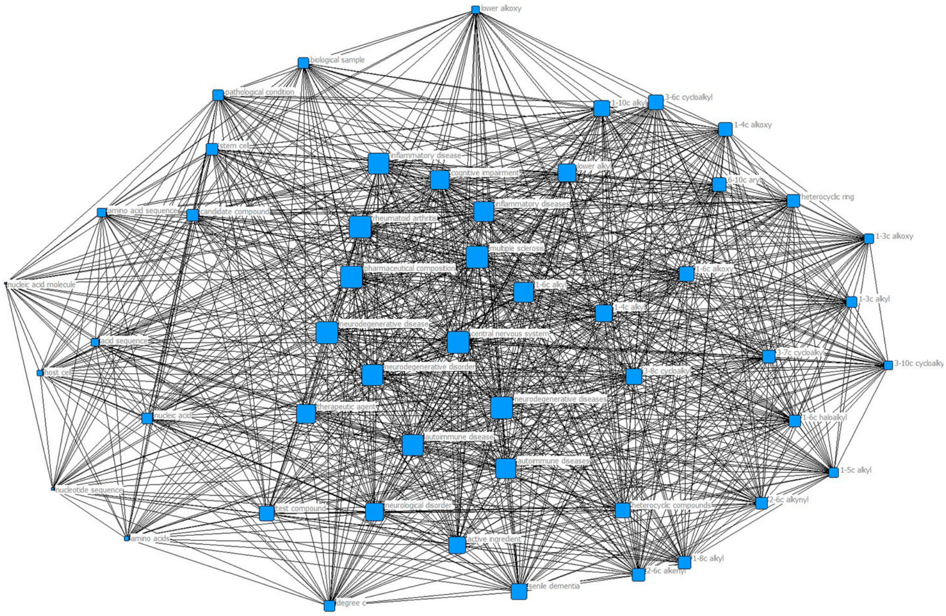

Figure 6. The co-occurrence analysis of main subjects based on top 50 keywords.

the nodes express institutions and inventors. The node's size refers to the number of publications; the bigger node presents a greater number of publications. The lines between nodes indicate the collaboration between the nodes; the thicker line presents more collaboration in publications. Moreover, the colour reveals the Sequence, for example, the red indicates the first Sequence, the green indicates the second Sequence, and the yellow indicates the third and follow-up Sequence. For instance, the institution numbers '123:102:452', means that during this period, the institution owns 677 patents in total, out of which, the institution applicants successfully 123 patents as the first patentee, 102 patents as the second patentee, and 452 patents as the other Sequence patentee.

This paper uses degree centrality in social network analysis to describe cooperation enthusiasm. Degree centrality is an index that measures the significance of the individual actors' position in the network, assessing the degree of the enterprise's network hub (Burt 2009), and the degree of resource acquisition and control (Wasserman and Faust 1994). The higher the degree centrality of network, the higher the enthusiasm for cooperation; the position of the actor is more at the core of the network. The lower value of degree of centrality means that the company is on the edge of the cooperation network.

4.5.1. Institution collaboration network

Figure 7 shows the cooperation network between the one hundred and fifty institutions in this field. The cooperation network between one hundred and forty-one institutions form a sizeable connected network, and the overall connections are tight – only nine institutions are not in the connected network (two institutions have cooperation between each other; the other seven institutions do not cooperate with the top one hundred and fifty institutions). The large connected network can be decomposed into four core networks which respectively are centred on HOFFMANN LA ROCHE & CO AG F (degree centrality is 0.0872), ASTRAZENECA AB (degree centrality is 0.0537), MERCK SHARP & DOHME CORP (degree centrality is 0.0671), PFIZER INC (degree centrality is 0.1141), and two edge networks which respectively are centred on UNIV CALIFORNIA (degree centrality is 0.1074) and INCYTE GENOMICS (degree centrality is 0.0537). The degree centrality of PFIZER INC is most significant, the enthusiasm for cooperation is highest. The total number of HOFFMANN LA

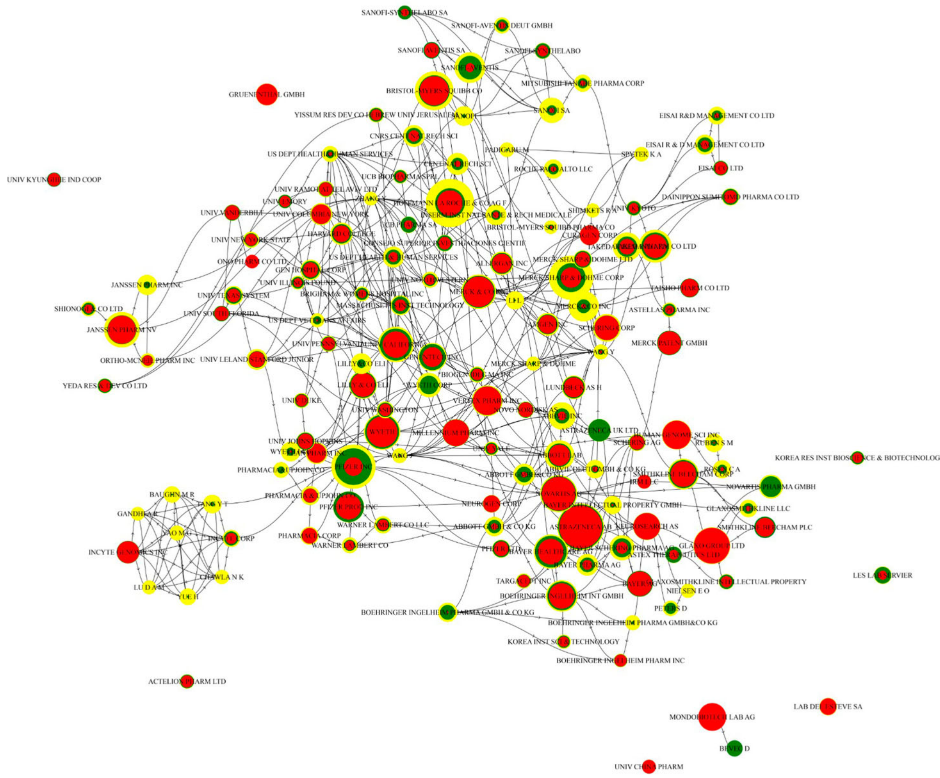

Figure 7. The collaboration network of institution on Alzheimer's disease.

ROCHE & CO AG F is 776 with first in ranking, whose turnnumber is 409;67;300, followed by ASTRAZENECA AB、MERCK SHARP & DOHME CORP、PFIZER INC, whose turnnumber respectively are 665;32; 32,300; 125; 226,272; 266; 111. Although HOFFMANN LA ROCHE & CO AG F has the largest number, the number of HOFFMANN LA ROCHE & CO AG F is the first author who has half fewer patents than ASTRAZENECA AB. Therefore, the contribution of ASTRAZENECA AB is not lower than that of HOFFMANN LA ROCHE & CO AG F in this field.

However, it is worth noticing that there is no much direct cooperation between top ten institutions, though only limited cooperation. For example, ASTRAZENECA AB mainly cooperates with NOVARTIS AG, and HOFFMANN LA ROCHE & CO AG F mainly cooperates with MERCK SHARP & DOHME. The top 150 institutions are mainly from US, European, and Japanese pharmaceutical companies and universities, while China has only one institution TAISHO TAISHO PHARM CO LTD (TAIS-C) ranking 66th and just cooperated with top 150 institutions MERCK & CO INC (MERI-C).

4.5.2. Inventor collaboration network

Figure 8 shows the inventor collaboration map. we can see that there are three clusters. The bigger cluster has the most productive and connected researchers and focused on WANG Y (degree centrality is 0.5906) with the largest number of patents; the turnnumber is 80;56;234. The other two clusters are centred on BACHER G (degree centrality is 0.0403), including seven inventors and PETERS D (degree centrality is 0.0201), including four inventors. The three clusters have no line to connect between each other, meaning that there is no intersection or cooperation. BACHER G ranks 6th concerning the total number of patents (261); however, considering the number of patents as the first sequence (258), BACHER G ranks first and is more than three times of WANG Y. BACHER G and BEVEC D are from the same institution MONDOBIOTECH LAB AG, GOLZ S, while BRUEGGEMEIER U

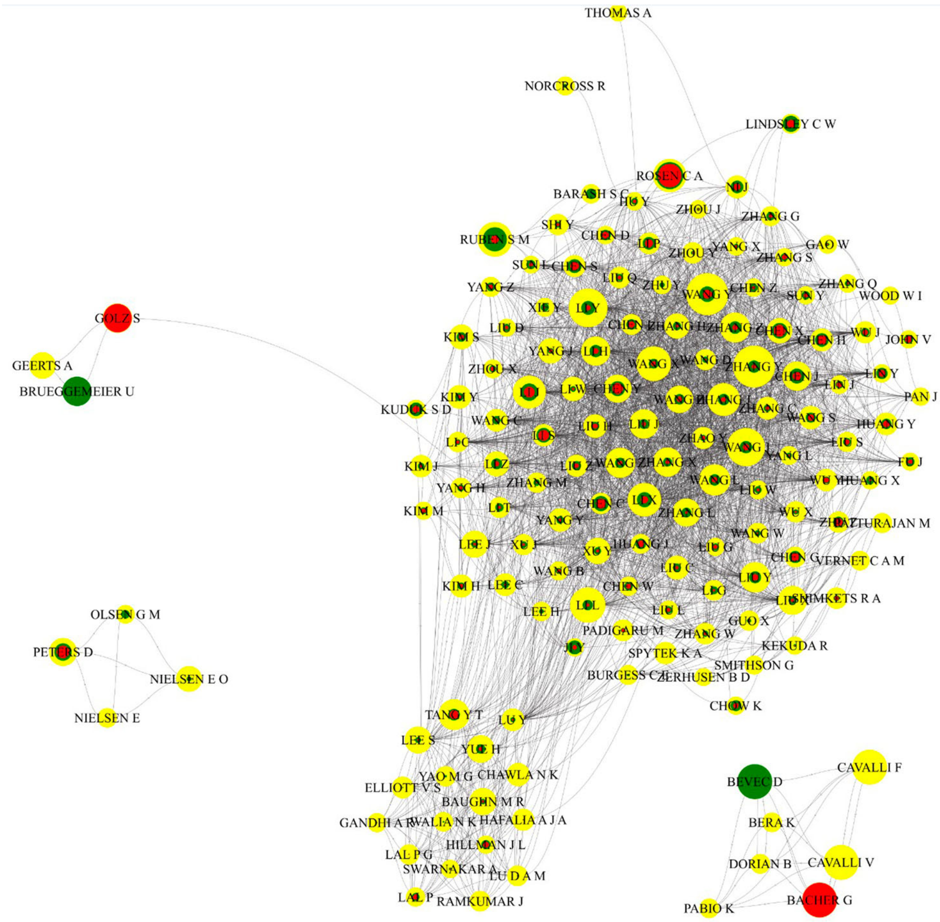

Figure 8. The collaboration network of inventor on Alzheimer's disease.

are from the same institution BAYER HEALTHCARE AG. It indicates that MONDOBIOTECH LAB AG and BAYER HEALTHCARE AG have strong research team.

5. Conclusion

This paper presents a systematic empirical analysis to (1) understand and monitor the technological development trends, (2) analyses the cooperation network more comprehensively and objectively, and (3) aims to facilitate more support for cooperation innovation. The findings of empirical research highlighted the following main concluding points and suggestions as follow:

(1) Inadequate cooperation between developed and developing countries.

It was observed that only two developing countries, China and India, are among the top ten patent-granting countries in Alzheimer's disease. This imbalance between developed and developing counties was further explored from the perspective of top ten patentees (institutions). This gap still evident in the list of top ten institutions, i.e. all belong to developed countries. Hence, systematic coordination is required from developed countries to stimulate support and collaboration with developing countries further.

(2) The technologies are concentrated on pharmaceutical activities in Alzheimer's disease, expanding interdisciplinary research and seeking new technological opportunities.

The technologies in Alzheimer's disease are focused on A61K, A61P, and C07D, totally account for 60%, which are related to preparations for medical, dental, and specific therapeutic activity. In addition, pharmaceutical ingredients such as 2-6c alkenyl, amino acid sequence are currently the hot subjects of research; however, traditional research subjects such as 1-6c alkyl are still the core subjects. To dig further and seek new technological opportunities, the institutions should expand interdisciplinary research and development in the next step.

(3) The current cooperation has great limitations, encourage more international cooperation and strong alliance.

The institutions tend to choose institutions that are easily accessible to cooperation, and there is no many direct cooperation between top ten institutions, yet only limited cooperation. Furthermore, the inventors are mostly from the same institutions. The current cooperation from the perspective of institutions and inventors has significant limitations; the institutions and inventors should carry out more international cooperation and encourage strong alliances to achieve more technological breakthroughs.

Acknowledgments

This work was supported by the General Program of National Natural Science Foundation of China under (Grant Nos.71774012, 71673024) and the strategic research project of the Development Planning Bureau of the Chinese Academy of Sciences (Grant No. GHJ-ZLZX-2019-42). The findings and observations in this paper are those of the authors and do not necessarily reflect the views of the supporters.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Rongrong Li is a doctor from the School of Economics and Management Beijing Institute of Technology. In addition, she is also a teacher at School of Economics & Management, China University of Petroleum (East China). Her specialty is e-commerce, technology assessment and data mining. Her current research is focused on e-commerce, text mining and forecasting innovation pathways.

Xuefeng Wang is professor at the School of Management and Economics, Beijing Institute of Technology, China. His specialty is technology innovation management, data mining and science and technology evaluation. His current research emphasises measuring, mapping and forecasting innovation pathways.

Yuqin Liu is professor at Beijing Green Printing and Packaging Industrial Technology Research Institute, Beijing Institue of Graphic Communication. His current research emphasises text mining, technology innovation assessment.

Shuo Zhang is a doctor from the School of Economics and Management Beijing Institute of Technology. Her current research is data mining and forecasting innovation.

Omer Hanif is a doctor from the School of Economics and Management Beijing Institute of Technology. His specialty is knowledge management and text mining.

References

Agrawal, Rakesh, Christos Faloutsos, and Arun Swami. 1993. "*Efficient Similarity Search in Sequence Databases*." Paper presented at the Foundations of data Organization and Algorithms, Berlin, Heidelberg.

Alzheimer's Association. 2018. "2018 Alzheimer's Disease Facts and Figures." *Alzheimer's & Dementia* 14 (3): 367–429.

Alzheimer's Disease International. 2019. "World Alzheimer Report 2018."

"Alzheimer's Medicines: Setbacks and Stepping Stones." In. 2018. Pharmaceutical Research and Manufacturers of America.

Anandarajan, Murugan, Chelsey Hill, and Thomas Nolan. 2019. "Cluster Analysis: Modeling Groups in Text." In *Practical Text Analytics*, 93–115. Cham: Springer.

Ansari, Mehtab Alam, Sumeer Gul, and Mohammad Yaseen. 2006. "Alzheimer's Disease: A Bibliometric Study."

Asghar, Ikram, Shuang Cang, and Hongnian Yu. 2017. "Assistive Technology for People with Dementia: an Overview and Bibliometric Study." *Health Information & Libraries Journal* 34 (1): 5–19.

Braun, T., W. Glänzel, and H. Grupp. 1995. "The Scientometric Weight of 50 Nations in 27 Science Areas, 1989–1993. Part I. All Fields Combined, Mathematics, Engineering, Chemistry and Physics." *Scientometrics* 33 (33): 263–293. doi:10.1007/BF02017332.

Bugge, Markus M., Teis Hansen, and Antje Klitkou. 2016. "What Is the Bioeconomy? A Review of the Literature." *Sustainability* 8 (7). doi:10.3390/su8070691.

Burt, Ronald S. 2009. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard university press.

Carrillo, Maria C, Kaj Blennow, Holly Soares, Piotr Lewczuk, Niklas Mattsson, Pankaj Oberoi, Robert Umek, Manu Vandijck, Salvatore Salamone, and Tobias Bittner. 2013. "Global Standardization Measurement of Cerebral Spinal Fluid for Alzheimer's Disease: an Update From the Alzheimer's Association Global Biomarkers Consortium." *Alzheimer's & Dementia* 9 (2): 137–140.

Chen, Huaqi, Yuehua Wan, Shuian Jiang, and Yanxia Cheng. 2014. "Alzheimer's Disease Research in the Future: Bibliometric Analysis of Cholinesterase Inhibitors from 1993 to 2012." *Scientometrics* 98 (3): 1865–1877.

Cross, Rob, Stephen P. Borgatti, and Andrew Parker. 2007. "Making Invisible Work Visible: Using Social Network Analysis To Support Strategic Collaboration." *California Management Review* 44 (2): 25–46. doi:10.2307/41166121.

Davidse, R. J., and A. F. J. Van Raan. 1997. "Out of Particles: Impact of CERN, DESY and SLAC Research to Fields Other Than Physics." *Scientometrics* 40 (40): 171–193. doi:10.1007/BF02457436.

de Oliveira, Meire Ramalho, Angela Emi Yanai, Diogo Soares Moreira, Cláudia Daniele de Souza, and Carlos Eduardo Gomes de Castro. 2018. "*Internet of Things (IoT): Technological Indicators from Patent Analysis*." Paper presented at the International Joint conference on Industrial Engineering and Operations management.

Dong, Rui, Hong Wang, Jishi Ye, Mingshan Wang, and Yanlin Bi. 2019. "Publication Trends for Alzheimer's Disease Worldwide and in China: A 30-Year Bibliometric Analysis." *Frontiers in Human Neuroscience* 13: 259.

Emmerich, Christiane. 2009. "Comparing First Level Patent Data with Value-Added Patent Information: A Case Study in the Pharmaceutical Field." *World Patent Information* 31 (2): 117–122. doi:10.1016/j.wpi.2008.06.003.

Gaitani, N., C. Lehmann, M. Santamouris, G. Mihalakakou, and P. Patargias. 2010. "Using Principal Component and Cluster Analysis in the Heating Evaluation of the School Building Sector." *Applied Energy* 87 (6): 2079–2086.

Godin, Benoît. 2006. "On the Origins of Bibliometrics." *Scientometrics* 68 (1): 109–133. doi:10.1007/s11192-006-0086-0.

Haunschild, Robin, Lutz Bornmann, and Werner Marx. 2016. "Climate Change Research in View of Bibliometrics." *PLoS One*, doi:10.1371/journal.pone.0160393. eCollection 2016.

Jain, Anil K., and Richard C. Dubes. 1988. *Algorithms for Clustering Data*. Vol. 6. Englewood Cliffs, NJ: Prentice Hall.

Jia, Jianping, Fen Wang, Cuibai Wei, Aihong Zhou, Xiangfei Jia, Fang Li, Muni Tang, Lan Chu, Youlong Zhou, and Chunkui Zhou. 2014. "The Prevalence of Dementia in Urban and Rural Areas of China." *Alzheimer's & Dementia* 10 (1): 1–9.

Jia, Jianping, Cuibai Wei, Shuoqi Chen, Fangyu Li, Yi Tang, Wei Qin, Lina Zhao, Hongmei Jin, Hui Xu, and Fen Wang. 2018. "The Cost of Alzheimer's Disease in China and re-Estimation of Costs Worldwide." *Alzheimer's & Dementia* 14 (4): 483–491.

Jones-Davis, Dorothy M, and Neil Buckholtz. 2015. "The Impact of the Alzheimer's Disease Neuroimaging Initiative 2: What Role Do Public-Private Partnerships Have in Pushing the Boundaries of Clinical and Basic Science Research on Alzheimer's Disease?" *Alzheimer's & Dementia* 11 (7): 860–864.

Lo Re, Michele. 2018. "La Network Analysis: proposta di un framework concettuale per le applicazioni economiche." *L'industria* 39 (4): 643–674.

Lo Re, Michele, and Eleonora Veglianti. 2017. "Among the Methodological Perspectives of Structural Analysis and the Inevitable Centrality of Economic-Social Facts." *L'industria* 38 (2): 241–268.

Madani, Farshad, Tugrul Daim, and Calvin Weng. 2017. "'Smart Building'technology Network Analysis: Applying Core–Periphery Structure Analysis." *International Journal of Management Science and Engineering Management* 12 (1): 1–11.

Medrano-Marqués, Nicolás Juris, and Bonifacio Martín-del-Brío. 1999. "*Topology Preservation in SOFM: An Euclidean Versus Manhattan Distance Comparison*." Paper presented at the international work-conference on Artificial Neural Networks.

Monarca, Umberto, Ernesto Cassetta, Michele Lo Re, and Linda Meleo. 2019. "A Network Analysis of the Intersectoral Linkages Between Manufacturing and Other Industries in China and Italy." *Global Journal of Emerging Market Economies* 11 (1-2): 80–97.

Newman, M. E. 2001. "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences of the United States of America* 98 (2): 404–409. doi:10.1073/pnas.98.2.404.

Nicolaisen, Jeppe. 2010. "Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics." *Journal of the American Society for Information Science & Technology* 61 (1): 205–207. doi:10.1002/asi.21181.

Oppenheim, Charles. 1981. "The Past, Present and Future of the Patents Services of Derwent Publications Ltd." *Science & Technology Libraries* 2 (2): 23–31.

Pettersson, Martin, Antonia F Stepan, Gregory W Kauffman, and Douglas S Johnson. 2013. "Novel γ-Secretase Modulators for the Treatment of Alzheimer's Disease: A Review Focusing on Patents from 2010 to 2012." *Expert Opinion on Therapeutic Patents* 23 (10): 1349–1366.

Raan, Anthony F. J. Van. 2005. "Fatal Attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods." *Scientometrics* 62 (1): 133–143. doi:10.1007/s11192-005-0008-6.

Reuther, Patrick. 2006. "Personal Name Matching: New Test Collections and a Social Network based Approach." 06-01: Available online: Accessed on January 2006. https://www.researchgate.net/publication/220046342_Personal_Name_Matching_New_Test_Collections_and_a_Social_Network_based_Approach.

Sampaio, Priscila Gonçalves Vasconcelos, Mario Orestes Aguirre González, Rafael Monteiro de Vasconcelos, Marllen Aylla Teixeira dos Santos, José Carlos de Toledo, and Jonathan Paulo Pinheiro Pereira. 2018. "Photovoltaic Technologies: Mapping From Patent Analysis." *Renewable and Sustainable Energy Reviews* 93: 215–224. doi:10.1016/j.rser.2018.05.033.

Selkoe, Dennis J. 2019. "Early Network Dysfunction in Alzheimer's Disease." *Science* 365 (6453): 540–541.

Shotorbani, Peyman Yazdizadeh, Farhad Ameri, Boonserm Kulvatunyou, and Nenad Ivezic. 2016. "*A Hybrid Method for Manufacturing Text Mining Based on Document Clustering and Topic Modeling Techniques*." Paper presented at the IFIP International Conference on Advances in Production management Systems.

Song, Min, Go Eun Heo, and Dahee Lee. 2015. "Identifying the Landscape of Alzheimer's Disease Research with Network and Content Analysis." *Scientometrics* 102 (1): 905–927.

Sorensen, Aaron A, Andrew Seary, and Kenneth Riopelle. 2010. "Alzheimer's Disease Research: A COIN Study Using Co-Authorship Network Analytics." *Procedia-Social and Behavioral Sciences* 2 (4): 6582–6586.

Theander, Sten S, and Lars Gustafson. 2013. "Publications on Dementia in Medline 1974–2009: A Quantitative Bibliometric Study." *International Journal of Geriatric Psychiatry* 28 (5): 471–478.

Veitch, Dallas P, Michael W Weiner, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack Jr, William Jagust, and John C Morris. 2019. "Understanding Disease Progression and Improving Alzheimer's Disease Clinical Trials: Recent Highlights From the Alzheimer's Disease Neuroimaging Initiative." *Alzheimer's & Dementia* 15 (1): 106–152.

Wang, Qiang, and Yong Chen. 2010. "Status and Outlook of China's Free-Carbon Electricity." *Renewable and Sustainable Energy Reviews* 14 (3): 1014–1025. doi:10.6/j.rser.2009.10.012.

Wang, Qiang, and Rongrong Li. 2016. "Natural gas From Shale Formation: A Research Profile." *Renewable and Sustainable Energy Reviews* 57: 1–6. doi:10.1016/j.rser.2015.12.093.

Wang, Xuefeng, Rongrong Li, Shiming Ren, Donghua Zhu, Meng Huang, and Pengjun Qiu. 2014. "Collaboration Network and Pattern Analysis: Case Study of Dye-Sensitized Solar Cells." *Scientometrics* 98 (3): 1745–1762. doi:10.007/s11192-013-1180-8.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Vol. 8. Cambridge: Cambridge University Press.

World Health Organization. 2018. "Global Health Estimates 2016." http://www.who.int/healthinfo/global_burden_disease/en/.

Xu, Jiachen, Xiangjun Kong, Lan Qiu, Xiaomei Geng, Yuanjia Hu, and Yitao Wang. 2014. "Research and Development of Anti-Alzheimer's Drugs: An Analysis Based on Technology Flows Measured by Patent Citations." *Expert Opinion on Therapeutic Patents* 24 (7): 791–800.