COLLNET Journal of Scientometrics and Information Management
ISSN : 0973 – 7766 (Print) 2168 – 930X (Online)
Vol. 18(2) December 2024, pp. 137–154
DOI : 10.47974/CJSIM-2023-0021

Research on exoplanets : A bibliometric study of a new research frontier

Xiaojun Hu

Zhixian Ye

Fangjie Xi

Xian Li

Ronald Rousseau

Xiaojun Hu @
Zhejiang University School of Medicine
Hangzhou 310058
China

Zhixian Ye †
School of Aeronautics and Astronautics
Zhejiang University
Hangzhou 310027
China

Fangjie Xi §
Zhejiang University School of Medicine
Hangzhou 310058
China

Xian Li ‡
College of Computer and Information Science
Southwest University
Chongqing 400037
China

Ronald Rousseau *
Faculty of Social Sciences
University of Antwerp
Antwerp
Belgium

and

Centre for R&D Monitoring (ECOOM) and Department MSI
KU Leuven
Leuven
Belgium

E-mail
@ xjhu@zju.edu.cn
† ye_zhixian@zju.edu.cn
§ fangjiexi@zju.edu.cn
‡ lx2009yet@126.com
* ronald.rousseau@uantwerpen.be; ronald.rousseau@kuleuven.be

* (Corresponding Author)

We conduct a systematic analysis of scientific publications in the new research frontier dealing with exoplanets. Our main methodology is Latent Dirichlet allocation (LDA) used to identify latent themes diachronically. In this way, we identify top themes, their trends, and leading journals. Moreover, co-occurrence networks of country-country collaboration and institution-institution collaboration are drawn. In this way, we determine the scientific frontiers of exoplanet studies, and themes that received the most citations were found. The size of research teams, as shown by co-authors per article, has been growing over time. Going beyond publications, we find that few articles on exoplanets have been cited by patents and also few patents result from exoplanet research.

Keywords: *Exoplanet research, LDA analysis, Theme evolution, Systematic analysis, Collaborative network, h-indices for topics.*

1. Introduction

1.1 *Earth and exoplanets*

Ever since humans realized they were not at the center of the universe, philosophers and scientists have been wondering if there exists life outside the Earth. Initial investigations focused on the Moon and Mars as possible places for the existence of life. [1] Once it was realized that the Galaxy (Milky Way) is just one among many galaxies (currently estimated at two trillion [2]), the odds became highly supportive of the existence of extraterrestrial life.

Considering life as we know it on Earth, one way to move on, besides in situ visits to the near planets, is to look for exoplanets, i.e., celestial objects orbiting stars other than the Sun. Among these, the most interest lies in habitable planets, those not too far away from their star (otherwise it would be too cold for life to exist) or not too close (because then it would be too hot). These planets are often referred to as belonging to the habitable zone, or with a term from a British fairy tale, the Goldilocks zone. In fact, in one version of this tale, a young girl named Goldilocks tastes three different bowls of porridge and finds that she prefers porridge that is neither too hot nor too cold, but has just the right temperature, a principle that also applies to habitable planets. One must realize that an absolute requirement for life is an energy source, and the notion of planetary habitability implies that many other geophysical, geochemical, and astrophysical conditions must be met before an astronomical body can support life. In this context, one uses the term "biosignature". This refers to any characteristic, substance, or feature that provides evidence of past or present life. In recent years, several dozen exoplanets have been identified within the habitable zone of their parent star, where surface temperatures may allow liquid water to be present. Scientists have now discovered that there are more planets than stars in our Galaxy.

This article presents a bibliometric study of the literature dealing with exoplanets. More concretely we will focus on the following topics:

(1) The evolution of publications on exoplanets from 1996 till 2022.
(2) The scientific frontiers on exoplanets as determined by LDA analysis [3]. Which themes received the most citations?
(3) For the three periods: 1996-2009, 2010-2015, and 2016-2022 we study the number of authors per article, the mean number of citations, and the international and national (within a country) collaboration.

Before embarking on the bibliometrics, we provide a brief description of exoplanets, how they are detected, and scientists' plans for studying exoplanets.

1.2 *What are exoplanets, and what is the importance of their study?*

In this study, we will not consider extraterrestrial life [4], leading to exobiology or xenobiology, such as studied e.g., in [5]. Instead, we focus on the study of exoplanets, regardless of their surface temperature or even composition as many exoplanets are not solid, like Earth, but are gaseous, i.e., mainly composed of helium and hydrogen (such as Jupiter and Saturn), or composed of 'ices', like ammonia and methane (such as Uranus and Neptune) [2].

The discovery of the first confirmed exoplanet orbiting a solar-type star was announced in 1995 and was credited to Michel Mayor and Didier Queloz from the Geneva Observatory in Switzerland. This planet was a giant planet like Jupiter but in a very close orbit around its parent star [2]. For this discovery, these two scientists received the Nobel Prize in Physics in 2019.

In [6], the Committee on Planetary Science and Astrobiology describes its plans for the next ten years. In this document, the authors formulate twelve priority science questions. While questions 9-11 deal with the search for life in our planetary system outside Earth, Question 12 specifically addresses exoplanets. Here the main question is: What do our planetary system and its circumplanetary systems of satellites and rings reveal about exoplanetary systems, and what can circumstellar disks and exoplanetary systems teach us about the solar system? This research question is subdivided into 11 sub-questions plus some supportive activities.

The authors of this report point out that over the past decades, planetary science has been revolutionized with the discovery of over 4,000 confirmed extrasolar planets of a variety of sizes and masses. These discoveries were enabled in large part by the Kepler mission and the Transiting Exoplanet Survey Satellite (TESS).

Nowadays, exoplanets offer a huge sample to provide statistical information about the factors that govern planetary evolution. An overarching goal for the next decade is to characterize the properties of the interiors and surfaces of solid exoplanets, leading to the development of "exogeoscience" [6].

In the next step, the authors of this report point out that the dynamical properties of exoplanets may have important implications for the study of potential habitability. Understanding the development of other planetary systems may have important implications for our understanding of planetary formation, evolution, and habitability.

Until now not a single planet similar to Earth, an Earth 2.0, has been found: a rocky planet in the habitable zone orbiting a star similar to the Sun. Not surprisingly, many countries, and in particular upcoming countries such as China and India, hope to be first. With that purpose, Mainland Chinese scientists plan to launch a rocket equipped with seven telescopes before the year 2027 [7].

1.3 *Which methods are used to detect exoplanets [2]?*

The method used to find the first exoplanet orbiting a Sun-like object was the Doppler effect [2]. We all know that when an object moves away from us, the sound waves are somewhat stretched, resulting in a lower pitch, while when this object moves in the direction of the observer the sound waves are compressed, resulting in a higher pitch. This is called the Doppler effect. The same effect applies to other waves, such as light waves, when movement is involved. When a source of light moves away from the observer then the light is a little bluer, while when the light source moves toward the observer, the light is a little redder. The amount of color shift depends on how fast the source is moving relative to the speed of propagation of light. It turned out that existing telescopes equipped with appropriate spectroscopes could find exoplanets.

A second general approach to detecting exoplanets is the transit technique [2]. If a planet moves between the observer and its star, the apparent brightness of the star is slightly reduced. Measuring these reductions in brightness with a ground-based telescope is nearly impossible, but has been done. Yet, this method can more easily be applied from space. The first such instrument was the French CoRoT satellite, launched in 2006 [2]. After that, the famous NASA Kepler mission started. This satellite, launched in 2009, has confirmed the existence of more than 3000 exoplanets.

We note that the science of exoplanets begins with their detection (by so-called planet hunters), but that is just a start. Scientists moreover want to know the chemical and physical characteristics of these planets, how they originated, how they might develop, and how they fit into the greater astronomical picture.

2. Data and methods

In this investigation, we apply the following approach.

2.1 *Search in the Web of Science (WoS) complete Core Collection.*

We started from an elementary query TS=exoplanet* which yielded 13,053 records. These data were updated on December 15, 2022, while the data used for the investigation based on LDA date from June 2022.

To make our investigation more precise, we restricted the records found by a simple TS=exoplanet* search (in June 2022), to those of article type. In this way, we focus on new results (by excluding reviews) and remove news items (as often published in journals such as *Nature* and *Science*). A total of 9,200 records were retained, referred to as the *Exoplanet Core Set*. This is the set investigated further on.

In order to observe the relationship between academic performance and technological invention in this area, we also conducted investigations using two well-known databases of patent records: the Lens (www.lens.org) and the Derwent Innovation Index (http://www-webofscience-com-s.webvpn.zju.edu.cn:8001/wos/diidw/basic-search).

2.2 *Academic performance in this area*

Based on the *Exoplanet Core Set* we study academic performance by co-author networks, international collaboration, and collaboration between academia and industry, and this over different periods.

2.3 *Core journals in exoplanet research using two aspects*

To find core journals focusing on exoplanet research, we analyze the source titles of records from the WoS from two aspects, the number of publications and their citations (only those publications in the *Exoplanet Set*, but including citations from all sources). For each journal, we rank all articles by the number of received citations in descending order and then calculate their h-index.

2.4 *LDA modeling for hot themes over time*

In order to provide a systematic landscape of the domain focusing on exoplanets, we employ Latent Dirichlet Allocation (LDA), a classical unsupervised learning algorithm based on a three-layer Bayesian probabilistic model, namely, the relations among the structure of words, topics, and text. For technical details and some applications, we refer the reader to [3], [8], [9]. This technique is applied to figure out the structure of themes about exoplanet publications included in the WoS and their evolution. For this, we used the Python library Gensim (https://pypi.org/project/gensim/). To be precise, we note that the titles and abstracts of the records are included in the text set to be analyzed.

Given a text set, it can be represented as its corresponding topic set and each topic as a particular polynomial distribution of words. Here, the appropriate number of topics for our dataset is determined by comparing the inter-topic similarity with LDAvis [10], a web-based interactive visualization tool for qualitative assessment of topic models, that provides an intuitive, yet profound, inspection of topic-term relationships in an LDA model [11].

3. Results and discussion

3.1 *Basic results*

The 13,053 records resulting from the search TS= exoplanet* are mostly of article type (71.5%) or contributions to conference proceedings (24.4%); 2.1% are review papers.

The records found belong for 83.2% to the research area Astronomy Astrophysics, 16.7% are in Optics, and another 11.6% are classified in Instruments. Further, we found 6.1% in Engineering, 4.7% in Physics, and 3.0% in Geology.

Leading countries/regions are: the USA with 57,2% of all records, followed by France (19.5%), England (19.0%), Germany (18.2%), and Switzerland (11.3%). China is only 15th, participating in 4.9% of all records. India is 29th in this ranking. Almost all of these publications, namely 99.86%, are written in English.

We note that very few articles in this area have been cited by patents (only 75 times in total, according to the database Lens), pointing to the fact that most of this research has no direct applications, which is also reflected by the fact that few patent records (only 22 in total) are obtained from the database Derwent Innovation Index.

3.2 *Academic research on the theme of exoplanets*

Table 1 provides an overview of exoplanet research from 1996 to 2022 in three periods. Although the earliest article about exoplanets was reported in 1996, the theme did not become hot until the year 2010 and has since obtained increased attention. Particularly, the size and amount of research teams in this area have been rapidly increasing as shown by the number of authors per article which went up to 11.3 during the period 2016-2022. Similar trends are observed for the number of institutions participating in this field, including the rate of international collaboration as well as academy-industry collaboration. This led us to include co-author networks in our investigations. We note that very few articles in this area

Table 1

An overview and trends of exoplanet publications in the *Exoplanet Core set*

	1996-2009	2010-2015	2016-2022
Number of publications	918	2,706	5,576
% within the Exoplanet Core Set	9.98	29.41	60.61
Number of received citations	58,791	140,968	87,322
Citations per publication	64.04	52.095	15.66
Sum of all authors (not necessarily different)	6,052	23,880	61,480
Authors per publication	6.59	8.82	11.03
Sum of institutions (not necessarily different)	2725	9597	24683
Institutions per publication	2.97	3.55	4.43
Sum of countries (not necessarily different)	1601	4672	10163
Countries per publication	1.74	1.73	1.82
International collaboration (in %)	35.62	36.73	40.33
Academy-industry collaboration (in %)	8.61	13.05	15.05

have been cited by patents (only 75 times in total according to the database Lens), pointing to the fact that most of this research has no direct applications, which is also reflected by the fact that few patent records (only 22 in total) are obtained from the database Derwent Innovation Index.

3.3 *Collaborative networks of exoplanet research*

Figures 1-2 are co-occurrence networks of country-country collaboration and institution-institution collaboration. The top 20 nodes (see Table 2) with the largest degree centrality in the networks are shown in the center position of figures, corresponding to the leading countries and institutions in exoplanet research. For more detailed information we refer the reader to Table 2.

The different colors in these figures represent different groups in the collaborative networks based on modularity analysis, the size of the nodes represents the degree centrality, and the edge thickness indicates link strengths. For example, the blue color in Figure 1 refers to the network of countries led by the USA, where the closest cooperative countries/regions are England, France, Italy, Canada, Australia, Japan, P.R. China, Switzerland, Chile, Portugal; the other group (marked in yellow) has Germany as its center. It includes the Netherlands, Spain, South Korea, Poland, and Austria, among others.

There are four groups shown in Figure 2. NASA is at the center of the biggest group (marked in red), surrounded by the University of California Santa Cruz, the Space Telescope Science Institute, Princeton University, the University of Texas Austin, MIT, the Harvard–Smithsonian Center for Astrophysics, the University of California at Berkeley and others. The second one (in yellow) consists of the Instituto de Astrofísica de Canarias, Leiden University, the University of Warwick, the University of Bern, the University of Porto, the

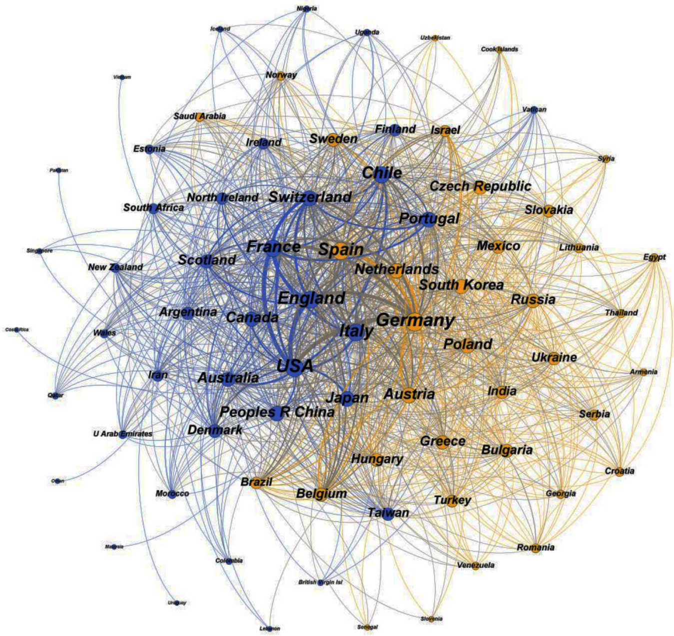

Figure 1

Top countries in the country collaboration network

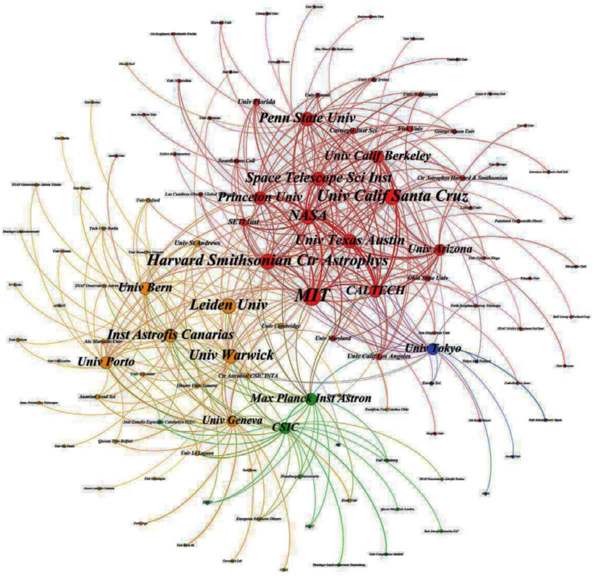

Figure 2

Top institutions in the institutional collaboration network

Table 2

Top countries and institutions in the collaborative network of exoplanet research

Ranking	Country/region	Degree	Institution	Degree
1	Germany	64	NASA	975
2	USA	63	CALTECH	902
3	Italy	59	MIT	713
4	Spain	59	Max Planck Inst Astron	655
5	England	59	Space Telescope Sci Inst	623
6	France	57	Univ Arizona	618
7	Chile	56	Univ Geneva	607
8	Switzerland	51	Inst Astrofis Canarias	599
9	Portugal	50	Princeton Univ	592
10	Austria	50	Univ Calif Berkeley	576
11	Japan	49	Penn State Univ	559
12	Peoples R China	48	Univ Texas Austin	551
13	Poland	47	Harvard Smithsonian Ctr Astrophys	541
14	Netherlands	47	Univ Porto	517
15	Canada	47	Univ Bern	513
16	South Korea	46	Univ Warwick	508
17	Australia	46	Leiden Univ	503
18	Russia	44	CSIC (Spain)	493
19	Scotland	43	Univ Calif Santa Cruz	487
20	Czech Republic	43	Univ Tokyo	477

University of Geneva, and other less important institutions. The group marked in green is connected through the Max-Planck-Institut für Astronomie and CSIC (Spain). Finally, the University of Tokyo leads the smallest group connecting to institutions not displayed in the figure.

One may note that Germany is the leading country in country collaboration but that the top three collaborating institutes are American ones. This observation suggests that the collaborative network of the US in this field is mainly based on domestic cooperation.

3.4 *The h-indices of core journals focusing on exoplanet research*

Contrasting to the traditional way of presenting journals, Table 3 shows the core journals in exoplanet research (only using the *Exoplanet Set*, *as defined earlier*) measured by the number of publications per journal and their citations. For example, if there are 112 publications (in the *Exoplanet Set)* of a certain journal with each at least 112 citations, then this journal's h-index is 112. The top ones (according to the h-index) and related data are shown. Note that, except for the multidisciplinary journals *Nature*, *Science*, and *PNAS*, most of them are from the fields Astrophysics and Space Science.

Table 3

The core journals of exoplanets and their h-indices

Journals	h-index	Number of articles	Total number of citations
ASTROPHYSICAL JOURNAL	112	1533	60185
ASTRONOMY & ASTROPHYSICS	84	1473	39212
MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY	76	1305	28981
ASTROPHYSICAL JOURNAL LETTERS	67	495	17834
ASTRONOMICAL JOURNAL	64	1036	21050
NATURE	51	116	10109
PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC	43	182	8524
ASTROPHYSICAL JOURNAL SUPPLEMENT SERIES	39	117	8047
ASTROBIOLOGY	34	121	3279
NATURE ASTRONOMY	26	100	1980
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA (PNAS)	25	47	2116
ICARUS	24	66	1655
SCIENCE	23	60	5584
JOURNAL OF ASTRONOMICAL TELESCOPES INSTRUMENTS AND SYSTEMS	21	135	2640
SPACE SCIENCE REVIEWS	16	41	951
CELESTIAL MECHANICS & DYNAMICAL ASTRONOMY	16	46	838

3.5 *Themes related to exoplanet research and its evolution*

3.5.1 The structure of themes dealing with exoplanet research

We found only 918 publications about exoplanet research published during the period 1996-2009. LDA modeling found eight themes (for details, see Appendix Table 7). Keywords for themes reflect the following points that received attention in this period: “design, engineering specifications and technical development of telescopes and observatories”, “observations of exoplanets using e.g., photometric and transit methods, and detection of physical properties”, “new observations and discoveries of stellar companions”, “orbital evolution, and stability of planet-exoplanetary systems”, “study of earth-like habitable condition on planets and exoplanets and their stability”, “research on the disk environment of habitable earth-like planets and exoplanets around the zone of the galaxy and solar system”, “observation on the atmospheric temperature and chemical composition of the extrasolar system”, “study of physical properties of giant planets”.

Tables 4a and 4b show the results of LDA modeling for publications in the period 2010-2022. The most important 17 themes are shown in descending way of importance. To promote the readability of our article we have subdivided this table into two parts (Table 4a and Table 4b) relegating Table 4b (the less important themes) to the appendix. Moreover, the top 30 words for each theme are listed in the right-hand column. Compared to the themes in 1996-2009, we note the following new themes on exoplanet research: "adaptive optics optimization and wavefront control of coronagraphs for exoplanet high-contrast imaging" (theme 1), "first discovery, detection results and status of exoplanets" (theme 3), "statistical analysis of exoplanet search and observation data samples" (theme 8), "spectral analysis of exoplanet atmosphere" (theme 9), "analysis of the formation mechanism of various giant planets" (theme 10), "study on the magnetic field environment of stellar wind in interplanetary space "(theme 13), "photometric observations and kinematic analysis of dwarfs" (theme 14), "detection of cosmic life and prediction statistics of exoplanets" (theme 16), "observation and discovery of the structure of exoplanet debris disk" (theme 17). High-loading keywords and most representative documents reveal that theme 1 (the most popular theme in this period) is about "adaptive optics optimization and wavefront control of coronagraphs for exoplanet high-contrast imaging".

Table 4a

LDA results for publications 2010-2022 (8,282 documents), Part a. Themes 1 to 9.

	Labels	The top 30 words
Theme 1	Adaptive optics optimization and wavefront control of coronagraphs for exoplanet high-contrast imaging	high; image; contrast; coronagraph; imaging; performance; system; instrument; control; optic; phase; light; optical; direct; spectrograph; wavefront; adaptive; coronagraphic; telescope; field; design; mask; error; speckle; pupil; calibration; detection; mode; mirror; present
Theme 2	Status, design, engineering specifications, and technical development of telescopes and observatories	mission; telescope; space; design; science; instrument; study; exoplanet; exoplanets; base; large; concept; characterization; observatory; development; ground; infrared; provide; goal; paper; capability; optical; discuss; future; program; NASA; class; include; next; scientific
Theme 3	First discovery, detection results, and status of exoplanets	planet; mass; star; orbit; period; radius; planetary; day; host; circle; transit; Jupiter; system; plus; orbital; size; around; super; Neptune; low; density; discovery; hot; two; detect; small; discover; candidate; giant; radial
Theme 4	Observation and analysis of the physical properties of exoplanets such as photometric spectrum, ephemeris, transit state, transit time, orbital parameters, etc.	transit; light; curve; wasp; eclipse; observation; variation; photometric; time; photometry; hat; data; secondary; exoplanet; depth; find; spot; spitzer; analysis; observe; present; timing; fit; telescope; base; measurement; precision; parameter; obtain; ground

Contd...

Theme 5	Orbital dynamics research on interactions, tidal motion, common-orbit resonance, orbital evolution, and stability of exoplanet multi-star systems	system; planet; orbit; orbital; planetary; eccentricity; tidal; body; dynamical; spin; motion; evolution; inclination; migration; resonance; stability; show; period; two; effect; close; eccentric; interaction; time; circular; find; may; dissipation; study; case
Theme 6	Measurement methods for radial velocity and orbital parameters of exoplanets	method; velocity; radial; stellar; parameter; measurement; data; signal; analysis; star; astrometric; precision; detection; aim; estimate; obtain; noise; technique; determine; measure; error; derive; different; astrometry; context; accuracy; high; show; line; accurate
Theme 7	Research on the atmospheric environment of habitable Earth-like planets	earth; planet; habitable; like; terrestrial; zone; surface; exoplanets; life; water; atmosphere; habitability; atmospheric; may; world; potentially; rocky; CO_2; liquid; sun; exo; climate; planetary; ocean; carbon; possible; detect; oxygen; condition; Venus
Theme 8	Statistical analysis of exoplanet search and observation data samples	search; data; new; exoplanet; survey; candidate; exoplanets; interferometer; system; object; detection; extrasolar; present; year; target; now; field; identify; include; discovery; observation; project; follow; multi; provide; discover; observe; detect; sky; available
Theme 9	Spectral analysis of exoplanet atmosphere	spectrum; wavelength; spectral; band; infrared; resolution; observation; spectroscopy; high; near; atmosphere; feature; transmission; optical; cloud; exoplanet; absorption; range; similar; telescope; ratio; atmospheric; extrasolar; detection; spectra; space; cover; spectrograph; Hubble;

3.5.2 The evolution of themes over time

When studying the dynamic distribution of the 17 themes on exoplanets research shown in Tables 4a and 4b, referring to the period from 2010 to 2022 we observed three points of interest (data not included):

(1) the percentages of themes 1-2 per year are gradually decreasing over time (with some fluctuations);

(2) the percentage of themes 3-11 and 17 per year is relatively stable;

(3) the percentage for themes 12-16 is gradually, but only slightly, increasing per year.

This result indicates some trends but generally, interest has not changed a lot. On the one hand, the most popular themes, such as “adaptive optics optimization and wavefront control of coronagraphs for exoplanet high-contrast imaging”, “status, design, engineering specifications and technical development of telescopes and observatories”, have somewhat decreased in popularity, while on the other hand, themes such as “atmospheric escape and atmospheric motion analysis of exoplanets”, “study on the magnetic field environment of stellar wind in interplanetary space”, “photometric observations and kinematic analysis of dwarfs", "atmospheric temperature and circulation models for exoplanets", "detection of cosmic life and prediction statistics of exoplanets" received more attention over the years.

3.6 *The top themes in exoplanet research*

Figure 3 shows the mean number of citations received by each theme (calculated as the total number of citations received by publications in the theme's subset divided by the number of publications in this subset) in the period 2010-2022 (the top left part) and the yearly number of received citations in the citation window corresponding to their publication years, indicating the trend of interest in the theme. Theme 3 "first discovery, detection results, and status of exoplanets", theme 9 "spectral analysis of exoplanet atmosphere", theme 12 "atmospheric escape and atmospheric motion analysis of exoplanets", and theme 10 "analysis of the formation mechanism of various giant planets" are the four top themes on exoplanets, each with more than 30 citations per article and in an increasing trend of yearly citations received. Themes 1-2 are listed at the bottom (among the citation curves as well as by the mean number of yearly citations).

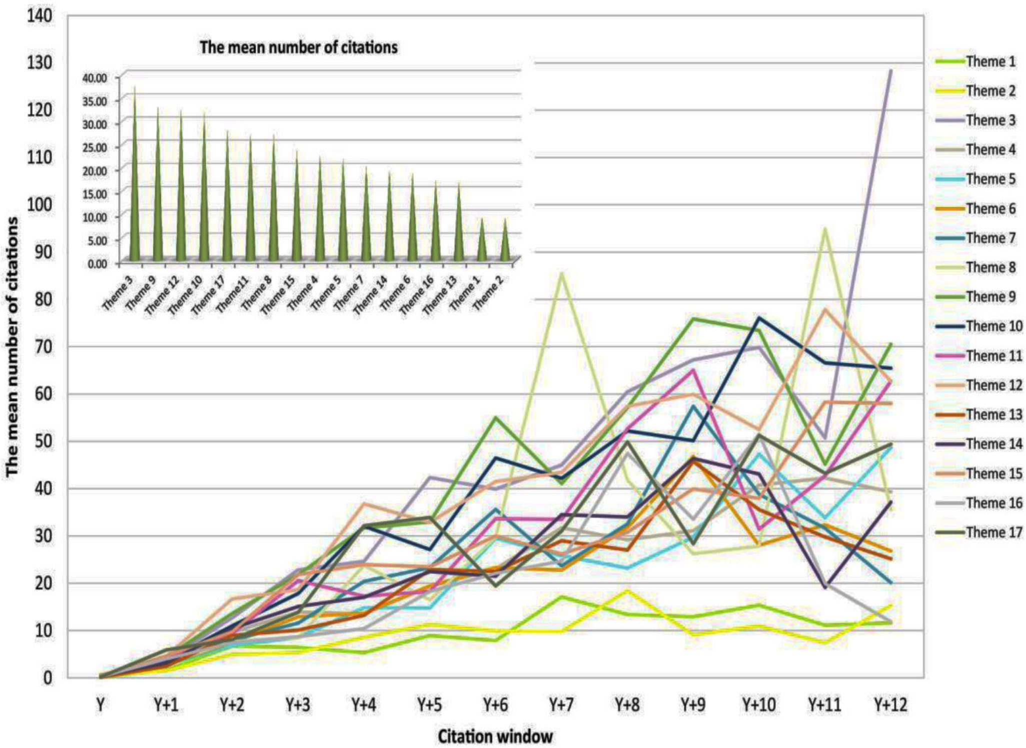

Figure 3

The mean number of citations for each topic during 2010-2022 and the yearly number of citations received per publication year

3.7 *Topics in exoplanet research, their frequencies, h-indices, and evolution*

We performed (TS=) searches in the complete WoS core collection on December 13, 2022. We used the term topic for the result of a single word (or term) search (or a simple combination of such searches), restricting the word theme for a more precise grouping obtained by an LDA analysis. We start with searches (over 5 non-overlapping periods) based on important terms related to exoplanets. Table 5 shows for each topic and each period (1996-2001, 2002-2006, etc.) the number of found publications in the period mentioned on the first line, followed by the h-index (in 2022). We recall that Banks [12] was the first to point out that the h-index can also be applied to topics.

Table 5

Number of articles per topic and their h-indices

Topic	1996-2001	2002-2006	2007-2011	2012-2016	2017-2022
Exoplanet*	85 - 20	532-71	2111-143	3561-144	6760-97
Astrobiology	109 – 22	485-64	691-72	910-73	1349-45
“Kepler mission”	4 – 3	27-16	189-68	352-71	312-37
“habitable zone* “	60-25	197-43	418-67	759-90	1005-60
Transit AND (star* OR planet*)	533-76	900-110	1951-149	2638-144	4225-87
“earth-like planet*	80-22	225-34	338-60	376-64	471-44

In the first period (1996-2001) the topic exoplanet* is ranked 5th among 6 (based on the h-index), while in the most recent period (2017-2022) it has become first, illustrating the growing interest in exoplanets. The Kepler telescope mission was launched in 2009, of course after many years of planning, and retired in 2018. This is clearly illustrated by the data in Table 5.

Next, we combined the previous searches with the search TS = exoplanet*, to obtain more focused results. Similarly, we combined this search with other searches that in themselves were not related to exoplanets, such as tidal, wavefront, and others. These extra terms were taken from the contents of the themes found earlier. We recall that CoRoT was a space telescope mission that operated from 2006 to 2013. Results are shown in Table 6.

Table 6

Focused results

Search: Exoplanet* AND	1996-2001	2002-2006	2007-2011	2012-2016	2017-2022
astrobiology	2 – 1	14 – 8	75 – 35	139 – 45	264 – 32
"Kepler mission"	0 – 0	2 – 1	61 – 34	127 – 45	107 – 26
"habitable zone"	9 – 3	41 – 14	174 – 44	400 – 67	658 – 51
Transit AND (star* OR planet*)	6 – 3	108 - 42	780 – 110	1194 – 120	2230 – 73
"earth-like planet*	3 – 1	32 – 9	122 - 42	168 - 43	297 – 37
coronagraph*	5 – 3	80 – 21	217 – 32	420 – 37	567 – 28
"giant planet*"	17 – 9	112 – 88	532 – 101	731 – 100	980 – 60
extrasolar	17 – 8	217 – 55	804 – 103	986 – 103	843 – 54
mission*	21 – 7	93 – 16	409 – 68	655 – 67	1280 – 49
wavefront*	2 – 2	23 – 7	61 – 16	203 – 24	373 – 18
tidal	2 – 1	24 – 18	142 – 64	227 – 54	342 – 31
"photometric observation*"	0 – 0	3 – 3	39 – 20	46 – 22	102 – 15
spectroscope*	9 – 5	101 – 40	470 – 87	961 – 108	1929 – 70
CoRoT	6 - 2	26 - 7	220 - 54	147 - 38	65 - 15

We found very few publications from the period 1996-2001 when exoplanets were not yet 'hot'. Among all publications dealing with exoplanets about 30% deal with the topic "transit AND (planet* or star*)", except for the first period. About 10% of publications dealing with exoplanets also deal with "habitable zone*". The role of the CoRoT mission and its results is also clearly visible, while the use of photometric and spectroscopic techniques in exoplanet research has increased in absolute frequency over the years.

4. Conclusion

Exoplanet research is a revolutionary field in astronomy, which not only speaks to the large majority of scientists but also to the interested intellectual. Since the detection of the first exoplanet in 1995, rewarded with a Nobel prize in Physics, interest in this field has grown. From our investigations one may predict that Theme 3, dealing with detection, will be the one drawing the most attention in the near future. This corresponds with the plans of the American, European, Chinese, and Indian (ISRO) astronomical agencies to increase the detection and consequently our knowledge of exoplanets. The fact that the emphasis is on detection shows that this field is still in its infancy. The more exoplanets are found, and the better the instruments, the more emphasis will shift to the physical and chemical properties of these exoplanets. With the international JWST (popularly known as the James Webb space telescope) launched on 25 December 2021, there are great expectations for huge progress in the study of exoplanets.

Research limitations – Although we used the Web of Science, together with the Lens and the Derwent Innovation Index, this is still a restricted number of databases, so our conclusions are not based on all existing knowledge.

We conclude by recalling the hope of many to find an exoplanet harboring life outside planet Earth. Whether that would be good or bad for humankind remains however to be seen.

Acknowledgments

This work was supported by the National Natural Science Foundation of China, Grant number: 71974167.

Appendix

Table 4b

LDA results for publications 2010-2022 (8,282 documents), Part b. Themes 10 to 17.

	Labels	The top 30 words
Theme 10	Analysis of the formation mechanism of various giant planets	giant; formation; planet; mass; gas; form; core; disc; abundance; metallicity; composition; disk; evolution; element; structure; super; interior; metal; solar; accretion; model; rich; large; planetary; exoplanets; mantle; high; dust; extrasolar; find
Theme 11	New observations and discoveries of stellar and substellar companions	star; mass; companion; binary; stellar; dot; age; similar; low; host; Jupiter; giant; period; circle; young; sin; object; sequence; main; massive; type; system; group; eff; high; oscillation; determine; determination; find; exoplanet
Theme 12	Atmospheric escape and atmospheric motion analysis of exoplanets	atmosphere; temperature; hot; Jupiter; line; atmospheric; thermal; emission; flux; loss; upper; hydrogen; high; energy; inversion; pressure; equilibrium; rate; find; absorption; ultraviolet; escape; bar; level; radiation; strong; layer; low; dominate; chemistry
Theme 13	Study on the magnetic field environment of stellar wind in interplanetary space	field; solar; magnetic; stellar; planetary; wind; environment; study; radio; exoplanets; interaction; evolution; state; energy; scale; wave; structure; radiation; discuss; dynamic; ray; also; may; principle; law; condition; relate; European; process; physic
Theme 14	Photometric observations and kinematic analysis of dwarfs	star; dwarf; type; around; host; activity; rotation; stellar; brown; sun; late; active; red; white; find; detect; low; nearby; event; flux; show; sample; rotational; similar; cool; variability; large; observe; rotate; near

Contd...

Theme 15	Atmospheric temperature and circulation models for exoplanets	model; phase; effect; radiative; simulation; albedo; cloud; function; side; code; dimensional; surface; change; equation; free; compare; thermal; increase; include; three; general; parameter; heat; two; present; transfer; calculation; different; scatter; temperature
Theme 16	Detection of cosmic life and prediction statistics of exoplanets	Kepler; planet; distribution; number; population; exoplanets; find; similar; per; less; sample; rate; probability; time; within; large; statistical; system; may; planetary; small; observational; possibility; bias; however; expect; propose; koi; initial; Plato
Theme 17	Observation and discovery of the structure of exoplanet debris disk	degree; disk; angle; lambda; separation; axis; inner; angular; similar; direct; large; debris; image; keck; project; central; Rossiter; Mclaughlin; limit; edge; alignment; ELT; planet; pair; system; observation; plane; ring; dust; tau

Table 7

LDA results for publications during the period 1996-2009 (918 documents)

Themes	Labels	The top 30 words
Theme 1	Design, engineering specifications, and technical development of telescopes and observatories	telescope; detection; high; image; instrument; design; phase; interferometer; optical; space; mission; technique; resolution; science; nulling; present; contrast; performance; imaging; planet; base; optic; direct; system; noise; study; coronagraph; interferometry; light; infrared
Theme 2	Observations of exoplanets using e.g., photometric and transit methods, and detection of physical properties	transit; star; COROT; photometric; observation; search; mission; light; space; data; detect; stellar; telescope; exoplanet; photometry; radius; field; time; planet; curve; base; observe; ground; planetary; project; ogle; variation; small; present; detection
Theme 3	New observations and discoveries of stellar companions	star; planet; velocity; radial; mass; companion; survey; dwarf; search; around; period; host; measurement; detection; similar; exoplanet; object; binary; low; type; brown; new; doppler; stellar; detect; planetary; sample; high; exoplanets; system
Theme 4	Orbital evolution, and stability of planet-exoplanetary systems	planet; system; orbit; orbital; mass; eccentricity; planetary; giant; migration; period; resonance; motion; large; Jupiter; show; inner; disk; body; exoplanetary; mean; around; massive; configuration; numerical; distribution; three; similar; find; formation; dynamical
Theme 5	Study of earth-like habitable conditions on planets and exoplanets and their stability	planet; solar; planetary; system; life; exoplanets; stellar; star; extrasolar; earth; discuss; terrestrial; like; study; stability; radio; formation; search; case; activity; emission; evolution; field; paper; surface; year; lead; condition; habitable; close

Contd...

Theme 6	Research on the disk environment of habitable earth-like planets & exoplanets around the zone of the Galaxy and solar system	star; earth; planet; disk; system; habitable; terrestrial; similar; main; microlensing; effect; like; dust; young; around; zone; within; exoplanets; find; large; sequence; galaxy; solar; cluster; small; frequency; region; detect; study; age
Theme 7	Observation of the atmospheric temperature and chemical composition of the extrasolar system	atmosphere; spectrum; atmospheric; infrared; line; model; planet; thermal; temperature; feature; abundance; spectral; emission; H_2O; water; profile; range; show; signature; presence; HD 209458b; key; flux; upper; star; CO_2; hot; present; wavelength; band; spectroscopy; class; observation; extrasolar; chemical; composition; absorption; comparison; cloud; large;
Theme 8	Study of physical properties of giant planets	Jupiter; exoplanets; mass; planet; giant; solar; model; hot; distance; VLTI; planetary; loss; core; type; rate; structure; radius; star; density; theoretical; gas; study; Saturn; wind; stellar; review; high; close; time; energy

References

[1] T. Sall, "Detection of extraterrestrial life - A review of the state of the art," Trans. New York Acad. Sci., vol. 26, no. 5, pp. 553–563, 1964. [Online]. Available: https://doi.org/10.1111/j.2164-0947.1964.tb01493.x

[2] K. Schrijver, One of Ten Billion Earths, Oxford University Press, 2018.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, no. 4/5, pp. 993–1022, 2003. [Online]. Available: https://dl.acm.org/doi/10.5555/944919.944937

[4] D. R. Hitchcock and J. E. Lovelock, "Life detection by atmospheric analysis," Icarus, vol. 7, pp. 149–159, 1967.

[5] M. Coccia, "General properties of the evolution of research fields: a scientometrics study of human microbiome, evolutionary robotics and astrobiology," Scientometrics, vol. 117, no. 2, pp. 1265–1283, 2018. [Online]. Available: https://doi.org/10.1007/s11192-018-2902-8

[6] National Academies of Sciences, Engineering, and Medicine, Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032, Washington, DC: The National Academies Press, 2022. [Online]. Available: https://doi.org/10.17226/26522

[7] Y. Ye, "China is hatching a plan to find Earth 2.0," Nature, vol. 604, no. 7906, p. 416, 2022. [Online]. Available: https://doi.org/10.1038/d41586-022-01025-2

[8] F.-J. García-Marco, C. G. Figuerola, and M. Pinto, "Análisis de la evolución temática de la investigación sobre Información y Documentación en español en la base de datos LISA mediante modelado temático (1978-2019)," Profes. Inf., vol. 29, no. 4, e290427, 2020. [Online]. Available: https://doi.org/10.3145/epi.2020.jul.27

[9] Y.-S. Lyu, M.-Q. Yin, F.-J. Xi, and X.-J. Hu, “Progress and knowledge transfer from science to technology in the research frontier of CRISPR based on the LDA model,” J. Data Inf. Sci., vol. 7, no. 1, pp. 1–19, 2022. [Online]. Available: https://doi.org/10.2478/jdis-2022-0004

[10] C. Sievert and E. K. Shirley, “LDAvis: a method for visualizing and interpreting topics,” in Proc. Workshop Interact. Lang. Learn., Visualization, and Interfaces, C. Chuang, G. Green, J. Hearst, J. Heer, and P. Koehn, Eds., Baltimore, MD, USA, 2014, pp. 63–70. [Online]. Available: https://doi.org/10.3115/v1/W14-3110

[11] Y. Miyata, E. Ishita, F. Yang, M. Yamamoto, A. Iwase, and K. Kurata, “Knowledge structure transition in library and information science: topic modeling and visualization,” Scientometrics, vol. 125, no. 1, pp. 665–687, 2020. [Online]. Available: https://doi.org/10.1007/s11192-020-03657-5

[12] M. G. Banks, “An extension of the Hirsch index: indexing scientific topics and compounds,” Scientometrics, vol. 69, no. 1, pp. 161–168, 2006. [Online]. Available: https://doi.org/10.1007/s11192-006-0146-5.