DSI DATA SCIENCE AND INFORMETRICS
VOL.3, NO.1, FEB. 2023

RESEARCH ARTICLE

Construction of portrait model of core professional researchers in the technical field and analysis of their research topics

Yuefen Wang[ab*], Tianao Xu[b], Yonghua Cen[a], Xue Yang[b]

a. Institute for Big Data Science, Tianjin Normal University, Tianjin, China
b. School of Economics and Management, Nanjing University of Science & Technology, Nanjing, China

ABSTRACT
In order to support scientific evaluation and decision-making, this paper characterizes and identifies core researchers and their research topics by modeling the portrait of core researchers in technical fields. In this paper, we constructed a portrait model of core researchers, explained the information attributes and their label data in each dimension of the portrait. Selected six identification indicators that can be measured, used the CRITIC method and gray correlation method to determine the weight coefficient and comprehensive score of the index, and designed quantitative identification methods for core researchers using Price's calculation formula and golden section coefficient. Took the field of artificial intelligence as an example for empirical analysis and result verification, selected representative core researchers to track the changes of research topics at the two levels of their teams and individuals, and characterize them based on the portrait model.

KEYWORDS
Portrait of core researchers; Identification of core researchers; Research topics; Artificial Intelligence

1 Introduction

After entering the era of "big data", network information exploded, which enriched and made life and work easier for people, but also caused data redundancy, data explosion and other problems. As a result, how to accurately capture target data and intelligence has become a major challenge in the current context of big data. Thus, the concept of "user portrait" has progressively emerged in people's vision. User portraits can help governments, businesses, researchers, etc., to set up precise information services, and their fields of application generally concern electronic commerce, health care, tourism and others.

Core researchers are the backbone of promoting the development of the technical fields and play the role of leader and guide. However, due to the complexity of components and influencing factors used to judge core researchers, it is difficult to accurately identify and

* Corresponding Author: yfwang@njust.edu.cn

track core researchers. With intensifying scientific and technological competition in different countries all over the world, it is urgent to scientifically evaluate talents and then improve national scientific and technological strength in a targeted manner. Therefore, this paper attempts to use the concept of "user portrait" to systematically characterize and identify core researchers and their research topics in the technical fields in a multi-dimensional way, so as to more accurately identify and grasp professional talents and their research status.

2 Related research

After studying the previous literature, this paper believes that the research on the identification of key personnel in the field of professional technology mainly focuses on the characterization of core researchers, core researchers identification and author topics analysis.

(1) Core researcher characterization

The idea of core researcher characterization comes from the concept of user portraits, which can help governments, enterprises, researchers and others to achieve accurate information services (Liu & Xu, 2021); however, the use in academia is limited to the construction of portraits of library users and researchers in a general sense, while less research has been conducted on the user portraits of core researchers. Relevant research on portraits of researchers has shown that the main dimensions of portraits of researchers are:

①Basic attribute dimensions, such as name, gender age, education, title, institution, correspondence address, etc. (Xie et al., 2019; Yao et al., 2020; Gao et al., 2019). ②Professional and academic dimensions, such as published literature, projects and awards (Gao et al., 2019; Fan, 2018). ③Research interest dimension, such as interest topic and research direction. ④Dimension of scientific research cooperation: such as collaborators, cooperative achievements, etc.

In contrast to researchers in the general sense, core researchers have their own particular descriptive dimensions, including extrinsic elements that are considered to be directly related to the status of core searchers, such as number of published articles, average citation (Costas & Bordons, 2008), citation (Zhang et al., 2018), H-index (Qiu & Zhou, 2008), etc., as well as the growth rules in time dimensions such as educational structure (Zong, 2005), growth cycle (Li & Yao, 2009) and mentoring relationship (Bu & Xu, 2009), and in spatial dimensions such as the number of institutional changes (McIntosh, 1989) and geographical mobility trends (Wang et al., 2015).

(2) Core researchers´ identification

At present, there are few strict studies on the recognition of core researchers, which are mostly expressed as the recognition of highly influential scholars, the influence evaluation of scholars, and the recognition of core scholars / authors, all of which are aimed at identifying the core and influential researchers in the field.

The methods used in existing studies are mostly based on different dimensions such as number of articles published, number of citations, average frequency of citations, h-index, Altmetrics, etc. Experts Grading method, entropy method, Factor analysis and other methods are used to calculate scholars' scores. The scholars with higher scores are the core researchers. Ma et al. (2011) constructed a comprehensive evaluation index system for core researchers by using the number of published articles, citation amount and citation frequency of authors, and determined the weight of each index by using expert consultation method. Guo and Xiao (2019) selected Altmetrics indicators from platforms such as CNKI, ScienceNet and WeChat Official accounts, then further filtered the indicators and determined their

weights through factor analysis, so as to calculate the Altmetrics score of each scholar.

In addition, the measurement method based on complex network is also commonly used. The author cooperation network is constructed according to the collaborative relationship between authors, and the importance of nodes is judged by the weight information of nodes and edges in the author cooperation network, so as to learn the importance of authors and identify core researchers (Gao & Zhang, 2016). Griffin et al. (2016) built a bipartite network of journal authors, and selected the top 1% authors in the network as core researchers according to the eigenvector centrality of nodes.

(3) Author topics analysis

Currently, national and international scholars use both bibliometric and text-mining methods to study author research topics. The bibliometric approach mostly uses keywords in the researcher's published literature to characterize the research interests of the researcher. For example, Sun (2012) selected a domestic library intelligence research field, constructed an author-keyword co-occurrence network within the field, and visualized this two-mode network so as to discover the research themes of researchers; Ren (2013) first identified the core authors in the field of archives based on their papers and citation counts, and then discovered their research topics by counting and analyzing the high-frequency keywords of the core authors. The model used in text mining methods is mostly the AT model and its base variants. The AT model (Author Topic Model) is an extension of the LDA topic model, which adds the author element to the LDA topic model and can extract each author's research topic. For example, He et al. (2019) extracted author research topics based on the AT topic model, constructed an author-topic association matrix and visualised it to generate a two-mode subject knowledge network of author-topic associations, and measured the influence of nodes through network metrics.

Through the above analysis, the current relevant research still focuses on the characterization of researchers in the general sense and has not paid attention to the importance of identification and representation of core researchers. Different from the previous research on common researchers, core researchers should have their own special description dimension. The existing research is relatively simple and scattered in terms of core researchers' identification and author's theme analysis. It is necessary to further improve the research methods for the identification and representation of core researchers and effectively expand its practical application direction.

3 Research design

3.1 Research idea

This paper focuses on core researchers and divides the main process of the study into four parts: the construction of core researchers' portrait model, the identification of core researchers, the analysis of core researchers' research topics, and the visualization of core researchers' portraits. Firstly, the portrait model of the core researcher is constructed from the perspective of the overall situation and the theoretical level. Take the field of artificial intelligence as an example, judge the comprehensive ability of researchers according to the characteristics of core researchers in academic research, so as to identify core researchers. Then, we use the bibliometric method to analyze the research topics of core researchers in the field of AI from the team and individual levels. Finally, based on the quantitative identifica-

tion of core researchers through the portrait model, the qualitative information is supplemented to build a visual portrait of core researchers to represent core researchers.

3.2 Core researcher portrait model construction

Different from previous studies in which the representation of researchers is mainly based on static indicators, this paper combines the composition of core researchers' elements with the growth law to determine the dimensions of core researchers' portrait. The basic model of the core researcher portrait is shown in Figure 1, including personal attribute dimension, personal growth dimension, research achievement dimension, research interest dimension and cooperation dimension.

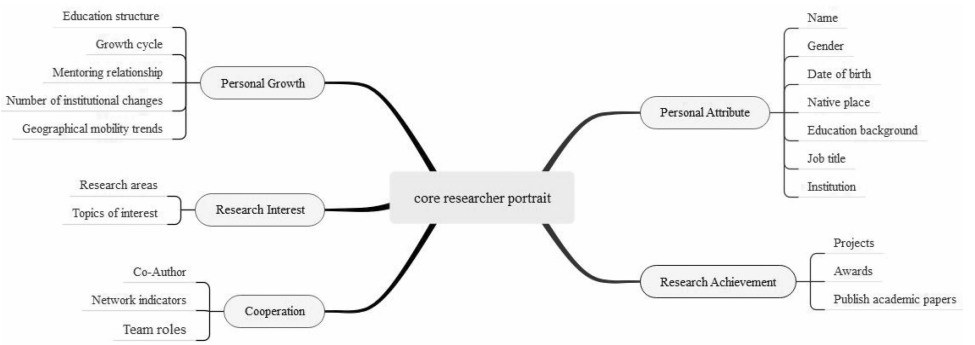

Figure 1 The basic model of the core researcher portrait

The following is a description of label information involved in each dimension:

(1) Personal Attribute Dimension

The personal attribute dimension of the core researcher portrait corresponds to the basic attribute dimension of the user portrait, which means the demographic information of the core researcher, including objective information such as name, gender, date of birth. The information mainly comes from personal homepage, encyclopedia, Aminer, Sciencenet and other platforms. This kind of label is concise, so the collected information can be directly used as the label data. The details are shown in Table 1.

Table 1 Personal attributes dimension information attributes

Label name	Secondary labels	Description
Name	——	Name of core researcher
Gender	——	Male, Female
Date of birth	——	Date of birth of core researcher
Native place	——	Country name + exclusive geographical distinction by country, e.g., China: Northern China, Eastern China … (Zhang, 2019)
Education background	First–record of college	First–record of college name
	First–record of college attribute	QS World University Rankings
	Highest education	Doctor, Master and Bachelor
	Highest degree university	Highest school record name
	Highest degree university attribute	QS World University Rankings
Job title	——	Assistant, Lecturer, Associate Professor, Professor
Institution	Name of Institution	Name of core researcher´s current institution
	Institution attribute	QS World University Rankings

(2) Personal Growth Dimension

The personal growth dimension of core researchers is an important dimension that is different from the portrait of common researchers in the existing research, which describes the past experience of core researchers from the two dimensions of time and space. The time dimension includes the academic structure, growth cycle and mentoring relationship of the core researchers; The spatial dimension includes the number of institutional changes and their geographical mobility trends. The personal growth dimension labels can be obtained by collecting the data of the researchers' educational background and educational experience through personal homepage, Baidu Encyclopedia and other websites. The required information is extracted from the data by means of statistical methods. Details are shown in Table 2.

Table 2 Personal growth dimension information attributes

Label name	Secondary labels	Description
Education structure	Interschool flow	the number of institutions attended
	Growth path	"Undergraduate–Postgraduate" mode, "Postgraduate–Doctor" mode, "Direct–entry PhD" mode, "Undergraduate–Postgraduate–Doctor"mode (Jiang et al., 2018)
	International exchange experience	None, exchange communication, undergraduate, postgraduate, PhD, postdoctoral researcher
Growth cycle	——	academic activities since graduation (time from graduation to winning various awards)
Mentoring relationship	——	basic information of tutor, tutor´s personal achievements
Number of institutional changes	——	Number of changes of workplace
Geographical mobility trends	——	China/USA: Intra–urban mobility, inter–urban mobility, inter–state mobility, international mobility; similar in other countries

(3) Research Achievement Dimension

Compared with user portraits, the research achievement dimension is a new dimension based on the internal characteristics of core researchers. This dimension mainly describes the level of scientific research achievements published by core researchers, and its data mainly comes from the literature database platform such as Web of Science, CNKI, Wanfang Data, etc. Through the personal homepage of core researchers, you can directly obtain the information about the core researchers' topics and awards for the label system.

Obtain the relevant data of the published documents through the document database platform, then use the document metrology method to analyze the number of published papers, the number of citations, the frequency of citations, the h index and other label information with the help of Python, R or bibliometric software. Details are shown in Table 3.

Table 3 Research achievement dimension information attributes

Label name	Secondary labels	Description
Projects	Number of projects	——
	Type/level of the project	national level, ministerial and provincial level, prefecture level, school level；such as USA——NSF fund, China——NSFC fund…

Label name	Secondary labels	Description
Awards	Number of awards	——
	Award type	Science and technology awards established by the government, such as the US National Science Award, the Swedish Crafoord Prize, the Chinese National Science and Technology Award, etc. Science and technology awards established by social forces, e.g. Nobel Prize, Fields Prize, Turing Award; most influential scholars in the field, etc. ……
Published literature	Issue time	publication time of scientific research achievements of core researchers
	Publications	journal: JCR journal divisions (Q1, Q2, Q3, Q4); conference: conference Level (A, B, C…)
	Number of articles issued	number of SCI papers published by core researchers
	Number of citations	total citation frequency of SCI papers published by core researchers
	Average citation frequency of articles	average frequency of citations per paper
	h index	at most h papers have been published and cited at least h times respectively
	…….	……

(4) Research Interest Dimension

The research interest dimension in the core researcher profile evolves from the interest dimension in the user profile and indicates the research areas and topics of interest of the core researcher. The research area refers to the academic field in which the core researcher's research topic is located and can be accessed directly through the core researcher's personal homepage, academic website, etc. Topics of interest are based on data analysis of keywords, titles and abstracts of literature published by core researchers. The degree of interest of scientific researchers in the research subject is reflected by the frequency of the use of subject words. At present, there are two main methods to generate subject words of researchers: 1) keyword characterization; 2) text mining, using LDA model or AT model and its extended model to mine research subjects. Details are shown in Table 4.

Table 4 Research interest dimension information attributes

Label name	Description
Research areas	academic field of the research subject, such as computer, medicine, etc.
Topics of interest	subject words

(5) Cooperation Dimension

The cooperation dimension in the core researcher portrait corresponds to the social dimension of the user portrait, which mainly describes the team cooperation situation of the core researcher. First, obtain the information of the co-authors according to the published literature of the core researchers. The network indicators of the core researchers in the collaborative network were then calculated with the help of Pajek social network analysis software using the centrality analysis method in social network analysis. In addition, the role played by the core researcher in the team is determined based on the number of collaborators in the team and network indicators. The co-author label is expressed as the name of the high-frequency co-author of the literature published by the core researcher, the network

indicator label is the calculated indicator value and the role label is expressed as the role title of the core researcher. The definition method will be described below and details are shown in Table 5.

Table 5 Cooperation dimension information attributes

Label name	Description
Co–Author	information about co–authors of the core researcher
Network indicators	degree, closeness, betweenness
Team roles	leaders, communicators, followers, free agents

3.3 Core researcher identification based on portrait models

On the basis of understanding the current status of research on core researcher identification and constructing a basic model for core researcher portrait, this section will select core researcher identification indicators with a focus on academic research from an operability perspective. The CRITIC method and grey correlation analysis method are adopted to build a quantitative identification model for core researchers, and the comprehensive scores of each researcher are calculated to identify core researchers.

(1) Core researcher identification indicators

In this paper, indicators that can effectively identify core researchers are selected from the research achievements dimension, research interest dimension and cooperation dimension of the core researcher portrait model.

When selecting indicators, the availability, reliability, distinguishability and quantification of indicators shall be fully considered. Each indicator must provide information independently and there must be no derivation of indicators from each other. Table 6 introduces a brief description, deeper meaning and occurrence of the selected indicators.

Table 6 Description of core researcher identification indicators

Indicator	Description	Meaning	Previous research
Number of articles	Number of SCI papers published by researchers (fraction counting method)	The number of publications is one of the most intuitive and traditional indicators used to measure the output of researchers	(Cortés et al., 2016; Wang et al., 2019)
Citation frequency per article	Total citations of SCI papers published by the researcher divided by the total number of papers	The average citation frequency is a good and stable indicator of the citation impact of the authors, since the influence of the total number of papers on the citation frequency is excluded	(Costas & Bordons, 2008)
h index	At most h papers have been published and cited at least h times respectively	The h–index combines "quantity" and " quality" to measure the academic output and impact of scholars	(Qiu & Zhou, 2008; Ma & Lv, 2011; Ma & Jiang, 2018)
Similarity of research interests	Similarity between researchers´ research interests and research interests in the whole field	Reflect the professional contribution and research frontier of researchers	new indicator
Weighted centrality	Weighted degree centrality takes into account the edge weight value (cooperation times). The weighted degree centrality of a node is the sum of the edge weight values connected to the point	In a social network, if a researcher has direct contact with many other researchers, the researcher will be in the center position and have greater "power" in the network	(Li & Zhu, 2008)

Indicator	Description	Meaning	Previous research
Betweenness centrality	An indicator that uses the number of shortest paths through a node to indicate the importance of the node	If a researcher is on many "paths", it can be considered that he is important in the overall connectivity of the network, because he has the ability to control the information/resource exchange between the other two researchers	(Li & Zhu, 2008)

(2) Calculation method of researchers´ comprehensive scores

The CRITIC grey correlation method is improved according to the research objectives based on the entropy-weighted grey correlation method proposed by Chen and Zheng (2013). Grey correlation analysis is a method for quantitatively analyzing the degree of correlation between two objects. The degree of correlation reflects the closeness of each evaluation object to the specific object. The closer the object is, the greater the degree of correlation is. CRITIC method is an objective weighting method superior to entropy weight method and standard deviation method. It comprehensively measures the objective weight of indicators based on the comparative strength of evaluation indicators and the conflict between indicators (Lin et al., 2018). In this paper, CRITIC method is selected as the method to determine the index weight, and CRITIC grey correlation method is proposed to calculate the comprehensive score of researchers.

(3) Quantitative identification model of core researchers

Referring to Price's calculation formula, this paper selects researchers whose number of published papers is equal to or greater than 0.749 times the square root of the number of papers published by the most productive researchers as candidate core researchers.

Core researchers are a subset of candidate core researchers, this paper uses the classic golden section coefficient τ to quantify and define the core set of researchers (Qian, 2014). According to the above RCI calculation method, calculate the RCI values of candidate core researchers and rank them in descending order, Define researchers who meet the requirements of $RCI_i > RCI_r$ as core researchers. The solution process of r $(1 \le i \le r)$ is as follows:

$$\sum_{i=1}^{r} RCI_i = (1-\tau)^{\kappa} \cdot \sum_{i=1}^{n} RCI_i \quad (1)$$

τ =0.618 , i=1~n , κ =1,2,3. The set of core researchers determined from κ =1 to r is known as the Zone 1 core researchers; the set of core researchers determined from κ =2 to r is known as the Zone 2 core researchers; and the set of core researchers determined from κ =2 to r is known as the Zone 3 core researchers.

4 Empirical analysis

4.1 Data collection and processing

This paper uses the core collection in the Web of Science (WoS) as the data source and WC="Computer Science, Artificial Intelligence" as the logical search formula to search for data in the artificial intelligence category in the field of computer science under the WoS subject classification. The retrieval period is set as 2009-2018, and the retrieval time is January 2019. The final retrieval result contains 467630 records. From these records,

bibliographic information of the literature in the field is obtained including: title, author, author's affiliation, keywords, abstract, date of publication of the literature, citations of the literature, etc. To address the issue of renaming between authors in the data, authors were disambiguated prior to analysis. The names of institutions are first cleaned according to the name of the country to which they belong, and then a distinction is made between authors with the same name according to whether or not there is an intersection between the institution to which they belong and their co-authors. After disambiguation, 656668 authors are obtained.

4.2 Core researcher identification and validation in the field of artificial intelligence

(1) Calculation of composite index of researchers

Considering the huge number of authors, if we analyze the full amount of data, it will cause a lot of waste of manpower and material resources. The number of publications issued is the most intuitive and traditional indicator to measure the output of authors. Therefore, this paper first identifies candidate core researchers based on Price's law. N_{max} is 134.7813, and the calculated value of M is 8.6955. A total of 2831 authors with 8.6955 publications and above were initially selected as candidate core researchers. On this basis, the original evaluation matrix of 2831*6 was established, the correlation coefficient of each evaluation index was calculated using grey correlation analysis, and the weights of each index were determined using the CRITIC method, the results are shown in Table 7. The values of the core researcher composite index were calculated according to Table 7, and they were ranked in descending order according to the RCI values, as shown in Table 8.

Table 7 Index weight distribution

Number of articles	Citation frequency per article	H index	Similarity of research interests	Weighted point degree centrality	Intermediary centrality
0.1177	0.1222	0.1883	0.2976	0.1459	0.1283

Table 8 Composite index of core researchers and its ranking

Researcher	RCI	Rank(RCI)	Institution	Country/Region
Zhang, Lei_4(张磊)	0.5681	1	Hong Kong Polytechnic University	China
Tao, Dacheng(陶大程)	0.5604	2	University of Technology Sydney	Australia
Vlahavas, Ioannis	0.5553	3	Aristotle University of Thessaloniki	Greece
Hassanien, Aboul Ella	0.5517	4	Cairo University	Egypt
Li, Xuelong(李学龙)	0.5300	5	Chinese Academy of Sciences	China
Shi, Yong_3(石勇)	0.5270	6	University of Nebraska	China
Abraham, Ajith_1	0.5215	7	University of Ostrava	Czech
Zhang, Shichao(张师超)	0.5177	8	Guangxi Normal University	China
Shi, Peng_3(石碰)	0.5153	9	Victoria University	Australia
Pedrycz, Witold_2	0.5133	10	University of Alberta	Canada

The core researcher sequence threshold r can be calculated according to equation (1), and the result (k=1, r=1020;k=2, r=368; k=3, r=130) indicates 1020 core researchers in Zone 1,

368 core researchers in Zone 2 and 130 core researchers in Zone 3.

(2) Validity test

In order to verify the scientificity and accuracy of the above results, the TOPSIS method and the RSR method were chosen to validate the results of the ranking of the researchers' composite index. The correlation between the ranking results of the three methods was calculated using SPSS 24.0 software, as shown in Table 9. The results show that the ranking results of the three methods are highly positively correlated, indicating that the RCI can reflect the characteristics of researcher's research behaviour and can be used to identify core researchers.

Table 9 Correlation between RCI ranking, TOPSIS method, RSR method ranking results

Pearson Correlation	RCI	TOPSIS	RSR
RCI	1	0.881	0.759
TOPSIS	0.881	1	0.840
RSR	0.759	0.840	1

4.3 Analysis of the research topics of the core researchers

In order to understand the research status of core researchers and provide help for scientific research talents, so as to promote the continuous development of fields that researchers belong to. This section provides further analysis of the research topics of the core researchers identified above. The analysis of the research topics of the core researchers is carried out at both team and individual levels, providing a reference for the evaluation and representation of researchers, the training of talents and the grasping of research trends. Firstly, based on the country or region where the core researchers are located, the core researchers are classified into two types, domestic researchers and foreign researchers. Core researchers from each type were selected as examples to analyse the thematic evolution of the individuals and teams to which they belong. We took Zhang, Lei from Hong Kong Polytechnic University as an example in domestic researchers, and his RCI ranking result is the highest. In foreign researchers, we took Pedrycz and Witold_2 from the University of Alberta as example, his RCI ranked 10th.

The analysis steps can be divided into the following four steps:

Step 1: Keyword data for all members of the team/the core researcher's articles are extracted from the downloaded data in WOS data format based on self-coded Python.

Step 2: Determine the high-frequency keywords (top 10) in the research period according to the frequency of keywords.

Step 3: Calculate the number of occurrences of the top 10 keywords in each year of publication (both team and individual) and plot them in a hierarchical stack.

Step 4: Interpret the evolution of team/individual topics according to graphics. On the basis of the quantitative analysis, the qualitative information of representative core researchers from the micro perspective is supplemented to draw their visual portraits.

(1) Team identification and team role determination

Taking the global data in the field of artificial intelligence as the research object, the Louvain community detection algorithm in Pajek is used to identify the research team in the field of artificial intelligence. The process is as follows :

First, construct a domain-wide co-authorship network based on the downloaded literature, involving a total of 656668 nodes and 2042924 edges. In order to improve the accuracy of the identification results, the co-authorship network was pruned. After deleting nodes in the original co-authorship network that have 1 publication and less than 100 citations, the new co-authorship network involves 186,997 nodes and 543,351 edges. Compared with the original co-authorship network, the overall co-authorship network is significantly reduced while maintaining the original important nodes. Continuously adjust the parameters by observing the familiar team divisions to verify the validity of the results. When the parameters are set to Resolution=290, Max Level=13, Max Iteration=13, a suitable team recognition result is achieved. In the end, 23,423 research teams were identified, involving a total of 186,997 researchers.

This paper uses the method of identifying team cooperation patterns in the study by Wang et al. (2020) to determine team roles for the identified research teams in terms of the cooperation of team members within the team and social network indicators.

The degree centrality D_i of node i in the team represents the number of authors cooperating with author i; The cooperation frequency of author i is represented by the weighted degree centrality WD_i of node i, denotes the number of posts by author i in pub_i and the total number of team members in Z. The core researcher roles are defined as Leaders, Communicators, Followers and Free Agents according to the position of the core researcher in their respective teams and the shape of the team network. The specific operational steps are designed as follows:

① Data pre-processing: Sort the nodes in descending order based on D_i, if $D_i=D_j$, further compare WD_i with WD_j; if $WD_i=WD_j$, further compare pub_i with pub_j.

② Identify team leaders: Determine the proportion between the number of authors in the team and the total number of team members in turn. Node i is the only team leader if the number of collaborators at the first node i satisfies $D_i/Z>=\alpha$ ($0<\alpha<1$); otherwise, judged in the following way, If there exist nodes i, j, k, satisfying $D_i>D_j>D_k$ (if $D_i=D_j$, then further compare WD_i with WD_j), and the number of nodes/$Z<=0.2$ such that $(D_i+D_j+D_k)/Z>=0.8$ (based on the two-eight law), then nodes i, j, k are all team leaders. It should be noted that the summation of D_i , D_j and D_k requires further exclusion of duplicate scholars in D_i , D_j and D_k , i.e., if nodes i, j and k all cooperate with node a, then a is only counted once.

③ Identify team roles: If there is only one leader in the team, it will be defined as the only leader. If there are multiple leaders in the team, they will be sorted according to the number of partners, and they will respectively be defined as the first leader, the second leader... At this time, the non leaders in the team are defined as followers; If there is no leader in the team, all members of that team are defined as free agents. A special category of researcher exists beyond these roles, namely the researcher who acts as a bridge between the two sub-networks in the team, defined as the team communicator.

In the process of determining team roles, the proportion between the number of collabrators in the team and the total number of team members has a decisive influence on the division results, which is called $Di/Z=\alpha$ threshold settings. After many experiments and combined with expert opinions, α is set to 0.8. According to the above methods, judge the team roles of the representative core researchers mentioned above, results are shown in Table 10.

Table 10 Team identification results and team roles of representative core researchers

Name of core researcher	Number of team members	Team Role
Zhang Lei（张磊）	34	Second leader
Pedrycz, Witold	52	only leader

(2) Team topic analysis

The keywords and publication time data of the above representative core researchers' teams were obtained, and the evolution of their research topics was indicated by the change in the frequency of high-frequency keywords in the time dimension. The team members' research topics were further analysed in relation to their individual homepages, titles of the team's published literature, and abstract data. The thematic evolution of the team to which Zhang Lei belongs is shown in Figure 2.

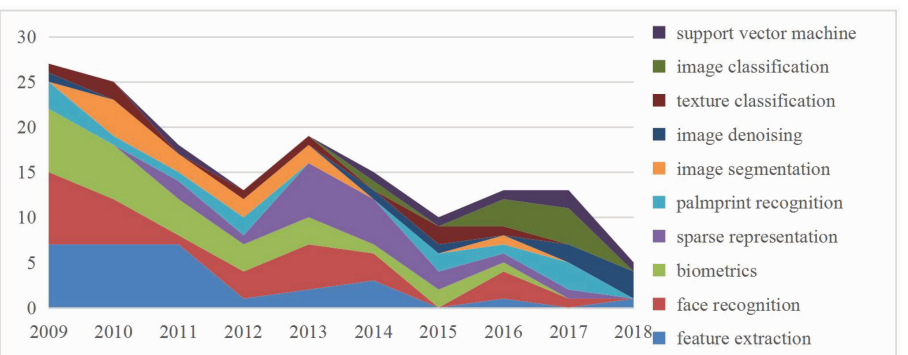

Figure 2 Evolution of the research topics of Zhang Lei's team

From 2009 to 2018, the main topics of interest for Zhang Lei's team are feature extraction, face recognition, biometrics, sparse representation, palmprint recognition, image segmentation, image denoising, texture classification, image classification and support vector machine. Further analysis shows that the team's research interest in the topics shown in Figure 2 is declining. Comparing the change of time dimension of each topic, it can be found that the team's research focus has gradually shifted from feature extraction, face recognition, biometrics to image segmentation and image classification.

The thematic evolution of Pedrycz and Witold's team is shown in Figure 3.

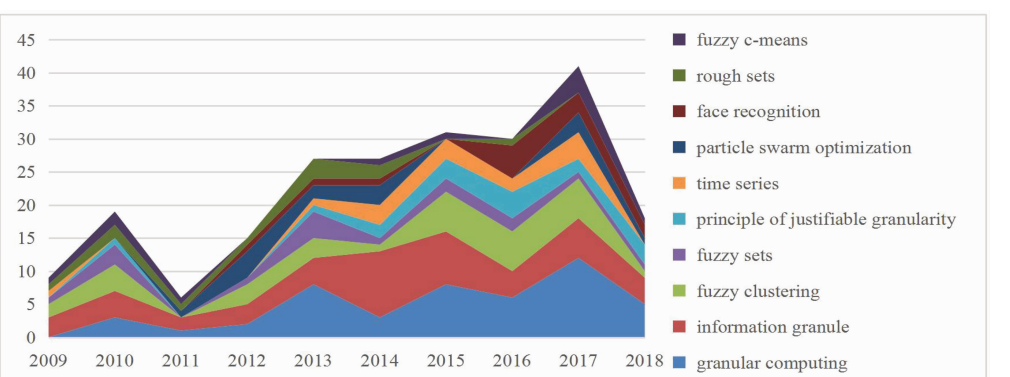

Figure 3 Evolution of the research topics of Pedrycz, Witold's team

It can be found that during 2009-2018, Pedrycz and Witold's teams mainly focused on the following topics: granular computing, information granule, fuzzy clustering, fuzzy sets, principle of reasonable grammar, time series, particle swarm optimization, face recognition, rough sets and fuzzy c-means. Overall, the team's research interest in these topics has shown a continuous upward trend, indicating an increasing level of interest and a more continuous focus on these topics. The team's core focus is on information grains and fuzzy sets, from which related research on algorithms and models has been derived.

Further analysis reveals that the team's research frontiers focus on three areas: time series, fuzzy cognitive maps and data mining, which specifically involve the application of fuzzy sets, feature space, information grains, time series, cognitive maps, fuzzy logic, and theories and methods such as rough sets, rational strength principle, and fuzzy C-mean clustering.

(3) Individual topic analysis

Similar to the team-level research topic analysis method, the keywords of the published literature of the representative core researchers are used to characterize their research topics, and the changes in high-frequency keywords are used to describe the evolution of their topics. Research topics are further interpreted in conjunction with the personal homepages of the representative core researchers, the titles and abstracts of the published literature and other data. The evolution of Zhang Lei's research topics is shown in Figure 4.

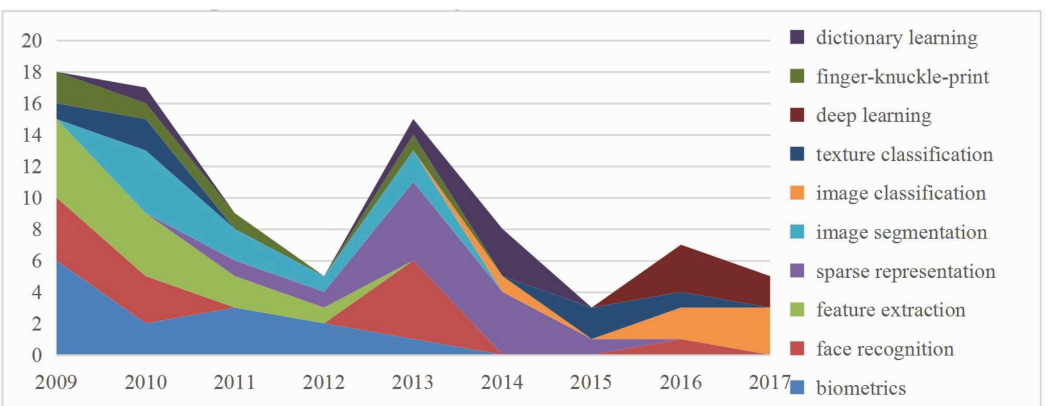

Figure 4 Evolution of the research topics of Zhang Lei

From 2009 to 2017, Zhang Lei's main topics of interest are biometrics, face recognition, feature extraction, sparse representation, image segmentation, image classification, texture classification, deep learning, finger-knuckle-printing and dictionary learning. The evolution of topics can be interpreted in three phases. During the period 2009-2012, Zhang Lei's main topics of interest were biometrics, face recognition and feature extraction. During the period 2012-2015, its research focus shifted to sparse representation and face recognition-related research increased further. During the period 2015-2017, he started to focus on deep learning, feature extraction, image classification, and face recognition was also relaunched, and these changes can characterize the evolution of its research focus in different phases.

As shown in Figure 5, the main topics of interest for Pedrycz, Witold during the period 2009-2018 were granular computing, information granules, fuzzy clustering, time series, fuzzy sets, principle of justifiable granularity, rough sets, particle swarm optimization, feature selection and fuzzy c-means. In comparison to their team, the core researcher's topics of in-

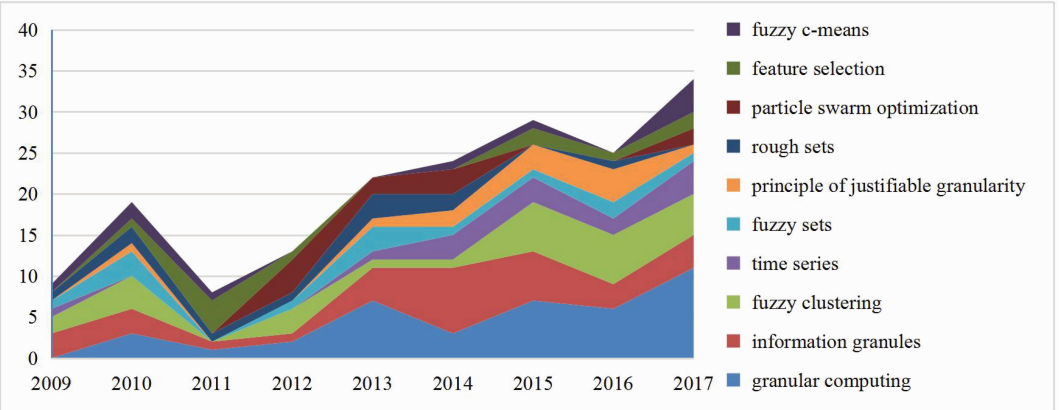

Figure 5 Evolution of the research topics of Pedrycz Witold

terest are essentially the same as those of their team, and the trends are similar, both increasing and more continuous. The topic with the highest attention intensity is information particles and their computation.

Combining the integrated resources on Microsoft Academic and Aminer websites, it can be found that Pedrycz Witold's recent research focuses on three aspects: "Neural network", "Fuzzy set" and "Pattern recognition".

4.4 Visualization of core researchers´ portraits

Before the visualization of the portrait can be carried out, it is necessary to obtain label data in five dimensions. The data sources of these labels in the personal attribute dimension and research achievement dimension are the same, including personal homepage, Baidu Encyclopedia, talent topic database, Google Academic, Bing Academic, Aminer development data set, etc. It should be noted that the dimension of personal growth needs to be analyzed based on the information obtained. For example, Zhang Lei entered the Department of Aeronautical Engineering of Shenyang Institute of Aeronautical Engineering in 1995 and

Figure 6 Visualization portrait of core researcher Zhang Lei

studied in the Department of Automatic Control of Northwestern Polytechnical University from 1995 to 2001. Therefore, he belongs to "Postgraduate-Doctor" mode. He formally joined the Department of Computing of the Hong Kong Polytechnic University in 2001 and was progressively promoted from Assistant Research Fellow to Professor until 2015, during which time he had three years of overseas study experience. Therefore the number of his institutional changes is two, and his geographical mobility trends include inter-provincial mobility and international mobility.

In order to more vividly represent the core researchers, set the background picture of the core researchers' portrait as their photos. The labels of five dimensions are imported into Tagul. Use different colors to represent different dimensions to form a visual portrait, as shown in Figure 6 and Figure 7.

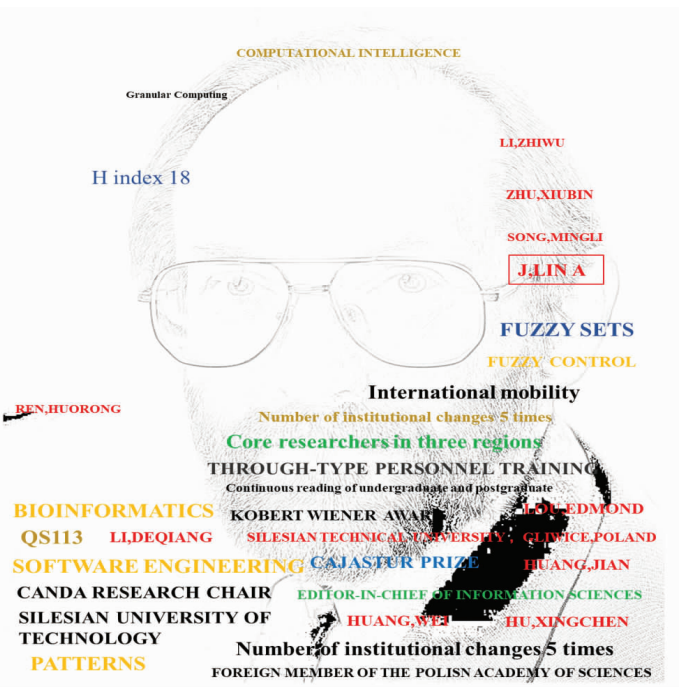

Figure 7 Visualization portrait of core researcher Pedrycz, Witold

5 Summary and Prospect

This paper constructs a comprehensive and detailed core researcher portrait model, selects indicators that can effectively evaluate core researchers, and proposes a core researcher identification method. On this basis, take the artificial intelligence field as an example to identify the core researchers in this field, then realize the research topic analysis of the core researchers from the team and individual levels. This paper can provide reference for scholars' representation, scholars' evaluation, talent training and grasping the research trends in this field. The details are as follows:

(1)Exploring the dimensions of the core researcher portrait, the results of the core researcher portrait can be a useful attempt to complement existing portrait studies.

(2)A quantitative identification method for core researchers is proposed and an empirical study is conducted to identify core researchers in the field of artificial intelligence.

(3)Exploring the changes in the research topics of core researchers at both team and individual levels. Based on this, this paper uses a portrait visualization tool to construct a representative core researcher portrait to more clearly characterize the core researcher.

In addition, there are certain limitations of this paper, which are mainly as follows:

① Only a core researcher portrait model was constructed, but unstructured data such as personal attribute data and personal growth data were obtained manually through academic websites such as Microsoft Academic. The ideal approach would be to automatically extract this through machine learning methods to accommodate the broader and more complex core researcher representation requirements. In the future, further research can be carried out to design tools that can automatically generate portraits of core researchers. ②Considering the research purpose and indicator characteristics, only some indicators are selected from the core researcher portrait label system for identification of core researchers. The academic characteristics are limited to the SCI papers of scholars, not involving patent inventions, books and fund applications. In addition, some emerging indicators, such as Altmetrics generated in the Internet era can also be considered as indicators for identifying core researchers, and identification indicators can be added according to research purposes in the future.

Acknowledgements

The authors are grateful to anonymous referees and editors for their invaluable and insightful comments, and the support from the National Social Science Fund of China (16ZDA224)

Author contributions

Yuefen Wang: Conceived and designed the analysis; Wrote the paper.
Tianao Xu: Collected the data and analysis; Wrote the paper.
Yonghua Cen: Conceived and designed the analysis.
Xue Yang: Contributed data or analysis tools; Wrote the paper.

Declaration of Competing Interest

The authors report no declarations of interest

References

Bu, X. Y., & Xu, F. (2009). A research on the mentoring relationship & status of China´s modern mathematic elites. *Studies in Philosophy of Science and Technology, 26* (4), 102–107.

Chen, W. J., & Zheng, Y. (2013). Research on the measurement of the author´s influence in the scientific collaboration network. *Information Studies: Theory & Application, 2013* (06), 85–88+67.

Cortés, L. M., Mora–Valencia, A., & Perote, J. (2016). The productivity of top researchers: A semi–nonparametric approach. *Scientometrics, 109* (2), 891–915.

Costas, R., & Bordons, M. (2008). Is g–index better than h–index? an exploratory study at the individual level. *Scientometrics, 77* (2), 267–288.

Fan, X. Y. (2018). *Research on the recommendations of major project team members based on multi–source technology management data* [Master´s thesis, Xidian University]. CNKI.

Gao, Y., Chi, X. H., Zhang, C. Z., & Kong, J. (2019). Precise user profile for domain–special talents: A case study of intelligent manufacturing. *Library Tribune, 39* (6), 90–97.

Gao, Z., & Zhang, Z. Q. (2016). Review of quantitative evaluation method for individual academic influence. *In–*

formation Studies: Theory & Application, 39 (1), 133–138.

Griffin, D. J., Bolkan, S., Holmgren, J. L., & Tutzauer, F. (2016). Central journals and authors in communication using a publication network. *Scientometrics, 106* (1), 91–104.

Guo, Y., & Xiao, X. T. (2019). The evaluation of Chinese scholars′ influence by altmetrics index. *Information Studies: Theory & Application, 42* (4), 64–70.

He, J., Guan, P., & Wang, Y. F. (2019). Construction and evolution analysis of discipline knowledge network based on author–topic association. *Information Science, 2019* (1), 56–62+67.

Jiang, L., Dong, W. C., & Liu, X. G.(2018). Research on the growth law of top innovative academic talents –– Based on the quantitative analysis of the group of young "Yangtze River Scholars". *China University Teaching, 2018* (1), 87–91.

Jiang, X., & Ma, H. Q.(2015). Detecting the intellectual structure of archival science in China based on author co–citation analysis. *Archives Science Study, 2015* (1), 22–29.

Li, L., & Zhu, Q. H. (2008). An Empirical study of coauthorship analysis using social network analysis. *Information Science, 2008* (4), 549–555.

Li, S. K., & Yao, Y. P. (2009). Analysis on the growth process and characteristics of China´s top young geological talents. *Forum on Science and Technology in China, 2009* (1), 98–101.

Lin, G. Q., Mo, T.W., Ye, X. J., Han, C., Lin, Z. Z., & Wang, Z. K. (2018). Critical node identification of power networks based on TOPSIS and CRITIC methods. *High Voltage Engineering, 2018* (10), 3383–3389.

Liu, L., & Xu, X. (2021). Study on big data–related position portrait based on recruitment data. *Data Science and Informetrics, 1* (4), 14–25.

McIntosh, R. P. (1989). Citation classics of ecology. *The Quarterly Review of Biology, 64* (1), 31–49.

Ma, H. Q., & Lv, H. (2011). Bibliometric analysis and research of information science from 2000 to 2009. *Information Science, 29* (6), 867–873.

Ma, H. Q., & Jiang, X. (2018). Research on the structure of core authors´ knowledge exchange of archival science in China based on author cross–citation analysis. *Archives Science Study, 2018*(06), 19–24.

Qian, G. (2014). *Study on the mining model and efficacy of indicators of core patent* [Master´s thesis, Jiangsu University]. CNKI.

Qiu, J. P., & Zhou, C. L. (2008). The method of selecting high influence authors using published papers and h–index: A case study of the library and information field. *Library Tribune, 28* (6), 44–49.

Ren, H. J. (2013). Analysis of the core authors of archival science in China and their research themes. *Archives Science Bulletin, 2013* (4), 30–32.

Sun, H. S. (2012). Author keyword co–occurrence network analysis: An empirical research. *Journal of Intelligence, 31* (9), 63–67.

Wang, F., Guo, H. L., & Zhang, R. (2015). The characteristics of the development of Chinese scientific elite in humanities and social sciences: An analysis based on Chang Jiang scholars data. *Renmin University of China Education Journal, 2015* (4), 128–145.

Wang, F. F., Jia, C. R., Wang, X. H., Liu, J. W., Xu, S., Liu, Y., & Yang, C. Y. Y. (2019). Exploring all–author tripartite citation networks: A case study of gene editing. *Journal of Informetrics, 13* (3), 856–873.

Wang, Y. F., Yang, X., Yu, H. Q., & Cao, J. J. (2020). The collaboration pattern and comparative analysis of research teams in the artificial intelligence field. *Library and Information Service, 64* (20), 14–22.

Xie, J., Qian, L., Shi, H. B., Kong, B. B., & Hu, J. Y. (2019). Designing framework for precise service of scholarly big data. *Data Analysis and Knowledge Discovery, 3* (1), 63–71.

Yao, Z. L., Chen, H. L., & Xu, X. (2020). Research on classification and grading evaluation of scientific research talents. *Journal of Southwest Minzu University (Humanities and Social Science), 41* (6), 234–240.

Zhang, M. Y. (2019). *Research on predicting scientific Elites based on growth law* [Master´s thesis, Nanjing University]. CNKI.

Zhang, N., Wan, S. S., Wang, P. L., Zhang, P., & Wu, Q. (2018). A bibliometric analysis of highly cited papers in the field of economics and business based on the essential science indicators database. *Scientometrics, 116* (2), 1039–1051.

Zong, N. (2005). Probe into the growth law of excellent and top talents—From the higher education experience of academicians of the Chinese Academy and the Chinese Academy of Sciences who graduated from universities after the reform and opening up. *China Higher Education, 2005* (Z2) , 15–16.