Academic Genealogy Analysis Based on Knowledge Graph

Chen Yongyue, Li Ruijing

School of Management, Wuhan Textile University, Wuhan, China

Email address:
46258848@qq.com (Chen Yongyue), 937625369@qq.com (Li Ruijing)
*Corresponding author

To cite this article:

Received: December 14, 2020; Accepted: February 27, 2021; Published: March 3, 2021

Abstract: [Background] In the process of development of science, scientific theories are constantly improved and evolved, and technologies are constantly changed. The process of passing on and developing knowledge, technology, and culture from generation to generation of scientists and technologists, as well as the process of the new generation of scientists and technologists stepping on the shoulders of their predecessors and creating new knowledge, has had a positive and facilitating influence on the development of science. From this perspective, it is necessary and urgent to sort out disciplinary genealogical relationships, clarify the lineage of knowledge transmission, and study the academic genealogy of contemporary Chinese scientists.[Purpose/significance] Through the study of academic genealogical reproduction capacity, it is possible to understand the evolutionary process of knowledge transformation and innovation in the intergenerational transmission of knowledge, to explore the impact on the intergenerational transmission of knowledge, to promote knowledge transformation, and to cultivate modern scientific and technological talents. [Method/process] This paper applies the theories and methods of bibliometrics, social network analysis, and co-word analysis, a visualization technique is used to construct a model for analyzing the factors influencing the reproductive capacity of scientists' academic genealogy. [Results/conclusions] The analysis targets senior scholar Qiu Junping and his descendants, constructs a knowledge map of the academic genealogy of scientists, and discusses the factors influencing the reproductive capacity of academic genealogy.

Keywords: Knowledge Mapping, Bibliometrics, Academic Genealogy, Visualization, Reproductive Capacity

1. Introduction

The progress and development of human society is inseparable from the transmission of knowledge and innovation. In particular, the history of modern scientific development is a history of intergenerational inheritance, continuous exploration and innovation of scientific and technological workers. For example, from 1948-2017, the CAS selected 68 mathematical academicians, in combing the experience of Chinese mathematics academicians, 68 academicians were found to have inherited from the six masters [1]; Karnigar through a study of the "Shannon Brodie Axelrod Snyder-Porter" chain of mentorship that yielded significant results, found that the chain of mentorship has had a profound impact on the transmission and development of knowledge in American medicine [2].

And in the process of knowledge transmission and development, academic transmission is an important means of transmitting knowledge today, it has contributed to the rapid progress and development of human technology. And academic heritage relationships are linked together, formed an academic community of scholars and scientists from different generations, forming the academic genealogy [3]. Currently in universities, research institutes, enterprises and other organizations, existence of a large number of scientific families with academic heritage relationships, they promote the transmission and innovation of scientific knowledge and technology. But in the midst of the reproductive of academic genealogies, some branches are flourishing, some branches are getting stronger, some branches and offspring, on the other hand, are in decline as they develop. The ability of academic genealogy to reproduce is largely a reflection of the ability of knowledge to be transformed and passed on from one generation to the next. Therefore, the study of academic genealogical reproduction is important to promote knowledge transmission and innovation.

The reproductive ability on academic lineage research,
targeting scientists and groups of scientists, theory and methods through sociology of science, bibliometrics, using visualization analysis tools and techniques, examining the development of academic genealogy from multiple perspectives, through the internal structure of the academic genealogy, mechanisms of collaboration, academic background of scientists, and outcome indicators, analysis of factors influencing the strength and weakness of branches and offspring [4]. The study of academic genealogical reproductive capacity is of great academic and applied value. Through the reproductive ability on academic lineage research, the ability to understand the evolution of knowledge transformation and innovation in the intergenerational transmission of knowledge, exploring the factors influence on knowledge intergenerational transmission and the reproductive ability on academic lineage [5]. Also through academic genealogical research, ability to guide the healthy development of the academic genealogy, exploring the rules of technology talent growth, facilitating knowledge transfer and transmission, cultivating modern technology talents.

This paper applies the theories and methods of bibliometrics, social network analysis, and co-word analysis, a visualization technique was used to construct a model for analyzing factors influencing the reproductive capacity of academic genealogies; and the senior scholar Qiu Junping and his descendants are the targets of the analysis, building a knowledge map of the academic genealogy of scientists [6-7], exploring the factors influencing the reproductive capacity of academic genealogies.

2. Academic Genealogy Reproduction Capacity Visualization and Mapping Analysis Model Construction

Analytical research based on the literature on academic genealogy, in conjunction with the content of this study, in this paper, an academic genealogy reproduction visualization mapping analysis model is constructed (Figure 1).

![Figure 1. The reproductive ability on academic lineage visualization analysis model.](image)

The reproductive ability on academic lineage visualization analysis model consists of four main parts: collecting and organizing data, pre-processing of data, building knowledge mapping of academic genealogical reproduction capacity,
analyzing of factors affecting academic genealogical reproduction capacity.

(a) Collecting and organizing data: according to the scientists studied, collecting the data that related to the results of research and related to the children. The data sources include: journal database, degree thesis database, research project data (National Natural Science/Social Science Project search platform), websites of universities and research institutions academic staff introduction and baidu academic, etc. Faculty relationships are established through dissertations and baidu's network of academic partnerships, and identify intergenerational relationships; According to journal databases, research project platform query system and related unit website scholar introduction, collecting research from members of all generations, cited frequency, H-index, major research projects, affiliations, job title and other information.

(b) Pre-processing of data: according to the data collected, proceeding attribution and aggregation of information, removing non-academic results information, and according to the relationship between the teacher and the unit, removal of duplicate author information, constructing an overall academic genealogy data set; Identify intergenerational relationships among members, aggregating scientific information across generations according to intergenerational relationships, forming information data sets across generations.

(c) Building knowledge mapping of academic genealogical reproduction capacity: according to the processed information, using visual analysis algorithms, and constructing an academic genealogical reproduction capacity visualization map. It consists mainly of academic genealogy atlas, academic genealogical mapping of generations and knowledge mapping of the evolution of genealogical research themes. Analysis of the reproductive capacity of academic genealogies based on these maps, including reproductive algebra, number of Offspring, offspring academic skills, clade strengths and evolutionary paths of research topics.

(d) Analyzing of factors affecting academic genealogical reproduction capacity: offspring from the academic genealogy map, the strength of the clade relationship, combined with related literature research and the information collected about each scholar, the main components are: number of outcomes, academic status of members, years of research, frequency of outcome citation, work background, project topics etc, analyzing the main factors and correlations that affect the reproductive capacity of academic genealogies.

3. Research Subjects and Data Collection & Processing

This paper is based on the academic genealogy of Qiu Junping, a well-known professor in the field of library intelligence at Wuhan University, to conduct an analysis of factors influencing the reproductive capacity of academic genealogies. Research data from CNKI journal articles, data about baidu's academic authors, wanfang degree thesis database, national natural sciences/social science project search platform, based on staff profiles on academic websites of universities and research institutions etc.

Members with whom Professor Qiu Junping has collaborated according to the literature partnership search, the second generation of faculty members will be identified based on their academic thesis and other information, third-generation members are identified based on the second-generation member's collaborative literature, dissertations, and other information, finding information on intergenerational relationships, then according to the intergenerational membership, organize all members' journal literature information, frequency of outcome citation, H-index, major research projects, affiliations, title and other information. And screening, weighting, attribution, and aggregation of data, constructing data sets for analysis, includes overall genealogical dataset and data set for each child generation.

<table>
<thead>
<tr>
<th>Number of Scholars</th>
<th>Number of Documents</th>
<th>Number of Journals</th>
<th>Publication Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Generation</td>
<td>1</td>
<td>674</td>
<td>1986-2019</td>
</tr>
<tr>
<td>Second Generation</td>
<td>141</td>
<td>5394</td>
<td>1987-2019</td>
</tr>
<tr>
<td>Third Generation</td>
<td>782</td>
<td>11489</td>
<td>1990-2019</td>
</tr>
<tr>
<td>Summary</td>
<td>924</td>
<td>17557</td>
<td>1986-2019</td>
</tr>
</tbody>
</table>

A total of 17,557 articles were retrieved from this study, after removing invalid data such as non-research outcome data, a total of 8,310 journal articles and 924 genealogical member records were collected (all data retrieved in this study are as of July 2019), as shown in Table 1.


4.1. Building an Academic Genealogy Visualization Map

According to the data set, based on collaboration between authors, constructing an academic genealogical map of the entire clan with Professor Qiu Junping as the core, First-, second-, and third-generation academic genealogy visualization mapping (Figure 2, Figure 3, Figure 4, Figure 5).

The entire genealogical map formed by the visualization of the core scientific results among the authors with Qiu Junping as the core can be seen as follows [8-9], faster reproduction of the entire academic lineage, shown by a core subgroup, some branches are more productive in research, the third generation is more prosperous, while some branches are less able to reproduce, not even third generation (Figure 2). Dispersion analysis of the map based on the collaboration of scholars across generations (Figure 3, Figure 4, Figure 5), Qiu Junping's collaboration with second-generation scholars
is relatively close, while his collaboration with third-generation scholars is sparse and unproductive, demonstrates a lack of collaboration between first- and third-generation scholars in academic families. Second- and third-generation scholars have a complex collaborative connection, demonstrates close collaboration between third- and second-generation scholars, close collaboration between third-generation and second-generation scholars is an important point of academic transmission, second-generation scholars carry on the research themes of one generation of scholars and pass on their knowledge to the next generation of scholars, it has played an important role in the transfer of academic knowledge and the development of science and technology.

According to Figures 3 and 4, it can be seen that there is a certain amount of cooperation between the two generations, but not very frequently, depending on when the collaborative paper was published, the partnership is primarily during graduate school, collaboration decreases dramatically after graduation. According to the analysis of the third generation sub population in Figures 4 and 5, it was found that there is little cooperation between third generations derived from different second generations, but the cooperation between three generations of the same second generation is more frequent. Across-sectional comparison between generations shows that lineage in the process of reproduction, as the
offspring continue to develop, fewer and fewer inter-generational collaborations between different clades.

4.2. Analysis of the Evolutionary Path of Academic Genealogy Research Themes

Academic research evolves and even mutates in its subject matter as it is passed down through the lineage. To investigate whether academic genealogical reproduction has an influence on the evolutionary paths of scholars and their descendants in research topics, this article is based on the time variation in the occurrence of keywords and subject terms, analyzing the evolution of the research themes of Qiu Junping scholars and their descendants in different time periods (Figures 6, Figure 7). According to Figure 6, the evolution of the research themes of the first generation of scholars can be divided into three stages:

The first phase 1986-1997 is the embryonic stage of using bibliometrics to analyze intelligence literature, it is proposed that bibliometrics is a discipline that uses bibliometric methods to explore the dynamics of scientific change through the analysis of scientific literature [10], and to statistically analyze the evolution of the topic of intelligence and the current state of development of bibliometrics [11]; The second period 1998-2012 is a period of rapid development in metrological theory research, with the rise of the web, the shift from bibliometrics to informatics in the study of metrics, a number of methods for analyzing information in the statistical literature are proposed [12-16]; The third phase 2013-present, is mainly the development phase of bibliometric applications, many methods of applied analysis have emerged, such as: social network analysis, visualization analysis and research with knowledge mapping [17-20].

As shown in Figure 7, the keyword clustering algorithm yields the main research themes of scholars in the academic genealogy: #0 social network analysis, #1 digital library, #2 world-class universities, #3 education management, #4 logistics information, #5 knowledge transfer, #6 standard system, #7 patent information, #8 online academic information, #9 Internet, #10 intelligence work, #11 information industry, #12 university libraries, #13 college of western studies. At the same time, by topic word emergence algorithm and keyword occurrence frequency, it is possible to derive the distribution of research hotspots based on time series in the academic genealogy: the main thematic hotspots in 1986-1990 were bibliometrics, humanities and social sciences, and public libraries; the main thematic hotspots from 1998-2007 were knowledge management, digital libraries, information management, information resources, networking, evaluation, e-commerce, knowledge services, and librarianship; the main thematic hotspots for 2008-2019 are social network analysis, knowledge mapping, visualization, cloud computing, cluster analysis, co-occurrence analysis, and citation analysis [21-22].

5. Analysis of Factors Influencing the Reproductive Capacity of Academic Genealogy

5.1. Strength of Academic Genealogical Reproduction Ability Analysis

According to the analysis of the academic genealogy figures 2 to 5, we know that 141 second-generation scholars have collaborated with Qiu Junping, second-generation scholars with children include Ma Haiqun, Zhao Rongying, Xie Xinzhou, Wen Tingxiao, Huang Xiaobin, Yang Tianping, Wang Weijun, Fu Lihong, Sha Yongzhong, Duan Yufeng, Yang Siluo, Liu Huancheng, Tan Chunhui, Wang Feifei, Chen Yuan, Zhang Yang, Yang Ruixian, Ma Ruimin, and a total of 29 scholars, the total number of second-generation scholars without descendants is 112. Generally speaking, the overall research capacity of scholars in terms of the number of research results and the number of offspring reproduced reflects the strength of academic reproduction, according to a visual analysis of Professor Qiu Junping's family lineage, getting a more reproductive branch (Table 2).

Second-generation scholars in Haiqun Ma have collaborated with 112 scholars. Ma Haiqun scholars have the highest number of publications, highest number of research topics, highest number of descendants, second generation scholars with the highest number of publications by their children; the number of offspring of Xie Xinzhou scholars is 109, the number of papers published by their descendants is over two hundred; the number of offspring of Huang Xiaobin scholars is 59, but the number of papers published by his children's scholars is over three hundred. The phenomenon of a small number of offspring but strong research capacity is not only present in academic genealogies but also in...
5.2. Analysis of Factors Influencing the Reproductive Capacity of Academic Genealogy

To further investigate the factors influencing academic reproductive capacity in academic genealogy, the author's analysis of the academic genealogy of the clades with significant reproductive capacity is based on the number of scholarly research results, years of academic research, scholar's professional position, scholars work background, h-index, number of offspring, national funded projects and research hot topics and other multiple dimensions to analyze (Table 3).

Table 3. Top 15 scholars cited in the academic genealogy.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Name</th>
<th>H-index</th>
<th>Literature Citation</th>
<th>Number of Outcomes</th>
<th>Years of Academic Research</th>
<th>Work Unit</th>
<th>Number of country-level projects</th>
<th>Total number of projects</th>
<th>Number of Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ma Haijun</td>
<td>637</td>
<td>726</td>
<td>654</td>
<td>34 years</td>
<td>Wuhan University</td>
<td>6/1/1</td>
<td>8</td>
<td>141</td>
</tr>
<tr>
<td>2</td>
<td>Zhao Rongying</td>
<td>32</td>
<td>399</td>
<td>337</td>
<td>21 years</td>
<td>Heilongjiang University</td>
<td>2/0/1</td>
<td>3</td>
<td>112</td>
</tr>
<tr>
<td>3</td>
<td>Huang Xiaobin</td>
<td>25</td>
<td>223</td>
<td>215</td>
<td>17 years</td>
<td>Wuhan University</td>
<td>2/0/0</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>Xie Xinzhou</td>
<td>32</td>
<td>214</td>
<td>148</td>
<td>21 years</td>
<td>Zhongshan University</td>
<td>2/0/0</td>
<td>2</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>Wen Tingxiao</td>
<td>31</td>
<td>158</td>
<td>173</td>
<td>33 years</td>
<td>Beijing University</td>
<td>9/0/2</td>
<td>11</td>
<td>109</td>
</tr>
<tr>
<td>6</td>
<td>Yang Tianping</td>
<td>18</td>
<td>146</td>
<td>148</td>
<td>17 years</td>
<td>Central South University</td>
<td>1/0/1</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Wang Weijun</td>
<td>20</td>
<td>143</td>
<td>145</td>
<td>21 years</td>
<td>Zhejiang Normal University</td>
<td>3/0/0</td>
<td>3</td>
<td>111</td>
</tr>
</tbody>
</table>

5.2.1. Analysis of Factors Influencing the Reproductive Capacity of Academic Genealogy

The analysis of factors influencing the reproductive capacity of academic genealogy can be divided into several categories:

1. Research Background: Scholars' academic positions, years of academic research, and their professional backgrounds.
2. Scientific Output: The number of published papers, citations, and other research outputs.
3. Collaborations: The number of co-authors and the number of collaborative projects.
4. Research Funding: The amount of national and international funding received.
5. Academic Inheritance: The influence of the mentor and the academic lineage.
6. Tendency of Research: The trend of research topics over time.
7. Network Analysis: The number of connections and collaborations with other researchers and institutions.
8. Other Factors: The influence of other factors such as economic conditions, political environment, and cultural background.

These factors can significantly influence the reproductive capacity of academic genealogy. For example, scholars with higher h-indices and more collaborative projects tend to have higher reproductive capacity. Additionally, scholars working in institutions with more research funding and a strong academic lineage are more likely to have a higher reproductive capacity.
The ability of academic genealogy to reproduce can be characterized by academic ability and the number of reproducing offspring. There are many indicators that characterize the strength of academic ability, such as the H-index, number of outcomes, literature citations, etc. Because the H-index provides a more comprehensive and accurate reflection of an individual's academic achievement [23-25]. Thus, the text uses the H-index to characterize the strength of academic ability, such as the number of offspring. There are many indicators that characterize by academic ability and the number of outcomes, literature citations, etc. Because the H-index provides a more comprehensive and accurate reflection of an individual's academic achievement [23-25]. Thus, the text uses the H-index to characterize academic ability. According to the statistics in Table 3, the factors influencing the reproductive capacity of academic genealogies are discussed in the following aspects.

(1) Effect of number of outcomes on the reproductive capacity of offspring

In terms of the number of scholarly findings, the number of research results of first generation scholar Qiu Junping is 654, the H-index63 is the highest among scholars on the academic spectrum. Among the second-generation scholars, Haqun Ma, Xiaobin Huang, and Xinzhou Xie have an H-index >30, number of research results >140, number of descendants >50 scholars, the H-index <20 for Ma Ruimin, Liu Yong and Anlu, number of research results <100, number of offspring <30 scholars. Additionally, from the central node of the knowledge graph know that the larger the circle node, the higher the number of citations in the scholarly literature, the greater the number of scholarly findings, the greater the scholars' ability to reproduce academically.

(2) Effects of work background on the reproductive capacity of offspring

To study the effect of a scholar's work background on offspring fecundity, according to the institutions to which scholars in the genealogy belong, there are four categories: double first college, "211 project" institutions, provincial key universities, general public colleges, seven scholars are working in "dual first-class" universities, two scholars working at "211" institutions, four scholars are working in key provincial universities, two of the scholars work in general public universities. The number of research results of scholars from "Project 211" institutions is >20, higher than the H-index for general college working scholars. The number of offspring of scholars at key provincial institutions is roughly >25, the number of offspring of scholars is higher than the number of scholars at regular colleges and universities. Among them, three of the five scholars who participated in the National Foundation's key projects are working in "dual first-class" universities, that is, the background of a scholar's work has some influence on academic achievement and academic productivity. The overall strength of academic reproduction is Double first-class institutions > "211 Project" institutions > Provincial key institutions > Ordinary public institutions.

(3) Effects of the project on the reproductive capacity of offspring

According to the scholars' applications for national funding programs, scholars who have led key national foundation projects include Qiu Junping, Ma Haiqun, Xie Xinzhou, Wen Tingxiao and Yang Siluo, three of them are working at "dual first-class" institutions, two scholars working at key provincial institutions. The number of children of scholars working in "dual first-class" institutions and leading national key fund projects is more than 100 in general, the number of offspring of scholars working in "dual first-class" institutions and leading non-key national fund projects is less than 60, less than 15 scholars working in ordinary universities and leading general national projects. The number of offspring with 5 or more projects is much higher than the number of offspring with 2 projects, the greater the number of children of scholars, the greater the scholarly productivity of the scholars. The higher the category of the national project and the higher the number of projects that the scholar leads, which have a positive effect on the number of offspring reproduction and academic reproduction.

(4) Effects of study years on offspring reproductive capacity

Academic research is a cumulative process over time, generally speaking, the longer the years of study, the greater the number of descendants in the transmission process, the
more research outcomes you accumulate. According to Table 3, it can be seen that the first generation of Qiu Junping scholars has 34 years of research experience, and his descendants number more than 140 scholars, number of research outcomes are 654. In the second generation, the same scholars with longer research years, relatively large number of progeny and fruitfulness. While three generations of scholars have a relatively low number of years of research experience, fewer research results and fewer descendants. Certainly outcomes and the number of children and grandchildren are affected by other factors as well, so from the data, some of the study years caused less significant variability.

(5) Effects of research topics on offspring reproductive capacity

Hot topics can be characterized by the number of current researchers and the results of their research. The greater the number of scholars currently working on a topic or the greater the number of research outcomes, indicates that the topic is a current research hotspot. According to figures 6 and 7, it can be seen that the thematic hotspots and trends are very similar between the first and third generations of the overall study over time, reflects the strong influence of first-generation scholars on the research directions, hotspots, and trends of scholars in the academic genealogy, embodying the transmission and transformation of knowledge across generations. Additionally, according to a comparative analysis of the research themes of first-generation and second-generation reproducing scholars in Table 2, it was found that most of the second generation of scholars are working in similar subject areas as the first generation, only Yang Tianping's research differs greatly, it is mainly related to his pre-master's research theme, but his later papers incorporate a bibliometric approach, and knowledge is passed on and developed. Thus, the research directions and themes of previous generations of scholars correspond to the hot topics and trends of the academic field over time, and it has an effect on the reproductive capacity of the next generation.

6. Conclusion

Research on academic genealogical reproduction capacity, discovering the transmission and development of knowledge across generations, facilitating knowledge translation and innovation, exploring the rules for growing technology talent, it is the important theoretical and applied value for cultivating modern scientific and technical talents. This paper uses the technique of visual knowledge mapping, constructing an analytical model and method for visualizing academic genealogical reproduction abilities and selected representative scientists for a case study, through academic genealogy atlas, analyzing the relationship between the strengths and weaknesses of the branches in a genealogy and in terms of the number of results, work background, project topics, years of research, research themes, etc, exploring the main factors that affect the reproductive capacity of academic genealogies, it provides a reference for theoretical research on academic genealogy.

7. Outlook

Academic genealogy can reflect the formation and development of an academic heritage, and advanced academic heritage is the internal motivation of whether an academic genealogy can have lasting vitality and keep reproducing, and it is also the key to a country's long-term maintenance of scientific and technological competitiveness. Research on the academic genealogy of scientists not only has academic value, but also has important practical significance, as it can integrate team strength, curb academic misconduct, and promote the formation of a good scientific tradition.

References

[1] How important is academic heritage? See the 6 main academic genealogies of CAS academicians [EB/OL]. https://new.qq.com/omn/20190921/20190921A0HTGJ00.html ?pc


