Research on the Growth Patterns and Attribute Classification of Disciplinary Hotspot Concepts: A Case Study of Library and Information Science in China (Postprint)

Authors: Gao Jinsong, Han Muzhe

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Growth regularities constitute dynamic characteristics of information and knowledge. Through examining growth patterns, the attributes of knowledge elements can be revealed to a certain extent, which holds practical significance for research on knowledge organization and knowledge network construction. [Method/Process] Following a brief review of literature growth regularities, a dataset was constructed based on annual frequency data of hot concepts in library and information science over the past 25 years for curve fitting. Concurrently, decomposition analysis of the fitting results was performed, and a measurement index for growth curve morphology was subsequently established based on the properties of logistic functions, followed by empirical analysis. [Results/Conclusion] The effectiveness of logistic growth regularities in describing information or knowledge growth was verified through curve fitting and decomposition analysis. Empirical analysis demonstrates that the proposed measurement index exhibits good discriminative power for attributes of disciplinary hot concepts. Accordingly, hot concepts in library and information science can be divided by quadrant into three groups: long-term foundational concepts, short-term aging concepts, and developing hot concepts.

Full Text

Preamble

Journal of Library and Information Science, Vol. 63, No. 20, October 2019

Research on the Growth Law and Attribute Sorting of Discipline Hot Concepts: A Case Study of Library and Information Science in China

Gao Jinsong, Han Muzhe

School of Information Management, Central China Normal University, Wuhan 430079

Abstract

[**Purpose/Significance**] Growth law represents a dynamic characteristic of information and knowledge. Revealing the attributes of knowledge elements through growth patterns holds practical significance for knowledge organization and knowledge network construction research. [**Method/Process**] Following a brief review of literature growth laws, this study constructs a dataset based on annual frequency data of hot concepts in library and information science over the past 25 years, performs curve fitting, and conducts decomposition analysis on the fitting results. A growth curve morphology measurement indicator is then established according to the properties of logistic functions, followed by empirical analysis. [**Result/Conclusion**] Curve fitting and decomposition analysis verify the effectiveness of logistic growth law in describing information or knowledge growth. Empirical analysis demonstrates that the proposed measurement indicator effectively sorts the attributes of discipline hot concepts. Accordingly, hot concepts in library and information science can be divided into three groups by quadrant: long-term basic concepts, short-term aging concepts, and developing hot concepts.

Keywords: hot concepts; library and information science; growth law; logistic curve

The growth, aging, and dispersion laws of published literature are considered the most fundamental laws marking the development of scientific literature. The exploration and application of information and knowledge growth and accumulation laws have long been important topics of continuous concern in library and information science.

In current information science research, traditional information processing and organization based on literature units can hardly meet analytical demands in the new environment. Researchers tend to break through the constraints of literature units and analyze literature content at finer granularities and knowledge units at deeper levels. Consequently, traditional bibliometrics has gradually shifted toward informetrics and knowledge metrics. The methodology and theoretical system of knowledge metrics are built upon various fundamental laws of information science, most of which are derived from traditional bibliometric methods. However, whether these laws remain applicable when research objects are refined to content units or knowledge units, and whether their practical significance based on bibliometrics can continue to apply, are questions worth further exploration. For scientific literature, keywords, subject terms, and feature words are all vocabulary extracted from text content that represents hot concepts in related fields. These words and their combinations have thematic revealing functions, and their identification process can be viewed as a kind of manual labeling and extraction work for knowledge elements constituting knowledge units.

From a macro perspective of a discipline, the element set composed of these vocabulary terms constitutes the basic knowledge units of that discipline. Mastering the evolution of hot concepts during the development process of a discipline can help grasp the dynamic evolution process and trends of the discipline' s main body. Previous studies have mostly examined discipline development trends based on the temporal distribution of concept frequencies, with few verifying and analyzing the temporal statistical morphological laws from the perspective of hot concepts themselves. In fact, this has important theoretical significance for the application of traditional bibliometric laws at the knowledge level. Temporal growth accumulation laws represent the dynamic characteristics of literature information flow, and the growth accumulation laws of hot concepts also point to knowledge concentration and dispersion phenomena in target discipline fields over time. This study collects annual frequency data of hot concepts in library and information science over the past 25 years and conducts statistical and regression analysis on their growth accumulation laws, attempting to explore the applicability of traditional scientific literature growth laws at the knowledge level, establish sorting indicators based on the mathematical properties of growth laws, analyze the differences in growth morphology among different discipline hot concepts, and attempt their practical application in knowledge organization.

2 Research on Literature Growth Laws

Research on information growth laws was initially summarized from scientific literature growth laws. Since the 1940s, bibliometricians represented by D. Price have proposed a series of classical laws to describe the development processes and distribution patterns of literature information. Literature growth law forms the foundation for subsequent information growth law research, with the most representative being exponential growth law and logistic growth law.

2.1 Exponential Growth Law

The exponential function is a monotonically increasing unbounded function. Its most primitive practical significance is that under ideal conditions without restrictions, the number of individuals in a system will double after a period of time. D. Price discovered through statistical analysis of the growth characteristics of dozens of journals and papers that literature grows exponentially over time. According to calculations, literature in normal scientific fields increases exponentially, roughly doubling every 10-15 years with an average annual growth rate of about 5%-7%. Taking literature quantity as the vertical axis and time as the horizontal axis, plotting literature quantities from different years point by point on the coordinate system and fitting them with a smooth curve yields an exponential curve representing scientific literature growth over time, known as the "Price curve." The mathematical expression of the exponential growth function is:

$$f(t) = ae^{bt}, \quad a > 0, b > 0 \quad (1)$$

where t is the independent variable representing time, $f(t)$ is the dependent variable representing literature quantity at time t, a is a conditional constant representing literature quantity at the initial statistical moment, and b is a time constant representing the continuous growth rate.

From formula (1), a quantitative indicator measuring literature growth speed can be derived: the time period required for literature quantity in a certain field to double, denoted as T:

$$T = \frac{\ln 2}{b}, \quad b > 0 \quad (2)$$

Since the growth rate of scientific literature is uneven, the doubling time cycle T varies across different disciplines. Relevant statistics show that some disciplines double every few years, while others may take more than a decade.

2.2 Logistic Growth Law

The logistic function is a mathematical model proposed by Belgian mathematician P.F. Verhulst when studying the relationship between population growth and resource carrying capacity. Its curve shape is S-shaped, and it is called the logistic function due to its rich logical reasoning implications. The logistic growth model for scientific literature was proposed by Soviet scientists V. Nalimov and G. Vladuch after experimental research. They found that scientific literature growth occurs in stages, with different growth patterns in each stage. In the initial stage of scientific literature accumulation, it follows exponential growth law, but over time, its growth rate gradually slows down, and the total accumulation eventually approaches a limit value. Thus, the logistic growth law of scientific literature has a growth speed that is initially slow, then fast, then slow again, with its accumulation curve showing an S-shape. The conventional mathematical expression of scientific literature logistic growth curve is:

$$f(t) = \frac{K}{1 + ae^{-bt}}, \quad a > 0, b > 0 \quad (3)$$

where t is the independent variable representing time, $f(t)$ is the dependent variable representing literature quantity at time t, $K = \lim_{t\to+\infty} f(t)$ represents the expected total accumulation of literature in the target field, and a, b are positive parameters.

Logistic growth is a theory of limited growth. The reason why unlimited exponential growth cannot be maintained in reality is mainly due to limitations imposed by the system environment. Regarding environmental resistance limiting scientific literature growth, V. Nalimov summarized it as factors including material conditions, financial sources, and author intelligence. For longer statistical periods, logistic growth law provides a more accurate description of literature growth accumulation status than exponential growth law and can better combine with life cycle theory to explain its practical significance. However, Qiu Junping pointed out that although logistic growth law "overcomes the 'divergence' defect of exponential curves, it also has the limitation of being 'bounded'." In reality, the slowing growth rate of scientific and technical literature does not entirely mean that the pace of scientific development will decline. In relevant statistics on scientific and technical literature growth, there are also phenomena where literature accumulation breaks through again during the plateau period and accelerates accumulation once more, leading some scholars to propose a multi-stage fitted logistic growth model.

2.3 Other Research on Literature Growth Laws

The above two laws are basic models describing scientific literature growth. Over the past half-century, scholars both domestically and internationally have continuously attempted to verify and improve them or propose new growth models to more accurately explain scientific literature growth laws. Representative models proposed by foreign scholars include: hierarchical sliding exponential model, transcendental function model, and Shebu growth model. Models proposed by domestic scholars mainly include: variable exponential model, Han Dongmei model, knowledge growth dynamics model, modified exponential growth model, system theory model, and knowledge growth model considering environmental constraints. Each of these models has its rationality and defects. Yang Hong and Zhao Bendong once reviewed the advantages and disadvantages of several domestic literature growth models, revised the exponential growth model according to the multiplication principle, and ultimately deduced the logistic growth model, thereby demonstrating the unity of exponential growth law and logistic growth law. The article argues that exponential growth law is a special form of logistic growth curve, and the practical significance of the K value in the logistic growth function of scientific literature is the "size of the research object field." When revolutionary changes occur, the K value (denoted as B in the original text) is a scale parameter that can continue to increase. This reasonably explains the "boundedness" problem of logistic growth law. However, the deduction of these models is all based on scientific literature quantity as the research object. Once the limitation of literature units is broken and research objects are refined to content units or knowledge units, whether the logistic growth law based on literature remains valid and whether the understanding of its practical significance based on bibliometrics can continue to apply are questions worth further discussion.

3 Growth and Accumulation Laws of Hot Concepts in Library and Information Science

This study examines hot concepts in library and information science over the past 25 years, investigates their temporal growth and accumulation status, conducts horizontal comparative analysis based on statistical indicators, and attempts to grasp the development laws of library and information science in recent years by analyzing differences in the growth accumulation status of various hot concepts and combining them with semantic information of the concepts themselves.

3.1 Data Collection and Preprocessing

This study uses publicly compiled resources from the CNKI database as the data source. The collection objects are subject terms presented in the Chinese academic journal database. These subject terms are high-frequency words automatically selected and organized by CNKI from paper keywords that meet search criteria. These high-frequency words have good representativeness and directivity for paper themes, thus serving as subject grouping basis to facilitate users' thematic retrieval of literature. The system also provides occurrence frequencies of subject terms. Considering data source representativeness and data processing scale, the author limits the data collection scope to journals indexed in the Chinese Core Journals Database and CSSCI database. The Chinese Journal Full-text Database currently only supports data collection after 1992, while 2018 data is incomplete, so the search time range is set as 1992-2017.

Data collection and preprocessing work were carried out alternately through the following steps:

(1) **Annual subject term collection and organization.** In the Chinese journal database, set the search scope: Literature Classification Directory - Information Science & Technology - Library and Information Science & Digital Libraries; set the literature source category: Core Journals & CSSCI; select the "Subject"category in the group browsing options; retrieve data for each year from 1992 to 2017 (25 years total), collect subject terms year by year, obtaining 806 annual subject terms; organize the collected annual subject term tables in Excel, set conditional formatting, screen out 703 duplicate subject terms, and obtain 103 different representative subject terms.

(2) **Initial subject term cleaning.** Clean the above 103 different representative subject terms based on word meaning, mainly including: removing fuzzy subject terms such as "buildings" and "various economies" that are aggregated from large numbers of similar but different subject terms and lack specificity, making them unsuitable for analysis; removing synonyms, such as "reading promotion" and "reading promotion activities" which have similar meanings and roughly identical growth accumulation trends, retaining "reading promotion" and directly screening out its synonymous

subject term without merging frequency data. A total of 18 subject terms were processed, leaving 87 subject terms after cleaning.

(3) **Subject term-frequency matrix establishment.** Since the subject terms selected by the system differ in quantity and content across different years, the author uses the 87 subject terms obtained after initial cleaning as keywords (fuzzy) to retrieve their annual frequency data in the Chinese journal database. The search scope is core journals and CSSCI journal papers in the Information Science & Technology category, representing the number of papers published in the target year with relevant subject terms (fuzzy) as keywords. A frequency matrix is constructed with rows as years and columns as subject terms.

(4) **Secondary subject term cleaning.** Organize the subject term-frequency matrix, screen out non-continuous subject terms (such as "number of publications" with discontinuous annual frequency data containing multiple zero values that cannot be used for regression analysis), and screen out subject terms with insufficient data volume (such as "digital humanities" with very short continuous frequency data, belonging to emerging concepts that are difficult to judge trends through regression analysis). A total of 8 subject terms were processed.

Through the above data collection and preprocessing steps, 79 subject terms representing hot concepts in library and information science from 1992-2017 and their frequency information were obtained.

3.2 Growth Accumulation Curve Fitting

The reason for fitting growth accumulation curves rather than directly using temporal frequency data is that growth accumulation curves are monotonically increasing curves. Compared with temporal frequency curves that have single or multiple peaks, their curve morphology is simpler and the data analysis models are more mature.

The preprocessed data is a temporal frequency matrix of 79 hot concepts in library and information science from 1992-2017. Before curve fitting, further processing is required: set the starting year of each hot concept as t_0 and assign $t_0 = 1$; for the same hot concept, any subsequent time data $t_n = t_0 + n$.

Accumulate the temporal frequency data of each hot concept, record the statistical frequency of each hot concept in each year as y, with the initial accumulation value being the frequency value in the starting year, denoted as $f(t_0) = y_0$; thus, the accumulated frequency value of the same hot concept in year k is $f(t_n) = \sum_{i=0}^{k} y_i$, where $k \leq n$.

Import the processed time-accumulation value data into data analysis software GeoGebra for bivariate regression analysis, with time t as the independent variable and annual accumulation values $f(t)$ as the dependent variable. The software includes various common regression models. This experiment selects models based on the scatter plot $(t, f(t))$ of relevant hot concepts, with the selection criterion being the coefficient of determination $R^2 \geq 0.95$. If a concept' s growth accumulation curve has two or more optional models, the one with the highest R^2 is selected.

Fitting the 79 datasets individually using the above method shows that 61 groups fit logistic curves, 11 fit exponential curves, and 7 fit cubic polynomial curves.

The unity of exponential and logistic growth laws for scientific literature has been discussed above. In the early stage of a field with fewer research results, logistic functions can approximate exponential function forms. In other words, exponential growth law can be considered a special form of logistic growth law. How to interpret the cubic polynomial fitting results and whether they indicate new laws in literature or knowledge growth accumulation require further decomposition and analysis of the logistic growth model combined with specific hot concepts.

3.3 Decomposition of Logistic Growth Model

The logistic function is a monotonically increasing function with a "slow-fast-slow" three-stage growth trend. Differentiating the logistic function (formula 3) yields its velocity function:

$$v(t) = f'(t) = \frac{Kabe^{-bt}}{(1 + ae^{-bt})^2} \quad (4)$$

Taking the second derivative of the velocity function (formula 4) and setting it to zero yields three key points in the logistic function growth process: two inflection points and one extreme point of the velocity function.

The extreme point' s abscissa is:

$$t_H = \frac{\ln a}{b}$$

The inflection points' abscissas are:

$$t_1 = \frac{\ln a - 1.317}{b}, \quad t_2 = \frac{\ln a + 1.317}{b}$$

These three key points correspond to the initial prosperity period, peak period, and late prosperity period of socioeconomic growth reflected by the logistic function. The two inflection points divide the growth process into three stages: gradual growth period $(0, t_1)$, rapid growth period (t_1, t_2), and slow growth period $(t_2, +\infty)$. Meanwhile, $f(t_H) \approx 0.5K$, meaning when the growth accumulation value reaches half of the expected scale, the logistic function' s growth speed is fastest, after which the growth rate begins to gradually slow down.

Below is a decomposition and analysis of the logistic curve combined with the actual situation of the hot concept "library management." Statistics show that "library management" appeared 5,053 times over 25 years, was in its infancy in the early statistical period, and reached its research peak between years 10-15. The data was fitted once overall and twice in segments. The results are shown in Figure 1 [Figure 1: see original paper].

Figure 1 [Figure 1: see original paper] Fitting Results for Hot Concept "Library Management"

The figure contains one scatter plot and four curves. Curve L_1 is the overall fitting curve for "library management," with a logistic fitting result and $R^2 = 0.9969$, indicating excellent fit: $L_1 : f(t) = \frac{1366.66}{1+74.32e^{-0.28t}}, (1 \leq t \leq 25)$.

Curve L_2 is the segmented fitting curve for the first half of "library management," using data from the first 11 years (the gradual growth period), with an exponential fitting result and $R^2 = 0.9795$, indicating good fit: $L_2 : g(t) = 19.1e^{0.26t}, (1 \leq t \leq 11)$.

Curve L_3 is the segmented fitting curve for the second half of "library management," using data from years 11-25 (continuous with L_2' s interval), with a cubic polynomial fitting result and $R^2 = 0.9974$, indicating excellent fit: $L_3 : h(t) = -0.32t^3 + 14.66t^2 - 134.84t + 438.96, (11 \leq t \leq 25)$.

Curve L_4 is the velocity curve of L_1: $L_4 : v(t) = f'(t) = \frac{28439.65e^{-0.28t}}{(1+74.32e^{-0.28t})^2}, (1 \leq t \leq 25)$. The three vertical lines in the figure represent the coordinates of three key time points: $t_1 = 10.6835$, $t_2 = 20.0906$, $t_H = 15.3871$, corresponding to the two inflection points and one extreme point of L_4.

Figure 1 shows that the overall logistic curve fitted for "library management" coincides with the exponential curve in its gradual growth period, while the curve from rapid growth period to early slow growth period approximates the S-shaped portion between the two extreme points of the cubic polynomial curve.

3.4 Analysis of Logical Growth Mechanism of Discipline Hot Concepts

Further calculation shows that within the domain $[t_1, t_2]$, the value range corresponding to function $f(t)$ is approximately $(0.21K, 0.79K)$. During the entire rapid growth period of logistic growth, the frequency accumulation ratio can account for about 58% of the expected scale. Time point t_1 can be roughly regarded as the watershed between exponential growth morphology and logistic growth morphology. Once frequency accumulation reaches 21% of the expected scale, environmental resistance begins to exert significant influence, preventing frequency accumulation from maintaining exponential growth momentum and gradually transitioning to logistic growth morphology.

For the hot concepts analyzed in this study, this environmental resistance cannot be attributed to material conditions, financial sources, and author intelligence as in scientific literature growth. The presentation of these hot concepts is induced from group academic research results and is not limited to the material and economic conditions and intentional choices of certain authors or author groups. The growth accumulation trends of these hot concepts correspond to the concentration trends of relevant knowledge within corresponding time periods, directly reflecting the development status of related disciplines and their major problems to be solved. However, any discipline has its own development trajectory and life cycle. The problems faced by a discipline in different eras are targeted and staged. In the course of discipline development, many staged problems will be solved or replaced by more important problems. This process of continuously scaling and refining scientific research directly manifests as the continuous accumulation of knowledge. Therefore, for this study, the direct reason preventing these hot concepts from growing indefinitely is the decline or shift in research heat of their corresponding problems. The factors causing this heat decline or shift are relatively complex, mainly including the resolution, aging, and major problem transfer of relevant research topics due to developments in the technological environment and social environment.

According to statistics, among the 11 groups of hot concepts fitted as exponential curves, the annual frequency difference values are all positive, not yet reaching their peak by the statistical cutoff. This characteristic corresponds to the gradual growth period of logistic growth. For the 7 groups fitted as cubic polynomial curves, they already had high presentation frequencies in the early statistical period, indicating these hot concepts had entered rapid or slow growth periods at the statistical starting point. These 18 datasets failed to display the complete life cycle of their corresponding hot concepts within the statistical interval. However, based on the currently displayed portions, their fitting results can all be regarded as segmented forms of logistic curves. Therefore, the growth morphology of all 79 datasets can be described using logistic growth models. The growth accumulation of hot concepts in Chinese library and information science also follows logistic growth law. It can be further inferred that logistic growth law applies not only to describing bounded literature growth but also to describing unbounded information growth or knowledge growth.

4 Concept Attribute Sorting Based on Logistic Growth Law

The identification and sorting of hot concept attributes are the basis and prerequisite for classification and organization. According to different needs, attributes of discipline hot concepts can be examined from various angles. Specialized research on attributes of discipline hot concepts is relatively scarce both domestically and internationally. However, in studies analyzing discipline structure detection and development trends based on keywords or subject terms, classifying vocabulary according to certain attributes is an indispensable part. Vocabulary attribute research mainly focuses on the natural language processing field. In bibliometrics, besides sorting attributes based on word frequency, co-occurrence, and citation situations in corresponding literature units, some scholars have recently used semantic attributes of vocabulary itself as classification criteria, though related semantic classification methods are not yet mature. Classifying concepts through the above attribute sorting methods and then analyzing information contained in concepts themselves through static or dynamic methods can help grasp the structure and development trends of target disciplines to a certain extent. However, dynamic development trends themselves can also serve as an attribute of things. Research using trend attributes as classification criteria mainly focuses on comparisons between disciplines or fields.

Observing the logistic growth curves fitted for various hot concepts on the same coordinate plane reveals that different hot concepts exhibit both differences and similarities in growth accumulation morphology, following certain patterns. This section examines the growth accumulation morphology of the above hot concepts as the target attribute, establishes corresponding morphology measurement indicators based on the mathematical properties of logistic functions, attempts to sort the growth accumulation morphology attributes of hot concepts, and explains the practical significance of this attribute sorting combined with the concepts themselves.

4.1 Establishment of Logistic Growth Morphology Measurement Indicators

Hot concepts conforming to logistic growth law have growth speeds following a three-stage pattern of slow-fast-slow. The rapid growth period of their velocity function corresponds to the most prevalent stage in the entire life cycle of the corresponding concept. According to the above calculation, the frequency accumulation ratio during this popular stage can account for about 58% of the expected scale. Therefore, this stage can also be defined as the popular stage of the relevant hot concept. The time span of this popular stage is denoted as $\Delta t = t_2 - t_1 \approx \frac{2.634}{b}$.

According to statistics, the average growth trend of the above hot concepts in library and information science can also be described by logistic functions. The popular stage of these concepts has no practical significance but can serve as a ruler to measure the length level of individual concepts' popular stages. The average popular stage duration of hot concepts in library and information science is about 14.633 years. This value is affected by the expected scale K of each hot concept and cannot directly represent the average level of popular stage duration. After grouping and weighting the K values representing the accumulation scale of relevant hot concepts, the balanced duration is about $0.8 \cdot \Delta t$, approximately 11.706 years, indicating that the average duration of the popular stage of hot concepts in library and information science is about 12 years. For a single concept, if its popular stage duration Δt_n is greater than $0.8 \cdot \Delta t$, it is called a long-term concept; otherwise, it is called a short-term concept.

Besides horizontally comparing the relative length of popular stages across hot concepts, measuring the growth speed of the same hot concept itself can also reflect its growth accumulation characteristics. The velocity curve of logistic curves is a single-peaked curve starting from 0. When t is in the domain $(0, 2t_H)$, the velocity function curve is a normal curve symmetric about $t = t_H$. When $t = 2t_H$, the velocity function value approaches 0 infinitely, and the tangent slope of the corresponding logistic function also gradually approaches 0. Without considering the possibility of multi-stage growth, the logistic curve after this stage becomes a line infinitely close to $f(t) = K$, and the velocity curve shows a long-tail effect.

According to the above calculation, during the entire rapid growth period of logistic growth, the frequency accumulation ratio can account for about 58% of the expected scale. That is, a hot concept' s accumulated growth scale during its popular stage Δt is about $0.58K$. For any hot concept conforming to logistic growth law, the length of the entire growth period from its emergence to before entering the long-tail stage should be $2t_H$, and the accumulated growth scale during this entire stage approaches the expected scale K infinitely. This allows calculation of the average growth speed of relevant hot concepts during their popular stage, $v_h = \frac{0.58K}{\Delta t}$, and the overall average growth speed, $\bar{v} = \frac{K}{2t_H}$.

In the analyzed dataset, 18 groups of data yielded exponential and cubic polynomial curves. The decomposition of logistic curves above has clarified that these 18 groups correspond to hot concepts whose growth accumulation laws conform to logistic growth law. However, due to statistical interval limitations and development stage constraints, the curves fitted from current data cannot display the full picture of logistic growth. Theoretically, methods like grey prediction could supplement relevant data samples to derive their corresponding logistic curves. However, since many complex factors influence a hot concept' s growth accumulation, simply predicting and supplementing based on existing data may yield subjective results. Therefore, subsequent experiments and analyses only target the 61 groups that can be directly fitted as logistic curves.

Based on the above analysis, the author proposes using the difference between any hot concept' s popular stage time span and the balanced average popular duration as the horizontal coordinate, and the difference between any hot concept' s popular-to-average speed ratio and the speed ratio when the three stages have equal time spans as the vertical coordinate, to construct a Cartesian coordinate system marking each hot concept' s position. For any hot concept, its corresponding point coordinates are $(\Delta t_n/0.8\Delta t - 1, \bar{v}/v_h - 0.5747)$. Thus, the four quadrants of this coordinate system represent four different logistic growth morphologies (see Figure 2 [Figure 2: see original paper]).

- **First quadrant**: Popular stage time span is greater than the balanced mean and greater than gradual and slow growth periods, belonging to long-term concepts with relatively stable overall growth speed.
- **Second quadrant**: Popular stage time span is less than the balanced mean but greater than gradual and slow growth periods, belonging to

short-term concepts with relatively stable overall growth speed.

- **Third quadrant**: Popular stage time span is less than the balanced mean and less than gradual and slow growth periods, belonging to short-term concepts with obvious surge attributes.
- **Fourth quadrant**: Popular stage time span is greater than the balanced mean but less than gradual and slow growth periods, belonging to long-term concepts with obvious surge attributes.

4.2 Sorting and Analysis of Hot Concepts in Library and Information Science

To sort the attributes of each hot concept using the above logistic growth morphology measurement method, after calculating the coordinates corresponding to the 61 hot concepts, the author additionally counted their discipline affiliations and preliminarily divided them into three types: internal concepts of library and information science, and external environmental concepts and cross-disciplinary concepts. This classification is manually categorized based on the discipline distribution of relevant hot concepts. If a hot concept has the highest occurrence frequency in the "Library and Information Science & Digital Libraries" category, it is considered an internal concept of library and information science; otherwise, it is classified as an external concept. If its highest-frequency discipline also belongs to the "Information Science & Technology" category, it is classified as a cross-disciplinary concept; otherwise, it is classified as an environmental concept. Statistics show that among the 61 hot concepts, there are 38 internal concepts of library and information science, 16 cross-disciplinary concepts among external concepts, and 7 environmental concepts.

The discipline affiliations and related data of the above hot concepts are shown in Table 1 . The results of displaying each representative hot concept point in the Cartesian coordinate system and adding identifiers according to discipline affiliation are shown in Figure 3 [Figure 3: see original paper].

Figure 3 [Figure 3: see original paper] Distribution of Attribute Sorting Results for Hot Concepts in Library and Information Science

The corresponding points of the above hot concepts are mainly distributed in the first and third quadrants, with fewer hot concepts in the second quadrant and only one hot concept near the y-axis in the fourth quadrant that is not representative. Therefore, the growth accumulation morphology of hot concepts in library and information science mainly has three types. Below, the growth accumulation laws of hot concepts in the three quadrants are analyzed separately in combination with their word meanings, statistical characteristics, and discipline affiliations.

First quadrant corresponds to long-term concepts with stable growth speed. Long-term concepts in this dataset are mainly internal concepts of library and information science and environmental concepts among external concepts, with only one cross-disciplinary concept. Hot concepts distributed in the first quadrant maintain relatively high influence in the discipline field for a long time, maintaining relatively stable research heat over extended periods. Combined with word meaning analysis, internal concepts distributed in this quadrant are mostly foundational and global, including basic hot concepts such as the discipline's main body, research objects, and main carriers of intelligence information. Problems involving these foundational concepts have always existed with the discipline and are continuously updated with its development. External concepts distributed in the first quadrant are mainly environmental concepts corresponding to recent socioeconomic environmental hot issues, involving various aspects of social science. These hot issues themselves have epochal characteristics, wide influence, and long duration. Compared with the single discipline field of library and information science, the existence of these environmental concepts that change with socioeconomic environment development is universal and persistent. Since the only hot concept in the fourth quadrant is not representative, hot concepts in the first quadrant can be collectively called long-term concepts. These long-term concepts have relatively gentle growth speeds in the short term, generally large K values, and most still show stable growth trends at the statistical time point. Since curve equations obtained through fitting are limited by data sample size and statistical intervals, there is reason to suspect that some long-term concepts may have larger actual growth scales and longer popular stages.

Second quadrant corresponds to short-term concepts with stable growth speed. Hot concepts in this quadrant have generally shown decline in growth speed during the statistical interval, with some already entering the long-tail period of logistic growth. Hot concepts in this quadrant are mainly internal concepts and cross-disciplinary concepts among external concepts, with one environmental concept being "knowledge economy." Combined with statistical data, hot concepts in the second quadrant have aged to varying degrees, mainly in several situations: (1) Concept-directed information aging, such as traditional literature services based on print media gradually transforming into information services based on digital information and online data platforms, with "readers" as library service objects generalizing into "information users," resulting in the concept "reader services" being mentioned less frequently; (2) Concept terminology aging, such as early use of the transliteration "Internet" to refer to the internet, which gradually standardized to "internet," making "Internet" rarely mentioned; (3) Aging caused by external concept foundationization, such as "computers" and "databases" emerging in multiple disciplines around the turn of the century, making digital library construction a key research focus in library and information science. After more than a decade of development, computer science has become one of the main cross-disciplines of library and information science, with research and practice on databases and digital library construction gradually deepening and refining. Previous broad theoretical research has transformed into current foundational achievements, and these hot concepts have become discipline branches with more refined structures, with research on them gradually internalized. In summary, hot concepts in the second quadrant are mainly aging short-term concepts. Within the statistical period, the data used for fitting basically covers the entire life cycle of these hot concepts. By the end of the statistical period, the expected scale K values of hot concepts distributed in the second quadrant have generally stabilized.

Third quadrant corresponds to short-term concepts with obvious surge attributes. Combined with statistical data observation, although the growth speed of most hot concepts has peaked during the statistical period, they still maintain growth momentum at the statistical cutoff, with many new hot concepts added to the third quadrant. Some concepts emerged relatively late, with an average of only 21.24 years of data for fitting. Hot concepts distributed in the third quadrant mainly involve specific research fields and discipline branches of library and information science, including objects, methods, concepts, and technologies, as well as cross-disciplinary concepts with computer science, management, and other fields. These concepts have specific meanings and correspond to research hot issues discussed frequently in recent years. Some concepts emerged from the discipline's own theoretical and technological innovations, while others were borrowed from related or cross-disciplines. These concepts attracted attention from library and information scholars from their emergence, thus showing very fast growth speeds and obvious surge attributes during their popular stages. Some of these hot concepts still maintain high research heat today. Discipline themes related to third-quadrant concepts such as "digital library development," "development of library and information science journals,""five metrics studies,""library reading promotion," and "library knowledge services" frequently appear in recent national social science fund project topic guidelines. However, the emergence of more cutting-edge research hot issues also continuously updates research hot spots, with traditional concepts' research heat being continuously dispersed by emerging concepts and their temporal growth speed gradually slowing down. Some concepts will gradually become foundational and transform into long-term concepts, while others may be replaced by new concepts and tend to age. From statistical data, hot concepts distributed in the third quadrant generally have insufficient fitting sample sizes, and mainstream data statistical analysis software such as GeoGebra and SPSS fit logistic functions by inferring K values from final values and then obtaining growth functions through linearization operations. This fitting method is suitable for functions whose data samples cover the entire growth cycle but may produce deviations for functions still in the growth stage. This means that the growth functions currently fitted for third-quadrant hot concepts are not stable. As time passes and data sample sizes increase, their growth functions will change and corresponding statistical characteristics will also change. The third quadrant only represents a transitional state shown at the current time node. With the development and maturation of relevant hot concepts, they will eventually stabilize in the first or second quadrant. However, the transitional growth functions and corresponding measurement indicators are not meaningless. Their good sorting ability has important practical significance and application value for determining hot concepts' attributes and states.

This study reviews previous scholars' research on logistic and exponential growth laws of literature, constructs a dataset based on annual frequency data of hot concepts in library and information science over the past 25 years, verifies the effectiveness of logistic growth law in describing information or knowledge growth through curve fitting and decomposition analysis, establishes a growth curve morphology measurement indicator based on logistic function properties, and demonstrates through empirical analysis that it effectively sorts attributes and development states of discipline hot concepts. Accordingly, hot concepts are grouped into three categories: long-term basic concepts in the first quadrant, short-term aging concepts in the second quadrant, and developing hot concepts in the third quadrant.

The hot concepts analyzed in this study are representative elements constituting the discipline' s related knowledge. These hot concepts, together with other knowledge elements, combine through certain association relationships to form the discipline's scientific propositions and corresponding knowledge. This knowledge forms the discipline's knowledge network through association combinations, which is the mapping of the discipline' s objective knowledge structure in reality. The main problems faced by a discipline differ across development stages, and discipline hot concepts also continuously change and update across time periods. This diachronic metabolism of discipline hot concepts is precisely the driving force promoting continuous discipline development. Hot concepts in a discipline may have different attributes, and their own development is unbalanced. In the composition of discipline knowledge, combinations of different concepts may also follow certain attribute matching patterns, which may further reveal association relationships constituting knowledge and provide a breakthrough for further research on knowledge element grouping, organization, and knowledge network construction. Concepts at different development stages may combine in fixed matching forms, and sorting them chronologically can further grasp the evolution process and dynamic mechanisms of knowledge in a discipline field. This study' s verification of hot concepts' growth accumulation laws and the proposed attribute sorting indicators for discipline hot concepts pave the way for deeper knowledge organization research.

This study still has aspects worth further research and improvement. The study decomposes logistic growth functions, intuitively explains from the graph why exponential and cubic polynomial growth appear in fitting results, and demonstrates the unity of exponential and logistic growth. However, it does not attempt to write logistic growth functions for all hot concepts based on the relationships among the three growth functions, resulting in the abandonment of 18 datasets in subsequent experiments and analyses. This consideration is partly due to data accuracy concerns and partly limited by the single function fitting method. Additionally, using logistic fitting results to distinguish between growth and aging states of concepts is not entirely reasonable. The reason the proposed measurement indicator can distinguish concept growth states is that third-quadrant curve fitting results are generally not yet stable due to limited sample sizes, and the fitting results themselves have a certain transitional and staged nature. In other words, from an overall perspective, the time span of the popular stage being longer than that of the gradual or slow growth period may be a fixed attribute of mature hot concepts that can be fitted with fixed logistic curves. However, this hypothesis may require more samples for verification.

References

[1] Song Yanhui, Wang Xiaoping. Evolution from literature unit and information unit to knowledge unit[J]. Science Research Management, 2015(1): 459-464.

[2] Zhao Rongying, Wei Mingkun. Analysis on the development and evolution of "Five Metrics Studies" in China[J]. Modern Intelligence, 2017(6): 157-161, 169.

[3] Xiao Ming, Li Guojun, Yang Nan. Research hotspots in domestic information science based on word frequency analysis (1998-2007)[J]. Journal of Intelligence, 2009, 28(8): 21-25.

[4] Wang Lianxi, Cao Shujin. Comparative analysis of network public opinion research themes from an interdisciplinary perspective: Taking domestic library and information science and journalism and communication as examples[J]. Journal of the China Society for Scientific and Technical Information, 2017(2): 53-63.

[5] Zhou Xin, Jiang Xun, Chen Yuanyuan. Analysis of development trends in American information science research from the perspective of word frequency change rate model[J]. Information Science, 2017(4): 169-175.

[6] Wei Wei, Guo Chonghui, Chen Jingfeng. Text mining and social change research on State Council Government Work Reports (1954-2017)[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(4): 70-85.

[7] Qiu Junping. Informetrics (II) Lecture 2: Literature information growth law and application[J]. Information Studies: Theory & Application, 2000, 2(23): 153-157.

[8] Price D. Science since Babylon[M]. Translated by Ren Yuanbiao. Shijiazhuang: Hebei Science and Technology Press, 2010.

[9] Verhulst PF. Mathematical investigations on the law of population growth[M]//Annotated readings in the history of statistics. Berlin: Springer, 2001: 69-75.

[10] Wang Changan. Research on logistic growth model of scientific and technical journals in China[J]. Information Science, 2004, 22(5): 554-557.

[11] Rescher N. The limits of science[M]. Pittsburgh: University of Pittsburgh Press, 1999.

[12] Qiu Junping, Wang Hongxin. Hierarchical analysis of the development of bibliometrics in the 20th century[J]. Library and Information Service, 2000, 20(4): 1-8.

[13] Su Yi. Variable exponential literature exponential growth law[J]. Journal of Intelligence, 1996(5): 11-13.

[14] Han Dongmei. A new model describing the growth law of scientific and technical literature[J]. Information Theory and Practice, 1997(5): 280-282.

[15] Lu Guiwu, Han Dongmei. Dynamic model of scientific and technical literature growth[J]. Journal of the China Society for Scientific and Technical Information, 2000, 19(1): 90-96.

[16] Jing Peidong, Kang Zhongyuan. On the mathematical model of scientific and technical literature growth[J]. Information Science, 1997(S1): 353-356.

[17] Wang Hongxin. System theory unification of classical literature growth models[J]. Journal of Intelligence, 2003, 22(5): 2-3.

[18] Lin Hui, Lin Wei. Growth and aging laws of scientific literature and their new general models[J]. Journal of Intelligence, 2010, 29(4): 22-25.

[19] Yang Hong, Zhao Bendong. Deduction of logistic growth model for scientific and technical literature[J]. Journal of Capital Normal University (Natural Science Edition), 2011, 32(4): 18-23.

[20] Cui Dangqun. Analysis and goodness-of-fit test of logistic curve equation[J]. Mathematical Statistics and Management, 2005, 24(1): 112-115.

[21] Yang Hongjin. Research on method based on keyword inspection attribute indicators[J]. Statistical Research, 2013, 30(8): 10-16.

[22] Zhou Xin, Chen Yuanyuan. Analysis of discipline research development trends from the perspective of keyword frequency changes: Taking domestic information science research as an example[J]. Journal of Intelligence, 2016, 35(5): 133-140.

[23] Feng Guohe, Kong Yongxin, Xiao Jieqiong. A new method for analyzing field hotspots and trends based on weighted keywords[J]. Library and Information Service, 2018, 62(18): 102-109.

[24] Liu Xiaobo. Visualization research on hotspots and trends of library science research in China based on keyword co-occurrence and word frequency statistics[J]. Library and Information Service, 2012, 56(7): 62-68.

[25] Han Muzhe. Research on knowledge diffusion based on citation network[D]. Qufu: Qufu Normal University, 2016.

[26] Liu Ziqiang, Wang Xiaoyue, Bai Rujiang. Research on semantic classification-based discipline topic evolution analysis method: Taking big data research in Chinese library and information science as an example[J]. Library and Information Service, 2016, 50(15): 76-85.

[27] National Office for Philosophy and Social Sciences. National Social Science Fund project topic guidelines[EB/OL]. [2018-10-16]. www.npopss-cn.gov.cn.

Author Contributions:

Gao Jinsong: Paper guidance, writing and revision;
Han Muzhe: Data collection, experimentation and paper writing.

Note: Figure translations are in progress. See original paper for figures.