

Postprint: Research on Critical Time Points of Emerging Research Topics Along Evolutionary Trajectories

Authors: Xu Haiyun, Huiling Zhang, Wu Huawei, Liu Ziqiang

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] This study investigates the differential characterization of influence of emerging research topics by various critical time points. [Method/Process] First, we review current application scenarios and acquisition methods for inflection point time, and construct an identification method for inflection points of emerging research topics along the innovation evolution path based on network node growth mechanisms and characteristics in knowledge diffusion. Second, we conduct a comparative analysis of the differences among first occurrence time, average time, and inflection point time to explore the earliest time point at which emerging research topics generate influence. Finally, using stem cell research topics as an empirical domain, we analyze the different representational capabilities of various critical time points for the influence of emerging research topics. [Results/Conclusion] Inflection point time can identify influential topics earlier than average time. The significance of first occurrence time, average time, and inflection point time differs significantly in topic development trajectories; determining the temporal distribution of emerging research topics in innovation pathways requires synthesizing these three distinct types of critical time points.

Full Text

A Study on Key Time-points of Emerging Research Topics on Their Evolutionary Path

Xu Haiyun¹, Zhang Huiling², Wu Huawei³, Liu Ziqiang⁴

¹School of Management, Shandong University of Technology, Zibo 255000

²Taiyuan Library, Taiyuan 030024

³Archives of Northwest Normal University, Lanzhou 730070

⁴School of Journalism and Communication, Nanjing Normal University, Nanjing 210023

Abstract

[Purpose/Significance] This study explores how different key time-points characterize the influence of emerging research topics. **[Method/Process]** First, we review current application scenarios and acquisition methods for turning point time, and construct a turning point identification method for emerging research topics on innovation evolution paths based on the growth mechanism and characteristics of network nodes in knowledge diffusion. Second, we comparatively analyze the differences among first appearance time, average time, and turning point time to explore the earliest time-point when emerging research topics generate influence. Finally, using stem cell research as an empirical field, we analyze the different representational capabilities of these key time-points for emerging research topic influence. **[Result/Conclusion]** Turning point time can identify influential topics earlier than average time. First appearance time, average time, and turning point time have significantly different meanings in topic development paths, and determining the distribution time of emerging research topics on innovation paths requires synthesizing these three different types of key time-points.

Keywords: Evolutionary Path, Innovation Path, Turning Point Time, Emerging Research Topic, Curve Fitting

1 Introduction

Emerging research topic identification serves as an important means for predicting scientific and technological innovation research and indicates the direction of such innovation. Currently, major countries and regions focused on technological innovation not only fund emerging research but also support the use of scientometric methods to evaluate the effectiveness of funded projects. For example, since 2000, the European Research Council (ERC) has funded two projects: Development and Verification of a Bibliometric Model for the Identification of Frontier Research (DBF) and Emerging Research Areas and their Coverage by ERC-supported Projects (ERACEP) [1]. The ERACEP project analyzed emerging fields covered by ERC-supported research in 2009 [2]. The Intelligence Advanced Research Projects Activity (IARPA) established the Foresight and Understanding from Scientific Exposition (FUSE) program in 2011, with a mission to generate transformative innovation by focusing on “high-risk/high-reward” projects rather than pursuing short-term benefits, particularly emphasizing early identification of major emerging research topics [3].

Proactively identifying emerging research topics is crucial for science and technology managers to deploy innovation strategies and optimize resource allocation. Consequently, emerging research topics have remained important objects

of academic study, increasingly appearing in news publications and academic journals. Meanwhile, scientometrics has developed various methods to monitor emerging research topics in scientific and technological innovation [3], with evolutionary paths being one important approach.

The evolutionary path of emerging research topics refers to the developmental trajectory of emerging topics, reflecting the generation, diffusion, and evolution process of technological innovation. Citation analysis is a commonly used method for obtaining technology evolution paths. Domain topics can acquire evolutionary paths through clustering based on citation data similarity along a timeline [4]. Through evolutionary paths of emerging topics, we can deeply understand the context of topic evolution and the mechanisms of breakthrough innovation, aiding in identifying and predicting breakthrough innovations. Additionally, evolutionary paths can help assess characteristics of emerging research topics that are not easily quantified, such as novelty, growth/persistence, and coherence, through visualization to analyze uncertainty and ambiguity indicators [5].

An emerging research topic may appear multiple times on an innovation evolution path during its lifecycle and may emerge at different time intervals, eventually forming a temporal distribution of the topic. This temporal distribution is not uniform or equivalent; rather, there are a few important “key time-points.” Key time-points refer to critical moments in topic generation, development, evolution, and breakthrough. How do different types of time-points on evolutionary paths differ in marking the lifecycle of emerging research topics? How should different time-points be used to improve monitoring and trend-tracking capabilities for emerging research topics? Currently, the average time value of scientific literature cumulative counts is typically used as the key time-point in science and technology roadmaps. However, the trend of paper publication and patent application is not linear accumulation but follows an approximate exponential development pattern, presenting a skewed distribution. Therefore, selecting average time as the key time-point on topic evolution path timelines will lead to lagged topic identification. Meanwhile, the earliest publication or application year may only represent sporadic research that has not yet formed significant influence.

In view of this, we analyze the differences among three types of key time-points in characterizing the developmental influence of emerging topics, particularly focusing on turning point time, which plays an important role in early identification of emerging technologies. We also explore the most influential key time-point for emerging research topics to more clearly display key time-points on evolutionary paths. Based on this, we comparatively analyze the differences between turning point time and other key time-points (first appearance time and average time). This study aims to improve the accuracy of topic evolution paths to enhance their predictive accuracy for emerging research topic identification.

2 Literature Review

Analyzing the difference in how key time-points characterize emerging topic development requires understanding three related issues: the evolutionary characteristics of emerging research topics, the application of turning point time in different disciplines, and methods for acquiring turning point time. We review existing research from these three aspects to provide a foundation for this study.

2.1 Evolutionary Curve Characteristics of Emerging Research Topics

The formation of scientific literature dissemination networks exhibits random characteristics in their initial stage. During early network formation, some time periods may see no new nodes while others may see multiple new nodes simultaneously. The growth pattern of new network nodes is related to network structure, with an overall trend of fast then slow growth. As the dissemination network scale expands, the number of newly added network nodes gradually increases. However, when a discipline develops to a certain stage, limited by researchers' knowledge levels and domain information resources, the growth rate of the dissemination network will slow down, eventually reaching a relatively stable state.

P. Young [6] collected numerous time series representing growth behaviors, classified them according to data distribution characteristics, and then attempted to determine optimal growth curve prediction models for different distribution types, enabling selection of optimal models for specific datasets before fitting. Some studies focus on stage issues in technology growth prediction processes and corrections to industrial development curve models. D. Kucharavy et al. proposed distinctions between short-term, medium-term, and long-term predictions based on S-curve three-stage theory [7], emphasized the definition of growth parameters in Logistic growth models [8], and together with R. D. Guio [9] discussed applications of simple Logistic, component Logistic, and Logistic substitution models based on Logistic growth functions in technology change prediction, contributing to long-term forecasting of emerging technologies. P. Nguimkeu [10] developed a simple model selection test between Gompertz and Logistic growth models to avoid prediction errors from using growth curves. N. Meade et al. [11] considered diffusion innovation in different countries at different stages, introduced marketing variables into model parameters, and standardized models to achieve continuous diffusion.

Additionally, technology growth models in emerging technology fields have gradually matured. T. Daim et al. [12] combined bibliometrics and patent analysis with renowned technology forecasting tools such as scenario planning, growth curves, and analogies to predict three emerging technologies: fuel cells, food safety, and optical storage. M. Kyebanbe et al. [13] utilized patent citation data from the US Patent and Trademark Office to derive predictive factors using patent citation methods, achieving 70% prediction accuracy for emerging technologies at least one year before their emergence, thereby validating the ef-

iciency of their proposed supervised learning automatic data labeling method.

Thus, current research has extensively studied the evolutionary curve characteristics of emerging research topics, which exhibit multiple evolution patterns, leading to various fitting and prediction models for different curve features.

2.2 Application of Turning Point Time

The connotation of turning points differs across disciplines. From a philosophical perspective, turning points are key points for resolving contradictions and transforming things [14-15], representing the point where quantitative change leads to qualitative change and gradual change leads to sudden change. Mathematically, a turning point is defined as the point where a plane curve's bending direction changes. If a curve changes from convex to concave or vice versa at a point, that point is a turning point where the tangent line crosses the curve. In turning point calculation, if a curve's function has a second derivative of zero or non-existent at a point, and the second derivative has opposite signs on either side of that point, then that point is a turning point of the function [16-18].

In economic management, turning points are significant points in development curves that can forecast future uncertainty and warn of potential risks. Although specific meanings vary across fields, the general significance of turning points is widely recognized as the inflection point where data or things transition between growth and decline, prosperity and decay, with inconsistent evolution trends before and after [19].

Different connotations across fields manifest as different independent and dependent variables in curve fitting. In philosophy, independent variables are mostly ideology in superstructure, while dependent variables are social development. Mathematics focuses on formula models for turning point measurement. Economic management focuses on market and financial turning points for trend prediction, typically defining prices, output, and sales as independent variables, and economic development levels and industry development levels as dependent variables. Some experts have also applied turning points to public opinion prediction, enriching the application scope.

Representative studies include: In economic management, experts often use turning points to detect development trends in economic indicators. For example, the "Lewis turning point" [20-22] uses marginal productivity of agricultural labor to predict turning points and thereby judge labor supply trends. The "Kuznets turning point" [23-25] judges income distribution gaps based on the relationship between income distribution and economic growth. The "Mongers turning point" calculates the equilibrium point between marginal revenue and marginal cost [26].

During application, many experts have further enriched turning point types by combining domain characteristics with specific application scenarios. Shi Lukui et al. [27] defined upper and lower turning points in stock price range prediction

research using the time-price (y) relationship: an upper turning point satisfies $(y - y_1)/y > R$ and $(y - y_1)/y > R$, while a lower turning point satisfies $(y_1 - y)/y > R$ and $(y_1 - y)/y > R$, where R is a set threshold. Zheng Buqing et al. [19] divided turning points into stage turning points and extreme turning points in network public opinion prediction, with stage turning points dividing evolution into three stages: latent, outbreak, and decline.

Thus, turning point time has different connotations and applications across disciplines, though different fields generally agree that “turning point time can characterize important transition time-points for research objects.”

2.3 Turning Point Acquisition Methods Based on Curve Fitting

In recent years, methods such as Bayesian panel VAR models [28-30] and hidden Markov models [31-33] have been applied to turning point prediction. F. Canova et al. [34] proposed a turning point prediction method based on VAR Bayesian panels for output growth rate turning points in G7 countries. M. Billio et al. [35] combined VAR and Markov models to generate turning point prediction methods for business cycle analysis. L. Koskinen et al. [36] used hidden Markov models to predict turning points in Swedish and US economies. Subsequently, R. F. Peláez [37] proposed a Logit model that could more accurately predict business cycle turning points. N. Kulendran and K. K. F. Wong [38] combined Logit and Probit models to predict economic cycle turning points.

Additionally, ARIMA models [39-40] and chaotic time series analysis [41-43] in time series prediction models have been widely applied to turning point prediction. Economists found that turning point changes correlate with growth cycles, but time series models cannot significantly present growth trends and cycles. Therefore, growth function models [38, 44-45] are also commonly applied to turning point research. A typical application of Logistic curve analysis in prediction is M. King Hubbert’s analysis of US oil industry reserves and production. In 1962, Hubbert combined Logistic curves measuring domestic oil cumulative proven reserves and cumulative production to predict the 1968 peak in US oil production, two years ahead of the actual peak [46]. Since then, scholars have begun using Logistic curves for turning point analysis. Zhan Xinyu used Logistic curve equations to predict the turning point of Guangxi’s industrial structure serviceization, finding that growth curve models provide high accuracy in turning point prediction [45].

With deepening knowledge connections, grey prediction [47-49] and neural networks [50-52] have opened new directions for turning point prediction methods. Tian Hongli et al. [47] used similar grey models to predict stock price turning points, solving the problem of low grey prediction model accuracy in oscillating sequences. O. Claveria et al. [50] compared three different artificial neural network technologies (multilayer perceptron networks, radial basis function networks, and Elman networks) for tourist demand turning point prediction, finding that multilayer perceptron networks have higher accuracy in complex

predictions, and that increasing dimensions can improve turning point prediction accuracy.

Since products and services have lifecycles including germination, development, maturity, and decline, determining a product's position in its lifecycle is another method for predicting turning points [53]. P. C. Guo et al. [54] found based on lifecycle theory that when the logistics index exceeds equilibrium value, the logistics industry produces an amplification effect on the economy and welcomes its development turning point. P. N. Golden et al. [55] found that sharp growth in some product sales curves mostly occurs during the transition from new product introduction to growth stages, calling this phenomenon "Take Off" and the transition point the "take-off point."

Different turning point identification and prediction models all aim to improve turning point prediction accuracy in specific fields. Because different fields and even different objects in the same field have different lifecycle curves, no single method is optimal—specific analysis based on the data distribution characteristics of the field is required.

3 Key Time-points on Innovation Paths and Acquisition Methods

3.1 Turning Point Time of Topic Distribution

Emerging research topics are persistent topics on innovation paths. During this development process, the distribution of emerging research topics on innovation paths is consistent with scientific discovery and technological innovation patterns, not uniformly distributed on the timeline. Just as breakthroughs in scientific and technological innovation often present cluster characteristics, topics on innovation paths show various patterns in their outputs: gradually increasing, gradually decreasing, rising then falling, falling then rising, and other combinations. Therefore, how to display the dynamic changes of topics on the timeline deserves in-depth analysis.

We use the number of scientific documents contained in a topic as an indicator of attention and influence on that topic. Typically, when significant progress is made or breakthroughs occur, the number of papers increases substantially. We attempt to use growth function (Logistic Function) curves for scatter plot curve fitting, but the results are unsatisfactory. Through experimentation, we find that a cubic function can better fit the scatter distribution in this study; therefore, we adopt cubic function fitting for scatter plots, plotting function curves on the graph and outputting the fitted function.

The method for acquiring turning points on innovation paths is as follows:

First, extract sample data from raw data files, count the number of papers (y) published each year (x) for each topic, and draw scatter plots for each topic's data.

Second, select fitting functions to fit the scatter plots from the first step, visualize the fitted curves, and output the fitted functions.

Third, calculate turning point values based on the fitted curves and plot turning point positions on the graph.

3.2 Analysis of Turning Point Types

Based on the interval range where turning point time appears, we classify turning point types into four categories: turning points appearing within the analysis time interval, before the analysis time interval, after the analysis time interval, and not appearing within the analysis time interval.

3.2.4 Comparative Analysis of Three Key Time-points To further confirm the effectiveness of turning points in predicting topic development trends, we compare the differences between turning points in analyzing emerging versus non-emerging research topics, and distinguish the functions of first appearance time, average time, and turning point time in topic prediction analysis. The calculation formulas for turning point, average time, and earliest time are shown in Table 1 .

Table 1 Formulas for Turning Point, Average Time, and Earliest Time

Time Type	Formula	Meaning
First Appearance Time	$t = t, i = 1$	Indicates the earliest starting time
Average Time	$t = (\sum t) / n$	Indicates the central position where time points are relatively concentrated
Turning Point Time	For $f(x) = ax^3 + bx^2 + cx + d$, any cubic function has a turning point when $a \neq 0$	Indicates when the topic first has significant influence

Where t is time, n is the total number of time points, i is the starting time point, and a, b, c, d are parameters.

4 Empirical Analysis

4.1 Dataset and Statistical Analysis

Stem cells are a class of cells with self-renewal and multi-directional differentiation capabilities, representing important research objects in biomedicine. Due to their significant value in disease treatment and regenerative medicine and their enormous development prospects, stem cells have attracted attention in

life science and medical research. We select the stem cell field as the empirical domain for this method, using Web of Science (WoS) to obtain analysis data. The retrieval date was October 20, 2018, yielding 422,101 research papers related to stem cell research.

The specific retrieval strategy was: $TS = ((\text{“stem cells” OR “stem cell”}) \text{ NOT } (\text{“stem cellulose” or “stem Cellular” or “cello” or “cellar” or “cellphone”})) \text{ OR } TS = ((\text{ESCoriPSoriPGCorMSCorCSCorLSCorTSCorADSCorHSC}) \text{ near } (\text{cells OR cell})) \text{ OR } TS = ((\text{“totipotent” or “pluripotent” or “multipotent” or “unipotent” or “progenitor” or “precursor”}) \text{ NOT } (\text{“non-pluripotent”})) \text{ ADJ } (\text{cells OR cell})) \text{ NOT } TS = (\text{“fuel cell” or “inplane switching” or “Intrusion Prevention System”})$.

Community identification is a common method for analyzing large complex network structures. This paper identifies meaningful research topics through citation networks. V. A. Traag et al. developed the Leiden algorithm [58], which can discover better community divisions, improve effective connections between communities, and run faster. Under the Leiden clustering algorithm, all papers in the WoS database are divided into three levels: 22 topics at the macro level, 868 topics at the meso level, and 4,047 topics at the micro level. We adopt the micro-level topic classification based on the Leiden algorithm and further analyze emerging research topics on this basis.

To avoid the impact of random fluctuations such as database expansion or contraction, we use smoothed annual publication volumes, selecting 2004-2018 as the research period and setting 11 time slices with every 5 adjacent years as a group. Within the research period, the number of published papers shows exponential growth (see Figure 3 [Figure 3: see original paper]). The micro-level topic classification of 422,101 records involves changes across different time slices; for example, there were 1,584 micro-level topics in 2004-2008 and 2,058 micro-level topics in 2013-2017 [58] (see Figure 3).

4.2 Topic Identification and Visualization Analysis

The distribution of stem cell research topics is shown in Figure 4 [Figure 4: see original paper]. Node size is proportional to the number of papers on stem cell research; larger nodes indicate more papers on that topic. Smaller distances between topics indicate more similar research content. The numbers on nodes in Figure 4 are not the numbers of papers involved but the micro-level classification label numbers of that topic in the entire WoS database. Stem cell research-related topics are mainly located in the “biological health” area and involve many topics. Although the Leiden classification algorithm [58] is built on the entire WoS database and helps more accurately discover global topics, it lacks domain specificity. To obtain more targeted stem cell topics, we further obtained the proportion of stem cell papers in each topic under the Leiden classification and deleted topics containing only a small number of stem cell papers.

4.3 Turning Point Characteristics Analysis of Emerging Research Topics

Using a cubic function, we determine turning points through the value of the third derivative. If the third derivative of $f(x)$ is not zero, then that point is a turning point. Considering that topics in early development stages may also have turning points outside the research period, this study does not impose threshold restrictions on turning point values, such as limiting them within the data analysis range. If a cluster has fewer than 4 samples, function fitting cannot be performed, and turning points cannot be calculated based on data model fitting. After obtaining turning point data, we must further analyze whether the results are all reasonable.

4.3.1 Classification of Emerging Research Topic Turning Point Types

(1) Turning points appearing within the analysis time interval. Nineteen emerging research topics have turning points within the analysis time interval: #2, #60, #107, #142, #254, #353, #461, #581, #648, #710, #814, #867, #921, #965, #1046, #1095, #1142, #1199, #1460. These topics established the earliest influential time-points within the analysis period. Using topics #581 and #107 as examples, we plot evolution curves with turning points; the circled dots pointed to by arrows are turning point times (see Figure 5 [Figure 5: see original paper]). The later the turning point appears within the time period, the more likely the topic is an emerging research topic. Topic #581 shows a turning point around 2009, while #107 shows one around 2014.

(2) Turning points appearing before the analysis time interval. Two emerging research topics have turning points before the time interval: #161 and #221 (see Figure 6 [Figure 6: see original paper]). Combined with domain expert analysis and fitted curve analysis, we find that prediction errors for this type stem from the assumption of using a cubic function model. In fact, these two topics are still in early development stages and have not yet reached their turning point time.

(3) Turning points appearing after the analysis time interval. Two emerging research topics have turning points after the time interval: #727 and #469 (see Figure 7 [Figure 7: see original paper]), indicating that turning points can be expected to appear in the future. However, we find that the future turning point time prediction for topic #469 may have problems. Combined with domain expert analysis and data characteristics, topic #469 is indeed in a rapid development stage, but the publication volume at the turning point may exceed predicted results. Therefore, predictions of turning points appearing after the analysis period may also have errors, but turning point predictions can still reflect that the topic is currently in a development stage before the turning point appears.

(4) No turning point appearing in the analysis time interval. Three topics could not obtain turning point times using this method: #1290, #2276,

and #261. The evolution curves for #1290 and #2276 are shown in Figure 8 [Figure 8: see original paper]. The main reason for no turning point appearing is that these topics are still in early development stages, making turning point values impossible to determine.

4.3.2 Comparison of Turning Point Time and Average Time for Emerging Research Topics

The comparison of turning point and average time for 26 emerging research topics demonstrates that turning point time can identify influential topics earlier than average time. Column 5, “Time Difference_{INF}-AVG,” is the difference between “Average Time_{AVG}” and “Turning Point Time_{INF}.” A positive “Time Difference_{INF}-AVG” value indicates that “Turning Point Time_{INF}” is earlier than “Average Time_{AVG}”; a zero value indicates they are equal; a negative value indicates “Turning Point Time_{INF}” is later than “Average Time_{AVG}.” Based on the understanding that turning point values indicate when a topic first has significant influence, for emerging research topics, a positive “Time Difference_{INF}-AVG” indicates that turning point time can identify the topic’s influence earlier than average time within the statistical period; a negative value indicates the topic has not yet reached a stage of significant influence.

From the “Time Difference_{INF}-AVG” column, excluding topics #261, #1290, and #2276 without turning point times, among the remaining 23 topics, 6 have negative “Time Difference_{INF}-AVG” values, 3 have zero values, and 17 have positive values. This demonstrates that in stem cell research, turning point time can identify influential topics earlier than average time. Two emerging research topics have differences greater than 5 years: topics #221 and #161, both with turning points before 1998, which upon verification are data fitting errors. These two topics are actually still in early development stages.

4.3.3 Classification of Non-Emerging Research Topic Turning Point Types

Consistent with emerging research topics, we also classify turning point types for 28 non-emerging research topics into four categories. In this empirical analysis, for non-emerging research topics, there are no cases of turning points appearing after the analysis time interval, which relates to the characteristics of non-emerging research topics themselves, as they may be topics with slow recent growth during the screening process.

(1) Turning points appearing within the analysis time interval. Twenty-two non-emerging research topics have turning points within the analysis period. Using topics #899 and #23 as examples (see Figure 9 [Figure 9: see original paper]), these topics established the earliest influential nodes within the analysis period. The later the turning point appears within the period, the more the topic is an emerging research topic. #899 shows a turning point around 2002, while #23 shows one around 2014.

(2) Turning points appearing before the time interval. Given that stem

cell research began in the 1990s, most research topics are still in emerging fields, and early signals are insufficient, making it reasonable for turning points to be outside the data. In this study, four non-emerging research topics have turning points before the analysis period: #203, #285, #1623, and #1638.

4.3.4 Turning Point Effectiveness for Emerging vs. Non-Emerging Research Topics

Through comparison, we find that turning points are more effective for analyzing emerging research topics, with fewer data fitting errors. This is mainly because emerging research topics show continuous growth trends, and cubic function fitting effects are more ideal. For non-emerging research topics, their development trend patterns may include more modes beyond stable or rapid growth over a period, such as stabilization or decline. Therefore, for non-emerging research topics, turning point time distribution is more uncertain, and prediction function error probability is higher. Overall, turning points provide more stable identification effects for emerging research topics.

4.4 Comparison and Application Strategy of Three Key Time-points

The advantages and disadvantages of first appearance time, average time, and turning point time distributions are shown in Table 5 . While turning point time is important, relying solely on turning point values cannot fully achieve topic prediction analysis. Research topics appear in different time periods of roadmaps, and the meanings of different time-points differ significantly. Determining topic distribution time in innovation paths requires synthesizing three time types. First appearance is the earliest signal of a topic, suitable for quantitative analysis but difficult to attract attention due to scarce data. When a topic reaches its turning point, vigilance is needed to consider early deployment. Average time can characterize whether a topic has become a research hotspot.

Table 5 Advantages and Disadvantages of Three Time Distributions

Time Type	Advantages	Disadvantages
First Appearance Time	Inevitably exists and is easy to obtain; is the earliest signal of topic appearance	Insufficient influence, difficult to attract attention
Average Time	Inevitably exists and is easy to obtain; can represent topic influence from quantitative characteristics	Since topic paper publication generally follows exponential growth patterns, average time tends to lag, making it difficult to assess this influence

Time Type	Advantages	Disadvantages
Turning Point Time	Can better characterize changes in topic development influence. If turning point time is earlier than average time, it can predict topic influence earlier; if later, it indicates the topic is still in early development stage	May not exist—either from the data itself or from data fitting errors. Different fitting functions may also have prediction errors

This study focuses on analyzing differences in topic development across different time characteristics, especially the turning point time of growth curves, which plays an important role in early identification of emerging research topics. We comparatively analyzed the differences among first appearance time, average time, and turning point time, exploring the earliest time-point when emerging research topics have influence. This method applies not only to publication volume but also to other indicators reflecting topic growth curve turning point analysis, such as citation curve turning point analysis. Using stem cell research topics as an empirical case, we obtained innovation development evolution paths for identified emerging research topics through curve fitting and calculated turning points in the research development path. The research results help improve the accuracy of topic evolution paths and enhance their predictive accuracy for forward-looking judgment and layout of emerging research topics.

The main conclusions are as follows: First, we classify turning point types into four categories based on the interval range where turning point time appears: within the analysis interval, before the analysis interval, after the analysis interval, and not appearing in the analysis interval. Through turning point characteristic analysis of emerging and non-emerging research topics, most turning points appear within the analysis time interval, indicating that turning points have good utility for predicting emerging research topics. Second, comparative analysis of turning point time and average time for emerging versus non-emerging research topics shows that turning point time can identify influential topics earlier than average time. Third, comparing turning point effectiveness between emerging and non-emerging research topics reveals that turning points are more stable and effective for emerging research topic analysis with fewer fitting errors; for non-emerging research topics, turning point time distribution is more uncertain with higher prediction error probability. Finally, first appearance time, average time, and turning point time have significantly different meanings in topic development paths, and determining topic distribution time in innovation paths requires comprehensive application of all three time types.

This study has several limitations. First, we cannot provide rigorous justification for the assumption of using a cubic function, focusing instead on finding evidence from scientometrics and analyzing meaning and application value from scientometric practice. This study has not considered other data distribution types, and not all curves follow power function patterns, which is the main cause of turning point identification errors. Second, cubic functions cannot effectively identify situations with multiple turning points, and logistic growth curves are not suitable for all topic data fitting, leading to some degree of distortion in measurement results. The reason for these situations is that each topic has its own development characteristics. Future research will attempt to use multiple indicators combined with specific domain data characteristics for analysis. Third, this study's data source is primarily scientific papers, which mainly reflect basic research results in a field, while patents can reflect applied development trends. Current technology field development often involves coordinated progress in basic and applied research, so a single literature type cannot comprehensively measure field development trends. Future research will attempt to combine multi-source data types such as patent data and industrial economic data to form more comprehensive curve simulations for field innovation development and more accurate turning point identification. Additionally, we will attempt to intelligently fit scatter points through an adaptive process to identify the optimal fitting function for each topic distribution curve. We will also collect full-process data for specific fields, especially during topic germination stages, to achieve more accurate curve fitting and turning point identification through these measures.

References

- [1] European Research Council. Identification of “Frontier Research” and “Emerging Research Areas” in research proposals: a bibliometric approach [EB/OL]. [2020-02-10]. <https://erc.europa.eu/event/identification-frontier-research-and-emerging-research-areas-research-proposals-bibliometric>.
- [2] REIS S T, VIGNOLA-GAGNÉ E, KUKK P, et al. ERACEP-emerging research areas and their coverage by ERC-supported projects [R]. Belgium: Technical Report European Research Council, 2013.
- [3] Foresight and Understanding from Scientific Exposition (FUSE) [EB/OL]. [2020-02-10]. <https://www.iarpa.gov/index.php/research-programs/fuse>.
- [4] WEI L, XU H Y, HU Z Y, et al. Multiple-pattern analysis and prediction of topic evolution path based on topic correlation: a case study of information science research [J]. *Library and information service*, 2016, 60(3): 71-81.
- [5] XU, H Y, YUE Z H, LUO R, et al. A study on the multidimensional scientometric indicators to detect the emerging topics [C]//Proceedings of the 17th international conference on scientometrics and informetrics (ISSI 2019). Italy: Sapienza University of Rome, 2019: 2518-2519.

- [6] YOUNG P. Technological growth curves: a competition of forecasting models [J]. *Technological forecasting & social change*, 1993, 44(4): 375-389.
- [7] KUCHARAVY D, GUIO R D. Problems of forecast [J]. *ETRIA TRIZ future*, 2005: 219-235.
- [8] KUCHARAVY D, GUIO R D. Application of S-shaped curves [J]. *Procedia engineering*, 2011, 9: 559-572.
- [9] KUCHARAVY D, GUIO R D. Logistic substitution model and technological forecasting [J]. *Procedia engineering*, 2011, 9: 402-416.
- [10] NGUIMKEU P. A simple selection test between the Gompertz and Logistic growth models [J]. *Technological forecasting & social change*, 2014, 88(Oct.): 98-105.
- [11] MEADE N, ISLAM T. Modelling and forecasting the diffusion of innovation-a 25-year review [J]. *International journal of forecasting*, 2006, 22(3): 529-545.
- [12] DAIM T U, RUEDA G, MARTIN H, et al. Forecasting emerging technologies: use of bibliometrics and patent analysis [J]. *Technological forecasting and social change*, 2006, 73(8): 981-1012.
- [13] KYEBANBE M N, CHENG G, HUANG Y Q, et al. Forecasting emerging technologies: a supervised learning approach through patent analysis [J]. *Technological forecasting and social change*, 2017, 125: 236-244.
- [14] SHAO S. Uncertainty, economic forecasting and turning points [J]. *Financial expo*, 2019(4): 31-33.
- [15] QIU G T. The “turning point” of development: analysis from the perspective of scientific development concept [J]. *Study & exploration*, 2009(3): 26-31.
- [16] ZHANG X S. “Turning point” in economics and “turning point” in mathematics [J]. *Higher mathematics studies*, 2012, 15(5): 29-31.
- [17] XIE Y Y. Also on the definition of turning point [J]. *Higher mathematics studies*, 2003, 6(3): 11.
- [18] WASOW W. *Linear turning point theory* [M]. New York: Springer-Verlag, 1985.
- [19] ZHENG B Q, ZOU H X, HU X J. Research on network public opinion prediction based on turning points [J]. *Computer science*, 2018, 45(S2): 539-541, 575.
- [20] CAI F. Demographic transition, demographic dividend, and lewis turning point in China [J]. *Economic research journal*, 2010, 3(2): 107-119.
- [21] ROSTOW W, LEWIS W. *The evolution of the international economic order* [M]//*International economic policies & their theoretical foundations*. New York: Academic Press, 1982: 15-37.

- [22] WATANABE S. The lewisian turning point and international migration: the case of Japan [J]. *Asian and Pacific migration journal: APMJ*, 1994, 3(1): 119-147.
- [23] EGLI H, STEGER T M. A dynamic model of the environmental kuznets curve: turning point and public policy [J]. *Environmental & resource economics*, 2007, 36(1): 15-34.
- [24] HUA Z W, ZHENG W. Kuznets turning point determines China' s future economic direction [J]. *Financial expo*, 2011(6): 31-33.
- [25] WEI J Y, WU Y P, WU Z J. Empirical study on the relationship between urban-rural income gap and economic growth in China based on Kuznets turning point [J]. *China agricultural resources and regional planning*, 2015, 36(6): 112-119.
- [26] ZHU X H. Mengers turning point and its theoretical principle [J]. *Financial expo*, 2019(2): 37-39.
- [27] SHI L K, QIN Z J, YAN H Q. Stock market turning point prediction method based on minimum variance [J]. *Application research of computers*, 2017, 34(11): 3373-3378.
- [28] CHRISTOU C, CUNADO J, GUPTA R, et al. Economic policy uncertainty and stock market returns in pacific rim countries: evidence based on a bayesian panel VAR model [J]. *Journal of macroeconomics*, 2017, 40: 92-102.
- [29] PETROVSKA M, KRSTEVSKA A, NAUMOVSKI N. Forecasting macedonian business cycle turning points using qualVar model [J]. *Nephron clinical practice*, 2016, 5(3): 61-78.
- [30] SCHREIBER S, SOLDATENKOVA N. Anticipating business-cycle turning points in real time using density forecasts from a VAR [J]. *Journal of macroeconomics*, 2016, 47: 166-187.
- [31] ANDERSSON E, BOCK D, FRIESÉN M. Some statistical aspects of methods for detection of turning points in business cycles [J]. *Journal of applied statistics*, 2006, 33(3): 257-278.
- [32] ANDERSSON E, BOCK D, FRISÉN M. Detection of turning points in business cycles [J]. *Journal of business cycle measurement and analysis*, 2004(1): 93-108.
- [33] FUKUDA K. Forecasting growth cycle turning points using US and Japanese professional forecasters [J]. *Empirical economics*, 2009, 36(2): 243-267.
- [34] CANOVA F, CICCARELLI M. Forecasting and turning point predictions in a bayesian panel VAR model [J]. *Journal of econometrics*, 2004, 120(2): 327-359.
- [35] BILLIO M, CASARIN R, RAVAZZOLO F, et al. Combination schemes for turning point predictions [J]. *The quarterly review of economics and finance*,

2012, 52(4): 402-412.

[36] KOSKINEN L, OLLER L E. A classifying procedure for signalling turning points [J]. *Journal of forecasting*, 2004, 23(3): 197-214.

[37] PELÁEZ R F. Ex ante forecasts of business-cycle turning points [J]. *Empirical economics*, 2007, 32: 239-246.

[38] KULENDRAN N, WONG K K F. Determinants versus composite leading indicators in predicting turning points in growth cycle [J]. *Journal of travel research*, 2011, 50(4): 417-430.

[39] JUDGE G. Turning-point errors and directional errors in forecasting: the case of Japan [J]. *Journal of macroeconomics*, 1985, 7(2): 261-271.

[40] KULENDRAN N, WONG K K F. Predicting quarterly Hong Kong tourism demand growth rates, directional changes and turning points with composite leading indicators [J]. *Tourism economics*, 2009, 15(2): 307-322.

[41] DENG H L, LI X Q. Stock price turning point prediction method based on chaotic time series analysis [J]. *Statistics & decision*, 2007(9): 19-20.

[42] REISS J D. *The analysis of chaotic time series* [D]. Atlanta: Georgia Institute of Technology, 2001.

[43] LI X, DENG Z. Prediction of turning points for chaotic time series using ensemble ANN model [C]//2008 7th world congress on intelligent control and automation. Chongqing: IEEE, 2008: 3459-3464.

[44] SCHARF J H. The position of the turning points in the growth curves during the intra-uterine period in homo sapiens [J]. *Acta anatomica*, 1974, 88(4): 520-540.

[45] ZHAN X Y. Guangxi industrial structure serviceization turning point prediction and countermeasures—based on growth curve model [J]. *Guangxi social sciences*, 2011(4): 24-28.

[46] HUBBERT M K. *Energy resources: a report to the committee on natural resources of the national academy of sciences—National Research Council* [R]. Washington: National Academy of Sciences-National Research Council, 1962.

[47] TIAN H L, LI C Q, YAN H Q. Research and application of 缠论 and similar grey model prediction methods in stock price turning point prediction [J]. *Application research of computers*, 2020, 37(6): 1666-1669, 1678.

[48] JACKSON B P. A turning point for Europe' s east [J]. *Policy review*, 2010(160): 49-62.

[49] QU P P. Mobile communication service income prediction method based on grey buffer operator theory [J]. *Grey systems: theory and application*, 2014, 4(2): 250-259.

- [50] CLAVERIA O, MONTE E, TORRAS S. Tourism demand forecasting with neural network models: different ways of treating information [J]. *International journal of tourism research*, 2015, 17(5): 492-500.
- [51] HU Y. Neural network turning point prediction method in financial markets [J]. *Finance & economics*, 2017(18): 81-82.
- [52] CHEN J H. Research on artificial neural network prediction of enterprise development turning points [J]. *Technology economics & management research*, 2003(1): 31-33.
- [53] MA F F, WANG J B. Analysis of three types of stock turning points based on RBF neural network [J]. *Science & technology information (academic research)*, 2008(6): 211-214.
- [54] GUO P C, WEI L, CHANG W X, et al. The contribution of logistics industry to economic growth based on logis model [C]//2010 international conference of information science and management engineering. Shanxi: IEEE, 2010: 489-492.
- [55] GOLDEN P N, TELLIS G J. Will it ever fly? Modeling the takeoff of really new consumer durables [J]. *Marketing science*, 1997, 16(3): 256-270.
- [56] ANDERSEN B. The hunt for S-shaped growth paths in technological innovation: a patent study [J]. *Journal of evolutionary economics*, 1999, 9(4): 487-526.
- [57] XU H, WINNINK J, YUE Z, et al. Multidimensional scientometric indicators for the detection of emerging research topics [J]. *Technological forecasting and social change*, 2021, 163: 120490.
- [58] TRAAG V A, WALTMAN L, VAN ECK N J. From louvain to leiden: guaranteeing well-connected communities [J]. *Scientific reports*, 2019, 9(1): 1-12.
- [59] CAO Y W, XU H Y, WU H W, et al. Breakthrough prediction of emerging research topics based on citation curve fitting—taking the stem cell field as an example [J]. *Library and information service*, 2020, 64(5): 100-113.

Author Contributions

Xu Haiyun: Conceived research ideas, designed research framework, wrote paper; **Zhang Huiling:** Conducted literature review and wrote part of literature review; **Wu Huawei:** Performed data processing and visualization, revised paper; **Liu Ziqiang:** Conducted part of data processing.

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

Source: ChinaXiv –Machine translation. Verify with original.