Postprint: Re-validation of Patent Forward Citations Following a Logistic Diffusion Model

Authors: Zhang Xian, Tian Pengwei, Ru Lijie, Xu Haiyun

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] Patent citation is a mechanism for achieving technology diffusion, and studying the characteristics of patent forward citation behavior is an important perspective for measuring technology diffusion patterns. [Method/Process] Using the graphene sensor technology field as a case study, this paper conducts Logistic regression analysis on patent forward citation trends to validate the diffusion model of the technology field, and compares it with related research methods and results. [Results/Conclusion] The study validates that patent forward citation conforms to the Logistic diffusion model and represents a reliable perspective for studying technology diffusion; it further extends the existing research conclusion that "forward citation of basic core patents in a certain field follows the Logistic diffusion model" to "forward citation of patents in a certain field follows the Logistic diffusion model."

Full Text

Re-validation of the Logistic Diffusion Model for Patent Forward Citations

Zhang Xian[1], Tian Pengwei[1,2], Ru Lijie[1,2], Xu Haiyun[1]

[1]Institute of Scientific and Technical Information of China, Beijing 100038
[2]University of Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] Patent citation is a behavior that reveals technology diffusion patterns, and studying the characteristics of patent forward citation provides an important perspective for measuring technology diffusion modes. [**Method/Process**] This paper takes the graphene sensor technology field as a case study, conducts Logistic regression analysis on patent forward citation trends to validate the diffusion model of the technology field, and compares the methodology and results with related research. **[Result/Conclusion]** The study verifies that patent forward citations conform to the Logistic diffusion model, confirming it as a reliable perspective for technology diffusion research. Furthermore, it extends the existing conclusion that "forward citations of basic core patents in a certain field follow the Logistic diffusion model" to "forward citations of patents in a field follow the Logistic diffusion model."

Keywords: technology diffusion; patent citation; Logistic model; graphene sensor
Classification Number: G306

Technology diffusion theory, proposed by E. M. Rogers, conceptualizes technology diffusion as the process through which a new technology spreads via channels to eventual adopters or users [1]. While classic technology development models suggest a simple linear path from basic R&D to commercial application, in practice, technology diffusion often manifests as a complex iterative process [2]. Numerous studies have compared the effectiveness of different models in technology diffusion research [3-5], with empirical results indicating that because technology diffusion is more strongly influenced by internal factors than external ones, the Logistic growth model—which emphasizes internal influence—is more suitable for studying the diffusion patterns of technological innovation [6]. Patents represent one important pathway for technology diffusion, and forward citation behavior in particular reflects the adoption trajectory of patented technology after its introduction, a behavior widely recognized by scholars as having significant technology diffusion implications [7-8]. Patent citation data has been regarded by many researchers as an objective and mature indicator for measuring technology diffusion [9-11]. Utilizing patent forward citation behavior to measure technology diffusion has become a crucial research perspective for understanding technology diffusion pattern characteristics.

This study employs patent forward citation behavior as an objective representation of technology diffusion activities, using the graphene sensor field as a case study. Based on seed patents in the field and their forward and backward citations, we constructed a technology domain patent collection and performed Logistic regression analysis on the forward citation trends of all patents within the domain, thereby validating that forward citations of patents in the field follow the Logistic diffusion model. Through comparison with the research of M. H. Fallah and E. Fishman et al. [12] and Zhang Xiaoqiang et al. [13], this study achieves three objectives: First, it re-verifies that patent forward citations conform to the Logistic diffusion model, confirming it as a reliable research perspective for technology diffusion. Second, it experimentally demonstrates that the conclusion from Zhang Xiaoqiang et al.— "forward citations of basic core patents in a certain field follow the Logistic diffusion model" —can be extended to "forward citations of patents in a field follow the Logistic diffusion model." Third, it discusses the experimental methodology and results of Logistic regression fitting studies.

1.1 Patent Citation and Technology Diffusion

The citation relationship between patents constitutes a behavior that achieves technology diffusion effects [14]. When subsequent patented technology cites prior patented technology, it generates flows, transfers, and diffusion of technology and knowledge, thereby promoting technology development and commercialization. A. B. Jaffe et al. were among the earliest to use patent data for technology diffusion research, employing patent citation information to analyze knowledge spillovers between countries [11]. Multiple studies have demonstrated correlations between patent citations and both patent value and technology diffusion. For instance, S. B. Chang et al. inferred the correlation between forward citations, technology diffusion, and patent value [15].

Huang Lucheng et al. [16] comprehensively summarized the current state of technology diffusion research based on patent citation relationships, categorizing existing research into six main aspects: (1) analysis of inter-country knowledge spillovers and diffusion using patent citation information; (2) the impact of patent flows on productivity and R&D output; (3) knowledge flows and technology diffusion within or between industries; (4) technology diffusion research combining patent citation and network analysis methods; (5) technology diffusion curve studies using patent or patent citation data to reflect technology diffusion stages; and (6) technology diffusion prediction based on diffusion models. They also analyzed shortcomings and limitations in existing research, including: failure to adequately reflect dynamic changes in the technology diffusion process; insufficient quantity and depth of research on diffusion within or between technology (industry) fields compared to inter-country technology diffusion studies; and inadequate exploration and prediction of diffusion prospects in potential application fields or industries.

1.2 Logistic Diffusion Model

The Logistic equation was first proposed by Belgian mathematician P. F. Verhulst in 1838 and gained attention from biologists and statisticians in the 1920s. It effectively describes certain bounded growth phenomena and has been widely applied in forecasting, information science, biology, agriculture, and economics [17]. The Logistic equation can be expressed as:

$$Y(t) = \frac{L}{1 + e^{-k(t-B)}}$$

where $Y(t)$ is the performance parameter at time t, representing the degree of diffusion in technology diffusion research; L is the growth upper limit of parameter Y, representing the saturation level of technology diffusion; t is time; B is the curve inflection point, representing the turning point of growth diffusion;

and k is the curve slope, representing the diffusion rate. Parameters B and k are derived from regression equations.

The Logistic model has been applied to technology diffusion trajectory comparison, technology diffusion pattern characteristic research, technology diffusion influencing factor analysis, and trend forecasting [6, 18-19]. In 2009, M. H. Fallah and E. Fishman et al. selected the Top 5 highly cited patents in three fields—biotechnology, telecommunications technology, and alternative energy technology—and conducted fitting analyses based on their forward citation frequencies using linear, quadratic, S-curve, and Logistic models. They concluded that the Logistic model showed low fitting significance, while the other three models demonstrated high fitting degrees [12]. In 2014, Zhang Xiaoqiang et al. conducted Logistic regression analysis on one basic core patent in the giant magnetoresistance field, concluding that "forward citations of basic core patents in a certain field follow the Logistic diffusion model" [13].

2 Research Hypothesis

This paper argues that the studies by M. H. Fallah and E. Fishman and Zhang Xiaoqiang et al., which selected highly cited patents and basic core patents as research objects respectively, have not rigorously demonstrated whether the fitted technology diffusion characteristics truly reflect the overall technology diffusion trends of the entire technology field. In reality, highly cited patents or basic core patents represent only a tiny minority within a field, while the vast majority consists of numerous low-frequency cited patents. Therefore, this paper contends that while these two studies can address whether the technology diffusion characteristics of basic (or core) patents in a technical field satisfy the Logistic diffusion model, they have not effectively validated that the overall patented technology diffusion trends of the field conform to the Logistic diffusion model.

Consequently, this study proposes to represent the degree of technology diffusion in a technical field through the forward citation development trends of all patents within that field and puts forward the research hypothesis: forward citations of all patents in a technology field follow the Logistic diffusion model. This hypothesis will be tested in the following sections.

3 Experimental Study

This paper selects the graphene sensor field as the research object. Due to its unique structure and exceptional electrical, mechanical, optical, chemical, and thermal properties, graphene is considered a promising material for the "post-silicon era" of nanoscale transistors and circuits. Its potential application areas include high-speed transistors, optical modulators, (flexible) transparent electrodes, printed electronics, novel composite materials, ultrasensitive sensors, new catalysts, gene sequencing, and energy storage devices [20]. Currently, graphene has become one of the hottest research topics in physics and materials science, with countries worldwide establishing it as a long-term strategic development direction. Patent applications in this field are highly active, and analysis of patent activity characteristics has attracted significant attention. Moreover, graphene technology has application potential in numerous fields, with related patents already covering six main areas: electronic devices, energy, optoelectronics, materials, chemistry, and biomedical applications [21-22]. Therefore, citation relationships in graphene-related patents can reflect rich technology diffusion patterns.

3.1.1 Data Source Selection This study uses Thomson Reuters' Derwent Innovations Index (DII) as the data source. DII includes over 10 million basic invention patents and more than 30 million patents from over 40 patent authorities worldwide, with data traceable to 1963. All patent documents are organized by patent family, enabling comprehensive comparative analysis of major countries/regions.

3.1.2 Dataset Construction Strategy This study constructs the analytical dataset through the following steps: (1) Identify a batch of patents with highly relevant technical themes to establish a seed patent collection; (2) Extract forward and backward citation relationships of each seed patent, collecting both prior cited patents referenced by these seed patents and subsequent citing patents that reference these seed patents; (3) Merge seed patents, cited patents, and citing patents to form the data sample collection.

Graphene patents first emerged in 2000, with a major breakthrough in preparation technology achieved in 2004. Considering the time lag between patent application and publication dates, the necessary technological and market development process required for prior patents to be cited by subsequent patented technologies, and the time lag between subsequent citing behavior occurrence and publication, this study limits the application year range of seed patents to 2000-2011 to ensure richer citation information. Additionally, to avoid impacts from differences in patent application and granting regulations across countries (organizations), this study restricts the research object to U.S. patents.

3.1.3 Sample Set Construction The retrieval strategy is shown in Table 1 , yielding 149 original patent data items. After assessing content relevance, 126 items were selected as seed patents for graphene sensors in this study. Forward and backward citations of these 126 seed patents were extracted for U.S. patents. To ensure the data samples could adequately reflect the technology diffusion chain, two generations of prior cited U.S. patents were collected. Merging seed patents with citation patents resulted in a total of 26,537 U.S. patents as the data analysis sample collection for this study.

3.2 Data Characteristics Observation

Patents were grouped by application year. The 26,537 U.S. patents covered application years from 1961 to 2015, forming 55 patent groups. For each group, annual citation frequencies from the application year to the present were calculated, yielding 55 sets of forward citation trend data. To avoid impacts from patent data publication lags, 50 groups of citation frequency change data for patents applied between 1961 and 2010 were selected as observation sample values.

Table 2 presents the annual citation frequency statistics for the 50 patent groups in each year after application. Table 3 shows the cumulative citation frequency statistics for the 50 patent groups in each year after application.

Scatter plot observations revealed that the forward citation frequency development trends of the 50 patent groups conform to technology diffusion model characteristics: the statistics for each period (annual citations) follows a bell-shaped curve (as shown in Figure 1 [Figure 1: see original paper]), while the cumulative statistics (cumulative citations) follows an S-shaped curve (as shown in Figure 2 [Figure 2: see original paper]).

The 50 sets of observation data (see Table 3) reflect that the diffusion rates of graphene sensor field patented technologies vary and are influenced by patent age. For example, early published patented technologies may have limited diffusion speed due to being in the embryonic stage with a small patent volume, while later technologies, though potentially having broader impact due to larger volume, may have shorter citation chains due to their recent introduction.

Therefore, this study selected the middle 5 groups from the 50 observation datasets (1986-1990 group) and used the sum of the 5 groups' observations as experimental variables (see Table 4) to conduct patented technology diffusion curve fitting analysis for the graphene sensor field, aiming to better reflect the steady-state characteristics of domain technology diffusion.

3.3.2 L Value Estimation

Numerous methods exist for estimating Logistic equation parameters (including maximum value L) [23]. This study employs the trial method. A number slightly larger than all observed Yi values is selected as the initial L value, which is then increased by a certain step size. For each Li value set, corresponding parameter estimates are calculated, and the fitting results of the corresponding function model are compared until optimal fitting is achieved.

Based on the observations in Table 4, the trial method was used to set the L value at 210,000.

3.3.3 Curve Regression Fitting Process

In SPSS 19.0 software, the curve estimation regression function was selected. Following the prompts, Y(t) was input as the dependent variable and t value as the independent variable, the Logistic model was selected, the estimated maximum value parameter L was entered, and variance analysis was selected to output test results (display ANOVA table). The execution program yielded the following experimental fitting results:

L = 210,000, b0 = 0.001, b1 = 0.739, with the model showing strong fitting significance. According to the SPSS analysis results, the Logistic model fitting effect is good in this experimental case, empirically verifying that forward citations of patents in the technology field indeed follow the Logistic technology diffusion model.

4.1 Representativeness of Data Object Selection

In the study by M. H. Fallah and E. Fishman, the top 10 highly cited patents in the field were selected as analysis objects, based on the assumption that highly cited patents represent key breakthrough inventions in the field. Zhang Xiaoqiang et al. selected one basic core patent with fundamental importance in the field, arguing that it has strong domain representativeness and its forward citations can illustrate the development and diffusion degree of the field, while its strong applicability provides reverse promotion effects on technology development, thus ensuring strong consistency between the diffusion of the patented technology and technological innovation diffusion. These two studies essentially used individual patents within a technology field to represent the entire domain, using the forward citation trends of these individual patents to represent the overall technology diffusion trends of the field.

This paper argues that while these two studies can address whether the technology diffusion characteristics of basic (or core) patents in a field satisfy the technology diffusion model, they lack sufficient rigor in proving that the entire field (especially the numerous low-frequency cited patents) satisfies these characteristics. This study constructs a relevant technology patent collection based on seed patents and their forward and backward citations to represent the overall technology field, analyzing domain technology diffusion trends based on the forward citation characteristics of all patents in the collection. Therefore, in terms of patent selection, the data object selection method in this study better represents the overall technology field.

4.2 Stability of Citation Trend Development

M. H. Fallah and E. Fishman selected Top 5 and Top 1 highly cited patents in their field for analysis. Due to the limited scale of analysis objects, it is difficult to exclude interference from random factors on citation frequency change trends. The cumulative citation volume trend charts in their paper (see Figure 4 [Figure

4: see original paper] and Figure 5 [Figure 5: see original paper]) reveal that potential outlier samples may have influenced curve morphology.

This study selected forward citation cumulative volumes from mid-range (1986-1990) patent outputs in the graphene sensor field' s 50-year (1961-2010) patent production as analysis samples, effectively avoiding interference from random factors on sparse sample volumes. Simultaneously, using the total patent output over five years captures trend characteristics across years within the interval while smoothing through group summation overcomes potential outlier samples caused by random interference in individual years or patents, making model fitting more standardized and the fitting results more representative of the overall field.

4.3 Impact of Maximum Parameter Estimation on Model Fitting Effect

Zhang Xiaoqiang et al. noted that M. H. Fallah and E. Fishman' s model fitting showed low significance because they did not set the Logistic upper limit. During this study' s implementation, attempts were made to use the grey system GM(1,1) model modeling method for parameter estimation, but the fitting results based on the obtained L predicted values differed significantly from actual observations, confirming Zhang Xiaoqiang et al.' s finding that maximum parameter estimation significantly impacts model fitting results.

Grey system theory is a new method for studying uncertain problems with small data and poor information, with wide application scope for grey prediction. However, it still has applicability issues that must be determined based on the characteristics of the prediction problem itself. For example, when the prediction problem has internal mechanisms, such as data conforming to certain functional characteristics, grey prediction may not be the most suitable method, and fitting or regression methods should be selected instead [24]. Additionally, data for grey prediction models should have certain monotonicity, and the magnitude of increase or decrease should also have certain monotonicity—these are theoretical bases for determining whether data are suitable for GM(1,1) models [25]. The experimental attempts in this study reflect the limitations of the grey prediction method in this research scenario.

4.4 Potential Applications of Patent Citation Logistic Curve Research

The Logistic model is an important application of growth curve methodology, essentially a method that mechanically extrapolates past data change trends. It embodies diffusion theory and social learning theory, reflecting characteristics of social simulation, communication, and exchange.

When prior patented technology is cited by subsequent patents, it represents that the prior invention' s innovative ideas have gained attention and even acceptance from subsequent applicants. The forward citation development trend of patents reflects, to a certain extent, the process by which patented technology is accepted by the public and market after its introduction, as well as the growth process of the technology field. Therefore, Logistic model fitting analysis of patent forward citation trends, in addition to studying technology diffusion behavior characteristics, can also be applied to technology field growth characteristic research. Based on citation development trends and combined with Logistic curve parameters such as maximum value, inflection point, and time, it can analyze and predict technology field growth limits, development turning points, and aging rates. Combined with more technical theme characteristic items, it can also be used for technology evolution characteristic analysis and technology or product maturity prediction.

5 Conclusion

Building upon the research of M. H. Fallah and E. Fishman and Zhang Xiaoqiang et al., this study re-validates that patent forward citations follow the Logistic diffusion model. Compared with existing research, this paper' s more optimally designed experimental validation demonstrates that not only highly cited patents in a technology field but the diffusion behavior of the entire technology field conforms to Logistic diffusion model characteristics. Therefore, this study extends the previous research conclusion that "forward citations of basic core patents in a certain field follow the Logistic diffusion model" to "forward citations of patents in a field follow the Logistic diffusion model." The experimental methodology and specific implementation results designed in this paper validate the effectiveness of this hypothesis, though further experimental validation in more different technology fields is needed. Additionally, based on the fundamental principles of growth curve methodology, this paper proposes that the characteristic of "patent forward citations following the Logistic diffusion model" may also play a role in other problem scenarios such as technology maturity and technology evolution analysis, with its research and application significance warranting deeper exploration.

References

[1] ROGERS E M. Diffusion of innovation[M]. 3rd ed. New York: The Free Press, 1983.
[2] EDQUIST C, JACOBSSON S. Flexible automation: the global diffusion of new technology in the engineering industry[M]. Oxford: Basil Blackwell, 1988.
[3] CHU W L, WU F S, KAO K S, et al. Diffusion of mobile telephony: an empirical study in Taiwan[J]. Telecommunications policy, 2009, 33(9): 506-520.
[4] LUNDWALL B A. National systems of innovation: towards a theory of innovation and interactive learning[M]. London: Anthem Press, 1992.
[5] TENG J T C, GROVER V, GUTTLER W. Information technology innovations: general diffusion patterns and its relationships to innovation characteristics[J]. IEEE transactions on engineering management, 2002, 49(1): 13-27.
[6] WONG C Y, THIRUCELVAM K, RATNAVELU K. Diffusion trajectories

of emerging sciences in Malaysian R&D system[J]. Technological forecasting & social change, 2010, 77(7): 1109-1125.
[7] NARIN F, HAMILTON K S, OLIVASTRO D. The increasing linkage between US technology and public science[J]. Research policy, 1997, 26(3): 317-330.
[8] LAI K K, ZHANG S B. Constructing a technology diffusion model for business methods: integrating patent citations and Bayesian models[J]. Journal of Technology Management, 2004, 9(1): 1-34.
[9] BREITZMAN A F, MOGEE M E. The many applications of patent analysis[J]. Journal of information science, 2002, 28(3): 187-205.
[10] MEYER M. Does science push technology? Patents citing scientific literature[J]. Research policy, 2000, 29(3): 409-434.
[11] JAFFE A B, TRAJTENBERG M, HENDERSON R. Geographic localization of knowledge spillovers as evidenced by patent citations[J]. Quarterly journal of economics, 1993, 108(3): 577-598.
[12] FALLAH M H, FISHMAN E, REILLY R R. Forward patent citations as predictive measures for diffusion of emerging technologies[C]//Portland: PICMET, 2009: 420-429.
[13] ZHANG X Q, DAI W S, YANG J Y. Do patent forward citations follow the Logistic diffusion model[J]. Journal of Intelligence, 2014, 33(6): 40-43, 65.
[14] WANG X S. The dissemination mechanism and mode of patent intelligence[J]. China Information Review, 1995(7): 19-20.
[15] CHANG S B, LAUI K, CHANG S M. Exploring technology diffusion and classification of business method patents: using the patent citation network[J]. Technological forecasting & social change, 2009, 76(1): 107-117.
[16] HUANG L C, WANG N. A review of technology diffusion research from the patent perspective[J]. Science of Science and Management of S.&.T., 2011, 32(10): 27-34.
[17] ZHU Z Y, CHEN W H, CHEN F. A comparative study of Logistic curve and Gompertz curve[J]. Mathematics in Practice and Theory, 2003(10): 66-71.
[18] CHEN Y H, CHEN C Y, LEE S C. Technology forecasting and patent strategy of hydrogen energy and fuel cell technologies[J]. International journal of hydrogen energy, 2011, 36(12): 6957-6969.
[19] FRANK L D. An analysis of the effect of the economic situation on modeling and forecasting the diffusion of wireless communications in Finland[J]. Technology forecast & social change, 2004, 71(4): 391-403.
[20] Science News Staff. Breakthrough of the year: the runners-up[J]. Science, 2009, 326(5960): 1600-1607.
[21] MA T C, WAN Y, FENG R H. Analysis of international R&D trends in graphene patent technology[J]. Science Focus, 2012, 7(3): 25-36.
[22] CHEN C Y, WU H Z. Visual analysis of patent information in China' s graphene field[J]. Modern Intelligence, 2014, 34(3): 120-124.
[23] ZHANG Y M, GAI J Y. Parameter estimation of Logistic model[J]. Journal of Sichuan College of Animal Science and Veterinary Medicine, 1994, 8(2): 47-52.
[24] LU Y. Research and application of grey prediction models[D]. Hangzhou:

Zhejiang Sci-Tech University, 2014.
[25] CHEN Z J, WANG F L, LU S X. Applicability analysis of grey prediction model GM(1,1) and its application in fire risk prediction[J]. Engineering Science, 2007, 9(5): 91-94.

Author Contributions:
Zhang Xian: Research design, implementation, and paper writing;
Tian Pengwei: Data analysis;
Ru Lijie: Data analysis;
Xu Haiyun: Research design supplementation.

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