

Transfer and distribution of knowledge creation activities of bio-scientists in knowledge space

Xiang Liu · Feicheng Ma

Received: 14 May 2012 / Published online: 4 August 2012
© Akadémiai Kiadó, Budapest, Hungary 2012

Abstract In order to explore the rule of knowledge creation activities at both temporal and spatial scales, this paper makes statistical analysis of the time interval and spatial displacement of consecutive knowledge creation activities of high-yield, low-yield, and ASFP (all the scientists published at least four papers), respectively. The research shows that, for high-yield scientists, the time interval of knowledge creation activities obeys heavy-tailed distribution and embodies bursting features, with both long-time silence and intensive burst of creation activities. The time interval distribution of low-yield scientists is approximate to exponential distribution, and is often randomly and occasionally distributed. For ASFP, the spatial distribution of creation activities also embodies heavy-tailed features, where their activities are intensively confined to a certain knowledge field, and where long-distance exploration across the knowledge fields has also been made in knowledge creation activities.

Keywords Knowledge space · Transfer · Knowledge creation activity · Human dynamics · Heavy-tailed distribution

Introduction

Exploring the transfer and distribution of knowledge creation activities of scientists is an effective way to study the knowledge generation and knowledge creation processes. Brookes (1981) proposed the concept of “knowledge space” a long time ago, and believed that knowledge created by humankind is distributed in multi-dimensional space and manifests in different structures and forms. Questions addressed in this report include: how do the knowledge creation activities of scientists transfer in the knowledge space, and how are these transfer processes distributed at the temporal and spatial scales in the knowledge space?

X. Liu · F. Ma (✉)
Center for Studies of Information Resources of Wuhan University,
Wuhan 430072, People’s Republic of China
e-mail: fchma@whu.edu.cn

X. Liu
e-mail: xianglew@gmail.com

It is generally considered that long-time and concentrated study in a small knowledge field is conducive to in-depth exploration and discussion in this field, while interdisciplinary study can expand the research scope and often trigger creative ideas and distinctive thoughts. Since the Renaissance, science has been developing rapidly. The knowledge accumulation and scientific exploration of humankind have covered a wide range of fields. Disciplines and knowledge domains are constantly subdivided. Since every person has limited energy, the knowledge creation activities of scientific researchers are becoming increasingly more specialized and in-depth. It is very difficult for a modern person to make extraordinary achievements in a variety of knowledge domains. For instance, it is generally believed in the mathematical community that the mathematician Henri Poincaré of the late nineteenth and early twentieth centuries has been the most recent person to display comprehensive knowledge about mathematics and its applications. The modern and contemporary mathematician is at most an expert in a specific area or branch of mathematics. However, many people believe that when scientific researchers, in this case, transfer from one knowledge domain to another, it is necessary to arouse the interest and enthusiasm of those scientists working in their field. Some scholars have become experts in multiple knowledge domains. Moreover, the transfer of research domains often can spark unexpected inspiration. For example, the twentieth-century scholar Herbert Simon frequently transferred his research interests from economics, psychology, and behavioral science to computational science, and ultimately made remarkable achievements in numerous fields.

It is thus very interesting to explore the transfer of knowledge creation activities of scientific researchers. In the actual process of knowledge creation, relevant questions include: What is the probability of enacting small domain research and long-distance exploration across knowledge fields, and how large is the scope of such field transfers? Do knowledge creation activities remain stable for a long time, or are they random and disorderly? This paper focuses on these problems and explores the transfer and distribution of knowledge creation activities at temporal and spatial scales.

Literature review

The movement and migration of human beings or creatures is often considered as disorderly and random wandering. Their movements are confined in a certain scope where long-time stay and long-distance transfer rarely happen. Erlang distribution can be used to describe this kind of time lag (Li 2003). The time of arrival is considered a Poisson process that only depends on time without lag effect and obeys exponential distribution. Moreover, since displacement distance is time-dependent ($\text{distance} = \text{time} \times \text{speed}$, where speed is a constant value), the spatial distribution of transfer also obeys exponential distribution. However, current studies have noted deviations. Viswanathan et al. (1999) found that the flying time and range of the albatross obeys power-law distribution; Nakamura et al. (2007) studied human movement at the physiological level and discovered similar distribution rules. It was also found that different types of people, such as depressed patients and healthy people, have the same power exponent, but the movement of depressed patients is made up of longer time intervals and shorter spatial displacements on average. Meanwhile, Brockmann et al. (2006), Gonzalez et al. (2008), Song et al. (2010) and Wang and Han (2010) have discussed the migration of human beings by such media as E-Check and mobile phone, and found that the spatial scope of human movement also obeys power-law distribution with a discrete index. Similar rules have been found in other behaviors of human beings, where Barabási (2005), Zhou et al. (2008), Guo (2011), and Fan et al.

(2010) indicated that the time interval distribution of emails, letters, movies on demand, blog comments, and book lending also have heavy-tailed and bursting features.

Heavy-tailed distribution, also called fat-tailed, thick-tailed, or long-tailed distribution, refers to a type of distribution without exponential order moments (Embrechts et al. 1997), i.e., $\forall \lambda > 0, Ee^{\lambda X} = \int_0^{\infty} e^{\lambda x} dF(x) = \infty$. Compared with Poisson distribution, heavy-tailed distribution has a slower rate of decay and a higher occurrence probability of larger observed value. This is reflected by bursting features of time interval distributions, i.e., the coexistence of long-time silence and short-time intensive bursts. Embrechts et al. (1997) clearly defined the heavy-tailed distribution in the following way: if the density function decays from power exponent to “0”, the distribution function is heavy-tailed; if the density function decays from exponential function to “0”, the distribution function is light-tailed. As shown in its distribution photograph, heavy-tailed distribution turns out to be a straight line in double logarithmic coordinates, while light-tailed distribution turns out to be an arc line. It is thus clear that the time interval and spatial scope of many human movements have heavy-tailed features. Is it true, however, that the transfer of knowledge creation activities of scientific researchers is also characterized by heavy-tailed distribution? To address this issue, we may first explore the dynamic mechanism for heavy-tailed distribution of animal behaviors.

Regarding the dynamical mechanism of heavy-tailed distribution in human or animal behaviors, Barabási (2005) proposed the priority-based task queuing model, where tasks are selected from the task queue according to priority, so that tasks with higher priority are accomplished at faster speeds and tasks with lower priority are postponed; Gabrielli and Caldarelli (2007) accurately analyzed this model and arrived at the conclusion that the model could generate the power-law distribution of waiting time under circumstances of both variable and invariable queue length. Blanchard and Hongler (2007) discussed the task queuing with deadlines using real-time queuing theory. According to their analytical results, when the deadline obeys heavy-tailed distribution, the distribution of waiting time also embodies heavy-tailed features. Han et al. (2008) proposed the self-adaptive interest model. According to this model, people’s interest in a certain behavior tends to increase with time. When their interest reaches a specific threshold value, a certain behavior will occur. However, frequent behaviors will undermine continuing interest. When the interest value reaches the threshold again, the behavior will tend to be silent. This process is continual. The temporal distribution of the simulation model obeys heavy-tailed distribution. Malmgren et al. (2009) observed the send mode of emails and pointed out that power-law distribution of time intervals result from periodical behaviors of human beings. He constructed the process model based on Cascaded Non-homogeneous Poisson Distribution, with the non-homogeneous Poisson process in the outer layer and the homogeneous Poisson process in the inner layer, thereby obtaining periodical heavy-tailed distribution.

According to the dynamic mechanism, knowledge creation activities of scientific researchers can also be considered as the priority-based task queue. Knowledge creation activities in a familiar knowledge field can soon be achieved, while knowledge creation activities in unfamiliar fields often require long-time deliberations. Knowledge creation activities also satisfy the interest rule to some extent, where researchers often show great interest in a new knowledge field. After continuous study for a certain period of time, these researchers may make numerous satisfactory achievements. However, as they carry on their studies in this field, they may become bored and choose to explore other knowledge fields. Knowledge creation activities also periodically change and transfer among different knowledge fields.

To explore the problem of scientists transferring and distributing the knowledge space during knowledge creation activities, this paper provides statistics on knowledge creation activities of scientists by analyzing the temporal and spatial distribution of their activities, and explores differences among various groups of creators (high-yield creators, low-yield creators, and ASFP) so as to reveal their transfer of research interests and changes in creation activities.

Research design

The published works of scientific researchers, reflecting achievements they made through their knowledge creation activities, can be regarded as their “footprint” in the knowledge space. Statistical analysis of these footprints can reveal the authors’ transfer processes and position distributions in the knowledge space. The time intervals and spatial displacements between two consecutive knowledge creation activities of scientific researchers can reflect the intensive degree of knowledge creation activities at temporal and spatial scales. In this paper, the distance between two consecutive steps of scientists represents the step length of creators. This paper mainly adopts scientific papers as the footprint of scientists in the knowledge space.

It is rather simple to directly measure the time interval between two consecutive creation activities of scientists. But the walking rule of scientists in the knowledge space is quite complicated. The statistical method for the walking rule in the knowledge space has been constructed as follows. There are two approaches to determine the exact spatial position of the footprint. One approach is to adopt the keyword vector of papers written by scientists. As the keywords of literatures can reflect the main features of these articles and represent core knowledge content of the literatures, it is appropriate to adopt the keyword vector. The second approach is to adopt the reference vector of papers, where the references of a thesis represent the authors’ theoretical basis and sources of their knowledge creation activity. For example, every reference in this paper can be regarded as a feature item, and all of these references can then form a feature vector, which can also be used for spatial positioning of scientists’ footprints.

While step length can only indicate the distance between two consecutive steps of scientists, it is incapable of measuring the walking range of scientists. Their walking range at the temporal level is the accumulative amount of time intervals between consecutive knowledge creation activities, and the distribution of accumulative time can be derived from the distribution of time intervals. On the other hand, it is more complicated to measure walking range at the spatial scale. Since knowledge space is multi-dimensional space, the walking range of scientists can be measured by the distance between their current footprint and initial footprint at the starting point. To simplify the process, this distance can be cast as one-dimensional space to calculate the distance between each current and initial footprint, which in turn obtains the statistical features of walking range distributions at the spatial scale.

Moreover, different groups of scientists have different knowledge creation behaviors. Three groups of scientists are analyzed below. The first is a high-yield group that has published numerous papers, so individual analysis can be made on their step length distributions at both temporal and spatial scales. The second is a low-yield group that has published a few papers. The third group is ASFP. As it is quite difficult to detect statistical features of individuals, several individual creators can be gathered together to obtain collective statistics and observation. Since the temporal and spatial transfer of three (or

fewer) knowledge creation activities happens only twice or even less often, it is quite difficult to observe the features of these scientists' knowledge transfers. Therefore, the statistics for ASFP group is based on the principle that the scientists shall publish at least four papers.

In the following section, actual data is applied to demonstrate the statistical features of the above two indicators, such as the walking features of scientists during knowledge creation activities at both temporal and spatial scales, and different footprint distributions of knowledge creation activities among different groups of scientists (high-yield group, low-yield group, and ASFP group).

Data collection

Data

The bibliographical descriptions below are collected from *Cell* and its sub-publications (including *Cell*, *American Journal of Human Genetics*, *Structure*, *Immunity*, *Neuron*, *Developmental Cell*, *Chemistry & Biology*, *Current Biology*, and *Molecular Cell*) in the *SCIE Database* (December 1998–December 2010) on ISI. Two document types, the Article and Proceedings Paper, are selected by means of refined subsets, while other types such as Editorial Material and Review are excluded. The reason for selecting the refined subset is that, the former two types can better reflect the innovation and inheritance of knowledge and these documents have a uniform writing model, while reviews belonging to summaries show significant differences from previous types. This paper thus selects only two document types and obtains 21,992 valid data items. The data is handled in the following manner.

Consistent processing of data: to keep the consistency of data type in the network, data references are filtered to exclude reference types other than main types, such as books and websites. The following data obtained includes 19,698 records in total, 373,697 references in total, an average number of references 18.97; 76,708 authors in total, an average number of authors 3.89; 40,408 keywords in total, and an average number of keywords 2.05; the time span of bibliography = 13 years (1998–2010). The most high-yield scientist is JESSELL TM, who completed 42 papers in total.

Approach

Keyword vector can clearly define the main content of a paper. While the keyword vector is too short to distinguish differences between papers, the reference vector is quite long and suitable for defining a citing publication. These two approaches have been taken into consideration below. Since both keywords and references have high and low frequencies, high-frequency keywords and highly-cited references contribute little to distinguishing differences between knowledge contents, and reflect fewer features of author footprint distance. In this case, the TF-IDF rule in vector space model (VSM) (Salton and Lesk 1968; Salton and McGill 1983) can be introduced for weighted processing of each feature item. The weighted formula is $w_i = f_i \times \log(N/d_i)$, where f_i is the frequency of term i occurring in the document, N is the total number of documents in the database, and d_i is the number of documents containing word i in the whole database. Logarithmic operation is conducted on the ratio between N and d_i , with an aim to reduce the impact of d_i and N on the final weighted value w_i . The above formula is called TF \times IDF: term frequency

$f_i \times$ inverse document frequency $\log(N/d_i)$. This weighted approach is proved to be justified and unshakeable (Robertson 2004). As both keywords and references are determined in papers, the term frequency is stable, and the difference lies only in the inverse document frequency. The low-weighted value of high-frequency keywords and references obtained by inverse document frequency can weaken their capability to distinguish footprint distance, but they highlight the distinguishing capacity of low-frequency keywords and references.

The footprint represented by No. t paper of scientist is recorded as vector $v(t)$. The step length is the sum of different feature items between No. t paper and No. $t - 1$, i.e., $d = \sqrt{\langle v(t) - v(t - 1), v(t) - v(t - 1) \rangle}$; that is, the inner product of vector difference is beneath the radical sign. Likewise, the distance between any footprint and start footprint indicates the walking range, i.e., $s = \sqrt{\langle v(t) - v(0), v(t) - v(0) \rangle}$; that is, $v(0)$ indicates the starting footprint. Limited data volume makes it difficult to define the initial footprint of the author, so that $v(0)$ is defined below as the earliest paper published by the author within the statistical time interval. And the corresponding walking range is the movement range of scientists within this time interval.

Data analysis and results

Time interval and spatial displacement of continuous knowledge creation activities

Time interval and spatial displacement of high-yield scientists

With respect to the distribution of time intervals, three high-yield scholars are used: JESSELL TM, TESSIER LAVIGNE M, and NASMYTH K. They have published 42, 41, and 41 research papers (including papers in which they are not the first authors) from 1998 to 2010, respectively, as shown in Fig. 1. The horizontal ordinate indicates the serial number of consecutive knowledge creation activities, e.g., No. n Paper. The vertical ordinate indicates the time interval between two adjacent knowledge creation activities, e.g., the publication time interval between No. n Paper and No. $n - 1$ Paper.

As shown in Fig. 1, the time interval between two consecutive knowledge creation activities of each scientist is unevenly distributed. Their knowledge may be created after a long time or within a short time period. For example, JESSELL published No. 28–No. 32 papers consecutively within a short period of time, and then spent a longer time publishing No. 33–No. 41 papers. NASMYTH published No. 30–No. 33 papers consecutively at a fast speed, but spent a longer time publishing No. 34–No. 37 papers.

With respect to spatial displacement, the footprints of scientist TESSIER LAVIGNE M are investigated. Figure 2 reflects the continuous displacement of keyword vectors and reference vectors from top to bottom. The horizontal ordinate indicates the number of papers published continuously, while the vertical ordinate indicates the degree of difference in the footprint vector (dimensionless) and the distance d of two continuous footprints:

The statistical result of JESSELL TM and NASMYTH K is also similar to those of TESSIER LAVIGNE M. We do not list those figures, but this can be confirmed by Fig. 4. As shown in Fig. 2, the displacement of keyword vectors is distributed in a concentrated manner, mainly between 20 and 30. High deviations occur occasionally, such as the difference degrees of No. 25 and No. 31 papers are 5 and 35, respectively. One possible

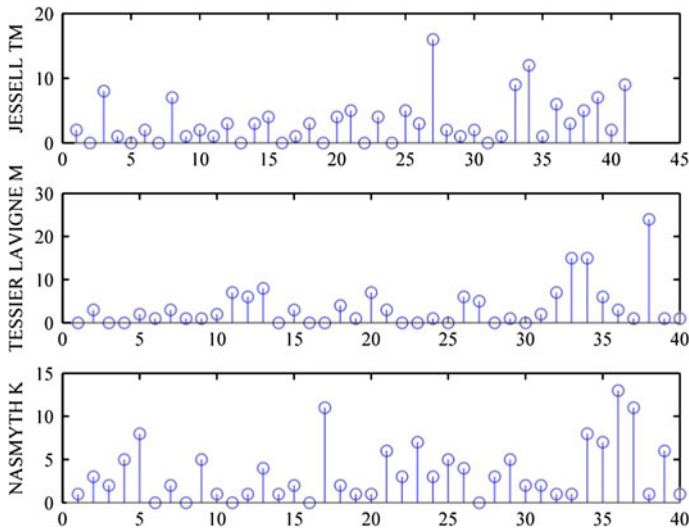


Fig. 1 Time interval between consecutive knowledge creation activities of high-yield scientists

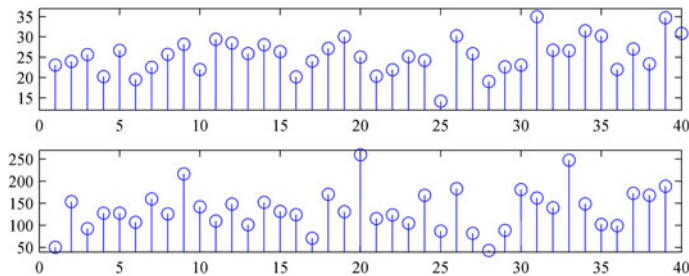


Fig. 2 Spatial displacement between consecutive knowledge creation activities of high-yield scientist of TESSIER LAVIGNE M, the figure shows continuous displacement of keyword vector and reference vector from top to bottom

reason for this is the short and fixed length of the keyword vector itself, which has restricted the occurrence of high deviations and has limitations for distinguishing two different papers. Displacement of the reference vector has larger amplitude, ranging from 0 to 270, with huge ups and downs and uneven distributions.

Time interval and spatial displacement of different groups of scientists

The question remains as to what kind of distribution the scientists’ two consecutive knowledge creation activities can obey. Due to the short time period of statistics, it is impossible to derive obvious rules from individual data. For this reason, statistical analysis is carried out on the collective data. Statistics are made on high-yield creators and low-yield creators, respectively. 936 scholars who have published at least 17 papers are selected as high-yield scientists, while 971 scholars who have published seven papers are selected as low-yield scientists. The data volume for both cases is around 950. As seen in Fig. 3, data is selected to ensure the comparability of figures. Statistical analysis is also made on

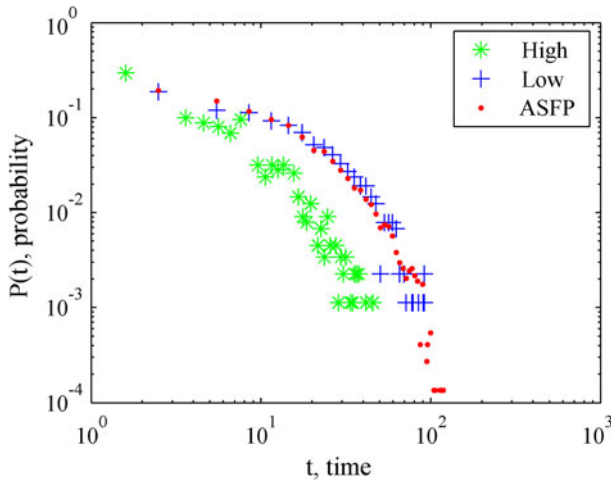


Fig. 3 Time interval distribution of continuous knowledge creation activities

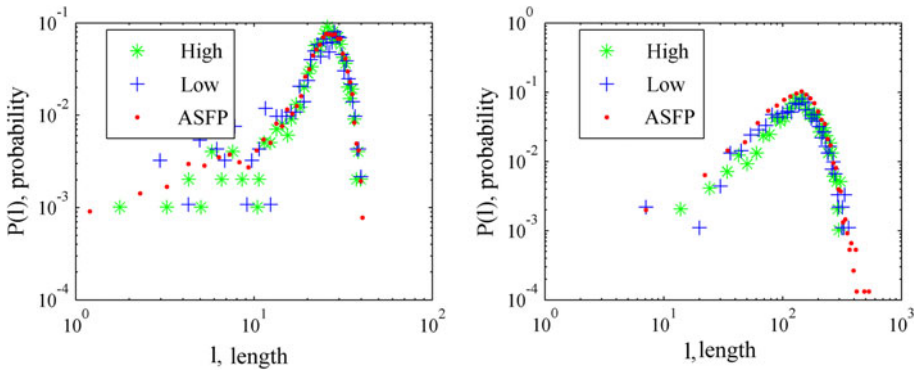


Fig. 4 Distribution of continuous step length

ASFP group, including 7,743 authors, along with statistics on the time interval of continuous knowledge creation activities of all creators who have published at least 4 papers:

As shown in Fig. 3, for high-yield scientists, the time interval distribution of continuous knowledge creation activities is in a straight line in the double logarithmic coordinate, which may obey heavy-tailed distribution. For low-yield scientists, the time interval distribution is shown in an arc line, which is approximate to exponential order distribution. For ASFP group, the time interval distribution is also shown in an arc line, which is approximate to exponential order distribution. Therefore, high-yield scientists have experienced periods of both long-time silence and intensive bursts. Low-yield creators have more random knowledge creation activities. Since the majority of scientists are low-yield creators, the general distribution of time interval obeys exponential order distribution. In addition, Fig. 3 also indicates that the time interval distribution of continuous knowledge creation activities of high-yield creators is to the left of that of low-yield creators. This means the time intervals of continuous knowledge creation activities of high-yield creators are shorter than that of low-yield creators.

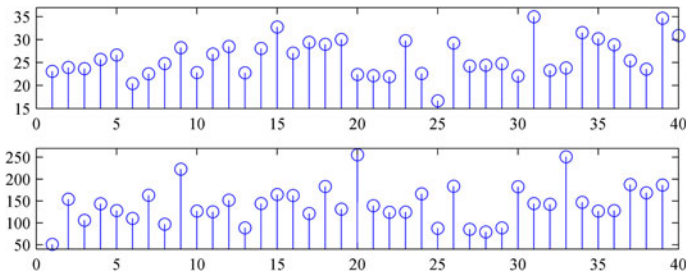


Fig. 5 Walking range of high-yield scientist of TESSIER LAVIGNE M, the figure shows continuous displacement of keyword vector and reference vector *from top to bottom*

Statistics are made on the step lengths of high-yield, low-yield, and ASFP. Under the same principle, authors who have published at least 17 papers are selected as high-yield scientists, authors who have published 7 papers are selected as low-yield scientists, and authors who have published at least 4 papers are selected as ASFP group, as seen in Fig. 4.

As shown in Fig. 4, the distribution curves of high-yield, low-yield, and ASFP overlap at the tail of the curves. All these curves obey heavy-tailed distribution. The distribution curve of step length described by keyword vector has a larger slope at the tail section, which indicates a higher concentration degree of distribution. This is consistent with the analysis in Fig. 2. The distribution curve of step length described by reference vector has a relatively smaller slope, which indicates a higher degree of uneven distribution.

Walking range distribution of continuous knowledge creation activities

Walking range of high-yield scientists

The footprints of scientist TESSIER LAVIGNE M are investigated. Figure 5 reflects the distance from the current foot print to the start point, described by the keywords vector and reference vector from top to bottom. The horizontal ordinate indicates the number of papers published continuously, while the vertical ordinate indicates the degree of difference in the footprint vector, and the distance *s* of two continuous footprints:

As shown in Fig. 5, the walking range described by the keyword vector is distributed in a concentrated manner, with high deviations occurring occasionally. The walking range described by reference vector has larger amplitude, with huge ups and downs and uneven distribution.

Walking range of different groups of scientists

Statistics are also made on the walking range of high-yield, low-yield, and ASFP. Under the same principle, authors who have published at least 17 papers are selected as high-yield scientists, authors who have published 7 papers are selected as low-yield scientists, and authors who have published at least 4 papers are selected as ASFP group, as seen in Fig. 6:

As shown in Fig. 6, the distribution curves of high-yield, low-yield, and ASFP overlap at the tail of the curves. All these curves obey heavy-tailed distribution. The distribution curve of walking range described by the keyword vector has a larger slope at the tail segment, which indicates a higher concentration range of creators walking. The

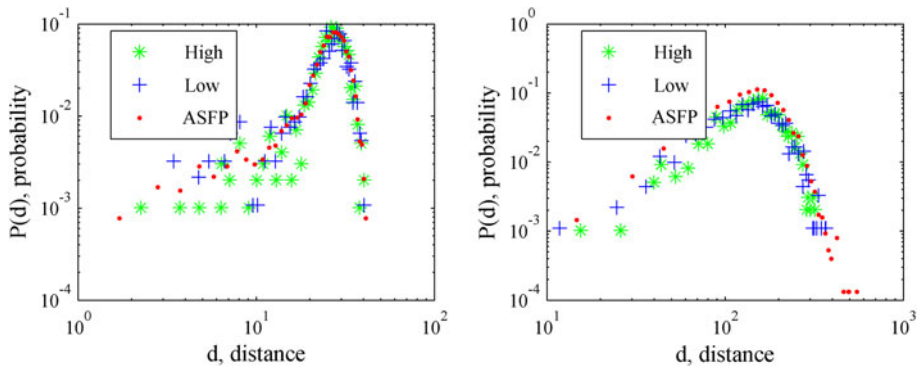


Fig. 6 Distribution of walking range

distribution curve of walking range described by the reference vector has a relatively smaller slope, which indicates a higher range of uneven walking.

Discussion

Statistical analysis on the time interval and spatial displacement of consecutive knowledge creation activities of high-yield, low-yield, and all the bio-scientists are noted, respectively. Discussion points for this paper include:

(1) For low-yield scientists, the time interval distribution of continuous knowledge creation activities is approximate to exponential distribution at the tail segment, where the distribution is more random than that of high-yield scientists. The time interval of continuous knowledge creation activities of low-yield scientists is longer than that of high-yield scientists. One possible reason is that low-yield scientists have poorer systematical research planning and approaches than high-yield creators. The time interval distribution curves of low-yield and ASFP generally overlap at the tail segment. This indicates the self-similarity of these curves.

(2) The spatial distribution of knowledge creation activities embodies a heavy-tailed feature, and distribution curves of high-yield, low-yield, and ASFP generally overlap at the tail segment. Both long-distance field transfer and short-distance intensive creation occur during the continuous knowledge creation activities. One possible explanation for this phenomenon is that, after in-depth study in a certain field for a period of time, scientists begin to explore other knowledge fields. As a result, long-distance field exploration happens frequently. With respect to the dynamic model in this regard, the Adaptive Interest Model proposed by Han et al. (2008) can explain the phenomenon to some extent, but interest in a certain topic should be deemed as interest in multiple knowledge fields;

(3) For the high-yield scientists, the time interval of knowledge creation activities is unevenly distributed and embodies a heavy-tailed feature, with both long-time silence and sudden efficient output. One possible explanation is the periodic feature of human behaviors, such as when sudden inspiration is drawn from long-term accumulation and thinking, efficient output lasts for a certain period of time. Since this process consumes much energy, long-time accumulation, thinking, and relaxation are required once again before gaining another inspiration. This phenomenon can be explained by Cascaded Nonhomogeneous Poisson Process Model proposed by Malmgren et al. (2009). This model

constructs the cascade modes of the external nonhomogeneous Poisson process and the internal homogeneous Poisson process, and reflects the periodical, repetitive, and heavy-tailed features of human behavior. However, as shown in Fig. 1, the knowledge creation activities of high-yield scientists have no obvious periodical features.

(4) The walking range “ d ” is recorded as the distance between the authors’ current footprint and start footprint. This cannot, however, reflect the breadth of knowledge creation activities to any extent. For this reason, the distances between all other footprints and the start footprint are quite long, but the distances among these footprints (excluding the start footprint) are quite short. In this case, the knowledge creation activities of scientists are confined in a concentrated and narrow scope. So the practical value of “ d ” may be inconsistent with the statistical value of “ d ”; and

(5) Keyword vectors and reference vectors have limited lengths. So there are some limitations to defining knowledge creation activities using these two vectors. Generally speaking, a citing publication has no more than 10 keywords, and the current paper often has about 35 cited references (Biglu 2008). Both the keyword and reference vectors have limited lengths, so high deviation from the mean value seldom occurs in statistical results, and it is difficult to measure the range of spatial displacement.

Conclusion

The following conclusion can be drawn from the statistical analysis on temporal and spatial distribution of knowledge creation activities carried out by high-yield, low-yield, and ASFP groups, as well as the exploration on the human behavior rules during knowledge creation activities:

(1) For high-yield scientists, the time interval of knowledge creation activities obeys heavy-tailed distribution, embodying a bursting feature, with both a long-time silence and intensive burst of creation activities. The corresponding distribution of low-yield scientists is approximate to exponential distribution, and it is often randomly and occasionally distributed; and

(2) The spatial distribution of knowledge creation activities embodies a heavy-tailed feature. Knowledge creation activities are often intensively confined to a certain knowledge field. Long-distance exploration across the knowledge fields is also made in knowledge creation activities. For high-yield, low-yield, and ASFP groups, the spatial distribution curves have overlapped tails, where the spatial transfer of knowledge creation activities is self-similar.

The above findings provide some basis for grasping the laws of scientist creation activities and forecasting scientists’ creation behaviors. The time interval and spatial migration are uneven in high-yield scientists’ creation activities. They generally have no stable knowledge output rate and are not always limited in small domain research. Low-yield scientists’ creation activities is more randomness and sporadic in time. However, their spatial distributions are similar to high-yield scientists, where this phenomenon reflects the characteristics of collectivity and similarity of human knowledge creation activities.

This paper still has some shortcomings, such as the descriptions of 4 and 5 in the discussion section. In addition, the walking range of scientists is cast on the one-dimensional space in this report, and the relevant analysis is irrelevant to the background knowledge. A more effective way to approach this would be to analyze these distances against the background knowledge. Moreover, the statistics should be made on distances

between two consecutive footprints. In this paper, works published by authors include those in which the author is not the first author. No difference has thus been made between the author group and the author himself or herself. Further studies shall be carried out in this regard.

Acknowledgments This research is supported by the Project of National Natural Science Foundation of China, notably, “Research on Formation Mechanism and Evolution Laws of Knowledge Networks” (No.: 71173249).

References

- Barabási, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*, 207–211.
- Biglu, M. H. (2008). The influence of references per paper in the SCI to Impact Factors and the Matthew Effect. *Scientometrics*, *74*(3), 453–470.
- Blanchard, P., & Hongler, M. O. (2007). Modeling human activity in the spirit of Barabási’s queueing system. *Physical Review Letters*, *75*(2), 026102.
- Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, *439*, 462–465.
- Brookes, B. C. (1981). The foundations of information science IV. *Journal of Information Science*, *3*, 3–12.
- Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer.
- Fan, C., Guo, J. L., Ji, Y. L., et al. (2010). Analysis of human behavior scaling law based on library loans. *Library and Information Service*, *54*(15), 35–39.
- Gabrielli, A., & Caldarelli, G. (2007). Invasion percolation and critical transient in the Barabási model of human dynamics. *Physical Review Letters*, *98*(20), 208701.
- Gonzalez, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*, 779–782.
- Guo, J. L. (2011). Empirical study and modeling of human behavior dynamics of comments on blog posts. *Application Research of Computers*, *28*(4), 1422–1424.
- Han, X. P., Zhou, T., & Wang, B. H. (2008). Modeling human dynamics with adaptive interest. *New Journal of Physics*, *10*(7), 073010.
- Li, X. (2003). *Foundations of probability theory 2* (pp. 129–130). Beijing: High Education Press.
- Malmgren, R. D., Stouffer, D. B., Campanharo, A. S. L. O., et al. (2009). On universality in human correspondence activity. *Science*, *325*, 1696–1705.
- Nakamura, T., Kiyono, K., Yoshiuchi, K., et al. (2007). Universal scaling law in human behavioral organization. *Physical Review E*, *99*(13), 138103.
- Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation*, *60*(5), 503–520.
- Salton, G., & Lesk, M. E. (1968). Computer evaluation of indexing and text processing. *Journal of the ACM*, *15*(1), 8–36.
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill.
- Song, C. M., Qu, Z. H., Blumm, N., et al. (2010). Limits of predictability in human mobility. *Science*, *327*, 1018–1021.
- Viswanathan, G. M., Buldyrev, S. V., & Havlin, S. (1999). Optimizing the success of random searches. *Nature*, *401*, 911–914.
- Wang, B. H., & Han, X. P. (2010). The dynamics and statistical mechanics of human behaviors. *Physics*, *39*(1), 28–37.
- Zhou, T., Kiet, H. A. T., Kim, B. J., et al. (2008). Role of activity in human dynamics. *Europhysics Letters*, *82*(2), 28002.