



Science as exploration in a knowledge landscape: tracing hotspots or seeking opportunity?

Feifan Liu^{1,2}, Shuang Zhang^{1,3} and Haoxiang Xia^{1,2,3,4*} 

*Correspondence:
hxxia@dlut.edu.cn

¹Institute of Systems Engineering,
Dalian University of Technology,
No. 2 Linggong Road, Dalian,
116024, Liaoning, China

²Institute for Advanced Intelligence,
Dalian University of Technology,
Linggong Road, Dalian, 116024,
Liaoning, China

Full list of author information is
available at the end of the article

Abstract

The selection of research topics by scientists can be viewed as an exploration process conducted by individuals with cognitive limitations traversing a complex cognitive landscape influenced by both individual and social factors. While existing theoretical investigations have provided valuable insights, the intricate and multifaceted nature of modern science hinders the implementation of empirical experiments. This study leverages advancements in Geographic Information System (GIS) techniques to investigate the patterns and dynamic mechanisms of topic-transition among scientists. By constructing the knowledge space across 6 large-scale disciplines, we depict the trajectories of scientists' topic transitions within this space, measuring the flow and distance of research regions across different sub-spaces. Our findings reveal a predominantly conservative pattern of topic transition at the individual level, with scientists primarily exploring local knowledge spaces. Furthermore, simulation modeling analysis identifies research intensity, driven by the concentration of scientists within a specific region, as the key facilitator of topic transition. Conversely, the knowledge distance between fields serves as a significant barrier to exploration. Notably, despite potential opportunities for breakthrough discoveries at the intersection of subfields, empirical evidence suggests that these opportunities do not exert a strong pull on scientists, leading them to favor familiar research areas. Our study provides valuable insights into the exploration dynamics of scientific knowledge production, highlighting the influence of individual cognition, social factors, and the intrinsic structure of the knowledge landscape itself. These findings offer a framework for understanding and potentially shaping the course of scientific progress.

Keywords: Scientists' exploration; Knowledge space; Topic-transition behavior; Gravity model; Radiation model

1 Introduction

Throughout their academic careers, scientists must confront a multitude of choices when it comes to selecting their research topics. These decisions wield a substantial influence over their academic productivity, impact, and overall career trajectory. Nobel laureate Chen Ning Yang shared a valuable insight during a symposium at the University of Chi-

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

nese Academy of Sciences [1]. He emphasized that, particularly for emerging scientists, the decision to persist in a particular field may not directly dictate their career's level of achievement. However, the wise selection of research topics and research directions holds paramount significance. In his words,

Pursuing a direction that leads to an impasse can be a treacherous endeavor, as the deeper one delves, the more arduous it becomes to alter course. Diverting from an unproductive trajectory is no simple feat, making persistence in a barren direction a most regrettable choice.

On a broader scale, the choices made by scientists in terms of topic selection and transition impact the development of the entire scientific ecosystem. Understanding the intricate motivations and multifaceted influences that guide scientists' decisions in the process of selecting research topics presents a substantial challenge in unraveling the behavioral patterns and internal mechanisms that underlie these choices.

Scientists' choices of topics can be illuminated as the persistent endeavors of cognitively constrained individuals within the intricate expanse of knowledge [2]. This pursuit adheres to the principle of "no free lunch". Owing to the inherent tension between accumulating academic accomplishments and fostering innovation, scientists grapple with the delicate task of balancing conventional and pioneering research fields [3]. Diverse strategies employed in the process of topic selection yield markedly distinct outcomes, impacting both personal development [4–6] and scientific progress [7]. Consequently, various levels of behavioral risk must be contemplated. To unravel these intrinsic conundrums, prior investigations have empirically validated and dissected the trade-offs scientists encounter during their exploration, focusing primarily on individual scientists' topic selection and their relationship with academic performance within their respective research fields [8, 9].

The exploration within the realm of knowledge reflects a complex interplay of scientists' decision-making behaviors. The selection of research topics is shaped by individual volition and concurrently influenced by the collective dynamics within the specific knowledge field. In contrast to the early days of modern scientific development, characterized by a limited number of scientists who primarily pursued research based on personal interests, contemporary scientific progress has witnessed a proliferation of participants and a diversification of topic matter [10, 11]. This expansion inevitably renders the process of selecting research topics susceptible to the impact of social factors. As government entities, corporations, and diverse social organizations have increasingly assumed central roles in funding scientific research, the defining characteristics of the scientific establishment have become more pronounced. In this era of 'big science', scientists' choice of topics is not solely propelled by personal aspirations and inclinations. It is equally shaped by a spectrum of social behaviors such as following, learning, emulating, and conforming to prevailing trends.

Aligning research interests within scholarly groups has the potential to accelerate scientific outputs, increase scholarly impact, and improve access to scholarly resources. This, in turn, serves the advancement of individual scholarly careers. However, it is important to remain vigilant that the advancement of science depends on groundbreaking discoveries and trendsetting contributions. An overemphasis on conforming to popular trends and

crowd-sourced research selection may lead to stagnation within the broader scientific research and innovation ecosystem [12], potentially resulting in a scenario where resources are allocated without commensurate progress.

The central question is whether scientists should opt for popular research areas that attract widespread attention or explore an uncultivated territory of research fields. It concerns the patterns of behavior that scientists exhibit when moving between topics within or across the research field. Can these patterns be quantified and further explained by a simple mechanistic model of group behavior? A comprehensive understanding of these issues can shed light on the strategic choices and risk preferences of scientists, provide deep insights into the underlying mechanisms of scientific development, and serve as a valuable basis for the design of research management policies.

To gain a deeper understanding of knowledge spaces and scientists' exploratory behaviors within them, we draw inspiration from Geographic Information Systems principles. The analysis of human mobility patterns in physical space has provided valuable insights [13]. Recent advancements in machine learning, especially in representation learning algorithms, have opened up opportunities for measuring knowledge distance between research subfields and help us better quantify the intricate and abstract knowledge spaces of disciplines [14], underpinning the empirical study of the collective mobility behavior of scientists.

Therefore, to bridge the gap in understanding scientists' topic selection and transition patterns at the population level, this study builds on the foundation of constructing a scientific knowledge space as a research field map, and attempts to integrate complex network analysis methods, machine learning algorithms, and geographic information analysis theories to understand the collective knowledge creation process in the scientific ecosystem. The main research contributions of this paper are as follows:

(1) Within the framework of constructing a knowledge space, scientists' papers are embedded in this space based on the topical distance. The knowledge space is partitioned into the grid and the Voronoi diagram subfields, using both equidistant and equal-density approaches. Scientists' trajectories are constituted of published papers and merge into Origin-Destination (OD) flows that effectively encapsulate scientists' exploration patterns in the knowledge space. The analysis of these topic selection and transition trajectories, when rooted in the entire scientific field space, provides novel insights for quantifying scientists' topic-changing. Including activities such as online socializing, web searching, and gaming, all of which involve complex and abstract spaces, the methodological approach in this study can potentially be extended to quantify individual-level or population-level mobility in virtual spaces with fine granularity.

(2) When exploring the flow of scientists' publication trajectories across different subfields within the knowledge space, it is evident that the distance traveled by scientists as they move between topics follows a log-normal distribution. This observation is particularly pronounced in the context of Voronoi diagram-based field partitioning. This broad, "heavy-tailed" distribution suggests that scientists' inter-field movement patterns, while predominantly characterized by short-range transits, also include occasional long-range transitions. It is noteworthy, however, that these patterns do not exhibit a "scale-free" behavior, underscoring that the majority of scientists tend to change their subfields with cautious, short-range transits.

(3) Intriguingly, the study reveals that the gravity model, which takes into account factors such as population size and the distances between starting and ending points, offers a more robust explanation and prediction of scientists' topic selection and transition within the knowledge space. In the quest to unravel the underlying mechanisms governing scientists' topic-transition patterns at the group level, this study introduces two distinct group exploration models: the distance-based "gravity" model and the opportunity-based "radiation" model. Our finding implies that the fundamental driving force behind scientists' topic selection and change is the research hotspots generated by the density of scientists in a given region. Conversely, the inhibiting factor is the knowledge distance between distinct fields. While research opportunities may exist at the intersection of subfields, this factor does not significantly influence scientists' decisions to change their research focus.

In Sect. 2, we describe the use of the dataset, the framework for constructing a knowledge space, the tessellated diagram types of spatial partitioning, the gravity model, the radiation model, and corresponding evaluation metrics. In Sect. 3, we use complex network and representation learning techniques to construct a knowledge space for physics using the American Physical Society (APS) dataset and identify paper positions. We then use the grid and Voronoi diagram to delineate sub-field regions, capturing the population-level mobility of scientists in the knowledge spaces. To disclose the underlying mechanism of scientists' inter-field OD flow, we introduce the gravity model and the radiation model. Then we test and validate the explanatory and predictive capabilities of these models on the mobile patterns of scientists in the knowledge space. In Sect. 4, we discuss our findings with studies on human mobility patterns in real and virtual spaces and other related works. Finally, in Sect. 5, we summarize our main findings, highlight research limitations, and suggest future directions.

2 Materials and methods

2.1 Dataset

The major part of this paper focuses on the field of physics and utilizes the journal literature dataset provided by the APS [15]. In exploring the topic-transition behavior patterns of scientists, more than 258,000 papers published in APS journals from 1985 to 2009 were used. Taking into account the impact of authors and the percentage of the number of papers, 13,720 scientists in the field of physics with more than or equal to 16 publications, involving 450,290 publication records, were eventually selected. Author and paper records were preprocessed and provided by Sinatra et al [16]. The selection of scientists is based on the fact that although the number of scientists with 16 or more publications accounts for only 13.1% (13,720/104,483) of the dataset of this study, the number of their papers accounts for 82.4% (209,473/254,117).

Our findings have also been further extended to Computer Science, Chemistry, Biology, Social Science, and Multidisciplinary Science with Microsoft Academic Graph (MAG) [17]. Leveraging the comprehensive "fields of study" classification system provided by the MAG [18], we extract a dataset encompassing 4,752,206 authors and 4,391,220 papers associated with the label "Computer Science", spanning from 1948 to 2019. Subsequently, we focus on a subset of 180,339 highly productive scientists, each with a minimum of 10 published papers within the domain. The Chemistry dataset encompassed 9,568,741 authors and 6,916,260 papers labeled "Chemistry", covering the period until 2019. We focus our analysis on 117,960 prolific scientists who had published at least 30 papers, totally involved

with 4,048,890 papers. The Biology dataset, comprising 9,731,092 authors and 7,157,231 papers categorized as “Biology” in MAG, covered the same timeframe. We finally identify 164,871 highly active scientists, whose papers count greater than or equal to 30, and their 4,701,836 papers. The Social Science dataset consisted of 740,196 authors and 765,709 papers published in journals belonging to the SAGE publishing group, spanning the period from 1965 to 2019. Our analysis focuses on 19,105 scientists, whose number of published papers is larger than or equal to 10, and their 237,278 papers in this domain. Furthermore, we construct a multidisciplinary dataset encompassing scientific publications from five prominent journals representing diverse research areas: Nature, Science, Proceedings of the National Academy of Sciences, Nature Communications, and Science Advances. This dataset comprises 948,180 authors and 562,998 papers published between 1869 and 2019. We identify 22,842 scientists, who had published at least 10 papers, contributing to a collective body of 295,888 papers in this area.

2.2 Construction of knowledge space

In the context of the scientific innovation system, a crucial aspect of the collective behavior of scientists corresponds to their decisions and transitions in research directions within the epistemic landscape. The establishment of an accurate and valid knowledge space serves as the basis for determining the distance at which scientists’ interests change. Given the stable characteristic of most physical subfields [19], we construct a knowledge network of physics disciplines by utilizing the co-occurrence relationship between Physics and astronomy classification scheme(PACS)codes and their co-occurrence frequency in each paper published in APS journals. This network consists of 874 secondary PACS codes as nodes and co-occurrence relationships between PACS codes as connected edges. Considering the elimination of the influence of the absolute difference in frequency between PACS codes, we further take the square root of the inverse of the joint probability of PACS code i and PACS code j appearing in a paper at the same time as the weight value w_{ij} of the network, and the calculation process is shown in Eq. (1):

$$w_{ij} = \frac{1}{\sqrt{\left(\frac{f_{ij}}{f_i} \cdot \frac{f_{ij}}{f_j}\right)}} = \frac{\sqrt{f_{ij}}}{f_{ij}}, \quad (1)$$

where the f_i and f_j are the cumulative edge frequencies in the network connected to node i and node j , respectively. The network’s modularity, calculated at approximately 0.506 through a community detection algorithm [20], signifies the presence of distinct community structures within the field of physics. This implies that physics can be divided into several closely related subfields with relatively sparse interconnections between them. We then apply Node2Vec [21] and the UMAP manifold learning algorithm [22] to create a knowledge map of physics.

Furthermore, to eliminate the potential influence of choosing representation methods for our observed patterns in this study, we utilize Doc2Vec [23], a widely used document embedding technique, to extract high-dimensional features from the title and abstract of research papers belonging to the other five disciplines. This approach ensures consistency across different disciplines and minimizes bias introduced by specific representation learning methods. The constructed map represents the research field and benefits from

representation learning to uncover knowledge structure and manifold learning for virtual spatial analysis. Overall, this approach facilitates embedding and visualizing the scientific landscape and offers a foundation for quantifying scientific research movements within the knowledge space.

2.3 Tessellated models of space: grid and Voronoi diagram

To comprehensively analyze the topic selection and transition of scientists, the following step involves partitioning the knowledge space into distinct regions and identifying the “geographic units”. In real-world geographic spaces, people often adopt administrative districts as their fundamental research units. However, these pre-defined districts do not exist within the realm of knowledge spaces. Consequently, in this section, the knowledge space is divided into spatial regions based on the principles of “equal distance” and “equal density”, with subsequent comparison of scientists’ behavioral patterns. Tessellated models of space, including grid and voronoi diagrams, serve as potent tools for the representation and analysis of spatial arrangements [24]. They offer a unified research framework for comprehending the knowledge space. In this study, we employ those two distinct spatial region delineation approaches to understand the impact of the knowledge space delineation method on our research conclusions.

The grid diagram approach involves partitioning the entire knowledge field map into a series of grid regions, with each grid region spanning a 1° interval in knowledge space. This results in a total of 90 grid regions arranged in a 10×9 configuration, of which 73 available non-empty grid regions were associated with the specific research areas addressed in this study.

On the other hand, the voronoi partitioning approach utilizes the spatial distribution of high-frequency PACS codes within co-occurrence networks to define the knowledge space. Initially, we identify the top 10 high-frequency PACS codes within each subfield region and designate their centroid positions as the focal points in the voronoi diagram field. These 90 positions were instrumental in generating the boundaries of the voronoi diagram.

The main difference between these two methods is their spatial division approach. The grid diagram method divides space into uniform grid points, maintaining an isometric structure. On the other hand, the voronoi diagram, determined by the high-frequency PACS code, divides space based on isodensity, aligning with the heterogeneous distribution of the population. In this study, we will perform statistical analyses of scientists’ group mobility OD flows and use predictive modeling to analyze trajectory patterns under both tessellated modes of knowledge spatial region.

2.4 Models of OD flow prediction: gravity model and radiation model

The measure of “OD flow distance” is based on the geographical distance between two points, a metric extensively applied in the field of human mobility research [25, 26]. Whenever this study involves operations that depend on distance or area, we consistently employ a projected coordinate reference system (CRS) with the authority code “EPSG:4326”. This ensures that all operations are conducted on a plane.

The Gravity Model [27] and the Radiation Model [28] are two prominent mathematical models employed in human mobility and migration studies. These models aim to elucidate the population-level patterns of movement between different locations. The Gravity

Model is predominantly distance-based, while the Radiation Model additionally incorporates factors like competition for destinations and accessibility.

Specifically, the gravity model, inspired by Newton's gravitational formula, suggests that the flow of exploration by groups in different regions is directly proportional to the size of the regional group and inversely proportional to the square of the distance accessible between regions [29]. The model was also the firstly used in the field of geography to explain group migration. The mathematical expression of the general gravity model is shown in Eq. (2):

$$T_{ij} = \frac{(m_i^\alpha)(n_j^\beta)}{f(d_{ij})}, \quad (2)$$

where T_{ij} denotes the flow of people between location i and location j , m_i and n_j denote the total population of location i and location j , respectively. d_{ij} denotes the distance between locations i and j . α and β are adjustable exponential variables and $\alpha = \beta = 1$ in our settings to keep the gravity model simple. $f(d_{ij})$ is a damping function set according to different empirical data, such as a power-law function $f(d_{ij}) = d_{ij}^\gamma$ or exponential function $f(d_{ij}) = e^{(\gamma \cdot d_{ij})}$. Depending on the constraints, gravity models can also be categorized into models under one-way and two-way constraints. This type of constrained model can more accurately estimate and predict total inter-regional flows by fixing the population from location i to location j (output model) or the number of people entering (attraction model). The gravity model estimates the parameters using the flow data provided as input, employing a Generalized Linear Model (GLM) that utilizes Poisson regression, as introduced in [13, 30].

Inspired by the opportunity model, Simini et al. [28] propose a radiation model that more accurately predicts population movement. They claim that the radiation model not only predicts the average flow between two locations but also captures the variability of the flow compared to the gravity model. Specifically, the mathematical expression of the radiation model is given in Eq. (3):

$$\langle T_{ij} \rangle = \frac{T_i(m_i n_j)}{(m_i + s_{ij})(m_i + n_j + s_{ij})}, \quad (3)$$

where $\langle T_{ij} \rangle$ denotes the average population flow between location i and location j and $T_i \equiv \sum_{(j \neq i)} T_{ij}$. Compared to the gravity model, an additional parameter s_{ij} has been introduced. This parameter represents the population (or employment opportunities) outside of locations i and j within a distance of d_{ij} . It signifies the potential opportunities within the range from location i to location j that attract people to move.

The gravity model is a one-way constraint model that predetermines the population size at the origin while incorporating power-law and exponential damping functions to capture varying distance effects. In contrast, the radiation model is a parameter-free model, and we directly apply Eq. (3) for conducting simulation experiments.

2.5 The evaluation metrics of the population-level human mobility model

To quantify the performance of population-level models in this study, we then introduce a set of evaluation metrics. Human mobility model evaluation metrics are specifically designed to gauge the level of consistency between a model and actual human mobility data

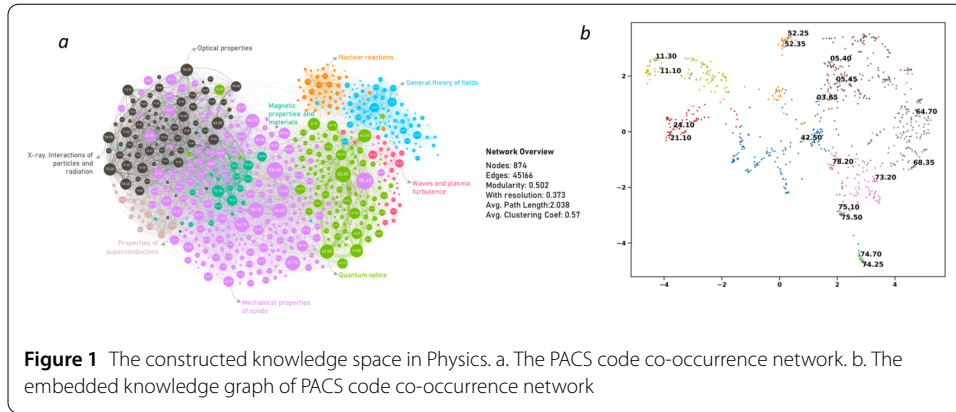


Figure 1 The constructed knowledge space in Physics. a. The PACS code co-occurrence network. b. The embedded knowledge graph of PACS code co-occurrence network

within spatial contexts. Beyond the common metrics such as R-squared, root mean square error, Spearman’s correlation coefficient, and Pearson’s correlation coefficient, the evaluation metrics for human mobility behavior models also encompass distinctive measures for assessing the convergence of human mobile activities [31].

These measures include the Common Part of Commuters (CPC), which quantifies the proportion of individuals with overlapping trajectories, the Common Part of Commuters’ Distance (CPC_d), which represents the fraction of overlapping distances traveled, and the Common Part of Links (CPL), which indicates the extent of overlap in mobility paths. Detailed formulas for computing these three metrics can be found in Eqs. (4)–(6):

$$CPC(T, \tilde{T}) = \frac{\sum_{(i,j=1)}^n \min(T_{ij}, \tilde{T}_{ij})}{N} = 1 - \frac{1}{2} \frac{\sum_{(i,j=1)}^n |T_{ij} - \tilde{T}_{ij}|}{N}, \tag{4}$$

$$CPC_d(T, \tilde{T}) = \frac{\sum_{(k=1)}^\infty \min(N_k, \tilde{N}_k)}{N}, \tag{5}$$

$$CPL(T, \tilde{T}) = \frac{2 \sum_{(i,j=1)}^n \mathbf{1}_{(T_{ij}>0)} \cdot \mathbf{1}_{(\tilde{T}_{ij}>0)}}{\sum_{(i,j=1)}^n \mathbf{1}_{(T_{ij}>0)} + \sum_{(i,j=1)}^n \mathbf{1}_{(\tilde{T}_{ij}>0)}}. \tag{6}$$

Among the three formulas mentioned earlier, the symbols T and \tilde{T} represent the actual flow and model-predicted flow values between locations *i* and *j*, respectively. *N* refers to the overall population flow, while *N_k* denotes the number of individual movements occurring between distances in the range of 2*k*-2 to 2*k*. The variable 1_{*x*} takes on a value of 1 when condition *x* is met, and it is 0 otherwise.

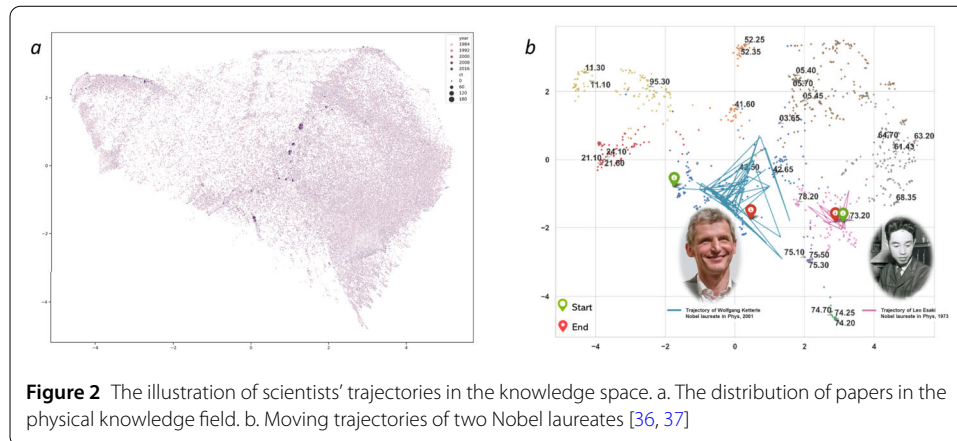
These indicators evaluate the precision of the model’s fitting or predictions, considering three essential factors: the population size, the knowledge distance, and the particular routes. These scores are instrumental in identifying the model’s strengths and limitations, as well as its adaptability for a specific human movement context at the population level.

3 Results

3.1 Knowledge space and trajectories in physics

Using the embedded PACS code co-occurrence network as a foundation, we create a knowledge space within the field of physics. By merging the node PACS code labels and the community tagging data, the results are depicted in Fig. 1.

In Fig. 1(a), the physical subfields that share a community not only show remarkable proximity but also exhibit distinct clustering characteristics on the knowledge map. Each

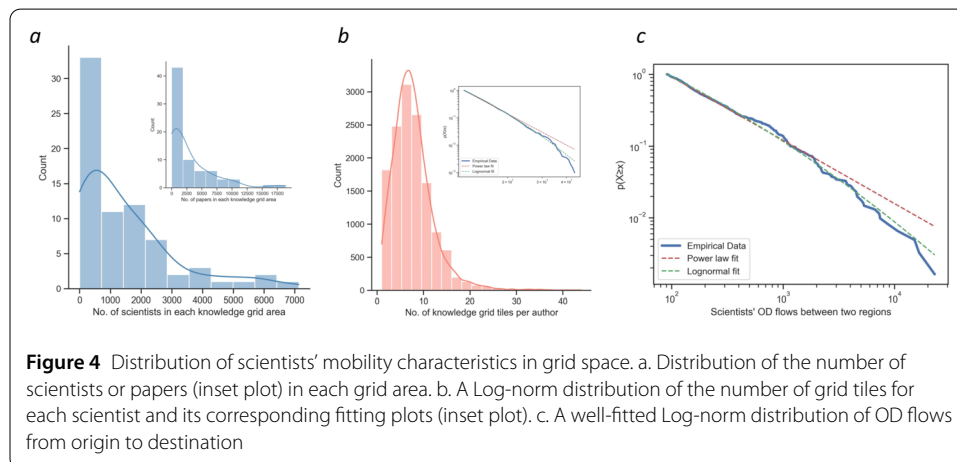
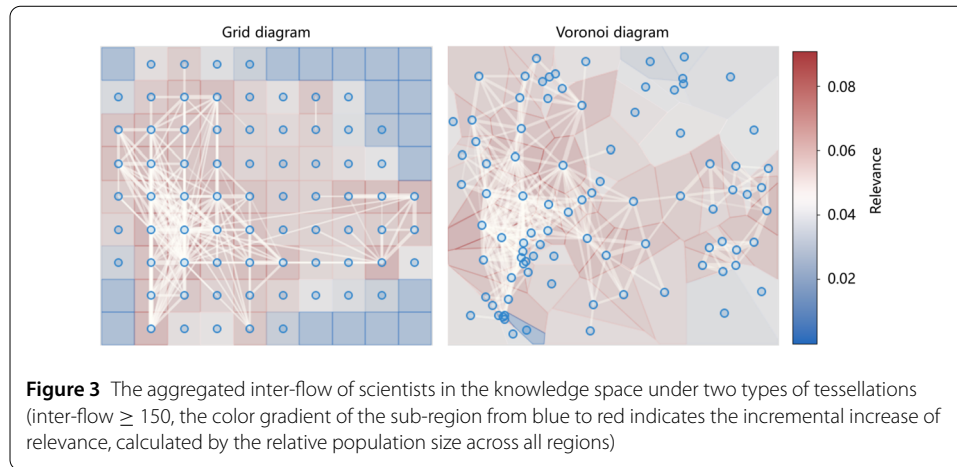


node in Fig. 1(a) corresponds to a PACS code, where the node's size is determined by the number of connecting edges. The nodes are distinguished by different colors representing the identified 9 subfields. In this context, a higher co-occurrence frequency between PACS codes translates into a shorter distance in the network, thus indicating a closer knowledge relationship between those specific PACS codes. This is evident in the network as nodes belonging to the same community or a particular subfield are grouped closely together.

In addition, the knowledge space is established based on the Node2vec algorithm with the parameters of dimensions = 64, walk length = 30, and number of walks = 200. We further test the stability of the node2vec algorithm with various parameters and metrics provided in [32]. As shown in Fig. 1(b), it effectively preserves the distinctions between different subfields. For example, the left side of the overall space is dominated by subfields related to condensed matter and statistical physics, and the right side is characterized by two subfields representing nuclear physics and astrophysics. It demonstrates that it is reasonable and effective to use the graph-embedded method to construct a knowledge map of physics.

After establishing the PACS code coordinates, we extract labeling information connecting authors' papers with PACS codes. Using this data, we calculate the center of mass for each paper, allowing us to position them on the knowledge map.

The distribution of papers in the physical field within the knowledge space is depicted in Fig. 2. In the knowledge map of Fig. 2(a), scattered dots represent papers and colors indicating 9 subfields in physics. The topological structure of the field knowledge space, along with the location information of each paper on the map, serves as the foundational basis for quantitatively analyzing scientists' topic-transition. Figure 2(b) illustrates the publication trajectories of two Nobel Prize laureates, Wolfgang Ketterle (left, blue) and Leo Esaki (right, pink), within the physics field knowledge space. Wolfgang Ketterle's Nobel Prize-winning contributions are in the realm of trapping cold atoms and reaching absolute zero, fundamental to the study of condensed matter within atomic physics. By observing his publication trajectory, we observe that his research encompasses nearly all subspaces of atomic physics. Leo Esaki's significant accomplishment lies in the discovery of the quantum tunneling effect in semiconductor materials, a key component of the superconductivity subfield in physics. In contrast to Wolfgang, Esaki's scientific exploration appears more focused on his research trajectory.

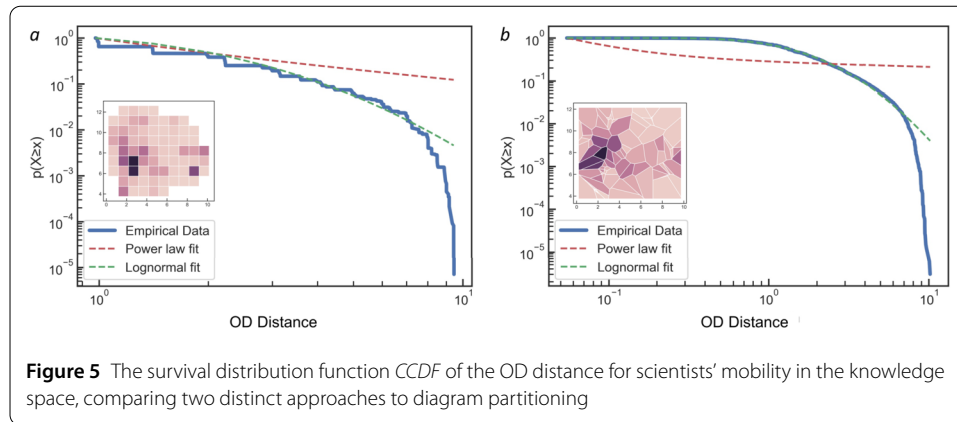


These findings underscore the divergent topic-transition trajectories of scientists within physics, despite their significant contributions to the field. This variation is likely attributed to the distinct research fields they inhabit. For the physics community as a whole, it remains fascinating to unravel the statistical patterns governing the selection and transition of research topics.

3.2 The non-scale-free pattern of the aggregated inter-flow of scientists in the knowledge space

When a researcher's paper transitions from one region of knowledge space to another, we can trace a sequence of origin and destination points within the region, mapping a trajectory from point i to point j . As we introduced before, we employ a partitioning of the knowledge space into two categories: the grid diagram and the Voronoi diagram, following the spatial division principles of Geographic Information System analysis.

Figure 3 illustrates these divisions: solid lines demarcate boundaries, circles signify central positions, while white connecting edges represent OD flows between regions, where the volume of flow is larger than 150. In addition, the color gradient of the sub-region from blue to red indicates the incremental increase in relative population size, *relevance*, compared to the other regions.



In Fig. 4, we present essential statistics on scientists' mobility within a grid space. It includes the distribution of the number of scientists or papers at each grid region (see Fig. 4(a)), the distribution of the number of scientists' knowledge tiles (see Fig. 4(b)), and OD flows between two regions (see Fig. 4(c)). Moreover, as depicted in the inset plot of Fig. 4(b) and Fig. 4(c), using the power-law distribution fitting method proposed by Alstott et al. [33], our analysis reveals that the number of grid tiles associated with each scientist, and the corresponding OD flow patterns, exhibit log-normal distributions rather than scale-free characteristics.

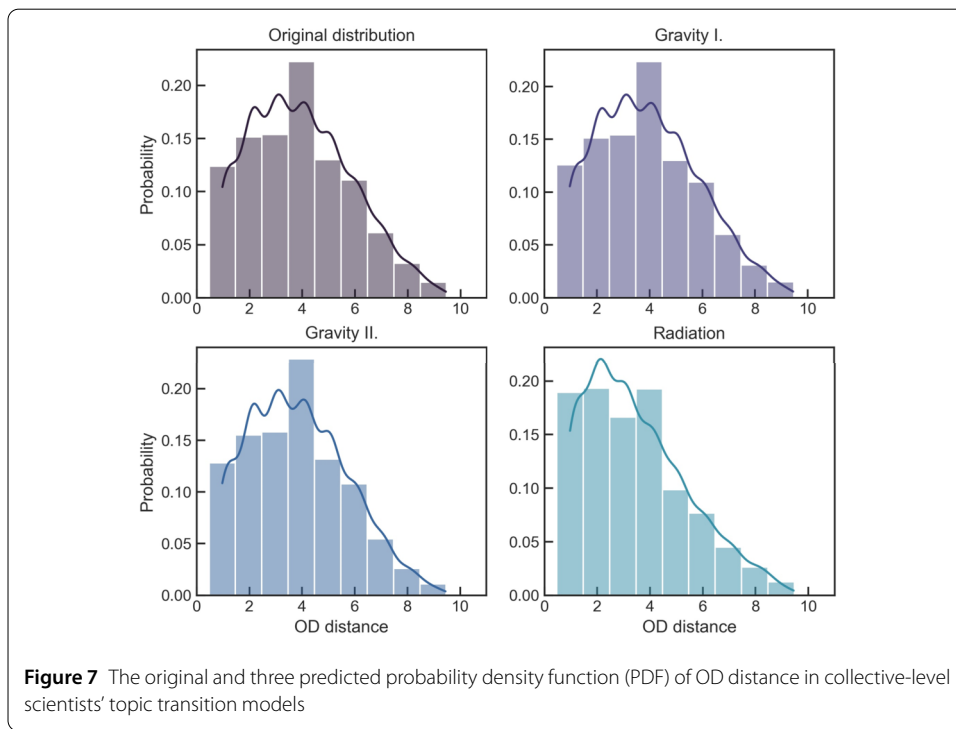
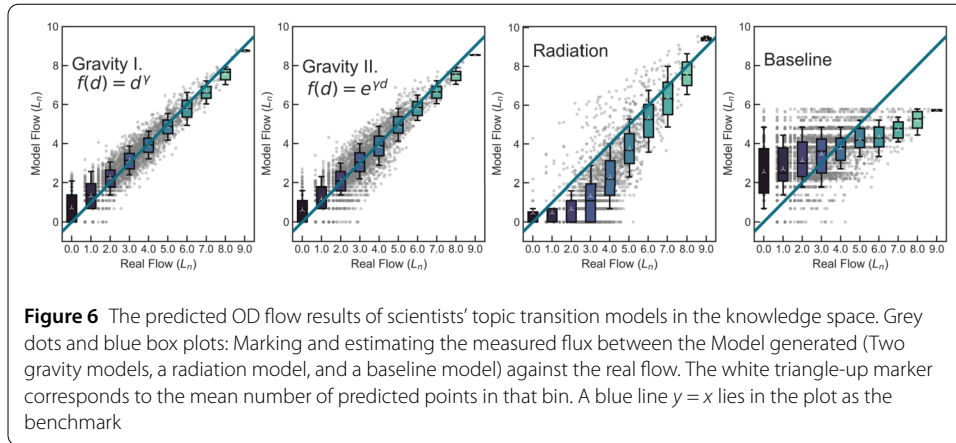
Figure 5(a)-(b) depicts the distribution of OD flow distances originating from and ending at scientists' locations under grid and Voronoi diagram partitioning methods. We also apply power-law and log-normal function fitting to the Complementary Cumulative Distribution Function (*CCDF*) of these OD flow distances. Furthermore, the insets in Fig. 5 illustrate the density distribution of people within each spatial region.

Our analysis reveals that scientists' OD flow distance distribution exhibited more log-normal features than power-law characteristics under both the grid diagram and the Voronoi diagram methods. Notably, the Voronoi diagram partitioning method yields superior log-normal distribution fitting results compared to the power-law fit. This heavy-tailed distribution suggests that scientists' inter-field exploration patterns are not notably 'scale-free', despite being characterized by short-distance transitions for the majority and long-distance transitions for the minority.

3.3 Models of scientists' topic-transition behavioral patterns

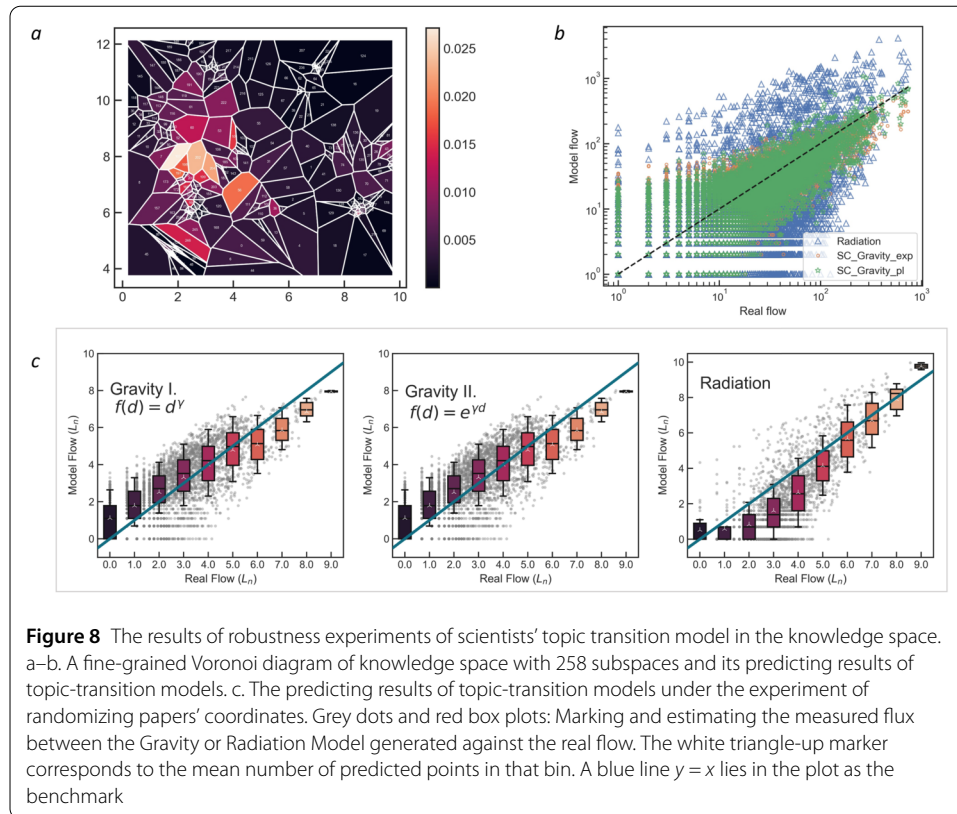
Delving into the social factors that influence scientists' decisions to change their research topics is key to understanding the dynamics of scientific progress. To what extent can we predict scientists' topic-transition? Addressing this question requires a deep exploration of the behavioral mechanisms underlying group-level mobility patterns within the knowledge space. Building upon the established knowledge space and scientists' publication trajectories, we introduce two models within the framework of GIS analysis methodology: the gravity model and the radiation model.

Figure 6 presents a comparison between actual OD flows and model-predicted flows across various types and parameters of population-level models. Gray points represent the correspondence level between observed and predicted flows for scientist topic-transition behaviors at each pair of starting and ending points. Box plots illustrate the 0.5-fold interquartile ranges, offering insights into data concentration intervals. White upward tri-



angular symbols pinpoint the mean values of this dataset, and a green diagonal reference line represents a perfect alignment between actual and model results. The baseline model, where the damping function employs a γ parameter set to 0, effectively nullifying the impact of distance difference, performs the poorest in prediction accuracy. In contrast, both gravity models outperform the radiation model. The Box plot reveals that the exponential damping function in the gravity model yields superior predictions compared to the power-law damping function.

Figure 7 displays the observed OD distance density distributions in knowledge space alongside three model-predicted distributions. Our analysis reveals that the gravity model again offers a superior capability of explanations and predictions for the patterns of scientists' topic-transition within the knowledge space, compared to the radiation model. To ensure the robustness and consistency of our findings, we conduct experiments involving



adjustments to the division scale of the field knowledge space and introduce randomized experiments in various contexts. These results serve to scrutinize the model predictions further.

In our scale-reconfiguration experiments (see Fig. 8), we alter the scale of subfield regions by different multiples and subsequently reevaluate the topic-transition pattern of scientists as well as the predictions from the simulation model. Figure 8(a) illustrates the subdivision of the Voronoi diagram into smaller segments, expanding the high-frequency 10 PACS codes from each subfield community to 30 PACS codes, creating 258 non-empty subspaces. Figure 8(b) compares actual OD flows with model predictions at this scale setting, showing the continued superiority of gravity models over the radiation model.

In the null model experiments, three scenarios were tested: (1) randomizing authors' publication date order to remove sequential timing effects, (2) random perturbation of paper coordinates in the knowledge space, and (3) maintaining the author's publication frequency while randomly selecting the same number of papers. Figure 8(c) demonstrates the diminished results of scenario 2 in the randomized experiment, highlighting the significance of keeping original publication coordinates in the knowledge space for predicting OD flows. The simulation results of scenarios 1 and 3 are not shown but close to scenario 2.

It's important to note that the key distinction between the gravity and radiation models lies in key factors that drive scientists' mobility in the knowledge space. The gravity model emphasizes the impact of distance between subfield regions on topical transition, while the radiation model focuses on attraction or repulsion based on potential research gaps between subfield regions. Our findings suggest that the distance between subfields and

the number of scientists in research subfields significantly influence scientists' movement more than the potential research 'opportunities' between subfields. Although peripheral research areas between subfield regions are crucial for scientific progress but pose risks, as their outcome is unpredictable. This uncertainty may contribute to the radiation model's reduced predictive accuracy, while the gravity model aligns with most scientists' conservative and 'hot-spot-tracing' research strategy when selecting or transiting research topics.

3.4 Null model experiments and robustness test of results

We systematically assess the effect of different parameters or experimental settings on model performance, including subfield region division granularity, damping function types, and randomized permutations in authors' trajectories. In addition, we introduce multiple model evaluation indices to compare experimental results comprehensively.

As summarized in Table 1, we deploy experiments with specific groups to evaluate model predictions against real results under various experimental conditions. Experiment groups 1–4 and 17–20 correspond to basic experimental settings depicted in Figs. 5–8. Experiment groups 5–10 and 21–26 involve randomized experiments with grid-based diagram and Voronoi-based diagram division, respectively, aligning with the above null model experiments. Experiment groups 11–16 explore model evaluation with grid region granularity reduced and expanded by a factor of 1. Experiment groups 27–32 pertain to modeling the Voronoi diagram subregions, involving adjustments to the number of high-frequency PACS codes and corresponding sub-regions. Furthermore, we consider the impact of coordinate scale transformations on experimental predictions, with experiments 33–35 representing scaled experiments.

Cross-validating across different model evaluation metrics minimizes bias inherent to a single metric. Of particular interest is the CPC indicator, widely used in the studies of human mobility behavior at the collective level, measuring explorer's overlap trajectories between origins and destinations in real or model-predicted data. The colored (blue) number indicates the best-performing results of the models within the same group. The group is categorized based on data input and model settings, including the baseline model(BSL), publishing order randomized experiments(Rand), and tessellation scaled experiments(Scale). By comparing various model evaluation metrics in Table 1, we deduce five key findings:

(1) Regardless of the grid partition type and subregion granularity, two gravity models significantly outperform the radiation model, predicting over 30% more real OD flows and 25% more trajectories. The baseline model, which does not consider distance factors, produces the poorest predictive results, with CPC indices of only 0.391 and 0.424 in the grid and Voronoi diagram cases, respectively.

(2) In the scale experiments, while the predictive power of the gravity model decreases with a smaller unit area granularity and increases with a larger granularity, overall, the scaling of the model does not significantly impact predictive performance. The minimum CPC index remains around 0.75.

(3) Regarding the three sets of null model experiments, only the model generated by shuffling the order of authors' publications shows a slight decrease in predictive performance compared to the baseline model, with a decrease of only 0.01 in the CPC index. However, the two models created by randomly shuffling all paper coordinates exhibit a noticeable drop in predictive performance for real OD flows, with a reduction of 0.23 in the CPC index.

Table 1 The aggregated results of the evaluation indexes of two population-level models and null models

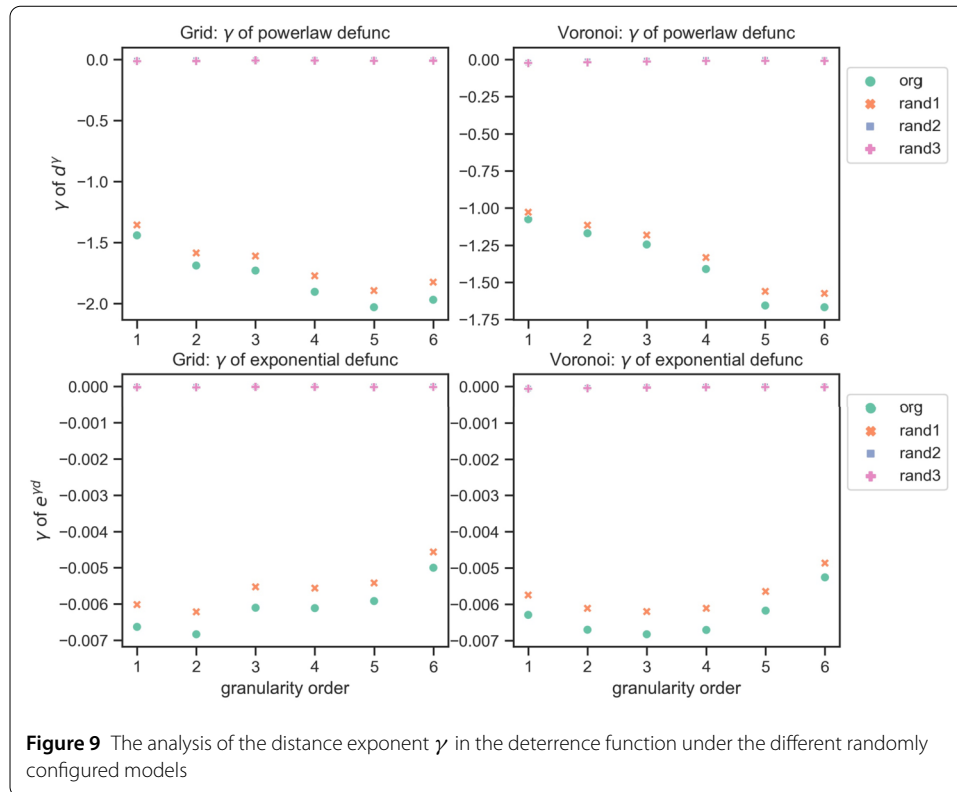
Id	Network Type	Robust Exp.	Model	Model Para.	R^2	RMSE	Spearman Coef.	Pearson Coef.	CPC	CPCd	CPL
1	Grid	BSL	Baseline	-	0.088	269.57	0.475	0.324	0.391	0.006	1
2	Grid	BSL	Gravity	exp	0.888	94.625	0.906	0.944	0.8	0.011	1
3	Grid	BSL	Gravity	pl	0.887	94.778	0.935	0.943	0.82	0.011	1
4	Grid	BSL	Radiation	-	-0.541	350.498	0.843	0.729	0.534	0.007	0.755
5	Grid	Rand1	Gravity	exp	0.878	98.478	0.913	0.94	0.794	0.011	1
6	Grid	Rand1	Gravity	pl	0.892	92.805	0.932	0.945	0.818	0.011	1
7	Grid	Rand2	Gravity	exp	0.527	1914	0.761	0.73	0.572	0.011	1
8	Grid	Rand2	Gravity	pl	0.527	191	0.761	0.731	0.572	0.011	1
9	Grid	Rand3	Gravity	exp	0.529	193.76	0.762	0.732	0.573	0.011	1
10	Grid	Rand3	Gravity	pl	0.53	193.635	0.762	0.733	0.573	0.011	1
11	Grid	Scale1	Gravity	exp	0.828	22.865	0.775	0.919	0.76	0.065	1
12	Grid	Scale1	Gravity	pl	0.866	20.201	0.797	0.931	0.785	0.069	1
13	Grid	Scale1	Radiation	-	-2.284	100.007	0.645	0.67	0.431	0.039	0.494
14	Grid	Scale2	Gravity	exp	0.878	502.72	0.954	0.938	0.839	0.001	1
15	Grid	Scale2	Gravity	pl	0.851	555.99	0.968	0.923	0.842	0.001	1
16	Grid	Scale2	Radiation	-	0.18	1304.34	0.92	0.713	0.603	0.001	0.923
17	Voronoi	BSL	Baseline	-	0.092	128.097	0.421	0.304	0.424	0.008	1
18	Voronoi	BSL	Gravity	exp	0.836	54.514	0.857	0.917	0.79	0.017	1
19	Voronoi	BSL	Gravity	pl	0.769	64.61	0.866	0.879	0.77	0.018	1
20	Voronoi	BSL	Radiation	-	-1.685	220.29	0.827	0.679	0.488	0.011	0.761
21	Voronoi	Rand1	Gravity	exp	0.821	56.869	0.863	0.911	0.784	0.018	1
22	Voronoi	Rand1	Gravity	pl	0.781	62.963	0.866	0.884	0.768	0.018	1
23	Voronoi	Rand2	Gravity	exp	0.4	104.114	0.648	0.642	0.539	0.015	1
24	Voronoi	Rand2	Gravity	pl	0.401	104.039	0.648	0.642	0.539	0.015	1
25	Voronoi	Rand3	Gravity	exp	0.401	103.997	0.649	0.643	0.539	0.015	1
26	Voronoi	Rand3	Gravity	pl	0.403	103.886	0.649	0.644	0.54	0.015	1
27	Voronoi	Scale1	Gravity	exp	0.745	14.502	0.766	0.874	0.746	0.08	1
28	Voronoi	Scale1	Gravity	pl	0.708	15.523	0.76	0.845	0.748	0.088	1
29	Voronoi	Scale1	Radiation	-	-5.169	71.326	0.669	0.581	0.38	0.046	0.476
30	Voronoi	Scale2	Gravity	exp	0.836	54.287	0.857	0.917	0.79	0.017	1
31	Voronoi	Scale2	Gravity	pl	0.776	63.388	0.868	0.883	0.771	0.019	1
32	Voronoi	Scale2	Radiation	-	-1.702	220.143	0.829	0.682	0.488	0.011	0.762
33	Voronoi	Scale3	Gravity	exp	0.836	54.349	0.857	0.917	0.79	0.017	1
34	Voronoi	Scale3	Gravity	pl	0.769	64.412	0.866	0.879	0.769	0.018	1
35	Voronoi	Scale3	Radiation	-	-1.701	220.249	0.827	0.681	0.488	0.011	0.762

Note: 1. Spearman and Pearson coefficients in the experiment p-value are less than 0.001. 2. Abbreviations of BSL: benchmark experiment under the initial setting, exp: exponential function, pl: power-law function, Rand: randomized experiment, Scale: scale expansion/reduction experiment. 3. The bold number indicates the best-performing results of the models within the same group.

(4) When uniformly reducing the coordinate scale by a factor of 10 without changing the grid partition granularity, the predictive power of the model remains largely unchanged. The experimental results of groups 33–35 show only minor differences compared to groups 18–20.

(5) In terms of the damping function type in the gravity model, the exponential function model under the grid partition is slightly inferior to the power-law function model in predicting results, whereas the results are reversed under the Voronoi diagram partition.

Furthermore, we analyze the relationship between different levels of granularity in knowledge space partitioning, including three different random experiments, and the γ index in the gravity model damping function. As shown in Fig. 9, the analysis reveals that in the context of real scientists' topic selection and transition within the knowledge space, the absolute value of the distance decay factor γ between scientists in different regions

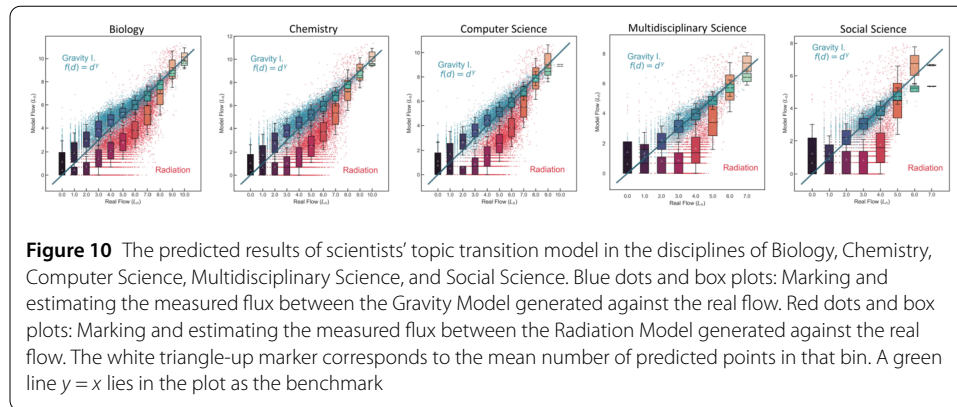


exceeds that in three other random experimental scenarios. This result underscores a significant bounded characteristic in the transition of scientists' interests. The conserved characteristic is influenced by mixed factors such as modularized knowledge structure, individual knowledge attributes, exploration preference patterns, or inter-domain knowledge barriers as scientists move in the knowledge space.

3.5 The generalizability of scientists' knowledge exploration pattern to other disciplines

To assess the generalizability of our findings beyond the discipline of physics, we test the performance of the gravity model and the radiation model across diverse disciplines. As depicted in Fig. 10, the results demonstrate the robustness of our proposed gravity model compared to the radiation model across various fields, including Biology, Chemistry, Computer Science, Multidisciplinary Science, and Social Science. Detailed descriptions of the dataset and the method for constructing the knowledge space based on the Doc2vec algorithm are provided in the Materials and Methods.

Across all disciplines with significant distinct research areas, the gravity model (blue dots in Fig. 10) consistently outperforms the radiation model (red dots in Fig. 10) in predicting scientists' actual mobility patterns within the knowledge space. However, further examination of the simulation results depicted in the grid diagram reveals a significant variance in model performance across disciplines. Social Science exhibits the lowest R-squared metric of 0.746 ($p < 0.001$), while Chemistry achieves the highest R-squared metric of 0.874 ($p < 0.001$). The observed disciplinary discrepancies reveal diverse patterns in scientists' exploration paths within the knowledge space.



4 Discussion

In this study, we utilize a knowledge space to map the trajectories of scientists' publications in chronological order, shedding light on their patterns of topic selection and transition within this knowledge space. We subdivide this space into grid or Voronoi diagram subfields using density and equidistant approaches. Our analysis reveals an overall log-normal distribution of scientists' topic-transition distances at the origins and destinations. To delve into the mechanisms governing these topic transitions at a group level, we introduce two movement behavior models: the gravity and radiation models. Our findings indicate that the gravity model, driven by factors such as population size and knowledge distance, outperforms considerations of research gap areas in explaining and predicting scientists' topic-transition behaviors. To enhance our insights, we compare our results to three key aspects related to existing studies:

1. Comparison with human commuting patterns in real geographic space: We find that scientists' explorations in the knowledge space are more influenced by 'distance' and regional 'population' factors than 'opportunity' factors. This mirrors the patterns observed in human commuting within administrative regions in a city, albeit without predefined sub-field spaces in our knowledge space.

2. Comparison with human movement patterns in virtual space: Scientists' exploratory behavior in the knowledge space exhibits similarities to human behaviors in virtual spaces. The log-normal distribution of exploration trajectories aligns with patterns seen in the game and website access behaviors [34]. Although the space construction frameworks differ, the underlying psychological mechanisms for resource search and acquisition appear to share commonalities [35].

3. Comparison with other models of scientists' topic-changing or switching behavior: We emphasize a collective rather than individual-level perspective on scientists' topic selection and transition, and find that knowledge distance and population size are two key social factors in explaining scientists' exploration patterns in the knowledge space, suggesting a typical hotspot-tracing tendency for the majority of scientists.

In summary, our research advances the understanding of scientists' topic transition by accounting for social influences and distance heterogeneity in the constructed knowledge space. Our findings suggest that most scientists tend to make cautious topic transitions, guided primarily by the number of scientists in their field and the knowledge distance between fields, rather than by 'gaps' or 'opportunities'. This cautious approach may have

significant implications for the efficiency and effectiveness of the scientific innovation system.

5 Conclusion

Our study deploys quantitative analysis methods to investigate scientists' topic selection and transitions, offering insights into the underlying mechanisms at the group level. We find that scientists' movements within the knowledge space exhibit heterogeneity, characterized by an overall log-normal distribution of OD flow distances. It indicates that, in essence, most scientists tend to make prudent and short-range transitions in their research interests. Our analysis identifies key social factors, including subfield population size, research gaps or opportunities, and knowledge distances, as instrumental in shaping scientists' topic transition.

The mechanistic analysis reveals a prevailing tendency towards hotspot-tracing and opportunity-seeking within the academic field, akin to animal foraging behavior, where resource distribution influences foraging strategies. In the competitive realm of scientific research, adopting a conservative strategy appears safe for scientists. Most scientists tend to follow a hotspot-tracing tendency rather than proactively exploring research opportunities between subfields and connecting knowledge from different domains. This conservatism can lead to issues like resource concentration, reduced research originality, and decreased research efficiency for the whole scientific enterprise. Understanding this conservative strategy reveals valuable insights into the dynamics of scientists' knowledge-creation within the innovation system, and provides empirical support for science policy-makers.

In future research, we plan to refine existing population-level models by incorporating additional factors that influence scientific mobility, such as individual career aspirations, hotspots' knowledge structures, and the evolving landscape of scientific research, optimize model performance by exploring various machine learning algorithms, and investigate the nuances of scientific mobility across diverse disciplines and career stages, utilizing academic datasets spanning a broad range of fields and historical periods.

Abbreviations

GIS, Geographic Information System; OD, Origin-Destination; APS, American Physical Society; MAG, Microsoft Academic Graph; PACS, Physics and Astronomy Classification Scheme; CRS, Coordinate Reference System; CPC, Common Part of Commuters; CPC_d , Common Part of Commuters' Distance; CPL, Common Part of Links; CCDF, Complementary Cumulative Distribution Function; BSL, Baseline experiment under the initial setting; Rand, Randomized experiment; Scale, Scale expansion/reduction experiment; exp, exponential function; pl, power-law function.

Acknowledgements

Not applicable.

Author contributions

HX and FL conceived the study. FL and SZ designed the research. FL and SZ performed the experiments. All authors contributed to the analysis of the results and writing of the manuscript. All authors read and approved the final manuscript.

Funding

This work is supported by the National Natural Science Foundation of China under Grant Nos. 72371052 and 71871042 (to HX), and by the Humanities and Social Science Project of the Ministry of Education of China Grant No 18YJA630118 (to HX).

Data availability

The APS data are available at <https://journals.aps.org/datasets> by submitting a request. The MAG data used in this paper was downloaded via the Microsoft Academic Graph APIs. However, the Microsoft Academic website and underlying APIs have been retired in 2021. All other materials used in this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Systems Engineering, Dalian University of Technology, No. 2 Linggong Road, Dalian, 116024, Liaoning, China.

²Institute for Advanced Intelligence, Dalian University of Technology, Linggong Road, Dalian, 116024, Liaoning, China.

³Center for Big Data and Intelligent Decision-Making, Dalian University of Technology, Linggong Road, Dalian, 116024, Liaoning, China. ⁴Key Laboratory of Social Computing and Cognitive Intelligence, Ministry of Education of China, Linggong Road, Dalian, 116024, Liaoning, China.

Received: 12 December 2023 Accepted: 21 March 2024 Published online: 02 April 2024

References

1. Chen-Ning Y (2019) My study and research experience. University of Chinese Academy of Sciences
2. Weisberg M, Muldoon R (2009) Epistemic landscapes and the division of cognitive labor. *Philos. Sci.* 76(2):225–252. <https://doi.org/10.1086/644786>
3. Besancenot D, Vranceanu R (2015) Fear of novelty: a model of scientific discovery with strategic uncertainty. *Econ. Inq.* 53(2):1132–1139. <https://doi.org/10.1111/ecin.12200>
4. Jia T, Wang D, Szymanski BK (2017) Quantifying patterns of research-interest evolution. *Nat. Hum. Behav.* 1(4):0078. <https://doi.org/10.1038/s41562-017-0078>
5. Yu X, Szymanski BK, Jia T (2021) Become a better you: correlation between the change of research direction and the change of scientific performance. *J. Informetr.* 15(3):101193. <https://doi.org/10.1016/j.joi.2021.101193>
6. Huang S, Huang Y, Bu Y, Luo Z, Lu W (2023) Disclosing the interactive mechanism behind scientists' topic selection behavior from the perspective of the productivity and the impact. *J. Informetr.* 17(2). <https://doi.org/10.1016/j.joi.2023.101409>
7. Azoulay P, Graff-Zivin J, Uzzi B, Wang D, Williams H, Evans JA, Jin GZ, Lu SF, Jones BF, Börner K, Lakhani KR, Boudreau KJ, Guinan EC (2018) Toward a more scientific science. *Science* 361(6408):1194–1197. <https://doi.org/10.1126/science.aav2484>
8. Zeng A, Shen Z, Zhou J, Fan Y, Di Z, Wang Y, Stanley HE, Havlin S (2019) Increasing trend of scientists to switch between topics. *Nat. Commun.* 10(1):3439. <https://doi.org/10.1038/s41467-019-11401-8>
9. Aleta A, Meloni S, Perra N, Moreno Y (2019) Explore with caution: mapping the evolution of scientific interest in physics. *EPJ Data Sci.* 8(1):27. <https://doi.org/10.1140/epjds/s13688-019-0205-9>
10. Milojević S (2015) Quantifying the cognitive extent of science. *J. Informetr.* 9(4):962–973. <https://doi.org/10.1016/j.joi.2015.10.005>
11. Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, Vespignani A, Waltman L, Wang D, Barabási AL (2018) Science of science. *Science* 359(6379):0185. <https://doi.org/10.1126/science.aao0185>
12. Bhattacharya J, Packalen M (2020) Stagnation and scientific incentives. *SSRN* 58(12):7250–7257. <https://doi.org/10.3386/w26752>
13. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: models and applications. *Phys. Rep.* 734:1–74. <https://doi.org/10.1016/j.physrep.2018.01.001>
14. Aceves P, Evans JA (2023) Mobilizing conceptual spaces: how word embedding models can inform measurement and theory within organization science. *Org. Sci.* <https://doi.org/10.1287/orsc.2023.1686>
15. American Physical Society (2018) APS data sets for research. <https://journals.aps.org/datasets>
16. Sinatra R, Wang D, Deville P, Song C, Barabási AL (2016) Quantifying the evolution of individual scientific impact. *Science* 354(6312):5239. <https://doi.org/10.1126/science.aaf5239>
17. Sinha A, Shen Z, Song Y, Ma H, Eide D, Hsu B-JP, Wang K (2015) An overview of Microsoft Academic Service (MAS) and applications. In: Proceedings of the 24th international conference on world wide web—WWW'15 companion, pp 243–246. <https://doi.org/10.1145/2740908.2742839>
18. Wang K, Shen Z, Huang C, Wu C-H, Eide D, Dong Y, Qian J, Kanakia A, Chen A, Rogahn R (2019) A review of Microsoft academic services for science of science studies. *Front. Big Data* 2:45. <https://doi.org/10.3389/fdata.2019.00045>
19. Pan RK, Sinha S, Kaski K, Saramäki J (2012) The evolution of interdisciplinarity in physics research. *Sci. Rep.* 2(1):551. <https://doi.org/10.1038/srep00551>
20. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008(10):10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
21. Grover A, Leskovec J (2016) Node2vec: scalable feature learning for networks. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, vol 13–17. ACM, New York, pp 855–864. <https://doi.org/10.1145/2939672.2939754>
22. McInnes L, Healy J, Saul N, Großberger L (2018) Umap: uniform manifold approximation and projection. *J. Open Sour. Softw.* 3(29):861. <https://doi.org/10.21105/joss.00861>
23. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th international conference on neural information processing systems. Curran Associates, Lake Tahoe, pp 3111–3119
24. Gold C (2016) Tessellations in gis: part I—putting it all together. *Geo-Spat. Inf. Sci.* 19(1):9–25. <https://doi.org/10.1080/10095020.2016.1146440>
25. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci.* 109(29):11576–11581. <https://doi.org/10.1073/pnas.1203882109>
26. Williams NE, Thomas TA, Dunbar M, Eagle N, Dobra A (2015) Measures of human mobility using mobile phone records enhanced with gis data. *PLoS ONE* 10(7):1–16. <https://doi.org/10.1371/journal.pone.0133630>
27. Anderson JE (2011) The gravity model. *Ann. Rev. Econ. Sci.* 3:133–160. <https://doi.org/10.1146/annurev-economics-111809-125114>

28. Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100. <https://doi.org/10.1038/nature10856>
29. Zipf GK (1946) The p1 p2/d hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* 11(6):677–686
30. Pappalardo L, Simini F, Barlacchi G, Pellungrini R (2022) scikit-mobility: a python library for the analysis, generation, and risk assessment of mobility data. *J. Stat. Softw.* 103(1):1–38. <https://doi.org/10.18637/jss.v103.i04>
31. Lenormand M, Bassolas A, Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. *J. Transp. Geogr.* 51:158–169. <https://doi.org/10.1016/j.jtrangeo.2015.12.008>
32. Hacker C, Rieck B (2022). On the surprising behaviour of node2vec. <https://github.com/aidos-lab/node2vec-surprises>
33. Alstott J, Bullmore E, Plenz D (2014) powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(1):85777. <https://doi.org/10.1371/journal.pone.0085777>
34. Szell M, Sinatra R, Petri G, Thurner S, Latora V (2012) Understanding mobility in a social Petri dish. *Sci. Rep.* 2:1–6. <https://doi.org/10.1038/srep00457>
35. Wang X, Pleimling M (2017) Foraging patterns in online searches. *Phys. Rev. E* 95(3):032145. <https://doi.org/10.1103/PhysRevE.95.032145>. [arXiv:1703.03901](https://arxiv.org/abs/1703.03901)
36. Wikimedia Commons (2020) File:Ketterle.jpg—Wikimedia Commons, the free media repository. [Online; accessed 13-November-2023]. <https://commons.wikimedia.org/w/index.php?title=File:Ketterle.jpg&oldid=450723481>
37. Wikimedia Commons (2020) File:Leo Esaki 1959.jpg—Wikimedia Commons, the free media repository. [Online; accessed, 8-December-2023]. https://commons.wikimedia.org/w/index.php?title=File:Leo_Esaki_1959.jpg&oldid=512082827

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.