Information Processing and Management 61 (2024) 103702

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Neural embeddings of scientific mobility reveal the stratification of institutions in China

Yongshen He [a], Yurui Huang [a], Chaolin Tian [a], Shibing Xiang [a,b], Yifang Ma [a,*]

[a] Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong, China
[b] Peng Cheng Laboratory, Shenzhen, Guangdong, China

ARTICLE INFO

Keywords:
Scientific mobility
Embedding
China
University stratification
Computational social science

ABSTRACT

We are trying to reveal the status of Chinese universities in the global talent circulation system and how university prestige affects mobility patterns in China. Other than geographic distance, we configured a two-dimensional Euclidean space of the neural embeddings of institutions from scientists' mobility trajectories. To be specific, we acquired the scientific mobility trajectories based on about 3,055,409 authors from 108 thousand affiliations and used the word2vec neural embedding model to encode affiliations into a continuous 100-dimensional embedding space and then further map the 55,298 distinct embeddings into the two-dimensional space. We found that in the current talent circulation system, China is furthest from the scientific center with a closeness of 0.08 and has a median diversity of 0.15 among the top 10 countries with the most affiliations, indicating relatively less communication with other international institutions and moderate diversity of communication patterns within the country. Furthermore, using the regression models, we found that for the universities on China's mainland, the effects of prestige are positive on the volumes of scientific mobility (all *p values* < 0.001) and on the uniqueness of universities (all *p values* < 0.1), revealing the stratifications of universities in the talent circulation. The interaction of prestige and regional economic status is also discussed in our results. This work gives policy implications to university development, knowledge circulation, and career development in science.

1. Introduction

With the trend of globalization, scientific mobility has been a universal activity in past decades (Ackers, 2005; Azoulay et al., 2017; Jałowiecki & Gorzelak, 2004; Zou et al., 2023). Notably, the number of bilateral migration flow (i.e., the number of moved scientists) between China and the US surged from 23 in 1974 to 14,910 in 2014 (Czaika & Orazbayev, 2018; Zou et al., 2023).

There are several reasons behind the movement of talents. First, researchers can promote their career progress thanks to overseas study experience. Petersen (2018) found as much as a 17 % increase in citations of mobile researchers compared to their non-mobile counterparts. Liu and Hu discovered (Liu & Hu, 2022) the increase in scientists' research quantity and collaboration of new partners. Nevertheless, some researchers claim that the situation is not so optimistic. Deville et al. (2014) found subsequent slight descends in citations of researchers migrating to inferior institutions and neglectable variation of citations of researchers whose destination institutions are superior. Li and Tang (2019) discovered the slowdown of late-phase career promotion affected by international academic

* Corresponding author.
E-mail address: mayf@sustech.edu.cn (Y. Ma).

https://doi.org/10.1016/j.ipm.2024.103702
Received 8 August 2023; Received in revised form 8 January 2024; Accepted 28 February 2024
Available online 3 March 2024
0306-4573/© 2024 Elsevier Ltd. All rights reserved.

mobility. More importantly, opening nations with more emigrating researchers tend to produce more papers with high impact (Wagner & Jonkers, 2017). Both the host and donor countries can benefit from the linkage of scientists (Gibson & McKenzie, 2014). However, only a few researchers are concerned with the properties of institutions based on scientific mobility data (Dakota et al., 2020; Wagner & Jonkers, 2017).

For statistical models on human mobility, geographic distance is a fundamental metric. Not only other metrics in human mobility, such as radius of gyration (Fixman, 1962) and effective speed (Tranter, 2012) but also flow generation models, including Gravity (Zipf, 1946) and Radiation (Simini et al., 2012), are based on geographic distance. However, geographic distance is in low-dimensional space, meaning they have just captured partial factors of scientific mobility. This is because scientific mobility is related to not only geographic distance, but also culture, language (Dakota et al., 2020), policy (Chinchilla-Rodríguez et al., 2018), and other factors. Previous work found that university prestige plays an important role in scientists' job placement. For example, Clauset et al. (2015) analyzed 19,000 regular faculties in computer science, business, and history and found that destination institutions can be better explained by doctoral prestige than a U.S. News & World Report rank.

For a long time, universities have developed latent brain circulation (Daugėlienė & Marcinkevičienė, 2009) patterns among universities with different types (Yuan et al., 2009), prestige, and geographical locations. For example, there are frequent movements of scientists between affiliations in the same city or the same province, and similarity, the international movements of scientists are concentrated among minor countries due to their sufficient educational resources. This prompts us to ask: how the scientific mobility patterns differ in different countries, and how the university stratification in terms of prestige (Weakliem et al., 2012), geo-locations (Hu & Vargas, 2015), and administrative divisions (Alessandretti et al., 2020), etc.? The challenge in answering these questions is how to fully capture the collective mobility patterns from all the scientists' affiliated institution records.

To solve the problem, we applied neural embedding to represent high-dimension and sparse mobility data in low-dimension space without the necessity of location characteristics. The embedding technique is also applied widely in Natural Language Processing (Devlin et al., 2018; Elnagar et al., 2023; Tomas et al., 2013) (NLP), Computer vision (Trinh et al., 2019; Xie et al., 2016) (CV), knowledge graph (Wang et al., 2023), time series analysis (Tabassum et al., 2022) and recommendation system (Liu et al., 2023). Some recent research employed this novel technique in both scientific mobility and collaboration and demonstrated that embeddings can capture not only geographic but also cultural, linguistic, and prestige information (Dakota et al., 2020; Liu et al., 2023).

Specifically, we apply the word2vec (Tomas et al., 2013) embedding method to scientific mobility trajectories to represent affiliations in a dense and continuous embedding space, so that the embedding distances between affiliations are based on the mobility pattern. The contributions of our study are as follows. (1) we represent affiliations in a low-dimensional vector space, visualize the global affiliation landscape, and rank the top 10 countries that have the most affiliations by centrality and diversity. (2) we analyze the relationship between affiliation vectors and university types and geographic locations in China. (3) we demonstrate that universities of high prestige are large-scale and distinctive in China.

2. Research objective

We utilize a novel method to represent affiliations. It is noteworthy that no (or few) existing works investigate affiliation representation in the case of China. Therefore, our research has the following aims. (1) To locate the position and status of China in the scientific knowledge circulation with affiliation vectors. (2) To analyze the essential information embedded in the affiliations of China's Mainland, including regional details and university classifications. (3) To examine whether the affiliation vectors can capture the prestige of universities in China's Mainland.

3. Literature review

3.1. Chinese university classification and prestige

Building internationally competitive universities and raising the overall levels of Chinese universities have always been the focus of the Chinese government. In preparation for achieving this purpose in the 21st century, Project 211 was launched by the Ministry of Education of China in December 1995, and approximately 100 universities were selected (Lixu, 2004). To cultivate the top universities of Project 211 and found world-class universities in the 21st century, Project 985 was named for first being announced on May 4, 1998, by the General Secretary of the Chinese Communist Party at the 100th anniversary of Peking University (Ying, 2011). In the past decade, Chinese universities have made considerable progress in academic fields. In October 2015, the Double First Class University Plan, a novel university hierarchy with incentive and restraint mechanism, was published by the Chinese Government and replaced previous higher education projects aiming to develop elite universities comprehensively (Peters & Besley, 2018). The plan of Double-First-Class Universities involves more universities than Project 211 at present.

At the same time, the prestige of universities is usually an international regime. International university ranking system is established by academic field, e.g., Times Higher Education (Baty, 2004) (THE) and commercial institutions, e.g., Quacquarelli Symonds (2004) (QS) World University Rankings, and US and World News Report (Washington, 1933) (US News) rather than the government. Especially QS ranking has gained wide acknowledgment in China (Allen, 2021).

University ranking is popular in the higher education system as well as in public. For students, the prestige of their universities concerns academic career identity attainment (Yang et al., 2022), so they will consult university ranking when choosing universities (Wut et al., 2022). For universities, their prestige plays a crucial role in exclusive licensing (Shen et al., 2022) and transaction prices in the technology market (Shen et al., 2023).

3.2. Scientific mobility

Scientific mobility has been a subject of growing interest in the academic literature. Numerous studies have explored the patterns, drivers, and consequences of researcher migration across geographic locations and academic institutions. Research in this area has provided valuable insights into the dynamics of knowledge exchange, collaboration networks, and the impact on scientific advancements. In scientific fields, researchers are intrigued by the motivation of talent migration (Gureyev et al., 2020). Zanon (2021) found most Austrian researchers left their home countries for better research conditions. Return migration is influenced positively by home linkage (Baruffaldi & Landoni, 2012) and academic ability (Gaulé, 2014). Gender (Jöns, 2011) is also an important factor in scientific mobility. Gargiulo and Carletti (2014) applied complex networks to demonstrate the home position, and the linguistic and historical similarity between countries are good predictors for the next moving affiliations. In this paper, we explore the relationship between scientific mobility and its factors at the institutional level.

3.3. Models of human mobility

Human mobility models mainly focus on predicting or generating crowd flows and individual trajectories. They can be divided into statistical models and neural embedding models.

3.3.1. Classical statistical models

Statistical models just take a few elements into account to predict flows and trajectories. Barbosa et al. (2018) noted, some random processes, such as random walk and Levy process, could be used when modeling individual patterns. Alessandretti et al. (2020) considered the attractiveness of locations in different levels of "containers" constructed by hierarchical clustering when predicting moving probabilities. At the popularity level, Zipf (1946) believed the average migratory flow between two places in proportion to their populations and the reciprocal of their distances. Simini et al. (2012) thought the average number of travelers related to the opportunities of both the origin and the destination as well as the sum of other opportunities closer to the origin than the destination. However, these methods are too simple and cannot capture most information. Schläpfer et al. (2021) deemed the population density of the area in proportion to its attractiveness and the reciprocal of the square of radius times the in and out frequency of each trajectory and used the attractiveness of two areas to predict their flows.

3.3.2. Neural embedding models

By comparison, neural embedding models are complicated architectures. Neural embeddings, a sort of dimension reduction technique, use neural networks to map high-dimensional data to a low-dimensional continuous space. They have revolutionized natural language processing and machine learning, capturing semantic relationships between words, sentences, and even images. Despite impressive advancements, challenges in fine-tuning and bias need further attention. The field continues to evolve rapidly, with potential for innovative applications in various domains of artificial intelligence. In the NLP area, neural embedding has been successfully applied in pre-training models of the large-scale corpus, such as word representation (Ji et al., 2015; Pennington et al., 2014; Tomas et al., 2013) and document representation (Devlin et al., 2018; Peters et al., 2018; Radford & Narasimhan, 2018). In the CV field, neural embedding is utilized when the majority of images lack labels, like clustering (Xie et al., 2016) and pretraining (Han et al., 2020; Trinh et al., 2019). However, (Bagheri et al., 2018) showed neural embedding does not show competitive enhancement compared to statistical models and methods based on knowledge base.

Researchers have successfully applied the GRU (Cho et al., 2014) and LSTM (Hochreiter & Schmidhuber, 1997) to predict the trajectory of an individual's movement and the flows of populations. Chen et al. (2020) applied RNNs (Rumelhart et al., 1985) and feed-forward neural networks with GRU (Cho et al., 2014) and attention mechanism (Vaswani et al., 2017) to encode time, locations, and users, and extract context of sequential dependency, space, periodicity, and society. Yang et al. (2020) proposed Flashback to predict sparse traces. Xu et al. (2021) proposed a two-stage temporal-spatial preference model to predict users' next location consumption locations. Zhang et al. (2023) adopted federated learning methods (Sattler et al., 2020) to cluster the feature extracted by LSTMSeq2Seq (Sutskever et al., 2014) and adopted graph neural networks (Kipf & Welling, 2016; Wu et al., 2020) in the last step to complete the task of trajectory prediction. As for modeling the flow, Ren et al. (2020) combined both LSTM (Hochreiter & Schmidhuber, 1997) and ST-ResNet (Zhang et al., 2017) to predict the spatial-temporal flow volumes. Tian (2020) combined long-term effects extracted by ConvLSTMs (Shi et al., 2015) with attention mechanism with short-term effects extracted by a single ConvLSTM. Nevertheless, these constructions lack interpretability and hardly uncover the latent structure of locations. Terroso-Saenz et al. (2022) use Multi-Layer Perceptron, LSTM (Hochreiter & Schmidhuber, 1997) to demonstrate Twitter displacements can reduce the errors when forecasting the flows.

4. DATA and method

4.1. Dataset description

OpenAlex (Priem et al., 2022) is a large-scale fully-open bibliographic database. The database includes the entities of works, authors, journals, institutions, disciplines, publishers, and funders. In our research, we used the whole work and institution information. The dataset contains about 238 million publications with detailed authorship information (author, institution, etc.) and publication year as of 2021. Each author or institution has a unique ID, The dataset of institutions contains about 108,607 affiliations with detailed geographical information (city, country, etc.) as of 2021. Other information on universities in China's Mainland can also be acquired on the website. University type can be obtained in ShanghaiRanking (Consultancy, 2003). Lists of Chinese universities with high prestige can be found according to Wikipedia (Donal, 2001) and the official websites and QS ranking (Symonds, 2004).

4.2. Data preprocessing

In summary, we use the large-scale scientific corpus, OpenAlex, to extract the scientific trajectories, that is, each author corresponds to a trajectory, and each element in the trajectory represents an affiliation. Then, we use the word2vec model to embed affiliations as low-dimensional vectors in preparation for analyzing their internal structure. We extract the scientific trajectories of scholars from the OpenAlex (Priem et al., 2022) database. For each author, we extract the mobility records from the affiliated institutions in her/his publications. In each trajectory, affiliations are sorted according to the ascending publication years, if multiple affiliations appear within a year in a trajectory, then their affiliation IDs are placed in random order. To ensure the length and accuracy of trajectories, we filter the authors whose papers are less than 10 or who are attached to only one affiliation. After that, 3055,409 authors are retained containing 55,298 distinct affiliations. Fig. 1. shows the migration trajectories between and within the top 10 countries that have the most affiliations. We can see that the share of China has been increasing in the past decade. However, the mobility of China is more and more inclined to internal mobility, and the proportion of mobility between China and other countries except the United States is sharply declined.

4.3. Affiliation embedding method

4.3.1. Word2vec embedding

We embed trajectories by analogizing them to sentences and locations to words. For each trajectory, we convert each affiliation (Dakota et al., 2020) to a continuous representation via the word2vec embedding model (Tomas et al., 2013). Based on prior work, the skip-gram model is chosen (McCormick, 2016), the vector dimension is set to 100, the width of the window is set to 1, and the minimum count is set to default 5. Eventually, we obtained 42,183 affiliation embeddings, each affiliation is embedded with a 100-dimensional vector.

4.3.2. Dimension reduction

To visualize affiliations in a two-dimension space and prepare for filtering the noise, we employ a manifold-based dimension

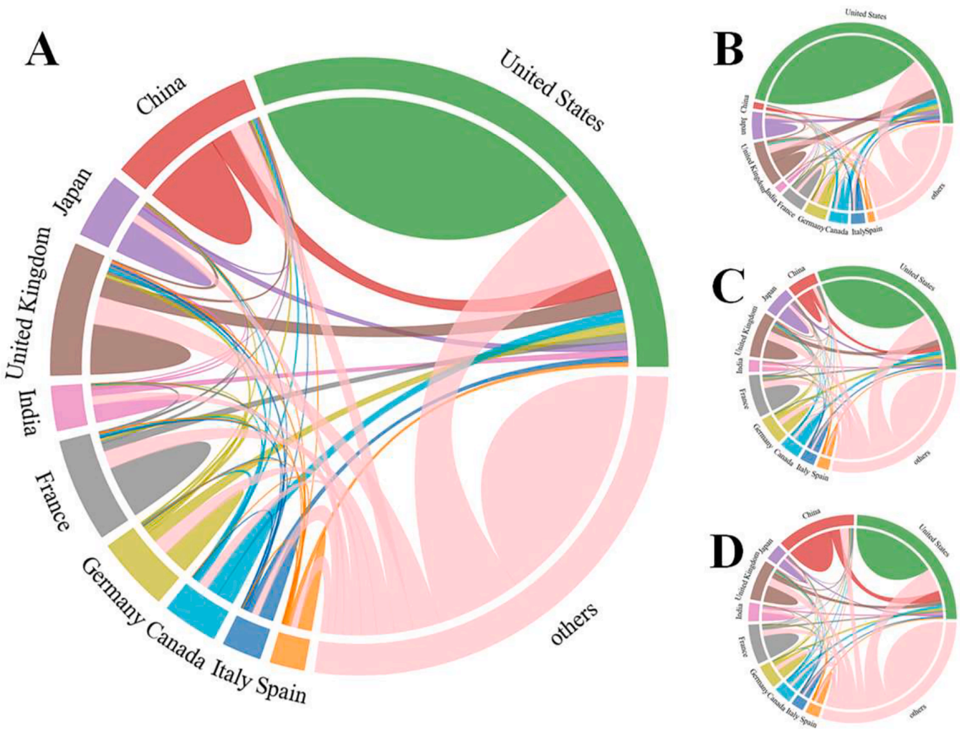

Fig. 1. Bilateral migration between and within the top 10 countries that have the most affiliations. In Panel A, the period is as of 2021. In Panel B, the period is as of 2001 and the affiliation records in trajectories of that period account for 13.31 %. In Panel C, the period is from 2002 to 2011 and the affiliation records in trajectories of that period account for 25.75 %. In Panel D, the period is from 2012 to 2021 and the affiliation records in trajectories of that period account for 60.94 %.

reduction method (UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2018)) to obtain two-dimension affiliation vectors. The aim of UMAP is to minimize the cross entropy of the similarity before and after dimensionality reduction. Thanks to the sparsity of the loss function, UMAP is very efficient in the large sample case, and the result is more distinguishable compared with TSNE (T-Distributed Stochastic Neighbor Embedding) (Liu et al., 2016). Based on the properties of word2vec embeddings, we choose cosine distance to measure the difference between 100-dimension embeddings, while the Euclidean distance is chosen for two-dimension vectors. The process of data preprocessing and neural embedding is shown in Fig. 2.

4.3.3. *Noise filtering*

After further dimension reduction, we center affiliation representations and select the top 10 countries according to the affiliation numbers within them (24,001 affiliations are included), comprising the United States, United Kingdom, Germany, India, China, Japan, France, Canada, Italy, and Spain. To filter the noise, we apply the DBSCAN (Ester et al., 1996) algorithm to detect the hub for each country. After DBSCAN, 20,997 affiliations are retained and over 75 % of affiliations are reserved in each country. We compute the silhouette score of our methods and the result is 0.6600, while that of TSNE and DBSCAN is just 0.3165. demonstrating that the affiliation after UMAP is more distinguishable than those after TSNE.

4.4. *Measurements of affiliations*

4.4.1. *Closeness*

Closeness measures how close a country is to the global scientific center. In two-dimensional Euclidean space, a country with a high closeness score should be near the center of global affiliation vectors. Thus, we define the closeness of a country j as

$$closeness_j = \frac{\min_k \| \boldsymbol{\mu}_k - \boldsymbol{\mu} \|_2}{\| \boldsymbol{\mu}_j - \boldsymbol{\mu} \|_2}. \tag{1}$$

Here, μ_j and μ_k are the mean vector of affiliations in the country j and k respectively. μ is the mean vector of affiliation in all countries. $\|$ ·AptCommand2016;$_2$ is the L-2 norm. The numerator makes sure this index is in [0,1]. If the position of a country is close to the center, its closeness is close to 1.

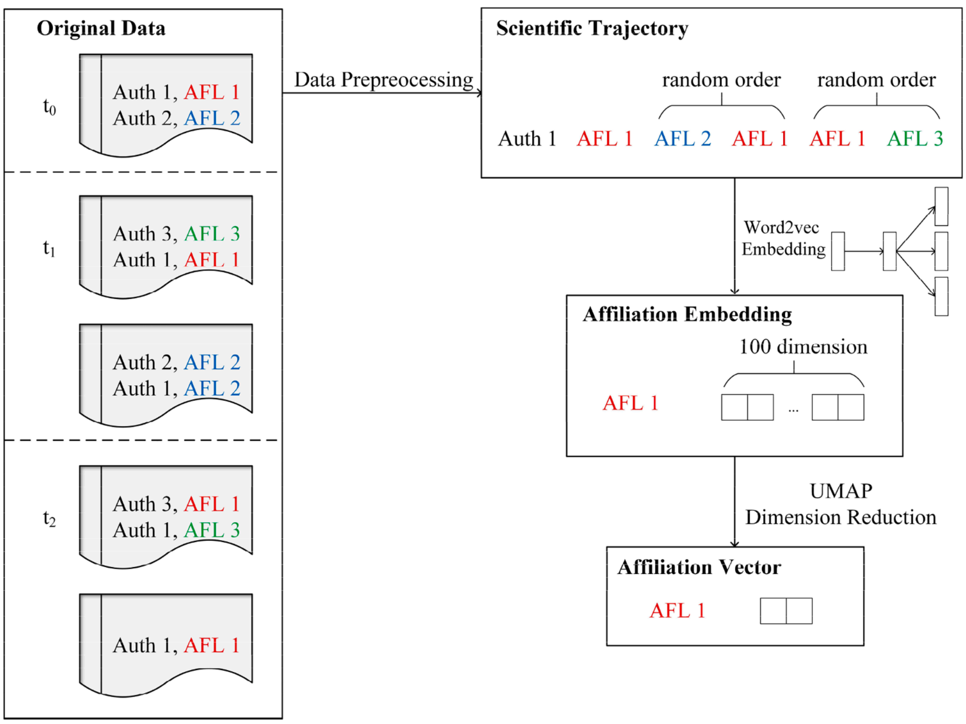

Fig. 2. Data processing and neural embedding. Originally, the affiliation information of authors is attached to papers. To make it clear, we present papers of author Auth 1 whose affiliations are AFL 1, AFL 2, and AFL 3 in the order of publication year t_0, t_1, and t_2. After data preprocessing, we obtain affiliation trajectories of authors sorted by publication year, in which their affiliations are placed in random order if over one affiliation appears within a year in a trajectory. Then we use the word2vec model to acquire 100-dimension affiliation embedding. Next, we utilize the UMAP method to gain 2-dimension affiliation vectors.

4.4.2. Diversity

Diversity measures how divergent affiliations are located within a country. If the essences of affiliations are distinctive in a country, such as preponderant disciplines and collaborators, this will influence the diversity of scientific mobility patterns and eventually respond to the variance of each dimension of affiliation vectors. Thereby, we define the diversity of the country j as

$$diversity_j = \frac{\frac{1}{|c_j|-1}\sum_{i\in c_j}\| \mathbf{x}_i - \boldsymbol{\mu}_j \|_2^2}{\max_k \frac{1}{|c_k|-1}\sum_{i\in c_k}\| \mathbf{x}_i - \boldsymbol{\mu}_k \|_2^2}. \tag{2}$$

Here, μ_j and μ_k are the mean vectors of affiliations in the country j and k respectively. c_j is the set of affiliations of the country j. $|c_j|$ is the number of the set c_j. x_i is the ith affiliation vector. The denominator guarantees the index falling into [0,1].

4.4.3. Volume

Volume counts the total affiliation appearances of all the scientific trajectories. The volume of affiliation i is formulated as

$$volume_i = \sum_{d\in D}\sum_{j\in d} I_{\{i=j\}} \tag{3}$$

Here, i and j are affiliations. $I_{\{i=j\}}$ is the indicator function whose value equals 1 if i and j are the same affiliation and otherwise equals 0. d is a scientific trajectory. D is the set of all the scientific trajectories. The large volume of a university characterizes two features of the university: the considerable quantity of mobile researchers and the large number of publications per author affiliated with the university.

4.4.4. Uniqueness

Uniqueness indicates how distinctive an affiliation is from another in the two-dimensional vector space. To be specific, a unique affiliation should have only a few neighbors within a short distance. Therefore, we defined the uniqueness of affiliation i as

$$uniqueness_i = \log\frac{\max_k \widehat{f}_h(\mathbf{x}_k)}{\widehat{f}_h(\mathbf{x}_i)}. \tag{4}$$

Here, x_i and x_k are the ith and kth affiliation vectors respectively. $\widehat{f}_h(\mathrm{x}_i)$ and $\widehat{f}_h(\mathrm{x}_k)$ are the kernel density estimates of x_i and x_k respectively. Gaussian Kernel and Euclidean distance is chosen for this estimate. Bandwidth is calculated via the Scott method (Stone, 2008). The minimum of this index is 0. If an affiliation is similar to many affiliations, its uniqueness is close to 0.

4.5. Shifted lognormal and gamma distributions

Shifted lognormal and gamma distributions are used to fit the indexes of volume and uniqueness respectively. The probability density function of shifted lognormal distribution is

$$f(x|s,\mu,\sigma) = \frac{1}{s(x-\mu)\sqrt{2\pi}}\exp\left(-\frac{\log^2\left(\frac{x-\mu}{\sigma}\right)}{2s^2}\right), x>\mu, s>0. \tag{5}$$

Here, μ is the location parameter. σ is the parameter. s is the shape parameter. If we take the logarithm on both sides of Eq. (5), we can see that *logf* $(x|s,\mu,\sigma)$ is the quadratic function of $\log(x-\mu)$ for lognormal distribution.

The probability density function of shifted gamma distribution is

$$f(x|s,\mu,\sigma) = \frac{(x-\mu)^{s-1}e^{-\frac{x-\mu}{\sigma}}}{\Gamma(s)\sigma^s}, x>\mu, s>0. \tag{6}$$

Here, μ is the location parameter. σ is the parameter. s is the shape parameter. If we take the logarithm on both sides of Eq. (6), we can see that $\log f(x|s,\mu,\sigma)$ is the linear combination of $x-\mu$ and $\log(x-\mu)$ for gamma distribution.

5. Results

5.1. Model validation

We use the gravity (Zipf, 1946) model to evaluate the accuracy of our neural embedding model. In our case, the gravity model tells the law that the average flux between two affiliations is in proportion to the product of their volumes divided by the power of their distances. If we take the logarithm of the average flux and divide it by the product of volumes, that is,

$$\log\frac{\widehat{T}_{i,j}}{m_i m_j} = \log C - \alpha \log r_{i,j}. \tag{7}$$

Here, $\widehat{T}_{i,j}$ is the estimate of flux between affiliation i and j. m_i and m_j are the volumes of affiliation i and j respectively. The intercept C is a constant. $r_{i,j}$ is the distance between affiliation i and j. α is the power of $r_{i,j}$ in the gravity model.

According to Eq. (7), we establish the linear regression model and compare the goodness of fit using the embedding distance and geographic distance within and between countries respectively. We confine large-scale flows greater than 500 and gain 3389 flows within countries and 246 flows between countries. The regression plots are in Fig. 3. with all $p < 0.01$. All the results of regressions are consistent with gravity models except that with geographical distance between countries. Furthermore, the correlation of flows between countries is stronger than that within countries.

The results also demonstrate the rationality of indexes of closeness and diversity. Fig. 3A implies that the global scientific center has the maximum attractiveness taking some kind of average. Fig. 3B suggests that the attractiveness of a country with a high diversity will not concentrate on only a few affiliations.

5.2. Scientific landscape in country level

After the neural embedding and further dimension reduction, affiliations are embedded into a two-dimensional space and the distance between two affiliation vectors indicates the tendency of scientists' mobility between the two locations. Based on the embedding, we analyzed how affiliations are allocated in the world and where China is through the affiliation map of the top 10 countries.

In Fig. 4, we can see that the affiliations within a country are located together on the map. The US is at the center and covers the largest area among the ten countries. The UK is also very close to the center and has some overlapping areas with the US. Canada, Italy,

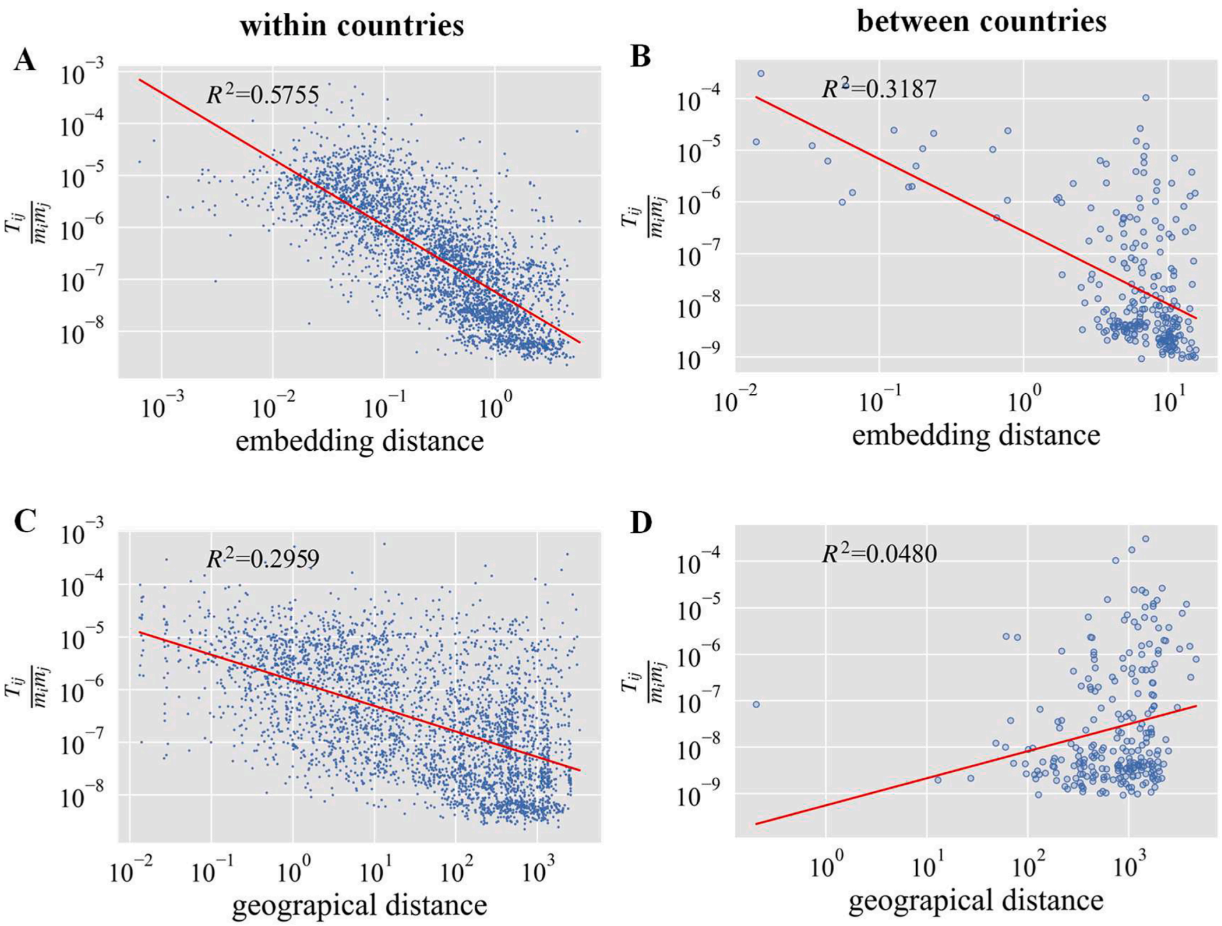

Fig. 3. Examination of gravity models. In Panel A, the independent variable is the embedding distance within countries. In Panel B, the independent variable is the embedding distance between countries. In Panel C, the independent variable is geographical distance within countries. In Panel D, the independent variable is the geographical distance between countries.

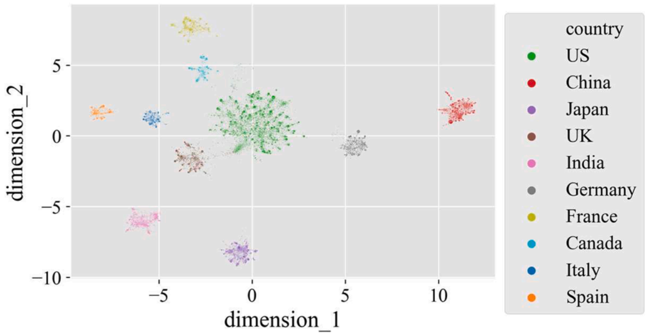

Fig. 4. Affiliation map of two-dimensional affiliation vectors (dimension 1, dimension 2) of the top 10 countries after DBSCAN. Sizes of points indicate times of the affiliation appearance of all the trajectories.

and Germany are in the second circle. On the third circle, France is next to Canada and Spain is near Italy, while Japan and India are at the periphery of the map. China is far from the other 9 countries and just like the pole of the map. The relative positions of the countries reveal the underlying mobility frequencies of scientists between different countries.

To further quantify the mobility patterns among the affiliations in the top 10 countries, we compute closeness and diversity and rank these two indexes of the top 10 countries, which is shown in Fig. 5. We can see that the US has the biggest closeness and diversity (equal to 1), which is followed by the UK, with a closeness of 0.24 and a diversity of 0.27. The high closeness and diversity of the US is aligned with the case that it is located in the center of the affiliation map in Fig. 4. Additionally, China, of which the closeness is 0.08, ranking at the bottom, and diversity is 0.15, ranking fifth among the ten countries. This result shows that for the circulation of knowledge via scientists' mobility, there is still a long way for China to go to outstrip the United States to become a scientific center.

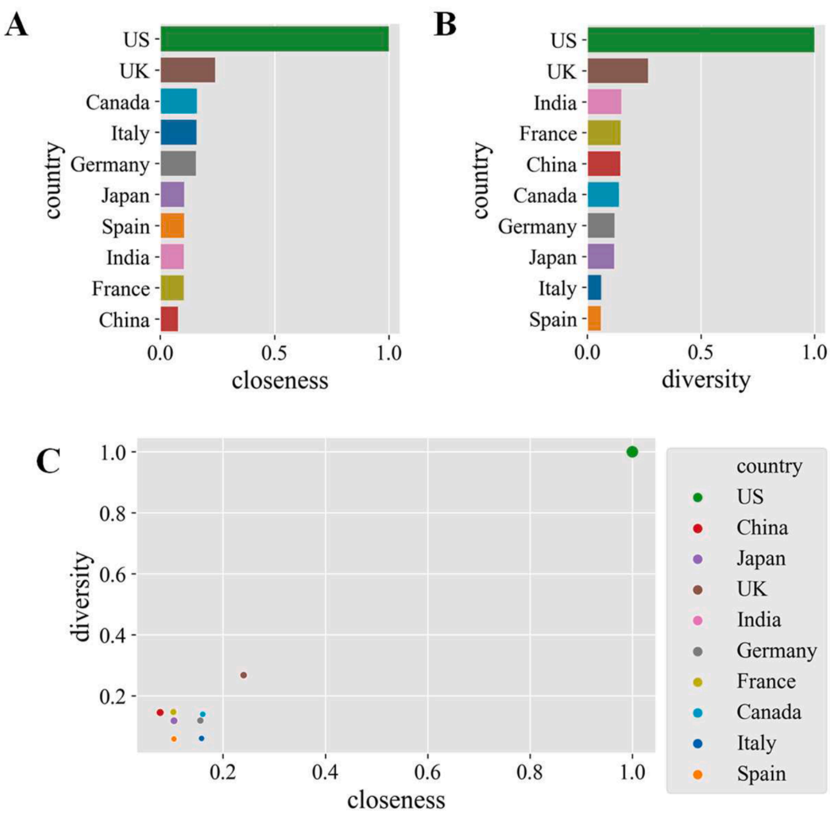

Fig. 5. Closeness and diversity of the top 10 countries. Panels A and B rank closeness and diversity of the top 10 countries respectively in descending order. Panel C illustrates the high correlation between closeness and diversity among the top 10 countries, in which the sizes of points indicate times of the affiliation appearance of all the trajectories.

5.3. Embeddings capture university type and geographic information

Here we focus on the 2232 universities in China by exploring the top 364 Chinese mainland affiliations with the highest mobility frequencies to see the difference in affiliation vectors of diverse types and geographic information.

First, we analyze the distribution of affiliation vectors in different regions. We aggregate the province vector as the mean vector of affiliation vectors within each province to illustrate the hierarchy of provinces, which is shown in Fig. 6A. Based on the hierarchical tree, it is suitable to divide the 30 provinces into 2 regions. The first region contains 10 provinces, which are Guangdong, Hunan, Hainan, Zhejiang, Yunnan, Guangxi, Jiangxi, Hubei, Fujian, and Jiangsu. Most of them are in the southeast of China, which is shown in Fig. 6B. The rest 20 provinces are in the second region and the majority are in the north of China. We further visualize the affiliation vectors in the two regions to illustrate the accuracy of clustering, which is shown in Fig. 7A. The distinction between affiliation vectors

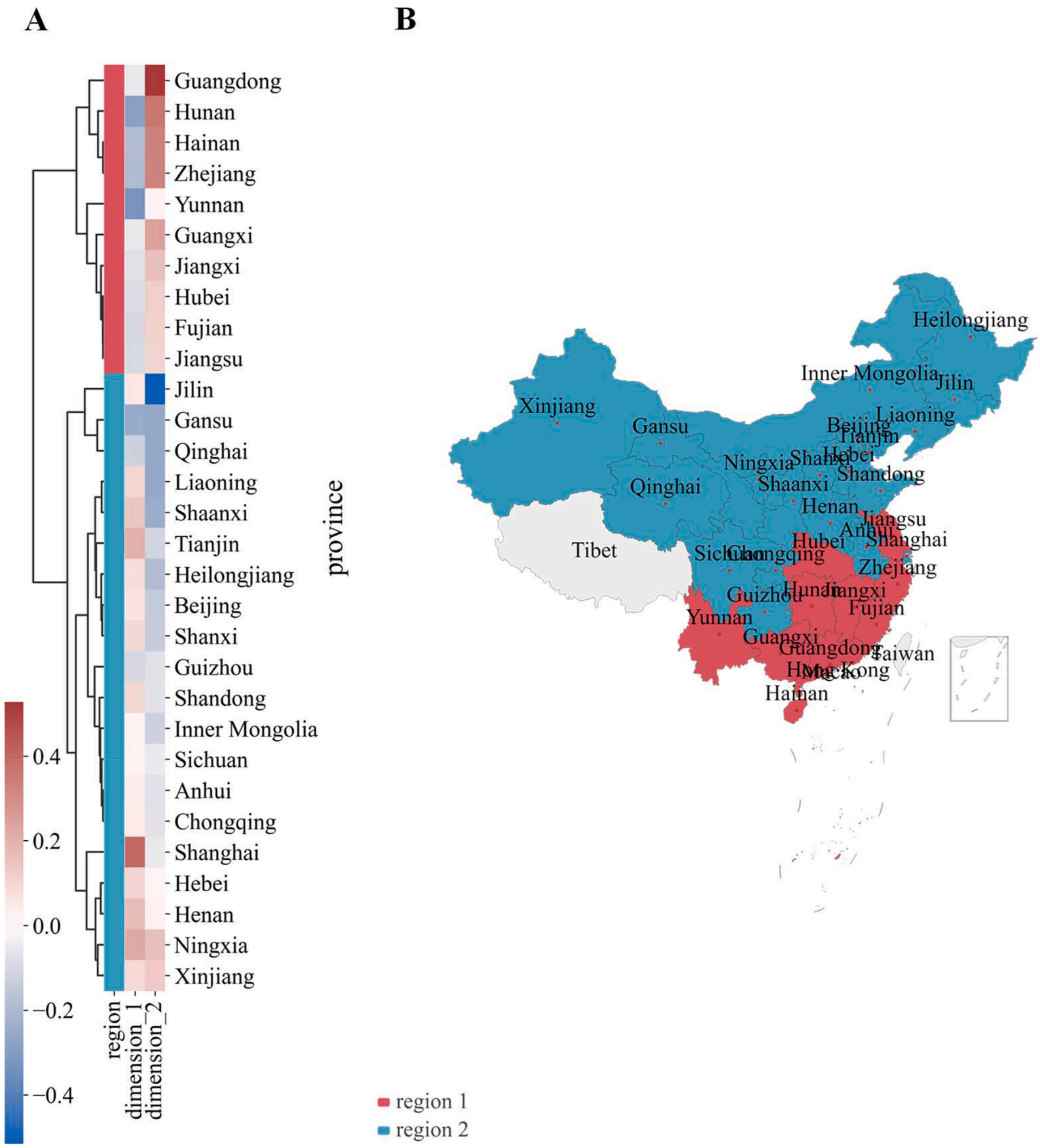

Fig. 6. Regional stratification of 30 provinces of China's mainland. In Panel A, two-dimensional vectors (dimension 1, dimension 2) have been centered. The color of the cell in the "region" column is consistent with that in Panel B. Euclidean distance is chosen to measure the difference between province vectors. Wald linkage is utilized when groups of province vectors agglomerate. Panel B is the map of the 30 provinces, in which colors represent the two regions.

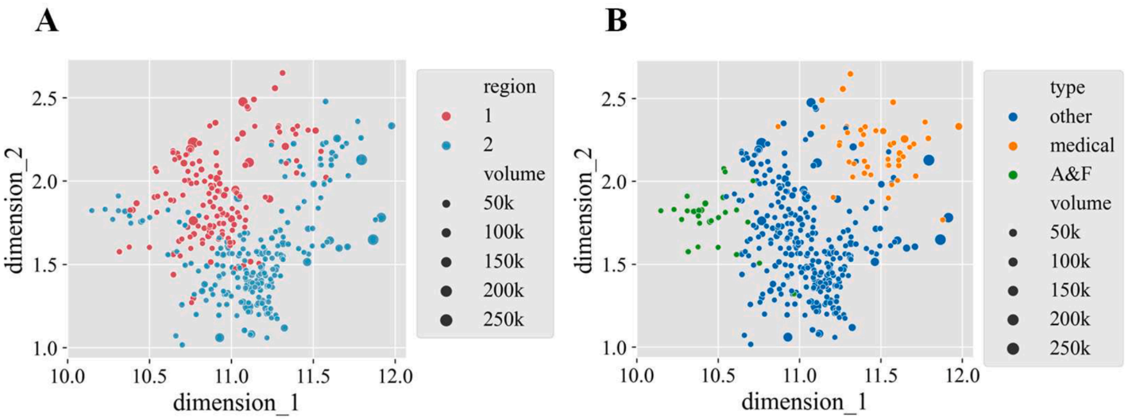

Fig. 7. Scatter plot of the 364 affiliation vectors (dimension 1, dimension 2) of different regions and types in China. Colors indicate the two regions in Panel A and types in Panel B. Sizes indicate times of the affiliation appearance of all the trajectories.

in the two regions is also significant statistically (Pillai's trace = 0.46, F (2361) = 150.92, $p < .001$).

Then, we analyze the distribution of affiliation vectors in different types. Based on the data of ShanghaiRanking (Consultancy, 2003), universities in China's Mainland can be divided into the following types according to preponderant disciplines: comprehensive type, science and technology, normal, agriculture and forestry (A&F), medicine, finance and economics. The counts of the above types of universities are more than five in our dataset. However, many universities have disciplines of science and technology, normal and finance and economics. Therefore, we separate medical universities and A&F universities from other ones. These universities have a high degree of specialization and talents are more inclined to move to the universities of the same type. We visualize the affiliation vectors of these three different types, which are shown in Fig. 7B. We can see significant gaps between the affiliation vectors of different types (Pillai's trace = 0.74, F (4722) = 106.50, $p < .001$).

5.4. *Mobility embeddings capture university prestige*

How do the mobility embeddings correlate with the universities' prestige? To answer this question, in this section, we compare the different university prestige classifications with features from the neural embeddings, we measured the difference between universities of different prestige levels in terms of volume and uniqueness through regression.

For university prestige, we used four different classifications to define the high-prestige university class. (1) Project 211, (2) Project 985, (3) The Double First Class, (4) The QS world university ranking. Here, we set the top 1000 QS ranking as the QS high prestige class. Table 1 shows the number of universities for each of the four different classifications.

We use the volume and uniqueness to capture the importance of universities in the embedding space, compute these two measurements of the 364 affiliations, and fit these two indexes with shifted lognormal and gamma distribution respectively, as shown in Fig. 8 (See Sections 4.4.3, 4.4.4., and 4.5). The estimated coefficients of both volume (K-S = 0.0364, p= .7048) and uniqueness (K-S = 0.0893, p= .0056) yield the distributions we assume at 0.001 significant levels.

According to 5.2, university types and provinces are related to the distribution of affiliation vectors, and therefore these two confounders should be considered when analyzing the relationship between prestige and the two characteristics of importance. To avoid functional form misspecification, some causal inference techniques are applied to eliminate the impacts of confounders. Given the unbalanced distribution of provinces, we utilize the overlap weighting (Li et al., 2018) method. The weight of each affiliation equals one minus the propensity score. We use overlap weighting rather than inverse probability weighting to mitigate the impact of extreme observations. The propensity score is predicted by categorical Naïve Bayes (C.D. Manning, 2008). The prior parameter alpha is set to 1/2. After the weighting, from Table 2, we can see that each pair of prestige and confounder is independent (p>.1).

Then we establish weighted generalized linear models to estimate the effects of local economic status and prestige on $\mathbf{log}(\boldsymbol{volume} - \boldsymbol{\mu})$ and $\boldsymbol{uniqueness} - \boldsymbol{\mu}$. The deviances of regressions in Table 3 are

Table 1
Several prestigious universities are based on different university classifications: Project 985, Project 211, the Double First Class, and the QS Top 1000.

	Number of universities
Project 211	106
Project 985	41
Double First Class	129
QS Top 1000	57

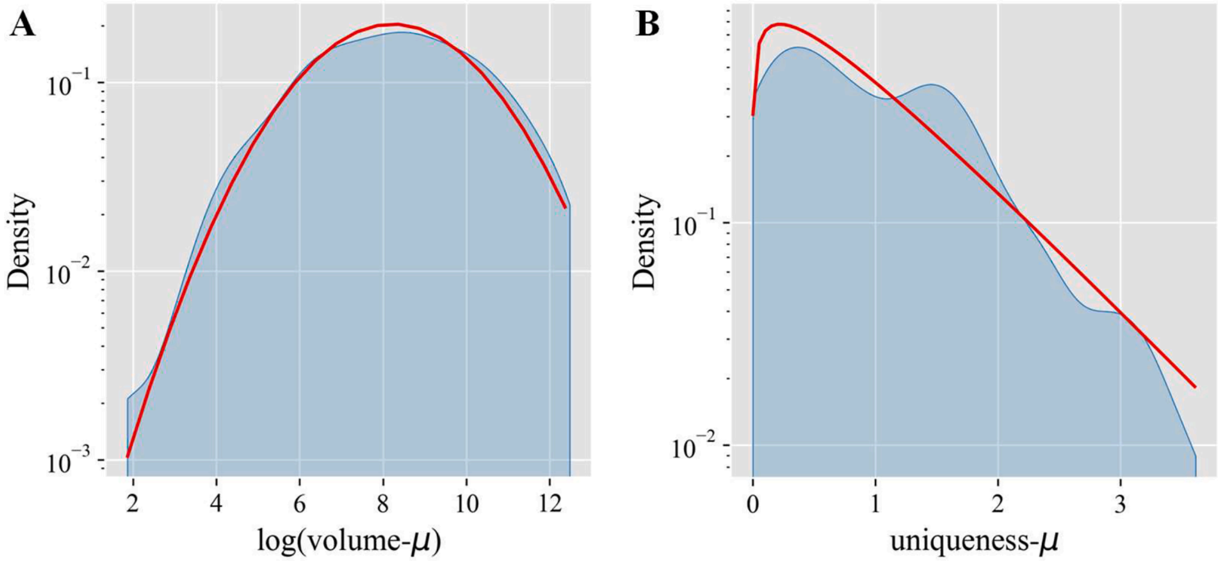

Fig. 8. Distribution of $\log(\boldsymbol{volume} - \boldsymbol{\mu})$ (Panel A) and $\boldsymbol{uniqueness} - \boldsymbol{\mu}$ (Panel B) in logarithmic scale. Blue shade areas cover the Gaussian kernel densities in both panels. In Panel A, the red curve draws the function of Eq. (5) with the location parameter $\boldsymbol{\mu} = 767$. In Panel B, the red curve plots the function of Eq. (6) with the location parameter $\boldsymbol{\mu} = -0.0037$. All the parameters of Eqs. (5) and (6) are calculated by maximum likelihood estimation.s (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Results of independence tests between prestige and confounders after overlap weighting.

Prestige	confounder	Pearson's χ^2	p	df
Project 211	Province	3.9956	1.0000	29
	Univ. type	0.3930	0.8216	2
Project 985	Province	20.2155	0.8861	29
	Univ. type	0.6638	0.7176	2
Double First Class	Province	2.8588	1.0000	29
	Univ. type	0.2717	0.8730	2
QS Top 1000	Province	19.0448	0.9201	29
	Univ. type	0.4926	0.7817	2

$$\sum_{i=1}^{n} w_i \left(\log(volume_i - \mu) - \mathbf{x}_i{}^T \boldsymbol{\beta}\right)^2. \tag{8}$$

Here, $\boldsymbol{n}$ is the number of observations equal to 364. $\boldsymbol{w_i}$ is the propensity score of university $\boldsymbol{i}$. $\boldsymbol{volume_i}$ is the volume of university $\boldsymbol{i}$. $\boldsymbol{\mu}$ is the location parameter equal to 767. $\mathbf{x}_i$ contains the constant term, the interaction of prestige and the log GDP of the city in 2021, region, and type of university $\boldsymbol{i}$. $\boldsymbol{\beta}$ is the coefficient vector.

The deviances of regressions in Table 4 are

$$2\sum_{i=1}^{n} w_i \left(\log\left(\frac{uniqueness_i - \mu}{e^{\mathbf{x}_i{}^T \boldsymbol{\beta}}}\right) + \frac{uniqueness_i - \mu}{e^{\mathbf{x}_i{}^T \boldsymbol{\beta}}} - 1\right). \tag{9}$$

Here, $\boldsymbol{n}$ is the number of observations equal to 364. $\boldsymbol{w_i}$ is the propensity score of university $\boldsymbol{i}$. $\boldsymbol{uniqueness_i}$ is the uniqueness of university $\boldsymbol{i}$. $\boldsymbol{\mu}$ is the location parameter equal to -0.0037. $\mathbf{x}_i$ contains the constant term, the interaction of prestige and the GDP of the city in 2021, region, and type of university $\boldsymbol{i}$. $\boldsymbol{\beta}$ is the coefficient vector.

From Table 3, we can see that university prestige has a significant positive effect on volume in all of the four models (all $p < 0.001$), consistently demonstrating that universities of high prestige are large-scale in mobility outputs in general. Effects of the economic status of cities on volumes are also significantly positive ($p < 0.001$) in most cases. The only exception is that for universities of Project 985 and QS Top 1000, the effects of GPD are not significant ($p > 0.1$), which is shown in Fig. 9, indicating that the economy of geographic regions is not crucial for the volumes of top elite universities.

Correlations between prestige and uniqueness are also consistently significantly positive in Table 4 (all $p < 0.1$) and visualized in Fig. 10. The effect of Project 211 is smaller than others. This is because Project 211 includes some relatively inferior universities compared to Project 985 and QS Top 1000, and Double First Class incorporates some distinctive universities that have preponderances

Table 3
Estimated coefficients of the generalized linear model of which the deviance is shown in Eq. (8). The distribution family used in the regression is Gaussian, and the link function is identity. Location parameter $\mu = 767$. Sample size $N = 364$.

Dependent variable	Model 1 $\log(volume - \mu)$	Model 2 $\log(volume - \mu)$	Model 3 $\log(volume - \mu)$	Model 4 $\log(volume - \mu)$
Project 211	2.5140 *** (0.1492)			
Project 985		3.1014*** (0.1448)		
Double First Class			2.5292*** (0.1498)	
QS Top 1000				3.0869*** (0.1400)
log(GDP)	0.3974*** (0.1023)	0.4994*** (0.0909)	0.3133** (0.1045)	0.2916** (0.0906)
Prestige * [log(GDP) -mean]	0.0241 (0.1599)	−0.6839*** (0.1753)	0.0601 (0.1637)	−0.4863** (0.1719)
Region	Yes	Yes	Yes	Yes
University Type	Yes	Yes	Yes	Yes
R^2	0.4982	0.6149	0.4953	0.6122
N	364	364	364	364

0.1,.
* $p<.05$,.
** $p<.01$,.
*** $p<.001$.

Table 4
The estimated coefficients of the generalized linear model of which the deviance is shown in Eq. (8). The distribution family used in the regression is Gamma, and the link function is logarithm. Location parameter $\mu = -0.0037$. Sample size $N = 364$.

Dependent variable	Model 5 $uniqueness - \mu$	Model 6 $uniqueness - \mu$	Model 7 $uniqueness - \mu$	Model 8 $uniqueness - \mu$
Project 211	0.1556· (0.0814)			
Project 985		0.3957*** (0.0842)		
Double First Class			0.2580** (0.0794)	
QS Top 1000				0.2768** (0.0857)
log(GDP)	0.1755** (0.0558)	0.1150* (0.0529)	0.0856 (0.0554)	0.1661** (0.0554)
Prestige * [log(GDP) -mean]	−0.0013 (0.0873)	0.0828 (0.1020)	0.0783 (0.0868)	0.0112 (0.1056)
Region	Yes	Yes	Yes	Yes
University Type	Yes	Yes	Yes	Yes
Pseudo R^2	0.1556	0.1403	0.1968	0.1330
N	364	364	364	364

0.1,.
* $p<.05$,.
** $p<.01$,.
*** $p<.001$.

in specific fields, such as the University of Chinese Academy of Sciences, Nanjing University of Information Science, etc. Effects of the economic status of cities on uniqueness are also significantly positive ($p < 0.05$) in Models 5,6 and 8, but not significant in Model 7, meaning that the economy of the geographic region is a relatively less important element impacting uniqueness compared to university prestige. All the interactions of university prestige and log GDP are not significant in models 5–8 (all $p > 0.1$).

6. Implication

6.1. Theoretical implication

Firstly, based on the previous research (Dakota et al., 2020) that examines the gravity model of the international flow of most countries, we specifically examine the gravity model to explain the relationship between the main flow between major countries and geographical distance and find the gravity model does not set up, while the embedding distance still works. We further propose two

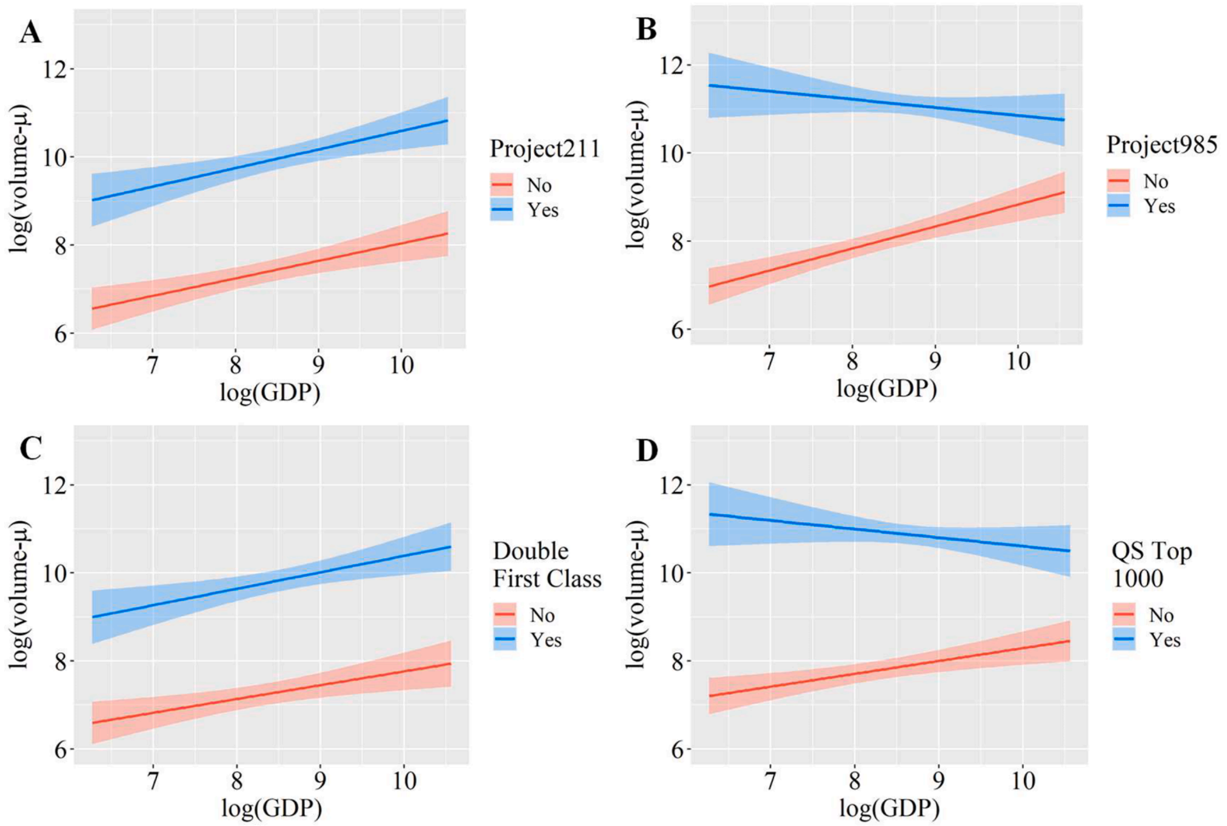

Fig. 9. Visualize regression results of Table 3. Panel A-D corresponds to Models 1–4. The X-axis is the logarithm of the GDP of the city. The Y-axis shows the estimate of $\log(volume - \mu)$. Colors represent university prestige. Solid lines represent estimated trends. Ribbons show the 95 % confidence intervals.

novel indexes, to measure how close the country is to the global scientific center and the diversity of academic mobility.

Secondly, unlike the previous research (Dakota et al., 2020) that discusses the mobility pattern in the United States via affiliation vectors, we investigate the distribution of affiliation vectors in China's Mainland and discover that in addition to geographical reasons, university types divided by preponderant disciplines are also crucial for positions of affiliation vectors. This property of Chinese university accord with China's national condition since our country experienced college and department adjustment in the 1950s.

Thirdly, based on our reasonable finding that prestige has positive effects on university volumes and the uniqueness of universities, we further investigate the interaction of prestige and regional economic status. Our research shows that volumes of universities are associated with the GDP of cities for universities of Project 985 and QS Top 1000, while for non-Project 985 and non-QS Top 1000 universities. The relationship between the volumes of universities and the GDP of cities is not statistically significant.

6.2. *Practical implication*

Firstly, based on the position of the global scientific landscape and the score of closeness and diversity, national policymakers can gain information on the gap between academically powerful countries and their countries, and take actions to foster integration with global scientific circulation and promote internal prosperity in knowledge dissemination.

Secondly, the two indexes, volumes, and uniqueness, provide insights into assessing the research impact of universities, which is associated with university prestige and regional economic status. These results provide information to university leaders that universities should not only stress being big and strong but also focus on preponderant disciplines.

7. Discussion

In this paper, we studied the stratifications of scientific institutions in terms of the scientific mobility of scientists. we recovered the scientific mobility trajectories of scientists from the large-scale Open Alex dataset and built the word2vec neural embedding model and UMAP dimension reduction method to represent and visualize affiliations in a two-dimensional space. Our results reveal the status of Chinese universities in the global talent circulation system and how the mobility patterns reflect the university prestige in China.

Our research also highlights the advantage of neural embeddings. First, embedding distance is consistent with geography distance

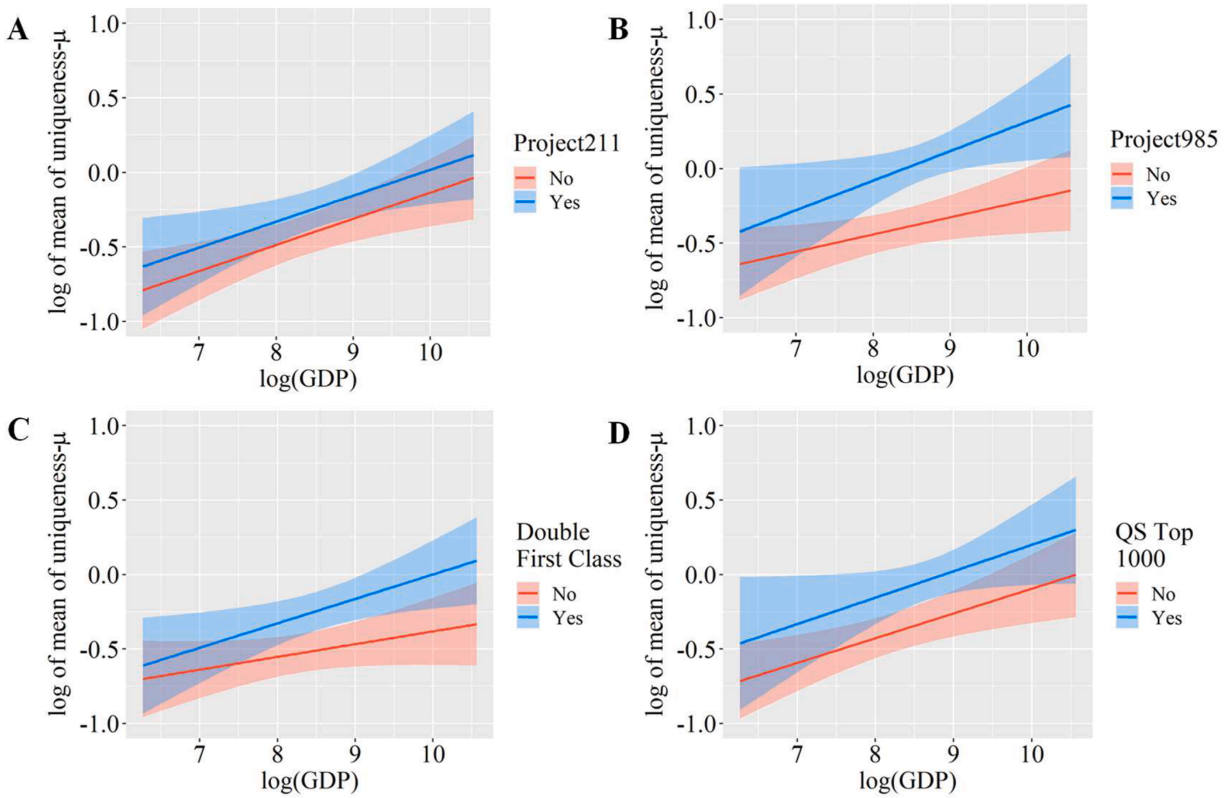

Fig. 10. Visualize regression results of Table 4. Panel A-D corresponds to Model 5–8. The X-axis is the logarithm of the GDP of the city. The Y-axis shows the estimate of *uniqueness* $-\ \mu$. Colors represent university prestige. Solid lines represent estimated trends. Ribbons show 95 % confidence intervals.

in most cases. From the perspective of the world, affiliations can be divided according to their countries. When we zoom into China, most universities in the south can also be separated. Furthermore, our model can capture features that cannot be revealed by geographical information, such as university type and distinctiveness related to the rank of universities.

Our study contributes valuable insights into the dynamics of scientific mobility, global collaboration, and the relationship between university prestige and research output. It provides a basis for informed decision-making to enhance scientific productivity and foster collaborative efforts for scientific advancements on a global scale. Especially in China, as the study shows the correlation between university prestige and mobility outputs (volume) signifies the significance of prestigious institutions in attracting and collaborating with researchers. Policymakers and funding agencies can use this information to assess the research impact of universities and consider it while making decisions about research funding and recognition. Moreover, by recognizing the importance of mobility in driving scientific collaboration, policymakers can promote policies and initiatives that facilitate international researcher exchanges, joint research projects, and collaborative networks. This fosters cross-pollination of ideas and expertise, contributing to advancements in various fields.

There are also some unavoidable limitations of this study. First, due to the dynamic nature of scientific collaboration and mobility, the study's findings may only represent a snapshot of the scientific landscape at a specific time. Changes and developments occurring after the data collection could impact the patterns of mobility and collaboration, making the study's conclusions time-sensitive. Moreover, the criteria used to classify universities based on prestige may introduce selection bias, as different classification methods may emphasize certain characteristics over others. This bias could affect the interpretation of the relationship between university prestige and research output. Third, although the large scale of OpenAlex, which also contains a small fraction of papers written in Chinese, may lack full coverage, especially in the field of humanity and social science. To overcome these limitations, future research should consider integrating scientific corpus in different languages, employing longitudinal data to track changes over time, and utilizing more comprehensive and unbiased methods for evaluating university prestige and collaboration patterns. Taking these steps would enhance the robustness and applicability of the study's findings in a broader context.

For further study, integrating scientific mobility embedding with text-based embedding holds great potential to uncover deeper insights into the intricate relationship between researcher migration and the transformation of research topics. For example, conducting longitudinal analyses will enable researchers to track the evolution of research topics and mobility patterns over extended periods. This longitudinal approach can reveal how the interplay between mobility and research themes evolves and how it shapes the overall scientific landscape. It is also interesting to investigate how migration influences the emergence and shift of new research topics can provide valuable insights into the process of knowledge creation and dissemination. This exploration can help predict emerging scientific trends and contribute to the identification of emerging research hubs.

Funding information

This work was supported by grants from the National Natural Science Foundation of China No. NSFC62006109 and NSFC12031005, the 13th Five-year plan for Education Science Funding of Guangdong Province No. 2020GXJK457, the Stable Support Plan Program of Shenzhen Natural Science Fund No. 20220814165010001, and partly sponsored by SUSTech Research Series No. SUSTECH2020C007.

CRediT authorship contribution statement

Yongshen He: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Yurui Huang:** Supervision, Writing – original draft. **Chaolin Tian:** Writing – original draft, Writing – review & editing. **Shibing Xiang:** Writing – original draft, Writing – review & editing. **Yifang Ma:** Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors have no competing interests.

Data availability

We used the OpenAlex dataset, which is publicly available.

Acknowledgments

We thank OpenAlex for the scientific corpus dataset. The computation in this study was supported by the Center for Computational Science and Engineering of the Southern University of Science and Technology.

References

Ackers, L. (2005). Moving people and knowledge: Scientific mobility in the European Union1. *International Migration, 43*(5), 99–131.
Alessandretti, L., Aslak, U., & Lehmann, S. (2020). The scales of human mobility. *Nature, 587*(7834), 402–407. https://doi.org/10.1038/s41586-020-2909-1
Allen, R. M. (2021). Commensuration of the globalised higher education sector: How university rankings act as a credential for world-class status in China. *Compare: A Journal of Comparative and International Education, 51*(6), 920–938.
Azoulay, P., Ganguli, I., & Zivin, J. G. (2017). The mobility of elite life scientists: Professional and personal determinants. *Research Policy, 46*(3), 573–590.
Bagheri, E., Ensan, F., & Al-Obeidat, F. (2018). Neural word and entity embeddings for ad hoc retrieval. *Information Processing & Management, 54*(4), 657–673. https://doi.org/10.1016/j.ipm.2018.04.007
Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., et al. (2018). Human mobility: Models and applications. *Physics Reports, 734*, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001
Baruffaldi, S. H., & Landoni, P. (2012). Return mobility and scientific productivity of researchers working abroad: The role of home country linkages. *Research Policy, 41*(9), 1655–1665. https://doi.org/10.1016/j.respol.2012.04.005
Baty, P. (2004). Times Higher Education World University Rankings. Retrieved 2023/7/4 from www.timeshighereducation.com/world-university-rankings.
Chen, Y., Long, C., Cong, G., & Li, C. (2020). Context-aware deep model for joint mobility and time prediction. In *Proceedings of the 13th international conference on web search and data mining*. https://doi.org/10.1145/3336191.3371837
Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., Costas, R., & Sugimoto, C. R. (2018). Travel bans and scientific mobility: Utility of asymmetry and affinity indexes to inform science policy. *Scientometrics, 116*(1), 569–590. https://doi.org/10.1007/s11192-018-2738-2
Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., & Schwenk, H. et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances, 1*(1), Article e1400005. https://doi.org/10.1126/sciadv.1400005
Consultancy, S. (2003). ShanghaiRanking Retrieved 2003/8/4 from www.shanghairanking.com.
Czaika, M., & Orazbayev, S. (2018). The globalisation of scientific mobility, 1970–2014. *Applied Geography, 96*, 1–10.
Dakota, M., Jisung, Y., Sadamori, K., Rodrigo, C., Woo-Sung, J., & Staša, M. et al. (2020). Unsupervised embedding of trajectories captures the latent structure of mobility. arXiv preprint arXiv:2012.02785.
Daugėlienė, R., & Marcinkevičienė, R. (2009). Brain circulation: Theoretical considerations. *Inžinerinė Ekonomika*, (3), 49–57.
Deville, P., Wang, D., Sinatra, R., Song, C., Blondel, V. D., & Barabási, A. L. (2014). Career on the Move: Geography, Stratification and Scientific Impact. *Scientific Reports, 4*(1), 4770. https://doi.org/10.1038/srep04770
Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Donal, J. (2001). wikipedia. Retrieved 2023/8/4 from wikipedia.org.
Elnagar, A., Yagi, S., Mansour, Y., Lulu, L., & Fareh, S. (2023). A benchmark for evaluating Arabic contextualized word embedding models. *Information Processing & Management, 60*(5), Article 103452. https://doi.org/10.1016/j.ipm.2023.103452
Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. kdd.
Fixman, M. (1962). Radius of gyration of polymer chains. *The Journal of Chemical Physics, 36*(2), 306–310.
Gargiulo, F., & Carletti, T. (2014). Driving forces of researchers mobility. *Scientific Reports, 4*(1), 4860. https://doi.org/10.1038/srep04860

Gaulé, P. (2014). Who comes back and when? Return migration decisions of academic scientists. *Economics Letters, 124*(3), 461–464. https://doi.org/10.1016/j.econlet.2014.07.014

Gibson, J., & McKenzie, D. (2014). Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific. *Research Policy, 43*(9), 1486–1495. https://doi.org/10.1016/j.respol.2014.04.005

Gureyev, V. N., Mazov, N. A., Kosyakov, D. V., & Guskov, A. E. (2020). Review and analysis of publications on scientific mobility: Assessment of influence, motivation, and trends. *Scientometrics, 124*(2), 1599–1630. https://doi.org/10.1007/s11192-020-03515-4

Han, S., Park, S., Park, S., Kim, S., & Cha, M. (2020). Mitigating embedding and class assignment mismatch in unsupervised image classification. In *Proceedings of the European conference on computer vision*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hu, A., & Vargas, N. (2015). Economic consequences of horizontal stratification in postsecondary education: Evidence from urban China. *Higher Education, 70*, 337–358.

Jöns, H. (2011). Transnational academic mobility and gender. *Globalisation, Societies and Education, 9*(2), 183–209. https://doi.org/10.1080/14767724.2011.577199

Jałowiecki, B., & Gorzelak, G. J. (2004). Brain drain, brain gain, and mobility: Theories and prospective methods. *Higher Education in Europe, 29*(3), 299–308.

Ji, S., Yun, H., Yanardag, P., Matsushima, S., & Vishwanathan, S. (2015). Wordrank: Learning word embeddings via robust ranking. arXiv preprint arXiv:1506.02761.

Kipf, T.N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Li, F., & Tang, L. (2019). When international mobility meets local connections: Evidence from China. *Science and Public Policy, 46*(4), 518–529. https://doi.org/10.1093/scipol/scz004

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association, 113*(521), 390–400.

Liu, M., & Hu, X. (2022). Movers' advantages: The effect of mobility on scientists' productivity and collaboration. *Journal of Informetrics, 16*(3), Article 101311. https://doi.org/10.1016/j.joi.2022.101311

Liu, S., Maljovec, D., Wang, B., Bremer, P. T., & Pascucci, V. (2016). Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics, 23*(3), 1249–1268.

Liu, X., Wu, K., Liu, B., & Qian, R. (2023). HNERec: Scientific collaborator recommendation model based on heterogeneous network embedding. *Information Processing & Management, 60*(2), Article 103253. https://doi.org/10.1016/j.ipm.2022.103253

Lixu, L. (2004). China's higher education reform 1998–2003: A summary. *Asia Pacific Education Review, 5*(1), 14–22. https://doi.org/10.1007/BF03026275

McCormick, C. (2016). Word2vec tutorial-the skip-gram model. Apr-2016.[Online]. Available: http://mccormickml.-com/2016/04/19/word2vec-tutorial-the-skip-gram-model.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing*. EMNLP.

Peters, M. A., & Besley, T. (2018). China's double first-class university strategy: 双一流. *Educational Philosophy and Theory, 50*(12), 1075–1079. https://doi.org/10.1080/00131857.2018.1438822

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). Deep contextualized word representations. In *Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers)*.

Petersen, A. M. (2018). Multiscale impact of researcher mobility. *Journal of The Royal Society Interface, 15*(146), Article 20180580.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint arXiv:2205.01833.

Radford, A., & Narasimhan, K. (2018). Improving language understanding by generative pre-training.

Ren, Y., Chen, H., Han, Y., Cheng, T., Zhang, Y., & Chen, G. (2020). A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. *International Journal of Geographical Information Science, 34*(4), 802–823.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. San Diego La: Institute for Cognitive Science, University of California.

Sattler, F., Müller, K. R., & Samek, W. (2020). Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Transactions on Neural Networks and Learning Systems, 32*(8), 3710–3722.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., et al. (2021). The universal visitation law of human mobility. *Nature, 593*(7860), 522–527. https://doi.org/10.1038/s41586-021-03480-9

Shen, H., Coreynen, W., & Huang, C. (2022). Exclusive licensing of university technology: The effects of university prestige, technology transfer offices, and academy-industry collaboration. *Research Policy, 51*(1), Article 104372. https://doi.org/10.1016/j.respol.2021.104372

Shen, H., Coreynen, W., & Huang, C. (2023). Prestige and technology-transaction prices: Evidence from patent-selling by Chinese universities. *Technovation, 123*, Article 102710. https://doi.org/10.1016/j.technovation.2023.102710

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., & Woo, W.C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 28.

Simini, F., González, M. C., Maritan, A., & Barabási, A. L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96–100. https://doi.org/10.1038/nature10856

Stone, C.J. (2008). an asymptotically optimal histogram selection rule.

Sutskever, I., Vinyals, O., & Le, Q.V. (2014). Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.

Symonds, Q. (2004). QS ranking website Retrieved 2023/6/20 from https://www.topuniversities.com/.

Tabassum, N., Menon, S., & Jastrzębska, A. (2022). Time-series classification with SAFE: Simple and fast segmented word embedding-based neural time series classifier. *Information Processing & Management, 59*(5), Article 103044. https://doi.org/10.1016/j.ipm.2022.103044

Terroso-Saenz, F., Flores, R., & Muñoz, A. (2022). Human mobility forecasting with region-based flows and geotagged Twitter data. *Expert Systems with Applications, 203*, Article 117477. https://doi.org/10.1016/j.eswa.2022.117477

Tian, C., Zhu, X., Hu, Z., & Ma, J. (2020). Deep spatial-temporal networks for crowd flows prediction by dilated convolutions and region-shifting attention mechanism. *Applied Intelligence, 50*(10), 3057–3070. https://doi.org/10.1007/s10489-020-01698-0

Tomas, M., Kai, C., Greg, C., & Jeffrey, D. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

Tranter, P. (2012). Effective speed: Cycling because it's faster. *City Cycling, 57*, 74.

Trinh, T.H., Luong, M.T., & Le, Q.V. (2019). Selfie: Self-supervised pretraining for image embedding. arXiv preprint arXiv:1906.02940.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., & Gomez, A.N. et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Wagner, C. S., & Jonkers, K. (2017). Open countries have strong science. *Nature, 550*(7674), 32–33. https://doi.org/10.1038/550032a

Wang, J., Zhang, Q., Shi, F., Li, D., Cai, Y., Wang, J., et al. (2023). Knowledge graph embedding model with attention-based high-low level features interaction convolutional network. *Information Processing & Management, 60*(4), Article 103350. https://doi.org/10.1016/j.ipm.2023.103350

Washington, D. C. (1933). U.S. News & World Report. www.usnews.com.

Weakliem, D. L., Gauchat, G., & Wright, B. R. E. (2012). Sociological stratification: Change and continuity in the distribution of departmental prestige, 1965–2007. *The American Sociologist, 43*(3), 310–327. https://doi.org/10.1007/s12108-011-9133-2

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*.

Wut, T. M., Xu, J., & Lee, S. W. (2022). Does university ranking matter? Choosing a university in the digital era. *Education Sciences, 12*(4), 229. https://www.mdpi.com/2227-7102/12/4/229.

Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. In *Proceedings of the international conference on machine learning*.

Xu, S., Pi, D., Cao, J., & Fu, X. (2021). Hierarchical temporal–spatial preference modeling for user consumption location prediction in Geo-Social Networks. *Information Processing & Management, 58*(6), Article 102715. https://doi.org/10.1016/j.ipm.2021.102715

Yang, D., Fankhauser, B., Rosso, P., & Cudre-Mauroux, P. (2020). Location prediction over sparse user mobility traces using rnns. In *Proceedings of the 29th international joint conference on artificial intelligence*.

Yang, J., Wu, Q., & Wang, C. (2022). Research networks and the initial placement of PhD holders in academia: Evidence from social science fields. *Scientometrics, 127*(6), 3253–3278. https://doi.org/10.1007/s11192-022-04394-7

Ying, C. (2011). A reflection on the effects of the 985 project. *Chinese Education & Society, 44*(5), 19–30. https://doi.org/10.2753/CED1061-1932440502

Yuan, C., Jia, Y., Fang, K., & Liu, Z. (2009). Effects of university type on university knowledge transfer in China: A resource−based view. *Science of Science and Management of S&T, 30*(7), 134–138.

Zanon, G. (2021). Who Is Leaving?-An Investigation of the European Research Landscape Universidade NOVA de Lisboa (Portugal)].

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI conference on artificial intelligence*.

Zhang, X., Wang, Q., Ye, Z., Ying, H., & Yu, D. (2023). Federated representation learning with data heterogeneity for human mobility prediction. *IEEE Transactions on Intelligent Transportation Systems, 24*(6), 6111–6122. https://doi.org/10.1109/TITS.2023.3252029

Zipf, G. K. (1946). The P 1 P 2 D hypothesis: On the intercity movement of persons. *American Sociological Review, 11*, 677.

Zou, B., Wang, Y., Kwoh, C. K., & Cen, Y. (2023). Directed collaboration patterns in funded teams: A perspective of knowledge flow. *Information Processing & Management, 60*(2), Article 103237.