https://doi.org/10.51408/issi2025_165

Productivity and Impact Patterns in Scientific Careers

Kaile Gong

gongkaile@njnu.edu.cn
School of Journalism and Communication, Nanjing Normal University, Nanjing, Jiangsu 210097 (China)

Introduction

The patterns of scientific careers have long been of interest to scientometrics (Sinatra et al., 2016). Mobility, especially international mobility, is widely recognized to have a significant impact on the development of scientific careers (Netz, Hampel & Aman, 2020). Although existing research has made many beneficial discoveries, they often rely on some small samples of elite scientists (e.g. Nobel Prize winners), with insufficient exploration of broader patterns and the role of international mobility in them. Therefore, this study takes PubMed as the data source, adopts the method of time series clustering to reveal multiple patterns of productivity and impact in the academic careers of 67,201 scientists, and then uses the Chi-square test to analyze the influence of international mobility.

Methodology

Data

The data was collected from the PubMed Knowledge Graph 2.0 (PKG 2.0), which is an open dataset built by Xu et al. (2024). PKG 2.0 provides PubMed-indexed papers published before 2024 and has high-quality author disambiguation with an F1 score of 96.24%. More importantly, it integrates multi-source data and maps partial PubMed authors to Orcid scholars, which offers accurate information about scientists' education and employment. Thus, both publications and international mobility can be identified and analyzed based on PKG 2.0. Since PubMed is a biomedical and life science database, this study's findings are applicable to this field.

In order to ensure that the selected scientists have a sufficiently long and continuous career and that their mobility can be identified via Orcid, this study draws on the approach of Sinatra et al. (2016) and applies four inclusion criteria: (1) the scientists should have at least a 30-year publication career; (2) the scientists should author at least one paper every 5 years; (3) the scientists should publish at least 30 papers; and (4) the scientists should have education and employment records in Orcid. Finally, the samples for analysis include 67,201 scientists and 8,769,452 papers they published from 1936 to 2023.

Productivity, impact and international mobility

Productivity refers to the number of papers published within a certain time range, so for each scientist, the number of papers published each year is counted to obtain the yearly publication sequence.

Impact refers to the citation impact of the most cited paper published by an author within a time range, specifically, it's defined as the highest 5-year citation count of a paper published within that time. Thus, for each scientist, the 5-year citation count of all papers published before 2018 is first counted by considering the 5-year citation window, then the most cited paper in each year is found and its 5-year citation count is used as the impact indicator in that year. Finally, the yearly impact sequence of each scientist is obtained.

International mobility is identified for each scientist based on the presence of two or more different countries in their Orcid education and employment records.

Time series clustering

Dynamic time warping (DTW) and K-medoids are combined as the time series clustering method to detect productivity and impact patterns in scientists' careers. DTW is the most popular and widely accepted method for measuring the similarity between time series data with different lengths (Ao et al., 2023). K-Medoids is a clustering algorithm similar to K-means, but it selects real points existing in the dataset as cluster centroids instead of calculating the average of all points, which makes K-Medoids more robust in handling noise and outliers (Arora & Varshney, 2016).

Let's take productivity patterns clustering as an example to briefly explain the clustering process: firstly, the Python package TAIDistance (Meert et al., 2022) is used to calculate the pairwise DTW distance between the yearly publication sequence of all scientists to form the distance matrix, and then the distance matrix is input into the K-medoids to implement clustering. The elbow method based on inertia value is used to determine the number of clusters, which is 3 in the clustering of productivity patterns and 4 in the clustering of impact patterns.

The influence of international mobility

The Chi-square test is used to analyze whether there are significant differences in the distribution of mobile and non-mobile scientists in different productivity (impact) patterns.

Results

The productivity and impact patterns in scientific careers are respectively shown in Fig. 1 and Fig. 2. The green lines in the figures are the sequences of K-Medoids centroids and the red lines are the modified centroids using the DTW Barycenter Averaging (DBA) algorithm, which can better represent each pattern.

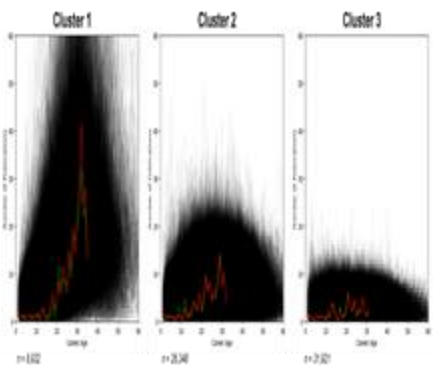

Figure 1. Productivity patterns in.

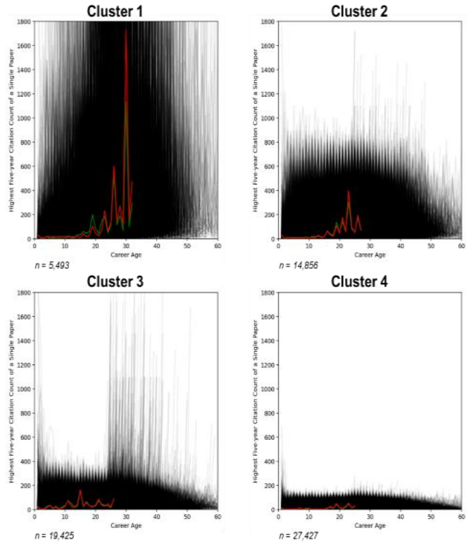

Figure 2. Impact patterns in scientific careers.

Table 1. Influence of mobility on productivity patterns.

		Productivity patterns			χ^2	p
		Cluster 1	Cluster 2	Cluster 3		
Mobility types	Mobility	6,311	16,240	16,659	1172.362	0.000***
	%Mobility	16.1%	41.4%	42.5%		
	%Cluster *i*	70.7%	61.6%	52.2%		
	Non-mobility	2,621	10,108	15,262		
	%Non-mobility	9.4%	36.1%	54.5%		
	%Cluster *i*	29.3%	38.4%	47.8%		

*** p<0.001

Table 2. Influence of mobility on impact patterns.

		Impact patterns				χ^2	p
		Cluster 1	Cluster 2	Cluster 3	Cluster 4		
Mobility types	Mobility	3,809	9,869	11,536	13,996	1285.608	0.000***
	%Mobility	9.7%	25.2%	29.4%	35.7%		
	%Cluster *i*	69.3%	66.4%	59.4%	51.0%		
	Non-mobility	1,684	4,987	7,889	13,431		
	%No-mobility	6.0%	17.8%	28.2%	48.0%		
	%Cluster *i*	30.7%	33.6%	40.6%	49.0%		

*** p<0.001

It can be concluded from Fig. 1 that the productivity patterns include three types: high peak (Cluster 1), moderate peak (Cluster 2), and low fluctuation (Cluster 3), and peaks often appear in the late stage of scientific careers. According to Fig. 2, the impact patterns include four types: high peak (Cluster

1), moderate peak (Cluster 2), low peak (Cluster 3), and flat (Cluster 4), and the age of the peak is advanced with the peak value decreasing. In addition, it is found that no matter productivity or impact, there's little difference between various patterns during the first 1/3 career, but after that, there's a clear divergence. The potential policy implication of the findings is that we need to be more patient with scientific career development, and what triggers the divergence deserves future attention.

Table 1 and Table 2 show that mobile and non-mobile scientists have significant differences in the distribution of productivity and impact patterns, and mobile scientists are more likely to achieve relatively higher peaks, which can be inferred that mobility is beneficial to scientific career development.

Acknowledgments

This study was funded by the National Social Science Fund of China (No. 23CTQ032).

References

Ao, W., Lyu, D., Ruan, X., Li, J., & Cheng, Y. (2023). Scientific creativity patterns in scholars' academic careers: Evidence from PubMed. *Journal of Informetrics, 17*(4), 101463.

Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science*, 78, 507-512.

Meert, W., Hendrickx K., Van Craenendonck, T., Robberechts, P., Blockeel, H., & Davis, J. (2022). DTAIDistance (Version v2). *Zenodo*, http://doi.org/10.5281/zenodo.5901139.

Netz, N., Hampel, S., & Aman, V. (2020). What effects does international mobility have on scientists' careers? A systematic review. *Research Evaluation, 29*(3), 327-351.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science, 354*(6312), aaf5239.

Xu, J., Yu, C., Xu, J., Ding, Y., Torvik, V. I., Kang, J., ... & Song, M. (2024). PubMed knowledge graph 2.0: Connecting papers, patents, and clinical trials in biomedical science. *arXiv preprint arXiv:2410.07969*.