SCIENCE CHINA
Information Sciences

• MOOP •

October 2018, Vol. 61 104101:1–104101:3
https://doi.org/10.1007/s11432-018-9469-5

CareerMap: visualizing career trajectory

Kan WU[1], Jie TANG[1*], Zhou SHAO[2], Xinyi XU[3], Bo GAO[1] & Shu ZHAO[4]

[1]*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084*, China;*
[2]*School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China;*
[3]*International School, Beijing University of Posts and Telecommunications, Beijing* 100876*, China;*
[4]*School of Computer Science and Technology, Anhui University, Hefei* 230601*, China*

Received 10 January 2018/Revised 3 April 2018/Accepted 22 May 2018/Published online 31 August 2018

Citation Wu K, Tang J, Shao Z, et al. CareerMap: visualizing career trajectory. Sci China Inf Sci, 2018, 61(10): 104101, https://doi.org/10.1007/s11432-018-9469-5

This study introduces a system named CareerMap that visualizes a scholar's career trajectory. As an online demonstration of CareerMap, we have shown the visualization result by CareerMap for the AMiner 2016 most influential scholars in machine learning (ML)[1)]. Each trajectory path on the map represents the movement of a scholar between different places (affiliations). The heatmap reveals the geographic distribution of the most influential ML scholars; a larger hotspot means a larger immigration of scholars into an affiliation. The right sidebar is a list of all scholars. When the user selects (clicks on) a scholar, the trajectory path of that scholar is highlighted in the map. The bottom bar shows the timeline. When the user selects a specific year, a textbox at the bottom displays the most important work (paper) published in that year by a scholar in the right-hand list. This example provides the track records of over half of the most influential ML scholars at the east and west coasts of the USA, and in west Europe. By zooming in, the user can also check the city-level results or obtain finer details.

CareerMap can benefit many applications. For example, if the movements of all experts in artificial intelligence (AI) worldwide were displayed on a visual map, government strategy departments could better understand the talent distribution and accordingly design wise AI strategies. Similarly, CareerMap can assist the design of smart recruiting plans by human resource (HR) departments of companies seeking talented employees. Individual users such as students can use the map to locate the best advisors for their Ph.D. studies.

A scholar's trajectory information is extracted from scientific publication data in AMiner [1], which has collected 130000000 researcher profiles and more than 200000000 papers from multiple publication databases. Since operations began in 2006, the system has attracted more than 8000000 independent IP accesses from over 200 countries/regions. Our extracted trajectory information is represented as a five-tuple of $\langle \text{name}, \text{affi}, \text{year}, \text{longitude}, \text{latitude} \rangle$, where name and affi represent the scholar's name and affiliation (extracted from the publication papers published by the scholar), respectively, year is the publication year of the extracted paper, and longitude and latitude are the geographic location information inferred from the extracted affiliation using Google map API. The main problem is extracting all the accurate affiliation information of all authors.

This problem is much more challenging than might be expected. First, the affiliation information is usually unavailable for various reasons. For example, even in some very professional publication databases such as ACM digital library and IEEE Xplore digital library, only 50% of the papers include complete affiliation information. The second challenge is name ambiguity, which arises

* Corresponding author (email: jietang@tsinghua.edu.cn)
1) https://www.aminer.cn/mostinfluentialscholar/ml.

when different scholars have the same name or the same person uses different names [2]. The last challenge lies in the visualization itself: how to determine the visualization granularity when the trajectories of many scholars are displayed on the same map.

Although the existing literature includes a few works on graph visualization, it scarcely reports on career trajectories. For example, Shi et al. [3] presented a method that visualizes the influence graph summaries on citation networks, but neglects the career trajectory information. Yang et al. [4] proposed a search-based visualization of the relationships between advisor and advisees. The present paper presents a novel system for visualizing the career trajectories of scholars. We first introduce the architecture of CareerMap and its major functions, and then present several intriguing discoveries from the visualization results. Finally, we produce a dynamic trajectory map of the top 10000 MI scholars (ordered by the h-index) in the world. The map includes all dynamic information (e.g., when a scholar moved to a different affiliation). A video demonstration of the dynamic map is publicly available online[2)].

Architecture and main features. Figure 1(a) shows the architecture of CareerMap. The system comprises two subsystems: career trajectory extraction and analytic visualization. Career trajectory extraction (bottom block in Figure 1(a)) aims to extract the career trajectory path of each scholar. As mentioned above, this subsystem must overcome data incompletion and name ambiguity. Name ambiguity is a common problem in applications such as scientific literature management and information integration. In AMiner, this problem is treated by a unified probabilistic model [2], which provides an overall performance (pairwise F1-score) of 88.80%.

To address the data incompletion problem, we propose a spatio-temporal factor graph model (STFGM), which infers the missing affiliations from spatio-temporal correlations. The STFGM was evaluated on two ground-truth datasets that were manually labeled by annotators. On the first dataset from AMiner and the second dataset from Microsoft academic graph (MAG), the overall precisions of the proposed STFGM were 80.07% and 76.68%, respectively. We then introduced a smoothing technique that improved the performances of the AMiner and MAG datasets by +5% and +10%, respectively. The details of the proposed model are beyond the scope of this article and will be reported elsewhere.

As an example, we plot the career trajectory of a Turing award winner, professor Tim Berners-Lee, extracted from his AMiner profiling page[3)]. By combining the trajectory of all experts, we can generate a dynamic trajectory map. CareerMap also offers rich statistics for the generated trajectory map. Figure 1(b) shows the statistics of the trajectory paths of the most influential ML scholars in 2016. The statistics reveal several interesting phenomena: (1) more than half of the movements (affiliation changes) occurred in the USA; (2) the USA received many scholars from Canada and the UK; and (3) there were large movements from USA to Israel, which probably explains why Israel consistently produces cutting-edge research.

Analysis and discussions. We present further analyses and discussions of the top 10000 scholars in AMiner as a case study. Figure 1(c) plots the distribution of the number of active experts over time. An active expert is defined as a scholar who has published one or more papers in the corresponding year. For example, if all papers of a given scholar were published between 1940 and 1950, that scholar was an active expert in the 1940–1950 period only. The peak shows that most of the top experts have remained active over the past ten years.

Map hotspots were generated by clustering active experts based on the distances between their locations. More specifically, we set the maximum radius of each hotspot to 50 km[4)]. We emphasize that the threshold radius (which dictates the hotspot area) is constant for different populations and affiliation densities of countries. Thereby, we can observe the high-density spots that attract a large number of experts. Figure 1(d) shows the number distribution of the hotspots over time. Intriguingly, the two largest declines coincide with the largest economic recessions (followed by funding cuts) worldwide.

Figure 1(e) shows another interesting result: the expert immigration trends into several big cities. Clearly, Boston received the largest number of scholars, followed by Baltimore, but the immigration trends of both cities turned downward after 2005. In contrast, the immigration trend of Beijing has significantly increased in recent years.

Summary. This study introduced CareerMap, a system that visualizes the career trajectories of scholars. The technologies developed in this system are quite general and applicable to other data

2) http://static.aminer.org/lab-datasets/trajectory/top10000career-trajectory.mp4.
3) https://www.aminer.cn/profile/tim-berners-lee/5485f381dabfae9b40133a8b.
4) The maximum radius of 50 km was empirically set after trialing different values and receiving user feedbacks.

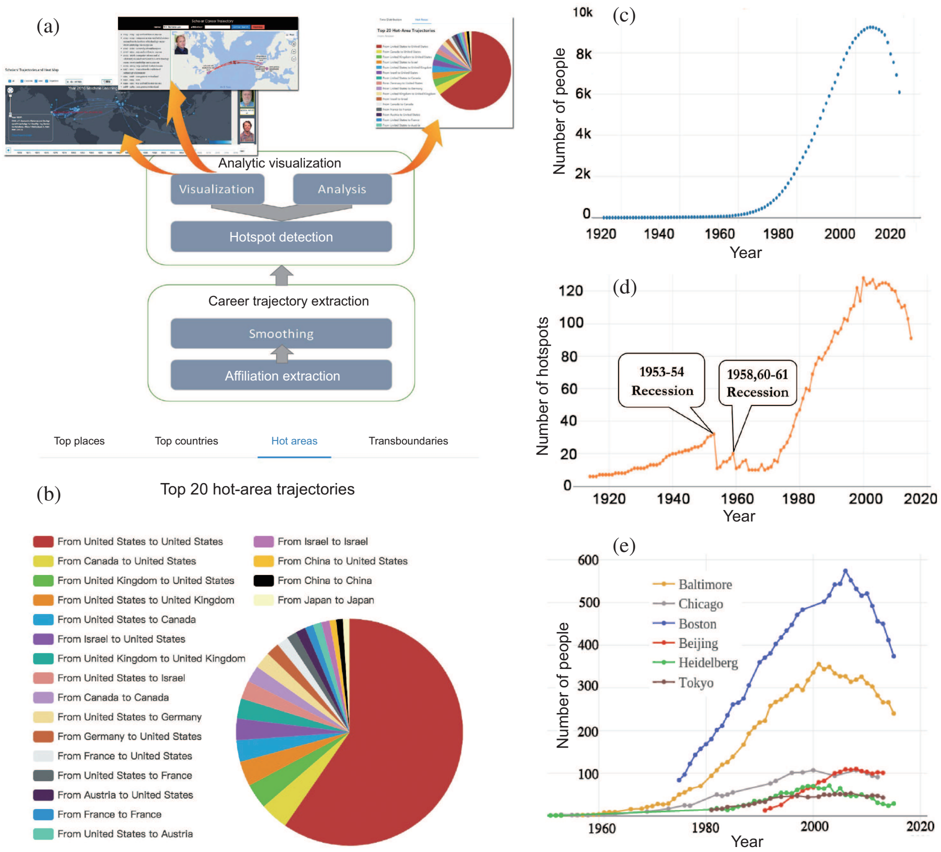

Figure 1 (Color online) (a) System architecture of CareerMap; (b) statistics of the trajectory paths of the 2016 most influential scholars in ML; (c) temporal distribution and (d) hotspot distribution of top 10000 scholars in the AMiner database; (e) scholar immigration dynamics of several big cities.

containing personal career information. The system is available online and all feedbacks are welcome.

Acknowledgements This work was supported by National Natural Science Foundation of China (NSFC) (Grant No. 61561130160) and Royal Society-Newton Advanced Fellowship Award.

Supporting information Videos and other supplemental documents. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Tang J, Zhang J, Yao L M, et al. ArnetMiner: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, 2008. 990–998
2 Tang J, Fong A C M, Wang B, et al. A unified probabilistic framework for name disambiguation in digital library. IEEE Trans Knowl Data Eng, 2012, 24: 975–987
3 Shi L, Tong H H, Tang J, et al. VEGAS: visual influence graph summarization on citation networks. IEEE Trans Knowl Data Eng, 2015, 27: 3417–3431
4 Yang Z, Tang J, Wang B, et al. Expert2Bole: from expert finding to bole search. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2009