

Scholar Trajectory: Visualizing the Migration of Talents

Zhou Shao
Nanjing University of Science and
Technology
shaozhou0001@gmail.com

Jie Tang*
Tsinghua University
jietang@tsinghua.edu.cn

Yutao Zhang
Tsinghua University
zhangyutao1106@gmail.com

Bo Gao
Tsinghua University
elivao@gmail.com

Yongli Wang
Nanjing University of Science and
Technology
yongliwang@mail.njust.edu.cn

ABSTRACT

In this paper, we present Scholar Trajectory to visualize and analyze the migration of research scholars. We extract the temporal and spatial footprints of scholars from the affiliation strings in their papers. The visual analyses are conducted at both individual level and group level. At individual level, we illustrate the location statistic and the academic achievements of each individual scholar. At group level, we explore the collective migration pattern of a research community. Empirical case studies verifies the informativeness and intuitiveness of the system.

ACM Reference Format:

Zhou Shao, Jie Tang, Yutao Zhang, Bo Gao, and Yongli Wang. 2018. Scholar Trajectory: Visualizing the Migration of Talents. In *Proceedings of 24th ACM SIGKDD Conference (KDD 2018)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nmmnnnnn.nmmnnnnn>

1 INTRODUCTION

The temporal and spatial trajectories exhibit the career mobility of a person, which is valuable for talent management, recruitment, and trend analysis. In this paper, we present Scholar Trajectory¹ for mining and visualizing migration trajectories of research scholars in different fields.

Scholar Trajectory is a data-driven temporal-spatial mining system based on AMiner Academic Graph². We generate migration trajectory for researchers based on the affiliations mentioned on their publications. There are three main challenges to build such a system: 1) Author disambiguation [2, 4], i.e. constructing scholar profiles from publication databases by distinguishing authors with same names. 2) Geo-location extraction [1], i.e. mapping the affiliation string mentioned on publications to a specific geographical coordinate. 3) Trajectory visualization, i.e. designing informative and intuitive visualizations to illustrate the temporal-spatial patterns of individual scholars and research communities. In this work,

*Corresponding author (email: jietang@tsinghua.edu.cn)

¹<https://traj.aminer.cn/trajectory-index>

²<https://www.aminer.cn/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD 2018, August 19-23 2018, London, United Kingdom

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nmmnnnnn.nmmnnnnn>

we leverage the disambiguated scholar profiles from AMiner [3] and focus on solving the latter two problems. Figure 1 gives an overview of the main user interface, where the yellow curves indicate the migration trajectories.



Figure 1: An overview of Scholar Trajectory System

2 SYSTEM OVERVIEW

There are two main components in the system, namely Individual Trajectory and Group Trajectory. Individual Trajectory focuses on the temporal-spatial footprint of an individual scholar, while Group Trajectory analyses the collective migration pattern of multiple scholars. We generate the migration trajectories of scholars based on AMiner Academic Graph, which consists of more than 130 million researchers and 270 million publications. Specifically, we extract the name of the institution mentioned in the affiliation string, and leverage Google Map API to map the institution name to its corresponding geographic location. This is non-trivial due to the inherent ambiguity and variability of natural language. To this end, we formulate the problem as a sequential labeling task. An affiliation string is broken into a quadruplet $\langle scholarId, institution, year, geoLocation \rangle$ using CRF model, where *scholarId* is the unique identifier of a researcher, *year* is the publication year, *institution* is the extracted institution name from the affiliation string, and *geoLocation* is the corresponding geographic coordinates.

2.1 Visual Design

To intuitively illustrate the temporal and spatial patterns, we design three essential visual elements: *Point*, *Line*, and *Heat*. Figure 2 gives an example of the three visual elements. *Point* denotes a geographical location of scholars, where the color of a *Point* indicates the number of the scholars. *Line* is a directed curve indicating the migration route of scholars from one location to another in a certain year. *Heat* shows the degree of influence of the scholars within certain area. Figure 2 gives an example of the three visual elements.

