Journal of Informetrics 9 (2015) 69–78

Contents lists available at ScienceDirect

Journal of Informetrics

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

Understanding and modeling diverse scientific careers of researchers

Tanmoy Chakraborty[a,*], Vihar Tammana[b], Niloy Ganguly[a], Animesh Mukherjee[a]

[a] Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur 721302, India
[b] Microsoft Corporation, Bellevue, WA, United States

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form
16 November 2014
Accepted 18 November 2014
Available online 9 December 2014

Keywords:
Scientific career
Diversity
Stochastic model

ABSTRACT

This paper analyzes the diverse scientific careers of researchers in order to understand the key factors that could lead to a successful career. Essentially, we intend to answer some specific questions pertaining to a researcher's scientific career – What are the local and the global dynamics regulating a researcher's decision to select a new field of research at different points of her entire career? What are the suitable quantitative indicators to measure the diversity of a researcher's scientific career? We propose two entropy-based metrics to measure a researcher's choice of research topics. Experiments with large computer science bibliographic dataset reveal that there is a strong correlation between the diversity of the career of a researcher and her success in scientific research in terms of the number of citations. We observe that while most of the researchers are biased toward either adopting diverse research fields or concentrating on very few fields, a majority of the prominent researchers tend to follow a typical "scatter-gather" policy – although their entire careers are immensely diverse with different types of fields selected at different time periods, they remain focused primarily in at most one or two fields at any particular time point of their career. Finally, we propose a stochastic model which, quite accurately, mimics the notion of field selection process observed in the real publication dataset.

1. Introduction

> *"It is really important to do the right research as well as to do the research right. You need to do 'wow' research, research that is compelling, not just interesting."*
>
> – Richard M. Reis, Stanford University

Of all the decisions we make as an emerging scientist, none is more important than identifying the right research area, and in particular, the right research topic. The success of scientific career gets determined by these two choices. Change in scientific research career can be defined as any major change in work-role requirements or work context (Brett, 1982, 1984; Nicholson, 1984) and as a process that may result in a change of job, profession, or a change in one's orientation of work while continuing in the same job (Albert, Ashforth, & Dutton, 2000; Evered & Louis, 1981). People believe that many factors act as an active role to regulate these changes. For instance, researchers might try to align themselves with the cutting-edge

* Corresponding author. Tel.: +91 9475444030.
E-mail address: its_tanmoy@cse.iitkgp.ernet.in (T. Chakraborty).

http://dx.doi.org/10.1016/j.joi.2014.11.008
1751-1577/

research at the current time and as a result of this a change in scientific research career becomes unavoidable (Passi & Mishra, 2004). On the other hand, this career shift might be described as an effect of "saturation" in the field of a researcher leading to a switch to the other fields (Pettigrew, 1990).

In this paper, we use a massive dataset of scientific publications in computer science domain and attempt to analyze the local and the global dynamics regulating a researcher's decision to select new field of research over the entire career. In particular, we investigate how the "relatively more successful" researchers make choices to select their fields of research at different points in their career. A remarkable observation is the "relatively more successful" researchers unlike the rest of the lot tend to behave in a "scatter-gather" fashion, i.e., they seem to work in diverse research areas over their entire career span; however in each time-slice of their career they pump in their most concentrated efforts in only one particular area of research. We provide extensive evidence for this through appropriate quantification of local and global diversity measures for the choice of research topics followed by rigorous empirical analysis of a large volume of citation data to corroborate the above observation.

This observation further motivates us to build a stochastic model that can reproduce the real-world phenomenon of field selection process. Evaluations of our model through the real-world data lead us to conclude that our model, quite accurately, mimics the field selection process for all the researchers present in the dataset. Note that, we use the terms "author" and "researcher" interchangeably in the rest of the paper. We make our experimental codes available in the spirit of reproducible research: https://github.com/centrality-multiplex/Modeling-ResearcherCareer.git.

2. Related work

Recently, research on citation, co-citation and co-authorship networks has gained interest in information sciences (Börner, Chen, & Boyack, 2003; Börner, Dall'Asta, Ke, & Vespignani, 2005; Chen, 2003) and in statistical physics (Barabási et al., 2002; Newman, 2001; Redner, 2005). The accumulation of published articles enables also the drawing of evolutionary tree-like structures of referencing over time. One famous example is the idea of a "historiograph" proposed by Garfield (Garfield, 1977, 2004; Garfield, Pudovkin, & Istomin, 2003).

Field mobility, or field migration (Vlachy, 1981), is defined as scientists moving into new research topics. Field mobility can be measured by identifying different research topics (fields or subfields), estimating the activity of scientists in these fields, and following the activity of scientists over time to mark the transitions. Field mobility has been investigated already since the 1980s (Le Pair, 1980; van Houten, van Vuren, Le Pairs, & Dijkhuis, 1983). Field mobility has been discussed as the driving force for the exploration of new territories in the "landscape" of science (Scharnhorst, 2001; Urban, 1982). More specifically, field mobility has been modeled as an exchange mechanism between research fields leading to a co-evolution or coupled growth of scientific specialties (Chen, Börner, & Fang, 2013; Ebeling & Feistel, 1986). Hellsten, Lambiotte, Scharnhorst, and Ausloos (2007) introduce a new approach to detecting scientists' field mobility by focusing on an author's self-citation network, and the co-authorships and keywords in self-citing articles.

Changing patterns of scientific activity have been also discussed in the context of interdisciplinarity. Attempts to measure interdisciplinarity rely on citation and publication patterns (see e.g., Rinia, van Leeuwen, Bruins, van Vuren, & van Raan, 2002). However, some studies also follow certain authors through their publication records (Pierce, 1999; Urata, 1990). Some studies use interviews and surveys to trace academic careers but this approach is restricted to rather small case studies (van Houten et al., 1983; Wagner-Döbler & Berg, 1993). Career moves of scientists are also a topic of science history or sociology research (see for e.g., an earlier research Gilbert, 1977). Currently, there are no automated techniques for quantitatively measuring scientists' field adaptation/mobility.

On the other hand, Zhou, Ji, Zha, and Giles (2006) illustrate how topic evolution and social interaction lead to build an author's research career. Biryukov and Dong (2010) make an attempt to analyze an author's scientific career through exploration of scientific communities based on some features and use them to compare the sets of top ranked conferences with the low ranked ones. Despite well-documented literature on bibliographic dataset, the last few decades have witnessed a scarcity of empirical research on career change. The key questions related to the dynamical process of career change along with the associated outcomes remain mostly uninvestigated.

In this paper, we attempt to investigate the following research questions – (i) How a researcher decides to select her field of research at different points in her career? (ii) What are the suitable quantitative indicators to measure the diversity of a researcher's scientific career? (iii) Can we mimic the real-world field selection process of researchers using a computational model? To the best of our knowledge, this is the first attempt to analyze the scientific career of researchers extensively using new quantitative indicators. Moreover a new computational model is proposed to imitate the field selection process of researchers that unfolds the real-world dynamics controlling the shift of research career.

3. Dataset

In this experiment, we used the DBLP dataset of the computer science domain developed by Chakraborty et al. (Chakraborty, Sikdar, Ganguly, & Mukherjee, 2014; Chakraborty, Sikdar, Tammana, Ganguly, & Mukherjee, 2013). The dataset contains 702,973 valid papers and 495,311 authors. The attributes of each paper are as follows: the name of the research paper, a unique index of the paper, the list of author(s), the year of publication, the publication venue, the list of research papers the given paper cites and (in some cases) the abstract and the keywords of the paper. After a series of preprocessing