The Evolution of Scientific Productivity of Junior Scholars

Chun-Hua Tsai, University of Pittsburgh, Pittsburgh
Yu-Ru Lin, University of Pittsburgh, Pittsburgh

Abstract
Publishing academic work has been recognized as a key indicator for measuring scholars' scientific productivity and having crucial impact on their future career. However, little has been known about how the majority of researchers progress in publishing papers across disciplines. In this work, using a collection consisting of over five millions academic publications across 15 disciplines, we study how the scientific productivity patterns of junior scholars change across different generations and different domains. Our study results help understand the evolution of the competitive "publish or perish" academic culture.
Keywords: innovation; history of science; scientific revolutions
Citation: Tsai, C-H., Lin, Y-R. (2015). The Evolution of Scientific Productivity of Junior Scholars. In *iConference 2015 Proceedings*.

Contact: cht77@pitt.edu

1 Introduction

In academia, publishing academic work rapidly and continually has crucial impact on a scholar's future career. Prior work has shown that successful publications not only facilitate young scholars' job-seeking in academia and other research oriented institutions, but also bring funding opportunities to sponsor their research when they progress in their academic career (van Dijk, Manor, & Carey, 2014). Based on this work, scholars' publications are recognized as a key indicator for measuring their scientific productivity and subsequently determines their career opportunities. Moreover, recent studies on scholars' publications have brought important insights into the scientific success (Jones, 2010) and the emergence of creativity (Uzzi, Mukherjee, Stringer, & Jones, 2013; Lungeanu, Huang, & Contractor, 2014).

While scientific progress largely depends on successful scholars, little has been known about how the majority of researchers progress in publishing papers. Recently, a 16-year study from Ioannidis, Boyack, and Klavans (2014) found that less than 1% of scientists continue publish papers annually. This leads to an interesting question: *What is the publication figure of a junior scholar in general?*

Understanding the productivity of a junior scholar, however, is challenging. Most of the prior work on measuring productivity and impact of scholars relied on certain cut-off of publication quantities, citation number or academic indexes (e.g., h-index) (Ioannidis et al., 2014). These measures are not suitable for quantifying the productivity of junior scholars because their early publications are likely to be insufficient, and therefore, cannot give informative comparison. Another challenge is that the required resource, cultural support and collaboration environment for publishing papers tend to vary by disciplines, and hence it is not fair to compare scholars through fixed productivity indicators.

In this work-in-progress paper we take the first initiate to study the scientific productivity of junior scholars and how the productivity patterns change across different generations and different domains of junior scholars. Rather than emphasizing on the effect of productivity in the competitive "publish or perish" academic culture, we seek to understand the evolution of this competitive environment and how the competition varies across different disciplines. We study over one million of junior scholars and compare the productivity patterns over twelve years and across 15 different domains, and investigate the association between productivity and collaboration structure. We found that, over time, the productivity in medical domain rises more steeply than that in STEM and non-STEM (including arts and humanities, economics and other social sciences) domains. In STEM domains, the association between productivity and collaboration structure is the highest. The collaboration structure in medical domain has weaker correlation with scholars' productivity on their early stage, but the collaboration effect increases more rapidly than that in non-STEM domains. We discuss the implications of our study and future work.

(a) Publications in each categories

Category	Domain	Publications	Authors
Med	medicine	1,721,580	421,952
non-Stem	agriculturescience	63,201	10,681
non-Stem	artshumanities	1,146	152
non-Stem	economics	511,988	118,683
non-Stem	geosciences	1,179	195
non-Stem	multidisciplinary	9,854	1,395
non-Stem	socialscience	543	59
Stem	biology	102,772	22,411
Stem	chemistry	274,214	57,351
Stem	computerscience	952,492	179,480
Stem	engineering	46,042	9,696
Stem	materialscience	7,687	1,123
Stem	mathmatics	622	172
Stem	physics	54,172	12,277
Other	Other	1,262,399	152,336

(b) Publication trend

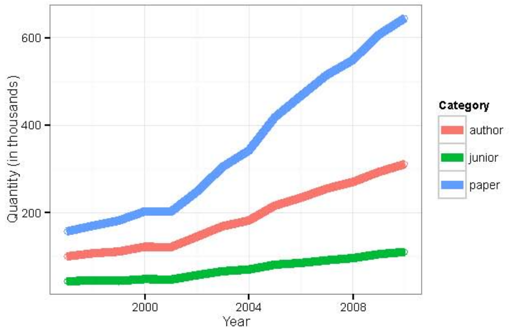

Figure 1: (a) Number of publications in 15 different domains. We group them into three categories. (b) Publication trend: the numbers of publications, authors and junior authors from 1997 to 2010.

2 Approach

2.1 Research Questions

In this paper, we seek to explore the scientific productivity patterns of junior scholar. Our research questions included: 1) How did junior scholars starting at different time continuously publish papers? 2) What is the domain and generation differences in junior scholars' publication patterns? 3) What is the relationship between collaboration and publication productivity across domains and time?

2.2 Data

We retrieved 5 million publications from Microsoft Scholar Database[1]. These papers are published from 1997 to 2010, included title, author, discipline information. In this data-set, the main proportion disciplines are medicine, economic and computer sciences within 15 different domains (see Figure1(a)). We group these disciplines into three categories: "Med" (short for medical), "STEM" (abbreviation of Science, Technology, Engineering and Math related domains) and "non-STEM" (the remaining disciplines). Figure2 (a) and (b) show the numbers of publications and junior authors for three categories in each year.

2.3 Analysis

Retention Rate We characterize how junior scholars continuously publish academic work by computing the "retention rate" of a set of junior scholar as follows. Let J_y be the set of authors who published their first paper in year y. The retention rate is computed based on whether or not the scholar has at least one paper within the next three or more years. Let J_y be the set of authors in J_y who published continually at least one paper within each year between years $y + 1$ and $y + d$. We define the retention rate R_y as:

$$R_y^{(d)} = \frac{|J_y^{(d)}|}{|J_y|}.$$

[1]http://academic.research.microsoft.com/

(a) Publications (b) Junior scholars

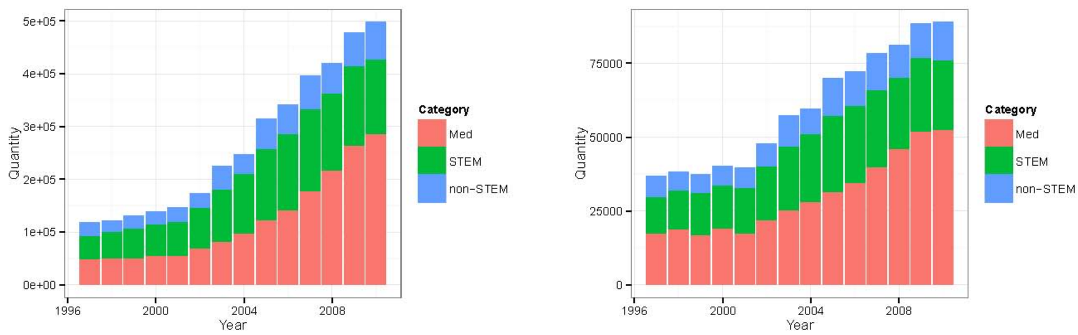

Figure 2: (a) Number of publications within the three categories from year 1997 to 2010; (b) Number of junior authors within the three categories from year 1997 to 2010.

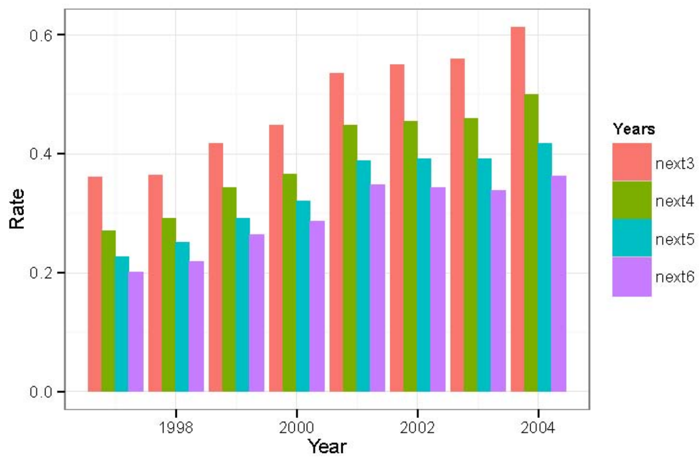

Figure 3: Retention rate for next 3, 4, 5, and 6 for scholars starting at year between 1997 and 2004.

Productivity We characterize scholars' productivity by the number of papers they published within a given period. Because the junior scholars are distinguished by their starting years and domains, it allows us to investigate the relationship between annual productivity and years of continuous publishing across generation and domains.

Collaboration Structure We characterize the collaboration of a scholar by the "collobaration size" – the total number of co-authors within a given period. Then, we qunatify the relationship between scholars' collaboration and their productivity using correlation. Due to the skew distribution in scholars' publications and collaboration sizes, we use Kandall's τ rank correlation.

3 Preliminary Result

3.1 Retention Rate

Figure 3 shows the pattern of juniors who continually publishing papers in academia in next 3, 4, 5, and 6 years. Retention rate decreases naturally because the retained scholars in the next $k + 1$ years is a subset of retained scholars in the next k years (they have kept publishing in an additional year). With the passage of time, there are less and less scholars remain in academia. For junior scholars starting at 2004, 61.37% of them continued to publish in the next 3 years and only 36.18% continued to publish in the next 6 years. This gap increased from around 17% in 1997 to around 25% in 2004.

3.2 Productivity Analysis

For those who survival in academia after a long period, their productivity (average publication of that year) continue to grow in their career. Figure 4(a) shows the average annual productivity of junior scholars starting at different years. The longer who stay in academia will lead to higher productivity. Moreover, the scholars starting at recently years (after 2003) had more publications in their early age. In other words, the group starting at 1998–2003 has less publications than the group starting at 2004–2008 in their same early career time. This suggests there is higher publication demanding for junior scholars lately.

To compare productivity across domains, in Figure 4(b), (c) and (d), we show the annual productivity by three categories: STEM, non-STEM and Med. The pattern of STEM related domains are similar to the overall trend. However, the productivity of scholars who stay more than 5 years in STEM domain, exceed the average productivity of lately scholars with the same seniority years. For example, the junior scholars from 1998 to 2001 with higher productivity after 5 years in their career than the junior scholar start from 2002 to 2004. In Figure 4(c), the annual productivity of non-STEM related domains showed a different pattern. For early career of non-STEM scholars, they are with less publications compared to the STEM domains. Moreover, the scholars who stay longer are not guarantee with higher productivity. The productivity for those scholars stay more than 5 year is dynamic. This might due to the non-STEM domain is not easy to accumulate their previously works for new researches in late career. In Figure 4(d), the pattern of Medical related domains are relatively stable. The early stage of scholars have higher productivity than the recently scholars. The scholars who stay longer will with the higher productivity, but it is not easy to boost their productivity in their early research stage. This domain difference might due to the medical research required a long term training to become mature. Each scholars who can survive in the long term training process could achieve a certain level of outputs.

3.3 Collaboration and Productivity

Prior work has show the importance of collaboration in scholars' productivity (Uzzi et al., 2013). Here we study the relationship of collaboration and productivity across different domains. We use Kandall's τ correlation to quantify the association between scholars' collaboration size and their productivity. Table1 presents the results for junior scholars starting at year 2004. Overall, there is a positive and significant association across all domains. The correlation increases with longer academic seniority. For instance, the correlation coefficients of STEM domains increase from 0.5293 to 0.6221 within next 4 to 8 years from junior year.

For STEM related domains, the correlation coefficient is bigger than the Medicine and non-STEM related domains in all three time period. In other word, the productivity of STEM related domains correlates more by collaboration size than the rest two domains.

Interestingly, the correlation for Med domain is averagely smaller than that in STEM and non-STEM domains initially (within next 4 and 6 years). However, the correlation in Med domains increases rapidly. After 8 years, the order of correlations shift from STEM > non-STEM > Med to STEM > MED > non-STEM. In MED area, the effect of collaboration in productivity is not strong initially, compared to the other two groups. The impact of collaboration appears stronger over time, suggesting that scholars in Med domains benefited more from more co-authors in their later stage.

Period	STEM	nonSTEM	MED
Next4	0.5293**	0.4895**	0.4679**
Next6	0.5803**	0.5450**	0.5416**
Next8	0.6221**	0.5743**	0.5927**

Table 1: Correlation of scholars' annual productivity and their collaboration size within next 4, 6 and 8 years, computed based on Kandall's tau rank correlation. All the correlation coefficients are significant with p-value < 0.01.

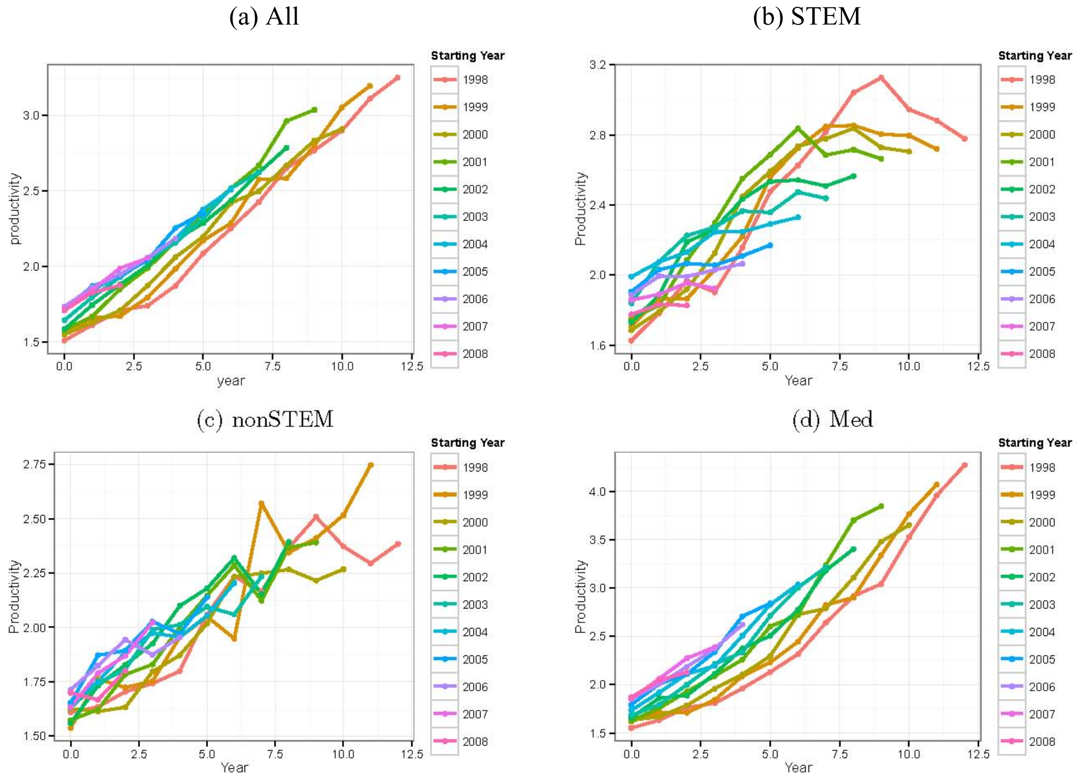

Figure 4: Correlation between productivity and time in academia: (a) All 15 domains data; (b)STEM domains; (c)non-STEM domains; (d)Medical domains. Our research data-set cover from year 1997 to 2010, so the starting year from 1997 will cover 12 hops to 2010. However, the starting year from 2008 will only provide 3 hops to 2010. All curves are aligned by time-period of 0 -12 years.

4 Discussion and Future Works

In this paper, we studied the productivity of junior scholars based on their academic publications. We analyzed junior scholar's retention rate, productivity and the collaboration structure across different domains. The result indicates less than 36.18% scholars will remain in academia after 6 years. Moreover, the scholars who stay longer in academia will gain more productivity in STEM and medicine related domain. Besides, more collaborators generally relate to higher productivity. In particular, the STEM domain can benefit more based the the collaboration with more scholars. However, in non-STEM domains, the larger coauthors size is not guarantee of the higher productivity.

Our finding contribute to the understanding of junior scholars' early career path across different disciplines, which has implication on developing more informative measures for evaluating scholars' success. In our future works, we plan to leverage this preliminary findings to develop predictive models for predicting junior scholars' progress and long-term success.

References

Ioannidis, J. P., Boyack, K. W., & Klavans, R. (2014). Estimates of the continuously publishing core in thescientific workforce. *PloS one*, *9*(7), e101698.Jones, B. F. (2010). Age and great invention. , *92* (1), 1–14. Retrieved 2014-08-22, from http://www .mitpressjournals.org/doi/abs/10.1162/rest.2009.11724

Lungeanu, A., Huang, Y., & Contractor, N. S. (2014). Understanding the assembly of interdisciplinary teamsand its impact on performance. , *8* (1), 59–70. Retrieved 2014-08-19, from http://www.sciencedirect.com/ science/article/pii/S1751157713000837 doi: 10.1016/j.joi.2013.10.006

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. ,*342* (6157), 468–472. Retrieved 2014-08-19, from http://www.sciencemag.org/content/342/6157/468 doi: 10.1126/science.1240474
van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market.*Current Biology*, *24*(11), R516–R517.