CrossMark

Standing on the shoulders of giants: the effect of outstanding scientists on young collaborators' careers

Mengjiao Qi[1] · An Zeng[1] · Menghui Li[2] · Ying Fan[1] ·
Zengru Di[1]

Abstract Collaboration is one of the key features in scientific research. With more and more knowledge accumulated in each discipline, individual researcher can only be an expert in some specific areas. As such, there are more and more multi-author papers nowadays. Many works in scientometrics have been devoted to analyze the structure of collaboration networks. However, how the collaboration impacts an author's future career is much less studied in the literature. In this paper, we provide empirical evidence with American Physical Society data showing that collaboration with outstanding scientists (measured by their total citation) will significantly improve young researchers' career. Interestingly, this effect is strongly nonlinear and subject to a power function with an exponent <1. Our work is also meaningful from practical point of view as it could be applied to identifying the potential young researchers.

Introduction

The past decade has witnessed the rapid development of the Internet technology and electronic media, which significantly accelerates the publication of scientific results and makes the access to these scientific papers much easier. In this context, measuring the the

---

✉ An Zeng
anzeng@bnu.edu.cn

Ying Fan
yfan@bnu.edu.cn

1 School of Systems Science, Beijing Normal University, Beijing 100875, China

2 Beijing Institute of Science and Technology Intelligence, Beijing 100044, China

quality and quantity of science becomes a very challenging problem (Leydesdorff and Milojevic 2015). Accordingly, Scientometrics has attracted attention from researchers with different backgrounds including physics, mathematics, computer science and social science (Chen et al. 2013; Henriksen 2016; Biesenbender and Hornbostel 2016; Ma and Guan 2005). The recent development of the complex network theory also provides effective tools for modeling and analyzing the scientific publication data. For instance, the PageRank has been applied to the citation networks for ranking the significance of papers (Yan 2014).

Scientific collaborations are indispensable for a scientist's academic life, and co-authorship has been increasing rapidly in both the natural sciences and the social sciences (Leydesdorff and Wagner 2009; Adams 2012; Cronin 2001; Wuchty et al. 2007; Ossenblok et al. 2014). Because knowledge is better transferred and combined by collaboration, co-authored papers are found to be cited more frequently (Adams 2006, 2012). Consequently, collaboration between authors has been intensively studied. Based on scientists' collaboration networks, many findings have been made. Examples include basic statistical properties such as the clustering coefficients and degree distributions of these networks (Albert and Barabasi 2002; Dorogovtsev and Mendes 2002; Newman 2001, 2003), the division of community structure (Albert and Barabasi 2002; Newman 2001, 2003), the dynamics of these growing networks (Golosovsky and Solomon 2012; Jeong et al. 2003; Peterson et al. 2010; Medo et al. 2011), and the evolution of policies and empirical networks with time (Zhao and Zhao 2016; Makkonen and Mitze 2016). Most related works aim to investigate the structural properties of these networks. Some other works also focus on the influence of the collaboration on scientists' research outcomes.

It is argued that collaboration and the structure of the collaboration network have some influence on a scholar's research. Some empirical studies have been conducted on this issue (Arajo et al. 2014; Damien et al. 2016; Sooryamoorthy 2014). The number of coauthors is proven to be a strong predictor of research productivity (Lee and Bozeman 2005), but this result does not mean that "more is better" (Levitt and Thelwall 2016). Another study (Kemp 2013) concluded that collaboration and competition can be regarded as an ultimately cyclical process in different stages of one's academic career. The structural position in the network has a significant impact on scientific performance (Ebadi and Schiffauerova 2015; Damien et al. 2016) and the funding the scientist can receive (Ebadi and Schiffauerova 2015). As a very important type of collaboration, doctoral education and the influence of the supervisor have been discussed in the literature (Horta and Santos 2016a; Larivire 2012; Waaijer et al. 2016). Collaboration with a supervisor is shown to be a significant driving factor in junior researchers' publication activity and career development (Pfeiffer et al. 2016; Horta and Santos 2016b; Pinheiro et al. 2012).

Investigating the characteristics of researchers' scientific career is actually an important problem, as it is closely related to a number of practical issues including quantifying the success of scientists (Hirsch 2005; Dorogovtsev and Mendes 2015) and predicting the impact of scientists' papers (Acuna et al. 2012; Revesz 2014). Related results are fruitful. Doctorate holders are found to represent a crucial human resource for research and innovation, and their labor market is more internationalized and can be represented by a network (Auriol 2007, 2010; Auriol et al. 2007). Some researchers pay close attention to the mobility of scientists' career in academia and geography (Solimano 2008; Deville et al. 2014), the influence factors of career choice (Petersen and Penner 2014; Petersen et al. 2012), the age dynamics in scientific creativity (Jones and Weinberg 2011; Lazer et al. 2009), the reputation and impact in scientists' career development (Solimano 2008), as well as the persistence and uncertainty in the academic career (Petersen et al. 2012). Also,

there are some works about predicting a scientist's performance (Vespignani 2009; Clauset et al. 2015).

In this paper, we focus on the influence of outstanding scientists on young collaborators' career development. We use the publication data from American Physical Society (APS) journals and find evidence that a young researcher tends to have a more proficient career research productivity in the future if he/she has collaborated with outstanding scientists in his/her early career. We consider different definitions of outstanding scientists, such as highly cited scientists and Nobel laureates, and similar evidence is identified. Interestingly, we also find that the influence of outstanding scientists on young collaborators' careers is highly nonlinear and subject to a power function with an exponent <1. By studying the evolution of the APS data, we find that the positive effect of outstanding scientists on young collaborators is actually becoming stronger over time. These findings are meaningful for identifying young researchers with high potential.

Method

The database used in this paper is from the American Physical Society (APS) journals in the period from 1893 to 2009. The journals include the Physical Review series, and Reviews of Modern Physics. In total, the data include 458,584 papers from 325,491 authors. For each author, we can obtain his/her name and all the papers he/she has published. For each paper, the information about DOI, authors, publication date and the DOI of citing papers are all available in the database.

We first briefly introduce the definitions and notations in this paper. For each author i, his/her total number of citations and publications by the end of the data (i.e., the year 2009) are respectively denoted as $C(i)$ and $P(i)$, respectively. As we mainly study the effect of outstanding collaborators on young scholars, we must define young scholars in this paper. We define the year when an author published his/her first paper as the beginning of his/her scientific career (Deville et al. 2014; Petersen and Penner 2014; Petersen et al. 2012). In addition, we regard the first three years of his/her scientific career as his/her young scholar period. While a young scholar i may have several collaborators in his/her young scholar period, we will focus on the most outstanding collaborator. To judge whether a collaborator is outstanding or not, we compute the total citations of this collaborator up to year t (i.e., the year young scholar i published his/her first article). We denote the most highly cited collaborator of author i as u_i, and his/her total number of citations and publications are accordingly denoted as $C(u_i)$ and $P(u_i)$.

We used the following two calculation methods in this paper: full counting and fractional counting. Full counting is the most commonly used, but the characteristic of allocating the credit for the article to authors while ignoring the number of authors may be misleading. In the fractional counting system, the total credit for a co-authored article is the same as for a single author article, but all the collaborative authors share equal fractions of the total credit for this paper. It has been presented that fractional counting is preferable over full counting when constructing bibliometric networks in which a small number of nodes having a very large degree (e.g., a publication with many co-authors) (Perianes-Rodriguez et al. 2016). The results of using these two methods in qualitative analysis are consistent. For simplicity, we show the results calculated in the full counting system in our paper, and present the results using fractional counting in the Supplementary Information.

The problem addressed in this paper is how outstanding collaborators influence young scholars' career development, which is straightforwardly illustrated in Fig. 1. Consider two scholars i and j who published the first paper in the same year and published 3 and 2 papers, respectively, in their young scholar period, where the most outstanding collaborator of i is cited 300 times while the most outstanding collaborator of j is cited 1000 times by the beginning of the two scholars' scientific career. Our question is who will be more likely to do better in their future career (i.e., who will tend to be more cited in the future).

Results

We first focus on the APS data from 1940 to 1990, during which 103,701 scholars published their first paper. For each scholar, we focus on their young scholar period (i.e., the first three years of the career) and identify the most highly cited collaborator u_i for each of them. We are mainly interested in the relation between $C(i)$ and $C(u_i)$. Here, $C(u_i)$ is the total citation of the most highly cited collaborator of a young scholar i, which measures whether i has a truly outstanding collaborator. $C(i)$, on the other hand, is the total citations of i, which measures whether i will be highly cited in his/her future career until the year 2009.

The probability density distribution of $C(i)$ and $C(u_i)$ is presented in Fig. 2a, b. In log-log scale coordinates (i.e., the illustrations in Fig. 2a, b), one can see that the distributions all obey the power-law distribution, which is consistent with the existing research conclusions. However, the difference between the two power-law distributions' exponent means that the relationship between $C(i)$ and $C(u_i)$ is not simply linear. We then study the
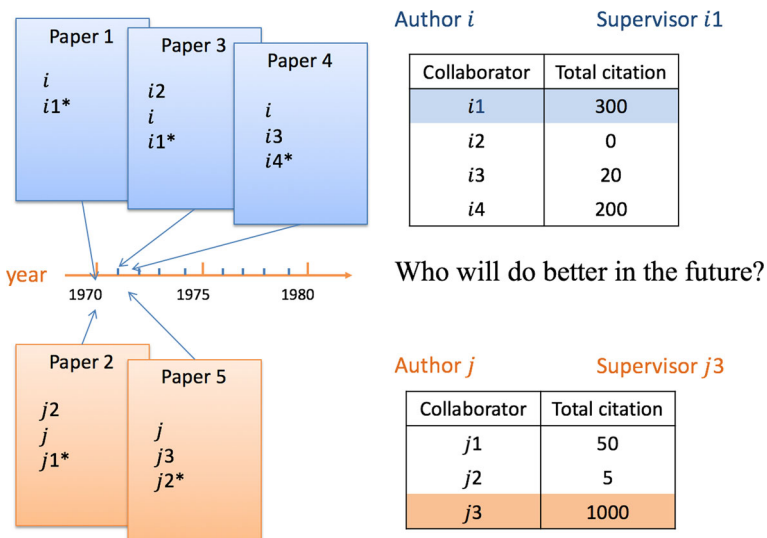
Fig. 1 (Color online) Illustration of the problem studied in this paper. We consider two scholar i and j who both publish their first paper in 1970. In their young scholar period (i.e. 1970–1972), scholar i published 3 papers with 4 collaborators and scholar j published 2 papers with 3 collaborators. The most highly cited collaborators of i and j are respectively cited 300 times and 1000 times until 1970. This paper aims to find out whether there is a tendency for j to outperform i in their future career (measured by their future total citations)
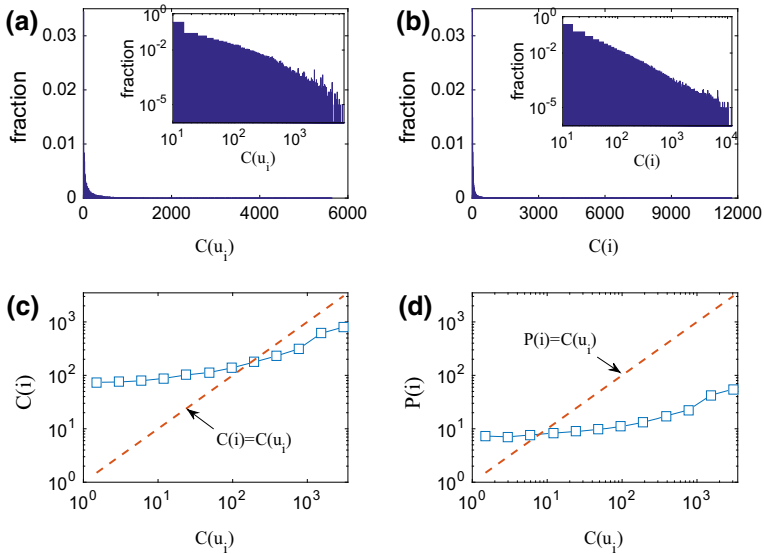
Fig. 2 (Color online) **a**, **b** The probability density distribution of the total citation of young scholar's most highly cited collaborator in his/her young scholar period $C(u_i)$ and the total citation of young scholars $C(i)$. The illustration is in log-log scale coordinates. One can see that they all obey the power-law distribution with different exponent which indicates that the relationship between them is not linear. **c**, **d** The relationship between the total citation of young scholar's most highly cited collaborator in his/her young scholar period $C(u_i)$ and the young scholar's academic performance (i.e. total citation $C(i)$ and total publication $P(i)$). The *red dotted lines* represent the standard line $y = x$. Both of $C(i)$ and $P(i)$ increase with $C(u_i)$, and the relationship can be subject to power function with linear fitting coefficient (i.e. the index of power function) 0.3097 and 0.2574 in log-log scale coordinates

relation between $C(i)$ and $C(u_i)$ in Fig. 2c, d. To explore and better quantify the relationship, We do not directly show the scatter plot but take a nonlinear average of the data, and the results are shown in Fig. 2c, d. For the xth data point, its corresponding $C(i)$ is averaged over the $C(i)$ of all the original data points whose $C(u_i)$ is in the range of $[2^{x-1}, 2^x - 1]$. The $C(u_i)$ of the xth data point is $(2^{x-1} + 2^x - 1)/2$. In Fig. 2c, one can see that $C(i)$ indeed increases with $C(u_i)$, and the relation between $C(i)$ and $C(u_i)$ is highly nonlinear and subject to a power function with an exponent <1. The power relation between $C(i)$ and $C(u_i)$ indicates that selecting a more outstanding collaborator will significantly improve a young scholar's future career. However, this effect is more obvious when the collaborator is an ordinary scientist. If the collaborator is a top cited scientist, selecting a more outstanding collaborator will not produce a significant improvement in the young scholar's future career. In Fig. 2d, we can see the same results in the relation between $P(i)$ and $C(u_i)$. These findings may also be helpful for young scholars when choosing their supervisors. An extremely outstanding scientist should always be the first choice. However, if such scientists are not available, it is also acceptable to select a supervisor who is slightly weaker than the most outstanding. As for the young scholars who must select their supervisors from among ordinary scientists, it is beneficial to find a scientist who is better than average. The position in the co-authors list can reflect an author's contribution to their article (Ponomariov and Boardman 2016), and it is suggested that the research self-efficacy is an important factor when predicting a scholar's academic productivity (Hemmings and Kay 2016; Horta et al. 2016). In advanced analysis of young

scholars collaborating with an outstanding scientist, we find that in both the full counting and the fractional counting systems, the highly cited young scholars have larger fractions on important positions (the first and the last) in the co-authors list than less cited young scholars (see Supplementary Information).

We then confirm the above finding using the data of Nobel prize winners. We identify 18 Nobel laureates in physics fulfilling the following three conditions: (i) they won the Nobel prize between 1950 and 2000; (ii) they published their Nobel prize winning papers in APS journals; and (iii) they collaborated with at least one young scholar in APS journals. For each Nobel laureate i, we then identify the scholars (denoted by set v_i) with whom the Nobel laureate has collaborated during their young scholar period. To compare the performance of the Nobel laureates' young collaborators with other young scholars during that period (denoted by set w_i), we compute the following four quantities for each Nobel laureate: (i) the average total citations of his/her young collaborators by 2009, $\overline{C}(v_i)$; (ii) the average total publications of his/her young collaborators by 2009, $\overline{P}(v_i)$; (iii) the average total citations of other young scholars during the same period, $\overline{C}(w_i)$; and (iv) the average total publications of other young scholars during the same period, $\overline{P}(w_i)$. The results are reported in Table 1. $\overline{C}(v_i)$ and $\overline{P}(v_i)$ are shown to be, respectively, larger than $\overline{C}(w_i)$ and $\overline{P}(w_i)$, indicating that the young scholars who collaborated with the Nobel laureates performed better in their subsequent career than young scholars without Nobel laureates as collaborators. However, this finding needs to be interpreted cautiously. It should be noted that collaboration behavior is a two-way selection process. The Nobel prize laureates also select excellent young scholars as collaborators, just as young scholars

Table 1 The academic performance of 18 Noble prize winners' young collaborators

ID	Winner's name	Award year	$\overline{C}(v_i)$	$\overline{P}(v_i)$	$\overline{C}(w_i)$	$\overline{P}(w_i)$
1	John Bardeen	1956	452.62	24.77	289.35	23.82
2	Tsung-Dao Lee	1957	1658.17	110.33	276.03	22.38
3	Chen-Ning Yang	1957	664.18	18.00	290.20	23.37
4	Owen Chamberlain	1959	105.94	5.56	327.38	26.09
5	Charles H. Townes	1964	184.73	14.55	328.24	26.13
6	Richard P. Feynman	1965	472.67	18.00	263.31	22.29
7	Julian Schwinger	1965	503.00	47.00	309.28	25.65
8	Philip W. Anderson	1977	874.55	37.31	213.93	18.73
9	Steven Weinberg	1979	996.00	23.33	302.51	24.73
10	Sheldon L. Glashow	1979	136.00	10.00	329.66	26.52
11	Arthur L. Schawlow	1981	10.57	3.14	202.60	17.86
12	N.Bloembergen	1981	326.89	13.56	264.17	22.20
13	Kenneth G. Wilson	1982	1244.00	34.50	235.66	20.17
14	William A. Fowler	1983	83.33	9.33	206.65	17.42
15	Klaus von Klitzing	1985	224.59	12.59	168.53	15.34
16	K. Alex Mller	1987	264.40	33.60	143.33	13.71
17	Leon M. Lederman	1988	379.00	38.00	196.41	17.29
18	Jerome I. Friedman	1990	1443.20	95.00	179.91	16.10

choose Nobel prize laureates. So the reason for this finding may include that Nobel prize laureates prefer to choose the best young scholars as collaborators.

We then investigate the evolution of the young scholars' careers after collaboration or lack thereof with the outstanding scientists during their young scholar period or not. To quantify the success of their career, we use both the number of total citations and the number of total publications. We no longer compute the citation number and publication number at the end of the data set, but calculate these two numbers up to year t, where t is the tth year of the scholar's academic career (taking the year when he/she published his/her first paper as the beginning of his/her career). The number of citations and number of publications up to year t for scholar i are denoted as $c_t(i)$ and $p_t(i)$, respectively. In contrast to the above analysis, where the outstanding scientists are defined as the Nobel laureates, here the outstanding scientists in a certain year are defined as the scientists whose total number of citations is within the top 10% among all authors in this year in APS. Thus, in every year, by defining outstanding scientists, we can divide the young scholars who publish their first papers in this year into two types: (i) individuals who collaborated with outstanding scientists during their young scholar period and (ii) individuals who did not collaborate with outstanding scientists. Then, we sort the two types of young scholars separately by their total citations up to 2009 (i.e., $C(i)$). We then compute the mean $c_t(i)$ and $p_t(i)$ of the top 10% young scholars of the first type as $\langle c_t \rangle$ and $\langle p_t \rangle$. For comparison, in the same year, we also compute $\langle c_t \rangle$ and $\langle p_t \rangle$ of the young scholars who are the top 10% in the second type. The results for the year 1990 (i.e., these scholars published their first paper in 1990) are presented in Fig. 3a, b. Although they show top performance in each type, one can see that both $\langle c_t \rangle$ and $\langle p_t \rangle$ of the outstanding scientists' young collaborators are higher than for their counterparts and increase faster during the early stage of their careers. In addition, we randomly select 10 young scholars from the top 50 of the first type and 10 young scholars who eventually achieved similar $\langle c_t \rangle$ and $\langle p_t \rangle$ in the second type. There are actually more than 10 such young scholars fulfilling this condition in the second type. We focus again on the evolution of $\langle c_t \rangle$ and $\langle p_t \rangle$, as shown in Fig. 3c, d. Interestingly, though these two types of young scholars have similar $\langle c_t \rangle$ and $\langle p_t \rangle$ values, the $\langle c_t \rangle$ and $\langle p_t \rangle$ of the young scholars who have outstanding scientists as collaborators increase faster during the early stage of their scientific career, indicating that these young scholars generally achieve a greater research productivity earlier. We see consistent phenomena in many other years (see Supplementary Information).

The above analyses show that there is an obvious positive effect of outstanding collaborators on young scholars' future careers. Finally, we study how strong this effect is in different years. To this end, we make use of the data from 1960 to 1990, within which there are 87,665 young scholars publishing their first papers. We then identify the most highly cited collaborator for each of them. According to the definition above, the citation number and publication number until year 2009 of i are $C(i)$ and $P(i)$, respectively. The number of citations and the number of publications of i's most highly cited collaborator are denoted as $C(u_i)$ and $P(u_i)$. In a specific year t, we denote all the young scholars who published their first paper in this year as year t's young scholars. We extract three groups from these young scholars in year t according to $C(u_i)$ (i.e., the citations of their most highly cited collaborators). The three groups are respectively the young scholars whose $C(u_i)$ belongs to the top 10% (group 1), which means that they collaborated with at least one outstanding scientist; the middle 10% (group 2), which means that they collaborated with ordinary scientist; and the bottom 10% (group 3), which means that they collaborated with scientist with lower performance. We aim to investigate the mean $C(i)$ and $P(i)$ of these three groups of young scholars in different years t. For a fairer comparison in different years, we
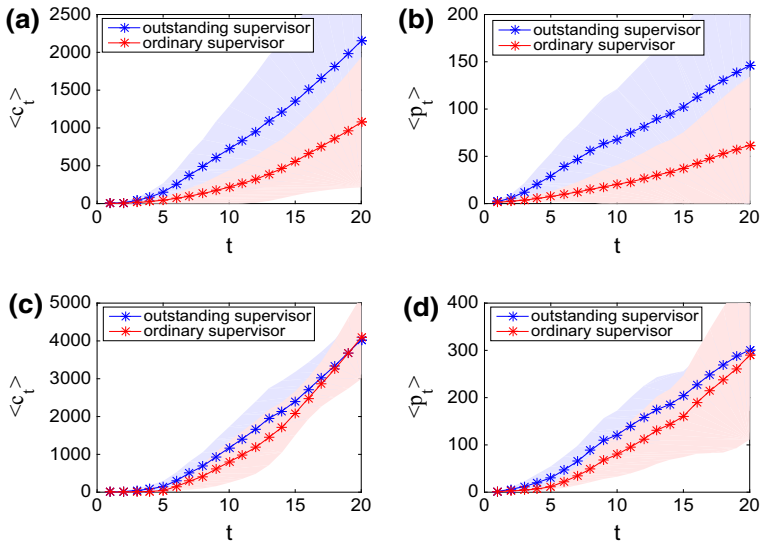
Fig. 3 (Color online) The evolution of the academic performance of young scholars who published his/her first paper in 1990. **a**, **b** The evolution of the averaged total citations and publications of top 10% young scholars for each type. **c**, **d** The evolution of the averaged total citations and publications of 10 young scholars of each type who reached similar academic performance by the end of the data. As $c_t(i)$ and $p_t(i)$ represents respectively the total citation and total publication of scholar i at year t, the averaged total citations and publications are simply obtained by averaging $c_t(i)$ and $p_t(i)$ over the above-mentioned 10% or 10 young scholars. The *shadow* is the standard deviation of the selected young scholars

do not directly compare $C(i)$ and $P(i)$, as the researchers who began their career earlier (i.e., young scholars with smaller t) tends to have higher $C(i)$ and $P(i)$. Therefore, we compute the mean rank of $C(i)$ and $P(i)$ of these three groups among all the young scholars of year t. The resultant mean ranks are denoted as $R_{C(i)}$ and $R_{P(i)}$, respectively. The smaller the mean ranks are, the better the young scholars' academic performance by 2009 is, and the stronger the positive effect is in year t. The results are reported in Fig. 4, where one can see clear trends for these three groups. For group 1, $R_{C(i)}$ decreases with t for group 1, increases slightly with t for group 2, and increases slightly with t for group 3. This trend is consistent when we look at $R_{P(i)}$. As $R_{C(i)}$ and $R_{P(i)}$ are the mean ranks in the year 2009, small values represent excellent scientific performance. The results indicate that the scientific performance of young scholars who collaborated with outstanding scientists in their young scholar period has become greater with time. So the influence of an outstanding scientist grows stronger with time, and collaboration with outstanding scientists is more necessary than before if a young scholar wants to make a difference in his/her field. The number of doctorate recipient has dramatically increased in recent decades (Horta and Santos 2016b), and higher education calls for greater competition and greater accountability for students (Bgelund 2015). Collaboration with supervisor is suggested to be a significant driving factor in junior researchers' publication activity and career development (Horta and Santos 2016a; Waaijer et al. 2016; Pinheiro et al. 2012). In the current academic environment, the supervisor is increasingly important to the young researchers.
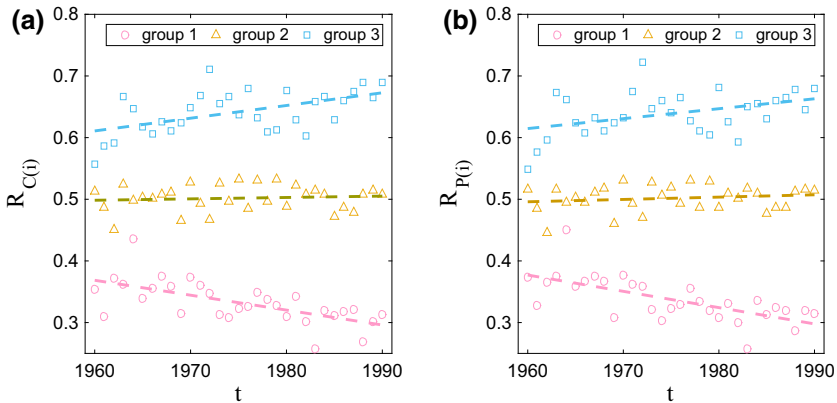
Fig. 4 (Color online) The relationship between year and young scholars' relative rankings in total citation, $R_{C(i)}$ and total publication, $R_{P(i)}$. In each year, the young scholars are divided into three groups. Group 1 are the young scholars who have collaborated with top 10% scientists. Group 2 are the young scholars who mainly collaborated with ordinary scientists. Group 3 are the young scholars who only collaborated with scientists with relatively low performance (bottom 10%)

Discussion

In this paper, we demonstrate the positive influence of outstanding scientists on young collaborators' future career. We find that this effect is strongly nonlinear and subject to a power function with an exponent <1. In addition, we investigate the evolution of the young scholars' citation and publication number after collaboration with outstanding scientists. The results show that the advantage from an outstanding scientist is more obvious in the early stage of a young scholar's career, but gradually becomes weaker as the time goes on. Finally, we study this effect in different years and find that this effect has been strengthening.

In fact, the results revealed in this paper can be easily understood. The young scholars as defined in this paper can be regarded as young students. The collaborators can be interpreted as their supervisors. The results indicate that selecting an outstanding supervisor is very important to one's future academic career. However, due to the nonlinearity observed in this paper, if the supervisor is "super" outstanding, the positive effect will not be "super" as well. This result is probably because when the supervisors' academic ability reaches a certain level, they can provide their young students with almost equal academic help (e.g., ideas, techniques, money, connections), even when their academic achievements are not exactly the same. In this sense, our paper provides some guidance for young students in selecting their supervisors. In addition, we find that this positive effect is more significant in the early stage of the young scholar's career. This positive effect could be attributed by many reasons. For instance, the young scholars learn some advanced techniques from the outstanding collaborators so that they can solve more difficult problems in their future careers. It might also be true that these young scholars' papers are more easily to be accepted due to the good reputation of the outstanding collaborators. We show in the last part of the paper that this positive effect tends to become stronger over time. In the current academic environment, the supervisor is more and more important to the young researchers. As mentioned above, a good supervisor can provide different aspects of support for young students' academic careers, such as new ideas, advanced techniques,

sufficient money, and a wide range of social contacts. In contrast to the past, as more and more young scholars appear in various fields, young scholars who want to stand out must rely more on external help from their supervisors.

Through the findings above, we show that there is a positive influence of outstanding scientists on young collaborators' future careers. In other words, collaborating with an outstanding scientist can make the academic career easier for a young scholar. However, it should be noted that this positive influence is more significant in the early stage of young scholars' academic career. Currently, supervisors are seeking economically viable and efficient research rather than high-quality research for certain practical reasons (Bgelund 2015). If the young scholar continues collaborating with an outstanding scientist, he/she may never leave the shadow of this scientist, never be an independent scholar in his/her own right and never be truly recognized for his/her own accomplishments. This collaboration may then become a negative experiences in the young scholar's academic development (Salabubar and Castell 2017).

This paper may inspire some extensions. For instance, in studying the collaboration network, most previous research has examined topological properties of the whole network, considering that the links in the network to be homogeneous. For further study, we can divide these links into two parts: the links of scientific research cooperation and the links of supervisor guidance, and then analyze their distribution in the network. We can also verify empirically whether it will be easier for a scholar to be cited in top journals if he co-authored with an outstanding scientist, and whether there is an increase in recognition after this co-authorship. There are some local fluctuations in our analysis of the trends over time, so we are very interested in the impact of key events in the field of physics may have on the trends of our findings. In this work, we focus on the statistical patterns rather than on specific cases. We will conduct a case study of some large events in future work. Our empirical analysis is limited to the field of physics because we do not have access to the databases of other disciplines. Our approach is general, so it is possible to expand this study to other fields when we obtain the relevant databases. These questions call for further research in the future.

References

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, *489*(7415), 201–202.

Adams J. (2006). International Partnerships of Research Excellence. Available at http://go.nature.com/zzwn8z.

Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, *490*(7420), 335–336.

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex network. *Reviews of Modern Physics*, *74*(1), 47–97.

Arajo, E. B., Moreira, A. A., Furtado, V., Pequeno, T. H. C., & Andrade, J. S. (2014). Collaboration networks from a large cv database: Dynamics, topology and bonus impact. *PLoS ONE*, *9*(3), e90537.

Auriol, L. (2007). Labour market characteristics and international mobility of doctorate holders: Results for seven countries. *OECD STI Working Papers*, *2*, 37.

Auriol, L. (2010). Careers of doctorate holders: Employment and mobility patterns. *Oecd Science Technology & Industry Working Papers*, *29*(11–05), 15–23.

Auriol, L., Felix, B., & Fernandez-Polcuch, E. (2007). Mapping careers and mobility of doctorate holders: Draft guidelines, model questionnaire and indicators. OECD Science Technology & Industry Working Papers.

Bgelund, P. (2015). How supervisors perceive PhD supervision—And how they practice it. *International Journal of Doctoral Studies*, 10, 39–55.

Biesenbender, S., & Hornbostel, S. (2016). The research core dataset for the German science system: Developing standards for an integrated management of research information. *Scientometrics*, 108(1), 401–412.

Chen, Y., Borner, K., & Fang, S. (2013). Evolving collaboration networks in scientometrics in 1978–2010: A micro-macro analysis. *Scientometrics*, 95(3), 1051–1070.

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, 1(1), e1400005.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52(7), 558–569.

Damien, C., Arnaud, D., Catherine, L., & Mlanie, P. (2016). The impact of a researcher's performance: An empirical analysis. *PLoS ONE*, 11(8), e0161281.

Deville, P., Wang, D., Sinatra, R., Song, C., Blondel, V. D., & Barabai, A. L. (2014). Career on the move: Geography, stratification, and scientific impact. *Scientific Reports*, 4, 4770.

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51(4), 1079–1187.

Dorogovtsev, S. N., & Mendes, J. F. F. (2015). Ranking scientists. *Nature Physics*, 11, 882–883.

Ebadi, A., & Schiffauerova, A. (2015). How to become an important player in scientific collaboration networks? *Journal of Informetrics*, 9(4), 809–825.

Ebadi, A., & Schiffauerova, A. (2015). How to receive more funding for your research? Get connected to the right people!. *PLoS ONE*, 10(7), e0133061.

Golosovsky, M., & Solomon, S. (2012). Stochastic dynamical model of a growing citation network based on a self-exciting point process. *Physical Review Letters*, 109(9), 297–311.

Hemmings, B., & Kay, R. (2016). The relationship between research self-efficacy, research disposition and publication output. *Educational Psychology*, 36(2), 347–361.

Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455–476.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.

Horta, H., & Santos, J. M. (2016a). The impact of publishing during PhD studies on career research publication, visibility, and collaborations. *Research in Higher Education*, 57(1), 28–50.

Horta, H., & Santos, J. M. (2016b). An instrument to measure individuals? Research agenda setting: The multi-dimensional research agendas inventory. *Scientometrics*, 108(3), 1243–1265.

Horta, H., Cattaneo, M., & Meoli, M. (2016). PhD funding as a determinant of PhD and career research performance. *Studies in Higher Education*. doi:10.1080/03075079.2016.1185406.

Jeong, H., Neda, Z., & Barabai, A. L. (2003). Measuring preferential attachment in evolving networks. *EPL*, 61(4), 567–572.

Jones, B. F., & Weinberg, B. A. (2011). Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(47), 18910–18914.

Kemp, A. (2013). Collaboration vs. individualism: What is better for the rising academic? *Qualitative Report*, 18(100), 1–8.

Larivire, V. (2012). On the shoulders of students? The contribution of phd students to the advancement of knowledge. *Scientometrics*, 90(2), 463–481.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Life in the network: The coming age of computational social science. *Science*, 323(5915), 721–723.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science: An International Review of Research in the Social Dimensions of Science and Technology*, 35(5), 673–702.

Levitt, J. M., & Thelwall, M. (2016). Long term productivity and collaboration in information science. *Scientometrics*, 108(3), 1103–1117.

Leydesdorff, L. & Milojevic, S. (2015). Scientometrics. In: J.D. Wright (editors-in-chief), *International Encyclopedia of Social and Behavioral Sciences* (2nd Edition), Section 8.5: Science and Technology Studies, Subsection 85030. Micheal Lynch (ed.), 21, 322–327.

Leydesdorff, L., & Wagner, C. S. (2009). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4), 317–325.

Ma, N., & Guan, J. (2005). An exploratory study on collaboration profiles of chinese publications in molecular biology. *Scientometrics*, 65(3), 343–355.

Makkonen, T., & Mitze, T. (2016). Scientific collaboration between 'old' and 'new' member states: Did joining the european union make a difference? *Scientometrics*, *106*(3), 1193–1215.

Medo, M., Cimini, G., & Gualdi, S. (2011). Temporal effects in the growth of networks. *Physical Review Letters*, *107*(23), 238701.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404–409.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

Ossenblok, T. L. B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities. *Journal of the Association for Information Science & Technology*, *65*(5), 882–897.

Perianes-Rodriguez, A., Waltman, L., & Eck, N. J. V. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195.

Petersen, A. M., & Penner, O. (2014). Inequality and cumulative advantage in science careers: A case study of high-impact journals. *EPJ Data Science*, *3*(1), 24.

Peterson, G. J., Pressé, S., & Dill, K. A. (2010). Nonuniversal power law scaling in the probability distribution of scientific citations. *Proceedings of the National Academy of Sciences of the United States of America,107*(37), 16023–16027.

Petersen, A. M., Riccaboni, M., Stanley, H. E., & Pammolli, F. (2012). Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5213–5218.

Pfeiffer, M., Fischer, M. R., & Bauer, D. (2016). Publication activities of german junior researchers in academic medicine: Which factors impact impact factors? *BMC Medical Education*, *16*(1), 190.

Pinheiro, D., Melkers, J., & Youtie, J. (2012). Learning to play the game: Student publishing as an indicator of future scholarly success. *Technological Forecasting & Social Change*, *81*(1), 56–66.

Ponomariov, B., & Boardman, C. (2016). What is co-authorship? *Scientometrics, 109*(3), 1939–1963.

Revesz, P. Z. (2014). A method for predicting citations to the scientific publications of individual researchers. In: International Database Engineering & Applications Symposium (pp. 9–18). ACM.

Salabubar, A., & Castell, M. (2017). Exploring the relationship between doctoral students' experiences and research community positioning. *Studies in Continuing Education, 39*(1), 16–34.

Solimano, A. (2008). *The international mobility of talent.* Oxford: Oxford University Press.

Sooryamoorthy, R. (2014). Publication productivity and collaboration of researchers in South Africa: New empirical evidence. *Scientometrics*, *98*(1), 531–545.

Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science*, *325*(5939), 425–428.

Waaijer, C. J. F., Benot, M., Sugimoto, C. R., & Vincent, L. (2016). Stability and longevity in the publication careers of U.S. doctorate recipients. *PLoS ONE*, *11*(4), e0154741.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Yan, E. (2014). Topic-based PageRank: Toward a topic-level scientific evaluation. *Scientometrics*, *100*(2), 407–437.

Zhao, Y., & Zhao, R. (2016). An evolutionary analysis of collaboration networks in scientometrics. *Scientometrics*, *107*(2), 759–772.