

Standing on the Shoulders of Giants

Tehmina Amjad^{1,2}, Ying Ding^{2,6,8}, Jian Xu³, Chenwei Zhang², Ali Daud^{7,1}, Jie Tang⁴, Min Song^{5*}

¹Department of Computer Science and Software Engineering, Islamabad, 44000, Pakistan

²School of Library and Information Science, Indiana University, Bloomington, 47405, USA

³School of Information Management, Sun Yat-sen University, Guangzhou, 510275, China

⁴Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China,

⁵Department of Library and Information Science, Yonsei University, Seoul, 120-140, Korea

⁶School of Information Management, Wuhan University, Wuhan, China

⁷Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

⁸Library, Tongji University, China

[*min.song@yonsei.ac.kr](mailto:min.song@yonsei.ac.kr)

ABSTRACT

Young scholars in academia always seek to work in collaboration with top researchers in their field in pursuit of a successful career. While success in academia can be defined differently, everyone agrees that training with a well-known researcher can help lead to an efficacious career. This study aims to investigate whether standing on the shoulders of giants does, in fact, improve junior scholars' chances of success. If not, what makes young scientists soar in their academic careers? We investigate this question by analyzing the effect of collaboration with a known-star on success of a young scholar. The results were surprising: working with leading experts can lead to a successful career, but it is not the only way. Researchers who were not fortunate enough to start their career with an elite researcher could still succeed through hard work and passion. These findings emerged from analyses of two discrete sets of well-known scholars on the career of new comers, suggesting their strength and validity.

1. Introduction

Success is not the same for everybody. Individuals must delineate what success means to them and to define one's own values, powers, abilities, aspirations, goals, and enticements. For successful career in academia, young scholars want to study at world's top universities. It is generally believed that studying with the outstanding professors at Ivy League universities will guarantee a successful start, which naturally leads to a successful career. In other words, young scholars can benefit from standing on the shoulders of giants who can elevate them to soaring career heights.

As an extension of scholarly tailgating, Adegbola (2013) defines "Newton's premise of standing on the shoulders of giants" as the process where a scholar makes a quantum jump in their career by collaborating with prominent experts in the healthcare domain. Adegbola's notion of scholarly tailgating embraces both horizontal and vertical relationships among scholars. Because these relationships are often mutual and developmental, participating scholars in the collaborative network can propel their careers to previously unknown heights (Quatman, Chelladurai, & others, 2008). Collaborating with international researchers allows scholars to actuate their scholastic activities (Kram & Isabella, 1985; Sweitzer, 2009). This, in turn, enables them to benefit from the profound knowledge and experience of respected scholars in a domain. Those scholars become rising stars by standing on the shoulders of giants, which forcefully moves them from inertia to dynamic and explosive scholarly heights (Adegbola, 2010).

2. Related Work

A review of the literature regarding collaboration behavior reveals that collaboration emphasizes significant impact upon academic success. Relationships and collaborations are an equivalently rightful constituent of doctoral education and formulation of a professoriate and academic career. Baker and Pifer (2011) have provided a theoretical framework for studying connection between collaborations and learning. The framework focuses on socio-cultural impact of learning and developmental networks. The collaborations and associations that came into existence when a new researcher prepares for academic practices, are considered very significant in the learning and identity development process of these young scholars (Baker & Pifer, 2011). Relationships at the initial stage of a career can be between advisors and advisees, between young scholar and other senior researchers, or among peers. Such collaborations influence the identity development process during the phase of transition from a doctoral student to an independent researcher. Students who

are able to build new contacts have the opportunity to work on new projects and have more co-authorship prospects. Highly productive researchers not only have more connections themselves, but also connect other researchers in a network (Ebadi & Schiffauerova, 2015). With application of multiple regression models, Ebadi and Schiffauerova (2015) established that previous achievements of the scholars and their acquired funding plays an important role in establishing and enhancing their position in co-authorship network. It was determined that young researchers who possess intermediate positions within a collaboration network tend to get connected with scholars from different communities and turn out to be the reason of information diffusion across the network (Ebadi & Schiffauerova, 2015). Weidman, Twale, and Stein (2001) showed that the connections of researchers from within and outside of the academic community are indispensable for resoluteness and professional success.

Quantitative measures, like the number of publications and citations, have been used in literature to assess the accomplishments and standing of researchers. Collaboration of young scholars with respected senior researchers can determine their position within a network. Number of citations is highly influenced by the centrality of scientists in a network and is indicative of the visibility of their work in future (Sarigöl, Pfitzner, Scholtes, Garas, & Schweitzer, 2014). The work of scholars is influenced by the work of their collaborators, more so if the collaborators are senior (Amjad, Daud, Akram, & Muhammed, 2016; Amjad, Daud, Che, & Akram, 2015; X.-L. Li, Foo, Tew, & Ng, 2009). The quantitative methods that were developed with the standard PageRank algorithm (Page, Brin, Motwani, & Winograd, 1999) used different weighting functions by incorporating various factors such as the number of publications, impact of journals where they publish, h-index and the influence of collaborating researchers into the weighting function. Furthermore, amount of funding received and the prestige of the PhD supervisor plays a significant role in predicting the future impact of a young researcher (Acuna, Allesina, & Kording, 2012). The teamwork is not anymore restricted among the humans but, the collaboration of machines and humans can bring forward innovative experiences for assembling and reviewing related research work and receive alerts on their activities and interactions (Ding & Stirling, 2016).

The dynamics of high-impact scientific work can foresee the success of a researcher early in their career, considering some open challenges (L. Li & Tong, 2015). These include topic of research, rank of venue, and author's rank. Success in academia is predictable and can be quantified using

several features, such as the number of published papers, the journal's impact factor, number of highly cited papers, the gender of researcher, and ranking of the institute (van Dijk, Manor, & Carey, 2014). Extrication of quality from quantity of scientific output is essential for the academic tasks such as hiring and promotion, and funding decisions mostly depend on a range of influence measures that are biased by factors such as specialty and academic age (Kaur, Ferrara, Menczer, Flammini, & Radicchi, 2015). The analysis and study of patterns of the progress of young researchers can give a clue about the rising stars, who have potential to flourish very quickly (Daud, Abbasi, & Muhammad, 2013; Daud, Ahmad, Malik, & Che, 2015; X.-L. Li et al., 2009). Dong, Johnson, and Chawla (2015) evaluated that the researcher's authorization on topic of publication and its venue of publication are crucial factors that can increase the h-index of primary author. However, recognition of the topic among community and the co-author's h-indices unexpectedly have less influence on impact of author and academic success.

Estimation of potentially influential literature is also of great significance for choosing important research references. It is beneficial to spot influential literature to craft improved utilization of "giant shoulders" as compared to traditional way to referencing papers (Yan, Huang, Tang, Zhang, & Li, 2012).

From the study of related literature, it was revealed that researchers have studied the impact of collaboration, positioning of a scholar within a scholarly network and their influence of their collaborators. Several quantitative measures have also been studied to find impact of an author as well as to predict impact of authors in future (Acuna et al., 2012; Daud et al., 2015; Dong, Johnson, & Chawla, 2016; Sarigöl et al., 2014; Sayyadi & Getoor, 2009). Our methodology is different from the existing methods, since we do not apply the quantitative measures or algorithms to calculate or predict scholarly impact, rather we analyze the relationship between academic success of young scholars and their collaboration with a well-known researcher. The goal of the present study is to investigate, whether standing on the shoulders of giants leads to success for junior scholars. If not, what makes young scholars soar in their academic careers? We study the correlation between success and co-authorship with a well-known researcher of a field. We are to verify the general assumption that prevails in societies that standing on the shoulders of a giant can lead to success.

3. Method

To this end, we compiled the dataset from AMiner that covers most important conferences and journals from the domain of computer science (Tang, Zhang, et al., 2008). AMiner is considered as a widely used and one of the best curated databases for computer science articles (Dong et al., 2015, 2016; Gollapalli, Mitra, & Giles, 2011; Moreira, Calado, & Martins, 2011; Tang, Jin, & Zhang, 2008; Zhang, Tang, Liu, & Li, 2008). The citation data is extracted from different scholarly sources including DBLP and ACM. The dataset is disambiguated (Arif, Ali, & Asger, 2014; Chen, Guo, Lan, Cao, & Cheng, 2014; Ferreira, Gonçalves, & Laender, 2012; Shoaib, Daud, & Khiyal, 2015) and includes 2,092,356 papers, 1,712,433 researchers, and 8,024,869 citation relationships ranging from 1936–2014. The coauthor network constructed with these data has 1,712,433 vertices (authors) and 4,258,615 edges (collaboration relationships). This dataset allows us to study the interaction, scientific collaboration, and influence among scientists. The analysis is conducted on the scholars whose first publication was published between year 2000 and 2004. However, as the papers cite both old and new articles, to calculate the number of citations we need a larger dataset.

The results are interesting, partially in favor of the hypothesis and unfolding the hidden realities against the hypothesis. Collaboration with a well-known researcher helps a lot to obtain success, but it is not the only possible way. Many individuals are innately gifted with different abilities, talents, and qualities required to be successful, but it does not ensure that they would always become successful. Their own effort towards success can help them catch the attention of a senior researcher. Our results show that 70% to 75% of successful scholars are those who started career without a top-class researcher; their abilities and struggle made them successful and they later collaborated with top-class researchers. From analysis of the results, it was found that if junior researchers, in early part of their careers, have a chance to benefit from the experience of a senior researcher, they can become a rising star very soon. However, researchers who were not fortunate enough to work with a well-respected scientist of their field in the early part of their careers can still gain the attention of senior researchers and flourish later in their careers. To affirm the validity of our results, we performed the analysis on collaborators of two discrete sets of senior researchers. In this study we call them the Authority Authors (AAs). We further divided AAs into two sub-categories, elite *authority authors* and *extended authority authors*. The criteria of selection of elite authority authors from the dataset were higher than the criteria of selection of extended authority

authors. Hence, there were fewer scholars in elite group than in extended group. As a result, more people had a chance to collaborate with an extended authority author than with elite authority authors. Our results are valid for both datasets of authority authors and their collaborators.

Before we proceed with the details of the analysis conducted, in this section we introduce the indicators and filters that we applied on the data for digging up hidden truths in data.

3.1 Indicators

Analyses conducted in this study were based on several parameters used as indicators. Here we introduce all these indicators.

- *Publications*: Publications is the number of research articles an author has published. This number normally represents the productivity of an author.
- *Citations*: Citation count is the number of citations received by the work of an author. The number of citations is an indicator of the impact of an author.
- *H-index*: h-index is a measure of productivity as well as citations impact of an author. It is based on the distribution of citations received by publications of a given scholar (Hirsch, 2005). For this study, h-index is calculated based on the ACM dataset; any citations that are not included in this dataset are not calculated. Because ACM does not collect all the papers and citations of AAs, the h-index that is shown in our dataset is lower than their actual h-index.
- *Sociability*: Sociability is the measure of collaboration of an author. It refers to how many coauthors an author has. We explain the calculation of sociability by a simple example given in Table 1. Here we calculate sociability of author A1 in three years separately and at the end of three years. Suppose in year 1, A1 has three papers: paper 1 with authors a2 and a3, paper 2 with authors a2 and a4, and paper 3 with authors a5 and a6. So in year 1, A1's sociability is 5. To find his total sociability at the end of three years we count all distinct coauthors of A1.

Table 1. Calculation of Sociability for author a1.

	Year 1	Year 2	Year 3	Total at end of three years
A1	P1(a2, a3) P2(a2, a4) P3(a5,a6)	P1(a2, a4) P2(a3, a5) P3(a6, a7)	P1(a4, a5, a7) P2(a4, a8)	a2, a3, a4, a5, a6, a7, a8
Sociability	5	6	4	7

In this study we have retrieved two subsets for the purpose of analysis and comparison. First subset comprises the publications from 2000 to 2004 and second contains publications from 2005 to 2014. Sociability from 2000-2004 represents that how much social or interactive was an author in the start of his or her career.

- *First publication, Last publication, and Longevity:* The first two indicators are the year of first publication of an author and the year of last publication of an author in the given dataset. These two indicators give us the total life span of an author. We call the career length of an author as longevity of an author. We calculated longevity by subtracting the year of first publication from year of last publication.
- *Publications from 2000-2004 and 2005-2014:* This indicator is the publication counts of the authors in the Identification Part (IP) and the Verification Part (VP) of the dataset (details are in section 4). These indicators give us a measure that how productive an author was at the beginning of his or her career and how the productivity increased or decreased in the second subset spanning from 2005 to 2014.

3.2 Filters

Along with the indicators described above we have applied some filters on our dataset to test our hypothesis. With the help of these filters, we are able to distinguish the authors who belong to different categories. These filters are as follows:

- *Senior researcher filter:* We consider the authors who have many publications and have a high h-index as senior researchers. In this study we call them the Authority Authors (AAs).
- *Author collaboration:* We filtered the authors to find the collaborators of an author. We intended to identify researchers who have collaborated with an AA. The collaboration of a new researcher with a well-known author can be an indicator that this young author could be a rising star.
- *Junior researcher filter:* We considered junior authors to be authors who have published some papers only in recent years. Specifically, we considered the authors to be junior authors if they had their first publication between 2000 and 2004.

4. Results and Discussions

In this section we discuss the analysis process that we performed on the selected dataset to identify patterns that can lead a researcher towards success.

First of all we applied the *senior researcher filter* to identify “hot” scientists. We name them the Authority Authors (AAs) and select them based on the h-index measure. For the purpose of experiments, we consider authors having h-index greater than or equal to 40 as “Elite AAs.” Out of 1,712,433 authors, we came across only 35 authors who fulfill the criteria for being an elite AA. Due to this low prevalence of elite AAs, finding a chance to work in collaboration with an elite AA is not possible for all scholars. To find more realistic essence of impact of collaboration, we extracted another group of AAs with different criteria. We selected the authors who have h-index between 25 and 39 as “Extended AAs.” As a result we came up with an additional 422 AAs. Naturally, there will be a larger number of collaborators for 422 authors than 35 authors. After identifying the AAs, both elite and extended, we calculated all the coauthors of the AAs from the whole dataset. In this way, we divided the dataset into two parts for each type of AA, authors who worked in collaboration with an AA and authors who didn't:

- Group A: authors who worked with an authority author (elite or extended) at least once
- Group B: authors who never worked with an any authority author (elite or extended)

We further divided the dataset into two parts based on the time duration, the Identifying Part (IP) (2000-2004) and the Verifying Part (VP) (2005-2014). We named first phase as IP because we intended to identify junior researchers in this time slot. Because a PhD scholar takes around 5 years to graduate, we selected a period of 5 years for the IP. In the VP, we wanted to analyze the progress of researchers identified in the IP to see whether they had become rising stars or not. Authors are assumed to be junior researchers if they had their first publication within the time range of the IP. For analysis, we analyzed different features, such as the number of publications, sociability, the number of citations, longevity, and h-index. We investigated these measures to find out whether rising stars are present among the candidates who were identified in the IP. This can be assessed by looking at their productivity and received citations in the VP.

Fig 1 gives an overview of the conducted analysis for coauthors of elite and extended AAs. Table 2 gives the statistics of the dataset after identifying AAs and junior scholars. There are 21,590

authors in our dataset who collaborated with an elite AA present in the dataset. Out of these, 1,291 are the scholars who had their first publication during the IP. We identified them as Candidate Rising Stars (CRS). There were 8,141 authors who had collaborated with an Extended AA and had their first publication within the IP. On the other hand, there is a long list of authors who never worked with an AA. Among them, 158,389 authors that had their first publication within the IP. Hence, we have 158,389 CRS in Group B. The number of researchers in group B is much more than the number of researchers in group A. For the purpose of comparison, we need to pick a random sample from group B. For being fair, with the help of a random number generating function we extracted three samples from group B. After some analysis, we came to know that researchers who appear in all these groups show similar statistics, so instead of showing results of all three groups with group A, we selected only one of these groups as group B and we used it for the comparison with group A. For the purpose of analysis, we divided Group A further into two sub-groups, Group A1 and Group A2. Group A1 includes scholars who had their first publication with an AA, while Group A2 worked with an AA later but not on their first publication. From Group A2, we further need to find whether these scholars collaborated with an AA in the start of the career or in a later part, so we divided it into two more sub-groups, Group A2-1 and Group A2-2. Group A2-1 contains scholars who worked with an AA in the early part of their careers, and Group A2-2 includes the scholars who collaborated with an AA in later part of their career.

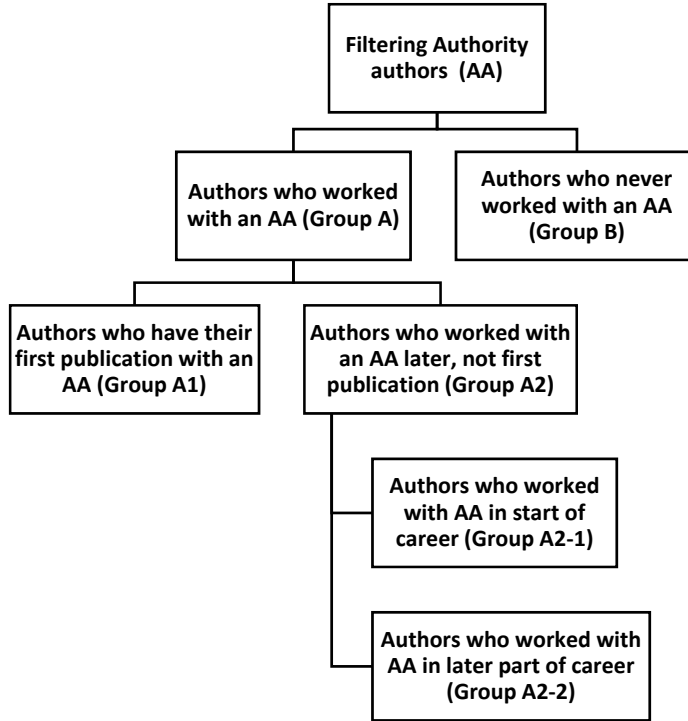


Fig 1. Overview diagram of analysis process performed for two types of AA, Elite and Extended

Table 2. Statistics of Dataset showing Group A and B w.r.t IP and VP

	Statistics
Total number of papers	2,092,356
Year range Part 1: Identifying part (IP)	2000-2004
Year range Part 2: Verifying part (VP)	2005-2014
Number of Elite authority authors (AA)	35
Number of Extended authority authors (AA)	422
Number of CRS in Group A Elite in IP	1,291
Number of CRS in Group A Extended in IP	8,141
Number of authors in Group B (authors who never worked with AA in whole dataset)	1,706,231
Number of CRS in Group B in IP	158,389
Randomly selected in Group B for comparison with Group A Elite	1,291
Randomly selected in Group B for comparison with Group A Extended	8,141
Number of CRS in Group A1 Elite	613
Number of CRS in Group A1 Extended	4,591
Number of CRS in Group A2 Elite	678
Number of CRS in Group A2 Extended	3,549
Number of CRS in Group A2-1 Elite	197
Number of CRS in Group A2-1 Extended	906
Number of CRS in Group A2-2 Elite	481
Number of CRS in Group A2-2 Extended	2,643

Fig 2 illustrates an example of typical authors from different groups (e.g., Group B, Group A1, and Group A2). They have similar h-index (13, 14, and 14). Each line corresponds to an article.

Blue lines represent papers with no AA, while green bars mean the paper author has collaborated with an AA in that specific year. It is obvious from Fig 2(c) that the productivity and citation of author Chang-shui Zhang has increased since he began to collaborate with the authority author.

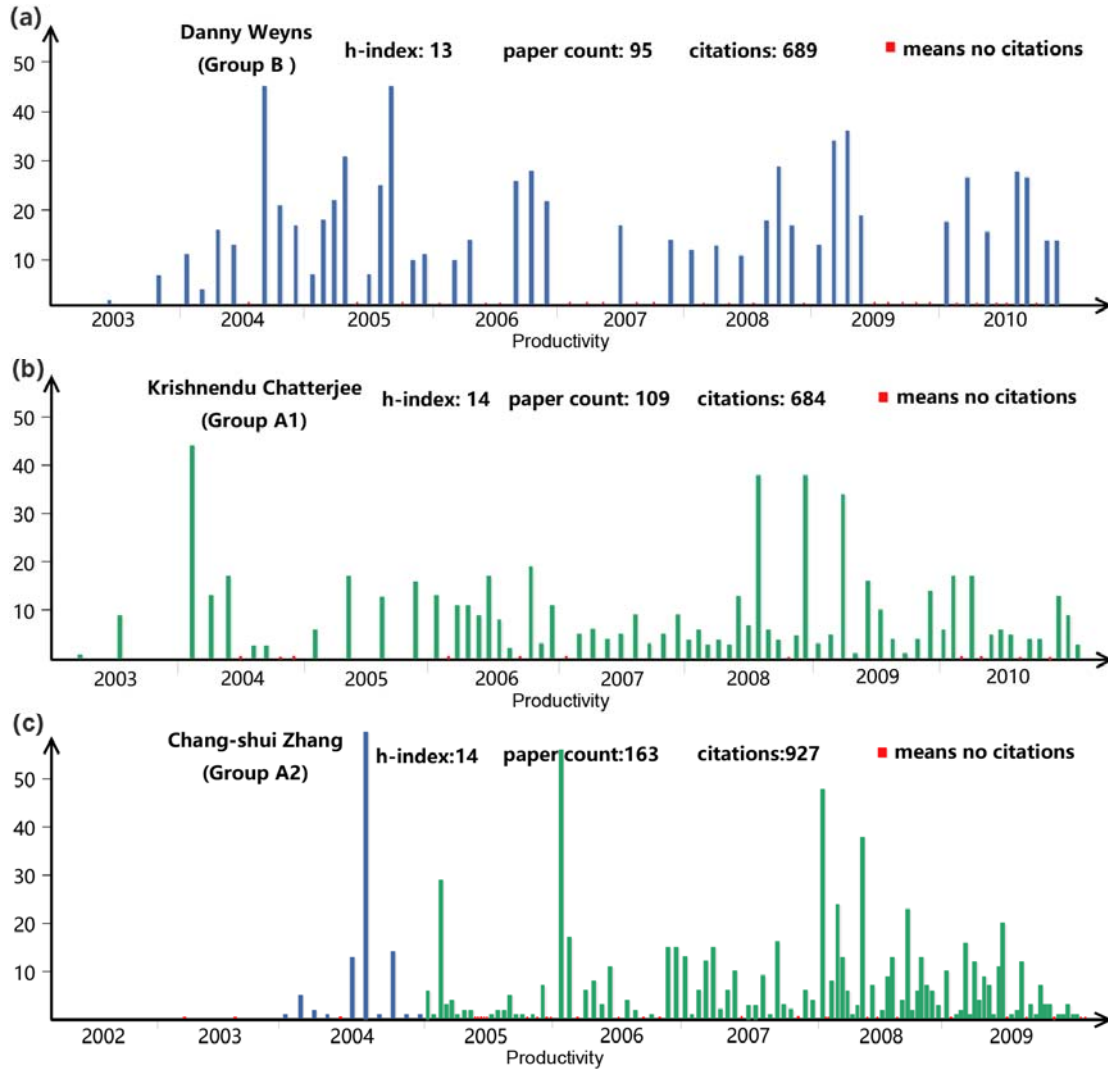


Fig 2. Comparison of productivity picking typical authors from Group B, A1 and A2(a) author selected from Group B; (b) author selected from Group A1 of elite AA; (c) author selected from Group A2 of elite AA.

4.1 Group A vs. Group B.

The results for Group A and B are shown in Table 3. We observed that authors from Group A who worked with an elite AA in the IP are more productive in the VP. They have more publications, they are more social, and they collaborated with more people in the VP. These authors also have longer career life spans. Table 3 supports our hypothesis showing that people in Group A are, on

average, more successful than people in Group B. There are 27% people in Group A who have career lengths greater than 10 years, while in Group B there are only 5% authors with comparable career longevity. This reflects that people who never work with an AA tend to perish quickly. Fig 3 shows the statistics of both groups in terms of percentage. We can see that CRS from Group A are more successful authors in terms of productivity, citation, and sociability.

Table 3: Statistics for comparison of Group A of elite AA and Group B

Candidates	Group A	Percentage	Group B	Percentage	Difference in %
Number	1,291		1,291		
Candidates having...					
Productivity					
10 < Publications < 20	240	18.6%	79	6.1%	12.50
Publications >20	416	32.2%	25	1.9%	30.30
5 < IP Publications < 10	181	14.0%	46	3.6%	10.40
IP Publications >10	50	3.9%	4	0.3%	3.60
10 < VP Publications < 20	148	11.5%	139	10.7%	0.80
20 < VP Publications < 30	78	8.2%	25	1.9%	6.30
VP Publications>30	127	9.8%	24	1.9%	7.90
Impact					
20 < citations < 40	101	7.8%	92	7.1%	0.70
citation >40	1,008	78.1%	93	7.2%	70.90
5 < h-index < 10	399	30.9%	44	3.4%	27.50
10 < h-index < 15	179	13.9%	4	0.3%	13.60
h-index>15	72	5.6%	0	0%	5.60
Sociability					
5 < IP Sociability< 10	416	32.2%	203	15.7%	16.50
10 < IP Sociability< 15	171	13.2%	38	2.9%	10.30
IP Sociability>15	297	23.0%	23	1.8%	21.20
5 < VP Sociability < 10	115	8.9%	133	10.3%	1.40
10 < VP Sociability < 15	110	8.5%	57	4.4%	4.10
VP Sociability >15	658	50.9%	99	7.7%	43.20
Career					
Longevity >10	354	27.2%	67	5.2%	22.0

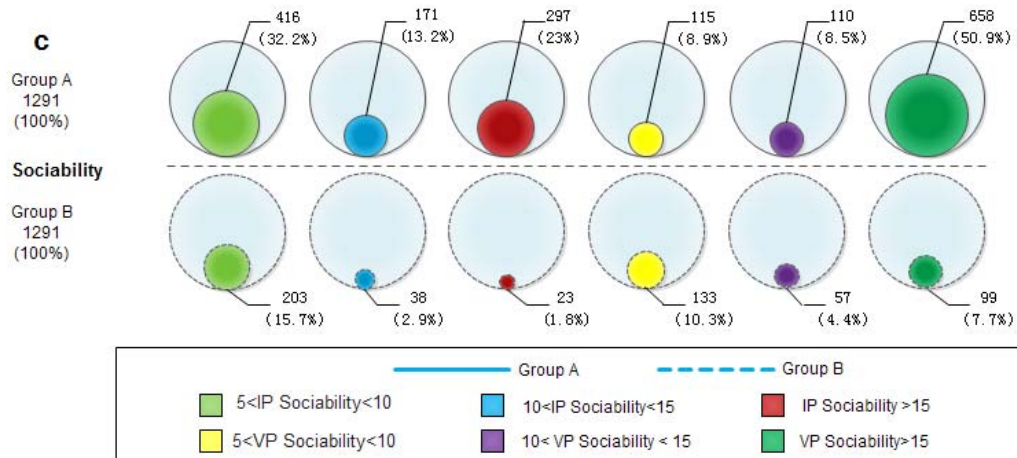
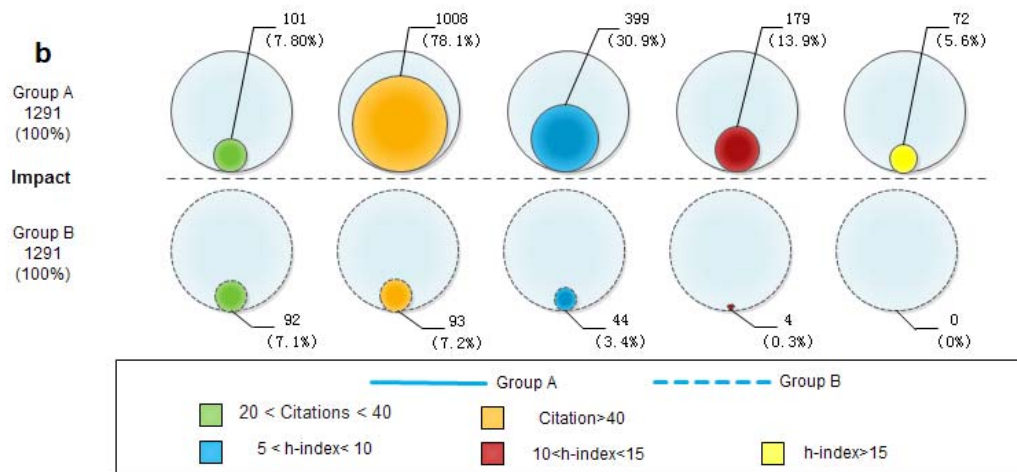
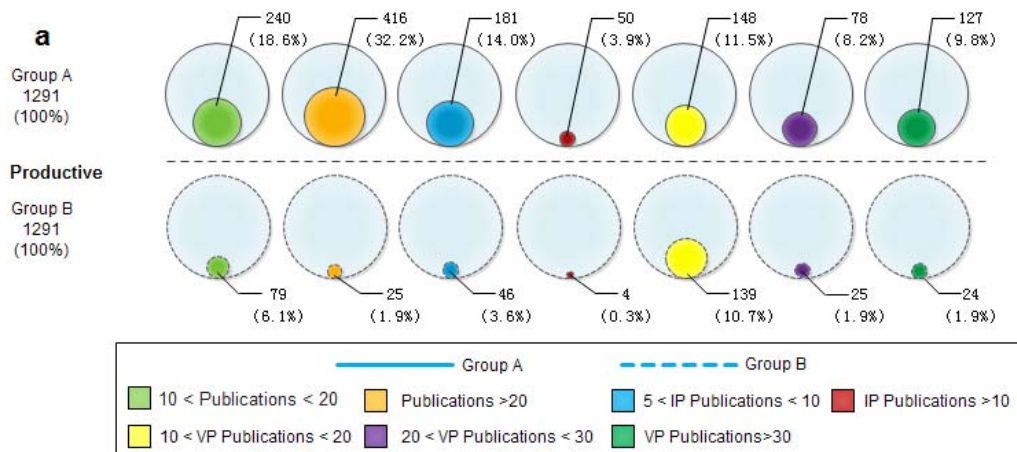


Fig 3. The comparison of a) Productivity, b) Impact and c) Sociability of Group A of elite AA and Group B

4.2 Group A1 vs. Group A2

While results of experimentation so far support the hypothesis that collaborating with top researchers can propel a young scholar’s career, we need to identify more patterns that exist in the career path of these successful authors in Group A. For this purpose, we divided the authors in Group A into two sub groups based on whether they published their first article in collaboration with an AA or they started independently and later on collaborated with an AA.

- Group A1: The authors who started their career with an AA. This can be the case when a young doctoral scholar worked under the supervision of an elite AA.
- Group A2: The authors who started the career independently, but later had a chance to work with an AA. This collaboration can be in the VP or the IP.

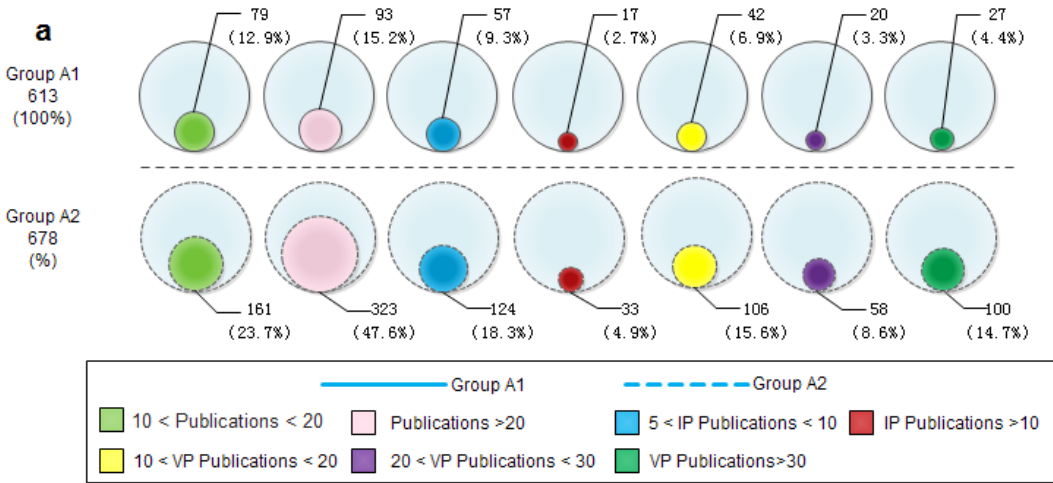
This division helped us determine how much an author’s progress is influenced by whether they started their career in collaboration with an elite AA. Table 4 shows the comparison of Group A1 and A2 of elite AA. We can see surprising results in Group A2. They have greater productivity and impact than the authors of Group A1. This shows that authors managed to produce a few publications without an AA early in their careers and later on began working with an elite AA were more productive than any other group. The authors in Group A1 are those who were part of a university associated with an elite AA and were lucky enough to start their work with them. However, the people of Group A2 started independently, under supervision of a non-AA supervisor, struggled hard, and eventually get connected with an elite AA and excelled. We noticed that sociability of both groups are close to each other, however, sociability is not a strong indicator independently. We also saw interesting differences between the two groups in career length. We calculated career length from a time span of 15 years (2000 to 2014). Approximately 42% of authors in Group A2 have a career length more than 10 years, showing that a large number of authors in this group are non-disruptive and are capable of producing publications over a longer time span, whereas in Group A1 only 11% of researchers have a longer time span.

Table 4: Statistics for comparison of Groups A1 and A2 of Elite AA

Candidates	Group A1	Percentage	Group A2	Percentage	Difference in %
Number	613	47.5%	678	52.5%	5.0
Candidates having...					
Productivity					
10 < Publications < 20	79	12.9%	161	23.7%	10.80

Publications >20	93	15.2%	323	47.6%	32.40
5 < IP Publications < 10	57	9.3%	124	18.3%	9.00
IP Publications>10	17	2.7%	33	4.9%	2.20
10 < VP Publications< 20	42	6.9%	106	15.6%	8.70
20 < VP Publications< 30	20	3.3%	58	8.6%	5.30
VP Publications>30	27	4.4%	100	14.7%	10.30
Impact					
20 < citations < 40	60	9.8%	41	6.0%	-3.80
citation >40	414	67.5%	594	87.6%	20.10
5 < h-index < 10	125	20.4%	274	40.4%	20.00
10 < h-index < 15	52	8.5%	127	18.7%	10.20
h-index>15	17	2.7%	55	8.1%	5.40
Sociability					
5 < IP Sociability< 10	180	29.4%	183	27.0%	-2.40
10 < IP Sociability< 15	69	11.3%	102	15.0%	3.70
IP Sociability>15	156	25.4%	141	20.8%	-4.60
5 < VP Sociability < 10	54	8.8%	61	9.0%	0.15
10 < VP Sociability < 15	49	7.9%	61	9.0%	1.10
VP Sociability >15	169	27.6%	489	72.1%	44.50
Career					
Longevity >10	70	11.4 %	284	41.9 %	30.5

Fig 4 shows the values of indicators (productivity, impact, and sociability) in percentages. We can see that in Group A2 a greater percentage of people are productive and have a high impact, however, the sociability of two groups are close to each other.



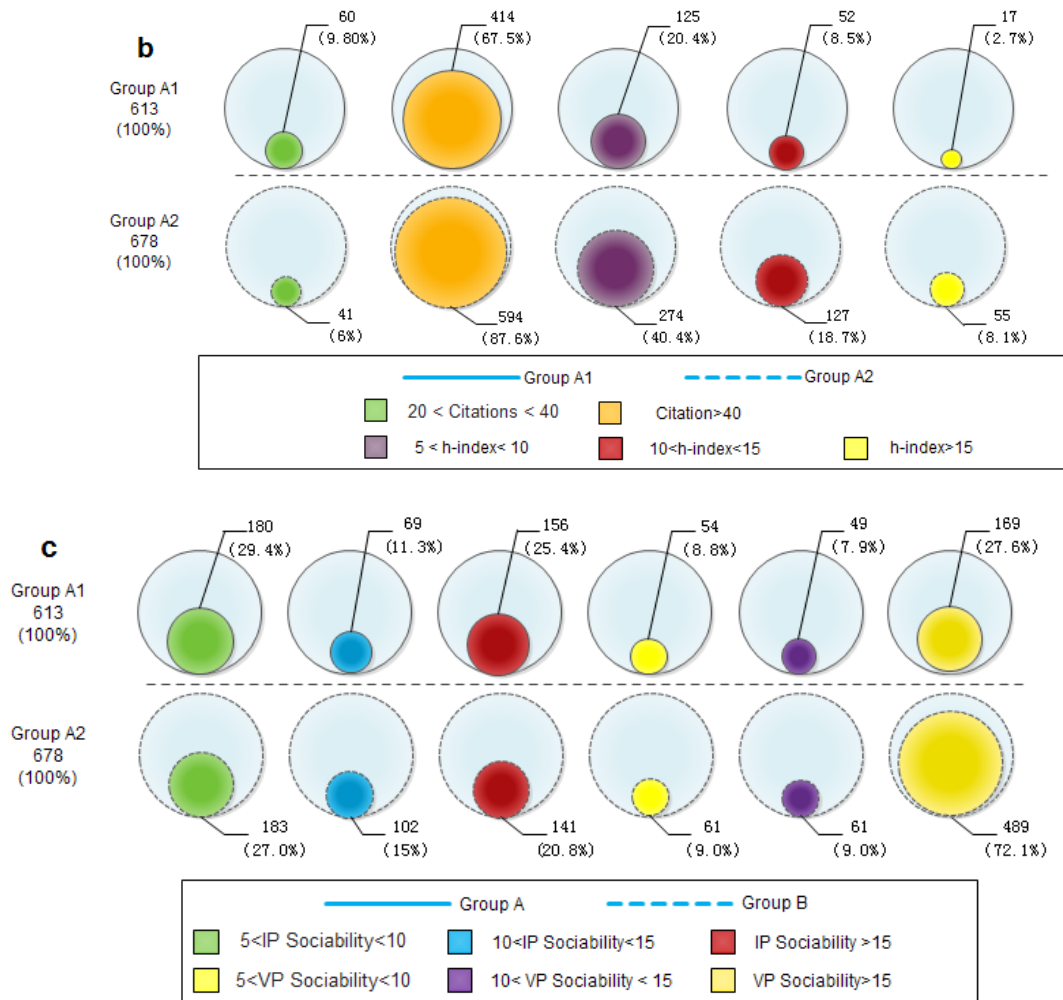


Fig 4: The comparison of a) Productivity, b) Impact and c) Sociability of Groups A1 and A2 of Elite AA

In Fig 5, we observed that the density center of Group A of elite AAs is higher and more towards the right side, which means authors from Group A are more likely to have higher productivity and citations than authors from Group B. Within Group A, we can see that there are two density centers for Group A1 and one for Group A2 (Fig 6). In general, the density center of Group A2 is slightly higher than the second density center of Group A1, and obviously more up-right than the second density center of Group A1, which means authors in Group A2 have higher probability of producing more and higher impact articles.

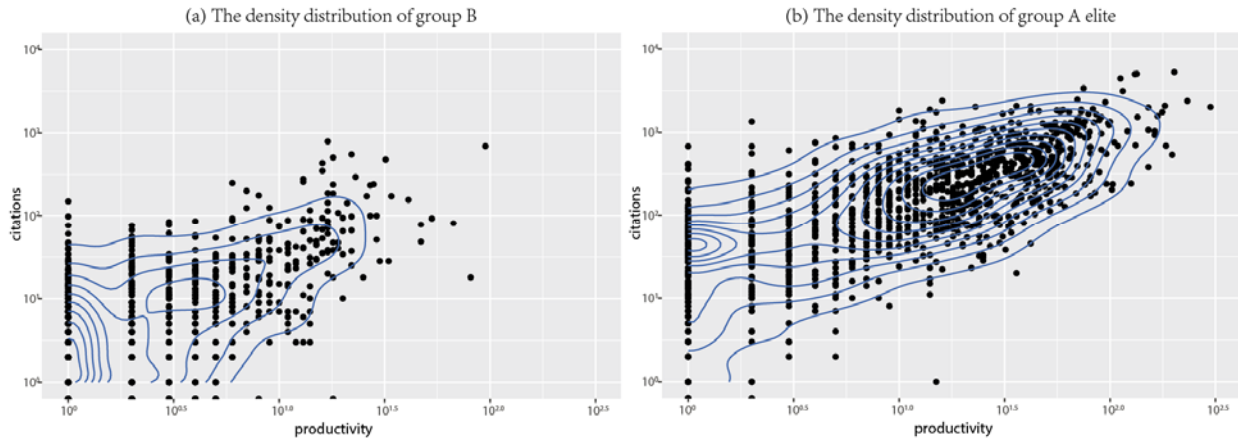


Fig 5. Density distribution of Group B and Group A of elite AA

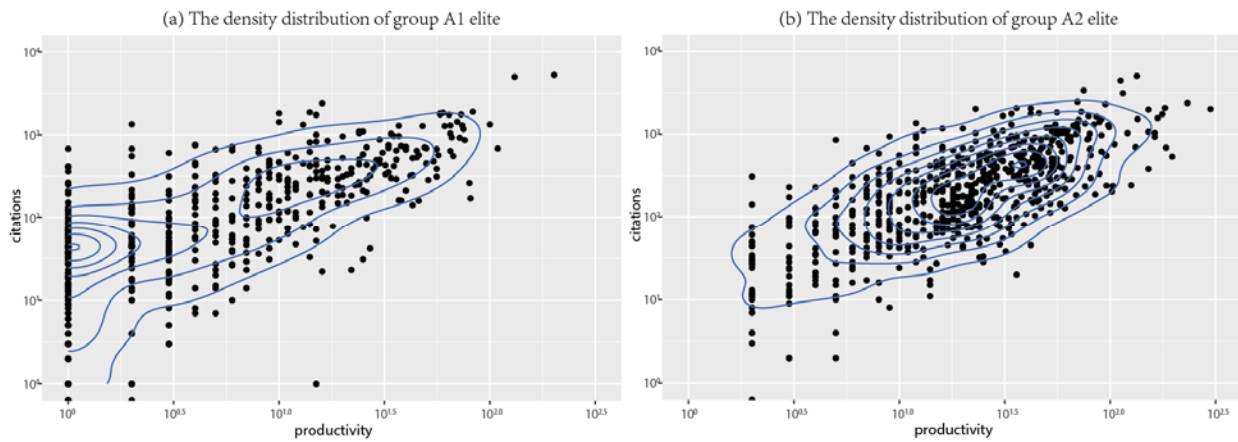


Fig 6. (a) Density distribution of Group A1 of Elite AA; **(b)** Density distribution of Group A2 of Elite AA

Our results indicate that collaboration with an elite AA has the power to advance a new researcher’s career. Researchers who work independently in the initial phase of their career and later on collaborate with an AA have the greatest chance of success, according to our analysis. However, it’s likely that an AA will only be willing to collaborate with them if they have shown some initial progress.

Authors categorized as elite AAs in our dataset are “cream of the crop” of scholars. Their h-index values calculated from publications and citations within the dataset is greater than or equal to 40. Extended AAs are also highly productive researchers, though they are not as productive as elite AAs, with an h-index between 25 and 39. We tested whether the effect of collaboration with an extended AA follows the same path as is observed for collaboration with an elite AA. In reality, we believe that every young researcher does not have an equal chance to collaborate with an elite

researcher. Does that mean that only a few can become a rising star, while others do not have this opportunity? By lowering the criteria, now we have a greater number of AAs, and thus more researchers who had a chance to collaborate with an AA.

4.3 Group A vs. Group B (Extended)

To analyze the authors who collaborated with an extended AA, we applied the same filters we used for our analyses of scholars who collaborated with elite AAs. These findings are shown in Table 5 and Fig 7, and they are highly congruent with the findings of Table 3 and Fig 3. Once again we can conclude that authors who worked with an AA in the IP were more productive in the VP. The results show that, like collaboration with an elite AA, collaboration with an extended AA can lead to a successful career. The percentage columns of Table 3 and Table 5 show that difference between performance statistics of Groups A and B for elite AA and extended AA are in agreement. Fig 3, Fig 7, Table 3, and Table 5 strongly support our hypothesis showing that people in Group A are more successful than people in Group B. Fig 7 illustrates that Group A, authors who collaborated with extended AA, are successful.

Table 5: Statistics for comparison of Group A of extended AA and Group B

Candidates	Group A	Percentage	Group B	Percentage	Difference in %
Number	8,141		8,141		
Candidates having...					
Productivity					
10 < Publications < 20	1,506	18.5%	509	6.2%	12.30
Publications >20	2,058	25.3%	276	3.4%	21.90
5 < IP Publications < 10	1,278	15.7%	221	2.7%	13.00
IP Publications>10	362	4.4%	37	0.5%	3.90
10 < VP Publications< 20	1,290	15.8%	394	4.8%	11.00
20 < VP Publications< 30	712	8.7%	112	1.4%	7.30
VP Publications>30	1,029	12.6%	112	1.4%	11.20
Impact					
20 < citations < 40	967	11.9%	544	6.7%	5.20
citation >40	5,070	62.3%	766	9.4%	52.90
5 < h-index < 10	2,320	28.5%	301	3.7%	24.80
10 < h-index < 15	748	9.2%	36	0.4%	8.80
h-index>15	153	1.9%	7	0.08%	1.82
Sociability					
5 < IP Sociability< 10	2,424	29.8%	1,336	16.4%	13.40
10 < IP Sociability< 15	926	11.4 %	279	3.4%	8.00
IP Sociability>15	872	10.7%	132	1.6%	9.10
5 < VP Sociability < 10	881	10.8%	664	8.2%	2.60
10 < VP Sociability < 15	699	8.6%	321	3.9%	4.70
VP Sociability >15	3,372	41.4%	682	8.4%	33.00
Career					

Longevity >10	1,849	22.7%	496	6.1%	16.6
---------------	-------	-------	-----	------	------

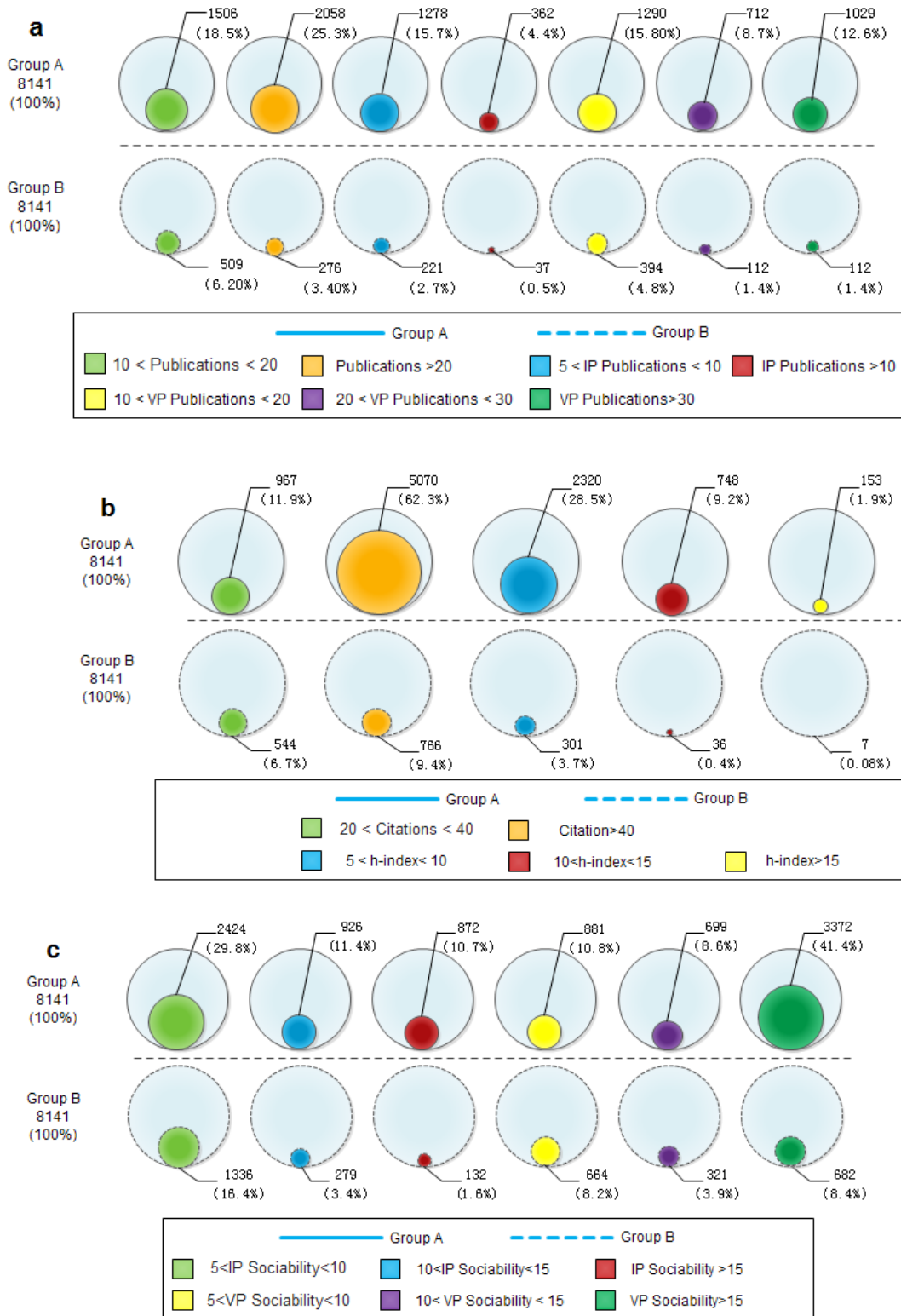


Fig 7: The comparison of a) Productivity, b) Impact and c) Sociability of Group A of extended AA and Group B

4.4 Group A1 vs. Group A2 (Extended)

To determine whether similar patterns exist within Group A of extended AAs as were seen in Group A of elite AAs, we divided the Group A into A1 and A2. The results are shown in Table 6 and Fig 8. As was the case for the Group A1 and A2 of elite AAs, the authors for Group A2 have a lead over the authors in Group A1. In Groups A1 and A2 for extended AAs, the productivity statistics are close, still, the Group A2 shows lead over A1. The percentage columns of Table 4 and Table 6 are not in total agreement. In case of elite AA, Group A2 show higher percentages for productivity, impact and sociability. However, in case of extended AA, productivity percentage of Group A2 is considerably higher but Group A1 are closer to group A2 in case of impact and sociability. If we specifically look at sociability of these groups, we came to know that sociability has increased in the VP as compared to the IP.

Table 6: Statistics for comparison of Groups A1 and A2 of extended AA

Candidates	Group A1	Percentage	Group A2	Percentage	Difference in %
Number	4591	56.4%	3549	43.6 %	12.8
Candidates having...					
Productivity					
10 < Publications < 20	554	12.1%	952	26.8%	14.70
Publications >20	643	14.0%	1,415	39.9%	25.90
5 < IP Publications < 10	529	11.5%	749	21.1%	9.60
IP Publications>10	149	3.2%	213	6.0%	2.80
10 < VP Publications< 20	434	9.5%	856	24.1%	14.60
20 < VP Publications< 30	224	4.9%	488	13.8%	8.90
VP Publications>30	296	6.4%	733	20.6%	14.20
Impact					
20 < citations < 40	577	12.6%	390	11.0%	1.60
citation >40	3,266	71.1%	2,703	76.2%	5.10
5 < h-index < 10	846	18.4%	1,474	41.5%	23.10
10 < h-index < 15	297	6.5%	451	12.7%	6.20
h-index>15	73	1.6%	80	2.3%	0.70
Sociability					
5 < IP Sociability< 10	1,371	29.9%	1,052	29.6%	0.30
10 < IP Sociability< 15	477	10.4%	449	12.7%	2.30
IP Sociability>15	519	11.3%	353	9.9%	1.40
5 < VP Sociability < 10	485	10.6%	396	11.2%	0.60
10 < VP Sociability < 15	285	6.2%	414	11.7%	5.50
VP Sociability >15	1,079	23.5%	2,293	64.6%	41.10
Career					
Longevity >10	574	12.5%	1,275	35.9%	23.4

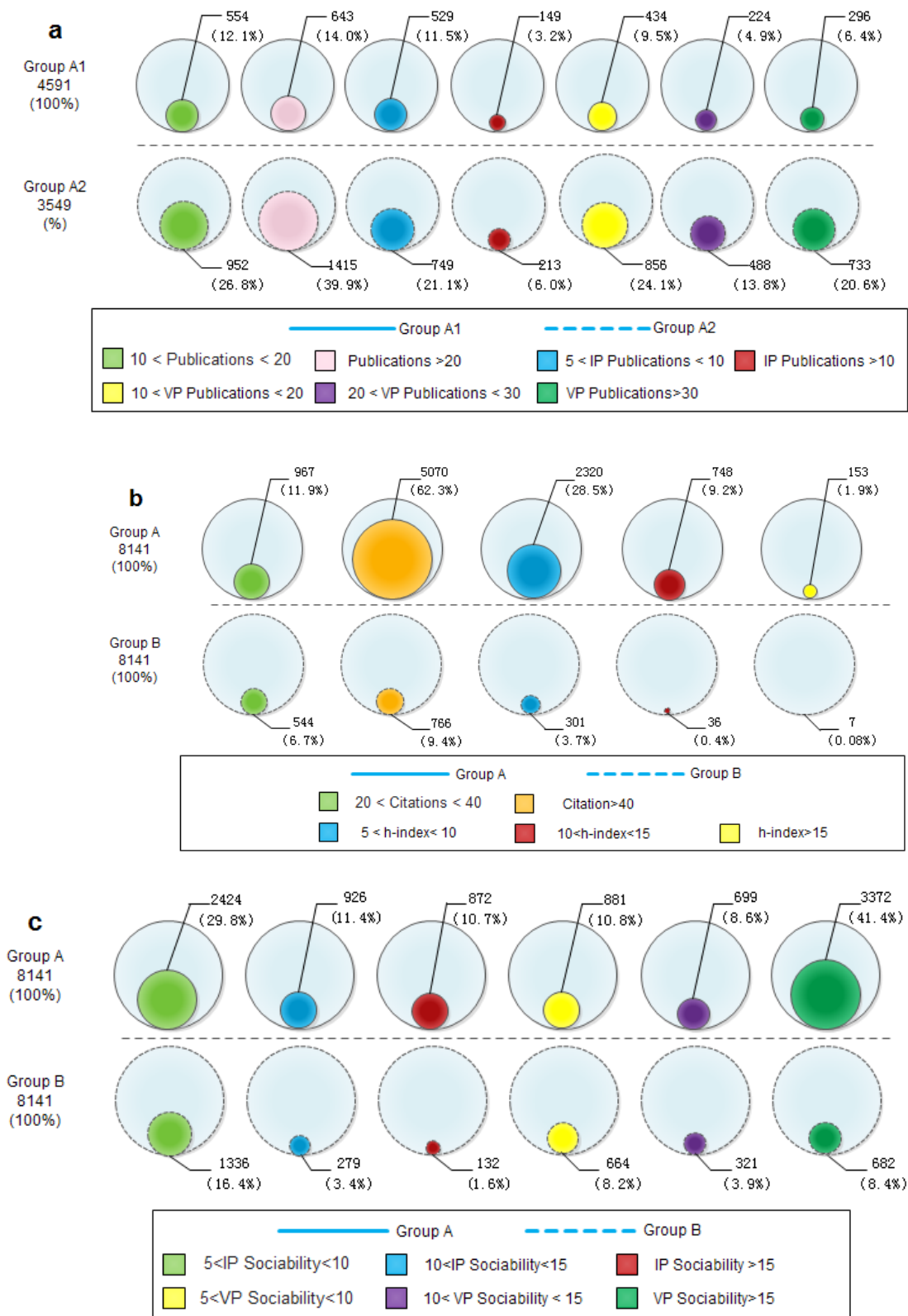


Fig 8: The comparison of a) Productivity, b) Impact and c) Sociability of Groups A1 and A2 of extended AA

Fig 9 shows a similar distribution pattern as Fig 6, which means the output performance was influenced by not only the chance of collaboration with AAs, but also the number of collaborations. The difference of h-index of AAs had limited impact on the performance of rising stars.

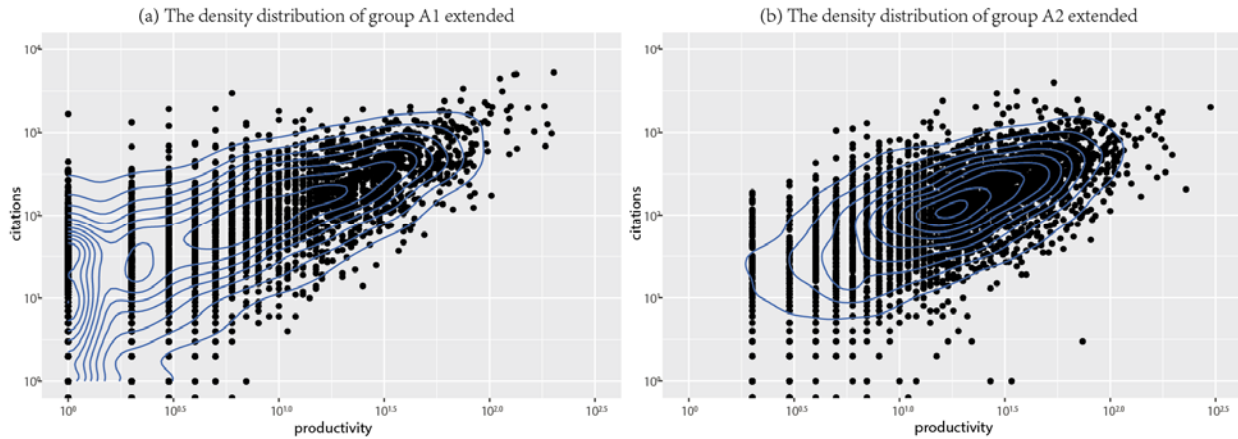


Fig 9. (a) Density distribution of Group A1 of extended AA; (b) Density distribution of Group A2 of extended AA.

Our analyses show that people who started independently, and later on collaborated with an AA were the most successful researchers. These results are surely surprising for people who believe that going to the world’s prestigious universities to work with top professors is necessary for a successful future. Our results show that even if a researcher does not collaborate with an AA at the very start, they may get a chance at any later stage of their career where they can catch the attention of an AA. The relationship between collaborations with AAs and future career success of young scholars thus needs more exploration.

4.5 Group A2-1 vs. Group A2-2

As mentioned in previous section, we divided the dataset into two phases: the IP and the VP. Normally, a PhD scholar takes 4 to 6 years to graduate. We assume that the researchers are somewhere in phase of their PhD training during the IP, while in the VP they have graduated and are excelling in their research fields. Authors of Group A1 collaborate with an AA on their very first publication, while authors of Group A2 do not. To analyze Group A2 in more depth, as they have not started working with an AA and are still more successful than people in A1, we further divided Group A2 into two parts. Group A2-1 consists of authors who collaborated with an AA

during the IP, while Group A2-2 includes the authors who collaborated with an AA for the first time during the VP.

For elite AAs, Group A2-1 consists of 197 authors (29.1%) and A2-2 includes 481 authors (70.9%). Thus, more successful researchers collaborated with an AA later in their career. Findings from the extended group are similar: Group A2-1 consists of 906 authors (25.5%) and A2-2 includes 2,643 authors (74.5%). Thus, for both, elite and extended AA groups, the majority of CRS have worked with an AA in the VP. This shows that they initially worked hard to prove themselves and earned the opportunity to work with an AA. The percentage of researchers in Group A2-2 for both elite AAs and extend AAs is higher than Group A2-1, but the percentages of productivity, impact, sociability, and longevity are similar between Group A2-1 and Group A2-2. This shows that scholars of both groups show similar performance, only there are more scholars who caught the attention of an AA in later part of career. Hence, it does not matter whether a scholar collaborates with an elite AA or an extended AA; it is the collaboration with any AA that matters, especially at the late stage of career (during the VP). These statistics are shown in Table 7, Table 8, and Fig 10.

Table 7: Statistics for comparison of Groups A2-1 & A2-2 of elite AA

Candidates (Elite AA)	Group A2-1	Percentage	Group A2-2	Percentage	Difference in %
Number	197	29.1%	481	70.9%	41.80
Candidates having...					
Productivity					
10 < Publications < 20	42	21.3%	119	24.7%	3.40
Publications >20	91	46.2%	232	48.2%	2.00
5 < IP Publications < 10	74	37.6%	92	19.1%	18.50
IP Publications>10	41	20.8%	30	6.2%	14.60
10 < VP Publications< 20	33	16.6%	115	23.9%	7.30
20 < VP Publications< 30	23	11.7%	68	14.1%	2.40
VP Publications>30	51	25.9%	141	29.3%	3.40
Impact					
20 < citations < 40	8	4.1%	33	6.9%	2.80
citation >40	179	90.9%	415	86.3%	4.60
5 < h-index < 10	70	35.5%	204	42.4%	6.90
10 < h-index < 15	44	22.3%	83	17.3%	5.00
h-index>15	26	13.2%	29	6.0%	7.20
Sociability					
5 < IP Sociability< 10	50	25.4%	133	27.7%	2.30
10 < IP Sociability< 15	46	23.3%	56	11.6%	11.70
IP Sociability>15	80	40.6%	61	12.7%	27.90
5 < VP Sociability < 10	21	10.7%	40	8.3%	2.40
10 < VP Sociability < 15	21	10.7%	40	8.3%	2.40

VP Sociability >15	115	58.4%	374	77.8%	19.40
Career					
Longevity >10	98	49.7%	186	38.7 %	11.00

Table 8: Statistics for comparison of Groups A2-1 and A2-2 of extended AA

Candidates (Extended AA)	Group A2-1	Percentage	Group A2-2	Percentage	Difference in %
Number	906	25.5%	2,643	74.5%	49.00
Candidates having...					
Productivity					
Publications >10 and <20	220	24.3%	732	27.7%	3.40
Publications >20	334	36.9%	1081	40.9%	4.00
IP Publications >5 and < 10	333	36.8%	416	15.7%	21.10
IP Publications>10	120	13.2%	93	3.5%	9.70
VP Publications>10 and < 20	145	16.0%	711	26.9%	10.90
VP Publications>20 and <30	78	8.6%	410	15.5%	6.90
VP Publications>30	187	20.6%	546	20.7%	0.10
Impact					
Having citations >20 and <40	77	8.5%	313	11.8%	3.30
Having citation >40	739	81.6%	1,964	74.3%	7.30
h-index >5 and < 10	333	36.8%	1,141	43.2%	6.40
h-index>10 and <15	150	16.6%	301	11.4%	5.20
h-index>15	32	3.5%	48	1.8%	1.70
Sociability					
5 < IP Sociability < 10	322	35.5%	730	27.6%	7.90
10 < IP Sociability < 15	213	23.5%	236	8.9%	14.60
IP Sociability >15	221	24.4%	132	4.9%	19.50
5 < VP Sociability < 10	292	32.2%	104	3.9%	28.30
10 < VP Sociability < 15	104	11.5%	310	11.7%	0.20
VP Sociability >15	448	49.4%	1,845	69.8%	20.40
Career					
Longevity >10	374	41.3%	901	34.1%	7.20

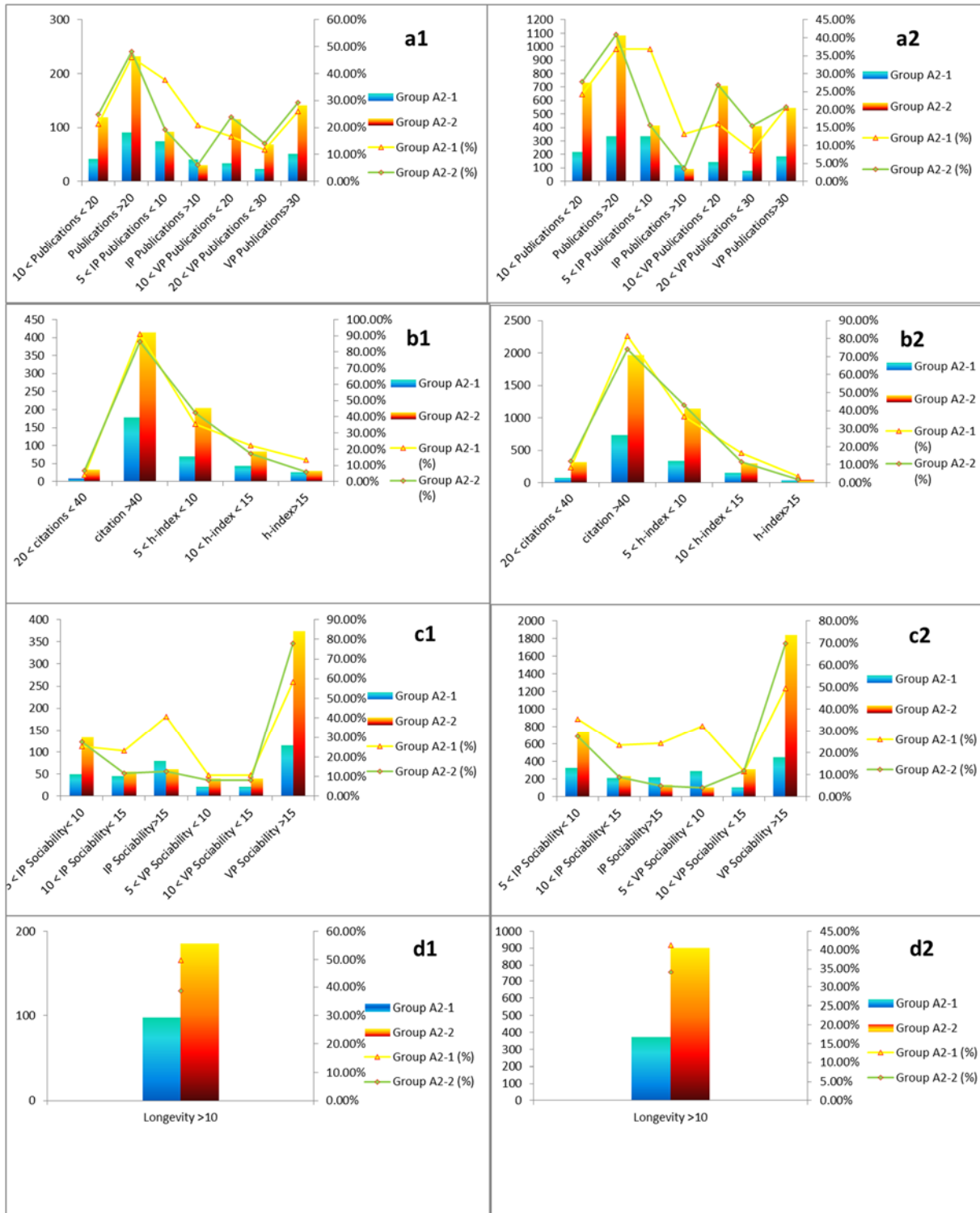


Fig 10. The comparison of a1, a2) Productivity, b1, b2) Impact c1, c2) Sociability and d1, d2) Career of Group A2-1 and A2-2 for elite and extended AA respectively.

5. Conclusion and Future Work

This study investigates the reality behind the hypothesis “Does standing on shoulders of a giant can lead you to success?” We study the correlation between “success: and co-authorship with a well-known researcher of a field. AMiner dataset from Computer Science domain was used as test bed ranging from 1936-2014. The results are very interesting, partially in favor of the hypothesis and unfolding some hidden realities against the hypothesis. We have analyzed a factor that how the collaboration with well-known authors of a certain field definitely influences on progress of a junior researcher. With help of the results of this analysis we came to conclude that among junior researchers there are hidden gems who can become visible in future if their capabilities are polished by a senior researcher. If junior researchers, in early part of their career, have a chance to get benefit from the experience of a senior researcher, they can become a rising star very soon. However, this is not the only way of being successful. Our results show that there are more number of successful researchers, who have started without an AA, and their progress was the reason that they got attention of an AA who collaborated with them in later stages of their career. Among the scientists who were lucky enough to work with a ‘hot scientist’ there are two further categories: (1) those who started their career with a senior (for example a PhD candidate who starts his research in supervision of a senior researcher) and (2) those who started independently or with supervisors who were not well-known, but later they got opportunity to work with a well-known senior. The later ones were found to be more productive and hence were able to receive more attention from scholarly society in terms of citations and sociability.

The results of the present study are based on the researchers from the computer science domain. The researcher’s activities can vary from one domain to another. For example, the scholars from Philosophy may dedicate most of their time in writing books and/or giving lectures instead of focusing on conference and journal articles. Thus in domains other than Computer Science, the impact of collaboration can be significantly different and needs to be studied in future. In addition, we plan to extend the present study to reflect the distance of a scholar from the giants in examining the scholarly impact of the scholar.

The causality problem is hard but worth investigating as a follow-up study. For studying causality, we will identify and analyze what factors impact scientific success. In particular, we plan to examine whether the researchers are successful because they co-author with elite researchers or

they were chosen by elite researchers because they were successful. With a deliberate research design, we will attempt to identify whether there exists the endogeneity bias in the causality study.

References:

- Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, 489(7415), 201–202.
- Adegbola, M. (2010). Nurses collaborating with cross disciplinary networks: starting to integrate genomics into practice. *Journal of National Black Nurses' Association: JNBNA*, 21(1), 46–49.
- Adegbola, M. (2013). Scholarly tailgating defined: A diverse, giant network. *The ABNF Journal: Official Journal of the Association of Black Nursing Faculty in Higher Education, Inc*, 24(1), 17.
- Amjad, T., Daud, A., Akram, A., & Muhammed, F. (2016). Impact of mutual influence while ranking authors in a co-authorship network. *Kuwait Journal of Science*, 43(3). Retrieved from <http://journals.ku.edu.kw/kjs/index.php/KJS/article/view/941>
- Amjad, T., Daud, A., Che, D., & Akram, A. (2015). MuICE: Mutual Influence and Citation Exclusivity Author Rank. *Information Processing & Management*.
- Arif, T., Ali, R., & Asger, M. (2014). Author name disambiguation using vector space model and hybrid similarity measures. In *Contemporary Computing (IC3), 2014 Seventh International Conference on* (pp. 135–140). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6897162
- Baker, V. L., & Pifer, M. J. (2011). The role of relationships in the transition from doctoral student to independent scholar. *Studies in Continuing Education*, 33(1), 5–17.

- Chen, Z., Guo, J., Lan, Y., Cao, L., & Cheng, X. (2014). Name Disambiguation by Collective Classification. In *Asia Information Retrieval Symposium* (pp. 406–417). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-12844-3_35
- Daud, A., Abbasi, R., & Muhammad, F. (2013). Finding rising stars in social networks. In *Database Systems for Advanced Applications* (pp. 13–24). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-37487-6_4
- Daud, A., Ahmad, M., Malik, M. S. I., & Che, D. (2015). Using machine learning techniques for rising star prediction in co-author network. *Scientometrics*, *102*(2), 1687–1711.
- Ding, Y., & Stirling, K. (2016). Data-driven Discovery: A New Era of Exploiting the Literature and Data. *Journal of Data and Information Science*, *1*(4), 1–9.
- Dong, Y., Johnson, R. A., & Chawla, N. V. (2015). Will This Paper Increase Your h-index? In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 259–263). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-23461-8_26
- Dong, Y., Johnson, R. A., & Chawla, N. V. (2016). Can Scientific Impact Be Predicted? *IEEE Transactions on Big Data*, *2*(1), 18–30.
- Ebadi, A., & Schiffauerova, A. (2015). How to become an important player in scientific collaboration networks? *Journal of Informetrics*, *9*(4), 809–825.
- Ferreira, A. A., Gonçalves, M. A., & Laender, A. H. (2012). A brief survey of automatic methods for author name disambiguation. *Acm Sigmod Record*, *41*(2), 15–26.
- Gollapalli, S. D., Mitra, P., & Giles, C. L. (2011). Ranking authors in digital libraries (pp. 251–254). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1998123>

- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Kaur, J., Ferrara, E., Menczer, F., Flammini, A., & Radicchi, F. (2015). Quality versus quantity in scientific impact. *Journal of Informetrics*, 9(4), 800–808.
- Kram, K. E., & Isabella, L. A. (1985). Mentoring alternatives: The role of peer relationships in career development. *Academy of Management Journal*, 28(1), 110–132.
- Li, L., & Tong, H. (2015). The child is father of the man: Foresee the success at the early stage. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 655–664). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2783340>
- Li, X.-L., Foo, C. S., Tew, K. L., & Ng, S.-K. (2009). Searching for rising stars in bibliography networks. In *Database Systems for Advanced Applications* (pp. 288–292). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-00887-0_25
- Moreira, C., Calado, P., & Martins, B. (2011). Learning to rank for expert search in digital libraries of academic publications. In *Progress in Artificial Intelligence* (pp. 431–445). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-24769-9_32
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Retrieved from <http://ilpubs.stanford.edu:8090/422>
- Quatman, C., Chelladurai, P., & others. (2008). Social network theory and analysis: A complementary lens for inquiry. *Journal of Sport Management*, 22(3), 338.

- Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, 3(1), 1–16.
- Sayyadi, H., & Getoor, L. (2009). FutureRank: Ranking Scientific Articles by Predicting their Future PageRank. (pp. 533–544). SIAM. Retrieved from <http://epubs.siam.org/doi/abs/10.1137/1.9781611972795.46>
- Shoaib, M., Daud, A., & Khiyal, M. S. H. (2015). Improving Similarity Measures for Publications with Special Focus on Author Name Disambiguation. *Arabian Journal for Science and Engineering*, 40(6), 1591–1605.
- Sweitzer, V. B. (2009). Towards a theory of doctoral student professional identity development: A developmental networks approach. *The Journal of Higher Education*, 80(1), 1–33.
- Tang, J., Jin, R., & Zhang, J. (2008). A topic modeling approach and its integration into the random walk framework for academic search (pp. 1055–1060). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4781224
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). Arnetminer: extraction and mining of academic social networks (pp. 990–998). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1402008>
- van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market. *Current Biology*, 24(11), R516–R517.
- Weidman, J. C., Twale, D. J., & Stein, E. L. (2001). *Socialization of Graduate and Professional Students in Higher Education: A Perilous Passage? ASHE-ERIC Higher Education Report, Volume 28, Number 3. Jossey-Bass Higher and Adult Education Series*. ERIC. Retrieved from <http://eric.ed.gov/?id=ED457710>

Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012). To better stand on the shoulder of giants. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (pp. 51–60). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2232831>

Zhang, J., Tang, J., Liu, L., & Li, J. (2008). A mixture model for expert finding. In *Advances in Knowledge Discovery and Data Mining* (pp. 466–478). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-68125-0_41