ISSI
2019

SAPIENZA
UNIVERSITÀ DI ROMA

17th INTERNATIONAL CONFERENCE ON SCIENTOMETRICS & INFORMETRICS

ISSI2019

with a Special STI Indicators Conference Track

2-5 September 2019
Sapienza University of Rome, Italy

PROCEEDINGS

VOLUME II

EdizioniEfesto

PROCEEDINGS OF THE 17TH CONFERENCE OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS

© Edizioni Efesto - ISBN 978-88-3381-118-5 - August 2019
Printed in Italy

Editors: *Giuseppe Catalano, Cinzia Daraio, Martina Gregori, Henk F. Moed and Giancarlo Ruocco*

Graphic cover design: *Francesco Manzo* | *graframan.com*
Cover photo: ©*Fayee - stock.adobe.com*

An empirical analysis on the relationship between publications and academic genealogy

Rogério Mugnaini[1], Rafael J. P. Damaceno[2] and Jesús P. Mena-Chalco[3]

[1] *mugnaini@usp.br*
University of São Paulo, School of Communication and Arts, Dept. of Information and Culture, Av. Prof. Lúcio Martins Rodrigues 443, São Paulo, SP 05508-020 (Brazil)

[2] *rafael.damaceno@ufabc.edu.br*
Federal University of ABC, Center for Mathematics, Computation and Cognition, Av. dos Estados 5001, Santo Andre, SP 09210-580 (Brazil)

[3] *jesus.mena@ufabc.edu.br*
Federal University of ABC, Center for Mathematics, Computation and Cognition, Av. dos Estados 5001, Santo Andre, SP 09210-580 (Brazil)

Abstract

In the belief that collaboration between advisor and students is a means of following a scientific path for the discovery of new knowledge, this study examines the relationship between bibliometric indicators for publications and academic genealogy. In this study, we analysed the curricular information of more than 40,000 PhD advisors, registered in a huge Curriculum Vitae dataset. This involved displaying different patterns of academic fecundity and publications, concerning several areas and mentoring ages. It was found that productivity in co-authorship with academic sons is an established practice in the hard sciences, while in the soft sciences it is only a reality for researchers until the 25 years (mentoring age). In addition, in the case of the output produced without the participation of the students, there was a constant distribution among the mentoring age groups (with the exception of Agricultural Sciences and Engineering, where there was a gradual decline over the period). Finally, there was a number of advisors that performed best in fecundity but worst in production, which suggests that the involvement in mentoring impairs the advisor's capacity for research. It can be concluded that a separate analysis of the researchers' output is needed, since student participation may be important for an assessment of scholarly performance.

Introduction

Brazilian scholarly output is concentrated in universities, particularly in public universities, where almost all of the Graduate Programs are run. The Coordination of Improvement of Higher Level Personnel (CAPES) has been the national agency responsible for the funding, evaluation, and support of Brazilian graduate education since 1976. The accredited researcher in this system will have his/her academic performance evaluated on the basis of teaching experience, the training of master´s and doctoral students and scientific production. Moreover, although the production of the researcher is the main object of the assessment, the students´ output is also evaluated, as well as their collaboration with a supervisor in producing co-published articles (Oliveira & Amaral, 2017).

Scholarly output can be observed in different ways. In this paper, we focused in both publications and academic mentoring in the training of young researchers. We believe that these terms are compatible when we look at the scholarly output as well as the relations formed by student-advisor pairs over time, which allows us to identify and analyse patterns in the academic genealogy.

Studies based on different approaches have focused on the relationship between publications and academic mentoring, to find out if the students are now publishing articles in greater quantity (Green & Bauer, 1995; Pinheiro, Melkers & Youtie, 2014). They are also concerned with analysing the relationship of co-authored publications between advisor and student (Tuesta et al, 2015) and investigating the involvement and positive influence of advisors in doctoral studies (Horta & Santos, 2016).

In light of this, Larivière (2012) analysed the involvement of more than 27,000 doctoral students in peer review publications. The author states that the grouping of doctoral students in research teams both assists and and encourages students to take part in different research projects. At the same time, Qui et al (2017) showed that collaboration with first-class scientists significantly improves young researchers' careers. They provided evidence that the benefits of working with an outstanding scientist are more noticeable in the early stages of a young student's career. These factors explain why it is important for advisors to give encouragement to students to conduct scientific research very early on.

Apart from the questions of publications, as well as the output that can derive from the mentoring process, many factors can lead to academic success (Reskin, 1977; Bäker, 2015). For this reason, the concern of a supervisor and of scientific policies should be to ensure the continuity of scientific knowledge throughout the generations, which may have more value than the publications or awards that a scholar might receive (Andraos, 2005). In light of this, it is not a question of stimulating the academic fecundity of advisors (or academic fathers), and thus ensuring that their students (or academic sons) are, for example, productive, but of encouraging advisors to foster productive and also fecund sons (Malmgren, Ottino & Amaral, 2010; Heinisch & Buenstorf, 2018). This means that the research output is no longer the target, but rather, the focus is on the widening and transmission of capacity (Bozeman, Dietz & Gaughan, 2001).

This study has sought to describe the relationship between the quantitative indicators of publications and academic genealogy. However, although these factors should be correlated, there is a need for comparing its behaviour, not only among the different areas of knowledge but also among scholars of different age groups. Despite the correlation, this supports the hypothesis that there might be advisors whose performance in publications and academic mentoring performance are antagonistic, because there are those who concentrate their efforts, in one area, to the detriment of another.

We have used a dataset of academic curricula nationwide, which contains the student-advisor pairs, permitting to establish the academic genealogy and respective metrics. Our decision was not to carry out a longitudinal study (Levin & Stephan, 1991), which despite being ideal for measuring the effect of specific factors over time, would be impossible for comparing groups of researchers from different generations – these have been subjected to different scientific policies.

Material and Methods

In this study, we measured both the scientific publications and the academic mentoring relationships of PhD researchers working in Brazil. With regard to scientific publications, we count the total number of publications of each advisor in scientific journals, conferences, book chapters, and books and determine how many of these publications are co-authored with their academic sons. Concerning the academic mentoring, i.e., the training of new researchers, three genealogical metrics were used, namely, academic fecundity, descendants (offspring) and the genealogical index.

Fecundity (F) refers to the number of mentoring relationships that a researcher has already established. Descendants (D) indicate the number of mentoring relationships established with the students, and also the future relationships of these students with their own students. It takes into account all the generations of a researcher, i.e., it also includes the direct academic sons, the indirect relationships (grandsons, great-grandsons, and so on). The genealogical index (GI) of an academic is defined as the largest number of *g* sons of an academic that have at least *g* sons (Rossi et al., 2017).

The procedure shown in Figure 1 is adopted to analyse the academic genealogy combined with scientific publications. It also contains a flowchart divided into five stages: (i) collecting and cleaning, (ii) extraction of scientific publications, (iii) merging, (iv) selection of researchers and (v) analysis (see Figure 1). The following sections describe these stages, which generate a dataset containing researchers with information regarding their academic genealogy and scientific publications.

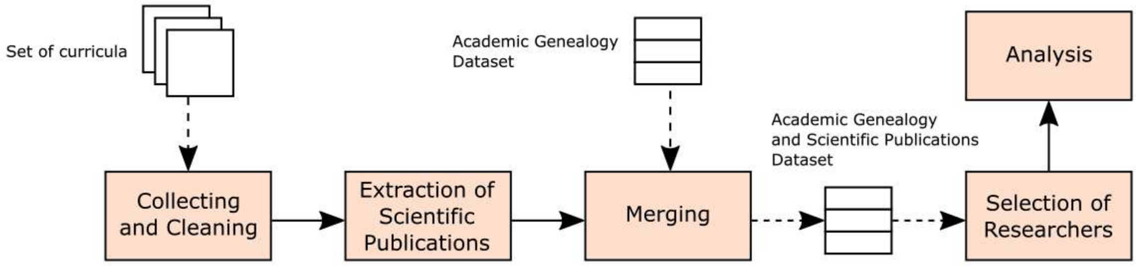

Figure 1. A flowchart that shows the five stages of the method applied in this work: collecting and cleaning, extraction of scientific publications, merging, selection of researchers and analysis.

Collecting and Cleaning

The study started from an existing dataset consisting of the PhD researchers working in the Brazilian Graduate Program, which was compiled in our previous study (Damaceno et al. 2019). The process of compiling this dataset involved drawing on information collected from the set of academic curricula of PhD researchers registered in the Lattes Platform (a large Curriculum Vitae dataset). In this dataset, there is information on each researcher's field of study (or areas of knowledge), individual identifier, academic degrees, and mentoring relationships, as well as the genealogical metrics of academic fecundity, the descendants and genealogical index. This dataset also contains the full names and curricular identifiers of the researchers' students (academic sons).

When adding information about the scientific publications in the genealogical dataset, we had to collect the same set of curricula that was to form the academic genealogy. It is essential to use the same set of curricula, i.e., obtained at the same time as the original dataset was formed, since the scientific publications must correspond with the same time as the information about the academic genealogy. After we collected these curricula (dated August 2017), we carried out the data cleaning. All diacritic marks were excluded as well as the characters with accents were transformed to English alphabet. All the characters were transformed to the lowercase. Any articles, books or book chapters without a title were not taken into account.

Extraction of Scientific Publications

The curriculum of each researcher has a section called “Bibliographic Production”, from which we extracted all the information that refers to articles, books or book chapters. In the case of articles, we only took into account the full papers published in journals or in the proceedings of a conferences (expanded abstracts were not included). In the case of books or book chapters, only full texts, encyclopedias, catalogs or collections in both printed or digital versions, were included. The total number of scientific publications of a researcher was calculated as the sum of all the articles, books and book chapters that he/she had published.

Additionally, this study counts the works produced in collaboration with academic sons. Each researcher's publication has a list of its co-authors' names (the initial of the forenames and the complete surname) and its co-authors' identifiers (IDs that identify their curricula in the dataset). A comparison was made between the identifiers and names to check if a researcher's publication was co-authored with some of the researcher's academic sons. With regard to each researcher's publication, the co-authors' identifiers must be identical to some identifier in the list of the researcher son's identifiers (collected from the dataset) to ensure it was co-authored with academic sons. If a co-author does not have an identifier, his/her name must be in the list of the researcher son's names (initial of the first name and the complete surname). The same publication was only counted once since a researcher could have co-authored it with two or more students.

Merging

The process of incorporating data about scientific publications into the genealogical dataset relied on the individual identifier of each researcher in the Lattes Platform (also included in the genealogical dataset). Each scientific publication obtained in the last stage is linked to an individual identifier - the same that is included in the genealogical dataset. Hence, the process of including this information resulted in a dataset consisting of both the academic genealogy and the scientific publications for each researcher in the original dataset.

Selection of Researchers

The academic genealogy dataset, together with the scientific publications added to it, contains information regarding 271,370 PhD researchers. We only analysed a proportion of these that met two requirements: (i) researchers that have at least one mentoring relationship completed in the doctoral studies (ruling as an advisor), and (ii) researchers that have at least one publication since the year they completed their first mentoring relationship.

Analysis

The researchers were separated into eight groups that represent the eight areas of knowledge defined by CAPES, which are as follows: Agricultural Sciences (AGR), Biological Sciences (BIO), Engineering (ENG), Exact and Earth Sciences (EXA), Health Sciences (HEA), Humanities (HUM), Linguistics, Language & Literature and Arts (LIN) and Applied Social Sciences (SOC). We analysed the areas of knowledge globally, and in accordance with the mentoring age defined in this study as the time passed (in years) since a researcher has finished the mentoring of his/her first PhD student. There are ten mentoring age groups, which are as follows: 1 to 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, 31 to 35, 36 to 40, 41 to 45 and 46 to 50 years.

Two metrics were used to analyse the scientific publications: Production with Academic Sons (PAS) and the remaining part of the Total Production (TP), calculated by TP - PAS. TP refers to all the work published by a researcher, since he/she completed the first mentoring relationship (ruling as an advisor). TP - PAS refers to the part of these scientific publications that a researcher co-authored with his/her students. For both measurements, we only took into account book chapters or entire books and the full papers published in journals or conferences. Further, we calculated a coefficient, that is the ratio between each scientific publication metrics (PAS and TP - PAS) and the “Fecundity” metric.

Dataset

The dataset obtained as a result of the five stages previously described, contains information about the knowledge area of the researchers, such as their mentoring age, the total number of their scientific publications and the percentage of these publications that was undertaken with students. These data are divided into two groups: N1, which represents all the academics that met all the requirements set out in the “Selection” section, and N2, a subset of N1. This includes meeting another requirement: researchers that have a score higher than zero in the genealogical index (or those that have at least one grandson).

Table 1 shows the total number of researchers and the median and average values for the mentoring age and of the researchers for N1, grouped by their area of knowledge. N1 represents the main dataset containing 40,368 researchers. Information about N2 is also shown, in which there are 10,996 researchers.

Table 1. Number and percentage of academics and their average and median mentoring age for each area of knowledge. The last three columns show the number of academics and include a sub-dataset consisting of all the academics with genealogical index greater than or equal to 1.

	N1				*N2*		
Area	*Academics*		*Mentoring Age*		*Academics*		
	N	*%*	*Avg.*	*Med.*	*N*	*%*	*% N1*
AGR	4,012	9.94	11.79	10	1,065	9.68	26.54
BIO	6,023	14.92	12.44	10	1,703	15.49	28.27
ENG	4,371	10.83	13.11	12	1,224	11.13	28.00
EXA	6,693	16.58	13.16	11	1,804	16.41	26.95
HEA	7,004	17.35	12.40	11	2,032	18.48	29.01
HUM	6,337	15.70	11.09	9	1,698	15.44	26.79
LIN	2,223	5.51	11.40	10	579	5.26	26.04
SOC	3,705	9.18	10.63	9	891	8.10	24.05
All	40,368	100.00	12.13	10	10,996	100.00	27.24

Results and discussion

First of all, we analysed the publication profile of the academics in the eight areas of knowledge (see Table 2). The TP does not reveal notable differences between the areas, while PAS shows a trend of HUM, LIN, and SOC to publish in a smaller quantity with students.

Table 2. TP, FP and percentage of FP for each area of knowledge.

Area	*TP*		*PAS*		*TP - PAS*	
	Avg.	*Med.*	*Avg.*	*Med.*	*Avg.*	*Med.*
AGR	64.45	39	26.46	9	37.99	25
BIO	49.36	29	19.27	6	30.09	20
ENG	78.77	51	32.95	12	45.82	31
EXA	52.65	30	18.53	5	34.12	21
HEA	62.46	38	19.67	6	42.79	27
HUM	38.38	23	6.46	1	31.92	20
LIN	29.75	17	2.27	0	27.49	16
SOC	43.14	25	7.77	1	35.38	22
All	53.49	31	17.41	4	36.08	22

We then proceeded to analyse the behaviour of publication coefficients concerning the genealogical metrics of groups of academics of different mentoring age groups. As can be seen in Figure 2b, the distributions of the different age groups have a very similar profile when account is taken of the total number of publications and fecundity, meaning there is no increase in productivity.

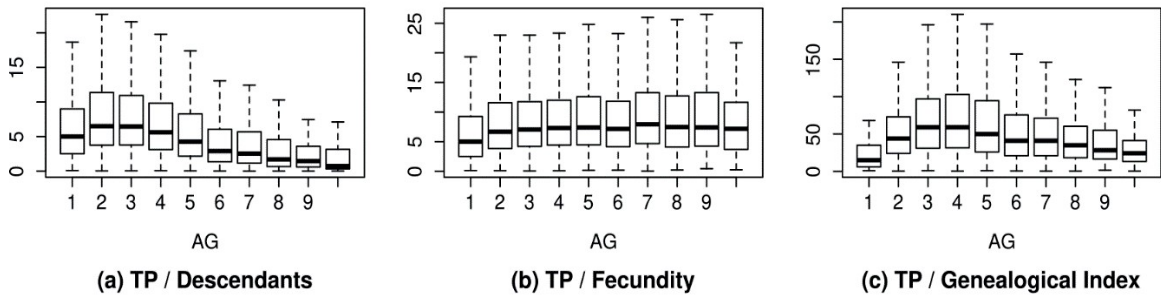

Figure 2. Distribution of academics by Age Group (AG) and the three coefficients, as follows: (a) TP / Descendants, (b) TP / Fecundity, and (c) TP / Genealogical Index. The X axis represents the mentoring age groups and the Y axis the boxplot of the respective coefficient.

Figures 2a and 2c, consider the coefficients by weighing the publications, in terms of the number of descendants and genealogical index respectively. The profile shown is different, since older researchers accumulate more people in their lineage, and scientific publications do not increase in the same proportion (perhaps because co-publication largely occurs with their academic sons, who are their direct descendants). A decline was noted in both cases, beginning from the fourth age group (16-20 years) when this includes the number of offspring; and from the fifth age group (21 to 25 years), in the case of the genealogical index.
As these are cumulative variables, it should be noted what happens to each one, independently. On the one hand, some features of the distributions, such as the coefficients obtained with the descendants and fecundity metrics (Figures 2a and 2b), show absolute values in a very similar range. On the other hand, there is a difference shown by the descendants, which, in addition to declining for the older age groups, reduces their dispersion (mainly among researchers with a low number of scientific publications, denoting a more marked asymmetry).
This behaviour reveals that in the case of Figure 2a, the cumulative effect of descendants reduces the coefficient. However, in the case of Figure 2b, there was an increase in the advisors' productivity. In other words, if their extra-mentoring publications showed a significant growth, the effect of the coefficients on the age groups would be inversely proportional (in terms of increase and dispersion). In view of this scenario, we thought it would be of value to broaden the analysis of the fecundity variable, separating the scientific production in two parts: Production with Academic Sons (PAS) and the remaining publications of the fathers (TP - PAS), or extra-mentoring publications (those not co-authored by the academic sons).
Figure 3 illustrates that, in general, there is a clear difference between the range of variables, for hard and soft sciences. In the case of soft sciences, co-authorship with students is very low, which may be due both to the low level of collaboration in these areas and to the fact that the academic sons are less involved in the advisor's research (Larivière, 2012). SOC is the area with the largest range of production without academic sons' participation, among all the areas. This is due to the tenth age group (more than 45 years), whose productivity is significant - this profile is usually not observed in most studies, which restrict publications to journal articles. Among the hard sciences, BIO has the highest proportion of production in co-authorship with students, and EXA the lowest.
In light of the distributions of box-plots from hard sciences, it is clear that production without students' participation (BIO, EXA and HEA) is increasing in the first age groups, and remains constant throughout most of the groups (from the fifth age group). On the other hand, AGR fluctuates between the intermediate groups, while ENG decreases from the fourth age group - both show a significant reduction in dispersion among the older researchers. With regard to co-authorship with their academic sons, there is less productivity for the first age groups,

which may be due to the fact that the younger advisors have not yet consolidated into groups or formed a research network that makes it easier for the students to be incorporated.

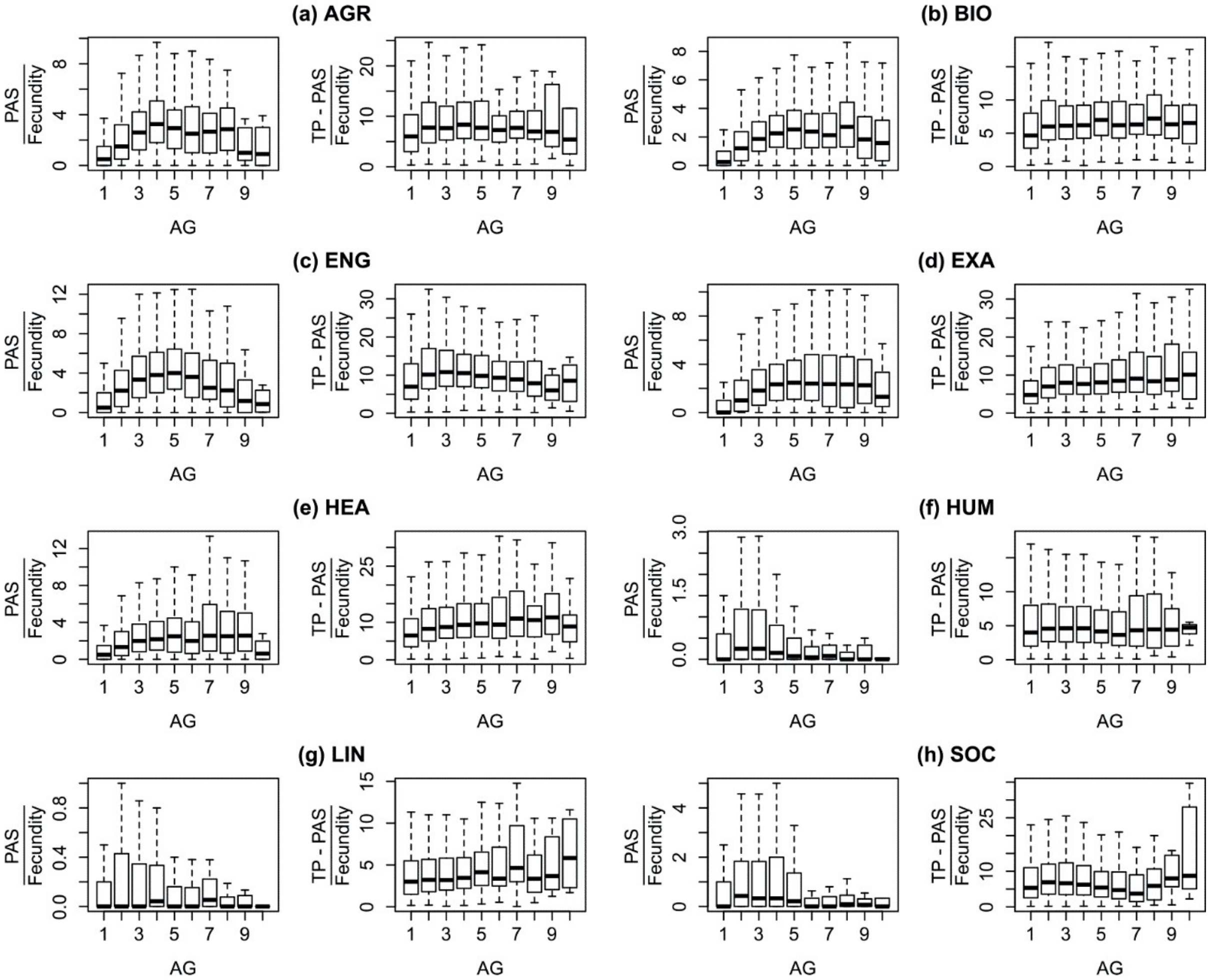

Figure 3. Distribution of academics by Age Group (AG), and coefficients of Production with Academic Sons (PAS) and and the remaining part of the production (TP - PAS). The academics are discretized in major areas of knowledge: (a) AGR, (b) BIO, (c) ENG, (d) EXA, (e) HEA, (f) HUM, (g) LIN and (h) SOC.

The soft sciences have an opposite profile, in which productivity in co-authorship with their academic sons is more pronounced among the younger researchers (up to 20 or 25 years), which can be attributed to the criteria governing the graduate scientific policy, which were laid down in 1998 and encourages co-authorship with students.

Thus, with regard to production without students' participation, it can be seen that, in addition to BIO, EXA and HEA, the soft sciences also showed a more constant distribution among the age groups. The fact that these areas represent about 80% of the total number of researchers explains the behaviour observed in Figure 2b.

Figure 4 shows the performance of the same pair of variables examined above, where the medians are in the scatter plots, with an arrow identifying the oldest age group. The series on the left side (y scale) shows the values for the parents in collaboration with their sons (PAS), while the series on the right side (y scale) relates to the remaining papers of the fathers (TP - PAS). There is an increase in the relationship between fecundity and the production variables that can be observed through almost all the age groups, with regard to the hard sciences. The main difference is that (except for ENG), while the fecundity declines significantly in the older age groups, the production in co-authorship with the students decreases more than the production without them. Among the social sciences: HUM shows a clear linear growth through almost all the age groups, in the case of production without the students, while LIN and SOC show a decline in fecundity among the older age groups. With regard to the production with the students, there is an increase until the third age group to HUM, and the fourth to SOC, while LIN shows than the median of production is zero for almost all the age groups.

It should be noted that, in contrast with the longitudinal approaches, which revealed that during academic life there is a decline in productivity - in terms of scientific publications (Levin & Stephan, 1991), this study covers the entire scientific publications of academics. For this reason, it is not possible to analyse the academics' careers. The comparisons between age groups are made with different groups of academics, which causes an increase in the number of publications originating from the growth in the number of scientific publications produced by the oldest academics.

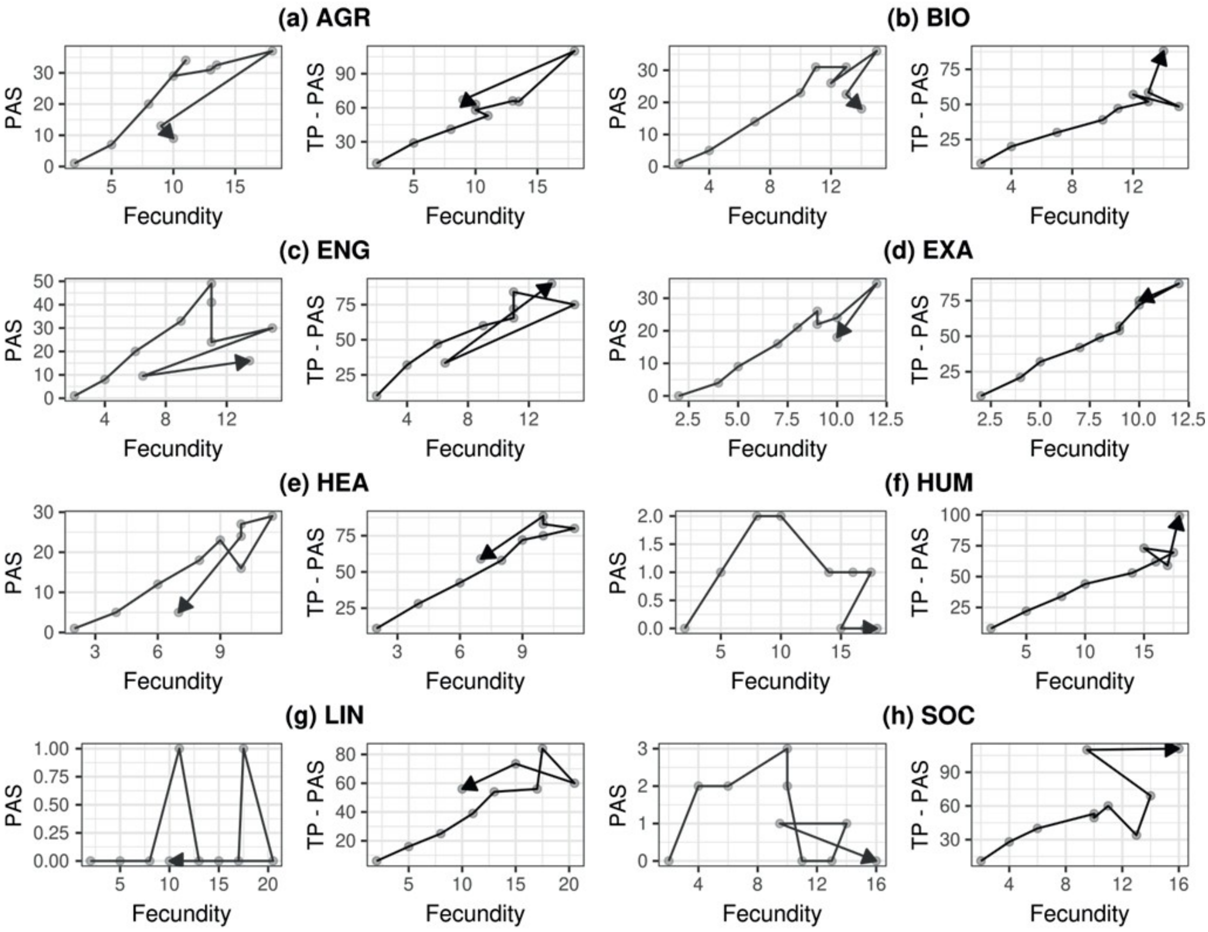

Figure 4. Distribution of academics by Fecundity (X), PAS (Y), and TP - PAS (Y). The academics are discretized in major areas of knowledge: (a) AGR, (b) BIO, (c) ENG, (d) EXA, (e) HEA, (f) HUM, (g) LIN and (h) SOC. Each point corresponds to the median of X and Y for the ten mentoring age groups concerned. The arrow represents age group 10.

Finally, some analytical factors should be noted with regard to one of the objectives of this study, which refers to researchers whose performance in production and fecundity shows a contrast (i.e., Q1 for the former and Q4 for the latter, or opposite). Moreover, the percentages of researchers whose performance in each of the measures is similar (i.e., Q1 or Q4 in both

measures) are also displayed, since they represent the expected relationship between the variables (which is the association between them, since the more mentoring relationships there are, the higher the scientific output derived from this relationship).
An examination of Table 3, shows that the respective percentages of production without students' participation (TP - PAS) and fecundity, when performance is better (Q1 for both) and worse (Q4 for both), are between 13.06% and 18.03%. As for production in co-authorship with the academic sons, these percentages are between 16.22% and 23.26% - with bigger percentages in the last quartile of both variables. Additionally, it can be seen that the association between the variables is stronger when the production is co-authored with students.
When the areas are compared, it is evident that ENG, HUM and LIN show the lowest values concerning Q4 (both variables), and production without the participation of academic sons; while BIO, EXA and HEA are the biggest. In the case of Q1 (both variables), there is less variability, with HUM and LIN performing best.
When the production in co-authorship with the students is analysed, the percentages are slightly higher, with BIO and HEA having the biggest percentage in Q4 (both variables) and BIO and ENG in Q1 (both variables).
As noted in Table 2, the medians of PAS for the soft sciences are very low (at most, one), making it impracticable to determine Q4, which explains the presence of empty cells in Table 3 - and the same occurred with EXA, which even had a median of 5.

Table 3. Percentage of academics in the Q1 and Q4 with regard to: (a) (TP - PAS) vs. Fecundity (F) and PAS vs. Fecundity (F).

Area	$(TP - PAS)_{Q1} \wedge F_{Q4}$	$(TP - PAS)_{Q4} \wedge F_{Q1}$	$(TP - PAS)_{Q4} \wedge F_{Q4}$	$(TP - PAS)_{Q1} \wedge F_{Q1}$	$PAS_{Q1} \wedge F_{Q4}$	$PAS_{Q4} \wedge F_{Q1}$	$PAS_{Q4} \wedge F_{Q4}$	$PAS_{Q1} \wedge F_{Q1}$
AGR	0.52	2.04	14.98	13.06	0.17	0.17	20.09	17.07
BIO	0.50	1.61	17.8	13.76	0.33	0.08	23.04	18.28
ENG	0.41	1.49	16.93	13.09	0.41	0.09	22.19	17.25
EXA	0.49	2.29	18.03	13.67	-	-	-	-
HEA	0.50	1.83	17.99	13.34	0.43	0.16	23.26	16.22
HUM	0.38	1.59	16.10	14.69	-	-	-	-
LIN	0.40	1.80	15.29	15.38	-	-	-	-
SOC	0.92	2.97	17.27	13.09	-	-	-	-

Finally, when the opposite kinds of behaviour are analysed in Table 3, it can be seen the first and fifth columns, with production (Q1) and fecundity (Q4), have the smallest values. This suggests that higher productivity is less probable when it is less fecund - and obviously, this situation in more pronounced in the production that is co-authored with students. It is clear that SOC has the highest percentage in the first column, followed by AGR, BIO and EXA. In the production with the participation of the students (fifth column) HEA and ENG are highlighted.
The opposite situation is more pronounced, when there are higher percentages of researchers that perform worse in production, even though they perform best in fecundity. This is more pronounced in the production without students, suggesting that the effort in mentoring disables the advisor's research productivity. It should also be noted that SOC has the highest percentage in the second column, followed by EXA and AGR. Regarding the scientific

publication with the participation of the students (sixth column) AGR and HEA are positively highlighted.

Conclusion

On the one hand, scientific productivity has been measured in the past, to a significant extent, by means of bibliometric measures, i.e., those based on the production of papers, books, and scientific citations, among other factors. On the other hand, recent works have measured scientific output also in terms of academic genealogy, i.e., through the formation of human resources. In this study, we conducted an empirical analysis of the relationship between scientific publications and academic genealogy of the PhD researchers who participated in the formation of scholars related to Brazilian science.

The evidence of a relationship between publications and genealogical metrics has made it possible to observe that fecundity is more closely related to publications. This result suggests that the stimulation of the scientific policy may be contributing to the research conducted by scholars. Yet, this may in some way be limited to the research conducted by the students, or else an increase in productivity would be observed. On the other hand, it might be owing to a strong involvement of students in the advisor's research, which is not necessarily the case in some areas, since productivity with academic offspring is declining among the age groups. As was expected, the main differences were found between the hard and soft sciences, and this is worth noting because productivity in co-authorship with sons is only a reality for young researchers in the latter category (i.e. soft sciences). Additionally, some specific features should be highlighted in these areas: Biological Sciences showed the highest proportion of production in co-authorship with academic, which may be the result of the students being more closely involved in their advisor's research; in contrast, the Social Sciences had the largest coefficient for productivity without the participation of the sons, in absolute numbers.

Finally, the analysis of ´antagonism´ in the performance of advisors with regard to publications and academic mentoring, revealed the following: higher productivity is less probable when it is less fecund, and is a factor that is more pronounced in production with students; in sharp contrast, and in a more pronounced way, it was found that there were higher percentages of researchers that perform worse in terms of production, even though they are best in fecundity. It was even more evident in the production without students, which suggests that the involvement in mentoring impairs the advisor's capacity for research.

In subsequent studies, it may be useful to find out if the other genealogical metrics (descendants and genealogical index) are related to the impact measured in citations, as an outcome of the indirect relationship established by the genealogy. Studies of genealogical metrics and their relationships with publications and citation impact may offer a wider perspective and could thus be included in the evaluative processes such as those found in Brazil.

Acknowledgements

This work would not have been possible without the financial support of São Paulo Research Foundation (FAPESP), grants #2012/00255-6; #2019/07118-3. The authors thank anonymous reviewers for constructive comments on improving the presentation of this work.

References

Andraos, J. (2005). Scientific genealogies of physical and mechanistic organic chemists. *Canadian journal of chemistry*, 83(9), 1400-1414. doi:10.1139/v05-158

Bäker, A. (2015). Non-tenured post-doctoral researchers' job mobility and research output: An analysis of the role of research discipline, department size, and coauthors. *Research Policy*, 44(3), 634-650. doi:10.1016/j.respol.2014.12.012

Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: an alternative model for research evaluation. *International Journal of Technology Management*, 22(7-8), 716-740. doi:10.1504/IJTM.2001.002988

Damaceno, R. J. P., Rossi, L., Mugnaini, R., Mena-Chalco, J. P. (2019). The Brazilian academic genealogy: Evidence of advisor-advisee relationships through quantitative analysis. *Scientometrics.* 119(1), 303-333. doi:10.1007/s11192-019-03023-0

Green, S. G., & Bauer, T. N. (1995). Supervisory mentoring by advisers: Relationships with doctoral student potential, productivity, and commitment. *Personnel Psychology*, 48(3), 537-562. doi:10.1111/j.1744-6570.1995.tb01769.x

Heinisch, D. P., & Buenstorf, G. (2018). The next generation (plus one): an analysis of doctoral students' academic fecundity based on a novel approach to advisor identification. *Scientometrics*, 117(1), 351-380. doi:10.1007/s11192-018-2840-5

Horta, H., & Santos, J. M. (2016). The impact of publishing during PhD studies on career research publication, visibility, and collaborations. Research in Higher Education, 57(1), 28-50. doi:10.1007/s11162-015-9380-0

Larivière, V. (2012). On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics*, 90(2), 463–481. doi:10.1007/s11192-011-0495-6

Levin, S. G., & Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. The American Economic Review, 81(1), 114-132.

Malmgren, R. D., Ottino, J. M., & Amaral, L. A. N. (2010). The role of mentorship in protégé performance. *Nature*, 465(7298), 622. doi:10.1038/nature09040

Oliveira, T. M. de, & Amaral, L. (2017). Public Policies in Science and Technology in Brazil: challenges and proposals for the use of indicators in evaluation. In: Mugnaini, R.; Fujino, A.; Kobashi, N. Y. (Orgs.), *Bibliometrics and scientometrics in Brazil: scientific research assessment infrastructure in the era of Big Data* (pp.189-217), São Paulo: ECA/USP. doi:10.1344/BiD2018.40.19

Pinheiro D., Melkers J., & Youtie J. (2014). Learning to play the game: Student publishing as an indicator of future scholarly success. *Technological Forecasting & Social Change*, 81, 56–66. doi:10.1016/j.techfore.2012.09.008

Qi, M., Zeng, A., Li, M., Fan, Y., & Di, Z. (2017). Standing on the shoulders of giants: The effect of outstanding scientists on young collaborators careers. *Scientometrics*, *111*(3), 1839-1850. doi:10.1007/s11192-017-2328-8

Reskin, B. F. (1977). Scientific productivity and the reward structure of science. American sociological review, 42(3), 491-504. doi:10.2307/2094753

Rossi, L., Freire, I. L., and Mena-Chalco, J. P. (2017). Genealogical index: A metric to analyze advisor-advisee relationships. Journal of Informetrics, 11(2), 564-582. doi:10.1016/j.joi.2017.04.001

Tuesta E. F., Delgado K. V., Mugnaini R., Digiampietri L. A., Mena-Chalco J. P., & Pérez-Alcázar J. J. (2015). Analysis of an Advisor–Advisee Relationship: An Exploratory Study of the Area of Exact and Earth Sciences in Brazil. PLoS ONE, 10(5), e0129065. doi:10.1371/journal.pone.0129065