

RESEARCH

Open Access



# Software engineering research in Brazil from the perspective of young researchers: a panorama of the last decade

Arilo C. Dias-Neto<sup>1\*</sup>, Rafael Prikladnicki<sup>2</sup>, Márcio de O. Barros<sup>3</sup> and Leonardo G. P. Murta<sup>4</sup>

## Abstract

**Background:** After graduating and starting a career as a professor, a young researcher usually finds him/herself lost due to a huge amount of new obligations and opportunities. Choosing the best strategy to guide his/her career into a productive and successful end is not an easy task to everyone, leading sometimes to anxiety and frustration. Previous studies on this topic looked only at published papers as the main source of analysis, lacking an analysis from the researchers' perspective. Such analysis would be able to identify relevant researchers in the field and which decision and actions they took along their careers.

**Methods:** We surveyed more than 30 researchers using snowball sampling and mining their profiles.

**Results:** We identified patterns related to the universities and regions that formed and hired the most prominent Brazilian young researchers in software engineering. We also found patterns related to research areas (within software engineering), vehicles where research results were published, citations, and joint publication.

**Conclusions:** We observed that the current generation of prominent researchers has graduated in the most important universities in Brazil and are still working (most of them) in Brazilian federal universities. Analyzing publication patterns, we observed that they target high-quality conferences and journals and usually collaborate strongly with a large number of peers. They also tend to establish themselves in a given research area and propose and develop workshops and conferences to promote the expansion of research on that area.

**Keywords:** Software engineering research; Career development; Snowball sampling

## Background

After finishing a Ph.D. and starting a career as a professor, a young researcher usually finds him/herself lost due to the huge amount of new obligations and opportunities. As a Ph.D. student, the main activities one faces are related to research and, sometimes, teaching. After assuming the role of professor, besides the usual research and teaching, other activities are introduced in different degrees depending on the university. These new activities include writing grant proposals, supervising students, serving in committees, assuming administrative roles, running projects with the industry, among others.

Each of these activities requires decisions and actions that may lead a capable professor to produce relevant results and, consequently, be appreciated by a given community. For instance, professors performing administrative roles may improve the means of organizing and managing the university, thus increasing the productivity of professors and administrative workers. Capable professors dedicated to teaching may produce good courses, eventually new teaching methodologies, among other useful results from a student's point-of-view. In this paper, we focus on research activities, particularly those performed within the field of software engineering.

Identifying the focus of prominent researchers and the results they have achieved allows recording successful cases that may help future generations to make informed decisions throughout their careers. In this sense, Silveira-Neto et al. [1] summarize the findings from a mapping study

\* Correspondence: arilo@icomp.ufam.edu.br

<sup>1</sup>Instituto de Computação, Universidade Federal do Amazonas (UFAM), Manaus, AM, Brazil

Full list of author information is available at the end of the article

about the first 25 editions of the Brazilian Conference on Software Engineering (SBES) (1987–2011). This study allows understanding the past of software engineering area in Brazil, including the main researchers or research groups (USP, UFPE, UFRJ, PUC-RJ, and so on), subareas (requirements, testing, and methodologies), distribution of papers per university/state (centralized in the southeast, northeast, and south), among other perspectives. It also points out the difficulties and challenges to be met by young software engineering researchers, such as difficult interaction with companies, the bureaucracy of the funding agencies, high student turnover, pressure for high productivity, and unfair and erroneous evaluation by conferences [2].

The former study departed from the published papers to draw a picture of 25 years of software engineering research in Brazil. However, we lack an analysis from the researchers' perspective, in order to determine what relevant researchers have achieved and which decision and actions they took along their careers led to such achievements. These actions and decisions are not frequently recorded in documents and most can only be retrieved by means of interviews with researchers. Nevertheless, despite of the best efforts, even such an interview may be biased towards those actions that led to successful results, disregarding ineffective ones and possibly drawing an incomplete history about the researcher. A way to avoid such bias is through a peer perception of their importance.

The goal of this paper is to identify the most prominent Brazilian young researchers in software engineering and highlight some patterns related to the path they chose to build a carrier and achieve prominence. We surveyed more than 30 software engineering researchers using snowball sampling [3], asking whom they considered the most prominent Brazilian young researchers in software engineering. After that, we identified patterns related to universities and regions that formed and hired them. We also explored patterns related to research areas within software engineering, vehicles where results were published, citations, and joint publication.

This paper is organized into six sections besides this introduction. "Methods" section presents the process that we used to identify the young and prominent software engineering researchers analyzed in our study. "Collected data" section discusses how we have collected and organized the data. "Contextual analysis and discussion" section presents the obtained results in terms of research universities where young researchers have concluded their Ph.D. and are currently working, research area, citations, as well as some patterns regarding how young researchers reported the results of their work, accounting for both conferences and journals, the sequence in which the venues were addressed, and joint publications. "Threats to validity" section presents threats to the validity of our results, while

"Related work" section presents related works. Finally, "Conclusions" section concludes the paper, summarizes the obtained results, and outlines our future work.

## Methods

As previously stated, the goal of this paper is to identify the most prominent Brazilian young researchers in software engineering and highlight some patterns related to the path they chose to build a carrier and achieve prominence. This leads to two important questions:

- Q1: What means "young" in terms of research?
- Q2: How can we point out the most prominent Brazilian young researchers in software engineering?

### Q1: What means "young" in terms of research?

Regarding the first question, we established a lower and upper boundary in terms of research age. The lower boundary is the Ph.D. defense. Students do not face the bureaucratic issues of being a professor and are sometimes strongly helped by their advisors. On the other hand, we defined the upper boundary as 10 years after the Ph.D. defense, which is the usual threshold separating junior and senior researchers for grants conceded by the most important Brazilian research agencies (such as CAPES, CNPq, and state-run agencies like FAPERJ, for instance). The selection of such boundaries attenuates the problem of comparing people's performance in different contexts. For instance, International Conference on Software Engineering (ICSE) submission and acceptance rates in the 90s were respectively around 200 and 20 %. However, in the last years, it raised to 400 and 15 %. In other words, in the 90s, a paper would need to beat other 160 papers to be accepted. Now, this number has increased to 340.

---

*Dear Prof. Minerva McGonagal<sup>1</sup>, we are running a study focused on identifying (nationally and/or) internationally prominent BRAZILIAN young researchers in Software Engineering. This work continues the previous research conducted by Professors Leonardo Murta (UFF) and Márcio Barros (UNIRIO) in the international context. It consists on recursively collecting indications and composing a graph based on these indications. This will allow us to analyze the most cited prominent Brazilian young researcher's curriculum, trying to figure out the most important decisions that led them towards their top position. Prof. Albus Dumbledore, from Hogwarts School of Witchcraft and Wizardry, has indicated you as a young and successful Software Engineering researcher. Can you please list up to five other Brazilian young Software Engineering researchers, which hold a Ph.D. for at most a dozen years or so, whom you would classify as the most successful? We would like to contact them in the same fashion we are contacting you, informing your indication (if it is ok for you) and asking the same question. Thanks in advance, Prof. Arilo Claudio Dias Neto, UFAM Prof. Leonardo Murta, UFF Prof. Márcio Barros, UNIRIO Prof. Rafael Prikladnicki, PUCRS*

---

**Q2: How can we point out the most prominent Brazilian young researchers in software engineering?**

The second question is tougher to answer because it depends on perception. People usually have a partial and obstructed vision about the field, intended or not. Due to that, we could not rely on a single, individual opinion. To attenuate this problem, we adopted snowball sampling [3], an iterative process based on individual perceptions to build a collective picture of the field. This process consisted of asking a researcher to name up to five young researchers in software engineering that he/she considered the most successful. For each nominee, we recursively asked the same question. The stop condition for this recursive process was the absence of answers or having all answers to already cited researchers. Each nominated researcher received the following e-mail (in Portuguese):

In this study, we did not involve senior researchers in the research identification process. We understand that to be regarded as a young prominent researcher within the Brazilian software engineering community, a researcher should also have the recognition of the older members. However, we followed a recursive process involving only young researchers, as described above. The participation of senior researchers in this indication process can be performed in the future as an evolution of the results described in this work.

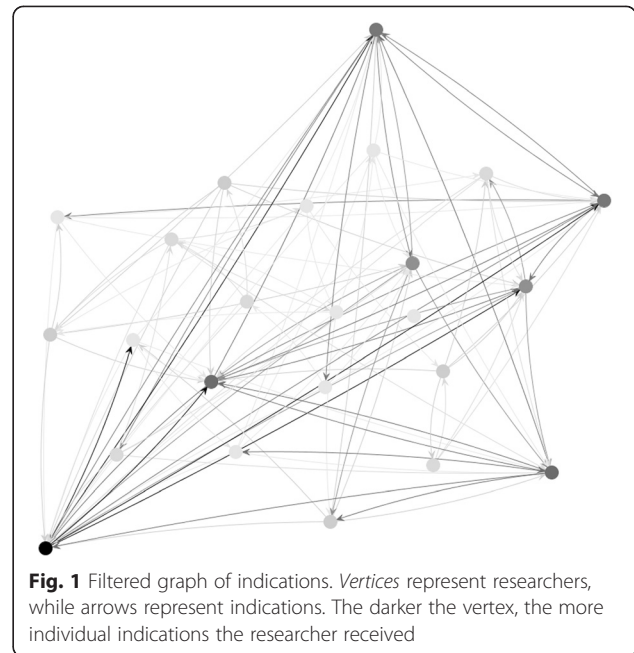
In order to start the recursive process, we selected one Brazilian researcher that is actively participating in the Brazilian and international software engineering communities, leading to a representative set of seeding indications. He already published full papers at ICSE and SBES, was member of the ICSE and SBES program committee, and chaired important software engineering conferences. He was independently and unanimously selected by all authors of this paper. At that moment, he qualified as “young” and was actually indicated more than once by other researchers from our pool. The process was considered finished when no new researcher was indicated and after 15 days without further answers.

**Collected data**

We ran the snowball sampling from January to March 2013 following the process described in “Methods” section. This process resulted in a graph of indications, depicting researchers as vertices and indications as edges. We received a total of 30 answers, summing up 144 indications

**Table 1** Distribution of researchers according to the number of indications

Number of indications	2	3	4	9	11	12	21
Number of researchers	8	5	4	2	2	2	1



(edges in the graph) to 35 researchers (vertices in the graph), totaling a response rate about 85 %.

The resulting graph was filtered to eliminate researchers who do not work with software engineering (2) or that received a single indication (9), ending up with 24 young researchers in software engineering. Seven (7) out of the 24 researchers received nine or more indications and the most cited researcher stands clearly apart, having received 21 individual indications (see Table 1). Due to that, we grouped researchers into two groups: 7 researchers who received nine or more indications and 17 researchers who received up to 4 indications.

Figure 1 shows the filtered graph of indications using a grayscale visualization to highlight the young researchers who were indicated often (darker gray). Despite of presenting only researchers with at least two indications, notice that some vertices have a single incoming arrow. This represents a situation in which the related researcher has received at least one indication from a researcher that received a single indication and, thus, does not participate in the graph.

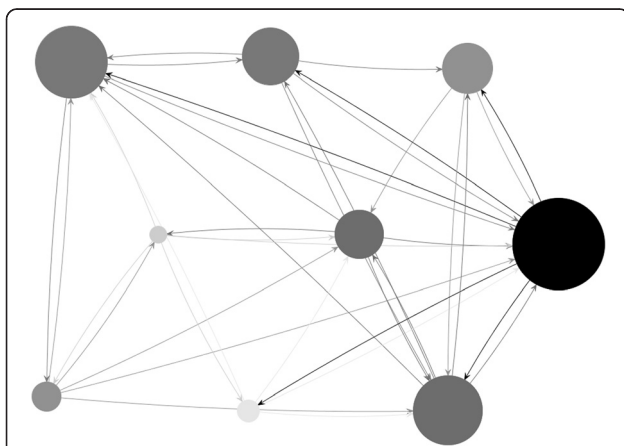
It is interesting to notice that the seven most cited researchers compose alone a fully connected graph (all vertices can be reached from any given one), denoting mutual recognition of the other researcher’s contribution to the field. These researchers alone received 85 indications (59 % of the total number of indications).

It is also interesting to observe mutual indications, that is, pairs of vertices (A, B) which present an edge from A to B and another in the opposite direction. We have accounted 21 mutual indications (42 out of 144 indications), involving 21 different researchers. All seven

researchers included in the topmost group on regard of indications participate in mutual indications. In fact, 9 out of the 21 mutual indications occur exclusively among these top researchers.

This first analysis grouped the researchers in terms of received indications, which can be seen as *in-degree* measure of graph centrality. This measure indicates the popularity of a researcher among his/her peers. However, other more elaborate centrality measures can be used to help analyzing the importance of a researcher in the whole graph of indications. We performed a similar analysis using *eigenvector centrality measure* [4] to capture the importance of researchers. This measure assigns a relative score to a researcher according to the scores of researchers who indicated him/her. In other words, if important researchers indicate a researcher, the *eigenvector score* of the indicated researcher increases. This way, indications from high-scoring researchers contribute more than indications from low-scoring researchers.

The computation of *eigenvector scores* considered the whole indication database (144 indications). This led to 35 young researchers, including those with one or more indications. Only 9 out of the 35 researchers achieved an eigenvector score above the mean ( $\mu = 0.0286$ ), denoting a strongly left-skewed distribution that concentrates most individuals below the mean and has a few individuals dragging the mean upwards. Figure 2 shows the filtered graph of indications considering only the researchers with *eigenvector score* above mean. As our former graph, the graph in Fig. 2 uses grayscale visualization to highlight the young researchers who were indicated more often (darker gray). Moreover, we used the size of the shape to indicate the *eigenvector score* (the bigger, the higher). It is possible to visually identify some correlation between color and size. In fact, the correlation between the in-degree and the *eigenvector score* is 0.94, which is extremely high.



**Fig. 2** Filtered graph of indications on which *vertices* represent researchers, *arrows* represent indications, and *vertex size* represents eigenvector centrality

The topmost group regarding *eigenvector score* (above mean) comprises all researchers who formed the topmost group regarding in-degree analysis. Six of them achieved an *eigenvector score* that exceeds the mean by at least one standard deviation ( $\sigma = 0.0426$ ). The last researcher from the topmost group formed in the in-degree analysis has an *eigenvector score* between  $\mu$  and  $\mu + \sigma$ , together with two researchers who were part of the second group regarding in-degree analysis. This shows a convergence among the indications of the most important researchers: in average 3.5 indications from researchers comprising the topmost group were given to other researchers pertaining to the same group.

This convergence of indications within the topmost group may be interpreted as the establishment of a very cohesive group with mutual favors or a natural effect of the relevance of the members belonging to that group. Aiming at providing a deeper understanding of indication concentrations in subgroups, we analyzed the formation of cliques in the graph. The classic clique definition applies only for undirected graphs. However, as our graph is directed, we adopted two alternative definitions for clique: (1) a *weak clique* in a directed graph  $G = (V, E)$  is a subset of the vertex set  $C \subseteq V$ , such that for every two vertices in  $C$  there exists some edge connecting the two, and (2) a *strong clique* in a directed graph  $G = (V, E)$  is a subset of the vertex set  $C \subseteq V$ , such that for every two vertices in  $C$  there exists two edges connecting the two in both directions.

With these definitions in hand, we searched for all maximal weak cliques in the graph, which are cliques that cannot be extended by including one more adjacent vertex. We could find 35 maximal weak cliques, which represent the bigger groups in which each member recognized or was recognized as relevant by all other members. Out of these 35 maximal weak cliques, only 1 clique has only members of the topmost group. On the other hand, 4 cliques do not have any member of the topmost group and all other 30 cliques have a mixture of both, in different distributions (16 cliques have more members from the topmost group, 7 cliques have less members from the topmost group, and 7 are evenly distributed). It is worth to mention that the three maximum weak cliques, which are cliques of the largest possible size in the graph, have 6 members that belong to both groups. One of these maximum weak cliques has 5 members from the topmost group and the other two have 4 members from the topmost group. This shows that, although there is a natural concentration of indications to researchers belonging to the topmost group, there is mutual recognition between both groups.

We could also find 17 maximal strong cliques, which represent cohesive groups of recognition, where all members of the group recognized all other members as

relevant. Out of these 17 maximal strong cliques, 6 cliques have only members from the topmost group. On the other hand, 6 cliques do not have any member from the topmost group, and the remaining 5 are evenly distributed. We could find three maximum strong cliques, with 3 members each. All these maximum cliques have only members from the topmost group. Similarly to the previous analysis, we observe that there is mutual recognition between both groups. Moreover, the large number of maximal and maximum cliques with few members in both analyses indicates that there is no individual concentration of indications. However, there are multiple localized concentrations, especially with members of the topmost group, as should be expected.

## Results and discussion

After collecting indications from peer researchers, we enriched the data with contextual information about the researchers: the gender, the year they concluded their Ph.D., institution and region/state where they did their Ph.D., institution and region/state where they currently work, software engineering areas of interest, citations record, the main venues where they have been publishing their works, and joint publications among the selected young researchers. These data were used in the analysis presented in the next subsections.

### Gender

From the 24 indicated young researchers, 22 are men and only 2 are women. This data indicates that there are few young women involved with software engineering research. This is also the trend if we analyze the supervisors of all indicated young researchers. From the 17 different supervisors, only 3 are women. This is at the

same time a challenge and an opportunity. The challenge is how to motivate more women to study computer science and, specifically, software engineering. Some women are reluctant to break into what many consider a man's field, creating a gender stereotype for the area. However, the opportunity is that more employers are eager to diversify their tech departments, and this generates opportunities for women. Some say that there has never been a better time for women to enter computer science and consequently software engineering. But unfortunately, as for now, our data also follow the trend of having few women in the area.

### Ph.D. universities and regions

Table 2 presents the universities where the most cited young Brazilian software engineering researchers have finished their Ph.D. The "Count" column depicts the number of researchers who have finished their Ph.D. on each institution, while the "Indications" column presents the total number of indications received by researchers who have concluded their Ph.D. on that institution. The "Count topmost" column depicts the number of researchers in the topmost group of prominent young software engineering researchers (the ones that received nine or more indications) that have finished their Ph.D. at that particular university.

As can be observed on Table 2, no university can be sought as a single source of successful young software engineering researchers—they are spread among a series of different universities. It is interesting to notice that a large part of these researchers made their Ph.D. in Brazilian universities. On the other hand, we still observe a strong concentration on universities residing on the southern part of the country, UFPE and UFCG being the exceptions to the rule.

**Table 2** Distribution of the selected Brazilian young software engineering researchers by the university where they have finished their Ph.D.

Ph.D. university	State/Country	CAPES's classification	Count	Indications	Count (topmost)
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)	RJ/Brazil	7	5	38	2
Federal University of Pernambuco (UFPE)	PE/Brazil	6	4	36	3
Federal University of Rio de Janeiro (UFRJ/COPPE)	RJ/Brazil	7	4	22	1
University of São Paulo (USP São Carlos)	SP/Brazil	6	4	11	-
Pontifical Catholic University of Rio Grande do Sul (PUCRS)	RS/Brazil	5	1	9	1
University of Campinas (UNICAMP)	SP/Brazil	7	1	4	-
University of Southampton	UK	ND <sup>a</sup>	1	3	-
University of California, Irvine	USA	ND <sup>a</sup>	1	3	-
Federal University of Campinas Grande (UFCG)	PB/Brazil	4	1	2	-
Lancaster University	UK	ND <sup>a</sup>	1	2	-
University of Illinois at Urbana-Champaign	USA	ND <sup>a</sup>	1	2	-

<sup>a</sup>ND means "Not Defined," because CAPES assesses only Brazilian Graduate Programs

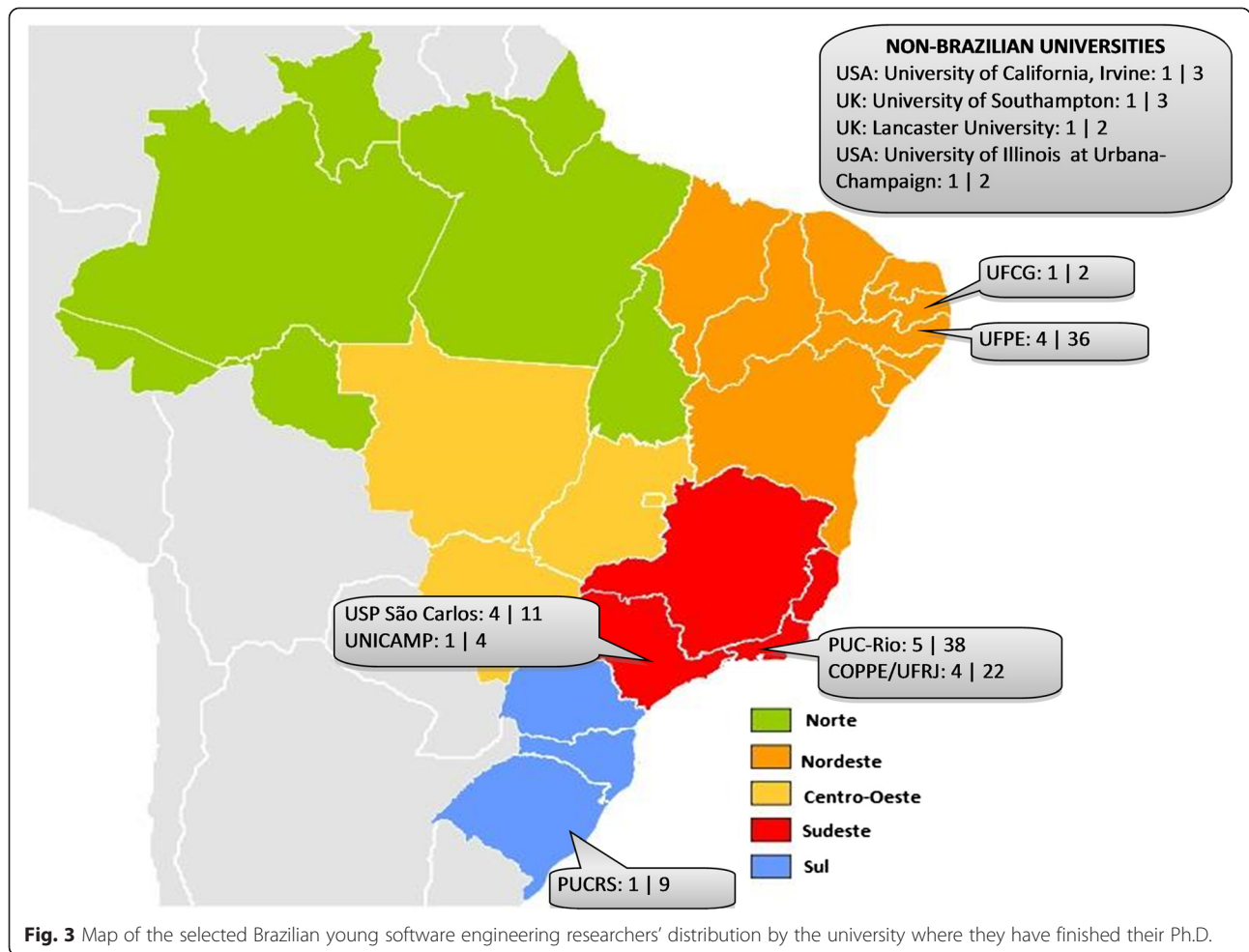


Figure 3 shows the distribution of researchers per Brazilian state or country where resides the university where they finished their Ph.D. We observe a concentration in seven (7) Brazilian universities, located in five (5) states, conveying 80 % of the selected researchers. Just few researchers (4) finished their Ph.D. in non-Brazilian universities, being two from the USA and two from UK.

Analyzing Table 2 according to the Graduate Programs' assessment system proposed by CAPES, all universities that have produced the young researchers who comprise the topmost group (those having 9 or more indications) are excellence graduate programs in Brazil, being classified in the highest levels (5, 6, and 7<sup>2</sup>) of the CAPES system.

Young researchers in the second group (those having up to 4 indications) were formed, mainly, in foreign universities, what may suggest that being abroad may have influenced in their visibility in the young researchers' community, since they might not have participated in local conferences and, thus, may have experienced difficulties in creating a research network with other Brazilian young researchers. Thus, it seems that someone does not have to

leave the country to do his Ph.D. to be considered prominent. Several Brazilian universities offer high-quality Ph.D. programs and given the investments in this area, one may expect a continuous increase in the quality standard of current and new programs.

Another analysis that we have carried out was regarding the profile of the Ph.D. supervisor. Six different professors

**Table 3** Distribution of selected Brazilian young software engineering researchers by the university where they did their sandwich

Sandwich university	Country	Count	Indications
University of California, Irvine	USA	2	13
Universität Mannheim	Germany	1	12
University of Waterloo	Canada	1	11
University of Victoria	Canada	1	9
Massachusetts Institute of Technology	USA	1	9
Newcastle University	UK	1	4
University of Texas at Dallas	USA	1	2
Lancaster University	UK	1	2

supervised the seven young researchers from the topmost group, where four (67 %) of them are Level 1 CNPq<sup>3</sup> researchers (one of them is level 1A and 3 of them are level 1D).

We also analyzed the influence of developing part of the Ph.D. course abroad, known as sandwich. Among the 20 young researchers who did their Ph.D. in Brazil, 9 (45 %) went to a sandwich period abroad. Table 3 presents the universities where the most cited young Brazilian software engineering researchers went for their sandwiches. The “Count” column depicts the number of researchers who have done their sandwiches on each institution, while the “Indications” column presents the total number of indications received by researchers who have done their sandwiches on that institution. We observe that the percentage of researches in topmost group that stayed abroad for a sandwich period during their Ph.D. is significantly higher (71 %). This clearly indicates that the sandwich

**Table 4** Distribution of selected Brazilian young software engineering researchers by the university where they currently work

University	State/ Country	Count	Indications
Federal University of Pernambuco (UFPE)	PE/Brazil	3	18
Federal University of Amazonas (UFAM)	AM/Brazil	3	11
Federal University of Bahia (UFBA)	BA/Brazil	2	14
Federal University of Rio Grande do Norte (UFRN)	RN/Brazil	2	13
Federal University of Goiás (UFG)	GO/Brazil	2	4
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)	RJ/Brazil	1	21
Fluminense Federal University (UFF)	RJ/Brazil	1	11
Federal University of Campinas Grande (UFMG)	PB/Brazil	1	9
Pontifical Catholic University of Rio Grande do Sul (PUCRS)	RS/Brazil	1	9
University of São Paulo (USP São Carlos)	SP/Brazil	1	4
Federal University of Rio de Janeiro State (UNIRIO)	RJ/Brazil	1	3
University of Brasília (UnB)	DF/Brazil	1	3
Federal University of São Carlos (UFSCAR)	SP/Brazil	1	3
Federal University of Pará (UFPA)	PA/Brazil	1	3
University of São Paulo (USP São Paulo)	SP/Brazil	1	2
Federal University of Minas Gerais (UFMG)	MG/Brazil	1	2
Federal University of São Paulo (UNIFESP)	SP/Brazil	1	2

period has a positive influence in the career of the young researchers.

#### Working universities and regions

Table 4 follows the same structure of Table 2 and presents the universities where the most cited Brazilian young software engineering researchers currently work (data for December 2013). As in Table 2, universities are sorted primarily by the number of researchers they have hired and secondly by the number of indications received by these researchers.

Most of the indicated young researchers (even those receiving fewer indications) already established themselves in new universities and formed their own research groups. Figure 4 depicts the information provided in Table 4 in the Brazilian map. It is interesting to observe that prominent researchers are more dispersed throughout the country while performing their jobs than while developing their Ph.D. theses. This is expected, since many of these researchers have left their families and hometowns to perform their Ph.D., returning home after graduating. Some of them were already professors in universities that gave them grants and due licenses for completing their Ph.D. in other universities located in other states or abroad. All indicated researchers are working in Brazilian universities; most of them are working in public (federal or state) ones. In fact, only two universities in this list are private (PUC-Rio and PUCRS).

These prominent researchers are also distributed in all five Brazilian regions. This result indicates that research on software engineering in Brazil is being renewed and carried on in locations formerly lacking of professionals working on the field. The prominent researchers are currently working in 12 different states (that is, 44 % of Brazilian states have prominent SE researchers). The states with more researchers from our pool are São Paulo (4), Rio de Janeiro (3), Pernambuco (3), and Amazonas (3). Regarding these states, only Amazonas has not formed prominent researchers. Pernambuco, Rio de Janeiro, and São Paulo have already consolidated and traditional software engineering groups in Ph.D. programs.

#### Ph.D. conclusion year

Figure 5 depicts the number of researchers and indications by year of Ph.D. conclusion. As can be observed, we lack prominent researchers in 2003 and 2012. In 2007 and 2009, it is possible to observe highest number of prominent researchers (5 each one), followed by 2004 (4), and 2006 and 2008 (3 researchers each). However, we can observe that from 2004 to 2007 fewer prominent researchers concentrate most indications (7.38 indications per researcher). On the other hand, it is possible to notice a significant drop in the number of indications from 2008 to 2011 (3.27 indications per researcher).

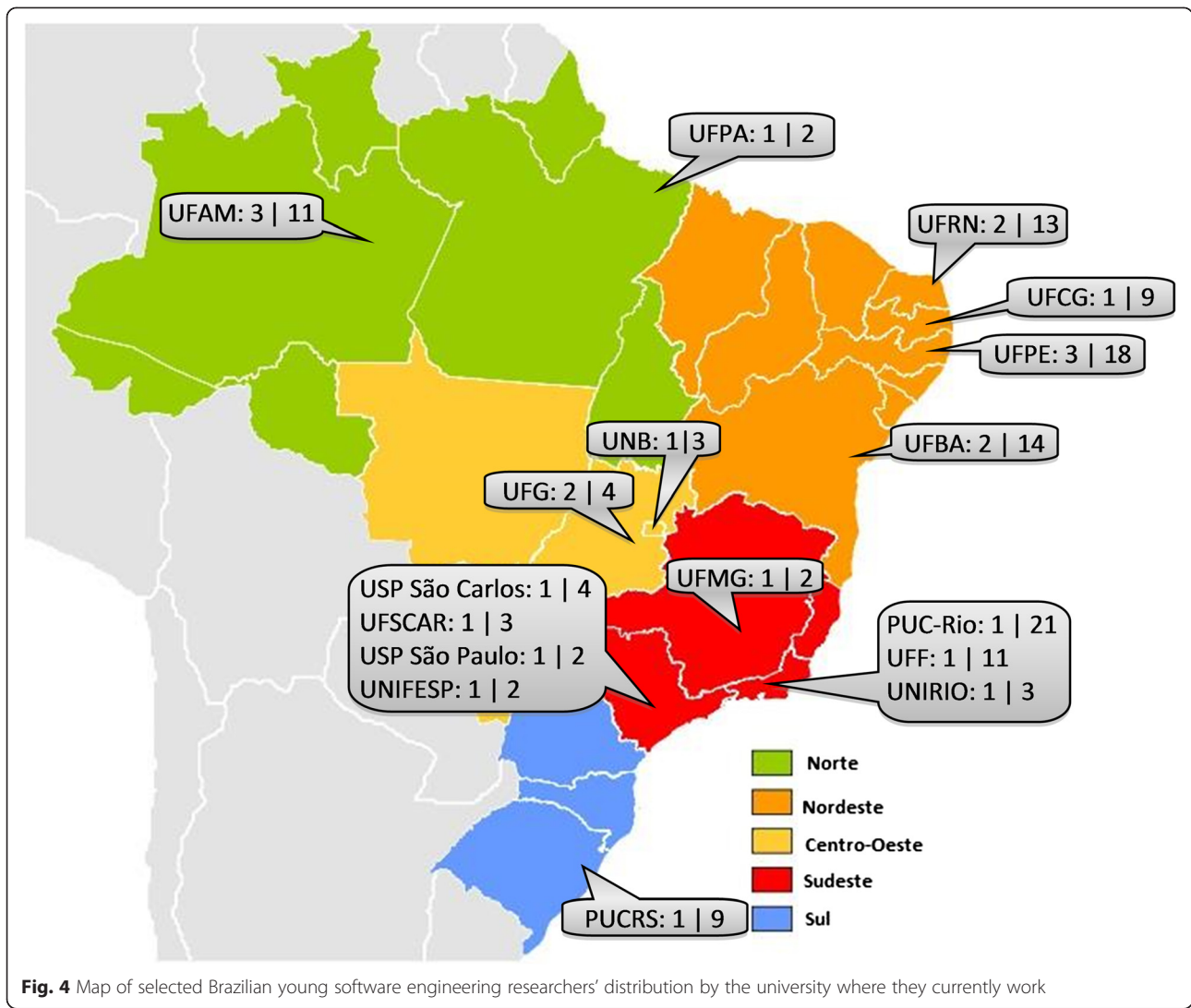


Fig. 4 Map of selected Brazilian young software engineering researchers' distribution by the university where they currently work

Moreover, six out of the seven topmost researchers finished their Ph.D. from 2004 to 2007.

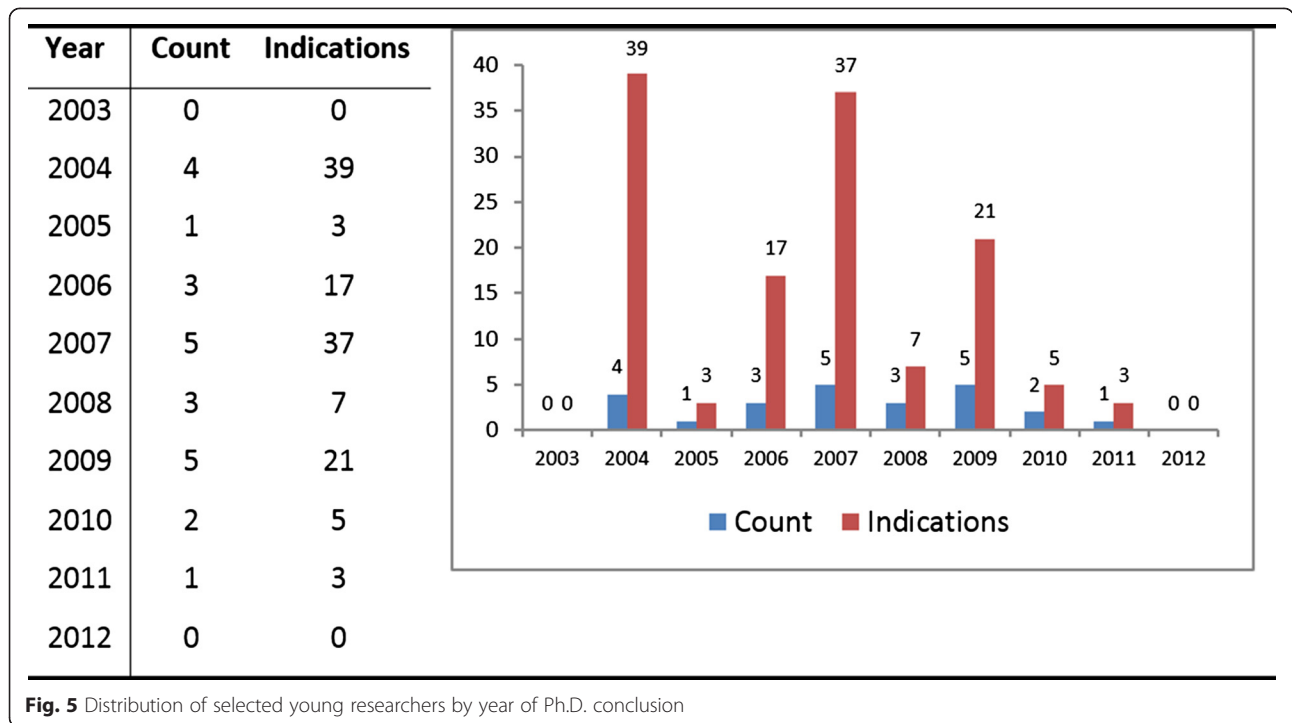
**Software engineering research area**

We have used information from Lattes CV and Google Scholar to identify the research areas addressed by the selected researchers. In these systems, a researcher may select a set of areas on which he/she works. We have collected data on December 2013. By then, only 22 out of the 24 selected researchers had a scholar profile but all had a Lattes CV. We have filtered the extensive list of research areas to those selected by at least two researchers from our group. Each researcher was associated with up to five research areas. Table 5 presents both the number of researchers working on each area and the number of indications they have received. Areas are sorted according to the number of researchers.

Most prominent Brazilian young researchers are working with software testing (ST) and aspect-oriented development (AOD) and this is a very connected group in terms of indications. They have regular workshops, which have been running for 7 years (in 2013) and are co-located with CBSOft<sup>4</sup> in both areas (SAST in ST and WMod<sup>5</sup> in AOD) from their very start.

It is also noticeable the interest on empirical studies in software engineering, being dealt by researchers from different groups. Programming languages, software product line, software maintenance, software design, search-based software engineering, and source-code analysis are also prominent research areas, evidencing a trend of interest towards low-level (design and code) concerns in software engineering. Agile development methodologies, model-driven development, fault tolerance, human-computer interfaces, and software process/metrics are also noticeable areas.





**Fig. 5** Distribution of selected young researchers by year of Ph.D. conclusion

**Citations**

Besides collecting research area data, we have also collected citation data from Google Scholar. As reported in “Ph.D. conclusion year” section, by December 2013, 22 out of the

24 selected researchers had a Google Scholar profile. For those, we have collected six different pieces of information:

**Table 5** Research areas within software engineering in which the selected researchers perform their work

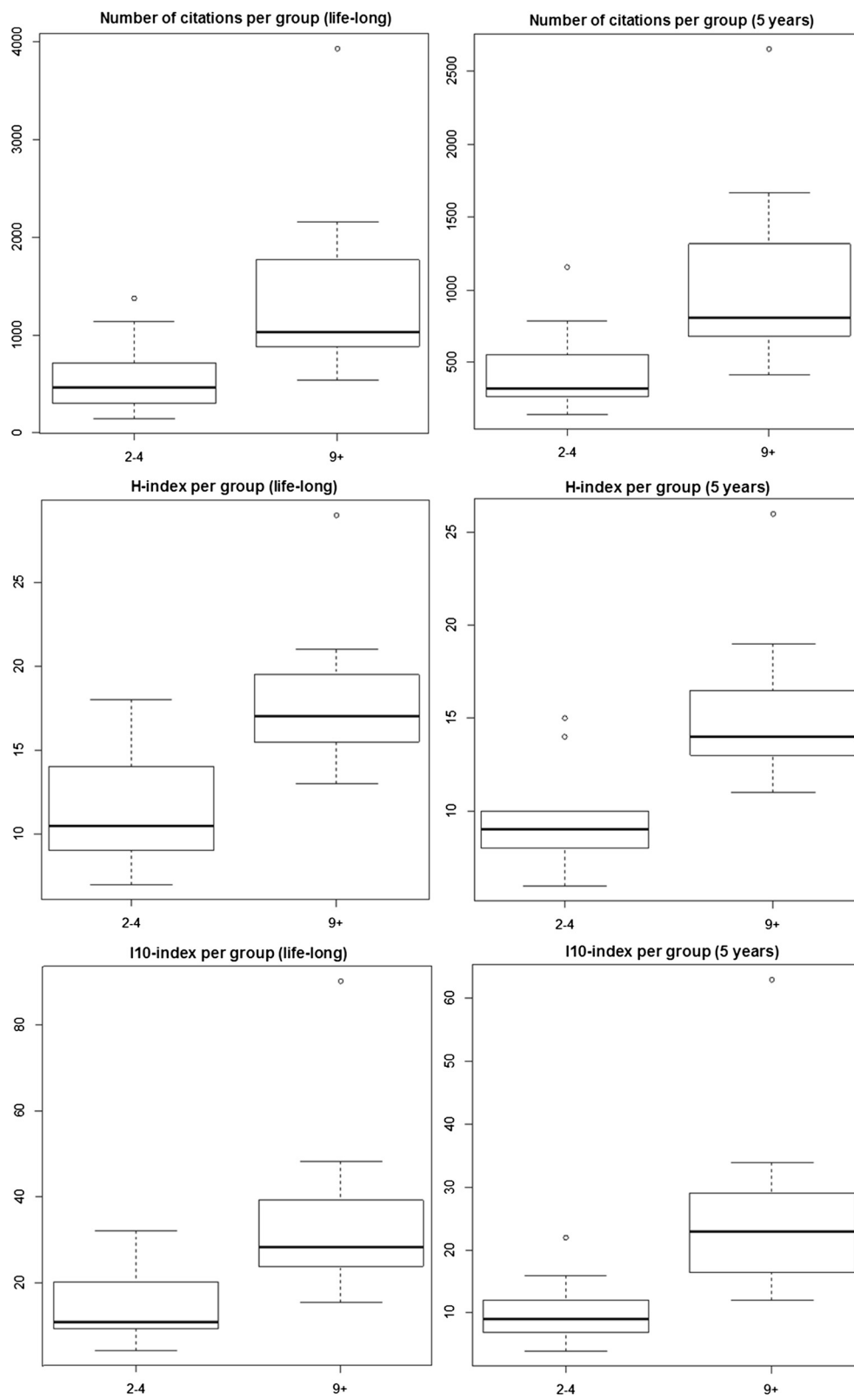
Software engineering research area	Acronym	Count	Indication
Software testing	ST	10	42
Aspect-oriented development	AOD	9	58
Empirical software engineering	ESE	7	35
Programming languages	PL	5	43
Software quality	SQ	5	22
Software product lines	SPL	4	44
Software maintenance	SM	4	27
Software design	SD	4	26
Search-based software engineering	SBSE	4	19
Source code analysis	SCA	4	15
Agile development methodology	ADM	3	15
Model-driven development	MDD	3	14
Fault tolerance	FT	2	25
Human-computer interfaces	HCI	2	7
Software process	SP	2	7
Software metrics	SME	2	4

- The number of citations identified by Google Scholar for all publications of the researcher, both for all times ( $C_{ALL}$ ) and for the last 5 years ( $C_{5Y}$ );
- The H-index, which is the largest integer number  $h$  such that  $h$  publications of the researcher have at least  $h$  citations, both for all times ( $H_{ALL}$ ) and for the last 5 years ( $H_{5Y}$ );
- The I10-index, which is the number of publications from the researcher that have at least 10 citations, both for all times ( $I10_{ALL}$ ) and for the last 5 years ( $I10_{5Y}$ ).

Table 6 presents descriptive statistics for these indicators. It is interesting to observe that the mean is always larger than the median (41 % larger for the number of citations, 9 % larger for the H-index, and 29 % larger for I10). This indicates a right-skewed distribution, on which few individuals having high values for the measures

**Table 6** Descriptive statistics for citation data

	C-All	C-5y	H-All	H-5y	I10-All	I10-5y
Minimum	177	173	7	7	4	4
Median	630	556	14	10	19	12
Mean	956.6	719.6	14.2	12.1	22.8	17.2
Standard deviation	893.7	623.5	5.3	4.6	19.8	14.4
Maximum	4129	2836	30	27	94	69



**Fig. 6** Box-plots of citation indexes received by young researchers

**Table 7** Correlation between Google Scholar indexes and the number of indications received by the selected researchers and their eigenvector scores

	C-All	C-5y	H-All	H-5y	I10-All	I10-5y
Number of indications	0.36	0.43	0.41	0.40	0.42	0.43
Eigenvector centrality	0.75	0.81	0.69	0.81	0.70	0.82

under interest drag the mean towards the distribution’s upper extreme, leaving most individuals below the mean (for instance, 63 % of the selected researchers are below the mean for the life-long citation count measure). High standard deviations also denote disperse data distributions.

Figure 6 presents box-plots for the distribution of the aforementioned citation indexes. The charts are plotted separately for researchers having 9 or more indications and researchers with 2–4 indications. Citation, H-index, and I10-index seem to be good proxies for reputation, though at least one highly indicated researcher fell below the inter-quartile range for all charts.

We have used the non-parametric *Spearman rank-order correlation index* to calculate the correlation between these indexes and the number of indications received by a researcher. We have also calculated correlations between indexes and *eigenvector scores*. Table 7 presents the results of our calculations.

Surprisingly, none of the indexes presented strong correlation with the number of indications. This result was not expected because it is generally accepted that the number of citations is a good predictor to the quality of research work and, thus, prominent researchers might be those with a strong number of citations. On the other hand, correlations with *eigenvector scores* are higher and

more significant than correlations based on the number of indications. We observe strong correlation between the centrality score and citations indexes, especially those collected from 2009 to 2013. Therefore, *eigenvector score* seems to be a better predictor of researcher influence than the number of indications itself.

As would be expected for young researchers, there is high correlation between any index for the lifetime and its counterpart for 5 years. This correlation is 0.97 for the number of citations, 0.93 for H-index, and 0.93 for the I10-index. It is also interesting to notice that all top-most researchers receive productivity grant from CNPq (one level 1D and all other level 2, by 2013). However, only 4 researchers out of the remaining 17 (23 %) have productivity grants from CNPq (all level 2).

**Venues**

Another interesting aspect to analyze is the venues that successful young researchers usually publish the results of their work. We used DBLP<sup>6</sup> to collect this information for each young researcher of the selected group. Aiming at differencing full and short papers in venues that accept papers with different number of pages, we considered as full paper all papers with at least eight pages. Table 8 shows the most popular venues ordered by the percentage of young researchers who published at least one full paper in the venue. We also present information regarding the classification of all venues according to the QUALIS system.<sup>7</sup> We observe a tendency to prioritize well-established conferences and journals.

As expected, the SBES, the main Brazilian conference in this field, appears in the first position: 80 % of the

**Table 8** Venues where young researchers published full papers

Conferences	QUALIS	Support	Journals	QUALIS	Support
SBES	B2	80 %	Journal of the Brazilian Computer Society	B2	52 %
ICSE	A1	56 %	Journal of Systems and Software	A2	48 %
SEKE	B1	52 %	Information and Software Technology	A2	36 %
SAC	A1	40 %	ACM Software Engineering Notes	B4	32 %
OOPSLA	A1	36 %	Software: Practice and Experience	A2	32 %
SBCARS	B4	32 %	Journal of Universal Computer Science	B1	32 %
SPLC	A2	28 %	CLEI Electronic Journal	B5	24 %
ECOOP	A2	28 %	IEEE Software	A1	24 %
EASE	B4	28 %	Electronic Notes in Theoretical CS	B1	24 %
ECSA	B3	28 %	IEEE Transaction on Software Engineering	A1	20 %
ICSME	A2	28 %			
CSMR	A2	28 %			
AOSD	A1	24 %			
ASE	A1	24 %			
ICSR	B1	24 %			

young researchers have published in the conference. The ICSE, the main international conference, appears in the second position (56 %), indicating that the young researchers have been concerned with giving international visibility to their results.

Other general conferences, such as the International Conference on Software Engineering and Knowledge Engineering (SEKE) and the Symposium on Applied Computing (SAC), appear in this list with high popularity among young researchers. Some field-specific conferences also present high popularity, with special emphasis to the Conference on Object-Oriented Programming, Systems, Languages and Applications (OOPSLA) and Brazilian Symposium on Software Components, Architectures and Reuse (SBCARS), with more than 30 % of popularity. Only three venues not classified among the QUALIS's top 4 strata (A1 to B2) are present in the results: SBCARS, EASE, and ECSA.

Regarding journals, the Journal of the Brazilian Computer Society (JBCS) appears in the first position, with 52 % popularity. However, it is still less popular than SBES, ICSE, and SEKE among young researchers. The Journal of Systems and Software appears in the second position, with 48 %. Both can be justified because in recent years, SBES has selected the distinguished papers to submit extended versions to these journals (JBCS and JSS). The same happened with CLEI Electronic Journal, which besides having low impact according to QUALIS's system has published extended versions of distinguished papers from some South American conferences.

It is interesting to notice that some of the most prestigious journals in software engineering do not appear in the list, such as empirical software engineering and ACM Transactions on Software Engineering and Methodology (TOSEM). Moreover, one of the most prestigious journals in software engineering, IEEE Transaction on Software Engineering, does appear in the list, but in the last position.

Analyzing the distribution of published papers by the selected young researchers according to QUALIS levels<sup>8</sup> (see Table 9), we observe a high concentration of papers in the top strata of the QUALIS system (from level A1 to B1), particularly for journals (58 %). A smaller percentage of papers published in top strata conferences (32 %) may be due to constant participation on local

workshops and theme-focused conferences, which do not usually attain high levels in the QUALIS system despite of their importance to build local research communities and develop international relationships, respectively. Overall, this result suggests that the selected researchers are concerned to publish the results of their research on venues with high impact in the scientific community. Given that the most traditional Brazilian venue for software engineering papers currently attains level B2, these results also suggest a focus in international venues.

Thus, a young research aspiring for prominence and recognition in the Brazilian software engineering community should keep the old adage and try publishing papers on the best conferences and journals. Those venues increase the visibility for the researcher's work and, consequently, the number of citations. Citations correlate well with recognition by top researchers (as shown in Table 7).

We have also computed the number of points obtained by each young researchers according to the QUALIS system, where papers have weights according to the classification of the venue where they were published (A1 = 1.0; A2 = 0.85; B1 = 0.7; B2 = 0.5; B3 = 0.2; B4 = 0.1; B5 = 0.05). Table 10 presents descriptive statistics for this indicator for three groups: (1) all researchers; (2) researchers with 2–4 indications; and (3) researchers with 9 or more indications. It is interesting to observe that summary data for researchers from the topmost group present a clear difference if compared to researchers with 2–4 indications, suggesting an association between the visibility of these researchers (number of indications received from other researchers) and the number of points obtained by them according to the QUALIS system (mild Spearman rank-order correlation of 0.58). For instance, the minimum value obtained by researchers on this group is larger than the mean/median obtained by researchers with 2–4 indications.

Figure 7 presents box-plots for the distribution of the aforementioned papers points. The charts are plotted separately for researchers having 9 or more indications and researchers with 2–4 indications.

**Publication history**

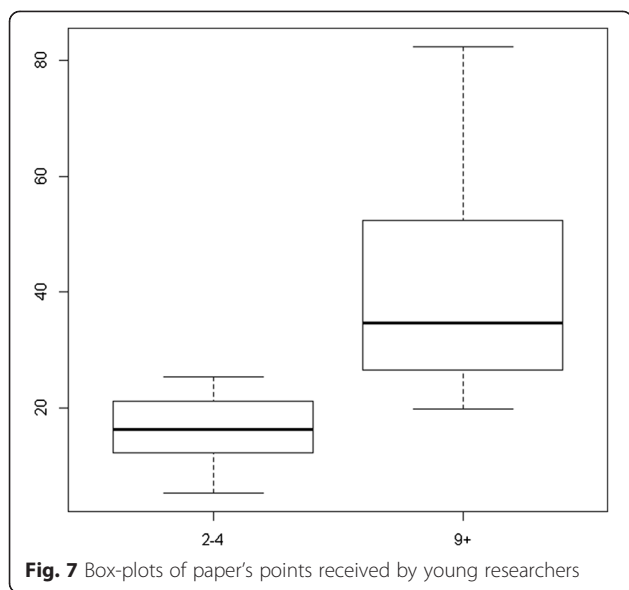
We also collected the sequence of publications for each young researcher from DBLP. Each sequence contains

**Table 9** Distribution of papers published by young researchers according to CAPES QUALIS system's levels

Venue		A1	A2	B1	B2	B3	B4	B5
Conferences	Count	86	90	140	202	160	211	99
	%	8.7	9.1	14.2	20.4	16.2	21.4	10.0
Journals	Count	24	65	43	37	18	18	22
	%	10.6	28.6	18.9	16.3	7.9	7.9	9.7

**Table 10** Descriptive statistics for citation data

	QUALIS All	QUALIS 2:4	QUALIS +9
Minimum	5.2	5.2	19.8
Median	19.9	16.1	34.7
Mean	24.2	16.3	42.1
Standard deviation	17.56	5.52	22.6
Maximum	82.5	25.3	82.5



all publications of a given young researcher grouped by year. These sequences together built a database used as input for sequence mining, with support at 24 %. This threshold was selected because it produced a comprehensive, though not too extensive, list of sequences. As a result, we could observe some publication patterns among these successful young researchers, as shown in Table 11. In this table, we present sequential patterns of size two, in the form  $A \rightarrow B$ , together with three measures: support ( $s\%$ ), confidence ( $c\%$ ), and lift ( $L$ ). This indicates that the pattern of publishing at  $B$  after publishing at  $A$  occurred for  $s\%$  of the young researchers. Moreover,  $c\%$  of the young researchers who published at  $A$  also published at  $B$  in the future. Finally, publishing at  $A$  increased in the frequency of publishing at  $B$  by  $L$ .

For example, the first pattern shown in Table 11 indicates that 48 % of the young researchers published two (or more) papers at SBES in different years (as shown in the *support* column). At a first glance, this seems to be a positive pattern. However, we can observe that, among the young researchers that published papers at SBES (80 %), only 60 % ( $48 \div 80$  %) of them published again a paper at SBES in the future (as shown in the *confidence* column). This way, this pattern is in fact negative, because publishing a paper at SBES decreased the frequency of publishing a new paper at SBES from 80 (see Table 8) to 60 % (or about 25 %, as shown in the *lift* column). The second, third, and fourth patterns are different. They show that 44 % of the young researchers published a paper at a new venue (JBCS, SEKE, and ICSE, respectively) after publishing a paper at SBES. Moreover, for the second and third patterns, publishing a paper at SBES increased in 6 % the frequency of publishing at JBCS and SEKE, respectively.

**Table 11** Venues sequence followed by young researchers

Number	Venue sequence	Support	Confidence	Lift
1	SBES → SBES	48 %	60 %	0.75
2	SBES → JBCS	44 %	55 %	1.06
3	SBES → SEKE	44 %	55 %	1.06
4	SBES → ICSE	44 %	55 %	0.98
5	JSS → SBES	40 %	83 %	1.04
6	JSS → SBCARS	32 %	67 %	2.08
7	ICSE → JBCS	32 %	57 %	1.10
8	ICSE → ICSE	32 %	57 %	1.02
9	EASE → SBES	28 %	100 %	1.25
10	SBCARS → SBCARS	28 %	88 %	2.73
11	SBES + SBES → ICSE	28 %	70 %	1.25
12	JSS → JUCS	28 %	58 %	1.82
13	JSS → JBCS	28 %	58 %	1.12
14	JSS → SEKE	28 %	58 %	1.12
15	JSS → ICSE	28 %	58 %	1.04
16	SEKE → JBCS	28 %	54 %	1.04
17	ICSE → ECOOP	28 %	50 %	1.79
18	ICSE → OOPSLA	28 %	50 %	1.39
19	ICSE → SEKE	28 %	50 %	0.96
20	ICSE → SBES	28 %	50 %	0.63
21	CSMR → SBCARS	24 %	86 %	2.68
22	SBCARS → JUCS	24 %	75 %	2.34
23	ACM SEN → JBCS	24 %	75 %	1.44
24	JUCS → JBCS	24 %	75 %	1.44
25	IST → SBES	24 %	67 %	0.83
26	JSS → ICSM	24 %	50 %	1.79
27	JSS → ECSA	24 %	50 %	1.79
28	JSS → OOPSLA	24 %	50 %	1.39
29	SEKE → SEKE	24 %	46 %	0.89

For the fourth pattern, publishing a paper at SBES decreased in just 2 % the frequency of publishing at ICSE. On the other hand, publishing two papers at the same edition of SBES (line 11) increased in 25 % the frequency of publishing at ICSE. Moreover, the 20th pattern says that 50 % of the researchers who published a paper at ICSE also published a paper at SBES in the future. Thus, publishing a paper at ICSE decreased in 37 % the frequency of publishing a paper at SBES, suggesting that after publishing a paper at ICSE, the focus of young researchers may become oriented towards international venues.

Other interesting results can be observed in Table 11. In the ninth sequential pattern, we can observe among the young researchers that published papers at EASE (28 %), all of them (100 %) published a paper at SBES in the future (as shown in the *confidence* column).

Moreover, publishing a paper at EASE increased in 25 % the frequency of publishing at SBES (as shown in the *lift* column). Analyzing the tenth sequential pattern, we observed that publishing a paper at SBCARS increased in 173 % the frequency of publishing a second paper at the same conference, indicating an active participation of a young researchers' subgroup (around 28 % of the indicated ones) in this conference. Moreover, publishing a paper in SBCARS increases in 134 % the frequency of publishing a paper at JUCS. As the best papers of SBCARS are usually invited to submit to JUCS special issues, this pattern is completely expected.

### Joint publication

Considering the selected researchers, we have identified each joint publication from DBLP, that is, a publication in which at least two young researchers could be counted as authors. In average, young researchers have published papers with five other researchers pertaining to the selected group. This is a very polarized sample, comprised of 10 researchers (including most of the topmost ones from our graph analysis) who have interacted with 9 or more researchers from the selected group, while the remaining has interacted with just a few (up to four) researchers. When considering only the researchers they have indicated, the selected researchers have published, in average, with two of their nominees.

We also calculated the correlation (*Spearman*) between the number of researchers in the group with which a given researcher has published and the number of indications received by the researcher. Data presented small correlation (0.06), which was again surprising since we would expect more indications from people that had interacted among themselves. Next, we analyzed the effects of participating and eventually meeting other researchers at conferences and producing joint research afterwards. We first crossed the data about the venues that each researcher published papers over the years and the venues where they published papers together, as co-authors. Then, we ran sequence mining over this information and observed a total of 60 joint publications, which represents 22 % of the pairwise combinations of researchers.

We have also observed that when two researchers had a joint publication, the chances of a second joint publication among them increased 3 times. Moreover, 21 % of the first joint publication among a pair of researchers occurred after they had independent papers published in the same edition of the same venue in the past. This percentage is not high, and actually the chances of publishing together are halved when researchers meet before in conferences. Thus, the reason for starting collaborations seems not to be related to meeting in conferences. Nevertheless, after a new collaboration is started and the

first publication together comes out, chances drastically increase for additional publications.

Finally, we investigated the level of dependence of the young researchers on their previous supervisors, through joint publications. For each researcher, we computed the percentage of publications with the previous supervisor and observed that the mean percentage of the topmost group is a bit higher (47 vs. 42 %), which is counterintuitive. Eleven out of the 24 young researchers have from 40 to 50 % of their publications also signed by their supervisor. The remaining researchers are close to uniformly distributed on both sides of this band: 6 out of 13 have from 20 to 40 % of their papers with their former supervisor, while 4 out of 13 have from 50 to 70 % of their publications on similar situations. Thus, there is no evidence that the selected young researchers are divided into two different groups according to their dependence on former supervisors.

We also calculated the correlation (*Spearman*) between the percentage of papers with the previous supervisor and the number of indications received by the researcher. Data presented a very small negative correlation ( $-0.02$ ), which was once again surprising since we would expect more indications to people that are independent of their previous supervisor.

Going in a different direction, we computed the correlation (*Spearman*) between the percentages of papers with the previous supervisor and holding or not productivity grant from CNPq. We observed a correlation of  $-0.20$ , which provide subtle evidence that the lesser the level of dependence on the previous supervisor, the higher the probability of having the aforementioned grant. Another interesting finding is that the young researchers who did their Ph.D. abroad have co-authored in average only 28 % of their publications with their former Ph.D. supervisors. Although this might show a sign of independence, none of these young researchers is part of the topmost group of young researchers.

### Threats to validity

No study is free of threats to its validity and the present work is not an exception. We have identified some concerns to the validity of our results, which we summarize in the following paragraphs along with the actions taken to prevent these issues to affect our observed results.

The structure of our data collection procedure may be the single most important source of threats to our observations. First, there is the issue of using a single seed to start the process, along with the selection of the seed research. While designing the present study, we have decided to build a closed graph of connections, on which every researcher indicating someone or being indicated by someone should comply with our definition of young Brazilian software engineering researcher.

Therefore, our seed researcher could not be a senior researcher or a foreigner young and prominent researcher. Thus, we decided to select a seed researcher who would probably end up being indicated as prominent. The selected seed was an independent and unanimous selection of all authors of this paper. In fact, the seed researcher became the most indicated researcher of our selected group.

Completeness is another source of validity threat. Not all researchers answered our call for indications and therefore we cannot expect that the present analysis encompass all possible young and prominent software engineering researchers. However, we believe the selected researchers form a representative group of the target population, especially considering the number of CNPq grants given to these researchers, the number of papers they have published, the venues that they have achieved their participation and organization of conferences, and other indicators of being active on research.

Finally, we must consider the possibility of research in the short-term future to become significantly different from present time's research. Such might render our analysis much less useful for the generations of researchers to come. Such a change was probably felt with the introduction of the Internet and the increased reach researchers have acquired to each other works a couple decades ago. New events might come to change the way research is performed and evaluated and an analysis such as the one proposed in this paper may become obsolete.

### Related work

Many papers evaluate software engineering research in a given context by means of examining conference and journal publications. DBLP [5] is an important and frequently adopted source of information to this end. Biryukov and Dong [6] have examined the average lifetime of researchers publishing in top computer science conferences and then investigated the career of long-living researchers with more than 10 years participating in such conferences. They discovered that most of these researchers have been engaged in two or more research areas and that most of their papers were published from 5 to 10 years of their first paper in a top venue.

Bird et al. [7] also used DBLP to extract a collaboration graph among researchers in computer science. They applied topology metrics over the extracted graph to identify how centralized, integrated, and cohesive the research areas are. Moreover, they observed how research areas change over time. They found some interesting patterns, such as that the overlap of researches working in both software engineering and database areas was high in the 80s (10 %) but drastically decreased to less than 1 % in the last decades. Currently, the area that overlaps the most with software engineering is programming languages (7 %).

Martins et al. [8] have evaluated the usefulness of journal-oriented publication quality metrics to address the importance of research conferences. They propose variants of the well-known impact factor metrics to consider the importance of longevity, size, periodicity, and prestige of conferences. They have shown the usefulness of the proposed metrics by comparing them to the opinion of a set of researchers on the importance of a set of computer science conferences, similarly to what we have done to capture the young and prominent researchers.

Elmacioglu and Lee [9] examined the characteristics of questionable and reputable computer science conferences on regard of their TPC. They found that reputable conferences tend to have a smaller TPC formed by active (having many publications) and prominent researchers when compared to questionable ones. As in our paper, prominence was measured using a graph centrality metric based on co-authorship data collected from ACM Digital Library.

In a more specific context, Silveira-Neto et al. [1] examined the 25-year history of the most important software engineering conference in Brazil (i.e., SBES), showing the main researchers who published in this conference, the most frequently addressed topics, most engaged universities, and the distribution of accepted publications throughout the extensive Brazilian territory. Besides evaluating the conference itself, the paper shows the evolution of software engineering research in Brazil, a proxy for the evolution of computer science research itself in the country, expanding from the coast-side of the country to its interior and decentralizing from a few think tanks for a broader and more distributed corpus of researchers.

Although conference and journal publications can help on evaluating software engineering research, this approach is not unanimous in the literature. Some researches argue that it suffers from a bias of evaluating research only by looking at academic contribution. For instance, Lionel Briand discussed in his keynote address at ICSM 2011 the importance of interacting with practitioners to bind research problem definition and solution evaluation to reality. Moreover, Bertrand Meyer added in his blog that difficulties of finding research agencies to support practical needs of software engineering researcher hinders building real software and limits academic research to demo versions. He correlates this characteristic with the limited contribution that academic research has given to software engineering evolution. Carlo Guezzi also observed, in a keynote speech at ICSE 2009, that industry has little participation in top software engineering conferences. He presented some limitations of citations as an instrument to measure the quality and impact of a given research: inability to capture indirect citations, different average citation numbers across different science branches, among others. Finally, David Rosenblum presented some concerns in his keynote address at APSEC 2012 about the under-

representation of some research areas within software engineering. He observed that, although the distribution of themes and subareas is very broad within the field, areas such as specification, testing, and debugging have considerable dominance. However, his conclusion is that the software engineering research area, as a whole, seems to be healthy.

Besides the existence of multiple papers analyzing the software engineering research community, we could not find papers with the same contribution of ours. Our paper has a specific focus on Brazilian young researchers. It is worth to notice that we have used DBLP as data source for some of our analysis, but we also extracted data from Google Scholar, Lattes CV, a survey with the researchers, and the researchers' home page. This information, altogether, helped us on better understanding of what a young and prominent Brazilian software engineering researcher looks like and how they behaved to reach their top position.

## Conclusions

In this paper, we report on the results of a study that identified prominent Brazilian young software engineering researchers by collecting peer indications from a seed researcher clearly recognized as a prominent representative of the field. We have identified 24 prominent young researchers and afterwards collected information about their formation, research work, the venues where they publish, and the collaborations among each other.

The results of this study revealed interesting information. For instance, we observed that the current generation of prominent researchers has graduated in the most important universities in Brazil and are still working in Brazilian institutions (most of them in federal universities). This indicates that there are Brazilian researchers able to produce new generations of researchers for the field. Analyzing publication patterns, we observed that prominent researchers target high-quality conferences and journals and usually collaborate strongly with a large number of peers. They also tend to establish themselves in a given research area and propose and develop workshops and conferences to promote the expansion of research on that area.

As future works, we believe that there are plenty of opportunities to evolve the analysis of this collected data. One of them is interviewing the topmost young researchers in order to extract some practical guidance for future researchers. Another opportunity is to replicate this analysis in the future and compare the results in order to evaluate how the area has evolved. Finally, we believe that the methodology and process presented in this paper could be replicated in other areas with the same type of analyses and findings.

## Endnotes

<sup>1</sup>Names used in this e-mail do not represent real researchers. They were collected from the Harry Potter series for the sake of example.

<sup>2</sup>Seven (7) is the maximum level in the scale proposed by CAPES to assess Graduate Programs.

<sup>3</sup>CNPq is the National Council of Technological and Scientific Development and has a system to recognize the most productive researchers in Brazil, in a ranking that has 5 levels: 1A (higher), 1B, 1C, 1D, and 2 (lower).

<sup>4</sup>Brazilian Conference on Software, where resides SBES, the most important Brazilian conference on software engineering.

<sup>5</sup>Former LA-WASP.

<sup>6</sup>DBLP Computer Science Bibliography: <http://www.informatik.uni-trier.de/~ley/db/>

<sup>7</sup>The QUALIS system, managed by CAPES, classifies the most important computer science conferences and journals according to a ranking system consisting of the 7 levels: A1 (higher), A2, B1, B2, B3, B4, and B5 (lower).

<sup>8</sup>Only papers published in venues classified by QUALIS system were considered in this analysis.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

LGPM and MOB conceived the objectives and methodology to conduct the research, already performed by them in another context, and supported the results analysis. ACDN is mainly responsible for conducting the survey, inviting the subjects, collecting the results, and also supporting the results analysis. RP worked closely supporting the survey execution and providing support for the interpretation of results obtained from the subjects. All authors contributed with important observations to technically improve the analysis and the presentation of the work. All authors reviewed the final version of paper and agreed with its submission to the *Journal of the Brazilian Computer Society*.

## Acknowledgements

This research is partially funded under the Brazilian Law 8.248/91. Funding also comes from the Rio de Janeiro, Rio Grande do Sul, and Amazonas States funding agencies (FAPERJ, FAPERGS, and FAPPEAM). Leonardo, Márcio, and Rafael thank CNPq for their research grants.

## Author details

<sup>1</sup>Instituto de Computação, Universidade Federal do Amazonas (UFAM), Manaus, AM, Brazil. <sup>2</sup>Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, RS, Brazil. <sup>3</sup>PPGI, Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, RJ, Brazil. <sup>4</sup>Instituto de Computação, Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil.

Received: 10 September 2014 Accepted: 6 August 2015

Published online: 26 August 2015

## References

1. Silveira-Neto PA, Gomes JS, Almeida ES, Leite JC, Batista TV, Leite L (2013) 25 years of software engineering in Brazil: beyond an insider's view. *J Syst Softw* 86(4):872–889. doi:10.1016/j.jss.2012.10.041, <http://dx.doi.org/10.1016/j.jss.2012.10.041>
2. Price E (2014) The NIPS Experiment. Available online at the following URL: <http://blog.mrtz.org/2014/12/15/the-nips-experiment.html>, last accessed in January/2015
3. Goodman LA (1961) Snowball sampling. *Ann Math Stat* 32:148–170



4. Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 2(1):113–120
5. Ley M (2009) DBLP—some lessons learned. Proceedings of the Very Large Database Conference, Lyon, pp 24–28
6. Biryukov M, Dong C (2010) “Analysis of Computer Science Communities Based on DBLP”, research and advanced technology for digital libraries. *Lect Notes Comput Sci* 6273:228–235
7. Bird C, Barr E, Nash A, Devanbu P, Filkov V, Su Z (2009) Structure and dynamics of research collaboration in computer science. Proc. of the 9th SIAM International Conference on Data Mining, USA, pp 826–837
8. Martins WS, Gonçalves MA, Laender AHF, Ziviani N (2010) Assessing the quality of scientific conferences based on bibliographic citations. *Scientometrics* 83:133–155
9. Elmacioglu E, Lee D (2009) Oracle, where shall i submit my papers? *Commun ACM* 52(2):115–118

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---