Scientometrics (2019) 118:93–108
https://doi.org/10.1007/s11192-018-2967-4

Co-authorship networks and research impact in large research facilities: benchmarking internal reports and bibliometric databases

Fabio S. V. Silva[1] · Peter A. Schulz[1] · Everard C. M. Noyons[2]

Received: 14 March 2018 / Published online: 26 November 2018

Abstract

In this paper, we address the main differences of the scientific production between internal and external researcher groups of a synchrotron radiation facility. Through the construction and analysis of their co-authorship networks, we could see the structural variations in the way these two different kinds of research groups collaborate. We also evaluated the scientific impact of each group and found surprising similarities, which led us to create two hypotheses that might contribute to the comprehension of the scientific assessment of large-scale research facilities. We found that, as the review criteria the studied synchrotron adopts to select external scientific projects is very effective; the quality of the external research is at least as good as the internal. Therefore, evaluating the internal scientific output appears to be an appropriate representation of the impact of the whole laboratory.

Keywords Research impact · Large-scale facilities · Co-authorship networks · Collaboration dynamics

Introduction

The interest in evaluating scientific research activities has roots going back to nineteenth century with the emergence of different concepts of research productivity (Godin 2009). Bibliometrics relies on the concept of research productivity as output, notably published papers, but the concepts of productivity as efficiency and outcome became of relevance after World War II, when funding of an increasing scientific community with the concomitant growth and spread of research institutions became a major issue (Narin and Hamilton 1996).

The conceptual changes in respect to research productivity over time, however, must be seen also in the context of changing research organizations. A founding work of

✉ Peter A. Schulz
peter.schulz@fca.unicamp.br

[1] Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas, Limeira, SP 13484-350, Brazil

[2] Centre for Science and Technology Studies (CWTS), Leiden University, PO Box 905, 2300 AX Leiden, The Netherlands

Scientometrics, “Little Science, Big Science” (Solla Price 1963), set in the very title two opposite poles of the scientific endeavor, which poses different challenges to a compreehensive research evaluation. On one side, the research activities at universities are to a large extent perceived as Little Science and papers and theirs citations counting gained a central role, which is debated in extensive literature regarding the impact evaluation of these institutions. On the other side, Big Science, strongly identified with High Energy Physics facilities, like the present large Hadron Collider at CERN, is characterized by completely different research motivations and organizations, with collaboration networks of over thousand people, leading also to completely different institutional assessment challenges (Manganote et al. 2016). Moreover, in the past few decades, a new kind of research structure with their related institutions, the so-called large-scale research facilities (LSRFs), has emerged between these two poles. Examples of LSRFs are synchrotron laboratories (or Synchrotron Light Sources—SLS, hereafter) (Hallonsten 2013), astronomical observatories (Trimble 2010), Neutron scattering sources (Lauto and Valentin 2013), and even archaeological complexes (Lozano et al. 2014).

In the present work, we chose to focus on SLS, a kind of facility now widespread in several countries and of growing importance in the development of many fields in natural sciences. The SLS were initially spinoffs of former high-energy physics accelerators. These spinoffs, identified as the first generation, evolved to entirely dedicated facilities to the new mission of innovative experimental investigations in materials sciences and biology undertaken in several research stations, the so-called synchrotron beamlines, connected to the core particle accelerators. SLS are complex and costly infrastructures that other institutions, like universities, could normally not maintain within their organizational structure divided in schools and departments. Hence, in view of the demand for using such facilities, SLS have policies for external users’ research projects applications, which share the use of the beamlines with internal staff researchers.

Until recently, “LSRFs have been object of only a few studies” according to Lauto and Valentin (2013). Considering SLSs, there are a growing number of historical accounts (Hallonsten 2015; Heinze et al. 2017; Kramer 2018). From a sociological and ethnographic point of view the interest on SLSs has been apparently pioneered by Doing (2009). An analysis on the “formation and expansion of a new organizational field (summarized in the previous paragraph) in experimental science” is provided by Hallonsten and Heinze (2015) and references therein.

Nevertheless, literature regarding the challenges of the scientific assessment of such large-scale research facilities is still incipient, but the special characteristics of such a research ecosystem lead to the proposal of the neologism “facilitymetrics” (Hallonsten 2013). From the scientometric perspective, Hallonsten (2013) states that a measure of productivity and quality of these facilities based on paper counting is not adequate and a qualitative assessment should be more appropriate. However, few quantitative data have been presented so far (Hallonsten 2016b) and the present paper contributes to fill this gap.

The necessity of filling the quantitative gap is given by, first, the fact that individual researchers, either from the SLS staff or external users, associated frequently to university departments, are still tied to the concept of productivity as output, hence counting papers and citations. Furthermore, the research reports from these large-scale facilities list the outputs (papers) from both, those related to entirely external projects, hereafter External Outputs (EO) and the ones resulting from the research efforts involving the co-authorship of internal staff, Internal Outputs (IO). Hence, the research organization of SLSs has to deal with scientific outputs from two distinct communities with asymmetric traceability, as depicted below.

Despite the reservations concerning quantitative assessments, the impact of the outputs (papers) originated at SLS, considering benchmarks to outputs from other kinds of research organizations, could hardly escape from assessing their citations. However, a simple division of scientific production into the two classes of outputs, IO and EO, causes a further challenge if one wants to evaluate the impact of a facility of this kind. Using, for instance, the Web of Science (WoS) for searching the name of one facility in the address field would only result in IO, permitting their citations retrieval. On the other hand, the use of the facility by external projects leading to EO (showing other institutional addresses) is supposed to be registered in the acknowledgements section of the articles, but this information is not structurally available so far, through the search tools of the databases. In WoS, for instance, the acknowledgements are provided only if a grant organization and grant number are mentioned as funding. Even though the use of the SLS facilities is an essential part for the generation of the EO, they are not accounted as its scientific output, and their impact by citations cannot be retrieved. Furthermore, even if WoS would provide an "acknowledgement search tool", not every paper, reporting a work that used SLS, acknowledges it. In many cases, this use is mentioned only in the body of the paper without, however, indicating the name of the facility. Hence, the only way to know the total scientific output related to the use of SLS facilities is through research reports provided by the facilities themselves.

Having in mind the possibility of an actual access to the total scientific output, as will be addressed in the present work, a further essential complementary approach to the bibliometric analysis is the study of the co-authorship networks that emerge in such large-scale research facilities. These networks are relevant to depicting the structure of the knowledge production in SLS and providing possible insights related to the differences in collaboration patterns for IO and EO cases, the aforementioned distinct communities. Despite collaboration networks have been intensively studied for a while (Newman 2004), the collaborations in SLS have not been investigated yet. Co-authorship networks in such research organizations may also shed some light on the possible relations between number of authors in a collaboration and research productivity (Lee and Bozeman 2005) and impact (Larivière et al. 2015).

In this article, hence, we combine bibliometric analyses and co-authorship networks, for both classes of SLS' user communities, leading either to IO or to EO. Thus, we were able to investigate the actual impacts of one of the SLS, from the perspective of research productivity defined by output, adding a sound and interesting quantitative assessment framework in the realm of "faciltymetrics". In what follows we first sketch the methodological pathways and afterwards results for co-authorship networks and a bibliometric analysis of scientific outputs and their impacts are presented. We complete the paper with a discussion and conclusion section.

Methodology

In this section, we briefly describe the main steps used through the present work, divided into two parts; (a) data collection and validation of these data with the framework for bibliometric analysis, which will be further addressed in the results section; and (b) the main procedures in the construction of the co-authorship networks.

Data collection, validation and bibliometrics

The data for both bibliometric and network analysis are collected from two different sources. As already stated above, the mandatory starting point is provided by properly discriminated publication lists of internal research reports. Therefore, we chose the Swiss Light Source (SLS, hereafter this acronym will refer to the chosen large-scale facility) of Paul Scherrer Institute (PSI), due to the completeness and accuracy of their reports in the period investigated. The second source is the Web of Science database, through an enhanced version hosted at the Center for Science and Technology Studies (CWTS) designed for bibliometric analyses, to validate data from the internal reports of the SLS and to proceed with the analysis of these data afterwards. CWTS's effort to create more sophisticated indicators enabled us to analyze more in-depth the impact of an institution. It should be noted that in this study only articles, letters and reviews are considered to calculate the indicators. More information about the CWTS database and methods can be found at Olensky et al. (2016) and Waltman et al. (2012).

As a first step, we compiled a publication list including IO and EO of the SLS within a 9 years period (2007–2015). This timespan is a compromise of two aspects. (1) An important characteristic of the research organization of a Synchrotron Lab is the commissioning of new beamlines and decommissioning of old ones, affecting both IO and EO during these transitions. A shorter timespan could not consider such effect in the output and impact indicators. (2) The chosen timespan of 9 years of the recent past maps, from one side, recent research activity and, on the other hand, the literature ageing: considering only citations of very recent articles could not deliver a sound evaluation of their impact (Bouabid and Larivière 2013; Zhang and Glänzel 2017). As already mentioned in the introduction, there is no direct procedure within WoS tools to capture EO through the addresses. Hence, the need of the primary records provided by PSI and only afterwards matching them against WoS. On the PSI's website, it is possible to find all publications, either IO or EO, sorted by the respective beamlines where experiments have been realized. We exported the outputs of each beamline to a spreadsheet, resulting in a list of non-standardized reference entries. An algorithm was created to standardize the entries and to capture the essential information of each reference (e.g., first author's surname and initials, title, journal, etc.), which were needed as input for an algorithm created by CWTS capable of matching articles from different databases to its own.

We used CWTS's matching algorithm to retrieve a list of unique WoS publication identifiers, linked to the source publications. The list was imported to CWTS database to facilitate the access of other information of the publications. To determine which of them were from the IO group and which were from the EO one, we developed a Structured Query Language (SQL) script to search the articles by the name of the institution. If at least one institution name in a reference entry was similar to "Paul Scherrer Inst", the publication was considered as IO. Since PSI is the whole institution and SLS is one of its many laboratories, there could be authors from PSI not allocated at the SLS, but since the related outputs are listed in the beamlines reports, they were considered as IO. Following this procedure, EO are the remaining ones, i.e., those that were not carrying the PSI address.

After successfully identifying and sorting the publications, we could collect every essential information to proceed with the proposed analysis. First, we collected names of the authors and publication years. Subsequently, we collected and calculated the bibliometric indicators. As mentioned before, only articles, reviews and letters were used in

Table 1 Example of co-authorship matrix

Author	Author		
	A	B	C
A	3	2	1
B	2	2	0
C	1	0	1

For this example with three authors and three publications, we are considering that A and B are authors of publication X, A and C of Y, A and B of Z. Therefore, A and B, for instance, have two publications in common. Publications X, Y and Z are not displayed on the table; they are part of the example to illustrate the collaboration matrix

our study, which consisted of 3796 documents of the whole set of 4120 publications for the chosen period (2007–2015).

Collaboration networks

With respect to the networks construction, we decided to create two types of co-authorship networks: (a) a first one for each year, each network consisting of documents published within that year; and (b) another one time aggregated, consisting of publications from all the years. As our period of analysis is 9 years (2007–2015) and this process was made for both IO and EO, there are 18 networks (1 for each year; for each outputs group) and 2 aggregated networks (1 for each group). The construction of the network itself followed a rather simple procedure: if two authors collaborate in the same publication, they are co-authors. Every co-authorship is addressed in a matrix and the values in matrix spots are the number of collaborations between a pair of authors, as illustrated for a simple example (three authors and three publications) in Table 1.

To create the co-authorship networks, we imported the matrixes to Gephi, a network analysis, visualization and manipulation software (Bastian et al. 2009). The Gephi software provides different network layouts and the calculation of network characteristics, which will be detailed in the results section.

Results

We begin with the presentation of co-authorship networks, since the qualitative differences between IO and EO related networks reveal to be a useful guide for the further analysis of research production and impact indicators.

The co-authorship networks are based on the retrieval of the outputs related to the SLS divided into the two groups, IO and EO. The retrieval and sorting, following the methodological procedures described above, including a first glance to the productivity and impact parameters (citations), lead to the results given in Table 2. It is worth anticipating these figures here, since the ones for IO and EO are comparable. Hence, the prevailing perception that such facilities have the fundamental role to provide opportunities to external users (Hallonsten 2016a) is not entirely reflected in the output indicators shown here. In what follows, the characteristics and importance of both groups, IO and EO, will be analyzed in depth.

Table 2 Number of publications, authors and sum of citations for each class of outputs

	Internal	External
No. authors	5076	10,306
No. publications	1476	2320
No. citations	23,796	55,790

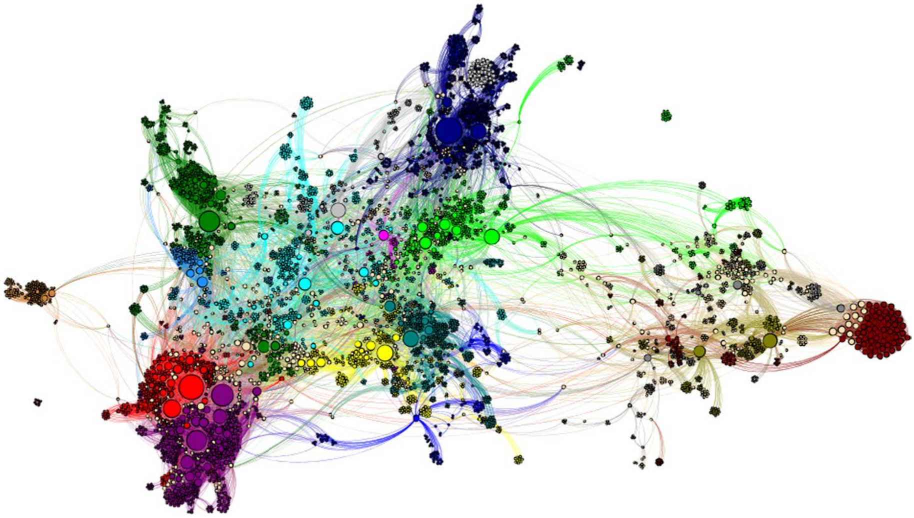

Fig. 1 Co-authorship network of SLS's internal outputs. Each node represents one author. The nodes size is proportional to the number of co-authors one author has. A line connecting two nodes indicates they have one or more publications in common. The colors are representing the beamline where the authors produced more publications over the years. The network layout, called Force Atlas, is useful for displaying groups formation. (Color figure online)

Co-authorship analysis

The first finding of this research is related to the differences between the two aggregated networks. Figure 1 shows the co-authorship network given by IO. Each node in this network represents an author and each line connecting two nodes represents that they are co-authors of at least one published output in the considered list. The nodes size is proportional to the number of co-authors one author has. Furthermore, the nodes colors were pre-defined by the beamline where the authors produced more publications over the years. If, for example, one author published 10 papers in two beamlines, 5 in each, we considered that he/she does not have a main beamline. As there are 16 beamlines, but several authors with unidentified beamline, there are 17 colors in this network. The beige nodes that are spread over the network are those authors with unidentified beamline.

The network layout was created using a force-directed algorithm, called Force Atlas. Such algorithms are usually good for graphs visualization, because they separate nodes that are more collaborative from those that are not. In this case, other layouts were not as effective showing the group formations. Although some areas of the graph are blurred due to the intense mix of nodes and edges, we can notice the collaboration between beamlines (clusters with different colors). The proximity, or even overlap of some groups, indicates

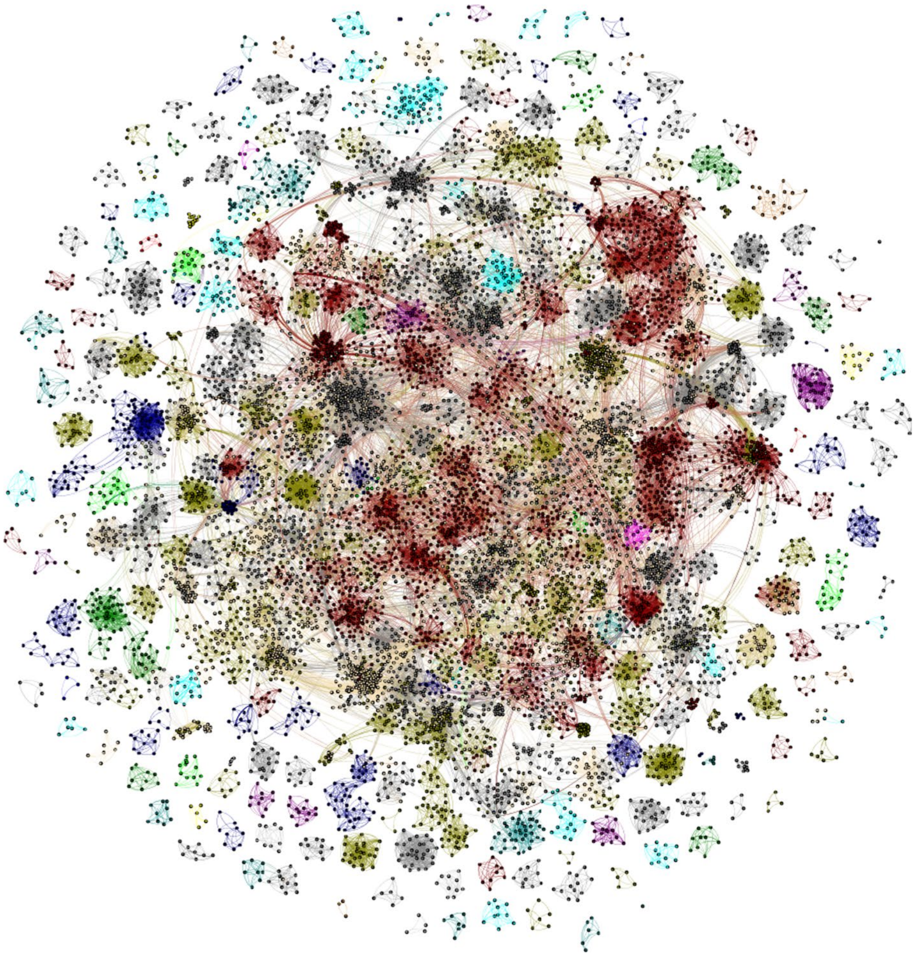

Fig. 2 Co-authorship network of SLS's external outputs. Each node represents one author. The nodes size is proportional to the number of co-authors one author has. A line connecting two nodes indicates they have one or more publications in common. The colors are representing the beamline where the authors produced more publications over the years. The network layout, called OpenOrd, is useful for organizing clusters in large networks, due to the computational speed. (Color figure online)

that their research areas may not differ significantly. We examined possible explanations for this fact and we found that some researchers take part of the research teams of two different beamlines, so they are likely to bridge them in the network. The brown community on the right side is an outlier and can be considered a type of noise. It regards a single article with about 70 authors, which is considerably above the average number of authors per publication in the network, as will be seen below. Considering the number of authors related to IO given by Table 2, we may conclude that they are highly collaborative, involving authors from different external addresses, representing interinstitutional collaborations (Larivière et al. 2015).

The EO related network illustrated in Fig. 2 shows a completely different structure. It has many more and often isolated communities and the color (beamline) distribution is

Table 3 Network measurements obtained from each year's network, for both IO and EO

Years	Internal			External		
	No. nodes	IC	GC (%)	No. nodes	IC	GC (%)
2007	474	8	66	691	65	10
2008	544	11	83	863	62	15
2009	780	12	87	1271	87	8
2010	779	14	81	1353	105	8
2011	932	15	87	1799	111	26
2012	996	11	89	1955	128	12
2013	1007	11	78	2415	132	12
2014	1364	8	85	2270	119	21
2015	1428	9	94	2133	122	8

IC isolated communities, *GC* giant component

different as well. In general, beamlines that do not have a significant presence in the IO network (i.e., the brown community) are the ones with higher frequency in the EO. This fact suggests that the synchrotron beamlines show different shares within internalized (related to IO) and externalized (related to EO) research groups. It is worth mentioning here that Figs. 1 and 2 have different layouts. Whereas the IO network (Fig. 1) uses a Force-Atlas representation, the EO network (Fig. 2) uses OpenOrd, which is also a force-directed algorithm. We chose OpenOrd in this case due to the speed of the algorithm. As the EO network has more nodes, it increased significantly the processing time of the layout. As we were less interested in the possible collaboration among beamlines, there was no need to visualize those groups as shown in Fig. 1, although the color code is the same.

A visual inspection of these aggregate networks is not sufficient to capture the differences between the collaborative characteristics involved in IO and EO, particularly as different layouts were used. Therefore, a finer analysis of the network parameters was undertaken for the networks taken year-by-year (not depicted here).

To follow the yearly development of the networks, we chose some network components, indicating their size and connectivity in Table 3. For the sake of clarity, the term "giant component" indicates the percentage of nodes that are joined together forming the largest component of the network (Newman 2003). Looking at the IO columns, we can see that, despite the increasing number of authors, the number of isolated communities remains small and the size of the giant component fluctuates around 85%. In other words, we can interpret that collaborations between authors within the same beamline and of different beamlines occur regularly, leading to a highly connected network, granted by the participation of SLS staff members. Although we have not analyzed the continuous growth of the networks over the years, it is reasonable to suggest that time has played an important role to create this giant component: authors that were once connected in the past are likely to be connected again in the future. On the other hand, this is not true for the co-authorships related to EO. Looking at Table 3, we see that as the network grows, the number of isolated communities increases and the size of the giant component is usually below 15%. Therefore, different research teams do not have a high frequency of collaboration between them, which is expected since it is more difficult to create connections between teams formed for specific collaborations without SLS internal staff members, coming from very different institutions and using the facilities with different purposes.

Table 4 Network measurements obtained from the aggregated networks of each class of output

	Internal	External
No. nodes	5076	10,306
Diameter	7	22
IC	5	146
GC (%)	99	86

IC isolated communities, *GC* giant component

We investigated both aggregated networks in a similar way as the previous analysis for separated years and the results are shown in Table 4. We added here the diameter of the network, which is the length of the longest of the shortest paths between any two nodes (Newman 2003). The size of the giant component in the IO network confirms the important role of time in its creation. It is even more impressive to see that, despite of its size, the diameter of the network is small, confirming that it is indeed highly connected. Surprisingly, the size of the giant component of the EO network is also large. However, looking at its diameter, we see that it is not highly connected. We can argue that authors from different research teams, who once were connected in the past, are less likely to be connected again in the future. Hence, the networks structures reveal that, from the internal scientific production point of view, the beamlines do not behave as isolated department-like units, while EO shows the structure of what is expected from such large-scale research facility open to external independent research proposals.

Scientific impact analysis

The second section of the analysis is dedicated to the overall scientific impact of the SLS related outputs, still separating between internal and external outputs. This separation is essential, as will be discussed below. We first checked the distribution of publications according to the number of authors that have collaborated in each output (Fig. 3). Looking over the years it is possible to inspect the existence of collaboration trends (Uddin et al. 2012).

For the sake of clarity, we introduce the IO chart in Fig. 3 depicting that the percentage of publications with 2–4 authors in 2007 was around 5%; with 5–7 around 45% and so on. Both IO and EO charts show remarkable similarities, particularly in recent years. Publications with 5–7 and with 8–10 authors represent together more than half of the total in both graphics. The number of publications with even more authors is slightly increasing over the years; while the ones with less authors decrease in EO chart in Fig. 3 and show an initial growth at the beginning of the period followed by a decrease in recent years in IO chart. The percentage of single-author papers is insignificant. The growing number of publications with more than 10 authors follows a general trend of increasing collaboration in several scientific fields and kinds of institutions. This trend, in the present case, could be related to the increasing complexity of experiments, which would require more work from more researchers (Katz and Martin 1997). It is worth mentioning that the figures shown in Fig. 3 are consistent with the results for team sizes found for Medicine and Natural Sciences over a wider spectrum of outputs and institutions presented by Larivière et al. (2015).

The evolution over the years of the scientific outputs, as well as their impacts, are depicted in Fig. 4. The impact indicators are obtained from the enhanced WoS database

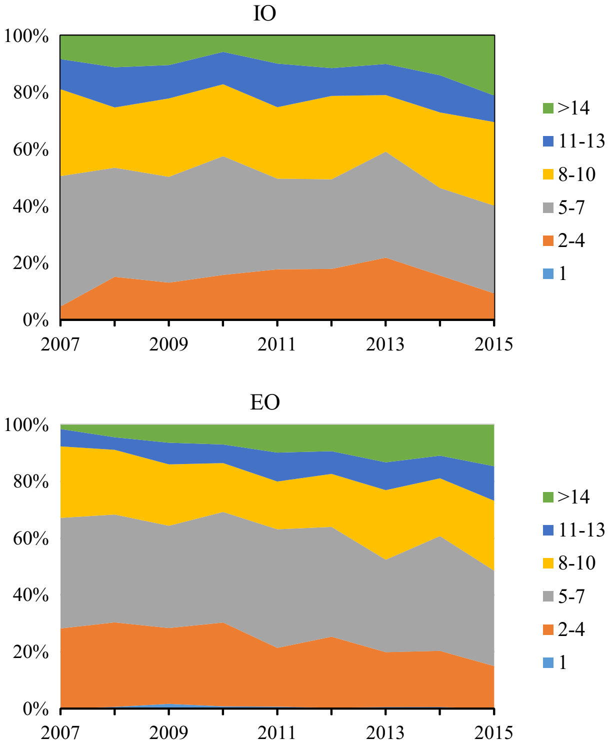

Fig. 3 Frequency of publications with x number of authors, where *x* indicates a range of values according to the caption. *IO* internal outputs, *EO* external outputs. Through the years, the number of publications being produced by more than 14 authors is visibly increasing, whereas the majority of them are still being produced by 5–10 authors

from CWTS, as mentioned above. Prior to the discussion and for the sake of clarity we first define these indicators shown in Fig. 4:

- Mean Normalized Citation Score (MNCS): average number of citations normalized to correct differences between different fields and between publications of different ages. If an MNCS indicator equals 2, it means that on average the publications have been cited twice as frequently as would be expected.
- Mean Normalized Journal Score (MNJS): similar interpretation to the previous definition, however it focuses on evaluating the journals used.
- PP (top 10%): proportion of papers belonging to the top 10% most highly cited considering research field and year of publication.

It will be also important to compare these indicators with the Mean Citation Score (MCS), average number of citations per publication, which is the usual metric presented by WoS to the general user of this database.

The number of publications of both research groups shows a similar growth pattern in Fig. 4. EO has higher values over the years. This difference is easily explained by the allocation of beamline times between groups with external users and groups with internal users. According to PSI's website, external projects have access to 70% of the total

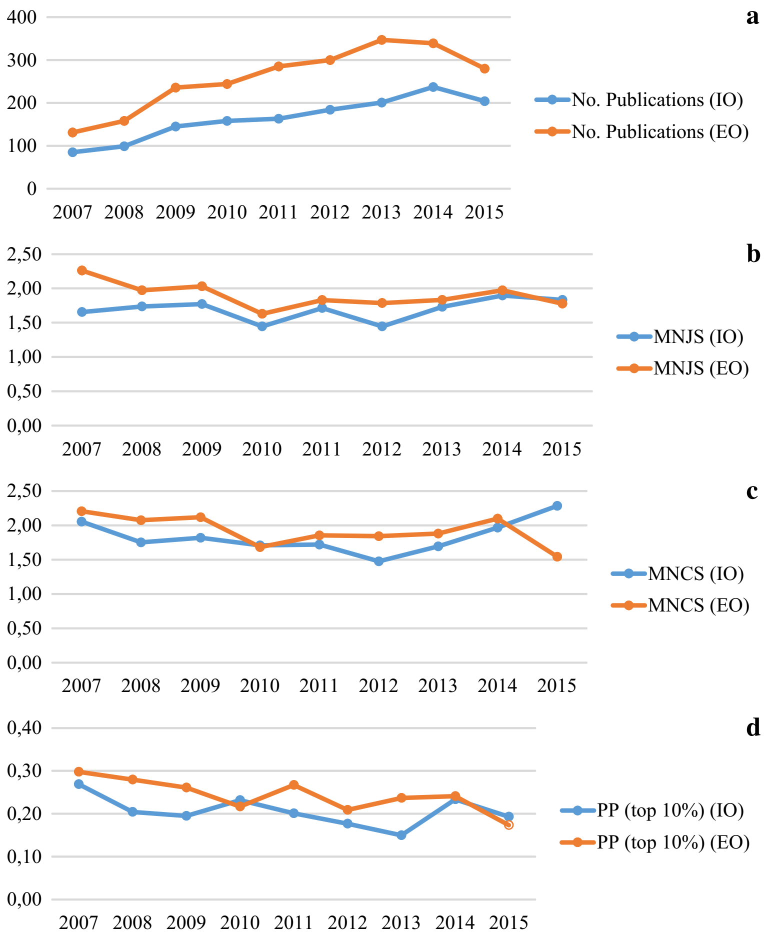

Fig. 4 Number of published documents (mainly papers) (**a**) and some scientific impact indicators (**b**–**d**) for internal outputs and external outputs, over the years. *MNJS* mean normalized journal score, *MNCS* mean normalized citation score, *PP (top 10%)* proportion of papers belonging to the top 10% most highly cited

beamline time.[1] Indeed, the number of publications from IO and EO (Table 2) scales with this time division. The values of MNCS and MNJS are considerably stable over the years and they do not differ significantly between IO and EO, both groups showing well above average citation scores.

PP (top 10%) are significant for both groups, fluctuating around 25% of the total outputs. The fluctuations in PP are in some extent related to the ones for other indicators, in particular for 2010. A further analysis could associate these fluctuations to commissioning (decommissioning) of new (old) beamlines, hence gaps in availability for experiments.

Table 5 provides values of the total scientific impact and the results are directly correlated to the yearly analysis. The higher number of external outputs consequently generates

[1] https://www.psi.ch/sls/users-proposals.

Table 5 Scientific impact indicators of each output group

	Internal	External
No. publications	1476	2320
No. citations	23,796	55,790
MCS	16.12	24.05
MNCS	1.83	1.90
MNJS	1.70	1.87
PP (top 10%)	0.20	0.24

MCS mean citation score, *MNCS* mean normalized citation score, *MNJS* mean normalized journal score, *PP* (*top 10%*) proportion of papers belonging to the top 10% most highly cited

higher number of citations, but these values do not scale entirely, suggesting that EO would have a higher impact, indeed the widespread used MCS seem to support this suggestion. However, the proximity of CWTS indicators reveals that internal and external outputs have actually similar scientific impacts. Even considering the differences in collaboration structure, revealed by the co-authorship analysis, there are no differences in the measured impacts. Both IO and EO have been cited, on average, almost twice as would be expected (MNCS); and the MNJS has a similar pattern. The difference between the MCS values are probably due to external publications belonging to fields that are usually more cited. This is an important example for the inappropriateness of using bare citations counting in measuring and comparing research impact.

This hypothesis concerning the publication pattern leading higher average citation counts (MCS) for EO compared to IO, seems to be supported by inspection of the journals in which the papers are published. The IO (EO) are spread over 327 (353) journals with 155 common to both groups. Furthermore, the most frequent journals used by IO and EO are completely different. The most frequent ones for IO are in Physics and Instrumentation categories, while for EO the most frequent categories are in fields related to Biology, indeed categories with very different referencing and citing cultures. Among the journals common to both groups, it should be noticed that the frequency is also very different. The most frequent journals are given in Fig. 5. In Fig. 5, some important (frequent) journals are missing, because they are not common to both communities. Optic Express (important journal for relevant aspects of Synchrotron technology), for instance, published 27 IO and none EO. On the other hand, Molecular Cell (leading journal in Biology) published 37 EO and none IO.

Discussion and conclusion

The purpose of this study was to investigate the differences between the scientific production of research groups associated to IO and EO of a synchrotron laboratory. We could see through their co-authorship networks that their collaborations are significantly different. Whereas most of the IO can be associated to a main beamline, collaborations among different beamlines are very frequent and could occur for various reasons, such as similar research themes or the existence of common members to two different research teams. Analyzing evolution over time of the networks, we could also see that authors who collaborated once in the past, are likely to collaborate again in the future. On the other hand,

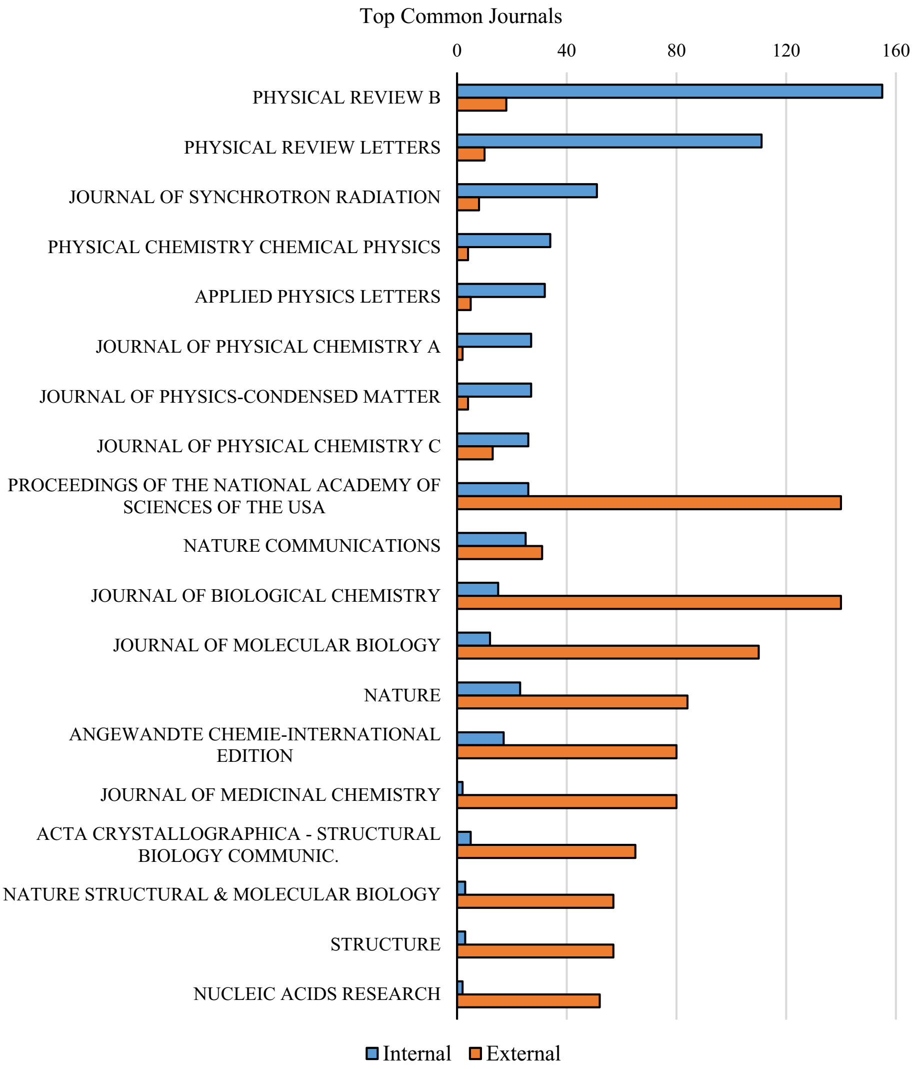

Fig. 5 Most frequent Journals used by Internal Outputs (blue) and External Outputs (red) related communities. The numbers indicate the number of published papers over the period 2007–2015. Only journals that are common for both groups are being displayed here. (Color figure online)

we see that the EO related co-authorship networks are less connected, as the elevated number of isolated communities and the diameter of the aggregated network prove. We also see that collaborations between authors from different research teams do happen, but they have a smaller chance of happening again in the future. We may attribute this to the way research is done by external users: they already have a team; come from different locations; and they have different interests in using the synchrotron.

Although the two kinds of research groups have structural differences in their co-authorships networks, they share some similarities; the first of them related to the number of authors collaborating in each paper. The vast majority of publications from both groups have 5–10 authors and the percentage of publications with even more authors is increasing over time. As this behavior is similar for both, we can deduce that IO and EO are related to research efforts that probably share the same level of complexity, as well as following a more general trend of increasing number of authors per paper (Larivière et al. 2015).

Furthermore, we analyzed the scientific impact of each research group, which are surprisingly similar. EO are more frequent because SLS allocates more time for external use. Despite that, the Mean Normalized Citation Score of IO and EO are 1.83 and 1.9, respectively; hence are almost equally cited within a proper normalized view. We can raise two hypotheses regarding the previous fact:

1. If one wants to evaluate the scientific impact of large-scale research facilities using, for instance, the WoS, taking only the internal production might be a good representation of the total. It is important to mention again the difficulty of retrieving the external production using only the WoS.
2. The apparent uncontrolled growth of the EO related network is not hindering the overall quality of the external research. This is probably due to the quality of criteria SLS adopts for selecting which projects are going to be developed there.

For the second hypothesis, we sought to understand how the project selection process happens. PSI's website has a section describing some procedures related to external users' proposals.[2] We discovered that an international external committee is responsible for evaluating proposals and they do so according to the IUPAP recommendations for large-scale facilities.[3] The selection and prioritizing of experiments are based on the following criteria which are graded from 1 (rejected) to 5.5 (excellent proposal): scientific merit; technical feasibility; capability of the experimental group; availability of the resources required. This review process bares similarities with those undertaken by research funding agencies, although in the present case the "grants" regard beamtime for experiments rather than money. Hence, figures for projects rejection rates by the review committees would be an interesting information, but they are not readily available for the SLS. Still, the annual report for 2016 of a similar facility, namely the European Synchrotron Radiation Facility, discloses rejection rates of 60% (i.e. success rate of 40%) of the applications.[4] Such rates are indeed comparable to research funding agencies in Europe (van Leeuwen and Moed 2012). Our results strongly suggest that the selection criteria applied by review committee are successfully fulfilling their task.

Taking into account the difficulty of evaluating the real scientific impact of large-scale research facilities, the merge of the two hypotheses above provides an interesting alternative. If the facility has a proposal review committee composed by experts in related areas, guided by similar selection criteria, then the quality of the selected projects is probably going to be as high as of the projects developed internally. Thus, scientific indicators of the internal production obtained from the WoS might be a good indication of the average total scientific impact. There are, nevertheless, two evident restrictions to this approach. The first is that we cannot find the total number of publications; however, if more time is

[2] https://www.psi.ch/sls/evaluation.

[3] http://www.iupap.org/wp-content/uploads/2013/12/Recommendations22ndGeneralAssembly.pdf.

[4] http://www.esrf.eu/Apache_files/Highlights/HL2016.pdf (see Fig. 168).

allocated for external users, it is likely that they will sum up more publications. The second is related to the Mean Citation Score, frequently used by research activities stakeholders. If we analyze the MCS of the internal production obtained from the WoS, the MCS of the external can be much higher than IO. However, the fairness of this analysis will be hindered by using the MCS, instead of the MNCS, as our results show.

The present results suggest further developments. First, there are some occurrences of different nodes in the networks representing the same author. A great effort would be necessary to avoid this, but we believe that the result would not be significantly different for the purpose of the present analysis. Nevertheless, a further scrutiny of this effect could bring to light finer details of the collaboration dynamics. Second, we focused on a quantitative analysis. A more in-depth qualitative analysis would provide an interesting complement to this project, such as investigating how the external proposals review committee makes their analysis and how the steering committee evaluates internal projects and how it drives and motivates the collaboration between different beamlines. It is worth mentioning, however, that the interplay between IO and EO reveals a more complex scenario than the one assumed for these facilities in the literature (Hallonsten 2016b). For future research, it would be interesting to explore these aspects, as well as extending this analysis to different large-scale research facilities. Nevertheless, is should be noted that the present results suggest a relevant and so far unprecedented quantitative approach to analyze the scientific production of this kind of large facilities.

As a final remark, the present results suggest the relevance of acknowledgments to the Synchrotron by external users, mentioned in the introduction, for research assessment policies. Following the funding agencies initiatives for demanding acknowledgement in the articles resulting from their support, large-scale research facilities could also better monitor such practice. On the other hand, bibliometric databases could enhance “acknowledgemetrics” beyond funding, as the importance of indexing acknowledgements is being increasingly recognized (Paul-Hus et al. 2017).

Acknowledgements This work was supported by Grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil.

References

Bastian, M., Heymann, S., & Jacomi, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Web and Social Media, 8,* 361–362.

Bouabid, H., & Larivière, V. (2013). The lengthening of papers’ life expectancy: A diachronous analysis. *Scientometrics, 97,* 695–711.

de Solla Price, D. J. (1963). *Little science, big science*. New York: Columbia University Press.

Doing, P. (2009). *Velvet revolution at the synchrotron: Biology, physics, and change in science*. Cambridge: MIT Press.

Godin, B. (2009). The value of science: changing conceptions of scientific productivity, 1869 to circa 1970. *Social Science Information*, *48*(4), 547–586

Hallonsten, O. (2013). Introducing ‘facilitymetrics’: A first review and analysis of commonly used measures of scientific leadership among synchrotron radiation facilities worldwide. *Scientometrics, 96*(2), 497–513.

Hallonsten, O. (2015). The parasites: Synchrotron Radiation at SLAC, 1972–1992. *Historical Studies in Natural Science, 45*(2), 217–272.

Hallonsten, O. (2016a). Use and produxtivity of contemporary, multidisciplinary Big Science. *Research Evaluation*, *25*(4), 485–495.

Hallonsten, O. (2016b). Use and productivity of contemporary, multidisciplinary big science. *Research Evaluation, 25*(4), 486–495.

Hallonsten, O., & Heinze, T. (2015). Formation and expansion of a new organization filed in experimental science. *Science and Public Policy, 42,* 841–854.

Heinze, T., Hallonsten, O., & Heinecke, S. (2017). Turning the ship: The transformation of DESY, 1993–2009. *Physics and Pespective, 19*(4), 424–451.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

Kramer, K. C. (2018). Lightening Europe: Establishing the European Synchrotron radiation Facility (ESRF). *History and Technology, 33*(4), 396–427.

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology, 66*(7), 1323–1332.

Lauto, G., & Valentin, F. (2013). How large-scale research facilities connect to global research. *Review of Policy Research, 30*(4), 381–408.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration and scientific productivity. *Social Studies of Science, 35*(5), 673–702.

Lozano, S., Rodríguez, X.-P., & Arenas, A. (2014). Atapuerca: Evolution of scientific collaboration in an emergente large-scale research infrastructure. *Scientometrics, 98*(2), 1505–1520.

Manganote, E. J. T., Schulz, P. A., & Brito Cruz, C. H. (2016). Effect of high energy physics large collaborations on higher education institutions citations and rankings. *Scientometrics, 109*(2), 813–826.

Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics, 36*(3), 293–310.

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

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences, 101*(1), 5200–5202.

Olensky, M., Schmidt, M., & van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the web of science. *Journal of the Association for Information Science and Technology, 67*(10), 2550–2564.

Paul-Hus, A., Díaz-Faes, A. A., Sainte-Marie, M., Desrochers, N., Costas, R., & Larivière, V. (2017). Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences. *PLoS ONE, 12*(10), e0185578.

Trimble, V. (2010). A generation of astronomical telescopes, their users, and publication. *Scientometrics, 84*(1), 21–34.

Uddin, S., Hossain, L., Abbasi, A., & Rasmussen, K. (2012). Trend and efficiency analysis of co-authorship network. *Scientometrics, 90*(2), 687–699.

van Leeuwen, T. N., & Moed, H. F. (2012). Funding decisions, peer review, and scientific excellence in physical sceinces, chemistry, and geosciences. *Research Evaluation, 21,* 189–198.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., et al. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63,* 2419–2432. https://doi.org/10.1002/asi.22708.

Zhang, L., & Glänzel, W. (2017). A citation-based cross-disciplinary study on literature ageing: Part II—diachronous aspects. *Scientometrics, 111,* 1559–1572.