



Characterization of Women's Scientific Participation in Brazil

Monique de Oliveira Santiago^(✉), Thiago Magela Rodrigues Dias,
and Felipe Affonso

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG),
Belo Horizonte, Brazil

moniqueosantiago@gmail.com, thiagomagela@gmail.com,
felipe-affonso@hotmail.com

Abstract. Several researchers have concentrated efforts to study the participation of women in scientific and technological careers, seeking to profile their trajectory and academic performance in science. In this context, this study aimed to analyze the participation of the group of doctor's degree who have curricula registered in the Lattes Platform and whose registered gender is female. After data collection, the stage of selection of curricula by gender criteria was performed, and data processing obtained a set of 149,850 registered curricula with female gender and maximum completed doctor's degree distributed in its various areas of scientific knowledge. The doctor's degree data were grouped regarding academic background, publications, productions, orientations, major areas of expertise and it was possible to analyze the evolution of scientific and technological production of the set in a temporal way. Studying the various aspects of gender difference in general and particularly in science and technology, as well as being relevant, can be a source of inspiration for government policies and programs that seek to promote change, encourage and value women's participation in science.

Keywords: Women in science · Lattes Platform · Scientific and technological production

1 Introduction

Scientific production has grown significantly in recent years, using the Web as one of the means that facilitates access and dissemination of its content. The data regarding scientific production has provided studies that seek to understand how science has evolved and how scientific collaboration occurs. Among the various studies applied to this set, one that has gained prominence is gender studies. Gender can be defined as the connection of two propositions: “[...] gender is a constitutive element of social relationships based on perceived differences

between the sexes, and gender is a primary way of signifying relationships of power" [9, p. 1067], therefore, corresponds an interdisciplinary field that has as its theme the identity and representation of men and women in society. This field includes the subfield Women's Study, which covers, among various dimensions of interest, women and their various relationships with science [5].

Because it is an interdisciplinary and comprehensive theme, studies that focus on women and their varied relationships with science have several approaches that seek to outline a career trajectory and female performance. Despite progress in women's participation in various segments in their academic and scientific careers, there is still a gender gap in science around the world that needs to be better understood. There is clear evidence of gender inequality in Brazilian Science, which probably has deep institutional and cultural roots [8]. Thus, conducting a study to investigate the scientific participation of women in Brazil is a step to understand the current scenario, and through this understanding contributes to the adoption of measures that promote gender equality. There may be differences in the performance of men and women in Brazilian science, but these differences are related to their presence in the field and the clipping that is given to the analyzes [5]. Reflecting on and searching for new ways to investigate women's productivity in science is a step towards minimizing existing career inequalities.

One of the biggest difficulties in analyzing a country's scientific production may be related to the acquisition of data, which is usually present in several repositories. However, this process can be facilitated by using the Lattes Platform curriculum data repository. The Lattes Platform is considered an important set of Brazilian scientific data, in which it provides high-quality information and enables researching data from individuals who are registered there, such as academic background and scientific production, among others [4]. This data set integrates into a single system the databases of Curriculum, Research Groups and Educational Institutions of the country and has curriculum information of the academic trajectory of the entire Brazilian scientific community.

Studying large data repositories becomes a complex task because the amount of data to be analyzed and the characteristics of each repository are unique and most have no definite pattern. In this scenario, bibliometrics emerges that seeks to quantify the processes of written communication, using methods for statistical analysis of the production and dissemination of knowledge applied to scientific data sources [1]. Therefore, bibliometric analyses are used as indicators of scientific production to provide indicators for national planning for the evolution of scientific research.

In this context, this paper aims to analyze the scientific and technological women's participation, investigating how their research has been carried out and how their academic career has evolved over the years from bibliometric analyses performed on curriculum data available on the Lattes Platform. This study, besides presenting an overview of women who have been conducting research, aims to contribute to the generation of national scientific indicators and the management of information in the scientific and technological areas.

One of the factors that motivate the study or understanding of women's scientific participation is the opportunity to understand the current scenario, and through this way to implement measures to promote reproductions between genders, minimizing possible career inequalities. Since, by recognizing that “[...] a science is gender-neutral, revealing the values and social characteristics attributed to women are devalued in the production of knowledge, and that gender inequalities pervade the field. scientific [...]” [10, p. 464, our translation]¹, admits a need to break down as barriers to inequality and thus ensures research excellence by retaining its best researchers and innovators [8] and thus enables “[...] to create a science more human, free from the transformations caused by centuries of exclusion of half of humanity as women” [5, p. 150, our translation]².

2 Development

All curriculum information for the Lattes Platform is included by the researcher himself and is freely available on the internet. This broad data set contains the entire record of the researcher's professional, academic and scientific career. Due to its wealth of information and as this data source has not been widely analyzed, it is therefore justified to choose the Lattes Platform as a data source to measure and evaluate women's national scientific performance.

The process of extracting and selecting of curriculums data from the Lattes Platform was performed through the LattesDataXplorer framework [3], which has a set of techniques and methods responsible for collecting, selecting, processing and analyzing the data. Thus, the framework collection module was responsible for collecting all curriculum in March 2019, surpassing 6,126,000 records.

LattesDataXplorer was specifically used for the collection and selection of the Lattes Platform of curriculums data, which obtained the Doctor's Degree Curriculums Repository in XML. After this process, it was necessary to select data from this repository and thus obtain the repository of all women's curriculums with completed academic/doctor's degree. Finally, the data were processed and visualized to analyze the scientific participation of women (Fig. 1).

To carry out a detailed analysis of women's national scientific participation, it was decided to limit the data through the level of academic education/degree, reducing the set for individuals who have completed the doctor's degree level. Although this set is not the most significant among the levels of education, they account for 74.51% of papers published in journal article and 64.67% published in conference paper, besides have generally updated date of their curriculums recently and notably are responsible for the highest level of training, namely,

¹ [...] a ciência é neutra com relação às questões de gênero, revelando que os valores e as características socialmente atribuídos às mulheres são desvalorizados na produção do conhecimento, e que desigualdades de gênero perpassam o campo científico [...].

² [...] criar uma ciência mais humana, livre das transformações causadas pelos séculos de exclusão de mais da metade da humanidade, as mulheres.

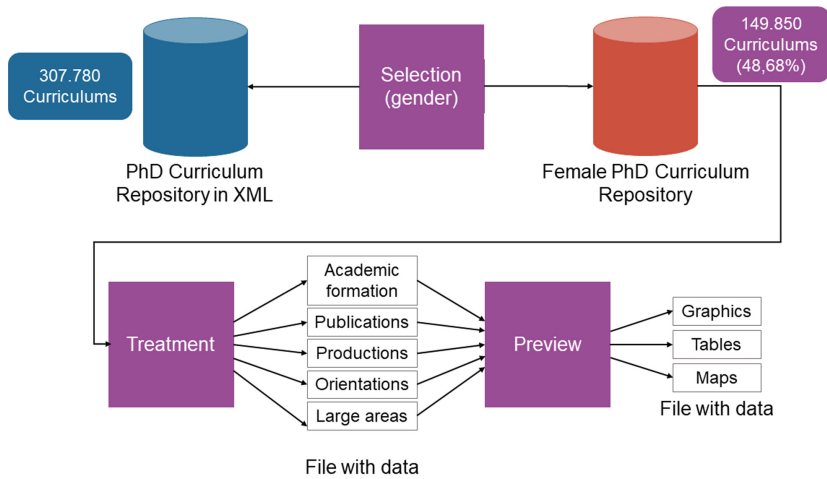


Fig. 1. Data selection, treatment and preview process.

masters and doctor’s degree [3]. Therefore, after the acquisition of all curriculums registered in the Lattes Platform, the LattesDataXplorer selection module was used to select, among them, the curriculums that have completed the doctor’s degree, thus totaling a set of 307,780 (5,02%) curriculums.

After the curricular data collection and doctor’s degree selection stages, the selection by gender criteria was performed, as some fields collected by the Lattes Platform are not displayed in the public consultation, such as CPF, gender, color or race, birth data, identity, passport, membership, home address, and more. Thus, we selected the female-owned curriculum from the Phd Curriculum Repository in Xml yielded 149,850 female curricula. This total corresponds to 2.45% of all Lattes Platform data.

The XML files extracted from the Lattes Platform have a well-defined structure, with structured information and data diversity, such as level of academic education/degree, large areas of expertise, research and extension projects, bibliographic and technical productions such as articles published in annals congress and papers in journal, presentation of papers, participation in newsstands, events, orientations, among others. As each curriculum has a specific amount of this information, treatment was performed to improve understanding of the data. Therefore, the preprocessing step was performed by accessing each XML curriculum, obtaining that curriculum-specific information, and storing it in a collection of structured files. In the end, these data were grouped in terms of academic background, publications, productions, orientations, major areas, among others. After grouping the data, a characterization was performed to facilitate the preview and analysis of the data. In this step, for each collection of structured files were generated charts, tables, maps, among others.

3 Results

3.1 General Characterization

Data collected from the Lattes Platform using the LattesDataXplorer framework in March 2019 totaled more than 6,126,000 records. From this total, the records with the level of academic doctor's degree completed were selected, totaling 307,780 (5.02%) curriculums from the various areas of scientific knowledge. These same data were selected by gender criteria, in which 149,850 (48.68%) correspond to curriculums of doctor's degree female and 157,942 (51.31%) to curriculums of doctor's degree male.

This data shows significant growth regarding the data provided by the Lattes Panel [7]. The last extraction performed by Lattes Panel took place on November 30, 2016, being the same made available for public consultation, which provided data regarding the panel Distribution by gender, age and area corresponding to 63,853 (47.50%) doctor's degree curriculums female and 70,567 (52.49%) doctor's degree curriculums male. Thus, by comparing these data provided by the Lattes Panel with the current data collected from the Lattes Platform, we can see the growth of 128.97% of researchers who completed their doctor's degree in more than two years. Another relevant aspect refers to the percentage of doctor's degree training, in which females increased by 1.18% and males decreased by -1.38%. Even with this considerable percentage, it was not enough for female doctor's degree to outnumber male doctor's degree.

With the general characterization of the data, it was possible to visualize the representativeness of the selected set within the total set of records. As the focus of this work is to characterize the Brazilian female scientific participation, the next analyzes will be used only the data set corresponding to the doctor's degree curriculums female (149,850 curriculums), which corresponds to 2.45% of the total of all curriculum registered in the Lattes Platform.

3.2 Women's Orientations Profile

Relevant information for analysis corresponds to the completed and ongoing orientations of women, as they are the doctor's degree responsible for the education of students at different levels of education. The completed orientations correspond to all the orientations made by the doctors since the beginning of their career and which are already finalized, while the orientations in progress are those that are under development and not yet finalized. Thus, by quantifying these data we have a total of 3,577,801 completed orientations and 341,607 ongoing orientations (Fig. 2).

An important aspect in the graphs above refers to the levels of postgraduate education: master's degree, doctor's degree, and postdoctorate. At these three levels, the sum of the percentages for the completed orientations graph is 17%, while for the ongoing orientations graph this sum is 47%. One hypothesis for this percentage difference concerns the fact that, as doctor's degree are responsible for student education in major postgraduate programs *stricto sensu* in Brazil,

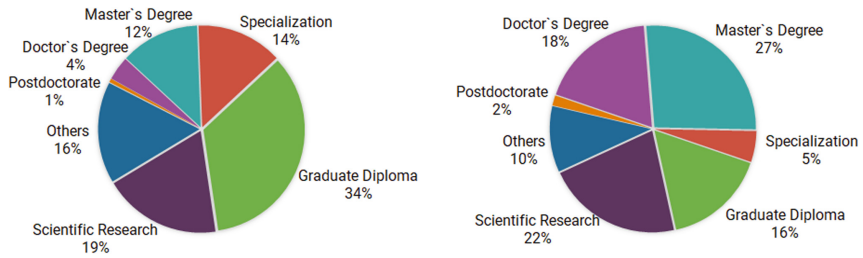


Fig. 2. Completed orientations (left) and ongoing orientations (right).

they tend to mentor more undergraduate students and fewer postgraduate students at the beginning of their careers and over the years after completing their doctorate the number of undergraduate counseling decreases and the postgraduate counseling increases. However, as all orientations are considered throughout their history, the completed orientations have the lowest level, as undergraduate degrees being more representative.

3.3 Representativeness of Women’s Bibliographic and Technical Production

The doctor’s degree are responsible for most of the works registered in the Latex Platform [3], so a survey of the main bibliographical, technical and other artistic/cultural productions of women was carried out. Thus, verifying the scientific and technological production of the data set of female doctor’s degree, we arrive at the quantitative, presented in (Table 1), corresponding to the main productions.

Table 1. Quantitative data by bibliographical and technical productions of womans doctor’s degree.

Production type	Production
Conference paper	5.092.804
Journal article	2.274.378
Other technical productions	1.375.610
Technical works	824.913
Book chapters	593.484
Texts in newspapers and magazines	314.212
Other bibliographic productions	313.423
Books	200.515

The previous table shows that the number of papers published in a conference paper and journal article is higher than the other types of productions, both

of which are the most common for publication and dissemination of content. From this data and taking into account the productivity of the doctor’s degree per year, a quantitative was performed (Fig. 3) with the most relevant types of productions.

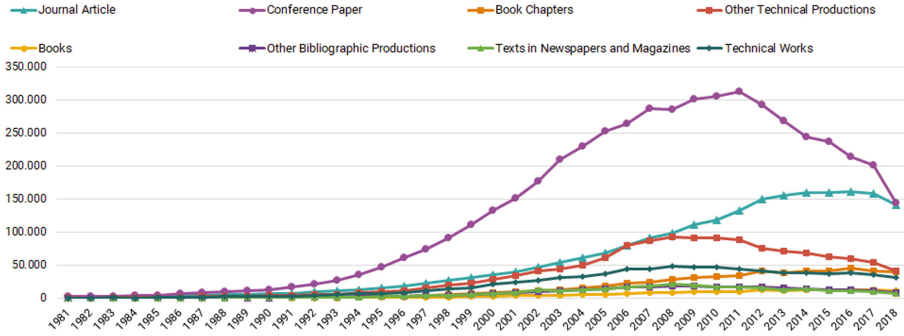


Fig. 3. Quantitative of scientific productions per year.

The first twenty years (1961 to 1981) scientific production remained constant for all types of productions. Possibly, a hypothesis to explain this fact may be related to the launch and standardization of the Lattes Platform curriculum that took place in August 1999, in which data referring to publications prior to this period may not have been disclosed by the doctor’s degree. Another hypothesis that may influence this fact is related to the number of doctor’s degree, who were few at the time.

Since the early 1980s, there has been an increase in all types of scientific and technological production. The productions that emerged from the others correspond to journal article and conference paper. Regarding journal article, it presented a significant growth until 2013, remaining constant after this year and falling in 2017. The production of conference paper had a considerable increase with the apex in 2011 and a significant decrease after that year. This same behavior is presented in the study of [3], for all doctor’s degree with curriculums registered in the Lattes Platform. This steep decline in conference paper was so significant that from its peak until 2018, it fell by 53.62%, reaching the end of last year with a total value of articles close to the total of journal article.

Different hypotheses may be related to the steep slope of the conference paper from 2011. One of them refers to the classification of scientific production used by CAPES, if we consider that the production evaluation system influences the actions of individuals. This fact is driven by the non-consideration of articles in conference paper in postgraduate program evaluations. Thus, journal article becomes more interesting for publication, as they influence the concepts of the programs in which the doctor’s degree participate and their efforts for this type of publication.

3.4 Distribution of Women in Their Major Areas

The information that describes the professional ties can be registered in the Lattes Platform curricula through the “Practice/Practice Areas” menu, allowing to choose and inform numerous knowledge areas and subareas. The large area can be defined as a “clustering of diverse areas of knowledge by virtue of the affinity of their objects, cognitive methods and instrumental resources reflecting specific sociopolitical contexts” [2]. The Lattes Platform has nine options for large areas that aggregate their respective areas, following the CNPq classification³.

To perform this analysis, the first option filled out by the female doctor’s degree was selected considering that this is the most relevant area among the others. Thus, using the two most significant means of production to group the doctor’s degree curriculum by major practice areas (Fig. 4), as expected, resulted in the most significant major human-related fields and the least significant corresponding to engineering. These data showing that men predominate in exact careers and women predominate in biological areas and health [6].

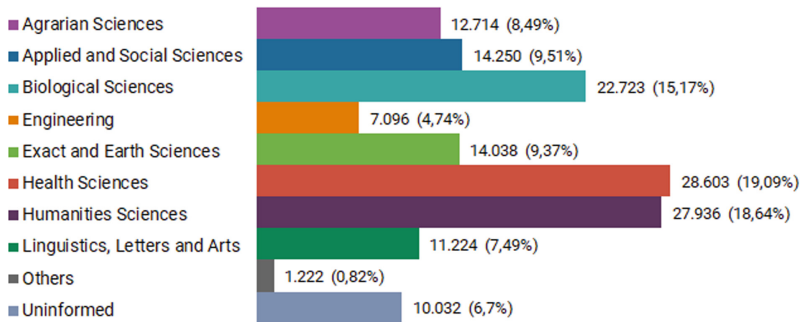


Fig. 4. Major areas of expertise.

A significant number of curriculums do not have information related to their area of expertise (10,032 curriculums). This can be explained by the fact that, as it is not mandatory information, when entering a new academic formation there is no concern to fill in the data of this field, or even there may be forgetfulness on the part of doctors. Another aspect that can be taken into account refers to the difficulty factor in associating the course with the area of expertise, especially for interdisciplinary courses and recent postgraduate courses, which results in not informing data for this field.

As journal article and conference paper are the two most significant means of production, they were used to verify scientific production by a large area of expertise (Fig. 5).

³ CNPq Knowledge Area Table: <http://lattes.cnpq.br/documents/11871/24930/TabelaAreasdoConhecimento.pdf/d192ff6b-3e0a-4074-a74d-c280521bd5f7>.

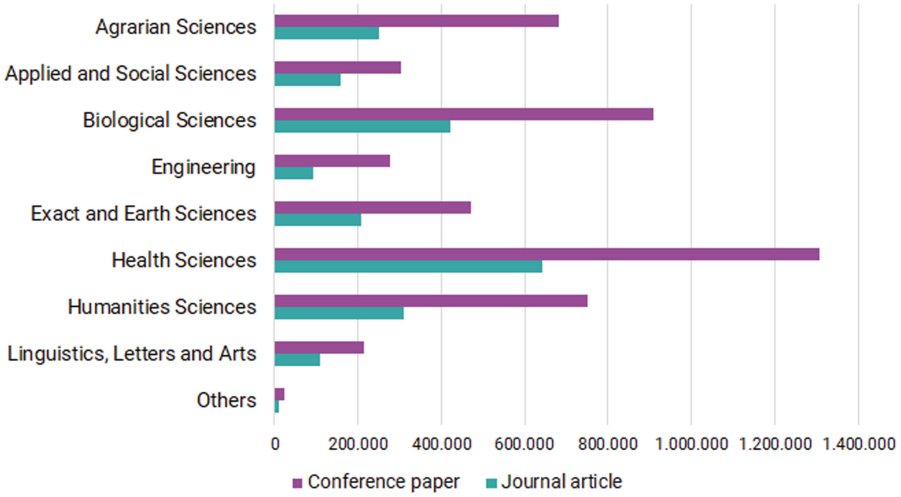


Fig. 5. Journal article and conference paper by area of activity.

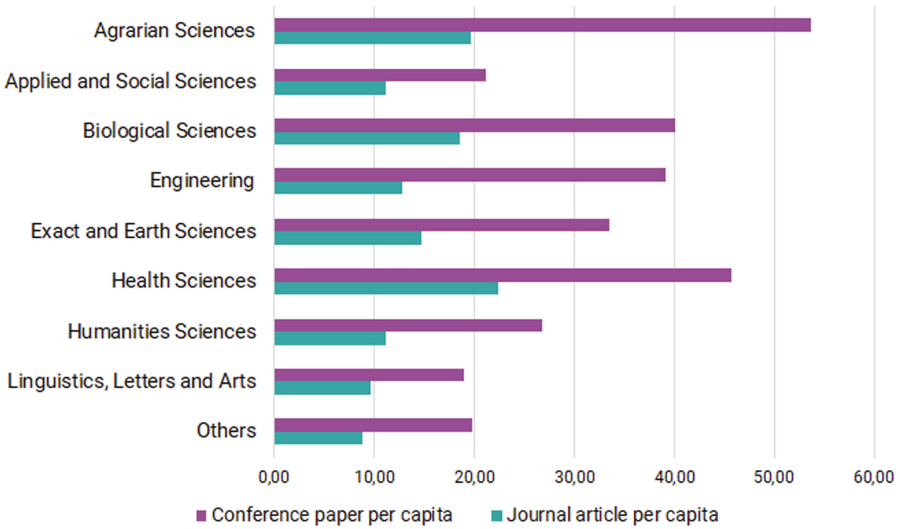


Fig. 6. Journal article and conference paper per capita by area of activity.

By dividing the total number of articles by the total number of individuals according to the corresponding large area of activity, we have a number of productions per capita (Fig. 6). This analysis becomes more interesting because it is possible to verify the production without the influence of a large number of authors from some areas.

Comparing the two previous figures, we can see a considerable change of positions between some areas of expertise in the production of journal article and conference paper. For the articles in conference paper, the first area that stands out about the others is the Agrarian Sciences, in which Fig. 5 presents itself as the fourth-largest producer of articles and Fig. 6 appears first. Health Sciences and Biological Sciences, which were first and second in Fig. 5, remain quite significant in Fig. 6, second and third place respectively. The Engineering area had a surprising and unexpected change from the second to last (Fig. 5) to fourth place (Fig. 6). The Humanities Sciences that was in third, moved to sixth place.

For journal article, the change of positions appears less significant than for articles in conference paper. Featured for Agrarian Sciences from a fourth-place (Fig. 5) appears second in Fig. 6., and the Engineering area had a surprising and unexpected change from the last (Fig. 5) to the fifth (Fig. 6).

These data reveal to us the importance of the scientific production of the doctor's degree in several major areas. It was possible to identify the current scientific production scenario and to verify that even the areas with less female participation, such as engineering, per capita production presents the opposite scenario.

3.5 Evolution of the Academic Formation of Female Doctor's Degree

All academic history can be informed in the Lattes Platform curricula, allowing to verify the academic trajectory of the doctor's degree. For doctor's degree, master's degree and professional master's degree, one can inform the major areas of knowledge and choose from nine options by selecting the areas and then the subareas that relate to the postgraduate course.

As it is possible to fill in more than one area of knowledge for each postgraduate informed, for this analysis was selected only the first area filled by the female doctor's degree, considering to be the most relevant among the areas registered. Thus, it is possible to verify the evolution of the major areas for the doctorate academic formation (Fig. 7).

As some fields in the Lattes Platform curriculum are not required, data for the starting year, ending year, or large area of knowledge that did not contain information was characterized as Uninformed and these records were not used for analysis. Noting that for the doctorate 45,510 (30.37%) records were categorized as Not informed.

The 71 years of history regarding the formation of doctors in the Lattes Platform were divided into four periods. In these periods, the major areas that formed the most doctor's degree per period alternated between Exact and Earth Sciences, Humanities Sciences, Biological Sciences, and Health Sciences.

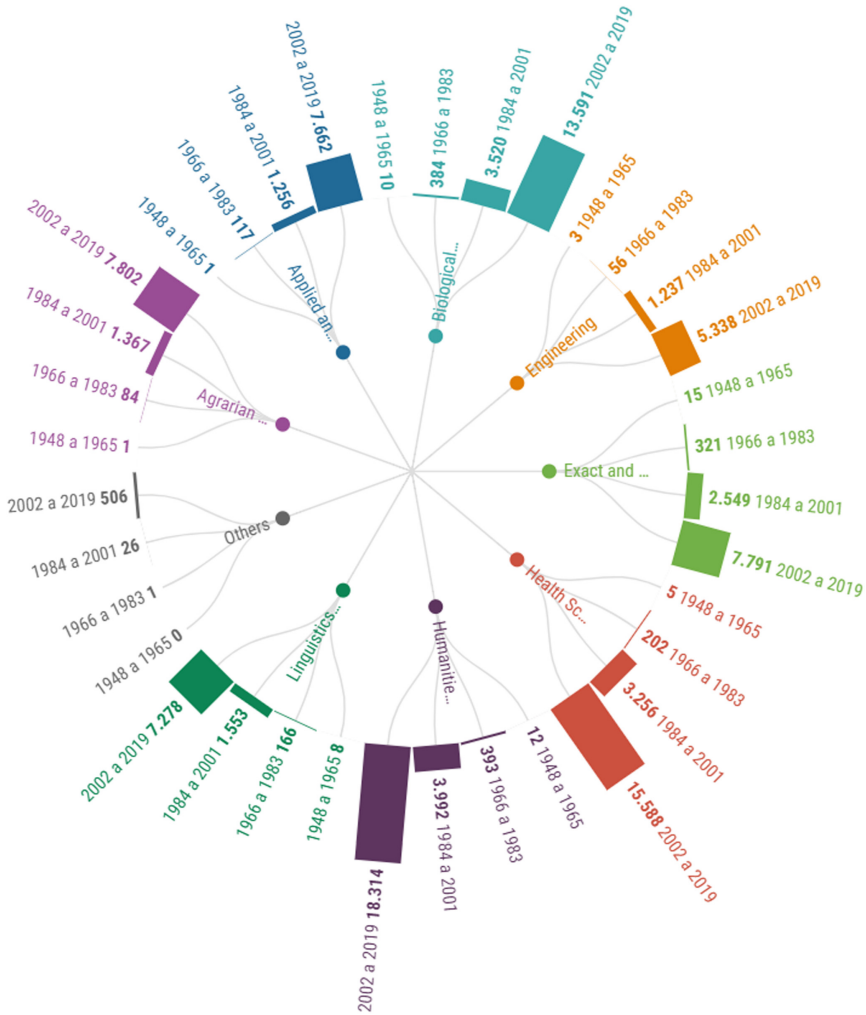


Fig. 7. Evolution of the major areas for the doctorate academic formation.

4 Final Considerations

Studies that focus on the various aspects of gender difference in work in general, and particularly in science and technology, are relevant and can be a source of inspiration for government policies and programs that promote change and lead to equal participation for women and men. As this study presents the full potential offered by the Lattes Platform for the understanding of women’s scientific participation in Brazil, all the results presented aimed to present a view of female participation in a temporal way on the data set. As future work is

expected to perform an analysis of the scientific collaboration of doctor's degree, CNPq research productivity scholarship holders and the scientific production of doctor's degree working in postgraduate programs.

Supported: The authors thank CAPES for their research assistance.

References

1. Araújo, C.A.A.: Bibliometria: evolução histórica e questões atuais. Em *Questão* **12**(1), 11–32 (2006)
2. CAPES. Tabela de Áreas de Conhecimento/Avaliação, Brasil (2019). <https://www.capes.gov.br/avaliacao/sobre-a-avaliacao/mestrado-e-doutorado-o-que-sao/44-cont-eudo-estatico/index.php?option=comcontent&view=article&id=6831:tabela-de-areas-de-conhecimentoavaliacao&catid=91:avaliacao-capes&Itemid=759>. Acesso 04 dez 2019
3. Dias, T.M.R.: Um Estudo da Produção Científica Brasileira a partir de Dados da Plataforma Lattes. Centro Federal de Educação Tecnológica de Minas Gerais, Doutorado em Modelagem Matemática e Computacional (2016)
4. Lane, J.: Let's make science metrics more scientific. *Nature* **464**(7288), 488 (2010)
5. Leta, J.: Mulheres na Ciência Brasileira: desempenho inferior? *Revista Feminismos* **3**(2), 139–152 (2014)
6. Olinto, G.: A inclusão das mulheres nas carreiras de ciência e tecnologia no Brasil. *Inclusão Soc.* **1**(5), 68–77 (2011)
7. Painel Lattes: Distribuição por Sexo, Faixa Etária e Grande Área de Atuação. <http://estatico.cnpq.br/painelLattes/sexofaixaetaria/>
8. Valentova, J.V., Otta, E., Silva, M.L., McElligott, A.G.: Underrepresentation of women in the senior levels of Brazilian science. *PeerJ* **5**(1), 1–20 (2017)
9. Scott, J.W.: Gender: a useful category of historical analysis. *Am. Hist. Rev.* **5**(91), 1053–1075 (1995)
10. Silva, F.F., Ribeiro, P.R.C.: Trajetórias de mulheres na ciência: “ser cientista” e “ser mulher”. *Ciência Educação* **2**(20), 449–466 (2014)