

A VIEW OF THE ACADEMIC FAMILY TREES IN BRAZIL

*Adilson Luiz Pinto**

Universidade Federal de Santa Catarina

*Thiago Magela Rodrigues Dias***

Universidade Federal de Santa Catarina

*Tales Henrique José Moreira****

Universidade Federal da Paraíba

*José Antonio Moreira González*****

Universidad Carlos III de Madrid

*Washington Luís Ribeiro de Carvalho Segundo******

Instituto Brasileiro de Informação em Ciência e Tecnologia

Abstract: The adviser-advisee relationship, which is characterized as a form of knowledge propagation, has the possible effect of promoting an increase in the scientific production of both. Specifically, in graduate programs, where this relationship is more active, either in the preparation of the thesis or in the derivation of works related to this type of study, where this is the central point of this research. In this study, orientation data extracted from curricula registered in the Lattes Platform are used, with the aim of obtaining, in addition to an overview of the orientation process in Brazil, an analysis of the bibliographic production. As a result, it was possible to observe that some individuals, especially those who work in postgraduate programs, tend to obtain a greater amount of direct guidance and, consequently, family trees with a greater number of generations. It was also verified that the areas of activity of the advisers are determinant for more dense family trees. Therefore, it is concluded that some areas such as Biological Sciences and Human Sciences hold the academic family trees with the highest number of generations and individuals, to the detriment of others such as Exact Sciences and Engineering. Innovative, the study makes genealogy into fifteen levels, a fact that previous studies only reach seven levels of guidance families. It also has a generation of computing technology and metrics to calculate everything and brings together analysis on a single platform.

Keywords: Academic genealogy; Lattes Platform; Social networks analysis; Scientometrics.

Título: PANORAMA DE LOS ÁRBOLES GENEALÓGICOS ACADÉMICOS EN BRASIL.

Resumen: La relación director-discípulo, que se caracteriza por ser una forma de difusión del conocimiento, tiene como posible efecto propiciar un incremento en la producción científica de ambos. Específicamente, en los programas de posgrado, donde esta relación es más activa, ya sea en la elaboración de la tesis o en la derivación de trabajos relacionados de este tipo de estudios, donde eso es el punto central de esta investigación. En este estudio se utilizan datos de orientación extraídos de currículos registrados en la Plataforma Lattes, con el objetivo principal de obtener, además de un panorama del proceso de orientación en Brasil, un análisis de la producción bibliográfica. Como resultado, fue posible observar que algunos individuos, especialmente aquellos que trabajan en programas de posgrado, tienden a obtener una mayor cantidad de orientación directa y, en consecuencia, árboles genealógicos con mayor número de generaciones. También se verificó que las áreas de actuación de los asesores son determinantes para árboles genealógicos más densos. Por tanto, se concluye que algunas áreas como las Ciencias Biológicas y las Ciencias Humanas ostentan los árboles genealógicos académicos con mayor número de generaciones e individuos, en detrimento de otras como las Ciencias Exactas y las Ingenierías. De innovador, el estudio hace genealogía en quince niveles, hecho que los estudios anteriores solo llegan a siete niveles de familias de orientación. También tiene una generación de tecnología de computación y de métricas para calcular todo, bien como reunir análisis en una única plataforma.

Palabras clave: Genealogía académica; Plataforma Lattes; Análisis de redes sociales; Cienciometría.

Copyright: © 2023 Servicio de Publicaciones de la Universidad de Murcia (Spain). Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

* adilson@cin.ufsc.br

** thiagomagela@gmail.com

*** tales.info@gmail.com

**** jamore@bib.uc3m.es

***** washingtonsegundo@ibict.br

Recibido: 28-02-2023; 2ª versión: 29-03-2023; aceptado: 31-02-2023.

1 INTRODUCTION

For some time the World Wide Web (Web) has been consolidating as an important source of information on different areas of knowledge. This allows for the creation of enormous data repositories. Such data, for example, may originate from tools such as digital libraries, academic social networks, and sites for individual record of academic production in which, beyond accessing their content, their users may be responsible for recording their academic data, as well as their own academic production (Dias, 2016), such as ResearchGate, ORCID, and Lattes Platform in Brazil (<http://lattes.cnpq.br>).

Data about the process of academic advising are an important object of study in the investigation of the academic education process through the study on academic genealogy, as they enable understanding and analyzing the propagation of knowledge. For Sugimoto (2014), academic genealogy is a quantitative study of intellectual heritage through the adviser-advisee relationship. For Ferreira, Furtado, and Silveira (2009), the adviser-advisee binomial (or dyad) is undoubtedly the basis for Graduate Programs (GP), which determines the growth and expansion of these courses and the advising demand. Furthermore, the authors stress that the GP student is a potential researcher in advanced development stage, that is to say, on the way to scientific autonomy, but yet dependent on a tutor, which justifies the advising activities as effectively necessary.

Family trees can be defined as a structure representing all or part of the history of ancestors of an individual. Such a structure is a graphical representation that presents in a hierarchical form the ancestors, which may or may not contain complementary information aiming to allow for a better understanding of an individual's history. Thus, the academic family trees are characterized as content that hierarchically represents the history of an adviser and his students. Consequently, characterizing an academic family tree, it is possible to observe how knowledge was passed on over time.

Bearing in mind the possibilities of access to and understanding of advising history and, therefore, diffusion of knowledge, carrying out the modeling and characterization of academic family trees arises as an interesting alternative for analyzing how Brazilian science has propagated using for this purpose, advising data.

In this context, academic family trees can be characterized, facilitating the understanding of the education process with the analysis of these trees. In general, the characterization of a tree occurs by a graph composed of a set of nodes (vertices) and connections (edges) between the nodes, which may be directed or not (Szwarcfiter, 1984). This approach had its origin in sociology, social psychology, and anthropology (Scott, 2012). The analysis of academic genealogy trees is of interest to diverse researchers, regardless of scope of practice, in that it is an aid to the understanding of these researchers' impact on their social environment.

Thus, we stress that data about advising, as well as about the individual characteristics of each of the individuals to be analyzed, are of utmost importance for the analysis and understanding of the adviser-advisee relationship. In the present research, the data used are resumes making up the Lattes Platform under the coordination of the National Council of Scientific and Technological Development (CNPq).

For this purpose, resumes registered in the Lattes Platform of individuals that have acted as advisers are analyzed with the objective of extracting information of interest and, in this way, obtaining a general view of the process of advising of this set, in which it is possible to verify how knowledge has propagated in the different areas of knowledge.

Several studies have been developed with a focus on academic genealogy and using as sources of data other web platforms. The Mathematics Genealogy Project¹ and the Neurotree² project are important sources of genealogical data for Mathematics and Neuroscience, respectively, as well as the scientific community of Physicists³ and PhDTree⁴, for the academics that hold a doctorate (Rossi y Mena-Chalco, 2015). Such projects, of utmost importance for the historical documentation of their members, are specialized in certain areas of knowledge or international data repositories. Thus, they do not represent the whole set of individuals, for example, specific to a particular country.

Thus, seeking a greater understanding of how Brazilian academic guidance has been developed in the various areas of knowledge, this research has as its main objective to characterize and analyze the academic genealogical trees with the guidance carried out in postgraduate programs based on network analysis and academic family tree based on curriculum data available on the Lattes Platform.

2 RELATED RESEARCH

Lane (2010), in an article published in the *Nature* journal, describes that measuring and assessing academic performance is already a reality. This measurement goes from a mere ranking to that that influences research funding at universities. Although important, the existing measurement systems are limited, while several problems arising from the use of current metrics are reported. The author presents a range of efforts toward constructing reliable infrastructures, which, despite useful, are laborious to keep. A good example of good practices cited is the Brazilian experience with the Lattes Platform, describing several efforts that make it one of the cleanest academic data systems that exist, providing quality data.

Mugnaini, Leitea, and Leta (2012) stress that while not holding indexing and cataloging of journals as occurs in many indexing databases, the Lattes Platform is an inexhaustible source of information about Brazilian science broadly. It is highlighted by the authors that although there is a great amount of information on the platform, few scientometric studies are observed. This reflects the limitations and difficulty imposed by the platform to the recovery and extraction of information, which proves to be an obstacle to these studies. Conversely, the great amount of personal, academic information and different types of production, accessed freely, must be a stimulus to the use of this repository by Brazilian researchers of the area. Furthermore, the fact that the Lattes Platform gathers in a single repository, information of all Brazilian scientific production, which allows for analyses that at times, would only be possible through international repositories, also stands out. It is possible to highlight diverse studies that have used the Lattes Platform as main source of data (Dias; Dias y Moita, 2022; Coimbra y Dias, 2021; Mascarenhas; Dias y Dias, 2020; Moreira; Dias y Moita, 2017).

According to Ferraz et al. (2014), currently, there is no resume database as complete as the Lattes Platform, which makes it unique. Although there are similar databases in other countries, for example, Portugal and Latin-American countries, in Brazil it is strongly used, as registration is mandatory both for the granting of funding and employment relationship officialization in degree courses. Differently from the Lattes Platform, where all resume data of an individual are recorded, most existing repositories freely provide only reference data and secondary information.

Mena-Chalco, Digiampietri and Cesar-Jr (2012) present the following two important reasons that reinforce the Lattes Platform importance: It has become a national standard (Brazil) and has been used in assessing Brazilian researchers' activities; and that researchers from the most diverse areas of knowledge are registered in the Platform.

In Rossi and Mena-Chalco's (2014) research, the authors try to characterize a select group of doctorate holders in Mathematics that obtained their degree in Brazil. Records available in the Mathematics Genealogy Project (MGP) were used in this study. The authors stress the academic advising importance for the broadening of scientific communities, directly contributing in the growth of individuals and their respective groups. Describing the community of mathematicians educated in Brazil is important for documenting history and analysis of the academic trajectory, relevance, and influence of a select academic area from Brazil.

Dores and Laender (2016) opted to use in their study the NDLTD⁵ (Networked Digital Library of Theses and Dissertations) database to provide a series of analyses from structures of generated family trees. Despite being a relevant database, it has some disadvantages relative to the Lattes Platform resumes, which are the lower number of records, as this is a database of dissertations and theses, excluding the other levels of training, and dependence upon the institutions submitting their data, different from the Lattes Platform, in which each individual is responsible for updating their data; disambiguation problem, while this also occurs with the resumes registered in the Lattes Platform, but, on a lower scale, as in the Lattes Platform there is the possibility of establishing links between individuals.

One of the first studies found that treat the academic genealogy on the basis of the Platform's data is proposed by Miyahara (2011). In the study, family trees of researchers are made considering for this purpose the advising relationships. According to the author, family trees can indicate the whole history of a given researcher. The proposed tool uses the resumes registered in the Lattes Platform for the trees characterization.

Tuesta et al. (2012) present a temporal analysis of the adviser-advisee relationship, with a case study about the productivity of researchers with a PhD in the area of Computer Science, in which they extract the data of analysis from resumes registered in the Lattes Platform. In the research, the main characteristics of the group and the co-authorship relationship are analyzed. The authors identify that the collaboration's average duration is more than three years after the first publication, concluding that the average time for the completion of a doctorate has decreased, which is possibly due to current facilities brought mostly by the Internet. Another interesting result is the fact that it was identified that

within the period of relationship, the student may be benefited to the effect that he may have a higher number of publications when the period of relationship between him and the adviser is higher.

Elias, Floeter-Winter and Mena-Chalco (2016) used Lattes Platform's data to identify genealogical graphs of researchers respected in Brazil in a certain area of knowledge. In this study, the data were collected in a semi-automatic way. The results tend to greater accuracy because the study's researchers have nationwide relevance and tend to have more updated resumes. The manual/semi-automatic process can also help correct imprecise data in the resumes but imposing limitations if considering large sets of data.

Thus, using data from resumes of the Lattes Platform for the advising process analysis, characterizing to this end trees and forest of academic genealogy, makes it possible to obtain a broad view of how the process of advising has evolved in Brazil, especially at the highest levels of training. By the analysis of the whole set of resumes, and thus proposition of techniques and methods for processing of large volumes of data, an unprecedented view of the Brazilian advising is presented, which includes details of how such a process occurs.

3 MATERIALS AND METHODS

All the procedures described below are essential to demonstrate the necessary steps to achieve the proposed objective of this study of presenting a characterization and analysis of how the guidance process occurs in Brazil, with emphasis on academic family trees, using the areas of performance of individuals as the main element of research.

The main source of data of the present study are the resumes registered in the Lattes Platform. The resumes have become a national standard in the register of the early and current lives of students and researchers from Brazil. Today they are adopted by most of the country's funding agencies, universities, and research institutes. Through its richness of information and growing reliability and comprehensiveness, the platform has become an element indispensable and compulsory for the analysis of merit and competence in decisions about funding in science and technology (CNPQ, 2017).

The Lattes Platform gained international projection in recent years, since then having great relevance for Brazilian technical-scientific production, which enabled several partnerships with countries from America and Europa, forming the ScienTI Network (Da Silva y Do Nascimento, 2006).

In possession of the resumes, it is possible to apply to them several transformations and process them in order to obtain important results. For the collection of resumes used in the present research, we used a framework called LattesDataXplorer (Dias, 2016) (Figure 1).

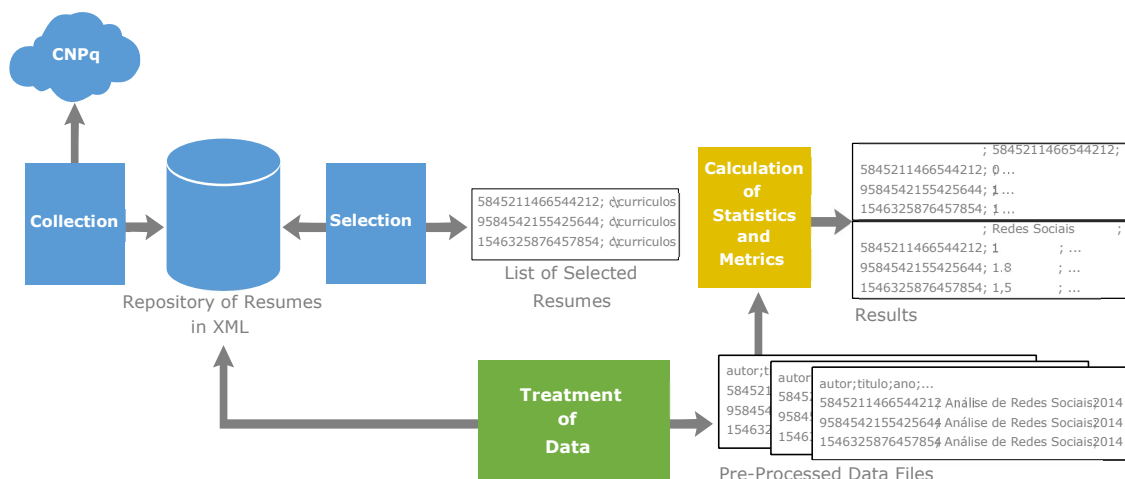


Figure 1: General view of LattesDataXplorer. Source: (Dias, 2016).

LattesDataXplorer is responsible for encompassing the whole set of techniques and methods for the collection, treatment, and analysis of data used in the present study. It is composed of several modules responsible for all the process of collection and treatment of data.

Considering the need for analyzing specific groups, such as researchers of a given institution or scope of practice, for example, a selection module was developed with the objective of producing a list of resumes relative to certain consultation criteria. This module facilitates the creation of groups on the basis of any criteria, as long as these are found in the resumes, and includes the possibility of using logic operators (AND OR) for a more specific search.

Thus, an expansion of the used framework was carried out to meet the needs of the present study (Figure 2).

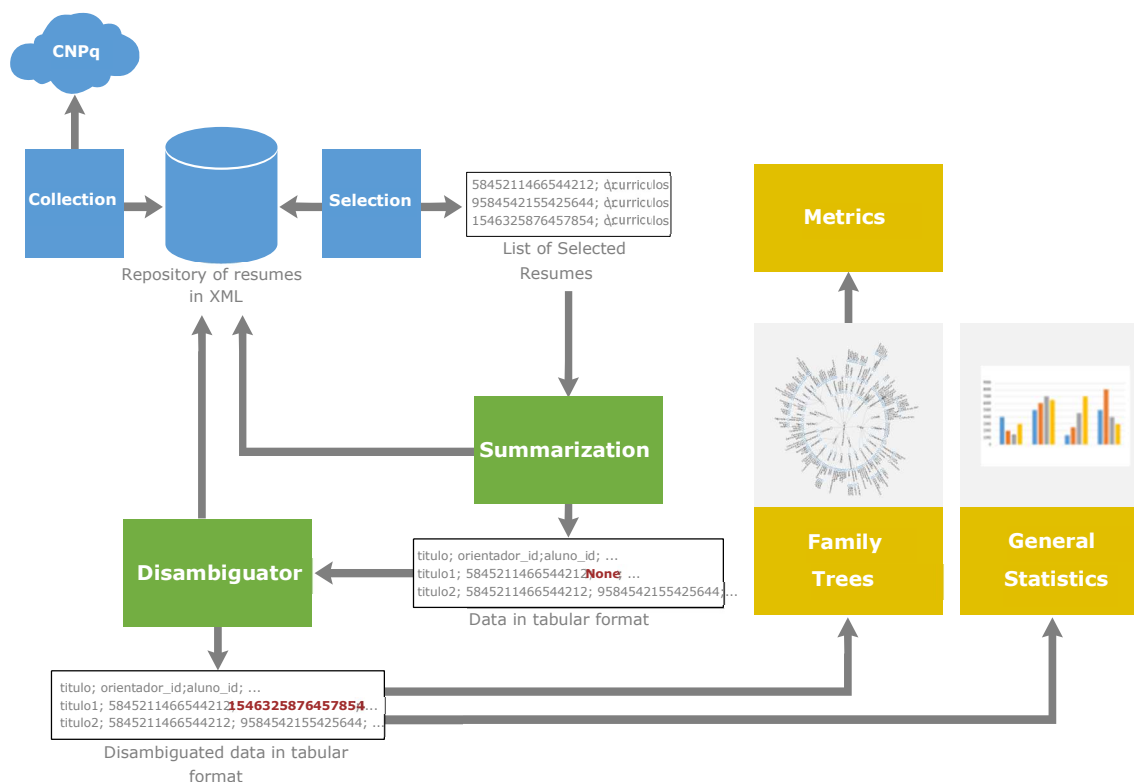


Figure 2: General view of the expanded LattesDataExplorer. Source: Data of the research.

The initially proposed framework, which encompasses the whole part of extraction and selection of data, was expanded in such a way that it started to incorporate modules for data summarization, disambiguation of advising records, calculation of general statistics, and the production of family trees, which are fundamental for the obtaining of the results presented in this work.

Such modules, absent in the original framework, are fundamental for disambiguating the data of academic advising, and consequently, producing academic family trees that include individuals before unknown.

4 RESULTS

The data here utilized were collected in September 2017 and correspond to 5,152,148 resumes registered in the Lattes Platform distributed in the most diverse areas of knowledge and the most diverse levels of training. In possession of the resumes registered in the Lattes Platform, it was possible to carry out the characterization of all advising records at all levels of advising. Among these, 481,624 individuals advised at some level of education, that is, only 9.348% of the total of individuals. The advising records considered for this characterization cover the period from 1900 (oldest registered advising) to December 2016, while this period was considered bearing in mind the collection carried out in 2017. Therefore, although the collection had been carried out in September 2017, only publications until the end of the year 2016 were considered. In this way, it is expected that all advising records until the year 2016 are inserted in the analyzed resumes.

In this section are shown data of the advisers that most advise according to the records in their resumes. Due to the quantity of advising data, which makes the forests denser, in this research only advising data from graduate courses are considered, facilitating not only the analysis but also the viewing of the trees. Finally, we used 1,196,916 advising records of the whole set identified.

Applying the Selection Module previously described for the obtaining of only stricto sensu graduate records, it is possible to rank the individuals of the set according to number of advising records. Thus, it is presented in Table I a ranking of individuals with the highest number of advising records found in their resumes, in addition to other information related to academic genealogy, such as the size of its tree and its number of generations.

N	Institution	Number of Direct Advising Records	Generations	Size of the Network
1	UFSC	399	5	1,396
2	PUC-SP	351	5	1,219
3	UFV	323	3	594
3	UFV	263	7	2,592
5	COPPE - UFRJ	259	5	1,542
6	UFRJ	246	5	1,129
7	PUC-CAMPINAS	236	8	2,132
8	PUC-SP	229	4	1,541
9	UFV	224	5	1,275
10	UFV	223	3	494

Table I: Ranking of the individuals with the highest number of graduate advising records.

University Acronyms: UFSC = Federal University of Santa Catarina; PUC-SP = Pontifical Catholic University of São Paulo; UFV = Federal University of Viçosa; COPPE-UFRJ = Alberto Luiz Coimbra Institute of Graduate Studies and Engineering Research, Federal University of Rio de Janeiro; UFRJ = Federal University of Rio de Janeiro; PUC-Campinas = Pontifical Catholic University of Campinas.

The highest numbers of advising records are not always directly related to the propagation of its knowledge to other generations. In cases such as individual 1 of Table I, we observe that there is a great volume of direct advising. Initially, one can presume that it is one of the greatest networks found. But, its family tree holds five generations of descendants, far less than the highest found, with eleven generations.

Such behavior is directly related to adviser's area of knowledge. In this context, some areas show different behaviors when assessing number of advising records. Thus, students of some areas tend to remain in the academic environment. This differs in other areas, in which their careers tend to be developed in an environment where advising is not part of the context. Generally, students of graduate courses, especially at doctorate level, tend to remain in the academic environment, which results in new advisers passing on their knowledge to other students and in the consequent increase of their advising network in new generations.

Table II presents the largest networks identified and the largest quantity of descendants, ranked by highest number of generations of descendants and both composed of many indirect advising records, resulted from the high number of generations.

Type	Major Area	Generations	Size of the Network
Highest Number of Generations	Biological Sciences	15	5,716
	Biological Sciences	15	5,442
	Biological Sciences	15	5,426
Maximum Size of Network	Human Sciences	6	8,318

Table II: Individuals with the largest networks in the set of data.

Aiming for a broader analysis, we also characterized the forest with the 100 largest trees identified in Lattes Platform's resumes. Therefore, the largest trees were taken into account bearing in mind their quantity of individuals. Thus, considering the root nodes of these trees it is possible to view their distribution through the advisers' major areas of activity (Figure 3).

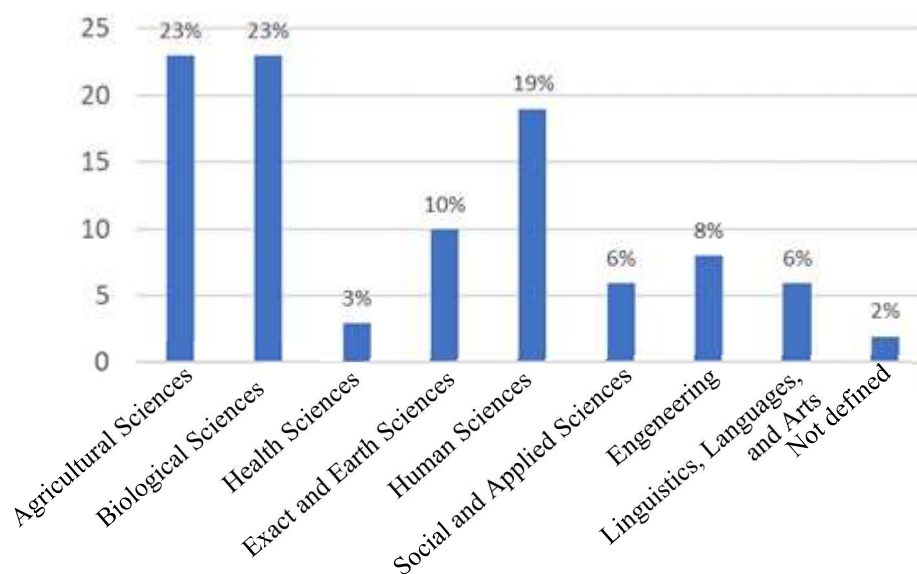


Figure 3: Distribution of root nodes through the advisers' major areas of activity.

As can be observed, the major areas of Agricultural Sciences and Biological Sciences positively stand out, representing 46% of the main roots of the 100 largest trees identified. The major area of Human Sciences also stands out. The major area of Health Sciences has low representativeness with only 3%.

Such distribution can be related to the fact that some technical areas, for example, Engineering, are characterized by having the majority of professionals with technical profiles, different from other areas, where the activity is basically academic and advising is in general frequent, especially at higher levels of training.

In addition to the distribution of root nodes concentrated on the main trees identified through the major areas of knowledge, it was also possible to identify their distribution by states of the federation (Figure 4).

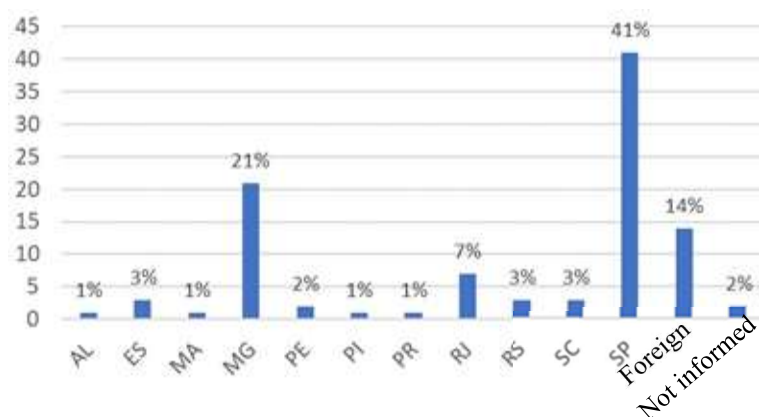


Figure 4: Distribution by state of origin of the main advisers.

Acronyms of Brazilian States: AL = Alagoas; ES = Espírito Santo; MA = Maranhão; MG = Minas Gerais; PE = Pernambuco; PI = Piauí; PR = Paraná; RJ = Rio de Janeiro; RS = Rio Grande do Sul; SC = Santa Catarina; SP = São Paulo.

The states of São Paulo (SP), Minas Gerais (MG), and Rio de Janeiro (RJ) are observed to be the states with the highest representativeness among the main roots of the largest trees. Such a fact is related to the concentration of the country's main public education institutions that concentrate the main graduate courses. Such courses, because they are traditional in their respective institutions, have a long history of education, resulting in representative propagation of their predecessors' knowledge.

Furthermore, it is noted that the states that make up Brazil’s North and Center-West regions do not have roots in the 100 largest trees making up the forest. One explanation for this low representativeness is related to the fact that there are few graduate courses in the institutions of these states, in which the existing courses are in general recent.

Beyond this distribution by states of the federation of origin of the root nodes of the main trees identified, it was also possible to verify that there exist a considerable number of researchers that are of foreign origin (Table III).

Position among the 100 largest trees	City	Country	Major Area	Area
4	New York	United States	Biological Sciences	Biochemistry
11	Quito	Equator	Agricultural Sciences	Agronomy
17	Basel	Switzerland	Biological Sciences	Biochemistry
24	Budapest	Hungary	Health Sciences	Medicine
26	Naples	Italy	Biological Sciences	Biochemistry
28	Bahia Blanca	Argentine	Agricultural Sciences	Animal Husbandry
33	Callalli	Peru	Agricultural Sciences	Agronomy
41	Leiha	India	Agricultural Sciences	Agronomy
58	Not Informed	Austria	Social and Applied Sciences	Economics
66	Buenos Aires	Argentine	Biological Sciences	Pharmacology
67	Santiago	Chile	Exact and Earth Sciences	Chemistry
69	Wazirabad	Pakistan	Engineering	Civil Engineering
73	New York	United States	Linguistics, Languages, and Arts	Linguistics
93	Marietta	United States	Biological Sciences	Ecology

Table III: Foreign advisers present in the forest.

Of the 100 largest trees identified, 14 roots are of foreign advisers. United States of America, followed in smaller quantities by Argentine, stands out among the foreign countries. Analyzing these foreign advisers, such a fact is perceived to be related to the training of excellent researchers in Brazil, who were advised by researchers of these countries, and these, in turn, have resumes registered in the Lattes Platform. Furthermore, we stress that the scope of practice of a large part of the advisers that are roots of the main trees are Biological Sciences and Agricultural Sciences.

From the analysis of the root nodes of the ten largest trees, it is possible to observe the cities, states, and countries of origin of the greatest advisers, as well as their major areas and areas of practice (Table IV).

Descendants	Generations	City	State	Country	Major Area	Area
8,318	6	Santo Antônio de Posse	SP	Brazil	Human Sciences	Education
7,162	10	São Paulo	SP	Brazil	Biological Sciences	Genetics
7,114	14	Itápolis	SP	Brazil	Agricultural Sciences	Agronomy
6,341	14	New York	NY	United States	Biological Sciences	Biochemistry
5,716	15	Rio de Janeiro	RJ	Brazil	Biological Sciences	Biochemistry
5,442	15	Rio de Janeiro	RJ	Brazil	Biological Sciences	Immunology
5,426	15	Santa Bárbara	MG	Brazil	Biological Sciences	Biochemistry
4,860	13	Oliveira	MG	Brazil	Agricultural Sciences	Agronomy
4,424	8	Morada Nova de Minas	MG	Brazil	Exact and Earth Sciences	Chemistry
4,379	10	Araguari	MG	Brazil	Biological Sciences	Genetics

Table IV: The forest’s ten largest roots.

The major area of Biological Sciences has great representativeness with six advisers (roots) among the ten largest trees. Among these, the areas of Biochemistry, Genetics, and Immunology stand out. Furthermore, we stress the area of Agronomy and Chemistry, beyond the area of Education, which contains the root node of the largest characterized tree containing 8,318 descendants.

The major area of Biological Sciences is perceived to hold the most representative advisers in the analysis of the number of generations in the ten largest trees. Of these, three related to this area contain fifteen generations.

Therefore, an important study here presented is to verify the relationship between the quantity of generations and the quantity of descendants of the largest trees (Figure 5).

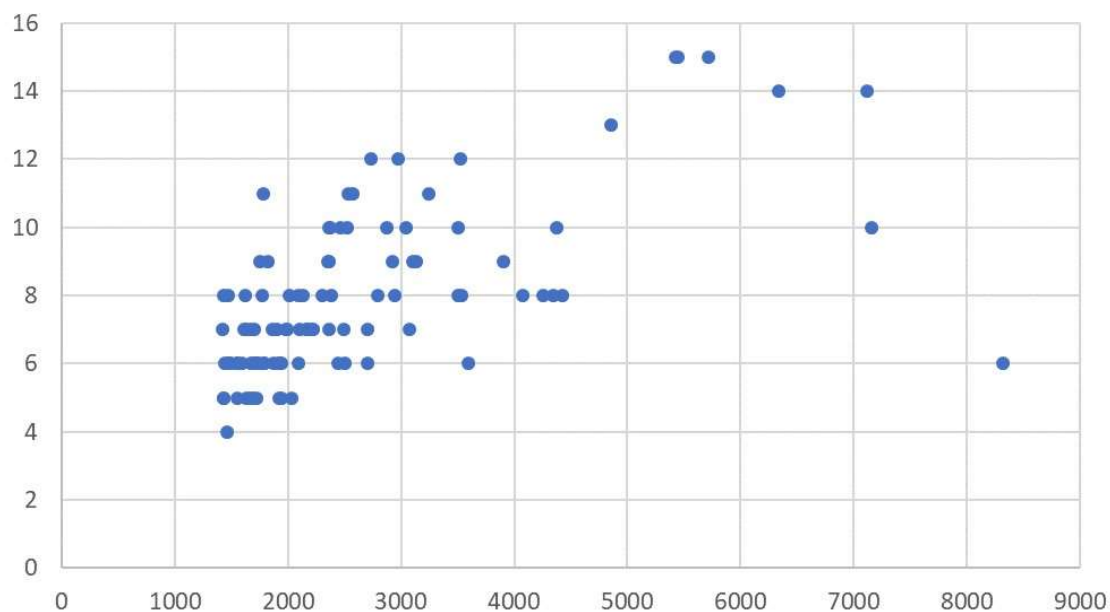


Figure 5: Relationship between the size of trees and number of generations.

We notice that in general, there is a direct correlation between the number of generations and number of individuals in the largest trees identified (0.62 Pearson coefficient), in which the trees with the highest numbers of generations have a large number of individuals. Consequently, within the analyzed set, the trees with the lowest numbers of generations hold a small number of descendants.

After the analysis of the 100 largest trees identified by the method proposed in this research, it was possible to characterize the forest that encompasses the greatest advisers of the set considered (Figure 6).

Therefore, it is possible to observe how all individuals integrating the 100 largest trees are connected. One highlights by the colors the major areas of activity of each individual, as well as their interconnections.

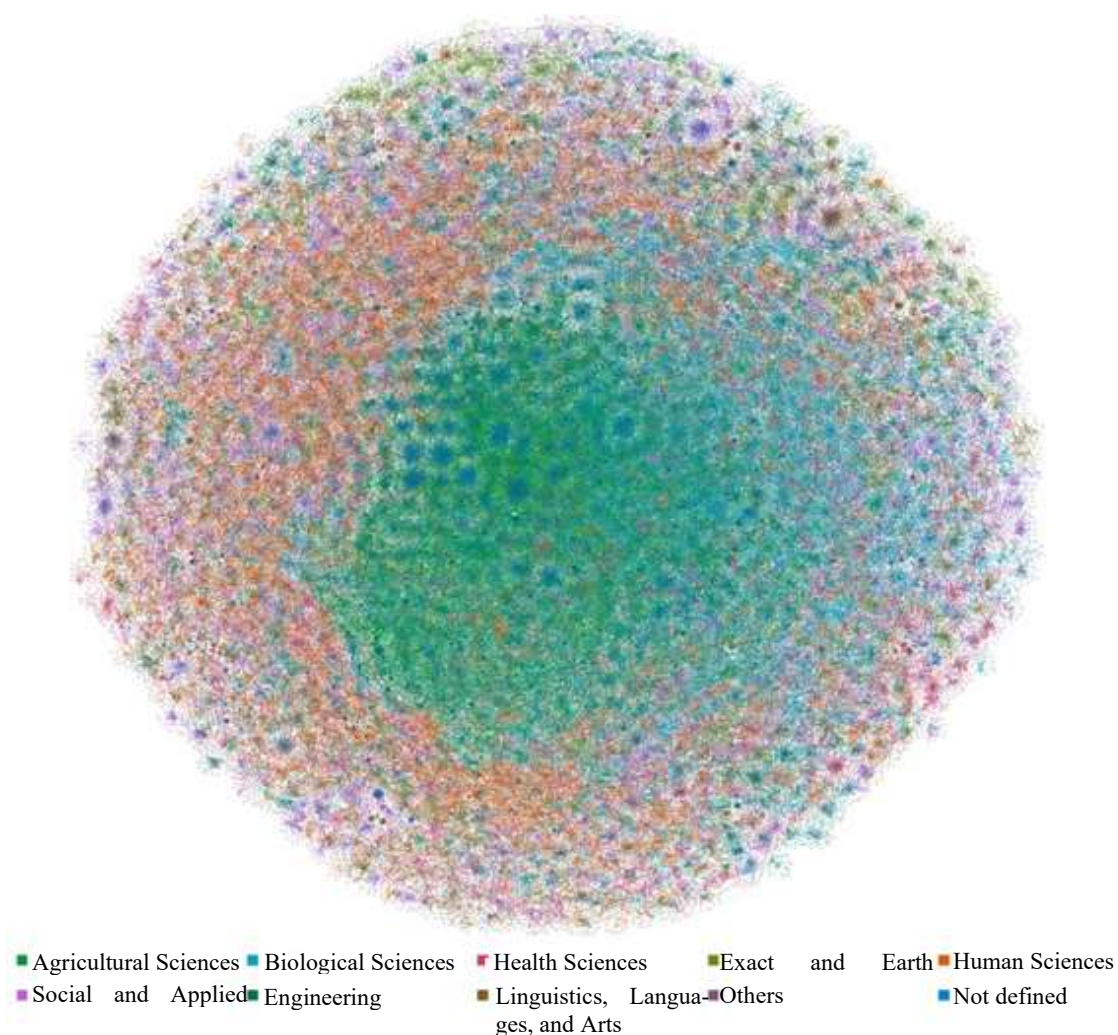


Figure 6: Forest with the 100 largest trees identified.

In the face of such characterization, it was possible to apply several classic network analysis metrics for understanding how the forest containing Brazil’s main advisers is structured (Table V).

Description	Value
Nodes	194,188
Edges	206,573
Average Degree	1.064
Diameter of the Network	15
Density of the Graph	0.00000548
Length of Average Path	3.752
Quantity of Nodes in the Giant Component	100%
Quantity of Edges in the Giant Component	100%

Table V: Metrics of the characterized forest.

The forest here analyzed is responsible for encompassing approximately 3.8% of all individuals with resumes registered in the Lattes Platform. This set has a total of 206,573 connections, which in this study mean advising records, showing, true to form, a small density of the forest (0.00000548) and average degree close to one (1.064). The diameter of the network is equal to fifteen but with a minimum average path equal to 3.75, which means that in the set analyzed, a certain individual is on average very close to the others.

Furthermore, we highlight that the giant component of the forest is responsible for encompassing all individuals selected. That is to say, when considering the 100 largest trees, no individual that makes them up is disconnected.

Thus, we observe that by verifying the largest trees identified, they are interconnected in some way, which results in a forest that contains a single connected component. This information is very relevant since we can conclude that regardless of the area of activity of the individuals analyzed, at some point there was interdisciplinary advising that resulted in the union of distinct trees.

5 CONCLUSIONS

In this study all resumes registered in the Lattes Platform were analyzed for obtaining the characterization of records of academic advising informed in the resumes not only of advisers in their sections of concluded advising, but also of advisees in the sections of academic education in the different levels of education concluded.

To achieve the general objective of this study, which is to characterize the process of academic guidance in Brazil, it was possible to verify how the academic family trees are different between the areas of knowledge. The trees in some areas closely linked to the teaching and training process tend to be much denser and with more generations than other areas with a more application bias, as expected. The areas of Human Sciences and Biological Sciences, due to their characteristics, have many more generations than the academic family trees of other areas, such as those of Engineering, in which their elements do not tend to follow an academic career and, therefore, do not have several levels of training. Consequently, the branches of these trees tend to be much smaller.

All the proposed tools use techniques with low computational cost to carry out the entire disambiguation process of the considered data. In addition, when checking the list of individuals' orientations and their academic training, this allows cross-validation of data, enabling the characterization of genealogical networks with several levels, which was not possible before.

Finally, to reach the level of this study, at least fifteen generations were considered in the genealogy, which could get more or less than seventy years of hierarchies in the training and publication segment jointly between adviser and advisee.

6 NOTES & ACKNOWLEDGEMENTS

¹ Mathematics Genealogy Project: <http://genealogy.math.ndsu.nodak.edu>

² Neurotree: <http://neurotree.org/neurotree>

³ Physicists' Scientific Community: <http://academictree.org/physics>

⁴ PhDTree: <http://phdtree.org>

⁵ Networked Digital Library of Theses and Dissertations: <http://www.ndltd.org/>

We would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - CNPq. Process: 402042/2022-0 – 200937/2022-7, Protocol 7763808095369450

Acción financiada por la Comunidad de Madrid a través de la línea de "Excelencia del Profesorado Universitario" en el marco del V PRICIT (V Plan Regional de Investigación Científica e Innovación Tecnológica)

7 REFERENCES

- CNPQ. *Sobre a plataforma Lattes*. 2017. Available at: <http://lattes.cnpq.br/> [Consulted: 28 February 2023]
- COIMBRA, F.S. and DIAS, T.M.R. Use of open data to analyze the publication of articles in scientific events. *Iberoamerican Journal of Science Measurement and Communication*, 2021, vol. 1, nº 3, p. 1-13. Available at: <https://doi.org/10.47909/ijsmc.123>. [Consulted: 28 February 2023]
- DA SILVA, L.C.R. and DO NASCIMENTO, H.A.D. Visualizando bases curriculares. 2006.
- DIAS, T.M.R. Um Estudo Sobre a Produção Científica Brasileira a Partir de Dados da Plataforma Lattes. 181 p. Tesis (Doctorado) — Centro Federal de Educação Tecnológica de Minas Gerais, 2016.
- DIAS, T.M.R.; DIAS, P.M. and MOITA, G.F. Collection and integration of scientific data for production analysis Brazilian scientific in open access journals. *Advanced Notes in Information Science*, 2022, vol. 2, p. 34-42. Available at: <https://doi.org/10.47909/anis.978-9916-9760-3-6.105>. [Consulted: 28 February 2023]

- DIGIAMPIETRI, L.A. et al. Minerando e Caracterizando Dados de Currículos Lattes. In: *I Brazilian Workshop on Social Network Analysis and Mining (BrasNAM)*. [s.l.: s.n.], 2012. p. 12.
- DORES, W. and LAENDER, A.H.F. Extracting Academic Genealogy Trees from the Networked Digital Library of Theses and Dissertations. In: *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries - JCDL '16*. New York, New York, USA: ACM Press, 2016. p. 163–166. Available at: <<http://dl.acm.org/citation.cfm?doid=2910896.2910916>>. [Consulted: 28 February 2023]
- ELIAS, M.; FLOETER-WINTER, L.M. and MENA-CHALCO, J.P. The dynamics of brazilian protozoology over the past century. *Memórias do Instituto Oswaldo Cruz, SciELO Brasil*, 2016, vol. 111, nº 1, p. 67–74.
- FERRAZ, R.R.N. et al. Análise e gestão de análise de redes de colaboração entre pesquisadores de programas de pós-graduação stricto sensu com a utilização da ferramenta computacional scriptlattes. *Perspectivas em Gestão & Conhecimento*, 2014, vol. 4, p. 133–147.
- FERREIRA, L.M.; FURTADO, F. and SILVEIRA, T.S. Relação Orientador-Orientando. O Conhecimento Multiplicador. *Acta Cirúrgica Brasileira*, 2009, vol. 24, nº 3, p. 170–172.
- LANE, J. Let's make science metrics more scientific. *Nature, Nature Publishing Group*, 2010, vol. 464, nº 7288, p. 488–489.
- MASCARENHAS, H.A.D.; DIAS, T.M.R. and DIAS, P.M. Adoption of Network Analysis Techniques to Understand the Training Process in Brazil. *AWARI*, 2020, vol. 1, nº 1, p. e004. Available at: <https://doi.org/10.47909/awari.63>.
- MENA-CHALCO, J.P.; DIGIAMPIETRI, L.A. and CESAR-JR, R.M. Caracterizando as redes de coautoria de currículos Lattes. In: *BraSNAM - Brazilian Workshop on Social Network Analysis and Minig*. [s.l.]: Brazilian Computing Society, 2012. p. 12.
- MIYAHARA, E.K. Genealogia Acadêmica Lattes. P. 27. Tesis (Doctorado) — Universidade de São Paulo, 2011.
- MOREIRA, T.H.J.; DIAS, T.M.R. and MOITA, G.F. An Overview of the Advice Process and the Scientific Production of the Adviser-Advised Relationship in the Areas of Engineering. *International Journal of Educational and Pedagogical Sciences*, 2017, vol. 11, nº 10, p. 2410-2414.
- MUGNAINI, R.; LEITE, P. and LETA, J. Fontes de informação para análise de internacionalização da produção científica brasileira. *PontodeAcesso*, 2012. vol. 5, nº 3, p. 87–102.
- ROSSI, L. and MENA-CHALCO, J.P. Aos ombros de gigantes: um estudo de genealogia acadêmica dos matemáticos no Brasil. In: *Simpósio de Pesquisa do Grande ABC (SPGABC)*. [s.l.: s.n.], 2014. p. 1–2.
- ROSSI, L. and MENA-CHALCO, J.P. Índice-h genealógico expandido: Uma medida de impacto em grafos de orientação acadêmica. In: *IV Brazilian Workshop on Social Network Analysis and Mining (BrasNAM 2015)*. [s.l.], 2015. p. 12.
- SCOTT, J. Social network analysis. [s.l.]: Sage, 2012. ISBN 1446259455.
- SUGIMOTO, C.R. Academic genealogy. In: *Beyond bibliometrics. Harnessing multidimensional indicators of scholarly impact*. MIT Press, 2014. p. 365–382.
- SZWARCFITER, J.L. *Grafos e Algoritmos Computacionais*. [s.l.]: Editora Campus Ltda. Rio de Janeiro, 1984.
- TUESTA, E.F. et al. Análise temporal da relação orientador-orientado: um estudo de caso sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação. In: *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BrasNAM)*. [s.l.], 2012. p. 11.