



Invisible Colleges and Technical Production: A Study of Patents in Brazil

Felipe C.¹, Thiago M. R. Dias¹, Patrícia M. Dias²

¹*Federal Center for Technological Education of Minas Gerais
Rua Álvares de Azevedo 400 - Bela Vista – Zip-Code: 35503-822 - Divinópolis - MG - Brazil
acad.charlesjr@gmail.com, thiagomagela@cefetmg.br*

²*State University of Minas Gerais
Av. Paraná, 3001 - Jardim Belvedere – Zip-Code: 35501-170 - Divinópolis – MG - Brazil.
patriciamdias@gmail.com*

Abstract. Analyzes of scientific collaboration networks have been extensively explored in research from different areas of knowledge, in view of their ability to identify how groups of researchers have carried out their work collectively. Such studies make it possible to identify how collaboration between individuals occurs through analyzes based on social network metrics. In this context, new studies have been proposed in order to analyze collaboration in the development of technical products, with data on patents being studied in most studies. This type of analysis is relevant because it makes it possible to understand the collaboration process in the proposal of new inventions. In this work, initially a general characterization of the group of individuals analyzed is presented, and afterwards, a global and temporal analysis of the collaboration network is performed in the proposal of patents of Brazilian individuals with curricula registered in the Lattes Platform. For that, all the patents registered in the curricula of these individuals were used for the identification and characterization of the collaboration networks. As a result, it is possible to see how collaboration in the proposed inventions of the analyzed set has been intensified over the years, with an emphasis on the institutions and areas of expertise of each inventor.

Keywords: Patentometrics, Lattes Platform, Data Analysis.

1 Introduction

The implementation of new technologies becomes a fundamental factor for the propulsion of the development and growth of a State. According to [1], the innovation process requires various types of technology and knowledge from different sources, including industry, companies, laboratories, research and development institutes, academia and consumers.

It can be seen that most of the inventions and innovations generated originate from research initiated in educational institutions, which makes them a great pole of national innovation. This factor boosted the creation of institutes that aim to protect the intellectual property generated in universities and also centers that aim to assist in the process of creating patents and in creating bridges between the academic and marketing sectors.

Currently, there are several online repositories that make it possible to search for published scientific productions, such as DBLP (Digital Bibliography & Library Project), ArnetMiner, Google Scholar and Microsoft Academic Search. It is also important to highlight the CNPq's Lattes Platform, which is an extremely

data-rich instrument for studies on Brazilian scientific and technical production. The Lattes Platform curricula present the academic trajectory of researchers and therefore are considered a national standard for analyzing their competences and merits [2]. As it is constantly used as a parameter for evaluation in research funding notices, researchers strive to keep their curricula up to date. However, even with the free availability of data and information within the scope of technical productions, such as patents, the Lattes Platform database has not yet been extensively analyzed.

The Lattes Platform has received special attention in the national academic sector as one of the main repositories for records of information from scientific publications in several areas of knowledge. Even with a high data repository, the platform only allows simple consultation of CVs, making it impossible to build analyzes of the entire platform, such as areas with the highest volume of publications and possible integrations between researchers who address topics with high synergy.

In addition to information regarding scientific productions, the platform also provides data on patents registered by employees, either of their own authorship or carried out in collaboration with companies or academic institutions. As stated [1], patents are industrial property titles on inventions, granted by the State as a reward to the inventor, guaranteeing him security in the negotiations for the sale of technologies and processes created.

In [3] a case study was developed with analysis of social networks and interviews with executives of a business group to examine a network of inventors and patents deposited between 1978 and 2008 for this group. It is a case study, with semi-structured interviews that were necessary for the first phase of the research (qualitative) and a second phase (quantitative) with a method until then little used in international business research, the use of network analysis. (ARS). The data used in the work were obtained from secondary and primary sources. The main secondary sources were internal publications by the Group and some of its main players. For the case study, semi-structured interviews were used as a primary source, carried out with a marketing executive and with a product development engineering manager. As a result of this work, the relationship between social network theory and actors involved in the generation of patents in a Brazilian multinational could be evaluated. In this way, it was possible to recognize the strategic relevance of the subsidiary and how it improved the mechanisms for generating patents and disseminated it to the parent company.

In [4] mappings of networks of inventors of a cosmetics company were made, with the objective of identifying patterns in the organization of the innovation network, in addition to identifying ways to promote the capacity for innovation through interconnectivity. The research was carried out using case study methods, and for this, mappings of inventors were made, based on previously researched patent records, taking into account the link (internal or external) with the company. of each inventor and also the number of patent citations. For data collection, the SpaceNet patent database was used. A greater hierarchy was identified in the networks with the presence of collaborators external to the company, as well as a possible greater technological content, since the number of citations was higher than that of other networks.

From this context, this work aims to broaden the understanding of nationally constructed patenting activities, seeking to assess the main actors, collaboration networks and the result that these inflict on the evolution of science. After learning descriptions, identifying patterns and generating clusters, it will be possible to present quantitative and temporal results on the patents proposed by Brazilians and registered in the Lattes Platform curricula.

2 METHODOLOGY

In this work, the analysis of data contained in the curricula registered in the Lattes Platform of several researchers from all over Brazil was carried out. The choice of the Lattes Platform for data extraction is related to the fact that it has a vast amount of data, as it deals with the integration of scientific data from curricula and institutions in the S&T area, recording academic, technical and production data. scientific research, also allowing the updating of individual data to be carried out by the researchers themselves.

To extract the data contained in the Lattes Platform, the framework called LattesDataXplorer was used

(Figure 1).

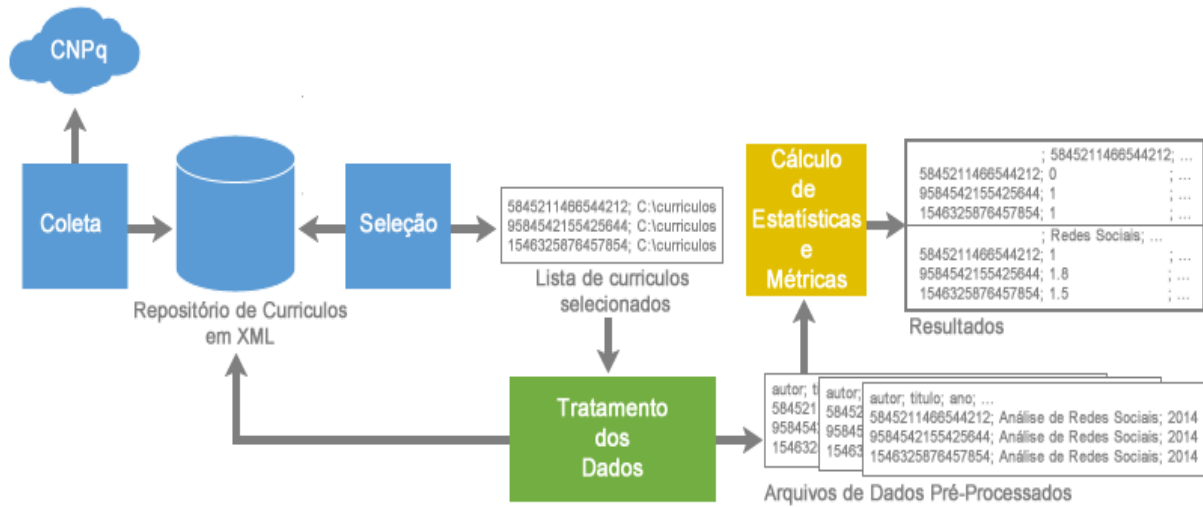


Figure 1. General architecture of LattesDataXplorer [5]

LattesDataXplorer is responsible for encompassing a set of techniques to obtain the entire set of curricula registered on the Lattes Platform. With the aid of the collection method, each of the curricula is accessed and later retrieved for local storage. In addition to the collection method, it also encompasses another series of methods that allow the performance of various bibliometric analyzes having the curricula as a data source.

With all the curricula stored locally in .XML format, the dataset goes through some phases that make it possible to achieve the expected results (Figure 2).

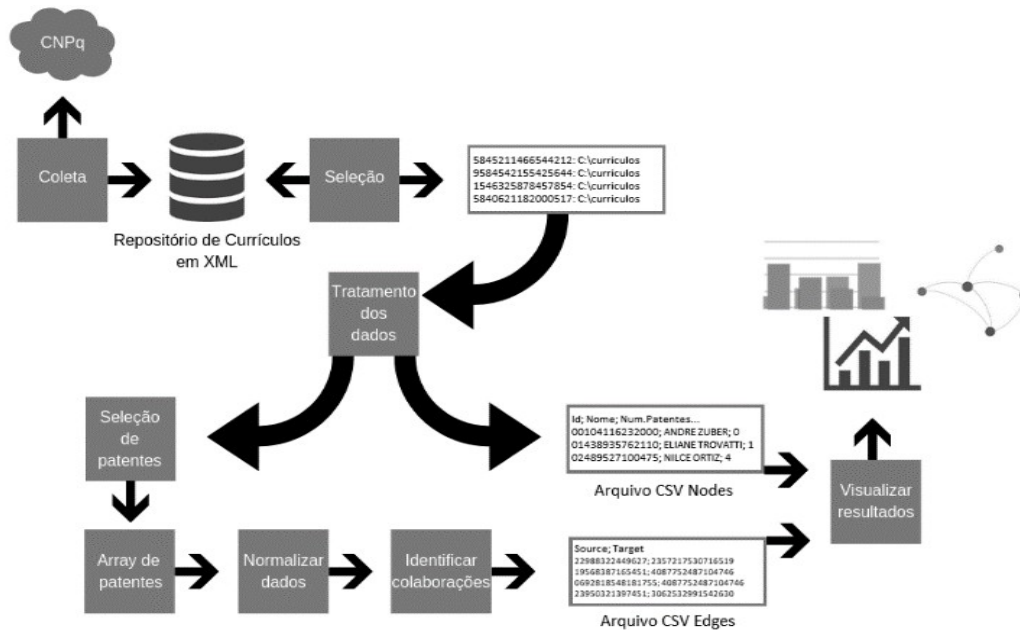


Figure 2. Steps in the network characterization process

After obtaining the data, they were processed in order to characterize the collaboration networks in patents between researchers. This step was carried out by a script developed in python language, a language that allows easier writing of scripts by presenting several modules with specific utility for each case and an easy and malleable syntax. The steps performed are basically:

1. Select patents that were published within the period of time - informed directly in the code - that is being analyzed. Throughout this process, data is also collected from all researchers who have filed patents.
2. Create a list whose index is the name of the patent, and the value is the set of identifiers of all its depositors. In this way, it becomes possible to identify collaborations between researchers.
3. Normalize patent names before using them as indexes of the list, thus avoiding data redundancy. As an example, there is the case of the same patent being registered in two curricula with slightly different names. In this case, normalization would cause the match between the two records to be identified. This result is achieved through several processes, such as radicalization, removal of “stop-words” and transformation of all letters to capital letters.
4. Identify collaborations after the patent list has been generated. An array is created containing, in each position, two identifiers, indicating a collaboration between A and B. To assemble this array, all permutations of two elements, without repetition, of the set of identifiers of depositors for each patent are obtained.

The processed information was stored in .CSV files. To obtain a visualization of these data, the software Gephi was used, which helped in the construction of graphs, as well as the matplotlib library of python to generate graphs.

3 Results

In view of the data analyzed in this work, it was possible to quantify the evolution of Brazilian technical production, specifically, the patents registered in the curricula registered in the Lattes Platform in a temporal way (Figure 3).

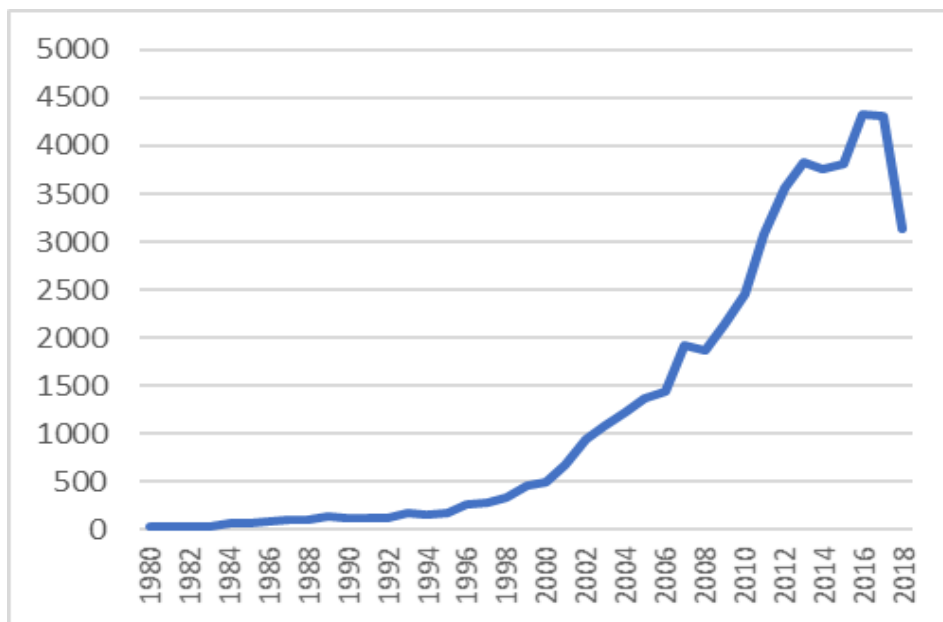


Figure 3. Time evolution of the patent proposal

As can be seen, from the year 2000 onwards, there has been a constant growth in the number of registered patents, with a small reduction in the years 2007 and 2012. Afterwards, there is a considerable drop from 2017 onwards, which may be justified by a possible lack of updating of curricula, given that data collection was carried out in 2019. This phenomenon has also been identified in the production of scientific articles, as presented in [5].

In order to assess collaboration in the patent proposal, it was necessary to characterize the technical collaboration networks. Therefore, the strategy devised by [6] was applied to patent titles in order to identify technical collaborations. As a result, the networks were characterized considering all recovered patents.

In Figure 4 it is possible to observe the institutions with the highest intensities of collaborations (degree). For this, we considered those institutions that collaborate directly with others in the elaboration of patents, regardless of the period of collaboration. Therefore, for the analyzes carried out here, the current institutional link of the individual registered in his Lattes Platform curriculum was considered.

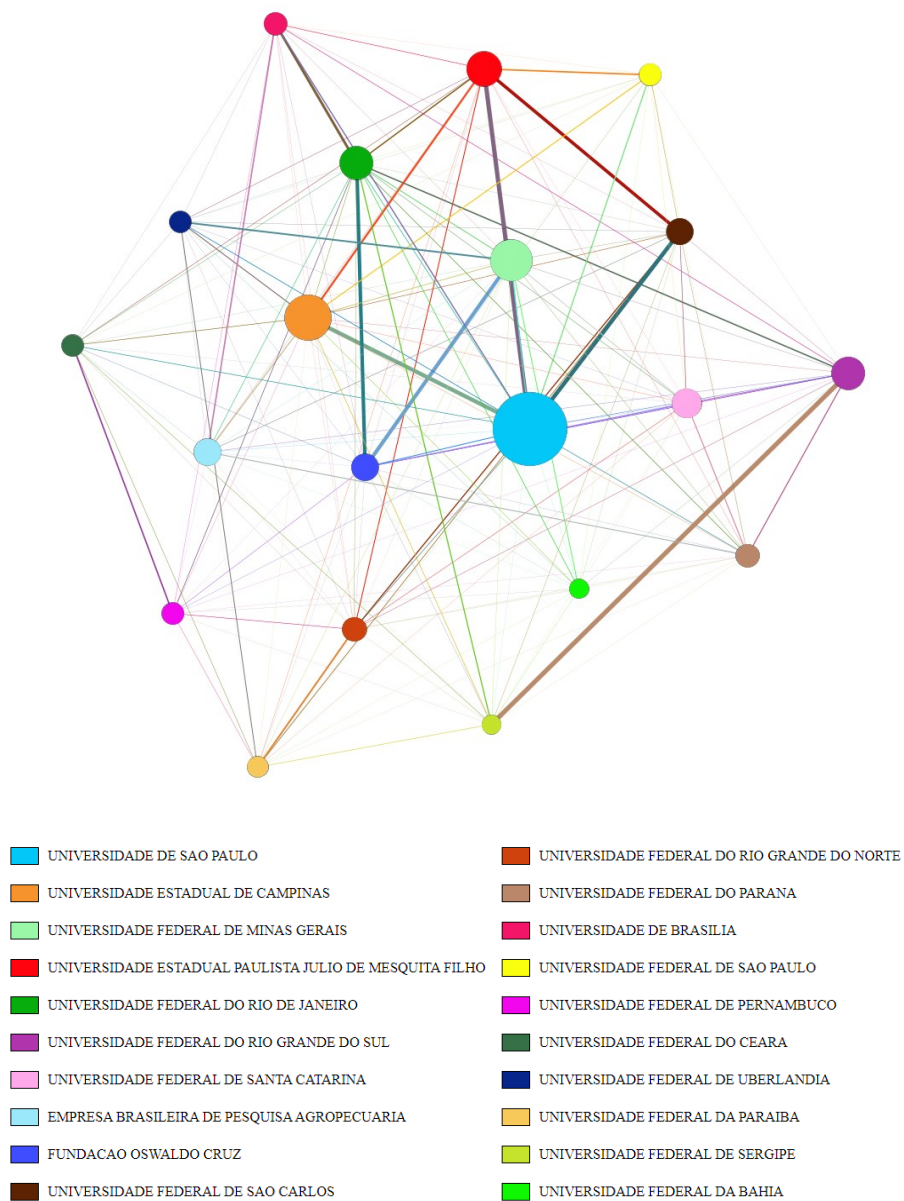


Figure 4. Characterized collaboration network

As can be seen in this network, the diameter of the vertices characterized by the institutions of the proponents takes into account the number of patents that their researchers have registered in their curricula. Therefore, the University of São Paulo, State University of Campinas and Federal University of Minas Gerais stand out with the largest number of patents, institutions that also appear as the most collaborative.

Such values are directly influenced by highly productive individuals as already presented. In view of this, new studies that evaluate in detail such individuals may provide interesting results about the patent registration process in Brazil.

4 Conclusion

It was possible to observe that the number of patents has been growing steadily over the last few years, except in the years 2017 and 2018, which may be influenced by the lack of updating of some curricula. In addition, it was possible to observe the most representative institutions, and also, how they have collaborated with other institutions in the registration of patents. It was observed that collaboration happens more intensely between institutions that are geographically close and that collaborate in the proposal of scientific works or in the academic training process of their researchers.

References

- [1] Amadei, J. R. P., & Torkomian, A. L. V. (2009). As patentes nas universidades: análise dos depósitos das universidades públicas paulistas (1995-2006). *Ciência da Informação*, 38(2), 9-18.
- [2] Amorin, C. V. (2003). Organização do currículo: Plataforma Lattes. *Pesquisa Odontológica Brasileira, SciELO Brasil*, v. 17, p. 18-22
- [3] João, B. D. N. (2009). Redes em subsidiárias de multinacionais: um estudo de caso com análise de redes sociais de inventores e patentes. *Revista de Administração Pública*, 43(5), 1037-1066.
- [4] Serrano, B. P.; & Junior, J. A. G. (2014). Redes de inovação: mapeamento de inventores de patentes em uma empresa do setor de cosméticos. *Revista GEPROS*, v. 9, n. 1, p. 101.
- [5] Dias, T. M. R. (2016). Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes. 2016. Tese (Doutorado em Modelagem Matemática e Computacional) CEFET-MG, Minas Gerais.
- [6] Dias, T. M. R., Moita, G. F. (2015). A Method for Identification of Collaborations in Large Scientific Databases. *Em questão. Porto Alegre*, v. 21. N. 2, mai./jul.