Technological Forecasting & Social Change 182 (2022) 121876

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

An exploratory analysis of Brazilian universities in the technological innovation process

Eron Passos Andrade [a,b,*], Jadiel dos Santos Pereira [a], Angela Machado Rocha [b,c], Marcio Luis Ferreira Nascimento [b]

[a] CETENS-Energy and Sustainability Science and Technology Center, Federal University of Recôncavo da Bahia, Rua Godofredo Rebello de Figueiredo Filho, 697, SIM, 44085-132 Feira de Santana, BA, Brazil

[b] PEI-Graduate Program in Industrial Engineering, Polytechnic School, Federal University of Bahia, Rua Aristides Novis, 2, Federação, 40210–630 Salvador, BA, Brazil

[c] ProfNIT-Graduate Program in Intellectual Property and Technology Transfer for Innovation, Federal University of Bahia, Rua Barão de Jeremoabo, 147, Campus Universitário de Ondina, 40170-115 Salvador, BA, Brazil

A R T I C L E I N F O

Keywords:
Technological innovation
Intellectual property
Universities
Principal component analysis
Similarity
Hierarchical clustering

A B S T R A C T

The purpose of this paper is to use hierarchical clustering (HC) and principal component analysis (PCA) for determining the key institutions and variables in a multidimensional data set to visualize Triple Helix (TH) relationships between industry, academia and government. This is a huge task, essential to better understand technological innovation, an interactive process that creates knowledge in an integrated way, reducing the number of variables. For this task we analyzed the data from eight Brazilian universities between 2008 and 2015 considering median of twelve parameters so diverse as the number of research groups; researchers; teaching staff; innovation projects in collaboration; papers; patents; technology transfer agreement; money generated from technology transfer and financing. From HC it was possible to identify four main university clusters considering all variables. PCA also shown four groups on main component mapping, in agreement with HC. The financing, the existence of innovation environments and specific innovation legislation, and the regional context explain clustering. PCA suggests that much of the data variability can be summarized in three principal components, presenting industry, academia and government interrelationships, in agreement with HC. So PCA and HC could be considered as a new view of investigation to quantify the TH, statistically mapping this model.

1. Introduction

Given the growing incentive for technological innovation in the modern economy and the greater involvement of public universities in the process, studies that contribute to a better understanding of the topic are justified. Eveleens (2010) highlighted the growing trend of the emergence of models that seek to describe the technological innovation process, its implications and patterns. An approach using multivariate analysis is gaining prominence in the literature on the subject in works such as those by Coccia (2005); Hardeman et al. (2013); Loi and Di Guardo (2015); Zhao et al. (2015); Jovanović et al. (2019); Alnafrah and Zeno (2020); Li et al. (2020).

To cluster and classify national innovation systems, Alnafrah and Zeno (2020) used machine learning classification and principal component analysis (PCA). This study included 36 indicators from 54 countries, which are divided into six groups, that represent the different national system of innovation dimensions. To analyze European research and innovation policy discourse, Hardeman et al. (2013) developed a framework for the analysis of 40 national research systems using PCA to aggregate four variables into one composite indicator. In Italy, Coccia (2005) used PCA analysis to pinpoint the main typologies operating in the national system of innovation, the author used ten variables to analyze the public research bodies of the National Research Council of Italy. These three works are focused on national systems of innovation and not on the relationships between the Triple Helix (TH) actors.

In the context of an increasing demand for knowledge to support technological innovation, the TH relationships between industry, academia and government are mode of collaboration that is seen as important interorganizational knowledge network (Etzkowitz and

* Corresponding author at: PEI-Graduate Program in Industrial Engineering, Polytechnic School, Federal University of Bahia, Rua Aristides Novis, 2, Federação, 40210–630 Salvador, BA, Brazil.

E-mail address: eronpassos@ufrb.edu.br (E.P. Andrade).

https://doi.org/10.1016/j.techfore.2022.121876
Received 18 August 2021; Received in revised form 11 June 2022; Accepted 11 July 2022

Leydesdorff, 1997; Etzkowitz, 2008; Etzkowitz and Zhou, 2017). Universities, when creating interorganizational ties, have a greater mobility of knowledge in terms of their ability to establish collaborative Research and Development (R&D) relationships (Petruzzelli et al., 2010). These can promote the creation of partnerships to leverage the joint development of new technologies and innovative solutions to products, services and process, bringing benefits to the university, industry and government (Philpott et al., 2011; Perkmann et al., 2013; Etzkowitz and Zhou, 2017).

The positive impact of research carried out at universities on industrial productivity and technological innovation in different sectors and in different countries has been largely demonstrated in recent times (Carayannis et al., 2014; Romano et al., 2014; Sydow et al., 2016; Sá et al., 2018; Oliva et al., 2019; Dooley and Gubbins, 2019; Johnston, 2020). These studies were born from the awareness of the dissemination of an academic model, not only oriented to the creation and diffusion of knowledge, but to the promotion of an entrepreneurial culture; exploring opportunities for innovation and development; strengthening ties with industries, government and other local organizations; support of new businesses; exploration of the results of scientific research; and the growth of the territory in which it operates (Del Giudice et al., 2012).

Just for comparison, as far authors know, Loi and Di Guardo (2015) and Zhao et al. (2015) performed a research using hierarchical cluster analysis (HC) using dendrograms. Loi and Di Guardo (2015) analyzed 75 Italian universities for a qualitative classification and found four clusters based on computed keyword frequencies. They used a qualitative content analysis, this study has some of the weaknesses that usually characterize qualitative research, especially regarding subjective interpretations or unlikely replications.

Zhao et al. (2015) considered eight dimensions (or variables) for regional classifications of universities, firms and research centers in China, finding six clusters of regional innovative collaboration. The authors used ordinal multidimensional scaling (OMDS), however the interpretation in OMDS method is based on visual inspection and they work with the projections of the solution on bi-dimensional sub-spaces. It should be noted that two-dimensional maps are not easy to interpret as the points are positioned on the map by means of a projection therefore two points that appear to be physically close to each other in the two-dimensional representation may not necessarily be close to each other in space. Thus, to assess real proximity between two points in the space the authors used HC.

Also, in China Li et al. (2020) designed a questionnaire using a five-point Likert scale and used proprietary survey data collected from industrial companies, universities and government agencies to examine the influence of the Triple Helix system of collaboration on regional entrepreneurship in five regions of China. Finally, they applied the confirmatory factor analysis to evaluate convergent validity of measures. The problem with that, according to Fávero and Belfiore (2019), is the use of Likert scale statistics as input to quantitative analysis – which should be avoided.

Jovanović et al. (2019) applied a multivariate method named two-step Composed *I*-distance, based an objective set of 20 indicators, to rank 32 OECD country members according to their socio-economic development. But they present an approach to measuring TH management performance and measured activities that are related to one aspect of R&D activity.

However, it is still noticeable that the identification of the most appropriate model for each situation in an incipient stage (Jovanović et al., 2019). In other words, there was a need to find a new analysis framework to explain the way in which new technological innovations emerge and diffuse and how these innovations influence the overall socio-economic performance. Briefly speaking, it is still difficult to observe quantitative relations between industry, academia and government helices.

The purpose of this paper is use HC and PCA for determining the key institutions and variables in a multidimensional data set and visualize TH relationships between industry, academia and government. So, this work aims to use a quantitative approach to better visualize such relations in the light of TH proposal and characterize the dynamic behavior of university parameters based on statistical tools, trying to mapping TH. They are exploratory multivariate statistical techniques for simplifying complex databases (Hair et al., 2019) as for example data from eight Brazilian universities, thus permitting a characterization of their underlying dynamical behavior as visual maps.

The present procedure is different from previous works. Note that, unlike the others that opt for one or another analysis, we used two multivariate analyzes (PCA and HC) in a combined way to investigate interactions between universities, industry and government under the light of TH theory. In addition to increasing the security of the results, this can be considered as a methodological contribution to prior literature, considering both as integrative views to TH model, as described below.

Both PCA and HC analyzes are unsupervised pattern recognition methods used when one has a large amount of data. However, this excess of data ends up making treatment (processing and storage) difficult. When these methods are used in combination, the main advantage is the elimination of most of the experimental noise (Hair et al., 2019). This is possible because the noises constitute a population of random errors, that is, they are not correlated with the information contained in the data matrix. Providing a dataset with relevant information, that is, only useful data for analysis.

For this task we considered only quantitative parameters as follow: the number of research groups; researchers; PhD researchers; teaching staff; teaching staff with PhD; innovation projects in collaboration; papers published; patents filed; patents granted; technology transfer agreement; money generated from technology transfer and financing. This approach is also new for universities in Brazil, but it can be used in any country to analyze the interactions of any actors within the scope of TH.

This paper is divided into six parts, the first one presents the paper and explain why this work is done. The second contextualizes the university in the process of technological innovation. The third shows the sequence of steps taken for the analysis. The fourth part describes the results, the fifth presents the discussion and the last part contains the conclusions.

2. Universities and the technological innovation process

The ancient word *tekhnología* comes from two Greek words: τέχνη (*tékhnē*, "craft", "skill", "art") and *λογία* (logía, "study"), that is related to the first ways of weaving or fabricating things as clothes, and is still recognized as one of the first skillful processes. In the modern economy, knowledge has been consolidated as an intangible factor that determines economic growth and, thus, it is possible to observe the emergence of new concepts of knowledge and its dissemination through the technological innovation process (Prahalad and Hamel, 1990; Nonaka and Takeuchi, 1995; Etzkowitz and Leydesdorff, 1997; Prusak, 1997). In this sense, universities are considered key institutions in the knowledge economy (Philpott et al., 2011; Perkmann et al., 2013; Etzkowitz and Zhou, 2017).

Rosemberg (1974) portrayed the technological innovation as a process as a set of activities linked through complex feedback loops named chain model. According to the author technological innovation was seen as a process of trial and error, the result of an interactive and collective process within a set of connections between people and institutions that evolve time.

Technological innovation is a dynamic and interactive process in which people, units, groups, organizations or nations influence each other at sectoral, regional and national levels to create new knowledge essential to be more productive and efficient (Balle et al., 2019; Alnafrah and Zeno, 2020; Johnston, 2020; Li et al., 2020; Basso et al., 2021). Far from reflecting any kind of automaticity, its achievement involves investments and deliberate efforts in the construction of technological capabilities (Sá et al., 2018; Dooley and Gubbins, 2019; Oliva et al., 2019; Johnston, 2020). The definition still in construction, due to ingenious new procedures promoted both by academy and industry, but not limited by them.

The growing interest about the connections that lead to knowledge and to the interaction between different institutions concerning technological innovation process were the bases of the National System of Innovation (NSI), a conceptual structure proposed by Freeman (1987), Lundvall (1992), Nelson (1993) and Edquist (1997). NSI is a multilevel concept (Alnafrah and Zeno, 2020), where national, regional and sectoral innovation system can coexist and coevolve together in the same country (Coccia, 2005; Hardeman et al., 2013; Alnafrah and Zeno, 2020).

Lundvall (1992) has provided the most clarifying definition of NSI. He was the first to include in it not only organizations directly involved in the innovation process but also all the aspects of institutional structures that influence learning, accumulation of knowledge and the search for technological innovation. The complex network of elements operating within a NSI can be interpreted referring to the Triple Helix (Coccia, 2005).

The TH model represents the creation of arrangements between government, industry and academia that promote better conditions for the existence of cooperation in the search for technological innovation. The interaction between the three helices characterizes a recursive process, that is, it repeats itself infinitely to the point of being represented by a spiral (Etzkowitz, 2008; Etzkowitz and Zhou, 2017).

The government formulates public policies to encourage technological innovation, provides resources and funds on scientific and technological research, promoting the reduction of uncertainties in the macroeconomics and stimulating other agents to invest in innovation. In addition, it can create institutions that regulate the productive and financial sectors and also promote the use of fiscal, monetary and exchange rate policies in favor of the process of technological innovation (Etzkowitz, 2008; Etzkowitz and Zhou, 2017).

Industry is directly responsible for technological innovation, that is, for the practical application of knowledge. Therefore, it is necessary to produce knowledge internally, or, when necessary, to seek from outside agents the necessary information to create a place for generating ideas and new knowledge that support technological innovation. It is up to them, among other activities, to capture the scientific and technological knowledge generated in universities and other research institutions, to produce and commercialize, offering to society new products, processes and services, generating economic benefits (Etzkowitz, 2008; Etzkowitz and Zhou, 2017).

Academia is responsible for human resources training, conduct scientific and technological research and developing prototypes of innovative technologies. They are source of knowledge and technology from which they can originate the knowledge and technology transfer process to the industries (Del Giudice, 2008; Etzkowitz, 2008; Etzkowitz and Zhou, 2017).

The engagement of universities in technological innovation activities in Triple Helix began with the establishment of collaborations between government, industry and academia through the creation of joint research projects between the public and private sectors (Etzkowitz and Leydesdorff, 1997; Etzkowitz, 2008; Etzkowitz and Zhou, 2017). According to this approach, universities not only bring new knowledge and technology, but do so as an economic perspective (Coccia, 2005).

The new task of the university is to act as an economic performer, creating and promoting an entrepreneurial culture, exploring opportunities for technological innovation and development, establishing stronger ties with industry, government, institutions and other actors in the territory, creating conditions for the beginning of activities of new ventures, taking technology from its limits, valuing and exploring scientific knowledge (Del Giudice, 2008; Etzkowitz, 2008; Philpott et al., 2011; Perkmann et al., 2013; Etzkowitz and Zhou, 2017). These activities were generally supported by Industrial Liaison Offices (ILO), which are institutions responsible for organizing the interaction between a department or a research unit and a group of interested industries (Carayannis et al., 2014; Romano et al., 2014).

In addition, the universities began to reap the benefits of such contributions, organizing activities for the exploitation of their intellectual property (Del Giudice, 2008; Petruzzelli et al., 2010; Romano et al., 2014; Philpott et al., 2011; Perkmann et al., 2013; Sydow et al., 2016; Sá et al., 2018; Dooley and Gubbins, 2019; Oliva et al., 2019; Johnston, 2020). That is the reason why patent policies were implemented and Technology Transfer Offices (TTO) were created in academic institutions (Romano et al., 2014; Wang and Lu, 2021). The main objective of a TTO is to facilitate the transfer of technologies developed by university to market, protecting intellectual property through patents and copyrights and, subsequently, licensing protected intellectual properties to companies outside university (Soares and Torkomian, 2021).

The most recent Triple Helix thesis is that universities are no longer playing a secondary role in the process of technological innovation in providing higher education and scientific and technological research but are assuming a primary role as a generator of new companies. As industrial societies are supplanted by knowledge, it is increasingly applied simultaneously in theory and practice, due to its multipurpose nature. Knowledge and technology transfer based on academic discoveries take place throughout their inventors, giving them the possibility to participate in both research and process of technological innovation (Etzkowitz and Zhou, 2017).

The TH is a conceptual framework on the interaction between three NSI actors (government, industry and academia) to drive the dynamics of technological innovation. The theoretical feature is being criticized for its purely conceptual approach. According to Jovanović et al. (2019), TH model as an analytical tool for a systemic explanation of dynamic and complex government, industry and academia relations could be beneficial, but concrete quantitative approach is a necessity.

Efforts to insert a quantitative approach in TH model according to research carried out by Zhao et al. (2015), Jovanović et al. (2019) and Li et al. (2020) mentioned above, in the introduction. Zhao et al. (2015) noted that the support from the government is important to fuelling regional innovation systems, as R&D and technological innovation is often capital intensive. They also realized that research institutes generated the highest research output (patents and creation and dissemination knowledge), so regions with a higher number of research institutions tend to generate better innovation outputs. Finally, different regions present different organisational mindsets, and the competitive regions are also the ones presenting more entrepreneurial behavior. These are the regions that present a business orientated mindset.

The key implication of study done by Zhao et al. (2015) was that collaborations can bring wealth to the regions that are engaged, yet all collaborations have to compromise the motivations of the different TH actors that pursue innovation and balance the allocation of resources. In order to exploit these differences, policy makers should engaged in an audit of skills, capabilities and capacities over all actors. According to Jovanović et al. (2019), R&D activities are marked as the catalysts of TH performance that foster and boost the third mission of the universities. They also realized that collaboration has the leading role among the three actors, while the main role among indicators is reserved for the technology balance of payments revenues resulting from patent activities, which reflects the significant result of innovative activities of the country.

Li et al. (2020) results highlight the importance of integrating university knowledge, industry needs, and government resources to ensure effective flow and reasonable configuration of entrepreneurship resources, to foster a more supporting entrepreneurial environment, and to ultimately stimulate innovation and entrepreneurship. Li et al. (2020) suggested that it is better for regional policymakers to tailor their designs of regional policies and strategies to be attuned to and embedded in the specific needs and resource availability of respective regions. Only in this way, they can establish a region-specific approach to better utilize localized capabilities with a path-dependent development.

The growing importance of the TH model has led to the emergence of a body of theoretical and empirical research to discuss other innovation models, such as the Quadruple and Quintuple Helix. The Quadruple Helix adds civil society as a fourth helix, specifically defined as culture-based and media-based public. The Quintuple Helix visualizes the environment, combines knowledge, know-how, and the natural-environment-system together in collective interaction (Carayannis and Campbell, 2010).

Authors such as Carayannis and Campbell (2010) and Bikse et al. (2016) point out that it is necessary to improve networking with different institutions and society, ensuring the protection of the environment, applying the Quintuple Helix model, providing eco-innovation and eco-entrepreneurship that must be understood in a broader understanding of knowledge and innovation.

Despite the rise of studies on new theoretical approaches to helices, still difficulty in understanding how the new helices are represented. The adding of the fifth helix of the natural environment to knowledge production is too ambiguous to measure empirically. This raises questions about empirical applications due to its theoretical ambiguity and measurement issue (Yoon et al., 2017). In agreement with Yoon et al. (2017), we decided to use the TH model in this work.

2.1. Brazilian universities as drivers of technological innovation

Brazilian public universities are also undergoing profound changes, especially with regards to their traditionally established teaching and research functions. Such agents are also recognized as drivers of technological innovation, as they concentrate most of R&D infrastructure (Cross et al., 2017; Dalmarco et al., 2019; Marques et al., 2019; Basso et al., 2021). However, its recognition in the technological innovation process is quite recent when compared to universities in Europe and the United States.

Since the 1980s, Brazilian public universities have been important players in Science and Technology Parks. They are formed by geographically close organizations with an entity in charge of building and managing common areas (Balle et al., 2019). Through planned and structured actions, they bring companies and academic institutions, such as universities, to promote organisational performance and employment, as well as science, technology, innovation and entrepreneurship (González-Masip et al., 2019). These parks promote advantage for organizations, regions and even nations on improving new technologies (Carayannis et al., 2014).

The institutionalization of the Technological Innovation Centers (NIT), a structure similar to Technology Transfer Offices (TTO), is a legal obligation for all Brazilian universities since Law n°. 10.973/2004, that aims to centralize an institutional technological innovation policy. When dealing with the process of creation, protection, negotiation and commercialization, NIT becomes responsible for establishing a favorable environment. Other NIT functions are to promote dialogue between university and productive sectors and to disseminate the portfolio of technologies (Castro and Souza, 2012; Soares and Torkomian, 2021).

In 2007, Brazilian universities were benefited from Support Program for the Restructuring and Expansion of Federal Universities (acronym for REUNI: *Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais*). This program aimed to create conditions for the expansion of higher education, by making better use of the structure and human resources existing in federal universities and the acquisition of new resources. In the REUNI implementation period, between 2008 and 2012, there was a considerable increase in the number of permanent teaching staff with the authorization of 21,786 new contracts and 64 % reduction in adjunct professors, mainly due to temporary contracts as well as retirements. There was also an increase of approximately 22 % in the number of visiting professors (Cross et al., 2017; Souza et al., 2015).

The number of teaching staff with PhD increased from 50.95 % to 68.78 % between the years 2003–2012. Numbers of teaching staff with just a masters fell 25.45 % while those with PhD increased by 68.78 %. This growth was a strategy to qualify teaching staff in higher education, since hiring PhD guarantees more committed teaching, research and extension (Cross et al., 2017; Souza et al., 2015). The observations made by Navas et al. (2020) for the Colombian higher education system, showing that PhD teaching staff had the greatest impact on research efficiency, they recommend that universities specialized in research should increase the number of PhD professors.

In Brazil, only in 2015, the NSI was officially created, which became known as the National Science, Technology and Innovation System (*Sistema Nacional de Ciência, Tecnologia e Inovação* - SNCTI). The Triple Helix model was used to create the SNCTI. In the SNCTI, universities were classified as Science, Technology and Innovation (ST&I) operators, and their leading role is in creating innovation ecosystems involving and promoting integration with other actors (Dalmarco et al., 2019; Basso et al., 2021).

This denotes a differentiation of the SNCTI compared to the NSI countries of the Organization for Economic Co-operation and Development (OECD). It is worth remembering that, in Brazil, the leading role in the activities of developing new technologies, particularly in areas recognized as high technology, belongs to universities, more than companies. In this context, when the number of companies that develop activities focused on innovation is still small, universities assume an important strategic role in terms of scientific and technological production (Dalmarco et al., 2019; Basso et al., 2021).

The aforementioned changes are helping to consolidate Brazilian universities as drivers of technological innovation. Along these lines, Dalmarco et al. (2019) carried out a comparative study between the agricultural and aerospace sectors in Brazil and the Netherlands, from the perspective of the Triple Helix, and noticed limited alignments between academia and industry, which continue to develop their activities separately, working together only in a limited and sporadic way. The authors point out that government initiatives, in turn, did not have a significant impact. As a result, the national innovation system has not yet reached the maturity reached by other countries, which has had a negative impact on technological innovation.

Basso et al. (2021) analyzed, also from the perspective of the Triple Helix, the technological cooperation networks of public universities in the State of São Paulo that originated patents. The results revealed that the main partnerships were academia-academia, in academia-industry collaborations it was noted that these partnerships are sporadic. Government institutions, such as support foundations, had low participation, thus showing that in the academia-government relationship, the focus is on research funding. The authors highlighted that these financial transfers are not intended for the development of innovations, but for the cost of new research that eventually ended up generating assets subject to patenting, with an uncertain future for absorption by the market.

It is important to note that the research carried out by Dalmarco et al. (2019) and Basso et al. (2021) were based exclusively on patent data. Despite the relevance of these studies, patent data does not capture all complex aspects of the TH relations. We are going beyond of already exists in the literature on TH in Brazil, including other parameters such as researchers and technology transfer, as will be detailed below. In this paper we are contributing to a literature using two quantitative multivariate analyses (PCA and HC) to understanding the relevant parameters of Brazilian universities as actors in technological innovation.

3. Methodology

This research work was conducted in the form of an applied investigation. The approach was predominantly quantitative since mathematical and statistical tools will be used. To study the relationship between parameters, exploratory multivariate techniques were used to develop a diagnosis of the data behavior under analysis and obtain information about the phenomenon under observation.

To start the data collection and processing, it was necessary to select the objects for study. The number of patents filed by the National Institute of Industrial Property (or *Instituto Nacional de Propriedade Industrial*, INPI) were the criteria used (INPI - National Institute of Industrial Property, 2018). Initially, 16 public universities were selected from among those with the highest scores in the year 2017 according to INPI - National Institute of Industrial Property (2018). However, some universities did not publish all entire data, therefore, data from eight institutions in the initial sample were analyzed in this work: Unesp; UFPR; UFRGS; UFSCar; Unicamp; UFBA; UFV and UFPE, located at different regions, according to Fig. 1.

Data were extracted from reports or statistical yearbooks indicated in references (UFBA - Federal University of Bahia, 2019; UFPE - Federal University of Pernambuco, 2019; UFSCar - Federal University of São Carlos, 2019; UFV - Federal University of Viçosa, 2019; UFPR - Federal University of Paraná, 2019; UFRGS - Federal University of Rio Grande do Sul, 2019; Unesp - São Paulo State University "Júlio de Mesquita Filho", 2019; Unicamp - State University of Campinas, 2018). When there is no data about financing available in these references, a new search was done considering other publication sources, such as sectoral management reports (CNPq - National Council for Scientific and Technological Development, 2019; CAPES - Coordination for the Improvement of Higher Education Personnel, 2019).

Twelve parameters were observed, based on literature: (1) number of research groups (Cross et al., 2017; Jovanović et al., 2019); (2) number of researchers (Cross et al., 2017; Jovanović et al., 2019); (3) number of PhD researchers (Cross et al., 2017; Jovanović et al., 2019); (4) number of teaching staff (Cross et al., 2017; Navas et al., 2020); (5) number of teaching staff with PhD (Cross et al., 2017; Navas et al., 2020); (6) number of innovation projects in collaboration (Luengo-Valderrey et al., 2020); (7) number of papers published (Cross et al., 2017); (8) patent filed (Zhao et al., 2015; Li et al., 2020); (9) patent granted (Zhao et al., 2015; Li et al., 2020); (10) number of technology transfer agreements (Zhao et al., 2015; Jovanović et al., 2019; Li et al., 2020); (11) money generated from technology transfer (Zhao et al., 2015; Jovanović et al., 2019; Li et al., 2020); (12) financing (Coccia, 2005; Navas et al., 2020).

One of the major problems to define innovation in universities and research centers is due to their specific differences and particular characteristics on size, structure, financing, among others. In fact, it was possible to observe similarities between Brazilian universities never done before, due to the large number of variables, as well as their characteristic diversities.

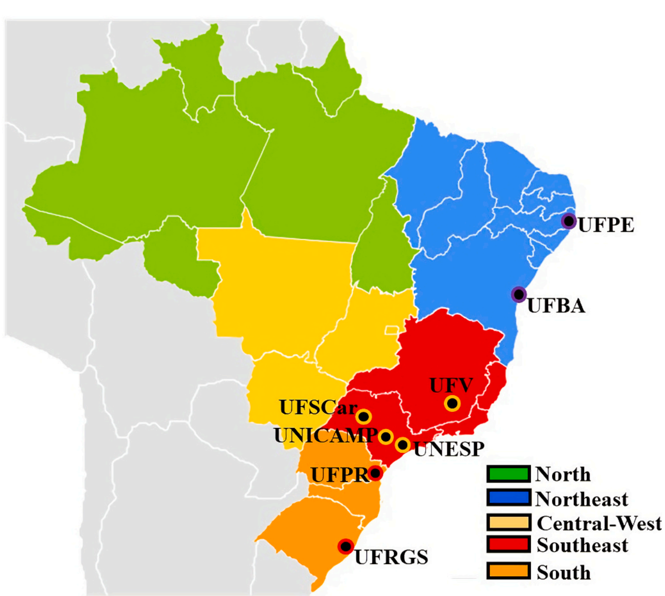

Fig. 1. The eight Brazilian universities in the study, distributed across five Brazilian regions: North, Northeast, Central West, Southeast and South.

In this sense, the parameters selected for this work reflect the performance of Brazilian universities in the innovation process, considering their performance and interaction with the other actors in the Triple Helix model. The first nine selected parameters in the present study describe characteristics of Brazilian universities such as their size and research structure.

Other variables are related to government and industries, i.e., other TH helices. For example, funding shows the main interaction between Brazilian universities and the Brazilian government. In Brazil, the federal government is the main source of funding for research and innovation (Cross et al., 2017). Number of innovation projects in collaboration, number of technology transfer agreements and money generated from technology transfer shows the interaction between universities and industries.

Data were organized using Microsoft Excel®. PyCharm™, software based on Python language, developed by the Czech company JetBrains™, was used to elaborate the dendrograms. PyCharm™ is an integrated development environment used in computer programming. The construction of dendrograms was done according to the theory described in Appendix A and the algorithm was developed based on Ward's method (Ward, 1963), in the following steps:

1. Calculate the matrix of distances (Eq. (2A)) between all normalized (Eq. (1A)) pairs of individuals;
2. Select the pair of closest individuals (clusters) with minimum distance;
3. Calculate the distance (Eq. (2A)) of this cluster to all other individuals and groups already formed;
4. Rebuild the distance matrix;
5. Reiterate the process until all individuals are clustered;
6. Construct the dendrogram.

OriginPro®, software developed by OriginLab®, was used to perform PCA analysis, according to the theory described in Appendix B, with the following steps:

1. Calculate the covariance (Eq. (4A));
2. The covariance was used to build the matrix algebra (Eq. (5A)). The solution of Eq. (5A) is related to a polynomial equation for each eigenvalue;
3. The total and cumulative percentages were calculated, and the respective eigenvalues of each axis considered (Eq. (6A));
4. Each principal component was defined as a linear combination of standardized data parameters (Eq. (7A));
5. Following, it was calculated the correlation coefficient (Eq. (8A));
6. Finally, the biplot was constructed.

In this paper we are filling some methodological gaps left by Coccia (2005), Hardeman et al. (2013), Loi and Di Guardo (2015), Zhao et al. (2015), Jovanović et al. (2019), Alnafrah and Zeno (2020) and Li et al. (2020) as mentioned above, in the introduction. Differently from what was done by Coccia (2005), Hardeman et al. (2013), Zhao et al. (2015) and Alnafrah and Zeno (2020), we proposed to use two combined multivariate techniques in an integrative way. In addition, our data deals with several parameters of the technological innovation process, such as technology transfer and interactions with other helices (government and industry) not just to one aspect of R&D activity as was done by Jovanović et al. (2019). Unlike Li et al. (2020) and Loi and Di Guardo (2015), we are using just quantitative data.

4. Results

4.1. Data presentation

This work presents data from eight different Brazilian universities, named as State or Federal (i.e., National) institutions: Unesp; UFPR; UFRGS; UFSCar; Unicamp; UFBA; UFV and UFPE. Three of them (Unesp, Unicamp and UFSCar) are located at the richest Brazilian State, São Paulo, and the first two are state universities. These three plus UFV are situated at the Brazilian Southeast region. Other two, UFPR and UFRGS, are located at South, and the remaining two, UFBA and UFPE, at the Northeast. Fig. 1 also shows that most universities are located near Brazilian cost, due to its historical expansion since Brazil's discovery in 1500. This has some similarity with countries as China (Li et al., 2020), that also presents substantial regional differences. China eastern coastal areas are far more developed than central and western provinces.

Table 1 reports the median of original data $\boldsymbol{X_i}$ between 2008 and 2015 of all twelve parameters observed for every k institution: ($\boldsymbol{X_1}$) number of research groups;($\boldsymbol{X_2}$) number of researchers; ($\boldsymbol{X_3}$) number of PhD researchers; ($\boldsymbol{X_4}$) number of teaching staff; ($\boldsymbol{X_5}$) number of teaching staff with PhD; ($\boldsymbol{X_6}$) number of innovation projects in collaboration; ($\boldsymbol{X_7}$) number of papers published; ($\boldsymbol{X_8}$) patent filed; ($\boldsymbol{X_9}$) patent granted; ($\boldsymbol{X_{10}}$) number of technology transfer agreements; ($\boldsymbol{X_{11}}$) money generated from technology transfer; ($\boldsymbol{X_{12}}$) financing. In agreement with Li et al. (2020), the utilization of quantitative survey data and the incorporation of regional heterogeneity in the present work design enable us to augment prior studies by systematically investigating and revealing some factors driving within-country differences.

It is possible to observe, from Table 1, that Unesp was the institution with the largest number of research groups in the period from 2008 to 2015, with a median of 1015. UFV was the institution with the smallest number of research groups (310 median). This groups are engaged in R&D and this is important for technological innovation in Brazil. These groups are made up of researchers, mostly PhD researchers, as shown in Table 1. It is observed that Unicamp, Unesp and UFRGS presented the highest values and UFSCar and UFV had the lowest values.

Report prepared by Clarivate Analytics in 2017 shows the predominance of research in Brazilian public universities (Cross et al., 2017). Jovanović et al. (2019) noted, in OECD countries, that R&D activities in universities are performance catalysts to promote the technological innovation process.

In Table 1 is possible to observe the number of professors and PhD professors from all institutions. They are very similar. Unesp stood out again in both observed variables (6163 and 5397 medians, respectively). UFV and UFSCar had the slowest numbers of professors and PhD professors. Considering the period between 2008 and 2015, many federal institutions received new professors, most of them with PhD, as UFPR, UFRGS, UFSCar, UFBA, UFV and UFPE. REUNI was primarily responsible for the increase in the number of teaching staff. As Unicamp and Unesp are state universities, they were unable to join this federal program. Other institutions not considered in this study were also part of this program (Cross et al., 2017; Souza et al., 2015).

Following, Table 1 shows the number of innovation projects in collaboration with other Triple Helix institutions, university-company helices (Etzkowitz and Leydesdorff, 1997; Etzkowitz, 2008; Etzkowitz and Zhou, 2017). UFRGS stood out from the other institutions, with median of 70 projects. It should be noted that between 2008 and 2015 UFPR had no projects in cooperation. It is also important to note that just one company, Petrobras, participated in 14 % of all universities-companies collaborative output during the period 2015–2017 (Web of Science, 2019), mainly on oil and gas research. For comparison, according to Luengo-Valderrey et al. (2020) the university-companies cooperation went from 13.55 % in 2010 to just over 4 % in 2015. However, business cooperation with technology centers increased sharply, followed by public research organizations and consultants.

The number of papers published is showed in Table 1, stands out

Table 1
Median of Brazilian university parameters between 2008 and 2015: ($\boldsymbol{X_1}$) number of research groups;($\boldsymbol{X_2}$) number of researchers; ($\boldsymbol{X_3}$) number of PhD researchers; ($\boldsymbol{X_4}$) number of teaching staff; ($\boldsymbol{X_5}$) number of teaching staff with PhD; ($\boldsymbol{X_6}$) number of innovation projects in collaboration; ($\boldsymbol{X_7}$) number of papers published; ($\boldsymbol{X_8}$) patent filed; ($\boldsymbol{X_9}$) patent granted; ($\boldsymbol{X_{10}}$) number of technology transfer agreements; ($\boldsymbol{X_{11}}$) money generated from technology transfer; ($\boldsymbol{X_{12}}$) financing (in R$). The subscript k indicates that distance calculations were done between respective universities.

k	Parameters	$X_{1,k}$	$X_{2,k}$	$X_{3,k}$	$X_{4,k}$	$X_{5,k}$	$X_{6,k}$	$X_{7,k}$	$X_{8,k}$	$X_{9,k}$	$X_{10,k}$	$X_{11,k}$	$X_{12,k}$
–	Triple helix axis[a]	A	A	A	A	A	I	A	A	A	I	I	G
–	Universities	Research Groups	Researchers	PhD Researchers	Teaching Staff	Teaching Staff with PhD	Innovation Projects in Collaboration	Papers Published	Patent Filed	Patent Granted	Technology Transfer Agreements	Money Generated from Technology Transfer	Financing
1	UNESP	1015	6163	5396.63	3589	3357.5	1	5128	20	5	11	204,799	338,624,442
2	UFPR	461	3087	2494.63	2217	1620	0	4016	48	0	3	0	97,242,494
3	UFRGS	736	4536	3648.38	2340	1955.5	70	2522	45	2	2	28,500,746	189,614,477
4	UFSCAR	413	2023	1802.5	978.5	920	2	887	14	7	3	516,949	125,303,157
5	UNICAMP	722	4449	3976	1744.5	1712.5	12	4362	78	13	9	476,188	304,959,118
6	UFBA	503	3255	2259.5	2245	1524.5	18	697	17	0	0	0	118,043,000
7	UFV	310	1857	1553.1	1063	805	0	1165	29	5	1	0	85,713,799
8	UFPE	573	3315	2564.88	2462.5	1971	13	1113	19	1	0	0	139,240,311

[a] Note: A describes the characteristics of academia following Triple Helix theory; I shows the interaction between academia and industry; and G shows the interaction between academia and government.

Unesp (5128 median), followed by Unicamp, UFPR and UFRGS (4362, 4016 and 2522 in median, respectively). UFSCar, UFBA, UFV and UFPE presented smaller numbers of published articles. Brazil ranks 13th in the world in terms of its output of research articles and reviews indexed in the Web of Science, just behind Italy, (8th), India (10th), Spain (11th) and South Korea (12th), and ahead of Russia (15th) and South Africa (21th). The first place in this rank belongs to USA, followed by China. In 2018 Brazilian researchers published more than fifty thousand papers, one third co-authored with researchers from other countries (Web of Science, 2019). Just 15 universities, according to Cross et al. (2017) and Web of Science (2019), produce over 60 % of the total paper production in Brazil.

Table 1 presents the number of patents filed and granted by institutions. Statistics released annually by INPI show that, among residents in Brazil, many patents filed and granted are from universities (INPI - National Institute of Industrial Property, 2018).

Following this parameter, Unicamp presented the highest data in the period analyzed, with median of 78. UFSCar was able to maintain a demand for patents in the period analyzed. Regarding the patents granted, Table 1 shows that Unicamp generally presented the highest number of patents granted, with a median of 13 and an outlier of 35 granted patents in 2015. Unesp granted 23 patents in 2013 and UFPR, the UFRGS and UFPE, in 2015, with eight, seven and three, respectively. In the period observed, UFBA had only one patent granted in 2012. All these numbers are extremally low when compared with papers published, also presented in Table 1.

Just for comparison, the average patent application in developing Chinese provinces as Fujian and Hubei in 2017 was of 52258, with an average of 33135 patent grants (Li et al., 2020). However, Zhao et al. (2015) informed the generation of an average of 580.60 $\pm$ 68.52 exploitable Chinese patents from 57 universities of 30 provinces in 2012 (with a minimum of 2 and a maximum of 2222), indicating that some regions are unable to generate relevant innovation outputs. The comparison with Brazil would be done considering regions as South-Southeast, according to Fig. 1, that are far below Chinese provinces (i. e. under-developed). From this view, we agree with Boardman (2009) that is necessary a major collaboration among government, universities, and the industry. This procedure would be highly beneficial for university technology transfer and the development of university entrepreneurship in Brazil.

Table 1 shows the number of technology transfer agreements signed with companies. This marks the iteration between academia and industry when considering the Triple Helix model (Etzkowitz, 2008; Etzkowitz and Zhou, 2017). Following these data is possible to observe that, in Brazil, the technology transfer is still deficient. Unesp and Unicamp presented medians of 11 and 9, respectively. Also, it was observed that UFPE and UFBA had no contracts signed between 2008 and 2015. It is important to cite UFRGS with 121 contracts in 2015.

It is also important to note that the present analysis is a quantitative study, and quite a few were done in this way, as pointed by Zhao et al. (2015). For these authors, many provinces in China are not involved in collaborations, thus losing the opportunity to benefit from regional technological innovation projects. It is also possible to observe in the results achieved by the authors that universities are the institutions that most establish collaborative innovation projects.

The low technology transfer rate reflects directly in the amount collected. Jovanović et al. (2019) mentioned that the technological balance of payments receipts indicator reflects innovative activities. Table 1 shows that UFBA and UFPE obtained no values from technology transfer. However, Unesp collected more than R$[1] 1.2 million in 2011. UFRGS presented the best performance, significantly increased in the last three years of the period considered, when obtained R$ 47 million in 2012, with a median of R$ 28 million. The next universities were, in order of their medians, UFSCar (R$ 0.517 million), Unicamp (R$ 0.476 million) and UFRGS (R$ 0.205 million). Li et al. (2020) reported that there was an average of contractual values of RMB[2] 6.203 billion in advanced Chinese provinces in 2017, as Beijing, Guangdong and Shanghai. By comparison with Brazilian regions, such expended values from China are very high.

Regarding public funding received by each institution, shown in Table 1, it is possible to observe a large variation in the volume of funds received for most institutions in Brazilian reais (R$). Financing marks the main iteration of universities-government when considering the Triple Helix model (Etzkowitz, 2008; Etzkowitz and Zhou, 2017). UFPR and UFV presented lower values received in the period analyzed, below US$ 100 million (median). Unesp, Unicamp and UFRGS presented the highest values. Other institutions showed similarities in terms of the average values received.

In Colombia as well as in Brazil, according to Navas et al. (2020), investment in R&D comes from public resources. Comparing the results of Navas et al. (2020) with those of Brazilian universities, presented here, can be observed that the resources coming from private companies, measured through the amount collected with technology transfer, is still much lower than the public resources received by universities, measured by financing.

Also in Italy, according to Coccia (2005) many institutes carrying out research activities are public and financed by the government, who may wish to maximize the added value for society. The public financing to research activities is justified by the fact that, besides being a product that enhances society, scientific production is also an investment that generates effects in terms of scientific-technological progress and therefore also in terms of a bigger wealth produced by the nation in the medium–long term.

4.2. HC and PCA analysis results

Following procedure shown in appendices, Table 2 presents the corresponding correlation matrix of standardized data. According to Hair et al. (2019), only factors having eigenvalues >1 are considered significant, as presented in Table 3. Other criterion is selecting enough factors to achieve a prespecified communality for each of the variables, near 93 % of cumulative percentage, also shown in Table 3. A third criterion (the scree or elbow test) is derived by plotting eigenvalues against the number of factors in their order of extraction. The shape of the resulting curve is used to evaluate the cut-off point, that follows a tangent method, as shown in Fig. 2. All such criteria were satisfied by the first three principal components.

Table 3 shows the total and cumulative percentages and respective eigenvalues of each axis considered (PhD researchers, innovation projects and patent filed, in order), whereas Table 4 presents the eigenvector values of these respective parameters for the corresponding three PC axis. In fact, from Table 3, most data were characterized by the first three axis (related to PhD researchers, representing 57.53 % of variability; innovation projects, representing 20.15 % of variability; and patent filed, representing 15.32 % of variability, respectively.

Fig. 3 summarizes the clustering formation based on Euclidean distance (in agreement with Eq. (2A)), i.e. similarity, observed from standardized data. Dendrogramming is able to capture multi-dimensionality and complex relationships from many variables, as that presented in Table 1 (Hair et al., 2019). The dendrogram presented in Fig. 3 illustrates how the clustering takes place and is in agreement with PCA analysis shown in Fig. 4, as described below. Four clusters were established from this dendrogram, obtained by taking into account the largest distance leap, according to Fávero and Belfiore (2019): (Unesp), (Unicamp), (UFRGS) and (UFSCar, UFV, UFPR, UFBA, UFPE). From these data it was possible to verify that UFBA and UFPE were the most similar

[1] R$, or real, is the official currency of Brazil.

[2] RMB, or Renminbi, is the official currency of China.

Table 2
Correlation matrix of standardized data (Z_i) according to Eq. (8A). Grey cells represent high correlated parameters, and black cells non-correlated parameters.

	$Z_{1,k}$	$Z_{2,k}$	$Z_{3,k}$	$Z_{4,k}$	$Z_{5,k}$	$Z_{6,k}$	$Z_{7,k}$	$Z_{8,k}$	$Z_{9,k}$	$Z_{10,k}$	$Z_{11,k}$	$Z_{12,k}$
	Research Groups	**Researchers**	**PhD Researchers**	**Teaching Staff**	**Teaching Staff with PhD**	**Innovation Projects in Collaboration**	**Papers Published**	**Patent Filed**	**Patent Granted**	**Technology Transfer Agreements**	**Money Generated from Technology Transfer**	**Financing**
Research Groups	1	0.98634	0.98481	0.81786	0.93218	0.2663	0.71709	0.17684	0.2173	0.75417	0.26465	0.91255
Researchers	0.98634	1	0.98407	0.85279	0.93524	0.28467	0.75809	0.24554	0.15445	0.73102	0.27295	0.88397
PhD Researchers	0.98481	0.98407	1	0.78065	0.90857	0.19919	0.81874	0.29553	0.29437	0.83017	0.2207	0.93738
Teaching Staff	0.81786	0.85279	0.78065	1	0.95543	0.15561	0.56179	−0.10061	−0.32225	0.4206	0.1192	0.56577
Teaching Staff-PhD	0.93218	0.93524	0.90857	0.95543	1	0.11241	0.67945	−0.02001	−0.05938	0.63323	0.11315	0.75858
Innovation Projects in Collaboration	0.2663	0.28467	0.19919	0.15561	0.11241	1	−0.08826	0.20502	−0.20929	−0.22975	0.95577	0.07121
Papers Published	0.71709	0.75809	0.81874	0.56179	0.67945	−0.08826	1	0.57149	0.34743	0.86173	0.01296	0.75445
Patent Filed	0.17684	0.24554	0.29553	−0.10061	−0.02001	0.20502	0.57149	1	0.5184	0.38646	0.21034	0.36107
Patent Granted	0.2173	0.15445	0.29437	−0.32225	−0.05938	−0.20929	0.34743	0.5184	1	0.65223	−0.17329	0.5858
Technology Transfer Agreements	0.75417	0.73102	0.83017	0.4206	0.63323	−0.22975	0.86173	0.38646	0.65223	1	−0.14668	0.90774
Money Generated from Technology Transfer	0.26465	0.27295	0.2207	0.1192	0.11315	0.95577	0.01296	0.21034	−0.17329	−0.14668	1	0.07271
Financing	0.91255	0.88397	0.93738	0.56577	0.75858	0.07121	0.75445	0.36107	0.5858	0.90774	0.07271	1

Table 3
Total and cumulative percentages and respective eigenvalues of each axis considered. The eigenvalue sum is twelve, as expected by theory (n = 12).

PC axis (n)	Eigenvalue	Total percent (%)	Cumulative percent (%)
1	6.90375	57.53 %	57.53 %
2	2.4180	20.15 %	77.68 %
3	1.83847	15.32 %	93.00 %
4	0.63667	5.31 %	98.31 %
5	0.18014	1.50 %	99.81 %
6	0.01969	0.16 %	99.97 %
7	0.00327	0.03 %	100.00 %
8	0	0.00 %	100.00 %
9	0	0.00 %	100.00 %
10	0	0.00 %	100.00 %
11	0	0.00 %	100.00 %
12	0	−0.00 %	100.00 %

due to the smallest Euclidean distance considering all 12 variables, promoting thus the first clustering stage. There is a correspondence about hierarchical and PCA analyzes, as Unesp and Unicamp are at fourth (lower right) quadrant and UFRGS is isolated at first (upper right) quadrant. In opposition, UFBA and UFPE are close at upper left quadrant, as well as UFV and UFSCar, at lower left quadrant, and all these grouped near UFPR.

Fig. 4 shows the mapping distribution of principal component 1 (**PC$_1$**, related mainly to PhD researchers) and principal component 2 (**PC$_2$**, related mainly to innovation projects). The third principal component (**PC$_3$**), is mainly related to patent filed, as described below.

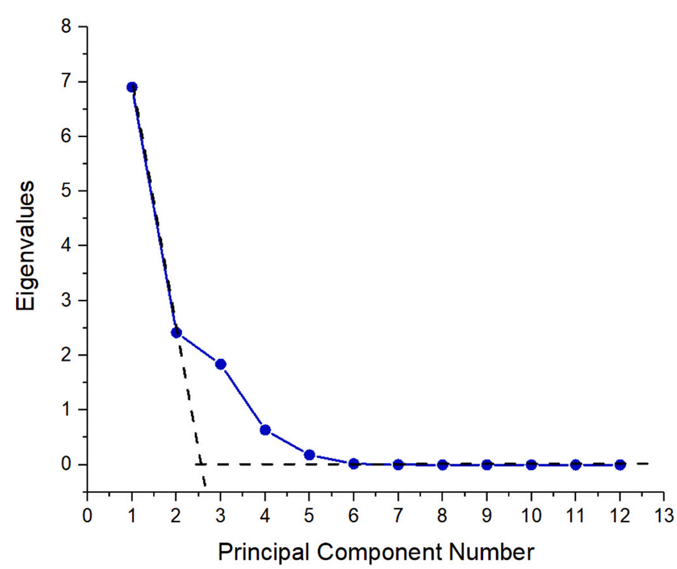

Fig. 2. Eigenvalue plot for scree (or elbow) test criterion, that considers all twelve parameters of this work. Starting with the first factor, the plot slopes steeply downward initially and then slowly become an approximately horizontal line. An inflection point termed as the "elbow" arises from two tangent lines, indicating near three relevant eigenvalues, near three in the present case.

Table 4
Eigenvector values of PhD researchers, innovation projects and patent filed of the corresponding three PC axis and considering the correlation matrix.

n	Parameter	PC_1	PC_2	PC_3
		$\alpha_{j,1}$	$\alpha_{j,2}$	$\alpha_{j,3}$
1	Research Groups	0.37021	0.1045	−0.04013
2	Researchers	0.37159	0.12233	−0.03261
3	PhD Researchers	0.37954	0.03288	−0.00199
4	Teaching Staff	0.29297	0.25757	−0.33786
5	Teaching Staff with PhD	0.34206	0.15201	−0.26498
6	Innovation Projects	0.05975	0.5037	0.43241
7	Papers published	0.32448	−0.17026	0.05193
8	Patents Filed	0.12901	−0.19666	0.53918
9	Patents Granted	0.12049	−0.47294	0.35039
10	Innovation Contracts	0.32436	−0.30968	0.01906
11	Money Tech Transfer	0.06823	0.47362	0.45189
12	Financing	0.3579	−0.13537	0.07228

From this map is possible to note that Unesp and Unicamp are at fourth (lower right) quadrant, UFRGS is isolated at first (upper right) quadrant. UFBA and UFPE are close at upper left quadrant, as well as UFV and UFSCar, at lower left quadrant. UFPR is closer to UFBA-UFPE group. So, it is possible to consider the following groups: (Unesp), (Unicamp), (UFRGS) and (UFSCar, UFV, UFPR, UFBA, UFPE).

From this figure the TH helices are visible and also grouped, presenting industry, academia and government interrelations. Previous studies cited in introduction also promoted similar approaches, but in this work, it is important to highlight that two different statistical techniques presented close results, in an integrative way as shown below. For instance, UFRGS is close to TH helices related to innovation projects and money tech transfer parameters. Unicamp is close to patent granted, patents filed and innovation contracts parameters. Unesp is close to PhD researchers, papers published and financing parameters.

As briefly explained in appendices, positive eigenvalues occurred for $\mathbf{PC_1}$ in Table 4 due to the use of matrix algebra applied to standardized data, in agreement with Eq. (7A). The first component $\mathbf{PC_1}$ can be expressed in terms of standardized Z_i variables from this table, as:

$$\mathbf{PC_1} = 0.37021\mathbf{Z_1} + 0.37159\mathbf{Z_2} + 0.37954\mathbf{Z_3} + 0.29297\mathbf{Z_4} + 0.34206\mathbf{Z_5} + 0.05975\mathbf{Z_6} + 0.32448\mathbf{Z_7} + 0.12901\mathbf{Z_8} + 0.12049\mathbf{Z_9} + 0.32436\mathbf{Z_{10}} + 0.06823\mathbf{Z_{11}} + 0.3579\mathbf{Z_{12}} \tag{1}$$

The coefficients of first principal component in Eq. (1) are related in order to PhD researchers, innovation projects and patent filed and to their respective eigenvectors. That is to say, $\mathbf{PC_1}$ will be high if all Z_i are high, and represents 57.53 % of the variation in the data. Low Z_6 and Z_{11} coefficients means that the values such variables do not affect $\mathbf{PC_1}$. Therefore, the first principal component is by far the most important for representing the variation in the twelve parameters of the eight Brazilian universities.

The second principal component can be interpreted in a similar way:

$$\mathbf{PC_2} = 0.1045\mathbf{Z_1} + 0.12233\mathbf{Z_2} + 0.03288\mathbf{Z_3} + 0.25757\mathbf{Z_4} + 0.15201\mathbf{Z_5} + 0.5037\mathbf{Z_6} - 0.17026\mathbf{Z_7} - 0.19666\mathbf{Z_8} - 0.47294\mathbf{Z_9} - 0.30968\mathbf{Z_{10}} + 0.47362\mathbf{Z_{11}} - 0.13537\mathbf{Z_{12}} \tag{2}$$

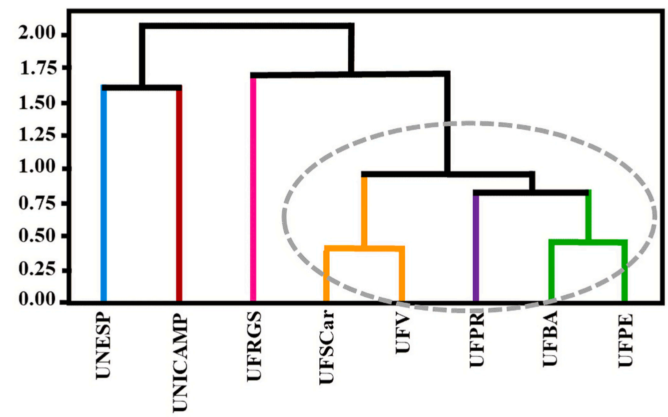

Fig. 3. Dendrogram computational results for eight Brazilian universities considering standardized Z_i of twelve median parameters, following Eq. (1A). Four clusters were established from this dendrogram, obtained by taking into account the largest distance leap: (Unesp), (Unicamp), (UFRGS) and (UFSCar, UFV, UFPR, UFBA, UFPE), as presented in Table 1, in agreement with PCA analysis.

The second principal component has half positive and half negative coefficients. In particular, the most representative parameter was about innovation projects. $\mathbf{PC_2}$ will be high if all Z_i with positive coefficients are high and all Z_i with negative coefficients are low, excepting Z_3, that has a small coefficient compared to others. With Eqs. (1) and (2) it is possible to plot Fig. 4 considering the standardized Z_i data. The first two components account for 77.68 % of the variance allowing most of the information to be visualized in two dimensions. All first three PC account for 93 % of the variance.

There are *highly* correlations (> 0.8) between variables that explains the reduction to just three principal components from twelve parameters, as for example the correlations between Z_1 with Z_2, Z_3, Z_4, Z_5 and Z_{12}. All these parameters are related to research groups, researchers, teaching staffs and financing. This means that it was noted strong correlations between academia and government helices (identified as Z_{12}) as described in Table 2. There is also an obvious relation between PhD researchers (Z_3) with papers published (Z_7), technology transfer agreements (Z_{10}) and financing (Z_{12}). Published papers from such universities (Z_7) were also highly correlated with technology transfer agreements (Z_{10}), with a coefficient correlation $\mathbf{Corr}(Z_7, Z_{10}) = 0.86173$ considering all the data available in Table 2, and in agreement with Eq. (8A). The money generated from technology transfer (Z_{11}) was correlated with innovation projects in collaboration (Z_6) with a coefficient correlation $\mathbf{Corr}(Z_6, Z_{11}) = 0.95577$. Finally, technology transfer agreements (Z_{10}) were correlated to financing (Z_{12}) with a coefficient correlation $\mathbf{Corr}(Z_{10}, Z_{12}) = 0.90774$.

Almost no correlations (near zero) were observed between papers

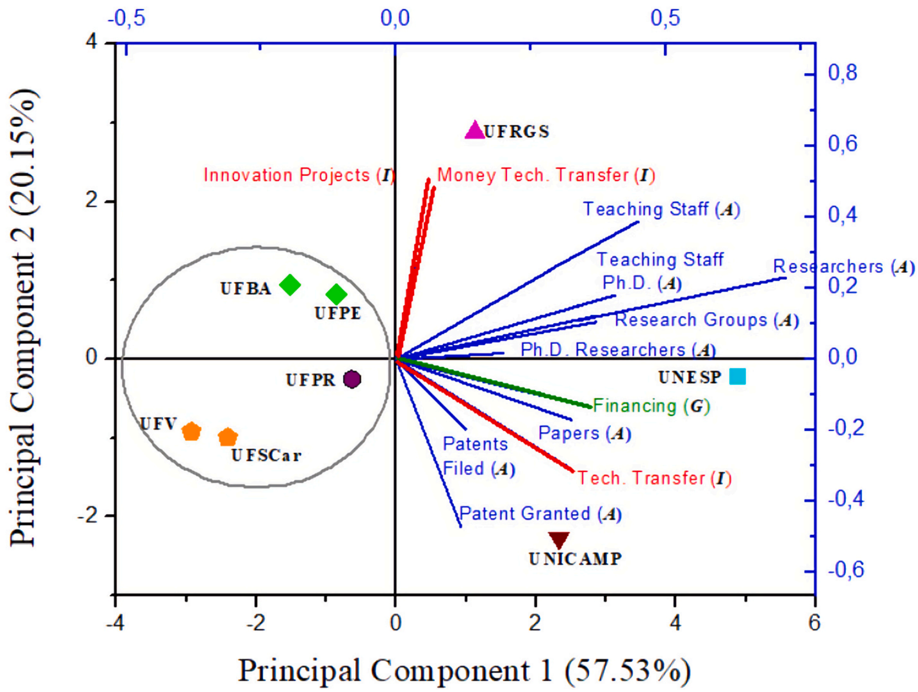

Fig. 4. Biplot of principal component 1 (related to 57.53 % of data) versus principal component 2 (20.15 % of data) considering all twelve parameters and the correlation matrix. The ellipsis represents the biggest group (UFSCar, UFV, UFPR, UFBA, UFPE), and the positions of the Brazilian in this bidimensional map are in agreement with hierarchical cluster distances. From this graph the helices from TH are visible, and also grouped, presenting industry, academia and government interrelations.
*Note: ***A*** describes the characteristics of academia following Triple Helix theory; ***I*** refers to the interaction between academia and industry; and ***G*** shows the interaction between academia and government.

published (Z_7) and innovation projects in collaboration (Z_6), as well as money generated from tech transfer (Z_{11}), with a coefficient correlation $\mathbf{Corr}(Z_7,Z_{11}) = 0.01296$. No correlations between patens filed (Z_8) and granted (Z_9) with teaching staff with PhD (Z_5) were observed, due to the coefficient correlations $\mathbf{Corr}(Z_5,Z_9) = -0.02001$ and $\mathbf{Corr}(Z_5,Z_9) = -0.05938$. Finally, no correlation was observed between financing (Z_{12}) and innovation projects in collaboration (Z_6), neither with money generated from tech transfer (Z_{11}).

PCA is usually based on the calculation of eigenvalues of the correlation matrix. However, it is also possible to calculate PCA considering the eigenvalues of the covariance matrix. In this particular case, we found just one difference on principal components: financing as one of the main factors for $\mathbf{PC_1}$, followed by innovation projects as $\mathbf{PC_2}$ and patents filed as $\mathbf{PC_3}$. This result is reasonable because, from covariance calculations, the parameter that presented the highest variability was financing. The mapping distribution of principal components 1 and 2 considering the covariance mode (not shown in this work) is similar to Fig. 4 because both $\mathbf{PC_1}$ results are highly correlated.

Thus, no significant change was observed from using the different procedures to obtain the first PCs, if one compares correlation or covariance modes, as described by some references, as Flury (1997). We obtained results of 57.53 % and 57.94 % for $\mathbf{PC_1}$, 20.15 % and 19.91 % for $\mathbf{PC_2}$ and 15.32 % and 14.60 % for $\mathbf{PC_3}$, respectively.

The application of the principal component analysis can capture the trends of different innovation parameters related to time series from different universities with significant results. This method can be associated with numerical taxonomy, another multivariate technique that also reduces multidimensional datasets.

All clusters presented in dendrogram from Fig. 3: (Unesp), (Unicamp), (UFRGS) and (UFSCar, UFV, UFPR, UFBA, UFPE) have corresponding groups on principal component mapping from Fig. 4. According to Hair et al. (2019), distance measures from HC focus on magnitude values and portray as similar the unities that are closer, even including their different patterns across the variables. In contrast, correlation measures from PCA focus on the patterns across the variables and do not consider the magnitude of the differences between Brazilian universities. One relevant result is that, from PCA map, the correlations between UFBA and UFPE, UFV and UFSCar presented a similarity with calculated distances considering HC, showing agreement between them in an integrative way. Also, Unicamp, Unesp and UFRGS are sparse, but UFPR are close to (UFV, UFSCar) and (UFBA, UFPE) groups.

5. Discussion

5.1. The similarity between Brazilian universities

In detail, the similarity between Unesp and Unicamp is evident in Figs. 3 and 4 and is in agreement with the data presented in Table 1. It should also be noted that, since 2014, universities in São Paulo have participated in the São Paulo System of Innovation Environments, formed by the São Paulo Technology Parks System, the São Paulo Technology-Based Business Incubators network, the Rede of Technological Innovation Centers of São Paulo and the Network of Technological Innovation Centers of São Paulo (Castro and Souza, 2012). This ecosystem, according to Balle et al. (2019), González-Masip et al. (2019) and Carayannis et al. (2014), benefits both the companies and universities in the technological innovation process.

The state of São Paulo can be considered a leader in terms of creating a formal technological innovation environment. It is also relevant to mention that, in Brazil, the SNCTI was created in 2015, one year after the creation of the São Paulo system. São Paulo also has a state law and is the only Brazilian state with specific legislation on innovation.

Following these aspects, such as superior financing, the existence of innovation environments and specific legislation, it is possible to explain the leadership of state universities in São Paulo in the technological innovation process. Add to that the state's predominantly industrial economy that contributes to better results in innovation. The innovation environment in state of Rio Grande do Sul does not follow this structure, but UFRGS stands out for trying to offer technological innovation integration.

Considering UFRGS, we highlight the creation of the Innovative Entrepreneurship Center in 2012, composed of teachers and technicians, which aims to awaken the culture of entrepreneurship and innovation among its teachers and students. The UFRGS Science and Technology Park (ZENIT Park) also started operating in 2012, this boosted the transfer of technology and increased the amount collected.

UFBA and UFPE also showed similarity in the data presented in Table 1. Some factors explain the similarity between the institutions in

Figs. 3 and 4. Both were officially created in the same year (1946), are located in Brazilian northeast, in neighboring states. The economies of Bahia and Pernambuco are composed of agriculture, industry, mining, tourism and services. It is worth mentioning that none of the states has specific legislation for innovation, so the innovation environment in both states is still poorly integrated. In addition, the results observed in the parameters obtained are also similar for the two institutions.

UFPR is located in the southern region of Brazil, so a performance similar to that of UFRGS was expected. The data presented in Table 1 shows many differences for UFRGS and greater similarity with UFBA and UFPE (Figs. 3 and 4). This is explained because the state of Paraná also lacks legislation and an innovation environment. Although the economy of Paraná is quite diversified, one can find a well-developed industrial park, as well as a service sector linked to urban centers, but the major driver of the economy of Paraná is also agribusiness.

The similarity between UFV and UFSCar is clear from Figs. 3 and 4, in agreement with data presented in Table 1. These two universities can be considered similar and presented the worst results among the analyzed institutions. UFV is located in Minas Gerais, this state does not have specific legislation for innovation. UFSCar, located in the state of São Paulo, follows the São Paulo Innovation Environment System, and similarity with Unicamp and Unesp was expected.

According to Alnafrah and Zeno (2020) the structural differences among the innovation systems characterize each innovation approach also have an impact and need to be considered. In this sense, the analysis carried out by this work revealed similarity in the technological innovation process between Brazilian universities located in the same geographical region. In China, the results verified by Zhao et al. (2015) and Li et al. (2020) also demonstrate that variations in the regional context affect the functioning and processes of the innovation system, as well as patterns of interaction and processes that reinforce each other among innovation agents.

The results performed by Loi and Di Guardo (2015) shows statistically significant association between geographical position of Italian universities. In the north of Italy was verified the profile of *exploitation*, focused on patent disclosure. In the central regions of Italy was verified the profile of *openness*, readiness to participate in external change and to satisfy external needs. In the South of Italy was verified the profile of *need for coherence*, focused on balancing public functions and technological innovation activities. The *old school* profile, focused on entrepreneurial activities as a source of funding, is uniformly distributed throughout the Italian territory.

5.2. *The triple helix in Brazilian universities*

As presented below, our results show agreement with theoretical Triple Helix framework (Etzkowitz, 2008; Etzkowitz and Leydesdorff, 1997). Also, as presented in Results section, Fig. 4 provides a graphical representation of the helices themselves. In the same way as achieved by Li et al. (2020) the results indicate that trilateral collaboration between universities, companies and the government is essential to create a favorable environment and stimulate the technological innovation. They also highlighted the importance of networking in promoting exchange and long-term cooperation and joint development between universities, industries and government agencies in the innovation system.

According to multidimensional analysis, many variables can be merged as simple principal components and viewed in light of TH model as their helices. It is thus possible to map TH by means of variable reduction, i.e., reducing the number or dimensions of parameters in our dataset. Despite of that, the interactions of Brazilian universities with other institutions of the Triple Helix is still far from the importance reported in the literature by authors such as: Del Giudice (2008); Petruzzelli et al. (2010); Philpott et al. (2011); Perkmann et al. (2013); Carayannis et al. (2014); Romano et al. (2014); Sydow et al. (2016); Sá et al. (2018); Dooley and Gubbins (2019); Johnston (2020); Wang and Lu (2021). These studies argue that the university can explore opportunities for innovation by establishing stronger ties with companies, government and other actors in the territory, taking technology from its limits, valuing and exploring scientific knowledge. Fig. 4 illustrates how the helices from TH are visible considering eight Brazilian universities, presenting some industry, academia and government interrelations between, as well as among university groups.

In Spain, Luengo-Valderrey et al. (2020) carried out an empirical analysis of a stratified sample of >5000 medium and high technological companies. The authors analyzed 21 indicators through using the structural equation method and covariance analysis to establish causal relationships between the innovative performance of companies and the information obtained from the Triple Helix.

In the Spanish context, Luengo-Valderrey et al. (2020) observed that the university, although considered a driving force and the most important source of information for technological innovation, loses ground to technology centers, public research organizations and consultancies. The authors argue that this contradiction may result from the difference in the response time of universities for the development of technology in relation to the time demanded by companies. These results of Luengo-Valderrey et al. (2020) can be also partially applied in Brazil.

From literature it was possible to note previous studies that considered both PCA and HC, but in this work is presented a new approach that surpasses some limitations when taking into account such separated statistical tools that would give wider practice and policy implications. Data taken from Brazilian institutions can be used as examples for further applications in other countries.

In the same way as Basso et al. (2021), our results show that in the relationship between academia and government, the main role of government is the funding of scientific and technological research. It is also noticed that academia and industry continue to develop their activities separately, working together only in a limited and sporadic way. It is worth noting that in Brazil the concepts of the Triple Helix model were used to create the national innovation system. However, the actual country's economic system, low investment in universities and the small number of national companies are factors that hinder scientific and technological progress and, consequently, technological innovation.

Historically, Brazilian industry has shown a preference for improving mature technologies and has rarely acquired research results and technologies from its universities and other public research institutions. In academia-industry collaboration, companies lacked the internal capacity to absorb and benefit from the knowledge generated by universities and, to avoid investing in these internal capacities, there was a clear tendency to import technology. Analyzing the technological cooperation networks of public universities in São Paulo for patent filings, Basso et al. (2021) and Dalmarco et al. (2019) corroborated this perception.

6. Conclusions

The comparative analysis presented and based on HC and PCA correspond to a new proposal to understand and visualize such complex data, associating many variables, parameters and/or dimensions. To deal with many parameters, variables and conditions, such exploratory techniques are some of the best to face multivariate problems, that identify similarities (e.g., from HC) and simplifies key variables (e.g., from PCA), helping to advance the complex dynamics of TH actors. Particularly, the helices from TH theory are visible and also grouped considering such statistical tools that taking into account industry, academia and government interrelations from Brazilian universities.

So, a complex system like the TH model could be illustrated by a figure, presenting industry, academia and government interrelations. Also determining the key institutions and variables in a multidimensional data set, being thus possible to map TH. In brief, our results confirm that PCA can find a reduced set of variables that are useful for understand the most relevant parameters from Brazilian universities and their innovation. These variables were PhD researchers, innovation projects and patent filed, in order, covering 93 % of the total variability, thus capturing most of the information. The main advantage of the use of PCA in this work was to enable simple graphic criteria to classify different Brazilian universities, that are in agreement with hierarchical analysis. All clusters presented in dendrogram have corresponding groups on principal component mapping.

This work presents a new analysis framework to explain the way in which new technological innovations emerge and diffuse in TH model. There were previous studies considering PCA and HC, but in this work is presented a new approach that surpasses some limitations when taking into account such statistical tools separated, mapping TH in an integrative way.

One limitation observed from previous studies was that they not considered exploratory analyses as complementary tools. Another one is to interpret graphical representations as helices themselves. Graphs can help to summarize multivariate analyses, providing a low-dimensional representation of the data. Following these procedures, at least for some cases, many variables can be merged as simple principal components and viewed in light of TH model as their helices (or part of), that can be applied to other contexts and even other countries. In such a way exploratory analyses as PCA and HC can be used as graphical criteria to match helices in the context of TH model.

This paper contributes to the regional innovation literature by improving a multi-dimensional analysis of innovative capabilities in Brazilian universities, mainly in terms of staff, collaborations and infrastructure. Our results should also be examined in light of the limitations of this work. For example, our work is unable to reveal other Brazilian universities, their amount of innovation investments and supportiveness of local government policies. Future research could consider larger databases to investigate other institutions and compare with other countries. However, the present mathematical approach also shed light on traditional and classical analysis on innovation studies, offering new possibilities, as minimizing some subjectivities in innovation classification, particularly done in universities and research centers.

CRediT authorship contribution statement

Eron Passos Andrade: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization, Supervision, Project Administration, Funding Acquisition. **Jadiel dos Santos Pereira:** Software, Data Curation. **Angela Machado Rocha:** Resources, Writing - Review & Editing.; **Marcio Luis Ferreira Nascimento:** Methodology, Validation, Formal Analysis, Resources, Writing - Review & Editing, Visualization, Funding Acquisition.

Funding

This work was supported in part by the Coordination for the Improvement of Higher Education Personnel (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES*) [Finance Code 001]; by the National Council for Scientific and Technological Development (*Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq*) [304705/2015-2, 404004/2016-4 and 305331/2018-3]; and by the Bahia Research Foundation (*Fundação de Amparo à Pesquisa do Estado da Bahia* - FAPESB) [BOL0151/2017].

Acknowledgment

We sincerely thank the anonymous reviewers whose comments and suggestions helped to improve and clarify this work.

Appendix A. Brief review of hierarchical analysis

Hierarchical Analysis, also known as dendrogramming, is a quantitative method for classification, can be viewed as a kind of numerical taxonomy. Dendrograms are pictographic data representations with many dimensions or variables, which are contrasted as a function of distances between the values of such variables. As a computational technique, takes place in the form of a phenotype matrix of dendrodistant elements $d_{i,j}$ (Sokal and Michener, 1958; Sneath and Sokal, 1962; Sokal and Sneath, 1963) considering a matrix $\boldsymbol{X}$ of raw data with t rows (or labels or objects) and n columns.

First it is important to consider the data transformed by a standardizing matrix $\boldsymbol{Z}$ from an original data matrix $\boldsymbol{X}$ with t rows (observations or labels) and n columns. From a practical point of view, raw data from Table 1 could be represented by a matrix $\boldsymbol{X}$ composed of $n = 12$ variables ($\boldsymbol{X}_1$ as the number of research groups; $\boldsymbol{X}_2$ as the number of researchers; $\boldsymbol{X}_3$ as the number of PhD researchers…) and $t = 8$ universities (or labels). The main goal is thus to determine which one of these twelve variables is more relevant, if any, assuming data from eight universities.

Every $\boldsymbol{Z_i}$ (for a fixed k or university) can be described by:

$$\boldsymbol{Z}_i = \frac{\boldsymbol{X}_i - \boldsymbol{X}_i^{min}}{\boldsymbol{X}_i^{max} - \boldsymbol{X}_i^{min}} \tag{1A}$$

where $\boldsymbol{X_i^{min}}$ is minimum and $\boldsymbol{X_i^{max}}$ is maximum values for each $\boldsymbol{Z_i}$.

It takes into account the Euclidean metrics, that associates two standardized elements, $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$ from a dataset, characterized by t rows and n columns (variables or parameters, according to Hair et al. (2019)):

$$d_{i,j} = \sqrt{\sum_{k=1}^{t} \left(Z_{i,k} - Z_{j,k}\right)^2} \tag{2A}$$

where k is the label for each of t universities from n characteristic values and $d_{i,j}$ is the Euclidean distance between normalized elements $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$ defined by Eq. (1A).

In other words, $Z_{i,k}$ is the value of variable Z_k for object i, and $Z_{j,k}$ is the value of the same variable for object j. From Eq. (2A) it is possible to observe that the distance is zero considering the same object, and the distances between $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$ are equal to $\boldsymbol{Z_j}$ and $\boldsymbol{Z_i}$.

The procedure provides levels of (dis)similarities due to respective distances between sub-sets through dotted horizontal lines in dendrograms. This was used to analyze innovation at some Brazilian universities, in a similar way as that proposed by Sokal and collaborators (Sokal and Michener, 1958; Sneath and Sokal, 1962; Sokal and Sneath, 1963). This technique deals with classification and ordering of universities with similar characteristics.

Appendix B. Brief Review of PCA Theory

PCA extracts relevant information when reducing the dimensionality of a data set, and thus determining the most relevant variables, termed Principal Components (PCs), which account for the majority of the variability in the data (Jolliffe, 2002).

Given t observations on n variables, the goal of PCA is to reduce the dimensionality of a data matrix $\boldsymbol{X}$ by finding p new variables with $p < n$. Thus, the first PC is the direction throughout the data that explains the highest variability. The second and subsequent PCs must be orthogonal to the previous PC and describe the maximum amount of the remaining variability (Hair et al., 2019; Jolliffe, 2002).

PCA is a singular case of transforming the original data into a new coordinate system with fewer variables and in order of their importance in terms of the variation in the data. If the original data involves n diverse variables that corresponds to diverse parameters and each column corresponds to one university, then each datum may be considered as a point in a multidimensional vector space. The PCA fundamentals are credited to Pearson (1901) and Hotelling (1933).

Briefly, a variance $\mathbf{Var}(\boldsymbol{X_i},\boldsymbol{X_i})$ is a "measure of data spread" considering a unique variable $\mathbf{X}_i$, $1 < i < n$:

$$\mathbf{Var}(X_i, X_i) = \sum_{i=1}^{t} \frac{(X_i - \overline{X}_i)^2}{t-1}. \tag{3A}$$

where $\overline{X}_i$ corresponds to the average (or mean) of the variable considered.

Covariance is similar to variance, but it considers data from different variables, i.e., $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$, with $\boldsymbol{i \neq j}$:

$$\mathbf{Cov}(X_i, X_j) = \sum_{i=1}^{t} \frac{(X_i - \overline{X}_i)(X_j - \overline{X}_j)}{t-1} \text{ for every } \boldsymbol{i \neq j}. \tag{4A}$$

The covariance indicates the direction of the linear relationship between variables. The procedure of PCA can be expressed concisely in terms of matrix algebra (Jolliffe, 2002) given below.

The eigenvalue equation can be described as:

$$\begin{matrix} \mathbf{Cov}(Z_1, Z_1) - \lambda & |\mathbf{Cov}(Z_1, Z_2)| & \dots & \mathbf{Cov}(Z_1, Z_{12}) \\ \mathbf{Cov}(Z_2, Z_1) & \mathbf{Cov}(Z_2, Z_2) - \lambda & \dots & \mathbf{Cov}(Z_2, Z_{12}) \\ \dots & \dots & \dots & \dots \\ \mathbf{Cov}(Z_{12}, Z_1) & \mathbf{Cov}(Z_{12}, Z_2) & \dots & \mathbf{Cov}(Z_{12}, Z_{12}) - \lambda \end{matrix} = 0 \tag{5A}$$

The solution is related to a polynomial equation on λ, with twelve roots: λ_1, λ_2 … and λ_{12} (for $n = 12$). The relevance percentage of every $\mathbf{PC}_i$ is obtained by (Jolliffe, 2002):

$$\frac{\lambda_i}{\sum_i^n \lambda_i} = \frac{\mathbf{Var}(\mathbf{Z}_i, \mathbf{Z}_i)}{\sum_i^n \mathbf{Var}(\mathbf{Z}_i, \mathbf{Z}_i)} \tag{6A}$$

From this procedure, the eigenvectors linked to every $\lambda_i = \mathbf{Var}(\boldsymbol{Z_i},\boldsymbol{Z_i})$ represent the cosine directors (or the contribution which each one of original axes gives to the composition of the new axes), named main components. The eigenvalues, in turn, correspond to the amount of original variance for the respective eigenvectors, following an order of relevance related to every PC: $\lambda_i > \lambda_j > \ldots > \lambda_n$ and with $\lambda_i + \lambda_j + \ldots + \lambda_n = n$.

Each principal component is a linear combination of $\boldsymbol{Z_i}$ variables, defined as (Hair et al., 2019):

$$\mathbf{PC}_k = \sum_{j=1}^{n} \alpha_{j,k} \boldsymbol{Z_j} \tag{7A}$$

where $\boldsymbol{Z_j}$ is the standard j component and the weight $\alpha_{j,k}$ is the j^{th} coefficient for the k^{th} principal component.

Finally, in the statistical sense, two variables, $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$, each one corresponding to n variables and t labels, and their respective averages, $\overline{\boldsymbol{Z}}_i$ and $\overline{\boldsymbol{Z}}_j$, can give results in terms of a correlation coefficient as follows as (Hair et al., 2019; Jolliffe, 2002):

$$\mathbf{Corr}(Z_i, Z_j) = \frac{\mathbf{Cov}(\mathbf{Z}_i, \mathbf{Z}_j)}{\sqrt{\mathbf{Var}(\mathbf{Z}_i, \mathbf{Z}_j)}\sqrt{\mathbf{Var}(\mathbf{Z}_i, \mathbf{Z}_j)}}, \text{for every } i \neq j \tag{8A}$$

for $\boldsymbol{i = j}$, the $\mathbf{Corr}(\boldsymbol{Z_i},\boldsymbol{Z_i})$ is 1, by definition.

From Eq. (8A) it is possible to observe that the correlation coefficient is the covariance of two variables $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$ divided by the product of their squared root variances. Correlation, which depends on covariance, indicates both the strength and direction of the linear relationship between two variables. However, unlike covariance, it is dimensionless.

The correlation coefficient is credited to Galton (1890) and Pearson (1896) and its value should lie between −1 and + 1. A coefficient of +1 specifies that the two variables are perfectly positively correlated: as one variable increases, the other also increases by a comparable quantity. However, this does not mean that the variation in one variable causes the other to change, only that their changes coincide. On the other hand, a coefficient of −1 shows a perfect negative relationship: if one variable increases, the other decreases by a comparable amount. A coefficient of zero implies that there is no linear relationship between the variables.

Both HC and PCA are exploratory multivariate techniques that can be used in any study in which the researcher aims to understand the relationship between variables without the need to estimate data behavior predictions. The main objectives of exploratory models refer to the reduction or structural simplification of data, the classification or grouping of observations and variables and the existence of correlation between metric variables. These are relevant techniques for developing diagnoses about the behavior of data and observations (Fávero and Belfiore, 2019).

While HC is usually used when wants to study similar behaviors between observations in relation to certain metric variables, PCA is usually used to create new variables that capture the joint behavior of the original metric variables. Commonly, they are used independently, but together, these techniques increase the safety and accuracy of the analysis and allow the ordering and allocation of observations in groups that are internally homogeneous and heterogeneous among themselves, in addition to allowing a structural reduction in the number of variables under analysis (Fávero and Belfiore, 2019). It is possible to affirm that such different tools can used in a complementary way to better interpret TH results, as presented in the corresponding paragraphs.

References

Alnafrah, I., Zeno, B., 2020. A new comparative model for national innovation systems based on machine learning classification techniques. Innov. Dev. 10 (1), 45–66.

Balle, A., Steffen, M., Curado, C., Oliveira, M., 2019. Interorganizational knowledge sharing in a science and technology park: the use of knowledge sharing mechanisms. J. Knowl. Manag. 23 (10), 2016–2038.

Basso, F.G., Pereira, C.G., Porto, G.S., 2021. Cooperation and technological areas in the state universities of São Paulo: an analysis from the perspective of the triple helix model. In: Technology in Society, 65.

Bikse, V., Lusena-Ezera, I., Rivza, B., Volkova, T., 2016. The transformation of traditional universities into entrepreneurial universities to ensure sustainable higher education. J. Teach. Educ. Sustain. 18 (2), 75–88.

Boardman, C., 2009. Government centrality to university-industry interactions: university research centers and the industry involvement of academic researchers. Res. Policy 38, 1505–1516.

CAPES - Coordination for the Improvement of Higher Education Personnel, 2019. Georeferenced information system. Available at: https://geocapes.capes.gov.br/geocapes/. (Accessed 6 September 2019).

Carayannis, E., Campbell, D.F.J., 2010. Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other? A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology. Int. J. Soc. Ecol. Sustain. Dev. 1 (1), 41–69.

Carayannis, E.D., Giudice, M.D., Peruta, M.R., 2014. Managing the intellectual capital within government-university-industry R&D partnerships. J. Intell. Capital 15 (4), 611–630.

Castro, B.S., Souza, G.C., 2012. "O papel dos Núcleos de Inovação Tecnológica (NITs) nas universidades brasileiras" [The role of technological innovation centers in Brazilian universities]. Liinc em Revista 8 (1), 125–140.

CNPq - National Council for Scientific and Technological Development, 2019. Tabular plane. Available at: http://dgp.cnpq.br/planotabular/index.jsp. (Accessed 6 September 2019).

Coccia, M., 2005. A taxonomy of public research bodies: a systemic approach. Prometheus 23 (1), 63–82.

Cross, D., Thomson, S., Sinclair, A., 2017. Research in Brazil. Clarivate Analytics, London.

Dalmarco, G., Hulsink, W., Zawislak, P.A., 2019. New perspectives on university-industry relations: an analysis of the knowledge flow within two sectors and two countries. Tech. Anal. Strat. Manag. 31 (11), 1314–1326.

Del Giudice, M., 2008. L'impresa pensante. Giappichelli, Torino.

Del Giudice, M., Carayannis, E.G., Della Peruta, M.R., 2012. Culture and cooperative strategies: knowledge management perspectives. In: Cross-cultural Knowledge Management. Springer, New York, NY.

Dooley, L., Gubbins, C., 2019. Inter-organisational knowledge networks: synthesising dialectic tensions of university-industry knowledge discovery. J. Knowl. Manag. 23 (10), 2113–2134.

Edquist, C., 1997. Systems of Innovation. Pinter, London.

Etzkowitz, H., 2008. The Triple Helix: University–Industry–Government Innovation in Action. Routledge, London and New York, NY.

Etzkowitz, H., Leydesdorff, L., 1997. Universities in the Global Knowledge Economy: A Triple Helix of Academic-Industry-Government Relations. Cassell, London.

Etzkowitz, H., Zhou, C., 2017. The triple helix: innovation and entrepreneurship university–industry–government. Adv. Stud. 31 (90), 23–48.

Eveleens, C., 2010. Innovation management: a literature review of innovation process models and their implications. Science 800 (2010), 900.

Fávero, L.P., Belfiore, P., 2019. Data science for business and decision making. Academic Press, Cambridge.

Flury, B., 1997. A first course in multivariate statistics. Springer, New York, NY.

Freeman, C., 1987. Technology Policy and Economic Performance: Lessons From Japan. Pinter, London.

Galton, F., 1890. In: Kinship and Correlation. N. Am. Rev., 150, pp. 419–431.

González-Masip, J., Martín-de Castro, G., Hernández, A., 2019. Inter-organisational knowledge spillovers: attracting talent in science and technology parks and corporate social responsibility practices. J. Knowl. Manag. 23 (5), 975–997.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2019. Multivariate Data Analysis. Cengage, London.

Hardeman, S., Van Roy, V., Vertesy, D., Saisana, M., 2013. An analysis of national research systems (I): a composite indicator for scientific and technological research excellence. RC Sci. Pol. Rep. EUR 26093 EN.. . DOI: 10.2788/95887.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. J. Ed. Psychol. 24 (6), 417–441.

INPI - National Institute of Industrial Property, 2018. Industrial Property Indicators. Available at: https://www.gov.br/inpi/pt-br/acesso-a-informacao/dados-abertos/arquivos/documentos/indicadores-de-propriedade-industrial. (Accessed 11 October 2020).

Johnston, A., 2020. Open innovation and the formation of university–industry links in the food manufacturing and technology sector: evidence from the UK. Eur. J. Innov. Manag. 24, 89–107.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer, New York, NY.

Jovanović, M.M., Rakićević, J.D., Jeremić, V.M., Levi Jakšić, M.I., 2019. How to measure triple helix performance? A fresh approach. In: Abu-Tair, A., et al. (Eds.), Proceedings of the II International Triple Helix Summit, Lecture Notes in Civil Engineering, 43, pp. 245–261.

Li, M., He, L., Yongxiang, Zhao, 2020. The triple helix system and regional entrepreneurship in China. Entrep. Region. Dev. 32 (7–8), 508–530.

Loi, M., Di Guardo, M.C., 2015. The third mission of universities: an investigation of the espoused values. Sci. Public Policy 42, 855–870.

Luengo-Valderrey, M.J., Pando-García, J., Periáñez-Cañadillas, I., Cervera-Taulet, A., 2020. Analysis of the impact of the triple helix on sustainable innovation targets in Spanish technology companies. Sustainability 12 (8), 3274–3294.

Lundvall, B.Å., 1992. National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. Pinter Publishers, London.

Marques, J., La Falce, J., Marques, F., De Muylder, C., Silva, J., 2019. The relationship between organizational commitment, knowledge transfer and knowledge management maturity. J. Knowl. Manag. 23 (3), 489–507.

Navas, L.P., Montes, F., Abolghasem, S., Salas, R.J., Toloo, M., Zarama, R., 2020. Colombian higher education institutions evaluation. Socio Econ. Plan. Sci. 71 (100801).

Nelson, R.R., 1993. National Innovation Systems. Oxford University Press, Oxford.

Nonaka, I., Takeuchi, H., 1995. The knowledge-creating company: how Japanese companies creates the dynamics of innovation. Oxford University Press, New York, NY.

Oliva, F.L., Semensato, B.I., Prioste, D.B., Winandy, E.J.L., Bution, J.L., Couto, M.H.G., Bottacin, M.A., Mac Lennan, M.L.F., Teberga, P.M.F., Santos, R.F., Singh, S.K., da Silva, S.F., Massaini, S.A., 2019. Innovation in the main brazilian business sectors: characteristics, types and comparison of innovation. J. Knowl. Manag. 23 (1), 135–175.

Pearson, K., 1896. VII. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philos. Trans. R. Soc. 187, 253–318.

Pearson, K., 1901. On lines and planes of closest fit to Systems of Points in space. Philos. Mag. 2, 559–572.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., 2013. Academic engagement and commercialisation: a review of the literature on university–industry relations. Research Policy 42 (2), 423–442.

Petruzzelli, A.M., Albino, V., Carbonara, N., Rotolo, D., 2010. Leveraging learning behavior and network structure to improve knowledge gatekeepers' performance. J. Knowl. Manag. 14 (5), 635–658.

Philpott, K., Dooley, L., O'Reilly, C., Lupton, G., 2011. The entrepreneurial university: examining the underlying academic tensions. Technovation 31 (4), 161–170.

Prahalad, C.K., Hamel, G., 1990. The core competence of the corporation. Harv. Bus. Rev. 68 (3), 79–91.

Prusak, L., 1997. Knowledge in Organizations. Butterworth-Heinemann, Boston, MA.

Romano, M., Del Giudice, M., Nicotra, M., 2014. Knowledge creation and exploitation in italian universities: the role of internal policies for patent activity. J. Knowl. Manag. 18 (5), 952–970.

Rosemberg, N., 1974. Science, invention and economic growth. Econ. J. 84 (333), 90–108.

Sá, E., Dias, D., Sá, M.J., 2018. Towards the university entrepreneurship mission: Portuguese academics 'self-perpective' of their role in knowledge transfer. J. Furth. High. Educ. 42 (6), 784–796.

Sneath, P.H.A., Sokal, R.R., 1962. Numerical taxonomy. Nature 193, 855–860.

Soares, T.J., Torkomian, A.L.V., 2021. TTO's staff and technology transfer: examining the effect of employees' individual capabilities. Technovation 102, 102213. https://doi.org/10.1016/j.technovation.2020.102213.

Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38 (22), 1409–1438.

Sokal, R.R., Sneath, P.H.A., 1963. Principles of Numerical Taxonomy. W. H. Freeman and Company.

Souza, C.D., Filippo, D., Casado, E.S., 2015. "Impacto do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais Brasileiras (REUNI) na atividade investigativa: Crescimento, qualidade e internacionalização" [Impact of the Support Program for Restructuring and Expansion Plans of Brazilian Federal Universities (REUNI) in the investigative activity: Growth, quality and internationalization]. Em Questão 21 (13), 336–367.

Sydow, J., Schubler, E., Muller-Seitz, G., 2016. Managing Inter-organisational Relations. Palgrave Publishing, London.

UFBA - Federal University of Bahia, 2019. Reports. Available at: https://proplan.ufba.br/documentacao-legislacao/relatorios-gestao. (Accessed 6 September 2019).

UFPE - Federal University of Pernambuco, 2019. Reports. Available at: https://www.ufpe.br/proplan/relatorios-de-gestao. (Accessed 6 September 2019).

UFPR - Federal University of Paraná, 2019. Reports. Available at: http://www.proplan.ufpr.br/portal/relatorio-de-gestao/. (Accessed 6 September 2019).
UFRGS - Federal University of Rio Grande do Sul, 2019. Reports. Available at: http://www.ufrgs.br/ufrgs/a-ufrgs/relatorios. (Accessed 6 September 2019).
UFSCar - Federal University of São Carlos, 2019. Reports. Available at: http://www.spdi.ufscar.br/documentos/relatorio_contas/. (Accessed 6 September 2019).
UFV - Federal University of Viçosa, 2019. Reports. Available at: https://www.dti.ufv.br/relatorioufv/. (Accessed 6 September 2019).
Unesp - São Paulo State University "Júlio de Mesquita Filho", 2019. Statistical Yearbooks. Available at: https://ape.unesp.br/anuario/. (Accessed 6 September 2019).
Unicamp - State University of Campinas, 2018. Statistical Yearbooks. Available at: https://www.aeplan.unicamp.br/anuario/2018/anuario2018.pdf. (Accessed 6 September 2019).
Wang, W., Lu, S., 2021. University-industry innovation community dynamics and knowledge transfer: evidence from China. Technovation 106 (in press).
Ward Jr., J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58 (301), 236–244.
Web of Science, 2019. Research in Brazil: funding excellence. Clarivate Analytics, London.
Yoon, J., Yang, J., Park, H., 2017. Quintuple helix structure of sino-korean research collaboration in science. Scientometrics 113 (1), 61–81.
Zhao, S.L., Cacciolatti, L., Lee, S.H., Song, W., 2015. Regional collaborations and indigenous innovation capabilities in China: a multivariate method for the analysis of regional innovation systems. Technol. Forecast. Soc. Chang. 94 (1), 202–220.

Eron Passos Andrade is bachelor in Production Engineering and Systems from State University of Santa Cruz (2010), M.Sc. in Industrial Engineering from Federal University of Bahia (2016) and a Ph.D. student in Graduate Program in Industrial Engineering from Federal University of Bahia. Now is Lecturer at Energy and Sustainability Science and Technology Center, Federal University of Recôncavo da Bahia. He has experience in Production Engineering, with emphasis in Organisational Engineering, working mainly on the following topics: innovation, strategic planning, and planning and control of production.

Jadiel dos Santos Pereira is bachelor in Physics from State University of Feira de Santana (2009), M.Sc. in Physics from Federal University of Bahia (2011) and Ph.D. in Physics from Federal University of Bahia (2018). Now is Assistant Professor at Energy and Sustainability Science and Technology Center, Federal University of Recôncavo da Bahia. He has experience in Physics with emphasis in thermology, carbon, nanotubes and nanotechnology.

Angela Machado Rocha is bachelor in Chemical Engineering from Federal University of Bahia (1988), Ph.D. in Energy and Environment from Federal University of Bahia (2013), with a sandwich research at Georgia Institute of Technology, United States. Now is Assistant Professor at the Institute of Health Sciences and Advisor to the Coordination of Innovation at Federal University of Bahia. She leads the Sustainable Innovation and Entrepreneurship Research Group at UFBA.

Marcio Luis Ferreira Nascimento is a Brazilian physicist from University of São Paulo (USP), Ph.D. in Materials Science and Engineering from Federal University of São Carlos (2004). Now he is Professor at Federal University of Bahia (UFBA). He teaches math, applied statistics, chemical engineering and materials science for undergraduate and graduate students. Currently he has a keen interest in subjects as glasses, innovation, history of sciences, multivariate analysis, machine learning, *Big Data* and in particular, mathematics for laymen.