Innovation, Networks and Emerging Jobs: A Bibliometric Study Of the Web of Science's Database, From 1945 To 2020

Leandro Rachel Arguello[1,a], Michel Angelo Constantino[g], Antonio Carlos Dorsa[g], Diego Bezerra de Souza[g], Flávio Henrique Souza de Araújo[g], Tiago Teixeira Pereira[g], Cristiane Martins Viegas de Oliveira[g]

[a]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil [b]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil [c]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil [d]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil
[e]Federal University of Grande Dourados, Mato Grosso do Sul, Brazil [f]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil [g]Catholic University Dom Bosco, Mato Grosso do Sul, Brazil

Abstract

This research presents a bibliometric study of scientific productions involving the themes of innovation, networks and emerging jobs in the English language, available in the Web of Science's database, dating from 1945 to 2020. Seeking to clarify some findings, in the statistical study of Web of Science's publications, this research demonstrates the orientation of the reasoning of the investigated authors, regarding innovation policies. Of the 49 publications found in a previous screening, we obtained the number of 44 publications in English, which were submitted to statistical, Reinert and Similitude analyses, using the IRAMUTEQ software. During the research, it was possible to notice that this group of words presented a relation in the context of the texts available in the base; that is: the tool presented, statistically, four groups of words with greater relevance, connecting the three themes. In the analysis, IRAMUTEQ presented a division in four clusters (using Reinert's method) that list the groups of words by their relation of higher incidence and correlation during the study (development, literature, information and form). According to this mapping, it is possible to conclude that the correlation of themes refers to a group of researchers who, in their discourses, connect technological and economic development to relations in an innovation network, and that 50% of the published texts dealt with issues involving the four largest groups of words: development, network, innovation and business.

Keywords: Innovation, Development, Economics, Statistical data, IRAMUTEQ.

1. Introduction

The constant transformation of the technological environment is part of a natural movement in the search for better productive capacity and new processes for it; according to [8], investments in technology are necessary to maintain the economic competitiveness of nations.

5 The innovation of production processes and the need to remain economically competitive on the global stage lead nations to develop challenging and highly integrated environments for such economic growth. In this sense, [4] points out, in the paper "Innovation Network", the complex policy attitudes of countries that have achieved and are achieving economic success, and of nations that will surely succumb to the global development movement.

Investigations on the advanced integrated system that drives technological evolution reverse to a country or to

10 a specific region countless benefits in an integrated chain of agents of the local scenario. According to [1], this comprehension of innovative factors refers to new job vacancies and improved economic productivity. [7] presents the opinion that a study on innovation and its application in practice contributes to understanding the movement of the labor supply market and its possible substitutions.

Studies related to technological development are also linked to the economy and to the development of specific

15 locations and even large nations;[7] states that such studies and investments have a positive impact on a country's economic development, agreeing with [4] — who, during their studies, present convincing justifications about the importance of investments in technology and innovation for the progress of technological development of nations.

One of the examples of this developmentalist movement is that of investments in mobile internet technology connection. According to 2019 data from the Big Data of *Our World in Data*, in the last five years, access to 4G

20 mobile internet technology has grown more than four times. This has opened space even for investments and the promotion of transformations in the so-called Digital Economy, which deals with the intensive use of Information and Communication Technologies (ICTs) in all branches of economy, including new ways of functioning and interaction between individuals, companies and the State.

The promotion of this Digital Economy foresees the interaction of innovation, networks and future skills for the

25 new generation of 5G technologies and Internet of Things, as presented by [11]. In order to understand the conceptual and investigative approaches to science, this study has brought together the three themes to assess how research is directing its scientific contributions.

Therefore, this research aims to present a bibliometric study of scientific productions involving the themes of innovation, networks and emerging jobs in English, available in the Web of Science's database, dated from 1945 until

30 2020.

Applying a traditional bibliometric technique, IRAMUTEQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) — a textual statistical analysis tool — was used.

Considering 44 publications in English, from the Web of Science's database, abstracts of each publication were extracted, and the text analysis tools were applied. As a result, they have presented, through graphs and tables, 35 significant data of the words with the highest incidence in the authors' writings.

The results of this production were divided into an initial study of the orientation of origin and areas of study of these scientific productions, and, later, they were presented through a Reinert Analysis, a Similitude Analysis and word cloud disclosure.

2. Related Studies

40 2.1. Bibliometry

Bibliometry consists of a technique for mapping information or data pertinent to a study in discussion, applying mathematical models to achieve more accurate results. As stated by [15], this study has been gaining visibility and becoming fundamental in scientific studies. Complementary, according to [13], this technique, originated in library and information science (LIS), became fundamental in the methodology of scientific research.

45 It is not, therefore, a new methodology, but a format developed between the 15th and the 19th centuries, according to [15]. Furthermore, this method, which seeks to quantify and organize research data, has been used in several areas, such as management, economics, health, innovation, among others, as [13] state.

The objective of using bibliometrics in scientific research has, as a final product, a more objective analysis of the research [5], presenting the trajectory of this research technique.

50 2.2. Innovation

Innovation has been a topic discussed since the beginning of studies on the behavior of human evolution itself. However, it started to be more debated and studied from the great revolutions that transformed the world and the forms of production: the innovation that follows a permanent behavior is directly linked to the need to improve traditional production processes and the use of resources. [8] incite their readers to observe these factors from an economic 55 perspective, which is, however, expanded in many other areas of knowledge.

Daron Acemoglu and Pascual Restrepo address deep specificities among the themes cited in their paper on artificial intelligence and work [3], and in the study about men and machines [2], paying attention to problems that may arise due to the need for growth and the development of technology.

In the era of fast connections, social networks and global connectivity — as it has already been happening in the

60 third decade of the 21st century —, the innovation process needs to be understood and guided so that we have the understanding between man and the relationship of advancement in innovation technology, states [7], while presenting studies of the Italian model.

The optimization of resources is also present in a large part of the innovation process, as stated by [16], seeking suggestions to integrate a sustainable behavior between government, industry and society.

65 In this context, innovation is linked to the transformation of existing processes, which, in turn, make use of new technologies to achieve better results regarding the effectiveness of this transformation.

The economist Daron Acemoglu applies this distinction, connecting technology to the necessary transformations in activities aimed at innovation, for countries that seek better performance in reaction to global competitiveness.

2.3. Network

70 As discussed in the previous subtopic, the innovation process is connected to a network of events and attitudes, as [4] state, when, in their studies, they deal with the relationship between the development of a country and the deep connection with investments in technological advances.

Within the comprehension of a complex and integrated network, it is possible to learn more by observing the great revolutions, which, in turn, are characterized by developments in different sectors and areas of society, leveraged by 75 strong technological development.

According to Manuel Castells Oliván, these complex connections that form the development are called innovation networks. In addition, the author deals with each cycle of this digital revolution, exposing how each period and its complex connections were structured.

This network is reinforced by the understanding of the effect new technologies have in the relationship between

80 man, society and work, as presented in the papers of [7], regarding the Italian model of digitization of the production process, and [12], on the relationship of Artificial Intelligence and the future of jobs.

The comprehension of networks also leads to a systemic understanding, which allows the identification of vulnerabilities in a given cycle of transformation, from public policies to direct cultural impacts, state [12], when listing factors of this development, while dealing with the complexity of the relationships of the factors that promote the 85 advance of innovation and development.

In [7], these policies appear more clearly, and the author also discusses the impacts on employment. When observing these networks, a high level of complexity and connection between them is noticeable, according to [11] and [2], as they debate networks and the impact on the work and development of a country.

2.4. Emerging Jobs

90 Another theme, also connected to the previous ones, relates results and impacts regarding employment.

According to [8], these impacts are seen in all revolutions; jobs are replaced, and some professions end up emerging amid these cycles, according to [3], as they discuss the relationship of Artificial Intelligence and impacts on employment.

These data have already become the reality, according to the emerging professions report of [10] - the current largest social network for relationships between professionals and companies. In this document, among the fifteen

95 emerging professions listed, there are the job positions of social media manager, cybersecurity engineer, data scientist, data engineer, AI specialist, JavaScript programmer, IT specialist recruiter and Agile methodology coach; all of these positions have emerged from technological evolution.

These and other professions are related to the process of digitalization of the world and its connections, according to [7], when it comes to the evolution of the market and the environment in which we operate. According to [4], this 100 evolution of occupations is a market need to absorb demand for qualified and highly specialized labor in technology.

3. Methodology

The study of bibliometrics consists of a technique of selection and specific analysis of a given text or group of selected information. Bibliometric analysis requires the selection of a base, delimiting the field of sampling and searches of the selected subjects, as state [14].

105 The data of the bibliometric analyses of this paper were extracted directly from the database Web of Science (hereinafter WoS), a platform originally called Web of Science Knowledge, which was developed in order to index (and, therefore, concentrate) scientific publications. Previously, WoS was maintained by the Institute for Scientific Information (ISI), and is currently managed by Clarivate Analytics, another company in the segment of concentration of scientific journals.

110 The WoS database tools allow the use of variation filters such as year, language, relevance, author, subject and availability, and the selection of databases, such as the *Science Citation Index Expanded*, the *Social Sciences Citation Index*, *Arts Humanities Citation Index*, *Emerging Sources Citation Index, Book Citation Index* and *Conference Proceedings Citation Index*.

The platform also includes consultations on patent projects and studies in development, which makes the brand a 115 strong reference in scientific-academic research.

4. Corpus constitution

The WoS search base has more than 160 million publications. Considering this universe of possibilities, this study started with a previous screening of productions published in the entire available base of WoS. The search for the keywords 'innovation', 'network' and 'emerging jobs', carried out on February 6, 2020, presented 49 results; after a

120 new filter for the English language, 44 publications remained, as it is shown in table 1. Some articles, therefore, were not considered, for not meeting the prerequisite of the English language.

Tabela 1: Research table

Keywords	Total of titles in the Web of Science's database
Innovation; Network; Emerging jobs	Total of titles in the last 85 years = 49
	Total of publications = 49
	Total of publications in English = 44

Source: The author, 2020.

5. Analysis methods and tools

The statistical analysis of the text under study is part of the process of building the ideas and the investigation of

125 the facts. The review, previously carried out using non-systemic methodologies, and with total dependence on the interpretation of the author or researcher, could have flaws and biased interpretations. With the advent of the microcomputer and the accelerated technological movement, means have emerged to make more creative and systematized research in this context [9].

The current practice of discourse analysis, according to [15], consists in the application of statistical models in

130 the content of the corpus, repeating the opportunity to visualize other forms of systematic interpretation. [14] affirm that qualitative analyses have been widely used as alternatives to a different understanding of the texts, having, as the main foundation, the structuring and the organization of the data, which, in the future, will be part of the analysis of the textual corpus.

To the data, three types of analyses were applied. The first consists of a statistical view of distribution, in a global

135 perspectives of publications, through a map of trees; it shows representativeness by areas and lists the countries of the journals in discussion, allowing a clear view of those which have more publications.

In the second, we present an analysis of Reinert, demonstrating a verification of cluster formation. According to [6], through a reading of the discourse's vocabulary, it is possible to sort the groups based on how often they appear.

Finally, from a Similitude analysis, we present the correlation of the texts graphically, therefore demonstrating the

140 links between publications and terms with a higher incidence of writing. Also, according to [6], the Similitude analyses allow researchers to identify the correlations between the most important words in the discourse in debate.

For these analyses, the software IRAMUTEQ *(Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires)* [6], which was developed by Pierre Ratinaud in France in 2008, was chosen. The software is available free of charge, and has been widely used as a support tool in textual statistical analysis.

145   6. Results and analyses

The first analysis lists, in a tree graph, publications, and shows a global distribution of scientific research on the previously selected subjects. This representation includes six countries with three or more publications, totaling thirty publications for the six identified countries.

As shown in figure 1, firstly, there is the United States (with thirteen publications), leading with a number three

150 times greater than the second, England (four publications); subsequently, France, Germany, India and Romania have three publications each.

Figura 1: Countries and number of publications
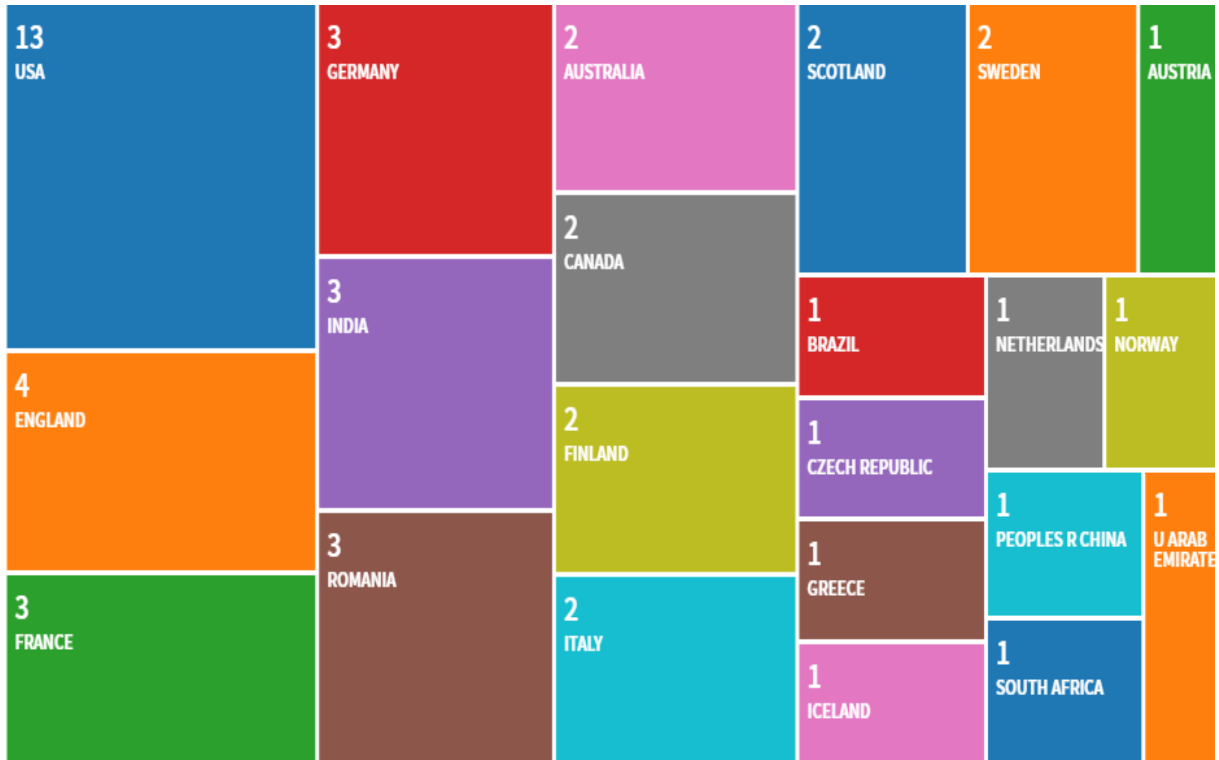
Source: Web of Science, 2020.

By analyzing the areas with more publications, it is possible to verify that a quarter of them stand out for their relevance in publications; first, at the top of the ranking, 36.36% of the base is focused on the management area. This is followed by business (25% of base) and economics (13.66% of published studies), as shown in table 2.

Tabela 2: Publication category

Category	Number of publications	% of 44 publications
Management	16	36,36 %
Business	11	25%
Economics	6	13,63%
Computer Sciences Information Systems	4	9,09%
Computer Sciences Software Engineering	3	6,81%
Computer Sciences Theory Methods	3	6,81%
Education and Educational Research	3	6,81%
Communication	2	4,45%
Development Studies	2	4,45%
Applied Psychology	2	4,45%
Social Sciences Interdisciplinary	2	4,45%

Sociology	2	4,45%
Telecommunications	2	4,45%
Computer Artificial Intelligence	1	2,27%
Computer Hardware Architecture	1	2,27%
Computer Sciences Interdisciplinary Applications	1	2,27%
Environmental Engineering Sciences	1	2,27%
Environmental Sciences	1	2,27%
Geography	1	2,27%
Physical Geography	1	2,27%
Multidisciplinary Geosciences	1	2,27%
Hospitality, Leisure, Sport and Tourism	1	2,27%

Source: The author, 2020.

155 When combined, the three areas represent more than 60% of the 44 analyzed papers. This demonstrates the relevance of the facts to management and business areas and can imply the strong connection of these fields with the processes of economic and productive development, justified by [4] as they affirm that there is a strong correlation between a developed nation and the continuous improvement of its investment process in new production technologies and business transformation.

160 In turn, [8] indicate the investments in technology production as a very important factor in the development of the countries' economies, considering the examples of industrial revolutions - which, according to the authors, are seen as milestones in the evolution of production processes and in the economic models.

Most of the base refers to studies related to economics and business, with 61.36% of published studies, which presents a strong feature of the productions for an economic vision.

165 6.1. Primary data analysis

44 texts, containing 308 segments responsible for connections of the discourses, and 11,286 occurrences, with 2,016 forms and 954 heptax (words that appear only once in the entire analysis) were analyzed, as it is shown in graph 2

Figura 2: Frequency and words amount
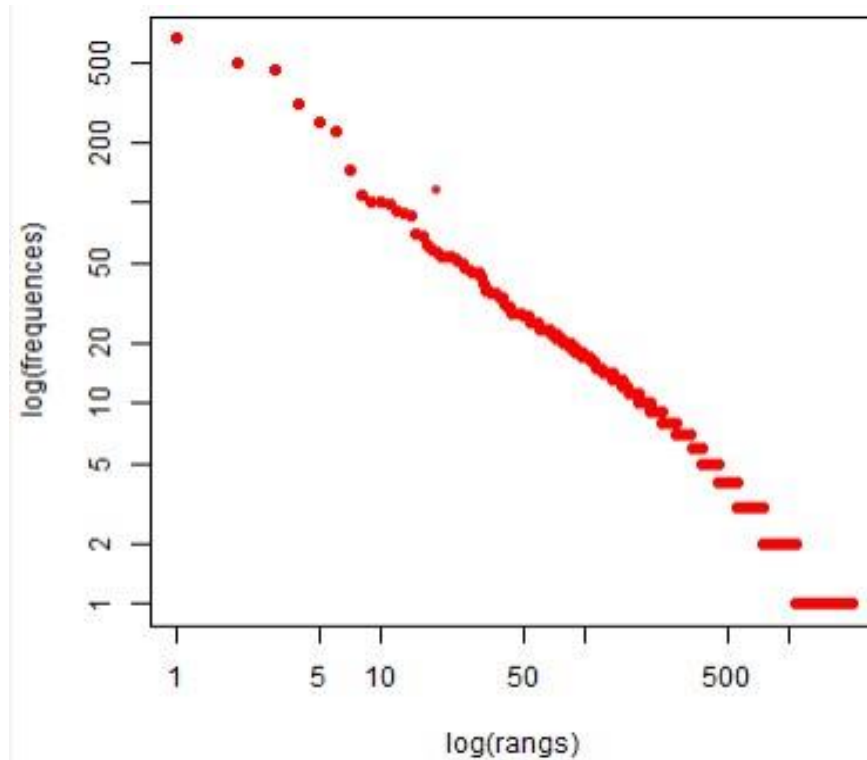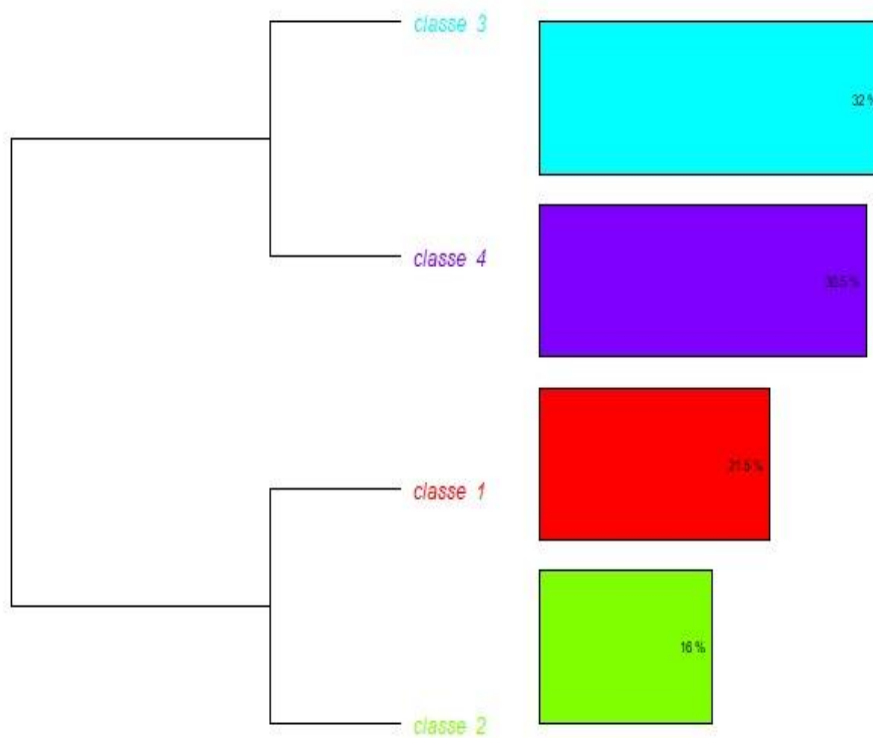
Source: The author, 2020.

This graph shows the crossing of frequencies with the amount of words; a word, for example, appeared more
than

170 500 times, while more than 600 words were cited only once in the textual corpus, justifying the data
stratified by the system.

6.2. Reinert's method analysis

In the analysis using the Reinert's method, out of the 44 texts, the 308 classes (again presented) were
divided into four new classes of groupings, with 256 segments classified in each.

175 These four analyzed classes presented the following results (GRAPH 3): the fourth class had 32.06% of
relevance, with 86 word correlations; the third class had 30.47%, totaling 76 correlations; the first class
presented 21.48% (53) correlations; the second class had 16.2%, with 41 word correlations. Classes 3 and
4 presented more than 60% of the method's relevance.

Figura 3: Dendogram graph
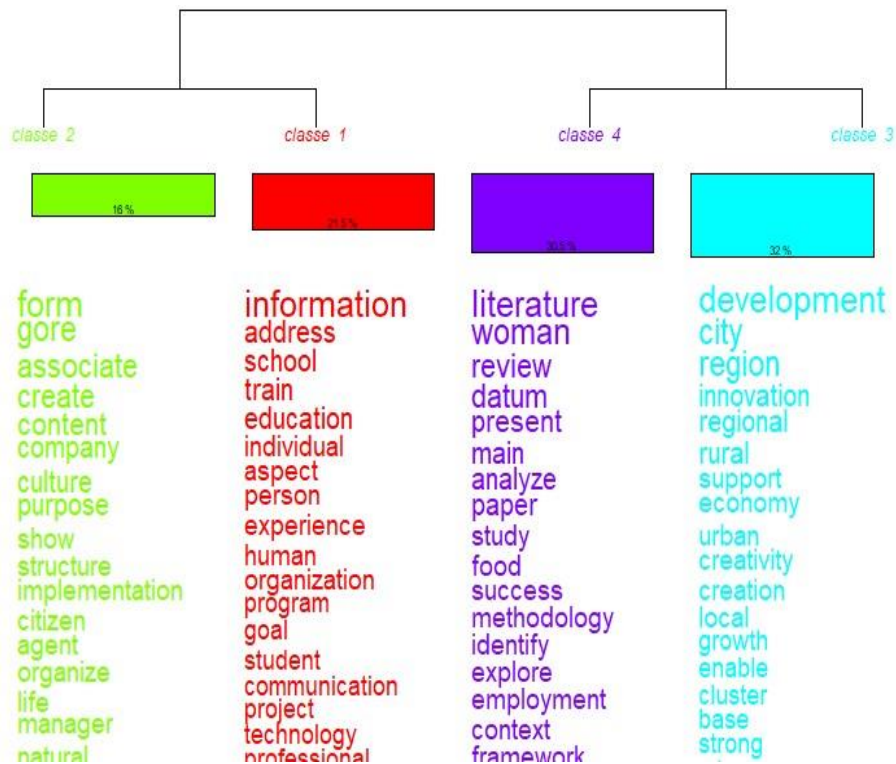
Source: The author, 2020.

The representation by grouping of words of class 3 presented, as its first five words, 'development', 'city', 'region', 180 'innovation' and 'regional' . In this class, it is possible to identify the word 'innovation' in the fourth position, referring to the first keyword of the basic search for textual corpus, as shown in graph 4.

Figura 4: Classes of word clusters

Source: The author, 2020.

For the fourth class, the word 'employment' appears in the 12th position, connected to the term 'emerging jobs' and the keywords of the search's source. The same term also belongs to the class that mentions 'literature', which is strongly linked to 'woman' and continues as the second greater incidence in the text.

185 For the two most relevant classes to the analysis, it is possible to assimilate the context of regionality and development, which, connected, refer to the knowledge formation and connectivity, according to the data in table 2, which shows the percentage of 'management', 'business' and 'economy'.

In classes 2 and 1, it is possible to connect 'information' to 'form' and 'gore'. The classes of observations, together, also refer to aspects that direct the corpus to the interaction structures of technology and implementation in

190the environment.

The relationships of the fourth and the third groups reinforce concepts of development, as they strongly refer to 'growth', 'local' and 'employment' - aspects that make up part of the success in developing a specific location.

6.3. Similitude analysis

From the Similitude analysis - or graph analysis -, it is possible to point out and identify the link between the most

195 representative words in the textual corpus and groups that stand out from each other, connecting different discourses and dialectics about given subjects.

During the research, the model presented four major written references: 'network', 'innovation', 'development' and 'emerge'. The greatest emphasis is on the first two, as it is shown in graph 5.

Figura 5: Word reference clusters
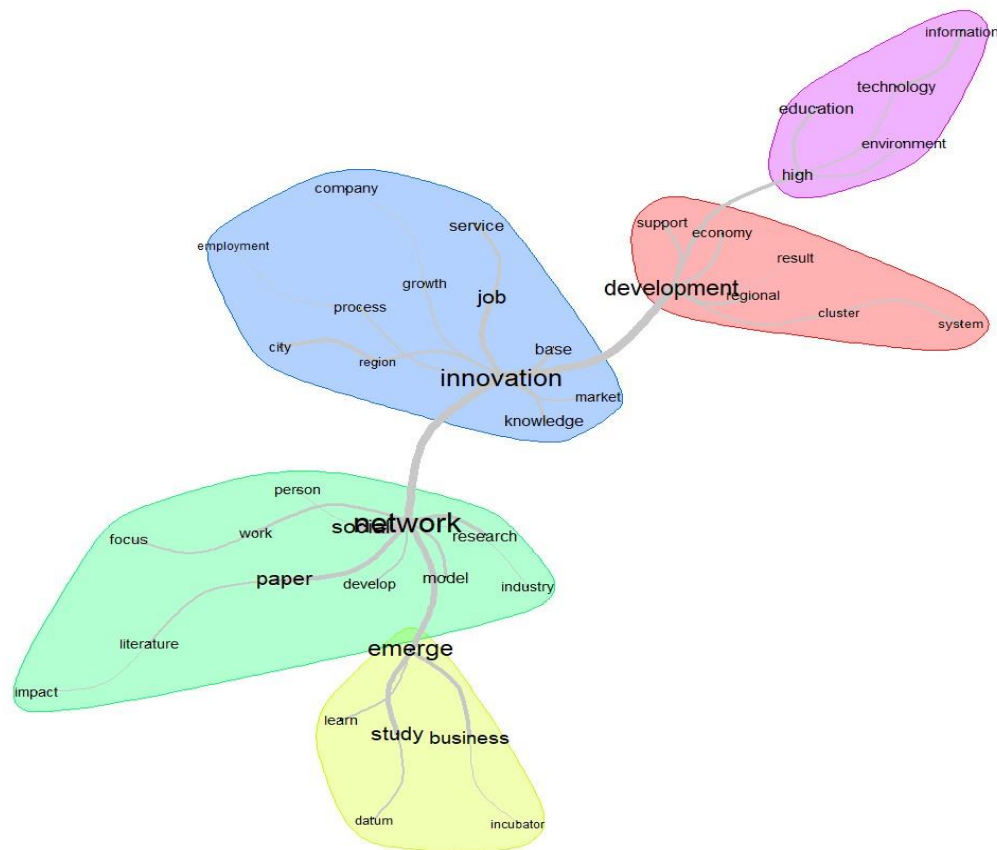
Source: The author, 2020.

When analyzing, from the largest to the smallest groups, it is possible to notice that 'network' presented the 200 highest concentration of words ('society', 'paperwork', 'person', 'focus', 'develop', 'model', 'research', 'industry', 'literature' and 'impact'). These terms, when grouped, refer to the sense of research and development, connected with society, producing development-focused models.

The second largest grouping of words has as its main reference 'innovation', which is connected to ten other written references ('job', 'base', 'market', 'knowledge', 'growth', 'process', 'employment', 'service', 'region' and

205 'company'). As they are grouped, the terms present an employment context linked to knowledge, reinforcing the growth dialectic with services that are driven by the market.

The third prominent group has the word 'development' connected to 'economy', 'regional', 'result' and 'support', a discourse geared towards the formation of clusters that develop direct economic and driving factors of a system, oriented to results.

210 Next to 'network', 'emerge' was identified; it also brings in its context 'business'and 'study', an orientation to learning ('learn') that is directly connected to words grouped in the 'network' cluster.

On the opposite side of the others, a fifth group of words presents a connection that involves 'education', 'technology' and 'information', which directly addresses the relationship with the other groupings.

6.4. Relevant words cloud

215 The results of the analyses connect different papers in relevant groups of words. Among the 44 analyzed publications, the words that stood out the most were 'innovation', 'network', 'development', 'emerge' and 'jobs', which form the group with the greatest weight of inferences, as it is possible to visualize in the word cloud shown in graph

6.

The word cloud criterion is to group, by minimum number, words that appear at least twenty times throughout the 220 textual corpus, only pointing out words with the highest number of incidences.

A second group, consisting of the terms 'technology', 'business', 'study' and 'knowledge' is connected in a concept aimed at development and network economy, regarding most publications in management, business and economy.

For more than 50% of the publications that deal with subjects involving the largest groups of words - 'deve-
225 lopment', 'network', 'innovation', 'business' and 'emerge' -, this group can be interpreted as a representation of the relationship between the development of business innovation networks directly, justifying the connection of the researched subjects.

Following the line of thought of authors such as Daron Acemoglu, Daniela Freddi and Jani Lukkarinen, this conjunction makes up the relationship of the factors that project the direction of nations for development and growth.

Figura 6: Word cloud

Source: The author, 2020.

230   7. Final considerations

The analysis indicates, in a systemic view, the connection of the discourses, and evidences the deep relationship of networks in the innovation process, as the studies of the aforementioned authors affirm the connection between economic growth and reallocation of human resources employed in production.

We conclude, therefore, that in all the analyses the same group of words - innovation, development and network 235 appears with a higher frequency of connection in the written architecture.

This fact refers to the need to understand the complex system that makes up the structure of innovation policies and their relationship with employment; comprehending its progress when addressing policy direction measures adopted to improve the evolution process impacts on the whole environment - and not just in the economy.

The group of analyzed texts demonstrated the complex link between factors and attitudes that involve the innova-

240 tion process; this network of connected words gives meaning to an orientation towards the attitude of innovating and developing innovation.

Therefore, corroborating the cited authors, the initially researched words pointed to an integrated innovation model, in which research and learning can be considered important for social, technological and economic development.

Referências

245   [1] , . Políticas "complicadas" - casos de três países sobre comprometimento de longo prazo e enraizamento das metas de política de er 13. URL: http://dx.doi.org/10.3390/en13061351, doi:10.3390/en13061351.

[2] Acemoglu, D., Restrepo, P., 2018. The race between man and machine: Implications of technology for growth, factor shares, and employment. American Economic Review 108, 1488–1542.

[3] Acemoglu, D., Restrepo, P., 2019. The wrong kind of AI? Artificial intelligence and the future of labor demand. Technical Report. National 250 Bureau of Economic Research.

[4] Acemoglu, Daron e Akcigit, U.e.K.W.R., . Rede de inovação. Anais da Academia Nacional de Ciências 113, 11483–11488.

[5] Araújo, C.A., 2006. Bibliometria: evolução histórica e questões atuais. Em questão 12, 11–32.

[6] Camargo, B.V., Justo, A.M., 2013. Iramuteq: um software gratuito para análise de dados textuais. Temas em psicologia 21, 513–518. [7] Freddi, D., . Digitalização e emprego na fabricação. AI   SOCIETY 33, 393–403.

255 [8] Freeman, C., Soete, L., 1997. The economics of industrial innovation. Psychology Press.

[9] de Freitas, H., da Cunha Júnior, M., Moscarola, J., 1997. Aplicação de sistema de software para auxílio na análise de conteúdo. Revista de Administra&ccdeil; ão da Universidade de São Paulo 32.

[10] In, L., 2020. Emergin jobs 2020. Linked In 1, 1–25.

[11] Lee, C., Kim, H., 2018. The evolutionary trajectory of an ict ecosystem: A network analysis based on media users' data. Information & 260 Management 55, 795–805.

[12] Lukkarinen, J., Berg, A., Salo, M., Tainio, P., Alhola, K., Antikainen, R., 2018. An intermediary approach to technological innovation systems (tis)—the case of the cleantech sector in finland. Environmental innovation and societal transitions 26, 136–146.

[13] Merigó, J.M., Yang, J.B., 2017. A bibliometric analysis of operations research and management science. Omega 73, 37–48.

[14] Mozzato, A.R., Grzybovski, D., 2011. Análise de conteúdo como técnica de análise de dados qualitativos no campo da administração:

265 potencial e desafios. Revista de Administração Contemporânea 15, 731–747.

[15] Santos, R.N.M.d., Kobashi, N.Y., 2009. Bibliometria, cientometria, infometria: conceitos e aplicações .

[16] Silvestre, B.S., Țîrca, D.M., 2019.˘ Innovations for sustainable development: Moving toward a sustainable future. Journal of Cleaner Production 208, 325–332.