Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Analysis model of scientific production in Postgraduate Programs based on Interaction Networks: A Case Study in Environmental Sciences

Modelo de análise da produção científica em Programas de Pós-Graduação com base em redes de interação: estudo de caso nas Ciências Ambientais

Washington de Jesus Sant'Anna de Franca-Rocha[1], *Rodrigo Nogueira Vasconcelos*[1], *Joselisa Maria Chaves*[1], *Patricia Bilotta*[2], *Isabel Jurema Grimm*[3,4], *Sónia Maria Carvalho Ribeiro*[5], *Rodrigo Affonso de Albuquerque Nobrega*[5], *Maria do Carmo Martins Sobral*[6], *Arlindo Philippi Junior*[7], *Carlos Alberto Cioce Sampaio*[8,7,9]

ABSTRACT

The ongoing evaluation of graduate programs (GP) is an important tool for improving the landscape of scientific knowledge and technological development in Brazil. The objective of this article was to develop a model for analyzing the scientific production of GP, based on interaction network resources with a focus on environmental sciences, for the purpose of detecting patterns and connections existing among these networks. The scope of the analysis was broad, encompassing all 112 GP, in addition to courses in the field of Environmental Sciences offered by the Coordination for the Improvement of Higher Education Personnel (CAPES), during the period from 2013 to 2016. The methodology was divided into four stages: 1. data collection and database construction (public information obtained by consulting the Sucupira Platform — CAPES); 2. data mining and processing, and the creation of an overarching network to represent the most relevant terms and themes common to all GP materials produced (theses, dissertations, research projects, articles, books, and book chapters); 3.

RESUMO

A avaliação permanente dos Programas de Pós-Graduação (PPG) constitui importante instrumento para melhoria no cenário de conhecimento científico e no desenvolvimento tecnológico no Brasil. O objetivo do artigo foi desenvolver um modelo para análise da produção científica dos PPG, utilizando recursos de redes de interação, com recorte para as ciências ambientais, visando detectar padrões e conexões entre eles. O recorte da pesquisa abrangeu todos os 112 PPG e os cursos da Área de Ciências Ambientais da Coordenação de Aperfeiçoamento de Pessoas de Nível Superior (CAPES), no período de 2013 a 2016. A metodologia foi dividida em quatro etapas: 1. coleta de dados e construção de um banco de dados (informações públicas obtidas em consulta à Plataforma Sucupira — CAPES); 5. mineração e processamento dos dados e a produção de uma rede geral para representar os termos e os temas de maior relevância comuns a todas as produções dos PPG (teses, dissertações, projetos de pesquisas, artigos, livros e capítulos de livros); 3. análise das redes semânticas; 4. geração de produtos. Os resultados

[1]Universidade Estadual de Feira de Santana – Feira de Santana (BA), Brazil.
[2]Centro Universitário Augusto Motta – Rio de Janeiro (RJ), Brazil.
[3]Instituto Superior de Administração e Economia do Mercosul – Curitiba (PR), Brazil.
[4]Universidade Santa Úrsula – Rio de Janeiro (RJ), Brazil.
[5]Universidade Federal de Minas Gerais – Belo Horizonte (MG), Brazil.
[6]Universidade Federal de Pernambuco – Recife (PE), Brazil.
[7]Universidade de São Paulo – São Paulo (SP), Brazil.
[8]Fundação Universidade Regional de Blumenau – Blumenau (SC), Brazil.
[9]Universidade São Judas Tadeu – São Paulo (SP), Brazil.
Correspondence address: Washington de Jesus Sant'Anna de Franca-Rocha – Avenida Transnordestina, s/n – Novo Horizonte, UEFS, Módulo V/ PPGM – CEP: 44036-900 – Feira de Santana (BA), Brazil. E-mail: wrocha@uefs.br

Conflicts of interest: the authors declare no conflicts of interest.
Funding: National Council for Scientific and Technological Development (CNPq), grant n. 407907/2022-0 (author WJSFR); Coordination for the Improvement of Higher Education Personnel (CAPES); National Foundation for the Development of Private Higher Education (FUNADESP); Ânima Institute; São Paulo Research Foundation (FAPESP).
Received on: 05/04/2023. Accepted on: 08/26/2023.
https://doi.org/10.5327/Z2176-94781619

analysis of semantic networks; and 4. generation of outcomes. The analysis yielded results such as geographical proximity and cluster maps, which allowed for an integrated analysis of GP production in the field of Environmental Sciences with respect to their central themes. The methodology employed proved to be robust and suitable for evaluating graduate programs in Brazil, as well as for identifying research gaps and emerging areas on a national scale, culminating in a proposed model based on semantic networks that analyze scientific production for four-year periods.

Keywords: evaluation of graduate programs; Sucupira Platform; semantic networks.

da pesquisa retornaram como produtos mapas de proximidade geográfica e de agrupamentos, que permitiram analisar a produção dos PPG das Ciências Ambientais em relação aos seus temas centrais, de forma integrada. A metodologia empregada se mostrou bastante robusta e apropriada para contribuir na avaliação dos programas de pós-graduação brasileiros, bem como para identificar lacunas na pesquisa e áreas emergentes em escala nacional, culminando em uma proposta de modelo que, a partir das redes semânticas, analisem as produções científicas nas avaliações quadrienais.

Palavras-chave: avaliação de Programas de Pós-Graduação, Plataforma Sucupira, rede semânticas.

Introduction

Although development and education are abstract concepts, they include practical aspects that make them suitable for application in real-world contexts. The concept of development assumes the pursuit of well-being in specific areas with the aim of improving the quality of life for humans and the environment across various timeframes. This requires the establishment of a sustainable environment that encompasses multiple dimensions (Sachs, 2004; Max-Neef et al., 2012; Florit et al., 2021). Conversely, education, more precisely higher education in the form of graduate programs (GP), plays a pivotal role in the training of skilled individuals capable of generating intellectual contributions that impact the scientific community (Brasil, 2010).

The Coordination for the Improvement of Higher Education Personnel (*Coordenação de Aperfeiçoamento de* Pessoal *de Nível Superior* — CAPES) exists to promote and consolidate *Stricto Sensu* Postgraduate Programs (GP) in all Brazilian states. Its mission is to aid in the training of highly qualified professionals for work in both the academic and non-academic sectors of the Brazilian economy (CAPES, 2022). The quality of GP is evaluated according to criteria established by CAPES, guided by instructions from the National Postgraduate Plan (PNPG) (Cabral et al., 2020), and field-specific documents (CAPES, 2022). The evaluation carried out in 2017 covered the four-year period from 2013 to 2016 (CAPES, 2017a).

Research conducted within the scope of the Environmental Sciences GP fundamentally focuses on understanding and generating solutions to contemporary challenges that are inherently socio-economic and technological. The adoption of an interdisciplinary perspective is an essential aspect of the nature of this field since such a perspective stimulates the formation of research networks (both intra- and inter-institutional) that collectively address complex environmental phenomena and avoid redundancies (CAPES, 2019; Sampaio and Phillip Junior, 2021).

The current study demonstrates the value of semantic analysis tools for guiding strategic actions and formulating public policy. Such tools provide extensive resources to identify significant patterns and attributes, which in turn, support meaningful studies (Pereira et al., 2011; Fariña Garcia et al., 2021; Yang et al., 2022). When combined with spatial analysis techniques, these resources can significantly aid in comprehending the intricate relationships inherent to multidisciplinary fields such as Environmental Sciences and evaluating their impact on society (Nobrega et al., 2018).

The objective of this article is to develop a model for analyzing scientific and technological output from GP, a process that involves the utilization of interaction network resources to identify patterns and connections among various elements. The scope of the study includes 112 GP classified by CAPES as being within the field of Environmental Sciences. Data was extracted from the Sucupira Platform. Since the research began mid-term within the 2017–2020 quadrennium, a comprehensive dataset from the preceding quadrennium (2013–2016) was employed.

The study resulted in fresh insights that revealed the need for metrics to assess the impact of the GP on society in order to identify which areas are a priority for introducing new programs, allocating human talent and financial resources to vulnerable sectors, fomenting new research networks and collaboration among GP, and providing meaningful social feedback on theses and dissertations that frequently involve communities and organizations (both public and private) as subjects or objects of study.

Literature Review

Within the context of the National Postgraduate System (SNPG), the field of Environmental Sciences (CiAmb) is located within the larger domain of multidisciplinary studies, which comprises eight other fields. It is part of the College of Exact Sciences, Technology, and Multidisciplinary Studies (Exact Sciences, Earth Sciences, and Engineering), the Colleges of Life Sciences (Agriculture, Biology, and Health) and Humanities (Humanities, Applied Social Sciences, and Linguistics, Letters, and Arts) (Sampaio and Phillip Junior, 2021).

As of 2023, a total of 141 GP were included in CiAmb, disseminated across all Brazilian states. These incorporate 101 academic master's programs, 33 professional master's programs, 47 academic doctoral programs, two professional doctoral programs, and three collaborative network programs (Graduate Program in Development and Environment [PRODEMA], Stricto Sensu Master's Degree in Management and Regulation of Water Resources [PROFÁGUA], and National Network for Teaching Environmental Sciences Master's Program [PROFCIAMB]) (Sampaio and Phillip Junior, 2021; CAPES, 2022). Most of these programs are distributed across the states, a situation that transformed the field of Environmental Sciences into a laboratory suitable for experimenting with methodologies to assess the impact of GP on society while bridging the gap between education and development (CAPES, 2019).

Data from CAPES (2017b) revealed that during the 2013–2016 quadrennium (the period studied in this work), Environmental Sciences GP were present in all regions of the country (Southeast: 31%; South: 21%; Northeast: 20%; Central-West: 26%; North: 12%). A total of 2,252 permanent faculty members (20.8% of whom were research or technological development and innovative extension grant recipients) and 4,709 graduate students were involved in these programs (Figure 1) (CAPES, 2017b).

A qualitative and semi-quantitative analysis of common themes in the programs was conducted during the initial preparations for the quadrennium assessment involving program coordinators. The analysis considered terms extracted from research line descriptors, which were then used to categorize the GP along four thematic axes: 1. "Technology, Modeling, Geoprocessing" (77%); 2. "Planning, Management, Environmental Public Policies" (53%); 3. "Development, Sustainability, and Environment" (46%); and 4. "Natural Resource Utilization" (37%) (CAPES, 2016). The recognition of these thematic axes, in conjunction with the analysis of other indicators, aids in the understanding of trends and identification of thematic gaps (such as ecological economics, environmental education, environmental law, and appropriate technologies) that can be integrated into existing program research lines or proposed for new courses (CAPES, 2016).

Since its inception, CiAmb has undergone various evaluation cycles within CAPES, experiencing changes that resulted from the maturation of assessment criteria (Bilotta et al., 2022). As the evaluation of PG conducted by CAPES analyzes scientific production by both faculty and students as evaluation dimensions, the fostering of initiatives that optimize such production is of vital importance. Such initiatives should be empowered by and conditioned to a comprehensive planning and understanding of what is generated within the programs. Reflecting on the evolution and history of CiAmb, as well as the effects of modifications to GP assessment cycles, poses a substantial challenge, but the use of semantic networks can aid in interpreting the data.

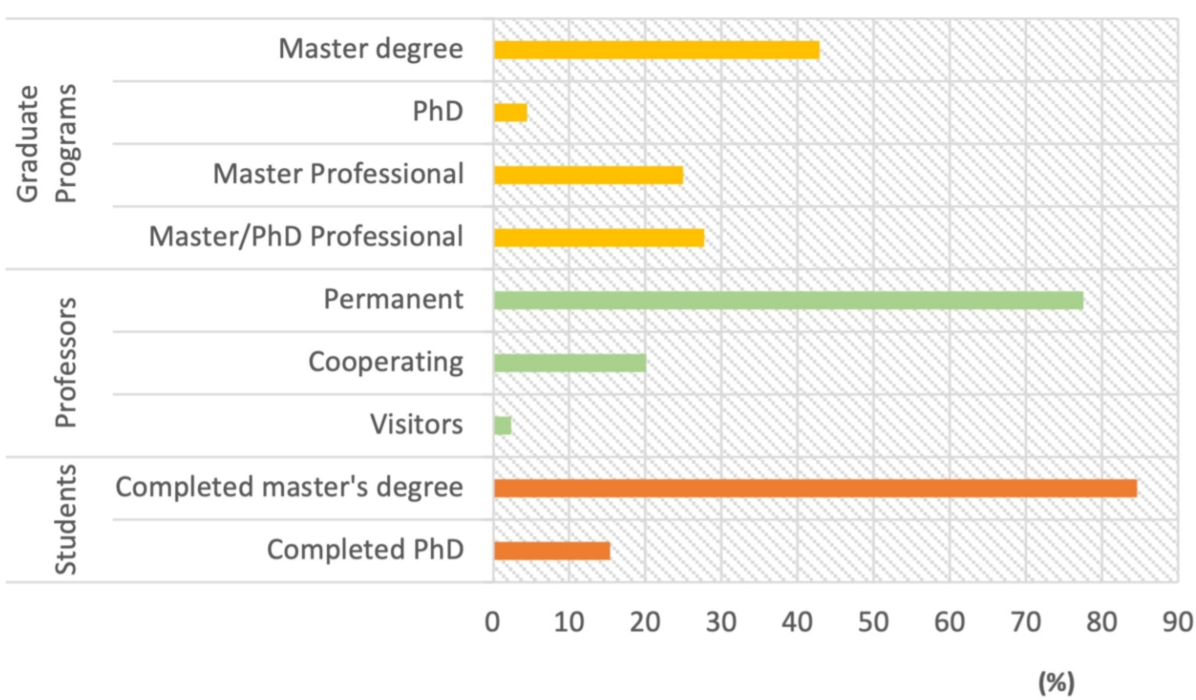

Figure 1 – Overview of Environmental Sciences graduate programs during the 2013–2016 quadrennium.
Source: CAPES (2017)

The term "semantic network" refers to the graphical systematization of information (data, concepts, ideas, events, etc.), linked by means of linguistic symbols (words), with the goal of representing connections (van Eck et al., 2006, 2010; van Eck and Waltman, 2007, 2011; Pereira et al., 2022; Kirby, 2023). Thus, a semantic network is a graphical representation model in which vertices (nodes) represent concepts (words or terms) and edges depict the relationships that exist between these concepts (Lehmann, 1992; van Eck and Waltman, 2007, 2010, 2011). Graphs employ binary logic between vertices and edges, utilizing elements that primarily describe two categories of information: 1. the "concept" and 2. the "conceptual relation" (Santos Junior et al., 2014).

Hence, semantic networks aim to visually express identified relationships among objects of inquiry (position, importance, significance) (Santos Junior et al., 2014; Pereira et al., 2022) and ascertain the intensity of existing connections by observing the clustering coefficient — the number of links to neighboring vertices (Garg and Kumar, 2018; Türker and Sulak, 2018; Shimada et al., 2019; Quispe et al., 2021). A higher coefficient indicates more significant clustering and stronger connections (Rosa, 2016).

This tool has found applications across a diverse range of fields (education, chemistry, biology, physics, psychology, engineering, mathematics, and computer science) (Santos Junior et al., 2014; Pereira et al., 2022). It models the interaction of domain objects in systems characterized by high complexity and interrelationships across time and space. Networks facilitate the construction of nonlinear, open, decentralized, dynamic, and non-hierarchical connection structures (Caldeira, 2005). Some authors (Rosa, 2016; Garg and Kumar, 2018; Türker and Sulak, 2018; Shimada et al., 2019; Pereira et al., 2022; Quispe et al., 2021) classify semantic network relationships into four main types: 1. "is-a" (connects a subcategory with a supercategory); 2. "instance-of" (connects a field with a supercategory); 3. "part-of" (connects a part to a whole); and 4. "is-capable-of" (connects a category with a corresponding attribute). Supercategories represent broader concepts, whereas subcategories encompass more specific concepts.

Relationships within a semantic network can be categorized under the headings hierarchy, inclusion, equivalence, or opposition. A straight line signifies a symmetric connection in the representation of the network, while an arrow denotes an asymmetric connection (van Eck and Waltman, 2007, 2010, 2011). Labels describe network attributes, namely, how the information of one edge relates to other edges (van Eck and Waltman, 2007, 2010, 2011; Rosa, 2016).

A specific type of semantic network, known as a co-occurrence term network, functions as a connection indicator; it is commonly employed to identify emerging topics, explore related research areas, and deepen the understanding of knowledge structures within a specific field (Pereira et al., 2022). These networks map the interconnection between terms based on their co-occurrence frequency within a given text corpus. In co-occurrence term networks, nodes represent terms, and edges indicate the joint occurrence of terms in documents. By analyzing these networks, it is possible to identify term association patterns and obtain valuable insights into the dynamics and interrelationships of concepts within a specific knowledge domain (Van Eck et al., 2006, 2010; van Eck and Waltman, 2010).

Co-occurrence term networks are crucial in identifying term clusters that exhibit high frequencies of common occurrences across a set of documents. Co-occurrence term frequencies are used in the elaboration of these networks, capturing proximity and association between terms based on their joint occurrence within a specific context. An analysis of these networks allows for visualization and comprehension of term interconnections, as well as the identification of significant co-occurrence patterns, thus revealing implicit and insightful relationships among studied concepts (van Eck et al., 2006, 2010; van Eck and Waltman, 2010).

Metodology

This study is analytical and descriptive, employing data mining techniques applied to a set of qualitative secondary data (textual expressions) in order to determine semantic relationships in text. Consistent with the nature of the present study, the data mining method is useful for identifying patterns, establishing significant relationships, and predicting behaviors within extensive information sets (Avelar et al., 2017). The research stages, primarily supported by data mining and semantic text relationships (clusters and word clouds), are outlined in Figure 2 and elaborated upon below.

The first stage consisted of building the database with terms and words extracted from the defined search strategy. The second stage (Figure 2) involved data processing, beginning with a preprocessing substage that edited terms and words within GP outputs in order to prepare the database for subsequent analyses. The following criteria were applied: 1. exclusion of articles (both definite and indefinite) and pronouns; 2. standardization of terms and words (uppercase, lowercase); 3. removal of characters such as accents, periods, quotation marks, exclamation marks, commas, umlauts, parentheses, brackets, and other punctuation-related characters, including special characters and derivatives; 4. exclusion of Roman and Arabic numerals and/or any nominal derivations; 5. removal of blank spaces; 6. indentation (line spacing) of production associated with each GP; and 7. exclusion of derivative forms of the same term in distinct verb tenses within the same sentence.

Subsequently, two .txt files were generated in a textual format, allowing for the observation of morphological, syntactical, semantic, and discursive aspects of the outputs. In the first file, all GP information was aggregated by entity, creating a semantic network of words and terms describing the main themes addressed in their outputs. In the second file, a structure was established in which the general production of each GP was associated with an identifier (institution of origin).

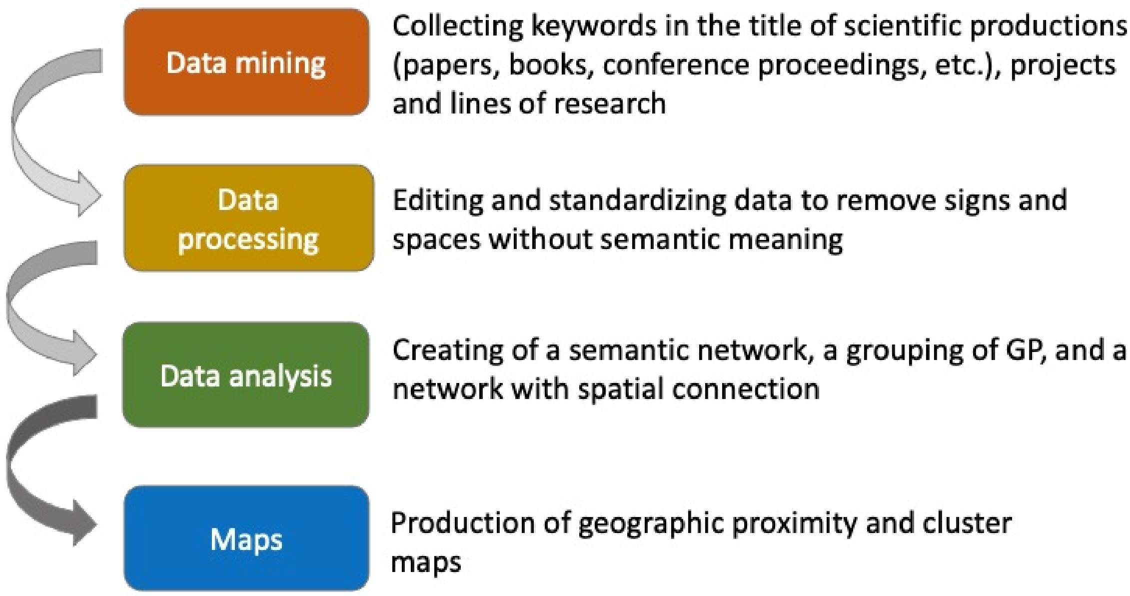

Figure 2 – Research stages.

The third stage (Figure 2) involved the construction of networks derived from combining all CiAmb outputs from the 2013–2016 quadrennium. This resulted in the creation of two networks: one with nodes and edges (nodes representing terms and/or words, and edges denoting connections between nodes), forming sets with similar and distinct colors in order to emphasize the occurrence of shared terms or thematic groups of similar and distinct production; and another network derived from the same dataset and analysis, in which data was visualized differently, using a density map of themes associated with chromatic patterns. In this type of density map (also known as a heat map), warm colors symbolize more significant themes, whereas cold colors represent the opposite extreme. Chromatic patterns are expressed in terms of a scale that emphasizes the importance of each word or theme in network construction.

The VOSviewer tool (van Eck and Waltman, 2009) was utilized to visualize bibliometric networks. A color pattern was applied in regions with cluster formation to accentuate the presence of jointly used words or thematic groups of similar and distinct production. The Ward's method (Valli, 2012; Tokuda et al., 2022) was used to create content clusters based on similarity and term sharing associated with the search results on the Sucupira Platform database.

A dissimilarity matrix was also developed, featuring a function that calculated the amount of dissimilarity between contents, taking into account word frequency and presence. This substage was conducted to avoid bias in the absence and frequency matrix calculation.

The fourth and final stage (Figure 2) involved the generation of products related to GP: cluster maps and geographical distribution maps.

A dendrogram was used for cluster production, highlighting group formation among GP. Clusters were created based on the similarity and sharing of words and terms from each GP. After analysis, quantitative metrics were extracted in order to generate graphs and tables related to the refined production. A function was developed based on the presence and frequency of words and terms to calculate the amount of dissimilarities to avoid bias in calculating the matrix due to absences and frequencies. The Ward's clustering method was employed using the R program (R Core Team) with the RStudio GUI (RStudio Team) and the "hclust" function from the base package (R Core Team, 2013). Aiming to evaluate the number of clusters and thereby generate the results, an optimization function was established based on parsimony calculation, in order to test the significance of the number of clusters (Cai et al., 2023). More specifically, a recursive function was developed to test different cluster numbers, returning the most significant outcome.

A map was prepared for spatial visualization of GP production, using the results table generated from cluster analysis to create an adjacency matrix based on cluster formation.

The geographical distribution network was elaborated using the Gephi program version 0.92 (Bastian et al., 2009) and exported in vector format. Subsequently, a map was generated to visualize the spatial network associated with production from Environmental Sciences GP, using the QGis program (version 3.26). GP from the same Higher Education Institution (HEI) were grouped under a single network node to simplify and refine the result visualization on the geographical distribution map.

Results and Discussion

this section presents the results and discussion regarding the patterns identified using the semantic network map, density map, cluster analysis, and geographical distribution network. The results of the semantic network for scientific production and research projects from the GP from 2013 to 2016 are shown in Figure 3. This figure presents the relationships between the sentences extracted in the data mining stage. It may be seen that the constructed network highlights the strong interdisciplinary nature that characterizes the field of environmental sciences, as identified by integrating clusters of high-frequency terms. Different colors indicate the formation of distinct clusters of sentences, that is, the nucleation of themes (words or terms) in the network. Similar colors demonstrate the proximity of sentences that revolve around one or more central themes.

The vertex (node) size in the semantic network is proportional to the number of times a specific sentence appeared in the data mining stage as compared to other sentences. Therefore, the larger the vertex, the more a given sentence occurred in scientific production and research projects involving Environmental Sciences GP during the period from 2013 to 2016. The edges, in turn, indicate the presence of sharing between sentences, i.e., existing interactions between nodes. These interactions are represented by the clustering coefficient, which indicates the number of connections between neighboring vertices. The higher the coefficient, the greater the clustering and, consequently, the stronger the connection between descriptors (Rosa, 2016).

Clusters associated with significant themes can be observed in Figure 3; these include themes such as “environmental vulnerability” (identified by nodes in dark blue), “analysis of land use and land cover dynamics (LULC)” (nodes in red), “use of new technologies” (yellow nodes), “biodiversity” and its relationship with “human health” (green nodes), “Amazon” (orange nodes), which includes not only the biome but also “chemical analysis” (in purple), “biomonitoring” and “genetics” associated with “aquatic ecosystems” and “biochemical analysis of essential oils” (salmon-colored nodes).

While the themes mentioned above form a core of relevant topics and approaches, the less-represented peripheral clusters indicate the inclusion of new themes, areas of knowledge, and/or development. This can be observed in the connection of less relevant nodes with more relevant nodes or clusters, such as the blue cluster associated with “production systems” and “biofuels” (Figure 3). These emerging themes need validation and should be subjects of analysis for future maturation regarding emerging topics in literature or intellectual production.

Figure 3 – Semantic network applied to terms extracted in the data mining stage.

The density map of themes in scientific production and research projects related to Environmental Sciences GP from 2013 to 2016 is shown in Figure 4. The brighter the yellow color, the higher the frequency of sentences identified through data mining. In other words, the most recurrent terms in the density map (quantitative indicator) are those most researched by Environmental Sciences GP faculty, resulting in a larger volume of scientific production recorded on the Sucupira Platform.

The following terms stood out in the density map: water, climate change, environmental vulnerability, Amazon, aquatic ecosystems, genetics, essential oils, biodiversity, Cerrado (savanna), new technologies, LULC, environmental impact, and land degradation. Therefore, these are the most recurrent themes among Environmental Sciences GP in Brazil. Identifying these predominant themes makes it possible to associate them with United Nations Sustainable Development Goals numbers 6 (Clean Water and Sanitation), 13 (Climate Action), 14 (Life Below Water), and 15 (Life on Land) of the 2030 Agenda, as well as to the thematic axes of "Development, Sustainability, and Environment" and "Use of Natural Resources" in the predominant lines of research in Environmental Sciences GP.

The density map also shows the intensity of relationships among different terms researched by faculty and identified in the data mining stage. A denser clustering indicates closer proximity between themes; in other words, terms that occur closer on the map appear together more often in the same research project or publication. For instance, sentences such as "environmental vulnerability," "climate change," "risk analysis," and "public health" demonstrate significant interaction among themselves, as may be seen by the fact that they cluster together. The clustering feature allows for the extraction of information that is valuable for interdisciplinary research since it identifies the most relevant facets surrounding the analysis of a specific subject. The results of this analysis are supported by the study conducted by Schmitt et al. (2022), who demonstrated the connection between interdisciplinarity and Environmental Sciences, in the context of preventing future pandemics.

These results are significant since they provide an understanding of how lines of research in the field of Environmental Sciences GP address the vast diversity of topics stemming from the relationship between the environment, economy, society, institutions, and their regional specificities. Moreover, by identifying and analyzing clusters and their hierarchy of importance, it is possible to gain a detailed view of already established trends and emerging themes, which appear as marginal clusters.

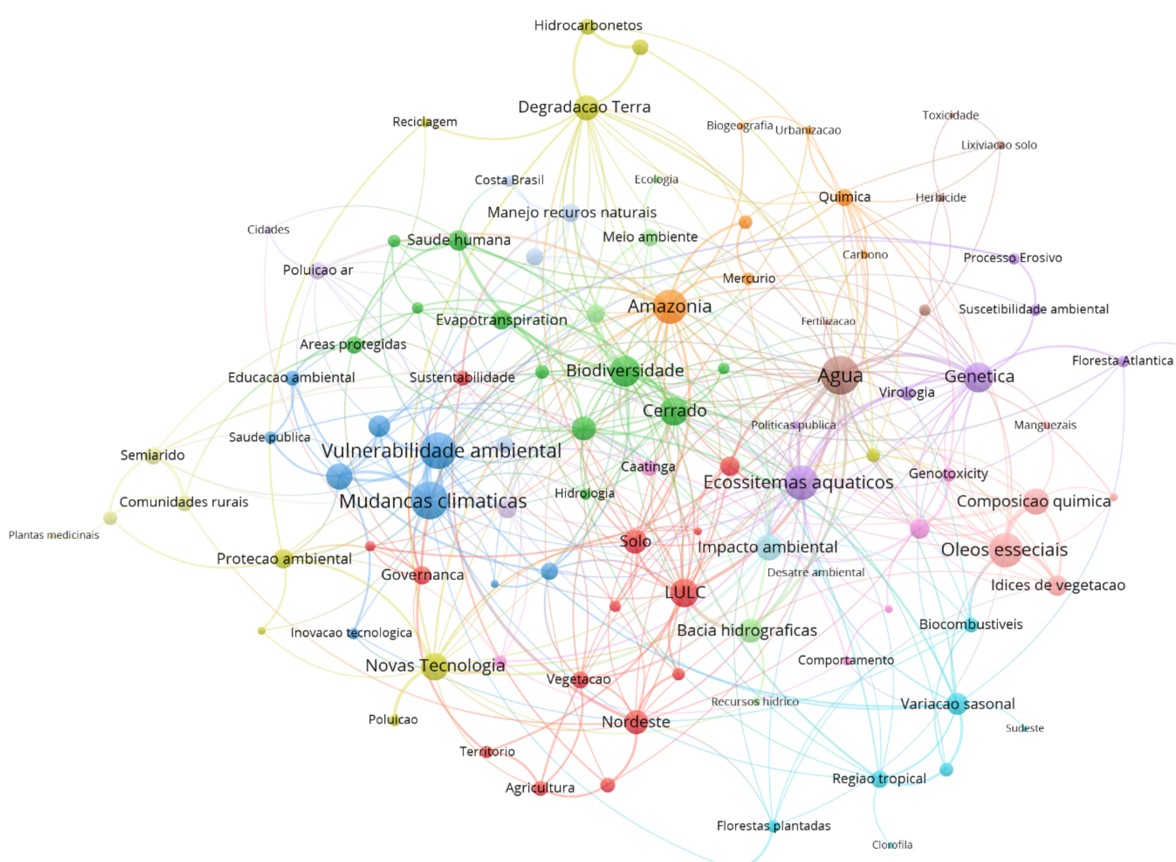

Figure 4 – Density map of terms extracted in the data mining stage.

According to Fariña García et al. (2021), the results obtained from semantic network analysis align with qualitative analyses conducted by experts. To validate this assertion, the results of this method were compared with those of a previous qualitative analysis carried out by CiAmb Coordination (CAPES, 2016) that grouped the GP into thematic axes. Although the qualitative analysis conducted by experts within CiAmb indicated that the Environmental Sciences GP is predominantly concentrated on the "Technology, Modeling, and Geoprocessing" thematic axis (CAPES, 2016), these terms were not evident in either the density map or the semantic network. This can be explained by the fact that geotechnologies are generally used as data analysis tools and, therefore, are only described in the methodology of articles or research projects, being infrequently mentioned in academic work titles or keywords.

The density map and the semantic network can be very useful tools for funding agencies for identifying socio-environmental themes of great regional or national interest that require induction, as well as for guiding research support grant announcements and providing the basis for the development of indicators to assess the impact of academic work on society and plan strategies for the Environmental Sciences field in the coming years.

In the final block of analyses, network diagrams were produced to study the relationships between the GP and the analyzed descriptors.

The dendrogram resulting from clustering Environmental Sciences GP by institution is represented in Figure 5. Each identified cluster is highlighted in the same color, indicating substantial similarity among the themes researched by faculty from those institutions. Eighty-five higher education institutions offering Environmental Sciences GP were identified by 2016, forming 30 content similarity clusters.

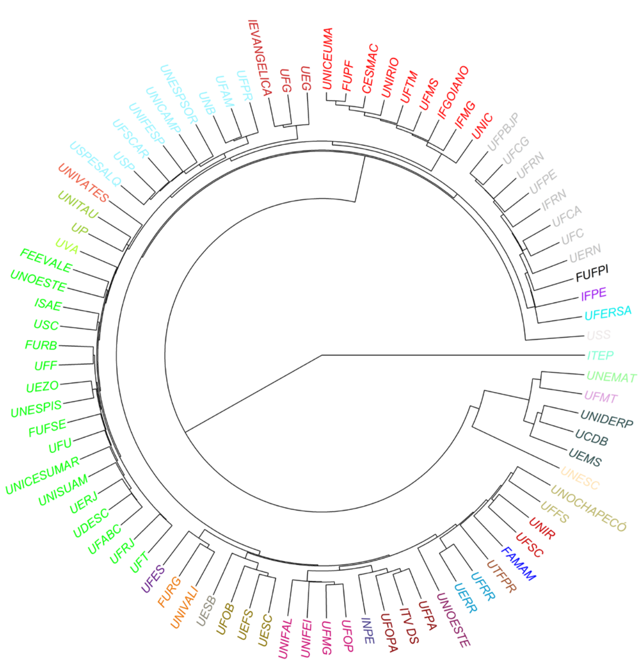

Figure 5 – Dendrogram of Environmental Sciences graduate programs production.

The dendrogram (or tree diagram) is a graphical representation of data grouped by level of similarity. The connection between two or more fields denotes similarity in content (words, terms, sentences). For example, the group represented by the color pink exhibits a high degree of similarity among the themes researched by the faculty. However, the Federal University of Ouro Preto (UFOP) and the Federal University of Minas Gerais (UFMG) form a subgroup, identified by the branching that unites these two institutions. Likewise, the Federal University of Alfenas (UNIFAL) and the Federal University of Itajubá (UNIFEI) form another subgroup. This means that within the same cluster (color designation), faculty research may show varying degrees of content similarity.

As shown in Figure 5, the interpretation of the cluster analysis for GP based on the similarity of words and terms found in their production revealed seven well-defined groups. These clusters form entities that represent distinct groups. This result shows that such GP share and/or interact and exchange knowledge regarding common or complementary themes and approaches. These clusters can also be explained by both the age of the programs and their geographic isolation. One example is the formation of a group of institutions in the Southeastern and Southern regions of the country and the more complete isolation among GP in the Northeastern and Northern regions. This pattern may also indicate greater exchange within each region of the country and an inverse relationship between inter-regional sharing. Regarding the spatialization of GP based on the network associated with word, term, and theme similarities, the most striking isolation of the Northern and Northeastern regions is evident compared to the South, Southeast, and even the Central-West regions.

The dendrogram is helpful, allowing one to visualize the potential for interaction among Environmental Sciences GP focusing on mutually interesting topics. The establishment of collaboration among GP researching similar themes should certainly be strongly encouraged since it facilitates faculty and student exchange (joint courses, projects, and scientific production), resource sharing among institutions (laboratory equipment, research materials, access to licensed software, etc.), and the acquisition of data and information. Thus, the results presented in Figure 5 show that inter-institutional approximations exist among Environmental Sciences GP. However, these need to be strengthened and new partnerships encouraged, as seen in the ongoing initiative to analyze and promote the formation of thematic clusters (USP, 2022). The observed institutional interaction in the field of environmental sciences may be supported by the fact that GP spontaneously seek alternatives to meet their requirements for specialists in areas that exceed the capacity of their faculty and physical infrastructure, due to the much greater complexity of interdisciplinary studies in analyzing human and biotic–abiotic interactions. In this regard, the periodic holding of national events can stimulate new institutional partnerships, as seen in the case of the UrbanSus I Academic Meeting on the Impact of Environmental Sciences on the UN 2030 Agenda, held in March 2021 by the Institute of Advanced Studies (IEA) of the University of São Paulo (USP) (Sampaio and Phillip Junior, 2021), as well as the II Academic Meeting, held in June 2022, and the III Meeting scheduled for November 2023.

Figure 6 presents the main clusters formed involving the distribution of scientific production and research projects by Environmental Sciences GP faculty for the 2013–2016 quadrennium.

The analysis shown in Figure 6 highlights the formation of five large clusters around several institutions that offer postgraduate courses in environmental sciences, with the predominance of three GP groups representing the South, Southeast, and Northeast regions having the highest number of programs in the field. These results reinforce the perspective of spontaneous proximity between themes researched by faculty from different GP and regions of the country. Inter-institutional partnerships promoted within the clusters, such as the ongoing formation of 14 thematic clusters around teaching, research, and extension activities that reconcile academic excellence and fulfillment of Sustainable Development Goals (SDGs) included in the 2030 Agenda (USP, 2022), have various types of positive impact on society. They enhance the exchange of experiences, the sharing of human talents (faculty) and infrastructure (laboratories, equipment, etc.), and student exchange (the development of new skills), taking into account the natural, economic, social, and cultural differences found across the Brazilian territory.

Figure 7 presents the GP geographical distribution map obtained from the terms identified in data mining (sentences or words). Each "node" (point) represents all Environmental Sciences GP in a particular state. At the same time, the "edges" (lines) indicate the interactions among various GP, observing similarities among the themes researched by their faculty. This map, in addition to showing the geographical distribution of GP, also presents the themes and interactions among the programs, regardless of their geographic proximity.

The Southeast, South, and Central-West regions concentrate on forming GP networks, while the Northern region exhibits a more complete isolation of GP. For example, the PRODEMA network stands out in the Northeast, demonstrating strong interaction among GP in this region around shared research themes.

There is evidence that the formation of collaborative networks benefits programs in isolated regions, such as the GP in the state of Rondônia which is linked to a network based in the Southern region and compensates for its geographic isolation.

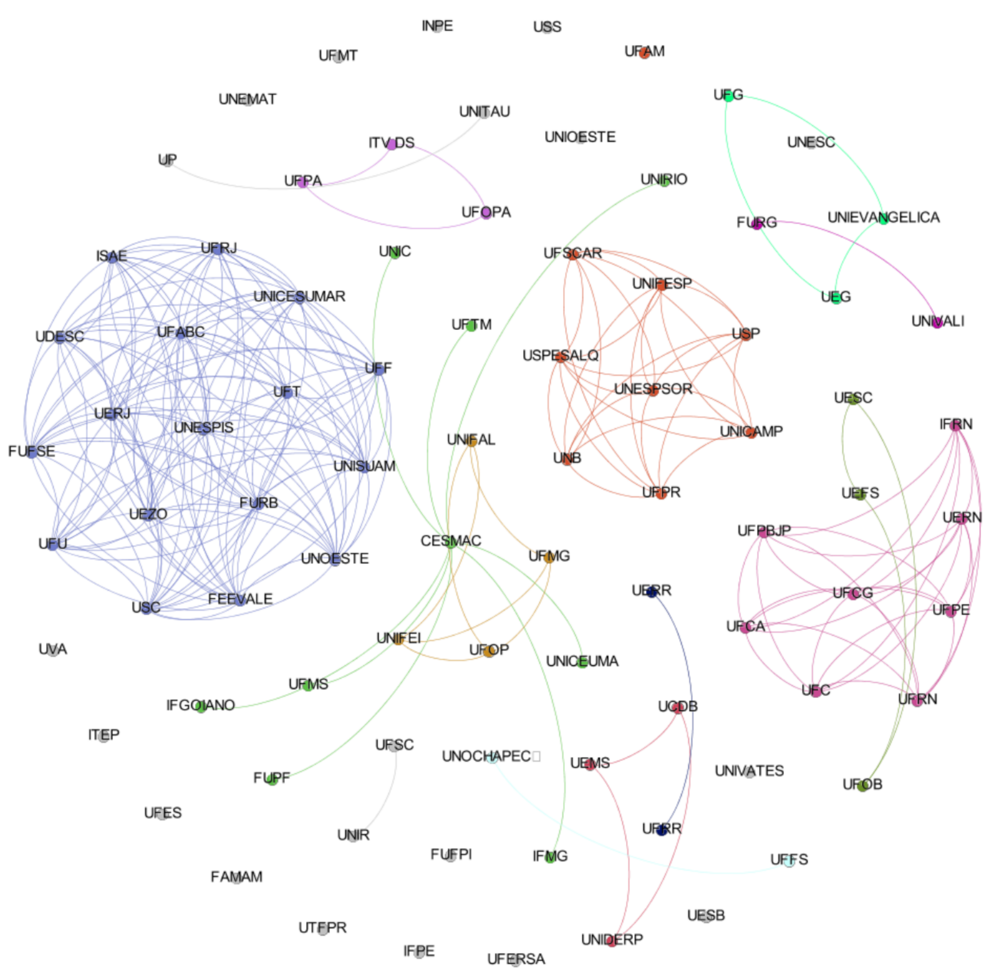

Figure 6 – Clusters of Environmental Sciences graduate programs.

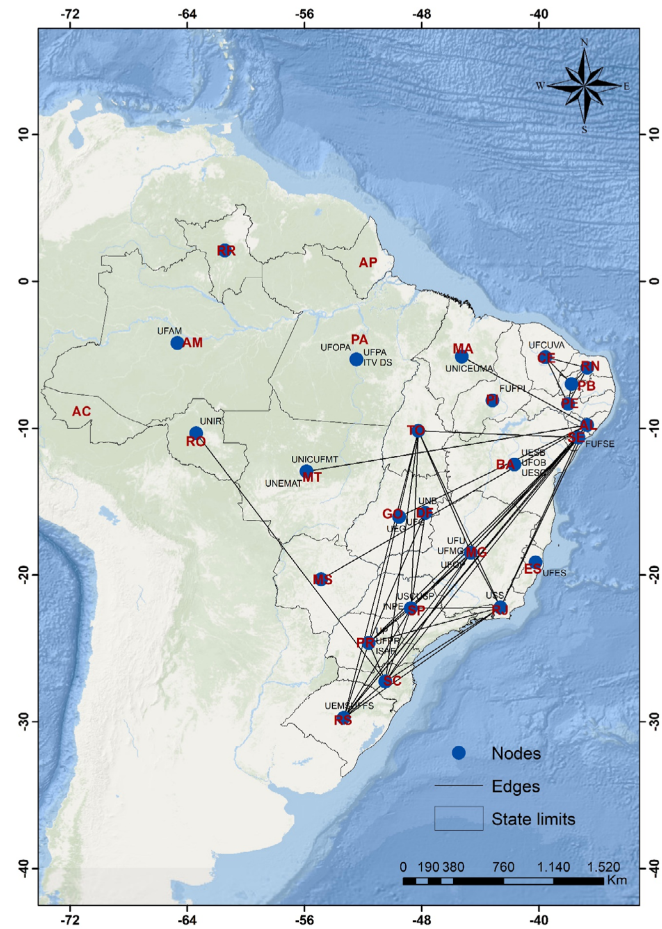

Figure 7 – Geographical distribution map of themes researched by Environmental Sciences graduate programs.

Conclusions

The results presented regarding the prospecting of impact indicators in society, based on density maps, semantic networks, cluster analysis, and geographical distribution networks in the analysis of scientific production in Environmental Sciences GP, indicate that postgraduate education assessment programs can make use of such tools to correlate, or rather measure the development of higher education.

As demonstrated in the above discussions, and in view of the results of this study, the proposed analysis model based on semantic networks is promising and should be expanded to include a broader range of data intersections, such as international projects and partnerships, information regarding faculty members (such as education and field of expertise), and students (background, education, integration in society), and types of products generated by theses and dissertations, among others.

The results obtained in this study can aid the Environmental Sciences field in fostering new research networks among postgraduate programs, collaboration with programs located in more vulnerable regions, allocation of human talents and financial resources toward preferred areas (those showing greater socio-economic and environmental fragility), and determination of territories that should be considered a priority for the introduction of new courses. Moreover, the study of semantic networks in the territorial analysis can provide the appropriate social feedback from undergraduate, master's, or doctoral (both academic and professional) theses, whose object or subject of research is often communities and both public and private organizations.

Finally, some of the weaknesses of this methodology may be associated with semantic networks when a large dataset is used in the territorial analysis, such as the excessive amounts of time required for data preprocessing, the lack of a formal structure that hinders the representation of more complex domains (being unsuitable for objects with many relationships), and the inheritance of attributes that visually represents the only way to make inferences about the representation. These limitations should be studied in greater detail in order to minimize their impact on the tool's applicability.

The proposed analysis model, based on semantic network analyses, has proven to be a tool with much potential for diagnosing the impact of research, extension courses, and education of postgraduate programs on society. This model preserves the dominance and physical, social, and political integrity of the studied area, making it possible to identify key concepts and relationships between different areas of study in the Environmental Sciences, as well as the connections that exist between researchers, institutions, and countries involved in this type of scientific production. The authors recommend that this analysis method be improved by incorporating new analyses and extracting indexers to produce more robust analysis elements based on more sophisticated statistical treatments. The authors plan to continue their research, incorporating data from more than one analysis period in order to obtain comparison and identification parameters of the effects of policies and actions that have been introduced. In this regard, it is also recommended that an automation system be developed and implemented to reduce errors and streamline the process as a whole.

Other network initiatives in environmental sciences are currently underway, such as 1. the inclusion of new evaluation indicators for GP Territorial Excellence for the 2021–2024 quadrennium; 2. the construction of evaluation indicators for the impact of GP on society through research involving participatory action; 3. the systematic and participatory formation of thematic clusters associated with teaching, research, and extension activities related to GP and their connection with the various dimensions of SDGs; and 4. formulation of an eco-formative proposal based on the systemic analysis of GP-PROFCIAMB.

Contribution of authors:

fRANCA-ROCHA, W.J.S.: conceptualization, investigation, project administration, methodology, writing – original draft, writing – review & editing. VASCONCELOS, R. N.: data curation, formal analysis. CHAVES, J. M.: conceptualization, investigation, project administration, writing – original draft, writing – review & editing. BILOTTA, P.: project administration, writing – original draft, writing – review & editing. GRIMM, I. J.: original draft, Writing. RIBEIRO, S. M. C.: original draft, writing. NOBREGA, R. A. A.: original draft, writing. SOBRAL, M. C. M.: writing – review & editing. PHILIPP JUNIOR, A.: supervision. SAMPAIO, C. A. C.: – supervision.

References

Avelar, C.F.P.; Rocha, T.A.H.; Cruz, F.J.S., 2017. Mineração de dados: uma revisão da literatura em administração. Revista Vianna Sapiens, v. 8, (2), 30-54. https://doi.org/10.31994/rvs.v8i2.232

Bastian, M.; Heymann, S.; Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. Thris Internacional AAAI Conference on Weblogs and Social Media, v. 3, (1), 361-362. https://doi.org/10.1609/icwsm.v3i1.13937

Bilotta, P.; Carbone, A.S.; Corbari, S.D.; Duleba, W.; Chaves, J.M.; Kniess, C.; Grimm, I.J.; Pregnolato, L.A., 2022. Environmental sciences and SDGs: Brazilian and international cases. SDGs in the Americas and Caribbean region. Springer, Cham, International Publishing, Switzerland. https://doi.org/10.1007/978-3-031-16017-2_27

Brasil, 2010. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Plano Nacional de Pós-Graduação: 2011–2020. Capes, Brasília.

Cabral, T.L.O.; Silva, F.C.; Pacheco, A.S.V.; Melo, P.A., 2020. A Capes e suas Sete Décadas: trajetória da pós-graduação stricto sensu no Brasil. Revista Brasileira de Pós-graduação-RBPG, v. 16, (36), 2358-2332. https://doi.org/10.21713/rbpg.v16i36.1680

Cai, J.; Hao, J.; Yang, H.; Zhao, X.; Yang, Y. 2023. A review on semi-supervised clustering. Information Sciences, v. 632, 164-200. https://doi.org/10.1016/j.ins.2023.02.088

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2016. Documento de área: Ciências Ambientais (Accessed July 30, 2019) at:. www.gov.br/Capes/pt-br/centrais-de-conteudo/49_CAMB_docarea_2016_publ2.pdf

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2017a. CAPES - Sucupira: coleta de dados, docentes de pós-graduação stricto sensu no Brasil 2013 a 2016 (Accessed July 30, 2019) at:. https://dadosabertos.capes.gov.br/dataset/35eab2f8-5a64-4619-b3f1-63a2e6690cfa/resource/b871db42-a86e-43d3-bf21-2ea2e85d0acd/download/metadados_docentes_2013a2016.pdf

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2017b. Relatório de avaliação – Ciências Ambientais (Accessed July 30, 2019) at:. www.gov.br/Capes/pt-br/centrais-de-conteudo/20122017CIENCIASAMBIENTAISquadrienal.pdf

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2019. Documento de área – Área 49: Ciências Ambientais. Diretoria de Avaliação, 24 p. (Accessed July 12, 2023) at:. https://www.gov.br/capes/pt-br/centrais-de-conteudo/49_CAMB_docarea_2016_publ2.pdf/view

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2022. Relatório de avaliação – Ciências Ambientais (Accessed March 15, 2023) at:. www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19122022_49.CAMB_Quadrienal_Relatorio_Final.pdf

Fariña García, M.C.; De Nicolás, V.L.; Yagüe Blanco, J.L.; Labrador Fernández, J. 2021. Semantic network analysis of sustainable development goals to quantitatively measure their interactions. Environmental Development, v. 37, 100589. https://doi.org/10.1016/j.envdev.2020.100589

Florit, L.F.; Sampaio, C.A.C. Philippi Junior, A., 2019. O desafio da ética socioambiental. In: Florit, L.F.; Sampaio, C.A.C.; Philippi Junior, A., Ética socioambiental. Barueri, Manole, pp .3-16.

Garg, M.; Kumar, M., 2018. The structure of word co-occurrence network for microblogs. Physica A: Statistical Mechanics and Its Applications, v. 512, 698-720. https://doi.org/10.1016/j.physa.2018.08.002

Kirby, A., 2023. Exploratory bibliometrics: using VOSviewer as a preliminary research tool. Publications, v. 11, (1), 10. https://doi.org/10.3390/publications11010010

Lehmann, F., 1992. Semantic network. Computers & Mathematics with Applications, v. 23, (2-5), 1-50. https://doi.org/10.1016/0898-1221(92)90135-5

Nobrega, R.A.A.; Ribeiro, S.M.C.; Costa, E.L.; Macedo, D.R.; Bilotta, P.; Grimm, I.J.; Sampaio, C.A.C.; Schypula, A.; Chaves, J.M.; Rocha, W.J.S.F.; Vasconcelos, R.N., 2018. Destaque territorial: proposta de modelagem socioeconômica e ambiental para avaliar a inserção social nos Programas de Pós-Graduação em Ciências Ambientais. Brazilian Journal of Environmental Sciences, v. 49, 34-50. https://doi.org/10.5327/Z2176-947820180372

Max-Neef, M.; Elizalde, A.; Hopenhayn, M., 2012. Desenvolvimento á escala humana. EdiFurb, Blumenau.

Pereira, H.B.B.; Fadigas, I.S.; Senna, V.; Moret, M.A., 2011. Semantic networks based on titles of scientific papers. Physica A: Statistical Mechanics and its Applications, v. 390, (6), 1192-1197. https://doi.org/10.1016/j.physa.2010.12.001

Pereira, H.B.B.; Grilo, M.; Fadigas, I.S.; Souza Junior, C.T.; Cunha, M.V.; Barreto, R.S.F.D.; Andrade, J.C.; Henrique, T., 2022. Systematic review of the "semantic network" definitions. Expert Systems with Applications, v. 210, 118455. https://doi.org/10.1016/j.eswa.2022.118455

Quispe, L.V C.; Tohalino, J.A.V.; Amancio, D.R., 2021. Using virtual edges to improve the discriminability of co-occurrence text networks. Physica A: Statistical Mechanics and Its Applications, v. 562, 125344. https://doi.org/10.1016/j.physa.2020.125344

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Ambientais (Accessed March 26, 2023) at:. http://www.R-project.org/

Rosa, M.G., 2016. Modelo empírico para analisar a robustez de redes semânticas. Doctoral Thesis, Programa Multidisciplinar e Multi-Institucional em Difusão do Conhecimento da UFBA, LNCC, UEFS, IFBA, SENAI-CIMATEC, FACED, IHAC. Feira de Santana, Bahia.

Sachs, I. 2004. Desenvolvimento: includente, sustentável, sustentado. Garamond, Rio de Janeiro.

SANTOS JUNIOR, R.P.; RODRIGUES, A.A.A.O.; LOPES, C.R.; DÉJARDIN, I.P.; PEIXOTO, J.L.B.; CUNHA, M.V., 2014. Análise de diferentes conceitos de educação por meio das redes semânticas. DataGramaZero: Revista de Informação, v.15, (3), A03.

Sampaio, C.A.C.; Phillip Junior, A., 2021. Impacto das Ciências Ambientais na Agenda 2030 da ONU. São Paulo: Instituto de Estudos Avançados, Universidade de São Paulo, v. 1. https://doi.org/10.11606/9786587773186

Schmitt, J.L.; Rocha, C.Y.M.; Galvincio, J.D.; Almeida, J.S., 2022. Interdisciplinaridade em ciências ambientais: monitoramento ambiental na prevenção de futuras pandemias. Historia Ambiental Latinoamericana Y Caribeña (HALAC) Revista De La Solcha, v.1 2, (1]0, 322-352. https://doi.org/10.32991/2237-2717.2022v12i1.p322-352

Shimada, Y.; Tatara, M.; Fujiwara, K.; Ikeguchi, T., 2019. Formation mechanisms of local structures in language networks. EPL (Europhysics Letters), v. 127, (5), 56003. https://doi.org/10.1209/0295-5075/127/56003

Tokuda, E.K.; Comin, C.H.; Costa, L.F., 2022. Revisiting agglomerative clustering. Physica A: Statistical Mechanics and its Applications, v. 585, 126433. https://doi.org/10.1016/j.physa.2021.126433

Türker, İ.; Sulak, E.E., 2018. A multilayer network analysis of hashtags in twitter via co-occurrence and semantic links. International Journal of Modern Physics B, v. 32, (4), 1850029. https://doi.org/10.1142/S0217979218500297

Universidade de São Paulo (USP), 2022. Centro de Síntese USP Cidades Globais. UrbanSus: II Encontro Acadêmico Impacto das Ciências Ambientais da Agenda 2020 da ONU. USP/IEA, São Paulo.

Valli, M., 2012. Análise de custer. Augusto Guzzo Revista Acadêmica, (4), 77-87. https://doi.org/10.22287/ag.v0i4.107

van Eck, N.J.; Waltman, L., 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. Journal of the American Society for Information Science and Technology, v. 60, (8), 1635-1651. https://doi.org/10.1002/asi.21075

van Eck, N.J.; Waltman, L., 2007. Bibliometric mapping of the computational intelligence field. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, v. 15, (05), 625-645. https://doi.org/10.1142/S0218488507004911

van Eck, N. J.; Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, v. 84, (2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N.J.; Waltman, L., 2011. Text mining and visualization using VOSviewer. ISSI Newsletter, v. 7, (3), 50-54. (paper, preprint, supplementary material)

van Eck, N.J.; Waltman, L.; Noyons, E.C.M.; Buter, R.K., 2010. Automatic term identification for bibliometric mapping. Scientometrics, v. 82, (3), 581-596. https://doi.org/10.1007/s11192-010-0173-0

van Eck, N.J.; Waltman, L.; van den Berg, J.; Kaymak, U., 2006. Visualizing the computational intelligence field. IEEE Computational Intelligence Magazine, v. 1, (4), 6-10. https://doi.org/10.1109/CI-M.2006.248043

Yang, H.; Liu, Q.; Zhang, J.; Ding, X.; Chen, C.; Wang, L., 2022. Community detection in semantic networks: a multi-view approach. Entropy, v. 24, 1141. https://doi.org/10.3390/e24081141