National Mapping of the Brazilian Academic Cybersecurity Community

Kauane Cordeiro[1], **Reinaldo Cézar de Morais Gomes**[1]

[1] National Research and Development Network (RNP)
Rio de Janeiro, RJ 22290-906, Brazil

{kauane.cordeiro,reinaldo.gomes}@rnp.br

Abstract. *This paper presents a methodological workflow for mapping the Brazilian academic cybersecurity community. The approach integrates multiple national databases through a pipeline that encompasses data collection, filtering, curation, automated processing, and the application of data analysis and visualization tools — including word-cloud exploration and a geospatial interactive map. The resulting visualizations reveal regional asymmetries, research clusters, and gaps in academic engagement, enabling an integrated view of the Brazilian cybersecurity ecosystem. The developed framework is reproducible and provides a structured overview that supports strategic planning, community strengthening, and the advancement of collaborative initiatives in cybersecurity.*

1. Introduction

Cybersecurity has become a key strategic field for Brazil's scientific, technological, and social development. Despite its growing relevance, significant gaps remain in the structured understanding of the Brazilian academic community in this domain. The absence of a consolidated perspective hinders national coordination, impedes the construction of institutional partnerships, and diminishes the international visibility of Brazilian research.

To address this scenario, the Brazilian National Research and Education Network (RNP), in collaboration with the Special Committee on Information and Computational Systems Security (CE-Seg) of the Brazilian Computer Society (SBC), launched an integrated initiative to map the Brazilian academic cybersecurity community.

The main objective of this initiative is to identify, organize, and disseminate information about researchers, research groups, and graduate programs, as well as to map the academic output in the field. This mapping effort reveals areas of concentration, regional gaps, scientific trends, and potential axes of collaboration among universities, research centers, institutes, and other strategic actors.

This effort is aligned with RNP's role as Brazil's focal point for the academic engagement axis (*think tank*) of the BRICS Working Group on the Security of ICT Use. A first version of the results was presented during the BRICS event held in April 2025, in Brasília. Since then, the initiative has undergone updates and refinements, expanding the quality, scope, and accuracy of the mapping.

2. Related works

Mappings are effective instruments for characterizing academic communities and ecosystems, enabling the identification of key actors, relationships, research themes, and insti-

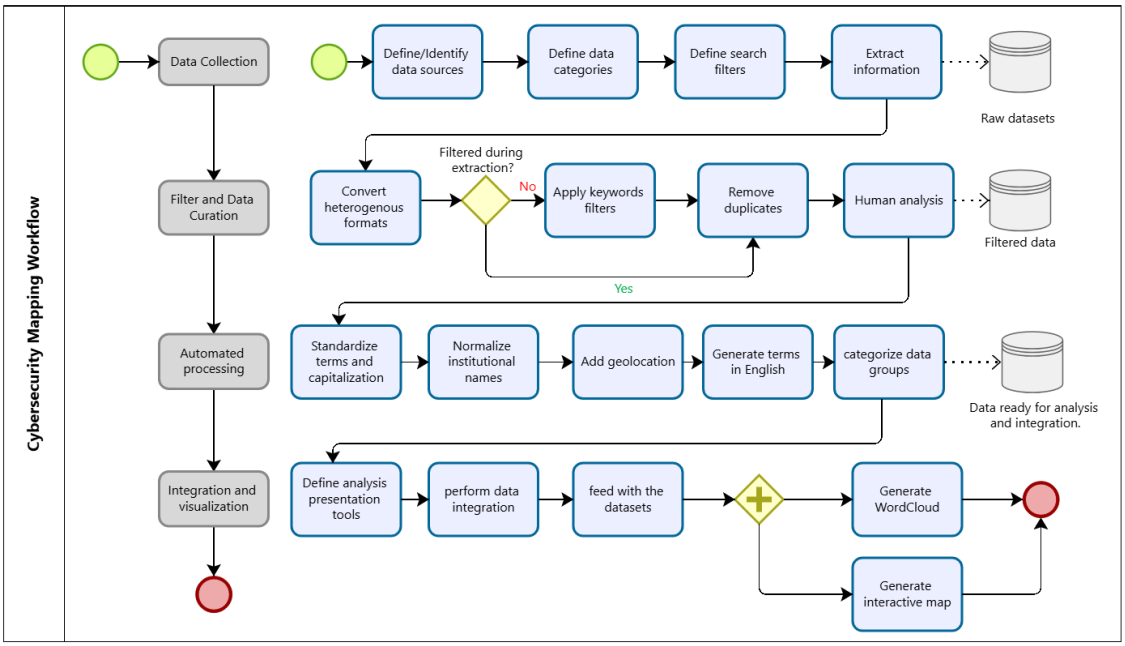

Figura 1. Cybersecurity Mapping Workflow

tutional structures, while supporting cooperation and strategic planning.

In the European context, the European Cybersecurity Mapping 2025 [European Champions Alliance 2025] provides a large-scale overview of the cybersecurity ecosystem, organizing information on startups, organizations, and innovation trends. Similarly, the Cybersecurity Competence Community Map [Directorate-General for Communications Networks, Content and Technology 2025] and the ENISA National Cyber Security Strategies Interactive Map [European Union Agency for Cybersecurity 2025] employ georeferenced visualizations and structured datasets to represent research capacities, governance models, and strategic priorities across Europe. Together, these initiatives demonstrate the relevance of visual and data-driven approaches for analyzing complex ecosystems.

Academic studies on scientific community analysis usually employ quantitative techniques and clustering models to investigate collaboration and research dynamics. Initiatives such as GRENMap [GNA-G 2024] highlight the potential of georeferenced platforms to represent institutional connectivity, while sector-specific mappings, such as the CESeg initiative [CESeg 2020], reinforce the role of open cataloging in increasing visibility, collaboration, and expertise identification. More recently, Mapping Scientific Communities at Scale [Barbier and Jeangirard 2025] proposed a scalable methodology based on enriched bibliographic metadata for detecting and prioritizing scientific relationships.

Despite these contributions, there is still no integrated, public, interactive, and continuously updated mapping focused on the Brazilian academic cybersecurity community. Existing international initiatives do not address the Brazilian context, while national studies remain fragmented and disconnected from researchers, graduate programs, research groups, and scientific results. This work addresses this gap by consolidating multiple data sources into a unified visualization of the national cybersecurity research ecosystem.

3. Methodology

The methodology for constructing the National Mapping of the Brazilian Academic Cybersecurity Community was structured into four main and complementary dimensions, as presented in Figure 1: data collection, curation and filtering, automated processing and integration and visualization. Each of these steps is detailed below.

3.1. Data Collection

Multiple reliable and nationally recognized sources were used to ensure the completeness and accuracy of the data:

- **Researchers:** The initial list of researchers was composed of the members of the Program Committee (PC) of the 2025 Brazilian Symposium on Cybersecurity (SBSeg). These participants completed a form in which they indicated their areas of expertise, affiliated institutions, and cybersecurity-related research lines.
- **Research Groups:** Research groups were identified using the CNPq Research Group Directory, by applying filters based on keywords related to cybersecurity;
- **Scientific Articles:** The analyzed scientific production was obtained from the Digital Library of the Brazilian Computer Society (SOL base [Brazilian Computer Society (SBC) 2025]), SBC's official repository that compiles articles published in all events and journals organized by the society. We considered different thematic tracks from all editions of SBSeg, thereby providing both a historical and a contemporary perspective on the evolution of cybersecurity research in the most important national conference in this domain.
- **Graduate Programs, Theses, and Dissertations:** Data from CAPES were used to identify graduate programs with research lines related to cybersecurity, as well as theses and dissertations classified by relevant keywords or correlated areas.

3.2. Data Curation and Filtering

The data curation stage posed significant challenges due to the need to integrate information from multiple sources, each with varying formats, structures, and levels of detail". This scenario required meticulous processing to identify correspondences, eliminate duplicates, and prepare the datasets for the subsequent stages.

Python scripts were developed to convert multiple data sources and formats into a unified format and to automate the corresponding filtering and cleaning procedures. The process began with a systematic filtering of related terms, encompassing variants in Portuguese and English, both with and without diacritics. Subsequently, human evaluation focused on the qualitative assessment and manual validation, ensuring thematic relevance and alignment with the cybersecurity context. The keywords used in this process were: "cibersegurança", "ciberseguranca", "segurança cibernética", "segurança cibernetica", "segurança da informação", "seguranca da informacao", "cybersecurity", "cyber security", "information security".

These terms were applied both during data collection, when repositories supported keyword-based filtering, and after collection in sources without native filtering.

3.3. Automated Processing

The curated datasets underwent an additional processing stage focused on cleaning, normalization, and standardization, transforming heterogeneous data sources into consistent structures suitable for analysis and visualization.

This stage included capitalization and terminological normalization, harmonization of institutional references — for example, unifying "UFPE," "Univ. Fed. de Pernambuco," and "Universidade Federal de Pernambuco" into a single representation — as well as data enrichment procedures, such as geolocation attribution and English translation of terms to support interoperability and comparative analyses.

All procedures were automated through Python scripts to ensure reproducible processing across datasets. Specialized libraries were employed for tabular manipulation (**pandas** [Pandas Development Team 2025]), geolocation and geocoding (**geopy** [Geopy Developers 2025] and **geocoder** [Geocoder Contributors 2025]), automatic translation (**googletrans** [Googletrans Contributors 2025]), spreadsheet and CSV processing (**openpyxl** [OpenPyXL Authors 2025] and **csv** [Python Software Foundation 2025a]), and textual normalization using regular expressions (**re** [Python Software Foundation 2025b]).

The resulting dataset provided a consistent basis for geographic and analytical visualization of the Brazilian cybersecurity research community.

3.4. Integration and Visualization

The final stage of the methodology focused on defining the presentation strategies for the results. Two complementary visualizations were selected — the Word Cloud and the Interactive Map — as each highlights distinct aspects of the academic cybersecurity community. All processing stages were designed to produce a flexible dataset capable of supporting different visualization formats and subsequent analyses.

Although both visualizations were based on the same curated datasets, each required additional preparation procedures due to their analytical purposes and technical characteristics. The corresponding steps are detailed below.

- **WordCloud**: was generated through a Python script developed with specialized libraries, including **WordCloud and STOPWORDS**[Mueller 2025] for cloud generation and stopword management, **re**[Python Software Foundation 2025b] for text cleaning and standardization **Pillow (PIL)**[Pillow Developers 2025] for image processing, and **Matplotlib**[Matplotlib Development Team 2025] for visualization rendering and adjustment.
  The script was designed to process different fields from the constructed datasets, automatically generating visual representations of the most representative terms in the textual corpus. Additionally, a human-assisted refinement stage was conducted to remove low-semantic-load terms, improving the visualization and emphasizing the main recurring themes and research trends in national cybersecurity studies.
- **Interactive Map**: Based on the previously processed data, this stage focused on defining the most relevant relationships among the different categories of information, with the goal of enabling their interactive visualization through the ChameleonMap platform developed by RNP.

Data insertion and relationship creation were automated with Python scripts that performed bulk loading and fed the tool's database. This procedure preserved structural consistency and enabled an interconnected representation of the academic community under analysis.

4. Results

The steps carried out resulted in the definition of a systematic, standardized, and replicable process for the treatment, organization, and integration of data, enabling distinct visualizations aligned with each analytical objective and with the tools employed.

In the context of constructing the word clouds, separate visualizations were generated, each linked to a specific dataset, with a focus on the researcher dataset and the scientific production dataset. Using these two sources allows the identification of complementary characteristics of the Brazilian academic community and enables an examination of how research output in the cybersecurity field has evolved.

4.1. Word Clouds

The researchers' dataset was used to generate a word cloud of declared topics of interest, aiming to identify the most recurrent research themes within the national academic cybersecurity community.

For the scientific production dataset, word clouds were generated from the complete set of articles published in the Brazilian Symposium on Cybersecurity (SBSeg), considering both annual subsets and the entire publication period. This approach enables the identification of thematic shifts over time, highlighting topics that remained central to the field as well as those that gained or lost prominence.

To illustrate these changes, the word clouds for SBSeg 2018 (Figure 2-a in Appendix A) and SBSeg 2024 (Figure 2-b in Appendix A) are presented. Terms such as "Networks," "Data," and "Detection" remained prominent, while "Blockchain" and "IoT" lost visibility, and "Attack," "Vulnerability," and "AI" became more recurrent.

Overall, these visualizations highlight the thematic domains that concentrate the largest volume of scientific production and declared research interest within the Brazilian cybersecurity academic community.

4.2. Interactive Map

The interactive map[1] is organized into two tabs with explorable sections. The first tab consolidates the integrated mapping data and supports individual, regional, and longitudinal analyses. Each section functions as a complementary layer connected to the processed datasets and relationships identified throughout the workflow.

- **Brazilian Cybersecurity Researchers / Researchers by State**: presents the geographic distribution of researchers and information such as areas of interest, Lattes CV, affiliated institutions, research groups and graduate programs.
- **Topics of Interest**: Based on information provided by the researchers, the topics-of-interest visualization (Figure 3 in Appendix A) connects themes to associated

[1] https://cybersecmap.rnp.br/

researchers and institutions. Unlike the synthetic overview provided by the word cloud, the interactive version enables thematic filtering and regional exploration of the mapped data. Selecting a topic allows the identification of associated researchers and institutions

- **Cybersecurity Research Groups (CNPq) / Research Groups by State**: presents research groups registered in the CNPq directory, including their institutions and leadership. The "By State" version enables regional filtering and comparative analyzes across states.
- **Graduate Programs (CAPES) / Graduate Programs by State**: consolidates the graduate programs identified during the dataset integration process, together with their associated academic outputs.
- **Theses and Dissertations by State**: presents the distribution of theses and dissertations across states, highlighting regions with higher academic output and supporting regional comparisons and identification of spatial patterns in scientific production.
- **Articles by Year**: shows the evolution of cybersecurity-related scientific production over the years, enabling a detailed temporal analysis of the data used in constructing the scientific-production word cloud. This representation broadens the analytical scope by highlighting the annual concentration of publications and their regional distribution. Together, these perspectives provide insight into both historical trends and the dynamics of growth and dispersion of cybersecurity research.

The second tab brings together national initiatives relevant to the cybersecurity ecosystem, including research centers, working groups, and strategic projects.

5. Conclusion

The national mapping of the Brazilian academic cybersecurity community presented in this work represents a strategic step toward consolidating an understanding of the field by integrating data on researchers, research groups, graduate programs, and scientific production. The proposed methodology — combining automated collection, data curation, and interactive visualization — proved effective and replicable for consolidating information from multiple sources into a single interactive platform.

The results highlight the evolution of national cybersecurity research, evidenced by the growth in scientific production and diversification of research lines. The approach enables the observation of geographic and institutional distributions, as well as the identification of thematic trends, regional gaps, and emerging areas of study.

From a strategic perspective, the mapping supports coordination among institutions, researchers, and public policymakers, providing a basis for actions in research, development, and human resource training. The integration of this information into an open and accessible environment also increases the visibility of Brazilian scientific production and strengthens the country's international presence.

Future work includes periodic database updates, the incorporation of new data sources, and the development of analytical indicators on the evolution of cybersecurity research in Brazil.

Referências

Barbier, V. and Jeangirard, E. (2025). Mapping scientific communities at scale. https://www.researchgate.net/publication/388180021_Mapping_scientific_communities_at_scale. Accessed: 2025-01-10.

Brazilian Computer Society (SBC) (2025). Sbc open library: Index of publications. Available at: https://sol.sbc.org.br/index.php/indice. Accessed: 2025-12-12.

CESeg (2020). Research groups in security. https://www.researchgate.net/publication/388180021_Mapping_scientific_communities_at_scale. Accessed: 2025-01-10.

Directorate-General for Communications Networks, Content and Technology (2025). Cybersecurity competence community map. https://european-champions.org/european-cybersecurity-mapping-2025. Accessed: 2025-01-10.

European Champions Alliance (2025). European cybersecurity mapping 2025. https://european-champions.org/european-cybersecurity-mapping-2025. Accessed: 2025-01-10.

European Union Agency for Cybersecurity (2025). National cyber security strategies interactive map. https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map. Accessed: 2025-01-10.

Geocoder Contributors (2025). geocoder: Simplified geocoding library. https://geocoder.readthedocs.io/. Accessed: 2025-01-10.

Geopy Developers (2025). geopy: Geocoding library for python. https://geopy.readthedocs.io/. Accessed: 2025-01-10.

GNA-G (2024). Grenmap. https://www.gna-g.net/join-working-group/gren-mapping/. Accessed: 2025-01-10.

Googletrans Contributors (2025). googletrans: Free and unlimited google translate api for python. https://py-googletrans.readthedocs.io/. Accessed: 2025-01-10.

Matplotlib Development Team (2025). Matplotlib: Visualization with python. https://matplotlib.org/stable/. Accessed: 2025-01-10.

Mueller, A. (2025). word_cloud: A word cloud generator for python. https://amueller.github.io/word_cloud/. Accessed: 2025-01-10.

OpenPyXL Authors (2025). openpyxl: Excel spreadsheet reader and writer for python. https://openpyxl.readthedocs.io/. Accessed: 2025-01-10.

Pandas Development Team (2025). pandas: Python data analysis library. https://pandas.pydata.org/. Accessed: 2025-01-10.

Pillow Developers (2025). Pillow (pil fork): Python imaging library. https://pillow.readthedocs.io/en/stable/. Accessed: 2025-01-10.

Python Software Foundation (2025a). csv: Csv file reading and writing. https://docs.python.org/3/library/csv.html. Accessed: 2025-01-10.

Python Software Foundation (2025b). re: Regular expression operations. `https://docs.python.org/3/library/re.html`. Accessed: 2025-01-10.

A. Appendix: Visualization Samples

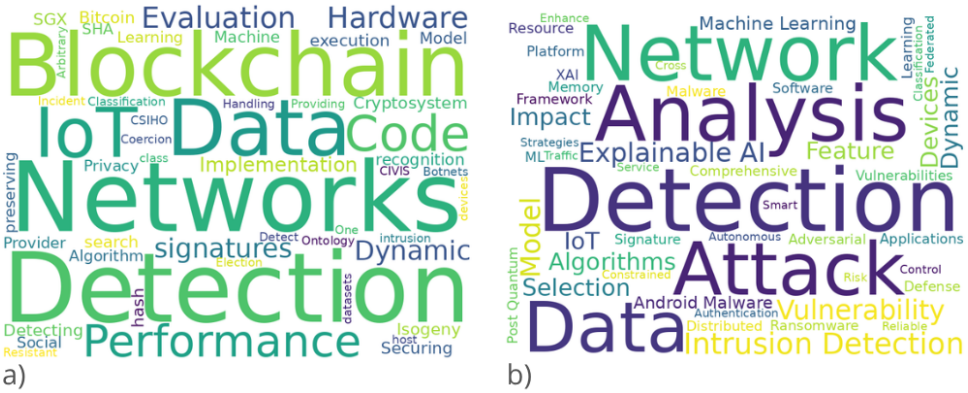

Figura 2. Recurring terms in the scientific production of SBSeg in 2018 (a) and 2024 (b)

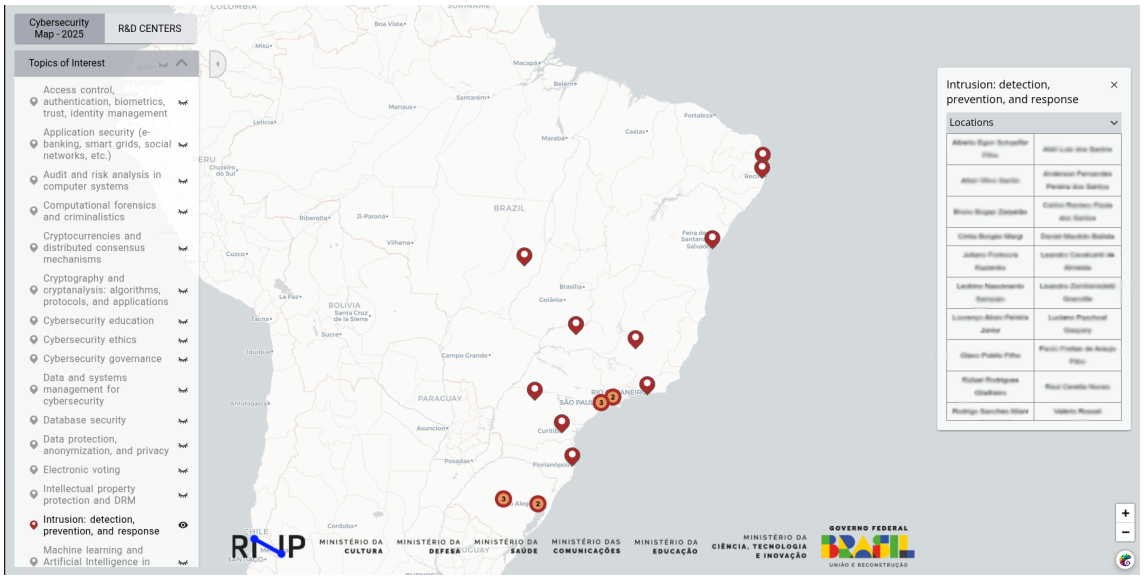

Figura 3. The section of researchers' topics of interest