Institutional Performance and Collaboration in Ageing Research: A Spatial Perspective

Cédric Roussel[1[0000-0002-7302-5431]], Amir Zahedidarehshoori[1], Klaus Böhm[1[0009−0008−8492−7650]] and Julian Stockemer[1[0009-0004-5976-39311]]

[1] i3mainz - Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences, 55128 Mainz, Germany
cedric.roussel@hs-mainz.de,
amir.zahedidarehshoori@students@hs-mainz.de,
{klaus.boehm, julian.stockemer}@hs-mainz.de

Abstract. Population ageing is one of the most significant demographic developments worldwide, stimulating rapid growth in scientific research across the biomedical, social and interdisciplinary fields. At the same time, scientific research has become more collaborative and distributed across international borders, resulting in complex institutional networks that influence research output and impact. Despite bibliometric approaches being widely used to evaluate scientific performance, most existing studies focus on publication counts, citation metrics or thematic trends, paying limited attention to the spatial organization of institutions and geographical collaboration patterns. This study addresses this gap by integrating bibliometric analysis, network modelling and geospatial visualization to investigate institutional performance and collaboration structures in ageing research. Using publications retrieved from PubMed over the last five years, a spatially enriched database containing authorship, affiliation, and collaboration information is constructed. Institutional performance is assessed through publication volume and a weighted authorship scoring metric that differentiates between leading and contributing roles in scientific publications. Geocoded affiliations are then used to create an interactive geovisual analytics dashboard representing institutions as spatially embedded nodes and collaborations as network connections. This framework allows to explore institutional hotspots, collaboration intensity and international linkages, offering new insights into the spatial dynamics of ageing research and demonstrating the added value of geospatial methods for bibliometric analysis and research evaluation.

Keywords: Bibliometric Analysis, Geovisualization, Ageing.

1 Introduction

As a central topic in numerous scientific disciplines, population ageing has become one of the most significant demographic trends of recent decades. Research on population ageing has grown substantially, with an 18.17% annual growth rate from 1985 to 2022

[1]. As research in this area grows, it is becoming increasingly important to understand how knowledge is produced, distributed and developed collaboratively. The institutional and spatial organization of research plays a crucial role in shaping scientific output and impact, beyond thematic content alone.

Alongside this thematic growth, scientific research has become more collaborative and international. Contemporary research is frequently carried out by multi-institutional, cross-border teams, resulting in complex networks of co-authorship and affiliation. Understanding how knowledge is organized and distributed within national research systems is therefore essential for assessing research capacity, performance and structural strengths.

Bibliometric methods are widely used to evaluate scientific performance. Common indicators for example are total publications and total citations or author specific metrics such as the h-index [2-4]. While these indicators provide important quantitative insights, they often treat publications as non-spatial entities and overlook the geographical distribution of research institutions. Furthermore, traditional bibliometric analyses often oversimplify collaboration, counting co-authorships without explicitly considering spatial relationships or distinguishing between institutional roles within publications. For example, Lai et al. [5] used keywords to visualize a network map, or a geographic map showing the worldwide scientific production. While these approaches lead to new insights, they do not demonstrate the performance of individual institutions or collaborations. Kanilmaz et al. [6] provide a more detailed analysis than a country-by-country breakdown, visualizing collaboration hubs as nodes and edges. Although they provide a more detailed version, their visualization is not interactive. When working with such complex big data, it is necessary to provide options for exploring the data and interacting with visualizations.

Spatial proximity, regional research clusters and cross-border linkages can influence the intensity, visibility and innovation potential of collaboration. However, analyses that explicitly consider spatial factors in relation to institutional performance in ageing research remain limited, particularly with respect to Germany as a major European research hub.

This study aims to address these gaps by combining bibliometric analysis, network modeling and interactive geospatial visualization to examine institutional performance and collaboration patterns in ageing research. Our focus is on recent publications from the last five years, and we investigate the spatial structure of research institutions active in this field. Specifically, we address the following analysis questions:

- Which are the top 25 research institutions in Germany specializing in ageing?
- Where are these institutions based?
- How well do these institutions perform?
- Who do they collaborate with?

According to Kirk [7], we focus especially on data transformation and data exploration. During transformation, we focus on "optimizing the condition, content and form of your data" [7, p. 27]. We provide a more differentiated view of institutional involvement, introducing a weighted authorship metric that distinguishes between leading and contributing roles in collaborative publications. This approach moves beyond equal

counting methods and enables a more nuanced assessment of institutional research performance. We then use interactive geovisualizations as data exploration to display the achieved results and to get a "deeper intimacy with the potential qualities and insights" [7]. This study contributes to the emerging field of spatial bibliometrics by integrating bibliometrics, network analysis and geovisual analytics. It also highlights the added value of geospatial methods for evaluating institutional research performance and scientific collaboration.

The structure of our study is as follows: In Section 2, we present the methodology for gathering the necessary data, processing it, and creating the planned geovisualization. Section 3 shows our results with the interactive geovisualization. Section 4 concludes the study and provides an outlook for future work.

2 Methodology

Our data for the study is based on PubMed, a large library of biomedical literature (https://pubmed.ncbi.nlm.nih.gov/, accessed 3 February 2026). This library contains millions of articles with metadata and is publicly available via an API. Our methodology to answer the stated research questions is structured in four steps:

Identify. Firstly, we gather all relevant literature from the PubMed database. The focus is to analyze the current situation and will not consider older relationships or the performance of institutions. For this, the Institute of Molecular Biology (https://www.imb.de/, accessed 10 February 2026) considered the last five years and provided the following keywords: ageing (aging for US spelling); age-related/age-associated disease; healthy ageing; longevity; lifespan; resilience; senescence; elderly; geriatric disease; older; older age; (super)centernarian; life expectancy; biological age; rejuvenation; regeneration; regenerative; epigenetic clock; immortality; anti-ageing; systemic approach; systems biology; systems medicine; mRNA therapy; translational; multimorbidity; interdisciplinary; interventions; longitudinal cohorts; ageing immune system; immunosenescence; inflammaging; prevention/preventive; stem cell exhaustion; immuno-ageing. To extract relevant research from PubMed, we analyzed different approaches in terms of feasibility and stability. Our analysis showed that the PubMed API is not feasible, as extracting large amounts of data results in server errors and incomplete data. For this reason, we are considering an alternative approach to the API. First, we use PubMed identifiers (PMIDs) to save all relevant IDs. We then compare these IDs with the results from the PubMed website. Finally, we download all articles that match exactly. This ensures that no articles are missed.

Extract. After gathering all the relevant literature from PubMed, we extract and store the features necessary for our analysis. The data tree structure in XML or JSON can differ for each piece of literature. We analyze each stored data tree and extract the following information: authors, affiliations, journal or conference, and year. We then store our relevant data in MongoDB, using different collections for authors and affiliations. We use the PMID as a unique index to ensure that there are no duplicates.

Encode. This step involves two key tasks. The first is geo-encoding the affiliations. Affiliations typically include a country, city and name. All of this information needs to be encoded and assigned to a new attribute in the database. To achieve this, we use the offline geographical database GeoNames (https://www.geonames.org/, accessed 12 February 2026), which provides details of millions of locations. An offline database is necessary as we are dealing with millions of affiliation records. Using an online database with an API would not be feasible due to the high volume of requests required. Additionally, the coordinates need to be encoded from the city in order to visualize the results geographically later on. For missing values, we manually search for geographical information. We also standardize city names based on similarity, as some may be written slightly differently depending on the language or publication style sheet. Using our geo-encoding approach, we receive coordinates for the city in which each institute is located, rather than exact coordinates. Encoding exact coordinates would require an online database and would probably result in server errors due to the number of requests.

After geo-encoding the affiliations, we check for duplicates. To do this, we apply vector embeddings. Additionally, we grouped the affiliations by city and applied a similarity method within each city. By using this procedure to merge duplicates, we significantly reduced our database to one million four hundred thousand articles, resulting in a standardized, consistent database with which to continue our analysis.

The second task involves implementing a scoring metric. While institutions could have the same number of publications, the order in which authors are listed on a paper can indicate differences in contribution. Typically, the first author receives the most credit for a publication. For this reason, we implement a flexible scoring metric. In our prototype implementation, institutions with the first author receive five points, while the other institutions receive one point. This metric is used to analyze the institutional performance. Different scoring systems can be evaluated and adapted later to achieve a more realistic approach. Our example using five and one points is prototypical.

Visualize. The final step involves geovisualization. We will use a collaboration graph in which institutions are the nodes and mutual publications are the edges. Due to geo-encoding, institutions in a city are assigned the same coordinate pair. We apply a small jittering approach to distinguish between multiple institutions in one city. The visualization will show institutions on a map, either according to the number of publications or the calculated score. The dashboard will also be made interactive to improve exploratory analysis of the results.

Additionally, we have decided to implement an AI agent in the form of a chatbot to enable further queries of the database. We use two local large language models for this purpose. The first model, Qwen2.5-14B, converts the user's question into a query that can be used with our database. This query retrieves the relevant data, which is then processed by another LLM. The second model, llama3.1-8b, then generates a response for the user. We show the final dashboard in the next Section 3.

3 Results

In this Section, we present our final results in the form of a geovisualization. As mentioned in our methodology, we use a graph visualization with nodes and edges, similar to a co-word network visualization such as in Donthu et al. [2]. The nodes represent institutions and the edges represent the relation between nodes [7], in our case the collaborations in publications. The thickness of the edges shows the number of collaborations. We also provide options to adjust the parameters of the geovisualization. Figure 1a shows the graph visualization for the top 25 institutes in ageing research, using our implemented score metric for authorship placement on publications.

In some cases, the total number of publications might be more relevant than our score metric. For this reason, we have included the option to change this parameter and alter the visualization according to the number of publications for each institution. This can be seen in Figure 1b. Compared to Figure 1a, some institutions appear where others disappear and the size of the node's changes. According to the score metric, the largest node was an institute in Berlin, whereas according to the total number of publications, the largest node is an institute in Heidelberg.

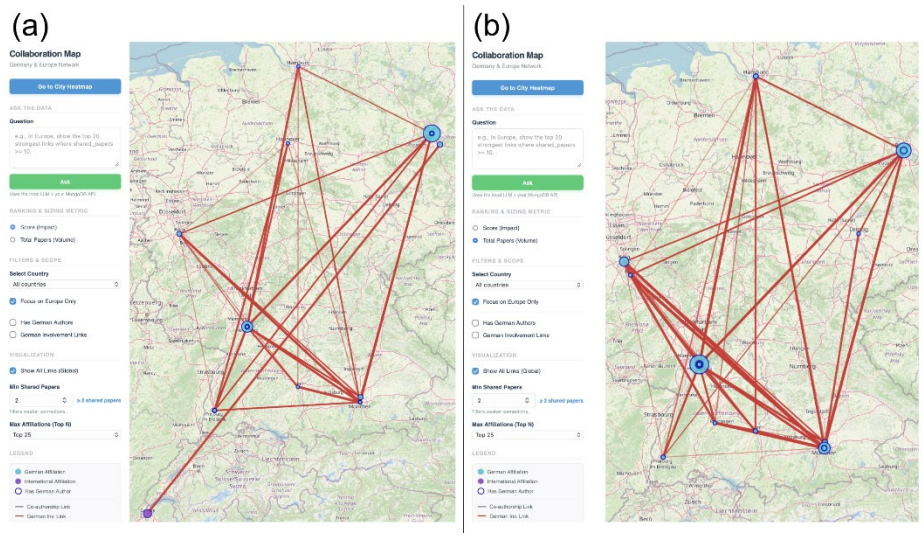

Fig. 1. Top 25 institutions in Germany by impact score metric (a) and by publication volume (b) in ageing research (Dashboard overview).

As well as the option to change from the score metric (impact) to the total number of publications (volume), we also provide further exploratory analysis options. Our dashboard allows to focus on the whole world, Europe only, or a specific country. When focusing on countries other than Germany, it is also possible to only show institutions that collaborated with a German institution. The collaborations, visualized by the edges, can also be filtered by increasing the number of shared publications. For example, it would be possible to only show collaborations between institutions that have published

at least n publications together. In Figure 1, this option is set to two. Furthermore, it is possible to focus on more than just the top 25 institutions, e.g. shown in Figure 2 with the top 100 institutions.

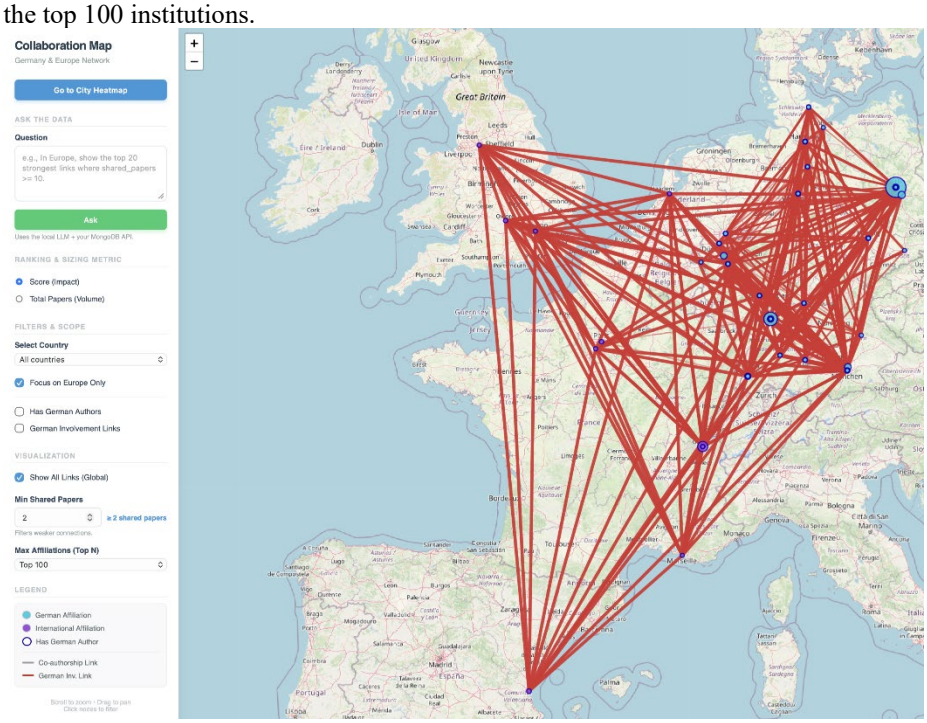

Fig. 2. Top 100 institutions in Europe by impact score metric in ageing research (Dashboard overview).

Our dashboard has been developed to be easily extendable for further analysis options. For example, we can provide heatmaps or charts such as the one in Figure 3, showing which countries collaborated the most with German institutions. In this case, we can see that the US has the highest number of collaborative publications. However, the US also has the most institutions. The ratio of publications to institutions is much higher in Switzerland, for example. Using our methodology to gather data, create a dataset and visualize it in a graph network allows us to successfully answer our stated analysis questions.

In addition to the visualization and the parameters on the left, we have a text box to enter queries for our AI agent. An example of a query could be: "List the top 10 countries by total_shared_papers_with_de". The output is then displayed in a popup window. Figure 4, shows the output for this query, which produces the same results as the bar chart in Figure 3 and demonstrates the accuracy of our agent.

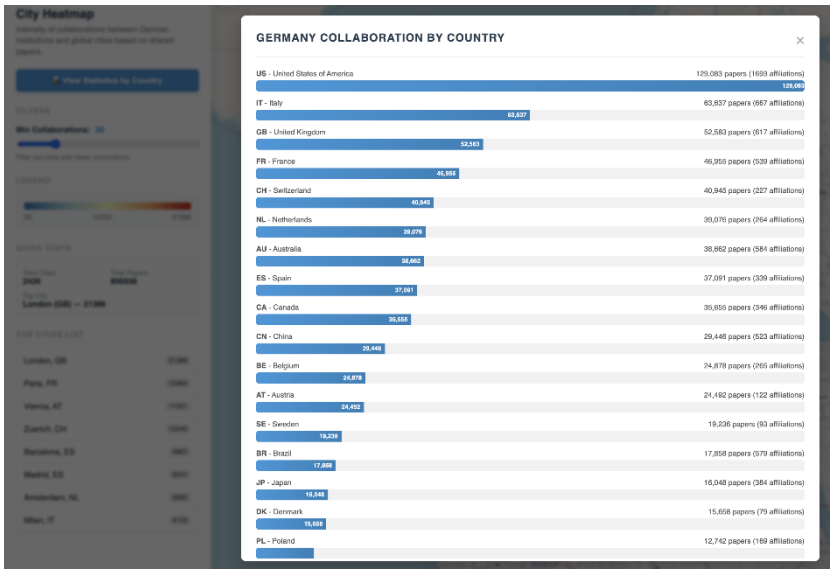

Fig. 3. Bar chart showing the total number of countries collaborating with Germany.

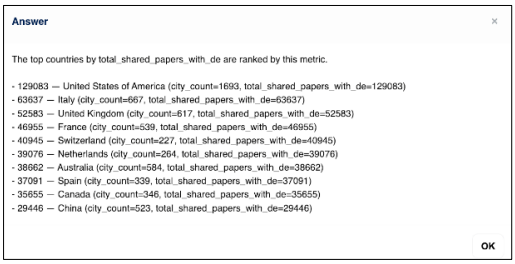

Fig. 4. Answer of the AI agent to the query: "List the top 10 countries by total_shared_papers_with_de".

4 Conclusion and Future Work

In this study, we analyzed the performance and collaborations of institutions in Germany in the field of ageing research. To do so, we considered published articles from the last five years. We gathered data from the publicly available database PubMed, extracting relevant information and features to encode for the final geovisualization implementation. For the prototypical implementation various technical challenges were addressed.

Our final results show that geovisualization can provide many new insights into institutional performance and collaboration. Not only can hot spots be identified, but also a collaboration network, which cannot be seen without a spatial perspective.

In the future, we want to improve our AI agent so that it can be used for chatting with stored messages. The AI agent will also be able to adjust the dashboard

visualization according to the user's query. We will then evaluate the dashboard through user testing to improve the design and identify potential new features. To achieve this, we consider usability, which is divided into three categories: efficiency, effectiveness and satisfaction. To evaluate this, we might use quantitative assessments through Likert-scale questions, as well as qualitative assessments through discussions with users. We will also enrich the database with additional data, such as the impact factor of journals or the number of citations for each paper. This will allow us to analyze the quality of institutions' output more effectively than if we only considered our implemented score metric.

Acknowledgments. The research for this paper is part of the AI-Lab at Mainz University of Applied Sciences, which is part of the project "Trading off Non-Functional Properties of Machine Learning" at the Johannes-Gutenberg-University Mainz. The authors further thank The Institute of Molecular Biology (IMB) for providing relevant keywords for the bibliometric analysis. This research was funded by the Carl Zeiss Foundation, grant number P2021-02-014.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mazza, R., Pace, R., & Paterno, A.: Themes and policies on population ageing: a bibliometric approach. RIEDS - Rivista Italiana di Economia, Demografia e Statistica - The Italian Journal of Economic, Demographic and Statistical Studies, 77(2), 33-43 (2023)
2. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., Lim, W.M.: How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296 (2021). doi: 10.1016/j.jbusres.2021.04.070
3. L., M.K., George, R.J., P.S., A.: Bibliometric analysis for medical research. Indian Journal of Psychological Medicine, 45(3), 277-282 (2023). doi: 10.1177/02537176221103617
4. Passas, I.: Bibliometric analysis: The main steps. Encyclopedia, 4(2), 1014–1025 (2024). doi: 10.3390/encyclopedia4020065
5. Lai, C.K., Mustafa, M.C., & Mahat, H.: Mapping the landscape of spatial literacy research: Bibliometric analysis. International Journal of Publication and Social Studies, 9, 12-27 (2024). doi: 10.55493/5050.v9i1.5095
6. Kanilmaz, U.N., Schneidergruber, T., Schmidt, S., Dehghan, R., Scholz, J.: Uncovering regional structures and collaboration hubs in Baden-Württemberg through clustering a company hyperlink network. AGILE: GIScience Series, 6, 29 (2025). doi: 10.5194/agile-giss-6-29-2025
7. Kirk, A.: Data visualisation: A handbook for data driven design - hardcover. SAGE Publications Ltd (2016)