ENS 491-492 – Graduation Project
Final Report

Project Title:

Science of Science Research for EU Research Landscape

Group Members:

Ayşegül Güler - 29440

Fikret Kayra Yılmaz - 29371

Supervisor(s): Onur Varol

Date: December 29, 2024

1. EXECUTIVE SUMMARY

This project aims to contribute to the growing field of Science of Science, which seeks to understand and optimize the processes of scientific discovery and collaboration. Focusing on the Turkish Academy of Sciences (TÜBA), the Science Academy Society of Turkey (BA), and the European academies, including the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Austrian Academy of Sciences (OeAW), the study builds a comprehensive database integrating member and works data. Using OpenAlex, a free and expansive platform for scholarly data, we systematically gathered, cleaned, and analyzed data to uncover trends, strengths, and gaps within and across these academies.

The project addresses key challenges, such as data consistency, multilingual barriers, and the integration of multiple author IDs, to ensure accurate representation of academic contributions. Methodologies include advanced web scraping, integration with OpenAlex for bibliometric insights, and detailed analyses of publication trends, citation impacts, research themes, and alignment with global goals like the Sustainable Development Goals (SDGs). Additionally, the study examines open-access practices, collaboration networks, and academic productivity through indices like H-Index and I10-Index.

The findings highlight significant variations in research productivity, thematic focus, and knowledge dissemination efforts across the academies. These insights not only enhance the visibility of Turkish contributions within the European research landscape but also contribute to broader discussions on optimizing academic networks and collaborations globally. This study offers a scalable framework for exploring other academies, providing valuable input for future Science of Science initiatives.

2. PROBLEM STATEMENT

The scientific landscape is an interconnected ecosystem, where equitable participation is crucial for global innovation and progress. However, disparities in visibility, resource allocation, and collaboration networks have hindered the integration of underrepresented scientific communities into the broader research ecosystem. Specifically, Turkish scientific contributions to the European research landscape have faced challenges in gaining recognition and influence.

This project builds on the principles outlined in “The Science of Science” (Fortunato et al., 2018), which emphasizes the importance of analyzing science as a dynamic system to identify biases, promote equity, and foster innovation. Motivated by the need to enhance the visibility of Turkish contributions, this project examines the networks, citation metrics, and institutional collaborations of two major Turkish academies (TÜBA and BA), alongside comparable European academies (Netherlands and Austria). The goal is to understand the structural barriers and opportunities within Turkish science, ultimately fostering a stronger integration with the European research community.

The Netherlands and Austrian academies were selected for analysis as these countries rank among the top 20 most collaborated countries for TÜBA and BA. This strategic selection ensures a focused comparison that highlights existing collaborative ties and explores opportunities to strengthen partnerships between Turkish and European academies.

By leveraging a data-driven approach, this study contributes to the broader understanding of science as a complex system, offering actionable insights to address systemic disparities while aligning with European benchmarks.

2.1. **Objectives/Tasks**

- **Objective 1: Data Cleaning and Preprocessing*:*** Prepare structured datasets by cleaning and standardizing data from TÜBA, BA, KNAW and OeAW to ensure consistency and reliability for comparative analysis, with the intended result of creating a comprehensive and uniform dataset to facilitate accurate and meaningful cross-academy comparisons.
- **Objective 2: Feature Extraction:** Extract key features such as authorship metrics, institutional affiliations, publication details, citation data, research themes and impact indicators to enable a deeper understanding of research outputs and trends, with the intended result of providing detailed insights into research productivity, influence, and thematic focus across academies.
- **Objective 3: Comparative Analysis:** Conduct a comparative analysis between Turkish academies (TÜBA and BA) and European counterparts (NL and AT academies) to highlight similarities, differences, and unique contributions, with the intended result of identifying key

differences and strengths in Turkish academies' outputs relative to their European counterparts, offering actionable insights.

- **Objective 4: Authorship and Collaboration Networks:** Analyze authorship networks, institutional collaborations, and long-term trends in co-authorship across Turkish and European academies, with the intended result of identifying collaboration patterns, gaps, and integration opportunities for Turkish academies in the European research landscape.
- **Objective 4: Research Impact and Themes**: Examine citation metrics, research visibility indicators, and publication trends. Compare research themes, productivity, and strengths across TÜBA, BA, KNAW, and OeAW, with the intended result of highlighting the relative impact, thematic focus, and opportunities for Turkish science within the broader European research context.

2.2. **Realistic Constraints**

- **Data Accessibility:** Accessing comprehensive and uniform datasets from all academies posed significant challenges. OpenAlex, while a valuable resource, had limitations such as authors having multiple IDs or some authors not being detectable by name-based searches. These challenges required manual verification and corrections, which were feasible for smaller academies but not scalable for larger ones like KNAW and OeAW. Details of these efforts are discussed in the Methodology section.
- **Data Consistency:** Variability in data formats, completeness, and quality across academies created integration and analysis difficulties. To ensure alignment with the newly added NL and AT academies, we revisited the previously prepared datasets for TÜBA and BA. This required removing most of the manually added author IDs to avoid potential biases in the dataset, ensuring all academies were processed under the same code. By prioritizing consistency and scalability, the project maintained uniformity across all datasets, facilitating reliable comparisons. Inconsistencies such as distorted entries from language-specific characters (e.g., Turkish, German, and Dutch characters) were resolved using preprocessing workflows. The methodology section provides further details on these processes.
- **Time Constraints:** The inclusion of the NL and AT academies was introduced late in the project timeline. This limited the depth of analysis for these academies, as much of the focus had already been directed toward the initial Turkish academies (TÜBA and BA). While

significant progress was made in incorporating these additional academies, further analysis of the chosen academies and inclusion of more European academies could enhance the project's scope in the future.

3. METHODOLOGY

3.1. Data Collection and Integration

3.1.1. Web Scraping and OpenAlex Integration [1]

Main goal of the project was to build a comprehensive database about science academies. The main objective of the project was built on data gathered from OpenAlex source which is a free platform that provides wide data that are useful for scientific research purposes. The data collection process for BA, TÜBA, KNAW and OeAW was conducted systematically, starting with web scraping from the academies' websites.

For TÜBA, KNAW, and OeAW, all members were listed under a single section on their respective websites. Therefore, based on the web structure of each academy, member information -such as member name and title, institutional affiliations, research areas, profile links- was fetched into structured JSON files.

In contrast, BA presented a unique challenge. Its website organized members into three separate sections: Asli (full) members, Honorary members, and Deceased members. For BA, separate scripts were developed to fetch data for "Asli" and "Honorary" members. After fetching, a duplicate check was performed to ensure that each member appeared only once. The Deceased members section, consisting of a smaller number of entries, was manually added to the consolidated JSON file containing Asli and Honorary members.

Once the members' information from all academies was successfully extracted, a custom script was used to query the OpenAlex API for additional details. This script took the JSON files containing member data as input and fetched the following:

- **Author ID:** A unique identifier for the researcher.
- **Works count:** The number of works published by the author.

[1] Relative codes can be accessed via github links in the appendix section.

- **Citation count:** The total number of citations the author received.

The enriched data was then saved as author_ids.json files for each academy. However, for cases where OpenAlex did not return an Author ID, manual searches were conducted, and the missing IDs were added along with their respective metrics.

H-index and I10-index were critical metrics added to the dataset to evaluate the academic impact of members.

- **H-index** measures a researcher's productivity and citation impact, indicating the number of publications with at least the same number of citations.
- **I10-index** counts the number of publications with at least 10 citations.

Another script was developed to retrieve these metrics from OpenAlex using the Author IDs obtained earlier. The enriched files now included comprehensive metrics for each member, facilitating deeper insights into their academic contributions.

3.1.2. Challenges in OpenAlex Data Integration

As discussed earlier in **Section 2.2 Realistic Constraints**, OpenAlex presents challenges where authors often have multiple IDs. Some of these IDs may be incomplete, blank, or associated with fewer works, while others may already encompass works included in a different ID for the same author. Addressing these issues requires manual verification to ensure data accuracy. This manual approach was feasible for BA and TÜBA, where the member count is relatively manageable.

However, for larger academies such as KNAW and OeAW, which each contain more than 600 members, manual checks become impractical. To avoid introducing bias into the dataset and ensure consistency across all academies, we standardized our approach by applying the same automated script to all datasets.

The script for fetching Author IDs prioritizes selecting a single ID that includes the most works and highest citation counts, which can reasonably be considered the author's primary ID. For cases where the script failed to fetch an ID, manual additions were made. In the final dataset, only a few authors have multiple IDs, and their relative metrics (e.g., works count, citations) have been carefully combined across these IDs to preserve data integrity.

With the finalized version of our code, all essential information has been reliably collected, ensuring consistency and minimizing errors across the datasets. This unified approach lays the groundwork for accurate analysis while maintaining the scalability required for large datasets.

3.1.3. Language Challenges and Unicode Decoding

Language and character encoding posed another significant challenge during the data collection phase. The members' names and institutional affiliations often included special characters such as **ü, ö, and ş** in Turkish and similar diacritics such as in German and Dutch. These were frequently distorted during web scraping or integration with OpenAlex.

A custom script was implemented to address this issue. The script applied automated Unicode decoding and character mapping techniques, ensuring that all names and affiliations matched their original formatting. This code was applied across the members' datasets and the works datasets to maintain consistency.

3.1.4. Works Data Fetching[2]

Following the integration of member data with OpenAlex, a secondary script was used to fetch **works data** for each author. Using the Author IDs from the previous phase, this script queried the OpenAlex API to retrieve detailed publication data for each member.
The works data included:

- **Work ID**: A unique identifier for each publication, essential for linking works to authors and enabling further analysis.
- **Title**: The publication title, offering a concise overview of the research focus.
- **DOI**: A persistent digital identifier, facilitating easy access and citation of the publication.
- **Publication Year**: Vital for analyzing temporal trends in research output across academies.
- **Cited by Count**: A crucial metric for evaluating the academic impact and influence of individual works.
- **Authorships**: Detailed information on all authors of a publication, including their affiliations and institution types. This attribute is pivotal for analyzing collaboration networks and institutional diversity.
- **Primary Topic**: Hierarchical categorization of research into fields, subfields, and domains, enabling thematic analysis across academies.

[2] Relative codes can be accessed via github links in the appendix section.

- **Concepts**: Specific subject tags associated with the work, providing granularity in thematic classification and aiding in research focus analysis.
- **Open Access Status**: Details on the accessibility of the publication, relevant for assessing the dissemination and reach of research.
- **Sustainable Development Goals (SDGs)**: Links between works and global challenges, offering insights into the societal relevance of research outputs.
- **Referenced Works**: Citations included in the publication, useful for exploring the intellectual lineage and influence of the research.

The script also managed pagination and API rate limits to fetch large datasets efficiently. Log files were created during the fetching process to track progress and ensure that any interruptions or errors in data retrieval could be easily identified and corrected. For cases where the primary topic was missing, the data was flagged, but broader categories such as concepts and fields ensured robust thematic classification. Each work retrieved was carefully filtered to include only relevant attributes, maintaining a focused and manageable dataset. The output data was structured into JSON format for seamless integration with member datasets, with a CSV version generated for easier inspection and analysis.

3.2 Analysis and Insights[3]

a) Research Productivity by Academy

This part analyzes the research productivity of the four academies, focusing on the total number of works, the number of authors, and the average works per author. Additionally, the publication trends over time are examined to understand temporal shifts in productivity and highlight active periods for each academy.

	Academy	Total Authors	Total Works	Average Works per Author
0	/CONTENT/KNAW	622	100437	161.47
1	/CONTENT/AUSTRIAN	950	127365	134.07
2	/CONTENT/BA	275	40388	146.87
3	/CONTENT/TUBA	219	29445	134.45

Figure 1. Academy-Level Statistics of Authors and Works

[3] Relative codes can be accessed via github links in the appendix section.

Figure 1 showcases the total number of authors, total works, and average works per author across the four academies. The KNAW and OeAW academies exhibit significantly higher numbers of total works compared to BA and TÜBA, reflecting their larger scale and productivity levels. Notably, the average works per author are relatively consistent across academies, suggesting comparable individual author contributions despite the size differences.

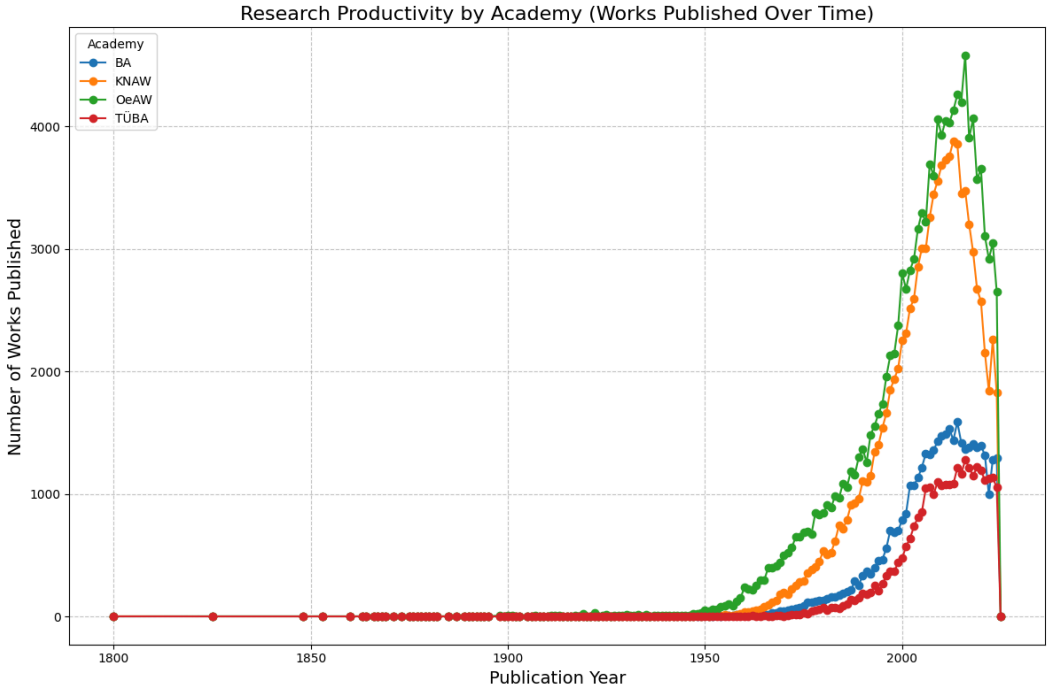

Figure 2. Research Productivity by Academy Over Time

Figure 2 illustrates research productivity trends over time by showing the number of works published per year across the four academies. A steep rise in publications can be observed starting in the mid-20th century, peaking towards the early 2000s for all academies, with OeAW and KNAW leading in total publications. This trend highlights the expansion of research activities and increased scholarly output over time.

b) **Citation Impact**

This part analyzes the citation impact of works produced by the members of each academy. Citation metrics serve as a key indicator of research influence, reflecting the extent to which an author's work contributes to the broader academic community. By

comparing total citations and average citations per work across TÜBA, BA, KNAW, and OeAW, we aim to identify variations in research impact among the academies. This analysis provides insights into the quality and visibility of the research outputs, highlighting academies with stronger academic influence and uncovering potential areas for improvement.

	Academy	Total Citations	Total Works	Average Citations per Work
0	TÜBA	2180270	29445	74.05
1	BA	2748914	40388	68.06
2	KNAW	11092659	100437	110.44
3	OeAW	10110976	127365	79.39

Figure 3. Total Citations, Works, and Average Citations per Work by Academy

Figure 3 provides a statistical overview of citation metrics for the four academies. KNAW leads with the highest total citations (11,092,659), followed by OeAW with 10,110,976 citations. TÜBA and BA have significantly lower total citations, at 2,180,270 and 2,748,914 respectively. However, when average citations per work are considered, KNAW again stands out with 110.44 citations per work, followed by OeAW with 79.39. TÜBA and BA have relatively lower averages of 74.05 and 68.06 citations per work, respectively, emphasizing the higher impact and influence of works from KNAW and OeAW. This figure highlights disparities in research visibility and impact across the academies.

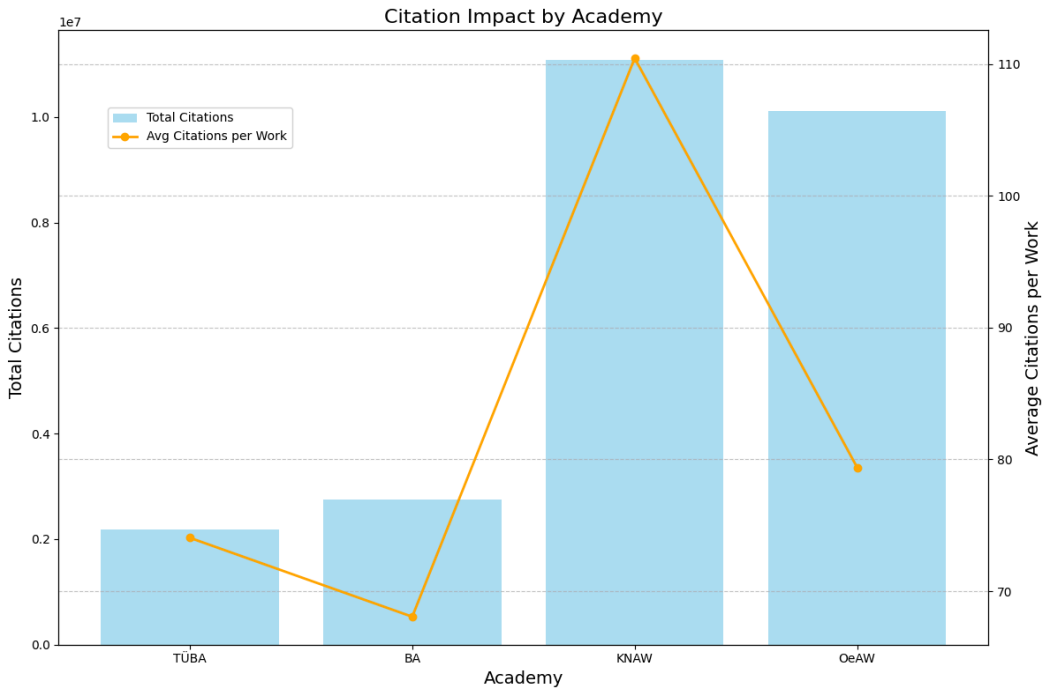

Figure 4. Citation Impact by Academy (Total Citations and Average Citations per Work)

Figure 4 presents the citation impact of each academy, highlighting the total citations and average citations per work. KNAW leads significantly with the highest average citations per work at approximately 110, suggesting a notable research impact. While OeAW follows with substantial total citations, its average citations per work are comparatively lower at around 80. TÜBA and BA exhibit lower total citations and averages, reflecting their smaller publication scale and possibly different research focus areas.

c) **Research Fields and Concepts**

This part delves into the thematic focus and specialization areas of each academy by analyzing their research fields, subfields, and associated concepts. Understanding these dimensions provides valuable insights into the academic strengths and priorities of each institution, as well as their contributions to different domains of knowledge. This analysis aims to uncover thematic strengths, highlight areas of specialization, and provide a basis for comparing the scope and diversity of research efforts across academies.

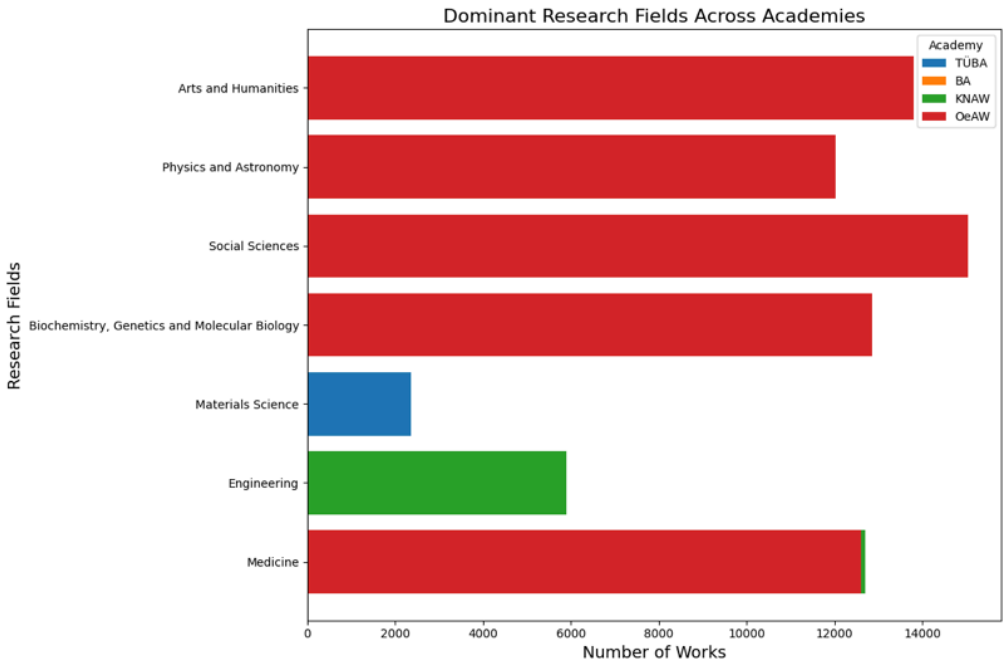

Figure 5. Dominant Research Fields Across Academies

Figure 5 illustrates the distribution of dominant research fields across the four academies (TÜBA, BA, KNAW, and OeAW). The visual dominance of the red bars (representing OeAW) across most fields, such as Medicine, Social Sciences, and Biochemistry, indicates that OeAW has a broader and more extensive research output in these areas compared to other academies. This finding suggests that OeAW has a stronger thematic focus or more resources dedicated to these fields. In contrast, TÜBA and BA show more localized contributions in fields like Materials Science and Engineering.

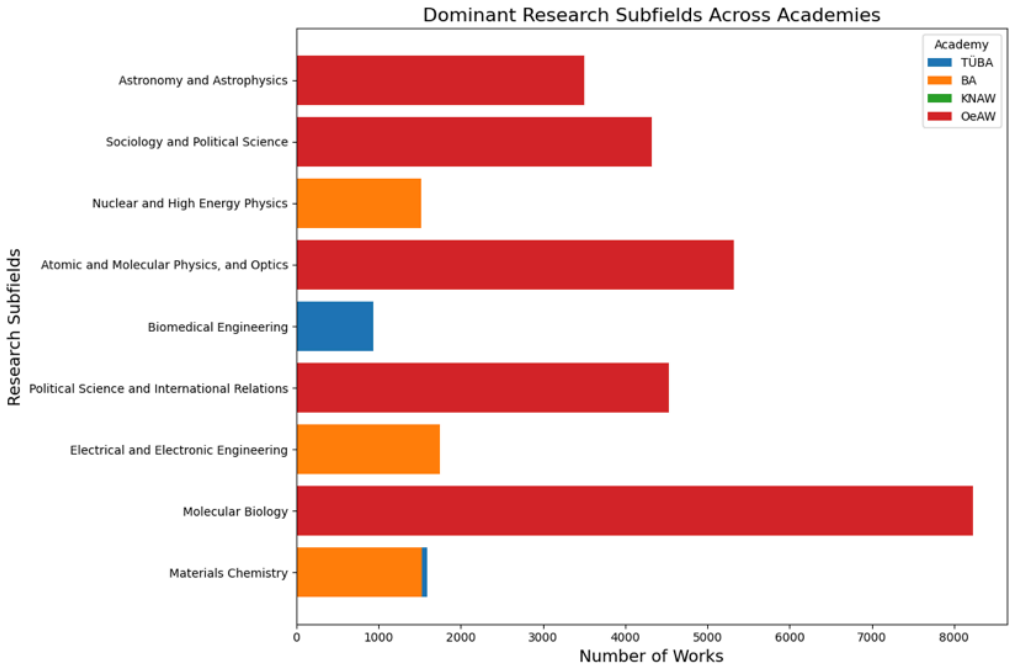

Figure 6. Dominant Research Subfields Across Academies

Figure 6 delves into research subfields, revealing a more granular view of academic contributions. Here, OeAW continues to dominate subfields like Molecular Biology, Political Science, and Astronomy, which aligns with its broader impact seen in the research fields. Notably, TÜBA shows a strong presence in Biomedical Engineering, highlighting its niche specialization, while BA contributes significantly to Nuclear and High Energy Physics. These subfield distributions underscore each academy's thematic strengths and areas of expertise.

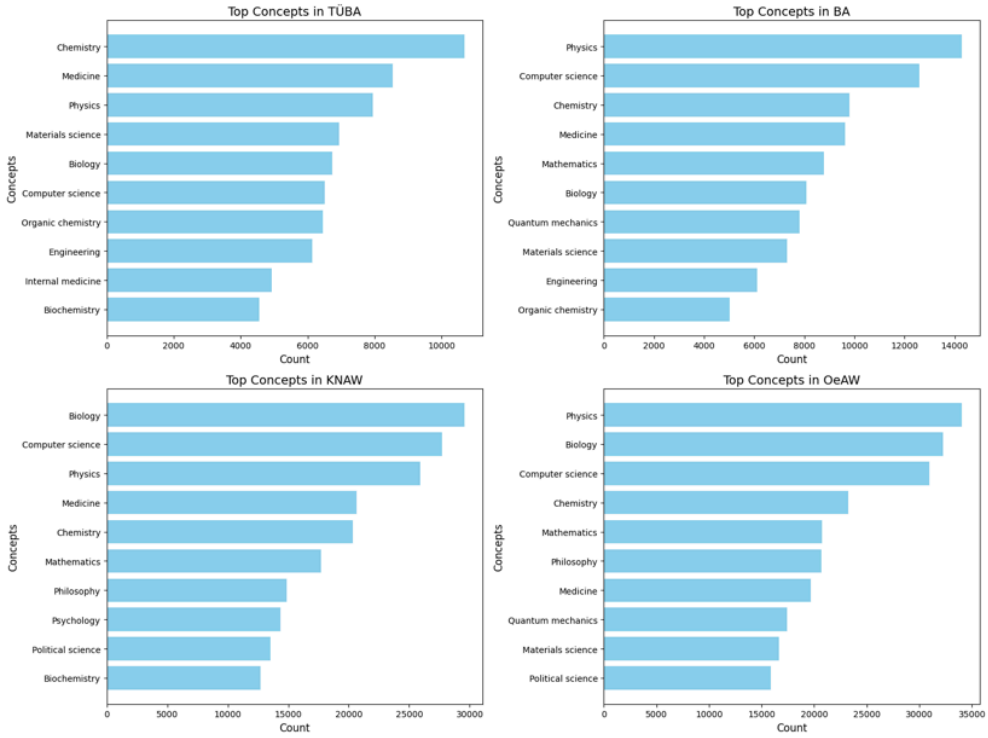

Figure 7. Top Concepts Across Academies

Figure 7 presents the top concepts explored by works from each academy. The red bar dominance in all subplots confirms that OeAW's works cover a wider range of concepts, such as Physics, Biology, and Computer Science, with a significant focus on fundamental sciences and interdisciplinary areas. TÜBA and BA, however, show strengths in fields like Chemistry and Materials Science, reflecting their contributions to applied research. This figure emphasizes the diversity in conceptual focus across the academies and highlights potential collaboration opportunities in overlapping areas such as Physics and Computer Science.

d) **Alignment with Sustainable Development Goals (SDGs)**

The alignment of research outputs with Sustainable Development Goals (SDGs) provides valuable insights into the societal relevance and global impact of academic works. SDGs represent global challenges such as quality education, climate action, and sustainable cities, offering a framework to assess how research addresses critical issues. By analyzing the alignment of works from TÜBA, BA, KNAW, and OeAW with SDGs, we aim to

identify thematic contributions and assess how each academy's research efforts contribute to these global goals.

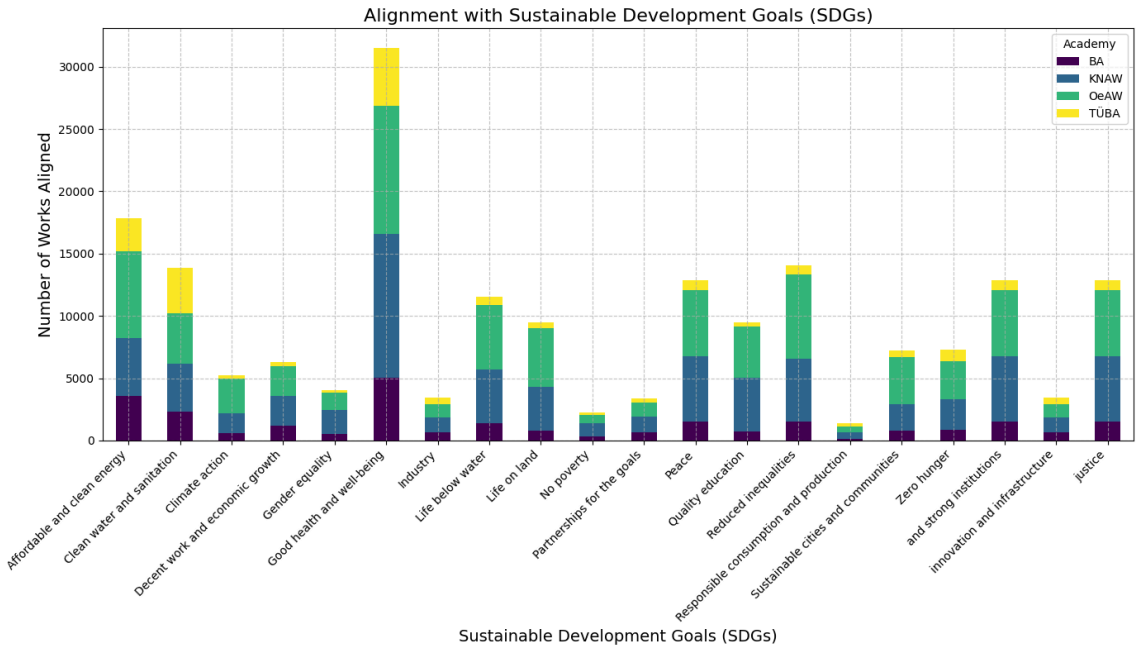

Figure 8. Top Concepts Across Academies

Figure 8 visualizes the alignment of works from each academy with the United Nations' Sustainable Development Goals (SDGs). The stacked bar chart highlights the number of works from each academy contributing to specific SDG categories. Each bar represents a particular SDG, with the stack segments indicating the contributions from TÜBA, BA, KNAW, and OeAW.

The figure reveals several key insights:

- **Dominance in "Industry, Innovation, and Infrastructure":** This SDG category shows the highest alignment across all academies, with significant contributions particularly from KNAW and OeAW.
- **Diverse Contributions Across Goals:** While all academies show engagement across various SDGs, OeAW has a notably larger contribution across a wider range of SDGs compared to other academies.
- **TÜBA's Focus Areas:** TÜBA demonstrates substantial alignment with "Clean Water and Sanitation" and "Affordable and Clean Energy," indicating a focus on environmental and sustainability-related research.

- **Smaller Contributions from BA:** BA has lower representation across SDGs compared to the other academies, suggesting a more focused or limited scope of research output in alignment with these global goals.

e) **Open Access Trends**

Open access (OA) status is crucial for understanding the accessibility of research outputs and their contribution to global knowledge dissemination. **Figure 9** presents the distribution of OA statuses—gold, green, hybrid, bronze, diamond, and closed—across BA, TÜBA, KNAW, and OeAW, highlighting trends in accessibility.

Explanation of Open Access Statuses

- **Gold**: Fully open access publications where authors typically pay an article processing charge (APC) to ensure their work is freely available to everyone.
- **Green**: Self-archived works by authors in an institutional or subject repository, often following an embargo period.
- **Bronze**: Free-to-read content on the publisher's website but without a clearly defined open access license.
- **Diamond**: Fully open access without any fees charged to authors or readers, often supported by institutions or consortia.
- **Hybrid**: Subscription-based journals that offer authors the option to make individual articles open access upon payment.
- **Closed**: Works that are not freely available and require subscription or payment to access.

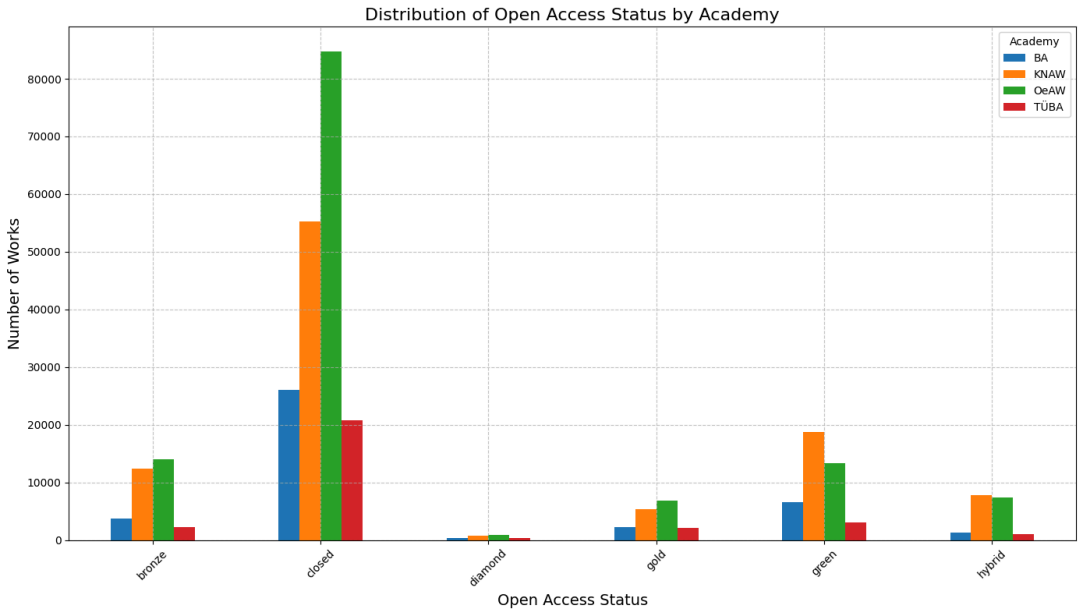

Figure 9. Distribution of Open Access Status by Academy

Figure 9 reveals that a significant portion of works, particularly from KNAW and OeAW, are classified as **closed access**, limiting public accessibility. BA and TÜBA, on the other hand, demonstrate a stronger presence in **bronze** and **green** categories, showcasing efforts toward making research freely available. However, fully open access categories like **gold** and **diamond** remain underrepresented across all academies, indicating limited adoption of fully accessible publishing models. Hybrid publications are also moderately present, reflecting a reliance on subscription-based journals with optional OA.

f) **Author H-Index and I10-Index Distribution**

The H-Index and I10-Index are widely recognized metrics for evaluating academic influence and productivity. The H-Index measures the number of publications (h) that have received at least h citations, reflecting both the quantity and quality of an author's work. The I10-Index, on the other hand, represents the number of publications with at least 10 citations, emphasizing consistent contributions over time. This analysis explores the distribution of these indices across authors from TÜBA, BA, KNAW, and OeAW, providing insights into the academic influence and productivity patterns of each academy.

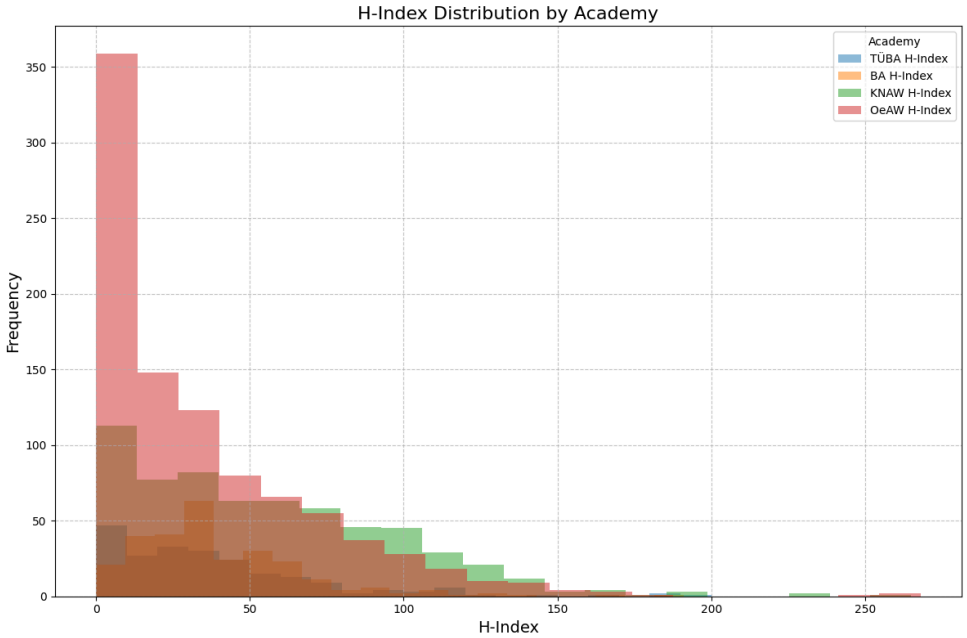

Figure 10. H-Index Distribution by Academy

Figure 10 shows the distribution of H-Index values for authors across the four academies. The majority of authors have lower H-Index values, indicating a concentrated academic impact among a smaller subset of authors with high H-Index scores. OeAW demonstrates the widest spread in H-Index values, suggesting a higher number of influential authors compared to other academies. TÜBA and BA show relatively lower H-Index distributions, reflecting a smaller pool of highly cited works.

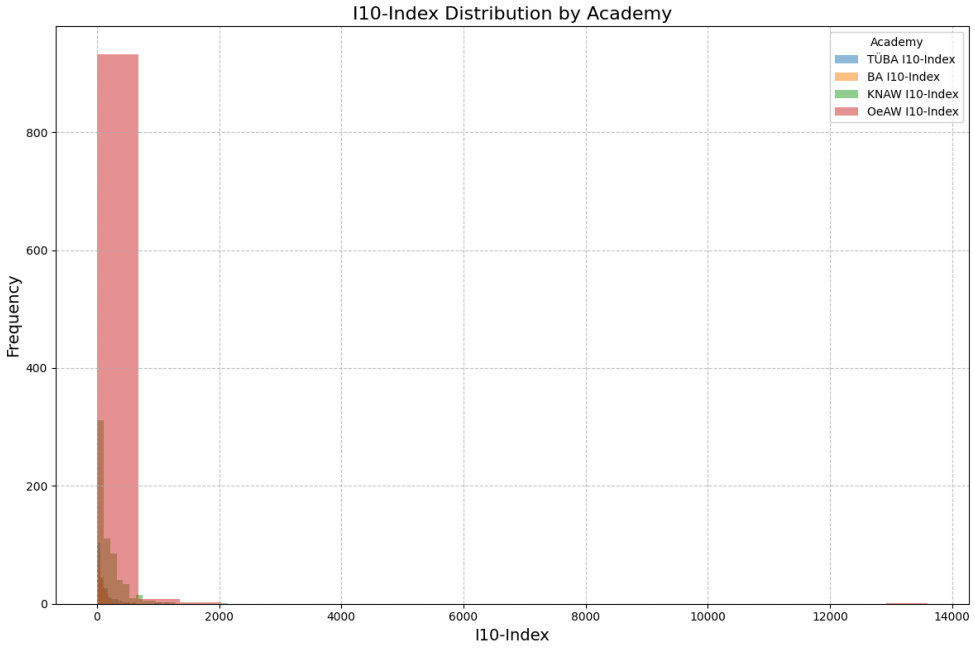

Figure 11. I10-Index Distribution by Academy

Figure 11 presents the distribution of I10-Index values, illustrating the number of publications with at least 10 citations. Similar to the H-Index distribution, OeAW exhibits the highest concentration of authors with significant I10-Index values, reinforcing its dominant role in research productivity. KNAW follows closely, while TÜBA and BA show narrower distributions, reflecting a smaller scale of publications meeting the I10-Index criteria.

g) **Authorship and Collaboration Networks**[4]

Collaborations between authors and institutions play a vital role in advancing knowledge and innovation. Network visualization is a powerful approach to display complex scientific relationships between co-authors and institutions. The Gephi tool was useful for creating network visualizations and analysis. Network graphs are one of the best ways to display methods of how authors and institutions collaborate. Furthermore, it shows connectivity and openness to collaboration between authors in science academies.

In Python, a file was created showing collaborations between authors from the same science academy. After creating a required file that stores network information such as “Source”, “Target”, and “Weight” in Python, that file was imported into Gephi. (“Source” means the author, “Target” means the author that

[4] Figures can be accessed through github link provided a the appendix section.

collaborated with the “Source” author, and “Weight” means how many works that they collaborated in.) In Gephi, “Fruchterman Reingold” Layout model was selected. Nodes were colored with the “Modularity Class-Partition” method. Nodes’ size were arranged relative to their “Degree” ranking.

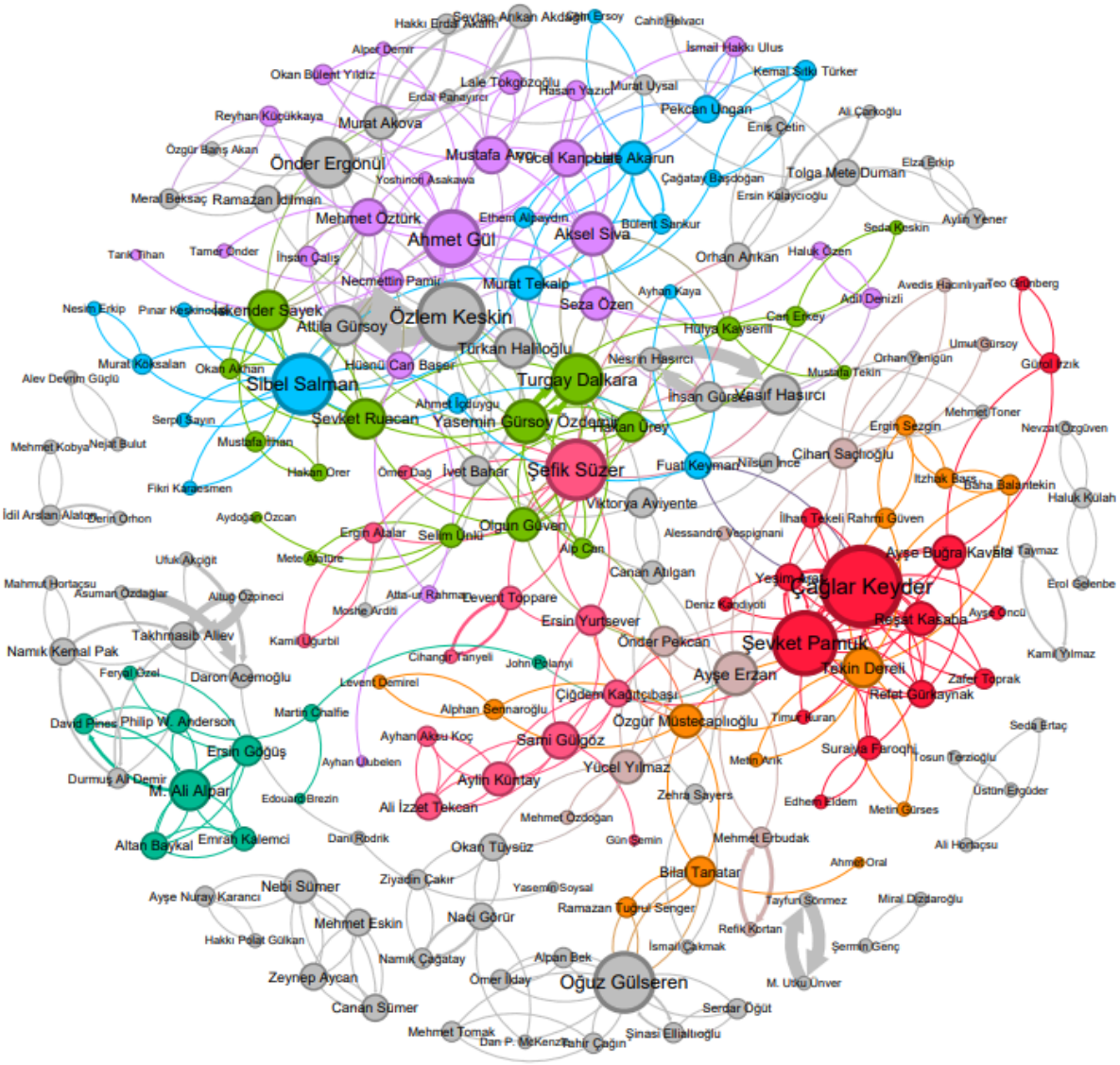

Figure 12. BA Co-authorship Network

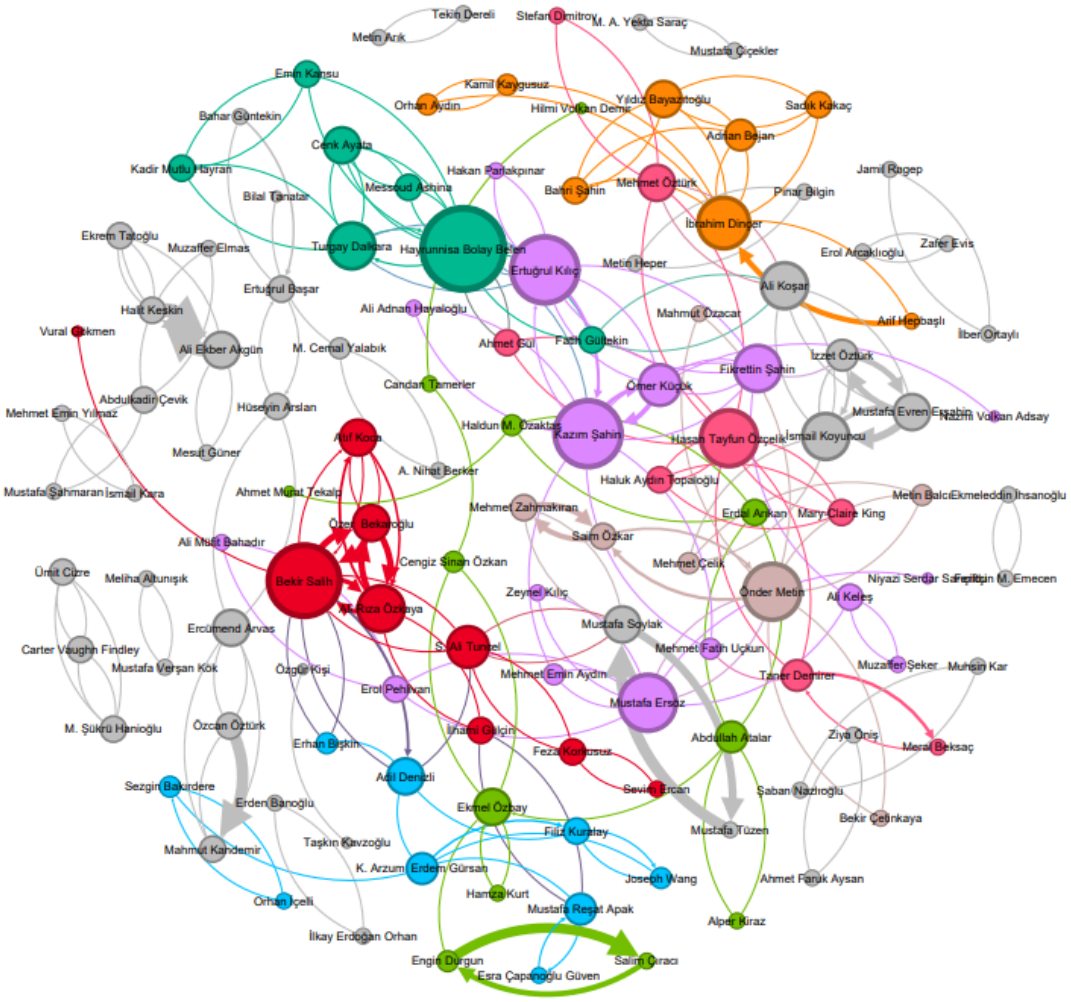

Figure 13. TÜBA Co-authorship Network

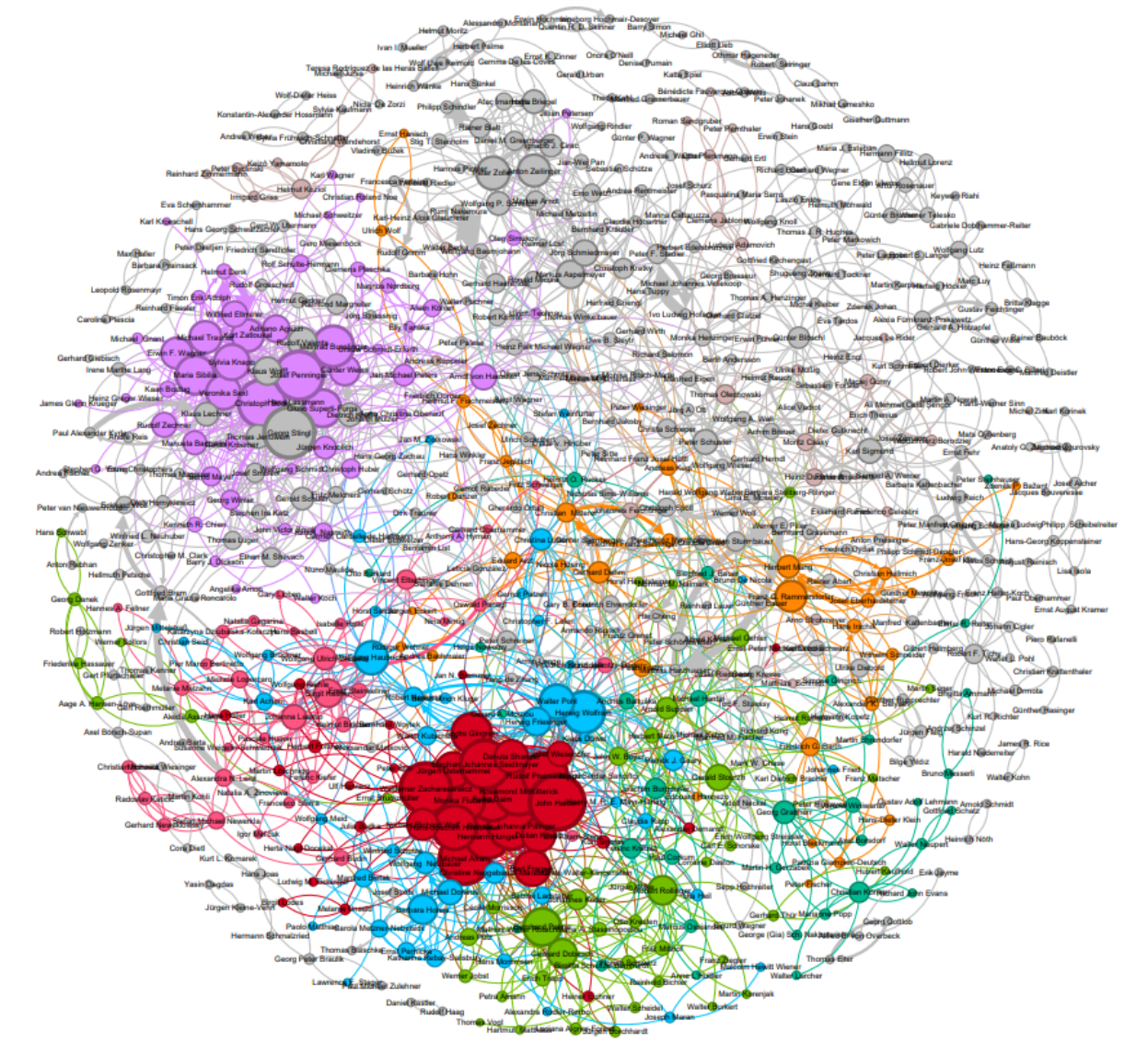

Figure 14. OeAW Co-authorship Network

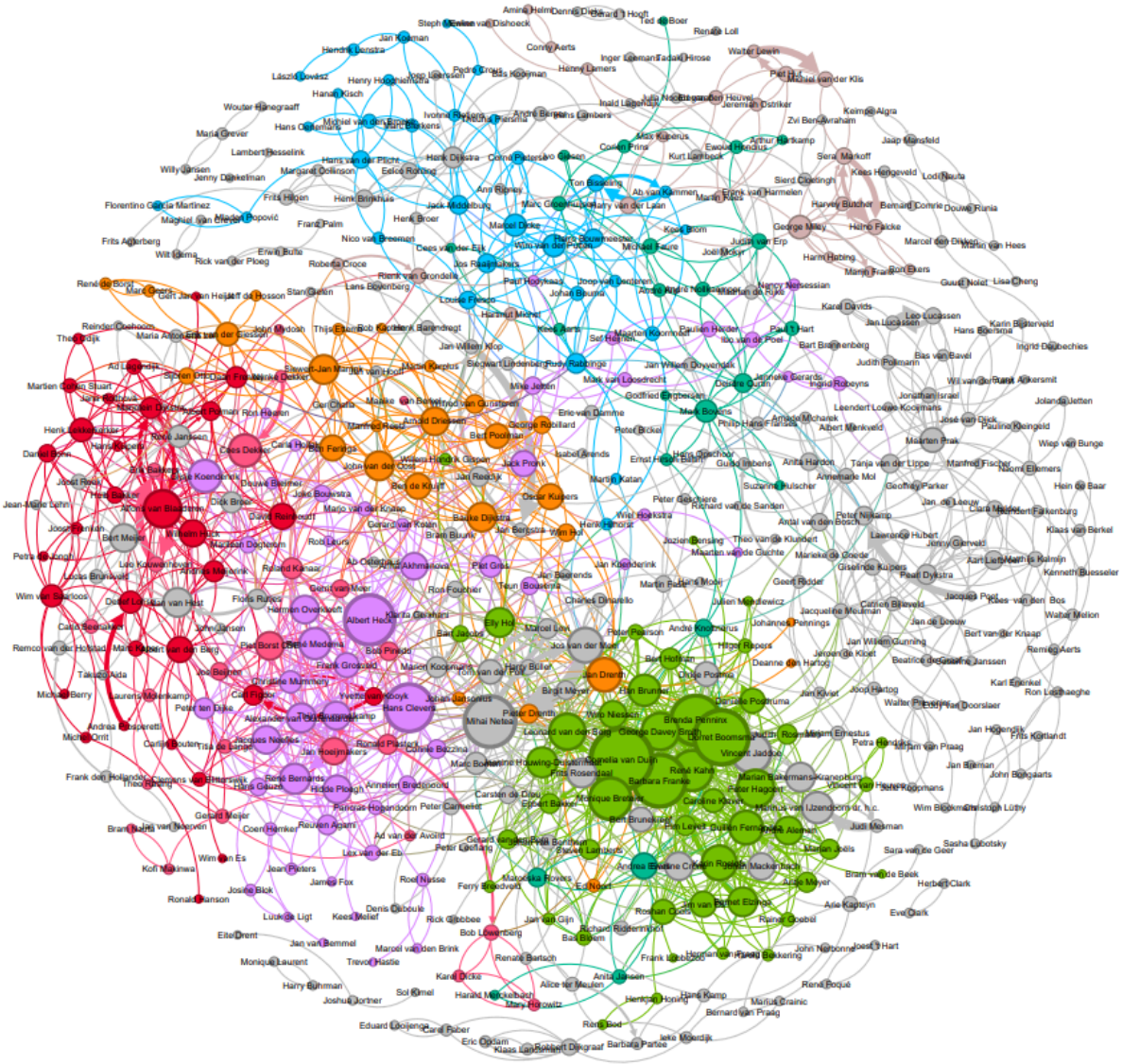

Figure 15. KNAW Co-authorship Network

h) Institution Collaboration Networks[5]

With the same method, collaborations of institutions were created. For better readability and understanding, we only included collaborations that have collaboration weights more than 40.

[5] Figures can be accessed through github link provided at the appendix section.

Figure 16. Collaborations of Institutions Network of BA

Figure 17. Collaborations of Institutions Network of TÜBA

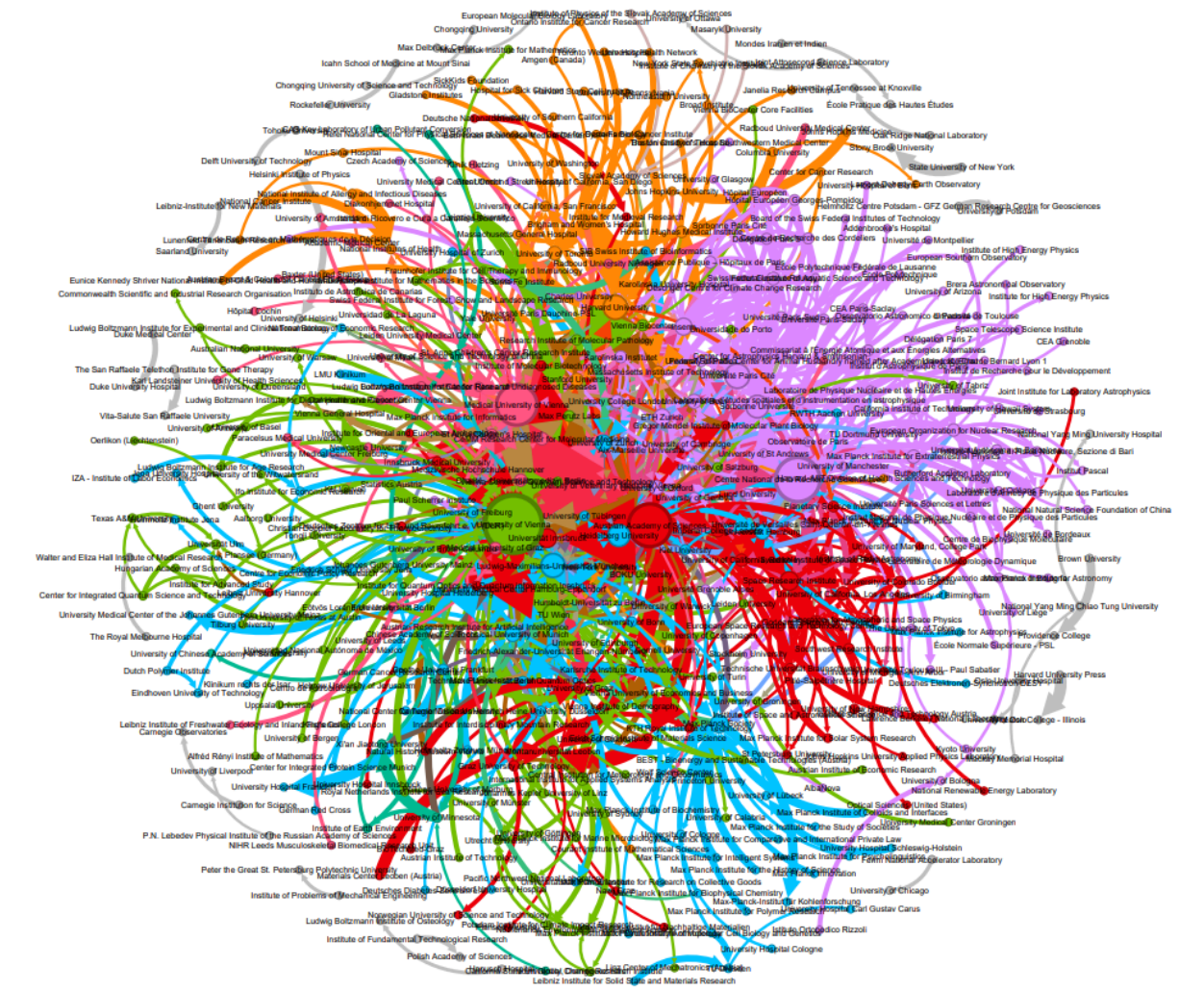

Figure 18. Collaborations of Institutions Network of OeAW

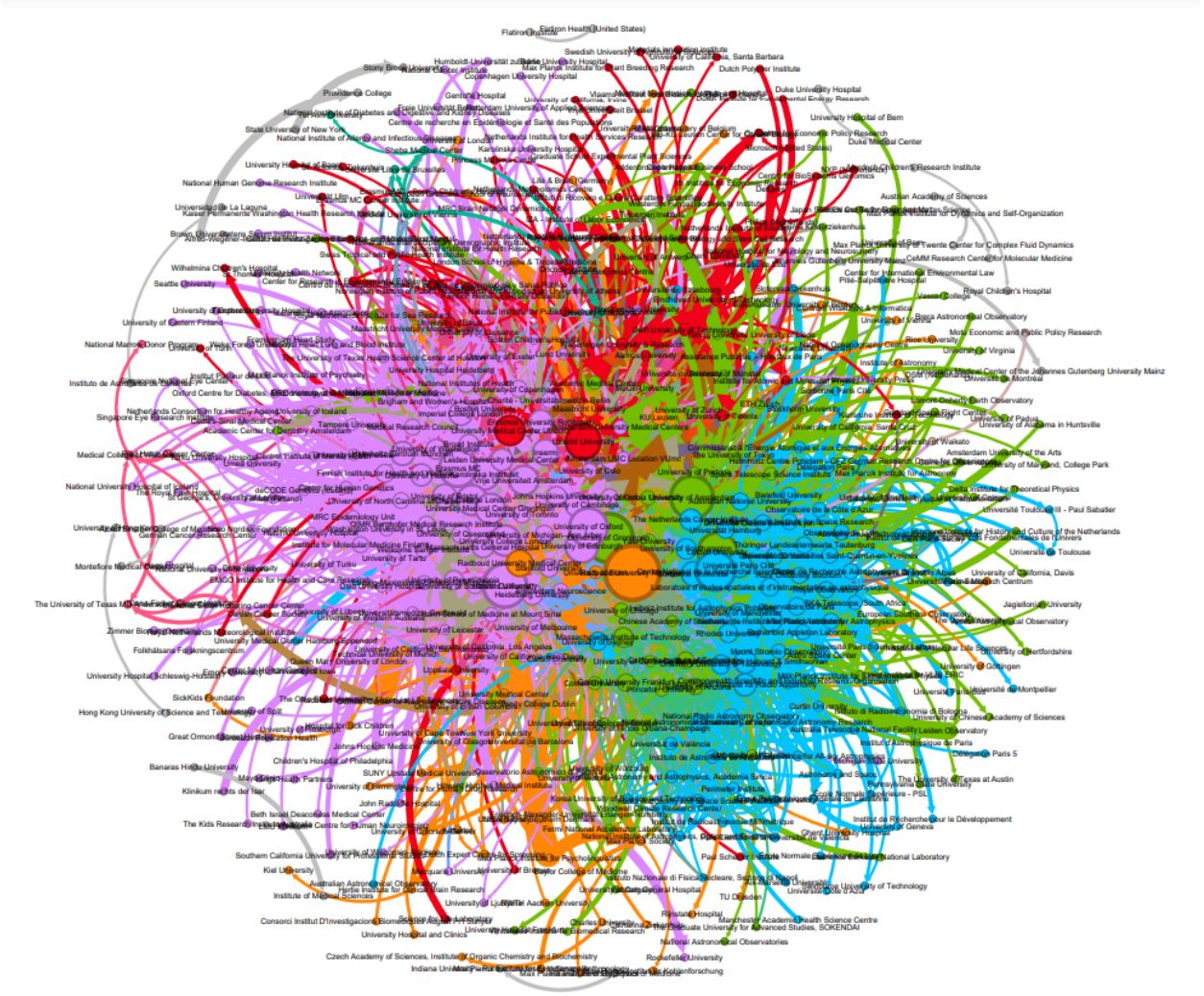

Figure 19. Collaborations of Institutions Network of KNAW

For better readability, we filtered the file again, and only selected collaborations with more than 100 collaboration weights.

Figure 20. Collaborations of Institutions Network of OeAW (weight >= 100)

Figure 21. Collaborations of Institutions Network of KNAW (weight >= 100)

4. RESULTS & DISCUSSION

The project aimed to analyze and compare the research productivity and impact of four prominent science academies: TÜBA, BA, KNAW, and OeAW. The primary objectives included evaluating the total number of works, the number of authors, the average works per author, and the citation impacts of each academy's contributions. Additional goals were to explore thematic specialization, alignment with Sustainable Development Goals (SDGs), and collaborative networks across these institutions.

Both KNAW and OeAW exhibit higher overall research productivity compared to TÜBA and BA, as indicated by the total number of works and average works per author. However, when

comparing the Turkish academies, **TÜBA** appears more aligned with the European standards in terms of total output, showing a closer similarity in scale to the smaller European academy, though still significantly behind.

In terms of citation impact, KNAW and OeAW significantly outpace both TÜBA and BA. However, TÜBA has a slightly higher average citation per work compared to BA, indicating that in terms of quality and influence of research outputs, **TÜBA** may be more comparable to European academies, albeit still at a lower scale.

The thematic focus shows OeAW with a broad range of disciplines, excelling in fields like Medicine, Social Sciences, and Biochemistry, similar to the range seen in larger research institutions. TÜBA shows strength in niche areas such as Biomedical Engineering. **BA**, on the other hand, aligns slightly more with European academies in terms of diversity in research themes, particularly in fundamental and applied sciences like Nuclear and High Energy Physics, suggesting a broader thematic alignment with European standards.

European academies exhibit dense and extensive collaboration networks. TÜBA and BA, while having their networks, would likely be less complex given their smaller size and output. Yet, **TÜBA** potentially aligns closer with European patterns due to its participation in international collaborations, as often seen in larger research outputs.

Regarding open access trends, European academies show a substantial number of closed-access works, with TÜBA and BA having a stronger inclination towards open access models. Here, **BA** reflects a trend more akin to European practices by maintaining a balance between open access and traditional publishing models, indicative of a shift towards greater accessibility in line with global academic trends.

European academies, particularly KNAW and OeAW, show significant contributions to a broad range of SDGs, reflecting their extensive research agendas. TÜBA's focused contributions to SDGs like Clean Water and Sanitation suggest niche alignment, whereas **BA's** limited but focused output on specific SDGs could indicate a strategic alignment that mirrors the targeted approach seen in some European contexts.

With a wider range of partnerships with internationally recognized academic institutions and research facilities like Harvard University, Lawrence Berkeley National Laboratory, and the Centre National de la Recherche Scientifique (CNRS), **TÜBA** exhibits a stronger alignment with European institutions. Its ties across a wide range of nations and disciplines point to a more

globalized and multidisciplinary research network. In contrast, **BA** has strong connections to universities such as the University of Oxford, Sorbonne Université, and Radboud University. Although **BA**'s partnerships have a significant impact, **TÜBA**'s wider and more varied network is more similar to the institutional partnerships of European academies that prioritize cross-border cooperation, such as **KNAW** and **OeAW**.

In summary, while **TÜBA** exhibits some similarities with European academies in terms of citation impact and collaboration networks, **BA** aligns more closely in thematic diversity and open access trends. Overall, each Turkish academy shares certain characteristics with the European models but also maintains distinct differences primarily due to scale, focus, and resource allocation.

5. IMPACT

The project titled "Science of Science Research for EU Research Landscape" has made significant contributions scientifically by making comparisons between Turkey (TÜBA and BA) and Europe (KNAW and OeAW) science academies. That comparison includes research outputs, citation metrics, and collaboration networks across Turkish (TÜBA and BA) and European (KNAW and OeAW) academies. The project identifies similarities and differences by deeper analysis such as research trends, collaboration approaches. Metrics such as author productivity, authors' citation numbers, and co-authorship strength creates a brand new framework.

From a technological standpoint, the project employed advanced data collection and analysis techniques using OpenAlex, Python, and network analysis tools. The visualization of collaboration and citation networks exemplifies the project's innovative use of data-driven technologies, making complex relationships easier to understand.

Socio-economically, the project fosters transparency in academic contributions and collaboration networks, encouraging inclusivity and equitable recognition in the scientific community. By analyzing trends in research outputs and collaborations, the findings could inform policies aimed at strengthening Turkey's position in global research and enhancing collaborative efforts with European academies.

In terms of Freedom-to-Use (FTU), there are no constraints. The data utilized in this project is publicly available via OpenAlex and academy websites. The methodologies, visualizations, and outcomes are freely accessible and can be extended or adapted by other

researchers. The project's open-source tools and approach further ensure that its findings can be replicated, validated, and enhanced by the academic community.

6. ETHICAL ISSUES

This project has been conducted in accordance with ethical guidelines, and no significant ethical concerns have been identified. The data used is open-access and publicly available via academies' websites and OpenAlex system, ensuring that no sensitive or personal information is involved.

All sources will be included in the "9. References" section to adhere to licensing requirements, avoiding any potential copyright issues. Additionally, the project avoids any examinations that could lead to biases or misrepresentation of the data.

The findings and tools developed in this project are intended to foster collaboration and inclusivity in the academic research community. There are no ethical concerns regarding the use or dissemination of the results.

7. PROJECT MANAGEMENT

The initial project plan was focused primarily on a comparative analysis between the Türkiye Bilimler Akademisi (TÜBA) and Bilim Akademisi (BA). At the beginning, team members were not fully aware of how the European research landscape could be integrated into the project. However, following valuable explanations and guidance from the supervisor, the scope of the project evolved to include a broader perspective.

Under the supervisor's guidance, the team embraced the mission of incorporating data from at least two additional European academies. This adjustment aimed to provide a more comprehensive analysis of the collaboration and impact of scientific contributions on a European scale. As a result, the project shifted from a bilateral comparison to a multi-academy exploration, which required redefining tasks, timelines, and data collection processes.

This change in scope presented several challenges, including identifying and gathering data from European academies with diverse structures and data formats. Despite these difficulties, the team adapted by refining their workflow, dividing responsibilities, and prioritizing efficient collaboration.

Through this process, team members gained valuable insights particularly in adapting to evolving objectives and additional tasks. The experience underscored the importance of flexibility, clear communication, and leveraging guidance from supervisors to align project goals with broader research objectives. This approach not only ensured the project remained on track but also enriched the final outcomes by incorporating a more nuanced and expansive analysis.

8. CONCLUSION AND FUTURE WORK

This project has provided a comprehensive analysis of the research outputs and collaboration networks of Turkish (TÜBA and BA) and European (KNAW and OeAW) science academies. By leveraging advanced data analysis and visualization techniques, the team successfully highlighted the scientific contributions, citation trends, and collaborative structures of these academies. The findings underline key similarities and differences between academies, offering valuable insights into the strengths and areas for improvement in the academic research landscape.

One of the most significant outcomes is the creation of a structured database and the development of metrics to evaluate the performance of academies. These tools have laid the groundwork for a broader comparative analysis, providing a replicable framework for future studies. The analysis of co-authorship networks and institutional collaborations has revealed critical patterns in research partnerships, fostering a deeper understanding of how Turkish academies take part in the global academic network.

The logical next steps for this project include extending the analysis to more science academies. This would not only broaden the comparative scope but also provide a clearer perspective on how TÜBA and BA align with European research standards. Moreover, incorporating additional metrics, such as societal impact and interdisciplinary research patterns, would further enrich the study.

Future iterations of the project could be adapted to examine the research contributions of individual institutions or countries, offering more granular insights. If pursued by other students or researchers, the project has the potential to significantly contribute to the understanding of regional and global academic landscapes.

9. APPENDIX

- GitHub links to Python notebooks and other related files used in the project: https://github.com/fkayray/ENS492

  https://github.com/guleraysegul/SciOSci_Graduation_Project

10. REFERENCES

- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., … & Barabási, A.-L. (2018). Science of science. *Science, 359*(6379). https://doi.org/10.1126/science.aao0185
- OpenAlex. (n.d.). About us. Retrieved December 15, 2024, from https://help.openalex.org/hc/en-us/articles/24396686889751-About-us