Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #48180

Understanding Research Dynamics at the University of Arizona: An AI-Driven Metadata Analysis

Dr. Iqbal Hossain, The University of Arizona

Dr. Iqbal Hossain is the Director of Research Data Science at the University of Arizona and a lead data scientist with expertise in network analysis, visualization, algorithms, natural language processing, and applied generative AI. He joined the University of Arizona as a postdoctoral researcher in 2016 and invented the UArizona Knowledge Map (KMap). KMap leverages various types of research metadata, analyzes them, and visualizes the data to enable a Retrieval-Augmented Generation (RAG) system for research-related inquiries at the University of Arizona.

Dr. Hossain has published over two dozen peer-reviewed articles in areas including data science, computer algorithms, graph theory, network visualization, information retrieval, information visualization, machine learning, natural language processing, and database systems. He actively collaborates with external groups, students, and researchers at the University of Arizona on a wide range of research projects.

With over 20 years of professional experience in research, IT systems development, team management, and innovation, Dr. Hossain is passionate about designing data science systems and leading efforts to solve the university's data science problems to help researchers and the university's administration.

Thomas Harman, University of Arizona

Thomas Harman has been an IT Business Analyst with the KMap team for nearly two years. In this role, he is responsible for data collection and processing, site analytics, and supporting various campus projects using KMap's powerful tools. His primary focus is on processing and understanding institutional data, such as grant proposals, faculty publications, and campus-wide collaborations to better assist the UA community in its research and academic environment. Thomas earned his degree in Information Technology from the University of Arizona and is currently pursuing a Master of Science in Management Information Systems in his free time.

Wesley Nguyen, University of Arizona

Wesley Nguyen is a web developer for the KMap team, where he builds and maintains the KMap website with a focus on performance, accessibility, and usability. He works closely with backend systems and institutional data sources to support research discovery and collaboration across the university. His work spans frontend development, database integration, and designing tools that help users explore faculty expertise and global engagement. Wesley earned his degree in Geography Data Science from the University of Washington and spends his free time tinkering with servers and listening to new original soundtracks.

Ravneet Chadha, The University of Arizona

Understanding Research Dynamics at the University of Arizona: An AI-Driven Metadata Analysis

Iqbal Hossain, Thomas Harman, Wesley Nguyen, Ravneet Chadha
University of Arizona Knowledge Map (KMap)
University of Arizona
Emails: {hossain, harman, wesngu28, rschadha}@arizona.edu

Abstract

This study explores the complex research landscape of the University of Arizona, which boasts over $955 million in annual research expenditures. By analyzing an extensive dataset of 190,000 publications, 6,000 researchers, 24,000 internal collaborations, 50 funding agencies, 40,000 funded projects, and 23,000 development research proposals, we reveal valuable insights into the institution's research strengths and emerging trends.

The methodology involves systematically collecting, processing, and analyzing diverse research metadata from multiple sources. We address the challenges of managing large-scale, unstructured data to provide a comprehensive view of the university's research activities. Key findings include: (a) the role and diversity of researchers, (b) interconnections between departments and colleges in collaborative research, (c) the university's research strengths in grants, patents, proposal writing, and publications, and (d) an analysis of the institution's collaboration network.

The findings have been integrated into the interactive University of Arizona Knowledge Map (KMap) platform, which maps research strengths and collaborations across the university. These insights offer valuable guidance to researchers, administrators, and policymakers aiming to enhance the university's research strategy and impact.

INTRODUCTION

In large organizations, research and scholarly metadata are distributed internally and are also publicly available on the internet. Examples of such metadata include publications, proposals, research projects, patents, course abstracts, grant projects, and biographies. The question is: Are we utilizing this metadata effectively in the decision-making process? How can we make this metadata more accessible to help determine strategy and data-driven decision-making?

Collecting, managing, and extracting useful information from these metadata is challenging. The University of Arizona Knowledge Map (KMap) addresses this challenge by collecting, connecting, and building systems using this information. The KMap system is easy to navigate, similar to Google Maps, allowing users to zoom and pan to explore departments,

Corresponding author

researchers, collaborations, and clusters. It also provides an organizational overview with features like heatmaps, to enable visualization of different types of activities. In this paper, we analyze a significant amount of data from the KMap system.

This paper addresses the following questions about the organization:

If we analyze all internal collaborations and projects, including co-authorship of papers, PI-Co-PI relationships, patent authorship, and student research supervision, what does the internal collaboration network look like within the organization? Even failed or rejected projects contribute to collaboration.

Do we have multiple large groups of researchers, or is there a single, giant collaboration network that connects most researchers?

How has collaboration within the organization changed over the last 10 years?

How do various departments collaborate?

Which colleges are highly interconnected or more isolated in terms of collaboration?

Where are the research strengths of the organization?

Which areas of research receive significant funding?

The main challenge addressed in this paper is the collection of diverse research information within the organization, integrating it into a cohesive dataset, and analyzing it to extract meaningful insights.

RELEVANT WORK

We have not discovered any work that uses extensive scholarly metadata, such as publications, proposals, research projects, patents, course abstracts, grant projects, and biographies within an organization, to understand the research landscape of a single organization. Hence, we focus on reviewing general research and the analysis of scholarly metadata, as well as internal institutional collaboration analysis, to cover similar work that has been done.

As research continues to move online, scholarly metadata is key to understanding the dynamics of research activity. Numerous studies have explored the role of scientific events, such as conferences and workshops, finding that events are crucial in encouraging collaboration and communication, particularly in fields such as computer science [1]. Institutional repositories and online libraries have infrastructure and metadata that support collaboration and research analysis, but improvements are needed in certain areas [2], [3].

A form of scholarly metadata, the H-index, a trackable metric for researcher productivity, has also been the focus of discussion. The H-index is often discussed both for its ability to indicate productivity and serve as a point of comparison between an institution's departments or individual researchers [4], [5], [6]. While its importance in assessing research units is

recognized, there is broad agreement that the metric could be refined to better reflect the complexities of research impact.

Alongside the analysis of scholarly metadata, significant attention has also been given to institutional collaboration. Collaboration among researchers, universities, industries, and institutions can influence productivity, with its effectiveness shaped by factors like partnership type, proximity, and academic discipline [7], [8]. For example, a case study at Middle Tennessee State University highlights how collaborations between university libraries and academic colleges can enhance the preservation and visibility of scholarly output [9]. While collaboration is clearly recognized as beneficial, studies also emphasize areas that could be improved, including aligning subject matter, matching researchers interests, implementing strategic policy, and improving academic infrastructure to further streamline establishing and growing collaborative networks.[10], [8], [11]. Institutional efforts to foster collaboration, such as seed grants and retreats, also vary in their effectiveness as what may work for key researchers may not work as well for others [12].

To recognize collaborative efforts, institutions tend to release institutional research retrospectives to help explain and visualize the research output. These often manifest in the form of general overviews that highlight large numbers and key researchers [13], [14], [15], [16]. While these overviews are valuable for a broad assessment of an institution's research landscape, it fails to capture deeper insights into their research and collaboration networks.

In recognition of this shortcoming, several institutions have leveraged network analysisfor a more nuanced look into their research networks. For instance, researchers at the University of New Mexico applied network analysis to map how university courses interconnect within degree programs, identifying "crucial" courses that significantly impact student progress and graduation timelines [17]. Similarly, the University of North Texas explored the role of metadata in search query success within their institutional repository, demonstrating the importance of metadata in enhancing visibility [18].

Social network analysis (SNA), a specific application of network analysis, has become important as a means of evaluating scholarly collaboration and understanding research within institutions [19]. At the University of Kentucky's Markey Cancer Center, SNA was used to assess interdisciplinary co-authorship networks, revealing how policies and funding can help foster collaboration and diversify scholarly output [20]. Similarly, researchers at Covenant University's computer science department utilized SNA to develop a collaboration recommendation system based on co-authorship network properties to identify potential partnerships and collaborations [21].

Building upon the research into scholarly metadata and network analysis, extensive work has been dedicated to developing systems to help manage and visualize scholarly metadata. Researchers at the University of Arkansas developed "CollaborationViz", an interactive tool that analyzes research collaboration networks by leveraging grant metadata metrics like centrality and clustering coefficients to evaluate collaboration strength and structure [22]. Likewise, the NcoVis framework offers a novel approach to visualizing academic collaboration networks, focusing on their structure and evolution to analyze changes over time [23]. Furthermore, Korona, a knowledge-based framework, uses semantic similarity within

knowledge graphs to uncover scholarly networks [24]. These tools demonstrate the potential of metadata analysis and visualization, but they often focus on specific metrics or datasets, whereas our approach aims to provide a perspective that combines multiple diverse data sources.

There have been previous interactive, web-based approaches to presenting data that help make analyzing research output more easy to visualize and accessible like the Global Research Activity Map (GRAM) [25]. Using data from Google Scholar profiles, GRAM creates a weighted topic graph to explore research topics and their co-occurrences. Being browser based enables it to offer interactive features like semantic zooming and map overlays, allowing users to analyze research and scholarly output at various levels, from local to global. However, due to its broad scope and reliance on externally managed data, it faces challenges with metadata format inconsistencies and the potential consumption of inaccurate information.

DATA

The datasets used in this analysis provide a comprehensive view of the university's research ecosystem, encompassing data on faculty, researchers, patents, student academic output, and grants. Each dataset is meticulously sourced from reliable systems, including the university's HR system, Tech Launch Arizona, the campus library repository, and external funding agencies. By integrating these diverse data streams, the system ensures that the information remains current, accurate, and reflective of the dynamic research environment. These datasets collectively enable a deeper understanding of the university's intellectual contributions, ongoing research activities, and potential areas of future growth. The details of the datasets are outlined below:

- **Current Faculty and Researchers**: This dataset is automatically sourced from the university's HR system to include only active faculty members and researchers. It ensures that analyses reflect the current research community by dynamically updating as new researchers join or existing ones leave the university. The dataset contains information such as researcher names, titles, departments, and statuses.
- **Patents**: Managed by **Tech Launch Arizona**, this dataset provides up-to-date information on the university's patents and technologies. The system integrates with their database to maintain current records. It includes details such as patent titles, brief descriptions, inventors/authors, and publication years.
- **Student Theses and Dissertations**: Data on student theses and dissertations is sourced from the university's campus repository, managed by the University Library. This dataset includes the titles of works, degree levels (e.g., Ph.D., Master's, or Bachelor's), supervisors, abstracts, and other related information.
- **Grant Abstracts**: Grant abstract data is collected from various external sources such as **grants.gov** and **NIH Reporter (https://reporter.nih.gov/)**, along with internal sources like the university research office. This ensures a comprehensive view of funded research activities.
- **ORCID Database**: ORCID identifiers are collected manually, via user input, and through the university's **ORCID@Arizona system** (https://orcid.arizona.edu/). The system connects with the ORCID API to pull detailed researcher information, including publications, projects, and their associated metadata.

- **Researcher Biographies**: Biographical data about researchers is gathered from the faculty reporting system and through manual updates. This dataset provides an overview of each researcher's background and expertise.
- **Publications**: Publication data is sourced from multiple platforms, including **Google Scholar**, ORCID, and the faculty reporting system. The dataset captures publication type (e.g., journal articles, conference papers, books), titles, abstracts, publishing venues, and other metadata.
- **Proposal Details**: The system analyzes **on-development proposals** to identify researchers' future interests. This dataset includes details such as the principal investigator (PI) and co-PI information, project titles, and brief descriptions or summaries.

The KMap system automatically collects data from various sources in diverse formats through an advanced data collection pipeline. This process involves periodically connecting with external RDBMS systems, making API calls, performing HTML scraping, and processing data from file repositories. The pipeline is capable of handling multiple data formats, including JSON, XML, text files, HTML, CSV, PDF, DOC, and XLS, ensuring seamless integration and efficient processing of diverse datasets.

PROCESSING

In the data section, we discussed the variety of unstructured information sources utilized in this study. In this section, we aim to explain how we processed and extracted meaningful insights from unstructured data. The process involved several major steps, as outlined below:

A. Extracting the Collaboration Network

To create a comprehensive collaboration network, we utilized data from publications, proposals, grants, patents, and project collaborations. Each dataset was processed to identify who collaborated with whom, along with details such as the timing and nature of the collaboration. We then filtered this information to include only researchers actively affiliated with the university. This ensured that the analysis excluded data from researchers who had left the institution. Conversely, the processing system automatically included newly hired faculty and research staff using the HR database.

During network generation, we built a network where nodes represent researchers from the University of Arizona, adhering to specific inclusion and exclusion criteria:

- Researchers must be full-time employees of the university.

- Faculty members are automatically included.

- Staff members are included only if they are associated with an external research grant reported in the university's research reporting systems.

- Researchers who do not meet these criteria can request inclusion.

Edges between nodes represent collaborations, such as co-authorships in publications, joint grants, or shared patents. The research data includes fields such as "Title," "Year,"

and “KMapId” (a unique user identifier). A connection is established between researchers if they are found to have participated in a research activity. To focus solely on research-related collaborations, HR-based connections, such as supervisor-supervisee relationships, are excluded.

Using these nodes and edges, we constructed an unweighted, undirected network of the university’s research collaborations with the NetworkX library. The initial network contained 6,186 nodes and 49,289 edges, making it too large for meaningful analysis. To examine the core collaboration network, we removed all nodes and edges not connected to the largest connected component, resulting in a single connected network spanning the university. This refined network was visualized using the sfdp [26] layout algorithm, with nodes colored red for faculty and green for other researchers. The resulting graph serves as a snapshot of the University of Arizona’s collaboration network and forms the basis for subsequent network-related analyses.

B. Map Building Process

Although the detailed map-building process is beyond the scope of this paper, we provide a brief overview. After creating the university-wide collaboration network, we enriched it with additional information for each node, such as department, college, and basic metadata. We computed inter-departmental collaboration and used graph embedding algorithms to optimize department positions. Each department was represented as a polygon, with the polygon size proportional to the number of members in the department. We then optimized the geometric positioning of researchers within each department.

An interactive application was developed using the Mapbox framework to allow users to zoom in and out and explore the network. The technical details of the geometric algorithms, infrastructure, and technology used in the map are omitted from this paper.

C. Identifying Research Areas for Activities

We utilized a large language model to classify research activities and associate them with specific research areas. This analysis helped us understand the university’s research focus. To achieve this, we employed topic modeling techniques inspired by prior work [27].

D. Research Areas for Individual Researchers

We determined each researcher’s area of interest by analyzing all their research activities. This information is integrated into the KMap search engine (not described in this paper), enabling users to find researchers in specific research areas and display these details on each researcher’s profile.

E. Aggregating Research Activities

To analyze the level of activity at the researcher and department levels, we aggregated various metrics. Examples include: Total grant dollars received by an individual. The primary

funding sources for a department. Funding distribution by funding agency. These aggregated insights are displayed as overlays on the interactive map, providing a comprehensive view of research activity at multiple levels.

RESULTS

Internal Collaboration

We have created a map embedding of the collaboration network by clustering researchers based on their associated departments. This map provides high-level information about the organization. A full description of the map-building process is beyond the scope of this paper. Instead, we offer a brief overview of how we visualize the broader picture of research activities at the University of Arizona.

The base map was created by analyzing collaboration patterns and departmental affiliations. For each department, we generated a custom polygon (referred to as a "city" on the map) and optimized the positions of all researchers within the corresponding polygon. The positions of departments on the map were further optimized based on their collaboration with other departments.

This base map supports interactive navigation features, such as zooming and panning, similar to Google Maps. At the top level, the map displays departments, and as users zoom in, individual researchers become visible. Additional data associated with individual researchers, such as awarded grants, can also be visualized. For instance, if we compute the total awarded grants and represent them as circles, the map would illustrate grant activity across the campus.

For a full interactive experience, readers are encouraged to explore our live system at https://kmap.arizona.edu/.

Fig. **??** shows a few screenshots of the system with overlay information. In base-map, all departments are visible user can zoom in to see researchers in the department (in polygon). The next three pictures shows grants, patents, and H-index (impacts) of overall university.

In Fig. 2, each dot represents a researcher at the University of Arizona. Red dots signify faculty members, while green dots indicate researchers without faculty titles. All three visualizations depict the same number of researchers and faculty members, whom we identified as being actively engaged in research. In each case, we observe large, interconnected components, highlighting a high degree of collaboration across the university. The isolated nodes on the periphery correspond to researchers who are either new, not yet integrated into the core collaborative network, or whose collaborators are outside the organization. Additionally, we see several smaller clusters, representing groups of researchers focused on specific projects or activities. This is interesting to see university of Arizona's largest collaboration comes from proposal development activities. The Fig. 2b shows the overall collaboration for proposal development activities.

Next we analyze the change of the research collaboration on campus over the past 10 years by creating snapshots of the network for each year. This allows us to both visualize the change in the maximum connected components over the years and analyze graph statistics to create a network wide understanding of how collaboration has changed over time.

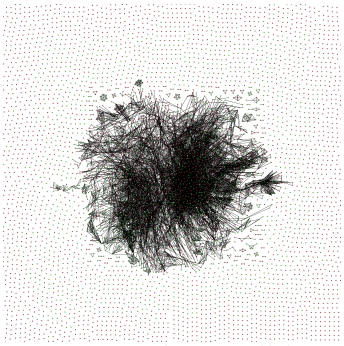
(a) Base Map (b) Grant Dollars

(c) Patents (d) H-index

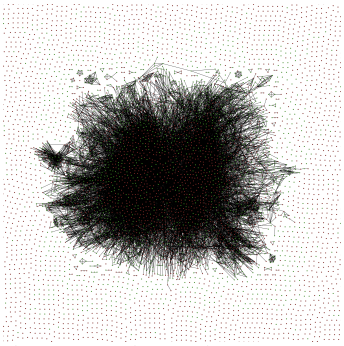
Fig. 1: Screenshots from the University of Arizona Knowledge Map system illustrating organization-wide research activity. (a) The base map, where each polygon represents a university department. The lower-left half includes many medical-related departments, while the upper-right half includes engineering, arts, law, etc. (b) Overlay showing grant dollars allocated to departments. (c) Overlay showing the number of patents filed by departments. (d) Overlay showing the H-index of departments, indicating research impact.

Starting with the base network created, we go back every year to filter nodes and edges that did not exist at this point in time. For each year's snapshot, we isolate the maximum connected component to exclude any nodes or edges that are no longer part of the central network. To ensure consistency and allow a direct comparison, we retain the same positions for all remaining nodes and edges in the graph across all years. This approach allows us to clearly visualize and analyze how the network's structure has evolved over time. An important note in this analysis is researchers who are no longer employed at the university are excluded from all graphs, even for years when they were previously employed. For example, a faculty member employed between 2015 and 2020 will not appear in any graphs, including those for 2015–2020, as they are not currently part of the institution.

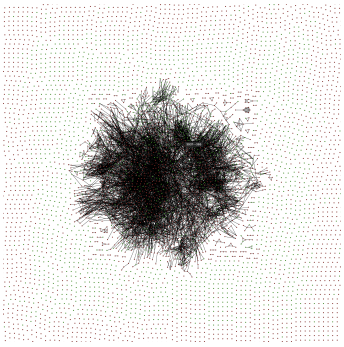
Fig. 2: Grants, proposals, and publication collaboration activities at the University of Arizona.



(a) Grant Projects Collaboration



(b) Proposal Development Activities



(c) Publications Collaboration

In addition to constructing these yearly networks, we calculate and store network statistics for each snapshot, including the number of nodes, number of edges, diameter, and density. We then plot some of these statistics over time to visualize trends, such as the growth or overall shifts in the network.

The collaboration graphs for 2015, 2020, and 2024 reveal a significant change in the university's research network including both a dramatic increase in nodes and edges. In 2015, the network was relatively sparse, consisting of 1,280 nodes and 5,255 edges. By 2020, the network grew substantially, almost doubling the number of nodes and increasing the number of edges by nearly 500%. By 2024 this growth continued by adding over 1,500 researchers and 23,000 connections. It's worth mentioning that this analysis includes only current researchers. That is, many researchers who joined the university and later left are not taken into account. This trend highlights the number of active collaborations that have formed in recent years and that the vast majority of collaborations on campus are relatively recent. While this graph does not provide a complete representation of past collaboration at that point in time, it is likely that collaborative activities have increased over time; it is expected that that change is not as extreme as depicted in the networks. These inferences lead to the idea that the university's collaboration network is rapidly evolving, with newer, more active collaborations emerging to replace researchers who have left the institution.

Dividing the university-wide network graph into two distinct groups: faculty and non-faculty highlights distinct differences in collaboration patterns across campus. Primarily, the faculty network has a larger size and higher connection frequency, with an average of 5.46 connections per node compared to 3.31 in the non-faculty network. Additionally, the faculty network contains a large, densely populated cluster at its center, with highly connected nodes. In contrast, the non-faculty network has a more sparsely populated central network, containing smaller pockets of highly connected nodes, indicating frequent collaboration in isolated settings.

The comparison shows that the faculty network comprises the majority of both nodes and edges in the graph, as well as the overall structure and density of the faculty network.

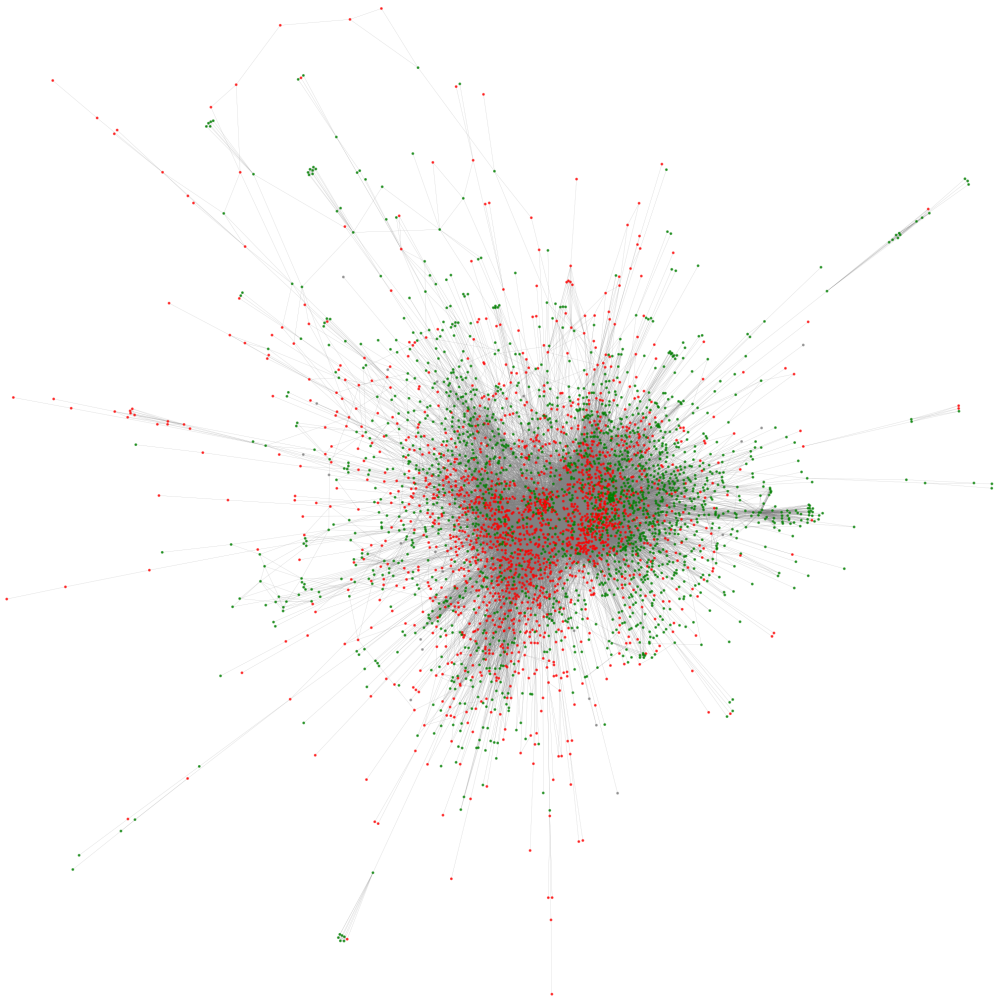


Fig. 3: The largest connected component in 2024, excluding small groups and isolated researchers, consists of 3,987 researchers with 49,160 connections. Red dots represent faculty members, while green dots represent staff. This network forms the core collaboration hub of currently active researchers at the University of Arizona.

F. Intra-departmental collaboration

The goal of this section is to visualize collaborations between departments within the university, providing insights into how departments interact and work together.

To generate this visualization, the top 100 departments are selected based on their total node counts, calculated by aggregating and grouping the number of researchers associated with each department stored within their node metadata. The researcher-to-researcher connections are then mapped to department-to-department connections using the same metadata. Duplicate connections are removed, for example, if a researcher collaborates with multiple researchers in another department, it is recorded as a single connection between the two departments.

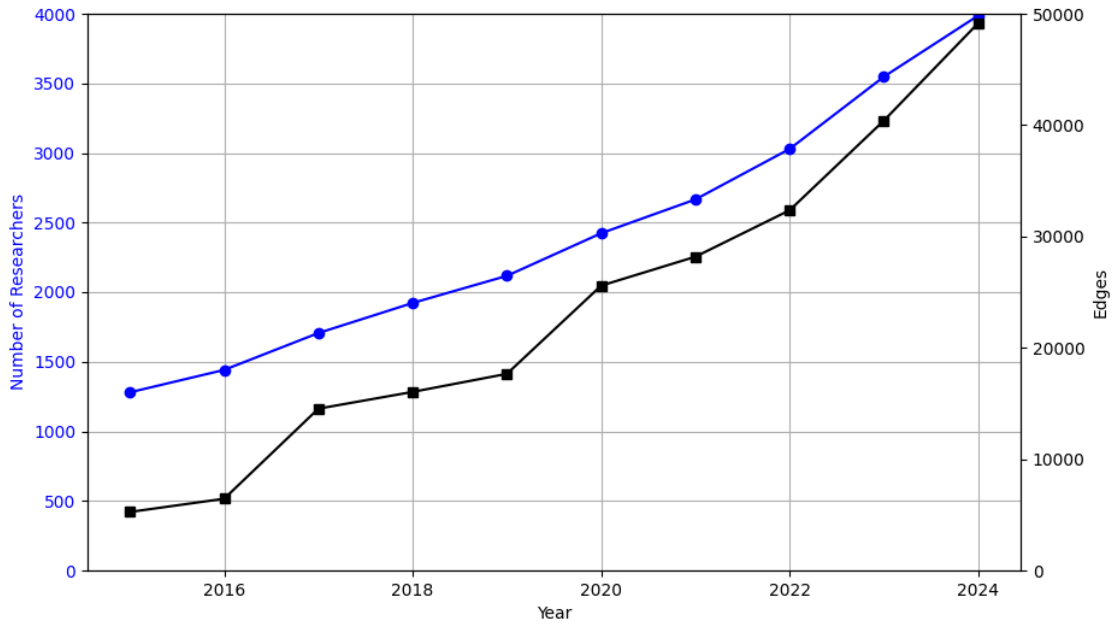


Fig. 4: The chart represents the cumulative number of new researchers who joined the core collaboration network and the cumulative number of connections (edges) formed in the network over the last 10 years.

This processing results in a network of the top 100 departments and the collaboration network between them.
The filtered data is then used to create a normalized connection weight for each pair of departments. This weight represents the strength of their collaboration as a percentage of their total nodes.

$$\text{Normalized Weight} = \left(\frac{\text{Number of all connections between two departments}}{\text{Total number of researchers in two departments}} \right) \times 100$$

Organizing these weighted connections into an agency matrix we then apply the Louvain community detection algorithm to cluster closely connected departments together. We arrange the departments by cluster and create a final matrix to generate a visual heat map of department collaboration.

The visualization in Fig. 7, we see a clear line down the middle of the graph indicating that the strongest department collaborations happen internally. Additionally we see clear borders around our clustering indicating that our clusters have worked and show that it is often a group of departments collaborating heavily between each other much less across other departments.

The boxes outlined in the visualization highlight a few clear clusters along with additional insights. By examining boxes F, A, and B, we can identify distinct clusters of connected departments on campus. Additionally, boxes C, D, and E reveal sub-clusters of highly connected departments located close together within an already defined group. Finally, box G shows that the Department of Biostatistics has a strong connection with many colleges across campus and is highly integrated within its clusters, even more so than its nearby departments.

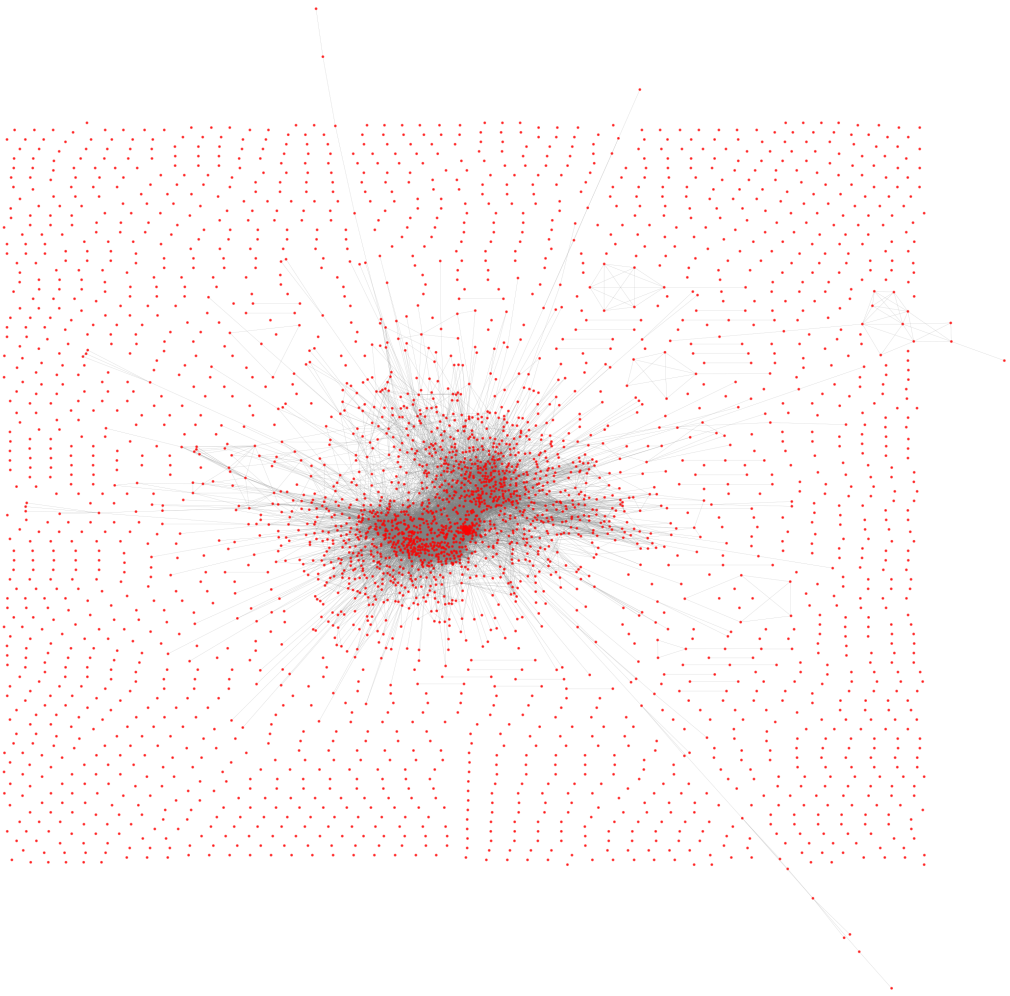


Fig. 5: The collaboration map of all university faculty members in 2024 including nodes that are unconnected. The network consists of 3,677 researchers with 20,069 connections between them.

G. Collaboration with college vs outside of collage

The goal of this section is to visualize internal and external college collaboration within the university in order to understand the college level collaboration across campus.

To generate the college-level collaboration visualization, researchers are first mapped to their home colleges using metadata stored with their nodes. Researcher-to-researcher connections are then used to create college-to-college connections using mapping. Duplicate researchers to college connections are then dropped. For instance, if multiple researchers from one college collaborate with researchers from another college, it is recorded as a single connection between those colleges.
A normalized connection weight is then calculated for each pair of colleges to calculate a

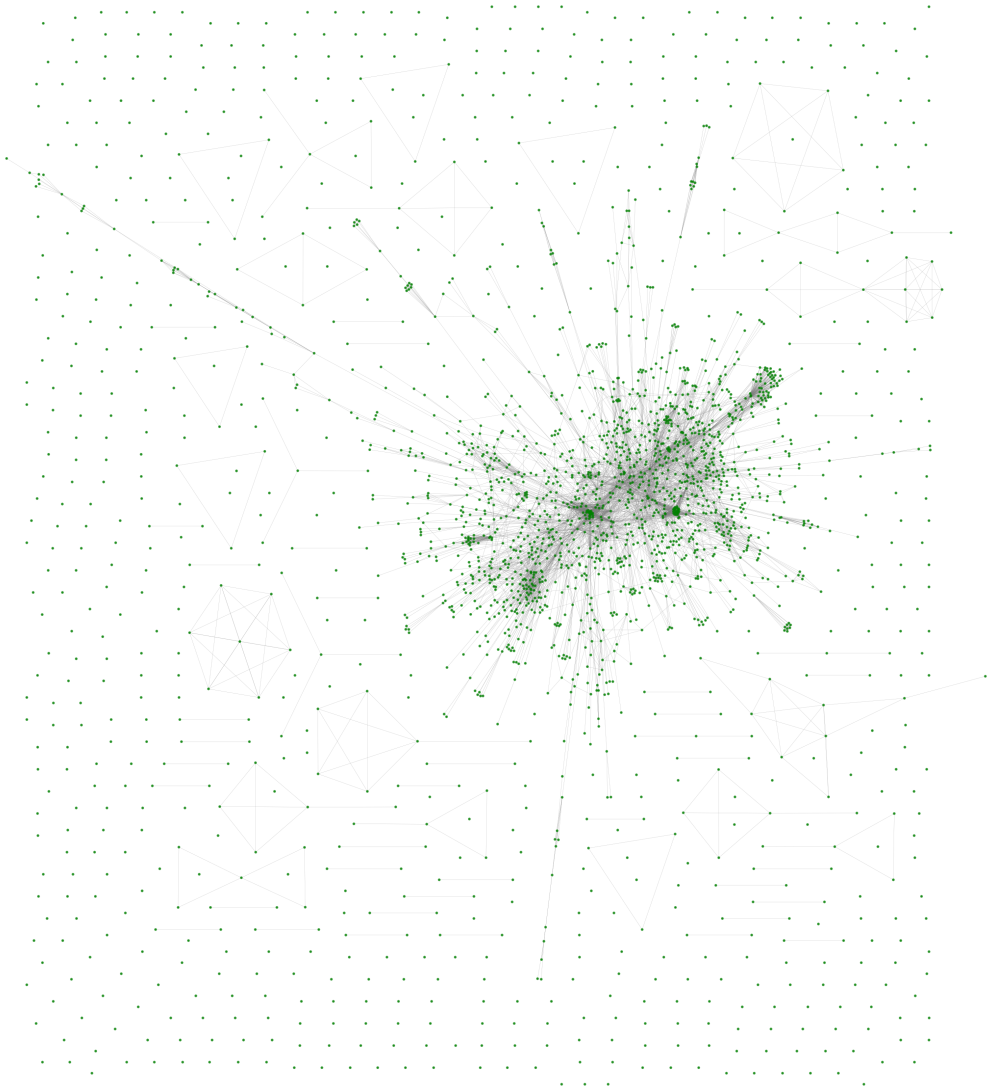


Fig. 6: The collaboration map of all university non-faculty members in 2024 including nodes that are unconnected. The network consists of 2,509 researchers with 8,301 connections between them.

weighted strength of their collaboration. The weight is calculated as a percentage of the total researchers in the destination college. We then take the top 10% of weighted connections to focus in on strong collaboration and filter out weaker connection. Internal connections, where the source and destination are the same college, are then pulled into a separate piece of data where these internal connection weights are visualized as a sorted bar chart, highlighting the relative collaboration within each college.
External collaborations are isolated to focus on relationships between different colleges. A Sankey diagram is created to visualize these college collaborations, with each college represented as both a source and a destination. The width of each link in the diagram reflects the weighted connection weight to represent the difference in connection strength.

The analysis reveals intriguing insights into how colleges collaborate externally. Some

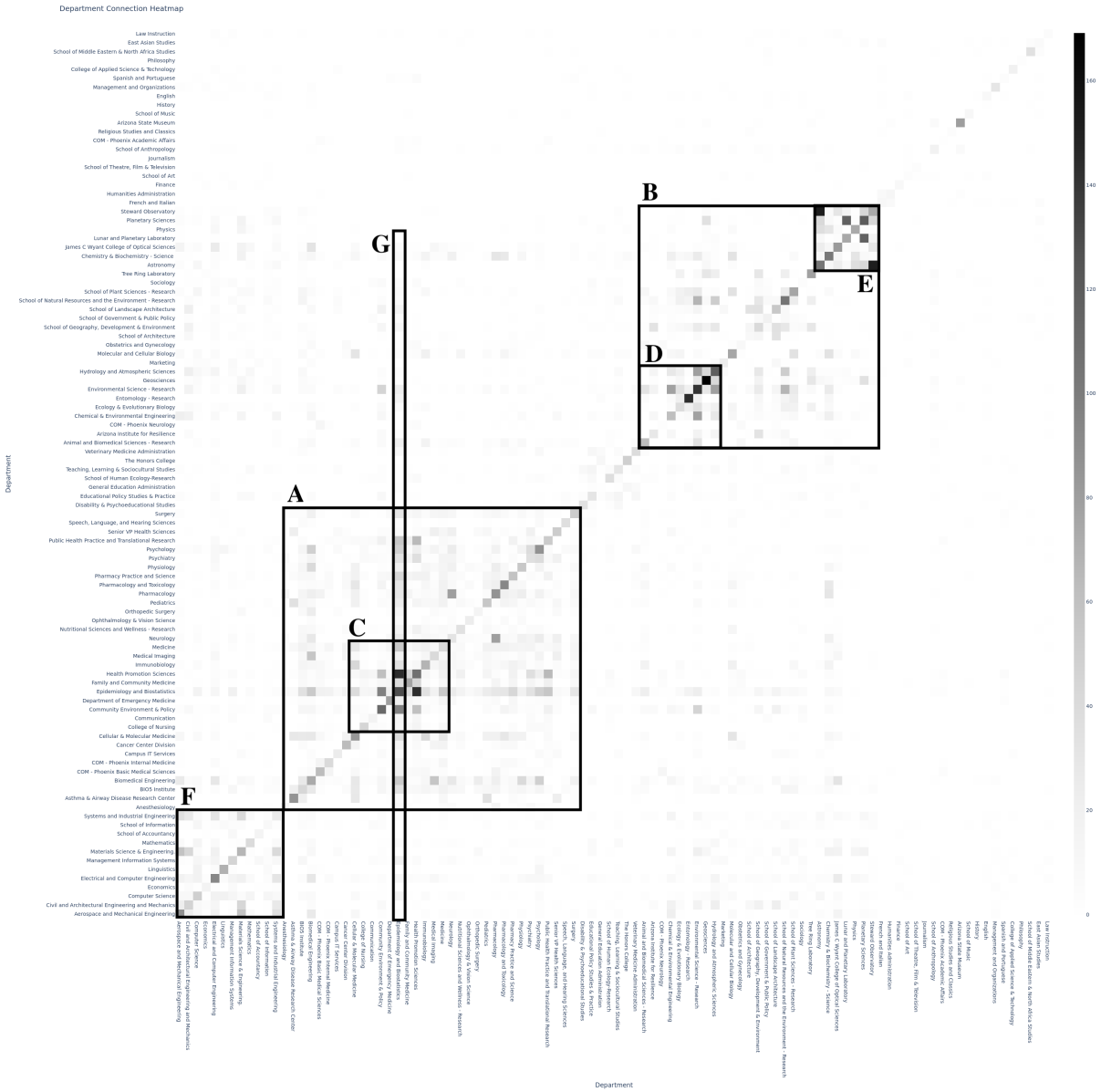


Fig. 7: Intra-departmental collaboration: The diagonal line represents collaboration within departments, normalized by the number of researchers in each department. Area A in the figure includes departments related to medical and medicine fields, while Area B includes two main groups of departments: i) environment-related, and ii) physics, chemistry related departments. Area F includes Mathematics, Computer Science and other Engineering

colleges, such as the College of Law, exhibit a broad network of collaborations, indicating partnerships with a wide range of colleges across campus. In contrast, other colleges demonstrate a narrower focus with fewer connections. We can also see that some colleges show strong connections with others in similar or related fields, highlighting the influence of disciplinary alignment on collaborative patterns.

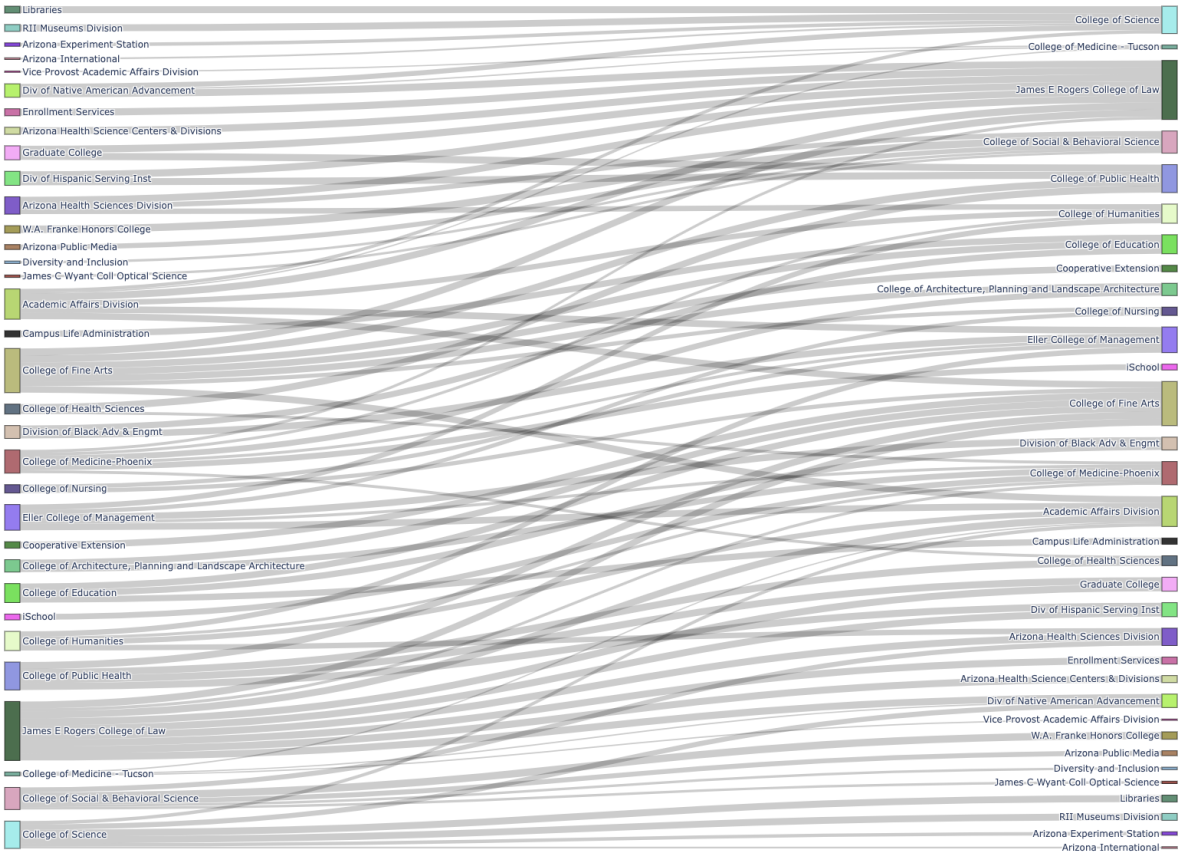


Fig. 8: Top 10% strongest research connections between colleges.

H. Research Strengths

The goal of this section is to use topic prediction on grant and publication data to understand research strengths within the university. This can provide valuable insights into the university's areas of strength while also highlighting potential opportunities for increased research focus in underrepresented fields.

In our topic prediction process, we identify all unique grants and publications at the university. Each title is then processed using OpenAI API, which is prompted to generate a list of five topics associated with the grant or publication. From this mapping unique research works and topics we group by topics and count the number of unique research works for each topic. Topics with fewer than 100 occurrences are excluded and for the remaining topics with more than 100 occurrences, we calculate the unique number of researchers involved in any grants or publications associated with those topics. For instance, if a researcher is found on multiple research works in a topic they are only counted once. We then create a scatter plot to visualize the relationship between the number of unique research works and the number of unique researchers for each topic. The higher resolution of the figures 11 and 12 can be found https://kmap.arizona.edu/kmap-topic-report/index.html

This plot shows the relationship between the number of unique researchers and the total number of works across various research topics. It highlights how certain topics, such as education, attract significant research activity, with 845 researchers contributing to a total of 1,661 research works. We can also observe topics like particle physics, where a small

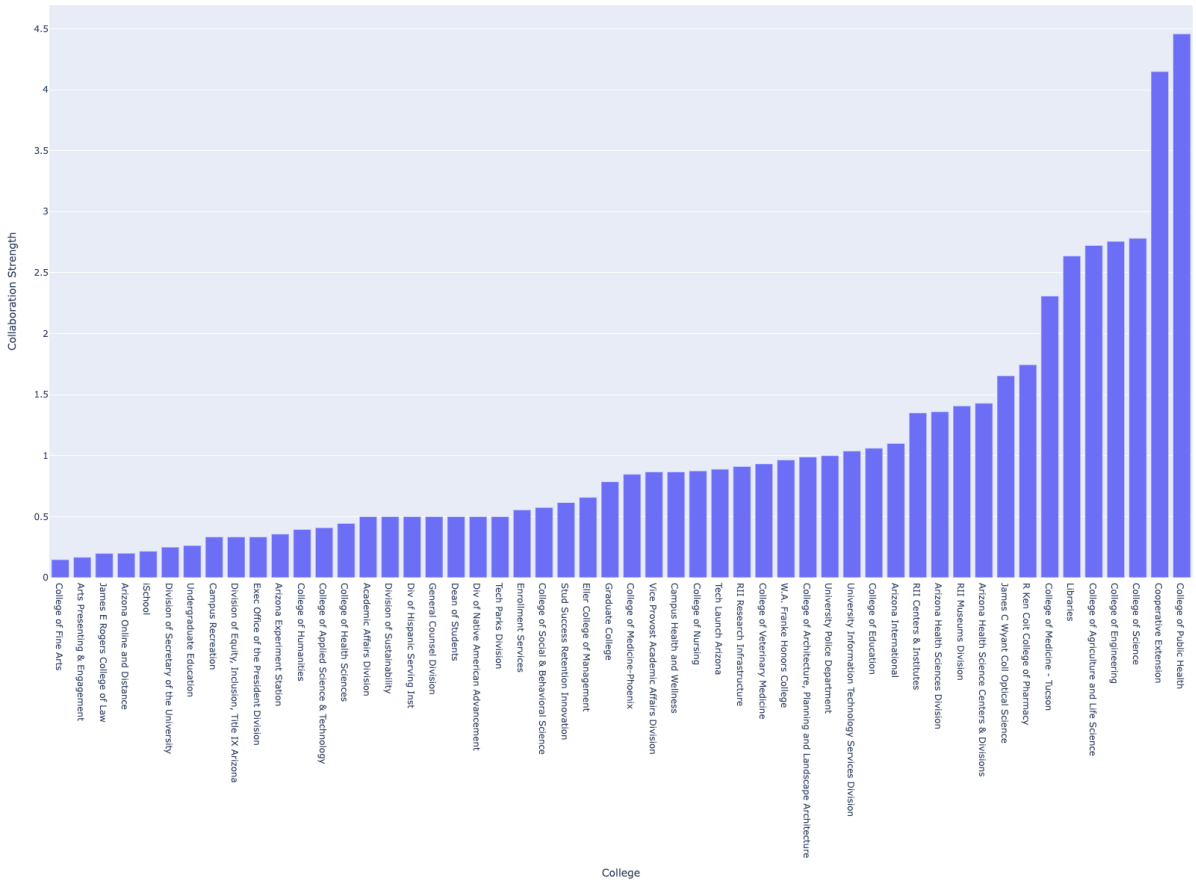


Fig. 9: Average collaboration within the college by excluding the researchers who don't have any connection at all

number of researchers produce a disproportionately high volume of research work. This suggests that while the topic is highly specialized and explored by relatively few researchers these researchers are very productive.

Figure 12 shows that medicine leads with a significant rise in awards, surpassing 300 by 2022, while education and social sciences also demonstrate steady growth. Emerging fields like data science and climate change show gradual increases, reflecting their growing importance. Meanwhile, topics like physics and space maintain stable levels of awards, indicating consistent but a lesser amount of grants awarded. These trends highlight medicine and education as key focus areas, with opportunities to further support emerging disciplines like data science.

LIMITATIONS

In this section, we discuss several limitations of this work:

a) While analyzing the core network, we only considered current researchers. However, in some cases, researchers joined the university and later left within the last 10 years. Since these data are not properly tracked, such transitions are not accounted for in the analysis.

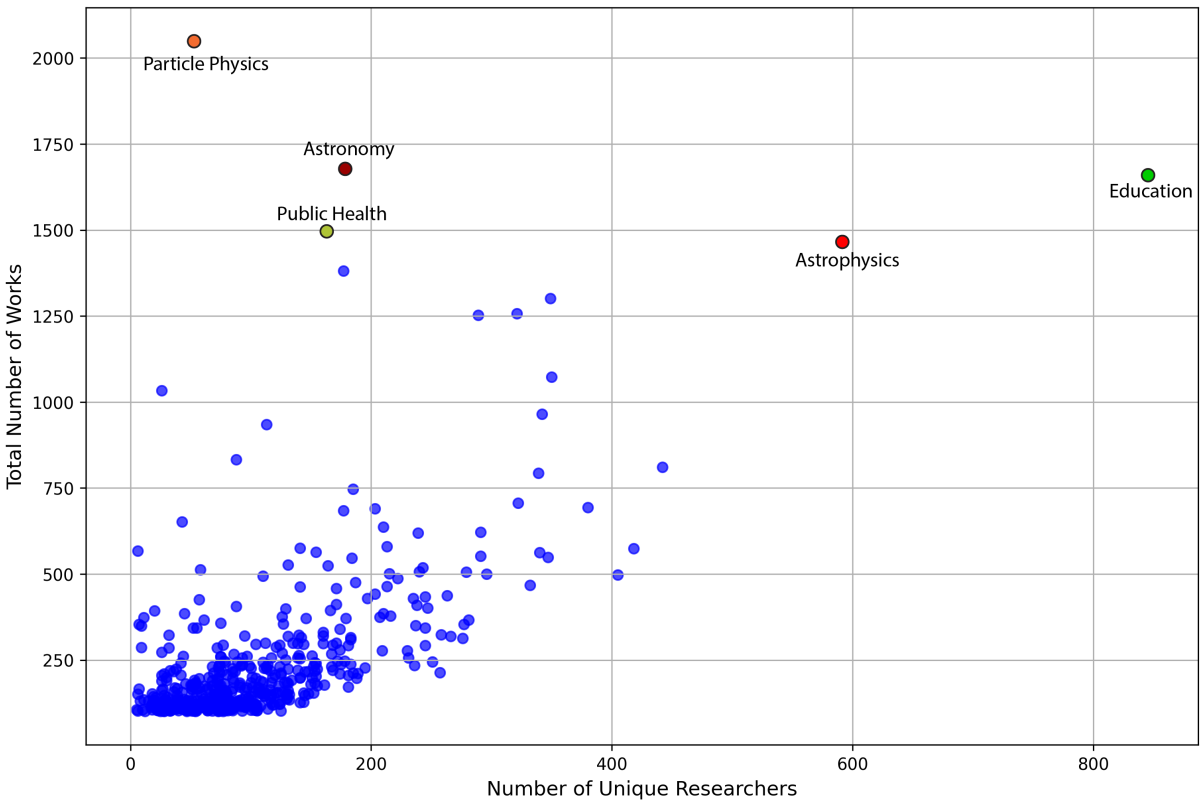


Fig. 10: Chart highlighting key research strengths across different areas.

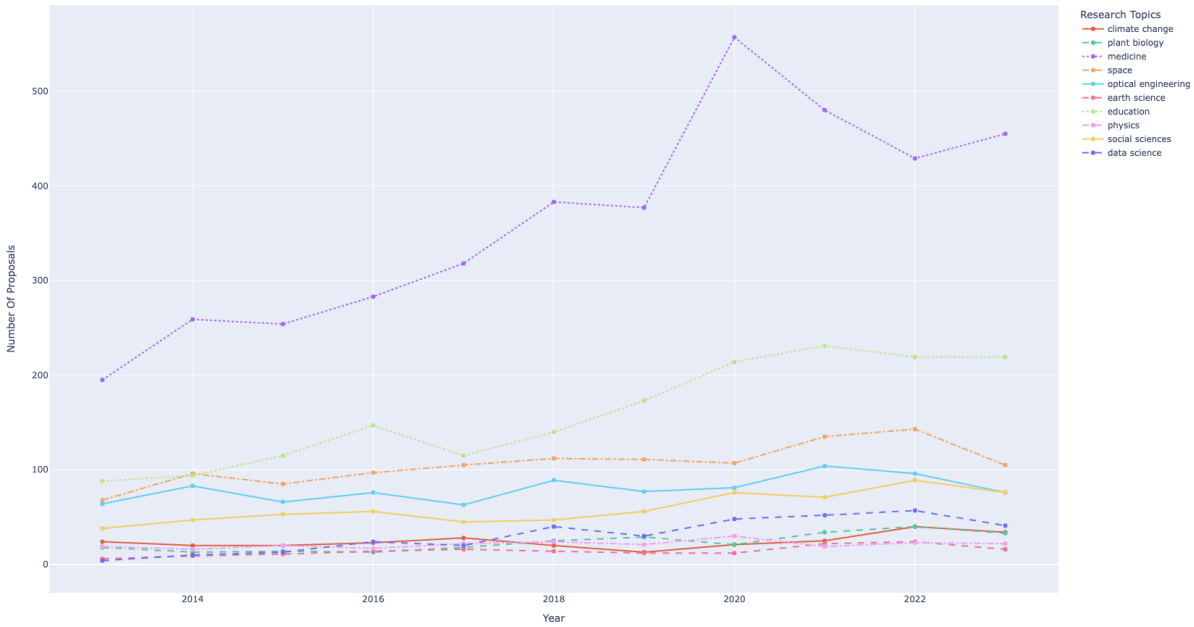


Fig. 11: Proposal writing activities per research area over the last 10 years.

b) When deriving statistics on proposal writing activity across different research areas, we relied on a keyword-based search approach. Some relevant keywords might have been missed, leading to incomplete results.

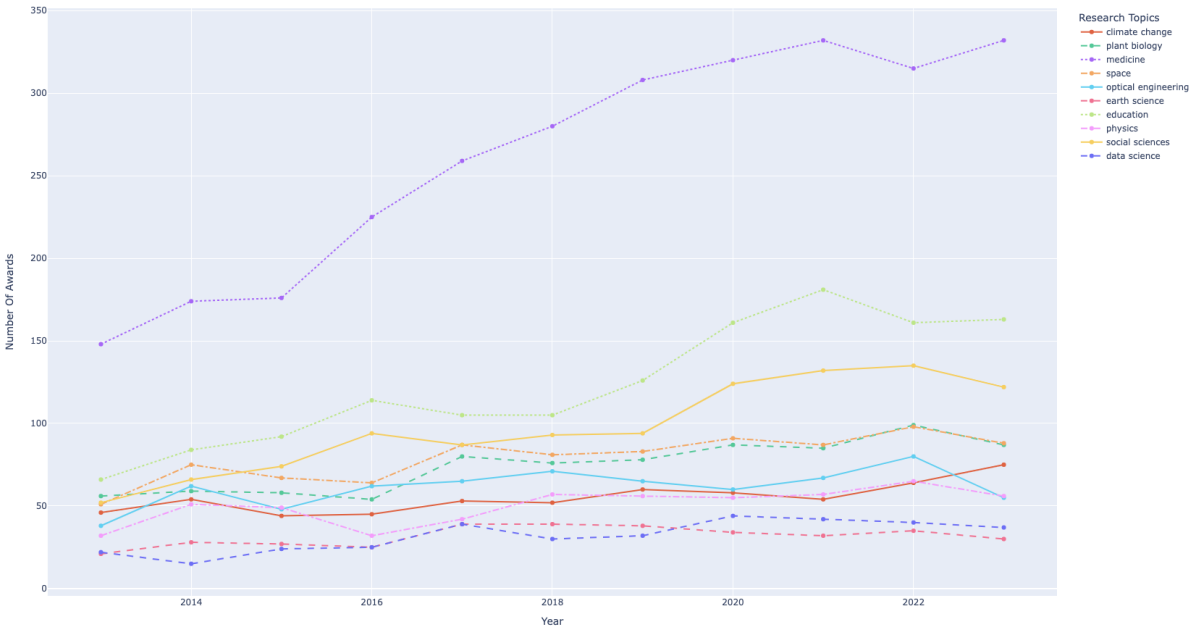


Fig. 12: Number of grant awards per research area over last 10 years

c) Although we utilized various data sources, there are instances where researchers do not maintain profiles on platforms like Google Scholar, faculty reporting systems, or ORCID. Consequently, some research activities may be missing from the analysis.

d) Research topic networks and neighborhoods are inherently complex [27]. For certain analyses, we used large language models (LLMs) to extract relevant research areas. Although these models are advanced, their individual outputs may not always be entirely accurate. However, because these outputs are aggregated, the impact of any individual mistake is minimized when considering the overall picture.

e) The researcher count is determined by the discovery of their research work. There may be cases where researchers were excluded simply because no research work of theirs was discovered in the data collection process.

f) Students are excluded in this analysis.

CONCLUSION

In this paper, we analyzed various research metadata from the University of Arizona's internal and external sources. The goal of the analysis was to understand the university's internal collaborations, research activities, and strengths. This analysis, spanning multiple dimensions of research, provides a clear representation of the internal research culture. The key takeaways from the analysis are as follows:

- As an R1 research university, a significant number of participants in research activities include staff in addition to research faculty.
- The university engages in extensive proposal-writing activities, which play a crucial role in connecting different parts of the institution.

- Over the last 10 years, collaborations have been consistent and robust across various research domains.
- Although the entire university forms a giant interconnected network, there are three main clustered areas of focus: medical sciences, engineering, and physical sciences.
- Certain colleges, such as the College of Law, make significant contributions to the research ecosystem, with connections spanning across the entire campus.
- Based on our findings, the university demonstrates considerable strength in research areas such as climate change, medicine, space science, optical sciences, social sciences, data science, and plant biology.

REFERENCES

[1] S. Fathalla, S. Vahdati, S. Auer, and C. Lange, "Metadata analysis of scholarly events of computer science, physics, engineering, and mathematics," in *Digital Libraries for Open Knowledge*, E. Méndez, F. Crestani, C. Ribeiro, G. David, and J. C. Lopes, Eds. Cham: Springer International Publishing, 2018, pp. 116–128.

[2] E. Dagienė, "Mapping scholarly books: library metadata and research assessment," *Scientometrics*, vol. 129, no. 9, pp. 5689–5714, 2024. [Online]. Available: https://doi.org/10.1007/s11192-024-05120-1

[3] A. Mierzecka, "The role of academic libraries in scholarly communication. a meta-analysis of research," *Studia Medioznawcze*, vol. 19, pp. 42–55, 05 2019.

[4] T. Lazaridis, "Ranking university departments using the mean h-index," *Scientometrics*, vol. 82, no. 2, pp. 211–216, Feb 2010. [Online]. Available: https://doi.org/10.1007/s11192-009-0048-4

[5] L. Waltman and N. J. van Eck, "The inconsistency of the h-index," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 2, pp. 406–415, 2012. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.21678

[6] L. Bornmann, C. Ganser, and A. Tekles, "Simulation of the h index use at university departments within the bibliometrics-based heuristics framework: Can the indicator be used to compare individual researchers?" *Journal of Informetrics*, vol. 16, no. 1, p. 101237, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1751157721001085

[7] R. Landry, N. Traore, and B. Godin, "An econometric analysis of the effect of collaboration on academic research productivity," *Higher Education*, vol. 32, no. 3, pp. 283–301, Oct 1996. [Online]. Available: https://doi.org/10.1007/BF00138868

[8] M. Smith, Y. Sarabi, and D. Christopoulos, "Understanding collaboration patterns on funded research projects: A network analysis," *Network Science*, vol. 11, no. 1, p. 143–173, 2023.

[9] A. Miller, "A case study in institutional repository content curation," *Digital Library Perspectives*, vol. 33, no. 1, pp. 63–76, Jan 2017. [Online]. Available: https://doi.org/10.1108/DLP-07-2016-0026

[10] D. A. Munoz, J. P. Queupil, and P. Fraser, "Assessing collaboration networks in educational research," *International Journal of Educational Management*, vol. 30, no. 3, pp. 416–436, Jan 2016. [Online]. Available: https://doi.org/10.1108/IJEM-11-2014-0154

[11] Z.-L. He, X.-S. Geng, and C. Campbell-Hunt, "Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a new zealand university," *Research Policy*, vol. 38, no. 2, pp. 306–317, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048733308002813

[12] J. Birnholtz, S. Guha, Y. C. Yuan, G. Gay, and C. Heller, "Cross-campus collaboration: A scientometric and network case study of publication activity across two campuses of a single institution," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 1, pp. 162–172, 2013. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.22807

[13] U. of Texas Arlington, "Fy24 annual sponsored project and research report," 2024. [Online]. Available: https://cdn.web.uta.edu/-/media/project/website/research/_downloads/annual-reports/annual-research-report-2024.ashx

[14] U. of Michigan, "Fy24 research annual report," 2024. [Online]. Available: https://research.umich.edu/news-and-issues/research-annual-reports/fy24-research-annual-report/

[15] O. S. University, "2024 research and innovation annual report," 2024. [Online]. Available: https://research.oregonstate.edu/2024-research-and-innovation-annual-report

[16] D. University, "Fy24 annual sponsored project and research report," 2024. [Online]. Available: https://research.duke.edu/about-ori/ori-annual-report-2023-2024/

[17] A. Slim, J. Kozlick, G. L. Heileman, J. Wigdahl, and C. T. Abdallah, "Network analysis of university courses," in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14 Companion. New York, NY, USA: Association for Computing Machinery, 2014, p. 713–718. [Online]. Available: https://doi.org/10.1145/2567948.2579360

[18] L. Waugh, H. Tarver, M. E. Phillips, and D. G. Alemneh, "Comparison of full-text versus metadata searching in an institutional repository: Case study of the unt scholarly works," https://digital.library.unt.edu, February 2015, accessed January 9, 2025. [Online]. Available: https://digital.library.unt.edu/ark:/67531/metadc725823/

[19] M. A. Tasleem Arif, Rashid Ali, "Scientific co-authorship social networks: A case study of computer science scenario in india," *International Journal of Computer Applications*, vol. 52, no. 12, pp. 38–45, August 2012. [Online]. Available: https://ijcaonline.org/archives/volume52/number12/8257-1790/

[20] J. Fagan, K. S. Eddens, J. Dolly, N. L. Vanderford, H. Weiss, and J. S. Levens, "Assessing research collaboration through co-authorship network analysis," *Journal of Research Administration*, vol. 49, no. 1, pp. 76–99, Spring 2018.

[21] I. T. Afolabi, A. Ayo, and O. A. Odetunmibi, "Academic collaboration recommendation for computer science researchers using social network analysis," *Wireless Personal Communications*, vol. 121, no. 1, pp. 487–501, Nov 2021. [Online]. Available: https://doi.org/10.1007/s11277-021-08646-2

[22] J. Bian, M. Xie, T. J. Hudson, H. Eswaran, M. Brochhausen, J. Hanna, and W. R. Hogan, "Collaborationviz: interactive visual exploration of biomedical research collaboration networks," *PLoS One*, vol. 9, no. 11, p. e111928, Nov 2014.

[23] Y. Yu, Y. Wu, X. Liang, C. Ma, and Q. Lu, "Ncovis: A visual analysis framework for exploring academic collaboration networks under new collaborative relationships," in *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2022, pp. 1203–1208.

[24] S. Vahdati, G. Palma, R. J. Nath, C. Lange, S. Auer, and M.-E. Vidal, "Unveiling scholarly communities over knowledge graphs," in *Digital Libraries for Open Knowledge*, E. Méndez, F. Crestani, C. Ribeiro, G. David, and J. C. Lopes, Eds. Cham: Springer International Publishing, 2018, pp. 103–115.

[25] R. Burd, K. A. Espy, M. I. Hossain, S. Kobourov, N. Merchant, and H. Purchase, "Gram: global research activity map," in *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, ser. AVI '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3206505.3206531

[26] Y. Hu, "Efficient, high-quality force-directed graph drawing," *Mathematica journal*, vol. 10, no. 1, pp. 37–71, 2005.

[27] R. Burd, K. A. Espy, M. I. Hossain, S. Kobourov, N. Merchant, and H. Purchase, "Gram: Global research activity map," in *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, 2018, pp. 1–9.