



Exploring Trends in CMU Grant Data

Kyra Balenzano, Melissa Dy, Michael Li, and Veda Lin

Project Advisor: Peter Freeman Project Supervisors: Huajin Wang and Sarah Young

Carnegie Mellon University
Libraries

Background

- Professors and researchers often seek the help of the Carnegie Mellon University Libraries to acquire information about the university’s research grants and research collaborations.
- Current methods to answer these inquiries require large amounts of manual labor and analysis, as there are no tools at the CMU Libraries’ disposal to process and glean insights from the research grant data they have.
- The goal of our research is to create such a tool—specifically, an R Shiny app—that can visualize and analyze the data in many ways, including those that address these questions:
 - What funding agencies have funded research at CMU and how has this changed over time? How is this related to fields/categories of research?
 - Which authors are associated with a certain field/category of research?
 - How do funded research topics change over time?

By addressing these questions, the CMU Libraries will be better equipped to answer grant-related questions, therefore empowering university’s active research community.

Data Description

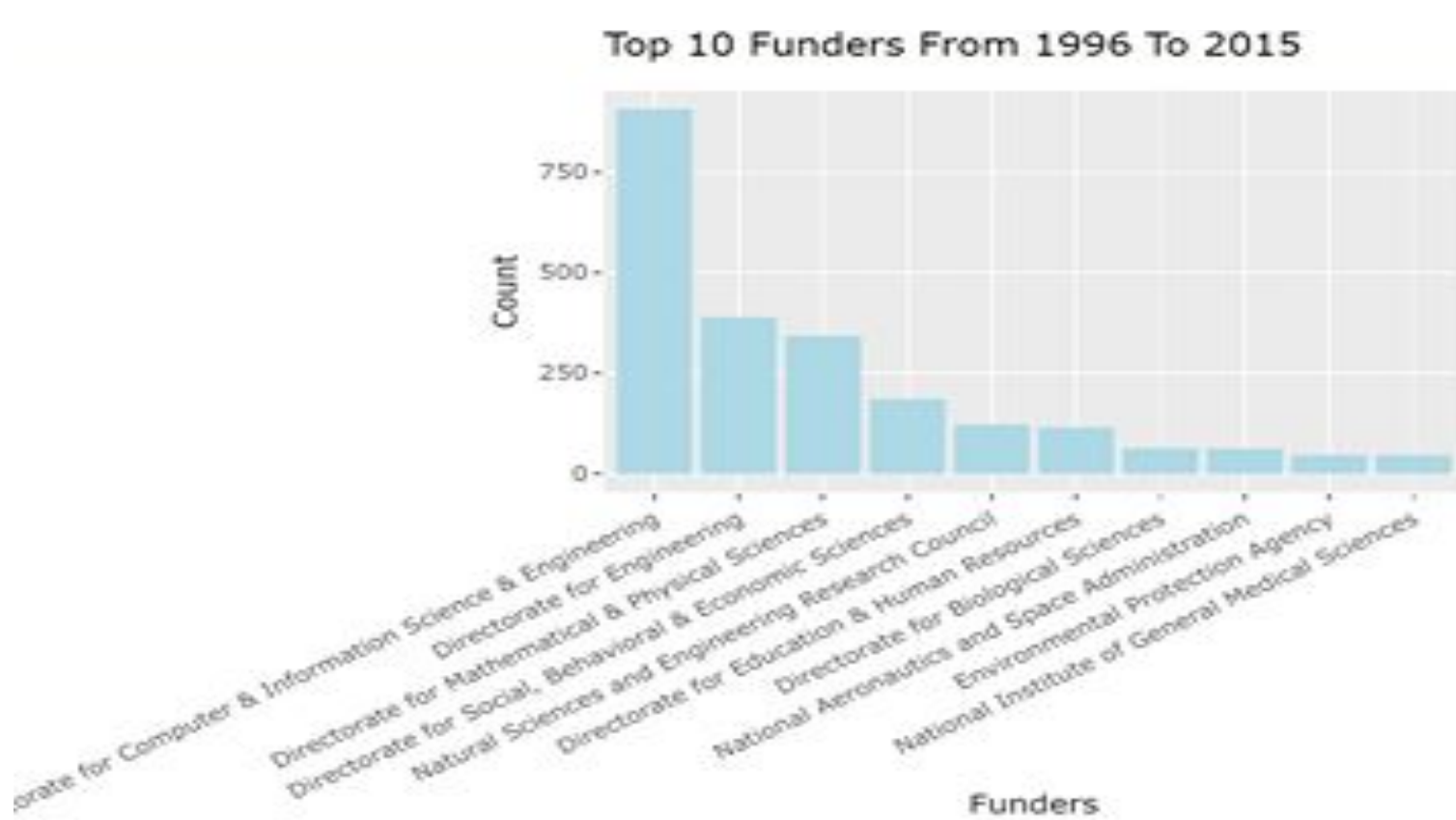
Our dataset, pulled from Dimensions AI, a research grants database that links grants to publications, consists of 6,128 research grants affiliated with CMU and 28 primary information fields, which can be summarized in three main groups:

General	Funding	Categorizations (System Author)
<ul style="list-style-type: none">Grant ID, NumberTitleAbstractResearchersResearch OrganizationResearch City, State, CountryStart Date, YearEnd Date, YearResulting PublicationsDimensions AI Link	<ul style="list-style-type: none">Funding AmountFunderFunder GroupFunder CountryFunder Grant Link	<ul style="list-style-type: none">Fields of Research Categories (ANZSRC)Research, Condition, and Disease Categories (NIH)Health Categories (HRCS)Health Research Activities (HRCS)Cancer Types (ICRP)Common Scientific Outline Categories (ICRP)Units of Assessment (REF)Sustainable Development Goals (UN)

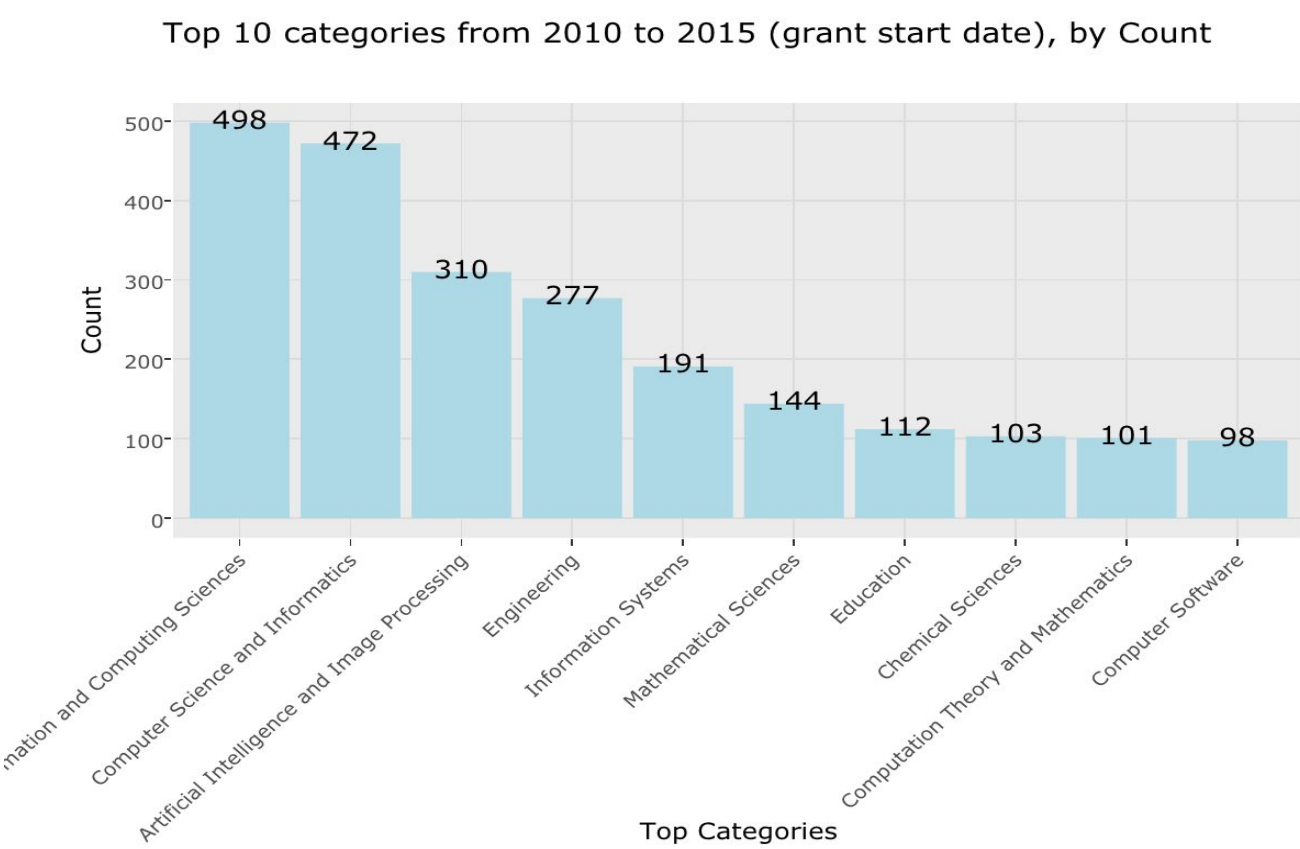
In the pre-processing stage, we combined the categorizations into one column for easier access and created a mapping to combine similar categories.

Categories & Funding

- Our R Shiny app allows users to visualize and analyze trends in categories of research and funding sources for a selected time frame using bar graphs.

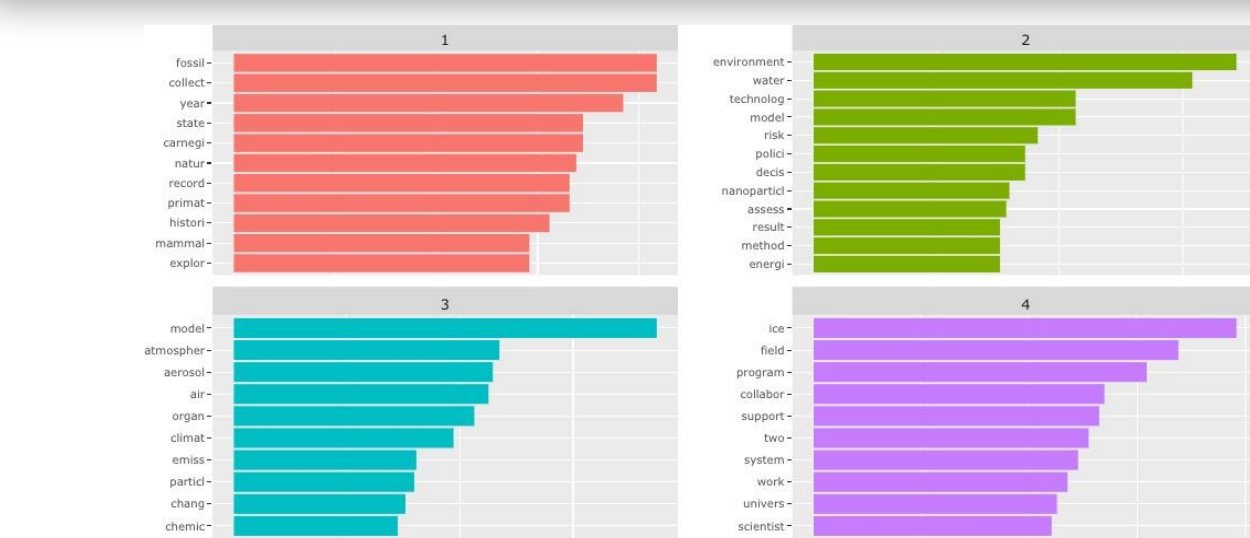


Top funders from 1996 to 2015



Top categories from 2010 to 2015

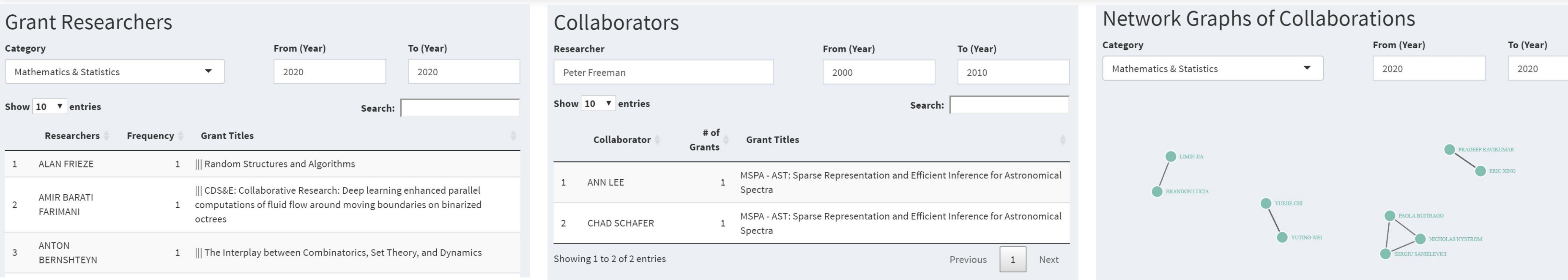
Topic Modeling



LDA with $k=4$ for all grants with an Earth Science categorization. A reasonable interpretation is that topic 1 corresponds with Paleontology, topic 2 with Hydrology, topic 3 with Atmospheric Science and Emissions, and topic 4 with collaborative Glaciology efforts.

- Within a research category, there is no further granularity of research topics.
- Leveraging Natural Language Processing (NLP) techniques on the grant titles and abstracts proves to be promising in identifying these topics.
- Specifically, we use Latent Dirichlet Allocation (LDA) with Gibbs Sampling, an unsupervised learning model, to find topic clusters.
- Users can freely adjust the number of clusters to find the most personally usable results, and the year range to see changes over time. Additionally, details of corresponding grants can be displayed.

Researchers & Collaborations



(From left to right)
List of researchers
in a given category
and time frame

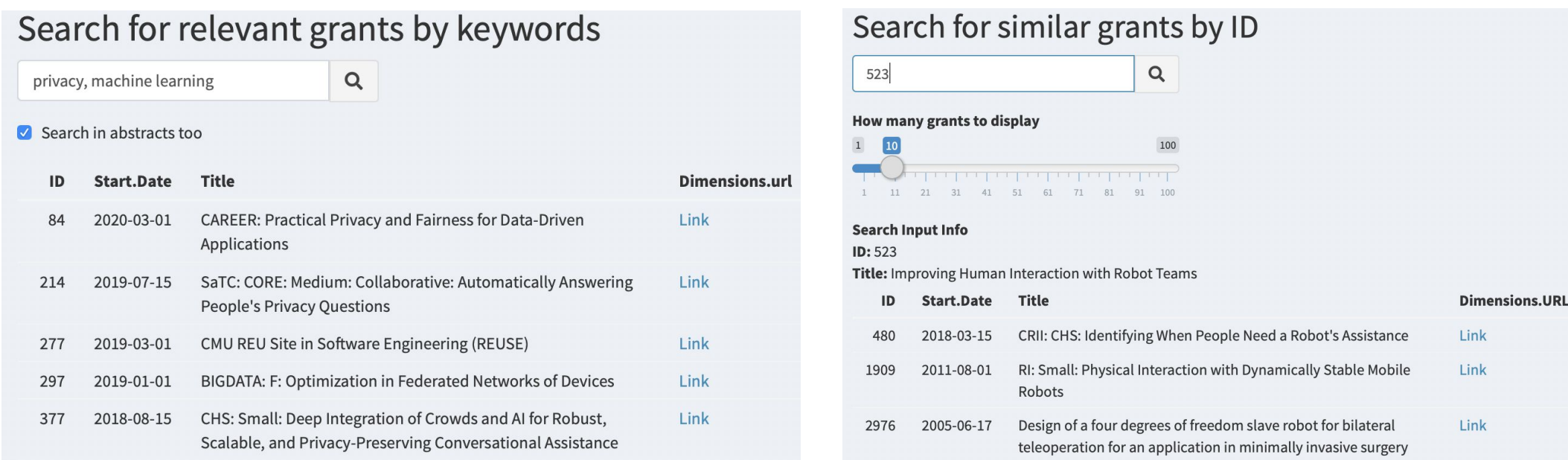
List of collaborators
for a given
researcher in a
selected time frame

Network graph of
collaborations in a
given category and
time frame

- Our R Shiny app also allows users to investigate which researchers are associated with a given field and their collaborations with each other.

Advanced Searching

- Users can easily search for relevant research using keywords and phrases.
- Alternatively, users can provide a specific grant of interest and explore similar works.



Searching for grants by keywords

Searching for related grants

Conclusions & Future Work

- Our app displays a wide variety of insights that can be used by the CMU Libraries to help professors and researchers find the resources they need to succeed.
- About 1,700 grants do not have any category labelings, therefore limiting the functionality of some of our application’s tabs. Initial attempts to categorize these uncategorized grants based on patterns in titles and abstracts were unsuccessful, so we defer the determination of optimal methods for categorizing these grants to future work.



Exploring Trends in CMU Grant Data

Kyra Balenzano, Melissa Dy, Michael Li, and Veda Lin

Project Advisor: Peter Freeman *Project Supervisors:* Huajin Wang and Sarah Young

Carnegie Mellon University
Libraries

Background

- Professors and researchers often seek the help of the Carnegie Mellon University Libraries to acquire information about the university's research grants and research collaborations.
- Current methods to answer these inquiries require large amounts of manual labor and analysis, as there are no tools at the CMU Libraries' disposal to process and glean insights from the research grant data they have.
- The goal of our research is to create such a tool—specifically, an R Shiny app—that can visualize and analyze the data in many ways, including those that address these questions:
 1. *What funding agencies have funded research at CMU and how has this changed over time? How is this related to fields/categories of research?*
 2. *Which authors are associated with a certain field/category of research?*
 3. *How do funded research topics change over time?*

By addressing these questions, the CMU Libraries will be better equipped to answer grant-related questions, therefore empowering university's active research community.

Data Description

Our dataset, pulled from Dimensions AI, a research grants database that links grants to publications, consists of 6,128 research grants affiliated with CMU and 28 primary information fields, which can be summarized in three main groups:

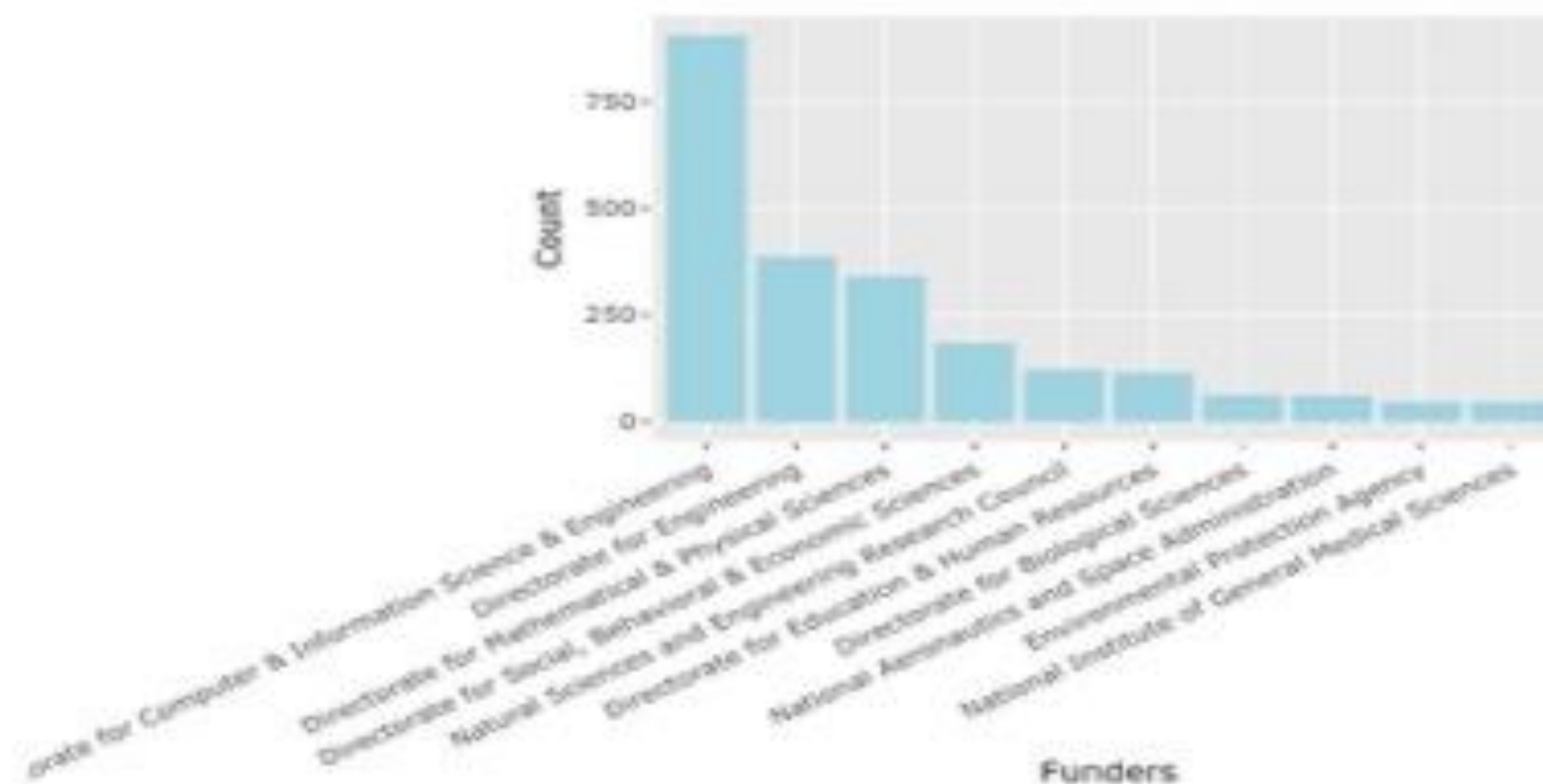
General	Funding	Categorizations (<i>System Author</i>)
<ul style="list-style-type: none">• Grant ID, Number• Title• Abstract• Researchers• Research Organization• Research City, State, Country• Start Date, Year• End Date, Year• Resulting Publications• Dimensions AI Link	<ul style="list-style-type: none">• Funding Amount• Funder• Funder Group• Funder Country• Funder Grant Link	<ul style="list-style-type: none">• Fields of Research Categories (<i>ANZSRC</i>)• Research, Condition, and Disease Categories (<i>NIH</i>)• Health Categories (<i>HRCS</i>)• Health Research Activities (<i>HRCS</i>)• Cancer Types (<i>ICRP</i>)• Common Scientific Outline Categories (<i>ICRP</i>)• Units of Assessment (<i>REF</i>)• Sustainable Development Goals (UN)

In the pre-processing stage, we combined the categorizations into one column for easier access and created a mapping to combine similar categories.

Categories & Funding

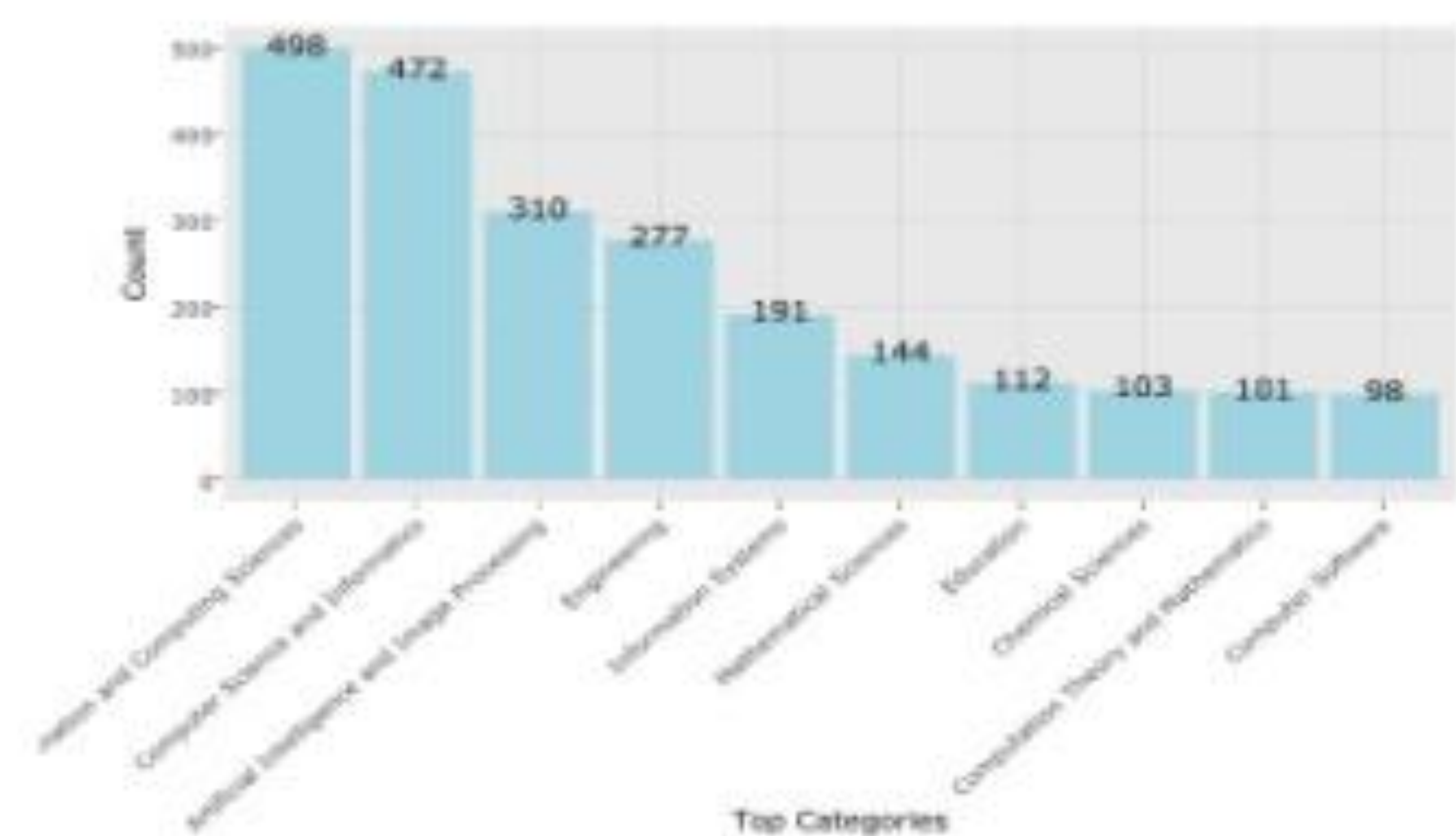
- Our R Shiny app allows users to visualize and analyze trends in categories of research and funding sources for a selected time frame using bar graphs.

Top 10 Funders From 1996 To 2015



Top funders from 1996 to 2015

Top 10 categories from 2010 to 2015 (grant start date), by Count



Top categories from 2010 to 2015

Topic Modeling



LDA with $k=4$ for all grants with an Earth Science categorization. A reasonable interpretation is that topic 1 corresponds with Paleontology, topic 2 with Hydrology, topic 3 with Atmospheric Science and Emissions, and topic 4 with collaborative Glaciology efforts.

- Within a research category, there is no further granularity of research topics.
- Leveraging Natural Language Processing (NLP) techniques on the grant titles and abstracts proves to be promising in identifying these topics.
- Specifically, we use Latent Dirichlet Allocation (LDA) with Gibbs Sampling, an unsupervised learning model, to find topic clusters.
- Users can freely adjust the number of clusters to find the most personally usable results, and the year range to see changes over time. Additionally, details of corresponding grants can be displayed.

Researchers & Collaborations

Grant Researchers

Category: Mathematics & Statistics From (Year): 2008 To (Year): 2020

Show 10 entries Search:

	Researchers	Frequency	Grant Titles
1	ALAN PRITTY	1	[[Random Structures and Algorithms
2	AMIR SABAHI FARMAH	1	[[CDS&E: Collaborative Research: Deep learning enhanced parallel computer-aided fluid flow around moving boundaries on binarized pictures
3	ANTON BERNSTEIN	1	[[The Interplay Between Combinatorics, Set Theory, and Dynamics

Collaborators

Researcher: Peter Freeman From (Year): 2000 To (Year): 2010

Show 10 entries Search:

	Collaborator	# of Grants	Grant Titles
1	AMIN LEE	1	NSF - AST: Sparse Representation and Efficient Inference for Astronomical Spectra
2	CHAD SCHAFER	1	NSF - AST: Sparse Representation and Efficient Inference for Astronomical Spectra

Showing 1 to 2 of 2 entries Previous 1 Next

Network Graphs of Collaborations

Category: Mathematics & Statistics From (Year): 2008 To (Year): 2020

(From left to right)

List of researchers in a given category and time frame

List of collaborators for a given researcher in a selected time frame

Network graph of collaborations in a given category and time frame

- Our R Shiny app also allows users to investigate which researchers are associated with a given field and their collaborations with each other.

Advanced Searching

- Users can easily search for relevant research using keywords and phrases.
- Alternatively, users can provide a specific grant of interest and explore similar works.

Search for relevant grants by keywords

privacy, machine learning

Search in abstracts too

ID	Start Date	Title	Dimensions URL
84	2020-03-01	CAREER: Practical Privacy and Fairness for Data-Driven Applications	Link
214	2019-07-13	SAFC: CORE: Medium Collaborative, Automatically Answering People's Privacy Questions	Link
277	2019-03-01	CRU: DEU: Site in Software Engineering (REUSE)	Link
297	2019-01-01	BIGDATA: P: Optimization in Federated Networks of Devices	Link
377	2018-08-15	CHS: Small: Deep Integration of Context and AI for Robust, Scalable, and Privacy-Preserving Conversational Assistance	Link

Searching for grants by keywords

Search for similar grants by ID

128

How many grants to display

Search input info

84-523

Title: Improving Human Interaction with Robot Teams

ID	Start Date	Title	Dimensions URL
480	2018-03-15	CHS: CHS: Identifying When People Need a Robot's Assistance	Link
2009	2011-09-01	RI: Small: Physical Interaction with Dynamically Stable Mobile Robots	Link
2075	2009-06-17	Design of a four degrees of freedom slave robot for bilateral teleoperation for an application in minimally-invasive surgery	Link

Searching for related grants

Conclusions & Future Work

- Our app displays a wide variety of insights that can be used by the CMU Libraries to help professors and researchers find the resources they need to succeed.
- About 1,700 grants do not have any category labelings, therefore limiting the functionality of some of our application's tabs. Initial attempts to categorize these uncategorized grants based on patterns in titles and abstracts were unsuccessful, so we defer the determination of optimal methods for categorizing these grants to future work.

Thank you!