Board 410: Tracing the Evolution of NSF REU Research Priorities and Trends

Dr. Yanxia Jia, Arcadia University

Dr. Yanxia Jia is an Associate Professor of Computer Science in the Department of Computer Science and Mathematics at Arcadia University. She earned her doctoral degree in Computing Science from University of Alberta, Canada. Dr. Jia's research interests include data science, machine learning, computer science education and computer networks.

Tiantian Wang, The University of Texas at San Antonio

Tiantian Wang is currently a PhD student in the Department of Electrical and Computer Engineering at the University of Texas at San Antonio (UTSA). Her research interests are in the field of single-cell analysis, specifically in identifying and analyzing cell types. By integrating advanced technologies and computational methods, she aims to uncover new insights into cellular functions and interactions, which could potentially lead to breakthroughs in biological research and personalized health control.

Chaomei Chen, Drexel University

Dr. Chaomei Chen is a Professor of Information Science in the College of Computing and Informatics at Drexel University in the USA. He is the Editor-in-Chief of Information Visualization and the Field Chief Editor of Frontiers in Research Metrics and Analytics. He is the author of a series of books on visualizing the evolution of scientific knowledge, including Representing Scientific Knowledge: The Role of Uncertainty (Springer 2017), The Fitness of Information: Quantitative Assessments of Critical Information (Wiley, 2014), Turning Points: The Nature of Creativity (Springer, 2011), Information Visualization: Beyond the Horizon (Springer 2004, 2006), Mapping Scientific Frontiers: The Quest for Knowledge Visualization (Springer 2003, 2013). He is the creator of the widely used visual analytics software CiteSpace for visualizing and analyzing structural and temporal patterns in scientific literature.

Yu-Fang Jin, The University of Texas at San Antonio

Dr. Yufang Jin got her Ph.D from University of Central Florida in 2004. After her graduation, she joined the University of Texas at San Antonio (UTSA). Currently, she is a Professor with the Department of Electrical and Computer Engineering at UTSA. Her research interest focus on applications of artificial intelligence, interpretation of deep learning models, and engineering education.

Tracing the Evolution of NSF REU Research Priorities and Trends

Abstract

The Research Experiences for Undergraduates (REU) program plays a crucial role in fostering research interests among undergraduate students, motivating them to pursue advanced degrees in Science, Technology, Engineering, and Mathematics (STEM) fields, and developing a diverse, skilled workforce for STEM careers. Annually, the National Science Foundation (NSF) awards approximately 170-190 REU grants. The funding for REU sites often reflects current trends in research. Our study aims to examine REU sites' contributions in terms of scholarly publications and student training over the past six years. Additionally, we explore the research themes of these REU sites and compare them with those in the Web of Science (WoS) database.

The NSF award database provides details about 3,500 REU awards, including project titles, abstracts, funding periods, and NSF directories. All REU award information is reformatted into the WoS citation format for thorough analysis using a literature analysis tool CiteSpace. Utilizing CiteSpace, we create and visualize topic clusters based on terms and keywords of REU titles and abstracts. Outcome data of REU sites is extracted from the 'Disclaimer/Publications' sections found in the Project Outcomes Reports on NSF award webpages. Quantifiable metrics are extracted, including the number of REU trainees and underrepresented and/or minority students, the number of publications produced, and the number of students who advanced to graduate studies.

Distribution of REU awards across various NSF directories is summarized, highlighting the emphasized areas of REU programs. Examining the quantified outcomes of the REU projects, such as the number of trainees, underrepresented trainees, publications, and students joining graduate school, facilitates a quantitative evaluation of the impact of REU programs and verifies REU sites' efforts to meet the goals of NSF REU program. Research themes of REU awards and engineering, science and technology-related publications from WoS are represented through the creation of topic clusters. Shared research themes from REU programs and WoS publications suggest that REU sites are keeping pace with the current and emerging trends in scientific research and that the REU program is an effective vehicle for contributing new knowledge to the research community. The analysis of the REU outcome data shows that REU sites made effective efforts in increasing the percentage of underrepresented students.

This study represents the first systematic and quantitative analysis of REU grants in terms of their research trends and outcomes. The insights gained will provide valuable information on the evolution of REU research areas and the scholarly impacts of REU programs, benefiting and aspiring REU principal investigators, grant administrators, and a broader range of researchers.

Objective and Motivation

Undergraduate research stands out as one of the most effective methods for enhancing college students' performance. Research Experiences for Undergraduates (REU) encompass both curriculum-based and co-curricular-based experiences. Curriculum-based experiences involve capstone courses and integrated research components within the curriculum. Co-curricular-based experiences are obtained through research activities outside the classroom, such as participating in summer research internships, fellowships, programs, and guided research projects. Undergraduate research experience serves as a valuable platform for fostering students' interest in research, attracting more individuals to pursue advanced degrees in Science, Technology, Engineering, and mathematics (STEM)-related fields, and cultivating a well-trained and diverse workforce in STEM careers. The impacts and benefits of the REU program have been confirmed by several large-scale surveys [1-6].

Recognizing the importance of involving undergraduate students in meaningful research and scholarly activity alongside faculty members [6], numerous REU sites have garnered support from various agencies such as National Science Foundation (NSF), the National Institute of Health, the Department of Energy, the Department of Defense, and the Department of Education. Additionally, private foundations contribute to various scholarships to bolster undergraduate research endeavors. The significant efforts from both government and private foundations have played a crucial role in enabling 45% of STEM undergraduate students to participate in research in 2023 [7]. The number of participants has been increasing steadily in the past decades [8-10].

The NSF has a longstanding commitment to offering research opportunities to undergraduates, particularly those from underrepresented groups, through its REU sites. The NSF REU program seeks to attract a diverse pool of talented students into careers in science and engineering through disciplinary, interdisciplinary, or educational research and to help ensure that they receive the best education possible. An REU site typically comprises a cohort of approximately ten undergraduates who engage in research within the programs of the hosting institution. Each student is assigned to a specific research project and collaborates closely with faculty members and other researchers.

A search of the NSF award database in December 2023 yielded approximately 3,500 REU sites, with the earliest REU site commencing in June 1982. This indicates a substantial and persistent contribution and effort from the NSF in promoting undergraduate research. From 2017 to 2022, the NSF awarded approximately 170-190 REU grants annually. Each year, about 6,000 undergraduates participate at NSF REU sites to perform research spanning 19 subject areas such as astronomical sciences, cyberinfrastructure, engineering, and STEM education. Most REU sites are hosted by institutes in the United States while some research is performed abroad, including a marine science project at Bermuda and some international REU sites such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) project. The continuous support from NSF has led to a total of 1,077 NSF REU sites in the United States from 2017 to 2022.

REU programs typically integrate educational and research activities with the goal of preparing a capable workforce to meet the demands arising from technological development. Previous research has underscored the significant impacts of REU programs, including stimulating interest in STEM fields among undergraduates [1, 6], promoting diversity in the workforce[11],

encouraging enrollment in graduate schools [1], and enhancing educational outcomes [4, 12, 13]. However, there is a deficiency in systematically analyzing the outcomes of REU programs and assessing whether the research themes of REU sites align with the evolving high-impact topics evidenced in literature.

To bridge the gap, the primary goal of this study is to examine the impacts of the NSF REU program. We collected and analyzed quantitative REU outcome data from the "Disclaimer" in the "Project Outcomes Report" section on REU award webpages. The "Disclaimer" provides detailed outcomes of an REU site, including the number of undergraduate students trained, the number of underrepresented undergraduate students trained, the number of REU trainees joining graduate school post REU training, and the number of publications reported by the REU site. We also compared the research focus of REU sites from 2017 to 2022 with publications in the Web of Science (WoS). The insights gained from this study would provide valuable information on the evolution of REU research areas and the scholarly impacts of the NSF REU program, offering significant benefits to REU principal investigators, aspiring participants, and grant administrators.

Data Collection

Data Collection for REU Outcome Analysis

To analyze REU outcomes, we collected outcome data for expired REU awards funded during fiscal years 2017 ~ 2019 from the NSF REU award websites, noting that REU awards funded in fiscal year 2019 (10/01/2019-09/30/2020) should expire in 2023, assuming a 3-year funding period. Furthermore, we selected 2017 as the starting year, hoping to identify recent research trends without an excessive amount of literature from the Web of Science (WoS).
For REU publication outcome data, we originally examined the NSF Public Access Repository (NSF-PAR) database for all publications, with "REU" as a keyword. The query led to a total of 510 records, including 143 conference publications, 1 conference proceeding, 3 datasets, and 363 journal articles from both active and expired REU grants. Among the 510 records, 83 publications have "REU Site" in the title.

Due to the insufficient publication records from the PAR database, we examined the "Disclaimer" of all expired REU awards funded from 2017~2019 in the NSF award webpage. After combining identical disclaimers for collaborative REU sites and removing 46 expired awards without a "Disclaimer" section in their award page, outcomes of a total of 243 REU awards were retained for analysis. Quantitative outcomes of an REU Site include the number of undergraduate students trained, the number of underrepresented undergraduate students trained, the number of REU trainees joining graduate school post REU training, and the number of publications reported by the REU site.

Data Collection for REU Research Topic Analysis

To compare the research themes of REU projects from fiscal year 2017 to 2022 with those of WoS publications during the same period of time, we retrieved detailed records of both active and expired REU awards for fiscal years 2017 to 2022 from the NSF award database [14].

A query into the NSF award database, filtered by "REU" in the project title and with the start date between 10/01/2017 and 9/30/2023, results in approximately 1,100 awards. Full records of these REU awards were downloaded and saved as a CSV file. The columns contain award number, title, start date, end date, NSF directorate, award amount to date, state of awarded organizations, and abstract of the REU award.

For pre-processing, we combined identical records of collaborative REU projects involving multiple universities and removed REU/RET sites and records that are not REU grants. As a result, a total of 1,077 unique REU projects were extracted.

Data Collection for WoS Publications Topic Analysis

The Web of Science is a subscription-based platform that provides access to multiple databases of scholarly publications, containing bibliographic and citation data from academic journals, conference proceedings, and other documents in various academic disciplines. These disciplines are broadly categorized into Arts & Humanities, Life Sciences & Biomedicine, Physical Sciences, Social Sciences, and Technology, encompassing 14, 77, 18, 26, and 22 areas respectively, as classified by WoS. It covers recorded publications since 1900 and holds over 79 million records in its core collection and 171 million records across its platform. Each data entry in WoS is assigned to at least one category [15]. Each WoS record includes information such as title, keywords, abstract, authors, and references. The frequency of a keyword across all records usually indicates the prominence of related research topics associated with that keyword, while citations of a record reflect the impact of the research documented within it.

A query of WoS by selecting topics in Engineering, or Science, or Technology from January 2018 to December 2023 yields a total of 1,602, 854 records of articles, proceeding papers, book chapters, or books in Core WoS collections. Applying the filter of "Highly Cited Papers" results in 13,564 records. All records were exported to "Plain Text File" with "Full Record and Cited References". The saved text file serves as an input to the literature analysis package, CiteSpace [16].

Methods

To meet our first goal of analyzing quantitative REU outcome data, we visualized the funding distribution from 2017 to 2022 across seven NSF directorates, including Mathematical and Physical Science (MPS), Biological Science (BIO), Computer and Information Science and Engineering (CSE), Engineering (ENG), Geoscience (GEO), Social, Behavioral and Economic Sciences (SBE), and STEM Education (EDU**)**. For our second goal, which involves comparing the primary research themes of all REU sites with those in WoS publications during the same period of time, we used a literature analysis software CiteSpace.

CiteSpace – a Literature Analysis Tool

CiteSpace is a visual analytic tool for visualizing and analyzing the literature of a scientific domain and has been used in scientometric studies of scientific and technological research. It takes bibliographic information such as keywords, terms (noun phrases) and cited references

from the Web of Science and Scopus databases and generates interactive visualizations. Specifically, CiteSpace generates and visualizes a network of clusters based on keywords and references from scholarly publications to reveal major research topics in a field of research.

In addition to WoS-format data, CiteSpace also supports datasets from non-WoS resources, such as REU grant data in a CSV format through a converting tool. This feature enables the creation of a network of clusters based on keywords and/or terms (noun phrases) extracted from titles and abstracts of REU awards. Specifically, for each REU award, two different approaches were adopted to extract terms and keywords. Terms were extracted from the titles and abstracts using CiteSpace. Technical keyword phrases focusing on research contents of REU awards were extracted by use of ChatGPT Application Programming Interface (API). Subsequently, a network of clusters was created based on the extracted terms and keywords. These clusters reveal the main topics of all REU projects in the dataset.

Based on the above-mentioned clusters generated from REU and WoS datasets, we compare the research focus observed in REU abstracts with those in Web of Science publications.

Natural Language Processing with ChatGPT

Unlike the WoS publication dataset, the REU dataset does not include keyword information. Considering that REU sites are specifically designed to train undergraduate students in research, the abstracts of REU grants, although including information about research content, place significant emphasis on aspects related to training in undergraduate research. Consequently, to make a fair comparison between the research themes in REU projects and those in WoS publications, we extract keyword phrases focusing on research topics and methods from REU abstracts. This process is executed through programming with the ChatGPT API.

Assessment Methods

1. REU Keywords Extraction

To evaluate the quality of keywords generated by ChatGPT, approximately ten percent of the 1,077 REU award abstracts were randomly selected for manual inspection. The analysis focused on the accuracy and relevance of the keyword phrases produced via the ChatGPT API. Findings indicate that the generated keyword phrases exhibit a high level of quality, demonstrating effectiveness in capturing the essential research topics embedded in the REU award abstracts.

2. Cluster Evaluation

We evaluate the clustering performance using two methods: silhouette values of clusters and the cluster exploration tool embedded in CiteSpace. A Silhouette coefficient measures the quality of a cluster in terms of homogeneity, based on the similarity of a member to other members in its own cluster compared to members of other clusters. The value ranges from -1 to 1. The higher the value, the better the quality of clustering performance.

Second, we made use of the cluster exploration tool in CiteSpace and looked at cluster details, such as cluster size, and top terms/keywords in the cluster. We also reviewed node details. For example, for a keyword node in a REU cluster, we examined its cluster label, the number of REU abstracts it appears in, and the years of appearance. This approach helps us assess the quality of the cluster labels and provides us with valuable insights into the cluster composition.

Results

Distribution of REU Awards

A total of 1,077 REU awards funded by the MPS, BIO, CSE, ENG, GEO, SBE, and EDU directorates from the fiscal year 2017 to 2022 were examined and illustrated in Figure 1. The median funding amounts of REU awards showed an increasing trend from 2017 to 2022 in MPS, CSE, and ENG directorates. There was a significant decrease of median funding amounts in 2022 in BIO and GEO directorates. Interestingly, the figure showed smaller ranges of funding amounts in CSE and ENG, compared with MPS and BIO, while the largest funding amount range was observed in MPS and GEO during these years.

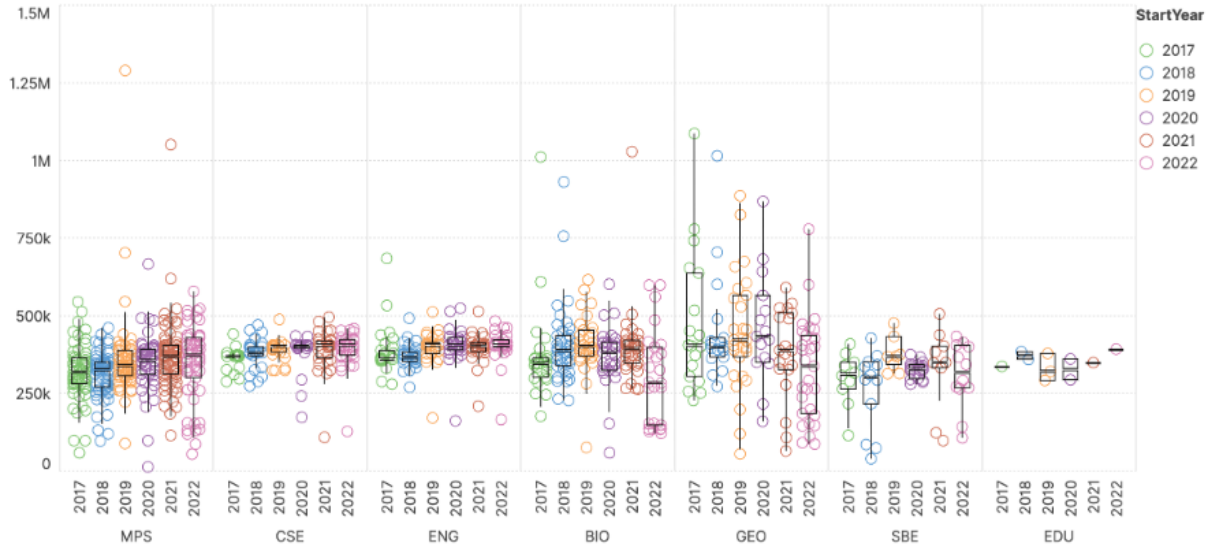

Figure 1. Illustration of the funding amounts of 1,077 REU awards from fiscal years 2017 to 2022, categorized by funding years and NSF directorates. Each circle denotes an REU award. The funding amount is depicted on the vertical axis, with funding year and directorate represented on the horizontal axis. The box-and-whisker plot representing the funding amount for each directorate per fiscal year is displayed in black. The boxes illustrate the interquartile range and highlight the median funding amount, while the upper and lower whiskers indicate the maximum and minimum funding amounts, respectively.

The total amount awarded and the number of REU awards funded by each NSF directorate per year were also examined and illustrated in Figure 2. Interestingly, there is an increasing trend in both the total award amount and the number of REU awards in the CSE directorate, while a decreasing trend is observed in the ENG directorate. The funding amount and

number of REU awards funded by other directorates during this period did not exhibit a clear pattern.

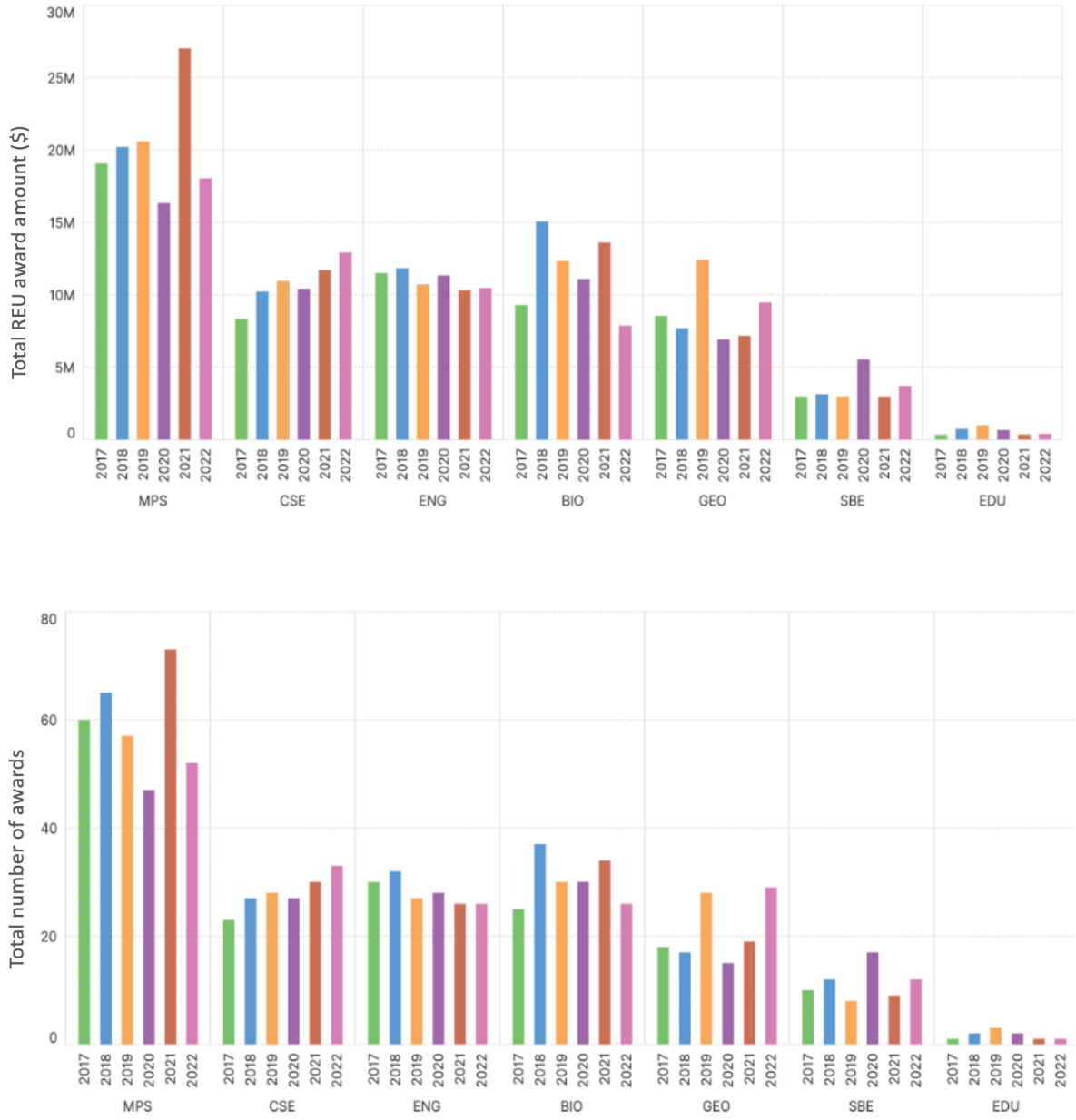

Figure 2. The total award amount of REU awards (upper) and the number of REU awards (lower) per NSF directorate, including MPS, BIO, CSE, ENG, GEO, SBE, and EDU, from 2017 to 2022. The MPS directorate stands out with the largest total award amount and number of REU awards, while SBE and EDU show much smaller amounts and numbers of REU awards compared to other NSF directorates during the period from 2017 to 2022.

Outcomes of REU Awards

Table 1 lists outcomes of 243 REU awards funded in fiscal years 2017 and 2018 (152 in 2017 and 91 in 2018). These 243 awards reported "Disclaimer" in their award web pages by December 2023. REU awards funded in 2019 are not included due to insufficient information, since very few of them are listed in the NSF award database as expired grants at the time of our search. The MPS directorate has the highest number of student participants. Table 1 also highlights the effort to attract a diverse pool of talented students into science and engineering research, which is a key goal of the NSF REU program. Four directorates, CSE, ENG, GEO, and SBE, show an increasing percentage of participants from underrepresented groups. The GEO and ENG Directorates are the two with the highest percentage of underrepresented trainees in 2018. The ENG and SBE directorates have the highest percentage of students attending graduate school. Overall, the ENG directorate demonstrated a relatively larger percentage of Underrepresented Minority (URM) trainees and REU trainees advancing to graduate school after their training, compared to other directorates.

Table 1. Outcomes of 243 REU awards funded in fiscal years 2017 and 2018.

NSF Directorate	Fiscal Year	Total # of trainees	% of under-represented /minority Trainees	% of trainees join graduate school	Total # of publications
MPS	2017	1704	28.35%	19.25%	185
	2018	1183	25.27%	10.40%	160
CSE	2017	721	27.74%	9.29%	103
	2018	640	30.78%	1.88%	124
ENG	2017	771	33.59%	13.36%	79
	2018	317	46.37%	29.34%	37
BIO	2017	875	30.63%	12.69%	71
	2018	487	24.44%	8.42%	32
GEO	2017	748	15.51%	7.35%	90
	2018	220	50.00%	5.91%	15
SBE	2017	153	0.00%	33.33%	17
	2018	124	37.10%	31.45%	6

There is a substantial amount of data missing. For instance, some awards funded in 2017 and 2018 did not include disclaimer information at all. For those with disclaimer information, there is still a significant amount of missing data in the "Disclaimer" section, as shown in Table 2.

A total of 243 awards were included for outcome analysis, including 152 funded in fiscal year 2017 and 91 in 2018. The notably smaller number of funded awards in 2018 than 2017 is likely due to some REU awards funded in 2018 still pending status updates. In addition, no report from

the EDU division was disclaimed, considering very few REU awards were funded by EDU division as shown in Figure 2.

Table 2. Number (percentage) of REU sites that reported outcome data.

Year	# of funded sites included in the analysis	# of sites reported URM or female student counts	# of sites reported publication counts	# of sites reported counts for students attending graduate school
2017 & 2018 combined	243	108 out of 243 (44%)	156 out of 243 (64%)	75 out of 243 (31%)

Analysis of REU Research Trends

To examine the research topics embedded in REU training, we use CiteSpace to analyze their keywords and phrases from 1,077 abstracts of REU awards funded in fiscal years 2017-2022, as well as such information embedded in WoS literature. The analysis results are shown in Table 3 and supplementary Figures S1 and Figure S2. The weighted mean silhouette values for the REU research topic clusters based on keywords and for the WoS publication topic clusters from CiteSpace are 0.9555 and 0.8928 (Figures S1 and S2), respectively, suggesting good performance of the clustering algorithm.

Table 3. Comparison of hot topics from WoS and research topics from REU awards.

Hot topics from WoS publications (titles, abstracts and keywords)	Topics from REU awards (titles and abstracts)
Digital Transformation in Supply Chain Management	Artificial Intelligence
Environmental Sustainability through Technological Innovation	Computational Biology
Advanced Photocatalytic Material Synthesis	Condensed Matter Physics
Advancements in Electrocatalysis Engineering	Mathematical Biology
COVID-19 Research and Implications	Climate Change
Internet of Things and Blockchain Technologies	Emerging Technologies
Advancements in Solar Energy Technology	Biogeochemical Cycling
Advanced Materials and Nanotechnology Applications	Ecosystem Storage
Advanced Battery Technology Development	Energy Storage
Machine Learning and Deep Learning Applications	Machine Learning Algorithms
Advanced Optimization and Modeling Techniques	Interdisciplinary research*
Environmental Science and Technology Advance	Scientific Communications*
	Scientific Conferences*
	Multidisciplinary Research*
	Professional Development Workshops*
	High-Quality Research Experience*
	Research Projects*

Not surprisingly, REU topics include not only technical topics but also research training-related topics, such as scientific communication, high-quality research experiences, and professional development workshops. Table 3 listed the WoS and REU research topics based on a descending order of cluster sizes. The technical topics and non-technical topics (denoted with *) were listed as two separate sections.

Related topics in REU and WoS are highlighted with the same color. For example, the REU topics of "Artificial Intelligence" and "Machine learning" are directly related to the WoS research topic of "Machine Learning and Deep Learning Applications". Similarly, the WoS research topic "Advanced Battery Technology Development" is related to the REU research topic "Energy Storage". If an REU topic has multiple related WoS topics, only one match (highlighted color) is selected to simplify the interpretation. As can be observed in Table 3, REU research topics in general align with hot topics in WoS publications, with the matching performed by ChatGPT and human inspection.

Research topic clustering on REU projects and WoS publications indicates that the REU research topics reflect the highly cited research themes of WoS publications. This suggests that REU sites are keeping pace with the current and emerging trends in scientific research and that the NSF REU program is an effective vehicle for contributing new knowledge to the research community.

Keywords	Strength	Begin	End	2017 - 202
scientific communication	2.11	2019	2020	
mathematical biology	1.29	2018	2019	
autonomous vehicles	1.05	2019	2020	
cognitive science	1.05	2019	2020	
big data	1.05	2019	2020	
career opportunities	1.05	2019	2020	
nanoscience and nanotechnology	1.04	2017	2018	
computational chemistry	1.04	2017	2018	
enzyme function	1.03	2020	2022	
molecular interactions	1.03	2020	2022	
condensed matter	1.03	2020	2022	
ethical conduct of research	1.01	2018	2019	
systems biology	0.99	2018	2019	
data analysis	0.97	2018	2019	
computer systems	0.7	2019	2020	
biological mechanisms	0.7	2019	2020	
behavioral pharmacology	0.7	2019	2020	
interdisciplinary research training	0.7	2019	2020	
interdisciplinary science	0.7	2019	2020	
drug delivery	0.69	2017	2018	
nanoscale materials discovery	0.69	2017	2018	
commercializable innovations	0.69	2017	2018	
interdisciplinary research opportunitie	0.69	2017	2018	
molecular dynamics	0.69	2017	2018	
nano-scale materials	0.69	2017	2018	
biogeochemical cycling	0.69	2017	2018	

Keywords	Strength	Begin	End	2017 - 2022
air quality	0.69	2017	2018	
ecosystem functioning	0.69	2017	2018	
human-robot interaction	0.69	2017	2018	
graduate studies	0.69	2017	2018	
chemistry research	0.69	2017	2018	
numerical modeling	0.69	2017	2018	
materials chemistry	0.69	2017	2018	
machine learning	0.69	2020	2022	
human-centered research experiences	0.69	2020	2022	
geospace research	0.69	2020	2022	
medical imaging	0.69	2020	2022	
interdisciplinary chemical research	0.69	2020	2022	
knot theory	0.69	2020	2022	
number theory	0.68	2017	2018	
computational sensing	0.67	2018	2019	
international scientific collaborations	0.67	2018	2019	
molecular ecology	0.67	2018	2019	
independent research	0.67	2018	2019	
discipline-based education research	0.67	2018	2019	
computer programming skills	0.67	2018	2019	
applied mathematics	0.67	2020	2022	
professional development workshops	0.66	2020	2022	
energy generation	0.51	2017	2019	
ecosystem services	0.51	2019	2022	
chemical biology	0.51	2019	2022	
astrophysics research	0.5	2018	2020	

Figure 3. Top 52 keywords with the strongest bursts of occurrence frequency.

Results of REU Keyword Bursts

Figure 3 shows the top 52 REU keyword phrases with the strongest occurrence bursts. They are the most popular keyword phrases used in REU abstracts during 2017-2022. The keywords were sorted in descending order of burst strengths within the usage period shown with red bars. The higher a keyword phrase' strength value is, the more REU projects it appears in. These keyword phrases also align with the future technologies recently identified by the National Science and Technology Council [17].

Conclusion and Discussion

This study represents the first systematic and quantitative analysis of REU grants in terms of their research trends and outcomes. We presented the distribution of REU awards across 7 NSF directorates in a six-year span from 2017 to 2022 fiscal years based on public data released by NSF. Analysis shows that the MPS directorate has the highest number of awards and total award amount in all years from 2017 to 2022. Based on available REU outcome data in 2017 and 2018, MPS is the directorate with the highest number of participants. Additionally, efforts by REU sites to increase the percentage of underrepresented students has been observed, with four directorates, CSE, ENG, GEO, and SBE, showing an increasing percentage of participants from underrepresented groups from 2017 to 2018. However, missing and incomplete disclaimer information might have impacted the accuracy of this conclusion. The significant amount of missing information in the disclaimer section for quantitative analysis suggests the need for consistency in submission of disclaimers.

The match between research topics obtained from REU awards and WoS publications indicates that the REU research topics reflect the highly cited research themes in WoS publications. This also suggests that REU sites are keeping pace with the current and emerging trends in scientific research and that the REU program is an effective vehicle for contributing new knowledge to the research community. Additionally, the 52 most frequently occurring keyword phrases identified from REU awards align with the future technologies recently highlighted by the National Science and Technology Council.

The insights gained from this study will provide valuable information on the evolution of REU research areas and the scholarly impacts of REU programs, benefiting and aspiring REU principal investigators, grant administrators, and a broader range of researchers.

Acknowledgement

The authors acknowledge the support by the NSF Award #2051113.

Supplements

Figure S1 illustrates clusters of REU research topics based on keyword phrases extracted from titles and abstracts while Figure S2 shows clusters of research topics based on keywords, titles, and abstracts from WoS literature.

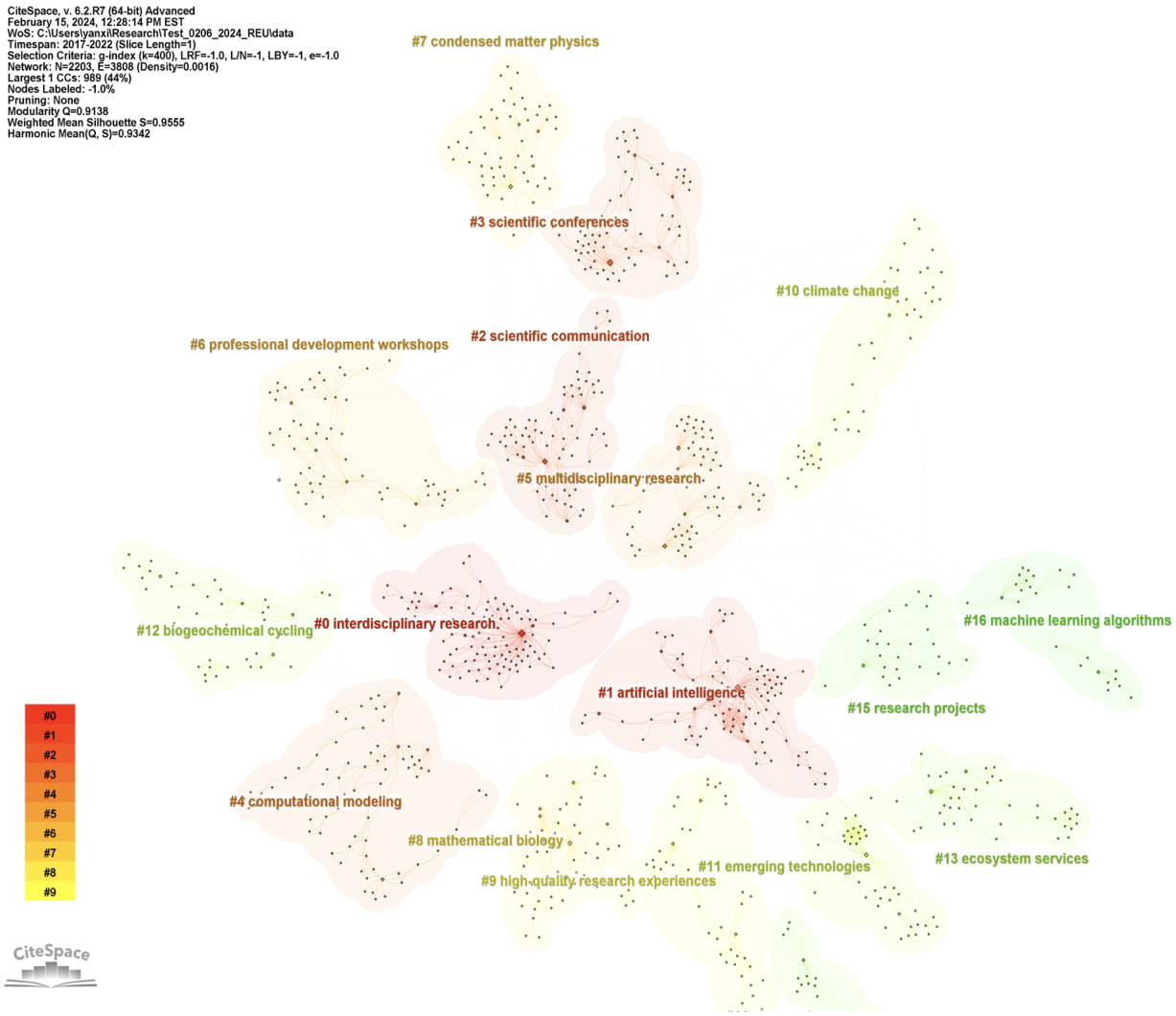

Figure S1. REU topic clusters based on keywords extracted from titles and abstracts.

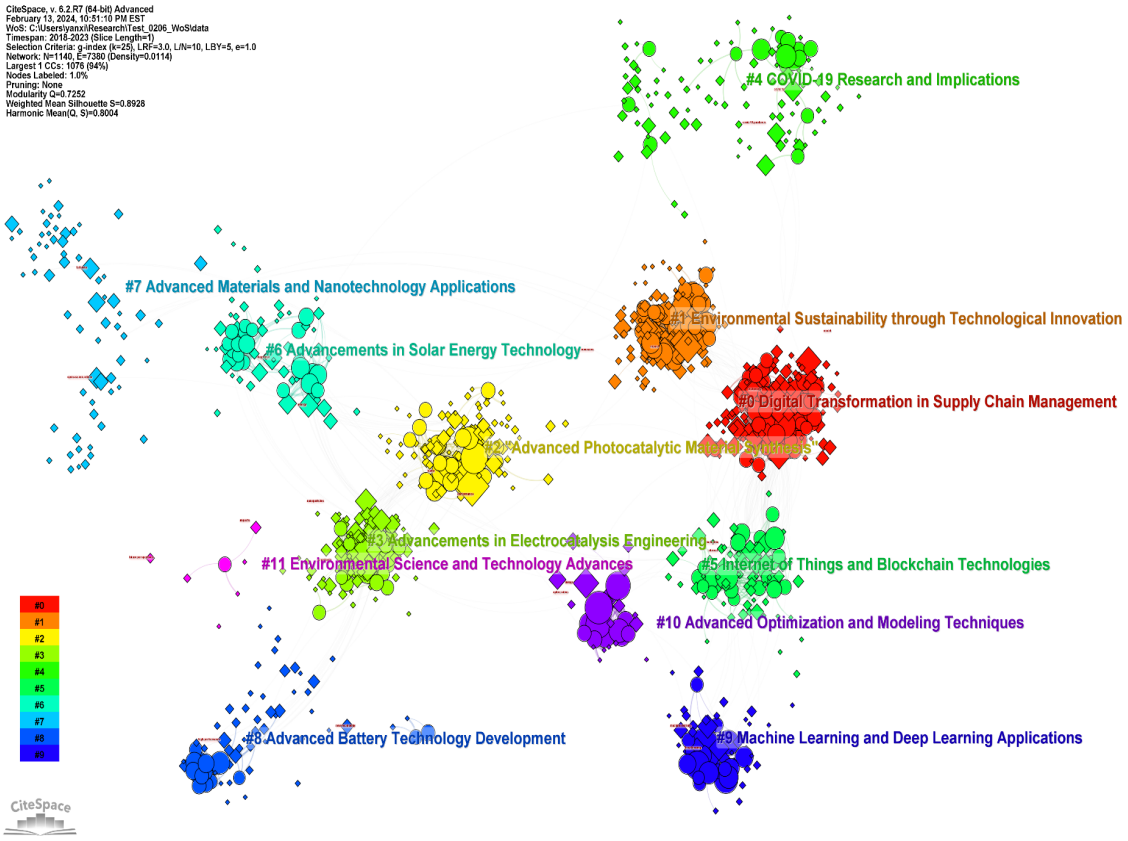

Figure S2: WoS topic clusters based on keywords, titles, and abstracts.

References

[1] S. H. Russell, M. P. Hancock, and J. McCullough, "Benefits of undergraduate research experiences," *Science,* vol. 316, no. 5824, pp. 548-549, 2007.

[2] D. Lopatto, "Survey of Undergraduate Research Experiences (SURE): first findings," (in eng), *Cell Biol Educ,* vol. 3, no. 4, pp. 270-7, Winter 2004, doi: 10.1187/cbe.04-07-0045.

[3] D. Lopatto, "Undergraduate research experiences support science career decisions and active learning," (in eng), *CBE Life Sci Educ,* vol. 6, no. 4, pp. 297-306, Winter 2007, doi: 10.1187/cbe.07-06-0039.

[4] D. Lopatto, "Exploring the benefits of undergraduate research experiences: The SURE survey," in *Creating effective undergraduate research programs in science.* New York, NY, US: Teachers College Press, 2008, pp. 112-132.

[5] M. C. Linn, E. Palmer, A. Baranger, E. Gerard, and E. Stone, "Undergraduate research experiences: Impacts and opportunities," *Science,* vol. 347, no. 6222, p. 1261757, 2015.

[6] N. S. F. D. o. U. Science and M. Education, *Report on the National Science Foundation Disciplinary Workshops on Undergraduate Education: Recommendations of the Disciplinary Taskforces Concerning Critical Issues in US Undergraduate Education in the Sciences, Mathematics and Engineering* (no. 3). National Science Foundation, 1989.

[7] J. Lindner, "Undergraduate Research Statistics," in "Education Statistics," Gitnux Market Data Report 2024, December 16, 2023. [Online]. Available: https://gitnux.org/undergraduate-research-statistics/

[8] R. Ignace, "Quantifying Undergraduate Participation in Scholarly Activities," *Scholarship and Practice of Undergraduate Research,* vol. 3, no. 2, pp. 62-70, 2019.

[9] A. Wilson, "Using the National Survey of Student Engagement to measure undergraduate research participation," *Council on Undergraduate Research Quarterly,* vol. 32, no. 3, pp. 9-15, 2012.

[10] J. Mervis, "Student Research: What Is It Good For?," *Science,* vol. 293, no. 5535, pp. 1614-1615, 2001, doi: doi:10.1126/science.293.5535.1614.

[11] P. R. Hernandez, A. Woodcock, M. Estrada, and P. W. Schultz, "Undergraduate research experiences broaden diversity in the scientific workforce," *BioScience,* vol. 68, no. 3, pp. 204-211, 2018.

[12] Y.-F. Jin, "Closing the Loop: A 10-year Follow-up Survey for Evaluation of an NSF REU Site," in *ASEE annual conference*, 2022.

[13] M. J. Graham, J. Frederick, A. Byars-Winston, A.-B. Hunter, and J. Handelsman, "Increasing persistence of college students in STEM," *Science,* vol. 341, no. 6153, pp. 1455-1456, 2013.

[14] National Science Foundation. "NSF Award Search." https://www.nsf.gov/awardsearch/ (accessed April 10, 2024).

[15] Web of Science. "Web of Science Search Platform." https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/ (accessed April 10, 2024).

[16] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of the American Society for Information Science and Technology,* vol. 57, no. 3, pp. 359-377, 2006, doi: https://doi.org/10.1002/asi.20317.

[17] National Science and Technology Council, "Critical and Emerging Technologies List Updates," 2022. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2022/02/02-2022-Critical-and-Emerging-Technologies-List-Update.pdf