

Project Proposal Title

Zachary Caterer^{1,2,3}

¹Department of Computer Science, University of Colorado Boulder

²Department of Chemical and Biological Engineering, University of Colorado Boulder

³Interdisciplinary Quantitative Biology Program, University of Colorado Boulder

February 7, 2025

Proposal

The National Science Foundation (NSF), as an independent federal agency, represents one of the largest public funders of fundamental research in the United States, investing over \$8.8 billion annually across science and engineering disciplines. The NSF serves as a crucial funding source and knowledge disseminator for public and private universities and research institutions across the United States, supporting research that spans from mathematics and computer science to social sciences and engineering.

Research funding through the NSF spans various mechanisms, from CAREER awards supporting early-career investigators to standard research grants funding established projects. As modern research becomes increasingly complex and interdisciplinary, the NSF has adapted its funding structures to support collaborative research through mechanisms like collaborative research grants and cross-disciplinary program solicitations, recognizing that breakthrough discoveries often require diverse expertise and cross-institutional collaboration.

This study aims to analyze how collaboration patterns in NSF-funded research have evolved over the past 25 years, particularly comparing recent awards (2020-2025) to those from 1995-2000. Using data from the NSF Awards database, we will construct and analyze two temporal networks representing these distinct periods. In our network structure, research institutions receiving NSF funding will serve as nodes, while collaborative grants between institutions will form edges. Edge weights will represent the number of shared grants between institutions, and we will allow self-loops only for cross-departmental collaboration within institutions. This analysis will provide a comprehensive view of how fundamental science research collaboration has evolved in the United States.

Our analysis will employ Python's NetworkX library to compute various network metrics. At the network level, we will examine global clustering coefficients, average path lengths, and network density to understand the overall collaborative landscape. For individual institutions,

we will analyze degree centrality, betweenness centrality, and eigenvector centrality to identify key players and their roles in the research ecosystem. Additionally, we will investigate whether certain research disciplines or programs tend to foster more collaborative relationships than others.

We anticipate several key findings from this analysis. First, we expect to observe increased network density and clustering in the recent period, reflecting greater collaborative intensity in modern research. Second, we anticipate higher average node degree in the recent network, indicating that institutions are forming more partnerships than in the past. Third, we expect to see the emergence of highly central "hub" institutions that bridge multiple research communities, particularly those successful in securing funding across different NSF directorates. Finally, we anticipate greater cross-disciplinary collaboration as evidenced by more diverse funding mechanisms and research areas.

To conduct this analysis, we will utilize the publicly available NSF Awards database as our primary data source. Python will serve as our main analytical tool, with NetworkX and Pandas libraries handling the network analysis and data manipulation respectively. For visualization purposes, we will employ either Gephi or Python's Matplotlib/Seaborn libraries to create clear and informative network visualizations that highlight key findings and patterns in the data.

This research will provide valuable insights into the evolution of collaborative research patterns across fundamental science over the past quarter-century, potentially informing future funding strategies and research policy decisions. The comparison between historical and contemporary collaboration networks may reveal important trends in how research institutions have adapted to increasingly complex scientific challenges through collaborative approaches, and how NSF funding mechanisms have influenced these collaboration patterns.

Questions I Might Want to Ask & Answer

- What are the institutes driving the most collaboration in the NSF-funded research ecosystem?
- Who are the key researchers that are driving collaboration across different research disciplines?
- How has the structure of NSF-funded research collaboration changed over the past 25 years?
- Are certain research disciplines or programs more likely to foster collaborative relationships?
- Is there a greater propensity for cross-disciplinary collaboration for inter-institutional grants compared to intra-institutional grants?
- Is there a correlation between the monetary value of grants and the number of collaborating institutions? Do larger grants tend to involve more complex collaborative networks?
- How does geographical proximity influence collaboration patterns between institutions? Are there regional clusters or do modern communication technologies enable truly distance-independent collaboration? Has this changed over time?
- What is the relationship between institution size/type (R1 universities, liberal arts colleges, research institutes) and their collaboration patterns? Do smaller institutions tend to collaborate more to access resources?