Understanding the NERSC Collaboration Network through Graph Analysis

Coby Sandholtz
Brigham Young University
Provo, UT, USA
cobysandholtz@gmail.com

Camelia D. Brumar
Tufts University
Medford, MA, USA
camelia_daniela.brumar@tufts.edu

Sumaiya Azad
University of Utah
Salt Lake City, UT, USA
sumaiya.azad@utah.edu

Kevin Gott
Lawrence Berkeley National Laboratory
Berkeley, CA, USA
kngott@lbl.gov

Rebecca Hartman-Baker
Lawrence Berkeley National Laboratory
Berkeley, CA, USA
rjhartmanbaker@lbl.gov

***Abstract*—NERSC users collaborate, forming a network. Which users occupy the most structurally influential positions in the NERSC collaboration network? Using 2020–2026 NERSC allocation records, a proxy collaboration graph was analyzed to reveal the collaboration network's dynamics. Two measures of user connectivity were compared: unweighted degree and repository-size-adjusted (RSA) strength. Degree distributions, program-office affiliation, and user roles were analyzed. Results show that RSA strength is a more accurate measure of user connectivity than unweighted degree. The strongest connectors in the network are career scientists and users associated with High-Energy Physics. These findings suggest that in HPC network data, unweighted connectivity is not a reliable measure of network influence or cross-disciplinary reach. The findings also suggest that allocation networks reflect field-specific collaboration patterns and may therefore provide a useful proxy for coauthorship networks within the same field.**

I. Introduction

The body of researchers who use NERSC resources forms a scientific collaboration network. This analysis seeks to achieve a greater understanding of this network through graph analysis of a proxy network derived from 2020–2026 allocation records. These data include repository name, program office, user ID, and user organizational role.

II. Method

A network graph was constructed by connecting users who share a project allocation in the same year. A 2025 network and a cumulative 2020–2026 network were constructed and analyzed using cuGraph and igraph.

For comparison, graphs representing scientific collaboration networks based on publication data from three databases were used [1]:

- Stanford Public Information Retrieval System (SPIRES): a repository of preprints and papers in high-energy physics.
- Networked Computer Science Technical Reference Library (NCSTRL): a repository of preprints in computer science.
- Los Alamos e-Print Archive (LANL): a repository of preprints in physics.

Two analyses were performed: a degree distribution analysis [1] to investigate program-office heterogeneity across connectivity levels, and an edge weighting comparison [3] to reveal user role dynamics within the network.

The cumulative degree distribution was fitted with a power law with an exponential cutoff,

$$P(z) \propto z^{-\tau} e^{-z/z_c}, \tag{1}$$

then compared with published fits. The distribution was also examined using logarithmic degree bins [2].

To assess the accuracy with which different measures of user connectivity represent ties within the NERSC collaboration network, unweighted degree [1] was compared with repository-size-adjusted (RSA) strength. Unweighted degree was chosen to enable comparison with unweighted network data in the literature, while RSA strength accounts for the inflation of connection strength within large projects. Under RSA weighting, edge weights are inversely proportional to project size, and node strength is the sum of incident edge weights [3]. We evaluated the rankings by the proportion of career scientists—faculty, professional staff, scientists, and postdocs—among the highest-ranked users and by role composition across logarithmic bins.

III. Results

The cumulative NERSC graph's share of highly connected users is higher than that of any other scientific network considered. Qualitatively, the cumulative NERSC graph most closely resembles the HEP-focused SPIRES network (Fig. 1).

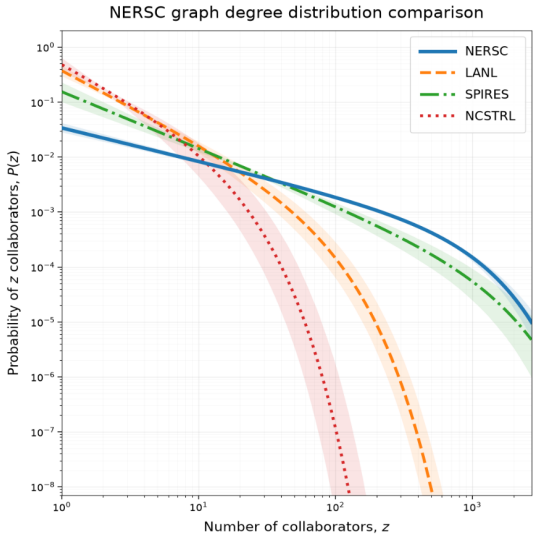


Fig. 1. NERSC graph degree distribution comparison.

Researchers who work on projects in the High-Energy Physics (HEP) program office dominate the tier of users with many connections. While 23.4% of all users appeared on a project in HEP, 96.1% of the top 5% of users by degree appeared on an HEP project (Fig. 2).

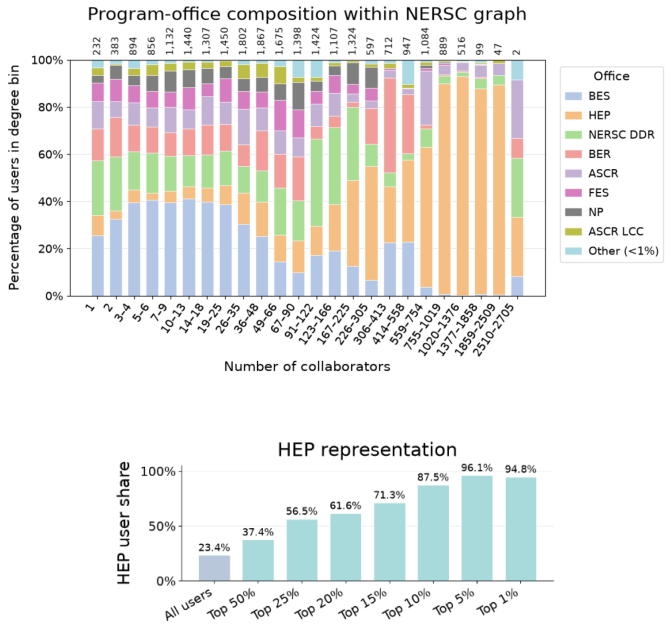


Fig. 2. Program-office composition within the NERSC graph and HEP representation by connectedness.

While career scientists make up 67 of the top 100 users by unweighted degree, they account for 86 of the top 100 users by RSA strength—a 19-percentage-point increase (Fig. 3).

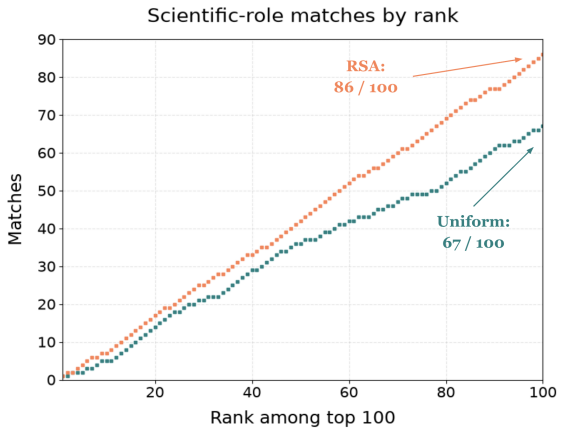


Fig. 3. Scientific-role matches by rank.

Role composition is largely stable across unweighted-degree bins but shifts toward career scientists as RSA strength increases (Fig. 4).

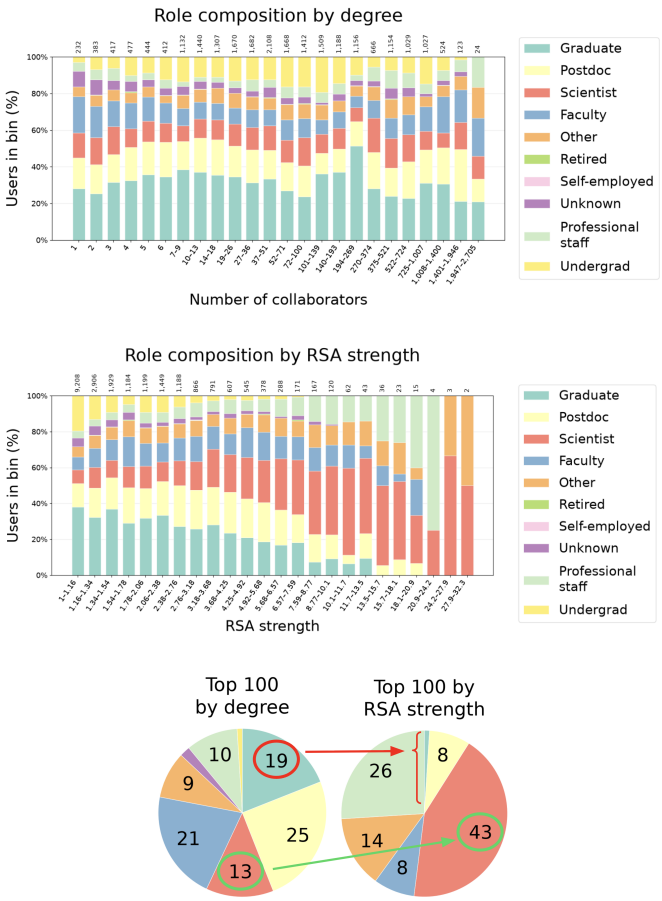


Fig. 4. Role composition by degree, role composition by RSA strength, and the top 100 users under each ranking.

IV. Conclusions and Future Work

Allocation-based network construction likely inflates the number of connections in a network, but allocation records

may reflect some field-specific patterns in how connectivity is distributed in coauthorship data. RSA strength is a more accurate measure of user connectivity than unweighted degree. Additionally, career scientists are the strongest connectors in the network, while students are among the weakest. Finally, users associated with HEP form more connections than users associated with any other program office. These findings suggest that in HPC network data, unweighted connectivity is not a reliable measure of network influence or cross-disciplinary reach.

Next steps include using RSA edge weighting in the degree distribution analysis focused on program offices to further investigate HEP dominance and analyzing the composition of code usage in degree bins.

ACKNOWLEDGMENTS

Gratitude goes to Kevin Gott and Rebecca Hartman-Baker for their generous mentorship, to NERSC staff for providing context and feedback, and to the Computing Sciences Summer Program for support.

REFERENCES

[1] M. E. J. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 404–409, 2001, doi: 10.1073/pnas.021544898.

[2] M. E. J. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review E*, vol. 64, Art. no. 016131, 2001, doi: 10.1103/PhysRevE.64.016131.

[3] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Physical Review E*, vol. 64, Art. no. 016132, 2001, doi: 10.1103/PhysRevE.64.016132.