How the NERSC User Community Connects

Finding communities, bridge projects, and candidates for outreach in allocation records

Sumaiya Azad[1], Camelia D. Brumar[2], Coby Sandholtz[3], Kevin Gott[4], Rebecca Hartman-Baker[4]

[1]University of Utah, [2]Tufts University, [3]Brigham Young University, [4]Lawrence Berkeley National Laboratory

INTRODUCTION

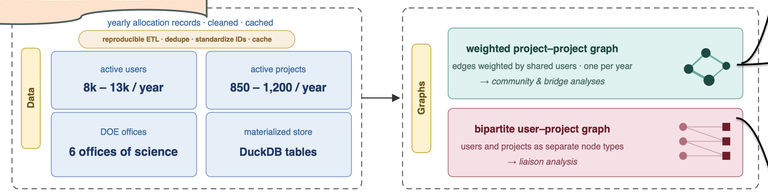
HPC facilities, like NERSC, support a large and diverse user base with a variety of knowledge, needs, support requirements, and expertise across the scientific community. Staff must decide who to train, who to warn before a system change, and where to focus limited support effort, but they have no map of who works with whom. Can available user data reveal that map?

Using anonymized allocation records from 2020 through April 2026, we construct user networks to explore whether actionable community structures can be found, such as whether communities are stable across offices and time, if key projects can be identified using bridge analysis and whether network analysis can reliably identify users acting as community liaison candidates for targeted outreach. Communities matter because their members share codes, queues, and failure modes: a need raised by one is likely shared by the rest.

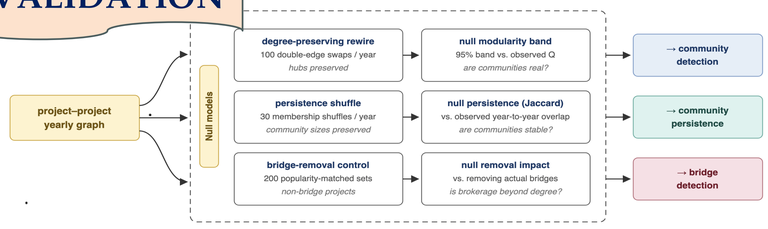
NETWORK CONSTRUCTION

We processed records with an ETL pipeline that standardizes project, user, and year identifiers across the analysis. Analyses used NetworkX [1]: Louvain community detection to find groups, modularity to measure how cleanly the network splits into them, betweenness centrality to find projects sitting on paths between many others, and Shannon entropy to measure how office-mixed a community is.

OVERVIEW

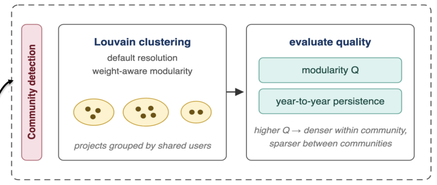


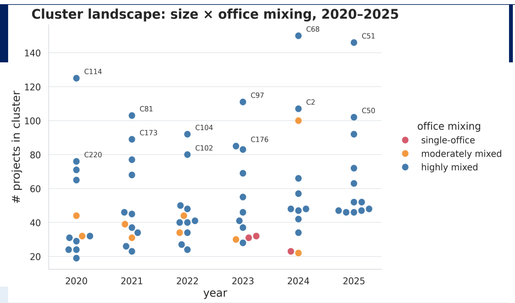
VALIDATION

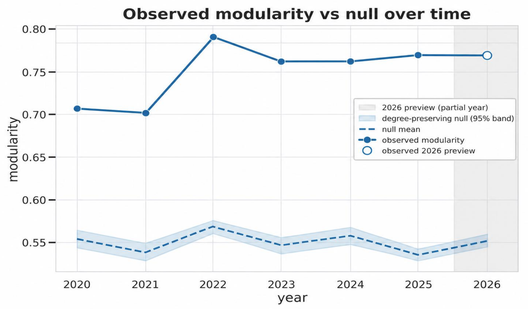


[1] Hagberg, Schult & Swart. Exploring network structure, dynamics, and function using NetworkX. SciPy, 2008.

COMMUNITY STRUCTURE





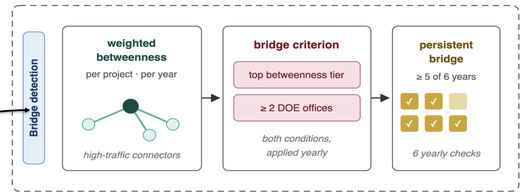


- Observed modularity stays above the 95% null band in every year, that is, the same network reshuffled at random, indicating that the detected **communities** are **real**.
- The resulting size distribution in every year is a **small number** of very **large communities**, a middle tier of medium-sized communities, and a longer tail of smaller ones.
- The **largest communities** are consistently **more office-mixed** than typical smaller communities, so major collaboration groupings do not simply follow office boundaries.
- The largest communities **remain recognizable** through 2025, while the shuffled membership baseline is near zero.
- More **cross-office communities** tend to have **higher GPU share**.



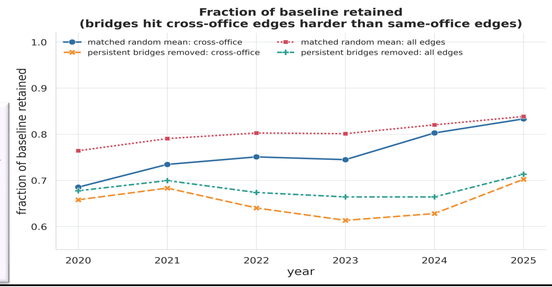
Project-sharing is **non-random**, cross-office communities remain **identifiable** across years.

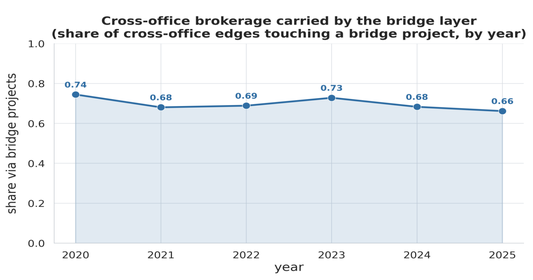
KEY BRIDGES



- About **10% of projects** meet the bridge definition in a given year, so cross-office connectivity is concentrated in a **small layer of bridge projects**; yet these projects touch **66-74% of cross-office edges**.
- According to the definition, **30 projects qualify** as **persistent bridges**, so the same bridge projects recur over time.

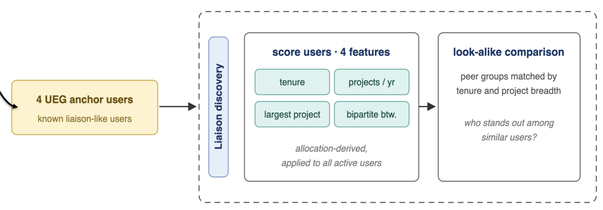
Removing persistent bridges **reduces cross-office connectivity more** than removing popularity-matched non-bridge projects, indicating that bridge importance is not explained by **degree alone**.

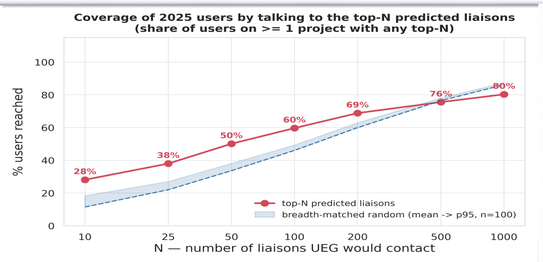


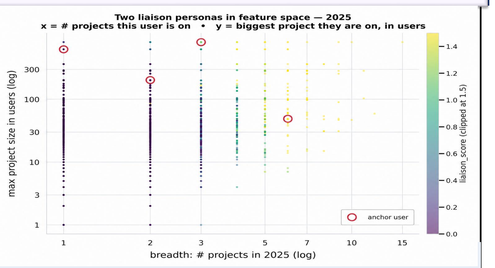


LIAISON DISCOVERY

A short, data-derived shortlist reaches most active users, enabling **data-targeted outreach**.







- The four anchor users separate into **two liaison types**: a **large-project-centered pattern** and a **multi-project bridge pattern**.
- The same idea that flags bridge projects surfaces a list of **top-ranked liaison candidates** from each **look-alike group** for efficient outreach
- In 2025, the top 50 liaison candidates reached ~50%, top **100** reached **~60%**, and top 200 reached ~69% of active users.

CONCLUSION

- Network analysis turns allocation records into a way to see how projects, users, and programs connect over time, which can support planning, collaboration, and outreach.
- Persistent bridge projects play an outsized role in maintaining cross-office connectivity.
- This analysis is useful if treated as an ongoing decision-support tool.

FUTURE STEPS

- Validating the liaison shortlist with the ticket and training data.
- Testing alternative community resolutions and weighted temporal network formulations.
- Packaging the workflow into a lightweight dashboard.

NOVELTY & CONTRIBUTION

- Application of community detection to HPC allocation data, expanding methods well established for scientific collaborations.
- Definition and validation of persistent bridge projects using null models and matched-popularity removal tests.
- Extension of the network framework to user-level liaison discovery.

ACKNOWLEDGMENTS

Supported by the Computing Sciences Summer Program at NERSC, LBL.
Special thanks to Kevin Gott and Rebecca Hartman-Baker for mentorship, and to the NERSC User Engagement Group for framing the problem.

ARTIFACT