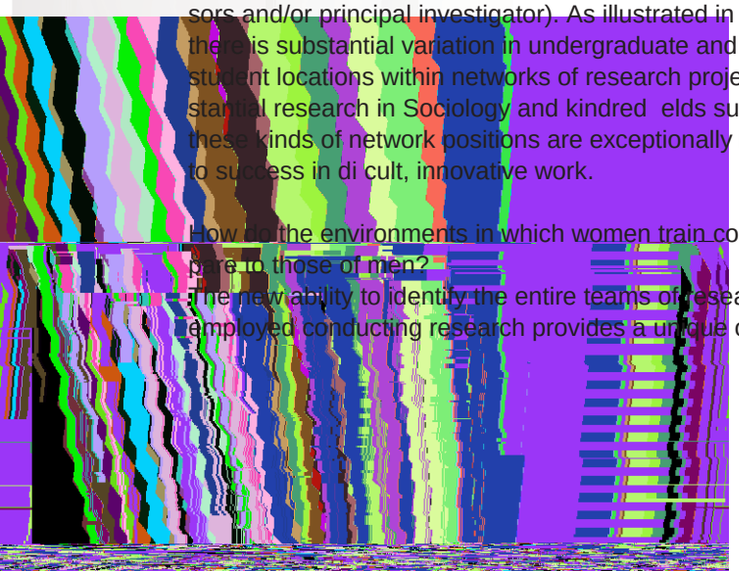
# UMETRICS as a Tool for Quantifying the Value of Research and Assessing Underrepresentation

Barbara McFadden Allen (Committee on Institutional Cooperation)
Julia I. Lane (New York University)
Rebecca Rosen (New York University)
Jason Owen Smith (University of Michigan)
Bruce A. Weinberg (Ohio State University, IZA, and NBER)

Figure 1 shows one faculty member (the large blue node), all of the people supported on grants with him, and all the people supported directly on grants with those people. In this image nodes represent individuals (red indicates female and blue indicates male) and shapes represent occupations. The  gure shows that among the faculty (circles) and graduate students (squares), many of the women are connected to each other and less central to this portion of the network. While this  gure represents just a small portion of the network at a single university, which may or may not be representative, such methods allow researchers to characterize the collaborative structure of science in very  ne detail. We argue

simply associating a principal investigator with a particular trainee. It is possible to quantify the number of faculty, sta , and postdoctoral researchers working with male and female undergraduate and graduate students on one or multiple projects throughout the students• training experience. It is also feasible to measure the gender, race, ethnicity, and national origin of trainees and whether they match advisors and/or principal investigator). As illustrated in Figure 1, there is substantial variation in undergraduate and graduate student locations within networks of research projects. Substantial research in Sociology and kindred elds suggests that these kinds of network positions are exceptionally important to success in di cult, innovative work.



How do the environments in which women train compare to those of men?

The new ability to identify the entire teams of researchers employed conducting research provides a unique opportunity to identify how the environments in which women train compare to those of men. Preliminary analysis shows that women graduate students are more likely to be employed on teams with other women and on grants with women as principal investigators, suggesting the potential for sizeable di erences in training environments.

The data also make it possible to compare training environments along a wide range of dimensions. For instance, researchers can examine whether women graduate students are employed on grants with a larger share of fellow graduate students or more sta or more faculty. Researchers can compare the number of grants on which women are supported, the type and source of these grants, the length of time, and

Documenting career pathways

The data allow new fundamental questions to be answered about how these various aspects of training environments relate to career outcomes, both in terms of career pathways and research production.

University data is being linked to strictly protected Census Bureau data on people and businesses to allow for the characterization of establishments (in academia, industry, or government) that hire people after they leave research, particularly the industries in which they operate, their geographic location (e.g. in the state in which the person trained or elsewhere), their size, age, growth, and wages.

The Census data also contain information on entrepreneurship. The Integrated Longitudinal Business Database (iLBD) combines administrative records and survey-based data for all nonfarm employer and nonemployer business units in the United States and hence provides information about the dynamics of rm growth and transitions from nonemployer to employer status

The people-centered approach emphasizes that where people go is critical for di using knowledge throughout the economy; the integration with Census data permit documentation of the extent to which research doctorates (and others employed on research projects) enter the broader economy and determine which aspects of the training environment matter for placement.

In particular, researchers can estimate how the careers paths of men and women compare to each other holding constant the rich characteristics of the training environment already discussed and also permit the identi cation of the long-term rami cations of any di erences in training environments. Are women more or less likely to obtain academic jobs versus go into industry? Are women who go into industry more or less likely to work at smaller rms or higher wage rms or more quickly growing rms or rms that are in di erent industries than observationally equivalent men? And, how much of any di erences can be explained by training?

Blume-Kohout nds that women supported on industry funded postdocs are more likely to participate in entrepreneurship.[3] The new ability to identify the mechanisms on which people were supported as graduate students and postdocs and then trace them through to subsequent activity can shed additional light on the decision to enter entrepreneurship and on success probabilities for a large number of researchers.

Research production

There are few economically important activities where the output of people are as readily available and as measurable as the journal articles that researchers publish. (Athletics might be another example.) The public nature of journal publications (and patents and public research funding) provide a rare opportunity to obtain fundamental estimates of how training environments relate to actual productivity. And the data are ideally suited to quantifying the research achievements of women and men, how they di er, and how any di erences close or widen over the career.

The sample frame based on people employed on grants, as opposed to people listed as coauthors on publications, is unique and particularly powerful way of studying the determinants of authorship. Speci cally, researchers can examine the publication patterns of all the people who were employed on a project as well as their jobs and time charged to it. In this way, researchers can quantify the extent to which women are less (or more) likely to appear on coauthors on articles and assess how the ordering of authors di ers controlling for a wide range of measures of involvement.

Conclusion

The new data infrastructure constitutes an important opportunity for breakthrough research on science and innovation that can inform many aspects of science policy. In addition to issues related to underrepresentation of women and other groups, they will support a wide range of analyses of the creation, transmission, and utilization of ideas and at an unprecedented level. They will rely on algorithmic, •big dataŽ methods to combine and mine data from a wide range of sources at low burden. And, the resulting, con dentiality protected, data will be made available to the research community through the newly founded Institute for Research on Innovation and Science (IRIS) (http://iris.isr.umich.edu0).

References

[1] http://www.cnet.com/news/women-in-tech-the-numbers-dont-add-up/#ftag=CADf328eec

[2] J. Robert Oppenheimer as quoted in Anonymous. •The Eternal ApprenticeŽ Time 52 (November 8, 1948): pp. 81.

[3] Margaret E. Blume-Kohout. 2014. •Understanding the Gender Gap in STEM Fields EntrepreneurshipŽ MBK Analytics, LLC.