THE UNIVERSITY OF GEORGIA
DEPARTMENT OF STATISTICS
Colloquium Series

4:00 PM, Thursday, November 17, 2022, via Zoom

https://zoom.us/j/94036794821?pwd=Qk14OEhaZjljbXBua2ErVkgyOURBUT09

Meeting ID: 940 3679 4821 Passcode: 185475

Dr. Rebecca Steorts
Department of Statistical Science, Duke University

Almost All of Entity Resolution

Whether the goal is to estimate the number of people that live in a congressional district, to estimate the number of individuals that have died in an armed conflict, or to disambiguate individual authors using bibliographic data, all these applications have a common theme—integrating information from multiple sources. Before such questions can be answered, databases must be cleaned and integrated in a systematic and accurate way, commonly known as structured entity resolution (record linkage or deduplication). Here, we review motivational applications and seminal papers that have led to the growth of this area. We review modern probabilistic and Bayesian methods in statistics, computer science, machine learning, database management, economics, political science, and other disciplines that are used throughout industry and academia in applications such as human rights, official statistics, medicine, and citation networks, among others. Last, we discuss current research topics of practical importance.

About the Speaker

Dr. Rebecca C. Steorts received her B.S. in Mathematics in 2005 from Davidson College, her MS in Mathematical Sciences in 2007 from Clemson University, and her PhD in 2012 from the Department of Statistics at the University of Florida under the supervision of Malay Ghosh. She was a Visiting Assistant Professor in 2012--2015, where she worked closely with Stephen E. Fienberg. She is currently an Associate Professor in the Department of Statistical Science at Duke University. Rebecca was named to MIT Technology Review's 35 Innovators Under 35 for 2015 as a humanitarian in the field of software. Her work was profiled in the September/October issue of MIT Technology Review and she was recognized at a special ceremony along with an invited talk at EmTech in November 2015. In addition, Rebecca is a recipient of a NSF CAREER award, a

with the Laboratory of Analytic (LAS) at NC State University, a Metaknowledge Network Templeton Foundation Grant, the University of Florida (UF) Graduate Alumni Fellowship Award, the U.S. Census Bureau Dissertation Fellowship Award, and the UF Innovation through Institutional Integration Program (I-Cubed) and NSF for development of an introductory Bayesian course for undergraduates. Rebecca was the recipient for Honorable Mention (second place) for the 2012 Leonard J. Savage Thesis Award in Applied Methodology. Her research interests are in large scale clustering, record linkage, privacy, network analysis, and machine learning for computational social science applications.