Abstract:
The growth of availability and variety of healthcare data sources has provided unique opportunities for healthcare data integration and evidence synthesis, which can potentially accelerate knowledge discovery and enable better clinical decision making. However, many practical and technical challenges, such as data privacy, high-dimensionality and heterogeneity across different datasets, remain to be addressed. In this talk, I present several methods for effective integration of electronic health records and other healthcare datasets. Specifically, we develop communication-efficient distributed algorithms for joint analyses of multiple datasets without the need of sharing patient-level data. Our algorithms do not require iterative communication across sites and are able to account for heterogeneity across different datasets. We provide theoretical guarantees for the performance of our algorithms, and examples of implementing the algorithms to real world clinical research networks, including the observational health data sciences and informatics (OHDSI) and the national patient-centered clinical research networks (PCORnet).