THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

TYLER H. MCCORMICK

Departments of Statistics and Sociology
University of Washington

Using Aggregated Relational Data to Feasibly Identify Network Structure Without Network Data

WEDNESDAY, February 19, 2020 at 4:00 PM
Jones 303, 5747 S. Ellis Avenue

ABSTRACT

Social connections are vital in understanding many health and economic outcomes. Examples include referral networks for jobs, peer influence in adopting/maintaining health behaviors, or informal insurance to mitigate financial shocks. Despite the critical role played by these connections, empirical research on networks is limited because collecting network data is financially or logistically prohibitive in many settings. Network mapping typically requires (1) enumerating a census, (2) eliciting the names of all network links for each individual, (3) matching the list of social connections to the census, and (4) repeating (1)-(3) across many networks. In settings requiring field surveys, steps (2)-(3) can be very expensive. In populations likely to face discrimination or stigma, privacy concerns may render (2)-(3) impossible. Both restrict the accessibility of high-quality networks research to investigators with considerable resources. We propose an inexpensive and feasible strategy for network elicitation using Aggregated Relational Data (ARD) -- responses to questions of the form "How many of your social connections have trait k?" Our method uses ARD to recover the parameters of a general network formation model, which in turn, permits the estimation of any arbitrary node- or graph-level statistic. The method works well in simulations and in matching a range of network characteristics in real-world graphs from 75 Indian villages. Moreover, we replicate the results of two field experiments that involved collecting network data. We show that the researchers would have drawn similar conclusions using ARD alone. We further evaluate the method by formally characterizing the class of statistics that we expect to recover with high probability using ARD. Finally, using calculations from fieldwork done by the Abdul Latif Jameel Poverty Action Lab, we show that in rural India, for example, ARD surveys are 80% cheaper than full network surveys.