

*Optimal design of experiments in the presence of network interference*

Edoardo M. Airoidi  
Associate Professor of Statistics  
Director of Graduate Studies  
Harvard University

Causal inference research in statistics has been largely concerned with estimating the effect of treatment (e.g. personalized tutoring) on outcomes (e.g., test scores) under the assumption of "lack of interference"; that is, the assumption that the outcome of an individual does not depend on the treatment assigned to others. Moreover, whenever its relevance is acknowledged (e.g., study groups), interference is typically dealt with as an uninteresting source of variation in the data. In many applications, however, from quantifying the influence of peers on learning, to understanding the role of social relations on the success of health intervention in rural villages in Africa and South America, from word-of-mouth advertising and viral marketing, to political campaigns aimed at increasing volunteering, donations, and supporters turnout on election day, and from estimating the effects of accidents on congestion in transportation networks, to assessing how social structure affects labor market dynamics, the lack of interference assumption is not tenable. Not only interference is present in these situations, and is an important aspect of the problem that cannot be abstracted away, but we are often interested in estimating the causal effect of such interference. In this talk, we present statistical methodology for working with interference. We will review challenges in defining the inferential target, review assumptions that facilitate estimation, and state alternative problem formulations for situations when the available network data is not well aligned with the target notion of interference. We will then introduce a two-stage strategy to define an optimal set of randomizations for estimating interference on large social and information networks.