Estimation of causal effects in network-dependent data

Oleg Sofrygin[1,*]**, Mark J. van der Laan**[1]

1. University of California, Berkeley

*Contact author: sofrygin@berkeley.edu

We describe two *R* packages which facilitate causal inference research in network-dependent data: **simcausal** package for conducting network-based simulation studies; and **tmlenet** package for the estimation of various causal effects in **simcausal**-simulated, or real-world network datasets. In addition to the estimation of various causal effects, the **tmlenet** package implements several approaches to estimation of standard errors for dependent (non-IID) data with known network structure. Both packages implement a new syntax that repurposes the list indexing operator '`[[...]]`' for specifying complex network-based data summaries of the observed covariates. For example, `sum(A[[1:Kmax]])` will specify a network-driven summary, evaluated for each unit i as a sum of the variable `A` values for all "friends" of i. This new syntax is fully generalizable towards any type of user-defined functions and any type of networks. The practical applicability of both packages is then illustrated with a large-scale simulation study of a hypothetical highly-connected community with an intervention that aimed to increase the level of physical activity by (i) educating a simulated study population of connected subjects, and/or (ii) by intervening on the network structure itself. We will describe how our work can be extended to complex network processes that evolve over time, and discuss possible avenues for future research on estimation of causal effects in longitudinal network settings.