The Impact of Incomplete Data on Inferred Network Topology

Catherine A. Bliss

Department of Mathematics & Statistics, Center for Complex Systems
& the Vermont Advanced Computer Center, University of Vermont, Burlington, VT 05405

February 12, 2011

ABSTRACT Complex networks underlie a variety of social, biological, physical, and virtual systems. A formidable, unsolved problem is that network data is often incomplete: in many applications, it is impossible to observe all interactions between members of a population. Previous work on this problem, which is surprisingly limited, shows that subsamples of network data, in the form of missing nodes, missing links, or missing interactions, can lead to gross over- or under- estimates of network statistics describing the overall network topology. Our aim is to explore and generate scaling methods for inferring the 'true' topology of a network given varying degrees of subsampling. In testing our approaches, we employ data from both simulated networks and Twitter, a rich and evolving large-scale social network.