Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem

Eric D. Kolaczyk*
Boston University, Boston, MA, USA kolaczyk@bu.edu

Yaonan Zhang
Boston University, Boston, MA, USA

Bruce D. Spencer
Northwestern University, Evanston, IL, USA

Networks are a popular tool for representing elements in a system and their interconnectedness. Many observed networks can be viewed as only samples of some true underlying network. We study the problem of how to estimate the degree distribution of a true underlying network from its sampled network, under various common network sampling designs. We show that it can be formulated as an inverse problem that is, in many cases, ill-posed. Accordingly, we offer a penalized least-squares approach to solving this problem, with the option of additional constraints. The resulting estimator is a linear combination of singular vectors of a matrix, relating the expectation of our sampled degree distribution to the true underlying degree distribution, which is defined entirely in terms of the sampling plan. Choice of the penalization parameter is made through a Monte Carlo version of Stein's unbiased risk estimation. We present the results of a simulation study, characterizing the performance of our proposed method, and we illustrate its use in the context of monitoring large-scale social media networks.

Key Words: Network inference, penalized least squares, social media networks.