Statistical Inference on Errorfully Observed Graphs

Carey E. Priebe
Johns Hopkins University
Department of Applied Mathematics and Statistics
Baltimore, Maryland 21218-2682 USA

Statistical inference on graphs is a burgeoning field in the applied and theoretical statistics communities, as well as throughout the wider world of science, engineering, business, etc. In many applications, we are faced with the reality of errorfully observed graphs. That is, the existence of an edge between two vertices is based on some imperfect assessment. In this paper, we consider a graph $G = (V, E)$. We wish to perform an inference task – the surrogate inference task considered here is “vertex classification”. However, we do not observe $G$; rather, for each potential edge $uv \in (V^2)$ we observe an “edge-feature” which we use to classify $uv$ as edge/not-edge. Thus we errorfully observe $G$ when we observe the graph $\tilde{G} = (V, \tilde{E})$. Moreover, we face a quantity/quality trade-off regarding the edge-features we observe – more informative edge-features are more expensive, and hence the number of potential edges that can be assessed decreases with the quality of the edge-features. We derive the optimal quantity/quality operating point for subsequent graph inference in the face of this trade-off.