Mathematics, Statistics, and Computer Science   **@ UIC**

Computer Science Seminar

Learning Communities in the Presence of Errors

Aravindan Vijayaraghavan (Northwestern)

Abstract: The Stochastic Block Model or the Planted Partition Model is the most widely used probabilistic model for community detection and clustering graphs in various fields, including machine learning, statistics, and social sciences. Many existing algorithms successfully learn the communities or clusters when the data is drawn exactly according to the model, but they do not work well in the presence of errors. In this talk, I will address the following question: Can we design robust polynomial time algorithms for learning probabilistic models for community detection that work in the presence of adversarial modeling errors?

I will present robust algorithms for (partially) recovering communities or clusters in graphs drawn from the Stochastic Block Model, in the presence of modeling errors or noise. These algorithms allow two types of adversarial errors: edge outlier errors (where an adversary can corrupt an arbitrary \epsilon fraction of the edges), and any amount of Feige-Kilian or monotone errors. Mossel, Neeman and Sly (STOC 2015) posed an open question about whether an almost exact recovery is possible when the adversary is allowed to add $o(n)$ edges. Our work answers this question affirmatively even in the case of $k>2$ communities.

Monday, May 2 at 3:00 PM in SEO 612

Finally, I will describe how our algorithms recover the clusters, even when the modeling error is captured using Kullback-Leibler (KL) divergence: these algorithms work when the instances come from any distribution of graphs that is \epsilon.m close to the Stochastic Block Model in the KL divergence (this result also handles adversarial errors).

This is based on joint work with Konstantin Makarchev and Yury Makarychev.

Monday, May 2 at 3:00 PM in SEO 612