Stanford University
Department of Statistics

DEPARTMENTAL SEMINAR

4:15pm, Tuesday, February 14, 2012
Sequoia Hall Room 200
Cookies served at 3:45pm, 1st Floor Lounge.

Speaker: Michael W. Mahoney, Stanford Mathematics

Title: Approximate computation and implicit regularization in large-scale data analysis

Abstract:

Statisticians and computer scientists adopt very different perspectives on algorithms and data. A concept that lies at the heart of this disconnect is that of regularization, a notion that has to do with how robust is the output of an algorithm to the noise properties of the input. Although it is nearly completely absent from computer science, which historically has taken the input data as given and modeled algorithms discretely, regularization in one form or another is central to nearly every application domain that applies algorithms to noisy data. Typically, it is implemented by adding some sort of norm constraint to an objective function and then exactly optimizing the modified objective function. In this talk, I will describe two examples illustrating how approximate computation, in and of itself, can implicitly lead to regularization.

The first example is a very precise theoretical characterization of how approximation procedures like computing the PageRank of a graph (a problem that originally arose to rank nodes in Web-scale graphs) exactly solve a regularized version of the problem of computing the leading nontrivial eigenvector of a graph Laplacian. Although this is a graph problem, it can be given a statistical interpretation (analogous to the manner in which Ridge regression and Lasso regression can be interpreted in terms of a Gaussian prior or a Laplace prior) which I will also describe. The second example is an empirical illustration of how spectral and flow-based approximation algorithms for computing approximations to the intractable graph partitioning problem (a problem that arises when trying to find communities in small and large graphs) implicitly provide more regular solutions than would be provided by an algorithm that exactly solved the intractable problem. These two examples grew out of our large-scale empirical analysis of the properties of large social and information networks, and so I will explain how understanding the statistical properties implicit within scalable worst-case approximation algorithms was essential for obtaining our downstream conclusions.