Matrix Computations & Scientific Computing Seminar

Organizer: James Demmel & Ming Gu

Wednesday, 11:00AM–12:00Noon, 380 Soda

Oct. 30   **Ravi Kannan**, Microsoft India
                   Algorithms for Distributed Data

We consider problems where the input data is split across many servers. We design algorithms for three common problem areas for massive data problems. (i) We solve Frequency Moment problems with low communication -three rounds of constant number of keys - with a simple importance sampling technique. Our results contrast with known lower bounds for the Streaming Model. (ii) Using the recently developed technique of subspace embeddings, we design algorithms for low rank approximations. (iii) We design algorithms for Clustering, in particular for an effective algorithm we call "Project and Cluster". We briefly compare resource bounds for our algorithms with other distributed models.

Joint work with Santosh Vempala and David Woodruff