



The University of Chicago
Department of Statistics
Statistics Colloquium Series

INDERJIT S. DHILLON

Department of Computer Science
University of Texas at Austin

Fast and Memory-efficient Dimensionality Reduction of Massive Graphs

MONDAY, October 17, 2011, at 4:00 PM

133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110.

ABSTRACT

In this talk, I will present a fast and memory-efficient procedure for the dimensionality reduction of massive graphs. The procedure involves a fast clustering of the graph followed by approximation of each cluster separately using existing methods, e.g. the singular value decomposition, or stochastic low-rank approximations. The clusterwise approximations are then combined to approximate the entire graph. This approach has several benefits: (1) important structure of the graph is preserved due to the clustering; (2) accurate low rank approximations are achieved; (3) the procedure is efficient both in terms of computational speed and memory usage. A salient feature of the approximation scheme is that it attempts to preserve the sparsity of the input graph. The recently introduced stochastic algorithms can be incorporated into our clustered low rank approximation framework to get theoretical bounds for the approximation error. To conclude, I will present experimental results of our algorithms on social networks as large as 3.8 million nodes and 65 million edges.

This is joint work with Berkant Savas.

For further information and inquiries about building access for persons with disabilities, please contact Dan Moreau at 773.702.8333 or send him an email at dmoreau@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website:
<https://lists.uchicago.edu/web/arc/statseminars>.