title: ””

Announcements

- Lecture notes 2 & 3 updated; Homework 1 posted. —-

Course Description

Designing algorithms for efficient processing of large data sets poses unique challenges. This course will discuss algorithmic paradigms that have been developed to efficiently process data sets that are much larger than available memory. We will cover streaming algorithms and sketching methods that produce compact data structures, dimension reduction methods that preserve geometric structure, efficient algorithms for numerical linear algebra, graph sparsification methods, as well as impossibility results for these techniques.

Tentative Syllabus

- **Sketching and Streaming** algorithms for basic statistics:
  - Distinct elements, heavy hitters, frequency moments, p-stable sketches
- **Dimension Reduction**
  - Johnson Lindenstrauss lemma, lower bounds and impossibility results
- **Graph stream algorithms**
  - Connectivity, cut/spectral sparsifiers, spanners, matching, graph sketching
- **Lower bounds** for Sketching and Streaming:
  - Communication complexity: Equality, Index and Set-Disjointness
- **Locality Sensitive Hashing**
  - Similarity estimation, approximate nearest neighbor search, data dependent hashing
- **Fast Approximate Numerical Linear Algebra**
  - Matrix multiplication, low-rank approximation, subspace embeddings, least squares regression
- **Massively Parallel Computing Models**
  - MST, connected components, matching, and submodular optimization in the MapReduce model

Previously numbered CS 369G: web.stanford.edu/class/cs369g/