Algorithms and Models for Big Data
Fall 2014

Course outline

Instructor: Periklis Papakonstantinou, e-mail: `papakons@tsinghua.edu.cn`, office: FIT 4-608-7

Teaching Assistant: Yuanxi Dai

Lectures: Monday 1.30-4.05pm (3x45min) @ Tsinghua Xuetang 112

Web-page: `http://iiis.tsinghua.edu.cn/~papakons/teaching/bigdataF14/`

Audience

Senior-level undergraduate research-oriented course in the intersection of advanced mathematical techniques with analysis and management of very large amounts of data.

Course description, objectives & prerequisites

Recently the buzzword "Big Data" appears everywhere from seminar lectures to popular books. Theoretical Computer Science and Statistics, and especially over the last 20 years, has developed a number of tools and deep algorithmic and mathematical machinery that puts the topic in the "right" context. A concrete connection with the engineering aspect has not been established yet, although practicioners have independently shown remarkable developments (e.g. map-reduce). This course discusses extracts from the well-developed fields of: (i) property testing and (ii) Fourier analysis of the boolean cube, (iii) computing over data-streams, (iv) finite metric embeddings, (v) high-dimensional statistics and connections to geometry, and if time permits (vi) large scale randomized numerical techniques. *All these will be presented at the appropriate (undergraduate) level starting from elementary principles.* At the same time emphasis will be given in connections with the actual real-world problems. Big Data analytics and management is not a purely theoretical topic and it cannot be presented in isolation to the real-world problems. In particular, some of the term-projects given in this class specifically target the interface between these advanced mathematical machinery and the existing systems and architectures.

Remark. The material from Statistics, Fourier Analysis, and High-dimensional geometry will be presented at an introductory level before each of the applications.

Grading

Here is the breakdown of the grade:

- **3-4 assignments: 30%**

- **Term project (including oral exam): 70%**

* selected projects among those that are in the interface between system engineering and theory will be presented at the end of the class – I anticipate that these will constitute new acquired practical knowledge.

Collaboration, missed assignments & remarking requests

- The assignments should be done individually by each student. You are not only allowed but also encouraged to form study groups. Your assignment report must be prepared solely by you (avoid plagiarism).
  What type of collaboration is not considered plagiarism: during your meetings to collaborate for an assignment (i) no electronic collaboration is allowed (you can only meet in person), (ii) you should not discuss the very details of the solutions, and (iii) you are not allowed to take any transcript out of your meeting; i.e. you cannot take with you any notes or any form of electronic record. Then, you let some time (e.g. one hour) pass in between this meeting and you starting preparing your report.

- No late assignments accepted. If there is an acceptable and well-documented reason I will arrange for redistribution of marks.