Fall 2018: Minicourse: High dimensional probability in data analysis

Instructor: Xiuyuan Cheng   Email: xiuyuan.cheng@duke.edu
Time: 10/29/2018 - 11/30/2018, MW 4:40PM - 5:55PM
Classroom: Gross 318

The minicourse will follow Ch. 4, 5 of the recent book R. Vershynin,

"High-Dimensional Probability: An Introduction with Applications in Data Science"
(https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf)

and selected topics in Ch. 7-10.

The course will start with preliminaries of random variables and the epsilon-net argument for analyzing the spectrum of a large random matrix. We then go to the isoperimetric inequality and the concentration results implied (e.g. that of Lipschitz functions of Gaussian random vectors or those uniformly distributed on a high-dimensional sphere). We will also introduce counterpart results for independent sums of random matrices instead of scalar random variables. These results have direct applications in data analysis and computation, including spectral clustering, community detection, the analysis of random projections applied to high-dimensional vectors, randomized linear algebra and so on, which will be discussed. Time permitting, we will briefly discuss techniques based on random processes (Slepian's inequality, Sudakov-Fernique, Talagrand's comparison) and how they can obtain new estimates as well as improve some previous ones.

Other references:
[1] M. Ledoux, "The concentration of measure phenomenon"
[2] J. Tropp, "An Introduction to matrix concentration inequalities"