Seminar on Combinatorial Computing
April 2, Wednesday, 6:30 p.m.
Room 6417, Graduate Center
365 Fifth Avenue, New York

Nearest neighbor search in high dimensions

Laura Heinrich-Litan
Freie Universität, Berlin

Abstract

We consider the *nearest-neighbor problem*, which is defined as follows: given a fixed set P of n data points in some metric space X, build a data structure such that for each given *query point* q a data point from P closest to q can be found efficiently. The underlying metric space is usually the d-dimensional real space $\mathbf{R}^d$ together with one of the L_p-metrics, $1 \leq p \leq \infty$. In many applications, the dimension d of the search space is quite high and can reach several hundreds or even several thousands. Therefore, running times and storage requirements exponential in d are prohibitive in these cases. Because of their exponential dependence on the dimension, all known techniques for exact nearest-neighbor problem are in fact in high dimensions not competitive with the brute-force method, which just determines the distance of q to each point in P and selects the minimum. This talk presents algorithms for solving the high-dimensional exact nearest-neighbor problem with respect to the L_∞-distance. These algorithms considerably improve the brute-force method, they are simple and easy to implement, and have a linear storage requirement.

For further information contact Janos Pach at pach@cims.nyu.edu,
or visit our website: http://math.gc.cuny.edu/Seminars/Combinatorics.html