Data Depth

Questions from Statistics have inspired fundamental research in Computer Science. One example involves the notion of data depth - depth defined via a given set of points. Let $a_1, \ldots, a_n$, $a_i \in R$ be a set of n distinct inputs. The points divides the real line into $n+1$ intervals and the depth of $x \in R$ is defined to be

$$d(x) = \min(|\{a_i : a_i \leq x\}|, |\{a_i : a_i \geq x\}|),$$

the fewest number of intervals that x meets as $|x| \to \infty$. A median is a point μ of maximal depth. It is familiar that a median has depth $\lfloor (n+1)/2 \rfloor$ and that it may be computed in linear time.

Many applications require an analogous notion for the depth of points $x \in R^d$. Several interesting generalizations of one-dimensional depth have been suggested. They pose a variety of challenges to computer science. On the algorithmic side, it is important to understand the complexity of computations involving depth, and to have efficient algorithms for these tasks. At the same time the various depth notions lead to combinatorial questions about arrangements of points and hyperplanes in $d-$dimnsional space.

I will describe several depth notions, some recent results about them (algorithms, complexity, and combinatorics), and mention open problems that remain. No prior background will be needed.

Biography:

William Steiger is Professor of Computer Science at Rutgers University. He received a PhD in Statistics from the Australian National University, where he was also Lecturer in Pure Mathematics. He has held visiting positions at the University of Sydney, University of Bonn, Polytechnic University of Catalunya, University of California at Berkeley, Princeton University, and Hong Kong University of Science and Technology. His research interests include discrete and computational geometry, probability, and the probabilistic method.