Sublinear Time and Space Algorithms 2016B – Problem Set 2

Robert Krauthgamer

Due: May 22, 2016

General instructions: Please keep your answers short and easy to read. You can use results, calculations or notation seen in class without repeating them, unless asked explicitly to redo them.

1. We saw in class how to construct a pairwise independent family H of hash functions $h : [n] \to [M]$, when $M \geq n$ is a prime. Specifically, each hash function was of the form $h_{p,q}(i) = pi + q \pmod{M}$, and can be stored using $\log_2 |H| = O(\log M)$ bits.

   Extend this approach to construct a k-wise independent family H_k for any $3 \leq k \leq n$. How many bits are needed to store a hash function?

   Hint: Use higher-degree polynomials, and rely on the determinant of a Vandermonde matrix.

2. Suppose the input is a dynamic graph (stream of edge insertions and deletions) on vertex set $V = [n]$. Design a streaming algorithm with storage requirement $\tilde{O}(n)$ to determine (with high probability) whether the graph is bipartite.

   Hint: Use the connectivity algorithm to propose a bipartition (2-coloring) of the vertices, and an independent ℓ_0-sampler to find an edge that "violates" this bipartition.

3. We saw in class a streaming algorithm for graph connectivity in the model of edge insertions and deletions.

   Extend this algorithm (and its analysis) to the st-connectivity problem, where at the end of the stream, the algorithm is given two vertices $s, t \in [n]$ and should report whether they are connected in the current graph.

Extra credit:

4. The Precision Sampling Lemma seen in class used precisions $u_i \sim \text{Exp}(1)$, i.e., drawn from exponential distribution. Would the same (or similar) conclusion hold if we used instead precisions $u'_i \sim U[0, 1]$, i.e., drawn from uniform distribution?

   Hint: First check if u'_i is stochastically dominated by u_i.