CS 671: Graph Streaming Algorithms and Lower Bounds **Rutgers: Fall 2020**

Problem set 5

Due: 11:59PM, October 13, 2020

Problem 1. The arboricity of an undirected graph $G = (V, E)$, denoted by $\alpha(G)$, is the minimum number of spanning forests needed to cover all the edges of G. By Nash-Williams theorem, we alternatively have:

$$\alpha(G) = \max_{U \subseteq V} \left\lceil \frac{|E(U)|}{|U| - 1} \right\rceil, \qquad (1)$$

where $E(U)$ is the set of edges with both endpoints in U.

The goal of this problem is to design a semi-streaming algorithm in *dynamic* streams for $(1\pm\varepsilon)$-approximation of the arboricity (albeit in exponential time). Throughout this problem, let n and m denote the number of vertices and edges, respectively, and assume $\varepsilon \in (0, 1)$ is a sufficiently small constant.

(i) Suppose $m \geq c \cdot \varepsilon^{-2} \cdot n \log n$ for a sufficiently large constant $c > 0$. Let $p := \Theta(\varepsilon^{-2} \cdot \log n) \cdot \frac{n}{m}$. Prove that if H is obtained from G by sampling each edge independently with probability p, then $\alpha(H)$ is a $(1 \pm \varepsilon)$-approximation of $\alpha(G)$.

Hint: Use the formula for $\alpha(G)$ in Eq (1). Prove that the "density" of any set U in H, i.e., $|E_H(U)|/|U| - 1$ is a good "proxy" for its density in G—this claim can only be true for sets U with sufficiently large density (i.e., at least $m/(n-1)$); for other choices of U, prove that this value is sufficiently smaller than the new density (in H) of the originally high density sets (in G).

(ii) Use part (i) to design a semi-streaming algorithm for estimating $\alpha(G)$ to within a $(1\pm\varepsilon)$-approximation in dynamic streams.

Hint: You should find a way of using ℓ_0-samplers to "simulate" sampling *every* edge independently and uniformly at random with probability p as in part (i).