

Pan-Private Graph Streaming Algorithms

CODY FREITAG¹

DIMACS REU

Summer 2016

¹Supported by the NSF

Overview

Streaming Algorithms

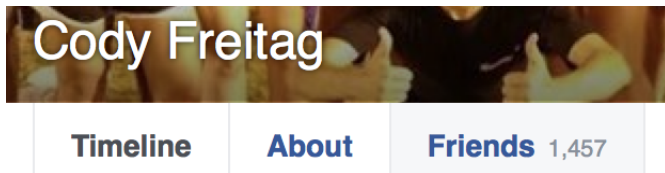
Pan-Privacy

My Project

Streaming Algorithms

- ▶ Data: $\vec{a}(t) = [a_1(t), \dots, a_n(t)]$
 - ▶ initially, $\vec{a}(0) = [0, \dots, 0]$
- ▶ Updates: (i_t, c_t)
 - ▶ $a_{i_t}(t) = a_{i_t}(t-1) + c_t$
- ▶ Query: estimate some function $f(\vec{a})$
 - ▶ Approximate: output is ε -close to the true value
 - ▶ Randomized: δ probability of success
- ▶ Goal: design an algorithm for various functions f that requires sublinear ($O(\text{polylog}(n))$) storage, update time, and query time

Streaming Algorithms – Facebook Friends



- ▶ Data: a map from Facebook users to number of friends.
- ▶ Updates: (Cody Freitag, +1) or (Cody Freitag, -1)
- ▶ Point Query: estimate how many friends Cody Freitag has.
- ▶ Heavy Hitters Query: output the people with lots of friends.

Streaming Algorithms – Graphs



- ▶ Data: a vector representation of an adjacency matrix
- ▶ Updates: $(\{u, v\}, +1)$ or $(\{u, v\}, -1)$
- ▶ Connectivity Query: is the graph connected?
- ▶ Semi-Streaming: can use $O(|V|\text{polylog}|V|)$ space.
 - ▶ Most graph queries have provable $\Omega(|V|)$ lower bounds.

Pan-Privacy

Informal Definition:

- ▶ Two data streams are “close” iff they only differ in updates for a single index.
- ▶ A streaming algorithm is *pan-private* iff for any two “close” data streams it takes as input, a computationally unbounded intruder can’t distinguish the internal state or outputs for each stream.

Pan-Privacy – Drug Testing Athletes



- ▶ Data: a map from MLB players to drug test results
- ▶ Updates: (Alex Rodriguez, true) or (Yu Darvish, false)
- ▶ Density Query: estimate the fraction of MLB players using PEDs.
- ▶ Want to encourage participation by guaranteeing information-theoretic privacy for participants.

My Project

- ▶ Formulation: understand pan-privacy for graph streaming algorithms.
- ▶ Upper Bounds: develop pan-private algorithms for graph queries.
- ▶ Lower Bounds: explore the limitations of pan-privacy on graph problems.

Questions?