Abstract

We study the problem of estimating the maximum matching *size* in graphs whose edges are revealed in a streaming manner. We consider both *insertion-only* streams, which only contain edge insertions, and *dynamic* streams that allow both insertions and deletions of the edges, and present new upper and lower bound results for both cases.

On the upper bound front, we show that an α-approximate estimate of the matching size can be computed in dynamic streams using $\tilde{O}(n^2/\alpha^4)$ space, and in insertion-only streams using $\tilde{O}(n/\alpha^2)$-space. These bounds respectively shave off a factor of α from the space necessary to compute an α-approximate matching (as opposed to only size), thus proving a non-trivial separation between approximate estimation and approximate computation of matchings in data streams.

On the lower bound front, we prove that any α-approximation algorithm for estimating matching size in dynamic graph streams requires $\Omega(\sqrt{n}/\alpha^{2.5})$ bits of space, *even* if the underlying graph is both *sparse* and has *arboricity* bounded by $O(\alpha)$. We further improve our lower bound to $\Omega(n/\alpha^2)$ in the case of *dense* graphs. These results establish the first non-trivial streaming lower bounds for *super-constant* approximation of matching size.

Furthermore, we present the first *super-linear* space lower bound for computing a $(1+\epsilon)$-approximation of matching size *even* in insertion-only streams. In particular, we prove that a $(1+\epsilon)$-approximation to matching size requires $\mathsf{RS}(n) \cdot n^{1-O(\epsilon)}$ space; here, $\mathsf{RS}(n)$ denotes the maximum number of edge-disjoint *induced matchings* of size $\Theta(n)$ in an n-vertex graph. It is a major open problem with far-reaching implications to determine the value of $\mathsf{RS}(n)$, and current results leave open the possibility that $\mathsf{RS}(n)$ may be as large as $n/\log n$. Moreover, using the best known lower bounds for $\mathsf{RS}(n)$, our result already rules out any $O(n \cdot \text{poly}(\log n/\epsilon))$-space algorithm for $(1+\epsilon)$-approximation of matchings. We also show how to avoid the dependency on the parameter $\mathsf{RS}(n)$ in proving lower bound for dynamic streams and present a near-optimal lower bound of $n^{2-O(\epsilon)}$ for $(1+\epsilon)$-approximation in this model. Using a well-known connection between matching size and *matrix rank*, all our lower bounds also hold for the problem of estimating matrix rank. In particular our results imply a near-optimal $n^{2-O(\epsilon)}$ bit lower bound for $(1+\epsilon)$-approximation of matrix ranks for dense matrices in dynamic streams, answering an open question of Li and Woodruff (STOC 2016).