Abstract

The famous shortest path tree lemma states that, for any node s in a graph $G = (V, E)$, there is a subgraph on $O(n)$ edges that preserves all distances between node pairs in the set $\{s\} \times V$. A very basic question in distance sketching research, with applications to other problems in the field, is to categorize when *else* graphs admit sparse subgraphs that preserve distances between a set P of p node pairs, where P has some different structure than $\{s\} \times V$ or possibly no guaranteed structure at all. Trivial lower bounds of a path or a clique show that such a subgraph will need $\Omega(n+p)$ edges in the worst case. The question is then to determine when these trivial lower bounds are sharp; that is, when do graphs have *linear size distance preservers* on $O(n + p)$ edges? In this paper, we make the first new progress on this fundamental question in over ten years. We show:

1. All G, P has a distance preserver on $O(n)$ edges whenever $p = O(n^{1/3})$, even if G is directed and/or weighted. These are the first nontrivial preservers of size $O(n)$ known for directed graphs.
2. All G, P has a distance preserver on $O(p)$ edges whenever $p = \Omega\left(\frac{n^2}{rs(n)}\right)$, and G is undirected and unweighted. Here, $rs(n)$ is the Ruzsa-Szemerédi function from combinatoric graph theory. These are the first nontrivial preservers of size $O(p)$ known in any setting.
3. To preserve distances within a subset of s nodes in a graph, $\omega(s^2)$ edges are sometimes needed when $s = o\left(\frac{n^{2/3}}{2^{\Theta(\sqrt{\log n \cdot \log \log n})}}\right)$ even if G is undirected and unweighted. For weighted graphs, the range of this lower bound improves to $s = o(n^{2/3})$. This result reflects a polynomial improvement over lower bounds given by Coppersmith and Elkin (SODA '05).

An interesting technical contribution in this paper is a new method for "lazily" breaking ties between equally short paths in a graph, which allows us to draw our new connections between distance sketching and the Ruzsa-Szemerédi problem.