

Sublinear Algorithms and Lower Bounds for Estimating MST and TSP Cost in General Metrics

Tuesday, July 11, 2023 4:20 PM (20 minutes)

Sanjeev Khanna, Yu Chen and Zihan Tan

Abstract: We consider the design of sublinear space and query complexity algorithms for estimating the cost of a minimum spanning tree (MST) and the cost of a minimum traveling salesman (TSP) tour in a metric on n points. We start by exploring this estimation task in the regime of $o(n)$ space, when the input is presented as a stream of all $\binom{n}{2}$ entries of the metric in an arbitrary order (a metric stream). For any $\alpha \geq 2$, we show that both MST and TSP cost can be α -approximated using $\tilde{O}(n/\alpha)$ space, and moreover, $\Omega(n/\alpha^2)$ space is necessary for this task. We further show that even if the streaming algorithm is allowed p passes over a metric stream, it still requires $\tilde{\Omega}(\sqrt{n/\alpha p^2})$ space.

We next consider the well-studied semi-streaming regime. In this regime, it is straightforward to compute MST cost exactly even in the case where the input stream only contains the edges of a weighted graph that induce the underlying metric (a graph stream), and the main challenging problem is to estimate TSP cost to within a factor that is strictly better than 2. We show that in graph streams, for any $\epsilon > 0$, any one-pass $(2 - \epsilon)$ -approximation of TSP cost requires $\Omega(\epsilon^2 n^2)$ space. On the other hand, we show that there is an $\tilde{O}(n)$ space two-pass algorithm that approximates the TSP cost to within a factor of 1.96.

Finally, we consider the query complexity of estimating metric TSP cost to within a factor that is strictly better than 2 when the algorithm is given access to an $n \times n$ matrix that specifies pairwise distances between n points. The problem of MST cost estimation in this model is well-understood and a $(1 + \epsilon)$ -approximation is achievable by $\tilde{O}(n/\epsilon^{O(1)})$ queries. However, for estimating TSP cost, it is known that an analogous result requires $\Omega(n^2)$ queries even for (1, 2)-TSP, and for general metrics, no algorithm that achieves a better than 2-approximation with $o(n^2)$ queries is known. We make progress on this task by designing an algorithm that performs $\tilde{O}(n^{1.5})$ distance queries and achieves a strictly better than 2-approximation when either the metric is known to contain a spanning tree supported on weight-1 edges or the algorithm is given access to a minimum spanning tree of the graph.

Prior to our work, such results were only known for the special cases of graphic TSP and (1, 2)-TSP.

In terms of techniques, our algorithms for metric TSP cost estimation in both streaming and query settings rely on estimating the `\em cover advantage` which intuitively measures the cost needed to turn an MST into an Eulerian graph. One of our main algorithmic contributions is to show that this quantity can be meaningfully estimated by a sublinear number of queries in the query model. On one hand, the fact that a metric stream reveals pairwise distances for all pairs of vertices provably helps algorithmically. On the other hand, it also seems to render useless techniques for proving space lower bounds via reductions from well-known hard communication problems.

Our main technical contribution in lower bounds is to identify and characterize the communication complexity of new problems that can serve as canonical starting point for proving metric stream lower bounds.

Presenter: CHEN, Yu

Session Classification: Track A-4