Research Notes

Asaf Etgar

October 6, 2022

Chapter 1

Metric Repair

1.1 Graph Metric repair

Algorithms are faster when the distance between datapoints satisfy a metric. Since data is noisy, usually don't satisfy metric - we want to repair as few entries as possible. Let $G = (V, E, w)$ be a weighted graph where $w : V \to \mathbb{R}_{\geq 0}$ is a weight function. We think of w as a symmetric $n \times n$ matrix with 0 diagonal.

Definition 1.1.1 (Graph Metric Repair)**.** Let G be as described above, and let $\Omega \subset \mathbb{R}$. The GMR problem is

$$\underset{W \in Sym_n(\Omega)}{\operatorname{argmin}} \quad \|W\|_0 \text{ such that } w + W \text{ is a metric on } G.$$

The *support* of a solution W is the set of edges e with $W(e) \neq 0$. A solution is simply W that satisfies $w + W$ is a metric, and a solution is optimal if it also minimizes ℓ_0. When $\Omega = \mathbb{R}_{\geq 0}$ ($\mathbb{R}_{\leq 0}$) this is the *increase (decrease) only* setting.

A cycle C is said to be *broken* if there exists $h \in E(C)$ such that $w(h) > \sum_{e \in C \setminus h} w(e)$. In such case we call h a *heavy* edge, and other edges are *light*. An important observation is if the shortest path between two neighbors is not the edge connecting them, then this edge is a heavy edge in a cycle.

1.1.1 Results

In [FGR+19] it is shown that:

- The decrease only problem can be solved in cubic time.
- The support of a solution (to increase, and regular problem) must contain at least one edge from any broken cycle.
- $MULTICUT \leq_{APX,p} MR(G, \mathbb{R}_{\geq 0})$ - showing it is NP-hard ,and APX hard.
- $MR(G, \mathbb{R}_{\geq 0}) \leq_{APX,p} MR(G, \mathbb{R})$.
- $L - LB - CUT \leq MR(G, \mathbb{R}_{\geq 0}$.
- Fixed parameter analysis for ς-chordal graphs.
- Approximation algorithms for increase and regular cases.

A relaxation of this problem, in some sense, is to find the ℓ_0 distance to the metric cone:

Definition 1.1.2 (Metric Violation Distance)**.** Given $x \in \mathbb{R}_{\geq 0}^{\binom{[n]}{2}}$, find $\underset{y}{\operatorname{argmin}} \left\|x - y\right\|_0$.

In [CADK+22] they provide an $O(\log n)$ approximation algorithm that runs in cubic time for this problem, which is tight (assuming a standard fine-grained complexity conjecture)

1.2 Tree Fitting

Definition 1.2.1 (ℓ_p-fitting)**.** In the ℓ_p-fitting problem, we get as an input a set S with a metric $\mathcal{D} : \binom{S}{2} \longrightarrow \mathbb{R} > 0$ and look for an output tree metric T that spans S and fits $\mathcal{D}$ is the sense of minimizing $\|T - \mathcal{D}\|_p$ (when only considering points in S in the definition).

In [CADK+22] they give an $\Theta(1)$ approximation for the case of ℓ_1 (and claim that it does not generalize nicely to other norms). These problems are APX hard, so this is the best we could hope for. Their algorithm uses (unweighted minimizing disagreements) Correlation Clustering on K_n [BBC04]. Nevertheless, for a general p we can get $O(\log n \log\log n)$ approximation.

Appendix A

Extra Definitions

Definition A.0.1 (Ultrametric). a metric d is said to be an Ultrametric if the triangle inequality can be strengthened to

$$\forall x, y, z \quad d(x, y) \leq \max \left\{ d(x, z), d(z, y) \right\}$$

Definition A.0.2 (Tree Metric). Given a set S, a tree metric on S is a (positively) weighted tree T such that $S \subset V(T)$.

Definition A.0.3 (MULTICUT). Let G be a weighted graph, and $P = \{(s_i, t_i)\}_{i=1}^n$ pairs of vertices. The MULTICUT problem asks to find a minimal weight $M \subset E$ s.t $G \setminus M$ disconnects all pairs in P.

Remark. Known to be $\mathrm{NP\text{-}hard}$ to approximate if Unique Game Conjecture is true.

Definition A.0.4 (LB-CUT). Given L and a graph G with specified (s, t), find a minimum size subset $M \subset E$ such that $d_{G \setminus M}(s, t) \geq L$.

Remark. Fixing L, $\mathrm{NP\text{-}hard}$ to approximate within $\Omega(\sqrt{L})$.

Bibliography

[BBC04] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. Machine learning, 56(1):89–113, 2004.

[CADK+22] Vincent Cohen-Addad, Debarati Das, Evangelos Kipouridis, Nikos Parotsidis, and Mikkel Thorup. Fitting distances by tree metrics minimizing the total error within a constant factor. In 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), pages 468–479. IEEE, 2022.

[FGR+19] Chenglin Fan, Anna C Gilbert, Benjamin Raichel, Rishi Sonthalia, and Gregory Van Buskirk. Generalized metric repair on graphs. arXiv preprint arXiv:1908.08411, 2019.