Advanced Algorithms	09/26, 2017

Lecture 1

Lecturer: Mohsen Ghaffari — *Scribe: Michelle Sweering*

1 Graph Sparsification — Preserving Distances, i.e., Spanners

In this lecture we want to sparsify graphs while preserving distances. We call these sparser subgraphs *spanners*. The formal definition is as follows:

Definition 1 ((α, β)**-spanner**). *Given values $\alpha, \beta \geq 1$ and an arbitrary undirected unweighted graph with n-nodes $G = (V, E)$, we call a spanning subgraph $G' \subseteq G$ an (α, β)-spanner of G if for every pair s, t of vertices, we have*

$$dist_G(s,t) \leq dist_{G'}(s,t) \leq \alpha \cdot dist_G(s,t) + \beta.$$

In particular $(1, 0)$-spanners preserve the graph exactly. Higher values of α and β let us delete more edges, but we don't want them to be very high either. In this lecture we focus on the two extremes: multiplicative sparsification ($\beta = 0$) and additive sparsification ($\alpha = 1$).

2 Multiplicative Spanners

In this section, we focus on purely multiplicative spanners, i.e., (α, β)-spanners where $\beta = 0$. We refer to an $(\alpha, 0)$-spanner as an α-multiplicative spanner.

Theorem 2. *(Althofer et al. [ADD+93]) For every $k \geq 1$, every n-node graph G has a $(2k-1)$-multiplicative spanner $G' \subseteq G$ with $O(n^{1+1/k})$ edges.*

Proof. First we observe that it is okay to remove an edge, if there is a path of length $2k - 1$ or shorter between its endpoints. Instead of removing edges one by one, we remove all edges and then add them back in one by one, unless there is a path of length at most $2k-1$ between their endpoints. If such path exists, there is no need to add an edge. The subgraph G' that we obtain, is a $(2k - 1)$-multiplicative spanner. Now we only have to show it does not have too many edges.

Note that the resulting graph has girth at least $2k + 1$, i.e. G' does not have a cycle of length $2k$ or smaller. Suppose G' has more than $2n^{1+1/k}$ edges. We want to generate an induced subgraph $G'' \subset G'$ whose vertices all have degree at least $n^{1/k} + 1$. Remove repeatedly all vertices with degree less than $n^{1/k} + 1$. Each time we remove less than $n^{1/k} + 1$ edges. So in total we remove less than $n^{1+1/k} + n$ edges. However G' has $2n^{1+1/k}$ edges and some edges remain. Once we stop, all vertices have at least $n^{1/k} + 1$ neighbors. G'' has two properties:

1. The minimal degree of G'' is at least $n^{1/k}$.
2. G'' does not contain a cycle of length $2k$ or smaller.

We pick a vertex and count the number of vertices at distance at most k. Their number is at least

$$1 + (n^{1/k} + 1) + (n^{1/k} + 1)n^{1/k} + (n^{1/k} + 1)n^{2/k} + ... + (n^{1/k} + 1)n^{(k-1)/k} = 1 + \frac{(n^{1/k} + 1)(n - 1)}{n^{1/k} - 1} > n.$$

This is a contradiction, as there are only n vertices in G. Therefore G' has less than $2n^{1+1/k}$ edges. $\square$

This algorithm solves the problem in polynomial time. There is also an algorithm that solves this problem in near linear time, which we will see in the problem set of this lecture.

The sparsity obtained by the above algorithm for stretch $\alpha = 2k-1$ is conjectured to be optimal. The concrete conjecture is actually about the existence of some graphs, which cannot be sparsified, without stretching their edges by at least a $2k$ factor:

Conjecture 3. *(Girth Conjecture of Erdős [Erd64]) For every $k \geq 1$, there exists an n-node graph with $\Omega(n^{1+1/k})$ edges and girth at least $2k + 2$.*

Notice that since the graph has girth $2k + 2$, removing any single edge of it would stretch at least one edge by a factor of $2k$ (why?). The conjecture remains widely open and has been proven only for for small values of k, e.g., $k = 1, 2, 3, 5$.

3 Additive Spanners

In this section, we focus on purely additive spanners, i.e., (α, β)-spanners where $\alpha = 1$. We refer to 1 $(1, \beta)$-spanner as a β-additive spanner.

Theorem 4. *(Aingworth et al. [ACIM99]) Every n-node graph G has a 2-additive spanner $G' \subseteq G$ with $\tilde{O}(n^{3/2})$ edges.*

Proof. We partition the vertices of G into two types[1]: *light vertices*, which have degree at most $n^{1/2}$, and *heavy vertices*, which have degree more than $n^{1/2}$. Now, we remove all edges except those incident to a light vertex and call this new graph G'.

For every pair of vertices (s, t) consider a path $P_{s \to t}$ of minimal length in G. If all edges of $P_{s \to t}$ are in G', the distance between s and t remains the same. Suppose we found a set S such that all remaining paths $P_{s \to t}$ contain a point in S. Now adding the BFS trees for the vertices in S—i.e., one BFS tree rooted in each vertex of S—ensures that a shortest path exists from s to t exists.

Each heavy vertex has at least $n^{1/2}$ neighbors. If instead of on the paths themselves, we can choose the S elements in the neighorhoods of the paths, we can reduce the size of S, while increasing the distances by at most 2. For that, we do as follows.

We build S by including each $v \in V$ in S with probability $p = 10 \log n / \sqrt{n}$, independently. The expectation of the number of vertices sampled in the neighborhood of $P_{s \to t}$ is at least $\sqrt{n} \cdot p \geq 10 \log n$. Therefore, by Chernoff bound, with high probability there is at least one sample vertex in $N(P_{s \to t})$. Union bounding over all $\binom{n}{2}$ pairs (s, t) gives that with high probability, for each pair (s, t) that have a shortest path $P_{s \to t}$ with at least one heavy vertex, there is at least one sampled S-vertex in $N(P_{s \to t})$. Furthermore, a simple application of the Chernoff bound shows that, with high probability, $|S| = O(np) = O(\sqrt{n} \log n)$. Hence, the number of the edges that we add to the spanner because of adding the BFSs rooted in S is at most $O(n\sqrt{n} \log n)$. □

Theorem 5. *(Chechik [Che13]) Every n-node graph G has a 4-additive spanner $G' \subseteq G$ with $\tilde{O}(n^{7/5})$ edges.*

Proof. The proof is somewhat similar to that of theorem 4. Again we keep all edges of light vertices, but this time light means degree at most $n^{2/5}$. For the shortest paths with heavy vertices we distinguish two cases:

1. Paths $P_{s \to t}$ with at least $n^{1/5}$ heavy vertices. For this case, we first argue that for any such path, we must have $|N(P_{s \to t})| > n^{3/5}/3$. First, notice that the summation of the neighborhood sizes of the heavy vertices on $P_{s \to t}$ is at least $n^{1/5} \cdot n^{2/5}$, simply because we have $n^{1/5}$ heavy vertices and each contributes at least $n^{2/5}$. However, we might have double-counting, in the sense of counting each node in the neighborhood of the path many times. We next argue that each node can be counted at most 3 times. For the sake of contradiction, suppose that there is a vertex v that is adjacent to four vertices v_1, v_2, v_3, v_4 on the path (in this order).

$$\begin{aligned} dist(s,t) &\leq dist(s,v_1) + dist(v_1,v) + dist(v,v_4) + dist(v_4,t) \\ &= dist(s,v_1) + 2 + dist(v_4,t) \\ &< dist(s,v_1) + dist(v_1,v_2) + dist(v_2,v_3) + dist(v_3,v_4) + dist(v_4,t) \\ &= dist(s,t) \end{aligned}$$

   This is a contradiction. Therefore $|N(P_{s \to t})| > n^{3/5}/3$. We now sample vertices in V with probability $p = n^{-3/5} \log n$. With high probability the sample will have a vertex in the neighborhood of every shortest path with a heavy vertex as above. Add the BFS trees of the vertices in the sample. Distances are increased by at most two, as we say in the proof of the previous theorem. Moreover, with high probability, we have at most $O(n \cdot n^{-3/5} \log n)$ sampled vertices. Since we add a full BFS, which has at most $n - 1$ edges, for each sampled vertex, this is at most $O(n^{7/5} \log n)$ edges, in total, that are added to the spanner.

2. Paths $P_{s \to t}$ with less than $n^{1/5}$ heavy vertices
   Pick $O(n^{3/5} \log n)$ heavy vertices at random. We call them the *heavy centers*. With high probability each heavy vertex has a heavy center neighbor. For each pair of heavy vertices (c_1, c_2) we add a

[1]We note that in this scheme, one can slightly optimize the bounds and achieve a spanner with $n\sqrt{n \log n}$ edges, by picking $\sqrt{n \log n}$ as the threshold for heavy vertices. However, for simplicity, we do not focus on these logarithmic factors.

shortest path between c_1 and c_2 among all paths that have at most $n^{1/5}$ internal (i.e., ignoring endpoints) heavy vertices, if such a path exists. Notice that we are not necessarily taking the shortest path among the two centers, as that might have many more heavy vertices.

Next, we argue that for any such path $P_{s\to t}$ that has less than $n^{1/5}$ heavy vertices, the constructed spanner includes a path with length at most $dist(s,t)+4$. Let h_1 be the first heavy node on $P_{s\to t}$ and h_ℓ be the last heavy vertex on it. Let c_1 and c_ℓ be the corresponding heavy centers, which exists with high probability, as argued above. Then, the spanner includes a path between c_1 and c_ℓ that has length at most $dist(h_1,h_\ell)+2$. The reason is that, the path $(c_1,h_1), h_1 \to h_\ell, (h_\ell, c_\ell)$ is one path connecting c_1 to c_ℓ that has at most $n^{1/5}$ internal heavy vertices and has length $dist(h_1,h_\ell)+2$, and we add the shortest such path connecting c_1 to c_ℓ to the spanner. Then, there is a path in the spanner connecting s to t by taking the segment of $P_{s\to t}$ from s to h_1, then the edge (h_1,c_1), then the afformentined path from (c_1) to (c_ℓ), the edge (c_ℓ,h_ℓ), and finally the segment of $P_{s\to t}$ from h_ℓ to t. This path has length at most $dist(s,t)+4$.

□

If we relax the desired additive stretch, we can achieve even a sparser spanner graph, as captured by the following theorem, which we do not prove in the class.

Theorem 6. *(Baswana et al. [BKMP05]) Every n-node graph G has a 6-additive spanner $G' \subseteq G$ with $\tilde{O}(n^{4/3})$ edges.*

Given the trend shown in the previous three theorems, one would expect that further increasing the additive stretch of spanners to larger constants should allow for even less edges. This is however not the case as stated in the following theorem.

Theorem 7. *(Abboud and Bodwin. [AB16]) For any constant $\varepsilon > 0$, there exists an n-node graph that has no 6-additive spanner with $O(n^{4/3}-\varepsilon)$ edges. In fact, for any small fixed $\varepsilon > 0$, there is a $\delta > 0$ and an n-graph that has no n^δ-additive spanners with at most $n^{4/3-\varepsilon}$ edges.*

That is, perhaps surprisingly, $O(n^{4/3})$ is the best possible sparsity for any constant additive spanner. It remains an open problem whether a 4-additive spanner with $O(n^{4/3})$ exists:

Open Problem 8. *Prove (or disprove) that every n-node graph G has a 4-additive spanner $G' \subseteq G$ with $\tilde{O}(n^{4/3})$ edges.*

References

[AB16] Amir Abboud and Greg Bodwin. The 4/3 additive spanner exponent is tight. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 351–361. ACM, 2016.

[ACIM99] Donald Aingworth, Chandra Chekuri, Piotr Indyk, and Rajeev Motwani. Fast estimation of diameter and shortest paths (without matrix multiplication). *SIAM Journal on Computing*, 28(4):1167–1181, 1999.

[ADD+93] Ingo Althöfer, Gautam Das, David Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9(1):81–100, 1993.

[BKMP05] Surender Baswana, Telikepalli Kavitha, Kurt Mehlhorn, and Seth Pettie. New constructions of (α, β)-spanners and purely additive spanners. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 672–681. Society for Industrial and Applied Mathematics, 2005.

[Che13] Shiri Chechik. New additive spanners. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 498–512. SIAM, 2013.

[Erd64] Paul Erdős. Extremal problems in graph theory. In *IN THEORY OF GRAPHS AND ITS APPLICATIONS, PROC. SYMPOS. SMOLENICE.* Citeseer, 1964.