Additive Spanners: A Simple Construction [5]

Seminar Advanced Algorithms and Data Structures

Basil Fürer

November 2, 2018

1 Motivation

As we all know simple graphs can have up to $\mathcal{O}(n^2)$ edges, this means that algorithms which depend on the number of edges (eg. breadth-first search) can be costly. To overcome this problem one could try to approximate a graph with a sparser representation, however this means that we will lose information. If one is trading loss of information a sparser representation, one is likely interested to do so in a way which makes some guarantees.

One very natural thing would be, to argue about how distances between two vertices behave when comparing the sparse approximation to the original graph. Requiring that any distance will only be incremented by a constant amount leads us to the formalisation of additive spanners – such a spanner is a subgraph in which any distance differs by at most a constant additive term k compared to the original graph – also called k-spanner.

Imagine a large graph describing certain routes, a route in the approximating graph which is only incremented by some constant amount is more likely to be interesting than one where the amount of time gets multiplied by a constant factor. While our intuition tells us that such an approximation is less likely to exist, it has been shown $[2, 3, 5, 6]$ that such k-spanners exist for all graphs and that these representations are interesting since there provable upper bounds on the number of edges. In fact, compressing graphs by using k-spanners finds application in various settings, such as graph compression itself, synchronisation in distributed systems, routing schemes or distance oracles[1] $[2, 6]$.

2 k-Spanners

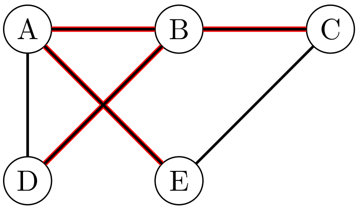
In the following we will talk only about unweighted, undirected graphs G and H shall always be a subgraph thereof. For a list of the notations used please refer to the last section *Formulas and Notations*.

Definition: A k-spanner of G is a subgraph H, such that for all $u, v \in V(G)$ it holds that $d_H(u, v) \leq d_G(u, v) + k$.

[1]For this or similar applications where one is only interested in the distance itself, one might want to drop the requirement of the approximation being a subgraph of the original graph. These kinds of approximations are called emulators.

Figure 1 shows a graph G with black edges, the subgraph consisting of the red edges is a 2-spanner of G.

The table (note that the edges are undirected) shows all the shortest paths in the original (black) graph and the differences (red) of the new distances. As we can see 2 is the largest difference for the path $C \to E$ which makes this a 2-spanner.



	A	B	C	D	E
A	×	0	0	**1**	0
B	1	×	0	0	0
C	2	1	×	0	**2**
D	1	1	2	×	**1**
E	1	2	1	2	×

Figure 1: Example of a 2-spanner

3 k-spanner-completion

The main contribution of the paper [5] is the following very elegant idea for an algorithm, referred to as *k-spanner completion*:

Start off with some subgraph H of G (we will discuss a good choice thereof later). And as long as there is a pair of vertices - say u and v - for which the distance is larger than $d_G(u, v) + k$, find a shortest path from between these vertices and add it to H. Formally:

while $\exists\, u, v \in V(G)\ .\ d_H(u, v) > d_G(u, v) + k$ **do**
  $(e_1, \ldots, e_t) \leftarrow$ `shortestPath`(u, v)
  add edges $e_1, \ldots, e_t$ to H
end

It's easy to see that this process will terminate (at each iteration at least one edge gets added) and when it does, it holds that $d_H(u, v) \leq d_G(u, v) + k$ for all pairs of vertices $u, v \in V(G)$. Hence an algorithm implementing this very idea will upon termination have produced a k-spanner H.

As already mentioned the choice of the initial subgraph H is left open and it will have a crucial influence on the resulting k-spanner. Since the goal of approximating graphs with spanners is to obtain a sparser representation of a graph we're interested in upper bounds of the number of edges $E(H)$. The paper gives two proofs:

Theorem 1: By setting $H = \big(V(G), \tilde{E}\big)$ where $\tilde{E}$ is a set of edges such that for each node there are $\lfloor n^{1/3} \rfloor$ edges of the original graph (if the degree is less than that, add all) and running a 6-spanner completion, we will get a 6-spanner with at most $O(n^{4/3})$ edges.

Theorem 2: By setting $H = \big(V(G), \varnothing\big)$ and running a 2-spanner completion on H we will obtain a 2-spanner with at most $\mathcal{O}(n^{3/2})$ edges.

4 Proof of Theorem 2

Let us prove[2] Theorem 2, for this we define two functions on the set of subgraphs of G:

[2] This proof slightly differs from the one given in the original paper [5] in that it is more verbose and fixes a mistake.

Namely v – the value function – which measures how good of an approximation H is, it is a sum over all shortest paths where each addend contributes a value from $[0, 3]$ (3 if the distance in H is the same as in G, 2 if they differ by 1, 1 if they differ by 2 and zero in any other case). Intuitively speaking, any distance between pairs that contribute 0 has yet to be improved by adding edges to H.

The other function c – the cost function – measures the number of edges in G by summing over all degrees squared:

$$v(H) = \sum_{u,v \in V(G)} \max\{0, d_G(u,v) - d_H(u,v) + 3\} \quad \text{and} \quad c(H) = \sum_{v \in V(G)} \deg_H(v)^2$$

We have $0 \le v(H)$ and since $d_G(u,v) \le d_H(u,v)$ it holds that $v(H) \le \sum\limits_{u,v \in V(G)} 3 = 3 \cdot \frac{n \cdot (n-1)}{2}$.

Furthermore

$$4 \cdot |E(H)|^2 = \big(2 \cdot |E(H)|\big)^2 \overset{HS}{=} \Big(\sum_{v \in V(G)} 1 \cdot \deg_H(v)\Big)^2 \overset{CS}{\le} \sum_{v \in V(G)} 1^2 \cdot \sum_{v \in V(G)} \deg_H(v)^2 = n \cdot c(H)$$

which implies $2 \cdot |E(H)| \le \sqrt{n \cdot c(H)}$.

We know $v(H) = \mathcal{O}(n^2)$, so if we prove that $\Delta = c(H) - 12 \cdot v(H)$ will never increase in any step of the 2-spanner completion, we know that $c(H) = \mathcal{O}(n^2)$ (because $v(H)$ initially is 0) and thus $|E(H)| = \mathcal{O}(n^{3/2})$.

Proof that Δ never increases: For this we consider an arbitrary step in the 2-spanner completion, let graph H_0 be the subgraph before adding edges where we found a shortest path from u to v of length t and $d_{H_0}(u,v) > d_G(u,v) + 2$ holds. H shall denote the subgraph after adding the edges of the shortest path from u to v. Let the vertices of the path be $P := \{w_i\}_{i=0}^t$ with the convention of writing w_0 as u and w_t as v.

Per node on the path we add maximally 2 new edges, so for all $i = 0 \dots t$:

$$\deg_H(w_i) - 2 \le \deg_{H_0}(w_i) \implies \big(\deg_H(w_i) - 2\big)^2 \le \deg_{H_0}(w_i)^2$$

$$\implies \sum_{i=0}^{t} \big(\deg_H(w_i) - 2\big)^2 \le c(H_0)$$

$$\implies c(H) - c(H_0) = \sum_{i=0}^{t} \big(\deg_H(w_i) - \deg_{H_0}(w_i)\big) + \sum_{v \in V(G) \setminus P} \underbrace{\big(\deg_H(v)^2 - \deg_{H_0}(v)^2\big)}_{=0}$$

$$\le \sum_{i=0}^{t} \underbrace{\big(\deg_H(w_i)^2 - (\deg_H(w_i) - 2)^2\big)}_{= 4 \cdot (\deg_H(w_i) - 1)} \le 4 \cdot \sum_{i=0}^{t} \deg_H(w_i)$$

Let us define the set of vertices on the path or adjoint to it as N, ie. $N := \bigcup\limits_{i=0}^{t} \Gamma_H(w_i)$.

Now for any such vertex $z \in N$ we have $d_H(u,z) + d_H(z,v) \le d_G(u,v) + 2$, on the other hand we also know that $d_G(u,v) + 2 < d_{H_0}(u,z) + d_{H_0}(z,v)$ since there would be a shorter path in H_0 otherwise, so we get

$$d_H(u,z) + d_H(z,v) < d_{H_0}(u,z) + d_{H_0}(z,v).$$

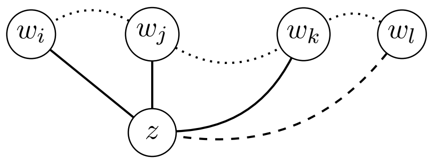


Figure 2: If the dashed edge gets added, the dotted path would not be the shortest.

For the sake of a contradiction, let us assume that $d_{H_0}(u,z) \leq d_H(u,z)$ and $d_{H_0}(z,v) \leq d_H(z,v)$ hold which leads us to $d_{H_0}(u,z)+d_{H_0}(z,v) \leq d_H(u,z)+d_H(z,v)$. We just showed the opposite is true and therefore at least one of the following must hold

$$d_{H_0}(u,z) > d_H(u,z) \quad \text{or} \quad d_{H_0}(z,v) > d_H(z,v).$$

From this we can deduce that for any $z \in N$

$$\begin{gathered}\sum_{w\in V(G)} \max\{0, d_G(u,z) - d_H(u,z) + 3\} > \sum_{w\in V(G)} \max\{0, d_G(u,z) - d_{H_0}(u,z) + 3\} \\ \text{or} \\ \sum_{w\in V(G)} \max\{0, d_G(z,v) - d_H(z,v) + 3\} > \sum_{w\in V(G)} \max\{0, d_G(z,v) - d_{H_0}(z,v) + 3\}\end{gathered} \tag{1}$$

will hold because $d_H(x,y) \leq d_{H_0}(x,y)$ is always true for any $x, y \in V(G)$ (this is a trivial fact of H_0 being a subgraph of H). It should be noted that the left terms will differ by at least 1 due to the addends being natural numbers.

Now we deploy a simple, yet powerful counting argument: If we sum over the degrees of all vertices on the path P in H, we will end up counting every vertex adjacent to or on the path (ie. the vertices of N) at most 3 times since otherwise P would not be the shortest path (see Figure 2), hence:

$$\sum_{i=0}^{t} \deg_H(w_i) \leq 3 \cdot |N|$$

One of the inequalities (1) holds for all nodes z in N for which there are at least $\frac{1}{3}\sum_{i=0}^{t} \deg_H(w_i)$ as just established, therefore we have for $v(H) - v(H_0)$:

$$\begin{aligned} &\underbrace{\sum_{z\in N} \sum_{w\in\{u,v\}} \big(\max\{0, d_G(z,w) - d_H(z,w) + 3\} - \max\{0, d_G(z,w) - d_{H_0}(z,w) + 3\}\big)}_{\geq 1} \\ &\underbrace{+ \sum_{\substack{z \in V(G)\setminus N \\ w \in V(G)\setminus\{u,v\}}} \big(\max\{0, d_G(z,w) - d_H(z,w) + 3\} - \max\{0, d_G(z,w) - d_{H_0}(z,w) + 3\}\big)}_{\geq 0} \\ &\geq \sum_{z\in N} 1 = |N| \geq \frac{1}{3}\sum_{i=0}^{t} \deg_H(w_i) \end{aligned}$$

Overall, we derived:

$$\frac{1}{3}\sum_{i=0}^{t}\deg_H(w_i) \le v(H) - v(H_0) \implies \underbrace{c(H) - c(H_0)}_{\le 4\cdot\sum_{i=0}^{t}\deg_H(w_i)} - \underbrace{12\cdot(v(H) - v(H_0))}_{\ge 12\cdot\frac{1}{3}\sum_{i=0}^{t}\deg_H(w_i)} \le 0$$

So $\Delta = c(H) - 12 \cdot v(H)$ does never increase when we adding paths according to the 2-spanner completion which concludes the proof. □

5 Conclusion

At the time the paper was published it has been shown [6] that for all k there are graphs with n nodes such that the $(2k-1)$-spanner has $\Omega(n^{1+1/k})$ edges, therefore the 2-spanner completion is optimal.

A year after this paper was released Amir Aboud and Greg Bodwin proved [1] that for all $\epsilon > 0$ there is a $\Delta > 0$ and an infinite family of graphs, such that for any subgraph H with $|E(H)| = \mathcal{O}(n^{4/3-\epsilon})$ there are at least two vertices u, v with $d_H(u,v) = d_G(u,v) + \Omega(n^{\Delta})$. This means that there can't be an additive spanner[3] with less than $\mathcal{O}(n^{4/3})$ edges, therefore the 6-spanner completion is optimal in the sense that there is no sparser representation.

This begs the question what other optimal algorithms for additive (+2 and +6) spanners exist and how do they compare:

k	time	Algorithm
2	$\mathcal{O}(n^2)$	D. Dor et al. [3]
	$\mathcal{O}(m \cdot \sqrt{n})$	M. Elkin and D. Peleg [4]
	$\mathcal{O}(m \cdot n)$	M.B. Tejs Knudsen (this paper) [5]
6	$\mathcal{O}(m \cdot n)$	S. Baswana et al. [2]
	$\mathcal{O}(m \cdot n)$	M.B. Tejs Knudsen (this paper) [5]

Even though this algorithm is not the fastest, given that it has been proved optimal and that it's a really concise and elegant algorithm, it is well worth knowing it.

[3] In fact, the statement is stronger and even holds for emulators or multiplicative spanners too.

References

[1] Amir Abboud and Greg Bodwin. The 4/3 additive spanner exponent is tight. *J. ACM*, 64(4):28:1–28:20, 2017.

[2] Surender Baswana, Telikepalli Kavitha, Kurt Mehlhorn, and Seth Pettie. New constructions of (alpha, beta)-spanners and purely additive spanners. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 672–681. SIAM, 2005.

[3] Dorit Dor, Shay Halperin, and Uri Zwick. All-pairs almost shortest paths. *SIAM J. Comput.*, 29(5):1740–1759, 2000.

[4] Michael Elkin and David Peleg. (1+epsilon, beta)-spanner constructions for general graphs. *SIAM J. Comput.*, 33(3):608–631, 2004.

[5] Mathias Bæk Tejs Knudsen. Additive spanners: A simple construction. *CoRR*, abs/1403.0178, 2014.

[6] David P. Woodruff. Lower bounds for additive spanners, emulators, and more. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 389–398. IEEE Computer Society, 2006.

Formulas and Notations

Notation	**Meaning**
$d_G(u,v)$	Distance from u to v (or vice versa) in G
$\deg_G(u)$	Degree of vertex u in G
$\Gamma_G(u)$	Neighbours of vertex u in G
$V(G)$	Set of vertices in G
$E(G)$	Set of edges in G

CS: Cauchy-Schwarz inequality: $\forall x, y \in \mathbb{R}^n \, . \, (\sum_{i=1}^{n} x_i \cdot y_i)^2 \leq \sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2$

HS: Hand Shaker's Lemma: $\sum_{v \in V(G)} \deg_G(v) = 2 \cdot |E(G)|$