CS 466/666: Algorithm Design and Analysis **University of Waterloo: Fall 2023**

Lecture 15

November 1, 2023

Instructor: Sepehr Assadi

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Topics of this Lecture

1 Graph Spanners 1

1.1 Density of Graphs without Short Cycles 3

2 Asymptotic Optimality of Proposition 4 for $k = 2$ 4

2.1 A "Weak" Construction Using Random Graphs 4

2.2 An Asymptotically Optimal Construction 5

1 Graph Spanners

A general family of questions in theoretical computer science (and beyond) are "graph compression" problems: how to compress a graph—in particular, *sparsify* its edges—while still preserving some of its "useful properties", perhaps approximately only. We already saw an example of this question in Homework 2 when designing *cut sparsifiers*. In this lecture, we consider another notion of graph compression, this time for preserving *distances*.

Let $G = (V, E)$ be any undirected graph. For any pairs of vertices $u, v \in V$, we use $\text{dist}_G(u, v)$ to denote the **distance** between u and v in G, i.e., the number of edges in the shortest path between u and v.

Definition 1. For any undirected graph $G = (V, E)$, and any $\alpha \geqslant 1$, an α-**spanner** of G is any spanning subgraph H of G such that for all pairs of vertices $u, v \in V$:

$$\text{dist}_G(u, v) \leqslant \text{dist}_H(u, v) \leqslant \alpha \cdot \text{dist}_G(u, v).$$

(Note that the left inequality is redundant here given that H is a subgraph of G.)

It is easy to see that every graph admits an α-spanner for any $\alpha \geqslant 1$: simply let H be the same as G. Moreover, the only 1-spanner of any graph G is itself; even if do not include a single edge (u, v) from G, then the distance between u and v will become more than 1 in the spanner, which violates the 1-spanner property. The trick however is to allow for some *approximation* of $\alpha > 1$ and obtain a *sparse* spanner, i.e., a one that has substantially less edges than G itself.

We can look at spanners from different points of view. An *algorithmic* approach for instance would be to find the *sparsest* α-spanner of a given graph G and parameter α. However, in this lecture (and generally in graph compression questions), we are interested in an *extremal* question: can we prove that every graph

G admits an α-spanner with $O(f(n))$ edges for some function $f(n)$? We will answer this question in this lecture by proving the following theorem.

Theorem 2. *For every integer $k \geqslant 1$, every graph G admits a $(2k-1)$-spanner with $O(n^{1+1/k})$ edges.*

For instance, Theorem 2 implies that there is already a 3-spanner for every graph with $O(n^{3/2})$ edges, much smaller than $O(n^2)$ edges that is the total number of possible edges in G. Similarly, by setting $k = \log n$, we obtain that every graph G admits an $O(\log n)$-spanner with $O(n)$ edges.

It is worth asking why we are only interested in *odd* values of approximation factor α in α-spanners, i.e., $(2k-1)$-spanners in Theorem 2. The answer to this question hopefully will become clear from the proof of Theorem 2 but you are encouraged to think about it on your own also (in particular, what is a difference between a $2k$-spanner and a $(2k-1)$-spanner in a bipartite graph?).

We prove Theorem 2 by presenting a simple algorithm, called the *greedy spanner*, for computing the spanner of the input graph. We will then bound the total number of edges in this spanner.

Algorithm 1. The **greedy spanner** algorithm for computing a $(2k-1)$-spanner.

1. Let $H \leftarrow \emptyset$ and $e_1, \ldots, e_m$ be any fixed ordering of the edges of G.
2. For $i = 1$ to m: if $H \cup \{e_i\}$ does not have a cycle of length $\leqslant 2k$, add e_i to H.
3. Return H as the spanner of G.

Let us first prove that H output by Algorithm 1 is a spanner of G.

Lemma 3. *The subgraph H output by Algorithm 1 is a $(2k-1)$-spanner of the input graph G.*

Proof. We say that an edge $(u,v) \in G$ is *stretched* by α iff $\text{dist}_H(u,v) = \alpha$, i.e., in place of the edge (u,v) in H, we now have a path of length α. Any edge $(u,v) \in G \setminus H$ is stretched by at most $(2k-1)$ because the reason we did not add (u,v) to H was existence of a cycle of length $\leqslant 2k$ in H (that inevitably includes both u and v); but this means that there is a path of length $\leqslant 2k-1$ between u and v in H. This proves the spanner property for every pairs of vertices u,v that are an edge in G. We now extend this to all pairs. We do this using two (slightly differently written) proofs. See Figure 1 for an illustration.

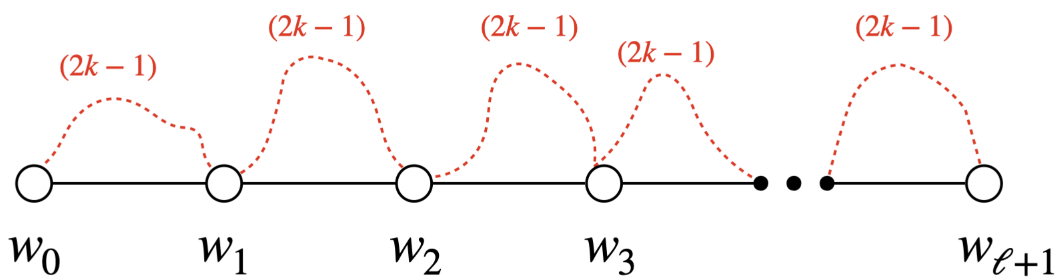

Figure 1: A simple illustration of the proof of Lemma 3. As we only skip edges of G in Algorithm 1 that are part of a $\leqslant 2k$ cycle, every edge of a shortest path between two vertices x,y in G is stretched by at most $(2k-1)$ in H; this means that the length of x-y shortest path in H is at most $(2k-1)$ times larger than G.

Proof 1. Consider a pair of vertices $x,y \in V$ and the shortest path $x = w_0 \rightarrow w_1 \rightarrow w_2 \ldots w_\ell \rightarrow y = w_{\ell+1}$ in G with length ℓ. Every edge in this path is stretched by at most $(2k-1)$ in H and thus there is a *walk* of length $(2k-1)\cdot\ell$ in H between x,y. This immediately implies also that $\text{dist}_H(x,y) \leqslant (2k-1)\cdot\text{dist}_G(x,y)$ as well as we can shortcut the edges of the walk into a path. Hence, H is a $(2k-1)$-spanner.

Proof 2. Again consider a pair of vertices $x, y \in V$ and the shortest path $x = w_0 \to w_1 \to w_2 \dots w_\ell \to y = w_{\ell+1}$ in G with length ℓ. We have,

$$\begin{aligned}
\mathrm{dist}_H(x,y) &\leqslant \sum_{i=1}^{\ell+1} \mathrm{dist}_H(w_{i-1},w_i) && \text{(by triangle inequality of distances)} \\
&\leqslant \sum_{i=1}^{\ell+1} (2k-1) \\
& \text{(as } (w_{i-1},w_i) \text{ is an edge in } G \text{ and we have established the spanner property on edges)} \\
&= (2k-1) \cdot \mathrm{dist}_G(x,y).
\end{aligned}$$

This concludes the proof. $\square$

Lemma 3 is kind of obvious by the design of Algorithm 1; we literally only skipped adding an edge to the spanner if we already provided a good approximation for it in the spanner. The main part of the analysis of the greedy spanner is then to bound its number of edges. Notice that a key property of the spanner H output by the algorithm is that it does not have any cycle of length $\leqslant 2k$ (the "last" edge of the cycle will never be inserted to the spanner). Now, we know that a graph without *any* cycle has to be sparse, i.e., can only have $n-1$ edges. Can we generalize this to graphs that do not have "short" cycles?

1.1 Density of Graphs without Short Cycles

To do this, we rely on a classical result in graph theory referred to as the **Moore Bound**.

Proposition 4 (**Moore Bound**). *For any integer $k \geqslant 1$, any graph $G = (V, E)$ with no cycle of length less than or equal to $2k$ can have at most $n^{1+1/k} + n$ edges.*

Proof. We first claim that G has an induced subgraph $\tilde{G}$ with minimum degree $\delta(\tilde{G}) > m/n$ (more than half the average degree). This can be proven as follows: if $\delta(G) > m/n$ we are done already by setting $\tilde{G} = G$; if not, pick any vertex $v \in G$ with $\deg(v) \leqslant m/n$ and remove it from the graph. As removing a vertex of degree $\leqslant m/n$ reduces number of edges by $\leqslant m/n$, this will reduce the average degree of the remaining graph to:

$$\geqslant \frac{2 \cdot (m - m/n)}{n-1} = \frac{2m \cdot (1-1/n)}{n \cdot (1-1/n)} = 2m/n.$$

As such, this process never decreases the average degree and hence needs to terminate in a non-empty subgraph $\tilde{G}$ with $\delta(\tilde{G}) > m/n$.

We now have a subgraph $\tilde{G}$ with minimum degree $d > m/n$ over at most n vertices. Consider growing a BFS tree of depth k from some arbitrary vertex v of $\tilde{G}$. The vertices of this tree cannot be visited in more than one path from v as otherwise any vertex x reachable from v via more than two paths of length $\leqslant k$, creates a cycle of length $\leqslant 2k$ already, contradicting $\tilde{G} \subseteq G$ not having any $\leqslant 2k$-cycle (see Figure 2).

But this means that we should visit

$$1 + d + d \cdot (d-1) + d \cdot (d-1)^2 + \dots + d \cdot (d-1)^{k-1} \geqslant (d-1)^k$$

distinct vertices in this BFS tree. As the number of vertices of $\tilde{G}$ is at most n, we have,

$$(d-1)^k \leqslant n \implies d \leqslant n^{1/k} + 1 \implies \frac{m}{n} \leqslant n^{1/k} + 1 \implies m \leqslant n^{1+1/k} + n,$$

concluding the proof. $\square$

Theorem 2 now follows immediately from Lemma 3 and Proposition 4.

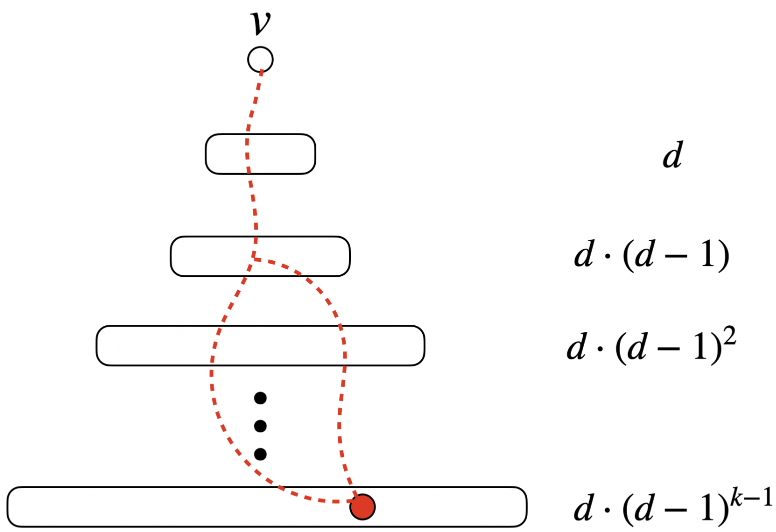

Figure 2: A simple illustration of the proof of Proposition 4. The vertices in the first k layer of the BFS tree from v are distinct, i.e., only have one path from v—otherwise, we will find a cycle of length $\leqslant 2k$ already (as shown by the red dashed edges).

Notice that the bound we obtained on the size of the spanner was a direct corollary of Proposition 4; thus, if we could improve the bounds there with a better analysis, we would immediately obtain a better bound on the size of the spanner as well. It is thus worth asking if we could improve Proposition 4? The answer turns out to be *No* in the following sense: for some small values of k, we already *know* this is impossible (to show case this, we will prove this for $k = 2$ in the next section); for larger values of k, improving this bound or proving this is not possible is a longstanding open question in graph theory:

Remark. The **Erdős-Girth** conjecture states that:

for every integer $k \geqslant 1$, there are graphs with $\Omega(n^{1+1/k})$ edges with no cycle of length $\leqslant 2k$;

in other words, the Moore bound is asymptotically tight. This is a very important conjecture when studying spanners (among others) but at this point, it is only known to be true for several small choices of k. The general conjecture is still open and the best known construction imply a lower bound of $\Omega(n^{1+1/2k-1})$ edges (a random graph of average degree $\Theta(n^{1/2k-1})$ can be used to obtain this).

2 Asymptotic Optimality of Proposition 4 for $k = 2$

We are going to prove that there are graphs with no cycle of length $\leqslant 4$ which still have $\Omega(n^{3/2})$ edges. We first see a more "standard" approach using random graph theory (that we briefly saw in Lecture 5 and homeworks) that does *not* give the optimal bounds, and then see how to use a more clever (and "magical") construction that does the trick.

2.1 A "Weak" Construction Using Random Graphs

Create a *bipartite* graph $G = (L, R, E)$ with $|L| = |R| = n$ by picking each edge independently with probability $p \in (0, 1)$ for some p to be determined later. Firstly, the expected number of edges in such a graph is:

$$\mathbb{E}\,|E| = p \cdot n^2.$$

Since G is bipartite, it cannot have a 3-cycle for us to worry about. What is a probability that this graph has a 4-cycle? We can calculate this probability by iterating over all possible choices of picking two vertices from both of L and R, and checking if all the edges between them appear in G. Thus,

$$\begin{aligned}
\Pr(G \text{ has a 4-cycle}) &\leqslant \sum_{u,v\in L, x,y\in R} \Pr(u,v,x,y \text{ form a cycle}) && \text{(by union bound)} \\
&= \sum_{u,v\in L, x,y\in R} p^4 && \text{(as all the 4 edges should appear independently)} \\
&= \binom{n}{2}^2 \cdot p^4 && \text{(by the number of choices of two vertices from either } L \text{ or } R\text{)} \\
&\leqslant n^4 \cdot p^4.
\end{aligned}$$

For this probability to be less than 1, we need to pick $p < 1/n$. But for $p < 1/n$, that means that the expected number of edges in G is only going to be n, which is very small to be of any use.

The trick however is to do a *two step* analysis. Let C denote the number of 4-cycles in G. By the same exact calculation as above, we have that

$$\mathbb{E}[C] \leqslant p^4 \cdot n^4.$$

Now consider the random variable Y which is equal to the number of edges of G that do not appear in any 4-cycle. As each cycle counted in C contains 4 edges, we have $Y \geqslant |E| - 4 \cdot C$. Thus, by linearity of expectation,

$$\mathbb{E}[Y] \geqslant \mathbb{E}|E| - 4 \cdot \mathbb{E}[C] \geqslant p \cdot n^2 - 4 \cdot p^4 \cdot n^4.$$

Now, suppose we pick p such that

$$4 \cdot p^4 \cdot n^4 = \frac{1}{2} \cdot p \cdot n^2,$$

i.e., set

$$p := \frac{1}{2 \cdot n^{2/3}}.$$

Then, we have

$$\mathbb{E}[Y] \geqslant \frac{1}{2} \cdot p \cdot n^2 = \frac{1}{4} \cdot n^{4/3}.$$

This implies that there should exist a bipartite graph with $\Omega(n^{4/3})$ edges and no 4-cycles (and no 3-cycles).

Remark. As an aside, the type of proof we did is called a **probabilistic method**: we used a probabilistic process to create a graph and then showed that in expectation it has the "right" number of edges for us. This implies that there should exists at least one graph with the right number of edges also otherwise the expected value of the above process could only become lower.

2.2 An Asymptotically Optimal Construction

We now present a bipartite graph with no 4-cycles and $\Omega(n^{3/2})$ edges. Let p be a prime number and define $\mathbb{F}_p$ as the field of integers modulo p, i.e., $\{0, 1, \ldots, p-1\}$, where addition, subtraction, and multiplication is done mod p.

Consider the following bipartite graph $G = (L, R, E)$ with the following vertices:

- Let $L := \mathbb{F}_p \times \mathbb{F}_p$ interpreted as *points* (x, y) in the two-dimensional plane ${\mathbb{F}_p}^2$.
- Let $R := \mathbb{F}_p \times \mathbb{F}_p$ interpreted as *lines* (a, b) in the two-dimensional plane ${\mathbb{F}_p}^2$.

- Let E be the set of edges that capture *incidence* in L and R meaning there is an edge between any $(x, y) \in L$ and $(a, b) \in R$, if the point (x, y) is on the line (a, b); formally, if

$$a \cdot x + b = y \pmod{p}.$$

See Figure 3 for an illustration.

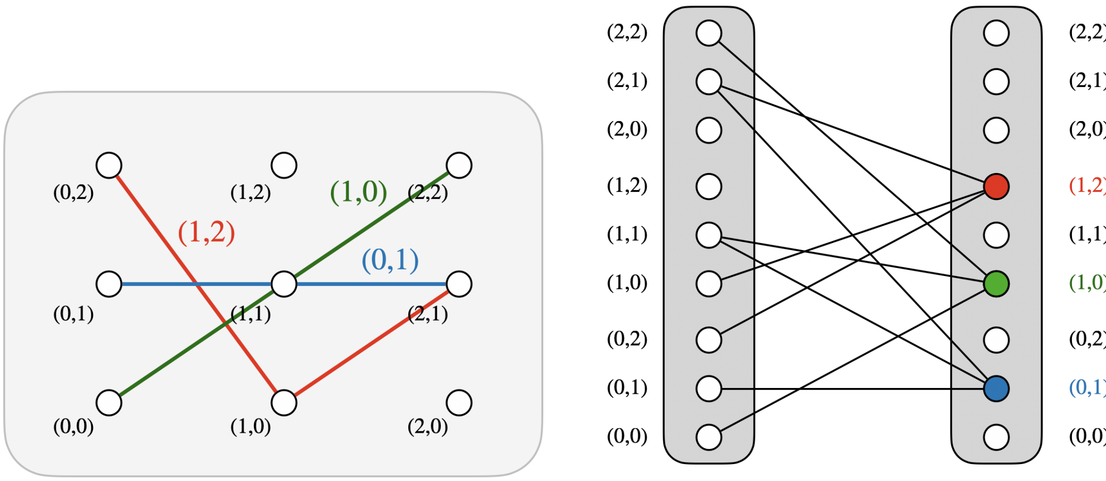

Figure 3: The plane used in the definition of the graph for $p = 3$ on the left, and the graph itself on the right. The points in $\{0, 1, 2\} \times \{0, 1, 2\}$ are all drawn – these correspond to vertices in L. Only three of the lines are drawn, which correspond to three marked vertices in R (the edges of remaining vertices in R are not drawn). A note that since the calculations are done modulo p, the lines (such as the red one) are not necessarily straight "geometric" lines.

We claim that this graph G satisfies all the properties we want.

Firstly, notice that degree of every vertex in this graph is exactly p. Consider a vertex $(x, y) \in L$. For any $a \in \mathbb{F}_p$, there exists a unique $b = y - a \cdot x$, thus, (x, y) has a p edges to vertices $(a, y - a \cdot x) \in R$ by ranging $a \in \mathbb{F}_p$. A similar analysis shows that degree of each $(a, b) \in R$ is also p (as each line in $\mathbb{F}_p \times \mathbb{F}_p$ contains p points inside it).

Secondly, G does not have any 4-cycles. Consider any pairs of vertices (x_1, y_1) and (x_2, y_2) in L. Since there is exactly one line (a, b) that passes through both these two points[1], (x_1, y_1) and (x_2, y_2) only have one shared neighbor. Thus, they cannot be part of a 4-cycle.

Finally, since $n = p^2$, we obtain a bipartite graph with $2n$ vertices and $n^{3/2}$ edges with no 4-cycles.

[1] Given that again our lines are defined mod p and not "geometrically", this may not be obvious at first glance. However, a simple linear algebra proves this: fixing both (x_1, y_1) and (x_2, y_2) gives us two linearly independent equations $a \cdot x_1 + b = y_1$ and $a \cdot x_2 + b = y_2$ for the two variables (a, b). As these two equations lead to a unique solution for (a, b), we can only have one line passing through both points.