Embedding series-parallel graphs into ℓ_1

1 Introduction

Recall that a *series-parallel graph* consists of an edge, or of two series-parallel graphs connected in series or in parallel. A graph can be decomposed into series-parallel blocks if and only if it has treewidth 2. It is easy to see that these graphs cannot, in general, be embedded isometrically into ℓ_1: consider, for example, the n-cycle. In this lecture, we shall show two different constant-distortion embeddings of series-parallel graphs into ℓ_1, both due to [2].

2 Direct construction

Theorem 1 (Main theorem, [2])**.** *Let $G = (V, E)$ be a weighted graph with treewidth 2, and let d be the corresponding metric. Then there exists a map $f: V \to \ell_1$ and a constant $c < 14$ such that for every $u, v \in V$,*

$$\tfrac{1}{c} d(u,v) \le \|f(u) - f(v)\|_1 \le d(u,v).$$

Moreover, this embedding preserves the lengths of edges and can be computed in polynomial time.

We shall use the following characterization of treewidth-2 graphs while proving theorem 1.

Fact 1. *Any treewidth-2 graph can be constructed as follows: start with a single edge e_0, and repeatedly attach a new vertex x to both endpoints of an arbitrary edge (s,t) in the existing graph (this edge is said to be the parent of x). Finally, delete some set of edges.*

It is no loss of generality to assume that no edges are deleted, because we can replace any deleted edge with an edge whose length equals the distance between its endpoints in the graph metric. In constructing the embedding, we shall consider the sequence of graphs $e_0 = G^2, G^3, \dots, G^n = G$ occurring during the composition procedure described above.

Observe that the metric induced by each of these intermediate graphs agrees with the metric in the final graph, i.e., $d_{G^i} = d|_{G^i}$. Therefore, to prove the theorem, it suffices to show how to extend an ℓ_1-embedding of G^i into one for G^{i+1}.

Proof of theorem 1. In the base case $G^2 = (u, v)$, embed u and v as 0 and $d(u, v)$, respectively.

For the inductive case, assume that we have a map $g: G^{i-1} \to \ell_1^d$ satisfying the conditions of the theorem, and that G^i is obtained by attaching a new vertex x to the endpoints of the edge (s,t). Consider the following three embeddings of G^i

$$\begin{array}{ccc} G^{i-1} \mapsto g(G^{i-1}) & \qquad G^{i-1} \mapsto g(G^{i-1}) & \qquad G^{i-1} \mapsto 0 \\ x \mapsto g(s) & \qquad x \mapsto g(t) & \qquad x \mapsto 1 \end{array}$$

Each of these is an ℓ_1 embedding, and therefore, any weighted sum of them is also an l_1-embedding. In order to preserve edge distances in G^{i-1} we must assign weights P_s and P_t to the first two embeddings such that $P_s + P_t = 1$. Call the weight for the cut metric embedding δ. Imposing the conditions that the lengths of the edges (x, s) and (x, t) are preserved, we can compute all these weights:

$$P_s = \frac{-d(s,x) + d(t,x) + d(s,t)}{2d(s,t)} \quad P_t = \frac{d(s,x) - d(t,x) + d(s,t)}{2d(s,t)} \quad \delta = \frac{d(s,x) + d(t,x) - d(s,t)}{2}.$$

Observe that all these quantities are nonnegative. Since this embedding preserves edge distances, it is clearly nonexpanding. By Lemma 2 below, the contraction is at most 14. $\square$

Before stating and proving a bound on the edge contraction, we introduce a new interpretation of the weights described above. Instead of composing the graph starting from one edge, we can think of decomposing the graph by deleting vertices in the opposite order. To delete a vertex x with parent edge (s, t), we take away the cut metric separating x from the remaining graph, and then 'collapse' x to s or t, with probabilities P_s and P_t respectively.

Lemma 2 (Bounding the contraction)**.** *Let $x, y \in G$. Then for any constant $\xi \in (1/2, 1)$, we have*

$$\|f(x) - f(y)\|_1 \geq \tfrac{(1-\xi)(2\xi-1)}{1+\xi} d(x, y).$$

In particular, for $\xi = 3/4$, the contraction is at most 14.

Proof. An edge e is said to be an *ancestor* of a vertex x if it is either the parent edge (s, t) of x, or it is an ancestor of either s or t. We shall prove this lemma first in the special case when y is incident on an ancestor edge of x, and then in the general case.

When y is incident on an ancestor edge of x, view this embedding as a random process as described earlier, and consider the sequence of parent edges $(s_1, t_1), \ldots, (s_k, t_k)$ involved in collapsing x to $y = s_k$. For notational convenience, denote $s_0 = t_0 = x$. Define

$$L_i = d(s_i, t_i) \qquad \alpha_i = d(s_{i-1}, s_i) \qquad \beta_i = d(t_{i-1}, t_i),$$

let P_s^{i-1} and P_t^{i-1} be the collapse probabilities at the ith step, and let Δ_i denote the expected sum of the weights of the cut metrics removed while collapsing x to (s_i, t_i).

We shall analyze the case when $t_i = t_{i-1}$ but $s_i \neq s_{i-1}$ (the other possibility is symmetric). Then the following claim holds:

Claim 3. *Under the above conditions,*

(a) if $P_s^{i-1} \geq \xi$, then $\|f(x) - f(s_i)\|_1 \geq \|f(x) - f(s_{i-1})\|_1 + (2\xi - 1)\alpha_i$.

(b) if $P_t^{i-1} \geq \xi$, then $\|f(x) - f(s_i)\|_1 \geq \|f(x) - f(t_{i-1})\|_1 + (2\xi - 1)L_i$.

(c) otherwise, $\|f(x) - f(s_i)\|_1 + \frac{2\xi}{1-\xi}(\Delta_i - \Delta_{i-1}) \geq \|f(x) - f(s_{i-1})\|_1 + \alpha_i$.

We shall construct a path π from s_k to s_0 as follows: suppose we have reached vertex $s_i \neq s_{i-1}$ (other cases similar). If $P_t^{i-1} > \xi$, then add (s_i, t_{i-1}) to π, otherwise, add (s_i, s_{i-1}) to the path instead. On reaching s_1 or t_1, add one more edge to reach $x = s_0$ and end the path.

For every edge (π_{j-1}, π_j) in this path, claim 3 tells us that

$$\|f(x) - f(\pi_j)\|_1 - \|f(x) - f(\pi_{j-1})\|_1 + \tfrac{2\xi}{1-\xi}(\Delta_{\pi_j} - \Delta_{\pi_{j-1}}) \geq (2\xi - 1)d(\pi_{j-1}, \pi_j)$$

(here, Δ_{π_j} is defined to be Δ_i, where i is the least index such that $\pi_j \in \{s_i, t_i\}$). Summing these inequalities along the path, we get

$$\|f(x) - f(y)\|_1 + \tfrac{2\xi}{1-\xi}\Delta_k \geq (2\xi - 1)(\text{path length}) \geq (2\xi - 1)d(x, y). \tag{1}$$

The result follows, because $\|f(x) - f(y)\|_1 \geq \Delta_k$.

In the general case, let $f = (s,t)$ be the last common ancestor of x and y. Then either (A) f separates x and y (i.e., every $x - y$ path passes through either s or t), or (B) there is a vertex q whose parent is (s,t) such that (s,q) is an ancestor of x and (t,q) is an ancestor of y.

For sub-case (A), let P_s and P_t denote the probabilities that x collapses to s and t respectively, and let Δ denote the expected weight of the cut metrics removed in the collapse. Let P'_s, P'_t, and Δ' denote the corresponding quantities for y. Then

$$\|f(x) - f(y)\|_1 = \Delta + \Delta' + (P_s P'_t + P_t P'_s)d(s,t).$$

We can check (by elementary computations) that $P_s P'_t + P_t P'_s \geq \frac{1}{2}\min\{P_s + P'_s, P_t + P'_t\}$. Using this fact and assuming without loss of generality that the minimum is achieved at t, we get

$$\|f(x) - f(y)\|_1 \geq \tfrac{1}{2}(\|f(x) - f(s)\|_1 + \Delta) + \tfrac{1}{2}(\|f(y) - f(s)\|_1 + \Delta').$$

But it follows from our path-length inequality (1) that

$$\tfrac{1}{2}(\|f(x) - f(s) + \Delta) \geq \tfrac{(1-\xi)(2\xi-1)}{1+\xi}d(x,s),$$

and a similar result holds for y. This lets us reduce the distortion of the $x - y$ distance to the previous case.

The proof for case (B) is a simple modification of this argument, and will not be pursued in these notes. $\square$

Proof of claim 3. In all three cases, we have

$$\begin{aligned}
\|f(x) - f(s_{i-1})\|_1 &= \Delta_{i-1} + P_t^{i-1}L_{i-1};\\
\|f(x) - f(t_{i-1})\|_1 &= \Delta_{i-1} + P_s^{i-1}L_{i-1};\\
\|f(x) - f(s_i)\|_1 &= \Delta_{i-1} + P_t^{i-1}L_i + P_s^{i-1}\alpha_i.
\end{aligned}$$

For case (a), rewrite the last of these as

$$\begin{aligned}
\|f(x) - f(s_i)\|_1 &= \Delta_{i-1} + P_t^{i-1}(L_i + \alpha_i) + (P_s^{i-1} - P_t^{i-1})\alpha_i\\
&\geq \Delta_{i-1} + P_t^{i-1}L_{i-1} + (P_s^{i-1} - P_t^{i-1})\alpha_i\\
&= \|f(x) - f(s_{i-1})\|_1 + (2\xi - 1)\alpha_i.
\end{aligned}$$

The argument for (b) is very similar. For (c), let δ_{i-1} be the weight of the cut removed while collapsing s_{i-1} to (s_i, t_i). Then

$$\Delta_i - \Delta_{i-1} = P_s^{i-1}\delta_{i-1} + P_s^{i-1}\frac{\alpha_i + L_{i-1} - L_i}{2},$$

and it follows that

$$\|f(x) - f(s_i)\|_1 + \frac{2P_t^{i-1}}{P_s^{i-1}}(\Delta_i - \Delta_{i-1} = [\Delta_{i-1} + P_t^{i-1}L_i + P_s^{i-1}\alpha_i] + [P_t^{i-1}(\alpha_i + L_{i-1} - L_i)]$$
$$= \|f(x) - f(s_{i-1})\|_1 + \alpha_i.$$

□

We note that due to a recent tighter analysis by William Evans and MohammadAli Safari, it can be shown that this embedding actually achieves a distortion of at most 6 [1].

3 Embedding into tree metrics

The previous section discussed how to embed series-parallel graphs into ℓ_1 directly. Here we discuss an alternative approach inspired by a constant-distortion embedding of outerplanar graphs into probability distributions over dominating tree metrics. As tree metrics are isometrically embeddable into ℓ_1, providing low-distortion embeddings into tree metrics also provides low-distortion embeddings into ℓ_1.

Definition 1. *A mapping f from a metric space (X, d) to another metric space (X', d') is said to be an* embedding into a dominating tree metric *if (X', d') is a tree metric and $d(x, y) \leq d'(f(x), f(y))$ for all $x, y \in X$. A probability distribution $\mathcal{F}$ over embeddings into dominating tree metrics D-*approximates *(X, d) if each $f \in \mathcal{F}$ is an embedding into a dominating tree metric and $E_{f\in\mathcal{F}}[d'(f(x), f(y))] \leq D \cdot d(x, y)$ for all $x, y \in X$.*

When (X, d) is a metric supported by some graph G and the support of $\mathcal{F}$ only contains spanning trees of G, the expansion of distances is maximized over some edge. Thus the condition that no pair of vertices has its distance expanded by a factor of more than D can be replaced by the condition that no edge is expanded by a factor of more than D. In our applications we will always be dealing with distributions over spanning trees of graph metrics.

Definition 2. *A graph is said to be* outerplanar *if every vertex of the graph lies on the outside face. An outerplanar graph G can be constructed by repeatedly applying one of the following two composition rules to create G_i from G_{i-1} starting from a path or cycle that we refer to as G_0 and ending with $G_m = G$.*

- *Choose an edge e_i on the outer face of G_{i-1} and attach a path P_i to both of e_i's endpoints.*
- *Choose a vertex u_i on the outer face of G_{i-1} and attach a path P_i to u_i.*

We add the further restriction that each edge e has at most one path joined to its ends throughout the course of the composition procedure.

In the above definition, a path P_i is called *slack* if either P_i is attached to a single vertex, or the edge e_i it is attached to has at most half the length of P_i. A slack composition of G is a composition for G where every P_i is slack.

Lemma 4. *If G is an outerplanar graph with a slack composition then there is a distribution of embeddings into dominating tree metrics that 4-approximates G.*

Proof. The distribution is over spanning trees of G and is defined inductively. Recall by an earlier argument that we need only bound the expected expansion on edges of G. Let $G_0 = P_0, P_1, \ldots, P_m$ be a slack composition.

If G_0 is a path, then the distribution contains only one tree (G_0). If G_0 is a cycle then we randomly delete an edge from G_0 where each edge is deleted with probability proportional to its weight. If an edge (u, v) has weight w and the total weight of all edges in the cycle is W, then the expected distance between u and v in the resulting path is

$$\frac{w}{W} \cdot (W - w) + \frac{W - w}{W} \cdot w = 2w - 2\frac{w^2}{W} \leq 2w$$

Now when going from G_i to G_{i+1}, if P_i attaches to a single vertex then we include all of P_i in our tree with probability 1. Otherwise, we delete an edge from P_i with probability proportional to its weight. If the edge P_i is attached to has weight d in G, we can inductively assume its expected distance in the distribution over trees thus far is at most $4d$. Let w be the weight of an edge e of P_i and W be the total weight of P_i. Then the expected distance between the endpoints of e in the tree distribution is at most

$$\frac{w}{W} \cdot (4d + W - w) + \frac{W - w}{W} \cdot w \leq w \left(4 \cdot \frac{d}{W} + 2\right)$$

Using that the composition is slack, we have $d/W \leq 1/2$, giving that the expected distance between the endpoints of the edge is at most $4w$. □

Lemma 5. *For any outerplanar graph G, there is an outerplanar graph H with slack composition such that G embeds into H with distortion at most 2.*

Proof. Let $G_0 = P_0, P_1, \ldots, P_m$ be a composition procedure for G. We show a procedure for gradually modifying the P_i so that we end with a slack composition for an outerplanar graph H dominated by G where no distances shrink by a factor more than 2.

We initially set the composition procedure for H to be $H_0 = Q_0, Q_1, \ldots, Q_m$ with $Q_i = P_i$ for each i. We then visit the Q_i in order for $i = 1, 2, \ldots, m$. If Q_i is slack or connects to a single vertex, we continue to Q_{i+1}. Otherwise, Q_i is attached to some edge e with weight $w(e)$ and the length L of Q_i satisfies $w(e) \leq L < 2w(e)$. We multiply the weight of every edge of Q_i by $w(e)/L$ to obtain a new path Q'_k with length $w(e)$. Now we remove Q_i from the composition procedure from H, and if e was added to H by the path Q_k then we modify Q_k to use Q'_k in place of e. Since the only path adjoined to e was Q_i, our composition procedure for H remains valid. Now, suppose e' is an edge of Q_i that has just been shrunk in weight. There may be edges added in future paths that now have weights larger than the distance between their endpoints in H. For all such edges we decrease their weights to match the new shortest path lengths. This changing of edge weights has no actual effect on the shortest path metric.

In this procedure edges either are removed or are decreased in weight, so G dominates H. Since each edge has its weight altered in a way that affects the shortest path metric at most once, is never increased, and is decreased by a factor of at most 2, it follows that distances in G are at most twice those in H. □

Combining lemmas 4 and 5 shows that any outerplanar graph is 8-approximated by a distribution of embeddings into dominating tree metrics.

4 Using bundles to embed treewidth-2 graphs into ℓ_1

Definition 3. *A graph is said to be* biconnected *if every two edges in the graph lie on a common simple cycle (here simple refers to edges not being repeated, but vertices may repeat themselves). The maximal biconnected subgraphs of a graph G are called its* blocks *or* biconnected components.

Lemma 6. *If for each block B_i of a connected graph G we can embed the metric supported by B_i into (X, d) with no contraction and distortion at most D_i, then we can embed the metric supported by G into (X, d) with distortion at most $\max_i\{D_i\}$.*

Proof. We are given that for each B_i there is an embedding f_i into (X, d) with distortion D_i. We extend each f_i to an embedding $\tilde{f}_i$ on all vertices of G by setting $\tilde{f}_i(x) = f_i(y)$, where y is the vertex of B_i closest to x in G.

Also, we imagine a tree where each B_i is represented by a node and there is an edge from one node to another if there is a an edge between the corresponding blocks. For each edge between blocks we define a cut metric between the nodes on opposite ends of the edge, and the weight of this cut metric will be the weight of that edge in G. Our final embedding for G is then the concatenation of all these cut metric embeddings together with all the $\tilde{f}_i$. □

Bundles are a special type of series-parallel graph which will be useful for embedding general treewidth-2 graphs into ℓ_1. Recall that treewidth-2 graphs are those whose biconnected components are series-parallel, so by Lemma 6 it suffices to only consider embeddings of series-parallel graphs. We then show how to embed a series-parallel graph under constant approximation into a distribution over graphs whose biconnected components are bundles, then how to embed bundles into ℓ_1 with distortion 2.

Definition 4. *A* bundle *is a series-parallel graph such that any path between its terminals has the same length. A bundle then has a well-defined* length *which is the length of any path between its terminals.*

Lemma 7. *If G is a bundle then G can be embedded into ℓ_1 with distortion at most 2.*

Proof. We use the embedding from Lemma 2. The analysis is similar but simpler. In the case where node s is an ancestor of node x, there is no distortion. This can be proved by induction, noticing that the cut metrics defined in the embedding have 0 weight for bundles. The result for this case then carries over using exactly the same proof technique to the case where neither vertex is on an ancestor edge of the other to give a distortion of 2 in that case. □

Now we introduce a relaxed composition procedure that also can be used to create any treewidth-2 graph. We first start with P_0 a path, and at each step we are allowed to add a new path P_i either to an existing vertex or to the endpoints of an existing edge. This is identical to the composition procedure for outerplanar graphs, except it is not required that the vertices of the new path are on the outer face.

Recall that a slack path is a path added to either a vertex or to an edge of at most half its length. A *taut* path is a path adjoined to an edge of length equal to the path's length. If a series-parallel graph has a slack-taut composition, i.e. a composition where each added path is slack or taut, we can reinterpret the composition as being slack where we add bundles at each step instead of paths. When a taut path is attached to an edge, we can incorporate the path into the bundle already containing that edge. Slack paths in the composition give us the starts of new bundles. Now, in a similar vein to outerplanar graphs, when adding a slack bundle of length L with terminals s, t we choose a number $r \in [0, L]$ uniformly at random and remove all edges from the bundle crossing the point at a distance r from s. This gives us an embedding of slack-taut graphs into graphs whose biconnected components are bundles, and the analysis is identical to that in the outerplanar case, giving a distortion upper bound of 4. The bundles can then be embedded into ℓ_1 with distortion at most 2 and we apply Lemma 6.

What's left to show is that we can embed any series-parallel graph into a series-parallel graph with slack-taut composition. Using a proof similar to that of Lemma 5, we can obtain such an embedding with distortion at most 2. In fact, the only difference in the proof is that when a path in the composition is not slack, we cannot remove the edge it is attached to after scaling the path's length downward. This is because, unlike in outerplanar graph compositions, series-parallel graph compositions may attach many paths to a single edge. Thus, deleting an edge may make our composition procedure invalid. Instead we observe that keeping the edge present makes the scaled-down path taut, and so we still maintain that our composition is slack-taut. We can thus state the following theorem.

Theorem 8. *Metrics supported by treewidth-2 graphs embed into ℓ_1 with distortion at most* 16.

References

[1] Will Evans, MohammadAli Safari. On the ℓ_1 embedding of series-parallel graphs. Manuscript. 2006.

[2] Anupam Gupta, Ilan Newman, Yuri Rabinovich, Alistair Sinclair. Cuts, Trees and ℓ_1-Embeddings of Graphs. *Combinatorica* 24(2): 233–269. 2004.