COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA 24 (2) (2004) 233–269

CUTS, TREES AND ℓ_1-EMBEDDINGS OF GRAPHS*

ANUPAM GUPTA†, ILAN NEWMAN, YURI RABINOVICH,
ALISTAIR SINCLAIR‡

Received November 27, 2000
Revised August 7, 2002

Motivated by many recent algorithmic applications, this paper aims to promote a systematic study of the relationship between the topology of a graph and the metric distortion incurred when the graph is embedded into ℓ_1 space. The main results are:

1. Explicit constant-distortion embeddings of all series-parallel graphs, and all graphs with bounded Euler number. These are the first natural families known to have constant distortion (strictly greater than 1). Using the above embeddings, algorithms are obtained which approximate the sparsest cut in such graphs to within a constant factor.
2. A constant-distortion embedding of outerplanar graphs into the restricted class of ℓ_1-metrics known as "dominating tree metrics". A lower bound of $\Omega(\log n)$ on the distortion for embeddings of series-parallel graphs into (distributions over) dominating tree metrics is also presented. This shows, surprisingly, that such metrics approximate distances very poorly even for families of graphs with low treewidth, and excludes the possibility of using them to explore the finer structure of ℓ_1-embeddability.

1. Introduction

Let $G=(V,E)$ be an undirected graph. Each assignment of non-negative weights to the edges of G naturally defines a metric space (V,μ),[1] where

Mathematics Subject Classification (2000): 05C12, 05C85, 68R10, 90C27

* A preliminary version of this work appeared in *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, 1999, pp. 399–408.

† This work was done while the author was at the University of California, Berkeley.

‡ Supported in part by NSF grants CCR-9505448 and CCR-9820951.

[1] More correctly, a *semi-metric* space, since we allow $\mu(x,y)=0$ even when $x\neq y$.

for each pair of vertices $x,y \in V$, $\mu(x,y) = d_G(x,y)$ is the shortest-path distance between them. We say that the metric μ is *supported on* (or *generated by*) G. Let (S,ρ) be another metric space. An *embedding* of G into (S,ρ) is a mapping $\phi : V \to S$. The *distortion* of ϕ is the smallest value $c \geq 1$ such that

$$d_G(x,y) \leq \rho(\phi(x),\phi(y)) \leq c\, d_G(x,y) \qquad \forall x,y \in V.$$

Thus the distortion measures the maximum factor by which any distance is stretched in the embedding. (This is a slightly restricted definition, in which we assume that no distances are shrunk. See Section 2 for a general definition.)

In recent years, the idea of embedding a graph into a "nice" metric space with low distortion has emerged as a useful ingredient in the design and analysis of algorithms in a variety of domains. "Nice" metric spaces are those with well-studied structural properties, such as Euclidean or ℓ_1 space, or weighted trees and distributions over them. A very incomplete list of applications includes approximation algorithms for graph and network problems, such as sparsest cut [27, 2], minimum bandwidth [18, 8], low-diameter decompositions [27], and optimal group Steiner trees [20, 10], and online algorithms for metrical task systems and file migration problems [4, 6]. These applications, together with its intrinsic mathematical interest, have made the study of low-distortion embeddings a significant field in its own right.

Most of the embeddings considered in the literature, notably [9, 4, 27], have been for metrics supported on general graphs, and give results that bound the worst-case distortion over all graphs. However, when the input graph has some special structure, it is plausible that better embeddings can be found. This is quite intuitive: it is clear that *any* metric is generated by the complete graph on its points, while only a very limited set of metrics can be generated by weighting the edges of, say, a tree. Thus the complexity of a metric generated by a graph G intrinsically depends on the topology of G. At present, very little is known about this interplay between the topological and metrical properties of the graph; the search for connections between the two is emerging as an intriguing and challenging area. This paper focuses in particular on the relationship between the topology of graphs and their optimal (or near-optimal) embeddings into ℓ_1 (i.e., real space of arbitrary dimension endowed with the ℓ_1 metric).

Embeddings into ℓ_1 have been widely studied, and are of special importance due to their intimate connection with the problem of finding a *sparsest cut* in multicommodity flow networks, which in turn is a key ingredient in approximate solutions of many other problems in such areas as VLSI layout, network routing and efficient simulations of one network by another (see, e.g., [7,26,24]). Although finding the exact sparsest cut is a computationally hard problem, efficient approximation algorithms for it can be obtained by embedding a natural metric associated with the optimal multicommodity flow into ℓ_1; the approximation ratio depends essentially on the distortion.

One motivation behind this paper is the intriguing conjecture that any metric supported on a *planar* graph (henceforth called a *planar metric*) can be embedded into ℓ_1 with constant distortion. More generally, we conjecture that this holds for *any* family of graphs which excludes a fixed minor. There is some evidence to suggest that planar metrics are better behaved than general metrics with respect to ℓ_1-embeddability. In an interesting recent development, Rao [34] has given an $O(\sqrt{\log n})$-distortion embedding of n-point planar metrics into ℓ_1, while the lower bound for general metrics is $\Omega(\log n)$. This result, and the decomposition lemma of [23] on which it is based, attest to the special structure of planar metrics.

Despite this promise, current techniques are apparently inadequate to resolve the above conjecture. For embeddings into ℓ_1, a celebrated result of Bourgain [9] tells us that *any* metric supported on an n-vertex graph (i.e., any metric on n points) can be embedded into ℓ_1 with distortion $O(\log n)$; unfortunately, the embedding technique is not sensitive to the topology and incurs a $\Omega(\log n)$-distortion even for the metric generated by the unit-weighted path P_n. Similarly, the method of Konjevod *et al.* of finding distributions over dominating trees is limited by a lower bound of $\Omega(\log n)$ for embedding the $n \times n$ grid [1,25]. Lastly, Rao gives embeddings into ℓ_1 by first embedding into ℓ_2, an approach that is limited by a lower bound of $\Omega(\sqrt{\log n})$ for embedding even series-parallel graphs into ℓ_2 [28].

In this paper, we systematically explore how the topology of a graph affects the distortion incurred by ℓ_1-embeddings of metrics supported on it. Using the intimate connection between ℓ_1-embeddability of metrics supported on a graph and multicommodity flow problems defined on it, one can show that graphs all of whose metrics are *isometrically* embeddable into ℓ_1 (i.e., embeddable with distortion 1) are exactly the graphs which exclude $K_{2,3}$ as a minor, which essentially corresponds to the class of *outerplanar* graphs. This fact, which rests on a theorem of Okamura and Seymour [30], is our starting point. As a natural next step, we consider the family of graphs which have K_4 as an excluded minor. These are graphs with treewidth 2, and essentially correspond to the familiar class of *series-parallel* graphs. Our first main result is an explicit ℓ_1-embedding of these graphs with small constant distortion. This is the first natural family known to have a constant distortion strictly bigger than 1. In addition, our construction implies a simple polynomial time algorithm for finding a sparsest cut within a constant factor of optimal in series-parallel graphs. In a similar vein, we also show that any family of graphs with bounded Euler characteristic can be embedded into ℓ_1 with constant distortion. The technique we use for these results is to explicitly construct a set of *cut metrics* whose sum approximates the original graph metric very closely. Cut metrics arise naturally in the study of ℓ_1-embeddability since any ℓ_1-embeddable metric[2] can be represented as a sum of cut metrics with non-negative coefficients, and vice versa [15].

We then go on to study the approximation of a metric by a probability distribution over *(dominating) tree metrics*. Since tree metrics are ℓ_1-embeddable (and so are their non-negative combinations), this gives us an alternative to the cut metrics approach. Furthermore, embeddings based on such metrics have proved particularly easy to work with, and possess additional properties that have been exploited in devising approximation algorithms and online algorithms for many problems (see, e.g., [4,6,3,20,37, 10,12]). It is natural to ask if we can obtain the above embeddability results for outerplanar and series-parallel graphs using these more restricted metrics. The answers are mixed. On the one hand, we show that this is possible for outerplanar graphs, at a small cost: we give an explicit embedding for such graphs into a distribution over dominating tree metrics with distortion 8 (compared to distortion 1 obtained using cuts). On the other hand, we prove a complementary negative result by exhibiting a family of series-parallel graphs for which *any* distribution over dominating tree metrics must necessarily incur a distortion of $\Omega(\log n)$.

Thus we see that the tree metrics approach breaks down at a surprisingly early stage (even for graphs of treewidth 2), which suggests that such embeddings by themselves offer little hope for exploring the finer structure of ℓ_1-embeddings. However, our results also indicate that combining dominating tree metrics with cut metrics is a potentially powerful technique. Indeed, the graphs which give the lower bound for tree embeddings mentioned above can be shown to have extremely simple ℓ_1-embeddings using cuts. Combining these cut metric embeddings with tree embeddings in a careful fashion leads us to an alternative constant distortion embedding for series-parallel graphs.

The organization of the paper closely follows the above outline. After a short section containing some definitions and notation, we briefly illuminate the connection between flows and ℓ_1-embeddings in Section 3. The embeddings of series-parallel graphs and graphs with small Euler number

[2] We shall use the unqualified term "ℓ_1-embeddable" to mean "isometrically embeddable into ℓ_1".

are described in Section 4. Finally, in Section 5, we present our positive and negative results on embeddings into tree distributions, as well as the alternative embedding for series-parallel graphs.

2. Definitions and notation

Metrics. Let X be a set. A function $d: X\times X\to\mathbb{R}^+$ is called a *semi-metric* if it is symmetric, i.e., $d(x,y)=d(y,x)$ for all $x,y\in X$, and $d(x,x)=0$ for all $x\in X$, and also satisfies the *triangle inequality*, i.e., $d(x,z)\le d(x,y)+d(y,z)$ for all $x,y,z\in X$. If, in addition, $d(x,y)=0$ holds only when $x=y$, then d is a *metric*. In this paper, we shall only consider finite semi-metrics. The number of points will usually be denoted by n. Without risk of confusion, the distinction between metrics and semi-metrics may sometimes be blurred. For more details on many of the metric concepts used here, see the book of Deza and Laurent [15].

Given two metric spaces, (V,ν) and (W,μ), and a map $f: V\to W$, define the following quantities:

$$\|f\| = \max_{x,y\in V} \frac{\mu(f(x),f(y))}{\nu(x,y)};$$
$$\|f^{-1}\| = \max_{x,y\in V} \frac{\nu(x,y)}{\mu(f(x),f(y))}.$$

We say that f has *contraction* $\|f^{-1}\|$, *expansion* $\|f\|$ and *distortion* $D(f)=\|f\|\cdot\|f^{-1}\|$. We say that (W,μ) r*-approximates* (V,ν) (or that the *distortion* between μ and ν is at most r) if there exists a map $f: V\to W$ with $D(f)\le r$. Often we shall consider two distance functions μ and ν over the same vertex set V. In such cases, we shall assume that f is the identity map. Also, μ will be said to *dominate* ν if for all $x,y\in V$, $\mu(x,y)\ge\nu(x,y)$.

Let $G=(V,E)$ be an undirected graph. A metric (V,μ) is *supported on* (or *generated by*) G if it is the shortest path metric of G w.r.t. some non-negative weighting of the edges E. Unless specified otherwise, we shall assume that the edge-weights $w(\cdot)$ satisfy $w(e)=\mu(e)$, where μ is the shortest-path metric of G with weights w. Observe that if it is not the case, the edge e can be removed without affecting the metric; such an e will be called *redundant*.

For a set $S\subseteq V$, the *cut metric* δ_S on V is defined by $\delta_S(x,y)=1$ if $|S\cap\{x,y\}|=1$, and $\delta_S(x,y)=0$ otherwise. An important observation is that the ℓ_1-embeddable metrics on V are precisely those metrics which can be written as a sum of cut metrics on V with non-negative coefficients [15]. One implication of this is that if two metrics μ_1 and μ_2 on the same underlying set are ℓ_1-embeddable, then so is their sum $\mu_1+\mu_2$.

Finally, we use the following simple observation throughout the paper: if each block (i.e., biconnected component) G_i of a graph G can be embedded into ℓ_1 with distortion D_i, then G can be embedded into ℓ_1 with distortion $\max_i D_i$. This immediately implies, in particular, that any metric supported on a tree T can be embedded isometrically into ℓ_1. (For a more direct proof of this latter fact, let $(S_e,\overline{S}_e)$ be the cut obtained by deleting an edge e in T; it can be verified that $\mu=\sum_{e\in T} d_T(e)\cdot\delta_{S_e}$ is isometric to the tree metric d_T [15, Prop. 11.1.4].)

Multicommodity flows. A *multicommodity flow network* (V,E,P) is specified by an undirected graph $G=(V,E)$, where E is the set of edges along which flow can be routed, and a set P of unordered pairs of vertices in V between which demands can be placed. In the unrestricted case, where P consists of all pairs of vertices, we shall omit explicit mention of P and refer to the network simply as $G=(V,E)$. Assigning non-negative *capacities* C to the graph edges E and *demands* D to the pairs P gives us a particular instance (V,C,D) of the *multicommodity flow* problem. For background, see the survey by Shmoys [36].

The optimal solution to this problem is the maximum value λ such that there is a multicommodity flow f respecting the edge capacities that satisfies a multiple λ of each demand. We shall refer to λ as *MaxFlow*(V,C,D). Its value (as well as an actual flow f which realizes it) can be found in polynomial time by linear programming.

A closely related problem is the *sparsest cut* problem, which entails finding a partition $(A,\overline{A})$ of V that minimizes the ratio

$$\kappa(A)=\frac{\text{Capacity}(A,\overline{A})}{\text{Demand}(A,\overline{A})}=\frac{C\cdot\delta_A}{D\cdot\delta_A}.$$

(To make sense of the inner products, note that C,D and the cut metric δ_A can all be viewed as elements of the vector space $\mathbb{R}^{\binom{|V|}{2}}$.) We shall refer to $\kappa=\min_A\kappa(A)$ as *MinCut*(V,C,D).

In the sequel it will be convenient to use the following identities (see, e.g., [27] or [15, page 135] for the proofs):

$$\text{MaxFlow}(V,C,D)=\min_{\delta\in M(V)}\frac{C\cdot\delta}{D\cdot\delta};\quad \text{MinCut}(V,C,D)=\min_{\delta\in M_1(V)}\frac{C\cdot\delta}{D\cdot\delta}, \tag{2.1}$$

where $M(V)$ is the set (in fact, a convex cone) of all metrics over V, and $M_1(V)$ is the set (again, a convex cone) of all ℓ_1-embeddable metrics over V.

As $M_1(V) \subseteq M(V)$ (the inclusion being strict for V of size ≥ 5), it is always the case that MaxFlow $\leq$ MinCut.

In contrast with the case when there is just one commodity, the MinCut is not equal to the MaxFlow in general. The ratio $\gamma \geq 1$ between the MinCut and the MaxFlow is called the *gap* of the instance (V,C,D). From the computational point of view, computing the value of the MinCut (and hence also the value of γ) is an NP-hard problem.

Graphs and Minors. An *outerplanar* graph G is a planar graph with an embedding in the plane so that every vertex lies on the outer (unbounded) face. A *series-parallel* graph $G=(V,E)$ with *terminals* $s,t \in V$ is either a single edge (s,t), or a *series combination* or a *parallel combination* of two series-parallel graphs G_1 and G_2 with terminals s_1,t_1 and s_2,t_2. The series combination of G_1 and G_2 is formed by setting $s=s_1$, $t=t_2$ and identifying $s_2=t_1$; the parallel combination is formed by identifying $s=s_1=s_2$, $t=t_1=t_2$.

The graph $G=(V,E)$ has an *H-minor* if there exists a sequence of edge-deletion and edge-contraction operations on G which results in a graph G' that is isomorphic to H. Note that each vertex of G' corresponds to a (connected) set of vertices of G which were contracted to it. For $U \subseteq V$, we say that G has an *H-minor w.r.t. U* if it has an H-minor G' such that for every vertex of G', the corresponding set of vertices of G contains a vertex from U. Finally, we say that G is *H-free* (w.r.t. U) if it has no H-minor (w.r.t. U).

It is well known that K_4-free graphs are those whose blocks are series-parallel graphs [16, p. 185], and that $K_{2,3}$-free graphs are those whose blocks are either outerplanar or isomorphic to K_4 [16, p. 81].

Finally, the *Euler number* of an undirected connected graph G is defined as $\chi(G) = |E(G)| - |V(G)| + 1$. (Throughout this paper, the symbol $\chi(G)$ denotes the Euler number and *not* the chromatic number.)

3. Multicommodity flows, metrics and graphs

Multicommodity flows have long been an object of study in combinatorial optimization (see [19] for a historical survey). The classical theory was concerned mainly with the following question: Under what conditions on the flow network (V,E,P) is the MaxFlow equal to the MinCut for every setting of capacities C and demands D? As it turns out, this question is equivalent to the following question concerning the ℓ_1-embeddability of metrics: What are the conditions on (V,E,P) such that, for every metric μ supported on $G=(V,E)$, there exists an ℓ_1-embeddable metric ν on V such that μ dominates ν, and $\mu=\nu$ on P? [35, Section 3]

In light of this equivalence, the classical results about flows (in cases where the gap $\gamma = 1$) have consequences for ℓ_1-embeddability and *vice versa.* For instance, a well-known theorem due to Okamura and Seymour [30] says that if $G = (V, E)$ is a planar graph with outer face F, and P consists only of pairs of vertices in F, then the MaxFlow and MinCut are equal for all instantiations of C and D. Taking $G = (V, E)$ to be an outerplanar graph, letting P consist of all pairs in V and using the above equivalence, we can infer that all metrics supported on outerplanar graphs can be isometrically embedded into ℓ_1. (See also [14] for a direct argument.)

To state this and other such results succinctly, let us introduce some notation. For a metric μ, let $c_1(\mu)$ be the minimum distortion between μ and ρ, where ρ ranges over all ℓ_1 metrics, and let $c_1(G)$ be the maximum value of $c_1(\mu)$ for all metrics μ supported on G. Hence, we have just seen that $c_1(G) = 1$ for every outerplanar graph G.

In fact, this turns out to be almost a characterization of graphs G with $c_1(G) = 1$. The full picture is that $c_1(G) = 1$ iff G is $K_{2,3}$-free. On the one hand, as mentioned earlier, each block of a $K_{2,3}$-free graph is either outerplanar or isomorphic to K_4, and a graph is ℓ_1-embeddable iff each of its blocks is. We have already seen that outerplanar graphs are ℓ_1-embeddable; it is also well known that the same holds for any metric on four points [15, Example 11.1.8]. Thus, for every $K_{2,3}$-free graph G, $c_1(G) = 1$. Conversely, it is well known that the metric of the unit-weighted $K_{2,3}$ is *not* ℓ_1-embeddable [15, Example 6.3.2]. Now if G has a $K_{2,3}$-minor, consider the sequence of edge contractions and deletions which turn G into $K_{2,3}$. Assigning ∞ to each deleted edge, 0 to each contracted edge, and 1 to the remaining edges, we obtain a semi-metric supported on G and coinciding (as a metric space) with that of the unit-weighted $K_{2,3}$. Thus, $c_1(G) \geq c_1(K_{2,3}) > 1$. Hence we have the following characterization:

Proposition 3.1. *The class of graphs for which $c_1(G) = 1$ is exactly the class of $K_{2,3}$-free graphs.*

Much recent research on multicommodity flows has been directed towards the case where equality does not hold, and to finding good bounds on the ratio γ between the MinCut and the MaxFlow. This study was pioneered in the paper of Leighton and Rao [26], and the results presented there were extended in a long sequence of papers by several authors (see [36] for a detailed account). The best results known [27,2] show that for any flow network (V, E, P), the gap between the MaxFlow and the MinCut can never be more than $O(\log |P|)$, and hence $O(\log n)$. This bound is tight when $G = (V, E)$ is a constant-degree expander, all edge capacities are unity and there is unit demand between all pairs of vertices. Better results have been obtained for planar graphs, showing that in such graphs the gap γ never exceeds $O(\sqrt{\log n})$ [34], and in fact is bounded by a constant in the special case of uniform demands [23].

An intimate relationship between the gap γ and $c_1(G)$ holds even in the case where the MaxFlow is not equal to the MinCut, and provides a compelling motivation for studying the quantity $c_1(G)$.

Theorem 3.2. *For any graph $G=(V,E)$, the worst possible gap γ attained by a multicommodity flow problem on G is exactly $c_1(G)$.*

Proof. The direction $\gamma \le c_1(G)$ was shown already in [27]. Indeed, by definition of c_1, for every metric μ supported on G, there exists an ℓ_1-embeddable metric δ which distorts μ by at most $c_1(G)$. But then, by definition of distortion, $\frac{C\cdot\delta}{D\cdot\delta} \le c_1(G)\,\frac{C\cdot\mu}{D\cdot\mu}$, and in view of (2.1) we are done.

For the other, apparently new, direction $\gamma \ge c_1(G)$, it will be convenient to use an equivalent dual definition of $c_1(\mu)$ for a metric μ on V:

$$c_1(\mu) = \max_{(C,D)} \frac{D\cdot\mu}{C\cdot\mu}\,, \tag{3.2}$$

where the maximum is taken over all non-negative vectors C,D indexed by ordered pairs of vertices of V which satisfy the restriction $\frac{D\cdot\delta}{C\cdot\delta} \le 1$ for any ℓ_1-embeddable metric δ on V. The proof of this equality follows from general facts about convex cones, and is deferred to the appendix.

By this dual definition, there exists a metric μ supported on G, and non-negative vectors $C,D \subseteq \mathbb{R}^{\binom{|V|}{2}}$, such that $\frac{D\cdot\mu}{C\cdot\mu} = c_1(G)$, while for any ℓ_1-embeddable metric δ we have $\frac{D\cdot\delta}{C\cdot\delta} \le 1$. First we claim that, without loss of generality, one may assume that C vanishes outside $E(G)$. Indeed, assume that for some pair of vertices $\{i,k\} \notin E(G)$, the value $C(i,k)$ is strictly positive. Since μ is supported on G, there exist edges $e_1 = (j_0,j_1), e_2 = (j_1,j_2),\dots,e_q = (j_{q-1},j_q)$ in G such that $j_0 = i, j_q = k$ and $\mu(j_0,j_q) = \mu(j_0,j_1) + \dots + \mu(j_{q-1},j_q)$. Define a new vector C' by

$$\begin{aligned} C'(i,k) &= 0, \\ C'(j_{r-1},j_r) &= C(j_{r-1},j_r) + C(i,k) \qquad &&\text{for each } r = 1,2,\dots,q, \text{ and} \\ C'(u,v) &= C(u,v) &&\text{otherwise.} \end{aligned}$$

Now, the pair C',D can replace the pair C,D in the above definition of $c_1(G)$. Clearly, for any metric δ on V we have $C'\cdot\delta \ge C\cdot\delta$; in particular, for any ℓ_1-embeddable δ we have $(D\cdot\delta)/(C'\cdot\delta) \le (D\cdot\delta)/(C\cdot\delta) \le 1$, as required by (3.2). On the other hand, for μ, the "worst" metric supported on G, we have the equality $C' \cdot \mu = C \cdot \mu$, and thus $(D \cdot \mu)/(C' \cdot \mu) = (D \cdot \mu)/(C \cdot \mu) = c_1(G)$. Repeating this updating procedure for all non-edges of G, we arrive at a vector C that vanishes outside $E(G)$.

Employing such a pair C, D and bearing in mind the definitions of MinCut and MaxFlow given in (2.1), we conclude that

$$\gamma \geq \gamma(V, C, D) = \frac{\mathrm{MinCut}(V, C, D)}{\mathrm{MaxFlow}(V, C, D)} \geq \frac{\min_{\delta \in M_1(V)} (C \cdot \delta)/(D \cdot \delta)}{(C \cdot \mu)/(D \cdot \mu)}$$
$$\geq \frac{D \cdot \mu}{C \cdot \mu} = c_1(G). \qquad \blacksquare$$

Recall that by Proposition 3.1, the graphs for which $c_1(G) = 1$ are exactly the $K_{2,3}$-free graphs. It is no coincidence that this characterization involves excluded minors. Observe that the graph-theoretic function c_1 is minor-monotone, i.e., if H is a minor of G then $c_1(G) \geq c_1(H)$. Indeed, edge deletion corresponds to assigning the edge the value ∞, while edge contraction corresponds to assigning it the value 0. The principal consequence of this observation is that $\mathcal{F}_c$, the family of all graphs G with $c_1(G) \leq c$, is minor-closed for any c. Hence, by a celebrated theorem of Robertson and Seymour, any $\mathcal{F}_c$ can be characterized in terms of forbidden minors (see, e.g., [16, Cor. 12.5.3]).

Another consequence of monotonicity of $c_1(G)$ is that the set $\{c_1(G)\} \subset \mathbb{R}$, where G ranges over all finite graphs, contains no infinite descending sequence. Indeed, assume that $c_1(G_1) > c_1(G_2) > c_1(G_3) > \dots$ is an infinite descending sequence. By a theorem of Robertson and Seymour, there must exist G_i and G_j with $j > i$ such that G_i is a minor of G_j (see, e.g., [16, Thm. 12.5.2]), contradicting the monotonicity of c_1. In particular, every point of $\{c_1(G)\}$ contains a unique "next to the right" point. Currently, we only know that the smallest point of this set is 1, and the second smallest is $c_1(K_{2,3})$, which can be shown to be $4/3$.

An intriguing conjecture, and one of the main motivations behind this paper, is that for any non-trivial minor-closed family $\mathcal{F}$ of graphs, there exists a constant $c_{\mathcal{F}} \geq 1$ such that for all $G \in \mathcal{F}$, $c_1(G) \leq c_{\mathcal{F}}$.

The results in the next section provide some evidence in support of this conjecture. We consider the next natural minor-closed class of graphs containing $K_{2,3}$, namely the class of series-parallel graphs, and show that they are ℓ_1-embeddable with constant distortion. In addition, we bound the distortion $c_1(G)$ of a graph in terms of its Euler characteristic alone, and thus establish an infinite sequence of natural minor-closed families with constant distortion, namely those with bounded Euler characteristic.

4. Constant-distortion embeddings for some graph families

In this section, we shall present explicit constant-distortion embeddings into ℓ_1 of the natural minor-closed families of series-parallel graphs, and of graphs with bounded Euler characteristic. These are the first non-trivial results exhibiting (necessarily) non-isometric embeddings of graph families with constant distortion.

4.1. Series-parallel graphs

Our goal will be to show that any metric supported on a series-parallel graph is embeddable in ℓ_1 with constant distortion. In fact, our argument is presented for the slightly more general class of treewidth-2 graphs, i.e., graphs whose blocks are series-parallel graphs. Recall that this is a minor-closed family with K_4 as the excluded minor. We have not attempted to achieve the best possible constant distortion, which we believe is rather less than the value of (just under) 14 shown here.

Theorem 4.1. *Let $G=(V,E)$ be a weighted graph with treewidth 2, and let $\mu=\mu_G$ be the metric induced by the edge weights of G. Then there exists an ℓ_1-embeddable metric $\widetilde{\mu}$ and a constant $c<14$ such that for every $u,v\in V$,*

$$\frac{1}{c}\,\mu(u,v) \leq \widetilde{\mu}(u,v) \leq \mu(u,v).$$

Moreover, this embedding preserves the length of edges, i.e., for every $(u,v)\in E$, $\widetilde{\mu}(u,v)=\mu(u,v)$. Finally, $\widetilde{\mu}$ can be computed in polynomial time.

Before proving the theorem, let us briefly discuss some properties of treewidth-2 graphs and the metrics generated by them. According to one of the many alternative definitions, treewidth-2 graphs can be constructed using the following composition procedure. Start with a single edge e_0, and repeatedly attach a single new vertex to the endpoints of an already existing edge (which we call the *parent* edge of the vertex); finally, after all the vertices have been attached, remove an arbitrary subset of the edges. We shall consider a weighted treewidth-2 graph G together with the sequence of intermediate weighted graphs $G^2,G^3,\dots,G^n=G$ occurring during its composition, where G^2 is the initial edge e_0. Each new edge $e=(u,v)$ will be endowed with weight $\mu(u,v)$, where μ is the metric induced by G. Observe that, w.l.o.g., we may assume that no edges are removed in the second stage of the construction, since removing a non-essential edge e (one with weight $\mu(e)$) has no effect on μ.

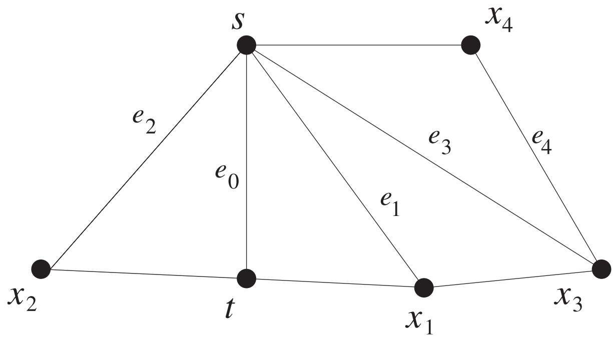

Figure 1. Ancestor and related edges.

The manner in which G was constructed implies that the metric μ^i induced by an intermediate graph G^i on $V(G^i) \subseteq V(G)$ agrees with μ restricted to these vertices, i.e., $\mu^i = \mu|_{V(G_i)}$. A closer look at the structure of G reveals more information about μ. Let us define the notions of ancestor and related edges of a vertex. The definition is recursive: the *ancestor* edges of $x \in V(G)$ include the parent edge $e = (s,t)$ of x, and the ancestor edges of s and t. The first edge e_0 is an ancestor edge of both its endpoints, and thus of all x in $V(G)$. A *related* edge of a vertex is an edge both of whose endpoints lie either on ancestor edges of x, or coincide with x. In particular, all ancestor edges of x are also related edges of x.

An example is shown in Figure 1, in which the vertices were added in the order x_1, x_2, x_3, x_4. The parent edge of x_4 is e_3, its ancestor edges are $\{e_0, e_1, e_3\}$, while $\{(t,x_1),(x_1,x_3),(s,x_4),(x_3,x_4)\}$ are its related non-ancestor edges.

Let e be an ancestor edge of x. Define $G^{x,e}$, a subgraph of G, as the union of all the related edges of x which were introduced after e, plus edge e itself. (For example, in Figure 1 the graph G^{x_4,e_1} is the subgraph induced by the vertices $\{s, x_1, x_3, x_4\}$.) The subgraph $G^{x,e}$ has a particularly simple structure: it is constructed by starting from e, marking it, and repeatedly attaching a single new vertex to the endpoints of the currently marked edge, upon which the marked edge is unmarked and one of the newly added edges is marked. The order of composition of $G^{x,e}$ is induced by that of G. The graph $G^{x,e}$ will simplify our later analysis; for the moment, observe that the distance between any pair of vertices in $G^{x,e}$ is equal to their original distance in G.

For a pair of vertices x, y, the *last common ancestor* edge $f = (s,t)$ of x, y is the common ancestor edge of x and y which was added last in the composition of G. When neither x nor y lies on an ancestor edge of the other, two possibilities may occur: either f *separates* x and y (i.e., every x-y path passes through either s or t), or there exists a vertex q whose parent edge is f, such that (s,q) is an ancestor edge of x (but not of y) while (t,q) is an ancestor edge of y (but not of x).

We are now ready to embark on the proof of the theorem.

Proof of Theorem 4.1. We start with the inductive construction of the approximating metric $\widetilde{\mu}$. The construction follows the composition procedure for G, first defining $\widetilde{\mu}$ on G^2, then extending it to G^3, G^4, etc. in turn. In the base case, G^2 is a single edge $e_0=(a,b)$, and we set $\widetilde{\mu}(a,b)=\mu(a,b)$. For the inductive step, we assume that $\widetilde{\mu}$ is already defined on $V(G^{i-1})$. Assume also that G^i is obtained from G^{i-1} by attaching a new vertex x to the endpoints of the edge (s,t). Let

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

Now, the value of $\widetilde{\mu}(x,\cdot)$, where $\cdot$ stands for any vertex of G^{i-1}, is defined as

$$\widetilde{\mu}(x,\cdot)=\delta+P_s\,\widetilde{\mu}(s,\cdot)+P_t\,\widetilde{\mu}(t,\cdot). \tag{4.3}$$

The definition of $\widetilde{\mu}$ immediately implies that it is computable in polynomial time.

The argument that $\widetilde{\mu}$ is ℓ_1-embeddable is inductive. The base case is that $\widetilde{\mu}$ on G^2 is trivially ℓ_1-embeddable. For the inductive step, observe that $\widetilde{\mu}$ on G^i is a positive linear combination of three metrics: the cut metric $\delta_{\{x\}}$ (with coefficient δ), the metric $\widetilde{\mu}$ on G^{i-1} with x at distance 0 from s (with coefficient P_s), and the metric $\widetilde{\mu}$ on G^{i-1} with x at distance 0 from t (with coefficient P_t). The cut metric is ℓ_1-embeddable; $\widetilde{\mu}$ on G^{i-1} is ℓ_1-embeddable by the induction hypothesis, and identifying the vertex x with either s or t does not affect this. Thus, by induction, the restriction of $\widetilde{\mu}$ to each G^i (and hence to $G^n=G$) is a sum of ℓ_1-embeddable metrics, and hence is ℓ_1-embeddable.

The next fact to prove is that $\widetilde{\mu}$ is dominated by μ. Since μ is the shortest path metric of G, the expansion of $\widetilde{\mu}$ is bounded by its expansion on the edges of G; thus it suffices to prove the stronger statement that every edge of G maintains its length under $\widetilde{\mu}$, i.e., for every $e=(u,v)$, $\widetilde{\mu}(u,v)=\mu(u,v)$. This stronger statement is again established by an inductive argument. The claim obviously holds for G^2. Assume that the vertex x is attached to the edge $(s,t)\in E(G^{i-1})$. By the inductive assumption, the claim holds for G^{i-1}, and in particular for (s,t). Consider, e.g., the new edge (x,s); by (4.3), $\widetilde{\mu}(x,s)=\delta+P_t\,\widetilde{\mu}(s,t)=\delta+P_t\,\mu(s,t)$ which, by definition of δ and P_t, equals $\mu(x,s)$.

Bounding the contraction of $\widetilde{\mu}$ will be the hardest part of the proof. In preparation for this, let us give an equivalent but more intuitive "backwards" description of $\widetilde{\mu}$. We envisage the process of constructing $\widetilde{\mu}$ as starting from the final vertex, and collapsing the current "last" vertex onto one of the endpoints of its parent edge. More precisely, if the edge (s,t) is the parent of x, we remove the cut metric corresponding to x (with weight δ), and then collapse the vertex x onto either s or t, with probabilities P_s and P_t respectively. (Note that P_s and P_t sum to 1, and both are non-negative by the triangle inequality.) Upon reaching G^2, we simply remove the corresponding cut metric, thus collapsing the entire graph to a single point. The metric $\widetilde{\mu}$ is just the *expected* sum of the (weighted) cut metrics removed in this process. In what follows, we shall make repeated use of this view of $\widetilde{\mu}$ as the expected result of a random process.

The bound we will prove on the contraction of $\widetilde{\mu}$ is stated in the following lemma.

Lemma 4.2. *Let x and x^* be any two vertices of G. Then, for any $\xi\in(\frac{1}{2},1)$, we have*

$$\widetilde{\mu}(x,x^*)\geq\frac{(1-\xi)(2\xi-1)}{1+\xi}\,\mu(x,x^*).$$

Theorem 4.1 follows at once from this lemma: we simply choose ξ optimally to be $\sqrt{3}-1$, and conclude that the contraction (and hence the distortion) of $\widetilde{\mu}$ is $(2-\sqrt{3})^{-2}$, which is at most 13.93.

We will split the proof of Lemma 4.2 into two cases:

Case (i): *x^* lies on an ancestor edge of x.*

Case (ii): *Neither x nor x^* lies on an ancestor edge of the other.*

Proof of Lemma 4.2, Case (i). In this case $x^*=s$ lies on an ancestor edge $e=(s,t)$ of x. Consider the graph $G^{x,e}$ as defined above, and let $\langle(s_1,t_1),\dots,(s_k,t_k)=(s,t)\rangle$ be the sequence of ancestor edges of x up to the edge on which s lies. (See Figure 2.) For convenience, set also $s_0=t_0=x$. For $1\leq i\leq k$, define

$$L_i=\mu(s_i,t_i);\quad \alpha_i=\mu(s_{i-1},s_i);\quad \beta_i=\mu(t_{i-1},t_i).$$

Note that for each $i\geq 2$, either $t_{i-1}=t_i$ with $\beta_i=0$, or $s_{i-1}=s_i$ with $\alpha_i=0$.

Denote by P_s (resp., P_t) the probability (under the random-process definition of $\widetilde{\mu}$) that, when x collapses to the edge (s,t), it collapses onto s

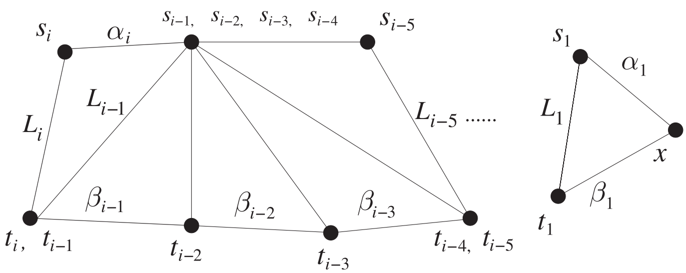

Figure 2. Proof of Lemma 4.2, Case (i).

(resp., t). Let Δ be the expected sum of the weights of the cuts removed under all collapses of x up to and including this time. Then we have $\widetilde{\mu}(x,s)=\Delta+P_t\widetilde{\mu}(s,t)$, and therefore, by the edge preservation property of $\widetilde{\mu}$,

$$\widetilde{\mu}(x,s) = \Delta + P_t\,\mu(s,t)\,. \tag{4.4}$$

Note also that not only is the actual distance $\mu(x,s)$ equal in G and in $G^{x,e}$, but the same holds for the approximated distance $\widetilde{\mu}(x,s)$: this is clear from (4.4) since the quantities Δ and P_t must be equal in G and in $G^{x,e}$. Thus in what follows we may restrict our attention to the subgraph $G^{x,e}$.

Now let P_s^i (resp., P_t^i) be the probability that, when x collapses to the edge (s_i,t_i), it collapses onto s_i (resp., t_i), and let Δ^i be the expected sum of the weights of the cuts removed under all collapses of x up to and including this time. Assume also that $t_i=t_{i-1}$ while s_i,s_{i-1} are distinct, as in Figure 2. (The other case is handled symmetrically.) The following claim establishes three inequalities relating the value of $\widetilde{\mu}(x,s_i)$ to the values of $\widetilde{\mu}(x,s_{i-1})$ and $\widetilde{\mu}(x,t_{i-1})$.

Claim 4.3. *Let $\xi\in(\frac{1}{2},1)$. Then, in the above situation,*
(a) *If $P_s^{i-1}\geq\xi$, then $\widetilde{\mu}(x,s_i)\geq\widetilde{\mu}(x,s_{i-1})+(2\xi-1)\alpha_i$.*
(b) *If $P_t^{i-1}\geq\xi$, then $\widetilde{\mu}(x,s_i)\geq\widetilde{\mu}(x,t_{i-1})+(2\xi-1)L_i$.*
(c) *Otherwise, if $1-\xi\leq P_s^{i-1}\leq\xi$, then $\widetilde{\mu}(x,s_i)+\frac{2\xi}{1-\xi}(\Delta^i-\Delta^{i-1})\geq\widetilde{\mu}(x,s_{i-1})+\alpha_i$.*

Proof. The proof is elementary but somewhat technical. Arguing as in the derivation of (4.4), we obtain

$$\begin{aligned}\widetilde{\mu}(x,s_{i-1}) &= \Delta^{i-1}+P_t^{i-1}\,L_{i-1};\\ \widetilde{\mu}(x,t_{i-1}) &= \Delta^{i-1}+P_s^{i-1}\,L_{i-1}.\end{aligned} \tag{4.5}$$

Keeping in mind the edge preservation property of $\widetilde{\mu}$, and conditioning on whether x collapsed onto s_{i-1} or t_{i-1}, we can express $\widetilde{\mu}(x,s_i)$ as

$$\widetilde{\mu}(x,s_i) = \Delta^{i-1}+P_t^{i-1}L_i+P_s^{i-1}\alpha_i. \tag{4.6}$$

Performing a formal manipulation, we get

$$\begin{aligned}\widetilde{\mu}(x,s_i) &= \Delta^{i-1} + P_t^{i-1}(L_i+\alpha_i) + (P_s^{i-1}-P_t^{i-1})\alpha_i \\ &\geq \Delta^{i-1} + P_t^{i-1}L_{i-1} + (P_s^{i-1}-P_t^{i-1})\alpha_i \\ &= \widetilde{\mu}(x,s_{i-1}) + (P_s^{i-1}-P_t^{i-1})\alpha_i,\end{aligned}$$

where we have used the triangle inequality $L_{i-1} \leq L_i + \alpha_i$, and (4.5). This implies (a).

Similarly,

$$\begin{aligned}\widetilde{\mu}(x,s_i) &= \Delta^{i-1} + P_s^{i-1}(L_i+\alpha_i) + (P_t^{i-1}-P_s^{i-1})L_i \\ &\geq \Delta^{i-1} + P_s^{i-1}L_{i-1} + (P_t^{i-1}-P_s^{i-1})L_i \\ &= \widetilde{\mu}(x,t_{i-1}) + (P_t^{i-1}-P_s^{i-1})L_i,\end{aligned}$$

implying (b).

In order to show (c), consider the change in Δ. Let δ^{i-1} be the weight of the cut removed while collapsing s_{i-1} to (s_i,t_i). Then

$$\Delta^i - \Delta^{i-1} = P_s^{i-1}\cdot\delta^{i-1} = P_s^{i-1}\cdot\frac{\alpha_i+L_{i-1}-L_i}{2}.$$

Substituting this expression for the value of $(\Delta^i-\Delta^{i-1})$, and using (4.6) and (4.5), we get

$$\begin{aligned}&\widetilde{\mu}(x,s_i) + \frac{2P_t^{i-1}}{P_s^{i-1}}(\Delta^i-\Delta^{i-1}) \\ &\quad = \left[\Delta^{i-1} + P_t^{i-1}L_i + P_s^{i-1}\alpha_i\right] + \left[P_t^{i-1}(\alpha_i+L_{i-1}-L_i)\right] \\ &\quad = \widetilde{\mu}(x,s_{i-1}) + \alpha_i\,. \qquad \blacksquare\end{aligned}$$

We are now in a position to bound $\widetilde{\mu}(x,s)$ from below in terms of $\mu(x,s)$. For this purpose, we will construct a path between x and s in $G^{x,e}$, and show that every edge on this path makes a substantial contribution to $\widetilde{\mu}(x,s)$. Since the length of the path is at least $\mu(x,s)$, this will yield the desired lower bound.

The path Π from $s=s_k$ to x in $G^{x,e}$ will be defined as follows. Assume we have already constructed some initial segment of Π, and have reached an endpoint of the edge (s_i,t_i), but have not yet reached the edge (s_{i-1},t_{i-1}). Assume also, w.l.o.g., that s_i,t_i are again situated as in Figure 2; the other case is treated in a symmetrical manner. Then we must have reached s_i. Consider the value of P_t^{i-1} defined above. If $P_t^{i-1} > \xi$, we add to Π the edge (s_i,t_{i-1}) of length L_i and continue; otherwise, we add to Π the edge (s_i,s_{i-1}) of length α_i and continue. Upon reaching (s_1,t_1), we add the edge connecting x to (s_1,t_1) to complete the path Π.

Clearly, Π is a well-defined path from $s=s_k$ to x in $G^{x,e}$. Moreover, by our choice of Π and the preceding analysis (i.e., Claim 4.3), if Π is $\langle s_k=\pi_0 \to \pi_1 \to \pi_2 \to \ldots \to \pi_m = x\rangle$, then for every edge $(\pi_{j-1},\pi_j)\in\Pi$ we have

$$\widetilde{\mu}(x,\pi_j)-\widetilde{\mu}(x,\pi_{j-1})+\frac{2\xi}{1-\xi}\left(\Delta^{\pi_j}-\Delta^{\pi_{j-1}}\right)\geq(2\xi-1)\cdot\mu(\pi_{j-1},\pi_j),$$

where, with a slight abuse of notation, Δ^{π_j} stands for Δ^r where r is the smallest index such that $\pi_j\in(s_r,t_r)$. (Observe that $(\Delta^{\pi_j}-\Delta^{\pi_{j-1}})\geq 0$, so we may safely add this term for all j.)

Summing up these expressions, we arrive at

$$\begin{aligned}\widetilde{\mu}(x,s_k)+\frac{2\xi}{1-\xi}\,\Delta^k &\geq (2\xi-1)\cdot(\text{the }\mu\text{-length of }P)\\ &\geq (2\xi-1)\,\mu(x,s_k).\end{aligned}\tag{4.7}$$

Since clearly $\widetilde{\mu}(x,s_k)\geq\Delta^k$, this completes the proof of Case (i) of Lemma 4.2. ∎

Proof of Lemma 4.2, Case (ii). In this case, neither x nor x^* lies on an ancestor edge of the other. Let (s,t) be the last common ancestor edge of x and x^*. As mentioned before, there are two possibilities. The first is that (s,t) separates x and x^*. The second is that there is a triangle $T=(s,q,t)$ such that (s,q) is an ancestor edge of x but not of x^*, (t,q) is an ancestor edge of x^* but not of x, and both (s,q) and (t,q) separate x from x^*.

We start with the analysis of the first possibility. Let P_s (resp., P_t) denote the probability that when x collapses to (s,t), it collapses onto s (resp., t); the probabilities P_s^* (resp., P_t^*) are the corresponding values for x^*. Also, let Δ (resp., Δ^*) be the expected value of the sum of the weights of cut metrics removed in the process of collapsing x (resp., x^*) to the edge (s,t). By the random process definition of $\widetilde{\mu}$, the collapses of x and of x^* proceed independently of each other; keeping in mind that $\widetilde{\mu}$ is preserved on edges, we get

$$\widetilde{\mu}(x,x^*)=\Delta+\Delta^*+(P_sP_t^*+P_tP_s^*)\,\mu(s,t).\tag{4.8}$$

Moreover, it can be easily verified that

$$P_sP_t^*+P_tP_s^*\geq\tfrac{1}{2}\min\left\{P_s+P_s^*;\;P_t+P_t^*\right\}.\tag{4.9}$$

Substituting this into (4.8), assuming w.l.o.g. that the minimum is attained at t, and using (4.4), we get

$$\widetilde{\mu}(x,x^*) \geq \tfrac{1}{2}\,(\,\widetilde{\mu}(x,s) + \varDelta) + \tfrac{1}{2}\,(\,\widetilde{\mu}(x^*,s) + \varDelta^*)\,. \tag{4.10}$$

However, adding the inequality (4.7) times the positive constant $\zeta = \frac{1-\xi}{1+\xi}$ to the inequality $\widetilde{\mu}(x,s) - \varDelta \geq 0$ times the positive constant $(\frac{1}{2} - \zeta)$, gives

$$\tfrac{1}{2}\,(\,\widetilde{\mu}(x,s) + \varDelta) \geq \frac{(1-\xi)(2\xi-1)}{1+\xi}\,\mu(x,s)\,.$$

An analogous bound holds for $\widetilde{\mu}(x^*,s)$. These two bounds, together with (4.10) and the triangle inequality $\mu(x,x^*) \leq \mu(x,s) + \mu(s,x^*)$, imply the Lemma when the first possibility occurs.

We now look at the second possibility, i.e., when there is the triangle $T = (s,t,q)$. To compute the values of $\mu(x^*,x)$ and $\widetilde{\mu}(x^*,x)$ in the original graph G, it suffices to look instead at the random process restricted to the graph H obtained by taking the graphs $G^{x,(s,q)}$ and $G^{x^*,(t,q)}$ and attaching them to the triangle $T = (s,q,t)$. (This follows by the same reasoning as in Case (i), when we argued that the values of $\mu(x,s)$ and $\widetilde{\mu}(x,s)$ in G could be computed by restricting our attention to $G^{x,e}$.)

The random process goes as follows: the graph H is first collapsed onto T, the vertex q is then collapsed onto either s or t, and finally the resulting $\{s,t\}$-cut is removed. Let us define a new random process, which collapses H onto T as before, but then collapses t onto (s,q) and removes the resulting $\{s,q\}$-cut. Our claim is that the value of $\widetilde{\mu}(x,x^*)$ is the same in both processes. Indeed, the two processes differ only in the final step, and it is simple to check that, given a triangle, the random process generates the same metric regardless of which vertex is collapsed onto its opposite edge.

Now, in this new order that we have introduced, the last common ancestor edge of x, x^* is (s,q), and this edge separates x and x^*. At this point, the argument for the first possibility applies, and the claim follows. ∎

This completes the verification of both cases in the proof of Lemma 4.2, and hence the proof of Theorem 4.1. ∎

Having proved the main theorem of this section, let us state some corollaries and observations.

Much of the complication in the proof arises from the need to account for both the cuts removed and the collapses made at each step. Let us consider for the moment the important special situation in which no cuts are removed, i.e., when the input series-parallel graph G has the property that for all x, for all ancestor edges (s,t) of x we have $\mu(x,s) + \mu(x,t) = \mu(s,t)$. (Observe that this property can be restated in a simpler form: for all x, we have $\mu(x,a)+\mu(x,b)=\mu(a,b)$, where a, b are the terminals of G. We shall point out an interesting application of these graphs in Section 5.4.)

For such graphs a stronger version of Lemma 4.2 is true: namely, $\widetilde{\mu}(x,x^*)\geq\frac{1}{2}\mu(x,x^*)$. Moreover, the proof is much simpler than in the general setting. To see this, consider first Case (i) (when $x^* = s$ lies on an ancestor edge of x); in this case we actually have that $\widetilde{\mu}(x,s)=\mu(x,s)$, and this follows directly from the definition of $\widetilde{\mu}$ using induction on the composition of G. Indeed, assume that x is attached to (s_1,t_1), and the claim has already been established for s_1,t_1. By definition of $\widetilde{\mu}$,

$$\widetilde{\mu}(x,s)=\frac{\mu(x,s_1)}{\mu(s_1,t_1)}\cdot\widetilde{\mu}(t_1,s)+\frac{\mu(x,t_1)}{\mu(s_1,t_1)}\cdot\widetilde{\mu}(s_1,s)\,.$$

By the inductive hypothesis,

$$\widetilde{\mu}(t_1,s)=\mu(t_1,s)=\mu(t_1,s_1)+\mu(s_1,s)\,;\qquad \widetilde{\mu}(s_1,s)=\mu(s_1,s)\,.$$

Combining the equations, we get $\widetilde{\mu}(x,s)=\mu(x,s)$ as claimed. Case (ii) of Lemma 4.2 can now be strengthened to $\widetilde{\mu}(x,x^*)\geq\frac{1}{2}\mu(x,x^*)$. This follows from (4.10), keeping in mind that $\Delta=\Delta^*=0$ and using the stronger version of Case (i) given above. Thus, we can conclude:

Lemma 4.4. *For the special series-parallel graphs described above, $\frac{1}{2}\mu\leq\widetilde{\mu}\leq\mu$.*

Returning now to the gap γ in multicommodity flow instances, Theorems 3.2 and 4.1 imply:

Corollary 4.5. *Let $G=(V,E)$ be a graph with no K_4-minor. Then, for every assignment of edge capacities C and demands D in G, the gap $\gamma=$ MinCut/MaxFlow is less than 14.*

With the aid of a little graph-theoretic machinery, this corollary can be generalized as follows. The proof is somewhat orthogonal to our main development, and is omitted (see [29]).

Theorem 4.6. *Let $G=(V,E)$ be a graph, and let the set of demand pairs be a subset of pairs from U, for some $U\subseteq V$. If G contains no K_4-minor w.r.t. U, then for every assignment of edge capacities C and demands D in G, the gap $\gamma=$ MinCut/MaxFlow is less than 28.*

4.1.1. Approximating the sparsest cut in series-parallel graphs The iterative procedure used in the above proof can be exploited to find a near-optimal sparsest cut in series-parallel graphs in polynomial time. Previously, this result was known only for the special case of uniform demands [33, 31, 23]. Observe that Corollary 4.5 alone does not immediately imply the existence of a polynomial time procedure for finding a good cut.

Theorem 4.7. *There is a polynomial time 14-approximation algorithm for the Sparsest Cut problem on series-parallel graphs.*

Proof Sketch. To approximate the MinCut in a series-parallel graph, we first solve the corresponding multicommodity flow problem, and find the metric μ minimizing $\frac{C\cdot\mu}{D\cdot\mu}$ (see the discussion following Theorem 3.2). By Theorem 4.1, we can find in polynomial time an ℓ_1-metric $\widetilde{\mu}$ that 14-approximates μ. Recall the manner in which $\widetilde{\mu}$ is built (see equation (4.3) and the description following it): at each step, it is a positive linear combination of three ℓ_1-metrics $\widetilde{\mu}_1, \widetilde{\mu}_2$ and $\widetilde{\mu}_3$. Consequently, at least one of these metrics must yield a value $\frac{C\cdot\widetilde{\mu}_i}{D\cdot\widetilde{\mu}_i}$ which is at most $\frac{C\cdot\widetilde{\mu}}{D\cdot\widetilde{\mu}}$. Choosing this minimizing metric and continuing with the corresponding subgraph, we will eventually reach a point where the remaining metric is a cut metric. This cut achieves the desired approximation ratio. ∎

4.2. Embedding graphs with few edges

Recall that for a graph $G=(V,E)$, the *Euler characteristic* $\chi(G)$ is defined as $|E|-|V|+1$. It is easy to see that, for each $c\in\mathbb{Z}^+$, the family of graphs $\mathcal{F}_c=\{G\,|\,\chi(G)\le c\}$ is minor-closed. The following theorem shows that graphs with low $\chi(G)$ can be embedded with low distortion into ℓ_1:

Theorem 4.8. *A metric supported on an arbitrary graph G can be embedded into ℓ_1 with distortion $O(\log\chi(G))$, where $\chi(G)$ is the Euler characteristic of G.*

Proof. The embedding will be similar in flavor to that of Theorem 4.1, though much simpler. As before, we assume that G is 2-connected; if not, we can apply the argument to each of its blocks. We also assume that G is not a cycle, since the cycle metric embeds isometrically into ℓ_1, as can be deduced from Proposition 3.1 (or for a direct proof see [27, Prop. 5.10]).

Define an *isolated path* to be a maximal path in G, each of whose internal vertices has degree 2. Hence each of its endpoints has degree at least 3. Call an isolated path B *tight* if its length is equal to the distance between its endpoints. We first decompose d_G, the shortest-path metric of G, into two simpler metrics: $\widetilde{\mu}$, which is the shortest-path metric of a graph G' with the same vertices and edges as G but which has only tight isolated paths, and $\widetilde{\mu}'$, which is a sum of cut metrics.

For this, let us consider a weighted cycle C, assuming that the weight of any edge is just its shortest-path length. Let $e=(u,v)$ be an edge on C. Since C is ℓ_1-embeddable, the metric d_C can be written as a positive linear combination of cut metrics. Let d_0 be the sum of all those cuts that separate u and v, and d_1 be the sum over the remaining cuts; clearly, $d_C=d_0+d_1$. Observe that the sum of d_0-lengths of all the edges in $E(C)-\{e\}$ is necessarily exactly equal to the length of e, or, in other words, the length of the path $P=C-\{e\}$ under d_0 is equal to the length of e; note also that $d_0(e)=d_C(e)$. Concerning d_1, observe that no cut metric δ_S in d_1 separates u and v, so we may assume w.l.o.g. that the corresponding set S satisfies $S\subseteq V(C)-\{u,v\}$.

All this leads to a decomposition of G into G' plus an ℓ_1 metric. Suppose G has isolated paths that are not tight. To the endpoints u and v of each isolated path B, add an edge $e=(u,v)$ of length $d(u,v)$; this forms a cycle with B. The shortest path metric of each such cycle can be decomposed into d_0 and d_1 as above. Each of the cut metrics in d_1 naturally extends to the whole of G, and hence d_1, being their weighted sum, also extends to an ℓ_1-embeddable metric on G. Call this $\widetilde{\mu}'$. By the preceding discussion, $d_G=d_{G'}+\widetilde{\mu}'$, where G' has the same vertices and edges as G, but all isolated paths in G' are now tight (as in d_0). This is the desired decomposition.

Since this phase involved no distortion, it suffices for the proof of the theorem to show that any graph G with tight isolated paths can be embedded into ℓ_1 with distortion $O(\log\chi(G))$. We will denote the length of an isolated path B by $d(B)$.

Let $\widetilde{G}$ be a minor (multigraph) of G obtained by the following random procedure: for each isolated path B with endpoints u_B and v_B, choose a value r_B uniformly and independently from the interval $[0,d(B)]$, and collapse all vertices in B at distance less than r_B from v_B to this endpoint, and all the other vertices in B to u_B. The length of the newly created edge $(u_B,v_B)\in E(\widetilde{G})$ is defined as $d(B)=d_G(u_B,v_B)$, so that the distance between u_B and v_B remains unchanged. Clearly, the minimum degree of $\widetilde{G}$ is now at least 3. Define $\widetilde{\mu}(\cdot,\cdot)=\mathrm{E}\left[d_{\widetilde{G}}(\cdot,\cdot)\right]$; being a convex combination of metrics, $\widetilde{\mu}$ is a metric as well. We claim that $\widetilde{\mu}$ closely approximates d:

Claim 4.9. *For any two vertices x,y of G, the expected distance $\widetilde{\mu}$ between x and y in $\widetilde{G}$ satisfies*

$$\frac{1}{4}\,d(x,y)\le\widetilde{\mu}(x,y)\le d(x,y).$$

Proof. Let us start with two simple observations. Firstly, if neither x nor y is an internal vertex of an isolated path, the distance between them remains the same, i.e., $\widetilde{\mu}(x,y)=d(x,y)$. Furthermore, a simple calculation (involving the probability that x and y are collapsed to different endpoints of B) shows that the same is true for any x and y belonging to the *same* isolated path B. Thus $\widetilde{\mu}$ preserves the lengths of all the edges of G, and since d is the shortest-path distance in G, we infer that $\widetilde{\mu}$ is dominated by d.

Consider now the case when the vertices x,y lie on different isolated paths B and B'. Let s,t be the endpoints of B, and q,r the endpoints of B'. Define P_s and P_t to be the probabilities that x is contracted to s and t respectively. P_q and P_r are defined similarly, with respect to y. Clearly,

$$P_s = \frac{d(x,t)}{d(s,t)}\,; \qquad \text{and} \qquad P_t = \frac{d(x,s)}{d(s,t)}\,.$$

The expressions for P_q and P_r are analogous. By the definition of $\widetilde{\mu}$,

$$\begin{aligned} \widetilde{\mu}(x,y) &= P_sP_q \cdot d(s,q) + P_sP_r \cdot d(s,r) + P_tP_q \cdot d(t,q) + P_tP_r \cdot d(t,r) \\ &= P_s \cdot [P_q d(s,q) + P_r d(s,r)] + P_t \cdot [P_q d(t,q) + P_r d(t,r)]\,. \end{aligned} \tag{4.11}$$

A scaled version of (4.9) together with the triangle inequality implies that

$$\begin{aligned} &P_q d(s,q) + P_r d(s,r) \\ &\quad \geq \tfrac{1}{2} \min\left\{P_q\left[d(s,q)+d(s,r)\right] + d(s,r)\,;\; P_r\left[d(s,q)+d(s,r)\right] + d(s,q)\right\} \\ &\quad \geq \tfrac{1}{2} \min\left\{P_q d(q,r) + d(s,r)\,;\; P_r d(q,r) + d(s,q)\right\} \\ &\quad = \tfrac{1}{2} \min\left\{d(y,r) + d(s,r)\,;\; d(y,q) + d(s,q)\right\} \\ &\quad = \tfrac{1}{2}\, d(s,y). \end{aligned}$$

Similarly, $P_q d(t,q) + P_r d(t,r) \geq \frac{1}{2} d(t,y)$. Substituting these inequalities into (4.11), and using the scaled version of (4.9) again, we conclude that

$$d(x,y) \geq \tfrac{1}{2}\left\{P_s d(s,y) + P_t d(t,y)\right\} \geq \tfrac{1}{2} \cdot \tfrac{1}{2}\, d(x,y)\,.$$

This completes the proof of the claim. ∎

Thus d is 4-approximated by $\widetilde{\mu}$. To conclude the proof of the theorem, we show that $\widetilde{\mu}$ can be embedded into ℓ_1 with small distortion. Note that $\widetilde{\mu}$ is a convex combination of semimetrics, all of which are supported on G', the graph obtained from G by replacing each isolated path by an edge. The distortion of embedding $\widetilde{\mu}$ into ℓ_1 is no more than that of $d_{G'}$, so it suffices to bound the latter.

But G' has very few vertices. On the one hand, it has minimum degree ≥ 3; on the other hand, it is a minor of G, and since taking minors cannot increase the Euler number, $\chi(G) \geq \chi(G')$. Let $n' = |V(G')|$, and $m' = |E(G')|$. By a degree argument, $m' \geq \frac{3}{2}n'$, implying $\chi(G) \geq \chi(G') \geq \frac{1}{2}n' + 1$. Consequently, G' has at most $2\chi(G) - 2$ vertices, and hence $d_{G'}$ can be embedded into ℓ_1 (e.g., using Bourgain's technique [9]) with distortion $O(\log \chi(G))$. ∎

5. Embeddings via tree metrics

The algorithms for ℓ_1-embeddings described in the previous section were based on constructing an approximating set of cut metrics. A different approach for embedding a metric (V, μ) into ℓ_1 is to specify a probability distribution over *trees* containing V, such that the *expected* tree distance between any two vertices x and y in V approximates $\mu(x,y)$ well. Since trees can be embedded isometrically into ℓ_1, this also gives an ℓ_1-embedding. Of particular interest are embeddings into distributions over *dominating* trees, in which the distance function in each tree dominates μ. Finding low-distortion embeddings of this kind has consequences for the design of many approximation algorithms (e.g., [4, 3, 20, 37, 10, 12]) and online algorithms (e.g., [4, 6]). Formally:

Definition 5.1. A metric d_G supported on a graph G is *α-probabilistically approximated* by a distribution $\mathcal{D}$ over (dominating) trees if
(1) each tree T in the distribution $\mathcal{D}$ has $V(G) \subseteq V(T)$;
(2) for all $x, y \in V$ and T in the distribution, d_T *dominates* d_G, i.e., $d_G(x,y) \leq d_T(x,y)$;
(3) for all $x, y \in V$, the expected distance $\mathrm{E}_{\mathcal{D}}[d_T(x,y)] \leq \alpha \cdot d_G(x,y)$.

In this paper we will use only spanning subtrees of G, and thus (1) and (2) will automatically be satisfied. Since the expansion is always maximal on the edges of G, condition (3) can be replaced by the more convenient

(3′) *for all* edges $e = (x,y) \in E(G)$, *the expected distance* $\mathrm{E}_{\mathcal{D}}[d_T(x,y)] \leq \alpha \cdot d_G(x,y)$.

We shall also refer to this approximation as an *embedding of d_G with distortion α into a tree distribution $\mathcal{D}$.*

Distributions over trees were first studied by Karp, who showed that distances in the unweighted cycle C_n can be $2(1 - \frac{1}{n})$-probabilistically approximated by a distribution over its subtrees [22]. The distribution is very simple: each possible spanning tree of G is output with probability $1/n$. This is in sharp contrast to the deterministic case, where it can be shown that any tree (not necessarily a subtree) approximating the cycle has $\Omega(n)$

distortion [32]. This line of enquiry was further developed in several papers [1,4,5,25,11], where distributions over arbitrary dominating trees were considered. The state-of-the-art results show that any graph with n vertices can be embedded into tree distributions with distortion $O(\log n)$ [17]. In line with our general approach, we now study the embeddability of outerplanar and series-parallel graphs into tree distributions.

5.1. Tree embeddings for outerplanar graphs

The first result of this section shows that any metric supported on a $K_{2,3}$-free graph can be embedded into a tree distribution with distortion at most 8. Of course, we already know by Proposition 3.1 that such metrics are isometrically embeddable into ℓ_1. However, that result says nothing about the stronger requirement that the embedding be a distribution over dominating trees. Both the main result of this section and the method used play an essential part in later, more difficult constructions (see, e.g., Section 5.4, and the recent [13]).

As usual, it suffices to embed only the biconnected components of the $K_{2,3}$-free graph, which are either K_4 or outerplanar. It is easy to verify that approximating any metric on n points by its minimum-weight spanning tree incurs a distortion of at most $(n-1)$, so any 4-point metric can be embedded into a tree with distortion 3. Thus, it suffices to bound the distortion for 2-connected outerplanar graphs. As always, we assume w.l.o.g. that the length of any edge is equal to the distance between its endpoints.

We start with a composition procedure for outerplanar graphs which will form the basis for the embedding. Given such a graph G, one can define a sequence of outerplanar graphs $G_0, G_1, \ldots, G_t = G$, where G_0 is a path or a cycle, and the graph G_i is obtained by attaching a path P_i either to a single vertex u_i on the outer face of G_{i-1}, or to the endpoints of an edge $e_i = (u_i, v_i)$ lying on the outer face of G_{i-1}. In the latter case, since the length of any edge is equal to the distance between its endpoints in G, the path P_i is at least as long as e_i. This implies that the shortest-path metric of the graph G_i coincides with the metric induced by d_G on $V(G_i)$. Clearly, the composition of G is completely specified by G_0 and the sequence of paths $\{P_i\}$.

Given an outerplanar graph G with a specified composition procedure, the path P_i is called *slack* if either P_i is attached to a single vertex, or P_i is attached to an edge e_i and the length of P_i is at least *twice* the length of e_i. A composition is called *slack* if all the paths P_i in it are slack. We shall first provide an embedding procedure for an outerplanar graph G assuming that G has a slack composition, and then show how to extend this to all outerplanar graphs.

Lemma 5.2. *Given an outerplanar graph G and a slack composition for it, G can be embedded into a tree distribution $\mathcal{D}$ with distortion at most 4.*

Proof. The embedding is inductive and follows the composition. At stage i, we shall construct a random spanning tree T_i of G_i from a random spanning tree T_{i-1} of G_{i-1}, while maintaining property (3′) for T_i with $\alpha = 4$; i.e., with $\mathrm{E}[d_{T_i}(x,y)] \le 4d_{G_i}(x,y)$ for all edges $(x,y) \in G_i$.

In the base case, if G_0 is a path, we do nothing. If it is a cycle, we randomly pick an edge e of G_0 with probability proportional to its length, and delete it to get a random subtree of G_0. Let the length of e be l, and the length of G_0 be L. The expected distance between the endpoints of e in T_0 is

$$\left(\frac{l}{L}\right)\cdot(L-l)+\left(\frac{L-l}{L}\right)\cdot l \le 2l, \tag{5.12}$$

satisfying property (3′).

At stage i, we look at P_i. If it is attached to a single vertex u_i, we attach it to T_{i-1} at u_i to get T_i. Clearly, property (3′) continues to hold for T_i. On the other hand, if P_i is attached to an edge e_i, we randomly pick an edge e from P_i (again with probability proportional to the length of e) and set $T_i = T_{i-1} \cup (P_i - \{e\})$. It is clear that T_i is a spanning tree of G_i. Let us show that property (3′) is maintained. By the induction hypothesis, this is true for edges (x,y) of G_{i-1}, since

$$\mathrm{E}[d_{T_i}(x,y)] = \mathrm{E}[d_{T_{i-1}}(x,y)] \le 4d_{G_{i-1}}(x,y) = 4d_{G_i}(x,y).$$

Consider an edge $e = (x,y) \in P_i$; denote its length by l, and the length of P_i by L_i. Furthermore, assuming that P_i is attached at the edge (u_i, v_i), denote $d_{G_{i-1}}(u_i, v_i)$ by d. The expected distance between x and y in T_i is at most

$$\begin{aligned}\left(\frac{l}{L_i}\right)\cdot(4d+L_i-l)+\left(\frac{L_i-l}{L_i}\right)\cdot l &= \left(\frac{l}{L_i}\right)\cdot(4d+2(L_i-l))\\ &\le l\left(4\left(\frac{d}{L_i}\right)+2\right).\end{aligned}$$

Since the composition is slack, we have $d/L_i \le 1/2$, and hence the expression above is at most $4l$, as required. ∎

While it might be the case that an outerplanar graph G does not have a slack composition, we now show that G can always be converted into a graph H which *does* have a slack composition, at the cost of a small distortion.

Lemma 5.3. *Given an outerplanar graph $G=(V,E)$, there is an outerplanar graph $H=(V,E')$ (in fact, a subgraph of G) with a slack composition such that $d_G \geq d_H \geq \frac{1}{2} d_G$.*

Proof. The graph H will be a subgraph of G, with edge lengths no longer than in G and no shorter than half those in G. Let $\langle G_0 = P_0, P_1, \ldots, P_t \rangle$ be the composition defining G. Our goal is to produce a slack composition $\langle H_0 = Q_0, Q_1, \ldots, Q_{t'} \rangle$ for H, thereby defining H in the process.

The composition sequence for H is initially set to be the same as that for G; we then consider the lowest unmarked path Q_i, and while processing and marking the path Q_i, we modify possibly both the preceding (marked) and forthcoming (unmarked) paths. We maintain the following invariants during this process: H is always a connected spanning subgraph of G; at each stage, the distances may only decrease; and finally, the edge lengths never decrease by more than a factor of 2 from their original values.

To begin, Q_0 is marked. For each $i > 0$, if the path Q_i is attached to a single vertex, we mark it and go on. Otherwise, Q_i is attached to some edge $e_i = (u_i, v_i)$ lying on some Q_k with $0 \leq k < i$. If Q_i is slack at this point, we again mark it and continue. So assume that the current length of Q_i is less than twice the current length of the edge $e_i = (u_i, v_i)$. We then do the following:

1. Modify Q_i: Decrease the lengths of all the edges in Q_i by a factor of $1 \leq \text{length}(Q_i)/\text{length}(e_i) < 2$, so that the current length of Q_i becomes exactly the current length of e_i. Remove Q_i from the sequence for H. Note that the lengths of edges in Q_i are halved in the worst case. They will never be changed again (except that the edges may possibly be removed later).

2. Modify Q_k: Recall that Q_i was attached to the ends of e_i lying on some previously marked path Q_k with $k < i$. Since now $\text{length}(e_i) = \text{length}(Q_i)$, replace e_i in Q_k by the entire rescaled path Q_i to get Q'_k. This does not change any current distances in the graph.

3. Modify Q_j, $j > i$: Observe that shrinking the path Q_i may have resulted in some edges being longer than the current distance between their endpoints in the forthcoming (but *not* the preceding) paths. To overcome this problem, consider any such edge $e \in Q_j$. If there is a path $Q_{j'}$, with $j' > j$, that is attached to the endpoints of e (and there can be only one such path), replace e in Q_j with $Q_{j'}$ and remove $Q_{j'}$ from the sequence. If there is no such $Q_{j'}$, deleting e splits Q_j into two paths, each attached to a single point, and we replace the old Q_j in the composition with these two new paths. Again, note that this does not alter any current distances. We do not mark any paths in this modification.

The main properties of the above procedure are as follows. At each time step, we have connected spanning subgraphs of G. The edges surviving upon termination were modified at most once, and their lengths were decreased at that time by at most a factor of 2. No edge-length (and hence no distance between any pair of vertices) is ever increased. The final sequence is slack. The process terminates when we have marked all the paths, i.e., in at most $|E|$ steps.

Let H be the graph specified by the resulting slack sequence. It is a connected spanning subgraph of G, with edge lengths at least half those in G. This immediately implies the lower bound $d_H \geq \frac{1}{2} d_G$. The upper bound $d_H \leq d_G$ follows from the fact that none of the steps above caused distances to increase. ∎

Now the overall procedure for embedding an outerplanar graph G is as follows. First, we obtain the graph H with a slack composition as in Lemma 5.3, incurring a distortion of at most 2. The graph H (with the edge lengths doubled in order to dominate G) is then embedded into a tree distribution with distortion at most 4 using Lemma 5.2, giving a total distortion of at most 8.

Furthermore, note that all the trees in the distribution are dominating subtrees of H with doubled edge lengths, and thus also dominating subtrees of G. For each such tree T, restoring the length of an edge $e \in T$ to $d_G(e)$ can only decrease the distortion without changing the domination property. Hence we get the main result of this section:

Theorem 5.4. *For any metric d_G supported on a $K_{2,3}$-free graph G, there is an embedding of d_G into a tree distribution $\mathcal{D}$ with distortion at most 8. Moreover, the embedding uses only subtrees of G with their original edge lengths.*

5.2. Tree embeddings for graphs with few edges

Theorem 5.5. *Any graph G with Euler characteristic $\chi(G)$ can be embedded into a dominating tree distribution with distortion $O(\log \chi(G))$.*

Proof. The proof is very similar to that of Theorem 4.8. Recall that an isolated path in G is a path with endpoints of degree ≥ 3, and all internal nodes of degree 2. For every isolated path $B = \langle v_1, v_2, \ldots, v_k \rangle$ in G, we add to G a new edge e_B between the endpoints of B, of length $d_G(v_1, v_k)$, thus leaving the original metric unaffected. Now, for each such B, independently of other isolated paths, choose an edge e in B with probability proportional to the length of e, and delete it. We get a distribution over graphs G', where each G' consists of the same "core" (including all the newly added edges), and the "hairs" (the remnants of the isolated paths).

Each G' dominates G, and the expected expansion of any edge in B introduced by the above step is at most 2 (by an analysis very similar to (5.12)), implying that the distortion incurred by this distribution over G'-metrics is at most 2.

Finally, we have to embed each G' into a dominating tree distribution. It suffices to embed the core, since each hair is already a tree and can simply be attached to the random tree approximating the core. As in the proof of Theorem 4.8, we conclude that the number of vertices in the core is $O(\chi(G))$, and hence it can be embedded into a distribution over trees with distortion $O(\log \chi(G))$ by the general result of [17]. This completes the proof. ∎

5.3. Lower bounds for series-parallel graphs

In view of the results of the previous sections, Theorems 5.4 and 5.5 may inspire hope that embeddings into tree distributions with constant distortion exist for other minor-closed families, such as series-parallel graphs. Our next result shows that this is not so; we prove a lower bound of $\Omega(\log n)$ on the distortion for embedding series-parallel graphs into dominating tree distributions. This result extends those of Alon *et al.* [1] and Konjevod *et al.* [25], who gave a technically more involved lower bound for the n-vertex grid, and shows that approximating graph metrics by distributions over tree metrics already breaks down for families of graphs that are much simpler than grids.

Theorem 5.6. *There exists an infinite family of series-parallel graphs $\{G_k\}$ such that any α-approximation of the shortest-path metric of G_k by a distribution over dominating trees has $\alpha = \Omega(\log |V(G_k)|)$.*

The proof makes use of the following fact from [32]:

Theorem 5.7 ([32]). *The distortion of any embedding of the unit-weighted cycle C_n into an (arbitrary) tree is at least $n/3-1$.*

Proof of Theorem 5.6. The graphs G_k are defined recursively. G_0 is a single unit-weighted edge between terminals s_0 and t_0. Inductively, H_{i+1} consists of two copies of G_i in series, and G_{i+1} consist of two copies of H_{i+1} in parallel between terminals s_{i+1} and t_{i+1} (see Figure 3). The graph G_k has $n = 4^k$ edges and $\Theta(n)$ vertices. Observe that for any G_i with terminals s_i

and t_i, both the distance between the terminals and the size of a minimum s_i-t_i cut are 2^i.

Following a standard framework for establishing lower bounds for probabilistic constructions (see, e.g., [38,1,25]), it suffices to come up with a distribution $\mathcal{D}$ over the *edges* of G_k, such that any tree T with $V(G_k)\subseteq V(T)$ and $d_T \geq d_{G_k}$ has a large expected expansion, i.e., $\mathrm{E}_{e\in\mathcal{D}}[d_T(u_e,v_e)] \geq \Omega(\log|V(G_k)|)$, where u_e, v_e denote the endpoints of edge e. More concretely, it suffices to show that for any tree metric $d_T \geq d_{G_k}$ on $V(G)$ we have

$$\sum_{e\in E(G_k)} d_T(u_e,v_e) = \Omega(k)\cdot \sum_{e\in E(G_k)} d_{G_k}(u_e,v_e) = \Omega(k)\cdot 4^k,$$

since then the same must also hold for any distribution over dominating tree metrics, implying an expansion of $\Omega(k)=\Omega(\log|V(G_k)|)$.

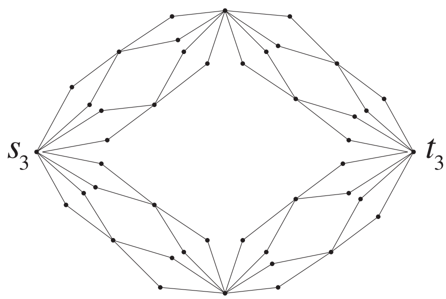

Figure 3. The graph G_3

Let T be a tree containing the vertices of G_k which dominates distances in G_k. For each $i\in[1,\dots,k]$, assign color i to all edges of G_k which suffer an expansion of at least $2^{i+1}/3-1$ in T. As a result, each edge in G_k has at least one color assigned to it, while some edges have multiple colors. Let $S_i\subset E(G_k)$ be the set of all edges that are assigned color i.

How large is S_k? Observe that any cycle which goes around the graph G_k (i.e., a simple cycle which includes the terminals s_k and t_k) has length 2^{k+1}, and therefore, by Theorem 5.7, contains an edge colored k. Thus S_k hits all such cycles, and consequently it must separate the terminals of at least one of the four copies of G_{k-1} that form G_k. Hence $|S_k|\geq 2^{k-1}$.

How large is S_{k-1}? Consider the four copies of G_{k-1} forming G_k. Arguing as before, we conclude that each of these copies must contain at least 2^{k-2} edges of color $k-1$. Hence, the size of S_{k-1} is at least $4\cdot 2^{k-2}$. Arguing in the same vein for each i, we get that $|S_i|\geq 4^{k-i}2^{i-1}=2^{2k-1-i}$.

For each $e \in E(G_k)$, let $\mathcal{C}_e$ be the set of colors assigned to e. The expansion of e is at least

$$\max_{i\in\mathcal{C}_e}\left(2^{i+1}/3-1\right) \geq \frac{1}{2}\sum_{i\in\mathcal{C}_e}\left(2^{i+1}/3-1\right).$$

Therefore,

$$\begin{aligned}\sum_{e\in E(G_k)} d_T(u_e,v_e) &\geq \frac{1}{2}\sum_{e}\sum_{i\in\mathcal{C}_e}\left(2^{i+1}/3-1\right)\\ &= \frac{1}{2}\sum_{i=1}^{k}|\{e\,|\,i\in\mathcal{C}_e\}|\cdot\left(2^{i+1}/3-1\right) = \frac{1}{2}\sum_{i=1}^{k}|S_i|\cdot\left(2^{i+1}/3-1\right)\\ &\geq \frac{1}{2}\sum_{i=1}^{k}2^{2k-i-1}\cdot\left(2^{i+1}/3-1\right) > \left(\tfrac{k}{6}-\tfrac{1}{4}\right)4^k.\end{aligned}$$

∎

Remark 5.8. After the preliminary version of this paper appeared, we were informed by Yair Bartal that Theorem 5.6 for the same family of graphs can also be inferred – albeit much less directly – from the result of Imase and Waxman [21] combined with the general framework of Bartal [4]. To see this, note that the Steiner tree problem is trivially 1-competitive on trees, and hence an α-probabilistic approximation of G_k by trees implies an α-competitive ratio on the graphs G_k [4, Theorem 4]. However, [21] establishes an $\Omega(k)$ lower bound for the competitive ratio for the Steiner problem on G_k, and hence $\alpha = \Omega(k)$.

5.4. An alternative embedding for series-parallel graphs

In light of the lower bound of the previous section, we cannot hope to embed general series-parallel graphs into tree distributions with constant distortion. However, by adding an extra ingredient (specifically, a cut-metric embedding of certain special series-parallel graphs which we call "bundles") to the tree metric technology, we will be able to come up with an alternative embedding of series-parallel graphs into ℓ_1 with constant distortion which is quite different from that of Section 4.1.

The new embedding proceeds along the same lines as the embedding of outerplanar graphs in Section 5.1. Given a series-parallel graph G, it first performs preprocessing and random edge deletion steps similar to those in Lemmas 5.3 and 5.2 to get a special tree-like series-parallel graph which we call a "tree of bundles" (i.e., a graph whose 2-connected components are bundles). This incurs a distortion of at most 8. The bundles are then embedded using the cut-metric technique with distortion 2, yielding an embedding with total distortion at most 16 for general series-parallel graphs. Although it has a marginally worse performance guarantee (at least in terms of the constant bounds we have established here), this second algorithm is conceptually simpler, and arguably more instructive than that of Theorem 4.1. Since much of the construction is similar to that for outerplanar graphs given in Section 5.1, we shall omit the recurring details and emphasize the differences.

As in Section 4.1, the construction is based on the composition procedure for G. The compositions allowed here are slightly less restrictive than before, in that we add paths of arbitrary lengths between the ends of some existing edge at each stage, rather than a single vertex (i.e., a path of length 2). Hence the composition consists of a sequence of graphs G_i, where $G_0 = P_0$ is a path, and G_i is obtained by attaching a path P_i to an already existing edge $e_i = (u_i, v_i)$. We require that the length of P_i be no less than the length of $e_i = (u_i, v_i)$, and that the lengths of all edges are equal to the actual distance between their endpoints in G. We shall further relax the composition by permitting P_i to be attached to just a single vertex; such a path will be called *free*.

Call a (non-free) path *slack* if its length L_i is at least twice d_i, the length of the edge $e_i = (u_i, v_i)$. Similarly, a path is called *taut* if $L_i = d_i$. (Note that it is possible for a path to be neither taut nor slack.) We say a composition is *slack-taut* if each (non-free) path is either slack or taut. The first observation is that we can define a preprocessing step similar to that in Lemma 5.3 for series-parallel graphs, which outputs a series-parallel graph with a slack-taut composition.

Lemma 5.9. *Given a 2-connected series-parallel graph $G = (V, E)$, there is a series-parallel graph $H = (V, E')$ with a slack-taut composition such that $d_G \geq d_H \geq \frac{1}{2} d_G$.*

The construction of H and the proof of its correctness are very similar to those of Lemma 5.3. One small difference is that whenever we reduced the length of P_i in the sequence defining an outerplanar graph, we could always remove the edge (u_i, v_i) to which P_i was attached. For series-parallel graphs, many paths can be attached to the same edge, so we cannot remove it. However, since the reduced path P_i is taut, leaving e_i in place satisfies the slack-tautness condition. Another small difference is that now we cannot remove a (forthcoming) edge which has become longer than the actual distance between its endpoints: this could contradict the technical requirement that paths must be attached to edges. To overcome this difficulty, we do not actually remove such an edge, but only mark it as "to be removed" and never touch it again until the end; then it is removed.

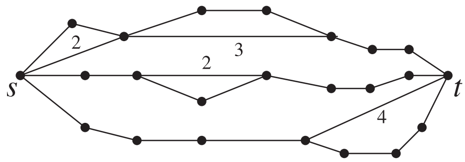

Figure 4. A bundle: all non-labeled edges have unit length.

Before stating the next lemma, let us formally define a *bundle* as a series-parallel graph such that all simple paths between its terminals are of the same length. Note that a bundle has a well-defined length, which is the distance between its terminals. Figure 4 shows an example of a bundle with terminals s and t.

Consider the slack-taut composition of H in Lemma 5.9. Observe that if P_j is a taut path attached to a preceding path P_i, and P_i is part of a bundle, then P_j also becomes a part of the same bundle. In this way we obtain the maximal bundles of the graph H. Note that if a maximal bundle B' is attached to two vertices on some other maximal bundle B (and in particular, B' cannot be considered a sub-bundle of B), then B' must be at least twice as long as the distance between its terminals. This view allows us to define another slack composition for H, in which we attach slack (maximal) bundles at each step (instead of adding slack paths).

Lemma 5.10. *Given a series-parallel graph H and a slack-taut composition for it, H can be embedded into a distribution over special subgraphs with distortion at most 4. The special subgraphs in this distribution have the property that all their maximal 2-connected components are bundles.*

The proof is similar to that of Lemma 5.2. Consider the slack composition, where a slack bundle is attached at each step. This is analogous to the slack composition for outerplanar graphs, and we shall use it in a similar way. Specifically, when adding a bundle of length L, we choose a value $r \in [0, L]$ uniformly at random and cut all the edges that cross a point at distance r from a fixed terminal of the bundle. The analysis of edge expansion is identical to that in the proof of Lemma 5.2. Since by cutting a bundle we create smaller bundles and some free paths, we obtain a "tree of bundles" at the end of the procedure.

The final step of the embedding has no outerplanar analog. Notice that bundles are precisely the special series-parallel graphs discussed in

Lemma 4.4. Thus they can be embedded into ℓ_1 with distortion at most 2 using the cut-metric technique.

Combining Lemmas 4.4, 5.9, and 5.10, we arrive at the main result of this section:

Theorem 5.11. *The procedure described in this section produces an embedding of series-parallel graphs into ℓ_1 with distortion at most 16.*

Acknowledgments

We are grateful to Gil Kalai, Alexander Karzanov, Mike Saks and David Zuckerman for insightful discussions. Many thanks to the referees for their detailed and helpful comments.

References

[1] Noga Alon, Richard M. Karp, David Peleg and Douglas B. West: A graph-theoretic game and its applications to the k-server problem, *SIAM Journal on Computing* **24(1)** (1995), 78–100.

[2] Yonatan Aumann and Yuval Rabani: An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm, *SIAM Journal on Computing* **27(1)** (1998), 291–301.

[3] Baruch Awerbuch and Yossi Azar: Buy-at-bulk network design, in *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science*, pages 542–547, 1997.

[4] Yair Bartal: Probabilistic approximations of metric spaces and its algorithmic applications, in *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science*, pages 184–193, 1996.

[5] Yair Bartal: On approximating arbitrary metrics by tree metrics, in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 161–168, 1998.

[6] Yair Bartal, Avrim Blum, Carl Burch and Andrew Tomkins: A polylog(n)-competitive algorithm for metrical task systems, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 711–719, 1997.

[7] Sandeep N. Bhatt and F. Thomas Leighton: A framework for solving VLSI graph layout problems, *Journal of Computer and System Sciences* **28(2)** (1984), 300–343.

[8] Avrim Blum, Goran Konjevod, R. Ravi and Santosh Vempala: Semi-definite relaxations for minimum bandwidth and other vertex-ordering problems, *Theoretical Computer Science* **235(1)** (2000), 25–42. (Preliminary version in *30th Annual ACM Symposium on Theory of Computing*, pages 100–105, 1998.)

[9] Jean Bourgain: On Lipschitz embeddings of finite metric spaces in Hilbert space, *Israel Journal of Mathematics* **52(1-2)** (1985), 46–52.

[10] Moses Charikar, Chandra Chekuri, Ashish Goel and Sudipto Guha: Rounding via trees: deterministic approximation algorithms for group Steiner trees and k median, in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 114–123, 1998.

[11] Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha and Serge A. Plotkin: Approximating a finite metric by a small number of tree metrics, in *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 379–388, 1998.

[12] Moses Charikar and Balaji Raghavachari: The finite capacity dial-a-ride problem, in *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 458–467, 1998.

[13] Chandra Chekuri, Anupam Gupta, Ilan Newman, Yuri Rabinovich and Alistair Sinclair: Embedding k-outerplanar graphs into ℓ_1, in *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 527–536, 2003.

[14] Victor Chepoi and Bernard Fichet: A note on circular decomposable metrics, *Geometriae Dedicata* **69(3)** (1998), 237–240.

[15] Michel Marie Deza and Monique Laurent: *Geometry of Cuts and Metrics*, Springer Verlag, Berlin, 1997.

[16] Reinhard Diestel: *Graph Theory*, Springer-Verlag, New York, 1997.

[17] Jittat Fakcharoenphol, Satish Rao and Kunal Talwar: A tight bound on approximating arbitrary metrics be tree metrics, in *Proceedings of the thirty-fifth ACM symposium on Theory of computing*, pages 448–455, 2003.

[18] Uriel Feige: Approximating the bandwidth via volume respecting embeddings, *Journal of Computer and System Sciences* **60(3)** (2000), 510–539.

[19] András Frank: Packing paths, circuits, and cuts – a survey. In Bernhard Korte, László Lovász, Hans Jürgen Prömel, and Alexander Schrijver, editors, *Paths, Flows and VLSI-Layout*, pages 47–100. Springer-Verlag, 1990.

[20] Naveen Garg, Goran Konjevod and R. Ravi: A polylogarithmic approximation algorithm for the group Steiner tree problem, *Journal of Algorithms* **37(1)** (2000), 66–84. (Preliminary version in *9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 253–259, 1998.)

[21] Makoto Imase and Bernard M. Waxman: Dynamic Steiner tree problem, *SIAM J. Discrete Math.* **4(3)** (1991), 369–384.

[22] Richard M. Karp: A $2k$-competitive algorithm for the circle. Manuscript, 1989.

[23] Philip Klein, Serge A. Plotkin and Satish B. Rao: Excluded minors, network decomposition, and multicommodity flow; in *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[24] Philip Klein, Satish Rao, Ajit Agrawal and R. Ravi: An approximate max-flow min-cut relation for undirected multicommodity flow, with applications; *Combinatorica* **15(2)** (1995), 187–202.

[25] Goran Konjevod, R. Ravi and F. Sibel Salman: On approximating planar metrics by tree metrics, *Information Processing Letters* **80(4)** (2001), 213–219.

[26] F. Thomas Leighton and Satish B. Rao: Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms, *Journal of the ACM* **46(6)** (1999), 787–832. (Preliminary version in *29th Annual Symposium on Foundations of Computer Science*, pages 422–431, 1988.)

[27] Nathan Linial, Eran London and Yuri Rabinovich: The geometry of graphs and some of its algorithmic applications, *Combinatorica* **15(2)** (1995), 215–245. (Preliminary version in *35th Annual Symposium on Foundations of Computer Science*, pages 577–591, 1994.)

[28] Ilan Newman and Yuri Rabinovich: A lower bound on the distortion of embedding planar metrics into Euclidean space, in *18th Annual ACM Symposium on Computational Geometry*, pages 94–96, 2002.

[29] Ilan Newman, Yuri Rabinovich and Michael Saks: unpublished notes.

[30] Haruko Okamura and Paul D. Seymour: Multicommodity flows in planar graphs, *Journal of Combinatorial Theory, Series B* **31(1)** (1981), 75–81.

[31] James K. Park and Cynthia A. Phillips: Finding minimum-quotient cuts in planar graphs (extended abstract), in *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 766–775, 1993.

[32] Yuri Rabinovich and Ran Raz: Lower bounds on the distortion of embedding finite metric spaces in graphs, *Discrete & Computational Geometry* **19(1)** (1998), 79–94.

[33] Satish B. Rao: Faster algorithms for finding small edge cuts in planar graphs, in *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 229–240, 1992.

[34] Satish B. Rao: Small distortion and volume preserving embeddings for planar and Euclidean metrics, in *15th Annual ACM Symposium on Computational Geometry*, pages 300–306, 1999.

[35] Alexander Schrijver: Homotopic routing methods, in Bernhard Korte, László Lovász, Hans Jürgen Prömel, and Alexander Schrijver, editors, *Paths, Flows and VLSI-Layout*, pages 329–371. Springer-Verlag, 1990.

[36] David B. Shmoys: Cut problems and their application to divide-and-conquer, in Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 192–235. PWS Publishing, 1997.

[37] Bang Ye Wu, Giuseppe Lancia, Vineet Bafna, Kun-Mao Chao, R. Ravi and Chuan Yi Tan: A polynomial time approximation scheme for minimum routing cost spanning trees, *SIAM Journal on Computing* **20(3)** (1999), 761–778. (Preliminary version in *9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 21–32, 1998.)

[38] Andrew Chi-Chih Yao: Probabilistic computations: Toward a unified measure of complexity, in *Proceedings of the 18th Annual IEEE Symposium on Foundations of Computer Science*, pages 222–227, 1977.

A. Appendix: Proof of equation (3.2)

Equation (3.2) follows from a general result concerning positive real vectors. Let $v,u\in\mathbb{R}^k$ be two positive vectors. Define

$$H(v,u) = \max_i \frac{u_i}{v_i} \cdot \max_j \frac{v_j}{u_j}\,.$$

If $S\subseteq\mathbb{R}^k$ is a closed set of positive vectors, define $H(v,S)$ as $\min_{u\in S} H(v,u)$.

Claim A.1. *If $K\subseteq\mathbb{R}^k$ is a closed convex cone, then*

$$H(v,K) = \max_{(C,D)} \frac{D\cdot v}{C\cdot v}\,, \tag{A.1}$$

where the maximum is taken over all non-negative vectors $D,C\in\mathbb{R}^k$ for which $\frac{D\cdot u}{C\cdot u}\le 1$ for any $u\in K$.

In the sequel, we use $\xi(v,K)$ to refer to the expression on the right hand side of (A.1). Before we prove Claim A.1, let us explain how it implies (3.2).

A metric (V,μ) on $|V|=n$ points can be viewed as a positive vector in $\mathbb{R}^{\binom{n}{2}}$, in which the value of the ij-th coordinate (for $i<j$) is $\mu(i,j)$. Since the set of l_1-embeddable metrics on a set V coincides with the set of non-negative combinations of cut metrics on V, they form a closed convex cone in $\mathbb{R}^{\binom{n}{2}}$, called the *cut cone* (see, e.g., [15] for more details). Denote the cut cone on V by $M_1(V)$.

Note that if v_μ is the vector corresponding to a metric (V,μ), then $H(v_\mu, M_1(V)) = c_1(\mu)$. Therefore, applying Claim A.1 to $K = M_1(V)$ and $v = v_\mu$, we obtain (3.2).

Proof of Claim A.1. One direction of the claim is easy: for any $u \in K$ and D, C as above,

$$\frac{D \cdot v}{C \cdot v} \le \max_i \frac{u_i}{v_i} \cdot \max_j \frac{v_j}{u_j} \cdot \frac{D \cdot u}{C \cdot u} \le H(v,u)\,.$$

Taking the "closest" $u \in K$ to v, we conclude that $\xi(v,K) \le H(v,K)$.

For the other direction, let $B_\delta(v) \subseteq \mathbb{R}^k$ be the set of all positive vectors $x \in \mathbb{R}^k$ such that $H(v,x) \le \delta$. Clearly,

$$B_\delta(v) = \{x \in \mathbb{R}^k \mid \forall_{r,q \in [1,\ldots,k]}\ \delta \cdot v_r x_q - v_q x_r \ge 0\}\,.$$

Observe that $B_\delta(v)$ is a closed convex cone containing v. By definition, $H(v,K)$ is the smallest δ such that $B_\delta(v) \cap K \neq \emptyset$. For this critical δ, we claim that there exists a vector $l \in \mathbb{R}^k$ such that

1. $l \cdot B_\delta(v) \ge 0$;
2. $l \cdot K \le 0$;
3. l is a non-negative combination of vectors $\Delta_{rq} \in \mathbb{R}^k$, $r,q \in [1,\ldots,k], r \neq q$, where Δ_{rq} has $-v_q$ in the r-th coordinate, δv_r in the q-th coordinate, and 0 in all other coordinates.

Indeed, the dual cone

$$B^*_\delta = \{y \in \mathbb{R}^k \mid \forall_{x \in B_\delta}\ \langle x, y\rangle \ge 0\}$$

is the convex hull of vectors $\{\Delta_{rq}\}$, and thus the normal vector to *any* supporting hyperplane of $B_\delta(v)$ separating it from K has the required properties.

Let l^+ and l^- be two non-negative vectors in $\mathbb{R}^k$ with $l^+ - l^- = l$, formed by taking the positive and the negative coordinates of l respectively. By the first two properties of l, for any $u \in K$, $\frac{l^+ \cdot u}{l^- \cdot u} \leq 1$, while $\frac{l^+ \cdot v}{l^- \cdot v} \geq 1$. In the rest of the argument, l^+ will play the role of D, while l^- will play the role of C.

Given an arbitrary form $(\sum_i d_i x_i)/(\sum_i c_i x_i)$ defined over non-negative $x \in \mathbb{R}^k$ with non-negative coefficients d_i and c_i, let us define a new form

$$\left(\frac{\sum_i d_i x_i}{\sum_i c_i x_i}\right)^{\#} = \frac{\sum_i (d_i - \min(d_i, c_i))\, x_i}{\sum_i (c_i - \min(d_i, c_i))\, x_i}\,.$$

Observe that if the value of the original form is ≥ 1, then the value of the new form exceeds that of the old one. Using this observation and the fact that $l = \sum \alpha_{rq} \Delta_{rq}$ for some non-negative α_{rq}'s, we can infer that

$$\xi(v, K) \geq \frac{l^+ \cdot v}{l^- \cdot v} = \left(\frac{\sum_{rq} \alpha_{rq} \Delta^+_{rq} \cdot v}{\sum_{rq} \alpha_{rq} \Delta^-_{rq} \cdot v}\right)^{\#} \geq \frac{\sum_{rq} \alpha_{rq} \Delta^+_{rq} \cdot v}{\sum_{rq} \alpha_{rq} \Delta^-_{rq} \cdot v} = \delta = H(v, K)\,,$$

which establishes the claim. ∎

Anupam Gupta

Department of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213-3891, USA
anupamg@cs.cmu.edu

Ilan Newman

Computer Science Department
University of Haifa
Haifa 31905, Israel
ilan@cs.haifa.ac.il

Yuri Rabinovich

Computer Science Department
University of Haifa
Haifa 31905, Israel
yuri@cs.haifa.ac.il

Alistair Sinclair

Computer Science Division, Soda Hall
University of California
Berkeley CA 94720-1776, USA
sinclair@cs.berkeley.edu