On the Geometry of Graphs with a Forbidden Minor

James R. Lee*
University of Washington
jrl@cs.washington.edu

Anastasios Sidiropoulos
University of Toronto
tasoss@cs.toronto.edu

ABSTRACT

We study the topological simplification of graphs via random embeddings, leading ultimately to a reduction of the Gupta-Newman-Rabinovich-Sinclair (GNRS) L_1 embedding conjecture to a pair of manifestly simpler conjectures. The GNRS conjecture characterizes all graphs that have an $O(1)$-approximate multi-commodity max-flow/min-cut theorem. In particular, its resolution would imply a constant factor approximation for the general Sparsest Cut problem in every family of graphs which forbids some minor. In the course of our study, we prove a number of results of independent interest.

- Every metric on a graph of pathwidth k embeds into a distribution over trees with distortion depending only on k. This is equivalent to the statement that any family of graphs excluding a fixed tree embeds into a distribution over trees with $O(1)$ distortion. For graphs of *treewidth* k, GNRS showed that this is impossible even for $k = 2$.

  In particular, our result implies that pathwidth-k metrics embed into L_1 with bounded distortion, which resolves an open question even for $k = 3$.

- We prove a generic *peeling lemma* which uses random retractions to peel simple structures like handles and apices off of graphs. This allows a number of new topological reductions. For example, if X is any metric space in which the removal of $O(1)$ points leaves a bounded genus metric, then X embeds into a distribution over planar graphs.

- Using these techniques, we show that the GNRS embedding conjecture is equivalent to two simpler conjectures: (1) The well-known planar embedding conjecture, and (2) a conjecture about embeddings of k-sums of graphs.

*Research supported by NSF CAREER award CCF-0644037.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
STOC'09, May 31–June 2, 2009, Bethesda, Maryland, USA.
Copyright 2009 ACM 978-1-60558-506-2/09/05 ...$5.00.

Categories and Subject Descriptors

F.2 [**Analysis of Algorithms and Problem Complexity**]: Miscellaneous

General Terms

Algorithms, Theory

Keywords

Embeddings, Geometry of Graphs, Forbidden Minors

1. INTRODUCTION

We view an undirected graph $G = (V, E)$ as a topological template that supports a number of different geometries. Such a geometry is specified by a non-negative length function $\mathsf{len} : E \to \mathbb{R}$ on edges, which induces a shortest-path pseudometric d_{len} on V, with

$$d_{\mathsf{len}}(u, v) = \text{length of the shortest path between } u \text{ and } v,$$

where a pseudometric might have $d_{\mathsf{len}}(u, v) = 0$ for some pairs $u, v \in V$ with $u \neq v$. From this point of view, we are interested in properties which hold simultaneously for all geometries supported on G, or even for all geometries supported on a family of graphs $\mathcal{F}$.

In the seminal works of Linial, London, and Rabinovich [19], and later Aumann and Rabani [1] and Gupta, Newman, Rabinovich, and Sinclair [11], the geometry of graphs is related to the classical study of the relationship between flows and cuts.

Multi-commodity flows and L_1 embeddings. For any metric space (X, d), we use $c_1(X, d)$ to denote the L_1 *distortion of* (X, d), i.e. the infimum over all numbers D such that X admits an embedding $f : X \to L_1$ with

$$d(x, y) \leq \|f(x) - f(y)\|_1 \leq D \cdot d(x, y)$$

for all $x, y \in X$. Here, we have $L_1 = L_1([0, 1])$, which can be replaced by the sequence space ℓ_1 when X is finite.

Corresponding to the preceding discussion, for a graph $G = (V, E)$ we write $c_1(G) = \sup c_1(V, d)$ where d ranges over all metrics supported on G, and for a family $\mathcal{F}$ of graphs, we write $c_1(\mathcal{F}) = \sup_{G \in \mathcal{F}} c_1(G)$. Thus for a family $\mathcal{F}$ of finite graphs, $c_1(\mathcal{F}) \leq D$ if and only if every geometry supported on a graph in $\mathcal{F}$ embeds into L_1 with distortion at most D.

On the other hand, one has the notion of a *multi-commodity flow instance in* G which is specified by a pair of non-negative mappings $\mathrm{cap} : E \to \mathbb{R}$ and $\mathrm{dem} : V \times V \to \mathbb{R}$. We write

$\mathsf{maxflow}(G; \mathrm{cap}, \mathrm{dem})$ for the value of the *maximum concurrent flow* in this instance, which is the maximal value ε such that $\varepsilon \cdot \mathrm{dem}(u,v)$ can be simultaneously routed between every pair $u, v \in V$ while not violating the given edge capacities.

A natural upper bound on $\mathsf{maxflow}(G; \mathrm{cap}, \mathrm{dem})$ is given by the *sparsity* of any cut $S \subseteq V$:

$$\frac{\sum_{uv \in E} \mathrm{cap}(u,v) |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{u,v \in V} \mathrm{dem}(u,v) |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}, \tag{1}$$

where $\mathbf{1}_S : V \to \{0,1\}$ is the indicator function for membership in S. In the case where $\mathrm{dem}(u,v) > 0$ for exactly one pair u, v, minimizing the upper bound (1) computes the minimum cut in G, and the max-flow/min-cut theorem states that this upper bound is achieved by the corresponding maximum flow.

In general, we write $\mathsf{gap}(G)$ for the maximum gap between the value of the flow and the upper bounds given by (1), over all multi-commodity flow instances on G. This is the *multi-commodity max-flow/min-cut gap for* G. Now we can state the fundamental relationship between the geometry of graphs and the flows they support:

THEOREM 1.1 ([19, 11]). *For every graph G, $c_1(G) = \mathsf{gap}(G)$.*

In particular, combined with the techniques of [17, 19], this implies that there exists a $c_1(G)$-approximation for the general Sparsest Cut problem on a graph G. Motivated by this connection, Gupta, Newman, Rabinovich, and Sinclair sought to *characterize* the graph families $\mathcal{F}$ such that $c_1(\mathcal{F}) < \infty$, and they posed the following conjecture.

CONJECTURE 1 ([11]). *For every family of finite graphs $\mathcal{F}$, one has $c_1(\mathcal{F}) < \infty$ if and only if $\mathcal{F}$ forbids some minor.*

We recall that H is a minor of G if it can be obtained from G via a sequence of edge contractions, edge deletions, and vertex deletions. The present paper is devoted to progress on this conjecture, through an understanding of the geometry of graphs which forbid a minor.

1.1 Simplifying the topology with random embeddings

A basic question is whether one can embed a graph metric G into a graph metric H with a simpler topology (for example, perhaps G is planar and H is a tree), where the embedding is required to have low distortion. The viability of this approach was ruled out by Rabinovich and Raz [23]. For instance, $\Omega(n)$ distortion is required to embed an n-cycle into a tree. In general (see [4]), if all metrics supported on a subdivision of some graph G can be embedded into metrics supported on a family $\mathcal{F}$, then G is a minor of some graph in $\mathcal{F}$, implying that we have not obtained a reduction in topological complexity.

On the other hand, a classical example attributed to Karp shows that random reductions might still be effective: If one removes a uniformly random edge from the n-cycle, this gives an embedding into a random tree which has distortion at most 2 "in expectation." More formally, if (X, d) is any finite metric space, and $\mathcal{Y}$ is a family of finite metric spaces, we say that (X, d) *admits a stochastic D-embedding into* $\mathcal{Y}$ if there exists a random metric space $(Y, d_Y) \in \mathcal{Y}$ and a random mapping $F : X \to Y$ such that the following two properties hold.

Non-contracting. With probability one, for every $x, y \in X$, we have $d_Y(F(x), F(y)) \geq d(x,y)$.

Low-expansion. For every $x, y \in X$,

$$\mathbb{E}\left[d_Y(F(x), F(y))\right] \leq D \cdot d(x,y).$$

For two graph families $\mathcal{F}$ and $\mathcal{G}$, we write $\mathcal{F} \rightsquigarrow \mathcal{G}$ if there exists a $D \geq 1$ such that every metric supported on $\mathcal{F}$ admits a stochastic D-embedding into the family of metrics supported on $\mathcal{G}$. We will write $\mathcal{F} \overset{D}{\rightsquigarrow} \mathcal{G}$ if we wish to emphasize the particular constant. Finally, we write $\mathcal{F} \not\rightsquigarrow \mathcal{G}$ if no such D exists. The relationship with Conjecture 1 is given by the following simple lemma.

LEMMA 1.2. *If $\mathcal{F} \overset{D}{\rightsquigarrow} \mathcal{G}$, then $c_1(\mathcal{F}) \leq D \cdot c_1(\mathcal{G})$.*

At first glance, $\rightsquigarrow$ seems like a powerful operation; indeed, in [11] it is proved that $\mathsf{OuterPlanar} \rightsquigarrow \mathsf{Trees}$, where $\mathsf{OuterPlanar}$ and Trees are the families of outerplanar graphs and acylic graphs, respectively. In general, if L is a finite list of graphs, we will write $\mathcal{E}L$ for the family of all graphs which do not have a member of L as a minor. The preceding result can be restated as $\mathcal{E}\{K_{2,3}\} \rightsquigarrow \mathcal{E}\{K_3\}$, where we use the notations K_n and $K_{m,n}$ to denote the complete and complete bipartite graphs, respectively. Unfortunately, [11] also showed that this cannot be pushed much further: $\mathcal{E}\{K_4\} \not\rightsquigarrow \mathcal{E}\{K_3\}$. Restated, this means that even graphs of treewidth 2 cannot be stochastically embedded into trees.

These lower bounds were extended in [4] to show that $\mathsf{Treewidth}(k+3) \not\rightsquigarrow \mathsf{Treewidth}(k)$ for any $k \geq 1$, where $\mathsf{Treewidth}(k)$ denotes the family of all graphs of treewidth at most k (see, e.g. [8] for a discussion of treewidth). Finally, in [5], these results are extended to any family with a weak closure property, which we describe next.

Sums of graphs. Suppose that H and G are two graphs and C_H, C_G are k-cliques in H and G respectively, for some $k \geq 1$. We define the *k-sum of H and G* as the graph $H \oplus_k G$ which results from taking the disjoint union of H and G and then identifying the two cliques C_H and C_G. We remark that the notation is somewhat ambiguous, as both the cliques and their identifications are implicit. The standard definition of a k-sum involves removing some subset of the edges after identification, but the stated notion is more appropriate in our setting. For a family of graphs $\mathcal{F}$, we write $\oplus_k \mathcal{F}$ for the closure of $\mathcal{F}$ under i-sums for every $i = 1, 2, \ldots, k$. With this notation in hand, we can state the following theorem.

THEOREM 1.3 ([5]). *If $\oplus_2 \mathcal{F} = \mathcal{F}$ and $\mathcal{G}$ is minor-closed, then $\mathcal{F} \rightsquigarrow \mathcal{G}$ implies $\mathcal{F} \subseteq \mathcal{G}$.*

This theorem implies, for example, that

$$\mathsf{Planar} \cap \mathsf{Treewidth}(k+1) \not\rightsquigarrow \mathsf{Treewidth}(k)$$

for any $k \geq 1$. We remark that planar graphs and bounded treewidth graphs are both closed under 2-sums. The mild assumptions of the preceding theorem suggest that even random embeddings are not particularly useful for reducing the topology, but in fact the main guiding principle of the present work is the following.

> **Guiding principle:** When $\oplus_2 \mathcal{F} \neq \mathcal{F}$, then $\mathcal{F}$ can be topologically simplified via stochastic embeddings.

In fact, this is suggested by a couple of existing reductions. In [6], it is proved that $\mathsf{Outerplanar}(k) \rightsquigarrow \mathsf{Trees}$, where $\mathsf{Outerplanar}(k)$ is the family of k-outerplanar graphs. Perhaps more surprisingly, it is shown in [13] that $\mathsf{Genus}(g) \rightsquigarrow \mathsf{Planar}$, where $\mathsf{Genus}(g)$ is the family of graphs embedded on an orientable surface of genus g, and $\mathsf{Planar} = \mathsf{Genus}(0)$ is the family of planar graphs. Note that while trees and planar graphs are closed under 2-sums, neither $\mathsf{Outerplanar}(k)$ nor $\mathsf{Genus}(g)$ are for $k \geq 1$ and $g \geq 1$. In some sense, the present work involves many different instantiations of the above principle, but each seems to require a genuinely distinct approach.

It should be noted that an extensive amount of work has been done on embedding finite metric spaces into distributions over trees, where the distortion is allowed to depend on n, the number of points in the metric space; see, e.g. [2, 3, 9]. These results are not particularly useful for us since we desire bounds that are independent of n.

1.2 Results and techniques

We now discuss the main results of the paper, along with the techniques that go into proving them.

Pathwidth. In [11], it is proved that $c_1(\mathsf{Treewidth}(2)) < \infty$, and later works [16, 5] nailed down the precise dependence $c_1(\mathsf{Treewidth}(2)) = 2$. Extending such a bound even to $\mathsf{Treewidth}(3)$ seems quite difficult, and is a well-known open problem. In fact, perhaps the simplest "width 3" problem involves the familiy $\mathsf{Pathwidth}(3)$, where we use $\mathsf{Pathwidth}(k) \subseteq \mathsf{Treewidth}(k)$ to denote the family of graphs of pathwidth at most k. These are fundamental in the graph minor theory (see e.g. [26, 20]); see Lemma 2.7 for an inductive definition.

We prove that for every $k \in \mathbb{N}$, $\mathsf{Pathwidth}(k) \rightsquigarrow \mathsf{Trees}$. In particular, this verifies Conjecture 1 for graphs of bounded pathwidth. By the results of Robertson and Seymour [26], this result can be equivalently stated as follows.

THEOREM 1.4. *If F is any forest, then $\mathcal{E}\{F\} \rightsquigarrow \mathsf{Trees}$.*

As a consequence, we resolve Conjecture 1 whenever $\mathcal{F}$ forbids some tree. Our proof, which appears in Section 2, involves a subtle method of load balancing. From the inductive definition of pathwidth k (see Lemma 2.7), there are only $\binom{k}{2}$ "active" edges at any point in time. At every step, k edges are released from the active set, with $k-1$ of them being cut. The entire algorithm rests on choosing the one edge out of k that will be saved. This edge is chosen in a probabilistic fashion: First a random set of elligible edges are constructed based on their respective lengths, and then the maximum rank elligible edge is saved. All other elligible edges have their ranks increased. The analysis rests on showing that (1) in any single phase between rank increases, the expected stretch of an edge is bounded, and (2) the total number of rank increases for any edge is bounded. Together, these complete the analysis.

In the full version, we complement our upper bound by showing that $\mathsf{Pathwidth}(k+1) \not\rightsquigarrow \mathsf{Pathwidth}(k)$.

The peeling lemma. There are some examples where a family $\mathcal{F}$ is not closed under 2-sums because it is specified by adding a bounded number of "small" structures to another graph family. A prevelant example is the class $\mathsf{Genus}(g)$ for $g \geq 1$, where in this case the structures are handles, and the base family is that of planar graphs. Note that even $\oplus_2\mathsf{Genus}(1)$ contains graphs of arbitrarily large genus. Another example involves *apices,* where one takes a base graph $G = (V, E)$ and attaches a single new vertex to an arbitrary subset of V.

In Section 3, we introduce the *peeling lemma* to show that these substructures can be randomly "peeled" off a graph, thereby reducing its topological complexity. As an illustrative special case, consider a planar graph G with a single apex a attached to a subset of vertices $Q \subseteq V$. Further specializing, suppose that $\mathsf{len}(a, v) = 0$ for every $v \in Q$. In this case, the new metric is an *arbitrary quotient* of a planar graph metric (i.e. the result of identifying all the vertices in Q, and taking the induced metric).

The peeling lemma allows us to make random cuts in the quotient so that the resulting graph is planar, and no edge is stretched by more than an $O(1)$ factor in expectation. Its proof is based on the notion of a *random Whitney decomposition* introduced in [15]. Given a metric space (X, d), this is a way of randomly partitioning the complement $X \setminus S$ of some subset $S \subseteq X$ so that as one gets closer to S, the diameter of the partition becomes finer and finer, and yet in a precise sense the boundary of the partition is "small."

To embed the quotient G/Q into a distribution over planar graphs, we create a planar graph which is composed of k copies of G, where $Q = \{q_1, q_2, \ldots, q_k\}$, and copy i is glued to the others at point q_i. Then a random Whitney decomposition is used to choose, for every $v \in V \setminus Q$, the layer in which v will be mapped. The basis for our random decomposition is the fundamental partitioning lemma of Klein, Plotkin, and Rao [14].

Face metrics. Another application of the Peeling Lemma which is of independent interest involves the embedding of *face metrics* into distributions over trees. As discussed, we know from [11] that $\mathsf{Outerplanar} \rightsquigarrow \mathsf{Trees}$. In the full version, we combine a number of ingredients to prove a strong generalization of this theorem, whose proof is significantly more involved, and is necessary for our reduction of the L_1 embedding conjecture. We show that there exists a constant C such that if G is any metric planar graph, and $F \subseteq V(G)$ is an arbitrary face of G, then $(F, d_G) \overset{C}{\rightsquigarrow} \mathsf{Trees}$. This result is a fundamental component in the reduction we discuss next. Its proof uses a variant of the Peeling Lemma, along with the Okamura-Seymour splitting geodesic from [5].

Reducing the L_1 embedding conjecture. Finally, in the full version, we use these techniques, combined with the powerful structure theory of Robertson and Seymour (see the survey of Lovász [20]), to reduce Conjecture 1 to two manifestly simpler conjectures. The first is simply the well-known planar embedding conjecture.

CONJECTURE 2 (PLANAR EMBEDDING CONJECTURE). *If Planar is the family of planar graphs, then $c_1(\mathsf{Planar}) < \infty$.*

This conjecture was first posed in published form in [11], but has been well-known since the initial publication of [19]. The conjecture has been emphasized in treatments by Indyk [12], Linial [18], and Matoušek [21]. The second conjecture is also quite natural, and involves the notion of k-sums discussed earlier.

CONJECTURE 3 (k-SUM CONJECTURE). *For any family of graphs $\mathcal{F}$, we have $c_1(\mathcal{F}) < \infty$ if and only if $c_1(\oplus_k \mathcal{F}) < \infty$ for every $k \in \mathbb{N}$.*

As a special case, this includes the conjecture that

$$c_1(\mathsf{Treewidth}(k)) < \infty$$

for every $k \in \mathbb{N}$, since the treewidth-k geometries are precisely those supported on $\oplus_k\{K_{k+1}\}$. We recall that the conjecture is true for $k = 1$. This fact will be used throughout the paper; its proof is a straightforward exercise.

LEMMA 1.5. *For any family $\mathcal{F}$, it holds that $c_1(\mathcal{F}) = c_1(\oplus_1\mathcal{F})$.*

The k-sum conjecture is not known to hold even in the case $k = 2$. The difficulty of the conjecture is hinted at by the fact that, for $k > 1$, equality cannot hold as for $k = 1$. For instance, $c_1(K_3) = 1$, while $c_1(K_3 \oplus_2 K_3) = 4/3$ [22], and $c_1(\oplus_2\{K_3\}) = 2$ [16, 5].

1.3 Sketch of the reduction

Here, we give an outline of the reduction from Conjecture 1 to Conjectures 2 and 3. First, we need the notion of a $(g, (k, w), a)$-*almost-embeddable graph.* Roughly speaking, such a graph is constructed by beginning with a graph embedded on a compact surface (orientable or non-orientable) of genus g, adding k "fringes" of width w (see the full version for a formal definition) to k faces of the genus-g graph, and then adding a additional vertices (called *apices*) arbitrarily to the resulting graph. We let $\mathsf{AE}(g, (k, w), a)$ denote the family of all such graphs.

A fundamental theorem of Robertson and Seymour [25] (see also the survey [20]) states the following.

THEOREM 1.6 ([25]). *For every graph H, there exists a number $h \in \mathbb{N}$ such that every $G \in \mathcal{E}\{H\}$ is a subgraph of some graph in $\oplus_h\mathsf{AE}(h, (h, h), h)$.*

Our reduction is based on the next theorem.

THEOREM 1.7. *For every $g, k, w, a \in \mathbb{N}$, there exists a number $h = h(g, k, w, a) \in \mathbb{N}$ such that*

$$\mathsf{AE}(g, (k, w), a) \rightsquigarrow \oplus_1(\mathsf{Planar} \cup \mathsf{Treewidth}(h)).$$

This leads to the reduction itself.

THEOREM 1.8. *Conjecture 1 is equivalent to the conjunction of Conjectures 2 and 3.*

PROOF. One direction is easy: If Conjecture 1 holds, then certainly Conjecture 2 holds. Furthermore, if $\mathcal{F}$ is such that $c_1(\mathcal{F}) < \infty$, then $\mathcal{F}$ forbids some minor, and thus $\oplus_k\mathcal{F}$ also forbids some minor, hence Conjecture 1 implies $c_1(\oplus_k\mathcal{F}) < \infty$ as well, yielding Conjecture 3.

For the other direction, suppose that Conjectures 2 and 3 hold. In this case, $c_1(\mathsf{Treewidth}(h)) < \infty$ for every $h \in \mathbb{N}$ since $c_1(\mathsf{Treewidth}(h)) = c_1(\oplus_h\{K_{h+1}\})$, and we have assumed Conjecture 3. Also Conjecture 2 implies $c_1(\mathsf{Planar}) < \infty$, hence combining Theorem 1.7 and Lemma 1.5, we see that for every $h \in \mathbb{N}$, $c_1(\mathsf{AE}(h, (h, h), h)) < \infty$. But then Conjecture 3 implies $c_1(\oplus_h\mathsf{AE}(h, (h, h), h)) < \infty$, and by Theorem 1.6, this implies Conjecture 1. $\square$

Our proof of Theorem 1.7 proceeds along the following lines.

$$\begin{aligned}
\mathsf{AE}(g, (k, w), a) &\overset{(1)}{\rightsquigarrow} \oplus_1\mathsf{AE}(g, (k, w), 0) \\
&\overset{(2)}{\rightsquigarrow} \oplus_1\mathsf{AE}(g, (1, 2^{O(k)}w), 0) \\
&\overset{(3)}{\rightsquigarrow} \oplus_1\left(\mathsf{Genus}(g) \cup \mathsf{Treewidth}(w \cdot 2^{O(g+k)})\right) \\
&\overset{(4)}{\rightsquigarrow} \oplus_1\left(\mathsf{Planar} \cup \mathsf{Treewidth}(w \cdot 2^{O(g+k)})\right).
\end{aligned}$$

Step (1) uses Theorem 3.6; and step (4) uses Theorem 3.7. Steps (2) and (3), as well as the proof itself appear in the full version.

1.4 Preliminaries

We now review some more basic definitions and notions which appear throughout the paper.

Graphs and metrics. We deal exclusively with finite graphs $G = (V, E)$ which are free of loops and parallel edges. We will also write $V(G)$ and $E(G)$ for the vertex and edge sets of G, respectively. A *metric graph* is a graph G equipped with a non-negative length function on edges $\mathsf{len} : E \to \mathbb{R}_+$. We will denote the metric space associated with a graph G as (V, d_G), where d_G is the shortest path metric according to the edge lengths. Note that $d_G(x, y) = 0$ may occur even when $x \neq y$, and also if G is disconnected, there will be pairs $x, y \in V$ with $d_G(x, y) = \infty$. We allow both possibilities throughout the paper. An important point is that *all length functions in the paper are assumed to be reduced,* i.e. they satisfy the property that for every $e = (u, v) \in E$, $\mathsf{len}(e) = d_G(u, v)$.

Given a metric graph G, we extend the length function to paths $P \subseteq E$ by setting $\mathsf{len}(P) = \sum_{e\in P}\mathsf{len}(e)$. For a pair of vertices $a, b \in P$, we use the notation $P[a, b]$ to denote the sub-path of P from a to b. We recall that for a subset $S \subseteq V$, $G[S]$ represents the induced graph on S. For a pair of subsets $S, T \subseteq V$, we use the notations $E(S, T) = \{(u, v) \in E : u \in S, v \in T\}$ and $E(S) = E(S, S)$. For a vertex $u \in V$, we write $N(u) = \{v \in V : (u, v) \in E\}$.

Cuts and L_1 embeddings. A cut of a graph is a partition of V into $(S, \bar{S})$—we sometimes refer to a subset $S \subseteq V$ as a cut as well. A cut gives rise to a pseudometric; using indicator functions, we can write the cut pseudometric as $\rho_S(x, y) = |\mathbf{1}_S(x) - \mathbf{1}_S(y)|$. A central fact is that embeddings of finite metric spaces into L_1 are equivalent to sums of positively weighted cut metrics over that set (for a simple proof of this see [7]).

A *cut measure on* G is a function $\mu : 2^V \to \mathbb{R}_+$ for which $\mu(S) = \mu(\bar{S})$ for every $S \subseteq V$. Every cut measure gives rise to an embedding $f : V \to L_1$ for which

$$\|f(u) - f(v)\|_1 = \int |\mathbf{1}_S(u) - \mathbf{1}_S(v)|\, d\mu(S), \qquad (2)$$

where the integral is over all cuts $(S, \bar{S})$. Conversely, to every embedding $f : V \to L_1$, we can associate a cut measure μ such that (2) holds. We will use this correspondence freely in what follows.

Embeddings and distortion. If $(X, d_X), (Y, d_Y)$ are metric spaces, and $f : X \to Y$, then we write

$$\|f\|_{\mathrm{Lip}} = \sup_{x\neq y\in X} \frac{d_Y(f(x), f(y))}{d_X(x, y)}.$$

If f is injective, then the *distortion of* f is defined by $\mathsf{dist}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$. If $d_Y(f(x), f(y)) \leq d(x, y)$ for every $x, y \in X$, we say that f is *non-expansive.*

Graph minors. If H and G are two graphs, one says that H is a *minor* of G if H can be obtained from G by a sequence of zero or more of the three operations: edge deletion, vertex deletion, and edge contraction. G is said to be *H-minor-free* if H is not a minor of G. We refer to [20, 8] for a more extensive discussion of the vast graph minor theory.

Equivalently, H is a minor of G if there exists a collection of disjoint sets $\{A_v\}_{v \in V(H)}$ with $A_v \subseteq V(G)$ for each $v \in V(H)$, such that each A_v is connected in G, and there is an edge between A_u and A_v whenever $(u, v) \in E(H)$. A metric space (X, d) is said to be *H-free* if it is supported on some H-minor-free graph.

Treewidth. The notion of *treewidth* involves a representation of a graph as a tree, called a tree decomposition. More precisely, a *tree decomposition* of a graph $G = (V, E)$ is a pair (T, χ) in which $T = (I, F)$ is a tree and $\chi = \{\chi_i \mid i \in I\}$ is a family of subsets of $V(G)$ such that (1) $\bigcup_{i \in I} \chi_i = V$; (2) for each edge $e = \{u, v\} \in E$, there exists an $i \in I$ such that both u and v belong to χ_i; and (3) for all $v \in V$, the set of nodes $\{i \in I \mid v \in \chi_i\}$ forms a connected subtree of T. To distinguish between vertices of the original graph G and vertices of T in the tree decomposition, we call vertices of T *nodes* and their corresponding χ_i's *bags.* The maximum size of a bag in χ minus one is called the *width* of the tree decomposition. The *treewidth* of a graph G, is the minimum width over all possible tree decompositions of G. A tree decomposition is called a *path decomposition* if $T = (I, F)$ is a path. The *pathwidth* of a graph G is the minimum width over all possible path decompositions of G.

Asymptotic notation. For two expressions E and F, we use $E = O(F)$ to denote that there exists a constant $C > 0$ which is *independent of the variables and parameters on which E and F depend* and such that $E \leq C \cdot F$. We also use the notation $E \lesssim F$ which is synonymous with $E = O(F)$.

Finally, we consider $\mathbb{N}$ to contain 0.

2. PATHWIDTH

2.1 Warm-up: Embedding pathwidth-2 graphs into trees

In this section, we prove that $\mathsf{Pathwidth}(2) \rightsquigarrow \mathsf{Trees}$, as a warmup for the general case in Section 2.2. The pathwidth-2 case does not involve a notion of "rank" which is required for the general case, but it does show the importance of using an inflation factor to blowup small edges in order to bound the expected stretch. Let $G = (V, E)$ be a metric graph of pathwidth 2.

Definition 2.1. *A sequence of pairs $(G_0, e_0), (G_1, e_1), \ldots, (G_k, e_k)$ is a* linear width-2 composition sequence for G *if $G_k = G$, G_0 is a single edge e_0, and (G_{i+1}, e_{i+1}) arises from (G_i, e_i) as follows. We arrive at G_{i+1} by attaching a new vertex v_{i+1} to the endpoints of $e_i = (u, v)$, and $e_{i+1} \in \{(u, v), (u, v_{i+1}), (v, v_{i+1})\}$.*

Lemma 2.2. *Every graph of pathwidth 2 is a subgraph of some graph possessing a linear width-2 composition sequence.*

By Lemma 2.2, it suffices to give a probabilistic embedding for a graph G possessing a linear width-2 composition sequence $(G_0, e_0), \ldots, (G_k, e_k)$. We will inductively embed G into a distribution over its spanning trees. Let T_i be a spanning tree of G_i, with $e_i \in E(T_i)$. We will produce a random spanning tree T_{i+1} of G_{i+1} with $e_{i+1} \in E(T_{i+1})$ as follows. Let $e_i = \{u, v\}$, and let v^* be the newly attached vertex. Let $\tau = 12$ be an *inflation factor.* There are two cases.

1. If $e_i = e_{i+1}$, we delete $\{u, v^*\}$ with probability
$$\frac{\mathsf{len}(u, v^*)}{\mathsf{len}(u, v^*) + \mathsf{len}(v, v^*)},$$
and otherwise we delete $\{v^*, v\}$.

2. If $e_i \neq e_{i+1}$, assume (without loss of generality) that $e_{i+1} = \{v, v^*\}$. In that case, we delete $\{u, v^*\}$ with probability
$$\min\left\{\frac{\tau \mathsf{len}(u, v^*)}{\mathsf{len}(u, v^*) + \mathsf{len}(u, v)}, 1\right\}, \tag{3}$$
and otherwise we delete $\{u, v\}$.

It is easy to see that if T_i was a spanning tree, then so is T_{i+1}. Furthermore, by construction $e_{i+1} \in E(T_{i+1})$. Let $T = T_k$ be the final tree. It remains to bound the expected stretch in T.

Let $x, y \in V(G_j)$, and let $i \geq j$. We define $K_i^{x,y}$ to be the expected stretch incurred on the pair x, y from the step i until the end of the algorithm. Formally, we set
$$K_i^{x,y} = \frac{\mathbb{E}[d_T(x, y)]}{d_{T_i}(x, y)}.$$

Lemma 2.3. *If $v^* \in \{x, y\}$, then $K_i^{x,y} \leq 3\tau \cdot K_{i+1}^{x,y}$.*

Proof. Assume, without loss of generality, that $v^* = y$, and that x is in the subtree of $T_i \setminus e_i$ rooted at u. Since the path between x and y in T_i passes through u, we have $K_i^{x,y} \leq K_i^{u,y}$, and therefore it suffices to consider the case $x = u$. If $e_{i+1} = e_i = \{u, v\}$, then it is easy to see that the expected stretch incurred on the pair x, y at step i is at most 3, and therefore in this case $K_i^{x,y} \leq 3 \cdot K_{i+1}^{x,y}$. Similarly, if $e_{i+1} = \{v, v^*\}$, then the expected stretch is inflated by at most a factor of τ, and therefore $K_i^{x,y} \leq 3\tau \cdot K_{i+1}^{x,y}$. Finally, if $e_{i+1} = \{u, v^*\} = \{x, y\}$, then $\{x, y\} \in E(T_{i+1})$, and therefore $K_i^{x,y} = K_{i+1}^{x,y}$. $\square$

Lemma 2.4. *If $v^* \notin \{x, y\}$, then $K_i^{x,y} \leq \max\{3, K_{i+1}^{x,y}\}$.*

Proof. Let T_u and T_v be the subtrees of $T_i \setminus e_i$ rooted at u and v respectively, where $e_i = \{u, v\}$. If x and y are both either in T_u, or in T_v, then $K_i^{x,y} = 1$. So, it suffices to consider the case $x \in T_u$ and $y \in T_v$. Observe further that since the unique path between x and y in T_i passes through $\{u, v\}$, it suffices to consider the case $x = u$, $y = v$.

If $e_{i+1} = e_i$, then the edge $\{u, v\}$ remains intact (i.e. $\{u, v\} \in E(T)$), and therefore no additional stretch is incurred between u and v (i.e. $K_i^{u,v} = 1$). Assume now that $e_{i+1} \neq e_i$, which means that we are in case (2) of the algorithm. Assume further that $e_{i+1} = \{v, v^*\}$. Recall that either $\{u, v\}$ or $\{u, v^*\}$ is cut.

Let $A = \mathsf{len}(u, v^*), B = \mathsf{len}(u, v), C = \mathsf{len}(v, v^*)$. With probability $p = \min\{1, \frac{\tau A}{A+B}\}$, the edge $\{u, v^*\}$ is cut, in

which case $d_T(u,v) = d_{T_i}(u,v)$. With probability $1-p$, the edge $\{u,v\}$ is cut, and the new path between u and v is u-v^*-v, so the distance between u and v is stretched to $A + C \le 2A + B$, and is eligible to be stretched by factor $K_{i+1}^{u,v}$ in the future.

Thus, if $A \ge B/(\tau-1)$, we have $K_i^{u,v} = 1$. We can therefore assume $A < B/(\tau-1)$. Then, the expected stretch incurred from level i onwards is at most

$$\begin{aligned} K_i^{u,v} &\le \frac{\tau A}{A+B} + K_{i+1}^{u,v}\left(1 - \frac{\tau A}{A+B}\right)\frac{2A+B}{B} \\ &\le \tau\frac{A}{B} + K_{i+1}^{u,v}\left(1 - \frac{\tau A}{2B}\right)\left(1 + 2\frac{A}{B}\right) \\ &\le \tau\frac{A}{B} + K_{i+1}^{u,v}\left(1 - \frac{\tau A}{3B}\right). \end{aligned}$$

But now one sees that

$$K_i^{u,v} \le \tau\frac{A}{B}(1 - K_{i+1}^{u,v}/3) + K_{i+1}^{u,v} \le \max\{3, K_{i+1}^{u,v}\}.$$

□

LEMMA 2.5. *For any $x, y \in V$, we have $\mathbb{E}[d_T(x,y)] \le 9\tau \cdot d_G(x,y)$.*

PROOF. By Lemma 2.3, the expected stretch of the tree obtained at step i, where both x and y appear in G_i appear for the first time, is at most 3τ. By Lemma 2.4, the expected stretch of all subsequent steps is at most 3, and therefore the total stretch is at most 9τ. □

2.2 Embedding pathwidth-k graphs into trees

DEFINITION 2.6 (LINEAR COMPOSITION SEQUENCE). *Let k be a positive integer. A sequence of pairs (G_0, V_0), (G_1, V_1), …, (G_t, V_t) is a* linear width-k composition sequence *for G if $G_t = G$, G_0 is a k-clique with vertex set V_0, and (G_{i+1}, V_{i+1}) arises from (G_i, V_i) as follows: Attach a new vertex v_{i+1} to all the vertices of V_i and choose $V_{i+1} \subseteq V_i \cup \{v_{i+1}\}$ so that $|V_{i+1}| = k$ and $v_{i+1} \in V_{i+1}$.*

LEMMA 2.7. *Every graph of pathwidth k is a subgraph of some graph possessing a linear width-k composition sequence.*

Let G be a graph of pathwidth k, and let $(G_0, V_0), \ldots (G_t, V_t)$ be a linear width-k composition sequence as in lemma 2.7. Our algorithm for embedding G into a random tree proceeds inductively along the linear width-k composition sequence. For each $i \in \{0, 1, \ldots, t\}$ we compute a subgraph H_i of G_i, whose only non-trivial 2-connected component is a k-clique on V_i (see Figure 1). More specifically, H_0 is just a clique on V_0. Given H_i, we compute H_{i+1} by adding all the edges between v_{i+1} and $V_{i+1} \setminus \{v_{i+1}\}$, and removing all the edges except for one between $V_{i+1} \setminus V_i$ and V_{i+1}.

The main part of the algorithm involves determining which edge in $(V_{i+1} \setminus V_i) \times V_{i+1}$ we keep in H_{i+1}. The high-level idea behind our approach is as follows. On one hand, we want to keep short edges so that the distance between $V_i \setminus V_{i+1}$ and V_{i+1} is small. On the other hand, keeping always the shortest edge leads to accumulation of the stretch. We avoid this obstacle via a randomized process that assigns a *rank* to each edge, which intuitively means that edges of lower rank are more likely to be removed. More specifically, at each step i, we pick a random threshold L and keep the highest

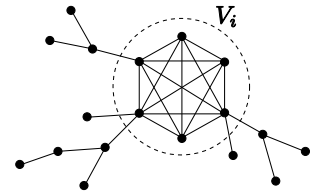

Figure 1: The graph H_i.

ranked edge of length at most L. We also update the ranks of the edges in the new graph appropriatelly.

Formally, let $\mathsf{rank}_i : V(G) \times V(G) \to \mathbb{Z}_{\ge 0}$ be an arbitrary function, with $\mathsf{rank}_0(u,v) = 0$, for each $u, v \in V(G)$. For $u, v \in V(H_i)$, let $P_i^{u,v}$ be the unique path between u and v in H_i that contains at most one edge in $E(V_i)$. Observe that $P_i^{u,v}$ is well-defined since V_i forms a clique. For an edge $e \in E(V_i)$ we set

$$\mathsf{edge\text{-}rank}_i(e) = \max_{u,v \in V(H_i): e \in P_i^{u,v}} \mathsf{rank}_i(u,v)$$

The randomized process for generating H_{i+1} and rank_{i+1} from H_i and rank_i is as follows. Let $\tau = 4k$ be an *inflation factor*.

Let w be the unique vertex in $V_i \setminus V_{i+1}$, and enumerate $E(w, V_{i+1}) = \{e_1, e_2, \ldots, e_k\}$ so that $\mathsf{len}(e_1) \le \mathsf{len}(e_2) \le \cdots \le \mathsf{len}(e_k)$.

Now, let $\{\sigma_j\}_{j=1}^{k-1}$ be a family of independent $\{0,1\}$ random variables with

$$\Pr[\sigma_j = 1] = \min\left\{1, \tau\frac{\mathsf{len}(e_j)}{\mathsf{len}(e_{j+1})}\right\}.$$

and define the set of eligible edges by

$$\mathcal{E} = \left\{e_j : \prod_{i=1}^{j-1} \sigma_i = 1\right\}.$$

In particular, $e_1 \in \mathcal{E}$ always. Let $e^* \in \mathcal{E}$ be any edge satisfying $\mathsf{edge\text{-}rank}_i(e^*) = \max_{e \in \mathcal{E}} \mathsf{edge\text{-}rank}_i(e)$.

Finally, we define H_{i+1} as the graph with vertex set $V(G_{i+1})$ and edge set (figure 2)

$$\begin{aligned} E(H_{i+1}) &= \{e^*\} \cup \{\{v_{i+1}, u\} : u \in V_{i+1}\} \\ &\quad \cup (E(H_i) \setminus E(w, V_{i+1})). \end{aligned}$$

We also define rank_{i+1} as follows. For any $u, v \in V(G)$

$$\mathsf{rank}_{i+1}(u,v) = \begin{cases} \mathsf{rank}_i(u,v) & \text{if } E(P_i^{u,v}) \cap \mathcal{E} = \emptyset \\ \mathsf{rank}_i(u,v) + 1 & \text{otherwise}. \end{cases}$$

It remains to analyze the expected stretch on H_i generated by the above process. First, we observe that the maximum rank of an edge is $O(k^2)$.

LEMMA 2.8. *For every $i = 1, 2, \ldots, t$ and every edge $e \in E(V_i)$, $\mathsf{edge\text{-}rank}_i(e) \le \binom{k}{2}$.*

PROOF. For each $i = 1, 2, \ldots, t$, and each $j = 1, 2, \ldots, \binom{k}{2}$, let $R_{i,j}$ be the j-th largest edge-rank of the edges in $E(V_i)$. That is, for each $i = 1, 2, \ldots, t$, $R_{i,1} \le R_{i,2} \le \ldots \le R_{i,\binom{k}{2}}$.

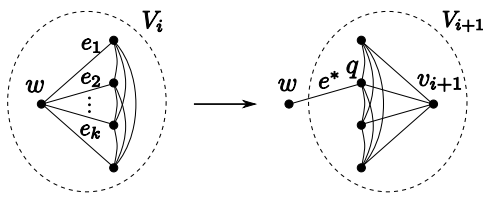

Figure 2: Transitioning from H_i to H_{i+1}.

We will prove by induction on i that for each $i = 1, 2, \ldots, t$, for each $j \in \binom{k}{2}$, $R_{i,j} \le j$. For $i = 1$, all the ranks are equal to 0, and the assertion holds trivially.

Assume now that the assertion holds for $i = l-1$, and consider the case $i = l$. Let $j \in \{1, \ldots, \binom{k}{2}\}$. If $R_{i,j} = 0$, then there is nothing to show. Otherwise, assume that $R_{i,j} > 0$. Observe that the maximum rank $R_{i-1,\binom{k}{2}}$ is removed, so $R_{i,j}$ is obtained by adding at most 1 to some rank $R_{i-1,j'}$, for some $j' \le j-1$. Therefore, $R_{i,j} \le R_{i-1,j'} + 1 \le j' + 1 \le j$. We obtain that for each $i \in \{1, \ldots, t\}$ the maximum rank is $R_{i,\binom{k}{2}} \le \binom{k}{2}$, and the lemma follows. □

For any $i \in \{0, \ldots, t\}$, $r \in \{0, \ldots, \binom{k}{2}\}$, let $K_i^{u,v}(r)$ be the expected stretch incurred from step i onward for a pair $u, v \in V(H_i)$ of rank r. That is

$$K_i^{u,v}(r) = \frac{\mathbb{E}[d_{H_t}(u,v) \,|\, \mathsf{rank}_i(u,v) = r]}{d_{H_i}(u,v)}.$$

Lemma 2.9. *Let $u, v \in V(H_i)$ with $\mathsf{rank}_i(u,v) = r < \binom{k}{2}$, such that $e_j \in E(P_j^{u,v})$. Then*

$$\begin{aligned} K_i^{u,v}(r) \;\le\; & \Pr[e_j \in \mathcal{E}] \left(1 + 2\frac{\mathbb{E}[\mathsf{len}(e^*) \,|\, e_j \in \mathcal{E}]}{\mathsf{len}(e_j)}\right) K_{i+1}^{u,v}(r+1) \\ & + \Pr[e_j \notin \mathcal{E}] \left(1 + 2\frac{\mathbb{E}[\mathsf{len}(e^*) \,|\, e_j \notin \mathcal{E}]}{\mathsf{len}(e_j)}\right) K_{i+1}^{u,v}(r) \end{aligned}$$

Proof. We have $\frac{d_{H_{i+1}}(u,v)}{d_{H_i}(u,v)} \le \frac{2\mathsf{len}(e^*)+\mathsf{len}(e_j)}{\mathsf{len}(e_j)}$. There are two possibilities: (1) $e_j \in \mathcal{E}$ occurs, and the rank of $\{u,v\}$ is increased by at least 1, (2) $e_j \notin \mathcal{E}$, and the rank of $\{u,v\}$ either increases, or remains the same. □

Lemma 2.10. *For any $j \in [k]$,*

$$\Pr[e_j \in \mathcal{E}] \left(1 + 2\frac{\mathbb{E}[\mathsf{len}(e^*) \,|\, e_j \in \mathcal{E}]}{\mathsf{len}(e_j)}\right) \le (4k)^k \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}.$$

Proof. The LHS is at most

$$\begin{aligned} & \tau^{j-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \left(1 + \frac{2}{\mathsf{len}(e_j)} \sum_{h=1}^{k} \mathsf{len}(e_h) \Pr[e^* = e_h | e_j \in \mathcal{E}]\right) \\ \le\; & \tau^{k-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \left(1 + \frac{2}{\mathsf{len}(e_j)} \sum_{h=1}^{k} \mathsf{len}(e_h) \Pr[e_h \in \mathcal{E} | e_j \in \mathcal{E}]\right) \\ \le\; & \tau^{k-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \left(1 + \frac{2}{\mathsf{len}(e_j)} \sum_{h=1}^{k} \mathsf{len}(e_h) \frac{\mathsf{len}(e_j)}{\mathsf{len}(e_h)}\right) \\ =\; & \tau^{k-1}(1+2k)\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \end{aligned}$$

□

Lemma 2.11. *For any $j \in [k]$,*

$$\Pr[e_j \notin \mathcal{E}] \left(1 + 2\frac{\mathbb{E}[\mathsf{len}(e^*) \,|\, e_j \notin \mathcal{E}]}{\mathsf{len}(e_j)}\right) \le 1 - \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}.$$

Proof. Let $I = \{h \in \{1, 2, \ldots, j-1\} : \mathsf{len}(e_{h+1}) > \tau \cdot \mathsf{len}(e_h)\}$. Observe that if $h \in \{1, 2, \ldots, j-1\} \setminus I$, then whenever $e_h \in \mathcal{E}$, we have also $e_{h+1} \in \mathcal{E}$.

$$\begin{aligned} & \Pr[e_j \notin \mathcal{E}] \left(1 + 2\frac{\mathbb{E}[\mathsf{len}(e^*) \,|\, e_j \notin \mathcal{E}]}{\mathsf{len}(e_j)}\right) \\ & \le \sum_{h=1}^{j-1} \Pr[e_h \in \mathcal{E} \text{ and } e_{h+1} \notin \mathcal{E}] \left(1 + \frac{2\mathsf{len}(e_h)}{\mathsf{len}(e_j)}\right) \\ & = \sum_{h \in I} \Pr[e_h \in \mathcal{E} \text{ and } e_{h+1} \notin \mathcal{E}] \left(1 + \frac{2\mathsf{len}(e_h)}{\mathsf{len}(e_j)}\right) \\ & \le \sum_{h \in I} \tau^{k-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_h)} \left(1 - \tau\frac{\mathsf{len}(e_h)}{\mathsf{len}(e_{h+1})}\right) \left(1 + \frac{2\mathsf{len}(e_h)}{\mathsf{len}(e_j)}\right) \\ & \le \sum_{h \in I} \tau^{k-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_h)} \left(1 - \tau\frac{\mathsf{len}(e_h)}{\mathsf{len}(e_{h+1})} + \frac{2\mathsf{len}(e_h)}{\mathsf{len}(e_j)}\right) \\ & = \sum_{h \in I} \tau^{k-1}\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_h)} \left(1 - \tau\frac{\mathsf{len}(e_h)}{\mathsf{len}(e_{h+1})}\right) + \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \sum_{h \in I} 2\tau^{k-1} \\ & \le 1 - \tau^k\frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} + \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}(2k\tau^{k-1}) \\ & = 1 + \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}\left(2k\tau^{k-1} - \tau^k\right) \\ & \le 1 - \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}. \end{aligned}$$

□

Theorem 2.12. *Every weighted graph of pathwidth k embeds into a distribution over trees with distortion $(4k)^{k^3+2}$.*

Proof. Let $u, v \in V(H_i)$. Combining lemmas 2.9, 2.10, and 2.11 we get that for i, r, $0 \le i < t$, $0 \le r < \binom{k}{2}$,

$$\begin{aligned} K_i^{u,v}(r) &\le (4k)^k \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} K_{i+1}^{u,v}(r+1) + \left(1 - \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)}\right) K_{i+1}^{u,v}(r) \\ &\le (4k)^k \frac{\mathsf{len}(e_1)}{\mathsf{len}(e_j)} \left(K_{i+1}^{u,v}(r+1) - K_{i+1}^{u,v}(r)\right) + K_{i+1}^{u,v}(r) \\ &\le (4k)^k \left(K_{i+1}^{u,v}(r+1) - K_{i+1}^{u,v}(r)\right) + K_{i+1}^{u,v}(r) \\ &\le \max\left\{(4k)^{k+1} K_{i+1}^{u,v}(r+1), K_{i+1}^{u,v}(r)\right\}. \end{aligned}$$

Therefore

$$\frac{\mathbb{E}[d_{H_t}(u,v)]}{d_G(u,v)} \le K_0^{u,v}(0) \le \left((4k)^{k+1}\right)^{\binom{k}{2}} < (4k)^{k^3+1}.$$

Now since the only non-trivial 2-connected component of H_t is V_t, it follows that by replacing the clique on V_t by a minimum spanning tree, we get a tree T such that $d_T(u,v) \le k \cdot d_{H_t}(u,v)$. The assertion follows. □

3. THE PEELING LEMMA

In this section, we prove a general theorem about embedding a graph into a distribution over simpler graphs which are derived from its subgraphs. This will allow us to "peel" small structures like apices and handles off of graphs. Roughly, if $G = (V, E)$ is a metric graph and $A \subseteq V$ is such that $d_{G[A]}(x,y) \approx d_G(x,y)$ for $x, y \in A$, then G embeds into

a distribution over a graph composed of 1-sums of $G[A]$ and $G[V \setminus A]$, provided that $d_{G[V\setminus A]}$ is "decomposable" in a sense that we now formalize.

3.1 Random partitions

First, we recall the notion of Lipschitz random partitions.

Definition 3.1 (Lipschitz random partition). *For a partition P of a set X, we write $P : X \to 2^X$ to denote the map which sends x to the set $P(x) \in P$ which contains x. A random partition P of a metric space X is Δ-bounded if*

$$\Pr\left[\forall C \in P, \operatorname{diam}(C) \le \Delta\right] = 1.$$

A Δ-bounded random partition P is β-Lipschitz if, for every $x, y \in X$,

$$\Pr[P(x) \neq P(y)] \le \beta \frac{d(x,y)}{\Delta}.$$

For a metric space X, we write β_X for the infimal β such that X admits a Δ-bounded β-Lipschitz random partition for every $\Delta > 0$, and we refer to β_X as the *decomposiability modulus of X*. The following result of Rao [24] and Klein, Plotkin, and Rao [14] will be crucial. The dependence on h is due to [10].

Theorem 3.2 (KPR decomposition). *If $G = (V, E)$ is a metric graph which excludes K_h as a minor, then $\beta_{(V,d_G)} = O(h^2)$. In particular, if some family of graphs $\mathcal{F}$ forbids a minor, then $\sup \beta_{(X,d)} < \infty$ where (X, d) runs over all metrics supported on $\mathcal{F}$.*

Following Lee and Naor [15], in the next section we will use random partitions to construct certain random retractions. Let (X, d) be a metric space with $A \subseteq X$. A mapping $R : X \to A$ is a *retraction of X onto A* if $f(x) = x$ holds for all $x \in A$. The retraction is *λ-local* if, for all $x \in X$, $d(R(x), x) \le \lambda \cdot d(x, A)$. The following theorem is from [15].

Theorem 3.3. *Let (X, d) be a finite metric space and suppose $A \subseteq X$. Then there exists a random $O(1)$-local retraction $R : X \to A$ such that for every $x, y \in X$,*

$$\mathbb{E}\left[d\left(R(x), R(y)\right)\right] \le O(\beta_X) d(x, y).$$

3.2 Random retractions

We now recall the proof of Theorem 3.3 in order to record some additional properties.

Theorem 3.4. *Let (X, d) be a metric space, and let $A \subseteq X$ be an arbitrary subset. Then there exists a random 5-local retraction $R : X \to A$ of X onto A such that, for every $x, y \in X$, we have*

$$\Pr\left[R(x) \neq R(y)\right] \le (\beta_X + 2) \frac{d(x,y)}{\min(d(x,A), d(y,A))},$$

where it is assumed that the bound is meaningless when $d(x, A) = 0$ or $d(y, A) = 0$.

Proof. Fix an arbitrary selection mapping $\Gamma : 2^X \to A$ such that for every $S \subseteq X$, $d(\Gamma(S), A) = d(S, A)$.

For each $k \in \mathbb{Z}$, let P_k be a 2^k-bounded β-Lipschitz random partition of X, and let $\alpha \in [0, 1]$ be a uniformly random value. We define a random retraction $R : X \to A$ as follows:

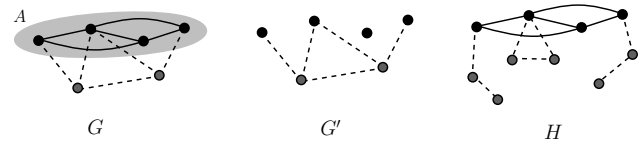

Figure 3: An example of the peeling lemma.

For every $x \in X$, put $R(x) = \Gamma\left(P_{\lceil \alpha + \log_2 d(x,A) \rceil}(x)\right)$. First, we verify that R is 5-local. For every $x \in X$, $d(x, R(x)) \le d(x, A) + \mathsf{diam}(P_{\lceil \alpha + \log_2 d(x,A) \rceil}(x)) \le 5 \cdot d(x, A)$.

Fix $x, y \in X$ and assume, without loss of generality, that $0 < d(x, A) \le d(y, A)$. Note that, by the triangle inequality, $d(y, A) \le d(x, A) + d(x, y)$. Now we have, $\Pr\left[\lceil \alpha + \log_2 d(x,A) \rceil \neq \lceil \alpha + \log_2 d(y,A) \rceil\right] \le |\log_2 d(x,A) - \log_2 d(y,A)| = \log_2 \frac{d(y,A)}{d(x,A)} \le \log_2\left(1 + \frac{d(x,y)}{d(x,A)}\right) \le 2\frac{d(x,y)}{d(x,A)}$. On the other hand, the β-Lipschitz property of P_k guarantees that $\Pr\left[P_k(x) \neq P_k(y)\right] \le \beta \frac{d(x,y)}{2^k}$. Note that if $k = \lceil \alpha + \log_2 d(x,A) \rceil = \lceil \alpha + \log_2 d(y,A) \rceil$, then $P_k(x) = P_k(y) \implies R(x) = R(y)$, thus taking a union bound over the preceding two inequalities yields $\Pr[R(x) \neq R(y)] \le (\beta + 2)\frac{d(x,y)}{d(x,A)}$, completing the proof. □

3.3 The peeling lemma

We now state and prove the peeling lemma. Given a metric graph $G = (V, E)$ and a subset $A \subseteq V$, we use

$$\mathsf{dil}_G(A) = \sup_{x,y \in A} \frac{d_{G[A]}(x,y)}{d_G(x,y)}$$

to denote the *dilation* of the induced subgraph on A in G.

Lemma 3.5 (Peeling Lemma). *Let $G = (V, E)$ be a metric graph, and $A \subseteq V$ an arbitrary subset of vertices. Let $G' = (V, E')$ be the metric graph with $E' = E \setminus E(G[A])$, and let $\beta = \beta_{(V,d_{G'})}$ be the corresponding modulus of decomposability. Then $G \overset{D}{\rightsquigarrow} H$, where $D = O(\beta \cdot \mathsf{dil}_G(A))$, and H is a 1-sum of isometric copies of the metric graphs $G[A]$ and $\left\{G[V \setminus A \cup \{a\}]\right\}_{a \in A}$. Furthermore, the embedding always has distortion at most $\mathsf{dil}_G(A)$ for pairs $x, y \in A$.*

In particular, using Lemma 1.5, this reduces the L_1 embedding problem on G to the embedding problem on $G[A]$ and $G[V \setminus A \cup \{a\}]$ for $a \in A$.

Proof Proof of Lemma 3.5. First, we specify a fixed metric graph $H = (V_H, E_H)$ into which we will randomly embed G. Enumerate the vertices $A = \{a_1, a_2, \ldots a_k\}$ and for $i \in \{1, 2, \ldots, k\}$, let H_i be a metric copy of $G[V \setminus A \cup \{a_i\}]$, such that the graphs $\{H_i\}_{i=1}^k$ are disjoint. To form H, we glue the H_i's to a disjoint metric copy F of $G[A]$, where H_i is naturally glued to F by identifying the copy of $a_i \in V(H_i)$ with the copy $a_i \in F$. See Figure 3 for an example of G, G', and H.

Now apply Theorem 3.4 to the metric space $(V, d_{G'})$ to obtain a random retraction $R : V \to A$ satisfying the consequence of the theorem with modulus β. We assume without loss of generality that $\beta \ge 1$.

Finally, we specify a random mapping $f : V \to V_H$ as follows. For every $x \in A$, $f(x) \in F$ is the canonical copy of x in $F \cong G[A]$. For every $x \in V \setminus A$, we put $f(x) = x^{(i)}$,

where $x^{(i)}$ is the copy of x in $H_i \cong G[V \setminus A \cup \{a_i\}]$, and i is such that $R(x) = a_i$.

We begin with a simple but important estimate. For any $x \in V$, we have

$$d_{G'}(x, R(x)) \leq 5 \cdot d_{G'}(x, A) = 5 \cdot d_G(x, A), \tag{4}$$

where the first inequality is because R is 5-local, and the second equality follows because the shortest path to A never requires an edge of $G[A]$. Combining this estimate with the geometry of H, we have for any $x, y \in V$,

$$\begin{aligned}
&d_H(f(x), f(y)) \\
&\leq d_{G'}(x, R(x)) + d_{G[A]}(R(x), R(y)) + d_{G'}(y, R(y)) \\
&\leq 5(d_G(x, A) + d_G(y, A)) + \mathsf{dil}_G(A) \cdot d_G(R(x), R(y)) \\
&\leq 5(d_G(x, A) + d_G(y, A)) \\
&\quad + \mathsf{dil}_G(A)\,[d_G(x, R(x)) + d_G(y, R(y)) + d_G(x, y)] \\
&\leq \mathsf{dil}_G(A)\,[d_G(x, y) + 6(d_G(x, A) + d_G(y, A))]\,,
\end{aligned} \tag{5}$$

where in the second and final lines, we used (4).

Now assume without loss of generality that $d_G(x, A) \leq d_G(y, A)$. If $d_G(y, A) \leq 2\, d_G(x, y)$, then from (5), we have $d_H(f(x), f(y)) \leq 25\mathsf{dil}_G(A) \cdot d_G(x, y)$. Thus we may assume that $d_G(y, A) > 2\, d_G(x, y)$. In particular, $d_G(y, A) \leq d_G(x, A) + d_G(x, y) \leq d_G(x, A) + d_G(y, A)/2$, which implies that

$$d_G(x, A) \leq d_G(y, A) \leq 2\, d_G(x, A). \tag{6}$$

Note that under our assumption $d_G(y, A) > 2\, d_G(x, y)$, it follows that the shortest path between x and y does not go through A, i.e. $d_{G'}(x, y) = d_G(x, y)$. This implies that if $R(x) = R(y)$, then $d_H(f(x), f(y)) = d_G(x, y)$. Also from (6) and (5), we always have

$$\begin{aligned}
d_H(f(x), f(y)) &\leq \mathsf{dil}_G(A)(d_G(x, y) + 18\, d_G(x, A)) \\
&\leq 19\, \mathsf{dil}_G(A)\, d_G(x, A).
\end{aligned}$$

Combining these estimates together yields

$$\begin{aligned}
&\mathbb{E}[d_H(f(x), f(y))] \\
&\leq d_G(x, y) + \Pr[R(x) \neq R(y)] \cdot (19\, \mathsf{dil}_G(A)\, d_G(x, A)) \\
&\leq d_G(x, y) + (\beta + 2)\frac{d_{G'}(x, y)}{d_{G'}(x, A)} \cdot (19\, \mathsf{dil}_G(A)\, d_G(x, A)) \\
&= d_G(x, y) + (\beta + 2)\frac{d_G(x, y)}{d_G(x, A)} \cdot (19\, \mathsf{dil}_G(A)\, d_G(x, A)) \\
&\leq O(\beta \cdot \mathsf{dil}_G(A))\, d_G(x, y).
\end{aligned}$$

□

Remark 3.1. *By inspecting the proof, one can observe the following slight strenghthening of the Peeling Lemma. Let $\partial A = \{v \in A : \exists (u, v) \in E \text{ with } u \notin A\}$ be the (inner) vertex boundary of A. Then one can take H to be a 1-sum of isometric copies of the metric graphs $G[A]$ and $\left\{G[V \setminus A \cup \{a\}]\right\}_{a \in \partial A}$, because one can enforce that the random retraction of Theorem 3.4 always maps points $x \notin A$ to a point of ∂A, by proper choice of the selection mapping Γ.*

3.4 Some applications

In this section, we give some relatively simple applications of the Peeling Lemma. In the full version, we present some more elaborate applications.

Given any family of graphs $\mathcal{F}$, we let $\mathsf{Apex}(\mathcal{F})$ denote the family of graphs which results from taking graphs $G \in \mathcal{F}$ and attaching a new vertex, called an *apex,* arbitrarily to the old vertices of G.

Theorem 3.6 (Removing apices). *For any family of graphs $\mathcal{F}$ that forbids some minor,*

$$\mathsf{Apex}(\mathcal{F}) \rightsquigarrow \oplus_1(\mathcal{F} \cup \{K_2\}).$$

Proof. Consider any $G \in \mathcal{F}$ equipped with some metric, and consider a metric graph $\hat{G} = (V \cup \{a\}, E \cup \hat{E})$ where $\hat{E} \subseteq \{(a, v) : v \in V\}$ with $\mathsf{len}_{\hat{G}}|_E = \mathsf{len}_G$. All graphs of $\mathsf{Apex}(\mathcal{F})$ are clearly of the form $\hat{G}$ for some $G \in \mathcal{F}$.

Now let $\widetilde{G}$ be the metric graph which results from $\hat{G}$ by placing a new vertex in the middle of every edge $e \in \hat{E}$, and let $A \subseteq V(\widetilde{G})$ be the set of all these new subdivision vertices, together with the apex a. By construction, it's clear that $d_{\widetilde{G}[A]}(x, y) = d_{\widetilde{G}}(x, y)$ for all $x, y \in A$. Also, since $\mathcal{F}$ forbids some minor, so does $\mathsf{Apex}(\mathcal{F})$, and thus by Theorem 3.2, there exists a constant C such that for every $G \in \mathcal{F}$, for every subgraph G' of $\widetilde{G}$, $\beta_{(V(G'), d_{G'})} \leq C$. Thus we can apply the Peeling Lemma to the pair $\widetilde{G}$ and A, completing the proof. □

Next, we recall that the family $\mathsf{Genus}(g)$ is the class of all graphs which can be embedded on a compact, orientable surface of genus $g \geq 0$. We let $\mathsf{noGenus}(g)$ be the set of all graphs which can be embedded on a compact, *non-orientable* surface of genus $g \geq 0$. The orientable case of the following theorem was proved in [13]. The Peeling Lemma provides an easy proof for the general case.

Theorem 3.7 (Removing handles). *For every $g \geq 0$, $\mathsf{Genus}(g) \cup \mathsf{noGenus}(g) \rightsquigarrow \mathsf{Planar}$.*

Proof. We begin with the following.

Lemma 3.8. *Let G be a metric graph embedded into an either orientable, or non-orientable surface $\mathbb{S}$ of positive genus. Let C be the shortest non-contractible cycle of G. Then, for any $u, v \in V(C)$, $d_C(u, v) = d_G(u, v)$.*

We ommit the proof from this extended abstract. The following lemma is folklore.

Lemma 3.9. *Let G be a graph embedded in either an orientable, or non-orientable surface of genus $g > 0$, and let C be a non-contractible cycle of G. Then, removing C from G leaves a graph with every connected component having orientable (resp., non-orientable) genus at most $g - 1$.*

To finish the proof, we can thus take a metric graph $G = (V, E)$ of orientable or non-orientable genus $g > 0$, and a shortest non-contractible cycle C in G. Assume, without loss of generality, that the minimum distance in G is 1. For every edge $e = \{u, v\}$ having exactly one endpoint, say $u \in C$, we replace e by a path u-w-v, where the edges $\{u, w\}$ and $\{w, v\}$ both have length $\mathsf{len}_G(u, v)/2$. Let S be the set of all these new vertices, and let $G' = (V', E')$ be the resulting metric graph. Note that $d_{G'}(x, y) = d_G(x, y)$ for all $x, y \in V$, therefore C is still geodesic in G'.

Let $A = C \cup S$, and observe that $\mathsf{dil}_{G'}(A) \leq 3$. Applying Lemma 3.9, each connected component of the subgraph $G[V \setminus C]$ has orientable (resp., non-orientable) genus at most

$g-1$, and thus the same is true for each connected component of $G'[V' \setminus A \cup \{a\}]$ for every $a \in A$. So applying the Peeling Lemma to G' and A (and using Theorem 3.2 to build β) yields $\mathsf{Genus}(g) \rightsquigarrow \oplus_1\mathsf{Genus}(g-1)$ and $\mathsf{noGenus}(g) \rightsquigarrow \oplus_1\mathsf{noGenus}(g-1)$. Repeating this g times yields the claimed result, using the fact that $\oplus_1\mathsf{Planar} = \mathsf{Planar}$. $\square$

4. REFERENCES

[1] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

[2] Y. Bartal. Probabilistic approximations of metric space and its algorithmic application. In *37th Annual Symposium on Foundations of Computer Science*, pages 183–193, Oct. 1996.

[3] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *30th Annual ACM Symposium on Theory of Computing*, pages 183–193, 1998.

[4] D. Carroll and A. Goel. Lower bounds for embedding into distributions over excluded minor graph families. In *Proceedings of the 12th European Symposium on Algorithms*, 2004.

[5] A. Chakrabarti, A. Jaffe, J. R. Lee, and J. Vincent. Embeddings of topological graphs: Lossy invariants, linearization, and 2-sums. In *IEEE Symposium on Foundations of Computer Science*, 2008.

[6] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Embedding k-outerplanar graphs into l_1. *SIAM J. Discrete Math.*, 20(1):119–136 (electronic), 2006.

[7] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[8] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.

[9] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 448–455, 2003.

[10] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *Proceedings of 6th Workshop on Approximation, Randomization, and Combinatorial Optimization*, volume 2764 of *Lecture Notes in Computer Science*, pages 36–46. Springer, 2003.

[11] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and l_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[12] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE Computer Society, 2001.

[13] P. Indyk and A. Sidiropoulos. Probabilistic embeddings of bounded genus graphs into planar graphs. In *Proceedings of the 23rd Annual Symposium on Computational Geometry*. ACM, 2007.

[14] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[15] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[16] J. R. Lee and P. Rhagevendra. Coarse differentiation and multi-flows in planar graphs. 2007.

[17] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

[18] N. Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press.

[19] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[20] L. Lovász. Graph minor theory. *Bull. Amer. Math. Soc. (N.S.)*, 43(1):75–86 (electronic), 2006.

[21] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[22] H. Okamura and P. D. Seymour. Multicommodity flows in planar graphs. *J. Combin. Theory Ser. B*, 31(1):75–81, 1981.

[23] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete Comput. Geom.*, 19(1):79–94, 1998.

[24] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

[25] N. Robertson and P. Seymour. Graph minors XVII. Taming a vortex. (77):162–210, 1999.

[26] N. Robertson and P. D. Seymour. Graph minors. I. Excluding a forest. *J. Combin. Theory Ser. B*, 35(1):39–61, 1983.