# Effective-Resistance-Reducing Flows, Spectrally Thin Trees, and Asymmetric TSP

Nima Anari
*EECS, UC Berkeley*
*Email: anari@berkeley.edu*

Shayan Oveis Gharan
*CSE, University of Washington*
*Email: shayan@cs.washgington.edu*

**Abstract**

We show that the integrality gap of the natural LP relaxation of the Asymmetric Traveling Salesman Problem is polyloglog($n$). In other words, there is a polynomial time algorithm that approximates the *value* of the optimum tour within a factor of polyloglog($n$), where polyloglog($n$) is a bounded degree polynomial of loglog($n$). We prove this by showing that any $k$-edge-connected unweighted graph has a polyloglog$(n)/k$-thin spanning tree.

Our main new ingredient is a procedure, albeit an exponentially sized convex program, that "transforms" graphs that do not admit any *spectrally* thin trees into those that provably have spectrally thin trees. More precisely, given a $k$-edge-connected graph $G = (V, E)$ where $k \geq 7\log(n)$, we show that there is a matrix $D$ that "preserves" the structure of all cuts of $G$ such that for a set $F \subseteq E$ that induces an $\Omega(k)$-edge-connected graph, the effective resistance of every edge in $F$ w.r.t. $D$ is at most polylog$(k)/k$. Then, we use our extension of the seminal work of Marcus, Spielman, and Srivastava [1], fully explained in [2], to prove the existence of a polylog$(k)/k$-spectrally thin tree with respect to $D$. Such a tree is polylog$(k)/k$-combinatorially thin with respect to $G$ as $D$ preserves the structure of cuts of $G$.

**Keywords**

Asymmetric Traveling Salesman Problem; Approximation Algorithms; Thin Tree Conjecture; Kadison-Singer Problem; Effective Resistance.

## I. Introduction

In the Asymmetric Traveling Salesman Problem (ATSP) we are given a set $V$ of $n := |V|$ vertices and a nonnegative cost function $c : V \times V \to \mathbb{R}_+$. The goal is to find the shortest tour that visits every vertex *at least* once.

If the cost function is symmetric, i.e., $c(u, v) = c(v, u)$ for all $u, v \in V$, then the problem is known as the Symmetric Traveling Salesman Problem (STSP). There is a 3/2 approximation algorithm by Christofides [3] for STSP.

There is a natural LP relaxation for ATSP proposed by Held and Karp [4],

$$\begin{aligned} \min \sum_{u,v\in V} & c(u,v)x_{u,v} & \\ \text{s.t.} \sum_{u\in S, v\notin S} & x_{u,v} \geq 1 & \forall S \subseteq V, \\ \sum_{v\in V} x_{u,v} &= \sum_{v\in V} x_{v,u} = 1 & \forall u \in V, \\ x_{u,v} &\geq 0 & \forall u, v \in V. \end{aligned} \tag{1}$$

It is conjectured that the integrality gap of the above LP relaxation is a constant, i.e., the optimum value of the above LP relaxation is within a constant factor of the length of the optimum ATSP tour. Until very recently, we had a very limited understanding of the solutions of the above LP relaxation. To this date, the best known lower bound on the integrality gap of the above LP is 2 [5].

Despite many efforts, there is no known constant factor approximation algorithm for ATSP. Recently, Asadpour, Goemans, Madry, the second author, and Saberi [6] designed an $O(\log n/\log\log n)$ approximation algorithm for ATSP that broke the $O(\log n)$ barrier from Frieze, Galbiati, and Maffioli [7] and subsequent improvements [8]–[10]. The result of [6] also upper-bounds the integrality gap of the Held-Karp LP relaxation by $O(\log n/\log\log n)$. Later, the second author with Saberi [11] and subsequently Erickson and Sidiropoulos [12] designed constant factor approximation algorithms for ATSP on planar and bounded genus graphs.

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.11

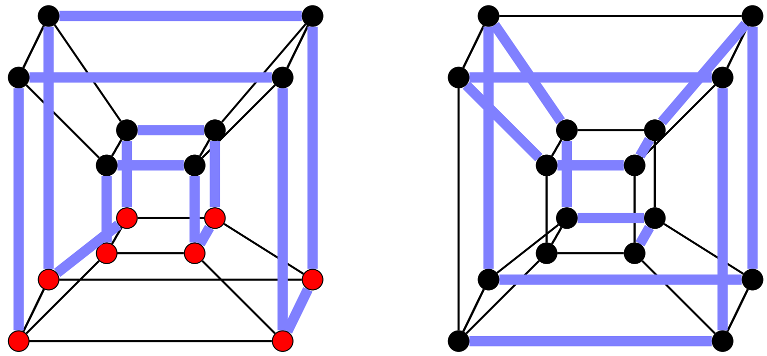

Figure 1: Two spanning trees of 4-dimensional hypercube that is 4-edge-connected. Although both of the trees are Hamiltonian paths, the left spanning tree is 1-thin because all of the edges of the cut separating red vertices from the black ones are in the tree while the right spanning tree is 0.667-thin.

*Thin Trees:* The main ingredient of all of the above recent developments is the construction of a "thin" tree. Let $G = (V, E)$ be an unweighted undirected $k$-edge-connected graph with $n$ vertices. Recall that $G$ is $k$-edge-connected if there are at least $k$ edges in every cut of $G$. We allow $G$ to have an arbitrary number of parallel edges, so we think of $E$ as a multiset of edges. Roughly speaking, a spanning tree $T \subseteq E$ is $\alpha$-thin with respect to $G$ if it does not contain more than $\alpha$-fraction of the edges of any cut in $G$.

**Definition 1.** *A spanning tree $T \subseteq E$ is $\alpha$-thin with respect to a (unweighted) graph $G = (V, E)$, if for each set $S \subseteq V$,*

$$|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|,$$

*where $T(S, \overline{S})$ and $E(S, \overline{S})$ are the set of edges of $T$ and $G$ in the cut $(S, \overline{S})$ respectively.*

One can analogously define $\alpha$-thin edge covers, $\alpha$-thin paths, etc. Note that thinness is a downward closed property, that is any subgraph of an $\alpha$-thin subgraph of $G$ is also $\alpha$-thin. In particular, any spanning tree of an $\alpha$-thin connected subgraph of $G$ is an $\alpha$-thin spanning tree of $G$. See Figure 1 for two examples of thin trees.

A key lemma in [6] shows that one can obtain an approximation algorithm for ATSP by finding a thin tree of small cost with respect to the graph defined by the fractional solution of the LP relaxation. In addition, proving the existence of a thin tree provides a bound on the integrality gap of the Held-Karp LP relaxation for ATSP.

Later, in [11] this connection is made more concrete. Namely, to break the $\Theta(\frac{\log(n)}{\log\log(n)})$ barrier, it suffices to ignore the costs of the edges and construct a thin tree in every $k$-edge-connected graph for $k = \Theta(\log(n))$.

**Theorem 2.** *For any $\alpha > 0$ (which can be a function of $n$) a polynomial-time construction of an $\alpha/\log(n)$-thin tree in any $\Theta(\log(n))$-edge-connected graph gives an $O(\alpha)$-approximation algorithm for ATSP. In addition, even an existential proof gives an $O(\alpha)$ upper bound on the integrality gap of the LP relaxation.*

The above theorem shows that to understand the solutions of LP (1) it is enough to understand the thin tree problem in graphs with low connectivity.

It is easy to show that any $k$-edge-connected graph has an $O(\log(n)/k)$-thin tree [13] using the independent randomized rounding method of Raghavan and Thompson [14]. It is enough to sample each edge of $G$ independently with probability $\Theta(\log(n)/k)$ and then choose an arbitrary spanning tree of the sampled graph.

Asadpour et al. [6] employ a more sophisticated randomized rounding algorithm and show that any $k$-edge-connected graph has a $\frac{\log(n)}{k\cdot\log\log(n)}$-thin tree. The basic idea of their algorithm is to use a correlated distribution, that is to sample edges almost independently while preserving the connectivity of the sampled set. More precisely, they sample a random spanning tree from a distribution where the edges are negatively correlated, so they get connectivity for free, and they only use the upper tail of the Chernoff types of bounds. The $1/\text{loglog}(n)$ gain comes from the fact that the upper tail of the Chernoff bound is slightly stronger than the lower tail,

Independently of the above applications of thin trees, Goddyn formulated the thin tree conjecture because of the close connections to several long-standing open problems regarding nowhere-zero flows.

**Conjecture 3** (Goddyn [15]). *There exists a function $f(\alpha)$ such that, for any $0 < \alpha < 1$, every $f(\alpha)$-edge-connected graph (of arbitrary size) has an $\alpha$-thin spanning tree.*

Goddyn's conjecture in the strongest form postulates that for a sufficiently large $k$ that is independent of the size of $G$, every $k$-edge-connected graph has an $O(1/k)$-thin tree. Goddyn proved that if the above conjecture holds for an arbitrary function $f(.)$, it implies a weaker version of Jaeger's conjecture on the existence of circular nowhere-zero flows [16]. Very recently, Thomassen proved a weaker version of Jaeger's conjecture [17], [18], but his proof has not yet shed any light on the resolution of the thin tree conjecture.

To this date, Conjecture 3 is only proved for planar and bounded genus graphs [11], [12] and edge-transitive graphs[1] [1], [19] for $f(\alpha) = O(1/\alpha)$. We remark that if Goddyn's thin tree conjecture holds for an arbitrary function $f(.)$, we get an upper bound of $O(\log^{1-\Omega(1)}(n))$ on the integrality gap of the LP relaxation of ATSP.

*Summary of our Contribution.:* In this paper, we show that any $k$-edge-connected graph has a polyloglog$(n)/k$-thin tree. Using Theorem 2 for $\alpha$ = polyloglog$(n)$ and $k = \log(n)$ this implies that the integrality gap of the LP relaxation is polyloglog$(n)$. Note that this does not resolve Goddyn's conjecture. Perhaps, one of the main consequences of our work is that we can round (not necessarily in polynomial time) the solutions of the LP relaxation exponentially better than the randomized rounding in the worst case.

The key to our proof is to rigorously relate the thin tree problem to a seemingly related spectral question that is known as the Kadison-Singer problem in operator theory [20] and then to use tools in spectral (graph) theory to solve the new problem. Until very recently, the best solution to the Kadison-Singer problem and the Weaver conjecture was based on the randomized rounding technique and matrix Chernoff bounds and incurred a loss of $\log(n)$ [21], [22]. Marcus, Spielman, and Srivastava [1] in a breakthrough managed to resolve the conjecture using spectral techniques with no cost that is dependent on $n$. As we will elaborate in the next section, the Kadison-Singer problem can be seen as an "$L_2$" version of the thin tree question, or thin tree question can be seen as an $L_1$ version of the Kadison-Singer problem. So, we can summarize our contribution as an $L_1$ to $L_2$ reduction.

We construct this $L_1$ to $L_2$ reduction using a convex program that symmetrizes the $L_2$ structure of a given graph while preserving its $L_1$ structure. More precisely, a convex program that equalizes the *effective resistance* of the edges while preserving the cut structure of $G$. We expect to see several other applications of this convex program in combinatorial optimization and approximation algorithms. In addition to that, we extend the result of Marcus, Spielman, and Srivastava to a larger family of distributions known as *strongly Rayleigh* distributions [2]. Strongly Rayleigh distributions are a family of probability distributions with the strongest forms of negative dependence properties [23]. They have been used also in a recent work of the second author, Saberi, and Singh [24] to improve the Christofides approximation algorithm for STSP on graph metrics. We refer the interested readers to [2] for more information.

*Subsequent Work:* Subsequent to our work, Svensson [25] employed a sophisticated cycle cover idea and designed a constant factor approximation algorithm for ATSP when $c(.,.)$ is the shortest path metric of an unweighted graph. It is unclear if a combination of the ideas in this work and [25] can lead to constant factor approximation algorithms for general ATSP.

The rest of this section is organized as follows: In Subsection I-A we overview the connections of the thin tree problem and graph sparsifiers and in particular the Kadison-Singer problem. Then, in Subsection I-B we present our main theorems. Finally, in Subsection I-C we highlight the main ideas of the proof.

### *A. Spectrally Thin Trees*

As mentioned before, thin trees are the basis for the best-known approximation algorithms for ATSP on planar, bounded genus, or general graphs. This follows from their intuitive definition and the fact that they eliminate the difficulty arising from the underlying asymmetry and the cost function. On the other hand, the major challenge in constructing thin trees or proving their existence is that we are not aware of any efficient algorithm for measuring or certifying the thinness of a given tree exactly. In order to verify the thinness of a given tree, it seems that one has to look at exponentially many cuts.

One possible way to avoid this difficulty is to study a stronger definition of thinness, namely the *spectral* thinness. First, we define some notation. For a set $S \subseteq V$ we use $\mathbf{1}_S \in \mathbb{R}^V$ to denote the indicator (column) vector of the set $S$. For a vertex $v \in V$, we abuse notation and write $\mathbf{1}_v$ instead of $\mathbf{1}_{\{v\}}$. For any edge $e = \{u, v\} \in E$ we fix an

[1]A graph $G = (V, E)$ is edge-transitive, if for any pair of edges $e, f \in E$ there is an automorphism of $G$ that maps $e$ to $f$.

arbitrary orientation, say $u \to v$, and we define $\mathcal{X}_e := \mathbf{1}_u - \mathbf{1}_v$. The Laplacian of $G$, $L_G$, is defined as follows:

$$L_G := \sum_{e \in E} \mathcal{X}_e \mathcal{X}_e^\intercal.$$

If $G$ is weighted, then we scale up each term $\mathcal{X}_e \mathcal{X}_e^\intercal$ according to the weight of the edge $e$. Also, for a set $T \subseteq E$ of edges, we write

$$L_T := \sum_{e \in T} \mathcal{X}_e \mathcal{X}_e^\intercal.$$

We say a spanning tree, $T$, is $\alpha$-spectrally thin with respect to $G$ if

$$L_T \preceq \alpha \cdot L_G, \text{ i.e., for all } x \in \mathbb{R}^n,\ x^\intercal L_T x \le \alpha \cdot x^\intercal L_G x. \quad (2)$$

We also say $G$ has a spectrally thin tree if it has an $\alpha$-spectrally thin tree for some $\alpha < 1/2$. Observe that if $T$ is $\alpha$-spectrally thin, then it is also $\alpha$-(combinatorially) thin. To see that, note that for any set $S \subseteq V$, $\mathbf{1}_S^\intercal L_T \mathbf{1}_S = |T(S, \overline{S})|$ and $\mathbf{1}_S^\intercal L_G \mathbf{1}_S = |E(S, \overline{S})|$.

One can verify spectral thinness of $T$ (in polynomial time) by finding the smallest $\alpha \in \mathbb{R}$ such that

$$L_G^{\dagger/2} L_T L_G^{\dagger/2} \preceq \alpha \cdot I,$$

i.e., by computing the largest eigenvalue of $L_G^{\dagger/2} L_T L_G^{\dagger/2}$. Recall that $L_G^\dagger$ is the pseudoinverse of $L_G$, and $L_G^{\dagger/2}$ is the square root of the pseudoinverse of $L_G$; $L_G^{\dagger/2}$ is well-defined because $L_G^\dagger \succeq 0$. So, unlike the combinatorial thinness, spectral thinness can be computed *exactly* in polynomial time.

The notion of spectral thinness is closely related to spectral sparsifiers of graphs, which have been studied extensively in the past few years [26]–[29]. Roughly speaking, a spectrally thin tree is a one-sided spectral sparsifier. A spectrally thin tree $T$ would be a true spectral sparsifier if in addition to (2), it satisfies $\alpha \cdot (1-\epsilon) x^\intercal L_G x \preceq L_T$ for some constant $\epsilon$. Until the recent breakthrough of Batson, Spielman, and Srivastava, all constructions of spectral sparsifiers used at least $\Omega(n \log(n))$ edges of the graph [26], [27], [29]. Because of this they are of no use for the particular application of ATSP. Batson, Spielman, and Srivastava [28] managed to construct a spectral sparsifier that uses only $O(n)$ edges of $G$. But in their construction, they assign different weights to the edges of the sparsifier which again makes their contribution not helpful for ATSP.

Indeed, it was observed by several people that there is an underlying barrier for the construction of spectrally thin trees and *unweighted* spectral sparsifiers. Many families of $k$-edge-connected graphs do not admit spectrally thin trees (see [19, Thm 4.9]). Let us elaborate on this observation. The *effective resistance* of an edge $e = \{u, v\}$ in $G$, $\mathcal{R}\text{eff}_{L_G}(e)$, is the *energy* of the electrical flow that sends 1 unit of current from $u$ to $v$ when the network represents an electrical circuit with each edge being a resistor of resistance 1 (and if $G$ is weighted, the resistance is the inverse of the weight of $e$). See [30, Ch. 2] for background on electrical flows and effective resistance. Mathematically, the effective resistance can be computed using $L_G^\dagger$,

$$\mathcal{R}\text{eff}_{L_G}(e) := \mathcal{X}_e^\intercal L_G^\dagger \mathcal{X}_e.$$

It is not hard to see that the spectral thinness of any spanning tree $T$ of $G$ is at least the maximum effective resistance of the edges of $T$ in $G$.

**Lemma 4.** *For any graph $G = (V, E)$, the spectral thinness of any spanning tree $T \subseteq E$ is at least* $\max_{e \in T} \mathcal{R}\text{eff}_{L_G}(e)$.

*Proof:* Say the spectral thinness of $T$ is $\alpha$. Obviously, by the downward closedness of spectral thinness, the spectral thinness of any subset of edges of $T$ is at most $\alpha$, i.e., for any edge $e \in T$,

$$L_{\{e\}} \preceq L_T \preceq \alpha \cdot L_G.$$

But, the spectral thinness of an edge is indeed its effective resistance. More precisely, multiplying $L_G^{\dagger/2}$ on both sides of the above inequality we have

$$L_G^{\dagger/2} \mathcal{X}_e \mathcal{X}_e^\intercal L_G^{\dagger/2} = L_G^{\dagger/2} L_{\{e\}} L_G^{\dagger/2} \preceq \alpha \cdot L_G^{\dagger/2} L_G L_G^{\dagger/2} \preceq \alpha \cdot I.$$

Since the matrix on the LHS has rank one, its only eigenvalue is equal to its trace; therefore,

$$\text{Tr}(\mathcal{X}_e^\intercal L_G^\dagger \mathcal{X}_e) = \text{Tr}(L_G^{\dagger/2} \mathcal{X}_e \mathcal{X}_e^\intercal L_G^{\dagger/2}) \le \alpha.$$

The lemma follows by the fact that $\mathcal{R}\text{eff}_{L_G}(e) = \text{Tr}(\mathcal{X}_e^\intercal L_G^\dagger \mathcal{X}_e)$. ■

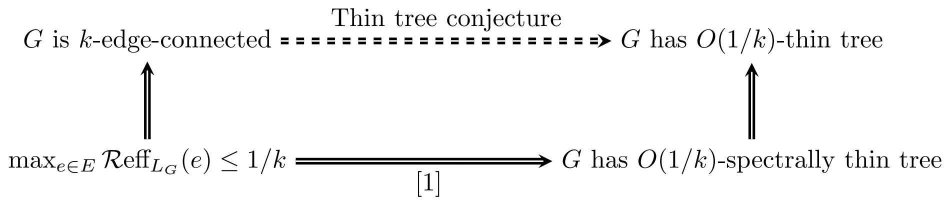


Figure 2: A summary of the relationship between spectrally thin trees and combinatorially thin trees before our paper.

In light of the above lemma, a necessary condition for $G$ to have a spanning tree with spectral thinness bounded away from 1 is that every cut of $G$ must have at least one edge with effective resistance bounded away from 1. In other words, any graph $G$ with at least one cut where the effective resistance of every edge is very close to 1 has no spectrally thin tree (see Figure 3 for an example of a graph where the effective resistance of every edge in a cut is very close to 1).

In a very recent breakthrough, Marcus, Spielman, and Srivastava [1] proved the Kadison-Singer conjecture. As a byproduct of their result, it was shown in [19] that a stronger version of the above condition is sufficient for the existence of spectrally thin trees.

**Theorem 5** ( [1])**.** *Any connected graph $G = (V, E)$ has a spanning tree with spectral thinness $O(\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e))$.*

See [19, Appendix E] for a detailed proof of the above theorem. It follows from the above theorem that every $k$-edge-connected edge-transitive graph has an $O(1/k)$-spectrally thin tree. This is because in any edge-transitive graph, by symmetry, the effective resistances of all edges are equal.

Let us summarize the relationship between spectrally thin trees and combinatorially thin trees that has been in the literature before our work. Goddyn conjectured that every $k$-edge-connected graph has an $O(1/k)$-thin tree. The result of [1] shows that a stronger assumption implies an stronger conclusion, i.e., if the maximum effective resistance of edges of $G$ is at most $1/k$, then $G$ has an $O(1/k)$-spectrally thin tree (see Figure 2).

We emphasize that $\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e) \leq 1/k$ is a stronger assumption than $k$-edge-connectivity. If $\mathcal{R}\mathrm{eff}_{L_G}(u, v) \leq 1/k$, it means that when we send one unit of flow from $u$ to $v$, the electric current divides and goes through at least $k$ parallel paths connecting $u$ to $v$, so, there are $k$ edge-disjoint paths between $u, v$. But the converse of this does not necessarily hold. If there are $k$ edge-disjoint paths from $u$ to $v$, the electric current may just use one of these paths if the rest are very long, so the effective resistance can be very close to 1. Therefore, if $\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e) \leq 1/k$, there are $k$ edge-disjoint paths between each pair of vertices of $G$, and $G$ is $k$-edge-connected, but the converse does not necessarily hold. For example in the graph in the top of Figure 3, even though there are $k$ edge-disjoint paths from $u_1$ to $v_1$, a unit electrical flow from $u_1$ to $v_1$ almost entirely goes through the edge $\{u_1, v_1\}$, so $\mathcal{R}\mathrm{eff}(u_1, v_1) \approx 1$.

As a side remark, note that the sum of effective resistances of all edges of any connected graph $G$ is $n - 1$,

$$\sum_{e\in E} \mathcal{X}_e^\intercal L_G^\dagger \mathcal{X}_e = \sum_{e\in E} \mathrm{Tr}(L_G^{\dagger/2} \mathcal{X}_e \mathcal{X}_e^\intercal L_G^{\dagger/2}) = \mathrm{Tr}\left(\sum_{e\in E} L_G^{\dagger/2} \mathcal{X}_e \mathcal{X}_e^\intercal L_G^{\dagger/2}\right) = \mathrm{Tr}(L_G^{\dagger/2} L_G L_G^{\dagger/2}) = n - 1.$$

In the last identity we use that $L_G^{\dagger/2} L_G L_G^{\dagger/2}$ is an identity matrix on the space of vectors that are orthogonal to the all-1s vector.

If $G$ is $k$-edge-connected, by Markov's inequality, at most a quarter of the edges have effective resistance more than $8/k$. Therefore, by an application of [1], any $k$-edge-connected graph $G$ has an $O(1/k)$-spectrally thin set of edges, $F \subset E$ where $|F| \geq \Omega(n)$ [19]. Unfortunately, the corresponding subgraph $(V, F)$ may have $\Omega(n/k)$ connected components. So, this does not give any improved bounds on the approximability of ATSP.

### *B. Our Contribution*

In this paper we introduce a procedure to "transform" graphs that do not admit spectrally thin trees into those that *provably* have these trees. Then, we use our recent extension of [1] to *strongly Rayleigh distributions* [2] to find spectrally thin trees in the transformed "graph". Finally, we show that any spectrally thin tree of the transformed "graph" is a (combinatorially) thin tree in the original graph. From a high level perspective, our transformation massages the graph to equalize the effective resistance of the edges, while keeping the cut structure of the graph intact.

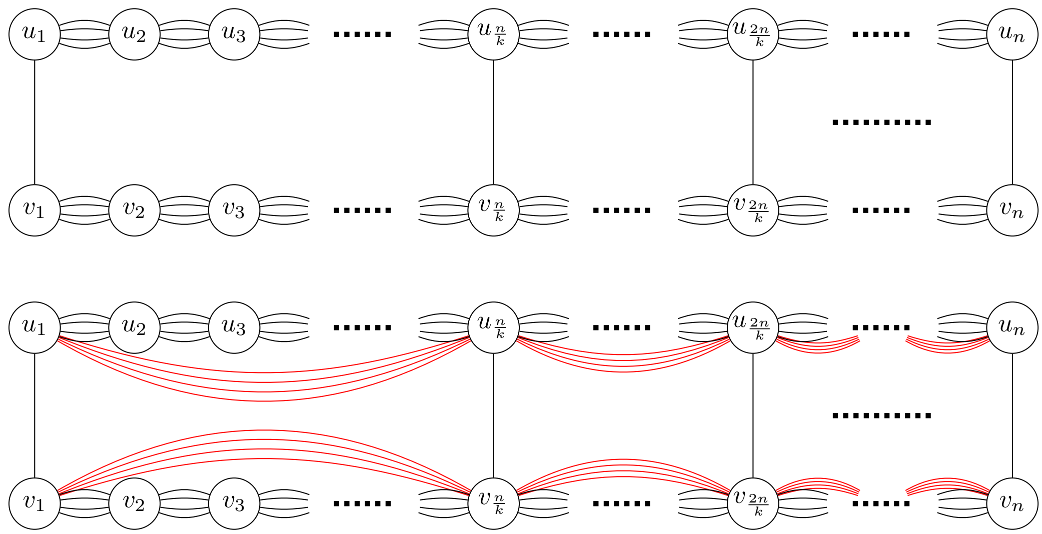


Figure 3: The top shows a $k$-edge-connected planar graph that has no spectrally thin tree. There are $k+1$ vertical edges, $(u_1, v_1), (u_{n/k}, v_{n/k}), \ldots, (u_n, v_n)$. For each $1 \leq i \leq n-1$ there are $k$ parallel edges between $u_i, u_{i+1}$ and $v_i, v_{i+1}$. The effective resistances of all vertical edges are $1 - O(k^2/n)$. The bottom shows a graph $G + D$ where the effective resistance of every black edge is $O(1/\sqrt{k})$. The red edges are edges in $D$ and there are $k$ parallel edges between the endpoints of consecutive vertical edges. Note that $L_D \preceq_\square L_G$ by construction.

For two matrices $A, B \in \mathbb{R}^{n \times n}$, we write $A \preceq_\square B$, if for any set $\emptyset \subset S \subsetneq V$,

$$\mathbf{1}_S^\intercal A \mathbf{1}_S \leq \mathbf{1}_S^\intercal B \mathbf{1}_S.$$

Note that $A \preceq B$ implies $A \preceq_\square B$, but the converse is not necessarily true. We say a graph $D$ is a *shortcut* graph with respect to $G$ if $L_D \preceq_\square L_G$. We say a positive definite (PD) matrix $D$ is a shortcut matrix with respect to $G$ if $D \preceq_\square L_G$.

Our ideal plan is as follows: Show that there is a (weighted) shortcut graph $D$ such that for any edge $e \in E$, $\mathcal{R}\text{eff}_{L_D}(e) \leq \tilde{O}(1/k)$. Then, use a simple extension of Theorem 5 such as [31] to show that there is a spanning tree $T \subseteq E$ such that

$$L_T \preceq_\square \alpha \cdot (L_G + L_D),$$

for $\alpha = O(\max_{e \in E} \mathcal{R}\text{eff}_{L_G + L_D}(e)) = \tilde{O}(1/k)$. But, since $L_D \preceq_\square L_G$, any $\alpha$-spectrally thin tree of $D + G$ is a $2\alpha$-combinatorially thin tree of $G$. In summary, the graph $D$ allows us to bypass the spectral thinness barrier that we described in Lemma 4.

Let us give a clarifying example. Consider the $k$-edge-connected planar graph $G$ illustrated at the top of Figure 3. In this graph, all edges in the cut $(\{v_1, \ldots, v_n\}, \{u_1, \ldots, u_n\})$ have effective resistance very close to 1. Now, let $D$ consist of the red edges shown at the bottom. Observe that $L_D \preceq_\square L_G$. The effective resistance of every *black* edge in $G + D$ is $O(1/\sqrt{k})$. Roughly speaking, this is because the red edges *shortcut* the long paths between the endpoints of vertical edges. This reduces the energy of the corresponding electrical flows. So, $G + D$ has a spectrally thin tree $T \subseteq E$. Such a tree is combinatorially thin with respect to $G$.

It turns out that there are $k$-edge-connected graphs where it is impossible to reduce the effective resistance of all edges by a shortcut graph $D$ (see Section IV for details). So, in our main theorem, we prove a weaker version of the above ideal plan. Firstly, instead of finding a shortcut graph $D$, we find a PD shortcut matrix $D$. The matrix $D$ does not necessarily represent the Laplacian matrix of a graph as it may have positive off-diagonal entries. Secondly, the shortcut matrix reduces the effective resistance of only a set $F \subseteq E$ of edges, that we call *good* edges, where $(V, F)$ is $\Omega(k)$-edge-connected.

**Theorem 6** (Main). *For any $k$-edge-connected graph $G = (V, E)$ where $k \geq 7\log(n)$, there is a shortcut matrix $0 \prec D \preceq_\square L_G$ and a set of good edges $F \subseteq E$ such that the graph $(V, F)$ is $\Omega(k)$-edge-connected and that for any edge*

$e \in F$,

$$\mathcal{R}\text{eff}_D(e) \leq \tilde{O}(1/k),^2$$

*where* $\mathcal{R}\text{eff}_D(e) = \mathcal{X}_e^\intercal D^{-1} \mathcal{X}_e$.

Note that in the above we upper bound the effective resistance of good edges with respect to $D$ as opposed to $D + L_G$; this is sufficient because $\mathcal{R}\text{eff}_{L_G+D}(e) \leq \mathcal{R}\text{eff}_D(e)$. We remark that the dependency on $\log(n)$ in the statement of the theorem is because of a limitation of our current proof techniques. We expect that a corresponding statement without any dependency on $n$ holds for any $k$-edge-connected graph $G$. Such a statement would resolve Goddyn's thin tree conjecture 3 and may lead to improved bounds on the integrality gap of LP (1). Finally, the logarithmic dependency on $k$ in the upper bound on the effective resistance of the edges of $F$ is necessary.

Unfortunately, the good edges in the above theorem may be very sparse with respect to $G$, i.e., $G$ may have cuts $(S, \overline{S})$ such that

$$|F(S, \overline{S})| \ll |E(S, \overline{S})|.$$

So, if we use Theorem 5 or its simple extensions as in [31], we get a thin set of edges $T \subseteq E$ that may have $\Omega_k(n)$ many connected components. Instead, we use a theorem, that we proved in our recent extension of [1], that shows that as long as $F$ is $\Omega(k)$-edge-connected, $G$ has a spanning tree $T$ that is $\tilde{O}(1/k)$-spectrally thin with respect to $D + L_G$.

**Theorem 7** ([2]). *Given a graph $G = (V, E)$, a PD matrix $D$ and $F \subseteq E$ such that $(V, F)$ is $k$-edge-connected, if for $\epsilon > 0$,*

$$\max_{e \in F} \mathcal{R}\text{eff}_D(e) \leq \epsilon,$$

*then $G$ has a spanning tree $T \subseteq F$ s.t.,*

$$L_T \preceq O(\epsilon + 1/k)(D + L_G).$$

Putting Theorem 6 and Theorem 7 together implies that any $k$-edge-connected graph has a $\text{polyloglog}(n)/k$-thin tree.

**Corollary 8.** *Any $k$-edge-connected graph $G = (V, E)$, has a* $\text{polyloglog}(n)/k$*-thin tree.*

We remark that in order to design a polynomial-time algorithm for finding thin trees, we need to have a constructive (in polynomial time) proof of Theorem 7. We also remark that, the above theorems do not resolve Goddyn's thin tree conjecture because of the dependency on $n$.

### *C. Main Components of the Proof*

Our proof has three main components, namely the thin basis problem, the effective resistance reducing convex programs, and the locally connected hierarchies. In this section we summarize the high-level interaction of these three components.

*The Thin Basis Problem:* Let us start by an overview of the proof of Theorem 7 which appears in full detail in [2]. The thin basis problem is defined as follows: Given a set of vectors $\{x_e\}_{e \in E} \in \mathbb{R}^d$, what is a sufficient condition for the existence of an $\alpha$-thin basis, namely, $d$ linearly independent set of vectors $T \subseteq E$ such that

$$\left\| \sum_{e \in T} x_e x_e^\intercal \right\| \leq \alpha?$$

It follows from the work of Marcus, Spielman, and Srivastava [1] that a sufficient condition for the existence of an $\alpha$-thin basis is that the vectors are in isotropic position,

$$\sum_{e \in E} x_e x_e^\intercal = I,$$

and for all $e \in E$, $\|x_e\|^2 \leq c \cdot \alpha$ for some universal constant $c < 1$.

The thin basis problem is closely related to the existential problem of spectrally thin trees. Say we want to see if a given graph $G = (V, E)$ has a spectrally thin tree. We can define a vector $y_e = L_G^{\dagger/2} \mathcal{X}_e$ for each edge $e \in E$. It turns out that these vectors are in isotropic position; in addition, if all edges of $G$ have effective resistance at most $\epsilon$, then

[2] For functions $f(.), g(.)$ we write $g = \tilde{O}(f)$ if $g(n) \leq \text{polylog}(f(n)) \cdot f(n)$ for all sufficiently large $n$.

$\|y_e\|^2 = \mathcal{X}_e^\intercal L_G^\dagger \mathcal{X}_e \leq \epsilon$. So, these vectors contain an $O(\epsilon)$-thin basis. It is easy to see that such a basis corresponds to an $O(\epsilon)$-spectrally thin tree of $G$ (see [2] for details).

As alluded to in the introduction, if $G$ is a $k$-edge-connected graph, it may have many edges of large effective resistance, so $\|y_e\|^2$ in the above argument may be very close to 1. We use the shortcut matrix $D$ that is promised in Theorem 6 to reduce the squared norm of the vectors. We assign a vector $y_e = (L_G + D)^{\dagger/2}\mathcal{X}_e$ to any good edge $e \in F$. It follows that

$$\|y_e\|^2 \leq \mathcal{X}_e^\intercal (L_G + D)^\dagger \mathcal{X}_e \leq \tilde{O}(1/k).$$

But, since the good edges are only a subset of the edges of $G$, the set of vectors $\{y_e\}_{e\in F}$ are not necessarily in an isotropic position; they are rather in a sub-isotropic position,

$$\sum_{e\in F} y_e y_e^\intercal \preceq I.$$

In [2] we prove a weaker sufficient condition for the existence of a thin basis. If the vectors $\{x_e\}_{e\in E}$ are in a sub-isotropic position, each of them has a squared norm at most $\epsilon$, and they contain $k$ disjoint bases, then there exists an $O(\epsilon + 1/k)$-thin basis $T \subset E$

$$\left\| \sum_{e\in E} x_e x_e^\intercal \right\| \leq O(\epsilon + 1/k).$$

Since, the set $F$ of good edges promised in Theorem 6 is $\Omega(k)$-edge-connected, it contains $\Omega(k)$ edge-disjoint spanning trees, so the set of vectors $\{y_e\}_{e\in F}$ defined above contains $\Omega(k)$ disjoint bases. So, $\{y_e\}_{e\in F}$ contains a $\tilde{O}(1/k)$-thin basis $T$; this corresponds to a $\tilde{O}(1/k)$-spectrally thin tree of $L_G + D$ and a $\tilde{O}(1/k)$-thin tree of $G$.

*Effective Resistance Reducing Convex Programs.:* As illustrated in the previous section, at the heart of our proof we find a PD shortcut matrix $D$ to reduce the effective resistance of a subset of edges of $G$.

It turns out that the problem of finding the best shortcut matrix $D$ that reduces the maximum effective resistance of the edges of $G$ is convex. This is because for any fixed vector $x$ and $D \succ 0$, $x^\intercal D^{-1} x$ is a convex function of $D$. The problem of minimizing the sum of effective resistances of all pairs of vertices in a given graph was previously studied in [32].

The following (exponentially sized) convex program finds the best shortcut matrix $D$ that minimizes the maximum effective resistance of the edges of $G$ while preserving the cut structure of $G$.

$$\begin{aligned} \min \quad & \mathcal{E}, \\ \text{s.t.} \quad & \mathcal{R}\text{eff}_D(e) \leq \mathcal{E} \quad \forall e \in E, \\ & D \preceq_\square L_G, \\ & D \succ 0. \end{aligned} \tag{3}$$

Note that if we replace the constraint $D \preceq_\square L_G$ with $D \preceq L_G$, i.e., if we require $D$ to be upper-bounded by $L_G$ in the PSD sense, then the optimum $D$ for any graph $G$ is exactly $L_G$ and the optimum value is the maximum effective resistance of the edges of $G$.

Unfortunately, the optimum of the above program can be very close to 1 even if the input graph $G$ is $\log(n)$-edge-connected. A bad graph is shown in Figure 4. In the full version of this work, [33], we show that the optimum of the above convex program for the family of graphs in Figure 4 is close to 1 by constructing a feasible solution of the dual.

To prove our main theorem, we study a variant of the above convex program that reduces the effective resistance of only a subset of edges of $G$ to $\tilde{O}(1/k)$. We will use combinatorial objects called locally connected hierarchies as discussed in the next paragraph to feed a carefully chosen set of edges into the convex program. To show that the optimum value of the program is $\tilde{O}(1/k)$, we analyze its dual. The dual problem corresponds to proving an upper bound on the ratio involving distances of pairs of vertices of $G$ with respect to an $L_1$ embedding of the vertices in a high-dimensional space. We refrain from going into the details at this point. We will provide a more detailed overview in Section IV.

*Locally Connected Hierarchies:* The main difficulty in proving Theorem 6 is that the good edges, $F$, are unknown a priori. If we knew $F$ then we could use (3) to minimize the maximum effective resistance of edges of $F$ as opposed to $E$. In addition, the $k$-th smallest effective resistance of the edges of a cut of $G$ is not a convex function of $D$. So, we cannot write a single program that gives us the best matrix $D$ for which there are at least $\Omega(k)$ edges of small effective resistance in every cut of $G$.

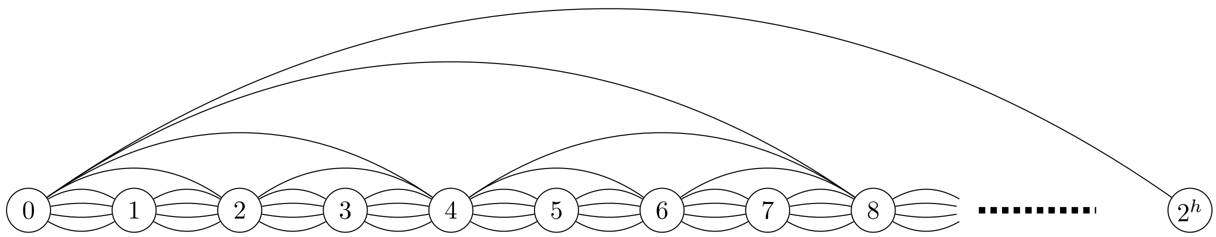

Figure 4: A tight example for (3). The graph has $2^h + 1$ vertices labeled with $\{0, 1, \dots, 2^h\}$. There are $k$ parallel edges connecting each pair of consecutive vertices. In addition, for any $1 \le i \le h$ and any $0 \le j < 2^{h-i}$ there is an edge $\{j \cdot 2^i, (j+1) \cdot 2^i\}$.

So, we take a detour. We use combinatorial structures that we call locally connected hierarchies that allow us to find an $\Omega(k)$-edge-connected set of good edges that may be very sparse with respect to $G$ in some of the cuts. Let us give an informal definition of locally connected hierarchies. Consider a *laminar* structure on the vertices of $G$, say $S_1, S_2, \dots \subseteq V$, where by a laminar structure we mean that there is no $i \neq j$ such that $S_i \cap S_j, S_i - S_j, S_j - S_i \neq \emptyset$. Modulo some technical conditions, if for all $i$, the induced subgraph on $S_i$, $G[S_i]$, is $k$-edge-connected, then we call $S_1, S_2, \dots$ a locally connected hierarchy.

Let $S_{i^*}$ be the smallest set that is a superset of $S_i$ in the family, and let $\mathcal{O}(S_i) = E(S_i, S_{i^*} - S_i)$ be the set of edges leaving $S_i$ in the induced graph $G[S_{i^*}]$. In our main technical theorem we show that for any locally connected hierarchy we can find a shortcut matrix $D$ that reduces the maximum of the average effective resistance of all $\mathcal{O}(S_i)$'s. In other words, the shortcut matrix $D$ reduces the effective resistance of at least half of the edges of each $\mathcal{O}(S_i)$. Unfortunately, these small effective resistance edges may have $\Omega(n)$ connected components.

To prove Theorem 6 we choose polyloglog$(n)$ many locally connected hierarchies adaptively, such that the following holds: Let the laminar family $S_1^j, S_2^j, \dots$ be the $j$-th locally connected hierarchy, and $D_j$ be a shortcut matrix that reduces the maximum average effective resistance of $\mathcal{O}(S_i^j)$'s. We let $F_j$ be the set of small effective resistance edges in $\cup_i \mathcal{O}(S_i^j)$. We choose our locally connected hierarchies such that $F = \cup_j F_j$ is $\Omega(k)$-edge-connected in $G$. To ensure this we use several tools in graph partitioning.

### D. Organization

The rest of the paper is organized as follows: We start with an overview of linear algebraic tools and graph theoretic tools that we use in the paper. In Section III we given a high-level overview of our approach; we formally define locally connected hierarchies and we describe the main technical theorem 18. Then in Section IV, we describe some of the key ideas of the proof.

## II. Preliminaries

For a set $S \subseteq V$, we use $G[S]$ to denote the induced subgraph of $G$ on $S$. For disjoint sets $S, T \subseteq V$ we write $E(S,T) := \{\{u, v\} : u \in S, v \in T\}$. For a set $S$ of elements we write $\mathbb{E}_{e \sim S}[.]$ to denote the expectation under the uniform distribution over the elements of $S$.

For a matrix $A \in \mathbb{R}^{m \times n}$ we write $A_i$ to denote the $i$-th column of $A$, $A^i$ to denote the $i$-th row of $A$ and $A_{i,j}$ to denote the $i,j$-th entry of $A$.

For an edge $e = \{u, v\}$ we use $\mathcal{X}_e = \mathbf{1}_u - \mathbf{1}_v$. We use $\mathcal{X} \in \mathbb{R}^{V \times E}$ to denote the matrix where the $e$-th column is $\mathcal{X}_e$.

*Balls and High-Dimensional Geometry:* For $x \in \mathbb{R}^d$ and $r \in \mathbb{R}$, an $L_1$ ball is the set of points at $L_1$ distance less than $r$ of $x$,

$$B(x, r) := \{y \in \mathbb{R}^d : 0 < \|x - y\|_1 < r\}.$$

Unless otherwise specified, any ball that we consider in this paper is an $L_1$ ball. We may also work with $L_2$ or $L_2^2$ balls and by that we are referring to a set of points whose $L_2$ or $L_2^2$ distance from a center is bounded by $r$.

A cut metric of $S$ is a mapping $X : S \to \{0, 1\}^h$ equipped with the $L_1$ metric. Note that any cut metric of $S$, for any two elements $u, v \in S$,

$$\|X_u - X_v\|_1 = \|X_u - X_v\|^2.$$

We can look at an embedding $X$ as a matrix where there is a column $X_u$ for any vertex $u$. We also write $\mathbf{X} = X\mathcal{X}$. Therefore, for any edge $e = \{u,v\} \in E$, $\mathbf{X}_e = X\mathcal{X}_e = X_u - X_v$.

*Facts from Linear Algebra:* A matrix $U \in \mathbb{R}^{n\times n}$ is called orthogonal/unitary if $UU^\intercal = U^\intercal U = I$. An orthogonal matrix is a nonsingular square matrix whose singular values are all 1. It follows by definition that orthogonal operators preserve $L_2$ norms of vectors, i.e., for any vector $x \in \mathbb{R}^n$,

$$\|Ux\| = \sqrt{(Ux)^\intercal Ux} = \sqrt{x^\intercal U^\intercal U x} = \sqrt{x^\intercal x} = \|x\|.$$

A (not necessarily square) matrix $U$ is called semiorthogonal if $UU^\intercal = I$, i.e. the rows are orthonormal, and the number of rows is less than the number of columns.

For two matrices $A, B$ of the same dimension we define the matrix inner product $A \bullet B := \mathrm{Tr}(AB^\intercal)$. If $A$, $B$ are positive semidefinite, then $\mathrm{Tr}(AB) \geq 0$. For any matrix $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$, $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

**Definition 9** (Matrix Norms)**.** *The trace norm (or nuclear norm) of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as follows:*

$$\|A\|_* := \mathrm{Tr}((A^\intercal A)^{1/2}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i,$$

*where $\sigma_i$'s are the singular values of $A$. The Frobenius norm of $A$ is defined as follows:*

$$\|A\|_F := \sqrt{\sum_{1\leq i\leq m, 1\leq j\leq n} A_{i,j}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.$$

The following lemma is a well-known fact about the trace norm.

**Lemma 10.** *For any matrix $A \in \mathbb{R}^{n\times m}$ such that $n \geq m$,*

$$\|A\|_* = \max_{\text{Semiorthogonal } U} \mathrm{Tr}(UA),$$

*where the maximum is over all semiorthogonal matrices $U \in \mathbb{R}^{m\times n}$. In particular,* $\mathrm{Tr}(A) \leq \|A\|_*$.

**Theorem 11** (Hoffman-Wielandt Inequality)**.** *Let $A, B \in \mathbb{R}^{n\times n}$ have singular values $\sigma_1 \leq \sigma_2 \leq \ldots \sigma_n$ and $\sigma'_1 \leq \sigma'_2 \leq \ldots \leq \sigma'_n$. Then,*

$$\sum_{i=1}^{n} (\sigma_i - \sigma'_i)^2 \leq \|A - B\|_F^2.$$

*Background in Graph Theory:* For a graph $G = (V, E)$, and a set $S \subseteq V$, we define $\phi_G(S) := \frac{\partial_G(S)}{d_G(S)}$ where $\partial_G(S) := |E(S, V - S)|$ is the number of edges that leave $S$, and $d_G(S)$ is the sum of the degrees (in $G$) of vertices of $S$. The expansion of $G$ is defined as follows: $\phi(G) := \min_{S\subset V} \max\{\phi_G(S), \phi_G(V - S)\}$, We say a graph $G$ is an $\epsilon$-expander, if $\phi(G) \geq \epsilon$. Recall that in an expander graph, $\phi(G) = \Omega(1)$.

An (unweighted) graph $G = (V, E)$ is $k$-edge-connected if and only if for any pair of vertices $u, v \in V$, there are at least $k$ edge-disjoint paths between $u, v$ in $G$. Equivalently, $G$ is $k$-edge-connected if for any set $\emptyset \subsetneq S \subsetneq V$, $\partial(S) \geq k$. There is a well-known theorem by Nash-Williams [34] that any $k$-edge-connected graph has at least $k/2$ disjoint spanning trees.

Given a graph $G = (V, E)$, and a set $S \subseteq V$, we write $G/S$ to denote the graph where the set $S$ is *contracted*, i.e., we remove all vertices $v \in S$ and add a new vertex $u$ instead, and for any vertex $w \notin S$, we let $|E(S, \{w\})|$ be the number of (parallel) edges between $u$ and $w$. We also remove any self-loops that result from this operation. The following fact will be used throughout the paper.

**Fact 12.** *For any $k$-edge-connected graph $G = (V, E)$ and any set $S \subseteq V$, $G/S$ is $k$-edge-connected.*

Throughout the paper we may use a natural decomposition of a graph $G$ (that is not necessarily $k$-edge-connected) into $k$-edge-connected subgraphs as defined below.

**Definition 13.** *For a graph $G = (V, E)$ a natural decomposition into $k$-edge-connected subgraphs is defined as follows: Start with a partition $S_1 = V$. While there is a nonempty set $S_i$ in the partition such that $G[S_i]$ is not $k$-edge-connected, find an induced cut $(S_{i,1}, S_{i,2})$ in $G[S_i]$ of size less than $k$, remove $S_i$ and add $S_{i,1}, S_{i,2}$ as new sets in the partition.*

## III. Overview of Our Approach

In this section we give a high-level overview of our approach. We will motivate and formally define locally connected hierarchies and we describe our main technical theorem. In this section we will not overview the proof of the main technical theorem 18, see Section IV for the explanation.

As alluded to in the introduction, it is not possible to reduce the maximum effective resistance of the edges of every $k$-edge-connected graph using a shortcut matrix.

The first idea that comes to mind is to reduce the maximum average effective resistance amongst all cuts of $G$. We can use the following convex program to find the best such shortcut matrix.

$$\begin{aligned} \min \quad & \mathcal{E} \\ \text{s.t.} \quad & \mathop{\mathbb{E}}_{e \sim E(S,\overline{S})} \mathcal{R}\text{eff}_D(e) \leq \mathcal{E} \qquad \forall \emptyset \subsetneq S \subsetneq V, \\ & D \preceq_{\square} L_G, \\ & D \succ 0. \end{aligned} \tag{4}$$

Note that if the optimum is small, it means that there are at least $k/2$ good edges in every cut of $G$, so the set $F$ of good edges is $\Omega(k)$-edge-connected and we are done. Unfortunately, even the optimum of the above convex program may be very close to 1 for $\Omega(\log(n))$-edge-connected graphs.

The above impossibility result shows that it is not possible to reduce the average effective resistance of all cuts of $G$. Our approach is to recognize families of subsets of edges for which it is possible to reduce the maximum average effective resistance.

In the first step, we observe that for any partitioning of the vertices of a $k$-edge-connected graph $G$ into $S_1, S_2, \ldots$ we can use a variant of the above convex program to reduce the maximum average effective resistance of the sets

$$E(S_1, \overline{S_1}), E(S_2, \overline{S_2}), \text{ and so on}$$

to $\tilde{O}(1/k)$. Next, we illustrate why this is useful using an example. Later, we will see that our main technical theorem implies a stronger version of this statement.

**Example 14.** *Assume that $G$ is defined as follows: Start with a $k$-regular $\epsilon$-expander on $\sqrt{n}$ vertices and replace each vertex with a cycle of length $\sqrt{n}$ repeated $k$ times where the endpoints of the expander edges incident to each cycle are equidistantly distributed. This graph is $k$-edge-connected by definition and all expander edges have effective resistance close to 1.*

*If we use the $\sqrt{n}$ cycles as our partition, by the above observation, we can reduce the average effective resistance of edges coming out of each cycle to some $\alpha = \tilde{O}(1/k)$. Let $F$ be the union of all of the cycle edges and the expander edges of effective resistance at most $2\alpha/\epsilon$. Now, we show that $F$ is $\Omega(k)$-edge-connected. For any cut that cuts at least one of the cycles, obviously there are at least $k$ cycle edges in $F$. For the rest of the cuts, at least $\epsilon$-fraction of the expander edges incident to the cycles on the small side of the cut cross the cut; among these edges at least half of them are in $F$, so $F$ has at least $\Omega(k)$ edges in the cut.*

We can use the above observation in any $k$-edge-connected graph repeatedly to gradually make $F$ $\Omega(k)$-edge-connected as follows: Start with partitioning into singletons; let $D_1$ be a shortcut matrix that reduces the average effective resistance of degree cuts to $\alpha = \tilde{O}(1/k)$, and let $F_1$ be the edges of effective resistance at most $2\alpha$. In the next step, let the partitioning $S_1, S_2, \ldots$ be a natural decomposition of $(V, F_1)$ into $k/2$-edge-connected components. Similarly, define $D_2$ and let $F_2$ be the edges connecting $S_1, S_2, \ldots$ of effective resistance at most $2\alpha$. This procedure ends in $\ell = O(\log n)$ iterations. It follows that $\cup_{i=1}^{\ell} F_i$ is $\Omega(k)$-edge-connected and the average of shortcut matrices, $\mathbb{E}_i D_i$, is a shortcut matrix that reduces the effective resistance of all edges of $F$ to $O(\ell \cdot \alpha)$. Therefore, if $\ell = \text{polyloglog}(n)$ we are done.

Unfortunately there are $k$-edge-connected graphs where the above procedure ends in $\Theta(\log n)$ steps because each time the size of the partition may reduce only by a factor of 2. Note that this procedure defines a laminar family over the vertices. Let $S_1, S_2, \ldots$ be all of the sets in all partitions; observe that they form a laminar family; let $S_{i^*}$ be the smallest set that is a superset of $S_i$. Also, let $\mathcal{O}(S_i) = E(S_i, S_{i^*} - S_i)$.

Suppose we write a convex program to *simultaneously* reduce the maximum average effective resistance of all $\mathcal{O}(S_i)$'s; then we may obtain a $k$-edge-connected set $F$ of good edges in a single shot. As we will see next, modulo some technical conditions, this is what we prove in our main technical theorem. Such a statement is not enough to get a $k$-edge-connected set of good edges, but it is enough to get $F$ in polyloglog$(n)$ steps.

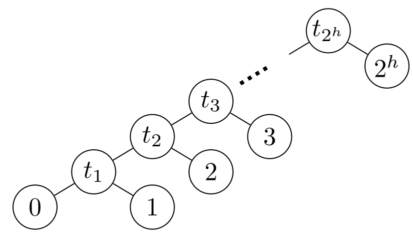


Figure 5: A $\mathcal{T}(k, 1/2, \{1, 2, \ldots, 2^h\})$-locally connected hierarchy of the graph of Figure 4.

*A. Locally Connected Hierarchies*

For a graph $G = (V, E)$, a hierarchy, $\mathcal{T}$, is a tree where every non-leaf node has at least two children and each leaf corresponds to a unique vertex of $G$. We use the terminology *node* to refer to vertices of $\mathcal{T}$. For each node $t \in \mathcal{T}$ let $V(t) \subseteq V$ be the set of vertices of $G$ that are mapped to the leaves of the subtree of $t$, $E(t)$ be the set of edges between the vertices of $V(t)$, and

$$G(t) = G[V(t), E(t)],$$

be the induced subgraph of $G$ on $V(t)$. Let $\mathcal{P}(t) := E(V(t), \overline{V(t)})$ be the set of edges that leave $V(t)$ in $G$. Throughout the paper we use $t^*$ to denote the parent of a node $t$. We define $\mathcal{O}(t) := E(V(t), V(t^*) - V(t))$ as the set of edges that leave $V(t)$ in $G(t^*)$. We abuse notation and use $\mathcal{T}$ to also denote the set of nodes of $\mathcal{T}$.

Let us give a clarifying example. Say $G$ is the "bad" graph of Figure 4. In Figure 5 we give a locally connected hierarchy of $G$. For each node $t_i$, $V(t_i) = \{0, 1, \ldots, i\}$. For each $1 \leq i \leq 2^h$, the set $\mathcal{O}(i)$ is the set of edges from vertex $i$ to all vertices $j$ with $j < i$. In addition, since $t_i$ has exactly two children, $\mathcal{O}(i) = \mathcal{O}(t_{i-1})$. Finally, $\mathcal{P}(i)$ is all edges incident to vertex $i$ and $\mathcal{P}(t_i)$ is the set of edges $E(\{0, 1, \ldots, i\}, \{i+1, \ldots, 2^h\})$.

For an integer $k > 1$, $0 < \lambda < 1$, and $T \subseteq \mathcal{T}$, we say $\mathcal{T}$ is a $(k, \lambda, T)$-locally connected hierarchy of $G$, or $(k, \lambda, T)$-LCH if

1) For each node $t \in \mathcal{T}$, the induced graph $G(t)$ is $k$-edge-connected.
2) For any node $t \in \mathcal{T}$ that is not the root, $|\mathcal{O}(t)| \geq k$. This property follows from 1 because $\mathcal{O}(t) = E(V(t), V(t^*) - V(t))$ is a cut of $G(t^*)$.
3) For any node $t \in T$, $|\mathcal{O}(t)| \geq \lambda \cdot |\mathcal{P}(t)|$. Note that unlike the other two properties, this one only holds for a subset $T$ of the nodes of $\mathcal{T}$.

We say $\mathcal{T}$ is a $(k, \lambda, \mathcal{T})$-LCH if $T$ is the set of all nodes of $\mathcal{T}$. For example, the hierarchy of Figure 5 is a $(k, 1/2, \{1, 2, \ldots, 2^h\})$-LCH of the graph illustrated in Figure 4. Condition 1 holds because there are $k$ parallel edges between any pair of vertices $i-1, i$, so $G(V(t_i))$ is $k$-edge-connected. Condition 2 holds because,

$$|\mathcal{O}(i)| = |\mathcal{O}(t_{i-1})| = |E(\{0, \ldots, i-1\}, \{i\})| \geq k.$$

Lastly, it is easy to see that condition 3 holds for any leaf node $i \in T$, $|\mathcal{O}(i)| \geq d(i)/2 = |\mathcal{P}(i)|/2$.

*Locally Connected Hierarchies and Good Edges:* Let $\mathcal{T}$ be a hierarchy of $G$. Let $t \in \mathcal{T}$ have children $t_1, \ldots, t_j$. Define

$$G\{t\} := G(t)/V(t_1)/V(t_2)/\ldots/V(t_j)$$

to be the graph obtained from $G(t)$ by contracting each $V(t_i)$ into a single vertex. We may call $G\{t\}$ an internal subgraph of $G$. Let $V\{t\}$ be the vertex set of $G\{t\}$; we can also identify this set with the children of $t$ in $\mathcal{T}$. Also, let $E\{t\}$ be the edge set of $V\{t\}$.

The following property of locally connected hierarchies is crucial in our proof. Roughly speaking, if a subset $F$ of edges of $G$ is $k$-edge-connected in each internal subgraph, then it is globally $k$-edge-connected.

**Lemma 15.** *Let $\mathcal{T}$ be a hierarchy of a graph $G = (V, E)$ and $F \subseteq E$. If for any internal node $t$, the subgraph $(V\{t\}, F \cap E\{t\})$ is $k$-edge-connected, then $(V, F)$ is $k$-edge-connected.*

*Proof:* Consider any cut $(S, \overline{S})$ of $G$. Observe that there exists an internal node $t \in \mathcal{T}$ such that $S$ crosses $V(t)$. Let $t_0$ be the deepest such node in $\mathcal{T}$ (root has depth 0). But then,

$$F(S, \overline{S}) \supseteq F(S \cap V(t_0), \overline{S} \cap V(t_0)),$$

and the size of the set on the RHS is at least $k$ by the assumption of the lemma. ■

To prove Theorem 6 we will find a good set of edges which satisfy the assumption of the above lemma. Note that the assumption of the above lemma does not imply that $F$ is dense in $G$. This is crucial because as we said in the previous subsection there is no shortcut matrix $D$ which has a dense set of good edges.

*Construction of an LCH for Planar Graphs:* In this section we give a universal construction of locally connected hierarchies for $k$-edge-connected planar graphs.

**Lemma 16.** *Any $k$-edge-connected planar graph $G = (V, E)$ has a $(k/5, 1/5, T)$-LCH $\mathcal{T}$ where $\mathcal{T}$ is a binary tree, and $T$ contains at least one child of each nonleaf node of $\mathcal{T}$.*

We will use the following fact about planar graphs, whose proof easily follows from the fact that *simple* planar graphs have at least one vertex with degree at most 5.

**Fact 17.** *In any $k$-edge-connected planar graph $G = (V, E)$, there is a pair of vertices $u, v \in V$ with at least $k/5$ parallel edges between them.*

The details of the construction are given in Algorithm 1. Observe that the algorithm terminates after exactly

---

**Algorithm 1** Construction of a locally connected hierarchy for planar graphs.

---

**Input:** A $k$-edge-connected planar graph $G$.
**Output:** A $(k/5, ., .)$-LCH of $G$.

For each vertex $v \in V$, add a unique leaf node to $\mathcal{T}$ and map $v$ to it. Let $W$ be the set of these leaf nodes. ▷ *We keep the invariant that $W$ is the nodes of $\mathcal{T}$ that do not have a parent yet, but their subtree is fixed, i.e., $V(t)$ is well-defined for any $t \in W$.*
**while** $|W| > 1$ **do**
  Add a new node $t^*$ to $W$.
  Let $G_{t^*}$ be the graph where for each node $t \in W$, $V(t)$ is contracted to a single vertex; identify each $t \in W$ with the corresponding contracted vertex. ▷ *Note that $G_{t^*}$ is also a planar graph, because for any $t \in W$, the induced graph $G[V(t)]$ is connected.*
  Let $t_1$ be a vertex with at most 5 neighbors in $G_{t^*}$. ▷ *$t_1$ exists by Fact 17.*
  Let $t_2$ be a neighbor of $t_1$ such that $\{t_1, t_2\}$ has the largest number of parallel edges among all neighbors of $t_1$. ▷ *Note that $t_1, t_2$ are not necessarily vertices of $G$, so parallel edges between them do not correspond to parallel edges of $G$.*
  Make $t^*$ the parent of $t_1, t_2$; remove $t_1, t_2$ from $W$, and add $t_1$ to $T$. ▷ *So, $V(t^*) = V(t_1) \cup V(t_2)$.*
**end while**
**return** $\mathcal{T}$.

---

$n - 1$ iterations of the loop, because any non-leaf node of $\mathcal{T}$ has exactly two children, so $|W|$ decreases by 1 in each iteration. We show that $\mathcal{T}$ is $\mathcal{T}(k/5, 1/5, T)$-LCH. First of all, for any non-leaf node $t$ of $\mathcal{T}$, $G(t)$ is $k/5$-edge-connected. We prove this by induction. Say, $t_1, t_2$ are the two children of $t^*$, and by induction, $G(t_1)$ and $G(t_2)$ are $k/5$-edge-connected. By the selection of $t_2$, there are at least $k/5$ parallel edges between $t_1, t_2$, so $G(t^*)$ is $k/5$-edge-connected. Secondly, we need to show that $\mathcal{O}(t_1) \geq \mathcal{P}(t_1)/5$. This is because by the selection of $t_2$, $1/5$ of the edges incident to $t_1$ in $G_{t^*}$ are $\{t_1, t_2\}$. This completes the proof of Lemma 16.

### *B. Main Technical Theorem*

Given a $(k, \lambda, T)$-LCH $\mathcal{T}$ of $G$, in our main technical theorem we minimize the maximum average effective resistance of $\mathcal{O}(t)$'s among all nodes $t \in T$. The following convex program finds a shortcut matrix $0 \prec D \preceq L_G$ that minimizes the maximum of the average effective resistance of edges in $\mathcal{O}(t)$ for all $t \in T$.

$$
\begin{aligned}
&\textbf{Tree-CP}(\mathcal{T} \in (k, \lambda, T)\text{-LCH}): \\
&\min \quad \mathcal{E} \\
&\text{s.t.} \quad \mathop{\mathbb{E}}_{e \sim \mathcal{O}(t)} \mathcal{R}\text{eff}_D(e) \leq \mathcal{E} \quad \forall t \in T, \\
&\qquad D \preceq_{\square} L_G, \\
&\qquad D \succ 0.
\end{aligned}
$$

**Theorem 18** (Main Technical)**.** *For any $k$-edge-connected graph $G$, and any $\mathcal{T}(k,\lambda,T)$-LCH, $\mathcal{T}$, of $G$, there is a PD shortcut matrix $D$ such that for any $t \in T$,*

$$\mathop{\mathbb{E}}_{e\sim\mathcal{O}(t)} \mathcal{R}\text{eff}_D(e) \le \frac{f_1(k,\lambda)}{k},$$

*where $f_1(k,\lambda)$ is a poly-logarithmic function of $k, 1/\lambda$.*

Note that the statement of the above theorem does not have any dependency on the size of $G$.

If we apply the above theorem to the $(k/5, 1/5, T)$-LCH $\mathcal{T}$ of a $k$-edge-connected planar graph as constructed in Algorithm 1, we obtain a shortcut matrix $D$ for which the small effective resistance edges are $\Omega(k)$-edge-connected. Let us elaborate on this. Let $F = \{e : \mathcal{R}\text{eff}_D(e) \le \frac{2f_1(k/5,1/5)}{k/5}\}$. First, note that by Lemma 16, $\mathcal{T}$ is a binary tree and at least one child of each internal node of $\mathcal{T}$ is in $T$. Say $t$ is an internal node with children $t_1, t_2$ and $t_1 \in T$. Then, by Markov's inequality

$$|F \cap \mathcal{O}(t_1)| \ge |\mathcal{O}(t_1)|/2 \ge \frac{k/5}{2}.$$

Since $t$ has only two children, this implies $G(V\{t\}, F \cap E\{t\})$ is $k/10$-edge-connected. Now, by Lemma 15, $(V,F)$ is $k/10$-edge-connected.

It is natural to expect that for every $k$-edge-connected graph $G$, one can find a locally connected hierarchy $\mathcal{T}$ such that one application of the above theorem produces a set $F$ of good edges such that for any $t \in \mathcal{T}$, $G(V\{t\}, F \cap E\{t\})$ is $\Omega(k)$-edge-connected. By Lemma 15 this would imply $(V,F)$ is $\Omega(k)$-edge-connected. However, the following example shows that this may not be the case.

**Example 19.** *Let $G = (V,E)$ be the $k$-dimensional hypercube ($n = 2^k$). Note that $G$ is $k$-edge-connected. Let $\mathcal{T}$ be a $(\Omega(k), ., .)$-LCH for $G$. Consider an internal node $t_0 \in \mathcal{T}$, all of whose children are leaves. By definition $G(t_0)$ is $\Omega(k)$-edge-connected. Consider a dimension cut of the hypercube that cuts $G(t_0)$ into $(S, V(t_0) - S)$. Imagine a solution $D$ of* Tree-CP($\mathcal{T}$) *which reduces the effective resistance of all edges except those in the cut $(S, V(t_0) - S)$. In such a solution, $\mathbb{E}_{e\sim\mathcal{O}(t)} \mathcal{R}\text{eff}_D(e)$ is small for all $t$. This is because each vertex $v \in G(t)$ has at most one of its $\Omega(k)$ neighboring edges in the cut $(S, V(t_0) - S)$. But note that the small effective resistance edges are disconnected in $G\{t_0\} = G(t_0)$.*

Consider a $(\Omega(k), ., .)$-LCH $\mathcal{T}$ of $G$ and let $t$ be an internal node. Theorem 18 promises that the average effective resistance of all degree cuts of the internal graph $G\{t\}$ are small. If $G\{t\}$ is an *expander* this implies that the good edges are $\Omega(k)$-edge-connected in $G\{t\}$. Therefore, if we can find a locally connected hierarchy whose internal subgraphs are expanding we can find an $\Omega(k)$-edge-connected set of good edges by a single application of Theorem 18. This is exactly what we proved in the case of planar graphs. The above hypercube example shows that such a locally connected hierarchy does not necessarily exist in all $k$-edge-connected graphs.

### *C. Expanding Locally Connected Hierarchies*

In this section we define expanding locally connected hierarchies and we describe our plan to prove Theorem 6 using the main technical theorem.

**Definition 20** (Expanding Locally Connected Hierarchies)**.** *For a node $t$ with children $t_1, \ldots, t_j$ in a locally connected hierarchy $\mathcal{T}$ of a graph $G = (V,E)$, an internal node $t$ (or the internal subgraph $G\{t\}$) is called $(\alpha,\beta)$-expanding, if $G\{t\}$ is an $\alpha$-expander and is $\beta$-edge-connected. A subset of the nodes $T$ is called $(\alpha,\beta)$-expanding iff each one of them is $(\alpha,\beta)$-expanding.*

For example, observe that the locally connected hierarchies that we constructed in Algorithm 1 for $k$-edge-connected planar graphs are $(1, k/5)$-expanding. In the full version of this article we construct an $(\Omega(1/k), \Omega(k))$-expanding $(\Omega(k), \Omega(1), \mathcal{T})$-LCH for any $k$-edge-connected graph where $k \ge 7\log n$. But Example 19 shows that this is essentially the best possible, as the $k$-dimensional hypercube does not have any $(\omega(1/k), \Omega(k))$-expanding locally connected hierarchy.

It follows that if $G$ has an $(\alpha, \Omega(k))$-expanding locally connected hierarchy then there is a shortcut matrix $D$ and an $\Omega(k)$-edge-connected set $F$ of edges such that

$$\max_{e\in F} \mathcal{R}\text{eff}_D(e) \le O(\text{Tree-CP}(\mathcal{T})/\alpha).$$

Recall the argument in Example 14 for details. Since the best $\alpha$ we can hope for is $O(1/\log n)$ this argument by itself does not work.

Our approach is to apply Theorem 18 to an adaptively chosen sequence of locally connected hierarchies. Each time we recognize the internal subgraphs of the locally connected hierarchy in which the set of good edges found so far are not $\Omega(k)$-edge-connected. Then, we apply Theorem 18 to the nodes in these internal subgraphs. We "refine" these internal subgraphs by a natural decomposition of the newly found good edges to get the next locally connected hierarchy. At the heart of the argument we show that this refinement procedure improves the expansion of the aforementioned internal subgraphs by a constant factor. Therefore, this procedure stops after $O(\log(1/\alpha)) = \text{polyloglog}(n)$ steps in the worst case.

We conclude this section by describing an instantiation of the above procedure in the special case of a $k$-dimensional hypercube for demonstration purposes. Let $G$ be a $k$-dimensional hypercube. We let $\mathcal{T}_1$ be a star, i.e., it has only one internal node and the vertices of $G$ are the leaves. This means that in Tree-CP($\mathcal{T}_1$) we minimize the maximum average effective resistance of degree cuts of $G$. Let $F_1$ be the edges of effective resistance at most twice the optimum of Tree-CP($\mathcal{T}_1$). It follows that half the edges incident to each vertex are in $F_1$. Now, we find a natural decomposition of the good edges $F_1$. In the "worst case", edges of $F_1$ form $k/2$ dimensional sub-hypercubes and all edges connecting these sub-hypercubes are not in $F_1$. Note that if we contract these sub-hypercubes, we get a $k/2$-dimensional hypercube which is a $2/k$-expander, twice more expanding than $G$. Of course, we cannot contract, because we need good edges having small effective resistance with respect to the original vertex set, but the expansion is our measure of progress.

In the next iteration, we construct a $(.,.,T_2)$-LCH $\mathcal{T}_2$ where the vertices of each $k/2$-dimensional sub-hypercube are connected to a unique internal node, and the root is connecting all internal nodes, i.e., $\mathcal{T}_2$ has height 2. We let $T_2$ be the set of all internal nodes (except the root). Note that if we delete the leaves, then $\mathcal{T}_2$ would be the same as $\mathcal{T}_1$ for a $k/2$-dimensional sub-hypercube. Similarly, we solve Tree-CP($\mathcal{T}_2$), and in the worst case the new good edges form $k/4$ dimensional sub-hypercubes. Continuing this procedure after $\log(k) = \log\log n$ iterations the good edges span an $\Omega(k)$-edge-connected subset of $G$.

In the next section, we describe the properties of the dual of our convex programs and we elaborate some of the ideas used in the analysis of the dual.

## IV. The Dual

In this section we prove the following proposition.

**Proposition 21.** *For any $k$-edge-connected graph $G = (V, E)$ and any set $F \subseteq E$, there is a PD shortcut matrix $D$ that reduces the average effective resistance of the edges of $F$ to $\tilde{O}(1/k)$.*

Although we do not directly use the above proposition in the proof of our main technical theorem, we do use the main tool of the proof, Lemma 23, as one of the key components of the proof for the main technical theorem. To prove the proposition, we show that the optimum value of the following convex program is at most $\tilde{O}(1/k)$.

$$\begin{aligned} \min \quad & \mathbb{E}_{e\sim F}\mathcal{R}\text{eff}_D(e) \\ \text{s.t.} \quad & D \preceq_{\square} L_G, \\ & D \succ 0. \end{aligned} \tag{5}$$

Before explicitly writing down the dual of the above program, let us give a few lines of intuition. We do this by writing down the dual of a few convex programs computing the maximum or average effective resistance of a number of pairs of vertices.

For a pair of vertices, $a, b \in V$, the optimum value of the following expression,

$$\max_{x:V\to\mathbb{R}} \frac{(x(a)-x(b))^2}{\sum_{u\sim v}(x(u)-x(v))^2}. \tag{6}$$

is exactly equal to $\mathcal{R}\text{eff}_G(a, b)$; in particular, if we fix $x(b) = 0, x(a) = \mathcal{R}\text{eff}(a, b)$, then the optimum $x$ is the *potential* vector of the electrical flow that sends one unit of flow from $a$ to $b$. It is an easy exercise to cast the above as a convex program.

Now, suppose we want to write a program which computes the maximum effective resistance of pairs of vertices $(a_1, b_1), \dots, (a_h, b_h)$. In this case we need to choose a separate potential vector for each pair, We use a matrix $X$

where the $i$-th row of $X$ is the potential vector associated to the $i$-th pair. The following program gives the maximum effective resistance of all pairs.

$$\max_{X\in\mathbb{R}^{h\times V}} \frac{\sum_{i=1}^h (X_{i,a_i}-X_{i,b_i})^2}{\sum_{i=1}^h\sum_{u\sim v}(X_{i,u}-X_{i,v})^2} = \max_{X\in\mathbb{R}^{h\times V}} \frac{\sum_{i=1}^h (X_{i,a_i}-X_{i,b_i})^2}{\sum_{u\sim v}(X_u-X_v)^2}$$

It is a simple exercise that the optimum of the above is the maximum effective resistance of all pairs $(a_1,b_1),\dots,(a_h,b_h)$. Recall that $X_u$ is the $u$-th column of $X$.

Note that the denominator of the RHS is coordinate independent, i.e., it is rotationally invariant. We can rewrite the numerator in the following way and make it rotationally invariant. Instead of mapping the $i$-th pair to the $i$-th coordinate, we map the $i$-th pair to $z_i$ where $\{z_1,\dots,z_h\}$ are $h$-orthonormal vectors. In other words, to calculate the numerator we need to find a coordinate system of the space such that the sum of the square of the projection of the edges on the corresponding coordinates is as large as possible

$$\max_{\substack{X\in\mathbb{R}^{h\times V},\\ \{z_1,\dots,z_h\}\text{ are orthonormal}}} \frac{\sum_{i=1}^h \langle z_i, X_{a_i}-X_{b_i}\rangle^2}{\sum_{u\sim v}(X_u-X_v)^2}.$$

Instead of choosing $z_1,\dots,z_h$ we can simply maximize over an orthogonal matrix $U\in\mathbb{R}^{h\times h}$ and let $z_1,\dots,z_h$ be the first $h$ rows of $U$,

$$\max_{X\in\mathbb{R}^{h\times V},\text{Orthogonal }U} \frac{\sum_{i=1}^h \langle U^i, X_{a_i}-X_{b_i}\rangle^2}{\sum_{u\sim v}(X_u-X_v)^2}, \tag{7}$$

where $U^i$ is the $i$-th row of the matrix $U$. The above program is equivalent to the dual of the following convex program

$$\begin{aligned} \min\quad & \mathcal{E}, \\ \text{s.t.}\quad & \mathcal{R}\text{eff}_D(a_i,b_i)\le\mathcal{E} \quad \forall 1\le i\le h, \\ & D\preceq L_G. \end{aligned}$$

When we replace the constraint $D\preceq L_G$ with $D\preceq_\square L_G$, we get the additional assumption that $X$ is a cut metric. This can significantly reduce the value of (7).

Next, we write a program which computes the expected effective resistance of pairs of vertices $(a_1,b_1),\dots,(a_h,b_h)$ with respect to a distribution $\lambda_1,\dots,\lambda_h$,

$$\sum_{i=1}^h \lambda_i\cdot\mathcal{R}\text{eff}(a_i,b_i) = \max_{X\in\mathbb{R}^{h\times V}} \sum_{i=1}^h \lambda_i\cdot\frac{(X_{i,a_i}-X_{i,b_i})^2}{\sum_{u\sim v}(X_{i,u}-X_{i,v})^2}. \tag{8}$$

where we simply used (6). Equivalently, we can write the above ratio as follows:

$$\max_{X\in\mathbb{R}^{h\times V}} \frac{\left(\sum_{i=1}^h\sqrt{\lambda_i}\cdot(X_{i,a_i}-X_{i,b_i})\right)^2}{\sum_{u\sim v}(X_u-X_v)^2}, \tag{9}$$

To see that the above two are the same, first, assume $X$ is normalized such that $\sum_{u\sim v}(X_{i,a_i}-X_{i,b_i})^2=1$ for all $i$. This simplifies (8) to $\sum_i\lambda_i(X_{i,a_i}-(X_i,b_i))^2$. Then let

$$Y^i = X^i\sqrt{\lambda_i}\cdot(X_{i,a_i}-X_{i,b_i}),$$

where as usual $Y^i$ is the $i$-th row of $Y$. Plugging in $Y$ in (9) gives the same value $\sum_i\lambda_i(X_{i,a_i}-X_{i,b_i})^2$.

Now, we are ready to write down the dual of (5). This is the rotationally invariant variant of (9). Recall that $\mathbf{X}=X\mathcal{X}$ is the matrix where for every edge $e=(u,v)$, $\mathbf{X}_e=X_u-X_v$.

$$\max_{\substack{X\in\{0,1\}^{h\times V},\\ \text{semiorthogonal }U\in\mathbb{R}^{E\times h}}} \frac{\frac{1}{|F|}\left(\sum_{e\in F}\langle U^e,\mathbf{X}_e\rangle\right)^2}{\sum_{e\in E}\|\mathbf{X}_e\|^2} = \max_{\substack{X\in\{0,1\}^{h\times V}\\ \text{semiorthogonal }U\in\mathbb{R}^{E\times h}}} \frac{(\mathbb{E}_{e\sim F}\langle U^e,\mathbf{X}_e\rangle)^2}{\frac{1}{|F|}\sum_{e\in E}\|\mathbf{X}_e\|^2}. \tag{10}$$

To prove Proposition 21 we need to upper bound the above ratio by $\tilde{O}(1/k)$.

First, let us describe how $k$-edge-connectivity blends into our proof. In the following simple fact we show that to lower bound the denominator it is enough to find many disjoint $L_1$ balls centered at the vertices of $G$ with large radii.

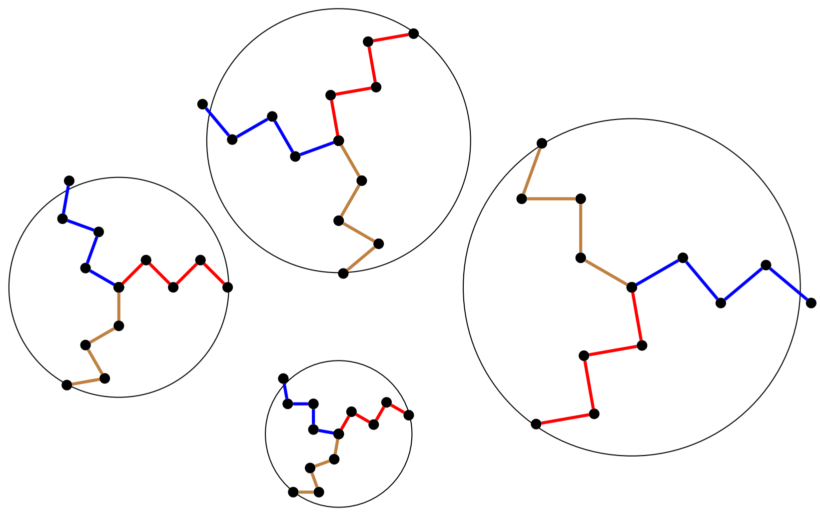

Figure 6: Sets of $k$ edge-disjoint paths in disjoint $L_1$ balls.

**Fact 22.** *For any $X : V \to \{0,1\}^h$ and any set of $\ell \geq 2$ disjoint $L_1$ balls $B_1, \dots, B_\ell$ centered at vertices of $G$ with radii $r_1, \dots, r_\ell$ we have*

$$\sum_{i=1}^{\ell} r_i \cdot k \leq \sum_{e \in E} \|\mathbf{X}_e\|_1 = \sum_{e \in E} \|\mathbf{X}_e\|^2 .$$

Since there are $k$ edge-disjoint paths connecting the center of each ball to the outside, by the triangle inequality, the sum of the $L_1$ length of the edges of the graph is at least $k$ times the sum of the radii of the balls. Note that if $\ell = 1$, i.e., if we have only one ball, the conclusion does not necessarily hold. This is because $B_1$ may contain all vertices of $G$.

Fix $X \in \{0,1\}^{h \times V}$ and a semiorthogonal matrix $U \in \mathbb{R}^{E \times h}$. It is enough to show that $\frac{(\mathbb{E}_{e\sim F}\langle U^e, \mathbf{X}_e\rangle)^2}{\frac{1}{|F|}\sum_{e\in E}\|\mathbf{X}_e\|^2} \leq \tilde{O}(1/k)$. Let $Y = UX$ and $\mathbf{Y} = Y\mathcal{X} = UX\mathcal{X}$. Note that since $U$ is semiorthogonal, $\|\mathbf{Y}_e\|^2 \leq \|\mathbf{X}_e\|^2$ for all $e$. Without loss of generality assume that

$$\frac{(\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e})^2}{\mathbb{E}_{e\sim F}\|\mathbf{Y}_e\|^2} \geq \alpha,$$

for $\alpha = \text{polylog}(k)/k$; otherwise we are done. In the following lemma we show that assuming the above inequality we can construct $b$ disjoint $L_2^2$ balls of radius $r$ centered at the vertices of the endpoints of edges of $F$ such that

$$r \cdot b \geq \frac{\alpha^\epsilon}{\text{poly}(\epsilon)} \cdot (\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e})^2 \, |F|.$$

On the other hand, since these balls are disjoint, by Fact 22,

$$r \cdot b \leq \frac{1}{k}\sum_{e\in E}\|\mathbf{X}_e\|^2 .$$

Note that we really need to apply Fact 22 to balls in the space of $X_v$'s, since $Y_v$'s do not necessarily satisfy the triangle inequality. However, given disjoint balls centered around $Y_v$'s, one can take the same balls around the corresponding $X_v$'s and they will remain disjoint, since $U$, the mapping from $X_v$ to $Y_v$, is a contraction.

Now, Proposition 21 simply follows by the above two inequalities for $\epsilon = \log k / \log\log k$.

**Lemma 23.** *Given $F \subseteq E$ and a mapping $Y \in \mathbb{R}^{E\times V}$ such that*

$$\Upsilon := \left(\mathop{\mathbb{E}}_{e\sim F}\mathbf{Y}_{e,e}\right)^2 \geq \alpha \cdot \mathop{\mathbb{E}}_{e\sim F}\|\mathbf{Y}_e\|_2^2 , \tag{11}$$

*for some $\alpha > 0$, for any $0 < \epsilon \leq 1$, there are $b$ disjoint $L_2^2$ balls $B_1, \dots, B_b$ with radius $r$ such that the center of each ball is an endpoint of an edge in $F$, $b \geq \alpha|F|/C_1(\epsilon)$, and*

$$r \cdot b \geq \frac{\alpha^\epsilon \cdot \Upsilon \cdot |F|}{C_1(\epsilon)},$$

*where $C_1(\epsilon)$ is a polynomial function of $1/\epsilon$.*

Before getting to the proof of the lemma, let us give an intuitive description of the statement of the lemma. The extreme case is for $\alpha \approx 1$. Observe that the inequality (11) enforces a very strong assumption on the mapping $\mathbf{Y}$. Since for any edge $e$, $\mathbf{Y}_{e,e} \leq \|\mathbf{Y}_e\|$, and $\alpha \approx 1$, the following two conditions must hold for $\mathbf{Y}$:

i) For most edges $e \in F$, $\mathbf{Y}_{e,e} \approx \|\mathbf{Y}_e\|$,

ii) For most pairs of edges $e, f \in F$, $\|\mathbf{Y}_e\| \approx \|\mathbf{Y}_f\|$.

The above two conditions essentially imply that the vectors $\{\mathbf{Y}_e\}_{e \in F}$ form an orthonormal basis up to normalizing the size of the vectors. It is an exercise to see that in this case one can select $\Omega(|F|)$ many $L_2^2$ balls of radius $\Omega(\Upsilon)$ around the endpoints of the edges in $F$; one can show that greedily picking balls that do not intersect each other works.

Our proof can be interpreted as a robust version of this argument. Due to limited space, we provide a proof of this lemma for the case $\epsilon = 1$; for a complete proof refer to [33].

*Proof of Lemma 23.* For a radius $r > 0$, run the following greedy algorithm. Scan the endpoints of the edges in an arbitrary order; for each point $Y_u$, if the $L_2^2$ ball $B(Y_u, r)$ doesn't touch the balls that we have already selected, select $B(Y_u, r)$. Suppose we manage to select $b$ balls. We say the algorithm succeeds if both of the lemma's conclusions are satisfied. In the rest of the proof we show that this algorithm always succeeds for some value of $r$.

Without loss of generality, in the rest of the proof we drop the columns of $\mathbf{Y}$ corresponding to edges $e \notin F$ and their corresponding rows and we assume $Y \in \mathbb{R}^{F \times F}$. Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{|F|}$ be the singular values of $\mathbf{Y}$. We can rewrite the assumption of the lemma as follows:

$$\left(\frac{1}{|F|}\sum_i \sigma_i\right)^2 \geq \left(\frac{\mathrm{Tr}(\mathbf{Y})}{|F|}\right)^2 = (\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e})^2 \geq \alpha \cdot \mathbb{E}_{e\sim F}\|\mathbf{Y}_e\|^2 = \frac{\alpha}{|F|}\|\mathbf{Y}\|_F^2 = \frac{\alpha}{|F|}\sum_{i=1}^{|F|}\sigma_i^2. \tag{12}$$

The first inequality follows by Lemma 10. Note that, for $\alpha = 1$, the LHS is always less than or equal to the RHS by the Cauchy-Schwarz inequality with equality happening only when $\sigma_1 = \dots = \sigma_{|F|}$. So, for large $\alpha$ the above inequality can be seen as a reverse Cauchy-Schwarz inequality.

In the next claim, we show that if the above algorithm finds a "small number" $b$ of balls for a choice of $r$, this means that $\sigma_b, \dots, \sigma_{|F|}$ are significantly smaller than $\sigma_1, \dots, \sigma_{b-1}$. In the succeeding claim we use the above reverse Cauchy-Schwarz inequality to show that this is impossible.

**Claim 24.** *Given $r > 0$, suppose that the greedy algorithm finds $b$ disjoint balls of radius $r$. Then*

$$r \geq \frac{1}{16|F|}\sum_{i=b}^{|F|}\sigma_i^2.$$

*Proof:* We construct a low-rank matrix $C \in \mathbb{R}^{F \times F}$. Then, we use Theorem 11 to prove the claim. Let $Y_{w_1}, \dots, Y_{w_b}$ be the centers of the chosen balls. Then, for any endpoint $v$ of an edge in $F$, let $c(v)$ be the closest center to $Y_v$, i.e.,

$$c(v) := \mathrm{argmin}_{w_i}\|Y_{w_i} - Y_v\|_2^2$$

We construct a matrix $C \in \mathbb{R}^{F \times F}$ such that the $e$-th column of $C$ is defined as follows: say the $\{u, v\}$-th column of $\mathbf{Y}$ is $Y_u - Y_v$ for $\{u, v\} \in F$; we let the $\{u, v\}$-th column of $C$ be $Y_{c(u)} - Y_{c(v)}$. By definition, $\mathrm{rank}(C) \leq b - 1$, since $C$'s columns are a subset of the differences between $b$ points.

First, notice that

$$\begin{aligned}
\|\mathbf{Y} - C\|_F^2 &= \sum_{\{u,v\}\in F}\left\|(Y_u - Y_v) - (Y_{c(u)} - Y_{c(v)})\right\|_2^2 \\
&\leq \sum_{\{u,v\}\in F}(\left\|Y_u - Y_{c(u)}\right\|_2 + \left\|Y_v - Y_{c(v)}\right\|_2)^2 \\
&\leq \sum_{\{u,v\}\in F} 2\left\|Y_u - Y_{c(u)}\right\|_2^2 + 2\left\|Y_v - Y_{c(v)}\right\|_2^2 \leq 16r \cdot |F|,
\end{aligned}$$

where the first inequality follows by the triangle inequality and the last inequality follows by the definition of greedy algorithm; in particular, for any point $v$, in the worst case there is a point $p$ in the $L_2^2$ ball about $c(v)$ such that $\|p - Y_v\|^2 < r$, so

$$(\|Y_v - Y_{c(v)}\|_2)^2 \leq (\|Y_v - p\| + \|Y_{c(v)} - p\|)^2 \leq (\sqrt{r} + \sqrt{r})^2 \leq 4r.$$

Now by Theorem 11,

$$16r \cdot |F| \geq \|\mathbf{Y} - C\|_F^2 \geq \sum_{i=b}^{|F|} \sigma_i^2.$$

where the second inequality uses the fact that $\text{rank}(C) \leq b - 1$. ■

All we need to show is that there is a value of $b \gtrsim \alpha|F|$ such that $\frac{b}{16|F|} \sum_{i=b}^{|F|} \sigma_i^2 \gtrsim \alpha\Upsilon \cdot |F|$.

**Claim 25.** *There is an integer $b \gtrsim \alpha|F|$ such that*

$$\frac{b}{16|F|} \sum_{i=b}^{|F|} \sigma_i^2 \gtrsim \alpha \cdot \Upsilon \cdot |F|.$$

*Proof:* We show the claim holds for $b = \alpha|F|/4$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left(\frac{1}{|F|}\sum_{i=1}^{|F|} \sigma_i\right)^2 &\leq 2\left(\frac{1}{|F|}\sum_{i=1}^{b-1} \sigma_i\right)^2 + 2\left(\frac{1}{|F|}\sum_{i=b}^{|F|} \sigma_i\right)^2 \\
&\leq \frac{2b}{|F|^2}\sum_{i=1}^{b-1} \sigma_i^2 + \frac{2}{|F|}\sum_{i=b}^{|F|} \sigma_i^2 \\
&= \frac{\alpha}{2|F|}\sum_{i=1}^{b-1} \sigma_i^2 + \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|} \sigma_i^2 \leq \frac{1}{2}\left(\frac{1}{|F|}\sum_{i=1}^{|F|} \sigma_i\right)^2 + \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|} \sigma_i^2,
\end{aligned}$$

where the equality uses the definition of $b$ and the last inequality uses (12). Therefore,

$$\frac{\Upsilon}{2} \leq \frac{1}{2}\left(\frac{1}{|F|}\sum_{i=1}^{|F|} \sigma_i\right)^2 \leq \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|} \sigma_i^2,$$

where the first inequality uses another application of (12). ■

Observe that the above claim implies Lemma 23. It is sufficient to run the greedy algorithm with the infimum value of $r$ such that the greedy algorithm returns at most $b$ balls. ■


## Acknowledgment.

We would like to thank James R Lee, Satish Rao, Amin Saberi, Luca Trevisan, and Umesh Vazirani for enlightening and stimulating discussions. We also would like to specially thank Lap Chi Lau for proofreading and extensive comments on earlier versions of this manuscript.


## References


[1] A. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem," 2013. [Online]. Available: http://arxiv.org/abs/1306.3969 1, 3, 5, 7

[2] N. Anari and S. Oveis Gharan, "The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP," 2014. [Online]. Available: http://arxiv.org/abs/1412.1143 1, 3, 5, 7, 8

[3] N. Christofides, "Worst case analysis of a new heuristic for the traveling salesman problem," Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA, Report 388, 1976. 1

[4] M. Held and R. Karp, "The traveling salesman problem and minimum spanning trees," *Operations Research*, vol. 18, pp. 1138–1162, 1970. 1

[5] M. Charikar, M. X. Goemans, and H. Karloff, "On the integrality ratio for the asymmetric traveling salesman problem," *MOR*, vol. 31, no. 2, pp. 245–252, 2006. 1

[6] A. Asadpour, M. X. Goemans, A. Madry, S. Oveis Gharan, and A. Saberi, "An $O(\log n/ \log\log n)$-approximation Algorithm for the Asymmetric Traveling Salesman Problem," in *SODA*, 2010, pp. 379–389. 1, 2

[7] A. M. Frieze, G. Galbiati, and F. Maffioli, "On the worst-case performance of some algorithms for the asymmetric traveling salesman problem," *Networks*, vol. 12, pp. 23–39, 1982. 1

[8] M. Bläser, "A new approximation algorithm for the asymmetric tsp with triangle inequality," in *SODA*, 2002, pp. 638–645. 1

[9] H. Kaplan, M. Lewenstein, N. Shafrir, and M. Sviridenko, "Approximation algorithms for asymmetric TSP by decomposing directed regular multigraphs," *J. ACM*, pp. 602–626, 2005. 1

[10] U. Feige and M. Singh, "Improved approximation ratios for traveling salesman tours and path sin directed graphs," in *APPROX*, 2007, pp. 104–118. 1

[11] S. Oveis Gharan and A. Saberi, "The asymmetric traveling salesman problem on graphs with bounded genus," in *SODA*, 2011, pp. 967–975. 1, 2, 3

[12] J. Erickson and A. Sidiropoulos, "A near-optimal approximation algorithm for asymmetric TSP on embedded graphs," in *SOCG*, 2014, p. 130. 1, 3

[13] M. X. Goemans, N. J. A. Harvey, K. Jain, and M. Singh, "A randomized rounding algorithm for the asymmetric traveling salesman problem," 2009. [Online]. Available: http://arxiv.org/abs/0909.0941 2

[14] P. Raghavan and C. D. Tompson, "Randomized rounding: a technique for provably good algorithms and algorithmic proofs," *Combinatorica*, vol. 7, no. 4, pp. 365–374, Dec. 1987. 2

[15] L. A. Goddyn, "Some open problems i like," 2004, available at *http://www.math.sfu.ca/ goddyn/Problems/problems.html*. 3

[16] F. Jaeger, "On circular flows in graphs, finite and infinite sets," *Colloquia Mathematica Societatis Janos Bolyai 37*, pp. 391–402, 1984. 3

[17] C. Thomassen, "The weak 3-flow conjecture and the weak circular flow conjecture," *Journal of Combinatorial Theory, Series B*, vol. 102, no. 2, pp. 521–529, 2012. 3

[18] L. M. Lovász, C. Thomassen, Y. Wu, and C. Zhang, "Nowhere-zero 3-flows and modulo k-orientations," *J. Comb. Theory, Ser. B*, vol. 103, no. 5, pp. 587–598, 2013. 3

[19] N. J. A. Harvey and N. Olver, "Pipage rounding, pessimistic estimators and matrix concentration," in *SODA*, 2014, pp. 926–945. 3, 4, 5

[20] N. Weaver, "The kadison-singer problem in discrepancy theory," *Discrete Mathematics*, vol. 278, no. 1-3, pp. 227–239, 2004. 3

[21] M. Rudelson, "Random vectors in the isotropic position," *Journal of Functional Analysis*, vol. 164, no. 1, pp. 60–72, 1999. 3

[22] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *Information Theory, IEEE Transactions on*, vol. 48, no. 3, pp. 569–579, 2002. 3

[23] J. Borcea, P. Branden, and T. M. Liggett, "Negative dependence and the geometry of polynomials." *Journal of American Mathematical Society*, vol. 22, pp. 521–567, 2009. 3

[24] S. Oveis Gharan, A. Saberi, and M. Singh, "A randomized rounding approach to the traveling salesman problem," in *FOCS*, 2011, pp. 550–559. 3

[25] O. Svensson, "Approximating atsp by relaxing connectivity," 2015, to appear in FOCS. [Online]. Available: http://arxiv.org/abs/1502.02051 3

[26] D. A. Spielman and S.-H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *STOC*, 2004, pp. 81–90. 4

[27] D. A. Spielman and N. Srivastava, "Graph sparsification by effective resistances," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1913–1926, 2011. 4

[28] J. D. Batson, D. A. Spielman, and N. Srivastava, "Twice-ramanujan sparsifiers," *SIAM Review*, vol. 56, no. 2, pp. 315–334, 2014. 4

[29] W. S. Fung, R. Hariharan, N. J. A. Harvey, and D. Panigrahi, "A general framework for graph sparsification," in *STOC*, 2011, pp. 71–80. 4

[30] R. Lyons and Y. Peres, *Probability on Trees and Networks.* Cambridge University Press, 2013, to appear. 4

[31] C. Akemann and N. Weaver, "A lyapunov type theorem from kadison-singer," 2013. [Online]. Available: http://arxiv.org/abs/1308.5276 6, 7

[32] A. Ghosh, S. Boyd, and A. Saberi, "Minimizing effective resistance of a graph," *SIAM Rev.*, vol. 50, no. 1, pp. 37–66, Feb. 2008. 8

[33] N. Anari and S. Oveis Gharan, "Effective-Resistance-Reducing Flows, Spectrally Thin Trees, and Asymmetric TSP ," 2014. [Online]. Available: http://arxiv.org/abs/1411.4613 8, 18

[34] C. S. J. A. Nash-Williams, "Edge disjoint spanning trees of finite graphs," *J. London Math. Soc.*, vol. 36, pp. 445–45, 1961. 10